Inverted Vocal Tract Variables and Facial Action Units to Quantify Neuromotor Coordination in Schizophrenia

Yashish Maduwantha H P E R S[1], Chris Kitchen[2], Deanna L. Kelly[2], and Carol Espy-Wilson[1]

[1] *University of Maryland College park*
[2] *University of Maryland School of Medicine*

Schizophrenia is a chronic mental disorder with heterogeneous presentations that can be seen mostly among the adult population. Symptoms of schizophrenia are broadly categorized as positive (e.g., hallucinations), negative (e.g. blunted affect) and cognitive (deficits in attention) [1][2]. Previous studies have found that individuals suffering from major depressive disorder (MDD) are subjected to neurophysiological changes that alter motor control, affecting mechanisms controlling speech production and facial expressions. Clinically, these changes are associated with psychomotor retardation, a condition of slowed neuromotor output causing a reduction in speaking rate, decreased movement and impaired cognitive functions [3]. Previous studies have shown promising results in identifying the severity of depression by using coordination features based on the correlation structure of the movements of various articulators [4]. This motivated us to investigate how neuromotor coordination is altered in schizophrenic patients who are markedly ill and exhibit positive symptoms by analyzing speech gestures and associated facial muscle movements.

A database collected by the University of Maryland School of Medicine and the University of Maryland College Park was used for this study. It contains video and audio data of four clinical interviews per subject, collected with 23 Schizophrenia patients, 18 MDD patients and 20 healthy controls over a six-week period. All of the MDD and schizophrenia patients were clinically diagnosed. Each interview session is 25-35 minutes long and every subject is assessed using standard depression severity measures and global psycopathology measures by a clinician. For this study we used the the 18-item Brief Psychiatric Rating Scale (BPRS) [5], where we took into account the total score and the group scores for psychosis, thought disorder and activation, and the Hamilton Rating Scale for Depression (HAMD) to select subjects. The results in this study are based on 6 Schizophrenic subjects, 6 Healthy controls and 3 MDD subjects.

We used a speech inversion (SI) system [6] [7] that maps the acoustic signal into vocal tract variables (TVs). TVs define the constriction degree and location of five distinct constrictors located along the vocal tract: lips, tongue tip, tongue body, velum, and glottis. The SI system estimates six TVs: Lip Aperture, Lip Protrusion, Tongue tip constriction degree, Tongue tip constriction location, Tongue body constriction degree and Tongue body constriction location. Coordination among the six TVs were estimated using the correlation structure features. These features are estimated by computing a channel delay correlation matrix using time delay embedding at a fixed delay scale [4][8]. Each correlation matrix has a dimensionality (MN × MN), based on M TV channels and N time delays per channel.

The video-based Facial Action Units (FAUs) provide a formalized method for identifying changes in facial expressions. We used the Openface 2.0: Facial Behaviour Analysis toolkit [9] to extract seventeen FAUs from the recorded videos of the subjects during the interviews.

From the correlation matrix R_i calculated for each sample i, the eigenspectrum (eigenvalues rank ordered in the descending order of magnitude) is computed. The eigenspectrum generated can be considered as a high level feature designed to characterize properties of coordination and timing from the low level features [8]. The eigenspectrum characterizes the within-channel and cross-channel distributional properties of the multivariate TV time series. The magnitude of the eigenvalues represent the average correlation in the direction of corresponding eigenvectors. Therefore the significance of the magnitude of eigenvalues indicate the number of independent dimensions that can be used to represent speech belonging to different groups. Therefore, a few significant eigenvalues imply a simpler articulatory coordination pattern whereas a large number of significant eigenvalues correspond to more complex articulatory coordination.

Figure 1 shows that the low rank eigenvalues are larger for MDD subjects relative to the schizophrenic patients and the healthy controls, and this trend is reversed towards the high rank eigenvalues. This pattern is a key observation associated with depression severity [10][8][4]. The magnitude of high rank eigenvalues

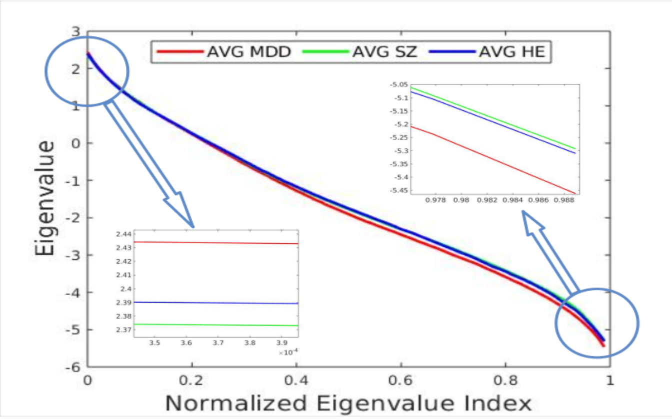

Figure 1: Averaged eigenspectrum for TVs

Table 1: Classification Accuracy (%)

	Index range	**Accuracy (%)**
FAUs	[0-0.02],[0.96-1]	65.63
TVs	[0-0.03],[0.95-1]	61.68

indicates the dimensionality of the time-delay embedded feature space. Thus, larger values in the high rank eigenvalues can be associated with greater complexity of articulatory coordination[4]. Thus, we can conclude that the schizophrenic subjects have a higher complexity than the healthy controls and the MDD patients, and the MDD patients have a simpler coordination relative to the healthy controls and the schizophrenic patients. These results are likely due to the negative symptoms of depression which results in psychomotor slowing (i.e., simpler coordination) and the strong positive symptoms of the schizophrenic patients such as activation that results in motor hyperactivity (i.e., complex coordination). We see this effect/trend in both the eigenspectra computed from FAUs and TVs.

To check the significance of the coordination features in identifying patients with schizophrenia, the 6 Schizophrenic subjects and the 6 Healthy controls were chosen to train a Support Vector Machine (SVM) classifier with radial basis function kernel. The eigenvalues computed are normalized across the two classes (schizophrenia and healthy) to generate normalized eigenspectra and the eigenvalues are averaged over multiple index ranges to be used as the input features to the classifiers. The features calculated are standardized across all instances before model training and testing. The data from all the 12 subjects were used to model a leave-one-subject-out-cross-validation scheme with a total of 12 folds. Table 1 shows the average accuracies obtained for the classification experiments.

[1] N. C. Andreasen and S. Olsen, "Negative v Positive Schizophrenia: Definition and Validation," *Archives of General Psychiatry*, vol. 39, no. 7, pp. 789–794, 07 1982.

[2] C. Demily and N. Franck, "Cognitive remediation: a promising tool for the treatment of schizophrenia," *Expert Review of Neurotherapeutics*, vol. 8, no. 7, pp. 1029–1036, 2008.

[3] J. S. Buyukdura, S. M. McClintock, and P. E. Croarkin, "Psychomotor retardation in depression: biological underpinnings, measurement, and treatment," *Prog. Neuropsychopharmacol. Biol. Psychiatry*, vol. 35, no. 2, pp. 395–409, Mar 2011.

[4] 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.

[5] https://www.smchealth.org/sites/main/files/file-attachments/bprsform.pdf?1497977629.

[6] G. Sivaraman, V. Mitra, H. Nam, M. K. Tiede, and C. Espy-Wilson, "Vocal tract length normalization for speaker independent acoustic-to-articulatory speech inversion," in *Proc. of Interspeech*, 2016, pp. 455–459.

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

[8] 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.

[9] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, 2018.

[10] 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," ser. AVEC '14, 2014, pp. 65–72.