

Considering Design for Automatic Speech Recognition in Use

Ben James Kraal, *BInfTech(Hons)*
School of Information Sciences and Engineering
University of Canberra ACT 2601

A thesis submitted in fulfillment of the requirements of the
Degree of Doctor of Philosophy.
April 2006

Certificate of Authorship of Thesis

Except where clearly acknowledged in footnotes, quotations and the bibliography, I certify that I am the sole author of the thesis submitted today entitled *Considering Design for Automatic Speech Recognition in Use*.

I further certify that to the best of my knowledge the thesis contains no material previously published or written by another person except where due reference is made in the text of the thesis.

The material in the thesis has not been the basis of an award of any other degree or diploma except where due reference is made in the text of the thesis.

The thesis complies with University requirements for a thesis as set out in <http://www.canberra.edu.au/secretariat/goldbook/forms/thesisrqmt.pdf>

Ben Kraal
Candidate

Michael Wagner
Chair of Supervision Panel

October 2, 2006

Acknowledgements

Academically, there are many people to thank. My supervisors, Michael Wagner, Penny Collings and Anni Dugdale are, above all, patient. It was Michael who led me to automatic speech recognition as an application area to study from the point of view of interaction design. He introduced me to Penny Collings who brought her wealth of experience in HCI to my supervision team. When it became clear that I was going down the path of using social methods to conduct much of my research it was Penny who introduced me to Anni Dugdale, sociologist and Actor-Network theorist. All of my supervisors were free with their time and rich with advice and guidance. Since the first day I arrived at the University of Canberra they have treated me with much respect.

I must thank the anonymous interview subjects who gave up their work time and allowed me to ask questions about their work, their lives and their injuries. I'd especially like to thank Sue Woodward for making it possible for me to meet my interview subjects. I'd like to thank the Hansard Editors who gave up their work time to tell me about their work and how it was done.

All the people at the ACT Magistrates Court were very generous with their precious time. Ron Cahill, Chief Magistrate (among other duties) was remarkably accessible for someone so busy and his staff and their colleagues were unfailingly happy for me to ask, what must have been to them, simple questions about their work and how they went about doing it.

My parents and sister were very supportive when I left the city of my birth to travel interstate to study. My inability to explain my research clearly left them in the dark as to what I was doing for the best part of four years.

Finally, I'd like to thank my wife, Marice for all her love and support, both emotional and financial, during the time I've taken to produce this thesis. She shared in my small victories and commiserated my defeats. Young Mr. Daniel, who came along towards the end, provided much distraction and laughter. I love them both.

List of Publications

1. Ben Kraal, Michael Wagner, and Penny Collings. Improving the design of dictation software. In *Australian Speech Science and Technology Conference*, University of Melbourne, Victoria, Australia, November 2002. ASSTA.
2. Ben Kraal, Penny Collings, Anni Dugdale, and Michael Wagner. An ethnography of speech recognition. In *OZCHI 2004*, University of Wollongong, NSW, November 2004. CHISIG.

Abstract

Talking to a computer is hard. Large vocabulary automatic speech recognition (ASR) systems are difficult to use and yet they are used by many people in their daily work. This thesis addresses the question: How is ASR used and made usable and useful in the workplace now?

To answer these questions I went into two workplaces where ASR is currently used and one where ASR could be used in the future. This field work was done with designing in mind. ASR dictation systems are currently used in the Australian Public Service (APS) by people who suffer chronic workplace overuse injuries and in the Hansard department of Parliament House (Hansard) by un-injured people.

Analysing the experiences of the users in the APS and at Hansard showed that using an ASR system in the workplace follows a broad trajectory that ends in the continued effort to maintain its usefulness. The usefulness of the ASR systems is 'performed into existence' by the users with varying degrees of success. For both the APS and Hansard users, they use ASR to allow work to be performed; ASR acts to bridge the gap between otherwise incompatible ways of working.

This thesis also asks: How could ASR be used and made usable and useful in workplaces in the future? To answer this question, I observed the work of communicating sentences at the ACT Magistrates Court.

Communicating sentences is a process that is distributed in space and time throughout the Court and embodied in a set of documents that have a co-ordinating role. A design for an ASR system that supports the process of communicating sentences while respecting existing work process is described.

Moving from field work to design is problematic. This thesis performs the process of moving from field work to design, as described above, and reflects the use of various analytic methods used to distill insights from field work data.

The contributions of this thesis are:

- The pragmatic use of existing social research methods and their antecedents as a corpus of analyses to inspire new designs;

- a demonstration of the use of Actor-Network Theory in design both as critique and as part of a design process;
- empirical field-work evidence of how large vocabulary ASR is used in the workplace;
- a design showing how ASR could be introduced to the rich, complicated, environment of the ACT Magistrates Court; and,
- a performance of the process of moving from field work to design.

Contents

Acknowledgements	iii
List of Publications	iv
Abstract	v
List of Figures	xii
1 Introduction	1
1.1 Research Origins	2
1.2 ASR in Popular Culture	5
1.2.1 Automatic Speech Recognition in Science Fiction	6
1.2.2 Reports on the Use of Automatic Speech Recognition in Non-Fiction	11
1.2.3 Summary	14
1.3 Research Questions	15
1.4 Thesis Statement	15
1.5 Thesis Plan	16
2 Literature Review	19
2.1 Motivation for looking at Speech Recognition	20
2.2 Usability Studies of Dictation Systems	23
2.3 Dialog-driven Interfaces	31
2.4 The Importance of Supporting the Task	33
2.5 Recognising “Natural” Speech	33
2.5.1 Broadcast Automatic Speech Recognition	34
2.5.2 Using Recognised “Natural” Speech	35
2.6 Summary	36

3	Tools to Think With: Methods	37
3.1	Fieldwork for Design	40
3.2	Analysis for Design	42
3.2.1	Using the Locales Framework	43
3.2.2	Using Actor-Network Theory	47
3.3	The Emergence of Design from Analysis	51
3.3.1	Concepts Used in Design	52
3.3.2	Ways to Design	55
3.4	Summary	57
4	Automatic Speech Recognition Users	59
4.1	The Hansard Department	60
4.1.1	Scenario: Paul the Hansard Editor	61
4.2	Injured ASR Users in the Public Service	67
4.2.1	Scenario: Kelly, the middle level manager	68
4.3	Comparison of Scenarios	72
5	The ACT Magistrates Court	73
5.1	Communicating Outcomes Defined	74
5.2	Court Workers	77
5.2.1	The Defendant's Folder	78
5.2.2	The Magistrate's Associate	80
5.2.3	The Magistrate	83
5.2.4	The Monitor	87
5.2.5	The List Clerk	89
5.2.6	The After Court	89
5.2.7	The Bail Office	91
5.3	Summary	94
6	Analysis of Ethnography	95
6.1	The Problem with ASR Applications	96
6.2	Deciding to Use	98
6.2.1	Reasons for Use	99
6.2.2	Previous Experience with Similar Tools	101
6.2.3	IT Savvy	103

6.3	Leaping into Use	106
6.3.1	Training	106
6.3.2	Hardware and Software and Environment	108
6.3.3	Fit with Work	111
6.4	Sustaining Use	114
6.4.1	Necessity of Use	114
6.4.2	Technical Support	115
6.4.3	Becoming the Conductor	119
6.5	Summary of Analysis	123
7	Analysis and Design for the Court	127
7.1	Scenarios for ASR in the Courtroom	129
7.1.1	Background	129
7.1.2	Scenario: Techno-Utopia	131
7.1.3	Scenario: Dysfunctional Dystopia	133
7.1.4	Explaining the Scenarios	135
7.2	Difficulties for ASR in the Court	136
7.2.1	Translating ASR and the Court	137
7.2.2	Interesting and Enrolling Actors	138
7.2.3	Points of Passage and Trials of Strength	143
7.3	Re-imagining ASR	144
7.3.1	Scenario: Interface for Court Audio Access	150
8	Reflections on Methodology	159
8.1	Social Research for ASR	163
8.2	Using the Locales Framework	165
8.3	Using Actor-Network Theory	167
8.4	Summary	169
9	Conclusion	171
9.1	Contributions	171
9.2	Implications for Design	174
9.3	Areas for Future Research	176
9.3.1	Automatic Speech Recognition	177
9.3.2	Other Technologies	177

9.3.3 Actor-Network Theory in Design	178
A Socially-Inspired Design	181
A.1 A Novel Way of Working	182
A.2 Contemporary Versions of that Work	182
A.3 Make the Practical Actions Clear	183
A.4 A New Design Emerges	183
A.5 Construction	184

List of Figures

3.1	A simple trajectory with a single start and end point.	45
3.2	Many trajectories, originating and ending in different locations, though having much in common.	46
3.3	The Actor-Network Theory Research Frame developed from Callon (1986b).	49
4.1	Plan diagram of Paul's Office.	62
5.1	Plan Diagram of Court Room One at the ACT Magistrates Court. Note: not to scale.	75
5.2	An example time-line for a defendant's appearance in court.	76
5.3	Dramatis Personae: Actors in the Performance of Communicating Outcomes in the Court.	77
5.4	A typical stamp used by a magistrate.	79
5.5	This stamp is applied to many bench sheets.	79
5.6	This stamp is used when the magistrate decides there is no case to answer.	79
5.7	Plan view of Claire's desk inside the Court.	83
5.8	Plan view of the Monitor's Booth between Court Room One and Two. Note: not to scale.	87
5.9	Plan view of the Bail Office. Note: Not to scale.	91
6.1	A model of the trajectory of using ASR in an organisation.	125
7.1	The touch-screen interface in the ready state.	132

- 7.2 Microphone placement in the ACT Magistrates Court. Red dots represent the microphones. See also figure 5.1, 75. . . . 139