DISTINCT PROSODIC CORRELATES FOR NINE DIMENSIONS OF MENTAL HEALTH SYMPTOMS

Maria K. Wolters[1], Alex S. Cohen[2], and Kristin Nicodemus[1]

University of Edinburgh[1], Louisiana State University[2]
maria.wolters@ed.ac.uk

ABSTRACT

People's lived experience of mental illness often includes symptoms associated with several different conditions. The Brief Symptom Inventory (BSI) is a well validated tool for collecting symptom self-reports that covers nine dimensions: depression, anxiety, phobia, paranoia, psychosis, OCD, hostility, somatisation, and interpersonal. In this paper, we investigate to what extent these dimensions are reflected in prosody. Prosody was characterised using five core principal components (PC) derived from GeMAPS analysis of a data set from 8 studies (14907 sound files, 990 participants). We used the data from a subset of 317 participants (5967 sound files, 4 studies) who had completed the BSI. Each BSI dimension shows a distinct pattern of correlations between the number of symptoms reported and our five PCs, but these patterns differ when comparing undergraduates (more Caucasians, more females) and a patient sample (more African-Americans, more males). We conclude that speech corpora for mental health studies need better demographic balance.

Keywords: prosody; GeMAPS; mental health.

1. INTRODUCTION

Just as emotions are reflected in speech [17], so are mental health conditions like depression, anxiety, or schizophrenia [12, 2, 4]. There is a rich literature that focuses on leveraging speech, in particular prosodic features, to detect the presence and/or severity of specific mental health conditions, such as depression [4].

In this paper, we examine the extent to which prosodic features co-vary with sets of symptoms that characterise the complex lived experience of mental wellbeing, going beyond the focus on a few specific diagnostic categories. In particular, we investigate whether well-defined sets of symptoms, such as hostility, interpersonal difficulties, somatisation, or anxiety, show different patterns in the way they correlate with prosodic features.

Symptoms are of interest for both theoretical and practical reasons. Within psychiatry, researchers are moving away from diagnosis towards a trans-diagnostic focus on both self-reported and objectively measurable symptoms of mental health [7]. Presence and severity of symptoms are also closely linked to quality of life [15, 9], because symptom patterns determine whether somebody with impaired mental wellbeing will actually be impaired in their functioning, as well. Finally, a symptom-based approach allows us to acknowledge the complex comorbidity patterns within mental health.

While the data used in our paper are a subset of the data from Cohen et al. [3] (used with permission), our work differs from the earlier paper in two key aspects: We used a standard acoustic feature set, GeMAPS [8] as the basis for deriving our principal component, and we do not use voice features to predict symptom scores; instead, we are interested in overall patterns of correlations.

The paper is structured as follows. In Section 2, we introduce the symptoms measure used in this study, the Brief Symptom Inventory (BSI; [6]). We then describe the data set and the statistical analysis method, while Section 3 summarises the distinctive correlation patterns we found for each of the nine BSI dimensions, and discusses them in the context of what is known about prosody and mental health. We conclude in Section 4 by arguing that speech corpora for the investigation of mental health and wellbeing should complement condition-specific questionnaires with more detailed assessments of mood and overall symptoms of potential psychopathology.

2. METHOD

2.1. Data

The full data set was taken from 8 studies involving people with and without a history of mental health problems. Participants were asked to produce speech in one of three different free speech tasks. These involved discussing

- daily routines, hobbies and/or living situations
- experiences and reactions to positive, neutral, or negative images from the International Af-

Table 1: Demographics for Participants with and without BSI information (BSI/No BSI) and with and without a history of mental health problems (Diag./No Diag.).

	BSI			No BSI		
	No Diag.	Diag.	All	No Diag.	Diag.	All
N	267 (85%)	47 (15%)	314	582 (86%)	94 (14%)	676
Age (M(SD))	20 (3)	43 (10)	24 (10)	21 (5)	41 (12)	23 (9)
Gender						
Female	169 (90%)	18 (10%)	187	351 (90%)	38 (10%)	389
Male	97 (78%)	28 (22%)	125	143 (72%)	56 (28%)	187
Not Spec.	1 (50%)	1 (50%)	2	88 (100%)	0	88
Ethnicity						
Afr.-Am.	30 (59%)	21 (41%)	51	57 (56%)	45 (44%)	102
Asian-Am.	4 (100%)	0 (0%)	4	14 (93%)	1 (7%)	15
Caucasian	221 (90%)	25 (10%)	246	398 (90%)	44 (10%)	442
Other/Not Spec.	12 (92%)	1 (8%)	13	113 (97%)	4 (4%)	117

fective Picture System (IAPS; [11] (e.g., door, lamp)

- autobiographical memories that were neutral in tone; e.g., life events or changes that were not inherently pleasant or unpleasant.

Instructions and stimulus presentation (e.g., IAPS slides) were automated and participants were encouraged to speak for the duration of the recording. The length of sound files varied between 20 and 90 seconds. The present data set is a subset of a larger corpus, discussed in [3], and used with permission.

2.1.1. Mental Health Symptoms

In four of the original studies, symptoms were measured using the Brief Symptom Inventory (BSI; [6]), which assesses a broad range of psychopathology and asks participants to focus on symptoms that they experienced during the past week. The BSI has 53 items and was derived from a larger 90 item scale [5] which has a nine-factor structure based on the diagnoses with which the symptoms that load on a factor are typically associated: depression, anxiety, hostility, somatization, obsessions/compulsions, interpersonal sensitivity, phobias, paranoia and psychosis.

2.1.2. Participants

The data set consists of 14907 sound files from 990 participants, recruited for eight studies (AD07, AD09, HR08, HR09, HR10, HR11, PT08, and PT10). Participants in two studies were patients with severe mental illness (PT08, PI10) and community controls (PT10); participants in the remaining six studies were undergraduates from a US university, some of whom (part of HR08) had been selected for high schizotypy. We have BSI scores from 314 (31.7%) of these participants, covering studies HR08, HR09, HR11, and PT08.

Table 1 shows that the demographics of the groups with and without BSI information are very similar. However, we do not have any gender information for 7% of all participants with no BSI data.

The group of participants for whom we have diagnostic information is mostly middle-aged, compared to the student group, and also consists of more men than women (60% male in both BSI and non-BSI groups, versus 36% male in the group with BSI and 29% in the group without diagnostic information), and more African Americans (11% in the baseline group versus 45% in the group with diagnostic information). While the number of patients in our sample, 47, may seem small, it is acceptable for a psychiatric study, and the pattern of comorbidities seen reflects real clinical practice.

All but 1 of the 47 participants with diagnostic information in the BSI group had been diagnosed with a mental illness previously. 9 (20%) only have one diagnosis, while the remaining 37 have a history of two or more diagnoses, including depression, schizophrenia, and psychosis.

2.1.3. Prosodic Features

We used the extended Geneva Minimalistic Acoustic Parameter Set (GeMAPS, [8]). GeMAPS is a standard feature set for detecting emotion and paralinguistic phenomena that was designed to be minimalistic. Its 88 features primarily comprise acoustic low-level descriptors and cover spectral, cepstral, prosodic and voice quality information. Computation does not require transcription of stimuli and is implemented through an open source tool kit, maximising replicability.

For each GeMAPS variable, all values that were an order of magnitude higher than the 95^{th} percentile, or an order of magnitude lower than the 5^{th} percentile, were labelled as outliers, and treated as missing data by converting them to `NA`.

2.2. Principal Components Analysis

Since many of the 88 GeMAPS features co-vary, we used exploratory Principal Component Analysis. The Promax rotation was used to account for the inevitable co-variation in prosodic measures. Each data point corresponded to a single sound file, thus, some of the principal components may contain speaker specific information.

The input for PCA were pairwise complete Spearman correlation coefficients between all 88 GeMAPS features, with outliers removed (c.f. Section 2.1.3. Spearman was used due to the non-normality of many GeMAPS features. The resulting correlation matrix was smoothed using the R function `cor.smooth` [16]

In order to examine the stability of the principal component solution, we performed PCA for three prespecified subsets: data from patients and community controls only ($n = 2854$ sound files), data from undergraduates, excluding HR08 ($n = 8645$), and data from the 20-second IAPS picture description task only ($n = 12117$).

The non-graphical Cattels' scree test, as implemented in [14], suggests that the optimal solution for each data set consists of 10–15 principal components. When examining the overlap between the GeMAPS variables loading highly on the main principal components of each data set, it appears that the first five principal components, which explain 59% of the observed variation in the full data set. In the rest of this paper, we will therefore use these components, which are summarised in Table **??** together with the three highest loading GeMAPS variables on each factor.

The Loudness/Rate factor (*Loud/Rate*, proportion: 0.17) comprises information about the distribution of loudness and formant amplitude (loudness), and voiced / unvoiced segment length statistics (rate). The variables that load on Loudness/Variation (*Loud/Var*, proportion: 0.14) describe the distribution of loudness, spectral flux, and loudness slope. The factor *Spectrum/VQ* (proportion: 0.11) includes cepstral coefficient distribution data as well as variation in the alpha ratio and the Hammarberg Index, which are related to voice quality. Variables describing the distribution of local jitter and shimmer load on the Jitter/Shimmer factor (*JitShim*) (proportion: 0.11), while the *Vocal Tract* factor consists mainly of data on the mean and standard deviation of F2 and F3 (proportion: 0.07).

2.3. Statistical Analysis

To establish robust correlations between BSI symptom scores and principal components, estimates and 99% confidence intervals for each estimate were computed using bootstrapping as implemented by the R function `spearman.cor.multcomp` [10] to correct for the effect of multiple comparisons.

3. RESULTS

Table 2 summarises the findings for the full data set, while Table 3 focuses on those participants for whom we have diagnostic information. In both groups, each set of symptoms has a distinctive pattern of correlations, and many overlap in the way they express themselves in speech. For example, symptoms of OCD, Depression, and Paranoia all result in decreased jitter/shimmer.

In the main group, which is dominated by female Caucasian undergraduates, the direction of trends is as expected from the literature. Depressive symptoms, for example, are characterised by slower speaking rate, and less jitter/shimmer, while symptoms of anxiety correlate with higher jitter/shimmer. However, significant correlations (level: $p < 0.01$, corrected for multiple comparisons) tend to be small, even though this group of participants was not screened for mental health.

In our patient group, which shows the comorbidities typical of severe mental illness, and is mostly male and almost half African-American, the loudness/rate variable is no longer significant for depressive symptoms. Instead, we see a strong correlation between variability in loudness and strength of symptoms. We find the same for anxiety, with an additional decrease in measures related to the shape of the spectrum, and increases in F2 and F3 mean and variability. Overall, correlations in the patient group are larger, and almost all symptom groups correlate negatively with Spectrum/Voice Quality and positively with the Vocal Tract factor.

Since the two participant groups differ with respect to age, gender, and ethnicity, we conducted a post-hoc fully factorial MANOVA with five outcome variables and four predictors to explore the effect of these differences. The outcome variables were each speaker's median scores on the five

Table 2: Correlations between symptoms and factors, entire data set. *: sig. at $p < 0.01$.

Symptoms	Loud/Rate	Loud/Var	Spectrum/VQ	Vocal Tract	JitShim
Anxiety	0.01	0.03	-0.04*	0.11*	0.08*
Depression	-0.06*	0.01	-0.01	0.03*	-0.08*
Hostility	-0.03*	-0.05*	-0.06*	0.19*	-0.03
Interpersonal	-0.04	-0.02	-0.06*	0.18*	0.1
OCD	-0.04	-0.05*	0.03	0.09	-0.06*
Paranoia	-0.06*	0.00	0.00	0.10*	-0.08*
Phobia	0.00	0.06	0.00	0.09*	0.01
Psychoticism	-0.05*	0.02	0.02	0.01	-0.07*
Somatisation	0.00	0.05*	0.00	0.21*	0.07*

Table 3: Correlations between symptoms and factors, participants with diagnosis information only. *: sig. at $p < 0.01$.

Symptoms	Loud/Rate	Loud/Var	Spectrum/VQ	Vocal Tract	JitShim
Anxiety	0.07*	0.12*	-0.25*	0.27*	0.03
Depression	-0.05	0.15*	-0.17*	0.06	-0.08*
Hostility	-0.01	0.0	-0.16*	0.36*	-0.04
Interpersonal	-0.06	0.15*	-0.23*	0.33*	0.07*
OCD	0.14*	0.12	-0.22*	0.19*	-0.15*
Paranoia	0.0	-0.14*	-0.27*	0.40*	-0.14*
Phobia	0.05	0.18*	-0.20*	0.30*	0.04
Psychoticism	0.04	0.11*	-0.27*	0.14	-0.08*
Somatisation	0.11*	0.17*	-0.25*	0.38*	-0.09*

prosodic factors, and the predictors were age, gender, group (patient versus student), and ethnicity (caucasian versus non-caucasian) for the 314 participants that contributed BSI values (c.f. Table 1). Four terms reached significance: age (F=8.1461, $p < 0.0001$), gender (F=12.1250, $p < 0.0001$), group (F=2.8081, $p < 0.03$) and Caucasian×age×group (F=2.24696, $p < 0.05$).

Individual ANOVAs show the expected effects of age and gender on the five prosodic factors. While Loud/Rate is not affected significantly by demographics, Spectrum/VT varies by group (F(1=7.373, $p < 0.01$), Vocal Tract is by gender (F(1)=131.829, $p < 0.0001$), group (F(1)=35.178, $p < 0.0001$), and age (F(1)=9.232, $p < 0.005$). JitShim is affected by gender (F(1)=16.818, $p < 0.0005$) and gender×group (F(1)=7.3, $p < 0.01$).

4. DISCUSSION AND CONCLUSION

The correlations observed between prosodic factors, as derived via PCA, and mental health symptoms, as measured by the BSI, are as expected for the undergraduate sample, but quite different for the patient sample. Initial post-hoc analysis indicates that while age and gender differences did lead to significant differences in the prosodic factors between undergraduates and participants with a history of mental illness, a clear group effect remains that cannot be reduced to demographics. The unit of analysis for both deriving the initial factors and the correlation analyses was the individual sound file, which is relevant for brief clinical screening in primary care. Our results indicate that we may require explicit models of inter-speaker variation to ensure generalisability across data sets. In further work, it would also be useful to study relevant intra-speaker variation, in particular the difference between neutral speech and speech about a topic or image designed to elicit emotion. In addition, the BSI could be supplemented with other measures of symptom burden, personality, quality of life, and current mentla state [1, 13].

Acknowledgements This analysis was funded by The Alan Turing Institute, EPSRC grant EP/N510129/1.

5. REFERENCES

[1] Benishek, L. A., Hayes, C. M., Bieschke, K. J., Stöffelmayr, B. E. 1998. Exploratory and confirmatory factor analyses of the brief symptom inventory among substance abusers. *Journal of Substance Abuse* 10(2), 103–114.

[2] Cohen, A. S., Alpert, M., Nienow, T. M., Dinzeo, T. J., Docherty, N. M. Aug. 2008. Computerized measurement of negative symptoms in schizophrenia. *Journal of psychiatric research* 42(10), 827–36. Citation Key: Cohen2008a.

[3] Cohen, A. S., Renshaw, T. L., Mitchell, K. R., Kim, Y. June 2016. A psychometric investigation of "macroscopic" speech measures for clinical and psychological science. *Behavior Research Methods* 48(2), 475–486. Publisher: Springer US.

[4] Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., Quatieri, T. F. Apr. 2015. A review of depression and suicide risk assessment using speech analysis. *Speech Communication* 71, 10–49. Citation Key: Cummins2015.

[5] Derogatis, L. R. 1992. Scl-90-r: Administration, scoring & procedures manual-ii for the (revised) version and other instruments of the psychopathology rating scale series. *Clinical Psychometric Research.* 1–16.

[6] Derogatis, L. R., Melisaratos, N. Aug. 1983. The Brief Symptom Inventory: an introductory report. *Psychological Medicine* 13(3), 595–605.

[7] Elvevåg, B., Cohen, A., Wolters, M., Whalley, H., Gountouna, V.-E., Kuznetsova, K., Watson, A., Nicodemus, K. 2016. An examination of the language construct in NIMH's research domain criteria: Time for reconceptualization! *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics* 171(6).

[8] Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., Andre, E., Busso, C., Devillers, L. Y., Epps, J., Laukka, P., Narayanan, S. S., Truong, K. P. Apr. 2016. The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. *IEEE Transactions on Affective Computing* 7(2), 190–202. Publisher: IEEE.

[9] Gao, K., Su, M., Sweet, J., Calabrese, J. R. Feb. 2019. Correlation between depression/anxiety symptom severity and quality of life in patients with major depressive disorder or bipolar disorder. *Journal of Affective Disorders* 244, 9–15.

[10] Herv'e, M. 2018. *RVAideMemoire: Testing and Plotting Procedures for Biostatistics*. R package version 0.9-70.

[11] Lang, P., Bradley, M. M. 2007. The international affective picture system (iaps) in the study of emotion and attention. *Handbook of emotion elicitation and assessment* 29.

[12] Laukka, P., Linnman, C., Åhs, F., Pissiota, A., Frans, O., Faria, V., Michelgrard, r., Appel, L., Fredrikson, M., Furmark, T. 2008. In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech. *Journal of Nonverbal Behavior* 32(4), 195–214. Publisher: Springer Citation Key: laukka2008nva.

[13] Müller, J. M., Postert, C., Beyer, T., Furniss, T., Achtergarde, S. Jun 2010. Comparison of eleven short versions of the symptom checklist 90-revised (scl-90-r) for use in the assessment of general psychopathology. *Journal of Psychopathology and Behavioral Assessment* 32(2), 246–254.

[14] Raiche, G. 2010. *an R package for parallel analysis and non graphical solutions to the Cattell scree test*. R package version 2.3.3.

[15] Renshaw, T. L., Cohen, A. S. 2014. Life satisfaction as a distinguishing indicator of college student functioning: Further validation of the two-continua model of mental health. *Social Indicators Research* 117(1), 319–334.

[16] Revelle, W. 2018. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University Evanston, Illinois. R package version 1.8.10.

[17] Scherer, K. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40(1–2), 227–256. Citation Key: scherer:03.