Identifying Schizophrenia Based on Temporal Parameters in Spontaneous Speech

Gábor Gosztolya[1], *Anita Bagi* [2,5], *Szilvia Szalóki* [3,5], *István Szendi* [3,5], *Ildikó Hoffmann* [2,4,5]

[1] MTA-SZTE Research Group on Artificial Intelligence, Szeged, Hungary
[2] University of Szeged, Department of Hungarian Linguistics, Szeged, Hungary
[3] Department of Psychiatry, University of Szeged, Hungary
[4] Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest, Hungary
[5] Prevention of Mental Illnesses Interdisciplinary Research Group, University of Szeged, Hungary

ggabor @ inf.u-szeged.hu

Abstract

Schizophrenia is a neurodegenerative disease with spectrum disorder, consisting of groups of different deficits. It is, among other symptoms, characterized by reduced information processing speed and deficits in verbal fluency. In this study we focus on the speech production fluency of patients with schizophrenia compared to healthy controls. Our aim is to show that a temporal speech parameter set consisting of articulation tempo, speech tempo and various pause-related indicators, originally defined for the sake of early detection of various dementia types such as Mild Cognitive Impairment and early Alzheimer's Disease, is able to capture specific differences in the spontaneous speech of the two groups. We tested the applicability of the temporal indicators by machine learning (i.e. by using Support-Vector Machines). Our results show that members of the two speaker groups could be identified with classification accuracy scores of between $70 - 80\%$ and F-measure scores between 81% and 87%. Our detailed examination revealed that, among the pause-related temporal parameters, the most useful for distinguishing the two speaker groups were those which took into account both the silent and filled pauses.

Index Terms: spontaneous speech, temporal parameters, schizophrenia, filled pauses

1. Introduction

According to Crow theory [1],schizophrenia (phenomenologically) could be a universal illness, which can be found in all the populations of the planet. Crow assumed that it might be closely related to the development of psychological structures and genetic changes that cause lateralization. The following criteria of symptoms represent the disease: (1) delusions; (2) hallucinations; (3) incoherent speech; (4) strikingly disintegrated or catatonic behavior; and (5) negative symptoms, i.e. emotional emptiness, alogia, or avolition [2].

Schizophrenia is characterized by several cognitive deficits including reduced information processing speed and impaired working memory [3]. Deficits in memory functions such as working memory, verbal fluency and episodic memory have been detected in patients with schizophrenia by neuropsychological tests [4, 5, 6]. Other authors have identified specific impairments in schizophrenic working memory and sustained attention [7, 8].

Patients with schizophrenia suffer from several impairments at different levels of speech and language [9]. Pawełczyk et al. found that schizophrenia patients scored significantly lower than controls in subtests measuring comprehension of implicit information, interpretation of humor, explanation of metaphors, inappropriate remarks and comments, discernment of emotional and language prosody and comprehension of discourse [10]. Differences were detected in prosody, while other findings indicate that the negative symptoms of schizophrenia may appear as a lack of tone and inflection [11, 12]. From the aspect of speech production, studies on spontaneous speech discuss the complexity of the communicated thought, which is less complicated in the case of schizophrenic patients than in the speech of healthy controls. However, in patients with higher performance, there is more involvement of depression and anxiety complications [13].

Several of these symptoms were analyzed by computational tools. Rosenstein et al. examined verbal memory by measuring recalled verbal processes using computational linguistic approaches [14]. Corcoran et al. found that automatized semantic and syntactic analysis could be used as a basis for diagnostic tools [15]. Prosodic abnormalities and potential characteristics were also examined [16, 17], and so were the continuity of speech or the quality and ratio of occlusive phenomena and pauses [18]. Other findings showed that patients with formal thought disorder (which could be a symptom in schizophrenia) made strikingly fewer filled pauses than controls did [19].

In this study we will focus on the deficits of memory processes reflected in spontaneous speech. We will do this by investigating directed spontaneous speech with a memory task ("tell me about your previous day"). We assume that temporal parameters of spontaneous speech will differ between healthy controls and schizophrenic speakers. We expect the most significant differences in the number and type of hesitations. We will perform our analysis in an automated way: we extract temporal speech parameters by utilizing Automatic Speech Recognition (ASR) techniques, and measure the utility of these parameters via applying statistical machine learning to distinguish the two speaker groups.

2. Temporal Speech Parameters

To investigate the spontaneous speech of schizophrenic patients and healthy controls, we calculated specific temporal parameters from their responses. We based our investigations on our previous studies [20, 21, 22], where we introduced temporal parameters focusing on hesitations in order to the early detection of Mild Cognitive Impairment (MCI). MCI, sometimes regarded as a prodromal stage of Alzheimer's Disease, is a mental disorder that is difficult to diagnose. MCI is known to influence the (spontaneous) speech of the patient in several aspects [23],

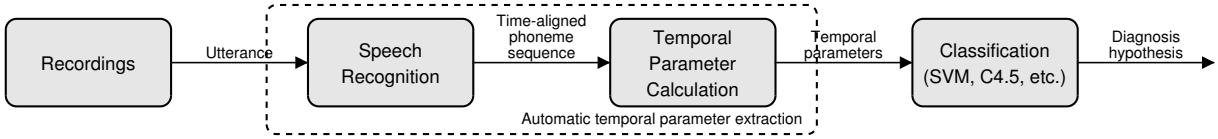

Figure 1: *Workflow of automatic temporal speech parameter calculation and analysis, based on the study of Tóth et al. [20].*

(1) **Articulation rate** was calculated as the number of phones per second during speech (excluding hesitations).

(2) **Speech tempo** (phones per second) was calculated as the number of phones per second divided by the total duration of the utterance.

(3) **Duration of utterance**, given in milliseconds.

(4) **Number of pauses** reflects the absolute number of pause occurrences.

(5) **Duration of pauses** was calculated as the total duration of pause occurrences.

(6) **Pause duration rate** was calculated by dividing the total duration of pauses by the length of the utterance.

(7) **Pause frequency** was calculated by dividing the number of pause occurrences by the length of the utterance.

(8) **Average pause duration** was calculated by dividing the total duration of pauses by the number of pauses.

Table 1: *The eight examined temporal speech parameters, based on the work of Hoffmann et al. [21] and Tóth et al. [22].*

from which we concentrated on the verbal fluency. In MCI, the verbal fluency of the patient tends to deteriorate, resulting in distinctive acoustic changes — most importantly, in longer hesitations and a lower speech rate [24, 25]. To exploit this, we developed a set of temporal parameters which mostly focus on the amount of hesitation in the speech of the subject.

Our set of temporal parameters can be seen in Table 1. Notice that parameters (4)–(8) all describe the amount of hesitation in the spontaneous speech of the subject by focusing on the number or duration of pauses in some way. At this point we may further clarify our terms regarding hesitation. The simplest form of pause is that of *silent* pause, i.e. the absence of speech. However, hesitation may manifest as *filled* pauses, i.e. vocalizations like “er”, “uhm”, “eh” etc. Clearly, both pause types indicate some sort of hesitation in spontaneous speech production. To be able to analyze both pause types, we included temporal parameters (4) to (8) in our set of temporal parameters examined by calculating them for silent pauses only, for filled pauses only, and for taking all pause occurrences into account regardless of type. This led to 18 temporal parameters overall.

2.1. ASR-Based Temporal Parameter Calculation

Calculating the above acoustic temporal indicators manually (as was done in some of our earlier studies such as in [25]) is quite expensive and labour-intensive. Therefore, we devoted our efforts to the automatic extraction of these temporal speech parameters. One choice could be to rely on signal processing methods as in e.g. [26]. Unfortunately, while it is relatively easy to distinguish silence from speech and other voiced parts of human speech via signal processing techniques, this approach cannot distinguish filled pauses from normal speech, and we would be unable to calculate articulation rate and speech tempo either.

After reflecting on the above aspects, we decided to apply Automatic Speech Recognition (ASR) techniques. (Our workflow was first presented in [20].) Evidently, an off-the-shelf ASR tool may be suboptimal for this task, mainly because standard speech recognizers are trained to minimize the transcription errors at the word level, while here we seek to extract nonverbal acoustic features like the rate of speech and the duration of silent and filled pauses. Luckily, though, the speech parameters in Table 1 do not require us to *identify* the phones; we need only to *count* them. Furthermore, while the filled pauses do not explicitly appear in the output of a standard ASR system, our set of temporal parameters requires them to be found. It is practically impossible to prepare a standard ASR system that it could handle errors such as these.

For these reasons, we decided to use a speech recognizer that provides only a phone sequence as output, treating filled pause as a special ‘phoneme’. Of course, omitting the word level completely (along with a word-level language model and a pronunciation vocabulary) can be expected to increase the number of errors at the phoneme level as well. However, as we pointed out, not all types of phone recognition errors harm the extraction of our temporal parameters; in our case only the number of phonemes and the two types of pauses (i.e. silent and filled) are important.

3. The Data

Ten subjects with schizophrenia were randomly selected from the currently available clinical research database. After random selection, eight healthy controls were matched in age and gender. The group of speakers having schizophrenia (*SCH*) and the matched group of healthy controls (*HC*) both had the same sex distribution of 50% males and 50% females. Of course, the number of speakers examined is rather low, but we plan to involve new subjects in our investigations in the near future.

The utterances were recorded between February 2016 and March 2017 at the Department of Psychiatry, Faculty of Medicine, University of Szeged. The study was approved by the Ethics Committee of the University of Szeged, and it was conducted in accordance with the Declaration of Helsinki. All the speakers were native Hungarian speakers. We used our S-GAP Test. We made the speakers perform spontaneous speech by asking them to talk about their previous day. The instruction was simply, “Tell me about your previous day!”. The subjects were then given roughly five minutes to complete the task. We used a Roland R-05 type recorder to record their replies.

The mean age was 39.9 years in the SCH group and 40.2

in the HC group. The education (in years) (t=-1.82, df=18, $p = 0.09$) and the age (t=0.06, df=18, $p = 0.96$) were not significantly different among the two speaker groups. We also performed Mini-Mental State Examination (MMSE, [27]) mental tests on our subjects, which had significantly different results for the two speaker groups (t=2.55, df=10.55, $p = 0.028$). The subjects with schizophrenia lost points mostly in the subtest of word recall; however, this was not only due to memory deficits, but also to their scattered attention.

4. Experimental Setup

4.1. Temporal Parameter Extraction

The acoustic model of the speech recognizer was trained on the BEA Hungarian Spoken Language Database [28]. This database contains spontaneous speech, which is quite important to us since filled pauses are only present in spontaneous speech. We used roughly seven hours of speech data from the BEA corpus. We made sure that the occurrences of filled pauses, breath intakes and exhales, laughter, coughs and gasps were present in the phoneme-level transcriptions in a consistent manner.

The ASR system was trained to recognize the phones in the utterances, where the phone set included the special non-verbal labels listed above. For acoustic modeling we applied a standard Deep Neural Network (DNN) with feed-forward topology. The DNN had 3 hidden layers with 1000 ReLU neurons each. We used our custom implementation, by which our team achieved the lowest phone recognition error rate published so far on the TIMIT database [29]. As a language model we employed a simple phoneme bigram (again, including all the above-mentioned non-verbal audio tags). The output of the ASR system is the phonetic segmentation and labeling of the input signal, which includes filled pauses. Based on this output, the temporal speech parameters of Table 1 can be easily extracted using simple calculations.

4.2. Evaluation Metrics

In the past, many studies in biomedical ASR applications relied on simple classification accuracy (e.g. [26, 30]). However, the frequency of the two types of subjects is quite imbalanced in the population, only 1-1.5% of the people being affected by schizophrenia; for such an imbalanced class distribution, accuracy is not a reliable metric at all. For this reason, we opted for the standard Information Retrieval metrics of *precision*, *recall* and their harmonic mean, *F-measure* (or F_1-score). Furthermore, we calculated the area under the ROC curve (the AUC metric) for the SCH class as well.

4.3. The Classification Process

Our classification process basically followed standard biomedical practices, and were similar to those of our earlier studies focusing on detecting MCI (i.e. [20, 22]). Using the above-listed temporal parameters, we trained a Support-Vector Machine (SVM, [31]), using the LibSVM [32] library. We used the nu-SVM method with a linear kernel; the value of C was tested in the range $10^{\{-5,\ldots,1\}}$.

From a machine learning perspective, we had an extremely small dataset, but the number of diagnosed patients is very limited. Having so few examples, we did not create separate training and test sets, but applied the common solution of speaker-wise cross validation (CV): we always trained our classifier model on the data of 17 speakers, and evaluated it on the remaining one. The C meta-parameter of SVM was set in *nested* cross-validation [33]: for the 17 speakers being in the training fold in the actual CV step, we performed *another* cross-validation. We chose the C value which led to the highest AUC score in this "inner" CV test; then we trained an SVM model on the data of these 17 speakers with this complexity value, and evaluated our model on the data of the 18th speaker. This way we ensured that there was no peeking, which would have created a bias in our scores if we had used standard cross-validation.

4.4. Data Preprocessing

In our experiments we could use only one recording from only 18 speakers. In order to increase the size of our dataset, we decided to utilize shorter utterance parts in our experiments. Our hypothesis was that our temporal speech parameters remain indicative even when they are calculated from relatively short utterances. With this in mind, we split our utterances into 30 second-long segments with a 10 second overlap (regardless of actual phonetic boundaries), and treated these examples independently. After this step, we ended up with 96 of these small, equal-sized segments, significantly increasing our SVM training set sizes. Of course, we still performed our classification experiments using the leave-one-speaker-out nested cross-validation scheme; that is, one fold always consisted of all the speech segments of one speaker.

Although reporting the various classification metrics makes sense for these 30 second-long audio clips, it would be better interpretable to translate these scores to the subject level. A straightforward solution could be to determine the category of each speaker by simply taking the class hypothesis which the majority of his segments had. This, however, would be pretty hard to interpret. Thus, we decided to aggregate our predictions into speaker-level values via another approach: we calculated a speaker-normalized confusion matrix by re-weighting each speech segment with $1/k$, k being the number of segments for the given speaker. That is, a healthy control speaker with 10 speech segments, from which 7 was correctly identified, counted as 0.7 true negative and 0.3 false positive cases. After repeating SVM training and evaluation for all the folds, we were readily able to calculate accuracy, precision, recall and F-measure from this, speaker-normalized confusion matrix. Of course, as the AUC score cannot be determined from this confusion matrix, we did not calculate the AUC values in this case.

5. Results

Table 2 contains the accuracy, precision, recall, F_1 and AUC scores obtained at the **segment level**. When including all 18 temporal speech parameters in our feature set, the 70.8% classification accuracy shows a fine performance and the F_1 value of 81.3% appears to be quite high in our opinion, especially if we consider the small number of training samples available. Examining the precision and recall scores we can see that the performance is not really balanced, though, as only 74% of the segments uttered by schizophrenia patients were found, but with a roughly 90% precision score. This issue could be handled by thresholding the example-wise output posterior values [34], but we think this falls outside the scope of the present study.

Examining the results obtained by using only a subset of temporal parameters, we observe that the classification scores improved in almost every case. Comparing the temporal parameters associated with either silent or filled pauses, we can

Table 2: *The segment-level accuracy scores obtained using the various parameter sub-sets*

	Accuracy (%)				
Feature Set	Acc.	Prec.	Rec.	F_1	AUC
Full	70.8	89.7	74.4	81.3	0.514
Silence-related	76.0	94.1	77.1	84.8	0.599
Filler-related	75.0	97.1	75.0	84.6	0.435
All pause-related	79.2	92.6	80.8	86.3	0.726
Tempo + silence	80.2	97.1	79.5	87.4	0.641
Tempo + filler	70.8	91.2	73.8	81.6	0.602
Tempo + all pauses	78.1	91.2	80.5	85.5	0.694

Table 3: *The speaker-level accuracy scores obtained using the various parameter sub-sets*

	Accuracy (%)			
Feature Set	Acc.	Prec.	Rec.	F_1
Full	60.8	60.0	88.0	71.4
Silence-related	68.3	65.0	93.0	76.6
Filler-related	65.7	62.1	98.3	76.1
All pause-related	77.2	74.4	90.0	81.5
Tempo + silence	73.4	68.7	95.6	80.0
Tempo + filler	61.0	59.9	89.7	71.9
Tempo + all pauses	76.5	74.1	88.6	80.7

see that the filled pauses are less useful than the silent ones for schizophrenia identification: the classification accuracy scores of 71-75% lag behind those of 76-80% obtained when we focused on silent pauses, and the F-measure and AUC scores are higher in the latter two cases as well. Examining the tendency of the metric scores obtained, in our opinion the most useful subset of the temporal parameters proposed was those which consisted of indicators calculated based on the occurrences of hesitations regardless of whether these were silent or filled pauses. Although using the silent pause related parameters along with articulation rate and speech tempo led to slightly higher accuracy and F_1 scores, the two cases when we considered all the pauses led to consistently high metric values, and to the two highest AUC scores.

Interpreting our classification results by normalizing the number of utterances **speaker-wise** instead of considering each segment independently (see Table 3), we can see a slight drop in the metric values. What might be even more interesting is that in the tendency of the precision and recall scores, we can see just the opposite trend as we found at the segment level: now we have lower precision and quite high recall values. This is probably because patients living with schizophrenia tended to describe their previous day in much more detail than healthy controls did, therefore recordings of schizophrenic subjects were significantly longer than those of healthy controls. This then resulted in an imbalance in the number of utterances: these appeared to be 68 and 28, patients with schizophrenia and healthy controls, respectively.

Examining the different subsets of temporal speech parameters applied, it is obvious in this case that we could identify the two speaker groups most efficiently with the parameters which took both pause types into account. Clearly, the classification accuracy scores of 77.2% and 76.5% are significantly higher than either those got by using only the silent pauses (68.3% and 73.4%) or only the filled pauses (65.7% and 61.0%). The F_1 scores got this way (81.5% and 80.7%) are also the highest ones measured (i.e. 76.6-80.0% and 76.1%-71.9%, silent and filled pauses, respectively).

Regarding the utility of the different temporal parameters, the fact that the recordings of schizophrenic subjects were significantly longer than those of healthy controls might be related to positive symptoms such as circumstantialism (over-detailed speech), thought rush and systematic self-referral. Higher number of silent pauses of subjects with schizophrenia can also be explained by other symptoms (related to executive and memory functions): confused thoughts and speech. People suffering from schizophrenia could have problems in organizing their thoughts, which might be reflected in the temporal parameters of spontaneous speech (such as number of silent or filled pause occurrences).

Overall, we found a significant difference in the temporal parameters of spontaneous speech for schizophrenic speakers and healthy controls. The examined temporal parameters, focusing on articulation rate, speech tempo and hesitations permitted an accurate distinction between the two speaker groups; of course, we plan to involve more speakers in our future studies to reinforce our findings. We also plan to continue analyzing spontaneous speech on the psychosis spectrum (including schizophrenia, bipolar disorder and schizoaffective disorder) in the near future.

6. Conclusions

In this study we assumed the presence of a difference in the temporal parameters of spontaneous speech for control subjects and schizophrenic patients. With automatic speech analysis and machine learning techniques, we were able to efficiently distinguish the members of the two speaker groups. Hesitations were expected to be the main distinctive features, which was justified by our test results: the classification accuracy scores of about 77% were significantly higher than either those obtained by using only silent pauses (68-73%) and those achieved by relying only on the filled pauses (61-66%).

Our work was a pilot study: we wanted to find out whether our automatic speech analysis process could be used in the temporal description of schizophrenic spontaneous speech. We also sought to contribute to the description of the linguistic characteristics of neurodegenerative disorders, or more specifically, a subset of suprasegmental attributes. Of course, obtaining more precise findings requires an increase of the number of speakers participating in our studies. We are already collecting recordings from further schizophrenic patients as well as speakers suffering from other neurodegenerative disorders, as we also plan to investigate the temporal speech parameters of other speaker groups belonging to the psychosis spectrum in the near future.

7. Acknowledgements

This research was supported by the EU-funded Hungarian grant EFOP-3.6.1-16-2016-00008. Gábor Gosztolya was funded by the National Research, Development and Innovation Office of Hungary via contract ID-124413.

8. References

[1] T. J. Crow, "Is schizophrenia the price that Homo sapiens pays for language?" *Schizophrenia Research*, vol. 28, no. 2–3, pp. 127–141, 1997.

[2] A. P. Association, *Diagnostic and statistic manual of mental disorders (DSM-5)*. American Psychiatric Publishing, 2013.

[3] P. Kochunov, T. R. Coyle, L. M. Rowland, N. Jahanshad, P. M. Thompson, S. Kelly, X. Du, H. Sampath, H. Bruce, J. Chiappelli, M. Ryan, F. Fisseha, A. Savransky, B. Adhikari, S. Chen, S. A. Paciga, C. D. Whelan, Z. Xie, C. L. Hyde, X. Chen, C. R. Schubert, P. O'Donnell, and L. E. Hong, "Association of white matter with core cognitive deficits in patients with schizophrenia," *JAMA Psychiatry*, vol. 74, no. 9, pp. 958–966, 2017.

[4] R. Heinrichs and K. Zakzanis, "Neurocognitive deficit in schizophrenia: a quantitative review of the evidence," *Neuropsychology*, vol. 12, no. 3, pp. 426–445, 1998.

[5] A. McCleery, J. Ventura, R. Kern, K. Subotnik, D. Gretchen-Doorly, M. Green, G. Hellemann, and K. Nuechterlein, "Cognitive functioning in first-episode schizophrenia: MATRICS consensus cognitive battery (MCCB) profile of impairment," *Schizophrenia Research*, vol. 157, no. 1–3, pp. 33–39, 2014.

[6] T. Zhang, H. Li, W. Stone, K. Woodberry, L. Seidman, Y. Tang, Q. Guo, K. Zhuo, Z. Qian, H. Cui, Y. Zhu, L. Jiang, A. Chow, Y. Tang, C. Li, K. Jiang, Z. Yi, Z. Xiao, and J. Wang, "Neuropsychological impairment in prodromal, first-episode, and chronic psychosis: assessing RBANS performance," *PLoS One*, vol. 10, no. 5, pp. 33–39, 2015.

[7] R. Chan, E. Chen, E. Cheung, and H. Cheung, "Executive dysfunction in schizophrenia: relationships to clinical manifestation," *European Archives of Psychiatry and Clinical Neuroscience*, vol. 254, no. 4, pp. 256–262, 2004.

[8] J. Huang, S. Tan, S. Walsh, L. Spriggens, D. Neumann, D. Shum, and R. Chan, "Working memory dysfunctions predict social problem solving skills in schizophrenia," *Psychiatry Research*, vol. 220, no. 1–2, pp. 96–101, 2014.

[9] A. Nagels and T. Kircher, "Symptoms and neurobiological models of language in schizophrenia," in *Neurobiology of Language*, G. Hickok and S. Small, Eds. Academic Press, 2016, pp. 887–897.

[10] A. M. Pawełczyk, M. Kotlicka-Antczak, E. Łojek, A. Ruszpel, and T. Pawełczyk, "Schizophrenia patients have higher-order language and extralinguistic impairments," *Schizophrenia Research*, vol. 192, no. Feb, pp. 274–280, 2017.

[11] M. A. Covington, H. Congzhou, C. Brown, L. Naçi, J. T. McClain, B. S. Fjordbak, J. Semple, and J. Brown, "Schizophrenia and the structure of language: The linguist's view," *Schizophrenia Research*, vol. 77, no. 1, pp. 85–98, 2005.

[12] V. Rapcan, S. D'Arcy, S. Yeap, N. Afzal, J. H. Thakore, and R. B. Reilly, "Acoustic and temporal analysis of speech: A potential biomarker for schizophrenia," *Medical Engineering & Physics*, vol. 32, no. 9, pp. 1074–1079, 2010.

[13] A. M. Moe, N. J. Breitborde, M. K. Shakeel, C. J. Gallagher, and N. M. Docherty, "Idea density in the life-stories of people with schizophrenia: Associations with narrative qualities and psychiatric symptoms," *Schizophrenia Research*, vol. 172, no. 1–3, pp. 201–205, 2015.

[14] M. Rosenstein, C. Diaz-Asper, P. W. Foltz, and B. Elvevag, "A computational language approach to modeling prose recall in schizophrenia," *Cortex*, vol. 55, no. Jun, pp. 148–166, 2014.

[15] C. M. Corcoran, F. Carrillo, D. Fernández-Slezak, G. Bedi, C. Klim, D. C. Javitt, C. E. Bearden, and G. A. Cecchi, "Prediction of psychosis across protocols and risk cohorts using automated language analysis," *World Psychiatry*, vol. 17, no. 1, pp. 67–75, 2018.

[16] J. S. Bedwell, A. S. Cohen, B. J. Trachik, A. E. Deptula, and J. C. Mitchell, "Speech prosody abnormalities and specific dimensional schizotypy features: Are relationships limited to males?" *The Journal of Nervous and Mental Disease*, vol. 202, no. 10, pp. 745–751, 2014.

[17] F. Martínez-Sánchez, J. A. Muela-Martínez, P. Cortés-Soto, J. J. G. Meilán, J. A. V. Ferrándiz, A. E. Caparrós, and I. M. P. Valverde, "Can the acoustic analysis of expressive prosody discriminate schizophrenia?" *The Spanish Journal of Psychology*, vol. 18, no. 86, pp. 1–9, 2015.

[18] M. Alpert, A. Kotsaftis, and E. R. Pouget, "At issue: Speech fluency and schizophrenic negative signs," *Schizophrenia Bulletin*, vol. 23, no. 2, pp. 171–177, 1997.

[19] K. Matsumoto, T. T. J. Kircher, P. R. A. Stokes, M. J. Brammer, P. F. Liddle, and P. K. McGuire, "Frequency and neural correlates of pauses in patients with formal thought disorder," *Frontiers in Psychiatry*, vol. 4, no. Oct, pp. 67–75, 2013.

[20] L. Tóth, G. Gosztolya, V. Vincze, I. Hoffmann, G. Szatlóczki, E. Biró, F. Zsura, M. Pákáski, and J. Kálmán, "Automatic detection of mild cognitive impairment from spontaneous speech using ASR," in *Proceedings of Interspeech*, Dresden, Germany, Sep 2015, pp. 2694–2698.

[21] I. Hoffmann, L. Tóth, G. Gosztolya, G. Szatlóczki, V. Vincze, E. Kárpáti, M. Pákáski, and J. Kálmán, "Beszédfelismerés alapú eljárás az enyhe kognitív zavar automatikus felismerésére spontán beszéd alapján (in Hungarian)," *Általános nyelvészeti tanulmányok*, vol. 29, pp. 385–405, 2017.

[22] L. Tóth, I. Hoffmann, G. Gosztolya, V. Vincze, G. Szatlóczki, Z. Bánréti, M. Pákáski, and J. Kálmán, "A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech," *Current Alzheimer Research*, vol. 15, no. 2, pp. 130–138, 2018.

[23] C. Laske, H. R. Sohrabi, S. M. Frost, K. L. de Ipiña, P. Garrard, M. Buscema, J. Dauwels, S. R. Soekadar, S. Mueller, C. Linnemann, S. A. Bridenbaugh, Y. Kanagasingam, R. N. Martins, and S. E. O'Bryant, "Innovative diagnostic tools for early detection of Alzheimer's disease (in press)," *Alzheimer's & Dementia*, 2015.

[24] B. Roark, M. Mitchell, J.-P. Hosom, K. Hollingshead, and J. Kaye, "Spoken language derived measures for detecting mild cognitive impairment," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2081–2090, 2011.

[25] I. Hoffmann, D. Németh, C. Dye, M. Pákáski, T. Irinyi, and J. Kálmán, "Temporal parameters of spontaneous speech in Alzheimer's disease," *International Journal of Speech-Language Pathology*, vol. 12, no. 1, pp. 29–34, 2010.

[26] K. L. de Ipiña, J. B. Alonso, J. Solé-Casals, N. Barroso, P. Henriquez, M. Faundez-Zanuy, C. M. Travieso, M. Ecay-Torres, P. Martínez-Lage, and H. Eguiraun, "On automatic diagnosis of Alzheimer's disease based on spontaneous speech analysis and emotional temperature," *Cognitive Computation*, vol. 7, no. 1, pp. 44–55, 2015.

[27] M. Folstein, S. Folstein, and P. McHugh, "Mini-mental state: A practical method for grading the cognitive state of patients for the clinician," *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189–198, 1975.

[28] M. Gósy, "BEA a multifunctional Hungarian spoken language database," *The Phonetician*, vol. 105, no. 106, pp. 50–61, 2012.

[29] L. Tóth, "Phone recognition with hierarchical Convolutional Deep Maxout Networks," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2015, no. 25, pp. 1–13, 2015.

[30] P. Garrard, V. Rentoumi, B. Gesierich, B. Miller, and M. L. Gorno-Tempini, "Machine learning approaches to diagnosis and laterality effects in semantic dementia discourse," *Cortex*, vol. 55, pp. 122–129, 2014.

[31] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[32] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 1–27, 2011.

[33] G. C. Cawley and N. L. C. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *Journal of Machine Learning Research*, vol. 11, no. Jul, pp. 2079–2107, 2010.

[34] W. Waegeman, K. Dembczynski, A. Jachnik, W. Cheng, and E. Hüllermeier, "On the Bayes-optimality of F-measure maximizers," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 3333–3388, 2014.