



Towards Supporting an Early Diagnosis of Multiple Sclerosis using Vocal Features

Monica Gonzalez-Machorro^{1,2*}, Pascal Hecker^{1,2*}, Uwe D. Reichel¹, Helly N. Hammer³, Robert Hoepner³, Lisa Pedrotti³, Alisha Zmutt³, Hesam Sagha¹, Johan van Beek⁴, Florian Eyben¹, Dagmar M. Schuller¹, Björn W. Schuller^{1,5,6}, Bert Arnrich²

¹audEERING GmbH, Gilching, Germany

²Digital Health – Connected Healthcare, Hasso Plattner Institute, University of Potsdam, Germany

³Department of Neurology, Inselspital, Bern University Hospital and University of Bern, Freiburgstrasse, Bern, Switzerland

⁴Biogen Switzerland AG, Baar, Switzerland

⁵Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany

⁶GLAM – Group on Language, Audio, & Music, Imperial College, UK

mgonzalez@audengineering.com, phecker@audengineering.com

Abstract

Multiple sclerosis (MS) is a neuroinflammatory disease that affects millions of people worldwide. Since dysarthria is prominent in people with MS (pwMS), this paper aims to identify acoustic features that differ between people with MS and healthy controls (HC). Additionally, we develop automatic classification methods to distinguish between pwMS and HC. In this work, we present a new dataset of a German-speaking cohort which contains 39 patients with low disability of relapsing MS and 16 HC. Findings suggest that certain interpretable speech features could be useful in diagnosing MS, and that machine learning methods could potentially support fast and unobtrusive screening in clinical practice. The study emphasises the importance of analysing free speech compared to read speech.

Index Terms: multiple sclerosis, speech analysis, dysarthria, German language, machine learning

1. Introduction

Multiple sclerosis (MS) is a neuroinflammatory disease that affects a population of around 2.8 million people worldwide [1]. In the course of this disease, the myelin sheath of the glia cells is getting attacked by the patient's immune system. The resulting lesions in the brain and spinal cord affect a variety of functions of the central nervous system.

Relapsing MS (RMS) is the most common course of MS with a prevalence of around 80%, and consists of alternating periods of asymptomatic and symptomatic phases. This course could transition to secondary progressive MS (SPMS) [2, 3], a stage in which symptoms worsen gradually until complete disability.

Among others, roughly 45% of affected patients suffer from dysarthria [4] – making it the most common speech disorder among people with MS (pwMS). Dysarthria is reported to be one of the earliest symptoms of the disease in pwMS [5], and is a mix of ataxic and spastic speech characteristics [6].

In this paper, we describe a unique dataset with speech of fully ambulatory pwMS in the German language. The aim is to assess whether the voice can be used to reliably screen for pwMS with low disability. To do so, we employ three well-known speech tasks 1) to identify acoustic features that differ

significantly between pwMS and healthy control (HC), and 2) to develop supervised machine learning (ML) methods to distinguish pwMS against HC.

Speech analysis for healthcare purposes yields the promise to enable fast and unobtrusive screening [7]. Small perturbations in the voice have been shown to correlate with various neurological disorders [8]. The research community has approached the quantification of neurological disorders with various challenges to assess the current state-of-the-art. Some exemplary approaches were to describe Parkinson's disease (PD) [9] and Alzheimer's disease (AD) [10].

In clinical practice, an initial screening system that supplements clinical routine shows great promise to support overburdened clinicians. This is because diagnosis encompasses multiple modalities, such as magnetic resonance imaging (MRI) and laboratory diagnostics [11]. In the context of MS, even though the disease manifests itself through specific brain lesions, the diagnosis usually depends on ruling out other diseases, since there are no specific screening tests [12].

1.1. Related work

Several studies have investigated speech changes in pwMS. Acoustic features that have been identified in pwMS are, among others, a loss of speech intensity and pitch control, vocal instability, higher jitter and shimmer, slower speech rate, compromised intelligibility, reduced breath support, as well as a decrease of the vowel space area [4, 8, 13]. An increase of harmonics-to-noise-ratio (HNR) has also been reported [13], which suggests a higher articulation effort. The maximum slope of the second formant (F2) is particularly affected in pwMS and is a good indicator to distinguish between mild and moderate dysarthria [14]. Studies have also concluded that dysarthria in pwMS is predominantly mild [15]. These findings suggest that certain speech features could be useful in diagnosing MS.

Further, the frequency and severity of dysarthria are associated with an expanded disability status scale (EDSS) score higher than 4 [16], even though speech impairment is present in all the disease stages [16]. The EDSS score depends on seven different functional systems: visual, brain-stem, pyramidal, cerebellar, sensory, bowel and bladder, and cerebral. An EDSS below 4 refers to patients who are ambulatory. An EDSS between 4 and 5 indicates a limited walking range. Finally, an EDSS of up to 8 relates to a requirement for walking assistance.

*The first two authors contributed equally and are co-first authors.

Table 1: Overview over the participants’ metadata across the data partitions.

	All	RRMS	HC
Whole dataset			
Participants	55	39	16
Gender	39F, 16M	29F, 10M	10F, 6M
Age	36.62±10.5	37.26±10.38	35.06±10.99
EDSS	-	1.08±1.0	-
Train set			
Participants	44	32	12
Gender	29F, 15M	22F, 10M	7F, 5M
Age	37.2±10.31	38.47±10.27	33.83±10.07
EDSS	-	1.2±1.0	-
Test set			
Participants	11	7	4
Gender	10F, 1M	7F, 0M	3F, 1M
Age	34.27±11.45	31.7±19.67	38.75±14.41
EDSS	-	0.81±1.19	-

Compared to other neurodegenerative diseases, such as AD and PD, automatic classification of MS patients from a healthy cohort using ML methods appears to be less explored in the literature. Rusz et al. 2018 reported 78% accuracy between asymptomatic pwMS and control speakers using a linear support vector machine (SVM) [17]. Gosztolya et al. 2022 used deep neural network (DNN) embeddings to classify some specific non-word utterances of pwMS and achieved the accuracy of 90% [18].

This contribution is divided as follows: in section 2, we describe the dataset employed and the experiment design. The outcomes of these experiments are presented in section 3 and put into context in section 4. Section 5 provides the conclusions.

2. Materials and Methods

2.1. Data

Data was collected at the Department of Neurology, Inselspital, Bern, Switzerland and the University of Bern, Switzerland. This observational clinical study was approved by the ethics committee of the Canton of Bern, Switzerland (approval no. 2021-02423, ClinicalTrials.gov identifier: NCT05561621), which unfortunately does not permit the publication of the recorded data. All participants provided their written informed consent.

The recording setup consisted of a laptop computer and a microphone system: an AKG 555L headset condenser microphone and a PreSonus Studio 18 audio interface for A/D conversion. The native sampling rate was 44,100 Hz. The framework employed for voice recording was AI SoundLab¹ [19], which is a web app, in which each patient could navigate through a voice recording session under the supervision of the study nurse. The voice recording session comprised several speech tasks, including verbal responses of the symbol digit modalities test (SDMT), sustained utterances, and diadochokinesis. Further included were free speech tasks and read speech tasks: the text passages ‘North Wind and the Sun (NWS)’ (German version) [20] and the ‘Buttergeschichte’, which are both widely used in the field of

phonetics). Each session took around 60 minutes.

For brevity, within the scope of this paper, we addressed the following three speech tasks: the picture description task (PDT) of the Western Aphasia Battery – Revised [21], the NWS (read twice during each session: once at the very beginning and once at the very end), and all read speech tasks combined (NWS and Buttergeschichte). We selected these tasks since they are commonly used in similar neurodegenerative diseases and contain more speech segments, which can be useful for ML-based analysis.

The dataset consisted of 55 speakers, 39 pwMS, and 16 HC participants without any known neurological disorders. The inclusion criteria for the MS patients were as follows: 1) diagnosis with relapsing MS (RMS) due to its prominence, 2) age range 18-60 years, 3) ambulatory patients with low disability (EDSS score lower than four), 4) the capability of written informed consent, and 5) fluency in the German language.

2.2. Data processing

For data processing, audio files were downsampled to 16 kHz and segmented by applying a voice activity detection (VAD) algorithm², which due to license restrictions is not available as open source, but its underlying architecture is based on the Speech & Music Interpretation by Large-space Extraction (openSMILE) interface. The minimum turn length was 0.76 s, the maximum turn length 6.0 s, the length of speech until a segment start was detected was set to 0.15 s, and the length of non-speech, until a segment end was detected, was set to 0.25 s. Manual audio quality checks were carried out before analyses. As a variable for the experiments, the signal volume was loudness-normalised with -24 dB LUFS.

We defined a fixed, hold-out test set, which was speaker-disjunct from the train set. We employ the variables sex and age to stratify each split. The test set corresponds to 20% of the data. An algorithm, that spawned 30 different combinations, was used to evaluate the quality of each split using the Jensen-Shannon divergence. The distance was measured between the target (MS/HC), sex, and age distribution in the train and test sets. The test split with the lowest distance from the train set was selected. The resulting test set has an information radius of 0.08. This approach was implemented to ensure that the test set was as representative as possible of the train set. Table 1 respectively summarises the participants’ metadata across the different dataset partitions. Using the VAD algorithm, the following number of speech segments was obtained: a) PDT: 652 segments (train: 528, test: 124); b) read speech tasks: 2,348 segments (train: 1,874, test: 474); c) NWS: 1,133 segments (train: 899, test: 234).

2.3. Methods

The extended Geneva minimalistic acoustic parameter set (eGeMAPS) [22] was extracted using the openSMILE feature extraction tool by the audEERING GmbH [23]. It contains 88 acoustic features and previous work has reported promising results in speech disorders [24]. Speech summary statistics of the eGeMAPS feature set were extracted. Summary statistics can be understood as aggregated information over an entire utterance. Examples of these are the mean, median, standard deviation (std), and skewness [22].

The statistical method employed for an exploratory feature analysis was a two-tailed Mann-Whitney U test, since val-

¹aisoundlab.audeering.com, powered by audEERING GmbH

²powered by audEERING GmbH

Table 2: Interpretable, significant acoustic features to distinguish MS from HC in the picture description task (PDT).

Features	Description	Interpretation	r	U_{MS}	p -value
F1 amplitude LogRel	Mean F1 amplitude relative to amplitude of first harmonic	Voice excitation strength	-0.40	9.025	< 0.001
F2 amplitude LogRel	Mean F2 amplitude relative to amplitude of first harmonic	Voice excitation strength	-0.39	7.65	< 0.001
Unvoiced Segment Length	Mean of unvoiced segment lengths	Phonatory control	0.38	6.26	< 0.02
Spectral Flux std	Standard deviation of the distance between spectra of adjacent frames	Speech rate variation and hesitations	0.28	4.78	< 0.05
Loudness peaks per second	Number of loudness peaks per second	Speech rate	-0.27	4.78	< 0.05

ues were not normally distributed. The significance level was set to 0.05. To identify significant features, we followed the methodology used in Bartl-Pokorny et al. 2021 [25] and calculated the correlation effect size as the absolute value of the Pearson product-moment correlation coefficient (r). For calculating r , we encoded HC as 0 and MS as 1. This way, positive correlations can be interpreted as higher feature values for MS. Those features, which differ significantly, are ranked based on effect size and identified as ‘relevant’.

We employ support vector machine (SVM), k-nearest neighbours (KNN), and random forest (RF) models through the scikit-learn [26] Python package. All models are optimised using grid search. For SVM, we optimise the cost parameter [3, 2.5, 2, 1.5, 1, 0.5, 0.1, 0.005, 0.05, 0.01, 0.001, 0.0001] and we selected a linear kernel to be consistent with related work [9]. Parameters optimised for RF are the number of estimators [50, 100, 300, 500, 800, 1000, 3000], the criterion [‘gini’, ‘entropy’], the minimum samples per split [2, 3], and bootstrapping [True, False]. For KNN, we optimise the leaf size in a range from 20 to 40, the number of neighbours from 1 to 15, and the power parameter [2, 3]. For all models, parameters are optimised on the train set using a leave-one-speaker-out (LOSO) cross-validation method. With the optimal hyperparameters, the respective model is fitted on the entire train set. The best performing model is then evaluated on the test set.

For evaluation, due to class imbalance, we report unweighted average recall (UAR). Area under the curve (AUC) and average precision score (APS) are reported, and the best-performing receiver operating characteristic (ROC) curve is presented. These metrics account for false positives and false negatives.

3. Results

3.1. Feature Analysis

Statistical analyses were performed on the train set for each speech task. Within the train set, 32 MS and twelve HC speakers are included (see Table 1). We found ten features for the PDT and one feature (arithmetic mean for F1 bandwidth) for the NWS for read speech respectively to distinguish between MS and HC. All of these features have an effect size higher than 0.20.

Due to space limitations, from the identified features the top 5 features are interpreted and discussed in this paper. For interpretability, the real r is reported. Table 2 shows the top

Table 3: Performance of the predictive modelling approaches. Unweighted average recall (UAR); picture description task (PDT); North Wind and the Sun (NWS); support vector machine (SVM); k-nearest neighbours (KNN); random forest (RF) receiver operating characteristic (ROC) area under the curve (AUC); average precision score (APS).

Task	Model	UAR	ROC AUC	AP
PDT	SVM	71.7%	76.0%	65.6%
	KNN	65.4%	70.2%	62.9%
	RF	56.8%	69.1%	61.7%
Read Speech	SVM	59.9%	62.8%	51.4%
	KNN	53.5%	58.7%	49.2%
	RF	50.0%	77.3%	65.8%
NWS	SVM	56.9%	61.3%	50.0%
	KNN	60.6%	68.9%	59.1%
	RF	55.0%	68.7%	60.0%

five features obtained in the PDT. We portray the features for this speech task in detail since it is one of the most commonly used ones in similar neurodegenerative diseases such as AD [27]. The top five features for PDT are: arithmetic mean of F1 relative amplitude and F2 relative amplitude, unvoiced segment lengths, spectral flux std normalised by mean, and loudness peaks per second. In line with previous work, we expect to find that pwMS present a more unstable phonatory control, lower voice excitation strength, decrease in loudness, and lower speech rate compared to healthy speech. Table 2 presents the feature description and interpretation, r value, U statistic corresponding with the MS sample, and p -value.

3.2. Predictive Modelling

Table 3 shows the results of the predictive modelling approaches. Using all eGeMAPS features, the best-performing model reaches an UAR of 71.1% using a SVM on the PDT. NWS performance is higher with KNN. Figure 1 depicts the ROC curve of the model with the highest UAR. Due to SVM obtaining the highest

Table 4: Performance of the predictive modelling approaches when only regarding the significantly correlating features described in section 3.1. Abbreviations as in Table 3.

Task	Model	UAR	ROC AUC	AP
PDT	SVM	67.2%	73.2%	60.6%
Read Speech	SVM	65.8%	74.1%	61.7%
NWS	SVM	64.3%	71.4%	59.0%

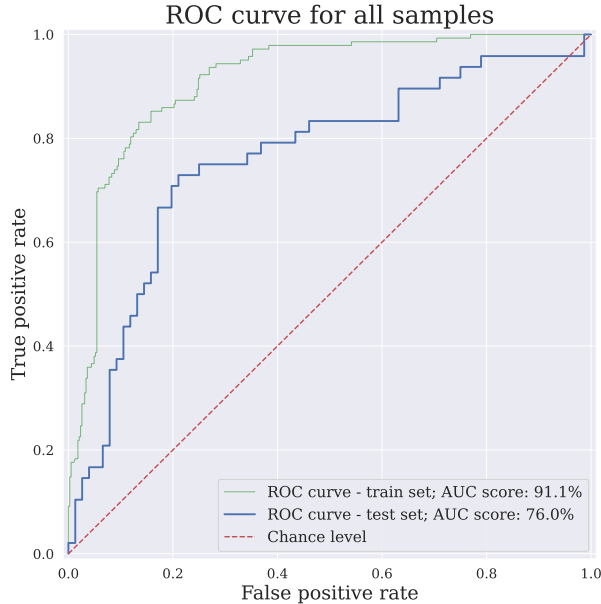


Figure 1: Receiver operating characteristic (ROC) curves of the best-performing model for the picture description task. 71.7% UAR, see Table 3. Green portrays the ROC curve of the model on all samples of the train partition and blue the one on all samples in the test partition.

performance in Table 3, we investigate the significant features identified in section 3.1. using only SVM. In Table 4, models with the subset of significant features described in section 3.1 were evaluated.

4. Discussion

In this study, we analysed acoustic features and ML models to distinguish between pwMS and HC when eliciting the PDT, NWS, and a combination of the NWS and the Buttergeschichte task. For initial screening, a dataset representing a cohort of patients that would be encountered in an ambulatory setting for initial diagnosis, is important. Therefore, we recruited a cohort of pwMS with a particularly low disability score (Table 1).

Feature interpretability is key in speech analysis for health-care, since it is needed for a potential clinical application, where patients and clinicians need to understand the speech changes in a particular disorder or disease. In this case, eGeMAPS is an adequate feature set, since it was curated by specialists in psychology, linguistics, medicine, and signal engineering [22].

Therefore, from the top five features presented in Table 2, we found that F1 and F2 relative amplitude are lower for MS speech, which suggests that pwMS present a lower voice exci-

tation strength than the HC. Due to muscle weakness being a common condition in MS, this would be the expected behaviour of MS speech. Furthermore, the lengths of unvoiced segments are higher in MS speech compared to healthy speech. This finding suggests that pwMS present lower phonatory control when switching from unvoicing to voicing. Nonetheless, a more indicative result could be found if the speech task would be more controlled, for example, a sustained utterance, as done in [25]. The std of spectral flux shows that pwMS have a higher speech rate variation, which suggests, among other things, that MS speech contains more filled pauses and unsystematic speech rate changes than HC. This finding is of special interest when regarding the PDT, since it hints at a higher cognitive load. This finding aligns with the fact that cognitive impairment is observed in MS even at early stages of the disease [28]. The std of spectral flux can also indicate varying amount of coarticulation. This might be explained by an increased proportion of articulatory target undershoot for pwMS. The final feature, loudness peaks per second, indicates a lower speech rate: results hint that MS presents lower speech rate and higher speech variation.

As for the ML experiments, models based on SVMs to distinguish pwMS and HC perform the best and reach an UAR of 71.7% for the PDT. This is slightly lower than the 78% accuracy, which was reported by [17]. Lower performance reported in our paper might be attributable to the use of diadochokinetic data in [17]. When regarding the subset of significant features (Table 4), performance decreased for the three speech tasks. This decline in performance might be due to excessively constraining the function space of the ML models.

Moreover, the PDT performed best across the ML models. This result is consistent with previous work for other neurodegenerative diseases, in which the PDT was described as a valuable clinical tool [27, 29]. Thus, our findings suggest that the PDT, a free speech task, better embodies speech differences between pwMS and HC than the employed read speech tasks.

5. Conclusions

In this study, we demonstrated that speech analysis can be used as a tool to distinguish between the speech of pwMS with low disability and HC. By using acoustic features, we reached an UAR of up to 71.1%. Feature interpretation of low disability MS speech is consistent with related work and dysarthric speech characteristics. Therefore, this paper can be considered as a step towards an assistive tool for clinicians, in which acoustic features could be used as additional tools to support MS diagnosis.

Future work will focus on including further participants and longitudinal data collection for monitoring MS progression. In order to have an initial screening tool based on speech analysis, it is important to investigate the difference between patients with low and high EDSS scores and to further investigate model interpretability to understand resulting predictions. It will also be valuable to identify specific features for pwMS in the German language and across multiple languages. Future work should also systematically study speech characteristics of different MS types.

6. Acknowledgements

We thank all speakers, who donated their data and the AI Sound-Lab team for customising the platform for this data collection. Thanks to Moritz Jacobshagen and Johan van Beek for coordinating Biogen's sponsorship of the study.

7. References

- [1] R. King, "Atlas of ms, 3rd edition," *The Multiple Sclerosis International Federation*, 2020.
- [2] M. Barzegar, S. Najdaghi, A. Afshari-Safavi, N. Nehzat, O. Mir-mosayyeb, and V. Shaygannejad, "Early predictors of conversion to secondary progressive multiple sclerosis," *Multiple Sclerosis and Related Disorders*, vol. 54, no. 103115, p. 103115, 2021.
- [3] J. W. L. Brown, A. Coles, D. Horakova, E. Havrdova, G. Izquierdo, A. Prat, M. Girard, P. Duquette, M. Trojano, A. Lugaresi, R. Bergamaschi, P. Grammond, R. Alroughani, R. Hupperts, P. McCombe, V. Van Pesch, P. Sola, D. Ferraro, F. Grand'Maison, M. Terzi, J. Lechner-Scott, S. Flechter, M. Slee, V. Shaygannejad, E. Pucci, F. Granella, V. Jokubaitis, M. Willis, C. Rice, N. Scolding, A. Wilkins, O. R. Pearson, T. Ziemssen, M. Hutchinson, K. Harding, J. Jones, C. McGuigan, H. Butzkueven, T. Kalincik, N. Robertson, and MSBase Study Group, "Association of initial disease-modifying therapy with later conversion to secondary progressive multiple sclerosis," *JAMA*, vol. 321, no. 2, pp. 175–187, 2019.
- [4] G. Noffs, T. Perera, S. C. Kolbe, C. J. Shanahan, F. M. C. Boonstra, A. H. Evans, H. Butzkueven, A. van der Walt, and A. P. Vogel, "What speech can tell us: A systematic review of dysarthria characteristics in multiple sclerosis," *Autoimmunity reviews*, vol. 17, no. 12, pp. 1202–1209, 2018.
- [5] A. V. Feijó, M. A. Parente, M. Behlau, S. Haussen, M. C. de Vecino, and B. C. d. F. Martignago, "Acoustic analysis of voice in multiple sclerosis patients," *Journal of Voice*, vol. 18, no. 3, pp. 341–347, 2004.
- [6] L. Hartelius, B. Runmarker, and O. Andersen, "Prevalence and characteristics of dysarthria in a multiple-sclerosis incidence cohort: Relation to neurological data," *Folia Phoniatrica et Logopaedica*, vol. 52, pp. 160–177, 2000.
- [7] M. Milling, F. B. Pokorny, K. D. Bartl-Pokorny, and B. W. Schuller, "Is speech the new blood? recent progress in ai-based disease detection from audio in a nutshell," *Frontiers in Digital Health*, vol. 4, 2022.
- [8] P. Hecker, N. Steckhan, F. Eyben, B. W. Schuller, and B. Arnrich, "Voice analysis for neurological disorder recognition—a systematic review and perspective on emerging trends," *Frontiers in Digital Health*, vol. 4, 2022.
- [9] B. Schuller, S. Steidl, A. Batliner, S. Hantke, F. Hönig, J. R. Orozco-Arroyave, E. Nöth, Y. Zhang, and F. Weninger, "The interspeech 2015 computational paralinguistics challenge: Nativeness, parkinson's & eating condition," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.
- [10] S. Luz, F. Haider, S. de la Fuente Garcia, D. Fromm, and B. MacWhinney, "Alzheimer's dementia recognition through spontaneous speech," *Frontiers in computer science*, vol. 3, 2021.
- [11] M. P. McGinley, C. H. Goldschmidt, and A. D. Rae-Grant, "Diagnosis and treatment of multiple sclerosis: a review," *Jama*, vol. 325, no. 8, pp. 765–779, 2021.
- [12] B. G. Weinshenker, B. Bass, G. P. Rice, J. Noseworthy, W. Carriere, J. Baskerville, and G. C. Ebers, "The natural history of multiple sclerosis: a geographically based study: I. clinical course and disability," *Brain*, vol. 112, no. 1, pp. 133–146, 1989.
- [13] P. Vizza, D. Mirarchi, G. Tradigo, M. Redavide, R. B. Bossio, and P. Veltri, "Vocal signal analysis in patients affected by multiple sclerosis," *Procedia Computer Science*, vol. 108, pp. 1205–1214, 2017.
- [14] K. M. Rosen, J. V. Goozée, and B. E. Murdoch, "Examining the effects of multiple sclerosis on speech production: does phonetic structure matter?" *J. Commun. Disord.*, vol. 41, no. 1, pp. 49–69, 2008.
- [15] V. Piacentini, I. Mauri, D. Cattaneo, M. Gilardone, A. Montesano, and A. Schindler, "Relationship between quality of life and dysarthria in patients with multiple sclerosis," *Archives of Physical Medicine and Rehabilitation*, vol. 95, no. 11, pp. 2047–2054, 2014.
- [16] G. Noffs, F. M. C. Boonstra, T. Perera, H. Butzkueven, S. C. Kolbe, F. Maldonado, L. E. Cofre Lizama, M. P. Galea, J. Stankovich, A. Evans, A. van der Walt, and A. P. Vogel, "Speech metrics, general disability, brain imaging and quality of life in multiple sclerosis," *European Journal of Neurology*, vol. 28, no. 1, pp. 259–268, 2021.
- [17] J. Rusz, B. Benova, H. Ruzickova, M. Novotny, T. Tykalova, J. Hlavnicka, T. Uher, M. Vaneckova, M. Andelova, K. Novotna, L. Kadrnokova, and D. Horakova, "Characteristics of motor speech phenotypes in multiple sclerosis," *Multiple Sclerosis and Related Disorders*, vol. 19, pp. 62–69, 2018.
- [18] G. Gosztolya, L. Tóth, V. Svindt, J. Bóna, and I. Hoffmann, "Using acoustic deep neural network embeddings to detect multiple sclerosis from speech," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 6927–6931.
- [19] P. Hecker, F. B. Pokorny, K. D. Bartl-Pokorny, U. Reichel, Z. Ren, S. Hantke, F. Eyben, D. M. Schuller, B. Arnrich, and B. W. Schuller, "Speaking corona? human and machine recognition of covid-19 from voice," in *Proceedings Interspeech 2021*, 2021, pp. 1029–1033.
- [20] I. P. Association, *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press, 1999.
- [21] A. Kertesz, "Western aphasia battery—revised," 2007.
- [22] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan, and K. P. Truong, "The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing," *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 190–202, 2016.
- [23] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: The munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM International Conference on Multimedia*. Association for Computing Machinery, 2010, pp. 1459–1462.
- [24] W. Xue, C. Cucchiari, R. van Hout, and H. Strik, "Acoustic correlates of speech intelligibility: the usability of the egemaps feature set for atypical speech," in *Proceedings SLATE 2019: 8th ISCA Workshop on Speech and Language Technology in Education*, 2019, pp. 48–52.
- [25] K. D. Bartl-Pokorny, F. B. Pokorny, A. Batliner, S. Amiriparian, A. Semertzidou, F. Eyben, E. Kramer, F. Schmidt, R. Schönweiler, M. Wehler, and B. W. Schuller, "The voice of COVID-19: Acoustic correlates of infection in sustained vowels," *The Journal of the Acoustical Society of America*, vol. 149, no. 6, pp. 4377–4966, 2021.
- [26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [27] A. Slegers, R.-P. Filiou, M. Montembeault, and S. M. Brambati, "Connected speech features from picture description in alzheimer's disease: A systematic review," *Journal of Alzheimer's Disease*, vol. 65, no. 2, pp. 519–542, 2018.
- [28] G. C. DeLuca, R. L. Yates, H. Beale, and S. A. Morrow, "Cognitive impairment in multiple sclerosis: clinical, radiologic and pathologic insights," *Brain Pathology*, vol. 25, no. 1, pp. 79–98, 2015.
- [29] K. D. Mueller, B. Hermann, J. Mecollari, and L. S. Turkstra, "Connected speech and language in mild cognitive impairment and alzheimer's disease: A review of picture description tasks," *Journal of Clinical and Experimental Neuropsychology*, vol. 40, no. 9, pp. 917–939, 2018.