Dysarthria Classification Using Acoustic Properties of Fricatives

Abner Hernandez*, Minhwa Chung**
*, **Seoul National University
*abner@snu.ac.kr, **mchung@snu.ac.kr

ABSTRACT

Speech is an essential mode of communication for many people. However, degenerative disorders like Parkinson's or ALS along with other neurological disorders (cerebral palsy) can greatly affect speech production. This study will examine the effectiveness of using acoustic measurements from fricatives as features for automatic classification of disordered speech, specifically dysarthria. Results show that a support vector machine (SVM) with fricative duration and spectral moments as features is able to classify dysarthric speech from healthy speech with an accuracy of 82%.

1. Introduction

Dysarthria is group of motor speech disorders resulting from neurological damage to the articulatory muscles that help produce speech. Individuals suffering from dysarthria can have impairments in respiration, phonation, resonance, prosody and articulation. Currently, speech pathologists perform subjective intelligibility assessments which can be costly and time consuming. Automatic classification of dysarthria can aid speech pathologist in their assessments by utilizing a data driven approach to distinguish healthy speech from dysarthric speech. Previous studies on impaired speech classification have used a variety of acoustic features such as Mel Frequency Cepstral Coefficients (MFCC), prosodic features (F0), voice quality features (jitter, shimmer) and glottal features [1, 3].

Our study will examine the effectiveness of fricatives as features to a classifier. Few studies have looked specifically at fricatives which have been found to be the most commonly mispronounced consonant class along with liquids and affricates [2]. Therefore, we hypothesize that fricatives may be a useful consonant class for dysarthric speech identification. Figure 1. shows a comparison between the fricative /s/ in the word 'said'. From the spectrogram we can see the dysarthric speech has a more irregular spectrum than healthy speech. There is no standard method of measuring fricatives, but spectral moments (mean spectral peak, variance, skewness, kurtosis) are commonly used in research with English fricatives [4]. The mean spectral peak refers to the frequency which divides the spectrum in a way that the top-half frequencies are equal to the low-half frequencies. Variation will tell us whether most energy is concentrated in a small band or dispersed over a wide range of frequencies. Skewness will measure the shape of the spectrum below the mean peak compared to the frequencies above the mean peak. Finally, kurtosis describes the peakness of energy distribution. A positive kurtosis suggests that spectral peaks are well defined while a negative kurtosis suggest a spectrum with a flat distribution.

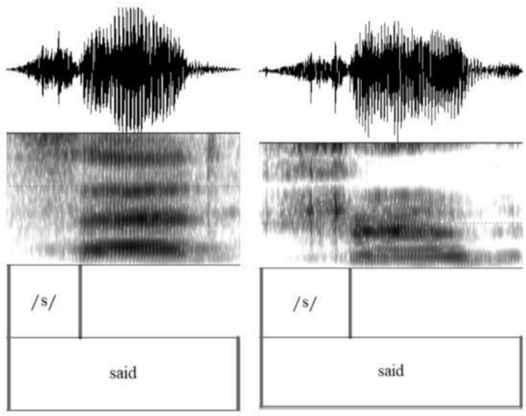

Figure 1. Coronal sibilant /s/ from a healthy speaker (left) and a dysarthric speaker (right).

2. Methods

Using the UA-Speech database we extracted fricative features from 10 dysarthric speakers with cerebral palsy along with 9 healthy speakers. The spectral moments were measured in a 20ms. hamming window at the centre of the fricative. Three different fricative initial words from the following sibilants were used: /s, z, ʃ/. Six of the 9 words were repeated three times leading to a total of 21 utterances per speaker. Each word is represented as a vector with five fricative measurements. These vectors are then used as input by a binary support vector machine classifier (SVM). Support vector machines have previously been used in impaired speech classification [1,5] and provide consistent performance even with small datasets. The goal of an SVM is to find the optimal hyperplane which maximally separates the data points of two or more classes. We specifically used a radial basis SVM with a margin parameter C and Gaussian kernel parameter γ. These parameters were determined through a grid search where C and γ vary from 10^{-3} to 10^{2}.

Table 1. Speaker for training and test set. CM refers to control speakers while M refers to dysarthric speakers.

Training set	Test set
CM01, CM04, CM08, CM09, CM10, CM12, M01(15%), M05(58%), M08(93%), M09(86%), M10(93%), M12(7.4%), M16(43%)	CM05, CM06, CM13, M07(28%), M11(62%), M14(90.4%),

To avoid overfitting, we split our data in way that no speaker in the test set was present in the training set. Table 1. displays how we split our data into a training and test sets with intelligibly rating in parenthesis. It is important to note that dysarthric speakers have intelligibility ratings ranging from 7.4% to 93%. Therefore, we split our data to have balanced averaged intelligibility ratings (train: 56%, test: 60%). Information regarding the database and intelligibly ratings can be found in [6]. We conducted three experiments: only spectral moments (SM), only duration (Dur), both spectral moments and duration (SM+Dur).

Table 1. Results of fricative-based dysarthria classifier.

Experiment	Accuracy	Precision	Recall
SM	72%	71%	76%
Dur	66%	63%	74%
SM+Dur	**82%**	**80%**	**84%**

3. Results

Results from table 1. show that fricative spectral moments together with duration perform the best. These results are promising in showing that phoneme level measurements as opposed to full word or sentence analysis can be used to achieve relatively high classification accuracy. Our classifier was able to solely use fricative measurements as features and classify dysarthric speech from healthy speech with some success. As can be seen from the confusion matrix in figure 2. while our model is able to make some accurate predictions it still contains some false positives and false negatives. Future research would benefit from taking into account different severity levels and examining a larger set of words with fricatives. It is possible that false negatives and positives arise when highly intelligible dysarthric speakers have measurements close to healthy speakers.

4. Conclusion

The present study has shown that phoneme level measurements can be used in classifying dysarthric speech. Mean spectral peak, variance, skewness, kurtosis and duration were used in a C-support vector classifier and achieved an accuracy of 82%, precision rate of 80% and recall of 84%. While these percentages are not as high as other studies, our study uses less data, only phoneme level measurements and minimal features to classify dysarthric speech. The current study is step towards exploring alternative speech features for classification when data is limited.

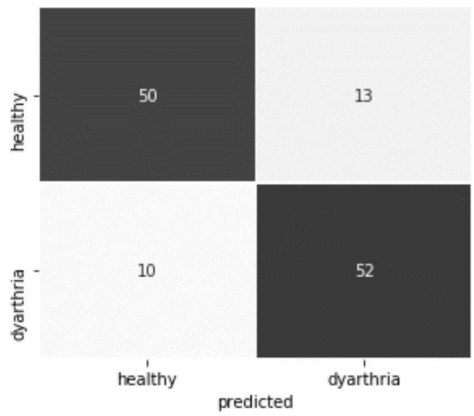

Figure 2. Prediction results for SVM classifier using fricative duration and spectral moments.

References

[1] Kim, J., Kumar, N., Tsiartas, A., Li, M., & Narayanan, S. S. (2015). Automatic intelligibility classification of sentence-level pathological speech. *Computer speech & language*, *29*(1), 132-144.

[2] Kim, H., Martin, K., Hasegawa-Johnson, M., & Perlman, A. (2010). Frequency of consonant articulation errors in dysarthric speech. Clinical linguistics & phonetics 24 (10), 759-770.

[3] Narendra, N. P., & Alku, P. (2018). Dysarthric Speech Classification Using Glottal Features Computed from Non-words, Words and Sentences. In *Interspeech* (pp. 3403-3407).

[4] Yoon, T. J. (2018). A corpus-based study on the effects of voicing and gender on American English Fricatives. *Phonetics and Speech Sciences*, *10*(2), 7-14.

[5] Orozco-Arroyave, J. R., Hönig, F., Arias -Londoño, J. D., Vargas-Bonilla, J. F., Skodda, S., Rusz, J., & Nöth, E. (2014). Automatic detection of Parkinson's disease from words uttered in three different languages. In *INTERSPEECH-2014*, 1573-1577.

[6] Kim, H., Hasegawa-Johnson, M., Perlman, A., Gunderson, J., Huang, T. S., Watkin, K., & Frame, S. (2008). Dysarthric speech database for universal access research. In *Ninth Annual Conference of the International Speech Communication Association*. (pp. 1741-1744).