How People Living With Motor Neurone Disease Use Personalised Automatic Speech Recognition Technology To Support Social Interaction

A thesis submitted for the degree of Doctor of Philosophy

University College London

Richard Cave

Student ID 17144322

July 2024

Language and Cognition Department

Division of Psychology and Language Science

University College London

Chandler House

2, Wakefield Street

London

WC1N 1PF

Declaration

I, Richard Cave

Confirm that the work presented in this thesis is my own.

Where information has been derived from other sources,

I confirm that this has been so indicated in the thesis.

Signature: Richard Cave Date: 28 July 2024

Abstract

Motor Neurone Disease (MND) is a progressive, ultimately fatal disease causing increasing muscular weakness. Most people living with MND (plwMND) experience acquired neurogenic motor speech difficulties (dysarthria) during the course of the condition. Although more than 80% of plwMND will become unable to communicate using natural speech, many express a desire to use speech for as long as possible, even when unintelligible to others.

Project Relate, an app from Google, is argued to support people with 'non-standard' speech to be better understood by others. Relate builds a personalised automated speech recognition model (ASR) that may better recognise speech with dysarthria, generating real-time captioning designed to help listener understanding.

This study videoed interactions between three plwMND and their communication partners over a year, to better understand the effects of personalised ASR captions and how the use of ASR may change as MND progresses.

This study assessed ASR caption accuracy and how well it preserved meaning. Conversation Analysis (CA) was used to identify participants' own organisational practices in the accomplishment of interaction in the presence of ASR captioning. Semi-structured interviews were completed and assessed using Thematic Analysis.

Trouble and repair were observed as a feature of the dyad extracts. Similarities and differences in allocation of trouble responsibility, as well as gaze direction were observed when comparing Relate to other forms of augmentative and alternative communication (AAC).

The expectations of all dyads were not met using Relate. Forecast caption accuracy was higher than actual and the percentage of captioned phrases that preserved meaning was frequently low. Issues were identified with app design, usability and support available. The high level of incorrect ASR captioning may reinforce perceptions of self-blame for voice change. The study provides wide-ranging recommendations, including user-centred review of app design, support, training for professionals and significant others.

Impact Statement

This is the first study that explores how people living with MND (plwMND) utilize personalised automated speech recognition (ASR), argued to better recognise speech with dysarthria and support everyday conversational communication.

This research studied the effect of ASR captioning from three perspectives:

- From the perspective of the ASR technology itself – how easy and understandable the technology is to train and to use, how it sets user expectations on ASR performance, how it measures its own performance, how accurately it captions speech and maintains meaning.
- From the perspective of conversation analysis (CA) – describing how each dyad conducted social interactions and how troubles and repair were managed, how responsibility was allocated and how gaze may differ when using ASR Captions compared to other types of AAC.
- Directly from the perspective of plwMND and their communication partners through semi-structured interviews, using Thematic Analysis (TA) for qualitative analysis of the interview transcripts.

The research reveals how plwMND with little confidence of technology can find Relate difficult to understand and to use, with limited instructions and support available. Recommendations are made to raise awareness and build skills, training and documentation.

The research showed how recording the mandatory 500 phrases to create a personalised ASR was fatiguing and providing little indication of reward for the effort. Suggestions are made for building and maintaining motivation.

Custom cards are argued to support better ASR recognition of words that plwMND use and this research showed (in a limited way) how ASR recognition of proper nouns improved, and recommended a larger study for further analysis.

The research showed that Relate sets overly optimistic expectations of model accuracy based on forecast word error rate (WER) in comparison to the actual WER measured in this study. Incorrectly captioned phrases that resulted in meaning loss were often more than 70% of all phrases in a session. There was no measure that included listener understanding, a key objective of Relate.

Using CA, the research showed that trouble and repair were features in the interactions, resulting in frequent instances of other-initiation of repair. Inaccurate captions can cause cascading troubles requiring extended repair. Communication partners frequently managed trouble responsibility as a function of supporting the management of their relationship with plwMND. Patterns of gaze appeared to differ using Project Relate compared to other AAC.

Thematic analysis indicated that plwMND and conversation partners were hopeful of the benefits of personalised ASR technology, though none felt it had met expectations. Some said it may help in particular contexts but many reported that caption accuracy was just not accurate enough to be useful.

This research underscores the practical benefits of understanding the needs of both plwMND and communication partners in product design.

Finally, even though the goal of Project Relate is to support successful conversation, it is not a replacement for therapeutic interventions such as conversation partner training.

The research has generated widespread interest and the project has been presented internationally.

UCL Research Paper Declaration Form

referencing the doctoral candidate's own published work(s)

For a research manuscript that has already been published (1/2)

a) **What is the title of the manuscript?**
How People Living With Amyotrophic Lateral Sclerosis Use Personalized Automatic Speech Recognition Technology To Support Communication

The published manuscript comprises content from chapters 1, 3-9.

b) **Please include a link to or doi for the work**
https://doi.org/10.1044/2024_JSLHR-24-00097

c) **Where was the work published?**
Journal of Speech, Language, and Hearing Research

d) **Who published the work?** (e.g. OUP)
ASHA

e) **When was the work published?**
11 July 2024

f) **List the manuscript's authors in the order they appear on the publication**
Richard Cave

g) **Was the work peer reviewed?**
Yes

h) **Have you retained the copyright?**
Yes

i) **Was an earlier form of the manuscript uploaded to a preprint server?** (e.g. medRxiv). If 'Yes', please give a link or doi)
https://pubs.asha.org/doi/10.1044/2024_JSLHR-24-00097

UCL Research Paper Declaration Form

referencing the doctoral candidate's own published work(s)

For a research manuscript that has already been published (2/2)

a) **What is the title of the manuscript?**
The use of speech recognition technology by people living with amyotrophic lateral sclerosis: a scoping review

The published manuscript comprises content from Chapter 2 (2.2 – 2.5, part of 2.8).

b) **Please include a link to or doi for the work**
10.1080/17483107.2021.1974961

c) **Where was the work published?**
Disability and Rehabilitation: Assistive Technology

d) **Who published the work?** (e.g. OUP)
Taylor and Francis

e) **When was the work published?**
11 September 2021

f) **List the manuscript's authors in the order they appear on the publication**
Richard Cave and Steven Bloch

g) **Was the work peer reviewed?**
Yes

h) **Have you retained the copyright?**
Yes

i) **Was an earlier form of the manuscript uploaded to a preprint server?** (e.g. medRxiv). If 'Yes', please give a link or doi)
https://discovery.ucl.ac.uk/id/eprint/10133835/1/17483107.2021.pdf

Acknowledgements

Thank you to Professor Steven Bloch and Professor Cathy Holloway for their wisdom, encouragement and ever-patient supervision throughout this project.

I am very grateful to the wonderful team at the MND Association who supported the project from the start.

Most of all, thank you to the people living with MND and their communication partners for their precious time and support for this research.

Table of Contents

1. Introduction 19
   1.1. Introduction to the chapter 19
   1.2. Motor Neurone Disease 20
   1.3. Augmentative and alternative communication (AAC) 21
   1.4. Identity 22
      1.4.1. Speech and identity 22
      1.4.2. AAC and identity 23
      1.4.3. Voice Banking and identity 26
   1.5. AAC Speed versus Accuracy 28
   1.6. Artificial Intelligence and AAC 29
   1.7. Automated Speech Recognition (ASR) 30
      1.7.1. Dysarthria and ASR 32
      1.7.2. Dysarthria and Word Error Rate 34
      1.7.3. Additional factors that may affect ASR 36
      1.7.4. ASR and Identity 38
      1.7.5. ASR and equality of service 39
      1.7.6. Project Euphonia 40
      1.7.7. Project Relate 42
   1.8. Conversation Analysis (CA) 45
      1.8.1. Trouble and Repair 46
      1.8.1. CA analysis of speakers with dysarthria 48
      1.8.1. AAC-in-interaction 48
      1.8.1. Allocation of Responsibility for Trouble 50
      1.8.1. Eye gaze 51
      1.8.1. CA analysis of personalised ASR in-interaction 52
   1.9. Introduction - Summary 53
   1.10. Research Questions 54
2. Previous Research 55
   2.1. Introduction to the chapter 55
   2.2. Background 56
   2.3. Objectives 57

2.4. Discussion 57
2.4.1. Increased use of machine learning technology in research 57
2.4.2. What do plwMND use ASR technology for? 58
2.4.3. How does ASR use and usefulness change as MND progresses? 59
2.4.4. WER remains the main measure for measuring ASR performance. 59
2.4.5. Recordings of plwMND remain in short supply or may be selectively used 60
2.4.6. Could ASR technology support diagnosis, therapeutic intervention? 62
2.4.7. What communication, physical and cognitive needs do participants have? 64
2.4.8. Could changing how plwMND speak support functional use of ASR technology? 64
2.4.9. ASR technology for non-speech or limited speech recognition 65
2.4.10. BCI technology with an approximation of natural speech 65
2.5. Conclusions 66
2.6. Research Direction 68
3. Objectives 70
4. Method 72
4.1. Participants 72
4.1. Recruitment 74
4.1.1. Informed consent and confidentiality 75
4.2. Participants 75
4.2.1. Pam and Carol 75
4.2.1. Bob and Adrian 76
4.2.1. Elsbeth and Tony 76
4.3. Equipment 77
4.3.1. Relate set-up 77
4.4. Assessment of dysarthria 80
4.5. Conversation analysis 80
4.6. ASR Accuracy - WER 81
4.7. ASR Accuracy - Meaning Retention 81
4.8. Thematic Analysis 84
4.8.1. Stakeholder involvement 85
4.8.2. Interview Analysis 85
4.8.3. Analysis approach and situating one's perspective 86
4.9. Additional records for each session 88

5. Results 89
5.1. Introduction to the chapter 89
5.2. ASR model performance 90
5.2.1. Pam – personalised ASR performance 91
5.2.2. FDA-2 Assessment 91
5.2.3. WER – Forecast vs Actual 91
5.2.4. Meaning Preservation 92
5.2.5. Custom Cards 94
5.2.6. Referencing the ASR Captions 95
5.2.7. Experience of using Relate 95
5.2.8. Bob – personalised ASR performance 96
5.2.9. FDA-2 Assessment 96
5.2.9.1. WER – Forecast vs Actual 96
5.2.9.1. Meaning Preservation 97
5.2.10. Custom Cards 98
5.2.11. Referencing the ASR Captions 99
5.2.12. Experience of using Relate 99
5.2.13. Elsbeth – personalised ASR performance 100
5.2.14. FDA-2 Assessment 100
5.2.14.1. WER – Forecast vs Actual 100
5.2.14.1. Meaning Preservation 101
5.2.15. Custom Cards 102
5.2.16. Referencing the ASR Captions 103
5.2.17. Experience of using Relate 103
5.3. Conversation Analysis 104
5.3.1. ASR captions contributing to efficient repair resolution 105
5.3.2. A little white board 106
5.3.3. Granddaughter getting married 109
5.3.4. Vegan Recipes 111
5.3.5. Some of the players are nearly eighty 113
5.3.6. ASR captions disrupting or lengthening repair resolution 115
5.3.7. Incorrectly captioned words, proper nouns, causing multiple troubles 115
5.3.8. Stereo 116

5.3.9. A walk and then coffee 120
5.3.10. The A259 125
5.3.11. Horatio 132
5.3.12. Extended repair when the captions perceived as accurate, but not 135
5.3.13. Movie Day 135
5.3.14. Vegetarian Shepherds 139
5.3.15. Mussolini 142
5.3.16. Repair of the ASR captions becomes the focus of interaction 144
5.3.17. Actually helpful 144
5.3.18. That wasn't bad actually 147
5.4. Thematic Analysis 151
5.4.1. Theme 1 - Speech and identity 153
5.4.2. Theme 2 - Strategies for communication 153
5.4.3. Theme 3- Hopes for personalised ASR 154
5.4.4. Theme 4 - Set-up and first use 155
5.4.5. First use of Relate 156
5.4.6. Design of Relate 157
5.4.7. Theme 5 - Overall experience of using ASR in practice 159
5.5. Research – Summary 160
6. Discussion 162
6.1. Introduction to the chapter 162
6.2. ASR model performance 165
6.2.1. AI and AAC – Speed and Accuracy 170
6.3. Conversation Analysis 173
6.3.1. Trouble and Repair 173
6.3.2. Allocation of responsibility 180
6.3.3. Gaze 183
6.3.4. The understandability of ASR as AAC 186
6.4. Thematic Analysis 190
6.5. Usability 193
6.5.1. Effective to use (effectiveness) 193
6.5.2. Efficient to use (efficiency) 193
6.5.3. Having good utility (utility) 194

6.5.4. Easy to learn (learnability) 195
6.5.5. Easy to remember how to use (memorability) 195
6.5.6. Enjoyable to use (satisfaction) 195
6.5.7. Safe to use (safety) 196
6.5.8. Usability summary 197
6.6. Discussion – Summary 197
7. Limitations 204
8. Recommendations 206
9. Implications for future research 208
10. Conclusion 210
References 213
Appendix 235
A1. Advert 235
A2. Participant Information Sheet 236
A3. Consent form 242
A4. Question Sheet 245
A5. Scoping Review 1990 (Jan 1) - 2020 (Nov 15) 246
Literature Eligibility Criteria 246
Information Sources 246
Search Strategy 247
Selection of sources of evidence 248
Summarizing the results 248
Results 1990 (Jan 1) - 2020 (Nov 15) 249
Table 3: Study types, participants, technology used, method. 258
Table 4: Participants' communication needs and studies' outcomes. 259
Table 5: Studies' advantages and disadvantages and future studies. 260
A6. Scoping Review 2020 (Nov 16) - 2024 (May 31) 261
- Results 2020 (Nov 16) - 2024 (May 31) 261
- Research to improve ASR accuracy 262
- Using ASR to support diagnosis 277
- Using ASR to support SLT assessment 279
- Using ASR as AAC 286

- How plwMND already use ASR technology 289
Table 7: Participants' communication needs and studies' outcomes. 296
Table 8: Studies' advantages and disadvantages and future studies. 300

Glossary

Term	Explanation
AAC	Augmentative and Alternative Communication. AAC is a set of tools and strategies designed to supplement speech and writing if these are impaired (ISAAC, 2022).
AI	Artificial Intelligence. A branch of computer science focused on developing machines capable of performing tasks that typically require human intelligence, such as learning, reasoning, and problem-solving (Russell, 2022).
ASR	Automated Speech Recognition. ASR technology is the process where a device such as a smartphone or computer converts a speech signal into words in text (Young & Mihailidis, 2010).
BCI	Brain-Computer Interface. A system enabling direct communication between the brain and an external device, bypassing traditional neuromuscular pathways, with potential applications in AAC for individuals with severe motor limitations.
CA	Conversation Analysis. A systematic data-driven methodology for examining the structural organization of naturally occurring conversations, identifying and describing the structures and procedures with which participants themselves achieve order in interaction (Schegloff, 1987).
CER	Character Error Rate. A metric used in ASR evaluation to measure the percentage of character errors (substitutions, insertions, and deletions) in the recognized text compared to the reference transcription.
DNN	Deep Neural Network. A type of artificial neural network with multiple layers, enabling the learning of complex patterns and representations from data, often used in ASR systems (Russell, 2022).
Dysarthria	Dysarthria is the collective term for a group of neurologic speech disorders that result in abnormalities in the movement required for the breathing, phonatory, resonatory, articulatory, or the prosodic aspects of speech production, including strength, speed, range, steadiness, tone, or accuracy (Duffy, 2020).
FDA-2	Frenchay Dysarthria Assessment - Second Edition. A standardized assessment tool used to evaluate the type and severity of dysarthria (Enderby & Palmer, 2008).
LLM	Large Language Model. An LLM is a type of artificial intelligence model designed to understand, generate, and interpret human language. These models are trained on vast amounts of text data to predict and generate text based on context (Humza Naveed et al., 2024).
MND	Motor Neurone Disease (also known as Amyotrophic Lateral Sclerosis or Lou Gehrig's disease) is a progressive and ultimately fatal neurological disease causing muscular weakness resulting in loss of function of limbs, trunk and neck. More than 80% of people with MND will become unable to use speech for communication (Oliver et al., 2017)
OIOR	Other-Initiated Other-Repair. When a recipient of the trouble source both initiates and resolves its repair.

Term	Explanation
OIR	Other-Initiated Repair. A conversational repair sequence where someone other than the speaker of the trouble source initiates the repair, often observed in interactions involving individuals with dysarthria.
OISR	Other-Initiated Self-Repair. When a recipient/hearer of the trouble source initiates its repair, but the speaker of the trouble source executes the repair himself/herself
PER	Phoneme Error Rate. A metric used to evaluate the accuracy of speech recognition systems at the phoneme level. It quantifies the difference between the captioned phoneme sequence and the reference (correct) phoneme sequence.
plwMND	Person/People living with Motor Neurone Disease.
PSV	Personalized Synthetic Voice. A synthetic approximation of a person's natural voice (Costello, 2016), for use in an AAC device, to provide an AAC device user with a personal voice that no other augmented communicator is likely to be using (Bunnell et al., 2005).
Repair	The term repair is used in Conversation Analysis to describe the practices used by participants to highlight and then deal with, or manage, problems or troubles within interaction. Repair itself is designed for efficiency and successful resolution of problems when they arise (Schegloff, 2020).
Responsibility	In Conversation Analysis, the term responsibility refers to the accountability a speaker takes or is perceived to take for the actions they perform within a conversation. Different relational stances can lead to different practices of repair initiation resulting in different behavioural consequences (Robinson, 2006).
SIOR	Self-Initiated Other-Repair – when the speaker of the trouble source is responsible for initiating the repair, but its resolution is brought about by a recipient/hearer
SISR	Self-Initiated Self-Repair. When both initiation and resolution of repair are carried out by the speaker of the trouble source.
Stance	Stance refers to ways that people position themselves in conversation, often in terms of politeness, certainty, or affect/emotion (Kiesling, 2022).
TA	Thematic Analysis. A qualitative research method for identifying, analyzing, and reporting patterns (themes) within qualitative data (Clarke & Braun, 2013).
Trouble	Everyday conversation is rarely produced without problems or errors - called 'troubles' in Conversation Analysis. It is common to observe troubles in speaking, hearing or understanding talk in conversation (Schegloff et al., 1977).
TTS	Text-to-Speech. Technology that converts written text into spoken language, often used in AAC devices.
VOCA	Voice Output Communication Aid. A VOCA is a type of AAC that is often provided to plwMND (Makkonen et al., 2018). A VOCA is operated by choosing letters, words, symbols, or sentences in the most accessible way for the person. The VOCA then 'speaks' the message using a synthetic representation of a generic human voice.

Term	Explanation
WER	Word Error Rate. WER is a common metric used to evaluate the performance of speech recognition systems. WER measures of the number of full word matches between the hypothesized model transcript and the ground-truth human transcript. Formally, it is calculated as the summation of substituted, deleted, and inserted words divided by the total number of words in the ground-truth transcript (Mishra et al., 2011).
WRR	Word Recognition Rate. WRR is the inverse of WER and measures the percentage of words that were correctly recognized by a system, as opposed to focusing on the errors (Mishra et al., 2011).

1. Introduction

1.1. Introduction to the chapter

This chapter presents a background to this research. It describes the difficulties people with MND (plwMND) face as their speech declines, how it may negatively affect their sense of personal identity and impacts on personal and social networks. PlwMND often use alternative assistive technology to communicate when speaking is difficult but it is usually much slower than natural speech and can present additional difficulties. Automatic speech recognition (ASR) technology may be a promising way forward to support plwMND to keep talking and be understood, however there are already known factors that may affect ASR accuracy and usefulness. Artificial Intelligence (AI) driven ASR technology such as Project Relate is argued to help people with impaired speech to be better understood. However it has not been assessed with plwMND in everyday conversation and over time and as the disease progresses. These are the objectives for this research. Conversation Analysis, an assessment approach used in this research is additionally discussed in general terms in the chapter.

1.2. Motor Neurone Disease

Motor neurone disease (MND), also known as Amyotrophic Lateral Sclerosis (ALS) or Lou Gehrig's disease is a progressive and ultimately fatal neurological disease causing muscular weakness resulting in loss of function of limbs, trunk and neck. At present, there is no cure and the average life expectancy from symptom onset is commonly two to three years, with approximately 25% of people surviving for five years and 10% surviving ten years (Oliver et al., 2017).

Approximately one third of people living with MND (plwMND) experience motor speech problems (known as *dysarthria*) as an initial symptom (Da Costa Franceschini & Mourão, 2015) and most experience dysarthria sometime during the course of the condition (Ball et al., 2004b). As a result, more than 80% of plwMND will become unable to communicate their daily needs using natural speech (Beukelman et al., 2011), and in time most will be unable to speak at all (Beukelman et al., 2007). Dysarthria is the collective term for a group of neurologic speech disorders that result in abnormalities in the movement required for the breathing, phonatory, resonatory, articulatory, or the prosodic aspects of speech production, including strength, speed, range, steadiness, tone, or accuracy (Duffy, 2020).

Speech decline can be rapid. Many plwMND will lose functional speech for communication within eighteen months from first symptoms, which may develop before an MND diagnosis is confirmed (Makkonen, Ruottinen et al., 2018). In addition to progressive dysarthria, and physical decline, approximately 15% plwMND will experience frontotemporal dementia (Bäumer et al., 2014) and up to 40% present with mild to moderate cognitive impairment (Pinto-Grau et al., 2018).

1.3. Augmentative and alternative communication (AAC)

Almost all plwMND eventually rely on augmentative and alternative communication (AAC) to support daily communication (Ball et al., 2004a). AAC is a set of tools and strategies designed to supplement speech and writing if these are impaired (ISAAC, 2022). AAC ranges from unaided systems not requiring technology or external equipment, such as signing, vocalizations, speech approximations and gesture, to aided systems which do require some form of equipment or technology, such as low-tech picture or letter charts, through to complex high-technology devices include such as on a phone, tablet computer and accessing a wide range of applications and digital communication media. These devices may be accessed directly by selecting items with a finger, toe, facial gesture or eyes (for example), or indirectly using a head mouse or switches, or indeed a combination of these (Beukelman, 2013; Communication Matters, 2020).

PlwMND typically use AAC that enables them to compose messages using a keyboard accessed through typing, eye gaze, or by scanning using a switch (Beukelman, 1998). Switch access refers to a specialist way of accessing a computer or communication aid using a simple switch that has two states – on and off. Switches are used by people who have a degree of physical impairment that means that they are unable to directly access a keyboard or any type of pointer control (BLTT, 2017).

The National Institute of Health Care and Excellence (NICE) guidelines for the assessment and management of MND, recommend that AAC equipment is provided without delay if appropriate to meeting the needs of a plwMND. As there are many different types of high and low technology AAC available, assessment is frequently conducted by a Speech and Language

Therapist (SLT) in partnership with the plwMND, the multidisciplinary team and significant others. The priority is that AAC should maximise participation in activities of daily living and maintain a person's quality of life (NICE, 2016).

A Voice Output Communication Aid (VOCA) (a type of AAC) is often provided to plwMND (Makkonen et al., 2018). A VOCA is operated by choosing letters, words, symbols, or sentences in the most accessible way for the person (which could be using fingers, a keyboard, pointer, adapted mouse, joystick, a switch, or eye tracking). The VOCA then 'speaks' the message using a synthetic representation of a generic human voice. NICE does not specify the type of voice that a VOCA should have (NICE, 2016).

Although VOCAs have been available since the 1980s, technology advances supported a rapid expansion in the choice and functionality from the early 2000s (Judge & Townend, 2013). It is argued that VOCAs improve increased quality of life by promoting autonomy, allowing for greater participation in the decision-making process, improving self-esteem and helping to support relationships (Brownlee & Palovcak, 2007).

1.4. Identity

1.4.1. Speech and identity

Speech is a powerful medium of identity (Bucholtz & Hall, 2005). It communicates mood, humour, geographical, social and educational background, health status, gender - as well as the content of the message (Nathanson, 2017). It contributes to a sense of self, while allowing the listener to derive multiple levels of meaning from a communication.

PlwMND often consider the likely loss of speech as one of the worst aspects of MND (Hecht et al., 2002). When it becomes difficult to communicate using speech, individual and social identity can be negatively affected, increasing the risk of social withdrawal (Yamagishi et al., 2012). Identity is constructed and mediated through social interaction and is a context for our actions and interactions with others (Wertsch, 1993) and a failed social exchange can carry the message of inferiority (Gee, 2010), resulting in reduced participation in social interactions (Shadden, 2005).

Discrepancies between how people are perceived and how they themselves feel based on their speech has been reported as creating an identity challenge (Miller et al., 2011). Speech is a means to express identity, and in turn how others react to us shapes our self-identity (Miller et al., 2011). Listeners may assign a different social and psychological persona to a person with impaired speech due to neurological impairment (Mariën et al., 2019).

Leite et al. identified a correlation between the increase in speech difficulties experienced by plwMND and a reduction in quality of life (QoL) (Leite Neto & Constantini, 2017). Changes to the quality of speech has also been associated with a negative perception by listeners in terms such as being less intelligent, likeable, trustworthy than those with unaffected speech (Amir & Levine-Yundof, 2013).

1.4.2. AAC and identity

Although AAC may help plwMND to communicate as talking becomes difficult to understand or when functional speech is lost, AAC cannot replace natural speech even with recent advances such as voice banking and can impact a person's sense of identity (Pullin et al., 2017).

Maintenance of identity can become more difficult as MND progresses and communication with others becomes more difficult (Kane et al., 2017). While high technology AAC such as a VOCA is argued to compensate for loss of speech, basic communication strategies without voice substitution are often preferentially used by plwMND, such as written communication using a pen and notepad, head-nodding, and letter boards (Simmons, 2013). This is because high technology AAC may not be appropriate given the individual's strengths and needs - for example when considering the psychological challenge of learning to communicate primarily via typed or selected text, or communicating with a synthetic human voice that may not align with the personal identity, accent or culture of the plwMND (Pullin et al., 2017).

Using speech to communicate may remain the first preference, even for those who are significantly difficult to understand (Smith & Connolly, 2008), even when speaking is effortful and they are often not understood (Beukelman, Fager, et al., 2007).

AAC devices are not well equipped to facilitate talk during verbal conversation due to the rapid timing demands of such interactions (Seale et al., 2020). This is because using an AAC device is slow, around 8–10 words per minute (wpm), perhaps 20 wpm when using word prediction, especially when compared to the average speaking rate of between 125 and 185 wpm (Kristensson et al., 2020). As a result, identity maintenance and relationship management can be impaired by the limitations of AAC devices as well as the disease (Kane et al., 2017).

People may use vocalisations and gesture to try to speed up communication with AAC (Seale et al., 2020). However, the effort required to speed up communication may influence whether a person chooses to communicate at all, and if they do, how much they should contribute to a conversation. Many may decide not to communicate beyond their basic needs into a more

extended conversation (Waller, 2019), and 'social closeness' - the maintenance and development of personal relationships - may be negatively impacted (Murphy, 2004).

Clarke et al. found the decisions whether to communicate or not using AAC were based on the perceived importance and value of the message, the time available to communicate, who the individual is speaking to, previous experience within that situation, and the personality of the individual. The inability to spontaneously contribute to conversation was perceived negatively, can be frustrating, and may influence whether an individual then decides to contribute (Clarke et al., 2019).

AAC may not be designed in a way that supports the observed communication preferences of people with dysarthria. For example, current AAC provision lacks the ability to easily augment existing speech and enable a person to combine other available modalities they have such as facial gesture or pointing (Clarke et al., 2023). This can lead to a person having to choose which modality to use to communicate. Many may choose to use speech first in many situations (and only speech), even if it is difficult for others to understand (Clarke et al., 2019).

As motor skills deteriorate due to the effects of MND, different physical movements may be needed to operate AAC (Ball et al., 2004a). This often increases the time needed to communicate, and the effort required to learn to operate AAC in new ways may mean fewer plwMND choose to use AAC at all (Felgoise et al., 2016; Linse et al., 2018).

Delays that extend beyond a few seconds increase the likelihood for partner inattention, frustration, and misunderstanding, presumptions of incompetence, stigma, and social isolation (Robillard, 1999). A pause of over three seconds can result in an awkward silence, while pauses of around 0.5 seconds can disrupt the flow of conversation (Jefferson, 1986)

As an example, medical professionals communicating with patients using AAC were reported to be less tolerant with the reduced speed of communication, frequently speaking to the carer instead, and made inappropriate assumptions about people using AAC and their cognitive level (Morris et al., 2013). Because of these difficulties, some plwMND may decide to restrict their interactions using AAC to a limited range of people (McNaughton et al., 2001).

Psychosocial factors are also argued to play an important role in successful use of AAC, including the confidence, motivation and attitude to communicate using AAC, and resilience in the face of difficulties or setbacks when using AAC (Light & McNaughton, 2014).

Given that plwMND frequently express a strong desire to use their own speech for as long as possible, even when unintelligible to others, it can be argued that strategies that support speech should always be encouraged and supported (Murphy, 2004), including AAC that could augment understandability when using speech.

1.4.3. Voice Banking and identity

Voice banking (also sometimes termed as 'Voice Cloning' or 'Voice Preservation') is a process for creating a 'personalised synthetic voice' (PSV), a synthetic approximation of a person's natural voice (Costello, 2016), for use in an AAC device, to provide an AAC device user with a personal voice that no other augmented communicator is likely to be using (Bunnell et al., 2005).

Although some commercially available synthetic speech models are argued to approach being perceptually indistinguishable from natural human speech (Bhagtani et al., 2024), this level of technology currently remains unavailable to be used offline within AAC devices. A PSV used in

AAC does not sound the same as how people communicate with natural speech even if the PSV has a similar tone to natural speech: a PSV cannot currently be dynamically adjusted to indicate happiness, anger, sadness, irony or sarcasm to name a few examples (Pullin & Hennig, 2015).

For a person used to speaking naturally, communicating using a synthesised voice can seem impersonal and unfriendly (Mckelvey et al., 2012) and is one of a number of factors linked to AAC acceptance and use (Linse et al., 2018). In a study where listeners compared the short sentences recorded by the same person using their natural voice and PSV, the natural voice recording was preferable perceptually to listeners even though the sentences were equally intelligible (Overton et al., 2017).

Voice banking is becoming increasingly popular for plwMND. For example, In 2023 more than 1,000 plwMND received funding from the MND Association for voice banking, and more are forecast to bank their voices in 2024 (Brown & Cave, 2024). This interest may be because they perceive a voice bank could help to maintain a sense of personal identity by preserving the way that they *speak* - as opposed to what a voice bank really does - preserve elements of the way their voice *sounds* for use on an AAC device that is significantly slower than talking (Cave & Bloch, 2021b).

People who choose to voice bank may also not appreciate that a PSV is unlikely to change the ease, speed or accuracy of communication once they use AAC (Cave & Bloch, 2021b).

However, the increasing demand for voice banking indicates that preservation of the characteristics of speech appears important to many plwMND as speech becomes progressively affected by dysarthria and harder to understand by others, and devices for communication start to be used (Cave & Bloch, 2021b).

1.5. AAC Speed versus Accuracy

Although most plwMND eventually rely on AAC to support daily communication (Ball et al., 2004a), for many there is nothing that could replace the ease, accuracy and speed of natural speech, citing the extended time it takes to spell a message with AAC (Murphy, 2004), and the slowness of using AAC to communicate can be a barrier to AAC adoption (Peters et al., 2015).

An AAC user may attempt to produce messages faster by typing short phrases, forming sentences that may not be syntactically complete, or using preprogrammed messages that may or may not reflect what they mean to convey. People may prioritise speed over accuracy of communication in an attempt to maintain their personal sense of identity and increase the depth and sophistication of social inclusion (Fried-Oken et al., 2023).

However, by prioritising speed, the communication itself may be less accurate, or the output may not fully overlap with their communication intent. Therefore, enhancement of speed may come at the expense of accuracy and potentially sense of identity (Fried-Oken et al., 2023).

Some AAC users may prefer to prioritise accuracy over speed. Accurate communication with an AAC device is argued to be one of the few ways a person with advancing MND can demonstrate competency through proper spelling, grammar, and pronunciation (Kane et al., 2017). The effort of ensuring accuracy of words, precision of punctuation and grammar when communicating with AAC is sometimes used as a way of demonstration how much a person cares for another person (Kelly et al., 2017). Accuracy of communication may also support preservation of a sense of personal identity – communicating precisely what it is a person wishes to share (Moore et al., 2018).

If an AAC device generates inaccurate, inappropriate or misleading communication, there may be a risk to valued relationships, reputation, as well as the risk of anticipated harms to health. These concerns are in addition to the frustration of generating inaccurate messages and the extended time of having to correct them (Fried-Oken et al., 2023).

Maintenance of personal identity is therefore argued as a reason to both prioritise speed over accuracy and accuracy over speed of communication. The sense of trade-off from the speed of natural speech to slower or potentially less accurate communication through AAC may feel particularly important to plwMND as they experience progressive communication challenges (Fried-Oken et al., 2023).

Despite decades of research, the technical problem of significantly increasing the speed of communication while maintaining accuracy has not been solved (Waller, 2019). As a result, there also appears little change to risks associated with significantly slow rates of communication when using AAC - social isolation, presumptions of incompetence, stigma, partner inattention, frustration, and misunderstanding, and ultimately the risk of negative impact to psychological wellbeing and sense of identity (Robillard, 1999).

For some, the best AAC system could still remain the human dyad - comprising speaking and non-speaking partners who know each other well (Waller, 2019).

1.6. Artificial Intelligence and AAC

Artificial intelligence (AI) is a field of computer science that focuses on the development of intelligent machines that can perform tasks that typically require human intelligence (Sennott et al., 2019). AI involves the use of algorithms and models to enable computers to learn from

data, recognize patterns, make decisions, solve problems autonomously, and can be used in various applications, such as speech recognition, natural language processing, medical diagnosis, and decision-making. (Azevedo et al., 2024).

AI technology is argued to have the potential to enable increased speed of AAC communication while also maintaining accuracy (Koch Fager et al., 2019; Light et al., 2019). AI may be able to maintain accuracy by automatically generating more contextually accurate text suggestions based on what the AAC device may hear, see, or contextually recognise, supporting the user to rely on AI-generated phrases that fit their communicative intent. Speed may be increased by reducing the requirement of manually typing phrases, or editing pre-stored messages (Valencia et al., 2023).

AI may also potentially support a reduction of the cognitive effort of using AAC in various ways: by automatically creating appropriate AAC-generated responses during an ongoing conversation (Valencia et al., 2023), or by identifying an individual's intrinsic needs and skills, as well as extrinsic variables in the environment, and suggesting appropriate, individual adaptations in real time (Koch Fager et al., 2019). For example, AAC may utilise AI to assess ambient noise and propose adjustments to speech volume and/or voice (Light et al., 2019).

1.7. Automated Speech Recognition (ASR)

The inclusion of AI technology is argued to enable personalized automated speech recognition (ASR) technology that better recognises the unique speech patterns of individuals with 'non-standard' speech, potentially including people living with MND (MacDonald et al., 2021). Used as AAC, ASR technology may provide an opportunity for plwMND to be better understood by others for longer while still using natural speech to communicate (Cattiau, 2021).

ASR technology is the process where a device such as a smartphone or computer converts a speech signal into words in text (Young & Mihailidis, 2010). ASR systems are driven by an acoustic model (AM), which maps events in the acoustic signal to speech (sounds or words), and a language model (LM), which predicts the frequency of common words or phrases (Tatman, 2021).

ASR technology has multiple uses: it can convert speech to text, find information online, enable hands-free computing such as dictating emails, automatic language translation and many other uses. It can be helpful for environmental control, such as changing the channel on a television, adjusting lighting and heating (Laricchia, 2020).

Voice interaction with devices is becoming ubiquitous. By the end of 2024, it is forecast people will be able to interact with over 8.4 billion devices using their voice - larger than the world's population and double the number of devices for 2020 (Moar & Escherich, 2020). Although smartphones are the largest group of devices that people will interact with using voice, cars and household devices such as televisions are forecast to have the highest rate of growth (Moar & Escherich, 2020).

ASR technology has rapidly developed over the past three decades. Launched in 1990, Dragon Dictate is regarded as the first consumer speech recognition product (Pinola, 2011). In 1993, 'Speakable items', was introduced as a built-in speech recognition and voice enabled control tool for Apple computers (Neuburg, 2000). Microsoft introduced ASR in the Windows operating system in 2007 (Littlejohn Shinder, 2007), and in 2011 Apple's Siri 'digital personal assistant' was integrated into the iPhone smart phone (Golson, 2011). In addition to being able to recognize speech, Siri was claimed to understand the meaning of what was said and take an

appropriate action (Daw, 2011). Amazon launched Alexa in 2014, claimed to be similar in functionality to Siri, as did Microsoft with the Cortana in the same year (Warren, 2014). 'Google Assistant' was launched in 2016 (Wallace, 2016). As at 2019, Amazon's Alexa digital assistant provided more than one hundred thousand different activities that can be controlled by ASR technology (Laricchia, 2020).

1.7.1. Dysarthria and ASR

Research on increased ASR accuracy for speech with dysarthria has been researched over the past two decades. The STARDUST (Speech Recognition for People with Severe Dysarthria) project developed a speech recognition model for people with severe dysarthria to understand a small vocabulary of ten to twelve individual words that were selected for easier access to environmental control devices (Parker et al., 2006). None of the participants were reported to have MND. Speaker dependent recognition models were trained with the speaker's own output and each target word could was represented by any phonetic realization as long as sufficient phonemic contrast was present to distinguish it from any other target word. A computerized training package was also used to build consistency of utterance of each target word production utilizing visual and auditory feedback. In this way, STARDUST did not require intelligibility of speech for machine recognition and was limited in the number of words it could recognize. The results showed that participants experienced a significant decrease in the time taken to complete control tasks, such as switching on the TV, and the results of the training package increased computer recognition rates by an average of 5%, though some participants observed no gain from training.

Hosom et al. devised Supplemented Speech Recognition (SSR), combining ASR optimized for dysarthric speech, alphabet supplementation, and word prediction. The first letter of the target word was typed and then the whole word spoken. Coupling ASR with first-letter identification decreased the dictionary set from which an ASR system had to match the acoustic signal. The set was further reduced by using word prediction algorithms. SSR put the most likely word into the line of text and provided five further probable word choices in word prediction boxes. The seven participants in the study had mixed or spastic dysarthria with variation in intelligibility ratings from mild to severe. None was reported to have MND. The results were measured in the number of keystrokes saved compared the total number of keystrokes to type each word fully. Keystroke savings using SSR typical averaged 68%, though the researcher gave assistance to accurately select from the touch-screen computer (Hosom et al., 2010).

The VIVOCA project (Hawley et al., 2013) combined training of an ASR for a small vocabulary of up to 47 individual words with predictive message building based on the spoken input. The results were passed to a speech synthesizer, producing intelligible spoken output. To increase the accuracy and consistency of the ASR model, participants were given training to motivate speech consistency and modification where helpful - for example encouragement to replicate their 'best attempt so far'. The participants in the study had moderate to severe dysarthria and low conversational intelligibility. No participant was reported to have MND. The study achieved a word recognition accuracy in excess of 85% for the words in the study, and subsequently the 'Vocatempo' app was developed to support accessibility and usability of the VIVOCA concept (Bright et al., 2019).

Voiceitt has a goal to integrate their ASR software for people with 'non-standard speech' with smart devices, enabling better access voice recognition tools including Amazon Echo or Google

Assistant, and are beta-testing their release (VoiceITT, 2020). Voiceitt are a partner in the Nuvoic project, a European Union funded initiative to develop ASR software to support people with mild dysarthria and ageing voices and better access connected home devices (VoiceITT, 2020).

More recently, Google's Euphonia research project has been focusing on ways to improve ASR accuracy in recognizing dysarthric speech (Cattiau, 2019; Green et al., 2021; MacDonald et al., 2021). Both VoiceITT and Euphonia projects may present an opportunity for people with dysarthric speech to be better understood by others, supporting the goal of AAC to retain, as far as possible, the speed and, ideally, the naturalness of spoken communication (Hawley et al., 2013).

ASR word accuracy for commercial systems is now as high as 95% for many speakers with unimpaired speech (Synnaeve et al., 2020), and has improved significantly over recent years due to increased computational power of deep learning systems and the availability of large training datasets (MacDonald et al., 2021). However, for people with dysarthria, ASR technology may still make too many word recognition errors to be of practical benefit (Ballati et al., 2018b).

1.7.2. Dysarthria and Word Error Rate

ASR model performance is frequently evaluated using Word Error Rate (WER) (Mishra et al., 2011). WER is a measure of the number of full word matches between the hypothesized model transcript and the ground-truth human transcript. Formally, it is calculated as the summation

of substituted, deleted, and inserted words divided by the total number of words in the ground-truth transcript.

Very high word error rates were measured for people with significant dysarthria when using three commercially available ASR systems (IBM Watson, Google Cloud, and Microsoft Azure Bing). For individuals with severely dysarthric speech the WER rates were between 78- 89%, resulted in a very low percentages of correctly transcribed sentences – between 0-1.2% (De Russis & Corno, 2019). The poor speech recognition was attributed to the diversity and complexity of atypical speech patterns (Duffy, 2020), and insufficient training data - when fewer examples are available to train the ASR system how a specific group speaks, ASR accuracy decreases and WER increases (Tatman, 2021).

However, voice-based access could be particularly helpful for people with dysarthria as many may have a co-occurring physical impairment. For MND, dysarthria can be observed as an initial sign in up to 30% of plwMND. Progressive muscular weakness of the limbs, trunk and neck is typical for most plwMND (Oliver et al., 2017) and almost all will have dysarthria in later stages (da Costa Franceschini & Mourão, 2015).

Dysarthria with co-occurring physical impairment is also common in other aetiologies. For example, up to 60% of people living with a stroke may have dysarthria (Safaz et al., 2016) and up to 75% of post-stroke patients may have functional limitations of the upper and lower limbs (Chen et al., 2019). Up to 65% people with Traumatic Brain Injury may have dysarthria (Mitchell et al., 2017), over 70% of people with Parkinson's disease (Müller et al., 2001) and up to 50% of people living with Multiple Sclerosis (Yorkston, 2010).

The challenge of ASR better recognising dysarthric speech may be a difficult problem to solve, as articulation, tempo, rhythm, and volume are frequently highly variable (Moore et al., 2018). While some levels of ASR accuracy for dysarthric speakers have been reported, they involved speakers with mild dysarthria or tasks with very limited vocabulary (Koch Fager et al., 2019) .

The speech of plwMND may be particularly difficult for ASR technology to interpret due to progressive dysarthria and the changing strategies plwMND may use to be understood (Caballero-Morales & Trujillo-Romero, 2014). Fatigue, frequently experienced by plwMND may also increase speech variability (Sanders et al., 2002).

However, the potential benefits of ASR technology for plwMND could be significant. Hands-free interaction has been cited as the most popular reason for using ASR devices (Olmstead, 2017) and for plwMND, this is an important consideration as around two-thirds of plwMND experience limb weakness as an initial and progressive symptom (Bäumer et al., 2014). ASR may help plwMND maintain some control of their home environment as more devices incorporate ASR technology, for example, using voice commands to switch lights on or off, or adjust the thermostat (Hawley et al., 2007) , or controlling their wheelchair (Simpson et al., 2008).

1.7.3. Additional factors that may affect ASR

In addition to dysarthria, other factors can affect ASR recognition accuracy. WER may increase significantly with non-native speakers of a language due to variation of pronunciation and accent features (Radzikowski et al., 2021). In one study, speakers who self-identified as having

Indian English accents experienced higher WER versus speakers who self-identified as having American English accents (Meyer et al., 2020).

Ethnicity and geographical background can affect ASR performance. For example, speakers from Scotland were found to have lower English ASR performance than speakers from New Zealand and the United States (Tatman, 2017).

ASR systems from Amazon, Apple, Google, IBM, and Microsoft had an average WER of 35% for black speakers compared with 19% for white speakers (Koenecke et al., 2020). The ASR was less able to recognise spoken words due to phonological, phonetic, or prosodic characteristics of African American Vernacular English. This may be because of insufficient audio data from black speakers when training the underlying ASR models. Grammatical or lexical differences did not result in significant WER differences between these groups (Koenecke et al., 2020).

Gender has been inconsistently associated with ASR performance in English, performing significantly better for male speakers over female speakers (Tatman, 2017), Female speakers over male speakers (Koenecke et al., 2020), or no significant difference (Meyer et al., 2020).

ASR technology has been argued to influence how people speak a language (Philip et al., 2021). If an English speaker does not speak with the accent deemed 'standard', the ASR may not consistently recognise the words and the speaker may feel compelled to mimic the dominant accent in order to be able to make use of the ASR (Philip et al., 2021). One study identified significant WER differences between speakers whose first language is English and those whose first language is not. The disparities were argued to be linked to the age at which an individual began speaking English, the environment they learned it in, and whether or not their first

language was Germanic, but also whether or not their birth country is a part of NATO, a political alignment with the United States (DiChristofano et al., 2022).

1.7.4. ASR and Identity

As ASR technology becomes increasingly mainstream, the effect of ASR performance gaps between groups may extend to a sense of negative psychological well-being as a result of a perceived discriminatory experience. Speakers are reported to sometimes link errors that an ASR makes to their own sense of identity, including racial, regional and location identity, and many participants modified their dialect and word choice in order to be understood by ASR (Mengesha et al., 2021).

Amazon claims that Alexa's understanding will improve as more people speak to it (Harwell, 2018). However, if one group finds more functional benefit from ASR technology than another, then that group is likely to use it more and the performance gap is likely to increase (Harwell, 2018). By selling products that work for one accent above others, technology companies make speakers of other accents less likely to adopt their products (DiChristofano et al., 2022).

For individuals with impaired speech, the more normalised voice interaction becomes, the more they risk being excluded from these current, commercially available technologies - even though they could benefit extensively (Koch Fager et al., 2019).

At a societal level, inequality in access to technology risks increasing the 'digital divide': worsening existing divisions and increased social exclusion (United Nations, 2016).

1.7.5. ASR and equality of service

The results of ASR inaccuracy and exclusion can be wide ranging. For example, patient harm as a result of inaccurate ASR transcription of patient notes (Topaz et al., 2018), or situational inequality when ASR is used by employers to automatically evaluate candidate interviews or by criminal justice agencies to automatically transcribe courtroom proceedings. (Koenecke et al., 2020).

ASR inequality may risk inequality in access to medical care: in Denmark, ASR technology is being used to listen to calls to emergency medical services (EMS) and output a prediction on the potential presence of cardiac arrest. The usage of ASR in this way has been demonstrated to increase earlier stroke recognition leading to better patient outcomes. (Cleve et al., 2020). However, there were no details in the study on accuracy of ASR in the calls and between callers, and if variations in WER could have had a variation in stroke recognition between speaker groups.

In recent years initiatives have been set up to help train ASR technology to better recognise diverse speech. For example, Common Voice is an open and free global database of donated voices with the goal of enabling ASR Technology to be more equitable and inclusive as a service to all of humanity (Mozilla, 2021) .

Other providers offer recordings of granular accent variation to test ASR effectiveness, for example recordings of twenty-nine different Spanish accents (Newswire, 2021).

However, inequality between groups may not be removed even when ASR models are trained with equal numbers of ethnic voices - AI prediction accuracy remained unequal even with equal numbers of African Americans and White Americans (Li et al., 2022). ASR technology may

unintentionally sustain or even worsen existing societal inequalities (Zou & Schiebinger, 2018), and illustrates the complexity of attaining equal access to ASR technology.

It remains unclear how the known ASR disparities for people with unimpaired speech may apply to plwMND with dysarthria. This is a matter for future research and the question may be to better understand how the factors detailed would not apply to people with dysarthria. Indeed, there may be added risk of inequality of ASR recognition for speakers with dysarthria due to an imbalance and number of recordings of different types of dysarthria made available to the ASR (MacDonald et al., 2021).

1.7.6. Project Euphonia

Google's Euphonia research project is focused on improving ASR accuracy in recognizing dysarthric speech (Cattiau, 2019). The project aim is for people that have speech with dysarthria could be better understood by others and supports the goal of AAC to preserve the speed and naturalness of spoken communication as far as possible (Hawley et al., 2013), and support 'social closeness' the preference to use own speech to communicate rather than through AAC (Murphy, 2004).

The Euphonia dataset has speech samples from over one thousand individuals with impaired speech, more than one million utterances, over 1300 hours of recordings (MacDonald et al., 2021). Using a two-step training approach that starts with a baseline 'standard' corpus and then fine-tunes the training with a personalized speech dataset, the resulting speech recognition model is argued to better understand non-standard speech (Google AI Blog, 2019). In this way, an individual is not required to train the speech recognition model word by individual word,

opening the possibility for near-real time ’translation’ of dysarthric speech as it is spoken in daily conversation.

Increasing the speed of AAC to support dysarthric speech comprehension may contribute to a reduction of the physical, cognitive and emotional effort required by an individual with dysarthric speech – factors that drive the potential for successful communication (Clarke et al., 2019).

Personalised ASR models generated by Euphonia for people who have recorded at least 300 phrases improved average WER by about 75%. Over 80% of these personalised models have WER below 15%, resulting in a WER close to speakers with unimpaired speech (MacDonald et al., 2021).

However, it is important to note that WER scores were assessed on shorter scripted phrases, not conversational speech, therefore the WER may not be reflective of how accurate Euphonia ASR is in everyday communication (Green et al., 2021).

Additionally, more than a quarter of the participants' speech resulted in WER greater than 10%, indicating that multiple factors may affect WER at even short phrase level, including severity of dysarthria, variability of prosody, articulation and phonation. Recording quality also matters, including low signal to noise ratio overlapping speech, environment noise, power of the phone/device and the number of recordings available to create individual ASR models (Green et al., 2021).

To create Euphonia, the team partnered with many non-profit groups to help raise awareness of Euphonia across communities, including MND Association, ALS-TDI, Team Gleason, ALS

Association, Canadian Down Syndrome Society, and LSVT Global, among others (MacDonald et al., 2021).

However, there is a significant imbalance between aetiologies in the collated Euphonia recordings. The largest group of people recording for Euphonia is from the MND community (29.9% of participants), followed by people with Down Syndrome (18.1%), Parkinson's Disease (11.4%), Stroke (1.6%) or traumatic brain injury (0.9%) (MacDonald et al., 2021). For context, the estimated number of MND cases diagnosed globally is about 145,000 each year, compared to Stroke, estimated at around 12.2 million (International Alliance of ALS/MND Associations, 2021; World Stroke Organization (WSO), 2022).

Additionally, Euphonia uses recordings of scripted speech. This approach reduces linguistic diversity and may not capture the words people use or speaking style in conversational speech. However, it does reduce transcription cost and time, a significant consideration when collecting such a large volume of recordings (MacDonald et al., 2021).

1.7.7. Project Relate

The Euphonia team created an Android app called Project Relate (Relate), free to download from Google's Play Store. Relate enables a person with 'non-standard' speech to create a personalised speech recognition model that may more accurately recognise their own speech (Cattiau, 2021). To be clear, the Euphonia dataset is used to create a 'foundation' ASR model and Relate is an application that people can use to create a 'personalised' ASR model building on the Euphonia foundation model. The Relate personalised ASR model can be used on a person's own Android smartphone or tablet to caption their speech.

Used as AAC, Relate is argued to reduce the trade-off between AAC speed and accuracy for plwMND as their speech declines, frequently observed using other forms of AAC (Fried-Oken et al., 2023). Relate is argued to achieve this by enabling a plwMND to keep talking (maintaining speed of communication) and be better understood by others due to the accuracy of the personalised speech recognition model (maintaining accuracy of communication) (Cattiau, 2021).

To date (July 2024), Relate is available in English only to adults in Australia, Canada, Ghana, Kenya, India, New Zealand, Nigeria, Singapore, United Kingdom and United States.

In order to create a personalised ASR, Relate requires a person to record a minimum of 500 (and up to 5,500) set phrases on their device. Once successfully completed, a personalised ASR model is downloaded to the person's smartphone or tablet to be used on-device.

Relate has three main features: 'Listen', 'Repeat', and 'Assistant'.

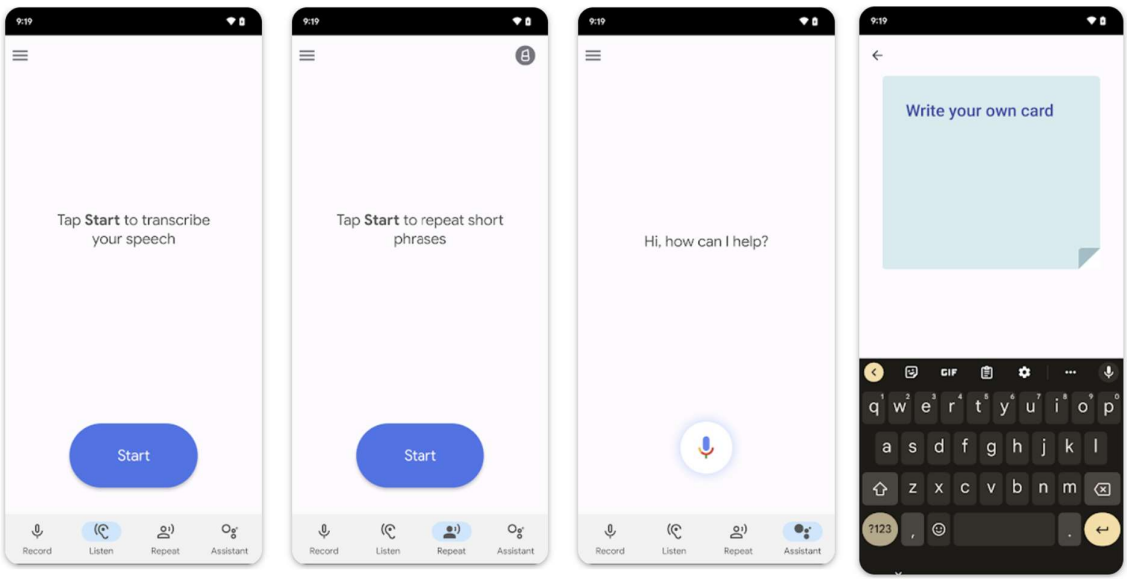


Figure 1: Relate features: ‘Listen’, ‘Repeat’, ‘Assistant’ and ‘Custom Cards’

Through the ‘Listen’ feature, Relate transcribes a person’s dysarthric speech to text in real time, displaying the text on the phone screen. This is to provide the listener with an additional modality to support understanding. The resulting transcription can also be replayed in a synthetic voice, or the text can be copied into other apps. However, this feature does not distinguish between speakers and will attempt to transcribe everything the microphone picks up. Thus, the words of all speakers in a conversation will be transcribed with no visible differentiation as to who has spoken.

‘Repeat’ restates a spoken short phrase of up to ten seconds duration using a synthesised voice. This speech-to-speech functionality is argued to support transactional speech-based conversation without the listener needing to reference text.

‘Assistant’ is argued to enable more successful interaction with Google Assistant using the transcribed speech from Relate.

To date, there has been no research on how Relate supports the communication of plwMND in everyday interaction, or how beneficial Relate may be to the listener.

Relate has the concept of ‘Custom Cards’, which is a way for a person to record a virtually unlimited number of proper nouns or personal phrases that they may say in daily life, such as names of family and friends, places they visit, names of objects, common phrases they say (Google, 2021). Custom cards are argued to support improved personalised ASR accuracy as many of these words may not be present in the general language model that Relate references (Google, 2021).

In the case of proper nouns, accurate ASR captioning with the support of custom cards may be particularly important for listener understanding as this word class is often unique and may not be predictable from the sequential context for people with dysarthric speech (Bloch & Saldert, 2020).

Custom card recording may be an important step as they are argued to positively affect ASR caption accuracy of 'personal vocabulary' and by implication Relate usability (Google, 2021).

The concept of custom cards is similar to other custom online supervised learning methods for specific subject domains. For example, ASR accuracy for captioning of live tennis television commentary was improved by creating a set of tennis words and terms from articles, interviews, video, players, and score information. Adding this domain specific vocabulary to the ASR model increased word recognition rate to nearly 70%, representing an increase of 5% in live TV captioning of tennis tournaments (Baughman et al., 2017).

However, as custom card recordings take a minimum of 24 hours to update the Relate ASR model (sometimes a few days), a user may need to plan ahead and record the words and phrases they think they may say in the future or add custom card recordings once an ASR caption error has been observed (Google, 2021).

1.8. Conversation Analysis (CA)

The research presented in this thesis uses Conversation Analysis (CA) to better understand how Relate may support social interaction between plwMND and their significant communication partners.

Conversation Analysis (CA) is a data-driven qualitative study of naturally occurring interaction. It attempts to describe conversation in terms of the management of interaction between two or more speakers. CA focuses on turns and sequences of turns in conversation as opposed to simply language as an abstract system of meanings: how they are designed, ordered, understood and achieve in action - such as storytelling or asking questions for example (Schegloff, 2007).

CA is has a number of assumptions, including: (1) everyday conversation in a naturally occurring encounter is how people normally conduct their everyday lives (Schegloff, 1992). (2) Conversational order is created and maintained by the participants themselves (Schegloff & Sacks, 1973), (3) Each turn is both context-shaped by its predecessor and context-renewing – forming a context for what is to be done at the next turn in any sequence. (4) Through each turn, a participant displays understanding or appreciation of the prior turn (Heritage, 1984), enabling another participant to assess how to proceed in their next turn at talk, especially if there has been a problem in understanding.

1.8.1. Trouble and Repair

Trouble sources describe what participants in conversation treat, or do not treat, as problems in talk. Anything within a turn can be treated as a trouble source even where there is no noticeable error or mistake (Schegloff, 1977). Additionally, delays between speaking turns may potentially be regarded as problematic or accountable (Clark, 1996; Sacks et al., 1974b).

Conversation is designed for understanding, therefore trouble is frequently repaired by the communicators themselves (Schegloff, 1992). The term 'repair' describes the process participants use to highlight and manage troubles within an interaction and is designed for efficiency and successful resolution of those troubles. Talk-in-interaction is prone to repair for a range of reasons, including issues related to the speaker, the recipient and their shared understanding of the talk (Schegloff, 1977).

The rules for turn taking (Sacks et al., 1974a) usually determine the first opportunity for repair goes to the speaker of the trouble source, and the next opportunity for repair is provided, in the next turn of talk, to other speakers (Schegloff, 1992).

Schegloff et al. (1977) report that repair revolves around two main practices: initiation (signalling something in preceding talk as a trouble) and outcome (the repair itself). Both initiation and outcome can be contributed by the speaker of the trouble source (self), or someone other than the trouble-source speaker (other), or both. Schegloff et al. (1977) identified a clear preference for the speaker of a trouble source to repair his or her own trouble (self-repair). The repair could occur within the same turn as a self-initiated self-repair (SISR) – perhaps restating an incorrectly produced word - or in another turn following identification of trouble by another person, called an other-initiated self-repair (OISR) - for example, when a conversation partner asks for repetition. OISR is regularly utilized when problematic talk in dysarthric conversations is encountered (Bloch & Wilkinson, 2009).

1.8.1. CA analysis of speakers with dysarthria

Most plwMND experience progressive dysarthria during the course of MND (Ball et al., 2004b). Conversation partners can play an increasingly active role to help clarify what a plwMND is attempting to convey and may methodically co-produce an utterance that would typically be produced by a single speaker (Bloch & Wilkinson, 2004). Resolution of trouble is usually resolved through the co-ordinated work of both participants, and once initiated, is often successful - even if it takes a longer time and a larger number of turns to resolve troubles when they arise. (Bloch & Wilkinson, 2011).

Repair may become the central feature of any interaction (Bloch, 2005). As MND progresses and speech becomes more affected by dysarthria, and reduced speech intelligibility is regularly located as a trouble source through other initiated repair followed by a self-repair attempt by the plwMND (OISR). Other-initiated repair can highlight which part of the plwMND's turn is creating difficulty, providing an opportunity for the plwMND to restate that part (Bloch, 2005; Bloch & Wilkinson, 2009).

1.8.1. AAC-in-interaction

The use of AAC does not preclude use of natural speech or gesture, indeed they are frequently all utilized to support each other in a highly organized manner (Bloch & Wilkinson, 2004), often to try and speed up communication (Clarke et al., 2023; Smith & Murray, 2016).

Timing troubles in turn production, turn transition, and repair may be problematic when AAC is used (Higginbotham & Wilkins, 1999), due to the relative slowness producing a message on

AAC resulting in increased difficulty understanding the sequential relationship between that utterance and what has preceded it (Bloch & Wilkinson, 2004).

Compared with speaking rates of between 125 and 185 words per minute (wpm) aided communication rates can fluctuate between 12–18 wpm (Higginbotham et al., 2012). The slower production of AAC-generated messages increases the potential for within-utterance permeability (G. Lerner & Kitzinger, 2012), in which conversation partners may insert further talk while an AAC-generated message is being constructed – increasing the risk of it being disrupted and/or not being completed (Wilkinson et al., 2011).

Speaking participants may experience problems understanding why the AAC user's turns were produced at particular points in the conversation (Bloch & Wilkinson, 2013a). This can be because of difficulties understanding of the sequential relationship between the AAC-generated message and the prior talk. Output may be mispronounced by the AAC or there could be a misspelling, or when replacing an earlier word with a subsequent one or restarting an utterance if the AAC user is attempting self-repair (Bloch & Wilkinson, 2013a). Even when turns produced using AAC are treated by recipients as intelligible (the words produced by the AAC device are identifiable) they may not be understandable in the context, why they are present at this point and what action they represent (Bloch & Wilkinson, 2004).

The relationship between an AAC-produced utterance and an earlier non-AAC-produced utterances may not be clear and the difficulty understanding an AAC-produced turn may lead to subsequent understanding problems by the recipient later in the conversation, resulting in prolonged repairs, and trouble understanding the AAC user's subsequent attempts at clarification (Bloch & Wilkinson, 2013a).

The understandability and intelligibility of AAC-produced speech may be affected by lack of stress, intonation, the pace and the accent of the synthesized voice (Wilkinson et al., 2011). Successful management of topic transition may not just rely on words, but also appropriate timing, prosody and non-verbal actions – and may have a significant impact on the relative success or otherwise of dysarthria and AAC-in-interaction (Bloch et al., 2015). A person using AAC may have difficulty indicating that they are initiating a new topic and may instead relate current talk to prior talk rather than starting something new (Bloch & Wilkinson, 2004).

1.8.1. Allocation of Responsibility for Trouble

Different relational stances can lead to different practices of repair initiation resulting in different behavioural consequences (Robinson, 2006). Stance refers to ways that people position themselves in conversation, often in terms of politeness, certainty, or affect/emotion (Kiesling, 2022).

As MND progresses, increased levels of dysarthria often becomes the location of a trouble source, frequently resolved through other initiated repair (OIR) (Bloch, 2005). OIR can make it difficult to avoid suggesting that someone else is responsible for the trouble. This happens because the structure of OIR is often context-independent, which can unintentionally imply that the issue lies with the other person (Sacks et al., 1974a). The risk is that the perceived meaning of an OIR implies disbelief in, disagreement with, challenge to, or rejection of the action in the trouble-source talk, and therefore may affect how the speaker of the trouble-source responds (Schegloff & Jefferson, 1977), and increase the risk of losing 'positive face' by communicating that they are responsible for repair-related trouble (Brown & Levinson, 1987).

Positive face is defined as the positive consistent self-image or personality claimed by interactants, including the desire that this self-image be appreciated and approved of (Brown & Levinson, 1987). In other research it was observed that AAC devices were sometimes treated as 'participants', partly by way of attributing blame on the voice output, as opposed to a conversation partner (Norén et al., 2022)

People constantly monitor acts, retrospectively and prospectively, for possible offensiveness, imagining the 'worst possible readings' (Goffman, 1971). OIR are 'practices for implementing action', including a practice for preserving the trouble-source speakers' positive face (Brown & Levinson, 1987).

1.8.1. Eye gaze

In addition to the importance of the eyes in displays of emotion and in the composition of facial expressions, gaze orientation during social interaction supports regulatory functions on turn - taking and the accomplishment of social action (Rossano, 2012). When people with dysarthria use AAC, they typically gaze where the graphic symbols are being used and not at the recipient, the recipient also regularly looks at the AAC device while the message is being constructed or produced (Wilkinson et al., 2011). The return of gaze to the recipient may mark the possible end of the utterance at that point, in the absence of other cues available when speech is used, such as prosody (Wilkinson et al., 2011). In this way, eye gaze is patterned quite differently compared to found in mundane face-to-face conversation (Antaki & Wilkinson, 2012). This research tracked eye gaze patterns across all the sessions to better understand how similar or

different patterns of eye gaze were when using Relate, compared to other AAC and face-to face conversation..

1.8.1. CA analysis of personalised ASR in-interaction

Personalised ASR technology, such as that used in Relate, is claimed to help people with dysarthria, or 'non-standard' speech be better understood in everyday conversation (Google, 2021) . The ASR technology displays a transcription of speech with dysarthria almost as it is spoken, on a screen such as a smartphone or tablet (Cattiau, 2021).

This appears quite different to what is seen in the slower and frequently sequential nature of AAC-based interaction. Could Relate help make a plwMND turn more understandable by reducing timing delays and being available in the intended moment of interaction? Or will there be fewer examples of conversation partners anticipating an AAC utterance completion (Bloch, 2011) or inserting further talk, such as a question (Bloch & Wilkinson, 2004)? There is no identified published research exploring the way participants may manage and make use of these different but co-occurring resources. CA analysis could be a useful approach to investigate these gaps in the knowledge base

1.9. Introduction - Summary

Many plwMND consider the change and loss of speech as one of the worst aspects of MND (Hecht et al., 2002), and when speech does decline it frequently has a negative impact on their sense of individual and social identity, increasing the risk of social withdrawal (Yamagishi et al., 2012). When speech becomes hard to understand, the AAC devices provided to plwMND are very slow for communication in comparison to natural speech (Kristensson et al., 2020), and the limitations of AAC devices can affect identity maintenance and relationship management as well as MND (Kane et al., 2017). As a result, some plwMND prefer to keep using speech even when they are difficult to understand (Smith & Connolly, 2008).

Project Relate, an AI-driven tool, is argued to help plwMND with dysarthric speech to be understood for longer when talking - by generating automated real-time captions of their words, supporting others to understand them (Cattiau, 2021). In doing so, Relate may be fulfilling a promise of AI for AAC by enabling speed of communication close to that of natural speech whilst retaining authenticity of the message (Koch Fager et al., 2019; Light et al., 2019). However, as the scoping review in the next chapter shows, there has been no research identified that examines how this technology works in practical use - 'in the wild'.

Therefore, the focus of this research is to assess how Relate supports plwMND to be understood by others in everyday conversation, and how it may change as speech declines during the course of the disease.

This goes beyond assessing Relate caption accuracy measured by WER and Conversation Analysis (CA) will be used to examine how the captions may help plwMND to be understood in

the wider context of interaction. This research will include the views of both plwMND and their communication partners on how useful they believe Relate captions are for supporting understanding. Their perceptions of conversational success are an important part of recognising how useful Relate may be in practice.

Additionally, the research will review the forecast prediction of understanding that Relate itself provides to users, based on estimated WER thresholds. The thresholds will be compared to the actual WER of the conversations recorded in this research. In this way the expectations Relate provides to users around being understood can be assessed - based on the measures Relate itself uses.

1.10. Research Questions

In summary, the following research questions describe the objectives of this research:

RQ1: How does the Relate app enable plwMND to be understood by others in everyday conversation?

RQ1a: To what extend does this change as speech declines during the course of the disease?

RQ2: How do the captions produced by Relate help plwMND to be understood in the wider context of interaction?

RQ3: How accurate are the predictions and measures Relate uses in its software?

The findings will provide insight into what degree Relate actually supports plwMND to be understood in everyday conversation with communication partners over time.

2. Previous Research

2.1. Introduction to the chapter

In this chapter, the results of two scoping reviews are discussed. Both examined how ASR technology is being used by plwMND and identified gaps in the evidence base. The original published scoping review covered three decades from 1990-2020 and the second updated the results to May 2024.

The literature eligibility criteria, list of information sources and the detailed results by paper are in the appendix. The discussion and conclusion are in this chapter.

In total 44 papers were identified across both reviews, 33 of which were published within the most recent three and a half years (2021-2024). The significant growth of research may be due to increased availability of machine learning technology, increased availability of recordings of plwMND with dysarthric speech, and growing interest in automated grading of dysarthria severity or identification of MND.

The majority of the studies used recordings, and measured the success of the technology by the number of captioning errors the ASR made. Two papers assessed how ASR captioning accuracy changes as dysarthria progresses by using recordings of short phrases across time.

However, no study examined how ASR performance assessed in this way related to the usability of ASR in functional use by plwMND, or sought feedback from plwMND (or who they may communicate with) on what acceptable ASR performance might be for them.

The results of both scoping reviews indicated many potential avenues for research, much of which is planned for future studies. For this research, the scope was limited to better

understand how plwMND and communication partners use ASR technology for live captioning in typical conversational settings talking about topics of their choice, and how the use of ASR may change over time as dysarthria progresses. This is because so little is known about how this technology is used in practice. The outcomes will provide recommendations for best practice and feedback to the technology providers based on user experience.

2.2. Background

Part of this chapter (2.4 and 2.5) and the Appendix (section A5) comprises content from the published scoping review: "The use of speech recognition technology by people living with amyotrophic lateral sclerosis: a scoping review", Disability and Rehabilitation: Assistive Technology, DOI: 10.1080/17483107.2021.1974961 (Cave & Bloch, 2021a). The scoping review covered the period 1990 (Jan 2) – 2020 (Nov 15).

For details on the information sources and search strategies please see the Appendix. Also in in the appendix is Table 1, containing the data extraction proforma for the identified studies. Tables 2-4 provide results of the data extracted from each study, and can be viewed online here, and are also in the Appendix.

An additional scoping review was completed in June 2024 to update the results of the first, covering the period 2020 (Nov 16) – 2024 (May 31) and is discussed in section 2.5 and 2.6. The detailed results are in the Appendix (section A6). Table 5 contains the data extraction proforma for the identified studies. Tables 6-8 provide the results of the data extracted from each study and can be viewed in the Appendix and online here.

2.3. Objectives

The field of ASR is rapidly developing in terms of technology and applications and these scoping reviews provided a state-of-the-art overview to highlight significant gaps in the evidence (Arksey & O'Malley, 2005) and point to future research in the field (Armstrong et al., 2011).

The objectives of both scoping reviews were to understand (1) what is known, from the existing literature about the use of ASR technology by plwMND, and (2) what gaps there are in the existing ASR-MND evidence base. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines were utilized for this study (Tricco et al., 2018).

2.4. Discussion

This section discusses the findings of both scoping reviews together. In the first scoping review eleven articles met the eligibility criteria from a total of 177 identified from electronic databases and grey literature searches across three decades. The second scoping review identified 33 articles from a total of 909 across three and a half years, representing significant growth in research in this area.

2.4.1. Increased use of machine learning technology in research

The increased availability and use of machine learning technology may be a factor driving the increase in research. 31 of the 33 papers referred to using machine learning techniques, compared to 6 of 11 in the first scoping review.

2.4.2. What do plwMND use ASR technology for?

Across both scoping reviews only seven studies investigated how plwMND may use existing ASR technology. Pradhan et al. investigated how useful the Alexa Echo may be to people with disability by assessing the contents of Amazon reviews (Pradhan et al., 2018), however it is difficult to draw any conclusions for plwMND as the authors did not indicate which of the participants had MND beyond citing two quotes.

Wallock & Cerny, (2021) focused on how plwMND use smart home technology as assistive technology. Smart home devices were reported to provide greater control over electronics, increasing functional independence and regaining the ability to complete some tasks they were unable to do. However, the study only had a small sample size and relied on self-reported data.

Ballati et al. (2018a) presented participant plwMND with a list of frequently used commands for Google and Alexa and asked if they would be useful in everyday life. The plwMND indicated that they generally would. However, this is not the same as understanding the ASR commands that plwMND actually use. Kim et al. recorded samples of everyday conversation, phrases that may not bear a relation to how plwMND actually use ASR devices (Kim et al., 2018).

Two of the studies utilized recordings from the TORGO database in their research (Ballati et al., 2018b; De Russis & Corno, 2019). However, as TORGO has no recordings of people interacting with ASR devices, samples were edited. Caves et al. used recordings of zip codes and number repetition, and it is difficult to interpret potential functional use of ASR from this dataset (Caves, 2007).

Without a clear understanding of what plwMND use ASR technology for, it is difficult to measure how successful ASR could be at meeting their needs. There remains a significant gap in understanding how effective ASR technology is for plwMND in a functional setting.

2.4.3. How does ASR use and usefulness change as MND progresses?

Only two studies across both scoping reviews assessed how ASR recognition error rates may change as MND progresses (Kim et al., 2018; Tomanek et al., 2023). Both papers used recordings of short phrases repeated over time. Although using recordings of short phrases provides a consistent basis for comparison across ASR devices, this approach is unlikely to reflect how plwMND may use ASR in conversation. Additionally, neither paper explored the change in usefulness or perceived benefit to plwMND across time. Research is needed to understand the changes in use and benefit of ASR for plwMND as their condition progresses.

2.4.4. WER remains the main measure for measuring ASR performance.

In the first scoping review 9 of the 11 studies assessed ASR performance using WER (word error rate), WRR (word recognition rate - which is the inverse of WER) or PER (phoneme error rate), and 12 of the 16 studies assessing ASR also used these measure in the second review. However no study examined how WER/WRR/PER relates to the usability of ASR devices for plwMND or any assessment of what level of WER/WRR/PER is acceptable to plwMND using ASR.

Tobin and Tomanek argued that 15% WER for commands to digital assistants and 20% WER in conversational speech would be of functional beneficial for people with dysarthria using an ASR (Tobin & Tomanek, 2022). However this was not validated with plwMND.

As an alternative to WER/WRR, one study used BERT (Bidirectional Encoder Representations from Transformers) as a measure of success (Tobin et al., 2022). The study found BERT was

more useful in measuring the semantic similarity between the ASR model's output and the intended meaning, especially where WER is misleading due to errors not altering meaning. This study assessed set phrases and not spontaneous conversational data the accuracy rates may be impacted in everyday environments with background noise. However, BERT can experience issues recognising semantic meaning with multilingual data and when words that are not in the LLM but nevertheless valid (for example some proper nouns).

Tomanek et al. (2024) argued that LATTEScore – a measure that uses tuned-LLM assessment of meaning preservation - may be used as a proxy for human evaluation when ASR captioning of speech with dysarthria. Their research proposed LATTEScore as a viable alternative to the slower and labour intensive method of manual assessment by SLTs. However, no comment was made about the content of the phrases used, whether conversational or transactional, short or long in length, and particularly how proper nouns that may not be present in the LLM are assessed. Additionally, LATTEScore evaluates meaning preservation only within individual sentences. Assessing preservation of meaning across multiple sentences or entire conversations better reflects the reality of conversational interaction.

No study assessed if a change in the measure used was significant enough for plwMND to use ASR devices any more or less.

2.4.5. Recordings of plwMND remain in short supply or may be selectively used

Across both scoping reviews nine studies utilized recordings from the TORGO database in their research. TORGO has recordings from eight participants, including only one plwMND - the remainder were diagnosed with cerebral palsy. In TORGO no details are provided about which

of participants has MND and which recordings are attributable to them. Therefore, it remains unclear how to interpret the results specifically for MND.

Ten studies utilized recordings from Euphonia, which, in contrast to TORGO, has more than 1300 hours of recorded dysarthric speech, including more than 170 plwMND who have recorded 300 phrases or more. Euphonia is claimed to contain the largest dataset of recordings of non-standard speech globally (MacDonald et al., 2021) . One problem with Euphonia is its availability - only research that is based at Google can access Euphonia (MacDonald et al., 2021). Another concern is that the Euphonia team prioritized high transcript conformity and reduced transcription costs over collecting recordings of the words people may actually say in conversational speech. As a result most of the recordings are repetitions of set words and short phrases (MacDonald et al., 2021).

Even with access to so many recordings, Euphonia-based research appears to have mainly focused on assessment of ASR recognition of single words or short phrases, as opposed to longer phrases, and conversational speech. Only one study specifically detailed inclusion of recordings of spontaneous conversational speech (Tobin & Tomanek, 2022) and the remaining nine used recordings of single words and short phrases.

However in the Tobin and Tomanek 2022 paper, some analysis is made of the 'organic test set', which contains limited recordings of plwMND in spontaneous conversation on a topic of their choosing. Only a chart of the results is shown in the paper but it indicates approximately 60% WER (at best) for two of the participants, and the best performing ASR model for any plwMND was around 20% WER.

It is difficult to be sure why the Euphonia-based published research has focused almost exclusively on single words and short phrases, especially considering that the main feature of Project Relate (built by the same team) is to caption conversational speech using a personalised ASR. In one paper Tomanek et al. refer to the 'confounding factors' associated with ASR for longer sentences and conversational speech (Tomanek et al., 2023). One factor may be the worse reported outcomes for conversational speech in terms of WER when compared to short phrases and single words. Another factor may be that 'successful' studies could be perceived as more valuable for publishers, editors and readers (Mlinarić et al., 2017) – and therefore more likely to be published. Research with positive results are reported to be more favourably cited (Duyx et al., 2017).

The relatively low availability of data outside Euphonia may have been a factor for some researchers to ask healthy participants to emulate how a plwMND with dysarthria may sound (Naeini et al., 2024). It is difficult to assess research outcomes for plwMND when the recordings do not replicate the full complexity and variation of dysarthric speech. Additionally, the same researcher created synthesised dysarthric speech, with 55 different accents without discussion on the similarities/differences of accent and dysarthria or the appropriateness of utilising accent to measure ASR effectiveness for dysarthria (Naeini et al., 2024).

2.4.6. Could ASR technology support diagnosis, therapeutic intervention?

Eighteen papers across both scoping reviews researched automated approaches for early detection of MND or measurement of dysarthria severity. Eleven papers explored how machine learning techniques could provide objective and reliable measurement of dysarthria that potentially could be used as a tool for intervention planning. One paper examined developing

a method for diagnosing bulbar involvement in plwMND using speech recordings (Tena et al., 2021), and Connaghan et al. used a smartphone app to remotely identify and track speech decline in MND (Connaghan et al., 2019). Stegmann reported detection of early speech changes and track speech progression in MND via automated algorithmic assessment of speech collected digitally (Stegmann et al., 2020). Research into automatic detection of MND from speech samples and the feasibility of detection even in pre-symptomatic stages is ongoing (An et al., 2018; Vieira et al., 2020; Wang et al., 2016), and one paper researched reliable identification of MND from recordings of plwMND compared to those with Parkinson's Disease and healthy participants (Bhattacharjee et al., 2021).

However, Wiepart et al. highlighted errors that an ASR model can itself generate when assessing dysarthria. For example whether an ASR model allows for non-words, as well as taking into account the general objectives of the ASR model (e.g., captioning intended meaning vs. sound-for-sound transcription) (Wiepert et al., 2023). Gutz et al., argued ASR systems should not be used for scoring of speech intelligibility or replacement of clinician dysarthria severity rating. This is because, for the ASR systems they assessed, there were many transcriptions errors, particularly with longer phrases, and for words that were not recognized but nevertheless valid.

Only two studies used recordings of conversational speech to assess dysarthria (Bhattacharjee et al., 2021, 2023). The remainder used recordings of vowel sounds, single words or short phrases, which may not be representative of dysarthria severity. This appears to be an area with much research interest, particularly to support or replace assessments of dysarthria by SLTs, potentially enabling earlier and more accurate assessment and intervention planning.

However, none of the researchers appeared to trial their studies with SLTs or in the environment where these assessments may take place.

2.4.7. What communication, physical and cognitive needs do participants have?

In the first scoping review only four studies described the type of dysarthria, and only four detailed the assessments used. In the second only eleven studies identified in detail the participant profile. For many studies, it was difficult to ascertain which of the outcomes related to plwMND. Without an understanding of the communication, physical and cognitive needs of the participant plwMND it is difficult to interpret research outcomes.

2.4.8. Could changing how plwMND speak support functional use of ASR technology?

Hird and Hennessey identified small gains in dictation speed using voice recognition technology as a result of therapy aimed at better regulation of flow of air and volume loss during speech, and practice at increasing loudness of speech to mitigate growing environmental noise (Hird & Hennessey, 2007). However, it is difficult to assess how this overall result applied to the plwMND in the study as individual results were not published. The paper did not explore the participant's own views on the relative functional benefits of the increase in speed or assess how speech recognition may change as the disease progresses. Some researchers argue that there are functional benefits by training a speaker to alter speech production to improve speech recognition of dysarthric speech (Hawley et al., 2013; Kotler & Thomas-Stonell, 1997). No identified research was found to measure the usefulness of this approach using more recent and popular ASR devices.

This prompts the wider question of how far a person should be obliged to change their own behaviour to mitigate poor design that lacks inclusiveness (Norman, 2013). More research is indicated to understand how acceptable it is to plwMND to need to adjust their own speech to bring about improved ASR performance.

2.4.9. ASR technology for non-speech or limited speech recognition

Cai et al.,(2021) demonstrated that ASR can be utilised for recognizing non-speech vocalisations. The researchers trained a switch to detect the vowel sound /a/ and link the sound to an action. The results showed that using ASR in this way could be useful for plwMND with severe dysarthria for environmental control. Along the same lines, Mulfari et al. argued that accurate single word ASR recognition can help interaction with virtual home assistants (VHAs), such as Amazon Alexa, supporting environmental control using voice, for example a smart home. Additionally, mapping single word speech to a customized speech message spoken aloud via a computer generated voice, could help with transactional communication to people as well as more complex instructions to VHAs (Mulfari et al., 2023). However both studies were very limited in the vowel sound or individual words that the ASR could accept, and therefore would benefit from analysis on how functionally beneficial to plwMND this research could be.

2.4.10. BCI technology with an approximation of natural speech

Angrick et al. (2024) used brain activity recorded with an implanted brain-computer interface (BCI) to synthesize speech. Recurrent neural networks (RNNs) were used to identify, decode

and synthesize speech from electrocorticographic (ECoG) signals from the motor, premotor and somatosensory cortices.

The synthesized speech closely resembled characteristics of the plwMND's voice profile for the six keywords it was trained on, and more than 80% of the synthesized words generated from neurological signal mappings were correctly recognized by human listeners.

The approach seems to be a promising way of adding personalization to communication using BCI technology, where the equivalent of the ASR is driven by neural signals rather than directly from speech. However, the paper did not detail feedback from the plwMND on their views of the system or what they thought of the synthesized voice, which is unfortunate.

2.5. Conclusions

The objective of this scoping review was to update the result of the earlier review to include research from November 2020 to May 2024. 33 papers were identified, in comparison to the eleven papers identified in the first scoping review that covered two decades.

There may be multiple factors driving the significant growth in research: the increased availability and use of machine learning technology, increased availability of recordings of dysarthric speech (for some researchers), and growing interest in automated grading of dysarthria severity, or identification of MND.

However, some trends remain very similar in both scoping reviews: for the 16 identified studies assessing ASR performance, 12 used WER or WRR as a measure of success of ASR, compared to 7 of the 11 studies in the earlier review. Similar to the first review, no study examined how

these measures relate to the usability of ASR devices for plwMND. Some researchers made assumptions on what level of WER may constitute acceptable performance, but no researcher appeared to involve plwMND to understand what acceptable WER/WRR might be to them, similar to the previous scoping review.

BERTScore and LATTEScore are argued to be more useful than WER in measuring meaning preservation in ASR captioning, however these measures are likely to be affected by the words it can recognise and are unable to assess meaning preservation at conversational level.

The availability of recordings of plwMND with dysarthria remains sparse for most researchers. Nine studies utilized recordings from the TORGO which has recordings of only one plwMND. The Euphonia dataset has more than 1300 hours of recorded dysarthric speech but only Google researchers can use it. However even then, only one Google study included recordings of spontaneous conversational speech.

Eighteen papers explored how machine learning techniques might support objective and reliable measurement of dysarthria or earlier detection of MND and argued how they may be used as a tool for earlier and more focused intervention planning for SLTs. However no researcher appeared to trial their studies with SLTs. Additionally, two papers raised significant issues about the accuracy and validity of using ASRs for objective measurement of dysarthria.

Two researchers assessed how ASR performance changes as dysarthria progresses due to MND however no assessment was made of the change in usefulness or perceived benefit as judged by plwMND themselves. Further research is needed to understand the changes in use and benefit of ASR for plwMND as their condition progresses, with or without additional recordings.

There is a paucity of research to understand how effective ASR technology is for plwMND in a functional setting, and what measures are most appropriate to measure benefits. Many studies used WER, WRR, BERT or LATTEScore as a measure of success of ASR but no study examined how these measures relate to the usability of ASR for plwMND.

All but two of the studies focused on a single snapshot in time, and research is needed to understand how the functional use and usefulness of ASRs changes as MND progresses and the severity of dysarthria changes – from the perspective of plwMND themselves.

Only two small studies examined how plwMND use ASR devices in daily life, how effectively their needs are met, and when they choose not to use it. It is important to understand the expectation of plwMND from ASR technology. Further work is needed exploring how effective the technology is in aiding communication in a functional setting.

In summary, despite the very significant advances in machine learning technology and the sizeable increase in research in recent years, there remain gaps in the research evidence base.

2.6. Research Direction

There are a number of directions further research could have gone as a result of these scoping reviews. For this study the focus was limited to better understanding how plwMND use ASR technology for live captioning in typical conversational settings, talking about topics of their choice, and how their use of ASR technology may change over time as dysarthria progresses. This research area was chosen because so little is known about how ASR technology is used in 'in the wild' by plwMND, what their lived experience is when talking about everyday things in everyday settings with ASR present, and how it may change over time. The outcomes will

provide recommendations for best practice and feedback to the technology providers based on user experience.

3. Objectives

To date, there is no published research on how plwMND use personalised ASR to support everyday spoken conversation with significant communication partners over time.

The objectives of this research are:

- To examine how Relate supports plwMND to be understood by others in everyday conversation. This goes beyond assessing Relate caption accuracy measured by WER and seeks to analyse how the captions may help plwMND to be understood in the wider context of each interaction, using the perspective of Conversation Analysis (CA),.
- How Relate supports plwMND to be understood as dysarthria progresses during the course of the disease over a period of up to 12 months. A significant decline in speech intelligibility can occur relatively quickly, even prior to confirmation of diagnosis of MND (Makkonen et al., 2018).
- Using Thematic Analysis (TA), this research will include the views of both plwMND and their communication partners on their experiences of using Relate. Their perceptions of conversational success are an important part of recognising how useful Relate may be in practice.
- The research will also review the expectations Relate itself provides to users for understanding, which is based on forecast WER thresholds. The thresholds will be compared to the actual WER of the conversations recorded in this research. In this way, the research can assess how realistic these expectations of understandability are based on the measures it uses. Additionally, as many researchers are examining retention of

meaning as an alternative to WER for measuring ASR caption success, this study will also assess how well Relate maintains meaning in practical use.

4. Method

This study was approved by the UCL Research Ethics Committee (ID 12607/003) on 20th June 2021.

4.1. Participants

Three plwMND participants were recruited via an open advert distributed at MND Association (MNDA) branches in England, Wales and Northern Ireland, also distributed via MNDA staff. Each plwMND was asked to nominate a communication partner with whom they communicate on a regular basis, knows them well, and would be willing to participate in the study.

Each plwMND was aware of Relate and was in the process of recording phrases to create a personalised ASR model or using it already.

All that had registered an interest in the research were emailed a link to view a five-minute online video providing a briefing of the study. All potential participants and their nominated significant communication partners were met by the researcher prior to the formal consent process in person, or virtually using UCL Microsoft Teams. This enabled potential participants to be provided with information in a clear and understandable form before they decide to progress.

Participants were included who:

- had a formal diagnosis of MND or were a communication partner of a person with MND
- The plwMND used speech predominantly and already uses Relate to some degree
- The communication partner reports difficulty understanding the natural speech of plwMND without AAC support. The researcher will have ascertained the level of

conversational breakdown and repair between plwMND and conversation partner at the interview.

- were over 18 years of age
- were resident in England, Wales or Northern Ireland
- Participants with English as an additional language (EAL) were accepted if they were able to understand English to a sufficient level as part of the interview process, as judged by the researcher at the information briefing prior to consent.
- There were no restrictions in terms of physical disability, sexual orientation, religion or ethnicity.

Up to 15% of people diagnosed with MND develop frontotemporal dementia (FTD), either at the same time or after diagnosis of MND, and up to 15% of people diagnosed with FTD may go on to develop MND (Bäumer et al., 2014) . People with MND and FTD may lack mental capacity to consent to the research or participate throughout the research. Mental capacity is assumed unless there is evidence to think otherwise, in compliance with the Mental Capacity Act 2014 (Office of the Public Guardian & Ministry of Justice, 2014). The researcher is a highly speech and language therapist experienced in completing mental capacity assessments and in working with people with complex communication disabilities. Where there is a concern about capacity to consent, or a concern about a participant lacking capacity during the course of the research (who initially had capacity), the participant will be excluded from the research in a sensitive manner.

No person was excluded from this study, and there was no participant where there were concerns about capacity to consent. Despite extended advertising from the MND Association

only three dyads were recruited (the goal was for five). This is likely because Relate is such a new technology and few plwMND may use it, or be aware of it.

After consent was obtained, data were collected using video recordings of each dyad in conversation at the home of the plwMND or their preferred location.

For this study between six and ten conversations between each plwMND and their conversation partner were video recorded for a period of up to a year. A year was chosen because one goal of the research is to observe how Relate supports plwMND to be understood as dysarthria progresses during the course of the disease. Many plwMND will lose functional speech for communication within eighteen months from first symptoms, which may develop before an MND diagnosis is confirmed (Makkonen, Ruottinen et al., 2018).

The goal was to record the conversations approximately monthly, though this was dependent on participant availability and was frequently closer to two monthly.

4.1. Recruitment

Opportunistic sampling was employed in order to recruit sufficient numbers of participants. Participants were recruited through either an advertisement (see appendix 13.1) placed in MND Association branches or through MNDA Area Support Co-ordinator (ASCs) in England, Wales and Northern Ireland. The ASCs were contacted by phone and email, provided with inclusion criteria by the researcher and requested verbal consent for the names and contact details of the plwMND and significant communication partner to be provided to the researcher.

4.1.1. Informed consent and confidentiality

A participant information sheet and consent form (see appendix 13.2 and 13.3) were provided to all participants identified by the ASCs or had responded to the advertisement. The researcher followed up to request verbal consent and obtained formal consent at interview. If potential participants were not able to sign a consent form because of the nature of their disability, they were asked to indicate consent on an audio recording.

4.2. Participants

Three dyads were recruited for this research and their details are below in table 7.

4.2.1. Pam and Carol

Pam had a diagnosis of MND for about 18 months prior to the beginning of this study. At the time of data collection, Pam was physically well and leading an active lifestyle – including going on long walks with friends and playing badminton at her local club. Her cognitive and language abilities remained intact.

Carol, Pam's communication partner for the study, has known Pam for more than 40 years and has no speech, language or cognitive difficulties. Pam and Carol usually meet up at least once a week for a coffee or a walk together and have done so for many years. Carol reported that she finds Pam's speech more difficult to understand in recent months and increasingly relies on Pam to write down key words to follow the conversation. During all of the conversations, natural speech was the primary mode of communication. Relate, and occasionally pen and paper was used as AAC. Pam chose for the video recordings to be at her home.

4.2.1. Bob and Adrian

Bob had a diagnosis of MND for about 15 months prior to commencement of this study. Throughout the period of research, he was working as an IT team leader, was physically well and his cognitive and language abilities intact.

Adrian is Bob's friend and work colleague and has known Bob for more than five years. Adrian reported that he finds Bob harder to understand over the past year, although they normally still rely on speech to communicate. Bob occasionally texts key words on his phone if helpful to the conversation. Bob chose for the video recordings to be at his place of work.

4.2.1. Elsbeth and Tony

Elsbeth is a PhD student and was diagnosed with MND for eight months from the start of this research. Her mobility is increasingly impaired and is becoming more reliant on others for activities of daily living. Her cognitive and language abilities are intact. Tony is her partner of more than 30 years. He reports that sometimes he cannot understand Elsbeth when she initiates a new topic without context, and it often helps when she writes key words. Elsbeth chose for the video recordings to be at her home.

Name	Relation	Months since diagnosis	Gender	Role	Primary communication method	AAC	Age
Pam	PlwMND	18	F	Retired	Speech	Pen and Paper	60-70
Carol	friend		F	Retired	Speech	n/a	60-70
Bob	PlwMND	15	M	Working	Speech	Text app on phone	40-50
Adrian	friend		M	Working	Speech	n/a	30-40
Elspeth	PlwMND	8	F	Student	Speech	Pan and Paper	50-60
Tony	partner		M	Retired	Speech	n/a	60-70

Table 11: Details of dyads participating in the research

4.3. Equipment

All recordings were in a quiet room, either at the person's home or place of work. The recordings were between 15-30 minutes in length, and the participants were free to stop at any time.

Video recording equipment was set up by the researcher in the room beforehand. The equipment to video was a Samsung 21 Ultra smartphone, mounted on a stand, and standard Android operating system screen recording software on the smartphone each dyad used to generate the Relate transcriptions from.

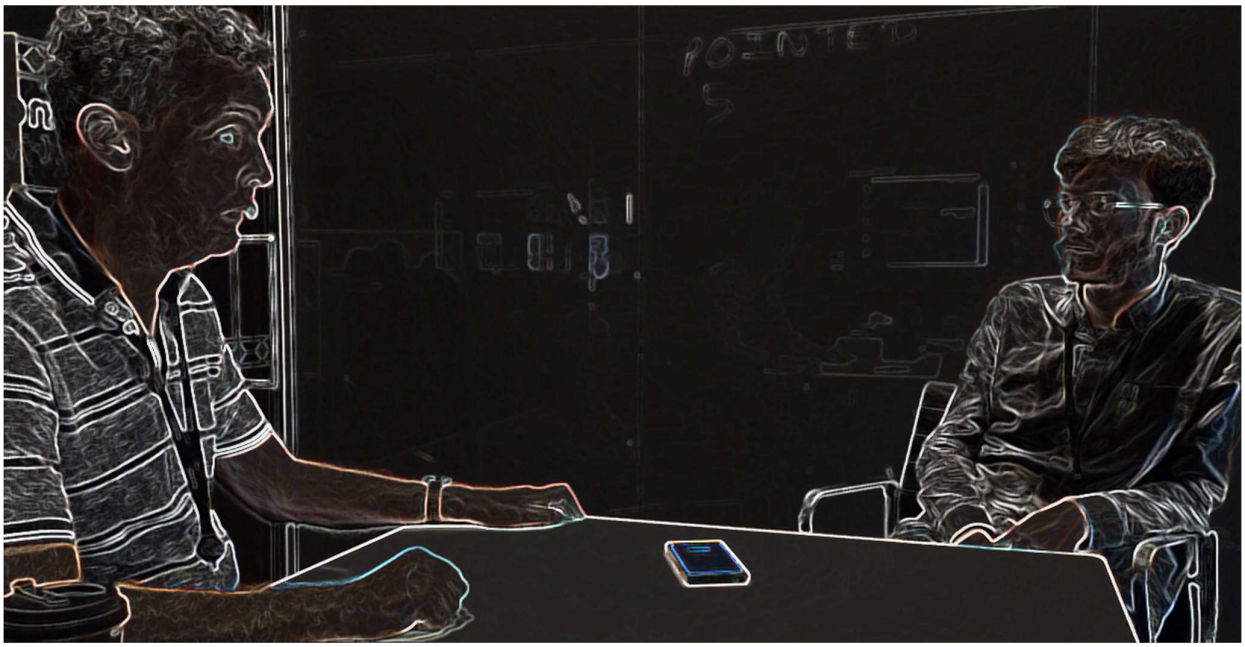
After each session, the two recordings (video and screen recorder) were merged using Microsoft ClipChamp software, to enable the participants and the Relate captions to be clearly observed on one screen.

The participants were reminded that video recording would take place at the beginning of the session. The researcher was not in the room to reduce to a minimum any researcher influence on the session and make the recording activity as unobtrusive as possible.

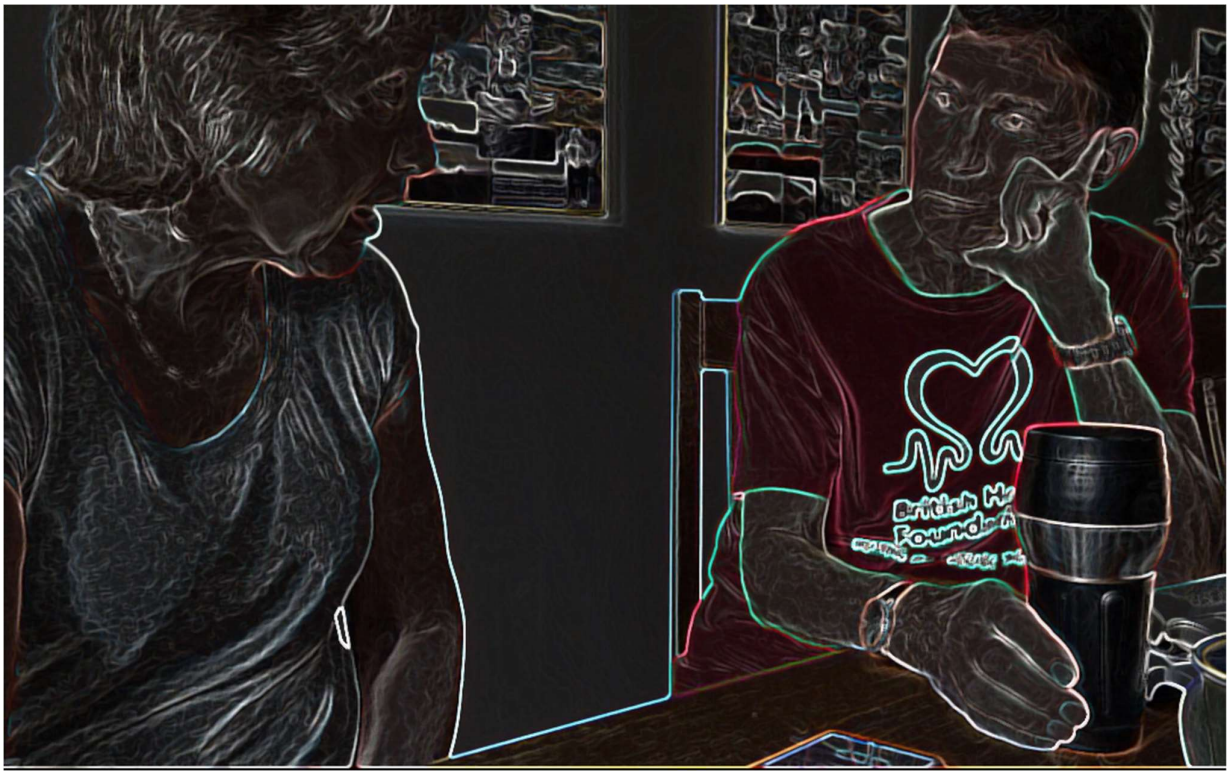
4.3.1. Relate set-up

Pam used Relate on a Google Pixel 6 smartphone. She positioned the phone screen to be viewed by Carol as she almost never looked at the captions herself when in conversation. The Pixel 6 was angled towards Carol with the top of the phone leaning on a glasses case and the bottom resting on the table between them.

Bob used a Pixel 6a smartphone and positioned the phone screen facing Adrian on a table they were sitting at. Bob could see the captions and quite frequently looked at them too.



Elsbeth used a Samsung A52 smartphone, positioning it on the kitchen table facing Tony. She looked at the captions occasionally.



For all three plwMND, Relate was switched on at the beginning of the session and left on throughout in 'Listen' mode.

4.4. Assessment of dysarthria

In addition to the online video data collection, the Frenchay Dysarthria Assessment - Second edition (FDA-2) (Enderby & Palmer, 2008) was administered to the participants with dysarthria as soon as feasible after recording at sessions one, three, six, nine as applicable to the number of sessions they actually had. These assessments were used to explore changes in speech and change in use of Relate over time. The FDA-2 assessments were administered and scored by the author. The intelligibility score for sentences is reported here as it is most relevant to the focus of the research.

All three plwMND used 'Listen' in conversation with their communication partners. To better understand how Relate was used by plwMND in their lived experience, there was no training administered between sessions unless help was specifically asked for by the plwMND.

4.5. Conversation analysis

Conversation Analysis (CA) was used in this research to better understand how Relate may support social interaction between plwMND and their significant communication partners.

Every session was videoed and sections were transcribed using Jeffersonian conventions (Jefferson, 2004), the standard for CA transcription. Distributed Open Transcription Environment (DOTE) transcriptions software was utilised (McIlvenny & Davidsen, 2022). Pseudonyms were used for all participants. All information that might lead to identification of participants was anonymised in transcripts. During analysis, visual details of participants were not masked.

4.6. ASR Accuracy - WER

When a personalised ASR model is created or updated, Relate provides the user with an expectation of 'High', 'Medium' or 'Low' ASR accuracy. To do this, Relate calculates a guideline WER score based on a sample of generic phrases tested against the updated personalised ASR model. A 'High' rating implies the guideline WER is under 15% (that is, 15 caption inaccuracies per 100 words spoken), 'Medium' between 15% and 35% and 'Low' is anything over 35%. However, the generic phrases used are unlikely to be reflective of what a person actually says while using Relate.

The actual WER for Relate was calculated for each recording session. A Google Sheets 'apps script' was used to compare the ground truth transcript with equivalent Relate caption and calculate WER.

Relate will attempt to transcribe every speaker it hears, plwMND and any other, therefore WER was only calculated for the speech of the plwMND in the recording. Phrases where speakers overlapped were also disregarded as Relate could not be expected to reliably transcribe only the speech of the plwMND.

Phrases were also disregarded where the researcher could not accurately transcribe the words spoken by the plwMND after extended analysis.

4.7. ASR Accuracy - Meaning Retention

WER is the most commonly used metric for evaluating speech recognition performance (Mishra et al., 2011). However, when calculating WER, all words are considered equally important, and

all errors (substitutions, deletions, insertions) have equal weight. WER does not consider whether some words may be more important to the meaning of the message and the impact of word errors may be also dependent upon the specific application in which ASR is being used (Favre et al., 2013).

A measure for ASR should have ecological validity that realistically simulates how ASR output would actually be used and how useful that output would be (Favre et al., 2013). People that use ASR for transcription reported less concern about measuring word-for-word accuracy between the spoken message and its transcript but whether the transcript produced by ASR captured the meaning of the spoken message (Mishra et al., 2011).

It is for these reasons the researcher additionally allocated a rating for accuracy of the Relate transcription to the meaning of what the plwMND actually said, similar to the approach used by Tomanek et al. (2023). Table 9 shows the rating criteria, and table 10 shows example phrases that were scored.

Score	Description
0	word for word transcription accuracy
1	some transcription inaccuracies. Still conveyed enough meaning to reflect what was said
2	serious transcription inaccuracies. Does not convey the meaning of what was said

Table 9: Score for accuracy of transcription

Where a score of 1 or 2 was given, and an indication of the main cause of the inaccuracy was detailed (there are often multiple causes in a single phrase, therefore a subjective assessment of the most significant and/or frequently occurring was taken as the main) : incorrect or misspelled proper noun, deletion, insertion, contraction, homophone, repetition, or word error

(generally incorrect transcription). Examples for each score are below, taken from Pam and Carol's first recording:

Score	plwMND talk	Relate Transcription	Main Error Type
0	It was all good	It was all good	none
1	I said you can talk it's allowed for you to talk she said I can't 'cos the noisy children	I said, you can talk it's allowed for you to talk. She said I can't because the noise is children.	word error
2	and that's for Maxine 'cos they Jim's moving there	I like that because it's a movie day and a house. And where was he?	proper noun

Table 10: Example score for accuracy of transcription

The transcription accuracy ratings were undertaken by the lead author and separately by specialist speech therapist working with plwMND not connected to this research. Any differences of score were discussed and a rating agreed.

A meaningfully inaccurate Relate caption may not imply poor understanding of a conversation. For example, the Relate captions may have no effect on understandability if neither party looks at the captions during some or all of talk, or relies on another modality to support their understanding.

Unlike other forms of AAC, a VOCA for example, Relate does not replace the speech of plwMND and the interaction does not need to go via the AAC. The ASR captions are generated

automatically as a person speaks and there is no requirement for the listener to utilize, or even look at the captions in conversation. This is one reason why it is difficult to link the level of ASR accuracy directly to the level of conversational understanding. The researcher therefore reviewed each videoed session to measure how many seconds either the conversation partner or plwMND looked at the ASR captions, to get an indication of how frequently they were referenced and what was being captioned when they did.

It should be added that a meaningfully accurate Relate transcription may not imply good conversational understanding – this requires comprehension of the meaning, grammar, or phonetic signal of the turn itself as well as perception of how the turn is constructed in relation to previous talk (Bloch & Wilkinson, 2004).

4.8. Thematic Analysis

Interviews were undertaken with plwMND and communication partner and designed to explore their feedback on using Relate in conversation. The interviews took place as soon as feasible after each recorded interaction using Relate at the preferred location of the plwMND and were conducted by the researcher.

A semi-structured interview approach was used. This allowed for the flexibility of covering key areas the researcher had planned for and enabled participants to discuss issues important to them that the researcher had not anticipated (Clarke & Braun, 2013).

An interview question schedule was created around a set of general questions corresponding to the overall aims of the study and informed by the author who is a senior specialist SLT working with plwMND. The questions themselves were written by the researcher and positioned to generate feedback on both the positive and negative aspects of everyday Relate use and not intended to bias responses one way or another. The researcher always reminded participants that that their candid views were most important –and that they were completely entitled to say anything they wished - positive or negative.

4.8.1. Stakeholder involvement

Stakeholder involvement at an early stage helps identify and prioritize user-relevant topics grounds research in day-to-day reality of users' experiences (Brett et al., 2014), and is consistent with the principles of public involvement in research – being carried out 'with' or 'by' members of the public rather than 'to', 'about' or 'for' them (Hayes et al., 2012). The proposed semi-structured interview was reviewed by a volunteer plwMND and a significant other. This helped to ensure the questions were meaningful to the participants, answered the research question and minimised problematic assumptions embedded the way questions were asked (Clarke & Braun, 2013). Their feedback resulted in minor wording changes for clarity.

4.8.2. Interview Analysis

The interview recordings were orthographically transcribed and pseudonymised with the participants represented as numbers. As no CA was conducted for the interviews, the Jeffersonian transcription system was not used. Thematic Analysis (TA) in the form described

by Braun & Clarke (2006) was used for qualitative analysis of the interview transcripts. TA is a recognised, widely used and accessible method of identifying themes to analyse concepts and ideas in qualitative data (Clarke & Braun, 2013).

4.8.3. Analysis approach and situating one's perspective

Researchers should distinguish if they are conducting an inductive or deductive thematic analysis as it will inform how themes are theorized (Braun & Clarke, 2006a). This analysis was inductive – there was no attempt to fit coding of the data into a pre-existing coding frame. However, as the researcher I acknowledge my familiarity with the topic and that I am not coding in isolation and as such play an active part in searching and identifying patterns and themes in the data. I have 16 years of experience of working with adults with acquired and progressive neurological conditions, including MND. The most recent five years have been working specifically with plwMND as national SLT project lead for assistive technology and voice banking on behalf of the MND Association.

TA needs a theoretical underpinning, and researchers should make their theoretical approach explicit (Terry et al., 2017). My epistemological position is a realist orientation - analysing the interview participants' words and language as though they are a direct link to the participants' experience. There are other positions a researcher may take, for example a Constructionist perspective (focusing on how themes are constructed and how meanings are produced through language and discourse, sometimes considering social, cultural, and power factors), or a Critical perspective (examining underlying power structures and social norms) (Braun & Clarke, 2006).

The research process broadly followed Braun and Clarke's (2006) checklist of criteria for good thematic analysis, including transcription of interviews, reviewed with equal attention given to each data item. NVivo 12 software was used as the analysis tool to code and group individual responses from the transcripts. Further review of the coding enabled sub-themes to be identified and eventually a hierarchy of distinctive main themes and sub-themes was created.

For the interviews, the transcriptions simply documented the words spoken. No CA analysis was completed.

4.9. Additional records for each session

The following additional records were made at each session:

- Number of recordings the plwMND had made to train Relate
- Number of 'custom cards' recorded by plwMND – for example to train Relate with proper nouns. A person can record as many custom cards as they wish to try and improve the recognition accuracy of specific nouns and phrases that they use in daily life. The guidelines encourage recording custom cards as an optional extra activity and at least 24 hours to update the model (Google, 2021). As there is no immediate improvement to ASR accuracy when a person records a custom card, planning ahead is important when considering what words to train.
- Any environmental, technical or situational factors that may affect Relate caption accuracy

All these factors may have an effect on the overall accuracy of Relate transcriptions. It may be feasible that Relate transcription accuracy could improve over time even if dysarthria is increasing, due to advances in Relate technology, the number of recordings a plwMND has made to train Relate, or the microphone or environment is optimised.

It may also be feasible for Relate transcriptions to become less accurate, possibly due to significantly increased dysarthria, a Relate upgrade that negatively benefits a particular type of dysarthria, or a noisier location.

5. Results

5.1. Introduction to the chapter

This section presents the result of the research. Pam was recorded in conversation with Carol ten times over twelve months and the average length of the sessions was just under 14 minutes. Bob was recorded with Adrian eight times over ten months, averaging 10 minutes per session. Elsbeth and Tony were recorded six times in eight months, averaging slightly more than 10 minutes per session.

The FDA-2 dysarthria assessment was administered for each plwMND after sessions 1, 3 and 6 for Elsbeth and Bob, and additionally after session 9 for Pam. Bob experienced a significant decline in his speech, and the other two remained stable throughout the study. For all three plwMND, the actual average WER percentage for each session was almost always higher than Relate's own forecast WER threshold, resulting in messaging from Relate that was likely over optimistic in setting expectations.

For Pam and Bob, between 50-70% of all Relate captions were judged to have lost the meaning of what they said For Elsbeth is was between 60-75%. Although the plwMND seemed to have low awareness of Custom Cards, they appeared to be beneficial when recorded for frequently used words. For Pam, average meaning loss improved from 55% to 24% in phrases where key words had been recorded in custom cards. For Bob the reduction was from 69% to 15% and Elsbeth 95% to 45%.

Pam and Carol gazed very infrequently at the Relate captions during their conversations – Pam referenced them on average 8 seconds per session and Carol 60 seconds. In contrast, Bob looked at the captions increasingly longer than Adrian (average 53 seconds), on average 50

seconds in sessions 1-4, increasing to 195 in sessions 5-8. Elsbeth and Tony displayed another difference in gaze pattern where Tony looked at the captions far more than Elsbeth (256 versus 62 seconds on average).

Using CA, repair of trouble was a frequently occurring feature across the dyad sessions extracts. 13 videoed vignettes were selected to illustrate the diverse sources of trouble and repair, including difficulties understanding speech with dysarthria, ASR caption inaccuracy and dynamically added punctuation by the ASR captions introducing ambiguity of meaning.

Following transcription of the semi-structured interviews, Thematic Analysis was used to identify five themes: speech and identity, strategies for communication, hopes for personalised ASR, set-up and first use, and overall experiences of using ASR in practice.

5.2. ASR model performance

In this study the researcher transcribed each recorded session and then calculated the WER for the session based on the words actually spoken by the plwMND. The calculation was completed using a Google Sheets 'apps script'.

Additionally, the researcher allocated a rating for the accuracy of the Relate transcription to the meaning of what was actually said. The 'accuracy of meaning' ratings were undertaken by the lead author and separately by a specialist speech therapist working with plwMND not connected to this research.

The following section presents the results by plwMND relating to their FDA-2 dysarthria scores, the Relate forecast of accuracy measured by WER compared to actual, meaning preservation

scores, use of custom cards and their effect on WER, how long each dyad referenced the Relate captions and the plwMND overall experience of using Relate.

5.2.1. Pam – personalised ASR performance

5.2.2. FDA-2 Assessment

Pam presented with moderate mixed spastic-flaccid dysarthria characterized by a harsh voice quality and mild hypernasality. The emergence of moderate articulatory weakness was also noted with moderately reduced tongue movement. Her FDA-2 assessment conversation intelligibility subsection was rated at grade C: 'Speech severely distorted; can be understood half the time. Very often has to repeat'.

5.2.3. WER – Forecast vs Actual

Pam recorded 3707 phrases and 70 unique words on custom cards over the period of 10 sessions (approximately 12 months).

Session available from	Total Recorded Phrases (cumulative)	Unique words recorded on custom cards (cumulative)	Forecast Relate WER
1	500		28.2
3	1401		23.0
6	3200	50	20.9
7	3400	70	21.7
8	3600	70	19.8
10	3707	70	21.1

Table 12: Pam – recorded phrases, custom cards and forecast WER

Throughout the period the Relate app categorized the accuracy of Pam's personalised ASR model recognition as 'Medium' (average WER between 15% and 35%). Relate describes a medium score as "the app has improved its ability to recognize your unique speech, but it's likely to make errors".

The researcher calculated the actual average WER per session by transcribing the words Pam said and comparing to the ASR captions. Over the sessions, the actual average WER per session varied between 38.2 – 53.2%, all greater than the 'Medium' category WER range of between 15-35%.

This difference is probably because the actual WER is based on Pam's conversational speech, not a sample of shorter generic phrases. ASR performance is likely to be degraded for longer utterances or spontaneous conversations when compared to shorter phrases (Green et al., 2021).

For Pam's actual WER scores, Relate would categorize them all as 'Low', recommending that Listen mode may work best with short phrases only, and encouraging the user to create custom cards (Listen mode was the only mode Pam used).

5.2.4. Meaning Preservation

For meaning preservation, the percentage of Relate captions judged to have preserved the meaning of what Pam actually said varied between the sessions from 26.5-49.1%. This indicates, at best, more than half of all the captioned phrases were not accurate enough to preserve meaning. At worst, nearly three-quarters of all phrases did not preserve the meaning of what was said.

PAM Session	Duration	FDA-2	Actial Average WER	Relate Forecast WER	Meaning Preserved %	Pam View screen (secs)	Pam View screen (% of session)	Carol View screen (secs)	Carol View screen (% of session)
Session 1	12:23	C	53.2	15-35%	26.5%	2	0.3%	70	9.4%
Session 2	17:35		38.2	15-35%	45.1%	26	2.5%	89	8.4%
Session 3	22:58	C	41.3	15-35%	49.1%	10	1.3%	77	9.9%
Session 4	13:37		48.8	15-35%	34.0%	4	0.5%	105	13.0%
Session 5	16:45		45.5	15-35%	45.7%	6	0.6%	53	5.3%
Session 6	21:15	C	41.8	15-35%	48.6%	0	0.0%	50	3.9%
Session 7	6:01		43.2	15-35%	43.7%	3	0.8%	22	6.1%
Session 8	13:15		44.0	15-35%	38.3%	7	0.9%	43	5.4%
Session 9	17:20	C	46.8	15-35%	47.0%	9	0.9%	78	7.5%
Session 10	8:31		42.1	15-35%	47.4%	4	0.8%	25	4.9%
All			44.5		42.5%	7	0.9%	61	7.4%

Table 13: Relate actual vs forecast WER, meaning preservation for Pam, duration, viewing time

Further analysis showed that accurate transcription of proper nouns seemed to be a particular difficulty for Relate, ranging between 44-83% meaning lost for all proper nouns spoken across the sessions. This is significant because proper nouns may not be easily predictable in the sequential context of an interaction to the listener, possibly needing strategies that take longer to repair than other word classes (Bloch & Saldert, 2020). In other words, proper nouns are harder to deduce if not understood, and could take longer to repair – and at the same time Relate seems to have particularly difficulty with accurate captioning.

5.2.5. Custom Cards

After session five, Pam recorded additional custom cards to try and improve the ASR caption accuracy of 50 proper nouns, including some of her family and friend's names. Pam asked for help from the researcher and Pam agreed for him to document the list of nouns recorded for analysis.

In session six, Pam produced a proper noun 53 times with an overall proper noun meaning loss of 55%. The meaning loss reduced to 24% for the 25 utterances containing a noun that Pam had recorded custom cards for, which is lower than the proper noun meaning loss for any other session (the lowest was session 3 with 44% proper noun meaning loss).

Some of the proper nouns with custom cards spoken by Pam in session 6 were also spoken in prior sessions. The overall meaning loss for those 54 utterances prior to custom cards was 70%.

Pam said the concept of custom cards was difficult to understand and had no idea at first that they may benefit her. As custom card recording is a separate process to recording standard phrases, she said she did not even know the option was there until it was pointed out. Even if a service is accessible, it is important that the system can be understood and navigated easily, effectively and efficiently (Combe et al., 2013).

5.2.6. Referencing the ASR Captions

The ASR captions were very often not referenced by Carol or Pam, and this is one reason why it is difficult to link the level of ASR caption accuracy directly to level of conversational understanding. The ASR captions do not replace Pam's speech and the conversation does not have to go through Relate, unlike other forms of AAC such as an eye-gaze communication device for example.

Table 13 shows, on average, Carol looked at Relate for about 7% of the time (average 60 seconds in total) for a session lasting on average a little under 14 minutes. Pam looked at Relate even less: about 1% of the time (average 8 seconds in total).

5.2.7. Experience of using Relate

Pam reported difficulty with the process of recording phrases and custom cards, and always enlisted her husband's help. She said the instructions were not easy to find or to print out. The experience of recording was also challenging due to fatigue from MND – Relate felt like hard work with little reward.

Pam more than once blamed herself for the inaccurate captions Relate produced attributing fault to her speech and lack of confidence with technology.

5.2.8. Bob – personalised ASR performance

5.2.9. FDA-2 Assessment

Bob initially presented with moderate mixed spastic-flaccid dysarthria characterized by mild hypernasality, moderate articulatory weakness was also noted with moderately reduced tongue movement. After session six Bob presented with severe spastic-flaccid dysarthria, moderate hypernasality, severe articulatory weakness and reduced tongue movement. Bob's FDA-2 conversation intelligibility subsection was initially rated as grade 'C' ('Speech severely distorted – can be understood half the time, very often has to repeat.'). After session six Bob's speech was graded as 'D' ('Occasional words decipherable') for the remainder of the study.

5.2.9.1. WER – Forecast vs Actual

Bob recorded 2284 phrases and 136 unique words on custom cards over the period of 8 sessions (approximately 10 months).

Session available from	Total Recorded Phrases (cumulative)	Unique words recorded on custom cards (cumulative)	Forecast Relate WER
1	500	0	23.6
2	1200	17	19.7
3	1300	22	20.7
4	2000	48	23.7
5	2054	75	24.0
6	2081	100	23.7
7	2137	110	23.7
8	2284	136	24.6

Table 14: Bob – recorded phrases, custom cards and forecast WER

Similar to Pam, the Relate app categorized Bob's personalised ASR model recognition accuracy as 'Medium' throughout the study (average WER between 15% and 35%).

The actual average WER per session was calculated and varied between 33.6 – 42.5%, all but one in the 'Low' category WER range (more than 35%), probably because actual WER is based on conversational speech, not a sample of shorter generic phrases.

5.2.9.1. Meaning Preservation

For meaning preservation, the percentage of Relate captions judged to have preserved the meaning of what Bob varied between the sessions from 27.3-50.0%. At best, half of all the captioned phrases did not preserve the meaning of what Bob said, at worst it was more than 72%.

BOB Session	Duration	FDA-2	Actual Average WER	Relate Forecast WER	Meaning Preserved %	Bob View screen (secs)	Bob View screen (% of session)	Adrian View screen (secs)	Adrian View screen (% of session)
Session 1	9:39	C	33.6	15-35%	50.0%	62	10.6%	57	9.8%
Session 2	8:17		39.4	15-35%	44.0%	22	4.5%	3	0.6%
Session 3	8:00	C	41.2	15-35%	28.9%	24	5.0%	4	0.8%
Session 4	8:49		38.9	15-35%	49.7%	90	16.9%	25	4.7%
Session 5	10:50		40.0	15-35%	29.6%	193	29.7%	85	13.1%
Session 6	11:45	D	41.1	15-35%	30.6%	200	28.4%	78	11.1%
Session 7	9:58		40.8	15-35%	31.4%	175	29.3%	89	14.9%
Session 8	11:59		42.5	15-35%	27.3%	210	29.2%	85	11.8%
All	9:59		39.7		36.4%	122	19.2%	53	8.4%

Table 15: Relate actual vs forecast WER, meaning preservation for Bob, duration, viewing time

5.2.10. Custom Cards

Bob regularly recorded custom cards to try and improve the ASR caption accuracy of 136 proper nouns. Bob completed the custom card recordings independently without support.

Most of the custom cards contained names of family, friends, people and places for social activities. There were a few names of colleagues at work and some technical terms used in the course of his professions.

For every session Bob chose to be recorded at his place of work with Adrian, his friend and work colleague. He did not often talk about his family of friends outside of work in the sessions, and surprisingly referenced few technical terms he had made custom cards for.

Across the sessions, Bob said 13 utterances containing one or more words that had already been recorded with a custom card, with an overall proper noun meaning loss of 15% (11 correct, 2 incorrect). For the 26 utterances with custom cards Bob said prior to the custom card recording, the overall proper noun meaning loss was 69% (8 correct 18 incorrect).

This highlights how contextual the content of conversation can be, and how guidance may be useful to help prompt custom card creation for the subject areas where conversation is likely to occur.

5.2.11. Referencing the ASR Captions

In contrast to Pam and Carol, Bob looked at the captions for much longer than Adrian, on average 9% of the time from sessions 1-4, increasingly so from session 5-8 to around 29%. Bob appeared to increasingly use Relate to check for himself how accurate the captions were for his speech. He said he felt his speech was declining, and the FDA-2 score after session six confirmed this. Adrian looked at the captions less than Bob, on average 9% of the time.

5.2.12. Experience of using Relate

Bob has a background in Information Technology and reported feeling comfortable with technology. He had no difficulty with the process of recording phrases, custom cards and using Relate in practice. However, he felt the time spent recording the phrases was not rewarded with significantly improved performance, which affected his motivation. He also indicated the process was very boring.

Bob said he felt positive about the potential for ASR technology but that it needs to be further developed to be more accurate.

5.2.13. Elsbeth – personalised ASR performance

5.2.14. FDA-2 Assessment

Elsbeth presented with moderate mixed dysarthria, moderate articulatory weakness and moderately reduced tongue movement. Elsbeth's conversation intelligibility rating on the FDA-2 remained at grade 'C' throughout the period of the study ('Speech severely distorted – can be understood half the time, very often has to repeat.').

5.2.14.1. WER – Forecast vs Actual

The Relate app categorized Elsbeth's personalised ASR model recognition accuracy as 'High' (average WER less than 15%). Relate describes a high score performance as "the app hopefully understands you relatively well! But it will make some errors".

Elsbeth recorded 1180 phrases and 120 unique words on custom cards over the period of 6 sessions (approximately 8 months). Her conversation intelligibility rating on the FDA-2 was grade 'C' throughout the period of the study ('Speech severely distorted – can be understood half the time, very often has to repeat.').

Session available from	Total Recorded Phrases (cumulative)	Unique words recorded on custom cards (cumulative)	Forecast Relate WER
1	500	50	8.9
2	561	83	10.1
3	660	83	9.4
4	1002	101	10.4
5	1007	101	10.4
6	1180	120	9.6

Table 16: Elsbeth – recorded phrases, custom cards and forecast WER

The actual average WER per session was calculated and varied between 30.6 – 34%, all within 'Medium' category WER range of between 15-35%. Medium performance is described as "the app has improved its ability to recognize your unique speech, but it's likely to make errors".

5.2.14.1. Meaning Preservation

For meaning preservation, the percentage of Relate captions judged to have preserved the meaning of what Elsbeth said varied from 13-40%. This indicates, at best, 60% of all the captioned phrases were not accurate enough to preserve the meaning of what was said. At worst it was nearly more than three-quarters of all phrases.

ELSBETH Session	Duration	FDA-2	Actual WER	Relate Forecast WER	Meaning Preserved %	Elsbeth View screen (secs)	Elsbeth View screen (% of session)	Tony View screen (secs)	Tony View screen (% of session)
Session 1	13:48	C	31.2	< 15%	37.2%	2	0.2%	225	27.1%
Session 2	6:05		30.6	< 15%	40.0%	93	25.5%	200	54.8%
Session 3	4:40	C	33.0	< 15%	17.8%	25	9.1%	121	44.0%
Session 4	10:23		34.5	< 15%	13.0%	90	14.6%	265	42.9%
Session 5	14:10		32.0	< 15%	18.8%	84	9.8%	353	41.4%
Session 6	14:49	C	34.0	< 15%	22.4%	79	8.9%	372	41.8%
All			32.6		24.9%	62	11.4%	256	42.0%

Table 17: Relate act vs forecast WER, meaning preservation for Elsbeth, duration, viewing time

5.2.15. Custom Cards

Elsbeth recorded 50 custom cards as part of the initial 500 phrases and recorded 70 more over the following six months.

For the initial 50 custom cards, Elsbeth recorded words predicting what she would likely talk about to her PhD supervisor – these were all highly technical terms and relevant to her subject area. In the end she chose not to use Relate in supervision, and the result was almost none of these highly technical words were spoken in the recorded sessions either.

Interestingly, even the 70 additional custom card words were based on specific topics discussed in sessions such as a trip to the Australian outback and nouns such as 'Cassowary' and 'Crocodile' where the ASR captions were incorrect. However, these words were almost never said in subsequent sessions

Across the sessions, Elsbeth said 11 utterances containing one or more words that had already been recorded with a custom card, with an overall proper noun meaning loss of 45% (6 correct, 5 incorrect). For the 44 utterances with custom cards spoken by Elsbeth spoken prior to the custom card recording, the overall proper noun meaning loss was 95% (2 correct 42 incorrect).

This highlights the difficulty of predicting what is going to be said in future conversations and contexts.

5.2.16. Referencing the ASR Captions

Elsbeth referenced the captions on average 11% of the time and appeared to occasionally check if the captions were accurate. However, Tony frequently referenced the ASR captions, on average 42% of the time in a session. He appeared to frequently check if the ASR accurately captioned Elsbeth's words, rather than supporting his understanding of Elsbeth.

5.2.17. Experience of using Relate

Elsbeth reported no difficulty with the process of recording phrases and custom cards. However, she had to take regular breaks because of fatigue from MND affecting her voice.

Elsbeth felt the accuracy of the ASR captions was unacceptably low and the custom cards provided little functional benefit. She tried using Relate in a variety of situations with her friends and family but usually resorted to using other means of AAC when not understood.

Elsbeth said the expectations that were set by a 'High' score were misleading and she felt disappointed with the experience.

5.3. Conversation Analysis

Repair of trouble was a frequently occurring feature for almost all the dyad extracts. Trouble arose from multiple sources, including:

- Conversation partners experiencing difficulties understanding speech with dysarthria
- Transcription inaccuracies in the ASR captions that significantly affected or lost the meaning of what the plwMND.
- Proper nouns appeared to be particularly difficult for the ASR to caption accurately.
- Dynamically added punctuation by the ASR captions that introduced ambiguity of meaning and reduced understandability.

Difficulties understanding speech with dysarthria co-occurring with incorrect captioning sometimes resulted in "cascading troubles" (G. Lerner & Kitzinger, 2012) with subsequently prolonged repairs, and increased risk of trouble understanding the AAC user's subsequent attempts at clarification (Bloch & Wilkinson, 2013a).

Additionally, trouble arising from the ASR caption inaccuracies sometimes resulted in becoming the focus of the interaction itself.

The following video extracts were selected to illustrate the diversity observed when managing trouble and repair with the ASR captions present. The first section provides examples of how the ASR captions contributed to efficient repair resolution, the second section shows how the ASR captions disrupted efficient repair resolution in three different ways: Incorrectly captioned words, contributing to extended or 'cascading' troubles and repair, and when repair becomes the focus of interaction.

Video extracts showing repair sequences have been grouped into the following categories:

Category	Extract
6.3 ASR captions contributing to efficient repair resolution	6.3.1 A little white board 6.3.2 Granddaughter getting married 6.3.3 Vegan Recipes 6.3.4 Some of the players are nearly eighty
6.4 ASR captions disrupting or lengthening repair resolution - Incorrectly captioned words causing multiple troubles	6.4.1.1 Stereo 6.4.1.2 A walk and then coffee 6.4.1.3 The A259 6.4.1.4 Horatio
- Extended repair when the captions may be perceived as accurate, but are not	6.4.2.1 Movie day 6.4.2.2 Vegetarian shepherds 6.4.2.3 Mussolini
- Repair of the ASR captions becoming the focus of interaction	6.4.3.1 Actually Helpful 6.4.3.2 That wasn't bad actually

Table 18: Video extracts grouped by category

5.3.1. ASR captions contributing to efficient repair resolution

The following four extracts show in different ways how the ASR captions appear to support repair, or could have been helpful if they were referenced, or demonstrate that a shift of eye gaze towards the captions (regardless of their accuracy) could be enough to indicate trouble and trigger self-repair.

In the first extract 'A little white board' Pam appears to over-articulate her speech, presumably to support Carol's understanding, and as an outcome the ASR caption that Carol looks at becomes accurate, supporting repair. In 'Granddaughter getting married' the ASR incorrectly captions Pam's speech, but appears to contribute to Carol's understanding to what Carol may have believed Pam actually meant, demonstrated by Carol initiating an other-initiated other repair (OIOR). In 'Vegan Recipes' the ASR Captions are accurate when Carol identifies trouble, but she does not look at them, and Carol initiates an other-initiated self-repair (OISR) directly with Pam instead. Finally in 'Some of the players are nearly eighty' Pam appears to be prompted to self-repair just by Carol looking at the ASR captions, regardless of caption accuracy.

5.3.2. A little white board

In this extract, Pam appears to over-articulate her speech, possibly to support Carol's understanding, and as an outcome the ASR caption that Carol looks at becomes accurate, supporting repair.

Pam is talking about a visit from her speech therapist (SLT). The SLT is going to be visiting 'on Thursday' (line 1) and the ASR captions her words correctly (line 2). Pam adds that the SLT will bring a little board (line 5) - referring to a small whiteboard, a type of AAC that Pam can use to handwrite words if someone doesn't understand what she is saying. This phrase is incorrectly captioned by Relate to be '***a little white boat***' (line 6), but Carol does not look at Relate at this point. After a pause of 1.0 second Carol follows with an other-initiated repair (OIR), **'A little?'** (line 8) which is a partial repeat of what Pam said on line 5, locating the trouble source in Pam's prior turn. Pam responds with a self-repair attempt **'a little white board'** (line 10) which this

time is captioned correctly by Relate. Carol looks at Relate and repeats the word 'board' (line 12), to which Pam replies **'yeah'** to confirm accuracy of repair and the action proceeds.

Extract 1: A little white board (the link to the transcript master is here)

```
Time   Line Speaker  Talk
00:00   1   Pam:     On Thurs[day                                    ]
        2   Relate:  on Thursday
                             [((Pam looks down at Relate and looks back))]
        3   Carol:           [oh right (.) °yeah°]
                             [((Carol nods))     ]
        4   Relate:  Right?
        5   Pam:     [She's bringing me a little board]
        6   Relate:  She's bringing me a little boat
                     [((Pam and Carol look down at the table, Pam gestures with her fingers and back
        7            (1.0)
        8   Carol:   A little?
        9   Relate:  A little?
       10   Pam:     A little [white board                                  ]
       11   Relate:  A little white board
                              [((Pam and Carol look at Relate, Pam looks back at Carol))]
       12   Carol:   (0.6) board
       13   Relate:  pulled
       14   Pam:     yeah
       15   Relate:  Yeah.
00:08                [((Carol looks back at Pam))]
```

Delays between speaking turns that exceed about one second (such as at line 7) may potentially be regarded as problematic or accountable (Sacks et al., 1974b). Person interaction places a demand on participants to produce talk in a timely manner, while maintaining mutual understanding (Enfield & Sidnell, 2017) .

Carol identifies trouble at line 8 and says **'a little?'**, a repair initiator which is a category-specific interrogative and Carol is claiming a grasp of the trouble-source, in this case a partial repeat (Schegloff & Jefferson, 1977). The repair solution is usually fitted to the format of the repair

initiation - namely trouble-source speakers generally respond to category-specific interrogatives with a repeat of the relevant categorical term (Kitzinger, 2012).

At line 10, Pam added **'white'**, presumably to be more specific that she was referring to a whiteboard, adding context. Sometimes a trouble-source turn is amended in repair to include full-form references (Schegloff, 1989).

Pam also appears to over-articulate her speech in line 10 when saying **'board'** when compared to line 6, self-repair is sometimes initiated by a modification in articulation (Schegloff, 2004). The action of repair for dysarthric speakers can go beyond simple repetition of a previously troublesome turn and extend to linguistic aspects (Rutter, 2009): speakers with dysarthria may adjust the phonetic characteristics of a repetition repair of a trouble source turns by increasing loudness, increasing duration, or expanding pitch range – all to 'upgrade' intelligibility (Curl, 2005). The accurate Relate caption of '***board***' (line 11) may have supported Carol and Pam to quickly repair trouble as part of an OISR sequence and to move on, though it is unclear whether Pam was aiming to benefit Carol's understanding directly, or to improve ASR accuracy, or both.

Nevertheless, the single self-repair effort by Pam was sufficient for accurate ASR captioning of the phrase and in turn was intelligible and understandable to Carol. The efficiency of this repair compares favourably with trouble resolution for non-speech-disordered conversation, which is often (but not always) achieved with a single repair effort (Schegloff, 2020).

5.3.3. Granddaughter getting married

In this extract, the ASR incorrectly captions Pam's speech, but appears to contribute to Carol's understanding to what Carol may have believed Pam actually meant, demonstrated by Carol initiating an other-initiated other repair (OIOR).

Pam is telling Carol about a friend whose granddaughter is 'getting married next Saturday' (line 1). The ASR caption at line 2 is incorrect '***he is going to get married next Saturday***', although Carol does not look at the captions. Carol indicates her understanding at line 3 'Oh nice'.

The friend is unwell, but he is now not undergoing chemotherapy. However, Carol appears to identify trouble at line 5 when Pam says **'they are off chemo'**. Carol looks at the ASR captions for 2.2 seconds which Pam appears to take as a cue to repeat what she may think is the trouble source at line 7 **'off it'**. However Pam initiates an other-initiated other repeat (OIOR) by repeating **'<u>he</u> is'** at line 11 and Pam confirms the repair at line 12 by saying **'yeah'**.

Extract 2: 'Granddaughter getting married' (the link to the transcript master is here)

Time	Line	Speaker	Talk
00:00	1	Pam:	His grand daughter's getting married next [Saturday]
	2	Relate:	***He is going to get married next Saturday***
	3	Carol:	[Oh nice]
	4	Relate:	***night***
	5	Pam:	They are off chemo
	6	Relate:	***so he is off chemo***
			(2.2) Carol looks down at Relate
	7	Pam:	[°off it°]
	8	Relate:	***after***
	9	Carol:	[right <u>he</u> is]
	10		[Carol nods and raises/lowers her left hand]
	11	Relate:	***he is***
	12	Pam:	yeah
00:13	14	Relate:	***here***

Relate did not accurately caption Pam's spoken words **'they are off chemo'** at line 5, instead captioning ('***so he is off chemo***'). The action of Carol looking at the ASR Captions, combined with the likely problematic and accountable delay of 2.2 seconds (Clark, 1996) after line 6 seemed to prompt Pam to self-initiate repair. Repair initiations can be accompanied by gestures such as a head tilt or head poke (Seo & Koshik, 2010), and these gestures can be used to initiate repair without verbal accompaniments (Kitzinger, 2012).

Even after looking at the (incorrect) captions, Carol appeared to understand that Pam actually said **'*they're* off the chemo'**, as she then initiated a OIOR at line 9 (**'right <u>he</u> is'**) to confirm it was only one person that has stopped having chemotherapy. Although the captions were incorrect, they may have supported Carol to repair what Pam presumably should have said – and the repair is confirmed at line 12 when Pam says **'yeah'**.

Bloch and Wilkinson report that mis-produced AAC output due to an error by a user or an incorrect/unintelligible output signal produced by the AAC device may lead to problems with

understandability, extending beyond the anticipated effects of unintelligibility (Bloch & Wilkinson, 2011), however in this case, the mis-produced ASR captions may have actually contributed to understandability.

5.3.4. Vegan Recipes

In this extract it is observed that Relate is not utilised by Carol even though it was accurate and may have been helpful for repair. Rather than referencing the ASR captions, Carol directly highlights to Pam the specific word causing trouble.

Pam is telling Carol about her friend Ella who is a vegan: **'vegan she looked at a vegan recipes'** (line 1). The ASR captions are mostly accurate at line 2: '***Vegan, and she looked up or vegan recipes.***' – captioning both references of 'vegan' correctly. However, Carol does not look at the ASR captions. Carol identifies trouble at line 3 evidenced by a 1.2 second delay and a repetition of she thinks she heard at line **'vegan? vegan?'**. Pam confirms repair at line 4 by saying **'yes yeah Ella's a vegan'**, though there may be still be trouble for Carol as there is an additional 1.5 second delay at line 6 and a question **'Ella is as well?'** which Pam confirms is the case at line 8 (**'yeah'**).

Extract 3: 'Vegan recipes' (the link to the transcript master is here)

Time	Line	Speaker	Talk
00:00	1	Pam:	Vegan she looked at a vegan recipes
	2	Relate:	***Vegan, and she looked up or vegan recipes.***
	3	Carol:	(1.2) °vegan?° [°vegan?°] ((Carol laughs))
	4	Pam:	[yes] yeah Ella's a vegan
	5	Relate:	***Yeah, Ellis is vegan***
	6	Carol:	(1.5) Ella is as well?=
	7	Relate:	***Eloise as well***
00:12	8	Pam:	=yeah

Carol's partial, virtually identical, final-rising-intoned repeat at line 3 (**'vegan? vegan?'**) usually locates just one component of the phrase as the repairable (Schegloff et al.,1977), but here it is unclear whether one or both references to 'vegan' on line 1 are the trouble source. For other-initiated repair (OIR) where the repair initiator specifies the trouble source as a particular word, the repair execution is often a reduced form of the turn containing the trouble source (Rutter, 2009).

Pam responds at line 4 by repeating vegan – **'yes Ella's a vegan'** which may be treating Carol's repair initiator as claiming she may not have adequately heard the words (Robinson & Kevoe-Feldman, 2010). When Carol produces a conditionally relevant response (Schegloff & Sacks, 1973) at line 6 **'Ella is as well?'**, Carol appears to treat Pam's repair (at line 4) as sufficiently resolving the trouble, despite a potentially accountable 1.5 second pause (Clark, 1996) at line 6. At no point does Carol look at Relate in this interaction, and yet Relate could have been helpful as it correctly transcribed 'vegan' three times.

5.3.5. Some of the players are nearly eighty

This extract shows that even when the ASR captions are wrong, just by looking at the ASR captions could be a contributory factor to prompt an OISR.

Pam is talking about when playing at her local Badminton club, some of the older members give her moral support: **'they praise me they cheer me on old ladies'** (line 1). The ASR captions incorrectly transcribe the words at line 2 ('***they praised me. They treat me. They are a lot.***') but Carol does not look at them at this point. Carol identifies trouble at line 3 (**'there some of them there are near nearly eighty'**), indicated by a delayed response of 1.7 seconds at line 5 where Carol looks at the ASR captions and back to Pam. Pam then restates the phrase in slightly different wording at line 6 (**'Some of the players are nearly eighty'**) and Carol again looks at the ASR Captions. They again do not accurately reflect what Pam says, but this time they may convey more of the meaning of what Pam means ('***someone that plays, so nearly 80***' line 7). Carol then appears to acknowledge repair by saying **'Gosh'** at line 8, and the action proceeds.

Extract 4: Some of the players are nearly 80 (the link to the transcript master is here)

Time	Line	Speaker	Talk
00:00	1	Pam:	They praise me (.) they cheer me on old ladies
	2	Relate:	***They praised me. They treat me. They are a lot***
	3	Pam:	There some of them there they [near nearly eighty]
			[((Carol looks at Relate and looks back at Pam))]
	4	Relate:	***lately that someone that they and they only edit***
	5		(1.7)
	6	Pam:	[Some of the players are nearly eighty]
			[((Carol looks at Relate and looks back at Pam))]
	7	Relate:	***someone that plays, so nearly 80***
	8	Carol:	Gosh
00:19	9	Relate:	***go***

Pam appears to take Carol's gaze at the ASR captions at line 5 combined with the accountable delay of 1.7 seconds (Clark, 1996) as a cue to restate what she thinks the trouble may be (**'Some**

of the players are nearly eighty' at line 6), although Pam at no point looks at the ASR captions herself. The presence of AAC may have a significant effect on how participants organize their interaction and how problems in interaction, when they arise, are managed (Bloch & Wilkinson, 2004) – including just when the AAC captions are briefly looked at, as appears to be the case here.

Although Carol appears to acknowledge repair when she says **'gosh'** at line 8, there is little in the response to confirm she has understood what Pam said at line 6. Using the term 'gosh' at this point seems a conditionally relevant response (Schegloff & Sacks, 1973) and appears to be enough indication of successful repair to let the action proceed, but does not contain content that shows Carol understood Pam's point. 'Gosh' may function as a 'change-of-state' token (similar to 'oh' in this context) in response to some prior action (Heritage, 2010) and may act as a 'news receipt' that acknowledges and assesses the new information that has 'just now' been noticed or realized (Heritage, 1998; Maynard, 1997). However it could also be used as an interactional resource rather than a direct index of the speaker's internal state - moving on with the conversation that has been temporarily stalled, where one chunk of talk has been brought to a possible completion and nothing else has yet been launched (Bolden, 2006).

5.3.6. ASR captions disrupting or lengthening repair resolution

The following extracts demonstrate different ways incorrect ASR Captions can disrupt and/or lengthen repair. The first four examples show how specific words that are incorrectly captioned, particularly proper nouns, can cause multiple troubles, requiring extended repair sequences, sometimes including gesture and additional forms of AAC. The following three extracts demonstrate how extended repair is required when the captions are perceived as accurate in some way, but are not. The final two extracts show how repeated attempts to repair incorrect ASR captions can become the focus of the interaction.

5.3.7. Incorrectly captioned words, proper nouns, causing multiple troubles

The first four examples show how multiple troubles can arise from specific words that are incorrectly captioned (frequently proper nouns), requiring extended repair sequences, sometimes including gesture and additional forms of AAC.

In 'Stereo', Bob reverts to using gesture and describing what the word 'stereo' means when Adrian does not understand him and the ASR captions are incorrect. However for Pam it is harder for her to use gesture or words that describe the proper noun 'Wyvale' In 'A walk and then coffee' when the ASR captions are repeatedly incorrect, and she eventually uses another form of AAC (a whiteboard and pen) to write the word. In 'The A259' after multiple attempts at repair, Carol can repeat the word Pam says ('A259') but she does not appear to understand the meaning of the term in the context of the interaction and Pam once again uses her whiteboard to write the word to complete repair (the ASR captions are always incorrect). In 'Horatio', Bob repeatedly focuses on correcting the ASR captions for the proper noun 'Horatio'

when it is apparent that Adrian has trouble understanding Bob. Eventually Adrian understands the word Bob says but it is unclear if he understands the point Bob has actually made.

5.3.8. Stereo

In this extract multiple troubles arise from a specific word ('Stereo') that is incorrectly captioned, requiring extended repair sequences, including use of gesture.

In 'Stereo', Bob reverts to using gesture and describing what the word 'stereo' means when Adrian does not understand him and the ASR captions are incorrect.

In this extract, Bob initially talks about a colleague that Bob and Adrian had recently met - **'it was good to to to see him in in New York'** (line 1) and Adrian agrees **'yeah…he's always a good person to talk to'**(lines 3-4).

After a 1.5 second pause (line 6) Bob says **'cool'** (line 7) and after a further 2.6 seconds (line 9) initiates a new topic by pointing at the ASR Captions and saying **'I mean if it if it had stereo'** (line 10), reverting to an earlier discussion in the session about how Relate should have stereo microphones to capture different voices (not part of this extract).

Adrian appears to identify trouble at line 12 evidenced by an extended pause of 2.9 seconds and both Adrian and Bob look at the ASR captions. However the captions are highly inaccurate: '***I mean if it even say go***' (line 11) and he looks back at Bob.

Bob repeats what he may think is the trouble source at line 13 (**'stereo'**). Bob and Adrian look at the ASR captions again and they are again inaccurate at line 14 ('***to you***') and Adrian looks back to Bob.

After 1.7 seconds, Bob repeats **'stereo'** once more (line 15) and makes a circular motion with his arm and looks at Adrian. Adrian looks at the ASR Captions, which are again wrong '***see you***' (line 16) and looks back at Bob.

Bob gestures with his hands an object at the side of his left and right ears and repeats **'stereo'** at line 17. Adrian again looks at the captions and they are incorrect once more at line 18 ('***working you'***) and looks back at Bob, this time slightly shaking his head.

After an additional 1.4 seconds Bob attempts to describe a stereo at line 19 ('it has two two speakers'). The ASR captions are incorrect again at line 20 ('***Know choose to to speakers***') but Adrian does not look at them this time.

After a 1.4 second delay Bob gestures 'two' using his fingers on his right hand and says **'two'** (line 21). The captions are incorrect once more at line 22 ('***to you***') but Adrian again does not reference them. At line 23 Adrian appears to confirm repair by saying **'oh yeah like stereo sound yeah'** and the action proceeds.

Extract 5: Stereo (the link to the transcript master is here)

```
00:00  1  Bob:      good .h it was it was good .h to to .h to see him .h in in New York
                    ((Bob looks at Relate and leans towards it))
       2  Relate:   Because it was good to see him in New York
       3  Adrian:   yeah [yeah yeah .h it were good                    ]
                         [((Bob points at Relate and Adrian looks too))]
       4            [He's always a good person to talk to (cannot decode)    ]
                    [((Adrian looks away to his left and back at Bob and nods))]
       5  Relate:   Yeah.Yeah. Yeah. Good. It ought to be best starter
       6            (1.5)
       7  Bob:      Cool
       8  Relate:   home
       9            [(2.6)                                                  ]
                    [((Bob looks and points at Relate and Adrian look at Relate))]
      10  Bob:      [I mean .h if it if it had stereo                    ]
                    [((Bob points at Relate and and Adrian looks at Bob)) ]
      11  Relate:   I mean if it even say go
                    ((Adrian looks at Relate and back at Bob))
      12            (2.9)
      13  Bob:      um .h um .h um .h um .h stereo
      14  Relate:   to you
                    ((Bob and Adrian look at Relate and Adrian looks back at Bob))
      15  Bob:      (1.7)[stereo                                                          ]
                         [((Bob looks at Relate, makes a circular motion with his arm and looks at Adrian))]
      16  Relate:   see you
                    ((Adrian looks at Relate and back at Bob))
                    ((Bob gestures an object to the side of his left and right ears using his hands))
      17  Bob:      (1.5) um .h um .h stereo
      18  Relate:   Working you
                    ((Adrian shakes his head slightly at Bob))
      19  Bob:      (1.4) it has (1.3) two .h two speakers
                    ((Bob gestures pointing to two areas in front of him ))
      20  Relate:   Know choose to to speakers
      21  Bob:      (1.4) [two                                              ]
                          [((Bob raises his right hand and indicates two fingers))]
      22  Relate:   to you
      23  Adrian    Oh yeah like stereo sound yeah
01:04 24  Relate:   Oh yea yes like stereo sound yeah
```

When Bob says **'Cool'** (line 7) the term may be an interactional resource ahead of initiating a topic change (in this case moving between the seemingly unlinked topics of discussing a mutual colleague to a missing feature in the Relate app). By saying 'Cool' Bob could be signalling he is

about to move on from a chunk of the conversation that has been brought to a possible completion and nothing else has yet been launched (Bolden, 2006).

However, successful management of topic transition may be more difficult for people with dysarthria – it may not just rely on words, but also appropriate timing, prosody and non-verbal actions (Bloch et al., 2015). It seems that Adrian experiences trouble right from the start of the topic change and indicates there is trouble by not responding after lines 10, 13, 15, 17 and 19 and on one occasion shaking his head (line 19). However at no point does he indicate what the trouble actually is. By not clarifying the nature of the trouble Adrian may be attempting to avoid communicating the stance that the issue is with Bob's speech being hard to understand (although the ASR Captions are also consistently inaccurate too). Other-initiated repair (OIR) can make it difficult to avoid suggesting that someone else is responsible for the trouble. This happens because the structure of OIR is often context-independent, which can unintentionally imply that the issue lies with the other person (Sacks et al., 1974). Bob nevertheless assumes it is the word 'stereo' causing trouble, perhaps because he sees it is consistently incorrectly captioned by Relate (although some other words are incorrectly captioned too, for example line 11 'had' was captioned as '***even***'). Initially Bob simply repeats what he believes is the trouble-source word. Trouble-source turns are often simplified on 'repeat' because elements of the trouble-source turn may be dispensed with on resaying (Schegloff, 2004b) and for speakers with limited motor speech abilities, focusing of motor effort on making one or two words more intelligible may well be easier than producing a whole sentence (Bloch & Wilkinson, 2011).

From the second and third repair attempts (lines 15 and 17), Bob adds gesture to try and support the repair. People using AAC may use gesture in addition to vocalization to try to speed

up communication (Seale et al., 2020). On the fourth and final repair attempt, Bob tries to describe the word stereo rather than say it again: **‘it has two two speakers’** (line 19). Elements of a trouble-source turn may be dispensed with on resaying - for example ‘technical terms’ can be replaced with ‘lay’ versions (Kitzinger & Mandelbaum, 2013).

Adrian appears to confirm repair at line 19 (**‘oh yeah like stereo sound yeah‘**) but it is unclear if repair was indeed achieved, as Bob's point was about how Relate would benefit from stereo recording, not playing in stereo sound. If Adrian confirmed repair without actually understanding Bob’s point, this may be another indication of Adrian’s unwillingness to allocate trouble responsibility to Bob.

5.3.9. A walk and then coffee

In this extract multiple troubles arise from a proper noun (‘Wyvale’), requiring extended repair sequences, that is eventually repaired by writing on a whiteboard - an additional form of AAC.

In ’Stereo’, Bob reverts to using gesture and describing what the word ‘stereo’ means when Adrian does not understand him and the ASR captions are incorrect. However for Pam it is harder for her to use gesture or words that describe the proper noun ‘Wyvale’ In ‘A walk and then coffee’ when the ASR captions are repeatedly incorrect, and she eventually uses another form of AAC (a whiteboard and pen) to write the word

Carol is asking Pam about the route of the walk they both plan to do soon **‘you got ideas for the 17th for walking?’** (line 1). It seems like there is some secrecy surrounding it **‘you going to**

tell me?' (line 1), **'I won't tell anybody heh heh'** (line 3). Pam says **'Green Lane'** (line 7) while pointing to her left, presumably in the general direction of the road. Carol appears to recognise the road by saying **'oh yes'** (line 9) and then Pam says **'to the 'Wyvale Centre'**, pointing to her left again (line 11), to **'have a coffee'** (line 13), gesturing drinking from a cup with her hand. Carol identifies trouble at line 15 and initiates an OISR (**'the what centre'**) and Pam repeats the part of the phrase Carol indicates she does not understand **'Wyvale'** (line 17). After a 0.5 second pause Carol appears to try guessing what Pam has said **'While? Why flower? Wild?'** (line 19) and Pam repeats **'Wyvale'** twice, the second time with emphasis (line 20).

Carol looks at the ASR Captions for the first time in the interaction but the captions are incorrect: '***Why are I fell?***' (line 21). There is a 7.6 second gap where Carol is looking at the ASR captions, and Pam reaches over to her whiteboard (a type of AAC) and writes 'Wyvale'. Upon reading the word, Carol accurately repeats **'Wyvale'** and apologises, saying it is 'difficult': **'Wyvale! Sorry yes ugh that's difficult Wyvale yes'** (line 23), and the action proceeds.

Extract 6: A walk and then coffee (the link to the transcript master is here)

Time	Line	Speaker	Talk
00:00	1	Carol:	You got ideas for the 17th for walking? (0.5) You going to tell me?
	2	Relate:	***You guys are 17th. Walking, you going to tell me?***
	4	Pam:	[yes]
	5	Carol:	[I won't] tell anybody °heh heh°
	6	Relate:	***I will tell you***
	7	Pam:	[Green Lane]
			[((Pam lifts her arm and points to her left))]
	8	Relate:	***Green Lane***
	9	Carol:	Oh (.) yes!
	10	Relate:	***Oh yes***
	11	Pam:	[To the why (0.5) to the Wyvale Centre]
			[((Pam stretches her arm pointing to her left and brings arm down))]
	12	Relate:	***You bye-bye centered***
	13	Pam:	[have a coffee]
			[((Pam gestures drinking from a cup with her hand))]
	14	Relate:	***I have a coffee***
	15	Carol:	The what centre?
	16	Relate:	***It's what center***
	17	Pam:	Wyvale
	18	Relate:	***Why the Well going?***
	19	Carol:	(0.5) [While? why flower? Wild?]
	20	Pam:	[cof cof] Wyvale .h Wyvale
	21	Relate:	***Why are I fell?***
			((Carol looks at Relate, Pam reaches over to her whiteboard and pen))
			((Carol then looks at the whiteboard as Pam writes 'Wyvale'))
	22		(7.6)
	23	Carol:	>Wyvale!< (.) Sorry (.) Yes (.) Ugh that's difficult (.) Wyvale (1.1) °yes°
00:35	24	Relate:	***Why fell? Sorry. Yes***

Carol identifies trouble at line 15 and says **'the what centre'**, an OISR which is a category-specific interrogative which claims a grasp of the trouble-source, in this case referring to a place (Schegloff & Jefferson, 1977). The repair solution is usually fitted to the format of the repair initiation - namely trouble-source speakers generally respond to category-specific interrogatives with a repeat of the relevant categorical term (Kitzinger, 2012), and this is what Pam does at line 17 (**'Wyvale'**).

However Carol produces a second repair initiation, treating the response to the first as inadequate for resolving the understanding problem. This time she proposes variations on repeats of the trouble-source turn which claim the capacity to reproduce it, but not necessarily to understand what is meant by it - at line 19 (**'While? Why flower? Wild?'**) (Robinson & Kevoe-Feldman, 2010).

This is a 'stronger' format for OIR than the category specific interrogative Carol first used, and usually indicating a better understanding of the trouble (Schegloff et al., 1977). Possibly Carol recognised the first syllable of Wyvale from Pam's repair attempt at line 17 but has not yet grasped the whole word.

Pam then attempts repair again at line 20, this time repeating **'Wyvale'** twice, the second time louder and with more emphasis. Speakers with dysarthria may adjust the phonetic characteristics of a repetition repair of a trouble source turns by increasing loudness, increasing duration, or expanding pitch range – all to 'upgrade' intelligibility (Curl, 2005), going beyond simple repetition of a previously troublesome turn and extend to linguistic aspects (Rutter, 2009).

However, proper nouns such as 'Wyvale' are often unique and may not be predictable from the sequential context, potentially requiring extended strategies (as is the case here) than might be the case with other classes of word (Bloch & Saldert, 2020).

Carol indicates that the repair attempt is again unsuccessful by not responding and instead looking at the ASR Captions (line 22), however the captions have been consistently inaccurate throughout the interaction, captioning 'Wyvale' as ***'bye-bye centered'*** (line 12), ***'why the Well going?'*** (line 18), ***'Why are I fell?'*** (line 21). This example demonstrates the difficulty Relate can

have recognizing proper nouns (however if a proper noun contains high-frequency commonly used terms then Relate may caption more accurately, for example ***'Green Lane'*** - line 8).

Pam then resorts to using handwriting on a whiteboard to repair trouble now that neither her speech nor the ASR captions have helped Carol with the repair. Once Pam has written 'Wyvale' (line 22) Carol confirms repair by accurately repeating **'Wyvale'**, accounting for the problem, saying it is 'difficult': **'Wyvale! Sorry yes ugh that's difficult Wyvale yes'** (line 23).

Handwriting-in-interaction has been reported to be less problematic than other forms of AAC (Bloch & Clarke, 2013) and it seems the case here too in comparison to the ASR Captions. This raises the question why Pam does not choose to use her whiteboard more when there is trouble. For Pam, her dysarthric speech appears to remain the preferred modality for communication, and handwriting seems to be the 'fallback' option to resolve extended repair sequences as needed. Even for those who are significantly difficult to understand due to changes in their speech, using speech to communicate may remain the first preference in most situations, (Smith & Connolly, 2008), even when speaking is effortful and they are often not understood (Beukelman, Fager, et al., 2007).

At line 23 Carol produces 'sorry' which is an apology that orients to her commission of, and accountability for, a possible offense (Robinson, 2004). Carol attributes 'Wyvale' to be a 'difficult' word, although she does not clarify why. Carol may be utilising an apology to mitigate the potential risk of supporting Pam's previously declared stance of self-blame when others do not understand her (Carol has reported that Pam frequently blames herself for others not understanding her).

5.3.10. The A259

This extract shows how Pam resorts to her whiteboard for repair after multiple unresolved troubles and attempts at repair using OISRs and/or referencing ASR incorrect captions.

Pam is telling Carol about a trip she made with others driving on the A259 (the name of a road) **'they took by A259'** (line 1) and Pam adds that a new crematorium is being built (by the road) **'they're building crematorium'** (line 1).

Carol identifies trouble at line 3 evidenced by her looking at the ASR captions and then back at Pam. The ASR captions are inaccurate: '***the device eight to five to nine and the beginning***' (line 2). Pam seems to take Carol's gaze at the ASR Captions as a OISR and restates what she thinks is the problem: **'they're building a crematorium'** (line 5). The ASR captions are again inaccurate ('***the building a grandma***' line 6) but Carol does not look at them at this point, and demonstrates repair by repeating the phrase Pam said: **'oh they're building a crematorium'** (line7). Carol explains her misunderstanding that she thought she **'caught the word but I didn't the connection'** (lines 8-10).

Pam then appears to explain the connection at line 12 **'yeah on the A259 so we were delayed'**. However Carol identifies trouble again at line 14 **'what did you say about a number?'** and Pam attempts repair at line 16 **'the A259 this road this road'**. The repair appears unsuccessful as Carol does not respond and looks at Relate for 7.5 seconds. Again the ASR captions are inaccurate: '***D39 is road, it's road***' (line 17). Pam looks at Relate too. Carol eventually checks that Pam is talking about a road name **'are you telling me the name of the road'** (line 19) and Pam confirms this to be the case, and attempts repair for the second time **'yeah I number A A259'** (line 22). Carol repeats what she has heard but appears not to have understood that this

is a road name **’80 5 9?’** (line 24). Pam tries repair for the third time by saying **‘yeah A A 259’** (line 26) but is again unsuccessful.

Pam then reaches for a whiteboard and pen and handwrites **‘A259’**. Carol confirms repair at line 29 **‘oh A2! It's that it's the A2 I didn’t get yeah A259’** and adds that the ASR captions also did not accurately caption the road name **‘and neither did that’** (line 29).

Extract 7: The A259 (the link to the transcript master is here)

The A259

Time	Line	Speaker	Talk
00:00	1	Pam:	they took by A25 (·h) 9 (·h) they're building crematorium
	2	Relate:	***the device eight to five to nine and the beginning***
	3	Carol:	[n:oh]
			[Carol looks at Relate and looks back to Pam]
	4	Relate:	(n/a)
	5	Pam:	they're building (·h) a crematorium
	6	Relate:	***the building a grandma***
	7	Carol:	oh they're building a cr(h)emat(h)orium
			[Carol looks at Relate and looks back at Pam]
	8		[I thought caught that word
	9		but I didn't get the c(h)onnection (·h)
	10		they're b(h)uilding a [crematorium]]
	11	Relate:	***that would I didn't get the connection, building a grandmother***
	12	Pam:	[yeah on the A259 (·h) so we were delayed]
	13	Relate:	(n/a)
	14	Carol:	what did you say about a number?
	15	Relate:	(n/a)
	16	Pam:	the A259 this road(.) (2.0) this road(.)
	17	Relate:	***D59 is road, it's road.***
			((Carol looks at Relate, Pam looks at Relate too))
	18		(7.5)
	19	Carol:	the [name you telling me the name of the road]
	20	Pam:	[A]
	21	Relate:	***So, you told me the name of the road***
	22	Pam:	yeah I number (·h) A A259
	23	Relate:	***I had a number 88259***
	24	Carol:	80 5 9?
	25	Relate:	***859***
	26	Pam:	yeah A A (·) 259
	27	Relate:	***888-8259***
	28		((Pam reaches over to her whiteboard and pen))
			((Carol then looks at the whiteboard as Pam writes 'A259' then looks at Relate))
	29	Carol:	oh A2! It's that it's the A2 I didn't get [yeah A259 (0.8) °and neither did that°]
	30	Pam:	[mmm]
01:05	31	Relate:	***Another day***

Pam appears to take Carol's gaze at the ASR captions at line 3 as a non-verbal OISR to restate what she thinks the trouble may be (**'they're building a crematorium'** at line 5). Gestures can initiate repair without verbal accompaniments (Seo & Koshik, 2010) and the presence of AAC may have a significant effect on how participants manage problems in interaction, when they arise (Bloch & Wilkinson, 2004) – including here, when the AAC captions are briefly looked at.

However, the ASR captions are highly inaccurate for either proper nouns 'A259' or 'crematorium': '***the device eight to five to nine and the beginning***' (line 2), demonstrating once again the difficulty the ASR Captions has at recognising proper nouns.

The reason why Pam made a reference to a crematorium at line 1 may not have been clear to Carol in the context of the story, and was only explained by Pam on line 12 **'yeah on the A259 so we were delayed'**. Even when turns are treated by recipients as intelligible (the words are identifiable) they may not be understandable in the context, why they are present at this point and what action they represent (Bloch & Wilkinson, 2004). Successful management of adding new information with limited context may be more difficult for people with dysarthria too (Bloch et al., 2015), and it is unclear what the source of the trouble was for Carol – Pam's dysarthric speech or provision of new information with little context, or both.

At line 8-9, Carol allocates trouble to her own understanding of where a crematorium fits into the story, and away from Pam's dysarthric speech production when she says **'I thought caught that word but I didn't get the connection'** (line 8-9). Responsibility of OIR is often allocated in context sensitive ways to claim ownership to 'self' or 'other', or some combination (Robinson, 2006).

Also at lines 8-9, Carol appears to use laughter as an emotional display in a way that is consequential for the interaction, and a resource for maintaining order (Wilkinson & Kitzinger, 2006). In this case, Carol is seeming to add laughter particles that may support the stance of her personal responsibility for a lack of understanding of how the crematorium fitted into Pam's story is to blame for the trouble – perhaps in a self-deprecatory way. Laughter is argued to be a systematically produced, socially organised activity that can be deployed for specific purposes in interaction, rather than it being a failure of self-regulation of emotion (Jefferson et al., 1987)

Pam subsequently reciprocates Carol's facial expression of smiling and generation of laughter particles, consistent with the pattern where a sequence of laughing alone can precede laughing together (Sacks, 1995). Facial expression seems to help ensure that the recipient is able to understand and reciprocate the stance taken in the telling (Ruusuvuori, 2012), and work to preserve mutual affiliation (Wilkinson & Kitzinger, 2006).

At line 14, Carol again identifies trouble: 'what did you say about a number?' and initiates an OISR in the form of a category-specific interrogative - claiming a grasp of the trouble - source (Schegloff & Jefferson, 1977) referring to line 12 **'yeah on the A259 so we were delayed'**. Pam repeated the trouble-source in a simplified way, adding that the term 'A259' is a road (**'the A259 this road** (2.0) **this road'** – line 16) . Trouble-source turns can be simplified (Schegloff, 2002) and can clarify 'technical' terms with 'lay' versions (Kitzinger & Mandelbaum, 2013).

Carol remains unable to resolve the trouble evidenced by her non-response and by looking at the ASR captions for a period of 7.5 seconds. The captions are again incorrect: '***D59 is road, it's road.***' (line 17). Eventually Pam looks at the ASR Captions too. Incorrectly captioned proper

nouns such as ‘A259’ (or ‘Crematorium’) can be generally problematic because they may not be predictable from the sequential context, potentially requiring extended strategies than might be the case with other classes of word (Bloch & Saldert, 2020), as happens here.

Carol initiates another category-specific interrogative OISR, this time claiming awareness that Pam may be talking about a road: **‘the name you telling me the name of the road’** (line 19) and Pam responds **‘yeah a number A A259’** (line 22). Carol then uses a ‘stronger’ format of OIR (Schegloff & Jefferson, 1977) by repeating the trouble-source turn **‘80 5 9’** (line 22) which shows she claims the capacity to reproduce it, but not to understand what is meant by it (Robinson & Kevoe-Feldman, 2010). Pam makes a third attempt at repair at line 26 **‘yeah A A 259’** and once again the ASR captions are incorrect ‘***888-8259***’ (line 26) - though Carol does not look at them at this point.

Pam then reaches for her whiteboard and pen and handwrites ‘A259’. As author, she self-initiates by reaching for the whiteboard and pen and bringing pen to paper and handwrites ‘A259’. The orientation of gaze and body towards the pad has functional parallels with other pre-beginning behaviours (Schegloff, 1996), such as inbreaths that can mark the beginning of a verbal speech turn. Handwriting-in-interaction may be characterized by three-part sequences, where the authored handwritten contribution (Part 1) is read aloud during the next turn (Part 2) and confirmed in the following turn (Part 3) (Bloch & Clarke, 2013). This sequence is present here at lines 28-30: Pam writes ‘A259’ (Part 1), Carol reads it aloud **‘oh A2! It's that it's the A2 I didn’t get yeah A259…‘** (Part 2) and Pam appears to confirm at line 30 **‘mmm’** (Part 3).

Pam and Carol appear to have no difficulty switching from speech to handwriting (and back to speech), supporting Bloch and Clarke’s hypothesis that and if an individual has acceptable

premorbid handwriting skills that are unaffected by MND, then handwriting may reasonably contribute to multimodal interaction and potentially a familiar modality with clear transition boundaries (Bloch & Clarke, 2013).

As at lines 8-9, Carol appears to be allocating responsibility for the trouble source to herself and away from Pam's dysarthric speech at line 29 (**'I didn't get that'**). Carol displays a stance of self-responsibility in a similar way to lines 8-9 where she allocates trouble to her own understanding of where a crematorium fits into the story, and not to Pam's dysarthric speech production. However at line 29, Carol also appears to allocate responsibility of OIR to multiple owners adding reference to 'other' - the ASR Captions **'and neither did that'** (line 29) in addition to herself. Responsibility of OIR can be allocated to claim ownership to some combination of 'self' or 'other' (Robinson, 2006).

5.3.11. Horatio

Relate again becomes the centre of attention in this example, and illustrates how trouble with a proper noun can take extended time to repair, in addition to showing how the ASR captions can be inaccurate with proper nouns.

Bob is talking to Adrian about some meetings he is arranging, and how he wants to **'reach out to people'** (line 1) and then **'get more meetings'** (line 5).

Bob says he **'wrote to to to Horatio yesterday'** (line 9) (Horatio is a work colleague), however Adrian identifies trouble at line 9 evidenced by a delay in response by 1.1 seconds and saying **'just a sec'**. Adrian looks at Relate but the captioning is incorrect: '***and info and spoke to two to radio yesterday***' (line 10). Bob attempts repair on the word he appears to think is the cause of the trouble at line 12 **'oh <u>Horatio</u>'**, adding stress. The ASR captions are incorrect again at line 13 ***'Working you'***. The trouble appears unresolved as Adrian does not respond and looks at Bob. Bob then makes a second attempt at repair **'oh oh oh oh oh <u>Horatio</u>'** (line 14), again adding stress to Horatio. The ASR captions are again incorrect '***Oh Oh Oh I wish you***' (line 15). Adrian looks at the ASR captions and back at Bob without response and Bob makes a third repair attempt **'oh oh hmm hmm ho <u>Horatio</u>'** (line 16) which again is incorrectly captioned '***Ho***' (line 17) and again Adrian does not respond.

Finally Bob appears successful his fourth attempt **'<u>Horatio</u>'** (line 18), despite the ASR Captions again being incorrect ***'I wish you'*** (line 19). At line 20 Adrian seems to confirm repair **'oh yeah yeah Horatio right yeah yeah it were it were good'** (line 20). However it is unclear how successful the repair actually was because, although Adrian confirms recognition of the name Horatio he does not necessarily confirm understanding of Bob's point (that he wrote to Horatio

for a meeting) as Adrian says ‘ it were good’ – it is unclear what Adrian is referring to that was ‘good’.

Extract 12: Horatio (the link to the transcript master is here)

```
00:00  1 Bob:    and then (1.3) I I want to .h  to reach out (0.6) to to people
       2 Relate: and then I want to reach out to people
       3 Adrian  yeah
       4 Relate:
       5 Bob:    and getting .h and get more more [meetings                                ]
                                                  [((Bob looks at Relate and back to Adrian))]
       6 Relate: and get going and get more more matics
       7 Adrian  hmm
       8 Relate:
       9 Bob:    um (0.7) I I wrote .h I wrote .h to .h to .h to Horatio yesterday
      10 Relate: and info and spoke to two to radio yesterday
      11 Adrian  (1.1) °just a sec°
                 ((Adrian looks at Relate))
      12 Bob:    (0.6) ho .h Horatio
      13 Relate: Working you
                 ((Adrian looks back at Bob))
      14 Bob:    (1.8) ho .h ho .h ho .h ho .h um .h ho Horatio
      15 Relate: Oh Oh Oh I wish you
                 ((Adrian down at Relate and back to Bob))
      16 Bob:    (1.2) ho .h ho .h hmm .h hmm ho Horatio
      17 Relate: Ho
      18 Bob:    (1.5)[Horatio                                  ]
                      [((Bob looks at Relate and back to Adrian))]
      19 Relate: I wish you
0:41  20 Adrian  >oh yeah yeah Horatio right yeah .h yeah yeah it were it were good<
```

At line 11, Adrian says **‘just a sec’** that indicates trouble but implying his own personal responsibility for some difficulty, processing the information provided, hearing what was said or understanding the words Bob said. Adrian’s stance appears to be that the responsibility for trouble belongs to himself. Bob takes Adrian’s ‘just a sec’ as a cue for self-repair, demonstrating that the use of OIR in actual, situated use can also manage trouble responsibility as a function of supporting the management of relationships, and preserving trouble-source speakers’ positive face (Brown & Levinson, 1987).

Adrian's 'just a sec' could be construed as an open-class OIR, communicating that there is repair related trouble but does not communicate the precise nature or location of the trouble (Schegloff & Jefferson, 1977). Bob appears to take the OIR as a cue for self-repair of one word: 'Horatio', even though the ASR captions on line 10 (which Adrian looked at) were incorrect for six out of the nine words Bob spoke on line 9. In response to an open-class repair initiation, the speaker of the trouble-source turn often produces a repeat or near-repeat of the trouble-source turn (Kitzinger, 2012), but in this case Bob appears to hypothesize that the trouble is to do with the proper noun.

Unlike in the 'Stereo' extract, where Bob was able to utilize gesture and alternative words to describe what 'stereo' is to support repair, with a proper noun like Horatio his options for gesture or alternative words may be more limited. Proper nouns may not be easily predictable in the sequential context of an interaction to the listener, and likely take longer to repair than other word classes (Bloch & Saldert, 2020). At the same time the ASR captioning of Horatio is totally inaccurate (as it is for many other words Bob says), again showing particularly difficulty with accurate captioning of proper nouns.

Bob over-articulates his speech at line 12, 14, 16 and 18 when saying 'Horatio' compared to line 9 (Curl, 2005; Rutter, 2009). For speakers with limited motor speech abilities, focusing of motor effort on making one or two words more intelligible may well be easier than producing a whole sentence (Bloch & Wilkinson, 2011) Although Adrian acknowledged repair at line 20, the 'cost' of such an extended repair sequence may have been loss of the thread of the conversation – which was about setting up a meeting – as Adrian says **'it were good'** (line 20) – which appears to not clearly link back to Bob's point. Bloch and Wilkinson highlight the risk that prolonged repair sequences could affect the ability of the recipient to understand aspects

of the AAC-mediated turn, or to understand the AAC user's subsequent attempts at clarification of aspects of that turn, or indeed by implication the thread of the interaction (Bloch & Wilkinson, 2013a).

5.3.12. Extended repair when the captions perceived as accurate, but not

In contrast to the previous section, these extracts show how repair may be extended when one or other of the dyad may perceive the ASR captions as accurate in some way, but are not. In 'Movie Day' Carol assumes accuracy of the ASR Captions and by doing so generates additional trouble, requiring extended repair directly with Pam. In 'Vegetarian Shepherds', even though the words are mostly transcribed correctly, it is the punctuation that the ASR Captions dynamically adds that results in reduced understandability, becoming a trouble source in itself.

Finally, In 'Mussolini', Tony appears to allocate the meaning of the incorrect captions as a reflection of his own contribution to the interaction, though the caption was generated from Elsbeth's speech.

5.3.13. Movie Day

This extract shows that an incorrect ASR caption can create multiple troubles, however if the ASR caption error sounds or looks similar to the spoken word it may be enough to support repair.

Pam is showing Carol a handmade gift she made for her friend Maxine, who is moving house. Carol identifies trouble at line 4, evidenced by a 1.2 second pause where she looks at the ASR

captions and repeats part of the incorrect caption as a candidate offer repair '**because it's a movie day?**'. This seems to become an additional source of trouble as Pam does not immediately respond with confirmation of repair.

Carol makes a second attempt to OISR at line 6, this time a restricted format of repair initiation, **'Why are you taking it to Maxine?'** and without referencing the ASR captions. But before Pam has the opportunity to respond, Carol almost immediately follows with another OIR candidate offer repair **'Moving!'** also on line 6.

Pam confirms the accuracy of this candidate offer repair on line 8 by stating that **'they're moving in Jedd Road they left Spain'**. Carol confirms her understanding on line 10 **'yeah I remember she left Spain'** and the action proceeds.

Extract 8: Movie Day (the link to the transcript master is here)

```
Time     Line  Speaker    Talk
00:00                     ((Pam shows Carol a gift she has for Maxine))
          1    Pam:       and that's for Maxine (·h) (0.7) 'cos they (.) Jim's moving there
          2               [.h and Di is away with them                                  ][(indistinct)            ]
                          [((Carol looks at Pam, frowns and then looks at Relate screen))]
          3    Relate:    I like that because it's a movie day and a house. And where was he?
          4    Carol:     (1.2)                                                          [Because it's a movie day?]
          5    Relate:                                                                    Because it's a movie day.
          6    Carol:     (0.9) [why are you taking it to Maxine? ]    [>moving!<       ]
                                  [((Pam lightly touches Pam's arm))]    [((lifts finger))]
          7    Relate:    Why are you taking it to Maxine moving?
          8    Pam:       They're moving in Jedd road (0.3) she left Spain
          9    Relate:    Just rude. And she left Spain?
         10    Carol:     yeah I remember she left Spain
00:22    11    Relate:    Yeah she left
```

This example demonstrates trouble generated by Carol's difficulty understanding Pam's speech-produced turn at line 1 was compounded by inaccurate ASR captioning at line 3, becoming an additional trouble source, creating "cascading troubles" (G. Lerner & Kitzinger, 2012) and prolonging the repair sequence.

Carol initially attempts an OISR by repeating what she believes is the trouble-source turn (at line 4 **'because it's a movie day?'**) – which claims the capacity to reproduce it, but may not understand what is meant by it (Robinson & Kevoe-Feldman, 2010). By doing so Carol, appears to accept that the ASR captions may authentically representing Pam's speech, however by repeating the phrase as a question could indicate that the words are intelligible to Carol they may not be understandable in the context. Even when turns produced using AAC are treated by recipients as intelligible (the words produced by the AAC device are identifiable) they may not be understandable in the context, why they are present at this point and what action they represent (Bloch & Wilkinson, 2004). When integrating different modalities, recipients of graphic utterances regularly produce a verbal formulation of what has been written revealing a recipient's understanding of what has been written, which may or may not match the

intended meaning (Bloch & Clarke, 2013). Although this finding may not necessarily apply for the graphic utterances generated by ASR Captions, the fact Carol repeated part of the caption provided Pam with the opportunity to confirm repair or otherwise.

With no immediate confirmation of repair from Pam, Carol reverts to a 'weaker' format of OIR at line 5 **'why are you taking it to Maxine?'**, a category-specific interrogative which claim a grasp of the trouble-source as referring to an action. In this way Carol appears to be reverting to her own understanding of what Pam said, and where the trouble-source may be located. Carol is implying some doubt on the ASR caption by no longer referencing them by referring to what Pam said on line 1, and also may explain why she moved from a stronger OIR format to a weaker one (Kitzinger, 2012).

However, the ASR captioning at line 3, inaccurate as it is, appears to support repair with a phonemically similar caption of 'movie' vs 'moving', which might be argued as a non-verbal 'near-repeat' of the trouble-turn (Schegloff, 2004b). Carol then generated a candidate understanding OIR **'moving!'** at line 6, and Pam appeared to confirm repair at line 7 by adding detail to where they are moving from and to. The repair solution is usually fitted to the format of the repair initiation, namely trouble-source speakers often respond to candidate understandings by providing clarification, as in this case (Kitzinger, 2012).

5.3.14. Vegetarian Shepherds

This extract shows how the punctuation that the ASR Captions dynamically adds can be a trouble source in itself, reducing understandability, even when the words are transcribed correctly. Later in the extract it was also observed how the ASR captions can generate potentially disatracting messaging when captioning incorrectly.

In this extract Pam is telling Carol about a meal she had with friends who are vegetarian. Pam is making the point that shepherd's pies are not vegetarian.

Carol identifies trouble at line 4, evidenced by a 1.8 second pause where she looks at the ASR

captions. Carol smiles and closes her eyes, then laughs loudly at the caption – probably '***Vegetarian Shepherds aren't edible***', which are the words Pam actually said but in separate phrases e.g. **'I mean when you're vegetarian, shepherds aren't edible'** (lines 1-2). The ASR captions additionally capitalizes '***Vegetarian Shepherds***' (line 3), potentially adding further ambiguity to the meaning. The captioned words are intelligible, but the added punctuation and capitalization reduces understandability by increasing ambiguity of meaning.

Carol laughs after reading the ASR captions but does not attempt repair, possibly because she understands the reference is about eating food, not shepherds. Pam laughs too but doesn't look at the Relate captions and may not have identified the trouble source or the reason for Carol's laughter.

After line 4, Carol does not look at the ASR captions. At lines 7 and 11 there is further inaccurate and potentially distracting captioning: '***when you fart them….and pregnant paid***', and '***Paul was upset because he said he didn't need to fire you***'. However, neither Pam nor Carol looked at Relate at this point.

Extract 9: Vegetarian Shepherds (the link to the transcript master is here)

```
Time   Line Speaker  Talk
00:00   1   Pam:     shepherds aren't edible ·h I mean when you're vegetarian ·h
        2           shepherd's aren't [edible                                    ]
                                      [((Carol looks at Relate, smiles, closes eyes))]
        3   Relate: veggie chairs are in the edible. I'm the way you are. Vegetarian Shepherds aren't edible.
        4   Carol:  (1.8)          [HEH HEH HEH °yes°]
        5   Pam:          when you [veg heh heh      ] enjoyed there and had fresh broccoli .h fresh carrots .h
        6           and Bragan paid for the [whole ·h] on my meal
        7   Carol:                          [↓°oh°   ]
        8   Relate: when you fart them and the fresh broccoli, fresh carrots and pregnant paid for there on my mill.
        9   Pam:    (0.6) [we       ] said Paul's upset ·h cos she said he ·h she didn't need to buy it
       10   Carol:        [very nice]
00:37  11   Relate: And we said, Paul was upset because he said he didn't need to fire you
```

By adding capitalization and punctuation to line 3 reduced understandability by increasing ambiguity of meaning. The words were intelligible and identifiable, but they may not be understandable in the context, why they are present at this point and what action they represent (Bloch & Wilkinson, 2004). It appears Carol may have initially taken the meaning of the ASR captions one way – and then realized Pam's talk is about eating food, not shepherds.

Carol may have found humour in the captioning errors in comparison to Pam's actual meaning which caused her to laugh. Carol does not attempt repair and may resonate with the pattern of delayed other-initiation of repair in non-serious talk of people with dysarthria (Bloch & Wilkinson, 2013b), but the difference here is that Carol is responding to what she reads in Relate – the ASR captions are the source of the non-serious message, not Pam.

Participants strive to adjust their visible emotional state in accordance with the demands (or the ongoing definition) of the situation (Goffman, 1972) . A momentary collapse of the role appropriate for the situation where individuals allow themselves, for example to burst into

laughter (Goffman calls this 'flooding out'), such as Carol's laughter on line 4 can be overlooked by other participants; they can be contagious; or they can result in changing the frame (Goffman, 1972). Pam subsequently reciprocates Carol's facial expression of smiling and generation of laughter particles, consistent with the pattern where a sequence of laughing alone can precede laughing together (Sacks, 1995), but it is unclear if Pam understands why Carol has laughed. Nevertheless, Carol's burst of laughter at line 4 appears not to have affected Pam continuing to discuss the meal she had with her friends at line 5.

The effect of Pam's laughter particles on line 5 probably affected the captioning accuracy on line 8 as Pam's speech is likely to vary and be harder to recognise when she is laughing at the same time.

In line 6, Pam says the proper noun 'Bragan' which the Relate captions as 'pregnant' (line 7) demonstrating again the difficulty Relate can have recognizing proper nouns. However, these captioning errors did not interfere with the interaction as neither Carol or Pam looked at the ASR Captions after line 4.

5.3.15. Mussolini

In this extract, Tony appears to allocate the meaning of the incorrect captions as a reflection of his own contribution to the interaction, though the caption was generated from Elsbeth's speech.

Elsbeth is talking to Tony about Mussolini. The ASR Captions at line 2 accurately reflect what Elsbeth says on line 1 ('***so Mussolini***'), and it seems Tony understood what Elsbeth said at line 3 because he points at the ASR captions and appears to confirm they are correct **'Ah! Got it'** (line 3). Both Elsbeth and Tony look at the ASR Captions.

Elsbeth then says **'right and'** on line 4, which is incorrectly captioned as '***crap and***' on line 5. At line 6 there is a 0.8 second pause and then both Tony and Elsbeth laugh loudly in response to the incorrect ASR caption. Tony then appears to allocate the meaning of the incorrect caption to himself when he says **'everything I said was crap it was translated everything I said it said every word was crap'** (line 8).

Extract 10: Mussolini (the link to the transcript master is here)

Time	Line	Speaker	Talk
00:00	1	Elspeth:	So Mussolini
	2	Relate:	***so mussolini***
			[((Tony looks at Relate and points at Relate))]
	3	Tony:	[(1.2) Ah! Got it]
			[((Elspeth looks at Relate))]
	4	Elspeth:	[↓°right° .h and]
	5	Relate:	***crap and***
	6	Tony:	(0.8) [HEH HEH HEH]
	7	Elspeth:	[HEH HEH HEH]
	8	Tony:	[everything I said was crap HEH HEH it was translated everything I said it said every word was crap]
	9	Elspeth:	[HEH HEH HEH]
00:19	10	Relate:	***and oh yeh, that's it. It's a word to be crap.***

In this extract, similar to 'Vegetarian Shepherds', Relate is again the trouble source of what seems like a non-serious message. Both Tony and Elsbeth 'flood out' (Goffman, 1972) into laughter as a result of the ASR caption's '***crap and***' at line 5, and as a result the focus of the interaction changes from Mussolini to the ASR Caption content.

The 0.8 second delay (line 6) prior to laughter by both Tony and Elsbeth may be an indication of 'doing being surprised' by the incorrect captions. Preceding surprise tokens, silence may not necessarily mark dispreference and may imply that the recipient is left without words because of surprise (Wilkinson & Kitzinger, 2006).

Tony appears to then continue the non-serious humour, as if the ASR captions had the sentience to comment on the content of his speech directly at line 8 - **'everything I said was crap it was translated everything I said it said every word was crap'**. However, Tony may have known it was simply a captioning error.

5.3.16. Repair of the ASR captions becomes the focus of interaction

The final two extracts show how the effort of repairing incorrect ASR captions can become the focus of the interaction, even if no trouble is indicated by the communication partner. In 'Actually Helpful' Bob makes multiple attempts to correct the ASR captions, apparently regardless of whether Adrian has understood or not, and repair is only resolved when the ASR Captions are accurate. In 'That wasn't bad actually', initiation of a change of topic in combination with inaccurate captions results in Bob focusing on the ASR Captions to attempt repair.

5.3.17. Actually helpful

This extract shows how attempts at OIR can be directed at Relate, and the focus of the interaction - seemingly regardless of whether the communication partner has identified trouble or not.

Bob is telling Adrian that he feels his speech is declining **'I tend to find my speech is getting worse'** (line 1), though when using the ASR captions they can help him be better understood **'but also I thought that when I use this actually it is actually helpful'** (line 3). Adrian does not indicate any trouble understanding, but Bob nevertheless makes multiple repair attempts of Relate at lines 7 (**'help <u>helpful</u>'**), 9 (**'help <u>helpful</u>'**) and 11(**'he <u>helpful</u>'**) because he notices the ASR captioning for 'helpful' is incorrect at lines 5 ('***couple***'), 8 ('***have am***') and 10 ('***am full***'). The trouble only resolves when Relate accurately captions '***helpful***' at line 12 (at the fourth attempt) and the conversation moves on.

Extract 11: Actually helpful (the link to the transcript master is here)

```
00:00  1 Bob:      I tend to find [this .h my speech is getting worse                ]
                                  [((Bob points at Relate and then points at himself))]
       2 Relate:   and two signs next nit my speech education words
       3 Adrian:   °hmm°
       4 Bob:      but [also .h um I thought that .h      ] when when [I use this .h actually .h it is actually helpful]
                       [((Bob looks and points at Relate))]          [(( Bob looks at Relate))                        ]
       5 Relate:   Also I though that when whenever I use this as a very has actually couple
       6 Adrian:   (0.8) °hmm°
                   ((Bob looks at Relate and leans towards it, Adrian looks at Relate))
       7 Bob:      (1.4) help helpful
       8 Relate:   Have am
       9 Bob:      (1.5) help helpful
      10 Relate:   am full
                   ((Bob moves closer to Relate))
      11 Bob:      (2.5) he helpful
      12 Relate:   helpful
      13 Bob:      ↓Ah
                   ((Bob slaps his hands on the table, smiles at Adrian, Adrian points at Relate and smiles and looks at Bob))
00:28 14 Ardian:   °yeah°
```

The three repair attempts of 'helpful' at lines 7, 9 and 11 are potentially examples of "recycling" repairs – where a speaker says again a stretch of talk that is less than a full TCU and usually just previously said (Schegloff, 2011). Getting it said "in the clear" is designed to deal with whatever trouble in hearing or understanding accompanied its involvement in hearing or understanding by the 'other' (Button & Lee, 1987).

In this case the 'other' appears to be the incorrect ASR captions, because there is no indication that Adrian has trouble with what Bob is saying in lines 1 and 4, and Adrian says **'hmm'** (a continuer) at lines 3 and 5 indicating alignment that as he, as the recipient of the Bob's story, (in this case Bob talking about his speech getting worse and how the ASR captions may help) supports the structural asymmetry of the activity by displaying an understanding that the story

teller has prioritized access to the floor (Stivers, 2008). A speaker who is aligned as recipient of an ongoing storytelling usually enacts this alignment by producing talk that is hearably relevant at the possible end of a unit of the ongoing story, and does not launch or participate in a competing action (Mandelbaum, 2012).

The ASR captions are incorrect at lines 2 and 5, but Adrian does not look at them until Bob does after line 6. It is Bob, not Adrian who identifies trouble at line 7 and enacts a OIR directly to the ASR, and again at lines 9 and 11. Only when Relate captions '***helpful***' correctly at line 12 does Bob acknowledge repair by saying **'ah'** at line 13 and slapping his hands on the table - evidence that Bob's OIR appears to be specifically produced to expose an error by Relate.

By redoing the same action in the same way, repeating the trouble source as Bob did on lines 7, 9 and 11 may indicate the action was likely both understandable and 'appropriate' in its original form (Robinson, 2006), after all Adrian shows no indication of trouble at this point.

Bob may be allocating the trouble type to that of the ASR incorrectly 'hearing' what Bob has said because Bob's repeat becomes more enunciated which is an additional practice for repairing hearing trouble, and after line 9, moves closer to produce a relatively stronger signal, which like enunciation, is an additional practice for repairing hearing trouble (Robinson, 2006).

It is interesting that Bob only highlights the incorrect caption of 'helpful' (line 5) for repair via repetition, even though seven words are incorrectly captioned in this phrase ('***Also I though that when whenever I use this as a very has actually couple***' line 5). It may be that Bob considers 'helpful' to be the key word to be captioned correctly, because his point to Adrian is that Relate is 'actually helpful' (line 3) as his speech is 'getting worse' (line 1), or perhaps he considers the other captioning errors to not materially changing the meaning of what he says,

or perhaps he considers Adrian to have understood his spoken words and this is a 'demonstration' of the helpfulness of the ASR captions, that has gone wrong and needs fixing.

Whatever the reason, Bob has moved away from discussing with Adrian the 'helpfulness' of Relate, to enacting multiple repair attempts of the incorrect ASR captions becoming the focus of the interaction.

5.3.18. That wasn't bad actually

In this final extract the accuracy of the ASR captions becomes the centre of attention and Bob looks at the captions for almost the entire time.

Bob is telling Adrian that the ASR captions are improving: **'I I think getting it I think getting it more that that you don't don't understand'** (line 1). The captions retain the meaning of Bob's words while removing sequential whole word repetitions (which Relate is designed to do): ***'It is getting it. It's getting more that you don't understand.'*** (line 2). Bob points at the ASR Captions and says **'that wasn't bad actually'** (line 5), almost sounding surprised and Adrian agrees at line 7: **'yeah ha ha yeah'**.

There is a subsequent gap of 1.1 seconds where both Bob and Adrian are looking at the captions and Bob then laughs loudly at line 9, presumably in recognition that the ASR captions have again accurately captioned his words from line 5: ***'that wasn't bad actually.'*** (line 6).

Following a 1.7 second gap, Bob changes the subject from commenting on the accuracy of the ASR captions to how they should be available as captions in Zoom (an app for online conferencing): **'Good but yes maybe if if if this was in incorporated in into Zoom it will be**

better' (line 11). However, this time the ASR captions are mostly inaccurate: ***'Yes, bit. Yes made me if it's if the food a in garbages it into children, it would be better.'*** (line 12), although the final four words of Bob's words on line 11 are sequentially accurate ('it would be better'). Adrian identifies trouble evidenced by a delay of 1.7 seconds and enacts a OISR by repeating what he may have understood, with an apology, perhaps to assign responsibility to his hearing or attention **'it would sorry'** (line 14).

Bob attempts repair on only part of the phrase: **'in incorporated into Zoom'** (line 15). Although the ASR captions are again incorrect (***'It copied into German'*** – line 16), Adrian appears to indicate successful repair at line 17: **'oh yeah yeah so when um you would have to type it out on zoom it wouldn't automatically play'**.

However, Adrian may not have understood Bob's point at line 11 and repeated at line 15 that the ASR captions would be better **'if this was in incorporated in into Zoom'** (line 11) and **'in incorporated into Zoom'** (line 15) because he suggests Bob would have to **'type it out on zoom it wouldn't automatically play'** (line 17) – making a different point that the ASR captions are not incorporated into Zoom. Bob's point is about incorporating the ASR captions – presumably to not require manual typing.

Extract 13: That wasn't bad actually (the link to the transcript master is here)

```
00:00  1  Bob:      [I I think getting it (0.8) I think getting it more ·h that that ·h you ·h don't don't understand]
                    [((Bob is looking at Relate))                                                                    ]
       2  Relate:   It is getting it. It's getting more that you don't understand.
       3            [(4.1)                                                                                  ]
                    [((Bob leans in to look at Relate closer and Adrian looks too, Bob points at Relate))]
       4  Adrian:   °ohh good°
       5  Bob:      That wasn't bad actually
       6  Relate:   That wasn't bad actually.
       7  Adrian:   yeah ha ha °yeah°
       8            (1.1)
       9  Bob:      [HA HA HA                                                                  ]
                    [((Bob points at Relate and Adrian moves his arm slightly as if to point too))]
      10            (1.7)
      11  Bob:      [Good (1.4) but (1.4) yes (1.2) maybe (1.2) if if if this was .h in incorporated ·h in into Zoom .h it will be better]
                    [((Bob is looking at Relate))                                                                                      ]
      12  Relate:   Yes, bit. Yes made me if it's if the food a in garbages it into children, it would be better.
      13            [(1.7)                                                                                  ]
                    [((Bob leans in to look at Relate closer and Adrian looks too, Bob points at Relate))]
      14  Adrian:   [°it would sorry°       ]
                    [((Adrian looks at Bob))]
      15  Bob:      [in incorporated .h into .h Zoom]
                    [((Bob is looking at Relate))   ]
      16  Relate:   It copied it into German.
      17  Adrian    oh yeah yeah so when (0.8) um (0.8) you would have to type it out on zoom it wouldn't automatically play
      18  Relate:   Yeah.Yeah. So when you would have quite outdoor we would other
00:53  19  Bob:      hmm
```

The laughter that Adrian expressed in line 7 and Bob in line 9 may have been indicative of a surprised response that that ASR captions were mostly correct at line 2 (**'It is getting it. It's getting more that you don't understand.'**) and line 6 (**'that wasn't bad actually.'**), and that they reflected the meaning of what Bob said on both occasions. A sudden shift in expectations and the resulting surprise can often provoke a feeling of humour and thus laughter (Glenn, 2009). This surprised response might also indicate that both Bob and Adrian have low expectations of the accuracy of the ASR captions generally. Additionally, Adrian's pointing at Relate while smiling and gazing at Bob could be a way to demonstrate understanding of Bob's point with the praising of Relate as helpful, bringing the talk back to topic.

After the 1.7 second pause at line 10, Bob says '**Good** (1.4) **but** (1.4) **yes** (1.2)' (line 7) perhaps an interactional resource that signals he is about to move on from a chunk of the conversation that has been brought to a possible completion and nothing else has yet been launched

(Bolden, 2006), in this case moving between talking about the accuracy of Relate to how the ASR Captions should be incorporated into Zoom.

Adrian says **'It would sorry'** at line 14 and by doing so generates a category-specific interrogative which claims a grasp of the trouble-source (Kitzinger, 2012), that not only initiates repair but also may attribute the OIR requirement to his own personal trouble, perhaps in hearing or understanding. 'Sorry' is an apology that orients to Adrian's commission of, and accountability for, a possible offense (Robinson, 2004), and perhaps Adrian said 'sorry' in an attempt to preserve the Bob's positive face (Brown & Levinson, 1987).

In a similar way to Adrian's confirmation of repair in 'Horatio' (line 20), it is also unclear here how successful the repair actually was. Even though Adrian confirms recognition of the term 'Zoom' he does not seem to confirm understanding of Bob's point (that the ASR captions would be better **'if this was in incorporated in into Zoom'** (line 11). Adrian instead seemed to make a different point that (as the ASR captions are not incorporated) **'you would have to type it out on zoom it wouldn't automatically play'** (line 17).

Both are candidate understandings that are re-sayings of the trouble-source (in this case 'Zoom') and in both, Adrian adds a response to what he perceives as the point Bob made. Bob has the opportunity to confirm repair or otherwise at the next turn (Schegloff & Jefferson, 1977). Here Bob responds with **'hmm'** (line 19), that does not appear to clearly confirm repair but is hearably relevant at the possible end of a unit of the ongoing story, and does not launch or participate in a competing action (Mandelbaum, 2012).

It may be that the time and effort Bob perceives that he has already taken to attempt repair, possibly combined with a perception of effort of repair from prior troubles can influence a decision to attempt further repair, or that it is 'not worth the effort' (Clarke et al., 2019).

5.4. Thematic Analysis

In Thematic Analysis (TA), a theme should capture something important about the data in relation to the research question, representing some level of patterned response or meaning within the data set (Braun & Clarke, 2006a) .

Following initial coding, five themes were identified: speech and identity, strategies for communication, hopes for personalised ASR, set-up and first use, and overall experiences of using ASR in practice.

All three plwMND felt the way they talk and what they say was linked to a sense of who they are and their different roles in life. To be able to keep talking and be understood by friends, family and work colleagues as speech declines was seen as very important, and a way of maintaining important social and work networks. Communication partners supported plwMND trying Relate, believing it may help maintain the confidence of plwMND to participate in activities and maintain networks.

For some, the application was not easy to understand and there was little support available beyond a short training video and family help. Some SLTs appeared to be aware of the app but were reported to did not have the skills to help with practical support. The effort required to record 500 phrases in the face of frequent fatigue (a common symptom of MND) without any indication of the likely usefulness in daily conversation was seen as a barrier.

None found the process of recording custom cards intuitive, despite appearing to be the best way for Relate better recognise proper nouns such as the names of people, places and things. One plwMND only found out about custom cards after recording the 'standard' phrases.

Because of the high perceived level of caption inaccuracy, plwMND reported they may only use Relate in specific contexts and with particular people, if at all. Communication partners voiced concern at looking away from the plwMND in mid-conversation to the ASR Captions – risking a perception that they were not listening , though this view was not shared by plwMND themselves.

For all participants the hoped-for benefits personalised ASR were not realized. The captions were seen as too inaccurate in practice and in the end some plwMND and conversation partners preferred to resolve any misunderstandings between themselves and not use the ASR captions.

One plwMND seemed to attribute responsibility for the ASR captioning errors to her own declining speech, without consideration of the technology shortcomings, risking potentially misplaced negative impacts their own sense of personal identity.

5.4.1. Theme 1 - Speech and identity

"MND has affected my voice. It hasn't affected my body. But it has affected everything else" [Pam] For both Pam and Elsbeth, expressing themselves through talk is an integral part of their sense of identity. ***"I always talked a lot it's who I am. I'm a talker I can talk for England but I'm hard to understand now"*** [Elsbeth]. For Bob, voice change and being less easily understood have led to reduced roles in life: **"I used to lead the training but I have handed that over now"** [Bob].

Communication partners also recognised the effects of voice change: ***"The less well you know someone the more embarrassed you can be, but we're past that stage now"*** [Carol]. ***I see some others find Bob harder to understand and they talk to him less*** [Adrian].

5.4.2. Theme 2 - Strategies for communication

Prior to using Relate, Pam, Elsbeth and Bob all used AAC to support communication when others could not understand what they said: **"I type the word on my phone if someone doesn't understand what I say"** [Bob]. ***"I write the words down when someone doesn't understand me"*** [Elsbeth]. ***"I use the whiteboard if I have to be understood. Paul says write the context word first on the whiteboard so I do that"*** [Pam].

Pam plans ahead when she needs to ask a question: ***"I have a note to show people for example can you tell me where the ladies long socks are"*** [Pam]. Or they get the help of the communication partner to 'translate': '***If someone wants to asks me a question they say 'where's Carol?' to help translate they don't understand me without her helping'*** [Pam].

Sometimes the plwMND use strategies to avoid talking altogether: ***"When we walk as a group walk ahead so no one has to talk to me"*** [Pam]. ***"I don't talk to my neighbours much now I don't want that look of sympathy"*** [Elsbeth].

Some situations are more difficult to avoid using talk. Elsbeth was completing a PhD and negotiating alternative arrangements for presenting to the upgrade panel: ***"They won't understand me so I asked to submit written responses in the upgrade but I am still waiting I have not heard back. It is very worrying"*** [Elsbeth]

When introduced to Relate all participants were hopeful it would help them to be better understood when talking: ***"If they see my words on the screen they will understand me"*** [Pam]. The idea that technology could help a person be better understood as they talk was appealing: ***"The app is amazing I want to talk just like normal and I want to be understood like normal"*** [Bob]. Elsbeth is hopeful that it will help her husband: ***"You know I told you about this app well it's going to help you understand me better"*** [Elsbeth]. Elsbeth also hoped it would help in her upgrade if the examiners insisted on a verbal interview: ***"I hope it won't happen but if I have to talk I will try to use it"*** [Elsbeth].

5.4.3. Theme 3- Hopes for personalised ASR

Communication partners were on the whole positive about the benefits Relate could bring: Tony is concerned that Elsbeth is withdrawing from her social circle ***"she was always arranging things maybe this will help bring her confidence back"*** [Tony], ***"This will help Pam - she is so worried about not being understood"*** [Carol]. Pam was hopeful that people won't finish her sentences for her when using Relate unlike when she was using the whiteboard: ***"Relate is***

quick and no one will wait or finish my sentences like when I'm on the whiteboard" [Pam]. However, Tony pointed out that there could be other reasons for not being understood: ***"You always tell me it's because I don't listen to you - is this going to make me listen better?"*** [Tony].

5.4.4. Theme 4 - Set-up and first use

Two of the plwMND needed some support to set up Relate and record the phrases in order to create a personalised model: ***"I didn't know how to do it I usually leave this sort of thing to Paul"*** [Pam]. There wasn't an obvious way to get support: ***"my speech therapist didn't know anything about it and there isn't an instruction manual"*** [Elsbeth]. Pam asked the researcher to visit to help her and her husband better understand how Relate works ***"Thanks I don't think I could have done it without you"*** [Pam]

Pam felt the number of phrases needed to train Relate was a big commitment: ***"five hundred phrases five hundred it's a lot of work and I get tired"*** [Pam]. Elsbeth felt the same: **"I had to take breaks because my voice got worse as I became tired"** [Elsbeth]. Bob found recording so many phrases dull "***I left it for a while as there were more important things to do it was only when you chased me I did anything at all"*** [Bob].

There was no indication of potential usefulness during the initial recording stage ***"It's boring to do and I don't really know what it will do for me***" [Pam].

There was no feedback if a mistake was made, making the task even longer. Pam found out she had made more than a hundred poor quality or empty recordings before she realized. Relate did not provide any indication on the quality of recordings ***"I thought everything was ok until***

I listened back to some of the recordings – they were silent or Paul was talking at the same time…why can't it tell me when I do it wrong?" [Pam].

The benefit of recording custom cards was not apparent ***"I recorded all the phrases Relate told me to but it didn't say I should record my own words too"*** [Pam]

5.4.5. First use of Relate

It was quickly seen by plwMND and communication partners that the ASR captions were frequently inaccurate. However, Carol said Relate sometimes helped her to recognize the correct word even if Relate was not accurate: ***"The phonic helps - the word is often wrong but I get the sound of what Pam was saying and that helped me guess"*** [Carol]. Even when few words are transcribed correctly it could be enough to understand the gist of what has been said: ***"A word can give a context - the flow of the sentence is wrong but there are words that are right and there are words that give context and some that sound similar that helps me to jog the meaning in my head"*** [Carol]

The plwMND felt Relate as helpful only in specific conversational contexts: ***"It is useful for people who don't understand me not with familiar people"*** [Bob], and ***"I would not use it in serious conversations"*** [Elsbeth]. Location mattered too: ***"I don't used it when I'm out in a crowd with lots of talking"*** [Pam], ***"It doesn't work well in noisy environments…one on one is when I would use it in a quiet room"*** [Bob], ***"It was too noisy on holiday to use Relate"*** [Pam]

5.4.6. Design of Relate

Relate cannot differentiate between speakers and will transcribe every voice it hears as if it is one speaker. This can result in a lot of words on the screen for the reader to review: ***"There is a lot of text on the screen and very difficult for me to look at quickly… hard to find my way around"*** [Adrian]

There were suggestions for how to limit the captions to the plwMND or distinguish who said what words: ***"It would help if there is an automatic way of clearing everything quickly from the screen to keep it focused on what is being currently said"*** [Adrian]. Another idea was colour-coding the captions to easier differentiate who said what: "***Can Relate change the colour on the screen for different speakers"*** [Adrian].

Bob suggested using a microphone ***"I need a directional microphone so it picks up only my voice not others"*** [Bob], or to switch Relate on only when he is speaking: "***I prefer to stop and start listen mode so people can read my words and not get confused by other people being transcribed"*** [Bob]

Adrian felt the screen captions were not easy to read: **"I needed to ask to adjust the brightness because I could not see clearly - perhaps this can be done automatically when the app opens"** [Adrian]. **"The font should be bigger – I can't read the words quickly"** [Adrian].

Carol felt the way the captions roll off the screen made it sometimes difficult to look back to check her understanding as the words were no longer visible ***"I needed to check something earlier to confirm my understanding but it had rolled off the screen and I couldn't see it"*** [Carol].

Conversation partners were concerned about how plwMND may feel when they look at the Relate screen. Carol believed there would be a negative perception when breaking eye contact to look at Relate: **"looking away to check the screen feels like it could be rude when you are talking to me"** [Carol].

Adrian was aware he was dividing his attention between Bob and the Relate screen: **"In some ways it feels like you are having a conversation with different people… I want to look at you and not stare at this"** [Adrian].

Carol felt it was more 'normal' to ask for clarification rather than look at the screen: **"It is just human nature to have the eye contact and say 'what did you say again?'…honestly I'm more inclined to say what did you say or go back a bit in a prompt, rather than look because it distracts from conversation because of the relationship of what conversation is"** [Carol].

Adrian felt the process of repair was delayed by him looking at Relate and then asking to Bob to repeat anyway: **"When I looked a couple of times I became conscious I looked at the screen - I would normally say sorry or ask to repeat something - because I looked at the screen it had been noticed - it feels a bit of a delay to then ask it"** [Adrian].

Indeed, Pam took it as a cue to repeat her words when Carol looked at Relate: ***"I feel I have got to repeat it again, go back and repeat the sentence"*** [Pam]

Overall, plwMND were more comfortable than their conversation partners when either looked at the screen mid-conversation:

"I don't mind if people look at the app but Adrian may not know that" [Bob], ***"I don't mind you looking down at Relate…it doesn't feel like a rejection of what I say"*** [Pam], ***"I saw him look at the screen but it didn't bother me"*** [Elsbeth].

Data privacy concerns were raised because Relate attempts to transcribe everyone it hears: ***"What I find off-putting is when people around me a proper conversation the words the other person's text to be there which I think like an invasion of privacy for them"*** [Bob].

Adrian questioned the need to keep so much transcription history at all: ***"You don't need all the history, especially off the screen but it is there. For data privacy concerns it is better to just show the last couple of sentences on the screen and then delete the rest - why have it there at all? It is not likely to be used in the conversation"*** [Adrian]

Some plwMND sometimes preferred other forms of AAC over Relate as they were more reliable: ***"When I use the whiteboard I write the words I want and people don't have any trouble understanding the word…I feel I'm in control when I write"*** [Pam], ***"I type the right words into my smartphone and show that"*** [Bob].

5.4.7. Theme 5 - Overall experience of using ASR in practice

Overall, using the ASR Captions in practice was a generally poor experience for both plwMND and communication partners: ***"It is accurate at the beginning and then after a while it goes crazy and is not accurate at all - it has a mind of its own"*** [Elsbeth]. Tony agreed ***"After a while I got a much better idea of what we are talking about directly with Elsbeth"*** [Tony], ***"It didn't help me at all really today. The words were wrong"*** [Pam]. **"Relate was amusing up to a point but some of the words I thought were easy and it should not have got it got them wrong"** [Elsbeth]

Carol felt Relate is not accurate enough ***"Relate does not help in a direct way - not a translation just a hint or clues"*** [Carol]. Adrian agreed: ***"Almost every time I looked the word wasn't***

accurate" [Adrian]. Tony felt it was not fit for purpose: ***"Out of five I score Relate a one"*** [Tony]. Elsbeth summarized her experience: ***"It just isn't accurate enough to be used"*** [Elsbeth].

For one plwMND, The negative experience of using Relate extended to a sense of personal failure: ***"This is my fault when I'm not understood by Relate"*** [Pam].

5.5. Research – Summary

Pam was recorded in conversation with Carol ten times over twelve months and the average length of the sessions was just under 14 minutes. Bob was recorded with Adrian eight times over ten months, averaging 10 minutes per session. Elsbeth and Tony were recorded six times in eight months, averaging slightly more than 10 minutes per session.

The FDA-2 dysarthria assessment showed that only Bob experienced a significant decline in his speech, and the other two remained stable throughout the study.

For all three plwMND, the actual average WER percentage for each session was almost always higher than Relate's own forecast WER threshold and between 50-75% of the Relate captions were judged to have lost the meaning. Custom Cards appeared to be beneficial when used with frequently used words.

Gaze patterns significantly when looking at Relate varied: Pam and Carol seldom looked at the captions, though Bob looked at the captions increasingly as his speech declined and Tony looked at the captions extensively as if to monitor their accuracy.

Using CA, repair of trouble was a frequently occurring feature across the dyad sessions extracts. And the vignettes illustrated multiple sources of trouble and repair, including difficulties

understanding speech with dysarthria, ASR caption inaccuracy, and dynamically added punctuation by the ASR captions introducing ambiguity of meaning.

Thematic Analysis identified five themes: speech and identity, strategies for communication, hopes for personalised ASR, set-up and first use, and overall experiences of using ASR in practice.

6. Discussion

6.1. Introduction to the chapter

This study is the first to examine how Relate is used in everyday conversation between three plwMND and their communication partners.

The goal of this research is to better understand how plwMND and significant communication partners use personalised ASR to support everyday spoken conversation over a period of up to 12 months.

This research studied the effect of ASR captioning from three perspectives:

- From the perspective of the ASR technology itself – how accurately it captions speech and maintains meaning, how it measures its own performance and sets user expectations, how easy and understandable it is to use.
- From the perspective of conversation analysis (CA) – describing how each dyad conducted social interactions and how troubles and repair were managed. CA does not explain the technology or features of the technology but of the social interaction when ASR captions are present.
- Directly from the perspective of plwMND and their communication partners through semi-structured interviews, using Thematic Analysis (TA) for qualitative analysis of the interview transcripts.

All three plwMND were hopeful that personalised ASR would help them communicate in daily life and were keen to try it. All three conversation partners were positive about the potential benefits of being able to better understand would support the plwMND maintain their relationship as well as social and work networks.

There were difficulties almost from the start when using Relate: it was not intuitive to use for people with little confidence around technology, the initial 500 recordings were tiring for some, boring and arduous for others, and with no indication of potential benefit. All three plwMND used 'Listen' even though there are other functions available which may have been helpful.

The app design implied that custom cards were an optional activity and therefore of little importance. This research indicated that Custom cards may indeed improve ASR recognition of proper nouns, but a much larger study is required to confirm if this is actually the case generally. That being said, the analysis showed that for some plwMND, for the words they recorded custom cards for were said very few times in the videoed conversations, demonstrating how difficult it may be to predict the content of future conversations. Guidelines on optimal use of custom cards is warranted.

The research demonstrated that the Relate app provided overly optimistic expectations of model accuracy based on forecast WER in comparison to the actual WER of the conversations recorded for this study. The percentage of captioned phrases with errors that lost the meaning of what was actually said by the plwMND was significant, at times more than 70% of all phrases in a session. There was no measure that included listener understanding, the key objective of Relate.

The research indicated that repair was a feature in the interactions between the dyads, with clearly occurring instances of other initiation of repair, similar to other research for people with dysarthria using AAC (Bloch, 2005). The ASR captions were reported to be sometimes helpful in repair even when inaccurate but phonemically close enough to prompt for the right word. It was observed how inaccurate captions can cause cascading troubles requiring extended repair (Bloch & Wilkinson, 2013a). At other times the captions themselves became the source of trouble even when the conversation partner gave little indication of trouble themselves. Sometimes inaccurate captions were a distraction and became the focus of the interaction.

One plwMND was seen to resort to another form of AAC (a white board) when trouble was not resolved despite multiple cycles of repair that Relate was unable to help resolve. However another plwMND resolutely focus on repairing the ASR Captions using speech, even when there was little indication of trouble from the communication partner.

The research showed a variation in how often the plwMND and communication partners looked at the Relate screen, potentially indicating different uses for the captions. Over the period of this study, Bob was the only plwMND that experienced a decline in speech as measured by FDA-2. Bob appeared to increasingly use Relate to check how accurate the captions were for his speech, perhaps to measure his speech change. He did this even when his communication partner appeared to understand what was said. Bob progressively looked at the captions more as the sessions progressed, average 9% of the time from sessions 1-4, increasingly so from session 5-8 to around 29%.

Thematic analysis indicated that plwMND and conversation partners were hopeful that personalised ASR technology would be helpful, but none felt it had met expectations. Many reported that the caption accuracy was just not accurate enough to be useful, although some said it may help in particular contexts, with particular people.

One plwMND in part seemed to blame their own speech for inaccurate ASR captions, risking negative effects on their own psychological wellbeing due to ASR performance.

6.2. ASR model performance

Actual ASR WER was higher than the forecasted WER for all plwMND and across almost every session. For Pam and Bob, Relate set the expectation that ASR caption accuracy would be in the 'Medium' range – that is between 15-35% WER. However, For Pam the actual average ASR WER per session varied between 38.2 – 53.2%, and for Bob between 33.6 – 42.5%, therefore all but one session had higher actual WER than the range of expectation set by Relate. For Elsbeth, her personalised ASR model recognition accuracy was forecast as 'High' (average WER less than 15%), however the actual average WER between 30.6 – 34%, significantly worse than the expectation of accuracy set by Relate.

The reason why actual WER was consistently worse than forecast may be because all three plwMND used it in conversation, whereas the Relate ASR forecasting process uses a sample of shorter generic phrases. ASR performance is likely to decline for longer utterances or spontaneous conversations (Green et al., 2021; Tobin & Tomanek, 2022). Additionally, the 'large language model' (LLM) that Relate is trained on may have few, if any, examples to learn

from for the names of people, places and things the three participants talk about in daily life, for example 'Wyvale' and 'Horatio'.

For Pam and Bob, the percentage of Relate captions judged to have lost the meaning of what they said was at best 50% in a session and at worst, more then 70%. For Elsbeth, at best, 60% of all the captioned phrases were not accurate enough to preserve the meaning of what she said. At worst it was nearly more than 75%.

However useful the WER or 'meaning lost' measures may be for assessing the accuracy of ASR captions, neither accurately measure for what Relate was built for: supporting listener understanding. This is because when ASR is captioning speech, listeners have more information to hand than the captions to understand meaning – the context, the spoken words themselves, tone of voice, pauses, mood, humour, geographical, social and educational background, health status, gender, as well as gesture – all in addition to the content of the words (Nathanson, 2017). Both WER and 'Meaning lost' measures would be more relevant for listener understanding if the ASR captions were the only modality a listener had for of understanding what was being said.

In practice 'forecast WER' is probably used by Relate because it can be calculated solely from the app, and listener understanding is not assessed because cannot be measured solely from the app – even if it does link directly to the overall purpose of Relate.

The multi-modality of communication may also be why less trouble was observed in interaction using CA than could be expected when frequently more than 70% of the transcribed phrases had lost the meaning to what was said. One possibility is that it could be that if an incorrectly transcribed word had similar phonemic properties to the actual word, it was sometimes enough

to capture the meaning of what was said (as demonstrated by Carol in 'Movie day'), or may be enough to capture the context if the ASR captioned a word correctly amongst incorrect words.

Another reason could be that as the ASR captions are not looked at by the listener for much of the time, they may have a less effect on the interaction than expected. For example, Carol only looked at the captions on average for 7% of the session time – about 60 seconds on average across the 10 sessions. Pam averaged only 8 seconds per session. It could be argued that caption accuracy didn't matter for much of the sessions as they were not being read anyway.

Conversely, the ASR captions may actually be an additional source of trouble. For example, Bob frequently looked at the captions to the point where he sometimes attempted repair of the incorrect captions even when Adrian showed no indication of trouble understanding (an example is 'Actually helpful'). Both Tony and Elsbeth frequently looked at the captions, seemingly to check the caption accuracy and on occasion when they were incorrect proved a distraction to the conversation itself (for example 'Mussolini').

Recording of custom cards is a process built especially for improving the ASR recognition of proper nouns. Custom cards may help reduce the frequency of extended repair sequences of nouns relative to other word classes (Bloch & Saldert, 2020). However, the process of creating custom cards is separate from the process of recording other phrases in Relate, and the app design indicates it is optional. Therefore, by design, the implication to the user may be that custom card creation is not very important.

Pam did not understand what Custom cards were and required support. Bob and Elsbeth did record custom cards independently, however the words chosen were mostly context dependent or specific to a distinct subject area, and as a result fewer than expected of the

words said in the sessions. It could be argued that this outcome was influenced by the lack of guidance on best practice for recording custom cards.

If custom cards were part of the same process of recording Relate phrases, and best practice guidelines were available for what words may be most helpful to record, it is possible that a higher number of words useful to the plwMND in everyday conversation could have been recorded, increasing the possibility of more functionally useful ASR captioning in practice.

Guidelines for custom cards are especially important because of the extended time needed for updating the ASR model. In order for custom cards to contribute to ASR recognition for a conversation, they need to be recorded and the ASR model updated before a conversation has happened. This process often takes a few days. The alternative is to create custom cards in response to trouble.

The required time to update the model was not apparent to any of the participants and they assumed the model would immediately improve once a custom card was recorded.

Pam reported finding the Relate app difficult to understand and she required the help of her more tech-savvy husband to support her. The incidence of MND is highest in people aged between 55-79 (Talbot et al., 2012) and a significant proportion of this age group lacks basic digital skills - around half of people aged 75 & over in the UK in 2023, around three in ten (29%) aged 65 to 74, and around one in seven (16%) aged 55 to 64 (Age UK, 2024). Therefore, how much support is available, how easy software is to navigate and how accessible it is are important considerations.

Given the difficulties Pam reported navigating Relate, it seems likely the process had not been designed with her technology needs in mind. Although it is not documented how much user

participation there was when designing Relate, the benefits of understanding user needs are well documented, including more useable products, safer systems and financial savings for suppliers (Ritter et al., 2014).

All participants used only the 'Listen' option in Relate throughout the research even though there are other functions available. Pam said she didn't know about the others, and it was unclear why Bob and Elsbeth never used them. It could be down to preference, or it could indicate a better understanding is needed about how the options may be useful, or something else. As communication occurs in different environments for different purposes it is possible that one function may be more appropriate than another depending on the context, but if the functions are not understood well, then this only serves to limit options available to the user.

In a systematic review, Baxter et al. identified multiple factors as barriers or facilitators to successful use of high tech AAC. Barriers included a lack of technical support, support staff awareness and training (such as the treating SLT), responses from other people and their attitude to perceived usefulness. All these factors appear to resonate with issues raised in this study for personalised ASR technology. The significantly increased speed of generating a message is likely to be a facilitator for usefulness, however it is unclear how beneficial it could be in practice in the presence of the multiple barriers identified here (Baxter et al., 2012).

Additionally, the ASR captions everything it hears without making distinction to who is speaking, and a person referencing the captions may require extra time to orient to the correct part of the captions relating the section of the conversation they want to better understand. Delays in message production can lead to frustration, misunderstanding, inattention, forgetting, and mis-hearings (Higginbotham & Wilkins, 1999).

User-centred design, implementation, training and support – they all matter, because poor implementation of these factors can have multiple unintended consequences that the developers almost certainly never envisaged. Pam more than once blamed herself for the inaccurate captions Relate produced. For Pam, the personalised ASR may have become a tool more to reinforce her perception of speech decline than a way to support communication. Rather than the intended positive user experience, using Relate appeared to increase the risk of negative impact to sense of well-being and sense of identity (Mengesha et al., 2021).

6.2.1. AI and AAC – Speed and Accuracy

The inclusion of AI technology in Relate is argued to enable personalized ASR that better recognises the unique speech patterns of individuals with 'non-standard' speech, potentially enabling plwMND to be better understood (Cattiau, 2021).

AI technology may have the potential for increased speed of AAC communication while also maintaining accuracy (Koch Fager et al., 2019; Light et al., 2019). Relate is argued to achieve this by enabling the user to keep talking (maintaining speed of communication) and be better understood by others due to the accuracy of the personalised speech recognition model (maintaining accuracy of communication) (Cattiau, 2021). Speed is also increased by potentially reducing the requirement of manually typing phrases, or editing pre-stored messages (Valencia et al., 2023).

However, for all three dyads in this study, their experience of using AI-generated ASR captions was different. Although there was clearly an increase in the speed of using AAC when using Relate in comparison to using a whiteboard and pen or texting messages on a smartphone, the

accuracy of the captions was frequently inaccurate. There was a high proportion of inaccurate ASR captions observed for all dyads, some misled the interaction (for example Movie Day Mussolini and Vegetarian Shepherds). Correct punctuation may be important when writing so that the writer appears competent (Kane et al., 2017), however the punctuation Relate dynamically added was seen at times to be not only incorrect but appeared to mislead the interaction (for example in Vegetarian Shepherds).

If an AAC device generates inaccurate, inappropriate or misleading communication, there may be a risk to valued relationships, reputation, as well as the risk of anticipated harms to health. These concerns are in in addition to the frustration of generating inaccurate messages and the extended time of having to correct them (Fried-Oken et al., 2023).

Fried-Oken et al. additionally reported that inaccurate messaging may result in concerns for anticipated harms to health, for example interfering with getting needed care for a medical condition from a doctor (Fried-Oken et al., 2023). This study highlighted an additional risk of potential harm to health: the risk to psychological well-being of the plwMND as a result of self-blame for incorrect captions.

The effects of captioning errors are illustrative of the practical risk of allocating initial responsibility to the ASR for caption accuracy. Unless the plwMND mitigates the risk in some way, for example hiding the screen from the listener until satisfied the captions reflect what they have said, then they have little immediate control over the quality of the captions that are automatically being broadcast. Relate does allow for manual correction of the captions but this is a timely process, requiring multiple steps after the captions have been created, and potentially after the listener has seen them.

Although it could be comparatively slower to communicate using a voice output communication aid (VOCA), whiteboard or texting messages on a smartphone, a plwMND better controls what is written or typed out. Additionally, in the case of a VOCA, prior research showed that dynamic synthetic voicing of texting also runs the risk of creating multiple repair cycles when the typed words are misspelled, or are not yet in the database (for example, foreign words or proper nouns) (Norén et al., 2022).

AI is argued to have the potential to reduce the cognitive effort of using AAC, by identifying an individual's intrinsic needs and skills (Koch Fager et al., 2019). Unfortunately, the experience of understanding and using Relate was reported as challenging and unlikely to have reduced cognitive effort, especially for Pam as she needed the help of her more tech-savvy husband to support her.

It would be helpful to be clear where and how AI is actually being used in a system. Relate 'is built on AI research' (Google - The Keyword, 2023), but it is important to understand what features of Relate incorporates AI technology – in this case the algorithms for personalized ASR, and not the design. The effect of reported poor user experience and challenging usability of the Relate app in functional use may diminish or even negate the benefits that AI technology may provide.

Similar to findings from Valencia et al., even though AI generated text offered the possibility to increase speed and reduce the amount of effort of using AAC, all participants were unsatisfied with the output as it was so often incorrect. The participants could edit the response to get the result they wanted, but rewriting is slow and participants were concerned about how incorrect

automatic phrase generation that was already seen by the listener could impact their relationships (Valencia et al., 2023).

6.3. Conversation Analysis

This section discusses trouble and repair as a feature of the dyad extracts, including analysis of the allocation of responsibility and gaze in trouble and repair observed in this research.

It also presents possible drivers for how the ASR Captions may be supporting the preference for speech and why alternative strategies such as alternative AAC were sometimes utilised only after extended trouble and repair. Additionally, there is a discussion on speech delay and permeability and how it may be a feature of the communication partner's interactions when orientating to the correct caption on the Relate screen. The section concludes with a comparison between the problems of understandability on AAC turn outcomes identified by Bloch & Wilkinson (2011) and those seen here.

6.3.1. Trouble and Repair

Repair was a frequently occurring feature in almost all the dyad's recorded conversations. Trouble arose from multiple sources, including:

conversation partners experiencing difficulties understanding speech with dysarthria (for example in 'Vegan Recipes' (line 3) and 'A little White Board' (line 8).

transcription inaccuracies in the ASR captions that significantly affected or lost the meaning of what the plwMND said, for example 'A walk and then coffee' (line 15).

dynamically added punctuation by the ASR captions that introduced ambiguity of meaning and reduced understandability, for example 'Vegetarian Shepherds' (line 3).

proper nouns, which appeared to be particularly difficult for accurate ASR captioning, for example 'Horatio' (lines 10, 13, 15, 17 and 19)

Difficulties understanding speech with dysarthria co-occurring with incorrect captioning sometimes resulted in "cascading troubles" (G. Lerner & Kitzinger, 2012) and prolonged the repair sequence, for example in 'The A259' (lines 14-27) and increased risk of trouble understanding the AAC user's subsequent attempts at clarification (Bloch & Wilkinson, 2013a).

Additionally, trouble arising from the ASR caption inaccuracies sometimes resulted in becoming the focus of the interaction itself, replacing the initial topic of discussion, for example in 'Mussolini' (lines 5-9) and 'Actually Helpful' (lines 4-14) – trouble caused by an incorrect/unintelligible output signal produced by the AAC device (Bloch & Wilkinson, 2011).

Trouble occurred when the ASR captions were intelligible but not understandable in the context of the interaction, for example 'Vegetarian Shepherds' (line 3) and 'Movie day' (line 3), a similar finding to Bloch & Wilkinson (2004).

Trouble related to the ASR Captions was sometimes raised by plwMND when the captions did not accurately transcribe their words – even when there was no indication the conversation partner had trouble understanding the plwMND for example Bob in 'Actually Helpful' (lines 4-13).

In repair, Bloch and Wilkinson reported that the speakers with dysarthria typically used AAC in an attempt to repair talk problems (Bloch & Wilkinson, 2004), however in this research the participants often continued using speech, sometimes with gesture, to repair AAC problems -

for example in ‘Stereo’ where Bob eventually uses alternative words to describe ‘stereo’, combined with gestures when the ASR Captions are repeatedly incorrect (line 16-23).

Frequently, the ASR captions, although available, were not referenced (for example ‘Vegan Recipes’), and when they were, they were often too inaccurate to be helpful – for example in ‘A walk and then coffee’ (lines 12-21).

The captions occasionally appeared helpful when referenced in repair even when they were inaccurate, for example in ‘Granddaughter getting married’ (line 6: ‘so he is off chemo’) – where the ASR captions did not reflect what Pam actually said, but may have prompted Carol to enact and OIOR to correct Pam to what she probably meant (line 9: ‘right he is’). Another example is in ‘Movie Day’ where the caption was incorrect (line 3: ‘I like that because it's a movie day and a house. And where was he?’ but may have been phonemically similar enough to what Pam actually said (line 1: ‘…Jim’s moving there’) to cue Carol and then to enact an OIOR ‘Moving!’ (line 6).

It was observed how plwMND appeared to adjust their speech, using over articulation, increased stress or increased volume to try and improve ASR Captions accuracy or support conversation partners in repair, or both – for example in ‘A little white board’ (line 10: ‘A little white board’).

Conversation partners often directed repair at plwMND and not referencing the ASR Captions, even though on occasion the captions were correct and may have been useful, for example in ‘Vegan Recipes’, when Relate accurately captioned ‘Vegan’ but Carol did not look at the screen (line 2: ‘Vegan, and she looked up or vegan recipes.’)

There were examples of extended sequences where multiple repair attempts were made using speech alone before alternative strategies were used (or not at all). For example, in 'A walk and then coffee' Pam made three repair attempts using only speech before successfully completing repair by using her whiteboard and pen as alternative AAC. For 'The A259' Pam made four repair attempts before using the whiteboard and pen. Bob kept using speech throughout 'Actually Helpful' (three attempts at repair) and 'Horatio' (four repair attempts). It was only in 'Stereo' (where he made five repair attempts in total) that Bob used alternative words and gesture for the final two repair attempts to describe 'stereo'. Elsbeth also did not use alternative AAC in any of the sessions either, relying on speech regardless of ASR caption accuracy.

It was observed how repair appeared to be treated as successful but may not have been. For example, when Adrian confirmed repair in 'That wasn't bad actually' (line 17) Bob has the opportunity to confirm repair or otherwise at the next turn (Schegloff & Jefferson, 1977), however Bob may have perceived further repair as 'not worth the effort' over the already lengthy and effortful repair of the current, including possibly the experience of prior lengthy repair (Clarke et al., 2019).

The ASR Captions were regularly inaccurate across the extracts for these dyads. They were frequently not referenced by either participant, and when they were, occasionally became such a distraction that they became the focus of the interaction (as in 'Mussolini') . A wider question could be whether the ASR captions provided much functional benefit to repair at all, or perhaps they just often added to trouble when referenced (as in 'Movie day') or delayed repair (as in 'The A259').

It could be argued the accuracy of the captions only mattered when they were looked at, and as the ASR Captions were not always referenced by some of the dyads, the implication is they may be less of a contributory factor to trouble than first appears. For example Carol did not look at the captions at all in 'Vegan Recipes'.

Additionally, in contrast to some other AAC devices, some types of trouble were not observed: the faster production of AAC-generated messages from the ASR captions reduced the potential for within-utterance permeability (G. Lerner & Kitzinger, 2012), where conversation partners may insert further talk while an AAC-generated message is being constructed – increasing the risk of it being disrupted and/or not being completed (Wilkinson et al., 2011).

However, there may be a counter argument for within-utterance permeability of the talk of conversation partners when they reference the ASR Captions – the time taken to orient to the correct caption on the screen may be an opportunity for plwMND to insert talk. As Relate captions everything it hears regardless of who is speaking, additional time may be required to orient to the place where the likely trouble source has been captioned on the screen. A possible example may be when Carol is looking at the ASR Captions in 'Some of the players are nearly eighty' at line 5, and Pam appears to take the extended time that Carol is looking at the captions as a cue of repair (although Carol does not explain the cause of the delay).

Delayed speech through the use of AAC may lead to interactional problems and 'mutual accusations' about character (Robillard, 1994) – though it is the communication partner that may be susceptible to delays due to orientation to the ASR captions mid-conversation. As an example, when Carol is looking at the ASR Captions it may not have been obvious to Pam how easy or difficult it may be to orient to the correct part of the captions as Pam almost never

looked at the screen herself. Therefore, when Carol is looking at the ASR Captions in silence the act is an individual one and in and of itself does not provide a public account of the passage of time (Higginbotham & Wilkins, 1999), the 'temporal imperative' (Clark, 1996).

It could be also argued that a contributory factor to extended repair sequences were the sometimes very limited strategies used by both plwMND and their conversation partners to repair trouble beyond simple repetition using dysarthric speech, for example in 'Horatio' (lines 12-18).

It was observed that when Pam wrote 'A259' and 'Wyvale' on her whiteboard, repair was confirmed immediately by Carol. This example illustrates one important difference between ASR captioning and AAC such as a whiteboard. ASR captioning is automatically generated and any errors it makes are shown on the screen and are usually available to the interaction. The ASR generates captions regardless of the level of confidence the internal algorithm has that they are correct. The ASR cannot 'not caption' – it will caption what it hears. The whiteboard, on the other hand was completely under the control of Pam – she wrote the word and by implication she had the opportunity to ensure the accuracy of the written word reflected what she meant as she considered what to write, as she wrote it, and could correct it herself immediately during and after writing it, and add to it any additional writing of her choice that may help with repair. It may have taken longer for Pam to write 'A259' and 'Wyvale' on the whiteboard than the ASR to (incorrectly) caption both words from her speech output, but Pam was in control of the whiteboard and pen as AAC, and the written output repaired both trouble-sources almost immediately.

Therefore a question could be asked: why did it take so many unsuccessful repair attempts using speech before Pam decided to use this form of AAC? For Bob too, it is unclear why did he not attempt another strategy beyond gesture (perhaps pen and paper) when he was repeatedly unable to complete repair in 'Horatio' or 'Stereo'.

One hypothesis could be that Pam and Bob are simply showing a preference for using speech. For those who are significantly difficult to understand due to changes in their speech, using speech to communicate may remain the first preference in most situations, (Smith & Connolly, 2008), even when speaking is effortful and they are often not understood (Beukelman, Fager, et al., 2007).

For a number of reasons, the ASR captions may also be supporting their preference for speech:

- First of all, the stated objective of Relate is for better understanding of 'non-standard' speech (Cattiau, 2021), possibly aligning with Pam and Bob's preferences. Therefore, speech might be encouraged when Relate is present, by implication.
- Speech is required for the ASR captions to work, therefore the presence of the ASR captions could encourage speech as a modality.
- The ASR captions do not require the either the plwMND or conversation partner to look at the ASR screen at any time – viewing this AAC is not mandatory for it to work - and therefore Relate could be more supportive to plwMND using speech alone than using alternative AAC such as a whiteboard and pen, which does not require speech to be used, but does require both participants to look at it.

- Finally the ASR captions may be perceived as supporting 'social closeness' through the preferential use of speech to maintain and/or develop personal relationships, maintain social etiquette and competency rather than through AAC (Murphy, 2004).

In summary, it may be that using AAC such as a whiteboard and handwriting to repair troubles may be perceived by people that want to talk to communicates as a less preferred option. However, Pam did choose eventually to use the whiteboard and pen in both 'The A259' and 'A walk and then coffee', but it remains unclear from the videoed interactions 'why now' and 'why not earlier' the decision was made in the moment to move from speech to AAC such as the whiteboard and pen.

A more mundane reason for the preference for speech might also be an artefact of this research selection criteria. Participants were recruited to be videoed using speech with the ASR Captions, not other forms of AAC. Therefore the participants may have been more biased to use speech and ASR captions rather than other strategies.

6.3.2. Allocation of responsibility

In this research, other initiated repair (OIR) was a common method for communication partners to initiate the process of repair, providing trouble-source speakers the opportunity to enact repair themselves.

OIR can make it difficult to avoid suggesting that someone else is responsible for the trouble. This happens because the structure of OIR is often context-independent, which can

unintentionally imply that the issue lies with the other person (Sacks et al., 1974). The vulnerability could imply disbelief in, disagreement with, challenge to, or rejection of the action in the trouble-source talk, and may affect how the speaker of the trouble-source responds (Schegloff & Jefferson, 1977), and increase the risk of losing 'positive face' (Brown & Levinson, 1987).

Carol, Adrian and Tony demonstrated in various ways how they managed trouble responsibility as a function of supporting the management of relationships. In 'A walk and then coffee' (**'Wyvale! Sorry Yes…'** line 23) Carol may be utilising and apology-based OIR that orients to her commission of, and accountability for, a possible offense (Robinson, 2004) and may mitigate the potential risk of the OIR supporting Pam's previously declared stance of self-blame when others do not understand her (Carol has reported that Pam frequently blames herself for others not understanding her). Carol also attributes 'Wyvale' to be a 'difficult' word (line 23), potentially attributing to her own personal trouble, perhaps in hearing or understanding.

In 'That wasn't bad actually' Adrian says **'It would sorry'** at line 14 and by doing so generates a OIR that not only initiates repair but also appears to attribute the OIR requirement to his own personal trouble, perhaps in hearing or understanding. Similarly, In 'Horatio' at line 11, Adrian says **'just a sec'** generating an OIR that may imply his own personal responsibility for some difficulty processing the information provided or understanding what was said. In both examples, Adrian's stance appears to be that the responsibility for trouble belongs to himself.

In 'Mussolini', Tony's stance also appears he is taking responsibility for the trouble generated by the incorrect ASR caption at line 5 (***'crap and'***), by saying at line 6 **'everything I said was crap**

HEH HEH it was translated everything I said it said every word was crap'. However the incorrect ASR caption was generated from Elsbeth's words at line 4 '°**right**° .h **and**'.

In contrast, Carol uses a questioning repeat (Jefferson 1978) in 'Movie Day' at line 4 **'because it's a movie day?'** to communicate a stance that responsibility for trouble belongs to Pam. Although Carol is repeating part of the (incorrect) ASR caption at line 3: ***'...because it's a movie day…'*** she nevertheless appears to attribute the trouble to Pam, evidenced by Carol looking at Pam and questioning her directly, perhaps assuming accuracy of the caption. However at line 6, after a 0.9 second delay Carol initiates another OISR without reference to the ASR caption: **'why are you taking it to Maxine?'** – referencing instead what Pam had said on line 1 **'and that's for Maxine'**. Carol may have rephrased the OISR because the 0.9 second delay at line 6 could have oriented her to more than one type of trouble being present – and by excluding reference to the ASR captions could indicate she allocates some responsibility for the trouble lies with the ASR captions. More than one type of trouble can mean that both participants are responsible for trouble (Robinson, 2006). In this case it seems that both the incorrect ASR captions (line 3) and Pam's unclear speech (line 1) had contributed to trouble responsibility. This may also show that the ASR can be treated as an additional participant in the interaction and can be allocated responsibility for trouble.

In 'Actually Helpful', Bob enacted a OIR directly to the ASR, by repeating **'helpful'** at lines 7, 9 and 11 – and by doing so allocated trouble responsibility to the ASR Captions. Bob's goal appeared to be on repairing the Relate captioning at lines 8, 10 and 12, and he only acknowledges repair at line 13 saying **'Ah'** at line 13 and slapping his hands on the table when Relate finally captions ***'helpful'*** accurately at line 12 - evidence that Bob's multiple OIRs appear to be specifically produced to expose an error by Relate. Bob's repeat becomes more

enunciated at lines 7, 9 and 11 compared to line 4, and he moves closer to the ASR device after line 9 to produce a relatively stronger signal which, like enunciation, is an additional practice for repairing hearing trouble (Robinson, 2006). By redoing the same action in the same way, repeating the trouble source as Bob did on lines 7 and 9 ma multiple owners, 'self' and 'other', as shown when Carol references both herself and the ASR captions (but not Pam) in 'The A259' y indicates the action was likely both understandable and 'appropriate' in its original form (Robinson, 2006).

Trouble responsibility of OIR is often allocated in context sensitive ways to claim ownership to 'self' or 'other', or some combination (Robinson, 2006). When the ASR captions are available, trouble responsibility may be allocated to multiple owners, 'self' and 'other', as shown when Carol references both herself and the ASR captions (but not Pam) in 'The A259' (line 28):

"oh A2! It's that it's the A2 I didn't get yeah A259 (0.8) °**and neither did that**° **"**

Trouble responsibility is argued to be an interactionally negotiated affair (Pomerantz, 2021), but in 'Actually helpful' there is no evidence of negotiation per se, as the interaction seems to be between Bob (self) and the ASR device (other) – Adrian appears to be not involved at this point. The ASR is automatically captioning what it hears, accurately or otherwise, and the only changes are to the signal strength when Bob moves closer to the ASR device or when he changes enunciation.

6.3.3. Gaze

In this research eye gaze patterns were seen to have more in common with face to face conversation than AAC use. For example, Pam and Carol gazed very infrequently to where the ASR captions were being generated. On average over the ten sessions Pam looked at the ASR

captions about 1% of the time (average 8 seconds) and Carol for about 7% of the time (average 60 seconds) per session. In 'The A259' Pam directed her gaze at Carol and not the ASR entirely through the extended repair sequence from line 1 to line 16, and Carol only briefly looked at the ASR captions at lines 3 and 8. At line 17, Pam only looks at the ASR captions when Carol has already been looking at them for four seconds, possibly following Carol's gaze.

In contrast, when Pam starts to use the whiteboard at line 27, both Pam and Carol look at the whiteboard without delay until Pam completes writing 'A259'. In this way, both Pam and Carol seem more typical in their pattern of gaze for people with dysarthria using AAC (Wilkinson et al., 2011) when using the whiteboard than when using the ASR captions.

One reason for this could be differences in footing (Goffman, 1981) between AAC such as a whiteboard and the ASR captions. When a whiteboard is used there is a separation between author and animator, with the dysarthric speaker being the author of the utterance but not its animator. Pam wrote the message which was 'animated' by the whiteboard. For the ASR captions there is less distinction between the author and animator because the ASR captions are generated automatically, at the same time as Pam's speech. To the listener, the dysarthric speaker could remain as both author and animator, with the ASR captions a supplementary and optional animator that is automatically generating captions that are available for reference, whether they are being looked at or not.

In addition, when using AAC such as the whiteboard, features which would typically accompany the spoken production of utterances, such as eye gaze toward a recipient, prosody and gesture are less likely to be available. For example when Pam wrote 'A259' there was no eye gaze towards Carol and no speech. In contrast, the ASR captions did not negatively affect Pam's eye-

gaze, prosody, gesture or other features of natural speech – because the ASR automatically captions speech as it is spoken, without delay, and without manual intervention required by the speaker.

The ASR captions may be supporting the preference to present gaze behaviour as 'ordinary' and for others not to others to find our gaze behaviour marked or deviant (Sacks, 2013). Looking at someone's face is argued to be more relevant to an interaction than looking elsewhere during a conversation and looking toward or looking away from the other participant is may be clue in terms of participants' engagement or otherwise in the conversation (Rossano, 2012).

In contrast to Pam and Carol, Bob looked at the captions for much longer than Adrian, 19% on average and increasing from session five to around 29%. Bob may have been increasingly gazing at the captions to 'audit' their accuracy (for example in 'Actually helpful'), a different use-case to supporting Adrian's understanding. Adrian looked at the captions less than Bob, on average 8% of the time.

For Elsbeth and Tony, it seemed that it was Tony that was auditing the captions to check on their accuracy of captioning Elsbeth's speech as opposed to using them to repair trouble. He looked at the ASR on average 42% of the time in a session. Elsbeth referenced the captions on average 11% of the time and also appeared to check if the captions were accurate as opposed to using them to support repair, though not as much as Tony.

In summary, the patterns of gaze observed across the three dyads appeared quite different to when people with dysarthria use AAC as those reported by Wilkinson (2011) with the exception of when Pam and Carol switched to using the whiteboard as AAC.

6.3.4. The understandability of ASR as AAC

Bloch and Wilkinson observed that the outcome of AAC turns were sometimes not fully understood by recipients – described as problems with understandability, extending beyond the anticipated effects of unintelligibility (Bloch & Wilkinson, 2011).

The categories of difficulties they describe and similarities and differences found in this research are discussed below.

Mis-produced AAC output due to an error by a user or an incorrect/unintelligible output signal produced by the AAC device (Bloch & Wilkinson, 2011).

- This research showed a high degree of mis-production of AAC output. However, it was usually not due to errors of user input, but the inaccuracy of the ASR captioning. Automation of captioning in real-time gives the AAC user less control over ensuring the accuracy of output prior to the communication partner accessing it, as is the case here. Pam rarely looked at the captions and did not attempt to control the accuracy or visibility of them. However, Bob did try to repair the ASR as shown in 'Actually Helpful', even when Adrian seemed to already understand.
- Troubles were observed arising from both talk and mis-produced output from Relate – sometimes creating a comp worked to resolve the multiple difficulties (Bloch & Wilkinson, 2013a) . An example is in 'A walk and then coffee' where Pam's lex scenario of cascading troubles in which both participants worked to resolve the multiple difficulties (Bloch & Wilkinson, 2013a). An example is in 'A walk and then coffee' where Pam's spoken attempts at repair and Relate's incorrect captioning caused additional troubles, resolved only when Pam wrote the correct word on her whiteboard.

- Although Relate can be trained to better recognise proper nouns, the process takes a few days for the speech recognition model to be updated for each word. Therefore, inaccuracies of proper noun captioning can take an extended time to be fixed, with the potential for the same word causing additional troubles in subsequent interactions. The longer a trouble takes to resolve, the greater the risk of assigning fault or apportioning blame for the failure (Bloch & Saldert, 2020) and may be a factor for AAC rejection (Judge & Townend, 2013).
- Relate has the option to 'speak' the captions in synthetic speech, similar to a VOCA, but none of the dyads used this feature and there were no examples of an incorrect or unintelligible output signal of this kind produced by the device.

An unclear relation between one AAC-produced word and another that follows it, since the combination of words does not appear to make sense (Bloch & Wilkinson, 2011).

This research observed the issue due to various causes:

- Relate always captions what it hears, regardless of the level of confidence the algorithm may have of accuracy, and this can lead to inaccurate, potentially misleading results, particularly for proper nouns. For example in 'A walk and then coffee', the proper noun 'Wyvale' was captioned incorrectly as 'bye-bye' (line 12), 'why the well going' (line 18), 'why are I fell' (line 21) and 'why fell?'(line 24).
- Relate makes no distinction between who is speaking – the person with dysarthria or communication partner. Instead, it captions all speech it hears as if it is the same person. This may lead to difficulty following who said what when looking at the Relate screen.

- No dyad looked at Relate captions all the time even though Relate generated captions throughout each conversation. Therefore, the clarity of the relation between an AAC-produced word and another may matter less if the AAC is not being referenced at all.

The relationship between AAC-produced utterance and earlier non-AAC-produced utterances may not be clear (Bloch & Wilkinson, 2011).

- In this research, problems with clarity of relationship between AAC and non-AAC produced utterances were less likely due to the extended time required for an AAC-produced turn (as the captions were generated almost simultaneous to speech), and more about the accuracy of automated captioning. An example is in 'Movie Day', when Relate captions 'because it's a movie day'. In this situation the relation between AAC produced utterance and previous non-AAC produced utterance is not clear because Relate is incorrect.

AAC turns in the context of self-repair can be treated as intelligible but not understandable by the conversation partner (Bloch & Wilkinson, 2004)

- In 'Vegetarian Shepherds', Relate accurately transcribed Pam's words, but the captioning amalgamated two phrases, adding punctuation and capitalization (line 3). The words are intelligible but less understandable due to increased ambiguity introduced by the punctuation and capitalization.

AAC was regularly used for a particular purpose and in a particular sequential context (i.e., as attempted self-repairs following other initiation of repairs by the conversation partner) (Bloch & Wilkinson, 2004)

- Pam almost never looked at the ASR captions and did not appear to use them for self-repair. Instead the ASR captions were used by Carol in the process of other initiation of repair.
- However, In 'Some of the players are nearly eighty' Pam takes a visual cue for OISR when she sees Carol looking at Relate – repeating the phrase (line 6).
- Additionally in 'Actually helpful' the ASR captions became the focus of attention and the trouble source, even when the communication partner seems not to have identified trouble.

6.4. Thematic Analysis

Retaining identity as speech changes was important for all three plwMND when dealing with their diagnosis of MND. To be able to keep talking and be understood by friends and family even as speech declines was seen as a way to help preserve identity, to maintain social networks and maintain different roles in life. Communications partners supported plwMND trying Relate, believing it may help maintain lifestyle and social networks.

Both plwMND and communication partners were hopeful that the ASR captions may delay or minimise the effect of voice deterioration – to keep talking and be understood for longer, to maintain a sense of identity and them to exert some control over the future. The communication partners were also hopeful that Relate will help to bring confidence back to participate in activities and maintain social networks.

Focusing on hopeful activities has been identified as a coping strategy for plwMND (Hirano et al., 2006) and Pam and Carol reported enjoyed the opportunity to talk when trialling Relate. Relate may have provided an opportunity to focus on something other than potentially distressing thoughts relating to MND (Kylmä et al., 2009).

Pam reported she was not comfortable with technology, quite common in her age group (Center for ageing better, 2018). Pam felt Relate was not easy to understand, there was no instruction manual and only a short training video, and she relied on her husband to help her with it. Her speech therapist had recommended Relate but could not help with the process either as she had no practical training.

The process of training Relate requires five hundred phrases to be recorded to build a personalised speech recognition model. Some found it tiring – fatigue is one of the most

commonly reported symptoms of MND (Gibbons, Thornton, & Young, 2013). One found it very dull and there was no indication if Relate would even be useful.

Relate can be trained to recognise less common words such as the names of people, places and things. Pam only found out about custom cards after recording the 'standard' phrases and felt this could have set her up for a negative experience when using Relate.

All participants felt the ASR captions were less accurate than expected. The ASR captions were inconsistent, frequently incorrect and unpredictable to the point where Elsbeth thought Relate had **'a mind of its own'**. Carol reported that the Relate captions can still be helpful when having similar phonemic properties to the word Pam said, or if there were some correct words amongst generally incorrect captioning. Carol positioned Relate as providing only **'hints and clues'** and not a direct transcription.

The plwMND reported they would only use Relate in specific contexts and with particular people. Elsbeth said it is not for **'serious conversations'**.

PlwMND and their communication partners frequently had different perceptions of the effect of looking at the ASR captions in conversation. Carol and Adrian felt looking at the Relate screen was an unnatural thing to do mid-conversation and was concerned that the plwMND may feel not listened to. On the other hand, no plwMND reported feeling ignored when others looked at the screen.

For Tony he said it was simpler just to ask what Elsbeth had said rather than check the screen, especially as the Relate captions were not always accurate. For usefulness, Tony said **'Out of five I score Relate a one'**.

In summary, for all participants the expected benefits personalised ASR were not realized. The captions were perceived as too inaccurate in practice, and some indicated was easier and quicker to resolve trouble without using them at all. Relate set-up was difficult to understand for some and recording mistakes were not picked-up as there was little feedback. The lack of an easy to access guide was unhelpful.

Rather than the ASR captions enhancing a sense of personal identity in communication, Pam attributed inaccuracy of ASR captioning to her own speech, and not the technology. This feeling of failure could have a negative impact on how their own sense of personal identity and has parallels with participants blaming the characteristics of their own 'non-standard' speech rather than the technology for inaccurate speech recognition (Mengesha et al., 2021).

6.5. Usability

The primary goal for Relate is to support people with non-standard speech to be better understood by others through the use of captioning from a personalised speech recognition model (Cattiau, 2021). Assessing the usability of Relate is one way of understanding better how closely it meets this goal, how easy an application Relate is to learn, how effective it is use, and how enjoyable the experience is in functional use (Rogers et al., 2023).

Rogers et al. break down usability into seven constituent parts, and the outcomes of this research is used to assess Relate in each of these areas.

6.5.1. Effective to use (effectiveness)

How good is Relate at doing what it is supposed to do? In this study Relate appears to be not accurate enough at captioning conversational speech, and did not meet the expectations of caption accuracy that Relate itself provided to the each plwMND. For Pam the actual average ASR WER per session varied between 38.2 – 53.2%, for Bob between 33.6 – 42.5%, and Elsbeth between 30.6 – 34%, all significantly worse than the expectation of accuracy set by Relate. When considering loss of meaning, Relate captions lost the meaning of what for Pam and Bob said between 50-70% of all phrases in the sessions. For Elsbeth, it was between 60-75%.

6.5.2. Efficient to use (efficiency)

Does Relate support people in carrying out the task of 'being better understood' in an easy and efficient manner without unnecessary keystrokes or menu navigation? In some ways yes: none

of the plwMND commented on difficulty of switching Relate app on and starting the automatic captions. However the communication partners did have difficulties: there was frequently a lot of text on the screen to navigate in the middle of a conversation due to Relate captioning everything it hears, including communication partners. Sometimes the captions had already rolled off the screen before they could be referenced. The brightness of the screen and the font size needed easier or automatic adjustment. Consideration of screen positioning was required as conversation partners believed there would be a negative perception when breaking eye contact to look at Relate.

6.5.3. Having good utility (utility)

Does Relate the provide appropriate functionality so that users can do 'being better understood'? The automatic captioning appeared easy to switch on and use for the plwMND, and the Custom Cards function appeared to support increased accuracy of recognition for individual words. However, none of the plwMND used the three other modes in Relate (Repeat, Assistant, Keyboard), even though they may have been helpful in some situations. Additionally, for one plwMND the benefits of Custom Cards were not apparent at first, and for the others the words added to the Custom Cards were frequently not spoken in later sessions, potentially indicating difficulty predicting what may be said in future conversations, or perhaps not fully understanding the functionality.

6.5.4. Easy to learn (learnability)

Is Relate easy to learn to use? It seems Relate was not easy to learn initially. Two of the plwMND needed help to set up Relate and record the phrases in order to create a personalised model, and the treating speech therapist did not have knowledge of Relate either. There was no indication of potential future benefit during the recording stage and no feedback if a mistake was made, making the task for some even longer. The benefit of recording custom cards was not apparent.

Pam and Elsbeth felt the number of phrases needed to train Relate was a big commitment, particularly when dealing with fatigue – a factor in MND. Bob found the process of recording so many phrases very dull and needed prompting to complete it.

6.5.5. Easy to remember how to use (memorability)

Is Relate easy to remember how to use, once learned? It seems that Relate was easy to remember how to use. Each plwMND set up Relate independently at the beginning of the videoed sessions and switched the captioning on. Once Pam had been shown how to record custom cards she did not need support to record more of them.

6.5.6. Enjoyable to use (satisfaction)

How acceptable is Relate when being used? How satisfied are people when using Relate?

Using Relate in practice appeared to be a poor experience for both plwMND and communication partners: the low caption accuracy appeared to be unacceptable to many, to

the point where some reported it was easier to talk directly with each other than use Relate. One conversation partner said the captioning was indirectly helpful – providing hints as to what was said – captioning similar sounding words or words with related semantic meaning. Overall most reported Relate was not accurate enough to be useful in many contexts.

6.5.7. Safe to use (safety)

Does Relate protect a person from undesirable situations? If an AAC device generates inaccurate, inappropriate or misleading communication, there may be a risk to valued relationships, reputation, as well as the risk of anticipated harms to health. (Fried-Oken et al., 2023). There were examples of inaccurate, and misleading Relate captions, for example in Vegetarian Shepherds and Mussolini. Additionally, Elsbeth may have realized the undesirable risk of incorrect captions when deciding not to use Relate in the meeting with her university supervisor.

For Pam, the inaccurate Relate captions extended to a sense of personal failure: “This is my fault when I'm not understood by Relate. Relate did not protect Pam from allocating potentially undue responsibility of captioning errors to herself, because Relate set expectations for ASR accuracy that were not reflected in lived experience, and the Relate guidelines were not clear on the limitations of this relatively new technology, ultimately risking negative effects on her own psychological wellbeing. Disability inclusion can be a ‘wicked problem’ – although the aim of the technology in itself might be intended as positive its application could have a perversely negative impact (Holloway, 2021) - as may be the case here.

6.5.8. Usability summary

Usability goals are typically stated as questions and can help an interaction designer with a concrete means of assessing various aspects of a product and the user experience (Rogers et al., 2023). Based on this limited research, it is apparent that Relate usability can be significantly improved in many different areas. It also shows how much can be learnt from end-user feedback on how a system works in practice, rather than how it is expected to work.

6.6. Discussion – Summary

This is the first research that explores how people living with MND (plwMND) utilize personalised ASR technology that is argued to better recognise speech with dysarthria and support everyday conversational communication (Cattiau, 2021).

Using speech for communication was important for all three plwMND in this study. Talk is a part of their lives and their sense of identity - 'I can talk for England' said one. What we say and how we say it is linked to how we perceive our identity, it shines a light to others on how we want to be represented (Cave & Bloch, 2021b).

When speech changes or is lost, some plwMND say this is the hardest experience of MND (Hecht et al., 2002). The loss can have a negative impact on identity, and in turn quality of life (Felgoise et al., 2016) . When MND progresses and speech declines often so do social networks, sometimes to the point where visits to the doctor or MND clinic are the most common form of social interaction (Gibbons et al., 2013) . Using devices that replace speech to communicate are significantly slower that many choose to speak less, only say what is necessary, have others speak on their behalf or say nothing (Clarke et al., 2023).

Many plwMND prefer to talk, even if difficult for others to understand (Beukelman, Fager, et al., 2007) and some people with significantly poor speech sound intelligibility can remain highly interactive and participative (Hartelius et al., 2008).

In this context, a tool that claims to help people with declining speech to be better understood (Cattiau, 2021) could support what many plwMND may want, and how they want to represent themselves, possibly delaying the negative impacts of speech change and loss.

All three plwMND were hopeful that personalised ASR would help them and were keen to try it. All three conversation partners were positive about the potential benefits of being able to better understand would support the plwMND maintain their sense of identity, their relationships as well as social and work networks.

This research studied the effect of ASR captioning from three perspectives:

- From the perspective of the ASR technology itself – how it sets user expectations on ASR performance, how it measures its own performance using forecast WER thresholds, how accurately it captions speech and maintains meaning.
- From the perspective of conversation analysis (CA) – describing how each dyad conducted social interactions and how troubles and repair were managed. CA does not explain the technology or features of the technology but of the social interaction when ASR captions are present.
- Directly from the perspective of plwMND and their communication partners through semi-structured interviews, using Thematic Analysis (TA) for qualitative analysis of the interview transcripts.

Participants reported that there were difficulties almost from the start when using ASR captions. Relate was not intuitive to use for people with little confidence around technology. The initial 500 recordings were tiring for some, boring and arduous for others, and with no indication of potential benefit. All three plwMND used 'Listen' even though there are three other options which may have been helpful. The app design implied that custom cards were an optional activity and therefore of little importance. Good app design should encourage exploration and tolerate mistakes as well as constrain and guide users to select appropriate actions when initially learning (Rogers et al., 2023).

Relate set expectations of model accuracy based on forecast WER that were overstated when compared to the actual WER of the conversations recorded for this study. The percentage of incorrectly transcribed captions that had lost the meaning of what the plwMND said was very high, at times more than 70% of all phrases in a session. The expected accuracy that the Relate messages to users only takes into account the forecast WER for shorter set phrases the user is unlikely to say, and is consequently a poor forecast for listener understanding, which is the key objective of Relate. Expectation management is the process of making sure that expectations of a product are realistic. If people feel the product has been misrepresented, then this will cause resistance and even rejection (Nevo & Wade, 2007).

The research indicated that Custom cards may indeed improve ASR recognition of proper nouns, but a much larger study is required to confirm if this is the case generally. That being said, the analysis showed that for some plwMND, for the words they recorded custom cards for were said very few times in the videoed conversations, demonstrating how difficult it may be to predict the content of future conversations. A key concept in software design is to help the user to avoid mistakes, and easily repair mistakes if they happen (Accessibility Guidelines

Working Group, 2023), however there appears to be little in the design of Relate that supports the user to help reduce ASR caption errors - the process of custom card creation and recording is not in the critical path, and when caption errors are made there is no process to quickly create custom cards in 'the moment' of the caption error being generated.

The research reported that repair was a feature in the interactions between the dyads, with clearly occurring instances of other initiation of repair, similar to other research for people with dysarthria using AAC. The ASR captions were reported to be sometimes helpful in repair even when inaccurate but phonemically close enough to prompt for the right word. The research observed how inaccurate captions can cause cascading troubles requiring extended repair. At other times the captions themselves became the source of trouble even when the conversation partner gave little indication of trouble themselves. Sometimes inaccurate captions were a distraction and became the focus of the interaction.

One plwMND resorted to another form of AAC (a white board) when trouble was not resolved despite multiple cycles of repair that Relate was unable to help resolve. However, another plwMND resolutely focused on repairing the ASR Captions using dysarthric speech, even when there was little indication of trouble from the communication partner.

This research showed a significant variation in how often the plwMND and communication partners looked at the Relate screen during interaction, potentially indicating different uses for the captions. Bloch et al. explored if AAC use itself generates a new framework within which interaction operates (Bloch & Wilkinson, 2013a): for Pam and Carol it seems their interaction was mostly operating independently of the framework of ASR captions as they rarely referenced them, but for Bob the accuracy of captioning appeared to be more relevant to the

interaction as time progressed and as his speech declined. Bob more than once attempted to correct the captions with over-articulation and louder speech, even when no trouble was indicated by his communication partner.

People with dysarthric speech use speech and AAC to communicate along a continuum rather than being mutually exclusive (Bloch & Tuomainen, 2017) and although Relate likely fits with this concept by being designed to be used in tandem with speech with dysarthria, based on this research, the overriding problem with Relate is it may not be perceived by many as accurate enough to be functionally beneficial.

Thematic analysis indicated that plwMND and conversation partners were hopeful that personalised ASR technology would be helpful, to help them to keep talking, be themselves, but none felt it had met expectations. Many reported that the caption accuracy was just not accurate enough to be useful and their expectations had been set incorrectly, although some said it may help in particular contexts, with particular people.

Trust in technology devices is argued to be largely based on the extent to which a device is perceived to properly perform its function (Muir & Moray, 1996) , displaying expert knowledge, technical facility, and everyday routine performance (Muir, 1987). The implication is that errors such as the significant rate of ASR inaccuracy observed here could strongly affect trust in Relate.

If there is unclear understanding of the technical limitations of Relate, there is a risk that responsibility for ASR inaccuracy that is rooted in the limitations of the technology is perceived as owned instead by the person using it. The Relate ASR captions are based on technology with known factors that can affect performance – for example limitations around proper noun word recognition amongst other factors.

What may be a misplaced allocation of responsibility for incorrect ASR captions is not without precedent - speakers are reported to sometimes link the errors that ASR makes to their own sense of identity, including racial, regional and location identity, and many participants modified their dialect and word choice in order to be understood by ASR (Mengesha et al., 2021).

Maintenance of personal identity can be impaired by the limitations of AAC devices as well as the disease (Kane et al., 2017), and the limitations of Relate caption accuracy may also have a negative impact on personal identity. One plwMND seemed to blame her own speech for inaccurate ASR captions without taking into consideration the limitations of the technology itself, risking negative effects on her own psychological wellbeing due to ASR performance. This research underscored the importance of setting correct expectations for ASR performance, including clarity of responsibility when the captions are incorrect – including clearly defining what the limitations are of this relatively new technology.

Without clarity on the technical limitations of ASR in practical use, and given the number of captioning errors the ASR makes, there may be an increased risk of misplaced responsibility and ownership of these errors and increased risk of negative impact on sense of personal identity when Relate is used.

This research also indicates the practical benefits of understanding the needs of both plwMND and communication partners in product design and argues that co-design creates better outcomes including more useable products and provide financial savings for suppliers.

Personalised ASR captioning is claimed to be high technology AAC that supports plwMND with speech with dysarthria to be better understood for longer (Green et al., 2021). But in practice,

for these participants at least, there are significant improvements required before it could better realize this goal. The impact of this misalignment of expectation can have negative consequences for trust in the technology as well as for the sense of personal identity for plwMND who use it.

The focus of this research was to assess how Relate supports plwMND to be understood by others in everyday conversation, and how it may change as speech declines during the course of the disease. The outcomes showed that Relate did not help plwMND to be understood as much as hoped for by plwMND and communication partners, or as much as the expectations set by Relate itself, and this did not change over time.

7. Limitations

MND is a relatively rare condition and for plwMND there are many priorities on their ever-decreasing time. Recruitment expectations were not met. Personalised ASR technology is relatively new, not well known, and as a result it was difficult to recruit plwMND using Relate and willing to participate in the research. Despite an attempt to recruit nationally, the participants were recruited mainly from the London area. The small size of the sample recruited is unlikely to balance across demographic or other characteristics. Further research is required to determine how representative these findings are across the wider population of plwMND and significant others.

Relate is still in development. As the technology underpinning Relate is updated it is quite possible the outcomes identified here will change.

In the period of this research the speech of two of the three plwMND did not significantly change. There was very limited opportunity to assess how the use and usefulness of ASR would change with speech decline.

Custom cards are claimed to improve ASR recognition of key words and phrases. Although this study found some benefit to recording custom cards, it was a very small sample.

All three dyads chose to record conversations in a quiet space. More research is required to explore how useful Relate may be in other situations and scenarios, for example in a group environment, or when out shopping.

This research focused entirely on one module of Relate, the Listen function as this was what the three dyads used. Further research would be needed to explore the use and usefulness of the other functions.

This study highlighted the need for support when using ASR technology and additional research would benefit including professionals such as SLTs and associated stakeholders, which would enhance the strength of any recommendations.

8. Recommendations

One goal of this research was to inform provision of information and training about Relate to plwMND, significant others, SLTs and other professionals. The following recommendations are made in the context of the limitations of this research already described.

- Realistic expectation setting – all participants expected Relate to be more accurate than it was. In this research the forecast WER was almost always lower than actual WER. It is important to provide plwMND with accurate information on the expected performance of Relate, to let them make their own decision whether training Relate is worth the effort or their time.
- Relate should be positioned as one of many AAC options available to plwMND, that it is not a 'one size fits all solution' and may be better suited in particular contexts, people, or locations, and/or used in combination with other AAC. Different AAC tools can be, or perhaps should be, used in parallel with Relate, as they may be quicker when repair is required.
- Development of a clear and accessible guide to Relate is needed to help meet the needs of plwMND unfamiliar with or lacking confidence in technology.
- Development of a guide for conversation partners, recognising that they are potential beneficiaries of this technology.
- The design of Relate should be reviewed, with feedback from plwMND and conversation partners. This research identified that the design was not intuitive for people less confident with technology, and there were a number of design issues highlighted by both plwMND and conversation partners that need to be taken into account.

- Some plwMND found recording the initial 500 phrases tedious and fatiguing, with no indication of the level of benefit it would give. There should be some way of showing tangible progress, perhaps progress markers for forecast ASR accuracy. Techniques such as gamification could be considered to help motivate plwMND to complete the recordings. Gamification has been shown to support motivation to complete some repetitive tasks (albeit with caveats) (Hamari et al., 2014).
- If custom cards do indeed improve ASR recognition of key words, then it is important that there is specific awareness raising and training on how to get the best from them and should be on the critical path of recording.
- The display of Relate should be easier for the reader to orient to the captioned words relevant to the point in the conversation – given that Relate captions everything it hears it would be helpful to be able to easily and quickly identify on the screen who is saying what.
- Privacy considerations should be reviewed. Captioning the words of people other than plwMND is a concern. A setting to adjust how much history of the conversation stays on the screen would enable the users to maintain more control.

9. Implications for future research

Many plwMND report that the loss of speech as one of the worst aspects of MND (Hecht et al., 2002), and all three plwMND in this research were hopeful that Relate could help them be better understood in their everyday conversations. Assistive technology that supports plwMND to use natural speech to communicate and to be understood for longer is an likely to be an important and functionally beneficial use-case for this community.

The following areas for future research are proposed:

Co-review and co-redesign of Relate: it would be useful for plwMND and their communication partners to co-review the objectives, design and useability of Relate with the developers, prioritizing any recommendations. This research should include the experiences of a larger group of plwMND who are using Relate or have chosen not to use Relate to understand the decision making process and the lived experience of using it. The research should include feedback from conversation partners, carers, speech therapists and other professionals, representing the 'full ecosystem' in which plwMND operate – to better understand the complexity and variety inherent in their lived experience (Holloway, 2021). The research should include clarity on the limitations as well as the strengths of this technology, to support plwMND to make an informed choice about if and when Relate may be helpful to them. In this way, Relate may be better designed, easier to learn, more effective to use - providing an enjoyable experience for the intended users (Rogers et al., 2023).

How Relate may support communication across different environments and contexts that people talk in: this thesis focused on conversation with familiar communication partners in a quiet space. It would be useful to understand how useful Relate may be in noisier environments, with unfamiliar conversation partners, in transactional as well as conversational contexts (for example when asking for directions or when buying something in a shop) – assessing Relate in the midst of 'messy human computer interactions of the real world' (Rogers, 2011).

Identify an alternative measure to WER for Relate to use when forecasting usefulness. It was apparent in this research that WER could not accurately measure the success or otherwise of an interaction where Relate was present. PlwMND and communication partners can rely on multiple modalities as well as awareness of the context but Relate can only use WER to provide a forecast indication of its usefulness. Therefore further research would be beneficial to identify a more appropriate measure that the Relate app can use to help plwMND and communication partners decide if and when Relate may be useful to them, or not.

The impact of custom cards on ASR use and usefulness in everyday conversational settings: identifying best practice guidelines for plwMND and those supporting them.

10. Conclusion

The goal of this research was to provide insight into what degree Relate actually supports plwMND to be understood in everyday conversation with communication partners over time. Four research questions were investigated, and a summary of results for each forms the conclusion to this thesis.

RQ1: How does the Relate app enable plwMND to be understood by others in everyday conversation?

The research shows that the Relate captions did not often support plwMND to be understood by others in everyday conversation. The caption accuracy was consistently lower than what was forecast by Relate to all three plwMND, and the percentage of captioned phrases with errors that lost overall meaning of the phrase was often very high, at times more than 70% of all phrases in a session.

However, on occasion the captions were reported to be helpful in repair - when they were phonemically close enough to prompt for the right word (but still inaccurate). Although at other times, the inaccuracy of the captions became a distraction and the focus of the interaction. It was also seen that the captions sometimes became the source of trouble themselves, with subsequent repair attempts by the plwMND - even when the conversation partner gave little indication of trouble themselves.

RQ1a: To what extend does this change as speech declines during the course of the disease?

As speech declined, the research showed that the Relate captions remained generally unhelpful to support plwMND to be understood by others in everyday conversation. However, over the period of this study, only one plwMND experienced a decline in speech and more research is needed. It was observed that the plwMND appeared to increasingly use Relate for a different purpose as his speech declined - to check how accurate the captions were for his speech, perhaps to measure his speech change. He did this even when his communication partner appeared to understand what was said.

RQ2: How do the captions produced by Relate help plwMND to be understood in the wider context of interaction?

The research indicated that the captions did not often help plwMND to be understood in the wider interactional context. Even though repair was a prominent feature in the interactions between all dyads, it was observed that the inaccurate captions themselves sometimes caused cascading troubles requiring extended repair, sometimes becoming the focus of the interaction. One plwMND was seen to resort to another form of AAC (a white board) when trouble was not repaired, despite multiple cycles of repair that Relate was unable to help resolve.

Patterns of eye gaze for both the plwMND and communication partners were observed to have more in common with face to face conversation than AAC use - the implication being that the Relate captions were sometimes not referenced very often to support repair of trouble.

RQ3: How accurate are the predictions and measures Relate uses in its software?

The accuracy of captions in everyday conversation for all three plwMND were consistently lower than forecast by Relate, and the percentage of captioned phrases with errors that lost overall meaning was frequently high, sometimes more than 70% of the phrases in a session.

The research indicated that Custom cards may indeed improve ASR recognition of proper nouns, but more research is needed to confirm if this is the case generally.

In summary, the research outcomes for this study showed that Relate did not help plwMND to be understood as much as hoped for by plwMND and communication partners, or as much as the expectations set by Relate itself, and this did not change over time.

References

Accessibility Guidelines Working Group. (2023). *WCAG 2 at a Glance | Web Accessibility Initiative (WAI) | W3C*. Web Accessibility Initiative. https://www.w3.org/WAI/standards-guidelines/wcag/glance/

Age UK. (2024). *Briefing Facts and figures about digital inclusion and older people*. www.ageuk.org.uk

Amir, O., & Levine-Yundof, R. (2013). Listeners' attitude toward people with dysphonia. *Journal of Voice*, *27*(4), 524.e1-524.e10. https://doi.org/10.1016/j.jvoice.2013.01.015

An, K. H., Kim, M., Teplansky, K., Green, J. R., Campbell, T. F., Yunusova, Y., Heitzman, D., & Wang, J. (2018). Automatic early detection of amyotrophic lateral sclerosis from intelligible speech using convolutional neural networks. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2018-September*, 1913–1917. https://doi.org/10.21437/Interspeech.2018-2496

Angrick, M., Luo, S., Rabbani, Q., Candrea, D. N., Shah, S., Milsap, G. W., Anderson, W. S., Gordon, C. R., Rosenblatt, K. R., Clawson, L., Tippett, D. C., Maragakis, N., Tenore, F. V., Fifer, M. S., Hermansky, H., Ramsey, N. F., & Crone, N. E. (2024). Online speech synthesis using a chronically implanted brain–computer interface in an individual with ALS. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-60277-2

Antaki, C., & Wilkinson, R. (2012). Conversation Analysis and the Study of Atypical Populations [Bookitem]. In *The Handbook of Conversation Analysis* (pp. 533–550). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118325001.ch26

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology: Theory and Practice*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Armstrong, R., Hall, B. J., Doyle, J., & Waters, E. (2011). "Scoping the scope" of a cochrane review. In *Journal of Public Health* (Vol. 33, Issue 1, pp. 147–150). Oxford University Press. https://doi.org/10.1093/pubmed/fdr015

Azevedo, N., Kehayia, E., Jarema, G., Le Dorze, G., Beaujard, C., & Yvon, M. (2024). How artificial intelligence (AI) is used in aphasia rehabilitation: A scoping review. In *Aphasiology* (Vol. 38, Issue 2, pp. 305–336). Routledge. https://doi.org/10.1080/02687038.2023.2189513

Ball, L. J., Beukelman, D. R., & Pattee, G. L. (2004a). Acceptance of augmentative and alternative communication technology by persons with amyotrophic lateral sclerosis. *AAC: Augmentative and Alternative Communication*, *20*(2), 113–122. https://doi.org/10.1080/0743461042000216596

Ball, L. J., Beukelman, D. R., & Pattee, G. L. (2004b). Communication effectiveness of individuals with amyotrophic lateral sclerosis. *Journal of Communication Disorders*, *37*(3), 197–215. https://doi.org/10.1016/j.jcomdis.2003.09.002

Ballati, F., Corno, F., & De Russis, L. (2018a). Assessing virtual assistant capabilities with Italian dysarthric speech. *ASSETS 2018 - Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, 93–101. https://doi.org/10.1145/3234695.3236354

Ballati, F., Corno, F., & De Russis, L. (2018b). *"Hey Siri, Do You Understand Me?": Virtual Assistants and Dysarthria*. https://doi.org/10.3233/978-1-61499-874-7-557

Baughman, A. K., Hammer, S., & Provan, D. (2017). Using language models for improving speech recognition for U.S. Open Tennis Championships [Article]. *IBM Journal of Research and Development*, *61*(4), 15-15:9. https://doi.org/10.1147/JRD.2017.2712019

Bäumer, D., Talbot, K., & Turner, M. R. (2014). Advances in motor neurone disease. In *Journal of the Royal Society of Medicine* (Vol. 107, Issue 1, pp. 14–21). SAGE Publications Ltd. https://doi.org/10.1177/0141076813511451

Baxter, S., Enderby, P., Evans, P., & Judge, S. (2012). Barriers and facilitators to the use of high-technology augmentative and alternative communication devices: A systematic review and qualitative synthesis. In *International Journal of Language and Communication Disorders* (Vol. 47, Issue 2, pp. 115–129). https://doi.org/10.1111/j.1460-6984.2011.00090.x

Beukelman, D. (1998). *Augmentative and alternative communication : management of severe communication disorders in children and adults* (Pat. Mirenda, Ed.; 2nd ed) [Book]. P.H. Brookes Pub.

Beukelman, D., Fager, S., & Nordness, A. (2011). Communication support for people with ALS. In *Neurology Research International* (Vol. 2011). Hindawi Publishing Corporation. https://doi.org/10.1155/2011/714693

Beukelman, D. R. (2013). *Augmentative & alternative communication : supporting children and adults with complex communication needs* (Pat. Mirenda, Ed.; 4th ed.) [Book]. Paul H. Brookes Pub. Co.

Beukelman, D. R., Fager, S., Ball, L., & Dietz, A. (2007). AAC for adults with acquired neurological conditions: A review. In *AAC: Augmentative and Alternative Communication* (Vol. 23, Issue 3, pp. 230–242). https://doi.org/10.1080/07434610701553668

Beukelman, D. R., Garrett, K. L., & Yorkston, K. M. (2007). *Augmentative communication strategies for adults with acute or chronic medical conditions* (D. R. Beukelman, K. L. Garrett, & K. M. Yorkston, Eds.) [Book]. Paul H. Brookes.

Bhagtani, K., Yadav, A. K. S., Bestagini, P., & Delp, E. J. (2024). *Are Recent Deepfake Speech Generators Detectable?* 277–282. https://doi.org/10.1145/3658664.3659658

Bhattacharjee, T., Jayakumar, A., Belur, Y., Nalini, A., Yadav, R., & Ghosh, P. K. (2023). Transfer Learning to Aid Dysarthria Severity Classification for Patients with Amyotrophic Lateral Sclerosis. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2023-August*, 1543–1547. https://doi.org/10.21437/Interspeech.2023-1542

Bhattacharjee, T., Mallela, J., Belur, Y., Atchayaram, N., Yadav, R., Reddy, P., Gope, D., & Ghosh, P. K. (2021). Effect of noise and model complexity on detection of amyotrophic lateral sclerosis and Parkinson's disease using Pitch and MFCC. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2021-June*, 7313–7317. https://doi.org/10.1109/ICASSP39728.2021.9413997

Bloch, S. (2005). *Co-constructing Meaning in Acquired Speech Disorders: Word and Letter Repetition in the Construction of Turns*.

Bloch, S. (2011). Anticipatory other-completion of augmentative and alternative communication talk: A conversation analysis study. *Disability and Rehabilitation*, *33*(3), 261–269. https://doi.org/10.3109/09638288.2010.491574

Bloch, S., & Clarke, M. (2013). Handwriting-in-interaction between people with ALS/MND and their conversation partners. *AAC: Augmentative and Alternative Communication*, *29*(1), 54–67. https://doi.org/10.3109/07434618.2013.767497

Bloch, S., & Saldert, C. (2020). Person reference as a trouble source in dysarthric talk-in-interaction. In *Atypical Interaction: The Impact of Communicative Impairments within Everyday Talk* (pp. 347–372). Springer International Publishing. https://doi.org/10.1007/978-3-030-28799-3_12

Bloch, S., Saldert, C., & Ferm, U. (2015). Problematic topic transitions in dysarthric conversation. *International Journal of Speech-Language Pathology*, *17*(4), 373–383. https://doi.org/10.3109/17549507.2014.979879

Bloch, S., & Tuomainen, J. (2017). Progressive dysarthria and augmentative and alternative communication in conversation: establishing the reliability of the Dysarthria-in-Interaction Profile. *International Journal of Language and Communication Disorders*, *52*(1), 3–9. https://doi.org/10.1111/1460-6984.12258

Bloch, S., & Wilkinson, R. (2004). The understandability of AAC: A conversation analysis study of acquired dysarthria. *AAC: Augmentative and Alternative Communication*, *20*(4), 272–282. https://doi.org/10.1080/07434610400005614

Bloch, S., & Wilkinson, R. (2009). Acquired dysarthria in conversation identifying sources of understandability problems. *International Journal of Language and Communication Disorders*, *44*(5), 769–783. https://doi.org/10.1080/13682820802546969

Bloch, S., & Wilkinson, R. (2011). Acquired dysarthria in conversation: Methods of resolving understandability problems. *International Journal of Language and Communication Disorders*, *46*(5), 510–523. https://doi.org/10.1111/j.1460-6984.2011.00076.x

Bloch, S., & Wilkinson, R. (2013a). *Multiple Troubles and Repair in Voice Output Communication Aid-Mediated Interaction*.

Bloch, S., & Wilkinson, R. (2013b). The accomplishment of non-serious talk in severe speech disability: An examination of recipient uptake and delayed other-initiated repair. *Journal of Interactional Research in Communication Disorders*, *4*, 45–70.

BLTT. (2017). *Switch Access: an introduction - Better Living Through Technology*. Better Living Through Technology. https://bltt.org/introduction-to-switch-access/

Bolden, G. B. (2006). Little words that matter: Discourse markers “So” and “Oh” and the doing of other-attentiveness in social interaction. *Journal of Communication*, *56*(4), 661–688. https://doi.org/10.1111/j.1460-2466.2006.00314.x

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brett, J., Staniszewska, S., Mockford, C., Herron-Marx, S., Hughes, J., Tysall, C., & Suleman, R. (2014). Mapping the impact of patient and public involvement on health and social care research: A systematic review. *Health Expectations*, *17*(5), 637–650. https://doi.org/10.1111/j.1369-7625.2012.00795.x

Bright, R., Gadgil, S., Hawley, M., Cunningham, S., & Fryer, K. (2019, March 1). *Vocatempo: development and evaluation of a voice-input voice-output communication aid app*. https://doi.org/10.13140/RG.2.2.10237.31207

Brown, & Levinson, S. (1987). *Politeness : some universals in language usage* (S. C. Levinson, Ed.) [Book]. Cambridge University Press.

Brown, R., & Cave, R. (2024). *Forecast demand for voice banking funded by the MND Association*.

Brownlee, A., & Palovcak, M. (2007). The role of augmentative communication devices in the medical management of ALS. In *NeuroRehabilitation* (Vol. 22). IOS Press.

Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. In *Discourse Studies* (Vol. 7, Issues 4–5, pp. 585–614). https://doi.org/10.1177/1461445605054407

Bunnell, H. T., Pennington, C., Yarrington, D., & Gray, J. (2005). *Automatic personal synthetic voice construction*. http://www.asel.udel.edu/speech/mt/

Button, G., & Lee, J. R. E. (1987). *TALK AND SOCIAL ORGANISATION* [Book].

Caballero-Morales, S. O., & Trujillo-Romero, F. (2014). Evolutionary approach for integration of multiple pronunciation patterns for enhancement of dysarthric speech recognition. *Expert Systems with Applications*, *41*(3), 841–852. https://doi.org/10.1016/j.eswa.2013.08.014

Cai, S., Lillianfeld, L., Seaver, K., Green, J. R., Brenner, M. P., Nelson, P. C., & Sculley, D. (2021). *A Voice-Activated Switch for Persons with Motor and Speech Impairments: Isolated-Vowel Spotting Using Neural Networks*.

Calvo, I., Tropea, P., Viganò, M., Scialla, M., Cavalcante, A. B., Grajzer, M., Gilardone, M., & Corbo, M. (2021). Evaluation of an Automatic Speech Recognition Platform for Dysarthric Speech. *Folia Phoniatrica et Logopaedica*, *73*(5), 432–441. https://doi.org/10.1159/000511042

Cattiau, J. (2019, May 9). *How AI can improve products for people with impaired speech*. Google Blog. https://www.blog.google/outreach-initiatives/accessibility/impaired-speech-recognition/

Cattiau, J. (2021). *A communication tool for people with non-standard speech*. https://blog.google/outreach-initiatives/accessibility/project-relate/

Cave, R., & Bloch, S. (2021a). The use of speech recognition technology by people living with amyotrophic lateral sclerosis: a scoping review. *Disability and Rehabilitation: Assistive Technology*. https://doi.org/10.1080/17483107.2021.1974961

Cave, R., & Bloch, S. (2021b). Voice banking for people living with motor neurone disease: Views and expectations [Article]. *International Journal of Language & Communication Disorders*, *56*(1), 116–129. https://doi.org/10.1111/1460-6984.12588

Caves, C. (2007). Developing an ASR for ALS dysarthric speech. *RESNA*.

Center for ageing better. (2018). *The digital age: new approaches to supporting people in later life get online Centre for Ageing Better*. *May*. https://doi.org/10.31077/ageing.better.2018.05a

Chen, Y.-J., Huang, Y.-Z., Chen, C.-Y., Chen, C.-L., Chen, H.-C., Wu, C.-Y., Lin, K.-C., & Chang, T.-L. (2019). Intermittent theta burst stimulation enhances upper limb motor function in patients with chronic stroke: A pilot randomized controlled trial [Article]. *BMC Neurology*, *19*(1), 69–69. https://doi.org/10.1186/s12883-019-1302-x

Chen, Z., Ramabhadran, B., Biadsy, F., Zhang, X., Chen, Y., Jiang, L., Chu, F., Doshi, R., & Moreno, P. J. (2021). *Conformer Parrotron: a Faster and Stronger End-to-end Speech Conversion and Recognition Model for Atypical Speech*.

Clark, H. H. (1996). *Using language* [Book]. Cambridge University Press.

Clarke, V., & Braun, V. (2013). *Successful qualitative research: a practical guide for beginners* [Book]. Sage Publications.

Clarke, Z., Judge, S., Fryer, K., Cunningham, S., Toogood, J., & Hawley, M. (2019). Is it worth the effort? Communicating with partially intelligible speech. *In Review*.

Clarke, Z., Judge, S., Fryer, K., Cunningham, S., Toogood, J., & Hawley, M. (2023). A qualitative study exploring the effect of communicating with partially intelligible speech. *AAC: Augmentative and Alternative Communication*. https://doi.org/10.1080/07434618.2023.2206910

Cleve, A., Devillers, D., Palladini, M., Paris, J., Michael, R., Faure, E., & Bonora R. (2020). *Detecting out-of-hospital cardiac arrest using artificial intelligence*.

Combe, N., Harrison, D., & Dong, H. (2013). LNCS 8014 - Designing Technology for Older People – The Role of Technical Self-confidence in Usability of an Inclusive Heating Control. In *LNCS* (Vol. 8014).

Communication Matters. (2020). *What is AAC?* https://www.communicationmatters.org.uk/what-is-aac/

Connaghan, K. P., Green, J. R., Paganoni, S., Chan, J., Weber, H., Collins, E., Richburg, B., Eshghi, M., Onnela, J. P., & Berry, J. D. (2019). Use of Beiwe smartphone app to identify and track speech decline in amyotrophic lateral sclerosis. *Proceedings of the Annual Conference of*

the International Speech Communication Association, INTERSPEECH, *2019-September*, 4504–4508. https://doi.org/10.21437/Interspeech.2019-3126

Costello, J. M. (2016). *Message Banking, Voice Banking and Legacy Messages*.

Curl, T. S. (2005). Practices in other-initiated repair resolution: The phonetic differentiation of "repetitions." In *Discourse Processes* (Vol. 39, Issue 1, pp. 1–43). Lawrence Erlbaum Associates Inc. https://doi.org/10.1207/s15326950dp3901_1

Da Costa Franceschini, A., & Mourão, L. F. (2015). Dysarthria and dysphagia in Amyotrophic Lateral Sclerosis with spinal onset: A study of quality of life related to swallowing [Article]. *NeuroRehabilitation (Reading, Mass.)*, *36*(1), 127–134. https://doi.org/10.3233/NRE-141200

Daw, D. (2011, October 24). *What Makes Siri Special?* . PCWorld. https://www.pcworld.com/article/477637/what_makes_siri_special_.html

De Russis, L., & Corno, F. (2019). On the impact of dysarthric speech on contemporary ASR cloud platforms. *Journal of Reliable Intelligent Environments*, *5*(3), 163–172. https://doi.org/10.1007/s40860-019-00085-y

DiChristofano, A., Shuster, H., Chandra, S., & Patwari, N. (2022). *Performance Disparities Between Accents in Automatic Speech Recognition*. http://arxiv.org/abs/2208.01157

Doshi, R., Chen, Y., Jiang, L., Zhang, X., Biadsy, F., Ramabhadran, B., Chu, F., Rosenberg, A., & Moreno, P. J. (2021). Extending Parrotron: An End-To-End, Speech Conversion and Speech Recognition Model for Atypical Speech. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2021-June*, 6988–6992. https://doi.org/10.1109/ICASSP39728.2021.9414644

Duffy, J. R. (2020). *Motor speech disorders : substrates, differential diagnosis, and management* (Fourth edition.) [Book]. Elsevier.

Duyx, B., Urlings, M. J. E., Swaen, G. M. H., Bouter, L. M., & Zeegers, M. P. (2017). Scientific citations favor positive results: a systematic review and meta-analysis. In *Journal of Clinical Epidemiology* (Vol. 88, pp. 92–101). Elsevier Inc. https://doi.org/10.1016/j.jclinepi.2017.06.002

ENDERBY, P. (1980). Frenchay Dysarthria Assessment. *International Journal of Language & Communication Disorders*, *15*(3), 165–173. https://doi.org/10.3109/13682828009112541

Enderby, P., & Palmer, R. (2008). *Frenchay dysarthria assessment–Second edition (FDA-2)*. Pro-ed.

Enfield, N. J., & Sidnell, J. (2017). Action from flexibility to accountability, across agents and scales, through contexts and time [Article]. *HAU Journal of Ethnographic Theory*, *7*(2), 443–450. https://doi.org/10.14318/hau7.2.036

Favre, B., Cheung, K., Kazemian, S., Lee, A., Liu, Y., Munteanu, C., Nenkova, A., Ochei, D., Penn, G., Tratz, S., Voss, C., & Zeller, F. (2013). *Automatic Human Utility Evaluation of ASR Systems: Does WER Really Predict Performance?*

Felgoise, S. H., Zaccheo, V., Duff, J., & Simmons, Z. (2016). Verbal communication impacts quality of life in patients with amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *17*(3–4), 179–183. https://doi.org/10.3109/21678421.2015.1125499

Fried-Oken, M., Kinsella, M., Stevens, I., & Klein, E. (2023). What stakeholders with neurodegenerative conditions value about speech and accuracy in development of BCI systems for communication. *Brain-Computer Interfaces*. https://doi.org/10.1080/2326263X.2023.2283345

Gee, J. P. (2010). *Introduction to Discourse Analysis* (3rd ed.) [Book]. Taylor & Francis Group. https://doi.org/10.4324/9780203005675

Geng, M., Xie, X., Ye, Z., Wang, T., Li, G., Hu, S., Liu, X., & Meng, H. (2022). *Speaker Adaptation Using Spectro-Temporal Deep Features for Dysarthric and Elderly Speech Recognition*. http://arxiv.org/abs/2202.10290

Gibbons, C. J., Thornton, E. W., Ealing, J., Shaw, P. J., Talbot, K., Tennant, A., & Young, C. A. (2013). Assessing social isolation in motor neurone disease: A Rasch analysis of the MND Social Withdrawal Scale. *Journal of the Neurological Sciences*, *334*(1–2), 112. https://doi.org/10.1016/J.JNS.2013.08.002

Gibbons, C. J., Thornton, E. W., & Young, C. A. (2013). The patient experience of fatigue in motor neurone disease. *Frontiers in Psychology*, *4*(October), 1–9. https://doi.org/10.3389/fpsyg.2013.00788

Glenn, P. (2009). Towards a social interactional approach to laughter. In *Laughter in Interaction* (pp. 7–34). Cambridge University Press. https://doi.org/10.1017/cbo9780511519888.003

Goffman, Erving. (1971). *Relations in public : microstudies of the public order* [Book]. Allen Lane.

Goffman, Erving. (1972). *Encounters : two studies in the sociology of interaction* [Book]. Penguin.

Goffman, Erving. (1981). *Forms of talk* [Book]. Blackwell.

Golson, J. (2011, October 4). *Siri Voice Recognition Arrives On the iPhone 4S*. MacRumors. https://www.macrumors.com/2011/10/04/siri-voice-recognition-arrives-on-the-iphone-4s/

Google. (2021). *Help center – Project Relate*. https://sites.research.google/relate/help/

Google - The Keyword. (2023, January 17). *AI is making the world a better place*. https://blog.google/technology/ai/7-ways-google-is-using-ai-to-help-solve-societys-challenges/

Google AI Blog. (2019). Project Euphonia's Personalized Speech Recognition for Non-Standard Speech. *Google AI Blog*.

Green, J. R., MacDonald, R. L., Jiang, P. P., Cattiau, J., Heywood, R., Cave, R., Seaver, K., Ladewig, M. A., Tobin, J., Brenner, M. P., Nelson, P. C., & Tomanek, K. (2021). Automatic speech recognition of disordered speech: Personalized models outperforming human listeners on short phrases. *Proceedings of the Annual Conference of the International Speech*

Communication Association, INTERSPEECH, *4*, 3051–3055. https://doi.org/10.21437/Interspeech.2021-1384

Gutz, S. E., Stipancic, K. L., Yunusova, Y., Berry, J. D., & Green, J. R. (2022). Validity of Off-the-Shelf Automatic Speech Recognition for Assessing Speech Intelligibility and Speech Severity in Speakers With Amyotrophic Lateral Sclerosis. *Journal of Speech, Language, and Hearing Research*, *65*(6), 2128–2143. https://doi.org/10.1044/2022_JSLHR-21-00589

Hahm, S., Heitzman, D., & Wang, J. (2015). Recognizing dysarthric speech due to amyotrophic lateral sclerosis with across-speaker articulatory normalization. *SLPAT 2015 - 6th Workshop on Speech and Language Processing for Assistive Technologies, Proceedings*, 47–54. https://doi.org/10.18653/v1/w15-5109

Hamari, J., Koivisto, J., & Sarsa, H. (2014). *Does Gamification Work?-A Literature Review of Empirical Studies on Gamification*.

Hartelius, L., Elmberg, M., Holm, R., Lövberg, A. S., & Nikolaidis, S. (2008). Living with dysarthria: Evaluation of a self-report questionnaire. *Folia Phoniatrica et Logopaedica*, *60*(1), 11–19. https://doi.org/10.1159/000111799

Harwell, D. (2018, July 19). *Why some accents don't work on Alexa or Google Home*. The Washington Post. https://www.washingtonpost.com/graphics/2018/business/alexa-does-not-understand-your-accent/

Hawley, M. S., Cunningham, S. P., Green, P. D., Enderby, P., Palmer, R., Sehgal, S., & O'Neill, P. (2013). A voice-input voice-output communication aid for people with severe speech impairment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *21*(1), 23–31. https://doi.org/10.1109/TNSRE.2012.2209678

Hawley, M. S., Enderby, P., Green, P., Cunningham, S., Brownsell, S., Carmichael, J., Parker, M., Hatzis, A., O'Neill, P., & Palmer, R. (2007). A speech-controlled environmental control system for people with severe dysarthria. *Medical Engineering and Physics*, *29*(5), 586–593. https://doi.org/10.1016/j.medengphy.2006.06.009

Hayes, H., Buckland, S., & Tarpey, M. (2012). *INVOLVE NIHR: Briefing notes for researchers: Public Involvement in NHS, public health and social care research*. http://www.invo.org.uk/wp-content/uploads/2014/11/9938_INVOLVE_Briefing_Notes_WEB.pdf

Hecht, M., Hillemacher, T., Gräsel, E., Tigges, S., Winterholler, M., Heuss, D., Hilz, M. J., & Neundörfer, B. (2002). Subjective experience and coping in ALS. In *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders* (Vol. 3, Issue 4, pp. 225–231). Martin Dunitz Ltd. https://doi.org/10.1080/146608202760839009

Heritage, J. (1998). Oh-prefaced responses to inquiry. In *Language in Society* (Vol. 27, Issue 3, pp. 291–334). Cambridge University Press. https://doi.org/10.1017/s0047404500019990

Heritage, J. (2010). A change-of-state token and aspects of its sequential placement. In *Structures of Social Action* (pp. 299–345). Cambridge University Press. https://doi.org/10.1017/cbo9780511665868.020

Higginbotham, D. J., Lesher, G. W., Moulton, B. J., & Roark, B. (2012). The application of natural language processing to augmentative and alternative communication. *Assistive Technology*, *24*(1), 14–24. https://doi.org/10.1080/10400435.2011.648714

Higginbotham, D. J., & Wilkins, D. P. (1999). Slipping Through the Timestream: Social Issues of Time and Timing in Augmented Interactions |. In M. Maxwell, D. Kovarsky, & J. F. Duchan (Eds.), *Constructing (in)competence* (pp. 55–88). Psychology Press. https://learning.oreilly.com/library/view/constructing-in-competence/9780805825909/xhtml/10_Chapter3.xhtml#Chapter3

Higginbotham, J., & Wilkins, D. (1999). Slipping Through the Timestream: Social Issues of Time and Timing in Augmented Interactions [Bookitem]. In *Constructing (in)competence* (pp. 55–88). Psychology Press. https://doi.org/10.4324/9780203763759-8

Hirano, Y. M., Yamazaki, Y., Shimizu, J., Togari, T., & Bryce, T. J. (2006). Ventilator dependence and expressions of need: A study of patients with amyotrophic lateral sclerosis in Japan. *Social Science and Medicine*, *62*(6), 1403–1413. https://doi.org/10.1016/j.socscimed.2005.08.015

Hird, K., & Hennessey, N. (2007). Facilitating use of speech recognition software for people with disabilities: A comparison of three treatments. *Clinical Linguistics and Phonetics*, *21*(3), 211–226. https://doi.org/10.1080/02699200601100249

Holloway, C. (2021). *Disability Interactions: Creating Inclusive Innovations* (1st ed.) [Book]. Springer Nature.

Hosom, J. P., Jakobs, T., Baker, A., & Fager, S. (2010). Automatic speech recognition for assistive writing in Speech Supplemented Word Prediction. *Proceedings of the 11th Annual Conference of the International Speech Communication Association, INTERSPEECH 2010*, 2674–2677. https://doi.org/10.21437/interspeech.2010-709

Hu, S., Liu, S., Xie, X., Geng, M., Wang, T., Hu, S., Cui, M., Liu, X., & Meng, H. (2022). *Exploiting Cross Domain Acoustic-to-articulatory Inverted Features For Disordered Speech Recognition*. http://arxiv.org/abs/2203.10274

Humza Naveed, Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2024). A Comprehensive Overview of Large Language Models [Article]. *ArXiv.Org*. https://doi.org/10.48550/arxiv.2307.06435

International Alliance of ALS/MND Associations. (2021). *What is ALS/MND?* https://www.als-mnd.org/what-is-alsmnd/

ISAAC. (2022). *ISAAC – What is AAC?* International Society for Augmentative and Alternative Communication. https://isaac-online.org/english/what-is-aac/

Jefferson, G. (1986). Notes on “Latency” in Overlap Onset. In *Source: Human Studies* (Vol. 9, Issue 2). https://about.jstor.org/terms

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation Analysis: Studies from the first generation* (pp. 13–31). Benjamins Publishing Company.

Jefferson, G., Sacks, H., & Schegloff, E. (1987). 6 Notes on Laughter in the Pursuit of Intimacy [Bookitem]. In *Talk and Social Organisation*. Multilingual Matters. https://doi.org/10.21832/9781800418226-008

Joshy, A. A., & Rajan, R. (2022). Automated Dysarthria Severity Classification: A Study on Acoustic Features and Deep Learning Techniques. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *30*, 1147–1157. https://doi.org/10.1109/TNSRE.2022.3169814

Judge, S., & Townend, G. (2013). Perceptions of the design of voice output communication aids. *International Journal of Language and Communication Disorders*, *48*(4), 366–381. https://doi.org/10.1111/1460-6984.12012

Kane, S. K., Morris, M. R., Paradiso, A., & Campbell, J. (2017). "at times avuncular and cantankerous, with the reflexes ol a mongoose": Understanding self-expression through augmentative and alternative communication devices. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 1166–1179. https://doi.org/10.1145/2998181.2998284

Kelly, R., Gooch, D., Patil, B., & Watts, L. (2017). Demanding by design: Supporting effortful communication practices in close personal relationships. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 70–83. https://doi.org/10.1145/2998181.2998184

Kiesling, S. F. (2022). Stance and Stancetaking. *Https://Doi.Org/10.1146/Annurev-Linguistics-031120-121256*, *8*, 409–426. https://doi.org/10.1146/ANNUREV-LINGUISTICS-031120-121256

Kim, M., Cao, B., An, K., & Wang, J. (2018). Dysarthric speech recognition using convolutional LSTM neural network. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2018-September*, 2948–2952. https://doi.org/10.21437/Interspeech.2018-2250

Kitzinger, C. (2012). Repair [Bookitem]. In *The Handbook of Conversation Analysis* (pp. 229–256). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118325001.ch12

Kitzinger, C., & Mandelbaum, J. (2013). Word Selection and Social Identities in Talk-in-Interaction. *Communication Monographs*, *80*(2), 176–198. https://doi.org/10.1080/03637751.2013.776171

Koch Fager, S., Fried-Oken, M., Jakobs, T., & Beukelman, D. R. (2019). New and emerging access technologies for adults with complex communication needs and severe motor impairments: State of the science. *AAC: Augmentative and Alternative Communication*, *35*(1), 13–25. https://doi.org/10.1080/07434618.2018.1556730

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J. R., Jurafsky, D., & Goel, S. (2020). Racial disparities in automated speech recognition [Article]. *Proceedings of the National Academy of Sciences - PNAS*, *117*(14), 7684–7689. https://doi.org/10.1073/pnas.1915768117

Kotler, A. L., & Thomas-Stonell, N. (1997). Effects of speech training on the accuracy of speech recognition for an individual with a speech impairment. *AAC: Augmentative and Alternative Communication*, *13*(2), 71–80. https://doi.org/10.1080/07434619712331277858

Kristensson, P. O., Lilley, J., Black, R., & Waller, A. (2020, April 21). A Design Engineering Approach for Quantitatively Exploring Context-Aware Sentence Retrieval for Nonspeaking Individuals with Motor Disabilities. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3313831.3376525

Kumar, C. V. T., Bhattacharjee, T., Belur, Y., Nalini, A., Yadav, R., & Ghosh, P. K. (2023). Classification of multi-class vowels and fricatives from patients having Amyotrophic Lateral Sclerosis with varied levels of dysarthria severity. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2023-August*, 146–150. https://doi.org/10.21437/Interspeech.2023-1455

Kylmä, J., Duggleby, W., Cooper, D., & Molander, G. (2009). Hope in palliative care: An integrative review. *Palliative and Supportive Care*, *7*(3), 365–377. https://doi.org/10.1017/S1478951509990307

Laricchia, F. (2020). *Amazon Alexa: skill growth 2016-2019 | Statista*. https://www.statista.com/statistics/912856/amazon-alexa-skills-growth/

Leite Neto, L., & Constantini, A. C. (2017). Dysarthria and quality of life in patients with amyotrophic lateral sclerosis. *Revista CEFAC*, *19*(5), 664–673. https://doi.org/10.1590/1982-021620171954017

Lerner, G., & Kitzinger, C. (2012). Research at the Intersection of Reference and Repair: Introduction to the Special Issue. *Research on Language and Social Interaction - RES LANG SOC INTERACT*, *45*, 111–115. https://doi.org/10.1080/08351813.2012.673371

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science*, *5*(1). https://doi.org/10.1186/1748-5908-5-69

Li, J., Bzdok, D., Chen, J., Tam, A., Ooi, L. Q. R., Holmes, A. J., Ge, T., Patil, K. R., Jabbi, M., Eickhoff, S. B., Yeo, B. T. T., & Genon, S. (2022). Cross-ethnicity/race generalization failure of behavioral prediction from resting-state functional connectivity [Article]. *Science Advances*, *8*(11), eabj1812–eabj1812. https://doi.org/10.1126/sciadv.abj1812

Light, J., & McNaughton, D. (2014). Communicative competence for individuals who require augmentative and alternative communication: A new definition for a new era of communication? In *AAC: Augmentative and Alternative Communication* (Vol. 30, Issue 1, pp. 1–18). https://doi.org/10.3109/07434618.2014.885080

Light, J., McNaughton, D., Beukelman, D., Fager, S. K., Fried-Oken, M., Jakobs, T., & Jakobs, E. (2019). Challenges and opportunities in augmentative and alternative communication: Research and technology development to enhance communication and participation for individuals with complex communication needs. In *AAC: Augmentative and Alternative Communication* (Vol. 35, Issue 1, pp. 1–12). Taylor and Francis Ltd. https://doi.org/10.1080/07434618.2018.1556732

Lin, Y., Wang, L., Dang, J., Li, S., & Ding, C. (2023). Disordered speech recognition considering low resources and abnormal articulation. *Speech Communication*, *155*. https://doi.org/10.1016/j.specom.2023.103002

Linse, K., Aust, E., Joos, M., & Hermann, A. (2018). Communication matters-pitfalls and promise of hightech communication devices in palliative care of severely physically disabled patients with amyotrophic lateral sclerosis. In *Frontiers in Neurology* (Vol. 9, Issue JUL). Frontiers Media S.A. https://doi.org/10.3389/fneur.2018.00603

Littlejohn Shinder, D. (2007, March 8). *Speech recognition in Windows Vista*. TechRepublic. https://www.techrepublic.com/article/speech-recognition-in-windows-vista/

MacDonald, R. L., Jiang, P. P., Cattiau, J., Heywood, R., Cave, R., Seaver, K., Ladewig, M., Tobin, J., Brenner, M. P., Nelson, P. C., Green, J. R., & Tomanek, K. (2021). Disordered speech data collection: Lessons learned at 1 million utterances from Project Euphonia. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *4*, 3066–3070. https://doi.org/10.21437/Interspeech.2021-697

Makkonen, T., Ruottinen, H., Puhto, R., Helminen, M., & Palmio, J. (2018). Speech deterioration in amyotrophic lateral sclerosis (ALS) after manifestation of bulbar symptoms. *International Journal of Language and Communication Disorders*, *53*(2), 385–392. https://doi.org/10.1111/1460-6984.12357

Mandelbaum, J. (2012). Storytelling in Conversation [Bookitem]. In *The Handbook of Conversation Analysis* (pp. 492–507). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118325001.ch24

Mariën, P., Keulen, S., & Verhoeven, J. (2019). Neurological Aspects of Foreign Accent Syndrome in Stroke Patients. In *Journal of Communication Disorders* (Vol. 77, pp. 94–113). Elsevier Inc. https://doi.org/10.1016/j.jcomdis.2018.12.002

Marini, M., Vigano, M., Corbo, M., Zettin, M., Simoncini, G., Fattori, B., D'Anna, C., Donati, M., & Fanucci, L. (2021). IDEA: An Italian Dysarthric Speech Database. *2021 IEEE Spoken Language Technology Workshop, SLT 2021 - Proceedings*, 1086–1093. https://doi.org/10.1109/SLT48900.2021.9383467

Maynard, D. W. (1997). The News Delivery Sequence: Bad News and Good News in Conversational Interaction [Article]. *Research on Language and Social Interaction*, *30*(2), 93–130. https://doi.org/10.1207/s15327973rlsi3002_1

McIlvenny, P., & Davidsen, J. (2022). *Distributed Open Transcription Environment (DOTE)* (1). Aalborg University.

Mckelvey, M., Evans, D. L., Kawai, N., & Beukelman, D. (2012). Communication styles of persons with ALS as recounted by surviving partners. *AAC: Augmentative and Alternative Communication*, *28*(4), 232–242. https://doi.org/10.3109/07434618.2012.737023

McNaughton, D., Light, J., & Groszyk, L. (2001). "Don't give up": Employment experiences of individuals with amyotrophic lateral sclerosis who use augmentative and alternative communication. *Augmentative and Alternative Communication*, *17*(3), 179–195. https://doi.org/10.1080/aac.17.3.179.195

Mendoza, V., Ac, R., Vasquez-Correa, J. C., Nöth, E., De Bodt Ac, M., Van, G., & Acd, N. (2022). *Automatic score of articulatory distortion in adults with dysarthria*. https://ssrn.com/abstract=4201851

Mengesha, Z., Heldreth, C., Lahav, M., Sublewski, J., & Tuennerman, E. (2021). "I don't Think These Devices are Very Culturally Sensitive."—Impact of Automated Speech Recognition Errors on African Americans [Article]. *Frontiers in Artificial Intelligence*, *4*, 725911–725911. https://doi.org/10.3389/frai.2021.725911

Meyer, J., Rauchenstein, L., Eisenberg, J. D., & Howell, N. (2020). *Artie Bias Corpus: An Open Dataset for Detecting Demographic Bias in Speech Applications* (pp. 6462–6468). https://aclanthology.org/2020.lrec-1.796

Miller, N., Taylor, J., Howe, C., & Read, J. (2011). Living with foreign accent syndrome: Insider perspectives. *Aphasiology*, *25*(9), 1053–1068. https://doi.org/10.1080/02687038.2011.573857

Mishra, T., Ljolje, A., & Gilbert, M. (2011). Predicting Human Perceived Accuracy of ASR systems. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 1945–1948. https://doi.org/10.21437/interspeech.2011-364

Mitchell, C., Bowen, A., Tyson, S., Butterfint, Z., & Conroy, P. (2017). Interventions for dysarthria due to stroke and other adult-acquired, non-progressive brain injury [Article]. *Cochrane Database of Systematic Reviews*, *2017*(1), CD002088–CD002088. https://doi.org/10.1002/14651858.CD002088.pub3

Mlinarić, A., Horvat, M., & Smolčić, V. Š. (2017). Dealing with the positive publication bias: Why you should really publish your negative results. In *Biochemia Medica* (Vol. 27, Issue 3). Biochemia Medica, Editorial Office. https://doi.org/10.11613/BM.2017.030201

Moar, J., & Escherich, M. (2020). *Hey Siri, how will you make money?* https://www.juniperresearch.com/press/number-of-voice-assistant-devices-in-use

Moore, M., Venkateswara, H., & Panchanathan, S. (2018). Whistle-blowing ASRs: Evaluating the need for more inclusive automatic speech recognition systems. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2018-September*, 466–470. https://doi.org/10.21437/Interspeech.2018-2391

Morris, M. A., Dudgeon, B. J., & Yorkston, K. (2013). A qualitative study of adult AAC users' experiences communicating with medical providers. *Disability and Rehabilitation: Assistive Technology*, *8*(6), 472–481. https://doi.org/10.3109/17483107.2012.746398

Motor neurone disease: assessment and Motor neurone disease: assessment and management management NICE guideline. (2016).

Mozilla. (2021). *Common Voice*. https://commonvoice.mozilla.org/en

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. In *Int. J. Man-Machine Studies* (Vol. 27).

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, *39*(3), 429–460. https://doi.org/10.1080/00140139608964474

Mulfari, D., Carnevale, L., & Villari, M. (2023). Toward a lightweight ASR solution for atypical speech on the edge. *Future Generation Computer Systems*, *149*, 455–463. https://doi.org/10.1016/j.future.2023.08.002

Mulfari, D., La Placa, D., Rovito, C., Celesti, A., & Villari, M. (2022). Deep learning applications in telerehabilitation speech therapy scenarios. *Computers in Biology and Medicine*, *148*. https://doi.org/10.1016/j.compbiomed.2022.105864

Müller, J., Wenning, G. K., Verny, M., McKee, A., Chaudhuri, K. R., Jellinger, K., Poewe, W., & Litvan, I. (2001). Progression of Dysarthria and Dysphagia in Postmortem-Confirmed Parkinsonian Disorders [Article]. *Archives of Neurology (Chicago)*, *58*(2), 259–264. https://doi.org/10.1001/archneur.58.2.259

Murphy, J. (2004). "I prefer contact this close": Perceptions of AAC by people with motor neurone disease and their communication partners. *AAC: Augmentative and Alternative Communication*, *20*(4), 259–271. https://doi.org/10.1080/07434610400005663

Naeini, S. A., Simmatis, L., Jafari, D., Yunusova, Y., & Taati, B. (2024). Improving Dysarthric Speech Segmentation With Emulated and Synthetic Augmentation. *IEEE Journal of Translational Engineering in Health and Medicine*, *12*, 382–389. https://doi.org/10.1109/JTEHM.2024.3375323

Nathanson, E. (2017). Native voice, self-concept and the moral case for personalized voice technology. In *Disability and Rehabilitation* (Vol. 39, Issue 1, pp. 73–81). Taylor and Francis Ltd. https://doi.org/10.3109/09638288.2016.1139193

Neuburg, M. (2000, August 28). *Bossing Your Mac with PlainTalk - TidBITS*. TidBITS. https://tidbits.com/2000/08/28/bossing-your-mac-with-plaintalk/

Nevo, D., & Wade, M. (2007). How to avoid disappointment by design [Article]. In *Communications of the ACM* (Vol. 50, Issue 4, pp. 43–48). ACM. https://doi.org/10.1145/1232743.1232746

Newswire. (2021, July 29). *Mind the (Accent) Gap: DefinedCrowd Contributing to More Inclusive Speech Technology | Newswire*. Newswire. https://www.newswire.com/news/mind-the-accent-gap-definedcrowd-contributing-to-more-inclusive-speech-21458402

Norén, N., Melander Bowden, H., & Evaldsson, A. C. (2022). Young students' treatment of synthetic voicing as an interactional resource in digital writing. *Classroom Discourse*, *13*(3), 241–263. https://doi.org/10.1080/19463014.2020.1814367

Norman, D. (2013). *The Design of Everyday Things* (1st ed.) [Book]. Basic Books.

Office of the Public Guardian, & Ministry of Justice. (2014). *Mental Capacity Act: making decisions - GOV.UK*. https://www.gov.uk/government/collections/mental-capacity-act-making-decisions

Okafor, O., & Ludi, S. (2022, October 22). Voice-Enabled Blockly: Usability Impressions of a Speech-driven Block-based Programming System. *ASSETS 2022 - Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. https://doi.org/10.1145/3517428.3550382

Oliver, D., Radunovic, A., Allen, A., & McDermott, C. (2017). The development of the UK National Institute of Health and Care Excellence evidence-based clinical guidelines on motor neurone disease. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *18*(5–6), 313–323. https://doi.org/10.1080/21678421.2017.1304558

Olmstead, K. (2017). *Voice assistants used by 46% of Americans, mostly on smartphones | Pew Research Center*. https://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/

Overton, K., Professor, M., Bourgeois, M., Barker, M. R., & Betancourt, K. (2017). *Perceptual Differences between Natural Speech and Personalized Synthetic Speech*.

Parker, M., Cunningham, S., Enderby, P., Hawley, M., & Green, P. (2006). Automatic speech recognition and training for severely dysarthric users of assistive technology: The STARDUST project. *Clinical Linguistics & Phonetics*, *20*(2–3), 149–156. https://doi.org/10.1080/02699200400026884

Peters, B., Bieker, G., Heckman, S. M., Huggins, J. E., Wolf, C., Zeitlin, D., & Fried-Oken, M. (2015). Brain-computer interface users speak Up: The virtual users' forum at the 2013 international brain-computer interface meeting. In *Archives of Physical Medicine and Rehabilitation* (Vol. 96, Issue 3, pp. S33–S37). W.B. Saunders. https://doi.org/10.1016/j.apmr.2014.03.037

Peters, M. D., Godfrey, C. M., & McInerney, P. (2015). *The Joanna Briggs Institute Reviewers'. Manual 2015: Methodology for JBI scoping reviews*. http://joannabriggs.org/assets/docs/sumari/ReviewersManual_Mixed-Methods-Review-Methods-2014-ch1.pdf.

Philip, K., Peters, B., Hicks, M., & Mullaney, T. S. (2021). Siri Disciplines [Bookitem]. In *Your Computer Is on Fire*. The MIT Press. https://doi.org/10.7551/mitpress/10993.003.0013

Pinola, M. (2011). *Speech Recognition Through the Decades: How We Ended Up With Siri | PCWorld*. https://www.pcworld.com/article/477914/speech_recognition_through_the_decades_how_we_ended_up_with_siri.html

Pinto-Grau, M., Hardiman, O., & Pender, N. (2018). The Study of Language in the Amyotrophic Lateral Sclerosis - Frontotemporal Spectrum Disorder: a Systematic Review of Findings and New Perspectives. In *Neuropsychology Review* (Vol. 28, Issue 2, pp. 251–268). Springer New York LLC. https://doi.org/10.1007/s11065-018-9375-7

Pomerantz, A. (2021). Attributions of Responsibility: Blamings [Bookitem]. In *Asking and Telling in Conversation*. Oxford University Press. https://doi.org/10.1093/oso/9780190927431.003.0010

Pradhan, A., Mehta, K., & Findlater, L. (2018). "Accessibility came by accident": Use of voice-controlled intelligent personal assistants by people with disabilities. *Conference on Human*

Factors in Computing Systems - Proceedings, *2018-April*. https://doi.org/10.1145/3173574.3174033

Pullin, G., & Hennig, S. (2015). 17 ways to say yes: Toward nuanced tone of voice in AAC and speech technology. In *AAC: Augmentative and Alternative Communication* (Vol. 31, Issue 2, pp. 170–180). Informa Healthcare. https://doi.org/10.3109/07434618.2015.1037930

Pullin, G., Treviranus, J., Patel, R., & Higginbotham, J. (2017). Designing interaction, voice, and inclusion in AAC research. *AAC: Augmentative and Alternative Communication*, *33*(3), 139–148. https://doi.org/10.1080/07434618.2017.1342690

Radzikowski, K., Wang, L., Yoshie, O., & Nowak, R. (2021). Accent modification for speech recognition of non-native speakers using neural style transfer [Article]. *EURASIP Journal on Audio, Speech, and Music Processing*, *2021*(1), 1–10. https://doi.org/10.1186/s13636-021-00199-3

Ritter, F. E., Baxter, G. D., & Churchill, E. F. (2014). Introducing User-Centered Systems Design. In *Foundations for Designing User-Centered Systems* (pp. 3–31). Springer London. https://doi.org/10.1007/978-1-4471-5134-0_1

Robillard, A. B. (1994). *Communication problems in the intensive care unit* [Book].

Robillard, A. B. (1999). *Meaning of a disability : the lived experience of paralysis* [Book]. Temple University Press.

Robinson, J. D. (2004). The sequential organization of "explicit" apologies in naturally occurring english. In *Research on Language and Social Interaction* (Vol. 37, Issue 3, pp. 291–330). Lawrence Erlbaum Associates Inc. https://doi.org/10.1207/s15327973rlsi3703_2

Robinson, J. D. (2006). Managing trouble responsibility and relationships during conversational repair. *Communication Monographs*, *73*(2), 137–161. https://doi.org/10.1080/03637750600581206

Robinson, J. D., & Kevoe-Feldman, H. (2010). Using full repeats to initiate repair on others' questions. *Research on Language and Social Interaction*, *43*(3), 232–259. https://doi.org/10.1080/08351813.2010.497990

Rogers, Y. (2011). Interaction design gone wild: Striving for wild theory [Article]. In *Interactions (New York, N.Y.)* (Vol. 18, Issue 4, pp. 58–62). Association for Computing Machinery. https://doi.org/10.1145/1978822.1978834

Rogers, Y., Sharp, H., & Preece, J. (2023). *Interaction Design: Beyond Human-Computer Interaction* (6th ed.) [Book]. John Wiley & Sons, Incorporated.

Rong, P. (2021). A novel hierarchical framework for measuring the complexity and irregularity of multimodal speech signals and its application in the assessment of speech impairment in amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research*, *64*(8), 2996–3014. https://doi.org/10.1044/2021_JSLHR-20-00743

Rossano, F. (2012). Gaze in Conversation [Bookitem]. In *The Handbook of Conversation Analysis* (pp. 308–329). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118325001.ch15

Rudzicz, F. (2011). Articulatory knowledge in the recognition of dysarthric speech. *IEEE Transactions on Audio, Speech and Language Processing*, *19*(4), 947–960. https://doi.org/10.1109/TASL.2010.2072499

Rudzicz, F. (2013). Adjusting dysarthric speech signals to be more intelligible. *Computer Speech and Language*, *27*(6), 1163–1177. https://doi.org/10.1016/j.csl.2012.11.001

Russell, S. J. (Stuart J. (2022). *Artificial intelligence : a modern approach* (P. Norvig, M.-W. Chang, J. Devlin, A. Dragan, D. Forsyth, I. Goodfellow, J. Malik, V. Mansinghka, J. Pearl, & M. J. Wooldridge, Eds.; Fourth edition.) [Book]. Pearson.

Rutter, B. (2009). Repair sequences in dysarthric conversational speech: A study in interactional phonetics. *Clinical Linguistics and Phonetics*, *23*(12), 887–900. https://doi.org/10.3109/02699200903062303

Ruusuvuori, J. (2012). Emotion, Affect and Conversation [Bookitem]. In *The Handbook of Conversation Analysis* (pp. 330–349). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118325001.ch16

Sacks H. (1995). *Lectures on Conversation* [Book]. Wiley-Blackwell. https://doi.org/10.1002/9781444328301

Sacks, H. (2013). Doing "Being Ordinary" [Article]. *Sotsiologicheski problemi*, *3–4*, 7–15.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974a). A Simplest Systematics for the Organization of Turn-Taking for Conversation. In *Source: Language* (Vol. 50, Issue 4).

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974b). A Simplest Systematics for the Organization of Turn-Taking for Conversation [Article]. *Language (Baltimore)*, *50*(4), 696–735. https://doi.org/10.1353/lan.1974.0010

Safaz, İ., Kesikburun, S., Adigüzel, E., & Yilmaz, B. (2016). Determinants of disease-specific health-related quality of life in Turkish stroke survivors [Article]. *International Journal of Rehabilitation Research*, *39*(2), 130–133. https://doi.org/10.1097/MRR.0000000000000156

Sanders, E., Ruiter, M., Beijer Hogeschool Arnhem, L., Strik, H., & Beijer, L. (2002). *Automatic Recognition of Dutch Dysarthric Speech: A Pilot Study GOBL: Games Online for Basic Language Learning View project TAPAS (Training Network on Automatic Processing of Pathological Speech) View project AUTOMATIC RECOGNITION OF DUTCH DYSARTHRIC SPEECH A PILOT STUDY*. https://www.researchgate.net/publication/221486790

Schegloff, E. (1989). Reflections on language, development, and the interactional character of talk-in-interaction. In M. H. Bornstein & J. S. Bruner (Eds.), *Interaction in Human Development* (pp. 139–153). Lawrence Erlbaum Associates.

Schegloff, E. (2002). *Complex Sentences in Grammar and Discourse : Essays in honor of Sandra A. Thompson*. (J. Bybee, Ed.; 1st ed.) [Book]. John Benjamins Publishing Company.

Schegloff, E. A. (1977). *The preference for self-correction in the organization of repair in conversation* (G. Jefferson & Harvey. Sacks, Eds.) [Book].

Schegloff, E. A. (1987). Analyzing Single Episodes of Interaction: An Exercise in Conversation Analysis [Article]. *Social Psychology Quarterly*, *50*(2), 101–114. https://doi.org/10.2307/2786745

Schegloff, E. A. (1992). Repair After Next Turn: The Last Structurally Provided Defense of Intersubjectivity in Conversation. In *Source: American Journal of Sociology* (Vol. 97, Issue 5). https://about.jstor.org/terms

Schegloff, E. A. (2004a). *Lectures on Repair and Categorization (244b/266 Conversational Structures II), UCLA*.

Schegloff, E. A. (2004b). On dispensability. In *Research on Language and Social Interaction* (Vol. 37, Issue 2, pp. 95–149). Routledge. https://doi.org/10.1207/s15327973rlsi3702_2

Schegloff, E. A. (2011). Ten operations in self-initiated, same-turn repair. In *Conversational Repair and Human Understanding* (pp. 41–70). Cambridge University Press. https://doi.org/10.1017/CBO9780511757464.002

Schegloff, E. A. (2020). The Relevance of Repair to Syntax-for-Conversation. In *Discourse and Syntax* (pp. 261–286). BRILL. https://doi.org/10.1163/9789004368897_012

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). *The Preference for Self-Correction in the Organization of Repair in Conversation* (Vol. 53, Issue 2).

Schegloff, E., & Sacks, H. (1973). Opening up Closings [Article]. *Semiotica*, *8*(4), 289–327. https://doi.org/10.1515/semi.1973.8.4.289

Seale, J. M., Bisantz, A. M., & Higginbotham, J. (2020). Interaction symmetry: Assessing augmented speaker and oral speaker performances across four tasks. *AAC: Augmentative and Alternative Communication*, *36*(2), 82–94. https://doi.org/10.1080/07434618.2020.1782987

Sennott, S. C., Akagi, L., Lee, M., & Rhodes, A. (2019). AAC and Artificial Intelligence (AI). *Topics in Language Disorders*, *39*(4), 389–403. https://doi.org/10.1097/TLD.0000000000000197

Seo, M. S., & Koshik, I. (2010). A conversation analytic study of gestures that engender repair in ESL conversational tutoring. *Journal of Pragmatics*, *42*(8), 2219–2239. https://doi.org/10.1016/J.PRAGMA.2010.01.021

Shadden, B. B. (2005). Aphasia as identity theft: Theory and practice. In *Aphasiology* (Vol. 19, Issues 3–5, pp. 211–223). https://doi.org/10.1080/02687930444000697

Shor, J., Emanuel, D., Lang, O., Tuval, O., Brenner, M., Cattiau, J., Vieira, F., McNally, M., Charbonneau, T., Nollstadt, M., Hassidim, A., & Matias, Y. (2019). *Personalizing ASR for Dysarthric and Accented Speech with Limited Data*. https://doi.org/10.21437/Interspeech.2019-1427

Simpson, R. C., LoPresti, E. F., & Cooper, R. A. (2008). How many people would benefit from a smart wheelchair? *Journal of Rehabilitation Research and Development*, *45*(1), 53–72. https://doi.org/10.1682/JRRD.2007.01.0015

Smith, M. M., & Connolly, I. (2008). Roles of aided communication: Perspectives of adults who use AAC. *Disability and Rehabilitation: Assistive Technology*, *3*(5), 260–273. https://doi.org/10.1080/17483100802338499

Smith, M., & Murray, J. (2016). *The silent partner? : language, interaction and aided communication* (M. Smith & J. Murray, Eds.) [Book]. J & R Press Ltd.

Soleymanpour, M., Johnson, M. T., & Berry, J. (2021). Increasing the Precision of Dysarthric Speech Intelligibility and Severity Level Estimate. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12997 LNAI*, 670–679. https://doi.org/10.1007/978-3-030-87802-3_60

Stegmann, G. M., Hahn, S., Liss, J., Shefner, J., Rutkove, S., Shelton, K., Duncan, C. J., & Berisha, V. (2020). Early detection and tracking of bulbar changes in ALS via frequent and remote speech analysis. *Npj Digital Medicine*, *3*(1). https://doi.org/10.1038/s41746-020-00335-x

Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction*, *41*(1), 31–57. https://doi.org/10.1080/08351810701691123

Synnaeve, G., Xu, Q., Kahn, J., Likhomanenko, T., Grave, E., Pratap, V., Sriram, A., Liptchinsky, V., & Collobert, R. (2020). *END-TO-END ASR: FROM SUPERVISED TO SEMI-SUPERVISED LEARNING WITH MODERN ARCHITECTURES A PREPRINT*. https://github.com/google/sentencepiece

Talbot, K., Turner, M. R., Marsden, R., & Botell, R. (2012). *Motor Neuron Disease* (Vol. 1). Oxford University Press. https://doi.org/10.1093/med/9780199547364.001.1

Tatman, R. (2017). Gender and Dialect Bias in YouTube's Automatic Captions. *EACL 2017 - Ethics in Natural Language Processing, Proceedings of the 1st ACL Workshop*, 53–59. https://doi.org/10.18653/V1/W17-1606

Tatman, R. (2021). *Sociolinguistic Variation and Automatic Speech Recognition: Challenges and Approaches*.

Tena, A., Claria, F., Solsona, F., Meister, E., & Povedano, M. (2021). Detection of bulbar involvement in patients with amyotrophic lateral sclerosis by machine learning voice analysis: diagnostic decision support development study. *JMIR Medical Informatics*, *9*(3). https://doi.org/10.2196/21331

Terry, G., Hayfield, N., Clarke, V., & Braun, V. (2017). Thematic Analysis. In C. Willig & W. S. Rogers (Eds.), *The SAGE Handbook of Qualitative Research in Psychology*. SAGE Publications Ltd. https://doi.org/10.4135/9781526405555

Tobin, J., Li, Q., Venugopalan, S., Seaver, K., Cave, R., & Tomanek, K. (2022). *Assessing ASR Model Quality on Disordered Speech using BERTScore*. http://arxiv.org/abs/2209.10591

Tobin, J., & Tomanek, K. (2022). PERSONALIZED AUTOMATIC SPEECH RECOGNITION TRAINED ON SMALL DISORDERED SPEECH DATASETS. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2022-May*, 6637–6641. https://doi.org/10.1109/ICASSP43922.2022.9747516

Tomanek, K., Seaver, K., Jiang, P.-P., Cave, R., Harrell, L., & Green, J. R. (2023). *An Analysis of Degenerating Speech Due to Progressive Dysarthria on ASR Performance*. 1–5. https://doi.org/10.1109/icassp49357.2023.10097195

Tomanek, K., Tobin, J., Venugopalan, S., Cave, R., Seaver, K., Green, J. R., & Heywood, R. (2024). *Large Language Models As A Proxy For Human Evaluation In Assessing The Comprehensibility Of Disordered Speech Transcription*. 10846–10850. https://doi.org/10.1109/icassp48485.2024.10447177

Topaz, M., Schaffer, A., Lai, K. H., Korach, Z. T., Einbinder, J., & Zhou, L. (2018). Medical Malpractice Trends: Errors in Automated Speech Recognition [Article]. *Journal of Medical Systems*, *42*(8), 153–154. https://doi.org/10.1007/s10916-018-1011-9

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., ... Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. In *Annals of Internal Medicine* (Vol. 169, Issue 7, pp. 467–473). American College of Physicians. https://doi.org/10.7326/M18-0850

United Nations. (2016). *Leaving no one behind: the imperative of inclusive development Report on the World Social Situation 2016*.

Valencia, S., Cave, R., Kallarackal, K., Seaver, K., Terry, M., & Kane, S. K. (2023, April 19). "The less I type, the better": How AI Language Models can Enhance or Impede Communication for AAC Users. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3544548.3581560

Venugopalan, S., Shor, J., Plakal, M., Tobin, J., Tomanek, K., Green, J. R., & Brenner, M. P. (2021). *Comparing Supervised Models And Learned Speech Representations For Classifying Intelligibility Of Disordered Speech On Selected Phrases*. http://arxiv.org/abs/2107.03985

Venugopalan, S., Tobin, J., Yang, S. J., Seaver, K., Cave, R. J. N., Jiang, P. P., Zeghidour, N., Heywood, R., Green, J., & Brenner, M. P. (2023). Speech Intelligibility Classifiers from 550k Disordered Speech Samples. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2023-June*. https://doi.org/10.1109/ICASSP49357.2023.10095933

Vieira, H., Costa, N., Sousa, T., Reis, S., & Coelho, L. (2020). Voice-Based Classification of Amyotrophic Lateral Sclerosis: Where Are We and Where Are We Going? A Systematic Review. In *Neurodegenerative Diseases* (Vol. 19, Issues 5–6, pp. 163–170). S. Karger AG. https://doi.org/10.1159/000506259

VoiceITT. (n.d.). *VoiceItt*. Retrieved October 4, 2023, from https://www.voiceitt.com/

Wallace, L. (2016, October 5). *The future is AI, and Google just showed Apple how it's done*. Cult Of Mac. https://www.cultofmac.com/447898/google-home-google-assistant-siri-ai/

Waller, A. (2019). Telling tales: unlocking the potential of AAC technologies. In *International Journal of Language and Communication Disorders* (Vol. 54, Issue 2, pp. 159–169). Wiley Blackwell. https://doi.org/10.1111/1460-6984.12449

Wallock, K. E., & Cerny, S. L. (2021). *Voices from the Field Benefits of Smart Home Technology for Individuals Living with Amyotrophic Lateral Sclerosis* (Vol. 15). www.atia.org/atob

Wang, J., Kothalkar, P. V., Cao, B., & Heitzman, D. (2016). Towards automatic detection of amyotrophic lateral sclerosis from speech acoustic and articulatory samples. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *08-12-September-2016*, 1195–1199. https://doi.org/10.21437/Interspeech.2016-1542

Warren, T. (2014, April 2). *The story of Cortana, Microsoft's Siri killer*. The Verge. https://www.theverge.com/2014/4/2/5570866/cortana-windows-phone-8-1-digital-assistant

Wertsch, J. V. (1993). Voices of the Mind [Book]. In *Cambridge, MA: Harvard U Press, 1991. viii+169 pp*. Harvard University Press. https://doi.org/10.2307/j.ctv1pncrnd

Wiepert, D. A., Utianski, R. L., Duffy, J. R., Stricker, J. L., Barnard, L., Josephs, K. A., Whitwell, J. L., Jones, D. T., & Botha, H. (2023). Not All Errors Are Created Equal: Evaluating The Impact of Model and Speaker Factors on ASR Outcomes in Clinical Populations. *2023 IEEE Automatic Speech Recognition and Understanding Workshop, ASRU 2023*. https://doi.org/10.1109/ASRU57964.2023.10389723

Wilkinson, R., Bloch, S., & Clarke, M. (2011). On the use of graphic resources in interaction by people with communication disorders. *Multimodality and Human Activity: Research on Behaviour, Action and Communication*.

Wilkinson, S., & Kitzinger, C. (2006). Surprise as an Interactional Achievement: Reaction Tokens in Conversation. In *Quarterly* (Vol. 69, Issue 2).

World Stroke Organization (WSO). (2022). *Global Stroke Fact Sheet 2022*. http://ghdx.healthdata.org/gbd-results-tool

Yamagishi, J., Veaux, C., King, S., & Renals, S. (2012). Speech synthesis technologies for individuals with vocal disabilities: Voice banking and reconstruction. In *Acoustical Science and Technology* (Vol. 33, Issue 1, pp. 1–5). https://doi.org/10.1250/ast.33.1

Yorkston, K. M. (2010). *Management of motor speech disorders in children and adults* (K. M. Yorkston, Ed.; 3rd ed.) [Book]. Pro-Ed.

Young, V., & Mihailidis, A. (2010). Difficulties in automatic speech recognition of dysarthric speakers and implications for speech-based applications used by the elderly: A literature review. *Assistive Technology*, *22*(2), 99–112. https://doi.org/10.1080/10400435.2010.483646

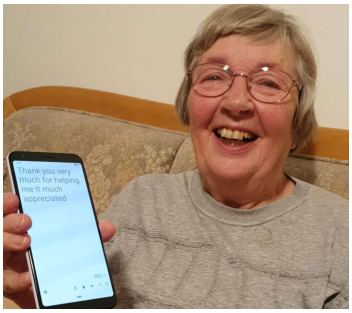
Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist - it's time to make it fair [Article]. *Nature (London)*, *559*(7714), 324–326. https://doi.org/10.1038/d41586-018-05707-8

Appendix

A1. Advert

This study has been approved by the UCL Research Ethics Committee (study ID 12607/003)

Would you like to contribute to research?



How people living with Motor Neurone Disease use Relate to support communication

This is postgraduate research project

Some people use the Relate app to assist others to understand what they are saying. We would like to learn more what happens in everyday conversation to deal with the problems of not being understood, and how Relate may help. The results will help provide recommendations on how best to use Relate and share good practice with health and social care professionals.

We are looking to recruit up to 5 people living with MND. This could be you if you use Relate to help others to understand. You will be asked to nominate a communication partner who you talk to often, knows you well, and is willing to participate in the study.

How does the process work? You will be asked to video record yourself in everyday conversation with your communication partner. This will be done using a computer at your home or at a place convenient to you approximately once a month over a twelve-month period.

With your consent the video recordings will be analysed and transcribed by the researcher and the results will be used to help other people living with MND benefit from Relate, and share good practice guidelines with health and social care professionals.

If you are interested in this project then please feel free to contact Richard Cave (postgraduate research student) Email: **Richard.Cave.17@ucl.ac.uk** phone: 07898 721278. The Principal Researcher is Dr Steven Bloch Email: s.bloch@ucl.ac.uk

Further information on how UCL uses participant information can be found in our 'general' privacy notice: For participants in health and care research studies, click here

A2. Participant Information Sheet

Participant Information Sheet For people living with MND and significant communication partners

UCL Research Ethics Committee Approval ID Number: 12607/003

YOU WILL BE GIVEN A COPY OF THIS INFORMATION SHEET

How people living with Motor Neurone Disease use Relate to support communication

Research Department of Language and Communication

Name and Contact Details of the Researcher: Richard Cave

Email: Richard.Cave.17@ucl.ac.uk

Tel: 07898 721278

Name and Contact Details of the Principal Researcher: Dr Steven Bloch

Chandler House

2 Wakefield Street

London WC1N 1PF

Email s.bloch@ucl.ac.uk

Tel 020 7679 4269

You are being invited to take part in a PhD research project. Before you decided it is important for you to understand why the research is being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information.

Take time to decide whether or not you wish to take part. Thank you for reading this.

1. What is the project's purpose?

To learn more about how conversations work when one person has speech difficulties (called dysarthria). Some people like you may find it difficult to make themselves understood, and you use the Relate app to assist others to understand what you are saying.

We want to understand what you do in everyday conversation to deal with the problems of not being understood when they happen, and how Relate may help. The results will be used to provide recommendations for people living with MND on how best to use Relate, and share good practice guidelines with health and social care professionals.

A person living with MND can put themselves forward and will be asked to nominate a communication partner with whom they communicate on a regular basis, knows them well, and may be willing to participate in the study.

The communication partner will be contacted separately by the research team as part of the application screening process.

2. Why have I been chosen?

You are either:

a) a person living with MND who sometimes finds it difficult to be understood because of speech difficulties (called dysarthria), and you sometimes use the Relate app to help others understand.
b) a communication partner nominated by the person living with MND.

3. Do I have to take part?

It is up to you to decide whether to take part. The researcher working on this project, Richard Cave, will arrange a video call to explain the research. You will have an opportunity to ask him questions. If you decide to take part, you will sign a consent form so that we have a record of your decision. You can stop taking part at any time without telling us why. If you decide to withdraw you will be asked what you wish to happen to the data you have provided up that point.

4. What will happen to me if I take part?

You will video record yourself in everyday conversation with your conversation partner. This will be done using Microsoft Teams on a computer at your home or at a place convenient to you approximately once a month over a twelve-month period. The exact dates and times of these recordings will be up to you. There will be twelve video conversations in total. Each lasting about 30 minutes.

You will be asked to do three things as part of this project

1. You will be asked to talk about how you deal with conversation with your conversation partner and what things help you. This will be recorded in a Microsoft Teams video call. You will be asked these questions at the beginning and at the end of the study.

2. You will be asked to repeat and read some sounds and words, describe a picture, and also talk in conversation with the researcher for about five minutes. This will be recorded in a Microsoft Teams video call. This will take place at the beginning, at session three, six, nine, and at the end of the study.

3. You will be asked to video record yourself in everyday conversation with your conversation partner. This will be recorded in a Microsoft Teams video call at your home or a place convenient to you. The researcher, Richard Cave will help set up the video recording.

We will show you how the Microsoft Teams video call works and give you written instructions. You will make the video when the researcher is not in the video call. We will keep your videos confidential.

Where possible the researcher will conduct video calls and not visit.

If you prefer a face-to-face visit and/or would like help with setting up the video recording equipment needed for the research, we will ask you to review an additional Information Sheet detailing the risks and responsibilities the researcher and participants have for minimizing the risk of COVID infection for a visit.

You will be required to complete a consent form prior to a visit, to indicate understanding of the risks and agreement of the required steps to reduce the risk of Covid infection.

You will be offered copies of the video recordings to keep for yourself if you want them – both the person living with MND and their communication partner should agree to this.

In total the project is running from December 2021 to December 2025. Your involvement will be for one year as early as possible within this period. Once your involvement is over, we will send you an update about our progress and results before the end of the project.

5. Will I be recorded and how will the recorded media be used?

The recordings will be transcribed (written down) and analysed using an observational method called Conversation Analysis. With permission from both person living with MND and communication partner we may also use parts of the recordings for conference presentations, lectures and teaching.

All video recordings, transcriptions and other project documentation will be securely stored at UCL. When we write articles and give presentations to academic meetings and conferences about the project, and when we label your video recordings, we will use a false name, never your real name. Nor will we ever reveal or share any of your personal details (address etc.)

All transcripts will be anonymized, and any identifying information will be edited out. Identifying information from your consent form will be stored separately from the potentially identifiable video data. No other use will be made of them without written permission from both the person living with MND and communication partner, and no one outside the project will be allowed access to the original recordings.

The person living with MND and the communication partner will each be asked to sign a consent form to show that you both agree with how we will keep and use your personal details and videos.

It is possible that a person attending an academic meeting or conference who sees your video might recognise you if they know you. The chance of this happening is actually very small. We do not blank out your face on the video, because we need to study your eyes and facial expression. If you do not want us to show your video data to anyone else then we will not do so.

We will write about the results in books and magazines and scientific journals. You can ask us for a copy of these once they have been published. If we write about you, both the person living with MND and communication partner will not be named – we will use false names.

6. What are the possible disadvantages and risks of taking part?

Talking about how your speech has changed and the difficulties of being understood can be distressing. If you do not feel ready to talk about this or if the subject causes you distress or worry, then this research study is not something you should consider.

7. What are the possible benefits of taking part?

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will improve how Relate may benefit people living with MND and improve the training of professionals working in this area.

8. What if something goes wrong?

In the first instance you should inform Dr Steven Bloch (his contact details are at the top of this information sheet). If you feel your complaint has not been handled to your satisfaction you can contact the Chair of the UCL Research Ethics Committee – ethics@ucl.ac.uk

9. Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any ensuing reports or publications.

10. Limits to confidentiality

Please note that confidentiality will be maintained as far as it is possible, unless during a conversation there is an indication that someone might be in danger of harm. In this situation relevant agencies may need to be informed.

11. What will happen to the results of the research project?

The results of the research are likely to be published in a report in 2026. You will be offered a copy if you would like one. You will not be identified in any report or publication. The data collected during the course of the project will be securely archived and might be used for additional or subsequent research. No personal data will be transferred outside of the EEA.

12. Data Protection Privacy Notice

The controller for this project will be University College London (UCL). The UCL Data Protection Officer provides oversight of UCL activities involving the processing of personal data, and can be contacted at data-protection@ucl.ac.uk

This 'local' privacy notice sets out the information that applies to this particular study. Further information on how UCL uses participant information can be found in our 'general' privacy notice:

For participants in health and care research studies, click here

The information that is required to be provided to participants under data protection legislation (GDPR and DPA 2018) is provided across both the 'local' and 'general' privacy notices.

The lawful basis that will be used to process your personal data are: 'Public task' for personal data and' Research purposes' for special category data.

Your personal data will be processed so long as it is required for the research project. If we are able to anonymise or pseudonymise the personal data you provide we will undertake this, and will endeavour to minimise the processing of personal data wherever possible.

If you are concerned about how your personal data is being processed, or if you would like to contact us about your rights, please contact UCL in the first instance at data-protection@ucl.ac.uk.

If you remain unsatisfied, you may wish to contact the Information Commissioner's Office (ICO). Contact details, and details of data subject rights, are available on the ICO website at: https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/

16. Contact for further information

For further information, please contact Dr Steven Bloch

Chandler House, 2 Wakefield Street, London WC1N 1PF

Tel 020 7679 4269

Email s.bloch@ucl.ac.uk

Thank you for reading this information sheet and for considering taking part in this research study.

A3. Consent form

CONSENT FORM FOR PEOPLE LIVING WITH MND AND SIGNIFICANT COMMUNICATION PARTNERS IN RESEARCH STUDIES

Please complete this form after you have read the Information Sheet and/or listened to an explanation about the research.

How people living with Motor Neurone Disease use Relate to support communication

Research Department of Language and Communication

Name and Contact Details of the Researcher:
Richard Cave
Email: Richard.Cave.17@ucl.ac.uk
Tel: 07898 721278

Name and Contact Details of the Principal Researcher:
Dr Steven Bloch
Chandler House
2 Wakefield Street
London WC1N 1PF
Email: s.bloch@ucl.ac.uk
Tel 020 7679 4269

Name and Contact Details of the UCL Data Protection Officer:
Alexandra Potts Data Protection & Freedom of Information Officer
Email: data-protection@ucl.ac.uk

This study has been approved by the UCL Research Ethics Committee: Project ID number: 12607/003

Thank you for considering taking part in this research. The person organising the research must explain the project to you before you agree to take part. If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time.

I confirm that I understand that by ticking/initialling each box below I am consenting to this element of the study. I understand that it will be assumed that unticked/initialled boxes means that I DO NOT consent to that part of the study. I understand that by not giving consent for any one element that I may be deemed ineligible for the study.

		Tick Box
1.	I confirm that I have read and understood the Information Sheet for the above study. I have had an opportunity to consider the information and what will be expected of me. I have also had the opportunity to ask questions which have been answered to my satisfaction and would like to take part in the following (please tick one or more) (a) video recorded individual interviews (b) video recorded conversations with a communication partner	
2.	I understand that the recordings will be stored securely on UCL OneDrive server, using password-protected software and will be used for specific research purposes and quality control.	
3.	I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason, I understand that if I decide to withdraw, any personal data I have provided up to that point will be deleted unless I agree otherwise.	
4.	I consent to the processing of my personal information, including name, address, diagnosis, date of diagnosis for the purposes explained to me. I understand that according to data protection legislation, 'public task' will be the lawful basis for processing, and 'research purposes' will be the lawful basis for processing special category data.	
5.	**Use of the information for this project only** I understand that all personal information will remain confidential. Although efforts will be made to ensure I cannot be identified, my face will not be blanked out on the video, because the research needs to study my eyes and facial expression, and there is a possibility that a person who sees the video might recognise me, if they know me. It will not be possible to identify me in any publications. If you do not want us to show your video data to anyone else then we will not do so. I understand that confidentiality will be maintained as far as possible, unless during our conversation the researcher hears anything which makes him worried that someone might be in danger of harm, the researcher might have to inform relevant agencies of this.	

6.	I understand that my information may be subject to review by responsible individuals from the University for monitoring and audit purposes.	
7.	I understand the potential risks of participating and the support that will be available to me should I become distressed during the course of the research.	
8.	I understand that the data will not be made available to any commercial organisations but is solely the responsibility of the researchers undertaking this study.	
9.	I understand that the information I have submitted will be published as a report and I wish to receive a copy of it. Yes/No	
10.	I am aware of who I should contact if I wish to lodge a complaint.	
11.	**Use of information for this project and beyond** I would be happy for the data I provide to be securely archived at UCL. I agree that my research data may be used by others for future research. We may also use parts of the recordings for conference presentations and academic meetings. I understand that all personal information will remain confidential. Although efforts will be made to ensure I cannot be identified we do not blank out your face on the video, because we need to study your eyes and facial expression, and there is a possibility that a person who sees your video might recognise you, if they know you.	

_____	_____	_____
Name of participant completing online)	Date	Signature (please type name if
_____	_____	_____
Researcher completing online)	Date	Signature (please type name if

A4. Question Sheet

There are no right or wrong answers, we just want to understand your views.

- Overall how have you been finding using Google Relate?
- Any particular highlight either positive or negative?
- What have been people's reactions when you used Google Relate?
- Did you find that the way you used the application changed over time?
- Are there any situations in which you thought Google Relate could be useful, but you don't think it's the case anymore?
- Are there any situations in which you thought Google Relate would not be useful, but was?
- For the conversation partner: How hard did you work to understand what was said? Remember, this is different than how many words you got right. For example, you could get all the words right but have to work very hard to do it (score out of 10, 10 being a lot of work).
- Is there anything else you would like to add?

A5. Scoping Review 1990 (Jan 1) - 2020 (Nov 15)

Literature Eligibility Criteria

All relevant literature regardless of study design was identified, as well as reviews, editorials, and relevant grey literature between 1st January 1990 - when the first consumer ASR product, Dragon Dictate was launched (Pinola, 2011) and 15th November 2020.

To update the results, an additional scoping review was completed for the period 16th November 2020 to 31st May 2024 using the same criteria.

Evidence from diverse sources can complement the findings from experimental studies and address questions beyond intervention effectiveness in emerging fields such as ASR (Levac et al., 2010).

Given the limited peer-reviewed research in this field, this open approach is essential to capturing and cataloguing the full range of evidence of how plwMND utilise ASR technology. Relevant data was extracted from the included sources, and a narrative summary of the findings created.

For purpose of this investigation, the term ASR is used to refer to voice recognition software that uses natural language input from a speaker (or a recording of a speaker) and carries out predetermined actions.

Information Sources

The following relevant electronic databases were searched in separate stages using identified keywords and indexed terms: PsycINFO; PubMed; Medline; CINAHL; Scopus; LLBA; Cochrane Library; Embase; Web of Science Core Collection; ACM digital Library.

To identify further studies, reference lists were scanned, and the grey literature was also searched. The review considered any existing primary research studies of quantitative, qualitative or mixed-methods design, as well as reviews, editorials, and relevant grey literature in English.

The following sources of grey literature were search: Global Health, HMIC - Health Management Information Centre, OpenDOAR, OpenGrey, PsycEXTRA, Social Care Online, OpenAIRE, Semantic Scholar, BASE: Bielefeld Academic Search Engine, Google Scholar.

Additionally, broader internet searches were undertaken using Google Chrome.

For the first scoping review, the searches were undertaken in October and November 2020 and the most recent search was completed on 15th November 2020. For the second scoping review, the searches were undertaken in May and June 2024 and the most recent completed 16th June 2024.

Search Strategy

The following search terms were used, representing the most commonly used brands of voice assistants using ASR on home-based devices [46], frequently used names and abbreviations for automatic speech recognition technology, known research projects developing ASR technology for dysarthric speech specifically including plwMND, and frequently used names and abbreviations for ALS.

Alexa, ALS, Amazon Transcribe, Amazon AWS, Amyotrophic lateral sclerosis, ASR, Automatic Speech Recognition, Bixby, BlackBerry Assistant, Cortana, Dragon Assistant, Dragon Naturally Speaking, Euphonia, Google Assistant, Intelligent personal assistant, Lou Gehrig's disease,

Microsoft Azure, Microsoft Speech Studio, MND, Motor neuron disease, Motor neurone disease, Siri, Speech Recognition, Voiceitt, virtual assistant, voice activated environmental control, voice assistants, voice recognition.

The search string was constructed in the following way:

("amyotrophic lateral sclerosis" OR " ALS" OR "motor neuron? disease" OR " MND" OR "Lou Gehrig*") AND ("ASR" OR "automat* speech recognition" OR “speech recognition” OR “intelligent personal assistant” OR "voice assistant*" OR "voice recognition" OR "Alexa" OR “Amazon AWS” OR “Amazon Transcribe” OR "Google assistant" OR “Microsoft Speech Studio” OR “Microsoft Azure” OR "Siri" OR "Cortana" OR "Bixby" OR “BlackBerry assistant” OR "virtual assistant*" OR "Dragon assistant" OR "Dragon naturally*" OR “Euphonia” OR “Voiceitt” OR “voice activated environmental control”)

The final search results were exported into Mendeley, and duplicates were removed.

Selection of sources of evidence

Literature screening was performed by the author, a UK registered Speech and Language Therapist (SLT) using the description of the ASR and how it related to plwMND, to understand if the study was eligible according to the criteria of this scoping review.

Summarizing the results

The following data extraction proforma has been adapted from the template provided in the Joanna Briggs Institute methodology manual (M. D. Peters et al., 2015) to capture relevant information from grey literature as well as peer-reviewed articles. This proforma was trialled

on the eligible studies from the search, and the categories were slightly revised. The charting was completed by the first author and reviewed by the second.

Results 1990 (Jan 1) - 2020 (Nov 15)

From a total of 177 records identified from electronic databases and grey literature searches, eleven articles met the eligibility criteria.

Table 1 contains the summary data extraction proforma for the eleven identified studies.

Tables 2-4 provide summaries of the data extracted from the 11 studies, comments on their outcomes, and suggestions for future studies.

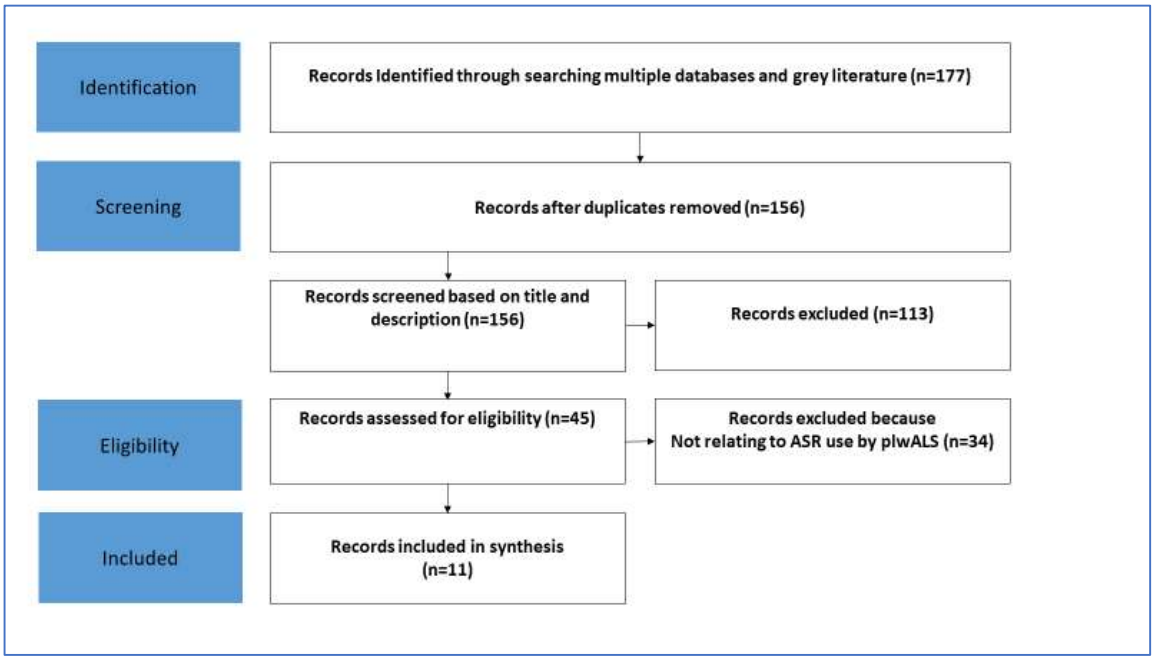


Table 1. PRISMA flow diagram of record selection 1990 - 2020.

Hird & Hennessey (2007) examined the relative benefit modifying a person's own speech to improve speech recognition software performance. Fifteen adults with dysarthria associated with a variety of aetiological conditions participated, including two plwMND. The plwMND were described as having moderate mixed dysarthria and low intelligibility following assessment using the Frenchay Dysarthria Assessment (Enderby, 1980). Type of MND was not specified. For the group as a whole, overall mean rate of dictation increased from 4.3 words per minute (wpm) to 6.3 wpm as a result of therapy aimed at better regulation of flow of air and volume loss during speech, practice at increasing loudness of speech to mitigate growing environmental noise. The authors concluded that the clinical significance of the treatment effect may be doubtful given the difference in mean dictation rates was small - suggesting modest practical benefit.

Rudzicz (2011) identified significant improvements to ASR accuracy for dysarthric speech when theoretical and empirical knowledge of the vocal tract for atypical speech were included in ASR model training, concluding that this approach should benefit assistive software for speakers with dysarthria. The recordings used were mainly from the MOCHA database of recordings of people without dysarthria and the authors own collection of dysarthric speech (it is unclear if any of these recordings were of plwMND). Recordings from the TORGO database were also used, which includes recordings of only one plwMND. The author highlighted that the assessed recordings were brief and reflected a snapshot in time, and some fundamental aspects of dysarthria such as dysfluency and reduced pitch control could not be reflected in the methods described in the paper.

Rudzicz (2013) later described a portable communication device that could potentially increase the intelligibility of speech with dysarthria by removing repeated sounds, inserting deleted

sounds, devoicing unvoiced phonemes, and adjusting the tempo and frequency of speech. Recordings from the TORGO database were used. In the paper, ASR word recognition rate increased from 72.7% to 87.9% relative to the original speech, and the recognition rate of human listeners increased from 21.6% to 41.2%.

Hahm et al. (2015) used 316 recordings of short phrases from five plwMND, three females and two males. All had been diagnosed a year or less, and their average speech intelligibility was 95%. There was no indication of what intelligibility assessment was used. Each plwMND produced up to four repetitions of twenty sentences used in daily conversation (for example, 'How are you?') or related to their medical condition ('This is an emergency'). The best reported ASR phoneme error rate (PER) of 31% was a result of combining articulatory and acoustic data with hidden Markov model (HMM) and deep neural network (DNN) technology. The authors acknowledged that the data in the study contained a small number of unique phrases from a small number of plwMND, and further research was needed with a larger dataset.

Kim et al. (2018) collected recordings from nine plwMND. Four of the plwMND made recordings on two or more occasions on average six months apart. Speech data was also collected from seven people with normal speech and no history of speech or language impairment. The speech intelligibility of each plwMND was measured by a speech pathologist for each session and varied between normal and severely unintelligible. For every recording session, each plwMND produced up to four repetitions of twenty unique sentences typical of daily conversation (for example 'How are you doing?'). Significant variation in speech intelligibility and ASR performance was measured across sessions within individual speakers, likely because of disease progression. For example, one participant declined from a 99% intelligibility rate to 0% and the phoneme error rate (PER) of the ASR increased from 14% to 75%. Factors were detailed

that negatively affected ASR accuracy as MNDS progressed, including increased dysarthria, long pauses between words, and reduced rate of speech. The authors predicted that as MND progresses the capability of plwMND to effectively use ASR would reduce. The best ASR performance observed with a PERs of 30.6% for plwMND and the authors argue their technical approach presents a possibility in effectively modelling dysarthric speaker-independent speech. The authors hypothesized that the ASR performance may improve with more recordings from plwMND displaying increased diversity of dysarthria severity.

Caves (2007) adapted ASR technology designed for military training simulation to create four ASR models trained on recordings from The incidence of MND is highest in people aged between 55-79 50 plwMND (21 male, 29 Female). The participants were described as having spastic, flaccid, or mixed dysarthria. There was no indication of severity, current communication needs or AAC use. Each participant recorded the digits 0 to 9 and read one hundred zip codes (postal codes). Based on an eleven-item digital recognition task, word accuracy was reported as 86% for individuals with mild dysarthria, 70-75% for moderate dysarthria and 51% for severe dysarthria. The word recognition rate for mild and moderate dysarthria models were claimed to be in line with the initial performance of off-the-shelf ASR systems for dysarthria.

Ballati et al. (2018) investigated the extent that Siri, Cortana and Google Assistant accurately interpreted 34 sentences in Italian spoken by eight plwMND and the appropriacy and consistency of the generated answers. Three of participants were categorized as having flaccid dysarthria, two spastic dysarthria, two unilateral UMN dysarthria and one mixed dysarthria. All participants were described as having moderate dysarthria, six scored level 3 – detectable

speech disturbance on ALS Functional Rating Scale (FRS-r) and two scored level 2 (intelligible with repeating). All were native Italian speakers.

Each participant recorded 34 phrases, selected from a collection of frequently asked questions for Amazon Echo and Google Home, modified to include all the phonemes of the Italian language (for example "When will the next Juventus match be?"). The recordings were then played to Siri, Cortana and Google Assistant.

Overall average word error rates (WER) for Google Assistant, Cortana, and Siri were 24.88%, 39.39% and 70.89% respectively. Correct responses were 63.96%, 39.7%, and 15.81%. Consistency in answers were 54.02%, 24.07%, and 60.47%.

The authors reported that WER is highly dependent on the level of dysarthria of the plwMND. In addition, the type of ASR device used also affected WER. In this study Google Assistant had the highest number of properly transcribed questions and provided a higher number and proportion of correct answers. Siri provided the highest consistency of correct answers across plwMND but for significantly fewer correct answers compared to Google Assistant

They concluded that Google Assistant was the only ASR device usable in functional tasks by people with dysarthria. However, they reported that usability is closely related to each individual's level of dysarthria and level of fatigue. ASR question comprehension was found not to be related to the different types of dysarthria. The authors observed that when there was an interval of silence due to the effort of speaking, Google Assistant frequently stopped 'listening', missing or misinterpreting the question.

The authors acknowledged some factors that may have affected the study outcomes: the phrases were pre-recorded, did not include the call phrases, for example “Hey Google”, and the recordings were replayed from a speaker to the ASR devices.

In a related study, Ballati et al. (2018) used recordings from seven participants from the University of Toronto’s database of acoustic and articulatory speech from speakers with dysarthria (TORGO). The recordings were used to examine how Siri, Google Assistant, and Amazon Alexa comprehends words and sentences spoken by people with dysarthric voices and the appropriacy of the generated answers.

One participant had been diagnosed with MND and the remainder diagnosed with cerebral palsy. No details were provided about the severity of dysarthria.

Five sentences were created from the TORGO dataset pronounced by seven speakers. The test phrases were limited by available recordings from TORGO that could be adapted to be possible requests to ASR devices, for example “Please open the window quickly”. The participants did not all use the same phrases or words due to the limitations of the TORGO recordings. The recordings made by each participant were not detailed.

The average WER for all participants was for Siri 69.41% and Google Assistant 15.38%. The WER for Amazon Alexa could not be measured as Alexa does not show a transcription. The overall average for appropriateness of answers was for Siri 58.82%, Google Assistant 64.7%, and Amazon Alexa 41.17%.

The authors concluded the three ASR devices assistants have comparable performance with a recognition percentage of around 50-60%. They recommended future research to evaluate a

larger number of sentences pronounced by people with dysarthria and a detailed profile of type and severity of dysarthria for each participant.

De Russis & Corno (2019) examined the WER of IBM Watson Speech-to-Text, Google Cloud Speech and Microsoft Azure Bing Speech ASR platforms for recognizing dysarthric speech. They assessed the most frequent transcription errors and if the transcription alternatives provided by each ASR platforms could be used to improve the overall recognition result. Recordings from the TORGO database were used from eight participants. One participant had been diagnosed with MND and the remainder were diagnosed with cerebral palsy. No details were provided about which participants had MND and which had cerebral palsy. 38 specific sentences were used and had been recorded by all five males and 13 sentences recorded by all three females. The speech intelligibility of the eight speakers with dysarthria ranged from “no abnormalities” (for two females and one male) to “severely distorted” (the remaining speakers, four males and one female).

The average WER for all participants for Google Cloud Speech was 59.81%, Microsoft Azure Bing Speech 62.94% and IBM Watson 67.35%. The most frequent transcription errors were “one or many” word errors with undetermined impact on semantic meaning. Using the most accurate transcription alternatives significantly improves WER (Google 5%, IBM 9%), although the alternatives were selected by the authors based on prior knowledge of the TORGO transcriptions.

The semantic impact of the 'single or multiple' word errors was reported as having little impact on understanding of the sentence, though the authors acknowledged the TORGO database phrases lacked ecological validity.

Pradhan et al. (2018) examined how users with disability use the Amazon Echo (a device that uses ASR). The authors completed thematic analysis of 346 Amazon Echo reviews on Amazon.com that include reference to users with a range of disabilities. Although the study did not detail which participants had MND there were two quoted reviews in the paper that related to plwMND. They also conducted semi structured interviews of 16 people with visual impairments (none had MND). The summary results indicated that users with impaired speech were generally positive about their experience with the Amazon Echo. Where the study did attribute a quotation from plwMND, one highlighted the ability of the device to control smart home appliances such as lights or thermostats in order to reduce caregiver burden. Another indicated Alexa was used to alert others that help was needed by switching the lights off and on in another room.

The authors acknowledged that as two-thirds of the reviews were written by third parties (significant others, carers for example) they may not be as accurate as first-person reviews in reflecting the experience of users with disabilities. The authors also hypothesized a potential sampling bias: users with more severe speech or physical impairments may not have thought to try the device and thus to write a review.

Shor et al. (2019) completed research on how to improve ASR accuracy for plwMND with mild-severe dysarthria, and for people with heavy accent. Seventeen plwMND completed a combined total of twenty-two hours of audio recordings. The participants were given sentences to read from The Cornell Movie-Dialogs Corpus, phrases used by text-to-speech voice actors, and a modified selection of sentences suggested for message banking from the Boston Children's Hospital. The plwMND had FRS-r speech scores between 1-3, categorized in the paper as having mild - severe dysarthria. No details were provided on an individual basis.

Using a test set of recorded phrases suggested for message banking (the list was not disclosed), overall WER was reduced from 33% to 11% for mild dysarthria (FRS-r score 3) and from 60% to 20% for moderate-severe dysarthria (FRS-r score 1-2). The study also identified patterns in phoneme deletion, insertion and substitution in the speech recorded by the group of plwMND and suggest further research might lead to early MND-detection techniques.

Table 3: Study types, participants, technology used, method.

Publication Details and Characteristics			
Citation	**Study Type**	**Objectives**	**Participant details (e.g. type of ALS age/sex and number), Technology used, Method**
Hird and Hennessey (2007) [48] Australia	Primary Research, peer reviewed	Therapy aimed at modifying a person's speech output to improve speech recognition software performance.	15 adults with dysarthria. Two had ALS. Type of ALS was not specified. One male age 46 and one female age. Power Secretary ASR technology was used 30 minutes of speech therapy treatment followed by 30 minutes of dictation. 15 sessions overall.
Caves et al. (2007) [50] USA	Conference Proceeding	To improve ASR accuracy for plwALS with mild-severe dysarthria.	50 plwALS (21 male, 29 Female). No ALS type or age of participant reported. An adapted military training simulation ASR model was used. Participants read digits 0 to 9 and one hundred "zip codes".
Ballati et al. (2018) [51] Italy	Conference Proceeding	Investigated the extent that Siri, Cortana and Google Assistant can comprehend words and sentences spoken by plwALS.	Eight participants (four female, four male), aged 67-83, ALS type not stated. Native Italian speakers. Google Assistant, Cortana, Siri were used.34 Italian sentences were recorded from 8 people and audio files from each participant were presented to each voice assistant.
Ballati et al. (2018) [52] Italy	Conference Proceeding	Investigate the extent that Siri, Google Assistant, and Amazon Alexa can comprehend words and sentences spoken by people with dysarthric voices and the accuracy and consistency of the answers generated.	Recordings from eight participants from the TORGO database. One participant had been diagnosed with ALS (remainder diagnosed with Cerebral Palsy). Siri, Google Assistant, Amazon Alexa were used.5 sentences were created from the TORGO dataset pronounced by 7 speakers. Performance of each system was measured by WER and appropriacy of answer.
De Russis and Corno (2019) [24] Italy	Primary Research, peer reviewed	(1) Are ASR platforms suitable for recognizing dysarthric speech? (2) What is the attained recognition rate? (3) What kinds of transcription errors are more frequent, in case of imperfect/partial recognition? (4) Can transcription alternatives (as provided by ASR platforms) improve the overall recognition result?	Recordings from eight participants(five males, three females) in English from the TORGO database. One participant had been diagnosed with ALS (remainder diagnosed with Cerebral Palsy). IBM Watson Google Cloud Speech, Microsoft Azure Bing Speech were used. 38 sentences recorded by all females, 13 sentences by all females. Performance measured by WER of each system, and the WER of the transcription alternative.
Pradhan et al (2020) [53] USA	Conference Proceeding	(1) To examine the accessibility of off-the-shelf Intelligent personal assistants (IPAs) (e.g., Amazon Echo) (2) To understand how users with disabilities are making use of these devices.	(1) Analysis of 346 Amazon Echo reviews on Amazon.com that include users with a range of disabilities, including ALS (the number, type, and gender unspecified). Amazon Alexa was used. (2) Semi structured interviews of 16 people with visual impairments (none had ALS).
Shor et al. (2020) [54] USA	Conference Proceeding	(1) To improve ASR accuracy for plwALS with mild-severe dysarthria, and for people with heavy accent.	17 plwALS. No details provided on type of ALS, age or sex. Personalized ASR models were used. The participants were given sentences to read from The Cornell Movie-Dialogs Corpus, phrases used by text-to-speech voice actors, and a modified selection of sentences from the Boston Children's Hospital.
Hahm et al. (2015) [56] USA	Conference Proceeding	(1) speaker independent recognition of dysarthric speech due to ALS using articulatory data.	5 plwALS. 3 females, 2 males. Average age 59. Recently diagnosed (within a year of the resarch) No details provided on type of ALS. Gaussian mixture model (GMM) - hidden Markov model (HMM), deep neural network (DNN) - HMM combined with articulatory data were used. plwALS produced up to four repetitions of twenty sentences used in daily conversation (for example, 'How are you?') or related to their medical condition ('This is an emergency').
Kim et al. (2018) [57] USA	Conference Proceeding	(1) speaker independent recognition of dysarthric speech due to ALS using spectral and temporal speech data.	9 plwALS. 6 females, 3 males. Average age 62.8. Recently diagnosed (within a year of the resarch). No details provided on type of ALS. Gaussian mixture model (GMM) - hidden Markov model (HMM), deep neural network (DNN) - HMM combined with articulatory data were used. plwALS produced up to four repetitions of twenty sentences used in daily conversation (for example, 'How are you?') or related to their medical condition ('I need to make an appointment'). 4 plwALS recorded with an average gap of 6 months between recordings.
Rudzicz F. (2011) [26] USA	Conference Proceeding	(1) Improve ASR for dysarthric speech by including articulatory information.	Recordings from MOCHA database of people without dysarthria, the authors own collection of dysarthric speech (not specified if any were of plwALS) and recordings from the TORGO database (One participant had been diagnosed with ALS and remaining seven diagnosed with Cerebral Palsy). Hidden Markov model (HMM), deep neural network (DNN), Support Vector Machines (SVMs), Dynamic Bayes Networks (DBNs) combined with articulatory models were used. Recordings from two speakers with dysarthria (male and female) and two speakers without (male and female). It is unclear if any were plwALS.
Rudzicz F. (2012) [58] UK	Primary Research, peer reviewed	(1)This paper presents a system that transforms the speech signals of speakers with physical speech disabilities into a more intelligible form that can be more easily understood by listeners.	Recordings from the TORGO database (One participant had been diagnosed with ALS and remaining seven diagnosed with Cerebral Palsy). human listeners attempt to identify words in sentence-level utterances under a number of acoustic scenarios. Sentences are either uttered by a speaker with dysarthria, modified from their original source acoustics, or produced by a text-to-speech synthesizer.

Table 4: Participants' communication needs and studies' outcomes.

Details/Results extracted from study		
Citation	**Communication needs: dysarthria, AAC Physical or cognitive impairments.**	**Outcomes for communication support / enviornmental control**
Hird and Hennessey (2007) [48] Australia	ALS participants had mod mixed UMN and LMN dysarthria, and low or intelligibility. AAC needs not detailed. Any physical or cognitive impairments were not specified.	The overall mean rate of dictation increased from 4.3 wpm to 6.3 wpm.
Caves et al. (2007) [50] USA	The participants were reported to have including spastic, flaccid, or mixed dysarthria. There was no grouping of severity, indication of current communication needs, or AAC use. Any physical or cognitive impairments were not specified.	Based on an eleven-item digital recognition task, word accuracy was 86% for individuals with mild dysarthria, 70-75% for moderate dysarthria and 51% for severe dysarthria.
Ballati et al. (2018) [51] Italy	Six participants had detectable speech disturbance, and two were intelligible with repeating. Three participants had flaccid dysarthria, two spastic dysarthria, two unilateral UMN dysarthria and one mixed dysarthria. Any physical or cognitive impairments were not specified.	Overall average word error rates (WER) for Google Assistant, Cortana, and Siri were 24.88%, 39.39% and 70.89% respectively. WER dependent on the level of dysarthria and ASR type
Ballati et al. (2018) [52] Italy	No details provided about the single plwALS and unable to identify which of the seven participants the plwALS is. AAC needs not detailed Any physical or cognitive impairments were not specified.	Overall average (all participants) WER for Siri 69.41%, Google Assistant 15.38%, Amazon Alexa could not be measured. Overall average appropriate answers from Siri 58.82%, Google Assistant 64.7%, Amazon Alexa 41.17%.
De Russis and Corno (2019) [24] Italy	Six participants had detectable speech disturbance, and two were intelligible with repeating. Any physical or cognitive impairments were not specified.	Google Cloud Speech average WER 59.81%, Microsoft Azure Bing Speech 62.94% and IBM Watson 67.35%. Using the most accurate transcription alternatives improves WER though they were manually selected based on knowledge of the TORGO transcriptions.
Pradhan et al (2020) [53] USA	Communication / AAC needs were not specified Any physical or cognitive impairments were not specified.	The study did not detail which participants had ALS. Users with impaired speech were generally positive about their experience with ASR technology. One plwALS able to control lights / thermostat.
Shor et al. (2020) [54] USA	Three participants had flaccid dysarthria, two spastic dysarthria, two unilateral UMN dysarthria and one mixed dysarthria. Any physical or cognitive impairments were not specified.	The overall WER for ALS speakers, on a test set of message bank phrases (not disclosed), was reduced from 33% to 11% for mild dysarthria and from 60% to 20% for moderate-severe dysarthria.
Hahm et al. (2015) [56] USA	Severity of dysarthria was mild with average speech intelligibility of 95%. No indication of assessment used or who conducted the assessment. Any physical or cognitive impairments were not specified.	The best reported phoneme error rate (PER) 31%, reduced from 46% baseline.
Kim et al. (2018) [57] USA	Severity of dysarthria was mild to profound with average speech intelligibility ranging from 100% to 0%. Assessed by a speech pathologist. Any physical or cognitive impairments were not specified.	Significant variation in speech intelligibility and ASR performance was measured across sessions likely due to ALS progression. Factors that may have negatively affected ASR accuracy included increased dysarthria, long pauses between words, and reduced rate of speech.
Rudzicz F. (2011) [26] USA	Level of dysarthria was not described Any physical or cognitive impairments were not specified.	Significant improvements to ASR accuracy for dysarthric speech when theoretical and empirical knowledge of the vocal tract for atypical speech were included in ASR model training.
Rudzicz F. (2012) [58] UK	Level of dysarthria was not described Any physical or cognitive impairments were not specified.	Human listener recognition rates increased from 21.6% to 41.2%) and ASR word recognition rate increased from 73% to 88%) relative to the original speech. Potential gains to EC derived from increased functional use of ASR device.

Table 5: Studies' advantages and disadvantages and future studies.

Summary			
Citation	**Advantages**	**Disadvantages**	**Future Studies**
Hird and Hennessey (2007) [48] Australia	A marginal increase in WPM was observed.	The authors concluded that the clinical significance of the treatment effect was doubtful.	Further research using a between groups design and control group with no concurrent treatment may show material gains in WPM from clinical intervention.
Caves et al. (2007) [50] USA	word recognition rate for mild/mod dysarthria was claimed to be in line with off-the-shelf ASR systems for dysrathria.	A small phraseset was tested, potentially of limited functional use.	Further research to assess the ASR model with a larger dysarthric speech dataset.
Ballati et al. (2018) [51] Italy	Google Assistant was the only ASR argued to be usable in functional tasks by people with dysarthria.	ASR useability related to dysarthria severity and fatigue.Few subjects involved in the experiment. Sentences played from recordings - not reflecting variation in speech.	Further research to evaluate a larger number of phrases. Compare the same speech samples spoken by people without dysarthria as a control. Research accessibility of ASR devices.
Ballati et al. (2018) [52] Italy	Siri, Google Assistant and Alexa have comparable performance.	Limited recordings assessed. The recordings were adapted from phrases not used with ASR.	Further research to evaluate more sentences and participants. Assess usefulness or ASR for different severities of dysarthria, environmental control and accessibility of ASR devices.
De Russis and Corno (2019) [24] Italy	Google Cloud Speech had the lowest average WER of 59.81%.	Even the lowest WER assessed was reported as significantly worse than the average WER of speakers without dysarthria.	Future work to identification of an algorithm able to improve the selection of accurate transcription alternatives, without reliance on a priori knowledge of the original text sentence.
Pradhan et al (2020) [53] USA	Potential for increased autonomy and reduced burden on caregivers.	Very limited data for plwALS use of IPAs in the study.	Further research to explore the use of IPAs for speech therapy,memory aids, and improved design of feature rich voice-based interfaces and apps.
Shor et al. (2020) [54] USA	Significantly lower WER is argued to enable plwALS to more successfully interact with smart home devices	WER remains significantly higher than for speakers without dysarthria.	The study identified common speech patterns and suggested research for early ALS detection techniques. Also research for building ASR models with limited data of dysarhtric speech.
Hahm et al. (2015) [56] USA	Significantly lower PER was achived by articulatory data to the ASR technology.	The data set used in the experiment contained only a small number of unique phrases collected from a small number of ALS patients.	Further studies with a larger vocabulary from more ALS patients are necessary to explore the limits of the proposed approach.
Kim et al. (2018) [57] USA	Best ASR performance observed a PERs of 30.6% for plwALS. Authors argue their approach may effectively modelling dysarthric speaker-independent speech.	The data set used in the experiment contained only a small number of unique phrases collected from a small number of ALS patients.	Further studies with a larger vocabulary from more ALS patients are necessary to explore the limits of the proposed approach.
Rudzicz F. (2011) [26] USA	The authors reported that improvements in ASR accuracy are possible through discriminative methods such as latent-dynamic conditional random fields.	The author highlighted the recordings were brief and a snapshot in time, and some fundamental aspects of dysarthria could not be reflected in the methods described in the paper.	Further studies with a larger vocabulary from more ALS patients are necessary to explore the limits of the proposed approach.
Rudzicz F. (2012) [58] UK	Potentially significant improvement to ASR functional usefulness for people with dysarthria.	Not tested on a broad range of people living with dysarthria.	Further studies with a larger vocabulary from more ALS patients.

A6.Scoping Review 2020 (Nov 16) - 2024 (May 31)

- Results 2020 (Nov 16) - 2024 (May 31)

From a total of 909 records identified from electronic databases and grey literature searches, 33 articles met the eligibility criteria.

Table 5 contains the summary data extraction proforma for the 33 identified studies.

Tables 6-8 provide summaries of the data extracted from the 33 studies, comments on their outcomes, and suggestions for future studies. Tables 6-8 can be viewed online here and are below.

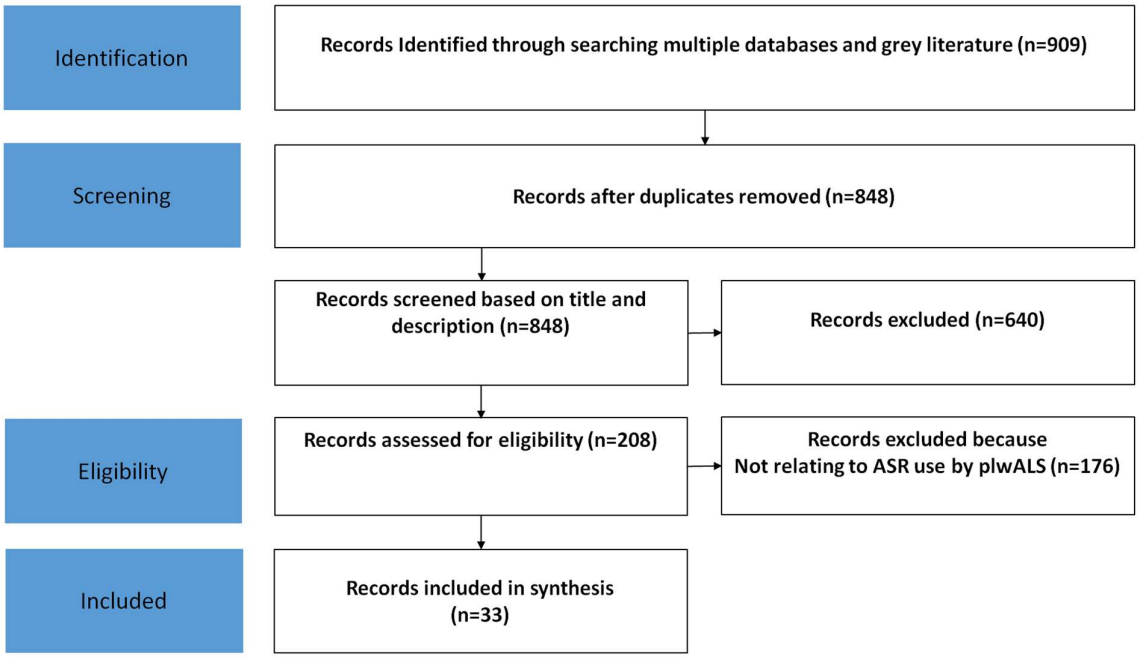


Table 5. PRISMA flow diagram of record selection 2020 (Nov 16) - 2024 (May 31)

From a total of 909 records identified from electronic databases and grey literature searches, 33 articles met the eligibility criteria.

Table 5 contains the summary data extraction proforma for the 33 identified studies.

Tables 6-8 provide summaries of the data extracted from the 33 studies, comments on their outcomes, and suggestions for future studies. Tables 6-8 can also be viewed online here.

The 33 papers are grouped into five categories:

(1) Research to improve ASR accuracy (17 papers)

(2) Using ASR to support diagnosis of MND and other conditions (2 papers)

(3) Replication and/or automation of SLT assessments (10 papers)

(4) Using ASR as AAC (3 papers)

(5) How plwMND use ASR technology already (1 paper)

- Research to improve ASR accuracy

MacDonald et al., (2021) discussed the lessons learnt from collating the Euphonia dataset of recorded dysarthric speech, comprising over one million utterances (1300 hours of disordered speech) from more than 1000 speakers. 179 plwMND contributed 300 or more have recorded phrases to this dataset. The Euphonia dataset is used to create personalised ASR models that are claimed to reduce WER by about 75% compared to standard off-the-shelf ASR models, and more than 80% of the personalized ASR models achieved WERs below 15% - however this varies by individual. The research is argued to demonstrate that personalized ASRs significantly improve recognition accuracy compared to standard ASR models.

The authors reported that mainly scripted speech and shorter phrases were collected for the Euphonia dataset. They justified this by arguing the approach allows for high transcript conformity and reduced transcription costs. However, it also potentially limits the potential usefulness of the Euphonia dataset for developing ASRs for spontaneous and extended conversational speech.

Additionally, as the dataset is not publicly available outside of Google Research due to participant consent terms, it may not support broader research and collaboration in the field. The planned research included developing speaker-independent ASR models capable of accurately recognizing disordered speech without the need for individual personalisation.

Green et al., (2021) selected short phrases of 2-4 words from a set of 172 recordings from plwMND from the Euphonia dataset in order to compare ASR accuracy between two speaker-independent ASR (SIASR) models and a personalised ASR (PASR) model. PASR showed a very significant performance gain over the others as measured by WER (4.6% median WER for the PASR vs 31.5% and 29.4% for the SIASRs). Additionally, the PASR median WER was approximately 9% lower when compared to human transcribers listening to the same recordings.

The researchers concluded that PASR models may have the potential to assist or even replace human listeners in certain applications, and can be effective in recognizing a wide range of speech impairments and severities.

The authors acknowledged that more research is needed to improve the robustness of PASR when speech is produced in longer phrases including everyday conversational communication, or when speech is characterized by low intelligibility and high production variability.

Tobin & Tomanek, (2022) investigated the performance of personalized ASR for recognizing disordered speech using small amounts of per-speaker data. Recordings from 195 participants in the Euphonia dataset were used including those from plwMND, although though the number was not detailed, or their profile. Personalized ASR models were trained with recordings ranging in size from under one minute to 18-20 minutes of speech data.

WER thresholds were selected to determine 'Success Percentage' (the percentage of personalized models reaching the target WER) in different application scenarios. For the 'home automation' use-case (recorded phrases that reflected frequently used commands to digital assistants), the researchers argued that a WER of 15% corresponded to a success rate of 80% when engaging with Google Assistant. For the 'conversational' use-case, a target of 20% WER was set due to the less structured, longer and more complex phrases typical of conversational speech. The researchers assumed both WER thresholds would achieve functional benefit for people with dysarthria when using ASR.

For the 'home-automation' use-case, average Success Percentage was 79% when models are personalized with 250 training utterances per speaker. With 50 utterances, the average Success Percentage was 63%. For the 'conversational' use-case, models personalized with 250 utterances saw a Success Percentage of 73%.

The researchers assumed both WER thresholds would achieve functional benefit to people using the ASR. However it remains unclear if this was indeed the case as there were no opportunities for end-users to validate these thresholds, and there was no testing with actual human-in-the-loop conversations. Success was measured using WER only.

For the 'conversational' dataset in Euphonia, participants are asked to repeat in isolation short to medium length phrase, for example ”Hey, why don’t you come over for dinner tomorrow”. This may not reflect conversational speech in daily life. For the ‘home-automation’ use-case a sample of commands to a digital assistants are used. There has been no research to understand if these are typical commands that plwMND actually use.

Future research is planned to identify how the data needed for speech model personalization can be further reduced and still be useful for task-specific utterance selection.

Tobin et al., (2022) investigated the use of BERTScore as an alternative metric to WER for assessing the accuracy ASR models when captioning speech with dysarthria. BERTScore assesses semantic similarity between the ASR output and the equivalent ‘ground truth’ transcription, which is argued to be a more informative measure then WER.

The research analysed the transcription errors made by the personalized ASR model created from Euphonia data for 15 individuals including plwMND. The transcription errors were given a WER score and SLTs then categorized each transcription error, assessing for severity based on degree of meaning loss. The study examined how well both BERTScore and WER aligned with the SLT assessments of meaning loss.

The SLTs assessed 3,473 transcription errors and assigned an ‘error type’ including deletion, contraction, normalization, homophone, spelling, proper noun, repetition. They then scored the ‘error severity’ assessment on a scale of 0, 1 or 2 depending on whether or not an error preserved meaning. 0 and 1 errors are judged as 'recoverable', where a communication partner would be able to understand what a speaker meant to say. An error level 2 represented a ‘communication breakdown’, where the perceived meaning is lost or drastically altered from a speaker’s intent.

The results showed BERTScore was more effective than WER in distinguishing between different levels of error severity and error types, and better than WER for predicting the SLT error assessments. BERTScore was also reported to be better than WER at recognizing when the overall meaning of a sentence is preserved, even with small errors. For example, it correctly assessed that "I'm" and "I am" mean the same thing, while WER counted this as an error. Additionally, BERTScore was more aligned with SLT judgments on error severity than WER and better differentiated between minor errors (like small typos) and major errors (that change the meaning completely). For example WER would count Beyoncé vs Beyonce as an error but BERTScore would not.

The researchers argued that BERTScore provides a more realistic measure of how useful an ASR system would be in practical use. BERTScore showed that the ASR transcription could still be useful to the reader if the errors didn't significantly affect understanding, even when WER score was high.

However, it was acknowledged that BERTScore has difficulty assessing words that are valid but not in the large language model (LLM) - for example some proper nouns of people places and

things, as well as multilingual phrases. Although BERT models can capture context and meaning from previous sentences, it has difficulty with longer passages. In this way research is needed to understand if BERT is useful in longer spontaneous speech.

The researchers also point out that BERTScore is more computationally expensive in terms of processing overhead than WER, and WER is more flexible – it can be used regardless of language or word selection. Finally, the researchers report that the BERT accuracy rates reported may not replicate in everyday environments with background noise.

Future research includes investigating if additional metrics including Character Error Rate (CER) may give insight into how an ASR model may be useful for speakers with dysarthria. Additionally, to assess if utilizing both WER and BERTScore together is more beneficial to assess ASR usefulness than using each measure individually.

Tomanek et al., (2023) examined the impact of degenerating speech on ASR accuracy for four plwMND with progressive dysarthria, assessed using perceptual speech severity ratings completed by two speech therapists and scored on a Likert scale ranging from 'typical' to 'anarthric'. Anarthria refers to the complete inability to articulate speech due to severe dysfunction of the muscles involved in speech production (Duffy, 2020). The recordings used were obtained from the Euphonia dataset collected over approximately 18 months, allowing for cases with declining speech to be identified. Only recordings of short phrases (3-5 words) were used and were typically commands to digital assistants, or generic phrases that a person was given to read.

The authors concluded that personalized ASR models become less effective over time as dysarthria becomes more severe, and emphasized the importance of continuous recording and ASR model retraining. The future research plan included recording a more controlled, longitudinal dataset with additional participants, and extending the study to additional aetiologies that lead to degenerating speech.

The authors acknowledged that only recordings of 3-5 word phrases were used and stated the reason was to avoid 'confounding factors' associated with ASR for conversational speech. Therefore, the research outcomes may not be applicable for plwMND with changing speech that use ASR in conversational settings.

Tomanek et al. (2024) explored how effective large language models (LLMs) may be used as a proxy for human evaluation when assessing preservation of meaning in ASR captioning of speech with dysarthria, using a measure called LATTEScore ("LLMs to Assess TranscripTion Errors Score"). LATTEScore was assessed alongside WER, BERTScore and other measure in comparison to 'ground truth' meaning preservation scores from SLTs. The results showed that LATTEScore can assess meaning preservation of ASR transcripts more effectively than WER and BERTScore. The researchers proposed LATTEScore as a viable method for automation of quality assessment of ASR model transcriptions, and an alternative to the slower and labour intensive method of manual assessment by SLTs. However, no comment was made about the content of the phrases used, whether conversational or transactional, short or long in length, and particularly how proper nouns that may not be present in the LLM are assessed and what the effect on the measure may be. Additionally, the research evaluated meaning preservation only

within individual sentences. Assessing preservation of meaning across multiple sentences or entire conversations may better reflect the reality of conversational interaction.

Future research includes assessment of ASR captions in languages other than English and determining a measure to assess meaning preservation at conversational level.

Chen et al., (2021) conducted research to improve the accuracy and speed of training for Parrotron, an ASR for non-standard speech that repeats (or 'Parrots') natural speech in a synthesised voice – this technology is known as voice conversion (VC). Recordings of eight plwMND were used in training the ASR model, as part of a larger corpus of more than two thousand participants. The results showed a significant relative reduction in WER of up to 53% for dysarthric speech. The time required for ASR training reduced by 22% without degradation in accuracy, supporting improved scalability to a large set of users.

The researchers pointed out that the majority of the data used in this study was derived from anonymised and transcribed Google voice search traffic, which may not be representative of how Parrotron could be used in practice. No future research was reported by the authors.

Doshi et al., (2021) extended the work of Chen et al., (2021) on the Parrotron ASR to evaluate the benefits of an additional text decoder and the effectiveness of a customized VC for severe dysarthria. They reported a significant reduction in WER from 71.3% to 16.9% on average for all the participants with dysarthria when compared to standard ASR models, achieved with relatively few recordings of dysarthric speech. Similar to Chen (2021), the majority of the data used in this study was derived from anonymised and transcribed Google voice search traffic,

which is unlikely to be representative of how Parrotron may be used in practice. Future research will be to explore prosody transfer and target voice personalization. Prosody transfer refers to emulating the prosodic characteristics of the source dysarthric speaker's speech in the Parrotron voice conversion (VC) and Target voice personalization refers to emulating the dysarthric speaker's voice in the Parrotron VC (without dysarthria).

Calvo et al., (2021) evaluated a new mobile and personal speech assistant (mPASS) ASR platform designed to caption speech with dysarthria. The study compared the accuracy of mPASS to Pocketsphinx, a commercially available speaker-independent ASR system when recognizing words and sentences spoken by individuals with and without dysarthria. Five plwMND with 10 others with dysarthric speech and 20 individuals with no dysarthria recorded 24 words and 5 short sentences in a clinical environment. mPASS was found to be significantly more accurate than Pocketsphinx. The word recognition rate (WRR) for mPASS was 88.7% and for Pocketsphinx was 63.7%. The researchers plan to using mPASS to explore creating individualized applications tailored to the users' needs.

However, the recorded samples were from the 'Speech Intelligibility Test' (SIT), comprising single words up to eight word phrases increasing in complexity. This is not the same as conversational speech and may not include words or commands that plwMND may use to interact with digital assistants. Therefore the results may not be indicative of plwMND may actually use ASR systems.

Future research included exploring how mPASS may be used to create individualized applications tailored to the users' needs.

Marini et al., (2021) created a database of dysarthric speech in Italian called IDEA, containing 16,794 recordings from 45 speakers of whom seven are plwMND. The database includes multiple recordings of 211 isolated common words chosen to cover all Italian phonemes. The researchers assessed two ASR systems trained with the IDEA dataset – one using Gaussian Mixture Model (GMM) and the other utilizing a Deep Neural Network (DNN). The GMM achieved an average WER of 11.75% and the DNN 14.99% WER when assessed with the unseen TORGO dataset.

However, as the research was assessed with the very few recordings of plwMND in the TORGO database, it is difficult to predict how accurate this ASR will be in everyday communication.

Future research includes increasing the number of recordings to train the ASR systems, and investigating including additional clinical information into the ASR training process.

Mulfari et al., (2022) and Mulfari et al., (2023) Investigated the use of deep learning-based approaches in ASR for people with dysarthria, specifically in the context of telerehabilitation-based speech therapy. They focused on building accurate ASR recognition of a set of 13 single words for native Italian speakers, selected by the researcher. When these words were trained in an individual personalised ASR, this resulted in an average 7% WER, comparing favourably with other ASR research that attained around 15% WER.

The researchers argue that the tool can be used for personalized articulation therapy that can be independently practiced with continuous monitoring and immediate feedback from the ASR.

Because the process can be remotely supervised by SLTs, the author argues that the tool will enhance the effectiveness, reach, and affordability of speech therapy for individuals with dysarthria.

The researchers also suggested that accurate single word ASR recognition can help interaction with virtual home assistants (VHAs), such as Amazon Alexa, supporting environmental control using voice, for example a smart home. Additionally, accurate ASR recognition of a single command in dysarthric speech could be mapped to a customized speech message spoken aloud via a computer generated voice, supporting multiword communication.

However, the researcher acknowledged that the functional benefit of the tool is limited because only 13 individual words pre-selected by the researcher are recognised by the ASR, and not generalizable to any other words.

Future research includes creating a central repository of speech samples from contributors to develop an ASR model that is more accurate for dysarthric speech without needing personalization, and to support multiple languages.

Geng et al., (2022) assessed methods for improving ASR accuracy for speech with dysarthria and elderly speech. They trialled 'Spectro-temporal deep feature adaptation' and used the TORGO dataset to assess with dysarthric speech. The best performing configuration achieved a statistically significant overall WER reduction from 9.47% to 7.87%. However the testing was completed on short phrases only and the researchers planned to assess their approach on conversational speech in a future study.

Hu et al., (2022) investigated the use of articulatory features to improve the accuracy of ASR for dysarthric speech. Articulatory movements are argued to be more consistent than the actual sounds a person makes, however it can be difficult to collect these data as intrusive electromagnetic articulography (EMA) or magnetic resonance imaging (MRI) is required. The authors proposed to generate articulatory features from speech recordings using cross-domain acoustic-to-articulatory (A2A) inversion. This approach estimates articulatory information from speech recordings using data-driven artificial neural network models.

The A2A model was trained on the TORGO corpus, and then adapted to the UASpeech corpus. They report that using A2A-generated articulatory features consistently improved ASR performance, at best achieving a WER 24.82% on the UASpeech dataset, beating the lowest published WER on this benchmark task.

However this research only referenced one set of recordings from a single plwMND in the TORGO database and limits the generalizability of these findings. Additionally, given that declining respiratory support is a common feature of MND and can contribute to the effect of dysarthria by reduced power of voice, and by not including this consideration the research may potentially overstate the benefits of the A2A model on ASR accuracy in everyday conversation.

Finally, the proposed approach was not tested by recordings of plwMND with differing severity of dysarthria. No future research was detailed in this paper.

Lin et al. (2023) utilized an approach called 'intended phonological perception method' (IPPM) to estimate the intended phonological features of dysarthric speech by analysing the articulatory features extracted from a speech signal. These estimated features were used to

correct ambiguous phoneme decisions in the ASR system. The authors evaluated their proposed method using the TORGO database and the UASPEECH corpus.

The authors also proposed a 'staged' knowledge distillation (KD) approach to address the relative paucity of data recordings of dysarthric speech to train ASR models. This technique involves transferring knowledge from a model with a large dataset of normal speech to a smaller model for dysarthric speech. The staged aspect attempts to avoiding overfitting to less relevant details while keeping the features of dysarthric speech.

IPPM combined with KD was found to be effective in improving ASR recognition of dysarthric speech. Using this approach on the TORGO dataset, the average phoneme error rate (PER) was 20.47%, comparing favourably with the baseline PER 33.08%. PER measures the percentage of phoneme errors made by the ASR system compared to the ground truth transcription.

The authors argued their approach can improve the accessibility and usability of ASR technology for people with dysarthria. However this research only referenced one set of recordings from a single plwMND in the TORGO database and limits the generalizability of these findings. Additionally, similar to Hu et al., (2022), the model did not take into account that declining respiratory support is a frequent feature of MND and can contribute to the effect of dysarthria by reduced power of voice. Finally, their solution was not tested by plwMND with differing severity of dysarthria.

For future research, the authors planned to focus on reducing ASR recognition errors resulting from 'improper breathing' by people with dysarthria.

In order to increase the data available to tune ASR models to better recognise dysarthric speech, Naeini et al., (2024) investigated using three datasets of recordings:

(1) healthy speakers imitating simple elements of dysarthric speech (variation of rate and loudness).

(2) synthesised dysarthric speech - 3,663 artificially produced human male and female using Google text-to-speech (TTS) tool and Tacotron2. 55 accents were added to the synthesised voices.

(3) recordings from 13 plwMND, and other aetiologies.

All three datasets contained multiple repetitions of a single sentence ‘‘Buy Bobby a Puppy’’ (BBP). The healthy voices and synthesised voices repeated the phrase BBP 1-3 times with normal volume and 3 speaking rates: slow, normal, and fast..

Three measures were used: WER, Correct Count Accuracy (CCA) (which is the percentage of recordings where the number of repetitions of the phrase BBP is correctly counted by the model), and Intersection over Union (IoU) – which is used to measure how well the computer's prediction of where words start and end in a speech recording matches the ‘ground truth’ manual annotation. A high IoU means the computer's prediction is close to ground truth.

Fine-tuning the ASR model with the imitated and synthesised dysarthric speech datasets improved its performance. The best performing ASR configuration achieved average WER 5.7%. vs baseline WER 17.1% (the pre-trained model without fine-tuning), CCA of 94.4%, vs baseline CCA 80.9% and IoU 89.62% vs baseline IoU 64.2%.

The researchers concluded that in the context of sparse availability of 'genuine' dysarthric speech recordings, use of imitated and synthetic dysarthric speech data may be a viable alternative for improving ASR performance.

However, the imitated/synthesised dysarthric speech was only a partial emulation of simple dysarthric symptoms and did not replicate the full complexity and variation of dysarthric speech – it is unclear how valid this approach is as a substitute for recordings of dysarthric speech. Additionally, 55 different accents were used in the synthesised speech generation – yet there was no discussion on the similarities/differences of accent and dysarthria or the appropriateness of utilising accent to measure ASR effectiveness for dysarthria. Finally, recordings of a single short phrases used. The results may not be indicative of performance for more realistic speech in conversational settings using a variety of words and phrases.

Future research was planned to compare the effect of ASR fine-tuning using different amounts of healthy control recordings with those with dysarthria.

Okafor & Ludi (2022) created a voice-enabled version of 'Blockly', a 'Block-based programming environment' (BBPE) educational tool used to help people learn programming by using visual blocks instead of typing code. They did this to help make Blockly more accessible to users with upper limb motor impairment (ULMI). The objectives were to valuate usability and collect feedback from participants. There was one participant with MND and eight others also with upper-limb motor impairments (ULMI) living with cerebral palsy or muscular dystrophy. Their overall experience (out of five – five being the best) and SUS (System Usability Scale) (out of 100 – 100 being the best) scores were collected. They also provided open-ended responses

about their likes, dislikes, and suggestions for the system. The averages SUS score was 66/100 and average experience score was 3.55.

However the plwMND scored SUS as 50 - the second lowest of the cohort and the and the experience as 2 out of 5 - the lowest of the cohort. The researchers do not provide details behind theses scores. Additionally, almost 70% of participants reported that speech recognition did not work properly.

Future plans are to add alternative voice commands to try and address difficulties with the ASR. After the changes have been implemented another usability study will be conducted.

- Using ASR to support diagnosis

Bhattacharjee et al., (2021) explored the use of voice recordings to detect MND and Parkinson's Disease (PD). The researchers focused on pitch and Mel-frequency cepstral coefficients (MFCCs) - MFCCs capture the unique sound patterns of a person's speech. They wanted to assess how well these features could tell the difference between healthy people and those with MND / PD, especially when the voice recordings had background noise. Recordings from 59 plwMND were used. The participants spoke six different languages in approximately equal proportion (Bengali, Hindi, Odiya, Tamil, Telugu and Kannada). Spontaneous speech recordings about one minute long were collected from each subject as they talked about a festival they celebrate and a place that they have recently visited.

When comparing pitch and MFCC features to classify MND vs. Healthy Controls (HC), the study found that pitch and MFCCs were equally good at detecting plwMND - when the recordings were clear. However, when the recordings were noisy, pitch was more reliable than MFCCs. As

pitch is easier to measure than MFCC the researchers contend this approach is a simpler to detect MND/PD using voice recordings, even in noisy environments. Additionally, as recordings of conversational speech were used across diverse languages, the authors argue the results may generalize across other languages too.

The study only assessed recordings with additive white noise. Other types of noise (including background talking, environmental sounds) was not researched. A future study was planned to examine the noise robustness of different speech features in noisy recordings.

Tena et al., (2021) developed a method for diagnosing bulbar involvement in plwMND using machine learning techniques that analyse the acoustic features of speech of plwMND and control subjects. The automated model showed promising results, accurately predicting bulbar involvement with an accuracy of 95.8%. The researchers suggest that the automated model could help with early detection and better management of MND, potentially leading to earlier intervention and treatment. However, the researchers recognized that the sample size of the study was small and limits the generalizability of the results. Additionally the study was limited to Spanish-speaking participants, and the findings may not be applicable to other languages.

Future research included identifying methods to improve identification of incorrectly undiagnosed bulbar-involvement in patients with MND and increasing the sample size of voice recordings for analysis.

- Using ASR to support SLT assessment

Venugopalan et al., (2021) developed an automated way of classifying intelligibility of dysarthric speech, grouped by five severity categories: mild, moderate, severe and profound. The recordings used had already been scored in the same groupings separately by SLTs. The recordings were from 661 participants, including 162 plwMND. Each participant recorded 29 phrases provided to them. The ASR System Encoder (ASR-enc) model performed best across the different assessments, agreeing with SLT severity categories in 77% of all recordings. The results showed that longer phrases may be more indicative of intelligibility deficits than single words. The researchers argued that using deep learning techniques may support faster and more consistent diagnosis of dysarthria. Although the results indicate that the models can detect mild disordered speech, the researcher indicated that this might also an indicator that the range of mild intelligibility scored by the SLTs is large. Future research includes assessing if conversational speech with dysarthria can be successfully classified in the same way.

Venugopalan et al., (2023), building on earlier work, further developed automated classifiers to detect dysarthric speech, grouping by severity into mild, moderate, severe and profound. The recordings used had been scored separately by SLTs in the same way.

The researchers trained three models on the Euphonia dataset using 551,176 disordered speech samples contributed by 468 speakers (nearly a quarter of the recordings were from plwMND). The trained models were then evaluated on unseen external datasets, including TORGO, UASpeech and ALS-TDI PMP. The models were reported to categorize severity well on these external datasets without further specific training . The best performing model achieved an accuracy of over 90% when compared to SLT scores. The researchers acknowledged that

the ASR models had difficulty classifying speakers who had dysarthria but typical speech intelligibility. Additionally, the models were not assessed for speech in noisy backgrounds.

Future research includes plans to include conversational unprompted speech, to diversify the dataset in terms of ethnicity and aetiology, and include non-speech samples as well as noisy backgrounds.

Soleymanpour et al., (2021) reported on an automated method of assessing dysarthria severity and speech intelligibility using machine learning at phrase level, as opposed to the more usual rating generalised to speaker level. Averaged Ranking Score (ARS) and Intelligibility Probability Distribution (IPD) were used as measures. ARS estimates dysarthria severity for individual utterances, between 1 (normal) and 4 (medium severity), and IPD assesses overall speaker intelligibility based on the distribution of intelligibility across their utterances. The TORGO and UASpeech datasets were used and recordings were scored for severity of dysarthria by a SLT using the Frenchay Dysarthria Assessment. ARS for dysarthria severity generally aligned with the labelled severity levels in the TORGO dataset. However, it also provided more granular assessment of dysarthria severity at phrase level.

This research referenced one set of recordings from a single plwMND in the TORGO database and may limit the generalizability of these findings for plwMND. Additionally recordings of a single words and phrases were used, and recorded in a quiet environment, therefore the results may not be indicative of performance in conversational settings or in noisier environments common in daily life. No future research was detailed in the paper.

Rong (2021) developed a method to assess dysarthria for plwMND, combining measurements of muscle activity, jaw movement, and speech sound recordings to differentiate between plwMND and healthy controls. Specific patterns such as decreased control of jaw movement and increased irregularity in muscle activity were identified for plwMND when compared to healthy controls. The researchers argue this method may be a useful tool for earlier detection of dysarthria in plwMND, potentially allowing for earlier intervention. The researchers acknowledged that there were a small number of participants in the research, limiting the generalizability of the findings and further research is needed to validate the findings and assess the applicability of the approach.

Future research included how to apply this framework to track speech decline over time for plwMND, and how it may be applied to different aetiologies.

Mendoza et al., (2022) aimed to develop an objective measure of the degree of phoneme distortion in individuals with dysarthria as a tool to support therapeutic intervention. Accurate assessment of dysarthria is argued to inform focused clinical intervention with the goal of improving speech intelligibility and communication. The researchers trained phoneme-dependent and phoneme-independent ASR models using Dutch speech samples from 35 subjects with mild to severe dysarthria, including two plwMND. The results showed that the phoneme-independent classification model demonstrated good accuracy (73.5%) in predicting the degree of phoneme distortion in dysarthric speech.

The researchers noted that there was very limited data to train the ASR model which limits the generalizability and reliability of the findings. Additionally, there was a risk of overestimation

of production accuracy, which may lead to unrealistic expectations and misinterpretation of the findings. The study was not peer-reviewed.

Future research includes improving the ASR prediction models, increasing the dataset size and diversity and investigating clinical utility and impact.

Joshy & Rajan, (2022) assessed different ways to automate assessment of dysarthria severity comparing deep learning models, including a Convolutional Neural Network (CNN) and a Deep Neural Network (DNN). Recordings from the TORGO database were used from eight participants. One participant had been diagnosed with MND and the remainder were diagnosed with cerebral palsy. No details were provided about which participant had MND, gender or level of dysarthria. The results show that the DNN outperforms other systems, and was consistent with TORGO's own dysarthria severity score for each recording 93.97% of the time. This research is argued to help SLTs evaluate dysarthria in a more consistent and reliable way, supporting focused assessment and intervention planning.

However this research only referenced one set of recordings from a single plwMND in the TORGO database and may limit the generalizability of these findings. Additionally, it is difficult to understand what the benefits may be for SLTs in practice as there was there was no validation with SLTs as part of the research.

Future research includes identifying ways to increase accuracy using additional technology.

Bhattacharjee et al.(2023) argue that dysarthria assessments by SLTs are time-consuming and subject to variability. They proposed an automated and more consistent dysarthria severity

prediction system. The research uses 'transfer learning' to address the limitations posed by data scarcity and improve the generalizability and accuracy of severity classification models (transfer learning is a machine learning technique where a model trained on one task is re-used on a different but related task). The objective was to explore the use of transfer learning to improve ASR dysarthria severity classification for plwMND when compared the manual assessments by SLTs.

Recordings of 120 plwMND patients were used and assessed for dysarthria severity by SLTs. 74 male 46 female, with an average age of 55 years. The type of MND was not specified. The participants spoke five different languages in roughly equal proportion (Bengali, Hindi, Tamil, Telugu and Kannada). Recordings of spontaneous speech recordings were used as each plwMND talked about a festival they celebrate and a place that they have recently visited for approximately one minute each in their native language.

The results showed that transfer learning improved the accuracy of dysarthria severity classification, particularly for mild cases. The best configuration achieved average dysarthria classification accuracy of 80.11% (11.03% higher than the alternative without transfer learning).

However, the researchers acknowledge that the distribution of dysarthria severity classes was uneven, with more plwMND in the ‘normal’ class than in the ‘mild’ or ‘severe’ classes. This imbalance may affect the model's performance, particularly in underrepresented classes. Additionally, it is not clear what the benefits may be for SLTs in practice as there was no validation with SLTs as part of the research.

Future plans were to explore different machine learning configurations and trial a five category dysarthria severity classification, mirroring the Revised Amyotrophic Lateral Sclerosis Functional Rating Scale (ALSFRS-R) speech assessment sometimes used by SLTs.

Kumar et al. (2023) Investigated the impact of dysarthria caused by MND on the discriminability of vowels and fricatives, comparing the performance of deep neural networks (DNN) and manual assessment for classification of severity. The goal was to develop an objective and automated assessment of dysarthria severity, potentially leading to the development of diagnostic tools to support SLT intervention planning. The research found that manual classification performed better than automatic classification for most severity levels. However, the difference between the two gradually reduced as the severity of dysarthria increase. No future research was detailed in the paper.

Gutz et al., (2022) assessed the clinical validity of Google Cloud ASR to evaluate severity of dysarthria for plwMND and to provide guidance for potential users of such systems in research and clinical settings. They found that Google Cloud ASR was inadequate for detection of early impairment of speech and grading of speech severity and should not be used for transcription of dysarthric speech, scoring of speech intelligibility or replacement of clinician dysarthria severity ratings. However the researchers added that these findings cannot be generalized to other ASR systems because the underlying technology may be different.

Future research was planned to explore how ASR systems perform on other clinical populations, including Parkinson's disease.

Wiepert et al., (2023) examined the relationship between speech abnormalities and reduced ASR performance. The researchers categorized ASR transcription errors based on their source, rating the degree of error per category, and assessing the relationship between the WER and different sources of error, to examine how well an ASR system can capture pathological speech abnormalities. Other factors were also explored, including how the choice of ASR model and utterance complexity of speech tasks may impact transcription accuracy in clinical populations. The research used 656 recordings from 236 participants with impaired speech. The number of plwMND included in the 236 participants was not detailed, or their profile.

Four ASR models were used to assess the impact of ASR model type on transcription errors. Wav2vec 2.0 architecture (Base, Large, and XLS-R) and the Whisper model. Transcriptions generated by the ASR models were compared with the target transcriptions, and transcription errors were classified into categories: speaker abnormality, normal variation, other audio issues, and model errors.

The results showed that the number of ASR Model errors were consistently larger than speaker abnormality errors, and varied between ASR models. WER also differed across ASR models, likely due to what was not transcribed rather than what was transcribed incorrectly. WER also increased with utterance complexity and varied across ASR models. The research concluded that although ASR can capture useful information about speaker abnormalities, errors generated by the ASR model itself are significant and need to be taken account of.

The authors acknowledge that the research was based on a limited sample of recordings and focused on specific ASR models only. The recordings used were 'clipped' from speech

assessments and may not be indicative speech in conversational settings using a variety of words and phrases.

Future research was planned to explore the interaction between perceptual domain of impairment (articulation, prosody, etc.) and utterance complexity and ASR model type. Additionally to examine the effect of noise on ASR performance. Finally, explore improving ASRs through fine-tuning of the model using transcriptions at character/phoneme level.

- Using ASR as AAC

Cai et al.,(2021) developed a voice-activated switch that was trained to detect the vowel sound /a/ produced by speakers with severe dysarthria. They trained the ASR for the switch on a large dataset of typical voices and then fine-tuned it on a smaller dataset of dysarthric voices.

This approach significantly improved the switch's accuracy, significantly reducing the rate of false positives per hour (FPPH) from three to around one. The trained switch was tested with a plwMND and concluded that it may be a useful aid to communication or environmental control for individuals with severe dysarthria. However, the researchers acknowledge that there were limited examples of the target vowel sound from impaired speakers, which may affect the performance of the ASR. Additionally, the switch can only trigger to a single vowel sound, limiting the range of utterances and actions that can be recognized in practice.

Future research is planned to train the detectors for other simple sounds, including multiple vowels. Additionally, training an ASR model tolerant to alterations in voice quality caused by

assisted breathing devices and progressive dysarthria. Finally, exploring the adaptation of the ASR to included keyword recognition for less-impaired speakers.

Valencia et al. (2023) explored the potential of large language models (LLMs) to speed and enhance AAC-based communication. They developed "speech macros," which are natural language prompts designed to transform abbreviated user input into full sentences using LLMs. The objectives of the research were to

(1) identify the benefits for AAC users, if any, of directly interacting with LLMs,
(2) How AAC users evaluate communication suggestions made by a LLM, and
(3) What concerns AAC users have about integrating LLMs into their own AAC devices.

There were 12 participants in the study, all AAC users. There were plwMND participants but no details were provided on the number, type of MND age or gender. A mixed-methods approach was used, combining qualitative and quantitative data collection and analysis.

Three interactive 'Speech Macro' web-based prototypes were developed,

(1) 'Extend Reply' - takes a short input from the participant and extends it with more detail.
(2) 'Reply with Background Information' - enables participants to input personal information about themselves beforehand. When asked a question during a conversation, the macro generates responses based on this pre-provided information.
(3) 'Turn Words into Requests': transforms single words into sentence level requests.

The researcher arranged a 90-minute remote video call where each participant interacted with the prototypes and provided feedback. Participants then completed a survey containing multiple-choice and open-ended questions.

Qualitative data analysis was performed on the transcripts of the video calls and survey responses, identifying themes related to feature suggestions, use cases, quality of suggestions, and concerns about AI in AAC.

Most participants found speech macros to be either 'very' or 'extremely' useful and many were positive about the potential to reduce the physical and cognitive effort required to communicate. Some participants envisioned using AI for various purposes beyond the tested speech macros, including interacting with voice assistants, managing routine tasks, and preparing for medical appointments.

However, many participants were concerned that AI-generated phrases may sound inauthentic or misrepresent their intentions. They were also concerned about data privacy and security of their personal data used by AI systems. Additionally, the AI model struggled to generate appropriate responses for sensitive topics or specific needs, such as medical requests.

There was no opportunity to trial the speech macros in everyday communication, and the prototype used in the study was intentionally simplified, therefore further analysis is needed to better understand how the app may work in the real world.

Future research includes (1) enabling users to control, override, or correct the AI's output whenever they wish, (2) assessing how participants to interact with the AI model in functional communication, (3) refining the speech macro to improve the quality and relevance of generated suggestions (4) co-design with people that use AAC to develop the interface interfaces.

Angrick et al. (2024) used an implanted brain-computer interface (BCI) to reconstruct and synthesize speech using brain activity recorded with intracranial electrodes. Recurrent neural networks (RNNs) were used to identify, decode and synthesize speech from electrocorticographic (ECoG) signals from the motor, premotor and somatosensory

cortices.

One plwMND with bulbar MND participated in the study. The research focused on six keywords that the participant could reliably produce using natural speech six months prior to the study commencing ("Left", "Right", "Up", "Down", "Enter", "Back", and "…" for silence). These were chosen to control a communication board.

Five and a half months prior to the study, spoken training data was collected for ten days over six weeks - during which the ECoG neurological signals were mapped to the speech produced. Each key word was spoken ten times (60 words in total). Later, during the study the ECoG mapping was used to generate synthesized output while the plwMND attempted to articulate the same keyword reading task. The intelligibility of the synthesized speech generated from the BCI was evaluated through listening tests with human listeners, who were able to correctly recognize 80% of the words.

Future research is planned to expand the limited vocabulary and explore different approaches to extrapolate to words that are not in the vocabulary of the training set.

- How plwMND already use ASR technology

Wallock & Cerny, (2021) surveyed 19 plwMND on the benefits of smart home technology as assistive technology. All reported using used various smart home devices on a daily basis. Smart home devices were reported to provide plwMND with greater control over electronics, increasing functional independence and regaining the ability to complete some tasks they were unable to do. PlwMND also used smart home devices daily for recreation, communication, and home management. However the researchers recognized that the sample size was small and may limit the generalizability of the findings. Additionally, the outcomes were based on self-reported data, which may be subject to biases and inaccuracies.

Future research is planned to investigate the impact of smart home technology on the emotional and physical well-being of plwMND.

Table 6: Study types, participants, technology used, method.

Group	N	Citation	Study Type	Objectives	Participant details (e.g. type of MND age/sex and number), Technology used, Method
ASR Accuracy	1	Disordered Speech Data Collection: Lessons Learned at 1 Million Utterances from Project Euphonia MacDonald et al., (2021) USA	Primary Research peer reviewed	(1) Provide an overview for Project Euphonia, which includes over 1 million utterances (>1300 hours) of disordered speech from individuals with impaired speech including plwMND (2) Review the key lessons learned from the data collection process (3) Explain the challenges for building personalised ASR models . (4) Discuss the objectives of Project Euphonia	(1) plwMND are the largest cohort recording phrases for Project Euphonia, with 173 plwMND recording over 300 phrases (total number of plwMND that recorded for Euphonia is not stated) (2) No MND type, gender or severity of dysarthria was specified. (3) A browser-based data collection tool was used to record individual scripted utterances usually without special equipment or intervention from the research team. (4) The recordings were often completed in participants' home environments using devices such as phones, tablets, laptops, or desktop computers.
ASR Accuracy	2	Automatic speech recognition of disordered speech: Personalized models outperforming human listeners on short phrases Green et al., (2021) USA	Primary Research peer reviewed	(1) evaluate accuracy of personalized ASR for recognizing disordered speech from short phrases from 432 people with dysarthria including 172 plwMND (2) benchmark personalized ASR models against expert human transcribers and speaker-independent ASR models trained on typical speech. (3) To assess the impact of participant characteristics, atypical speech characteristics, and technical factors on recognition accuracy.	(1) Recordings from 432 speakers, all of whom recorded 300 or more utterances, including from 172 plwMND. (2) No MND type, gender or severity of dysarthria was specified. (3) Short phrases of 2-4 words were chosen. (4) WERs were estimated for three different ASR models and for human transcribers and broadly across aetiologies. (5) Two speaker independent ASR systems were trained and optimized for typical speech, and ASR models personalized to the speech of each of the 432 participants.
ASR Accuracy	3	Personalised Automatic Speech Recognition trained on small disordered speech datasets Tobin & Tomanek, (2022) USA	Primary Research peer reviewed	(1) This study investigates the performance of personalized ASR for recognizing disordered speech using small amounts of per-speaker data.	(1) the number of plwMND included in the 195 participants was not detailed, or their profile. (2) Personalized ASR models were trained for 195 individuals with training sets ranging in size from <1 minute to 18-20 minutes of speech data. (3) Word error rate (WER) thresholds were selected to determine 'Success Percentage' (the percentage of personalized models reaching the target WER) in different application scenarios. (4) For the 'home automation' use-case (recorded phrases that reflected frequently used commands to digital assistants), 15% WER was argued to corresponded to a success rate of 80% when engaging with Google Assistant. For the conversational use-case, a target of 20% WER set due to less structured, longer more complex phrases typical of conversational speech. (5) The researchers assumed both WER thresholds would achieve functional benefit to people using the ASR.
ASR Accuracy	4	Assessing ASR Model Quality on Disordered Speech using BERTScore Tobin et al., (2022) USA	Primary Research peer reviewed	This study investigates the use of BERTScore as an alternative metric to WER for assessing ASR models on disordered speech. BERTScore is proposed as a more nuanced measure that captures semantic similarity between the ASR output and the ground truth transcript.	(1) Recordings of 15 individuals from Euphonia, including plwMND, Cerebral Palsy, Down Syndrome. (2) No details of plwMND were provided. (3) The research involved a detailed analysis of transcription errors from the personalized ASR models. SLTs categorized the errors and assessed their severity based on the preservation of meaning. The study then compared how well BERTScore and WER aligned with the SLT assessments. (4) SLPs assessed a total of 3473 transcription errors and determined error type and assigned an assessment rating of the error severity. Errors grouped by type e.g. deletion, contraction, normalization, homophone, spelling, proper noun, repetition. (5) The error severity assessment was rated on a scale of 0, 1 or 2 depending on whether or not an error preserved meaning. 0 and 1 errors are judged as 'recoverable', where a communication partner would be able to understand what a speaker meant to say. An error level 2 represents a communication breakdown, where the perceived meaning is lost or drastically altered from a speaker's intent.
ASR Accuracy	5	An Analysis of Degenerating Speech Due to Progressive Dysarthria on ASR Performance Tomanek et al. (2023) USA	Primary Research peer reviewed	(1) Analyse the change in performance of ASR over time in individuals with degenerating speech due to MND. (2) Explore mitigation strategies to optimize recognition throughout the progression of the disease and improve the ASR performance.	(1) Recordings from four adults with progressive dysarthria due to MND. (2) Type of MND and gender were not specified. (3) They were selected from the Euphonia dataset and chosen based on their significant decline in ASR performance over time and an increase in speech severity between their first and last recording sessions. (4) Short phrases of 3-5 words in length were chosen. (5) Word error rate (WER) was calculated per severity bin. (6) End-to-end ASR models based on the RNN-T architecture were used. The models included an encoder network with 8 layers and a predictor network with 2 layers of unidirectional LSTM cells. (7) The ASR models were trained on typical speech data and fine-tuned on the Euphonia dataset. (8) Adapted Speaker Dependent (A-SD) models were personalized using data from the specific speaker
ASR Accuracy	6	Large Language Models As A Proxy For Human Evaluation In Assessing The Comprehensibility Of Disordered Speech Transcription Tomanek et al. (2024) USA	Primary Research peer reviewed	This research explores the use of large language models (LLMs) as a proxy for human evaluation in assessing the comprehensibility of ASR captioning of speech with dysarthria. (1) to assess if LATTEScore (("LLMs to Assess Transcription Errors Score") is better at measuring the performance of ASR in transcribing disordered speech with dysarthria than WER, BERTScore and others.	(1) Recordings from 49 participants in the Euphonia dataset were used to test, including 8 plwMND. (2) No details were provided about the plwMND (3) Two SLTs provided severity ratings on the ASR transcriptions at phrase level on a 3 point scale, with 0 indicating full meaning preservation between ground truth and ASR transcript, 1 indicating errors but meaning mostly preserved, and 2 indicating significant change in meaning. (4) WER, BERTScore and the LLM-based models were assessed against the 'gold standard' SLT assessments of meaning preservation in the ASR Captions. (5) LATTEScore was used a measure to distinguish between the model results based on accuracy of meaning retention
ASR Accuracy	7	Conformer Parrotron: a Faster and Stronger End-to-end Speech Conversion and Recognition Model for Atypical Speech Chen et al.(2021) USA	Primary Research peer reviewed	Parrotron is an ASR for non-standard speech that can repeat natural speech in a synthesised voice - known as voice conversion (VC). The objectives are: (1) Improve Parrotron ASR/VC accuracy. (2) Improve speed of training of the ASR model to be scalable to a large number of users (3) Validate ASR/VC model updates on dysarthric speech conversion.	(1) Recordings of eight plwMND were used in training (over three hours of recorded speech) the ASR model, as part of a larger corpus of more than 2000 participants. (2) The Parrotron model was trained using the speech recordings, performance was analysed using Word error rate (WER) (3) An analysis of the trade-off between speed and accuracy of the ASR was completed.

Group	N	Citation	Study Type	Objectives	Participant details (e.g. type of MND age/sex and number), Technology used, Method
ASR Accuracy	8	Extending Parrotron: An End-To-End, Speech Conversion and Speech Recognition Model for Atypical Speech Doshi et al., (2021) USA	Primary Research peer reviewed	Parrotron is an ASR for non-standard speech that can also repeat natural speech in a synthesised voice - known as voice conversion (VC). The objectives are: (1) to evaluate the benefits of an additional text decoder (2) to evaluate a customized VC for severe dysarthria.	(1) Recordings of seven plwMND were used in training (over three hours of recorded speech) the ASR model, as part of a larger corpus of more than 2000 participants. (2) The Parrotron model was trained using the speech recordings, performance was analysed using Word error rate (WER)
ASR Accuracy	9	Evaluation of an Automatic Speech Recognition Platform for Dysarthric Speech Calvo et al., (2021) Italy	Primary Research peer reviewed	(1) This study evaluated mPASS (an ASR platform designed to caption speech with dysarthria) in comparison to Pocketsphinx, a commercially available speaker-independent ASR system. (2) The study assessed accuracy of each ASR for recognizing words and sentences spoken by individuals with and without dysarthria.	(1) Five plwMND (three male, two female) along with 10 others with dysarthric speech and 20 individuals with non dysarthric speech recorded 24 words and 5 sentences in a clinical environment. (2) Differences in recognition accuracy between two ASR systems were evaluated. (3) In addition, mPASS usability was assessed with a technology acceptance model (TAM) questionnaire. (4) word recognition rate (WRR) was used to assess the results.
ASR Accuracy	10	IDEA: An Italian Dysarthric Speech Database Marini et al., (2021) Italy	Primary Research peer reviewed	(1) IDEA is a database of dysarthric speech in Italian, containing 16,794 recordings from 45 speakers of which seven have MND. The database includes 211 isolated common words chosen to cover all Italian phonemes. The objectives are (2) To improve the performance of ASR systems for individuals with dysarthria. (3) To collect a comprehensive dataset of Italian dysarthric speech to facilitate research in assistive technologies. (4) To assess two ASR systems using a subset of the IDEA database	(1) Seven plwMND (three male, four female) (2) Each participant recorded 211 individual words three times each. The words were chosen to include at least one pronunciation of each Italian phoneme. (3) A subset of the recordings were used to train two different ASR models: Gaussian Mixture Model (GMM) based and Deep Neural Network (DNN) based (4) WER was used to assess the results.
ASR Accuracy	11	Deep learning applications in telerehabilitation speech therapy (Mulfari et al., 2022) Italy	Primary Research peer reviewed	(1) Investigate the use of deep learning-based approaches in ASR for people with dysarthria, specifically focusing on single word recognition for native Italian speakers. (2) To collecting recordings of dysarthric speech to train an ASR model to recognize keywords in atypical speech, (3) To evaluating the benefits of the trained ASR system in telerehabilitation speech therapy context.	(1) plwMND are the largest cohort recording phrases for the CapisciAMe dataset, recording around 3900 of 7269 phrases. (2) No MND type, gender or severity of dysarthria was specified. (3) For the 6 participants in the trial: individual testing datasets were extracted, consisting of ten repetitions of 13 keywords and subsequently 14 speech models were trained.
ASR Accuracy	12	Toward a lightweight ASR solution for atypical speech on the edge Mulfari et al., (2023) Italy	Primary Research peer reviewed	(1) Investigate the use of deep learning-based approaches in single word ASR for people with dysarthria / aphasia (2) To collecting recordings of dysarthric speech to train an ASR model to recognize keywords in atypical speech,	(1) plwMND are the largest cohort recording phrases for the CapisciAMe dataset of the total 24606 speech examples. (2) No MND type, gender or severity of dysarthria was specified. (3) For the 16 participants in the trial: individual testing datasets were extracted, consisting of ten repetitions for each of the 13 single keyword and 17 speech models were trained. (4) The results were assessed using WER.
ASR Accuracy	13	Speaker Adaptation Using Spectro-Temporal Deep Features for Dysarthric and Elderly Speech Recognition 'Geng et al., (2022) USA	Primary Research peer reviewed	(1) improve the accuracy of ASR for people with dysarthria and elderly individuals.	(1) Spectro-temporal deep feature adaptation is used to personalize the system for dysarthric and elderly speakers. (2) The TORGO dataset was used to evaluate the speaker adaptation technique. (3) TORGO has recordings from eight dysarthric speakers with varying levels of speech impairment, including one plwMND. (4) There are no details of MND type or gender.
ASR Accuracy	14	Exploiting Cross Domain Acoustic-to-articulatory Inverted Features For Disordered Speech Hu et al., (2022) USA	Primary Research peer reviewed	(1) Investigated the use of articulatory features to improve the accuracy of ASR for dysarthric speech. (2) Articulatory movements are argued to be more consistent than the actual sounds a person makes. (3) Articulatory features from speech recordings were generated using cross-domain acoustic-to-articulatory (A2A) inversion.	(1) Only one participant identified with MND in the dataset (TORGO). No MND type, gender or severity of dysarthria was specified. Other recordings from non-ALS participants were also used from TORGO and the UASPeech dataset (2) The A2A model was trained on the TORGO corpus, and then adapted to the UASpeech corpus. (3) WER was used to measure the ASR transcription accuracy of the UASpeech dataset
ASR Accuracy	15	Disordered speech recognition considering low resources and abnormal articulation Lin et al. (2023) USA	Primary Research peer reviewed	(1) The researchers used an approach called 'intended phonological perception method' (IPPM) to estimate the intended phonological features of dysarthric speech by analysing the articulatory features extracted from a speech signal. (2) These estimated features were used to correct ambiguous phoneme decisions in the ASR system. (3) 'Staged' knowledge distillation (KD) was used to address the relative paucity of data recordings of dysarthric speech to train ASR models. (4) KD involves transferring knowledge from a large dataset of normal speech to a smaller model for dysarthric speech. (5) The staged aspect attempts to avoiding overfitting to less relevant details while keeping the features of dysarthric speech.	(1) Only one participant identified with MND in the dataset (TORGO). No MND type, gender or severity of dysarthria was specified. Other recordings from non-ALS participants were also used from TORGO and the UASPeech dataset. (2) The experiments were conducted using the Kaldi speech recognition toolkit and a Transformer-based machine translation model. The input speech features were processed to capture different aspects of the sound. The models were trained to recognize phonemes and articulatory features. (3) Due to the limited amount of dysarthric speech data, the researchers used a technique called 'early stopping' to prevent the models from overfitting (becoming too specialized to the training data). They also averaged the parameters of multiple saved models to further avoid overfitting. (4) WER was used to measure the outcomes.

Group	N	Citation	Study Type	Objectives	Participant details (e.g. type of MND age/sex and number), Technology used, Method
ASR Accuracy	16	Improving Dysarthric Speech Segmentation With Emulated and Synthetic Augmentation Naeini et al., (2024) USA	Primary Research peer reviewed	(1) The researchers investigated the use of synthesised dysarthric speech and healthy speakers imitating possible elements of dysarthric speech (rate and loudness) to improve ASR performance.	(1) Recordings from 13 plwMND, age range 50-93, no details on type of MND or gender. Additional recordings from healthy participants and other aetiologies. (2) The study used three datasets: one of dysarthric speech, one of healthy individuals imitating possible elements of dysarthric speech (rate and loudness variation), and one of synthesised possible elements of dysarthric speech (3,663 artificially produced human voices using Google text-to speech (TTS) tool and Tacotron2. The synthesised voices of men and women had 55 various accents (3) All three datasets contain multiple repetitions of a single sentence ''Buy Bobby a Puppy'' (BBP). the healthy voices and synthesised voices repeated the phrase "Buy Bobby a Puppy" (BBP) 1-3 times with normal volume and 3 speaking rates: slow, normal, and fast.. (4) Three measures were used: (i) WER, (ii) Correct Count Accuracy (CCA) - the percentage of recordings where the number of repetitions of the phrase "Buy Bobby a Puppy" is correctly counted by the model and (iii) Intersection over Union (IoU) - used to measure how well the computer's prediction of where words start and end in a speech recording matches the manual annotation (ground truth). A high IoU means the computer's prediction is close to ground truth
ASR Accuracy	17	Voice-Enabled Blockly: Usability Impressions of a Speech-driven Block-based Programming System Okafor & Ludi (2022) USA	Primary Research peer reviewed	(1) Blockly is a 'Block-based programming environment' (BBPE) educational too and is used to help people learn programming by using visual blocks instead of typing code. (2) The researchers created a voice-enabled version of Blockly to make it accessible to users with upper limb motor impairment (ULMI) to enable increased accessibility and inclusion for learning new skills. (3) The objectives were to valuate usability and collect participant feedback	(1) One plwMND, female aged 63. Eight other participants also with upper-limb motor impairments (ULMI) living with cerebral palsy or muscular dystrophy. (2) The study was conducted remotely over Zoom, and participants completed tasks using their voice as the input modality. (3) Their overall experience and SUS (System Usability Scale) scores were collected, and they also provided open-ended responses about their likes, dislikes, and suggestions for the system.
Supporting Diagnosis	18	Effect of noise and model complexity on detection of amyotrophic lateral sclerosis and Parkinson's disease using Pitch and MFCC Bhattacharjee et al., (2021)	Primary Research peer reviewed	(1) This study explores the use of voice recordings to detect MND and Parkinson's Disease (PD) (2)The researchers focused on pitch and Mel-frequency cepstral coefficients (MFCCs) - which capture the unique sound patterns of a person's speech. (3) They wanted to see how well these features could be used to tell the difference between healthy people and those with MND / PD especially when the voice recordings had background noise.	(1) Recordings from 59 plwMND (38 male, 21 female, aged between 36-75). No type of MND was specified. (2) The participants spoke in roughly equal proportion six different languages (Bengali, Hindi, Odiya, Tamil, Telugu and Kannada). (3) Spontaneous speech recordings were collected from each subject as they talked about a festival they celebrate and a place that they have recently visited for approximately one minute each in their native language. (4) The researchers used pitch and MFCC features to classify MND vs. Healthy Controls (HC) and PD vs. HC. (5) The classifier operated on feature chunks obtained from speech segments. (6) The percentage of correct classifications made by the model was used to evaluate the ASR performance.
Supporting Diagnosis	19	Detection of Bulbar Involvement in Patients With Amyotrophic Lateral Sclerosis by Machine Learning Voice Analysis: Diagnostic Decision Support Development Study Tena et al., (2021) Spain	Primary Research peer reviewed	(1) To develop a method for diagnosing bulbar involvement in plwMND using speech analysis and machine learning techniques to analyse the acoustic features of speech recordings from plwMND patients and controls	(1) 45 plwMND (26 male and 19 female) aged from 37 to 84 years and 18 control subjects (9 males and 9 females). 14 pwlMND were diagnosed with MND with bulbar involvement (3 males and 11 females; aged from 38 to 84 years. 31 diagnosed with MND but did not display bulbar symptoms (23 male and 8 female, aged from 37 to 81 years. (2) The recordings used in the study were sustained samples of the Spanish vowels a, e, i, o, and u.
Supporting Assessment	20	Comparing Supervised Models And Learned Speech Representations For Classifying Intelligibility Of Disordered Speech On Selected Phrases Venugopalan et al., (2021)	Primary Research peer reviewed	(1) To find the most effective method for detecting and classifying disordered speech intelligibility using three different deep learning techniques	(1) 661 speakers of which 162 were plwMND (2) Each participant recorded the same 29 phrases provided to them. (3) SLTs listened to the recordings and assessed intelligibility using a five-point Likert scale. (4) The scale was mapped to five classes: typical, mild, moderate, severe, and profound. (5) The research used deep learning techniques to analyse and classify disordered speech. (6) Three different models were trained to classify the intelligibility of the speech. (7) A convolutional neural network (CNN) analysed the spectrogram of speech recordings predicting the intelligibility rating. (8) Pre-trained CNN models learned non-semantic speech representations. (9) Representations extracted from the encoder of an ASR system trained on typical speech and captured acoustic features (10) The models were evaluated on a dataset of 15,246 disordered speech samples from the 661 participants. The performance of the models was analysed based on different intelligibility classes of the speakers.
Supporting Assessment	21	Speech Intelligibility Classifiers from 550k Disordered Speech Samples Venugopalan et al., (2023) USA	Primary Research peer reviewed	(1) Develop automated classifiers for intelligibility of dysarthric speech grouping by severity into mild, moderate, severe and profound. (2) The classifiers attempted to accurately predict a 5-point rating of severity for each recording completed separately by speech and language therapists (SLTs)	(1) 551,176 disordered speech samples were contributed by a diverse set of 468 speakers. Nearly a quarter of the recordings were from plwMND. (2) No MND type, gender or severity of dysarthria was specified. (3) Classification models were based on different deep learning architectures. (4) The models trained on 550K samples to predict either a binary (typical, atypical speech) label, or the five class labels. (5) The performance of the models on dysarthric speech assessed using TORGO, specifically focusing on MND and CP.
Supporting Assessment	22	Increasing the Precision of Dysarthric Speech Intelligibility and Severity Level Estimate Soleymanpour et al., (2021) USA	Primary Research peer reviewed	(1) This research assessed an automated method of assessing dysarthria severity and intelligibility using machine learning at individual phrase level, as opposed to a generalised rating at speaker level. (2) Two outcome measures were used: Averaged Ranking Score (ARS): estimates dysarthria severity for individual utterances, between 1 (normal) & 4 (medium severity) Intelligibility Probability Distribution (IPD) assesses speaker intelligibility based on distribution of intelligibility across utterances.	(1) Only one participant identified with MND in the dataset (TORGO). No MND type, gender or severity of dysarthria was specified. Other recordings from non-ALS participants were also used from TORGO and the UASPeech dataset (2) Recordings were scored for severity of dysarthria by a SLT using the Frenchay Dysarthria Assessment. (3) A one-dimensional Convolutional Neural Network (1-D CNN) model was trained on the UASpeech dataset and then further trained on the TORGO dataset (4) A leave-one-speaker-out cross-validation procedure was used, meaning one speaker was excluded from training and used as unseen test data. (5) ARS and Intelligibility Probability Distribution were used to asses the model compared to the SLT assessment

Group	N	Citation	Study Type	Objectives	Participant details (e.g. type of MND age/sex and number), Technology used, Method
Supporting Assessment	23	A Novel Hierarchical Framework for Measuring the Complexity and Irregularity of Multimodal Speech Signals and Its Application in the Assessment of Speech Impairment in Amyotrophic Lateral Sclerosis Rong, (2021) USA	Primary Research peer reviewed	(1) To develop a new method for assessing dysarthria in plwMND by combining measurements of muscle activity, jaw movement, and speech sound. (2) To accurately distinguish between plwMND and healthy individuals. (3) To improve early detection and provide timely interventions for plwMND.	(1) 13 plwMND (eight men, five women; age: 38–74 years) and 10 neurologically healthy individuals. (2) The researchers collected data from muscle activity, jaw movement, and speech sound. The participants were asked to read sentences while their speech was recorded. The data was then analysed using machine learning algorithms to classify the speech samples. (3) Participants read three sentences ("Take today's tasty tea on the terrace"; "Clever Kim called the cat clinic"; "Buy Bobby a puppy") 3 times at their habitual loudness and rate, resulting in 69 speech samples.
Supporting Assessment	24	Automatic Score of Articulatory Distortion in Adults with Dysarthria Mendoza et al., (2022) USA	Primary Research not peer reviewed	(1) improve assessment and treatment of dysarthria (speech sound distortions), including for plwMND (2) Analyse the articulatory accuracy of speech sound production (3) Develop a new algorithm for objective estimation of phoneme distortion	(1) 2 plwMND (both Female, aged 84 and 85) out of 35 participants. No MND type specified. (2) The participants recorded a list of mono and bisyllabic words. The target phonemes were presented into initial, medial, and final positions within the stimulus words. (3)The articulatory accuracy of the speech sound production was rated by three SLTs, evaluating the general degree of distortion of the target consonants on a visual analogue scale (VAS) 0-10. (4) Different regression models and classification models were evaluated for predicting speech sound distortions.
Supporting Assessment	25	Automated Dysarthria Severity Classification: A Study on Acoustic Features and Deep Learning Techniques Joshy & Rajan, (2022) USA	Primary Research peer reviewed	(1) Focused on the classification of dysarthria severity levels using ASR deep learning techniques and acoustic features. (2) The researchers evaluate different technologies measuring specific features from prosody, articulation, phonation, and glottal functioning	(1) Only one participant identified with MND in the dataset (TORGO). (2) No MND type, gender or severity of dysarthria was specified. (3) The researchers used deep learning models, such as Convolutional Neural Network (CNN) and Deep Neural Network (DNN), to develop an automated system for assessing dysarthria severity. (4) Acoustic features extracted from speech recordings from TORGO and other datasets were used to train the models.
Supporting Assessment	26	Transfer Learning to Aid Dysarthria Severity Classification for Patients with Amyotrophic Lateral Sclerosis Bhattacharjee et al.(2023)	Primary Research peer reviewed	(1) The researchers argue that dysarthria assessments by SLTs are time-consuming and subject to variability. (2) They propose an automatic and consistent dysarthria severity prediction system is needed. (3) The study uses 'transfer learning' to address the limitations posed by data scarcity and improve the generalizability and accuracy of severity classification models. (4) Transfer learning is a machine learning technique where a model trained on one task is re-used on a different but related task. (5) Objective: to explore the use of transfer learning to improve ASR dysarthria severity classification with limited available data.	(1) 120 plwMND patients, assessed for dysarthria severity by SLTs. 74 male 46 female, average age 55 years, type of MND not specified. (2) The participants spoke in roughly equal proportion five different languages (Bengali, Hindi, Tamil, Telugu and Kannada). (3) Spontaneous speech recordings were collected from each subject as they talked about a festival they celebrate and a place that they have recently visited for approximately one minute each in their native language. (4) Additionally speech recordings from 88 healthy controls. (5) The study categorises dysarthria severity prediction into normal, mild, and severe. Machine learning was used using temporal statistics of mel-frequency cepstral coefficients (MFCCs) . Transfer learning approaches were trialled. (6) the automated dysarthria categorisations were compared to prior SLPs assessments using the speech component of the MNDFRS-R scale.
Supporting Assessment	27	Classification of multi-class vowels and fricatives from patients having Amyotrophic Lateral Sclerosis with varied levels of dysarthria severity Kumar et al. (2023) USA	Primary Research peer reviewed	(1) To investigate the impact of dysarthria caused by MND on the discriminability of different vowels and fricatives. (2) To compare the performance of automatic classification algorithms with manual classification (through listening tests) in classifying vowels and fricatives at different severity levels. (3) To assess the potential of deep neural networks and self-supervised speech representations for automatic classification of vowels and fricatives in dysarthria.	(1) 119 plwMND, 73 male 46 female, age between 23 and 81 years. (2) The plwMND spoke diverse languages including Bengali, Hindi, Tamil, Telugu, and Kannada. (3) Participants recorded sustained utterances of 4 vowels (/a/, /i/, /o/, /u/) and 3 fricatives (/s/, /sh/, /f/). (4) The research compared automatic and manual classification to analyse discriminability of vowels and fricatives in plwMND (5) Automatic Classification utilised Feature Extraction: Self-supervised speech representations were computed using pre-trained models. The models were trained on the speech data from plwMND and healthy controls. The performance of the classifiers was evaluated on a separate dataset used for manual listening tests. (6) Manual Classification: Each utterance was classified by three human listeners who provided a confidence score for their decision. The mode of the three decisions provided by the listeners was taken as the final manual classification result.
Supporting Assessment	28	Validity of Off-the-Shelf Automatic Speech Recognition for Assessing Speech Intelligibility and Speech Severity in Speakers With Amyotrophic Lateral Sclerosis 'Gutz et al., (2022) USA	Primary Research peer reviewed	(1) To assess the clinical validity of Google Cloud ASR to evaluate severity of dysarthria or speech intelligibility of plwMND. (2) Provide guidance for potential users of such systems in research and clinic.	(1) Speech samples collected from 52 plwMND and 20 healthy control speakers. (2) Recordings from The Speech Intelligibility Test (SIT) were used, varying from single words up to eight word phrases increasing in complexity. (2) Ten SLTs rated the speech samples and provided scores for severity of dysarthria and sentence intelligibility Word recognition rate (WRR)was compared between Google Cloud ASR to clinician provided judgement.
Supporting Assessment	29	Not All Errors Are Created Equal: Evaluating The Impact of Model and Speaker Factors on ASR Outcomes in Clinical Populations Wiepert et al. (2023) USA	Primary Research peer reviewed	(1) This study examined the assumption that speech analysis using ASR may have a close relationship between pathological speech abnormalities and reduced ASR performance. (2) The researchers categorized ASR transcription errors based on their source, rate the degree of error per category, and assess the relationship between the WER and different sources of error (3) The goal was to examine how well an ASR system can capture pathological speech abnormalities. (4) other key factors were explored, such as choice of ASR model and utterance complexity of speech tasks.	(1) The research used 656 recordings from 236 participants with impaired speech. The number of plwMND included in the 236 participants was not detailed, or their profile. (2) Recordings of three short phrases (5-6 words) 'clipped' from both typical and pathological speech assessments conducted in English were used, including recordings from speakers with dysarthria, apraxia of speech, dysphonia (3) recordings with noise distortion, strong regional accents, or non-English first language speakers were removed. (4) Four ASR models were used to assess the impact of ASR model type on transcription errors. Wav2vec 2.0 architecture (Base, Large, and XLS-R) and the Whisper model. (5) Transcriptions generated by the ASR models were compared with the target transcriptions and the manual transcriptions. Transcription errors were classified into categories: speaker abnormality, normal variation, other audio issues, and model errors. Each error category was rated on a scale from 0 to 3 based on the number of words impacted. (6) WER was calculated for each transcription
ASR as AAC	30	A Voice-Activated Switch for Persons with Motor and Speech Impairments: Isolated-Vowel Spotting Using Neural Networks Cai et al.,(2021) USA	Primary Research peer reviewed	(1) This research aims to switch access for individuals with severe/profound speech and motor impairments. (2) development of a voice-activated switch triggered by a specific vowel sound (/a/), which may be more accessible for people with severe / profound dysarthria	(1) four out of the five participants in the user testing had MND, while one participant had cerebral palsy. No gender information provided. (2) The researchers used a neural network ASR to detect this vowel sound and trained it with examples from typical speakers and impaired speakers. (3) The trained model can be used in a mobile app to activate preconfigured actions, such as alerting caregivers. (4) User testing showed that the switch successfully recognized the vowel sound and activated the desired actions.

Group	N	Citation	Study Type	Objectives	Participant details (e.g. type of MND age/sex and number), Technology used, Method
ASR as AAC	31	"The less I type, the better": How AI Language Models can Enhance or Impede Communication for AAC Users Valencia et al. (2023) USA	Primary Research peer reviewed	The researchers explored the potential of large language models (LLMs) to speed and enhance AAC-based communication. The authors developed a system of "speech macros," which are natural language prompts designed to transform abbreviated user input into full sentences using LLMs. The objectives were: (1) What are the benefits for AAC users, if any, of directly interacting with large language models (LLMs)? (2) How do AAC users evaluate communication suggestions made by an existing LLM? (3) What concerns do AAC users have about integrating LLMs into their own AAC devices?	(1) 12 participants, including plwMND. Age between 25-64, 7 male 3 female, 2 non-binary. (2) No details specifically provided on number of plwMND, type of MND age or gender. (3) A mixed-methods approach was used, combining qualitative and quantitative data collection and analysis. (4) Participants were introduced to the concept of LLMs and their potential applications in AAC (5) Three interactive Speech Macro web-based prototypes were developed, * Extend Reply - takes short input from the participant and extends with more detail, For example, if the conversation partner asks, "What did you do today?" and the participant types "work," the macro might suggest, "I worked from 8 AM to 6 PM." * Reply with Background Information: enables participants to input personal information about themselves beforehand. When asked a question during a conversation, the macro generates responses based on this pre-provided information. * Turn Words into Requests: transforms single words into full-fledged requests. (6) A 90-minute remote video call where each participant interacted with the prototypes and provided feedback. (7) Post-Study Survey: Participants completed a survey containing multiple-choice and open-ended questions (8) Qualitative data analysis was performed on the transcripts of the video calls and survey responses, identifying themes related to feature suggestions, use cases, quality of suggestions, and concerns about AI in AAC.
ASR as AAC	32	Online speech synthesis using a chronically implanted brain–computer interface in an individual with MND Angrick et al. (2024) USA	Primary Research peer reviewed	(1) The researchers explored the feasibility of using an implanted brain-computer interface (BCI) to synthesize intelligible speech for a plwMND. (2) For a plwMND using BCI to produce audible, intelligible words that closely resemble characteristics of the user's voice profile, with most synthesized words of sufficient intelligibility to be correctly recognized by human listeners.	(1) A male native English speaker in his 60s with chiefly Bulbar MND. (2) The BCI system utilized electrocorticographic (ECoG) signals recorded from parts of the brain to identify, decode, and synthesize speech using an ASR. (3) The study focused on six keywords that the participant could reliably produce using natural speech. ("Left", "Right", "Up", "Down", "Enter", "Back", and "…" for silence). These were chosen to control a communication board. (4) Training data was collected for ten days over six weeks - during which ECoG was recorded during multiple blocks of a speech production task where the key words were spoken ten times (60 words in total). (5) five and a half months later the synthesized output was played to the participant while he performed the same keyword reading task as in the training sessions. (5) As the participant could still produce speech, the researcher could reliably map his cortical activity during overt speech production processes and his voice's acoustic features. (6) The results were assessed through a listening test conducted with 21 native English speakers. The listeners were asked to listen to individual samples of the synthesized words and select the word that most closely resembled each sample.
How ASR is used	33	Voices from the Field: Benefits of Smart Home Technology for Individuals Living with Amyotrophic Lateral Sclerosis Wallock & Cerny, (2021) USA	Primary Research peer reviewed	(1) This paper examines the benefits of smart home technology as assistive technology for plwMND. (2) Data was collected through surveys from plwMND	(1) 19 participants with MND were surveyed - all used various smart home devices on a daily basis. (2) 12 males and 7 females, ages ranged from 28 to 86 (3) All individuals included in the sample had either spinal or limb-onset MND. (4) The duration of MND varied among the participants, with 47% having lived with MND for 2-3 years, 21% for 1-2 years, 16% for less than one year, and 16% for more than three years.

Table 7: Participants' communication needs and studies' outcomes.

Group	N	Citation	Communication needs: dysarthria, AAC Physical or cognitive impairments.	Outcomes for communication support / environmental control
ASR Accuracy	1	MacDonald et al., (2021)	(1) Recordings from plwMND with mild to severe dysarthria though not individually reported. (2) AAC needs not detailed. (3) Any physical or cognitive impairments were not specified.	(1) Compilation of a large dataset of disordered speech, claimed to be the largest known dataset of diverse disordered speech. (2) The dataset is used to create personalised ASR models that are overall claimed to reduce WER by about 75% but vary by individual. Over 80% of the personalized models achieved WERs below 15% based on internal assessment. (3) The research provided valuable insights into the considerations and challenges involved in collecting data for disordered speech, including privacy, consent, outreach to diversify the dataset, prompted vs spontaneous speech, lab vs home recordings, noise and artifact types, phrase quantity and repetitions, use-cases and domains, quality control, and metadata.
ASR Accuracy	2	Green et al., (2021)	(1) Recordings from MND participants with mild to severe dysarthria though not individually reported. (2) AAC needs not detailed. Any physical or cognitive impairments were not specified.	(1) Personalized ASR models outperformed speaker-independent models when measured using WER. The median word error rate (WER) for personalized models was 4.8%, while the median WER for speaker-independent models was 31%. They Also outperformed human speech recognition with a median gain of 9%. (2) personalized ASR models varied based on the severity of speech impairments and higher for the severe speech impairment group (median WER of 13%) compared to the normal, mild, and moderate groups (median WER of 3.8%). (3) Factors such as speech characteristics, signal-to-noise ratio, and training set size had an impact on personalized ASR performance.
ASR Accuracy	3	Tobin & Tomanek, (2022)	(1) No details provided for the MND participants in this study	(1) WER thresholds were selected to determine 'Success Percentage' (the percentage of personalized models reaching the target WER) in different application scenarios. (2) For the home-automation use-case, average Success Percentage was 79% when models are personalized with 250 training utterances per speaker. With 50 utterances, the average Success Percentage was 63%. (3) For the conversational use-case, models personalized with 250 utterances saw a Success Percentage of 73%.
ASR Accuracy	4	Tobin et al., (2022)	(1) No details of plwMND dysarthria, AAC needs, physical or cognitive impairments were provided in the paper.	(1) BERTScore was found to be more useful in measuring the semantic similarity between the ASR model's output and the intended meaning, especially in cases where WER might be misleading due to errors that don't significantly alter the meaning. (2) BERTScore was a better predictor of how SLTs would assess the severity of an error compared to WER.
ASR Accuracy	5	Tomanek et al. (2023)	(1) MND participants all progressed over time from mild to severe/profound dysarthria as assessed by two SLPs. (2) AAC needs not detailed. Any physical or cognitive impairments were not specified.	(1)WER increased with the severity of dysarthria. (2) The performance of the three ASR models (Unadapted Speaker Independent, Adapted Speaker Independent, and Adapted Speaker Dependent) degraded as speech became more impaired. (3) The Unadapted Speaker Independent models consistently performed the worst, Adapted Speaker Dependent models the best. (4) The findings emphasized the importance of continuous recording and model retraining when providing personalized models for individuals with progressive speech impairments.
ASR Accuracy	6	Tomanek et al. (2024)	(1) No details of dysarthria, AAC needs, physical or cognitive impairments were provided in the paper.	(1) A LATTEScore threshold of 80% was found to be optimal for distinguishing high-quality models (2) LATTEScore showed that tuned LLMs can effectively assess the comprehensibility of ASR transcripts, providing a better alternative to WER and BERTScore.
ASR Accuracy	7	Chen et al.(2021)	(1) The plwMND were reported to have dysarthria ranging from mild to severe. (2) There was no indication of current communication needs, or AAC use. (3) Any physical or cognitive impairments were not specified.	(1) Significant reduction in WER: between 24% and 32% relative on normal speech and up to 53% relative for dysarthric speech. (2) The time required for ASR training reduced by 22% without degradation in accuracy.
ASR Accuracy	8	Doshi et al., (2021)	(1) All plwMND were reported to have dysarthria ranging from mild to severe. (2) There was no indication of current communication needs, or AAC use. (3) Any physical or cognitive impairments were not specified.	(1) There was significant reduction in WER from 71.3% to 16.9% on average for all the participants with dysarthria.
ASR Accuracy	9	Calvo et al., (2021)	(1) The plwMND had moderate or severe dysarthria. (2) No indication of current communication needs, or AAC use. (3) Any physical or cognitive impairments were not specified.	(1) mPASS significantly more accurate than Pocketsphinx. The word recognition rate for mPASS was 88.7%, Pocketsphinx was 63.7%. (2) The TAM questionnaire results showed a mean score of 17.3 for perceived usefulness of mPASS (20 means strongly agree) and a mean score of 16.4 for perceived ease of use (20 means strongly agree).

Group	N	Citation	Communication needs: dysarthria, AAC Physical or cognitive impairments.	Outcomes for communication support / environmental control
ASR Accuracy	10	Marini et al., (2021)	(1) No details of plwMND dysarthria, AAC needs, physical or cognitive impairments were provided in the paper.	(1) The GMM-HMM ASR system achieved an average WER of 11.75% and the DNN-HMMASR 14.99% WER.
ASR Accuracy	11	Mulfari et al., (2022)	(1) For the six participants, 2 had mild, 2 moderate, 2 severe ratings. (2) No details on AAC, physical or cognitive impairments.	(1) Average 7% WER was measured when naming 13 individual words trained in the individual personalised ASR. (2) This compares favourably with other research that attained around 15% WER.
ASR Accuracy	12	Mulfari et al., (2023)	(1) For the 16 participants, 3 had mild, 8 moderate, 5 severe ratings. (2) No details on AAC, physical or cognitive impairments.	(1) Average 6.3% WER was measured when naming 13 individual words trained in the individual personalised ASR and 15.3% WER on a global model using recordings from other users.
ASR Accuracy	13	Geng et al., (2022)	(1) No details of dysarthria reported. (2) AAC needs not detailed. (3) Any physical or cognitive impairments were not specified.	(1) The technique consistently outperformed existing methods for creating ASR models measured by a statistically significant improvement in word error rate (WER) reduction.
ASR Accuracy	14	Hu et al., (2022)	(1) Recordings from the TORGO database were used from eight participants. (2) One participant had been diagnosed with MND and the remainder with cerebral palsy. (3) No details were provided about which participant had MND, gender or level of dysarthria.	(1) Using A2A-generated articulatory features consistently improved ASR performance, at best achieving a WER 24.82% on the UASpeech dataset, beating the lowest published WER on this benchmark task (25.21% WER).
ASR Accuracy	15	Lin et al. (2023)	(1) Recordings from the TORGO database were used from eight participants. (2) One participant had been diagnosed with MND and the remainder with cerebral palsy. (3) No details were provided about which participant had MND, gender or level of dysarthria.	(1) IPPM combined with KD was found to be effective in improving ASR recognition of dysarthric speech. (2) Using this approach on the TORGO dataset, the average phoneme error rate (PER) was 20.47%, comparing favourably with the baseline PER 33.08%. (PER measures the percentage of phoneme errors made by the ASR compared to ground truth transcription).
ASR Accuracy	16	Naeini et al., (2024)	(1) All 13 plwMND described as having mixed spastic/flaccid dysarthria. (2) No details of AAC needs, physical or cognitive impairments were provided in the paper.	(1) fine-tuning the ASR model with imitated and synthesised dysarthric speech improved its performance. (2) The best performing model configuration achieved - average WER 5.7%. vs baseline WER 17.1% (the pre-trained model without fine-tuning) - CCA of 94.4%, vs baseline CCA 80.9%. - IoU 89.62% vs baseline IoU 64.2%
ASR Accuracy	17	Okafor & Ludi (2022)	(1) No details of dysarthria, AAC needs, physical or cognitive impairments were provided in the paper.	(1) A majority of participants found the system easy to use after familiarizing themselves with the voice commands and the averages SUS score was 66/100. However the plwMND scored 50 - the second lowest of the cohort. (2) Most participants rated their overall experience above average, however the plwMND rated their overall experience with the system as 2 out of 5 - the lowest of the cohort.
Supporting Diagnosis	18	Bhattacharjee et al., (2021)	(1) No details of plwMND dysarthria, AAC needs, physical or cognitive impairments were provided in the paper.	(1) The study found that pitch and MFCCs were equally good at detecting the MND/PD when the recordings were clear. (2) However, when the recordings were noisy, pitch was more reliable than MFCCs.

Group	N	Citation	Communication needs: dysarthria, AAC Physical or cognitive impairments.	Outcomes for communication support / environmental control
Supporting Diagnosis	19	Tena et al., (2021)	(1) MND severity and its bulbar presentation also varied among participants, as assessed by the MND Functional Rating Scale-Revised (ALSFRS-R). (2) The MNDFRS-R score (0-48) was obtained from 12 survey questions that assess the degree of functional impairment, with the score of each question ranging from 4 (least impaired) to 0 (most impaired). (3) The scores of the 45 participants in this study ranged from 6 to 46 (mean 31.3,SD 8.6; 3 patients' scores were reported as not available).	(1) The automated model they developed showed promising results, accurately predicting bulbar involvement with an accuracy of 95.8% (2) The researchers suggest that the automated model could help with early detection and better management of MND, potentially leading to earlier intervention and treatment.
Supporting Assessment	20	Venugopalan et al., (2021)	(1) SLTs assessed the overall intelligibility of each speaker on a five-point Likert scale. (2) The scale was mapped to 5 classes- typical, mild, moderate, severe, and profound. (3) No indication of current communication needs, AAC or physical of cognitive impairments.	(1) Classifiers trained on representations extracted from an ASR performed the best in detecting and classifying disordered speech. (2) The ASR clustered speech by the spoken phrase and longer phrases more indicative of intelligibility deficits than single words. (3) This research demonstrated the potential of ASR and deep learning techniques to objectively assess and classify disordered speech. (4) This approach may support faster and more consistent diagnosis of dysarthria, and improve ASR system performance.
Supporting Assessment	21	Venugopalan et al., (2023)	(1) Recordings from plwMND with mild to severe dysarthria though not individually reported. (2) AAC needs not detailed. (3) Any physical or cognitive impairments were not specified.	(1) The models achieved an accuracy of over 90% when evaluated on a test set of nearly 94K utterances from 100 speakers
Supporting Assessment	22	Soleymanpour et al., (2021)	(1) No details of plwMND dysarthria, AAC needs, physical or cognitive impairments were provided in the paper.	(1) ARS for dysarthria severity generally aligned with the labelled severity levels in the TORGO dataset. (2) However the ARS for normal speakers was slightly higher than expected, with many scoring close to 2 (very low severity) rather than 1 (normal), suggesting had difficulty differentiating between normal speech and low dysarthria severity. (3) the speech intelligibility probability distribution for individual phrases showed clear differences in mean intelligibility between intelligible and unintelligible speakers.
Supporting Assessment	23	Rong, (2021)	(1) Six manifested bulbar symptoms and seven plwMND were absent of clinical bulbar symptoms but self-reported speech changes. (2) There was no indication of current communication needs, or AAC use. (3) Any physical or cognitive impairments were not specified.	(1) The automated model they developed showed promising results. (2) This suggests that the model could be a valuable tool for early detection and better management of MND.
Supporting Assessment	24	Mendoza et al., (2022)	(1) Both plwMND were reported to have severe mixed dysarthria. (2) There was no indication of current communication needs, or AAC use. (3) Any physical or cognitive impairments were not specified.	(1) The research aimed to develop an objective measure of the degree of phoneme distortion. (2) Different regression models and classification models were evaluated for predicting speech sound distortions. (3) The results showed that the models had varying levels of accuracy.
Supporting Assessment	25	Joshy & Rajan, (2022)	(1) Recordings from the TORGO database were used from eight participants. (2) One participant had been diagnosed with MND and the remainder were diagnosed with cerebral palsy. (3) No details were provided about which participant had MND, gender or level of dysarthria	(1) The DNN classifier using MFCC-based i-vectors outperforms other systems, achieving an accuracy of 93.97% dysarthria severity score (compared to the TORGO classification) using the speaker-dependent scenario.
Supporting Assessment	26	Bhattacharjee et al.(2023)	(1) The plwMND ranged between normal and severe dysarthria as measured by MNDFRF-R. (2) No details of AAC needs, physical or cognitive impairments were provided in the paper.	(1) Transfer learning improved dysarthria severity classification accuracy. (2) The best configuration achieved average dysarthria classification accuracy of 80.11% (11.03% higher than without transfer learning). (3) This improvement was primarily due to better classification of the 'mild' dysarthria category.
Supporting Assessment	27	Kumar et al. (2023)	(1) Dysarthria severity was assessed using the MNDFRS-R scale. (2) The scale ranges from 0 to 4 - 0 indicates complete loss of useful speech and 4 signifies normal speech. (3) The mode of the three ratings was taken as the final severity rating. (4) plwMND were divided into three severity groups based on their MNDFRS-R scores: (5) 39 plwMND were in severe dysarthric group (G1) with scores 0 and 1,	(1) The research found that, except for the highest severity case for vowels, manual classification performed better than automatic classification at all dysarthria severity levels. (2) However the difference between the two gradually reduced with an increase in severity.

Group	N	Citation	Communication needs: dysarthria, AAC Physical or cognitive impairments.	Outcomes for communication support / environmental control
Supporting Assessment	28	Gutz et al., (2022)	(1) 25 plwMND had spinal onset, 21 had bulbar onset, three had mixed onset, and three had unknown onset. (2) Mean MNDFRS-R score was 31. (3) No indication of current communication needs or AAC use. (4) Any physical or cognitive impairments were not specified.	(1) Clinician WRR achieved better accuracy than Google Cloud ASR WRR when categorizing dysarthria severity
Supporting Assessment	29	Wiepert et al. (2023)	(1) No details provided for the MND participants in this study.	(1) ASR Model errors were consistently larger than speaker abnormality errors. (2) ASR models 'wav2vec2 base and large' had a significantly lower proportion of model-only errors and a higher proportion of speaker-only errors compared to XLS-R and Whisper models. (3) WER differed across ASR models, with Whisper having a significantly higher WER than all other models. Differences between models likely dependent on what was not transcribed rather than what was transcribed incorrectly. (4) WER increased with utterance complexity, but the degree varied across ASR models. (5) ASR Model choice influenced the basis of any clinical analysis, for example if the ASR model transcribed non-words or ignored them.
ASR as AAC	30	Cai et al.,(2021)	(1) All had had moderate to severe dysarthria but were still able to produce the target vowel sound (/a/). AAC needs not detailed. (2) Physical or cognitive impairments were not specified.	(1) The results showed that the two-step ASR training approach led to substantial improvement in signal-detection quality. (2) The false positives per hour (FPPH) for the two-phase approach, was 5.5 times lower (better) compared to other methods of training.
ASR as AAC	31	Valencia et al. (2023)	Not detailed which participants had MND but of the twelve participants two used eye-gaze technology to access AAC, four switch users and six AAC users who used direct selection. All participants were reported to have no cognitive impairments. No other details provided	(1) Most participants found speech macros to be either 'very' or 'extremely' useful (2) Many were positive about the potential to reduce the physical and cognitive effort required to communicate (3) Participants highlighted the importance of AI-generated phrases aligning with their personal communication style, tone, and identity. (4) Participants envisioned using AI for various purposes beyond the tested speech macros, including interacting with voice assistants, managing routine tasks, and preparing for medical appointments. (5) Some expressed concern about losing control over their communication and sounding inauthentic. (6) Some were worried about the privacy and security of their personal data used by AI systems.
ASR as AAC	32	Angrick et al. (2024)	(1) The plwMND 'was still able to produce some words that were intelligible when spoken in isolation, but his sentences were often unintelligible'. The plwMND could articulate the six keywords in this research with a 'high degree' of intelligibility. (2) MNDFRS-R in a score of 26 out of 48 possible points (speech was rated at 1 point. Continuous speech mostly unintelligible (though individual words were	(1) The BCI system successfully synthesized words that closely resembled the participant's own voice and pacing. (2) The intelligibility of the synthesized speech was evaluated through listening tests with human listeners, who were able to correctly recognize 80% of the words. (3) The words "Left" and "Back" were responsible for the most of misclassified words. The researchers argued that the plwMND dysarthria likely made these two words less discriminable in the synthesis process. (4) The study's findings suggest that speech-impaired individuals with MND can utilize chronically implanted BCIs to produce synthesized words
How ASR is used	33	Wallock & Cerny, (2021)	(1) No details of dysarthria reported. (2) AAC needs not detailed. (3) Any physical or cognitive impairments were not specified.	(1) All plwMND used smart home technology devices daily, reporting greater control over electronics in their home environment. (2) Most participants (89%) reported that smart home technology allowed them to regain some functional independence in their home. (3) Over half plwMND (63%) felt smart home technology had enabled them to complete tasks that they previously lost the ability to do.

Table 8: Studies' advantages and disadvantages and future studies.

Group	N	Citation	Advantages	Disadvantages	Future Studies
ASR Accuracy	1	MacDonald et al., (2021)	(1) The research has compiled a large dataset of disordered speech and a resource for further research and development in the field of ASR. (2) The research provides insights into the considerations and challenges involved in collecting data for disordered speech.	(1) The data collection tool only recorded scripted speech. Although this allowed for high transcript conformity and reduced transcription cost it provided little data for conversational speech, or examples of the words people actually say. (2) English only (3) The entire dataset is not publicly available due to participant consent terms. This could potentially hinder further research and collaboration in the field.	Further research to develop speaker-independent ASR models capable of accurately recognizing disordered speech of unseen speakers.
ASR Accuracy	2	Green et al., (2021)	(1) Personalised ASR models outperformed speaker-independent models and were comparable to or even better than expert human transcribers. (2) Personalised ASR has the potential to assist or even replace human listeners in certain applications. (3) Personalised ASR models can be effective in recognizing a wide range of speech impairments and severities.	(1) WER was higher for participants with severe speech impairments compared to those with milder impairments. (2) Limited applicability to functional communication: the study focused on short phrases and home automation tasks.	Further research is needed to evaluate the performance of personalized ASR models with longer utterances and everyday conversation.
ASR Accuracy	3	Tobin & Tomanek, (2022)	The researchers argue that small amounts of training data can be enough for useful personalized ASR models, especially for home automation with limited vocabulary and less complex phrases.	(1) The researchers assumed both WER thresholds would achieve functional benefit to people using the ASR. However it remains unclear if this was indeed the case as there were no opportunities for end-users to validate these assumptions (2) There was no testing with actual human-in-the-loop conversations and success was measured using WER. (3) For the 'conversational' dataset in Euphonia users are asked to repeat in isolation short to medium length phrase, for example ""Hey, why don't you come over for dinner tomorrow". This may not reflect conversational speech in daily life. For home-automation use-case these are a sample of commands to a digital assistant that have not been selected based on frequency of use e.g. ("Turn on the lights".	To research how the amount of data needed for speech model personalization can be further reduced and still be useful for task-specific utterance selection.
ASR Accuracy	4	Tobin et al., (2022)	(1) BERTScore distinguishes error severity better than WER. (2) It also distinguishes error types better and fits SLT assessments better. (3) BERTScore is more useful when measuring semantic similarity vs WER	(1) The BERTScore is more computationally expensive in terms of processing overhead than WER. (2) WER can be used many applications regardless of language or word selection. (3) However, because BERT is assessing semantic meaning, it can run into issues with multilingual data and words that are not in the large language model but nevertheless valid (for example some proper nouns). (4) The research was based on set phrases. The results may not be indicative of performance for more realistic speech in conversational settings using a variety of words and phrases. (5) The accuracy rates reported may not replicate in everyday environments with background noise.	To investigate further metrics, such as Character Error Rate (CER) and BLEURT, that may give insight into when a model will be useful for speakers with disordered speech. To assess if utilizing both WER and BERTScore together is more beneficial to assess ASR usefulness than using each measure individually.
ASR Accuracy	5	Tomanek et al. (2023)	Improved Recognition Performance: Both the start-over and continued training strategies significantly improved WER compared to the baseline. Updating the ASR models with more recent recordings from the severe stages of speech decline can enhance recognition accuracy.	(1) Only recordings of short phrases used even though longer conversational recordings available, because of 'confounding factors' associated ASR for conversational speech. Therefore may be little implication for plwALS with changing speech using ASR for conversation. (2) Limited Training Data: In practice, it may be challenging to collect a substantial amount of data from individuals with severe dysarthria. In the severe stages of speech decline, the number of available recordings may be limited, impacting the effectiveness of the mitigation strategies (3) Individual Variability: The effectiveness of the mitigation strategies may vary across individuals due to the unique characteristics of their dysarthria. What works well for one individual may not necessarily work as effectively for another.	(1) Further investigate the effectiveness of the mitigation strategies and explore their generalizability to a larger population. (2) Explore additional mitigation strategies and compare their effectiveness in improving ASR performance. (3) Evaluate the performance of other ASR models, such as hybrid models or transformer-based models, in the context of degenerating speech. (4) Utilize a more controlled, longitudinal dataset with additional participants.
ASR Accuracy	6	Tomanek et al. (2024)	(1) LATTEScore is proposed as an practical measure for assessing meaning retention in ASR transcriptions of dysarthric speech. (2) This approach is argued to enable automation of quality assessment of ASR model transcriptions, and an alternative to the slower and labour intensive method of manual assessment by SLTs.	(1) No comment was made about the content of the phrases used, whether transactional, short or long in length, and particularly how words that may not be present in the LLM are assessed by LATTEScore - for example, proper nouns of people places and things were uttered. (2) The research evaluated meaning preservation only within individual sentences. However, assessing preservation of meaning across multiple sentences or entire conversations may better reflect the reality of conversational interaction.	(1) Research assessment of multi-lingual LLMs - beyond English only. (2) Determine a measure to assess whether meaning preservation at conversational level.
ASR Accuracy	7	Chen et al.(2021)	(1) Improved Accuracy - significant reductions in WER for both normal speech and dysarthric speech. (2) Faster ASR training, supporting improved scalability to a large set of users.	(1) Lack of Real-world Testing: The research presents results based on recordings on specific datasets.	(1) No future studies discussed in the paper.

Group	N	Citation	Advantages	Disadvantages	Future Studies
ASR Accuracy	8	Doshi et al., (2021)	The research shows that with relatively few recordings of speech with dysarthria, the Parrotron ASR model can achieve a 77% relative reduction in WER, potentially making it more useable.	(1) This outcomes of this research were not assessed by people with dysarthria and their conversation partners, therefore the researcher's claim that Parrotron "makes it easier for speakers with speech impairments to be understood by other people' is not proven. (2) The research was based on set phrases. The results may not be indicative of performance for more realistic speech in conversational settings using a variety of words and phrases. (3) The accuracy rates reported may not replicate in everyday environments with background noise.	(1) To explore prosody transfer and target voice personalization. (2) Prosody transfer refers to emulating the prosodic characteristics of the source dysarthric speaker's speech in the Parrotron voice conversion (VC) (3) Target voice personalization refers to emulating the dysarthric speaker's voice in the Parrotron VC (without dysarthria).
ASR Accuracy	9	Calvo et al., (2021)	(1) mPASS significantly outperformed a commercially available speaker-independent ASR system in recognizing both words and sentences spoken by individuals with dysarthria. (2) mPASS platform is designed to be user-friendly and this was supported by the TAM survey responses	(1) The research was based on single words and short 2-3 word phrases chosen by the researcher. The results may not be indicative of performance for more realistic speech in conversational settings using a variety of words and phrases. (2) mPASS was tested in a clinical environment; the accuracy rates reported may not replicate in other, potentially noisier environments.	To explore the use of mPASS software with the creation of individualized applications tailored to the users' needs.
ASR Accuracy	10	Marini et al., (2021)	The experiment shows that ASR systems based on GMM model perform better than DNN model for this amount of data.	(1) The research was based on single words. No everyday conversational speech was assessed. (2) recordings of only single words used. The results may not be indicative of performance for more realistic speech in conversational settings using a variety of words and phrases. (3) The ASR models were tested in a clinical environment; the accuracy rates reported may not replicate in other, potentially noisier environments.	To use more data to train ASR systems and to investigate how to include additional clinical information in the training process.
ASR Accuracy	11	Mulfari et al., (2022)	(1) the potential for personalized articulation therapy that can be remotely supervised by SLTs and independently practiced (2) continuous monitoring, immediate feedback, (3) opportunities data collection of speech with dysarthria (4) to enhance the effectiveness, reach, and affordability of speech therapy for individuals with dysarthria. (5) accurate single word ASR recognition can help interaction with virtual home assistants (VHAs), such as Amazon Alexa, supporting environmental control using voice (6) mapping single word speech to a customized speech message spoken aloud via a computer generated voice, could help with transactional communication	(1) Sparse availability of speech data across a broad variety of aetiologies increased WER when using grouped recordings vs relying on individual personal recordings per word. (2) Only a few individual words recognised by the ASR, and these are trained individually by each user and not generalizable to words not trained.	(1) To centrally store speech samples from contributors and many other types of data in open source big data no SQL solutions. (2) to add new functionalities to the proposed software prototype to support speech therapy for people with dysarthria (3) To support multiple languages.
ASR Accuracy	12	Mulfari et al., (2023)	(1) accurate single word ASR recognition can help interaction with virtual home assistants (VHAs), such as Amazon Alexa, supporting environmental control using voice (2) mapping single word speech to a customized speech message spoken aloud via a computer generated voice, could help with transactional communication	(1) Sparse availability of speech data across a broad variety of aetiologies increased WER when using grouped recordings vs relying on individual personal recordings per word. (2) Only a few individual words recognised by the ASR, and these are trained individually by each user and not generalizable to words not trained. (3) The research was based on a selection of single words chosen by the researcher. The results may not be indicative of performance for more realistic speech in conversational settings or when interacting with digital assistants . The words were alto: high , apri: open avanti: forward, basso: low, chiudi: close, destra: right entrata: entrance, indietro: back, okay: okay, sinistra: left tappa: stop, uscita: exit, volume: volume (4) The accuracy rates reported may not replicate in everyday environments with background noise.	(1) Collaborating with non-profits / patient associations to build the volume and diversity of available recordings. (2) Evaluating the concept with people living with dysarthria. The researchers acknowledge that they have not done this. (3) Develop an application usin the concept that can interface with smart home automation.
ASR Accuracy	13	Geng et al., (2022)	(1) Improved WER compared to existing methods for recordings of dysarthric speech (2) Argued to be generalizable to different languages and speech disorders.	(1) Not tested on conversational speech - only very few short phrases (2) Not tested by plwALS (3) Lack of user evaluation or subjective assessments.	(1) focus on conversational speech
ASR Accuracy	14	Hu et al., (2022)	(1) The A2A inversion approach may make it easier to incorporate articulatory features into ASR systems, even when limited articulatory data is available. This is argued to lead to accurate ASR recognition of dysarthric speech .	(1) Very few recordings of plwALS were used (from the TORGO database) and containing no conversational speech. (2) Given that declining respiratory support is a common feature of ALS and can contribute to the effect of dysarthria by reduced power of voice, by not including this consideration the research may potentially overstate the benefits of the A2A model on ASR accuracy in everyday conversation. (3) The solution was not tested by recordings of plwALS with differing severity of dysarthria. (4) In summary, it is difficult to forecast the level of functional benefit plwALS may have with this solution.	No specific future research is detailed in this paper.

Group	N	Citation	Advantages	Disadvantages	Future Studies
ASR Accuracy	15	Lin et al. (2023)	(1) The researchers argued their approach can improve the accessibility and usability of ASR technology for people with dysarthria.	(1) Very few recordings of plwALS were used (from the TORGO database) and containing no conversational speech. (2) Additionally, similar to (Hu et al., 2022) the model did not take into account that declining respiratory support is a frequent feature of ALS and can contribute to the effect of dysarthria by reduced power of voice. (3) Their solution was not tested by plwALS with differing severity of dysarthria. (4) In summary, it is difficult to forecast the level of functional benefit plwALS may have with this solution.	(1) The authors plan to focus on reducing ASR recognition errors resulting from 'improper' breathing by people with dysarthria (2) Incorporate of a language model to govern the selection of output words.
ASR Accuracy	16	Naeini et al., (2024)	(1) imitated and synthetic speech data can be used to improve the performance of ASR models for dysarthric speech, and may be a viable alternative for improving ASR performance in the context of sparse availability of dysarthric speech recordings.	(1) The imitated/synthesised dysarthric speech was a partial emulation of simple dysarthric symptoms and did not replicate the full complexity and variation of dysarthric speech. (2) 55 different accents were used in the synthesised speech generation - there was no discussion on the similarities/differences of accent and dysarthria or the appropriateness of utilising accent to measure ASR effectiveness for dysarthria. (3) recordings of a single short phrases used. The results may not be indicative of performance for more realistic speech in conversational settings using a variety of words and phrases (4) recordings with background noise were excluded - therefore the results may not be indicative of performance in noisier environments common in daily life.	To compare the effect of fine-tuning using different amounts of healthy control recordings with those with dysarthria to establish the optimal sample size for an ASR.
ASR Accuracy	17	Okafor & Ludi (2022)	(1) For some participants voice-enabled block-based programming increase accessibility	(1) Even though the plwALS gave almost the lowest scores in both assessments, suggesting poor useability, the researchers do not provide specific information about she liked or disliked. (2) Almost 70% of participants reported that speech recognition did not work properly (3) This research shows that even when the additional voice control modality was added to be more inclusive, not everyone benefits as expected, as in this case particularly for the plwALS.	(1) Alternative voice commands will be added to address the difficulty in detecting speech. (2) more intuitive and less ambiguous voice commands . (3) a more robust ASR. (4) reduce the time lag in system response, (5) After the changes have been implemented another usability study will be conducted.
Supporting Diagnosis	18	Bhattacharjee et al., (2021)	(1) As pitch was more reliable than MFCCs when the recordings were noisy, it may be particularly beneficial as voice recordings made in daily life often have background noise. (2) Additionally, using pitch is easier to measure vs MFCCs - leading to a potentially simpler ways to detect ALD/PD using voice recordings, even in noisy environments. (3) Recordings of conversational speech were used, and across diverse languages. As a result the authors contend the results may generalize across other languages too.	(1) A larger and more diverse dataset, including plwALS with varying degrees of disease severity across linguistic backgrounds, would strengthen the author's claim of generalizability. (2) The study only assessed additive white noise. Other types of noise (including background talking, environmental sounds) was not researched.	(1) Examine the noise robustness of different speech features in various additive noise conditions as well as real noisy recordings. (2) Trial using denoising algorithms
Supporting Diagn	19	Tena et al., (2021)	(1) automated signal analysis and machine learning for diagnosing bulbar involvement, which may improve the accuracy of human diagnosis of ALS.	(1) The study is limited to Spanish-speaking participants, the findings may not be directly applicable to other languages. (2) The sample size of the study is small which may limit the generalizability of the results. (3) it is unclear how the performance of the automated model may change over time as ALS progresses.	(1) improved identification of incorrectly undiagnosed bulbar-involvement in patients with ALS. (2) increasing the sample size of voice recordings for analysis
Supporting Assessment	20	Venugopalan et al., (2021)	(1) classifiers trained on representations extracted from an ASR system performed the best in detecting and classifying disordered speech. (2) The ASR system's representations clustered speech by the spoken phrase, and longer phrases were found to be more indicative of intelligibility deficits than single words. (3) This research demonstrates the potential of using ASR systems and deep learning techniques to objectively assess and classify disordered speech, which can have implications for diagnosing dysarthria.	(1) Although the results indicate that the models can detect mild disordered speech, this might just be an indicator that the range of mild intelligibility scored by the SLTs is large.	(1) Comparing the performance of these ASR models to Speech Therapists (2) Assess if this classification process can be carried out on general (conversational) speech.
Supporting Asses:	21	Venugopalan et al., (2023)	The Classifiers demonstrated generalization to different datasets, and showed good performance in assessing speech intelligibility for recordings of plwMND.	(1) The models had difficulty classifying speakers who had dysarthria but typical speech intelligibility. (2) The models were assessed using recordings of scripted phrases and not conversational speech or phrases that people may use in daily life. (3) The models were not assessed either for speech in noisy backgrounds	(1) Diversifying the dataset: The current dataset used in the research was reported to lack diversity in terms of race and aetiologies. (2) Including speech samples from different demographics, minority groups, and speakers of other languages and dialects. (3) To include non-speech samples and noisy backgrounds: (4) To include unprompted speech

Group	N	Citation	Advantages	Disadvantages	Future Studies
Supporting Assessment	22	Soleymanpour et al., (2021)	(1) The researchers argue that automatic and sentence assessment of dysarthria severity and speech intelligibility could offer a more objective and informative tool for clinical assessment and intervention planning for people with dysarthria. (2) The variation in speech intelligibility at phrase level could support detailed analysis of specific sound contrasts in dysarthric speakers, helping planning for clinical intervention.	(1) small number of participants, only one with ALS - limiting the generalizability of the findings. (2) recordings of a single words and phrases used. The results may not be indicative of performance for more realistic speech in conversational settings using a variety of words and phrases (3) results may not be indicative of performance in noisier environments common in daily life. (4) The study did not compare their approach with other automated methods measuring dysarthria severity and speech intelligibility. However as this is the first work to continuously assess dysarthria severity at phrase level and intelligibility there is no direct way to compare these results with the findings of other works.	No future research detailed.
Supporting Asses	23	Rong, (2021)	(1) The research introduces a new method to assess speech problems in plwALS, combining multiple measurements, such as muscle activity, jaw movement, and speech sound, (2) The method accurately distinguishes between plwALS and healthy controls (3) Early detection of speech impairment in ALS can lead to timely interventions.	(1) small number of participants, limiting the generalizability of the findings. (2) Further research is needed to validate the findings and assess the applicability of the approach (3) The study did not compare the new approach with existing methods for measuring speech variability in ALS. (4) Long-term follow-up was not conducted to track speech decline and evaluate treatment outcomes.	(1) To apply this framework to track speech decline over time for plwALS. (2) The multimodal approach may be applied to individuals with different neurologic diseases.
Supporting Assessment	24	Mendoza et al., (2022)	(1) The researchers argued there was a higher likelihood of objective estimation of dysarthria (2) Potential for automatic feedback in therapy, allowing real-time feedback on accuracy in producing specific phonemes during speech production. (3) Potential reduction of scoring time and subjectivity in articulation therapy. Challenges:	(1) There were challenges related to training requirements, reliability of reference ratings, phoneme-dependent prediction inaccuracies, and sensitivity to small changes	(1) improving the prediction models (2) increasing the dataset size and diversity (3)investigating the clinical utility and impact.
Supporting Asses	25	Joshy & Rajan, (2022)	This research can potentially help people with dysarthria in several ways: (1) Objective Assessment of dysarthria severity levels. This can assist SLTs evaluate progress of interventions and therapy strategies. (2) By consistently analysing specific speech characteristics and patterns, therapy can be customized to address the unique needs and challenges of each person, and monitors over time. (3) The findings from this research can contribute to the development of ASR assistive technologies, and voice synthesis tools (4) With automated and objective assessment, It may be possible to identify dysarthria at an earlier stage, allowing for earlier intervention.	(1) Only one plwALS recorded phrases in the TORGO database. These phrases included "How are you?" "This is an emergency." "Please open the window quickly." (2) These phrases were not recorded multiple times to reflect the progression of ALS or any speech decline (3) The phrases may not be representative of what plwALS actually say in daily life. (4) There are no recordings of plwALS in conversation, or recordings of specific commands to digital assistants (5) Therefore it is difficult to assess if the research can indeed assess dysarthria as it changes severity for an individual, or assess if the results are applicable to the words and phrases plwALS may use in daily life. All of which is likely to impact the robustness and reliability of the classification models. (6) There was also a lack of comparison with clinical assessments on the same data.	(1) To assess additional technology to identify ways to increase accuracy.
Supporting Assessment	26	Bhattacharjee et al.(2023)	(1) Transfer learning techniques improved ASR model classification accuracy of dysarthria severity . (2) Transfer learning supported creating more robust models despite the scarcity of training data. (3) An automatic dysarthria severity classification system may improve consistency and objectivity in assessments of dysarthria (4) Considering the plwALS participants spoke five different languages, the study may indicate this approach could be applicable across diverse linguistic backgrounds.	(1) The distribution of dysarthria severity classes is uneven, with more plwALS in the normal class than in the mild and severe classes. This imbalance may affect the model's performance, particularly in underrepresented classes. (2) No trials of the technology in clinical environments where assessments are normally conducted.	(1) Explore different machine learning configurations (2) To perfom 5-class dysarthria severity classification, reflecting the five groupings of dysarthria in the ALSFRS-R speech subscore.
Supporting Asses	27	Kumar et al. (2023)	(1) An improved understanding of dysarthria in ALS (2) Objective assessment of dysarthria severity, potentially leading to the development of diagnostic tools (3) Inform speech therapy approaches (4) Advance machine learning techniques for speech analysis. (5) Overall, potentially leading to improved quality of life for plwALS.	(1) Manual classification was better than automatic for all dysarthria severity levels, except the highest severity case for vowels, the difference between the two gradually reduces with an increase in severity.	(1) No future studies discussed in the paper.

Group	N	Citation	Advantages	Disadvantages	Future Studies
Supporting Assessment	28	Gutz et al., (2022)	This research provided clarity on the limits of ASR Google Cloud ASR for early detection of dysarthtia and grading dysarthria severity.	(1) Recorded samples were from the Speech Intelligibility Test (SIT), set phrases varying from single words up to eight word phrases increasing in complexity. Examples of sentences used in the SIT include: "They left for the beach early." "He lost his keys in the park." "Shoe" "Tree". This is not the same as conversational speech or commands that plwALS may use to interact with digital assistants. Therefore the results may not be indicative of plwALS may actually use ASR systems. (2) Google Cloud ASR was inadequate for detection of speech impairment and grading speech severity. (3) sentence length significantly impacted both Machine and Clinician WRR accuracy. (4) Researchers reported that Google Cloud ASR should not be used for transcription of dysarthric speech, scoring of speech intelligibility or replacement of clinician dysarthria severity rating. (5) Because ASR systems use different underlying technology these findings cannot be generalized to other ASR systems.	Potentially explore how ASR systems perform on other clinical populations. For example, Parkinson's disease.
Supporting Asses	29	Wiepert et al. (2023)	Although ASR can capture useful information about speaker abnormalities, errors generated by the ASR model itself need to be considered. For example whether a model allows for non-words, as well as the general objective of the model (e.g., transcribing intended meaning vs. sound-for-sound).	(1) The research is based on a limited sample size and focuses on specific ASR models only. (2) The research was based on recordings "clipped from speech assessments". The results may not be indicative of performance for more realistic speech in conversational settings using a variety of words and phrases. (3) The accuracy rates reported may not replicate in everyday environments with background noise, especially as recordings with background noise were excluded.	(1) To explore the interaction between perceptual domain of impairment (articulation, prosody, etc.) and the interactions with utterance complexity and ASR model type. (2) Examine the effect of noise, which ,may degrade ASR performance (3) Examine transcriptions at the character/phoneme level and explore improving ASRs through fine-tuning of the model.
ASR as AAC	30	Cai et al.,(2021)	(1) A voice-activated switch that can be triggered by a simple vowel sound (/a/), can increase accessibility (2) On-device inference allows for real-time detection and activation of preconfigured actions.	(1) there were limited examples of the target vowel sound (/a/) from impaired speakers, which may affect the performance of the neural network. (2) Relatively high false-positive rate, requiring confirmation through repeated activation of the model. (3) The switch can only trigger to a single vowel sound, limiting the range of utterances and actions that can be recognized.	(1) training detectors for other simple sounds (2) developing a multi-class vowel detector (3) training a model tolerant to alterations in voice quality caused by assisted breathing devices (4) exploring the adaptation of keyword spotting systems for less-impaired speakers.
ASR as AAC	31	Valencia et al. (2023)	(1) Speech Macros may reduce the physical and cognitive effort and increase speed of AAC-based communication potentially leading to faster interactions. (2) Participants envisioned using AI for various purposes beyond the tested speech macros, including interacting with voice assistants, managing routine tasks, and preparing for medical appointments.	(1) The study involved a small sample size and limited interaction time with speech macros, with no assessment in everyday communication. (2) The prototype used in the study was intentionally simplified, and further analysis on user Interface design is needed to better understand how the app may work in the real world. (3) The participants were concerned the AI-generated phrases may sound inauthentic or misrepresent their intentions. (4) Concerns were raised about data privacy and security of their personal data used by AI systems (5) The AI model struggled to generate appropriate responses for sensitive topics or specific needs, such as medical requests.	(1) Enable participants to interact with the AI model more extensively and in functional communication. (2) Refine the speech macro to improve the quality and relevance of generated suggestions. (3) Investigate enabling users to edit and customize their own macros for greater personalization. (4) Enable users to control, override, or correct the AI's output whenever they wish. (5) Develop more user-friendly interfaces through co-design with people that use AAC.
ASR as AAC	32	Angrick et al. (2024)	(1) ECoG BCIs utilizing online speech synthesis can serve as AAC devices for people living with ALS. (2) BCIs could be used to preserve or restore communication using a synthesised voice similar sounding to the natural one. (2) The speech synthesis used training data from five and a half months prior to testing and produced similar results over 3 separate days of testing, with recalibration but no retraining in each session. These findings suggest that the correspondence between neural activity in ventral sensorimotor cortex and speech acoustics were not significantly changed over this time period. (3) The researchers argue that direct synthesis of acoustic speech from ECoG is feasible, accurate, and stable, even in a person with dysarthria due to ALS. Given the increasing use of voice banking among people living with ALS, it may also be possible to improve the intelligibility of synthetic speech using this approach even in participants with unintelligible or absent speech.	(1) Only 6 words were assessed. The approach used here did not allow for synthesis of unseen words as it cannot capture fine-grained and isolated part-of-speech units, such as syllables or phonemes. (2) Additionally, where there was significant dysarthria in the natural voice recordings - these translated into dysarthric synthesized representations of BCI signals, leading to listener misunderstanding. (3) Longer time periods will be needed to confirm the stability of ECoG as a BCI signal source for speech synthesis. It remains to be seen how long this approach can produce intelligible speech as the neural responses change over time due to ALS.	(1) To expand the limited vocabulary used here (2) To explore to what extent similar or different approaches are able to extrapolate to words that are not in the vocabulary of the training set.

Group	N	Citation	Advantages	Disadvantages	Future Studies
How ASR is used	33	Wallock & Cerny, (2021)	(1) Greater Control: Participants reported that the use of smart home technology gave them greater control over their home environment. (2) Functional Independence: The majority of participants reported that smart home technology allowed them to regain some functional independence (3) smart home technology gave them the ability to complete tasks that they had previously lost the ability to do. (4) Improved Well-being: Users reported improved sense of safety, work simplification, energy conservation, and increased efficiency in daily living.	(1) Small sample size which may limit the generalizability of the findings. (2) The outcomes were based on self-reported data, which may be subject to biases and inaccuracies. (3) The study did not include a control group for comparison, making it difficult to determine the specific impact of smart home technology on the participants' outcomes.	(1) More research is needed to determine the impact of smart home technology on plwALS emotional and physical well-being.