INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Extended Study on the Use of Vocal Tract Variables to Quantify Neuromotor Coordination in Depression

Nadee Seneviratne[1], *James R. Williamson*[2], *Adam C. Lammert*[3], *Thomas F. Quatieri*[2], *Carol Espy-Wilson*[1]

[1]University of Maryland - College Park
[2]MIT Lincoln Laboratory
[3]Worcester Polytechnic Institute

nadee@umd.edu, jrw@ll.mit.edu, alammert@wpi.edu, quatieri@ll.mit.edu, espy@umd.edu

Abstract

Changes in speech production that occur as a result of psychomotor slowing, a key feature of Major Depressive Disorder (MDD), are used to non-invasively diagnose MDD. In previous work using data from seven subjects, we showed that using speech-inverted vocal tract variables (TVs) as a direct measure of articulation to quantify changes in the way speech is produced when depressed relative to being not depressed outperforms formant information as a proxy for articulatory information. In this paper, we made significant extensions by using more subjects, taking into account more eigenvalue features and incorporating TVs related to (1) place of articulation and (2) the glottal source. These additions result in a significant improvement in accuracy, particularly for free speech. As a baseline, we perform a similar analysis using higher-dimensional Mel Frequency Cepstral Coefficients (MFCCs).

Index Terms: speech production, vocal tract variables, psychomotor slowing, neuromotor coordination, depression, mental health, glottal

1. Introduction

Major Depressive Disorder (MDD), also known as clinical depression, is a mental health disorder that can be characterized by long-lasting depressed mood (sadness or hopelessness) or loss of interest in activities that will cause significant impairment in daily life. Around 264 million people worldwide suffer from MDD [1]. Depression is one of the most common precursors leading to suicidality which is the second leading cause of death in youth in the United States between 10 and 34 years of age [2]. Most of the previous work on depression classification and severity prediction focused on prosodic, source, and spectral features [3, 4, 5]. The work presented in this paper explores the possibility of improving the performance of depression detection task using articulatory representations of speech.

Psychomotor slowing is identified as a major characteristic of depression [6, 7]. Currently, it is viewed as a necessary feature of MDD and a key component in assessing and monitoring the severity of depression [8, 9, 10]. Effects of psychomotor slowing observed in speech include more and longer pauses, slowed responses and monotonic phrases [11]. The motivation for quantifying the articulatory coordination comes largely from these effects. These articulatory coordination features can be used to characterize the level of articulatory coordination and timing. To measure the coordination, assessments of the multi-scale structure of correlations among the time series signals were used [12, 13, 14]. This was extensively done using acoustic features consisting of the first three resonances of the vocal tract (formants). However this approach has been less extensively validated using direct articulatory speech features.

In a preliminary study by the authors [15], the use of speech-inverted vocal tract variables (TVs) as a direct measure of articulation to quantify changes in the way speech is produced by depressed and non-depressed subjects was explored. The TVs are based on Articulatory Phonology (AP) [16], which views speech as a constellation of overlapping gestures, and are defined by the constriction degree and location of five distinct constrictors (lips, tongue tip, tongue body, velum, and glottis) along the vocal tract. We used the Mundt database [17] for the experiments. In this pilot study, we used the eigenspectrum features computed from the corresponding time-delay embedded correlation matrices based on a subset of TVs to perform depression classification. Using only seven subjects, we showed that the coordination features computed over three TVs corresponding to constriction degree outperform those of three formants in classifying depressed vs. not depressed speech. For formants, accuracies of 57.1% and 42.9% were observed for read and free speech, respectively. For TVs, the respective accuracies were 64.3% and 71.43%. It was observed that the articulators of depressed speech have less complex coordination associated with more coupled movements which results in reduced variability (coarticulation and lenition) and high intelligible speech.

In this paper, we have extended the preliminary study by (1) including results from a more complete set of TVs (adding constriction location TVs and glottal TV), (2) using data from additional subjects in the Mundt database and (3) using a wider range of eigenspectrum features as inputs to the classification model. We show that including the location TVs further improves the accuracy of the classifier (77.22% for RS and 75.71% for FS). By incorporating periodicity and aperiodicity measures to represent the glottal TV, a significant accuracy improvement was observed for FS (81.77%).

In Section 2, we explain the dataset, the estimation of the TVs, computation of the coordination features, and the details of the classification experiments. Section 3 presents the results of classification experiments and graphical illustrations of coordination features. Finally, in Section 4 we interpret these results in detail and discuss the possible future directions.

2. Method

2.1. Dataset Description

For this study, we used a subset of the Mundt Database [17] which contains speech samples collected over a period of six weeks from thirty five physician-referred patients. The patients started on pharmacotherapy and/or psychotherapy treat-

http://dx.doi.org/10.21437/Interspeech.2020-2758

ment for depression close to the beginning of the study. The speech recordings were collected using interactive voice response (IVR) technology. Speech data collected through this study include read speech (the Grandfather passage) and spontaneous speech where patients describe how they feel emotionally, physically and their ability to function in each week. In addition to this, other elicited voice measures include sustained vowels (for 5 seconds), counting from 1 to 20, reciting the English Alphabet, and /pa-ta-ka/ repeated rapidly for 5 seconds.

We used the clinician-reported Hamilton Depression Rating Scale (HAMD) score to choose subjects for the depressed and non-depressed classes with a balanced distribution. In the case of read speech, we chose all speech when subjects are depressed (HAMD$\geq$20) and all speech when subjects are not depressed (HAMD$\leq$ 7). In the case of free speech, we used the same HAMD thresholds, but selected only those utterances that are less than 30 sec in duration for depressed speech to obtain a balanced distribution of two classes. For free speech (total of 26 subjects), there were 51 utterances for depressed speech and 66 utterances for non-depressed speech. For read speech (total of 30 subjects), there were 33 and 20 utterances for depressed and non-depressed speech, respectively. Note that in the preliminary study [15], we used only 7 utterances (from 7 subjects) for each class for both read and free speech.

2.2. Acoustic-to-Articulatory Speech Inversion (SI)

A speaker independent, DNN based SI system is used to compute the Vocal Tract Variables (TVs) that represent constriction location and degree of articulators located along the vocal tract [18, 19].

The model was trained using the Wisconsin X-Ray Microbeam (XRMB) database [20]. The XRMB recordings originally comprise of naturally spoken utterances along with XRMB cinematography of the mid-sagittal plane of the vocal tract with pellets placed at points along the vocal tract. The trajectory data are recorded for the individual articulators: Upper Lip, Lower Lip, Tongue Tip, Tongue Blade, Tongue Dorsum, Tongue Root, Lower Front Tooth (Mandible Incisor), Lower Back Tooth (Mandible Molar). We call these trajectories as pellet trajectories. The X-Y positions of the pellets are closely tied to the anatomy of the speakers. The quantification of the vocal tract shape is better performed by the location and the degree of these constrictions based on relative measures as opposed to the X-Y positions of the pellets. The TVs specify the salient features of the vocal tract area function more directly than the pellet trajectories [21] and are relatively speaker independent. Hence, the pellet trajectories were converted to TV trajectories using geometric transformations as outlined in [22] to define a corpus of 'ground truth' TV trajectories. The six TVs obtained from the seven pellet trajectories were – Lip Aperture (LA), Lip Protrusion (LP), Tongue Body Constriction Location (TBCL), Tongue Body Constriction Degree (TBCD), Tongue Tip Constriction Location (TTCL) and, Tongue Tip Constriction Degree (TTCD).

2.3. Glottal TV Estimation

Descriptions of speech articulation in Articulatory Phonology typically include TVs related to the glottal state. Due to the difficulty in acquiring ground-truth glottal TV data by placing the sensors near the glottis, the DNN based SI system could not be trained to estimate the glottal TVs. As an alternative to this, we used the periodicity and aperiodicity measure obtained from the Aperiodicity, Periodicity and Pitch (APP) detector developed in [23]. This program estimates the proportion of periodic energy and aperiodic energy in a speech signal along with the pitch period for the periodic component. This uses a time domain approach and is based on the distribution of the minima of the average magnitude difference function (AMDF) of the speech signal:

$$\gamma_n^k = \sum_{m=-\infty}^{\infty} |x(n+m)w(m) - x(n+m-k)w(m-k)| \quad (1)$$

where $x(n)$ is the input signal, $w(m)$ is a 20-ms rectangular window and k is the lag value, which varies from 0 to the sample value equivalent of 20 ms (eg., for the sampling rate of 16kHz, k will have the range of [0,320]).

2.4. Mel-Frequency Cepstral Coefficients (MFCCs) Estimation

We used higher-dimensional MFCCs as a proxy for actual articulatory features instead of formants as used in the preliminary study, to enable fair comparisons with the higher dimensional TV data. For this, 12 MFCC time series were extracted by using an analysis window of 20 ms with a 10 ms frame shift (1st MFCC coefficient was discarded).

2.5. Coordination Features

The correlation structure features [12] were used to estimate the coordination among three sets of time series data: 6 TVs (constriction location and degrees), 8 TVs (adding the glottal TVs to 6 TVs) and 12 MFCCs. For each speech signal, a channel-delay correlation matrix is computed from low-level multi-channel signals (TVs or MFCCs in this case), using a time-delay embedding at a constant delay scale (7 samples). The sampling rate of TVs and MFCCs was 100Hz, therefore the delay scale of 7 samples introduced delays to the signals in 70 ms increments. This correlation matrix is computed as an intermediate representation of the complexity of speech coordination. This compact representation provides more detail about which time series signal is correlated with which, and at which time delays, and is therefore rich with information about the mechanisms underlying the coordination level. Each correlation matrix R_j has dimensionality (MN x MN), based on $M = 6$, 8 or 12 channels and $N = 15$ time delays per channel. A rank ordered eigenspectrum is computed from the correlation matrix R_j , taking the form of an MN-dimensional (90-,120- or 180-dimensional) feature vector. The rank ordering is in descending order, such that the rank 1 eigenvalue is the largest and the rank MN eigenvalue is the smallest.

These time-delay embedded articulatory coordination features are useful in capturing the information related to temporal dynamics of multivariate time series data and can easily be extended to any number of channels.

Another interpretation to these eigenvalues which supports the above hypothesis can be found in [24]. The amplitude of each eigenvalue is proportional to the amount of correlation in the direction of their associated eigenvectors and the sum of the eigenvalues will remain constant. Additionally, depressed speech has few eigenvalues with significant magnitudes. Therefore, depressed speech can be represented using a few independent dimensions implying that there is less complexity associated with articulatory coordination and more coupled movements. In non-depressed speech, given that the magnitude of the

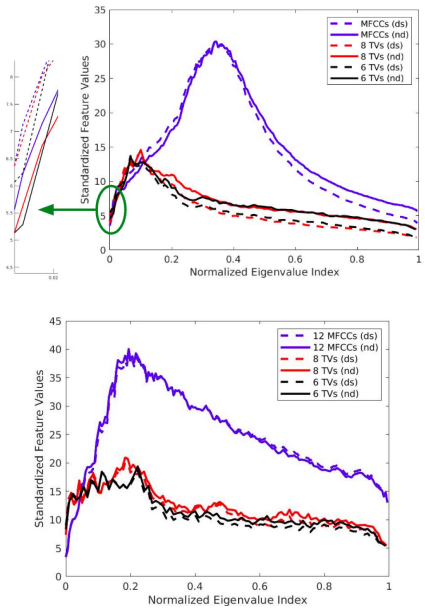


Figure 1: *Standardized feature values of coordination features in the not-depressed speech samples relative to those in the depressed speech samples (Free Speech (top), Read Speech (bottom))*

high-rank eigenvalues are higher, it can be thought of as more complex articulatory coordination that is associated with a large number of independent dimensions.

2.6. Classification Experiments

Experiments were conducted to understand how well these coordination features, computed over MFCCs and TVs could be used to train a model to classify depressed vs. not depressed speech. The features were individually standardized (i.e., z-scored) across all instances prior to model training and testing. In order to utilize more eigenspectrum features, instead of using two points in the spectrum like we did in the preliminary study, we averaged the eigenspectrum features in different index ranges to obtain a low-dimensional representation of the high dimensional eigenspectrum feature vector. Model training and testing were carried out within a leave-one-subject-out cross-validation scheme. When N number of subjects were present, at each fold, a Support Vector Machine (SVM) classifier was trained on data samples of $N-1$ subjects and used as the basis for estimating a label for the test utterances from the remaining subject. Classification accuracy of these estimated labels was calculated across all folds.

3. Results

We plotted the eigenspectrum features that are associated with depressed and non-depressed speech samples for the three cases we analyzed (see Figure 1). For visualization purposes we use the standardized feature-wise means as a function of the normalized eigenvalue feature index $(j-1)/MN$. For a given feature index j, the values of the curves plotted in Figure 1 were calculated according to:

$$\varepsilon_j = \frac{\mu_j^{\gamma}}{s_j} \tag{2}$$

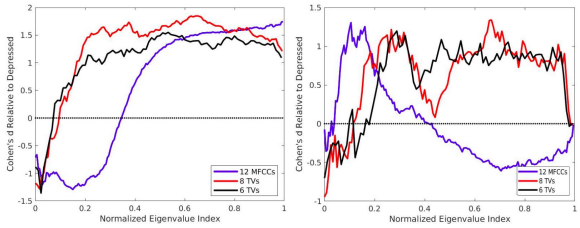


Figure 2: *Effect sizes between the feature-wise means (Cohen's d) of coordination features in the not-depressed speech samples relative to those in the depressed speech samples (Free Speech (top), Read Speech (bottom))*

where μ_j^{γ} is the mean feature value, $1/n_{\gamma}\sum_{i\in\gamma}\lambda_{i,j}$, for all samples taken in the state $\gamma \in$ not depressed (nd), depressed (ds) The quantity s_j is the pooled standard deviation, defined as:

$$s_j = \sqrt{\frac{(n_{ds}-1)s_j^{ds} + (n_{nd}-1)s_j^{nd}}{n_{ds}+n_{nd}-2}} \tag{3}$$

where s_j^{ds} and s_j^{nd} are variances in depressed and not-depressed classes respectively. Eigenspectrum features are shown for read and free speech, and for 6 TVs, 8 TVs, and 12 MFCCs. The magnitudes of low-rank eigenvalues for depressed speech are higher relative to non-depressed speech and the trend is reversed towards the high-ranked eigenvalues as explained in section 2.5.

The effect sizes relative to the depressed state can be computed by the Cohen's d equation:

$$d_j = \frac{\mu_j^{nd} - \mu_j^{ds}}{s_j} \tag{4}$$

The Cohen's d plots given in Figure 2 show the discrimination between the depressed and the non-depressed classes for each set of features and can be quantified using the largest magnitude and mean absolute magnitudes of Cohen's d values as shown is Table 1.

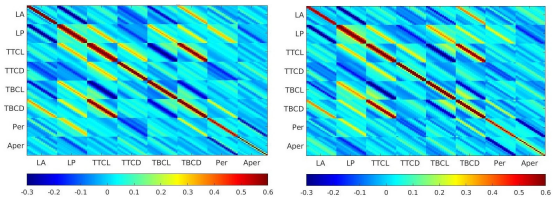
Table 1: *Largest magnitudes and mean absolute magnitudes for Cohen's d values, across all features for free speech (FS) and read speech (RS).*

Feature Set	Max (FS)	Mean (FS)	Max (RS)	Mean (RS)
6 TVs	1.48	1.03	1.19	0.71
8 TVs	1.85	1.42	1.34	0.74
MFCCs	1.75	1.21	1.30	0.44

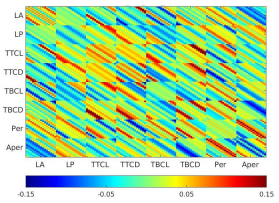
The accuracy results obtained for leave-one-subject-out cross-validation training procedure are included in the Table 2.

Table 2: *Classification accuracies (%) and index ranges over which the averages were calculated to obtain features for classification experiments.*

	6 TVs	8 TVs	MFCCs
RS Accuracy	**77.22**	**77.5**	**72.77**
Index Range	$\leq$ 0.68, [0.68-0.76] $\geq$ 0.76	$\leq$ 0.16, [0.16-0.6], $\geq$ 0.6	$\leq$ 0.08, [0.08-0.18], $\geq$ 0.18
FS Accuracy	**75.71**	**81.77**	**81.70**
Index Range	$\leq$ 0.46, [0.46-0.76], $\geq$ 0.76	$\leq$ 0.3, [0.3-0.47], $\geq$ 0.47	$\leq$ 0.58, $\geq$ 0.58

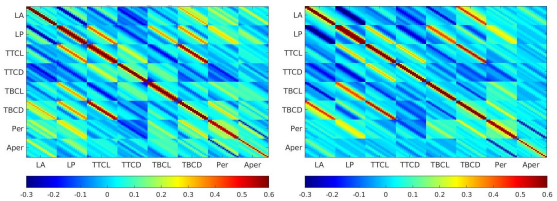


(a) *HAMD=24* (b) *HAMD=7*

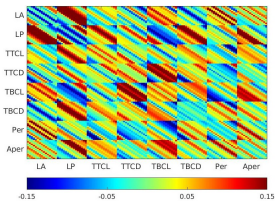


(c) *DS-ND*

Figure 3: *Time-Delay correlation matrix comparison for Read Speech – Subject 127 – from the Mundt database.*



(a) *HAMD=24* (b) *HAMD=7*



(c) *DS-ND*

Figure 4: *Averaged Time-Delay correlation matrix comparison for Free Speech – Subject 127 – from the Mundt database.*

4. Discussion

The low rank eigenvalues being larger for high HAMD session (high depression) relative to the low HAMD session and the trends being reversed towards the high rank eigenvalues is a signature observation associated with depression severity. The dimensionality of the time-delay embedded feature space can be indicated by the magnitude of high rank eigenvalues. Thus, larger values in the high rank eigenvalues indicate greater complexity of articulatory coordination [12]. This trend (low-rank eigenvalues being larger for the depressed class) is held in general for all three cases (except in the case of MFCCs based read speech eigenspectra). It can be seen that by adding the glottal TVs, TV based articulatory coordination features can achieve an accuracy of 81.77% which is about a 8% relative improvement compared to the best accuracy obtained using only constriction degree and location TVs.

Even though we observed comparable results for the MFCC based free speech depression classification, the read speech classification results underperform those obtained for the TVs. We can see that the corresponding standardized read speech eigenspectrum for MFCCs does not hold the general trend of coordination features and the deviation of non-depressed class eigenvalues relative to the depressed class is relatively low which might have caused the degradation in accuracy results. The MFCC based results show that feature dimensionality alone may not be helpful in improving the classification performance and TVs include better discriminative information in representing articulatory coordination in depressed speech with both free speech and read speech. Since depression results in changes in speech production and given our approach is articulatory based, we believe it is easier to understand what changes may be occurring when a person is depressed.

According to the Cohen's D plots, the discrimination between the depressed and the non-depressed classes are maximum (Table 1) in the case of 8 TVs (i.e. when the glottal TV is included) and hence a higher accuracy for both free speech and read speech. The APP detector based glottal TVs seem to provide additional source information related to differentiating articulatory coordination in depressed and non-depressed speech. These glottal measures are also indicative of the breathiness (aperiodic energy in the higher frequencies) of the speech signal. Therefore it is worthwhile to investigate if increased breathiness is a characteristic of the depressed speech in the future. These observations are inline with the results presented in [25].

In Figures 3 and 4, we show the TV based correlation matrices (includes all 8 TVs) corresponding to a single subject (127) in a depressed and non-depressed state, along with the difference matrix of the two correlation matrices. For read speech, a single speech sample is used and for free speech, the average across multiple files is considered. The difference plots indicate that there are relatively higher auto- and cross- correlations present among the TVs in the depressed state compared to the not-depressed state in both the cases. This is inline with our hypothesis of having simpler coordination when a subject is depressed. We also observe that there is considerably more correlation for free speech relative to read speech. This can be another reason for higher classification accuracies observed for free speech. Therefore, free speech can be useful in providing a better representation of the neuromotor coordination involved during speech production due to the increased cognitive load associated with it.

In our future work, we will explore if the TVs can be complemented by other speech features such as dynamic temporal features of TVs (velocity and acceleration) and pause related features, which may be helpful to increase the performance of the depression assessment models. We also plan to explore classifiers that combine MFCCs and TVs. We will extend the use of TV based articulatory coordination features in predicting the depression severity scores.

5. Distribution Statement & Disclaimer

Approved for public release. Distribution is unlimited. This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15- D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

6. Acknowledgements

This work was supported in part by a seed grant between the University of Maryland Medical School and the University of Maryland College Park. We also thank Dr. James Mundt for the depression database [17].

7. References

[1] World Health Organization (WHO). (2020) Depression. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression

[2] National Center for Injury Prevention and Control - Centers for Disease Control and Prevention -"WISQARS". (2018) 10 leading causes of death by age group, united states - 2018. [Online]. Available: {https://www.cdc.gov/injury/images/lc-charts/leading_causes_of_death_by_age_group_2018_1100w850h.jpg}

[3] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. F. Quatieri, "A review of depression and suicide risk assessment using speech analysis," *Speech Communication*, vol. 71, pp. 10 – 49, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167639315000369

[4] S. Scherer, G. Stratou, M. Mahmoud, J. Boberg, J. Gratch, A. Rizzo, and L. Morency, "Automatic behavior descriptors for psychological disorder analysis," in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 4 2013, pp. 1–8.

[5] N. Cummins, J. Epps, V. Sethu, M. Breakspear, and R. Goecke, "Modeling spectral variability for the classification of depressed speech," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, pp. 857–861, 01 2013.

[6] J. R. Whitwell, *Historical notes on psychiatry*. Oxford, England, 1937.

[7] G. Zilboorg, *A History of Medical Psychology*. W W Norton & Co., 1944.

[8] American Psychiatric Association, *Copyright*. Washington, DC, 2000. [Online]. Available: https://dsm.psychiatryonline.org/doi/abs/10.5555/appi.books.9780890425596.x00pre

[9] D. J. Widlöcher, "Psychomotor retardation: Clinical, theoretical, and psychometric aspects," *Psychiatric Clinics of North America*, vol. 6, no. 1, pp. 27 – 40, 1983, recent Advances in the Diagnosis and Treatment of Affective Disorders. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0193953X18308384

[10] C. B. Greden J.F., "Psychomotor function in affective disorders: an overview of new monitoring techniques," *The American Journal of Psychiatry*, vol. 131(11), pp. 1441–8, 1981.

[11] C. Sobin and H. Sackeim, "Psychomotor symptoms of depression," *The American journal of psychiatry*, vol. 154, pp. 4–17, 02 1997.

[12] J. R. Williamson, D. Young, A. A. Nierenberg, J. Niemi, B. S. Helfer, and T. F. Quatieri, "Tracking depression severity from audio and video based on speech articulatory coordination," *Computer Speech & Language*, vol. 55, pp. 40 – 56, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0885230817303510

[13] J. R. Williamson, T. F. Quatieri, B. S. Helfer, G. Ciccarelli, and D. D. Mehta, "Vocal and facial biomarkers of depression based on motor incoordination and timing," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, ser. AVEC '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 65–72. [Online]. Available: https://doi.org/10.1145/2661806.2661809

[14] J. R. Williamson, T. F. Quatieri, B. S. Helfer, R. Horwitz, B. Yu, and D. D. Mehta, "Vocal biomarkers of depression based on motor incoordination," in *Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge*, ser. AVEC '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 41–48. [Online]. Available: https://doi.org/10.1145/2512530.2512531

[15] C. Espy-Wilson, A. C. Lammert, N. Seneviratne, and T. F. Quatieri, "Assessing Neuromotor Coordination in Depression Using Inverted Vocal Tract Variables," in *Proc. Interspeech 2019*, 2019, pp. 1448–1452. [Online]. Available: http://dx.doi.org/10.21437/Interspeech.2019-1815

[16] C. P. Browman and L. Goldstein, "Articulatory Phonology : An Overview *," *Phonetica*, vol. 49, pp. 155–180, 1992.

[17] J. C. Mundt, P. J. Snyder, M. S. Cannizzaro, K. Chappie, and D. S. Geralts, "Voice acoustic measures of depression severity and treatment response collected via interactive voice response (ivr) technology," *Journal of Neurolinguistics*, vol. 20, no. 1, pp. 50 – 64, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0911604406000303

[18] G. Sivaraman, V. Mitra, H. Nam, M. Tiede, and C. Espy-Wilson, "Unsupervised speaker adaptation for speaker independent acoustic to articulatory speech inversion," *The Journal of the Acoustical Society of America*, vol. 146, no. 1, pp. 316–329, 2019. [Online]. Available: https://doi.org/10.1121/1.5116130

[19] G. Sivaraman, V. Mitra, H. Nam, M. K. Tiede, and C. Y. Espy-Wilson, "Vocal tract length normalization for speaker independent acoustic-to-articulatory speech inversion," in *Proceedings of Interspeech*, 2016, pp. 455–459. [Online]. Available: https://doi.org/10.21437/Interspeech.2016-1399

[20] J. R. Westbury, "Speech Production Database User ' S Handbook," *IEEE Personal Communications - IEEE Pers. Commun.*, vol. 0, no. June, 1994.

[21] R. S. McGowan, "Recovering articulatory movement from formant frequency trajectories using task dynamics and a genetic algorithm: Preliminary model tests," *Speech Communication*, vol. 14, no. 1, pp. 19–48, 1994.

[22] H. Nam, V. Mitra, M. Tiede, M. Hasegawa-Johnson, C. Espy-Wilson, E. Saltzman, and L. Goldstein, "A procedure for estimating gestural scores from speech acoustics," *The Journal of the Acoustical Society of America*, vol. 132, no. 6, pp. 3980–3989, 2012. [Online]. Available: https://doi.org/10.1121/1.4763545

[23] O. Deshmukh, C. Y. Espy-Wilson, A. Salomon, and J. Singh, "Use of temporal information: detection of periodicity, aperiodicity, and pitch in speech," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 776–786, 9 2005.

[24] K. Schindler, H. Leung, C. E. Elger, and K. Lehnertz, "Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG," *Brain*, vol. 130, no. 1, pp. 65–77, 11 2006. [Online]. Available: https://doi.org/10.1093/brain/awl304

[25] S. Sahu and C. Espy-Wilson, "Effects of depression on speech," *The Journal of the Acoustical Society of America*, vol. 136, pp. 2312–2312, 10 2014.