LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Methods of Signal Processing Applied to Speech and Language for Objective Assessment of Cognitive Function in Ageing and Psychiatry

A dissertation submitted to the University of Dublin for the degree of

Doctor of Philosophy

by

Viliam Rapčan, M.Sc.

Supervisor: Professor Richard B. Reilly

Trinity College Dublin, April 2013

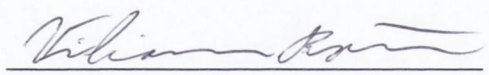
Neural Engineering Group, Trinity Centre for Bioengineering
Department of Electronic and Electrical Engineering
Trinity College Dublin

TRINITY COLLEGE
0 9 JUN 2014
LIBRARY DUBLIN

Thesis 10409

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree that the Library may lend or copy this thesis upon request.



Viliam Rapčan

6th March 2014

Abstract

Ageing is associated with a change in cognitive function. Impaired cognitive function is the biggest limiting factor to independence in older adults. Cognitive decline does not appear suddenly, but occurs gradually over varying periods of time.

Cognitive function is currently assessed by employing gold-standard neuropsychological tests and scales, such as the Positive and Negative Symptoms Scales, the Brief Psychiatric Rating Scale, the Mini-Mental State Examination, the Word Recall test, the Category Fluency test, etc. Although these gold standard scales and tools are well accepted, they represent very subjective manners for estimation of severity of cognitive impairment. The final score of these tests depends on the individual judgement and the quality of training of the person evaluating the subject. Currently, there are no objective biomarkers for diagnosis and monitoring of cognitive function in ageing and psychiatry.

This thesis focuses on the use of speech as an objective biomarker for assessment and monitoring of cognitive function. The production of speech is the end-product of a complex network of cognitive processes. Speech not only serves as a means of communication, but can also provide some other information about the speaker, such as their emotional state.

In this thesis, a novel method of using speech as a complementary and objective means of assessment and monitoring of cognitive function in ageing and psychiatry is presented. Novel algorithms for speech analysis were developed, which automatically and reliably (overall accuracy of 97.29%, sensitivity of 93.52%, specificity of 98.37%) extract temporal and acoustic features from recordings of speech acquired in real-life environments.

In a study of schizophrenia, schizophrenic patients and healthy controls were recorded reading out loud an emotionally neutral and semantically simple text extract from a children's story. Employing a Linear Discriminant Analysis (LDA) classifier to discriminate between schizophrenic patients and healthy controls using speech features, classification achieved an accuracy of 79.42% (sensitivity of 75.21%, specificity of 83.62%). The results demonstrate the potential of using speech as an objective biomarker of schizophrenia.

A study of ageing investigated the effectiveness of speech characteristics as measures of cognition. Speech features were automatically extracted from recordings of three speech tasks completed by a large cohort (N=172) of older adults. The speech task included a reading task and two picture description tasks. The cohort of older adults was divided into cognitively impaired and healthy groups. A Linear Discriminant Analysis (LDA) classifier trained on the

speech features extracted achieved classification accuracy of 80.4% in discriminating between cognitively impaired and healthy older adults. The results demonstrate the potential of using speech as an objective biomarker for cognitive function in older adults.

Another study demonstrated that the speech features employed in the assessment of cognitive function can be reliably extracted from telephone quality speech (overall accuracy of 93.2%, sensitivity of 97.3%, specificity of 89.5%). An Interactive Voice Response (IVR) technology was employed for remote automated delivery of cognitive function assessment interviews over the telephone. The developed IVR application was demonstrated to be functional and well received by the elderly population. Employment of the IVR technology in the cognitive assessment may allow for early detection of cognitive decline.

Another study investigated remote web-based assessment of cognitive function. A novel battery of tasks for assessment of memory and attention was developed. The assessment was undertaken by a cohort of older adults repeatedly for a period of eight weeks. The results of these tests were analysed for practice effect, test-retest reliability over time and validated against standard neuropsychological tests measures. The results suggest that these tests, while employing the same stimuli over time, do not suffer from practice effect and may potentially be administered at regular intervals over time. This feature is unique in cognitive assessment.

All studies received approval from the ethics committees of the hospitals and clinics involved, for both participant recruitment and data acquisition.

Speech features have been shown to discriminate between cognitively impaired and cognitively healthy individuals. Speech features may provide an objective measure of cognitive function and may be a complementary measure to existing neuropsychological cognitive assessments. Such cognitive assessment may be administered remotely, thus resulting in reducing some of the financial and labour intensive burden affecting healthcare teams and healthcare providers.

Indirect outcomes of the methods developed and presented in this thesis may lead to increased quality of life of older adults and may allow them to live as independently as possible, for as long as possible.

Acknowledgments

This thesis would not be possible without the support and contributions of many people.

First and foremost, I would like to thank Prof Richard Reilly for his guidance and unconditional support throughout my PhD studies. In particular, I am very grateful to Prof Richard Reilly for investing his time, for giving advice and encouragement, and for trusting me.

In addition, I would also like to thank Prof Ian Robertson for his valuable input into the research carried out in this thesis and for sharing his formidable expertise in neuropsychology.

I would like to express my gratitude to Dr Shona D'Arcy for her valuable input into the research carried out in this thesis and for her advice and help throughout my PhD studies.

I wish to thank Dr Jogin Thakore, Prof Brian Lawlor and Prof Rose Anne Kenny for their input into the research carried out in this thesis, in particular in providing clinical datasets.

Moreover, I would like to thank the current and former members of the Neural Engineering Group for creating a pleasant atmosphere to work in and for the lab social events. I owe gratitude to the following: Dr Edmund Lalor for sharing his impeccable knowledge of research, Ms June O'Reilly and Ms Melanie Apied for making the Neural Group run smoothly and for all their help throughout my PhD study, Ms Isabelle Killane and Mr Martin Holmes, for their feedback upon reading this thesis.

I wish also to thank the members of The Technology Research for Independent Living (TRIL) Programme and St. James's TRIL Clinic staff including Nils Pénard, Sian Counihan, Alessandra Galli, Nicola Burke.

I am also grateful to the members of the Neuro-Enhancement for Independent Lives (NEIL) Programme including Dr Sabina Brennan, Ms Caoimhe Hanningan, Ms Eleonora Greco, Ms Francesca Pigni.

Most importantly, this thesis would not have been possible if it were not for all the participants of all the studies included in this thesis.

I want to express my sincere gratitude to my family and friends for their support. My parents, Viliam and Edita, for their unconditional support, belief and encouragements throughout my PhD study. In addition my sisters, Edita and Katarína, and their families have been a wonderful sense of support. And my beautiful nieces and nephew for bringing joy into my life.

Table of Contents

Abstract iii

Acknowledgments v

Table of Contents vii

List of Figures xv

List of Tables xvii

Abbreviations xix

Chapter 1: Introduction 1

1.1 Thesis aims 7

1.2 Thesis outline 7

1.3 Contributions of this thesis 10

1.4 Publications arising from this thesis 11

1.4.1 Journals 11

1.4.2 Conferences 12

Chapter 2: Measuring changes in cognitive function 15

2.1 Changes in cognitive function in human ageing 15

2.2 Attention 16

2.2.1 Selective attention 17

2.2.2 Divided attention and attention switching 18

2.2.3 Sustained attention 19

2.2.4 Attention: Summary and implications 19

2.3 Memory 19

2.3.1 Working memory 19

2.3.2 Attentional resources 20

2.3.3 Speed of information processing 21

2.3.4 Inhibitory control 21

2.3.5 Working memory: Summary and implications 22

2.3.6 Long-term memory 22

2.3.7 Episodic memory 22
2.3.8 Semantic memory 24
2.3.9 Autobiographical memory 24
2.3.10 Procedural memory 24
2.3.11 Implicit memory 25
2.3.12 Prospective memory 25
2.3.13 Long-term memory: Summary and implications 25
2.4 Perception 26
2.5 Higher-level cognitive functions 27
2.5.1 Speech and language 27
2.5.2 Decision making 28
2.5.3 Executive control 28
2.6 Inter-individual variability in cognitive function 29
2.7 Cognitive function: Summary 30
2.8 Neurodegenerative conditions affecting cognitive function 30
2.8.1 Dementia 30
2.8.2 Mild cognitive impairment and Alzheimer's disease 30
2.9 Standard neuropsychological assessment of cognitive function 32
2.9.1 Assessment methods for cognitive function 32
2.9.2 Assessment of executive function 36
2.9.3 Assessment of memory 37
2.9.4 Assessment of attention 38
2.9.5 Assessment of mood, depression and personality 39
2.10 Limitations of current methods for cognitive function assessment for the older adults 43
2.10.1 MMSE variations 44
2.10.2 MMSE reliability 45
2.10.3 MMSE validity 46

2.10.4 Influence of social and demographic variables on MMSE score 46
2.10.5 Cut-off scores 47
2.11 Neuroimaging methods 48
2.11.1 Magnetic resonance imaging 48
2.11.2 Functional imaging 49
2.11.3 Electroencephalography 49
2.11.4 Limitations of imaging methods 50
2.12 Structural changes in the brain associated with impairment of cognitive function 50
2.13 Neural mechanisms of speech production 51
2.14 Physical effects of ageing on speech production 53
2.14.1 Changes in respiratory system 53
2.14.2 Changes in the larynx 55
2.14.3 Changes in the vocal tract 57
2.14.4 Neuromuscular control 58
2.15 Influence of impaired cognitive function on speech production in abnormal ageing ... 58
2.15.1 Dementia 58
2.15.2 Mild cognitive impairment 59
2.15.3 Alzheimer's disease 60
2.15.4 Other conditions associated with cognitive impairment 60
2.16 Influence of psychiatric conditions on speech production 61
2.16.1 Formal thought disorder 61
2.16.2 Schizophrenic language disorders 62
2.16.3 Schizophrenic language disorders vs. aphasia 63
2.16.4 Speech characteristics in schizophrenia 64
2.17 Summary 66
2.18 Research questions 67
2.18.1 Methodology questions to address research hypothesis 67

2.18.2 Speech characteristics in schizophrenia and their correlation to cognitive function 67

2.18.3 Changes in speech characteristics during ageing and their correlation to cognitive function 68

2.18.4 Remote assessment of cognitive function 68

Chapter 3: Signal processing of speech signals pertinent to cognitive function 71

3.1 Extraction of speech features 72

3.1.1 Speech/Non-Speech Threshold estimation 72

3.1.2 Automatic breath sounds detection and removal 73

3.1.3 Extraction of temporal features 78

3.1.4 Extraction of energy and fundamental frequency features 81

3.2 Classification of subjects into cognitively healthy and cognitively impaired cases 82

3.2.1 Classifier performance 82

3.2.2 Receiver operator characteristic (ROC) curves 83

3.3 Summary 84

Chapter 4: Speech as an objective biomarker for schizophrenia 85

4.1 Study 1 – Investigation of the speech characteristics of individuals with schizophrenia 86

4.1.1 Methods 87

4.1.2 Results 90

4.1.3 Discussion 93

4.2 Study 2 - Assessment of the performance of the algorithm for extraction of temporal features 96

4.2.1 Methods 96

4.2.2 Results 98

4.2.3 Discussion 100

4.3 Study 3 - Impact of the breath sounds detection and removal on the ability to discriminate between cognitively healthy and cognitively impaired cases 101

4.3.1 Methods 101

4.3.2 Results 102

4.3.3 Discussion 103

4.4 Summary 105

Chapter 5: Speech characteristics in ageing 107

5.1 Study 1 – Speech as an objective measure of cognitive function in older adults 109

5.1.1 Methods 109

5.1.2 Results 115

5.1.3 Discussion 119

5.2 Study 2 – Addressing the limitations of the Mini-Mental State Examination 125

5.2.1 Methods 125

5.2.2 Results 128

5.2.3 Discussion 131

5.3 Study 3 – Temporal features of speech as indicators of individual's memory state 134

5.3.1 Methods 134

5.3.2 Results 136

5.3.3 Discussion 139

5.4 Study 4 – Dynamic minimum pause threshold estimation 142

5.4.1 Methods 143

5.4.2 Results 146

5.4.3 Discussion 147

5.5 Summary 149

Chapter 6: Automated remote telephone assessment of cognitive function 151

6.1 Study 1 – Modifications of feature extraction algorithms for use with telephone speech recordings 154

6.1.1 Methods 154

6.1.2 Results 156

6.1.3 Discussion 157

6.2 Study 2 – Automated Interactive Voice Response application for remote assessment of cognitive function 158

6.2.1 Methods 159
6.2.2 Results 167
6.2.3 Discussion 173
6.3 Summary 177
Chapter 7: Automated remote web-based assessment of memory and attention 179
7.1 Methods 181
7.1.1 Cohort 181
7.1.2 Standard neuropsychological assessment 181
7.1.3 New battery of memory and attention tasks 183
7.1.4 Data acquisition 185
7.1.5 Data processing 187
7.1.6 Statistical analysis 188
7.2 Results 189
7.2.1 Validity 191
7.2.2 Analysis of practice effect 198
7.2.3 Test-retest reliability 199
7.3 Discussion 200
7.3.1 Validity 202
7.3.2 Analysis of practice effect 203
7.3.3 Test-retest reliability 205
7.3.4 Future directions 205
7.4 Summary 208
Chapter 8: Discussion 209
8.1 Thesis aims 209
8.1.1 Speech characteristics in schizophrenia and their correlation to cognitive function 210
8.1.2 Changes in speech characteristics during ageing and their correlation to cognitive function 211

8.1.3 Remote assessment of cognitive function 213
8.2 Future applications and directions 216
8.2.1 Long-term monitoring of cognitive function employing speech 216
8.2.2 Intervention and treatment monitoring 216
8.2.3 Existing speech data collections 217
8.2.4 Linguistic analysis 218
8.2.5 Influence of language on cognitive speech based measures 218
8.2.6 Other speech tasks 218
8.2.7 Other classification methods 219
8.2.8 Brain imaging methods 219
8.2.9 Transcranial Magnetic Stimulation 220
8.2.10 Brain modelling 221
8.2.11 Other aspects of cognitive function 221
8.3 Conclusion 221
Bibliography 225
Appendix A: Supporting material A-1
A.1 Heidi passage A-1
A.2 Picture Description A-2
A.3 Picture Taboo A-4

List of Figures

Figure 2-1: Stroop test [68] 17
Figure 2-2: Mini-Mental State Examination [30] 33
Figure 2-3: The Montreal Cognitive Assessment (MOCA) [31] 35
Figure 2-4: Center for Epidemiologic Studies Depression (CES-D) scale [144] 40
Figure 2-5: Hospital Anxiety and Depression Scale (HADS) [146] 41
Figure 2-6: Positive and Negative Affect Schedule (PANAS) [148] 42
Figure 3-1: Difference in amplitude envelope between breath sound and speech signal. 74
Figure 3-2: Breath sounds detection and removal steps. 78
Figure 3-3: Segmentation of the recording into speech and non-speech segments. Speech/Non-speech Profile represented by the red line. 79
Figure 3-4: Graphical User Interface for displaying the results of segmentation into speech and non-speech segments. 80
Figure 4-1: Artificially constructed speech recording processed with proposed feature extraction algorithm. 98
Figure 5-1: Sample picture of the Picture Taboo task. 112
Figure 5-2: Comparison of average feature values for Proportion of Recording in Silence 116
Figure 5-3: Accuracy in classifying subjects into cognitive classes as a function of classification method, feature set and speech type 131
Figure 5-4: Sample picture of the Picture Description task. 135
Figure 5-5: ROC curves representing the classification accuracy of the speech features extracted from the Picture Taboo (PT) task (blue line) and Picture Description (PD) task (green line). 139
Figure 5-6: Pause mixed lognormal distribution 145
Figure 6-1: The interview pack received by each participant. 160
Figure 6-2: Sample page of the interview guide used during the IVR assessment. 161
Figure 6-3: IVR application call flow diagram 165
Figure 6-4: Word Recall - Immediate and Word Recall – Delayed (WRD) completion rates for Group1 168
Figure 6-5: Word Recall - Immediate and Word Recall – Delayed (WRD) completion rates for Group 2 168
Figure 6-6: Percentage of complete VAAT tasks per group 169
Figure 7-1: Graphical user interface of the Word Recognition task. 184
Figure 7-2: Graphical user interface of the Shape Recognition task. 185

Figure 7-3: Website interface of automated remote web-based assessment of memory and attention. 186
Figure 7-4: Average number of correct answers for the Shape Recognition – Immediate test for Week 1 - 8 for both participants' groups. Including fitted linear regression lines. 191
Figure A-1: Image 1 of the Picture Description task. A-2
Figure A-2: Image 2 of the Picture Description task. A-3
Figure A-3: Image 2 of the Picture Description task. A-3
Figure A-4: Image 1 of the Picture Taboo task A-4
Figure A-5: Image 2 of the Picture Taboo task A-5
Figure A-6: Image 3 of the Picture Taboo task A-5
Figure A-7: Image 4 of the Picture Taboo task A-6
Figure A-8: Image 5 of the Picture Taboo task A-6

List of Tables

Table 3-1: Temporal features extracted from speech recordings 80
Table 3-2: Definitions of true positives/negatives and false positives/negatives. 83
Table 4-1: Patient profile, medication and psychiatric rating. 88
Table 4-2: Mean feature values (and standard deviations) for schizophrenic and control subjects. 90
Table 4-3: Linear Discriminant Analysis (LDA) classifier performance metrics. 91
Table 4-4: Correlation coefficients between features, symptom scales, dosages of medication and length of illness. 92
Table 4-5: True/false positives and true/false negatives definition for assessment of the performance of the feature extraction algorithm 97
Table 4-6: Combined performance metrics from three recordings and three listeners 99
Table 4-7: Combined performance metrics of the three recordings for each listener 99
Table 4-8: Impact of the breath sounds detection and removal on the LDA classifier performance 102
Table 4-9: Differences between average values of extracted features with/without prior breath sounds detection and removal 103
Table 5-1: Distribution of the cohort 110
Table 5-2: Correlation analysis between speech features of the Text Passage task and the MMSE group, and speech features and the cognitive tasks scores 117
Table 5-3: Correlation analysis between speech features of the Picture Taboo task and the MMSE group, and speech features and the cognitive tasks scores 118
Table 5-4: LDA classification results for cognitive and speech features 119
Table 5-5: Demographic information of the cohort 126
Table 5-6: Average speech feature values and standard deviations for the Text passage task using MMSE categorization of subjects 128
Table 5-7: Average speech feature values and standard deviations for the Text passage task using CCS categorization of subjects 129
Table 5-8: Average speech feature values and standard deviations for the Picture Taboo task using MMSE categorization of subjects 129
Table 5-9: Average speech feature values and standard deviations for the Picture Taboo task using CCS categorization of subjects 130
Table 5-10: Average temporal features values for Low and High memory performers groups and both speech based tasks 137

Table 5-11: Pearson's r correlation values between the variables education, age, MMSE scores, MMSE memory subscale scores and temporal features for the Picture Description task 137
Table 5-12: Pearson's r correlation values between the variables education, age, MMSE scores, MMSE memory subscale scores and temporal features for the Picture Taboo task 138
Table 5-13: Student's t-test analysis of speech features between participants groups.....146
Table 5-14: Wilcoxon Rank-Sum test analysis of speech features between participants groups 146
Table 5-15: Mean feature values for features that achieved statistically significant differences for both participants' groups 147
Table 5-16: Linear Discriminant Analysis classification.....147
Table 6-1: Average values comparison for clinic and telephone recordings. 156
Table 6-2: Participants group distribution. 159
Table 6-3: Reliability measures between F2F and IVR assessments.....171
Table 7-1: Demographic data for Low and High performers groups 181
Table 7-2: Recognition tasks - Mean values for each task measure and group separately.....189
Table 7-3: SART - Mean values for each task measure and group separately.....190
Table 7-4: Validation of Week 1 recognition tasks scores against standard neuropsychological tests scores 192
Table 7-5: Validation of Week 2 recognition tasks scores against standard neuropsychological tests scores 193
Table 7-6: Validation of average Week 1 – 8 recognition tasks scores against standard neuropsychological tests scores 194
Table 7-7: Validation of Week 1 SART task scores against standard neuropsychological tests scores 195
Table 7-8: Validation of Week 2 SART task scores against standard neuropsychological tests scores 196
Table 7-9: Validation of average Week 1 – 8 SART task scores against standard neuropsychological tests scores 197
Table 7-10: Analysis of practice effect - linear trends 198
Table 7-11: Magnitude of practice effect of the Word and Shape Recognition and SART tasks measures.....199
Table 7-12: Test-retest reliability of the Word and Shape Recognition and SART tasks measures.....200
Table 8-1: Status of the research questions 222

Abbreviations

A_{max}	Maximum amplitude value
ACAD	Automatic Cognitive Assessment Delivery
AD	Alzheimer's disease
ASR	Automatic speech recognition
AUC_ROC	Area under the ROC curve
BA	Broca's area of the brain
BPRS	Brief Psychiatric Rating Scale
CCS	Combined Cognitive Score
CDR	Clinical Dementia Rating scale
CES-D	Center for Epidemiologic Studies Depression scale
CF	Category Fluency
CoVE	Coefficient of Variation of Mean Energy per Second
$CoVF_0$	Coefficient of Variation of Fundamental Frequency
$CoVF_0A$	Coefficient of Variation of Fundamental Frequency Amplitude
CoV_RT	Coefficient of Variation of Reaction Time
d	Cohen's *d* coefficient
DA	Discriminant Analysis
DDK	Diadochokinetic rate
DS	Digit Span task
DSM	Diagnostic and Statistical Manual of Mental Disorders
DT-MRI	Diffusion tensor magnetic resonance imaging
DTMF	Dual tone modulated frequency
E_x	Energy of a discrete time signal
E_{log}	Log energy
EEG	Electroencephalography
FN	False Negative
fMRI	Functional Magnetic Resonance Imaging
FP	False Positive
GDS	Global Deterioration Scale
HADS	Hospital Anxiety and Depression Scale
IADL	Instrumental Activities of Daily Living
IVR	Interactive Voice Response
LDA	Linear Discriminant Analysis

MANOVA	Multivariate Analysis of Variance
MCI	Mild Cognitive Impairment
MES	Mean Energy per Second
MF_0	Mean Fundamental Frequency
MF_0A	Mean Fundamental Frequency Amplitude
MMSE	Mini-Mental State Examination
MOCA	Montreal Cognitive Assessment
MPD	Mean Pause Duration
MRI	Magnetic Resonance Imaging
MUD	Mean Utterance Duration
MySQL	Open Source Database
NA	Negative affect
NART	National Adult Reading Test
NP	Number of Pauses
PA	Positive affect
PANAS	Positive and Negative Affect Schedule
PANSS	Positive and Negative Syndrome Scale
PD	Picture Description
PET	Positron emission tomography
RGB	Red Green Blue scale
PR	Phoneme Rate
PRS	Proportion of Recording in Silence
PT	Picture Taboo
ROC	Receiver Operator Characteristic
rTMS	Repetitive Transcranial Magnetic Stimulation
RV	Reference Value
SANS	Scale for Assessment of Negative Symptoms
SART	Sustained Attention to Response Task
SDF_0	Standard Deviation of Fundamental Frequency
SDF_0A	Standard Deviation of Fundamental Frequency Amplitude
SD_RT	Standard Deviation of Reaction Time
SDMES	Standard Deviation of Mean Energy per Second
SNP	Speech/Non-Speech Profile
SPECT	Single photon emission computer tomography

SRec	Shape Recognition
SRec-D	Shape Recognition Delayed
SRec-I	Shape Recognition Immediate
SsLM	Standard score of Logical Memory
SsNART	Standard score of NART
SWF	Adobe Flash file format used for multimedia, vector graphics
TDEA	Triple Data Encryption Algorithm
Th_{GRAD}	Gradient Threshold
Th_{SN}	Speech/Non-Speech Threshold
TLP	Total Length of Pauses
TLU	Total Length of Utterances
TMS	Transcranial Magnetic Stimulation
TN	True Negative
TP	True Positive
TRIL	Technology Research for Independent Living Programme
TRT	Total Recording Time
VAAT	Vigilant Auditory Attention Task
WMS	Wechsler Memory Scale
WoZ	Wizard of Oz
WRD	Word Recall Delayed
WRec	Word Recognition
WRec-D	Word Recognition Delayed
WRec-I	Word Recognition Immediate
WRI	Word Recall Immediate

Chapter 1

Introduction

The ageing of the world's population - in developing and developed countries - is an indicator of improving global health. The world's population of people 60 years of age[1] [1] and older has doubled since 1980 and is forecast to reach 2 billion by 2050, representing 21% of the world's population [2]. In the middle of the 20th century there were just 14 million people on the whole planet aged 80 years or older. By 2050, there will be 100 million in this age group living in China alone, and 400 million people worldwide [3].

While this is a cause for celebration, it comes with special health challenges for the 21st century. It is estimated that 25-30% of people aged 85 or older have some degree of cognitive decline [3]. Ferri at al. [4] estimated that 24.3 million people had dementia in 2005, with 4.6 million new cases of dementia diagnosed every year (one new case every 7 seconds). The number of people affected was estimated to double every 20 years to 81.1 million by 2040. In 2005, most people with dementia lived in developing countries (60% in 2001, rising to 71% by 2040). Rates of increase are not uniform; numbers in developed countries are forecast to increase by 100% between 2001 and 2040, but by more than 300% in India, China, and their south Asian and western Pacific neighbours [4]. The prevalence of dementia increases continuously with age. Lobo et al. [5] found that in European population-based studies the prevalence of dementia was 0.8% in the group age of 65 to 69 years. At the age of 90 years and older, the prevalence of dementia increased to 28.5% [5]. With this increase in the prevalence of dementia with age, the rapid ageing of the population will inevitably result in the increase in the number of people with cognitive impairment.

[1] The United Nations has not adopted a standard definition of 'older adults', but they refer to adults of 60+ years as older adults.

Dementia is a chronic disease of ageing characterised by progressive cognitive decline that interferes with independent functioning [6] and affects a large and growing number of older adults [7, 8]. Dementia is the greatest cause of dependency (i.e. need for care) and disability in high-income countries and the second greatest cause of dependency and disability worldwide [3]. According to different estimates, between 2% and 10% of all cases of dementia start before the age of 65 [9]. The prevalence doubles with every five-year increment in age after 65. Dementia affects 20% of people aged over 80 years [10]. More recent data suggests that, in 2010, there were 35.6 million people living with dementia globally [11]. Epidemiological studies indicate that this number is expected to grow at an alarming rate. It is estimated that numbers will nearly double every 20 years, to 65.7 million in 2030 and 115.4 million in 2050 [3].

The economic impact of impaired cognitive function of older population represents a significant burden on the global economy. The total estimated worldwide cost of dementia in 2010 was US $604 billion [9]. The total monetary cost of dementia in 2010 in the United States was estimated between $157 billion and $215 billion [12]. Dementia is among the diseases that are most costly to society. The cost for dementia care purchased in the market-place ($109 billion) was similar to estimates of the direct health care expenditures for heart disease ($102 billion) and significantly higher than the direct health care expenditures for cancer ($77 billion) [12]. Dementia represents only one type of cognitive impairment. The global cost of cognitive impairment is significantly higher when the cost of psychiatric conditions associated with cognitive impairment is included. The overall 2002 cost of schizophrenia in USA only was estimated to be $62.7 billion [13].

While the numbers and the costs are daunting, the impact on those with the illness and on their caregivers and families is extreme – medically, psychologically and emotionally. The behavioural and psychological symptoms linked to cognitive impairment profoundly affect the quality of life of people with cognitive impairment and their caregivers [9]. Many of the very old lose their ability to live independently because of limited mobility, frailty or other physical or mental health problems. Many require long-term care, including home-based nursing, community, residential and hospital-based care [3].

Dementia has an immense impact on the lives of the family, and particularly the person who takes the primary role in providing care. Most care is provided by family and other informal support systems in the community. However, changing population demographics may reduce the availability of informal caregivers in the future. The provision of care to a

person with dementia can result in significant strain for those who provide most of that care. The stressors are physical, emotional and economic [9].

Social impacts of cognitive impairment may include a reduction in work hours or loss of employment, loss of relationships, time with friends and families and social activities, or the need to relocate or change living arrangements in order to provide care. Health impacts include depression, anxiety, stress, physical problems and sleep disruption. Additional stresses can occur if the family carer is older and in failing health themselves [14].

Studies in high income countries show that only one-fifth to one-half of cases of dementia are routinely recognised and documented in primary care facilities, as inpatients or visiting outpatients [15-19]. This 'treatment gap' is much greater in low and middle income countries, with one study in India suggesting 90% remain unidentified [20].

Providing early detection of the onset of decline in cognitive function and subsequent intervention through improvement in diagnosis is essential. Eating a healthier diet, exercising regularly [21], pharmacological and non-pharmacological interventions [11], such as cognitive training [22], may reverse or delay the progress of decline in cognitive function. Greenaway et al. [23] employed procedure-based memory compensation training models in a study of amnestic mild cognitive impairment (MCI)[2]. Subjects that received a 6-week memory training intervention showed a statistically significant improvement in activities of daily living over subjects that did not receive compensation training [23]. Several other randomized trials have also demonstrated that subjects with MCI that have undergone cognitive training show improved cognition, mood and psychological well-being [24-27]. A systematic review of the literature examined the efficacy of cognitive intervention programs in people with amnestic MCI [28]. Of the 15 studies that met the inclusion criteria, all found a statistically significant improvement on subjective and/or objective memory measures. Other novel methods for reversing cognitive decline are being developed. These methods combine optimized training protocols with traditional drug treatments [29], which may lead to further restoring of cognitive function.

Early interventions have the potential to improve cognitive function, delay institutionalisation, reduce carer strain and psychological illness and improve the quality of life [11]. Earlier diagnosis allows people with impaired cognitive function to plan ahead while they

[2] Mild cognitive impairment (MCI) causes a slight but noticeable and measurable decline in cognitive abilities, including memory and thinking skills. A person with MCI is at an increased risk of developing Alzheimer's or another dementia.

still have the capacity to make important decisions about their future care. In addition, they and their families can receive timely practical information, advice and support. Only through receiving a diagnosis can they get access to available drug and non-drug therapies that may improve their cognition and enhance their quality of life [11]. Improving the likelihood of earlier diagnosis of onset of cognitive decline can be enhanced through the introduction of accessible diagnostic methods [11].

The assessment of cognitive function in older adults is currently carried out by employing gold-standard neuropsychological tests, such as the Mini-Mental State Examination [30], Montreal Cognitive Assessment [31], Wechsler Memory Scale [32], and others. While the neuropsychological tests are well known and widely used, they face several major limitations. Firstly, these tests are time consuming to administer. The individual, whose cognitive function is being assessed, is usually administered a battery of neuropsychological tests. Typically, it can take 1-3h to administer the full battery [33]. Given the time needed for the cognitive assessment community neuropsychologists fail to diagnose dementia in over 50% of dementia cases, particularly in the earlier mild to moderate stages [34].

These tests need to be administered by trained neuropsychologists [33]. Even with the best intent of the neuropsychologists to acquire as objective test score as possible, the test-retest reliabilities of these test are affected by several factors impacting the objectiveness of the test [35]. Lack of explicit scoring criteria is one of the reported factors affecting objectivity of these tests [36]. The subjective nature of these tests may further be amplified by factors such as the current state (e.g. mood, physical health, fatigue, performance anxiety and others) of the individual being assessed [37]. The visit to the clinic and the associated "white coat effect" while the individual is being assessed may have an impact on the mood and anxiety of the individual. Furthermore, age, education, cultural and socioeconomic background can cause a considerable bias in the outcomes of these tests [38]. Additionally, these tests are not commonly designed for repeated administration and subject's performance is enhanced due to a practice effect [39]. All these factors may affect the outcomes of a test and may result in cognitively healthy person being considered as impaired [40].

Employing neuroimaging methods to assess functional or structural changes in the brain represents another means of cognitive function assessment in older adults. These methods include Magnetic Resonance Imaging (MRI), Electroencephalography (EEG), Single photon emission computer tomography (SPECT), and others [41]. Detection of pathologic features like senile plaques, neurofibrillary tangles, decreased synaptic density, neuron loss, and atrophy

relative to age-matched controls by MRI has been employed in studies of Alzheimer's disease[3] [42] with some limited success. Positron emission tomography and SPECT techniques have demonstrated specific regional abnormalities in brain perfusion among patients with established Alzheimer's disease [43, 44]. Quantitative EEG assessment of brain activity, employing frequency power analysis, event-related potentials and P300 latency analysis, has been used in studies of impaired cognitive function, such as dementia, Alzheimer's disease, schizophrenia [45-47].

Despite the advances in neuroimaging technology and the availability of facilities with neuroimaging equipment, the use of neuroimaging methods is generally expensive, time consuming and requires the individual to visit the hospital/clinic. Neuroimaging assessments may take place at different neuroimaging centres. The protocols for image acquisition may not be well defined and may differ between centres resulting in poor diagnosis sensitivity [48]. If not evaluated by a computer algorithm, substantial variability between clinical raters has been observed [48-50], which also suggests that a significant number of subjects may be misclassified [48]. Head position in MRI scanner has been found to increase variability across clinical raters [48]. All these factors may contribute to variable and suboptimal treatment decisions.

A biomarker is defined as "a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" [51]. Biomarkers, that would resolve the issues affecting objectivity of current methods for assessment of cognitive function, are actively being sought in the research of impaired cognitive function of older adults [52-55].

Structural and functional changes in the regions of brain involved in speech and language production in subjects with impaired cognitive function have been reported in the literature [42, 56-58]. These regions of brain include, among others, prefrontal cortex, Broca's area, left superior temporal sulcus and neural pathways. It is hypothesized that the structural and functional changes in these brain regions may lead to changes in speech characteristics during the process of speech generation.

The influence of impaired cognitive function on speech production in abnormal ageing has been studied and reported in the literature [59-61]. Dementia is characterised by a

[3] Alzheimer's disease is a progressive, degenerative disorder that attacks the brain's nerve cells, or neurons, resulting in loss of memory, thinking and language skills, and behavioral changes. Alzheimer's disease is the most common cause of dementia, or loss of intellectual function, among people aged 65 and older.

breakdown in intellectual and communicative functioning, manifesting itself as a communication disorder [62]. Articulation rate features have been investigated in a study of dementia [63]. Standardized pause rate (the ratio of words uttered to the number of pauses uttered) has been found to be statistically different in subjects with mild cognitive impairment compared to healthy subjects [61]. Qualitative changes in speech characteristics have been included in diagnostics manuals of conditions associated with impaired cognitive function [6].

Innovation and technology can help older adults in many ways. To better monitor health status and detect early signs of disease and connect older people to health care assessment. Use of new methods, such as eHealth and connected health, to ensure better data collection, develop new versions of diagnostic, monitoring and assistive devices, as well as to assist older adults with functional loss and to remain independent.

Current healthcare systems are often criticized for treating people as averages, not unique individuals. While comparing an individual against average measures in cross-sectional studies has been employed extensively in the research, it is the longitudinal monitoring of individual's measures that has the profound advantage over a cross-sectional approach. Baseline measures can be obtained for each individual and it is the deviation from these baseline measures that would inform of the onset of a worsening health condition. Longitudinal monitoring would also allow measurement of the progression of a disease. Appropriate treatment methods may be employed as a result of monitoring of treatment effects.

Continuously more wide-spread use of mobile devices, such as mobile phones, tablets, allow capture of speech, video, tactile information and, with add-on devices, specific biosignals from the user. Cognitive assessment tasks may be developed that will be deployed on these mobile devices and will be able to be administered by non-specialists, i.e. the general practitioners or even the users themselves. This will provide a non-invasive, "light-touch" manner of cognitive assessment delivery. Combining the use of mobile devices with the advancement and emergence of technologies, such as high performance computing, cloud storage, big data set analysis, will enable us to build predictive models for each individual and personalized health care. This would allow for remote, accurate, pervasive, uninterrupted, prevention-driven monitoring of individual's health. This would further reduce the logistical issues associated with cognitive function assessment mentioned above, as well as enable remote population scale screening at a very low cost.

1.1 Thesis aims

Expanding on the overall discussion presented above, the following aims were established for this thesis.

- To assess whether speech characteristics can provide objective quantitative measures to distinguish between cognitively impaired and healthy subjects.
- To assess whether the cognitive assessment can be administered remotely, in a "light touch" manner.

1.2 Thesis outline

Chapter 2

Chapter 2 provides a thorough review of the biology of cognitive function in older adults. At first, specific domains of cognitive function are presented to demonstrate its complexity. Following, conditions affecting cognitive function in older adults are addressed. Methods for standard neuropsychological assessment of cognitive function are presented and reviewed in this chapter. The limitations of these methods are addressed. Then, advanced neuroimaging methods and their use in studies of cognitive function are presented. Structural and functional changes in brains of cognitively impaired adults and their potential impact on communication are discussed. Description of physiological and neural mechanism of speech production follows. Effects of ageing on speech production are also presented in this chapter. Previous studies in the literature in the area of speech analysis of cognitively impaired individuals are discussed. Literature review of previous work on influence of psychiatric conditions on speech production is assessed. Concluding Chapter 2, a series of research questions are presented.

Chapter 3

Chapter 3 introduces a signal processing algorithm to identify cognitive function from speech recordings acquired in real-life environments. Description of an algorithm for breath sounds detection and removal from speech recordings acquired in real-life environments follows. Algorithms designed for energy and fundamental frequency estimation are also described in this chapter. The description of all feature extraction algorithms is followed by presentation of methods for classification of subjects into healthy and abnormal cases employing the speech features.

Chapter 4

Three studies are presented in Chapter 4. The first study investigates the use of speech as a possible biomarker for schizophrenia. Acoustic and temporal features are extracted from speech recordings of schizophrenic and healthy control subjects using the algorithms described in Chapter 3. An assessment of the ability to discriminate between schizophrenic and healthy control subjects employing the extracted speech features is carried out.

In Study 2, an assessment of the performance of the algorithm for extraction of temporal features from speech recordings acquired in real-life environments is carried out. Three tests are performed to assess the performance of this algorithm. In the first test, artificially constructed speech recording is employed. The results of segmentation into speech and pauses of three speech recordings by three listeners are compared to results of the developed algorithm in the second test. The final test compares segmentation results of 20 recordings hand-labelled by one listener to the segmentation results of the developed algorithm.

The final study of this chapter, Study 3, investigates the impact of breath sounds detection and removal on the ability to discriminate between healthy and abnormal cases. Two sets of temporal speech features are extracted from speech recordings with and without prior breath sounds detection and removal. A classifier is trained for each set of features separately. The ability of the classifier to discriminate between schizophrenic and healthy control subjects is compared for both sets of features.

Chapter 5

In Chapter 5, a speech corpus containing data from three speech based tasks is employed to study cognitive function of older adults. Acoustic and temporal features are extracted from the speech recordings using algorithms described in Chapter 3. Four studies are presented in the Chapter 5.

Study 1 presents analysis of speech recordings from a cohort of older adults producing read and spontaneous speech. The cohort is divided into cognitively impaired and cognitively healthy groups based on a standard neuropsychological measure. The ability to discriminate between cognitively impaired and healthy subjects employing the extracted speech features is assessed.

In Study 2, the use of a combined cognitive score as a novel neuropsychological measure is proposed. The cohort is split into cognitively impaired and cognitively healthy groups based on the combined cognitive score. The ability to discriminate between cognitively impaired and

cognitively healthy subjects employing the extracted speech features is assessed. Comparison of the discrimination abilities employing standard neuropsychological measure and combined cognitive score is carried out.

Study 3 investigates whether speech features are related to more specific cognitive domain, i.e. memory, rather than overall cognitive measure. Assessment of ability to discriminate between low and high memory performers employing the extracted speech features is carried out.

In the last study, Study 4, a dynamic minimum pause threshold estimation for extraction of temporal speech features is proposed. Three separate sets of features are extracted from each recording. The ability to discriminate between cognitively impaired and cognitively healthy older adults employing the temporal speech features is assessed. A comparison of discrimination ability of three sets of speech features is carried out.

Chapter 6

The design and evaluation of a novel protocol for a remote, fully-automated, telephone assessment of cognitive function in older adults is presented in Chapter 6. Test-retest reliability and practice effect of multiple administration of the cognitive assessment are assessed. Validation of the cognitive assessment is carried out. Investigation of cost effectiveness and user experience of the developed cognitive assessment is performed.

Chapter 7

In Chapter 7, a protocol and implementation of an automated, remote, web-based assessment of memory and attention is presented. A novel battery of memory and attention assessment tests is designed. Test-retest reliability and practice effect of multiple administration of the assessment to a cohort (N = 16) of older adults are assessed. Validation of the memory and attention assessment protocol and battery is carried out in this chapter.

Chapter 8

The final chapter addresses the aims, applications and contributions of this thesis and outlines some potential directions for future work.

1.3 Contributions of this thesis

Aspects of novelty in this thesis can be summarized as follows.

Chapter 3

- Development of novel algorithm for temporal features extraction from recordings of speech acquired in real-life environments.
- Development of novel algorithm for breath sound detection and removal from recordings of speech acquired in real-life environments.

Chapter 4

- Application of developed algorithms to recordings of speech of schizophrenic and healthy adults.
- Assessment of ability to discriminate between schizophrenic and healthy adults using temporal and acoustic features of speech.
- Assessment of the performance of the algorithm for extraction of temporal features.
- Investigation of impact of breath sounds detection and removal on the ability to discriminate between healthy and abnormal cases.

Chapter 5

- Generation of a speech corpus that contains data from three speech based tasks for the purpose of assessment of cognitive function in older adults.
- Novel phoneme rate speech features are proposed.
- Assessment of ability to discriminate between cognitively impaired and cognitively healthy older adults using temporal and acoustic features of speech.
- Investigation into ability of various speech based tasks to discriminate between cognitively healthy and cognitively impaired older adults.
- Investigation of dynamic minimum pause threshold estimation impact on ability to discriminate between cognitively impaired and cognitively healthy older adults using temporal features of speech.

Chapter 6

- Design and evaluation of novel protocol for remote fully-automated telephone assessment of cognitive function in older adults.
- Assessment of test-retest reliability of the remote fully-automated telephone cognitive assessment over time.
- Analysis of practice effect of the remote fully-automated telephone cognitive assessment over time.
- Validation of the remote fully-automated telephone cognitive function assessment protocol.
- Cost assessment of usage of the remote fully-automated telephone assessment of cognitive function.
- Investigation of user experience of the remote fully-automated telephone assessment of cognitive function.

Chapter 7

- Development of novel battery of tests for an automated remote web-based assessment of memory and attention.
- Development of novel tests for word recognition and shape recognition memory assessment.
- Assessment of test-retest reliability of the automated remote web-based assessment of memory and attention.
- Analysis of practice effect of the automated remote web-based assessment of memory and attention.
- Validation of the automated remote web-based assessment of memory and attention.

1.4 Publications arising from this thesis

A number of publications have been derived from the research described in this thesis.

1.4.1 Journals

S. D'Arcy, **V. Rapcan**, A. Galli, N. Burke, G.C. O'Connell, I.H. Robertson, R.B. Reilly, "A study into the automation of cognitive assessment tasks for delivery via the telephone: lessons for developing remote monitoring applications for the elderly", *Technology and Health Care*, vol. 21, pp. 387-396, 01/01/ 2013

J. Thakore, **V. Rapcan**, S. D'Arcy, S. Yeap, N. Afzal, R.B. Reilly, "Acoustic and temporal analysis of speech: a potential marker for Schizophrenia", *International Clinical Psychopharmacology*, vol. 26, p. e131, 2011.

V. Rapcan, S. D'Arcy, S. Yeap, N. Afzal, J. Thakore, R.B. Reilly, "Acoustic and temporal analysis of speech: A potential biomarker for schizophrenia", *Medical Engineering & Physics*, vol. 32, pp. 1074-1079, 2010.

1.4.2 Conferences

I. Rochford, **V. Rapcan**, S. D'Arcy, R.B. Reilly, "Dynamic Minimum Pause Threshold Estimation for Speech Analysis in Studies of Cognitive Function in Ageing", *Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp. 3700-3703.

N. Burke, **V. Rapcan**, A. Galli, S. D'Arcy, R.B. Reilly, I.H. Robertson, "Improving the sensitivity of the animal naming category fluency test using temporal speech features", *Proceedings of the 8th Annual Psychology, Health and Medicine Conference*, Galway, April 4th, 2011.

A. Galli, C. Garattini, N. Burke, **V. Rapcan**, S. D'Arcy, R.B. Reilly, I.H. Robertson, "Cognitive function monitoring using temporal speech features", *Proceedings of the International Association of Gerontology and Geriatrics-VII European Congress*, Bologna, April 2011.

V. Rapcan, S. D'Arcy, and R. B. Reilly, "Speech analysis as a monitoring tool for schizophrenia", *Proceedings of the 16th Bioengineering in Ireland Conference*, Dublin, Ireland, January 22-24, 2010.

S. D'Arcy, S. Counihan, N. Penard, F. van den Berg, **V. Rapcan**, M. Hansteen Izora, R.B. Reilly, I.H. Robertson, "Technology for remote assessment of cognitive function of the elderly", *Proceedings of the 16th Bioengineering in Ireland Conference*, Dublin, Ireland, January 22-24, 2010.

N. Pénard, S. Counihan, S. D'Arcy, **V. Rapcan**, R.B. Reilly, I.H. Robertson, "Investigating Speech as a Source of Biomarkers for Changes in Cognition, Executive Function and Mood", *The 47th Annual Meeting of the Academy of Aphasia*, Boston, Massachusetts, October 18-20, 2009.

V. Rapcan, S. D'Arcy, N. Penard, I. H. Robertson, and R. B. Reilly, "The Use of Telephone Speech Recordings for Assessment and Monitoring of Cognitive Function in Elderly People," *Proceedings of Interspeech 2009*, Brighton, UK, 2009, pp. 943-946.

V. Rapcan, S. D'Arcy, and R. B. Reilly, "Automatic breath sound detection and removal for cognitive studies of speech and language," *Proceedings of the 20th IET Irish Signals and Systems Conference 2009*, pp. 1-6.

S. D'Arcy, N. Penard, S. Counihan, **V. Rapcan**, I.H. Robertson, R.B. Reilly, "Corpora for assessment of cognitive function in older persons", *Proceedings of the 3rd Advanced Voice Function Assessment International Workshop*, Madrid, Spain, 18-20 May 2009.

S. D'Arcy, **V. Rapcan**, N. Pénard, M.E. Morris, R.B. Reilly, I.H. Robertson, "Speech as a Means of Monitoring Cognitive Function of Elderly Subjects", *Proceedings of Interspeech 2008*, Brisbane, Australia, 22-26 September 2008.

N. Pénard, S. D'Arcy, **V. Rapcan**, P. Dockree, C. Besani, K. O'Sullivan, R.B. Reilly, I.H. Robertson, "Investigating Speech as a Diagnostic Tool for Abnormal Ageing", *Proceedings of the 2008 British Psychological Society Annual Conference*, Dublin, April 2008.

Chapter 2

Measuring changes in cognitive function

The biggest limiting factor to independence in the older adults is impaired cognitive function and its consequences [64]. Such consequences include: accident proneness (falls, burns, bruising and cuts), self-neglect (missed medication, poor nutrition, poor hygiene), loss of initiative, diminished repertoire of activities and low mood. The assessment of cognitive function is expensive and labour intensive and hence non-viable for all but a tiny fraction of the elderly population.

The need for objective cognitive function assessment as outlined in Chapter 1 is increasing. This need is driven by the increasing rise in our older population. To measure cognitive function and specifically subtle changes in such function, one needs to understand the cognitive changes that occur during the ageing process.

2.1 Changes in cognitive function in human ageing

As people age, they change in many ways both biologically and psychologically. This section is focused on the neuroanatomical and neurophysiological changes that occur with age, and the mechanisms that account for them.

While there is evidence that alterations in brain structure and function are linked to alterations in cognitive function, the complexity of both the neural and cognitive functions make an exact mapping between brain and behaviour very difficult. Establishing such links between brain and cognition is the principal goal of cognitive neuroscience.

The purpose of this section is to review the changes in cognition that occur in normal human ageing, in order to provide a foundation against which changes can be assessed and

interpreted. The relationship between brain and cognition is a dynamic one and may change across the life course. The age-related changes in brain structure and function do not occur uniformly across the whole brain. Nor do they occur uniformly across older individuals. Age-related changes in cognition are also not uniform across all cognitive domains [65].

The cognitive functions most affected by age are attention and memory. There are many aspects to both of these functions and evidence suggests that some aspects of attention and memory remain intact with age while others show significant decline. Perception also shows significant age-related decline due to declining sensory capacities. Deficits at these early processing stages are considered to affect cognitive functions later in the processing stream. Higher-level cognitive functions such as language processing and decision-making may also be affected by age. These higher-level functions rely on more basic cognitive functions and these generally show deficits to the extent that those fundamental processes can become impaired. Moreover, complex cognitive tasks may also depend on a set of executive functions, which manage and coordinate the various components of the tasks. Considerable evidence points to impairment of executive function as a key contributor to age-related declines in a range of cognitive tasks [66].

Although the impression of old age is one of cognitive decline, enormous variability exists across individuals. There is considerable interest in the researcher community on what accounts for this variability. This section aims to highlight the cognitive domains that show the greatest decline with age and which are also the most variable. Theories of cognitive ageing that have developed in the literature within each cognitive domain are outlined and the brain regions to underlie these functions presented.

This section reviews evidence from the literature for age-related impairments in basic cognitive functions, focusing primarily on attention and memory.

2.2 Attention

Attention is a basic yet complex cognitive process that has multiple sub-processes specialized for different aspects of attentional processing. An aspect of attention is involved in all cognitive domains, except when task performance has become habitual or automatic. Declines in attention can therefore have broad-reaching effects on one's ability to function adequately and efficiently in everyday life. Attention has been partitioned in many ways by different researchers in the literature. The divisions described in the following sections are

those that have been most extensively investigated and reported in the literature with respect to normal ageing [67].

2.2.1 Selective attention

Selective attention refers to the ability to attend to one specific stimulus while disregarding other stimuli that are irrelevant. For example, in visual search tasks, subjects are asked to search a visual display for a target letter that is surrounded by other non-target letters. The task can be made more difficult by increasing the similarity of targets and distractors. For example, searching for an *F* in a background of *E*s, or by increasing the number of relevant or irrelevant features that are part of the search criteria. In another task, the Stroop task [68], the subject is asked to name the colour of the ink in which an incongruent colour word is printed. For example, the word "red" printed or displayed in green. Here, the word information tends to interfere with colour naming, causing errors and an increase in response times. To perform well in these kinds of tasks, subjects have to select the relevant stimulus or dimensions for processing and ignore the irrelevant ones. Although findings are not entirely consistent across studies and may differ across tasks, in general older adults appear to be slower than younger adults in responding to the targets, but are not differentially affected by distraction [67, 69]. Thus, deficits found in many of these tasks can be largely attributed to a general slowing of information processing in older adults rather than to selective attention deficits specifically.

BROWN	RED	BLUE	BLUE	RED
RED	PURPLE	GREEN	ORANGE	GREEN
ORANGE	BROWN	PURPLE	GREEN	PURPLE
GREEN	BLUE	ORANGE	BROWN	ORANGE
BLUE	ORANGE	RED	RED	BROWN
PURPLE	BROWN	BLUE	BLUE	GREEN
RED	PURPLE	BROWN	PURPLE	RED
GREEN	BLUE	BLUE	BROWN	PURPLE
BROWN	RED	RED	ORANGE	BLUE
ORANGE	BROWN	ORANGE	BLUE	GREEN
PURPLE	GREEN	PURPLE	BROWN	ORANGE
GREEN	ORANGE	GREEN	PURPLE	RED

***Figure 2-1**: Stroop test [68]*

2.2.2 Divided attention and attention switching

Divided attention has usually been associated with significant age-related declines in performance, particularly when tasks are complex. Divided attention tasks require the processing of two or more sources of information or the performance of two or more tasks at the same time. For example, subjects are asked to monitor stimuli at two different spatial locations, or they may be asked to make semantic judgments about visually presented words while simultaneously monitoring for the occurrence of an auditory presented digit [70]. The impact of dividing attention is assessed by comparing performance under dual task conditions to performance when the tasks are performed separately. Studies in the literature suggest that older adults are more affected by the division of attention than young adults, particularly when the attentional demands of the two tasks are high. In addition, older adults seem less able to allocate resources appropriately when instructions are given to vary task priority [71]. These findings cannot be completely accounted for by a general slowing of information processing, but instead are typically explained in the literature in terms of declining processing resources associated with normal ageing. Such limited resources are believed to be over-extended in older adults when attention must be divided between two or more sources. Similarly, the performance of older adults is slowed to a greater degree than that of young adults when attention must be switched from one task to another [69].

There is evidence in the literature that age deficits in divided attention and attention switching can be reduced by practice or extended training [72] and even by aerobic exercise [73]. However, the exact mechanism of such improvements is unclear. In the case of task-specific training, it is possible that some aspects of the tasks become automatic with practice, thus requiring fewer attentional resources.

Alternatively, subjects may develop strategies with extensive training that reduce the attentional demands of the tasks. It has been hypothesized that cardiovascular fitness may improve the efficiency of neural processes or may provide increased metabolic resources for task performance. Interestingly, the enhancement effects of aerobic exercise appear to be greatest on tasks involving executive control of attention [74], which depend largely on prefrontal cortex.

2.2.3 Sustained attention

Sustained attention refers to the ability to maintain concentration on a task over an extended period of time. Typically, vigilance tasks are used to measure sustained attention, in which people must monitor the environment for a relatively infrequent signal. In general, older adults are not impaired on vigilance tasks.

2.2.4 Attention: Summary and implications

Older adults show significant impairments on attentional tasks that require dividing or switching of attention among multiple inputs or tasks. They show relative preservation of performance on tasks that require selection of relevant stimuli; and although they are slower than young adults, they are not differentially impaired by distraction. They also are able to maintain concentration for an extended period of time. The tasks on which older adults show impairments tend to be those that require flexible control of attention, a cognitive function associated with the frontal lobes. Importantly, these types of tasks appear to be amenable to training and show benefits of cardiovascular fitness.

Attentional deficits can have a significant impact on an older person's ability to function adequately and independently in everyday life. Being able to measure these deficits will be needed in order to measure cognitive function changes.

2.3 Memory

The cognitive domain that has probably received the most attention in normal ageing is memory [75, 76]. Many older adults complain of increased memory lapses as they age, and a major focus of research has been to try to distinguish memory declines attributable to normal ageing from those that are indicative of pathological ageing, particularly Alzheimer's disease. As with attention, memory is not a unitary construct, as some forms of memory remain relatively intact with age while others show significant decline. There are many forms of memory and all have been studied and reported in the literature with respect to their impact on cognition.

2.3.1 Working memory

Working memory is a multifaceted aspect of cognitive function that has been hypothesized as the fundamental source of age-related deficits in a variety of cognitive tasks, including long-term memory, language, problem solving, and decision-making. In fact, the majority of theories in the literature of cognitive ageing seem to implicate working memory.

Although there are several models of working memory, all agree that it is a limited capacity system that involves the active manipulation of information that is currently being maintained in attention [75, 77, 78]. Short-term or primary memory, on the other hand, involves the simple maintenance of information over a short period of time. For example, one may maintain a phone number in short-term memory by simple rehearsal of the number. Older adults show minimal or no deficits in short-term memory and can typically hold about 7 ± 2 digits in mind as long as the digits are being rehearsed. Repeating the numbers backwards, however, requires an active reorganization or manipulation of the information held in short-term memory. This task thus requires working memory and this facet of memory shows impairments with age. Some researchers consider working memory a divided attention task, with the contents of short-term memory must be maintained while simultaneously being manipulated or processed for some other purpose. Given the previously discussed findings of divided attention deficits with increased age, it is not surprising that older adults are impaired in working memory.

In the original working memory model of Baddeley and Hitch [79], the manipulation of information in short-term memory was handled by a central executive, and deficits in working memory were viewed as deficits in executive control, a function attributed primarily to prefrontal cortex. Recent neuroimaging research [80] has confirmed a role for dorsolateral prefrontal cortex (PFC) in the manipulation and updating of information in working memory, with left PFC involved more in verbal tasks and right PFC in visuospatial tasks. In recent years, however, the role of the central executive has been expanded to cover a range of executive control functions other than those associated strictly with working memory.

Although there is a general consensus that working memory is impaired in older adults, there is disagreement concerning the mechanisms involved, and much of the research has focused on testing a variety of theories. Three theories of cognitive ageing have been articulated within the context of working memory deficits, although they may apply more broadly across other cognitive domains: (1) one theory proposes a reduction of attentional resources, (2) one focuses on reduced speed of information processing, and (3) one attributes problems to a failure of inhibitory control [81].

2.3.2 Attentional resources

Theories of age-related decline in working memory generally assume some reduction in processing resources. Craik et al. [82, 83] have suggested that the resource limitation is

attentional and reflects a reduction in mental energy. Tasks with high attentional demands show impairments, whereas tasks requiring little or no attention, that are relatively automatic, are largely intact. Working memory tasks by their very nature involve divided attention and are therefore more likely to strain the limited resources of older adults. This theory is intuitively appealing, but has been criticized for being more descriptive than explanatory. The construct of attentional resources is poorly defined; and although neurophysiological correlates such as arousal, measured by electrodermal activity, or neural activity, measured by electroencephalography have been suggested [67], they have not been consistently demonstrated.

2.3.3 Speed of information processing

Salthouse [84] has suggested that speed of processing may be considered a resource, and that age-related deficits in working memory and other cognitive tasks can be explained in terms of a general slowing of information processing. There is little disagreement that older adults are slower than younger adults and that slowing of fundamental cognitive processes may have detrimental effects on more complex tasks. The literature is divided on whether a generalized slowing can account for the empirical findings or if more process-specific components are also needed. Salthouse [85, 86] has demonstrated in numerous studies that slowing of information processing can account for a large proportion of the age-related variance in a variety of cognitive tasks, including working and long-term memory, and has suggested that speed of processing is a "cognitive primitive". Other investigators [87] have suggested that speed of processing and working memory provide independent contributions to higher-level cognition and that working memory deficits must therefore be accounted for in terms of something other than just speed of processing. Finally, slowed processing, like attentional resources, is more a descriptor of ageing cognition than an explanation for cognitive deficits. There is little in the literature on what causes slowing with age. Here too, discovery of neurophysiological correlates are required to provide further understanding of the underlying mechanisms.

2.3.4 Inhibitory control

Hasher, Zacks, and May [88, 89] proposed that a lack of inhibitory control may be responsible for cognitive deficits associated with ageing. Specifically, failure to suppress irrelevant information in working memory may effectively reduce its capacity, denying access to relevant information. For example, working memory tasks often involve the successive

presentation across a series of trials of increasingly long list of words or long strings of digits. Age deficits may be caused by the failure to delete from working memory digits or words from previous trials, thus reducing the "working space" for new stimuli [90]. Although considerable data suggest that older adults experience more interference from irrelevant information under some conditions [91], studies in the literature are mixed and other data fail to support an inhibitory deficit account [67]. It may be that there are a variety of inhibition age-related effects and that some may be task or paradigm specific.

2.3.5 Working memory: Summary and implications

Older adults exhibit significant deficits in tasks that involve active manipulation, reorganization, or integration of the contents of working memory. Although the mechanisms underlying these age-related deficits are as yet poorly understood, the effects of such deficits are very likely far-reaching. Many complex everyday tasks such as decision-making, problem-solving, and the planning of goal-directed behaviours require the integration and reorganization of information from a variety of sources. It seems probable that attention, speed of information processing, and the ability to inhibit irrelevant information are all important functions for effective performance of these higher-level cognitive tasks. The brain regions that are active during working memory tasks are also beginning to be identified in a variety of functional neuroimaging studies that are reported in the literature. Results suggest that different areas are activated in young and old adults, particularly within the prefrontal cortex, indicating that younger and older adults are performing these tasks differently [78]. An understanding of age-related neurophysiological changes may help to account for these differences.

2.3.6 Long-term memory

Long-term memory, unlike short-term and working memory, requires retrieval of information that is no longer present or being maintained in an active state. This information may have occurred a few minutes ago or been acquired many years ago.

2.3.7 Episodic memory

Episodic memory refers to memory for personally experienced events that occurred in a particular place and at a particular time. This kind of memory allows one to think back through subjective time [92] and it usually evokes an "I remember" response. Episodic memory may be distinctly human; it is the most advanced form of memory and is also genetically the latest to

develop. It also seems the most susceptible to brain damage and the most affected by normal ageing.

The episodic memory problems experienced by older adults may involve deficient encoding, storage, or retrieval processes. At the input stage, older adults may encode new information less meaningfully or with less elaboration, so that memories are less distinctive, more similar to others in the memory system, and thereby more difficult to retrieve [93]. Alternatively, older adults may attend to specific salient information but fail to take account of peripheral detail, or they may fail to integrate contextual aspects of an experience with central content. This is referred to as a source memory problem [94]. Many of the common everyday memory lapses reported by normal older adults, such as forgetting the location of keys or glasses, likely involve poor encoding. These kinds of memory failures have generally been attributed to reduced use of effortful encoding strategies, which depend particularly on prefrontal brain regions. Another possibility is that noticing and integrating the various aspects of an experience involve divided attention and require working memory.

Older adults may also experience problems at the level of storage or consolidation. This aspect of episodic memory critically depends on the medial temporal lobe, particularly the hippocampus. Consolidation is thought to involve the binding of various aspects of experience into a composite memory trace. What may be particularly important for episodic memory and impaired in older adults is the extent to which an event is bound to its spatial and temporal context.

Finally, considerable evidence points to retrieval as a source of episodic memory problems in ageing. Although it is clear that retrieval is at least partly dependent on encoding, with well-encoded information easier to retrieve, there are also effortful retrieval processes that appear to be impaired by ageing. Older adults tend to show deficits on tests of free recall, to a somewhat lesser degree in cued recall, but minimally in recognition memory. Craik [83] suggests that the requirement to self-initiate strategic search processes in memory recall overloads the limited resources of older adults. If environmental support can be provided at retrieval as well as at encoding, through the use of good multisensory cues, the resource demands of encoding and retrieval are reduced and age differences are minimal. Similarly, Jennings and Jacoby [95] have demonstrated that recollection, which requires effortful retrieval of episodic detail, is impaired with age, whereas the more automatic judgments of familiarity are intact. Evidence from functional neuroimaging and neuropsychological studies

suggests that these more strategic retrieval processes depend on the prefrontal cortex, as well as the hippocampus [96, 97].

2.3.8 Semantic memory

Semantic memory refers to the ability to store of general knowledge about the world, including factual information and knowledge of words and concepts. Such information is not tied to the space or time of learning, and its retrieval is generally prefaced with "I know." Normally ageing older adults do not have significant impairments in semantic memory. In fact, their knowledge of the world often exceeds that of young people. In addition, although access to information may be somewhat slower (particularly for words and names), the organization of the knowledge system has been reported to be unchanged with age [98]. Semantic memories are believed to be stored within regions in the posterior neocortex.

2.3.9 Autobiographical memory

Autobiographical memory involves memory for a personal past and includes memories that are both episodic and semantic in nature. Evidence suggests that recent memories are easiest to retrieve, while those from early childhood are most difficult to retrieve. It has been suggested that there is a monotonic decrease in retention from the present to the most remote past. This general pattern holds across all ages, suggesting that autobiographical memory is largely preserved with age [99]. More detailed analyses of the nature of the autobiographical information retrieved has suggested that although memory for personal semantics is intact in old age, memory for specific episodic or contextual details about a subject's personal past may be impaired. In a recent study, Levine et al. [100] observed that although older adults reported the overall story of autobiographical event memories as well as young people, they reported fewer details. There may be exceptions to this finding, however. Recent studies of memory have demonstrated that older adults remember as much as young adults about the details and circumstances surrounding highly emotional public events [101, 102].

2.3.10 Procedural memory

Procedural memory refers to knowledge of skills and procedures such as riding a bicycle, playing the piano, or reading a book. These highly skilled activities are acquired more slowly than episodic memories through extensive practice. Once acquired, procedural memories are expressed rather automatically in performance and are not amenable to description (i.e., it is not easy to say "how" one reads). When talking about procedural memory or knowledge, one

is likely to say, "I know how to." In general, older adults show normal acquisition of procedural skills in both motor and cognitive domains and retain them across the lifespan. With high levels of expertise, there is often little slowing of skilled performance with age (at least until the very oldest ages), although some individual components of the skill may decline. So, for example, although the finger movements of a skilled typist slow down with age, overall typing speed is maintained because other aspects of the skill adjust (e.g., scanning further ahead in the text to be typed) [103]. Procedural memory depends on several brain regions, but specifically on the basal ganglia and the cerebellum.

2.3.11 Implicit memory

Implicit memory refers to a change in behaviour that occurs as a result of prior experience, although one has no conscious or explicit recollection of that prior experience. For example, laboratory experiments have shown that it is easier to identify a degraded stimulus (e.g., from a brief exposure or partial information) if the stimulus was seen previously, even if one does not remember the prior occurrence. This "priming" probably occurs in everyday life and appears relatively intact in normal ageing, although there are some inconsistencies in the literature [104]. Implicit memory has been associated with left frontal and left temporal cortical regions.

2.3.12 Prospective memory

Much of what we have to remember in everyday life involves prospective memory, and involves remembering to do things in the future, such as keep appointments or pay bills on time [105]. Older adults do well on these daily tasks, using a variety of external aids such as calendars and diaries to remind themselves of these tasks and activities. However, certain habitual tasks such as taking medications at the appropriate times each day, may create difficulties for older adults. For these tasks, there often are no salient reminders or cues in the environment, and so the tasks require self-initiated activities that may to be problematic for older adults. Prospective memory may also rely on some aspect of working memory to maintain future intentions over time and likely also involves divided attention, both functions that show age-related deficits. Prospective memory and episodic memory tend not to be correlated and probably depend on different regions of the prefrontal cortex.

2.3.13 Long-term memory: Summary and implications

Ageing principally affects episodic memory, namely memory for specific events or experiences that occurred in the past. Although many older adults believe that their memories

for remote events are better than their memories for recent events, it is likely that older memories have become more semantic or blurred, retaining the general core information but lacking details, particularly spatial and temporal context. More problematic for older adults is remembering context or source information: where or when something was heard or read, or even whether something actually happened or was just talked about, what has been called "reality monitoring" [106]. Encoding and retrieval of these kinds of specific or peripheral details about a prior event may be particularly demanding of attentional resources, and good cues for the retrieval of such information may often be lacking. Although semantic memory is largely preserved in old age, the fact that what is retrieved from semantic memory is general knowledge, not specific detail, may contribute to the absence of age differences. The exception to this pattern may be the retrieval of a person's name or a specific word for a specific context, both of which show deficits in normal ageing. The specificity of the information to be retrieved may therefore be a critical determinant of age differences [107]. There is some suggestion that age-related deficits in memory may be reduced for emotionally arousing events or materials [102], and so emotional or personal investment in an experience may be an important variable in episodic memory in older adults. High levels of emotion or stress, however, generally have negative effects on memory.

2.4 Perception

Besides attention and memory, perception is also important in cognitive function. Neuroscience typically defines perception as a set of processes that occurs prior to cognition. However, the boundaries between perception and cognition are unclear, and much evidence suggests that these domains are interactive with top-down cognitive processes affecting perception and perceptual processing having a clear impact on cognition. Evidence indicates that perceptual function is reduced in older adults and is not always correctable by external aids [108]. The impact from this research is that care must be applied to and control for sensory and perceptual deficits when conducting cognitive experiments. Evidence from a range of large-scale ageing studies has demonstrated that a significant proportion of the age-related variance in several cognitive tasks can be accounted for by hearing and vision loss and that once these sensory differences are statistically controlled, there are no longer age differences in cognitive functioning [109]. Baltes and Lindenberger [109] proposed that overall neural degeneration may account for both sensory and cognitive deficits. Alternative explanations have also been proposed. Schneider and Pichora-Fuller [108] suggested that perception and cognition are part of a highly integrated system and draw on a common pool

of attentional resources. When parts of this system are stressed, such as when auditory or visual acuity are compromised and are essential to a task, other parts of the system will be negatively affected.

Declining sensory and perceptual abilities have important implications for the everyday lives of older adults. Hearing loss can isolate older adults, preventing them from engaging in conversation and social interactions. Visual impairments can limit mobility. As older adults develop strategies to compensate for declining sensory abilities, the ways in which they perform other cognitive tasks may also be altered and may be less efficient [65].

2.5 Higher-level cognitive functions

2.5.1 Speech and language

Speech and language processing are largely intact in older adults under normal conditions, although processing time may be somewhat slower than in young adults. In fact, there is evidence that discourse skills actually improve with age. Older adults often tell well-structured elaborate narratives that are judged by others to be more interesting than those told by younger adults [110]. They typically have a more extensive vocabulary. Although they exhibit the occasional word-finding difficulty, older adults are easily able to provide roundabout ways of explaining things to mask the problem. Older adults tend to skilled conversationalists and appear to have few difficulties in processing on-going speech. As noted above, however, some older adults have hearing loss and thus in conversational settings may be required to interpret a weak or distorted acoustic signal. Even under these conditions, older adults are able to maintain good levels of comprehension by effectively using context to interpret the conversation [111]. However, this compensatory top-down processing may have negative consequences for other cognitive operations and may be at least partly responsible for reducing the functional capacity of working memory. The inverse relationship has also been proposed, namely that the well-documented reduced working memory capacity in older adults limits the comprehension of syntactically complex text. The fact that comprehension of text is often measured by recall, a cognitive function known to be impaired in ageing, make the interpretation of comprehension deficits difficult. Older adults also experience problems with comprehension when individual words are presented at a very rapid rate, but they show sharply reduced impairments when such words form meaningful sentences. Older adults seem able to engage intact top-down processes to bolster deficiencies in bottom-up processing. They thus appear to retain good language skills well into older age. Deficits that occur under

difficult processing conditions seem primarily attributable to sensory loss or working memory limitations, not to impairments in basic language capacities itself [112].

2.5.2 Decision making

Relatively little research has been carried out on the effects of ageing on decision-making. Research in the literature has highlighted the potential impact of attentional and working memory limitations on the ability to make decisions, but also has incorporated ideas involving motivation, relevance, emotional investment, and prior knowledge as important moderators of those effects, particularly in real-life contexts. Decision-making seems to be a domain that makes clear demands on processing resources, but in everyday life those demands may be reduced by life-relevant knowledge or expertise in the problem-solving domain [113]. Older adults, possibly due to working memory limitations, tend to rely on expert opinion to a greater degree than young adults.

2.5.3 Executive control

In the past decade, there has been an increasing focus on executive control as a primary contributor to cognitive decline with age. Executive control is a multi-component construct that consists of a range of different processes that are involved in the planning, organization, coordination, implementation, and evaluation of non-routine activities. This "central executive" plays a key role in virtually all aspects of cognition, allocating attentional resources among stimuli or tasks, inhibiting distracting or irrelevant information in working memory, formulating strategies for encoding and retrieval, and directing all problem-solving, decision-making, and other goal-directed activities [79, 114]. Executive control is particularly important for carrying out new tasks for which a set of habitual processes is not readily available. Executive function depends critically on prefrontal cortex, which exerts its broad-reaching controlling influence via extensive reciprocal connections with posterior cortical regions. An overall explanation of cognitive ageing implies a causal role to executive control deficits, what is known in the literature as the frontal lobe hypothesis of ageing [115]. In support of this hypothesis, both structural and functional neuroimaging studies have revealed a preferential decline in older adults in volume and function of prefrontal brain regions [116].

2.6 Inter-individual variability in cognitive function

Although there are clear generalities and common principles that can be demonstrated in cognitive ageing, what is perhaps most compelling about age-related cognitive change is its variability. Cognitive decline is not inevitable. Some older adults retain excellent cognitive function well into their 70s and 80s and perform as well or better than younger adults. Others, although within the normal range, show signs of decline by age 60. In addition, decline is not uniform across cognitive domains. For example, some older adults have excellent episodic memory function but impaired executive function, and vice versa [117]. So, although there are clear interactions among cognitive domains, it seems evident that they also have some degree of independence and may be more or less susceptible to ageing in different individuals. From the literature it is unclear what accounts for this variability.

Inter-individual variability is most likely due to a range of factors and mechanisms: biological, psychological, health-related, environmental and lifestyle. One possibility is that variability is related to differential internal compensatory mechanisms. A number of recent functional neuroimaging studies have found different patterns of brain activation in older and younger adults while performing identical memory or working memory tasks. One such pattern involves greater bilateral activation in older adults for tasks that activate only unilateral brain regions in young adults [118, 119]. This increased activation has been observed particularly in a sub-group of high-functioning older adults [119], and has been interpreted by many as compensatory activity, representing perhaps some reorganization of the ageing brain. Others have suggested, however, that bilateral activation represents inefficient or less selective cognitive processing in older adults [120]. Another possibility is that such changes relate to declining sensory and perceptual abilities [108], which older adults compensate for in different ways [121].

Lifestyle variables have also been the focus of much recent research on factors related to differential cognitive ageing. Active lifestyles are generally associated with better outcomes, and aerobic exercise in particular has been shown to produce substantial benefits to cognitive function, particularly on those tasks requiring executive control [74]. Performance on these same kinds of non-automatic tasks is also particularly sensitive to circadian rhythms. For example, older adults perform better at their peak time of day, usually in the morning, on tasks requiring inhibitory control [89]. Interestingly, stimulants such as caffeine have been found to reduce the time-of-day effects on strategic memory tasks, by enhancing performance during non-peak times of day [122].

2.7 Cognitive function: Summary

Basic research in cognitive ageing has focused on attention and memory, and indeed it may be that deficits in these fundamental processes can account for much of the variance observed in higher-level cognitive processes. The mapping of cognitive processes onto neural structures constitutes a relatively recent research enterprise driven largely by advances in neuroimaging technology. Age-related changes in cognitive function vary considerably across individuals and across cognitive domains, with some domains of cognitive function appearing more susceptible than others to the effects of ageing. Research has focused on establishing brain regions associated with different kinds of cognitive performance and revealed that normally ageing older adults often appear to activate different brain structures than young people when performing cognitive tasks. The reasons for these differences are continually reported in the literature. Ultimately, the understanding of age-related changes in cognition will require a parallel understanding of the age-related changes in the brain and the underlying mechanisms responsible for those changes [65].

2.8 Neurodegenerative conditions affecting cognitive function

Besides the normal ageing process, older adults can be affected by abnormal ageing process. This abnormality effecting cognitive function is mostly due to neurodegenerative processes occurring in the brains of older adults.

2.8.1 Dementia

Dementia, and in particular Alzheimer's disease (AD), is a major concern for those individuals who live beyond the age of 70 years. While dementia is by no means a natural consequence of ageing, both its incidence and its prevalence increase dramatically with age [123, 124]. Consequently, a person aged 85 years has a one in three chance of having significant cognitive impairment. Ultimately, as therapies progress, early identification may lead to prevention of age related cognitive disabilities. All of this speaks to the importance of trying to identify impairment of cognitive function at its earliest stage.

2.8.2 Mild cognitive impairment and Alzheimer's disease

Mild cognitive impairment (MCI) precedes the onset of Alzheimer's disease. Individuals with MCI experience subtle cognitive deficits with largely intact cognition and activities of daily living. The most frequent presentation of this clinical condition is forgetfulness. Subtle memory impairment is the most common initial complaint of individuals with MCI [42].

There is a consensus that MCI is an important topic of study but disagreement as to the uniformity of its definition. In multiple-domain form of MCI, subjects may have slight memory impairment in conjunction with mild impairments in executive function and language. The diagnosis of multiple-domain MCI is a clinical judgment on the part of the clinician evaluating the subject and cannot be made solely on the basis of neuropsychological testing.

A form of MCI, known as single non-memory-domain MCI is characterized by the subject having a relatively isolated impairment in a single non-memory domain such as executive function, visuospatial processing, or language.

Typically, subjects with a very mild degree of impairment such as found in MCI often do not get referred to a dementia clinic for an evaluation. The subtle deficits that are found in these subjects are often overlooked by the subjects themselves, their families, and their examining physicians [42].

A study by Petersen et al. [125] demonstrated that subjects, who have fulfilled the clinical criteria for MCI, tend to progress to dementia or clinically probable AD at a rate of approximately 12% per year. While the normal control cohort developed MCI/AD at a rate of 1% to 2% per year. Approximately after six years, 80% of the MCI cohort had declined to dementia. This highlights the importance of early detection of cognitive decline.

There may be some individuals who have had a lifelong tendency to be "slow learners" and on entering older age appear to meet criteria for MCI but, in fact, are not progressing to dementia [42]. This illustrates the need for tasks of cognitive function to be highly specific to age related cognitive decline.

MCI is a useful concept because of its ability to predict AD in a rather specific fashion, but it lacks clear definition, particularly when the diagnosis is based on psychometric tests that themselves can lack objective scoring protocols [42]. These tests are often only one element which when combined with activity of daily living scales form the judgement of the clinician [42].

Clinical criteria for AD have been published and used worldwide. For example, the Diagnostic and Statistical Manual IV (DSM IV) [6] and National Institute for Neurological Communicative Disorders and Stroke / Alzheimer's Disease and Related Disorders (NINCDS / ADRDA) Diagnostic and Statistical Manual of Mental Disorders, 4th ed. [126]. These have been very useful, but they are designed to detect the disease in the reasonably well developed state. The criteria require impairments in memory and other cognitive domains of sufficient

severity to affect functional activities. There is a need for more objective methods of cognitive function assessment.

2.9 Standard neuropsychological assessment of cognitive function

There are a number of neuropsychological tests available to the clinician to assess cognitive function. Some of these have been extensively validated, others have been validated to a lesser extent.

2.9.1 Assessment methods for cognitive function

The tests described in this section are employed to assess state of overall cognitive function in older adults.

The Mini-Mental State Examination

The Mini-Mental State Examination (MMSE) [30] is often employed to screen for mental impairment, particularly in the older adults. The MMSE consist of a variety of questions, has a maximum score of 30 points, and can be administered in 5-10 minutes. It assesses orientation to time and place, attention, memory and ability to follow commands. The MMSE provides a quick quantitative assessment of an individual's cognitive state. Although the MMSE was designed for use with hospitalized patients [30], the scale has attained widespread use among clinicians and researchers concerned with primary care [127-129] and community settings [130]. The MMSE has also been adapted for use in paediatric settings [131].

The MMSE assesses orientation to time and place, registration, memory, attention and concentration, praxis, and constructional and language capacity. Orientation to time is tested by asking the subject the season, date, day of the week and month. Orientation to place is tested by asking the subject in what state, county, city or town, building and floor, and address they are in. Immediate registration is tested by asking the subject to repeat a word list of three common objects. Concentration or attention and calculation are tested by asking the subject to subtract seven from one hundred, and seven from the result, and so on for five computations ('Serial sevens' task). Alternatively, the subject is asked to spell the word 'world' backwards. Delayed recall is tested by asking the subject to recall the word list of three objects named in the immediate registration task. Language is assessed by asking the subject to name two common objects, repeat a phrase ('no ifs, ands or buts'), follow a three step command, read and obey a single command, write a sentence and copy a drawing of a two pentagons design [30].

MINI MENTAL STATE EXAMINATION (MMSE)

Patient's name:

Hospital number:

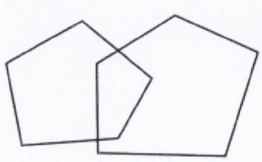
ONE POINT FOR EACH ANSWER	DATE			
ORIENTATION				
Year Month Day Date Time	____/5	____/5	____/5	____/5
Country Town District Hospital Ward	____/5	____/5	____/5	____/5
REGISTRATION				
Examiner names 3 objects (eg apple, table, penny) Patient asked to repeat (1 point for each correct). THEN patient to learn the 3 names repeating until correct.	____/3	____/3	____/3	____/3
ATTENTION AND CALCULATION				
Subtract 7 from 100, then repeat from result. Continue 5 times: 100 93 86 79 65 Alternative: spell "WORLD" backwards - dlrow.	____/5	____/5	____/5	____/5
RECALL				
Ask for names of 3 objects learned earlier.	____/3	____/3	____/3	____/3
LANGUAGE				
Name a pencil and watch.	____/2	____/2	____/2	____/2
Repeat "No ifs, ands, or buts".	____/1	____/1	____/1	____/1
Give a 3 stage command. Score 1 for each stage. Eg. "Place index finger of right hand on your nose and then on your left ear".	____/3	____/3	____/3	____/3
Ask patient to read and obey a written command on a piece of paper stating "Close your eyes".	____/1	____/1	____/1	____/1
Ask the patient to write a sentence. Score if it is sensible and has a subject and a verb.	____/1	____/1	____/1	____/1
COPYING				
Ask the patient to copy a pair of intersecting pentagons: 	____/1	____/1	____/1	____/1
TOTAL	____/30	____/30	____/30	____/30

Figure 2-2: *Mini-Mental State Examination [30]*

The score is the total number of correct answers. The maximum score for the MMSE is 30 points. The correct interpretation of the results is critical and can only be based on the correct administration of the test. This includes asking the question clearly and unambiguously. There is considerable anecdotal evidence that delivery is poor and that underestimates the level of cognitive impairment in the subject being tested [35].

The Montreal Cognitive Assessment

The Montreal Cognitive Assessment (MoCA) is a brief screening tool for mild cognitive impairment (MCI) [31]. MoCA, like MMSE, provides general overview of cognition. It assesses different cognitive domains: attention and concentration, executive functions, memory, language, visuo-constructional skills, conceptual thinking, calculations, and orientation. Time to administer the MoCA is approximately 10 minutes. The total possible score is 30 points; a score of 26 or above is typically considered normal [31].

MoCA tests visuo-spatial and executive function by asking the subject to complete an alternating trail making task and by copying a drawing of cube. Visuo-constructional skills are tested with a task requiring the subject to draw a clock showing the time "ten past eleven". The subject is asked to name three animals. Immediate memory is tested by asking the subject to repeat five words in two trials. Attention is tested by administration of Forward and Backward Digit Span, Vigilance and serial sevens tasks. During Forward Digit Span task the subject is asked to repeat five numbers in the same order as they were presented to the subject. The subject is requested to repeat in reverse order the set of three numbers during the Backward Digit Span task. Language is assessed by asking the subject to repeat two sentences articulated by the examiner. For the verbal fluency task, the subject is instructed to name as many words as he/she can think of that begin with letter F during 60 seconds. Abstraction is assessed by presenting the subject with three pairs of words and asking what these words have in common. Delayed memory is tested by administration of delayed recall of the five words used during test of immediate memory. If the subject cannot recall any of the five words, a category cue is presented. If the subject still can't remember, a multiple choice cue is given. Orientation to time is tested by asking the subject to state the current date, month and year. Orientation to place is tested by asking the subject in what place and city he/she is currently located.

The total score is calculated as a sum of the sub-scores from all tasks. One point is added for individuals that have 12 years or fewer of formal education. The possible maximum is 30

points. The MOCA is becoming more popular as an assessment of cognitive function [132, 133].

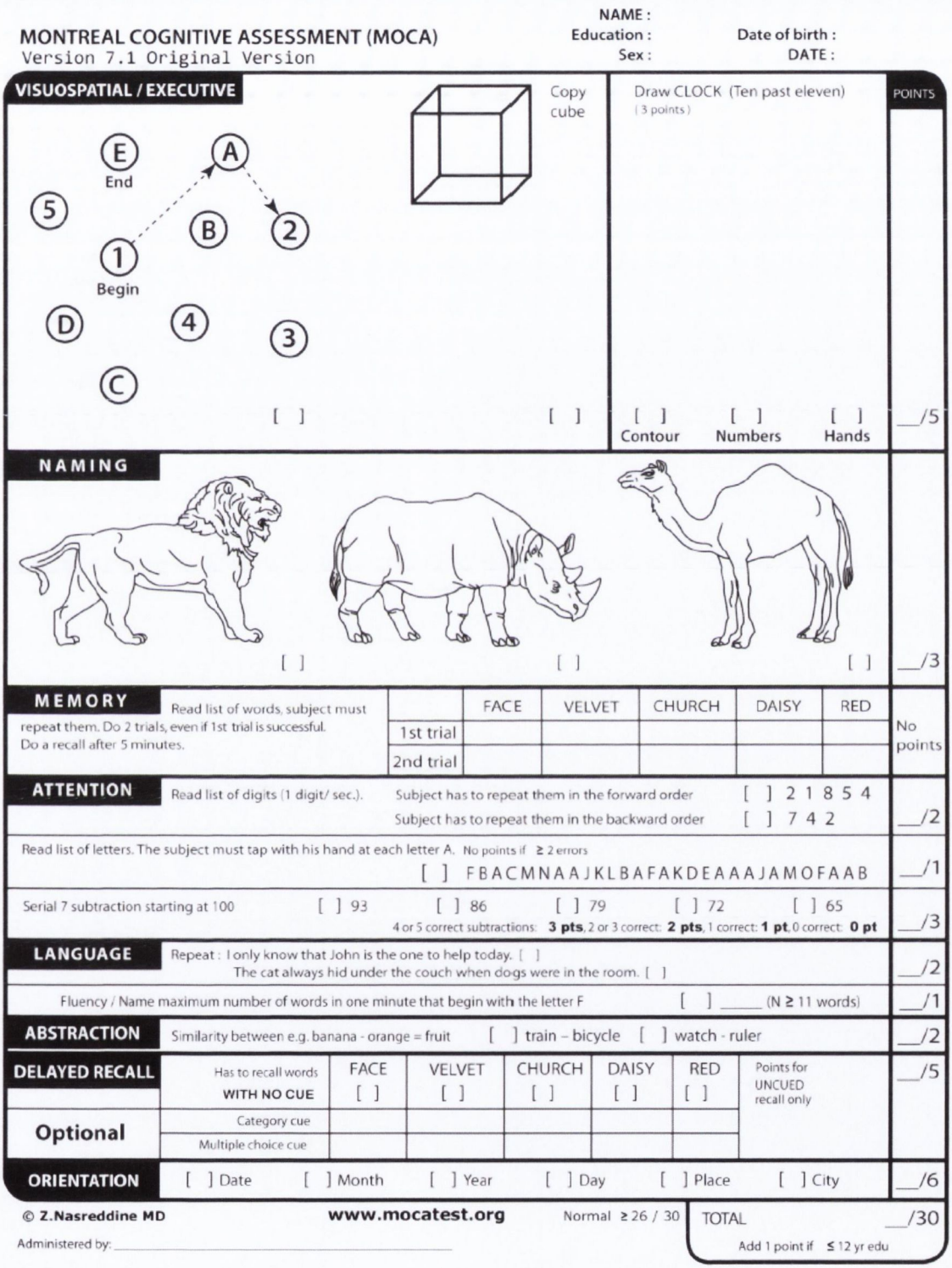


***Figure 2-3**: The Montreal Cognitive Assessment (MOCA) [31]*

National Adult Reading Test

The National Adult Reading Test (NART) [134] is a test of premorbid intellectual ability. The NART is an untimed measure and a reading test of 50 irregularly spelled words (e.g., ache, naive, thyme). It has promise as an assessment tool for the determination of premorbid intellectual function. Assuming that the patient is familiar with the word, accuracy of pronunciation is employed to predict IQ. As the words are short, patients are not required to analyse a complex visual stimulus, and as they are irregular, phonological decoding or intelligent guesswork will not provide the correct pronunciation [68]. Each word is presented individually and subjects are required to read each aloud.

The use of a pronunciation guide and a tape recorder is recommended to facilitate scoring. Each incorrectly pronounced word counts as one error. Slight variations in pronunciation are acceptable when they are due to regional accents. The total number of errors is tabulated. This is a subjective scoring of this test and using the recording of the subject's voice it is important to have a number of raters score the subject to remove any examiner bias. The use of non-native speakers as both subjects and examiners also raises concerns of the objectiveness of this test.

2.9.2 Assessment of executive function

As mentioned above in Section 2.5.3, executive function is vital element of cognitive function. The tests described below are employed in assessment of executive function.

Stroop task

The Stroop task assesses the ease with which a person can maintain a goal in mind and suppress a habitual response in favour of a less familiar one. This measure of selective attention and cognitive flexibility was originally developed by Stroop in 1935 [68]. The Stroop paradigm is one of the oldest and most widely used methods to assess attention [135].

There are four parts to the test. In Part 1, the subject reads randomized colour names (blue, green, red, brown, purple) printed in black type. In Part 2, the subject reads the colour names (blue, green, red, brown, purple) printed in coloured ink (blue, green, red, yellow), ignoring the colour of the print (the print colour never corresponds to the colour name), see Figure 2-1. In Part 3, the subject has to name the colour of squares (blue, green, red, brown, purple). In Part 4, the subject is given the card used in Part 2, but this time, the subject must name the colour in which the colour names are printed and disregard their verbal content.

Of major interest is the subject's behaviour when presented with coloured words printed in nonmatching coloured inks. Stroop reported that subjects with normal executive function can read colour words printed in coloured ink as fast as when the words are presented in black ink (Part 2 versus Part 1). However, the time to complete the task increases significantly when the subject is asked to name the colour of the ink rather than read the word (Part 4 versus Part 3). This decrease in colour-naming speed is called the "colour-word interference effect."

For each part, the examiner records both the time to complete and the number of errors. Spontaneous corrections are scored as correct. Researchers have typically relied on a difference score, defined as the difference in the amount of time required for the interference card (e.g., Part C) versus the colour card (e.g., Part D). Graf et al. [136] contend that a difference score is not independent of age-related slowing and recommend the use of a ratio index of interference (e.g., Card C/Part D). [68]

Category Fluency

Category Fluency [68] is a test employed to assess verbal fluency as another element of executive function. This test evaluates the spontaneous production of words under restricted search conditions (verbal association fluency). This task requires the time limited production of exemplars of a specific semantic category, e.g. animals, fruits, vegetables. The time limit is typically 60 seconds.

The score is the total of all admissible words for the semantic category. For animal category fluency, names of extinct, imaginary, or magic animals are admissible, but given names for animals like "Fido" and "Morris" are not. Inadmissible words under these instructions (e.g., proper names, wrong words, variations, repetitions) are not counted as correct [68]. This test is becoming more popular with the neuroscientists [137, 138].

2.9.3 Assessment of memory

As mentioned in Section 2.3, memory is a key element of cognitive function. The following tests are employed in assessment of memory in older adults.

Wechsler Memory Scale

The Wechsler Memory Scale (WMS) [32] has been designed to assess auditory and visual declarative memory and auditory and visual working memory abilities in adults and adolescents. This test is now in its third edition (WMS-III) [68].

Word list recall

A subtest of the Wechsler Memory Scale, Word recall immediate (WRI) [32] is a test of short-term memory. The subject is required to recall a list of ten words presented orally by the neuropsychologist. The list is repeated three times and the total number of the correct answers from the three iterations constitutes the score for this task.

Word recall delayed (WRD) [32], a subtest of the Wechsler Memory Scale, is a test of medium to long-term memory. Twenty minutes after completing the word recall immediate, the subject is asked to recall as many words as possible from the original list. The score for this task is the total number of words correctly recalled.

Digit Span

The Digit Span task [32] is a measure of working memory. Working memory refers to the cognitive system that allows temporary storage and manipulation of information. This task requires the recall of a sequence of digits in reverse order; the sequence is augmented each time the sequence is recalled correctly. The score is the length of the last sequence recalled correctly. The reverse nature of this task makes it more challenging and relies heavily on the manipulation of information.

2.9.4 Assessment of attention

As presented in Section 2.2, attention is another key element of cognitive function. The following tests are used for assessment of attention.

Sustained Attention to Response Task

The Sustained Attention to Response Task (SART) [139] is widely used as a behavioural measure of sustained attention failures. This test involves key presses to frequently visually presented non-targets, but with the requirement to withhold motor responses to occasional targets. The non-targets were represented by eight different shapes and the target by just one shape, the shape of a star. The stimuli are presented one by one in random order with at least two non-targets between targets. The capacity to withhold responses to some but not all instances of the no-go target is interpreted as reflecting lapsing attentional control over the response [140].

The reaction times, number of commission errors (key press for targets) and number of omission errors (missing key press for non-target) is collected for this test.

2.9.5 Assessment of mood, depression and personality

Low mood, loss of interest and diminished capacity for enjoyment, negative outlook with feelings of hopelessness may lead to depression and have an impact on the outcomes of cognitive function assessment [10]. To assess mood of older adults, the following tests are employed.

Instrumental Activities of Daily Living

The Instrumental Activities of Daily Living (IADL) [68] is a caregiver/informant-based rating scale designed to measure functional status. Two major categories of functional abilities are commonly distinguished: (a) activities of daily living, commonly known as ADLs, which focus primarily on overlearned self-care activities such as feeding, bathing, toileting, and basic mobility, and (b) instrumental activities of daily living, commonly known as IADLs, which are viewed as involving fairly complex cognitive abilities and include activities such as managing medication, managing finances, using transportation, using the telephone, maintaining one's household (housekeeping), meal preparation, and nutrition. It is the IADLs that are of primary interest to neuropsychologists, as loss of competence in complex tasks of daily living is a defining diagnostic feature of Alzheimer's disease (AD) and related dementing disorders [68].

From the literature, it can be observed that researchers differ in their scoring of the IADL. For example, according to Lawton and Brody [141], item scores of 1 reflect complete independence in performing the task, and scores of 0 reflect partial or complete dependence. By contrast, others [142, 143] employ a greater range of scores to increase sensitivity, with scores of 0 used to reflect complete independence and higher scores (e.g., 1 or 2) used to reflect partial or complete dependence. [68]

Center for Epidemiologic Studies Depression scale

The Center for Epidemiologic Studies Depression (CES-D) scale [144] is a short self-report scale designed to measure depressive symptomatology in the general population. The items of the scale are symptoms associated with depression which have been used in previously validated longer scales. CES-D scale is commonly used and despite its popularity, several recent investigations have called into question the robustness and suitability of the commonly used 4-factor 20-item CES-D model [145].

Center for Epidemiologic Studies Depression Scale (CES-D), NIMH

Below is a list of the ways you might have felt or behaved. Please tell me how often you have felt this way during the past week.

	During the Past Week			
	Rarely or none of the time (less than 1 day)	Some or a little of the time (1-2 days)	Occasionally or a moderate amount of time (3-4 days)	Most or all of the time (5-7 days)
1. I was bothered by things that usually don't bother me.	☐	☐	☐	☐
2. I did not feel like eating; my appetite was poor.	☐	☐	☐	☐
3. I felt that I could not shake off the blues even with help from my family or friends.	☐	☐	☐	☐
4. I felt I was just as good as other people.	☐	☐	☐	☐
5. I had trouble keeping my mind on what I was doing.	☐	☐	☐	☐
6. I felt depressed.	☐	☐	☐	☐
7. I felt that everything I did was an effort.	☐	☐	☐	☐
8. I felt hopeful about the future.	☐	☐	☐	☐
9. I thought my life had been a failure.	☐	☐	☐	☐
10. I felt fearful.	☐	☐	☐	☐
11. My sleep was restless.	☐	☐	☐	☐
12. I was happy.	☐	☐	☐	☐
13. I talked less than usual.	☐	☐	☐	☐
14. I felt lonely.	☐	☐	☐	☐
15. People were unfriendly.	☐	☐	☐	☐
16. I enjoyed life.	☐	☐	☐	☐
17. I had crying spells.	☐	☐	☐	☐
18. I felt sad.	☐	☐	☐	☐
19. I felt that people dislike me.	☐	☐	☐	☐
20. I could not get "going."	☐	☐	☐	☐

SCORING: zero for answers in the first column, 1 for answers in the second column, 2 for answers in the third column, 3 for answers in the fourth column. The scoring of positive items is reversed. Possible range of scores is zero to 60, with the higher scores indicating the presence of more symptomatology.

***Figure 2-4**: Center for Epidemiologic Studies Depression (CES-D) scale [144]*

Hospital Anxiety and Depression Scale

Hospital Anxiety and Depression Scale (HADS) [146] is an instrument for detecting states of depression and anxiety in the setting of an hospital medical outpatient clinic. The HADS is a fourteen item scale that generates ordinal data. It was divided into an Anxiety subscale (HADS-A) and a Depression subscale (HADS-D) both containing seven intermingled items. The anxiety and depressive subscales are also valid measures of severity of the emotional disorder [146]. HADS has been used extensively by the neuroscientists [147].

	Yes definitely	Yes sometimes	No, not much	No, not at all
1. I wake early and then sleep badly for the rest of the night.	3	2	1	0
2. I get very frightened or have panic feelings for apparently no reason at all.	3	2	1	0
3. I feel miserable and sad.	3	2	1	0
4. I feel anxious when I go out of the house on my own.	3	2	1	0
5. I have lost interest in things.	3	2	1	0
6. I get palpitations, or sensations of 'butterflies' in my stomach or chest.	3	2	1	0
7. I have a good appetite.	0	1	2	3
8. I feel scared or frightened.	3	2	1	0
9. I feel life is not worth living.	3	2	1	0
10. I still enjoy the things I used to.	0	1	2	3
11. I am restless and can't keep still.	3	2	1	0
12. I am more irritable than usual.	3	2	1	0
13. I feel as if I have slowed down.	3	2	1	0
14. Worrying thoughts constantly go through my mind.	3	2	1	0

Anxiety 2, 4, 6, 8, 11, 12, 14

Depression 1, 3, 5, 7, 9, 10, 13

***Figure 2-5**: Hospital Anxiety and Depression Scale (HADS) [146]*

Positive and Negative Affect Schedule

The Positive and Negative Affect Schedule (PANAS) is a 20-item self-report measure of positive and negative affect developed by Watson et al. [148]. The 20 items are words that describe different feelings and emotions. Negative affect (NA) and positive affect (PA) reflect dispositional dimensions, with high-NA epitomized by subjective distress and un-pleasurable

engagement, and low-NA by the absence of these feelings. By contrast, PA represents the extent to which an individual experiences pleasurable engagement with the environment.

1	2	3	4	5
Very Slightly or Not at All	A Little	Moderately	Quite a Bit	Extremely

_____ 1. Interested	_____ 11. Irritable
_____ 2. Distressed	_____ 12. Alert
_____ 3. Excited	_____ 13. Ashamed
_____ 4. Upset	_____ 14. Inspired
_____ 5. Strong	_____ 15. Nervous
_____ 6. Guilty	_____ 16. Determined
_____ 7. Scared	_____ 17. Attentive
_____ 8. Hostile	_____ 18. Jittery
_____ 9. Enthusiastic	_____ 19. Active
_____ 10. Proud	_____ 20. Afraid

***Figure 2-6**: Positive and Negative Affect Schedule (PANAS) [148]*

Each item is rated on a 5-point scale ranging from 1 = *very slightly* or *not at all* to 5 = *extremely* to indicate the extent to which the respondent has felt this way in the indicated time frame.

The scoring of PANAS is performed by calculating the total of the ten items for PA and the total of the ten items for NA. A higher score PA indicates more positive affect, or the extent to which the individual feels enthusiastic, active and alert. A higher NA score indicates more negative affect, or the extent to which the individual feels aversive mood states and general distress.

All the tests listed above are standard cognitive tests incorporated into any comprehensive cognitive assessment battery.

2.10 Limitations of current methods for cognitive function assessment for the older adults

The current cognitive assessment methods, despite being considered gold-standard, have limitations. To interpret correctly what a test or scale is reporting on a subject requires experience in using that test. In particular a number of factors can affect performance and ratings. The numerical result representing the final score cannot be considered as definitive verdict of a subject's state of cognition.

Factors affecting performance of a test may include poor vision, hearing, age, level of education, and gender. Sensory processes decline with age and this means that test stimuli must be designed to be appropriate for the older adults. It must be assured that the test materials can be clearly seen by the subject. The examiner needs to consult with the client to ensure that the instructions are heard and understood. Spectacles, hearing aids, and hearing devices may also have an impact on the outcome of a test.

Individual differences increase dramatically with age, making a wider range in older adulthood of what is normal. Ceiling and floor effects are therefore more likely in older adult groups where the range of ability is so great.

Test anxiety is an issue with regard to cognitive score interpretation, specifically if the test subjects feel that their mental capacities may be declining. They may be unused to being tested. It may have been many decades since they were last tested at anything if ever, as many older individuals have a basic primary level education. The subject may be fearful as to what the tests are going to reveal, and their implications. Sometimes excessive test anxiety rather than any age- or disease-related problem contribute to impaired test performance [149]. Lack of test anxiety may also impact the performance of the subject being assessed. It is important to know who referred the subject for screening and if the cognitive assessment tests are perceived as of any value by the individual being tested. Effort and motivation have to be considered regarding the test performance.

Fatigue can occur sooner during the assessment in older adults than young subjects. Due to decreased stamina, shorter assessment batteries have been recommended for older subjects [150].

Other factors affecting screening test performance include factors such as mood, depression, dysphasia, medication (psychotropic, social), psychosocial stressors, pain, physical

illness, and others [37]. All these factors need to be taken into account by researcher interpreting the cognitive screening.

2.10.1 MMSE variations

When relating to specific cognitive assessment tests there are variations in the performance due to the examiner's and the subject's abilities. The MSSE as discussed above is a gold-standard test for assessment of cognitive function. When delivering the MMSE there can be variations in the wording and content of some questions used by the examiner, as well as in the administration and scoring of the MMSE. This can result in some tests being more difficult than others due to subtle changes in some of the questions.

Since the MMSE was developed to assess hospital patients, the orientation questions of the MMSE questionnaire require the respondent to specify the name and floor where they are located in the hospital. However, alternative items frequently are substituted when the MMSE is administered outside the hospital, particularly in community surveys and epidemiological studies [38, 151].

The choice of words employed to assess a person's ability to learn and retain three words was left to the discretion of the examiner in the original version of the assessment test by Folstein [30]. When the MMSE was incorporated as a part of structured interview into the Diagnostic Interview Schedule[4] [152], the words 'apple', 'penny', and 'table' were employed. These three words are adopted by most subsequent studies and quoted in the literature. Exceptions to this practice have included words such as shirt, brown, honesty, flag, ball, tree, rose, ring, elephant, and dog [128, 153-156].

Folstein et al [30] routinely administered the Serial Sevens task (see Section 2.9.1) on every test. However, patients were permitted to spell the word 'world' backward if they could not or would not perform the Serial Sevens task. While using the 'world' spelling task as an alternative has been followed in many studies, several other procedures have been adopted. Some studies and applications employ only the 'world' task [156-160] while others employ only the Serial Sevens task [154, 161-163]. Others routinely include both tasks, which are scored in one of the following ways:

[4] The Diagnostic Interview Schedule is a structured interview designed to diagnose in a reliable and valid manner the major psychiatric disorders. This schedule was employed in the National Institute of Mental Health - Epidemiologic Catchment Area program that interviewed almost 20000 subjects.

1. The higher of the two scores is used [38, 153, 155, 164-167],
2. The two scores are combined [155, 166, 168, 169], or
3. Each task is analysed separately [128, 170, 171].

Variations also exist in how the 'world' spelling task is scored [30, 156, 160, 170, 172].

Use of just one or both tasks can have wide impact. The use of both tasks over just one task may have an impact on the Delayed Recall task due to time difference between administration of registration and delayed recall.

2.10.2 MMSE reliability

In cases where there are repeated administrations of the MMSE, several studies reported an increase of total scores upon retest. Presumably due to a practice effect [30, 38, 165, 173, 174]. Moreover, evidence exists suggesting that some participants "study" for cognitive assessment tests by rehearsing answers given on a previous administration of the MMSE [175]. The score of orientation to time and place tasks may be influenced by a subject's rehearsal of these tasks before the administration of the test. From another perspective, if the subject does not remember the exact name of the hospital/clinic in which they are located, it does not imply the subject does not know where they are.

In two studies reporting test-retest reliability estimates [36, 176], correlation coefficients were less than 0.50. Olin et al [36] attribute the decrease reliability to several psychometric problems, including regression to the mean, method for assessing attention/concentration, and lack of explicit scoring criteria for the 'copy two pentagons' task. The finding reported by O'Connor et al [38] that the second lowest kappa value for interrater reliability occurred for the 'copy two pentagons' task further suggest that the lack of scoring criteria may affect the stability scores. Regardless of their cause, the relatively low reliability coefficients that occurred for healthy subjects have important implications for using the MMSE with longitudinal assessment, suggesting that small changes in scores should be interpreted with caution.

The site of testing can also influence test scores [171]. Several studies reported that participants tested at home achieved significantly higher scores than when they were tested in a clinic [177, 178].

2.10.3 MMSE validity

The most frequently cited shortcoming of the MMSE relates to its lack of sensitivity to mild cognitive impairment and its failure to adequately discriminate patients with mild Alzheimer's disease from normal patients [35]. The MMSE has received criticism because of its insensitivity to progressive changes occurring with severe Alzheimer's disease. Inconsistencies in the way MMSE is administered, scored and interpreted make cross-study comparison difficult. The content of the MMSE is highly verbal, lacking sufficient items to adequately measure visuospatial and constructional praxis. Items designed to measure language functions also tend to be overly simplistic and tend to be insensitive to mild linguistic deficits, hence increasing the number of false negative errors [35].

2.10.4 Influence of social and demographic variables on MMSE score

Other factors that lead to variability in final scores of cognitive assessment test are education, age, race, ethnicity and social class.

Education

MMSE scores have been repeatedly shown to be related to educational attainment. The importance of education was revealed by regression analyses, showing that education accounted for more variance than other demographic variables, including gender, race, and social class [164, 179-181]. Education introduces a psychometric bias leading to misclassification of individuals from different educational backgrounds [35]. Low educational levels increase the likelihood of misclassifying normal subjects as cognitively impaired, in particularly in subjects with less than nine years of education [128, 182]. Fillenbaum et al [179] speculated that higher education levels may mask MCI and O'Connor et al [38] reported that all dementia patients with an MMSE score of 24 or greater had relatively high levels of education.

Age

Numerous studies have shown that MMSE scores decreases as age increases [38, 157, 164, 168, 182, 183]. Most of this age-related change begins about age 55 or 60 years and then dramatically accelerates over the age of 75 or 80 years [130, 184]. These age effect persist when subjects are stratified by education level [130, 151, 167, 184], demonstrating that the age effect is not simply due to cohort differences in educational attainment [35].

Crum et al [185] analysed MMSE scores of over 18,000 cognitively healthy adults in terms of age and education. This study established median scores for healthy adults at different ages and levels of education. For example, this study proposed that 60-70 year olds with less than 4 years of education had a median MMSE score of 22, while the same age group with more than 12 years of education was found to have a median MMSE score of 29.

Race/Ethnicity and social class

Contradictory findings were reported when studying the effects of race/ethnicity on the MMSE performance. Escobar et al [170] showed, on a sample of 3000 English and Hispanic residents of Los Angeles, that Hispanics performed significantly lower on many of the MMSE tasks. A study by George et al [130] indicates that race/ethnicity exerts a significant effect on the distribution of MMSE scores. Murden et al [128] reported that race does not affect the MMSE performance. Effects of social class and socioeconomic status on MMSE scores have been observed [38, 168, 180].

2.10.5 Cut-off scores

One cited disadvantage of the MMSE is the lack of agreement on the threshold for cognitive impairment on the 30-point scale [186]. An MMSE score of 24 is often cited in the literature as the threshold for dementia [35], while other studies have used 27 as the threshold for dementia [187]. Lerner and McKee rated cognitive function of subjects using the Clinical Dementia Rating (CDR) scale and compared the reliability of using different MMSE thresholds [188] for determining appropriate cut-offs for healthy cognition on the MMSE scale, and to identify normal subjects who may be at risk for cognitive decline. The authors found a threshold of 27 resulted in a more balanced trade-off between specificity and sensitivity than a threshold of 24. Lerner and McKee concluded that although an MMSE score of 24 is certainly the cut off for dementia, an MMSE score of less than 27 suggests a high risk of cognitive decline.

MCI rating scales

A source of confusion in the literature with respect to comparison of studies on MCI concerns the use of rating scales. Some studies define MCI using specific clinical criteria [125]. Other studies equate cognitive stages on rating scales with MCI. Popular rating scale for disease severity is the Clinical Dementia Rating (CDR) scale [189]. With this instrument, the summary score of CDR 0.5 has been believed to represent an intermediate stage of cognitive impairment or very mild dementia [42]. But a controversy exists as to whether all CDR 0.5

subjects have dementia or more specifically Alzheimer's disease [190], or they don't have a dementia, but have MCI [42].

Similarly to CDR, the Global Deterioration Scale (GDS) has been used to classify the severity of normal and demented subjects [191].The GDS ranges from 1 (normal) to 2 (normal with subjective complaints) on to stages 3 through 7, which represent increasing levels of impairment. Kluger et al [192] claim that the GDS of 3 is equivalent to MCI. Flicker et al [193] demonstrated that subjects with a GDS of 3 may have MCI or may be mildly demented. The mean GDS for subjects classified with MCI by clinical criteria was 2.7 in a study by Petersen et al. [125], where some of the subjects had the classification of GDS 2 and some GDS 3 [42].

2.11 Neuroimaging methods

Neuroimaging methods, such as magnetic resonance imaging and electroencephalography, are employed in cross-sectional case-control comparison studies, in which case are cognitively impaired individuals compared against cognitively healthy controls. Neuroimaging methods are employed in longitudinal studies of cognitive function, where baseline measures of an individual are taken at the beginning of the study. Structural and functional changes in brain of the individual are assessed over time in relation to progress of cognitive impairment. Such information would allow the design of specific tasks which may probe structural or functional areas which are sensitive to cognitive ageing. The ability of an imaging measurement to predict which members of the cohort will and which will not develop cognitive impairment is the focus of this research [42]. Review of selected literature on use of neuroimaging methods for cognitive function assessment is presented in this section.

2.11.1 Magnetic resonance imaging

Anatomic Magnetic Resonance Imaging (MRI) is employed in studies of ageing and dementia. MRI is employed for diagnosis and monitoring disease progression. Detection of pathologic features like senile plaques, neurofibrillary tangles, decreased synaptic density, neuron loss, and atrophy relative to age-matched controls by MRI is used in studies of Alzheimer's disease [42]. Neurofibrillary pathology of AD affects medial temporal lobe limbic areas, neocortical association areas, as well as primary sensory and motor areas.

Diffusion tensor magnetic resonance imaging (DT-MRI) is a powerful quantitative method with the ability to detect in vivo microscopic characteristics and abnormalities of brain tissue. DT-MRI has also been used to study patients with cognitive decline, mainly those with

Alzheimer's disease. Several image-analysis approaches have been employed, including region of interest, histogram, voxel-based analyses and DT-MRI-based tractography. Specific patterns of spatial distribution of tissue damage and correlations with neuropsychological measures have been reported [194].

2.11.2 Functional imaging

Positron emission tomography (PET) and single photon emission computer tomography (SPECT) have been employed in studies of progressive memory impairment, individuals with genetic risk factor or family history of AD [42]. PET and SPECT techniques have demonstrated specific regional abnormalities in brain perfusion among patients with established AD. Decreased perfusion or metabolism in the temporoparietal cortex was commonly found among mildly impaired patients with probable AD [43, 44].

2.11.3 Electroencephalography

Investigation of neuronal dynamics at the whole-brain level by measurements of resting electroencephalography (EEG) microstate parameters was performed in study of dementia, Alzheimer's disease and schizophrenia [195]. In a study of Streletz et al [45], frequency power analysis applied to EEG measures yielded significant differences between controls and subjects with dementia (AD or Huntington's disease). Dunkin et al [196] demonstrated that quantitative EEG coherence between brain areas linked by long cortico-cortical fibres (termed "fascicle" coherence) was differentially reduced in subjects with Alzheimer's disease, whereas coherence between brain areas linked by short cortico-cortical and cortico-subcortical fibres in postcentral areas (termed "visual" coherence) was differentially reduced in subjects with multi-infarct dementia.

Assessing EEG measures over the course of dementia progression, a significant increase of delta and theta power in conjunction with decrease of alpha and beta power over a period of 30 months from diagnosis have been found [197]. Analysis of EEG rhythms during photic stimulation indicated that the non-demented older adults generally demonstrated more absolute power than AD patients during photo stimulation [198].

Event-related brain potential measures have distinguished between subcortical and cortical dementias [46, 47], and P300 latency can help to distinguish individuals with dementia from those with depression-associated pseudodementia. The latter group shows only the delay in latency associated with normal aging [199, 200]. Discrimination between patients with early Alzheimer's disease and healthy individuals has also been reported [201, 202], with

simple tasks yielding the largest P300 difference between groups [203]. More sophisticated cognitive paradigms have been applied to these populations with some success, but no bioelectric disease markers have yet been established [204, 205].

2.11.4 Limitations of imaging methods

While the neuroimaging methods are candidates for a biomarker for cognitive function, the use of neuroimaging methods is generally expensive, time consuming and requires the individual to visit a hospital/clinic. Neuroimaging assessments may take place at different neuroimaging centres, the protocols for image acquisition may not be well defined and may differ between centres [48]. If not evaluated by a computer algorithm, substantial variability between clinical raters was observed [48-50], which also suggests that a significant number of subjects may be misclassified [48]. Head position and image registration has been found to increase variability across clinical raters [48]. All these factors may contribute to variable assessment of cognitive function in older adults.

2.12 Structural changes in the brain associated with impairment of cognitive function

The use of neuroimaging methods has allowed more detailed investigation of structural changes of the brains in older adults. In a study of selective preservation and degeneration within the prefrontal cortex in ageing and Alzheimer disease [56], it was found that subjects with Alzheimer disease had less total prefrontal cortex grey matter than did age-matched healthy subjects. Further, a significantly less volume in the inferior prefrontal cortex region, where the Broca's area is situated (BA 44, 45). These degenerations in prefrontal cortex area may predict changes to motor control of articulators and thus changes to speech parameters.

A significant reduction of grey matter volume in the hippocampal formation was found in patients with Alzheimer's disease [57]. The volumes of hippocampus, putamen and thalamus were reported to be significantly reduced in patients diagnosed with probable Alzheimer's disease [58]. It was found that the decrease in volume of these deep grey matter structures correlated linearly with impaired global cognitive performance. Especially, the left volumes of the hippocampus, putamen and thalamus formed the strongest predictors of cognitive performance.

Regions such as the superior temporal sulcus area contribute to higher order cognitive functions, including language, integration of sensory afferents, and some aspects of attention.

The development of Alzheimer neuropathological changes and neuronal loss later in the disease process therefore likely reflects the clinical observations that symptoms in these cognitive domains frequently occur after the stage of minimal cognitive impairment and are symptoms associated with the clinical diagnosis of Alzheimer's disease [42].

Regions of the brain, such as prefrontal cortex, Broca's area, left superior temporal sulcus are involved in speech and language production. The structural and functional changes in these areas of brain of subjects with impaired cognitive function may lead to changes in the process of speech production. These changes in speech production may then result in changes of speech content (i.e. What is being said) or changes of speech characteristics (i.e. How it is being said).

2.13 Neural mechanisms of speech production

The production of speech is the end-product of a complex network of cognitive processes [206]. The transformation of thoughts and intentions into sequences of movements and sounds is the core focus of the psycholinguistic tradition of research into speech production [207]. Through the study of errors [208] and timing of production [209] this tradition has identified distinct phases of planning and control processes – the components and systems underlying speech production [206].

These phases have been aggregated into formal neuropsychological frameworks of speech production such as those presented by Levelt [210] and Whitworth et al [211]. Levelt's model incorporates a flow of processes that impart syntactic, morphological, phonological and phonetic organisation to drive the progression from thought to speech. The language processing model presented by Whitworth et al [211], based on Patterson and Shewell's logogen model [212], is a more broad cognitive description for speech than Levelt's, describing the processes not only for spontaneous or propositional speech, but also establishing the connection between reading, listening, writing and speaking. The connection between reading and speaking is important. For this reason emphasis is placed on Whitworth et al's model [211]. Using a model such as Whitworth et al's [211] requires that a number of assumptions be made, including modularity, subtractivity and universality of cognitive systems. These assumptions are points of contention however they have been research of Coltheart et al [213].

In the literature, criticism of psycholinguistic models has been focused on the fact that they fail to emphasise the importance of the last step of speech production, articulation [206, 207]. Articulation has traditionally been assumed to be a passive process for the transmission of linguistic signals from prior planning stages [206]. However contrary to this approach, it is now accepted that articulation forms an important aspect of communication that is actively adjusted by speakers to communicate effectively. For example, style and precision are two aspects of articulation that are continuously subject to modification to adapt to social and environmental contexts and to respond to conversational demands [206].

The neglect of the motor control system of articulation in neuropsychological frameworks may reflect the distinction that has traditionally been maintained between cognitive functions and motor control functions. Hauert [214] cited in Georgopoulos [215] argued "in support of the view that motor control is a cognitive function". Hauert's assertion came before published research in the ensuing years into the interface between motor control and cognition [215]. Evidence to support the involvement of the motor cortex in the processing of cognitive information related to motor function has been established and Georgopoulos [215] offers an overview of the various related research.

Modularity is one of the assumptions implicit in the use of models such as Whitworth et al's [211]. Functional modularity is the assumption that the components of a cognitive model are characterised by a high degree of independence. Anatomical modularity extends this concept of independence into the anatomy of the brain, stating that the modules of the cognitive speech system are localised in different parts of the brain [211]. Functional localisation of language speech production is debated in the literature, however regardless of whether speech production involves activation of discrete localised or much larger regions of the cortex there is considerable evidence that specific areas of the brain, such as Broca's area and primary motor cortex, are central in the process of producing speech [216].

Bridging the gap between 19th century models of speech production and modern cognitive neuropsychological models such as Levelt's [209] and Whitworth et al's [211], Price proposed a new neurological and cognitive model of speech production based on research in functional imaging [217]. In Price's model both auditory (listening) and visual (reading) stimuli were modelled however for the purposes of this thesis only the visual stimuli aspect of the model will be discussed. The visual processing of text is asserted as activating the posterior fusiform and lingual gyri. Price asserts that when speech output is required the posterior temporal gyrus (Wernicke's area) and the left posterior inferior temporal cortex exhibit

activation. Wernicke's area has been found to be associated with non-semantically mediated speech, in contrast with the left posterior inferior temporal cortex, which is activated by a range of word retrieval task including reading. Reading has also been linked to the activation of the occipital areas, and more specifically the left occipito-temporal region [218]. Semantic processing then activates the left posterior temporal middle temporal, posterior middle temporal, posterior temporoparietal and anterior inferior temporal cortices. Traditionally asserted as being specific to visual word forms, the angular gyrus is asserted by Price as being engaged with the establishment of semantic associations. Articulatory planning activates the anterior insula and the frontal operculum rather than the third frontal convolution (Broca's area) as was traditionally asserted. The final step in speech production, articulation, involves the sensori-motor cortices. Circuits involved with general motor activity include the cerebellum, basal ganglia and the thalamus [219]. During articulation the process of hearing the speech output also activates the superior temporal gyri enabling speech feedback circuits to be employed [217, 220].

2.14 Physical effects of ageing on speech production

Several physical and physiological changes occur in human bodies with ageing. As described earlier, typical changes include decline in sensory processing, vision and hearing, but also weakening of muscles, mobility restrictions and weakened immune system. Similar to other body parts, organs in the human speech production mechanism also undergo age related changes such as reduction in the respiratory muscle strength, restricted vocal fold adjustments during phonation and difficulty in adjustments of tongue and lip shapes [221]. The rate at which voices age does not however depend only on the chronological age of a person, but also on other factors such as lifestyle, physiological condition, smoking habits and profession. [222]

2.14.1 Changes in respiratory system

The most significant changes seen in the respiratory system of older adults are the loss of lung elasticity, increase in the stiffness of the chest wall and decrease in the respiratory muscle strength [223, 224].

Lung recoil elasticity is the ease with which lungs rebound after having been stretched during inhalation. A decline in lung elasticity due to ageing has been reported by Mahler et al. [223]. The loss of lung elastic recoil with age is found to be faster in males as compared to females [225].

Due to the alterations in the muscles of the chest wall, the thorax becomes increasingly rigid with ageing [226]. This leads to a reduced movement in response to the respiratory muscle forces. Due to the degeneration of the upper and middle regions of the thoracic vertebral column, a pronounced curvature of the back is observed in some older adults. This phenomenon called Kyphosis, alters the shape of the thorax and may affect the amount of air that can be inhaled and exhaled.

Several research studies have reported weakening of respiratory muscles during old age [226, 227]. This leads to reduced respiratory forces during inhalation and exhalation. A decline in maximal respiratory pressure progressively beyond the age of 65 has been reported by Enright et al. [228]. The decline is more prominent in males compared to females. A loss in diaphragm strength leading to an average reduction of 25% of maximum transdiaphragmatic pressure in elderly group as compared to younger subjects has also been reported [229].

While the total lung volume remains unaltered in the older adults, the forced expiratory volume and the lung pressure are decreased. This leads to a decline in the amount of air that can be moved in and out of the lungs and the efficiency with which it can be moved [230, 231]. The rate of this decline accelerates with advancing age [223]. Also the amount of air left after exhalation known as 'Residual volume' has been found to increase by about 40% from the age of 20 to the age of 70 [222, 232].

The changes in the respiratory system of older adults may have an impact on speech production. Hoit et al. [233] demonstrated that participants in the age 75 group adjust their linguistic performance by using fewer syllables per breath during extemporaneous speaking than participants of younger age groups. Huber et al. [234] observed that older adult speakers exhibited greater difficulty relative to utterance length and loudness and planned in advance for these difficulties by using processes associated with pre-motor speech planning. With the structural and functional changes in speech production areas of the brain (see Section 2.12) and motor function impairment in subjects with cognitive impairment [143, 235], the changes in respiratory system may further emphasise the impact on temporal and energy features of speech in the impaired subjects.

2.14.2 Changes in the larynx

The parts of the larynx that form the vocal apparatus are the laryngeal cartilages (to which the vocal folds are attached), the vocal folds that play a key role in phonation, and the intrinsic muscles that regulate the vocal cord tension and the vocal fold opening [236]. Several anatomical changes are seen in these organs with ageing.

Among the several cartilages in the larynx, the thyroid, cricoid and arytenoid cartilages are the most significant from the speech production point of view. The thyroid and cricoid cartilages form the skeleton of the larynx. A pair of arytenoid cartilages is located on the upper edge of the cricoid cartilage. The vocal cords are attached posteriorly to the arytenoid cartilages and anteriorly to the thyroid cartilages. The cricoarytenoid joints allow the arytenoid and thus the vocal apparatus to move laterally or medially. The arytenoids can also glide on the surface of the cricoid and move closer or recede away from each other. The most significant change in the cartilages observed as an individual moves from adulthood to old age is the toughening of the soft tissue into bone like structure (ossification). This phenomenon is observed in both males and females. It occurs at an earlier age and is more prominent in males as compared to females. Each of the cartilages has its own pattern of ossification. Arytenoid cartilage ossifies only partially sparing the vocal process. Significant age-related changes have been reported in the cricoarytenoid joint [237, 238]. Changes include thinning of the joint surface, reduced collagen fibers in the cartilage matrix and surface irregularities. These changes are again more prominent in males compared to females and hamper overall positional or postural movements of the arytenoid cartilages. This leads to reduction in the degree and extent of vocal ligament closure and makes it difficult for vocal fold adjustments during phonation. The result of this is impaired vocal quality and reduced vocal intensity due to air leakage through incomplete vocal fold closure.

The vocal folds have a complex layered structure. They are comprised of five discrete histological layers: the Epithelium, three layers jointly called Lamina Propria and the Thyroarytenoid muscle. The thin layer of Epithelium forms the protective covering for the vocal folds. The epithelial cells are bound together firmly and form a smooth lining reducing the friction to the air flow. The superficial layer of Lamina Propria is a thin layer made of elastin fibres. This layer can be stretched in several directions. The intermediate layer which is formed of elastin and collagen fibres is more densely packed and can only be stretched in anterior-posterior direction. The deep layer is formed on collagen fibres and is least stretchable. This layer protects the vocal cords from over extension. The Thyroarytenoid

muscle lies below the Lamina Propria. They are mainly concerned with pulling together the thyroid and arytenoid cartilage, thus relaxing the vocal folds.

Several changes in the structure with ageing alter the biomechanical properties of the vocal folds [221]. Glandular changes in the laryngeal mucosa (the mucous lining of larynx) [230] cause drying of the epithelial tissue, increasing the stiffness of vocal cord cover. This increase in cover stiffness leads to instability of vocal fold vibration. Some investigations [239] have reported thickening of laryngeal epithelium progressively with age. Tissues age at varying rates and to varying extents and substantial structural changes need to occur before observing noticeable changes in voice [222].

In the Lamina Propria, several age related changes have been documented in all the three layers. The thickness of the superficial layers alters [239] and atrophy and degeneration of the elastic fibres in the layer has been observed [240]. Changes seen in the intermediate layer include thinning of the layer, decrease in the density of the fibres, atrophy of the fibres and changes in the contour of the layer [221]. The fibrous protein loses elasticity and the layer stiffens. The deep layer thickens with an increase in the collagen fibres. Such morphological changes in the fibres of the vocal folds contribute partially to the ageing of the voice.

The thyroarytenoid muscle also displays atrophy with ageing. Changes in muscle fibres have been reported [241]. A decrease in thyroarytenoid muscle activity has been reported [242] in older speakers than young speakers. This affects the fine control of the position of the arytenoid joint and thereby the fine control of the pitch of the voice.

Intrinsic laryngeal muscles are responsible for control of the vocal cords. The tension in the vocal cords is regulated by the cricothyroid muscle. The opening (abduction) of the vocal fold opening (called Rima Glottidis) is controlled by the posterior cricoarytenoid muscle and the closing (adduction) is controlled by the lateral cricoarytenoid and thyroarytenoid muscles. Regressive changes and atrophy have been reported in all these muscles with ageing [243, 244]. The changes include accumulation of fats, degeneration of muscle fibers and unusual variations in the cross sectional areas [221]. As a result, precise control of the vocal cord tension and complete abduction/adduction is affected [222]. Associated motor function impairment in subjects with cognitive impairment [143, 235] may have further impact on the precise control of the vocal cords in cognitively impaired older adults. This may lead to more pronounced changes in acoustic features of speech, such as the fundamental frequency, in the impaired subjects.

2.14.3 Changes in the vocal tract

The human vocal tract consists of all the organs above the vocal folds that are involved in speech production. It is comprised of the pharynx (throat), the oral cavity, the nasal cavity, soft palate (velum) and the articulators viz., the tongue and the lips. The human speech production mechanism can be viewed as a source-filter model. The lungs in conjunction with vocal cords act as the source and expel air into the vocal tract. Depending on the presence or absence of the vocal cord vibrations, the source is either voiced or unvoiced. This quasi periodic air then resonates in the pharynx, oral and nasal cavities to generate a rich timbre. The vocal tract thus acts as the filter.

The vocal tract can be broadly thought to be comprised of three resonating cavities, the pharynx, and the oral and nasal cavities. The pharynx is involved in the production of all speech sounds. The pharynx can change shape to a limited extent and thus alter the resonance patterns. The pharynx can be constricted, and it can be raised or lowered.

The position of the velum also alters the shape of the pharyngeal cavity. The velum controls the flow of air into the nasal cavity. During the production of nasal sounds such as /m/ and /n/, the velum is moved forward to open the air passage through the nasal cavity. The oral cavity is the most flexible among the three cavities in varying the shape. The resonating property of the oral cavity depends on the position of the temporomandibular joint, the shape of the tongue and the lips and the position of the velum.

Thinning of pharyngeal epithelium and degeneration of the pharyngeal muscles has been reported with ageing [221]. However these changes in the pharynx are not found to be extensive.

The temporomandibular joint is the joint at which the jaw is hinged to the skull. It is used in controlling the position of the jaw and hence influences the oral resonance during speech production. Jaw movement has a significant role to play in articulation of certain phonemes as well as in the co-articulation of adjacent phonemes.

With ageing, degenerative changes are observed in the temporomandibular joint [245]. Displacement of the temporomandibular joint disk is commonly observed leading to a lowering of the articulating surface. Xue and Hao [246] have reported increase in vocal tract dimensions in older speakers. The vocal tract volume of older speakers in particular is significantly higher compared to the younger speakers. This could lead to changes in the resonance patterns in older voices.

The tongue plays a major role in speech production. It is very flexible and can be moved up, down, forward and backward. By adjusting the shape of the tongue and the position of the tongue tip, the oral cavity's shape is modified affecting the resonance patterns and hence the sound produced. Significant changes have been reported in the tongue with ageing [Rother et al., 2002]. Decrease in the thickness of epithelium and glandular atrophy have been reported in people over 50 years of age [247].

However the most significant change in the tongue during ageing that affects the speech production is the atrophy of the tongue muscles. From ultrasound observations, decline in the tongue motor skills in the older adults in comparison to young adults were reported by Koshino et al. [248]. A decline in tongue strength has also been reported in older individuals [249]. These changes in the tongue could affect the articulatory patterns. [222]

2.14.4 Neuromuscular control

Another aspect that age effects relates to changes that take place in the peripheral and central nervous system that have implications for speech production. One of the changes in the peripheral neural system is the decline of motor neurons. This loss in the motor units has been implicated as the primary mechanism for muscle atrophy and loss of contractile strength in the muscles [250]. An average loss of 25% neurons has been reported from the second to the tenth decade of life. However this loss of motor units is partially compensated by an increase in the size of the motor units along with a slowing of the contractile speed. This affects various muscles involved in the speech production and is a possible cause of the slower speaking rate observed in older speakers. [222]

2.15 Influence of impaired cognitive function on speech production in abnormal ageing

Investigations employing speech as a measure of cognitive function have been reported in several studies [59, 61, 251]. The aim in these studies has been to establish changes in speech production as a way to probe an individual's cognitive function.

2.15.1 Dementia

Dementia is characterised by a breakdown in intellectual and communicative functioning, manifesting itself as communication disorders [62]. These disorders are present in 88-95% of all those with dementia, being most pronounced in those with Alzheimer's disease [62]. In terms of speech, the principal features of such communication disorders include word finding

deficits, paraphasias, circumlocution, comprehension impairments and impairments in discourse [62]. Each of these features has a profound impact on the performance of speech production, and in particular the fluency.

Lowit et al [63] investigated articulation rate, articulation/pause time ratio and percentage change in articulation rate from habitual to fast and habitual to slow reading condition. The cohort included controls, subjects with Parkinson's disease (PD) but no cognitive decline and subjects with early onset dementia. The Lowit study found the subjects with PD and subjects with early onset dementia were able to be distinguished from controls using several speech measures. Specifically - measures of articulation rate in normal, fast and slow speech and the rate change from normal to slow and normal to fast. The early onset dementia group was classified as having no dysarthria (lack of motor control), suggesting the speech changes are influenced by cognitive decline. This study is also one of the few studies to incorporate the use of read speech (produced while reading a text passage).

2.15.2 Mild cognitive impairment

Other studies have found spoken language markers to be indicative of Mild Cognitive Impairment (MCI). Roark et al [61] reported standardized pause rate (the ratio of words uttered to the number of pauses uttered) to be statistically different for two sets of older adults speakers; using a Clinical Dementia Rating (CDR), the cohort was separated into those with MCI and those without. While other speech duration parameters were investigated in the study including verbal rate (number of words per second), phonation rate (proportion of total time spent speaking) and mean duration of pauses, none of these were found be significantly different for the two populations.

Another study by Roark at al [60] investigated speech and linguistic metrics from a spontaneous speech task to discriminate between clinically defined cognitively impaired subjects and matched controls. This study recorded speech from a passage recall task from 74 participants from 82 to 95 years of age. The audio recordings were transcribed by hand and the transcriptions automatically parsed to extract linguistic and temporal speech features.

The Roark study [60] found several linguistic and temporal features that discriminated between the two groups. The speech features found to be most discriminative between the two groups of recordings were standardized pause ratio, total phonation time, phonation rate and transformed phonation rate. The standardized pause ratio is the number of words in the recording divided by the total number of pauses. The transformed phonation rate is the

arcsine of the square root of the phonation rate. Phonation rate is the number of phonemes per second of speech. The authors employed a support vector machine to classify participants and calculate the Area under the ROC curve (AUC_ROC). The highest AUC_ROC value achieved based only on linguistic and temporal speech features was 0.732. The AUC_ROC value obtained from combining an array of cognitive tests was 0.815, while including a subset of the temporal and linguistic features increased the AUC_ROC value to 0.861. This study demonstrated the effectiveness of speech and language features in discriminating between cognitively healthy and cognitively impaired older persons.

2.15.3 Alzheimer's disease

The possibility that early linguistic ability impairment may precede the onset of AD is apparent, for instance, when Butterworth analysed Ronald Reagan's speeches the year prior to his re-election as president of United States (*The Sunday Times*, 4th November 1994). Ten years before Reagan was diagnosed with Alzheimer's disease, Butterworth observed a remarkable impoverishment of language with simplified clauses and fragmentation of speech. Another case of early signs of dramatic decline can be found in the written texts, when available, like in Iris Murdoch's last book "Jackson's Dilemma". Garrard et al. [252] analysed the grammatical complexity and the language diversity of three of Murdoch's works (i.e., *"Under the net"*, *"The sea, the sea"* and *"Jackson's dilemma"*) and found clear signs of diminished ability to produce a creative and enriched language, with shorter sentences and lower density of meaning. The observed changes, such as fragmentation of speech and shorter sentences, may have an impact on temporal features of speech.

2.15.4 Other conditions associated with cognitive impairment

A study by Lieberman et al [253] investigated audio recorded from climbers ascending Mount Everest. The effect of altitude on the cognitive function of climbers was used to simulate that experienced by astronauts in space. Astronauts' cognition is impacted by hypoxic and cosmic ray-induced insult to the brain, as well as degraded cognitive performance resulting from task difficulty. The authors found increased vowel duration under these conditions and a hit rate in discriminating impaired from unimpaired performance on Mini-Cognitive tests of working memory and vigilance of 85%. This study demonstrated that even artificially induced impairment of cognitive function can be detected by analysis of speech characteristics.

2.16 Influence of psychiatric conditions on speech production

Besides the studies of speech production in ageing, research in the literature has also been reported in this area with regard to psychiatric illnesses. Symptoms such as affective flattening, blunted affect, the inability to experience or express a normal range of affective responses, emotional dullness, pervasive apathy, poverty of speech and psychomotor retardation have been constituents of clinical definitions of schizophrenia and have led to the negative-positive model of schizophrenia [254, 255]. In this model, delusions, hallucinations and florid formal thought disorders are conceptualized as pathological excesses, whereas negative symptoms are conceptualized as pathological deficits.

Based on this heuristic division, negative symptoms have been theorized to be unresponsive to antipsychotic medication, to be associated with a deteriorating course, and to reflect some structural brain abnormalities that would mediate the hypothesized association between negative symptoms and poor prognosis.

Recent findings, however, have indicated that positive symptoms are not uniformly responsive to treatment, and that negative symptoms are not universally immutable. Specifically, during recovery from an acute episode of schizophrenia, the key symptoms tend to disappear first while symptoms at lower levels persist longer. Non-specific and affective symptoms tend to recur as prodromal symptoms of relapse. Furthermore, several authors have reported a diagnostic non-specificity of negative symptoms, particularly with regard to various clinical forms of depression which occur relatively commonly in schizophrenic patients and there is also some overlap between the specific phenomena of depressive illness and negative symptoms in schizophrenia [251].

2.16.1 Formal thought disorder

In the psychiatric literature, many of the abnormalities of language in schizophrenia are grouped together as "formal thought disorder"; a disorder in the form of thought, not content. Perhaps most commonly it is the moment-to-moment, logical sequencing of ideas which is at fault. At other times, the mechanisms of language production appear to be disturbed, so that the meaning of individual words and phrases is obscured. While at other times, the disfunction seems to be at the level of discourse: individual words, sentences and sequences of thought make sense, but there is no discernible thread to longer verbal productions.

Formal thought disorder is a relatively uncommon finding in acute schizophrenia, though it is somewhat more common in chronic cases. Manifestations of formal thought disorder include poverty of content (failure to express sufficient information), loss of goal (slippage away from the intended topic), clanging (chaining together similar sounding words as if distracted by them), and other kinds of incoherence and unintelligibility.

2.16.2 Schizophrenic language disorders

The study of schizophrenic language disorder by linguists began with Chaika [256], who studied a single patient who spoke normally for weeks at a time, her deviant language coinciding with what her psychiatrists termed psychotic episodes. The abnormalities that Chaika observed were:

1. Failure to utter the intended lexical item.
2. Distraction by the sounds or sense of words, so that a discourse becomes a string of word associations rather than a presentation of previously intended information.
3. Breakdown of syntax and/or discourse.
4. Lack of awareness that the utterances are abnormal.

Of these, observation 2 is most characteristic of schizophrenia; observations 1 and 3 resemble ordinary speech errors, and observation 4 resembles some form of aphasia [256].

Fromkin [257] reported that except for the disruption of discourse which can be attributed to non-linguistic factors, all the features [of schizophrenic language] are prevalent in normal speech as exemplified by speech errors and "slips of the tongue". Mistaken lexical choices and minor scramblings of syntax are common in everyday speech. Indeed, speech errors are often triggered by the sounds or senses of recently uttered words, and speakers are commonly unaware of their fumbles [208, 258]. Thus, observations 1, 2, 3 and 4 of Chaika's core abnormalities are all removed, except for derailment of discourse, which Fromkin considers extra linguistic.

This claim has not held up. Although there are obvious similarities between Chaika's samples of schizophrenic language and Fromkin's corpus of speech errors, there are also obvious differences. Normal speakers make occasional errors like those seen in schizophrenia, but not a whole strings of errors. A representative patch of gibberish from Chaika's patient comprised of nine syllables, and uncorrected speech errors of such length and unintelligibility

do not occur in normal speech. Another aspect of importance is for normal speakers, when an error is pointed out, they immediately correct it, whereas speakers with schizophrenia do not.

Moreover, Chaika's patient would commonly string together 10 or 20 sentences connected only by word associations [256]. Another subjects, even when plagued by speech errors, do not do this. Normal speech errors are momentary deviations from a discourse plan that is immediately resumed.

In later work, Chaika [259] argued that schizophrenic language disorder is fundamentally a loss of voluntary control over the speech generation process. Indeed, according to Chapman [260], patients sometimes say in retrospect that this is exactly what happened—they could not control their speech. This is similar to the main theme of the Schneiderian first-rank symptoms [261, 262], which is loss of control over the train of thought. Note, however that Chaika's original patient apparently lacked such insight.

Chaika argued [259] that loss of voluntary control ties together a wide range of observed phenomena, depending on which part of language production goes out of control—most often discourse organization, but often lexical retrieval, and sometimes pronunciation or syntax. It fits well into a more general conception of schizophrenia as degradation of communication between mental subsystems.

2.16.3 Schizophrenic language disorders vs. aphasia

One question that needs to be addressed is whether language disturbances of schizophrenia resemble the aphasia caused by stroke, traumatic brain injury, or neurological conditions such as epilepsy.

Publications in the literature agree that there are important differences, but beyond that, discussion of the issue has been complicated by the heterogeneity of both schizophrenia and aphasia [263]. Aphasia-like symptoms are episodically observed in only a small proportion of subjects considered to be schizophrenics whereas the aphasia produced by stroke or brain injury is in most cases constantly present. Patients with aphasia have normal thoughts and express them with difficulty; those with schizophrenia have unusual thoughts (or disorganized discourse plans) and express them with comparative ease.

Pinard and Lecours [264] compared schizophrenic language to Wernicke's aphasia (including jargon aphasia), a disorder in which the patient speaks fluently but unintelligibly. Their main findings included:

1. Schizophasic discourse often has a preferred theme or preoccupation; aphasic discourse rarely does.
2. Speakers with schizophrenia often jump from one subject to another based on the sounds or associations of words they have uttered (association chaining or glossomania). This is seldom observed in jargon aphasia; it requires lexical mastery well beyond that of most aphasics, as well as remarkable control of prosody.
3. Schizophasic discourse often includes rare words, evidence of a large, intact vocabulary; jargon aphasia, even when very fluent, shows a restricted vocabulary.
4. Schizophrenic speech can include conscious creation of new words (neologisms) and consciously constrained discourse in which the speaker is well aware that the speech is unusual, whether or not others can understand it. Aphasic speakers who produce fluent unintelligible discourse do not seem to be fully aware of what they are doing, and if they create new words, it is as if by accident.

2.16.4 Speech characteristics in schizophrenia

Predominantly negative-symptom schizophrenia (NSZ) can be characterized by a number of atypical reductions in observed behaviour, including communication behaviour. Classically observed outward manifestations in spoken (e.g., linguistic production) and pragmatic/paralinguistic (e.g., body language, intonation, and prosody) aspects of communication are related to affective flattening, anhedonia, alogia, and avolition (the DSM IV summarized characterization of NSZ), implying an effect on both affect and cognition. Atypical communication can have an effect on clinical assessment, as it may influence rater scoring on a number of items from subjective behavioural rating scales (e.g. PANSS) including items such as blunted affect, emotional withdrawal, poor rapport, social withdrawal, reduced flow of conversation, and motor retardation [265]. Although these rating scales are important in their ability to help clinicians assess symptomatology, the addition of objective and quantifiable measures of disease severity and therapeutic treatment response are desirable and possible through speech and voice acoustic measurement.

In the most basic terms, physical quantitative measurements of communication behavior, using aspects of frequency, intensity, and time, support and extend the clinical impressions of atypical communication used clinically [265]. This adds clinical value by providing repeatable quantification of observed symptomatology.

The literature surrounding acoustic investigations in persons with schizophrenia has revealed a number of consistent themes. Flat affect, alogia[5], and asociality (as measured by the PANSS scale) are strongly related to restricted speech output, monotonous speech, pause in speech, energy variation, utterance duration, and inflection [265-267]. In addition, acoustic measures have shown great promise in identifying treatment response by demonstrating a larger treatment effect than those seen with traditional rating scales [268]. It has also been demonstrated that specific measures of acoustic inflection are sensitive enough to differentiate between antipsychotic medication (Olanzapine vs. Haloperidol; Remoxipride vs. Haloperidol), whereas rating scales were not able to detect this difference. Bidirectional changes in the speech acoustic characteristics between drug conditions have also led researchers to the conclusion that different mechanisms of drug action that may be at work, though rating scales were not able to make this distinction [269]. Acoustic measures were able to separate the different drug groups at outcome while the rating scales failed to show a difference.

In a recent acoustic investigation by Wisniecki et al. [270] have been successfully demonstrated measurable differences in cognitive behaviour and motor slowing by comparing persons with NSZ to a control group. Speech pause behaviour in a simple counting and picture description task have demonstrated that average pause length was an indication of motor retardation, whereas global measures of pause were indicative of the increased cognitive linguistic demands of the picture description task. [270]

Based on a sample of 42 chronic schizophrenic patients and 42 carefully matched controls, Stassen et al. [251] investigated potential relationships between acoustic variables on the one hand, and negative syndromes, positive syndromes and affective disturbances, on the other. A set of 12 acoustic variables automatically assessed in a standardized experimental setting allowed an almost perfect discrimination between schizophrenic patients and normal subjects. Acute side-effects of medication did not explain this finding. However, the question of whether the observed changes in speaking behaviour and voice sound characteristics were caused by long-term neuroleptic treatment, for example, as a consequence of tardive dyskinesia, was not able to be answered by the investigation. In view of a biological validation of the negative-positive model of schizophrenia, the reliability of various psychopathological subscales was tested through repeated assessments at 14 day intervals.

[5] Alogia , or poverty of speech, is a general lack of additional, unprompted content seen in normal speech. Alogia is often caused by a disruption in the thought process. In conversation, alogic subjects reply very sparsely and their answers to questions lack spontaneous content. Sometimes, they even fail to answer at all. Their responses are brief, generally only appearing as a response to a question or prompt.

Stassen et al. [251] observed most psychopathology scores to be sufficiently stable and reproducible over time, thus representing a suitable basis for the estimation of severity with respect to the negative and positive component of schizophrenia. Using the first measurements as training samples and the second measurements 14 days later as test samples, discriminant analysis yielded conclusive proof of a close relationship between acoustic variables and the severity of the negative and positive component of schizophrenia. In particular, by means of "objective" acoustic variables and under the constraint of reproducibility, 75.9% of patients were correctly classified as low or high scorers with respect to the negative syndrome, 71.9% of patients with respect to the positive syndrome, and 79.4% of patients with respect to their depressive symptomatology [251].

To asses speaking behaviour and voice sound characteristics 16 acoustic variables were used. All these variables have turned out to be highly stable over time and to be sufficiently sensitive to distinguish between emotionally neutral and emotionally charged texts read out loud by the same individuals.

2.17 Summary

This chapter defined the importance of cognitive function in quality of life and how cognitive functioning is affected in ageing. The specific domains, attention, memory, perception and higher-level cognitive functions, are all critical to ensuring that cognitive function is maintained. Neurodegenerative conditions, such as dementia, mild cognitive impairment and Alzheimer's disease affect cognitive function. Current gold-standard neuropsychological methods for cognitive function assessment have limitations that impact the objectivity of the cognitive assessment and its availability for older adults. More advanced neuroimaging methods (MRI, EEG, SPECT) for assessment of cognitive function are the favourable candidates to discover a biomarker for cognitive function. Despite this fact, these methods also face issues with objectivity and are time consuming, expensive and require the older adult to visit a hospital or a clinic. Structural and functional changes in regions of the brain associated with speech production of cognitively impaired subjects have been reported in the literature. Qualitative and quantitative changes in speech characteristics of subjects with cognitive impairment have been investigated in the literature. Qualitative changes in speech characteristics of subject with psychiatric cognitive disorders, such as schizophrenia, have also been reported in the literature. Alternative methods that provide objective assessment of cognitive function and would be highly accessible for the older adults are being actively sought.

2.18 Research questions

Following a thorough review of the current methods and limitations of assessing cognitive function and delivery of cognitive function assessment, a number of questions beset the literature.

- Can changes in an individual's cognitive function be objectively assessed by quantitative changes of their speech characteristics?
- Can remote monitoring of cognitive function be carried out through speech?

The hypothesis developed to address these questions is that speech provides a window into cognitive function. To prove this hypothesis a series of specific research questions were posed.

2.18.1 Methodology questions to address research hypothesis

1. Can the temporal and acoustic features of speech be extracted reliably from speech recordings taken in real life environments, while removing artefacts such as breath sounds and compensating for varying levels and sources of noise?

2. Can the removal of breath sounds from speech recordings prior to temporal and acoustic features extraction improve the classification performance between cognitively healthy and cognitively impaired individuals?

2.18.2 Speech characteristics in schizophrenia and their correlation to cognitive function

3. Does the status of individual's cognitive function correlate with temporal features of speech of individuals with/without schizophrenia extracted from a recording of reading out loud of a short text passage?

4. Does the status of individual's cognitive function correlate with acoustic features of speech of individuals with/without schizophrenia extracted from a recording of reading out loud of a short text passage?

2.18.3 Changes in speech characteristics during ageing and their correlation to cognitive function

5. Does the status of individual's cognitive function correlate with temporal features of speech of older adults extracted from a recording of reading loud a short text passage?

6. Does the status of individual's cognitive function correlate with acoustic features of speech of older adults extracted from a recording of reading loud a short text passage?

7. Does the status of individual's cognitive function correlate with temporal features of speech of older adults extracted from a recording of a picture description task?

8. Does the status of individual's cognitive function correlate with acoustic features of speech of older adults extracted from a recording of a picture description task?

9. Does the status of individual's memory correlate with temporal features of speech of older adults extracted from a recording of a picture description tasks?

10. Will the features extracted using dynamic pause threshold from speech recording of reading out loud of a short text passage increase the ability to discriminate between cognitively healthy and cognitively impaired older individuals?

2.18.4 Remote assessment of cognitive function

11. Can the speech features be extracted reliably from recordings of speech acquired over a telephone?

12. What is the protocol that will enable remote administration of clinical cognitive assessment over a telephone?

13. What are the design requirements for both protocol and implementation for a remote, longitudinal assessment of cognitive function of older individuals?

14. Can the cognitive assessment be fully automated for large scale deployments using existing telephone infrastructure?

15. Can the analysis of remotely collected speech recordings provide equivalent information about cognitive function compared to labour intensive gold-standard cognitive assessment?

16. Can speech be employed to identify changes in cognitive function over time?

17. Do the results of a face-to-face cognitive assessment correlate with the results of automated cognitive assessment?

18. Will the older individuals be comfortable using the automated cognitive assessment technology?

19. Will the cost of remote cognitive assessment be lower than the cost of gold-standard cognitive assessment?

20. Can an assessment of memory and attention be delivered automatically over a website?

21. Can a battery of tasks for the assessment of memory and attention be administered at regular intervals over time without inducing a practice effect?

22. Can a battery of tasks for the assessment of memory and attention employ the same stimuli over time without inducing a practice effect?

To address these research questions this thesis is organised into a specific series of cross-sectional and longitudinal studies involving a large cohort of cognitively healthy and cognitively impaired older adults. To undertake these studies a series of signal processing methods need to be developed. The next chapter, Chapter 3, describes details of these methods.

Chapter 3

Signal processing of speech signals pertinent to cognitive function

The hypothesis to use speech measures to assess and monitor changes in cognitive function was introduced in Chapter 2. An important step in testing this hypothesis is to extract speech based features which may be related to cognitive function for a variety of speech based tasks. It is necessary to accurately extract these features from speech recording carried out in real-life environments.

For the purpose of the assessment of cognitive function through use of speech measures, a novel algorithm for automatic extraction of temporal features from speech recordings was developed. Another novel algorithm was developed that detected artefacts such as breath sounds detection and removed them prior to the extraction of the features. The features can then be incorporated into a system for cognitive function assessment.

The analysis of speech described in this chapter was employed in all studies in the following chapters.

A number of publications have been derived from the research described in this chapter.

V. Rapcan, S. D'Arcy, N. Penard, I. H. Robertson, and R. B. Reilly, "The Use of Telephone Speech Recordings for Assessment and Monitoring of Cognitive Function in Elderly People," *Proceedings of Interspeech 2009*, Brighton, UK, 2009, pp. 943-946.

V. Rapcan, S. D'Arcy, and R. B. Reilly, "Automatic breath sound detection and removal for cognitive studies of speech and language," *Proceedings of the 20th IET Irish Signals and Systems Conference 2009*, pp. 1-6.

3.1 Extraction of speech features

Initial processing of any speech recording requires segmentation into speech and non-speech segments. Non-speech segments here were considered silences, breaths and other non-speech artefacts such as clicks, knocks and coughs.

All speech recordings (f_s = 44.1 kHz) were high-pass filtered at 80 Hz with 7th order type II Chebyshev filter (roll-off of 55dB/octave) prior to processing in order to remove the ambient noise existing in lower frequency bands and improve signal to noise ratio.

Pauses longer than 250ms were manually removed at the beginning and the end of each recording. This was carried out to ensure that analysis was performed from the moment the person starts to speak until the person stops speaking (otherwise the number of pauses would be increased in error by two, having a corresponding incorrect increase in duration of pauses).

3.1.1 Speech/Non-Speech Threshold estimation

A *Speech/Non-speech Threshold* is estimated at the beginning of feature extraction process, and is subsequently employed to separate the audio data into speech and non-speech segments.

The stages for *Speech/Non-speech Threshold* estimation are as follows.

1) Full-wave rectification is performed on the speech signal. Full-wave rectification is defined as

$$x_{rect}(n) = abs(x(n))$$

Eq. 1

2) The rectified signal is divided into non-overlapping frames of 50ms duration and the energy is calculated in each frame. The energy of a discrete time signal [271] is defined as

$$E_n = \sum_N x^2[n]$$

Eq. 2

3) Based on observations of the recordings of subjects reading aloud, at least 15% of the frames typically represent no speech samples. Therefore, 15% of frames with the lowest value of energy are selected for *Speech/Non-speech Threshold* estimation and maximum amplitude value (***A_{max}***) in each of these frames is stored.

$$A_{\max}(i) = \max(x_{rect(i)}(n))$$

Eq. 3

for i = 1, 2, ..., M, where M is number equal to 15% of the 50ms frames of the speech signal, e.g. a file of 5s will be divided into 100 frames and M will be equal to 15. And n = 1,2,..., N, where N is the number of samples in each frame.

4) Finally, the ***Speech/Non-speech Threshold*** (***Th_{SN}***) is calculated as a mean value of the stored maximum amplitude values.

$$Th_{SN} = \frac{1}{M}\sum_{i=1}^{M} A_{\max}(i)$$

Eq. 4

for i = 1, 2, ..., M, where M is number equal to 15% of the 50ms long frames of the speech signal.

3.1.2 Automatic breath sounds detection and removal

While speaking, pauses are inserted into speech to provide extra time with which to address the cognitive load of the current task. Breaths can be often substituted for silence to produce 'filled pauses' that achieve the same result. Therefore, breath sounds may theoretically be classified as a part of a pause. However, many feature extraction algorithms classify breath sounds as speech [251]. Thus, breath sounds detection and removal is required prior extraction of temporal features from speech signals.

Gradient Threshold

The ***Gradient Threshold*** (***Th_{GRAD}***) is a gradient value above which the signal is considered speech, where gradient is defined as

$$g_n = x_{rect(n+1)} - x_{rect(n)}$$

Eq. 5

If one inspects the amplitude envelope of breath sounds and compares them to the amplitude envelope of speech sounds, a clear difference in gradient can be observed. The gradient value of two adjacent speech samples is usually much higher than that of two adjacent breath sounds samples. See Figure 3-1.

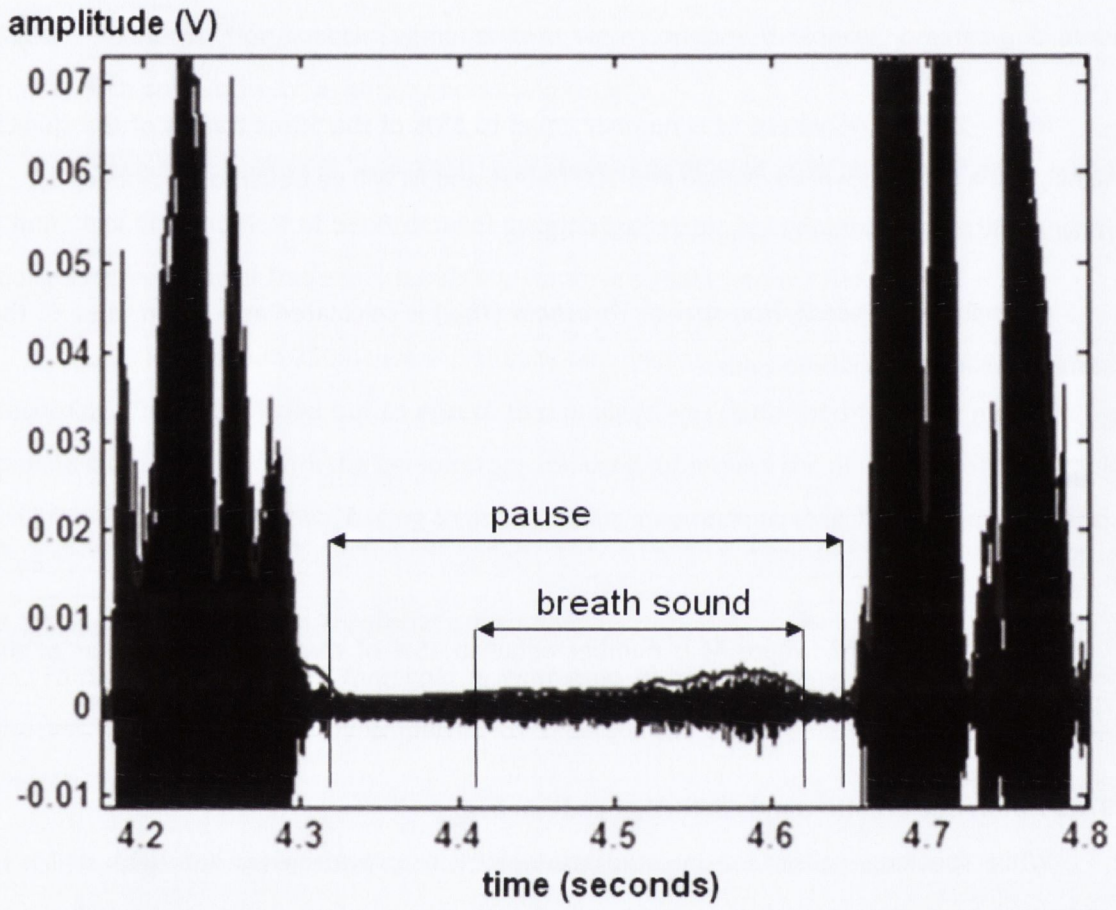


Figure 3-1: *Difference in amplitude envelope between breath sound and speech signal.*

Reference Value

The *Gradient Threshold* differs for each speaker, and so a ***Reference Value*** (***RV***) was calculated which can be employed to estimate this *Gradient Threshold*. This *Reference Value* was estimated experimentally by randomly selecting 10 speech files from all available recordings. For every file, the *Gradient Threshold* (Th_{GRAD}) which detected all breath sounds in the recording was estimated. The *Reference Value*, for each selected recordings, was calculated by dividing the *Gradient Threshold* (Th_{GRAD}) by the *Speech/Non-speech Threshold* (Th_{SN}) value for that particular recording.

$$RV(i) = \frac{Th_{GRAD}(i)}{Th_{SN}(i)}$$

Eq. 6

for i = 1, 2,..., *N*, where *N* is the number of speech files, i.e. 10 in this case.

Finally, the *Reference Value* was calculated as an average value of all 10 recording's reference values.

$$RV = \frac{1}{N}\sum_{i=1}^{N} RV(i)$$

Eq. 7

for *i* = 1, 2,..., *N*, where *N* is the number of speech files, i.e. 10 in this case.

Breath sounds detection

Initial identification of breath sounds is based on an amplitude envelope detection and is described in the following four steps below. This allows to detect breath sounds within the speech signal. The steps for Breath Sounds Detection are as follows:

1) The amplitude envelope of the waveform is generated by performing Hilbert transform and full-wave rectification (*Eq.* 1) on the speech signal. The sampling rate is then decreased by decimation of the signal and the resulting signal is low-pass filtered at 33 Hz.

The original sampling frequency of 44.1 (22.05) kHz requires the signal to be decimated by a factor of 30 (15) in order to achieve a smooth envelope.

2) The gradient between each adjacent sample of the envelope is calculated. The additional advantage of measuring envelope gradients, rather than an absolute amplitude values, is that if one changes the volume level during recording it will change the value of amplitude. However the value of the difference between two neighbouring samples will not be significantly affected.

3) Two conditions that must be satisfied so that the sample is marked as a breath sample are:

a) The value of the gradient has to be below the *Gradient Threshold*, which is defined as

$$Th_{GRAD} = RV \cdot Th_{SN}$$

Eq. 8

where Th_{SN} is the *Speech/Non-speech Threshold* for the particular recording that is being processed.

b) False positives can arise from individual gradient values (particularly at the peaks of the envelope) being less than the *Gradient Threshold*. The amplitude value of the sample must be below 120% of the *Speech/Non-Speech Threshold* to be considered a breath sound.

The *Speech/Non-Speech Threshold* is calculated on the segments of the smallest amplitude; however breath sounds have higher amplitude than the mean amplitude of noise and thus 120% was found experimentally to compensate for this difference.

4) Breath sounds are typically longer than 100ms [272], thus segments identified as possible breath sound shorter than 100ms are considered to be speech segments.

Breath onset/offset detection algorithm

The breath sounds detection does not always detect the whole breath sound, but only part of it. For this reason, it is important to more accurately detect the boundaries of the breath sound. Breath onset and offset detection is the process of maximizing temporal resolution in order to find the boundaries of breath sounds. The stages of onset/offset detection are as follows:

1) In order to detect full breath sound, the breath sound segment is extended by 200ms on both sides. The extended segment is divided into 10ms frames. The energy is calculated for each frame.

2) Energy contour is estimated by calculating the *log energy* (E_{log}) in every frame. The log energy is defined as

$$E_{\log} = \log_{10}(E_x)$$

Eq. 9

where E_x is the discrete time signal energy defined in *Eq.* 2. A three-point moving average filter is subsequently applied to smooth the energy contour.

3) A *maximum energy value* (***E**max*) is estimated over the original (non-extended) breath sound segment.

4) A *minimum energy value* (***E**min*) is estimated over the extended breath sound segment.

5) The *Energy range* (***E**range*) is calculated as

$$E_{range} = E_{\max} - E_{\min}$$

Eq. 10

6) Local minima are located over the energy contour and non-significant minima are excluded. A minimum is considered non-significant if the difference between E_{max} and the energy value of minimum is less than 70% of E_{range}.

7) Finally, onset and offset of the breath segment are calculated based on the position of significant minima closest to the position of the maximum of the original breath segment.

Breath sounds removal

After breath onset/offset detection, the breath segments are replaced with silence of the same length as is the length of the breath segment. The silence is represented by a signal with zero amplitude. The length of the breath segment is calculated from the onset and offset of the breath segment. The recording is then passed for the feature extraction phase.

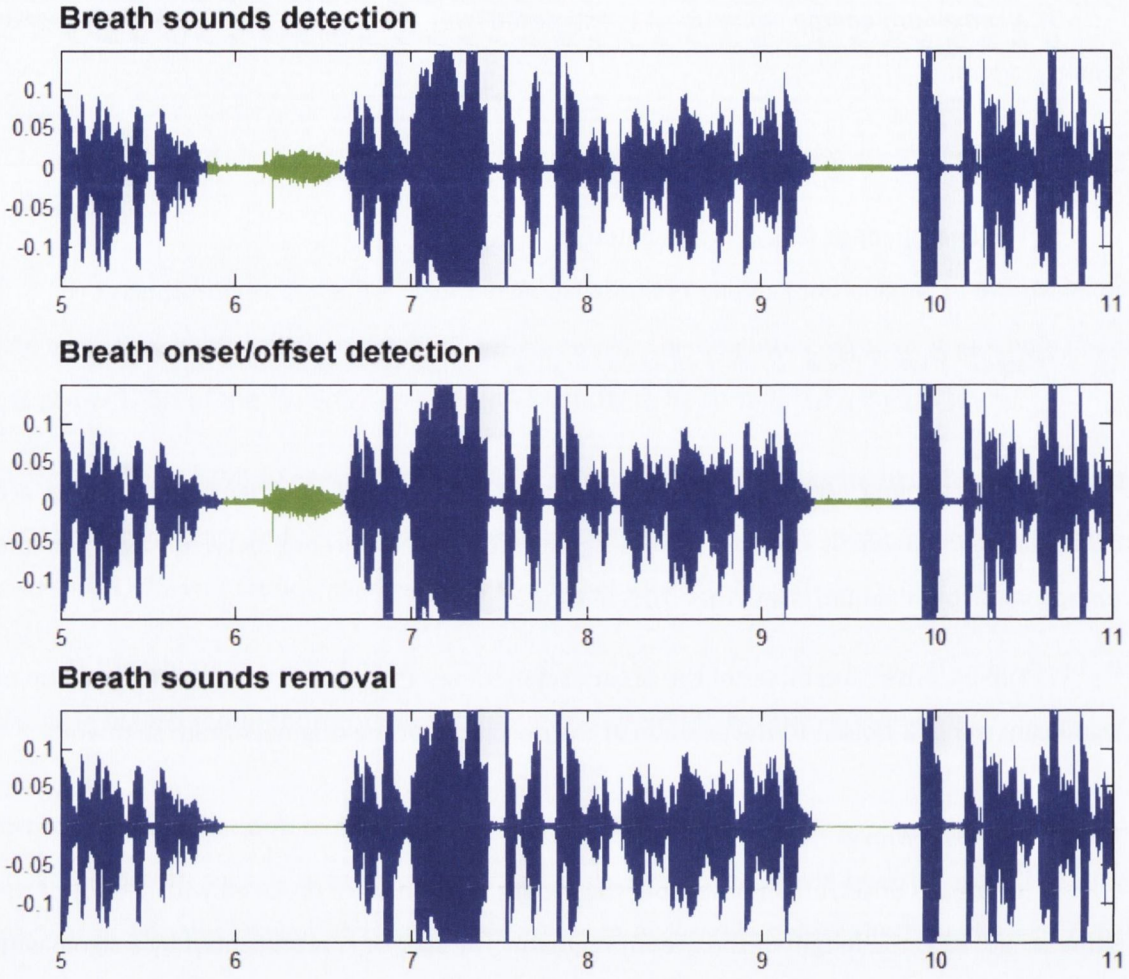


***Figure 3-2**: Breath sounds detection and removal steps.*

3.1.3 Extraction of temporal features

The observed changes in speech characteristics of individuals with impaired cognitive function were described in Sections 2.15 and 2.16. It was hypothesised that changes in cognitive function may be reflected in changes of temporal features of speech. With regard to the research hypothesis, it was deemed essential to extract temporal features of speech from all acquired recordings.

At the beginning of the feature extraction, all samples of the recording are classified as speech and non-speech samples in the first step. Samples above the *Speech/Non-speech Threshold* (Th_{SN}) are considered speech samples (labelled with value of 1); samples below the threshold are considered non-speech samples (labelled with value of -1).

$$SNP(n) = \begin{cases} 1, & abs(x(n)) \geq Th_{SN} \\ -1, & abs(x(n)) < Th_{SN} \end{cases}$$

Eq. 11

where ***SNP*** is the ***Speech/Non-speech Profile***, and *x(n)* is the original speech recording.

All non-speech segments shorter than 250ms are transformed to speech segments. Only non-speech segments of duration of 250ms or greater are considered pauses [273]. Anything shorter than this is considered a natural part of speech, where segments of low energy are a result of transitions from one word to another.

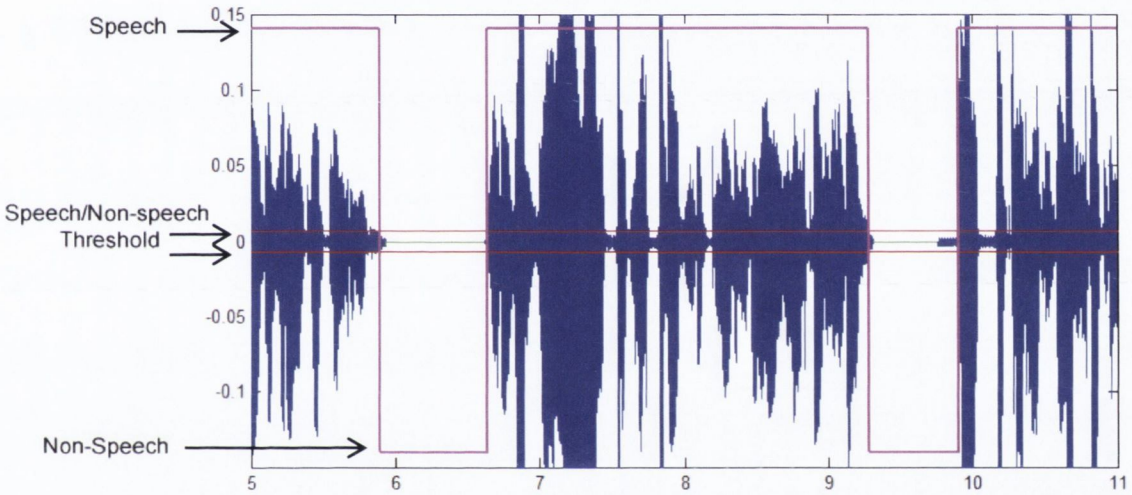


Figure 3-3*: Segmentation of the recording into speech and non-speech segments. Speech/Non-speech Profile represented by the red line.*

All speech segments shorter than 100ms were transformed to non-speech segments. All speech segments shorter than 100ms usually represents "clicks", "knocks" on the microphone, coughing, etc.

The result of this process is a segmentation of audio recordings into speech or non-speech parts. This process provides one of the important set of features for the investigation of cognitive function through speech.

Temporal features

Table 3-1 shows temporal features that were extracted for all recordings in studies in Chapters 4 – 6.

Table 3-1: Temporal features extracted from speech recordings

Temporal feature	Abbreviation	Units
Number of Pauses	NP	-
Mean Pause Duration	MPD	s
Proportion of Recording in Silence	PRS	-
Mean Utterance Duration	MUD	s
Total Recording Time	TRT	s
Total Length of Pauses	TLP	s
Total Length of Utterances	TLU	s

In Table 3-1, the following definitions apply

$$Mean\ Pause\ Duration = \frac{Total\ Length\ of\ Pauses}{Number\ of\ Pauses}$$

Eq. 12

$$Mean\ Utterance\ Duration = \frac{Total\ Length\ of\ Utterances}{Number\ of\ Utterances}$$

Eq. 13

Total Length of Pauses is the sum of durations of all pauses and *Total Length of Utterances* is the sum of durations of all utterances (i.e. not pauses) in the recording.

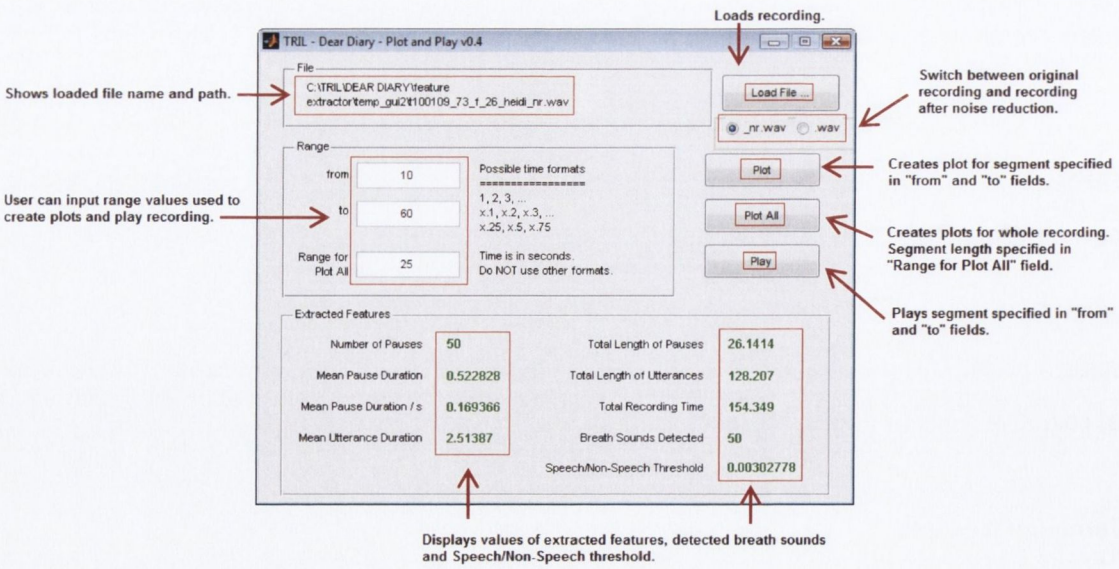


***Figure 3-4**: Graphical User Interface for displaying the results of segmentation into speech and non-speech segments.*

3.1.4 Extraction of energy and fundamental frequency features

Functional and morphological changes in areas of speech production and motor control in the brains of individuals with impaired cognitive function have been reported in Sections 2.11 and 2.12. Section 2.16 described qualitative changes in speech of individuals with impaired cognitive function. Considering these previous findings and changes in the brain areas of motor control that are responsible for the precise control of articulators, it was hypothesised that the energy and fundamental frequency features may be affected in individuals with impaired cognitive function (see Sections 2.14.1, 2.14.2, 2.16.4).

The following energy and fundamental frequency features were extracted from the speech recordings and provided another set of important features for the investigation of cognitive function through assessment of speech characteristics.

Energy features

The energy of a discrete time signal is defined in Eq. 2 (see Section 3.1.1).

Mean Energy per Second (***MES***) and ***Standard Deviation of Mean Energy per Second*** (***SDMES***) were estimated for all recordings. In order to take into account inter-speaker variance in the energy of the speech signal, the ***Coefficient of Variation of Mean Energy per Second*** (***CoVE***) was calculated.

$$CoVE = \frac{SDMES}{MES}$$

Eq. 14

Fundamental frequency estimation

During fundamental frequency (F_0) estimation, the signal was divided into 50ms frames and fundamental frequency was estimated in each of these frames. The fundamental frequency was estimated by finding the maximum amplitude of the signal cepstrum [271] within the range of 50 to 500Hz. ***Mean F_0*** (***MF_0***) was calculated by averageing F_0 estimates of all 50ms frames. Similarly, the ***Standard Deviation of F_0*** (***SDF_0***) was calculated from all 50ms frames. Finally, ***Coefficient of Variation of F_0*** (***$CoVF_0$***) was calculated.

$$CoVF_0 = \frac{SDF_0}{MF_0}$$

Eq. 15

Fundamental frequency amplitude

The amplitude of F_0 is the amplitude of the F_0 peak in the cepstrum. The ***Mean F_0 Amplitude (MF_0A)***, ***Standard Deviation of F_0 Amplitude (SDF_0A)*** and ***Coefficient of Variation of F_0 Amplitude ($CoVF_0A$)*** were calculated similarly as in the case of fundamental frequency features estimation.

3.2 Classification of subjects into cognitively healthy and cognitively impaired cases

To enable the use of temporal, energy and frequency speech features to be incorporated in a system to assess cognitive function, a pattern classifier was employed.

A Linear Discriminant Analysis (LDA) classifier [274] was employed to differentiate between cognitively healthy and cognitively impaired cases based on the distribution of extracted speech features. Cross-fold validation [275] was used to maximize training and determine classification accuracies.

In *k*-fold cross-validation, the dataset is randomly split into *k* mutually exclusive subsets (the folds) of equal size. In leave-one-out cross-validation, the classifier is trained employing all subsets, except one. After the training, the classifier's performance is tested employing the one subset not included in the training of the classifier. The training and testing is repeated *k*-times, every time leaving out different subset for testing. In this manner, the classifier is always tested employing a subset that has not been included in the training.

The variance of the performance estimates was decreased by averaging results from multiple runs of cross-validation where a different random split of the training data into folds is employed for each run. Number of repetitions and folds of cross-validation used to estimate classifier performance is specified for each study separately. For each run of cross fold validation the number of cognitively healthy and cognitively impaired cases was equal.

3.2.1 Classifier performance

Classifier performance was measured using sensitivity, specificity, positive predictivity, negative predictivity and the overall accuracy. These measures were calculated as per the definition of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) presented in Table 3-2.

Table 3-2: Definitions of true positives/negatives and false positives/negatives.

		True classification	
		Pathology	**Normal**
Predicted classification	**Pathology**	True Positive (TP)	False Positive (FP)
	Normal	False Negative (FN)	True Negative (TN)

Where performance metrics are defined as follows:

$$\textit{Overall Accuracy} = {}^{(TP+TN)}\!/_{(TP+TN+FP+FN)} \qquad \textit{Eq. 16}$$

$$\textit{Sensitivity} = {}^{TP}\!/_{(TP+FN)} \qquad \textit{Eq. 17}$$

$$\textit{Specificity} = {}^{TN}\!/_{(TN+FP)} \qquad \textit{Eq. 18}$$

$$\textit{Positive Predictivity} = {}^{TP}\!/_{(TP+FP)} \qquad \textit{Eq. 19}$$

$$\textit{Negative Predictivity} = {}^{TN}\!/_{(TN+FN)} \qquad \textit{Eq. 20}$$

3.2.2 Receiver operator characteristic (ROC) curves

Receiver operator characteristic (ROC) curves [276], which provide valuable information on the classifier's ability to discriminate between two classes over the complete spectrum of decision thresholds, were calculated. The ROC curve is a graph of sensitivity versus (100% - specificity), as the a-priori probabilities of the two classes are swept between zero and one. It provides information on clinical usefulness since it presents a trade-off in costs between false positives and false negatives and can be employed to decide the threshold for different clinical requirements e.g. screening vs. pre-surgical diagnosis. The *Area Under the ROC curve* (***AUC_ROC***) is a metric against which other classifier's configurations and/or features can be compared.

Key points

- Developed algorithm for fully-automated extraction of temporal features from speech recordings carried out in real-life environments.
- Developed algorithm for fully-automated breath sounds detection and removal from speech recordings carried out in real-life environments.
- Presented algorithms for energy and fundamental frequency features estimation.
- Presented method for classification to stratify subjects based on speech features.
- Presented method for estimation of classifier performance.

3.3 Summary

This chapter presented an algorithm that was developed for extraction of temporal features from speech recordings acquired in real-life environments. The algorithm included detection and removal of artefacts such as breath sounds. Methods that allow classification of subjects into healthy and impaired group based on their speech features were described. The developed algorithms and classification methods will be used in studies presented in the following chapters of this thesis.

The next chapter, Chapter 4, describes an investigation of speech characteristics in subjects with schizophrenia. Chapter 4 also provides assessment of the performance of the algorithm for extraction of temporal features from speech recordings. Investigation of the impact of the breath sounds detection and removal on the ability to discriminate between healthy and abnormal cases is presented as the last study of Chapter 4.

Chapter 4

Speech as an objective biomarker for schizophrenia

Cognitive impairment does not only affect older adults. Psychiatric conditions as discussed in Chapter 2, Section 2.16 can be accompanied by cognitive impairment. Included among these conditions is schizophrenia. The hypothesis with regard to schizophrenia is that speech characteristics of schizophrenics are different to those of healthy controls and these changes in speech characteristics would allow stratification of subjects into controls and schizophrenics. Building on methods developed in Chapter 3, the Study 1 of Chapter 4 investigated changes in speech characteristics of schizophrenics and healthy controls. Study 2 and Study 3 of Chapter 4 provide assessment of the performance of the algorithm described in Chapter 3. A number of publications have been derived from the research described in this chapter.

J. Thakore, **V. Rapcan**, S. D'Arcy, S. Yeap, N. Afzal, R.B. Reilly, "Acoustic and temporal analysis of speech: a potential marker for Schizophrenia", *International Clinical Psychopharmacology*, vol. 26, p. e131, 2011.

V. Rapcan, S. D'Arcy, S. Yeap, N. Afzal, J. Thakore, R.B. Reilly, "Acoustic and temporal analysis of speech: A potential biomarker for schizophrenia", *Medical Engineering & Physics*, vol. 32, pp. 1074-1079, 2010.

V. Rapcan, S. D'Arcy, and R. B. Reilly, "Speech analysis as a monitoring tool for schizophrenia", *Proceedings of the 16th Bioengineering in Ireland Conference*, Dublin, Ireland, January 22-24, 2010.

V. Rapcan, S. D'Arcy, and R. B. Reilly, "Automatic breath sound detection and removal for cognitive studies of speech and language," *Proceedings of the 20th IET Irish Signals and Systems Conference 2009*, 2009, pp. 1-6.

4.1 Study 1 – Investigation of the speech characteristics of individuals with schizophrenia

Schizophrenia is a serious mental illness characterized by a fundamental disturbance in perception, thought and communication [277]. Characteristic symptoms of schizophrenia are categorized into positive and negative symptoms. Positive symptoms, such delusions and hallucinations are those that appear to reflect an excess or distortion of normal functions [278]. Negative symptoms, such as loss of interest or asociality, are those that appear to reflect a diminution or loss of normal functions [279-281]. Disorganized speech, a positive symptom is characterized by tangential, loosely associated, or incoherent speech severe enough to substantially impair effective communication. Alogia, or poverty of speech, is a negative symptom characterized by lessening of speech fluency and productivity, thought to reflect slowing or blocked thoughts, and often manifested as short, 'empty' replies to questions. Another negative symptom, affective flattening, is the reduction in the range and intensity of emotional expression, including facial expression, voice tone, eye contact, and body language [282].

From a linguistic perspective, the duration of pauses and hesitations have been found to correlate strongly with the clinician's impressions of the patient's flat affect and alogia [265]. Alpert et al. [283] have shown that patients with this flat affect spoke with less inflection and have used less of their available time to describe recent experiences, i.e. were less fluent. Docherty et al. [284] have demonstrated that sustained attention impairment and impaired sequencing abilities (i.e. adequate words and phrases ordering for the communication of intended meanings), which are often found in schizophrenia, are highly predictive of communication failures related to language structure. Furthermore, Wisniecki et al. [270] have shown, using a simple counting and picture description task, that average pause length was indicative of motor retardation in those with negative symptom schizophrenia compared to controls. A number of different groups have used analysis of speech patterns so as to quantify the differences in length of pauses between those with mental illness and controls. These studies demonstrated that the length of pauses of the schizophrenic patients was longer in comparison with control subjects [285-287].

It has been demonstrated using MRI imaging, that structural changes post-onset can be found in the brains of those with schizophrenia [288-291] and that cognitive impairment affects up to 73% of schizophrenic patients [292]. Rapoport et al. [293] reported a 7% reduction of thalamus volume per year, in adolescents with schizophrenia. A similar volume reduction of the putamen and thalamus were reported in patients with probable Alzheimer's disease [58]. It was found that the decrease in volume of these deep grey matter structures correlated linearly with impaired global cognitive performance. The putamen, as a part of the striatum in the basal nuclei, connects with many other structures and pathways in the brain, forming a series of complex circuits between the cerebral cortex, basal nuclei and thalamus. One of the major circuits in this network is the motor circuit [294]. Significant reduction in grey matter in frontal areas in adolescent schizophrenic patients over a 3-5 year period was reported by Rapoport et al. [295]. The frontal lobes contain major motor control and speech production areas, including the primary motor cortex (Brodmann's area 4 (BA-4)), the premotor area (BA-6) and Broca's area (BA-44, BA-45), which are responsible for sequencing and controlling the motor movements required for the production of speech [294].

Two symptoms of schizophrenia in particular are associated with speech; alogia, which may reflect slowing or blocked thoughts, and cognitive impairment. The study presented here investigated temporal speech parameters such as the number, length of pauses and utterances as indicators of these symptoms. Affective flattening as a symptom of schizophrenia and structural changes in the brain of schizophrenics may be indicative of changes in functioning of articulators, hence this study investigated vocal pitch and energy parameters.

The aim of this study was to determine whether the speech characteristics of those with schizophrenia were significantly different from healthy controls, specifically to establish which components of speech can serve as a basis for objective analysis in schizophrenia.

4.1.1 Methods

Cohort

Informed consent was obtained from 39 (12 female) patients who fulfilled the Diagnostic and Statistical Manual of Mental Disorders IV criteria (DSM-IV) [6] using the Structured Clinical Interview for DSM [296] (Mean ± SD age = 42.3 ± 13.5 years) from the St. Vincent's Hospital Fairview catchment area in Dublin. The mean duration of illness (time from diagnosis) was 14.4 (± 11.2) years. All patients were physically healthy. Study exclusion criteria included

psychiatric diagnosis of co-morbid DSM IV [6] including alcohol or illicit drug abuse. The severity of illness was rated in all of the recruited patients, using the Brief Psychiatric Rating Scale (BPRS) [297] and the Scale for Assessment of Negative Symptoms (SANS) [254]. Patients were being prescribed a variety of different antipsychotic medications that have been converted to chlorpromazine (mg) equivalents in Table 4-1.

Control subjects comprised of 18 subjects (10 female) with a mean (± SD) age of 40.5 (± 12.9) years. All subjects were physically healthy and had no personal or family history of psychiatric illness and were recruited from within the local community. None of the control subjects were taking any form of prescribed or over the counter medication.

All schizophrenic and control subjects were literate. The study had approval from the Ethics Committee of St. Vincent's Hospital, Fairview. After complete description of the study to the subjects, written informed consent was obtained.

Table 4-1: Patient profile, medication and psychiatric rating.

	Female		Male	
	Mean	SD	Mean	SD
Number of subjects	12	-	27	-
Meds (CPZ) (mg)	818	347	708.4	346
In/Out Patient	2 in / 10 out	-	8 in / 19 out	-
BPRS	34.9	7.7	37.8	12.7
SANS	17.6	17.5	21.7	17.9
LOI (years)	18.4	12.4	12.6	10.4
Age	49.2	15.6	39.2	11.5

BPRS is the Brief Psychiatric Rating Scale, SANS is the Scale for Assessment of Negative Symptoms, CPZ is Chlorpromazine equivalents in mg, and LOI is Length of Illness in years.

Data acquisition

Each subject read aloud an extract from the children's story "Heidi" (see Appendix A.1). This passage is considered to be emotionally neutral and was particularly chosen for its verbal and semantic simplicity. The text passage has been previously employed in other cognitive studies [251]. The passage has a word count of 390 and, on average, takes 3-3.5 minutes to read aloud. Sample of the text is presented below and the full text can be found in Appendix A.1.

"The thing which attracted her most, however, was the waving and roaring of the three old fir trees on these windy days. She would run away repeatedly from whatever she might be doing, to listen to them, for nothing seemed so strange and wonderful to her as the deep mysterious sound in the tops of the trees. She would stand underneath them and look up, unable to tear herself away, looking and listening while they bowed and swayed and roared as the mighty wind rushed through them."

All audio files were recorded in a quiet room on a minidisc recorder (SONY MZ-B10) with direct digital 16-bit sampling and at a sampling rate of 22.05 kHz. Although every effort was taken to keep a constant distance during acquisition between the speaker and microphone, variation in acoustic amplitude exists between speakers due to individual variation in speaking volume.

Data processing

The temporal and acoustic features presented below were extracted from acquired speech recordings in this study. See Chapter 3, Section 3.1 for description of the feature extraction process.

Temporal features

- *Number of Pauses*
- *Mean Pause Duration*
- *Proportion of Recording in Silence*
- *Mean Utterance Duration*
- *Total Recording Time*
- *Total Length of Pauses*
- *Total Length of Utterances*

Acoustic features

- *Mean Energy per Second (MES)*
- *Standard Deviation of Mean Energy per Second (SDMES)*
- *Coefficient of Variation of Mean Energy per Second (CoVE)*
- *Mean F_0 (MF0)*
- *Standard Deviation of F_0 (SDF_0)*
- *Coefficient of Variation of F_0 ($CoVF_0$)*

Classification and statistical analysis

A two-tailed, unequal sample sizes, unequal variance, Student's t-test was employed to determine if there are statistically significant differences between the features of control and schizophrenic group. A one-tailed Fisher's F-test was employed to determine if correlation exists between the extracted speech features and the clinical variables – *Dosages of Medication*, symptom scales (*BPRS, SANS*), and *Length of Illness*.

Analysis of the histograms of all features extracted from all recordings for schizophrenic and control group showed an approximate Gaussian distribution. Therefore, a classifier based on Linear Discriminant Analysis (LDA) was chosen to differentiate between the control and schizophrenic subjects (see Chapter 3, Section 3.2). In this study, 39 repetitions of eighteen-fold cross-validation were used to estimate classifier performance. For each run of cross fold validation the number of normal and abnormal cases was equal.

4.1.2 Results

Feature values

Table 4-2 displays the average feature values for the patient and control groups. A clear difference in these values can be observed, with some features showing larger differences. The *Number of Pauses*, *Proportion of Recording in Silence* and the *Total Length of Pauses* were found to increase by 27%, 23% and 40% respectively for the patient group compared to the control group. The variations in their distributions are clearly smaller for the control group.

Table 4-2: Mean feature values (and standard deviations) for schizophrenic and control subjects.

	Schizophrenics		Controls	
	Mean	SD	Mean	SD
Number of Pauses (s)	68.72	19.77	54.05	8.94
Mean Pause Duration (s)	0.63	0.08	0.57	0.08
Proportion of Recording in Silence	0.19	0.04	0.16	0.02
Mean Utterance Duration (s)	2.69	0.63	3.02	0.48
Total Recording Time (s)	222.15	42.80	194.58	30.64
Total Length of Pauses (s)	42.80	12.42	30.56	6.10
Total Length of Utterances (s)	179.35	35.78	164.02	25.46
CoVE	0.67	0.21	0.58	0.13
$CoVF_0$	0.22	0.05	0.24	0.08

CoVE - Coefficient of Variation of Mean Energy per Second, $CoVF_0$ - Coefficient of Variation of Fundamental Frequency

Employing the Student's *t*-test, the following features were found to be statistically significant - *Number of Pauses* ($t(55)=3.87$, $p<0.0003$), *Mean Pause Duration* ($t(35)=2.62$, $p<0.01$), *Proportion of Recording in Silence* ($t(54)=4.81$, $p<0.00001$), *Mean Utterance Duration* ($t(43)=2.15$, $p<0.04$), *Total Recording Time* ($t(45)=2.77$, $p<0.008$), *Total Length of Pauses* ($t(55)=4.99$, $p<0.000007$), *Coefficient of Variation of Mean Energy per Second* ($t(49)=2.12$, $p<0.04$). The only features not achieving statistical significance were the *Total Length of Utterances* ($t(45)=1.85$, $p<0.07$) and the *Coefficient of Variation of Fundamental Frequency* ($t(24)=0.88$, $p<0.4$).

Classification results

To assess the ability of the LDA classifier to distinguish between schizophrenic patients and controls, the LDA classifier was trained using three feature sets. At first, the LDA classifier was trained on a combination of all the features listed in Table 4-2, and the performance of this classifier may be seen in Table 4-3, Feature Set 1.

Table 4-3: Linear Discriminant Analysis (LDA) classifier performance metrics.

	Feature Set 1	Feature Set 2	Feature Set 3
Sensitivity (%)	72.64	75.21	72.36
Specificity (%)	78.63	83.62	85.47
Positive predictivity (%)	77.86	82.02	83.39
Negative predictivity (%)	74.24	77.40	75.75
Overall accuracy (%)	75.64	79.42	78.92
AUC_ROC	0.79	0.82	0.80

AUC_ROC *- Area under the ROC curve,*
Feature Set 1 *- all features combined,*
Feature Set 2 *- Feature Set 1 minus Total Length of Utterances and Coefficient of Variation of Fundamental Frequency,*
Feature Set 3 *- Number of Pauses, Proportion of Recording in Silence, Total Recording Time and Total Length of Pauses*

The training and testing was repeated for each individual feature set and certain features were found to result in better classification accuracy than others. Predictably, the two feature sets which were found not to have significantly different data distributions resulted in the poorest classification performance. The classification result using *Total Length of Utterances* and *Mean Utterance Duration* was 60% and 62% respectively, compared to 75% and 77% classification for *Number of Pauses* and *Proportion of Recording in Silence*.

The combined feature set was then reduced removing *Total Length of Utterances* and *Coefficient of Variation of Fundamental Frequency*. The results of the LDA training/testing may be seen in the Table 4-3, Feature Set 2. Further reduction in the features set removed the next poorest performing features; *Mean Pause Duration* and *Mean Utterance Duration* and *Coefficient of Variation of Mean Energy per Second* (see Table 4-3, Feature Set 3).

The classification performance when training on the two reduced feature sets outperform the original classifier for all classifier metrics, however between the two reduced feature set the results are not as clear. While the Feature Set 2 (only omitting *Total Utterance Duration* and *Coefficient of Variation of Fundamental Frequency*) performs better than the Feature Set 3 (additionally removing *Mean Pause Duration* and *Mean Utterance Duration* and *Coefficient of Variation of Mean Energy per Second*) for nearly all classification results except Specificity and Positive Predictivity. The higher specificity suggests more controls being classified correctly while the slightly higher positive predictability suggests a higher likelihood of classifying a patient correctly.

Correlation analysis

Table 4-4 lists the correlation coefficients between extracted features and clinical variables - *Dosages of Medication*, symptom scales (*BPRS, SANS*), and *Length of Illness*.

Table 4-4: Correlation coefficients between features, symptom scales, dosages of medication and length of illness.

	Correl. with Meds	Correl. with BPRS	Correl. with SAND	Correl. with LOI
Number of Pauses	0.02	-0.19	-0.03	0.11
Mean Pause Duration	0.31	0.10	-0.01	0.04
Proportion of Recording in Silence	0.33[a]	-0.34[a]	-0.35[a]	0.43[b]
Mean Utterance Duration	-0.22	0.46b	0.37[a]	-0.46[b]
Total Recording Time	-0.13	0.25	0.38[a]	-0.28
Total Length of Pauses	0.18	-0.14	-0.02	0.11
Total Length of Utterances	0.22	0.35[a]	0.46[b]	-0.38[a]
CoVE	-0.34[a]	0.37[a]	0.37[a]	-0.44[b]
$CoVF_0$	-0.15	0.07	0.02	-0.18

CoVE - Coefficient of Variation of Mean Energy per Second, $CoVF_0$ - Coefficient of Variation of Fundamental Frequency

[a] *Statistically significant at p=0.05*

[b] *Statistically significant at p=0.01*

Employing the Fisher's F-test to determine if these correlation coefficients are statistically significant yielded following results. Each of the clinical variables has a correlation of statistical significance ($p<0.05$) with at least one of the temporal features. Highest correlation coefficients, which were statistically significant at $p<0.01$, were between *Length of Illness* and *Proportion of Recording in Silence* ($F(1,35)=8.33$), *Mean Utterance Duration* ($F(1,35)=9.74$) and *Relative Variation in Energy* ($F(1,35)=8.74$), further between *SANS* and *Total Length of Utterances* ($F(1,35)=10.08$) and also between *BPRS* and *Mean Utterance Duration* ($F(1,35)=9.90$) (see Table 4-4).

4.1.3 Discussion

From classification results of the two reduced feature sets it is clear that the pause related features were most significant in differentiating between schizophrenic and control group. Classification was carried out on each feature set individually and it was the *Number of Pauses*, *Proportion of Recording in Silence* and *Total Length of Pauses* that had the highest overall accuracy for classification. On average, patients with schizophrenia tend to insert more pauses in their read speech (+27%, see Table 4-2), which causes the *Proportion of Recording in Silence* and *Total Length of Pauses* to be higher for patients with schizophrenia than control subjects. *Mean Pause Duration* is also slightly greater (+10%, see Table 4-2) for patients with schizophrenia. This result confirms the hypothesis that the speech characteristics of patients with schizophrenia contain more pauses and longer duration of pauses.

These differences in temporal features of speech may be related to the cognitive impairment which affects people with schizophrenia [292]. Cognitive impairment is associated with psychomotor retardation and has deteriorating effects on the production of speech. Saykin et al. [298] compared two schizophrenic patients groups - first episode patients and previously treated patients, with healthy control group. Patients groups had nearly identical profiles showing generalized impairment, particularly in verbal memory and learning attention-vigilance, and speeded visual-motor processing and attention.

In terms of the changes in the brain areas of schizophrenic patients and the negative symptom, affective flattening, which is associated with the speech of schizophrenics, it was hypothesized that a lack of variation in energy and vocal pitch would be observed in speech of schizophrenics (see Section 2.16.4). Of the remaining features; utterance related features and energy/pitch variation, it is the *Coefficient of Variation of Mean Energy per Second* that had the highest success rate in classification. Given the variation in individual speaker volume, the

absolute values for *Mean Energy per Second* and *Standard Deviation of Mean Energy per Second* are not valid for inter-subject comparison. Relative measures of these features were calculated and *Coefficient of Variation of Mean Energy per Second* results in a classification accuracy of 63.3%. The distribution of the *Coefficient of Variation of Fundamental Frequency* was found not to be statistically significant (p=0.4) and was not a useful measure in differentiating between the two groups. No significant difference was found between speech parameters of in-patients and out-patients.

The ability to discriminate between healthy controls and schizophrenic subjects of our study was lower than the discriminating performance of 95.2% shown in the study of Stassen et al. [251]. This may be a result of Stassen's training of the discriminant analysis classifier. The classifier was trained on recordings from 84 subjects (42 schizophrenics, 42 controls). The same 84 subjects were re-recorded 14 days later and the speech parameters from these recordings tested on the trained classifier. Stassen et al. also demonstrated that speech parameters for healthy subjects [299] and schizophrenic patients [251] remain stable over the period of 14 days and are highly correlated. Using the same subjects in the training and test samples, and taking into account the stability of speech parameters over time, an increased accuracy of the discriminant analysis classifier may be expected. Also, Stassen et al. recorded data in an acoustically shielded room. The recordings reported in this study were carried out in a quiet room, more akin to what one might expect in real world clinical settings. Nevertheless, a good discrimination accuracy of 79.4% was achieved. Therefore, the results from the study presented herein demonstrate that the patients' voice characteristics can differentiate from those of healthy subjects.

The effect of antipsychotic medication on speech characteristics in schizophrenic subjects is a source of concern when interpreting the study results. To assess the effect of medication on speech additional scale scores reflecting the side-effects of medication are required. Stassen et al. [251] reported no significant relationship between most of the acoustic variables and measures of antipsychotic medication side-effects. Qualitative changes in speech characteristics are included in the diagnosis of first episode schizophrenics in the National Institute of Health - Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) [6]. During the initial psychiatric assessment, the first-episode schizophrenics are not taking antipsychotic medication. Therefore, it may be implied that the differences in speech characteristics reported in this study reflects the psychiatric condition rather than the effects of the medication. Further research is required to confirm this assumption.

A larger database of speech samples from schizophrenic patients displaying a range of symptoms, as well as samples from healthy controls would contribute to a more robust classification system and may possibly show better correlation between acoustic features and dosages of medication, symptom scales and/or length of illness.

Using speech for assessment of schizophrenia may enable development of remote and fully automated system. Studies by Moran et al. [300] and Wormald et al. [301] presented automatic assessment systems for monitoring vocal fold pathologies. In this way, high quality assessments may be carried out remotely, resulting in the use of available medical resources in an efficient and effective manner, as well as providing psychiatrists with up to date results with minimal processing. Monitoring of the effects of medication on cognitive functioning of schizophrenic patients using regular voice recordings may provide the psychiatrist with valuable information about the suitability of the prescribed medication.

Key points

- The *Number of Pauses, Proportion of Recording in Silence* and the *Total Length of Pauses* were found to increase by 27%, 23% and 40% respectively for the schizophrenic group compared to the control group.
- The speech of patients with schizophrenia contains more and longer pauses.
- Statistically significant differences were found in mean values between the patient group and control group for all speech features, except the *Total Length of Utterances* and the *Coefficient of Variation of Fundamental Frequency*.
- Ability to discriminate between schizophrenics and healthy controls of the LDA classifier trained with all speech features achieved classification accuracy of 75.6% (Sensitivity of 72.6%, Specificity of 78.6%)
- Training LDA classifier with the Feature Set 2 yielded classification accuracy of 79.4% (75.2%, 83.6%)
- Training LDA classifier with the Feature Set 3 yielded classification accuracy of 78.9% (72.4%, 85.5%)
- The pause related features were most significant in differentiating between schizophrenic and control group.
- Each of the clinical variables has a statistically significant correlation ($p<0.05$) with at least one of the temporal speech features.

4.2 Study 2 - Assessment of the performance of the algorithm for extraction of temporal features

In order to provide validation of the algorithm developed for automated extraction of temporal speech features from speech recordings described in Chapter 3, Section 3.1, the assessment of performance of this algorithm was carried out to estimate the accuracy of the algorithm. To achieve correct speech feature values that correspond with reality, the algorithm must correctly identify pauses and speech segments in the recording. This study focused on estimation of the ability of the feature extraction algorithm to correctly identify pauses and speech segments in the recording. Three tests were performed to assess the performance of the algorithm.

4.2.1 Methods

Artificially constructed speech recording

A speech file was artificially constructed by concatenating speech, silence and breath segments, each of 500ms duration to one 180s file. The file consisted of 120 pauses (which included 120 breath sounds) and 120 speech segments. The total length of pauses in the constructed file was 120s and the total length of speech segments was 60s. The file was split into 10ms long non-overlapping frames of speech and pauses, and the performance of the feature extraction algorithm to correctly identify speech and pause frames was estimated.

Three speech recordings hand-labelled by three listeners

To provide more realistic assessment of the performance of the automated feature extraction algorithm, as a second test, three recordings were randomly selected from all available recordings described in Section 4.1.1. In each of these recordings, pauses (which includes breath sounds) and speech frames (10ms long, non-overlapping) were independently hand-labelled by three experienced listeners. The listeners hand-labelled three recordings with total length of 6.9 minutes. The recordings contained 118 pauses, which included 110 breath sounds. The performance of the feature extraction algorithm to correctly identify speech and pause frames was estimated for these files for every listener, to calculate if there are significant differences in hand-labelling between all three listeners. This step was required to find out whether in the next test the recordings have to be hand-labelled with more listeners or just with one.

Larger corpora hand-labelled by one listener

To obtain a more accurate estimate of the performance of the feature extraction, 20 randomly selected recordings from the corpus described in Section 4.1.1 were hand-labelled by one listener (Listener 1) for speech and pause frames (10ms long, non-overlapping). The total length of all 20 recordings was 46.5 minutes and contained 872 pauses which included 634 breath sounds. Average performance of the feature extraction algorithm was estimated.

Assessment of the performance of the feature extraction algorithm

Performance was measured using sensitivity, specificity, positive predictivity, negative predictivity and overall accuracy as defined in Eq. 16 - Eq. 20. The true/false positives and true/false negatives were defined as shown in Table 4-5.

Table 4-5: True/false positives and true/false negatives definition for assessment of the performance of the feature extraction algorithm

		Listener identification	
		Pause	**Speech**
Algorithm identification	**Pause**	True Positive (TP)	False Positive (FP)
	Speech	False Negative (FN)	True Negative (TN)

Sensitivity is described by some authors [302] as "*non-speech hit rate HR0*" and specificity as "*speech hit rate HR1*". The non-speech hit rate is defined as

$$HR0 = \frac{N_{0,0}}{N_0^{ref}}$$

Eq. 21

where $N_{0,0}$ is the number of correctly classified non-speech frames, and N_0^{ref} is the number of real non-speech frames in the recording. And the speech hit rate is defined as

$$HR1 = \frac{N_{1,1}}{N_1^{ref}}$$

Eq. 22

where $N_{1,1}$ is the number of correctly classified speech frames, and N_1^{ref} is the number of real speech frames in the recording.

These two measures provide information about the proportion of correctly classified pause (speech) frames over all real pause (speech) frames. To obtain a better estimate of the performance of the feature extraction, in this study, the overall accuracy, positive and negative predictivity were also calculated, where positive (negative) predictivity provides information about the proportion of correctly identified pause (speech) frames over all identified pause (speech) frames.

4.2.2 Results

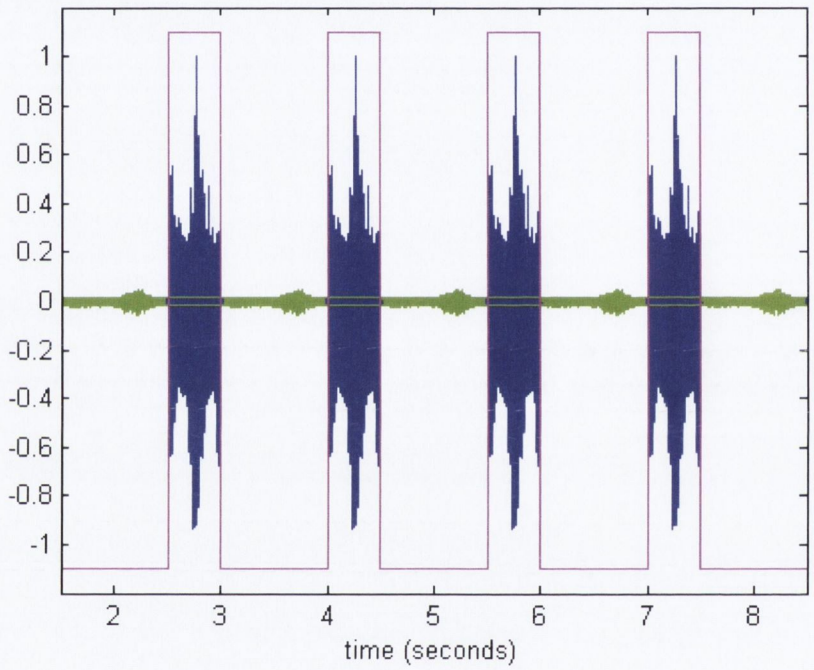


Figure 4-1: *Artificially constructed speech recording processed with proposed feature extraction algorithm.*
Pauses that include breath sounds are displayed in green colour, speech segments in blue colour. Final decision of the algorithm is represented by the red line.

Artificially constructed speech recording

The feature extraction algorithm performed very well on the artificially constructed speech recording and as expected identified all speech and pause frames correctly (see Figure 4-1). Following performance metrics were achieved:

- Overall accuracy: 100%
- Sensitivity: 100%
- Specificity: 100%
- Positive predictivity: 100%
- Negative predictivity: 100%

Three speech recordings hand-labelled by three listeners

Differences between hand-labelling performances of all three listeners were small (SD of accuracy = 0.40%, SD of sensitivity = 1.36%, SD of specificity = 0.51%). The segmentation of the recordings into pause and speech frames corresponded closely with that of the feature extraction algorithm. Results can be seen in Table 4-6 and Table 4-7.

Table 4-6: Combined performance metrics from three recordings and three listeners

	Average	SD
Overall accuracy (%)	97.84	0.40
Sensitivity (%)	96.40	1.36
Specificity (%)	98.03	0.51
Positive predictivity (%)	90.95	2.33
Negative predictivity (%)	99.36	0.26

Table 4-7: Combined performance metrics of the three recordings for each listener

	Listener 1 (avg/SD)	Listener 2 (avg/SD)	Listener 3 (avg/SD)
Overall accuracy (%)	97.61 (0.30)	97.82 (0.56)	98.09 (0.33)
Sensitivity (%)	94.98 (5.60)	96.67 (3.42)	97.55 (2.62)
Specificity (%)	98.05 (1.09)	97.93 (0.88)	98.11 (0.50)
Positive predictivity (%)	90.99 (4.67)	90.67 (0.93)	91.19 (0.72)
Negative predictivity (%)	99.08 (0.81)	99.42 (0.44)	99.57 (0.36)

Average values for all three listeners with standard deviations in parentheses.

Larger corpora hand-labelled by one listener

Given the close agreement in hand-labelling of the three listeners, Listener 1 hand-labelled another 20 recordings to obtain more accurate estimate of the feature extraction algorithm performance. Following results were achieved:

- Overall accuracy: 97.29%
- Sensitivity: 93.52%
- Specificity: 98.37%
- Positive predictivity: 92.92%
- Negative predictivity: 98.23%

4.2.3 Discussion

To evaluate the performance of the developed algorithm, three tasks were performed. At first, the feature extraction algorithm was tested against artificially created recording, which contained 120 pauses (which included breath sounds) and 120 speech segments of known length. The algorithm identified all speech and pause frames correctly.

As a second test, an agreement in hand-labelling of the recordings between different listeners was investigated. This was needed to find out whether in the next task the recordings have to be hand-labelled with more listeners or just with one listener. Very close agreement of all listeners was achieved with standard deviation of 0.40% in accuracy, 1.36% in sensitivity and 0.51% in specificity. Therefore, it was decided that, in the next task, hand-labelling of the recordings with just one listener will provide data with enough accuracy to estimate the performance of the feature extraction algorithm.

To obtain a better estimate of the performance of the feature extraction algorithm, Listener 1 hand-labelled additional 20 recordings and the average performance from all these recordings was calculated. Very good performance of the feature extraction algorithm was achieved with overall accuracy of 97.29%, sensitivity of 93.52% and specificity of 98.37%. Therefore, the feature values from automated extraction of temporal features should closely correspond with real values.

Key points

- The developed temporal feature extraction algorithm identified all speech and pause frames correctly (100% performance) for the artificially constructed speech recording.
- The developed temporal feature extraction algorithm achieved very good performance in speech and pause identification using 20 real speech recordings: overall accuracy of 97.29%, sensitivity of 93.52% and specificity of 98.37%.

4.3 Study 3 - Impact of the breath sounds detection and removal on the ability to discriminate between cognitively healthy and cognitively impaired cases

A spoken utterance contains not only words that are produced to make syntactic and semantic sense, but also many insertions of non-speech artefacts. Non-speech artefacts can be coughs, splutters and breath sounds.

The hypothesis of this study is that pauses are employed by speakers to deal with the cognitive load of reading aloud. The subject may attempt to fill pauses with breath sounds to manage the cognitive load imposed on them by the current task. Separation of breath sounds segments from speech segments in a recording will result in change of temporal features values. The aim of this study was to assess if this change would have an impact on ability of an LDA classifier to discriminate between schizophrenics and healthy subjects.

4.3.1 Methods

The audio corpora employed in this study contained 57 recordings of reading out loud of the text passage described in Chapter 4, Section 4.1.1 and Appendix A.1.

The temporal features presented below were extracted from acquired speech recordings in this study. See Chapter 3, Section 3.1 for description of the feature extraction process.

- *Number of Pauses*
- *Mean Pause Duration*
- *Proportion of Recording in Silence*
- *Mean Utterance Duration*
- *Total Recording Time*
- *Total Length of Pauses*
- *Total Length of Utterances*

The temporal features were extracted twice for each recording:

1. Without breath sounds detection and removal
2. With breath sounds detection and removal

Classification

A Linear Discriminant Analysis (LDA) classifier (see Chapter 3, Section 3.2) was employed to differentiate between control and schizophrenic cases. In this study, 39 repetitions of 18-fold cross-validation were used to estimate the classifier performance. For each run of cross fold validation the number of cognitively healthy and cognitively impaired cases was equal.

To evaluate influence of breath sounds detection and removal on discrimination ability between schizophrenics and control subject, LDA performance was estimated for features extracted with and without breath sounds detection and removal.

4.3.2 Results

The Table 4-8 shows the performance of the LDA classifier trained and tested employing the temporal features extracted without breath sounds detection and removal, and employing the temporal features extracted while employing breath sounds detection and removal.

Table 4-8: Impact of the breath sounds detection and removal on the LDA classifier performance

	Without breath sounds detection and removal	With breath sounds detection and removal	Difference
Sensitivity (%)	55.7	69.8	14.1
Specificity (%)	79.3	78.5	-0.8
Positive predictivity (%)	74.3	77.6	3.3
Negative predictivity (%)	64.0	71.8	7.8
Overall accuracy (%)	67.5	74.2	6.7
AUC_ROC	0.68	0.75	0.07

AUC_ROC - Area under the ROC curve

The difference in average values of temporal features extracted from recordings without and with prior breath sounds detection and removal, from all 57 recordings, can be seen in Table 4-9.

Table 4-9: Differences between average values of extracted features with/without prior breath sounds detection and removal

	Without breath sounds detection and removal	With breath sounds detection and removal
Number of Pauses	38.14	64.09
Mean Pause Duration (s)	0.50	0.61
Proportion of Recording in Silence	0.09	0.18
Mean Utterance Duration (s)	5.23	2.79
Total Length of Pauses (s)	19.06	38.94
Total Length of Utterances (s)	194.39	174.51

4.3.3 Discussion

The removal of the breath sounds has been shown to be an important factor in the analysis of temporal features of read speech. Non-speech segments of duration of 250ms or greater were considered pauses [273]. Anything shorter than this is considered a natural part of speech, where segments of low energy are a result of transitions from one word to another. Without breath detection short segments of non-speech, which lead into breath sounds, would be interpreted as speech. With breath detection a pause is identified in such a situation, thus it can be expected the number and duration of pauses to increase, as observed in Table 4-9. This increase in the number of pauses correspondingly has an effect on the mean utterance duration and total length of utterances.

The removal of breath sounds resulted in a new set of temporal features. To evaluate the effect of breath removal on the ability of the LDA classifier to discriminate between schizophrenics and control subjects, classifier performance was estimated for features extracted without prior breath sounds detection and removal, and features extracted with prior breath sounds detection and removal.

An increase in all performance metrics was achieved, except "Specificity" which slightly decreased by 0.8% from 79.3% to 78.5%. This decrease means an increase in false positive detections (control subjects classified as schizophrenics) and decrease in true negative detections (control subjects classified as controls) of the classifier. The overall accuracy of the LDA classifier increased by 6.7%, from 67.5% to 74.2%. The area under the ROC curve increased by 0.07, from 0.68 to 0.75. An increase in true positive detections (schizophrenics classified as schizophrenics) and decrease in false negative detections (schizophrenics classified as controls) resulted in highest increase in performance from 55.7% to 69.8%

(+14.1%) that was observed for Sensitivity. The positive (negative) predictivity increased by 3.3% (7.8%) from 74.3% (64%) to 77.6% (71.8%).

The hypothesis of this study is that pauses are employed by speakers to deal with the cognitive load of reading aloud. An increase in the number of pauses indicates that the speakers need more time to process and fulfil the task in front of them (in this case a reading task). The duration of pauses is an indication of the cognitive load being experienced by the subject. While breath sounds may be considered a normal part of speech, they may also be elongated by subjects attempting to fill pauses while trying to manage the cognitive load imposed on them by the current task. It is for this reason that classing breaths as pauses leads to a more discriminative feature for analysis.

Significant cognitive impairment is a common attribute of schizophrenia, affecting up to 75% of patients [303]. Reading tasks offer a very interesting method of assessment of cognitive impairment for this psychiatric condition. The text passage used in this study contains structural pauses, usually dictated by commas and full stops, which are cues to the reader to help separate distinct ideas. Additional pauses, hesitation pauses, are indicative of a person's inability to maintain the thread of a sentence or phrase for its duration. These pauses can often be filled with breath sounds, and typically breaths used to 'fill pauses' are longer than structural pauses. Including breath sounds as non-speech artefact increases the discriminatory nature of the classifier to provide assessment of cognitive impairment.

Key points

- Employing the breath sounds detection and removal, an increase in all performance metrics of the LDA classifier was achieved, except for specificity.
- Employing the breath sounds detection and removal, the overall accuracy of the LDA classifier increased by 6.7%, from 67.5% to 74.2%.
- Employing the breath sounds detection and removal, the area under the ROC curve increased by 0.07, from 0.68 to 0.75.
- Employing the breath sounds detection and removal, the sensitivity of the LDA classifier increased by 14.1%, from 55.7% to 69.8%.
- The results of this study suggest that the hypothesis that subjects may fill a pause with a breath to manage the cognitive load imposed on them by the current task is correct.

4.4 Summary

The studies in this chapter answered the research questions posed in Chapter 2, Section 2.18.2 and 2.18.1.

Study 1 of this chapter demonstrated that the temporal and energy features of speech correlate with cognitive function of individuals with/without schizophrenia. The results suggest the potential of speech characteristics in the assessment and monitoring of schizophrenia. The results suggest a potential for employing the speech features in complement with current gold-standard neuropsychological assessment of schizophrenia. The speech characteristics may represent an objective biomarker for schizophrenia, while allowing for remote monitoring, providing psychiatrists with up to date results and insight into the suitability of prescribed medication.

Study 2 investigated the performance of the developed algorithm for temporal features extraction. Very good performance of the feature extraction algorithm was demonstrated, proving the reliability of the feature extraction algorithm. This result proves the accuracy of the speech features employed in the studies of this thesis.

Study 3 demonstrated the importance of breath sounds detection and removal prior the speech feature extraction in studies of cognitive function. The results of this study suggest that the hypothesis that subjects may fill a pause with a breath to manage the cognitive load imposed on them by the current task is correct. Therefore, it is important to employ breath sounds detection and removal in studies of speech characteristics in cognitive function.

The next chapter describes investigation of speech characteristics of cognitively impaired and healthy older adults.

Chapter 5

Speech characteristics in ageing

This chapter describes the investigation into speech characteristics of older adults in relation to their cognitive function. As presented in Chapters 1 and 2, the number of people over 60 years is rapidly growing in the whole world [64]. The biggest limiting factor to independence in the older adults is impaired cognitive function and its consequences [64].

The current manner of identifying cognitive impairment using clinical neurocognitive batteries of standard neuropsychological tests (see Chapter 2, Section 2.9) can be very time consuming, and represent substantial labour and monetary burden. With the increasing population of older adults, the labour, temporal and monetary burden will only increase.

Besides these issues most of the standard neuropsychological tests lack objectivity. This is due to the tests being administered by different clinicians and some tests are affected by practice, age, education.

There is a clear need for the development of novel methods of cognitive function assessment that will address the limitations of current cognitive assessment methods and also meet the future demands for cognitive assessment of older adults. In particular, the novel methods should provide means of objective assessment of cognitive function. Ideally, these new methods would lower the cost and time required to undertake cognitive assessment, essentially allowing the clinicians more effective use of their time and resources.

Chapter 2, Section 2.12 reviewed research focused on the morphological and functional changes in the brains of individuals with cognitive impairment. Previous research into the speech characteristics of individuals in relation to the status of their cognitive function was also presented in Chapter 2, Section 2.15. This chapter presents four studies investigating

speech characteristics of older adults as a potential objective and quantitative biomarker of cognitive function. A number of publications have been derived from the research described in this chapter.

I. Rochford, **V. Rapcan**, S. D'Arcy, R.B. Reilly, "Dynamic Minimum Pause Threshold Estimation for Speech Analysis in Studies of Cognitive Function in Ageing", *Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp. 3700-3703.

N. Burke, **V. Rapcan**, A. Galli, S. D'Arcy, R.B. Reilly, I.H. Robertson, "Improving the sensitivity of the animal naming category fluency test using temporal speech features", *Proceedings of the 8th Annual Psychology, Health and Medicine Conference*, Galway, April 4th, 2011.

A. Galli, C. Garattini, N. Burke, **V. Rapcan**, S. D'Arcy, R.B. Reilly, I.H. Robertson, "Cognitive function monitoring using temporal speech features", *Proceedings of the International Association of Gerontology and Geriatrics-VII European Congress*, Bologna, April 2011.

N. Pénard, S. Counihan, S. D'Arcy, **V. Rapcan**, R.B. Reilly, I.H. Robertson, "Investigating Speech as a Source of Biomarkers for Changes in Cognition, Executive Function and Mood", *The 47th Annual Meeting of the Academy of Aphasia*, Boston, Massachusetts, October 18-20, 2009.

S. D'Arcy, N. Penard, S. Counihan, **V. Rapcan**, I.H. Robertson, R.B. Reilly, "Corpora for assessment of cognitive function in older persons", *Proceedings of the 3rd Advanced Voice Function Assessment International Workshop*, Madrid, Spain, 18-20 May 2009.

S. D'Arcy, **V. Rapcan**, N. Pénard, M.E. Morris, R.B. Reilly, I.H. Robertson, "Speech as a Means of Monitoring Cognitive Function of Elderly Subjects", *Proceedings of Interspeech 2008*, Brisbane, Australia, 22-26 September 2008.

N. Pénard, S. D'Arcy, **V. Rapcan**, P. Dockree, C. Besani, K. O'Sullivan, R.B. Reilly, I.H. Robertson, "Investigating Speech as a Diagnostic Tool for Abnormal Ageing", *Proceedings of the 2008 British Psychological Society Annual Conference*, Dublin, April 2008.

5.1 Study 1 – Speech as an objective measure of cognitive function in older adults

As discussed in Chapter 2, Sections 2.10 and 2.11, there are limitations in the current gold-standard neuropsychological assessment of cognitive function in older adults. Following Chapter 2, Section 2.15 discussed the influence of impaired cognitive function on speech production of older adults. This study aimed to investigate quantitative speech characteristics changes between older adults with intact and with impaired cognitive function and the power of the features in classification of cognitively intact and impaired older adults.

5.1.1 Methods

Cohort

Participants in this study were from a cohort of 208 older adults. The age ranged from 60 to 93 years of age. The subjects were recruited from the Technology Research for Independent Living (TRIL) Clinic in St. James's Hospital, Dublin, Ireland [304]. After complete description of the study to the subjects by a trained research staff, written informed consent was obtained. The study was approved by the St. James's Hospital Ethics Committee, Dublin.

During their attendance at the TRIL Clinic, subjects undertook a battery of cognitive tests and were recorded while reading and completing various tasks, i.e. generating spontaneous speech.

Personal information relevant to the study was collected.

- Gender
- Handedness
- Age
- Existing medical condition
- Consumption of alcohol per week
- If prone to falls
- If a smoker
- Number of years of education

However, it is known that gender, age, smoking, and medical conditions affect individual speech characteristics. While years of education, IQ, cognitive function and social interaction intuitively have implications for speech and language fluency, their significance has yet to be

quantitatively assessed. The characteristics may or may not be contributing factors in cognitive decline but it was considered important to gather this information.

Due to the fact that subjects with severe cognitive impairment cannot give consent, those subjects were not included in the corpus. This resulted in an uneven distribution of the Mini-Mental State Examination (MMSE) scores.

Exclusion criteria were severe hearing impairment, severe reading impairment, Alzheimer's disease, Parkinson's disease, stroke and/or transient ischemic attack, epilepsy, dyslexia, major psychiatric illness and psychiatric medication.

This study analysed data from a sample of 172 subjects from the cohort of 208 subjects. 36 failed to complete one or more of the assigned tasks. The age range of the sample was from 60 to 93 years. The average age is 72.4 years (SD 7 years); 62.5% female and 37.5% male[6]. Table 5-1 contains the age and Mini-Mental State Examination (MMSE) profile of the corpus. The cohort was split into two groups, *MMSE low* and *MMSE high*, according to MMSE score. Subjects that achieved an MMSE score of 26 and lower were considered cognitively impaired, i.e. *MMSE low* group. Those that achieved an MMSE score of 27-30 were considered cognitive healthy, i.e. *MMSE high* group. From the 172 subjects, 37 were classed as cognitively impaired (*MMSE low* group) and 135 were classed as cognitively healthy (*MMSE high* group).

Table 5-1: Distribution of the cohort

	Age (years)						MMSE (total score)						
	60-64	65-69	70-74	75-79	80-84	85+	30	29	28	27	26	25	24-
All	25	40	33	53	11	10	24	48	36	27	20	7	11
Male	8	19	15	19	4	5	8	25	13	9	9	1	6
Female	17	21	18	34	7	5	16	23	23	18	11	6	5

Numbers represent participant count for each condition (age group, MMSE score)

Battery of neurocognitive tests

All participants of the cohort completed a battery of standard neuropsychological tests that was administered by a trained clinical psychologist. The battery included the following tests:

[6] Gender imbalance caused by lower number of male participants deciding to participate in the study.

- Mini-Mental State Examination,
- National Adult Reading Test,
- Word List Recall,
- Category Fluency,
- Digit Span,
- Center for Epidemiologic Studies Depression scale,
- Hospital Anxiety and Depression Scale.

The details of these tests are described in Chapter 2, Section 2.9. These tests require the neuropsychologist to provide a detailed explanation of the task to the participant. They also require the neuropsychologist to score the tests manually, typically using pen and paper.

Speech tasks

In addition to the neurocognitive test battery, each participant was recorded performing three speech tasks. The neuropsychologist provided vocal instructions for these tasks and remained beside the subject throughout all tasks.

Reading aloud a Text Passage

The first task was a controlled speech task; each subject read aloud an extract from the children's story "Heidi". This passage is considered to be emotionally neutral and was particularly chosen for its verbal and semantic simplicity. The text passage has been previously employed in other cognitive studies [251]. The passage has a word count of 390 and, on average, takes 3-3.5 minutes to read aloud. A sample of the text is presented below, with the full text can be found in Appendix A.1.

"The thing which attracted her most, however, was the waving and roaring of the three old fir trees on these windy days. She would run away repeatedly from whatever she might be doing, to listen to them, for nothing seemed so strange and wonderful to her as the deep mysterious sound in the tops of the trees. She would stand underneath them and look up, unable to tear herself away, looking and listening while they bowed and swayed and roared as the mighty wind rushed through them."

Picture Taboo task

A *Picture Taboo (PT)* task was designed to specifically target the executive function domain of the cognitive function. Executive function refers to a set of cognitive abilities generally associated with the frontal lobes. These functions are responsible for monitoring and supervising more automated behaviours. Sustained attention, inhibitory control and planning are examples of executive functions. This task consisted of a set of five pictures, each displayed on a computer screen with a pair of words (see sample in Figure 5-1, other samples can be seen in Appendix A.3). The pair of words are the 'Taboo' words. These words are major themes of the images and the subject had 60 seconds to describe the image without using these Taboo words. The time was controlled with a stopwatch and a beep-sound marks the beginning and the end of the task. This task is designed to tap into executive function by forcing the subject to inhibit the use of the taboo words while searching for alternative descriptor words. The subject's speech was recorded for the duration of 60 seconds irrespective of whether the subject was talking or not.

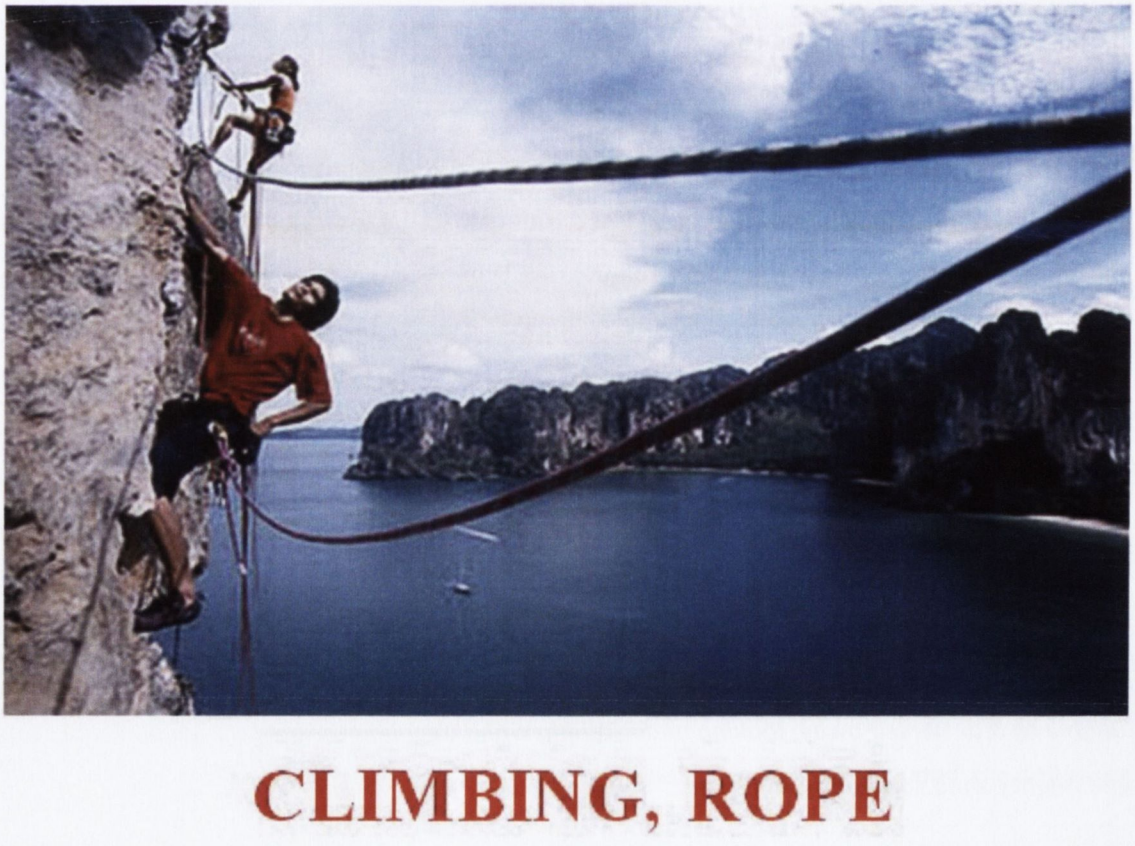


***Figure 5-1**: Sample picture of the Picture Taboo task.*

Data acquisition

A headset microphone was used to record speech from all participants; this was connected to an external USB sound card connected to a PC. All efforts were made to keep the microphone at the same distance from the mouth and to reduce background noise to a minimum. Speech was recorded at 44.1 kHz sampling frequency and 16 bit resolution.

The participants' speech was recorded for the *Text Passage* and *Picture Taboo (PT)* tasks described above.

Data processing

With regard to the structural and functional changes in regions of brain involved in speech production and the qualitative changes in speech characteristics reported in subjects with impaired cognitive function, it was deemed important to extract specific features from speech recordings.

The temporal and acoustic features presented below were extracted from acquired speech recordings in this study. See Chapter 3, Section 3.1 for description of the feature extraction process.

Temporal features

- *Number of Pauses*
- *Mean Pause Duration*
- *Proportion of Recording in Silence*
- *Mean Utterance Duration*
- *Total Recording Time*
- *Total Length of Pauses*
- *Total Length of Utterances*

Acoustic features

- *Mean Energy per Second (MES)*
- *Standard Deviation of Mean Energy per Second (SDMES)*
- *Coefficient of Variation of Mean Energy per Second (CoVE)*
- *Mean F_0 (MF0)*
- *Standard Deviation of F_0 (SDF_0)*
- *Coefficient of Variation of F_0 ($CoVF_0$)*

Phoneme Rate features

Education is known to affect ones performance on cognitive tasks and is also known to affect reading ability [305]. In this study, the feature set extracted from the *Text* Passage was transformed in terms of phone rate. *Phoneme Rate* is defined as the number of phones in the text passage divided by the *Total Length of Utterances*. The number of phones in the text passage was calculated from the phonetic transcription of the text passage. Dividing the absolute temporal speech features by the *Phoneme Rate (PR)* allowed to normalize the variations in reading rate between participants [306]. The *Standardized Pause Ratio* is the number of words in the *Text Passage* (390) divided by the *Number of Pauses*. This can be considered a measure of fluency, how many words can be maintained between pauses. The following set of features was generated.

- *Phoneme Rate (PR)*
- *Number of Pauses / PR*
- *Mean Pause Duration / PR*
- *Proportion of Recording in Silence / PR*
- *Standardized Pause Ratio*
- *Mean Utterance Duration / PR*
- *Total Pause Duration/ PR*
- *Total Utterance Duration / PR*
- *Total Recording Time / PR*

Classification and statistical analysis

In order to better understand the contribution each feature makes to the overall difference in speech between the cognitively impaired and healthy participants, Pearson's *r* correlation analysis was calculated between the speech features and the *Mini-Mental State Examination (MMSE)* group. Additionally, to investigate if any of the speech features targeted specific cognitive function domains, correlation analysis was carried out between the speech features and specific cognitive task scores. The cognitive tasks included *Word Recall Immediate (WRI), Word Recall Delayed (WRD), Category Fluency (CF), Digit Span (DS)* task.

A Linear Discriminant Analysis (LDA) classifier was employed to differentiate between healthy and cognitively impaired participants, based on the speech features presented above. See Chapter 3, Section 3.2 for details regarding the classifier performance estimation.

The LDA classifier was designed to classify classes of the same samples sizes. The cognitively impaired group contained 37 participants, while the cognitively healthy group contained 135 participants. Due to the difference in the numbers of participants in the healthy and cognitively impaired group and in order to ensure there was no effect of group, a repetitive classification protocol was employed.

A group of 37 participants was randomly selected from the 135 participants of the cognitively healthy group. Speech features values of these 37 cognitively healthy participants and of the 37 cognitively impaired participants were employed in the estimation of the LDA classification performance.

The random selection from the cognitively healthy group was repeated 135 times. Each time the estimation of the LDA classification performance was also repeated. The overall classification performance was calculated as the average of the 135 iterations.

5.1.2 Results

Figure 5-2 presents a comparison between the average values of *Proportion of Recording in Silence* extracted from read speech (*Text Passage*), from read speech normalized to phoneme rate (*Text Passage/PR*) and from spontaneous speech (*Picture Taboo*).

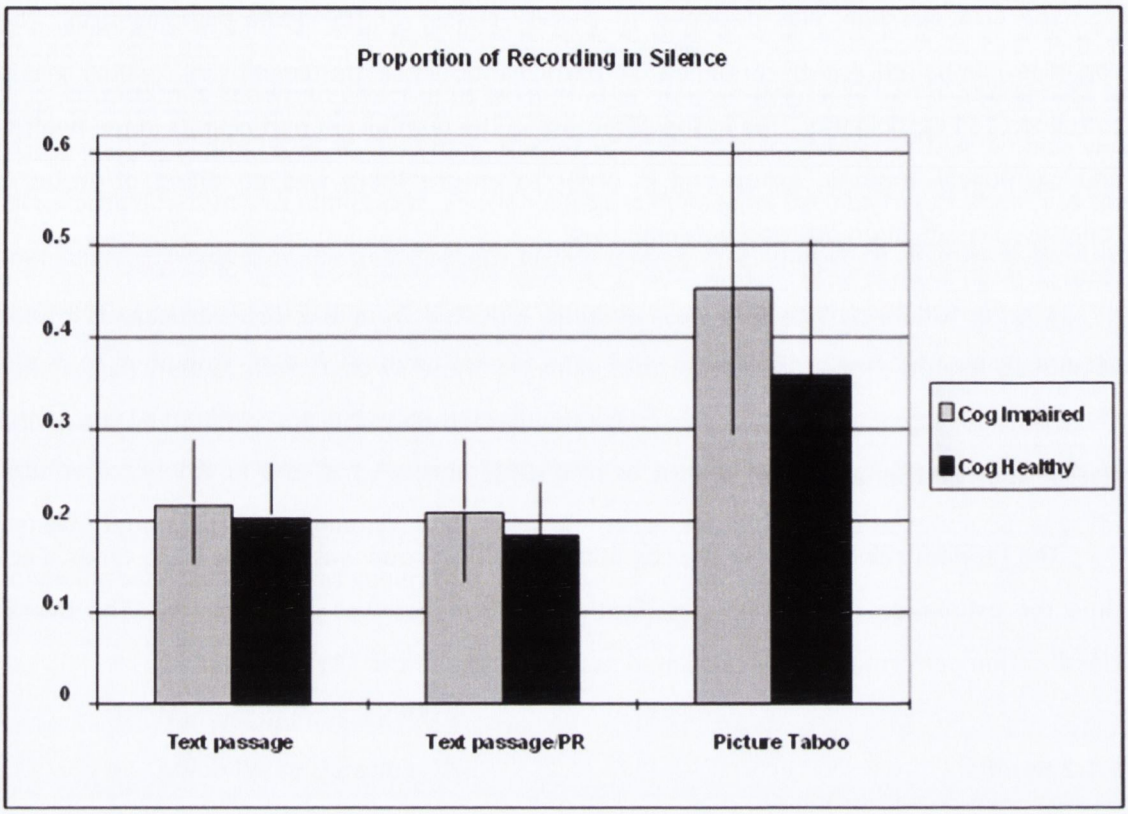


***Figure 5-2**: Comparison of average feature values for Proportion of Recording in Silence*

Text passage

Table 5-2 presents the correlation of each speech feature from the *Text Passage* with the MMSE group. In addition, the table presents correlations between speech features and specific cognitive tasks - *Word Recall Immediate (WRI), Word Recall Delayed (WRD), Category Fluency (CF), Digit Span (DS)*. Statistical significance is noted it the table.

Table 5-2: Correlation analysis between speech features of the Text Passage task and the MMSE group, and speech features and the cognitive tasks scores

Feature	MMSE Group	Word Recall Immediate	Word Recall Delayed	Category Fluency	Digit Span
Number of Pauses	-0.126*	-0.279**	-0.196*	-0.235**	-0.108
Mean Pause Duration	-0.088	-0.220**	-0.180*	-0.024	-0.028
Proportion of Recording in Silence	-0.102	-0.267**	-0.219**	-0.118	-0.071
Mean Utterance Duration	0.125	0.203**	0.159*	0.167*	0.081
Total Length of Pauses	-0.138*	-0.311**	-0.226**	-0.175*	-0.089
Total Length of Utterances	-0.133	-0.176*	-0.110	-0.228**	-0.047
Total Recording Time	-0.151*	-0.291**	-0.200**	-0.246**	-0.081
Phoneme Rate	0.161*	0.171*	0.093	0.235**	0.040
Standardized Pause Ratio	0.143*	0.258**	0.192*	0.262**	0.087
Number of Pauses / PR	-0.142*	-0.289**	-0.197*	-0.264**	-0.110
Mean Pause Duration / PR	-0.138*	-0.266**	-0.200**	-0.142	-0.052
Proportion of Recording in Silence / PR	-0.153*	-0.321**	-0.242**	-0.204**	-0.093
Mean Utterance Duration / PR	0.057	0.095	0.079	0.030	0.054
Total Length of Pauses / PR	-0.157*	-0.322**	-0.228**	-0.212**	-0.097
Total Length of Utterances / PR	-0.104	-0.168*	-0.111	-0.215**	-0.044
Total Recording Time / PR	-0.135*	-0.244**	-0.167*	-0.238**	-0.069
CoV Mean Energy	0.062	-0.024	0.005	0.080	0.071
CoV F0	0.121	-0.040	-0.061	-0.004	-0.019
CoV F0 Amplitude	0.120	0.207**	0.226**	0.023	0.152*

*Pearson's r correlation values. *Speech feature correlates with cognitive measure at p<0.01, ** Speech feature correlates with cognitive measure at p<0.001*

Picture taboo

Table 5-3 presents the correlation of each speech feature from the *Picture Taboo* task with the MMSE group. In addition, the table presents correlations between speech features and specific cognitive tasks - *Word Recall Immediate (WRI), Word Recall Delayed (WRD), Category Fluency (CF), Digit Span (DS).*

Table 5-3: Correlation analysis between speech features of the Picture Taboo task and the MMSE group, and speech features and the cognitive tasks scores

Feature	MMSE Group	Word Recall Immediate	Word Recall Delayed	Category Fluency	Digit Span
Number of Pauses	0.043	0.008	-0.065	-0.068	0.040
Mean Pause Duration	-0.218**	-0.345**	-0.251**	-0.388**	-0.261**
Proportion of Recording in Silence	-0.224**	-0.336**	-0.267**	-0.400**	0.255**
Mean Utterance Duration	0.173*	0.215**	0.202**	0.314**	0.160*
Total Length of Pauses	-0.226**	-0.335**	-0.269**	-0.403**	-0.250**
Total Length of Utterances	0.221**	0.342**	0.269**	0.394**	0.260**
CoV Mean Energy	-0.150*	-0.180*	-0.273**	-0.188*	-0.126
CoV F0	-0.044	0.107	0.063	0.142	0.070
CoV F0 Amplitude	0.105	0.213**	0.210**	0.040	0.074

*Pearson's r correlation values. *Speech feature correlates with cognitive measure at p<0.01, ** Speech feature correlates with cognitive measure at p<0.001*

Classification results

Classification performance was estimated for various combinations of the data sets. At first, the LDA classifier was trained employing the *Cognitive Battery* dataset. This dataset contained the scores of the cognitive tasks - *Word Recall Immediate (WRI), Word Recall Delayed (WRD), Category Fluency (CF), Digit Span (DS)*. The LDA classification performance was employed to assess the power of these cognitive tasks to discriminate between cognitively healthy and cognitively impaired older adults as determined by the MMSE score.

The remaining five datasets included speech features from the speech tasks separately and for the speech tasks combined. The estimation of the ability of the LDA classifier to discriminate between the cognitively impaired (*MMSE low*) and cognitively healthy (*MMSE high*) groups was repeated for each dataset to identify the optimal feature set for classification. The classification performance for feature combinations was investigated to assess if any particular type of features is significantly more discriminative than others.

The classification performance results are presented in Table 5-4 in terms of accuracy and the *Area Under the ROC Curve* (*AUC_ROC*) along with their standard deviations.

Table 5-4: LDA classification results for cognitive and speech features

Feature set	Accuracy (%)	(SD)	AUC_ROC	(SD)
Cognitive battery	73.5	(3.23)	0.77	(0.03)
All features from the Text Passage	68.9	(4.16)	0.75	(0.04)
All features from the Picture Taboo task	68.0	(4.40)	0.74	(0.04)
Acoustic features of both tasks	63.2	(4.20)	0.68	(0.04)
Temporal features of both tasks	72.2	(3.70)	0.78	(0.03)
Combing all features from both tasks	80.4	(3.70)	0.87	(0.02)

5.1.3 Discussion

Initial interpretation of the results investigated the correlation between individual speech features and overall cognitive score as well as specific cognitive domains.

Text passage

Correlation analysis between the speech features of the *Text Passage* and the cognitive tasks measures shown that there is some degree of correlation between the *Temporal features* and the memory and executive function tasks (i.e. *Word Recall – Immediate and Delayed, Category Fluency)*. Three of the *Temporal features* achieved statistically significant Pearson's *r* correlation with the MMSE group. These features are the *Number of Pauses, Total Length of Pauses* and *Total Recording Time*. While the correlations for these features are statistically significant, the correlations coefficients are small. All *Temporal features* achieved statistically significant correlation with the *Word Recall Immediate* measure. All *Temporal features*, except the *Total Length of Utterances*, achieved statistically significant correlation with the *Word Recall Delayed* measure. All *Temporal features*, except the *Mean Pause Duration* and *Proportion of Recording in Silence*, achieved statistically significant correlation with the *Category Fluency* measure. None of the *Temporal features* achieved statistically significant correlation with the *Digit Span* measure.

All *Phoneme Rate features*, except the *Mean Utterance Duration/PR* and *Total Length of Utterances/PR*, achieved statistically significant correlation with the MMSE group. Statistically significant correlation was observed between all *Phoneme Rate features* and the *Word Recall Immediate* measure, except the *Mean Utterance Duration/PR*. All *Phoneme Rate features*, except the *Phoneme Rate, Mean Utterance Duration/PR* and *Total Length of Utterances/PR*, achieved statistically significant correlation with the *Word Recall Delayed* measure. All *Phoneme Rate features*, except the *Mean Pause Duration/PR* and *Mean Utterance Duration/PR*, achieved statistically significant correlation with the *Category Fluency* measure.

As in the case of *Temporal features*, none of the *Phoneme Rate features* achieved statistically significant correlation with the *Digit Span* measure.

During the correlation analysis of the *Acoustic features* with the cognitive tasks measures, only the *Coefficient of Variation of F_0 Amplitude* achieved statistically significant correlations with the *Word Recall Immediate, Word Recall Delayed* and *Digit Span* measures.

The higher number of *Phone Rate features* correlating with the MMSE group, than was the number of *Temporal features,* suggests that it is important to compensate for reading rate in studies of speech characteristics in cognitive impairment. Although, the correlation was small, similar to that of *Temporal features.* Stronger and more significant correlation of the speech features with the *Word Recall Immediate, Word Recall Delayed* and *Category Fluency* measures was achieved than with the correlation with *MMSE group.* This suggests that the speech features potentially reflect specific cognitive domains, rather the overall cognitive function. The cognitive domains in this case were short- and long-term memory and word retrieval processes.

Picture Taboo

All *Temporal features*, except the *Number of Pauses*, achieved statistically significant correlation with the MMSE group, as well as with all cognitive tasks measures. *Number of Pauses* did not achieve statistically significant correlation with any of the cognitive measures, nor with the MMSE group.

From the *Acoustic features*, the *Coefficient of Variation of Mean Energy per Second* achieved statistically significant correlation with the MMSE group and all cognitive tasks measures, except the *Digit Span*. The *Coefficient of Variation of F_0* did not achieve statistically significant correlation with any of the cognitive measures, neither the MMSE group. The *Coefficient of Variation of F_0 Amplitude* achieved statistically significant correlation with *Word Recall Immediate* and *Word Recall Delayed* measures.

The correlation between the speech features from the *Picture Taboo* task and the measures of cognitive tasks is generally stronger than in the case of *Text passage*. The statistically significant correlation of the *Temporal* features with the *Digit Span* measure suggests that maintaining the taboo words in short term memory and manipulating responses around these words coincide with the working memory aspect of the *Digit Span* task.

While the correlation coefficients in Table 5-2 and Table 5-3 are significant, they are low. The variability within each feature is highlighted by the large standard deviations observed in the analysis, see Figure 5-2. Inter-speaker comparisons are always subject to high variability due to the inherent variability of speech. The features investigated in this study may be more appropriate for intra-speaker comparisons, due to the lower variability experienced in intra-speaker speech features. The features extracted automatically in this study may be suitable as a means of monitoring ones cognitive function via speech over a long period of term as opposed to a once off classification.

The analysis of the correlation for each feature highlights the need to take into account reading ability when extracting speech features from a reading task. This analysis also informs us about the characteristics of speech that are important in terms of MMSE and may provide insight for refining these features in future research.

Classification performance

The correlation analysis found several features of speech to be significantly correlated to measures of cognitive function. The next phase was to investigate if these features can be reliably employed to classify subjects in terms of their cognitive function.

Table 5-4 presents classification results for six datasets derived from the available data. The classification results illustrate that the feature sets from *Text Passage* task and *Picture Taboo* task achieved similar classification results, 69% and 68% respectively. However, combining the *Temporal features*, from both tasks, achieved higher classification accuracy when discriminating between cognitively healthy and impaired groups than the combined acoustic features, 72.2 % and 63.2% respectively. The classification results found that combining the *Temporal* features from both tasks achieved similar results to the classification results when employing the *Cognitive Battery.* It is the combination of speech features from both tasks (*Text Passage* and *Picture Taboo*) that achieves the highest classification results in this study (80.4%). This is a significant improvement on either speech task or feature type individually. Similar neural processes are activated for the production of speech for both *Text Passage* and *Picture Taboo* tasks. However, distinct processes (short-term memory vs. visual processing of text) are also recruited during these two discrete speech tasks. The significant improvement in classification accuracy achieved when features from the two speech tasks are combined suggests that both tasks are important in the assessment of overall cognitive function.

The classification results from the study by Roark et at [60] and the results presented in this study both demonstrate the effectiveness of employing features of speech to classify participants in terms of cognitive health. Similar baseline classifications in terms of a cognitive battery were achieved by Roark et al and the study presented here, *AUC_ROC* of 0.82 and 0.77 respectively. The battery of tests for the baseline of both studies targeted similar brain function, i.e. working memory, verbal fluency, and short and medium term memory. The study by Roark et al. [60] investigated features of speech and language to classify participants' class according to their Clinical Dementia Rating (CDR) [307] and found characteristics of language to be the most robust features for classifying participants on their data set. The difference is that the study presented here focused on temporal features of speech that are readily extracted automatically from recorded speech.

One significant difference in feature definition between the study presented here and other studies, which have not observed an increase in classification performance from speech features [60], is the minimum duration of non-speech that is considered a pause. Roark et al. [60] set the minimum pause duration to 1000ms. Setting the minimum pause duration to 250ms (as in this study) may make this analysis more sensitive to inserted pauses that represent pauses required for individuals with cognitive impairment and may result in improved classification using temporal features of speech.

Comparing the *Cognitive battery* classifier with speech features based classifiers indicates that there is a higher discriminative ability for the speech based classifiers to assess cognitive status.

In addition, the cost involved in administering the cognitive battery and calculating the cognitive scores is much greater than recording and automatic extraction of speech features. This balance of information derived and cost supports the use of speech for the purpose of differentiating between healthy and cognitively impaired older adults.

The results presented in this study demonstrate the effectiveness of employing features of speech to classify participants in terms of their MMSE group. While the MMSE is a common clinical tool and an outcome measure employed in many research studies, the score of MMSE is influenced by factors, such as age, education, ethnicity (see Chapter 2, Section 2.10). The exact interpretation of answers may vary between neuropsychologists administering the assessment resulting in a level of subjectivity in scoring of the MMSE. Automating the

extraction of speech features and analysis minimizes any variability that may occur in classification.

One often cited criticism of the MMSE is the lack of agreement on the threshold for cognitive impairment on the 30-point scale [186], as described in Chapter 2, Section 2.10. Despite these limitations, the MMSE is one of the most frequently used brief assessments of cognitive function. This study has demonstrated a difference in the measures of certain speech features between cognitively healthy and cognitively impaired participants, as determined by a dichotomized MMSE score. With this relationship established monitoring of these individual speech features may prove more indicative of a change in cognitive function than a single point reduction in an individual's MMSE score. A floor and ceiling effects of repeatedly (annually) administering the MMSE were found by Galasko et al [308]. An assessment incorporating speech based tasks can be administered repeatedly as there is no learning effect associated with speech generation and the array of speech based tasks from which to choose is large.

Conclusion

This study has established the potential of temporal and acoustic speech features as a means of monitoring cognitive health. Speech as a biomarker for cognitive health may complement current screening tools of cognitive function in older adults.

Key points

- The temporal and acoustic features automatically extracted from speech can be employed to classify older persons in terms of their cognitive status.
- *Temporal features* of speech of the *Text Passage* task correlate with *MMSE*, *Word Recall - Immediate & Delayed*, and the *Category Fluency* tasks measures.
- *Temporal features* of speech of the *Picture Taboo* task correlate with *MMSE*, *Word Recall - Immediate & Delayed*, *Category Fluency* and *Digit Span* tasks measures.
- The higher number of *Phoneme Rate features* correlating with the MMSE group, than was the number of *Temporal features,* suggests that it is important to compensate for reading rate in studies of speech characteristics in cognitive impairment.
- Training the LDA classifier with all speech features of the *Text Passage* task yielded higher accuracy (68.9%) in discriminating cognitively impaired and healthy subjects than the LDA classifier trained with all speech features of the *Picture Taboo* task achieved the accuracy of 68.0%.
- Training the LDA classifier with *Temporal features* of speech yielded higher accuracy (72.2%) in discriminating cognitively impaired and healthy subjects than was the accuracy (63.2%) of the LDA classifier trained with acoustic features of speech.
- Training the LDA classifier with combination of the *Text Passage* and the *Picture Taboo* task features yielded accuracy of 80.4% in discriminating cognitively impaired and healthy subjects.

5.2 Study 2 – Addressing the limitations of the Mini-Mental State Examination

The limitations of the MMSE as a stand-alone test for cognitive function have been discussed in Chapter 2, Section 2.10. In an attempt to generate a more objective measure of cognitive function with which to relate speech features to, a study to investigate the inclusion of additional cognitive scores to provide a broader cognitive profile of the subject and to correlate this new profile with speech derived characteristics was carried out.

Previous research in the literature analysed several combinations of cognitive tests for discriminating between expertly diagnosed Dementia patients and healthy controls. Jacqmin-Gadda et al [309] found the most discriminant combination of tests to be the Category Fluency (animals) task, MMSE and visual retention task. In the study presented herein, three additional tests were combined with the Category Fluency task and the MMSE, to produce a Combined Cognitive Score (CCS).

The aim of this study was to investigate if there are significant differences in speech features of older adults categorized into cognitively healthy and impaired groups based on their MMSE score and based on the CCS score. The hypothesis was that the CCS would provide more objective estimate of overall subject's cognitive function than MMSE. To test this hypothesis, the ability to discriminate between cognitively healthy and impaired older adults of an LDA classifier employing speech features was assessed separately for the CCS categorization of subjects and the MMSE categorization of subjects.

5.2.1 Methods

Cohort

This study was conducted on a large cohort of 208 subjects from 60 to 93 years of age. The average age is 72.4 years (standard deviation is 7 years); 62.5% female and 37.5% male. Seventeen subjects had MMSE scores of 24 or less (8%), 62 subjects had an MMSE score of between 25 and 27 (30%) and the remaining 106 (62%) had an MMSE score that are considered to be in the normal range (28-30). Table 5-5 contains the age, MMSE and education profile of the corpus. Of the 208, 13 subjects were omitted from the final analysis due to excess noise during the recording or failure to complete one of the tasks.

Table 5-5: Demographic information of the cohort

	Age (years)						MMSE (total score)							Education (years)				
	60-64	65-69	70-74	75-79	80-84	85+	30	29	28	27	26	25	24-	1	2	3	4	5
All	26	47	54	55	13	13	33	49	47	32	20	10	17	70	41	53	33	11
Male	8	20	21	19	5	5	9	26	16	10	8	2	7	37	12	13	10	6
Female	18	27	33	36	8	8	24	23	31	22	12	8	10	33	29	40	23	5

Numbers represent participant count for each condition (age group, MMSE score, years of education).

Combined Cognitive Score

Cognitive tasks contained in the original cognitive battery were chosen for use in designing a *Combined Cognitive Score (CCS)*. These tasks are *Word Recall Immediate* and *Delayed, Digit Span, Category Fluency* (see Chapter 2, Section 2.9 for the full details on each task). While *MMSE* provides measure of overall cognitive function, these tasks provide measures of specific domains of cognitive function. These tasks are more objective in their scoring protocols than the *MMSE*.

The *Combined Cognitive Score (CCS)* was calculated by assessing whether subject achieves the score within expected 'normal' ranges for each cognitive task. Each subject's *MMSE* score was compared to the expected norm for their age and education from the study published by Crum et al. [185]. If this score lay within one standard deviation of its expected *MMSE* score one point was given. An additional point was scored if the subject recalled four or more words for the *Word Recall - Immediate* task, three or more words for the *Word Recall - Delayed* task [310], named at least 14 animals for the *Category Fluency* task [311] and scored at least five out of 14 for the *Digit Span* task [312]. The highest possible *CCS* score was 5.

The subjects' cohort was divided into two splits. At first, the cohort was split based on the *MMSE* score into '*MMSE low*' group (MMSE < 27) and '*MMSE high*' group (MMSE > 27). The second split was based on the *CCS* score. '*CCS low*' group with *CCS* score of 1 or 2. And '*CCS high*' group with *CCS* score of 3 to 5.

Data acquisition

The subjects carried out two tasks: *Text Passage* and *Picture Taboo*. During *Text Passage* task, the subjects read aloud a short text passage. During the *Picture* Taboo task, the subjects described a picture without using two taboo words. See Section 5.1.1 of the previous study for details regarding the speech tasks. The speech from each subject carrying out these speech based tasks was recorded.

A headset microphone was used to record speech from all participants; this was connected to an external USB sound card connected to a PC. All efforts were made to keep the microphone at the same distance from the mouth and to reduce background noise to a minimum. Speech was recorded at 44.1 kHz sampling frequency and 16 bit resolution.

Data processing

Following temporal features were extracted from the acquired speech recordings in this study. See Chapter 3, Section 3.1 for description of the speech features extraction process.

- *Number of Pauses*
- *Mean Pause Duration*
- *Proportion of Recording in Silence*
- *Mean Utterance Duration*
- *Total Recording Time*
- *Total Length of Pauses*
- *Total Length of Utterances*

Classification and statistical analysis

Student's t-test was employed to investigate the difference in distributions of the features between the cognitively healthy and impaired groups. This analysis was performed separately for each categorization method of cognitive function.

A Linear Discriminant Analysis (LDA) classifier was employed to differentiate between healthy and cognitively impaired participants, based on the speech features presented above. To compare the differences in these two categorisation methods (MMSE/CSS), separate LDA classifiers were trained for each method. The *MMSE low* (i.e. cognitively impaired subjects) group included 47 subjects, the *MMSE high* (i.e. cognitively healthy subjects) group 161 subjects. The *CCS* low group included 39 subjects and the *CCS* high group 169 subjects. See Chapter 3, Section 3.2 for details regarding the classification methods.

5.2.2 Results

Text Passage

Average speech feature values for the *MMSE high* and *MMSE low* groups for the *Text Passage* task are presented in Table 5-6. The *Number of Pauses*, *Mean Utterance Duration* and *Total Length of Pauses* show largest differences in their mean values between the two groups. While none of the distributions are significantly different, the difference in the distributions for *Number of Pauses* and *Mean Utterance Duration* are found to approach statistical significance, p=.039, p=0.09 respectively.

Table 5-6: Average speech feature values and standard deviations for the Text passage task using MMSE categorization of subjects

Feature	MMSE low	MMSE high
Number of subjects	47	161
Number of Pauses	57 (18)	46 (14)
Mean Pause Duration (s)	0.7 (0.18)	0.6 (0.11)
Total Length of Pauses (s)	38.1 (19)	30.0 (13)
Proportion of Recording in Silence (%)	23 (7)	20 (5)
Mean Utterance Duration (s)	2.3 (0.52)	2.6 (0.61)
Total Length of Utterances (s)	124 (20)	115 (16)
Total Recording Time (s)	162 (33)	145 (23)

Average value (standard deviation).

Table 5-7 presents the average speech feature values for the categorization of subjects using the *CCS*. This method finds that all features except *Mean Pause Duration* have larger differences than those observed in Table 5-6. Applying the Student's t-test to these data sets found all features, except *Mean Pause Duration* to show statistically different distributions; *Number of Pauses* (p=.0003), *Total Length of Pauses* (p=.0008), *Proportion of Recording in Silence* (p=.01), *Mean Utterance Duration* (p=.0007), *Total Recording Time* (p=.03) and *Total Length of Utterances* (p=.004).

Table 5-7: Average speech feature values and standard deviations for the Text passage task using CCS categorization of subjects

Feature	CCS low	CCS high
Number of subjects	39	169
Number of Pauses	50.9 (16.7)	52.6 (13.3) §
Mean Pause Duration (s)	0.65 (0.15)	0.67 (0.10)
Total Length of Pauses (s)	33.5 (17)	36.2 (13) §
Proportion of Recording in Silence (%)	21 (6)	22 (5) †
Mean Utterance Duration (s)	2.4 (0.55)	2.4 (0.63) §
Total Length of Utterances (s)	118 (17)	121 (17) †
Total Recording Time (s)	152 (29)	158 (24) *

*Average value (standard deviation). *Statistically significant at p<0.05. †Statistically significant at p<0.01. § Statistically significant at p<0.001*

Picture Taboo task

Similarly to the results above, Table 5-8 and Table 5-9 present the average speech feature values for the two cohort splits (*MMSE* and *CCS*) for the *Picture Taboo* task. As each subject describes five pictures during the *Picture Taboo* task, the speech feature values were averaged over the five repetitions of the task.

Table 5-8: Average speech feature values and standard deviations for the Picture Taboo task using MMSE categorization of subjects

Feature	MMSE low	MMSE high
Number of Pauses	10.6 (2.0)	10.2 (1.7) †
Mean Pause Duration (s)	1.33 (0.58)	1.07 (0.48) §
Total Length of Pauses (s)	12.7 (4.5)	10.7 (4.6) §
Proportion of Recording in Silence (%)	42 (15)	36 (16) †
Mean Utterance Duration (s)	1.8 (0.72)	1.92 (0.73) †
Total Length of Utterances (s)	17.3 (4.5)	19.2 (4.7) §

Average value (standard deviation). †Statistically significant at p<0.01. § Statistically significant at p<0.001

Table 5-9: Average speech feature values and standard deviations for the Picture Taboo task using CCS categorization of subjects

Feature	CCS low	CCS high
Number of Pauses	10.2 (2.1)	10.4 (1.8)
Mean Pause Duration (s)	1.4 (0.67)	1.1 (0.47) *
Total Length of Pauses (s)	13.6 (5.2)	11.1 (4.4) *
Proportion of Recording in Silence (%)	45 (17)	37 (15)
Mean Utterance Duration (s)	1.8 (0.85)	1.97 (0.7) *
Total Length of Utterances (s)	16.5 (5.2)	18.9 (4.5) *

*Average value (standard deviation). *Statistically significant at p<0.05*

The difference in the feature values between the high and low cognition groups is clear from these results. The *CCS* categorization results in larger differences between the average feature values for four out of the six features. Statistically significant differences in the data distributions, using the *MMSE* categorization, are found for all features. *Number of Pauses* (p=.007), *Mean Pause Duration* (p=.001), *Proportion of Recording in Silence* (p=.008), *Total Length of Pauses* (p=.001), *Mean Utterance Duration* (p=.008), *Total Length of Utterances* (p=.001).

Using the *CCS* split, no significant differences were found in the data distributions for the *Number of Pauses* and the *Mean Utterance Duration*. Statistically significant differences were found for the *Mean Pause Duration* (p=.03), *Proportion of Recording in Silence* (p=.03), *Total Length of Pauses* (p=.03), and *Total Length of Utterances* (p=.02).

Classification results

Figure 5-3 presents the classification accuracy of the LDA classifier using a combination of all speech features, for both tasks (*Text Passage* and *Picture Taboo*) and for each cognitive categorisation method (*MMSE* and *CCS*).

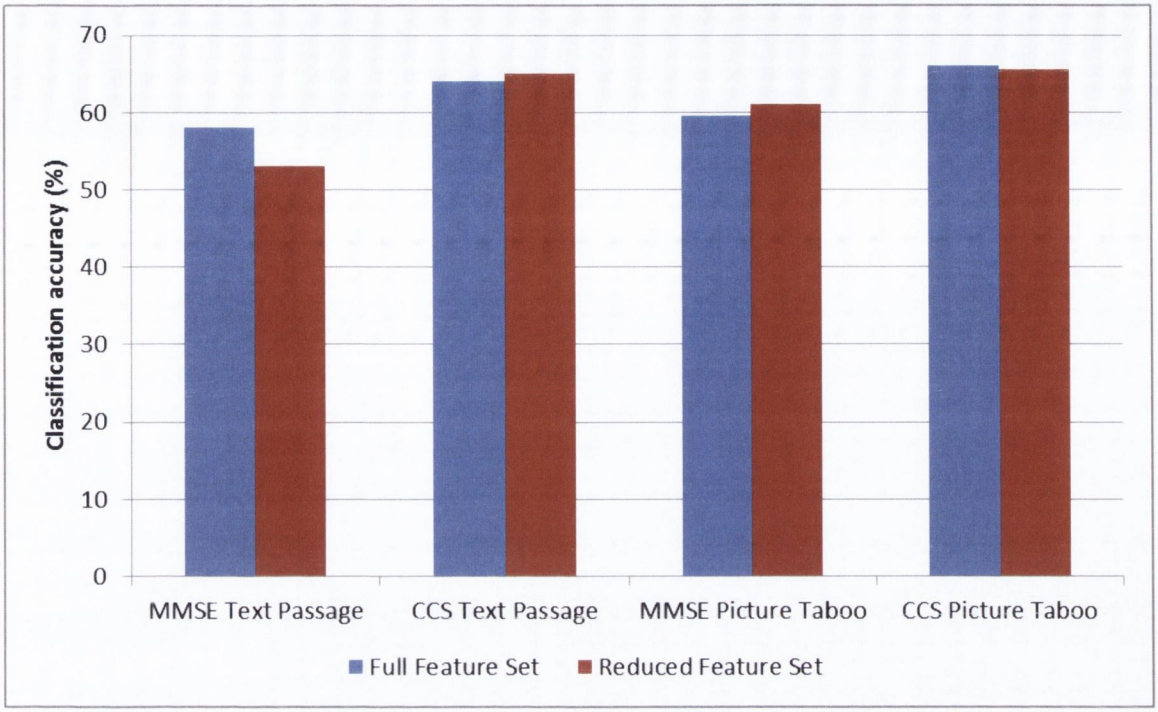


***Figure 5-3**: Accuracy in classifying subjects into cognitive classes as a function of classification method, feature set and speech type*

For the *Text Passage* task, classifying subjects based on the *CCS*, LDA achieve 65% accuracy. Similarly for *Picture Taboo* task, it was the categorisation based on the *CCS* that achieved the highest classification accuracy, 66%.

The classification was repeated using a reduced feature set. This set contained the features which were found to have statistically different data distributions based on the results presented hereinbefore. This reduced feature set either achieved lower or equivalent classification accuracy for all combinations of scales and tasks.

5.2.3 Discussion

As mentioned in Chapter 2, Section 2.10, the *MMSE* has many limitations, in particular the subjective scoring protocol of this test. In this study, the *MMSE* was compared against a *Combine Cognitive Score (CCS)* calculated employing cognitive tasks that targets specific cognitive domains and have more objective scoring protocols than the *MMSE*. The aim of this study was to assess the ability of speech based measures to correlate with this new combined cognitive score.

Feature values were compared between cognitively healthy and impaired groups in terms of the *CCS* categorization method. For the *Text Passage* task the groups categorized in

terms of *CCS* found more statistically significant difference than that observed for the groups categorized by *MMSE*. For the *Picture Taboo* task, it was the *MMSE* categorization that found more statistically significant differences between the two subjects' groups. The statistical analysis of the results of the *Text Passage task* found utterance based features to be more discriminatory between cognitive groupings. The variation in utterance durations and *Number of Pauses*, which were found to be highly significant, may be related to a subject's ability to maintain the thread of a sentence or phrase for its duration.

The performance of the LDA classifier was assessed for both cognitive categorization methods. Despite the more significant differences in between the groups categorized based on *CCS*, the LDA classification just slightly outperformed that of the classifier employing the *MMSE* categorization. To address any redundancy in the features, the classification was repeated using a reduced feature set containing only those features which exhibited statistical significance. However, this was not found to improve classification accuracy. This implies that each feature contributes to the classification performance.

The use of the CCS as an alternative cognitive measure is presented. Using 27 as a threshold on the MMSE scale, the low and healthy cognitive groups contain 47 and 161 subjects respectively compared to 39 and 169 subjects respectively for the CCS. On application of this new cognitive classification, all but five subjects were moved from the cognitively 'low' group to the cognitively healthy group. These five subjects only met the expected normal scores, for age and education, on a maximum of two cognitive tests. There is no absolute agreement in the literature that 27 be employed as the threshold for cognitive decline. The 32 subjects in this corpus with an MMSE score of 27 can be considered to be on the border of low cognitive function. If the MMSE threshold was reduced to 26 to account for this uncertainly, the five potentially cognitively low subjects, mentioned above, would be misrepresented as cognitively healthy.

The CCS was developed to overcome some of the limitations of using the MMSE on its own, and it is shown here to outperform the classification based on MMSE categorization alone (Figure 5-3). Classification accuracy of subjects categorised by CCS increased by 11% and 18% compared to subjects categorised by MMSE, for the full and reduced feature sets respectively for the *Text Passage* task. For the *Picture Taboo* task, CCS categorization increase classification of speakers by 10% and 7% for the full and reduced feature sets respectively.

Quantifying the advantage of the CCS over the MMSE categorization method would require a large sample longitudinal study. However, the principles behind its generation and the subsequent classification results here suggest it is a valid tool for classifying cognitive status.

Key points

- All temporal features, except *Mean Pause Duration*, for the *Text Passage* task achieved statistically significant differences between the cognitively impaired and healthy group for the subjects' cohort split employing the *CCS* categorization.
- All temporal features for the *Picture Taboo* task achieved statistically significant differences between the cognitively impaired and healthy group for the subjects' cohort split employing the *MMSE* categorization.
- Four out of six features (*Mean Pause Duration*, *Total Length of Pauses*, *Mean Utterance Duration*, *Total Length of Utterances*) for the *Picture Taboo* task achieved statistically significant differences between the cognitively impaired and healthy group for the subjects' cohort split employing the *CCS* categorization.
- Classification accuracy of LDA classifier for subjects categorized by the *CCS* increased in comparison with *MMSE* categorization for both tasks.

5.3 Study 3 – Temporal features of speech as indicators of individual's memory state

Addressing the limitations of the MMSE, Small et al. [313] showed how the MMSE memory subscale is relatively independent from education and furthermore is the best predictor of Alzheimer's disease (AD) of all the MMSE subscales. Other subscales include orientation to time and place, registration, attention and concentration, praxis, and constructional and language capacity. See Chapter 2, Section 2.9.1 for details of MMSE subscales.

For this reason, an investigation into the MMSE memory subscale for validation of the temporal speech features extracted from Picture Description (PD) and Picture Taboo (PT) tasks was carried out. These two tasks were chosen over the text passage task, due to their targeting of word retrieval processes.

The aim of this study was to investigate if a subject's performance on the MMSE memory subscale can be correlated with the characteristics of speech. A second aim of this study was to investigate if age and education correlate with the characteristics of speech. The last aim was to investigate if an LDA classifier trained using the speech features of older adults would allow discrimination between low and high memory performers.

5.3.1 Methods

Cohort

A sample of 40 older adults between 66 to 82 years old, mean age = 72.6, SD = 4.6, were selected for in this study. On the basis of the MMSE memory subscale, the participants were split into two groups. Participants recalling zero or one word out of three were classified as *Low memory performers (L)*, n = 28. Those recalling two or three words were classified as *High memory performers (H)*, n = 12. The two groups differed in MMSE total score; the *Low memory performers* showed a significantly lower MMSE total score, mean = 24.92, SD = 2.84, than the *High memory performers*, mean = 27.89, SD = 1.96, $t(38) = 3.82$, $p < 0.001$. However, the two groups did not differ neither in age, *Low memory performers*, mean = 74.00, SD = 5.34; *High memory performers*, mean = 72.54, SD = 5.34, $t(38) = -0.89$, $p = 0.3$, nor education level, $\chi 2\ (4, N = 40) = 2.58$, $p = 0.46$.

Speech tasks

The subjects carried out two tasks: *Picture Taboo* and *Picture Description*. During the *Picture Taboo* task, the subjects described a picture without using two taboo words. See Section 5.1.1 of the Study 1 for details regarding this speech task.

In addition to the *Picture Taboo (PT)* task, the participants' speech was recorded during the *Picture Description (PD)* task.

Picture Description task

For the *Picture Description (PD)* task, the participants were asked to describe the picture they saw displayed on a computer screen in front of them, in as much detail as possible in 30 seconds. Compared to *Picture Taboo* task, this task did not contain the two taboo words and the participants were allowed to use any words during the description of the picture in front of them. The time was controlled with a stopwatch and a beep-sound would mark the beginning and the end of the task. The participants were instructed to describe the picture in as much detail as possible. The experimenter provided vocal instructions and remained beside the subject throughout the test.

***Figure 5-4**: Sample picture of the Picture Description task.*

Data acquisition

Participants' speech was recorded for two tasks: *Picture Description (PD)* and *Picture Taboo (PT)*. A headset microphone was used to record speech from all participants; this was connected to an external USB sound card connected to a PC. All efforts were made to keep the microphone at the same distance from the mouth and to reduce background noise to a minimum. Speech was recorded at 44.1 kHz sampling frequency and 16 bit resolution.

Data processing

Following temporal features were extracted from the acquired speech recordings in this study. See Chapter 3, Section 3.1 for description of the speech features extraction process.

- *Number of Pauses*
- *Mean Pause Duration*
- *Proportion of Recording in Silence*
- *Mean Utterance Duration*
- *Total Recording Time*
- *Total Length of Pauses*
- *Total Length of Utterances*

Classification and statistical analysis

For each task, Multivariate Analysis of Variance (MANOVA) with the variable Group as between-subjects factor was performed to estimate significant changes in speech features between subjects' groups. Subsequently, an exploratory correlation analysis (Pearson's r correlation) was employed to estimate correlation between each temporal speech feature and education, age, MMSE total score and MMSE memory subscale.

A Linear Discriminant Analysis (LDA) classifier was employed to differentiate between the *Low memory performers* and *High memory performers* groups, based on the speech features. See Section 3.2 for details regarding the classification methods.

5.3.2 Results

The Multivariate Analysis of Variance (MANOVA) yielded the following results. For the *Picture Description* task, a statistically significant group effect was observed for *Mean Pause Duration*, $F(1,40) = 6.38$, $p < 0.05$, *Proportion of Recording in Silence*, $F(1,40) = 6.71$, $p < 0.05$, *Total Length of Pauses*, $F(1,40) = 6.69$, $p < 0.05$ and *Total Length of Utterances*, $F(1,40) = 6.73$,

$p < 0.05$ (see Table 5-10, *Picture Description* section for more details). In the *Picture Taboo* task, the MANOVA analysis showed a main effect of the group for the *Mean Utterance Duration*, $F(1,40) = 5.04$, $p<0.05$, *Proportion of Recording in Silence*, $F(1,40) = 4.99$, $p<0.05$, *Total Length of Pauses*, $F(1,40) = 4.99$, $p<0.05$ and *Total Length of Utterances*, $F(1,40) = 4.98$, $p<0.05$ (see Table 5-10, *Picture Taboo* section for more details).

Table 5-10: Average temporal features values for Low and High memory performers groups and both speech based tasks

	Picture Description		Picture Taboo	
	H group	L group	H group	L group
	Mean (SD) C.I.	Mean (SD) C.I.	Mean (SD) C.I.	Mean (SD) C.I.
Mean Pause Duration (s)	0.75 (0.22) 0.65 – 0.84	1.03 (0.46) 0.76 – 1.30	0.99 (0.28) 0.86 – 1.10	1.19 (0.45) 0.94 – 1.18
Proportion of Recording in Silence	0.25 (0.08) 0.21 – 0.28	0.35 (0.16) 0.25 – 0.44	0.33 (0.11) 0.28 – 0.38	0.42 (0.14) 0.34 – 0.50
Mean Utterance Duration (s)	2.38 (0.75) 2.06 – 2.70	1.95 (0.79) 1.50 – 2.41	2.15 (0.80) 1.80 – 2.49	1.61 (0.48) 1.33 – 1.89
Total Length of Pauses (s)	7.41 (2.57) 6.32 – 8.50	10.59 (5.02) 7.69 – 13.49	9.88 (3.32) 8.48 – 11.29	12.60 (4.06) 10.25 – 14.94
Total Length of Utterances (s)	22.61 (2.58) 21.52 – 23.70	19.42 (5.02) 16.52 – 22.32	20.13 (3.32) 18.73 – 21.54	17.43 (4.06) 15.08 – 19.77

H group *– High memory performers,* ***L group*** *– Low memory performers,* ***C.I.*** *– Confidence Interval*

Picture Description task

Pearson's correlation analysis yielded following results for the *Picture Description* task. This analysis showed that the MMSE total scores do not correlate with the different temporal features while the MMSE memory subscale does with the exception for the *Mean Pause Duration* (see Table 5-11 for details).

Table 5-11: Pearson's r correlation values between the variables education, age, MMSE scores, MMSE memory subscale scores and temporal features for the Picture Description task

	Mean Pause Duration	Proportion of Recording in Silence	Mean Utterance Duration	Total Length of Pauses	Total Length of Utterances
Education	-0.49*	-0.42**	0.28	-0.43**	0.42**
Age	0.47**	0.50**	-0.38*	0.50**	-0.50**
MMSE	-0.25	-0.23	0.08	-0.23	0.22
MMSE memory	-0.24	-0.33*	0.35*	-0.33*	0.33*

*Pearson's r correlation values. * Significant at p<0.05. ** Significant at p<0.01.*

The speech temporal measures of *Picture Description* task were correlated with age and level of education of the participants. *Mean Utterance Duration*, however, is the only feature not correlated with education (r = - 0.28, p = 0.19).

Picture Taboo task

Similarly, the same analysis performed on the *Picture Taboo* task showed no significant correlation between MMSE total scores and the different temporal speech features. Again, the MMSE memory subscale correlates with all the temporal features, with the exception for the *Mean Pauses Duration* (see Table 5-12 for details).

Table 5-12: Pearson's r correlation values between the variables education, age, MMSE scores, MMSE memory subscale scores and temporal features for the Picture Taboo task

	Mean Pause Duration	Proportion of Recording in Silence	Mean Utterance Duration	Total Length of Pauses	Total Length of Utterances
Education	-0.40*	-0.41**	0.28	-0.41**	0.42**
Age	0.37*	0.41**	-0.23	0.41**	-0.41**
MMSE	-0.22	-0.26	0.22	-0.26	0.26
MMSE memory	-0.19	-0.31*	0.50**	-0.32*	-0.32*

*Pearson's r correlation values. * Significant at p<0.05. ** Significant at p<0.01.*

Similarly to the *Picture Description* task, among the temporal features, *Mean Utterance Duration* appeared to be the most robust measure independent from age (r = .23, p =0.14) and education (r = .28, p = 0.07).

Classification results

The classifier's ability to discriminate between *High memory performers* and *Low memory performers* employing the temporal features, for both speech based tasks, was assessed. For the *Picture Description* task, the LDA classifier discriminated between the *Low memory performers* and *High memory performers* with accuracy of 68%, sensitivity of 63% and specificity of 73% (see Figure 5-5, green line). However, training the LDA classifier with only the *Mean Utterance Duration* feature, yielded following classification performance: accuracy of 65%, sensitivity of 75%, and specificity of 54%.

For the *Picture Taboo* task, the LDA classifier discriminated between the two participants' groups with accuracy of 73%, sensitivity of 79%, and specificity of 67% (see Figure 5-5, blue line). However, training the LDA classifier with only the *Mean Utterance Duration*, yielded

following classification performance: accuracy of 65%, sensitivity of 74%, and specificity of 56%.

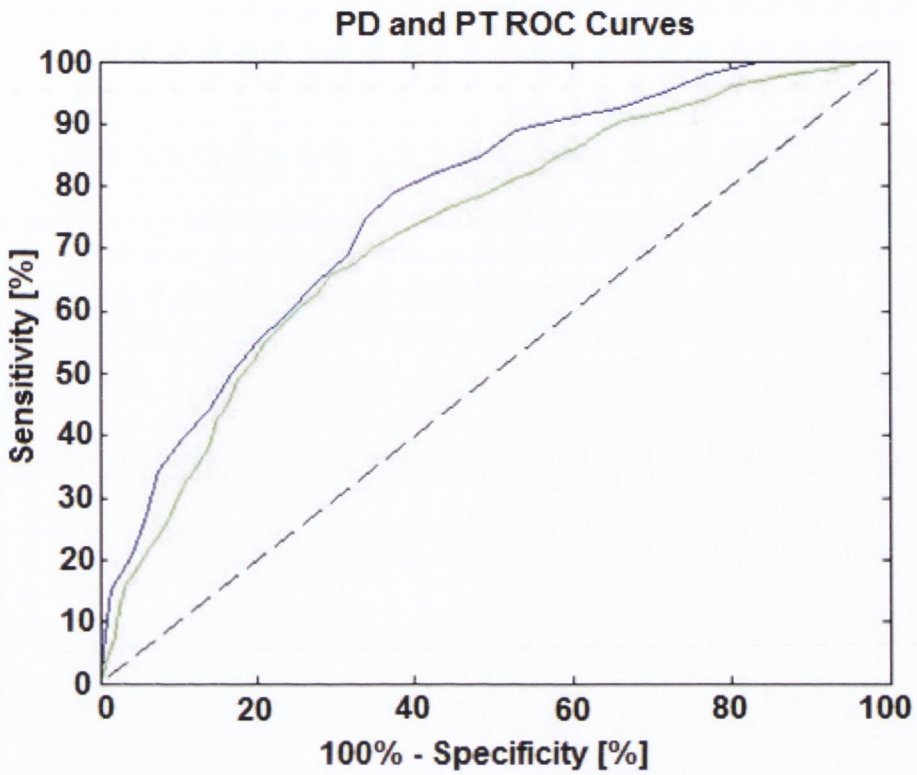


***Figure 5-5**: ROC curves representing the classification accuracy of the speech features extracted from the Picture Taboo (PT) task (blue line) and Picture Description (PD) task (green line). The dashed black diagonal line represents 50% of chance.*

5.3.3 Discussion

The Pearson's *r* correlation analysis confirmed the independence of the MMSE memory subscale from education ($\chi 2$ (4, N =40) = 2.58, p = 0.46), in agreement with the findings of Small et al. [313]. This independence is important since language is an acquired skill heavily influenced by the educational level of an individual. Therefore, a measure not biased by education is to be favoured when screening for cognitive decline through language.

An analysis of the relationship between speech features and the MMSE memory subscale was performed in order to provide a preliminary validation. The results suggest a relationship between the temporal features of speech and individual's memory.

The two tasks employed in this study, *Picture Description* and *Picture Taboo*, differ in difficulty. From the analysis of the temporal features, a main effect of the task was observed, where the *Picture Taboo* task elicited longer pauses and shorter utterances than the *Picture*

Description task. This is possibly due to constraints imposed by the taboo words, which increased the cognitive demand of the *Picture Taboo* task. The longer pauses can therefore be interpreted as time required for alternative word finding while maintaining control of the task.

The main effect of the group (*Low memory performers* and *High memory performers*) resulted in longer pauses and reduced amount and length of utterances for the *Low memory performers* compared to the *High memory performers*. This suggests a relation between the short-term memory functioning (reflected in the number of words recalled after a short delay) and temporal features of the speech. This is further suggested by the significant correlation achieved between speech features and MMSE memory subscale, which was not observed with the MMSE total score.

In particular, the *Mean Utterance Duration* appeared to be the most robust among all the speech features thanks to its independence from education observed in both tasks. Moreover, the ability of the all speech features to discriminate between the two groups only slightly outranks the classification accuracy of the *Mean Utterance Duration* when used alone in both tasks.

Conclusion

This study aimed at testing the potential of speech features from two picture description tasks for classifying subjects based on memory performance. The most promising results showed a correlation of four out of the five speech features (*Proportion of Recording in Silence, Mean Utterance Duration, Total Length of Pauses,* and *Total Length of Utterances*) with the MMSE memory subscale and not with the MMSE total score. This suggests that the temporal characteristics of speech may reflect a cognitive process that is more specifically linked with short-term memory mechanisms. In particular, the feature *Mean Utterance Duration* has shown particular strength thanks to its independence from education, and relatively high classification accuracy.

The results of this study are particularly important in terms of repeated assessment of cognitive function. While the word list employed for the memory task is a validated tool for assessing memory, it is not suitable for repeated administration due to a practice effect. Extracting features of speech from tasks, such as *Picture Description* and *Picture Taboo*, may provide a reliable, repeatable method of assessing memory.

Key points

- Four out of the five temporal features of speech (*Proportion of Recording in Silence, Mean Utterance Duration, Total Length of Pauses, Total Length of Utterances*) extracted from picture description tasks correlate with the MMSE memory subscale.
- The *Mean Utterance Duration* feature of the *Picture Taboo* task demonstrated independent from individual's education.
- Shorter utterances and longer pauses were observed during the *Picture Taboo* task than during the *Picture Description* task.
- Potential use of temporal features of speech for monitoring individual's memory over time.
- Employing speech features extracted from the *Picture Description* task, the LDA classifier discriminated between the *Low memory performers* and *High memory performers* with accuracy of 68%, sensitivity of 63% and specificity of 73%.
- Employing speech features extracted from the *Picture Taboo* task, the LDA classifier discriminated between the *Low memory performers* and *High memory performers* with accuracy of 73%, sensitivity of 79%, and specificity of 67%

5.4 Study 4 – Dynamic minimum pause threshold estimation

Despite the apparent significance of pause-oriented features, much of the research into pause detection has employed criteria that is either poorly defined or based on the speech performance of neurologically intact adults [314]. Minimum pause duration is consistently used as a criterion for detection of pause boundaries [314] and is furthermore one that varies considerably across the literature, reflecting the arbitrary nature in which most researchers have employed it [315, 316]. Goldmann-Eisler [273] advocated the use of a minimum pause duration of 250ms, and this threshold has proliferated many of the studies into pausing behaviour including Stassen et al. [251]. However, there are numerous examples of studies that have deviated from this 250ms value for minimum pause duration – a review of relevant publications conducted by Kirsner et al [315] yielded 32 different values, ranging from 100 up to 300ms and with a median of 250ms. Minimum pause duration thresholds encountered as part of this review included 40ms [317], 270ms [318], and 1000ms [61]. Given the variability in pause threshold values, it is not immediately evident what an appropriate value is for minimum pause duration.

Initiated by Kirsner et al [315], this traditional approach to pause identification has been readdressed, and emerging from this renewed interest is evidence that pause duration exhibits a two-component mixed lognormal distribution, one component associated with short pauses – products of articulatory processes and another associated with long pauses – associated with cognitive processes. Having successfully fitted this distribution it is possible to determine an optimal threshold value for each recording for differentiating between the short-pause component and long-pause component. As this threshold value can vary from speaker to speaker it is dynamic in nature.

Recent studies by Rosen et al [319], and Hird & Kirsner [320] investigated the correlation between the pause distribution parameters and Friedrich's ataxia (FRDA) and brain-damage induced aphasia respectively. Rosen et al [319] found a significant difference between the parameters for those with FRDA and control subjects and Hird & Kirsner [320] graphically demonstrated the variation between brain damaged subjects and control subjects. Given these findings, findings of Roark et al [61] and the findings presented in Sections 5.1 - 5.3 of this thesis that pause-related features are correlated with cognitive function, there is clearly potential in using pause distribution parameters to discriminate between cognitively healthy and cognitively impaired subjects.

This study readdresses the approach presented in the previous sections of this chapter (Sections 5.1 - 5.3), replacing the use of a static 250ms minimum pause duration with this dynamic threshold. The parameters of the pause and utterance distributions are furthermore assessed in their ability to discriminate between cognitively healthy and cognitively impaired speakers.

5.4.1 Methods

Cohort

187 older adults participated in this study, 73 (39.04%) of whom were male and 114 (60.96%) female, with a mean age of 72.44 (SD 7.01, range 60 - 80) years. All participants underwent cognitive assessment using the MMSE, the scores ranging from 20 to 30 with a mean MMSE score of 27.68 (SD 2.00). Based on their MMSE scores, the participants were segmented into two groups. With an MMSE score of 27 or higher, 150 of the participants (80.21%) were classified as cognitively healthy, and 37, with a score of 26 or lower (19.79%), were classified as cognitively impaired.

Data acquisition

Participants' speech was recorded for the *Text Passage* task. During *Text Passage* task, the subjects read aloud a short text passage. See Section 5.1.1 of the Study 1 for details regarding the speech tasks. The speech from each subject carrying out this speech based task was recorded.

A headset microphone was used to record speech from all participants; this was connected to an external USB sound card connected to a PC. All efforts were made to keep the microphone at the same distance from the mouth and to reduce background noise to a minimum. Speech was recorded at 44.1 kHz sampling frequency and 16 bit resolution.

Data processing

Following temporal features were extracted from the acquired speech recordings in this study. See Chapter 3, Section 3.1 for description of the speech features extraction process. These features, using the 250ms minimum pause duration threshold, were labelled as *Static temporal features* in this study.

- *Number of Pauses*
- *Mean Pause Duration*
- *Proportion of Recording in Silence*
- *Mean Utterance Duration*
- *Total Recording Time*
- *Total Length of Pauses*
- *Total Length of Utterances*

In addition to these *Static temporal features*, using the dynamic minimum pause threshold estimation, another set of temporal features was generated labelled as *Dynamic temporal features*.

Dynamic minimum pause threshold estimation

Following the initial segmentation temporal thresholds were applied to more accurately identify both pauses and utterances. Initial thresholds of 100ms for minimum utterance duration and 250ms for minimum pause duration were employed to provide the secondary segmentation of the signal into long utterances and pauses. Thresholds of 30ms and 20ms were then applied to the long utterances to break them into short utterances and pauses, providing the tertiary and final segmentation of the signal into speech and pauses. In this manner all pauses of duration greater than or equal to 20ms and all utterances of duration greater than or equal to 30ms were identified. Following the identification of the pauses and utterances, the pause and utterance distributions were generated. The dynamic minimum pause threshold, which was different for every recording, was estimated from the pause distribution as described below.

Pause and utterance distributions

Two-component mixed lognormal distributions were fitted to the pause duration data using maximum likelihood estimation (MLE). The two components of the distribution were termed the short-pause component and the long-pause component. The point at which the detection error was minimized for each component was found to be at the intersection of the two components. This intersection point was employed as the 'dynamic' minimum pause threshold and was used to discriminate between the long-pauses, which were of interest and the short-pauses which were to be discarded. MLE was also employed to fit a unimodal lognormal distribution to utterance duration data for each recording.

The following *Distribution features* were then extracted from the pause and utterance distributions:

- *Pause Mixing Proportion*
- *Short-Pause Mean*
- *Long-Pause Mean*
- *Short-Pause Standard Deviation*
- *Long-Pause Standard Deviation*
- *Utterance Mean*
- *Utterance Standard Deviation*

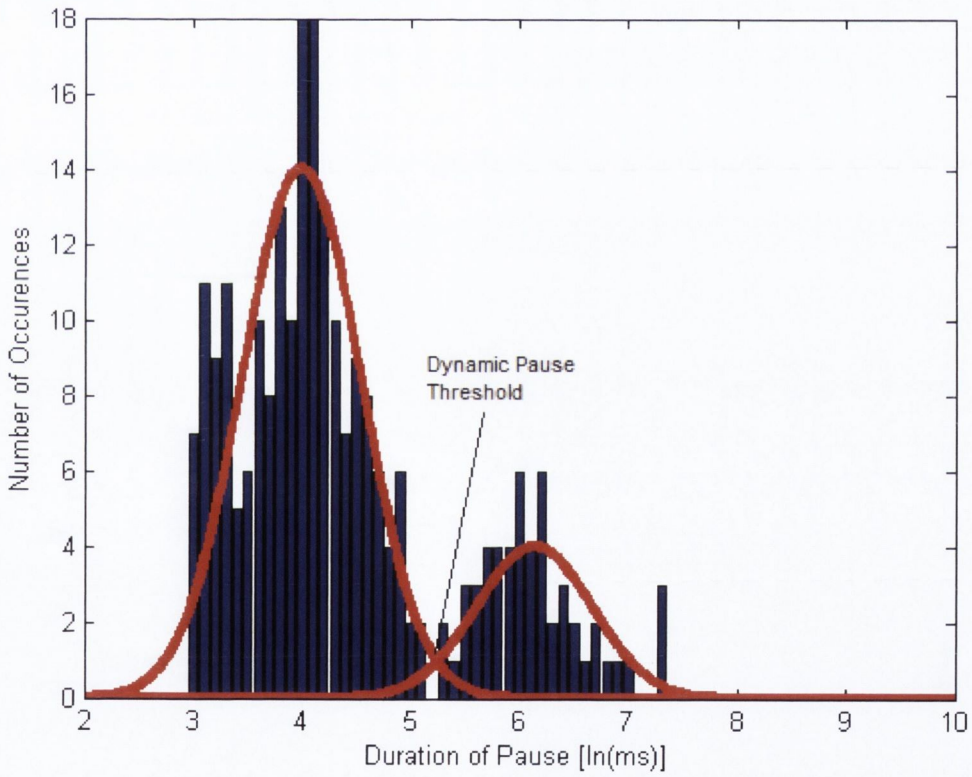


Figure 5-6: *Pause mixed lognormal distribution*

Classification and statistical analysis

The performance of these features in discriminating between cognitively healthy (MMSE >=27) and cognitively impaired (MMSE<27) participants was assessed using a combination of Student's t-test, Welch's t-test and Wilcoxon's Rank-Sum test. The selection of the appropriate test for each feature was based on the results of normality tests, variance tests, and visual inspection of quantile-quantile plots. The feature set under analysis comprised the parameters

of the pause distributions and utterance distributions, and the temporal features for the static and dynamic thresholds. The features that emerged as statistically significant at a significance level of $\alpha = 0.05$ were subsequently used in the classification procedure.

A Linear Discriminant Analysis (LDA) classifier was employed to differentiate between healthy and cognitively impaired participants, based on the speech features presented above. See Chapter 3, Section 3.2 for details regarding the classification.

5.4.2 Results

From the three sets of features (*Static temporal features, Dynamic temporal features,* and *Distribution features*) the variance analysis yielded seven statistically significant features (see Table 5-13 and Table 5-14). Inspection of the mean values of these features (see Table 5-15) revealed that the participants of the cognitively impaired group generated more pauses for the static threshold case, with a mean of 28.76s compared to 24.01s for the healthy group. The *(Dynamic) Total Length of Pauses* feature indicated that the impaired group paused on average 6.5s more than the healthy group. Both static and dynamic *Total Length of Utterances* demonstrated that the impaired group had longer utterances than the healthy group.

Table 5-13: Student's t-test analysis of speech features between participants groups

	Student's *t*-test		
Speech feature	***t***	***df***	***p***
(Distribution) Utterance Mean (ln(ms))	2.4824	185	0.0139
(Dynamic) Total Length of Utterances (s)	2.1258	185	0.0348
(Static) Proportion of Recording in Silence	-2.5708	185	0.0109
(Static) Total Length of Utterances (s)	2.2888	185	0.0260

Distribution – features extracted from lognormal distributions,
Dynamic – temporal features extracted employing dynamically estimated pause threshold,
Static – temporal features extracted employing static pause threshold

Table 5-14: Wilcoxon Rank-Sum test analysis of speech features between participants groups

	Wilcoxon Rank-Sum Test		
Speech feature	***U***	***z***	***p***
(Distribution) Pause Mixing Proportion	3214	-2.0883	0.0368
(Dynamic) Total Length of Pauses (s)	4488	2.0687	0.0386
(Static) Number of Pauses	4514	2.1556	0.0311

Distribution – features extracted from lognormal distributions,
Dynamic – temporal features extracted employing dynamically estimated pause threshold,
Static – temporal features extracted employing static pause threshold

Table 5-15: Mean feature values for features that achieved statistically significant differences for both participants' groups

Speech feature	Mean feature value	
	Cognitively healthy group	Cognitively impaired group
(Distribution) Utterance Mean (ln(ms))	5.77	5.93
(Distribution) Pause Mixing Proportion	0.76	0.71
(Static) Number of Pauses	24.01	28.76
(Static) Mean Pause Duration per Second (s)	0.13	0.15
(Static) Total Length of Utterances (s)	110.64	122.03
(Dynamic) Total Length of Pauses (s)	29.08	35.58
(Dynamic) Total Length of Utterances (s)	109.66	118.44

Distribution – features extracted from lognormal distributions,
Dynamic – temporal features extracted employing dynamically estimated pause threshold,
Static – temporal features extracted employing static pause threshold

Table 5-16: Linear Discriminant Analysis classification

Classification performance	Feature set		
	(Static) Temporal Features	*(Dynamic) Temporal features*	*Distribution features*
Overall accuracy (%)	65.39	61.97	68.66
Sensitivity (%)	63.98	58.25	64.20
Specificity (%)	66.79	65.69	73.12
ROC Area	0.69	0.58	0.74

ROC Area – Area under the Receiver Operating Characteristics (ROC) curve

Classification of the participants was then performed using the features that were found to be statistically significantly different between the participants' groups. The results of the classification can be seen in Table 5-16. Employing the *Dynamic temporal features* yielded decreases of 5.73% (to 58.25%) in the sensitivity, 1.10% (65.69%) in the specificity, and 3.42% (61.97%) in the accuracy when compared with the performance of the static temporal features.

When the classifier was trained using the pause and utterance *Distribution features*, the classification performance increased. The sensitivity of the LDA classifier increased by 0.22% (to 64.20%), specificity by 6.33% (73.12%) and the overall accuracy by 3.27% (68.66%).

5.4.3 Discussion

This study investigated if a dynamic estimation of minimum pause duration can improve the correlation of speech features with cognitive tasks measures. From the investigation of pause and utterance duration distribution data and their impact on the performance of an

LDA classifier, two distributional parameters, *Pause Mixing Proportion* and *Utterance Mean*, were found to be statistically significantly different between the cognitively healthy and cognitively impaired groups. These two parameters outperformed the temporal features in classifying the participants according to their level of cognitive function.

Contrary to expectations however, the use of a dynamic threshold derived from the distributional data had a negative impact on the classification performance of the temporal features. These findings would suggest that the distributional data are best employed directly, rather than using them to estimate the temporal features for each individual speaker via the dynamic threshold estimation.

Despite dynamic threshold estimation yielding no improvement on traditional methods, this study did highlight the potential in pause and utterance duration distribution parameters in improving classification of people according to their cognitive function.

Key points

- The *Pause Mixing Proportion* and *Utterance Mean* distribution parameters showed statistically significant differences between cognitively impaired and healthy group.
- The *Distribution features* outperformed the temporal features in classifying the participants according to their level of cognitive function.
- The use of a dynamic threshold derived from the distributional data had a negative impact on the classification performance of the temporal features.
- Employing the *Dynamic temporal features*, the sensitivity of the LDA classifier decreased by 5.73% (to 58.25%), specificity by 1.10% (65.69%) and the accuracy by 3.42% (61.97%).
- Employing the *Distribution features*, the sensitivity of the LDA classifier increased by 0.22% (to 64.20%), specificity by 6.33% (73.12%) and the overall accuracy by 3.27% (68.66%).

5.5 Summary

The studies in this chapter answered the research questions posed in Chapter 2, Section 2.18.2. This chapter has investigated the reliability of employing speech features as a means of assessing cognitive function in older adults.

The analysis focused on speech features extracted from two speech based tasks. While both tasks targeted the cognitive processes required to generate speech, these tasks also individually targeted different cognitive processes. The *Text Passage* task targeted visual processing of text and word retrieval processes. The *Picture Taboo* task targeted short-term memory and executive function.

The speech features extracted from these two tasks aim at measuring speech and non-speech segments durations achieved throughout the completion of the task. It was expected that the durations of these segments would be indicators of how fluently participants can generate speech for the specific tasks. For example a large *Number of Pauses* and short *Mean Utterance Duration* would indicate an inability to maintain focus on the task. Long *Mean Pause Duration* could indicate a need for longer cognitive processing required to complete the task.

The studies presented in this chapter investigated the correlation of various speech features and measures of cognitive function. The correlation analysis found that the speech features achieved stronger correlation with individual cognitive tasks measures than with the Mini-Mental State Examination (MMSE) score.

The classification analysis was employed to assess the ability of speech features to differentiate between cognitively healthy and cognitively impaired older adults. This analysis investigated various combinations of speech features to find the most accurate in classifying cognitive status as determined by the MMSE.

The MMSE is a cognitive screening tool of overall cognitive function while the other cognitive tasks employed in this analysis are considered measures of specific cognitive function domains. Classification of individuals in terms of MMSE was as successful when employing the temporal speech features as a classification when employing the scores of these cognitive tasks. This was the first result that suggested the ability to employ speech features to assess cognitive function in older adults.

The successive studies investigated different approaches to improving the classification performance when employing speech features to assess cognitive function. A Combined Cognitive Score (CCS), based on the results of achieving expected age and education norms for MMSE and the results of the cognitive tasks, was generated to investigate if including these tasks would improve classification performance. An improvement was observed, however, further analysis is required to determine if this is due to increased sensitivity of the CCS classification.

The results of the studies indicated that the speech features are better correlated to tasks measures of specific cognitive function domains. Therefore, speech features were assessed in terms of the memory subscale of the MMSE. The increased classification performance, and independence form age and education observed, demonstrated that speech features extracted from picture description tasks have significant potential as method for longitudinal monitoring of at least some domains of cognitive function. While no significant correlation was found between speech features and other subscales, further research should be carried out to identify other speech features that correlate with these subscales.

The final study investigated the extraction of non-speech segments and attempted to improve the sensitivity of this measure. Compared to the existing measures of non-speech segments, this analysis found distribution features to outperform static and dynamic temporal features when classifying participants in terms of MMSE.

The culmination of these studies is that speech features can be employed as a measure of cognitive function. Additional research may be required to refine the speech features and tasks to maximise the power of speech as a measure of cognitive function. However, the benefit of assessing intra-speaker variability in speech features in order to monitor cognitive function is clear.

The next chapter, Chapter 6 describes design and evaluation of an automated remote telephone based system for longitudinal monitoring of cognitive function of older adults.

Chapter 6

Automated remote telephone assessment of cognitive function

Following the clinical study of speech characteristics in older adults presented in Chapter 5, and taking into consideration the monetary, personnel burden and time requirements associated with the cognitive function assessment (see Chapter 1 and Chapter 2), a study investigating automated remote assessment of cognitive function in older adults was carried out.

Cognitive decline occurs at a very individual rate and while there are norms for age ranges [185] there is yet no way of determining at what rate one is going to exhibit signs of cognitive decline. Regular monitoring of cognitive function in individuals would enable the early identification of cognitive decline and facilitate early interventions that could reduce the impact of cognitive decline on an individual. Regular monitoring of cognitive function is the most reliable method to identify changes in cognitive function but it is all but impractical given the face to face nature of cognitive assessments. The clinical face to face cognitive assessment requires the neuropsychologist to administer the assessment. It is time consuming and generally expensive. This is particularly relevant in light of the predicted increase in the number of older adults. Regular monitoring of cognitive assessment is an ideal application for an automated system of longitudinal data collection.

Automation of cognitive assessments has been investigated with computer-based tools. Snyder et al [321] have produced a comprehensive review of sixteen assessments of cognitive impairment. The authors discuss the types of tests included in each assessment, the automation and repeatability of delivering the assessment via a PC and the sensitivity and

specificity of each measure for identifying individuals with Mild Cognitive Impairment (MCI). Many of these assessments were self-administered and did not require a trained professional. While the advantage of these computerized assessments is clear, they do suppose that participants have a computer and are comfortable using it.

Several studies [322-324] have investigated delivery of remote cognitive function assessment and these studies have employed the telephone to remotely separate the subject from the neuropsychologist. The studies reported in the literature indicate that these methods tend to be very reliable when compared to traditional in-person cognitive assessments [324, 325]. While these studies assessed cognitive function remotely, a neuropsychologist was required to administer the cognitive assessment tasks to the participant.

Interactive Voice Response (IVR) systems enable callers to interact with a computer system via a telephone. Pre-recorded prompts or speech synthesis are used to present information and options to callers. Dual tone modulated frequency (DTMF) (pressing the keys on the telephone keypad) or speech are employed to capture input from the caller. A typical IVR system asks participants to respond to presented questions by pressing the appropriate key on the alphanumeric keypad of the phone or responding by voice. Automation of cognitive assessments has been investigated using IVR technologies [325, 326]. A study by Mundt et al [325] modified common cognitive tasks to use DTMF input for each task and in another study using speech input. Mundt [326] employed simple questionnaires to gather input from participants.

Older adults have been shown to use technology if it has been correctly designed. This can be defined as being designed using universal design principles. When insured that the technology works well in human settings, honours people's needs, does not remove person's autonomy, does not invades the privacy of its users [327]. Almost everybody is familiar with telephones and does not feel unduly uncomfortable while using this technology. If proper user centred design principles [328] are employed the resulting IVR system may be cognitively un-demanding.

There is extensive evidence to support the administration of cognitive assessments over the telephone [329], a technology with which most older adults are familiar. Speech has traditionally not been used as the input mode for automated cognitive assessments such as this. However using speech in this manner provides us with a way to design very user centred

applications, thus improving the usability of the system particularly for the specific cohort in this study, i.e. the older adults. Speech in this context has the added advantage of users not having to remove the phone from their ear in order to press the correct button, potentially missing instructions or getting confused with the multi-modal aspect of the task.

Telephones are employed in the two studies presented in this chapter, as they are readily available to older adults and are a technology that this cohort is comfortable with. The first study describes the feature extraction algorithm employed with telephone quality data. The second study describes design, implementation and usability of an automated IVR application for remote delivery of cognitive function assessment.

Three publications have been derived from the research described in this chapter.

S. D'Arcy, **V. Rapcan**, A. Galli, N. Burke, G.C. O'Connell, I.H. Robertson, R.B. Reilly, "A study into the automation of cognitive assessment tasks for delivery via the telephone: lessons for developing remote monitoring applications for the elderly", *Technology and Health Care*, vol. 21, pp. 387-396, 01/01/ 2013

S. D'Arcy, S. Counihan, N. Penard, F. van den Berg, **V. Rapcan**, M. Hansteen Izora, R.B. Reilly, I.H. Robertson, "Technology for remote assessment of cognitive function of the elderly", *Proceedings of the 16th Bioengineering in Ireland Conference*, Dublin, Ireland, January 22-24, 2010.

V. Rapcan, S. D'Arcy, N. Penard, I. H. Robertson, and R. B. Reilly, "The Use of Telephone Speech Recordings for Assessment and Monitoring of Cognitive Function in Elderly People," *Proceedings of Interspeech 2009*, Brighton, UK, 2009, pp. 943-946.

6.1 Study 1 – Modifications of feature extraction algorithms for use with telephone speech recordings

A first step in the process of moving the cognitive assessment from clinic to subjects' homes was to ensure that the developed feature extraction algorithm (described in Chapter 3, Section 3.1) reliably extracts speech features from telephone quality recordings.

The study reported here presents modifications required to be performed on the feature extraction algorithms for reliable extraction of speech characteristics from telephone recordings. The results of applying the feature extraction algorithms to speech recorded over the telephone are presented.

6.1.1 Methods

Cohort

Nineteen older adults participated in this study and were taken from the larger cohort presented in Chapter 5, Section 5.1.1. Mean age (± SD) of the cohort was 68.6 (± 6.9) years. The cohort had mean Mini-Mental State Examination (MMSE) score (± SD) of 28.26 (± 1.69). Two subjects had MMSE score lower than 27. Twelve of the subjects were female, seven were male.

Data acquisition

Speech was recorded in St. James's Hospital clinic for all participants while reading out loud a *Text Passage* described in Chapter 4, Section 4.1.1.

A headset microphone was used to record speech from all participants; this was connected to an external USB sound card connected to a PC. All efforts were made to keep the microphone at the same distance from the mouth and to reduce background noise to a minimum. Speech was recorded at 44.1 kHz sampling frequency and 16 bit resolution.

In addition to these speech recordings, the participants were recorded reading out loud the same text passage in their own homes over a telephone during following home deployment. Each participant was called by a neuropsychologist and the conversation recorded onto a laptop via a USB recorder to which the phone line was connected. The recordings were sampled at 8 kHz, with 16-bit resolution and CCITT μ-LAW compression, as provided by the recording system.

Data processing

The recordings from St. James's Hospital clinic were processed employing the algorithms described in Chapter 3, Section 3.1. This algorithm was also employed in studies of Chapter 5 and Chapter 4. The telephone recordings were processed employing the same algorithms with the modifications described below.

Modifications of the feature extraction algorithm for use with telephone recordings

The telephone recordings differ from clinic recordings in sampling frequency, bandwidth, noise level, have variable volume level and contain more perturbations in form of "clicks", "knocks". As a result of these differences, for telephone recordings two modifications of the feature extraction algorithm were required. These modifications were required to correctly segment the recording into speech and pause segments.

Increased number of perturbations (clicks, knocks) in the recordings caused feature extraction algorithm to misinterpret pauses as speech segments. Therefore, as a first step, all telephone recordings were filtered with a seven-point moving average filter to smooth the signal which helped with suppressing of these unwanted effects. Further, different noise level between the clinic and telephone recordings required the *Reference Value* (see Chapter 3, Section 3.1.2) to be re-estimated for the telephone recordings.

The temporal features presented below were extracted from acquired speech recordings in this study. See Chapter 3, Section 3.1 for description of the feature extraction process.

- *Number of Pauses*
- *Mean Pause Duration*
- *Proportion of Recording in Silence*
- *Mean Utterance Duration*
- *Total Recording Time*
- *Total Length of Pauses*
- *Total Length of Utterances*

Two sets of features were generated, one set from the clinic recordings and the second set from the telephone recordings:

1. Clinic recordings
2. Telephone recordings

Classification and statistical analysis

The ability of the feature extraction algorithm to correctly detect pauses and speech segments in telephone quality recording was investigated. Performance was measured employing the same methods as described in Chapter 4, Section 4.2.1.

In addition to the algorithm performance estimation, the average values of the extracted temporal features of the two features sets were calculated. Student's t-test (2-tailed, 2-sample assuming equal variances) was employed to estimate the differences in means of the two features sets.

6.1.2 Results

The following performance of the feature extraction algorithm was achieved using the telephone recordings.

- Overall accuracy: 93.2%
- Sensitivity: 97.3%
- Specificity: 89.5%
- Positive predictivity: 91.1%
- Negative predictivity: 97.1%

Table 6-1 shows average values for the extracted temporal features from clinic and telephone recordings. This table also shows the results of the Student's t-test analysis.

Table 6-1: Average values comparison for clinic and telephone recordings.

	Clinic Recordings	Telephone Recordings	*t*	*df*	*p*
Number of Pauses	47	50	-1.04	36	0.30
Mean Pause Duration (s)	0.63	0.65	-0.69	36	0.49
Proportion of Recording in Silence	0.20	0.21	-0.83	36	0.41
Mean Utterance Duration (s)	2.51	2.48	0.19	36	0.85
Total Recording Time (s)	144.64	151.76	-1.34	36	0.19
Total Length of Pauses (s)	29.57	32.35	-0.97	36	0.34
Total Length of Utterances (s)	115.08	119.41	-1.15	36	0.26

Columns 'Clinic Recordings' and 'Telephone Recordings' show average feature values for all participants for both sets of features separately. The remaining three columns show Student's t-test analysis results.

6.1.3 Discussion

Only two modifications were required to the feature extraction algorithm to accurately segment telephone speech recordings into pause and speech segments. Speech can be easily remotely recorded over a telephone and, as shown above, speech features can be reliably (overall accuracy of 93.2%, sensitivity of 97.3%, specificity of 89.5%) extracted from the telephone recordings. Average differences between features extracted from clinic environment recordings and telephone recordings for 19 subjects are under 15% for all features, except the *Total Length of Pauses* where the difference is around 19%. These differences were found not to be statistically significant. The non-significant differences may be attributed to the natural intra-speaker variability over time. The results demonstrate that the telephone speech quality recordings can provide adequate speech features for cognitive function assessment when compared to those extracted from recordings acquired in clinic.

All recordings were conducted by a trained neuropsychologists telephoning the participants and asking the participants to perform specific tasks, such as *Word Recall, Digit Span, Category Fluency*, to assess their level of cognitive function. Further development of the protocol and automation of the tasks lead to a fully automated Interactive Voice Response (IVR) system for cognitive function assessment, similar to that presented in study of voice pathologies by Moran et al [300] and Wormald et al [301]. Such a system would not necessitate the need for trained human experts, allowing expertise to be redirected towards intervention as opposed to assessment of cognitive function, and will allow monitoring of cognitive function of large scale cohorts. Design and evaluation of such system is presented below.

Key points

- The temporal speech features can be reliably (overall accuracy of 93.2%, sensitivity of 97.3%, specificity of 89.5%) extracted from speech recordings made over a telephone.
- Non-significant differences were demonstrated between the high-quality clinic recordings and the lower quality telephone recordings.
- The results demonstrate that the telephone speech quality recordings can provide adequate speech features for cognitive function assessment when compared to those extracted from recordings acquired in clinic.

6.2 Study 2 – Automated Interactive Voice Response application for remote assessment of cognitive function

The previous study described the required modifications of the feature extraction algorithm for use with telephone quality recordings. This study describes the automation and remote delivery of cognitive function assessment over the telephone.

Automation and remote delivery of a clinically valid cognitive assessment would provide longitudinal monitoring for large cohorts of older adults at relatively low cost and provide a repeated measure of cognitive function.

Regular monitoring of cognitive function in individuals may enable the early identification of cognitive decline and facilitate early interventions that may reduce the impact of cognitive decline on an individual. Regular monitoring of cognitive function is a method to early identify changes in cognitive function. However the nature of the face to face cognitive assessments makes regular monitoring of cognitive function highly impractical, imposing financial and labour burden on the health care providers. This is particularly relevant in light of the predicted increase in the number of older adults. However, regular monitoring of cognitive function is an ideal application for an automated system of longitudinal data collection.

In the study presented here, cognitive assessment was implemented over the telephone using an Interactive Voice Response (IVR) system. Interactive Voice Response (IVR) systems enable callers to interact with a computer system via a telephone. The cognitive assessment in this study automated several neuropsychological tasks administered in a typical battery of cognitive tests. Speech is employed as the input mode to make the assessment more natural for the user. This study focused on the design, implementation, success and usability of an automated remote cognitive assessment with older adults in their own homes.

The hypothesis of this study was that the delivery of cognitive function assessment would provide the same quality of assessment as a face to face interview with a neuropsychologist, if the cognitive test employed were similar. Specific questions were asked to address this hypothesis.

This study investigated whether having an initial face to face clinical assessment prepares the participants for the automated assessment resulting in better completion rates. This study investigates the reliability of the task being administered repeatedly over time and employing different methods of administration.

6.2.1 Methods

Cohort

Sixty-six participants have been recruited for this study, from which 60 (41 women) completed the study with a mean age of 69.99 (± 5.98) years, Mini-Mental State Examination (MMSE) (see Chapter 2, Section 2.9.1) score of 28.15 (± 1.25) and 15.26 (± 3.73) years of education.

One aim of this study was to investigate whether having a face-to-face assessment prior to IVR assessments would influence one's performance. In order to do this the cohort was split into two groups and a crossover experimental design employed. Group1 performed two IVR based cognitive assessments followed by a face-to-face (F2F) assessment. Group2 started with a face-to-face assessment followed by two IVR based assessments (Table 6-2). The interval between two adjacent assessments ranged from five to seven days.

Table 6-2: Participants group distribution.

	Assessment 1	Assessment 2	Assessment 3	N
Group 1	IVR1	IVR2	F2F	28
Group 2	F2F	IVR1	IVR2	32

Group distribution of participants that completed all three assessments.
IVR1 – first IVR assessment, IVR2 – second IVR assessment, F2F – face-to-face assessment, N – number of subjects.

Prior to each IVR interview each participant received an information pack (see Figure 6-1). This pack contained a checklist of conditions under which to carry out the interview, i.e. choose a quiet room, keep distractions to a minimum (no TV or pets), have a glass of water to hand. Some tasks in the assessment required visual aids, e.g. images for picture description and phonological cues for the diadochokinetic rate task. The interview guide consisted of five laminated A4 size pages bounded on top for easy turning and page numbers for ease of reference (see Figure 6-2). The information pack also included a unique four-digit identification number (ID) and a toll-free telephone number with which to make a telephone call to the IVR system.

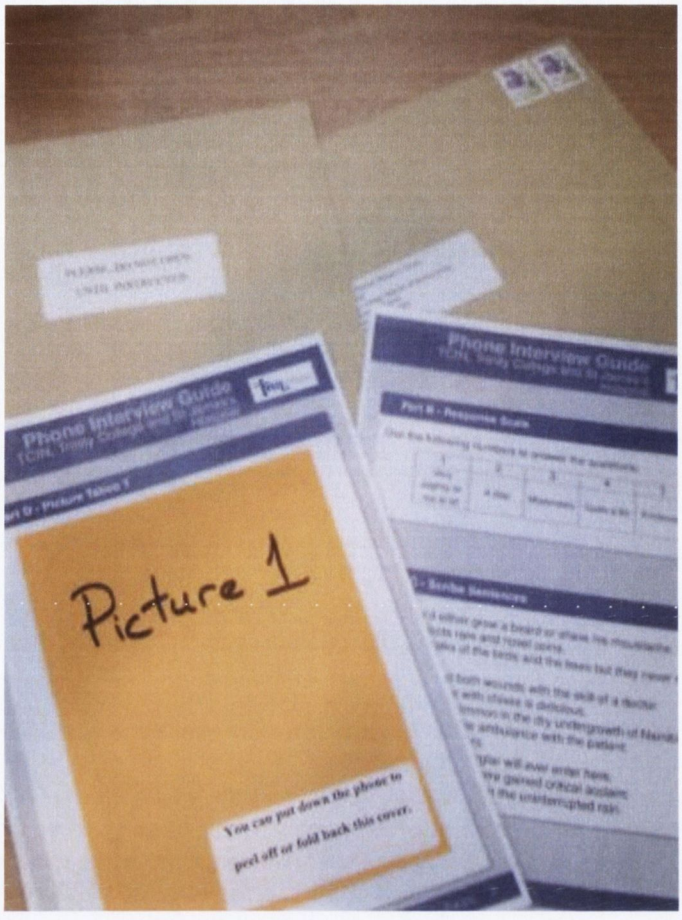


***Figure 6-1**: The interview pack received by each participant.*

The St. James's Hospital Ethics Committee, Dublin, approved the study. Participants completed the IVR assessments in their own homes and the face to face (F2F) and post assessment interviews in Trinity College Dublin. Trained neuropsychologist carried out all F2F assessments.

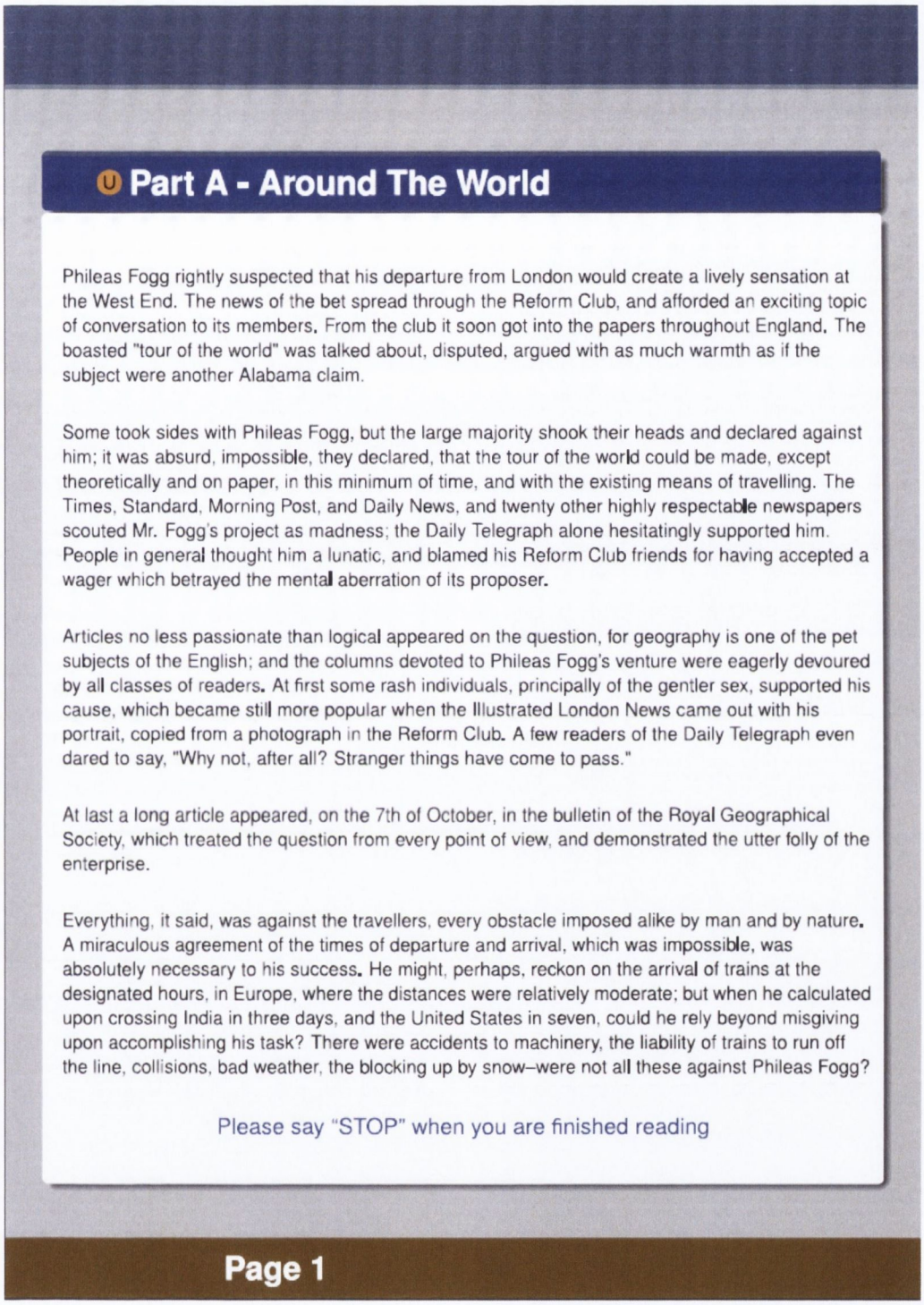


***Figure 6-2**: Sample page of the interview guide used during the IVR assessment.*

Tasks

Each participant was asked to complete an array of cognitive assessment tasks. These included cognitive tasks traditionally administered face to face by neuropsychologist in the clinic. The cognitive tasks included in the assessment aimed to test different brain functions as would be employed in a clinical battery of cognitive tasks. These specific tasks were:

The *Days of the week backwards* task is a modified task of Osberg et al. [330], during which the participant is required to name the days of the week in reverse order. This task is aimed at assessing declarative and working memory processes of the brain.

The *Word Recall – Immediate* and *Delayed* – were included as a test of short- and long-term memory (see Chapter 2, Section 2.9.3). To compensate for a practice effect, a different word list (animals, fruits, vegetables) was employed for each of the assessment interviews.

To assess semantic memory performance, the *Category Fluency (CF)* task was included in the assessment (see Chapter 2, Section 2.9.2). To compensate for a practice effect, a different category (animals, fruits, vegetables) was employed for each assessment interview.

The *Vigilant Auditory Attention Task (VAAT)* was included to assess participants' ability to concentrate. It was an auditory version of the Sustained Attention to Response Task (SART) presented in Chapter 2, Section 2.9.4. During this task, the participant was orally presented with numbers from one to nine, repeatedly 25 times. The participant was required to say 'tap' for every number except the number 'five'.

A short version (10 items) of the *Positive and Negative Affect Scale (PANAS)* as presented in Chapter 2, Section 2.9.5 was included to assess participant's mood state.

Several speech tasks were also included in this study to assess their suitability for inclusion in an automated assessment. Speech tasks were the *Picture Description* task introduced in Chapter 5, Section 5.3.1, the *Text Passage* tasks introduced in Chapter 5, Section 5.1.1 and also a task to assess articulation. The measure of articulation was the diadochokinetic (DDK) rate [331], required participants to repeat specified syllables (PAH, TAH, KAH) or set of syllables (PAH-TAH-KAH or KAH-TAH-PAH), as fast as possible and as clearly as possible, for the duration of one breath.

The IVR application design

The automated assessment was designed to be as user friendly as possible and to reduce the apprehension of carrying out a cognitive assessment and/or using an IVR system. The voice responses from the participants were employed to navigate through the assessment. This removes the need for the participant to find and press the required DTMF key on the telephone's keypad. Thus, the designed and developed IVR application was unimodal and reduced the cognitive load (switching between listening to the instructions and entering answers using the keypad) on the participant during the telephone call. Additionally, keeping

the telephone constantly by the ear means instructions presented by the IVR application are not missed by the participant.

At the beginning of each call, the participants were required to authenticate themselves by saying their four-digit ID provided to them in their information pack. The participants were allowed three attempts to authenticate themselves by voice. If the ID was not recognized by speech input, the participant had an opportunity to authenticate by entering the ID with DTMF input. Following successful ID authentication the participant was asked to verify his/her first name. After successful name confirmation, additional checks were carried out before the assessment began. Participants were asked if they had a glass of water and the interview guide to hand. Finally participants were given the opportunity to adjust the volume level of the pre-recorded prompts. Answers to each of these specific questions (simple yes/no answers) were provided by speech response only and processed by the speech recognition engine.

User-centred design

Each task in the cognitive assessment battery was designed as a separate module within the application. Each module was designed to be consistent with the manner in which the task is normally administered by a neuropsychologist as much as possible while adhering to standard voice user interface design principles [328]. Prior to implementation of the design, a Wizard of Oz (WoZ) user experience experiment [79] was carried out with 12 older persons. WoZ experimentation is a common tool in the fields of experimental psychology, human factors and usability engineering as part of an iterative design methodology. A typical WoZ experiment employs a "wizard" who, under laboratory conditions simulates the behaviour of a theoretical automated computer application.

In the WoZ study, the cognitive assessment tasks were simulated by employing pre-recorded instructions played over the telephone to the participants. Following each WoZ assessment, feedback on the IVR experience was collected from participants. The findings from these interviews had a direct influence on the user centred design of the automated IVR cognitive assessment. This exercise provided an opportunity to ensure that the prompt texts balanced the amount of information contained without being too long and that participants understood the requirements of each task. The WoZ exercise also allowed the assessment of time-out parameters (indications of when a person has finished speaking in order to proceed to the next prompt/task) and the speed at which prompts are delivered.

Technology

The IVR application was hosted on a dedicated server purchased from company Voxeo [332]. The server employs VoiceXML and CCXML scripting languages commonly used in IVR solutions. The application was designed using the Evolution software package provided by Voxeo. Voxeo's proprietary automatic speech recognition (ASR) engine processed the responses of participants. ASR grammars are generated from the list of possible answers to each question.

The instructions employed by the IVR application were recorded by a professional voice artist in a sound recording studio and normalized to five volume levels (0dB, -6dB, -12dB, -15dB, -18dB). Before the assessment began the participant was given the opportunity to change the volume level by saying "up" or "down", or keep the default volume level.

Call flow

At the beginning of each task, instructions on how to complete the task were played to the participant. The participant was then asked if they needed more instructions or if the instructions were clear enough for the task to begin. If the participant required more instructions, more detailed instructions were played to the participant. The participants were then asked to confirm that they understood these more detailed instructions, and then the task began. In the case where the participant did not understand the more detailed instructions the IVR application moved on to the next task. Each task had a pre-defined time-out interval implemented to allow for non-responsiveness.

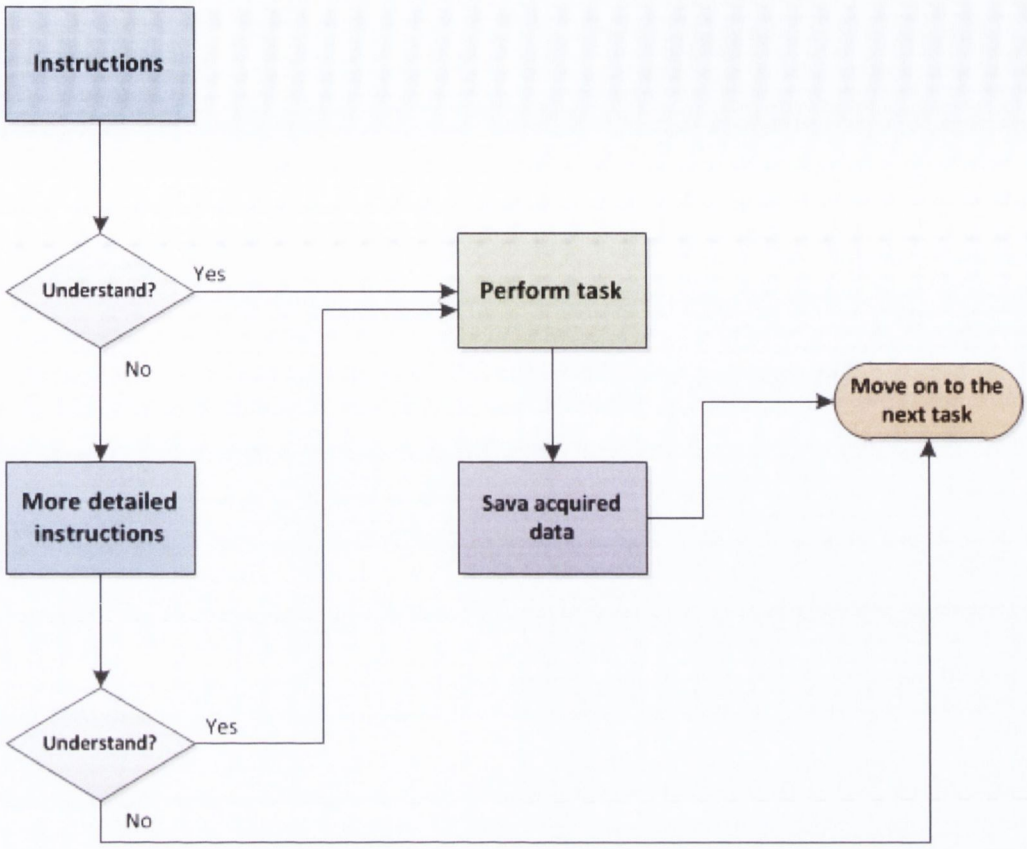


***Figure 6-3**: IVR application call flow diagram*

A pause feature was included in the IVR application and allowed the participant to put the telephone down to open the interview guide and say "start" in order re-start the interview. Reminder of the voice command for resuming the actions was re-played to the participant every 15 seconds. This facility was also used in other parts of the assessments to allow participants take a break and take a drink of water.

Data acquisition

The speech responses of the participant were recorded for all tasks during the cognitive assessment. Both channels of the call (participant and the IVR application) and separate channels in stereo were recorded (μ-law codec, 8 kHz sampling rate, 16-bit resolution). Each recording was saved with specified file name, which included participant's ID, date, and time when the recording was created, and identification code of each task. After the telephone call ended, all recordings were automatically transferred from the IVR server to a storage server.

As one of the aims of this study was to validate the hypothesis that the IVR assessment provided the same quality of results as an assessment with a neuropsychologist in the clinic, the clinical assessment followed exactly the same protocol as the IVR assessments and used

the exact same instructions. All speech responses from participant were recorded with 44.1 kHz sampling rate and 16-bit resolution in an uncompressed format.

Data processing

Scoring of each task was carried out through automated processing and also hand scoring. The features of interest from the speech tasks were extracted automatically. Automatic scoring of the *Vigilant Auditory Attention Task (VAAT)* and *Positive and Negative Affect Scale (PANAS)* was also possible. While the *Word Recall – Immediate (WRI)* and *Word Recall – Delayed (WRD)* tests and the *Category Fluency (CF)* tests were captured using speech input and a speech recognition engine transcribed the captured words.

Evaluation of reliability

Evaluation of reliability between assessments of the standard neuropsychological tests was carried out. The scores of the *Word Recall – Immediate, Word Recall – Delayed* and *Category Fluency* tasks were transformed to z-scores to compensate for use of different word lists during *Word Recall – Immediate* and *Word Recall – Delayed* tasks and different categories during *Category Fluency* task. Three different word lists and categories were needed due to close temporal administration of the face to face and IVR tests. Internal consistency of the individual tasks was assessed by calculating the Cronbach's α coefficients [333] and Pearson's correlations within each participant's group. A Cronbach's α coefficient higher than predefined threshold of 0.6 is considered reliable [334].

Evaluation of user experience

User experience interviews were conducted between the participants and the neuropsychologist the following day upon completion of all three interviews. The participants completed a user experience interview focused on three main areas; instructions, functionality and overall experience. This interview aimed to assess how well the IVR application performed and how well participants were able to navigate through the assessment.

Cost assessment

The cost assessment was estimated for a group of 400 individuals using 30 minutes long automated IVR cognitive assessment twice a year. The number 400 was estimated as an approximate number of subjects to which one psychologist may be capable of administering a cognitive assessment twice a year.

6.2.2 Results

Sixty participants successfully completed all three assessments. Of the six participants who dropped-out of the study, one participant was unable to participate due to worsening health condition. Two participants did not complete any assessment and dropped-out without specifying a reason. Three participants dropped-out due to technical difficulties with the IVR application during their first telephone call. All remaining participants completed the interview, however not all participants completed each task successfully.

The *Word Recall - Immediate* task was only considered complete when participants completed three recall iterations. Of the 28 participants in Group1 two did not attempt the *Word Recall - Immediate* task at all, and two did not complete one of the three iterations for the first IVR assessment. For the second IVR assessment one participant did not complete the *Word Recall - Immediate* task at all and one participant missed a recall iteration of *Word Recall - Immediate*. For the 32 participants in Group2, five missed one iteration of word recall in the first IVR and six missed one recall iteration in the second IVR. Only three participants in the whole cohort missed a single iteration in both IVR assessments. The completion rates for the *Word Recall* tasks are shown in Figure 6-4 for Group1 and Figure 6-5 for Group2.

For the *Word Recall - Delayed* task 6 participants from Group1 did not complete the task for the first IVR assessments and 8 for the second IVR assessment. For Group2, five participants did not complete the task for the first IVR assessments and six for the second IVR assessment, Figure 6-4 & Figure 6-5.

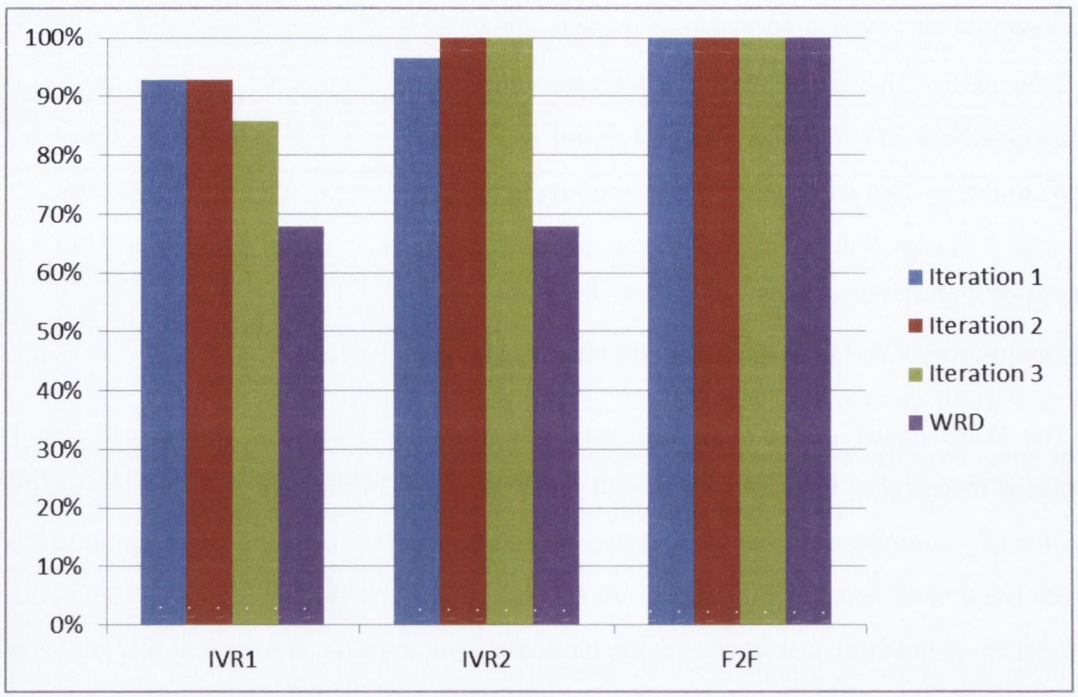


***Figure 6-4**: Word Recall - Immediate and Word Recall – Delayed (WRD) completion rates for Group1*

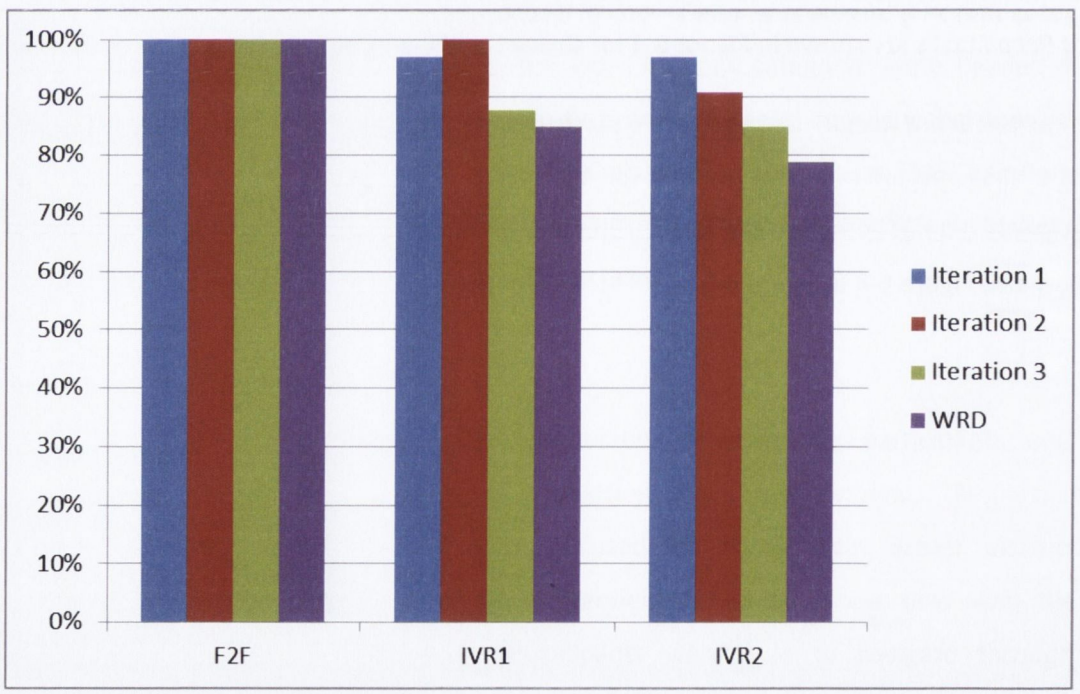


***Figure 6-5**: Word Recall - Immediate and Word Recall – Delayed (WRD) completion rates for Group 2*

For the category fluency test only one participant from Group1 failed to complete the task for the first IVR but completed it successfully in the second IVR. For Group2, two did not complete the *Category Fluency* task in the first IVR assessment and all completed the task in the second IVR assessment.

All participants completed the *Days of the Week backwards* task for all three assessments.

For the *Vigilant Auditory Attention Task (VAAT)* task, all participants in Group1 completed the practice and full task for the first IVR assessment and for the second assessment 27 completed the practice and 26 completed the full task. For Group2, 29 participants completed the practice and 28 completed full task for the first IVR and for the second IVR 26 completed the practice and 23 completed the full task, Figure 6-6.

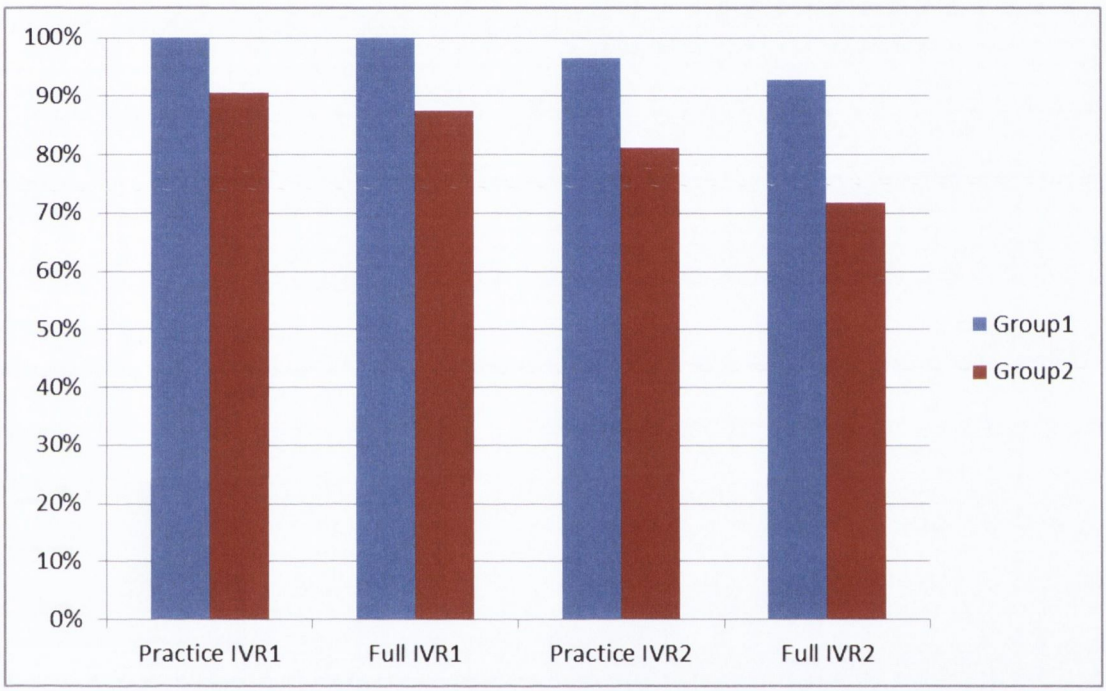


***Figure 6-6**: Percentage of complete VAAT tasks per group*

All participants completed the read and picture description tasks for all assessments regardless of the order in which they were administered.

For the diadochokinetic rate task all participants completed all versions of the task for the first IVR, regardless of the order of delivery of assessments. All participants in Group1 completed the task for the second IVR assessment but in Group2 three participants failed to complete the task in the second IVR assessment.

Evaluation of reliability

The evaluation of reliability between the clinical face to face assessment and the IVR based assessments yielded following results.

The results of the *Word Recall - Immediate* task were found to be reliable between all assessments for Group 2 and achieved Pearson's correlation coefficients statistically significant at $p<0.01$. The *Word Recall - Delayed* task was also found to be reliable between all assessments for Group 2 and also between the IVR assessments for Group 1. Statistically significant correlation was achieved between IVR assessments for both groups ($p<0.01$ for Group 1, $p<0.05$ for Group 2) and between IVR1 and F2F assessment for Group 2 ($p<0.01$).

The *Category Fluency* task achieved reliability between both the IVR assessments for both participants groups, and also between IVR2 and F2F assessment for Group 1. Pearson's correlation coefficients were statistically significant at $p<0.05$ between IVR assessments for both Group1 and Group2, and at $p<0.01$ between IVR2 and F2F assessment for Group 1.

The *Vigilant Auditory Attention Task (VAAT)* task was found to be reliable for both groups and between all assessments, except between F2F and first IVR assessment for the Group 1. Statistically significant correlation was achieved between IVR2 and F2F assessment for Group 1 ($p<0.05$), and between IVR assessments for Group 2 ($p<0.01$).

The *Days of the Week backwards* task was found to achieve a reliable Cronbach's α coefficients and statistically significant correlations ($p<0.01$) only for Group 1.

Table 6-3: Reliability measures between F2F and IVR assessments.

Task	Group	IVR1 vs. IVR2	IVR1 vs. F2F	IVR2 vs. F2F
WRI	Group 1	0.56 (0.39)	0.57 (0.40*)	0.56 (0.39*)
	Group 2	**0.81** (0.68†)	**0.79** (0.65†)	**0.81** (0.68†)
WRD	Group 1	**0.86** (0.78†)	0.28 (0.16)	0.54 (0.37)
	Group 2	**0.80** (0.67*)	**0.78** (0.64†)	**0.63** (0.48)
CF	Group 1	**0.65** (0.49*)	0.47 (0.31)	**0.75** (0.60†)
	Group 2	**0.70** (0.54*)	0.55 (0.38)	0.43 (0.27)
VAAT	Group 1	**0.78** (0.64)	0.32 (0.20)	**0.86** (0.80*)
	Group 2	**0.94** (0.99†)	**0.63** (0.79)	**0.82** (0.88)
Days of the week backwards	Group 1	**0.95** (0.90†)	**0.62** (0.51†)	**0.77** (0.67†)
	Group 2	0.13 (0.21)	0.44 (0.28)	-0.03 (-0.04)

All values are Cronbach's α (Pearson's correlation) coefficients.
WRI – Word Recall Immediate, WRD – Word Recall Delayed, CF – Category Fluency, VAAT – Vigilant Auditory Attention Task
IVR1 – first IVR assessment, IVR2 – second IVR assessment, F2F – face-to-face assessment
Reliable Cronbach's α coefficients in bold text.
** Pearson's r correlation significant at 0.05 level. † Pearson's r correlation significant at 0.01 level.*

Evaluation of user experience

From the original cohort 47 participants completed the user experience interviews with the neuropsychologist. The level of details presented in the instructions was queried: "Were the task instructions understandable?", "Were the error recovery strategies clear?", "Was the option of extra help for instructions clear?" and "Did the instructions make the participant feel confident in completing the tasks?".

All of the participants reported that, in general, they easily understand the instructions presented by the IVR application. Seven participants had difficulties understanding only one specific instruction for a task (different for each participant). All understood that more detailed instructions could have been requested from the IVR application and twenty one (44.7%) participants asked for more detailed instructions at least once during the assessment. Only one participant reported feeling lost or worried during the assessment (during the *Word Recall* task).

The functionality aspect was more concerned with how well the speech recognition module preformed and how seamlessly the overall assessment was presented. Participants were asked if they felt their voices were understood and if they experienced any moments of

uncertainty with regards what to do next. In terms of recognition performance, four participants reported that they felt their voice was not understood well by the automated system. Six participants had to repeat one answer during the whole assessment. Having to repeat an answer twice during the assessment was reported by four participants and seven participants reported repeating an answer more than twice. Too many pauses between the instructions and a need to wait during the assessment were issues reported by five participants, two of these participants did not feel stressed to wait longer.

The overall experience section of the user experience interview was concerned with how participants felt interacting with a machine for such a task. The potential for using such a system as a regular monitoring tool was also queried. All participants, except four, claimed to have enjoyed the study. There were mixed responses to how participants felt about working with a machine/automated system. Only two participants were adamantly unhappy about it and 13 participants would specifically prefer to undergo a face-to-face assessment. However three participants felt more relaxed carrying out the assessment at home; one participant reported "feeling like talking to a person" during the IVR assessment and another found it easier to undergo the IVR assessment rather than the FF assessment. Two participants stated that they preferred the anonymous nature of the IVR assessment and felt that they could answer honestly compared to face-to-face cognitive assessment. Finally, all participants, except three, would take part in such automated telephone assessment at regular intervals, if their doctor advised them to do so.

Cost assessment

The cost for delivering a 30-minute cognitive assessment twice a year to 400 participants using an automated IVR system was estimated. The operational cost for hosting the IVR platform is €10,990 per annum, which covers use of the platform, provision and maintenance of the server in the data centre. There is one-off charge of €1,570 for configuration work. The two toll-free lines cost €0.26 per minute of use. The total cost for the first year of operation would be €18,840. Each following year of operation would cost €17,270.

The IVR system is at present configured to process two telephone calls concurrently. This number of concurrent telephone lines can easily be increased; making the system scale exponentially with cost increased only by that of the extra telephone calls. The cost of the IVR assessments would be considerably lower than the cost of in-person assessments, when

compared to the median annual salary of €48,000 of an experienced clinical psychologist in Ireland as of December 2013 [335].

6.2.3 Discussion

This study has presented methodology for designing and implementing an automated cognitive assessment over the telephone using natural speech. The successful completion of common cognitive tasks implemented via an IVR system for a cohort of older adults was discussed. While several tasks, namely both *Word Recall* tasks and the *Vigilant Auditory Attention Task (VAAT)* task, were not completed in full, all participants managed to reach the end of the assessment regardless of whether they were able to complete each individual task. This indicates that the recovery strategies implemented are effective in dealing with errors in individual task modules.

The completion rate for tasks varied across the IVR assessment. More participants failed to complete the *Word Recall - Immediate*, *Word Recall - Delayed* and *Vigilant Auditory Attention Task (VAAT)* tasks more than any other task in the assessment. Several participants also failed to complete the full articulation task. There are several possible reasons for this, the completion of these tasks requires several interactions with the participant not only are they dealing with the fact that they are using an IVR but they are completing serious and sometimes complicated cognitive tasks. The more steps involved in completing a task, the more likely an error is to occur.

The study cohort was divided into two groups to investigate whether having an initial F2F interview would impact one's ability to complete the cognitive tasks in the IVR. Figure 6-4 and Figure 6-5 show that the completion rate for *Word Recall - Immediate* and *Delayed*, appear to improve with subsequent IVRs for Group1 but not for Group2. Potentially the transfer of the assessment from F2F to IVR has a significant effect on users. Another interesting finding is the completion rates for the *Vigilant Auditory Attention Task (VAAT)* and *diadochokinetic (DDK)* tasks appear to reduce with the second IVR for both groups. Participants considered these two tasks particularly boring and tiresome. The full completion rates for the first IVR and reduced completion for the subsequent IVR suggest that participants knew what the task was and did not what to complete it, while in the first IVR they were interested to know what they had to complete.

In terms of order of delivery; including a F2F assessment before IVRs does not appear to increase the chances of completing the cognitive tasks. For the word recall tasks an

improvement in completion rates is seen for Group1 between IVR1 and IVR2, while Group2 actually appears to do worse in IVR2 compared to IVR1. There is a gradual decrease in completion rates for the *Vigilant Auditory Attention Task (VAAT)* task for Group1 and a more significant decrease for Group2.

Another potential reason for the low completion rates observed for *Word Recall - Immediate* task is that this is the first task participants encounter in the IVR. Perhaps commencing the IVR assessment with a dummy task may provide some experience and confidence in using an IVR application. The improvement in completion rates for IVR2 (Group1) and the similarity in *Word Recall - Delayed* task completion rates between IVR assessments support this theory.

The analysis of the completion rates highlights certain tests that prove more difficult to complete than others in this automated environment. This information provides insight into which type of tests are more suitable for inclusion in such an automated assessment. Shorter tasks that would increase the task completion rate should be the focus of such automated assessments.

Evaluation of reliability

The *Word Recall - Immediate* and *Word Recall - Delayed* tasks achieved high reliability for Group 2 in all cases. Except between the IVR2 and F2F assessment where a lower, but still reliable, Cronbach's α coefficient was achieved. Group 1 achieved reliable Cronbach's α coefficient only between the IVR assessments for the *Word Recall - Delayed* task. An IVR effect has been found between the two groups. This effect observed that participants who carried out the IVR assessment first performed poorer than those who carried out the F2F assessment first. This is potentially due to the *Word Recall - Immediate* task being the first task presented to the participants and potentially confusing them. Undertaking one face-to-face assessment initially, during which the participant would get familiar with the cognitive tasks, prior the IVR assessments may be adequate. From the results of Group 2, we can see that using IVR technology is a stable manner of administering the assessment.

While the use of different categories for the *Category Fluency* task was compensated before calculating the reliability measures by producing z-scores, the *Category Fluency* task achieved reliability only between IVR assessments for both participant groups, and between IVR2 and F2F assessment for Group 1. Higher reliability may be achieved by repeating exemplar production of the same category over longer period of time, which would remove

any issue associated with a practice effect. Reliability measures would then be calculated by using the scores of the same category only.

High reliability was achieved for the *Vigilant Auditory Attention Task* between all assessments, except between IVR1 and F2F assessment for the Group 1. It should be noted that no commission errors were made during any of the three assessments by 45.8% of participants and 87.5% of participants scored no errors during at least one assessment. This task may be too easy to perform and may not challenge sufficiently participant's sustained attention.

The *Days of the week backwards* did achieve sufficient reliability between all assessments only for the Group 1. None of the participant had difficulties performing this task and it may not challenge the declarative and working memory processes of the brain of the participants recruited for this study. Therefore, the reliability results for this task may not be relevant.

Evaluation of user experience

The results of post assessment usability interviews have also been presented in this chapter. Feedback from Wizard of Oz (WoZ) and usability studies was used to ensure the application was designed specifically for the target users, i.e. older adults. The successful completion of the automated assessment for all participants implies that the final design accomplishes these goals.

The majority of participants had a positive experience with the automated IVR system and 93.6% of them would continue to take part in such cognitive assessment again, if their doctor advised it. The instructions, presented by the IVR application, appear to have been sufficiently well designed and only few participants reported difficulties with understanding instructions. This high satisfaction with the concept and functionality of the assessment is vital if similar remote assessments are to be used in everyday life.

While the user experience was positive, it has to be noted that only three participants had an MMSE score of 26 or lower. The score of 27 was defined in this study as the cut-off value for cognitive impairment. Future inclusion of larger numbers of cognitively impaired individuals may provide more valid overview of user experience and usability of this automated system by more cognitively impaired older adults. It should be noted that the three participants that would not undergo an IVR assessment again were among the cognitively healthy participants.

Longitudinal monitoring is the key to identifying the early onset of cognitive decline, in order to implement early interventions and improve quality of life. With the predicted increase of older population this will only be achieved by automation of cognitive assessments. This study demonstrated that the IVR technology can be a suitable platform for automated remote assessment of cognitive function. With further automation of scoring of included tasks, a fully automated system may be achieved. Such system would allow longitudinal monitoring of cognitive function for large scale cohorts and would complement the current methods for neuropsychological assessment of cognitive function in older adults.

Key points

- All participants managed to complete the IVR cognitive assessment interview.
- The implemented recovery strategies are effective in dealing with errors in individual task modules of the IVR application.
- Including a F2F assessment before IVR assessment does not appear to increase the chances of completing the cognitive tasks.
- Commencing an IVR assessment with a dummy task may provide some experience and confidence in using an IVR application for the participant.
- The use of the IVR technology for remote fully-automated delivery of cognitive function assessment interviews was found to be reliable across all sessions for three out of five cognitive tasks for Group 2 and for one task for Group 1.
- The majority of participants had a positive experience with the automated IVR system and 93.6% of them would continue to take part in such cognitive assessment again.
- The cost of the IVR assessments would be considerably lower than the cost of face-to-face assessments.

6.3 Summary

The studies in this chapter answered the Research Questions 11 - 19 posed in Chapter 2, Section 2.18.4.

The Study 1 of this chapter demonstrated that features extracted from speech in the assessment of cognitive function can be reliably extracted from telephone quality speech.

In the Study 2, the use of the IVR technology for remote automated delivery of cognitive function assessment interviews has been demonstrated to be functional and well received by the population of older adults. However the few shortcomings of the assessment will provide valuable lessons in designing cognitive tasks specifically for IVR administration.

This assessment is envisaged as a monitoring tool for cognitive function. Extracting measures of cognitive function on a periodic basis increases the likelihood of detecting the onset of cognitive impairment early and thus improves the outcome of intervention protocols. The advantage of an automated assessment such as this is that it can be deployed on a large scale at a relatively low cost.

Importantly, the employment of IVR technology may allow to obtain cognitive measures of older adults on periodic basis and have a positive economic impact on the assessment of the cognitive function of the ever-growing population of older adults. Employing IVR technology for cognitive assessment delivery may as well resolve issues with difficult access of older adults to healthcare providers' facilities.

The next chapter, Chapter 7, presents study of an automated remote web-based assessment of memory and attention, two critical elements of cognitive function.

Chapter 7

Automated remote web-based assessment of memory and attention

The previous chapter demonstrated that remote assessment of cognitive function can be carried out over a telephone in means of recording speech input from the individual. While the possibility of delivering a cognitive assessment remotely over a telephone is desirable, not all domains of cognition can be assessed in this manner. This chapter presents the design and implementation of a system for remote web-based monitoring of memory and attention. This study investigated the validity of these measures of memory and attention, and the reliability of these repeated measures over time.

As discussed in the previous chapter, repeated assessment is common in clinical practice to aid decisions regarding diagnosis; monitoring of deterioration or recovery of function; or evaluation of therapeutic, pharmacological, or surgical interventions. With an ageing population, serial neuropsychological evaluations need to be employed in the diagnosis and monitoring of such conditions as dementia, as well as in longitudinal research, with the aim of detecting significant or clinically meaningful changes or decline [39].

However, an on-going issue that is of concern to every neuropsychologist or researcher with regard to serial neuropsychological evaluation is the effect of practice, whereby an individual's retest performance might be enhanced as a result of better test-taking strategies, familiarity with the test environment or the actual test content rather than reflecting genuine improvement in the underlying ability. Practice effects of different magnitudes have been demonstrated in numerous studies over intervals of weeks, months, or years depending on the choice of tests and the cognitive domain concerned [39].

Although it is well established that there are substantial differences in practice effects between cognitive domains, the domains most susceptible to practice effects have not been consistently demonstrated across studies. For example, Rabbit et al. [336] reported larger practice effects on tests of fluid intelligence than on tests of episodic and semantic memory, whereas Ferrer et al. [337] found relatively larger practice effects on composite memory measures than on composite measures of spatial reasoning and speed. Comparison of changes in different cognitive abilities is further complicated by variability in practice effects among tests within the same domain, suggesting that practice effects are not domain specific, but test specific [338, 339], and influenced by the number of re-administrations [337, 339, 340]. In studies with three or more repeated assessments, practice effects were generally found to be greatest in the second assessment for memory and speed measures [337, 340], whereas continuous gains were reported for up to the fourth assessment in measures of fluid reasoning and executive function [39, 341].

Some researchers [341] advocate the use of parallel forms (same test, but employing different stimuli) to eliminate the effect of practice, especially when a short retest interval is expected. This was employed in the study in Chapter 6 for remote Interactive Voice Response (IVR) assessment of cognitive function. However, the use of parallel forms does not completely eliminate practice effects, though the effect may be attenuated [342]. The fact that practice effects prevail in the face of different items suggests that other factors such as familiarity with the testing environment and test strategy are still operating. Parallel forms also raise the question of equivalence of the specific items [39].

The aim of the study reported here was to develop a battery of tests for reliable repeated assessment of memory and attention, which would not be prone from practice effects over time and could be administered at regular intervals over time. The battery of tests was delivered remotely to the participant, with the possibility to undertake the testing at home. The developed tests employed the same stimuli every time to avoid complications with equivalence of the specific items. The hypothesis is that individuals with low memory status perform worse on these tests delivered in this manner than individuals with high memory status. The validity of the tests was assessed by correlating the tests scores with scores from standard neuropsychological tests administered clinically.

7.1 Methods

As a part of this study a cohort of older adults undertook a battery of memory and attention assessment tests on a dedicated website for the study [343]. Participants were asked to complete two phases of testing. Phase 1 required the participant to complete an online web-based testing once every week, for a period of eight weeks.

7.1.1 Cohort

Twenty six older adults were recruited for this study. From these, ten participants did not complete all eight test rounds and dropped out of the study. The mean age of the remaining 16 participants was 70.00 (SD = 3.27) years. From the 16 participants, eleven were female and five were male. All participants were right handed. The cohort was divided into two groups – *Low performers* and *High performers* - based on the status of individual's memory (see Section 7.1.2). The complete demographic data split by group can be seen in Table 7-1. After complete description of the study to the subjects, written informed consent was obtained from all participants. The study was approved by the Ethics Committee of the School of Psychology, Trinity College Dublin.

Table 7-1: Demographic data for Low and High performers groups

Group	Gender	Age	Full School Edu.	Part School Edu.	Full Univ. Edu.	Part Univ. Edu.
Low performers	7 female, 1 male	68.63 (2.83)	11.00 (2.56)	0.50 (0.93)	2.88 (4.16)	1.63 (1.77)
High performers	4 female, 4 male	71.38 (3.25)	10.75 (4.77)	0.38 (1.06)	1.75 (1.49)	1.25 (1.91)

All values are mean values (standard deviation) in years, except for Gender, which is represented by total numbers.
Full/Part School Edu. – years of full/part time school education.
Full/Part Univ. Edu. – years of full/part time university education.

7.1.2 Standard neuropsychological assessment

All participants underwent cognitive neuropsychological assessment employing gold-standard neuropsychological tests. The following tests were administered to the participants by a trained neuropsychologist (see Section 2.9 for tests details):

- Mini-Mental State Examination (MMSE) [30],
- National Adult Reading Test (NART) [134],
- Wechsler Memory Scale (WMS) III [32] subtests:
  - Logical Memory,
  - Verbal Pairs,
  - Digit Span,
- Stroop test [344],
- Category Fluency (CF) [68]
- Hospital Anxiety and Depression Scale (HADS) [146].

Memory groups

The participants were divided into two groups, *Low performers* and *High performers*, based on their memory performance relative to an estimate of their pre-morbid IQ. The following method was used to decide if participant was a *Low performer* or *High performer*.

Standard score of Logical Memory (SsLM) was calculated as

$$SsLM = \frac{Scale\ score - 10}{3}$$

Eq. 23

Where 'Scale score' is the raw score of the Logical memory subtest of the Wechsler Memory Scale (WMS) adjusted for age according to the WMS manual [32].

Standard score of NART (SsNART) was defined as

$$SsNART = \frac{x - 100}{15}$$

Eq. 24

Where x = 128 – (0.83 * number of errors in the NART).

Z-scores were used to relate their performance on a standardized story recall memory test (SsLM) to their NART estimated IQ (SsNART). When the *Standard score of NART* was higher than the *Standard score of Logical Memory* and this difference was more than one standard deviation, the subject was defined as a *Low performer* (page 28 of [345]).

7.1.3 New battery of memory and attention tasks

Two new tasks were designed to address the limitations of existing tasks and included in the battery of tasks. First task, called *Word Recognition*, was employed to assess the participant's verbal recognition memory. The other task, called *Shape Recognition*, assessed the participant's non-verbal recognition memory. Two versions of these tasks were administered to the participant, an *Immediate* version that assessed participant's short-term memory and a *Delayed* version that assessed long-term memory. A third task included in the battery was *Sustained Attention to Response Task (SART)* [139], which was employed to assess participant's attention.

The tasks were administered in the following order:

1. *Word Recognition - Immediate (WRec-I)*
2. *Shape Recognition - Immediate (SRec-I)*
3. *Sustained Attention to Response Task (SART)*
4. *Word Recognition - Delayed (WRec-D)*
5. *Shape Recognition - Delayed (SRec-D)*

Word Recognition - Immediate

The *Word Recognition - Immediate (WRec-I)* task can be divided into two phases – Presentation and Recognition. The *WRec-I* task employs a fixed list of 20 nouns with mean length of 4.95 (SD = 1.36, range: 3 – 8) letters and mean normative frequency of 82.45 (SD = 108.44) occurrences per million in The Sydney Morning Herald word database [346]. The words are displayed on screen (see Figure 7-1) in Arial font, with font size of 24 points and font colour R:15, G:15, B:15 in RGB scale, against background with dimensions of 750 by 400 pixels and background colour R:175, G:175, B:175. The background was delimited by a 3 points thick border line of R:105, G:105, B:105 colour.

At the beginning of the task, the order of words in the list was randomized. The first ten words of the randomized list are displayed to the participant during the Presentation phase. Each word was displayed for 2000 ms, followed by a 1000 ms pause. Following the last word, all ten words were presented one more time at the same speed.

The Presentation phase is followed by the Recognition phase. At the beginning of the Recognition phase, the list of all 20 words was randomized again. The words were presented one by one to the participant. The participant was required to answer whether the displayed word was from the original list (displayed during Presentation phase) or a new word (from the

ten words that were not displayed during Presentation phase). The participant's answer was followed by a 1000 ms pause during which the current word was removed from the display and after which another word was displayed.

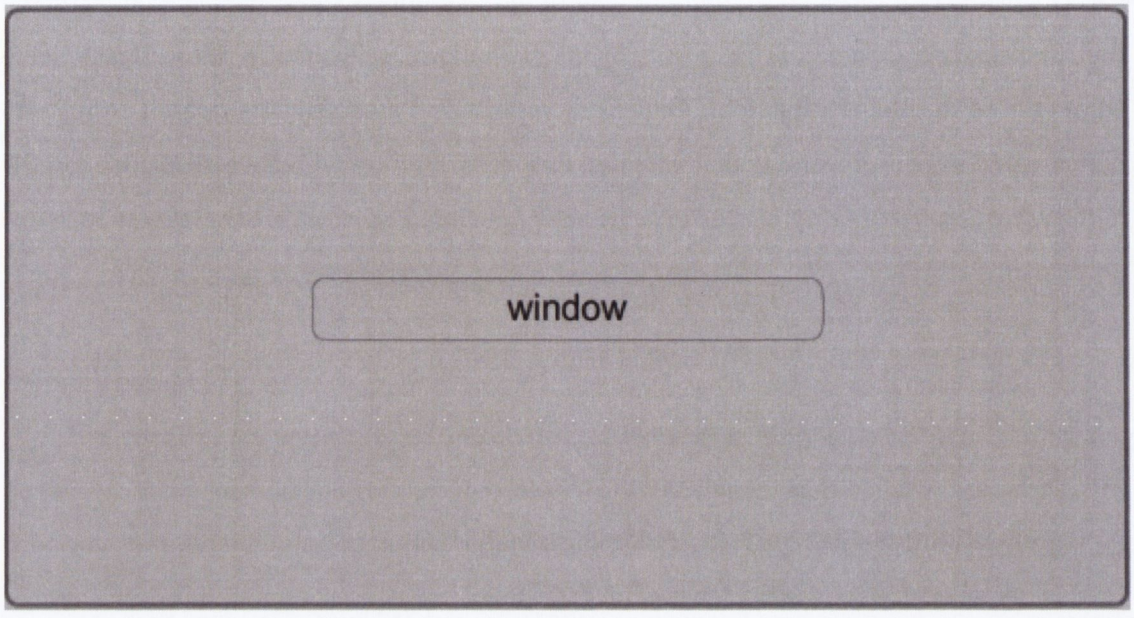


***Figure 7-1**: Graphical user interface of the Word Recognition task.*

Word Recognition - Delayed

The *Word Recognition - Delayed (WRec-D)* task included only Recognition phase of the *WRec-I* task. At the beginning of the task, the list of 20 words was randomized again, resulting in different display order of the words than during the initial Recognition phase of the *WRec-I* task. The task remained the same as during the Recognition phase of the *WRec-I* task.

The *Word Recognition - Immediate* and *Word Recognition - Delayed* tasks served as a measure of participant's verbal recognition memory, immediate and delayed.

Shape Recognition - Immediate & Delayed

The *Shape Recognition - Immediate (SRec-I)* and *Shape Recognition - Delayed (SRec-D)* tasks followed the same procedures as the *WRec-I* and *WRec-D* tasks, but instead of employing words these tasks employed shapes. The shapes had a width of 130 and height of 130 pixels. The line of the shape was 5 pixels wide and of black colour (R:0, G:0, B:0). The inside of the shape was fully transparent.

The shapes were generated in a manner that would not evoke association with words or common objects of the real world. Therefore, the task only assessed non-verbal recognition memory, immediate and delayed.

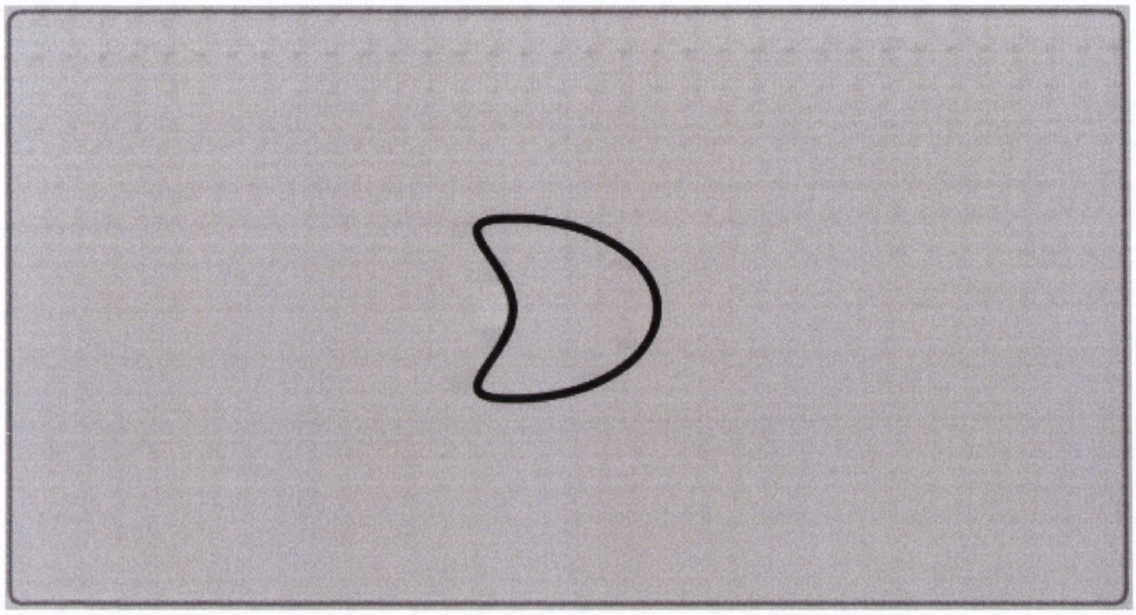


***Figure 7-2**: Graphical user interface of the Shape Recognition task.*

Sustained Attention to Response Task

To measure participant's sustained attention, the *Sustained Attention to Response Task (SART)* developed by Robertson et al. [139] was employed. The *SART* was adapted for use on the study's website. The *SART* required the subject to press key to frequently displayed non-targets, but with the requirement to withhold motor responses to occasional targets. The non-targets were represented by eight different shapes and the target by the shape of a star. The shapes were presented on the same background as was used in *Word Recognition* and *Shape Recognition* tasks. The height and width of each shape was 100 pixels. The line of the shape was 5 pixels wide and of black colour (R:0, G:0, B:0).

The non-targets were displayed 85 times and target 15 times, one by one in random order with at least two non-target shapes being displayed in between the target shapes. Both types of stimuli remained on screen for 500 ms, followed by a 1500 ms pause.

7.1.4 Data acquisition

All participants were contacted by telephone prior to the beginning of the study and received instruction by regular mail. The battery of tasks was delivered remotely using a dedicated website and the participants undertook the testing from their homes. Participants

were instructed to inform other persons living with them to minimise distraction while they carried out the web-based assessment.

The dedicated study's website (https://acad.tchpc.tcd.ie/) [343] was programmed in 'PHP: Hypertext Preprocessor' server-side scripting language [347]. The *Word Recognition*, *Shape Recognition* and *SART* tasks were programmed in ActionScript 3.0 language [348] used by the Adobe Flash Player platform [349] and embedded on the website in form of SWF files. All participants' data and tasks results were stored in a MySQL database [350], while sensitive participants' data were stored encrypted employing Triple Data Encryption Algorithm (TDEA) algorithm [351].

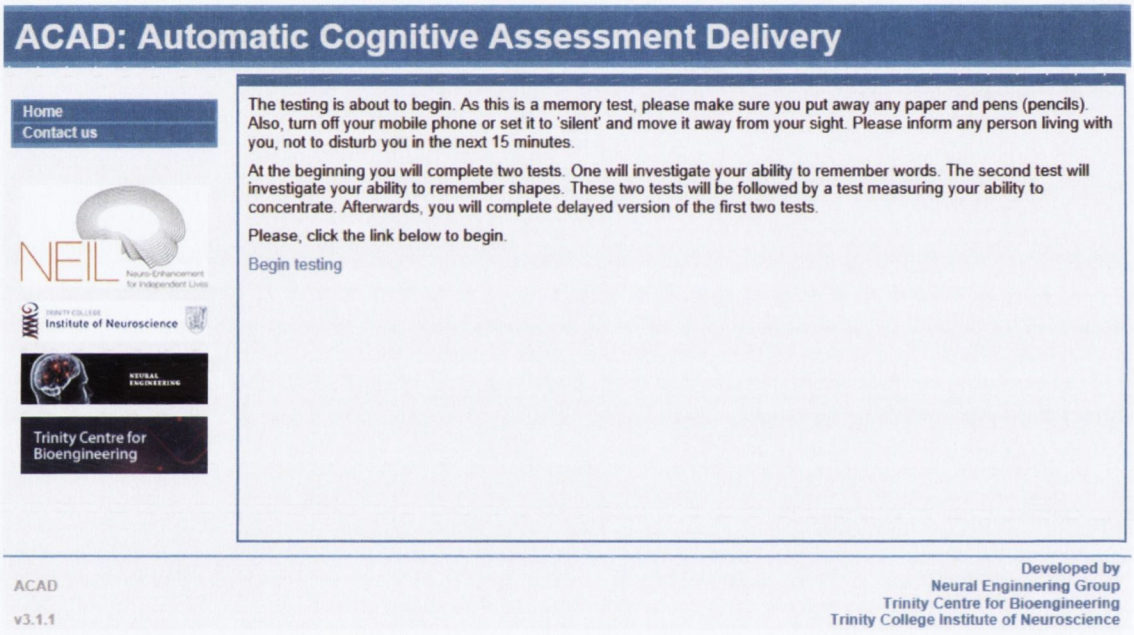


***Figure 7-3**: Website interface of automated remote web-based assessment of memory and attention.*

A dedicated username and password was generated for each participant. At their first login, the participants were requested to fill in short demographic form, collecting their age, gender, handedness, education details, mobile phone number, date and time of the week when they would like to undertake the testing. The mobile phone number was used to deliver automated reminders to the participants.

Following the login (or demographic form for the first time), the participant was directed to undertake the battery of tasks. Employing the Adobe Flash Player platform to run the tasks allowed reaction time measurements on participants' computers (instead of on the server side), while removing issues with delays occurring between participant's computer and the web-server. During all tasks, participants used 'Enter/Return' key to make positive answers

and 'Space' key for negative answers. Keys 'C' and 'R' were used for navigation through the instruction displayed on the screen. The participants were not required to use the mouse to avoid confusion with which device should be used for inputs.

On successful completion of all tasks, participants were directed to complete a subjective questionnaire for self-evaluation of memory and mood. This questionnaire was displayed as a standard HTML page, i.e. not using the Adobe Flash Player platform. The participants were requested to answer 'Yes', 'No' or 'Maybe' (drop-down menu selection) to seven queries. The following statements were used:

1. I am happy with my memory.
2. I am worried about my memory.
3. I feel stressed.
4. I feel depressed.
5. Someone told me my memory is getting worse.
6. I find it difficult to concentrate.
7. I think I need help with my memory.

The self-evaluation was followed by displaying information to the participants regarding date and time of the next testing.

The number of correct answers was collected for each of the recognition tasks. Reaction time for each response, number of *Omission errors* (missing key press to non-target) and *Commission errors* (key press to target) were collected for the SART task.

7.1.5 Data processing

The *percentage ratio of delayed to immediate correct answers* was calculated for each completed *Word Recognition (WRec-P)* and *Shape Recognition (SRec-P)* task.

$$WRec\text{-}P = 100 * \left(\frac{WRec\text{-}D}{WRec\text{-}I}\right)$$

Eq. 25

where *WRec-D* is the number of correct answers from the delayed version of the *Word Recognition* task, and *WRec-I* is the number of correct answers from the immediate version of the *Word Recognition* task.

$$SRec\text{-}P = 100 * \left(\frac{SRec\text{-}D}{SRec\text{-}I}\right)$$

Eq. 26

where *SRec-D* is the number of correct answers from the delayed version of the *Shape Recognition* task, and *SRec-I* is the number of correct answers from the immediate version of the *Shape Recognition* task.

Standard deviation of reaction time (SD_RT) and *Coefficient of Variation of reaction time (CoV_RT)* were calculated from all reaction times of the *SART* task for each task completed.

All measures of all tasks were automatically generated or calculated, effectively removing the need to calculate these measures manually and saving researcher's time.

7.1.6 Statistical analysis

To test the hypothesis if there is statistically significant difference in performance on the recognition tasks and the *SART* task between the *Low* and *High performers* group, paired two-tailed Student's *t*-test was applied to the mean values of correct answers and percentage ratios for each group and to each week of the study.

Validity

To assess validity of these online web-based tasks, Spearman's ρ rank correlation between all tasks measures and measures of standard neuropsychological tests was calculated. Tasks measures from *Week 1*, *Week 2* and mean value from *Week 1* to *Week 8* was used to calculate the correlation coefficients.

Analysis of practice effect

Assessment of practice effect was performed by fitting a linear regression line to the mean values of recognition tasks and *SART* measures from each week and for each group separately. Analysis of significance of the slope of the regression line was performed for all regression lines.

To estimate the magnitude of practice effect, Cohen's *d* coefficient [352] was calculated from tasks measures from each week. The practice effect was estimated between adjacent weeks, and between the *Week 8* and *Week 1*. The practice effect was considered minimal ($d<.15$), moderate ($.15<d<.70$) or large ($d>.70$) [39].

Significance of the regression line slopes and Cohen's *d* coefficients were calculated for each participant's group separately.

Test-retest reliability

Test–retest reliability of the recognition tasks and *SART* measures was estimated using non-parametric Spearman's ρ rank correlation between each week for each measure. The reliability of less than .60 was considered low, .60 - .69 marginal, .70 - .79 adequate and above .80 high [39, 68].

7.2 Results

Mean values for each week and each group separately were calculated for all recognition tasks measures. The values can be seen in Table 7-2.

Table 7-2: Recognition tasks - Mean values for each task measure and group separately

Task Measure	Group	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	*t*	*p*
WRec-I	Low	16.13	19.00	18.50	19.00	18.75	18.88	18.38	19.25	-0.83	0.430
	High	19.00	18.75	18.50	18.38	18.38	19.00	19.13	19.38		
WRec-D	Low	15.25	16.00	14.63	16.38	14.88	15.25	16.50	15.88	-4.89	0.002*
	High	18.38	18.00	17.38	17.00	16.88	16.25	17.13	17.00		
WRec-P	Low	96.26	83.97	79.60	86.07	78.70	80.81	89.65	82.27	-3.56	0.009*
(%)	High	96.88	95.73	93.79	92.43	92.07	85.50	89.19	87.79		
SRec-I	Low	17.50	17.00	17.63	17.13	17.25	17.63	17.25	16.88	-8.64	<0.001*
	High	18.88	17.88	18.25	18.00	18.00	18.38	18.50	18.38		
SRec-D	Low	15.00	15.25	15.13	15.50	15.63	15.50	16.50	14.50	-2.59	0.036*
	High	16.25	15.88	16.50	15.25	16.00	16.00	16.38	16.63		
SRec-P	Low	86.38	89.97	85.28	90.52	90.84	97.64	96.10	86.36	0.62	0.554
(%)	High	86.02	88.85	90.72	84.65	89.01	87.21	88.50	90.39		

Low = Low performers, High = High performers

** - task measure with statistically significant difference in mean values between groups, estimated employing paired 2-tailed Student's t-test*

Employing paired two-tailed Student's *t*-test, statistically significant differences were found in mean values of four measures between the *Low* and *High performers* group. The *Word Recognition – Delayed (WRec-D)* measure had statistically significant differences at p=0.002, *Percentage ratio* of delayed to immediate correct answers of the *Word Recognition (WRec-P)* task at p=0.009, *Shape Recognition – Immediate (SRec-I)* at p<0.001 and *Shape Recognition – Delayed (SRec-D)* at p=0.036. In all four measures the participants of the *Low*

performers group on average performed worse than the *High performers* group. Except for the *WRec-P* in *Week 7* and *SRec-D* in *Week 4* & *Week 7*, when the *Low performers* group on average outperformed slightly the *High performers* group.

The *Word Recognition – Immediate (WRec-I)* and *Percentage ratio* of delayed to immediate correct answers of the *Shape Recognition (SRec-P)* task did not show statistically significant differences between the two groups.

Mean values for each week and each group for the *Sustained Attention to Response Task (SART)* can be seen in Table 7-3.

Table 7-3: SART - Mean values for each task measure and group separately

Task		Week	Week	Week	Week	Week	Week	Week	Week		
Measure	Group	1	2	3	4	5	6	7	8	*t*	*p*
Comm.	Low	1.50	0.88	0.13	0.25	0.38	0.38	0.38	0.50	-2.98	0.021*
Errors	High	2.75	1.63	1.25	0.25	0.50	1.13	1.63	0.25		
Omission	Low	0.63	0.13	0.38	0.50	0.00	0.25	0.25	0.00	-2.30	0.055
Errors	High	1.13	0.63	0.38	0.75	0.25	0.00	0.50	0.13		
SD_RT	Low	114.8	125.3	130.7	99.1	98.0	107.1	86.9	112.9	1.20	0.271
(ms)	High	97.9	105.3	100.4	93.1	105.5	90.7	114.6	105.8		
CoV_RT	Low	0.24	0.24	0.25	0.19	0.21	0.23	0.19	0.22	1.59	0.156
	High	0.22	0.23	0.21	0.19	0.20	0.18	0.22	0.20		

Low = Low performers, High = High performers, SD_RT – standard deviation of reaction time, CoV_RT – coefficient of variation of reaction time

** - task measure with statistically significant difference in mean values between groups, estimated employing paired 2-tailed Student's t-test*

Employing paired two-tailed Student's *t*-test, statistically significant differences (p=0.021) were found in the mean values of *Commission Errors* measure between the two groups. The remaining three *SART* measures did not achieve statistically significant differences between the two participants' groups, but the *Omission Errors* measure approached significance (p=0.055).

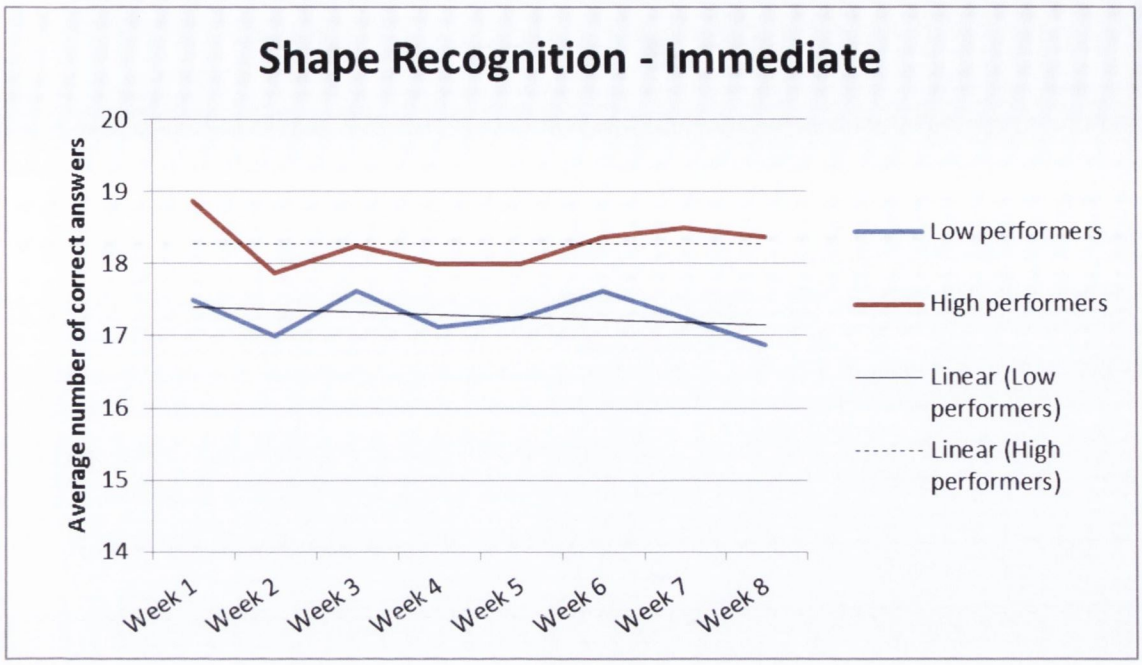


Figure 7-4: *Average number of correct answers for the Shape Recognition – Immediate test for Week 1 - 8 for both participants' groups. Including fitted linear regression lines.*

7.2.1 Validity

The results of assessment of validity of the new tasks scores against the standard neuropsychological tests scores can be seen in Table 7-4 to Table 7-9.

Table 7-4: Validation of Week 1 recognition tasks scores against standard neuropsychological tests scores

Week 1	Word Recognition			Shape Recognition		
	Immediate	Delayed	Percentage	Immediate	Delayed	Percentage
MMSE	0.094	0.191	-0.074	0.182	0.454	0.092
NART Errors	0.123	0.175	0.351	0.073	0.306	0.299
WMS Story 1 Immediate	0.248	0.107	-0.362	0.344	0.284	-0.043
WMS Story 2 Immediate 1	0.180	0.116	-0.271	0.266	0.179	-0.089
WMS Story 2 Immediate 2	0.192	0.367	0.052	0.221	0.411	0.133
WMS Story 1 Delayed	**0.570***	0.434	-0.129	0.403	0.249	-0.132
WMS Story 2 Delayed	0.286	0.426	0.06	0.291	0.355	0.093
WMS Story Recognition	-0.078	0.090	-0.169	0.313	**0.547***	0.252
WMS Pairs Immediate	0.266	0.307	-0.122	0.473	0.327	-0.116
WMS Pairs Delayed	0.471	0.495	0.039	0.424	0.198	0.017
WMS Pairs Recognition	0.402	0.367	-0.431	0.431	0.346	-0.367
WMS Digit Span	-0.099	-0.039	-0.021	-0.141	-0.147	0.016
Stroop test	0.014	-0.138	-0.304	0.225	0.131	0.060
Category Fluency	-0.422	-0.198	-0.124	-0.453	-0.348	-0.043
HADS Anxiety	0.055	-0.249	-0.243	0.044	-0.339	-0.401
HADS Depression	0.193	-0.243	**-0.584***	-0.010	-0.309	**-0.548***
SsNART	-0.171	-0.084	-0.168	-0.076	-0.167	-0.090
SsLM	0.288	0.403	0.166	0.243	0.304	0.027

*All values are Spearman's ρ coefficients. Strong correlation in bold. * - statistically significant correlation $p<0.05$*

Table 7-5: Validation of Week 2 recognition tasks scores against standard neuropsychological tests scores

Week 2	Word Recognition			Shape Recognition		
	Immediate	Delayed	Percentage	Immediate	Delayed	Percentage
MMSE	0.194	**0.694†**	**0.742§**	-0.170	0.134	0.144
NART Errors	0.124	0.138	0.134	0.039	-0.284	-0.294
WMS Story 1 Immediate	0.008	0239	0.318	0.217	0.358	0.211
WMS Story 2 Immediate 1	0.180	0.220	0.231	0.184	0.211	0.211
WMS Story 2 Immediate 2	0.185	**0.533***	**0.587***	0.244	0.229	0.062
WMS Story 1 Delayed	-0.026	0.322	0.391	0.221	0.367	0.207
WMS Story 2 Delayed	0.132	**0.521***	**0.571***	0.288	0.376	0.244
WMS Story Recognition	0.265	**0.542***	**0.567***	0.228	0.404	0.319
WMS Pairs Immediate	0.196	0.288	0.278	0.339	0.358	0.158
WMS Pairs Delayed	-0.035	0.103	0.151	0.308	0.399	0.281
WMS Pairs Recognition	0.409	0.396	0.337	-0.258	0.086	0.309
WMS Digit Span	-0.099	-0.191	-0.132	-0.020	0.278	0.202
Stroop	0.362	0.002	-0.121	-0.036	0.159	0.182
Category Fluency	-0.256	-0.085	-0.001	0.076	0.398	0.365
HADS Anxiety	-0.424	-0.394	-0.401	0.419	0.093	-0.180
HADS Depression	-0.092	-0.345	-0.257	-0.248	-0.427	-0.219
SsNART	-0.048	0.056	0.058	0.081	0.311	0.242
SsLM	-0.090	0.495	**0.597***	0.353	0.298	0.072

*All values are Spearman's ρ coefficients. Strong correlation in bold. * - statistically significant correlation p<0.05, † - statistically significant correlation p<0.01, § - statistically significant correlation p<0.001*

Table 7-6: Validation of average Week 1 – 8 recognition tasks scores against standard neuropsychological tests scores

Week 1-8 Average	Word Recognition			Shape Recognition		
	Immediate	Delayed	Percentage	Immediate	Delayed	Percentage
MMSE	0.227	0.478	**0.610***	0.018	0.273	**0.525***
NART Errors	-0.11	-0.139	-0.265	0.102	-0.213	-0.176
WMS Story 1 Immediate	0.030	0.156	0.320	0.294	0.295	0.107
WMS Story 2 Immediate 1	0.055	0.184	0.249	0.168	-0.070	-0.260
WMS Story 2 Immediate 2	0.255	**0.578***	**0.656†**	0.179	0.106	0.171
WMS Story 1 Delayed	0.301	0.371	0.354	0.365	0.218	0.043
WMS Story 2 Delayed	0.322	**0.591***	**0.640†**	0.299	0.153	0.087
WMS Story Recognition	-0.029	0.227	0.402	0.232	0.283	0.245
WMS Pairs Immediate	0.112	0.201	0.228	0.434	0.297	0.001
WMS Pairs Delayed	0.420	0.301	0.100	**0.530***	0.359	0.038
WMS Pairs Recognition	0.423	0.308	0.140	0.056	0.168	0.028
WMS Digit Span	0.045	0.170	0.311	0.027	0.318	0.326
Stroop	0.043	0.036	0.182	0.155	0.174	-0.068
Category Fluency	0.032	0.222	0.339	-0.316	-0.058	0.192
HADS Anxiety	-0.229	-0.351	-0.308	0.172	-0.183	-0.473
HADS Depression	-0.211	-0.322	-0.243	-0.134	-0.367	**-0.586***
SsNART	0.178	0.332	0.462	-0.080	0.180	0.259
SsLM	0.227	**0.582***	**0.702†**	0.307	0.185	0.171

*All values are Spearman's ρ coefficients. Strong correlation in bold. * - statistically significant correlation p<0.05, † - statistically significant correlation p<0.01*

Table 7-7: Validation of Week 1 SART task scores against standard neuropsychological tests scores

Week 1	Commission Errors	Omission Errors	SD RT	CoV RT
MMSE	-0.008	-0.120	-0.409	-0.272
NART Errors	-0.211	0.171	0.194	0.078
WMS Story 1 Immediate	0.151	0.057	-0.079	0.040
WMS Story 2 Immediate 1	0.043	-0.104	-0.083	-0.040
WMS Story 2 Immediate 2	0.160	0.128	-0.075	-0.004
WMS Story 1 Delayed	-0.066	0.118	-0.108	-0.046
WMS Story 2 Delayed	0.119	0.081	-0.202	-0.121
WMS Story Recognition	0.198	0.037	-0.236	-0.083
WMS Pairs Immediate	0.123	0.097	0.132	0.087
WMS Pairs Delayed	0.149	0.278	0.099	0.097
WMS Pairs Recognition	0.147	-0.332	-0.084	-0.252
WMS Digit Span	0.242	0.089	-0.064	0.001
Stroop	0.359	-0.029	0.256	0.205
Category Fluency	-0.053	-0.147	**-0.540***	-0.368
HADS Anxiety	-0.347	-0.066	0.252	0.068
HADS Depression	-0.160	-0.496*	0.061	-0.076
SsNART	0.239	-0.019	-0.183	0.003
SsLM	0.065	0.104	-0.206	-0.120

*All values are Spearman's ρ coefficients. Strong correlation in bold. * - statistically significant correlation p<0.05*

Table 7-8: Validation of Week 2 SART task scores against standard neuropsychological tests scores

Week 2	Commission Errors	Omission Errors	SD RT	CoV RT
MMSE	0.032	0.394	-0.189	0.253
NART Errors	0.240	-0.219	-0.142	-0.118
WMS Story 1 Immediate	-0.120	0.439	-0.201	0.135
WMS Story 2 Immediate 1	0.080	0.183	-0.444	-0.203
WMS Story 2 Immediate 2	-0.033	0.317	**-0.599***	-0.109
WMS Story 1 Delayed	-0.245	0.396	-0.236	0.149
WMS Story 2 Delayed	0.061	0.376	-0.424	-0.068
WMS Story Recognition	0.255	0.464	-0.363	-0.092
WMS Pairs Immediate	-0.140	0.465	-0.093	0.073
WMS Pairs Delayed	-0.105	0.155	-0.057	-0.094
WMS Pairs Recognition	-0.297	0.123	-0.364	-0.364
WMS Digit Span	-0.432	0.100	0.074	0.148
Stroop	0.081	0.336	0.323	0.130
Category Fluency	-0.363	-0.052	**-0.527***	-0.356
HADS Anxiety	-0.308	-0.063	0.114	-0.089
HADS Depression	-0.302	-0.375	-0.215	-0.053
SsNART	-0.060	0.288	-0.033	0.090
SsLM	0.038	0.351	-0.340	0.088

*All values are Spearman's ρ coefficients. Strong correlation in bold. * - statistically significant correlation $p<0.05$*

Table 7-9: Validation of average Week 1 – 8 SART task scores against standard neuropsychological tests scores

Week 1-8 Average	Commission Errors	Omission Errors	SD RT	CoV RT
MMSE	0.147	0.071	-0.317	-0.134
NART Errors	0.031	0.316	0.221	0.236
WMS Story 1 Immediate	-0.007	-0.094	-0.204	-0.191
WMS Story 2 Immediate 1	-0.004	-0.292	-0.290	-0.377
WMS Story 2 Immediate 2	0.101	-0.011	**-0.537***	-0.279
WMS Story 1 Delayed	-0.167	-0.135	-0.347	-0.359
WMS Story 2 Delayed	0.094	-0.057	-0.398	-0.339
WMS Story Recognition	0.189	0.210	-0.149	-0.041
WMS Pairs Immediate	0.004	0.131	-0.102	0.084
WMS Pairs Delayed	0.003	0.022	-0.020	-0.082
WMS Pairs Recognition	0.056	-0.231	-0.420	-0.420
WMS Digit Span	-0.051	-0.007	-0.258	0.010
Stroop	0.280	-0.051	0.245	0.222
Category Fluency	-0.393	-0.264	**-0.710†**	**-0.583***
HADS Anxiety	-0.482	-0.176	0.135	-0.121
HADS Depression	-0.187	**-0.756§**	-0.181	-0.391
SsNART	0.057	-0.214	-0.298	-0.205
SsLM	0.099	-0.009	-0.368	-0.319

*All values are Spearman's ρ coefficients. Strong correlation in bold. * - statistically significant correlation p<0.05, † - statistically significant correlation p<0.01, § - statistically significant correlation p<0.001*

7.2.2 Analysis of practice effect

Analysis of the slopes of the fitted linear regression lines to the mean values of all recognition tasks measures from each week, showed that the slopes of all measures were not significantly different from horizontal line, except the *WRec-D, WRec-P* and *Omission Errors* measures for the *High performers* group (see Table 7-10). All slopes showed slight decrease in performance over time.

Table 7-10: Analysis of practice effect - linear trends

Task Measure	Group	R^2	F	Slope (b)	t	p
WRec-I	Low	0.331	2.973	0.234	1.724	0.135
	High	0.221	1.704	0.071	1.305	0.239
WRec-D	Low	0.093	0.615	0.086	0.784	0.462
	High	0.583	8.400	-0.208	-2.898	0.027*
WRec-P	Low	0.131	0.905	-0.872	-0.952	0.378
	High	0.813	26.142	-1.447	-5.113	0.002*
SRec-I	Low	0.097	0.641	-0.036	-0.801	0.454
	High	<0.001	<0.001	0.000	0.000	1.000
SRec-D	Low	0.041	0.254	0.048	0.504	0.632
	High	0.086	0.564	0.052	0.751	0.481
SRec-P	Low	0.099	0.659	0.451	0.811	0.448
	High	0.100	0.669	0.271	0.818	0.445
Commission Errors	Low	0.322	2.856	-0.103	-1.690	0.142
	High	0.365	3.454	-0.210	-1.859	0.112
Omission Errors	Low	0.356	3.320	-0.055	-1.822	0.118
	High	0.548	7.269	-0.110	-2.696	0.036*
SD_RT	Low	0.305	2.631	-3.296	-1.622	0.156
	High	0.102	0.679	1.012	0.824	0.441
CoV_RT	Low	0.298	2.544	-0.006	-1.595	0.161
	High	0.235	1.844	-0.003	-1.358	0.223

Low = Low performers, High = High performers

** - statistically significant slope of the linear regression line*

To assess the magnitude of practice effect, Cohen's *d* coefficient was calculated from recognition tasks measures from each week. The results can be seen in Table 7-11.

Table 7-11: Magnitude of practice effect of the Word and Shape Recognition and SART tasks measures.

Task Measure	Group	Week 2 - Week 1	Week 3 - Week 2	Week 4 - Week 3	Week 5 - Week 4	Week 6 - Week 5	Week 7 - Week 6	Week 8 - Week 7	Week 8 - Week 1
WRec-I	Low	0.98†	-0.35	0.35*	-0.17	0.09	-0.36	0.65*	1.08†
	High	-0.24	-0.21	-0.11	0.00	0.41*	0.11	0.24*	0.38*
WRec-D	Low	0.25*	-0.58	0.77†	-0.49	0.13	0.62*	-0.28	0.22*
	High	-0.18	-0.27	-0.15	-0.05	-0.28	0.35*	-0.05	-0.84
WRec-P	Low	-0.92	-0.36	0.55*	-0.57	0.17*	1.11†	-0.89	-1.11
	High	-0.12	-0.21	-0.12	-0.03	-0.66	0.36*	-0.14	-1.09
SRec-I	Low	-0.25	0.41*	-0.32	0.07	0.22*	-0.24	-0.19	-0.27
	High	-0.69	0.26*	-0.18	0.00	0.26*	0.09	-0.10	-0.41
SRec-D	Low	0.14	-0.05	0.13	0.06	-0.06	0.46*	-1.02	-0.23
	High	-0.17	0.35	-0.64	0.34*	0.00	0.18*	0.10	0.17*
SRec-P	Low	0.37*	-0.46	0.43*	0.03	-0.38	0.84†	-0.90	0.00
	High	0.31	0.21*	-0.64	0.45*	-0.23	0.13	0.16*	0.44
Commission Errors	Low	-0.78	-1.17	0.30*	0.25*	0.00	0.00	0.19*	-1.53
	High	-0.41	-0.23	-0.91	0.40*	0.57*	0.25*	-0.78	-1.03
Omission Errors	Low	-0.49	0.43*	0.14	-0.66	0.76†	0.00	-0.50	-0.63
	High	-0.29	-0.25	0.28*	-0.37	-0.50	0.66*	-0.47	-0.66
SD_RT	Low	0.26*	0.09	-0.57	-0.04	0.31*	-0.70	0.79†	-0.05
	High	0.29*	-0.15	-0.22	0.34*	-0.37	0.51*	-0.17	0.22*
CoV_RT	Low	-0.12	0.17*	-0.85	0.44*	0.41*	-0.77	0.77†	-0.48
	High	0.20*	-0.25	-0.29	0.29*	-0.53	0.51*	-0.16	-0.32

Low = Low performers, High = High performers

All values are Cohen's d coefficients. Negative values show deterioration in time.

*Unmarked values show minimal practice effect, * - moderate practice effect, † - large practice effect.*

7.2.3 Test-retest reliability

Test–retest reliability of the recognition task measures was estimated using non-parametric Spearman's ρ rank correlation between each week and between *Week 8* & *Week 1* for each measure. The results can be seen in Table 7-12.

Table 7-12: Test-retest reliability of the Word and Shape Recognition and SART tasks measures.

Task Measure	Correl. / Signif.	Week 2 - Week 1	Week 3 - Week 2	Week 4 - Week 3	Week 5 - Week 4	Week 6 - Week 5	Week 7 - Week 6	Week 8 - Week 7	Week 8 - Week 1
WRec-I	ρ	0.134	0.177	0.633†	0.120	0.311	0.339	0.503	-0.042
	Sig.	0.621	0.512	0.009*	0.657	0.241	0.199	0.047*	0.877
WRec-D	ρ	0.580	0.539	0.470	0.663†	0.307	0.339	0.626†	0.145
	Sig.	0.018*	0.031*	0.066	0.005*	0.247	0.126	0.010*	0.592
WRec-P	ρ	0.338	0.529	0.347	0.545	0.086	0.294	0.532	0.447
	Sig.	0.200	0.035*	0.188	0.029*	0.752	0.268	0.034*	0.083
SRec-I	ρ	0.166	0.118	0.447	0.534	0.443	0.469	0.724§	0.443
	Sig.	0.539	0.664	0.083	0.033	0.086	0.067	0.002*	0.094
SRec-D	ρ	0.230	0.157	0.700§	0.320	0.238	0.810‡	0.667†	0.276
	Sig.	0.392	0.560	0.003*	0.227	0.375	<0.001*	0.005*	0.3
SRec-P	ρ	0.032	0.007	0.224	0.452	0.028	0.321	0.480	0.112
	Sig.	0.906	0.980	0.405	0.079	0.917	0.226	0.060	0.681
Commission Errors	ρ	0.432	0.265	0.045	0.241	0.394	0.620†	0.347	0.261
	Sig.	0.094	0.321	0.869	0.369	0.131	0.010*	0.188	0.329
Omission Errors	ρ	0.576	0.188	0.110	0.553	-0.098	-0.180	-0.123	0.553
	Sig.	0.020*	0.487	0.686	0.026*	0.719	0.504	0.649	0.026*
SD_RT	ρ	0.312	0.756§	0.403	0.650†	0.565	0.374	0.735§	0.426
	Sig.	0.240	0.001*	0.122	0.006*	0.023*	0.154	0.001*	0.099
CoV_RT	ρ	0.382	0.626†	0.294	0.606†	0.497	0.609†	0.762§	0.518
	Sig.	0.144	0.009*	0.269	0.013*	0.050*	0.012*	0.001*	0.040*

ρ = Spearman's ρ coefficient. Sig. = significance of the correlation (2-tailed).

** - statistically significant correlation.*

Unmarked ρ values show low test-retest reliability, † - marginal reliability, § - adequate reliability, ‡ - high reliability.

7.3 Discussion

The study presented in this chapter presented newly developed battery of online web-based tasks for monitoring of memory and attention, two critical elements of cognitive function. This battery of tasks is delivered to the participant over a website and allows for automated scoring of participant's responses and generation of tasks scores. This results in substantial time saving for the clinician or researcher and allows more time to interpret the data for the specific participant.

Statistically significant differences were found in mean values of four recognition tasks measures (*WRec-D, WRec-P, SRec-I, SRec-D*) between the *Low performers* and *High performers* groups (see section 7.2). The *Low performers* group achieved lower mean task scores than the

High performers group on each week of testing, except for the *WRec-P* measure in *Week 7* and the *SRec-D* measure in *Week 4 & Week 7* where the *Low performers* achieved slightly better mean score. This supports the hypothesis of this study that the *High performers* group should outperform the *Low performers* group.

Both participants' groups achieved relatively high score on the *WRec-I* task, which suggests that this task may not be enough cognitively challenging for the older adults. While the *SRec-I* and *SRec-D* mean scores for the *Low performers* is lower than for the *High performers* in accordance with the hypothesis, the *percentage ratio of delayed to immediate correct answers* (*SRec-P*) is similar for both groups. Due to these reasons, the *WRec-I* and *SRec-P* measures did not achieve statistically significant differences between the *Low performers* and *High performers* groups.

The *High performers* group achieved a lower mean number (mean difference of 1.14) of correct answers across all weeks on the *Shape Recognition – Delayed* task in comparison to the *Word Recognition- Delayed* task. This may support the results of previous research from Boyle et al. [353], in which the authors concluded that visual memory declines earlier in life than verbal memory.

Analysis of the *Sustained Attention to Response Task (SART)* results yielded one measure – *Commission Errors* – with statistically significant differences in mean values between participants' groups. On all weeks, the *Low performers* group generated less *Commission Errors* than the *High performers* group, except for *Week 8*. The other three *SART* measures (*Omission Errors, SD RT, CoV RT*) did not achieve significant difference in mean values between the participants' groups. Similarly to the *Commission Errors* measure, the *Low performers* group generated less *Omission Errors* than the *High performers* group across all weeks, except the *Week 6*. The performance of the two groups on the remaining two *SART* measures varied over the period of eight weeks.

7.3.1 Validity

Validating the *Word Recognition, Shape Recognition* and *SART* scores from *Week 1* against the standard neuropsychological tests scores yielded strong correlation between

- *WRec-I* and *WMS Story 1 Delayed* (ρ = 0.570, p = 0.021),
- *WRec-P* and *HADS Depression* (ρ = -0.584, p = 0.017),
- *SRec-D* and *WMS Story Recognition* (ρ = -0.547, p = 0.028),
- *SRec-P* and *HADS Depression* (ρ = -0.548, p = 0.028),
- *SART SD RT* and *Category Fluency* (ρ = -0.540, p = 0.031).

Validating the *Word Recognition, Shape Recognition* and *SART* scores from *Week 2* against the standard neuropsychological tests scores yielded strong correlation between

- *WRec-D* and *MMSE* (ρ = 0.694, p = 0.003),
- *WRec-D* and *WMS Story 2 Immediate 2* (ρ = 0.533, p = 0.034),
- *WRec-D* and *WMS Story 2 Delayed* (ρ = 0.521, p = 0.039),
- *WRec-D* and *WMS Story Recognition* (ρ = 0.542, p = 0.030),
- *WRec-P* and *MMSE* (ρ = 0.742, p = 0.001),
- *WRec-P* and *WMS Story 2 Immediate 2* (ρ = 0.587, p = 0.017),
- *WRec-P* and *WMS Story 2 Delayed* (ρ = 0.571, p = 0.021),
- *WRec-P* and *WMS Story Recognition* (ρ = 0.567, p = 0.022),
- *WRec-P* and *SsLM* (ρ = 0.597, p = 0.015),
- *SART SD RT* and *WMS Story 2 Immediate 2* (ρ = -0.599, p = 0.014),
- *SART SD RT* and *Category Fluency* (ρ = -0.527, p = 0.036).

Validating the *Word Recognition, Shape Recognition* and *SART* scores using average scores from *Week 1-8* against the standard neuropsychological tests scores yielded strong correlation between

- *WRec-D* and *WMS Story 2 Immediate 2* (ρ = 0.578, p = 0.019),
- *WRec-D* and *WMS Story 2 Delayed* (ρ = 0.591, p = 0.016),
- *WRec-D* and *SsLM* (ρ = 0.582, p = 0.018),
- *WRec-P* and *MMSE* (ρ = 0.610, p = 0.012),
- *WRec-P* and *WMS Story 2 Immediate 2* (ρ = 0.656, p = 0.006),
- *WRec-P* and *WMS Story 2 Delayed* (ρ = 0.640, p = 0.008),
- *WRec-P* and *SsLM* (ρ = 0.702, p = 0.002),

- *SRec-I* and *WMS Pairs Delayed* (ρ = 0.530, p = 0.035),
- *SRec-P* and *MMSE* (ρ = 0.525, p = 0.037),
- *SRec-P* and *HADS Depression* (ρ = -0.586, p = 0.017),
- *SART Omission Errors* and *HADS Depression* (ρ = -0.756, p = 0.001),
- *SART SD RT* and *WMS Story 2 Immediate 2* (ρ = -0.537, p = 0.032),
- *SART SD RT* and *Category Fluency* (ρ = -0.710, p = 0.002),
- *SART CoV RT* and *Category Fluency* (ρ = -0.583, p = 0.018).

7.3.2 Analysis of practice effect

Fitting linear regression lines to the mean values from all weeks of the recognition tasks and *SART* measures, and subsequent analysis of significance of the regression lines' slopes showed that three of the linear regression lines for the *High performers* group had slopes significantly different from horizontal line (see Section 7.2.1, Table 7-10). The *WRec-D, WRec-P* and *Omission Errors* slopes showed slight decrease in performance of the *High performers* group over the period of eight weeks. The remaining 17 regression lines' slopes were not statistically significant.

To further analyse the effect of practice on the results, Cohen's *d* coefficient was calculated for all measures between all adjacent weeks and between *Week 8* and *Week 1* for both participants' groups separately (see section 7.2.1, Table 7-11) to assess the magnitude of practice effect. Out of the 96 Cohen's *d* coefficients (6 measures x 2 groups x 8 between weeks estimations) calculated for recognition tasks measures,

- 63 (i.e. 65.63%, *Low* – 31 (64.58%), *High* – 32 (66.67%)) coefficients showed minimal practice effect,
- 28 (29.17%, *Low* – 12 (25.00%), *High* – 16 (33.33%)) coefficients moderate practice effect, and
- five (5.21%, *Low* – 5 (10.42%), *High* – 0 (0.00%)) coefficients large practice effect.

Out of the 64 Cohen's *d* coefficients (4 measures x 2 groups x 8 between weeks estimations) calculated for *SART* measures,

- 40 (62.50%, *Low* – 20 (62.50%), *High* – 20 (62.50%)) coefficients showed minimal practice effect,
- 21 (32.81%, *Low* – 9 (28.13%), *High* – 12 (37.50%)) coefficients moderate practice effect, and
- three (4.69%, *Low* – 3 (9.38%), *High* – 0 (0.00%)) coefficients large practice effect.

The minimal and moderate practice effect observed across nearly all measures over such short period of time (eight weeks) suggests that the performance on these tasks is not prone to improvement over time. Other studies reported in the literature have typically used longer intervals between two assessments, ranging from months to years [39, 336]. Employing longer intervals between assessments would further lower the magnitude of practice effect achieved in this study and potentially allow administering the battery at regular intervals over time.

As a part of on-going extension of this study, the participants will undertake the testing once every three months for the period of two years. This extension, known as Phase 2, will then give better insight on the practice effect size between two assessments over longer intervals of time.

Small sample size (eight individuals in each group) also may not be big enough to achieve adequate results of analysis of magnitude of practice effect. Increasing the sample size of both groups may provide more accurate estimates of practice effect.

As mentioned in the introduction of this chapter, the greatest practice effect in repeated memory assessment typically impacts the second assessment in a series of assessments [337, 340]. The only practice effect of $d = 0.98$ between *Week 1* and *Week 2* was found for the *Word Recognition - Immediate* (*WRec-I*) task for the *Low performers* group. All other tasks for both groups showed minimal or moderate practice effect between *Week 1* and *Week 2*. Analysis of differences of the means of scores between *Week 1* and *Week 2* for each participants' group approached significance (p=0.054, Student's t-test) only for the *WRec-I* task for the *Low performers* group, in accordance with the Cohen's d measure.

The *WRec-I* task is the first task the participants encounter, thus the participants were not yet familiar with the testing environment, which may have an impact on the final score of this task. At the time when the participants perform the *SRec-I* task, the participants are already familiar with the task. Therefore the difference in means of *SRec-I* scores between *Week 1* and *Week 2* may be smaller than the difference in means of *WRec-I* scores between *Week 1* and *Week 2*. This then reduce the effect of practice for *SRec-I* compared to *WRec-I*.

7.3.3 Test-retest reliability

To estimate test-retest reliability, Spearman's ρ rank correlation coefficient was calculated from the recognition tasks and *SART* measurements for each two adjacent weeks and between *Week 8* and *Week 1* (see Section 7.2.3, Table 7-12). Out of the 80 correlation coefficients,

- 65 (81.25%) coefficients showed low test-retest reliability,
- ten (12.50%) coefficients marginal reliability,
- four (5.00%) coefficients adequate reliability, and
- one (1.25%) coefficient high reliability.

Low test-retest reliability scores may be due to factors, including participants' mood, fatigue, performance anxiety and others. Due to these tasks being undertaken at participants' homes, it is not possible to create the controlled environment that is created in clinics and therefore it is not possible to achieve standardized administration conditions. This may have further impact on the test-retest reliability scores. Similar to estimation of practice effect, the small sample size may not be adequate to achieve adequate test-retest reliability results. Therefore further studies with bigger cohorts would be desirable.

Additionally, tests used in populations with high response variability, such as older adults, may invariably yield low reliability coefficients despite the best efforts of test developers [68]. As this study employed cohort of older adults, this fact may have further negative impact on the test-retest reliability results.

Research has shown that some cognitive domains are more reliable than others. For example, measures of verbal comprehension and attention have been found to be more reliable than memory tests in cognitively normal older adults [354]. Dikmenet al. [355] suggested that the lower reliabilities in memory tests reflect the variable nature of human memory, as poor reliabilities for verbal memory have been observed for most tests [356, 357].

7.3.4 Future directions

Analysis of the mean measurement scores of the recognition tasks results revealed that the participants scored a high number of correct answers on both of the *Word* and *Shape Recognition - Immediate* and *Delayed*. Future revision of the *Word Recognition* task may revise the word stimuli employed. To make the tasks more cognitively challenging, it is proposed to normalize the length and to lower the range of normative frequencies of the word stimuli.

For the *Shape Recognition* task, some of the stimuli will be upgraded to embody more abstract form. Despite the fact that the shapes were created in a manner that would not evoke association with objects or words, the feedback from participants suggests that some of the shapes were associated with words or objects. These shapes require upgrading to ensure that the *Shape Recognition* task will only target non-verbal memory.

Making the recognition tasks more cognitively challenging may further increase the performance gap between *Low performers* and *High performers* groups, allowing for better discrimination between the two groups.

Key points

- Developed novel battery of tasks for assessment of memory and attention remotely over a web-site allowing to undertake the testing at participant's home.
- Developed novel tasks for word recognition and shape recognition memory assessment, which employ the same stimuli for every assessment avoiding the need for use of parallel forms.
- Automated scoring of tasks' results providing substantial time saving for the researcher analysing the collected data.
- The remote administration method and automated scoring of tasks allow for massive scale deployments.
- Four out of six recognition tasks measures showed statistically significant differences between *Low performers* and *High performers* groups.
- *Low performers* group achieved worse performance on the recognition tasks than the *High performers* group in accordance with the hypothesis of the study.
- *SART Commission Errors* measure showed statistically significant differences between *Low* and *High performers* groups.
- 65.63% Cohen's *d* coefficients of recognition tasks measures and 62.50% of SART measures showed minimal practice effect over time.
- 29.17% Cohen's *d* coefficients of recognition tasks measures and 32.81% of SART measures showed moderate practice effect over time.
- Only 5.21% Cohen's *d* coefficients of recognition tasks measures and 4.69% of SART measures showed large practice effect over time.
- Analysis of validity yielded strong correlations between standard neuropsychological tests measures and two *Word Recognition*, two *Shape Recognition*, three *SART* measures.

7.4 Summary

This chapter answered the Research Questions 20 - 22 posed in Chapter 2, Section 2.18.4.

This chapter presented a study during which a novel battery of tasks for assessment of memory and attention was developed. This study demonstrated that the battery can be administered remotely over a dedicated website, allowing the testing to take place at participant's home. The remote administration method and automated scoring of tasks' results allow for deployment to very large scale cohorts at very low cost.

All tasks employ the same stimuli for every assessment and therefore overcome the complications with equivalency of specific items used in parallel forms tests. The small practice effect achieved, which may be further lowered by employing longer intervals between assessments, may potentially allow administration of the tasks at regular intervals over time. To confirm this assumption, testing over period longer than eight weeks while employing larger cohort of older adults is required.

The next chapter reviews comprehensively the findings of all the studies of this thesis. Future applications and directions of the research are also presented.

Chapter 8

Discussion

The overall objective of this thesis was to determine if changes in an individual's cognitive function may be objectively assessed by quantitative changes of their speech characteristics. From the elaborate list of research questions posed in Chapter 2 a series of methods were developed and a number of studies undertaken to address this objective.

8.1 Thesis aims

As speech is the focus of this research, the first target was to reliably extract temporal and acoustic features of speech from speech recordings acquired in real-life environments, while removing artefacts such as breath sounds and compensating for varying levels and sources of noise. A new algorithm for temporal features extraction from speech recordings was developed and presented in the Chapter 3. To achieve correct speech feature values that correspond with reality, the algorithm must correctly identify pauses and speech segments in the recording. The performance of the developed algorithm was assessed in Chapter 4, Section 4.2. Performance assessment of the algorithm on 20 real recordings was observed to achieve very good performance in speech and pause identification with overall accuracy of 97.29%, sensitivity of 93.52% and specificity of 98.37%. Therefore, the feature values from automated extraction of temporal features should closely correspond with real values.

A spoken utterance contains not only words that are produced to make syntactic and semantic sense, but also many insertions of non-speech artefacts, such as coughs, splutters and breath sounds. Therefore, the second specific research question objective to assess cognitive function in real-life (i.e. non-studio, without controlled conditions) environments was to assess if the removal of breath sounds from speech recordings prior to the speech

features extraction can improve the classification performance between healthy and impaired individuals. The impact of breath sounds detection and removal prior to the speech feature extraction improved the discrimination ability between healthy and impaired subjects (see Section 4.3). Employing the breath sounds detection and removal, an increase in all performance metrics of the classifier was achieved with the overall accuracy increasing by 6.7%, from 67.5% to 74.2%. The area under the ROC curve increased by 0.07, from 0.68 to 0.75. The sensitivity increased by 14.1%, from 55.7% to 69.8%. The removal of the breath sounds has been shown to be an important factor in the analysis of speech features in studies of impaired cognitive function. The results suggest that the duration of pauses is an indication of the cognitive load being experienced by the subject. While breath sounds may be considered a normal part of speech, they may also be elongated by subjects attempting to fill pauses while trying to manage the cognitive load imposed on them by the current task. It is for this reason that classing breaths as pauses leads to a more discriminative feature for analysis.

8.1.1 Speech characteristics in schizophrenia and their correlation to cognitive function

The population of older adults is not the only one affected with impairment in cognitive function. Psychiatric conditions, such as schizophrenia, are associated with cognitive disorders. The research question have asked if the status of individual's cognitive function correlates with speech features of individuals with/without schizophrenia extracted from recordings while reading loud a short text passage. This study found that the speech of patients with schizophrenia in comparison with healthy controls contains an increased number of pauses and also pauses of increased duration, which may reflect the impairment of cognitive function that is associated with schizophrenia. *The Number of Pauses*, *Proportion of Recording in Silence* and the *Total Length of Pauses* features were found to increase by 27%, 23% and 40% respectively for a patient group compared to a control group. Statistically significant differences were found in mean values between schizophrenics and healthy controls for all speech features, except the *Total Length of Utterances* and the *Coefficient of Variation of Fundamental Frequency*. The ability to discriminate between schizophrenics and healthy controls by a classifier trained with these speech features achieved classification accuracy of 75.6% (Sensitivity of 72.6%, Specificity of 78.6%). The pause related features were most significant in differentiating between schizophrenic and control group. Each of the clinical variables had a correlation of statistical significance ($p<0.05$) with at least one of the temporal speech features.

The speech characteristics may represent a complementary measure for the current methods of cognitive function assessment in schizophrenia. Speech may serve as an objective biomarker for schizophrenia, while allowing for remote monitoring, providing psychiatrists with up to date knowledge of cognitive function and insight into the suitability of the prescribed medication.

8.1.2 Changes in speech characteristics during ageing and their correlation to cognitive function

Six research questions investigated speech characteristics in cognitively healthy and impaired older adults in Chapter 5. The initial analysis investigated use of the speech features as a means of assessing cognitive function. This was carried out by calculating the correlation of speech features with cognitive measures. Speech features extracted from recordings of speech of reading loud a *Text passage* were found to correlate with measures of the gold-standard neuropsychological tests, in particular *Mini-Mental State Examination (MMSE),* and also *Word Recall – Immediate* and *Delayed*, and the *Category Fluency* task. Speech features extracted form recordings of the *Picture Taboo* task were found to correlate with *MMSE* and *Digit Span* task measures. These results demonstrate that the temporal and acoustic features extracted from speech can be employed to classify older adults in terms of their cognitive status. The higher number of *Phoneme Rate features* correlating with the MMSE group, than was the number of *Temporal features,* suggests that it is important to compensate for reading rate in studies of speech characteristics in cognitive impairment.

The classification analysis investigated the various combinations of speech features to achieve the best performance. Training the classifier with all speech features of the *Text passage* task yielded slightly higher accuracy (68.9%) in discriminating cognitively impaired and healthy subjects than was the accuracy (68.0%) of the classifier trained with all speech features of the *Picture Taboo* task. Training the classifier with temporal features of speech of both tasks yielded higher accuracy (72.2%) in discriminating cognitively impaired and healthy subjects than was the accuracy (63.2%) of the classifier trained with acoustic features of speech. Training the classifier with a combination of the speech features of the *Text passage* and the *Picture Taboo* tasks yielded 80.4% classification accuracy. The significant improvement in classification accuracy achieved when features from the two speech tasks are combined suggests that both tasks are important in the assessment of overall cognitive function, targeting different domains of cognitive function, specifically sustained attention and short-

term memory. This demonstrated the complexity of the cognitive processes employed during speech production.

The second study of Chapter 5 addressed the limitations of the Mini-Mental State Examination (MMSE) as a standalone test for cognitive function. Several cognitive tests with objective scoring protocols were employed in designing a Combined Cognitive Score (CCS). The impact of splitting the participant cohort into healthy and impaired groups based on their CCS score, instead of their MMSE score, was assessed. A classifier was trained employing speech features extracted from the *Text passage* and *Picture Taboo* task. Classification accuracy of the LDA classifier for subjects categorized by the CCS increased in comparison with MMSE categorization for both tasks. The CCS outperformed the classification based on MMSE categorisation alone. This is due to the fact that the elements making up the CCS are more targeted than the broad MMSE score.

The third study of Chapter 5 demonstrated the potential use of temporal features of speech for monitoring specific cognitive domain, i.e. memory, over time. Temporal speech features demonstrated potential to provide indicators of a subject's memory state. Temporal speech features of recording of picture description tasks were found to correlate with the MMSE memory subscale. Employing speech features extracted from the *Picture Description* task, the classifier discriminated between the Low and High memory performers with an accuracy of 68%, sensitivity of 63% and specificity of 73%. Employing the features of the *Picture Taboo* task, the classifier discriminated between the two participant groups with an accuracy of 73%, sensitivity of 79%, and specificity of 67%. Participants generated shorter utterances and longer pauses during the *Picture Taboo* task compared to the *Picture Description* task. This may reflect the added cognitive load of the *Picture Taboo* task associated with not being able to use the two taboo words, which represent the key elements of the picture. The longer pauses can therefore be interpreted as time required for alternative word finding while maintaining control of the task.

To answer the last research question of Chapter 2, Section 2.18.2, a literature review on the definition of the minimum pause duration in studies investigating temporal features of speech was performed. Given the ambiguous definition of minimum pause duration and the dynamic nature of speech production [358], the last study of Chapter 5 readdresses the approach presented in the previous sections of this chapter (Sections 5.1 - 5.3), replacing the use of a static 250ms minimum pause duration with a dynamic threshold. Three sets of speech features were extracted from the recordings - static temporal features, dynamic

temporal features and distribution features. The *Pause Mixing Proportion* and *Utterance Mean* distribution parameters showed statistically significant differences between cognitively impaired and healthy group. The distribution parameters outperformed the temporal features in classifying the participants according to their level of cognitive function. The sensitivity of the classifier increased by 0.22% (to 64.20%), specificity by 6.33% (73.12%) and the overall accuracy by 3.27% (68.66%). Employing the dynamic temporal features in classification yielded a decrease in classifier performance in comparison to employing static temporal features. The sensitivity of the classifier decreased by 5.73% (to 58.25%), specificity by 1.10% (65.69%) and the accuracy by 3.42% (61.97%). Despite the dynamic threshold estimation yielding no improvement on traditional methods, this study did highlight the potential in pause and utterance duration distribution parameters in improving classification of older adults according to their cognitive function

The findings of the studies into assessing cross-sectionally cognitive function highlight the advantage of using speech as a measure of cognitive function. The results suggest the potential for employing the temporal features of speech as a screening method for remote long-term monitoring of individual's cognitive function, allowing for early detection and prevention of cognitive decline. Speech features may provide objective measures of individual's cognitive function as opposed to the subjective gold-standard neuropsychological tests. The speech features developed and validated here may serve as a complementary method to the current methods of cognitive function assessment. An individual's baseline speech characteristics may be obtained and changes in these speech characteristics of that individual may serve as an indicator of onset of cognitive decline, effectively providing means for personalised medicine.

8.1.3 Remote assessment of cognitive function

While cross-sectional studies of cognitive function are important, longitudinal assessment of cognitive function represents a much more important method of detecting changes in an individual subject's cognitive function. Studies presented in Chapter 6 and Chapter 7 focused on development of tools that can employ speech characteristics for longitudinal assessment and monitoring of cognitive function. Current cognitive function assessments are associated with time requirements, monetary and administrative burden and it can be difficult for some to access the facilities of health care providers. To address these issues, an investigation of an automated remote telephone assessment of cognitive function was performed in Chapter 6.

The first step was to assess whether speech features can be reliably extracted from recordings of speech acquired over a telephone where the acoustic quality of the signal is reduced. The study reported found that the temporal speech features can be reliably (overall accuracy of 93.2%, sensitivity of 97.3%, specificity of 89.5%) extracted from speech recordings with telephone quality speech. Comparing the speech features acquired in clinic and over a telephone, non-significant differences were demonstrated between the high-quality clinic recordings and the lower quality telephone recordings. The non-significant differences may be attributed to the natural intra-speaker variability over time. The results demonstrate that the telephone speech quality recordings can provide adequate speech features for cognitive function assessment as those extracted from recordings acquired in clinic.

To build on this development with regard to telephone quality speech, an automated Interactive Voice Response (IVR) application was developed. The IVR application developed allows deployment of clinically valid cognitive assessment tasks to large cohorts of older adults and may serve as a tool for repeated assessment of cognitive function. Long-term remote periodic assessment of cognitive function in older adults may enable the early identification of cognitive decline and facilitate early interventions that may reduce the impact of cognitive decline on an individual.

The use of the IVR technology for remote automated delivery of cognitive function assessment interviews has been demonstrated to be functional and well received by a population of 60 older adults. The majority of participants reported a positive experience with the automated IVR system and the majority (93.6%) of them would continue to take part in such cognitive assessment again. A protocol was designed that implemented effective recovery strategies for dealing with errors in individual task modules of the IVR application. This resulted in all participants successfully completing the IVR cognitive assessment interview.

The use of the IVR technology for remote automated delivery of cognitive function assessment interviews was found to be reliable. Including a face-to-face assessment before IVR assessment did not appear to increase the chances of completing the cognitive tasks. However, commencing an IVR assessment with a practice task, the score of which would not be measured, may provide some experience and confidence in using an IVR application for the participant.

The positive experience of the older adults with the developed IVR application suggest that such systems may allow for non-invasive, remote, low-cost, population scale screening, while employing the existing infrastructure and technologies. This will enable pervasive, prevention-driven, personalized care for older adults.

Following the development of the automated IVR application for assessment of cognitive function in Chapter 6, the study presented in Chapter 7 addressed the research question whether an assessment of memory and attention can be delivered automatically over a website. A novel battery of tasks for memory and attention was developed. The battery can be delivered remotely from a dedicated website allowing to undertake the testing at participant's home. This may enable large scale deployments at low cost and effectively resolve issues associated with current cognitive function assessment.

For this web-based assessment, novel tasks (*Word Recognition* and *Shape Recognition* tasks) for word recognition and shape recognition memory assessment were developed. Four out of six recognition tests measures showed statistically significant differences between Low and High memory performers groups. For a test of sustained attention (SART), *Commission Errors* measure showed statistically significant differences between Low and High memory performers groups. 65.63% Cohen's *d* coefficients of recognition tests measures and 62.50% of SART measures showed minimal practice effect over time. The advantage here is that employing the same stimuli for every assessment avoids the need for use of parallel forms. 29.17% Cohen's *d* coefficients of recognition tests measures and 32.81% of SART measures showed moderate practice effect over time. Only 5.21% Cohen's *d* coefficients of recognition tests measures and 4.69% of SART measures showed large practice effect over time. Analysis of validity yielded strong correlations between standard neuropsychological tests measures and two *Word Recognition*, two *Shape Recognition* and three *SART* measures. The results suggest that these tests, while employing the same stimuli over time, do not suffer from practice effect and may potentially be administered at regular intervals over time. To confirm this assumption, testing over a period of multiple months while employing a larger cohort of older adults is required.

8.2 Future applications and directions

A number of extensions to this research can be considered.

8.2.1 Long-term monitoring of cognitive function employing speech

A multi-year long monitoring of cognitive function employing speech would allow for analysis of intra-speaker changes of speech characteristics, instead of analysis of inter-speaker changes in speech production. Employing intra-speaker analysis of speech changes over time for at risk older adults (MMSE < 27) would allow correlation of the changes in speech measures with the changes in cognitive functioning.

The developed Interactive Voice Response (IVR) system may be used for delivery of the speech task, such as the *Text passage* and *Picture Taboo*. These two speech tasks may also be included in an existing long-term study of older adults. The Irish Longitudinal Study on Ageing (TILDA) is a study of a cohort of over 8500 adults aged over 50 collecting physical, mental health and cognitive measures. The *Text passage* and *Picture Taboo* tasks may be employed to further study the influence of cognitive impairment on speech production over a period of several years. If included in the existing battery of neuropsychological tests, these two speech tasks would extend the existing battery of tasks by only a few minutes.

8.2.2 Intervention and treatment monitoring

The developed IVR system and the subsequent speech analysis may be employed as a biomarker in monitoring of suitability and effectiveness for cognitive impairment interventions and associated treatments through monitoring of speech characteristics. The analysis of remotely collected speech samples may complement current face to face neuropsychological assessments and provide more frequent assessment of the effects of intervention.

Dementia has multiple causes. Some are reversible and may be treated [359]. These reversible causes include, among others, the following:

- B1, B6, B12 vitamin deficiency,
- Endocrine disorders, e.g. hypothyroidism,
- Infections, e.g. AIDS, Syphilis
- Electrolytes imbalance
- Anaemia

If the progression from the causes of dementia is detected in its early stage, appropriate treatment may be selected and the progress of treatment may then be monitored objectively through changes in speech characteristics.

In the area of psychiatric cognitive disorders, such as schizophrenia and depression, the methods developed in this thesis may be employed in monitoring recovery from the psychiatric condition. The monitoring of the changes in speech characteristics would provide information on suitability and effectiveness of the selected treatment medication or behavioural change. Changes in fundamental frequency speech features have been employed in previous research to reflect the individual's time course of recovery from depression [360]. Repetitive speech analysis has been also employed to determine the onset of action of antidepressants in depressive patients [361].

The developed IVR application may be modified and employed for assessment of other neurological disorders, such as Parkinson's disease. Parkinson's disease does not only affect limb movement but also voice generation. Dysphonia measures have been employed in the literature to discriminate between healthy subjects and subjects with Parkinson's disease [362, 363]. Modifications of the current IVR system to deliver sustained phonation tasks for Parkinson's disease assessment would be straightforward and have been reported previously [300, 301]. In combination with signal processing and classification algorithms may provide a fully-automated low cost assessment of Parkinson's disease progress.

8.2.3 Existing speech data collections

The availability of digital media over the past two decades has resulted in considerable archived material and datasets. These have an important impact for research on cognitive function.

Similarly to the linguistic analysis of Ronald Regan's speech prior to his diagnosis with Alzheimer's disease and of Iris Murdoch's published work [252], the methods developed in the studies presented in this thesis may also be applied to archived speech recordings of newsreaders, politicians, and others in the broadcast industry. Recordings of newsreaders may provide similar speech data than the *Text passage* task employed in studies of this thesis if a transcript of the news story can be obtained. Recordings of politicians and others in the broadcast industry may provide speech content for analysis of emotions through speech characteristics and mood through words employed [364].

8.2.4 Linguistic analysis

Following on the analysis of Ronald Regan's speeches and of Iris Murdoch's published work, linguistic properties of speech demonstrate potential to be employed in studies of impaired cognitive function. Linguistic analysis of autobiographies of nuns demonstrated that early-life (mean age of 22 years) idea density, obtained through essay writing, was significantly related with late-life cognitive function [365]. Low idea density has been associated with greater impairment in cognitive function. Low idea density also has been significantly associated with lower brain weight, higher degree of cerebral atrophy, more severe neurofibrillary pathology, and the likelihood of meeting neuropathologic criteria for Alzheimer's disease. Roughly 80% of nuns whose writing was measured as lacking in linguistic density went on to develop Alzheimer's disease in old age; meanwhile, of those whose writing was not lacking, only 10% later developed the disease [365].

Overall, findings of the Nun Study suggest "that traits in early, mid, and late life have strong relationships with the risk of Alzheimer's disease, as well as the mental and cognitive disabilities of old age" [366].The Nun Study allowed investigation of a relatively homogeneous group (no drug use, little or no alcohol, similar housing and reproductive histories, etc.), minimizing the extraneous variables that may confound other similar research. A similar speech based study over the life course would be of enormous scientific benefit.

8.2.5 Influence of language on cognitive speech based measures

The developed methods presented in this thesis focused on the English language. These methods need to be assessed with other languages, to assess if the methods can be applied to other languages. Languages differ in speaking rate, rhythm, voicing onset times [367]. These factors need to be considered in the context of different languages.

Culture may have an effect on the speech tasks presented in this thesis. The image for the *Picture Taboo* and *Picture Description* tasks may need to be specifically selected to depict situations or environments related to the specific cultural group. Similarly, a suitable text for the *Text passage* task may need to be selected that would be similarly semantically simple in its content and emotionally neutral.

8.2.6 Other speech tasks

Additional speech tasks can be considered. One proposed tasks is for subjects to write a short paragraph on a specific theme using their own words and read it aloud afterwards. In this manner, the subjects employ vocabulary that they are familiar with. This approach would

also compensate for an effect of education as the text would not contain vocabulary the subject potentially may not use or understand. There are some factors to consider with this approach in terms of a comparison across subjects but the proposed self-generated text passage would provide a useful baseline for longitudinal monitoring of the subject.

8.2.7 Other classification methods

While the current aim of this research was to focus on features that were physiologically and cognitively relevant to the tasks, more sophisticated classifiers may be employed and their impact on the classification performance assessed. Such methods may include Quadratic discriminant analysis (QDA), Support vector machines (SVM), Neural networks (NN), and others.

8.2.8 Brain imaging methods

To further contribute to the hypothesis of using speech as a quantitative method in studies of cognitive impairment, the speech measures developed ideally need to be correlated with outcome measures of brain imaging methods such as electroencephalography (EEG) and functional Magnetic Resonance Imaging (fMRI). The high temporal resolution of EEG can provide more information about the time course of spoken word production. Several studies performed EEG acquisition during overt speech production [368]. In a picture naming task, Costa et al. [369] demonstrated that lexical access occurred around 200 ms after the onset of the picture. The same method may be employed to compare brain functioning of cognitively impaired and healthy individuals. Potential differences in the brain activation patterns between the cognitively impaired and healthy individuals correlating with the changes in speech characteristics would further confirm the hypothesis of employing speech in studies of cognitive function.

fMRI has made it possible to study the neurophysiological basis of language. fMRI was employed in areas such as speech recognition, parsing (lexical) semantic memory, sentence and discourse comprehension and production. Many studies employed fMRI to study speech production [370]. fMRI may be employed during speech tasks for comparison of cognitively impaired and healthy individuals.

While EEG provides high temporal resolution and fMRI provides high spatial resolution, both methods are susceptible to motion artefacts [371]. The speech tasks will need to be designed to overcome these limitations or methods for artefact removal, like Independent Component Analysis [372], will need to be employed.

The potential functional differences in speech production areas of the brain between cognitively impaired and healthy participants would contribute to the hypothesis of employing speech in studies of cognitive function.

8.2.9 Transcranial Magnetic Stimulation

Transcranial Magnetic Stimulation (TMS) is a powerful non-invasive method for brain stimulation, which allows us to interfere with speech and language processes in the brains of healthy participants. It provides new information on the neural and cognitive basis of language organization, studies of which were previously limited to patients with neurological impairment or to correlative functional imaging methods [373]. TMS provides a virtual lesion, the effects of which are brief and reversible. In addition to interfering with brain function, it can provide new correlates of brain function such as measures of cortical excitability. These measures may be more sensitive to brain function than measures of blood flow and electrical activity [373].

Applying high-frequency repetitive TMS (rTMS) [374] over brain language regions of healthy individuals, while disrupting the functions of these areas, may simulate cognitive impairment and its effects on speech production. This may give a better insight on how the cognitive impairment in older adults may affect speech production.

Low-frequency rTMS applied over Wernicke's area has shown to increase the response times on a word-picture matching tasks [375]. This method could be applied to simulate cognitive impairment in healthy adults and its effects on information processing during reading or picture description task.

Single-Pulse TMS can be used to measure motor excitability during language tasks identifying functional links between language and motor system [376]. The differences in excitability of language brain areas between cognitively healthy and impaired may give an insight of effects of cognitive impairment on speech production.

Opposite to simulating cognitive impairment, TMS may be used to improve speech production performance of cognitively impaired individuals. TMS has shown to increase reaction times in a simple semantic task of visually presented words [377]. In a similar manner, it may be employed to attempt to improve the speech production. In a study of Naeser et al. [378], significant improvements on naming accuracy and reaction times were observed in aphasic individuals following TMS treatment. This improvement persisted for another two months for all individuals and eight months for one individual.

8.2.10 Brain modelling

Modelling of brain functions have recently received considerable attention when the European Union announced to provide 1 Billion euro in funding to a human brain modelling project, called the Human Brain Project [379]. Chapter 2, Section 2.12 of this thesis mentioned morphological changes in brains of cognitively impaired individuals, which includes the neuronal loss in brain areas of speech production. Brain modelling may be employed to simulate effects of these morphological changes on speech production.

8.2.11 Other aspects of cognitive function

Other aspects of cognitive function besides those involved in speech production may be investigated in the individuals with cognitive impairment. Motor impairment is an important aspect of cognitive decline in older adults. Kluger et al. [380] demonstrated that motor assessment may be comparably sensitive to standard neuropsychological tests of cognitive function in individuals affected by the earliest stages of Alzheimer's disease. Writing is one of the skills of manual dexterity.[7] Analysis of the subject's writing properties, such as pen pressure, writing speed, may provide a window into the cognitive function of the subject.

Compared to standard neuropsychological tests, motor measures may enhance the identification of individuals at risk for cognitive impairment by utilizing methods that are less dependent on levels of education.

8.3 Conclusion

The research questions asked in this research have been answered over the course of ten studies. Speech features have been shown to discriminate between cognitively impaired and healthy individuals. Speech features may provide an objective measure as a biomarker of cognitive function and a complementary measure to existing neuropsychological assessments. This type of cognitive assessment may be administered remotely, thus resulting in a reduction in some of the financial and labour intensive burden affecting healthcare teams and healthcare providers.

Table 8-1 re-lists the research questions posed in Section 2.18 and points to sections where the research questions were addressed.

[7] Manual dexterity is the ability to make coordinated hand and finger movements to grasp and manipulate objects. Manual dexterity includes muscular, skeletal, and neurological functions to produce small, precise movements.

Table 8-1: Status of the research questions

#	Research Question	Status
1	Can the temporal and acoustic features of speech be extracted reliably from speech recordings taken in real life environments, while removing artefacts such as breath sounds and compensating for varying levels and sources of noise?	Addressed in Section 4.2.
2	Can the removal of breath sounds from speech recordings prior to temporal and acoustic features extraction improve the classification performance between cognitively healthy and cognitively impaired individuals?	Addressed in Section 4.3.
3	Does the status of individual's cognitive function correlate with temporal features of speech of individuals with/without schizophrenia extracted from a recording of reading out loud of a short text passage?	Addressed in Section 4.1.
4	Does the status of individual's cognitive function correlate with acoustic features of speech of individuals with/without schizophrenia extracted from a recording of reading out loud of a short text passage?	Addressed in Section 4.1.
5	Does the status of individual's cognitive function correlate with temporal features of speech of older adults extracted from a recording of reading loud a short text passage?	Addressed in Section 5.1.
6	Does the status of individual's cognitive function correlate with acoustic features of speech of older adults extracted from a recording of reading loud a short text passage?	Addressed in Section 5.1.
7	Does the status of individual's cognitive function correlate with temporal features of speech of older adults extracted from a recording of a picture description task?	Addressed in Section 5.1.
8	Does the status of individual's cognitive function correlate with acoustic features of speech of older adults extracted from a recording of a picture description task?	Addressed in Section 5.1.
9	Does the status of individual's memory correlate with temporal features of speech of older adults extracted from a recording of a picture description tasks?	Addressed in Section 5.3.
10	Will the features extracted using dynamic pause threshold from speech recording of reading out loud of a short text passage increase the ability to discriminate between cognitively healthy and cognitively impaired older individuals?	Addressed in Section 5.4.

Table 8-1 cont.: Status of the research questions

#	Research Question	Status
11	Can the speech features be extracted reliably from recordings of speech acquired over a telephone?	Addressed in Section 6.1.
12	What is the protocol that will enable remote administration of clinical cognitive assessment over a telephone?	Addressed in Section 6.2.
13	What are the design requirements for both protocol and implementation for a remote, longitudinal assessment of cognitive function of older individuals?	Addressed in Section 6.2.
14	Can the cognitive assessment be fully automated for large scale deployments using existing telephone infrastructure?	Addressed in Section 6.2.
15	Can the analysis of remotely collected speech recordings provide equivalent information about cognitive function compared to labour intensive gold-standard cognitive assessment?	Addressed in Section 6.2.
16	Can speech be employed to identify changes in cognitive function over time?	Addressed in Section 6.2.
17	Do the results of a face-to-face cognitive assessment correlate with the results of automated cognitive assessment?	Addressed in Section 6.2.
18	Will the older individuals be comfortable using the automated cognitive assessment technology?	Addressed in Section 6.2.
19	Will the cost of remote cognitive assessment be lower than the cost of gold-standard cognitive assessment?	Addressed in Section 6.2.
20	Can an assessment of memory and attention be delivered automatically over a website?	Addressed in Section 7.
21	Can a battery of tasks for the assessment of memory and attention be administered at regular intervals over time without inducing a practice effect?	Addressed in Section 7.
22	Can a battery of tasks for the assessment of memory and attention employ the same stimuli over time without inducing a practice effect?	Addressed in Section 7.

Bibliography

[1] World Health Organization. *Definition of an older or elderly person*. Accessed on: November 23, 2013. Available at: http://www.who.int/healthinfo/survey/ageingdefnolder/en/

[2] Population Division, DESA, United Nations, *World Population Ageing, 1950-2050*. New York, NY: United Nations Publications, 2002.

[3] World Health Organization, "Good health adds life to years. Global brief for World Health Day 2012," 2012.

[4] C. P. Ferri, M. Prince, C. Brayne, H. Brodaty, L. Fratiglioni, M. Ganguli, *et al.*, "Global prevalence of dementia: a Delphi consensus study," *Lancet,* vol. 366, pp. 2112-7, Dec 17 2005.

[5] A. Lobo, L. J. Launer, L. Fratiglioni, K. Andersen, A. Di Carlo, M. M. Breteler, *et al.*, "Prevalence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group," *Neurology,* vol. 54, pp. S4-9, 2000.

[6] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders.*, 4th ed. Washington: American Psychiatric Association, 1994.

[7] B. L. Plassman, K. M. Langa, G. G. Fisher, S. G. Heeringa, D. R. Weir, M. B. Ofstedal, *et al.*, "Prevalence of dementia in the United States: the aging, demographics, and memory study," *Neuroepidemiology,* vol. 29, pp. 125-32, 2007.

[8] R. Brookmeyer, S. Gray, and C. Kawas, "Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset," *Am J Public Health,* vol. 88, pp. 1337-42, Sep 1998.

[9] World Health Organization and Alzheimer's Disease International, *Dementia: a public health priority*: World Health Organization, 2012.

[10] J. Young, D. Meagher, and A. MacLullich, "Cognitive assessment of older people," *BMJ,* vol. 343, 2011-09-07 00:00:00 2011.

[11] M. Prince, R. Bryce, and C. Ferri, "World Alzheimer Report 2011: The benefits of early diagnosis and intervention," Alzheimer's Disease International2011.

[12] M. D. Hurd, P. Martorell, A. Delavande, K. J. Mullen, and K. M. Langa, "Monetary Costs of Dementia in the United States," *New England Journal of Medicine,* vol. 368, pp. 1326-1334, 2013.

[13] E. Q. Wu, H. G. Birnbaum, L. Shi, D. E. Ball, R. C. Kessler, M. Moulis, *et al.*, "The economic burden of schizophrenia in the United States in 2002," *J Clin Psychiatry,* vol. 66, pp. 1122-9, Sep 2005.

[14] G. Rees, A. Eayrs, and A. s. Australia, *Dementia: Facing the Epidemic : a Vision for a World Class Dementia Care System*: Alzheimer's Australia, 2009.

[15] V. Valcour, K. H. Masaki, J. Curb, and P. Blanchette, "THe detection of dementia in the primary care setting," *Archives of Internal Medicine,* vol. 160, pp. 2964-2968, 2000.

[16] M. Olafsdottir, I. Skoog, and J. Marcusson, "Detection of dementia in primary care: the Linkoping study," *Dement Geriatr Cogn Disord,* vol. 11, pp. 223-9, Jul-Aug 2000.

[17] C. H. Wilkins, K. L. Wilkins, M. Meisel, M. Depke, J. Williams, and D. F. Edwards, "Dementia Undiagnosed in Poor Older Adults with Functional Impairment," *Journal of the American Geriatrics Society,* vol. 55, pp. 1771-1776, 2007.

[18] M. Boustani, C. M. Callahan, F. W. Unverzagt, M. G. Austrom, A. J. Perkins, B. A. Fultz, *et al.*, "Implementing a screening and diagnosis program for dementia in primary care," *J Gen Intern Med,* vol. 20, pp. 572-7, Jul 2005.

[19] M. Löppönen, I. Räihä, R. Isoaho, T. Vahlberg, and S.-L. Kivelä, "Diagnosing cognitive impairment and dementia in primary health care – a more active approach is needed," *Age and Ageing,* vol. 32, pp. 606-612, November 1, 2003 2003.

[20] A. Dias and V. Patel, "Closing the treatment gap for dementia in India," *Indian J Psychiatry,* vol. 51 Suppl 1, pp. S93-7, Jan 2009.

[21] UK Department of Health, "Living well with dementia: A National Dementia Strategy," 2009.

[22] C. Basak, W. R. Boot, M. W. Voss, and A. F. Kramer, "Can training in a real-time strategy video game attenuate cognitive decline in older adults?," *Psychology and Aging,* vol. 23, pp. 765-777, 2008.

[23] M. C. Greenaway, G. E. Smith, E. G. Tangalos, Y. E. Geda, and R. J. Ivnik, "Mayo older americans normative studies: factor analysis of an expanded neuropsychological battery," *Clin Neuropsychol,* vol. 23, pp. 7-20, Jan 2009.

[24] S. Belleville, B. Gilbert, F. Fontaine, L. Gagnon, E. Menard, and S. Gauthier, "Improvement of episodic memory in persons with mild cognitive impairment and healthy older adults: evidence from a cognitive intervention program," *Dement Geriatr Cogn Disord,* vol. 22, pp. 486-99, 2006.

[25] G. Cipriani, A. Bianchetti, and M. Trabucchi, "Outcomes of a computer-based cognitive rehabilitation program on Alzheimer's disease patients compared with those on

patients affected by mild cognitive impairment," *Arch Gerontol Geriatr,* vol. 43, pp. 327-35, Nov-Dec 2006.

[26] S. Rapp, G. Brenes, and A. P. Marsh, "Memory enhancement training for older adults with mild cognitive impairment: a preliminary study," *Aging Ment Health,* vol. 6, pp. 5-11, Feb 2002.

[27] E. Talassi, M. Guerreschi, M. Feriani, V. Fedi, A. Bianchetti, and M. Trabucchi, "Effectiveness of a cognitive rehabilitation program in mild dementia (MD) and mild cognitive impairment (MCI): a case control study," *Arch Gerontol Geriatr,* vol. 44 Suppl 1, pp. 391-9, 2007.

[28] L. Jean, M. E. Bergeron, S. Thivierge, and M. Simard, "Cognitive intervention programs for individuals with mild cognitive impairment: systematic review of the literature," *Am J Geriatr Psychiatry,* vol. 18, pp. 281-96, Apr 2010.

[29] R.-Y. Liu, Y. Zhang, D. A. Baxter, P. Smolen, L. J. Cleary, and J. H. Byrne, "Deficit in Long-Term Synaptic Plasticity Is Rescued by a Computationally Predicted Stimulus Protocol," *The Journal of Neuroscience,* vol. 33, pp. 6944-6949, April 17, 2013 2013.

[30] M. F. Folstein, S. E. Folstein, and P. R. McHugh, ""Mini-mental state". A practical method for grading the cognitive state of patients for the clinician," *J Psychiatr Res,* vol. 12, pp. 189-98, Nov 1975.

[31] Z. S. Nasreddine, N. A. Phillips, V. Bedirian, S. Charbonneau, V. Whitehead, I. Collin, *et al.*, "The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment," *J Am Geriatr Soc,* vol. 53, pp. 695-9, Apr 2005.

[32] D. Wechsler, *WMS-III : Wechsler memory scale administration and scoring manual.* San Antonio, Tex.: Psychological Corp., 1997.

[33] H. J. Woodford and J. George, "Cognitive assessment in the elderly: a review of clinical methods," *QJM,* vol. 100, pp. 469-484, August 1, 2007 2007.

[34] J. Grace and M. M. Amick, "Cognitive screening of older adults," *Med Health R I,* vol. 88, pp. 8-11, Jan 2005.

[35] T. N. Tombaugh and N. J. McIntyre, "The mini-mental state examination: a comprehensive review," *J Am Geriatr Soc,* vol. 40, pp. 922-35, Sep 1992.

[36] J. T. Olin and E. M. Zelinski, "The 12-month reliability of the Mini-Mental State Examination," *Psychological Assessment: A Journal of Consulting and Clinical Psychology,* vol. 3, pp. 427-432, 1991.

[37] L. G. Chepenik, L. A. Cornew, and M. J. Farah, "The influence of sad mood on cognition," *Emotion,* vol. 7, pp. 802-811, 2007.

[38] D. W. O'Connor, P. A. Pollitt, J. B. Hyde, J. L. Fellows, N. D. Miller, C. P. Brook, *et al.*, "The reliability and validity of the Mini-Mental State in a British community survey," *J Psychiatr Res,* vol. 23, pp. 87-96, 1989.

[39] A. H. Lo, M. Humphreys, G. J. Byrne, and N. A. Pachana, "Test-retest reliability and practice effects of the Wechsler Memory Scale-III," *J Neuropsychol,* vol. 6, pp. 212-31, Sep 2012.

[40] C. Jagger, M. Clarke, J. Anderson, and T. Battcock, "Misclassification of dementia by the mini-mental state examination--are education and social class the only factors?," *Age Ageing,* vol. 21, pp. 404-11, Nov 1992.

[41] G. Thierry, "Neurophysiological examination methods: electrophysiology and neuroimaging," in *Cognitive Neurology: A clinical textbook* S. Cappa, J. Abutalebi, J.-F. Demonet, P. Fletcher, and P. Garrard, Eds., ed New York Oxford University Press, 2008.

[42] R. C. Petersen, *Mild Cognitive Impairment : Aging to Alzheimer's Disease*: Oxford University Press, USA, 2003.

[43] R. P. Friedland, T. F. Budinger, E. Ganz, Y. Yano, C. A. Mathis, B. Koss, *et al.*, "Regional cerebral metabolic alterations in dementia of the Alzheimer type: positron emission tomography with [18F]fluorodeoxyglucose," *J Comput Assist Tomogr,* vol. 7, pp. 590-8, Aug 1983.

[44] N. L. Foster, T. N. Chase, L. Mansi, R. Brooks, P. Fedio, N. J. Patronas, *et al.*, "Cortical abnormalities in Alzheimer's disease," *Ann Neurol,* vol. 16, pp. 649-54, Dec 1984.

[45] L. J. Streletz, P. F. Reyes, M. Zalewska, L. Katz, and R. G. Fariello, "Computer analysis of EEG activity in dementia of the Alzheimer's type and Huntington's disease," *Neurobiology of Aging,* vol. 11, pp. 15-20, 1990.

[46] C. Rosenberg, K. Nudleman, and A. Starr, "Cognitive evoked potentials (p300) in early huntington's disease," *Archives of Neurology,* vol. 42, pp. 984-987, 1985.

[47] D. S. Goodin and M. J. Aminoff, "Electrophysiological differences between subtypes of dementia," *Brain,* vol. 109 (Pt 6), pp. 1103-13, Dec 1986.

[48] E. Altay, E. Fisher, S. E. Jones, C. Hara-Cleaver, J. Lee, and R. A. Rudick, "Reliability of classifying multiple sclerosis disease activity using magnetic resonance imaging in a multiple sclerosis clinic," *JAMA Neurology,* vol. 70, pp. 338-344, 2013.

[49] N. Colledge, R. Sellar, J. Wardlaw, S. Lewis, G. Mead, and J. Wilson, "Interobserver Agreement in Magnetic Resonance Brain and Neck Imaging," *Journal of Neuroimaging,* vol. 16, pp. 47-51, 2006.

[50] P. Scheltens, L. Launer, F. Barkhof, H. Weinstein, and W. Gool, "Visual assessment of medial temporal lobe atrophy on magnetic resonance imaging: Interobserver reliability," *Journal of Neurology,* vol. 242, pp. 557-560, 1995/09/01 1995.

[51] A. J. Atkinson, W. A. Colburn, V. G. DeGruttola, D. L. DeMets, G. J. Downing, D. F. Hoth, *et al.*, "Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework*," *Clin Pharmacol Ther,* vol. 69, pp. 89-95, 2001.

[52] K. L. Poston, "Development of Multimodal Imaging Biomarkers for Cognitive Dysfunction in Parkinson's Disease," ed. Stanford , United States: The Michael J. Fox Foundation 2011.

[53] J. J. Gomar, M. T. Bobes-Bascaran, C. Conejero-Goldberg, P. Davies, T. E. Goldberg, and A. s. D. N. Initiative, "Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to alzheimer disease in patients in the Alzheimer's disease neuroimaging initiative," *Archives of General Psychiatry,* vol. 68, pp. 961-969, 2011.

[54] C. M. Roe, A. M. Fagan, E. A. Grant, J. Hassenstab, K. L. Moulder, D. Maue Dreyfus, *et al.*, "Amyloid imaging and CSF biomarkers in predicting cognitive impairment up to 7.5 years later," *Neurology,* Apr 10 2013.

[55] K. S. Sheinerman, V. G. Tsivinsky, F. Crawford, M. J. Mullan, L. Abdullah, and S. R. Umansky, "Plasma microRNA biomarkers for detection of mild cognitive impairment," *Aging (Albany NY),* vol. 4, pp. 590-605, Sep 2012.

[56] D. H. Salat, J. A. Kaye, and J. S. Janowsky, "Selective Preservation and Degeneration Within the Prefrontal Cortex in Aging and Alzheimer Disease," *Arch Neurol,* vol. 58, pp. 1403-1408, September 1, 2001 2001.

[57] T. Ohnishi, H. Matsuda, T. Tabira, T. Asada, and M. Uno, "Changes in Brain Morphology in Alzheimer Disease and Normal Aging: Is Alzheimer Disease an Exaggerated Aging Process?," *AJNR Am J Neuroradiol,* vol. 22, pp. 1680-1685, October 1, 2001 2001.

[58] L. W. de Jong, K. van der Hiele, I. M. Veer, J. J. Houwing, R. G. J. Westendorp, E. L. E. M. Bollen, *et al.*, "Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study," *Brain,* vol. 131, pp. 3277-3285, December 1, 2008 2008.

[59] S. Singh, R. S. Bucks, and J. M. Cuerden, "Evaluation of an objective technique for analysing temporal variables in DAT spontaneous speech," *Aphasiology,* vol. 15, pp. 571-583, 2001/06/01 2001.

[60] B. Roark, M. Mitchell, J. Hosom, K. Hollingshead, and J. Kaye, "Spoken Language Derived Measures for Detecting Mild Cognitive Impairment," *Audio, Speech, and Language Processing, IEEE Transactions on,* vol. 19, pp. 2081-2090, 2011.

[61] B. Roark, J. Hosom, M. Mitchell, and J. Kaye, "Automatically Derived Spoken Language Markers for Detecting Mild Cognitive Impairment," in *2nd International Conference on Technology and Aging (ICTA)*, Toronto, Canada, 2007.

[62] S. R. Bucks, S. Singh, J. M. Cuerden, and G. K. Wilcock, "Linguistic Analysis of Spontaneous Speech in Probable Alzheimer's Disease: A Comparison with Normal Older Adults," in *Clinical Phonetics and Linguistics*, W. Ziegler and K. Deger, Eds., ed London, England: Whurr Publishers, 1998, pp. 312 - 320.

[63] A. Lowit, B. Brendel, C. Dobinson, and P. Howell, "An investigation into the influences of age, pathology and cognition on speech production," *J Med Speech Lang Pathol,* vol. 14, pp. 253-262, Dec 1 2006.

[64] United Nations, "Report of the Second World Assembly on Aging," Madrid, Spain April 8-12, 2002.

[65] E. L. Glisky, "Changes in cognitive function in human aging," in *Brain aging: Models, methods, and mechanisms*, D. R. Riddle, Ed., ed Boca Raton, FL, US: CRC Press, 2007, pp. 3-20.

[66] E. P. Apfelbaum, A. C. Krendl, and N. Ambady, "Age-related decline in executive function predicts better advice-giving in uncomfortable social contexts," *Journal of Experimental Social Psychology,* vol. 46, pp. 1074-1077, 2010.

[67] J. M. McDowd and R. J. Shaw, "Attention and ageing: a functional perspective," in *The Handbook of Ageing and Cognition*, F. I. M. Craik and T. A. Salthouse, Eds., ed. Mahwah, NJ: Lawrence Erlbaum Associates, Incorporated, 2000, p. 221.

[68] E. Strauss, E. M. S. Sherman, and O. Spreen, *A compendium of neuropsychological tests : administration, norms, and commentary*. Oxford; New York: Oxford University Press, 2006.

[69] P. Verhaeghen and J. Cerella, "Aging, executive control, and attention: a review of meta-analyses," *Neuroscience & Biobehavioral Reviews,* vol. 26, pp. 849-857, 2002.

[70] J. M. McDowd and F. I. Craik, "Effects of aging and task difficulty on divided attention performance," *J Exp Psychol Hum Percept Perform,* vol. 14, pp. 267-80, May 1988.

[71] P. S. Tsang and T. L. Shaner, "Age, attention, expertise, and time-sharing performance," *Psychol Aging,* vol. 13, pp. 323-47, Jun 1998.

[72] A. F. Kramer, J. L. Larish, T. A. Weber, and L. Bardell, "Training for executive control: Task coordination strategies and aging," in *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application*, D. G. A. Koriat, Ed., ed Cambridge, MA, US: The MIT Press, 1999, pp. 617-652.

[73] H. L. Hawkins, A. F. Kramer, and D. Capaldi, "Aging, exercise, and attention," *Psychology and Aging,* vol. 7, pp. 643-653, 1992.

[74] S. Colcombe and A. F. Kramer, "Fitness Effects on the Cognitive Function of Older Adults: A Meta-Analytic Study," *Psychological Science,* vol. 14, pp. 125-130, March 1, 2003 2003.

[75] R. T. Zacks, L. Hasher, and K. Z. H. Li, "Human memory," in *The Handbook of Ageing and Cognition*, F. I. M. Craik and T. A. Salthouse, Eds., ed. Mahwah, NJ: Lawrence Erlbaum Associates, Incorporated, 2000, p. 293.

[76] J. D. Kester, A. S. Benjamin, A. D. Castel, and F. I. M. Craik, "Memory in elderly people," in *The Handbook of Memory Disorders*, A. D. Baddeley, M. D. Kopelman, and B. A. Wilson, Eds., 2 ed West Sussex, U.K.: Wiley, 2002, p. 543.

[77] D. C. Park and T. Hedden, "Working memory and ageing," in *Perspectives on Human Memory and Cognitive Ageing: Essays in Honour of Fergus Craik*, M. Naveh-Benjamin, M. Moscovitch, and H. I. Roediger, Eds., ed New York: Psychology Press, 2001, p. 148.

[78] P. A. Reuter-Lorenz and C.-Y. C. Sylvester, "The Cognitive Neuroscience of Working Memory and Aging," in *Cognitive Neuroscience of Aging*, R. Cabeza, L. Nyberg, and D. Park, Eds., ed Oxford: Oxford University Press, 2005, p. 186.

[79] A. Baddeley and G. Hitch, "Working memory," in *The psychology of learning and motivation: Advances in research and theory*. vol. 8, G. H. Bower, Ed., ed New York: Academic Press, 1974, pp. 47-89.

[80] T. D. Wager and E. E. Smith, "Neuroimaging studies of working memory: a meta-analysis," *Cogn Affect Behav Neurosci,* vol. 3, pp. 255-74, Dec 2003.

[81] D. C. Park, "The basic mechanisms accounting for age-related decline in cognitive function," in *Cognitive aging: A primer*, D. C. P. N. Schwarz, Ed., ed New York, NY, US: Psychology Press, 2000, pp. 3-21.

[82] F. I. M. Craik and M. Byrd, "Aging and cognitive deficits: The role of attentional resources," in *Aging and Cognitive Processes*, F. I. M. Craik and S. E. Trehub, Eds., ed New York: Plenum, 1982, pp. 191-211.

[83] F. I. M. Craik, "A functional account of age differences in memory," in *Human Memory and Cognitive Capabilities: Mechanisms and Performances*. vol. 5, F. Klix and H.

Hagendorf, Eds., ed Amsterdam, North-Holland: Elsevier Science Publishing Company, 1986, pp. 409-422.

[84] T. A. Salthouse, "Processing capacity and its role on the relations between age and memory," in *Memory performance and competencies: Issues in growth and development*, F. E. W. W. Schneider, Ed., ed Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc, 1995, pp. 111-125.

[85] T. A. Salthouse, "The aging of working memory," *Neuropsychology,* vol. 8, pp. 535-543, 1994.

[86] T. A. Salthouse, "The processing-speed theory of adult age differences in cognition," *Psychological Review,* vol. 103, pp. 403-428, 1996.

[87] D. C. Park, A. D. Smith, G. Lautenschlager, J. L. Earles, D. Frieske, M. Zwahr, *et al.*, "Mediators of long-term memory performance across the life span," *Psychology and Aging,* vol. 11, pp. 621-637, 1996.

[88] L. Hasher and R. T. Zacks, "Working memory, comprehension, and aging: A review and a new view," in *The Psychology of Learning and Motivation*. vol. 22, G. H. Bower, Ed., ed New York: Academic Press, 1988, pp. 193-225.

[89] L. Hasher, R. T. Zacks, and C. P. May, "Inhibitory control, circadian arousal, and age," in *Attention and Performance XVII*, D. Gopher and A. Koriat, Eds., ed Cambridge, MA: MIT Press, 1999, pp. 653-675.

[90] C. P. May, L. Hasher, and M. J. Kane, "The role of interference in memory span," *Mem Cognit,* vol. 27, pp. 759-67, Sep 1999.

[91] T. Hedden and D. Park, "Aging and interference in verbal working memory," *Psychology and Aging,* vol. 16, pp. 666-681, 2001.

[92] E. Tulving, "Episodic memory: from mind to brain," *Annu Rev Psychol,* vol. 53, pp. 1-25, 2002.

[93] F. I. M. Craik, D. a. Routh, and D. E. Broadbent, "On the Transfer of Information from Temporary to Permanent Memory [and Discussion]," *Philosophical Transactions of the Royal Society of London. B, Biological Sciences,* vol. 302, pp. 341-359, August 11, 1983 1983.

[94] E. L. Glisky, S. R. Rubin, and P. S. R. Davidson, "Source memory in older adults: An encoding or retrieval problem?," *Journal of Experimental Psychology: Learning, Memory, and Cognition,* vol. 27, pp. 1131-1146, 2001.

[95] J. M. Jennings and L. L. Jacoby, "An opposition procedure for detecting age-related deficits in recollection: Telling effects of repetition," *Psychology and Aging,* vol. 12, pp. 352-361, 1997.

[96] P. R. Davidson and E. Glisky, "Neuropsychological correlates of recollection and familiarity in normal aging," *Cognitive, Affective, & Behavioral Neuroscience,* vol. 2, pp. 174-186, 2002/06/01 2002.

[97] S. F. Nolde, M. K. Johnson, and M. D'Esposito, "Left prefrontal activation during episodic remembering: an event-related fMRI study," *Neuroreport,* vol. 9, pp. 3509-14, Oct 26 1998.

[98] L. L. Light, "The organization of memory in old age," in *The Handbook of Ageing and Cognition*, F. I. M. Craik and T. A. Salthouse, Eds., ed. Hillsdale, NJ: Lawrence Erlbaum Associates, Incorporated, 1992, p. 111.

[99] D. C. Rubin, "Autobiographical memory and ageing," in *Cognitive Ageing: A Primer*, D. Park and N. Schwarz, Eds., ed Philadelphia, PA: Psychology Press, 2000, p. 131.

[100] B. Levine, E. Svoboda, J. F. Hay, G. Winocur, and M. Moscovitch, "Aging and autobiographical memory: Dissociating episodic from semantic retrieval," *Psychology and Aging,* vol. 17, pp. 677-689, 2002.

[101] P. S. R. Davidson and E. L. Glisky, "Is flashbulb memory a special instance of source memory? Evidence from older adults," *Memory,* vol. 10, pp. 99-111, 2002/03/01 2002.

[102] P. S. Davidson, S. P. Cook, and E. L. Glisky, "Flashbulb memories for September 11th can be preserved in older adults," *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn,* vol. 13, pp. 196-206, Jun 2006.

[103] T. A. Salthouse, "Effects of age and skill in typing," *J Exp Psychol Gen,* vol. 113, pp. 345-71, Sep 1984.

[104] M. W. Prull, J. D. E. Gabrieli, and S. A. Bunge, "Age-related changes in memory: a cognitive neuroscience perspective," in *The Handbook of Ageing and Cognition*, F. I. M. Craik and T. A. Salthouse, Eds., ed. Mahwah, NJ: Lawrence Erlbaum Associates, Incorporated, 2000, p. 91.

[105] R. West, "The neural basis of age-related declines in prospective memory," in *Cognitive Neuroscience of Aging*, R. Cabeza, L. Nyberg, and D. Park, Eds., ed Oxford: Oxford University Press, 2005, p. 246.

[106] M. K. Johnson, S. Hashtroudi, and D. S. Lindsay, "Source monitoring," *Psychological Bulletin,* vol. 114, pp. 3-28, 1993.

[107] F. I. M. Craik, "Levels of processing: Past, present... and future?," *Memory,* vol. 10, pp. 305-318, 2002/09/01 2002.

[108] B. A. Schneider and M. K. Pichora-Fuller, "Implications of perceptual deterioration for cognitive ageing research," in *The Handbook of Ageing and Cognition*, F. I. M. Craik and T. A. Salthouse, Eds., ed. Mahwah, NJ: Lawrence Erlbaum Associates, Incorporated, 2000, p. 155.

[109] P. B. Baltes and U. Lindenberger, "Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging?," *Psychology and Aging,* vol. 12, pp. 12-21, 1997.

[110] S. Kemper and K. Kemtes, "Ageing and message production and comprehension," in *Cognitive Ageing: A Primer*, D. Park and N. Schwarz, Eds., ed Philadelphia, PA: Psychology Press, 2000, p. 197.

[111] A. Wingfield, "Speech perception and the comprehension of spoken language in adult ageing," in *Cognitive Ageing: A Primer*, D. Park and N. Schwarz, Eds., ed Philadelphia, PA: Psychology Press, 2000, p. 175.

[112] A. Wingfield and E. A. L. Stine-Morrow, "Language and speech," in *The Handbook of Ageing and Cognition*, F. I. M. Craik and T. A. Salthouse, Eds., ed. Mahwah, NJ: Lawrence Erlbaum Associates, Incorporated, 2000, p. 359.

[113] A. G. Sanfey and R. Hastie, "Judgment and decision making across the adult life span: a tutorial review of psychological research," in *Cognitive Ageing: A Primer*, D. Park and N. Schwarz, Eds., ed Philadelphia, PA: Psychology Press, 2000, p. 253.

[114] A. Baddeley, "Fractionating the Central Executive," in *Principles of Frontal Lobe Function* D. T. Stuss and R. T. Knight, Eds., ed Oxford: Oxford University Press, 2002, p. 246.

[115] R. L. West, "An application of prefrontal cortex function theory to cognitive aging," *Psychological Bulletin,* vol. 120, pp. 272-292, 1996.

[116] N. Raz, "Ageing of the brain and its impact on cognitive performance: integration of structural and functional findings," in *The Handbook of Ageing and Cognition*, F. I. M. Craik and T. A. Salthouse, Eds., ed. Mahwah, NJ: Lawrence Erlbaum Associates, Incorporated, 2000, p. 1.

[117] E. L. Glisky, M. R. Polster, and B. C. Routhieaux, "Double dissociation between item and source memory," *Neuropsychology,* vol. 9, pp. 229-235, 1995.

[118] C. L. Grady, L. J. Bernstein, S. Beig, and A. L. Siegenthaler, "The effects of encoding task on age-related differences in the functional neuroanatomy of face memory," *Psychology and Aging,* vol. 17, pp. 7-23, 2002.

[119] R. Cabeza, N. D. Anderson, J. K. Locantore, and A. R. McIntosh, "Aging Gracefully: Compensatory Brain Activity in High-Performing Older Adults," *NeuroImage,* vol. 17, pp. 1394-1402, 2002.

[120] J. M. Logan, A. L. Sanders, A. Z. Snyder, J. C. Morris, and R. L. Buckner, "Under-Recruitment and Nonselective Recruitment: Dissociable Neural Mechanisms Associated with Aging," *Neuron,* vol. 33, pp. 827-840, 2002.

[121] S. Daselaar and R. Cabeza, "Age-Related Changes in Hemispheric Organization," in *Cognitive Neuroscience of Aging*, R. Cabeza, L. Nyberg, and D. Park, Eds., ed Oxford: Oxford University Press, 2005, p. 186.

[122] L. Ryan, C. Hatfield, and M. Hofstetter, "Caffeine reduces time-of-day effects on memory performance in older adults," *Psychol Sci,* vol. 13, pp. 68-71, Jan 2002.

[123] S. Gao, H. C. Hendrie, K. S. Hall, and S. Hui, "The relationships between age, sex, and the incidence of dementia and Alzheimer disease: a meta-analysis," *Arch Gen Psychiatry,* vol. 55, pp. 809-15, Sep 1998.

[124] W. A. Kukull and M. Ganguli, "Epidemiology of dementia: concepts and overview," *Neurol Clin,* vol. 18, pp. 923-50, Nov 2000.

[125] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen, "Mild cognitive impairment: clinical characterization and outcome," *Arch Neurol,* vol. 56, pp. 303-8, Mar 1999.

[126] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, "Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease," *Neurology,* vol. 34, pp. 939-44, Jul 1984.

[127] T. J. Goldschmidt, R. Mallin, and C. N. Still, "Recognition of cognitive impairment in primary care outpatients," *South Med J,* vol. 76, pp. 1264-5, 1270, Oct 1983.

[128] R. A. Murden, T. D. McRae, S. Kaner, and M. E. Bucknam, "Mini-Mental State exam scores vary with education in blacks and whites," *J Am Geriatr Soc,* vol. 39, pp. 149-55, Feb 1991.

[129] E. G. Tangalos, G. E. Smith, R. J. Ivnik, R. C. Petersen, E. Kokmen, L. T. Kurland, *et al.*, "The Mini-Mental State Examination in general medical practice: clinical utility and acceptance," *Mayo Clin Proc,* vol. 71, pp. 829-37, Sep 1996.

[130] L. George, R. Landerman, D. Blazer, and J. Anthony, "Cognitive impairment," in *Psychiatric Disorders in America: The Epidemiologic Catchment Area Study*, L. N. Robins and D. M. Reiger, Eds., ed New York: The Free Press, 1991, pp. 291-327.

[131] R. A. Ouvrier, R. F. Goldsmith, S. Ouvrier, and I. C. Williams, "The value of the Mini-Mental State Examination in childhood: a preliminary study," *J Child Neurol,* vol. 8, pp. 145-8, Apr 1993.

[132] *MoCA News Update April 2, 2013*. Accessed on: April 26, 2013. Available at: http://www.mocatest.org/moca_news.asp

[133] Z. Ismail, T. K. Rajji, and K. I. Shulman, "Brief cognitive screening instruments: an update," *International Journal of Geriatric Psychiatry,* vol. 25, pp. 111-120, 2010.

[134] H. E. Nelson and J. R. Willison, *The Revised National Adult Reading Test - Test Manual*. Windsor, UK: NFER-Nelson, 1991.

[135] C. M. MacLeod, "Half a century of research on the Stroop effect: An integrative review," *Psychological Bulletin,* vol. 109, pp. 163-203, 1991.

[136] P. Graf, B. Uttl, and H. Tuokko, "Color- and picture-word Stroop tests: performance changes in old age," *J Clin Exp Neuropsychol,* vol. 17, pp. 390-415, May 1995.

[137] J. Henry and J. Crawford, "A meta-analytic review of verbal fluency deficits in schizophrenia relative to other neurocognitive deficits," *Cognitive Neuropsychiatry,* vol. 10, pp. 1-33, 2005/01/01 2005.

[138] J. D. Henry and J. R. Crawford, "A Meta-Analytic Review of Verbal Fluency Deficits in Depression," *Journal of Clinical and Experimental Neuropsychology,* vol. 27, pp. 78-101, 2005/02/23 2005.

[139] I. H. Robertson, T. Manly, J. Andrade, B. T. Baddeley, and J. Yiend, "'Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects," *Neuropsychologia,* vol. 35, pp. 747-758, 1997.

[140] J. A. Grahn and T. Manly, "Common Neural Recruitment across Diverse Sustained Attention Tasks," *PLoS ONE,* vol. 7, p. e49556, 2012.

[141] M. P. Lawton and E. M. Brody, "Assessment of older people: self-maintaining and instrumental activities of daily living," *Gerontologist,* vol. 9, pp. 179-86, Autumn 1969.

[142] R. M. Bertrand, S. L. Willis, and A. Sayer, "An Evaluation of Change Over Time in Everyday Cognitive Competence Among Alzheimer's Patients," *Aging, Neuropsychology, and Cognition,* vol. 8, pp. 192-212, 2001/09/01 2001.

[143] P. A. Boyle, R. A. Cohen, R. Paul, D. Moser, and N. Gordon, "Cognitive and motor impairments predict functional declines in patients with vascular dementia," *Int J Geriatr Psychiatry,* vol. 17, pp. 164-9, Feb 2002.

[144] L. S. Radloff, "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population," *Applied Psychological Measurement,* vol. 1, pp. 385-401, June 1, 1977 1977.

[145] R. N. Carleton, M. A. Thibodeau, M. J. N. Teale, P. G. Welch, M. P. Abrams, T. Robinson, *et al.*, "The Center for Epidemiologic Studies Depression Scale: A Review with a Theoretical and Empirical Examination of Item Content and Factor Structure," *PLoS ONE,* vol. 8, p. e58067, 2013.

[146] A. S. Zigmond and R. P. Snaith, "The hospital anxiety and depression scale," *Acta Psychiatr Scand,* vol. 67, pp. 361-70, Jun 1983.

[147] I. Bjelland, A. A. Dahl, T. T. Haug, and D. Neckelmann, "The validity of the Hospital Anxiety and Depression Scale: An updated literature review," *Journal of Psychosomatic Research,* vol. 52, pp. 69-77, 2002.

[148] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the PANAS scales," *J Pers Soc Psychol,* vol. 54, pp. 1063-70, Jun 1988.

[149] S. A. Beaudreau and R. O'Hara, "The association of anxiety and depressive symptoms with cognitive performance in community-dwelling older adults," *Psychol Aging,* vol. 24, pp. 507-12, Jun 2009.

[150] P. L. Ackerman and R. Kanfer, "Test length and cognitive fatigue: An empirical examination of effects on performance and test-taker reactions," *Journal of Experimental Psychology: Applied,* vol. 15, pp. 163-181, 2009.

[151] M. Folstein, J. C. Anthony, I. Parhad, B. Duffy, and E. M. Gruenberg, "The meaning of cognitive impairment in the elderly," *J Am Geriatr Soc,* vol. 33, pp. 228-35, Apr 1985.

[152] L. N. Robins, L. Cottler, K. Buchholz, and W. Compton, *The diagnostic interview schedule, version IV*. St. Louis, MO: Washington University, 1995.

[153] J. W. Ashford, P. Kolm, J. A. Colliver, C. Bekian, and L. N. Hsu, "Alzheimer patient evaluation and the mini-mental state: item characteristic curve analysis," *J Gerontol,* vol. 44, pp. P139-46, Sep 1989.

[154] W. W. Beatty and D. E. Goodkin, "Screening for cognitive impairment in multiple sclerosis. An evaluation of the Mini-Mental State Examination," *Arch Neurol,* vol. 47, pp. 297-301, Mar 1990.

[155] D. Galasko, M. R. Klauber, C. R. Hofstetter, D. P. Salmon, B. Lasker, and L. J. Thal, "The Mini-Mental State Examination in the early diagnosis of Alzheimer's disease," *Arch Neurol,* vol. 47, pp. 49-52, Jan 1990.

[156] E. L. Teng and H. C. Chui, "The Modified Mini-Mental State (3MS) examination," *J Clin Psychiatry,* vol. 48, pp. 314-8, Aug 1987.

[157] M. L. Bleecker, K. Bolla-Wilson, C. Kawas, and J. Agnew, "Age-specific norms for the Mini-Mental State Exam," *Neurology,* vol. 38, pp. 1565-8, Oct 1988.

[158] J. F. Farber, F. A. Schmitt, and P. E. Logue, "Predicting intellectual level from the Mini-Mental State Examination," *J Am Geriatr Soc,* vol. 36, pp. 509-10, Jun 1988.

[159] B. J. Mowry and P. W. Burvill, "A study of mild dementia in the community using a wide range of diagnostic criteria," *Br J Psychiatry,* vol. 153, pp. 328-34, Sep 1988.

[160] J. C. Morris, A. Heyman, R. C. Mohs, J. P. Hughes, G. van Belle, G. Fillenbaum, *et al.*, "The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease," *Neurology,* vol. 39, pp. 1159-65, Sep 1989.

[161] M. Roth, E. Tym, C. Q. Mountjoy, F. A. Huppert, H. Hendrie, S. Verma, *et al.*, "CAMDEX. A standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia," *Br J Psychiatry,* vol. 149, pp. 698-709, Dec 1986.

[162] G. G. Fillenbaum, A. Heyman, W. E. Wilkinson, and C. S. Haynes, "Comparison of two screening tests in Alzheimer's disease. The correlation and reliability of the Mini-Mental State Examination and the modified Blessed test," *Arch Neurol,* vol. 44, pp. 924-7, Sep 1987.

[163] E. A. Zillmer, P. C. Fowler, H. N. Gutnick, and E. Becker, "Comparison of two cognitive bedside screening instruments in nursing home residents: a factor analytic study," *J Gerontol,* vol. 45, pp. P69-74, Mar 1990.

[164] H. R. Bird, G. Canino, M. R. Stipec, and P. Shrout, "Use of the Mini-mental State Examination in a probability sample of a Hispanic population," *J Nerv Ment Dis,* vol. 175, pp. 731-7, Dec 1987.

[165] A. F. Jorm, R. Scott, J. S. Cullen, and A. J. MacKinnon, "Performance of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) as a screening test for dementia," *Psychol Med,* vol. 21, pp. 785-90, Aug 1991.

[166] A. F. Jorm, R. Scott, A. S. Henderson, and D. W. Kay, "Educational level differences on the Mini-Mental State: the role of test bias," *Psychol Med,* vol. 18, pp. 727-31, Aug 1988.

[167] J. Magaziner, S. S. Bassett, and J. R. Hebel, "Predicting performance on the Mini-Mental State Examination. Use of age- and education-specific equations," *J Am Geriatr Soc,* vol. 35, pp. 996-1000, Nov 1987.

[168] D. W. Kay, A. S. Henderson, R. Scott, J. Wilson, D. Rickwood, and D. A. Grayson, "Dementia and depression among the elderly living in the Hobart community: the effect of the diagnostic criteria on the prevalence rates," *Psychol Med,* vol. 15, pp. 771-88, Nov 1985.

[169] J. A. Frazier, J. N. Giedd, S. D. Hamburger, K. E. Albus, D. Kaysen, A. C. Vaituzis, *et al.*, "Brain anatomic magnetic resonance imaging in childhood-onset schizophrenia," *Arch Gen Psychiatry,* vol. 53, pp. 617-24, Jul 1996.

[170] J. I. Escobar, A. Burnam, M. Karno, A. Forsythe, J. Landsverk, and J. M. Golding, "Use of the Mini-Mental State Examination (MMSE) in a community population of mixed ethnicity. Cultural and linguistic artifacts," *J Nerv Ment Dis,* vol. 174, pp. 607-14, Oct 1986.

[171] G. Livingston, K. Sax, J. Willison, B. Blizard, and A. Mann, "The Gospel Oak Study stage II: the diagnosis of dementia in the community," *Psychol Med,* vol. 20, pp. 881-91, Nov 1990.

[172] L. E. Klein, R. P. Roca, J. McArthur, G. Vogelsang, G. B. Klein, S. M. Kirby, *et al.*, "Diagnosing dementia. Univariate and multivariate analyses of the mental status examination," *J Am Geriatr Soc,* vol. 33, pp. 483-8, Jul 1985.

[173] R. I. Pfeffer, T. T. Kurosaki, J. M. Chance, S. Filos, and D. Bates, "Use of the mental function index in older adults. Reliability, validity, and measurement of change over time," *Am J Epidemiol,* vol. 120, pp. 922-35, Dec 1984.

[174] L. J. Thal, M. Grundman, and R. Golden, "Alzheimer's disease: a correlational analysis of the Blessed Information-Memory-Concentration Test and the Mini-Mental State Exam," *Neurology,* vol. 36, pp. 262-4, Feb 1986.

[175] H. J. Keating, 3rd, ""Studying" for the Mini-Mental Status Exam," *J Am Geriatr Soc,* vol. 35, pp. 594-5, Jun 1987.

[176] M. Mitrushina and P. Satz, "Reliability and validity of the Mini-Mental State Exam in neurologically intact elderly," *J Clin Psychol,* vol. 47, pp. 537-43, Jul 1991.

[177] C. A. Peters, J. F. Potter, and S. G. Scholer, "Hearing impairment as a predictor of cognitive decline in dementia," *J Am Geriatr Soc,* vol. 36, pp. 981-6, Nov 1988.

[178] H. W. Ward, J. W. Ramsdell, J. E. Jackson, M. Renvall, J. A. Swart, and E. Rockwell, "Cognitive function testing in comprehensive geriatric assessment. A comparison of cognitive test performance in residential and clinic settings," *J Am Geriatr Soc,* vol. 38, pp. 1088-92, Oct 1990.

[179] G. G. Fillenbaum, D. C. Hughes, A. Heyman, L. K. George, and D. G. Blazer, "Relationship of health and demographic characteristics to Mini-Mental State examination score among community residents," *Psychol Med,* vol. 18, pp. 719-26, Aug 1988.

[180] R. F. Uhlmann and E. B. Larson, "Effect of education on the mini-mental state examination as a screening test for dementia," *J Am Geriatr Soc,* vol. 39, pp. 876-80, Sep 1991.

[181] C. Brayne and P. Calloway, "The association of education and socioeconomic status with the Mini Mental State Examination and the clinical diagnosis of dementia in elderly people," *Age Ageing,* vol. 19, pp. 91-6, Mar 1990.

[182] J. C. Anthony, L. LeResche, U. Niaz, M. R. von Korff, and M. F. Folstein, "Limits of the 'Mini-Mental State' as a screening test for dementia and delirium among hospital patients," *Psychol Med,* vol. 12, pp. 397-408, May 1982.

[183] L. Tsai and M. T. Tsuang, "The Mini-Mental State Test and computerized tomography," *Am J Psychiatry,* vol. 136, pp. 436-8, Apr 1979.

[184] C. E. Holzer, G. L. Tischler, P. J. Leaf, and J. K. Myers, "An epidemiologic assessment of cognitive impairment in a community population," *Research in Community & Mental Health,* vol. 4, pp. 3-32, 1984.

[185] R. M. Crum, J. C. Anthony, S. S. Bassett, and M. F. Folstein, "Population-based norms for the Mini-Mental State Examination by age and educational level," *JAMA,* vol. 269, pp. 2386-91, May 12 1993.

[186] A. W. Wind, F. G. Schellevis, G. Van Staveren, R. P. Scholten, C. Jonker, and J. T. Van Eijk, "Limitations of the Mini-Mental State Examination in diagnosing dementia in general practice," *Int J Geriatr Psychiatry,* vol. 12, pp. 101-8, Jan 1997.

[187] T. K. Rajji, D. Miranda, B. H. Mulsant, M. Lotz, P. Houck, M. D. Zmuda, *et al.*, "The MMSE is not an adequate screening cognitive instrument in studies of late-life depression," *J Psychiatr Res,* vol. 43, pp. 464-70, Jan 2009.

[188] A. J. Lerner and M. McKee, "P3-063: Receiver Operator Characteristics (ROC) of the Mini Mental State Examination (MMSE) relative to the Clinical Dementia Rating (CDR) scale," *Alzheimer's & dementia : the journal of the Alzheimer's Association,* vol. 2, p. S391, 2006.

[189] J. C. Morris, "The Clinical Dementia Rating (CDR): current version and scoring rules," *Neurology,* vol. 43, pp. 2412-4, Nov 1993.

[190] J. C. Morris, M. Storandt, J. P. Miller, D. W. McKeel, J. L. Price, E. H. Rubin, *et al.*, "Mild cognitive impairment represents early-stage Alzheimer disease," *Arch Neurol,* vol. 58, pp. 397-405, Mar 2001.

[191] B. Reisberg, S. H. Ferris, M. J. de Leon, and T. Crook, "The Global Deterioration Scale for assessment of primary degenerative dementia," *Am J Psychiatry,* vol. 139, pp. 1136-9, Sep 1982.

[192] A. Kluger, S. H. Ferris, J. Golomb, M. S. Mittelman, and B. Reisberg, "Neuropsychological Prediction of Decline to Dementia in Nondemented Elderly," *Journal of Geriatric Psychiatry and Neurology,* vol. 12, pp. 168-179, December 1, 1999 1999.

[193] C. Flicker, S. H. Ferris, and B. Reisberg, "Mild cognitive impairment in the elderly: predictors of dementia," *Neurology,* vol. 41, pp. 1006-9, Jul 1991.

[194] M. Bozzali and A. Cherubini, "Diffusion tensor MRI to investigate dementias: a brief review," *Magnetic Resonance Imaging,* vol. 25, pp. 969-977, 2007.

[195] K. Nishida, Y. Morishima, M. Yoshimura, T. Isotani, S. Irisawa, K. Jann, *et al.*, "EEG microstates associated with salience and frontoparietal networks in frontotemporal dementia, schizophrenia and Alzheimer's disease," *Clin Neurophysiol,* Feb 8 2013.

[196] J. J. Dunkin, A. F. Leuchter, T. F. Newton, and I. A. Cook, "Reduced EEG coherence in dementia: State or trait marker?," *Biological Psychiatry,* vol. 35, pp. 870-879, 1994.

[197] L. A. Coben, W. Danziger, and M. Storandt, "A longitudinal EEG study of mild senile dementia of Alzheimer type: changes at 1 year and at 2.5 years," *Electroencephalography and Clinical Neurophysiology,* vol. 61, pp. 101-112, 1985.

[198] M. Kikuchi, Y. Wada, and Y. Koshino, "Differences in EEG harmonic driving responses to photic stimulation between normal aging and Alzheimer's disease," *Clinical EEG Electroencephalography,* vol. 33, pp. 86-92, 2002.

[199] W. S. Brown, J. T. Marsh, and A. LaRue, "Event-related potentials in psychiatry: differentiating depression and dementia in the elderly," *Bull Los Angeles Neurol Soc,* vol. 47, pp. 91-107, 1982.

[200] J. V. Patterson, H. J. Michalewski, and A. Starr, "Latency variability of the components of auditory event-related potentials to infrequent stimuli in aging, Alzheimer-type dementia, and depression," *Electroencephalogr Clin Neurophysiol,* vol. 71, pp. 450-60, Nov-Dec 1988.

[201] L. E. Holt, A. Raine, G. Pa, L. S. Schneider, V. W. Henderson, and V. E. Pollock, "P300 topography in Alzheimer's disease," *Psychophysiology,* vol. 32, pp. 257-65, May 1995.

[202] J. Polich, "Probability and inter-stimulus interval effects on the P300 from auditory stimuli," *International Journal of Psychophysiology,* vol. 10, pp. 163-170, 1990.

[203] J. Polich and J. Corey-Bloom, "Alzheimer's disease and P300: review and evaluation of task and modality," *Curr Alzheimer Res,* vol. 2, pp. 515-25, Dec 2005.

[204] C. C. Duncan, R. J. Barry, J. F. Connolly, C. Fischer, P. T. Michie, R. Näätänen*, et al.*, "Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400," *Clinical Neurophysiology,* vol. 120, pp. 1883-1908, 2009.

[205] E. J. Golob, R. Irimajiri, and A. Starr, "Auditory cortical activity in amnestic mild cognitive impairment: relationship to subtype and conversion to dementia," *Brain,* vol. 130, pp. 740-752, March 1, 2007 2007.

[206] K. G. Munhall, "Functional imaging during speech production," *Acta Psychol (Amst),* vol. 107, pp. 95-117, Apr 2001.

[207] G. Hickok, "Computational neuroanatomy of speech production," *Nat Rev Neurosci,* vol. 13, pp. 135-145, 2012.

[208] V. Fromkin, *Speech Errors As Linguistic Evidence*: De Gruyter, 1973.

[209] W. J. M. Levelt, in *Speaking: From Intention to Articulation*, ed: MIT Press, 1989.

[210] W. J. M. Levelt, "Producing spoken language: a blueprint of the speaker," in *The Neurocognition of Language* C. M. Brown and P. Hagoort, Eds., ed Oxford, UK: Oxford University Press, 1999, pp. 83-122.

[211] A. Whitworth, J. Webster, and D. Howard, "A Cognitive Neuropscyhological Approach: Theories and Models," in *A Cognitive Neuropsychological Approach to Assessment and Intervention in Aphasia: A Clinician's Guide*, A. Whitworth, J. Webster, and D. Howard, Eds., ed Hove, England: Taylor & Francis, 2005, pp. 3-10.

[212] K. Patterson and C. Shewell, "Speak and Spell: Dissociations and Word-Class Effect," in *The Cognitive Neuropsychology of Language*, M. Coltheart, G. Satori, and R. Job, Eds., ed Hillsdale, NJ, England: Lawrence Erlbaum Associates, 1987, pp. 273-294.

[213] M. Coltheart, "Assumptions and Methods in Cognitive Neuropsychology," in *The Handbook of Cognitive Neuropsychology: What Deficits Reveal about the Human Mind*, B. Rapp, Ed., ed Philadelphia, USA: Psychology Press, Taylor & Francis Group, 2001, pp. 3-21.

[214] C. A. Hauert, "The relationship between motor function and cognition in the developmental perspective," *Ital J Neurol Sci,* vol. Suppl 5, pp. 101-7, Apr 1986.

[215] A. P. Georgopoulos, "Neural aspects of cognitive motor control," *Curr Opin Neurobiol,* vol. 10, pp. 238-41, Apr 2000.

[216] R. D. Kent, *The speech sciences*. San Diego: Singular Pub. Group, 1997.

[217] C. J. Price, "The anatomy of language: contributions from functional neuroimaging," *J Anat,* vol. 197 Pt 3, pp. 335-59, Oct 2000.

[218] S. Dehaene, *Reading in the Brain: The New Science of How We Read*: Penguin Group US, 2009.

[219] V. Mildner, *The Cognitive Neuroscience of Human Communication*: Lawrence Erlbaum, 2008.

[220] I. Rochford, "Speech Correlates of Cognitive Function in Older Speakers," MSc, Department of Electronic & Electrical Engineering, Trinity College Dublin, Dublin, Ireland, 2012.

[221] S. E. Linville, *Vocal Aging*: Singular Thomson Learning, 2001.

[222] R. Vipperla, "Automatic Speech Recognition for ageing voices," Doctor of Philosophy, Institute for Language, Cognition and Computation, School of Informatics, University of Edinburgh, Edinburgh, 2011.

[223] D. A. Mahler, R. A. Rosiello, and J. Loke, "The aging lung," *Clin Geriatr Med,* vol. 2, pp. 215-25, May 1986.

[224] A. Rossi, A. Ganassini, C. Tantucci, and V. Grassi, "Aging and the respiratory system," *Aging (Milano),* vol. 8, pp. 143-61, Jun 1996.

[225] F. R. Bode, J. Dosman, R. R. Martin, H. Ghezzo, and P. T. Macklem, "Age and sex differences in lung elasticity, and in closing capacity in nonsmokers," *J Appl Physiol,* vol. 41, pp. 129-35, Aug 1976.

[226] J. Kahane, "Anatomic and physiologic changes in the aging peripheral speech mechanism," in *Aging: communication process and disorders*, D. Beasley and G. Davis, Eds., ed New York: Grune and Stratton, 1981, pp. 21–45.

[227] L. F. Black and R. E. Hyatt, "Maximal respiratory pressures: normal values and relationship to age and sex," *Am Rev Respir Dis,* vol. 99, pp. 696-702, May 1969.

[228] P. L. Enright, R. A. Kronmal, T. A. Manolio, M. B. Schenker, and R. E. Hyatt, "Respiratory muscle strength in the elderly. Correlates and reference values. Cardiovascular Health Study Research Group," *Am J Respir Crit Care Med,* vol. 149, pp. 430-8, Feb 1994.

[229] K. Tolep, N. Higgins, S. Muza, G. Criner, and S. G. Kelsen, "Comparison of diaphragm strength between healthy adult elderly and young men," *Am J Respir Crit Care Med,* vol. 152, pp. 677-82, Aug 1995.

[230] S. E. Linville, *The Aging Voice*: The ASHA Leader, 2004.

[231] L. O. Ramig, S. Gray, K. Baker, K. Corbin-Lewis, E. Buder, E. Luschei, *et al.*, "The aging voice: a review, treatment data and familial and genetic perspectives," *Folia Phoniatr Logop,* vol. 53, pp. 252-65, Sep-Oct 2001.

[232] P. Lynne-Davies, "Influence of age on the respiratory system," *Geriatrics,* vol. 32, pp. 57-60, Aug 1977.

[233] J. D. Hoit and T. J. Hixon, "Age and Speech Breathing," *Journal of Speech and Hearing Research,* vol. 30, pp. 351-366, September 1, 1987 1987.

[234] J. E. Huber, "Effects of utterance length and vocal loudness on speech breathing in older adults," *Respiratory Physiology & Neurobiology,* vol. 164, pp. 323-330, 2008.

[235] K. K. Gnanalingham, E. J. Byrne, A. Thornton, M. A. Sambrook, and P. Bannister, "Motor and cognitive function in Lewy body dementia: comparison with Alzheimer's and Parkinson's diseases," *J Neurol Neurosurg Psychiatry,* vol. 62, pp. 243-52, Mar 1997.

[236] M. L. Pretterklieber, "Functional Anatomy of the Human Intrinsic Laryngeal Muscles," *European Surgery,* vol. 35, pp. 250-258, 2003/10/01 2003.

[237] F. P. Paulsen and B. N. Tillmann, "Degenerative changes in the human cricoarytenoid joint," *Arch Otolaryngol Head Neck Surg,* vol. 124, pp. 903-6, Aug 1998.

[238] R. A. Dedivitis, M. Abrahao, M. de Jesus Simoes, O. A. Mora, and O. Cervantes, "Cricoarytenoid joint: histological changes during aging," *Sao Paulo Med J,* vol. 119, pp. 89-90, Mar 2001.

[239] M. Hirano, S. Kurita, and S. Sakaguchi, "Ageing of the vibratory tissue of human vocal folds," *Acta Otolaryngol,* vol. 107, pp. 428-33, May-Jun 1989.

[240] K. Sato and M. Hirano, "Age-related changes of elastic fibers in the superficial layer of the lamina propria of vocal folds," *Ann Otol Rhinol Laryngol,* vol. 106, pp. 44-8, Jan 1997.

[241] T. Sato and H. Tauchi, "Age changes in human vocal muscle," *Mech Ageing Dev,* vol. 18, pp. 67-74, Jan 1982.

[242] K. K. Baker, L. O. Ramig, E. S. Luschei, and M. E. Smith, "Thyroarytenoid muscle activity associated with hypophonia in Parkinson disease and aging," *Neurology,* vol. 51, pp. 1592-8, Dec 1998.

[243] M. T. Rodeno, J. M. Sanchez-Fernandez, and J. M. Rivera-Pomar, "Histochemical and morphometrical ageing changes in human vocal cord muscles," *Acta Otolaryngol,* vol. 113, pp. 445-9, May 1993.

[244] A. Bach, F. Lederer, and R. Dinolt, "Senile changes in the laryngeal musculature," *Archives of Otolaryngology,* vol. 34, pp. 47-56, 1941.

[245] B. E. Weinstein, "The Biology of Aging," in *Geriatric Audiology*, B. E. Weinstein, Ed., ed: Thieme, 2000, pp. 15–40.

[246] S. A. Xue and G. J. Hao, "Changes in the human vocal tract due to aging and the acoustic correlates of speech production: a pilot study," *J Speech Lang Hear Res,* vol. 46, pp. 689-701, Jun 2003.

[247] M. Nakayama, "[Histological study on aging changes in the human tongue]," *Nihon Jibiinkoka Gakkai Kaiho,* vol. 94, pp. 541-55, Apr 1991.

[248] H. Koshino, T. Hirai, T. Ishijima, and Y. Ikeda, "Tongue motor skills and masticatory performance in adult dentates, elderly dentates, and complete denture wearers," *J Prosthet Dent,* vol. 77, pp. 147-52, Feb 1997.

[249] H. C. Crow and J. A. Ship, "Tongue strength and endurance in different aged individuals," *J Gerontol A Biol Sci Med Sci,* vol. 51, pp. M247-50, Sep 1996.

[250] T. J. Doherty, A. A. Vandervoort, and W. F. Brown, "Effects of ageing on the motor unit: a brief review," *Can J Appl Physiol,* vol. 18, pp. 331-58, Dec 1993.

[251] H. H. Stassen, M. Albers, J. Püschel, C. Scharfetter, M. Tewesmeier, and B. Woggon, "Speaking behavior and voice sound characteristics associated with negative schizophrenia," *Journal of Psychiatric Research,* vol. 29, pp. 277-296, 1995.

[252] P. Garrard, L. M. Maloney, J. R. Hodges, and K. Patterson, "The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author," *Brain,* vol. 128, pp. 250-260, February 1, 2005 2005.

[253] P. Lieberman, A. Morey, J. Hochstadt, M. Larson, and S. Mather, "Mount Everest: a space analogue for speech monitoring of cognitive deficits and stress," *Aviat Space Environ Med,* vol. 76, pp. B198-207, Jun 2005.

[254] N. C. Andreasen, *The Scale for the Assessment of Negative Symptoms (SANS)*. Iowa City: University of Iowa, 1983.

[255] N. C. Andreasen, *Scale for the Assessment of Positive Symptoms (SAPS)*. Iowa City, IA: University of Iowa, 1984.

[256] E. Chaika, "A linguist looks at "schizophrenic" language," *Brain and Language,* vol. 1, pp. 257-276, 1974.

[257] V. A. Fromkin, "A linguist looks at "A linguist looks at 'schizophrenic language'"," *Brain Lang,* vol. 2, pp. 498-503, Oct 1975.

[258] S. Freud and J. Strachey, *The psychopathology of everyday life*. New York: Norton, 1966.

[259] E. Chaika, *Understanding Psychotic Speech: Beyond Freud and Chomsky*: Thomas, 1990.

[260] J. Chapman, "The Early Symptoms of Schizophrenia," *The British Journal of Psychiatry,* vol. 112, pp. 225-251, March 1, 1966 1966.

[261] K. Schneider, *Clinical psychopathology*: Grune & Stratton, 1959.

[262] C. S. Mellor, "First Rank Symptoms of Schizophrenia," *The British Journal of Psychiatry,* vol. 117, pp. 15-23, July 1, 1970 1970.

[263] A. R. Lecours and M. Vanier-Clement, "Schizophasia and jargonaphasia. A comparative descrition with comments on Chaika's and Fromkin's respective looks at "schizophrenic" language," *Brain Lang,* vol. 3, pp. 516-65, Oct 1976.

[264] G. Pinard and A. R. Lecours, "The language of psychotics and neurotics," in *Aphasiology*, A. R. Ours, F. Lhermitte, and B. Bryans, Eds., ed London: Balliere Tindall, 1983, pp. 313–335.

[265] M. Alpert, A. Kotsaftis, and E. R. Pouget, "At issue: speech fluency and schizophrenic negative signs," *Schizophrenia Bulletin,* vol. 23, pp. 171-7, 1997.

[266] J. Püschel, H. H. Stassen, G. Bomben, C. Scharfetter, and D. Hell, "Speaking behavior and speech sound characteristics in acute schizophrenia," *Journal of Psychiatric Research,* vol. 32, pp. 89-97, 1998.

[267] R. H. Dworkin, S. C. Clark, X. F. Amador, and J. M. Gorman, "Does affective blunting in schizophrenia reflect affective deficit or neuromotor dysfunction?," *Schizophrenia Research,* vol. 20, pp. 301-306, 1996.

[268] M. Alpert, E. R. Pouget, C. Sison, M. Yahia, and E. Allan, "Clinical and acoustic measures of the negative syndrome," *Psychopharmacol Bull,* vol. 31, pp. 321-6, 1995.

[269] M. Alpert, R. J. Shaw, E. R. Pouget, and K. O. Lim, "A comparison of clinical ratings with vocal acoustic measures of flat affect and alogia," *Journal of Psychiatric Research,* vol. 36, pp. 347-353, 2002.

[270] A. Wisniecki, M. Cannizzaro, H. Cohen, and P. Snyder, "Speech Impairments in Neurodegenerative Diseases/Psychiatric Illnesses," in *Encyclopedia of Language & Linguistics*, ed: Elsevier, 2006, pp. 758-764.

[271] J. R. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*. Upper Saddle River, New Jersey 07458: Prentice-Hall, 1993.

[272] D. Ruinskiy and Y. Lavner, "An Effective Algorithm for Automatic Detection and Exact Demarcation of Breath Sounds in Speech and Song Signals," *Audio, Speech, and Language Processing, IEEE Transactions on [see also Speech and Audio Processing, IEEE Transactions on],* vol. 15, pp. 838-850, 2007.

[273] F. G. Eisler, *Psycholinguistics: experiments in spontaneous speech*. London; New York: Academic P., 1968.

[274] R. Duda, P. Hart, and D. Stork, in *Pattern Classif.*, S. Edition, Ed., ed: Wiley-Interscience, 2000.

[275] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," presented at the Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2, Montreal, Quebec, Canada, 1995.

[276] M. H. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine," *Clin. Chem.,* vol. 39, pp. 561-577, 1993.

[277] E. Bleuler, *Dementia Praecox, or the Group of Schizophrenias. (1911) Translated by J. Zinkin*. New York: International University Press, 1950.

[278] V. Peralta and J. C. Manuel, "Factor structure and clinical validity of competing models of positive symptoms in schizophrenia," *Biological Psychiatry,* vol. 44, pp. 107-114, 1998.

[279] M. A. Roy and X. DeVriendt, "[Positive and negative symptoms in schizophrenia: a current overview]," *Can J Psychiatry,* vol. 39, pp. 407-14, Sep 1994.

[280] T. J. Crow, "Brain changes and negative symptoms in schizophrenia," *Psychopathology,* vol. 28, pp. 18-21, 1995.

[281] J. J. Blanchard, K. T. Mueser, and A. S. Bellack, "Anhedonia, Positive and Negative Affect, and Social Functioning in Schizophrenia," *Schizophrenia Bulletin,* vol. 24, pp. 413-424, January 1, 1998 1998.

[282] United States. Public Health Service. Office of the Surgeon General, Center for Mental Health Services, National Institute of Mental Health, *Mental health : a report of the Surgeon General*. Rockville, Md.; Pittsburgh, PA: Dept. of Health and Human Services, U.S. Public Health Service, 1999.

[283] M. Alpert, "Prosody and lexical accuracy in flat affect schizophrenia," *Psychiatry Research,* vol. 97, pp. 107-118, December 27, 2000 2000.

[284] N. M. Docherty, M. E. Strauss, T. J. Dinzeo, and A. St-Hilaire, "The Cognitive Origins of Specific Types of Schizophrenic Speech Disturbances," *American Journal of Psychiatry,* vol. 163, pp. 2111-2118, December 1, 2006 2006.

[285] M. Alpert, A. Rosen, J. Welkowitz, C. Sobin, and J. C. Borod, "Vocal acoustic correlates of flat affect in schizophrenia. Similarity to Parkinson's disease and right hemisphere disease and contrast with depression," *Br J Psychiatry Suppl,* pp. 51-6, May 1989.

[286] J. F. Greden, A. A. Albala, I. A. Smokler, R. Gardner, and B. J. Carroll, "Speech pause time: a marker of psychomotor retardation among endogenous depressives," *Biol Psychiatry,* vol. 16, pp. 851-9, Sep 1981.

[287] A. M. Kring, M. Alpert, J. M. Neale, and P. D. Harvey, "A multimethod, multichannel assessment of affective flattening in schizophrenia," *Psychiatry Research,* vol. 54, pp. 211-222, 1994.

[288] M. Shenton, C. Dickey, M. Frumin, and R. McCarley, "A review of MRI findings in schizophrenia," *Schizophrenia Res.,* vol. 49, pp. 1-52, 2001.

[289] P. J. Harrison and D. A. Lewis, "Neuropathology of Schizophrenia," in *Schizophrenia (Second Edition)*, D. R. W. Steven R. Hirsch, Ed., ed, 2007, pp. 310-325.

[290] D. R. Weinberger and S. Marenco, "Schizophrenia as a Neurodevelopmental Disorder," in *Schizophrenia (Second Edition)*, D. R. W. Steven R. Hirsch, Ed., ed, 2007, pp. 326-348.

[291] B. R. Rund, "Is there a Degenerative Process Going on in the Brain of People with Schizophrenia?," *Front Hum Neurosci,* vol. 3, p. 36, 2009.

[292] B. W. Palmer, R. K. Heaton, J. Kuck, D. Braff, and et al., "Is it possible to be schizophrenic yet neuropsychologically normal?," *Neuropsychology,* vol. 11, pp. 437-446, 1997.

[293] J. L. Rapoport, J. Giedd, S. Kumra, L. Jacobsen, A. Smith, P. Lee, *et al.*, "Childhood-onset schizophrenia. Progressive ventricular change during adolescence," *Arch Gen Psychiatry,* vol. 54, pp. 897-903, Oct 1997.

[294] C. T. Ferrand, *Speech science : an integrated approach to theory and clinical practice*. Boston: Pearson/Allyn and Bacon, 2007.

[295] J. L. Rapoport, "Progressive Cortical Change During Adolescence in Childhood-Onset Schizophrenia: A Longitudinal Magnetic Resonance Imaging Study," *Archives of General Psychiatry,* vol. 56, pp. 649-654, July 1, 1999 1999.

[296] R. L. Spitzer, J. B. Williams, M. Gibbon, and M. B. First, "The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description," *Arch Gen Psychiatry,* vol. 49, pp. 624-9, Aug 1992.

[297] J. E. Overall and D. R. Gorham, "The Brief Psychiatric Rating Scale," *Psychol. Rep.,* vol. 10, pp. 799-812, 1962.

[298] A. J. Saykin, D. L. Shtasel, R. E. Gur, D. B. Kester, L. H. Mozley, P. Stafiniak, *et al.*, "Neuropsychological Deficits in Neuroleptic Naive Patients With First-Episode Schizophrenia," *Arch Gen Psychiatry,* vol. 51, pp. 124-131, February 1, 1994 1994.

[299] H. H. Stassen, "Affective state and voice: the specific properties of overtone distributions," *Methods Inf Med,* vol. 30, pp. 44-52, 1991.

[300] R. J. Moran, R. B. Reilly, P. de Chazal, and P. D. Lacy, "Telephony-Based Voice Pathology Assessment Using Automated Speech Analysis," *IEEE Transactions on Biomedical Engineering,* vol. 53, pp. 468-477, 2006.

[301] R. N. Wormald, R. J. Moran, R. B. Reilly, and P. D. Lacy, "Performance of an automated, remote system to detect vocal fold paralysis," *Annals of Otology, Rhinology and Laryngology,* vol. 117, pp. 834-8, Nov 2008.

[302] J. Ramirez, J. C. Segura, C. Benitez, A. de la Torre, and A. Rubio, "Efficient voice activity detection algorithms using long-term speech information," *Speech Communication,* vol. 42, pp. 271-287, Apr 2004.

[303] R. O'Carroll, "Cognitive impairment in schizophrenia," *Adv Psychiatr Treat,* vol. 6, pp. 161-168, May 1, 2000 2000.

[304] *Technology Research for Independent Living*. Accessed on: 2nd March 2013. Available at: www.trilcentre.org

[305] G. Bravo and R. Hebert, "Age- and education-specific reference values for the Mini-Mental and modified Mini-Mental State Examinations derived from a non-demented elderly population," *Int J Geriatr Psychiatry,* vol. 12, pp. 1008-18, Oct 1997.

[306] M. Cannizzaro, B. Harel, N. Reilly, P. Chappell, and P. J. Snyder, "Voice acoustical measurement of the severity of major depression," *Brain and Cognition,* vol. 56, pp. 30-35, 2004.

[307] L. Berg, "Clinical Dementia Rating (CDR)," *Psychopharmacol Bull,* vol. 24, pp. 637-9, 1988.

[308] D. R. Galasko, R. L. Gould, I. S. Abramson, and D. P. Salmon, "Measuring cognitive change in a cohort of patients with Alzheimer's disease," *Stat Med,* vol. 19, pp. 1421-32, Jun 15-30 2000.

[309] H. Jacqmin-Gadda, C. Fabrigoule, D. Commenges, L. Letenneur, and J.-F. Dartigues, "A cognitive screening battery for dementia in the elderly," *Journal of Clinical Epidemiology,* vol. 53, pp. 980-987, 2000.

[310] R. F. Cohen, G. R. Swanwick, C. Maguire, M. Kirby, B. Lawlor, J. B. Walsh, *et al.*, "Memory impairment in Alzheimer's disease : replication and extension of the delayed word recall (DWR) test," *The Irish Journal of Psychological Medicine,* vol. 13, pp. 55-58, 1996.

[311] S. J. Canning, L. Leach, D. Stuss, L. Ngo, and S. E. Black, "Diagnostic utility of abbreviated fluency measures in Alzheimer disease and vascular dementia," *Neurology,* vol. 62, pp. 556-62, Feb 24 2004.

[312] R. L. Hester, G. J. Kinsella, and B. Ong, "Effect of age on forward and backward span tasks," *J Int Neuropsychol Soc,* vol. 10, pp. 475-81, Jul 2004.

[313] B. J. Small, L. Fratiglioni, M. Viitanen, B. Winblad, and L. Backman, "The Course of Cognitive Impairment in Preclinical Alzheimer Disease: Three- and 6-Year Follow-up of a Population-Based Sample," *Arch Neurol,* vol. 57, pp. 839-844, June 1, 2000 2000.

[314] J. R. Green, D. R. Beukelman, and L. J. Ball, "Algorithmic Estimation of Pauses in Extended Speech Samples of Dysarthric and Typical Speech," *J Med Speech Lang Pathol,* vol. 12, pp. 149-154, Dec 2004.

[315] K. Kirsner, J. Dunn, K. Hird, T. Parkin, and C. Clark, "Time for a pause..." presented at the Ninth Intnernational Speech Science & Technology Conference, Melbourne, Australia, 2002.

[316] A. E. Hieke, S. Kowal, and D. C. O'Connell, "The Trouble with "Articulatory" Pauses," *Language and Speech,* vol. 26, pp. 203-214, July 1, 1983 1983.

[317] L. Castro and J. A. Moraes, "The temporal structure of professional speaking styles in Brazilian Portuguese," in *2nd ISCA Workshop on Experimental Linguistics*, Athens, Greece, 2008, pp. 57 - 60.

[318] S. Kowal, D. C. O'Connell, E. A. O'Brien, and E. T. Bryant, "Temporal Aspects of Reading Aloud and Speaking: Three Experiments," *The American Journal of Psychology,* vol. 88, pp. 549-569, 1975.

[319] K. Rosen, B. Murdoch, J. Folker, A. Vogel, L. Cahill, M. Delatycki, *et al.*, "Automatic method of pause measurement for normal and dysarthric speech," *Clinical Linguistics & Phonetics,* vol. 24, pp. 141-154, 2010.

[320] K. Hird and K. Kirsner, "Objective measurement of fluency in natural language production: A dynamic systems approach," *Journal of Neurolinguistics,* vol. 23, pp. 518-530, 2010.

[321] P. J. Snyder, C. E. Jackson, R. C. Petersen, A. S. Khachaturian, J. Kaye, M. S. Albert, *et al.*, "Assessment of cognition in mild cognitive impairment: A comparative study," *Alzheimer's & Dementia,* vol. 7, pp. 338-355, 2011.

[322] J. J. Manly, N. Schupf, Y. Stern, A. M. Brickman, M.-X. Tang, and R. Mayeux, "Telephone-Based Identification of Mild Cognitive Impairment and Dementia in a Multicultural Cohort," *Arch Neurol,* vol. 68, pp. 607-614, May 1, 2011 2011.

[323] S. E. Cook, M. Marsiske, and K. J. M. McCoy, "The Use of the Modified Telephone Interview for Cognitive Status (TICS-M) in the Detection of Amnestic Mild Cognitive Impairment," *Journal of Geriatric Psychiatry and Neurology,* vol. 22, pp. 103-109, June 1, 2009 2009.

[324] R. S. Wilson, S. E. Leurgans, T. M. Foroud, R. A. Sweet, N. Graff-Radford, R. Mayeux, *et al.*, "Telephone Assessment of Cognitive Function in the Late-Onset Alzheimer's Disease Family Study," *Arch Neurol,* vol. 67, pp. 855-861, July 1, 2010 2010.

[325] J. C. Mundt, K. L. Ferber, M. Rizzo, and J. H. Greist, "Computer-Automated Dementia Screening Using a Touch-Tone Telephone," *Arch Intern Med,* vol. 161, pp. 2481-2487, November 12, 2001 2001.

[326] J. C. Mundt, P. J. Snyder, M. S. Cannizzaro, K. Chappie, and D. S. Geralts, "Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology," *Journal of Neurolinguistics,* vol. 20, pp. 50-64, 2007.

[327] C. Bailey, C. Sheehan, and A. McLean, "Older Adults, ICT and Technology: Users' Perspectives," in *Centre for Excellence in Universal Design Launch Conference*, Dublin, 2007.

[328] M. H. Cohen, J. P. Giangola, and J. Balogh, *Voice User Interface Design*: Pearson Addison Wesley Prof, 2004.

[329] S. R. Rapp, C. Legault, M. A. Espeland, S. M. Resnick, P. E. Hogan, L. H. Coker, *et al.*, "Validation of a Cognitive Assessment Battery Administered over the Telephone," *Journal of the American Geriatrics Society,* vol. 60, pp. 1616-1623, 2012.

[330] P. Östberg, S.-E. Fernaeus, N. Bogdanović, and L.-O. Wahlund, "Word sequence production in cognitive decline: Forward ever, backward never," *Logopedics Phoniatrics Vocology,* vol. 33, pp. 126 - 135, 2008.

[331] R. D. Kent, G. Weismer, J. F. Kent, H. K. Vorperian, and J. R. Duffy, "Acoustic studies of dysarthric speech: Methods, progress, and potential," *Journal of Communication Disorders,* vol. 32, pp. 141-186, 1999.

[332] *Voxeo Europe Limited*. Accessed on: March 7, 2013. Available at: www.voxeo.com

[333] L. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika,* vol. 16, pp. 297-334, 1951.

[334] R. F. DeVellis, *Scale development: Theory and applications* vol. Volume 26. Newbury Park, CA: Sage Publications, Inc., 1991.

[335] *Clinical Psychologist Average Salary*. Accessed on: December 4, 2013. Available at: http://www.payscale.com/research/IE/Job=Clinical_Psychologist/Salary

[336] P. Rabbitt, M. Lunn, D. Wong, and M. Cobain, "Age and Ability Affect Practice Gains in Longitudinal Studies of Cognitive Change," *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences,* vol. 63, pp. P235-P240, July 1, 2008 2008.

[337] E. Ferrer, T. A. Salthouse, W. F. Stewart, and B. S. Schwartz, "Modeling age and retest processes in longitudinal studies of cognitive abilities," *Psychol Aging,* vol. 19, pp. 243-59, Jun 2004.

[338] P. Rabbitt, P. Diggle, F. Holland, and L. McInnes, "Practice and Drop-Out Effects During a 17-Year Longitudinal Study of Cognitive Aging," *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences,* vol. 59, pp. P84-P97, March 1, 2004 2004.

[339] R. S. Wilson, Y. Li, J. L. Bienias, and D. A. Bennett, "Cognitive decline in old age: separating retest effects from the effects of growing older," *Psychol Aging,* vol. 21, pp. 774-89, Dec 2006.

[340] A. Collie, P. Maruff, D. G. Darby, and M. McStephen, "The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals," *J Int Neuropsychol Soc,* vol. 9, pp. 419-28, Mar 2003.

[341] L. J. Beglinger, B. Gaydos, O. Tangphao-Daniels, K. Duff, D. A. Kareken, J. Crawford, *et al.*, "Practice effects and the use of alternate forms in serial neuropsychological testing," *Archives of Clinical Neuropsychology,* vol. 20, pp. 517-529, 2005.

[342] R. H. B. Benedict and D. J. Zgaljardic, "Practice Effects During Repeated Administrations of Memory Tests With and Without Alternate Forms," *Journal of Clinical and Experimental Neuropsychology,* vol. 20, pp. 339-352, 1998/06/01 1998.

[343] *Automated Cognitive Assessment Delivery (ACAD)*. Accessed on: January 25, 2013. Available at: https://acad.tchpc.tcd.ie/

[344] J. R. Stroop, "Studies of interference in serial verbal reactions," *Journal of Experimental Psychology,* vol. 18, pp. 643-662, 1935.

[345] S. Brennan, "Neurocognitive and Electrophysiological Indices of Cognitive Performance in Ageing," PhD, Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland, 2011.

[346] C. J. Davis, "N-watch: a program for deriving neighborhood size and other psycholinguistic statistics," *Behav Res Methods,* vol. 37, pp. 65-70, Feb 2005.

[347] *PHP: Hypertext Preprocessor*. Accessed on: January 25, 2013. Available at: http://php.net/

[348] *ActionScript 3.0*. Accessed on: January 25, 2013. Available at: http://www.adobe.com/devnet/actionscript/articles/actionscript3_overview.html

[349] *Adobe Flash Player*. Accessed on: January 25, 2013. Available at: http://www.adobe.com/products/flashplayer.html

[350] *MySQL*. Accessed on: January 25, 2013. Available at: http://www.mysql.com/

[351] W. C. Barker and E. B. Barker, "SP 800-67 Rev. 1. Recommendation for the Triple Data Encryption Algorithm (TDEA) Block Cipher," National Institute of Standards \& Technology2012.

[352] J. Cohen, "A power primer," *Psychological Bulletin,* vol. 112, pp. 155-159, 1992.

[353] E. Boyle, Jr., A. M. Aparicio, J. Kaye, and M. Acker, "Auditory and visual memory losses in aging populations," *J Am Geriatr Soc,* vol. 23, pp. 284-6, Jun 1975.

[354] R. J. Ivnik, G. E. Smith, J. A. Lucas, R. C. Petersen, B. F. Boeve, E. Kokmen, *et al.*, "Testing normal older people three or four times at 1- to 2-year intervals: defining normal variance," *Neuropsychology,* vol. 13, pp. 121-7, Jan 1999.

[355] S. S. Dikmen, R. K. Heaton, I. Grant, and N. R. Temkin, "Test-retest reliability and practice effects of expanded Halstead-Reitan Neuropsychological Test Battery," *J Int Neuropsychol Soc,* vol. 5, pp. 346-56, 1999.

[356] M. Mitrushina and P. Satz, "Effect of repeated administration of a neuropsychological battery in the elderly," *J Clin Psychol,* vol. 47, pp. 790-801, Nov 1991.

[357] M. E. Theisen, L. J. Rapport, B. N. Axelrod, and D. B. Brines, "Effects of practice in repeated administrations of the Wechsler Memory Scale Revised in normal adults," *Assessment,* vol. 5, pp. 85-92, Mar 1998.

[358] L. Deng, *Dynamic Speech Models: Theory, Algorithms, and Applications*: Morgan & Claypool Publishers, 2006.

[359] M. Tripathi and D. Vibha, "Reversible dementias," *Indian J Psychiatry,* vol. 51 Suppl 1, pp. S52-5, Jan 2009.

[360] S. Kuny and H. H. Stassen, "Speaking behavior and voice sound characteristics in depressive patients during recovery," *Journal of Psychiatric Research,* vol. 27, pp. 289-307, 1993.

[361] H. H. Stassen, S. Kuny, and D. Hell, "The speech analysis approach to determining onset of improvement under antidepressants," *European Neuropsychopharmacology,* vol. 8, pp. 303-310, 1998.

[362] Little, M. A., P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease," *Biomedical Engineering, IEEE Transactions on,* vol. 56, pp. 1015-1022, 2009.

[363] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate Telemonitoring of Parkinson's Disease Progression by Noninvasive Speech Tests," *Biomedical Engineering, IEEE Transactions on,* vol. 57, pp. 884-893, 2010.

[364] T. L. Nwe, S. W. Foo, and L. C. De Silva, "Speech emotion recognition using hidden Markov models," *Speech Communication,* vol. 41, pp. 603-623, 2003.

[365] K. P. Riley, D. A. Snowdon, M. F. Desrosiers, and W. R. Markesbery, "Early life linguistic ability, late life cognitive function, and neuropathology: findings from the Nun Study," *Neurobiology of Aging,* vol. 26, pp. 341-347, 2005.

[366] *The Nun Study, University of Minnesota*. Accessed on: April 24, 2013. Available at: www.healthstudies.umn.edu/nunstudy/faq.jsp

[367] M. L. Zampini and K. P. Green, "The Voicing Contrast in English and Spanish: The Relationship between Perception and Production," in *One Mind, Two Languages: Bilingual Language Processing*. vol. 2, J. Nicol, Ed., ed: John Wiley & Sons, 2001.

[368] L. Y. Ganushchak, I. K. Christoffels, and N. O. Schiller, "The use of electroencephalography in language production research: a review," *Front Psychol,* vol. 2, p. 208, 2011.

[369] A. Costa, K. Strijkers, C. Martin, and G. Thierry, "The time course of word retrieval revealed by event-related brain potentials during overt speech," *Proceedings of the National Academy of Sciences,* November 23, 2009 2009.

[370] C. J. Price, "The anatomy of language: a review of 100 fMRI studies published in 2009," *Ann N Y Acad Sci,* vol. 1191, pp. 62-88, Mar 2010.

[371] V. L. Gracco, P. Tremblay, and B. Pike, "Imaging speech production using fMRI," *NeuroImage,* vol. 26, pp. 294-301, 2005.

[372] T. P. Jung, S. Makeig, C. Humphries, T. W. Lee, M. J. McKeown, V. Iragui, *et al.*, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology,* vol. 37, pp. 163-78, Mar 2000.

[373] K. E. Watkins and J. T. Devlin, "Transcranial Magnetic Stimulation (TMS) as a Tool for Studying Language," in *Handbook of the Neuroscience of Language*, B. Stemmer and H. Whitaker, Eds., 1st ed: Academic Press, 2008.

[374] A. Pascual-Leone, J. R. Gates, and A. Dhuna, "Induction of speech arrest and counting errors with rapid-rate transcranial magnetic stimulation," *Neurology,* vol. 41, pp. 697-702, May 1991.

[375] S. Knecht, A. Floel, B. Drager, C. Breitenstein, J. Sommer, H. Henningsen, *et al.*, "Degree of language lateralization determines susceptibility to unilateral brain lesions," *Nat Neurosci,* vol. 5, pp. 695-9, Jul 2002.

[376] H. Tokimura, Y. Tokimura, A. Oliviero, T. Asakura, and J. C. Rothwell, "Speech-induced changes in corticospinal excitability," *Ann Neurol,* vol. 40, pp. 628-34, Oct 1996.

[377] J. T. Devlin, P. M. Matthews, and M. F. Rushworth, "Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study," *J Cogn Neurosci,* vol. 15, pp. 71-84, Jan 1 2003.

[378] M. A. Naeser, P. I. Martin, M. Nicholas, E. H. Baker, H. Seekins, N. Helm-Estabrooks, *et al.*, "Improved naming after TMS treatments in a chronic, global aphasia patient--case report," *Neurocase,* vol. 11, pp. 182-93, Jun 2005.

[379] *Human Brain Project*. Accessed on: March 23, 2013. Available at: http://www.humanbrainproject.eu/

[380] A. Kluger, J. G. Gianutsos, J. Golomb, S. H. Ferris, A. E. George, E. Franssen, *et al.*, "Patterns of Motor Impairment in Normal Aging, Mild Cognitive Decline, and Early Alzheimer' Disease," *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences,* vol. 52B, pp. P28-P39, January 1, 1997 1997.

Appendix A: Supporting material

A.1 Heidi passage

The thing which attracted her most, however, was the waving and roaring of the three old fir trees on these windy days. She would run away repeatedly from whatever she might be doing, to listen to them, for nothing seemed so strange and wonderful to her as the deep mysterious sound in the tops of the trees. She would stand underneath them and look up, unable to tear herself away, looking and listening while they bowed and swayed and roared as the mighty wind rushed through them.

There was no longer now the warm bright sun that had shone all through the summer, so Heidi went to the cupboard and got out her shoes and stockings and dress, for it was growing colder every day, and when Heidi stood under the fir trees the wind blew through her as if she was a thin little leaf, but still she felt she could not stay indoors when she heard the branches waving outside.

Then it grew very cold, and Peter would come up early in the morning blowing on his fingers to keep them warm. But he soon left off coming, for one night there was a heavy fall of snow and the next morning the whole mountain was covered with it, and not a single little green leaf was to be seen anywhere upon it.

There was no Peter that day, and Heidi stood at the little window looking out in wonderment, for the snow was beginning again, and the thick flakes kept falling till the snow was up to the window, and still they continued to fall, and the snow grew higher, so that at last the window could not be opened, and she and her grandfather were shut up fast within the hut.

Heidi thought this was great fun and run from one window to the other to see what would happen next, and whether the snow was going to cover up the whole hut, so that they would have to light a lamp although it was broad daylight. But things did not get as bad as that, and the next day, the snow having ceased, the grandfather went out and shovelled away the snow round the house, and threw it into such great heaps that they looked like mountains standing at intervals on either side the hut.

And now the windows and door could be opened, and it was well it was so, for as Heidi and her grandfather were sitting one afternoon on their three-legged stools before the fire there came a great thump at the door, followed by several others, and then the door opened.

It was Peter, who had made all that noise knocking the snow off his shoes; he was still white all over with it, for he had had to fight his way through deep snowdrifts, and large lumps of snow that had frozen upon him still clung to his clothes. He had been determined, however, not to be beaten and to climb up to the hut, for it was a week now since he had seen Heidi.

"Good-evening," he said as he came in; then he went and placed himself as near the fire as he could without saying another word, but his whole face was beaming with pleasure at finding himself there. Heidi looked on in astonishment, for Peter was beginning to thaw all over with the warmth, so that he had the appearance of a trickling waterfall.

A.2 Picture Description

This section includes samples of images used in the Picture Description task.

***Figure A-1**: Image 1 of the Picture Description task.*

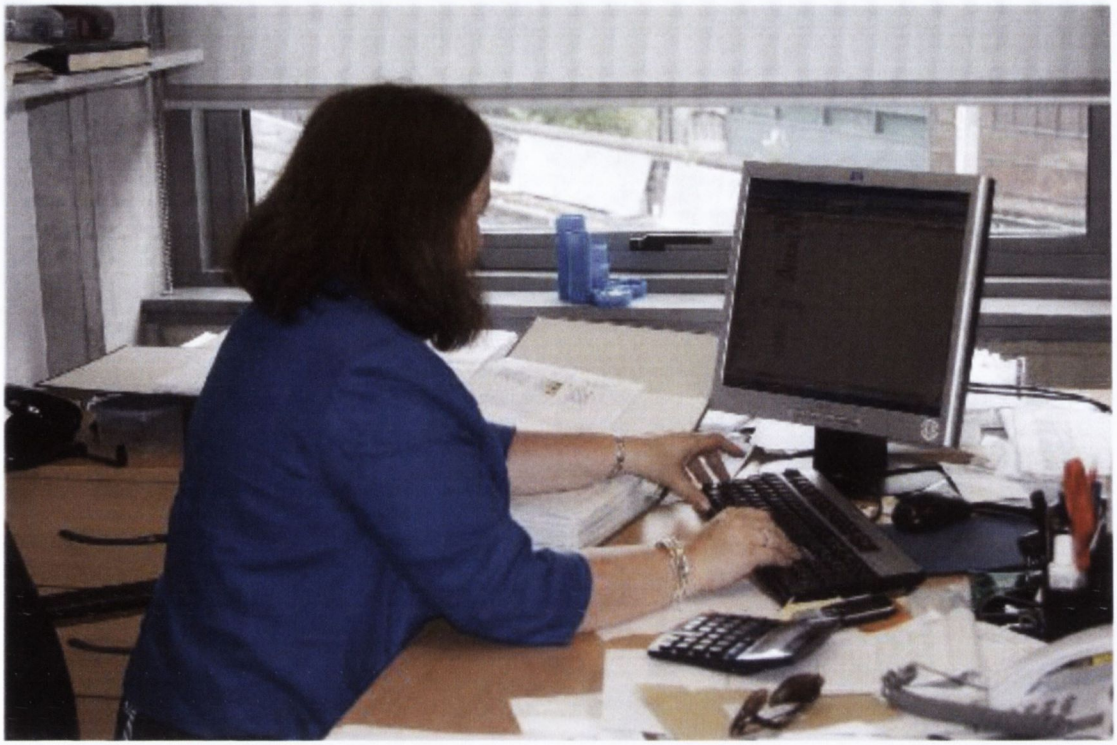


***Figure A-2**: Image 2 of the Picture Description task.*

***Figure A-3**: Image 2 of the Picture Description task.*

A.3 Picture Taboo

This section includes samples of images used in the Picture Taboo task.

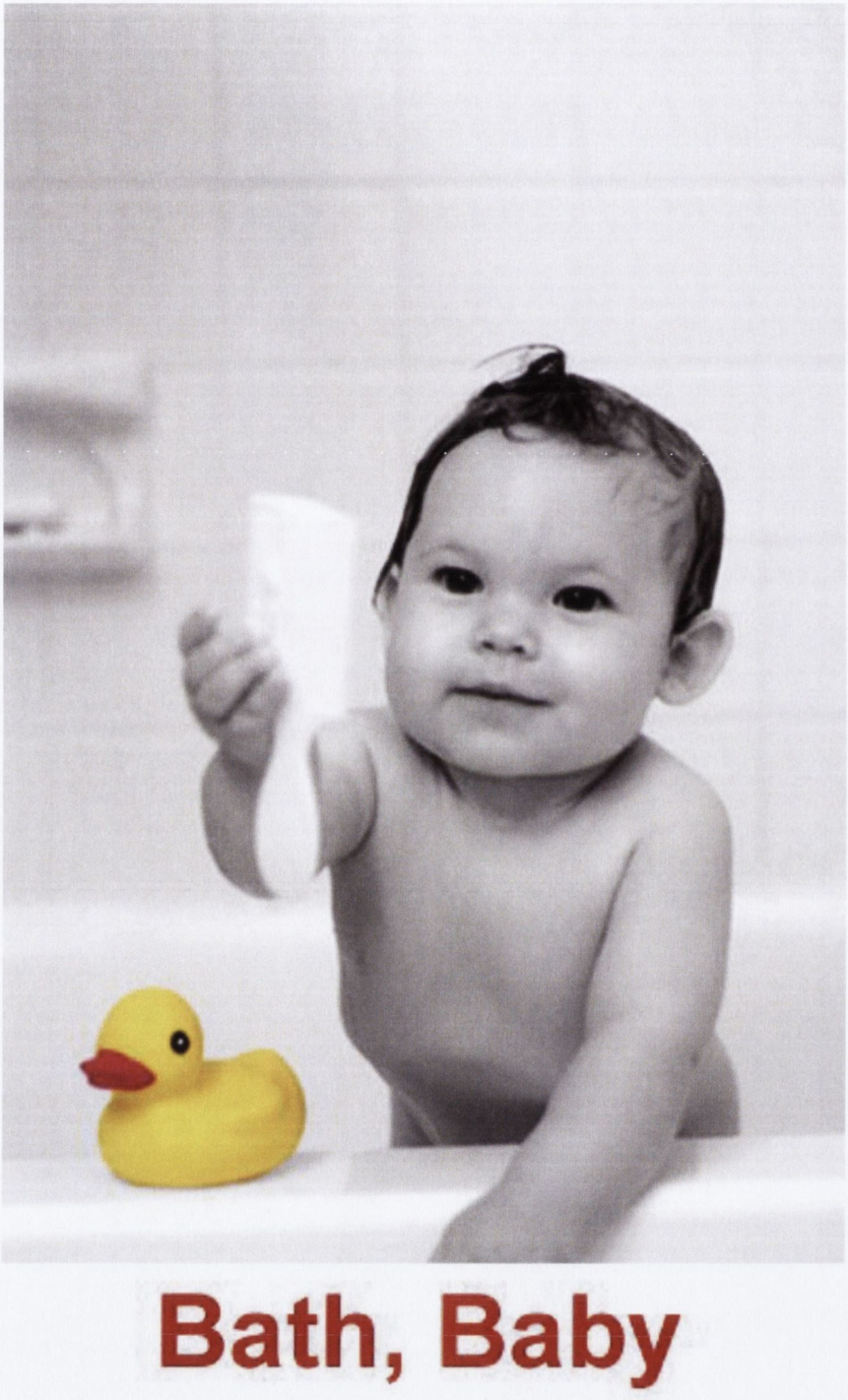


***Figure A-4**: Image 1 of the Picture Taboo task.*

The subject is not allowed to use the two words displayed below the picture during the description of the picture.

Carve, Man

***Figure A-5**: Image 2 of the Picture Taboo task.*

Stage, Play

***Figure A-6**: Image 3 of the Picture Taboo task.*

Proposal, Ring

***Figure A-7**: Image 4 of the Picture Taboo task.*

Cheese

***Figure A-8**: Image 5 of the Picture Taboo task.*