INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Computational Analysis of Acoustic Descriptors in Psychotic Patients

Torsten Wörtwein[1], *Tadas Baltrušaitis*[2], *Eugene Laksana*[3], *Luciana Pennant*[4], *Elizabeth S. Liebson*[4,5], *Dost Öngür*[4,5], *Justin T. Baker*[4,5], *and Louis-Philippe Morency*[2]

[1]Karlsruhe Institute of Technology
[2]Carnegie Mellon University
[3]University of Southern California
[4]McLean Hospital
[5]Harvard Medical School

torsten.woertwein@student.kit.edu, {tbaltrus, morency}@cs.cmu.edu, elaksana@ict.usc.edu, lpennant@mclean.harvard.edu, {eliebson, dongur, jtbaker}@partners.org

Abstract

Various forms of psychotic disorders, including schizophrenia, can influence how we speak. Therefore, clinicians assess speech and language behaviors of their patients. While it is difficult for humans to quantify speech behaviors precisely, acoustic descriptors, such as tenseness of voice and speech rate, can be quantified automatically. In this work, we identify previously unstudied acoustic descriptors related to the severity of psychotic symptoms within a clinical population (N=29). Our dataset consists of semi-structured interviews between patients and clinicians. Psychotic disorders are often characterized by two groups of symptoms: negative and positive. While negative symptoms are also prevalent in disorders such as depression, positive symptoms in psychotic disorders have rarely been studied from an acoustic and computational perspective. Our experiments show relationships between psychotic symptoms and acoustic descriptors related to voice quality consistency, variation of speech rate and volume, vowel space, and a parameter of glottal flow. Further, we show that certain acoustic descriptors can track a patient's state from admission to discharge. Finally, we demonstrate that measures from the Brief Psychiatric Rating Scale (BPRS) can be estimated with acoustic descriptors.

Index Terms: psychotic disorders, acoustic descriptors, brief psychiatric rating scale

1. Introduction

Many people are affected by a psychotic disorder during their lifetimes or know someone who is. For schizophrenia alone, 23.6 million cases were reported globally in 2013 [1]. Psychotic disorders affect how we speak [2] and how we express ourselves with facial expressions [3, 4]. Thus, medical assessments have included speech-related descriptors for a long time [5, 6]. Such differences in speech might be difficult for humans to assess objectively, but can be captured by computational acoustic descriptors. This brings the opportunity to support clinicians in assessing symptoms and allow for better decision-making.

Work on this topic is limited, and many computational acoustic descriptors have not been studied with clinical psychotic patients, including articulation rate, vowel space [7], speech volume, and glottal flow parameters. While an early measure of voice tenseness was associated with schizophrenia [8], more recent robust measures of voice quality such as peak slope [9] have not been investigated. Psychotic disorders are described by positive symptoms, exaggerations of normal functions (e.g., grandiosity and hallucination), and negative symptoms, declines of normal functions (e.g., emotional withdrawal and motor retardation) [10, 11]. While speech-related behaviors of negative symptoms have been studied through work on depression and PTSD, positive symptoms have not been paid much attention in computational studies.

In this work, we perform a computational study of acoustic descriptors to understand psychotic disorders and their positive symptoms better. This study is performed on a dataset of semi-structured interviews between clinical psychotic patients and clinicians. We investigate the following questions.

Q1: What are the acoustic descriptors related to overall psychotic severity? What are the acoustic descriptors related to specific positive symptoms?

Q2: Can we estimate positive symptoms and overall severity of psychotic disorders with acoustic descriptors?

2. Related Work

While the research community has studied how psychotic individuals perceive, e.g., speech [12] and emotions [13], there is less research on whether they express themselves differently through speech and language.

In a within-patient study of patients diagnosed with schizophrenia, a decrease of the fundamental frequency (pitch) and a better pronunciation of vowels, i.e., the first and second formants were closer to a reference pronunciation, were observed at discharge compared to admission [8]. In the same study, a tendency toward a more tense voice was observed for patients with schizophrenia, while the opposite has been seen in depressed patients at discharge [8]. Compared to this within-patient study, our paper investigates more acoustic descriptors, including a more robust estimator of voice tenseness, between patients. An exhibition of inadequate speech behaviors, e.g., in volume, in rate, and in pitch variation, was found in children diagnosed with schizophrenia compared to a control group of the same size and demographic [2]. The same study observed that children with schizophrenia were not identifiable by a single speech behavior, but that they often deviated more from the norm on many speech-related behaviors, e.g., speaking too loudly or too quietly. In contrast to our work, all speech- and language-related behaviors were manually assessed.

More recently, second formant variation was linked to the severity of negative symptoms [11], e.g., blunted affect and emotional withdrawal, among patients with schizophrenia [14]. Besides the second formant, the first formant showed a similar but not statistically significant trend. In other disorders, more

http://dx.doi.org/10.21437/Interspeech.2017-466

severe negative symptoms have been linked to a smaller vowel space for self-reported PTSD compared to a control group [7], i.e., similar first and second formants for different vowels, and to a more tense voice in self-reported depression [15]. Even though these two studies are based on many individuals, the individuals are not clinically diagnosed and are not hospitalized, i.e., we expect milder symptoms.

We contrast with previous research by investigating robust computational measures of acoustic descriptors, which have not been studied previously in psychotic disorders. Besides establishing relationships between acoustic descriptors and symptoms, we also investigate automatic estimation of psychotic severity, with a focus on positive symptoms.

3. Methods

Our dataset consists of audio and video recordings of 29 semi-structured interviews between clinicians and 20 unique psychotic individuals who are hospitalized in an inpatient service at a psychiatric hospital. The semi-structured interview protocol was designed to reflect the daily clinical encounters by patients and their clinicians. This dataset is a significant expansion of a previously-published dataset used to study facial expressions of psychotic patients [4]. Our new dataset includes multiple interviews from the same patient from admission to discharge to analyze temporal changes. While most patients are diagnosed with schizophrenia, some are diagnosed with bipolar or with mania. The average duration of these interviews is 8.33 minutes (SD=4.22). Of the 29 interviews, 17 interviews are with male and 12 interviews are with female patients. They are recorded using head-mounted microphones and one webcam for each facing the upper body of the patient and clinician.

After each of the interviews, the patients are assessed using the 24-item version of the Brief Psychiatric Rating Scale (BPRS) [6]. It was designed to measure the severity of relatively independent symptoms often found in psychiatric disorders [16]. The BPRS total score (M=42.4, SD=13.6) is the sum of all BPRS items, which are scored on an ordinal scale from 1 (not present) to 7 (extremely severe). Therefore, BPRS total ranges theoretically from 24 to 168. In our analysis, we focus on the total score as well as on positive items [10]. Further, we omit BPRS items that do not vary in our patient population (SD$<$1). This leaves the following six positive BPRS items: *grandiosity*, *elevated mood*, *hallucination*, *unusual thoughts*, *excitement*, and *motor hyperactivity*.

3.1. Speaker Diarization

A first step when computing acoustic descriptors is speaker diarization. Our experimental setup includes head-mounted microphones designed to reduced cross-over speech. Even in these good recording setups, we hear the other person talking. In this paper, we explore manual annotations and an automatic diarization for speaker diarization. The experiments with manual diarization allow studying computational acoustic indicators in the ideal case. Experiments using automatic diarization allow us to get closer to our goal of building decision support tools.

Our automatic speaker diarization is based on the time delay of arrival (TDOA). Since we have only two speakers each wearing a head-mounted microphone, we can distinguish speakers by TDOA between the two audio signals as estimated by the generalized cross-correlation with phase transform [17]. Patient and clinician, who are spatially separated by the recording setup, are approximately 3 meters apart. Therefore, we expect a delay of 8ms between the audio signals. TDOA might not be reliably estimated when the other microphone does not pick up an audio signal. Therefore, we rely on TDOA only if a voice is detected [18] in both audio signals. If TDOA is less than 5ms, and if a voice is detected in both signals, we assume that both patient and clinician are speaking. If speech is detected in one signal only, we assume that the corresponding person is speaking. A recording problem during six of the interviews made it impossible to recover the audio signal from the clinician microphone. For this reason, experiments with automatic diarization are performed on 23 interviews.

We calculate the diarization error rate (DER) of the automatic approach, over all interviews, based on the manual annotations. It is suggested to use a 250ms no-score collar around the annotated segment boundaries [19]. However, this would remove a significant amount of our annotations. Without this collar, we reach a DER of 20.10%, which is still comparable to DERs in similar settings [20] with the collar.

3.2. Computational Acoustic Descriptors

As mentioned in Section 2, limited prior work has investigated computational acoustic descriptors in interviews with psychotic patients. We use descriptors inspired from work on depression and PTSD behavior analysis [7, 15]. Our descriptors include the first Mel-frequency cepstral coefficient ($MFCC_0$) as a measure of volume, vowel space [7], formants (F_1 and F_2), fundamental frequency, voice quality descriptors from COVAREP [21], and articulation rate from Praat [22].

We remove parts of the audio signals where the patient is not voicing [23] since many descriptors can only be estimated for voicing. On average, we have 3.14 minutes (SD=2.08) of voicing for the patients. Then, we compute descriptive statistics of our descriptors, i.e., median and interquartile range (IQR). These two statistics are used because they are robust against outliers, which might occur due to the diarization.

Articulation rate is the ratio of the number of syllables and the phonation time over all speech segments according to the diarization. The variation of articulation rate is the IQR of the articulation rates per speech segment. We do not calculate the median for fundamental frequency or formants because they have only shown to be indicative in within-patient studies [8]. Speech volume (median of $MFCC_0$) is not used because it has been shown to be sensitive to the recording environment. This leads to 12 acoustic descriptors for computational analysis. All descriptors are mean-centered and normalized by their standard deviation.

3.3. Automatic Estimation of BPRS Items

In Q2 we want to estimate BPRS items. Since we have not too many interviews, we choose linear support vector regression (SVR) to estimate BPRS items. Experiments are performed in a speaker-independent fashion using the leave-one-patient-out method. Hyperparameters of the linear SVR, including descriptor selection, are determined automatically using a nested leave-one-patient-out validation on the training set.

For each training partition, we find a suitable subset of descriptors by conducting a greedy forward selection on the minimizing criteria $-corr(Y, \hat{Y})$ (Pearson's linear correlation), where $\hat{Y}$ are the estimated scores and Y the corresponding ground truth scores. The maximum number of descriptors is restricted to five descriptors to prevent over-fitting. During the descriptor selection, we validate the SVR's penalty parameter

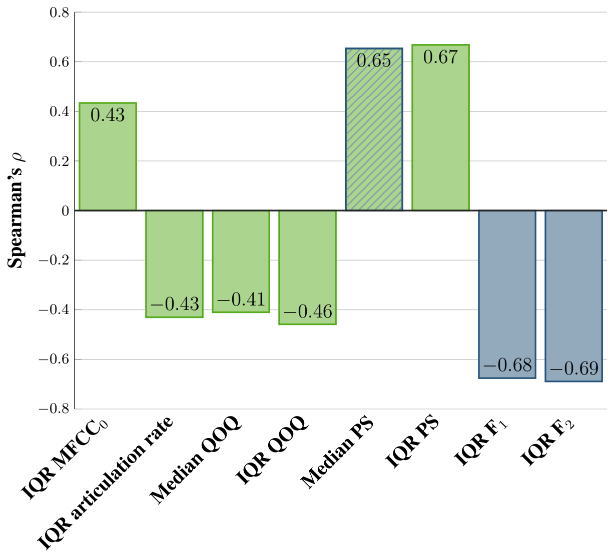

Figure 1: *Correlations between acoustic descriptors and BPRS total score on our dataset. Descriptors colored in blue have also been studied in previous work. Median PS is marked differently because an early descriptor of tenseness [25] has been investigated, but the opposite result was observed [8].*

C (between 0.001 and 100) with Bayesian optimization [24], which uses a Gaussian process to model $-corr(Y, \hat{Y})$ of the nested leave-one-patient-out validation.

We use two evaluation metrics in our experiments: Pearson's correlation coefficient (r) and the mean absolute error (MAE). Before these two metrics are calculated, estimations are clipped to valid BPRS scores. For comparison, we calculate the MAE of a naive mean estimation ($\text{MAE}_{\text{naive}}$) as a baseline, where the mean is calculated on each training fold.

4. Results and Discussion

In this section, we present our experiments to study the two research questions previously introduced: (Q1) correlation analysis of acoustic descriptor with positive BPRS items and BPRS total score, and (Q2) models to estimate BPRS items based on computational descriptors.

4.1. Q1 - Acoustic Descriptors of Psychotic Symptoms

We investigate acoustic descriptors related to the BPRS total score and positive BPRS items. Since the relationship between acoustic descriptors and BPRS might be non-linear, we use Spearman's rank correlation coefficient ρ. All descriptors are based on manual diarization to use all 29 interviews. Our significant correlation results ($p < 0.05$) are summarized in Figure 1 and Table 1. In the next paragraphs, we discuss these results.

Speech volume: IQR of MFCC_0, a measure related to the variation of speech volume, correlates positively with BPRS total score. In line with our result, it has been observed that patients diagnosed with schizophrenia deviate more from the norm for speech volume and other descriptors [2]. Individual positive items show no correlation with a variation in speech volume.

Articulation rate: We do not observe any correlations between articulation rate and BPRS. However, IQR of articulation rate correlates negatively with the BPRS total score and many positive items. For children with schizophrenia, it was observed

Table 1: *Significant correlations of acoustic descriptors for all positive BPRS items ($p < 0.05$).*

Positive symptom	Acoustic descriptor	ρ
Hallucinations	Median PS	0.43
	IQR F_1	−0.40
	IQR F_2	−0.37
Unusual thoughts	IQR PS	0.56
	Median PS	0.52
	Vowel space	0.41
	IQR F_2	−0.45
	IQR F_1	−0.55
Elevated mood	Median PS	0.40
	IQR F_2	−0.38
	IQR F_1	−0.47
	IQR articulation rate	−0.62
Grandiosity	IQR articulation rate	−0.40
	IQR F_1	−0.44
	IQR F_2	−0.44
Excitement	Vowel space	0.40
	IQR articulation rate	−0.52
Motor hyperactivity	Vowel space	0.45
	Median QOQ	−0.39
	IQR articulation rate	−0.46

that they have a more excessive variation in speech rate [2]. We found the opposite to be the case for our dataset.

Glottal flow: Quasi-open-quotient (QOQ) [26] measures the ratio of the opening time of the vocal folds. Median and IQR of QOQ correlate negatively with the BPRS total score. Larger BPRS total scores tend to be related to a smaller QOQ range and a shorter opening time of the vocal folds. The range of QOQ is often reduced for people with functional dysphonias [26], in combination with a low QOQ speaking loudly requires more effort and sounds more "stalled" [26].

Voice quality: Peak slope (PS) [9], a voice quality descriptor related to the breathy-modal-tense spectrum, correlates positively with many BPRS items. This indicates a more tense voice for more severe symptoms. A more tense voice was associated with clinical [8] and self-reported [15] depression, but the opposite was reported for patients with schizophrenia [8]. While the contradicting study observed less tense voice based on an early tenseness measure [25], this change was not statistically significant. IQR of PS also correlates positively with many BPRS items. A variation in breathy-modal-tense voice seems to be as indicative as the actual voice quality. A positive correlation of IQR of PS indicates that more severe symptoms tend to be associated with a less consistent voice quality. The consistency of voice quality has to our knowledge never been studied computationally in any clinical study.

Vowel space: It was found that vowel space correlates negatively with self-reported depression [7]. Depression is mainly characterized by negative symptoms. For positive items, e.g., *excitement* and *motor hyperactivity*, we observe a positive correlation with vowel space. This indicates that it is important to analyze positive and negative symptoms separately since effects could average out, i.e., we do not observe a correlation with overall symptoms.

Formants: IQR of the first two formants correlates negatively with almost all BPRS items, i.e., a smaller range cor-

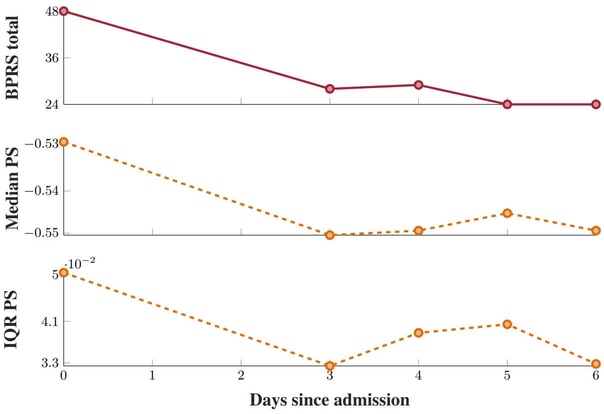

Figure 2: *BPRS total (solid line) and acoustic descriptors (dashed lines) from admission to discharge for one patient.*

relates with more severe symptoms. This has previously been observed for the variation of the second formant and indicated for the first formant for negative symptoms [14]. The first and second formant are mainly influenced by the position of the tongue and the extension of the jaw. It could be argued that psychotic patients with more severe symptoms do not move their tongue [14] and mouth as much.

We further investigate within-patient differences from admission to discharge for one patient. While, some acoustic descriptors, such as pitch in in-between studies, might seem to be indicative due to, e.g., a gender bias in a dataset, they might not be indicative in within-studies. Therefore, we would like to see that previously unstudied descriptors behave similarly over time as BPRS total within a patient from admission to discharge. For this, we plot the two peak slope statistics because they have the strongest correlations with BPRS total. As can be seen in Figure 2, peak slope's statistics behave similarly to BPRS total for this patient.

4.2. Q2 - Automatic Estimation of BPRS Items

Table 2 summarizes the estimation results for all positive BPRS items as well as the BPRS total score. Except for *hallucinations* and *unusual thoughts*, we can estimate the BPRS items well (high correlation and lower MAE than baseline). BPRS specifies to assess these two items based on only what individuals say not on how they speak. While we observe correlations with these two items, see Table 1, a linear SVR is in our case not able to estimate these two items well. All well estimable BPRS items have some relation to acoustic descriptors. While *motor hyperactivity* might not at first glance appear to be related to acoustic descriptors, the BPRS manual specifically advises taking rapid speech into account for it.

The performance comparison between descriptors based on the two diarization systems is summarized in Table 3. Even though all models are trained on descriptors based on manual diarization, we see a very similar performance with descriptors derived from the automatic diarization. This indicates that even though our automatic diarization is not perfect, we can robustly estimate our descriptors.

5. Conclusion

We discovered several acoustic descriptors related to psychotic symptoms while studying our first research question (Q1). Among them are a less consistent voice quality (larger variation in tenseness), a larger variation in speech volume, less variation in the opening time of vocal folds, a larger vowel space, and a smaller variation of speech rate. We also observe a smaller variation in formants for the overall severity and positive symptoms, which has been found in prior work for negative symptoms [14]. For tenseness of voice, our observations are in line with studies of depression and PTSD [7, 15]: a more tense voice is associated with more severe symptoms. We see small differences between positive symptoms and the overall severity. Namely, no variation in speech volume, not as prominent tense voice and not as prominent reduced variation in formants for positive symptoms. Compared to previous work on depression [7], which observed a smaller vowel space for more severe negative symptoms, we observe the opposite for the severity of positive symptoms. This emphasizes that positive and negative symptoms should be studied separately.

Based on these computational acoustic descriptors, we train models to estimate BPRS (Q2). For several of the BPRS items, we can robustly estimate the scores. Since we are able to achieve good performance even with automatic diarization, we can provide models to estimate BPRS items without the need of any manual annotations. One simple future work would be to contextualize these acoustic descriptors with the behaviors from the clinicians, including the sentiment and intimacy of the questions.

Table 2: *Estimated generalization error of BPRS items on 29 interviews. * and ** mark significantly smaller absolute errors ($p < 0.05$ and $p < 0.01$, Wilcoxon signed rank test).*

BPRS item	r_{our}	MAE_{our}	$\text{MAE}_{\text{naive}}$
Hallucinations	0.27	1.00*	1.34
Unusual thoughts	0.00	1.12	1.01
Elevated mood	0.73	0.64*	1.12
Grandiosity	0.81	0.36**	0.67
Excitement	0.71	0.60	0.91
Motor hyperactivity	0.74	0.55**	0.94
Total	0.57	9.73	12.47

Table 3: *Performance comparison between descriptors based on manual (m) and automatic (a) diarization of 23 interviews. * and ** mark significantly smaller absolute errors ($p < 0.05$ and $p < 0.01$, Wilcoxon signed rank test).*

BPRS item	r_{our}		MAE_{our}		$\text{MAE}_{\text{naive}}$
	m	a	m	a	
Hallucinations	0.23	0.29	1.15	1.84	1.44
Unusual thoughts	0.12	0.19	1.09	1.17	1.10
Elevated mood	0.78	0.68	0.62*	0.80*	1.21
Grandiosity	0.82	0.80	0.41**	0.46**	0.77
Excitement	0.76	0.72	0.53*	0.69	0.94
Motor hyperact.	0.73	0.67	0.51**	0.84	0.92
Total	0.47	0.49	11.02	10.35	12.56

6. Acknowledgments

This work is partially supported by the Carnegie Mellon University's BrainHub. Research by the first author is supported by a fellowship within the FITweltweit programme of the German Academic Exchange Service.

7. References

[1] T. Vos, R. M. Barber, B. Bell, A. Bertozzi-Villa, S. Biryukov, I. Bolliger, F. Charlson, A. Davis, L. Degenhardt, D. Dicker *et al.*, "Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the global burden of disease study 2013," *The Lancet*, vol. 386, no. 9995, p. 743, 2015.

[2] W. Goldfarb, N. Goldfarb, P. Braunstein, and H. Scholl, "Speech and language faults of schizophrenic children," *Journal of autism and childhood schizophrenia*, vol. 2, no. 3, pp. 219–233, 1972.

[3] T. Tron, A. Peled, A. Grinsphoon, and D. Weinshall, "Automated facial expressions analysis in schizophrenia: A continuous dynamic approach," in *International Symposium on Pervasive Computing Paradigms for Mental Health*. Springer, 2015, pp. 72–81.

[4] S. Vijay, T. Baltrušaitis, L. Pennant, D. Ongür, J. T. Baker, and L.-P. Morency, "Computational study of psychosis symptoms and facial expressions," in *Computing and Mental Health, Workshop at SIGCHI*, 2016.

[5] P. T. Trzepacz and R. W. Baker, *The psychiatric mental status examination*. Oxford University Press, 1993.

[6] V. J. Lukoff D., Nuechterlein KH., "Manual for expanded brief psychiatric rating scale," *Schizophrenia Bulletin*, vol. 12, pp. 594–602, 1986.

[7] S. Scherer, G. Lucas, J. Gratch, A. Rizzo, and L.-P. Morency, "Self-reported symptoms of depression and ptsd are associated with reduced vowel space in screening interviews," *IEEE Transactions on Affective Computing*, vol. 7, no. 1, pp. 59–73, 2015.

[8] F. Tolkmitt, H. Helfrich, R. Standke, and K. R. Scherer, "Vocal indicators of psychiatric treatment effects in depressives and schizophrenics," *Journal of communication disorders*, vol. 15, no. 3, pp. 209–222, 1982.

[9] J. Kane and C. Gobl, "Identifying regions of non-modal phonation using features of the wavelet transform," in *Proceedings of Interspeech*, 2011, pp. 177–180.

[10] I. R. Nicholson, J. E. Chapman, and R. W. Neufeld, "Variability in bprs definitions of positive and negative symptoms," *Schizophrenia research*, vol. 17, no. 2, pp. 177–185, 1995.

[11] S. R. Kay, A. Flszbein, and L. A. Opfer, "The positive and negative syndrome scale (panss) for schizophrenia," *Schizophrenia bulletin*, vol. 13, no. 2, p. 261, 1987.

[12] D. I. Leitman, J. J. Foxe, P. D. Butler, A. Saperstein, N. Revheim, and D. C. Javitt, "Sensory contributions to impaired prosodic processing in schizophrenia," *Biological psychiatry*, vol. 58, no. 1, pp. 56–61, 2005.

[13] C. G. Kohler, J. B. Walker, E. A. Martin, K. M. Healey, and P. J. Moberg, "Facial emotion perception in schizophrenia: a meta-analytic review," *Schizophrenia bulletin*, vol. 36, no. 5, pp. 1009–1019, 2009.

[14] M. A. Covington, S. A. Lunden, S. L. Cristofaro, C. R. Wan, C. T. Bailey, B. Broussard, R. Fogarty, S. Johnson, S. Zhang, and M. T. Compton, "Phonetic measures of reduced tongue movement correlate with negative symptom severity in hospitalized patients with first-episode schizophrenia-spectrum disorders," *Schizophrenia research*, vol. 142, no. 1, pp. 93–95, 2012.

[15] S. Scherer, G. Stratou, J. Gratch, and L.-P. Morency, "Investigating voice quality as a speaker-independent indicator of depression and ptsd," in *Proceedings of Interspeech*, 2013, pp. 847–851.

[16] J. E. Overall and D. R. Gorham, "The brief psychiatric rating scale," *Psychological reports*, vol. 10, no. 3, pp. 799–812, 1962.

[17] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[18] T. Drugman, Y. Stylianou, Y. Kida, and M. Akamine, "Voice activity detection: Merging source and filter-based information," *IEEE Signal Processing Letters*, vol. 23, no. 2, pp. 252–256, 2015.

[19] N. I. of Standards and Technology, "Spring 2006 (rt-06s) rich transcription meeting recognition evaluation plan," 2006, http://www.itl.nist.gov/iad/mig/tests/rt/2006-spring/docs/rt06s-meeting-eval-plan-V2.pdf.

[20] K. J. Han, P. G. Georgiou, and S. S. Narayanan, "The sail speaker diarization system for analysis of spontaneous meetings," in *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*. IEEE, 2008, pp. 966–971.

[21] G. Degottex, J. Kane, T. Drugman, T. Raitio, and S. Scherer, "COVAREP - a collaborative voice analysis repository for speech technologies," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, 2014, pp. 960–964.

[22] N. de Jong and T. Wempe, "Praat script to detect syllable nuclei and measure speech rate automatically," *Behavior Research Methods*, vol. 41, no. 2, pp. 385–390, 2009.

[23] T. Drugman and A. Abeer, "Joint robust voicing detection and pitch estimation based on residual harmonics," in *Proceedings of Interspeech*, 2011, pp. 1973–1976.

[24] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in neural information processing systems*, 2012, pp. 2951–2959.

[25] J. D. M. H. Laver, "Individual features in voice quality." Ph.D. dissertation, University of Edinburgh, 1987.

[26] T. Hacki, "Klassifizierung von glottisdysfunktionen mit hilfe der elektroglottographie," *Folia Phoniatrica*, vol. 41, no. 1, pp. 43–48, 1989.