

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

Yashish Maduwantha H.P.E.R.S¹, Chris Kitchen², Deanna L. Kelly², and Carol Espy-Wilson¹

¹University of Maryland College park, ²University of Maryland School of Medicine

1. INTRODUCTION

- 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. hallucination, delusions)
 - Negative (e.g. blunted effect, alogia)
 - Cognitive (e.g. disorganized thinking, slow thinking) [1][2]
- As with previous studies in Major Depressive Disorder (MDD), neurophysiological changes may affect speech production and facial movements [3][4]
- These changes may be captured by coordination features based on the correlation structure of the movements of various articulators [5]
- In this study, we investigate how neuromotor coordination is altered in schizophrenic patients who are **markedly ill** and exhibit strong **positive symptoms**

2. DATASET

- A database recently collected for a collaborative observational study

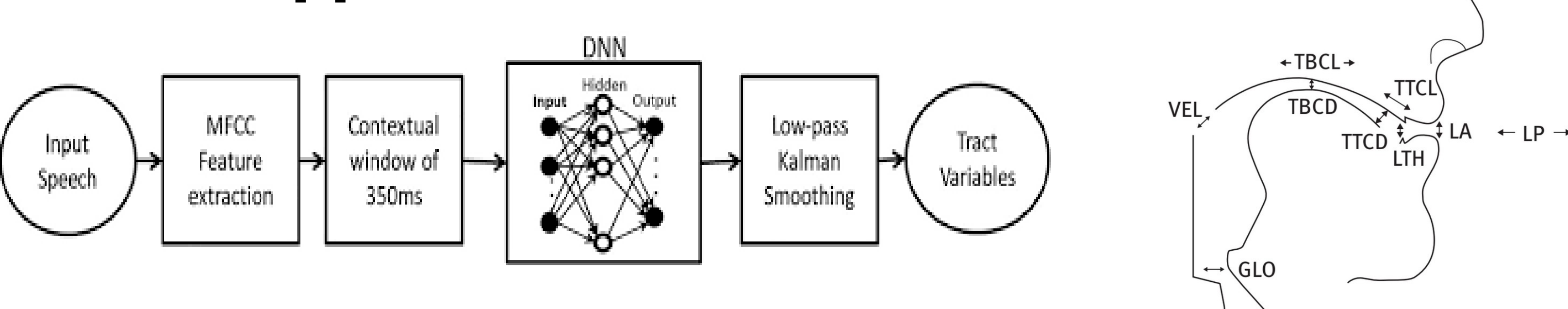
Longitudinal	5 weeks
Number of Subjects	31 M, 30 F
Demography	26 African American, 28 Caucasian, 5 Asian
Assessment	HDRS, MADRS, BPRS, CAPE-42 (Weeks 1,3,5)
Recording Type	Video and Audio
Session Length	10-50 mins
Medications	Antipsychotics, Antidepressants, Anticholinergics, Antihypertensives

- Details of the subject data used for the study

	SZ	HC	MDD
Number of Subjects	6	6	3
BPRS score range	45<score<=62	18<score<=23	18<score<=23
HAMD score range	0<score<14	0<score<7	20<=score
Mean session duration	35 min	18 min	38 min

3. SPEECH INVERSION SYSTEM AND VOCAL TRACT VARIABLES

- We used a Speech inversion (SI) system [7] [8] based on Articulatory Phonology that maps the acoustic signal into vocal tract variables (TVs)
 - Feed forward Network trained on Wisconsin X-ray Microbeam Database [9]



Constricted Organ	Tract Variable	Articulators
Lip	Lip Aperture (LA) Lip Protrusion (LP)	Upper Lip, Lower Lip, Jaw
Tongue Body	Tongue body constriction degree (TBCL) Tongue body constriction location (TTCL)	Tongue Body, Jaw
Tongue Tip	Tongue tip constriction degree (TTCD) Tongue tip constriction location (TTCL)	Tongue Body, Tip, Jaw
Velum	Velum (VEL)	Velum
Glottis	Glottis (GLO)	Glottis

4. FACIAL ACTION UNITS (FAUs)

- Videos of patients were used with the Openface 2.0 Facial Behavior Analysis toolkit [11] to extract seventeen FAUs

FAU No	FAU Name	FAU No	FAU Name
1	Inner brow raiser	14	Dimpler
2	Outer brow raiser	15	Lip corner depressor
4	Brow raiser	17	Chin raiser
5	Upper lid raiser	20	Lip stretcher
6	Cheek raiser	23	Lip tightner
7	Lid tightner	25	Lips part
9	Nose wrinkler	26	Jaw drop
10	Upper lip raiser	45	Blink
12	Lip corner puller		

5. TIME DELAY EMBEDDED CORRELATION FEATURES

- Correlation structure features estimated by a channel delay correlation matrix at a fixed delay scale (TVs : 7/100 = 70 ms and FAUs : 3/28 ~ 107ms)
 - Step 1** : Computation of the Channel delay correlation matrix
 - Time series signal shifted by multiples of the fixed delay scale
 - Auto and cross correlations are computed among shifted time series signals (Each Correlation matrix R_j has dimensionality of $MN \times MN$ where M is the number of channels and N is the number of time delays per channel)
 - Step 2** : Computation of the Eigenspectrum
 - Eigenvalues of the Correlation matrix are rank ordered in the descending order of magnitude
 - Few significant eigenvalues imply a simpler articulatory coordination pattern. Large number of significant eigenvalues correspond to more complex articulatory coordination [6][10][12]

6. ANALYSIS OF COORDINATION FEATURES

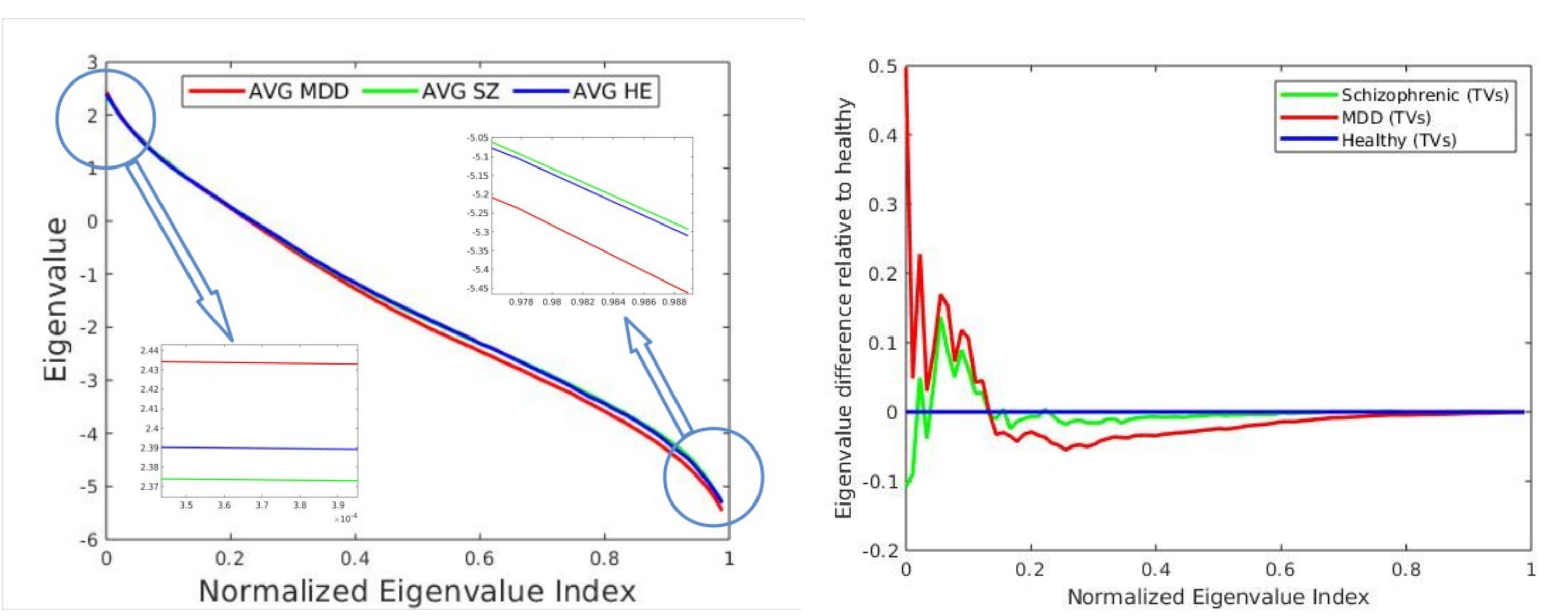


Figure 1: Averaged eigenspectrum for TVs (left) and corresponding difference plot (right).

- Large number of significant eigenvalues for schizophrenia suggests complex articulatory coordination.
- Few significant eigenvalues for MDD suggests simpler articulatory coordination

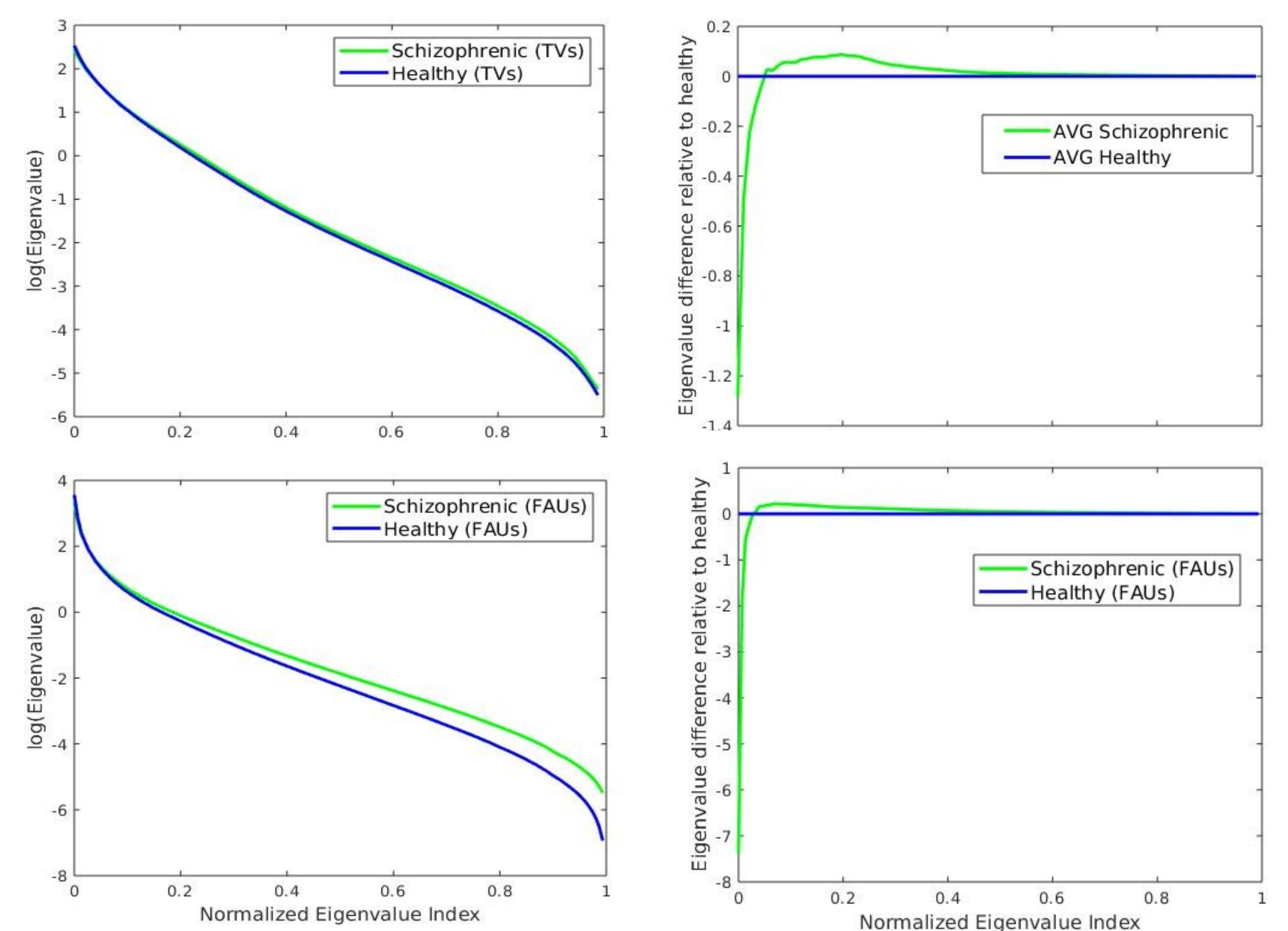


Figure 2: Averaged eigenspectra for TVs and FAUs (left) and corresponding difference plots (right) for classification experiments

7. CLASSIFICATION RESULTS

- Binary classification on subjects with schizophrenia and healthy controls (6 subjects for each group)
- Averaged across ranges from the Normalized eigenspectra [6] as low-dimensional features
- SVM classifier with RBF kernel trained on Leave-one-subject-out cross validation scheme
- Multi-modal fusion of TVs and FAUs with a stacking ensemble classifier

Method	Index range	Accuracy (%)
FAUs	[0-0.02],[0.96-1]	65.63
TVs	[0-0.03],[0.95-1]	61.68
Multi-modal	-	68.19

9. REFERENCES

[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] J. Girard, J. Cohn, M. Mahoor, S. Mavadati, and D. Rosenwald, "Social risk and depression: Evidence from manual and automatic facial expression analysis," 04 2013, pp. 1-8.

[5] C. Espy-Wilson, A. C. Lammer, N. Seneviratne, and T. F. Quatieri, "Assessing Neuromotor Coordination in Depression Using Inverted Vocal Tract Variables," in *Proc. Interspeech 2019*, 2019, pp. 1448-1452.

[6] <https://www.smchealth.org/sites/main/files/file-attachments/bprform.pdf?1497977629>.

[7] 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.

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

[9] 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, Jul 2019.

[10] 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.

[11] 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.

[12] 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.