University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Noffs, Gustavo

Title:
The relationship between disease status and acoustic analysis of speech in Multiple Sclerosis

Date:
2021

Persistent Link:
https://hdl.handle.net/11343/274367

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.

The relationship between disease status and acoustic analysis of speech in Multiple Sclerosis

Gustavo Noffs

ORCID identifier 0000-0001-8905-7879

Submitted in February 2021

Department of Audiology and Speech Pathology

Faculty of Medicine, Dentistry and Health Sciences

University of Melbourne

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy

in Medicine, Dentistry and Health Sciences

Abstract

Multiple Sclerosis (MS) is a common cause of disability in young adults. Modern MS treatment has markedly improved long-term outcomes for the most common type of MS, i.e., relapsing-remitting MS. However, optimal pharmacological treatment decisions depend on early detection of disease progression. Validated assessment of progression includes disability scales and standard structural magnetic resonance imaging (MRI) of the brain (brain volume changes), but these are not sensitive to the early disability worsening. More dynamic tools are required to increase the responsiveness of MS monitoring. Objective speech analysis offers several potential advantages for disease monitoring, including short assessment time, low patient effort, automation, remote assessment (via smartphone/mobile device), improved reliability via elimination of inter and intra-rater variability, and wide disability-wise suitability as speech continues to be produced even by severely disabled patients.

The general aims of this thesis were to (1) review and identify gaps in the literature regarding the validity of speech to monitor MS progression, (2) examine the relationship between acoustic metrics and general neurological status or cerebellar impairment in people with MS (pwMS), and (3) describe differences in health-related acoustic metrics between mobile devices' recordings and professional-grade microphones, both of which are commonly used in research settings.

Most previous studies described only differences between pwMS and healthy controls for a narrow set of acoustic speech features, with almost inexistent information on disability or neuroimaging correlates of speech impairment in MS. In the current study, we observed a moderate strength relationship between speech acoustic features and general disability, cerebellar ataxia, whole brain and cerebellar MRI-measured white matter volume and lesion load. Single-feature metrics independently contributed to an acoustic composite score that

differentiated between mild, moderate, and severe levels of disability and correlated with EDSS scores in people with or without overt dysarthria (subclinical speech impairment. A similar ataxia acoustic composite score predicted an abnormal time to complete the 9-hole peg test with 85% accuracy, including in a naïve dataset. Acoustic models were non-specific, with the same speech features compounding both the general and cerebellar models. Lastly, we observed wide differences in acoustic values for most speech features of interest between mobile devices and a professional recording set. Differences did not affect the relationship between acoustic scores and EDSS, although tested only in a small cohort.

The current studies describe various speech acoustic features in their relationship with clinical disability and neuroimaging metrics in the same participants and further explores subpopulations of interest according to disability level and presence of dysarthria. Present results support the validity of speech models to, at least in part, reflect neurological status in MS. Current and previous findings strongly suggest against the interchangeable use of recording methods, particularly between different microphones and different distances between microphone and sound source, but do not preclude the use of mobile devices for health-related speech data acquisition.

Declarations

This is to certify that:

1. The thesis comprises only my original work towards the PhD except where indicated in the Preface;

2. Due acknowledgement has been made in the text to all other material used;

and

3. The thesis is fewer than 100,000 words in length exclusive of tables, bibliographies and appendices.

Gustavo Noffs

Preface

This is a thesis with publication and was supervised by Prof Adam Vogel, A/Prof Anneke Van der Walt, Dr Thushara Perera, Dr Scott Kolbe and Prof Helmut Butzkueven. No work towards the thesis has been submitted for other qualifications or carried out prior to enrolment in the degree. No third party assisted in the editorial preparation of the thesis.

The work towards the thesis was financially supported by the Melbourne Research Scholarship and Fee Remission, and NHMRC grant number 1085461.

I acknowledge the contribution of co-authors to the published and unpublished articles included in Chapters 2-5, as outlined below.

Chapter 2

Article title: What speech can tell us: a systematic review of dysarthria characteristics in Multiple Sclerosis.

Authors: Gustavo Noffs, Thushara Perera, Scott C. Kolbe, Camille J. Shanahan, Frederique M.C. Boonstra, Andrew Evans, Anneke Van der Walt and Adam P Vogel.

Publication status: Published on Autoimmunity Reviews (journal), volume 17, issue 12, pages 1202-1209, in December 2018.

Author contributions: conception and design: AV, AVW, GN; database search and extraction: GN; critical interpretation: AV, GN; first manuscript draft: GN; manuscript review and further writing contribution: AE, AV, AVW, CS, FB, SK, TP; response to reviewers and manuscript revisions: GN.

GN was responsible for approximately 55% of the work for this chapter.

Chapter 3

Article title: Speech metrics, general disability, brain imaging and quality of life in MS**.**

Authors: Gustavo Noffs, Frederique M.C. Boonstra, Thushara Perera, Helmut Butzkueven, Scott C. Kolbe, Fernanda Maldonado, L. Eduardo Cofré Lizama, Mary P. Galea, Jim Stankovich, Andrew Evans, Anneke Van der Walt and Adam P Vogel.

Publication status: Published in the European Journal of Neurology, volume 28, issue, pages 259-268, in January 2021.

Author contributions: conception and design: AV, AVW, GN, HB, SK, TP; recruitment: AVW, FB, GN, HB, LL, MG; data collection: AVW, FB, GN, LL; data analysis: AV, FB, FM, GN, JS; first manuscript draft: GN; manuscript review and further writing contribution: AE, AV, AVW, FB, FM, HB, JS, LL, MG, SC, TP; response to reviewers and manuscript revisions: AV, AVW, FB, GN, HB, JS, SK, TP.

GN was responsible for approximately 55% of the work for this chapter.

Chapter 4

Article title: Acoustic speech analytics are predictive of cerebellar dysfunction in Multiple Sclerosis**.**

Authors: Gustavo Noffs, Frederique M.C. Boonstra, Thushara Perera, Scott C. Kolbe, Jim Stankovich, Helmut Butzkueven, Andrew Evans, Adam P Vogel and Anneke Van der Walt.

Publication status: Published in The Cerebellum (journal), volume 19, issue 5, pages 691-700, in October 2020.

Author contributions: conception and design: AV, AVW, GN, HB, SK, TP; recruitment: AVW, FB, GN, HB; data collection: AVW, FB, GN; data analysis: AV, FB, GN, JS, TP; first manuscript draft: GN; manuscript review and further writing contribution: AE, AV, AVW, FB, FM, HB, JS, SC, TP; response to reviewers and manuscript revisions: AV, AVW, FB, GN, JS, SK, TP.

GN was responsible for approximately 55% of the work for this chapter.

Chapter 5

Article title: The effect of mobile recording devices on speech acoustic metrics.

Authors: Gustavo Noffs, Matthew Cobler-Lichter, Scott C. Kolbe, Helmut Butzkueven, Frederique M.C. Boonstra, Andrew Evans, Anneke Van der Walt, Adam P Vogel and Thushara Perera.

Publication status: Submitted for publication to the Journal of Speech, Language and Hearing Research on the 11th of February 2021.

Author contributions: conception and design: AV, AVW, GN; recruitment: AVW, FB, GN, HB, MCL; data collection: AVW, GN, MCL; data analysis: AV, AVW, FB, GN, MCL, TP; first manuscript draft: GN; manuscript review and further writing contribution: AV, AVW, FB, HB, MCL, SC, TP. GN was responsible for approximately 60% of the work for this chapter.

Publications, presentations and awards

Peer-reviewed publications

- Iva P, Fielding J, Clough M, White O, Noffs G, Godic B, Martin R, van der Walt A, Rajan R. Speech discrimination impairments as a marker of disease severity in multiple sclerosis. Mult Scler Relat Disord. 2021 Jan;47:102608. DOI: 10.1016/j.msard.2020.102608. Epub 2020 Nov 1. PMID: 33189
- 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. Speech metrics, general disability, brain imaging and quality of life in multiple sclerosis. Eur J Neurol. 2021 Jan;28(1):259-268. DOI: 10.1111/ene.14523. Epub 2020 Oct 22. PMID: 32916031
- Magee M, Lewis C, Noffs G, Reece H, Chan JCS, Zaga CJ, Paynter C, Birchall O, Rojas Azocar S, Ediriweera A, Kenyon K, Caverlé MW, Schultz BG, Vogel AP. Effects of face masks on acoustic analysis and speech perception: Implications for peri-pandemic protocols. J Acoust Soc Am. 2020 Dec;148(6):3562. DOI: 10.1121/10.0002873. PMID: 33379897; PMCID: PMC7857500
- Noffs G, Boonstra FMC, Perera T, Kolbe SC, Stankovich J, Butzkueven H, Evans A, Vogel AP, van der Walt A. Acoustic Speech Analytics Are Predictive of Cerebellar Dysfunction in Multiple Sclerosis. Cerebellum. 2020 Oct;19(5):691-700. DOI: 10.1007/s12311-020-01151-5. PMID: 32556973
- Tamplin J, Morris ME, Marigliani C, Baker FA, Noffs G, Vogel AP. ParkinSong: Outcomes of a 12-Month Controlled Trial of Therapeutic Singing Groups in Parkinson's Disease. J Parkinsons Dis. 2020;10(3):1217-1230. DOI: 10.3233/JPD-191838. PMID: 32538865
- Boonstra FMC, Evans A, Noffs G, et al. OnabotulinumtoxinA treatment for MS-tremor modifies fMRI tremor response in central sensory-motor integration areas. *Mult Scler Relat Disord*. 2020;40:101984. DOI:10.1016/j.msard.2020.101984.

- Boonstra FM, Noffs G, Perera T, et al. Functional neuroplasticity in response to cerebello-thalamic injury underpins the clinical presentation of tremor in multiple sclerosis. *Mult Scler*. 2020;26(6):696-705. DOI:10.1177/1352458519837706
- Boonstra FMC, Perera T, Noffs G, et al. Novel Functional MRI Task for Studying the Neural Correlates of Upper Limb Tremor. *Front Neurol*. 2018;9:513. Published 2018 Jul 2. DOI:10.3389/fneur.2018.00513
- Perera T, Lee WL, Yohanandan SAC, et al. Validation of a precision tremor measurement system for multiple sclerosis. *Journal of Neuroscience Methods* 2018. DOI:10.1016/j.neumeth.2018.09.022
- 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;17:1202-1209. DOI:10.1016/j.autrev.2018.06.010

Podium presentations

Oral presentations

- Sep/2019: "Acoustic analysis of speech and Multiple Sclerosis disease status: towards tracking neurological progression", Karolinska Institutet, Stockholm, Sweden.
- Sep/2019: "Acoustic analysis of speech and Multiple Sclerosis disease status: towards tracking neurological progression", University of Gothenburg, Sweden.
- Nov/2018: "Objective analysis of speech in Multiple Sclerosis", Melbourne School of Health Sciences Graduate Research Colloquium, Melbourne, Australia.
- Oct/2018: "A comparison of human and digital speech analysis to describe and monitor cerebellar dysfunction and disease severity in multiple sclerosis". European Congress for the Treatment and Research of Multiple Sclerosis, Berlin, Germany.
- Oct/2017: "MS talks", Melbourne Brain Symposium, Melbourne, Australia.

- Oct/2017: "Objective analysis of speech correlates with disease severity in multiple sclerosis and differentiates groups with and without upper limb tremor", Progress in MS Research Conference, Sydney, Australia.

Conference papers and posters

- Noffs, G., Boonstra, F., Perera, T., Butzkueven, H., Kolbe, S., Maldonado, F., Lizama, E., Galea, M., Stankovich, J., Evans, A. and Vogel, A., 2020, December. Objective measurement of speech correlates with disease status and quality of life in people with MS without dysarthria. European Congress for the Treatment and Research of Multiple Sclerosis (online), 2020. Abstract published in *Multiple Sclerosis Journal*, Vol. 26, No. 3 SUPPL, pp. 176-176.
- Vogel, A., Noffs, G., Boonstra, F., Perera, T., Kolbe, S., Stankovich, J., Butzkueven, H., Evans, A. & Walt, AD (2020). Speech biometrics can predict cerebellar dysfunction in multiple sclerosis. International Congress of Parkinson's disease and movements disorders (online), 2020. Abstract published in *Movement Disorders*, 35, pp. S538-S538.
- Noffs G, Boonstra F, Boonstra FMC, Kolbe S, Perera T, Maldonado F, Iva P, Evans A, Butzkueven H, van der Walt A, Vogel A. Cognitively demanding speech tasks are predictive of cerebellar function and structure. European Congress for the Treatment and Research of Multiple Sclerosis, 2019, Stockholm, Sweden.
- Noffs G, Boonstra F, Kolbe S, Perera T, Shanahan C, Evans A, Butzkueven H, van der Walt A, Vogel A. Under three seconds: objective speech markers of ataxia in Multiple Sclerosis. International Congress of Parkinson's Disease and Movements Disorders, 2018, Hong Kong.
- Noffs G, Boonstra F, Kolbe S, Perera T, Shanahan C, Evans A, Butzkueven H, Vogel A, van der Walt A. Objective speech marker correlates with clinical scores in non-dysarthric MS. Progress in MS research conference 2017, Sydney, Australia.
- Noffs G, Boonstra F, Kolbe S, Perera T, Shanahan C, Evans A, Butzkueven H, Vogel A, van der Walt A. Subclinical speech signs correlate with MS disease severity and differentiate patients with

and without clinical cerebellar dysfunction. European Congress for the Treatment and Research of Multiple Sclerosis, 2017, Paris, France.

Awards

- Melbourne Research Scholarship and Fee Remission, 2017-2020
- Best 3 minutes "Blitz" presentation, Melbourne Brain Symposium (travel bursary), 2017

Acknowledgements

I dedicata this thesis to all my better versions. To Luciana, who brightens my path in so many ways every day. To my family, Damauri, Sergio and Guilherme, for their unwavering support and love, even from afar.

I am fortunate for my PhD supervisors, to whom I am sincerely grateful. Thank you, Professor Adam Vogel, for giving me the opportunity to join the team, for your trust, patience, constant availability, generous support, for your intellectual guidance and sharp sense of humour. Thank you, A/Prof Anneke Van der Walt, for the invaluable teachings and insights into Neurology and Neuroscience, for securing funding, for your patience, stimulating chats and warm advice, and for your generous career support in research and in Medicine. Thank you, Dr Thushara Perera, for your guidance in the world of measurements and in statistical methods and for multiple and well-humoured article revisions. Thank you, Dr Scott Kolbe, for introducing me to the importance of precision of thought (and writing) in science and for guidance and insights in neuroimaging. Thank you, Professor Helmut Butzkueven, for your calmness, trust, career advice and support, and insights into the "big picture" of what we do.

I would like to thank the University of Melbourne International Scholarship program for financially supporting my studies. I am thankful to my advisory committee, especially A/Prof Bryony Nayagam, for her constant support and always practical advice. I am also thankful to all the team in the Multiple Sclerosis and Neuroimmunology research office at the Royal Melbourne Hospital, especially Mrs Josephine Baker and Mrs Lisa Taylor, for patiently helping with participants' management. I am thankful to the co-authors in the articles included in this thesis for their intellectual contributions. Last but not least, I would like to thank all participants who not only donated their time and effort but who made the countless assessments enjoyable with their fantastic humour and friendliness.

Table of Contents

List of tables..... 5

List of figures..... 6

Chapter 1: Background and aims.....8

1.1. Multiple sclerosis..... 8

1.1.1. Epidemiology..... 8

1.1.2. Etiology .....**Error! Bookmark not defined.**

1.1.3. Clinical presentation and course..... 10

1.1.4. Pathology 12

1.1.5. Treatment 14

1.1.6. Measuring disease status and monitoring progression..... 15

1.2. Speech assessment in MS 21

1.2.1. Acoustic analysis of speech..... 22

1.3. Background summary..... 23

1.4. Aims..... 24

1.5. Hypotheses..... 25

1.6. Outline of following chapters..... 25

Chapter 2: What speech can tell us: a systematic review of dysarthria characteristics in Multiple Sclerosis..... 28

2.1. Abstract 29

2.2. Introduction..... 29

2.3. Search methods..... 31

2.4. Results 34

2.4.1. Overall dysarthria characteristics 34

2.4.2. Articulation 38

2.4.3. Voice 39

2.4.4. Respiratory support 41

2.4.5. Prosody 41

2.4.6. Dysarthria and MS disease course 42

2.4.7. Dysarthria and cognition 44

2.4.8. Treatment for dysarthria in MS 45

2.5. Considerations 46

2.6. Conclusion 47

2.7. Take-home messages 48

Chapter 3: Speech metrics, general disability, brain imaging and quality of life in MS 49

3.1. Illustrated highlights 50

3.2. Abstract 50

3.3. Introduction 51

3.4. Methods 53

3.4.1. Participants 53

3.4.2. Neurological assessment 53

3.4.3. Speech assessment 53

3.4.4. Self-assessed quality of life 57

3.4.5. MRI data acquisition and analysis 58

3.4.6. Statistical Methods 58

3.4.7. Data availability statement 60

3.5. Results 60

3.5.1. Speech impairment in MS 60

3.5.2. Relationship between speech measures and EDSS 61

3.5.3. Relationship between speech and structural MRI 63

3.5.4. Relationship between speech and quality of life 65

3.5.5. Acoustic composite score 65

3.6. Discussion 67

3.7. Conclusion 70

3.8. Acknowledgements 70

3.9. Funding 70

3.10. Authors contributions 70

3.11. Potential conflict of interest 72

3.12. Supplementary material 73

Chapter 4: Acoustic speech analytics are predictive of cerebellar dysfunction in Multiple Sclerosis 86

4.1. Abstract 87

4.2. Introduction 88

4.3. Methods 89

4.3.1. Participants 89

4.3.2. Neurological assessment 89

4.3.3. MRI data acquisition and analysis 90

4.3.4. Speech assessment protocol 90

4.3.5. Patient reported quality of life 94

4.3.6. Statistical Methods 95

4.3.7. Relationship between characteristics of speech and cerebellar impairment 95

4.3.8. Determination of an acoustic cerebellar score 95

4.3.9. Data availability 96

4.4. Results 97

4.4.1. Speech phenotype 97

4.4.2. Acoustic composite score for cerebellar dysfunction 97

4.4.3. Relationship of the acoustic composite score with clinical scores, MRI measures and quality of life 100

4.4.4. Prediction of upper hand dexterity by the acoustic composite score 100

4.5. Discussion 102

4.6. Conclusion 105

4.7. Acknowledgements 105

4.8. Funding 105

4.9. Authors roles 105

4.10. Possible conflicts of interest 106

4.11. Supplementary material 108

Chapter 5: The effect of mobile recording devices on speech acoustic metrics 112

5.1. Abstract 113

5.2. Introduction 114

5.3. Methods 117

5.3.1. Materials 117

5.3.2. Participants 117

5.3.3. Tasks 118

5.3.4. Recording protocol 118

5.3.5. Acoustic analyses 119

5.4. Results 122

5.4.1. Microphone characteristics 122

5.4.2. Agreement coefficients 123

5.5. Discussion 131

5.6. Conclusion 136

5.7. Open practices statement 136

5.8. Supplementary material 137

Chapter 6: General discussion 143

6.1. Future research directions 150

6.2. Conclusion 151

References 151

Appendices 167

List of tables

Table 2.1. Glossary of commonly used terms in the assessment of speech deficits 33

Table 2.2. Summary of main speech deficits found in MS. 37

Table 3.1. Speech items included in the analyses. 54

Table 3.2. Descriptive statistics for groups and subgroups. 61

Table 3.3. Correlation coefficients between perceptual speech measures, disability scores and quality of life scores. 62

Table 3.4. Correlation coefficients between acoustic speech measures, disability scores and quality of life scores. 64

Table 3.5. Variables in the regression model for EDSS. 66

Table 3.6. Group comparisons for perceptual analysis of speech. 73

Table 3.7. Group comparisons for acoustic analysis measurements. 78

Table 3.8. Correlation coefficients between speech and imaging measures. 82

Table 4.1. Speech variables grouped by domains and assessment method..... 92

Table 4.2. Demographics and clinical characteristics by groups. 98

Table 4.3. Percentage of participants who presented cerebellar lesions by location. 99

Table 4.4. Correlations between the acoustic composite score and measures of cerebellar impairment. 101

Table 4.5. Spearman's (rho) and Pearson's (r) correlation coefficients between speech features and general MS pathology and disability. 108

Table 4.6. Spearman's (rho) and Pearson's (r) correlation coefficients between cerebellar scores and individual speech items. 109

Table 4.7. Correlations between speech MRI measures. 111

Table 5.1. Summary list of speech measurements tested for agreement. 121

Table 5.2. Comparative response characteristics of microphones for the same sound stimulus. 123

Table 5.3. Bias and 95% confidence interval of agreement between Reference and tested microphones for frequency-dependent measurements. 127

Table 5.4. Bias and 95% confidence interval of agreement between Reference and tested microphones for timing measurements. 129

List of figures

Figure 1.1. Prevalence of MS by country 9

Figure 1.2. Prevalence of symptoms including all MS stages. 11

Figure 1.3. Schematic representation of EDSS step components and distribution of people with MS 17

Figure 2.1. PRISMA flow diagram of the reviewing process 32

Figure 2.2. Overview of impaired speech characteristics in relation to overall disability (EDSS scores), severity of dysarthria and disease course 36

Figure 3.1. Confidence intervals for global speech scores, naturalness and acoustic composite score across subgroups. 66

Figure 4.1. ROC curve of accuracy for predicting an abnormal 9HPT in the MS group with the acoustic composite score as predictor. 102

Figure 5.1. Power spectral density levels for mid-vowel /a/, healthy subject, female. 120

Figure 5.2. Representative Bland-Altman plot with limits of agreement, 95% CIs and 95% CIA. 121

Figure 5.3. Average power spectral density (PSD) levels (re 10-5 Pa) for voices of males and females, by microphone. 124

Figure 5.4. Bland-Altman plots and 95% CIA (dotted lines) for f0, showing good agreement between reference and test microphones. 125

Figure 5.5. Bland-Altman plots and 95% CIA (dotted lines) for monologue pause percentage, showing poor agreement between reference and test microphones. 125

Figure 5.6. Relationship between neurological disability status and acoustic score for the high-grade reference microphone set and the in-built microphone. 126

Chapter 1

Background and aims

1.1. Multiple sclerosis

Multiple sclerosis (MS) is a chronic, immune-mediated inflammatory disease of the central nervous system (CNS). The name *multiple sclerosis* derives from the French *sclerose en plaques* (meaning *stiff plaques*) used in autopsy examinations since 1868 by Charcot (Charcot, 1877), referring to the macroscopic appearance of scar-like plaques of hard consistency relative to the surrounding usually soft brain tissue and spinal cord. Rather than encompassing all current knowledge about MS, the following sections aim to highlight key concepts related to the disease's importance and phenotypes, the recent advancement in treatment efficacy, and strategies for monitoring progression, which underly the rationale for this thesis.

1.1.1. Epidemiology

MS is the most common cause of neurological disability in young adults in many countries (WHO, 2013). An estimated 2.3 million people worldwide live with MS. Prevalence varies from <1 to 189 per 100 000 inhabitants, being highest in Europe and North America (Figure 1). Within regions, higher prevalence is associated with greater geographical latitude. Females are affected two to three times as often as males, and disease onset can occur at any age but most often does between 20 and 40 years (WHO, 2013). The annual economic burden of MS was recently estimated at US$4.3 billion in the United States of America (Chen, Chonghasawat, & Leadholm, 2017) and €14.6 billion in Europe (Gustavsson et al., 2011). In Australia, 1 per 1000 people living with MS, and the average annual cost per person due to the

disease was estimated in AU$ 68.4 thousand, 32% of which being from productivity loss (Ahmad, Palmer, Campbell, van der Mei, & Taylor, 2018).

Figure 1.1. Prevalence of MS by country

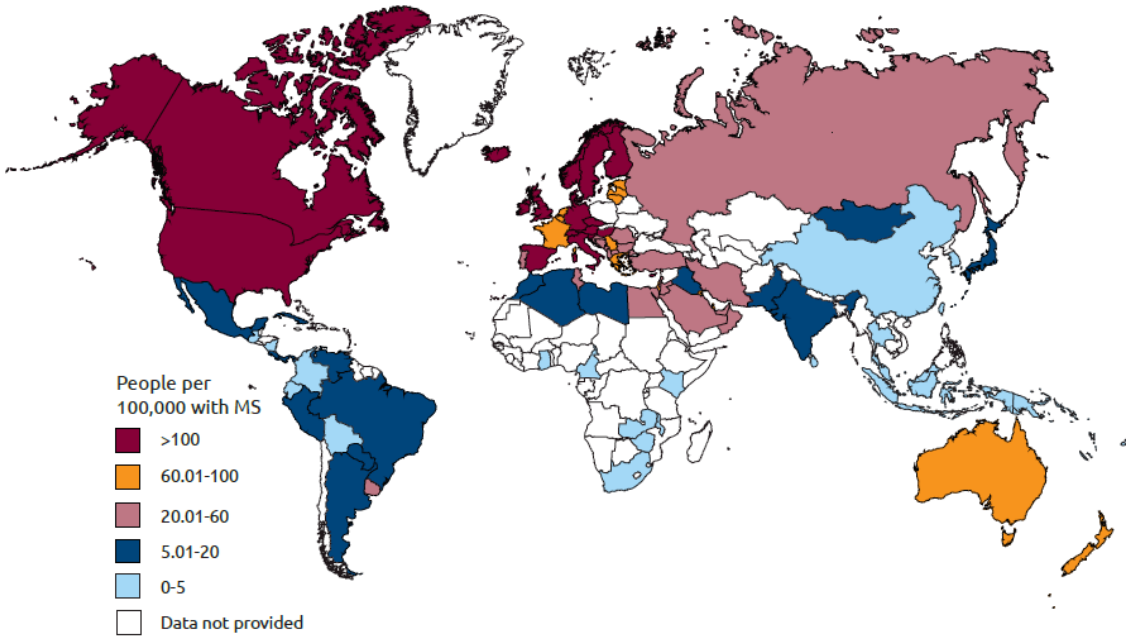


Image taken from the Atlas of MS (WHO, 2013).

1.1.2. Aetiology

The cause of MS is unknown, but a few genetic and environmental risk factors have been established. Genetic susceptibility is supported by shared MS diagnosis being more frequent between homozygotic "identical" twins in comparison to dizygotic twins (Fagnani et al., 2015). Additionally, several genetic regions have been associated with MS, with an approximate 20% genetic risk linked to specific haplotypes of the human leukocyte antigen (HLA) complex (Olerup & Hillert, 1991).

Support for environmental influence on the development of MS includes an inverse association with childhood exposure to sunlight (van der Mei et al., 2003) and adult vitamin D supplementation (Munger et al., 2004); and direct association with geographical latitude (WHO, 2013), positive serology for Epstein-Barr virus (Ascherio & Munger, 2010), obesity

during adolescence and tobacco smoking (Olsson, Barcellos, & Alfredsson, 2017). Environmental risk is significantly supported by an increase in the risk of developing MS for people who migrated early in life from a country with low to a country with a high prevalence of the disease (Munk Nielsen et al., 2019). Synergistic interactions occur between genetic and environmental factors, namely HLA haplotypes and smoking, obesity during adolescence or Epstein-Barr virus infection, where the combined risk is significantly greater than the sum of individual risks (Olsson et al., 2017).

1.1.3. Clinical presentation and course

MS is highly heterogeneous in symptomatology and disease course. In approximately 85% of cases, symptom onset is characterized by an acute or subacute neurological impairment (e.g., loss of sensation in the legs) that lasts more than 24 hours and does not have an apparent causative event (e.g., trauma, infection). That first attack is called clinically isolated syndrome (CIS). The first symptom in CIS will be somatosensory in 40% of cases, such as paresthesia (described as skin tingling or pins-and-needles), reduced proprioception (perception of the position of limbs), reduced sensation to touch, vibration or pain. Around 30% of CIS will initially present with a motor deficit such as paresis (significant reduction in strength of a muscle group), spasticity (deficit in muscle relaxation) and pyramidal signs (abnormal neurological reflexes). Another 25% will initially experience optic neuritis, typically characterized by partial or total vision loss in one eye (large black patch in the visual field) associated with ocular pain worsened by eye movement and dyschromatopsia (impairment in colour perception)(Filippi et al., 2018). Typically, the initial impairment worsens for a few days, followed by a variable level of recovery from minimal to total, which is reached within about six weeks from symptom onset. New attacks can occur (relapses) spaced by periods without change (remission), which characterizes conversion to relapsing-remitting MS (Brownlee, Hardy, Fazekas, & Miller, 2017). Around two-thirds of people with relapsing-remitting MS

(RRMS) will later experience ongoing worsening of symptoms independently from relapses, which configures conversion to secondary progressive MS (SPMS)(Scalfari et al., 2010). The remaining 15% of people with MS who do not have a CIS are classified as primary progressive MS (PPMS), which is defined by continuous neurological deterioration from onset without marked relapses or remission (Brownlee et al., 2017).

Figure 1.2. Prevalence of symptoms including all MS stages.

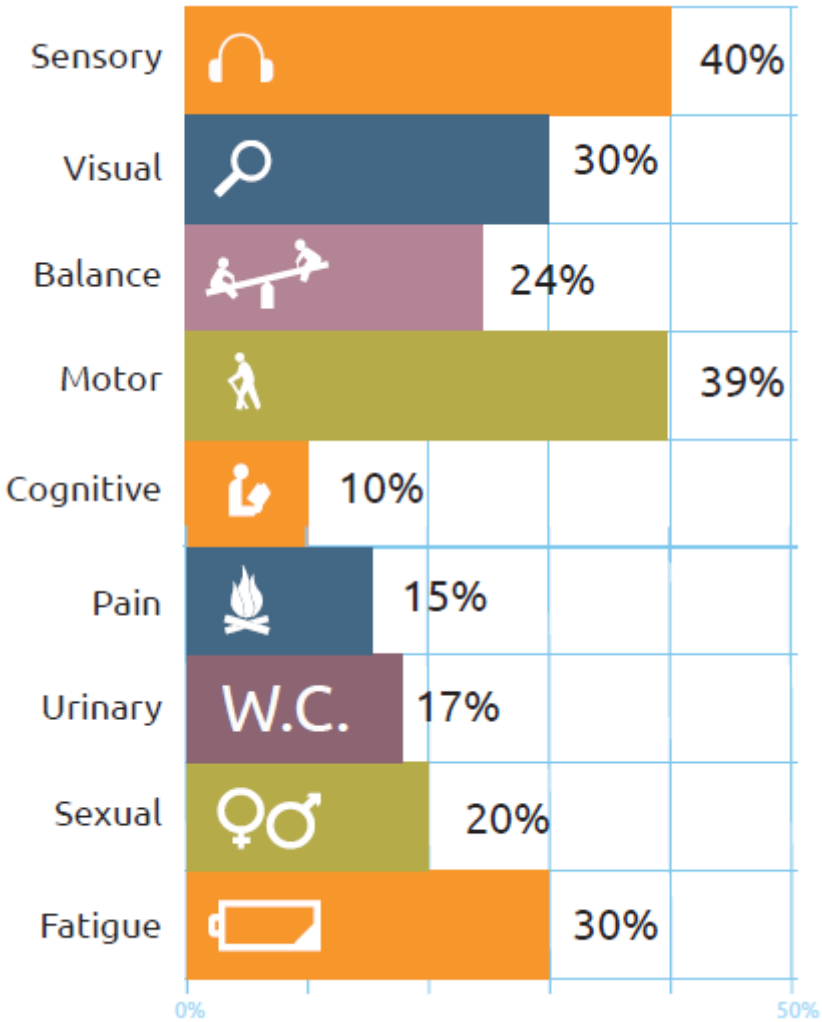


Image taken from the Atlas of MS (WHO, 2013). As the disease progresses, new deficits may occur and remit in variable degrees, resulting in countless possible individual combinations and consequent large clinical heterogeneity. However, the reported prevalence of symptoms depends on the instrument used to measure them, a limitation further discussed in section 1.6. (monitoring progression).

1.1.4. Pathology

The normal, disease-free CNS harbours very few immune cells (Prinz & Priller, 2017). These include mainly microglia in the parenchyma, memory T cells in the cerebrospinal fluid (CSF), and macrophages at the CNS barriers (choroid plexus, meninges, and perivascular space)(Prinz & Priller, 2017). Conversely, in MS, peripheral immune cells migrate into the CNS and effect damage to the myelin sheath and axons. The initial site of immune activation, whether inside the CNS (intrinsic model) or in the periphery (extrinsic model with cross-autoimmunity), has been discussed in pathogenesis models. However, the initial site of activation, trigger(s) and first cellular events that lead to activation and migration of immune cells into the CNS remain unknown.

The main cell types migrating to the CNS in people with MS are macrophages, cytotoxic and helper T-cells, and B-cells, in order of density. The relevance of cellular migration to disease activity is firstly supported by the preferential distribution of lesions being close to CNS barrier sites. After entering the CNS, the immuno-infiltrate form localized plaques in the spinal cord and/or brain white matter, most commonly around the brain ventricles (choroid plexus, CNS barrier)(Dendrou, Fugger, & Friese, 2015). Additionally, early organized meningeal infiltrates (meninges being another CNS barrier) are observed about as often as white matter plaques and are linked to damage of the underlying cortical neurons (Haider et al., 2016; Lucchinetti et al., 2011; Magliozzi et al., 2010).

At early stages, tissue between plaques is not infiltrated by peripheric immune cells but may already express activation of local microglia (Giannetti et al., 2015). The proinflammatory medium produced by migrated and resident cells, particularly via cytokines in the CSF, might be the direct driver of microglia activation. Injection of CSF from people with MS is sufficient to cause demyelination in the brain of mice, and the extent of damage in mice reflects the rate of progression in the human donor (Cristofanilli et al., 2014). Yet stronger

support for the crucial role of cellular migration and humoral inflammatory pattern on disease progression derives from the high efficacy of current pharmacological treatments which target those phenomena (summarized next)(Buzzard, Broadley, & Butzkueven, 2012).

The role of each cell type and interactions between them have increasingly been described in relation to migration, inflammatory up-regulation and maintenance, myelin and neuronal damage (i.e., pathophysiology), and myelin repair or failure thereof - see (Dendrou et al., 2015) for a review. The last active step in effecting structural damage seems to involve activated resident microglia, astrocytes and possibly migrated macrophages and T CD8+ cells, causing direct molecular disarrangement to the myelin sheath and to neurons through targeted release of oxidative radicals (Fischer et al., 2013; Lassmann, 2014). That oxidative damage is more frequently observed at the preferential lesion sites, i.e., superficial cortical layers and periventricular white matter. Conversely, neurons at deep cortical layers are less prone to suffer direct oxidative damage but rather exhibit retrograde (Wallerian) degeneration, which is otherwise observed after axonal transection (such as in brain concussion), and therefore presumed to be an indirect consequence of damage to axons located in the adjacent white matter (Haider et al., 2016). Even when not transacted, axons without myelin require considerably more energy to maintain the transmembrane electric potential or to repolarize after transmitting an impulse. That ongoing high demand becomes unmet by energy supply as mitochondrial function and structure are impaired by the pro-inflammatory medium, eventually resulting in cellular energy exhaustion. In that phase, neuroaxonal injury is similar to that observed in hypoxic diseases such as stroke or cerebral palsy. Without enough energy to sustain the Na+/K+ pump, Na+ accumulates inside the axon, leading to a secondary and sustained influx of Ca++, which in turn triggers apoptosis (Campbell et al., 2011; Trapp & Stys, 2009).

As the disease progresses, new and old plaques pass to coexist with a more diffuse inflammatory infiltrate between plaques (Kutzelnigg et al., 2005). Clinical progression is histologically linked to more irreversible neuroaxonal loss, with consequent gliotic substitution and brain atrophy (Dendrou et al., 2015).

1.1.5. Treatment

Pharmacological treatment for MS has evolved rapidly in the last two decades, both in efficacy and options available, with the development of disease-modifying treatments (DMTs). DMTs reduce the relapse rate, magnetic resonance imaging (MRI)-detected new lesions and disability accumulation in 20% to 50% on average for people with CIS, RRMS or SPMS (Comi, Radaelli, & Soelberg Sørensen, 2017). The mechanisms of action responsible for the efficacy of each well-established DMTs are only partially understood and include mainly reduction in migration of lymphocytes (e.g., natalizumab, fingolimod, dimethyl fumarate), shifting the immuno-response from proinflammatory Th1 to anti-inflammatory Th2 type (β-interferon, glatiramer acetate), mitigation of oxidative stress (dimethyl fumarate), and possibly modulation of astrocytes (fingolimod). Emergent drugs act by depleting selected lymphocytes (ocrelizumab, alemtuzumab, teriflunomide) and reducing the migration of lymphocytes (siponimod). However fragmented, knowledge of DMTs' pharmacodynamics point to potential drugs with similar action for testing (e.g., fingolimod and siponimod), and generates or supports pathophysiology hypotheses and vice-versa, which has characterized the current phase of accelerated discoveries (Buzzard et al., 2012; Comi et al., 2017; Kappos et al., 2018).

Optimizing treatment at the individual (clinical care) level represents one of the main challenges in the management of MS at present. Despite reasonably similar rates of success between DMTs at a group level, the efficacy of any given DMT is highly variable between individuals. Additionally, response to a DMT is not constant over time for the same individual and, in general, tend to decrease with higher disease duration or after conversion to a

progressive course (Comi et al., 2017). Heterogeneity among DMTs side effects also adds complexity to decision making. DMTs differ greatly in the frequency and severity of adverse events. Riskier but potentially more effective DMTs might be particularly beneficial early on to people with aggressive progression (Brown et al., 2019). Furthermore, the type of potential adverse effect of one DMT (e.g., hepatic failure) can make it less recommendable for patients with related comorbidities (e.g., hepatic insufficiency) or for certain demographic populations such as older patients or females in the reproductive age group. Such multi-variate treatment options coupled with the disease course's heterogeneity warrants personalised and dynamic treatment decisions (Bielekova & Martin, 2004; Comi et al., 2017; Gross & Corboy, 2019).

Recent advancements in the management of real-world data and in statistical methods might help refine personalised care. Consolidation of large international patient registries such as MSBase (Martin et al., 2006) has offered invaluable insights into MS disease course and treatment response (Trojano et al., 2017). So far, registries recorded mostly data from standard clinical appointments, some demographic information and standard MRI metrics. Coupled with statistical techniques for "big data" observational databases, these registries enabled the modelling of individual predictors for relapses (Kalincik et al., 2014; Tintore et al., 2015) and for response to different DMTs (Kalincik et al., 2017; Spelman et al., 2016). Such models of prognosis and treatment response have been used so far only to inform clinicians through journal publications rather than as an interactive tool. Future integration of emerging paraclinical assessments into databases and respective modelling will likely result in key advantages for personalized and more dynamic treatment decisions, as discussed next.

1.1.6. Measuring disease status and monitoring progression

Several methods are currently in use to assess MS disease status and progression, which in turn inform treatment decisions. The most likely reasons for changing DMT treatment are relapses, clinical deterioration, and MRI-detected breakthrough activity (Saccà et al.,

2019). The following section briefly discusses advantages and limitations of the main current clinical and MRI assessment methods in MS.

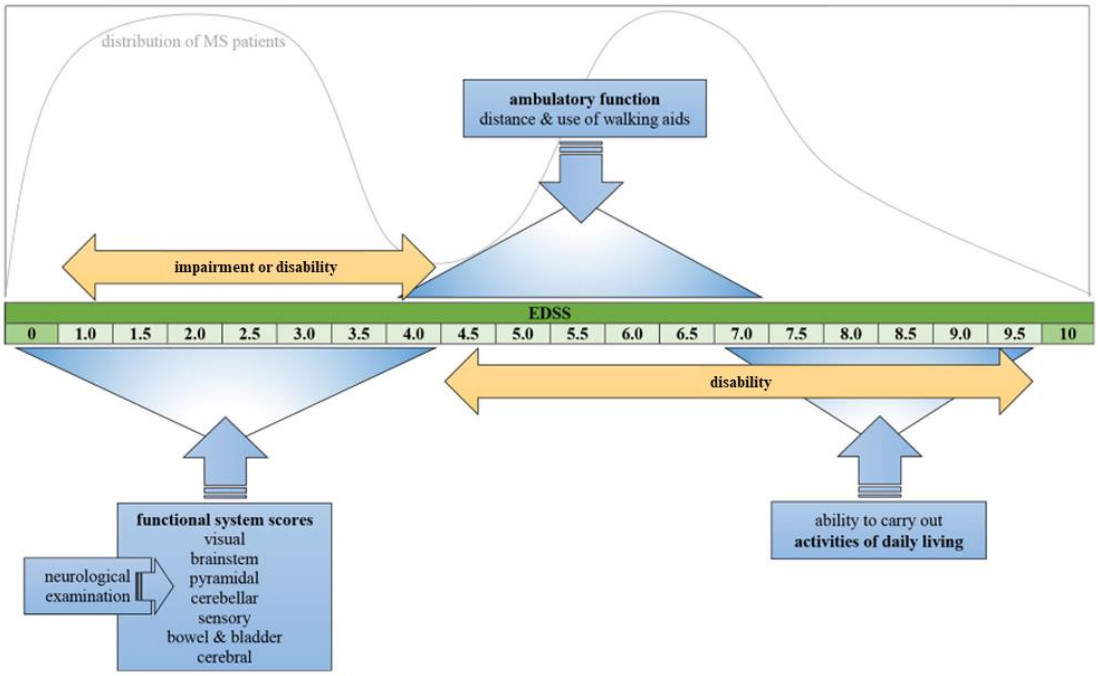
Clinical assessment of general neurological impairment and disability

Due to MS heterogeneity, a comprehensive assessment of multiple neurological systems is necessary to capture all impairment related to the disease. The Expanded Disability Status Scale (EDSS)(Kurtzke, 1983) is a scoring system often used for clinical grading of MS, where eight functional systems (FS) are independently scored, namely visual, brainstem, pyramidal, cerebellar, sensory, bowel and bladder, cerebral and ambulation. The scoring of each FS is based on landmark impairments within that system, reducing clinical subjectivity and streamlining the process. The "EDSS step" score is a summary score that takes into account all FS scores, and which ranges from zero (normal neurological exam) to ten (death due to MS) at half-point increments. EDSS scoring requires nearly no equipment and is composed mostly of tests from standard neurological examination, favouring its widespread clinical use. Due to its long-term and frequent use in research, its measurement properties (e.g. validity, reliability) and cross-domain associations (e.g. with quality of life or health-care costs) are very well-established, thus offering means for extensive historical and cross-studies comparisons (Hobart, Freeman, & Thompson, 2000; Meyer-Moock, Feng, Maeurer, Dippel, & Kohlmann, 2014; Naci, Fleurence, Birt, & Duhig, 2010; Patwardhan et al., 2005).

Several characteristics, however, limit EDSS' performance for measuring impairment and disability status. The scale largely overlooks cognitive impairment and fatigue despite those features translating into significant personal and social burdens in MS (van Munster & Uitdehaag, 2017). Additionally, the various neurological sub-systems integrate almost exclusively the lower range of the scale (EDSS step < 4), whereas ambulation alone largely defines scores between four and seven. Thus, higher EDSS scores often bear little association with impairment/disability level other than ambulation (Hobart et al., 2000). Not surprisingly,

EDSS scores present a bimodal distribution (figure 3)(van Munster & Uitdehaag, 2017), in practice behaving like two distinct scales artificially put in tandem. Its disproportionate composition, non-linearity and ordinal character impose the use of less powerful and less versatile non-parametric statistical methods in research.

Figure 1.3. Schematic representation of EDSS step components and distribution of people with MS



Adapted from van Munster & Uitdehaag (2017).

EDSS presents additional limitations for monitoring disability progression. Both EDSS and FS show suboptimal inter and intra-rater reproducibility (Hobart et al., 2000; Meyer-Moock et al., 2014). Poor reproducibility is even more evident at EDSS < 4 (Goodkin et al., 1992), a stage at which swift change in treatment has the greatest potential for preventing disability. Current recommendations to increase EDSS reliability include a minimum change of

1 EDSS point (2 steps) if previous EDSS < 6, and/or confirmation of EDSS change after six months (van Munster & Uitdehaag, 2017). Unfortunately, such stabilizing strategies tend to decrease even further the score's already poor responsiveness to change (Hobart et al., 2000; Sharrack, Hughes, Soudain, & Dunn, 1999). Despite its limitations, EDSS remains a very comprehensive, practical and widely used tool for registering disease status in MS.

Other clinical scales have been proposed to address shortcomings of the EDSS. The Multiple Sclerosis Functional Composite (MSFC)(Cutter et al., 1999), for example, produces linear and continuous scores (time to complete tasks or number of right answers), which favour parametric statistical testing. MSFC has better reliability in comparison with the EDSS and seems more responsive to change, particularly in the ambulation-based range EDSS 4-7 (Cohen et al., 2002). Like EDSS, MSFC baseline scores and change in scores are associated with quality of life scores, employment status, MRI metrics, and conversion to SPMS (Rudick et al., 2009). MSFC includes formal testing of cognitive function but neglects various other systems (e.g. sensory, autonomic, depression, fatigue), may be difficult to interpret clinically, is less well tolerated by patients and has a much smaller literature database.

Monitoring cerebellar dysfunction in MS

System and symptom-specific scales also have a role in measuring impairment and disability in MS, but with different advantages and limitations. The Scale for the Rating and Assessment of Ataxia (SARA)(T Schmitz-Hübsch et al., 2006), for example, is a 20 points (40 incremental steps) rating protocol used to clinically grade ataxia in various diseases. SARA produces one score per cardinal ataxic domain (e.g. truncal ataxia, gait, speech) and an overall (summative) ataxia score. Thus, SARA will not capture status or change in other systems affected by MS, making it unsuitable for stand-alone tracking of progression or treatment effect. Conversely, SARA offers a more detailed characterization of ataxia and a multi-point distribution in comparison with general scales, which is desirable to study disease progression

(Bürk et al., 2009) or specific relationships of ataxia with other domains (e.g. with localized brain pathology or with quality of life)(Hernandez-Castillo, Galvez, Diaz, & Fernandez-Ruiz, 2016). Additionally, because SARA is not specific to MS, scores can be compared between diseases. Alternatively, the nine-hole peg test (9HPT)(Mathiowetz, Weber, Kashman, & Volland, 1985) is often used to assess the severity of ataxias and treatment effects (Corben et al., 2010; Lynch, Farmer, Wilson, & Balcer, 2005; T. Schmitz-Hübsch et al., 2008; Zesiewicz et al., 2012) as it offers an objective assessment of upper limb dexterity and produce continuous values, i.e., time to complete the task. Furthermore, measurements from the 9HPT have good validity and reliability performance in MS in relation to both disease presence and severity of disability (Feys et al., 2017).

MRI assessment

MRI-derived brain volume and lesion load are currently among the most studied and used paraclinical measurements of disease state and progression in MS (Gajofatto, Calabrese, Benedetti, & Monaco, 2013). Calculation of MRI metrics is currently (mostly) automatable and offer objective and continuously distributed data and easily interpretable results (a priori, see limitations below). Baseline whole-brain MRI lesion load is predictive of conversion from CIS to definite MS (Barkhof et al., 1997; Fisniku et al., 2008; Tintore et al., 2010), while lesion load is weakly associated with disability status (Shiee et al., 2012) and progression in the following four years (Rovaris et al., 2003) to twenty years (Fisniku et al., 2008). Accumulation of lesion load within five years of disease onset has a slightly stronger relationship with long-term disability than baseline imaging alone, whereas that association seems to become weaker with longer disease duration (Fisniku et al., 2008). Brain volume (Johnen et al., 2019; Sanfilipo, Benedict, Sharma, Weinstock-Guttman, & Bakshi, 2005; Shiee et al., 2012) and, most consistently, change in brain volume (i.e. atrophy) also modestly correlate with disability progression as measured by EDSS or MSFC (Fisher et al., 2000; Zivadinov et al., 2013).

This lack of strong association is extensively documented and has been termed the clinico-radiological paradox in MS, which is a limitation of current standard MRI for monitoring disease progression. That paradox may be explained by limitations in measurements of clinical disability as previously discussed (e.g. EDSS, MSFC), technical limitations of standard MRI acquisition (e.g. the vast majority of cortical lesions not being detected in T1-T2 sequences)(Geurts, Bö, et al., 2005) and limitations in statistical modelling. Furthermore, standard structural imaging does not reflect biochemical and many histological changes which, in addition to plaques, are responsible for short-term functional impairment and long-term atrophy (e.g. diffuse inflammatory activity and neuronal damage in normal-appearing white matter)(Barkhof, 1999; Filippi et al., 2012; Kutzelnigg et al., 2005). Lastly, brain atrophy progresses slowly, whereas its measurement can be distorted by common transient disease events (e.g. reducing inflammatory swelling through treatment, which results in paradoxical pseudoatrophy) and by concurrent pathology (e.g. diabetes)(Wattjes et al., 2015).

Recent research addresses the clinico-radiological paradox with substantial success through advances in imaging techniques (Filippi et al., 2011; Geurts, Pouwels, et al., 2005) and advances in statistical modelling (e.g. weighted-combining lesion load and/or atrophy for segmented areas or whole brain (Hackmack et al., 2012; Popescu et al., 2013). Modern MRI techniques also begin to explore functional, microstructural and biochemical (fluid) composition.

Other non-MRI MS biomarkers measure morphological changes (e.g. optical coherence tomography), inflammatory and degenerative signalling (e.g. CSF and blood components) and neuro-transmission disruptions (electrophysiological testing, e.g. evoked potentials)(Gajofatto et al., 2013). While new MRI acquisition sequences and biomarkers are progressively validated, they substantially add scanning/assessment time, energy load, patient burden and financial cost, thus limiting frequent assessments required for swift adjustment in treatment.

Summary

MS-specific functional scales like EDSS assess multiple neurological sub-systems to capture the most common impairments and apply a summary score to standardize the quantification of disability in this heterogeneous disease. Conversely, syndrome-specific scales frequently offer a more detailed assessment of individual sub-systems, may be compared across different diseases but cannot be used in isolation for tracking MS disease progression. In this study, EDSS is used as a measure of general neurological disability, and SARA and 9HPT represent the main assessment tools of cerebellar ataxia. Standard MRI measurements offer a readily available *in vivo* assessment of the brain's morphology and pathology in pwMS through lesion metrics and brain volume. Those measures increase sensitivity and objectivity of diagnosis, measurement of disease status and tracking of progression. Recent imaging research seems to successfully address their main technical limitations, such as undetected cortical pathology, which might further increase the validity of MRI metrics for future clinical practice. In the current study, the standard MRI measurements of lesion load and brain volume were used to characterize neuro-anatomopathology in pwMS in relation to speech and other functional outcomes.

1.2. Speech assessment in MS

Impairment of speech production is only marginally assessed within the EDSS and not formally included in the MSFC (although timed oral responses in neuropsychological tests likely influenced by motor speech impairment, as discussed in chapter two). Various types of speech impairment may occur in people with MS (pwMS) given that multiple neurological networks are involved in speech production (e.g., cognitive, sensory, motor)(Hickok, 2012). "Telegraphic" or *stacatto* speech was one of the three major symptoms, along with intention tremor and nystagmus, described by Charcot as typical in pwMS (Charcot, 1877), while modern

assessments describe a plethora of speech deviations reviewed in detail in chapter two. Current evidence points to a prevalence of overall executive speech impairment, i.e. dysarthria, in about 50% of pwMS as judged by blinded expert listeners (Hartelius, Runmarker, & Andersen, 2000). Nonetheless, speech's widespread neurological reliance (e.g., cognitive, affective, sensory, motor) implies that some speech impairment, although perhaps subtle, is expected to occur in most pwMS, even at low disability levels. Concurrent with this study, emerging evidence from Parkinson's (Hlavnička et al., 2017), Huntington's Disease (Chan, Stout, & Vogel, 2019) and MS (Fazeli et al., 2020; J. Rusz, Benova, et al., 2018) suggest that changes in speech production can be objectively measured before overt dysarthria becomes noticeable. The ability to measure subclinical speech impairment would translate into an even larger prevalence of speech deficits in MS and could offer an opportunity for disease monitoring and treatment adjustments that do not depend on the occurrence of (often irreversible) disability.

1.2.1. Acoustic analysis of speech

Objective speech data could offer several desirable qualities for assessing disease status and progression, conditional to speech's validity as a proxy for neurological impairment in MS. In contrast to expert-based speech analysis, objective speech assessments could gauge subclinical changes in speech as it does not depend on overt speech impairment; improve reliability via elimination of inter and intra-rater variability; and facilitate scalability and reduce assessment costs, depending on the objective speech method selected. Furthermore, objective measurements, in general, permit the use of a much wider variety of, and more powerful parametric statistical methods, as discussed earlier. Continuous and normally distributed speech data could further be integrated with other objective measurements (e.g. hand dexterity, cognitive tests, time to walk a certain distance) for a comprehensive and objective representation of neurological impairment in MS.

Several objective methods assess speech-related structure and function. These include, but are not limited to, structural and functional imaging (e.g., functional speech MRI, endoscopic laryngoscopy, fluoroscopy), electrophysiological tests (e.g., laryngeal electromyography, glottography), mechanical tests (e.g., tongue force and speed), and acoustic analysis. Key advantages of acoustic analysis in relation to other methods derives from the signal under analysis, i.e., the speech soundwave, which is obtained non-invasively, demands low effort levels for production (spontaneously produced speech can also be analysed) and requires common technology for its recording and analysis (a priori). Another related advantage is the potential analysis of language in the same speech signal, which cannot be obtained from other analyses' methods (e.g., from MRI or electroneuromyography signals), which would increase the neurological representativeness of such assessments. Acoustic analysis is also arguably the most studied among objective speech assessments, offer a wealth of metrics, shares applicability across different fields (e.g., sound engineering, digital audio compression, geology, medical imaging, speech and hearing sciences) which accelerates its forward technical advancement. The current computational capability of portable devices, such as smartphones, and the widespread internet connectivity may further enable the use of automated acoustic analysis in remote settings and engage and empower patients through assisted self-monitoring. Thus, acoustic analysis is the objective method of speech measurement used throughout this study. Chapter 2 summarizes the specific speech features included in acoustic analyses in later chapters.

1.3. Background summary

Multiple Sclerosis is a common cause of progressive neurological impairment in young and middle-aged adults worldwide. Modern disease-modifying therapies can greatly slow MS progression. Optimizing treatment is, however, often not straightforward given the wide

heterogeneity of both disease and treatment options. Swift treatment adjustments would greatly benefit from personalized and dynamic motoring of disease progression. Clinical ratings, clinical relapses and MRI-measured lesions and brain volume constitute the current standard assessment of MS status, which often requires either confirmation after months or extensive new impairment/disability (both usually irreversible), thus limiting dynamic monitoring. Measurements from acoustic analysis of speech are objective, continuous and offer key advantages for frequent monitoring such as its succinctness, low cost and automation. The current worldwide easy access to smartphones and mobile devices might also allow for widespread remote or self-assessment of speech. Furthermore, measurement of subclinical speech impairment through acoustic analysis might extend the utility of speech assessments to people without dysarthria and favour early intervention. To assess the potential of acoustic analysis of speech to be used as a proxy for neurological status (alone or integrated with other measures), its validity and reliability must be described in detail.

1.4. Aims

The primary aim of this study is to determine the concurrent validity of acoustic speech metrics as a surrogate of disease status in MS. In addition to general MS, subpopulations of interest consist of people with low disability levels, i.e., when EDSS scores have the poorest reliability and responsiveness; people without dysarthria, to test the validity of subclinical speech impairment; and people with cerebellar impairment, for cerebellum's multiple contributions to speech production, and as a prototype of speech modelling of specific sub-systems. A secondary aim of this study is to assess interchangeability between speech acquisition equipment, a component of measurement reliability.

1.5. Hypotheses

i) Acoustic metrics can detect a progressive speech impairment across the spectrum of MS neurological status (chapter 3).

ii) The association between speech measurements and general neurological disability level is present in pwMS without overt dysarthria (chapter 3).

iii) A specific ataxia-modelled acoustic composite score correlates with cerebellar status as assessed by different functional and imaging tests (convergent validity, specificity to neurological sub-system, chapter 4).

iv) The effect size of associations between speech metrics and neurological impairment are robust to variations in speech recording protocols, particularly to the employment of mobile technology (chapter 5).

1.6. Outline of following chapters

Chapter 2: a systematic review of the literature on speech impairment in MS.

- The most consistently affected speech features in MS, according to research up to date, are described. That consistency of impairment is the main selection criterion for the inclusion of speech metrics in the subsequent chapters, from a large number of possible speech metrics. This pre-selection approach based on previous studies prioritizes the content validity of included speech features, and alternative approaches are briefly discussed in the last chapter of this thesis.
- Defines missing, conflicting and preliminary information necessary for validating speech as a marker of disease progression.

Chapter 3: a cross-sectional study of speech in 119 pwMS and 22 healthy controls (HC).

- The relationship between speech and current measurements of general neurological status in MS is characterized by correlations with EDSS, MRI-measured whole brain volume and lesion load, and self-assessed quality of life.
- Speech function measured through acoustic analysis and listener-based rating (perceptual analysis) to highlight communalities/divergences between methods and facilitate speech-clinical contextualization of acoustic metrics.
- Acoustic composite score modelled to reflect general neurological disability.
- Speech measures compared between groups of increasing EDSS scores to estimate the stage at which each speech impairment occurs in MS.
- Association between acoustic metrics and disease status tested in a subgroup of 77 pwMS without speech impairment.

Chapter 4: speech-cerebellar correlates.

- Cerebellar impairment assessed in 85 pwMS through FS score within the EDSS, SARA, 9HPT and cerebellar volumes and lesion load.
- Ataxia acoustic composite score modelled and tested to predict abnormal 9HPT results.
- Similarities between the ataxia-modelled and the general acoustic composite scores are discussed, as well as possible improvements for future acoustic composite models of ataxia.

Chapter 5: acoustic differences between recording devices.

- Speech from each of 18 participants simultaneously recorded a professional-grade set and three microphone configurations of a mobile device.
- Inter-equipment agreement intervals are compared to the expected change in speech for various neurological conditions.
- Frequency response and signal-to-noise measured and discussed.

- The chapter highlights the most robust speech metrics in that context and discusses the main confounders for the flexibilization of speech recording protocols for neurological monitoring.

Chapter 6: general discussion

- A summary of the main findings in previous chapters is presented.
- The combination of results is interpreted and discussed, as well as their relationship with past and emerging research.
- Directions for future advancement in neuro-speech research is outlined.

Chapter 2

What speech can tell us: a systematic review of dysarthria characteristics in Multiple Sclerosis

Published on Autoimmunity Reviews (journal), volume 17, issue 12, pages 1202-1209 in December 2018.

DOI: 10.1016/j.autrev.2018.06.010

AUTHORS

GUSTAVO NOFFS[1, 5], THUSHARA PERERA[3, 4], SCOTT C KOLBE[2], CAMILLE J SHANAHAN[2], FREDERIQUE M C BOONSTRA[2], ANADREW EVANS[5], HELMUT BUTZKUEVEN[2, 5], ANNEKE VAN DER WALT[2, 3, 5, 6, *] AND ADAM P VOGEL[1, 3, 5, 6, *]

1. Centre for Neuroscience of Speech, University of Melbourne, Australia
2. Melbourne Brain Centre, Department of Medicine, University of Melbourne, Australia
3. The Bionics Institute, East Melbourne, Australia
4. Department of Medical Bionics, The University of Melbourne, Australia
5. Department of Neurology, Royal Melbourne Hospital, Melbourne, Australia
6. Department of Neurodegeneration, Hertie Institute for Clinical Brain Research, University of Tübingen, Germany

* Authors contributed equally.

2.1. Abstract

IMPORTANCE: Multiple sclerosis produces neurological impairments that are variable in duration, severity and quality. Speech is frequently impaired, resulting in decreased communication skills and quality of life. Advancements in technology now make it possible to use quantitative acoustic assessment of speech as biomarkers of disease progression. OBSERVATIONS: Four domains of speech have been identified: articulation (slow articulation and imprecise consonants), voice (pitch and loudness instability), respiration (decreased phonatory time and expiratory pressure) and prosody (longer and frequent pauses, deficient loudness control). Studies also explored I) predictive models for diagnosis of MS and of ataxia using speech variables, II) the relationship of dysarthria with cognition and III) very few studies correlated neuroimaging with dysarthria. We could not identify longitudinal studies of speech or dysarthria in Multiple Sclerosis. CONCLUSION AND RELEVANCE: Refinement of objective measures of speech has enhanced our understanding of Multiple Sclerosis-related deficits in cross-sectional analysis while both integrative and longitudinal studies are identified as major gaps. This review highlights the potential for using quantitative acoustic assessments as clinical endpoints for diagnosing, monitoring progression and treatment in disease-modifying trials.

Key words: Multiple Sclerosis, Dysarthria, Speech, Dysphonia, Voice, Cognition, Quality of Life, Deep Brain Stimulation, Acoustics

2.2. Introduction

Multiple sclerosis (MS) is the most common neurological disease to cause disability in young adults (Rejdak, Jackson, & Giovannoni, 2010). Prevalence varies (<1 to 193 per 100 000) depending on ethnic and geographical variables (Rosati, 2001). Different autoimmune processes are implicated in the variability of the disease's expression (Ben-Ami, Miller, & Berrih-Aknin, 2014; Fraussen, de Bock, & Somers, 2016; Shabgah, Navashenaq, Shabgah,

Mohammadi, & Sahebkar, 2017), and symptoms are caused by inflammatory-induced structural (i.e. to the myelin sheath) and functional damage of neurons (Patejdl & Zettl, 2017). This damage occurs practically anywhere in the central nervous system (CNS)(Zettl, Stüve, & Patejdl, 2012) and result in acute or chronic deficits across a variety of neurological domains (Milo & Miller, 2014; Zettl et al., 2012).

Dysarthria (abnormal speech production) is the most common expressive communication deficit presenting in patients with MS (Hartelius, Runmarker, & Andersen, 2000), with prevalence of around 45% (Bauer, Aleric, & Jancic, 2015; Beukelman, Kraft, & Freal, 1985; Danesh-Sani et al., 2013; Darley, Brown, & Goldstein, 1972; Gerald, Murdoch, & Chenery, 1987; Hartelius, Runmarker, & Andersen, 2000; Hartelius & Svensson, 1994; Hartelius, Svensson, & Bubach, 1993; Klugman & Ross, 2002; Konstantopoulos, Vikelis, Seikel, & Mitsikostas, 2010; Piacentini et al., 2014; Yamout et al., 2009; Yorkston et al., 2003). There are some reports of cognitive language disorders such as anomia and aphasia, which share overlapping characteristics with other (non-language related) cognitive deficits and have an even less defined prevalence, however, they are believed to be much less common than dysarthria itself (Beukelman et al., 1985; Gerald et al., 1987; Hartelius, Runmarker, & Andersen, 2000).

Existing evidence suggests that dysarthria in people with MS (PwMS) is typically mild in nature, with patients rarely becoming unintelligible (Beukelman et al., 1985; Stipancic, Tjaden, & Wilding, 2016). Nevertheless, impaired speech is known to have a negative impact on employment status, social participation and overall quality of life in this population (Baylor, Yorkston, Bamer, Britton, & Amtmann, 2010; Beukelman et al., 1985; T. Chiara, D. Martin, & C. Sapienza, 2007; Klugman & Ross, 2002; Mackenzie & Green, 2009; Piacentini et al., 2014).

Emerging evidence strongly support a drastic and ongoing change in the treatment of Multiple Sclerosis (Calvo-Barreiro, Eixarch, Montalban, & Espejo, 2018; Merkel, Butzkueven,

Traboulsee, Havrdova, & Kalincik, 2017; Sharif et al., 2018). Current disease-modifying therapy aims at long periods free of disease activity and is recommended at early stages (Merkel et al., 2017; Ziemssen, Medin, Couto, & Mitchell, 2017). Yet, as presentation and disease progression are highly variable, a tailored approach is considered ideal. Thus, defining biomarkers and surrogate endpoints is the essential next step to enable efficient, tailored therapy in Multiple Sclerosis (Bielekova & Martin, 2004; Merkel et al., 2017; Ziemssen et al., 2017).

Recent refinement of speech assessment methods suggests that discrete speech features might be used as an additional tool to monitor disease status. This review aims to update and expand our current understanding of dysarthria in PwMS, and the use of speech as a marker of changes in disease state. We have included studies utilizing objective and subjective measures as well as the results from neuroimaging studies.

2.3. Search methods

Search and review processes are summarized in Figure .21, and a short glossary of common terminology is provided in Table 2.1. A total of 626 publications were found. Reference lists from selected studies were also examined.

Screening inclusion criteria were 1) human subjects; 2) speech/dysarthria as being the primary outcome of investigation OR related to the primary outcome(s); 3) qualitative and/or quantitative data of dysarthria or speech in the MS population; and 4) definite or probable diagnosis of MS. Studies were excluded after full-text analysis (eligibility) if they described: 1) speech perception by participants and auditory processing data only (i.e., hearing/comprehending by the participant as opposed to output production of speech); 2) animal studies; 3) books or book chapters; 4) non-original data (e.g., consensus, professional opinion, letters); and 5) practical duplicates (same set of participants AND very similar hypothesis/findings as another study already included).

Figure 2.1. PRISMA flow diagram of the reviewing process

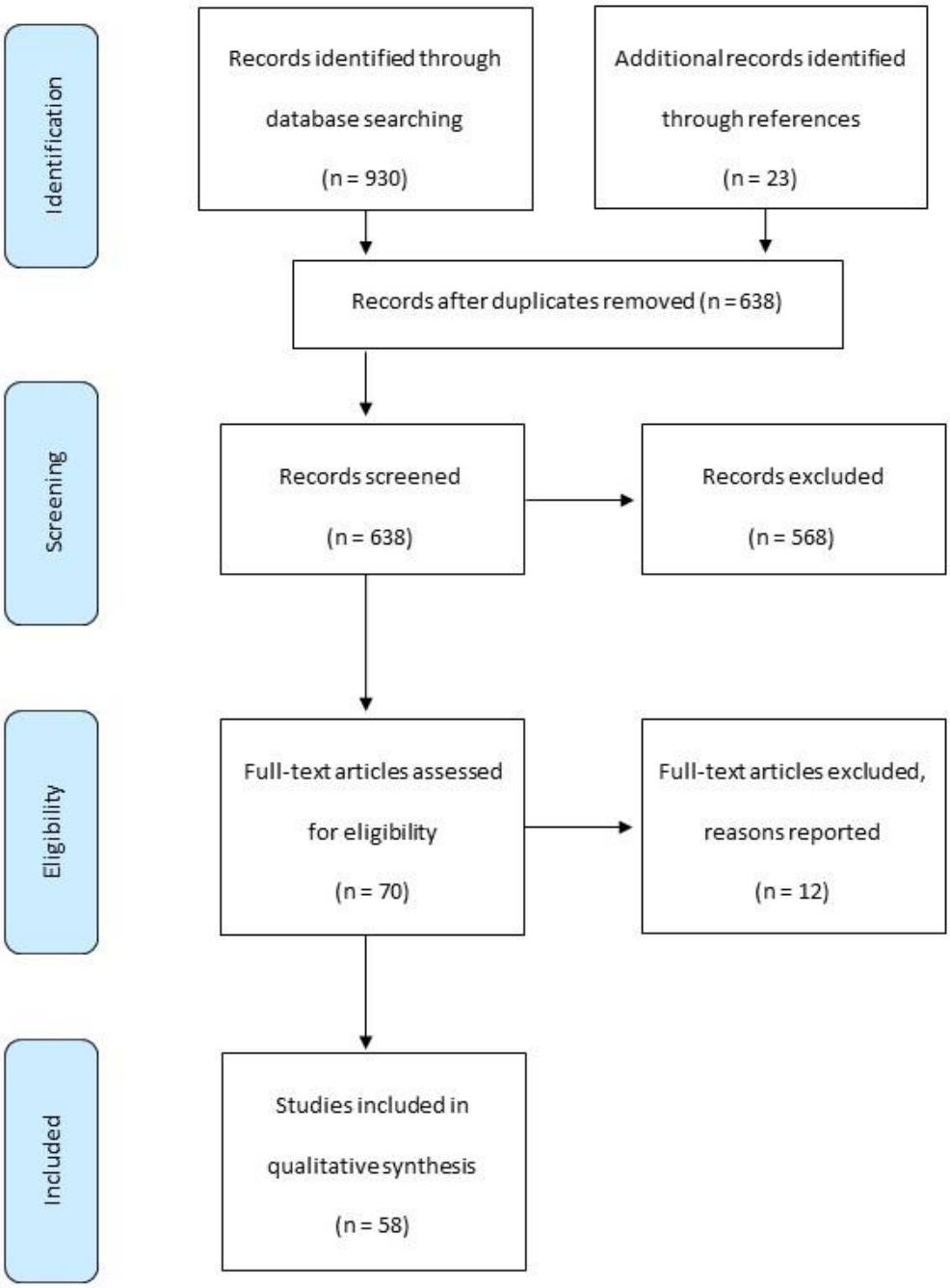


Table 2.1. Glossary of commonly used terms in the assessment of speech deficits

Asthenia, voice	**Perceptual measurement denoting low energy during phonation.**
Articulatory rate	Quantitative measurement from both perceptual and acoustic assessments. Speech (syllables, words) produced per time, excluding silences (in contrast with 'speech rate', which includes silences).
Dysarthria	Impairment of motor control of speech production. The current definition goes beyond its etymological 'disturbed articulation' to also include deficits in respiratory support, phonation and prosody(Darley, Aronson, & Brown, 1969).
Second formant (F2), from acoustic analysis	Formants are frequencies where there is high concentration of sound energy. Formants result from the resonance in the vocal tract (mostly pharynx, mouth and nasal cavities) of the sound produced in the vocal folds. Changes in F2 are associated with changes in the shape of the mouth cavity (e.g., during utterance of the word "you").
Fundamental frequency (f0)	Lowest frequency where sound energy concentrates. Produced by the vibration of the vocal folds in normal voicing. Acoustic analysis measurement frequently associated with perceptual pitch.
Imprecise consonants	Deviation from the expected perceptual sound of one or more consonants, impacting on intelligibility or requiring context to be understood.
Loudness	Perceptual quantification of sound volume in voice/speech.

Table 2.1. Continued

Pitch	**Perceived vocal tone, typically higher in children and women and lower in men.**
Prosody	Collection of interrelated speech features (e.g. rhythm, variation in pitch and loudness) intentionally produced to add to or finely tune the meaning of phrases (frequently adding emotional/social cues).
Strain, voice	Perceived physical effort to phonate/speak, higher than expected for the resulting loudness.
Syllable repetition rate	Number of syllables per time produced during continuous and sequential fast repetition of one, two or three prescribed syllables without semantic meaning.

2.4. Results

Removal of duplicates, application of the selection criteria and additional search for cross-references yielded a total of 68 original studies and one review (see diagram 1). Eleven original studies were then excluded as their data were later updated (i.e. new cases added) or simply duplicated in more recent publications by same authors. The only review was also removed as it reported speech findings from only two original studies, which are already included here. A critical summary of the remaining 58 articles is provided, organized by dysarthria domains and followed by a brief discussion.

2.4.1. Overall dysarthria characteristics

Dysarthria is considered the primary cause of communication deficit in MS, yet pwMS present with concurrent cognitive deficits that can interfere with effective communication. A series of recorded interviews with patients using open-ended questions (such as “What’s

communication like for you?”) (Yorkston et al., 2007; Yorkston, Klasner, & Swanson, 2001) described naming deficits (expressive language), attentional problems (cognition) and fatigue as possible causes of communication difficulties in MS. Although one-third of participants presented with dysarthria, there was little to no mention of dysarthria by patients. In contrast, self-reported prevalence of dysarthria ranges from 23% to 56% in structured questionnaires (Beukelman et al., 1985; Hartelius & Svensson, 1994; Klugman & Ross, 2002; Konstantopoulos et al., 2010; Piacentini et al., 2014; Yamout et al., 2009; Yorkston et al., 2003). A few non-controlled variables may have interfered in the descriptive content of these studies, such as a failure to exclude individuals with cognitive impairment and the lack of standardized questions targeting speech features specifically.

Intelligibility is marginally reduced in MS, with the degree of impairment consistently reported between studies. In a rater blinded study of 78 PwMS, intelligibility for single words was 96±0.03% for PwMS and 97±0.01% for healthy controls. Other work reported similar results with either marginally decreased intelligibility in comparison to normal controls (97% vs 98%, $F[1,71]=8.51$; $p=.005$) (Rodgers, Tjaden, Feenaughty, Weinstock-Guttman, & Benedict, 2013) or a non-significant trend towards lower intelligibility (Sussman & Tjaden, 2012).

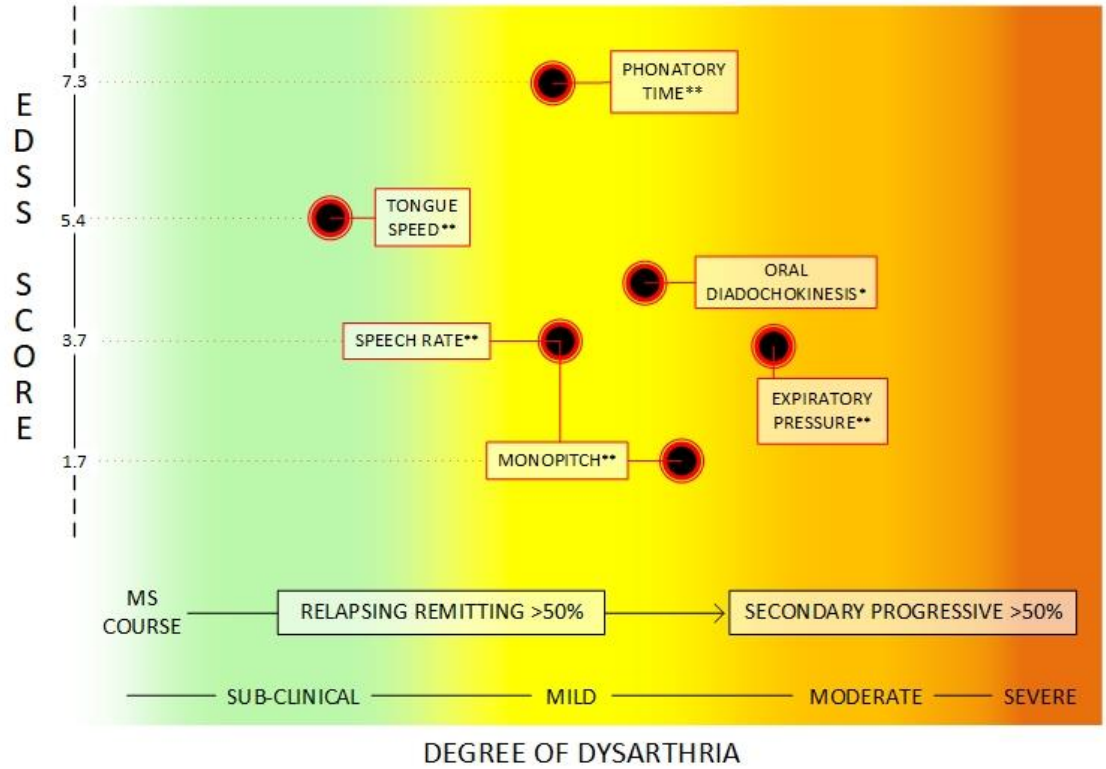
Data derived from subjective listener-based scales have yielded different results. Subjective tests where blinded listeners used a visual analog scale for perceived naturalness demonstrated worse scores than intelligibility assessments for both healthy controls and MS groups (Stipancic et al., 2016; Vitorino, 2009) with the MS group rated considerably worse than healthy controls (3.07±1.32 vs 1.07±0.25, on a scale from 1=normal to 4=severely disturbed) (Vitorino, 2009). This would suggest that the ordinal scale approach either overestimates the magnitude of deficit or is more sensitive to pathology.

In general, the progression of dysarthria parallels progression in other neurological systems in MS. As such, severe dysarthria is usually only present in PwMS with advanced

neurological disability. This observation is supported by the reports by Hartelius et al., where two comprehensive perceptual assessment protocols with multiple individual speech scores were condensed into one composite "overall dysarthria score" for each participant. Both composite dysarthria protocols showed a strong relationship with Expanded Disability Status Scale (Kurtzke, 1983) scores (EDSS, correlation coefficient of 0.6) (Hartelius, Runmarker, & Andersen, 2000).

Specific speech deficits are described below, and most prominent impairments are condensed in Table 2.2. Only six studies reported both EDSS and the general degree of dysarthria, all of which used acoustic assessments (Figure 2.2).

Figure 2.2. Overview of impaired speech characteristics in relation to overall disability (EDSS scores), severity of dysarthria and disease course



*p<0.05; **p<0.001.

Table 2.2. Summary of main speech deficits found in MS.

Deficit	Findings	Type	Studies
Flaccidity of oropharyngeal muscles, particularly the tongue	Long phonemes	PA	(Darley et al., 1972; Gerald et al., 1987; Hartelius, Runmarker, Andersen, & Nord, 2000)
		AA	(Darley et al., 1972; Rodgers et al., 2013; J. Rusz, Benova, et al., 2018; Tjaden & Watling, 2003b)
	Reduced F2 slope	AA	(Hartelius, Nord, & Buder, 1995; K. M. Rosen, Goozée, & Murdoch, 2008)
	Slow, weak tongue, lips and velopharyngeal	PA	(Hartelius & Lillvik, 2003; Vitorino, 2009)
		BM	(Murdoch, Spencer, Theodoros, & Thompson, 1998; Vitorino, 2009)
Glottal insufficiency and instability	Vocal asthenia	PA	(Bauer et al., 2013; Dogan et al., 2007)
	Voice breaks	PA	(Gerald et al., 1987; Hartelius et al., 1993; Yamout et al., 2013; Yamout et al., 2009)
	Reduced closed quotient	EP	(Vavougios, Doskas, & Konstantopoulos, 2018; Yamout et al., 2013)
	Phonatory instability	PA	(Darley et al., 1972; Gerald et al., 1987; Hartelius, Runmarker, & Andersen, 2000; Hartelius et al., 1993)
		AA	(Dogan et al., 2007; Feijó et al., 2004; Hartelius, Buder, & Strand, 1997; Konstantopoulos et al., 2010; Yamout et al., 2009)
	Breathiness, reduced loudness	AA	(Dogan et al., 2007; Tjaden & Watling, 2003b)
Decreased strength of expiratory muscles and/or control over voluntary respiratory movements	Respiratory support	PA	(Darley et al., 1972; Gerald et al., 1987; Hartelius, Runmarker, & Andersen, 2000)
	Expiratory pressure	BM	(T. Chiara et al., 2007)
	Max. expiratory time	BM	(Nordio et al., 2018)
Nonspecific. Might be related impairment of cognition (attention, language), respiratory support, or might reflect motor hesitancy	Frequent, longer and inappropriate pauses	PA	(Gerald et al., 1987; Hartelius, Runmarker, & Andersen, 2000; Hartelius et al., 1993)
		AA	(Hartelius, Runmarker, Andersen, et al., 2000; Kuo & Tjaden, 2016; Tjaden & Watling, 2003b)

Table 2.2. (continued)

Deficit	Findings	Type	Studies
Nonspecific. Early sign. Might be related to depressive mood	Monopitch	PA	(Darley et al., 1972; Gerald et al., 1987; Hartelius, Runmarker, Andersen, et al., 2000; Hartelius, Theodoros, Cahill, & Lillvik, 2003)
		AA	(Konstantopoulos et al., 2010; J. Rusz, Benova, et al., 2018; Vavougios et al., 2018)
Pneumo-phonatory incoordination, related to ataxia	Excessive loudness variation	PA	(Darley et al., 1972; Gerald et al., 1987; Hartelius, Runmarker, Andersen, et al., 2000; Hartelius et al., 2003)
		AA	(J. Rusz, Benova, et al., 2018)

AA= acoustic analysis; BM= biomechanical assessments; EP= electrophysiology; PA= perceptual assessments

2.4.2. Articulation

Articulation of speech was analysed in 576 PwMS across 14 studies. Articulation in MS is characterized by consonant imprecision, decreased word output rate and slow vowel transitions likely due to slow tongue movements. Listener rated scores reported that consonants were mispronounced by 26- 40% of PwMS (Gerald et al., 1987; Hartelius, Runmarker, & Andersen, 2000; Hartelius et al., 2003; Vitorino, 2009) accounting for 90% of intelligibility variance in one study (Gerald et al., 1987).

Articulatory rate was consistently reduced in both subjective and objective studies and correlates with non-speech disabilities. Longer duration of syllables was observed in different tasks, including spontaneous speech, reading and syllable repetition (Darley et al., 1972; Gerald et al., 1987; Hartelius, Runmarker, Andersen, et al., 2000; Rodgers et al., 2013; J. Rusz, Benova, et al., 2018; Tjaden & Watling, 2003b). Additionally, results from acoustic analysis showed an inverse moderate correlation between articulation rate and overall disability (EDSS, $r=0.5$, $p<0.001$), hand dexterity (9 peg hole test, $r=0.54$, $p<0.001$), and walking speed (timed 25-foot walk, $r=0.41$, $p<0.001$)(J. Rusz, Benova, et al., 2018).

Tongue movements are particularly affected in PwMS and can be detected even before overt dysarthria manifests. Studies that utilized acoustic analysis showed, through investigation of the maximum slope of the second formant (F2), that phoneme transitions requiring fast tongue or fast pharyngeal movements were slower in dysarthric PwMS (Hartelius et al., 1995; K. M. Rosen et al., 2008). When studied in isolation, tongue speed, strength and endurance were lower not only in dysarthric but also in non-dysarthric PwMS (Hartelius & Lillvik, 2003; Murdoch et al., 1998), whereas lip and velopharyngeal movements were abnormal in dysarthric but spared in non-dysarthric PwMS (Hartelius & Lillvik, 2003; Murdoch et al., 1998; Vitorino, 2009).

2.4.3. Voice

Voice function in pwMS is characterized by glottal inefficiency (i.e., the sound energy produced in relation to the amount of air passing between the vocal folds), decreased loudness control and both short and long-term (tremor-like) vocal instability.

Voice quality was studied in a total of 484 persons from 14 studies. Perceptual studies were mostly non-blind, without a matched control group and largely descriptive where speech-language pathologists rated PwMS as vocally impaired in 45% to 91% of the cases (Darley et al., 1972; Gerald et al., 1987; Hartelius et al., 1993; Yamout et al., 2009).

A considerable number of studies support the hypothesis that loss of glottal efficiency is a contributor to decreased voice quality in PwMS. Phonatory asthenia and strain were found to be different between MS and healthy control groups in two studies, present in up to a third of PwMS. Additionally, asthenia was strongly correlated with disease duration (r=0.53) and Voice Handicap Index scores (r=0.44)(Bauer et al., 2013; Dogan et al., 2007). No correlation with EDSS scores was found except for reported voice fatigue (Yamout et al., 2009). A quarter of PwMS had frequent voice breaks (Gerald et al., 1987; Hartelius et al., 1993; Yamout et al., 2009), which inversely correlated with Closed Quotients in an electroglottographic study (i.e.,

the proportion of time that the vocal folds sustain full contact during vocalization)(Yamout et al., 2013). The same electroglottographic variables were also used to create a discriminant regression model (equation), resulting in perfect classificatory for the diagnosis of MS, achieving perfect classificatory accuracy (sensitivity and specificity of 100%) for the experiments cohort of 64 PwMS and 64 matched controls (Vavougios et al., 2018). Some less commonly reported findings concerned glottal inefficiency manifesting in reduced mean loudness (Tjaden & Watling, 2003b) and increased breathiness as measured by the soft phonation index (Dogan et al., 2007).

A second contributor to poor voice quality is related to phonatory instability. Pitch (Gerald et al., 1987; Hartelius, Runmarker, & Andersen, 2000; Hartelius et al., 1993) and loudness (Darley et al., 1972; Gerald et al., 1987; Hartelius, Runmarker, & Andersen, 2000; Hartelius et al., 1993) control were rated as impaired in approximately one-third of PwMS in perceptual (subjective) studies. However, only loudness control remained statistically different from healthy controls in a protocol where raters were blinded to the diagnosis (Hartelius, Runmarker, & Andersen, 2000). Data from acoustic analysis suggest that instability of intensity (acoustic equivalent of loudness) and frequency (acoustic equivalent of pitch) are common in PwMS (Feijó et al., 2004; Konstantopoulos et al., 2010). Both jitter and shimmer (which reflect short term variation of frequency and intensity respectively, between two consecutive vocal pulses) were higher in PwMS than in healthy control speakers, most markedly for males (Dogan et al., 2007; Feijó et al., 2004; Konstantopoulos et al., 2010; Yamout et al., 2009). Longer-term “tremor-like” instability (taking into consideration several consecutive pulses, i.e. a few seconds of continuous vocalization) was also reported as a predictor differentiating PwMS from healthy individuals with sensitivity of 80-85% and specificity of 90-100% (Hartelius et al., 1997).

Data on fundamental frequency (acoustic measure related to perceived pitch) appeared inconsistent across studies, with some reporting either lower (Dogan et al., 2007; Konstantopoulos et al., 2010; Yamout et al., 2013) or higher (Feijó et al., 2004) when compared to matched controls.

2.4.4. Respiratory support

Respiratory support was described in a total of 415 PwMS from six studies. Respiration (also cited as ventilation) was impaired in one-third of PwMS in early studies (Darley et al., 1972; Gerald et al., 1987), and expiratory and phonatory times were inversely correlated with EDSS scores in a high disabled cohort of 50 PwMS (average EDSS of 7.3) in the only study to test for non-speech correlations (Nordio et al., 2018). Respiration (rather than either articulation, phonation, oral motor performance, prosody, and intelligibility) was found to be the speech-related domain that best differentiated MS from healthy participants in one study (Hartelius, Runmarker, & Andersen, 2000). Accordingly, maximum phonatory time, maximum expiratory time and maximum expiratory pressure were significantly reduced (T Chiara, D Martin, & C Sapienza, 2007; Dogan et al., 2007; Nordio et al., 2018; Yamout et al., 2009) or showed a trend towards reduction in PwMS (Hartelius et al., 1993) when compared to healthy controls.

2.4.5. Prosody

Prosody was described in at least 470 subjects from ten studies. Speech rate was perceived as reduced (slower) in 39% to 47% of individuals with MS (Darley et al., 1972; Gerald et al., 1987; Hartelius et al., 1993) and confirmed in objective measurements – means of 11% and 24% less syllables per second (Arnett, Smith, Barwick, Benedict, & Ahlstrom, 2008; K. M. Rosen et al., 2008) and around 14% less words per minute (T. Chiara et al., 2007; J. Rusz, Benova, et al., 2018) were produced by PwMS in comparison to matched controls.

Acoustic analysis showed that longer and more frequent pauses were observed both in reading and spontaneous speech tasks (Gerald et al., 1987; Hartelius, Runmarker, & Andersen, 2000; Hartelius, Runmarker, Andersen, et al., 2000; Hartelius et al., 1993; Kuo & Tjaden, 2016; Tjaden & Watling, 2003b). Particularly in ataxic PwMS, deficient "on-the-flight" timing adjustments were apparent. For these people, data showed lack of variation in syllable length within a single utterance – named intra-sentence syllable isochrony – but higher than normal variation between sentences, which could be argued to be a delayed overcorrection resulting in decomposition of rhythm, similar to what occurs for limb movements in ataxia. PwMS with the ataxic type of dysarthria were found to sacrifice the rhythmic pattern of stressing words in order to keep syllable length fixed, while normal speakers kept the rhythm constant by varying syllable length (Hartelius, Runmarker, Andersen, et al., 2000; Tjaden & Watling, 2003b).

While overall speech intonation (i.e. stressing the wrong part of a word or phrase) was found to be impaired in 34 to 43% of PwMS (Darley et al., 1972; Gerald et al., 1987; Hartelius et al., 1993), supporting data suggest that pitch and loudness control may reflect different neuro-networks, thus should be treated separately. Excessive loudness variation was found only in ataxic participants and correlated with hand dexterity, whereas monotonic reading (monopitch) strongly differentiated healthy control and no-disability MS groups (EDSS<2)(Konstantopoulos et al., 2010; J. Rusz, Benova, et al., 2018) but was not associated with hand dexterity or ataxia (J. Rusz, Benova, et al., 2018). Recently, acoustic analysis of pitch variation was included in two prediction models to classify speakers as healthy control or PwMS without disability, reaching accuracies of 78% and 100% (J. Rusz, Benova, et al., 2018; Vavougios et al., 2018).

2.4.6. Dysarthria and MS disease course

Clinically evident dysarthria often presents in more advanced stages of the disease, thus being significantly correlated with overall disease severity as rated in EDSS (Kurtzke,

1983). Accordingly, overt dysarthria is more frequently described in primary and secondary progressive subtypes of MS (Beukelman et al., 1985; Darley et al., 1972; Hartelius, Runmarker, & Andersen, 2000; Yamout et al., 2009)(Figure 2). However, it should be noted that MS-related disability is inconsistently reported in the published literature.

There are also several less common speech presentations reported within the literature. At least forty-nine patients with paroxysmal episodes of dysarthria have been described. The case reports describe acute onset of frank dysarthria lasting only for a few seconds per episode and recurring often and daily, frequently associated with other cerebellar/brain-stem symptoms (e.g., generalized ataxia). They occurred approximately 6-10 weeks after a classic relapsing episode. Six (more recent) cases were investigated at the time of the primary relapse with magnetic resonance imaging (MRI), reporting new enhancing lesions including one or more at or below the red nucleus. Repeat imaging at the time of onset of the paroxysmal dysarthria, however, did not show new demyelinating lesions. Most of these patients were treated with carbamazepine and the symptoms resolved within a few weeks or months, leaving no permanent functional deficit (Blanco, Compta, Graus, & Saiz, 2008; Goodwin & Carpenter, 2016; Iorio, Capone, Plantone, & Batocchi, 2014; Li, Zeng, & Luo, 2011; Marcel et al., 2010; Tüzün, Akman-Demir, & Eraksoy, 2001; Twomey & Espir, 1980; Valentino et al., 2011). Demographic data was reported for 24 of these patients, with 14 being male and overall mean age of 37.5 years.

Additional data describing the relationship between dysarthria and anatomical lesion location in PwMS are very limited. Besides the reports already mentioned, three cases of acute-onset persistent dysarthria were associated with lesions within the brain stem, cerebellum and motor cortex (Kishimoto et al., 2008; Kutzelnigg et al., 2007; Pender & Ferguson, 2007) and two reports of worsening dysarthria following thalamotomy and deep

brain stimulation of the thalamus (Alusi et al., 2001; Pützer, Wokurek, & Moringlane, 2017) for MS-associated upper limb tremor.

2.4.7. Dysarthria and cognition

PwMS are often tested for cognitive deficits, thus some studies have explored the relationship between dysarthria and abstract mental processes. In comparison to PwMS without overt speech impairment, PwMS with dysarthria were shown to have poorer performance in neuropsychological tests (NPT) where speech was required to gauge performance (De Looze et al., 2017; M. M. Smith & Arnett, 2007) (e.g. Paced Auditory Serial Addition Test – PASAT(Gronwall, 1977) and oral version of the Symbol Digit Modality Test – SDMT (A. Smith, 1982)). Moderate correlations have been found between NPT and speech intelligibility (Mackenzie & Green, 2009), speech rate (total output per time) (Arnett et al., 2008; Rodgers et al., 2013) and articulatory rate (i.e. length of phonemes) (Rodgers et al., 2013; J. Rusz, Benova, et al., 2018). Time and frequency of pauses have a stronger correlation than articulatory rate (Roberg, Somogie, Thelen, & Bruce, 2015; Rodgers et al., 2013). Pause time accounted for most of the difference between PwMS and control participants in NPT in one study (Roberg et al., 2015) where articulation speed had no influence. The correlation between motor control of speech and cognition scores was found in MS but not in healthy controls (Rodgers et al., 2013).

Despite those findings, most investigations have a clear confounding bias where speech rate is used at the same time as predictor (in dysarthria scores) and response (in timed oral NPT). Clinicians and researchers must be aware when interpreting oral cognitive tests in dysarthric people and choose an alternative assessment whenever where possible.

2.4.8. Treatment for dysarthria in MS

Resistive respiratory training showed better results than non-resistive exercises. One study tested an integrated protocol of respiratory, phonatory and articulatory exercises in supervised sessions of 45 minutes, four times per week also for eight weeks in 30 PwMS. No effect was observed over maximum phonation time, maximum expiratory time and dysarthria scores during the experiment (Nordio et al., 2018). Another intervention used an intensive protocol similar to limb strength training. Seventeen PwMS and fourteen healthy controls were instructed to blow through a modified simple pressure-controlling device (modified Threshold®PEP, Healthscan Products Inc, Marietta, USA) for five seconds, six consecutive times, resting for 30-60 seconds between sets and completing four sets per day, five days per week (under supervision once per week), for eight weeks. Maximal expiratory pressure increased 40% from baseline in PwMS and 29% in healthy controls. Small but significant improvement in measured reading rate and in self-reported dysarthria was also observed (reported only for the MS group). Results were more evident in the moderate versus mild subgroup (as per EDSS scores) and remained relevant four weeks after the end of the training period. Maximum phonation time was not significantly affected by training (T. Chiara et al., 2007). The lack of a non-training control group prevents exclusion of practising-effect bias and the effects of respiratory training on quality of life, and other speech measures need further exploration.

Behavioural interventions have also been trialled to improve dysarthria in MS. In a study involving 30 PwMS and 32 healthy individuals, participants had their voices individually recorded while reading a standardized passage and then were intensively instructed to speak (in turns) louder, clearer or slower. Improvements in both acoustic characteristics and intelligibility (40, 58-61) were observed. The “loud version” brought the largest increment of consonant distinctiveness and intelligibility. Additionally, a variable enhancement of F2 slopes

was very apparent in half PwMS. None of the behavioural interventions results were reported to last longer than the study session itself.

Successful treatment of phonatory dystonia by injection of botulinum toxin A in the thyroarytenoid muscle has been reported for three PwMS. Unilateral and very low doses (1-3 units) were sufficient to achieve the desired effect of fluid phonation without vocal spasms (Di Stadio, Bernitsas, Restivo, Alfonsi, & Marchese-Ragona, 2018). Little detail of speech, clinical and neuroimaging characteristics of those patients was reported.

Finally, a single case of acute relapse-related dysarthria (among other symptoms) was successfully treated with plasma exchange after failure to improve with corticosteroids (Ikeda, Lee, Fraser, Mirsattari, & Morrow, 2015). Schmidt and colleagues (Schmidt, 2013) reported good results using the potassium channel agonist, Fampridine, in three cases of severe dysarthria. These results were only reported in a brief letter to the editor, with non-blinded assessment, no objective measures of speech, and with declared conflicts of interest.

2.5. Considerations

Different assessment methods influenced the magnitude and generalizability of results. For example, where a speech characteristic was investigated through more than one type of assessment, the frequency and severity of abnormalities were most pronounced when recorded using instrumental measures (e.g., acoustic analysis), followed by professional perceptual ratings (speech pathologists > neurologists) and, lastly, patient self-scoring.

It is clear that greater use of objective tests (i.e. acoustic analysis, electrophysiology and imaging) along with the usual clinical assessments has the potential to address much of the discordance in naming and definitions, as well as the inherent dependency on training and influence from professional background observed in subjective investigations (De Bodt, Wuyts, Van de Heyning, & Croux, 1997).

The relationship between dysarthria and cognitive impairment in PwMS requires further exploration. It is reasonable to assume that any condition that slows speech rate would affect cognitive scores derived from verbal output. Similarly, where individuals present with concomitant cognitive and motor deficits, the competing demands of each process can place stress on each domain, potentially leading to exacerbation of the perceived deficit in a formal testing setting(Vogel, Fletcher, & Maruff, 2014; Vogel, Poole, et al., 2017). This is not unique to MS – people with Myasthenia Gravis (a disease of the neuro-muscular junction) scored below controls in all NPT requiring a verbal response (and in one test that required rapid hand response) but normally in other NPT (Paul, Cohen, Gilchrist, Aloia, & Goldstein, 2000). Furthermore, mean scores in the standard version of the written SDMT (rather than oral) in a large cohort of 811 PwMS, did not report lower than expected scores (Van der Walt et al., 2011).

2.6. Conclusion

Speech production (as a mean of transmitting information) is impaired in PwMS but is usually only clinically evident in more advanced stages of the disease. Although intelligible, speech in PwMS is often perceived as deficient by both the general population and specialists, having a negative impact on communicative participation and quality of life. The main dysarthric features in PwMS are slowness, increase of pauses (frequency, duration and inappropriate onset), deficient loudness control, monopitch, imprecise consonants, asthenic/strained voice and decreased respiratory capacity (Table 2.2). Most speech variables were studied in isolation from one another, from other disease characteristics (i.e., ambulation, cerebellar dysfunction, disease phenotype), and from progression and neuroimaging correlates.

Objective speech assessments offer greater accuracy, replicability, and feasibility in comparison to perceptual analysis. If coupled with additional meaningful outcomes such as measures of speech-related quality of life, objective assessments have the potential to assist decision making when tracking disease progression and treatment response in MS. Longitudinal studies are needed to define whether dysarthria and its measurement provide additional and unique insights into MS disease progression or as a subclinical surrogate marker of cerebellar network involvement.

2.7. Take-home messages

- Mild dysarthria is highly prevalent in MS and significantly impacts quality of life.
- Slow, imprecise, and monotonic speech are common findings in people with MS and can be associated with other neurological deficits. ·
- Objective speech assessments show high classification accuracy for early-MS versus no-MS in experimental cohorts.
- The potential to monitor disease progression has been demonstrated through correlation of objective speech measurements with clinical scores of accrued disabilities. Longitudinal studies are warranted.

Chapter 3

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

Published (online) in the European Journal of Neurology on 11 of September 2020.

DOI: 10.1111/ene.14523

AUTHORS

GUSTAVO NOFFS[1,2], FREDERIQUE M. C. BOONSTRA[3], THUSHARA PERERA[4,5], HELMUT BUTZKUEVEN[3], SCOTT C. KOLBE[3] FERNANDA MALDONADO[1], L. EDUARDO COFRÉ LIZAMA[6,7,8], 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[1,4,9,10,*]

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. School of Allied Health, Human Services and Sports, La Trobe University, Australia

9. Department of Neurodegeneration, Hertie Institute for Clinical Brain Research, University of Tübingen, Germany

10. Redenlab, Victoria, Australia

3.1. Illustrated highlights

- In multiple sclerosis, computer-based speech metrics correlate with neurological disability in people without dysarthria.
- Acoustic analysis of speech differentiated between groups of mild, moderate and severe neurological disability.
- Acoustic analysis of speech correlated with quality of life, white matter atrophy and whole brain lesion load.

3.2. 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.

3.3. Introduction

Dysarthria (impairment of speech execution) affects around 50% of people with multiple sclerosis (MS) at some stage of their illness (Hartelius, Runmarker, & Andersen, 2000; O'Connor, 2002). 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 (Noffs et al., 2018). The degree of impairment is typically mild, and intelligibility is marginally affected (Noffs et al., 2018).

Subtle impairment in speech production might not translate into overt dysarthria. Speech production relies not only on motor control but also affective (mood)(Cummins, Sethu, Epps, Schnieder, & Krajewski, 2015; Flint, Black, Campbell-Taylor, Gailey, & Levinton, 1992; Mundt, Vogel, Feltner, & Lenderking, 2012) and cognitive (Arnett et al., 2008; Rodgers et al., 2013) processes of the central nervous system. Yet, the reported prevalence of cerebellar, cognitive and affective symptoms (Kister et al., 2013; Martin et al., 2006; Panou, Mastorodemos, Papadaki, Simos, & Plaitakis, 2012) 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 (Fletcher & Munson, 1933). We can further perceive instability in loudness, often defined as voice tremor (Maryn, Leblans, Zarowski, & Barkmeier-Kraemer, 2019). Yet, acoustic measurement of *sound intensity* instability can separate individuals with MS from healthy controls (HC) with 90% accuracy (Hartelius et al., 1997) while experienced raters perceive such instability in only about 35% of pwMS (Hartelius, Runmarker, & Andersen, 2000). In a recent study (J. Rusz et al., 2019), 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 (Chan et al., 2019). 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 (Fazeli et al., 2020; Noffs et al., 2020; J. Rusz, Benova, et al., 2018) or brain volumetrics (Noffs et al., 2020; J. Rusz et al., 2019) 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.

3.4. Methods

3.4.1. Participants

People with a definite diagnosis of relapsing-remitting or secondary-progressive MS (Thompson et al., 2018) 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.

3.4.2. 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)(Kurtzke, 1983).

3.4.3. 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 (Noffs et al., 2018). Speech variables were organized into three major domains (Table 3.1) and assessed through both perceptual and acoustic methods.

Table 3.1. Speech items included in the analyses.

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) Loudness instability (energy CoV)	VOWEL
		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

Table 3.1. Continued

Multi-domain	Naturalness Intelligibility	Composite score	All tasks

Mathematical definitions of non-standard acoustic measures are shown between brackets.

CoV = coefficient of variance (i.e., standard deviation divided by the mean); f0 = fundamental frequency; F2 = second formant; SD = standard deviation.

Speech sample

We elicited speech using five standardized speech tasks which fit along a spectrum of automaticity (Adam P Vogel et al., 2014), 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). See Appendix A. Each task, except for the monologue, were elicited twice to reduce the effect of unfamiliarity (Vogel & Maruff, 2014). 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, 20Hz-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 (Boersma, 2002) 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.

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 (Vogel, Poole, et al., 2017) (Table 3.1 and Appendix B). 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.

Acoustic analysis

We selected acoustic characteristics that were previously shown to be altered in MS (e.g. fundamental frequency CoV (Hartelius et al., 1997), F2 slope (K. M. Rosen et al., 2008)) 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 (Noffs et al., 2018) 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 (Tjaden & Watling, 2003b). Similarly, the analysis of sustained vowels was confined to the middle 3 seconds for stability and consistency across participants (Ray D Kent, Vorperian, & Duffy, 1999).

Voicesauce (Y. L. Shue, Keating, & Vicenik, 2009) 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)(Wang, Kent, Duffy, & Thomas, 2009) to analyse DDK rate and variability; previously developed (Folker et al., 2012; Vogel, Poole, et al., 2017) 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 (Vogel, Maruff, Snyder, & Mundt, 2009) while the analysis window length was set to 30ms and 42ms. Window shift was fixed at 10ms.

3.4.4. Self-assessed quality of life

MS participants completed the Multiple Sclerosis Impact Scale (MSIS-29)(Hobart, Lamping, Fitzpatrick, Riazi, & Thompson, 2001), a questionnaire validated to assess the impact of MS on quality of life.

3.4.5. 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. (Boonstra et al., 2018).

3.4.6. 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)(Kurtzke, 1983); and 2) groups that represented different “phases” of disease progression (Amato & Ponziani, 1999; Hohol, Orav, & Weiner, 1999; Meyer-Moock et al., 2014; Ravnborg et al., 2005). 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% (Benjamini & Yekutieli, 2001) 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(Wasserstein & Lazar, 2016), 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 (Sussman & Tjaden, 2012). 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.

3.4.7. 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 are part of ongoing investigations by the researchers. The authors agree to share the study protocol and further details of statistical methods upon request.

3.5. Results

Participants included 119 pwMS and 22 HC. Sixty-eight MS participants agreed to undergo the MRIs protocol. Demographics are shown in Table 3.2.

3.5.1. 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 material Table 3.6 and Table 3.7).

Table 3.2. Descriptive statistics for groups and subgroups.

	HC	MS		MS subgroups		
				MILD	MODERATE	SEVERE
			MRI	(EDSS≤2.5)	(3≤EDSS≤5.5)	(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

*Mean±standard deviation for age, disease duration and MRI metrics, and median±interquartile range for EDSS and MSIS-29. EDSS = Expanded Disability Status Scale; HC = healthy controls; MRI = magnetic resonance imaging participants; MS = Multiple Sclerosis; *n=14.*

3.5.2. 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 3.3 and 3.4).

Table 3.3. Correlation coefficients between perceptual speech measures, disability scores and quality of life scores.

Speech domain	Speech item and method	EDSS		MSIS-29	
		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

Table 3.3. Continued

	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

EDSS = Expanded Disability status Scale; False discovery rate (FDR) adjusted critical p value = 0.008. MSIS-29 = Multiple Sclerosis Impact Scale (quality of life measure); p = p value for the given coefficient; r =correlation coefficient.

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.

3.5.3. Relationship between speech and structural MRI

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 material Table 3.8), and no correlation was observed between speech measurements and brain volumes for people without dysarthria.

Table 3.4. Correlation coefficients between acoustic speech measures, disability scores and quality of life scores.

Speech domain	Speech item and method	EDSS		MSIS-29	
		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

Table 3.4. Continued

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

EDSS = Expanded Disability status Scale; False discovery rate (FDR) adjusted critical p value = 0.008. MSIS-29 = Multiple Sclerosis Impact Scale (quality of life measure); p = p value for the given coefficient; r =correlation coefficient.

3.5.4. Relationship between speech and quality of life

MSIS-29 scores correlated moderately with perceptual prolonged intervals (Table 3.3), and with acoustic intensity decay and frequency instability (Table 3.4). 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.

3.5.5. Acoustic composite score

Multiple linear regression resulted in a model that included pause percentage FREE, speech rate READ and frequency instability VOWEL (Table 3.5), termed 'acoustic composite score', which accounted for 35% of EDSS variation (adjusted r^2=0.347, standard error of the estimate=1.8).

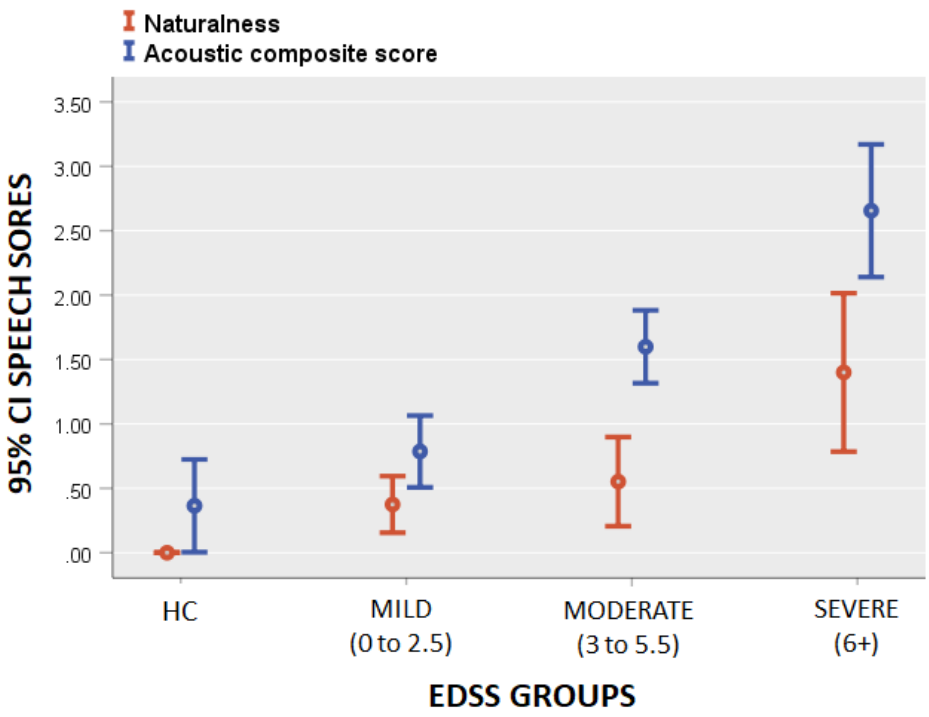
The acoustic composite score correlated with MSIS-29 (r=0.5, p<0.001), white matter volume (r=0.32, p=0.007) and total lesion load (r=0.34, p=0.008). Considering only people without dysarthria, the acoustic composite score correlated with EDSS (r=0.45, p<0.001) and with MSIS-29 (r=0.4, p=0.01). The acoustic composite score differentiated between the mild and moderate MS groups (p<0.001) and between moderate and severe MS groups (p=0.003, supplementary material Table 3.6 and Figure 1).

Table 3.5. Variables in the regression model for EDSS.

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

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

Figure 3.1. Confidence intervals for global speech scores, naturalness, and acoustic composite score across subgroups.



3.6. 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 (Hartelius, Runmarker, & Andersen, 2000; O'Connor, 2002) (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 (Hartelius, Runmarker, & Andersen, 2000)). In agreement with previous studies (Noffs et al., 2018; J. Rusz et al., 2019) 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. (J. Rusz, Benova, et al., 2018), 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 (Konstantopoulos et al., 2010; J. Rusz, Benova, et al., 2018; Vavougios et al., 2018). 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 (Cannizzaro, Harel, Reilly, Chappell, & Snyder, 2004) 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)(J. Rusz et al., 2019). 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 (J. Rusz et al., 2019; Vogel, Fletcher, Snyder, Fredrickson, & Maruff, 2011) 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 are 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.

3.7. 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.

3.8. Acknowledgements

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

3.9. 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).

3.10. 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.

3.11. 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.

3.12. Supplementary material

Table 3.6. Group comparisons for perceptual analysis of speech.

Perceptual analysis			Group comparison vs HC* Median score, inter-quartile range (p value)					Between MS subgroups p value	
Speech domain	Speech variable	Effect direction	HC	MS	EDSS mild	EDSS moderate	EDSS severe	MILD vs MODERATE	MODERATE vs SEVERE
Multiple	Naturalness		0, 0	0, 1 (0.001)	0, 1 (0.004)	0, 1 (0.003)	1, 2.75 (<0.001)	0.637	0.015
	Intelligibility	↓	0, 0	0, 0 (0.119)	0, 0 (0.407)	0, 0 (0.124)	0, 1 (0.006)	0.246	0.094
Timing	Prosody rate	↓	0, 0.25	0, 0 (0.046)	0, 1 (0.958)	0, 1 (0.065)	-1, -2 (<0.001)	0.067	0.002

Table 3.6. Continued

	DDK speed	↓	1, 1	0, 1 (0.007)	0, 1 (0.265)	0, 1 (0.023)	0, -1 (<0.001)	0.052	0.015
	Prolonged intervals	↑	0, 0	0, 1 (0.010)	0, 0 (0.237)	0, 1 (0.004)	1, 1 (<0.001)	0.016	0.055
	Variable rate	↑	0, 0	0, 0 (0.131)	0, 0 (0.675)	0, 1 (0.098)	1, 1 (0.001)	0.155	0.034
Voice quality	Nasality	↑	1, 1	1, 2 (0.223)	1, 1 (0.309)	1, 1.5 (0.135)	2, 1 (<0.001)	0.619	0.017
Control	Imprecise consonants	↑	0, 0	0, 0 (0.033)	0, 0 (0.407)	0, 0 (0.025)	0, 1.75 (0.001)	0.036	0.066
	Monopitch	↑	0, 0	0, 1 (0.119)	0, 0.75 (0.466)	0, 1 (0.100)	1, 2 (0.001)	0.325	0.037

Table 3.6. Continued

Monoloudness	↑	0, 0	0, 0 (0.096)	0, 0 (0.485)	0, 0.5 (0.054)	0, 1 (0.005)	0.139	0.265
Pitch breaks	↑	0, 0	0, 0 (0.182)	0, 0 (0.407)	0, 0 (1.000)	0, 1 (0.006)	0.341	0.002
Loudness decay	↑	0, 0	0, 0 (0.104)	0, 0 (0.143)	0, 0 (0.124)	0, 1 (0.006)	0.900	0.070
Distorted vowels	↑	0, 0	0, 0 (0.104)	0, 0 (0.407)	0, 0 (0.124)	0, 0.75 (0.014)	0.239	0.178
Voice tremor	↑	1, 0	1, 0 (0.689)	1, 0 (1.000)	1, 0 (0.316)	1, 1 (0.020)	0.474	0.130
DDK irregularity	↑	0, 1	0, 1 (0.938)	0, 0.75 (0.581)	0, 1 (0.935)	1, 2 (0.024)	0.611	0.012
critical p values level for FDR of 5%			0.007	0.001	0.005	0.024	0.0002	0.003

Table 3.6. Continued

Timing	Prolonged phonemes	-	0, 0	0, 0 (0.137)	0, 0 (0.407)	0, 0 (0.072)	0, 0 (0.030)	0.132	0.526
	False starts	-	0, 0	0, 0 (0.398)	0,0 (0.228)	0, 0 (0.843)	0, 0 (0.946)	0.294	0.789
Control	Audible inspiration	-	0, 1	0, 0 (0.812)	0, 0.75 (0.963)	0, 1 (0.864)	0, 1 (0.531)	0.822	0.641
	Reduced stress	-	0, 0	0, 0 (0.383)	0, 0 (1.000)	0, 0 (0.213)	0, 0 (0.294)	0.134	0.788
	Irregular articulatory breakdowns	-	0, 0	0, 0 (0.383)	0, 0 (1.000)	0, 0 (0.213)	0, 0 (0.133)	0.134	0.700
	Groping	-	0, 0	0, 0 (0.666)	0, 0 (1.000)	0, 0 (1.000)	0, 0 (0.294)	1.000	0.229

Table 3.6. Continued

	Phonemic errors	-	0, 0	0, 0 (0.666)	0, 0 (1.000)	0, 0 (0.384)	0, 0 (1.000)	0.294	0.406
Voice quality	Roughness	-	1, 0.25	1, 1 (0.629)	1, 1 (0.316)	1, 0.5 (0.855)	1, 1 (0.103)	0.366	0.049
	Breathiness	-	0, 1	0, 1 (0.216)	0, 1 (0.635)	1, 1 (0.187)	1, 1 (0.193)	0.362	0.991
	Strain	-	0, 0.25	0, 1 (0.537)	0, 0 (0.724)	0, 1 (0.166)	0, 1 (0.205)	0.055	0.962

** Mann-Whitney test; FDR = False Discovery Rate; HC = healthy controls; MS = Multiple Sclerosis.*

Table 3.7. Group comparisons for acoustic analysis measurements.

				Group comparison vs HC* Mean measurement ± standard deviation (p value)					Between MS subgroups* p value	
Acoustic analysis										
Speech domain	Speech variable	Task	Effect direction	HC	MS	EDSS mild	EDSS moderate	EDSS severe	MILD vs MODERATE	MODERATE vs SEVERE
Multiple	Acoustic composite score		↑	0.36±0.7	1.50±1.26 (<0.001)	0.79±0.7 (0.059)	1.6±0.72 (<0.001)	2.7±1 (<0.001)	<0.001	0.003
Control	Frequency variability (%)	READ	↓	19.3±47.6	14.9± 4.8 (<0.001)	14.4±4.1 (<0.001)	14±5.7 (0.003)	15.5±5.8 (0.010)	0.778	0.309

Table 3.7. Continued

Timing	Pause-length variability (ms)	READ	↑	172±70	249±198 (<0.001)	211±67 (0.004)	218±73 (0.005)	375±414 (<0.001)	0.684	0.149
	Pause percentage	FREE	↑	19.1±5.8	25.7±8.8 (0.001)	22.1±6.8 (0.098)	26.8±5.5 (<0.001)	34.1±11 (<0.001)	0.005	0.011
	Pause percentage	READ	↑	22.9±4.6	27.4±9.3 (0.010)	25.5±5.8 (0.082)	27.7±5.8 (0.003)	33.3±14.1 (0.002)	0.171	0.16
	Syllable period SD (ms)	DDK	↑	35.7±20.8	49.6±27.6 (0.023)	44.7±22 (0.118)	42.3±22.7 (0.369)	66.7±37.8 (0.004)	0.693	0.018
	Syllables/second	DDK	↓	6.2±1.1	5.5±0.9 (0.004)	5.8±0.8 (0.123)	5.6±0.9 (0.340)	4.8±1 (<0.001)	0.371	0.008
	Words/second	READ	↓	3.9±0.3	3.5±0.7 (0.030)	3.7±0.5 (0.192)	3.5±0.5 (0.009)	3.1±1.1 (0.007)	0.165	0.122
critical p values level for FDR of 5%					0.007	0.001	0.005	0.024	0.0002	0.003

Table 3.7. Continued

Timing	Pause-length variability (ms)	FREE	-	172±63	220±128 (0.089)	173±75 (0.866)	217±90 (0.055)	327±215 (0.031)	0.033	0.038
	Words/second	DAYS	-	3.9±1	3.8±1 (0.789)	4.1±1 (0.665)	3.7±1 (0.390)	3.6±0.9 (0.225)	0.155	0.662
	Articulatory time (s)	READ	-	21.9±2.2	23.5±6.5 (0.302)	22.1±3.3 (0.730)	22.8±3.4 (0.634)	25.9±9.7 (0.111)	0.647	0.117
	Articulatory time (s)	DAYS	-	3.5±1	3.5 ±0.7 (0.873)	3.3±0.9 (0.560)	23.6±0.8 (0.592)	3.9±1.1 (0.200)	0.191	0.291
Control	Frequency instability (%)	VOWEL	-	0.76±0.43	1.07±0.8 (0.087)	0.89±0.42 (0.546)	1.1±0.4 (0.019)	1.17±0.45 (0.110)	0.068	0.570
	Loudness instability (%)	VOWEL	-	14.5±4.4	17.4±9.7 (0.497)	15.6±6.1 (0.496)	18±10 (0.454)	20.5±14.3 (0.612)	0.593	0.896

Table 3.7. Continued

	F2 articulation (%)	READ	-	62.9±5.9	61.3±11.6 (0.291)	62.5±11.3 (0.846)	60.5±8.6 (0.269)	57.9±13.5 (0.136)	0.461	0.405
	Loudness decay	FREE	-		0.256	0.148	0.848	0.979	0.123	0.384
	Loudness variability	FREE	-	102±31.2	105±44.3 (0.925)	112±63.4 (0.813)	105±39.5 (0.992)	108±33 (0.361)	0.729	0.452
	Frequency variability	FREE	-	15.9±4.2	15±5.1 (0.062)	15.9±5.2 (0.939)	14.4±4.4 (0.026)	16±6 (0.3900	0.042	0.226
Voice quality	HNR	VOWEL	-	47.2±11.6	45.9±8.9 (0.533)	46.6±9.4 (0.827)	45.3±6.8 (0.473)	44.4±6 (0.110)	0.513	0.651
	CPP	VOWEL	-	24.7±3.6	24.3±3.7 (0.922)	24.7±3.2 (0.689)	23.8±3.8 (0.620)	25±4 (0.574)	0.522	0.301

** Independent samples t-test for normally distributed measurements and Mann-Whitney test for non-normally distributed measurements; CPP = cepstral peak prominence; DAYS = say the days of the week; DDK = speech oral diadochokinesis (pa-ta-ka task); FDR = False Discovery Rate; FREE = story-telling task; HC = healthy controls; HNR = harmonic to noise ratio; MS = Multiple Sclerosis; READ = reading a paragraph; VOWEL = holding a long /a/ utterance.*

Table 3.8. Correlation coefficients between speech and imaging measures.

Speech domain	Speech item and method		Brain volume		Gray matter volume		White matter volume		Total lesion load	
			r	p	r	p	r	p	r	p
Multi-domain	Naturalness	P	-0.20	0.082	-0.17	0.131	-0.22	0.047	0.14	0.270
	Intelligibility	P	-0.20	0.07	-0.14	0.205	-0.23	0.040	0.21	0.092
	Composite score	A	-0.28	0.019	-0.27	0.024	-0.32	0.007	0.34	0.008
Timing	Prosody rate	P	0.26	0.021	0.21	0.066	0.32	0.004	-0.34	0.005
	DDK speed	P	0.03	0.778	0.08	0.466	0.03	0.775	-0.25	0.044
	Prolonged intervals	P	-0.25	0.023	-0.23	0.039	-0.24	0.032	0.21	0.097
	Variable rate	P	-0.29	0.008	-0.22	0.046	-0.33	0.003	0.26	0.037
	DDK irregularity	P	-0.08	0.463	-0.02	0.830	-0.11	0.333	0.13	0.292
	Prolonged phonemes	P	-0.15	0.178	-0.14	0.223	-0.13	0.260	0.05	0.691
	False starts	P	-0.09	0.426	-0.13	0.257	0.05	0.669	-0.08	0.550

Table 3.8. Continued

	Pause percentage, FREE	A	-0.23	0.043	-0.21	0.065	-0.27	0.020	0.19	0.137
	Speech rate, READ	A	0.15	0.21	0.12	0.307	0.14	0.229	-0.23	0.078
	Speech rate, DDK	A	0.04	0.743	0.04	0.749	0.10	0.404	-0.22	0.073
	Pause variability, FREE	A	-0.16	0.164	-0.20	0.086	-0.13	0.264	0.18	0.165
	Pause variability, READ	A	-0.22	0.051	-0.19	0.098	-0.12	0.314	-0.04	0.763
	Speech rate, DAYS	A	0.07	0.555	0.03	0.778	0.12	0.312	0.05	0.682
	Articulatory rate, DAYS	A	-0.11	0.35	-0.08	0.481	-0.14	0.222	-0.05	0.702
	Arrticulatory rate, READ	A	-0.06	0.634	-0.04	0.713	-0.10	0.384	0.25	0.049
	Pause variability, READ	A	-0.21	0.074	-0.19	0.106	-0.18	0.131	0.06	0.671
	Articulatory variability, DDK	A	-0.03	0.814	-0.04	0.725	-0.05	0.663	0.12	0.364
Control	Imprecise consonants	P	-0.05	0.648	0.00	0.985	-0.15	0.197	0.05	0.696
	Distorted vowels	P	-0.21	0.066	-0.15	0.171	-0.21	0.066	0.12	0.343

Table 3.8. Continued

	Voice tremor	P	-0.06	0.611	-0.10	0.375	0.01	0.917	0.09	0.500
	Loudness decay	P	0.02	0.884	0.01	0.902	0.02	0.881	-0.04	0.759
	Monopitch	P	0.06	0.625	0.05	0.649	-0.03	0.784	0.13	0.297
	Monoloudness	P	-0.28	0.013	-0.23	0.039	-0.27	0.017	0.10	0.414
	Pitch breaks	P	-0.12	0.292	-0.08	0.483	-0.14	0.202	0.11	0.394
	Audible inspiration	P	0.15	0.177	0.10	0.400	0.18	0.114	-0.23	0.069
	Reduced stress	P	-0.10	0.357	-0.14	0.227	-0.06	0.615	0.20	0.107
	Irregular art. breakdown	P	-0.04	0.737	0.04	0.760	-0.13	0.270	0.01	0.971
	Phonemic errors	P	-0.11	0.312	-0.07	0.533	-0.08	0.478	-0.08	0.550
	Nasality	P	-0.19	0.087	-0.19	0.087	-0.17	0.125	0.13	0.285
	Frequency instability	A	-0.30	0.008	-0.29	0.010	-0.24	0.036	0.22	0.073
	Vowel articulation, READ	A	0.01	0.925	0.01	0.962	0.07	0.564	0.14	0.267
	Energy decay	A	-0.01	0.954	-0.04	0.747	0.04	0.756	0.05	0.706
	Energy variability	A	-0.30	0.007	-0.34	0.002	-0.26	0.021	0.10	0.448

Table 3.8. Continued

	Energy instability	A	-0.02	0.878	-0.01	0.923	-0.06	0.619	0.03	0.814
	Frequency varibility, READ	A	-0.04	0.745	-0.05	0.660	-0.06	0.627	0.13	0.316
	Frequency variability, FREE	A	0.11	0.32	0.09	0.427	0.07	0.537	0.00	0.976
Voice quality	Roughness	P	-0.24	0.031	-0.28	0.013	-0.14	0.205	0.09	0.456
	Strain	P	-0.11	0.318	-0.17	0.139	-0.03	0.796	-0.01	0.916
	Breathiness	P	-0.07	0.517	-0.06	0.581	-0.07	0.516	0.00	0.985
	Cepstral Peak Prominence	A	0.06	0.572	0.04	0.733	0.04	0.723	-0.02	0.885
	Harmonics-to-noise ratio	A	-0.01	0.905	0.00	0.999	-0.02	0.829	0.07	0.566

Critical p = 0.008. A = acoustic measurement; P = perceptual; p = statistical significance; r =correlation coefficient.

Chapter 4

Acoustic speech analytics are predictive of cerebellar dysfunction in Multiple Sclerosis

Published in The Cerebellum (journal), volume 19, issue 5, pages 691-700, on 18 of June 2020.

DOI: 10.1007/s12311-020-01151-5

AUTHORS

GUSTAVO NOFFS, MD[1,2] FREDERIQUE M. C. BOONSTRA, MSC[3] THUSHARA PERERA, PHD[4,5] SCOTT C. KOLBE, PHD[3,6] JIM STANKOVICH[3] HELMUT BUTZKUEVEN, PHD[3] ANDREW EVANS, MD[2,4] ADAM P. VOGEL, PHD[1,4,7,8*] AND ANNEKE VAN DER WALT, PHD[2,3,4*]

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. Florey Institute of Neuroscience and Mental Health, Australia
7. Department of Neurodegeneration, Hertie Institute for Clinical Brain Research, University of Tübingen, Germany
8. Redenlab Inc, USA

* Joint senior authors

MeSH descriptors: Multiple Sclerosis, disability evaluation, speech acoustics, magnetic resonance imaging, quality of life, biomarkers

4.1. Abstract

BACKGROUND: Speech production relies on motor control and cognitive processing and is linked to cerebellar function. In diseases where the cerebellum is impaired, such as multiple sclerosis (MS), speech abnormalities are common and can be detected by instrumental assessments. However, the potential of speech assessments to be used to monitor cerebellar impairment in MS remains unexplored. AIM: To build an objectively measured speech score that reflects cerebellar function, pathology, and quality of life in MS. METHODS: 85 people with MS and 21 controls participated in the study. Speech was independently assessed through objective acoustic analysis and blind expert listener ratings. Cerebellar function and overall disease disability were measured through validated clinical scores; cerebellar pathology was assessed via magnetic resonance imaging; and validated questionnaires informed quality of life. Selected speech variables were entered in a regression model to predict cerebellar function. The resulting model was condensed into one composite speech score and tested for prediction of abnormal 9-hole peg test (9HPT), and for correlations with the remaining cerebellar scores, imaging measurements and self-assessed quality of life. RESULTS: Slow rate of syllable repetition and increased free speech pause percentage were the strongest predictors of cerebellar impairment, complemented by phonatory instability. Those variables formed the acoustic composite score that accounted for 54% of variation in cerebellar function, correlated with cerebellar white matter volume (r=0.3, p=0.017), quality of life (r=0.5, p<0.001) and predicted an abnormal 9HPT with 85% accuracy. CONCLUSION: An objective multi-feature speech metric was highly representative of motor cerebellar impairment in MS.

4.2. Introduction

Cerebellar dysfunction is common in Multiple Sclerosis (MS), affects quality of life(Klevan et al., 2014) and predicts prognosis (Kalincik et al., 2014). Cerebellar function extends beyond the motor domain, with evidence suggesting prominent additional roles in cognitive-affective processes (Schmahmann, Guell, Stoodley, & Halko, 2019). Given the importance of cerebellar dysfunction in MS, effective assessment tools of cerebellar impairment are essential for clinical practice and clinical trials. Additionally, objectiveness in cerebellar assessment is particularly desirable for clinical trials. Nevertheless, clinical assessments of cerebellar impairment, such as the cerebellar functional sub-score of the Expanded Disability Status Scale (EDSS)(Kurtzke, 1983) or the Scale for the Rating and Assessment of Ataxia (SARA),(T Schmitz-Hübsch et al., 2006) are subjective and rely almost exclusively on motor symptoms to gauge cerebellar dysfunction. The nine-hole peg test, albeit being objective, has a relatively very low cognitive demand. Alternatively, magnetic resonance imaging (MRI) can provide objective serial monitoring of cerebellar volume(Jacobsen et al., 2014) but is slow to change, difficult to implement in clinical practice and only partially reflects cerebellar function (Ilg et al., 2018).

Speech is intrinsically linked to the motor control and cognitive functions of the cerebellum. The motor role of the cerebellum in speech production is demonstrated, firstly, by its timing control of speech in healthy participants during functional MRI (Riecker et al., 2005) and secondly, through speech deficits in people with cerebellar diseases (Folker et al., 2012; Vogel, Rommel, et al., 2017; Vogel et al., 2018; Vogel, Wardrop, et al., 2017). Additionally, evidence points to the cognitive role of the cerebellum in speech, with 'internal speech' being considered a cerebellar process involving both verbal memory and pre-motor control of speech production (Ackermann, Mathiak, & Ivry, 2004; Durisko & Fiez, 2010; Ravizza et al., 2006; Valentino et al., 2009). In MS, the relationship between motor speech performance

(mostly speech rate) and cognitive scores is well described (Arnett et al., 2008; Rodgers et al., 2013), but the contribution of cerebellar injury to those associations remains unexplored.

As speech production is reliant on both cognitive/premotor and motor roles of the cerebellum, speech analysis could offer a useful assessment of MS-associated cerebellar injury. It could be integrated into multi-domain tool to be used in clinical trials, clinical practice, and patient self-monitoring. The primary aim of this study was to build and test a speech model that represented cerebellar motor impairment. Our secondary aim was to estimate the contribution of cerebellar impairment to individual speech deficits.

4.3. Methods

4.3.1. Participants

People with a definite diagnosis of MS (Thompson et al., 2018) of disease course type relapsing-remitting or secondary progressive were prospectively and sequentially recruited from tertiary MS clinics. Exclusion criteria included: 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). The study was approved by the responsible Ethics Committee, and written informed consent was obtained from all participants.

4.3.2. Neurological assessment

All participants with MS underwent a detailed physical and neurological examination by a qualified neurologist who determined their overall disease severity using EDSS. Measures of cerebellar performance included 1) the cerebellar functional sub-score of the EDSS (FSS-cerebellar); 2) the overall score from SARA, which is the sum of eight sub-scores that assess deficits commonly observed in cerebellar diseases such as distal and proximal balance, gait deficits, dysarthria, upper limb tremor and dysmetria, and upper and lower limb movement

decomposition; and 3) shortest of two trials of time to complete the nine-hole peg test (9HPT)(Mathiowetz et al., 1985) with the dominant hand.

4.3.3. MRI data acquisition and analysis

Structural imaging included: 1) high-resolution 3D T1-weighted MPRAGE scan (Magnetization Prepared Rapid Acquisition Gradient Echo) with motion correction; repetition time=2530ms; echo time=2.5ms; inversion time=1260ms; field-of-view=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°; echo train length=625; FoV=144x220mm; voxel size=1.0x1.0x1.0mm). Data analyses were conducted using Freesurfer version 6.0 to determine whole cerebellar, cerebellar grey matter and cerebellar 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 toolbox (version 2.0.15) for statistical parametric mapping (SPM, The FIL Methods Group) as described in Boonstra et al. (Boonstra et al., 2018). Locations of the cerebellar lesions were determined by registration of the lesion images to the cerebellar mask from FSL (Diedrichsen, Balsters, Flavell, Cussans, & Ramnani, 2009).

4.3.4. Speech assessment protocol

The selection of speech variables was informed by findings in previous studies in MS (Noffs et al., 2018). We opted to limit the number of acoustic variables included in the experiment to reduce the probability of falsely rejecting the null hypothesis (type 1 error) and to reduce overfitting for the regression analysis. For acoustic analysis, we selected characteristics that were previously shown to be altered in MS (e.g., fundamental frequency CoV (Hartelius et al., 1997)) or variables which could reflect a perceptual characteristic previously shown to be altered in MS and that had a straight forward calculation (e.g., pause percentage in acoustic analysis as an acoustic proxy for prolonged pauses in perceptual

analysis (Noffs et al., 2018)). Speech variables from both perceptual and acoustic analyses were organized into three major domains and are summarized in Table 4.1.

Speech samples

We applied five standardized speech tasks to elicit the desired speech samples, see Appendix A. Tasks fit along a spectrum of automaticity (Adam P Vogel et al., 2014), from simple to complex (phonetically and/or cognitively): 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 quick and clear as possible for 10 seconds (diadochokinetic speech, i.e. fast alternation of movements, DDK), 4) reading a phonetically balanced paragraph at a comfortable/habitual pace (READ), and 5) telling a personal story from memory with positive content (one-minute unscripted monologue, FREE). Each task, except for the monologue, was elicited twice to reduce the effects of unfamiliarity (Vogel & Maruff, 2014). Only the second trial was used for analysis. All recordings were completed in a single session, in a quiet room without acoustic isolation, reflecting real-world clinical practice, using a Roland Quad-Capture recorder with an AKG C520 cardioid head-mounted condenser microphone (frequency range, 20Hz-20kHz; sensitivity, -43 dB) positioned approximately 8 cm from the mouth, at 45° angle laterally and inferiorly. Recordings were sampled at 44.1 kHz and quantized at 16 bits.

We used Audacity® v2.1.2 (Free Software Foundation, Boston, USA) for auditory and graphic manual screening of all speech samples to exclude or correct for procedural errors. Silences were trimmed before and after the end of speech, tasks using supervised automated editing based on sound energy, in Matlab© v2018b (The Mathworks Inc, US). Speech samples were then renumbered and randomized before perceptual analysis.

Table 4.1. Speech variables grouped by domains and assessment method.

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 rate (task time – total pause time)	DAYS, READ
		Speech rate (syllables/second)	DDK, DAYS, READ
		Pause variability (silences length SD)	READ, FREE
		Articulatory variability (syllables length SD)	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) Loudness instability (energy CoV)	VOWEL
		F2 articulation speed (f2 CoV/articulatory time)	READ
		Energy decay (mean of differences in energy between starting and finishing quarters 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	DDK, FREE, VOWEL

(Referent to Table 4.1.) A summary of acoustic items with calculations shown inside brackets. CoV = coefficient of variance (i.e. SD/mean); f0 = fundamental frequency; F2 = second formant; SD = standard deviation.

Perceptual analysis

One speech and language pathologist and one otolaryngologist, both specialized in neurological dysarthria, rated all audio recordings using a 26-item scoring protocol (Table 4.1 and Appendix B). Normal or unremarkable speech was scored as zero, 1=subclinical, 2 mild, 3=moderate, and 4=severe. Each rater scored all samples independently, with disagreements resolved by consensus between raters.

Acoustic analysis

We restricted the acoustic analysis of DDK to after the first train of *pa-ta-ka,* for the length of two seconds. Typically, in naturalistic speech, the first few syllables are produced with a much higher energy (loudness), adversely affecting the threshold to be used. Conversely, the initial few seconds of the task offer the most intrasubject repeatability (Tjaden & Watling, 2003b) and can be produced by most people despite the degree of dysarthria. The analysis of sustained vowels was confined to the middle 3 seconds for stability and consistency across participants (Ray D Kent et al., 1999). No segmentation was applied to other parts of the speech sample, except for removal of all audio data that preceded the first word or that followed the last word uttered by the participant for the analysed task.

For female and male, respectively, fundamental frequency boundaries were set to 100Hz and 300Hz, and 70 to 250Hz (Vogel et al., 2009), with the analysis window length was set to 30ms and 42ms. Window shift was fixed at 10ms. We used Snack in Voicesauce (Y.-L. Shue, 2010), a Matlab-toolbox, to compute the variables of *cepstral peak prominence* (CPP), *harmonic-to-noise ratio* (HNR, zero to 3,500Hz band), *fundamental frequency* mean and standard deviation, and a list of time-point values of the second formant. The latter was used to calculate the *second formant (F2) articulation speed* for the reading passage. As with other

included features, the speed of F2 movements has been found to be impaired (slowed) by MS and other neurological diseases (K. M. Rosen et al., 2008) by dividing time was defined by the total time in milliseconds that contained a positive value for F2 (i.e. the number of analysis windows with F2 multiplied by 10). Energy was also extracted using Snack in Voicesauce to calculate *energy decay*. First, utterances were automatically segmented using which was defined as the difference between the mean energy in the first quarter of a utterance and the mean energy in the last quarter of the utterance. First, as the difference between for energy-based first order calculations (means, standard deviations, perturbation, percentiles).

Features from the DDK task were extracted with MDVP© (Kay-Pentax)(Wang et al., 2009). By default, the threshold for detection of syllable boundaries was set to 40% of the maxima, with light smoothing. If the threshold line crossed the energy contour in multiple places or neglected small true peaks determined by manual checking, smoothing would be changed accordingly and lastly the threshold would be manually adjusted, as close as possible to 40% of maxima, to best reflect the true total syllable count.

All timing measures were extracted through automated scripts on Praat© v6.0.28 (Boersma, 2002). Boundaries between silence and speech were determined in the energy-time domain using an energy threshold of 0.65 the distance between the absolute minimum value for energy (i.e., floor) and the trimmed peak (i.e., the 95th percentile for energy). Minimum pause duration was set to 15ms and minimum speech duration to 30ms (K. Rosen et al., 2010).

4.3.5. Patient reported quality of life

Eighty-one participants with MS completed the Multiple Sclerosis Impact Scale (MSIS-29)(Hobart et al., 2001), a questionnaire validated to assess the impact of MS on quality of life, and the Voice Handicap Index (VHI)(Benninger, Ahuja, Gardner, & Grywalski, 1998), a 10-question self-assessment of the impact of vocal impairment on communication.

4.3.6. Statistical Methods

We performed all statistical analyses using IBM SPSS® Version 25.0 except for regression modelling and cross-validation, where we used Matlab (Mathworks). We reported descriptive statistics for age, sex, disease duration, EDSS scores and cerebellar lesion volume through means and standard deviations (SD) or medians and quartiles.

4.3.7. Relationship between characteristics of speech and cerebellar impairment

We used Spearman's or Pearson's correlation coefficients, where appropriate, to gauge the strength of relationships between speech measurements and non-speech variables (i.e., SARA, 9HPT, FSS-cerebellar, MSIS-29, VHI and MRI measures). In a previous unpublished study, we have observed correlations between the speech measurements under investigation here and the overall disease markers of EDSS and total brain volume (supplementary material, Table 4.5). Thus, to test if correlations were specific to cerebellar rather than generalised cerebral pathology, we further calculated correlations between speech and MRI measures controlling for overall brain volume and total lesion load. Because statistical controlling procedures were not appropriate for EDSS, we simply reported correlation coefficients between speech measures and EDSS to allow for comparison against correlations between speech and cerebellar scores.

4.3.8. Determination of an acoustic cerebellar score

We tested SARA and FSS-cerebellar for normality of distribution, kurtosis, and skewness. We then selected the clinical cerebellar score closest to parametric to be used as response variable for regression modelling. To avoid overfitting, we restricted the number of predictors to the eight acoustic variables with the strongest correlation with the selected response variable. The acoustic speech predictors were then entered into a forward stepwise linear regression modelling for the selected cerebellar response score and evaluated using 10-fold cross-validation repeated 50 times. The resulting model, composed by valid speech

variables multiplied by their *beta* coefficients, was recorded as a single dimension 'acoustic composite score'. To facilitate interpretation, values for the acoustic composite score are reported on a percentile scale of the trimmed 95% of results. Higher values denote stronger presence of speech characteristics that were associated with cerebellar impairment.

We tested and reported the redundancy and exchangeability between speech predictors in the model by excluding one at a time from the pool of predictors and remodelling. We then tested the correlation between the acoustic composite score and other validated measurements of cerebellar performance, namely 9HPT and FSS-cerebellar. To understand the contribution of cerebellar lesions to speech impairment, we compared the acoustic composite score between people with MS with cerebellar lesions and without cerebellar lesions or HC. We also assessed the correlation between the acoustic composite score and MRI cerebellar measurements or quality of life scores.

Finally, the composite speech score was used to predict an abnormal 9HPT, defined by two standard deviations away from mean value for their age and sex. We generated predictions for the original cohort in which the acoustic composite score was determined (MS only) and for all participants, including HC for which the model was naïve to. We used binary logistic regression for that purpose and reported accuracy and the ROC curve.

To account for multiple comparisons, we applied a family-wise false discovery rate of 5% (q value)(Benjamini & Yekutieli, 2001) to calculate the critical p value, which corresponds to $p = 0.05$ of single comparison analysis. A p value smaller than the critical p was used to describe valid correlations. We reported the original p values along with the respective critical p for each group of analysis.

4.3.9. Data availability

The data used in the current study comprises naturalistic speech samples which are easily identifiable. Audio recordings will not be made public for ethical safeguard of participants. A large part of the remainder data used in the current study is also part of other

ongoing investigations by the research group and will be made available when those studies are published. The authors agree to share the study protocol and further details of statistical methods upon request.

4.4. Results

Eighty-five people with MS participated in the study. Of those, sixty-eight people completed the study's MRI protocol. Twenty-one age and sex-matched healthy volunteers participated (HC). Table 4.2 describes demographic characteristics and neurological performance, and Table 4.3 provides a summary of cerebellar lesions.

4.4.1. Speech phenotype

Forty-three percent of people with MS and none of the HC presented dysarthria according to blinded perceptual rating. The degree of cerebellar dysfunction as measured by SARA was associated mainly with impaired naturalness and intelligibility, slow speech rate, prolonged pauses and inaccurate pronunciation of vowels and consonants. Various other speech measurements from both acoustic and perceptual analysis correlated with clinical cerebellar scores (supplementary material, Table 4.6).

4.4.2. Acoustic composite score for cerebellar dysfunction

Four acoustic variables were selected through a regression model for SARA. DDK speech rate (t=4.5) and frequency instability during sustained vowel (t=3.7) offered the largest contributions to the specific model, followed by pause percentage in free speech (t=3) and energy instability during sustained vowel (t=2.6). Model fitting (RMSE = 5.467) and cross-validation (RMSE = 4.883) both resulted in adjusted r^2=0.54. The acoustic composite score was determined by the following linear function:

33 + (0.251*pause percentage) + (2.72*frequency instability) + (0.158*energy instability) – (3.2*DDK rate)

Table 4.2. Demographics and clinical characteristics by groups.

	MS cohort			MRI cohort			HC		
N total	85			68			21		
Women	74%			74%			71%		
	Mean	±SD	Range	Mean	±SD	Range	Mean	±SD	Range
Age (years)	47.8	±11.2	22 to 70	46.7	±11.6	22 to 70	46	±14.3	21 to 67
Disease duration (years)	14.6	±8.4	1 to 37	13	±7.7	1 to 31			
9HPT (seconds)	36.8	±31.2	16 to 191	33.9	±29.6	16 to 191	18.2	±2.1	14 to 22
	Median		Range	Median		Range	Median		Range
SARA	8.5		0 to 36.5	7		0 to 23.5	0		0 to 3
FSS-cerebellar	2		0 to 4	2		0 to 4			
EDSS	4		1 to 8.5	3.5		1 to 6.5			

9HPT = nine-hole peg test, FSS = functional system score HC = healthy control, MRI = magnetic resonance imaging, N = number of participants; SARA = Scale for the Assessment and Rating of Ataxia, SD = standard deviation.

Test for redundancy by removal of 'pause percentage' during free speech or 'DDK speech rate' from the predictors resulted in a drop in explained variance of SARA scores to r^2=0.45 and r^2=0.42, respectively. Removal of either frequency or energy vowel instability from the predictors led to three possible alternative models with r^2≈0.51. Each of the alternative models included the previous 'free speech pause percentage' and 'DDK speech rate' added by

one measure from the reading task, 'vowel articulation', 'articulatory rate' or 'pause percentage'.

The initial four variables included in the regression using the complete set of predictors were then condensed into an 'acoustic composite score'. Values of the acoustic composite score were normally distributed. On average, every one-point increase in SARA equalled 5% increase in the acoustic composite score.

Table 4.3. Percentage of participants who presented cerebellar lesions by location.

	MRI cohort (n = 68)	
Cerebellar lesion	Absent (n = 30)	Present (n = 38)
Location		
Anterior lobe		55.3%
Posterior lobe		84.2%
Flocculonodular		23.7%
Vermis		34.2%
Mean whole brain lesion load in thousands of voxels (SD	8 (8.3)	10.3 (10.7)
Age, mean (SD)	44.2 (11.2)	48.8 (11.7)
Women	90%	61%
Disease duration in years, mean (SD)	12.4 (7.6)	13.4 (7.8)
SARA, median (interquartile range)	3.75 (6)	9 (10.2)
FSS-cerebellar	2 (1)	3 (2)
EDSS, median (interquartile range)	2.75 (2.4)	6.5 (2.5)

EDSS = Expanded Disability Status Scale, FSS = functional system score, MRI = magnetic resonance imaging, n = number of participants; SARA = Scale for the Assessment and Rating of Ataxia, SD = standard deviation.

4.4.3. Relationship of the acoustic composite score with clinical scores, MRI measures and quality of life

The acoustic composite score correlated with the other validated clinical measurements of cerebellar dysfunction, namely 9HPT and FSS-cerebellar (Table 4.4), and with MSIS-29 (but not with VHI after adjustment for multiple comparisons, critical $p = 0.0177$). Overall, every 5% increase in the composite speech score meant additional 2 points on the MSIS-29. We observed that all functional cerebellar scores (including the acoustic composite score) correlated similarly and modestly with cerebellar white matter volume.

The acoustic composite score was also associated with cerebellar lesion load but less so than were SARA and 9HPT (Table 4.4). People with MS and MRI detected cerebellar lesion had an acoustic composite score on average 26% worse than people with MS without cerebellar lesions ($p = 0.044$, 95% CI 7.5 to 52%) and 95% worse than HC ($p > 0.001$, 95% CI 52 to 138%).

No correlation was found between any of the functional cerebellar measures/scores (i.e., speech, SARA, 9HPT, FSS-cerebellar) and cerebellar grey matter volume. Additionally, no correlation was observed between functional measures and cerebellar white and grey matter volumes when controlling for the whole brain volumes. Correlation coefficients between individual acoustic and MRI measures are presented in supplementary material Table 4.6.

4.4.4. Prediction of upper hand dexterity by the acoustic composite score

Fifty-five percent of people with MS took longer than age and sex-adjusted normal time to complete 9HPT. Classification accuracy for detecting an abnormal 9HPT in the MS group was 83.3% where sensitivity=81.3%, specificity=85% and positive predictive value= 81.3±13.5% (±95% confidence interval, cut-off point at 51). Receiver operator characteristics (ROC curve) are depicted in figure 1. Classification of all participants at the same time (i.e., MS+HC, where 36% had an abnormal 9HPT) resulted in similar accuracy of 84.9% with

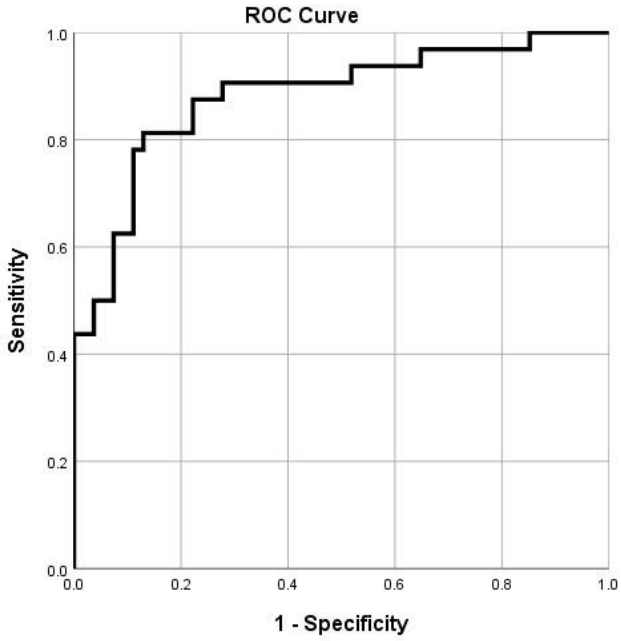
sensitivity=81.3%, specificity=87% and positive predictive value= 78.8±14% (cut-off point at 50.5).

Table 4.4. Correlations between the acoustic composite score and measures of cerebellar impairment.

r (p)			SARA		9HPT		FSS-cerebellar	
	Acoustic composite score		0.745	(<0.001)	0.513	(<0.001)	0.634	(<0.001)
MSIS-29	0.542	(<0.001)	0.535	(<0.001)	0.468	(<0.001)	0.502	(<0.001)
VHI	0.261	(0.028)	0.072	(0.520)	0.178	(0.127)	0.074	(0.510)
Cerebellum volume	-0.267	(0.043)	-0.230	(0.059)	-0.214	(0.085)	-0.124	(0.314)
Cerebellum WM volume	-0.312	(0.017)	-0.322	(0.007)	-0.326	(0.007)	-0.064	(0.606)
Cerebellum GM volume	-0.223	(0.092)	-0.208	(0.0880)	-0.150	(0.228)	-0.161	(0.191)
Cerebellar lesion load	0.310	(0.017)	0.524	(<0.001)	0.475	(<0.001)	0.361	(0.003)

For comparison, correlation coefficients between other functional cerebellar scores and imaging or quality of life are shown. Critical p = 0.017. 9HPT = nine-hole peg test; FSS = functional system score; GM = grey matter; MSIS-29 = Multiple Sclerosis Impact Scale with 29 items; p = "p value", i.e. probability of type 1 error; r = correlation coefficient; SARA = Scale for the Assessment and Rating of Ataxia; VHI = Voice Handicap Index; WM = white matter.

Figure 4.1. ROC curve of accuracy for predicting an abnormal 9HPT in the MS group with the acoustic composite score as predictor.



Area under the curve = 0.87, 95% confidence interval = 0.78 to 0.96.

4.5. Discussion

We developed an experimental composite speech score by combining acoustic speech measures with the strongest independent relationship with cerebellar function, as measured by the SARA score. This acoustic composite score was associated with other measurements of clinical cerebellar function, MS-affected quality of life and with MRI measurements. None of the acoustic measures in isolation correlated with cerebellar lesion load, whereas the composite score did. It is unlikely that a single speech measure will be sensitive enough to gauge the performance of a specific neurological domain. Thus, grouping the most relevant speech measures into one composite score, which is only possible with objective measurements (e.g., acoustic analysis) is likely to improve its validity as a surrogate for cerebellar deficits. In clinical practice or clinical trials, such composite speech score could assist in detecting subtle progressive cerebellar impairment in “real-time”.

Pause percentage during free speech and DDK speech rate was the most relevant features to the speech representation of cerebellar motor impairment as neither factor emerged as redundant in the regression analysis. In line with the compensation and reserve hypothesis of brain disorders (Fornito, Zalesky, & Breakspear, 2015), the requirement of maximum performance of speed in the DDK task and the cognitive-motor complexity of recounting a personal story in the free speech task was arguably responsible for these measurements' success in eliciting speech deficits. A similar detrimental effect on performance in MS was previously observed between concurrent speech and walking tasks (Hamilton et al., 2009), and might inform future research design that intends to elicit or amplify neurological dysfunctions.

Interestingly, pause percentage during free speech was the only functional measurement, not only amongst speech variables but also including clinical measures, SARA, 9HPT and FSS-cerebellar, that remained correlated with cerebellar white matter volume after controlling for overall brain white matter atrophy. This warrants further investigation on causal relationships between the cerebellum, increased pauses during speech and cognition.

Measures of instability during the prolonged vowel task were the remainder variables included in the acoustic composite speech model. Acoustic measures of vowel instability have been previously studied in isolation and shown to discriminate people with MS from controls with very high accuracy (Hartelius et al., 1997). We had also previously observed that frequency instability was one of the few measurements that correlated with whole brain volume in MS (supplementary material Table 4.5). However, their correlation with the severity of cerebellar impairment had not been tested before.

A limitation for gauging cerebellar impairment with any specificity is the heterogenous nature of the clinical presentation and pathophysiology of MS. Dysfunction in other domains including depression, cognitive impairment and visual difficulties (for reading) were not controlled for and may have influenced the results. It should be noted that the current study

compares speech measurements between a wide variety of people in an attempt to reflect real-world conditions, but which could limit conclusions. The observed gradient in speech measurements could be attributed to inter-subject differences not related to disease, such as age, sex, and educational level, given that speech measures also vary widely between healthy people. We have, however, previously examined this in a recent unpublished study which compared inter-subject variability of speech in HC against variability of speech across different MS stages and found that the deviations observed in speech produced by people with MS are much greater than that observed amongst controls.

Another limitation relates to the weak correlations observed between speech measures and neuroimaging, despite clear differences in acoustic composite scores between people with and without cerebellar lesions, which highlights two points of potential improvement. First, tailoring the selection of speech variables to be included in the composite score through correlations with both motor and cognitive cerebellar deficits might improve speech representativeness of the cerebellum. Secondly, the use of more specific imaging techniques is likely to yield correlation estimates closer to true correlations. It is also important to notice that the MRI protocol further included functional imaging for speech and upper limb movement as part of a large study in MS. The length of the imaging protocol was a frequent reason for participants with high disability to decline undertaking that protocol, which potentially decreased the acquisition of images of extreme cerebellar atrophy and lesion load.

Finally, as progression of isolated cerebellar dysfunction is rare in MS (J. Rusz, Benova, et al., 2018), we understand that it should be assessed in conjunction with other naturally occurring impairments, such as pyramidal deficits rather than in isolation. Larger sample sizes are likely to yield a clearer picture of the discrete contribution of the cerebellum to speech and to other phenomena in MS. In that sense, the use of emerging technology using continuous audio recordings from home or online data portals may allow for larger cohorts or more meaningful data. Despite the need for improvement in reliability of data acquisition via those

approaches, the immense potential contribution of coupling objective speech assessment with mobile technology cannot be overstated. Integration of speech analysis into wearables will permit frequent monitoring of neurological functions, real-time detection of events such as relapses and will empower patients through self-monitoring and lateral medicine (Rovini, Maremmani, & Cavallo, 2017).

4.6. Conclusion

Speech measurements such as an increase in pauses, slower maximum speech rate and subclinical voice tremor are associated with cerebellar dysfunction. When combined, those measures are highly predictive of cerebellar dysfunction in MS. Speech samples are easily collected in clinical practice, and its analysis is fully automatable. It presents potential in detecting subtle and/or progressive cerebellar decline. Longitudinal studies are necessary to assess its capacity to detect change over time and to assess its accuracy to predict worsening of cognitive or physical disability as well as responsiveness to treatment modifications.

4.7. Acknowledgements

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

4.8. Funding

This study was funded by NHMRC fellowship grants number 1085461 and 1082910.

4.9. Authors roles

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

Andrew Evans was responsible for the 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; review, critique and writing 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 the conception, organization and execution of the research project; design and execution of the statistical analysis; writing of the first draft.

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

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

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 the conception of the research project; design, review and critique of the statistical analysis; review, critique and writing of the manuscript.

4.10. Possible conflicts of interest

Adam Vogel is Chief Science Officer of Redenlab Inc. He 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.

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.

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

Thushara Perera has nothing to disclose.

4.12. Supplementary material

Table 4.5. Spearman's (rho) and Pearson's (r) correlation coefficients between speech features and general MS pathology and disability.

Speech feature	Brain volume		Brain lesion load		EDSS	
Perceptual	rho	p	rho	p	rho	p
Imprecise consonants	-0.180	0.148	0.119	0.343	0.422	0.013
DDK speed	0.046	0.716	-0.249	0.044	-0.464	<0.001
Prosody rate	0.320	0.009	-0.344	0.005	-0.548	<0.001
Naturalness	-0.149	0.226	0.135	0.273	0.375	0.009
Distorted vowels	-0.103	0.409	0.107	0.394	0.340	0.002
Intelligibility	-0.194	0.119	0.209	0.092	0.328	0.003
Prolonged Intervals	-0.233	0.059	0.206	0.097	0.421	<0.001
Variable rate	-0.284	0.021	0.257	0.037	0.330	0.002
Monoloudness	0.006	0.965	-0.039	0.759	0.259	0.019
Monopitch	0.002	0.986	0.085	0.500	0.321	0.003
Roughness	-0.214	0.085	0.093	0.456	0.340	0.002
Loudness decay	-0.279	0.023	0.102	0.414	0.251	0.023
Acoustic	r	p	r	p	r	p
Speech rate, DDK	0.121	0.338	-0.200	0.111	-0.474	<0.001
Speech rate, READ	0.132	0.306	-0.020	0.877	-0.411	<0.001
Pause percentage, FREE	-0.269	0.031	0.050	0.69	0.457	<0.001
Silence length SD, FREE	-0.154	0.220	-0.014	0.910	0.426	<0.001
Frequency instability, VOWEL	-0.179	0.151	0.046	0.716	0.412	<0.001

Table 4.5. Continued

Pause percentage, READ	-0.06	0.632	-0.089	0.482	0.233	0.036
Articulatory rate, READ	-0.108	0.394	0.035	0.785	0.307	0.005
Pause length SD, READ	-0.029	0.816	-0.065	0.605	0.267	0.016
Vowel articulation, READ	-0.016	0.900	0.178	0.155	-0.205	0.067

9HPT = nine-hole peg test; FSS = functional system score; p = "p value", i.e. probability of type 1 error; rho = Spearman's correlation coefficient; SARA = Scale for the Assessment and Rating of Ataxia.

Table 4.6. Spearman's (rho) and Pearson's (r) correlation coefficients between cerebellar scores and individual speech items.

Speech feature	SARA		9HPT		FSS-cerebellar	
Perceptual	rho	p	rho	p	rho	p
Imprecise consonants	0.535	<0.001	0.476	<0.001	0.521	<0.001
DDK speed	-0.511	<0.001	-0.519	<0.001	-0.401	<0.001
Prosody rate	-0.486	<0.001	-0.391	<0.001	-0.248	<0.001
Naturalness	0.464	<0.001	0.408	<0.001	0.350	<0.001
Distorted vowels	0.472	<0.001	0.400	<0.001	0.446	<0.001
Intelligibility	0.463	<0.001	0.368	0.001	0.383	<0.001
Prolonged Intervals	0.449	<0.001	0.381	0.001	0.294	0.007
Variable rate	0.358	<0.001	0.309	0.001	0.110	0.323
Monoloudness	0.332	0.002	0.300	0.001	0.292	0.008
Monopitch	0.325	0.003	0.350	0.001	0.143	0.199
Roughness	0.316	0.004	0.266	0.020	0.278	0.011
Loudness decay	0.314	0.004	0.288	0.011	0.302	0.006
Acoustic	rho	p	r	p	rho	p

Table 4.6. Continued

Speech rate, DDK	-0.536	<0.001	-0.484	<0.001	-0.339	<0.001
Speech rate, READ	-0.492	<0.001	-0.425	<0.001	-0.379	<0.001
Pause percentage, FREE	0.485	<0.001	0.496	<0.001	0.357	<0.001
Silence length SD, FREE	0.467	<0.001	0.495	<0.001	0.246	0.008
Frequency instability, VOWEL	0.375	<0.001	0.377	<0.001	0.291	0.002
Pause percentage, READ	0.343	0.002	0.259	0.028	0.366	<0.001
Articulatory rate, READ	0.335	0.003	0.327	0.005	0.177	0.059
Pause length SD, READ	0.331	0.003	0.290	0.013	0.217	0.020
Vowel articulation, READ	-0.319	0.004	-0.163	0.161	-0.351	<0.001

Only speech variables with correlation p values bellow critical p = 0.0249 are shown (highlighted in grey). 9HPT = nine-hole peg test; FSS = functional system score; p = "p value", i.e. probability of type 1 error; rho = Spearman's correlation coefficient; SARA = Scale for the Assessment and Rating of Ataxia.

Table 4.7. Correlations between speech MRI measures.

Measures	Cerebellar WM	Controlling for whole brain WM	Cerebellar GM	Controlling for whole brain GM
Speech	r (p)		r (p)	
Speech rate, READ	0.249 (0.051)	0.260 (0.045)	0.164 (0.203)	0.099 (0.453)
Articulatory rate, READ	-0.229 (0.066)	-0.256 (0.043)	-0.113 (0.368)	-0.051 (0.690)
Pause percentage, READ	-0.172 (0.170)	-0.174 (0.172)	-0.169 (0.179)	-0.153 (0.232)
Pause length SD, READ	-0.149 (0.235)	-0.151 (0.238)	-0.072 (0.570)	-0.021 (0.869)
Vowel articulation, READ	0.039 (0.758)	-0.028 (0.827)	-0.102 (0.419)	-0.237 (0.062)
Pause percentage, FREE	-0.411 (0.001)	-0.343 (0.006)	-0.293 (0.018)	-0.152 (0.234)
Pause length SD, FREE	-0.326 (0.008)	-0.281 (0.026)	-0.164 (0.191)	-0.013 (0.918)
Speech rate, DDK	0.231 (0.064)	0.176 (0.167)	0.160 (0.203)	0.133 (0.299)
Frequency instability, VOWEL	-0.083 (0.505)	0.094 (0.461)	-0.021 (0.864)	0.183 (0.149)
	critical $p < 0.01$		No adjustment necessary, all original $p > 0.05$	

Critical p = 0.008. 9HPT = nine-hole peg test; GM = grey matter; MRI = magnetic resonance imaging; p = "p value", i.e. probability of type 1 error; r = correlation coefficient; SARA = Scale for the Assessment and Rating of Ataxia; VHI = Voice Handicap Index; WM = white matter.

Chapter 5

The effect of mobile recording devices on speech acoustic metrics

GUSTAVO NOFFS[1], MATTHEW COBLER-LICHTER[1], FREDERIQUE M. C. BOONSTRA[2], SCOTT C. KOLBE[2], HELMUT BUTZKUEVEN[2], ANNEKE VAN DER WALT[2,3], THUSHARA PERERA[3,4,*] AND ADAM P. VOGEL [1,3,6,7,*]

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

*Authors contributed equally

POTENTIAL CONFLICT OF INTEREST

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

Andrew Evans has nothing to disclose.

Anneke van der Walt receives grant support from the National Health and Medical Research Council of Australia and Multiple Sclerosis Research Australia.

Frederique M.C. Boonstra has nothing to disclose.

Gustavo Noffs receives a scholarship from the University of Melbourne.

Helmut Butzkueven has nothing to disclose.

Matthew Cobler-Lichter has nothing to disclose.

Scott Kolbe receives grant income from the National Health and Medical Research Council of Australia.

Thushara Perera has nothing to disclose.

5.1. Abstract

PURPOSE: To assess consumer-grade mobile microphone configurations for their acoustic measurement agreement with a research reference microphone. METHODS: Speech was simultaneously recorded by one iPod touch (in-built microphone), two different iPod mount-on microphones and a high-quality reference microphone commonly used in research. Twelve speech-acoustic features were calculated using recordings from each microphone to determine the agreement intervals in measurements between microphones. Agreement intervals were then compared to expected deviations in speech in various neurological conditions. Each microphone's response to speech and to silence was characterized through acoustic analysis to explore possible reasons for differences in acoustic measurements between microphones. Additionally, acoustic metrics from consistent rather than interchangeable use of each microphone were explored through correlation coefficients and group comparisons. RESULTS: Microphones' response to speech favoured high frequencies (mean centre of gravity difference ≥ +175.3Hz) and transmitted more relative noise (mean signal-to-noise intensity difference ≤ -4.2dB) for the three consumer-grade microphones compared to the reference microphone. Between consumer-grade microphones, differences in relative noise were related to the distance between microphone and speaker's mouth. Agreement intervals between microphones remained under disease-expected deviations only for fundamental frequency (agreement interval ≤0.06Hz), fundamental frequency instability (agreement interval ≤0.05%) and for tracking of second formant movement (agreement

interval ≤1.4Hz/millisecond). Microphone's acoustic metrics were highly correlated with each other and resulted in equivalent rank-classifications of speech in relation to neurological disability (r=0.84, p=0.04). CONCLUSION: Comparison of research across studies or translation of research findings into clinical applications must account for differences in recording equipment, particularly consumer-grade microphones.

5.2. Introduction

Speech analysis can provide useful insights into the diagnosis and progression of various neurological conditions. Measurements obtained through acoustic analysis have been related to the severity of neurological impairment or were able to distinguish between healthy controls and people with diseases such as depression(Mundt et al., 2012), Multiple Sclerosis (Friedova et al., 2019; Konstantopoulos et al., 2010), Parkinson's disease (Vizza et al., 2019) and others (Daudet et al., 2017; Konig et al., 2019; Konig et al., 2018; Nevler, Ash, Irwin, Liberman, & Grossman, 2019; Starkstein, Jorge, Mizrahi, & Robinson, 2006). Acquisition of speech recordings for health-related analysis has become increasingly accessible with the use of mobile devices in research settings.

Real-world application of remote recordings for health-related analysis requires specific considerations. In contrast to research experiments, clinical practice introduces confounders such as background noise, variation in the distance and direction between a microphone and the sound source, and differences in microphones' properties. Titze and Winholtz (Titze & Winholtz, 1993) varied the distance between professional-grade microphones and a synthesized voice source from 4cm to 30cm to 1m. They observed that the recorded amount of false signal perturbation (i.e., noise; variation in the recorded signal while in the absence of variation in the original signal) progressively increased by almost one order of magnitude at each distance-step, for both frequency and amplitude. Furthermore, the use of consumer-grade microphones resulted in greater false perturbations and less

responsiveness to true perturbations when compared to professional-grade microphones. These findings were associated with the microphone's sensitivity to sound (Titze & Winholtz, 1993). The more a microphone is selectively sensitive to speech, the smaller the proportion of the recorded signal is occupied by noise (environmental and self-generated), i.e. larger signal-to-noise ratio. The same applies to the positioning of microphones where proximity to sound source results in a greater signal magnitude and consequently smaller relative noise. Relative noise interferes with various acoustic measurements related to variation and periodicity (Deliyski, Shaw, & Evans, 2005; Vogel & Maruff, 2008)(e.g., perturbations, standard deviation, harmonics-to-noise ratio, cepstral peak prominence).

Microphone sensitivity varies across frequencies. The distribution of sensitivity along the frequency spectrum is referred to as *frequency response*, which may influence the recorded signal in a predictable manner. Parsa et al. (Parsa, Jamieson, & Pretty, 2001) used loudspeaker playback of human voices to compare four professional-grade microphones and found that systematic divergencies in microphones' frequency responses accounted for small distortions in acoustic analysis, which were sufficient to decrease the accuracy of automated discrimination between normal and pathological voices. Because of the (mostly) systematic nature of the differences, a post-recording mathematical correction accounting for the microphone's frequency response improved accuracy for discriminating between healthy and pathological voices.

Studies that looked at simultaneous recordings of live human speech also showed that microphone quality is important for accurate speech data acquisition, particularly by smartphones and mobile devices. Vogel and Maruff (Vogel & Maruff, 2008) found that some speech measurements related to frequency were comparable while measures of perturbation linked to sound intensity were different between professional-level equipment and a consumer-grade USB-connected laptop microphone. The same authors later observed (Vogel & Maruff, 2014) that the effect of treatment for depression was associated with (landline)

telephone recording acoustics only when the same device was used. Most acoustic metrics from smartphones and a reference microphone showed moderate or high correlation coefficients, yet consistently differed between microphones even in a study with a sample size of only eleven (Lin, Hornibrook, & Ormond, 2012) and often exceeded 10% RMSE (root mean square error) in another (Vogel, Rosen, Morgan, & Reilly, 2014), argued to be unacceptable by the authors. Daudet et al. (Daudet et al., 2017) argued that the in-built microphones of mobile devices are insufficient to filter out noise and that a high-quality microphone coupled to a mobile device is necessary to achieve the high accuracy required for detecting traumatic brain injury in their study. Conversely, Rusz et al. (J. Rusz, Hlavnicka, et al., 2018) found that speech metrics from smartphone recordings were almost as sensitive as professional-grade microphones in capturing key abnormalities in speech of people at high risk of developing or already diagnosed with Parkinson's disease.

The pattern of measurement error (systematic or random) and agreement interval between microphones might also vary according to the acoustic feature under measurement, which could explain discrepancies in the literature. Furthermore, estimating the benefit of adding a consumer-grade mount-on microphone to mobile devices could inform strategies for translating research findings into the clinical use of smartphones and mobile devices for health-related speech analysis. To the best of our knowledge, no study has investigated those factors. In the current study, we compared energy and frequency-based characteristics of speech between three microphone configurations of the iPod Touch and a reference professional-grade microphone to determine the extent to which they can be used interchangeably in speech acquisition for health-related acoustic analysis. Agreement intervals between measurements must be examined in relation to the expected differences between conditions of interest (e.g., affected people vs healthy). Thus, to offer a perspective of the clinical relevance of our results, we compared differences in speech measurements between microphones to differences in speech between clinical conditions found by previous studies.

We hypothesized that agreement intervals between devices would be smaller than the expected clinical effect (thus acceptable) for timing measurements but not for frequency-based measurements, with the exception of fundamental frequency. Lastly, we explored how microphones would separately classify the speech of participants in relation to their neurological status.

5.3. Methods

The study was reviewed and approved by the Melbourne Health Ethics Committee.

5.3.1. Materials

We selected a high-quality configuration previously used in speech research as the Reference microphone, consisting of an AKG C520 head-worn cardioid condenser microphone coupled to a Roland Duo Capture EX USB Audio Interface and connected to a laptop. Characteristics of this microphone make it acoustically ideal for recording speech(Titze & Winholtz, 1993), including a minimum sensitivity of -43dB, a near-flat tonal frequency response, transducer type (condenser), pattern (cardioid) and positioning (close to the sound source). Consumer-grade smartphone-type microphones included three configurations of the 6th generation iPod Touch: (1) the in-built iPod Touch microphone (In-built); (2) Rode IXY-L mobile-mount cardioid condenser microphone (Directional); and (3) Sennheiser ClipMic Digital lapel omnidirectional condenser microphone (Lapel).

5.3.2. Participants

We purposefully recruited people with normal speech as well as people with dysarthria into a single group to explore if acoustic differences between microphones (biases) were proportional to mean measurements. Checking for such proportionality is one of the steps in measuring agreement intervals through the elected method further described under statistical analysis. In total, eighteen people (10 females) aged between 18 and 59 years participated in the study. Inclusion of people with dysarthria was achieved through five

participants with Multiple Sclerosis (two with mild dysarthria, one with signs only, and two with normal speech according to perceptual evaluation). The inclusion of participants with Multiple Sclerosis (pwMS) also served for visualizing the relationship between neurological disability and acoustic scores according to different microphones. All participants were informed of the purpose of the study as well as their right to withdraw at any point, and signed consent was obtained.

5.3.3. Tasks

We used four increasingly demanding speech tasks to elicit increasing complexity in acoustics. These tasks included (1) sustaining the vowel /a/ for ten seconds; (2) repeating /pa/ta/ka/ as fast as possible for ten seconds (speech diadochokinesis task, DDK); (3) reading a phonetically and linguistically balanced text, the grandfather passage; and (4) a one-minute monologue, consisting of recounting a pleasant memory (free speech). These tasks are designed to test specific aspects of speech and are commonly used in voice and speech research(Chan et al., 2019; Noffs et al., 2018; Rodgers et al., 2013; J Rusz, Cmejla, Ruzickova, & Ruzicka, 2011; Tjaden & Watling, 2003a; Vogel, Fletcher, & Maruff, 2010; Vogel & Maruff, 2008, 2014; Vogel et al., 2009; Vogel, Poole, et al., 2017; A. P. Vogel et al., 2014; Vogel, Shirbin, Churchyard, & Stout, 2012).

5.3.4. Recording protocol

Recordings occurred simultaneously through the four sets of equipment in a hospital outpatient clinic room in a quiet corridor away from main thoroughfare and reception areas.

The design and build of a microphone determine its fixed properties (e.g., self-generated noise, sensitivity, frequency response) but also to its intended positioning in relation to sound source (e.g., table-top microphone, hand-held, head-mount, lapel-clip). Thus, to emulate its use by a hypothetical consumer (patient or clinician), each microphone was positioned according to its commercial design as follows. The Reference microphone was

positioned at an angle of 45 degrees laterally to the sagittal plane, axially at the level of the mouth, at approximately 8cm from the participants' mouths, and supported by a back-of-head and around-eyers adjustable arch. The In-built and the Directional microphones were kept on top of a table directly in front of participants, at elbow height and approximately 50cm away from the participants' mouths. To ensure proper orientation of the Directional microphone towards the participant's mouth, its respective iPod was secured in place with a Joby Grip Tight Pro Mount tripod. The Lapel microphone was secured to the participant's clothes (e.g., shirt) over the anterior thorax in a position that would not cause friction or inadvertent motion artifact, at approximately 20cm from their mouth. Those fixed microphone positions were used to record each participant. Unless otherwise specified, these exact recording configurations (microphone + iPod/laptop + positioning) will be referred to as simply 'microphones' (or their specific designation, e.g., 'Lapel') from this point forward. Recording and pre-analyses procedures are further detailed in supplementary methods.

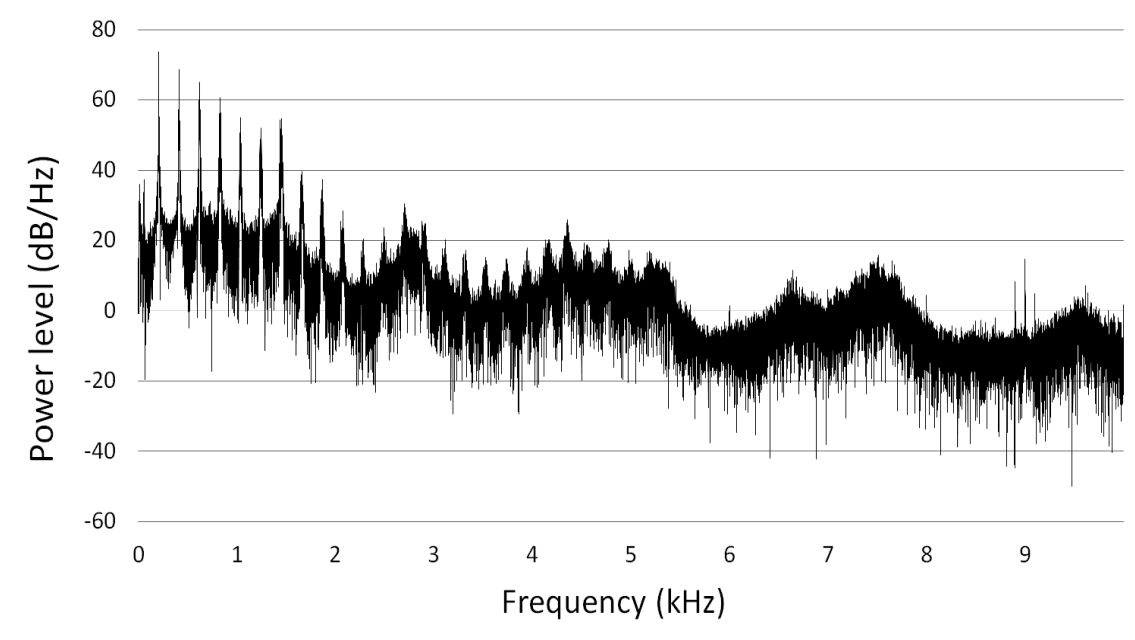
5.3.5. Acoustic analyses

Acoustic analysis was performed in the open-source software Praat (Boersma, 2002). Two groups of speech features were analyzed, one to describe microphones' responsiveness to speech and silence, and another to determine the agreement between measurements taken by different microphones.

We characterized microphones' responsiveness to voice frequencies instead of pure tones. Pure tone responsiveness curves are available online for Reference and Directional microphones. Responsiveness to voice accounts for possible interactions between components of the complex voice waveform and does not require extra data collection. Frequency responsiveness was characterized by power spectral density (PSD, Figure 5.1) and centre-of-gravity (CoG). Additionally, signal-to-noise ratio (SNR) was determined for each participant-microphone and reported as average and confidence interval per microphone. Lastly, we determined the inter-microphones limits of agreement (Figure 5.2) separately for twelve

acoustic measurements (Table 5.1). Methodological details on acoustic analysis, limits of agreement and statistical analysis are included in supplementary methods.

Figure 5.1. Power spectral density levels for mid-vowel /a/, healthy subject, female.



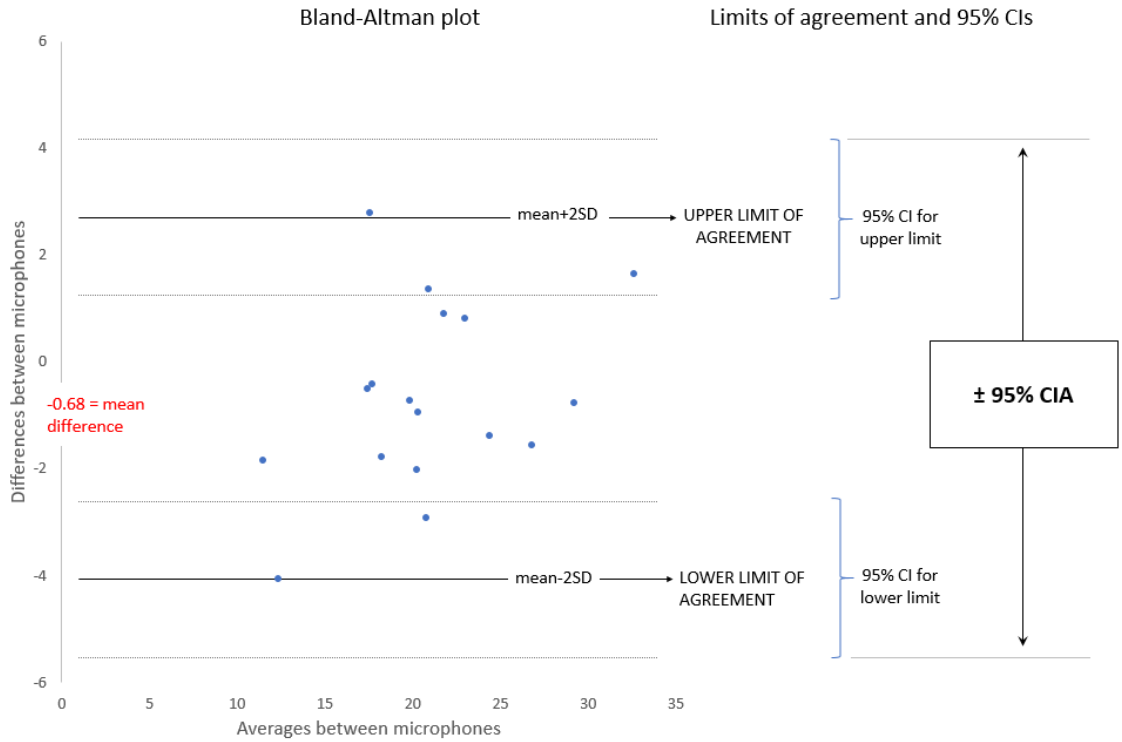
The first peak represents the fundamental frequency followed by peaks of other harmonics. Frequencies above 2kHz have low power peaks, whereas no distinct peaks are observable above 5kHz.

Limits of agreement and confidence intervals alone are insufficient to determine whether devices can be used interchangeably in research and clinical settings. For an overview of the clinical relevance of our findings, we reproduced differences between clinical conditions found in previous studies. In general, if the coefficient of agreement for a pair of microphones is greater than the expected difference between clinical conditions, there would be insufficient reliability for using those microphones interchangeably to measure differences between conditions.

Table 5.1. Summary list of speech measurements tested for agreement.

Frequency related measurements	**Timing measurements**
Mean fundamental frequency (F0)	Mean pause length
F0 coefficient of variation (F0 CoV)	Pause length standard deviation
Mean first formant (F1)	Pause percentage time
Mean second formant (F2)	Total speaking time
Median change in F2 intra-word (F2 slope)	
Mean third formant (F3)	
Harmonics-to-noise ratio (HNR)	
Cepstral peak prominence (CPP)	
Smoothed CPP (CPPS)	

Figure 5.2. Representative Bland-Altman plot with limits of agreement, 95% CIs and 95% CIA.



CI = confidence interval; CIA = confidence interval of agreement; SD = standard deviation.

Lastly, we descriptively explored how microphones would classify speech if the same microphone were used consistently instead of interchangeably. In previous work(Noffs et al.,

2021), we have modelled an acoustic composite score to reflect general neurological disability in pwMS consisting of pause percentage in the monologue, DDK rate, and frequency instability in the sustained vowel. We applied that acoustic model to the recordings in the current experiment separately for the reference microphone and the stand-alone iPod and plotted the results along with actual neurological scores for pwMS. The neurological status of participants with MS was previously determined by a neurologist through the Expanded Disability Status Scale (EDSS) score(Kurtzke, 1983). Given the limited number of pwMS, comparing graphs between microphones for this simulated neurological classification is merely exploratory.

5.4. Results

For simplicity, we excluded recordings of sustained vowel from one neurotypical female volunteer who produced intermittent but frequent vocal fry, which caused extraneous variation (halving) in fundamental frequency measurements. Additionally, recordings of two people with MS were summarily terminated by the device connected to the directional microphone (corrupted data) and thus unsuitable for acoustic analysis.

5.4.1. Microphone characteristics

The intensity of recorded signal during speech intervals (i.e., noise during silence) was greater for the in-built microphone relative to the Reference microphone (intensity floor, table 2). Higher relative noise for the in-built microphone can also be observed during phonation and in more detail on curves of power spectral density (PSD, figure 3) where the 25dB gap in responsiveness to the fundamental frequency between Reference and In-built (first marker in each curve, under 0.3kHz) disappear for frequencies commonly populated by noise (>5kHz). While the three consumer-grade microphones were proportionally more sensitive to higher frequencies in comparison to the Reference microphone (higher CoG, Table 5.2), that difference was more associated with responsiveness to the fundamental frequency and less with noise for both Lapel and Directional microphones (Table 5.2 and Figure 5.3).

Table 5.2. Comparative response characteristics of microphones for the same sound stimulus.

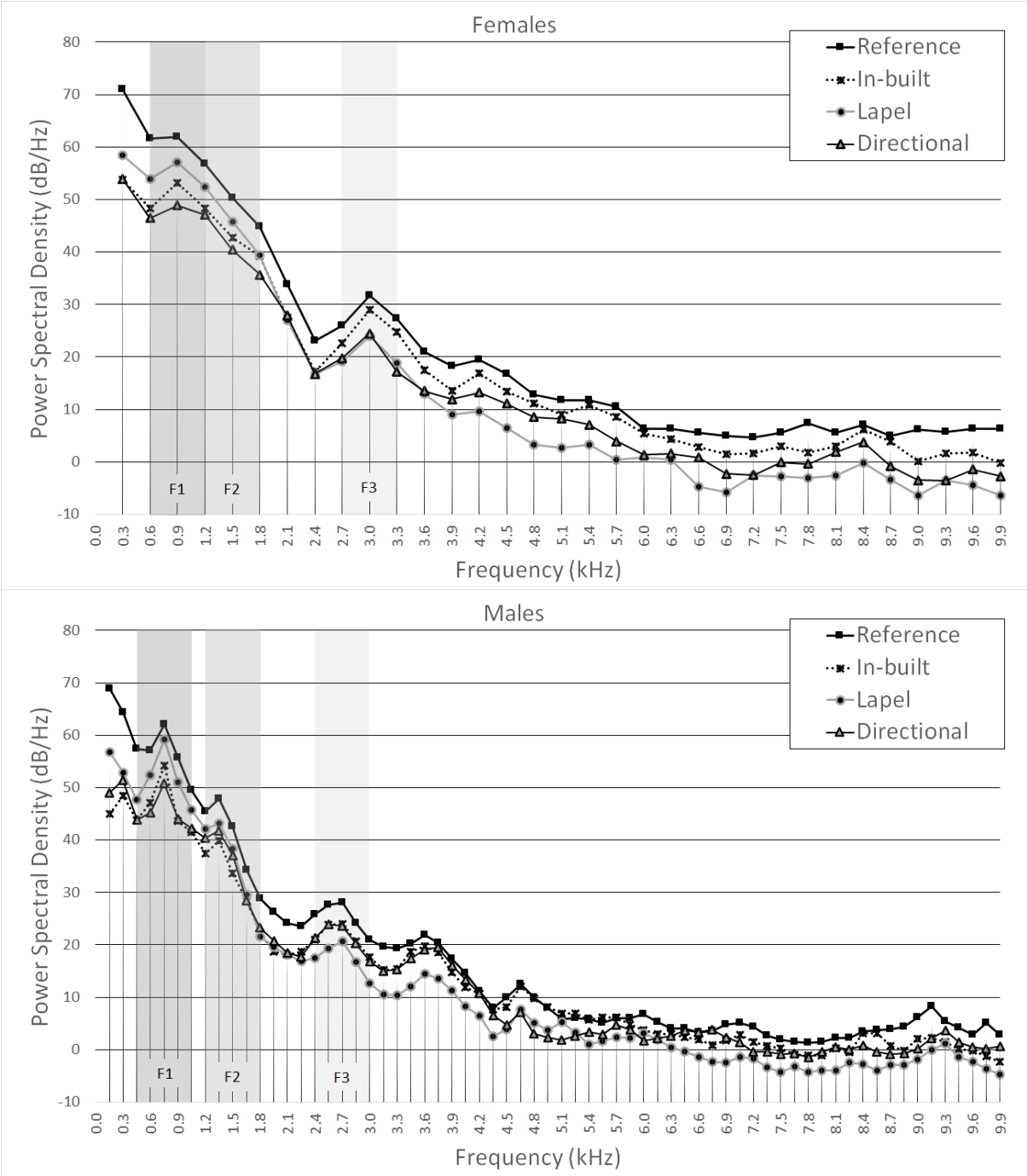
Measurement	Microphone			
	Reference	Lapel	Directional	In-built
Intensity floor (noise), mean ± 95% CI	24.7 dB ±2.6	29.9 dB ±0.5	31.4 dB ±0.5	33.8 dB ±2.4
			vs Reference	
SNR, mean ± 95% CI	50.9 dB ±3.0	-12.1±3.8 (p< 0.001)	-20.1±3.1 (p< 0.001)	-20.5±3.7 (p< 0.001)
SNR adjusted, mean ± 95% CI	50.9 dB ±3.0	-4.2±3.8 (p= 0.030)	-4.2±3.1 (p= 0.031)	-4.6±3.7 (p= 0.018)
Centre of gravity, mean	517.1 Hz ±80	+175.3±39.1[b] (p = 0.006)[c]	+228.3 ±45.7[b] (p = 0.0024)[c]	+254.3±43.6[b] (p = 0.0008)[c]

CI = confidence interval; dB = decibel; Hz = Hertz. [a] Paired t-test. [b] Near-normal distribution. [c] Wilcoxon rank test.

5.4.2. Agreement coefficients

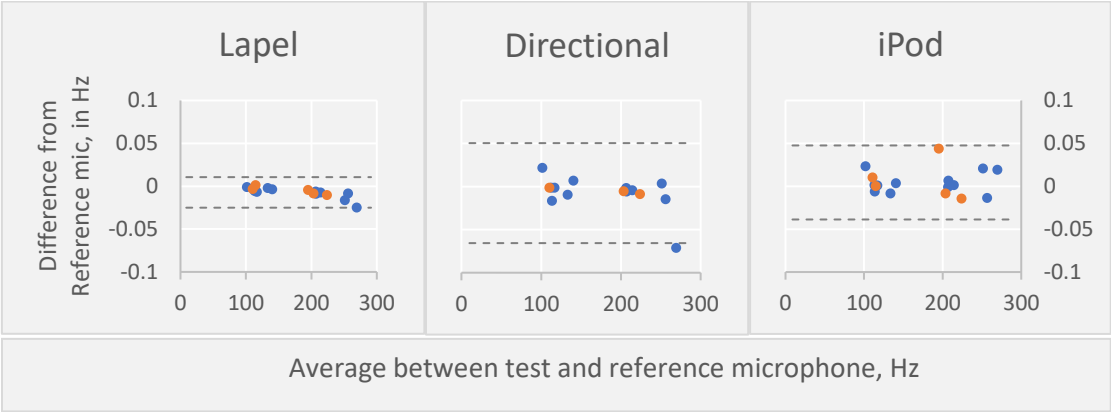
Agreement between microphones was highest for *f*0 measurements (Tables 5.3 and Figure 5.4). Coefficients of agreement and 95%CIAs remained under 0.1Hz for mean *f*0and under 0.05% for *f*0 CoV. Measurements of average formant frequency, HNR and CPP diverged 10% to 25% between microphones. All three formants' SD presented the largest 95% CIA, up to ten times as wide as the formant's average SD, i.e., 1000% divergence in measurements.

Figure 5.3. Average power spectral density (PSD) levels (re 10-5 Pa) for voices of males and females, by microphone.



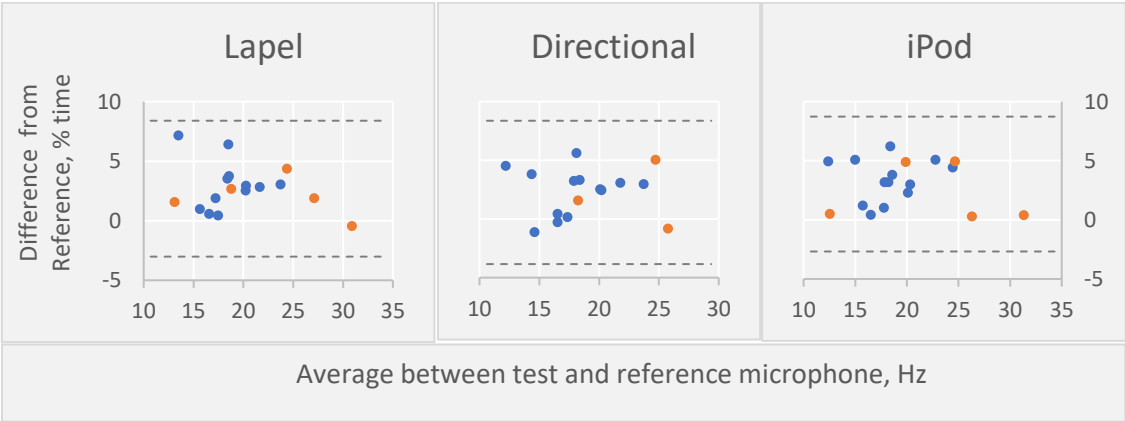
The first marker on each line represents the power in fundamental frequency, whereas shaded areas highlight sound power for formant frequencies (as determined by analysis of recordings by the Reference microphone). F1, F2, F3 = first three formants. Note the number of frequency-points (multiples of the fundamental) for males is twice the number as for females, which is a result of typical frequencies for the fundamental in each sex.

Figure 5.4. Bland-Altman plots and 95% CIA (dotted lines) for f0, showing good agreement between reference and test microphones.



PwMS are shown in red. The average difference between microphones is close to zero, i.e., not systematic (small measurement bias), and differences are independent of average measurement or of the presence of MS. Random disagreement is also small, highlighted by 95% CIA under 0.1% of the average measurement.

Figure 5.5. Bland-Altman plots and 95% CIA (dotted lines) for monologue pause percentage, showing poor agreement between reference and test microphones.



PwMS are shown in red. All test microphones systematically overestimated pause percentage in comparison to the reference microphone (measurement bias), but differences seem independent of average measurement or of the presence of MS. Non-systematic (random) disagreement is reflected in large 95% CIA, which are over 20% of the average measurement.

Among timing measurements (Table 5.4), utterance sum for the reading and free speech tasks presented the most agreement between microphones (5% to 7.5% as measured by 95%CIA) followed by pause length SD in the same tasks and pause percentage in all tasks

(20% to 30%, Figure 5.5). Other measurements had 95%CIAs wider than 30% of the mean for that measurement. Lastly, limits of agreement between the Directional and Reference microphones for DDK mean syllable period could not be calculated as differences were not normally distributed.

Raw and log-transformed differences between microphones for the acoustic composite score were not normally distributed, thus not suitable for calculation of agreement intervals through the elected method. The acoustic composite score determined from recordings by each microphone separately yielded a similar classification of neurological impairment of speech as visualized in Figure 5.6.

Figure 5.6. Relationship between neurological disability status and acoustic score for the high-grade reference microphone set and the in-built microphone.

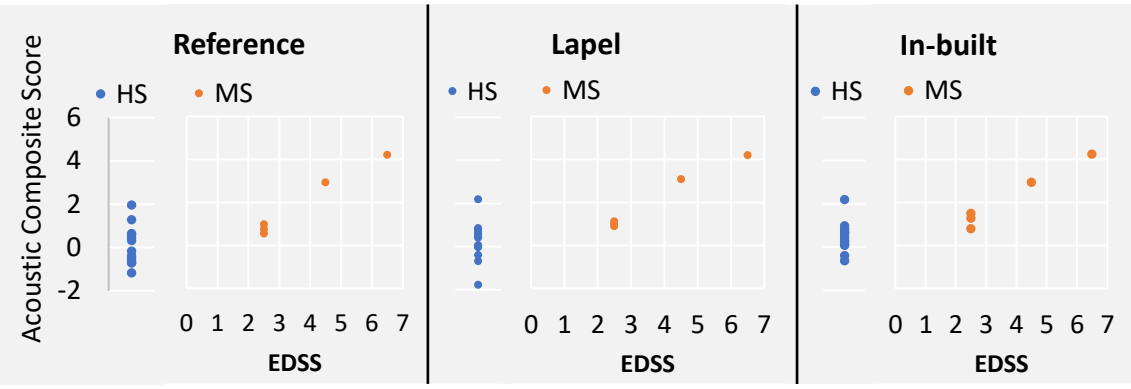


Table 5.3. Bias and 95% confidence interval of agreement between Reference and tested microphones for frequency-dependent measurements.

Measurement	Reference microphone (mean)	Unit	Test microphone	Bias (mean difference)	95% CIA	Previously observed difference
*f*0	177.3	Hz	Lapel	-0.007	**±0.018**	37.8 Hz, males, Parkinson disease(Gamboa et al., 1997)
			Directional	-0.059	**±0.058**	
			In-built	-0.601	**±0.043**	
*f*0 CoV	0.311	%	Lapel	-0.001	**±0.021**	0.4 to 1%, Multiple Sclerosis(Hartelius et al., 1997)
			Directional	-0.009	**±0.031**	
			In-built	-0.020	**±0.048**	
F1	708.6	Hz	Lapel	NA	NA	74Hz, intelligibility in Cerebral Palsy(Liu, Tsao, & Kuhl, 2005)
			Directional	50.43	±221.2	
			In-built	24.02	±247.8	
F2	1359.9	Hz	Lapel	-16.84	±296.7	
			Directional	60.88	±203.4	
			In-built	58.78	±290.7	
F2 slope	8.14	Hz/ms	Lapel	-0.024	**±0.831**	1.7 Hz/ms, Multiple Sclerosis(K. M. Rosen et al., 2008)
			Directional	0.290	**±1.088**	
			In-built	0.357	**±1.402**	

Table 5.3. Continued

F3	2790.2	Hz	Lapel	-225.20	±1099.7	
			Directional	-20.45	±355.6	
			In-built	-11.10	±385.6	
HNR	25.92	dB	Lapel	-2.976	±4.592	1 to 16dB, Post glottic cancer surgery(Motta, Cesari, Mesolella, & Motta, 2008)
			Directional	-6.455	±4.641	
			In-built	-6.789	±5.432	
CPP	28.06	dB	Lapel	-0.717	±6.498	1dB, Friedreich's Ataxia(Vogel, Wardrop, et al., 2017)
			Directional	-1.540	±4.595	
			In-built	-2.074	±7.012	
CPPS	20.81	dB	Lapel	-0.952	±1.929	1dB, dysphonia in benign pathologies(Brinca, Batista, Tavares, Gonçalves, & Moreno, 2014)
			Directional	-2.058	±2.284	
			In-built	-2.593	±2.666	

Confidence intervals (second last column) smaller than the expected clinical difference (last column) are highlighted. The last column refers to previous findings from studies where speech measurements were compared between clinical conditions or to disease severity scores. 95% CIA = 95% confidence interval of agreement; CPP(S) = cepstral peak prominence (smoothed); dB = decibel; f0 = fundamental frequency; F(1-3) = formants; HNR = harmonics-to-noise ratio; Hz = Hertz; ms = milliseconds.

Table 5.4. Bias and 95% confidence interval of agreement between Reference and tested microphones for timing measurements.

Measurement	Reference microphone (mean)	Unit	Test microphone	Bias (mean difference)	95% CIA	Previously observed difference
Pause %, reading	19.98	%	Lapel	3.791	±4.556	2.7%, Parkinson's disease(Skodda & Schlegel, 2008)
			Directional	2.415	±5.647	
			In-built	2.966	±5.844	
Mean pause length, reading	187.8	ms	Lapel	-10.600	±89.43	
			Directional	7.007	±77.9	
			In-built	6.292	±77.17	
Pause length SD, reading	210.0	ms	Lapel	9.427	±71.82	77ms, Multiple Sclerosis(Noffs et al., 2021)
			Directional	21.693	±78.78	
			In-built	22.130	±75.92	
Speaking time, reading	22.27	s	Lapel	-1.055	±1.381	
			Directional	-0.660	±1.313	
			In-built	-0.834	±1.604	
Pause %, free speech	18.10	%	Lapel	1.851	±11.63	6.6%, Multiple Sclerosis(Noffs et al., 2021)
			Directional	2.264	±7.781	
			In-built	3.030	±5.705	

Table 5.4. Continued

Mean pause length, free speech	161.1	ms	Lapel	9.82	±100.81	40ms, response to treatment for depression(Alpert, Pouget, & Silva, 2001)
			Directional	29.14	±96.86	
			In-built	27.48	±81.74	
Pause length SD, free speech	192.9	ms	Lapel	12.17	±107.7	
			Directional	27.06	±77.23	
			In-built	26.11	±62.1	
Speaking time, free speech	42.27	s	Lapel	-4.362	±34.8	
			Directional	-1.285	±3.17	
			In-built	3.582	±2.306	
Mean pause length, DDK	48.10	ms	Lapel	14.89	**±37.13**	50ms, Multiple Sclerosis(Tjaden & Watling, 2003b)
			Directional	20.38	±53.89	
			In-built	22.51	**±37.17**	
Mean speech rate, DDK	5.89	syl/s	Lapel	-0.043	±2.538	2.2syl/s, Multiple Sclerosis(Tjaden & Watling, 2003b)
			Directional	-0.200	**±2.109**	
			In-built	-0.231	±2.697	

Confidence intervals (second last column) smaller than the expected clinical difference (last column) are highlighted. The last column refers to previous findings from studies where speech measurements were compared between clinical conditions or to disease severity scores. 95% CIA = 95% confidence interval of agreement; ms = milliseconds; s = seconds; syl = syllables.

5.5. Discussion

We found that measurements of the fundamental frequency (*f*0) were robust to variation in recording device sets. In addition, F2 slope values had agreement intervals marginally smaller than clinical effect sizes. Other measurements, including all timing metrics, were largely influenced by the recording device type, which does not support our initial hypothesis.

The following characteristics of the fundamental frequency allow it to be estimated through simple methods, which might explain its robustness observed in the current and previous studies. The quasi-periodic vibration of the vocal folds (glottis) during phonation is responsible for *f*0, which is theoretically the lowest periodic frequency in the voice waveform. Because this glottal sound is the major energy source for the production of other sounds in voice, its energy relative to the total voice energy is great enough to be detected through direct analysis of the raw (unprocessed, untransformed) recorded signal. The fundamental frequency is determined by the auto-correlation method in Praat, which consists of simply making an identical copy of the signal and then calculating the correlation coefficient between original and copy each time the copy is slid forward by a chosen mount in the time domain. The predominantly periodic nature of voice means that the sound wave "repeats itself" at almost fixed intervals, at least in short time windows. Thus, sliding the copy by exactly one such "repeating interval" will yield a very high correlation coefficient between copy and original. The amount of displacement of the copy for which the highest correlation coefficient is found defines the size of the repeating interval, called fundamental period. The fundamental frequency, in turn, refers to the number of periods that occur per time (per second, in the case of Hertz). Because the auto-correlation method examines the whole waveform, it might be susceptible to inaccuracy for calculating the fundamental frequency when other components of the waveform (i.e., formants) change frequencies within the analysis window. This occurs, for example, during the production of more than one phonetic vowel within one utterance in

naturalistic speech. That would not necessarily affect its agreement between microphones, i.e., estimations would be “equally wrong” across microphones.

Other frequency-related measurements depend on the decomposition of the recorded signal to produce estimates. The sound energy of higher frequencies is often many times smaller than that of the fundamental frequency in quiet (conversational level) utterances. In the current experiment, tracking of F2 movement was less affected by different microphones than location estimates. Determining the location of formants, such as mean F2, depends on the prominence of semi-periodic high-frequency signal (above noise, which is often also high-frequency but aperiodic). Thus, both microphone’s frequency-specific responsiveness and noise levels may affect the estimation of the formant’s location. Estimates of a formant’s location might diverge between microphones more than estimates of that formant’s movement, e.g., F2 slope, as long as those estimates move in parallel between microphones (same direction and about the same magnitude).

Surprisingly, we found that timing measurements varied widely between recordings from different microphones. The timing analysis method in the current experiment is based on sound intensity in the time domain. The threshold for separating silence and speech is determined in relation to the maximum sound pressure level. There are several microphone attributes that could cause a lack of agreement for the measurement of timing and/or high frequencies. These include transient responses, i.e. the speed of microphones to respond to an abrupt sound (Peus, 2004). Different transient responses affect the onset time and onset slope of speech. As transient responses are very short by definition, they would affect more the measurement of precise (time-wise short) speech events, as consonant duration, for example, and only discretely the measurement of long speech events such as utterance duration, which consist the bulk of timing metrics analysed in the current study.

Secondly, microphone positioning might greatly affect speech measurements. Increasing the distance between sound source and microphone makes the signal more

susceptible to non-free field influences (e.g. reflection and refraction of sound in out-of-lab environments)(Eargle, 2003), which inflates measurements related to noise and perturbation or distortion (Titze & Winholtz, 1993). As expected, the decrement in SNR was proportional to the distance between microphone and mouth (source). The same decrement could be observed for clinical measurements that involve ratios between periodicity and noise, namely HNR, CPP and CPPS. The consequence of decreased SNR for timing analyses is briefly discussed below. The angle of placement in relation to the sound source can also influence measurements to a smaller degree(Titze & Winholtz, 1993), while the relation between the microphone and body parts can cause frequency-dependent distortions (Feigin, Barlow, & Stelmachowicz, 1990).

The third possible cause for disagreement between microphones concerns frequency responsiveness. In this study, the Reference microphone's response favoured low frequencies in comparison to the other microphones, as shown by PSD curves and the lower CoG. A microphone's responsiveness to frequencies that naturally have a higher speech energy density, particularly the fundamental frequency in vowels, will largely influence the overall measured sound intensity. This overall intensity is the basis for determining the silence-speech threshold used in timing analyses. However, the onset and offset of syllables are not necessarily populated by low-frequency energy-dense speech, which creates a mismatch between the automatically determined threshold and the actual amplitude of onset/offset. Thus, optimizing the silence-speech threshold for the Reference microphone causes it to be sub-optimal for another microphone and vice-versa. Frequency responsiveness further varies in relation to distance to sound source depending on the polar pattern of microphones. A noticeable boost in low frequencies is observed for cardioid microphones (e.g. the Reference and the Directional) when that distance is shortened, i.e., proximity effect (Clifford & Reiss, 2011).

Fourth, differences in relative noise level likely have also influenced timing measurements. Microphones vary in self-generated noise (Peus, 2004), electrical insulation and filtering of background noise. The polar pattern of microphones (directionality) also affects the amount of background noise captured (Vogel & Morgan, 2009). In our experiment, after discounting the theoretical loss in signal intensity due to the distance between microphone and mouth (source), a residual decrease in SNR of about 4dB remained for the three consumer-grade relative to the Reference microphone. That could be due to both reduced microphone sensitivity and increased recorded noise. The effect of SNR on HNR, CPP and CPPS was discussed above. In the current experiment, decrease in SNR means a reduced difference between speech intensity and "silence" intensity (i.e., background noise). That difference between intensities is the window for silence/speech detection. A smaller SNR results in a greater chance of misclassifying speech or silence for analysis based on intensity. Thus, even if the silence-speech threshold (timing analysis) were optimized individually for each microphone, its accuracy would be worse for the three consumer-grade microphones or any microphone in conditions of excessive noise.

Those four interrelated differences were present between the tested microphone settings in the current experiment and might explain, at least in part, the lack of agreement between them. Importantly, these discussed differences may be compounded by other commonly unaccounted variables in clinical settings, such as wide variations in the recording environment, bodily movements and differences in algorithms (software) used for analyses, which introduce multiple and interrelated confounders when comparing acoustic results from multiple recording protocols (Raymond D. Kent & Vorperian, 2018).

Interestingly, increased relative noise caused merely by differences in microphones and/or distance translated into more dysarthric-like values for acoustic metrics. In general, reduced SNR generated a higher pause percentage, higher pause variation, lower harmonicity (HNR and CPP) and lower DDK rate. A similar parallel between noise and dysarthria has been

observed for intelligibility as assessed by listener's transcription accuracy (Borrie, Baese-Berk, Van Engen, & Bent, 2017). An acoustic distortion of the speech signal by noise might explain its similarities with dysarthric speech, although perceptual intelligibility involves additional and more complex acoustic features than those included in the present study, thus warranting further investigation.

Limitations of the current study include low speech sample size, which increased the probability of non-normal distribution of measurements and large confidence intervals. We contacted the manufacturer (Apple Inc) asking whether the same in-built microphone sensor was present across other mobile devices (e.g., iPhones, iPads) but are yet to receive a response. Thus, although other mobile devices may produce similar agreement limits due to inherent constraints in size and power consumption, the specific agreement intervals and biases reported here relate to only the tested microphones and cannot be generalized.

In our opinion, speech research must be designed to mitigate the effect of common confounders related to hardware use. The only immediate and certain way to eliminate confounders is to standardize the speech assessment protocol, which has been advocated for many years (Svec & Granqvist, 2010; Titze, 1994). That includes the use of the same equipment between participants and across time, fixing the distance between microphone and participant, and avoidance of noise, among others. There is, however, a trade-off between the number of confounders that are controlled in this way and the flexibility in data acquisition, which translates into a trade-off between precision and clinical scalability. Another concurrent path relates to the development of robust methods of signal analyses, which include numerous recent efforts showing variable progress. The ideal outcome is to increase robustness without compromising validity, precision and accuracy. CPPS, for example, could be considered such an improvement over the original CPP for the estimation of dysphonia (Heman-Ackah et al., 2003). In the current experiment, we observed increased stability in measurement between different microphones for CPPS in comparison to CPP at the cost of a

smaller average measurement. A third path for optimizing reproducibility could be the use of combined measurements to gauge the desired outcome. Examples include principal component analysis, machine learning models or multiple regression models, such as the acoustic composite score used in the current study. Composite scores mitigate, in general, the effect of random error of its components once errors in opposite directions cancel one another. Such stabilizing effect might, however, also decrease the responsiveness of composite scores, which must be tested and weighed against its gain in reproducibility. The multiplicity of measurements used as predictors is not confined to their instantaneous qualities but could also include temporal patterns (in repeated measurement conditions). In general, the main requirements for such models are computational power (increasingly available and inexpensive) and large datasets. In order to produce high-quality research, we must consider the variations in signal acquisition highlighted in this study and develop strategies to overcome these where possible.

5.6. Conclusion

Measurement of the fundamental frequency of speech is robust to variation in recording equipment. Other tested acoustic metrics are sensitive to recording variations, which may relate to differences in signal-to-noise ratio (largely affected by mouth-microphone distance) and frequency responsiveness (mostly determined by the microphone's quality). Using the same device across repeated measurements for individual trajectories can help to mitigate measurement error and allow for the clinical use of speech analysis for monitoring neurological diseases.

5.7. Open practices statement

Data will not be made public for the ethical safeguard of participants as it may be identifiable. The authors agree to share the study protocol and further details of statistical methods upon request. The experiment was not pre-registered.

5.8. Supplementary material

Recording

Audacity was used to record all speech by the reference microphone, and the Rode Rec app was used to record all speech by each of the mobile devices. Before recording, digital filters and enhancers were turned off on the app. Recording parameters were kept constant for the entire experiment, including a sampling rate of 44.1kHz, 16-bit quantization and constant gain (tested and set before the start of recording for each participant).

At the beginning of each recording session, a pure-frequency beep (500Hz, 500ms) was played from an independent device, more than 1 meter away and equidistant from all devices, to serve as a reference time-point for later synchronization of recording files. The four microphones simultaneously recorded each participant, who produced each speech task only once. The door of the room was kept closed, and recordings were stopped, discarded, and re-initiated in case of interruptions (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 at 2 meters distance from the wall in front of them. For descriptive purposes only (rather than for room acoustic measurements), the room was rectangular and measured 3x5m, 3m high ceiling, with solid brick walls, carpeted floor, and furnished with office and research equipment.

Pre-analysis processing

Each microphone recorded one continuous audio file per participant, including all tasks. To ensure that subsequent analyses would be carried over the same speech excerpt (i.e. synchronization), the beginning of each file was set to match the starting time-point of the initial reference beep. To ensure that time between sound events were not significantly affected by differences in microphones' transient response or by possible mechanical-to-digital conversion delays (e.g. due to differences in CPU processing speed while recording), the interval between the start of the beep and the start of the sustained vowel was compared

between microphones. Additionally, the same beep-vowel interval was determined manually twice in each recording of two participants for all microphones (intra-screener error). The average difference between microphones was similar to the intra-screener error, on average of 10 samples (approximately 0.2ms). Thus, differences in time between speech events between microphones were too small for detection through manual screening and could be ignored.

Next, the beginning and end of each speech task, i.e. time-boundaries, were determined manually for the audio file recorded by the Reference microphone. The maximum intra-screener error for single time-boundaries was eight audio samples (approximately 0.2ms) and was estimated on twenty randomly selected time-boundaries, which were determined a second time after all time-boundaries had been determined once.

The originally determined time-boundaries for each task were then applied to the other three files (In-built, Directional and Lapel). This process was repeated separately for each participant. Files were divided so that each post-processing file contained only one task from one microphone and from one participant, resulting in 288 speech files (18 participants x 4 tasks x 4 microphones). The files were equal in starting and finishing time across the four recording devices for each participant.

To emulate procedures applied in previous research for the analysis of specific speech features, the vowel sound task (1) was trimmed for the middle three seconds(Hartelius et al., 1997), while the DDK files were trimmed to include exactly two seconds of that task after the first train of syllables (i.e. after the first 'pataka')(Ray D Kent et al., 1999; Tjaden & Watling, 2003a).

Microphone responsiveness to speech and silence

Frequency responsiveness was characterized by power spectral density (PSD). PSD here refers to the power level (in dB/Hz, relative $2x10^{-5}$ Pa) of the signal (i.e. speech) rather than referring to the absolute power (in Watts) of the sound source (participant). In short, PSD

quantifies how much “each frequency” is contributing to the total sound power. It also allows comparison of power between different frequencies (e.g. between harmonic and noise-related frequencies) and offers a visual representation of the components of the complex waveform (figure 1). Because our aim was to compare microphones rather than compare speakers’ outputs (i.e. speech signals), PSD was averaged across participants for each microphone and compared between microphones. In voice, peaks of sound power occur at intervals very close to the fundamental frequency(Kunieda, Shimamura, & Suzuki, 1996) (i.e. close to harmonics). For that reason, a window size slightly longer than the highest fundamental frequency among all participants of the same sex was used. Males and females were computed separately. For each participant, the highest peak (i.e. maximum power) within each such frequency window was selected. Then, the average maxima among all participants of the same sex was calculated for each frequency window. Otherwise, averaging the whole power spectrum across participants would result in an undesirable cancellation effect between peaks and troughs once each person produces power peaks at different frequencies. For visualization of differences between microphones, the average maxima for each frequency window were plotted in a graph. Differences between microphones for a given PSD frequency window reflect diversions in responsiveness to speech (or noise) in that frequency band. Responsiveness to frequencies of interest can then be compared (e.g. between frequencies commonly populated by the fundamental frequency, formants and noise).

The second feature used to characterize frequency responsiveness was centre-of-gravity (CoG), also determined from the power spectrum. CoG is measured in Hz and can be understood as the power-weighted mean frequency of a sound, where the sum of power in frequencies higher than CoG equals the sum of power in frequencies lower than CoG. It was determined for each participant (including all frequencies irrespective of peak presence), then averaged across all participants for each microphone. Average CoGs and their 95% confidence intervals were then statistically compared between microphones.

Lastly, signal-to-noise ratio (SNR) was defined as mean intensity in decibel (dB) of the vowel's middle three seconds minus the whole file intensity floor. SNR was calculated separately for each participant-microphone file. Then the difference in SNR between the Reference and each consumer-grade microphone was calculated per participant. Mean SNR differences, 95% confidence intervals (CI) and t-test p-values are reported. Because the distance between microphone and participant affects the intensity of the signal of interest (speech) but not the intensity of background noise, we would expect different SNR with different distances to participants. Considering the mouth as the speech source, the distance for the Lapel remained at approximately 20cm/8cm=2.5 times that of the Reference, while both In-Built and Directional microphones rested at 50cm/8cm=6.25 times the distance between Reference and source. As intensity is inversely proportional to the square of the distance to the sound source, distance alone would account for $10*Log10(2.5^2) = 7.96$dB loss of intensity for the Lapel and 15.92dB loss of intensity for In-Built and Rode. Differences between microphones in SNR corrected for distance (SNR adjusted) is also reported.

Determining agreement between microphones

The second group of analyses produced speech measurements that have been previously used in speech research related to neurological conditions and a few additional related measurements (see Table 5.1 for a summary list). We calculated the agreement between measurements derived from recordings of different microphones. We measured the frequency for the VOWEL sample, which contains only continuous voicing, to mitigate possible irregular tracking of frequencies, especially formants, by the spectrum algorithm during abrupt speech transitions (voicing onset/offset and consonant-vowel). The only measurement derived from the reading task was median F2 movement.(K. M. Rosen et al., 2008). Principles of phonetics and screening of all spectrograms directed our selection of candidate words for that measurement. The word "well" was chosen for its consistent change in F2 and for its

consistent continuity in fundamental frequency between participants (i.e. no voice breaks, no creaky voice).

Extraction of frequency features was automated in Praat. Fundamental frequency was determined through autocorrelation using minima and maxima fundamental frequencies of 70Hz and 250Hz for men and 100Hz and 300Hz for women.(Vogel et al., 2009) The analysis window was kept at 42ms and 30ms, respectively, i.e. three times the size of the longest wavelength, and window shift was fixed at 10ms. As per Praat guidelines and default settings, the maximum number of formants was kept at 5 with a maximum of 5500Hz for formants' detection. All other parameters were maintained at software default.

Timing measurements were also automated in Praat and separately calculated for the speech tasks DDK, reading and free speech. The detection of silence-speech and speech-silence transitions was done using a threshold on the energy domain relative to trimmed peaks as in previous works(Adam P Vogel et al., 2014; Vogel et al., 2011; Vogel et al., 2018). The threshold was set to 65% of the 95th percentile, with minimum silence length set to 20ms and minimum speech length to 30ms.

Statistical analysis

Descriptive results included means (standard deviations, SD) and medians (interquartile ranges, IQR). We estimated limits of agreement between microphones through the Bland-Altman method(Bland & Altman, 1986). Each mobile device was compared to the reference recording device and to each other, resulting in six pairwise comparisons for each acoustic measurement. In summary, we used scatter plots of pair-wise averages by pair-wise differences in measurements to screen for outliers, measurement bias (systematic errors), and error dependency (i.e., a relationship between error magnitude and mean measurement). Measurement bias was also quantified through mean differences between pairs of microphones.

We used histograms and Lilliefors tests to screen for deviations from the normal distribution in pair-wise differences. In cases of non-normal distribution, we used the histogram and the sample size adjusted z-score (termed moving criterion)(Van Selst & Jolicoeur, 1994) to screen for outliers. For our sample size, the z-score cut-off of 2.39 standard deviations corresponds to excluding the extreme 1% of differences (0.5% on each tail). We temporarily excluded the outlier. If the data then met normality assumptions, we re-included the outlier and continued with the process to determine agreement coefficients. Possible causes for the presence and degree of extreme differences are discussed later. Outliers were included in all following calculations of agreement in order to generate conservative rather than optimistic limits of agreement. When single outliers were not responsible for deviations from the normal distribution, skewness was treated by log transformation of measurements. If differences between log-transformed measurements were still not normally distributed, no further analysis would be undertaken, and that dataset was reported as technically inappropriate.

Limits of agreement were defined as mean difference ± 2SD, and the coefficient of agreement as 2SD in cases of non-log transformed data. For log-transformed data, the equivalent antilogs of mean difference ± 2SD give a proportional result, which we multiplied by the mean measurement to inform the limits of agreement. As with non-log transformed data, the coefficient of agreement is half the difference between the upper and lower limits of agreement and equivalent to 2 SD(Bland & Altman, 1986).

The larger the sample size used for estimation of limits of agreement, the more precise the estimates. To account for precision in estimates, we calculated 95% CI for the upper and lower limits of agreement. We then expanded the concept of coefficient of agreement to include the 95% CIs and reported that as 95% CI of agreement (95% CIA, figure 2). The smaller the 95% CIA, the smaller the differences between measurements (i.e. higher agreement between microphones).

Chapter 6

General discussion

Multiple Sclerosis is a common cause of disability in young adults. While current pharmacological treatments are highly effective in slowing disease progression, more personalized and dynamic monitoring is needed. Objective measurements of function offer various advantages for frequent assessment. The overarching aim of this study was to describe the concurrent validity of acoustic metrics to represent the overall neurological status in MS.

This chapter summarises the main findings and limitations of chapters 2-5, highlights the clinical implications or contributions of the work and discusses potential future directions that could be undertaken as a result of these findings.

Speech research in MS focus on single features and healthy subjects

In Chapter 2, we found that most studies compare speech between pwMS and HC, including three objective speech models to differentiate pwMS from HC (Hartelius et al., 1997; Konstantopoulos et al., 2010; J. Rusz, Benova, et al., 2018). Only two studies tested for associations between levels of speech impairment and severity of general neurological disability, one using an overall dysarthria score from perceptual analysis (Hartelius, Runmarker, & Andersen, 2000), the other using ten individually tested acoustic metrics (J. Rusz, Benova, et al., 2018). Few studies analysed associations and confounders between dysarthria and cognition. For the validation of objective speech metrics as markers of MS progression, the main gaps identified were the lack of neuroimaging studies (other than case reports), lack of longitudinal studies and lack of correlational studies between general neurological status and a multi-domain acoustic metric. Additionally, the findings by Rusz et al. (J. Rusz, Benova, et al.,

2018) warranted replication and extension, particularly on the relationship of subclinical speech impairment and general disability. After the publication of that systematic review, the same authors published an analysis of associations between speech metrics and MRI metrics (J. Rusz et al., 2019), which is briefly discussed below.

Speech impairment occurs across the spectrum of MS neurological status

Impairment of speech execution only becomes clinically relevant at advanced stages of the disease, *a priori* limiting its use as a marker of disease progression. Previous studies suggested that acoustic analysis might be more sensitive to impairment of speech than perceptual analysis, which could expand the concurrent validity of speech metrics to earlier stages of MS. The hypothesis that objective speech metrics are progressively affected across all MS disease stages was supported by differences in the acoustic composite score, but its interpretation requires a few considerations. In chapter 3, the level of speech impairment correlated moderately with general neurological disability. Additionally, an acoustic composite score differentiated between the three disability subgroups, with progressive scores from MS mild to MS severe, and correlated with MRI-measured white matter volume and total lesion load. Conversely, single-feature speech impairment affected almost exclusively the MS severe subgroup regardless of the method applied (perceptual or acoustic). That points to a probable maximum sensitivity of the tested speech metrics to gauge disease status between the moderate and severe subgroups (EDSS 3-8).

Our findings also support a distinction between disease diagnostic versus disease monitoring speech assessments in MS. Only acoustic monopitch differentiated each MS subgroup from matched HC (including MS mild), confirming monopitch's previously observed sensitivity to disease presence (Konstantopoulos et al., 2010). Acoustic monopitch did not, however, correlate with EDSS scores, confirming another recent observation in a different cohort (J. Rusz, Benova, et al., 2018). Conversely, the acoustic composite score differentiated between levels of disability despite not differing statistically between MS mild and HC. These

results suggest that identifying speech differences against HC might be irrelevant to prospective monitoring models, which would require alternative strategies for selecting speech variables.

Pause percentage during the self-generated monologue was the only other acoustic metric that differentiated MS moderate and HC. Differently from monopitch, pause percentage correlated with EDSS scores and MRI metrics. Pause percentage also tended to differentiate between the three severity groups and presented the largest weight (main independent variable) in the acoustic composite score. The relationship between increased speech pauses and disease status had not been previously reported in MS. However, increase in speech pauses is not specific to MS, featuring in and correlating with severity of common conditions such as depression (Mundt et al., 2012), Parkinson's Disease (J. Rusz, Hlavnicka, et al., 2018), sleep deprivation and the circadian cycle (Vogel et al., 2010). That lack of specificity to disease applies to the other tested speech features in different degrees, which highlights the challenge of using single-feature or short composite models to assist in disease diagnosis. Conversely, for tracking disease status, the adoption of controlling and adjusting procedures might suffice, such as screening out or statistically accounting for concurrent conditions (e.g., sleep disorder) that affect each speech metric, counterbalancing cyclical variations in speech by considering several measurements across multiple days or determining a threshold for change based on previously established statistics on the stability of metrics (Vogel et al., 2011).

Acoustic metrics correlate with disability in the absence of dysarthria

Still, in chapter 3, we applied a dedicated and extensive perceptual speech assessment where both pwMS and HC were randomly and blindly rated by two experienced raters, including the author. For conservative inference, people with signs of dysarthria were excluded from the non-dysarthric group. The main novel contribution of the present findings lies in correlations observed between acoustic speech metrics and EDSS scores in that exclusively non-dysarthric group. Rusz et al. recently found that a different four-variable acoustic model

(including monopitch) differentiated between HC and pwMS with EDSS < 2.0 (i.e., people without any disability, which by definition includes people with normal speech or with only signs of dysarthria)(J. Rusz, Benova, et al., 2018). Differences in outcome variables (MS presence vs disability level), tested predictors and modelling technique (linear regression vs support vector machine) might explain the diversion between studies' models. Nonetheless, both studies point to the presence of subclinical speech impairment which increases in severity along with general neurological dysfunction independently of dysarthria becoming overt.

We did not, however, observe correlations between speech and neuroimaging metrics in the non-dysarthric group. That lack of correlation might be due to the chain of weak correlations involved in the modelling and testing of the acoustic composite score, i.e., between standard neuroimaging and EDSS (the clinico-radiological paradox) and then between EDSS (outcome variable for modelling) and subclinical speech impairment. For the lack of correlation with MRI metrics, the current hypothesis was only partially or weakly supported.

The major limitation of this study relates to its cross-sectional design. Most neurological signs included in the EDSS, neuromuscular reflexes or visual acuity, for example, are not significantly influenced by transitory states like mood, or by personal characteristics, like extroversion or literacy, whereas the opposite occurs with speech metrics such as reading rate or pitch variation. In the current study, we controlled for common personal confounders such as age and gender and used adjusted statistical tests of significance and confidence intervals to estimate the contribution of neurological status to variation in speech in comparison to intra-disability level speech variation. Nonetheless, other unaccounted factors may have contributed substantially to the observed variation in speech metrics. Arguably, a cross-sectional model containing neurological signs gauges (almost) only disease status, whereas a cross-sectional speech model does not. In that sense, different weights and possibly different speech features are likely to integrate a speech model of disease progression, as measured longitudinally.

Our findings highlight the potential of speech analysis, unexplored in current disease progression protocols and largely overseen by speech-related research. Current results will inform longitudinal studies that people with mild or moderate disability, and people without dysarthria, are also likely to benefit from speech monitoring. Our results also introduce the validity of metrics from unscripted speech in MS, the concept of acoustic composite scores to better reflect general neurological status, and the distinction between disease diagnostic versus disease monitoring models of speech.

An ataxia acoustic score is not ataxia-specific

The hypothesis that an ataxia-modelled acoustic composite score correlates with cerebellar status as assessed by different functional and imaging tests was supported, but the ataxia-acoustic model was not specific to ataxia or cerebellar impairment. In chapter 4, the ataxia acoustic composite score correlated with all cerebellar functional assessments other than SARA (on which it was modelled) and differentiated between normal and abnormal 9HPT with 85% accuracy. That acoustic composite score also correlated with MRI cerebellar white matter volume and lesion load, while pwMS without cerebellar lesions did significantly better in that experimental acoustic score than people with detected cerebellar lesions. Thus, its concurrent validity was supported. Yet, the components in the ataxia composite score were the same as those in the general disability composite score described in chapter 3, added only by loudness instability. Despite a distinct weighting of speech features in each model (DDK rate being the most prominent for ataxia), the overlapping between models does not support the argument of specificity for ataxia. Noteworthy, a high co-variation between SARA and EDSS was also observed, the same occurred. Characteristics of MS pathology may not be conducive of modelling specific sub-systems. In the study conducted by Rusz et al.(J. Rusz, Benova, et al., 2018), for example, none of 141 PwMS presented only cerebellar disability. As summarized in chapter 1, after a few years of disease duration, plaques are rarely restricted to a small specific

site (e.g., cerebellum) and the inflammatory infiltrate becomes diffuse in the CNS of pwMS, possibly causing subtle progressive impairment in multiple neurological domains.

The main limitation in this predictive modelling study is the lack of acoustic metrics that might be key in ataxic speech. “Distorted vowels” is a typical example, where the expected vowel sound within a word or phrase is elongated, shortened, added or substituted by a different sound (frequently another vowel or semi-vowel) due to poor articulatory control. Thus, detecting distorted vowels requires analysis of language to first determine the expected sound. “Imprecise consonants” is another example that would require language analysis and manual extraction, in contrast to the automated method employed. As these perceptual features correlated with SARA and EDSS, the Inclusion of their acoustic counterparts as potential predictors in the acoustic model could have increased its predictive accuracy. Another limitation, common to both general and ataxia modelling, relates to only multiple linear regression being used for determining speech models, whereas other methods (e.g., SVM, trees, Bayesian hierarchical) might yield better predictive performance given that speech features change at different stages of the disease (non-linear change).

The present results highlight the supplementary value of various speech features to the assessment of cerebellar impairment in MS, particularly in protocols applying SARA and 9HPT. Short speech batteries could be recorded remotely, between clinical assessments, to better track longitudinal trajectories.

Most speech features are not robust to variation in recording equipment

In Chapter 5, we determined the agreement intervals between professional-grade recording equipment and iPods with three microphone configurations for 20 speech metrics and compared those intervals to previously observed effect sizes of diseases on speech. Most speech metrics were not robust to changes caused by variations in recordings. That confirms and adds a “clinical importance perspective” to previous findings that acoustic metrics from professional-grade microphones are significantly different from consumer-grade equipment

(Titze & Winholtz, 1993; Vogel & Maruff, 2008; A. P. Vogel et al., 2014). We explored frequency distribution and noise as potential mediators for the disagreement in metrics between microphones. These mediators are commonly affected by microphone quality and type, frequency responsiveness, distance to sound source, and filtering. Despite the divergence in acoustic values, we observed a high correlation between metrics from different microphones and an exact equivalence between microphones for rank correlations between speech scores and EDSS. While current and previous findings strongly suggest against the interchangeable use of recording methods (microphone and distance from the sound source), they do not preclude the use of mobile devices for health-related speech data acquisition. In fact, recent studies using constant recording equipment found that speech recorded by smartphones and tablets produced acoustic metrics capable of differentiating HC from people with Parkinson's Disease (J. Rusz, Hlavnicka, et al., 2018) and with mild traumatic brain injury (Daudet et al., 2017). Yet, the strength in the relationship between recording acoustics by mobile devices and disease status or progression warrants future investigation.

The small number of participants was the main weakness in the experiment described in chapter 5, which likely inflated confidence intervals of agreement between recording equipment. Another limitation relates to testing of iPods but not smartphones, the latter more common and thus more likely to be used as mobile recorders and as speech analysis processors (higher computational power). Nevertheless, large differences might also exist between the uncountable number of smartphone models, which is compounded by the accelerated turnover in mobile technology. Rather than gadget-specific, the findings in chapter 5 serve as a cautionary confirmation of previous observations about the interchangeability of recording devices (or lack thereof) for health-related acoustic analysis. Moreover, the validity of recordings by mobile devices to measure disability level when the same device is used warrants dedicated investigation.

The present findings will inform researchers on acoustic features that are most robust to variation in recording protocols, which is particularly relevant to real-world applications where such variations might not be rare.

6.1. Future research directions

Current findings would be substantially complemented by the following:

1) Objective speech metrics must be studied in their longitudinal trajectories in people with MS to determine their monitoring and predictive characteristics in relation to disease progression.
2) Speech models of disease progression should be refined through investigation of other (non-linear) modelling methods, as well integration of other speech metrics and latent speech features (e.g., through deep machine learning).
3) Assessment of MS status or progression involves measuring performance in various smaller domains through individual tasks (items). To be valid, each item must add significantly to the whole assessment, i.e., not be redundant. Thus, for further integration of speech into monitoring protocols, speech metrics should be tested for their unique contribution to multi-domain assessments.
4) Speech related to ataxia or other neurological sub-systems will benefit from a stratified sampling to recruit a high proportion of people with cerebellar-only or predominantly cerebellar impairment. Alternatively, speech models determined in purely cerebellar diseases should be validated in MS.
5) Imaging metrics that are more representative of disfunction, such as composite imaging scores and advanced imaging techniques, should be used to investigate anatomo-functional correlates of speech impairment in MS.

6.2. Conclusion

Objective measurement of speech production through acoustic analysis can detect speech impairment in all stages of MS disability. Well established speech features that are sensitive to MS presence are not necessarily sensitive to disability level. Experimental speech models for general and cerebellar disability were also valid for MRI metrics, particularly white matter volume, for self-assessed quality of life and for hand dexterity test (cerebellar model). The relationship between speech metrics and general disability also exists in people without perceivable dysarthria (progressive subclinical speech impairment). No evidence was found for the specificity of an ataxia speech model o for the exchangeability of recording devices for health-related acoustic analysis. Consistencies and limitations of this study should inform future longitudinal studies aimed at confirming the potential of speech measurement to assist in monitoring MS progression.

References

Ackermann, H., Mathiak, K., & Ivry, R. B. (2004). Temporal organization of "internal speech" as a basis for cerebellar modulation of cognitive functions. *Behav Cogn Neurosci Rev, 3*(1), 14-22. doi:10.1177/1534582304263251

Ahmad, H., Palmer, A. J., Campbell, J. A., van der Mei, I., & Taylor, B. V. (2018). Health economic impact of multiple sclerosis in Australia in 2017.

Alpert, M., Pouget, E. R., & Silva, R. R. (2001). Reflections of depression in acoustic measures of the patient's speech. *J Affect Disord, 66*(1), 59-69. doi:10.1016/s0165-0327(00)00335-9

Alusi, S., Aziz, T., Glickman, S., Jahanshahi, M., Stein, J., & Bain, P. (2001). Stereotactic lesional surgery for the treatment of tremor in multiple sclerosis. *Brain, 124*(8), 1576-1589.

Amato, M. P., & Ponziani, G. (1999). Quantification of impairment in MS: discussion of the scales in use. *Mult Scler, 5*(4), 216-219. doi:10.1177/135245859900500404

Arnett, P. A., Smith, M. M., Barwick, F. H., Benedict, R. H., & Ahlstrom, B. P. (2008). Oralmotor slowing in multiple sclerosis: relationship to neuropsychological tasks requiring an oral response. *J Int Neuropsychol Soc, 14*(3), 454-462. doi:10.1017/s1355617708080508

Ascherio, A., & Munger, K. L. (2010). Epstein-barr virus infection and multiple sclerosis: a review. *J Neuroimmune Pharmacol, 5*(3), 271-277. doi:10.1007/s11481-010-9201-3

Barkhof, F. (1999). MRI in multiple sclerosis: correlation with expanded disability status scale (EDSS). *Mult Scler, 5*(4), 283-286. doi:10.1177/135245859900500415

Barkhof, F., Filippi, M., Miller, D. H., Scheltens, P., Campi, A., Polman, C. H., . . . Valk, J. (1997). Comparison of MRI criteria at first presentation to predict conversion to clinically definite multiple sclerosis. *Brain, 120 (Pt 11)*, 2059-2069. doi:10.1093/brain/120.11.2059

Bauer, V., Aleric, Z., & Jancic, E. (2015). Comparing voice self-assessment with auditory perceptual analysis in patients with multiple sclerosis. *Int Arch Otorhinolaryngol, 19*(2), 100-105. doi:10.1055/s-0034-1397332

Bauer, V., Aleric, Z., Jancic, E., Knezevic, B., Prpic, D., & Kacavenda, A. (2013). Subjective and perceptual analysis of voice quality and relationship with neurological disfunction in multiple sclerosis patients. *Clin Neurol Neurosurg, 115 Suppl 1*, S17-20. doi:10.1016/j.clineuro.2013.09.015

Baylor, C., Yorkston, K., Bamer, A., Britton, D., & Amtmann, D. (2010). Variables associated with communicative participation in people with multiple sclerosis: a regression analysis. *Am J Speech Lang Pathol, 19*(2), 143-153. doi:10.1044/1058-0360(2009/08-0087)

Ben-Ami, E., Miller, A., & Berrih-Aknin, S. (2014). T cells from autoimmune patients display reduced sensitivity to immunoregulation by mesenchymal stem cells: role of IL-2. *Autoimmun Rev, 13*(2), 187-196. doi:10.1016/j.autrev.2013.09.007

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann Stat, 29*(4), 1165-1188.

Benninger, M. S., Ahuja, A. S., Gardner, G., & Grywalski, C. (1998). Assessing outcomes for dysphonic patients. *J Voice, 12*(4), 540-550.

Beukelman, D. R., Kraft, G. H., & Freal, J. (1985). Expressive communication disorders in persons with multiple sclerosis: a survey. *Arch Phys Med Rehabil, 66*(10), 675-677.

Bielekova, B., & Martin, R. (2004). Development of biomarkers in multiple sclerosis. *Brain, 127*(Pt 7), 1463-1478. doi:10.1093/brain/awh176

Blanco, Y., Compta, Y., Graus, F., & Saiz, A. (2008). Midbrain lesions and paroxysmal dysarthria in multiple sclerosis. *Mult Scler, 14*(5), 694-697. doi:10.1177/1352458507087846

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 1*(8476), 307-310.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot international, 5*.

Boonstra, F., Perera, T., Noffs, G., Marotta, C., Vogel, A. P., Evans, A., . . . Kolbe, S. (2018). Novel functional MRI task for studying the neural correlates of upper limb tremor. *Front Neurol, 9*, 513.

Borrie, S. A., Baese-Berk, M., Van Engen, K., & Bent, T. (2017). A relationship between processing speech in noise and dysarthric speech. *J Acoust Soc Am, 141*(6), 4660. doi:10.1121/1.4986746

Brinca, L. F., Batista, A. P., Tavares, A. I., Gonçalves, I. C., & Moreno, M. L. (2014). Use of cepstral analyses for differentiating normal from dysphonic voices: a comparative study of connected speech versus sustained vowel in European Portuguese female speakers. *J Voice, 28*(3), 282-286. doi:10.1016/j.jvoice.2013.10.001

Brown, J. W. L., Coles, A., Horakova, D., Havrdova, E., Izquierdo, G., Prat, A., . . . Robertson, N. (2019). Association of Initial Disease-Modifying Therapy With Later Conversion to Secondary Progressive Multiple Sclerosis. *Jama, 321*(2), 175-187. doi:10.1001/jama.2018.20588

Brownlee, W. J., Hardy, T. A., Fazekas, F., & Miller, D. H. (2017). Diagnosis of multiple sclerosis: progress and challenges. *Lancet, 389*(10076), 1336-1346. doi:10.1016/s0140-6736(16)30959-x

Bürk, K., Mälzig, U., Wolf, S., Heck, S., Dimitriadis, K., Schmitz-Hübsch, T., . . . Schulz, J. B. (2009). Comparison of three clinical rating scales in Friedreich ataxia (FRDA). *Mov Disord, 24*(12), 1779-1784. doi:10.1002/mds.22660

Buzzard, K. A., Broadley, S. A., & Butzkueven, H. (2012). What do effective treatments for multiple sclerosis tell us about the molecular mechanisms involved in pathogenesis? *Int J Mol Sci, 13*(10), 12665-12709. doi:10.3390/ijms131012665

Calvo-Barreiro, L., Eixarch, H., Montalban, X., & Espejo, C. (2018). Combined therapies to treat complex diseases: The role of the gut microbiota in multiple sclerosis. *Autoimmun Rev, 17*(2), 165-174. doi:10.1016/j.autrev.2017.11.019

Campbell, G. R., Ziabreva, I., Reeve, A. K., Krishnan, K. J., Reynolds, R., Howell, O., . . . Mahad, D. J. (2011). Mitochondrial DNA deletions and neurodegeneration in multiple sclerosis. *Ann Neurol, 69*(3), 481-492. doi:10.1002/ana.22109

Cannizzaro, M., Harel, B., Reilly, N., Chappell, P., & Snyder, P. J. (2004). Voice acoustical measurement of the severity of major depression. *Brain Cogn, 56*(1), 30-35. doi:10.1016/j.bandc.2004.05.003

Chan, J. C. S., Stout, J. C., & Vogel, A. P. (2019). Speech in prodromal and symptomatic Huntington's disease as a model of measuring onset and progression in dominantly inherited neurodegenerative diseases. *Neurosci Biobehav Rev, 107*, 450-460. doi:10.1016/j.neubiorev.2019.08.009

Charcot, J. M. (1877). Leçons sur les maladies du système nerveux: faites à la Salpêtrière . Tome 1.

Chen, A. Y., Chonghasawat, A. O., & Leadholm, K. L. (2017). Multiple sclerosis: Frequency, cost, and economic burden in the United States. *J Clin Neurosci, 45*, 180-186. doi:10.1016/j.jocn.2017.06.005

Chiara, T., Martin, D., & Sapienza, C. (2007). Expiratory muscle strength training: speech production outcomes in patients with multiple sclerosis. *Neurorehabilitation and neural repair*.

Chiara, T., Martin, D., & Sapienza, C. (2007). Expiratory muscle strength training: speech production outcomes in patients with multiple sclerosis. *Neurorehabil Neural Repair, 21*(3), 239-249. doi:10.1177/1545968306294737

Clifford, A., & Reiss, J. (2011). *Proximity effect detection for directional microphones.* Paper presented at the Audio Engineering Society Convention 131.

Cohen, J. A., Cutter, G. R., Fischer, J. S., Goodman, A. D., Heidenreich, F. R., Kooijmans, M. F., . . . Whitaker, J. N. (2002). Benefit of interferon β-1a on MSFC progression in secondary progressive MS. *Neurology, 59*(5), 679-687. doi:10.1212/wnl.59.5.679

Comi, G., Radaelli, M., & Soelberg Sørensen, P. (2017). Evolving concepts in the treatment of relapsing multiple sclerosis. *Lancet, 389*(10076), 1347-1356. doi:10.1016/s0140-6736(16)32388-1

Corben, L. A., Tai, G., Wilson, C., Collins, V., Churchyard, A. J., & Delatycki, M. B. (2010). A comparison of three measures of upper limb function in Friedreich ataxia. *J Neurol, 257*(4), 518-523. doi:10.1007/s00415-009-5352-7

Cristofanilli, M., Rosenthal, H., Cymring, B., Gratch, D., Pagano, B., Xie, B., & Sadiq, S. A. (2014). Progressive multiple sclerosis cerebrospinal fluid induces inflammatory demyelination, axonal loss, and astrogliosis in mice. *Exp Neurol, 261*, 620-632. doi:10.1016/j.expneurol.2014.07.020

Cummins, N., Sethu, V., Epps, J., Schnieder, S., & Krajewski, J. (2015). Analysis of acoustic space variability in speech affected by depression. *Speech Commun, 75*, 27-49.

Cutter, G. R., Baier, M. L., Rudick, R. A., Cookfair, D. L., Fischer, J. S., Petkau, J., . . . Willoughby, E. (1999). Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain, 122 (Pt 5)*, 871-882. doi:10.1093/brain/122.5.871

Danesh-Sani, S. A., Rahimdoost, A., Soltani, M., Ghiyasi, M., Haghdoost, N., & Sabzali-Zanjankhah, S. (2013). Clinical assessment of orofacial manifestations in 500 patients with multiple sclerosis. *J Oral Maxillofac Surg, 71*(2), 290-294. doi:10.1016/j.joms.2012.05.008

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *J Speech Hear Res, 12*(2), 246-269. doi:10.1044/jshr.1202.246

Darley, F. L., Brown, J. R., & Goldstein, N. P. (1972). Dysarthria in multiple sclerosis. *J Speech Hear Res, 15*(2), 229-245. doi:10.1044/jshr.1502.229

Daudet, L., Yadav, N., Perez, M., Poellabauer, C., Schneider, S., & Huebner, A. (2017). Portable mTBI Assessment Using Temporal and Frequency Analysis of Speech. *IEEE J Biomed Health Inform, 21*(2), 496-506. doi:10.1109/jbhi.2016.2633509

De Bodt, M. S., Wuyts, F. L., Van de Heyning, P. H., & Croux, C. (1997). Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual rating of voice quality. *J Voice, 11*(1), 74-80. doi:10.1016/s0892-1997(97)80026-4

De Looze, C., Moreau, N., Renie, L., Kelly, F., Ghio, A., Rico, A., . . . Petrone, C. (2017). Effects of cognitive impairment on prosodic parameters of speech production planning in multiple sclerosis. *J Neuropsychol*. doi:10.1111/jnp.12127

Deliyski, D. D., Shaw, H. S., & Evans, M. K. (2005). Adverse effects of environmental noise on acoustic voice quality measurements. *J Voice, 19*(1), 15-28. doi:10.1016/j.jvoice.2004.07.003

Dendrou, C. A., Fugger, L., & Friese, M. A. (2015). Immunopathology of multiple sclerosis. *Nat Rev Immunol, 15*(9), 545-558. doi:10.1038/nri3871

Di Stadio, A., Bernitsas, E., Restivo, D. A., Alfonsi, E., & Marchese-Ragona, R. (2018). Spasmodic Dysphonia in Multiple Sclerosis Treatment With Botulin Toxin A: A Pilot Study. *J Voice*. doi:10.1016/j.jvoice.2018.01.002

Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E., & Ramnani, N. (2009). A probabilistic MR atlas of the human cerebellum. *Neuroimage, 46*(1), 39-46. doi:10.1016/j.neuroimage.2009.01.045

Dogan, M., Midi, I., Yazici, M. A., Kocak, I., Günal, D., & Sehitoglu, M. A. (2007). Objective and subjective evaluation of voice quality in multiple sclerosis. *J Voice, 21*(6), 735-740. doi:10.1016/j.jvoice.2006.05.006

Durisko, C., & Fiez, J. A. (2010). Functional activation in the cerebellum during working memory and simple speech tasks. *Cortex, 46*(7), 896-906. doi:10.1016/j.cortex.2009.09.009

Eargle, J. M. (2003). Environmental Effects and Departures from Ideal Performance. In *Handbook of Recording Engineering* (pp. 65-73). Boston, MA: Springer US.

Fagnani, C., Neale, M. C., Nisticò, L., Stazi, M. A., Ricigliano, V. A., Buscarinu, M. C., . . . Ristori, G. (2015). Twin studies in multiple sclerosis: A meta-estimation of heritability and environmentality. *Mult Scler, 21*(11), 1404-1413. doi:10.1177/1352458514564492

Fazeli, M., Moradi, N., Soltani, M., Naderifar, E., Majdinasab, N., Latifi, S. M., & Dastoorpour, M. (2020). Dysphonia Characteristics and Vowel Impairment in Relation to Neurological Status in Patients with Multiple Sclerosis. *J Voice, 34*(3), 364-370. doi:10.1016/j.jvoice.2018.09.018

Feigin, J. A., Barlow, N. L., & Stelmachowicz, P. G. (1990). The effect of reference microphone placement on sound pressure levels at an ear level hearing aid microphone. *Ear and hearing, 11*(5), 321-326. doi:10.1097/00003446-199010000-00001

Feijó, A. V., Parente, M. A., Behlau, M., Haussen, S., de Veccino, M. C., & Martignago, B. C. (2004). Acoustic analysis of voice in multiple sclerosis patients. *J Voice, 18*(3), 341-347. doi:10.1016/j.jvoice.2003.05.004

Feys, P., Lamers, I., Francis, G., Benedict, R., Phillips, G., LaRocca, N., . . . Rudick, R. (2017). The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. *Mult Scler, 23*(5), 711-720. doi:10.1177/1352458517690824

Filippi, M., Bar-Or, A., Piehl, F., Preziosa, P., Solari, A., Vukusic, S., & Rocca, M. A. (2018). Multiple sclerosis. *Nat Rev Dis Primers, 4*(1), 43. doi:10.1038/s41572-018-0041-4

Filippi, M., Rocca, M. A., Barkhof, F., Brück, W., Chen, J. T., Comi, G., . . . Lassmann, H. (2012). Association between pathological and MRI findings in multiple sclerosis. *Lancet Neurol, 11*(4), 349-360. doi:10.1016/s1474-4422(12)70003-0

Filippi, M., Rocca, M. A., De Stefano, N., Enzinger, C., Fisher, E., Horsfield, M. A., . . . Comi, G. (2011). Magnetic resonance techniques in multiple sclerosis: the present and the future. *Arch Neurol, 68*(12), 1514-1520. doi:10.1001/archneurol.2011.914

Fischer, M. T., Wimmer, I., Höftberger, R., Gerlach, S., Haider, L., Zrzavy, T., . . . Lassmann, H. (2013). Disease-specific molecular events in cortical multiple sclerosis lesions. *Brain, 136*(Pt 6), 1799-1815. doi:10.1093/brain/awt110

Fisher, E., Rudick, R. A., Cutter, G., Baier, M., Miller, D., Weinstock-Guttman, B., . . . Simonian, N. A. (2000). Relationship between brain atrophy and disability: an 8-year follow-up study of multiple sclerosis patients. *Mult Scler, 6*(6), 373-377. doi:10.1177/135245850000600602

Fisniku, L. K., Brex, P. A., Altmann, D. R., Miszkiel, K. A., Benton, C. E., Lanyon, R., . . . Miller, D. H. (2008). Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain, 131*(3), 808-817. doi:10.1093/brain/awm329

Fletcher, H., & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *Bell Syst Tech J, 12*(4), 377-430.

Flint, A. J., Black, S. E., Campbell-Taylor, I., Gailey, G. F., & Levinton, C. (1992). Acoustic analysis in the differentiation of Parkinson's disease and major depression. *J Psycholinguist Res, 21*(5), 383-389. doi:10.1007/bf01067922

Folker, J. E., Murdoch, B. E., Rosen, K. M., Cahill, L. M., Delatycki, M. B., Corben, L. A., & Vogel, A. P. (2012). Differentiating profiles of speech impairments in Friedreich's ataxia: a perceptual and instrumental approach. *Int J Lang Commun Disord, 47*(1), 65-76. doi:10.1111/j.1460-6984.2011.00078.x

Fornito, A., Zalesky, A., & Breakspear, M. (2015). The connectomics of brain disorders. *Nat Rev Neurosci, 16*(3), 159-172. doi:10.1038/nrn3901

Fraussen, J., de Bock, L., & Somers, V. (2016). B cells and antibodies in progressive multiple sclerosis: Contribution to neurodegeneration and progression. *Autoimmun Rev, 15*(9), 896-899. doi:10.1016/j.autrev.2016.07.008

Friedova, L., Rusz, J., Motyl, J., Srpova, B., Vodehnalova, K., Andelova, M., . . . Uher, T. (2019). Slowed articulation rate is associated with information processing speed decline in multiple sclerosis: A pilot study. *J Clin Neurosci, 65*, 28-33. doi:10.1016/j.jocn.2019.04.018

Gajofatto, A., Calabrese, M., Benedetti, M. D., & Monaco, S. (2013). Clinical, MRI, and CSF markers of disability progression in multiple sclerosis. *Dis Markers, 35*(6), 687-699. doi:10.1155/2013/484959

Gamboa, J., Jiménez-Jiménez, F. J., Nieto, A., Montojo, J., Ortí-Pareja, M., Molina, J. A., . . . Cobeta, I. (1997). Acoustic voice analysis in patients with Parkinson's disease treated with dopaminergic drugs. *J Voice, 11*(3), 314-320. doi:10.1016/s0892-1997(97)80010-0

Gerald, F. J. F., Murdoch, B. E., & Chenery, H. J. (1987). Multiple sclerosis: Associated speech and language disorders. *Aust J Hum Commun Disord, 15*(2), 15-35.

Geurts, J. J., Bö, L., Pouwels, P. J., Castelijns, J. A., Polman, C. H., & Barkhof, F. (2005). Cortical lesions in multiple sclerosis: combined postmortem MR imaging and histopathology. *AJNR Am J Neuroradiol, 26*(3), 572-577.

Geurts, J. J., Pouwels, P. J., Uitdehaag, B. M., Polman, C. H., Barkhof, F., & Castelijns, J. A. (2005). Intracortical lesions in multiple sclerosis: improved detection with 3D double inversion-recovery MR imaging. *Radiology, 236*(1), 254-260. doi:10.1148/radiol.2361040450

Giannetti, P., Politis, M., Su, P., Turkheimer, F. E., Malik, O., Keihaninejad, S., . . . Piccini, P. (2015). Increased PK11195-PET binding in normal-appearing white matter in clinically isolated syndrome. *Brain, 138*(Pt 1), 110-119. doi:10.1093/brain/awu331

Goodkin, D. E., Cookfair, D., Wende, K., Bourdette, D., Pullicino, P., Scherokman, B., & Whitham, R. (1992). Inter- and intrarater scoring agreement using grades 1.0 to 3.5 of the Kurtzke Expanded Disability Status Scale (EDSS). *Neurology, 42*(4), 859-859. doi:10.1212/wnl.42.4.859

Goodwin, S. J., & Carpenter, A. F. (2016). Successful treatment of paroxysmal ataxia and dysarthria in multiple sclerosis with levetiracetam. *Mult Scler Relat Disord, 10*, 79-81. doi:10.1016/j.msard.2016.09.003

Gronwall, D. M. (1977). Paced auditory serial-addition task: a measure of recovery from concussion. *Percept Mot Skills, 44*(2), 367-373. doi:10.2466/pms.1977.44.2.367

Gross, R. H., & Corboy, J. R. (2019). Monitoring, Switching, and Stopping Multiple Sclerosis Disease-Modifying Therapies. *Continuum (Minneap Minn), 25*(3), 715-735. doi:10.1212/con.0000000000000738

Gustavsson, A., Svensson, M., Jacobi, F., Allgulander, C., Alonso, J., Beghi, E., . . . Olesen, J. (2011). Cost of disorders of the brain in Europe 2010. *Eur Neuropsychopharmacol, 21*(10), 718-779. doi:10.1016/j.euroneuro.2011.08.008

Hackmack, K., Weygandt, M., Wuerfel, J., Pfueller, C. F., Bellmann-Strobl, J., Paul, F., & Haynes, J. D. (2012). Can we overcome the 'clinico-radiological paradox' in multiple sclerosis? *J Neurol, 259*(10), 2151-2160. doi:10.1007/s00415-012-6475-9

Haider, L., Zrzavy, T., Hametner, S., Höftberger, R., Bagnato, F., Grabner, G., . . . Lassmann, H. (2016). The topograpy of demyelination and neurodegeneration in the multiple sclerosis brain. *Brain, 139*(Pt 3), 807-815. doi:10.1093/brain/awv398

Hamilton, F., Rochester, L., Paul, L., Rafferty, D., O'Leary, C. P., & Evans, J. J. (2009). Walking and talking: an investigation of cognitive-motor dual tasking in multiple sclerosis. *Mult Scler, 15*(10), 1215-1227. doi:10.1177/1352458509106712

Hartelius, L., Buder, E. H., & Strand, E. A. (1997). Long-term phonatory instability in individuals with multiple sclerosis. *J Speech Lang Hear Res, 40*(5), 1056-1072. doi:10.1044/jslhr.4005.1056

Hartelius, L., & Lillvik, M. (2003). Lip and tongue function differently affected in individuals with multiple sclerosis. *Folia Phoniatr Logop, 55*(1), 1-9. doi:10.1159/000068057

Hartelius, L., Nord, L., & Buder, E. H. (1995). Acoustic analysis of dysarthria associated with multiple sclerosis. *Clin Linguist Phon, 9*(2), 95-120.

Hartelius, L., Runmarker, B., & Andersen, O. (2000). Prevalence and characteristics of dysarthria in a multiple-sclerosis incidence cohort: relation to neurological data. *Folia Phoniatr Logop, 52*(4), 160-177. doi:10.1159/000021531

Hartelius, L., Runmarker, B., Andersen, O., & Nord, L. (2000). Temporal speech characteristics of individuals with multiple sclerosis and ataxic dysarthria:'Scanning speech'revisited. *Folia Phoniatr Logop, 52*(5), 228-238.

Hartelius, L., & Svensson, P. (1994). Speech and swallowing symptoms associated with Parkinson's disease and multiple sclerosis: a survey. *Folia Phoniatr Logop, 46*(1), 9-17.

Hartelius, L., Svensson, P., & Bubach, A. (1993). Clinical assessment of dysarthria: Performance on a dysarthria test by normal adult subjects, and by individuals with Parkinson's disease or with multiple sclerosis. *Scand J Logop Phoniatr, 18*(4), 131-141. doi:10.3109/14015439309101359

Hartelius, L., Theodoros, D., Cahill, L., & Lillvik, M. (2003). Comparability of perceptual analysis of speech characteristics in Australian and Swedish speakers with multiple sclerosis. *Folia Phoniatr Logop, 55*(4), 177-188.

Heman-Ackah, Y. D., Heuer, R. J., Michael, D. D., Ostrowski, R., Horman, M., Baroody, M. M., . . . Sataloff, R. T. (2003). Cepstral peak prominence: a more reliable measure of dysphonia. *Ann Otol Rhinol Laryngol, 112*(4), 324-333. doi:10.1177/000348940311200406

Hernandez-Castillo, C. R., Galvez, V., Diaz, R., & Fernandez-Ruiz, J. (2016). Specific cerebellar and cortical degeneration correlates with ataxia severity in spinocerebellar ataxia type 7. *Brain Imaging Behav, 10*(1), 252-257. doi:10.1007/s11682-015-9389-1

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nat Rev Neurosci, 13*(2), 135-145. doi:10.1038/nrn3158

Hlavnička, J., Čmejla, R., Tykalová, T., Šonka, K., Růžička, E., & Rusz, J. (2017). Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder. *Sci Rep, 7*(1), 12. doi:10.1038/s41598-017-00047-5

Hobart, J., Freeman, J., & Thompson, A. (2000). Kurtzke scales revisited: the application of psychometric methods to clinical intuition. *Brain, 123 (Pt 5)*, 1027-1040. doi:10.1093/brain/123.5.1027

Hobart, J., Lamping, D., Fitzpatrick, R., Riazi, A., & Thompson, A. (2001). The Multiple Sclerosis Impact Scale (MSIS-29): a new patient-based outcome measure. *Brain, 124*(Pt 5), 962-973.

Hohol, M. J., Orav, E. J., & Weiner, H. L. (1999). Disease steps in multiple sclerosis: a longitudinal study comparing disease steps and EDSS to evaluate disease progression. *Mult Scler, 5*(5), 349-354. doi:10.1177/135245859900500508

Ikeda, K. M., Lee, D. H., Fraser, J. A., Mirsattari, S., & Morrow, S. A. (2015). Plasma Exchange in a Patient with Tumefactive, Corticosteroid-Resistant Multiple Sclerosis. *Int J MS Care, 17*(5), 231-235. doi:10.7224/1537-2073.2014-078

Ilg, W., Branscheidt, M., Butala, A., Celnik, P., de Paola, L., Horak, F. B., . . . Timmann, D. (2018). Consensus Paper: Neurophysiological Assessments of Ataxias in Daily Practice. *Cerebellum, 17*(5), 628-653. doi:10.1007/s12311-018-0937-2

Iorio, R., Capone, F., Plantone, D., & Batocchi, A. P. (2014). Paroxysmal ataxia and dysarthria in multiple sclerosis. *J Clin Neurosci, 21*(1), 174-175. doi:10.1016/j.jocn.2013.01.031

Jacobsen, C., Hagemeier, J., Myhr, K. M., Nyland, H., Lode, K., Bergsland, N., . . . Zivadinov, R. (2014). Brain atrophy and disability progression in multiple sclerosis patients: a 10-year follow-up study. *J Neurol Neurosurg Psychiatry, 85*(10), 1109-1115. doi:10.1136/jnnp-2013-306906

Johnen, A., Schiffler, P., Landmeyer, N. C., Tenberge, J. G., Riepl, E., Wiendl, H., . . . Meuth, S. G. (2019). Resolving the cognitive clinico-radiological paradox - Microstructural degeneration of fronto-striatal-thalamic loops in early active multiple sclerosis. *Cortex, 121*, 239-252. doi:10.1016/j.cortex.2019.08.022

Kalincik, T., Buzzard, K., Jokubaitis, V., Trojano, M., Duquette, P., Izquierdo, G., . . . Butzkueven, H. (2014). Risk of relapse phenotype recurrence in multiple sclerosis. *Mult Scler, 20*(11), 1511-1522. doi:10.1177/1352458514528762

Kalincik, T., Manouchehrinia, A., Sobisek, L., Jokubaitis, V., Spelman, T., Horakova, D., . . . Butzkueven, H. (2017). Towards personalized therapy for multiple sclerosis: prediction of individual treatment response. *Brain, 140*(9), 2426-2443. doi:10.1093/brain/awx185

Kappos, L., Bar-Or, A., Cree, B. A. C., Fox, R. J., Giovannoni, G., Gold, R., . . . Dahlke, F. (2018). Siponimod versus placebo in secondary progressive multiple sclerosis (EXPAND): a double-blind, randomised, phase 3 study. *Lancet, 391*(10127), 1263-1273. doi:10.1016/s0140-6736(18)30475-6

Kent, R. D., & Vorperian, H. K. (2018). Static measurements of vowel formant frequencies and bandwidths: A review. *Journal of Communication Disorders, 74*, 74-97. doi:10.1016/j.jcomdis.2018.05.004

Kent, R. D., Vorperian, H. K., & Duffy, J. R. (1999). Reliability of the Multi-Dimensional Voice Program for the analysis of voice samples of subjects with dysarthria. *Am J Speech Lang Pathol, 8*(2), 129-136.

Kishimoto, R., Yabe, I., Niino, M., Sato, K., Tsuji, S., Kikuchi, S., & Sasaki, H. (2008). Baló's concentric sclerosislike lesion in the brainstem of a multiple sclerosis patient. *J Neurol, 255*(5), 760-761. doi:10.1007/s00415-008-0795-9

Kister, I., Bacon, T. E., Chamot, E., Salter, A. R., Cutter, G. R., Kalina, J. T., & Herbert, J. (2013). Natural history of multiple sclerosis symptoms. *Int J MS Care, 15*(3), 146-158. doi:10.7224/1537-2073.2012-053

Klevan, G., Jacobsen, C. O., Aarseth, J. H., Myhr, K. M., Nyland, H., Glad, S., . . . Farbu, E. (2014). Health related quality of life in patients recently diagnosed with multiple sclerosis. *Acta Neurol Scand, 129*(1), 21-26. doi:10.1111/ane.12142

Klugman, T. M., & Ross, E. (2002). Perceptions of the impact of speech, language, swallowing, and hearing difficulties on quality of life of a group of South African persons with multiple sclerosis. *Folia Phoniatr Logop, 54*(4), 201-221. doi:10.1159/000063194

Konig, A., Linz, N., Zeghari, R., Klinge, X., Troger, J., Alexandersson, J., & Robert, P. (2019). Detecting Apathy in Older Adults with Cognitive Disorders Using Automatic Speech Analysis. *J Alzheimers Dis, 69*(4), 1183-1193. doi:10.3233/jad-181033

Konig, A., Satt, A., Sorin, A., Hoory, R., Derreumaux, A., David, R., & Robert, P. H. (2018). Use of Speech Analyses within a Mobile Application for the Assessment of Cognitive Impairment in Elderly People. *Curr Alzheimer Res, 15*(2), 120-129. doi:10.2174/1567205014666170829111942

Konstantopoulos, K., Vikelis, M., Seikel, J. A., & Mitsikostas, D. D. (2010). The existence of phonatory instability in multiple sclerosis: an acoustic and electroglottographic study. *Neurol Sci, 31*(3), 259-268. doi:10.1007/s10072-009-0170-3

Kunieda, N., Shimamura, T., & Suzuki, J. (1996). *Robust method of measurement of fundamental frequency by ACLOS: autocorrelation of log spectrum.* Paper presented at

the 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings.
Kuo, C., & Tjaden, K. (2016). Acoustic variation during passage reading for speakers with dysarthria and healthy controls. *J Commun Disord, 62*, 30-44. doi:10.1016/j.jcomdis.2016.05.003
Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology, 33*(11), 1444-1452.
Kutzelnigg, A., Faber-Rod, J. C., Bauer, J., Lucchinetti, C. F., Sorensen, P. S., Laursen, H., . . . Lassmann, H. (2007). Widespread demyelination in the cerebellar cortex in multiple sclerosis. *Brain Pathol, 17*(1), 38-44. doi:10.1111/j.1750-3639.2006.00041.x
Kutzelnigg, A., Lucchinetti, C. F., Stadelmann, C., Brück, W., Rauschka, H., Bergmann, M., . . . Lassmann, H. (2005). Cortical demyelination and diffuse white matter injury in multiple sclerosis. *Brain, 128*(Pt 11), 2705-2712. doi:10.1093/brain/awh641
Lassmann, H. (2014). Multiple sclerosis: lessons from molecular neuropathology. *Exp Neurol, 262 Pt A*, 2-7. doi:10.1016/j.expneurol.2013.12.003
Li, Y., Zeng, C., & Luo, T. (2011). Paroxysmal dysarthria and ataxia in multiple sclerosis and corresponding magnetic resonance imaging findings. *J Neurol, 258*(2), 273-276. doi:10.1007/s00415-010-5748-4
Lin, E., Hornibrook, J., & Ormond, T. (2012). Evaluating iPhone recordings for acoustic voice assessment. *Folia Phoniatr Logop, 64*(3), 122-130. doi:10.1159/000335874
Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2005). The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *J Acoust Soc Am, 117*(6), 3879-3889. doi:10.1121/1.1898623
Lucchinetti, C. F., Popescu, B. F., Bunyan, R. F., Moll, N. M., Roemer, S. F., Lassmann, H., . . . Ransohoff, R. M. (2011). Inflammatory cortical demyelination in early multiple sclerosis. *N Engl J Med, 365*(23), 2188-2197. doi:10.1056/NEJMoa1100648
Lynch, D. R., Farmer, J. M., Wilson, R. L., & Balcer, L. J. (2005). Performance measures in Friedreich ataxia: potential utility as clinical outcome tools. *Mov Disord, 20*(7), 777-782. doi:10.1002/mds.20449
Mackenzie, C., & Green, J. (2009). Cognitive-linguistic deficit and speech intelligibility in chronic progressive multiple sclerosis. *Int J Lang Commun Disord, 44*(4), 401-420. doi:10.1080/13682820802697879
Magliozzi, R., Howell, O. W., Reeves, C., Roncaroli, F., Nicholas, R., Serafini, B., . . . Reynolds, R. (2010). A Gradient of neuronal loss and meningeal inflammation in multiple sclerosis. *Ann Neurol, 68*(4), 477-493. doi:10.1002/ana.22230
Marcel, C., Anheim, M., Flamand-Rouvière, C., Heran, F., Masnou, P., Boulay, C., . . . Roze, E. (2010). Symptomatic paroxysmal dysarthria-ataxia in demyelinating diseases. *J Neurol, 257*(8), 1369-1372. doi:10.1007/s00415-010-5534-3
Martin, C. L., Phillips, B. A., Kilpatrick, T. J., Butzkueven, H., Tubridy, N., McDonald, E., & Galea, M. P. (2006). Gait and balance impairment in early multiple sclerosis in the absence of clinical disability. *Mult Scler, 12*(5), 620-628. doi:10.1177/1352458506070658
Maryn, Y., Leblans, M., Zarowski, A., & Barkmeier-Kraemer, J. (2019). Objective Acoustic Quantification of Perceived Voice Tremor Severity. *J Speech Lang Hear Res, 62*(10), 3689-3705. doi:10.1044/2019_jslhr-s-19-0024
Mathiowetz, V., Weber, K., Kashman, N., & Volland, G. (1985). Adult norms for the nine hole peg test of finger dexterity. *Occup Ther J Res, 5*(1), 24-38.
Merkel, B., Butzkueven, H., Traboulsee, A. L., Havrdova, E., & Kalincik, T. (2017). Timing of high-efficacy therapy in relapsing-remitting multiple sclerosis: a systematic review. *Autoimmunity reviews, 16*(6), 658-665.
Meyer-Moock, S., Feng, Y. S., Maeurer, M., Dippel, F. W., & Kohlmann, T. (2014). 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 Neurol, 14*, 58. doi:10.1186/1471-2377-14-58

Milo, R., & Miller, A. (2014). Revised diagnostic criteria of multiple sclerosis. *Autoimmun Rev, 13*(4-5), 518-524.

Motta, S., Cesari, U., Mesolella, M., & Motta, G. (2008). Functional vocal results after CO2 laser endoscopic surgery for glottic tumours. *J Laryngol Otol, 122*(9), 948-951. doi:10.1017/s0022215107001211

Mundt, J. C., Vogel, A. P., Feltner, D. E., & Lenderking, W. R. (2012). Vocal acoustic biomarkers of depression severity and treatment response. *Biol Psychiatry, 72*(7), 580-587. doi:10.1016/j.biopsych.2012.03.015

Munger, K. L., Zhang, S. M., O'Reilly, E., Hernán, M. A., Olek, M. J., Willett, W. C., & Ascherio, A. (2004). Vitamin D intake and incidence of multiple sclerosis. *Neurology, 62*(1), 60-65. doi:10.1212/01.Wnl.0000101723.79681.38

Munk Nielsen, N., Corn, G., Frisch, M., Stenager, E., Koch-Henriksen, N., Wohlfahrt, J., . . . Melbye, M. (2019). Multiple sclerosis among first- and second-generation immigrants in Denmark: a population-based cohort study. *Brain, 142*(6), 1587-1597. doi:10.1093/brain/awz088

Murdoch, B. E., Spencer, T. J., Theodoros, D. G., & Thompson, E. C. (1998). Lip and tongue function in multiple sclerosis: A physiological analysis. *Motor Control, 2*(2), 148-160.

Naci, H., Fleurence, R., Birt, J., & Duhig, A. (2010). The impact of increasing neurological disability of multiple sclerosis on health utilities: a systematic review of the literature. *J Med Econ, 13*(1), 78-89. doi:10.3111/13696990903543085

Nevler, N., Ash, S., Irwin, D. J., Liberman, M., & Grossman, M. (2019). Validated automatic speech biomarkers in primary progressive aphasia. *Ann Clin Transl Neurol, 6*(1), 4-14. doi:10.1002/acn3.653

Noffs, G., Boonstra, F. M. C., Perera, T., Butzkueven, H., Kolbe, S. C., Maldonado, F., . . . Vogel, A. P. (2021). Speech metrics, general disability, brain imaging and quality of life in multiple sclerosis. *Eur J Neurol, 28*(1), 259-268. doi:10.1111/ene.14523

Noffs, G., Boonstra, F. M. C., Perera, T., Kolbe, S. C., Stankovich, J., Butzkueven, H., . . . van der Walt, A. (2020). Acoustic Speech Analytics Are Predictive of Cerebellar Dysfunction in Multiple Sclerosis. *Cerebellum*. doi:10.1007/s12311-020-01151-5

Noffs, G., Perera, T., Kolbe, S. C., Shanahan, C. J., Boonstra, F. M. C., Evans, A., . . . Vogel, A. P. (2018). What speech can tell us: A systematic review of dysarthria characteristics in Multiple Sclerosis. *Autoimmun Rev, 17*(12), 1202-1209. doi:10.1016/j.autrev.2018.06.010

Nordio, S., Bernitsas, E., Meneghello, F., Palmer, K., Stabile, M. R., Dipietro, L., & Di Stadio, A. (2018). Expiratory and phonation times as measures of disease severity in patients with Multiple Sclerosis. A case-control study. *Mult Scler Relat Disord, 23*, 27-32. doi:10.1016/j.msard.2018.04.010

O'Connor, P. (2002). Key issues in the diagnosis and treatment of multiple sclerosis. An overview. *Neurology, 59*(6 Suppl 3), S1-33. doi:10.1212/wnl.59.6_suppl_3.s1

Olerup, O., & Hillert, J. (1991). HLA class II-associated genetic susceptibility in multiple sclerosis: a critical evaluation. *Tissue Antigens, 38*(1), 1-15. doi:10.1111/j.1399-0039.1991.tb02029.x

Olsson, T., Barcellos, L. F., & Alfredsson, L. (2017). Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. *Nat Rev Neurol, 13*(1), 25-36. doi:10.1038/nrneurol.2016.187

Panou, T., Mastorodemos, V., Papadaki, E., Simos, P. G., & Plaitakis, A. (2012). Early signs of memory impairment among multiple sclerosis patients with clinically isolated syndrome. *Behav Neurol, 25*(4), 311-326. doi:10.3233/ben-2012-110201

Parsa, V., Jamieson, D. G., & Pretty, B. R. (2001). Effects of microphone type on acoustic measures of voice. *J Voice, 15*(3), 331-343. doi:10.1016/s0892-1997(01)00035-2

Patejdl, R., & Zettl, U. K. (2017). Spasticity in multiple sclerosis: Contribution of inflammation, autoimmune mediated neuronal damage and therapeutic interventions. *Autoimmun Rev, 16*(9), 925-936. doi:10.1016/j.autrev.2017.07.004

Patwardhan, M. B., Matchar, D. B., Samsa, G. P., McCrory, D. C., Williams, R. G., & Li, T. T. (2005). Cost of multiple sclerosis by level of disability: a review of literature. *Mult Scler, 11*(2), 232-239. doi:10.1191/1352458505ms1137oa

Paul, R. H., Cohen, R. A., Gilchrist, J. M., Aloia, M. S., & Goldstein, J. M. (2000). Cognitive dysfunction in individuals with myasthenia gravis. *Journal of the neurological sciences, 179*(1), 59-64.

Pender, M. P., & Ferguson, S. M. (2007). Dysarthria and dysphagia due to the opercular syndrome in multiple sclerosis. *Mult Scler, 13*(6), 817-819.

Peus, S. (2004). *Modern acoustic and electronic design of studio condenser microphones.* Paper presented at the Audio Engineering Society Convention 116.

Piacentini, V., Mauri, I., Cattaneo, D., Gilardone, M., Montesano, A., & Schindler, A. (2014). Relationship between quality of life and dysarthria in patients with multiple sclerosis. *Archives of physical medicine and rehabilitation, 95*(11), 2047-2054.

Popescu, V., Agosta, F., Hulst, H. E., Sluimer, I. C., Knol, D. L., Sormani, M. P., . . . Vrenken, H. (2013). Brain atrophy and lesion load predict long term disability in multiple sclerosis. *J Neurol Neurosurg Psychiatry, 84*(10), 1082-1091. doi:10.1136/jnnp-2012-304094

Prinz, M., & Priller, J. (2017). The role of peripheral immune cells in the CNS in steady state and disease. *Nat Neurosci, 20*(2), 136-144. doi:10.1038/nn.4475

Pützer, M., Wokurek, W., & Moringlane, J. R. (2017). Evaluation of Phonatory Behavior and Voice Quality in Patients with Multiple Sclerosis Treated with Deep Brain Stimulation. *J Voice, 31*(4), 483-489. doi:10.1016/j.jvoice.2016.10.022

Ravizza, S. M., McCormick, C. A., Schlerf, J. E., Justus, T., Ivry, R. B., & Fiez, J. A. (2006). Cerebellar damage produces selective deficits in verbal working memory. *Brain, 129*(Pt 2), 306-320. doi:10.1093/brain/awh685

Ravnborg, M., Blinkenberg, M., Sellebjerg, F., Ballegaard, M., Larsen, S. H., & Sørensen, P. S. (2005). Responsiveness of the multiple sclerosis impairment scale in comparison with the expanded disability status scale. *Mult Scler, 11*(1), 81-84.

Rejdak, K., Jackson, S., & Giovannoni, G. (2010). Multiple sclerosis: a practical overview for clinicians. *Br Med Bull, 95*, 79-104. doi:10.1093/bmb/ldq017

Riecker, A., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W., & Ackermann, H. (2005). fMRI reveals two distinct cerebral networks subserving speech motor control. *Neurology, 64*(4), 700-706. doi:10.1212/01.Wnl.0000152156.90779.89

Roberg, B. L., Somogie, M., Thelen, J. M., & Bruce, J. M. (2015). Articulation Time Does Not Affect Speeded Cognitive Performance in Multiple Sclerosis. *Cognitive and Behavioral Neurology, 28*(1), 33-38.

Rodgers, J. D., Tjaden, K., Feenaughty, L., Weinstock-Guttman, B., & Benedict, R. H. (2013). Influence of cognitive function on speech and articulation rate in multiple sclerosis. *J Int Neuropsychol Soc, 19*(2), 173-180. doi:10.1017/s1355617712001166

Rosati, G. (2001). The prevalence of multiple sclerosis in the world: an update. *Neurol Sci, 22*(2), 117-139.

Rosen, K., Murdoch, B., Folker, J., Vogel, A., Cahill, L., Delatycki, M., & Corben, L. (2010). Automatic method of pause measurement for normal and dysarthric speech. *Clin Linguist Phon, 24*(2), 141-154. doi:10.3109/02699200903440983

Rosen, K. M., Goozée, J. V., & Murdoch, B. E. (2008). Examining the effects of multiple sclerosis on speech production: Does phonetic structure matter? *J Commun Disord, 41*(1), 49-69.

Rovaris, M., Agosta, F., Sormani, M. P., Inglese, M., Martinelli, V., Comi, G., & Filippi, M. (2003). Conventional and magnetization transfer MRI predictors of clinical multiple sclerosis evolution: a medium-term follow-up study. *Brain, 126*(10), 2323-2332. doi:10.1093/brain/awg232

Rovini, E., Maremmani, C., & Cavallo, F. (2017). How Wearable Sensors Can Support Parkinson's Disease Diagnosis and Treatment: A Systematic Review. *Front Neurosci, 11*, 555. doi:10.3389/fnins.2017.00555

Rudick, R. A., Polman, C. H., Cohen, J. A., Walton, M. K., Miller, A. E., Confavreux, C., . . . Sandrock, A. W. (2009). Assessing disability progression with the Multiple Sclerosis Functional Composite. *Mult Scler, 15*(8), 984-997. doi:10.1177/1352458509106212

Rusz, J., Benova, B., Ruzickova, H., Novotny, M., Tykalova, T., Hlavnicka, J., . . . Horakova, D. (2018). Characteristics of motor speech phenotypes in multiple sclerosis. *Mult Scler Relat Disord, 19*, 62-69. doi:10.1016/j.msard.2017.11.007

Rusz, J., Cmejla, R., Ruzickova, H., & Ruzicka, E. (2011). Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *the Journal of the Acoustical Society of America, 129*(1), 350-367.

Rusz, J., Hlavnicka, J., Tykalova, T., Novotny, M., Dusek, P., Sonka, K., & Ruzicka, E. (2018). Smartphone Allows Capture of Speech Abnormalities Associated With High Risk of Developing Parkinson's Disease. *IEEE Trans Neural Syst Rehabil Eng, 26*(8), 1495-1507. doi:10.1109/tnsre.2018.2851787

Rusz, J., Vaneckova, M., Benova, B., Tykalova, T., Novotny, M., Ruzickova, H., . . . Horakova, D. (2019). Brain volumetric correlates of dysarthria in multiple sclerosis. *Brain Lang, 194*, 58-64. doi:10.1016/j.bandl.2019.04.009

Saccà, F., Lanzillo, R., Signori, A., Maniscalco, G. T., Signoriello, E., Lo Fermo, S., . . . Sormani, M. P. (2019). Determinants of therapy switch in multiple sclerosis treatment-naïve patients: A real-life study. *Mult Scler, 25*(9), 1263-1272. doi:10.1177/1352458518790390

Sanfilipo, M. P., Benedict, R. H., Sharma, J., Weinstock-Guttman, B., & Bakshi, R. (2005). The relationship between whole brain volume and disability in multiple sclerosis: a comparison of normalized gray vs. white matter with misclassification correction. *Neuroimage, 26*(4), 1068-1077. doi:10.1016/j.neuroimage.2005.03.008

Scalfari, A., Neuhaus, A., Degenhardt, A., Rice, G. P., Muraro, P. A., Daumer, M., & Ebers, G. C. (2010). The natural history of multiple sclerosis: a geographically based study 10: relapses and long-term disability. *Brain, 133*(Pt 7), 1914-1929. doi:10.1093/brain/awq118

Schmahmann, J. D., Guell, X., Stoodley, C. J., & Halko, M. A. (2019). The Theory and Neuroscience of Cerebellar Cognition. *Annu Rev Neurosci, 42*, 337-364. doi:10.1146/annurev-neuro-070918-050258

Schmidt, S. (2013). Fampridine improves severe dysarthria in secondary-progressive multiple sclerosis. *Mult Scler, 19*(4), 511-511.

Schmitz-Hübsch, T., Du Montcel, S. T., Baliko, L., Berciano, J., Boesch, S., Depondt, C., . . . Kang, J.-S. (2006). Scale for the assessment and rating of ataxia: development of a new clinical scale. *Neurology, 66*(11), 1717-1720.

Schmitz-Hübsch, T., Giunti, P., Stephenson, D. A., Globas, C., Baliko, L., Saccà, F., . . . Klockgether, T. (2008). SCA Functional Index: a useful compound performance measure for spinocerebellar ataxia. *Neurology, 71*(7), 486-492. doi:10.1212/01.wnl.0000324863.76290.19

Shabgah, A. G., Navashenaq, J. G., Shabgah, O. G., Mohammadi, H., & Sahebkar, A. (2017). Interleukin-22 in human inflammatory diseases and viral infections. *Autoimmun Rev, 16*(12), 1209-1218. doi:10.1016/j.autrev.2017.10.004

Sharif, K., Watad, A., Bragazzi, N. L., Lichtbroun, M., Amital, H., & Shoenfeld, Y. (2018). Physical activity and autoimmune diseases: Get moving and manage the disease. *Autoimmun Rev, 17*(1), 53-72. doi:10.1016/j.autrev.2017.11.010

Sharrack, B., Hughes, R. A. C., Soudain, S., & Dunn, G. (1999). The psychometric properties of clinical rating scales used in multiple sclerosis. *Brain, 122*(1), 141-159. doi:10.1093/brain/122.1.141

Shiee, N., Bazin, P. L., Zackowski, K. M., Farrell, S. K., Harrison, D. M., Newsome, S. D., . . . Reich, D. S. (2012). Revisiting brain atrophy and its relationship to disability in multiple sclerosis. *PLoS One, 7*(5), e37049. doi:10.1371/journal.pone.0037049

Shue, Y.-L. (2010). *The voice source in speech production: Data, analysis and models*: University of California, Los Angeles.

Shue, Y. L., Keating, P., & Vicenik, C. (2009). VOICESAUCE: A program for voice analysis. *J Acoust Soc Am, 126*(4), 2221-2221. doi:10.1121/1.3248865

Skodda, S., & Schlegel, U. (2008). Speech rate and rhythm in Parkinson's disease. *Mov Disord, 23*(7), 985-992. doi:10.1002/mds.21996

Smith, A. (1982). Symbol digit modalities test (SDMT) manual (revised) Western Psychological Services. *Los Angeles*.

Smith, M. M., & Arnett, P. A. (2007). Dysarthria predicts poorer performance on cognitive tasks requiring a speeded oral response in an MS population. *Journal of clinical and experimental neuropsychology, 29*(8), 804-812.

Spelman, T., Mekhael, L., Burke, T., Butzkueven, H., Hodgkinson, S., Havrdova, E., . . . Vucic, S. (2016). Risk of early relapse following the switch from injectables to oral agents for multiple sclerosis. *Eur J Neurol, 23*(4), 729-736. doi:10.1111/ene.12929

Starkstein, S. E., Jorge, R., Mizrahi, R., & Robinson, R. G. (2006). A prospective longitudinal study of apathy in Alzheimer's disease. *J Neurol Neurosurg Psychiatry, 77*(1), 8-11. doi:10.1136/jnnp.2005.069575

Stipancic, K. L., Tjaden, K., & Wilding, G. (2016). Comparison of Intelligibility Measures for Adults With Parkinson's Disease, Adults With Multiple Sclerosis, and Healthy Controls. *Journal of Speech, Language, and Hearing Research, 59*(2), 230-238.

Sussman, J. E., & Tjaden, K. (2012). Perceptual measures of speech from individuals with Parkinson's disease and multiple sclerosis: intelligibility and beyond. *J Speech Lang Hear Res, 55*(4), 1208-1219. doi:10.1044/1092-4388(2011/11-0048)

Svec, J. G., & Granqvist, S. (2010). Guidelines for selecting microphones for human voice production research. *Am J Speech Lang Pathol, 19*(4), 356-368. doi:10.1044/1058-0360(2010/09-0091)

Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., . . . Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol, 17*(2), 162-173. doi:10.1016/s1474-4422(17)30470-2

Tintore, M., Rovira, A., Arrambide, G., Mitjana, R., Río, J., Auger, C., . . . Montalban, X. (2010). Brainstem lesions in clinically isolated syndromes. *Neurology, 75*(21), 1933-1938. doi:10.1212/WNL.0b013e3181feb26f

Tintore, M., Rovira, À., Río, J., Otero-Romero, S., Arrambide, G., Tur, C., . . . Montalban, X. (2015). Defining high, medium and low impact prognostic factors for developing multiple sclerosis. *Brain, 138*(Pt 7), 1863-1874. doi:10.1093/brain/awv105

Titze, I. R. (1994). Toward standards in acoustic analysis of voice. *J Voice, 8*(1), 1-7.

Titze, I. R., & Winholtz, W. S. (1993). Effect of microphone type and placement on voice perturbation measurements. *J Speech Hear Res, 36*(6), 1177-1190. doi:10.1044/jshr.3606.1177

Tjaden, K., & Watling, E. (2003a). Characteristics of diadochokinesis in multiple sclerosis and Parkinson's disease. *Folia Phoniatr Logop, 55*(5), 241-259. doi:72155

Tjaden, K., & Watling, E. (2003b). Characteristics of diadochokinesis in multiple sclerosis and Parkinson's disease. *Folia Phoniatrica et Logopaedica, 55*(5), 241-259.
Trapp, B. D., & Stys, P. K. (2009). Virtual hypoxia and chronic necrosis of demyelinated axons in multiple sclerosis. *Lancet Neurol, 8*(3), 280-291. doi:10.1016/s1474-4422(09)70043-2
Trojano, M., Tintore, M., Montalban, X., Hillert, J., Kalincik, T., Iaffaldano, P., . . . Butzkueven, H. (2017). Treatment decisions in multiple sclerosis - insights from real-world observational studies. *Nat Rev Neurol, 13*(2), 105-118. doi:10.1038/nrneurol.2016.188
Tüzün, E., Akman-Demir, G., & Eraksoy, M. (2001). Paroxysmal attacks in multiple sclerosis. *Mult Scler, 7*(6), 402-404. doi:10.1177/135245850100700609
Twomey, J., & Espir, M. (1980). Paroxysmal symptoms as the first manifestations of multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry, 43*(4), 296-304.
Valentino, P., Cerasa, A., Chiriaco, C., Nistico, R., Pirritano, D., Gioia, M., . . . Gallo, O. (2009). Cognitive deficits in multiple sclerosis patients with cerebellar symptoms. *Mult Scler, 15*(7), 854-859.
Valentino, P., Nisticò, R., Pirritano, D., Bilotti, G., Del Giudice, F., Sturniolo, M., & Quattrone, A. (2011). Lamotrigine therapy for paroxysmal dysarthria caused by multiple sclerosis: a case report. *Journal of neurology, 258*(7), 1349-1350.
van der Mei, I. A., Ponsonby, A. L., Dwyer, T., Blizzard, L., Simmons, R., Taylor, B. V., . . . Kilpatrick, T. (2003). Past exposure to sun, skin phenotype, and risk of multiple sclerosis: case-control study. *Bmj, 327*(7410), 316. doi:10.1136/bmj.327.7410.316
Van der Walt, A., Stankovich, J., Bahlo, M., Taylor, B. V., Van der Mei, I. A., Foote, S. J., . . . Butzkueven, H. (2011). Heterogeneity at the HLA-DRB1 allelic variation locus does not influence multiple sclerosis disease severity, brain atrophy or cognition. *Mult Scler, 17*(3), 344-352. doi:10.1177/1352458510389101
van Munster, C. E., & Uitdehaag, B. M. (2017). Outcome Measures in Clinical Trials for Multiple Sclerosis. *CNS Drugs, 31*(3), 217-236. doi:10.1007/s40263-017-0412-5
Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology Section A, 47*(3), 631-650.
Vavougios, G. D., Doskas, T., & Konstantopoulos, K. (2018). An electroglottographical analysis-based discriminant function model differentiating multiple sclerosis patients from healthy controls. *Neurol Sci, 39*(5), 847-850.
Vitorino, J. (2009). Velopharyngeal function in adult speakers of Portuguese diagnosed with multiple sclerosis. *NeuroRehabilitation, 25*(4), 279-287.
Vizza, P., Tradigo, G., Mirarchi, D., Bossio, R. B., Lombardo, N., Arabia, G., . . . Veltri, P. (2019). Methodologies of speech analysis for neurodegenerative diseases evaluation. *Int J Med Inform, 122*, 45-54. doi:10.1016/j.ijmedinf.2018.11.008
Vogel, A. P., Fletcher, J., & Maruff, P. (2010). Acoustic analysis of the effects of sustained wakefulness on speech. *J Acoust Soc Am, 128*(6), 3747-3756. doi:10.1121/1.3506349
Vogel, A. P., Fletcher, J., & Maruff, P. (2014). The impact of task automaticity on speech in noise. *Speech Communication, 65*, 1-8.
Vogel, A. P., Fletcher, J., Snyder, P. J., Fredrickson, A., & Maruff, P. (2011). Reliability, stability, and sensitivity to change and impairment in acoustic measures of timing and frequency. *J Voice, 25*(2), 137-149. doi:10.1016/j.jvoice.2009.09.003
Vogel, A. P., & Maruff, P. (2008). Comparison of voice acquisition methodologies in speech research. *Behav Res Methods, 40*(4), 982-987. doi:10.3758/brm.40.4.982
Vogel, A. P., & Maruff, P. (2014). Monitoring change requires a rethink of assessment practices in voice and speech. *Logoped Phoniatr Vocol, 39*(2), 56-61. doi:10.3109/14015439.2013.775332
Vogel, A. P., Maruff, P., Snyder, P. J., & Mundt, J. C. (2009). Standardization of pitch-range settings in voice acoustic analysis. *Behavior research methods, 41*(2), 318-324.

Vogel, A. P., & Morgan, A. T. (2009). Factors affecting the quality of sound recording for speech and voice analysis. *Int J Speech Lang Pathol, 11*(6), 431-437. doi:10.3109/17549500902822189

Vogel, A. P., Poole, M. L., Pemberton, H., Caverle, M. W. J., Boonstra, F. M. C., Low, E., . . . Brodtmann, A. (2017). Motor speech signature of behavioral variant frontotemporal dementia: Refining the phenotype. *Neurology, 89*(8), 837-844. doi:10.1212/wnl.0000000000004248

Vogel, A. P., Rommel, N., Oettinger, A., Horger, M., Krumm, P., Kraus, E. M., . . . Synofzik, M. (2017). Speech and swallowing abnormalities in adults with POLG associated ataxia (POLG-A). *Mitochondrion, 37*, 1-7. doi:10.1016/j.mito.2017.06.002

Vogel, A. P., Rommel, N., Oettinger, A., Stoll, L. H., Kraus, E. M., Gagnon, C., . . . Synofzik, M. (2018). Coordination and timing deficits in speech and swallowing in autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS). *J Neurol, 265*(9), 2060-2070. doi:10.1007/s00415-018-8950-4

Vogel, A. P., Rosen, K. M., Morgan, A. T., & Reilly, S. (2014). Comparability of modern recording devices for speech analysis: smartphone, landline, laptop, and hard disc recorder. *Folia Phoniatr Logop, 66*(6), 244-250. doi:10.1159/000368227

Vogel, A. P., Shirbin, C., Churchyard, A. J., & Stout, J. C. (2012). Speech acoustic markers of early stage and prodromal Huntington's disease: a marker of disease onset? *Neuropsychologia, 50*(14), 3273-3278. doi:10.1016/j.neuropsychologia.2012.09.011

Vogel, A. P., Wardrop, M. I., Folker, J. E., Synofzik, M., Corben, L. A., Delatycki, M. B., & Awan, S. N. (2017). Voice in Friedreich Ataxia. *J Voice, 31*(2), 243.e249-243.e219. doi:10.1016/j.jvoice.2016.04.015

Wang, Y. T., Kent, R. D., Duffy, J. R., & Thomas, J. E. (2009). Analysis of diadochokinesis in ataxic dysarthria using the motor speech profile program. *Folia Phoniatr Logop, 61*(1), 1-11. doi:10.1159/000184539

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: context, process, and purpose. In: Taylor & Francis.

Wattjes, M. P., Rovira, À., Miller, D., Yousry, T. A., Sormani, M. P., de Stefano, M. P., . . . Xavier, M. (2015). Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis--establishing disease prognosis and monitoring patients. *Nat Rev Neurol, 11*(10), 597-606. doi:10.1038/nrneurol.2015.157

WHO. (2013). Atlas of MS 2013. Mapping Multiple Sclerosis Around the World. *Available at http://www.msif.org/wp-content/uploads/2014/09/Atlas-of-MS.pdf, Accessed August, 2020*.

Yamout, B., Al-Zaghal, Z., El-Dahouk, I., Farhat, S., Sibai, A., & Hamdan, A.-L. H. (2013). Mean Contact Quotient Using Electroglottography in Patients With Multiple Sclerosis. *Journal of Voice, 27*(4), 506-511.

Yamout, B., Fuleihan, N., Hajj, T., Sibai, A., Sabra, O., Rifai, H., & Hamdan, A.-L. (2009). Vocal symptoms and acoustic changes in relation to the expanded disability status scale, duration and stage of disease in patients with multiple sclerosis. *European Archives of Oto-Rhino-Laryngology, 266*(11), 1759-1765.

Yorkston, K. M., Baylor, C. R., Klasner, E. R., Deitz, J., Dudgeon, B. J., Eadie, T., . . . Amtmann, D. (2007). Satisfaction with communicative participation as defined by adults with multiple sclerosis: A qualitative study. *Journal of Communication Disorders, 40*(6), 433-451.

Yorkston, K. M., Klasner, E. R., Bowen, J., Ehde, D. M., Gibbons, L. E., Johnson, K., & Kraft, G. (2003). Characteristics of multiple sclerosis as a function of the severity of speech disorders. *Journal of Medical Speech-Language Pathology, 11*(2), 73-85.

Yorkston, K. M., Klasner, E. R., & Swanson, K. M. (2001). Communication in ContextA Qualitative Study of the Experiences of Individuals With Multiple Sclerosis. *American Journal of Speech-Language Pathology, 10*(2), 126-137.

Zesiewicz, T. A., Greenstein, P. E., Sullivan, K. L., Wecker, L., Miller, A., Jahan, I., . . . Perlman, S. L. (2012). A randomized trial of varenicline (Chantix) for the treatment of spinocerebellar ataxia type 3. *Neurology, 78*(8), 545-550. doi:10.1212/WNL.0b013e318247cc7a

Zettl, U. K., Stüve, O., & Patejdl, R. (2012). Immune-mediated CNS diseases: a review on nosological classification and clinical features. *Autoimmunity reviews, 11*(3), 167-173.

Ziemssen, T., Medin, J., Couto, C. A.-M., & Mitchell, C. R. (2017). Multiple sclerosis in the real world: a systematic review of fingolimod as a case study. *Autoimmunity reviews, 16*(4), 355-376.

Zivadinov, R., Bergsland, N., Dolezal, O., Hussein, S., Seidl, Z., Dwyer, M. G., . . . Horáková, D. (2013). Evolution of cortical and thalamus atrophy and disability progression in early relapsing-remitting MS during 5 years. *AJNR Am J Neuroradiol, 34*(10), 1931-1939. doi:10.3174/ajnr.A3503

Appendices

Appendix A: Speech tasks

1. PLEASE READ THE FOLLOWING PASSAGE:

You wish to know all about my grandfather. Well, he is nearly 93 years old, yet he still thinks as swiftly as ever. He dresses himself in an old black frock coat, usually several buttons missing. A long beard clings to his chin, giving those who observe him a pronounced feeling of the utmost respect. We have often urged him to walk more and smoke less, but he always answers, “Banana oil!” Grandfather likes to be modern in his language.

2. PLEASE SAY “AAAHHH” for AS LONG AS YOU CAN WHILE WE RECORD YOU.

3. PLEASE SAY THE DAYS OF THE WEEK, STARTING AT MONDAY, IN ONE BREATH

4.1. PLEASE REPEAT /pa/ta/ka/ - /pa/ta/ka/ AS FAST AND AS CLEARLY AS YOU CAN FOR 10 SECONDS.

4.2.NOW, PLEASE REPEAT /pa/ta/ - /pa/ta/ AS FAST AND AS CLEARLY AS YOU CAN FOR 10 SECONDS.

5. WHEN YOU ARE READY, I WILL ASK YOU TO TELL ME A NICE PERSONAL STORY OR FAVORITY MEMORY WHILE I RECORD YOU FOR ONE FULL MINUTE. PLEASE, CONTINUE SPEAKING UNTILL I TELL YOU TO STOP.

Appendix B: Perceptual analysis of speech scores

Participant:_____ Date:_____ Rater:_____

Domain	Feature	Severity (0-4)	Additional notes	Consensus rating (0-4)
General	Naturalness			
	Inteligibility			
Pitch	Monopitch			
	Pitch breaks			
	Voice tremor			
Respiration	Audible inspiration			
loudness	Monoloudness			
	Loudness decay			
prosody	Rate			
	Variable rate			
	Reduced stress			
	Prolonged intervals			
voice	Rough			
	Breathy			
	Strained			
articulation	Imprecise consonants			
	Prolonged phonemes			
	Repeated phonemes			
	Irregular articulatory breakdowns			
	Distorted vowels			
	Groping			

	Phonemic errors			
	False starts			
Resonance	Hypo/Hypernasality			
DDK	Speed			
	Regularity			