Concurrent Information Communication in Voice-based Interaction

By

Muhammad Abu ul Fazal

Supervised By

Dr. Sam Ferguson

Dr. Andrew Johnston

A Thesis Submitted in Fulfilment of the

DOCTOR OF PHILOSOPHY

School of Computer Science

Faculty of Engineering and Information Technology

University of Technology Sydney

July, 2019

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Muhammad Abu ul Fazal declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy in Information Technology, in the School of Computer Science, Faculty of Engineering and IT at the University of Technology Sydney.

This thesis is wholly my own work, unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature: Signature removed prior to publication.
Date: 23/07/2019

Acknowledgment

Firstly, I would like to express my sincere gratitude to my supervisors Dr Sam Ferguson and Dr Andrew Johnston. From my initial offer to be a visiting research student at the University of Technology Sydney (UTS) to the offer of completing a PhD from UTS with fully funded Scholarship, I am truly grateful for the opportunities. Each of my supervisors helped, supported, and guided me through my research at UTS in an exceptional and unforgettable manner. I shall always remain thankful to Dr Ferguson and Dr Johnston.

Dr Ferguson's insights taught me how to address a research problem scientifically, how to conduct standardised experiments, how significant it is to apply statistical tests on results, how to present results visually and also how to make an academic argument. These are but a few research skills from a, what seems, unending list that I learned from Dr Ferguson.

I am also thankful to the Faculty of the Department of Computer Science, Quaid-i-Azam University (QAU), Islamabad, Pakistan for offering me admission in PhD. I am thankful to my former research supervisor Dr Shuaib Karim for supervising me in QAU and guiding towards international opportunities. I am also thankful to the Higher Education Commission of Pakistan for awarding me an IRSIP scholarship to support my studies at UTS. I wish there were an official agreement of researching collaborations between the QAU and UTS, which would have enabled me to continue my PhD as a joint degree program from both the universities together.

I am thankful to the office staff at the School of computer science, UTS, in particular, Margot, Janet, Teraesa and Reshma. All team members I met within

the School, provided a conducive research environment. I am also thankful to the staff at Graduate Research School at UTS, in particular, Jing and Grandia, who always remained welcoming of my questions and queries, resolving my problems in a timely manner. I also pay thanks to the office staff at CS department at QAU for the support they provided me.

Regards

Muhammad. Abu ul Fazal

Dedication

My thesis is dedicated to the people who have supported my goals, inspired me and challenged me academically to make it to this day.

Reflecting on my path that led me to this day, after completing matriculation, circumstances forced me to abandon studies for what I thought would be a permanent arrangement. I provided support to my family and assisted my father run a medical store. Happenstance led a past teacher to the medical store and I spoke to him about resuming studies. He suggested I join his evening academy to prepare for 12th-grade exams while still helping my father. I joined the academy.

I fondly remember one particular day at the academy when I inadvertently mentioned my name and my aspiration to my childhood friend Adnan; "Hello, I am Professor Doctor Muhammad Abu ul Fazal". He asked *will you be*? I replied *maybe*! The title combined with my name pleasantly haunted my mind.

In 2003 when I was in my 3rd year of completing the Bachelor of Science, one of my teachers suggested that I extend my degree and complete the Master of Science so that I could later enrol in a PhD. Today, I submit my Doctoral thesis at the University of Technology Sydney. My earlier 'maybe' has become a 'yes'.

For this day today, first and foremost credit goes to my mother, who during a severe financial crisis when I asked about my possibility to study a Bachelors, she replied: "Fazal, take admission and do it". I saw her working days and nights taking on additional teaching activities to manage my education fees.

Thank you, Ammi. Thank you Great Abba G, you have always been an inspiration to me and supported a lot from the background. Thank you to my sisters, Saima Umar, Nafisa Umar, Shabiah Umar, Asma Junaid and my brothers

Junaid Umar, Safi Ullah, Akmal Ata, Waseem Asif, and lovely nieces and nephews for always remaining a great source of support and encouragement. Ghulam Mustafa, my dear friend, thanks to you too.

When I started my PhD my immediate family comprised of my wife Hadiah, our daughter Mahrukh and me. In the middle of my PhD, my son Sherdil joined our family. Without the support of Hadiah, I would have never been able to complete my doctorate studies abroad. Mahrukh, and Sherdil, I owe you a lot. The time I spent in Australia doing my PhD, you in your childhood living in Pakistan - you were the real owners of this time. Hadiah, I am eternally thankful to you. Thank you to my great father-in-law and mother-in-law and as well as siblings of my wife who offered well wishes for my studies and prayed for my success.

Abstract

Speech-based information is primarily communicated to users sequentially; however, users are capable of obtaining information from multiple sources concurrently. This fact implies that the sequential approach is under-utilising human perception capabilities and restricting users to perform optimally. In this research, two informal studies and two experiments were carried out for investigating concurrent communication of multiple voice-based information streams. The informal studies were carried out to understand users' interest and expectations in concurrent information communication and to examine whether users can comprehend concurrent information. In the first experiment, different designs for speech-based multiple information communication and the depth of comprehension by users in each design were tested. In the second experiment, various combinations of information streams presented concurrently and their viability regarding cognitive load were tested.

The results of the first study manifested user's interest in concurrent information communication design and supported the argument that users are able to discriminate and understand the concurrent voice streams using their selection and attention abilities. The results led to the second study, where users, including visually challenged users, expressed their expectations from such system and shared how would they prefer to interact with the systems providing concurrent information communication. Based on user's feedback, a web-based *'Vinfomize framework'* is designed to allow for concurrent communication of multiple information streams to users. Findings from the third study showed that concurrent speech-based information designs, involving intermittent form and a spatial dif-

ference in sources of the streams, provide satisfying comprehension of the content. The study further showed that users could comprehend both the main information and the detailed information. The fourth study showed that the perceived cognitive workload for the listening task in baseline condition and concurrent combinations remain the same; however, users response in preference and frequently using different combinations remain significantly lower than the baseline condition. The fourth study also showed that the combinations created with music were preferred the most by the users in concurrent combinations, followed by the song. From the information types providing speech-based information (non-music/song), result shows the intermittent form of communication creates the low cognitive workload in voice-based information communication.

Our research findings contribute to providing improvements in methods to communicate voice-based information efficiently under a large variety of application fields.

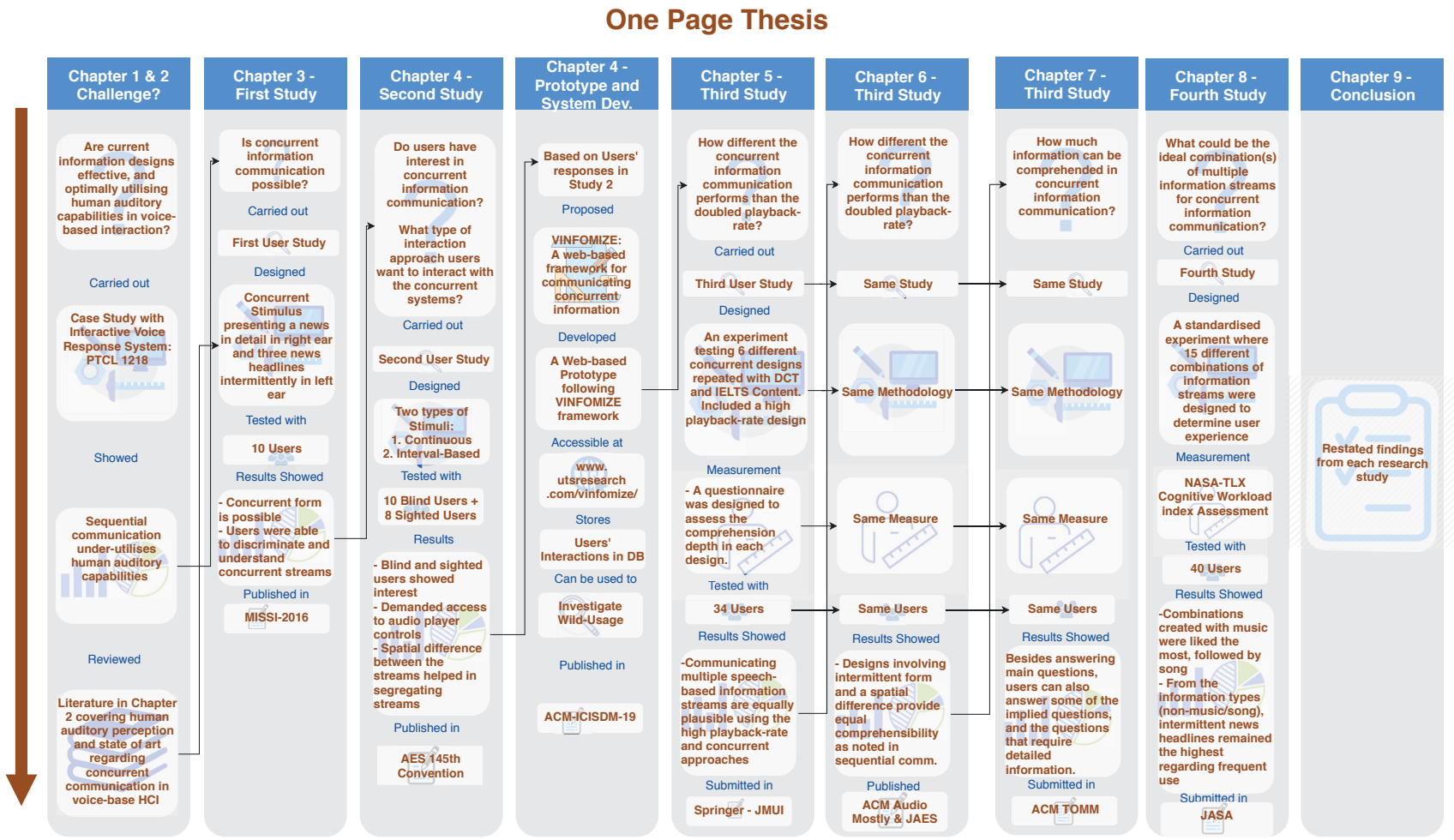

Figure 1 : **Research Overview:** Involved 102 users in total including visually challenged users in 4 different user studies, recorded 24000 user responses, and processed them to come up with the findings mentioned in this thesis.

List of Publications

1. M. A. u. Fazal and M. Shuaib Karim, "Multiple information communication in voice-based interaction," in *Advances in Intelligent Systems and Computing*. Springer, pp. 101–111

2. M. A. u. Fazal, S. Ferguson, M. S. Karim, and A. Johnston, "Concurrent Voice-Based Multiple Information Communication: A Study Report of Profile-Based Users' Interaction," in *145th Convention of the Audio Engineering Society*. Audio Engineering Society, 2018

3. M. A. u. Fazal, S. Ferguson, M. S. Karim, and A. Johnston, "Vinfomize: A framework for multiple voice-based information communication," in *Proceedings of the 2019 3rd International Conference on Information System and Data Mining*. ACM, 2019, pp. 143–147

4. ——, "Investigating Efficient Speech-based Information Communication - A Comparison between the High-rate and the Concurrent Playback Designs," *Journal on Multimodal User Interfaces (JMUI)*, vol. -, no. -, pp. 1–8, 2019, submitted

5. M. A. u. Fazal, S. Ferguson, and A. Johnston, "Investigating Concurrent Speech-based Designs for Information Communication," in *Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion*, ACM. New York, NY, USA: ACM, 2018, pp. 1–8

6. ——, "Investigating Concurrent Speech-based Designs for Efficient Information Communication - Extended Analysis," *Journal of the Audio Engineering Society (JAES)*, vol. -, no. -, pp. 1–8, 2019, submitted

7. M. A. u. Fazal, S. Ferguson, and A. Johnston, "Evaluation of Information Comprehension in Speech-based Designs for Concurrent Audio Streams," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. -, no. -, pp. 1–18, 2018, submitted

8. ——, "Investigating cognitive workload in concurrent speech-based information communication," *The Journal of the Acoustical Society of America (JASA)*, vol. -, no. -, pp. 1–20, 2019, submitted

9. A. Hussain, M. A. u. Fazal, and M. S. Karim, "Intra-domain user model for content adaptation," in *Smart Innovation, Systems and Technologies*. Springer, 2015, pp. 285–295

Contents

Publications **x**

List of Tables **xvii**

List of Tables xvii

List of Figures **xviii**

List of Figures xviii

1 Introduction **1**

1.1 Voice-based Interaction 1

1.2 Interactive Voice Response System: PTCL Case-Study 3

1.3 Human Abilities of Listening to Multiple Information Simultaneously 6

1.4 Multiple Information Communication in Voice-based Interaction . . 8

1.5 Motivating Scenarios 10

1.6 Research Aim 12

1.6.1 Research Questions 12

2 Literature Review **14**

2.1 Human Auditory Perception 14

2.2 Psychological Studies Exploring Concurrency 22

2.3 Contemporary Research Studies on Concurrent Speech Interface . . 29

2.4 Summary of the Angles Reviewed 36

2.5 Our Research 36

3 Viability of Concurrent Information Communication 38

3.1 Aims & Motivation . . . 38

3.2 Methodology . . . 39

3.2.1 Stimulus Material . . . 39

3.2.2 Design . . . 39

3.2.3 Participants . . . 40

3.2.4 Questionnaire . . . 41

3.3 Results . . . 41

3.4 Discussion . . . 46

3.5 Limitations & Future Work . . . 47

4 Viable Interaction Approach to Interact with the System Communicating Concurrent Information? 49

4.1 Aims & Motivation . . . 50

4.2 Investigation . . . 50

4.2.1 Participants . . . 51

4.2.2 Study 2-A - Continuous: Stimulus & Questionnaire . . . 52

4.2.3 Study 2-B - Continuous and Intermittent: Stimulus & Questionnaire: . . . 54

4.2.4 Protocol . . . 55

4.3 Results & Analysis . . . 57

4.3.1 Qualitative Analysis . . . 57

4.3.2 Quantitative Analysis . . . 59

4.4 Discussion . . . 60

4.5 Vinfomize Framework . . . 62

4.6 Web-based System Development based on the proposed Framework 67

4.7 Limitations & Future Work . . . 68

5 A Comparison between High Playback-rate and Concurrent Design 71

5.1 Aims & Motivation . . . 72
5.2 Method . . . 73
5.2.1 Participants . . . 73
5.2.2 Design . . . 73
5.2.3 Material . . . 77
5.2.4 Stimuli Information . . . 78
5.2.5 Measures . . . 79
5.2.6 Questionnaire . . . 79
5.2.7 Apparatus . . . 80
5.2.8 General Procedure . . . 81
5.3 Results . . . 81
5.3.1 Proportion Analysis . . . 82
5.3.2 Comprehension Performance Analysis . . . 83
5.3.3 Comprehension Depth Analysis . . . 85
5.4 Discussion . . . 87
5.5 Limitations and Future Work . . . 88

6 Investigating Concurrent Speech-based Designs for Information Communication 90

6.1 Aims & Motivation . . . 90
6.2 Methodology . . . 91
6.3 Results . . . 91
6.3.1 Concurrent Designs Analysis . . . 92
6.3.2 Intermittent Designs in Detail . . . 94
6.4 Discussion . . . 98
6.5 Limitations & Future Work . . . 99

7 Evaluation of Information Comprehension Depth in Speech-based Concurrent Designs 102

7.1 Aims & Motivation . . . 102

7.2 Method . . . 103

7.3 Results . . . 103

7.3.1 Overall Comprehension Comparison between Streams . . . 104

7.3.2 Comprehension Depth for Concurrent Condition . . . 106

7.3.3 Users' Experience . . . 108

7.4 Discussion . . . 109

7.5 Limitations and Future Work . . . 113

8 Evaluating Various Combinations of Information Streams in Concurrent Information Communication 115

8.1 Aims & Motivation . . . 115

8.2 Method . . . 116

8.2.1 Participants . . . 116

8.2.2 Design . . . 117

8.2.3 Material . . . 119

8.2.4 Stimuli Information . . . 119

8.2.5 Measures . . . 120

8.2.6 Apparatus . . . 121

8.2.7 General Procedure . . . 122

8.3 Results . . . 122

8.3.1 Baseline vs. Concurrent . . . 123

8.3.2 Concurrent Combinations . . . 125

8.3.3 Information Streams Impact in Concurrent Communication 133

8.3.4 Impact of Presentation in Left — Right Ears . . . 135

8.4 Discussion . . . 138

8.5 Limitations and Future Work 143

9 Conclusion **145**

Bibliography **149**

Appendix A A Letter to the Institute Requesting the Participation of Visually Challenged Persons in Study II **176**

Appendix B Vinformize-based System Interface **178**

Appendix C Ethics Application for Study III **180**

Appendix D Interface with Selected Playable Audio Files URLs, and Questionnaire for Study III **190**

Appendix E Ethics Application for Study IV **209**

Appendix F Interface with Selected Playable Audio Files URLs, and Questionnaire for Study IV **215**

Appendix G Publication 1 **221**

Appendix H Publication 2 **232**

Appendix I Publication 3 **238**

Appendix J Publication 4 [Submitted] **246**

Appendix K Publication 5 **257**

Appendix L Publication 6 [Submitted] **266**

Appendix M Publication 7 [Submitted] **276**

Appendix N Publication 8 [Submitted] **295**

Appendix O Publication 9 **346**

List of Tables

3.1 Participants Demography 40
3.2 Questionnaire & Users Responses 42

4.1 Users' Profiles 51
4.2 Profile-based User Groups 52
4.3 Questionnaire Study 2-A 54
4.4 Questionnaire Study 2-B: 56

5.1 Speech-based Concurrent Communication Designs 75
5.2 Results of the One-to-one Proportion Comparison 86

7.1 Proportion Comparison Test: 105
7.2 One-to-one Proportion Comparison Test Results 108

8.1 Combinations of Different Types 117
8.2 *Post hoc* Tukey HSD Analysis Comparing Concurrent Scales scales . 125
8.3 *Post hoc* Tukey HSD Analysis Comparing Speech-based Concurrent Combinations 128
8.4 *Post hoc* Tukey HSD Analysis Comparing Music-based Concurrent Combinations 131
8.5 *Post hoc* Tukey HSD Analysis Comparing Stream Types 133

List of Figures

1 Research Overview viii

1.1 Voice-based Interaction 2
1.2 IVR Case Study 4
1.3 Auditory Perceptual Dimensions 7
1.4 Auditory System Mechanisms and Processes 8
1.5 Overlay Example 9

2.1 Schematic View of the Periphery Auditory System 15

3.1 Distinguish Secondary Voice 44
3.2 Interesting News Topic 45
3.3 Multiple Information Preference 45

4.1 Continuous Stimulus Design for Study 2-A 53
4.2 Continuous and Intermittent Stimulus Design for Study 2-B 53
4.3 Group-wise Responses in Both Studies 59
4.4 Group-wise Responses in Basic and Advanced Questions 60
4.5 All Users' Responses in Both Studies 61
4.6 Vinfomize Framework 63

5.1 Stimuli Designs 76
5.2 The Proportion of User Responses 83

5.3 Percentage of Correct Answers 84
5.4 Comprehension Depth 86

6.1 Proportion of Users Responses 92
6.2 Percentage of Correct Answers 94
6.3 Competing and Non-competing Questions 95
6.4 Percentage of Correct Answers 96
6.5 Deep Intermittent Analysis 97

7.1 Percentage of Correct Answers 104
7.2 Users' Comprehension in each Stream w.r.t MIS, MII, DTS, DTI . . 107
7.3 Participants Preference 109

8.1 Concurrent Stimulus Design 118
8.2 User Experience in Baseline Condition 123
8.3 Users' Experience in Concurrent Communication 125
8.4 Users' Experience in each Combination 126
8.5 Users' Experience regarding each Information Type 134
8.6 Impact of Users' Experience (Stress) 137
8.7 Impact of Users' Experience (Acceptance) 137
8.8 Perceived Workload Index Score 138
8.9 Ratings for Frequent and Like Scales 139
8.10 Order of Combinations 140
8.11 Order of Information Types 141