

An Integrated Approach for Episodic Cognition Assessment

Submitted by

Raymond Bruce Wallace, B.E.Sc, M.Eng, P.Eng, SMIEE

A thesis submitted to
the Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Ottawa-Carleton Institute for Electrical and Computer Engineering

Department of Systems and Computer Engineering

Carleton University

Ottawa, Ontario, K1S 5B6

Canada

December 2016

Copyright ©

2016 - Raymond Bruce Wallace, B.E.Sc, M.Eng, P.Eng, SMIEE

Abstract

Dementia and Mild Cognitive Impairment (MCI) are significant health issues that are a rising cost to society. They are monitored with cognitive tests during clinical appointments that are limited by the healthcare system capacity and patient's ability/willingness to attend. Current cognitive tests use behavioural measures and not direct measures of underlying cellular change. This causes delay in identification through a patient's ability to compensate (reminder notes) to mask symptoms.

This work presents measurement methods for cognition between clinical appointments using an integrated approach for episodic cognition assessment. Methods assess the patient during Instrumental Activities of Daily Living (IADL) within an episodic measurement framework.

Electroencephalogram (EEG) / Event Related Potential (ERP) methods are presented as an emerging alternative means to detect changes in the brain. Recent consumer EEG devices make at home use a future possibility. ERP features for healthy and MCI volunteers are defined, analyzed and machine learning identified two features to distinguish the two cases with 1 False Positive and 1 False Negative error in a group of 32 subjects.

The measurement of two IADLs is presented: Computer game play and Driving. Two games were developed and piloted with MCI volunteers showing they could indicate cognitive change. The work presents game design needs including hint and measurement subsystems.

Driving is a complex task that combines executive cognitive tasks (navigation) with over-learned cognitive tasks (turn signal use). The work presents measures of driving behavior creating a driver unique signature. Machine learning techniques show that the features will allow two drivers of a shared vehicle to be distinguished from each other with an error rate as low as 1.5%. Navigational performance measures are presented for driver trip planning to indicate executive function

showing a Google maps derived reference provides best performance. Turn signal use is an over learned action that is measured through detection of turn signal use from dashboard video along with GPS and map methods to determine when signals were required.

The work presents big data analytics and methods to ensure the anonymity of volunteers is preserved through presentation of k-anonymity and differential privacy methods within the data sets.

The measures are combined through the episodic measurement framework for a more holistic view of the patient.

Acknowledgements

This work could never have been possible without the inspiration, support and acceptance of the long hours of work from my family including Thea Foss Wallace, Eric Foss Wallace, Gill Foss, Tara Foss, Aidan Foss and Suzanne Waldman.

I greatly appreciate many people that supported me through this work. My supervisors: Dr. Rafik Goubran and Dr. Frank Knoefel and research collaborators: Dr. Shawn Marshall, Dr. Michelle Porter, Dr. Holly Tuokko, Dr Anita Myers, Dr. Vanesa Taler, Dr. Michael Breau and Dr. Lisa Sweet. Fellow graduate students: Mike Rockwood, Vilas Joshi, Megan Holtzman, Stephanie Bennett, Luke Russell, Xiaochuan He, Andrew Smith, Rocio Lopez-Zunini, Anissa Shaddy, Zhaofen Ren, Anita McKee, Cassandra Morrison, Amanda Baker, Akram Alakel and Nate Fung and undergraduate students: Akshay Puli, Phil Masson, Hannah Hladkowicz, Madeline Harlow, Mihaela Petriu and Alex McAvoy. The team of research assistants and staff: Melissa Donskov, Lynn MacLeay, Jodie Taylor, Courtney Lord, Aaron MacCosham, Briana Allard, Caroline Gaudet, Fiatsogbe Dzuali and Miriam Goubran. The assistance provided by the System's Engineering staff members: Darlene Hebert, Jennifer Poll, Jessie Cartwright, Jenna Noseworthy, Madeleine Ibrahim, Danny Lemay, Narendra Mehta, Daren Russ, Jerry Buburuz and Kong Chiv.

I greatly appreciate the financial support from:

- Natural Sciences and Engineering Research Council (NSERC) and industrial and government partners, through the Healthcare Support through Information Technology Enhancements (hSITE) Strategic Research Network.
- Bruyère Research Institute - Bruyère Academic Medical Organization grant and the Ontario AHSC AFP Innovation Fund.
- Bruyère Research Institute Growth Fund awarded to Dr. Frank Knoefel and the Mitacs Accelerate Program awarded to Dr. Frank Knoefel, Dr. Vanessa

Taler and Dr. Rafik Goubran in partnership with DynamicBrain Inc. and Chartwell Retirement Residences Inc.

The Candrive study was funded by a Team Grant from the Canadian Institutes of Health Research (CIHR) entitled "The CIHR Team in Driving in Older Persons (Candrive II) Research Program" (grant 90429). Additional support was provided by the Ottawa Hospital Research Institute and the Toronto Rehabilitation Institute, University Health Network, and the University of Manitoba. The Candrive Investigators thank the Candrive cohort study participants for their dedication. Without their commitment, this publication would not have been possible. The Candrive Investigators thank Lynn MacLeay, Candrive Program Manager, for her role in managing and operationalizing the study for the Canadian sites. The Candrive Investigators thank the Research Associates for their dedication and contribution to the success of the study; Jennifer Biggs and Anita Jessup (Ottawa Coordinating Centre). The Candrive Investigators wish to thank the following individuals for their work associated with the in-vehicle driving data; Glenys Smith, Andrew Cull, Satoru Nakagawa, Sandra Webber and Maureen Babb (Winnipeg).

The work is in memory for my late wife Gwendolen Foss.

Table of Contents

Abstract	iii
Acknowledgements	v
List of Figures	xi
List of Tables	xviii
List of Appendices	xxii
Glossary	xxiii
Chapter 1: Introduction	1
1.1. Problem	1
1.2. Overview	1
1.3. Collaborators	4
1.4. Statement of Originality	5
1.4.1. Episodic measurement framework	5
1.4.2. EEG/ERP cognition measurement	5
1.4.3. Computer Games cognition measurement	6
1.4.4. Driving cognition measurement	6
1.4.5. Big Data Analytics	8
1.5. Organization	9
Chapter 2: Background Review	11
2.1. MCI and Dementia - the Medical Challenge	11
2.2. Clinical Measurement of Cognition	12
2.2.1. Neuropsychological Testing	12
2.2.2. Brain Bio-Markers	13
2.2.3. EEG and ERP	13
2.3. Activity of Daily Living Measures (ADL)	15
2.3.1. Overview	15
2.3.2. Physical Monitoring	16
2.3.3. Computer Games	16
2.3.4. Driving	18
2.3.5. Driving - Cognitive Adaption and Compensation	18
2.3.6. Driving - Measurement	19
2.3.7. Driving - Driver vs Vehicle Measurement	20
2.3.8. Driving - Signature	21
2.3.9. Driving - Sensors	22
2.4. Signal Processing Techniques	22
2.4.1. Anonymity/Privacy	22
2.4.2. Signal Integrity and Quality	23
2.4.3. Data Set Augmentation	23
2.4.4. Feature Identification and Classification	24
Chapter 3: Experimental Set-up	25
3.1. Gaming and EEG/ERP Study	25

3.1.1.	Student Volunteer Game Pilot	26
3.1.2.	BRI MCI Game and EEG/ERP Study	27
3.2.	Driving Study	29
3.2.1.	Research Team Drivers	29
3.2.2.	Candrive Data Set	30
3.3.	Platforms and Tools	32
Chapter 4:	Episodic Measurement Framework	33
4.1.	Clinical Assessment	33
4.2.	Episodic Framework	36
4.2.1.	Activities of Daily Living - (ADL)	37
4.2.2.	Instrumental Activities of Daily Living - Games	38
4.2.3.	Instrumental Activities of Daily Living - Driving	39
4.3.	Episodic Cognition Assessment Fusion	39
4.4.	Discussion	42
Chapter 5:	Data Preparation and Signal Processing Methods	43
5.1.	EEG/ERP	44
5.2.	Carleton Games	45
5.3.	Driving Studies Candrive Data Set	46
5.3.1.	Sensors and Data Collection	46
5.3.2.	GIS Data Augmentation	47
5.3.3.	Data Integrity and Data Format Processing	48
5.3.4.	Trip Inclusion and Outlier Filtering	50
5.3.5.	Fusion with Additional Data Sources	51
5.3.6.	Privacy and Anonymity Considerations	51
5.3.7.	Alignment to Cloud Computing Programming Models	56
5.4.	Discussion	57
Chapter 6:	EEG as Cognitive Indicator	59
6.1.	EEG and ERP Analysis	59
6.1.1.	Cognitive Task: n-Back	60
6.1.2.	Cognitive Task: Verbal Recognition	61
6.1.3.	Cognitive Task: Go - No-go	61
6.1.4.	HC and MCI Group Comparison	61
6.2.	LDA Classification	63
6.2.1.	EEG Sensor and Test Paradigm Selection	63
6.2.2.	Expanded Feature Identification	64
6.2.3.	LDA Classification	65
6.2.4.	LDA Classification - 1-Back ERP Measures Only	66
6.2.5.	LDA Classification - 1-Back Behavioural Measures Only	68
6.2.6.	LDA Classification - 1-Back ERP and Behavioural Measures	69
6.3.	Alternative Classifiers	70
6.4.	Discussion	77
Chapter 7:	Games as Cognition Measure	79

7.1.	Method	79
7.1.1.	Carleton Word Search Game (CWG)	80
7.1.2.	Carleton Sudoku Game (CSG)	81
7.2.	Experimental Results - Students	82
7.2.1.	Carleton Word Search Game (CWG)	82
7.2.2.	Carleton Sudoku Game (CSG)	84
7.3.	Experimental Results - MCI Patients	85
7.3.1.	Carleton Sudoku Game (CSG)	86
7.3.2.	Carleton Word Game (CWG)	87
7.3.3.	MoCA and RBANS Cognitive Test Scores	88
7.3.4.	Participant during cognitive decline	89
7.4.	Discussion	91
Chapter 8: Driving: Signature Analysis		93
8.1.	Method	94
8.1.1.	Driver Behaviour Attributes - 100 Trip Views	94
8.1.2.	Driver Behaviour Attributes - Month to Month Trends	98
8.1.3.	Sensors signals excluded from analysis	100
8.1.4.	Acceleration and Deceleration Events	101
8.1.5.	Distinguishing Features - t-tests	102
8.2.	Experimental Results	103
8.2.1.	Acceleration and Deceleration Events	103
8.2.2.	Two-Phase Relationship Feature Measures	114
8.2.3.	Distinguishing Features	118
8.3.	LDA Classification	119
8.3.1.	Features and Test Cases	120
8.3.2.	LDA Classification - single feature	120
8.3.3.	LDA Classification - two features	122
8.3.4.	LDA Classification - best performance	127
8.4.	Alternative Classifiers	129
8.4.1.	Driver 7 and 11	130
8.4.2.	Driver 3 and 6	132
8.5.	Discussion	135
Chapter 9: Driving: Navigation		139
9.1.	Method	140
9.2.	Experimental Results	144
9.2.1.	Trip Analysis	145
9.2.2.	Baseline Trip Analysis	150
9.2.3.	Validation Trip Analysis	151
9.3.	Discussion	154
Chapter 10: Driving: Vehicle Operation		155
10.1.	Method	155

10.1.1.	Dashboard Signal Lamp Detection	156
10.1.2.	Turn Detection in GPS Data	157
10.1.3.	GIS Turn by Turn Route Creation	158
10.1.4.	Correlation of GPS and GIS Turns with Dashboard Video	158
10.2.	Experimental Results	159
10.3.	Discussion	162
Chapter 11: Summary and Future Work		165
11.1.	Future Work	165
11.1.1.	Episodic Framework	165
11.1.2.	EEG/ERP	166
11.1.3.	Games	166
11.1.4.	Driving	166
References		167

List of Figures

Figure 2-1: Example ERP trace showing the location of the relevant maxima and minima that occur post stimulus.	14
Figure 3-1: Example image of camera placement to capture dashboard video from turn signals (two camera configuration).....	30
Figure 3-2: Block diagram of the data collection architecture and image of the Persentech OttoView-CD data collection device, [83 , 84].....	31
Figure 4-1: Flow chart of current cognitive assessment process.....	33
Figure 4-2: Episodic cognitive assessment framework showing various sensor sub-systems with the fusion framework and data analytics system..	36
Figure 4-3: Episodic cognitive assessments sensor sub-system details for driving and computer games as examples of two IADL.	38
Figure 4-4: Episodic cognitive assessments framework details for assessment of driving behaviour as one example IADL.....	41
Figure 5-1: Flowchart showing the processing steps required to convert raw vehicle sensor data to a reference repository with Differential Privacy and k-anonymization. Processing steps include the augmentation of the sensor data through fusion with GIS and weather reference data sets.	55
Figure 5-2: Example trip path presented using Intermediate Trust level where the data has been located back to the Ottawa area while continuing to not show the precise origin and destination of the user. Trip starts at the 'o' and ends at the 'x'. The posted limits on the roads traveled are 40km/hr (blue), 60km/hr (green), 80 km/hr (red) and 100km/hr (black). The regions highlighted in orange have road hazard or condition notifications.	56
Figure 6-1: Example ROC showing the area under the curve calculation for LDA (Quadratic) classifier.....	72
Figure 6-2: Example Scatter Plot for LDA (Quadratic). Orange - classified as MCI, Blue - classified as HC. Dot - classified correctly, X - classified incorrectly. Figure identifies the 4 participants from Table 6-5 that were incorrectly classified by one or more classifier.....	73
Figure 6-3: Decision scores for LDA (Quadratic) using best 2 features from Table 6-5 for each of the 32 participants. Blue - "Healthy" score and Yellow - "MCI" score with participant classified into the class with the larger score. 15 MCI subjects on the left and 17 HC subjects on the right.	74
Figure 6-4: Decision scores for Fine KNN using best 2 features from Table 6-5 for each of the 32 participants. Blue - "Healthy" score and Yellow - "MCI" score with participant classified into the class with the larger score. 15 MCI subjects on the left and 17 HC subjects on the right.	75
Figure 6-5: Decision scores for Weighted KNN using best 2 features from Table 6-5 for each of the 32 participants. Blue - "Healthy" score and Yellow -	

	"MCI" score with participant classified into the class with the larger score. 15 MCI subjects on the left and 17 HC subjects on the right. 75
Figure 6-6:	Decision scores for Medium Gaussian SVM using best 2 features from Table 6-5 for each of the 32 participants. Blue - "Healthy" score and Yellow - "MCI" score with participant classified into the class with the positive score. 15 MCI subjects on the left and 17 HC subjects on the right..... 76
Figure 6-7:	Decision scores for Logistic Regression using best 2 features from Table 6-5 for each of the 32 participants. Blue - "Healthy" score and Yellow - "MCI" score with participant classified into the class with the larger score. 15 MCI subjects on the left and 17 HC subjects on the right..... 76
Figure 7-1:	Example CWG game board showing word list on left, letters being selected for a new word in boxes, letters in previously found words in red and letters included within a hint (letter A) for the next word in magenta..... 80
Figure 7-2:	Example CSG game board showing given numbers in grey squares, solution numbers in large font with white background, user pencil marks in small font and a hint for a cell that can be solved highlighted in blue. 81
Figure 7-3:	CWG log trace showing the record of game play. Letters found shows the selection of letters by the user with 1 = letter in hidden word, -1 = letter not in hidden word (error). Letters de-selected shows letters being unclicked. Words found shows the time of each found word. Hints shows when they were provided. 83
Figure 7-4:	CSG results showing all the actions of game play. 85
Figure 7-5:	Left: CWG mouse track for 180 to 200 seconds in Figure 7-3. Right: CSG mouse track for 150 to 250 seconds in Figure 7-4. 85
Figure 7-6:	Mean and standard deviation of Game Completion Time metric for Sudoku game..... 86
Figure 7-7:	Mean and standard deviation of Error Rate metric for Completed Sudoku Games. 87
Figure 7-8:	Mean and standard deviation of Play Time game metric for word search game. 87
Figure 7-9:	Mean and standard deviation of Number of Words Found per Game metric for word search game..... 88
Figure 7-10:	Mean and standard deviation of Number of Words Found per Minute (Word Rate) metric for word search game..... 88
Figure 7-11:	MoCA test scores for participant P1 and P2. 89
Figure 7-12:	RBANS test scores for participant P1 and P2. 89
Figure 7-12:	Play Time game metric for word search game for participant P2 for each of the games played. Games shown in order played over the trial period. 89

Figure 7-13: Number of words found for word search game for participant P2 for each of the games. Games shown in order played over the trial period, no games were completed.	90
Figure 7-14: Game play time metric for Soduko games for participant P2 over the trial period. Games shown in order played over the trial period. Cyan - game not completed, Red - game successfully completed.....	90
Figure 7-15: Number errors made per minute for Sudoku games for participant P2 over the trial period. Games shown in order played over the trial period. Cyan - game not completed, Red - game successfully completed.	91
Figure 8-1: Histogram of trip distances for the 100 trips for each of the participants shown on two different scales. Plot (a) shows the detailed histogram for trip distances for all trips 40km or less while plot (b) includes the longer distance trips with enlarged vertical scale. Driver 1 - dark blue, 2 - light blue, 3 - yellow, 4 - red.	95
Figure 8-2: Histogram of trip durations for the 100 trips for each of the participants shown on two different scales. Plot (a) shows the detailed histogram for trip duration for all trips 40 minutes or less while plot (b) includes the longer duration trips with enlarged vertical scale. Driver 1 - dark blue, 2 - light blue, 3 - yellow, 4 - red.....	96
Figure 8-3: Number of trips taken by three sample drivers each calendar month for 1 year from their entry into the study.	98
Figure 8-4: Distance driven by three sample drivers on a calendar month basis for 1 year from their entry into the study.	99
Figure 8-5: Analysis of the velocity driven by three sample drivers showing calendar month averages of the velocity (km/hr) histogram distributions.	99
Figure 8-6: Analysis of the time of day for three sample drivers for trips driven showing calendar month averages of the time in the solar day (dawn, day, dusk, night).....	100
Figure 8-7: Distribution of deceleration events for one driver compared to the size of the velocity change and the duration of the event (N= 25016 events).	103
Figure 8-8: Monthly velocity change profiles for 1 year. (N= 24794 events in total).	104
Figure 8-9: Distribution of the monthly mean acceleration values for the deceleration events. (N= 24794 events in total).	104
Figure 8-10: Distribution of the monthly minimum acceleration (largest deceleration) values for each of the deceleration events. (N= 24794 events in total).....	105
Figure 8-11: Distribution of the deceleration events for one driver comparing the number of events to both the mean event acceleration and duration. (N= 25016 events).	105
Figure 8-12: Distribution of the monthly minimum acceleration (largest deceleration) values for an example driver for the deceleration events. (N= 40213 events).	106

Figure 8-13: Distribution of the mean acceleration for one driver against the size of the velocity change in the deceleration event. (N= 25016 events).	106
Figure 8-14: Distribution of the minimum acceleration for one driver against the size of the velocity change in the deceleration event. (N= 25016 events).	107
Figure 8-15: Distribution of the mean acceleration for one driver against the size of the velocity change in the deceleration event for driving on roads with speed limits >40 km/hr and <=70 km/hr (N= 16750 events).	107
Figure 8-16: Distribution of the mean acceleration against the size of the velocity change for events where the final velocity was a stop (0 km/hr). (N= 7685 events).	108
Figure 8-17: Comparison of the estimated best fit Gaussian mean for one driver to indicate the ridge lines: Figure 8-13, blue □, all decelerations; Figure 8-15, red ◇, city street; Figure 8-16 - yellow ○, final velocity of 0 km/hr.	108
Figure 8-18: Estimated standard deviation for best fit Gaussian distributions for one driver as measure of ridge width: Figure 8-13, blue □, all decelerations; Figure 8-15, red ◇, city street; Figure 8-16 - yellow ○, final velocity of 0 km/hr.	108
Figure 8-19: Distribution of the mean acceleration for a second example driver against the size of the deceleration for the event. (N= 22469 events).	109
Figure 8-20: Distribution of the minimum acceleration for a second example driver against the size of the deceleration for the event. (N= 22469 events).	110
Figure 8-21: Plots of the ridge lines for one driver for the mean surface in Figure 8-13 (blue) and minimum surface in Figure 8-14 (red) where the estimated peak is shown along with 95% confidence intervals. Best fit two phase relationship with optimal transition between the phases is shown by the lines.	110
Figure 8-22: Correlation coefficient compared to phase transition for the two phase linear interpolation for one driver for the surface ridge lines shown in Figure 8-21. Mean surface (blue) and Minimum surface (orange).	111
Figure 8-23: Plot of the ridge lines for four different drivers showing the two phase linear relationship for each of the drivers while also showing the variations in the curves that will be explored to determine if they allow the differentiation of drivers.	111
Figure 8-24: Distribution of the mean acceleration for one driver against the size of the velocity change in the acceleration event. (N= 25261 events).	112
Figure 8-25: Distribution of the minimum acceleration for one driver against the size of the velocity change in the acceleration event. (N= 25261 events).	113

Figure 8-26: Distribution of the mean acceleration for a second example driver against the size of the acceleration for the event. (N= 23022 events).	113
Figure 8-27: Distribution of the minimum acceleration for a second example driver against the size of the acceleration for the event. (N= 23022 events).	114
Figure 8-28: Plots of the ridge lines for one driver for the mean surface in Figure 8-25a (blue) and maximum surface in Figure 8-26a (red) where the estimated peak is shown along with 95% confidence intervals. Best fit two phase relationship with optimal transition between the phases is shown by the lines.....	114
Figure 8-29: Plot of the ridge lines for four different drivers showing the two phase linear relationship for each of the drivers while also showing the variations in the curves that will be explored to determine if they allow the differentiation of drivers.	115
Figure 8-30: Plot of the impact of the minimum number of decelerations events in each trip on mean correlation to the driver's two phase relationship for four sample drivers.....	117
Figure 8-31: Plot of the impact of the minimum number of decelerations events in each trip on mean correlation to the driver's two phase relationship for four sample drivers.....	117
Figure 8-32: Plot of the impact of the minimum number of accelerations events in each trip on mean correlation to the driver's two phase relationship for four sample drivers.....	117
Figure 8-33: Plot of the impact of the minimum number of accelerations events in each trip on mean correlation to the driver's two phase relationship for four sample drivers.....	118
Figure 8-34: Error performance for quadratic LDA classifiers using 2 features showing the best error provided by two features for each driver pair.	124
Figure 8-35: Error performance for quadratic LDA classifiers using 2 features showing the worst error provided by two features for each driver pair.	124
Figure 8-36: Error performance for quadratic LDA classifiers using 2 features showing the mean error provided by two features for each driver pair.	125
Figure 8-37: Error performance for quadratic LDA classifiers using 2 features showing the best error provided by two features for each feature pair.	126
Figure 8-38: Error performance for quadratic LDA classifiers using 2 features showing the worst error provided by two features for each feature pair.....	126
Figure 8-39: Error performance for quadratic LDA classifiers using 2 features showing the mean error provided by two features for each feature pair.....	127

Figure 8-40: Plot of the best-achieved error performance for the 91 driver pairs when trips have at least 10 acceleration and deceleration events..	128
Figure 8-42: Example Scatter Plot for LDA (Quadratic). Orange - classified as Driver 7, Blue - classified as Driver 11. Dot - classified correctly, X - classified incorrectly.....	131
Figure 8-43: Quadratic LDA classifier scoring for driver A=7 and B=11.	131
Figure 8-44: Medium Gaussian SVM classifier scoring for driver A=7 and B=11.	132
Figure 8-45: Bagged Tree classifier scoring for driver A=7 and B=11.	132
Figure 8-46: Example Scatter Plot for LDA (Quadratic). Orange - classified as Driver 3, Blue - classified as Driver 6 Dot - classified correctly, X - classified incorrectly.	133
Figure 8-47: Quadratic LDA classifier scoring for driver A=3 and B=6.	134
Figure 8-48: Weighted KNN classifier scoring for driver A=3 and B=6.	134
Figure 8-49: Subspace KNN classifier scoring for driver A=3 and B=6.	135
Figure 9-1: Example route trace for a training trip with stops identified in green and trip origin location in red. Trip starts at origin (stop 1) and follows through stops in order 2 through 5 and then returning to the origin.	144
Figure 9-2: Example stop identification for training trip showing the 4 stops with the first being between 2 and 5 minutes, stop 2 and 3 between 5 and 60 minutes and a final stop of greater than 60 minutes. The diagram also shows slow movement and short duration stops related to traffic.	145
Figure 9-3: Stop detection for Validation Trip 1 with direction of travel in clock-wise direction. Trip origin (red circle) and stops (green circle).	149
Figure 9-4: Stop detection for Validation trip 3 with direction of travel in counter clock-wise direction. Trip origin (red circle) and stops (green circle).	149
Figure 9-5: Stop detection for Validation trip 5 with direction of travel in clock-wise direction. Trip origin (red circle) and stops (green circle).	149
Figure 9-6: Trip trace for validation Trip 7 showing travel path and stops. Trip includes an extra stop. Trip origin (red circle) and stops (green circle).	151
Figure 9-7: Trip trace for validation Trip 10 showing travel path and stops. Trip used same home location and initial stop (lower right corner) as trips 1 through 8. It then included 3 additional stops in another area of the city near the home locations.....	153
Figure 10-1: Example right turn lamp image: (a) vehicle 1; (b) vehicle 2.....	159
Figure 10-2: Example of arrow detection algorithm image processing steps.....	160
Figure A-1: Histogram for time of day travel for each of the drivers showing the mean value for each of the 1 hour wide bins for the 100 trips. + signs show one standard deviation (a) - driver 1, (b) - driver 2, (c) - driver 3, (d) - driver 4.	180
Figure A-2: Histogram of velocity habits for the GPS report velocity showing the mean value for each of the 10km/hr hour wide bins for the 100 trips. +	

	shows one standard deviation. (a) - driver 1, (b) - driver 2, (c) - driver 3, (d) - driver 4.	181
Figure A-3:	Histograms of the ratio of velocity to posted speed limit for each of the drivers. Data excludes all samples where the speed limit is unknown. (a) - driver 1, (b) - driver 2, (c) - driver 3, (d) - driver 4.	182
Figure B-1:	Example time and frequency domain acceleration profiles. Blue-solid: GPS, Red-dashed: OBDII.	187
Figure B-2:	Example acceleration profiles for the 3 accelerometers with the mean signal. Blue-solid: Z10A, Red-dashed: Z10B Black-dashdot: Z10C, Green-dot: Mean.....	187
Figure B-3:	Example acceleration signals comparing GPS to OBDII for two difference formulas. Blue-solid: GPS, Red-dashed: OBDII.....	188
Figure B-4:	Example comparison of acceleration signals for the Central 2-point and Forward 3-point formulas showing the resulting effects of noise in the resulting acceleration signal. Blue-solid: Central 2-point, Red-dashed: Forward 3-point.	188

List of Tables

Table 3-1: Information captured by Candrive sensor system. All data captured at a 1Hz sampling rate.	31
Table 3-2: Ottawa Candrive participant demographic information at project entry.	32
Table 3-3: Summary of computing platforms used for the analysis	32
Table 4-1: Clinical and Episodic cognition measurement techniques.	40
Table 5-1: Task accuracy and number of outliers by participants and response type.....	45
Table 5-2: Sensor data captured by Persentech OttoView-CD sensor system....	47
Table 5-3: Example road hazard and road condition information included through the gis augmentation. These messages can occur individually or in combination.....	47
Table 5-4: Summary of the data integrity results for the transformation of the files for 13 Candrive Participants for 4 years of detailed driving data. The participants were chosen based on a sample of convenience.....	48
Table 5-5: Summary of the effects of differential precision for sunrise / sunset calculations for Ottawa On metropolitan area.....	52
Table 5-6: Differential Privacy Models.....	54
Table 5-7: Summary of performance impacts and the ability to parallelize the computation for the analysis of the all the data for Three Candrive participants.....	57
Table 6-1: Summary of behavioural test scores for the HC and MCI groups showing mean score achieved and the SPSS p-value measure for the significance of the statistical difference between the two groups.....	62
Table 6-2: Summary of n-back accuracy and reaction time results for the HC and MCI groups showing mean score (standard deviation) achieved and the SPSS p-value measure for the significance of the statistical difference between the two groups.....	62
Table 6-3: Summary of P2 latency measure for the HC and MCI groups showing mean value (standard deviation) and the SPSS p-value measure for the significance of the statistical difference between the two groups.....	62
Table 6-4: Summary of the performance of the LDA classifiers using an incremental search from best feature for only the ERP analysis features.....	67
Table 6-5: Summary of the performance of the LDA classifiers based an initial best pair search for only the ERP analysis features	68
Table 6-6: Summary of the performance of the LDA classifiers using an incremental search from best feature for only the 1-back behavioural test scoring features.....	69

Table 6-7: Summary of the performance of the LDA classifiers using an incremental search from best feature for the 1-back behavioural scoring and ERP analysis combined.....	69
Table 6-8: Summary of the performance of the LDA classifiers based on initial best pair search for the 1-back behavioural scoring and ERP analysis features combined	70
Table 6-9: Summary of the 23 classifier models	71
Table 6-10: Sensitivity and specificity data for each of classifier using the best pair of features from Table 6-8. For false positive and false negative cases, the incorrectly classified participant number is noted in (). Table also shows the area under the ROC curve for the classifier. . .	72
Table 7-2: CWG results showing measures of game performance calculated from the game log for two example users and the overall group (n=17)...	83
Table 7-3: CSG results showing measures of game performance calculated from the game log for two example users and the overall group (n=17)...	84
Table 8-1: Summary travel distance and time statistics for 100 trips for each of 4 drivers.	94
Table 8-2: Summary for 4 drivers showing percent of time each driver traveled by road choice based on posted limit where posted limit known.	97
Table 8-3: Summary for 4 drivers showing percent of time based on time of day and related traffic levels (Note statutory holidays counted as weekend days)	97
Table 8-4: Summary of the dataset for 14 Triple Stable Candrive Participants Analyzed	98
Table 8-5: Summary of the features calculated for all acceleration and deceleration events over the study period where a 4 km/hr or larger velocity change occurred.	102
Table 8-6: Deceleration correlation coefficients results for the two phase linear relationships for the 14 cognitive and physically stable drivers.....	112
Table 8-7: Summary of the trip level correlations for all drivers between the Deceleration events in each trip and the driver's two-phase relationship showing the effect of the minimum number of events in each trip.	115
Table 8-8: Summary of the trip level correlations for all drivers between the Acceleration events in each trip and the two-phase relationship for that driver showing the effect of the minimum number of events in each trip.	116
Table 8-9: t-test results each feature using all possible pairs of 2 drivers from the set of 14 drivers (91 unique pairs) showing percent of pairs distinguished (p<5%).	119
Table 8-9: Summary of features fore each trip used with classification algorithms.	121
Table 8-10: Single feature LDA classification performance statistics for the set of 91 driver pairs. For each feature, all 91 pairs were classified and the aggregate performance statistics for the feature were calculated for	

Linear and Quadratic LDA models. Columns compare stats for single features to other features. Rows compare aggregate performance for single feature classifiers.	123
Table 8-11: Summary of features for each driver pair used with classification algorithms.	123
Table 8-12: Summary LDA performance for quadratic classifier. Classifier used an adaptive number of features to achieve best possible performance. Effect of number of acceleration and deceleration events in the trips shown.	128
Table 8-13: Comparison of classifier accuracy, sensitivity and specificity along with area under ROC for the best three classifiers to distinguish Driver 7 from 11.	130
Table 8-14: Comparison of classifier accuracy, sensitivity and specificity along with area under ROC for the best three classifiers to distinguish Driver 3 from 6.	133
Table 8-15: Summary of behaviour measures for their performance to differentiate drivers.	136
Table 9-1: Example route optimization analysis for training trip	145
Table 9-2: Example comparison of segment trip information with crow flight and Google	146
Table 9-3: Summary of the driving trips captured.....	147
Table 9-4: Summary of performance of the stop detection algorithm for all stops within the validation data set. No stop detection errors were observed.	148
Table 9-5: Summary of the distance measurements for validation trips. Overall mean and standard deviation is provided and the mean and standard deviation for the difference between the measure and the trip as driven.	150
Table 9-6: Baseline trip summary attributes and features.	150
Table 9-7: Analysis of the 3 baseline trips as compared to the baseline summary trip.....	152
Table 9-8: Summary of the analysis results for the 7 validation trips as compared to the baseline trip formed from trips 1, 2 and 3.	152
Table 10-1: Summary of results for signal arrow detection showing all turn signal events were detected with no false positive or false negative errors.	160
Table 10-2: Summary performance for the GPS and GIS remapping algorithms in the detection of turn and merge events within the trips.	161
Table 10-3: Resulting association of signal lamp use with identified turns.	162
Table 10-4: Turn signal usage rates for the two drivers.	162
Table B-1: Comparison of finite difference formulas for derivative estimate.	184
Table B-2: Comparison of the smart phone acceleration profiles showing the maximum, minimum, mean correlation coefficient and standard deviation of mean. (n=16)	189

Table B-3: GPS and OBDII correlation results for the difference estimates (n=16)	190
Table B-4: Correlation results comparing the GPS signal for the difference formula derivatives with the smart phone acceleration signals. (n=16)	191
Table B-5: Correlation results comparing the OBDII signal for the difference formula derivatives with the smart phone acceleration signals. (n=16)	192

List of Appendices

Appendix A: Additional Driver Behaviour Attributes.....	179
Appendix B: Acceleration Measurement.....	183
B.1. Method.....	183
B.2. Experimental Results.....	186

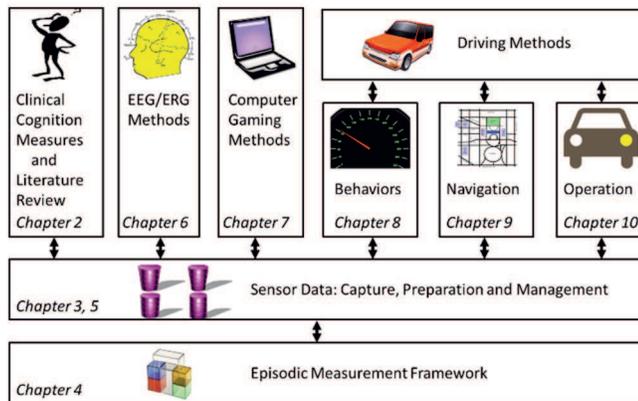
Glossary

ADL	Activities of Daily Living
ADAS	Advanced Driving Assistance Systems
API	Application Programming Interface
BADL	Basic Activities of Daily Living
BMP	Bruyère Memory Program
CSG	Carleton Sudoku Game
CWG	Carleton Word Search Game
CSV	Comma Separated Value
CT	Computed Tomography
DT	Decision Trees
DP	Differential Privacy
EEG	Electroencephalogram
EC	Ensemble Classifiers
ERP	Event Related Potential
FIR	Finite Impulse Response
fMRI	Functional Magnetic Resonance Imaging
IMU	Inertial Measurement Unit
IADL	Instrumental Activities of Daily Living
GIS	Geographic Information System
GPS	Global Positioning System
HC	Healthy Control
KNN	k-Nearest Neighbour
LDA	Linear Discriminant Analysis
LR	Logistic Regression
MRI	Magnetic Resonance Imaging
MCI	Mild Cognitive Impairment
MoCA	Montreal Cognitive Assessment
OBDII	On Board Diagnostics II
RFID	Radio Frequency Identification

RBANS	Repeatable Battery for the Assessment of Neuropsychological Status
ROC	Receiver Operating Characteristic
SD	Secure Digital
SPECT	Single-Photon Emission Computed Tomography
SPSS	Statistical Package for the Social Sciences
SHRP2	Strategic Highway Research Program 2
SVM	Support Vector Machines
URL	Uniform Resource Locator



Chapter 1: Introduction



1.1. Problem

As people age, a small amount of cognitive change can be expected, related to brain aging. In contrast, an increasing number of aging adults are developing functional impairments related to loss of cognition that are clinically described in the early stages of decline as Mild Cognitive Impairment (MCI). Early detection of MCI is essential as early treatment may delay transition to dementia. The measurement of cognition and change in cognition is currently limited to tests performed by clinicians during appointments, which occur infrequently because of their cost, time and inconvenience.

1.2. Overview

Dementia and age related cognitive decline is a significant health issue for individuals and their families and also a rising cost to the healthcare system and society to provide the care needed. More significant decline, described as MCI, may be an intermediate stage of cognition between normal healthy and dementia. A foundation for the diagnosis and treatment of cognitive decline is the ability to detect cognitive changes within patients that are beyond the normal effects of aging. It is most commonly identified through cognitive tests performed by clinicians during annual or biannual medical appointments.



These tests have a number of limitations: the capacity of the healthcare system and patient's ability/willingness to see their doctor limits the frequency that they can occur. Current cognitive testing is limited to measurement of behavioural changes and not direct measures of biological change that is causing decline. The ability of the brain and the patient to compensate (e.g. writing a reminder note) may mask the symptoms and delay identification of issues until long after they have started to occur. This leads to the need for alternative measurement methods for cognition that can occur between clinical appointments and also provide measures prior to significant behavioural symptoms have been observed.

This work presents an integrated approach for episodic cognition assessment for the ongoing measurement of cognition through the measurement of the patient as they go about their daily activities and routines. This work focuses on methods for the ongoing measurement of cognition within an older population. An Episodic Measurement Framework that brings a diverse set of Instrumental Activities of Daily Living (IADL) into a single summary for a patient is presented and then specific measurement methods are presented for the ongoing measurement of cognitive ability.

Electroencephalogram (EEG) / Event Related Potential (ERP) methods are presented as an emerging alternative means to detect cognitive change at the brain electrical activity level. The emergence of consumer EEG devices makes the use of EEG based testing at home a future possibility. The work defines and measures a set of features of the ERP and associated behavioural measure while healthy and MCI volunteers performed a 1-back test. The features were analyzed to identify features both alone and in combination that allowed the correct classification of the participants. A feature pair is identified that provides the best 2 feature performance (1 FP and 1 FN error in a 32 subject group) with a third feature providing additional improvement.

The ongoing measurement of IADLs between clinical appointments is explored through two different activities. Computer use and specifically game play is presented where two games were developed and piloted with a group of MCI volunteers. The work shows the potential for game based measurement of ability



and presents a number of aspects relevant to the choice of games including ensuring games are appropriate and familiar for the older target users (word search and Sudoku), they support the needs of the target users (hint system) and that they include measurement subsystems to capture the details of the game play. The games were pilot tested and the results suggest the games provide measures that could indicate cognitive change.

Driving and the measurement of driving is also presented as this is a complex task that combines many executive cognitive tasks (where am I going and how do I get there) along with many over-learned cognitive tasks (signaling an upcoming turn). The work specifically explores three main aspects of driving:

Signature Analysis where the driving behaviours of drivers are explored to identify features that differentiate between drivers. The work presents a method to measure many features of a driver's behaviour and then uses these features along with machine learning to show that the features will allow two drivers of a shared vehicle to be distinguished from each other with an error rate as low as 1.5%. Behavioural measures have application in emerging self-driving vehicle where owners want their car to drive more like them.

Navigational Analysis where the driver's choice for destination order within a trip and navigational choices between destinations provide indications of executive cognitive function. The work presents methods for the analysis of navigational performance for trips with comparisons against *a posteriori* generated reference measures for a trip. The work shows that a Google maps derived reference provides best performance.

Vehicle Operation, where a driver's over learned actions can be measured and monitored, are explored through the measurement of turn signal use. The work presents methods for the detection of turn signal use from dashboard video along with methods to determine *a posteriori* that signals were required through analysis of trip GPS data and Google map supporting trip turn instructions. The work shows that turn signal use can be measured.



All of the work depends on Big Data Analytics and methods that have been presented to process the vast amounts of information that are available from sensor systems such as the Candrive data set. This set includes 256 older drivers with on average over 18,000km/yr of detailed driving data being available for each participant. The work presents methods required to ensure the anonymity of volunteers is preserved through presentation of a k-anonymity and differential privacy methods.

1.3. Collaborators

My thesis supervisors: Dr. Rafik Goubran through his expertise in data analytics and engineering and Dr. Frank Knoefel through his expertise as a researching medical doctor with a practice focused on dementia and MCI provided significant guidance, ideas and asked lots of tough questions to focus and move the research forward.

Driving Study: Dr. Shawn Marshall through his role of principal investigator on the Candrive project and Dr. Michelle Porter who led the research design for the sensor system and methods for Candrive provided driving data, identified many challenges and provided many guiding insights into the Candrive data. Andrew Smith analyzed the Candrive participant assessments to identify the triple stable drivers used for this study. Dr. Holly Tuokko and Dr Anita Myers as part of the Candrive team, created the driving comfort scale measures. Mike Rockwood's thesis work associated with acceleration measurement through smart phone based sensors enabled the analysis of the acceleration measurement methods for the Candrive data. Akshay Puli, Phil Masson, Hannah Hladkowitz and Madeline Harlow as an undergraduate students and Akram Alahel as a graduate student provided their talents assist in the processing and analysis of the big data sets of Candrive data.

EEG/ERP and Gaming Study: Dr. Vanesa Taler and Rocio Lopez-Zunini provided the leadership and expertise on EEG/ERP measurements and through Dr. Taler's lab enabled the actual measurements for the studies. Dr. Michael Breau and Dr. Lisa Sweet provided their leadership to the MCI Games project as



practising clinicians including the assessment of patients as potential participants in the studies. Melissa Donskov, Courtney Lord, Aaron MacCosham, Cassandra Morrison, Briana Allard and Caroline Gaudet through their roles as research project managers and research assistants supported the MCI game study through recruitment, participant sessions and data analysis. Vilas Joshi and Anissa Shaddy analyzed the captured data and proposed features measured from the Carleton games, while Mihaela Petriu and Alex McAvoy created the Carleton games under my direction. Zhaofen Ren assisted in the processing of ERP data.

1.4. Statement of Originality

The contributions of this research are summarized below:

1.4.1. *Episodic measurement framework*

The ongoing measurement of cognition, through measurement of the performance of IADLs by the patient, creates the need for the fusion of these diverse sources of information into a single view of the patient as they will perform many different IADLs through-out their day and week. This work presents the episodic framework as a model for the fusion of the measures (refereed conferences [1]).

Major Contribution 1: Proposed an episodic measurement framework that allows ongoing measurement of cognition through periodic sensing, measurement and classification of IADLs.

1.4.2. *EEG/ERP cognition measurement*

The Electroencephalogram (EEG) used in conjunction with the Event Related Potential (ERP) technique is studied as it provides as one potential sensor and measurement method for the biological processes in the brain through measurement of the brain's electrical activity. The analysis and classification methods are presented for EEG/ERP measurements that distinguish healthy controls from patients with Mild Cognitive Impairment (MCI) (refereed conferences [2-7]).



Major Contribution 2: Identification of the best 2 features from a set of 314 ERP derived and 4 behavioural measurements that allow HC to be distinguished from MCI subjects.

Minor Contribution 3: Proposed signal processing techniques for noise filtering, outlier detection and removal, and spatial filtering of EEG to achieve ERP signals. The method for outlier detection is a simple technique to remove ocular effects that is generally applicable to the capture of EEG traces. A short test interval and subject supervision ensure no poor sensor contact related affects. Should they occur, the test can be redone.

Minor Contribution 4: Identification of the best 4 classification algorithms from a set of 23 that allow HC to be distinguished from MCI subjects with Fine KNN shown to provide the best performance.

1.4.3. Computer Games cognition measurement

The work presents two computer based games that were developed as a method to measure cognition. The games were specifically developed with the capabilities required for the measurement of cognition and the features to make the games accessible to an older user population. The games were chosen as they are appropriate for the target group of older users (refereed conferences [8, 9]).

Major Contribution 5: Proposed potential performance measures within game performance and identified measures that were indicative of cognitive ability and change.

Major Contribution 6: Proposed generalized feature requirements for the design of games for use within studies of MCI patients. Requirements include the instrumentation of games to capture the necessary research data, a hint system to assist in play and a detailed results recorder for the capture of game performance.

Minor Contribution 7: Proposed specific signal processing techniques for outlier removal, filtering and measurement of game performance.

1.4.4. Driving cognition measurement

Driving is a very complex cognitive task including executive function tasks such as trip and route planning through to highly trained and learned tasks such as turn



signal use. The measurement of driving is presented in three sections with an initial focus of measurement of driving behaviours that are unique to the driver, the measurement of navigational performance as an indication of cognitive ability and the measurement of turn signal use as one measure of vehicle operation (refereed journals [10] and conferences [12-21]).

Major Contribution 8: Proposed two-phase behaviour models for deceleration and acceleration that provide driver unique behavioural features. These features were identified through analysis of large data sets of longitudinal acceleration records and provide a model that can be used for driver differentiation and for personalization of self-driving vehicles so that the drive reflects more the owner's preferences.

Major Contribution 9: Identified features from a set of 162 driving behaviour measurements that allowed each of 91 pairs of drivers to be differentiated from each other. Identified features that provide no/limited value for any pair. Showed that optimal features differ between driver pairs.

Major Contribution 10: Identification of the best classification algorithms from a set of 23 to allow driver pairs to be differentiated. Again showed that algorithm performance varied between driver pairs.

Major Contribution 11: Proposed measurement models for navigational performance, including methods to generate comparative distance measures for trip, identification of optimal comparative driving paths, method to create a baseline reference and comparative method for trip to baseline along with scoring. Identified a generalized method based on Google maps *a posteriori* reference data that provided the best performance that can be applied to road navigational research.

Major Contribution 12: Proposed measurement models for turn signal use and performance, including algorithm and method to detect turn signal use from dashboard lamps, algorithm to detect turns from GPS data, algorithm to re-reference GPS trace to turn-by-turn instructions and pull signal use and method to re-align temporal ordered list to route driven.



- Proposed generalized method for detection of image information captured by video from a display where camera orientation and position cannot ensure co-planarity, constant position or no vibration.
- Proposed generalized method for the identification of longer-term events within a video stream where the long term event consists of a sequence of non-continuous shorter term events. Applied to the specific case of a turn signal being indicated by ongoing lamp flashing on a dashboard.
- Proposed generalized method using *a posteriori* Google maps reference data to find turn information for road navigational research.
- Proposed method to fuse results from turn by turn analysis of GPS data with turn information from Google navigational references to provide a single turn signal reference.

Minor Contribution 13: Identification and analysis techniques for large data sets of driving data for features that identify a driver through their behaviours, including acceleration and deceleration profile curves, velocity models, time of day, duration, distance and road choice.

Minor Contribution 14: Proposed novel methods for the detailed analysis of the measurement of acceleration from GPS signals with the resulting categorization as acceleration and deceleration events. Features of these events were measured allowing events to be contrasted on many different attributes leading to two-phase behaviour models for deceleration and acceleration that allow drivers to be distinguished.

1.4.5. Big Data Analytics

The study of the ongoing performance of IADLs through sensors leads to the capture of longitudinal data for the patient and these datasets can become quite large as they capture the evolution of the patient's ability over months and even years. The work presents contributions that address the needs for the preparation of the large datasets for analysis including quality assurance to ensure data integrity, anonymization techniques to ensure privacy and augmentation of



longitudinal sensor data post capture from reference data sources such as digital map data-bases (refereed conferences [11, 20] and journals [10]).

Major Contribution 15: Proposed generalized method to ensure privacy and anonymity within longitudinal data sets of GPS, GIS and weather information and specifically presented its application within long-term driving data set. Proposed method ensures differential privacy and k-anonymity of the data sets.

Minor Contribution 16: Proposed method for the analysis of long-term longitudinal driving data. The methods include general techniques that can be applied to GPS data sets for outlier detection and removal and associated automated quality analysis of large datasets.

Minor Contribution 17: Proposed implementation methods for the data fusion and augmentation with additional resources such as GIS mapping systems.

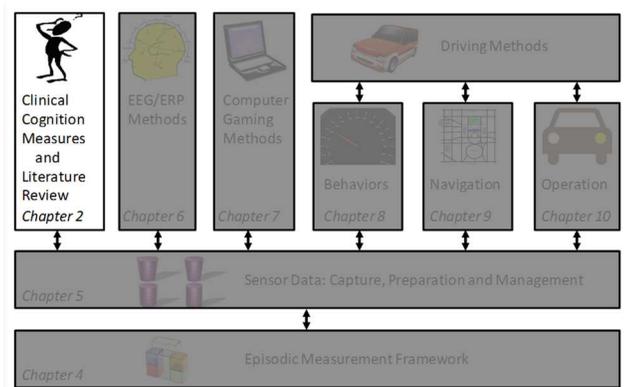
1.5. Organization

After providing an overview on cognition measurement methods within a clinical setting and reviewing the current literature associated with the monitoring and measurement of ADLs and IADLs, this thesis presents the work on cognition measurement through ongoing measurement of patient activities and behaviours. It starts by providing an overall framework for cognition measurement and then explores each of the specific measurement methods proposed, starting with EEG/ERP techniques and then computer game based techniques. Driving is presented through three distinct measurement models that include behaviours that allow drivers to be distinguished followed by measurement of driver navigation and vehicle operation.





Chapter 2: Background Review



2.1. MCI and Dementia - the Medical Challenge

As people age, a small amount of cognitive change can be expected, related to brain aging. In contrast, an increasing number of aging adults are developing excessive impairments related to loss of cognition that are clinically described in the early stages of decline as Mild Cognitive Impairment (MCI) which is an intermediate stage of cognition between normal healthy that may transition to dementia with its associated functional impairments [22]. Eventually, 10% of the patients identified with MCI transition to dementia annually. The most frequent cause for dementia is Alzheimer's Disease. The World Health Organization estimates that worldwide, 47.5 million people are living with dementia and by 2030 it is projected to be 75.6 million [23]. In Canada, the number of adults with dementia is expected to more than double from 250,000 (1994) to 592,000 (2021) [24]. Dementia though is a significant health risk for individuals as it affects all aspects of their quality of life, leads to costly burdens on the patient's family and/or healthcare system to provide care and supports that the patients require to live safely.

It has been shown [25,26,27] that early intervention (e.g. increase cognitive activity through social interaction and activities) is important to achieve optimal outcome for dementia patients. This leads to the need to identify and intervene during the MCI phase but standards for diagnosis of MCI [25, 26] remain a



challenge. Most dementia clinics perform a base-line series of blood tests and brain scans, and repeat neuropsychological testing from 6 months to annually for persons with MCI to determine if there has been progression.

2.2. Clinical Measurement of Cognition

2.2.1. Neuropsychological Testing

The measurement of cognitive ability within a clinical setting is typically done through the application of standardized clinical tests such as Mini-Mental State Examination [28], Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) [29], Trail Making Test A & B [30], and Montreal Cognitive Assessment (MoCA) [31]. Each of these tests is performed by a trained practitioner and test the patient's memory, language and other cognitive functions helping lead to diagnosis.

The tests have different but compatible diagnostic objectives. The MoCA test is a 10-minute test that can be used by primary care physicians to help detect MCI. It includes a number of simple subtests such as visuospatial (drawing an analog clock with a specific time), naming (animal identification), memory (word lists), attention (letter identification), language (creating a list of words), abstraction (word similarity) and orientation (date and location). The MMSE is a 5 to 10-minute test using a questionnaire that includes registration, calculation, attention, language, recall, ability to follow commands and orientation. The Trail Making Test A & B consists of two tests where the subject is asked to connect a set of labeled dots in numerical order (A test) and in a 1-A-2-B-3 etc. order (B test), This provides a measure of visual search speed, scanning, processing speed and set-shifting ability. The RBANS test is a more detailed test of cognition that takes approximately 30 minutes to complete. The test provides an overall score for the subject that can be compared to published norms and in addition provides sub scores in five specific areas: Immediate Memory, Visuospatial, Language, Attention and Delayed Memory. Another key advantage of the design of RBANS test is that alternate versions allow for repetition of the test with minimum practice effect. Assuming that clinicians have finite amounts of time for work, the challenge is choosing between



a long and sensitive test performed frequently on a small number of patients vs. a short less sensitive test performed less frequently and treating a larger number of patients.

MCI is a clinical diagnosis that benefits from formal cognitive testing using neuropsychological tests, however it is known that these results are impacted by language skills, education and stress [32]. The frequency of these tests is also limited by the ability for the patient to attend a clinical appointments and for the practitioner to be able to perform the test. In addition, specific clinical tests may not be great at predicting performance for complex real-life tasks [33] such as driving. There continues to be much debate in the literature about which cognitive tests are best at determining fitness to drive. Hence, work continues on searching for alternative measurement methods. Ideally, the cognitive test would be low cost, repeatable and optionally performed outside of the clinic.

2.2.2. Brain Bio-Markers

As mentioned earlier, numerous tests are performed at Memory Clinics. These include relatively low cost and non-invasive blood tests, medium cost and more invasive lumbar spinal punctures, and very high cost brain scans (computed tomography (CT), magnetic resonance imaging (MRI) and single-photon emission computed tomography (SPECT)). None of these tests by themselves or in combination are currently able to definitively diagnose MCI or dementia. Hence, there is interest in finding a relatively cheap test that can be repeated over time that could help.

2.2.3. EEG and ERP

The Electroencephalogram (EEG) provides a measure of brain electrical activity [34] through the application of a series of electrodes on the patient's head. This technique provides excellent temporal resolution, on the order of milliseconds. A key challenge with the measurement of the EEG is that each sensor provides a measure of the aggregate electrical activity of the brain, and given the large amount of brain activity, the identification of a specific response to a stimulus can be difficult to distinguish from the background brain activity.



The Event Related Potential (ERP) [34] technique has emerged as a method to measure cortical response to a stimulus through repeated stimulus presentation and subsequent averaging of the resulting responses for each electrode by time aligning the responses relative to the stimulus onset. The averaging has the effect of removing noise and signals due to activity that is uncorrelated with the stimulus, allowing the activity that is correlated to the stimulus to be observed.

An example ERP response curve is shown in Figure 2-1 and it is typically analyzed [35] based on the location (latency) and amplitude of the peaks and valleys within the resulting waveform. The naming convention for the peaks and valleys is N (negative peak) and P (positive peak) along with a number that is either 100, 200, 300 etc. which typically occur at 100 msec, 200 msec etc. after the stimulus. In this work, given that patients with MCI often have delayed latencies so that the P300 peak may actually appear 400 msec after the stimulus, the shortened form of P100 (P1), and N100 (N1) etc. will be used in this work.

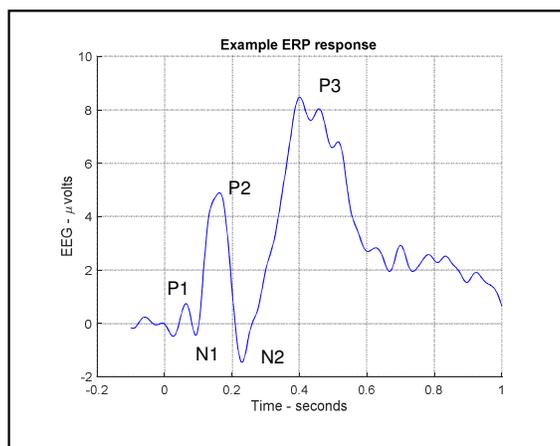


Figure 2-1: Example ERP trace showing the location of the relevant maxima and minima that occur post stimulus.

The relationships between ERP measures and cognitive abilities such as Memory Encoding and Retrieval [36] and Attention [37] have been previously shown. The P300 signal is believed to be an indicator for fixation on a target and has been shown [43] to be sensitive to changes related to aging and cognitive change [44].

This work extends previous ERP studies through the application classification techniques and the 1-back test to identify ERP features that distinguish MCI



patients from healthy controls. The 1-back test [38, 39] is an example of a test that can be used with the ERP technique where the subject is presented with a sequence of symbols and asked to press a button if the symbol presented is the same as the previous symbol and to do nothing if the symbol is different. This provides two different measures of the brain response that can be captured through an ERP. First, when the button should be pushed, the subject has had to make the decision to press the button and then signal their hand to do the physical act of pushing the button – hence two interconnected cognitive processes. On the other hand, when a button should not be pushed, the subject needs to make the decision to not press a button and therefore there is no message to the hand – hence different cognitive tasks.

In the typical ERP response shown in Figure 2.1 for a subject that receives the stimulus event at time 0 with a series of subsequent maxima and minima. The data collected prior to the stimulus establishes a zero baseline while outlier removal (such as eye blinks) is required for individual traces. The calculation of an ERP [34, 40] stimulus aligns repeated traces, removes outliers and the resulting EEG set are averaged reducing uncorrelated signal content with subsequent low-pass filtering for additional noise reduction [41, 42].

2.3. Activity of Daily Living Measures (ADL)

2.3.1. Overview

The differentiation of MCI and dementia depends on the intactness or impairedness of Instrumental Activities of Daily Living (IADL). Clinicians have to rely on self-reporting or family reporting on the patient's abilities regarding IADLs for diagnosis. Basic ADLs (BADL) include basic hygiene activities (e.g. dressing, mobility, bathing, self-feeding etc) [45] while IADLs involve more complex tasks (e.g. food preparation, shopping, driving, social activities, using a computer) and these can provide a direct indication of an individual's cognitive abilities and also indicate changes if they are no longer or significantly less capable of performing tasks.



Patients that maintain a high level of social engagement and activities are shown [46] to have slower progression of cognitive decline and hence interventions that maintain (or increase) social activity are important. Loss of mobility such as ability to drive can have significant detrimental effects on long term prognosis and outcomes [47 , 48].

Solutions to measure a patient's cognitive ability between appointments through sensor-based observation are an alternative to periodic clinical appointment cognition tests. Ongoing measurement of the performance of IADLs provides the opportunity for episodic data to monitor for changes and trends.

2.3.2. Physical Monitoring

The monitoring of physical well being has been explored extensively through the applications of various sensor solutions to monitor physical activity levels [49, 50 , 51] including monitoring the physical well-being of subjects in their home and activity levels for activities such as getting out of bed. Kitchen use and sensor systems specifically focused on safety of the patient through the use of smart apartments [52]. Physical activity is important for a patient's well-being but does not get into the quality of performance of the activities such as efficiency, speed, effectiveness and at best provides an indirect indication of cognitive ability.

2.3.3. Computer Games

Methods to measure cognitive performance at home have been proposed [53 , 54 , 55] through instrumentation of computer interactions such as games or web usage. Researchers have developed models of brain plasticity that refers to the brain's lifelong capacity for physical and functional change; it is this capacity that enables experience to induce learning throughout life [56 , 57]. Computer based gaming applications provide the potential to maintain or build cognitive ability and include a number of programs that allow the user to improve memory, speed of processing and reasoning. A number of large randomized control trials have shown that healthy older adults are able to improve their cognitive functioning using programs [57]. This thesis focuses on MCI patients as there have been very few studies on computer use for brain training in older adults with MCI. Studies have



primarily focused on healthy elderly adults tracking browser and internet use along with overall performance (time to complete) for a number of computer games including word jumbles and memory games [53 , 54 , 55].

This thesis provides pilot study results for games such as Sudoku and word search as there have been no studies designed to show that they help brain performance even though they are very appropriate for the target population; in fact, these types of games are sometimes used in the control group. On the other hand [58], a recent Cochrane review found that in some trials, members of these control groups showed similar improvement to the treatment groups and results for older adults have shown some benefit in working memory [59] through computer based perceptual discrimination training through commercial internet cognition training game from Posit Science [60].

Computer activity monitoring [61] has been used to discover information about people with early dementia as computer based games require the user to apply many different aspects of brain function, such as learning the game, measuring of speed and accuracy, and finally motor skills to move the mouse around or for entry of information along with [53 , 54 , 55]. Games based on numerical or word based foundations use different brain regions along with areas for search or logical deduction. Macro performance (time to complete, success) for games provides an indication of the ability of the user, and measured trends provide an indication of changing cognitive ability. This thesis extends these through the measurement of more detailed user interactions with a game including player speed, errors the user makes and associated correction or non-correction, need for hints to reach a solution, solution strategies used and mouse usage.

A key challenge in the use of gaming for cognitive decline measurement or therapy is to design the game to measure/monitor the user capability. One group proposed the use of EEG while playing Sudoku [52] to measure cognitive function. PC Keyboard [63] use has also been proposed, using the login process as a cognitive test and the actual measurement of typing as a motor skill test. Both of these could be monitored over time. Analysis of mouse movement during web browsing [53] has also been presented. However, it has not been used to measure



game playing ability. Mouse use has also been combined with keyboard use during [64, 65, 66] Solitaire and word jumble games.

2.3.4. Driving

In addition to computer use and home based activities of daily living, automobile driving is an activity that requires high levels of cognitive function. Driving requires navigation, trip planning, and safe vehicle operation including signaling turns, checking blind spots, observing other vehicles, obstacles, road signs, pedestrians etc. Driving as an everyday competence has been reported [81]. The ability to drive has also been reported as an enabler for social engagement and activities that have the effect of slowing cognitive decline [67, 68]. In fact poorer health [69] has been shown for older adults that stop driving. Medical professionals face challenges currently, as many jurisdictions require that they determine driving risk. Current best practice for this assessment includes clinical and driving history, physical examination, and cognitive tests. While on-road driving tests are considered the gold standard for driving risk assessment, they are difficult to arrange and costly, and only truly reflect driving at that point in time under those conditions.

2.3.5. Driving - Cognitive Adaption and Compensation

Reviews of research on older drivers [70, 71] have highlighted a number of issues related to the variable effects of aging and age-related cognitive decline on driving ability. Self-regulation and avoidance of driving after dark or driving on expressways [72, 73] are reported as coping mechanisms for reduced ability or confidence but there are issues with driver compliance. Surveys of driving habits for subjects with cognitive decline [74, 75] showed that drivers can reduce the variety of destinations they choose, reduce the driving distance by no longer going as far from home and using closer destinations or they reduce the complexity of trips by making fewer stops on a given trip. A tendency to choose familiar routes instead of perhaps a more optimal route was shown [76, 77]. Compensation may include returning home [78] between destinations allowing navigation decisions to



always start from the home reference. Studies have also explored older drivers and their increased crash rate [79 , 80].

Driving requires significant spatial orientation to perform and hence declines in cognitive function will affect the ability to perform the task of driving. This work explores the potential for the repetitive nature of trips (same grocery store, friends, social clubs or family) to lead trips being measured and compared over time to identify changes as an indication of cognitive change. Virtual reality mazes [82] have been used to measure spatial orientation as a measure for cognitive performance.

2.3.6. *Driving - Measurement*

The measurement of driving, allows real world data to be collected while drivers go about their daily activities. It has advantages over alternatives such as requiring the subject to drive a vehicle other than their own, or to operate a simulator. It allows the drivers to drive in familiar neighborhoods instead of simulated or scripted/closed driving courses that might prevent the testing from being an accurate representation of actual driving ability.

The measurement of executive cognitive function for navigation [83, 84, 70, 71] has been presented for studies of older drivers where the vehicles have been outfitted with sensor technology providing insight into executive cognitive decisions such as trip frequency, duration, distance and time of day. The work presents methods for the automation of driver identification within longitudinal driving studies as a new area of research that requires the application of measurement techniques to identify distinguishing features. Previous driving studies [70] and the Strategic Highway Research Program 2 (SHRP2) [85, 86], all use in cabin video that includes the driver's face to allow manual driver identification.

The Candrive study [83, 84] is a study of older drivers to create and evaluate the predictive validity of tools for assessing fitness to drive. At entry into the study, drivers that indicated they shared their vehicle were identified and Radio Frequency Identification (RFID) key tags were provided for the drivers to carry on their key chains to identify the participant driver. Drivers with infrequent alternate drivers



were not provided RFID tags and instead were asked to keep track in a log when the vehicle was driven by someone else. The resulting collected data are then sorted based on the RFID tags or driver logs. In both cases there is the potential for reporting error as a driver may have left the RFID tag at home or loaned their keys to another driver resulting in an RFID reporting error. The manual logs can also contain inaccuracies as the driver could have forgotten to record an alternate driver.

The long term and continuous measurement of sensor information using participant vehicles introduces design challenges for the sensor system. The GPS and OBDII sensors do not provide a direct measurement of acceleration, but the addition of an accelerometer would add deployment complexity and cost due to the additional sensor and increased storage requirements for the data generated. Additionally maintaining accelerometer orientation [90] will be difficult in long-term deployments as will issues of long-term drift and calibration [91]. The performance of GPS systems [92, 93] can be augmented through the use of additional sensors such as accelerometers. Some recent applications of GPS sensor information include the use of Kalman filtering of GPS information to predict the future position of a vehicle based on its position and velocity [94] and to support GPS navigation with Inertial Measurement Unit (IMU) sensor information for ship navigation [95]. This work presents methods for the use of GPS data for acceleration measurement avoiding the need for this additional sensor.

Turn signal use is one of many cognitive functions associated with driving as drivers perform (or skip) based on their automatic actions through training and experience. Eye focus areas, another automatic/trained action, can be analyzed to ensure the driver is focusing on the forward direction of travel, dashboard, mirrors and blind-spots for the appropriate amount of time and at the appropriate times.

2.3.7. *Driving - Driver vs Vehicle Measurement*

The study of older driver's habits and ability is an emerging area of research as concerns related to the effects of normal aging on driving ability can become more significant if the driver has any cognitive or physical decline. Drivers have been



studied through simulator use [87] and using in-vehicle trained observers [88]. These observation methods however influence actual driving habits, as they introduce aspects that are unfamiliar and/or different for the driver.

There is the potential to also collect data from drivers other than the participant when collecting driving data through on-board vehicle sensor systems, as most personal vehicles are also driven by friends or family members. Driver logs [83, 70, 71], where the participant maintains a record of the drivers of the vehicle, have been used and the accuracy of the log is highly dependent on participants remembering and accurately maintaining them. The use of RFID tags has also been proposed as this allows for easy post sorting of the data, since the RFID information is captured with the data, but still has potential for errors due to lost or loaned keys.

Facial recognition or other biometric information has also been proposed. These are challenged by the need to include the biometric sensors such as a camera, the need for it to be placed safely within the vehicle, the accuracy of the automated processing of the result, and the ethics of filming drivers.

2.3.8. *Driving - Signature*

This work presents methods to differentiate drivers through a driving signature, which is a set of attributes that allows different drivers of a vehicle to be distinguished from each other as alternative to RFID tags which depend on drivers not sharing keys. Characteristic behaviours of drivers [96] were studied where vision tracking was used to measure driver intentionality for associated driving actions as a contribution towards Advanced Driving Assistance Systems (ADAS). Instrumentation, such as neuromuscular admittance, has been proposed to measure racecar driver actions as they relate to associated driving behaviour models [97, 98]. Driving event classification models [99, 100] and driver style classification [101] using inertial sensors has been presented. Inertial sensors [102] has been proposed to identify and classify driving events and explored their application for driver identification. The need and potential have been presented for characteristic behaviours of individual drivers to be applied to automated driving



systems to make them more natural and personalized leading to drivers being more accepting of these systems [101, 103].

2.3.9. Driving - Sensors

GPS technologies [104, 105, 106] have been presented as an effective sensor solution for location tracking within an urban environment demonstrating the effectiveness of these devices for the measurement of location while proposing solutions to common issues such as measurement errors and loss of GPS signal.

The emergence of low cost and easy to deploy sensors and large data storage devices enables the study of participants over long periods. Some research has shown the applications of GPS and inertial measurement sensors [110, 111, 112] for the tracking of vehicle position with applications to fuel consumption measurement and insurance risk related driving behaviours. These studies also explore the identification and classification of specific driving behaviours and the issues related to GPS outlier errors.

2.4. Signal Processing Techniques

The analysis of sensor data to measure cognitive ability requires many signal processing techniques for the various types of analysis that must be performed and also for the proper management of the data.

2.4.1. Anonymity/Privacy

This work presents solutions for the management of privacy of the participants within captured sensor data is a significant challenge and a priority as this is an absolute requirement for the data due to research ethics guidelines and privacy laws. In particular it is necessary for large databases to be potentially widely shared with researchers and the privacy considerations [113] required for the securing of large cloud storage.

Privacy needs can range from personal use only applications such as a lifestyle tracking applications [114] through to aging in place studies [115], where sensor data has been centralized. In this case, k-anonymity [116, 117] is required, such that participants cannot be identified through information from the combination of k



elements from the data stream. Differential privacy (DP) [118, 119] in the anonymization of data sets allows the trust level for a given user of the data to lead to different levels of precision and privacy within the data.

This work presents specific solutions for the privacy of GPS related data from driving studies while a participant's name is easily removed from data, identifying details such as places of work or home can still be easily determined within the location data.

2.4.2. *Signal Integrity and Quality*

The long term and continuous measurement of sensor information using participant vehicles introduces design challenges for the sensor system. There are a number of challenges with the capture of large amounts of sensor data over long periods of time. Issues of data validity [120] as the sensor log files may contain errors, such as incomplete sample records caused by random operational events. In addition, there may also be failure of sensor operation and resulting missing or imprecise data.

The work presents a solution for the automated formatting and preprocessing of the raw sensor data, such that they are available in formats that are easily processed within chosen analytical tools and cloud computer architectures.

2.4.3. *Data Set Augmentation*

The use of data fusion techniques within a big data framework can also allow for the extension of the sensor data through augmentation of information from other sources and databases. In the case of vehicle operation, other data information, such as Geographic Information System (GIS) information can be included. GIS can provide information such as speed limits, road hazards or school zones. Weather information from local weather reporting stations [121] can be added. Traffic conditions including congestion, construction and accidents can be added providing access to an appropriate source. Sunrise/sunset for night/day driving indication [72, 121] can also be added through use of the time of day and longitude/latitude of the vehicle. This work presents methods for this augmentation while still considering the issues related to anonymity of the participant.



2.4.4. Feature Identification and Classification

Linear Discriminant Analysis (LDA) is a generalization of the Fisher technique [122] for pattern recognition and machine learning. Adaptive LDA training algorithms allow systems with large numbers of potential features to be analyzed and the training algorithms can search subsets of the feature space to identify the features that provide the best performance while excluding features that provide no or limited classification value. The classifiers may use either linear or quadratic decision models.

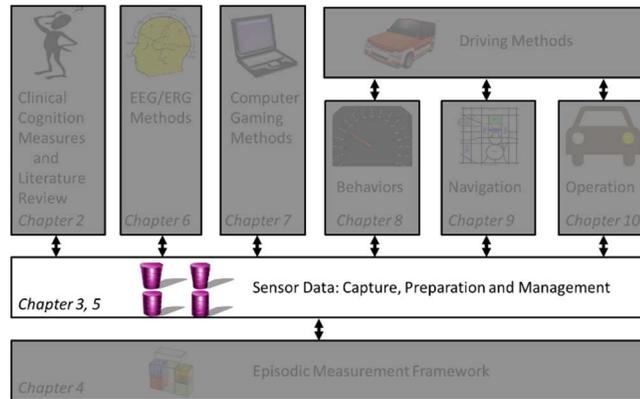
This work applies LDA classification to EEG/ERP and driving data sets where numerous features are potentially measured and the LDA algorithm provides the mechanism to reduce the dimensionality by determining the subset of the features that enable classification while discounting others. It also compares the results for other machine learning techniques including Support Vector Machines (SVM) [123], k-Nearest Neighbour (KNN) [124] and Decision Tree methods [125].

Classifier tools have been applied to cognitive measurement and decision making [126] where driver cognition responses captured through EEG were classified, while [127] using classification of fMRI data for vowel clusters as an indication of mental state. The diagnostic utility of EEG biomarkers for Alzheimer's disease, including ERP measures [128], was explored and LDA classifiers applied to EEG data to distinguish emotional state [129]. LDA classification was studied for EEG measures of ability within a number of instrumental activities of daily living (IADL) tasks, such as telephone use and basic handling of finances, [130] and applied to the classification of mental states from EEG signals [131].

Comparisons of the performance of different classifiers have been reported for the detection of epilepsy within EEG with over 97% accuracy [143], for the study of ERPs associated with facial vs non-facial image recognition with at 74.4% accuracy [144] and detection of a driver's cognitive state such as sleepiness within EEG measurements with 80% accuracy [145]. The relative performance of the various classifiers reported was varied between the papers showing that the classifier needs to be identified for each specific machine learning problem.



Chapter 3: Experimental Set-up



Objective: The research presented in this thesis uses multiple experimental and data collection methodologies because of the differing cognitive measurement techniques being studied and associated participant tasks being performed. This chapter presents the experimental set-up for each including research ethics clearance, data collection protocols, sensor systems and computational environment.

3.1. Gaming and EEG/ERP Study

The study of computer-based games and EEG/ERP for cognition measurement presented in Chapters 6 and 7 used two studies with volunteers using two computer games (Word Search and Sudoku).

Initially student volunteers used the games in a pilot trial to validate the operation of the games and capture data for algorithm development. A second study used the games with MCI volunteers from the Bruyère Memory Program at the Elisabeth Bruyère Hospital.

The MCI pilot project was designed to see if older adults with MCI would be able to play computer games three times a week for 9 weeks, it measured the participant's cognition at study start and end using standard clinical cognitive assessments and EEG/ERP tests to baseline their ability and to measure any change that occurred during the study.



3.1.1. Student Volunteer Game Pilot

Ethics Approvals

Research Ethics Board	Protocol Number	Initial approval date
Carleton	# 13684	August 28, 2013
Bruyère Continuing Care	# M16-13-045	September 24, 2013

Data Collection and Recruiting Protocol

Participants for the student pilot trial were recruited through in-class announcements, Emails to students and postings on campus bulletin boards. A total of 17 participants were recruited including 6 females and 11 males, with a mean age of 24.7 years and a standard deviation of 9.9 years.

Subjects participating in the study played the two games:

- 1) Carleton Sudoku Game (CSG) – A variation of the highly popular newspaper game that requires the participant to complete missing numbers in a 9 x 9 matrix.
- 2) Carleton Word Search Game (CWG) - A variation of the highly popular newspaper game that requires the participant to find words in a large matrix of random letters.

Procedures

Research set-up

- 1) Subjects were asked to read and sign the consent form.
- 2) Rules and use of each of the games explained.
- 3) Overall plan for the data collection phase was explained.

Data collection

- 1) Subjects were provided a quiet and comfortable (desk / chair) environment in which to play the games.
- 2) Round 1: Learning phase: Subjects played each game for 15 minutes followed by a 15-minute break.
- 3) Round 2: Optimal performance: Subjects played each game for 15 minutes followed by a 15-minute break.



- 4) Round 3: Distraction 1: Subjects played each game for 15 minutes while experiencing cognitive distraction simulating decreased concentration. The room contained highly variable noise to distract the participants.
- 5) Subjects were given a 30-minute break.
- 6) Round 4: Distraction 2: Repeat round 3.
- 7) Round 5: Return to optimal: Repeat round 2.
- 8) Subjects were thanked for their help in the study.

3.1.2. BRI MCI Game and EEG/ERP Study

Ethics Approvals

Research Ethics Board	Protocol Number	Initial approval date
Carleton	# 100794	November 22, 2013
Bruyère Continuing Care	# M16-13-043	November 19, 2013

Participant Recruitment

Participants for the MCI pilot were identified from patients of the Bruyère Memory Program (BMP) that have a diagnosis of MCI and have indicated they were willing to being contacted for research. If they consented to participate, they received cognitive testing and a baseline EEG. The neuropsychological tests selected for baseline testing were: RBANS, Trail Making Test A & B, and the MoCA.

Three research team members, one physician and two neuropsychologists, reviewed all cases to ensure participants had a diagnosis of amnesic MCI.

Study Groups

This pilot had a two-arm intervention trial, where participants were randomly assigned to the 1) BrainHQ and 2) Carleton Games. The groups were balanced for age, sex and education.

1) BrainHQ

Posit Science Corporation is a company based in San Francisco, CA founded by Neuro-scientists from the University of California, San Francisco and research partners from across the USA. Their product, BrainHQ, has modules they have shown to improve memory, attention and brain processing



speeds in healthy adults, as well as modules on navigation, intelligence and people skills. The modules have increasing difficulty and improving results are tracked on a scorecard. The first group of participants received 9 weeks of supervised BrainHQ training three times per week for 60 minutes. BrainHQ was chosen specifically because this particular program helps to achieve the project objectives to engage individuals with MCI in computer-based games that are programmed to 1) allow off-line monitoring of improvement or decline and 2) strengthen cognitive function through improvement in game performance. The program provides a scorecard that tracks an individual's scores and overall progress or difficulty with the various game tasks and so allows for the off-line monitoring of improvement or decline that this project needs.

2) Carleton Word and Number Games

CSG and CWG were used by this group and the participants received 9 weeks of supervised and monitored Carleton Games training three times per week for 60 minutes.

Participants were also asked to optionally complete a preliminary pilot study examining EEG/ERP.

No more than four participants at a time met, three times a week for 60 minutes. Each session was divided into 25 minutes of computer activity, followed by 5 minutes of break and discussion about the experience, followed by another 25 minutes of computer activity.

At the end of the 9-week period, cognitive testing and EEG testing was repeated.

Control Group for EEG Comparison

The participants recruited for the healthy control (HC) group were not patients from the Bruyère Memory Program, but instead were healthy adults over the age of 65. This healthy baseline group was recruited on a volunteer basis. Interested participants provided a health history after giving their verbal consent. These participants completed the same set of cognitive and EEG tests as the MCI group.



After cognitive testing, all classifications of MCI and HC were confirmed by a clinical committee. MCI participants consisted of seven males and eight females. They had a mean age of 75.7 years old and a mean education of 14.7 years. There were six male and eleven female HC participants. They had a mean age of 72.4 years old and a mean education of 15.6 years. The two groups did not differ significantly in age or education.

EEG Testing

The EEG was recorded using a 32 tin electrode nylon cap (Electro-Cap International Inc., Eaton, OH, USA) with the addition of four electrodes to measure eye movements. The EEG was amplified with NeuroScan NuAmps (NeuroScan, El Paso, TX, USA) at a sampling rate of 500 Hz in a DC to 100 Hz bandwidth. The 1-back task was presented to the participant using the E-Prime application (Psychology Software Tools, Sharpsburg, PA, USA) that was interconnected with the EEG recording so that stimulus events were automatically included in the EEG files as time markers for post processing.

3.2. Driving Study

The study of driving as a source of sensor information that could be used for cognitive measurement was completed through two separate data collection processes. One involved only members of the research group while the second was through collaboration with the CanDrive research project.

3.2.1. Research Team Drivers

The results for the data collected using research team drivers is presented in Chapters 9 and 10.

Data collection Protocol

GPS location technology was used to locate the subject's vehicle using an iPhone 4GS as the GPS location device. The iPhone ran a GPS tracking application [134] that logs the real-time GPS information to a network server (time, latitude, longitude, velocity, bearing, altitude) with a sampling rate of 0.2Hz when vehicle velocity is greater than 20km/hr, 1/30 Hz if velocity less than 20km/hr and



every 1/60 Hz if vehicle is stopped. The slower sampling reduces wireless data network utilization at low speeds.

The vehicles used were personal small to mid-sized 4 door vehicles that were outfitted with a camera capturing the dashboard so that turn signal use could be studied as shown in Figure 3-1. Dashboard was recorded with 480x720, 30fps video using one or two cameras depending on the dashboard layout and turn signal lamp placement. Drivers drove a pre-determined course.

The use of the iPhone application allows GPS data to be collected without the driver interacting with the device or reporting locations through other methods.

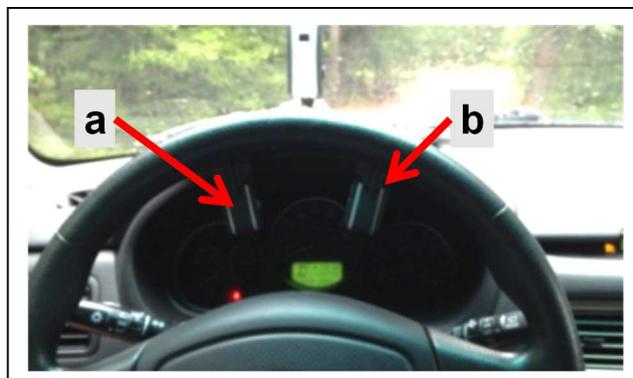


Figure 3-1: Example image of camera placement to capture dashboard video from turn signals (two camera configuration).

3.2.2. *Candrive Data Set*

The results for the Candrive dataset analysis is presented in Chapter 8.

Ethics

Research Ethics Board	Protocol Number	Initial approval date
Ottawa Health Sciences Network	# 2008610-01H	November 26, 2013 date Carleton University activity added
Carleton	# 100680	November 4, 2013
Bruyère Continuing Care	# M16-13-062	January 14, 2014

Data Collection Protocol

The data collection methodology and protocol [83 , 84] recruited volunteer drivers aged 70 and higher at entry to the project and had their vehicles outfitted with a Persentech OttoView-CD data recording device shown in Figure 3-2. The



block diagram shows how the device interfaced to the engine through the On Board Diagnostics II (OBDII) interface that has been standard on North American vehicles since 1996 and provides access to parameters captured by the engine control computer in addition to providing power for the data recorder. The recorder also has a GPS antennae providing GPS location information for the vehicle. The recorder has an optional RFID reader that can sense the proximity of an RFID tag on the driver's key chain to identify the driver.

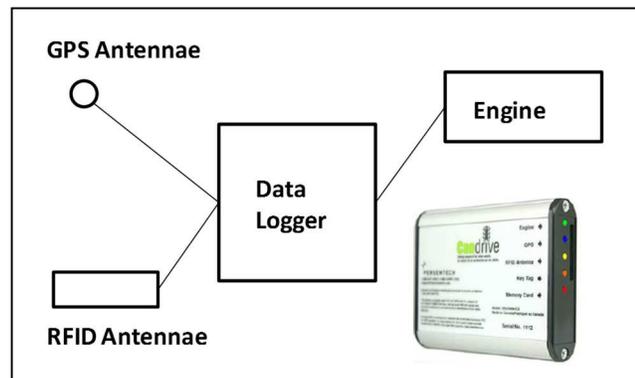


Figure 3-2: Block diagram of the data collection architecture and image of the Persentech OttoView-CD data collection device, [83 , 84].

Table 3-1: Information captured by Candrive sensor system. All data captured at a 1Hz sampling rate.

Parameter	Measure Value	Sensor
Time	Date/time (second)	GPS
Location	Latitude/Longitude, Fix accuracy	GPS
Velocity	km/hr	GPS
Speed Limit	km/hr	GPS/GIS mapping
Alerts	text (e.g., school zone)	GPS
Trip Data	Trip data, RFID tag #	OBDII recorder
Engine data	Engine RPM, Coolant and Intake air temperature, Absolute throttle position	OBDII recorder
Speed	Vehicle speed sensor (dashboard)	OBDII recorder

The parameters tracked and recorded are shown in Table 3-1 and these are stored locally on a memory card at a 1Hz sampling rate for all trips taken by the vehicle over the study period. Participants meet with the study team every four months and at that time, the memory card is swapped for a blank one allowing the data to be retrieved.



Table 3-2: Ottawa Candrive participant demographic information at project entry.

Number of participants	256
Participant age at entry	
Mean	76.3
Std Deviation	4.5
Range	70 - 92
70-74	106
75-79	90
80-84	47
85-89	12
90+	1
Gender F / M	36% / 64%

In this work, the data for the Ottawa based participants in the Candrive study were used and the attributes of the user group is summarized in Table 3-2.

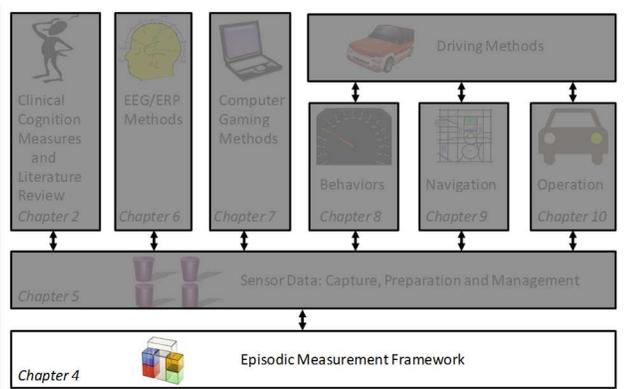
3.3. Platforms and Tools

Table 3-3: Summary of computing platforms used for the analysis

Analysis use	CPU	OS	Applications
Candrive anonymization	Intel i7-4500U @1.80Ghz, 8GB RAM	Windows 8 - 64bit	Dev C++ v5.7.1
Candrive analysis, ERP and games	Intel i7-4770 @3.4Ghz, 16GB RAM		Matlab 2014b and 2016a
Navigation and Signal use	Intel Q9400 @2.66Ghz, 4GB RAM	Windows 7 - 64bit	Matlab 2013a



Chapter 4: Episodic Measurement Framework



Objective: This chapter identifies challenges with the current clinical appointment based assessment process that achieves only annual or biannual assessments of patients. The chapter proposes a model for between appointment assessments that can utilize the cognition measurement methods presented in subsequent chapters. Topics covered include:

- Current clinical assessment process.
- Episodic Framework for sensor fusion from diverse sensor based measurement methods.
- Models for the inclusion of measurements from ERP, gaming and driving as examples of cognition measurement.

4.1. Clinical Assessment

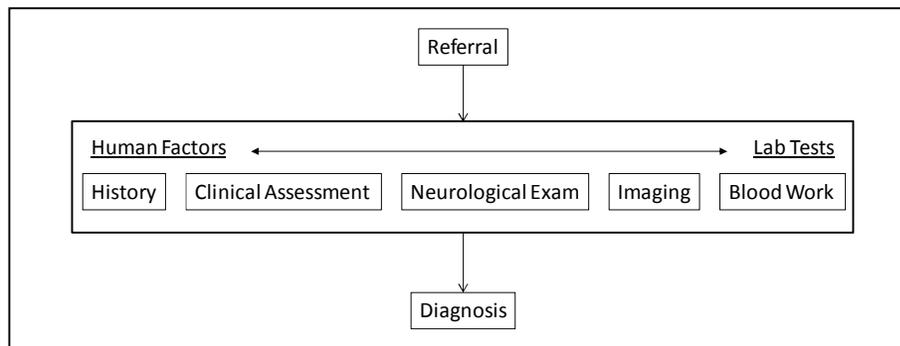


Figure 4-1: Flow chart of current cognitive assessment process.



An example diagnostic flow for a patient presenting at a memory clinic is shown in Figure 4-1 [132] where the patient has been referred to the memory specialist typically because someone has noted some form of memory issue and has raised the concern with the family physician. This is likely long after the cognitive decline has started as the brain is able to adapt and a person can modify behaviour (writing reminder notes) to mask early symptoms. The physician uses a combination of assessments to lead to a diagnosis. Fundamental to this process is that the impairments have led to tasks no longer being performed as well as previously.

The physician's assessment is limited to discrete measurements that they are able to perform in their clinic, request from labs or discrete reporting that occurs during the visit. As part of the process, the physician must also ensure that they identify and either diagnose or rule out other medical issues that include cognition impacting symptoms. The range of assessments used to diagnose MCI/dementia and distinguish from other causes that include cognitive symptoms include a range of human factors and lab tests [132]:

- **History:** The specialist will gather as good a history as they can through conversation with the patient and available family members or care givers to identify changes that are beyond normal aging. This assessment will include both questions about specific areas of the cognition, such as memory (remembering details of conversations), orientation to time (knowing the current date and upcoming events), executive functioning (ability to organize an event), and functional ability (balancing a cheque book, making a meal) with a focus on identification of abilities they can no longer perform but previously could.
- **Clinical Assessment:** The measurement of cognitive ability within a clinical setting can be performed through testing of specific cognitive areas, such as remembering lists of words for memory or copying a design for visuo-spatial construction. Usually these are measured as part of a formal test, such as MMSE [28], RBANS [29] and the MoCA [31]. These standardized tests allow the comparison of the subject to societal norms. Cognition can be tested indirectly using a functional test such as creating change for a purchase.



- **Neurological & Physical Exam:** There are no neurological or physical exam components that can be used to diagnose MCI/dementia. Rather, the physical examination of patients is used to identify conditions that could provide identifying contributing factors to the differential diagnosis, such as vascular disease or Parkinson's disease.
- **Blood Work:** There is currently no blood test to diagnose MCI/dementia, but again blood tests should be performed to rule out health conditions that could mimic dementia, such as hypothyroidism or a severe B12 deficiency.
- **Medical Imaging:** Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) is again used to rule out other potential causes of cognitive decline, such as brain tumors, normal pressure hydrocephalus or strokes.

The physician uses the combination of these assessments of history, biological factors and objective measures to lead to a diagnosis of healthy aging, subjective cognitive decline, MCI, dementia, or perhaps another cause that has presented with cognition related symptoms. Current research is exploring the possibility that waiting for dementia to be present is too late to begin treatment, so some studies are treating MCI [25, 27].

The measurement of cognitive ability within a clinical setting is typically done through the application of standardized clinical tests with these tests performed by a trained practitioner and include testing of memory, language and executive functioning helping lead to diagnosis.

Cognitive testing is heavily influenced by life experience and education as a patient's results are compared against societal norms, performance measures derived from a composite of a large population sample. Exceptions that cause a given subject to differ from the norm sample set such as heritage, immigrant and non-English speaker, very high or low levels of education or ability can cause the norms to not apply. In addition [32] challenges such as patient tiredness, focus or time of day that affect clinical measurement of cognition and cognitive decline in early stage dementia introduce variability in the results of discrete clinical memory tests. In addition, the amount of testing is limited by the clinician's schedule – for every additional test the next patient in the waiting room will be delayed further.



4.2. Episodic Framework

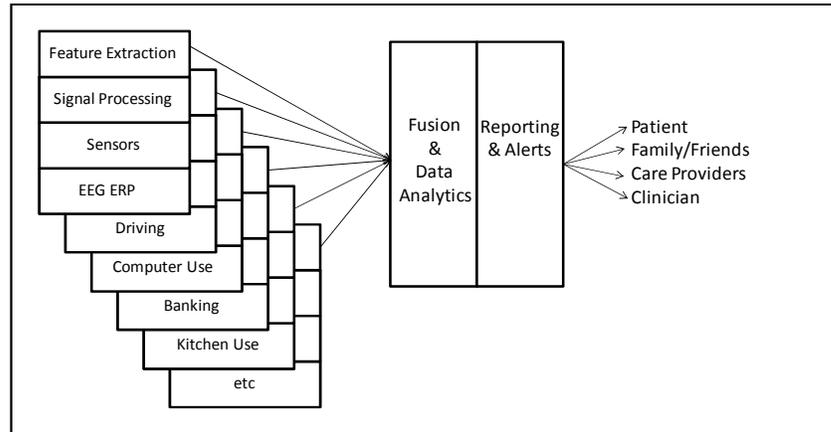


Figure 4-2: Episodic cognitive assessment framework showing various sensor sub-systems with the fusion framework and data analytics system.

The emergence of new sensors, signal processing, and big data analytic techniques lead to improved measurements and knowledge about the patient. These new assessment strategies can provide the physician with additional insight through assessment between clinical appointments leading to a better diagnosis, including more insight into the changes and issues that the specific patient is experiencing in their day-to-day life. Consumer EEG sensors are even becoming available such as the EPOC+ (Emotiv Inc, San Francisco, U.S.A.) and Muse (Muse Inc, Toronto, Canada) leading to the potential for EEG based measures to be completed outside the clinical setting. Improved information about a patient's well-being and cognitive ability can be acquired through the inclusion of more frequent episodic assessments outside of their clinic appointments. This thesis proposes an episodic framework for cognitive assessment and identifies the elements required for its successful deployment. The episodic cognition assessment framework is shown in Figure 4-2 and consists of three major blocks of capability.

- **Sensor sub-systems:** For each of the sources of cognition assessment, a specific set of sensors is required with the signal processing and feature extraction algorithms.
- **Fusion and Data Analytics:** This is key to bringing the results of the diverse set of sensor sub-systems together to create a holistic view of the patient. A key feature of this layer is the management of the episodic nature of the



individual sensor sub-systems as neither driving nor computer game play are likely to occur every day. Trends of various activities would be required to identify change over time.

- **Reporting and Alerts:** A key benefit of the episodic assessment of cognitive ability is providing better information to the physician. The reporting function will generate this summary information, providing the physician with a much more detailed, complete and accurate patient history. In addition, the system would also generate alerts to the patient as reminders or a patient's family/caregivers should significant changes or issues be identified.

4.2.1. *Activities of Daily Living - (ADL)*

The challenges with the accurate reporting of changes in a patient's functional abilities within the history could be addressed through the episodic measurement of cognition. Changes within Basic Activities of Daily Living - (BADL) that include basic hygiene activities (dressing, mobility, bathing, self-feeding, etc) and Instrumental Activities of Daily Living (IADL) that involve more complex tasks (food preparation, shopping, driving, social activities, using a computer) could be measured. Clinicians gather insight into the intactness of a patient's ability to perform BADLs and IADLs as part of the patient history. Changes in a patient's ability to perform a task are usually an indication of cognitive change, assuming the clinician has ruled out physical changes as a cause. The direct and ongoing measurement of these activities [140, 141] through the application of sensor systems and associated signal processing and data analytics can overcome limitations on the frequency of clinical cognitive testing through observation of their ongoing capability.

The deployment of sensor solutions within smart apartments has been presented [50, 51, 141] where a patient's ongoing BADLs and IADLs are monitored by detecting specific activities such as appliance usage, washroom use and kitchen use. Sensor based reminder systems have been applied to hand washing [142]. As cognitive impairment advances, first IADLs are affected and ultimately BADLs are affected. Thus monitoring of IADLs will provide earlier symptoms for diagnosis of cognitive decline.



4.2.2. Instrumental Activities of Daily Living - Games

Computer use is an example IADL and measurement of mouse movements, game play and web browsing has been presented [53, 55]. Specifically, [53] mouse movement was measured during web browsing, but has not been used as a measure of ability within game play. The assessment framework details for games are shown in Figure 4-3 and can be applied to various games. Two such games are based on numerical or word based foundations are presented in Chapter 7, allowing the different regions of the brain to be used, along with the need for brain functions for search or logical deduction. Macro performance (time to complete, successful completion) for games provides an indication of the ability of the user where changes over time can provide an indication of changing cognitive ability. This is only a portion of the information that can be measured.

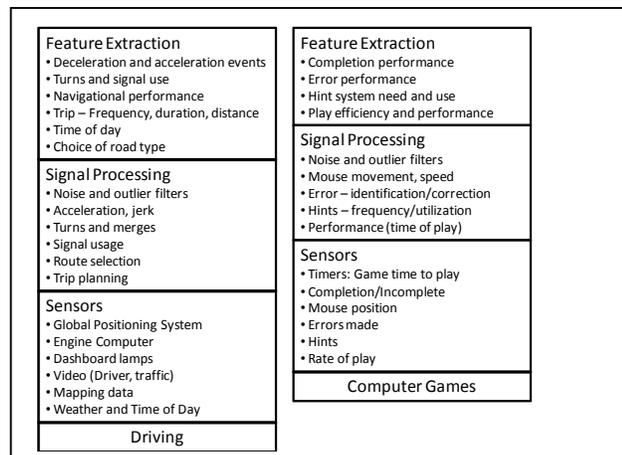


Figure 4-3: Episodic cognitive assessments sensor sub-system details for driving and computer games as examples of two IADL.

Games can be an indirect way to measure a patient's cognitive well-being, providing they are easy to learn, play and age appropriate for older patients. Changes in game play ability, such as smooth and effective motion of the mouse becoming more random and inefficient, can be an indication of cognitive change which may be a proxy for changes in ability in the kitchen [57]. The measurement of a user's interaction with a game includes many other aspects, such as rate that individual steps are performed, errors the user makes and associated correction or



non-correction, need for hints to reach a solution, solution strategies used and mouse usage as indication of focus area within a game.

4.2.3. Instrumental Activities of Daily Living - Driving

In addition to computer use and home based activities of daily living, automobile driving is an activity that requires high levels of cognitive function, including executive function for navigation and trip planning tasks and self-monitoring for over-learned tasks like blind spot checks, mirror use and vehicle operation (signal use, velocity, braking, acceleration). All drivers must perform these activities on most trips and hence these skills can be measured. The framework can indicate changes in ability as summarized in Figure 4-3.

The deployment of sensors within personal vehicles allows direct measurement while a driver goes about their daily activities. This is an advantage over other methods of driving testing as the driver has the comfort of driving their own vehicle on the roads that they normally choose while going to their regular destinations demonstrating their normal behaviour and activity level. This allows focus on measures related to vehicle operation, high level navigation performance through to day to day levels of activity.

Trip level decision making can be a measure of executive cognitive function for older drivers using vehicles with installed sensor technology [83, 84, 70] by providing insight into driver decision regarding trips including frequency, duration, distance and time of day. Measurement methods are presented for driver behaviours, including identification of unique driver behaviours in Chapters 8, 9 and 10.

4.3. Episodic Cognition Assessment Fusion

The details of the cognition assessment data analytics and fusion block for just driving information is shown in Figure 4-4. The primary input to this block is the features generated by the IADL driving sensor system that will produce a data structure of features for each trip. This block is responsible for tracking the ongoing ability and performance of the patient within a given task and the identification of significant trends such as deteriorating ability or even the cessation of an activity.



The block will use filtering, classification and data fusion techniques to bring together the results into a single view. Filtering techniques such as Kalman filters can be applied to the data to identify and remove outlier trips from the analysis as simple trips like moving a car in a driveway should be discarded while one off outlier events, like a late night trip to pick someone up from an airport, would not represent a behaviour change. Statistical analysis and classification techniques such as linear discriminant analysis and adaptive neural networks along with sensor fusion will allow for the generation of 4 key measures of the driver.

Table 4-1: Clinical and Episodic cognition measurement techniques.

Method	When is the test performed	Where is the test performed	How is cognition measured	Who performs assessment	What is patient doing	Status of the method	
Clinical Assessments							
History	Discrete measures of ongoing behaviour	Interview of subject or care giver	Behavioural based indications	Clinician	BADL & IADL	In use	
Cognitive Assessment	Discrete test protocol	In clinic	Discrete cognitive areas		Specific tasks		
Neurological Exam			Rule out other illness		Rule out diagnosis		
Blood Test							
Imaging			Specialist				
Episodic Measurements							
BADL Kitchen	Ongoing day to day	In home	Behavioural based indications	Unattended	BADL & IADL	Transition	
BADL Hygiene							
IADL Games					In vehicle	IADL	Emerging
IADL Driving							
IADL Banking		In home/store					
EEG/ERP with NBT		In home	Direct biomarker and behavioural based indications		Specific Test	Concept	

Quantitative Statistics: Foundational statistical measures for how much the person drives, how often they drive and how many different places they go on their trips.



- **Risk Measure:** A specific measure to look at trends in the driver's choices that would indicate increases or decreases in exposure to risk. Drivers may choose to compensate for decline by reducing higher risk driving, such as at night.
- **Complexity Measure:** This provides a measure of the driver's choices around the complexity of their driving, such as multi-stop combination errands becoming multiple single stop trips.
- **Behaviour Measure:** This measure focuses specifically on how they operate the vehicle and can be an indication of insight into one's abilities. For instance, using the flasher to indicate lane change is an over-learned activity. If a driver forgets to use the flasher, this may cause the driver behind to honk in annoyance. It requires self-monitoring to make the connection between the forgotten flasher and the honking.

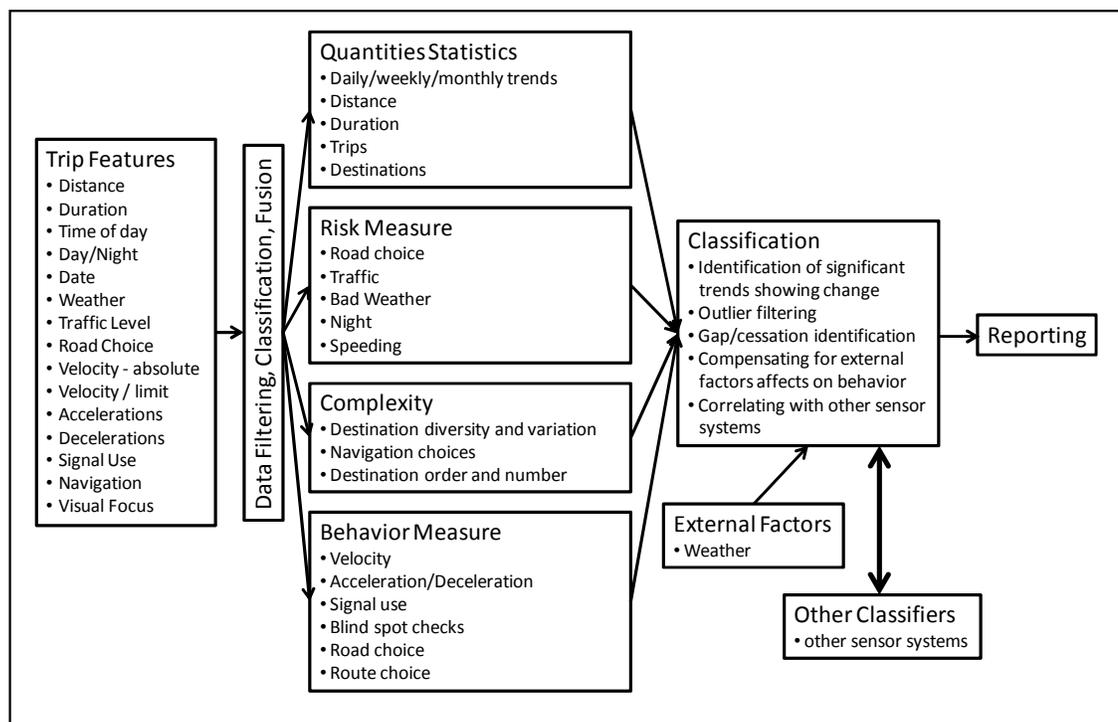


Figure 4-4: Episodic cognitive assessments framework details for assessment of driving behaviour as one example IADL.

In the last stage, additional information is brought into the classification stage, such as driving cessation. This behavioural issue reflects judgment, for instance a



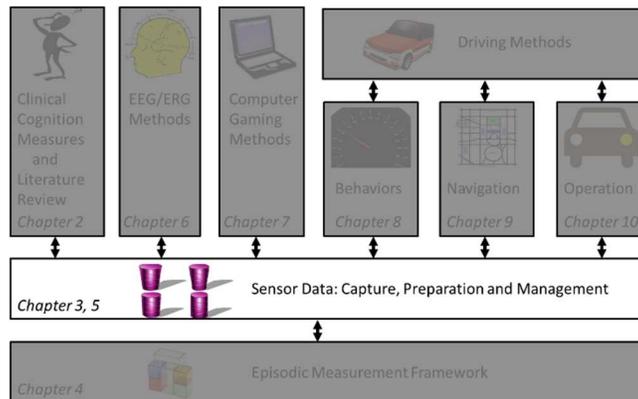
major snow storm should affect the driving routine. Access to the information from the other sensor systems at this stage can also allow reduced driving to be correlated with increased computer use for a patient that is ill and choosing to stay at home to recover or to identify a common decline trend across sensors. The Episodic Cognition Framework could take the results and combine them to provide a better indication of changes such as poorer performance measures associated with game play and a reduction in trip complexity and/or navigational performance could be combined to provide an indication of a significant change.

4.4. Discussion

The episodic cognition assessment framework has the potential to provide physicians with significantly better information on which to assess the well-being of patients. The framework applies emerging sensor technologies, signal processing and data analytics techniques to measure performance of IADLs as indicators of cognition. The ongoing nature of these measures between appointments while the patient goes about their normal daily routine could help to objectively show the evolution of cognition decline. As an example, gaming based measurements, with irregular or regular play and associated measurements of progression of skill, can show a lot as they provide a consistent measure against the person as their own norm. The result is objective measurements that can support subjective reporting by the subject such as "I used to be able to do this and now I can't".



Chapter 5: Data Preparation and Signal Processing Methods



Objective: The collection of data from sensor systems from volunteer study participants introduces many challenges for the data capture, collection and subsequent processing. This includes issues related to noise, outliers, data size, privacy and anonymity. This section will identify each of the issues with the collected data, explore potential solutions. Topics covered include:

- EEG/ERP
 - ERP measurement and foundation signal processing techniques.
 - Outlier detection
 - Noise filtering.
- Computer Games
 - Data preparation techniques to deal with variation and unexpected user use models.
- Driving
 - Sensors for driving measurement and available signals.
 - Data integrity processing required for long term longitudinal studies.
 - Augmentation of measured data from secondary sources such as GIS and weather data.



- Outlier filtering of measured GPS data.
- Trip exclusion filtering for non-meaningful trips.
- Privacy and anonymity solutions.
- Computation scalability for large data set analysis.

5.1. EEG/ERP

This section presents the preparation of the raw EEG/ERP sensor data for use in research, where the continuous EEG files were initially segmented into 1100 msec epochs for each stimulus event, including 1000 msec post-stimulus and 100 msec pre-stimulus. The pre-stimulus period provides a reference to establish a zero baseline for each of the trials.

Individual EEG epochs contain significant information related to uncorrelated and/or unrelated activities by the participant during the test, such as ocular activity as simple as an eye blink. These signals represent outlier EEG traces and need to be removed before the ERP signals are calculated and can be identified by significantly larger signal ranges [34] than other trials for the participant. In this work, the individual EEG traces for a participant were analyzed to measure the signal range for each EEG trace leading to a mean and standard deviation for the range for the participant. An exclusion threshold of three standard deviations from the mean range for the participant was applied to the individual signals. This method proved to be a robust outlier removal method for the EEG traces being analyzed and is generally applicable where EEG measurements do not suffer from intermittent or poor sensor contacts such as those completed in a controlled and supervised environment and also over short periods.

The remaining EEG traces for the participant were then filtered to reduce the signals that are uncorrelated with the stimulus event by time-aligning the traces, using the stimulus event for alignment and averaging the EEG traces for the participant across the trials. This resulted in an ERP response for the participant that was then low-pass filtered to remove additional high frequency noise using a Blackman window low pass linear phase Finite Impulse Response (FIR) linear



phase filter with a pass band corner frequency of 12.5Hz and a stop band frequency of 25Hz.

Table 5-1 summarizes the number of trials that were available for the 32 participants. For the response required trials, each participant performed 60 trials, while for the no response required trials, each participant performed 120 trials. Table II shows statistics for task accuracy by trial type for each participant, as well as number of outliers identified.

Table 5-1: Task accuracy and number of outliers by participants and response type.

Signal	Case	Number of correct EEGs		Outliers	
		Mean (st dev)	Min / Max	Mean (st dev)	Min / Max
CPz & Pz	Response required	57.3 (3.2)	44 / 60	0.75 (0.57)	0 / 2
	No response required	119.2 (1.3)	114 / 120	1.44 (1.11)	0 / 5

5.2. Carleton Games

The user activity log files captured within the Carleton Sudoku Game (CSG) and Word Search Game (CWG) required minimal processing to prepare them for analysis within Matlab. The only issue within the files that was identified that prevented the direct loading of the files was within the pencil mark feature of the games. When a user made use of the pencil mark capability, all of the actual text they placed in the field was included in the log for future analysis. It was incorrectly assumed in the design specifications that users would only place the numbers 1-9 but in actual play it was found that some users entered commas between the numbers and in one case the "\" appeared regularly. Both of these prevented the log file from being loaded directly as Comma Separated Value (CSV) table of numbers. A pre-parser C++ application was created that removed all non-numeric characters from the pencil mark field and any log entry that was modified had a flag set to indicate the modification for future reference if required.



5.3. Driving Studies Candrive Data Set

This section presents the preparation of the raw Candrive in-vehicle sensor data for use in research and a theoretical open access model that achieves the necessary participant anonymity.

5.3.1. *Sensors and Data Collection*

The Candrive project [83] Research Assistants installed an OttoView-CD data recording device (Persentech, Winnipeg, MB, Canada) into participant vehicles. The recorder captured 1Hz sampled data from a GPS location antenna and the OBDII data from the engine computer. On subsequent meetings with the participant, they verified the installation while changing the SD card to collect the data. The recorder included an RFID reader that could detect an RFID key tag as a mechanism to distinguish between multiple drivers for a shared vehicle. The sensor has an OBDII interface based on the SAEJ1979 technical standard that powers the device in addition to the engine computer parameters. The GPS receiver and antennae is a 20-channel, 1575 MHz SiRF Star II e/LP with 10m position and 1.0km/hr velocity accuracy. The RFID reader uses passive tags at 125 KHz. The recorder supports 2 GB SD memory.

The captured GPS and engine parameters are summarized in Table 5-2. The OBDII port does not provide access to turn signal, braking or anti-lock brake system information as these are controlled outside the engine computer and are not easily accessible on a standardized interface. The OBDII port was designed with a primary purpose of engine diagnosis and maintenance, but to minimize memory use many other engine specific parameters are not recorded.

For the Candrive study, data are locally stored on a SD card to avoid need for wireless network connection. Participants met with the study team approximately 3 times per year and the data on the memory card was retrieved.



Table 5-2: Sensor data captured by Persentech OttoView-CD sensor system.

Source Sensor	Measured Value	Units / format
Persentech device	Sensor Serial Number	Integer
Persentech device	Trip number	Integer - increments for every vehicle start
GPS	Date & Time	Date string e.g. "18 January 2013 10:05:10"
GPS	Longitude Latitude	Degrees
GPS	GPS Fix Status	Text string e.g. "3D Fix"
GPS	GPS Degrees of Precision	Floating Point
GPS	Velocity	km/hr
GIS	Posted Limit	km/hr
GIS	Alerts	Text String - see table xxx
RFID	RFID tag number	Hexadecimal serial number
OBDII	Engine Coolant Temperature	Degrees C
OBDII	Engine RPM	Integer
OBDII	Velocity (vehicle speed sensor)	km/hr
OBDII	Air Intake Temperature	Degrees C
OBDII	Ambient Air Temperature	Degrees C
OBDII	Absolute Throttle Position	Floating point
OBDII	Relative Throttle Position	Percentage

5.3.2. GIS Data Augmentation

The data from the SD card was initially processed by software tools provided by the sensor manufacturer that converted the data from a proprietary storage format to Comma Separated Value (CSV) format.

Table 5-3: Example road hazard and road condition information included through the gis augmentation. These messages can occur individually or in combination.

Crosswalk
School Zone
Red Light Camera
Playground
Deer Crossing
Hazardous Intersection



The post processing application included GIS map databases for the regions in the study and CSV samples were augmented with the posted speed limit for the road if known and any road hazards/conditions summarized in Table 5-3.

The tool generates a detailed 1Hz sample record and summary records for each trip including date, distance traveled, duration, and trip start and end time defined as the engine turning on and off respectively. The resulting CSV files included the fields in Table 5-2 and blanks for OBDII values that were not recorded resulting in 141 fields for each sample.

5.3.3. Data Integrity and Data Format Processing

The resulting CSV files were further analyzed for data integrity issues and data reformatted to ease subsequent processing. The data integrity issues included:

- Identification of any incomplete sample records: Infrequently single samples were discarded that did not include all fields potentially because of power failure within the recording device, OBDII interface issues or other unknown causes. Subsequent signal processing included considerations for missing samples.
- Missing GPS information: This may occur when the vehicle was in a tunnel, parking garage or other area with poor GPS reception or due to GPS failure such as sensor disconnection. These samples were tagged to allow subsequent inclusion or exclusion as appropriate [121].

Table 5-4: Summary of the data integrity results for the transformation of the files for 13 Candrive Participants for 4 years of detailed driving data. The participants were chosen based on a sample of convenience.

	Minimum	Mean	Maximum	Mean Std Dev
Total Data Size for CSV files for participant (MB)	1350	2250	4010	822
GPS location failure (% of samples)	0.0%	5.0%	91.8%	11.7%
GPS excluding outlier files (% of samples)	0.0%	2.9%	10.3%	1.8%



Table 5-4 provides a summary of the data archive files for a sample of convenience from the Ottawa Candrive dataset and shows the significant size of the detailed sensor data that has been collected for each of the drivers. For this set of participants, there were no incomplete samples that had to be discarded. The table also shows the performance of the GPS sensor. One of the vehicles had 4 archive files with between 31% and 91.8% failure while another vehicle had a single archive with 73% GPS failure for a total of 5 of the 152 archive files. The cause is likely a faulty sensor connection to the recording device as the faults extended over periods of days.

To ease subsequent processing, the CSV files needed to have numerical only data, string data were converted to a numerical representation:

- Date/Time stamp information that was provided by the sensor system in a "18 January 2013 10:05:10" format was converted to fields for each of year, month, day, hour, minute and second using a 24-hour clock.
- GPS precision information such as "3D Fix" was converted to a numerical code representation for each of the precision cases.
- Road hazard information as shown in Table 5-3 was converted to a unique numerical code. The hazards were frequently seen in combination such as a Crosswalk within a School Zone so a unique code was created for each unique combination resulting in 22 codes.

The processing applications for this stage were coded in the C++ language, which is ideally suited to process the ASCII CSV files. To enable automation of the conversion, the tools used an adaptive list technique for the text strings conversion. Specifically, the set of road hazards / road conditions was unknown a priori so the programs compared every GIS message to a table of messages observed so far. If a new message was found, a new table entry was created with an associated new code number.



5.3.4. Trip Inclusion and Outlier Filtering

Non-meaningful driving information such as warming up a vehicle or moving it within a driveway was excluded as all trips <100m in length or <120 seconds in duration were excluded from the analysis.

An outlier detection and correction algorithm was applied to the GPS timestamp, longitude/latitude and velocity data where the relative difference between the two velocity measures was calculated per equation 5-1:

$$\begin{aligned}
 & \text{if } (V_{GPS}[n] == 0 \text{ and } V_{OBDII}[n] == 0) && [5-1] \\
 & \quad \text{relativeDifference}[n] = 0; \\
 & \text{else} \\
 & \quad \text{relativeDifference}[n] = \\
 & \quad (V_{GPS}[n] - V_{OBDII}[n]) / \text{mean}(V_{GPS}[n], V_{OBDII}[n]);
 \end{aligned}$$

The standard deviation of the relative differences was calculated and GPS outlier samples were identified as those samples with relative difference greater than +/- 3 standard deviations. In these cases, the OBDII velocity sample was used as a replacement velocity measure.

Outlier date and time stamps were detected by ensuring date/time stamps were monotonic increasing and removing any samples outside 24 hours of the trip mode date allowing a trip to cross over midnight once.

Outlier latitude and/or longitude samples were detected by first rejecting all samples that indicated invalid GPS through the GPS fix signal and then through identification of any individual samples not on the trip path as indicated by samples that were significantly off (>100km) from prior and subsequent samples. These samples were removed from trip distance travelled calculations.

The GPS outlier detection algorithm impact was found to be highly dependent on the driver, as a driver's regular driving patterns may typically include or exclude areas with poor GPS reception. For many drivers and for most trips, there were no outlier values. The impact to the measures for driving distance, driving time and number of deceleration events for most drivers was minimal or nil.



5.3.5. Fusion with Additional Data Sources

In addition to GIS information, the data can be augmented with data from other sources. In many regions, historical weather data are available for local weather reporting stations [121]. These data are available from government and private sources in both summary form, such as daily summary, and more time precise information, including hourly reports. The geographic location of the vehicle can be used to identify the nearest weather recording station, and this weather data can now be added to the records with attributes such as temperature, precipitation amounts for rain and snow, visibility, and wind speed and direction.

The time of day and location information can be used to determine solar day for sunrise and sunset allowing classification as night, dawn, day, or dusk.

5.3.6. Privacy and Anonymity Considerations

The rich set of data amassed for a single driver or vehicle is substantial and hence the issues of privacy for the individual driver are significant. In addition to requirements for anonymity that are required through research ethics, many other applications that capture driving data will need to ensure the privacy of the driver. This is an example of a general problem that is increasing in prevalence with the increasing deployment of GPS sensors (smart phones) and associated ability for this information to be tracked, monitored and recorded. Privacy and anonymity issues presented in the literature and discussed in section 2.4.1 introduce the concepts of k-anonymity and differential privacy within other applications and the challenge of GPS sensors. This work needs to apply these concepts to the anonymity in recorded GPS data and associated augmentations of the data from GIS and other sources. As a result, a number of privacy attributes were analyzed based on ensuring k-anonymity across the data with differential privacy options:

- Direct driver identification: All driving files are stored without any demographic or personal information (name, sex, birth date, address). A participant ID allows multiple files for a participant to be associated.
- GPS location differential privacy: The data are stored using a zero longitude/latitude origin where the first sample in each trip is set to 0,0



and all subsequent samples relative to this. This allows three levels of differential privacy the files:

- 1) "as is" provides no indication of actual location
- 2) "original region" (e.g. Ottawa, Canada) allows a shift to the general region but not to the precise location, or
- 3) "absolute origin reference" allows the files to attain the full precision of the original GPS data.

The privacy management of the location information is absolutely critical. The full precision data can easily allow identification of the participant's home (where the vehicle spends most nights), place of work or other frequent daily destinations and other common locations allowing identification and knowledge of private details.

Table 5-5: Summary of the effects of differential precision for sunrise / sunset calculations for Ottawa On metropolitan area

Attribute	Value
Population	1.2M
Metropolitan Area Size	~100km East - West ~50km North - South
Variation in SunRise/Sunset time from Center to East/West limits	+/- 4 mins
Variation in Solar Day SunRise to Sunset time from Center to North/South limits	+/- 1 min

Within the dataset, the seven parameters that in combination potentially allow the re-identification of the user are longitude, latitude, region, sunrise, sunset, weather data and road hazards. The differential privacy technique is applied to the parameters to achieve k-anonymity for any combination of these parameters to ensure that they cannot be used to rebuild the vehicle location:

- Sunrise and Sunset information: The date with sunrise and sunset times can provide the trip location as the difference between sunrise and sunset indicates latitude while the absolute value of sunrise or sunset indicates relative longitude within a time zone. As an example - for Ottawa, Canada shown in Table 1-4 (45.4214° N, 75.6919° W), a 1 second precision



sunrise allows longitude to be determined within ~200m (100km width has ~500 second difference in sunrise time).

- Weather information can also be used to identify a region through archived online weather reports. In the differential privacy model, this data may only be shared where there is some level of trust in the user such that they can know the driver's region.
- Road hazard/Road condition information can identify a more precise location for a vehicle through comparison with GIS map data. The relative location for example of 2 school zones and the associated road network connections could be used to precisely relocate the trip. To prevent this, the numerical code and conversion to text would only be shared with trusted data users.

As a result, k-anonymity is required in the sharing of the above attributes such that weather information can only be shared with users that are trusted to some level, as this information provides an alternative mechanism to determine the regional location of the driver. Sunrise/sunset information can be shared using a differential privacy model where it is calculated not for the absolute location of the driver but instead for the general region for the trip.

The use of differential privacy on the sunrise and sunset calculations is shown in Table 5-5 for an example metropolitan area where the center of the region has the effect of changing the sunrise and sunset by at most 4 minutes while the impact on the total length of the solar day is 1 minute. These small impacts on the precision of solar cycle information will have minimal impact on the use of the data, while ensuring the privacy of the study participant. Data could have additional noise added so that there is a larger uncertainty in the solar day length.

The differential privacy models are summarized in Table 5-6, where 3 models have been provided for “fully trusted”, “intermediate trust”, and “open access” users and their access to the data. With 3 levels of trust for the data, the open access model provides the highest level of privacy, while the intermediate trust level provides additional details. The models also provide guidance for maintaining privacy in published results. A single master big data archive of the data with the



origin of trips set to zero longitude/latitude and a set of supplemental files allows the data to be shared while ensuring privacy at the three trust levels.

Table 5-6: Differential Privacy Models

	Fully Trusted User	Intermediate Trust User	Open Access User
Dataset Access	Zero referenced Longitude/Latitude trip Data		
Regional Weather Data	Yes	Yes	No
Sunrise/Sunset Data	Yes	Yes for region center	Yes with added noise
Location of trip origin	Absolute origin	Regional central reference only	No
Road Hazards	Codes and code meanings	Codes only	Codes only
Latitude Accuracy	GPS limited +/- 10m typical	Defined by size of the region for Ottawa +/-25km	+/- 100km
Longitude Accuracy	GPS limited +/- 10m typical	Defined by size of the region for Ottawa +/-50km	+/- 100km but for every time zone

Open access users of the data will at best be able to calculate the actual latitude and longitude of the user to a controlled precision that can be set to +/- 100km with minimal impact on the data. Anonymity is further enhanced as this position will be relative to the boundary of an unknown time zone.

Users with intermediate trust on their use of the data can be provided additional supporting information that enable more accurate location information, while still ensuring the absolute privacy of a participant's home, place of work or other destinations. Fully trusted users have the absolute location to the limit of the GPS precision and hence identify a participant's home and other frequent destinations.

These techniques for differential privacy and data augmentation are generally applicable to any study of location based behaviours of individuals or groups. That would include studies of driving related activity such as individual driving studies or studies of groups of vehicles such as fleets. It is applicable to studies beyond



vehicular use as it could be applied to other modes of transportation such as cycling and also walking.

Figure 5-1 shows the resulting automated workflow for the processing and augmentation of the sensor data through the stages required to convert the raw sensor data into a sharable repository. Figure 5-2 shows the plot of an actual trip for one participant at the intermediate trust level as it has been repositioned to the Ottawa region only. This trip is 13.4 minutes in duration and 7.7km in length. The trip has been shown with the GIS map data augmentation including speed limits and road hazard / road condition information.

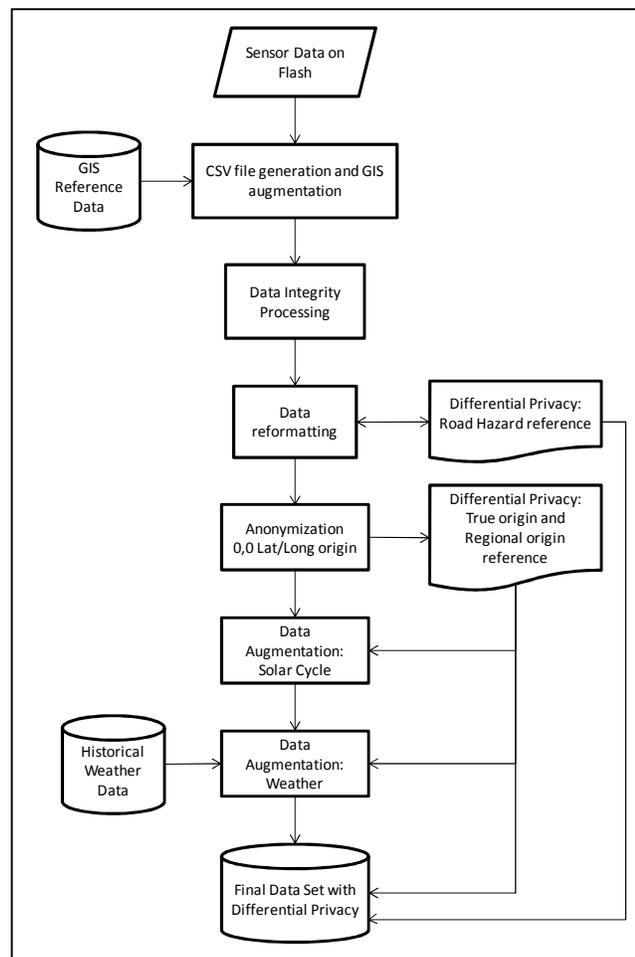


Figure 5-1: Flowchart showing the processing steps required to convert raw vehicle sensor data to a reference repository with Differential Privacy and k-anonymization. Processing steps include the augmentation of the sensor data through fusion with GIS and weather reference data sets.

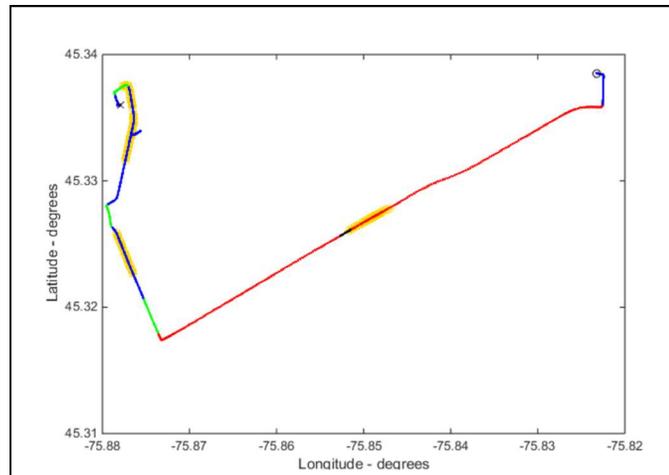


Figure 5-2: Example trip path presented using Intermediate Trust level where the data has been located back to the Ottawa area while continuing to not show the precise origin and destination of the user. Trip starts at the 'o' and ends at the 'x'. The posted limits on the roads traveled are 40km/hr (blue), 60km/hr (green), 80 km/hr (red) and 100km/hr (black). The regions highlighted in orange have road hazard or condition notifications.

5.3.7. Alignment to Cloud Computing Programming Models

The Candrive dataset aligns with the scalability attributes required by the Hadoop architecture for cloud processing. The trip data sets could be easily distributed within a Hadoop distributed file system, while the processing is highly scalable across cloud computation since the processing of the data for individual trips and/or drivers provides simple foundations for parallel processing.

Table 5-7 provides a demonstration of the scalability of the processing of the driving data making use of a multi-core architecture. An identical extensive feature extraction and analysis algorithm was used for both of the test cases without change and four years of data records for three users were used as a test data set. The data sets for the three drivers captured approximately 111,000km, 43,000km and 102,000km of driving and 5899, 2217 and 9813 trips respectively. The algorithm was configured to process a batch set of Candrive records where: Case 1: a single Matlab image had all of the files assigned to it; Case 2: the files were divided such that each Matlab image was assigned 25% of the files while running within the same system sharing a common file storage system. The result is that by distributing the task across four Matlab images the total elapsed time was



reduced by 61% while the total combined computation time for the four parallel tasks was only increased by 25%.

Table 5-7: Summary of performance impacts and the ability to parallelize the computation for the analysis of the all the data for Three Candrive participants.

Attribute	Case 1	Case 2
Software	Windows 8.1 64bit Matlab 2014 64bit	
Hardware	Intel Core i7-4770 CPU @ 3.4GHz (quad core)	
Storage	Solid State SATA	
Processing Model	Single Matlab instance	Four Matlab instances
Matlab application	No use of Matlab Parallel Processing Toolbox	
Data Assignment	All data processed by single instance	25% of the data files assigned to each Matlab Instance
Processing elapsed time	0.92 hrs	Range 0.28-0.35 hrs mean 0.29 hrs maximum 0.35 hrs
Total CPU time	0.92 hrs	1.14 hrs

5.4. Discussion

Tracking of longitudinal driving data has many applications, whether it be the driving behaviours of older adults that is the purpose of the Candrive study, the management of fleet vehicles to optimize utilization and maintenance strategies, or the study of traffic flows and congestion by planners exploring the need for changes to the road network. All of these applications are dependent on the accessibility to big data sets of longitudinal sensor records.

This work shows that longitudinal data, once captured with “simple-to-deploy” vehicle-based recording systems, can be automatically processed to ensure the integrity of the sample records. It also shows the automatic augmentation of the data with additional information through fusion with other data sources to add sun cycle, weather, road speed limit and hazard information. These come from GIS sources, enhancing the potential analysis and research questions that may be applied to the data.

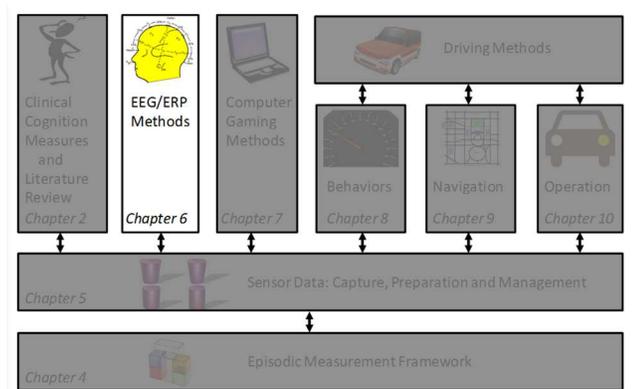


It then presents general solutions to enable a differential privacy based model for sharing the data of GPS based driving logs with other researchers that allows for use by researchers with different trust levels. Specifically, it presents an example of three trust levels for the data. The data set has been analyzed based on a k-anonymity model to ensure that data set parameters individually and in combination have been captured and are shared so the participant anonymity is maintained. This general solution for privacy can be applied to many studies of location based behaviour.

Lastly, it shows that the resulting data set supports the distribution of analysis processing across multiple processors within a cloud-computing model. The result is large-scale research repositories for use by multiple researchers while still ensuring research participant privacy and anonymity.



Chapter 6: EEG as Cognitive Indicator



Objective: This chapter proposes EEG based ERP techniques as measures of brain function to allow healthy subjects to be distinguished from MCI patients. The work specifically explores potential test methods that can be used in conjunction with the ERP technique, potential feature measurements of an ERP leading to identification of distinguishing features and comparison of the performance of various machine learning classifiers. Topics covered include:

- Application of neuropsych cognitive assessment techniques such as n-back tests in conjunction with EEG/ERP measurement.
- ERP analysis and feature measurement methods within a test group of 32 volunteers including both healthy and MCI subjects,
- Application of classification techniques to measured features providing classifiers and feature sets that distinguish between the two groups.

6.1. EEG and ERP Analysis

The Event Related Potential (ERP) based on the Electroencephalogram (EEG) is a method where the cortical response to a stimulus is measured through repeated presentation of a stimulus with capture of the responses. The set of responses are then filtered for noise, outliers removed (such as eye blinks) and averaged after being time aligned to the stimulus event. The averaging has the effect of removing noise and signals due to activity that is uncorrelated with the stimulus, allowing the activity that is correlated to the stimulus to be observed.



The potential for ERP results to be used in conjunction with or in place of neuropsychological cognitive testing was studied through the comparison of the ERP testing in a group of 32 older adults where 15 were patients with MCI and 17 cognitively intact healthy older adults presented in Section 3.1.2. Cognition was tested using the MoCA [31]; RBANS subtests for immediate and delayed memory, attention, visuo-spatial and language [29] abilities; Trails Making Test A and B [30]. Participants were also asked to perform a number of cognitive tasks to elicit working memory relevant ERPs.

6.1.1. Cognitive Task: n-Back

The 1-back test is an example of an n-back task [38] that can be used with the ERP technique. In the 1-back task, the subject is presented with a sequence of symbols, and asked to press a button if the symbol presented is the same as the previous symbol and to do nothing if the symbol is different. This provides two different measures of the brain response that can be captured through an ERP, as the participant must make a decision and also make the response action or take no action.

For the 1-back assessment, participants were presented with 180 trials where they were asked to press the mouse key when the same number appeared consecutively. Numbers were presented for 1.0 sec with 1.7 sec between trials. Within the 180 trials, 60 trials required a button press and 120 trials did not. This resulted in data categorized into four cases based on whether a response action was required and whether the participant took the correct action or not.

The 0-back test uses a similar model to the 1-back except in this case, the participant is given a specific symbol at the start of the trial and is requested to press the button when that symbol is presented. This test presents a lower cognitive load as the target symbol is unchanging. The 2-back test is similar to the 1-back, except the participant is asked to press / not press the button based on the current symbol matching the symbol prior to the previous symbol. This represents an increased working memory cognitive load, as the participant must keep track of more historical symbols.



6.1.2. Cognitive Task: Verbal Recognition

Participants also performed a verbal recognition task that consisted of the presentation of a series of words on the computer screen that were presented one at a time and the participant indicates if they have seen this word previously (repeated word) or if it is a new word.

6.1.3. Cognitive Task: Go - No-go

In this test, the research subject is presented with a sequence of 2 symbols such as two different letters of the alphabet and is asked to press a button when a one symbol is presented (go: about 80% of time) and not press the key when the alternate symbol is presented (no-go: about 20% of time).

6.1.4. HC and MCI Group Comparison

The cognitive assessment scores were analyzed with IBM Statistical Package for the Social Sciences (SPSS) to determine if they showed statistical differences. The results shown in Table 6-1 demonstrate the differences between the two groups as there were significant differences between all neuropsychological test results such as mean Trails B time for MCI was 173.5 seconds compared to 76.4 seconds in the HC ($p < 0.001$). Within the cognitive tasks shown in Table 6-2, the behavioural scoring of the performance also demonstrated some of the differences in the two groups such as the HC group having more correct responses. The reaction time for HC was faster than the MCI patients in all 3 n-back conditions ($p = 0.002$).

To determine the initial potential for EEGs and specifically ERPs to indicate if the two groups were MCI or HC, the ERP amplitude and latency for the maxima near 200 msec, 300 msec and minima near 400 msec were analyzed using Brain Analyzer 2.0. ERP results in the n-back revealed: there was an increase in latency in the 200msec maxima as shown in Table 6-3 with a decrease in 400 msec post stimulus amplitude in MCI patients when compared to HC in all n-back conditions ($p = 0.04$).



Table 6-1: Summary of behavioural test scores for the HC and MCI groups showing mean score achieved and the SPSS p-value measure for the significance of the statistical difference between the two groups.

Test	MCI	HC	p
MoCA	22.3	27.4	0.00
RBANS Total	79.2	113.3	0.00
Immediate Memory	71.1	107.6	0.00
Delayed Memory	59.2	110.3	0.00
Visuospatial & Constructional	107.2	124.9	0.01
Language	89.8	99.3	0.06
Attention	93.5	103.1	0.09
Trails A (sec)	57.4	35.9	0.04
Trails B (sec)	173.5	76.4	0.00

Table 6-2: Summary of n-back accuracy and reaction time results for the HC and MCI groups showing mean score (standard deviation) achieved and the SPSS p-value measure for the significance of the statistical difference between the two groups.

	ACCURACY - %			REACTION TIME - msec		
	MCI	HC	p	MCI	HC	p
0-back	97.3 (2.4)	96.8 (2.9)	0.7	480 (69)	424 (43)	0.002
1-back	88.9 (7.6)	95.0 (2.0)	0.02	562 (70)	467 (41)	0.002
2-back	52.9 (14.6)	74.3 (11.5)	0.002	628 (77)	532 (44)	0.002

Table 6-3: Summary of P2 latency measure for the HC and MCI groups showing mean value (standard deviation) and the SPSS p-value measure for the significance of the statistical difference between the two groups.

Condition	MCI - msec	HC - msec	p
0-back	196 (24)	177 (24)	0.04
1-back	196 (21)	175 (21)	0.04
2-back	195 (25)	174 (25)	0.04



6.2. LDA Classification

The potential for the application of classification technologies to be used with the ERP test results as a means to distinguish between healthy and MCI participants was explored. This required the systematic analysis of the potential large space of options:

- Selection of EEG sensors to use within the 32 lead EEGs captured.
- Selection of the best test paradigm(s) to use from the N-back (0,1 and 2 back), Go/no-go, and Verbal recognition [38,39,133] test paradigms performed.
- Definition and measurement of potential features of the ERPs.
- Systematic exploration of these features to identify the features or sets of features that provided the best performance.

6.2.1. EEG Sensor and Test Paradigm Selection

The ERPs captured by the NeuroScan NuAmps system included the analysis by those tools of a number of ERP features across the set of EEG leads as reported in [133] that included a base set of features and signal leads. EEG signals included Cz, CPz, Pz, Fz and FCz and the features measured included P2, N2, P3 for N-back, P3 for Go/no-go, and N4 and Late Positive Complex (LPC) for Verbal recognition.

Adaptive LDA classifiers were trained using a leave one out train and test method for these ERP components and measurement paradigms. Given the goal was the identification of potential ERP signals and components, the behavioural performance measures associated with the test paradigms were excluded for this analysis.

For this group of MCI and HC, the N-back paradigm provided the best performance with an error rate of 15.6% using the 1-back P2 latency at Pz electrode with the CPz electrode showing the second best performance. The 0-back and 2-back did not improve on the 1-back results. For the Go/NoGo, the best performance was achieved with an error rate of 21.9% using P3 amplitude at Fz and FCz electrodes. The Verbal recognition paradigm provided the same performance



using two features including the N4 amplitude in repeated words at CPz. No combination of the features from the three tests improved results.

6.2.2. Expanded Feature Identification

Based on these results a focused analysis of CPz and Pz electrodes for the set of six ERP responses for the 1-back test paradigm for each participant was performed. This ERP set consisted of three ERP responses for each of the two electrodes for three cases: where a response was required, where no response was required and the combined response for these two cases.

A series of features were then calculated for the CPz and Pz responses for each participant:

- P1, N1 - latency and amplitude of the first maximum and minimum in the combined response respectively post stimulus.
- P2, N2 - latency and amplitude of the second maximum and minimum in the combined response respectively post stimulus.
- P3 - latency and amplitude of the third maximum in the combined response post stimulus.
- The amplitude for the response required and no response required ERPs at the same latencies for each of the above maxima and minima.
- A series of mean ERP response features were also calculated where the mean ERP signal is calculated for +/- 20msec window for each of the identified maxima and minima for each of the 6 ERPs for the participant.
- The ERP does not always show clear maxima and minima beyond 400 msec post stimulus; thus, a series of mean ERP response features for 400, 450 and 500 msec +/- 20msec window post-stimulus for each of the 6 ERPs for each participant was calculated.
- The correlation was calculated between the response required and no response required ERP for each of the two electrodes (CPz and Pz) providing a similarity measure.



These maxima and minima were located using a local peak [34] technique to identify a maximum/minimum within a window of samples. A local peak requires that the identified maximum/minimum must not be in the furthest right or left position within the window such that it must have samples on either side that were lower/higher than the chosen value.

From this set of results, a number of additional measures were calculated including:

- Latency and amplitude differences between each of the chosen maxima/minima and the subsequent minimum/maximum respectively. (i.e. N1 to P1).
- Latency and amplitude differences between each of the chosen maxima/minima and the subsequent maximum/minimum respectively. (i.e. N1 to N2).
- The magnitude of the ERP response may vary between individuals, so a set of relative magnitude measures was generated by scaling the amplitudes of the calculated maxima, minima and means for each participant by the maximum amplitude in that participant's post-stimulus ERP response.

This results in 314 features for each of the ERP responses to the 1-back test paradigm in addition to the 4 behavioural measures associated with the 1-back test.

6.2.3. LDA Classification

The development of classifiers was performed using Matlab's quadratic LDA classifier models and the 32 participants. Models were developed and tested using a leave one out trial model where the results of all 32 leave one out cases were averaged to provide an error rate for the model.

The Matlab "*sequentialfs*" and "*classify*" tools were used for this study to allow for the exploration of the measured features to identify the subsets that provide the best performance. The developed feature set includes many distinct features that are correlated with each other to varying known and unknown degrees, complicating the use of "*sequentialfs*", because subsets of features that are too



correlated can lead to failures within the classifier algorithms as they assume independence of the features.

The default implementation of "*sequentialfs*" creates optimal feature sets through evaluation of all features individually and the choosing the best individual feature. It then tries all remaining features with that feature to select a second feature. This procedure is repeated and features added until no further improvement occurs but this leads to the potential evaluation of subsets of features that are too correlated. The application of *sequentialfs* was enhanced to ensure that any feature when added to the candidate set that caused this issue was adaptively excluded. The exclusion list was generated uniquely and automatically for each trial so that once features were selected, any subsequent additional feature that caused issues was excluded.

6.2.4. LDA Classification - 1-Back ERP Measures Only

The Matlab classification tools were applied within the following cases:

- Individual feature performance - The performance for each of the generated features to classify the participants was measured.
- Incremental from best individual feature performance - Using the enhanced Matlab *sequentialfs* tool, the feature set was evaluated to determine the incremental classification value of adding additional features to the identified best individual feature.
- All 2 feature case evaluation - The performance for all pairs of generated features to classify the participants was measured to determine if a pair existed that provided better performance (i.e., a pair that excludes the best individual performing feature).
- Incremental from best feature pair performance - Using the Matlab *sequentialfs* tool, the feature set was evaluated to determine the incremental classification value of additional features to the identified best feature pair.

The exploration of all feature pairs is a required step to overcome a potential limitation of the *sequentialfs* tool to create optimal feature sets as it evaluates all



features and chooses the best individual feature and then trials all remaining features with that feature to select a second feature. This method will never find potential cases where an initial choice of two features that excludes the best individual provides better performance, and so all features pairs were evaluated.

Table 6-4: Summary of the performance of the LDA classifiers using an incremental search from best feature for only the ERP analysis features

Case	Features	Performance (%)			
		TN	TP	FN	FP
Single best feature	(1) No response required CPz P2 amplitude	82	67	33	18
Adding second feature to single best	Feature (1) plus: (2) Response required Pz N2 to P3 amplitude difference	82	73	27	18
Adding third feature	No improvement compared to features (1) and (2) only	n/a			
Adding fourth feature	Features (1) and (2) plus: (3) Response required CPz P1 to P2 latency (4) Combined CPz P1 relative amplitude	94	80	20	6

The initial classification results for the ERP trials using only the ERP derived features is summarized in Table 6-4 showing the analysis of the calculated features and identifying the best-performing feature. Once this feature was selected, the analysis was repeated combining this feature with all other features to identify pairs of features that improved the model's performance. This method was repeated again through the addition of a third feature and then a fourth feature. No improvement was found with the addition of a fifth feature.

The best single feature had 3 false positive (FP) / 5 false negative (FN) errors, indicating that almost one-third of the participants were incorrectly classified. A small improvement to 3 FP / 4 FN errors and 1 FP / 3 FN errors was achieved through the use of 2 and 4 features respectively. Combinations of 3 features did not improve on the performance of the 2 feature case.



All possible pairs of features were tested to determine if there was a pair of features that in combination outperformed the best performance found for two features including the best individual feature. This case now allows for the use of feature pairs that do not include the feature that provided the best overall performance individually. The result of this analysis is summarized in Table 6-5. One pair of signals that provides 2 FP / 2 FN errors was identified, which is an improvement over the 2-feature result shown in Table 6-4 and comparable to the 4-feature case in Table 6-4. The inclusion of the mean amplitude in the 450msec post stimulus range for the classification of individual participants aligns with the results in López Zunini et al. [133]. Classification models were then analyzed adding incremental features to the two-feature pair. The addition of a third feature to the best pair was analyzed and it improved performance to 2 FP / 0 FN errors. Additional features did not further improve performance.

Table 6-5: Summary of the performance of the LDA classifiers based an initial best pair search for only the ERP analysis features

Case	Features	Performance (%)			
		TN	TP	FN	FP
Best pair	(1) Response required Pz 450ms mean amplitude (2) Pz correlation	88	87	13	12
Best Pair + third feature	Features (1) and (2) plus: (3) Response required Pz N1 to N2 latency	88	100	0	12

6.2.5. LDA Classification - 1-Back Behavioural Measures Only

The results for the classification of the participants into participant groups (Healthy or MCI) using only the 1-back behavioural test scoring measures is summarized in Table 6-6. The model exhibits relatively poor performance, with classification errors of 2 FP/ 5 FN errors when the best 2 of the 4 available features are used and 1 FP / 7 FN when only 1 feature is used. No further improvement was found if 3 or all 4 features were used.



Table 6-6: Summary of the performance of the LDA classifiers using an incremental search from best feature for only the 1-back behavioural test scoring features.

Case	Features	Performance (%)			
		TN	TP	FN	FP
Single Feature	(1) 1-back Omission errors	94	53	47	6
Two Features	Feature (1) plus: (2) 1-back Accuracy	88	67	33	12

6.2.6. LDA Classification - 1-Back ERP and Behavioural Measures

The potential for the combination of the 1-back behavioural and ERP analysis features combined is summarized in Table 6-7 using an incremental search starting from the base single performing feature, as performed with the ERP only case. These results show that 1 FP / 3 FN errors was achieved using 3 features which is poorer than the best found in Table 6-2. No additional performance improvement was found using a fourth feature.

Table 6-7: Summary of the performance of the LDA classifiers using an incremental search from best feature for the 1-back behavioural scoring and ERP analysis combined

Case	Features	Performance (%)			
		TN	TP	FN	FP
Single feature	(1) 1-back Omission errors	94	53	47	6
Two features	Feature (1) plus: (2) Response required CPz N1 to P1 amplitude difference	100	67	33	0
Three features	Features (1) and (2) plus: (3) Response required Pz P2 to N2 latency	94	80	20	6

Again a search of all feature pairs using the combined feature set was performed. This search is summarized in Table 6-8: the best pair of features from the combined set yielded 1 FP / 1 FN errors, and through an incremental search from this pair, a



third feature was identified that yielded 1 FP / 0 FN errors. The inclusion of additional features beyond 3 did not improve the error rate.

Table 6-8: Summary of the performance of the LDA classifiers based on initial best pair search for the 1-back behavioural scoring and ERP analysis features combined

Case	Features	Performance (%)			
		TN	TP	FN	FP
Best Pair	(1) 1 back Response Time (2) CPz correlation	94	93	7	6
Best Pair + third feature	Features (1) and (2) plus: (3) Response required Pz P1 to N1 latency	94	100	0	6

6.3. Alternative Classifiers

The LDA classifier based on discriminant analysis is one of many classification methods that could be applied. Other classifiers include those based on decision trees (DT), logistic regression (LR), support vector machines (SVM), k-nearest neighbour (KNN) and ensemble classifiers (EC). The Matlab "*classificationLearner*" provides 23 different classifiers as shown in Table 6-9 and allows their relative performance to be compared. The features identified in Table 6-5 were used with the "*classificationLearner*" tools with initial focus on the use of only the best pair of features identified.

The results for the 23 classifiers have been summarized on Table 6-10 where 4 classifiers in addition to LDA (Quadratic) are shown. These classifiers are the only cases from the 23 that provided similar or better performance than the LDA (Quadratic) classifier. The table shows the specificity and sensitivity results for each of the classifiers and in the case of classification errors shows the participant(s) that were incorrectly classified. The results for LDA (Quadratic) and Medium Gaussian SVM are identical with the same error rate and the errors occurring on the same participants. Weighted KNN had similar performance also but the False Negative error occurred on a different participant. Logistic Regression performance was also similar but the errors differed with 2 false negative and no



false positives. In all of the above cases, participant 7 was misclassified. Fine KNN performed the classification with no errors.

Table 6-9: Summary of the 23 classifier models

Classifier Family	Classifier Model
Decision Trees	Complex Tree
	Medium Tree
	Simple Tree
Discriminant Analysis	LDA (Linear)
	LDA (Quadratic)
Logistic Regression	Logistic Regression
Support Vector Machines	Linear SVM
	Quadratic SVM
	Cubic SVM
	Fine Gaussian SVM
	Medium Gaussian SVM
	Coarse Gaussian SVM
k-Nearest Neighbour	Fine KNN
	Medium KNN
	Coarse KNN
	Cosine KNN
	Cubic KNN
	Weighted KNN
Ensemble Classifiers	Boosted Trees
	Bagged Trees
	Subspace Discriminant
	Subspace KNN
	RUSBoosted

One figure of merit for the performance of a classifier is the area under the ROC curve with an area of 1.0 representing perfection, while an area of 0.5 implies performance no better than random. An example plot of the ROC curve for the LDA (Quadratic) classifier is shown in Figure 6-1 and the area calculations for each of



the classifiers is shown in Table 6-10. As expected, the results for this figure of merit are very similar for the classifiers that have similar error results.

Table 6-10: Sensitivity and specificity data for each of classifier using the best pair of features from Table 6-8. For false positive and false negative cases, the incorrectly classified participant number is noted in (). Table also shows the area under the ROC curve for the classifier.

Classifier	True Positive (MCI) N=15	True Negative (HC) N=17	False Positive	False Negative	Area Under ROC
LDA (Quadratic)	14	16	1 (7)	1 (25)	0.97
Fine KNN	15	17	0	0	1.00
Weighted KNN	14	16	1 (7)	1 (18)	0.99
Medium Gaussian SVM	14	16	1 (7)	1 (25)	0.98
Logistic Regression	13	17	2 (7,14)	0	0.93

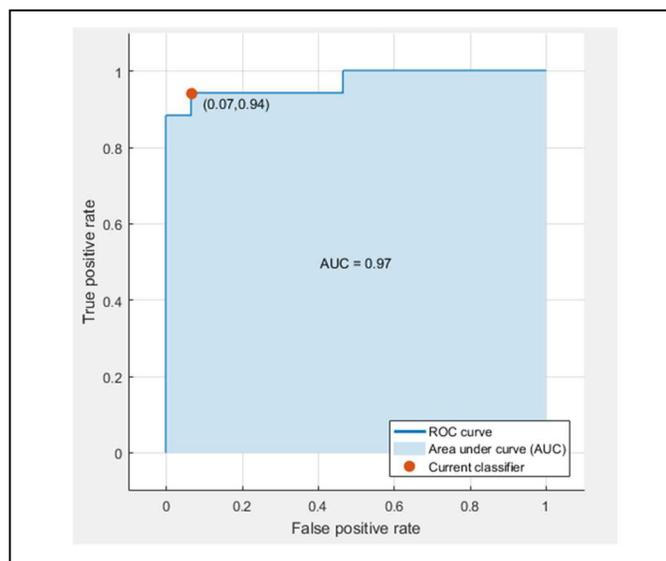


Figure 6-1: Example ROC showing the area under the curve calculation for LDA (Quadratic) classifier.

The classification results for the LDA (Quadratic) classifier are shown as a scatter plot in Figure 6-2 where the 32 participants are shown against the measures



for the best two features in Table 6-8. A split between the two classes is shown with the HC participants in the lower left triangle and the MCI participants in the upper right triangle. The close proximity of the two groups is clear, as are the two classification errors for this classifier. The figure shows the 4 participants (7,14,18,25) that were misclassified by one or more of the classifiers and all of these as expected reside near the transition between the two triangular regions.

Each of the classifiers produces decision scores to provide an indication of the relative likelihood that the data represents each of the two classes. For LDA, KNN and LR, the score is presented as percentage likelihood with the larger value indicating the predicted class and values with larger spread indicate confidence in the decision. For SVM, the scores are two values with differing sign, the positive value indicates the predicted class, and the magnitude provides a measure of the confidence in the prediction. The score results for each of the classifiers are shown in Figures 6-3 to 6-7.

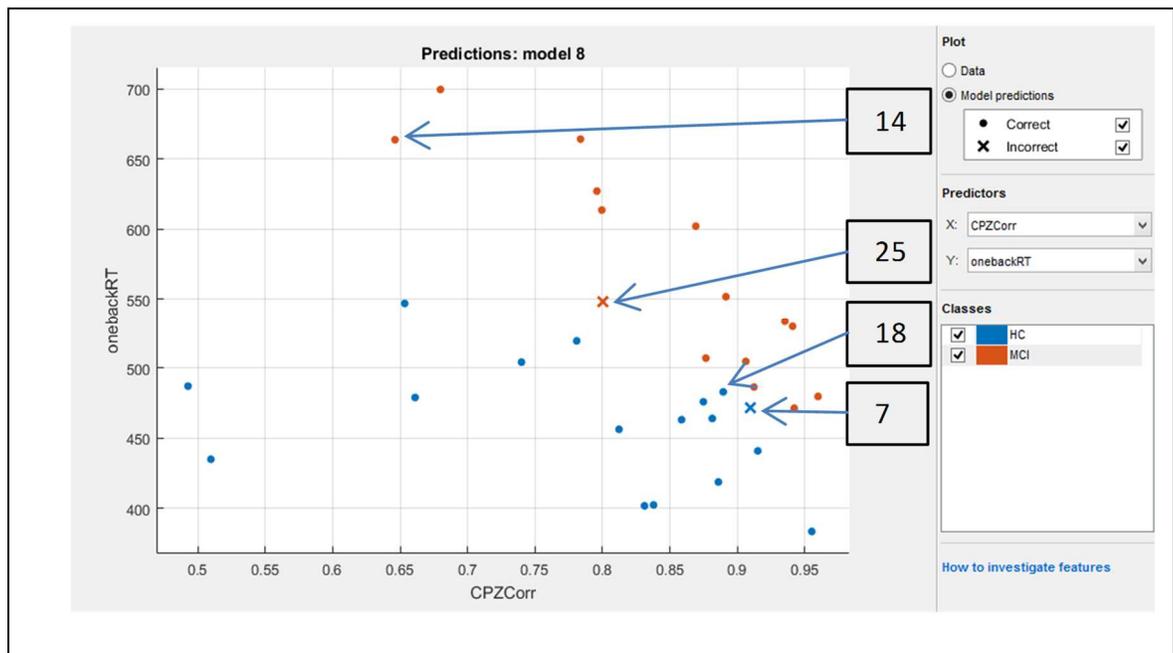


Figure 6-2: Example Scatter Plot for LDA (Quadratic). Orange - classified as MCI, Blue - classified as HC. Dot - classified correctly, X - classified incorrectly. Figure identifies the 4 participants from Table 6-5 that were incorrectly classified by one or more classifier.

The scores for the LDA (Quadratic) classifier in Figure 6-3 shows in general the large differences in the scores with the key exceptions being the two decision errors



with participants 7 and 25, with a close but correct score for participant 17. In Figure 6-4, the scores for Fine KNN results are shown with this classifier providing a binary score of 0 or 1. The Weighted KNN scores shown in Figure 6-5 show participants 7, 17 and 25 as the three with scores closest to even with the same errors as LDA. The Medium Gaussian SVM scores are shown in Figure 6.6 where the four participants with scores closest to even are 7, 13, 17 and 25. The LR scores are shown in Figure 6.7 and again these are binary scores with the two errors on participants 7 and 14. The introduction of the third feature from Table 6-8 did not significantly change the results for any of the alternate classifiers.

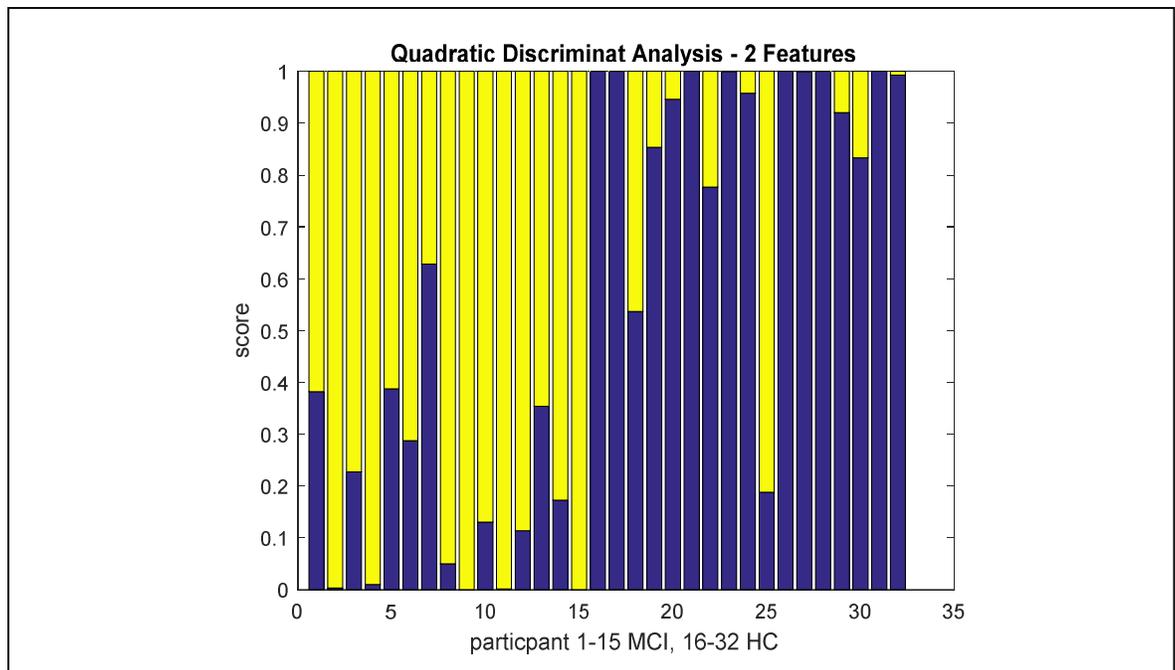


Figure 6-3: Decision scores for LDA (Quadratic) using best 2 features from Table 6-5 for each of the 32 participants. Blue - "Healthy" score and Yellow - "MCI" score with participant classified into the class with the larger score. 15 MCI subjects on the left and 17 HC subjects on the right.

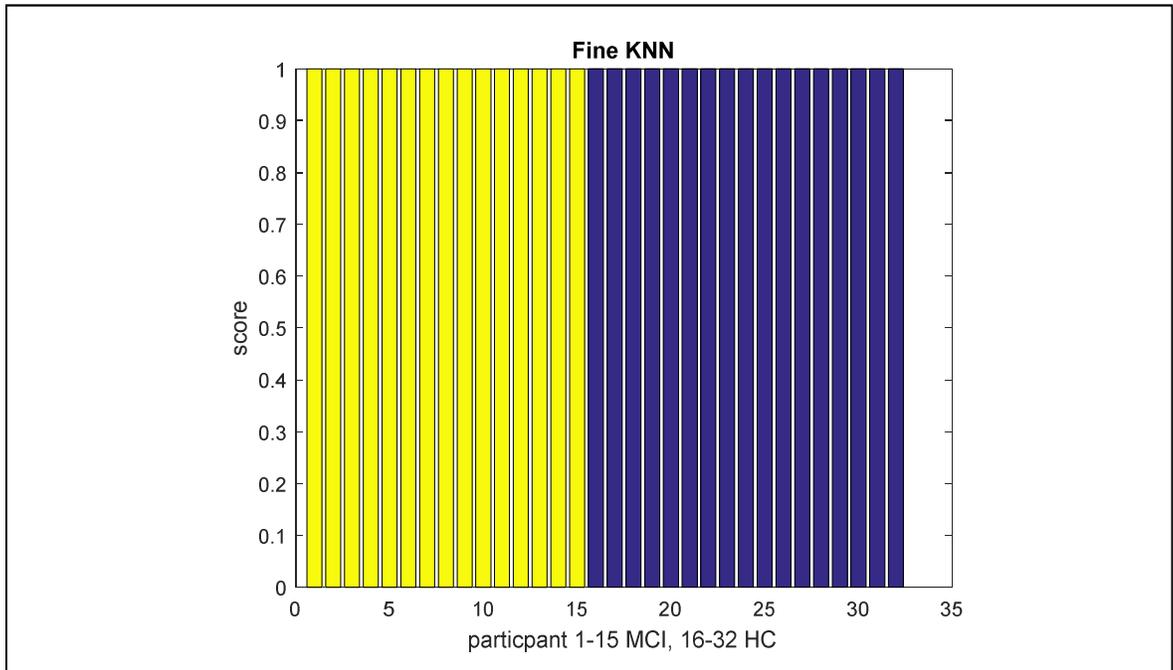


Figure 6-4: Decision scores for Fine KNN using best 2 features from Table 6-5 for each of the 32 participants. Blue - "Healthy" score and Yellow - "MCI" score with participant classified into the class with the larger score. 15 MCI subjects on the left and 17 HC subjects on the right.

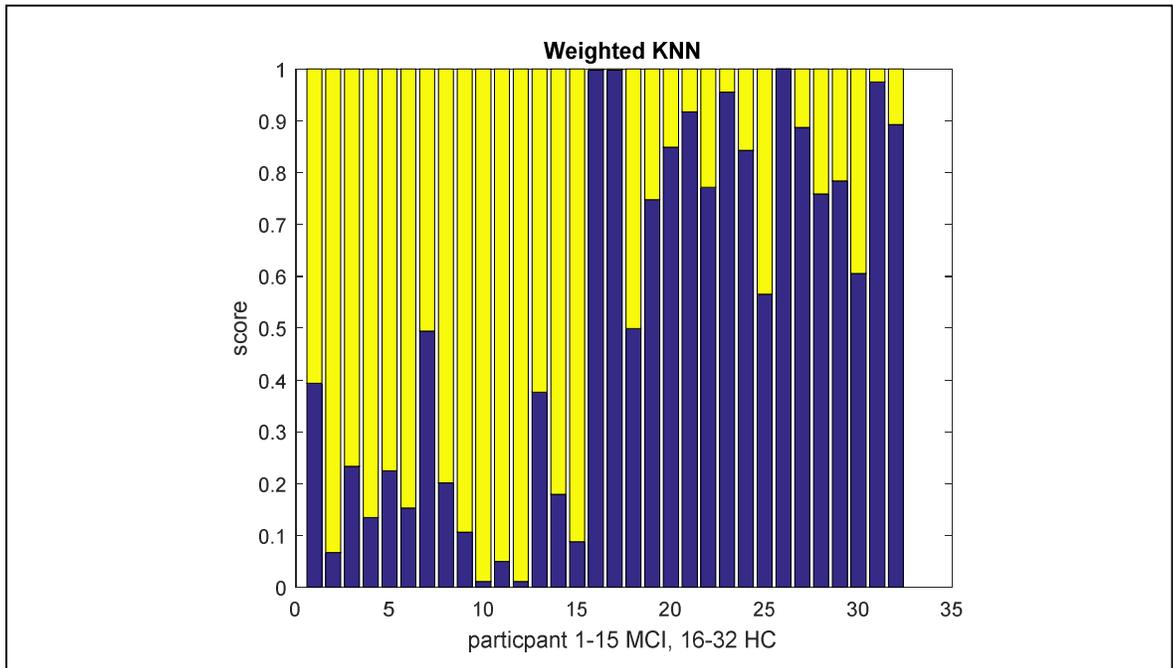


Figure 6-5: Decision scores for Weighted KNN using best 2 features from Table 6-5 for each of the 32 participants. Blue - "Healthy" score and Yellow - "MCI" score with participant classified into the class with the larger score. 15 MCI subjects on the left and 17 HC subjects on the right.

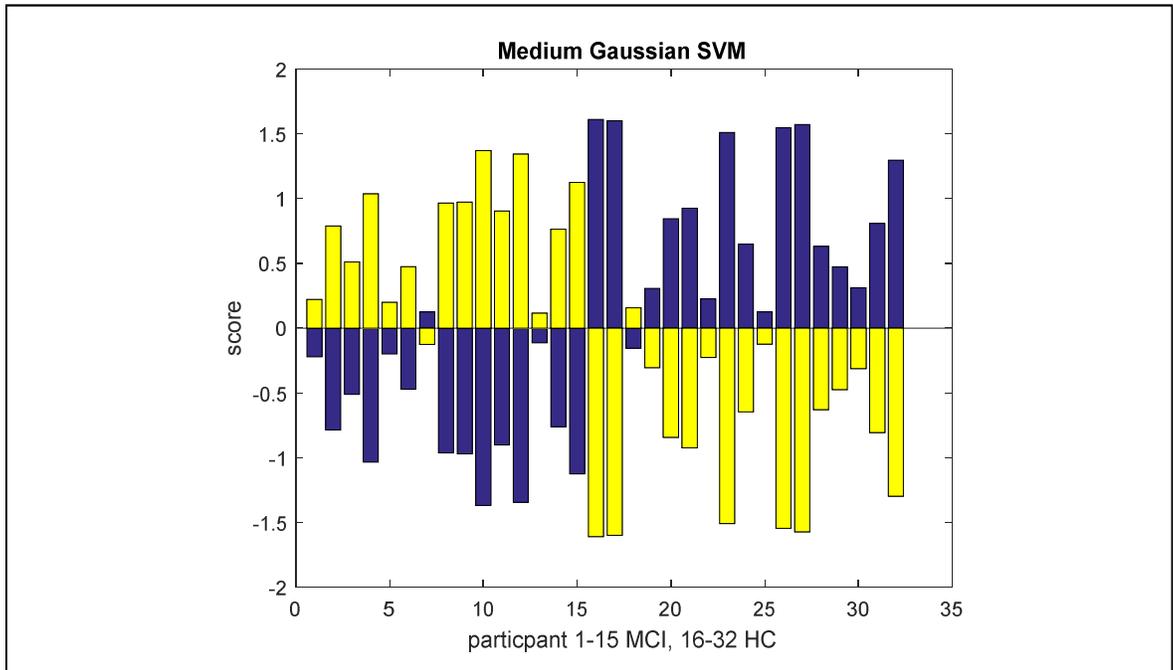


Figure 6-6: Decision scores for Medium Gaussian SVM using best 2 features from Table 6-5 for each of the 32 participants. Blue - "Healthy" score and Yellow - "MCI" score with participant classified into the class with the positive score. 15 MCI subjects on the left and 17 HC subjects on the right.

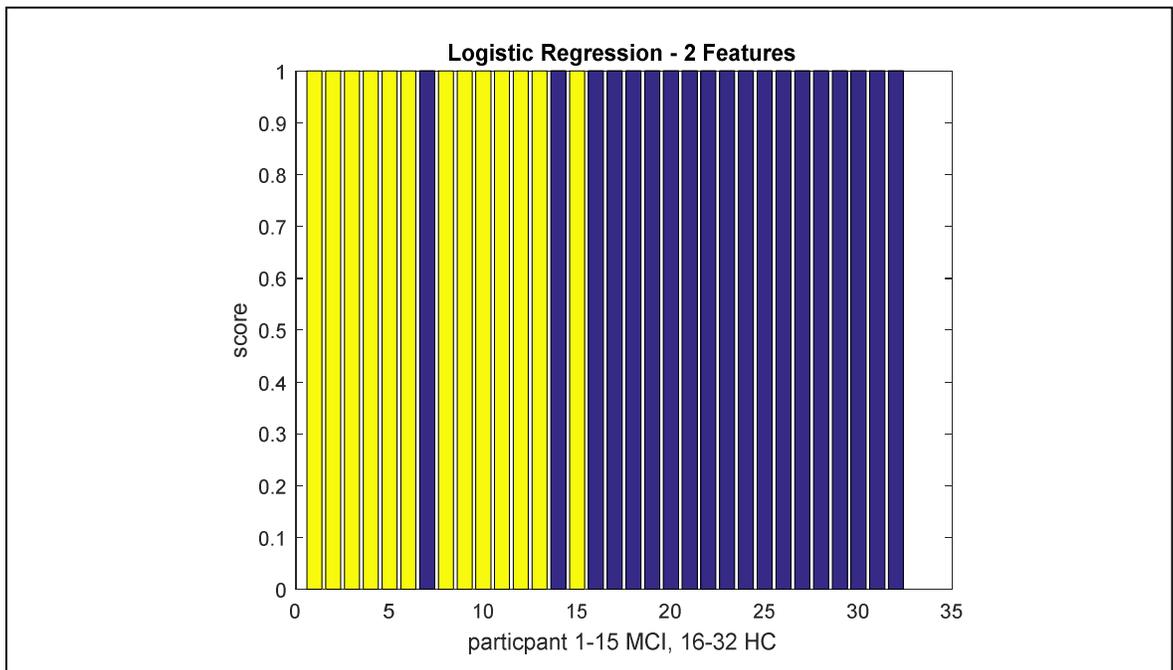


Figure 6-7: Decision scores for Logistic Regression using best 2 features from Table 6-5 for each of the 32 participants. Blue - "Healthy" score and Yellow - "MCI" score with participant classified into the class with the larger score. 15 MCI subjects on the left and 17 HC subjects on the right.



6.4. Discussion

This work explores the potential for ERP responses to help with the identification of MCI in patients. It found that measures of the participant's ERP responses along with their associated behavioural measures during a 1-back task outperformed the results from the behavioural measures only and also outperformed the results ERP and behavioural measures for the other two N-back tests, Go/No-Go and Verbal Recognition tests.

The work further explores the development of classifiers to distinguish HC and MCI participants with the analysis of 314 ERP measurement features and 4 behavioural measures for the 1-back test. Using LDA classification, the 3 features that provide the best classification are identified with the 3 providing 1 FP / 0 FN errors and using only 2 features, an error rate of 1 FP / 1 FN was found. The potential for other classifiers to improve the performance was then analyzed where 23 classifiers were tested with the best features with 4 classifiers providing similar or better performance than LDA with Fine KNN identified as a classifier that performed with no errors.

Classifiers and machine learning techniques identified a decision rule based on the two features represented by a diagonal decision rule between the two participant groups that is not possible with a single feature. The decision rule for the classifier shows that healthy subjects exhibited faster response times and a larger difference in their ERPs for the response required and not required cases.

The response time measure is related to processing speed for the participant and it would be expected that healthier subjects should be able to respond faster. In a clinical cognitive test, the participant is asked to do a timed task such as draw an analog clock once. This is scored once by the observing clinician based on time and accuracy. Should the participant make an error unrelated to their ability, this could lead to diagnostic consequences. The 1-back cognitive test uses a simpler cognitive task: click or not-click a mouse button based on a presented number, and the test is repeated and click speed measured each time by the test computer. The simpler task and repetition allows the participant to focus and perform at their ability



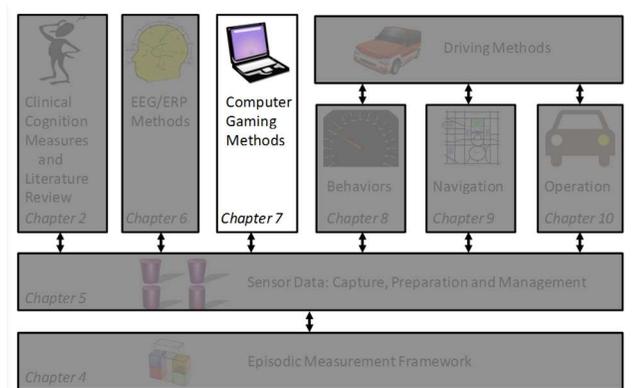
level with averaging ensuring effects of a single outlier or natural variation in a single result measure reduced. The participant also does not have the stress of direct observation and measurement by the clinician.

The ERP is a measure of the cortical response signal for the participant with the two cases within the 1-back test requiring different actions by the participant; with a mouse click action required in only one case. The cortical signal measured by ERP is an aggregate of the cognitive pathways associated with the processing for each of the two response cases. Physiologically, one of the impacts of cognitive decline and reduced brain ability is a reduction in the diversity of cortical pathways within the brain. The proposed correlation measure indicates that the ERP for the response required and not required cases for an individual have more similarity for the MCI participants than HC. This could be an indication that healthy brains have more options for routing information to perform a task through the use of different pathways for the two response cases. On the other hand, MCI causes patients have higher similarity in the ERP indicating less diversity in the brain response between the two cases.

Future studies of 1-back testing with ERP measurement should explore the extensibility of these results through study of a larger population of both MCI and HC control participants with the possible inclusion also of mild dementia patients. Studies are also required that explore the variation in these measures for an individual, such as day to day or time of day, or impact of fatigue or caffeine, while also managing for practice effects that may also occur.



Chapter 7: Games as Cognition Measure



Objective: This chapter explores methods for the measurement of cognition through computer based games. Computer gaming represents a method for cognition measurement that can be easily deployed to patient's homes allowing them to easily perform between appointments. The chapter explores the features and capabilities required in games. Topics covered include:

- The features and capabilities of games designed for the measurement of cognition with older adults is presented including the need to be interesting, relevant and at an appropriate ability level for the group. Ability is supported through hint systems to support play.
- The instrumentation of the games to provide the needed sensor information for cognition measurement is presented.
- The design of two games developed for this work is presented.
- The functional validation of the games with volunteer students and cognition distraction presented showing the measures that are possible with the games.
- The results from pilot trials of the games with MCI patients is presented.

7.1. Method

Two custom games were developed for use with MCI patients as both a mechanism to stimulate cognitive ability and to detect cognitive change over time of use. There are many games that could be used including card games (solitaire,



memory) or word jumbles [54, 55]. CWG (Word Search) and CSG (Sudoku) were chosen and developed as research games because they always have a solution whereas many card games have luck of the deal and high potential for no winning solution. Word search games focus on language centers and the brain functions to search the grid for the hidden word. Sudoku's use of numbers and logic makes use of different brain centers.

7.1.1. Carleton Word Search Game (CWG)

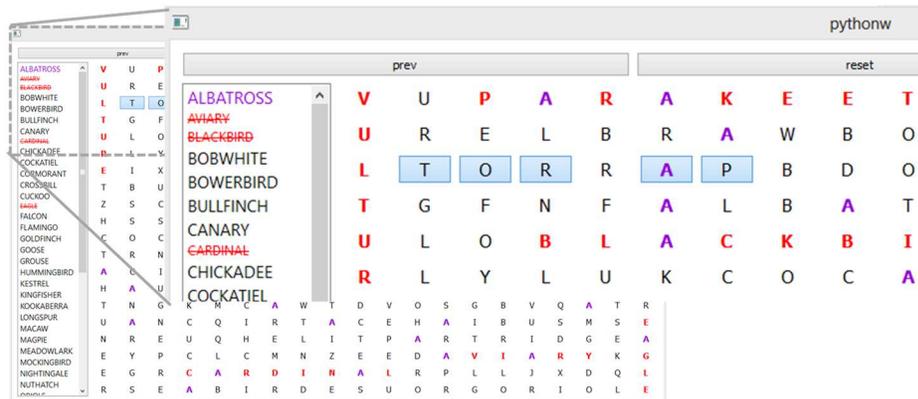


Figure 7-1: Example CWG game board showing word list on left, letters being selected for a new word in boxes, letters in previously found words in red and letters included within a hint (letter A) for the next word in magenta.

The game presents a grid of letters with the hidden words along with the list of words. A screen shot is shown in Figure 7-1. Key features of CWG include:

- Hint system: It is important that MCI patients are able to make progress to avoid frustration. After no progress for a set time limit, the game highlights all letters that match the first letter of the next word to be found. It will then highlight the second and third letters in turn based on the timer.
- Detailed record of user activity: In addition to overall time to complete, a detailed time log for the following user activities is created:
 - All mouse movements
 - Letter selection and de-selection
 - Words found
 - Hints provided



7.1.2. Carleton Sudoku Game (CSG)

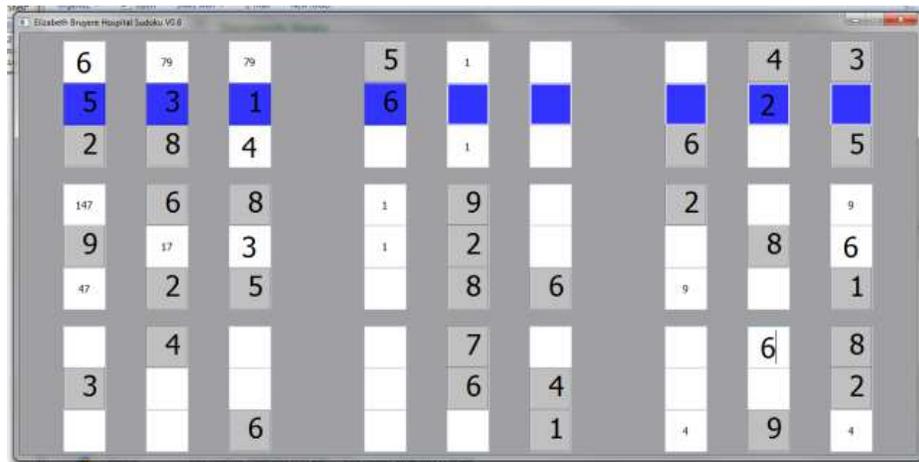


Figure 7-2: Example CSG game board showing given numbers in grey squares, solution numbers in large font with white background, user pencil marks in small font and a hint for a cell that can be solved highlighted in blue.

The game presents the user with a game grid that has some prefilled cells. An example game screen is shown in Figure 7-2. Key features of CSG include:

- Pencil Marks: A user can enter a single value in a cell as the solution value but many game players need to make notes in cells of potential values as they work the logic of the puzzle. CSG allows both solution values and pencil marks.
- Hint System: Two types of hints are provided by CSG, if a user has any cells where the entered number is incorrect, the game will highlight those cells in red after an adjustable time limit of no progress. If there are no errors, the game will instead provide a solution hint by highlighting a row, column or 3x3 square where there is a cell that can be solved. Hints will cycle through all three hints for a given cell so that once all are shown; the hinted cell is uniquely located for the user.
- Detailed Record of User Activity: The specific measures included by CSG in the time log include:
 - All mouse movements
 - Solution number entries and erasures
 - Pencil mark entries and erasures
 - Hints provided including hint type



7.2. Experimental Results - Students

The CWG and CSG were operationally validated with volunteers and then used within a pilot study of MCI patients. The volunteer study allowed the game functionality and the associated measurement system to be validated. It also provided data for the development of analysis algorithms and through the use of cognitive distraction, data with varied cognitive performance for the same individual. The study with MCI patients provided pilot data on the actual use and performance of the games with a target population.

7.2.1. Carleton Word Search Game (CWG)

Operational validation of the games was completed through a study with healthy students and staff as summarized in Section 3.1.1. Figure 7-3 shows an example timeline plot for one participant showing many key aspects of game play. The selection and de-selection of letters by the user including clearly distinguishing selection of letters that are in hidden words and those that are not (errors). The log also shows when words were found and these allow exploration of the variation in a given user's play. The last part of the figure shows when the user received hints.

Figure 7-5 plots a portion of the mouse track for CWG play showing how the mouse is used as a visual aid for the user as they search the grid, including a word being found (loop) and selected (line) near (925, 200). Application of signal processing techniques to these traces could identify typical search strategies along with regions on random activity. Table 7-2 shows example results for 2 players and the student group. These show a number of the measures of game play that can be derived from the game log, including various measures of user activity such as average time between each found word indicating search skill, and how quickly the letters in a word are selected measuring motor control. These preliminary results show the differing performance of the users, such as their use of the hint system and the number of errors made. The ability of the measures to show differing performance is indicated by the lower number of words found and the larger average time between words found when distracted. The word rate standard



deviation indicates the expected variation in rate as players find obvious words quickly and then slow as they search for other words.

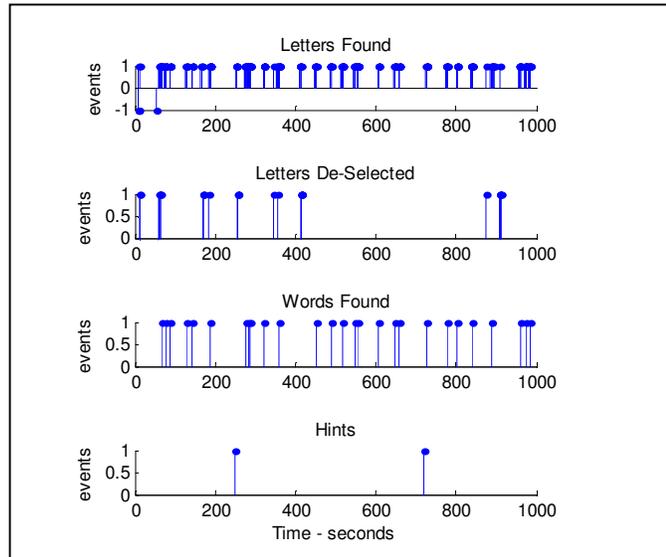


Figure 7-3: CWG log trace showing the record of game play. Letters found shows the selection of letters by the user with 1 = letter in hidden word, -1 = letter not in hidden word (error). Letters de-selected shows letters being unclicked. Words found shows the time of each found word. Hints shows when they were provided.

Table 7-2: CWG results showing measures of game performance calculated from the game log for two example users and the overall group (n=17).

User	Distraction	Words found (sec/word)		Letter selection (sec/letter)		Game actions (sec/action)		Words found	Hints	Errors	Errors fixed	Letters found	Extra Letters
		mean	st dev	mean	st dev	mean	st dev	count	count	count	count	count	count
1	N	27.8	20.2	0.48	0.37	3.12	9.22	26	1	2	2	251	7
1	Y	30.2	30.1	0.66	1.29	2.80	8.06	26	1	9	9	201	17
2	N	22.8	17.1	0.38	0.34	2.12	6.66	41	1	0	0	282	3
2	Y	21.4	15.8	0.37	0.16	2.23	7.30	36	1	1	1	245	0
all	N	38.0	13.1	0.61	0.23	3.5	1.2	24.4	3.3	3.9	3.9	191	15.6
all	Y	39.4	11.8	0.55	0.17	3.9	1.0	22.5	3.8	2.7	2.6	169	8.8



Table 7-3: CSG results showing measures of game performance calculated from the game log for two example users and the overall group (n=17).

User	Distraction	Solution Entry (sec/number)		Game actions (sec/action)		Correct Entries	Entry Errors	Errors Hints	Error Hint Used	Regular Hints	Regular Hint Used
		mean	st dev	mean	st dev	count	count	count	count	count	count
1	N	10.68	10.36	5.25	8.70	5	10	0	1	1	5
1	Y	19.53	22.12	6.41	9.61	6	0	0	1	0	6
2	N	12.46	14.89	6.96	16.58	2	2	0	6	0	2
2	Y	10.33	10.85	4.82	7.96	12	0	0	0	0	12
all	N	23.0	22.6	7.2	4.1	4.9	5.2	0.9	5.4	0.8	4.9
all	Y	25.3	28.0	7.9	5.6	4.8	4.8	0.9	3.5	0.5	4.8

7.2.2. Carleton Sudoku Game (CSG)

Figure 7-4 shows a timeline plot of CSG game play. The entry and erasure of solution entries shows the user's progress towards the solution and the log indicates if each entry is correct or incorrect. The Pencil Mark entry information provides an indication of activity by the player as they work through the logic of the puzzle. The last portion of the figure shows when the user received hints and whether the hint was for an incorrect entry or showing a cell that could be solved.

Table 7-3 shows summary CSG results for 2 players and the student group. Analysis of the game log information for the player action following a hint related to the hint such as a player being shown an incorrect cell and choosing to correct it (hint used) or working elsewhere in the grid. The ability of the measures to indicate performance differences when distracted is shown in the increased time between actions. Figure 7-5 shows a portion of the mouse track for the game play and this shows how the mouse is moved around the grid to aid focus and between the cells as results are entered.

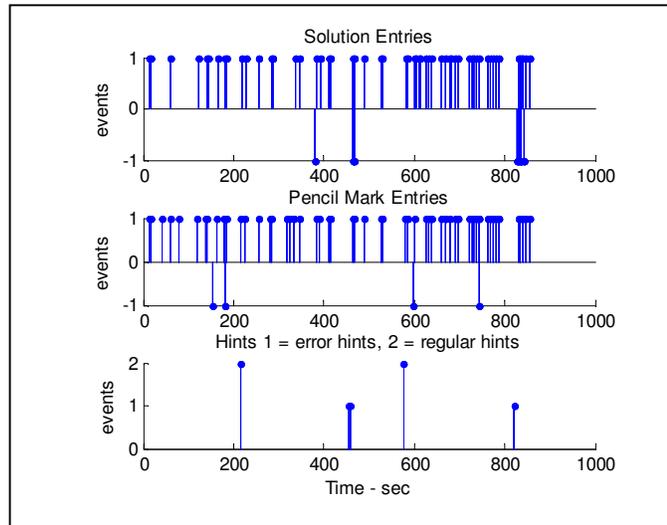


Figure 7-4: CSG results showing all the actions of game play.
 Solution entries: 1=correct entry into cell -1= incorrect
 Pencil mark entries shows: 1 = pencil mark entry -1 = pencil mark erase
 Hints shows type: 1 = error hint 2 = solution hint

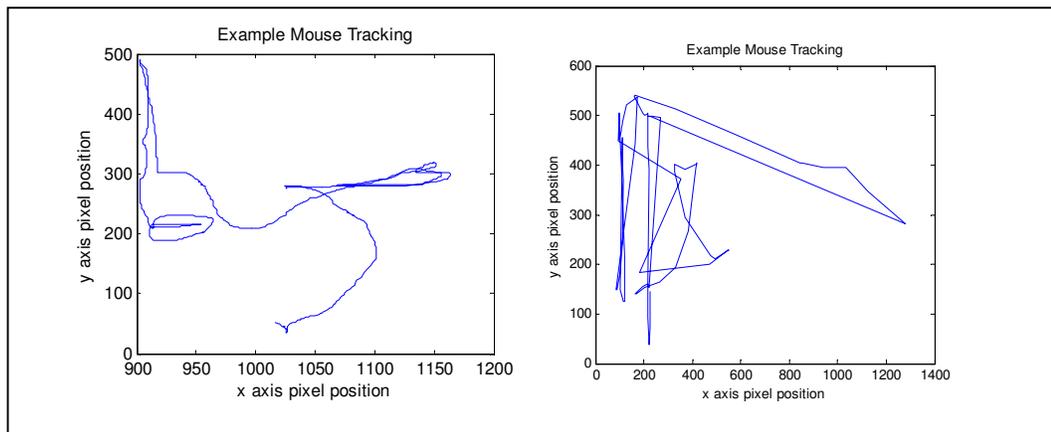


Figure 7-5: Left: CWG mouse track for 180 to 200 seconds in Figure 7-3.
 Right: CSG mouse track for 150 to 250 seconds in Figure 7-4.

7.3. Experimental Results - MCI Patients

The 15 MCI participants summarized in Section 3.1.2 were split into two groups with 7 completing a series of sessions playing CSG and CWG while the other 8 played an online cognitive game (BrainHQ). In this section the performance metrics plots of two of these participants will be summarized. Specifically, the results for two players (P1, P2) are shown. In the study group, six of the participants were cognitively stable over the over the trial period and showed stable play at their given



ability level. An example participant (P1) from this group is presented in this work in contrast to the one participant that suffered significant cognitive decline (P2) over the trial leading to conversion from MCI to Dementia post trial. This work show how this methodology and metrics may be used to monitor changes in participant's cognitive ability.

7.3.1. Carleton Sudoku Game (CSG)

Figure 7-6 shows the bar plots of the CSG game completion time metric for two participants over a 9-week study with 3 sessions per week. A lower mean game completion time indicates that the participant was able to finish the games more quickly and hence represents better performance. Conversely longer completion time indicates that the participant took more time to complete the game. It is clear from the figure that participant P1 tended to complete the games more quickly that participant P2.

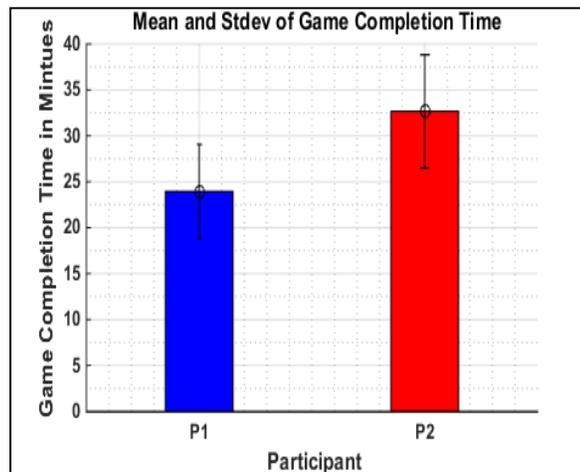


Figure 7-6: Mean and standard deviation of Game Completion Time metric for Sudoku game.

Figure 7-7 shows the bar plots of the Error Rate metric for completed games for both participants that indicate that the participant P1 was able to solve puzzles with fewer errors than P2. The higher standard deviation of the Error Rate for P1 indicates that P1's performance varied to a larger degree over the trial period. It is



clear from Figure 7-6 and 7-7 that the participant P1 outperforms participant P2 in both of the CSG performance metrics.

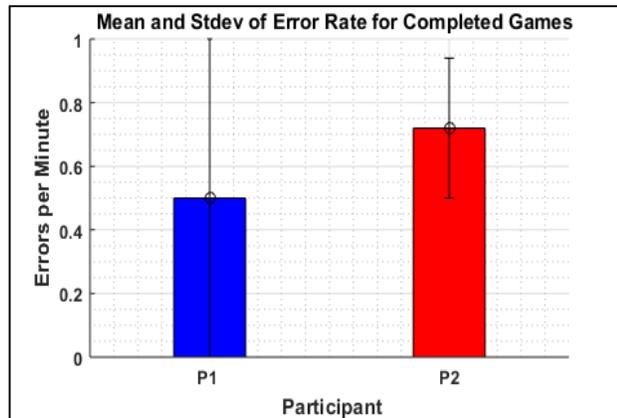


Figure 7-7: Mean and standard deviation of Error Rate metric for Completed Sudoku Games.

7.3.2. Carleton Word Game (CWG)

Figure 7-8 shows the bar plots of CWG game completion time metric for both of the participants showing that for this metric, both of the participants, P1 and P2, were almost same. Figure 7-9 shows a plot of the number of words found per game metric for both participants. Participant P1's mean for words found per game is 20.75 while for participant P2 it is 4.48. There is significant performance difference in number of words found metric between the participant P1 and P2. The participant P1 outperforms P2 by a factor of 4.

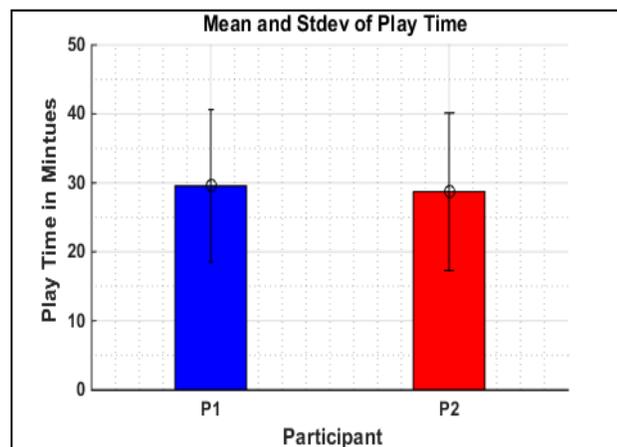


Figure 7-8: Mean and standard deviation of Play Time game metric for word search game.



Figure 7-10 shows the number of words found per minute metric for P1 and P2 and it is again clear that participant P1 performed significantly better compared to participant P2 on word rate metric.

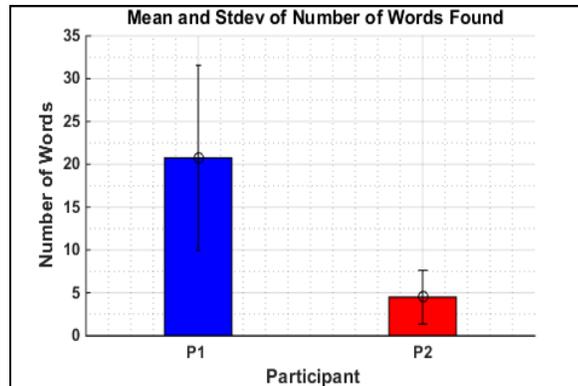


Figure 7-9: Mean and standard deviation of Number of Words Found per Game metric for word search game.

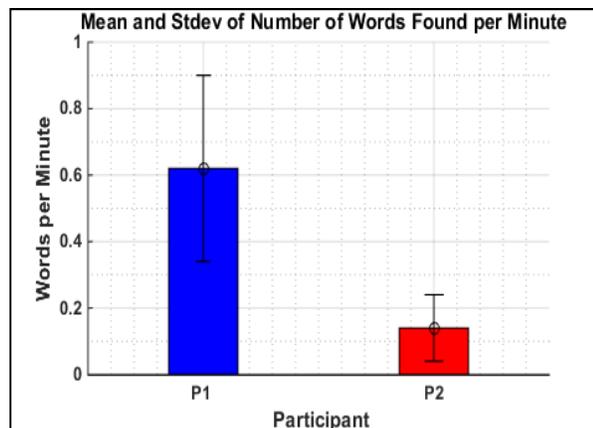


Figure 7-10: Mean and standard deviation of Number of Words Found per Minute (Word Rate) metric for word search game.

7.3.3. MoCA and RBANS Cognitive Test Scores

The participants P1 and P2 both underwent MoCA and RBANS cognitive tests before start of the trial to establish baseline cognitive ability. Figure 7-11 shows the test scores for MoCA test and Figure 7-12 shows the test scores for RBANS test. P1's results are just below the threshold to be considered healthy and not MCI for both MoCA and RBANS tests (Healthy is 26 or higher for MoCA and mean of 100+/- 15 for RBANS). These test results show that participant P2 has more severe cognitive impairment than the participant P1 that is clinically significant and correlates with the lower performance observed in the game play for participant P2.



Figure 7-11: MoCA test scores for participant P1 and P2.

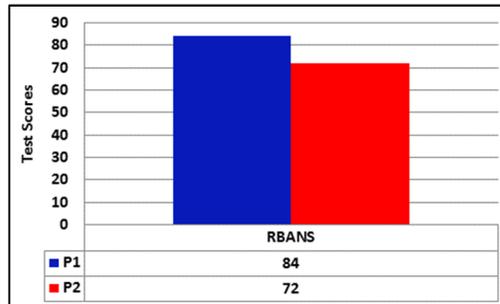


Figure 7-12: RBANS test scores for participant P1 and P2.

7.3.4. Participant during cognitive decline

Participant P2 coincidentally underwent significant cognitive decline during the period that ultimately resulted in the participant being re-diagnosed as full dementia shortly after the trial. This presented a unique opportunity to explore the changes in game play performance during this transition. Figure 7-12 shows the play time for CWG while Figure 7-13 shows the number of words found and this shows a negative trend in play with a drop off in the last few sessions.

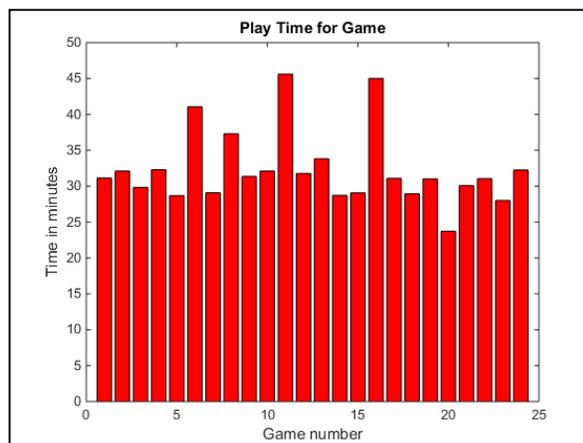


Figure 7-12: Play Time game metric for word search game for participant P2 for each of the games played. Games shown in order played over the trial period.

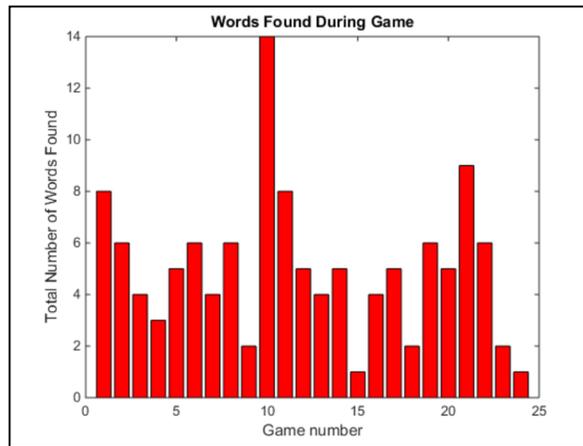


Figure 7-13: Number of words found for word search game for participant P2 for each of the games. Games shown in order played over the trial period, no games were completed.

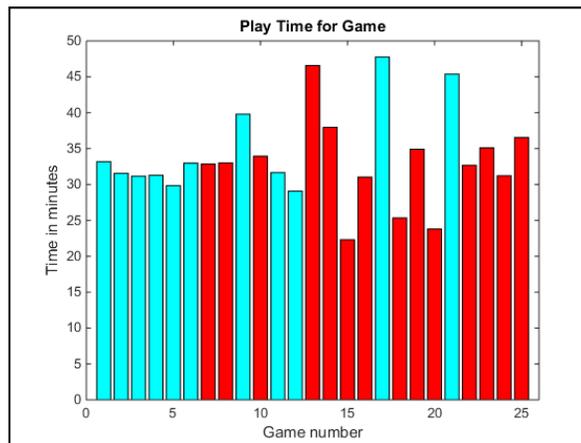


Figure 7-14: Game play time metric for Soduko games for participant P2 over the trial period. Games shown in order played over the trial period. Cyan - game not completed, Red - game successfully completed.

Figure 7-14 shows the CSG game completion time plot over the trial period while Figure 7-15 shows the error rate for the participant over the same sessions. The plots show that in the first few sessions, the user had a learning curve with the game and did not complete a game until game 7 followed by more regular success. Although in the early sessions, the puzzles were not solved, the error rate in Figure 7-15 is at its lowest level for this player. In later sessions, the error rate has a general trend towards higher levels and the user demonstrated an inability to recover from errors in cases when the puzzle was not completed. This suggest that



the participant was not able to use the hint system or otherwise recover from the error and instead continued to make additional errors compounding the problem.

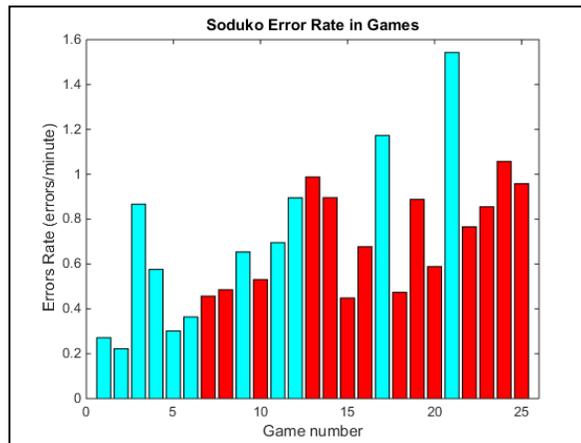


Figure 7-15: Number errors made per minute for Sudoku games for participant P2 over the trial period. Games shown in order played over the trial period. Cyan - game not completed, Red - game successfully completed.

7.4. Discussion

This section presents the design and operational validation of two games that have been designed for use with cognitively impaired subjects. The paper shows the key aspects required in game design for this user group including a hint system so users can be assisted to ensure they progress in the games. The game log that is included in both games provides a detailed record of the user interactions that will allow insight into the player game play actions, including overall performance and detailed performance for both logic phases and fine motor control phases of the solution.

The performance of the participants from the Bruyère Memory Program was compared using the proposed performance metrics for Sudoku and word search game. Both the cognitive test results and computer based game results show that the participant P2 is cognitively more impaired than participant P1 and provided indications of the changes being experienced by participant P2. The results indicate that computer based games and its metrics data may be used to measure patient's cognitive performance outside the clinical setting on a regular basis.

The game based measures that have been shown to indicate cognitive ability and change include:



- Measure of errors in play and specifically trends that show changing performance.
- Number of words found and the rate words are found.

The game based measures that have been shown to have limited value:

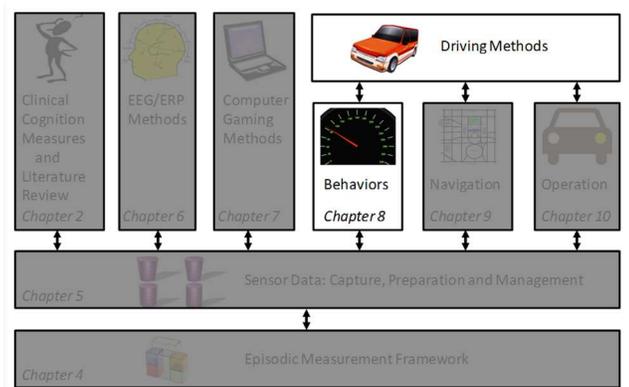
- Game completion time as variation in specific difficulty of specific starting grids causes natural variation in player performance.

This may allow for unobtrusive monitoring of patient cognitive status over a period of time. The data could be collected on a schedule established by the clinician, e.g. weekly or monthly, and allowing for tracking as well as early detection of improvement or decline in the patient's status. Clinicians may use the variations in weekly/monthly statistics like mean and standard deviation to look at trends in performance. Longitudinal measurements of the game performance can then be provided for long term tracking as well as monitoring effectiveness of treatments patients might be having.

Future work on games should explore games more appropriate for subjects with lower cognitive ability such as dementia patients as CSG and CWG proved to be challenging for many of the MCI users and it is likely that they will be too difficult for subjects with even lower cognitive ability.



Chapter 8: Driving: Signature Analysis



Objective: This chapter explores methods for the measurement of driving behaviour which has three main applications:

1. Changes in behaviour may provide an indication of changes in the patient's health (cognitive or physical).
 2. Driver behaviours provide measures that are unique to a given driver that allow drivers of a shared vehicle to be distinguished from each other allowing the driver to be identified.
 3. Driving behaviours provide measures that could allow the personalization of self-driving vehicles to drive more like their owner, improving owner acceptance.
- . The measurement of driving behaviour includes:
- Measurement of trip level attributes for each trip taken by a driver within a multiyear longitudinal study including distance, time of day, rush hour and many other attributes.
 - Detailed second by second trip measurements, including analysis of driver's velocity, acceleration and relative velocity to speed limit.
 - Trend analysis for driver behaviours.
 - Acceleration and deceleration event measurements identified that provide direct measure of a driver unique driving behaviour.



- Big data analytic techniques to process extremely large datasets of longitudinal driving information.
- Ability of the features to distinguish drivers is measured.
- Performance of machine learning based classification methods to distinguish drivers.

8.1. Method

This section presents methods that were used to identify and measure behaviours of drivers leading to their use as features to distinguish between drivers.

8.1.1. Driver Behaviour Attributes - 100 Trip Views

Initial analysis of the Candrive study data (2008 to 2016) focused on the Ottawa based participants. Their demographic information is summarized in Table 3-2. The initial focus was on the identification of features, algorithms and methods that could distinguish between drivers of a vehicle, with a goal of building classifiers that allow for drivers of the same vehicle to be differentiated from each other. This initial analysis focused on the data for 100 trips for four example sole drivers from different vehicles.

Table 8-1: Summary travel distance and time statistics for 100 trips for each of 4 drivers.

	Driver 1	Driver 2	Driver 3	Driver 4
Trip length (km)				
Mean	5.2	26.2	6.4	5.5
Std Dev	5.1	41.8	6.2	4.7
Min	0.50	0.35	0.02	0.02
Max	27.3	207.9	29.3	24.9
Trip duration (min)				
Mean	11.0	23.7	16.1	11.4
Std Dev	6.5	30.1	12.1	8.0
Min	3.6	2.1	3.4	2.0
Max	32.1	154.4	73.9	53.0

The summary trip statistics for the drivers is shown in Table 8-1. Drivers 1, 3 and 4, show similar results, with mean trip lengths being 5-7km and with a maximum trip length being less than 30km. This contrasts with the results of Driver 2. Trip duration data shows more variation than distance for drivers 1, 3 and 4. Drivers 3



and 4 had longer maximum trip times than driver 1 indicating that they may leave their vehicle idling more frequently, travel at slower speeds either by choice or because of travel at times with higher congestion, or the speed limits of the roadways used.

Data analytics techniques allows for a number of different attributes to be considered, including: time of day, day of the week, and driving at peak (rush hour) traffic times, and day/night/twilight driving. GIS information, including posted limits, allows for the determination of a driver's chosen driving velocity relative to the posted speed limit and road choice habits. The use or avoidance of expressways (typical 100 km/hr limit in study area), highways (80 km/hr limit) or city streets (50 km/hr limit) indicates driving speed preference, which may influence chosen route.

The date and longitude/latitude information with solar cycle [137] for a region indicates the times the driver was driving during daylight hours, at dawn, at dusk, or during night hours. Dawn and dusk driving are defined as driving within 1 hour prior to and 1 hour after sunrise and sunset times respectively. As the selected Candrive participants reside in greater Ottawa, a northern latitude city, there is significant variation in sunrise/sunset over the year.

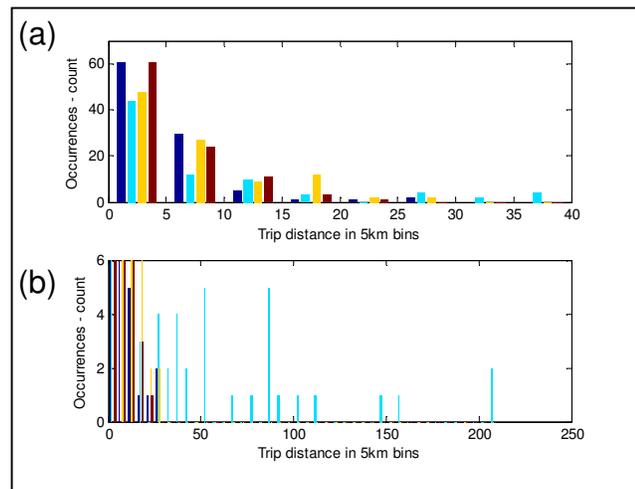


Figure 8-1: Histogram of trip distances for the 100 trips for each of the participants shown on two different scales. Plot (a) shows the detailed histogram for trip distances for all trips 40km or less while plot (b) includes the longer distance trips with enlarged vertical scale. Driver 1 - dark blue, 2 - light blue, 3 - yellow, 4 - red.



Driving during high traffic times may generate different behaviours because of the influence of the other vehicles and is classified using equation 8-1.

Rush Hour : [8-1]

if (Saturday or Sunday or holiday) {not rush hour}

else if (time between 7am and 9am) {rush hour}

else if (time between 4pm and 6pm) {rush hour}

else {not rush hour}

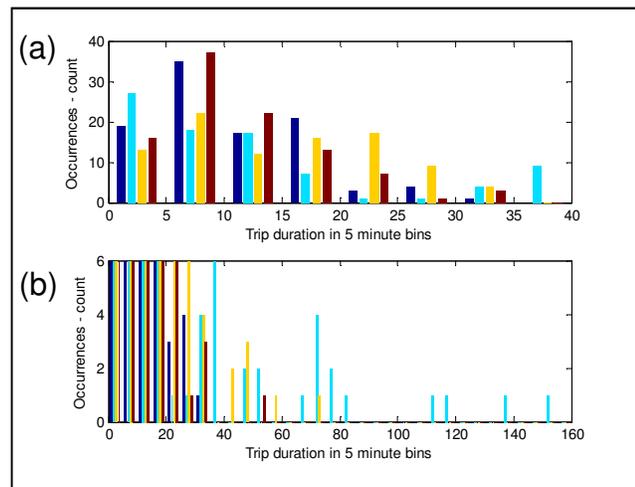


Figure 8-2: Histogram of trip durations for the 100 trips for each of the participants shown on two different scales. Plot (a) shows the detailed histogram for trip duration for all trips 40 minutes or less while plot (b) includes the longer duration trips with enlarged vertical scale. Driver 1 - dark blue, 2 - light blue, 3 - yellow, 4 - red.

Figures 8-1 and 8-2 provide a more detailed histogram analysis of the distances and trip durations for the 100 trips for each of the four participants. Both figures show the tendency for driver 2 to do longer trips with a number of trips greater than 50km and 40 minutes in duration. For all drivers, the most frequent trips are between 0 and 5km in distance, with drivers 1, 3 and 4 trips most frequently lasting 5-10 minutes while driver 2's most frequent trips are 0-5 minutes in duration.

Many older drivers choose to restrict their driving through avoidance of higher risk situations, such as the use of highways or driving at busy times. The type of road is indicated by the posted speed limit and Table 8-2 provides a summary of the percent of time each of the drivers used the various road types. The table excludes times where the speed limit is unknown, indicating that the driver is using



non-roads (e.g. parking lots) or driving in areas outside the region where GIS data was available, such as the long distance trips for driver 2. Drivers 2 and 4 choose to make use of highways, while drivers 1 and 3 do not use 100 km/hr expressways. Driver 1 also makes minimal use of 70-90km/hr highways. Except driver 2, the drivers use city streets for over 85% of their travel.

Table 8-2: Summary for 4 drivers showing percent of time each driver traveled by road choice based on posted limit where posted limit known.

	Posted Limit	Driver 1	Driver 2	Driver 3	Driver 4
city	40km/hr	40.4	0.0	6.0	5.9
	50km/hr	50.6	69.0	69.2	44.6
	60km/hr	8.2	2.6	20.3	35.9
	Total	99.2	71.6	95.5	86.4
highway	70km/hr	0.2	1.4	0.1	0.1
	80km/hr	0.2	3.5	4.3	3.3
	90km/hr	0.0	6.7	0.0	0.0
	100km/hr	0.4	16.8	0.1	10.2
	Total	0.8	28.4	4.5	13.6

The time of day that a driver chooses to drive allows them to manage risk and avoid conditions related to both traffic levels and visibility (i.e. night). Table 8-3 summarizes the timing of trips against the typical traffic patterns where there is a typical morning and evening rush hour on weekdays. The table shows that all the drivers generally avoid early morning rush hour driving and the preponderance of their driving is during the daylight hours. Additional analysis of these 4 drivers is included in Appendix A.

Table 8-3: Summary for 4 drivers showing percent of time based on time of day and related traffic levels (Note statutory holidays counted as weekend days)

Rush Hour	Driver 1	Driver 2	Driver 3	Driver 4
Weekday - over night	7.0	2.0	1.3	4.2
Weekday - morning rush	2.2	4.6	2.5	0.0
Weekday - mid day	56.3	46.9	60.6	49.2
Weekday - evening rush	9.5	19.5	3.1	10.6
Weekend - over night	0.0	2.0	1.3	4.0
Weekend - morning rush	1.0	1.6	0.0	0.0
Weekend - mid day	24.0	17.1	28.8	24.1
Weekend - evening rush	0.0	6.2	2.5	7.9



8.1.2. Driver Behaviour Attributes - Month to Month Trends

Fourteen drivers that demonstrated stable general, physical and cognitive health over their first year in the Candrive study [135], from the Ottawa region and that did not share their vehicle were chosen for detailed analysis. Stable general, cognitive, and physical health allows for analysis while controlling for any influence on their driving behaviours caused by changing health. As sole drivers, all data for the vehicle can be associated with the specific driver. The demographics of these 14, drivers are summarized in Table 8-4. The table shows that the participants are all older drivers that make regular use of their vehicle.

Table 8-4: Summary of the dataset for 14 Triple Stable Candrive Participants Analyzed

Demographic Data	Total 14: Male 9, Female 5
	Age at entry: mean 75.7 yrs, st dev 5.4 yrs
Trips / yr	mean 1567, st dev 647
Distance / yr	mean 20,200 km, st dev 10,600 km
Average Trip Duration	mean 15.0 min, st dev 3.1 min
Average Trip Distance	mean 13.6 km, st dev 7.0 km

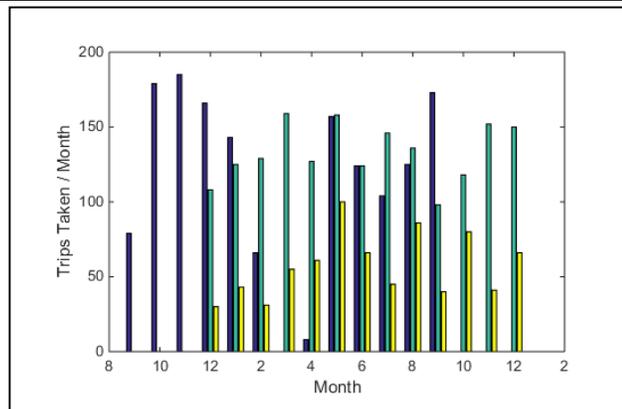


Figure 8-3: Number of trips taken by three sample drivers each calendar month for 1 year from their entry into the study.

Figures 8-3 and 8-4 show the general driving habits for three example drivers from the 14 drivers showing the number of trips taken and the distance driven for each month respectively over their first year in the study. One of the drivers joined the study in September while the other two joined in December. The figures show some of the attributes of the drivers, such as the almost 2-month period where one



of the drivers was away (March-April), otherwise the drivers drove every month with some variation. Two of the drivers typically drive more than 100 trips per month. The distance per month indicates different patterns; although two of the drivers have similar trips/month counts, they have very different distance driving patterns.

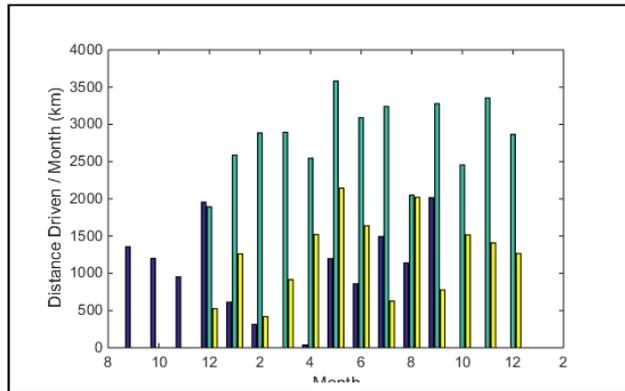


Figure 8-4: Distance driven by three sample drivers on a calendar month basis for 1 year from their entry into the study.

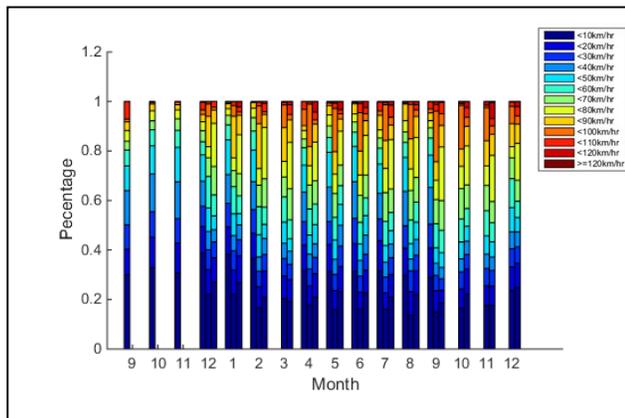


Figure 8-5: Analysis of the velocity driven by three sample drivers showing calendar month averages of the velocity (km/hr) histogram distributions.

Figure 8-5 shows example histograms of percent of time in 13 velocity ranges for each month for the same three sample drivers. The large amount of time spent at lower velocities is expected, since this would include time spent in slow traffic, at traffic signals, slow moving areas and when accelerating from and decelerating to a stop. The data shows some of the expected differences, such as a driver spending more of their driving time at the lower velocities. The figure also shows that drivers may have a habit to exceed the speed limit as there are no roads with



speed limits above 100 km/hr in the region driven and it shows one driver that spends more of their time at higher speeds compared to the other two.

Figure 8-6 presents data on the time of day when driving occurred and specifically compares time of day to the solar cycle. Although there appears to be variation in night, dawn, and dusk driving during the winter months of the study period, this change is likely caused by the seasonal solar variation in the northern hemisphere (Ottawa, Canada). Drivers that habitually make trips at the same time of day will have trips shifted into the dawn, dusk, and night periods during the winter months. This feature has limited value to distinguish specific trips between drivers, as an average 15-minute trip spans a limited portion of the solar day with all drivers completing trips during each period of the solar day and 97.2% of the trips completed within a single solar period.

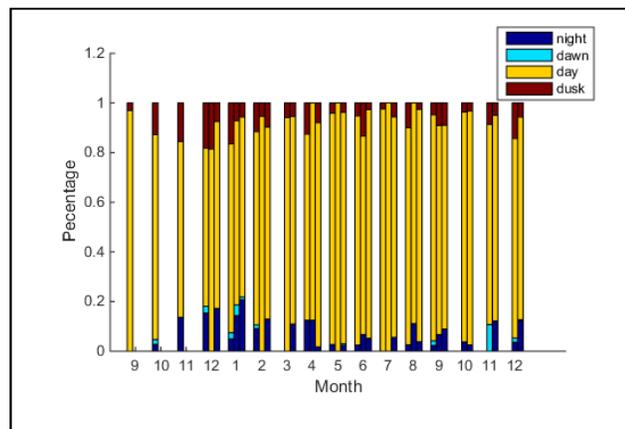


Figure 8-6: Analysis of the time of day for three sample drivers for trips driven showing calendar month averages of the time in the solar day (dawn, day, dusk, night).

8.1.3. Sensors signals excluded from analysis

The Candrive sensor system captures many signals associated with the details of the engine operation and behaviour such as engine RPM and other direct measures of the engine and these at best provide very indirect measures of the driver's behaviour. As an example, with the prevalence of automatic transmission cars, a driver's choice of speed which is a direct behavioural decision is translated into a transmission gear choice and associated engine RPM. Both of these influenced by the car design. The goal of this work is the identification of measures



of the driver behaviour and as a result measures available in the Candrive set that provide vehicle measures and not driver choice alone were excluded.

8.1.4. Acceleration and Deceleration Events

Measuring driver acceleration characteristics is important as it could lead to insights into driver performance and potentially help indicate who is driving. Appendix B compares the acceleration signals derived from the GPS and OBDII velocity sensors with accelerometer measures and the Central 2-point derivative (equation 8-2) where h is the sample period (1 second) of the vehicle GPS velocity.

$$f'[n] = \frac{f[n+1] - f[n-1]}{2h} \quad [8-2]$$

The analysis of behaviours associated with acceleration requires a focused analysis of acceleration and deceleration events. Data analytics techniques used to identify deceleration and acceleration events that met the specification in equation 8-3 & 8-4:

Deceleration event specification: [8-3]

Acceleration[n] < 0 for n=k, k+1, ... m AND

VGPS [k] - VGPS [m] >= 4 km/hr AND

No sample gaps > 2 seconds (i.e. max 1 sample gap)

Acceleration event specification: [8-4]

Acceleration[n] > 0 for n=k, k+1, ... m AND

VGPS [m] - VGPS [k] >= 4 km/hr AND

No sample gaps > 2 seconds (i.e. max 1 sample gap)

This defines events as a continuous set of measures with a total velocity change of 4 km/hr or more while ensuring that events that contain sample gaps are split or discarded. This removes events with very small changes in velocity to allow focus on events where the driver is more likely to be making the change through direct choice or need. It also ensures events do not falsely span gaps in the data.



Table 8-5: Summary of the features calculated for all acceleration and deceleration events over the study period where a 4 km/hr or larger velocity change occurred.

Acceleration Attributes	Mean acceleration Minimum acceleration (deceleration events) Maximum acceleration (acceleration events) Actual acceleration values (vector)
Velocity Attributes	Velocity change Initial & Final velocity for event Actual velocity values (vector)
Duration Attributes	Duration of event
Time Attributes	Time of day Date of event Solar cycle position Traffic level (Rush hour or not)
Road Network	Posted speed limit

The data analytics classification filters were applied for the first year of study participation leading to the identification of every deceleration and acceleration event for each trip. Table 8-5 summarizes the analytics data structure generated for each event. All trips driven by the driver were analyzed to identify every event within the trips. This leads to a set of acceleration and deceleration events for each trip.

Initially all the identified events for each driver were pooled into a single set, sorted and summarized based on these various attributes, including the use of moving averages and histograms. Histograms were further analyzed using the Matlab fit function for a Gaussian distribution to provide measures of the mean and standard deviation of the resulting distributions. Linear regression techniques were applied to identify relationships within the analyzed results.

8.1.5. Distinguishing Features - t-tests

The measurements of a trip feature for two different drivers can be compared through a 2-tail t-test assuming unequal standard deviations to determine if measurements for two drivers are different, where an alpha value of <5% has been used to reject the hypothesis that the two drivers have the same behaviour. The sequences for each driver are independent of each other and represent a series of measures of the behaviour of each driver. The t-test, as applied, indicates that the



two drivers have different or similar behaviours and hence measures the ability of the feature to distinguish the two drivers. The 14 sole drivers allow for 91 distinct pairs of drivers to be formed.

8.2. Experimental Results

This section presents experimental results showing how they can be used to identify and measure behaviours of drivers and provide features to distinguish between drivers.

8.2.1. Acceleration and Deceleration Events

The patterns in the deceleration events for the drivers is explored starting in Figure 8-7 where deceleration events are compared with total velocity drop for the event and the event duration for an example driver. As expected, there is some correlation $R^2 = 0.749$ between duration and velocity drop. Figures 8-8, 8-9, and 8-10 show different views of the monthly distributions of events. The Figures all show a vacation period with limited driving but otherwise do not show any significant variation. Winter/summer seasonal variation could have been expected, but there are minimal differences between months 12-3 (winter) to 6-9 (summer).

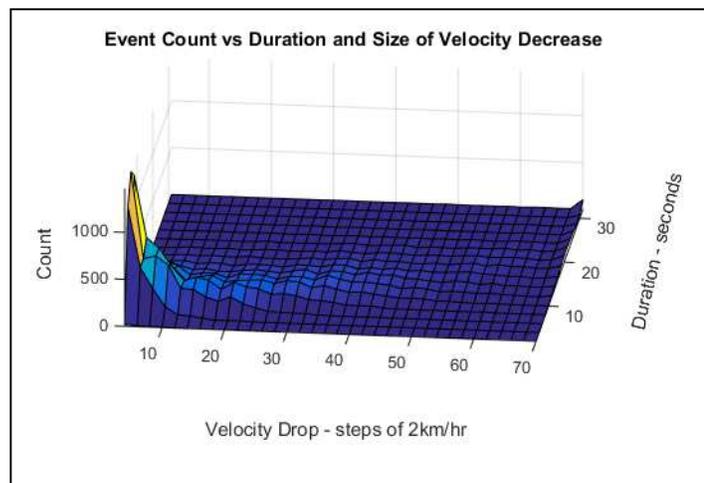


Figure 8-7: Distribution of deceleration events for one driver compared to the size of the velocity change and the duration of the event (N= 25016 events).

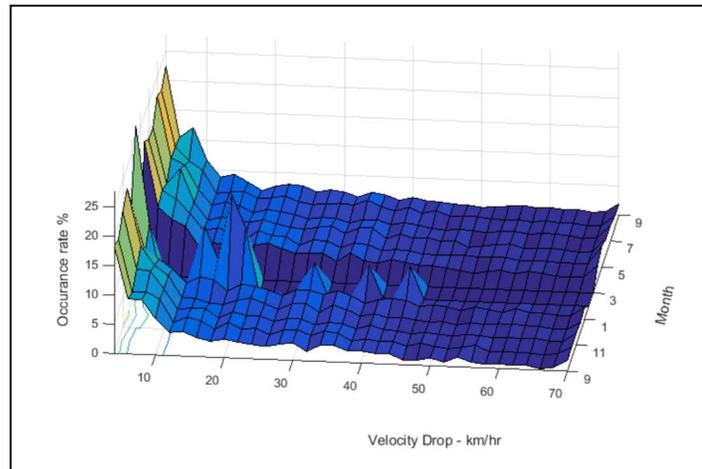


Figure 8-8: Monthly velocity change profiles for 1 year. (N= 24794 events in total).

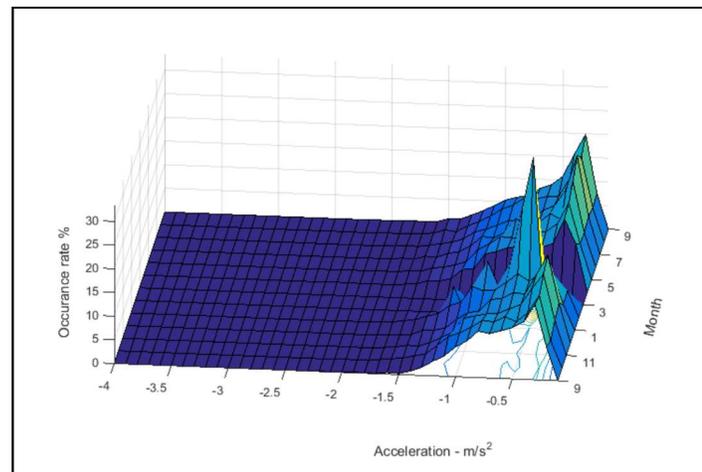


Figure 8-9: Distribution of the monthly mean acceleration values for the deceleration events. (N= 24794 events in total).

Figure 8-11 compares the number of deceleration events identified using equation 8-3, against the event mean deceleration and duration, showing, the high occurrence of short duration, small deceleration events. As well, results show that mean deceleration has an almost constant value once an event is of sufficient duration. This indicates the potential for a two phase deceleration relationship.

Figure 8-12 shows monthly distributions of minimum (largest) deceleration within deceleration events and again there are no significant seasonal variations.

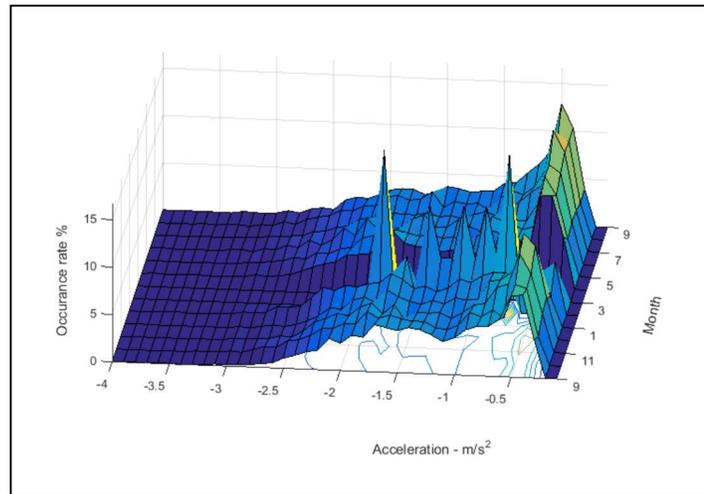


Figure 8-10: Distribution of the monthly minimum acceleration (largest deceleration) values for each of the deceleration events. (N= 24794 events in total).

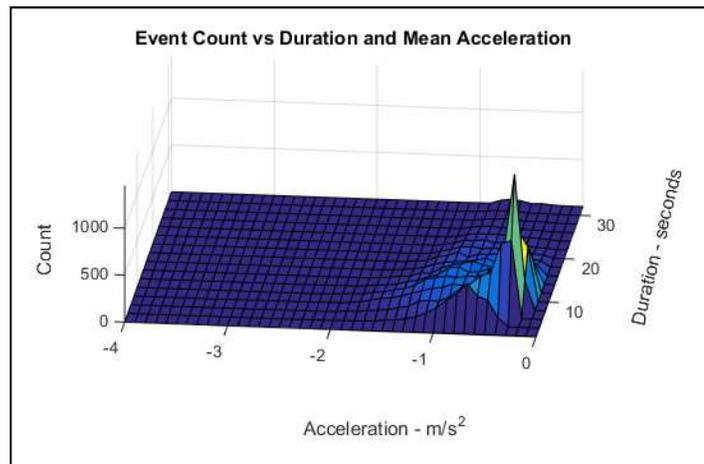


Figure 8-11: Distribution of the deceleration events for one driver comparing the number of events to both the mean event acceleration and duration. (N= 25016 events).

Figure 8-13 and 8-14 explore the relationship between driver mean and minimum acceleration and velocity drop size. In both cases the plots show a clear relationship with a ridge line in the distributions. The mean acceleration for a given velocity change starts near zero and increases with the size of the velocity drop until growth slows for the largest decelerations. This indicates two possible phases for decelerations: smaller changes where the driver might be influenced by external factors such as traffic, and larger changes where the driver's preference or habit may dominate. The same characteristic dual phase deceleration relationship is



present in Figure 8-26 for the minimum acceleration value, with similar attributes other than the size of the acceleration values.

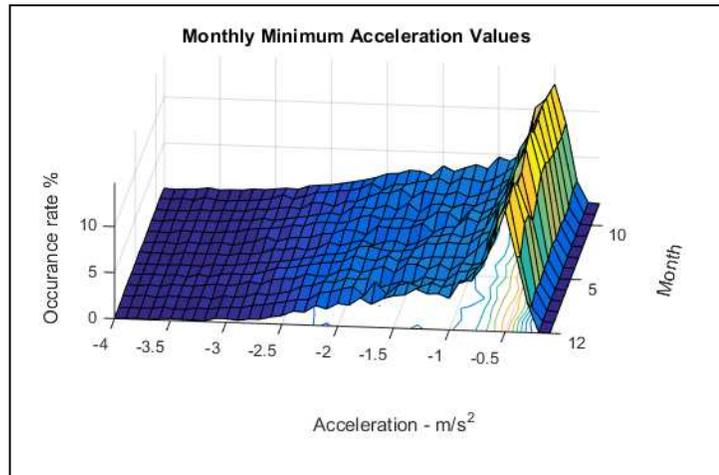


Figure 8-12: Distribution of the monthly minimum acceleration (largest deceleration) values for an example driver for the deceleration events. (N= 40213 events).

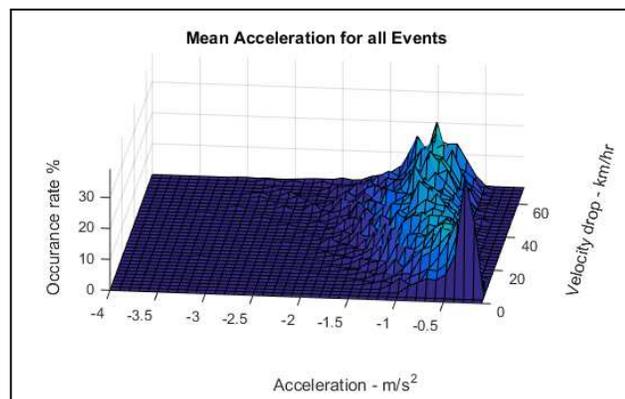


Figure 8-13: Distribution of the mean acceleration for one driver against the size of the velocity change in the deceleration event. (N= 25016 events).

These characteristics were explored on a number of subsets of the data such as looking at only trips where the posted limit indicates the driver was on a street with a velocity of 40-70 km/hr - typical for city streets in the study area (Figure 8-15) with very similar ridge line attributes to Figure 8-13. The distribution of deceleration events ending with a stopped vehicle is shown in Figure 8-16. Again a ridge line is present and similar to Figure 8-13. The same observation was found for minimum acceleration values and was found across all examined data subsets for both mean and minimum acceleration. The similarity of the Figures 8-13, 8-15 and 8-16 is



shown in Figures 8-17 and 8-18 where a best fit Gaussian distribution was calculated for each of the acceleration histograms to provide a measure for the ridge line position indicated by the best fit mean and the ridge line variation indicated by the standard deviation of the best fit mean.

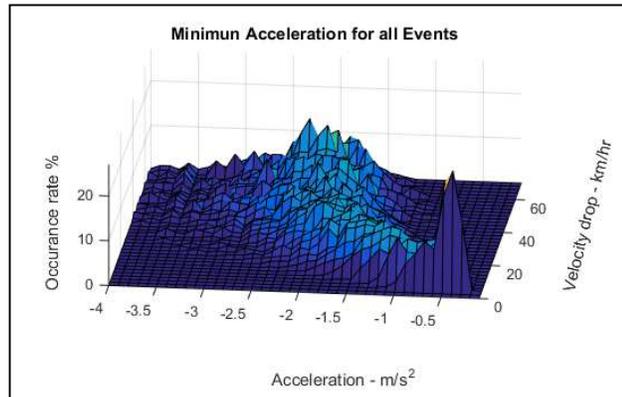


Figure 8-14: Distribution of the minimum acceleration for one driver against the size of the velocity change in the deceleration event. (N= 25016 events).

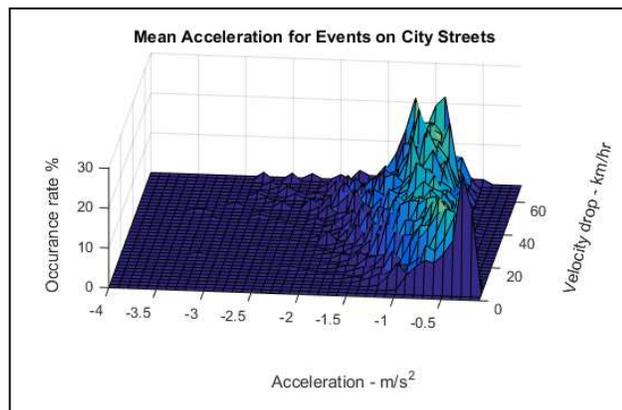


Figure 8-15: Distribution of the mean acceleration for one driver against the size of the velocity change in the deceleration event for driving on roads with speed limits >40 km/hr and ≤ 70 km/hr (N= 16750 events).

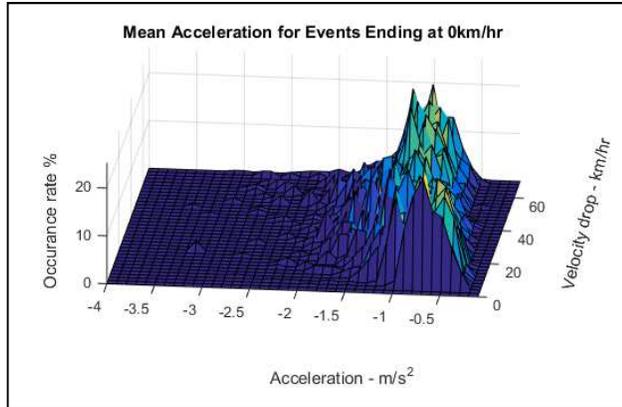


Figure 8-16: Distribution of the mean acceleration against the size of the velocity change for events where the final velocity was a stop (0 km/hr). (N= 7685 events).

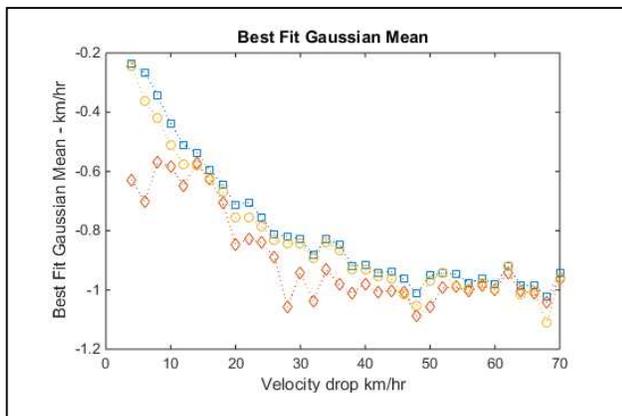


Figure 8-17: Comparison of the estimated best fit Gaussian mean for one driver to indicate the ridge lines: Figure 8-13, blue \square , all decelerations; Figure 8-15, red \diamond , city street; Figure 8-16 - yellow \circ , final velocity of 0 km/hr.

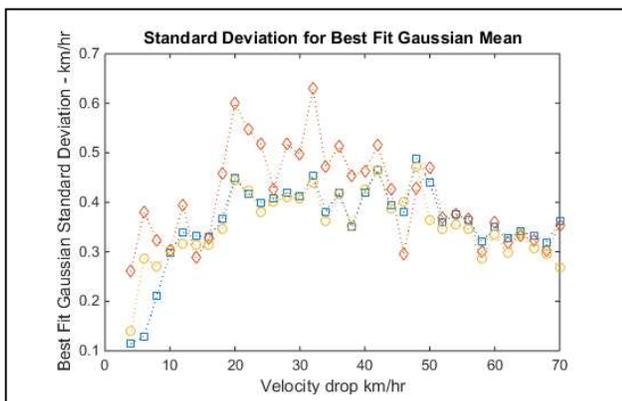


Figure 8-18: Estimated standard deviation for best fit Gaussian distributions for one driver as measure of ridge width: Figure 8-13, blue \square , all decelerations; Figure 8-15, red \diamond , city street; Figure 8-16 - yellow \circ , final velocity of 0 km/hr.

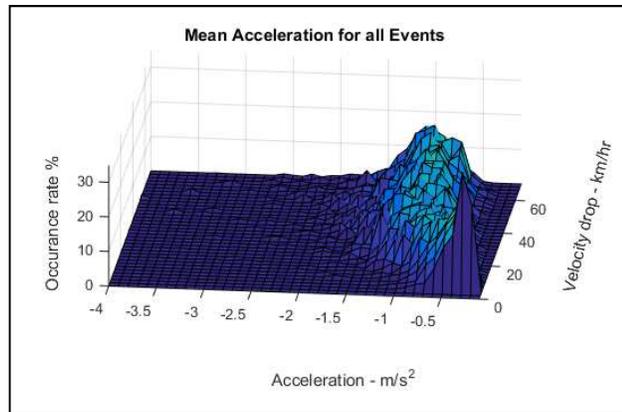


Figure 8-19: Distribution of the mean acceleration for a second example driver against the size of the deceleration for the event. (N= 22469 events).

The other 13 drivers were analyzed for these same characteristics and similar patterns were found in all of the drivers. An example second driver is shown in Figure 8-19 for deceleration event mean and Figure 8-20 for deceleration event minimum. Although the similar pattern was observed, there are differences between the drivers as to the location of the ridge lines and transitions vary between the drivers. This leads to the further exploration of these two phase relationships as a means to differentiate drivers.

The Two Phases of Deceleration

Figure 8-21 explores the ridge lines for the distributions shown in Figures 8-13 and 8-14. The best fit Gaussian mean for each histogram is shown along with the 95% confidence intervals for each of these mean estimates with a similar pattern of two linear phases: one for velocity drops between 4 km/hr through to the 25-30 km/hr region. The second linear phase then starts and continues through to the largest continuous velocity drops that were larger than 70 km/hr.

The optimal fit for the two phases for both the mean and minimum acceleration curves was found using linear regression within each phase. The location of the optimal phase transition is presented in Figure 8-22. The resulting regression lines for each phase were combined for a correlation coefficient for the resulting paired lines. The figure shows the results for one driver where a peak correlation of 0.989 was achieved for the mean acceleration ridge line and 0.993 for the minimum acceleration ridge line with a transition between the phases at 27.3 km/hr and 27.8



km/hr respectively. These optimal fit lines are shown in Figure 8-21. The best fit for each of the two phases provides a feature that characterizes the driving habits for the driver.

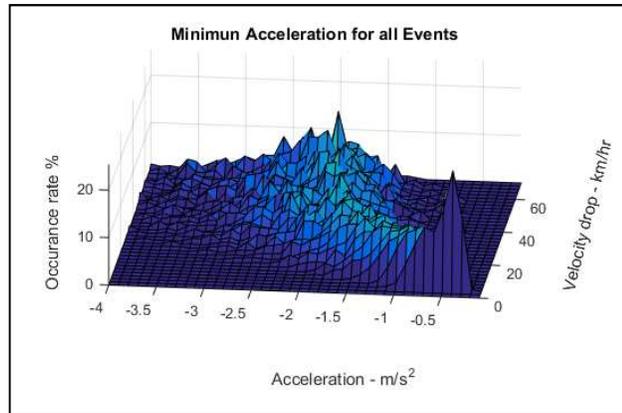


Figure 8-20: Distribution of the minimum acceleration for a second example driver against the size of the deceleration for the event. (N= 22469 events).

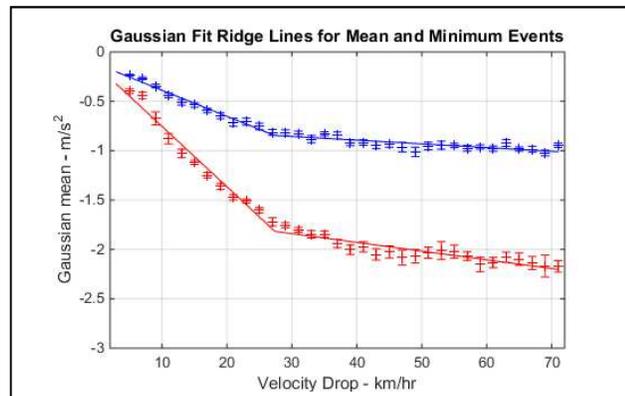


Figure 8-21: Plots of the ridge lines for one driver for the mean surface in Figure 8-13 (blue) and minimum surface in Figure 8-14 (red) where the estimated peak is shown along with 95% confidence intervals. Best fit two phase relationship with optimal transition between the phases is shown by the lines.

Figure 8-23 shows the mean and minimum lines found for 4 example drivers from the set of 14 drivers. This shows that although some drivers can have very similar relationships, such as the ones shown in blue and red in the 25-30 km/hr region on the minimum line, the other two drivers are much more distinctive. A general pattern for the drivers to have more diversity in the minimum lines than in the mean lines is apparent and this pattern is followed across all 14 drivers and



may be an indication that personal behaviour manifests itself more in that feature as it has less normative effects.

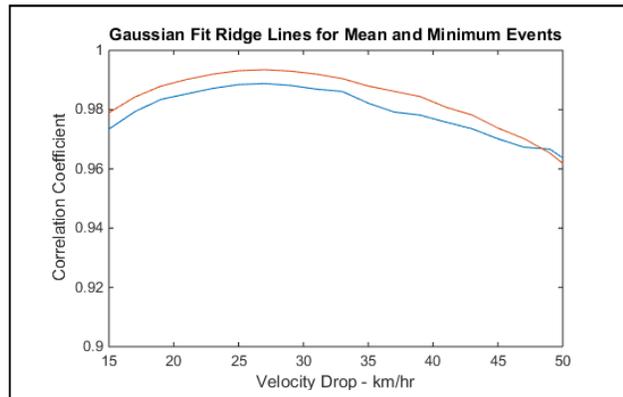


Figure 8-22: Correlation coefficient compared to phase transition for the two phase linear interpolation for one driver for the surface ridge lines shown in Figure 8-21. Mean surface (blue) and Minimum surface (orange).

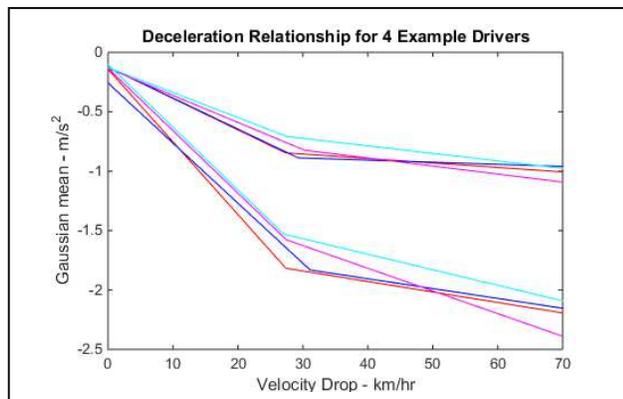


Figure 8-23: Plot of the ridge lines for four different drivers showing the two phase linear relationship for each of the drivers while also showing the variations in the curves that will be explored to determine if they allow the differentiation of drivers.

Table 8-6 shows the summary results for all 14 drivers and the high correlations that were achieved for the ridge lines for each of the drivers. Testing of the ridge lines derived from all driver events to subsets of the events is also presented in the table. The correlations with day time driving had minimal change while there was more change for night time driving as may be expected as drivers may change their behaviours. The large change in the non-rush hour data was caused by one driver's low correlation for mean (0.498) while still having a high correlation for minimum (0.946).



The Two Phases of Acceleration

Similar to deceleration, acceleration events provide another potential measure of a driver's behaviour. Figure 8-24 and 8-25 explore the relationship between driver mean and maximum acceleration and velocity drop size respectively, for identified acceleration events. The plots again show a clear relationship with a ridge line for the distributions and again indicate two possible phases for accelerations: smaller changes where the driver might be influenced by external factors, and larger changes where the driver's preference or habit may dominate.

Table 8-6: Deceleration correlation coefficients results for the two phase linear relationships for the 14 cognitive and physically stable drivers.

	MEAN LINE			
	MIN	MAX	MEAN	ST DEV
All	0.965	0.996	0.989	0.0080
Day	0.962	0.994	0.986	0.0081
Night	0.761	0.973	0.889	0.0531
Rush Hour	0.767	0.981	0.912	0.0784
Off Rush Hour	0.498	0.988	0.939	0.1279
	MINIMUM LINE			
All	0.981	0.998	0.994	0.0046
Day	0.980	0.998	0.992	0.0047
Night	0.743	0.981	0.924	0.0641
Rush Hour	0.797	0.988	0.931	0.0638
Off Rush Hour	0.927	0.995	0.975	0.0195
N (EVENT COUNT)	17593	60153	30156	11010

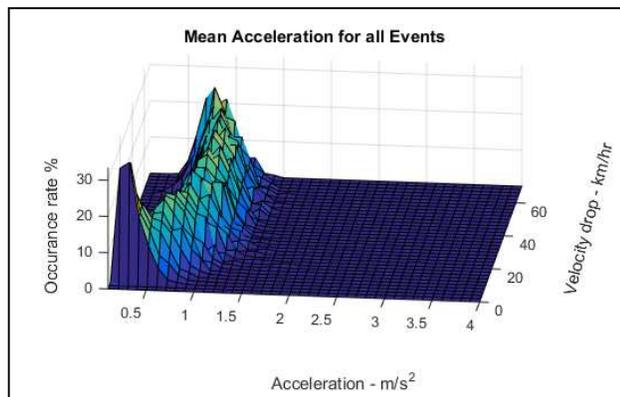


Figure 8-24: Distribution of the mean acceleration for one driver against the size of the velocity change in the acceleration event. (N= 25261 events).

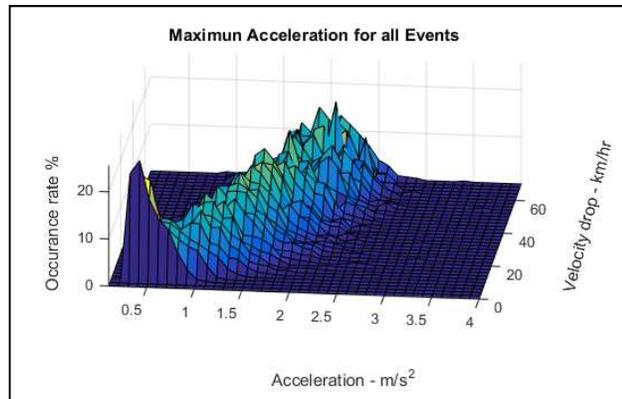


Figure 8-25: Distribution of the minimum acceleration for one driver against the size of the velocity change in the acceleration event. (N= 25261 events).

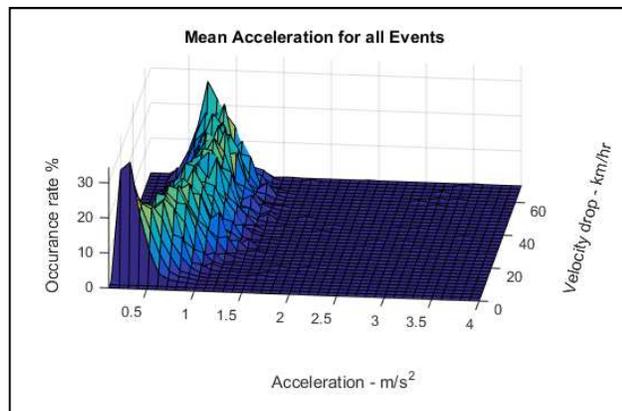


Figure 8-26: Distribution of the mean acceleration for a second example driver against the size of the acceleration for the event. (N= 23022 events).

The other 13 drivers again showed the same characteristics, and an example second driver is shown in Figure 8-26 and Figure 8-27. Although the similar pattern was observed, again there are variations between the drivers that will be explored to determine if they allow drivers to be distinguished.

Figure 8-28 explores the ridge lines for the distributions shown in Figures 8-24 and 8-25. The best fit Gaussian mean for each histogram is shown along with the 95% confidence intervals for each of these mean estimates. Similar to decelerations, the two sets of data have a similar pattern where there are two linear phases: one for velocity drops between 4 km/hr through to the 25-30 km/hr region, the second linear phase then starts and continues through to the largest continuous velocity drops that were larger than 70 km/hr.

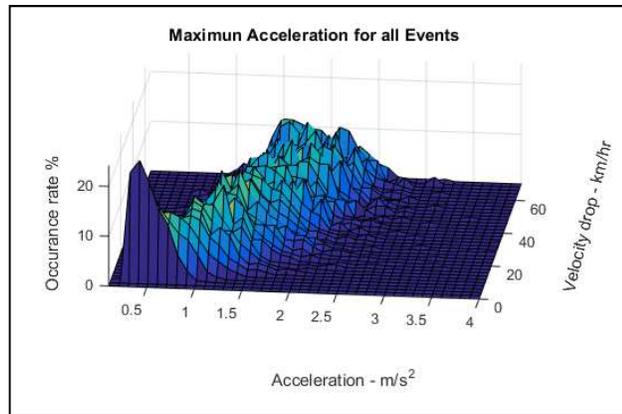


Figure 8-27: Distribution of the minimum acceleration for a second example driver against the size of the acceleration for the event. (N= 23022 events).

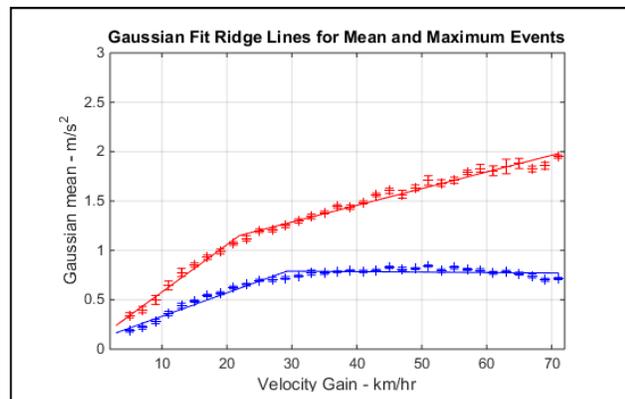


Figure 8-28: Plots of the ridge lines for one driver for the mean surface in Figure 8-25a (blue) and maximum surface in Figure 8-26a (red) where the estimated peak is shown along with 95% confidence intervals. Best fit two phase relationship with optimal transition between the phases is shown by the lines.

The optimal fit for the two phases for both the mean and minimum acceleration curves was found using linear regression within each of the phases. Figure 8-29 shows the mean and maximum lines found for 4 example drivers from the set of 14 drivers and this again shows that although some drivers can have very similar relationships, there are differences. A general pattern for the drivers to have more diversity in the maximum lines than in the mean lines is apparent.

8.2.2. Two-Phase Relationship Feature Measures

The performance of the two phase deceleration relationships as a measure of their typical driving behaviour was analyzed through comparison of the summary behaviour for each driver and the events within each trip that they drove. For this



analysis, the set of events associated with a particular trip is compared with the two phase relationships derived for the pooled set of events. The results are summarized in Table 8-7. Within a given trip, the number of deceleration events can be highly varied, as typically shorter duration trips will have fewer events but also extremely long duration trips can also have a relative low number of deceleration events if they include long periods of highway driving.

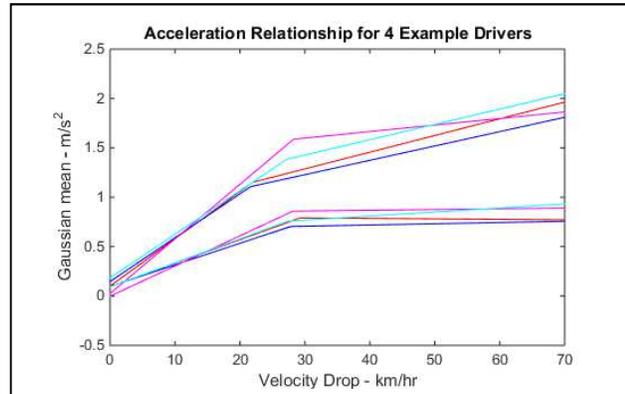


Figure 8-29: Plot of the ridge lines for four different drivers showing the two phase linear relationship for each of the drivers while also showing the variations in the curves that will be explored to determine if they allow the differentiation of drivers.

Table 8-7: Summary of the trip level correlations for all drivers between the Deceleration events in each trip and the driver's two-phase relationship showing the effect of the minimum number of events in each trip.

NUMBER OF DECELERATION EVENTS	MEAN LINE		MINIMUM LINE	
	MEAN	ST DEV	MEAN	ST DEV
>=2	0.500	0.0421	0.701	0.0386
>=4	0.512	0.0435	0.714	0.0379
>=6	0.527	0.0467	0.728	0.0380
>=8	0.544	0.0503	0.741	0.0396
>=10	0.559	0.0549	0.751	0.0411
>=12	0.573	0.0578	0.758	0.0426
>=14	0.582	0.0598	0.762	0.0432
>=16	0.588	0.0635	0.764	0.0430

The data were further analyzed based on the number of deceleration events in the trip and the results show that the trip level correlation increases with the number of deceleration events in the trip until it reaches a plateau around 8 events. 84.6% of the analyzed trips had 8 or more deceleration events. Table 8-7 also shows how



the performance of the minimum deceleration measure is much higher than the mean measure which given the greater diversity in these relationships indicates more potential for these to differentiate between drivers.

Table 8-8 shows the same performance comparison for acceleration events and similar to the minimum measure from deceleration events, the maximum measure for acceleration events is better than the mean measure. Given the greater diversity in these relationships between drivers, this again indicates more potential for these measures to differentiate between drivers.

Table 8-8: Summary of the trip level correlations for all drivers between the Acceleration events in each trip and the two-phase relationship for that driver showing the effect of the minimum number of events in each trip.

NUMBER OF ACCELERATION EVENTS	MEAN LINE		MAXIMUM LINE	
	MEAN	ST DEV	MEAN	ST DEV
>=2	0.464	0.0632	0.689	0.0479
>=4	0.476	0.0658	0.705	0.0441
>=6	0.492	0.0709	0.722	0.0426
>=8	0.507	0.0752	0.737	0.0414
>=10	0.519	0.0780	0.748	0.0396
>=12	0.530	0.0784	0.756	0.0399
>=14	0.541	0.0776	0.761	0.0397
>=16	0.549	0.0757	0.765	0.0386

Figures 8-30 and 8-31 show the specific results for 4 example drivers from the set of 14 and show the distinct maximums in the correlations for each of the drivers for deceleration events and also show the higher performance for the minimum relationship over the mean for each of the drivers. Figures 8-32 and 8-33 show the acceleration events.

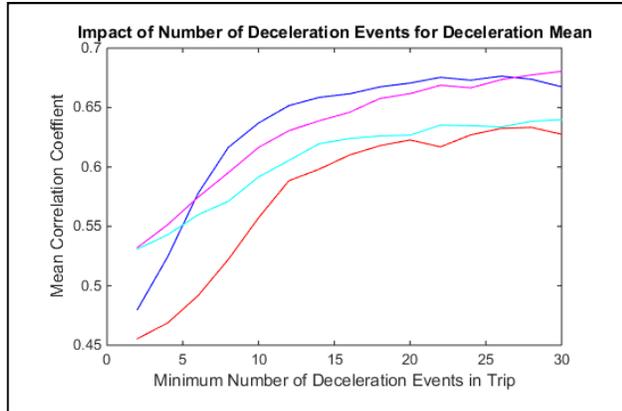


Figure 8-30: Plot of the impact of the minimum number of decelerations events in each trip on mean correlation to the driver's two phase relationship for four sample drivers.

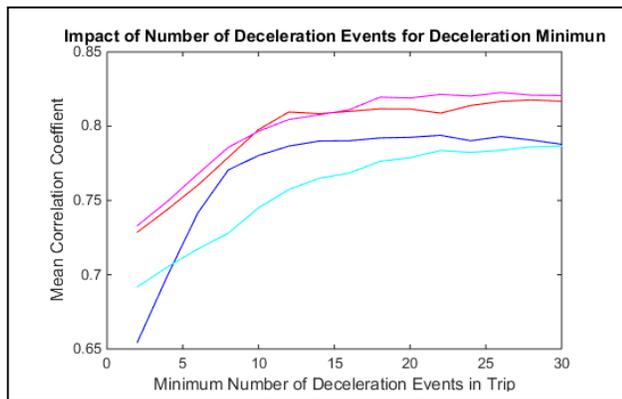


Figure 8-31: Plot of the impact of the minimum number of decelerations events in each trip on mean correlation to the driver's two phase relationship for four sample drivers.

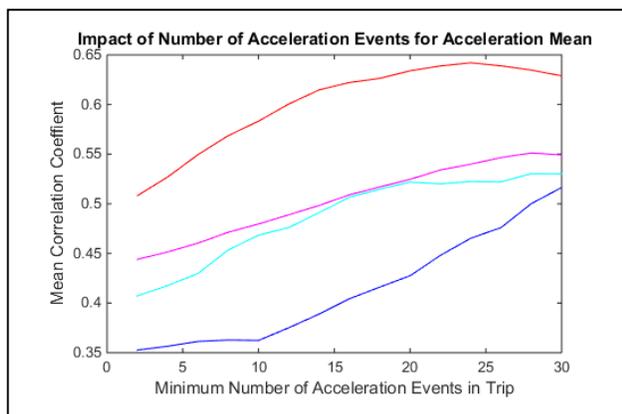


Figure 8-32: Plot of the impact of the minimum number of accelerations events in each trip on mean correlation to the driver's two phase relationship for four sample drivers.

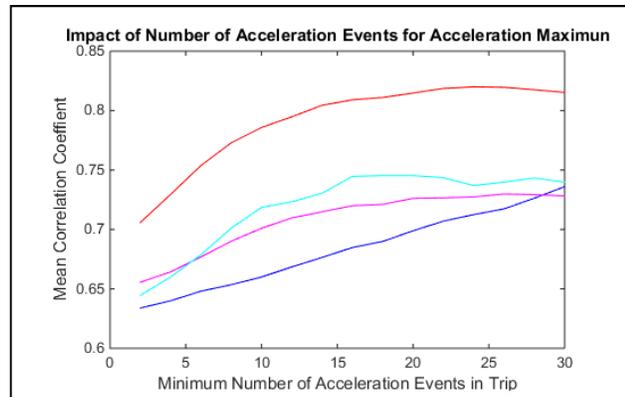


Figure 8-33: Plot of the impact of the minimum number of accelerations events in each trip on mean correlation to the driver's two phase relationship for four sample drivers.

8.2.3. Distinguishing Features

The 14 sole drivers allow for 91 distinct pairs of drivers to be formed, where every driver is independent of each of the others and the distinguishing ability of the features is compared through t-test as shown in Table 8-9. The measurements for most of the measurement features have many attributes that cause non Gaussian attributes [146] in their distributions. Such asymmetry in the distributions is caused where there are no possible negative values but positive tails can be very large compared to a mean such as an extremely long trip with trip distance or duration measures. The trip distance and duration features show that 90.1% (distance) and 90.1% (duration) of the 91 cases demonstrate distinct behaviours. Distance or duration showed distinct behaviours all but 1 of the 91 pairs.

A t-test was also performed on each of the velocity histogram bins for the 13 velocity bands, showing that these have varying potential to be used to differentiate between drivers. The poorest performance is in the higher velocity bands, as they provide no value to differentiate drivers where neither one travels at this speed, but do where only one driver does and may if both drivers do.

The performance of the two phase relationships to provide a measure to differentiate drivers was performed by using each of the 91 driver pairs and using a t-test on the correlation for the trip level events to each of the two phase relationships for both drivers as distinct measures. The result is 8 total tests for each driver pair given that there is a mean and min/max relationship for both



deceleration and acceleration events for each driver in the pair. The results are summarized in Table 8-13 and show that 83.5% of the mean and 85.7% of the minimum relationships had a p value of <5% for deceleration events and 84.6% of the mean and 80.2% of the maximum relationships. For the 91 driver pairs, at least one of the four deceleration relationships provided a p value of <5% indicating that at least one has the potential to distinguish between the drivers.

Table 8-9: t-test results each feature using all possible pairs of 2 drivers from the set of 14 drivers (91 unique pairs) showing percent of pairs distinguished ($p < 5\%$).

Feature		Percent t-Test $p < 5\%$
Trip measures	Length - km	90.1%
	Duration - min	90.1%
Velocity bin - % of trip	<10km/hr	93.4%
	>10 & <20km/hr	91.2%
	>20 & <30km/hr	91.2%
	>30 & <40km/hr	95.6%
	>40 & <50km/hr	93.4%
	>50 & <60km/hr	82.4%
	>60 & <70km/hr	90.1%
	>70 & <80km/hr	86.8%
	>80 & <90km/hr	93.4%
	>90 & <100km/hr	86.8%
	>100 & <110km/hr	83.5%
	>110 & <120km/hr	85.7%
	>120km/hr	60.4%
Deceleration Event - m/s^2	Event Mean	83.5%
	Event Minimum	85.7%
Acceleration Event - m/s^2	Event Mean	84.6%
	Event Minimum	80.2%

8.3. LDA Classification

As presented for the EEG / ERP classification in Chapter 6, machine learning tools based classifiers provide a potential mechanism to combine the broad set of trip features that have been developed leading to a classification between two potential drivers.



8.3.1. Features and Test Cases

The work has presented the analysis and measurement of each of the trips driven providing a large number of features for each trip. The resulting set of 162 features is summarized in Table 8-9 and includes 8 features derived from the acceleration and deceleration events within the trip. An additional 2 features are provided by the overall trip duration and 43 features are available related to the time of day for the trip. The type(s) of road chosen by the driver provide another 5 features while velocity and associated acceleration and jerk measures provide 26, 35 and 43 features respectively.

Similar to the t-test analysis presented previously, the 14 triple stable sole drivers provide a set of 91 gold standard cases that can be used for analysis as the trips for two drivers are mixed to create a "shared" vehicle, the actual driver for every trip is known. Given the large number of trips driven by the 14 drivers within the year as shown in Table 8-4, thousands of trips were available for training and testing of the resulting classifier. A 90/10 test and train method was used where the given data for a case was divided randomly into 10 groups containing 10% of the trips each and all 10 cases where 90% of the data was used to train and 10% used for testing with the performance based on the average across the 10 cases.

8.3.2. LDA Classification - single feature

The potential for each of the 162 identified features was initially explored by evaluating the classification performance for LDA classifiers for both linear and quadratic for all 91 driver pairs using the 90/10 test and train methods for each of the driver pairs. The results demonstrated the highly variable performance of features across the driver pairs.



Table 8-9: Summary of features fore each trip used with classification algorithms.

Feature Group	Features	Number of features
Acceleration Events	<ul style="list-style-type: none"> • Trip events correlation to self/other driver for Mean/Max 	4
Deceleration Events	<ul style="list-style-type: none"> • Trip events correlation to self/other driver for Mean/Min 	4
Trip Attributes	<ul style="list-style-type: none"> • Length of trip in seconds and kilometers 	2
Time of Trip	<ul style="list-style-type: none"> • Time of day histogram (hourly bins) • Rush hour driving • Solar cycle driving • Day of week 	24 8 4 7
Posted Limit on Road	<ul style="list-style-type: none"> • ≤ 40km/hr, < 70km/hr, ≤ 90km/hr, > 90km/hr, unknown 	5
Velocity	<ul style="list-style-type: none"> • Actual velocity (10km/hr bins) • Velocity ratio to posted (10% bins) 	13 13
Acceleration	<ul style="list-style-type: none"> • Trip minimum and Maximum • Actual Acceleration histogram bins (-3m/s^2 to 3m/s^2) 	2 33
Jerk	<ul style="list-style-type: none"> • Trip minimum and Maximum • Actual Jerk histogram bins (-5m/s^3 to 5m/s^3) 	2 41
Total		162

A number of features provided no classification value at all, but this was expected as these features measured driving attributes that were atypical of the older drivers in the sample. Features that failed for all driver pairs were middle of the night driving between 1am and 5am while late night (10pm-1am) and early morning (5am-8am) failed for some driver pairs. Similarly driving at dawn failed for some driver pairs. Older drivers typically do not drive during these times of day and hence the features



contained minimal information. Night driving because of early sunset in Ottawa did contain enough information for each driver to provide some classification value. Other features that failed in some or all cases include the highest and lowest bins for acceleration and jerk, especially the large positive jerk bins as these histograms were designed to span the values observed in the drivers and some drivers again did not exhibit behaviours at the extremes. The imbalance in the jerk values is expected as hard braking is a more frequent event than matching levels of hard acceleration. The result, up to 22 of the features provided no distinguishing value in some or all of the driver pairs.

The feature group that provided the good overall performance with error rates as low as 3.4% for one driver pair was road type choice as measured through posted speed limit information. Although these provided strong classification performance for some driver pairs, the average error performance for these features individually across all 91 pairs was 30-43% with worst case error rates of over 50%.

Another group that shows low error rates performance for some pairs is the velocity ratio features that have best errors rates for some driver pairs between 5 and 16% while again having average error rate performance between 30-40% and worst case error performance over 50%.

No other feature group exhibited best error performances better than 10%, although many provided error rates between 10 and 25% for best cases and the summary shown in Table 8-11 shows the number of features providing various levels of best performance.

These results lead to the conclusion that there is no one feature that can be used to distinguish every driver pair and that combinations of features are likely going to be required to achieve the best performance.

8.3.3. LDA Classification - two features

To further explore the classification potential of the features, LDA classification models similar to the single feature models in the previous section were run for all pairs of features across all 91 driver pairs using the same 10 repetitions of the 90% train, 10% test models. The results are presented for the quadratic LDA classifiers.



Table 8-10: Single feature LDA classification performance statistics for the set of 91 driver pairs. For each feature, all 91 pairs were classified and the aggregate performance statistics for the feature were calculated for Linear and Quadratic LDA models. Columns compare stats for single features to other features. Rows compare aggregate performance for single feature classifiers.

Attribute	LDA (Linear)				LDA (Quadratic)			
	Lowest Error Rate	Mean Error Rate	Standard Deviation	Largest Error Rate	Best Error Rate	Mean Error Rate	Standard Deviation	Largest Error Rate
Best	3.4%	29.2%	3.5%	51.9%	4.9%	29.9%	3.5%	49.7%
Mean	19.2%	42.8%	10.7%	66.6%	19.0%	42.0%	10.0%	62.5%
Largest	37.6%	55.5%	17.6%	81.8%	37.6%	55.7%	17.2%	83.2%

Table 8-11: Summary of features for each driver pair used with classification algorithms.

Error Rate	LDA (Linear)	LDA (Quadratic)
$\leq 5\%$	2	2
$> 5\%$ and $\leq 10\%$	13	11
$> 10\%$ and $\leq 15\%$	23	27
$> 15\%$ and $\leq 20\%$	42	47

The results for the 2 feature classifiers are presented in Figures 8-34 through 8-36 where the performance is presented on the basis of the driver pairs. The diagonal line on the plots can be ignored as both drivers being the same is not a valid test case (applies to Figures 8-34 through 8-41). Figure 8-34 shows the best performance from a 2 feature classifier and a number of attributes can be observed. Driver 3 appears to be more easily distinguished from all of the other drivers and the driver 10 and 14 combination is the most difficult to distinguish. The large number of combinations with error rates higher than 15% indicates the need for classifiers using larger number of features. Figure 8-35 shows the worst performance level for the driver pair excluding features pairs that included features



that completely failed individually (100% error rate) as discussed previously. This figure shows that in most cases the worst performance was above 50%, which is worse than guessing. Figure 8-36 shows the mean error performance for 2 features for each of the driver pairs.

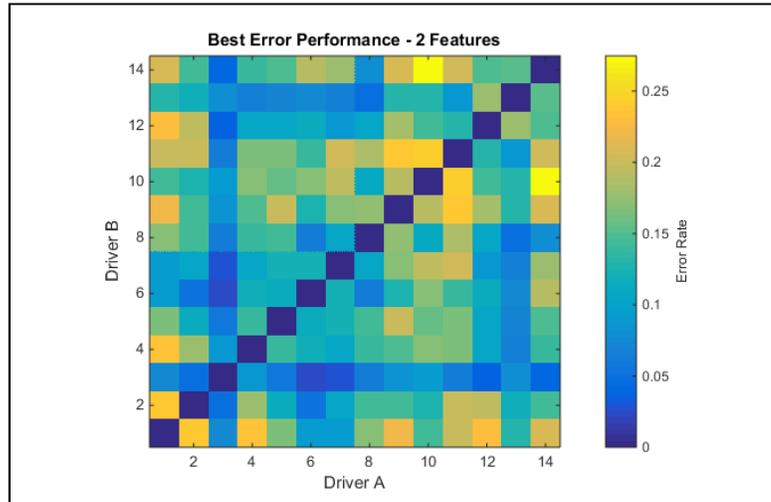


Figure 8-34: Error performance for quadratic LDA classifiers using 2 features showing the best error provided by two features for each driver pair.

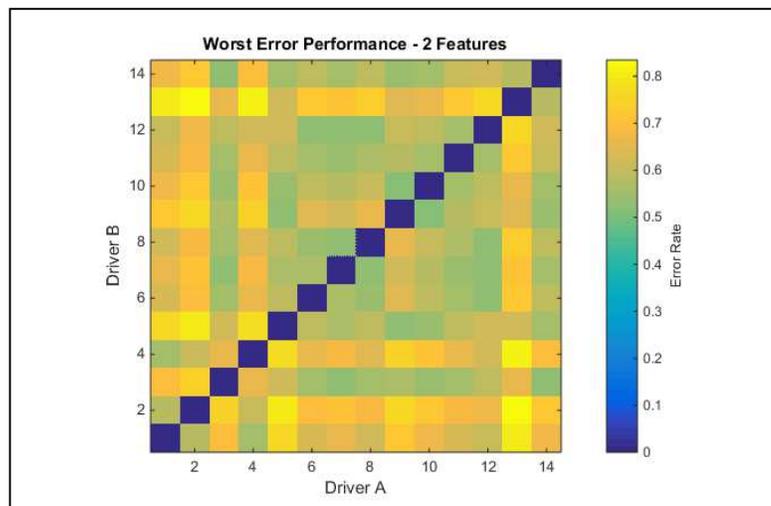


Figure 8-35: Error performance for quadratic LDA classifiers using 2 features showing the worst error provided by two features for each driver pair.

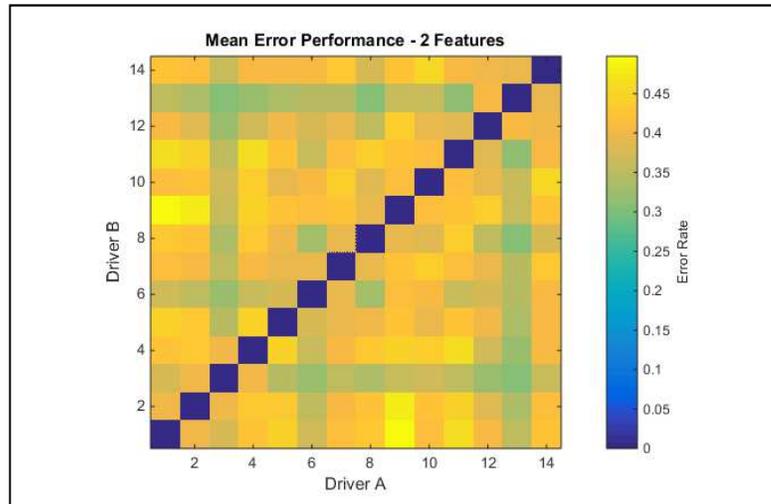


Figure 8-36: Error performance for quadratic LDA classifiers using 2 features showing the mean error provided by two features for each driver pair.

The results for the 2 feature classifiers are presented in Figures 8-37 through 8-39 where the performance is presented on the basis of the feature pairs. In these plots, all features are presented including those that provide no classification ability. In all three figures, the set of yellow lines shows features that failed to provide any classification as single features as older drivers do not exhibit those behaviours. Figure 8-37 shows the best performance from a 2 feature classifier and the darker blue bands show the feature bands that provide the best classification performance. Again, these bands align with the features groups that provided best single feature performance. Figure 8-38 shows the worst performance for each feature pair and the lack of any dark blue indicates that no feature pair provides good performance across all driver pairs. Figure 8-39 shows the mean performance for the feature pairs and this also shows that no feature pair provides strong performance across all driver pairs.

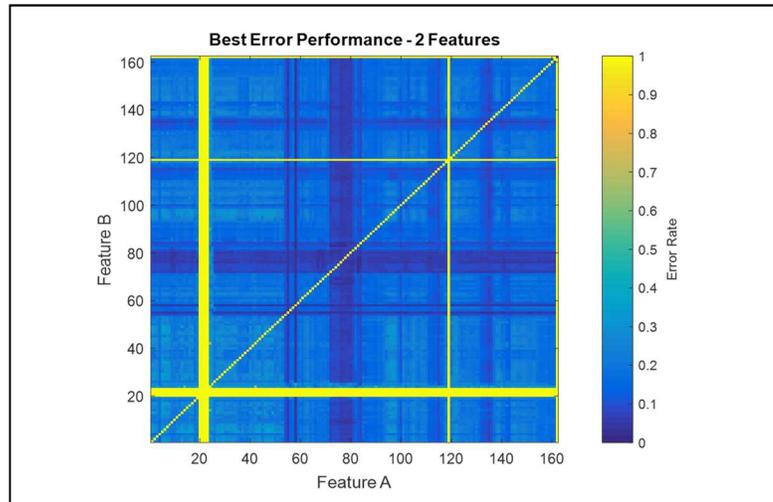


Figure 8-37: Error performance for quadratic LDA classifiers using 2 features showing the best error provided by two features for each feature pair.

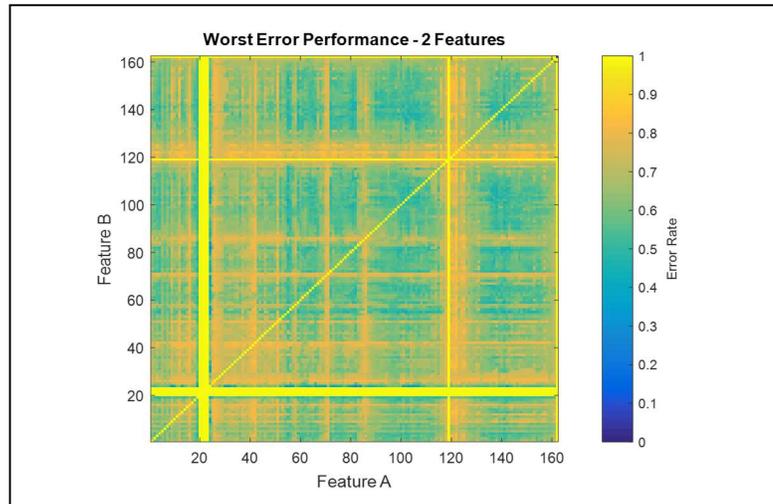


Figure 8-38: Error performance for quadratic LDA classifiers using 2 features showing the worst error provided by two features for each feature pair.

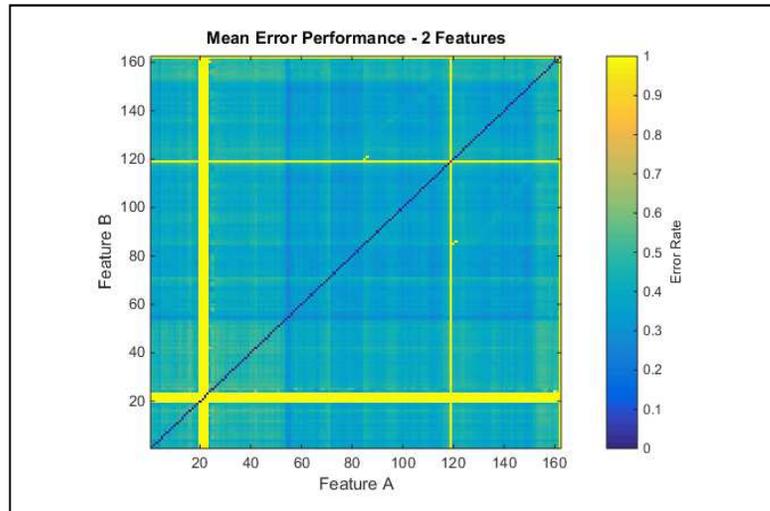


Figure 8-39: Error performance for quadratic LDA classifiers using 2 features showing the mean error provided by two features for each feature pair.

8.3.4. LDA Classification - best performance

The results for the individual features indicates that the best features to distinguish two drivers from each other will be highly dependent on the behaviours of the two drivers and that there will not be a single classifier model, let alone set of classification features, to distinguish any two drivers from each other. The "sequentialfs" capability of Matlab, including the modifications for highly correlated features presented in Chapter 6, were used to analyze each of the 91 possible driver pairs where the best classifier was formed for each driver pair using as many or as few features as needed to achieve the best error performance.

The acceleration and deceleration event features are based on a correlation measure between the events within a given trip and the two phase models for the two drivers. The value of these measures will likely depend on the number of events within a trip so the models were developed using trips with 5 or more event and again for trips with 10 or more events. As for the single feature studies data was segmented so that 10 different 90% train and 10% validate cases were performed using quadratic models.

Table 8-12 and Figures 8-40 and 8-41 summarize the results for the 91 driver pairs showing the ability for a quadratic LDA classifier to distinguish the drivers. The best performance observed has 1.5% error rate across year 1 trips. The



poorest performance was an error rate of over 22.0% with a mean of just over 8.6%. The performance for trips with greater than 10 events was slightly better than trips with greater than 5 events. The figures show that the performance for the classifiers has some variation with only a couple of driver pairs that had poor performance.

Table 8-12: Summary LDA performance for quadratic classifier. Classifier used an adaptive number of features to achieve best possible performance. Effect of number of acceleration and deceleration events in the trips shown.

Error Rate	Acc/Dec events ≥ 5	Acc/Dec events ≥ 10
Best	1.6%	1.5%
Worst	23.5%	21.5%
Mean	8.7%	8.6%
standard deviation	4.4%	4.0%

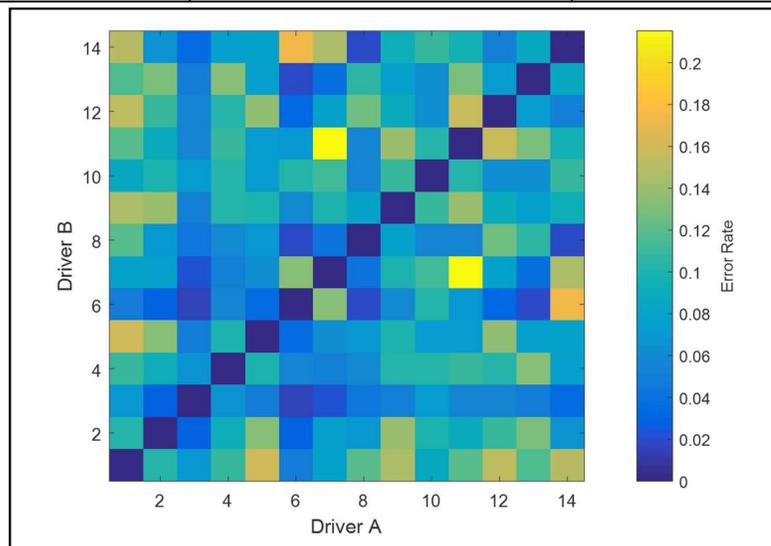


Figure 8-40: Plot of the best-achieved error performance for the 91 driver pairs when trips have at least 10 acceleration and deceleration events.

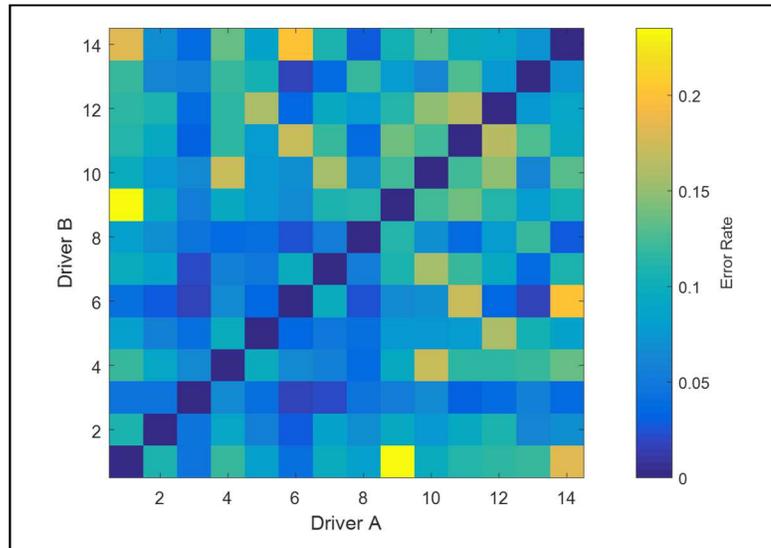


Figure 8-41: Plot of the best-achieved error performance for the 91 driver pairs when trips have at least 5 acceleration and deceleration events.

8.4. Alternative Classifiers

Similar to the classification analysis of the ERP data, the LDA classifier based on discriminant analysis is one of many classification methods that could be applied. So again the 23 different classifiers that are available in the Matlab "*classificationLearner*" toolbox, as shown in Table 6-9, were tested to determine if there was a potential for any performance improvement. Figures 8-40 and 8-41 present 182 potential classifier designs cases and for this section, the focus will be on two cases. The classifiers in Figure 8-40 with the poorest performance (Driver 7 and 11) and with the best performance (Driver 3 and 6) are presented. In each case, the features chosen for the best performance from the LDA classifier were used and the tested classifiers includes those based on decision trees (DT), logistic regression (LR), support vector machines (SVM), k-nearest neighbour (KNN) and ensemble classifiers (EC). For this analysis, the results are presented for a train on 75% of the data and the results for the 25% of the features used for test are shown. The "*classificationLearner*" performs a single train and test, and does not present the aggregate of doing multiple train and tests models such that all cases are used in a test group.



8.4.1. Driver 7 and 11

Driver 7 and 11 proved to be difficult to distinguish from each other as demonstrated by the low accuracy achieved using Quadratic LDA classifiers and a sequential search classification design. These two drivers appear to have very similar driving styles based on the features measured. Figure 8-42 shows the two features that provide the best separation between the drivers. Clearly, there is significant overlap for these two features. The best performing Quadratic LDA classifier used 6 features and these features were provided to all 23 classifiers.

Table 8-13 summarizes the results where 2 classifiers were identified that performed slightly better than the Quadratic LDA with Medium Gaussian SVM improving the accuracy to 82.0% from 79.3%. All three of the classifiers had similar sensitivity and specificity with a general balance of the errors between the two cases.

Figures 8-43 through 8-45 show the scoring results examples for each of the 3 classifiers and as expected the errors in classification are clear and there are also many cases of correct decision for scores close to the 50/50 decision point.

Table 8-13: Comparison of classifier accuracy, sensitivity and specificity along with area under ROC for the best three classifiers to distinguish Driver 7 from 11.

Classifier	Accuracy	True Driver 7	True Driver 11	False Driver 7 as 11	False Driver 11 as 7	Area Under ROC
LDA (Quadratic)	79.3%	153	177	45	41	0.84
Medium Gaussian SVM	82.0%	162	179	36	30	0.89
Bagged Tree	81.5%	161	176	37	40	0.88

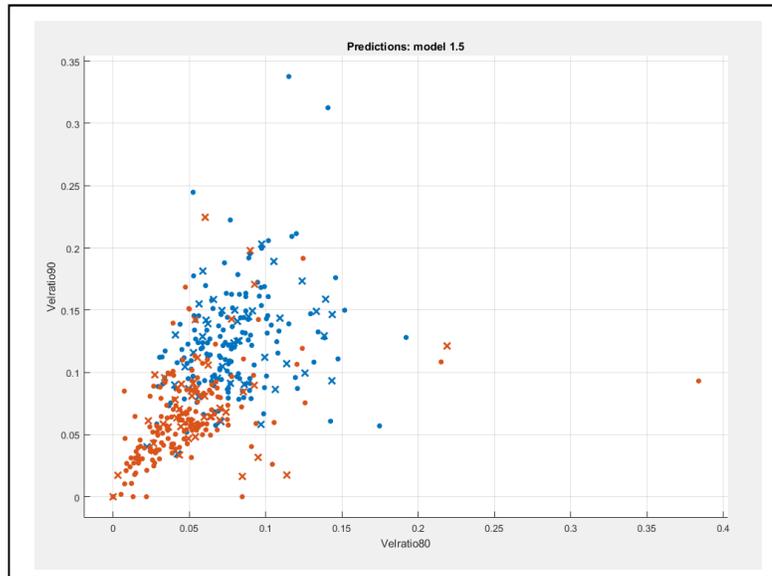


Figure 8-42: Example Scatter Plot for LDA (Quadratic). Orange - classified as Driver 7, Blue - classified as Driver 11. Dot - classified correctly, X - classified incorrectly.

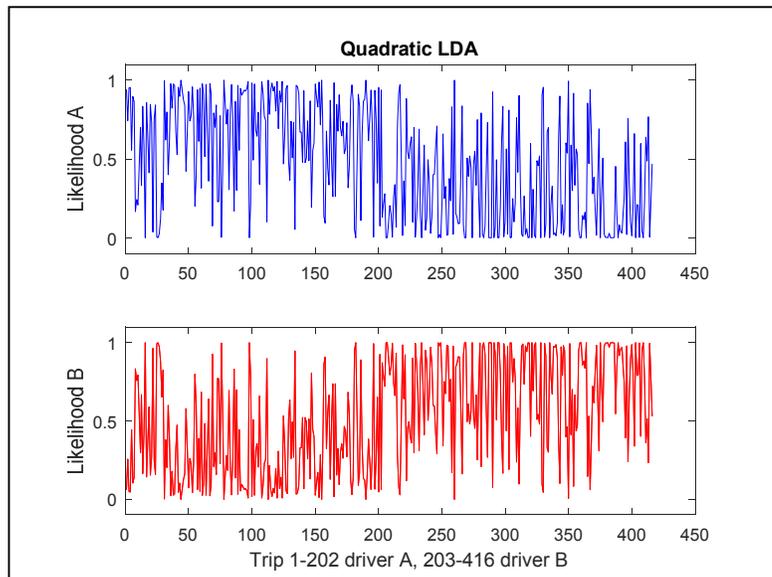


Figure 8-43: Quadratic LDA classifier scoring for driver A=7 and B=11.

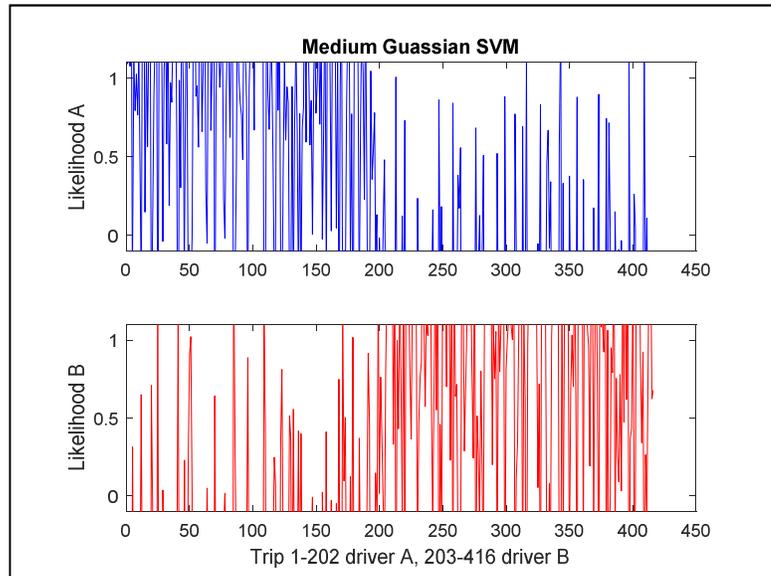


Figure 8-44: Medium Gaussian SVM classifier scoring for driver A=7 and B=11.

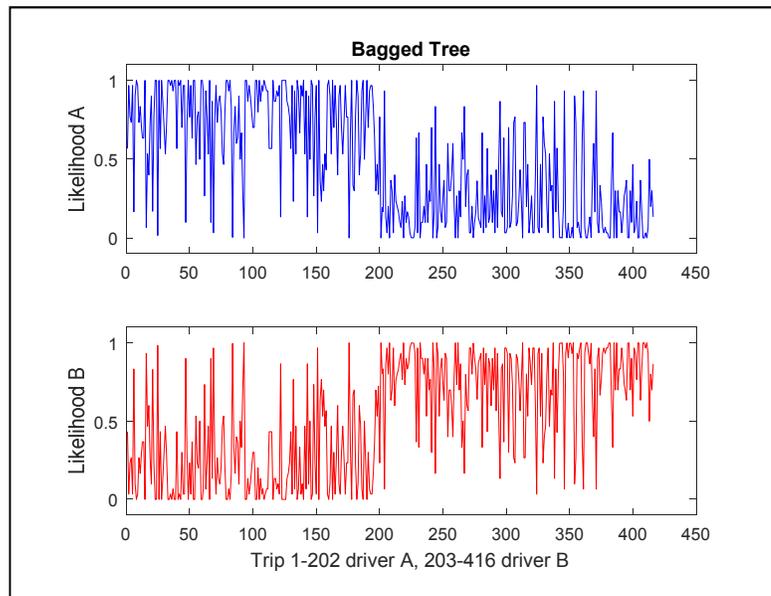


Figure 8-45: Bagged Tree classifier scoring for driver A=7 and B=11.

8.4.2. Driver 3 and 6

Driver 3 and 6 is an example of a pair that could be easily distinguished from each other as demonstrated by the high accuracy achieved using Quadratic LDA classifiers. These two drivers appear to have very distinct driving styles based on the features measured. Figure 8-46 provides an example two features that provide a very clear separation between the drivers. The best performing Quadratic LDA classifier used 8 features and these features were provided to all 23 classifiers.



Table 8-14: Comparison of classifier accuracy, sensitivity and specificity along with area under ROC for the best three classifiers to distinguish Driver 3 from 6.

Classifier	Accuracy	True Driver 3	True Driver 6	False Driver 3 as 6	False Driver 6 as 3	Area Under ROC
LDA (Quadratic)	98.9%	254	208	2	3	0.99
Weighted KNN	98.7%	253	208	3	3	0.99
Subspace KNN	98.7%	254	207	2	4	0.99

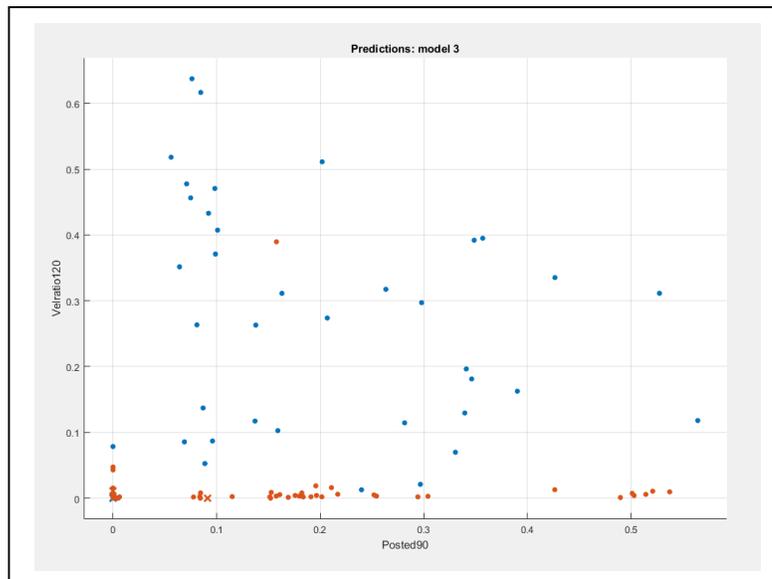


Figure 8-46: Example Scatter Plot for LDA (Quadratic). Orange - classified as Driver 3, Blue - classified as Driver 6 Dot - classified correctly, X - classified incorrectly.

Table 8-14 summarizes the results with no classifiers outperforming the Quadratic LDA. The table does show the next best pair of classifiers that both provided a slightly lower accuracy of 98.7% compared to the best of 98.9%. All three of the classifiers again had similar sensitivity and specificity with a general balance of the errors between the two cases.

Figures 8-47 through 8-49 show the scoring results examples for each of the 3 classifiers with Quadratic LDA showing a clear score with few 50/50 cases while Subspace KNN has many examples near 50/50.

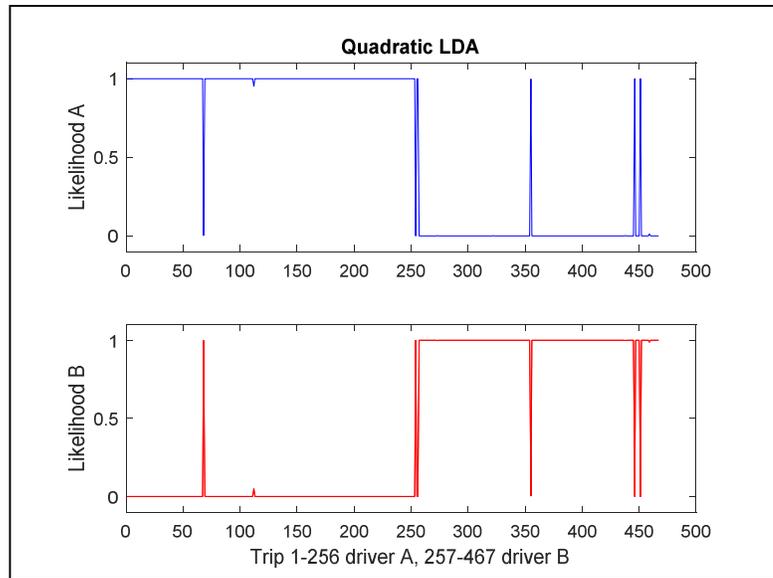


Figure 8-47: Quadratic LDA classifier scoring for driver A=3 and B=6.

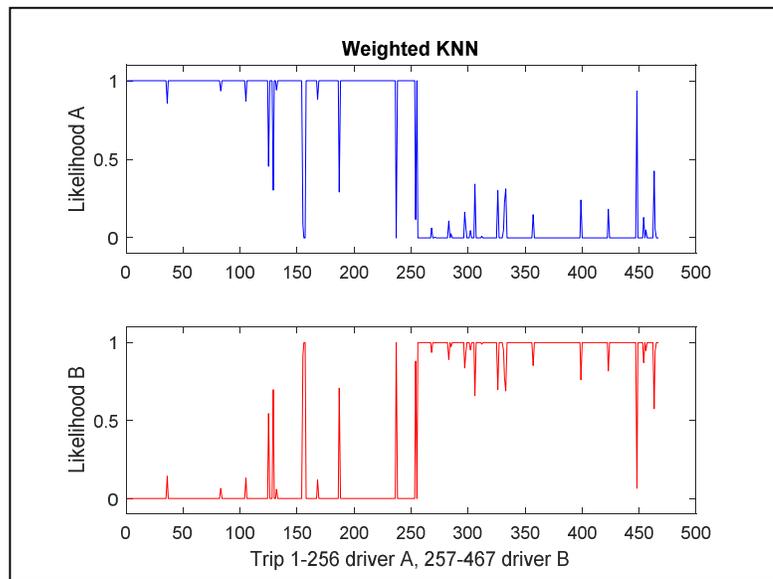


Figure 8-48: Weighted KNN classifier scoring for driver A=3 and B=6.

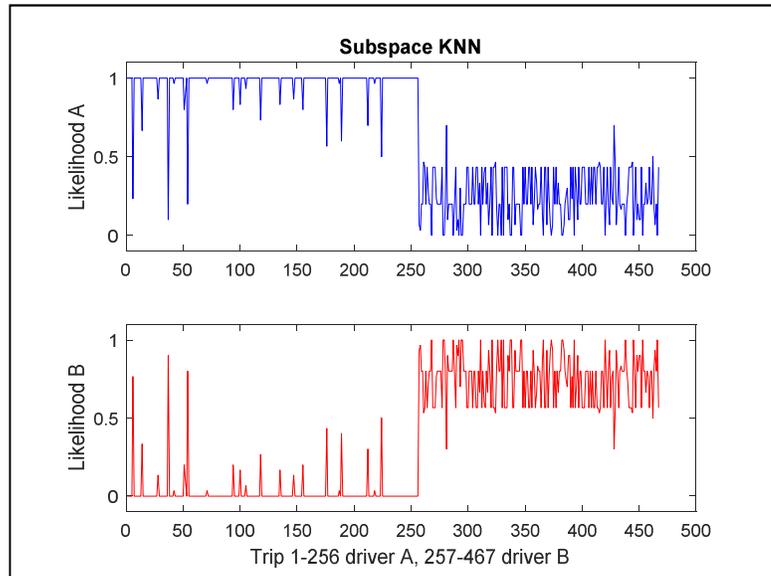


Figure 8-49: Subspace KNN classifier scoring for driver A=3 and B=6.

8.5. Discussion

This section provides an overview of some potential analysis techniques that can be applied to the driving data captured within the Candrive dataset. The results show how the various analysis techniques create features that distinguish the differing driving habits and tendencies of drivers and their behaviours including how they drive, when they drive and road types chosen. Specifically, the analysis of 100 trips by four different drivers shows the potential for road choice and time of day of travel to differentiate between drivers based on the tendency to avoid risks associated with higher speed or driving in the dark. Velocity data and velocity/posted limits indicate the different driver habits including their tendency to drive at or above the posted speed limit. Acceleration analysis provides indications of drivers' habits for fast or slow accelerations.

The work demonstrates that acceleration information derived from GPS position sensors and OBDII velocity sensors provides a valid measurement of acceleration when compared to direct acceleration measurement. This provides the foundation for the use of these signals for the analysis of acceleration and deceleration behaviours of drivers.

This work analyzes the data set for a set of 14 older drivers that have stable general, cognitive and physical health over a one-year period and do not share their



vehicle. The analysis focused on identification of features of driving behaviour that have the potential to allow for the differentiation between two drivers and shows that distance and duration for each trip along with histogram summaries of the actual velocity traveled have potential to differentiate drivers. The time of day of a trip within the solar cycle has limited value to differentiate between drivers. The analysis has shown that these drivers may not have seasonal variation in their driving patterns.

Table 8-15: Summary of behaviour measures for their performance to differentiate drivers.

Features that fail in some or all cases	Best Performing Features
Overnight driving - between 1am and 5am (all drivers) - between 10pm-1am and 5am-8am (some drivers) - dawn driving (some drivers) Acceleration and jerk - highest and lowest acceleration levels (all drivers)	Road type choice - measured through posted limits Velocity - both actual velocity choices and velocity relative to posted limit. Acceleration - through comparison to acceleration and deceleration event models.

Deceleration and acceleration habits of the drivers are shown to have two distinct phases based on the size of the net velocity change within the event. For the lower velocity net changes, the mean and min/max acceleration values within the event change with the size of the velocity drop while for larger net velocity changes, the attributes are much more stable and only change slightly as the size of the velocity increases. This pattern has been shown for all 14 drivers within the group, while it has also been shown each that each driver has distinct features.

The performance of the two phase relationships was then tested through comparison with specific trips for the driver, and it was found that the performance reached a maximum performance providing the trip has at least 8 deceleration events. At this level of performance, the minimum measure was better at distinguishing drivers than that of the mean measure and combined with the greater variation on the minimum relationships between the various drivers provides the potential for this feature to be used to differentiate between two drivers. It is also



possible that changes in a driver's behaviours could provide an indication of cognitive or physical health change.

Machine learning classification tools were evaluated for their ability to distinguish between gold standard reference driver pairs. LDA classifiers trained for each of the 91 gold standard pairs had a mean error rate of 8.6% and a best error rate of 1.5%. Other machine learning algorithms provided some performance improvement with limited impact for the driver pair with the best error performance and a 2.7% improvement for the lowest error performance pair. The potential remains for the identification of additional driving behaviour features that could improve the performance for these challenging cases. One example would be more detailed exploration of behaviours of the drivers under particular driving conditions such as features derived for driving on only city streets, only on expressways or during particular weather events such as rain or snow.

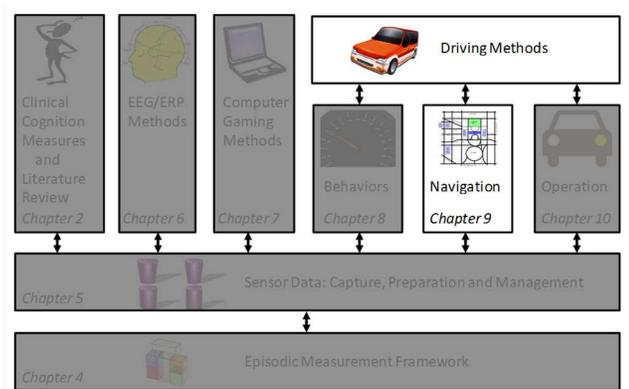
This work augments previous works that focus on identification of specific driving behaviours through the association of behaviours with particular drivers. In insurance applications, features that measure risk related behaviour or diversity of drivers for a vehicle could lead to increased or decreased premiums.

Future work on driver unique behavior measurement includes identification of additional driving behaviour features to improve the performance, especially for driver pairs that prove to be difficult to distinguish based on the already identified driving features. In addition, the future work should explore drivers that are known to having changing health (cognitive, physical or general) and determine if the identified behaviour feature measures are stable or change in conjunction with the health changes.





Chapter 9: Driving: Navigation



Objective: This chapter explores methods to measure the navigational abilities of a driver as a measure of executive cognitive function. This required the creation of methods to measure and compare a driver's navigational and trip planning decisions through comparison against navigational references generated *a posteriori* from the trip data. Topics covered include:

- GPS route information analysis - Analysis of the GPS sensor driving routes to determine trip events including stop detection, driving distance.
- Navigational reference generation - To provide a navigational performance reference measure, crow flight and GIS navigational routes for trips retroactively generated automatically.
- Optimal trip identification - a performance reference generated from navigational information for multi-stop trips.
- Performance measurement - as driven information compared with segment and multi-stop trip level references to provide measure of driver performance.
- Trend measurement - measurement of driver performance and patterns over time to identify changes in driving behaviours and ability over longitudinal study.



9.1. Method

The widespread and cost effective availability of GPS technology provides a sensor mechanism to measure, track and record vehicle movements. An iPhone 4GS running a GPS tracking application [134] logs the real-time GPS information (time, latitude, longitude, velocity, bearing, altitude) to a network server. A network based GPS system provides the ability for trip information to be automatically centrally stored and analyzed with an option for the results to be conveyed to relevant parties. Local storage on the device requires subsequent retrieval (as in Candrive) delaying analysis but eliminating the need for network access and the risk of data loss when the network is not available. The iPhone application allowed GPS data to be collected without driver interaction.

This section reports on data collected from healthy volunteers from the research team that drove to a series of familiar locations. The trips included efficient, although not necessarily optimal, routes that can be used to establish the baseline performance, and various inefficient routes (such as backtracking to home location) to demonstrate various coping mechanisms.

The collected data is analyzed through a series of steps:

- Stop filter: GPS measurement error and noise appears as jitter in the position. When the vehicle is not moving (velocity = zero), the data contains multiple measurements for a single location that can be found through an averaging (low pass filter) per equation (9-1).

$$\begin{aligned}
 & \text{if } velocity(i:i+n) == 0 && [9-1] \\
 & \quad lat(i:i+n) = mean(lat(i:i+n)) \\
 & \quad long(i:i+n) = mean(long(i:i+n))
 \end{aligned}$$

This removes potential errors in the detection of vehicle turns just prior to or after a stop caused by short distance noise related segments.

- Latitude and longitude data traces are then analyzed using the algorithm shown in (5.2) to identify vehicle stops that may be traffic related or actual destinations. A threshold of 2 minutes was used to distinguish traffic related stops from destinations.



[9-2]

```

if (lat(i:i+k)==constant
    and long(i:i+k)==constant)
or velocity(i:i+k)==0
    stoptime= time(i+k)-time(i)
    case stoptime
        >60 minutes → long errand stop
        >5 minutes → medium errand stop
        >2 minutes → short errand stop
        otherwise traffic related stop
  
```

- The GPS data is then analyzed for the actual distance driven for each of the trip segments. The distance between location points is calculated using the Haversine formula for great circle paths shown in (5.3). Providing actual distance driven for each trip segment per (5.4).

[9-3]

```

for each longitude-latitude point pairs
    dlat=lat(i+1)- lat(i)
    dlon=long(i+1)- long(i)

a = sin2(dlat/2) + cos(lat(i))*cos(lat(i+1)) *sin2(dlon/2);
c=2*atan2(sqrt(a),sqrt(1-a))

distance=R*c; // Where R=6371km (mean earth radius)

segmentdistance=∑distances for point pairs
  
```

[9-4]

- Route options: Based on trip destinations a set of five alternative route options were identified. Crow flight distances - direct minimum distance between the destinations is calculated without reference to the road network. Road network distances - used actual road network optimal routing. The Google maps API [138] uses Dijkstra's algorithm to search the known road network for optimal routing between locations providing a gold standard reference. Crow flight and Google Maps derived navigation provide two different independent comparators with Crow flight being very simple to



calculate but failing to include the actual road network restrictions; hence typically underestimating travel time and distance.

- Crow flight - as driven: For destinations in the order chosen by the driver – overall and segment distance using (9-3).
- Crow flight - optimal destinations order: Keeping the initial location fixed, the balance of the destinations are placed in optimal order to minimize the overall direct distance for trip. A search of all the potential orders for the minimum distance was used for this travelling salesman problem.
- Google maps API - as driven: Crow flight distances do not deal with road network limitations, so Google maps [138] API was used to automatically calculate the road network based distance reference for the stops in the order driven. A URL containing the destination locations is automatically formed and the API response provides a table for A→B and B→A distances and times which allows the distance for the as driven route.
- Google maps API - optimal distance order: The Google maps API response includes distances between all the destinations in both directions allowing for asymmetry in the road network. This gold standard reference allows the optimal (minimum distance) order to be found again using a travelling salesman search.
- Google maps API - optimal distance order: The Google maps API provides the estimated gold standard travel times. These times are used to calculate the optimal (minimum time) order for the destinations.
- These destination order sets and trip attribute measures are compared with the actual trip attributes, and variation from the reference routes provides an indication of the overall route planning capability for the individual. It is not expected that individuals will perform optimally, but that initial performance will give a baseline indication of their ability and typical choices.
- The segment distances as driven are compared to the references for those segments, providing an indication of the driver's navigation capability. Again



it is not expected that individuals will perform optimally but that initial performance will provide a baseline.

- Establish baseline: A set of trips were identified as representative of baseline performance and were used to create reference attributes. From the trips the following baseline attributes are determined:
 - Baseline trip summary - Mean and standard deviation each of (3-10)
 1. Number of stops
 2. Stop location (longitude and latitude)
 3. As-driven total distance
 4. Crow flight as-driven total distance
 5. Crow flight optimal total distance
 6. Google as-driven total distance
 7. Google optimal total distance
 8. Google time optimal total distance
 9. Segment ratio - As-driven segments / Crow flight segments
 10. Segment ratio - As-driven segments / Google segments
 - Comparison to Baseline: Measure to compare to baseline:
 - Trip complexity change:
 - increase or decrease in the number of stops
 - presence of each baseline destination (within 100m of baseline stop)
 - identification of additional (non baseline) stops
 - Trip distance change:
 - Increase/decrease in overall trip distance: actual change, percentage of baseline distance and relative to baseline trip standard deviations.
 - Trip level planning and navigation performance: comparing trip overall distance driven using the 5 reference trips to the baseline trip with



percentage of baseline and number of standard deviation difference calculated.

- Segment level detailed navigation performance: comparing the segment level distances for the trip under consideration against the Crow flight and Google Maps reference distances for those segments with percentage of baseline and number of standard deviation difference calculated.

Trip complexity provides an indication of driver choice such as a reduction in the number of destinations or variety of destinations, indicating a coping mechanism. Similarly trip distance change will indicate drivers coping by staying closer to home. Trip level planning provides an indication of high level cognitive ability to determine an efficient order for destinations while segment level navigation shows the ability of the driver to get between two locations efficiently.

9.2. Experimental Results

This research problem includes 3 major steps:

1. Collection and analysis of the GPS data for each trip.
2. Calculation of baseline reference trip features and attributes.
3. Comparison of trip with baseline to identify performance changes.

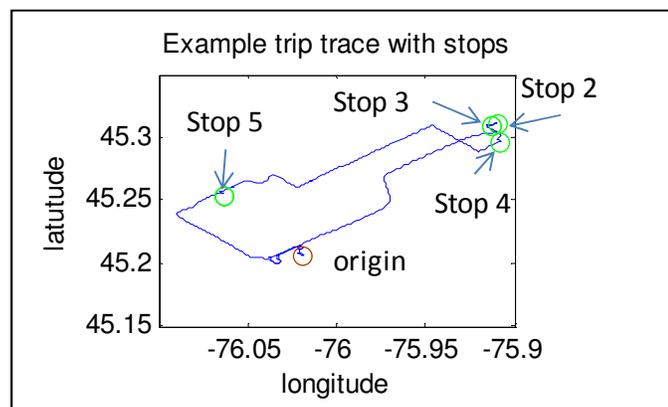


Figure 9-1: Example route trace for a training trip with stops identified in green and trip origin location in red. Trip starts at origin (stop 1) and follows through stops in order 2 through 5 and then returning to the origin.



9.2.1. Trip Analysis

Driver 1 drove three trips in another region of the city to provide data for algorithm development with validation against the 10 trips.

Figure 9-1 shows an example plot of the de-jittered GPS data per equation (1) for a training trip. The trip began at the red circle (origin – stop 1) with four destinations (stop 2 through 5) returning to the trip origin.

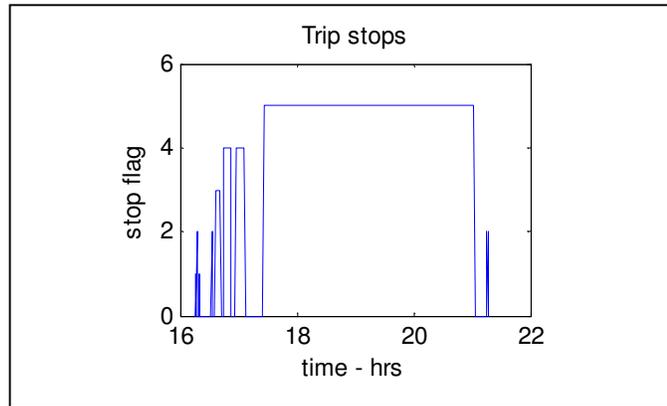


Figure 9-2: Example stop identification for training trip showing the 4 stops with the first being between 2 and 5 minutes, stop 2 and 3 between 5 and 60 minutes and a final stop of greater than 60 minutes. The diagram also shows slow movement and short duration stops related to traffic.

stop_flag = 0 - not stopped ; 1 - <3km/hr;
 2 - <2 min; 3 - >2min and <5min;
 4 - >5min and <60min; 5 - >60min.

Table 9-1: Example route optimization analysis for training trip

Case	Distance (km)	Stop Order
As-driven	54.3	1, 2, 3, 4, 5
Crow flight as driven	35.9	1, 2, 3, 4, 5
Crow flight optimal	34.9	1, 4, 2, 3, 5
Google as driven	54.5	1, 2, 3, 4, 5
Google optimal distance	53.4	1, 4, 3, 2, 5
Google optimal time	53.4	1, 4, 3, 2, 5

Figure 9-2 shows an example time line for the same trip with the vehicle stops and slow moving segments (9-2) identified. The four stops (destinations) are clearly visible with 3 short initial stops, one less than five minutes and 2 longer than five



minutes around 17:00 and one long duration stop. There are also a number of traffic related stops or slow movements correctly classified.

Table 9-1 shows an optimization analysis for the overall trip with the actual trip distance and each of the five references. The as-driven distance is slightly shorter as Google calculates distance using more precise measures of road curves compared to 5 second sampled GPS data with piecewise linear approximation.

Table 9-2 shows the analysis for the training trip on a per segment basis showing the measured travel distance and time for each trip segment in comparison to the Crow flight, Google optimal route and optimal time. Although the Google distances are similar to the as-driven values, the time differs as Google routing assumes speeds on some rural roads that differ significantly from posted limits, but this will be consistent between trips on those roads.

For this training trip, the results show that a trip can be measured, identifying the stops and route driven, and the results can be compared with derived optimal routes using crow flight and Google maps information. The road network for this trip requires some back tracking near the trip origin and this is not detected in the crow flight analysis but is noted on Google routes. In this case, the Google data finds a small improvement would occur with a reordering of the destinations. It shows that trips can be analyzed to create a baseline driver performance relative to gold standard navigation information from Google maps.

Table 9-2: Example comparison of segment trip information with crow flight and Google

Origin stop	Destination stop	As-driven (km)	Crow flight (km)	Google route (km)	As-Driven (min)	Google route (min)
1	2	21.26	14.48	21.19	19.80	19.93
2	3	0.56	0.46	1.14	2.50	3.85
3	4	2.01	1.59	2.24	4.92	5.17
4	5	17.30	13.03	17.31	18.82	22.92
5	1	12.71	6.34	12.62	14.82	18.28



Table 9-3 summarizes the trip data collected for a series of trips to typical destinations (shopping, recreation, and library) within a region of the city. Trips were repeated over a number of days for a total of 10 trips.

A trip with 6 destinations (including start point) was repeated 6 times with an example trace shown in Figure 9-3 for Trip 1. Four trips were performed with the stops in near optimal order and efficient navigation between the destinations. Two additional trips were captured for those same 6 stops to provide examples of backtracking to home (Trip 5) - common coping mechanism, a trip (Trip 6) that has a small routing error (wrong turn leading to a quick U-turn on a side street). Trip 7 and 8 capture two other expected variations in trips where an extra stop (Trip 7) was included or a stop was skipped (Trip 8) indicating trip variation.

Table 9-3: Summary of the driving trips captured

Trip	Use	Attributes	Driver
1	Baseline	6 stops, clockwise	Driver1
2	Baseline	6 stops, clockwise	Driver1
3	Baseline	6 stops, counter-clockwise	Driver1
4	Test	6 stops, clockwise	Driver2
5	Test	6 stops, clockwise, backtracking to home	Driver1
6	Test	6 stops, counter-clockwise, small routing error	Driver2
7	Test	7 stops, counter-clockwise	Driver2
8	Test	5 stops, counter-clockwise, stop 4 skipped	Driver2
9	Test	5 stops, stop 2 from baseline + 3 new stops, navigation errors	Driver1
10	Test	5 stops, same as trip 9, no navigation errors	Driver1

The performance of the stop detection algorithm is provided in Table 9-4 and shows that all 58 stops were detected successfully by the algorithm with no false positives or negatives. The algorithm was able to correctly reject all traffic related stops including extremely slow moving traffic (road construction in trip 1). Three



example route traces for the validation trips are shown in Figures 9-3 – 9-5. Figures 9-3 and 9-4 show the trip with stops in optimal order and optimal routing between. Figure 9-3 had the stops in a clock-wise order while Figure 9-4 had the stops in a counter clock-wise order. Figure 9-5 has the stops in an optimal order but routing between includes backtracking to home.

Table 9-5 summarizes the distance measurements for the validation trips. The as-driven distances from the GPS data correspond closely to the Google as-driven data as expected. The Google data underestimates the total distance as it does not include any driving that was not road network restricted, such as movements through parking lots for each stop. The Crow flight references for both as driven stop order and optimal stop order grossly under estimates the total trip distance as expected.

Table 9-4: Summary of performance of the stop detection algorithm for all stops within the validation data set. No stop detection errors were observed.

Attribute	Count
Total stops on trips	58
Stops found by algorithm	58
Extra stops found (false positive)	0
Stops missed (false negative)	0

Both of the Google optimal measures provide good references for the as driven trip. They both underestimate the total travel distance as both drivers chose road route options that were close to the shortest time route over options that were shorter in distance. The challenge with the use of travel times as a method to detect slower driving as a coping mechanism was demonstrated in Trip 1 where travel was greatly slowed for one segment because of construction delays. Although the travel distance was not affected, the travel time was greatly affected. Also variability in traffic lights could obscure any driver related variance.

Two final trips were captured that were of similar complexity (number of stops) to the baseline trip but included stops in a different direction from the home location. This provides examples of trip variation. In one case (Trip 9), the route included navigation errors, while Trip 10 used efficient routing.

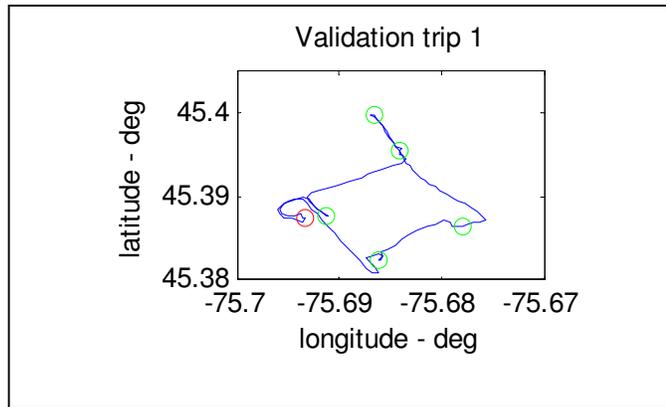


Figure 9-3: Stop detection for Validation Trip 1 with direction of travel in clock-wise direction. Trip origin (red circle) and stops (green circle).

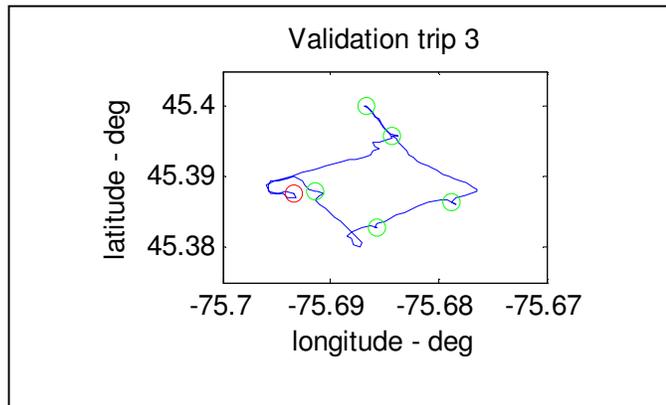


Figure 9-4: Stop detection for Validation trip 3 with direction of travel in counter clock-wise direction. Trip origin (red circle) and stops (green circle).

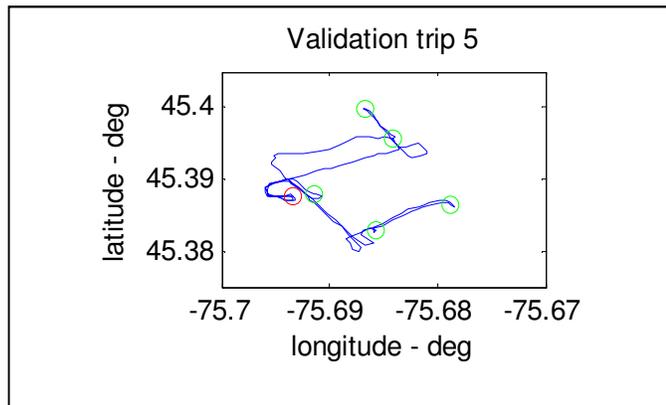


Figure 9-5: Stop detection for Validation trip 5 with direction of travel in clock-wise direction. Trip origin (red circle) and stops (green circle).



Table 9-5: Summary of the distance measurements for validation trips. Overall mean and standard deviation is provided and the mean and standard deviation for the difference between the measure and the trip as driven.

Measure	Optimization	Mean (km)	Std Dev (km)	Delta Mean (km)	Delta Std Dev (km)
Trip	as driven	8.7	1.8	---	----
Crow flight	as driven	5.2	0.9	3.4	1.4
Crow flight	distance optimal	5.1	0.9	3.6	1.6
Google	as driven	8.3	1.0	0.4	1.1
Google	distance optimal	7.8	1.0	0.8	1.7
Google	time optimal	8.3	0.7	0.4	1.7

9.2.2. Baseline Trip Analysis

Table 9-6: Baseline trip summary attributes and features.

Baseline Attribute	Units	Mean	St Dev	Mean	St Dev
Number of stops on trip		6			
Stop 1 longitude/latitude	degrees	-75.6935	6.11E-05	45.3875	4.16E-05
Stop 2 longitude/latitude	degrees	-75.6913	8.81E-05	45.3878	8.85E-05
Stop 3 longitude/latitude	degrees	-75.6840	4.48E-05	45.3957	1.05E-04
Stop 4 longitude/latitude	degrees	-75.6867	1.11E-04	45.3999	1.33E-04
Stop 5 longitude/latitude	degrees	-75.6784	2.76E-04	45.3862	1.39E-04
Stop 6 longitude/latitude	degrees	-75.6859	1.39E-04	45.3826	1.39E-04
As driven distance	m	7559	263		
Crow flight distance	m	4836	116		
Crow flight optimal distance	m	4719	49.7		
Google as driven distance	m	7742	227		
Google optimal distance	m	7470	64.1		
Google time optimal distance	m	7470	64.2		
Segments as driven/crow flight	m/m	2.00	1.36		
Segments as driven/Google	m/m	0.971	0.0389		

Trips 1, 2 and 3 were chosen as the baseline, leaving the 7 trips to be analyzed. In practice, it would be expected that a larger number of trips for a given subject would be used to generate the baseline. The training trips included travel in both a clockwise and counter clockwise path between the destinations, which has minimal effect on route as compared to optimal route and represents a driver choice. The summary analysis of the 3 baseline trips is shown in Table 9-6, where key summary features have been identified along with measures of the variation in these features. The crow flight features with their large variation from actual travel distance show the limitation of this reference.



9.2.3. Validation Trip Analysis

Figures 9-5 and 9-6 show the trip routes for trips 5 and 7 respectively, and the backtracking in Trip 5 to the start location is clearly visible. This is a common navigational coping mechanism for patients with cognitive decline through use of familiar routes. The additional stop in Trip 7 is in the lower right hand corner.

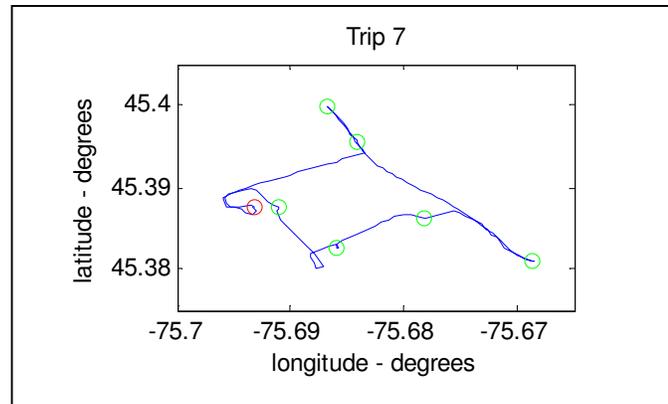


Figure 9-6: Trip trace for validation Trip 7 showing travel path and stops. Trip includes an extra stop. Trip origin (red circle) and stops (green circle).

Table 9-7 shows the comparison of the 3 baseline trips to the baseline attributes for each of the proposed measurement criteria while Table 4 shows the results for the 7 validation trips. The broad range of measures was used so that attributes and measures that were more effective at identifying changes could be determined. As expected, the training trips are all in close agreement with the baseline trip, but the table does give an indication of the size of variation that can be observed in the measures for similar trips.

The comparisons to crow flight routing was problematic as expected. Table 9-7 shows that 50% and greater variation, and as a result this measure is discarded from further exploration.

Actual travel time and a comparison to optimal travel time was also determined to be an ineffective measure as the actual travel time for city travel has many effects that are beyond a drivers control such as traffic congestion, traffic lights and other controls and lastly construction delays. None of these reflect driver ability but cause variation in the measured data.



Table 9-7: Analysis of the 3 baseline trips as compared to the baseline summary trip.

Attribute \ Trip		Base 1	Base 2	Base 3
Actual stops		6	6	6
trip stop difference from baseline		0	0	0
Number of baseline stops missed		0	0	0
Baseline stop missed	stop #	none	none	none
Count of non baseline stops		0	0	0
Overall trip delta from baseline	m	165	138	-304
Overall trip distance change	%	2.2%	1.8%	-4.0%
	#sd	0.63	0.53	1.15
Trip variation from Google as-driven route	%	-0.23%	-0.58%	-6.28%
	#sd	0.08	0.20	2.14
Trip variation from Crow flight as-driven route	%	60%	59%	50%
	#sd	24.8	24.6	20.8
Trip variation from Crow flight optimal route	%	64%	63%	54%
	#sd	60.4	59.9	51.0
Trip variation from Google distance optimal route	%	3.41%	3.05%	-2.87%
	#sd	3.97	3.55	3.34
Trip variation from Google time optimal route	%	3.41%	3.04%	-2.87%
	#sd	3.96	3.54	3.34
Segment variation from baseline crow flight	%	0.59%	0.68%	-1.26%
	#sd	0.01	0.01	0.02
Segment variation from baseline Google segments	%	-0.09%	0.37%	-0.28%
	#sd	0.022	0.092	0.069

Table 9-8: Summary of the analysis results for the 7 validation trips as compared to the baseline trip formed from trips 1, 2 and 3.

Attribute \ Trip		Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10
Actual stops		6	6	6	7	5	5	5
trip stop difference from baseline		0	0	0	1	-1	-1	-1
Number of baseline stops missed		0	0	0	0	1	4	4
Baseline stop missed	stop #	none	none	none	none	4	3 4 5 6	3 4 5 6
Count of non baseline stops		0	0	0	1	0	3	3
Overall trip delta from baseline	m	-266	4904	482	1596	-1266	3060	1461
Overall trip distance change	%	-3.5%	64.9%	6.4%	21.1%	-16.7%	40%	19%
	#sd	1.01	18.63	1.83	6.06	4.81	11.6	5.5
Trip variation from Google as-driven route	%	-5.80%	60.98%	3.87%	18.25%	-18.71%	37%	17%
	#sd	1.98	20.78	1.32	6.22	6.37	12.7	5.62
Trip variation from Crow flight as-driven route	%	51%	158%	66%	89%	30%	120%	87%
	#sd	21.1	65.5	27.5	37.1	12.5	49.7	35.9
Trip variation from Crow flight optimal route	%	55%	164%	70%	94%	33%	125%	91%
	#sd	51.8	155.7	66.8	89.2	31.7	119	86.5
Trip variation from Google distance optimal route	%	-2.36%	66.85%	7.66%	22.56%	-15.74%	42%	21%
	#sd	2.75	77.87	8.92	26.28	18.34	49.1	24.2
Trip variation from Google time optimal route	%	-2.37%	66.84%	7.65%	22.56%	15.75%	42%	21%
	#sd	2.75	77.76	8.90	26.24	18.32	49.0	24.1
Segment variation from baseline crow flight	%	-12.7%	25.33%	0.48%	-7.33%	9.61%	3.3%	0.31%
	#sd	0.19	0.37	0.01	0.11	0.14	0.40	0.0046
Segment variation from baseline Google segments	%	-5.48%	41.72%	1.00%	-2.67%	-4.83%	10.1%	3.2%
	#sd	1.37	10.40	0.25	0.67	1.20	2.52	0.79



Trip 4 was driven by a different driver and although the training trip route was followed exactly, the driver had small variations in decisions such as parking locations within lots at destinations. The resulting trip was within baseline range for comparative trip level attributes, but showed segment level variations.

Trip 5 includes a large amount of backtracking and this is clearly visible in most of the attributes for this trip. Trip 6 included a small navigational error that is detected in the features even though this effect is minimal.

Trip 7 included an additional stop that has been detected by the stop level analysis, so although this leads to variation in many of the trip level attributes, the additional stop allows it to be distinguished from a trip with poor routing decisions alone. The segment level variation is comparable to the variation seen in Trip 4, providing an indication of trip to trip expected variation and not an indication of a performance decline trend. Trip 8 excluded stop 4 and this missing stop has been identified but again the trip level analysis shows changes relative to the baseline that are much smaller scale than for backtracking for segment level comparison.

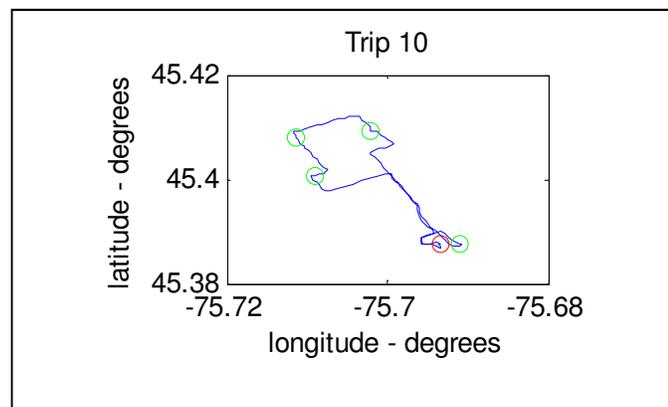


Figure 9-7: Trip trace for validation Trip 10 showing travel path and stops. Trip used same home location and initial stop (lower right corner) as trips 1 through 8. It then included 3 additional stops in another area of the city near the home locations.

Trips 9 and 10 clearly indicate the significant changes in destination set and as expected, this resulted in large changes in the overall trip measures given that the trips are so different from the baseline trip. The segment level analysis for Trip 10 shown in Figure 9-7 shows again similar variation to that observed in Trips 4, 7 and



8 indicating expected variations that could be seen in trips over time whereas the segment level analysis for Trip 9 clearly indicates the navigational errors.

9.3. Discussion

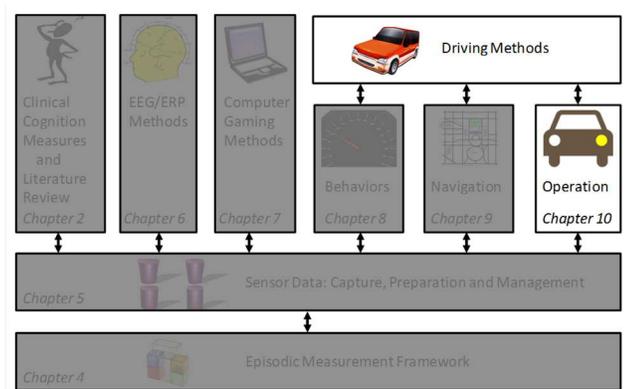
This section proposes and validates algorithms for the measurement of driving navigation behaviours as an example of an instrumental activity of daily living that requires high level cognitive function. The work specifically provides an algorithm for the detection of destination stops within a GPS trace for a trip that operated without error. Google Maps is used to provide gold standard reference measures for as-driven and optimal travel distances for identified destinations.

The work showed that use of Crow flight measures does not provide a useful reference. It demonstrates that the navigational decisions for drivers can be detected through sensing and analysed to identify changes in navigational performance. Similarly changes in trip complexity, including number and variety of destinations chosen by the driver can be detected. A driver's baseline performance measurement is proposed that provides a reference for the driver's ability. Google Maps distance is shown to be an effective gold standard reference for navigational performance while a Crow Flight reference is shown to be ineffective.

With the nature of many trips taken, there is a high repetition in destinations (work, shopping, family, social activities) and this work shows that these trips can be analyzed to form a summary performance set of features. The work then shows that new trips can be compared against this baseline and trips with normal variations of destination choice can be distinguished from trips that are indicative of the coping mechanisms typical for cognitive decline.



Chapter 10: Driving: Vehicle Operation



Objective: This chapter explores methods for the measurement of specific vehicle operation tasks performed by a driver as a measure of an over learned cognitive task. The example explored is the use of turn signals with the goal to have methods be able to measure use of turn signals by the driver along with methods to determine when they were needed leading to a possible performance measurement. Topics covered include:

- Turn signal use - detected through image processing and event analysis.
- Actual vehicle turns - detected through analysis and event detection of as-driven GPS location data.
- GIS turn augmentation - use of GIS information to retroactively identify turns and merges in a trip using GIS navigational databases.
- Sensor fusion - merging GPS and GIS turn information with turn signal use events requiring time alignment of the three sources.
- Performance measurement - Providing a measure of driver ability.

10.1. Method

The measurement of turn signal use as an indication of autonomous cognition while driving requires sensors to detect both signal use and the events to signal. The chosen sensors consisted of dashboard video of the signal lamps and a smart phone based GPS sensor. The sensor data are combined with navigation analysis from chapter 9 to create as-driven GIS turn information from Google Maps in



addition to direct GPS path analysis. The dashboard video had to accommodate dashboards that had the turn signal lamps widely separated (two cameras) and those with closely placed lamps (one camera). CANbus signal information uses manufacturer proprietary protocols and could not be used. Signal lamp voltage was also rejected to avoid potential impact to operation of the signals and the need for no permanent modification on the vehicle.

10.1.1. Dashboard Signal Lamp Detection

Identification of the turn indicator status is an example of a general image processing problem where the event being detected has known features (arrow shape in this case) but unknown size/scale, position, orientation (tilted or not coplanar with camera), noise (vibration) and the detectable feature (lamp lit) repeats, indicating a longer duration event (signal engaged). The algorithm requires the analysis of the video to detect that the lamp and signal are “on” between the first detected “on” transition through to the last detected “off” transition. The algorithm must adapt to the location of the signal lamps on the dashboard, support variation in dashboard lamp placement (one camera or two), it needs to distinguish arrows from other signal lamps and determine the direction of the arrow. Grey scale converted video images (Figure 10-1 shows source image examples) represented by unsigned integers were used in the work and analysis was performed on forward and backward 5 frame difference images and associated measures of the variance and energy in the delta images with the frame gap allowing time for lamp to transition on or off.

$$ForDelta = Image[n] - Image[n - 5] \quad [10-1]$$

$$BackDelta = Image[n - 5] - Image[n] \quad [10-2]$$

$$ForVar = Var (ForDelta) \quad [10-3]$$

$$BackVar = Var (BackDelta) \quad [10-4]$$

$$ForEnergy = \sum ForDelta^2 \quad [10-5]$$

$$BackEnergy = \sum BackDelta^2 \quad [10-6]$$



Regions in the image were identified by looking for localized areas of high energy in the delta images:

$$\textit{Turning on: } ForVar > 150 * ForEnergy \quad [10-8]$$

$$\textit{Turning off: } BackVar > 150 * BackEnergy \quad [10-9]$$

The location of the lamp was observed to move within the image through vibration or by bumping causing a step move in position. A candidate window is chosen in the region with highest energy. Given that the signals cannot be guaranteed to be horizontally aligned with the camera frame, the correlation is measured for the candidate arrow from -15° to $+15^\circ$ rotations in 1° steps. At each rotation of the candidate arrow, the image is cropped based on Matlab canny edge detection and the correlation of candidate arrow with ideal left and right arrows scaled to same size as candidate is measured. If maximum correlation of left vs. right is not more than 2% different, candidate is rejected as not an arrow removing lamps other than arrows. Turn signal state is then created by processing the lamp on-off states where the signal is on between the first lamp “on” detection to a lamp off detection that does not have a lamp on within 0.5 seconds.

10.1.2. Turn Detection in GPS Data

The GPS data is analyzed to locate vehicle turns. At each GPS location point, two vectors are formed; A for segment entering the location and B for segment leaving the location. To avoid noise from GPS jitter while stopped, short segments are omitted. The magnitude of the direction change is determined from a dot product, while cross product provides direction.

$$\text{angle} = \cos^{-1}((A \cdot B)/(|A| |B|)) \quad [10-10]$$

$$\text{direction} = \text{sign}((A \times B)_{z \text{ component}}) \quad [10-11]$$

Candidate corners are then identified as changes greater than 30° , which are then filtered to ensure higher velocity turns (ramps) only have the first of consecutive detections accepted.



10.1.3. GIS Turn by Turn Route Creation

Google Maps, as presented in chapter 9, provides a GIS database accessible through APIs that can also be used to find the location of turns and merges within a route retroactively. Each trip segment is represented by a start and stop location (longitude, latitude) and up to 8 (Google API limitation) intermediate locations that are equally spaced in time, resulting in a set of routing directions. The route query is repeated for 1 fewer intermediate locations providing a second measure as different intermediate points may prevent the vehicle from being placed on the wrong road, for instance a small GPS measurement error near a bridge.

The XML file returned by Google Maps Directions API includes turn by turn directions for the trip along with trip summary information. The trip distance estimate for the two sets of directions is compared to the actual as-driven distance, and the result closest to actual driven distance is chosen. The XML turn and merge events set (longitude, latitude, direction (left/right) and type (turn/merge)) from the directions is a series that has a known order of occurrence and a known location, while the GPS trip data is a time series of longitude and latitude samples. The turn events need to be time aligned to the actual as-driven trip through a minimum distance error measure between the as-driven location samples and the turn event locations. A tree search is used to find the ordered as-driven locations that provide the minimum total distance error for the turn events.

$$\text{Error Distance} = \sum_{k=1}^n |\text{Distance errors}| \quad [10-12]$$

10.1.4. Correlation of GPS and GIS Turns with Dashboard Video

The now time aligned GPS data and GIS identified turns must be sampling rate aligned with lamp events. This is an example of a general sampling rate transformation problem where data is not only at highly differing sampling rates, but in the case of GPS data, the sampling rate is variable and the key features of the higher sampling rate must be maintained (signal events).

The video events are down sampled to align with GPS/GIS information, where each new sample reflects the state of the turn signal lamps in preceding interval (left, right or off). GPS and GIS turns data and lamp status information is then



compared, resulting in three event sets {Signal events associated with turns, Signal events not associated with turns, Turns with no associated signals}.

10.2. Experimental Results

Data was collected for a total of ten trips driven by two healthy drivers, each driving their personal vehicle. The drivers were both male, ages 48 (driver 1) and 51 (driver 2). The two vehicles had different dashboard configurations, with one having adjacent turn signal lamps (driver 2), allowing one camera to capture both signals, while the other (driver 1) required separate cameras for each signal.

Video capture of the dashboard proved challenging as car designs include shade structures to prevent sunlight from getting onto the dash, in addition to the steering wheel blocking many camera positions. Only 2 locations were found to give an acceptable image of the turn signals. The first, mounting near the driver's head, was considered unsafe as it could interfere with driver head movements and vision. Hence, the second, hanging the cameras from the top of the dashboard between the steering wheel and dashboard, was chosen. The cameras used were extremely small in size (66mm x 29mm x 15mm) and minimally obscured the dashboard, while the location caused them to be mounted upside down, requiring analysis to account for the image inversion.

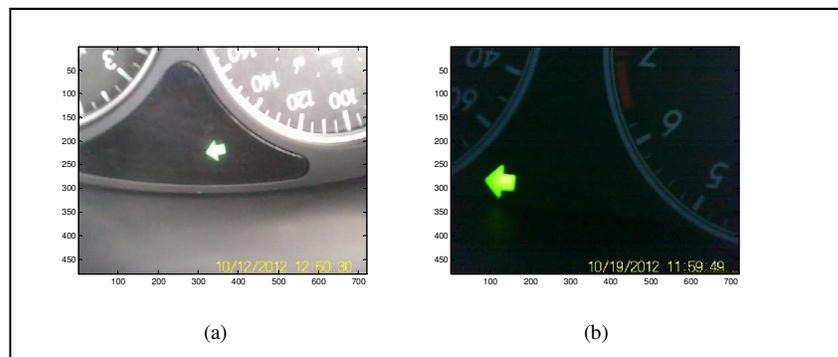


Figure 10-1: Example right turn lamp image: (a) vehicle 1; (b) vehicle 2.

An example dashboard image is shown in Figure 10-1 for each of the vehicles. The video included a time stamp, aligning the video and GPS data, which was excluded from all image analysis. The turn signal events captured on video include 225 separate uses of the signals. These were almost evenly split between left and



right (118 right, 107 left) and ranged from a single flash of the lamp through to events of over 100 flashes. The performance of the detection algorithm is shown in Table 10-1 where 100% of the signal events were detected, with no missed or false positive detections. The algorithm proved to be robust to the effects of vibration or incidental bumps of the camera. The addition of rotation to the algorithm was key to achieving the performance.

Table 10-1: Summary of results for signal arrow detection showing all turn signal events were detected with no false positive or false negative errors.

Signal	Detected	False Positive	False Negative
Right signal (n=118)	100%	0%	0%
Left signal (n=107)	100%	0%	0%
Combined (n=225)	100%	0%	0%

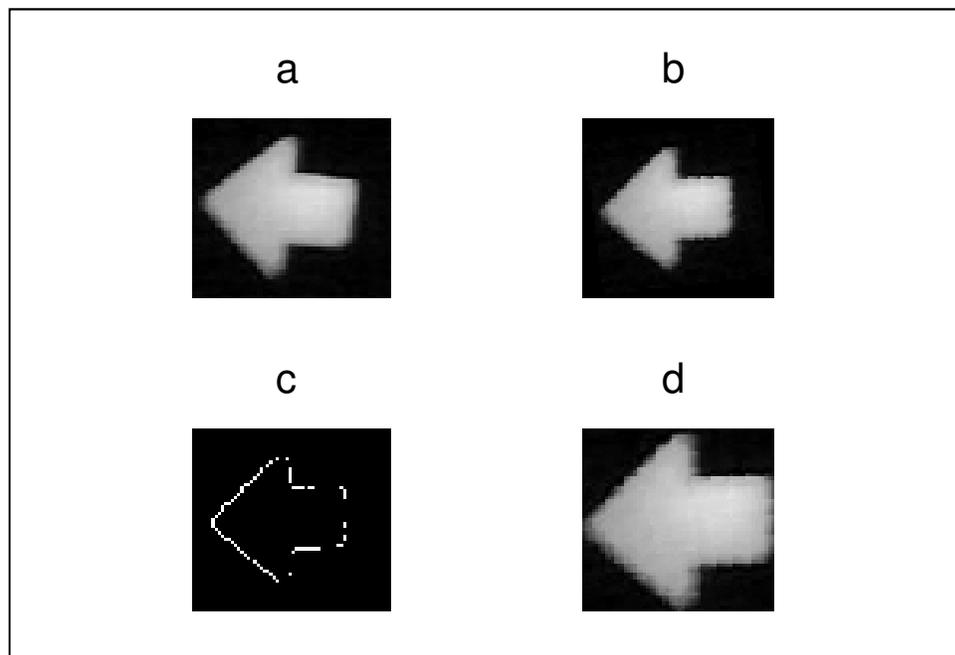


Figure 10-2: Example of arrow detection algorithm image processing steps.

- a – Difference image for raw window containing potential signal arrow.
- b – Candidate image rotated to position with optimal correlation.
- c – Edge detection results for rotated candidate arrow.
- d – Cropped candidate arrow used for correlation.

Examples of the algorithm steps are shown in Figure 10-2. The candidate arrows are incrementally rotated before correlation to determine the optimal result, with



grey-scale isolating the algorithm from variations in lamp colour. To be able to easily identify both the on and off transitions, a delta image provided the best reference. The delta image excluded most of the background other than vibration effects and a 5 frame delta was found to ensure that at least 2 deltas were calculated between full on and full off for each flash, allowing time to change state.

The results of the GPS analysis and Google based GIS remapping are summarized in Table 10-2. Of the 215 turns taken, 73% of them were correctly detected by the GPS turn detection algorithm, while only 58% were detected by the Google GIS algorithm. A key difference in the two algorithms is that the Google GIS algorithm performed well locating turn and merge events on the road network, but failed off the road network (parking lots). The GPS algorithm was able to locate non-road events, but missed merges and some turns into/out of parking lots, where parking allowed minimal travel distance pre/post turn.

Table 10-2: Summary performance for the GPS and GIS remapping algorithms in the detection of turn and merge events within the trips.

Turns	GPS only	Google only
Right turns (n=131)		
Detected turns	73%	62%
False Positive errors	3%	5%
False Negative errors	14%	23%
Left turns (n=86)		
Detected turns	72%	55%
False Positive errors	2%	6%
False Negative errors	15%	24%
Total (n=215)		
Detected turns	73%	58%
False Positive errors	3%	4%
False Negative errors	14%	21%

The false positive detection rate on roads is relatively low at 2 - 6% for identified corners that are not actual corners. Most of these detections are associated with long curved segments of a road, where traffic causes the driver to slow. The extra false positive errors for the Google data are mostly associated with two segments,



where there was a drop out in the GPS data, so an alternative path for a portion of the segment was returned by the API.

The correlation of the turn signal usage events identified in the video signals is shown in Table 10-3. On a standalone basis, the GPS algorithm performs better than the Google GIS algorithm. This is expected because the Google algorithm does not identify non-road events. When the signal events for the two algorithms are combined, the performance improves.

Table 10-3: Resulting association of signal lamp use with identified turns.

Signalled Turns	GPS only	Google only	Combined
Right turns	81	74	89
Left turns	55	46	62
Total	136	120	151

Table 10-4: Turn signal usage rates for the two drivers.

Signalled Turns	Driver 1	Driver 2
Right turns	67%	70%
Left turns	68%	77%
Total	67%	73%

Table 10-4 shows the performance measures for the two drivers, where Driver 2 demonstrates a higher tendency to signal turns than Driver 1. This provides a different measure of driving performance that can be combined with navigational performance. It is not expected that a driver will remember to use their signal at every turn, but it is expected that a driver will have relatively consistent performance, and any long term change in this performance could indicate cognitive change.

10.3. Discussion

This work demonstrates that an automatic over-learned cognitive function associated with driving can also be measured, adding to the more complex executive cognitive functions reported previously. A fusion algorithm was created for sensor signals captured from a GPS location recording system in conjunction

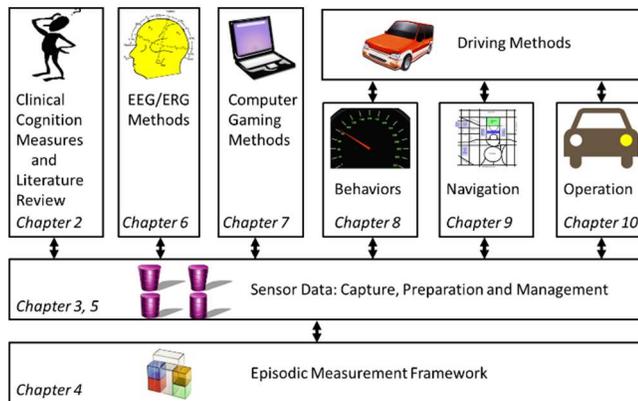


with GIS mapping information. The work specifically shows methods to remap the driving behavior *a posteriori* so that turn and merge requirements for the trip driven can be determined. This information augments the turns identified from analysis of the GPS path. The GIS data provides better performance for areas hard to detect in the path such as merges while GPS analysis allows turns off the road network to be detected such as parking lots. Turn signal use is detected through analysis of dashboard video. The algorithm provides a measure of the driver's performance for use of turn signals for actual turns performed. The result is an algorithm that can analyze a given driver's performance over time as an indirect indication of change in driver cognitive ability.





Chapter 11: Summary and Future Work



This work has presented the episodic cognition assessment framework that provides potential for ongoing assessment of patient well-being through IADL assessment. The work presents methods for cognition assessment for 3 sub-systems within the framework:

- Methods for classification of EEG / ERP features has been shown to allow for MCI subjects to be distinguished from healthy controls.
- Methods for the measurement of performance within two computer games has been shown to allow for the longitudinal measurement of ability allowing for change in ability to be detected.
- Methods for the measurement of driving behaviours have been proposed that allow drivers of a shared vehicle to be distinguished and also for change in navigational or vehicle operations to be detected over longitudinal studies.

11.1. Future Work

There are a number of areas for future exploration:

11.1.1. Episodic Framework

- Expand the Episodic Framework to include additional monitoring subsystems.



- Development of real time implementation of Episodic framework for array of sensor sources.
- Capture of longer term study data using multiple cognition measurement techniques for a set of subjects.

11.1.2. EEG/ERP

- A study of ERP measures for a larger group of participants and also include longitudinal study of the participants to explore stability of EEG/ERP as a measure for both stable and subjects undergoing cognitive change.
- Exploration of the potential for consumer EEG devices to be used that allow for simpler and hence more frequent EEG/ERP measurements.

11.1.3. Games

- Extension of gaming studies to include dementia patients with the potential need for alternative game designs for these patients.
- Explorations of game designs that allow for the integration of EEG / ERP measurements.

11.1.4. Driving

- Development of real time deployment implementations for real time in car analysis.
- Exploration of the effects that health changes such as cognitive, physical or general health have on driving behaviours.
 - Exploration of the long term effects of cognitive change on driving behaviours and the required adaption of a signature over time.
 - Use of driving signature change as indication of driver change (cognitive or physical health).



References

1. B. Wallace, F. Knoefel, R. Goubran, "Cognition Assessment: A Framework for Episodic Measurement ", *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 385-390, 2016.
2. F. Knoefel, R. Lopez Zunini, V. Taler, M. Breau, L. Sweet, R. Wallace, R. Goubran, "Mild Cognitive Impairment: Cognitive testing and EEG changes", *35th AGM of CGS*, 2015.
3. F. Knoefel, R. Lopez Zunini, V. Taler, M. Breau, L. Sweet, C. Lord, R. Wallace, R. Goubran, "Comparing MCI Patients to Healthy Controls using Three ERP Paradigms", *EPIA Pre-Conference Meeting of the Alzheimer's Association International Conference*, Washington DC, USA, July 18, 2015.
4. F. Knoefel, R. Lopez Zunini, V. Taler, M. Breau, L. Sweet, C. Lord, R. Wallace, R. Goubran, "Can Event Related Potentials Using a Language Paradigm Help Diagnose Mild Cognitive Impairment?", *AGM of the Regional Geriatric Program of Eastern Ontario*, Oct 2015.
5. C. Morrison, F. Knoefel, R. Lopez Zunini, V. Taler, M. Breau, L. Sweet, C. Lord, R. Wallace, R. Goubran, S. Allard, "Mild Cognitive Impairment: Cognitive Testings and EEG Changes with Go-NoGo", *8th Canadian Conference on Dementia*, Oct 2015.
6. A. MacCosham, R. Wallace, F. Knoefel, R. Lopez Zunini, V. Taler, M. Breau, L. Sweet, C. Lord, R. Goubran, "Event Related Potentials Elicited During Cognitive Tasks: Biomarkers For Mild Cognitive Impairment", *8th Canadian Conference on Dementia*, Oct 2015.
7. B. Wallace, F. Knoefel, R. López Zunini, M. Breau, L. Sweet, R. Goubran, V. Taler, " Evoked Response Potentials: Which Features Help Distinguish MCI from Healthy Controls", *Alzheimer's Association International Conference AAIC*, 2016.
8. B. Wallace, R. Goubran, F. Knoefel, M. Petriu, A. McAvoy, "Design of games for measurement of cognitive impairment", *2014 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 117-120, 2014.
9. V. Joshi, B. Wallace, A. Shaddy, F. Knoefel, C. Lord, R. Goubran, "Metrics to Monitor Performance of Patients with Mild Cognitive Impairment using Computer Based Games", *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 521-524, 2016.
10. B. Wallace, A. Puli, R. Goubran, F. Knoefel, S. Marshall, M. Porter, A. Smith, "Measurement of Distinguishing Features of Stable Cognitive and Physical Health Older Drivers", *IEEE Transactions on Instrumentation and Measurement* Vol. 65, Iss. 9, pp. 1990-2001, 2016.



11. B. Wallace, R. Goubran, F. Knoefel, S. Marshall, M. Porter, M. Harlow, A. Puli, "Automation of the Validation, Anonymization, and Augmentation of Big Data from a Multi-year Driving Study", *2015 IEEE International Congress on Big Data (BigData Congress)*, pp. 608-614, 2015.
12. B. Wallace, R. Goubran, F. Knoefel, S. Marshall, M. Porter, "Measuring variation in driving habits between drivers", *2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1-6, 2014.
13. B. Wallace, M. Rockwood, R. Goubran, F. Knoefel, S. Marshall, M. Porter, "Measurement of vehicle acceleration in studies of older drivers from GPS position and OBDII velocity sensors", *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 391-396, 2015.
14. M. Rockwood, B. Wallace, R. Goubran, F. Knoefel, S. Marshall, "Adaptive drift calibration of accelerometers with direct velocity measurements", *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 240-244, 2015.
15. B. Wallace, A. Puli, R. Goubran, F. Knoefel, S. Marshall, M. Porter, A. Smith, "Big data analytics to identify deceleration characteristics of an older driver", *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 89-94, 2015.
16. B. Wallace, R. Goubran, F. Knoefel, "Measurement of driving routes and correlation to optimal navigation paths", *2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1465-1470, 2013.
17. B. Wallace, R. Goubran, F. Knoefel, "Cognitive change measurement through driving navigation ability sensing and analysis", *2013 IEEE International Symposium on Medical Measurements and Applications Proceedings (MeMeA)*, pp. 164-169, 2013.
18. R. Wallace, R. Goubran, F. Knoefel, "Scoring Cognitive Change Through Sensing and Analysis of Changing Driving Ability", *CMBEC36*, 2013.
19. R. Wallace, R. Goubran, F. Knoefel, "Measurement of Signal Use and Vehicle Turns as Indication of Driver Cognition", *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3747-3750, 2014.
20. B. Wallace, R. Goubran, F. Knoefel, S. Marshall, M. Porter, A. Smith, "Driver Unique Acceleration Behaviours and Stability over 2 Years", *2016 IEEE International Congress on Big Data (BigData Congress)* pp. 230-235, 2016.
21. A. Alakel, B. Wallace, M. Porter, H. Tuokko, A. Myers, R. Goubran, F. Knoefel, S. Marshall, "Do Older Drivers Use Cruise Control?", *Human-Computer Interaction (HCI) International 2016* (invited paper).
22. Petersen, et al., "Mild cognitive impairment: ten years later", *Archives of Neurology*, Vol. 66, pp. 1447-1455, 2009.



23. "Dementia", World Health Organization Fact Sheet, <http://www.who.int/mediacentre/factsheets/fs362/en/>, April 2016.
24. "Canadian Study of Health and Aging Working Group. Canadian Study of Health and Aging: Study methods and prevalence of dementia", *Canadian Medical Association Journal*, Vol. 150, pp. 899-913, 1994.
25. R. Petersen, J. Stevens, M. Ganguli, E. Tangalos, J. Cummings, S. DeKosky, "Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology", *Neurology*, Vol. 56, pp. 1133-1142, 2001.
26. R. Petersen, "Mild Cognitive Impairment", *NEJM*; Vol. 364, pp. 2227-2234, 2011.
27. P. Doraiswamy, L. Kaiser, F. Bieber, R. Garman, "The Alzheimer's Disease Assessment Scale: Evaluation of Psychometric Properties and Patterns of Cognitive Decline in Multicenter Clinical Trials of Mild to Moderate Alzheimer's Disease", *Alzheimer Disease & Associated Disorders*, Vol. 15, no. 4, pp. 174-184, 2001.
28. M. Folstein, L. Robins, Jo. Helzer. "The mini-mental state examination", *Archives of general psychiatry*, Vol. 40, Iss. 7, pp. 812-812, 1983.
29. C. Randolph, M. Tierney, E. Mohr, T. Chase, " The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Preliminary Clinical Validity", *Journal of Clinical and Experimental Neuropsychology*, Vol. 20, Iss. 3, pp. 310-319, 1998.
30. T. Tombaugh, "Trail Making Test A and B: normative data stratified by age and education", *Archives of clinical neuropsychology*, Vol. 19.2, pp. 203-214, 2004.
31. Z. Nasreddine, et al. "The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment", *Journal of the American Geriatrics Society*, Vol. 53.4, pp. 695-699, 2005.
32. K. Ritchie, S. Artero, J. Touchon, "Classification criteria for mild cognitive impairment: A population-based validation study", *Neurology* Vol. 56, no. January, pp. 37-42, 2001.
33. M. Bédard, et al. "The SIMARD Screening Tool to Identify Unfit Drivers Are We There Now?" *Journal of primary care & community health*, Vol. 2.2, pp. 133-135, 2011.
34. S. Luck, "An introduction to the event-related potential technique", MIT Press, 2014.
35. T. Picton, O. Lins, M. Scherg. "The recording and analysis of event-related potentials", *Handbook of neuropsychology*, Vol 10, pp. 3-71, 1995.



36. D. Friedman, R. Johnson, "Event-related potential (ERP) studies of memory encoding and retrieval: a selective review. *Microscopy research and technique*, Vol. 51, Iss 1, pp. 6-28, 2000.
37. G. Killane, R. Browett, R. Reilly, "Measurement of attention during movement: Acquisition of ambulatory EEG and cognitive performance from healthy young adults". *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pp. 6397-6400, 2013.
38. P. Missonnier, et al., "Working memory load-related electroencephalo-graphic parameters can differentiate progressive from stable mild cognitive impairment". *Neuroscience*, Vol. 150, Iss. 2, pp. 346-356, 2007.
39. L. McEvoy, M. Smith, A. Gevins, "Dynamic cortical networks of verbal and spatial working memory: effects of memory load and task practice". *Cerebral Cortex*, Vol. 8, Iss. 7, pp. 563-574, 1998.
40. A. Ivannikov, T. Karkkainen, T. Ristaniemi, H. Lyytinen, "Extraction of ERP from EEG data". *2007 9th International Symposium on Signal Processing and Its Applications*, pp 1-4, 2007
41. C. McGillen, J. Aunon, K. Yu, "Signals and noise in evoked brain potentials". *IEEE Transactions on Biomedical Engineering*, Vol. 12, pp. 1012-1016, 1985.
42. T. Rondík, J. Ciniburk, J. "Comparison of various approaches for P3 component detection using basic methods for signal processing". *4th International Conference on Biomedical Engineering and Informatics (BMEI)*, 2011, Vol. 2, pp. 698-702, 2011.
43. L. Fernandes, D. Ferreira, P. Almeida, N. Dias, "Aging and attentional set shifting on WCST: An event-related EEG study", *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp.1088-1091, 2015.
44. H. Devillez, et al., "The P300 potential for fixations onto target object when exploring natural scenes during a visual task after denoising overlapped EFRP", *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp.1024-1027, 2015.
45. S. Katz, "Assessing self-maintenance: Activities of daily living, mobility, and instrumental activities of daily living", *Journal of the American Geriatrics Society*, Vol. 31(12), pp. 721-727, 1983.
46. R. Marottoli, C. Mendes de Leon, T. Glass, C. Williams, L. Cooney, Jr., L. Berkman, "Consequences of Driving Cessation: Decreased Out-of-Home Activity Levels", *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, Vol. 55, no. 6, pp. S334-S340, 2000.
47. S. Bassuk, T. Glass, L. Berkman, "Social disengagement and incident cognitive decline in community-dwelling elderly persons", *Annals of Internal Medicine*, Vol. 131, no. 3, pp. 165-173, 1999.



48. D. Hulstsch, C. Hertzog, B. Small, R. Dixon, "Use it or lose it: Engaged lifestyle as a buffer of cognitive decline in aging?" *Psychology and Aging*, Vol. 14, no. 2 pp. 245-263, 1999.
49. T. Hayes, S. Hagler, D. Austin, J. Kaye, M. Pavel, "Unobtrusive assessment of walking speed in the home using inexpensive PIR sensors", *EMBC 2009. Annual Int. Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 7248-7251, 2009.
50. A. Arcelus, I. Veledar, R. Goubran, F. Knoefel , "Measurements of Sit to-Stand Timing and Symmetry from Bed Pressure Sensors", *IEEE Trans. on Inst. and Measurement*, Vol. 60, no. 5, pp. 1732-40, 2011.
51. M. Taylor, D. McEwen, R. Goubran, H. Finestone, F. Knoefel , "Assessing Standing Stability of Older Adults using Pressure Sensitive Arrays", *Proc. of IEEE International Workshop on Medical Measurements and Applications*, Budapest, Hungary, pp.1-5, 2012.
52. R. Sobolewski, R. Reilly, S. Finnigan, P. Dockree, K. O'Sullivan, I. Robertson, "Monitoring of cognitive processes in older persons", *Neural Engineering, 2009. NER '09. 4th International IEEE/EMBS Conference on*, pp. 132-135, 2009.
53. S. Hagler, H. Jimison, M. Pavel, "Unobtrusive in-home assessment by means of everyday computer mouse usage", *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3816-3819, 2011.
54. H. Jimison, M. Pavel, K. Wild, P. Bissell, J. McKanna, D. Blaker, D. Williams, "A Neural Informatics Approach to Cognitive Assessment and Monitoring", *3rd International IEEE/EMBS Conference on Neural Engineering, 2007 (CNE '07)*, pp. 696-699, 2007.
55. H. Jimison, M. Pavel, T. Le, "Home-based cognitive monitoring using embedded measures of verbal fluency in a computer word game", *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008 (EMBC)*, pp. 3312-3315, 2008.
56. F. Unverzagt, L. Kasten, K. Johnson et al. "Effect of memory impairment on training outcomes in ACTIVE", *J IntNeuropsychol Soc*; Vol. 13, pp.953–960, 2007.
57. S. Willis, S. Tennstedt, M. Marsiske, et al. "Long-term effects of cognitive training on everyday functional outcomes in older adults", *JAMA*; Vol. 296, pp. 2805-2814, 2006.
58. M. Martin, L. Clare, A. Altgassen, M. Cameron, F. Zehnder, "Cognition-based interventions for healthy older people and people with mild cognitive impairment", *Cochrane Database Syst Rev 1.1*, 2011.
59. A. Berry, T. Zanto, W. Clapp, J. Hardy, P. Delahunt, H. Mahncke, A. Gazzaley. "The influence of perceptual training on working memory in older adults", *PLoS one*, Vol. 5(7), pp. 1-8, 2010.



60. www.brainhq.com
61. H. Nikamalfard, Z. Huiru, W. Haiying, P. Jeffers, M. Mulvenna, P. McCullagh, S. Martin, J. Wallace, J. Augusto, W. Carswell, B. Taylor, K. McSorley, "Knowledge discovery from activity monitoring to support independent living of people with early dementia", *2012 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 910-913, 2012.
62. E. Cha, J. Wood, J. Finkelstein, "Using gaming platforms for telemedicine applications: A cross-platform comparison", *2012 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 918-921, 2012.
63. D. Austin, et al. "Measuring motor speed through typing: a surrogate for the finger tapping test", *Behavior research methods*, Vol. 43.4, pp. 903-909, 2011.
64. H. Jimison, M. Pavel, J. Pavel. "Adaptive interfaces for home health", *Proceedings of the International Workshop on Ubiquitous Computing for Pervasive Healthcare*, 2003.
65. H. Jimison, et al. "Unobtrusive monitoring of computer interactions to detect cognitive status in elders", *IEEE Transactions on Information Technology in Biomedicine*, Vol. 8.3, pp. 248-252, 2004.
66. H. Jimison, et al. "Monitoring Computer Interactions to Detect Early Cognitive Impairment in Elders", *1st Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare, 2006. D2H2*, IEEE, pp. 75-78, 2006.
67. T Seeman, D. Miller-Martinez, S. Stein Merkin, M. Lachman, P. Tun, A. Karlamangla, "Histories of social engagement and adult cognition: midlife in the U.S. study", *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, Vol. 66 suppl 1, pp. i141-i152, 2011.
68. M. Zunzunegui, B. Alvarado, T. Del Ser, A. Otero, "A Social networks, social integration, and social engagement determine cognitive decline in community-dwelling Spanish older adults", *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, Vol. 58, no. 2, pp. S93-S100, 2003.
69. J. Edwards, M. Lunsman, M. Perkins, G. Rebok, D. Roth, "Driving cessation and health trajectories in older adults", *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, Vol. 64, no. 12, pp. 1290-1295, 2009.
70. D. Eby, et al. "Driving behaviors in early stage dementia: A study using in-vehicle technology", *Accident Analysis & Prevention*, Vol. 49, pp. 330-337, 2012.
71. D. Eby and L. Molnar. "Older adult safety and mobility issues and research needs", *Public Works Management & Policy*, Vol. 13, Iss. 4, pp. 288-300, 2009.



72. L. Molnar, and D. Eby. "The relationship between self-regulation and driving-related abilities in older drivers: an exploratory study", *Traffic injury prevention* Vol. 9, Iss. 4, pp. 314-319, 2008.
73. M. Rapoport, et al. "The relationship between cognitive performance, perceptions of driving comfort and abilities, and self-reported driving restrictions among healthy older drivers", *Accident Analysis & Prevention*, Vol. 61, pp. 288–295, 2013.
74. K. Anstey, J. Wood, S. Lord, J. Walker, "Cognitive, sensory and physical factors enabling driving safety in older adults", *Clinical Psychology Review*, Vol 25, no. 1, pp. 45-65, 2005.
75. R. Marottoli, C. Mendes de Leon, T. Glass, C. Williams, L. Cooney, Jr., L. Berkman, "Consequences of Driving Cessation: Decreased Out-of-Home Activity Levels", *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, Vol. 55, no. 6, pp. S334-S340, 2000.
76. R. De Raedt, I. Ponjaert-Kristoffersen, "Can strategic and tactical compensation reduce crash risk in older drivers?", *Age Ageing*, Vol. 29, no. 6 pp. 517-521, 2000.
77. R. Dubinsky, A. Stein, K. Lyons, "Risk of driving and Alzheimer's disease (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology", *Neurology*, Vol. 54, June, pp. 2205-2211, 2000.
78. L. Hunt, A. Brown, I. Gilman, "Drivers With Dementia and Outcomes of Becoming Lost While Driving", *Am J Occup Ther*, Vol. 64, March, pp. 225-232, 2010.
79. N. Silverstein, et al. "Do Memory-Impaired Drivers and Their Family Members Agree on Driving Ability and Behaviors?", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2265, pp. 200-206, 2011.
80. T. Dingus, R. Hanowski, and S. Klauer. "Estimating crash risk", *Ergonomics in Design: The Quarterly of Human Factors Applications*, Vol 19, pp. 8-12, 2011.
81. W. Lindstrom-Forneri, et al. "Driving as an everyday competence: a model of driving competence and behavior", *Clinical Gerontologist*, Vol 33, pp. 283-297, 2010.
82. F. Morganti, G. Riva, "Spatial orientation decline in elderly population", *2011 International Conference on Virtual Rehabilitation (ICVR)*, pp.1-2, 2011.
83. S. Marshall, et al. "Protocol for Candrive II/Ozcandrive, a multicentre prospective older driver cohort study", *Accident Analysis & Prevention*, Vol. 61, pp. 245–252, 2013.
84. S. Marshall, et al. "The Canadian Safe Driving Study–phase I pilot: Examining potential logistical barriers to the full cohort study", *Accident Analysis & Prevention*, Vol. 61, pp. 236–244, 2013.



85. S. Hallmark, et al. "Initial analyses from the SHRP 2 naturalistic driving study: addressing driver performance and behavior in traffic safety", *Transportation Research Board, No. SHRP 2 Safety Project S08*, 2013.
86. K. Campbell, "The SHRP 2 naturalistic driving study: Addressing driver performance and behavior in traffic safety", *Transportation Research (TR) News*, Iss. 282, pp 30-35, 2012.
87. M. Bédard, M. Parkkari, B. Weaver, J. Riendeau, M. Dahlquist, "Assessment of driving performance using a simulator protocol: Validity and reproducibility", *American Journal of Occupational Therapy*, Vol. 64, pp. 336-340, 2010.
88. M. Porter, et al. "Older Driver Estimates of Driving Exposure Compared to In-Vehicle Data in the Candrive II Study", *Traffic Injury Prevention*, Vol. 16, Iss. 1, pp. 24-27, 2015.
89. Z. Vlahodimitrakou, et al. "Development and evaluation of a Driving Observation Schedule (DOS) to study everyday driving performance of older drivers", *Accident Analysis & Prevention*, Vol. 61, pp. 253-260, 2013.
90. M. Tundo, E. Lemaire, N. Baddour, "Correcting Smartphone orientation for accelerometer-based analysis", *2013 IEEE International Symposium on Medical Measurements and Applications Proceedings (MeMeA)*, pp.58,62, 2013.
91. T. Beravs, J. Podobnik, M. Munih, "Three-Axial Accelerometer Calibration Using Kalman Filter Covariance Matrix for Online Estimation of Optimal Sensor Orientation", *IEEE Transactions on Instrumentation and Measurement*, Vol.61, no.9, pp.2501-2511, 2012.
92. I. Skog, and P. Handel. "In-car positioning and navigation technologies-A survey", *IEEE Trans. on Intelligent Transportation Systems*, Vol. 10, pp 4-21, 2009.
93. J. Jun, R. Guensler, J. Ogle. "Smoothing methods to minimize impact of Global Positioning System random error on travel distance, speed, and acceleration profile estimates", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1972.1 pp. 141-150, 2006.
94. C. Barrios, Y. Motai, "Improving Estimation of Vehicle's Trajectory Using the Latest Global Positioning System With Kalman Filtering", *IEEE Transactions on Instrumentation and Measurement*, Vol. 60, no. 12, pp. 3747-3755, 2011.
95. P. Silson, "Coarse Alignment of a Ship's Strapdown Inertial Attitude Reference System Using Velocity Loci", *IEEE Transactions on Instrumentation and Measurement*, Vol. 60, no. 6, pp. 1930-1941, 2011.
96. S. Beauchemin, M. Bauer, T. Kowsari, J. Cho, J. (2012). "Portable and scalable vision-based vehicular instrumentation for the analysis of driver intentionality", *IEEE Transactions on Instrumentation and Measurement*, Vol. 61, Iss. 2, pp. 391-401, 2012.



97. D. Katzourakis, E. Velenis, D. Abbink, R. Happee, E. Holweg, "Race-car instrumentation for driving behavior studies", *IEEE Transactions on Instrumentation and Measurement*, Vol. 61, Iss. 2, pp. 462-474, 2012.
98. D. Katzourakis, D. Abbink, E. Velenis, E. Holweg, R. Happee, "Driver's Arms' Time-Variant Neuromuscular Admittance During Real Car Test-Track Driving", *IEEE Transactions on Instrumentation and Measurement*, Vol. 63, Iss. 1, pp. 221-230, 2014
99. C. D'Agostino, A. Saidi, G. Scouarnec, L. Chen, "Learning-Based Driving Events Recognition and Its Application to Digital Roads", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, Iss. 4, pp. 2155-2166, 2015.
100. C. D'Agostino, A. Saidi, G. Scouarnec, L. Chen, "Learning-based driving events classification", *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, 2013.
101. M. Canale, S. Malan, "Analysis and classification of human driving behaviour in an urban environment", *Cognition, Technology & Work*, Vol. 4 Iss. 3, pp. 197-206, 2002.
102. M. Van Ly, S. Martin, M. Trivedi, "Driver classification and driving style recognition using inertial sensors", *2013 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1040-1045, 2013.
103. S. Kraus, M. Althoff, B. Heissing, M. Buss, "Cognition and emotion in autonomous cars", *Proc. 2009 IEEE Intelligent Vehicles Symposium*, pp.635-640, 2009.
104. K. Jo, K. Chu, M. Sunwoo, "Interacting Multiple Model Filter-Based Sensor Fusion of GPS With In-Vehicle Sensors for Real-Time Vehicle Positioning", *IEEE Transactions on Intelligent Transportation Systems*, Vol 13, no 1, pp. 329-343, 2012.
105. E. Mok, G. Retscher, W. Chen, "Initial test on the use of GPS and sensor data of modern smartphones for vehicle tracking in dense high rise environments", *Ubiquitous Positioning, Indoor Navigation, and Location Based Service (UPINLBS) 2012*, pp. 1-7, 2012.
106. E. Tzoreff, B. Bobrovsky, "A Novel Approach for Modeling Land Vehicle Kinematics to Improve GPS Performance Under Urban Environment Conditions", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, no. 1, pp. 344-353, 2012
107. A. Bergeron, N. Baddour, "Design and Development of a Low-Cost Multisensor Inertial Data Acquisition System for Sailing", *IEEE Transactions on Instrumentation and Measurement*, Vol. 63, no. 2, pp. 441-449, 2014.
108. P. Handel, "Discounted Least-Squares Gearshift Detection Using Accelerometer Data", *IEEE Transactions on Instrumentation and Measurement*, Vol. 58, no. 12, pp. 3953-3958, 2009.



109. D. Eby et al. "Fitness to drive in early stage dementia: an instrumented vehicle study", *M-CASTL*, Report No. MCASTL-2009-03, 2009.
110. I. Skog, P. Handel, "Indirect Instantaneous Car-Fuel Consumption Measurements", *IEEE Transactions on Instrumentation and Measurement*, Vol.63, no.12, pp.3190,3198, Dec. 2014.
111. I. Skog, P. Handel, "In-Car Positioning and Navigation Technologies—A Survey", *IEEE Transactions on Intelligent Transportation Systems*, Vol.10, no.1, pp.4,21, March 2009.
112. J. Nilsson, I. Skog, P. Handel, "Aligning the Forces—Eliminating the Misalignments in IMU Arrays", *IEEE Transactions on Instrumentation and Measurement*, , Vol.63, no.10, pp.2498,2500, Oct. 2014.
113. M. Nassar, B. al Bouna, Q. Malluhi, "Secure Outsourcing of Network Flow Data Analysis", *2013 IEEE International Congress on Big Data (BigData Congress)*, pp.431-432, 2013.
114. Y. Jeong, H. Choi, Y. Kwon, K. Kang, J. Lee, H. Kim, H. Park, J. Kim, "Semi-automated lifestyle manager for obesity", *2014 International Conference on Big Data and Smart Computing (BIGCOMP)*, pp.229-230, 2014.
115. A. Chakravorty, T. Wlodarczyk, C. Rong, "A Scalable K-Anonymization Solution for Preserving Privacy in an Aging-in-Place Welfare Intercloud", *2014 IEEE International Conference on Cloud Engineering (IC2E)*, pp.424-431, 2014.
116. X. Zhang, C. Yang; S. Nepal, C. Liu; W. Dou; J. Chen, "A MapReduce Based Approach of Scalable Multidimensional Anonymization for Big Data Privacy Preservation on Cloud", *Third International Conference on Cloud and Green Computing (CGC)*, pp.105-112, 2013.
117. X. Zhang, C. Liu; S. Nepal, C. Yang; W. Dou; J. Chen, "Combining Top-Down and Bottom-Up: Scalable Sub-tree Anonymization over Big Data Using MapReduce on Cloud", *2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pp.501-508, 2013.
118. D. Mir, S. Isaacman, R. Caceres, M. Martonosi, R. Wright, "DP-WHERE: Differentially private modeling of human mobility", *2013 IEEE International Conference on Big Data*, pp.580-588, 2013.
119. N. Manfredi, D. Mir, S. Lu, D. Sanchez, "Differentially private models of tollgate usage: The Milan tollgate data set", *2014 IEEE International Conference on Big Data (Big Data)*, pp.46-48, 2014.
120. A. Artikis, M. Weidlich, A. Gal, V. Kalogeraki, D. Gunopulos, "Self-adaptive event recognition for intelligent transport management", *2013 IEEE International Conference on Big Data*, pp.319-325, 2013.
121. G. Smith, "Seasonal variation in older adults' driving trip distances", *Master's Thesis, University of Manitoba*, <http://hdl.handle.net/1993/23427>, 2014.



122. J. Friedman, "Regularized discriminant analysis". *Journal of the American Statistical Association*, Vol. 84, Iss. 405, pp. 165-175, 1989.
123. J. Suykens, J. Vandewalle, J. "Least squares support vector machine classifiers", *Neural processing letters*, Vol. 9, Iss. 3, pp. 293-300, 1999.
124. J. Keller, M. Gray, J. Givens, J. A. "A fuzzy k-nearest neighbor algorithm", *IEEE transactions on systems, man, and cybernetics*, Vol. 4, pp. 580-585, 1985.
125. T. Dietterich, T. G. "Ensemble methods in machine learning", In *International workshop on multiple classifier systems*, Springer Berlin Heidelberg, pp. 1-15, 2000.
126. S. Liang, et al., "Classification of driver's cognitive responses from EEG analysis". *IEEE International Symposium on Circuits and Systems (ISCAS 2005)*, pp. 156-159, 2015.
127. M. Aberg, J. Wessberg, "An Evolutionary Approach to the Identification of Informative Voxel Clusters for Brain State Discrimination", *IEEE Journal of Selected Topics in Signal Processing*, Vol.2, no.6, pp.919-928, 2008.
128. C. Cecere, C. Corrado, R. Polikar, "Diagnostic utility of EEG based biomarkers for Alzheimer's disease". *2014 40th Annual Northeast Bioengineering Conference (NEBEC)*, pp. 1-2, 2014.
129. A. Bhardwaj, et al., "Classification of human emotions from EEG signals using SVM and LDA Classifiers", *2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)*, pp.180-185, 2015.
130. Y. Ou, et al., "Instrumental activities of daily living (IADL) evaluation system based on EEG signal feature analysis", *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp.1-4, 2013.
131. M. Mahanta, A. Aghaei, K. Plataniotis, "A Bayes optimal matrix-variate LDA for extraction of spatio-spectral features from EEG signals", *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp.3955-3958, 2012.
132. A. Moore et al. "Fourth Canadian Consensus Conference on the Diagnosis and Treatment of Dementia Recommendations for family physicians", *Canadian Family Physician*, Vol. 60, Iss. 5, pp. 433-438, 2014.
133. R. López Zunini, F. Knoefel, C. Lord, M. Breau, L. Sweet, R. Goubran, V. Taler, "P300 amplitude alterations during inhibitory control in persons with Mild Cognitive Impairment", *Brain Research*, Vol. 1646, pp. 241-248, 2016.
134. www.instamapper.com
135. A. Smith, et al. "Stability of physical assessment of older drivers over 1 year", *Accident Analysis & Prevention*, Vol. 16, Iss. 1, pp. 261-266, 2013.



136. S. Chapra, "Applied Numerical Methods with MATLAB for Engineers and Scientists, 6th edition", *McGraw-Hill Science/Engineering/Math*, 2006.
137. "Appendix E, Almanac for Computers", *Nautical Almanac Office, U.S. Naval Observatory*, 1978.
138. <https://developers.google.com/maps/documentation/distancematrix/>
139. M. Rockwood, V. Joshi, K. Sullivan, R. Goubran, "Using a Real-Time Operating System for Multitasking in Remote Patient Monitoring", *2014 IEEE International Symposium on Medical Measurements and Applications Proceedings (MeMeA)*, pp. 475-479, 2014.
140. T. Grant, et al. "Measuring sit-to-stand timing variability over time using under mattress pressure sensor technology", *Proc. of 2014 IEEE International Symposium on Medical Measurements and Applications Proceedings (MeMeA)*, 2014.
141. E. Aguirre, et al. "Implementing context aware scenarios to enable smart health in complex urban environments", *Proc. of 2014 IEEE International Symposium on Medical Measurements and Applications Proceedings (MeMeA)*, 2014.
142. J. Hoey, et al. "Automated handwashing assistance for persons with dementia using video and a partially observable markov decision process", *Computer Vision and Image Understanding*, Vol. 114. Iss. 5, pp. 503-519, 2010.
143. K. Vidyasagar, M. Moghavvemi and T. Prabhat, "Performance evaluation of contemporary classifiers for automatic detection of epileptic EEG", *2015 International Conference on Industrial Instrumentation and Control (ICIC)*, pp. 372-377, 2015.
144. N. Kafshgari, R. Davoodi Kahaki, M. Moradi and A. Younesi, "An ERP study on visual attention to facial stimuli; N170 component", *2014 22nd Iranian Conference on Electrical Engineering (ICEE)*, pp. 1976-1979, 2014.
145. C. Chuang, P. Lai, L. Ko, B. Kuo and C. Lin, "Driver's cognitive state classification toward brain computer interface via using a generalized and supervised technology", *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-7, 2010.
146. S. Shapiro, M. Wilk. "An analysis of variance test for normality (complete samples)", *Biometrika* Vol. 52, no. 3/4 pp. 591-611, 1965.



Appendix A: Additional Driver Behaviour Attributes

Figure A-1 provides a further analysis of the driving patterns for each of the four drivers showing a histogram of the time of day using 1-hour bins for each trip taken by the driver and the histogram for each of the 100 trips is then averaged. The figure clearly shows that all drivers drive during the day and early evening avoiding the overnight, very early morning and late evening. The wide spread in the standard deviations show large trip to trip variation that is expected as many of the trips are short in duration so the histogram for a particular trip tends to occur within only 1 or 2 bins.

Figure A-2 shows histograms of the GPS reported velocity for each of the drivers. The data for drivers 1 and 3 further demonstrates the avoidance of high-speed (80+km/hr) roads as the velocity profiles have minimal actual driving at 80km/hr and above. The use of highways by driver 4 is visible in the increased travel at 80 and 90km/hr and aligns with the higher use of 100km/hr highways in Table 4-7 but also indicates a tendency for travel below the posted speed limit. Driver 3 data shows the much higher use of highways and driving at 100km/hr. The figure also shows that there is reduced variation in the trip-to-trip velocity profiles through the narrowed standard deviation spread.

Figure A-3 further explores the velocity choices of drivers by comparing their choice of velocity with the posted speed limit on the road where known. Drivers 1 and 3 did not make use of 80km/hr highways but they do have a much higher occurrence of travel above the posted speed limit as shown by the non-zero results for the 110% and >120% of the posted limit while drivers 2 and 4 do not show this tendency indicating different driving habits for the drivers. The detailed data as summarized in Table 8-2 allows the speed limit of the road to be used in conjunction with the drivers chosen speed. Again, the standard deviation spreads show the limited trip-to-trip variation in this measure.

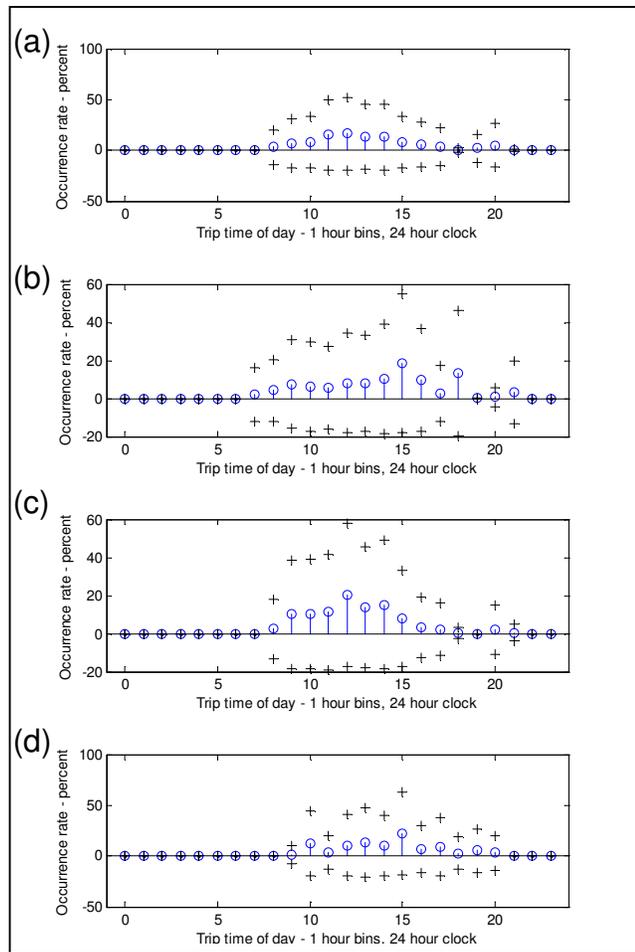


Figure A-1: Histogram for time of day travel for each of the drivers showing the mean value for each of the 1 hour wide bins for the 100 trips. + signs show one standard deviation (a) - driver 1, (b) - driver 2, (c) - driver 3, (d) - driver 4.

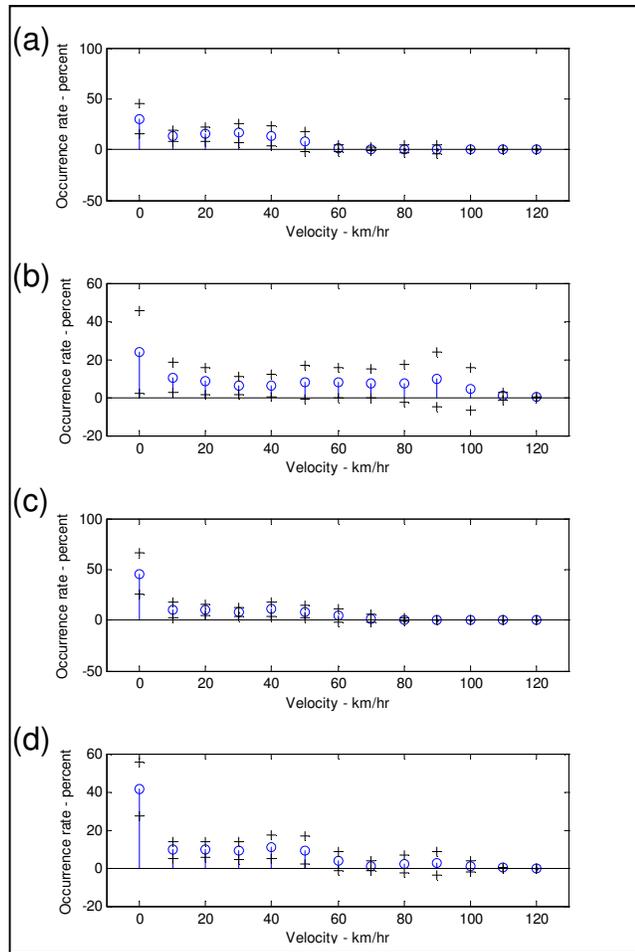


Figure A-2: Histogram of velocity habits for the GPS report velocity showing the mean value for each of the 10km/hr hour wide bins for the 100 trips. + shows one standard deviation. (a) - driver 1, (b) - driver 2, (c) - driver 3, (d) - driver 4.

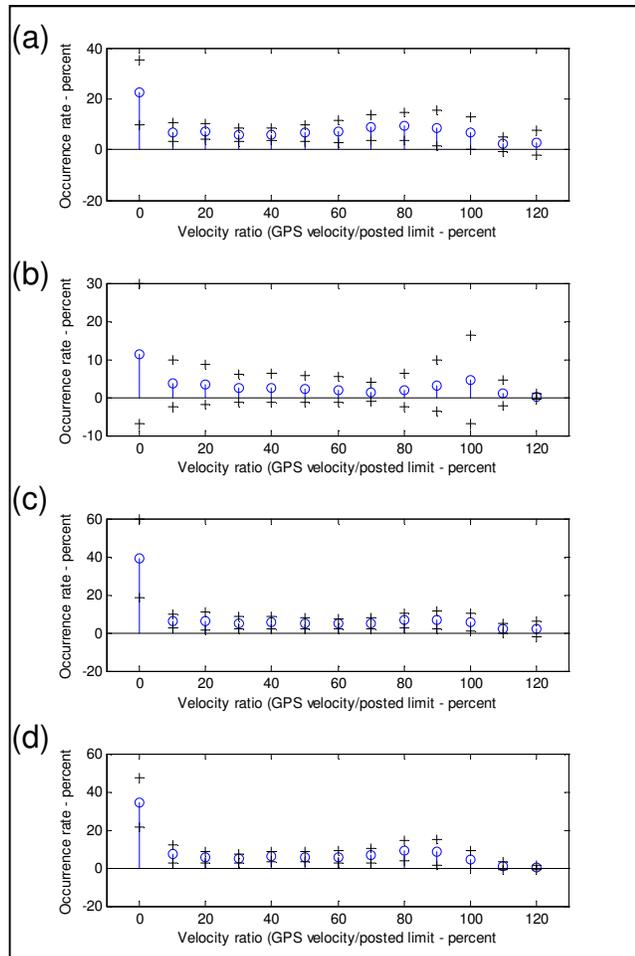


Figure A-3: Histograms of the ratio of velocity to posted speed limit for each of the drivers. Data excludes all samples where the speed limit is unknown. (a) - driver 1, (b) - driver 2, (c) - driver 3, (d) - driver 4.



Appendix B: Acceleration Measurement

B.1. Method

Measuring driver acceleration characteristics is important as it could lead to insights into driver performance and potentially help indicate who is driving through different characteristic behaviours. The Candrive sensors do not directly capture acceleration information, as no accelerometer was included in the sensor set due to concerns about added complexity, cost, and storage requirements. The sensor system does record the vehicle velocity from the engine computer (Dashboard velocity) inertial sensor, and GPS system velocity.

Acceleration Measurement Techniques

The derivative of the GPS and OBDII velocity signals provides two acceleration measures, and Table B-1 summarizes difference formulas that can be used to estimate the derivative of discrete data [136]. This section explores the optimal choice of difference formula and sensor source so that the derived acceleration best corresponds to direct acceleration measurement.

The authors' vehicle was outfitted with the Candrive sensor and with 3 Blackberry Z10 smart phones (each containing 3 axis accelerometers). An application [139] was deployed on each of the smart phones that allowed local recording of the accelerometer readings. The smart phones were set to a 40Hz sampling rate to provide over sampling from the 1Hz GPS and OBDII velocity data. The accelerometer and velocity sensors were recorded during 4 acceleration / deceleration profiles with the smart phone in the following positions:

- All 3 phones oriented with top in direction of travel.
- 2 phones with bottom / 1 with top in direction of travel.
- All 3 phones oriented with right side in direction of travel.
- 2 phones with left / 1 with right side in direction of travel.



Table B-1: Comparison of finite difference formulas for derivative estimate.

	Newton Fwd/Bwd 2-point	Fwd/Bwd 3-point	Central 2-point	Central 4-point
Filter Type	high pass linear phase	high pass non- linear phase	band pass linear phase	high pass linear phase
Interpolation	Linear	2 nd Order	2 nd Order	4 th Order
Symmetry	Asymmetric	Asymmetric	Symmetric	Symmetric
Noise free Error Order	$O(h)$	$O(h^2)$	$O(h^2)$	$O(h^4)$

The 16 acceleration/deceleration profiles and a mean accelerometer signal was produced by time aligning the smart phone signals and averaging the readings. The estimated acceleration from the velocity signals was estimated using difference formulas with the Newton 2-point method shown in equation B-1 and estimates the derivative using the next point (Forward) or previous point (Backward) with the estimate based on the slope of the interpolating line.

The Newton 2-point technique is extended to 3 points evaluating the derivative of the interpolating quadratic polynomial with three cases: Forward 3-point using the next 2 points, (equation B-2) Backward 3-point using the previous 2 points, (equation B-3) and Central 2-point using 1 point on either side (equation B-4). Central 4-point formula (equation B-5) is a fourth method derived through the application of Richardson extrapolation [136] to the Central 2-point formula and is equivalent to measuring the derivative of an interpolating fourth order polynomial.

$$f'[n] = \frac{f[n+1]-f[n]}{h} \quad [B-1]$$

$$f'[n] = \frac{-f[n+2]+4f[n+1]-3f[n]}{2h} \quad [B-2]$$

$$f'[n] = \frac{-3f[n-2]+4f[n-1]-f[n]}{2h} \quad [B-3]$$

$$f'[n] = \frac{f[n+1]-f[n-1]}{2h} \quad [B-4]$$

$$f'[n] = \frac{-f[n+2]+8f[n+1]-8f[n-1]+f[n-2]}{12h} \quad [B-5]$$



The two central difference formulas are symmetric through the use of a balanced set of points on either side of the reference point, while the Newton 2-point and Forward/Backward 3-point formulas are asymmetric.

Each of these methods has a noise free measurement error term [136] associated with their accuracy in measuring the derivative of a noise free signal. In all cases, the error is heavily influenced by the distance between the intervening points (h) (Table B-1). The actual signals measured will include noise and measurement error, so their impact on the derivative estimates must be evaluated. The measured signal from each of the sensors is shown in equation B-6, where the measured signal $f[n]$ is the combination of the actual signal $g[n]$ and a Gaussian noise/error term $Z[n]$.

$$f[n] = g[n] + Z[n] \quad [B-6]$$

The substitution of this noisy signal into each of equations B-1 through B-5 leads to equations B-7 through B-11, respectively.

$$f'[n] = \frac{g[n+1]-g[n]}{h} + \frac{2Z[n]}{h} \quad [B-7]$$

$$f'[n] = \frac{-g[n+2]+4g[n+1]-3g[n]}{2h} + \frac{4Z[n]}{h} \quad [B-8]$$

$$f'[n] = \frac{-3g[n-2]+4g[n-1]-g[n]}{2h} + \frac{4Z[n]}{h} \quad [B-9]$$

$$f'[n] = \frac{g[n+1]-g[n-1]}{2h} + \frac{Z[n]}{h} \quad [B-10]$$

$$f'[n] = \frac{-g[n+2]+8g[n+1]-8g[n-1]+g[n-2]}{12h} + \frac{1.5Z[n]}{h} \quad [B-11]$$

The predicted noise performance of each of the finite difference formulas is quite different where the derivative estimate is the sum of the desired value and a noise term that is based on the signal noise multiplied by a gain. The Central 2-point method has the best predicted performance, with a gain of $1/h$, while the Central 4-point method is 50% greater at $1.5/h$ with the others showing higher gains. The result is that the performance prediction for each of the finite difference formulas for noisy signals is different than for noise free signals, with Central 2-point having the best predicted noise performance.



The similarity of the signals from each of the sensors for each of the acceleration/deceleration profiles was measured through the maximum correlation coefficient (equation B-12) where $P[n]$ and $K[n]$ are two measures of the profile being compared. This equation assumes that both signals have the same sampling rate and the shift parameter n allows time alignment of the signals to provide the maximum correlation.

$$R = \max_{-3 \leq n \leq 3} \left(\frac{\sum_m P^*[m]K[m+n]}{\sqrt{E_P E_K}} \right) \quad [\text{B-12}]$$

$$\text{where } E_P = \sum_m (P[m])^2 \quad [\text{B-13}]$$

The correlations of the smart phone signals with the GPS and OBDII signals require that the higher sampling rate acceleration signal be decimated through a two-step process where the signal is low-pass anti-alias filtered (ideal, non-causal, linear-phase, Hann window with length 950, unity gain, zero delay, pass band corner 0.40Hz, stop band corner 0.48Hz).

$$R_{\text{decimate}} = \max_{\substack{-3 \leq n \leq 3 \\ 0 \leq l \leq 40}} \left(\frac{\sum_m P_{LP}^*[40m+l]K[m+n]}{\sqrt{E_{PLP} E_K}} \right) \quad [\text{B-14}]$$

Here, P_{LP} is decimated through the $40m$ term with the l shift parameter, allowing for micro-alignment at the 40Hz sampling rate and the n term allowing time alignment at the 1Hz sampling rate to ensure maximum correlation between signals (equation B-14).

B.2. Experimental Results

A typical velocity profile for the OBDII and GPS velocity sensors is shown in Figure B-1 with the upper plot shows the time domain, while the lower plots showing the power spectrum. The accuracy of GPS velocity measure is +/- 0.4km/hr [93] while the OBDII accuracy is specified at 10% (or +/- 4km/hr at 40km/hr). Automakers typically ensure that the speedometer readings (OBDII) are never lower than actual velocity so the actual vehicle calibration has an unknown offset. The two figures show the strong similarity of the velocity profiles, but also demonstrate that the OBDII signal contains a small amount of higher frequency noise and the calibration offset in the OBDII signal.

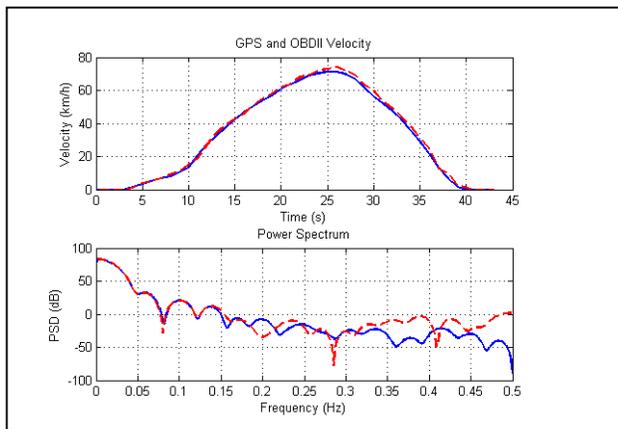


Figure B-1: Example time and frequency domain acceleration profiles. Blue-solid: GPS, Red-dashed: OBDII.

Figure B-2 compares the various smart phone accelerometer signals for the example GPS and OBDII profile in Figure B-1. The smart phone signals are shown after each has been low-pass filtered in preparation for decimation. Figure B-9 shows the similarity between the signals while also highlighting the variation between accelerometers in the smart phone devices.

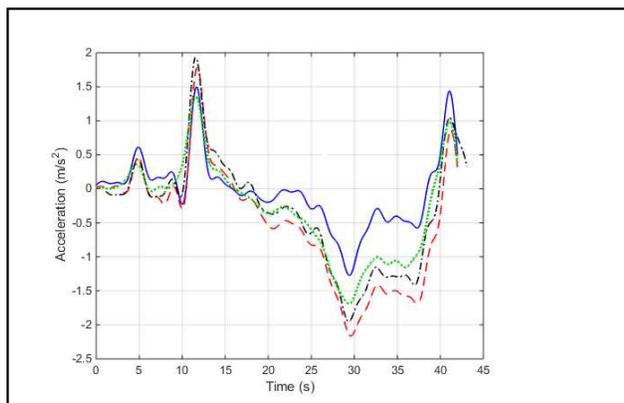


Figure B-2: Example acceleration profiles for the 3 accelerometers with the mean signal. Blue-solid: Z10A, Red-dashed: Z10B Black-dashdot: Z10C, Green-dot: Mean

Figures B-3 and B-4 provide comparisons between the Central 2-point and Forward 3-point difference formulas for a typical acceleration/deceleration profile. In the upper plot of Figure B-3, the acceleration signal was derived from both the GPS and OBDII velocity data using the Central 2-point formula, while in the lower plot, the acceleration signals were obtained using the Forward 3-point formula. The error in the Central 2-point acceleration estimates can be found (equation B-10) of



0.1m/s^2 for the GPS and 1 m/s^2 for the OBDII given an h of 1 second. The errors for the other derivative formulas scale per the noise gains.

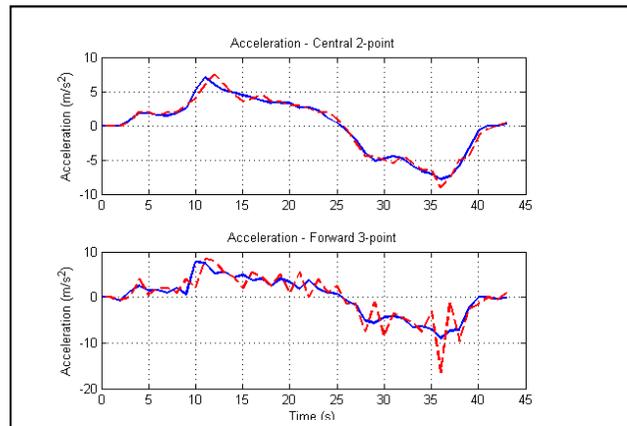


Figure B-3: Example acceleration signals comparing GPS to OBDII for two difference formulas. Blue-solid: GPS, Red-dashed: OBDII.

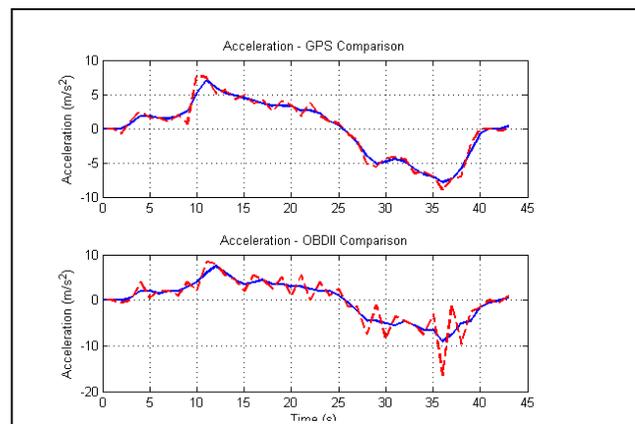


Figure B-4: Example comparison of acceleration signals for the Central 2-point and Forward 3-point formulas showing the resulting effects of noise in the resulting acceleration signal. Blue-solid: Central 2-point, Red-dashed: Forward 3-point.

Figure B-4 compares the derived signals with GPS signals being very similar while the OBDII derived signals differ significantly. The OBDII Central 2-point is very similar to both of the derived GPS signals, but the Forward 3-point signal demonstrates the effect of the higher frequency noise (shown in Figure B-1) being amplified by the difference formula (as predicted by equation B-9). Similar effects were seen in the Backward 3-point, Central 4-point and Forward/Backward 2-point as noise amplification varied as predicted by the differing noise gains.



Table B-2: Comparison of the smart phone acceleration profiles showing the maximum, minimum, mean correlation coefficient and standard deviation of mean. (n=16)

		Z10 A	Z10 B	Z10 C	Z10 mean
Z10 A	Max		0.974	0.973	0.984
	Mean		0.949	0.956	0.976
	Min		0.875	0.930	0.955
	StDev		0.025	0.012	0.008
Z10 B	Max	0.974		0.998	0.998
	Mean	0.949		0.973	0.990
	Min	0.875		0.831	0.931
	StDev	0.025		0.041	0.016
Z10 C	Max	0.973	0.998		0.998
	Mean	0.956	0.973		0.994
	Min	0.930	0.831		0.974
	StDev	0.012	0.041		0.006

The comparison of the correlation coefficient for the 16 acceleration / deceleration events between the 4 accelerometer signals is shown in Table B-2. For each profile, the correlation coefficient was measured using equation B-15 and down sampled 1Hz signals showing that the smart phone accelerometers provide results that are highly correlated, despite the signal variation.

Table B-3 shows the correlation coefficient comparison of the OBDII and GPS derived acceleration signals for each of the difference formulas with the highest correlations when Central 2-point is used for both signals even though the theory for noise-free signals predicts that Central 4-point should provide the best result. This is in line with the noise effects predicted in equations B-7 through B-11. The effect of the higher frequency content in the OBDII signal is very evident in the table, but the smaller effect of the lower noise GPS signal is also evident. The results for Forward and Backward 2-point are identical as these two acceleration signals are only 1 sample shifted. One method to reduce the effect of the noise within the formulas is to introduce a low pass filter ahead of the difference formula, but the Central 2-point formula has this filter “built-in” as it acts as a band pass filter.



Table B-3: GPS and OBDII correlation results for the difference estimates (n=16)

GPS	OBDII	Central 2-point	Central 4-point	Forward 3-point	Backward 3-point	Fwd/Bwd 2-point
Central 2-point	Max	0.991	0.985	0.887	0.878	0.955
	Mean	0.986	0.977	0.815	0.812	0.927
	Min	0.978	0.963	0.739	0.732	0.895
	StDev	0.004	0.006	0.043	0.042	0.018
Central 4-point	Max	0.989	0.983	0.890	0.880	0.957
	Mean	0.985	0.976	0.816	0.811	0.926
	Min	0.974	0.960	0.743	0.732	0.896
	StDev	0.004	0.007	0.043	0.042	0.019
Forward 3-point	Max	0.974	0.971	0.869	0.876	0.943
	Mean	0.966	0.959	0.814	0.807	0.917
	Min	0.937	0.924	0.752	0.739	0.855
	StDev	0.009	0.011	0.036	0.037	0.023
Backward 3-point	Max	0.981	0.979	0.907	0.871	0.954
	Mean	0.971	0.965	0.812	0.815	0.918
	Min	0.950	0.939	0.733	0.748	0.887
	StDev	0.009	0.010	0.048	0.036	0.019
Forward/Backward 2-point	Max	0.987	0.983	0.888	0.879	0.953
	Mean	0.980	0.971	0.814	0.814	0.929
	Min	0.970	0.960	0.734	0.731	0.892
	StDev	0.005	0.007	0.042	0.043	0.017



Table B-4: Correlation results comparing the GPS signal for the difference formula derivatives with the smart phone acceleration signals. (n=16)

		Central 2-point	Central 4-point	Fwd 3-point	Fwd 2-point	Bwd 3-point	Bwd 2-point
Z10 A vs GPS	Max	0.875	0.873	0.873	0.873	0.858	0.872
	Mean	0.710	0.709	0.689	0.704	0.721	0.709
	Min	0.639	0.635	0.619	0.634	0.655	0.633
	StDev	0.065	0.062	0.064	0.061	0.048	0.065
Z10 B vs GPS	Max	0.872	0.867	0.851	0.867	0.868	0.868
	Mean	0.823	0.822	0.800	0.817	0.828	0.822
	Min	0.706	0.706	0.708	0.707	0.699	0.705
	StDev	0.040	0.038	0.035	0.037	0.040	0.040
Z10 C vs GPS	Max	0.943	0.938	0.923	0.937	0.922	0.936
	Mean	0.838	0.837	0.812	0.832	0.843	0.837
	Min	0.790	0.791	0.757	0.784	0.806	0.788
	StDev	0.038	0.036	0.039	0.037	0.029	0.038
Z10 mean vs GPS	Max	0.879	0.876	0.872	0.876	0.859	0.874
	Mean	0.810	0.808	0.785	0.803	0.815	0.810
	Min	0.760	0.762	0.742	0.758	0.779	0.757
	StDev	0.034	0.031	0.035	0.031	0.021	0.034

Tables B-4 and B-5 provide the correlation coefficient comparisons of the smart phone acceleration signals to the GPS and OBDII accelerations. Once again the high frequencies in the OBDII reduced the correlations due to the noise amplification on the difference formulas. The tables show the strong correlation between the acceleration signals derived from the GPS position sensor and OBDII velocity sensor in comparison to direct acceleration measurement.



Table B-5: Correlation results comparing the OBDII signal for the difference formula derivatives with the smart phone acceleration signals. (n=16)

		Central 2-point	Central 4-point	Fwd 3-point	Fwd 2-point	Bwd 3-point	Bwd 2-point
Z10 A vs OBDII	Max	0.856	0.845	0.712	0.815	0.711	0.813
	Mean	0.712	0.705	0.594	0.676	0.613	0.678
	Min	0.622	0.613	0.500	0.579	0.535	0.580
	StDev	0.068	0.067	0.058	0.062	0.063	0.067
Z10 B vs OBDII	Max	0.875	0.862	0.765	0.832	0.788	0.844
	Mean	0.821	0.813	0.681	0.775	0.700	0.783
	Min	0.700	0.696	0.611	0.684	0.614	0.684
	StDev	0.044	0.044	0.049	0.042	0.054	0.045
Z10 C vs OBDII	Max	0.941	0.931	0.759	0.879	0.786	0.887
	Mean	0.836	0.827	0.688	0.787	0.710	0.792
	Min	0.771	0.759	0.609	0.729	0.632	0.729
	StDev	0.043	0.043	0.049	0.041	0.053	0.044
Z10 mean vs OBDII	Max	0.869	0.859	0.739	0.819	0.767	0.842
	Mean	0.808	0.801	0.667	0.761	0.687	0.768
	Min	0.742	0.732	0.590	0.699	0.616	0.698
	StDev	0.040	0.040	0.049	0.039	0.051	0.044