TITLE

Speech metrics, general disability, brain imaging and quality of life in MS

RUNNING HEAD

Speech analytics of MS disease severity

AUTHORS

Gustavo Noffs[1,2], Frederique M. C. Boonstra[3], Thushara Perera[4,5], Helmut Butzkueven[3], Scott C. Kolbe[3] Fernanda Maldonado[1], Eduardo C. Lizama[6,7], Mary P Galea[6,7], Jim Stankovich[3], Andrew Evans, MD [2,4], Anneke van der Walt, PhD [2,3,4,*] and Adam P. Vogel, PhD [1,4,8,9,*]

1. Centre for Neuroscience of Speech, University of Melbourne, Victoria, Australia
2. Department of Neurology, Royal Melbourne Hospital, Australia
3. Department of Neuroscience, Central Clinical School, Monash University, Australia
4. The Bionics Institute, Australia
5. Department of Medical Bionics, University of Melbourne, Australia
6. Department of Medicine, University of Melbourne, Australia
7. Australia Rehabilitation Research Centre, Royal Melbourne Hospital, Australia
8. Department of Neurodegeneration, Hertie Institute for Clinical Brain Research, University of Tübingen, Germany
9. Redenlab, Victoria, Australia

* joint senior authors with equal contribution

Corresponding author

Gustavo Noffs

550 Swanston Street, University of Melbourne, Carlton 3053, Victoria, Australia

Gustavo.noffs@unimelb.edu.au

Phone: +61 39 9035 3322

Title character count including spaces: 75

Running head character count including spaces: 39

Word count

Abstract (with headings): 198

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/ENE.14523

Main text (excluding table heads and captions): 3,463

Number of references: 47

Tables: 5 (+3 supplementary)

Figures: 1

DR. GUSTAVO NOFFS (Orcid ID : 0000-0001-8905-7879)

DR. ADAM P VOGEL (Orcid ID : 0000-0002-3505-2631)

Article type : Original Article

ABSTRACT

BACKGROUND: Objective measurement of speech has shown promising results to monitor disease state in Multiple Sclerosis. In this study, we characterize the relationship between disease severity and speech metrics, through perceptual (listener based) and objective acoustic analysis. We further look at deviations of acoustic metrics in people with no perceivable dysarthria. METHODS: Correlations and regression were calculated between speech measurements and disability scores, brain volume, lesion load, and quality of life. Speech measurements were further compared between three subgroups of increasing overall neurological disability – mild (as rated by the Expanded Disability Status Scale ≤2.5), moderate (≥3 and ≤5.5) and severe (≥6). RESULTS: Clinical speech impairment occurred majorly in people with severe disability. An experimental acoustic composite score differentiated mild from moderate ($p<0.001$) and moderate from severe subgroups ($p=0.003$), and correlated with overall neurological disability ($r=0.6$, $p<0.001$), quality of life ($r=0.5$, $p<0.001$), white matter volume ($r=0.3$, $p=0.007$) and lesion load ($r=0.3$, $p=0.008$). Acoustic metrics also correlated with disability scores in people with no perceivable dysarthria. DISCUSSION: Acoustic analysis offers a valuable insight into the development of speech impairment in MS. These results highlight the potential of automated analysis of speech to assist in monitoring disease progression and treatment response.

1. INTRODUCTION

Dysarthria, (impairment of speech execution), affects around 50% of people with multiple sclerosis (MS) at some stage of their illness.[1, 2] Studies using standardized assessments by trained raters, herein perceptual analysis, highlighted that dysarthria in people with MS (pwMS) is often characterized by slow speech rate, increased frequency of pauses and poor respiratory support[3]. The degree of impairment is typically mild and intelligibility is marginally affected[3].

Subtle impairment in speech production might not translate into overt dysarthria. Speech production relies not only on motor control but also affective (mood)[4-6] and cognitive[7, 8] processes of the central nervous system. Yet, the reported prevalence of cerebellar, cognitive and affective symptoms[9-11] are disproportionally larger than the prevalence of perceptually assessed dysarthria in MS. Arguably, impairment of the former systems affects speech but remain undetected by conventional perceptual assessments.

Measurements from objective acoustic analysis of speech, related but not identical to clinical characteristics of dysarthria, seem to be more sensitive to neurological dysfunction. What we perceive as the loudness of sound, for example, is a non-linear function of variation in air pressure, termed sound intensity in acoustic analysis.[12] We can further perceive instability in loudness, often defined as voice tremor.[13] Yet, acoustic measurement of sound intensity instability can separate individuals with MS from healthy controls (HC) with 90% accuracy[14] while experienced raters perceive such instability in only about 35% of pwMS.[1] In a recent study[15], frequency instability and slow articulation rate were observed in pwMS without any clinically detectable dysarthria. Evidence from Huntington's disease, which has a well-defined pre-symptomatic stage, points to objective changes in speech occurring before not only dysarthria but any disease-related symptom becomes overt.[16] Thus, acoustic measures appear to be more sensitive than perceptual scores to the presence of neurological damage.

Through objective measurement of speech, acoustic analysis can inform on change of neurological function. The relationship between specific acoustic metrics and neurological dysfunction[17-19] or brain volumetrics[15, 19] was recently described in MS. The primary aim of the current study was to describe the relationship between speech measurements and general neurological impairment, brain volume, brain lesion load and quality of life in MS in a single cohort. The secondary aims were to determine the association between acoustic metrics and

neurological dysfunction in non-dysarthric pwMS, and to estimate at which level of neurological disability each speech metric changes. The current study also expands on recent findings by including the analysis of unscripted speech recordings, desirable for frequent and non-invasive monitoring.

2. METHODS

2.1. Participants

People with a definite diagnosis of relapsing-remitting or secondary-progressive MS[20] were sequentially recruited from tertiary MS clinics. All participants were recruited to participate regardless of speech symptoms. Exclusion criteria were: 1) the presence of other neurological or neuromuscular disorder; 2) MS relapse within the last 3 months; 3) impaired vision or hearing resulting in an inability to complete the testing protocol; and 4) speech impairment not related to MS (e.g. stuttering, vocal tics). Age and sex-matched volunteers with no history of neurological or muscular diseases were recruited as healthy controls (HC). The study was approved by the relevant Ethics Committees and written informed consent was obtained from all participants.

2.2. Neurological assessment

All participants underwent a detailed physical and neurological examination by a qualified neurologist who determined all MS participants' Expanded Disability Status Scale score (EDSS)[21].

2.3. Speech assessment

To avoid overfitting and mitigate type 2 error, we restricted the analysis of speech variables to only those shown to be impaired in MS by previous studies[3]. Speech variables were organized into three major domains (Table 1) and assessed through both perceptual and acoustic methods.

[INSERT TABLE 1 ABOUT HERE]

2.3.1. Speech sample

We elicited speech using five standardized speech tasks which fit along a spectrum of automaticity[22], from simple to complex (phonetically and/or cognitively), including: 1) sustaining the vowel 'a' for 10 seconds (VOWEL), 2) saying the days of the week in order, beginning at Monday (DAYS), 3) repeating the syllables pa-ta-ka as fast as possible for 10

seconds (diadochokinetic speech, i.e. fast alternation of movements, DDK), 4) reading a phonetically balanced paragraph (READ), and 5) telling a personal story from memory (one-minute unscripted monologue, FREE). Each task, except for the monologue, were elicited twice to reduce the effect of unfamiliarity[23]. The first was a practice trial while the second was used for analysis.

We used a Roland Quad-Capture recorder with an AKG C520 cardioid head-mounted condenser microphone (frequency range, 20-20KHz; sensitivity, -43 dB) positioned 8 cm from the mouth, at 45° angle laterally/inferiorly. Recordings were sampled at 44.1 KHz and quantized at 16 bits. Recordings for each participant (HC and pwMS) were completed in a single session, in a quiet room without acoustic isolation, reflecting real-world clinical practice. The door of the room was kept closed, and recordings stopped and re-initiated in case of extraneous noises (i.e. loudspeaker announcement, knocking on the door). Participants' positioning in the room was also standardized. Participants were recorded while sitting, facing the centre of the room and two meters from the wall in front of them. As relative noise can affect acoustic analysis, we determined the signal-to-noise ratio (SNR) in Praat[24] v6.0.28. SNR was calculated by subtracting the intensity in dB of the recorded background noise immediately before and after the sustained vowel from the mean intensity during the sustained vowel. SNR mean was 39dB (±6dB).

We used Audacity v2.1.2 (Free Software Foundation, Boston, USA) for auditory-graphic manual screening of all speech sample, to exclude procedural errors. Pre and post-task silences were deleted. Recordings from sustained vowels under 5 seconds were excluded (2%) but those with duration from 5 to 9 seconds (21%) were included in the subsequent analyses. Speech files were renumbered and randomized before perceptual analysis.

2.3.2. Perceptual analysis

One speech and language pathologist (FM) and one otolaryngologist (GN) rated all audio recordings blinded to presence of MS and disease severity. Twenty-six speech characteristics were analysed[25] (Table 1). Briefly, a normal or unremarkable speech feature was scored as zero while increasing deviation is scored using a four-point ordinal scale (e.g. speech naturalness, where 0=normal and 4=severely unnatural/bizarre). A single score was attributed to each perceptual feature, i.e. scores were not distributed across tasks. Each rater scored all samples from a single participant simultaneously and independently. Overall, 80% of initial perceptual scores showed perfect agreement between raters and an additional 18% diverged in only one point. Exponentially weighted kappa coefficient averaged at 0.514. Independent rating was

immediately followed by a discussion of scores and replay of recordings until consensus was reached. Only consensus scores were used in subsequent analysis.

2.3.3. Acoustic analysis

We selected acoustic characteristics that were previously shown to be altered in MS (e.g. fundamental frequency CoV[14], F2 slope[26]) or which had a corresponding perceptual characteristic previously shown to be altered in MS and that had a simple and straight forward acoustic calculation (e.g. prolonged pauses in perceptual analysis[3] with corresponding pause percentage in acoustic analysis).

The two seconds after the first train of pa-ta-ka were analysed. Typically in naturalistic speech, the first few syllables are produced with a much higher intensity (i.e. louder), followed by a short period with high intrasubject stability[27]. Similarly, the analysis of sustained vowels was confined to the middle 3 seconds for stability and consistency across participants[28].

Voicesauce[29] was used to compute cepstral peak prominence (CPP), harmonic-to-noise ratio (HNR, zero to 3,500Hz band), fundamental frequency and second formant; MSP© (Pentax)[30] to analyse DDK rate and variability; previously developed[25, 31] automated scripts on Praat for all other timing measures; and free-standing short scripts on Matlab© v2018b (The Mathworks Inc, US) for intensity-based first order calculations (means, standard deviations, perturbation, percentiles). For women and men, respectively, fundamental frequency boundaries were set to 100Hz and 300Hz, and 70 to 250Hz[32] while the analysis window length was set to 30ms and 42ms. Window shift was fixed at 10ms.

2.4. Self-assessed quality of life

MS participants completed the Multiple Sclerosis Impact Scale (MSIS-29)[33], a questionnaire validated to assess the impact of MS on quality of life.

2.5. MRI data acquisition and analysis

Structural imaging included: 1) high-resolution 3D T1-weighted MPRAGE scan with motion correction (TR=2530ms; TE=2.5ms; TI=1260ms; FOV=176x256mm; voxel size=1.0x1.0x1.0mm); and 2) high resolution 3D T2-weighted double inversion recovery (DIR) sequence (TR=7400ms; TE=324ms; TI=3000ms; flip angle=120°; ETL=625; FOV=144x220mm; voxel size=1.0x1.0x1.0mm). MRIs were administered within two weeks of behavioural testing. Data analyses were conducted using Freesurfer version 6.0 to determine

the whole brain, total grey matter and total white matter volumes from the MPRAGE images. Volumes were then standardized as a percentage of total intracranial volume. Lesion load was derived from the DIR images using the lesion prediction algorithm within the lesion segmentation tool toolbox (version 2.0.15) for statistical parametric mapping (SPM[c]) (The FIL Methods Group) as described in Boonstra et al[34].

2.6. Statistical Methods

Statistical analyses were performed using IBM SPSS[R] 25.0. To estimate the neurological disability level at which each speech measurement differs from controls, we divided data from the MS group into three subgroups according to EDSS scores: mild (EDSS<2.5), moderate (EDSS 3 to 5.5) and severe (EDSS>5.5). The criteria for EDSS-score boundaries were: 1) clinical intragroup identity (i.e. clinical similarities in their EDSS-scores definitions)[21]; and 2) groups that represented different "phases" of disease progression[35-38]. To match subgroups and HC for age and sex, we sequentially excluded participants from each subgroup starting with the most deviant participant until $p>0.1$ for age and sex.

We compared the general MS group with HC. We then compared each EDSS subgroup against HC and performed inter-subgroup comparisons of speech characteristics. We used independent samples t-test for parametric and Mann-Whitney for non-parametric pairs, for evaluating the null hypothesis (p values).

To account for multiple comparisons, we applied a maximum false discovery rate (q value) of 5%[39] and determined a reference p value (critical p) which is equivalent to a single comparison $p = 0.05$. In this method, the critical p value varies according to characteristics of the set of p values entered (e.g. the number of p values, their magnitude and dependency). As with conventional p values[40], the critical p should be used only as a reference value for interpreting results rather than a fixed threshold for significance. We reported the respective critical p value along with each set of results.

To estimate the strength of correlations, we calculated Spearman's coefficients (r) between speech measurements and the EDSS, imaging measurements and MSIS-29. To test if associations between speech metrics and EDSS were dependent on dysarthria, we calculated the same coefficients including only pwMS without dysarthria, as defined by naturalness = 0 in blinded ratings. Naturalness is a global measure of dysarthria and more sensitive than intelligibility.[41] We further calculated correlation coefficients for people without specific perceptual impairment for selected variables.

We entered all acoustic variables in a forward stepwise multiple linear regression to determine the best fit modelling for EDSS, for all pwMS regardless of the presence of dysarthria, with $p \leq 0.05$ to enter and >0.1 for variables to leave the model. We used the variables and parameters from the linear regression model as an 'acoustic composite score', i.e. the sum of all speech variables that formed the model multiplied by their respective beta values. Rather than predicting EDSS, the main purpose of the regression was to create an overall acoustic measure directed towards neurological dysfunction. In the perceptual analysis method, 'naturalness' is used by the trained rater in a similar way, considering only speech deviations relevant to the detection and grading of neurological impairment (dysarthria). The resulting acoustic composite score was then entered in the same group comparisons described before and tested for the correlations with imaging and quality of life. Further, 95% confidence intervals (CI) of the acoustic composite score and perceptual naturalness were used to gauge overall speech status on each EDSS subgroup.

2.7. Data availability statement

Data will not be made public for ethical safeguard of participants as it may be identifiable. Most of the remaining data in the current study is part of ongoing investigations by the researchers. The authors agree to share the study protocol and further details of statistical methods upon request.

3. RESULTS

Participants included 119 pwMS and 22 HC. Sixty-eight MS participants agreed to undergo the MRIs protocol. Demographics are shown in table 2.

[INSERT TABLE 2 ABOUT HERE]

3.1. Speech impairment in MS

None of the HC and 35% of pwMS presented with dysarthria as defined by naturalness ≥ 1 in blinded ratings (20% signs only, 8% mild, 6% moderate and 1% severe). Communalities between perceptual and acoustic analyses revealed slower speech rate ($p \leq 0.007$), increased variation in speech rate ($p \leq 0.004$), increase in pauses ($p \leq 0.002$), and smaller pitch variation (or monotonic speech, $p \leq 0.010$) during connected speech for the general MS cohort in comparison with HC (Supplementary Table A1 and Table A2).

3.2. Relationship between speech measures and EDSS

Dysarthria frequency and severity was associated with EDSS (r=0.41, p<0.001, figure 1) and affected the majority of pwMS with EDSS≥4. The three characteristics with the strongest associations with EDSS were the same for perceptual and acoustic analysis, and included DDK rate, speech rate and increase in pauses/intervals. DDK rate and speech rate were affected mainly in the severe group whereas the increase in pauses was observed for both the moderate and severe groups (Tables 3 and 4).

Speech exclusively from pwMS without global dysarthria (i.e. naturalness=0, n=77) correlated with EDSS for the individual acoustic metrics of pause percentage (r=0.36, p=0.002) and frequency instability (r=0.3, p = 0.01). Acoustic pause percentage correlated with EDSS (r=0.36, p=0.001) when considering only pwMS who scored zero for prolonged pauses in the blinded perceptual rating (subclinical increase in pauses). EDSS also correlated with frequency instability (r=0.35, p<0.001) for people without voice tremor (subclinical voice instability). Eighty-four percent of pwMS and without global dysarthria and 80% of HC presented minor deviations in perceptual scores but none correlated with EDSS.

[INSERT TABLE 3 AND 4 ABOUT HERE]

3.3. Relationship between speech and structural neuroimaging

MRI brain volumes correlated with perceptually assessed prosody rate and variable rate, and with acoustic measurements of energy variability and frequency instability. All correlations were weak to moderate ($r \approx 0.3$, $p \leq 0.008$, supplementary table A3) and no correlation was observed between speech measurements and brain volumes for people without dysarthria.

3.4. Relationship between speech and quality of life

MSIS-29 scores correlated moderately with perceptual prolonged intervals (Table 3), and with acoustic intensity decay and frequency instability (Table 4). For people without dysarthria, intensity decay (r=0.40, p=0.006) and decreased frequency variability (r=0.40, p=0.007) were the only variables that correlated with MSIS-29.

3.5. Acoustic composite score

Multiple linear regression resulted in a model that included pause percentage FREE, speech rate READ and frequency instability VOWEL (Table 5), termed ‘acoustic composite score’, which accounted for 35% of EDSS variation (adjusted r^2=0.347, standard error of the estimate=1.8).

[INSERT TABLE 5 ABOUT HERE]

The acoustic composite score correlated with MSIS-29 ($r = 0.5$, $p<0.001$), white matter volume ($r=0.32$, $p=0.007$) and total lesion load ($r=0.34$, $p=0.008$). Considering only people without dysarthria, the acoustic composite score correlated with EDSS ($r=0.45$, $p<0.001$) and with MSIS-29 ($r=0.4$, $p=0.01$). The acoustic composite score differentiated between the mild and moderate MS groups ($p<0.001$) and between moderate and severe MS groups ($p=0.003$, supplementary table A1, figure 1).

[INSERT FIGURE 1 ABOUT HERE]

4. DISCUSSION

This cross-sectional study confirmed that speech impairment was present in all disability stages but clinically significant only in groups with moderate or severe neurological impairment (EDSS≥3). An acoustic composite score included pause percentage, frequency instability and DDK speech rate that explained 35% of the variance in EDSS. Both the acoustic composite score and perceptual naturalness, global measurements of speech function, were associated with EDSS-defined disease severity. In addition, only the acoustic composite score differentiated between the three EDSS subgroups and correlated with quality of life and EDSS in people without dysarthria. Results suggest that an objective speech biomarker, incorporating multiple speech features, is likely to better contribute to the representation of overall neurological impairment in pwMS than pure perceptual or isolated speech tests.

The proportion of pwMS with overt dysarthria was smaller in comparison to previous reports[1, 2] (35% vs 50%), which is likely due to differences in disability level of included participants (median EDSS of 2.5 in our study vs 6 in Hartelius et al[1]). In agreement with previous studies[3, 15] pwMS presented primarily with slower speech rate, prolonged pauses and, to a minor degree, impairment of articulatory accuracy (consonants and vowels) and loss of speech intensity (monoloudness and loudness decay). Such impairments were mostly evident in the EDSS≥6 subgroup. In the EDSS≤2.5 subgroup, we observed speech very similar to that of HC, regardless of the method of analysis.

Larger statistical differences were observed with acoustic analysis in relation to perceptual for group comparisons. A proportion of that difference in favour of acoustic metrics is likely attributable to the nature of scores, ordinal and course for perceptual and continuous for acoustic. Continuous measurements also permit the use of more powerful and numerous statistical procedures, such as conventional regression and principal component analysis, in addition to being objective and automatable.

Acoustic variability of frequency was unrelated to disease severity, in agreement with Rusz et al[17], or to brain volumetrics. It was, however, the only individual speech feature that differentiated each of the MS subgroups from HC, complementing findings between HC and groups of pwMS with a broad spectrum of disability status[17, 42, 43]. Observing both results in a single cohort suggests that decreased variability of frequency (monopitch) might be more relevant to disease detection than for tracking disease progression in MS. Conversely, the positive correlation between decreased variability of frequency and quality of life might reflect depressive states[44] but requires dedicated validation.

Acoustic metrics of frequency instability and energy variability were associated with MRI whole brain volume in agreement with Rusz et al. (termed frequency variability and excessive loudness variation in that study)[15]. Although we also found age, disability status, MRI acquisition and MRI feature extraction very similar between studies, we did not find an association between whole brain volumes and acoustic speech rate. Methodologically, Rusz et al. used the average of two trials for acoustic analysis, whereas we used the second of two trials. Nonetheless, it is unlikely that this difference alone could explain results' differences between studies given the short-term reliability of these measurements[15, 31] and further investigation is needed.

The association between quality of life and speech should not be interpreted as evidence for causation nor that individual speech metrics could predict the impact of MS in a person's well-being. Interestingly, speech features that correlated with quality of life were mostly those observed in the EDSS severe subgroup. This association possibly reflects the burden of disease-related disability rather than the impact of speech impairment on quality of life. Nonetheless, observing correlations of individual speech metrics may assist in selection of speech variables for future work related to quality of life.

Limitations in this study included restriction in the number of variables included in analyses, both in speech measurements and MS-related neurological dysfunction. However, we used a well-studied and consolidated set of perceptual measurements. For the acoustic analysis, we chose to observe tasks separately, which may have increased the number of redundant measures such as pause percentage during free speech and speech rate during reading. To limit the number of statistical tests, acoustic analysis of characteristics, which are (arguably) not well established in the literature were not included here. However, the broader representativeness of the perceptual set showed that imprecision of consonants and hypernasality were present in the severe subgroup, possibly representing advanced pyramidal, cerebellar and brainstem dysfunctions. Other domains were not assessed through either method, such as maximum

phonatory time. Thus, it is possible that acoustic representation of more speech characteristics would increase the sensitivity of the composite score for neurological impairment. Secondly, we presented only whole brain volumetric and lesion measures for the purpose of complementing the representation of overall disease severity measured through EDSS and MSIS-29. Studies of specific brain regions, neural tracts and functional imaging may show stronger correlations with speech features and add feature-location specificity.

Limitations of the acoustic analysis procedure must also be considered. While we screened and corrected for errors, some mismatch between obtained and real values is likely to occur with any automated analysis.

It is also important to note that, despite differences between groups being much larger than within groups, speech characteristics varied considerably between individuals with the same level of disability or between healthy controls. Differences in speech metrics between persons is intuitive and unavoidable. In that sense, it seems likely that the next phase of longitudinal measurements of speech will produce possibly stronger correlations with disability scores and other clinical endpoints.

5. CONCLUSION

The degree of speech impairment moderately parallels that of overall MS-related neurological impairment and was not restricted to advanced stages of MS. An experimental acoustic composite score was sensitive to change in clinical scores, MRI measurements and quality of life. Future studies should test the accuracy of acoustic scores to longitudinally track speech changes.

6. ACKNOWLEDGEMENTS

The authors acknowledge the immeasurable contribution of participants who voluntarily donated their time and patience for science.

7. FUNDING

This work was supported by the National Health and Medical Research Council (grant numbers 1085461 and 1082910); and by the University of Melbourne (graduate research scholarship).

7. AUTHORS CONTRIBUTIONS

Adam Vogel was responsible for conception, organization and execution of the research project; design of the statistical analysis; review, critique and writing of the manuscript.

Andrew Evans was responsible for conception of the research project; review and critique of the manuscript.

Anneke van der Walt was responsible for conception, organization and execution of the research project; review and critique of the statistical analysis; securing funding; review, critique and writing of manuscript.

L. Eduardo Cofré Lizama participated of the organization of the research project, data collection and review of the manuscript.

Fernanda Maldonado was responsible for data analysis; review and critique of the manuscript.

Frederique Boonstra was responsible for conception, organization and execution of the research project; review, critique and writing of the manuscript.

Gustavo Noffs was responsible for conception, organization and execution of the research project; design and execution of the statistical analysis; writing of the first and subsequent drafts.

Helmut Butzkueven was responsible for conception of the research project; review and critique of the statistical analysis; review, critique and writing of the manuscript.

Jim Stankovich was responsible for design, critique and review of the statistical analysis;

Mary Galea was responsible for organization of the research project; review and critique of the manuscript.

Scott Kolbe was responsible for conception and organization of the research project; review and critique of the statistical analysis; review, critique and writing of the manuscript.

Thushara Perera was responsible for conception of the research project; design, review and critique of the statistical analysis; review, critique and writing of the manuscript.

8. POTENTIAL CONFLICT OF INTEREST

Adam Vogel is Chief Science Officer of Redenlab. Adam receives grant and fellowship funding from the National Health and Medical Research Council of Australia.

Andrew Evans received honoraria from Novartis for giving presentations and providing consultancy services. He has participated in scientific advisory board meetings for Novartis, UCB Pharma, Allergan, and Boehringer Ingelheim. He has received conference travel support from Boehringer Ingelheim.

Anneke van der Walt has received travel support and served on advisory boards for Novartis, Biogen, Merck Serono, Roche and Teva. She receives grant support from the National Health and Medical Research Council of Australia.

L. Eduardo Cofré Lizama has nothing to disclose.

Frederique M.C. Boonstra has nothing to disclose.

Gustavo Noffs has nothing to disclose.

Helmut Butzkueven served on scientific advisory boards for Biogen, Novartis and Sanofi-Aventis and received conference travel support from Novartis, Biogen and Sanofi Aventis. He serves on steering committees for trials conducted by Biogen and Novartis received research support from Merck, Novartis and Biogen.

Mary Galea has nothing to disclose.

Scott Kolbe receives grant income from the National Health and Medical Research Council of Australia and has received honoraria from Novartis, Biogen and Merck.

Thushara Perera has nothing to disclose.

10. REFERENCES

1. Hartelius L, Runmarker B, rn o, Andersen O. Prevalence and characteristics of dysarthria in a multiple-sclerosis incidence cohort: relation to neurological data. Folia phoniatrica et logopaedica. 2000;52(4):160-77.
2. O'Connor P. Key issues in the diagnosis and treatment of multiple sclerosis. An overview. Neurology. 2002 Sep 24;59(6 Suppl 3):S1-33.
3. Noffs G, Perera T, Kolbe SC, et al. What speech can tell us: A systematic review of dysarthria characteristics in Multiple Sclerosis. Autoimmun Rev. 2018 Dec;17(12):1202-9.
4. Mundt JC, Vogel AP, Feltner DE, Lenderking WR. Vocal acoustic biomarkers of depression severity and treatment response. Biological psychiatry. 2012 Oct 1;72(7):580-7.
5. Flint AJ, Black SE, Campbell-Taylor I, Gailey GF, Levinton C. Acoustic analysis in the differentiation of Parkinson's disease and major depression. Journal of Psycholinguistic Research. 1992;21(5):383-99.
6. Cummins N, Sethu V, Epps J, Schnieder S, Krajewski J. Analysis of acoustic space variability in speech affected by depression. Speech Communication. 2015;75:27-49.
7. Arnett PA, Smith MM, Barwick FH, Benedict RH, Ahlstrom BP. Oralmotor slowing in multiple sclerosis: Relationship to neuropsychological tasks requiring an oral response. Journal of the International Neuropsychological Society. 2008;14(03):454-62.
8. Rodgers JD, Tjaden K, Feenaughty L, Weinstock-Guttman B, Benedict RH. Influence of cognitive function on speech and articulation rate in multiple sclerosis. Journal of the International Neuropsychological Society. 2013;19(02):173-80.
9. Kister I, Bacon TE, Chamot E, et al. Natural history of multiple sclerosis symptoms. Int J MS Care. 2013 Fall;15(3):146-58.

10. Panou T, Mastorodemos V, Papadaki E, Simos PG, Plaitakis A. Early signs of memory impairment among multiple sclerosis patients with clinically isolated syndrome. Behavioural neurology. 2012;25(4):311-26.
11. Martin CL, Phillips BA, Kilpatrick TJ, et al. Gait and balance impairment in early multiple sclerosis in the absence of clinical disability. Multiple sclerosis (Houndmills, Basingstoke, England). 2006 Oct;12(5):620-8.
12. Fletcher H, Munson WA. Loudness, its definition, measurement and calculation. Bell System Technical Journal. 1933;12(4):377-430.
13. Maryn Y, Leblans M, Zarowski A, Barkmeier-Kraemer J. Objective Acoustic Quantification of Perceived Voice Tremor Severity. Journal of speech, language, and hearing research : JSLHR. 2019 Oct 25;62(10):3689-705.
14. Hartelius L, Buder EH, Strand EA. Long-term phonatory instability in individuals with multiple sclerosis. Journal of Speech, Language, and Hearing Research. 1997;40(5):1056-72.
15. Rusz J, Vaneckova M, Benova B, et al. Brain volumetric correlates of dysarthria in multiple sclerosis. Brain Lang. 2019 Jul;194:58-64.
16. Chan JCS, Stout JC, Vogel AP. Speech in prodromal and symptomatic Huntington's disease as a model of measuring onset and progression in dominantly inherited neurodegenerative diseases. Neuroscience and biobehavioral reviews. 2019 Dec;107:450-60.
17. Rusz J, Benova B, Ruzickova H, et al. Characteristics of motor speech phenotypes in multiple sclerosis. Multiple sclerosis and related disorders. 2018;19:62-9.
18. Fazeli M, Moradi N, Soltani M, et al. Dysphonia Characteristics and Vowel Impairment in Relation to Neurological Status in Patients with Multiple Sclerosis. Journal of voice : official journal of the Voice Foundation. 2020 May;34(3):364-70.
19. Noffs G, Boonstra FMC, Perera T, et al. Acoustic Speech Analytics Are Predictive of Cerebellar Dysfunction in Multiple Sclerosis. Cerebellum (London, England). 2020 Jun 18.
20. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. The Lancet Neurology. 2018;17(2):162-73.
21. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology. 1983 Nov;33(11):1444-52.
22. Vogel AP, Fletcher J, Maruff P. The impact of task automaticity on speech in noise. Speech Communication. 2014;65:1-8.
23. Vogel AP, Maruff P. Monitoring change requires a rethink of assessment practices in voice and speech. Logopedics, phoniatrics, vocology. 2014 Jul;39(2):56-61.
24. Boersma P. Praat, a system for doing phonetics by computer. Glot international. 2002;5.

25. Vogel AP, Poole ML, Pemberton H, et al. Motor speech signature of behavioral variant frontotemporal dementia: Refining the phenotype. Neurology. 2017 Aug 22;89(8):837-44.
26. Rosen KM, Goozée JV, Murdoch BE. Examining the effects of multiple sclerosis on speech production: Does phonetic structure matter? Journal of Communication Disorders. 2008;41(1):49-69.
27. Tjaden K, Watling E. Characteristics of diadochokinesis in multiple sclerosis and Parkinson's disease. Folia Phoniatrica et Logopaedica. 2003;55(5):241-59.
28. Kent RD, Vorperian HK, Duffy JR. Reliability of the Multi-Dimensional Voice Program for the analysis of voice samples of subjects with dysarthria. American journal of speech-language pathology. 1999;8(2):129-36.
29. Shue YL, Keating P, Vicenik C. VOICESAUCE: A program for voice analysis. The Journal of the Acoustical Society of America. 2009;126(4):2221-.
30. Wang YT, Kent RD, Duffy JR, Thomas JE. Analysis of diadochokinesis in ataxic dysarthria using the motor speech profile program. Folia phoniatrica et logopaedica : official organ of the International Association of Logopedics and Phoniatrics (IALP). 2009;61(1):1-11.
31. Vogel AP, Fletcher J, Snyder PJ, Fredrickson A, Maruff P. Reliability, stability, and sensitivity to change and impairment in acoustic measures of timing and frequency. Journal of Voice. 2011;25(2):137-49.
32. Vogel AP, Maruff P, Snyder PJ, Mundt JC. Standardization of pitch-range settings in voice acoustic analysis. Behavior research methods. 2009;41(2):318-24.
33. Hobart J, Lamping D, Fitzpatrick R, Riazi A, Thompson A. The Multiple Sclerosis Impact Scale (MSIS-29): a new patient-based outcome measure. Brain. 2001 May;124(Pt 5):962-73.
34. Boonstra F, Perera T, Noffs G, et al. Novel functional MRI task for studying the neural correlates of upper limb tremor. Frontiers in neurology. 2018;9:513.
35. Amato MP, Ponziani G. Quantification of impairment in MS: discussion of the scales in use. Multiple Sclerosis Journal. 1999;5(4):216-9.
36. Hohol M, Orav E, Weiner H. Disease steps in multiple sclerosis: a longitudinal study comparing disease steps and EDSS to evaluate disease progression. Multiple Sclerosis Journal. 1999;5(5):349-54.
37. Ravnborg M, Blinkenberg M, Sellebjerg F, Ballegaard M, Larsen SH, Sørensen PS. Responsiveness of the multiple sclerosis impairment scale in comparison with the expanded disability status scale. Multiple Sclerosis Journal. 2005;11(1):81-4.

38. Meyer-Moock S, Feng Y-S, Maeurer M, Dippel F-W, Kohlmann T. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. BMC neurology. 2014;14(1):58.
39. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. The annals of statistics. 2001;29(4):1165-88.
40. Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. Taylor & Francis; 2016.
41. Sussman JE, Tjaden K. Perceptual measures of speech from individuals with Parkinson's disease and multiple sclerosis: Intelligibility and beyond. Journal of Speech, Language, and Hearing Research. 2012;55(4):1208-19.
42. Vavougios GD, Doskas T, Konstantopoulos K. An electroglottographical an alysis-based discriminant function model differentiating multiple sclerosis patients from healthy controls. Neurological Sciences. 2018;39(5):847-50.
43. Konstantopoulos K, Vikelis M, Seikel JA, Mitsikostas D-D. The existence of phonatory instability in multiple sclerosis: an acoustic and electroglottographic study. Neurological Sciences. 2010;31(3):259-68.
44. Cannizzaro M, Harel B, Reilly N, Chappell P, Snyder PJ. Voice acoustical measurement of the severity of major depression. Brain and cognition. 2004 Oct;56(1):30-5.

Table 1. Speech items included in the analyses. Mathematical definitions of non-standard acoustic measures are shown between brackets.

Speech domain	**Perceptual items**	**Acoustic items and tasks**	
Timing	Prolonged intervals Prolonged phonemes Prosody rate Variable rate DDK speed DDK irregularity False starts	Pause percentage (total pause time/task time)	READ, FREE
		Articulatory time (task time – total pause time)	DAYS, READ
		Speech rate (syllables or words/second)	DDK, DAYS, READ
		Pause variability (silences length SD)	READ, FREE
		Articulatory variability (syllables length)	DDK
Control	Voice tremor Loudness Decay Monoloudness Monopitch Audible inspiration Distorted Vowels Imprecise consonants Repeated phonemes Irregular articulatory breakdowns Groping Pitch breaks Reduced stress Nasality Phonemic errors	Frequency instability (f0 CoV)	VOWEL
		Loudness instability (energy CoV)	
		F2 articulation speed (f2 CoV/articulatory time)	READ
		Energy decay (mean of differences in energy between starting and finishing quarters of time of each utterance)	FREE
		Loudness variability (energy CoV)	FREE
		Frequency variability (f0 CoV)	READ, FREE

Voice quality	Roughness Breathiness Strain	Harmonics-to-noise ratio Cepstral Peak Prominence	VOWEL
Multi-domain	Naturalness Intelligibility	Composite score	All tasks

CoV = coefficient of variance (i.e. standard deviation divided by the mean); F0 = fundamental frequency; F2 = second formant; SD = standard deviation.

Table 2. Descriptive statistics for groups and subgroups. Mean±standard deviation for age, disease duration and MRI measurements, and median±interquartile range for EDSS and MSIS-29.

	HC	MS	MS subgroups			
			MRI	MILD (EDSS≤2.5)	MODERATE (3≤EDSS≤5.5)	SEVERE (EDSS≥6)
N	22	118	68	33	30	20
Female	73%	75%	74%	82%	67%	80%
Age, years	45.2±14.5	45.5±11.6	46.7±11.6	47.1±11.6	49.2±9.8	50.8±6.7
Disease Duration		11.9±8.3	13±7.7	8.5±5.4	14.7±7.7	18.1±7.2
EDSS		2.5±4	3.4±3	2±1	4±2	6.5±0.5
MSIS-29		73.5±44.25		52±35.5	73.5±43.25	93±44.25
Brain volume (%)	73.6±5*		70.8±7	73.4±5.8	69.9±7.9	67.4±6.8
Lesion load (voxels)			9320±9747	6461±6797	10713±9462	15521±14424

EDSS = Expanded Disability Status Scale; HC = healthy controls; MRI = magnetic resonance imaging participants; MS = Multiple Sclerosis; *from n=14.

Table 3. Correlation coefficients between perceptual speech measures, EDSS and MSIS-29 (quality of life). FDR "adjusted" critical p value = 0.008.

Speech domain	Speech item and method	EDSS		MSIS	
		r	p	r	p
Multi-domain	Naturalness	0.41	<0.001	0.28	0.011
	Intelligibility	0.36	<0.001	0.19	0.083
Timing	Prosody rate	-0.41	<0.001	-0.39	<0.001
	DDK speed	-0.42	<0.001	-0.37	0.001
	Prolonged intervals	0.37	<0.001	0.43	<0.001
	Variable rate	0.28	0.002	0.29	0.009
	DDK irregularity	0.25	0.006	0.12	0.303
	Prolonged phonemes	0.18	0.056	0.26	0.021
	False starts	0.14	0.141	0.16	0.145
Control	Imprecise consonants	0.39	<0.001	0.17	0.124
	Distorted vowels	0.37	<0.001	0.36	0.001
	Voice tremor	0.37	<0.001	0.19	0.093
	Loudness decay	0.33	<0.001	0.26	0.018
	Monopitch	0.31	0.001	0.04	0.759
	Monoloudness	0.31	0.001	0.20	0.068
	Pitch breaks	0.26	0.005	0.23	0.040
	Audible inspiration	0.23	0.010	0.19	0.083
	Reduced stress	0.23	0.012	0.15	0.170
	Irregular art. breakdown	0.19	0.041	0.14	0.200
	Phonemic errors	0.09	0.359	0.06	0.581
	Nasality	0.06	0.503	0.07	0.519
Voice quality	Roughness	0.32	<0.001	0.24	0.033
	Strain	0.29	0.002	0.24	0.030
	Breathiness	0.13	0.174	0.21	0.066

p = p value for the given coefficient; r =correlation coefficient.

Table 4. Correlation coefficients between acoustic speech measures, EDSS and MSIS-29 (quality of life). FDR "adjusted" critical p value = 0.008.

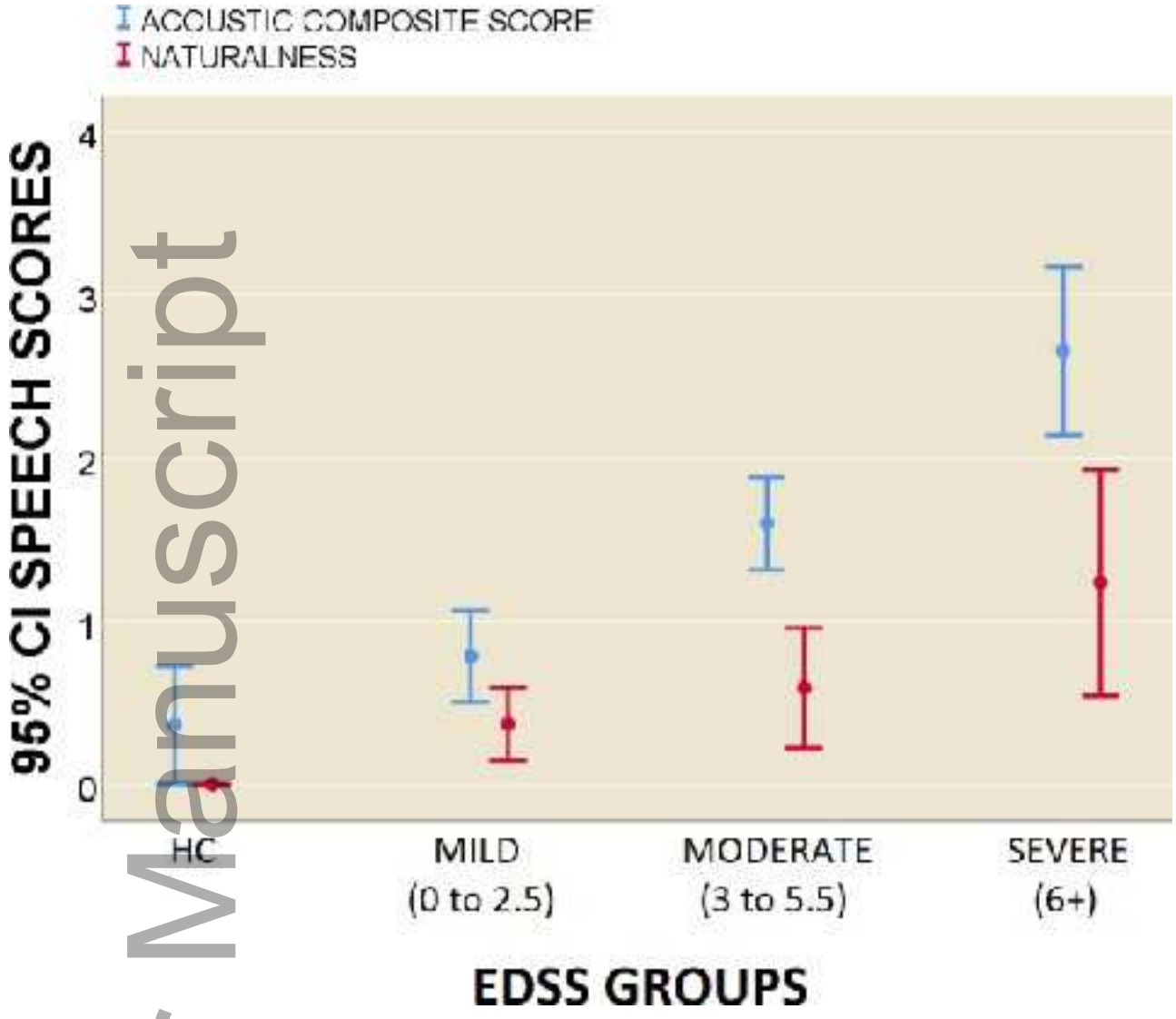
Speech domain	Speech item and method	EDSS		MSIS	
		r	p	r	p
Multi-domain	Acoustic composite score	0.60	<0.001	0.50	<0.001
Timing	Pause percentage, FREE	0.46	<0.001	0.27	0.015
	Speech rate, READ	-0.40	<0.001	-0.28	0.014
	Speech rate, DDK	-0.37	<0.001	-0.35	0.002
	Pause variability, FREE	0.35	<0.001	0.36	0.001
	Pause variability, READ	0.30	0.001	0.17	0.139
	Speech rate, DAYS	-0.24	0.008	-0.14	0.233
	Articulatory time, DAYS	0.23	0.012	0.10	0.368
	Articulatory time, READ	0.22	0.019	0.26	0.021
	Pause variability, READ	0.16	0.080	0.27	0.019
	Syllable variability, DDK	0.03	0.762	0.14	0.233
Control	Frequency instability	0.40	<0.001	0.40	<0.001
	F2 articulation speed, READ	-0.26	0.005	-0.24	0.034
	Energy decay	0.22	0.014	0.44	<0.001
	Energy variability	0.12	0.202	-0.02	0.864
	Energy instability	0.11	0.258	0.15	0.180
	Frequency variability, READ	-0.06	0.525	-0.22	0.055
	Frequency variability, FREE	0.05	0.623	-0.07	0.565
Voice quality	Cepstral Peak Prominence	0.07	0.466	-0.11	0.345
	Harmonics-to-noise ratio	0.00	0.992	-0.05	0.662

p = p value for the given coefficient; r =correlation coefficient.

Table 5. Variables in the regression model for EDSS. Beta coefficients were used to calculate the composite score whereas standardized beta can be used to compare the 'weight' of each variable for the model.

Variables	Beta	Standard error	Standardized beta	t	p
Constant	3.488	1.297	NA	2.690	0.008
Pause percentage, FREE	0.083	0.020	**0.340**	4.105	<0.001
Frequency instability, VOWEL	0.711	0.224	**0.257**	3.170	0.002
Speech rate, DDK	-0.513	0.178	**-0.241**	-2.879	0.005

DDK = speech diadochokinesis task, FREE = unscripted monologue speech task, VOWEL = sustained utterance of the vowel /a/ task.



ene_14523_f1.png

Author/s:
Noffs, G; Boonstra, FMC; Perera, T; Butzkueven, H; Kolbe, SC; Maldonado, F; Cofre Lizama, LE; Galea, MP; Stankovich, J; Evans, A; van Der Walt, A; Vogel, AP

Title:
Speech metrics, general disability, brain imaging and quality of life in multiple sclerosis

Date:
2020-10-22

Citation:
Noffs, G., Boonstra, F. M. C., Perera, T., Butzkueven, H., Kolbe, S. C., Maldonado, F., Cofre Lizama, L. E., Galea, M. P., Stankovich, J., Evans, A., van Der Walt, A. & Vogel, A. P. (2020). Speech metrics, general disability, brain imaging and quality of life in multiple sclerosis. EUROPEAN JOURNAL OF NEUROLOGY, 28 (1), pp.259-268. https://doi.org/10.1111/ene.14523.