



COVER SHEET

This is the author-version of an article published as:

Kraal, Ben J and Dugdale, Anni and Collings, Penny (2006) Scenarios for Embracing Errorful Automatic Speech Recognition . In *Proceedings OZCHI 2006*, pages pp. 341-344, Sydney.

Accessed from <http://eprints.qut.edu.au>

Copyright 2006 ACM Press

Scenarios for Embracing Errorful Automatic Speech Recognition

Ben Kraal^{1,3}, Anni Dugdale² and Penny Collings¹

¹School of Information Sciences and Engineering

²School of Sociology
University of Canberra ACT Australia
{Anni.Dugdale,
Penny.Collings}@canberra.edu.au

³Queensland University of Technology
School of Design

b.kraal@qut.edu.au

ABSTRACT

Errorful speech recognition can be embraced in the design of automatic speech recognition (ASR) support for the Magistrates Court. In this paper we describe processes and scenarios that led to a design by examining work practices and considering a more realistic understanding of ASR technology than is promoted in ASR literature.

This paper also uses scenarios in a novel way to package and communicate field work data in a way that is accessible to a wide range of stakeholders.

Author Keywords

Automatic Speech Recognition; social worlds; scenarios; field work

ACM Classification Keywords

H.5.1 Multimedia Information Systems

H.5.2 User Interfaces (D.2.2, H.1.2, I.3.6)

H.5.3 Group and Organization Interfaces

INTRODUCTION

A group that I worked with while completing my PhD was approached by the Chief Magistrate of the Magistrates Court (the Court) to investigate the introduction of Automatic Speech Recognition (ASR) technology to the courtroom for use by the magistrate in the process of communicating outcomes, or decisions (Kraal, Collings et al. 2004; Kraal 2006).

The Chief Magistrate asked for an ASR system that could replace his existing manual system of handwriting and rubber stamps. When recording an outcome, a magistrate has the option of using a one or a combination of large rubber stamps (see Figure 1) and handwriting to record a sentence. They will also speak the sentence aloud. The Chief Magistrate thought that, since he was speaking the

sentence, an ASR system could be employed to record what he had said and remove the need for him to record decisions on paper. Writing outcomes down is time consuming, particularly as one defendant may be appearing on many charges, each of which will require a decision from the magistrate.

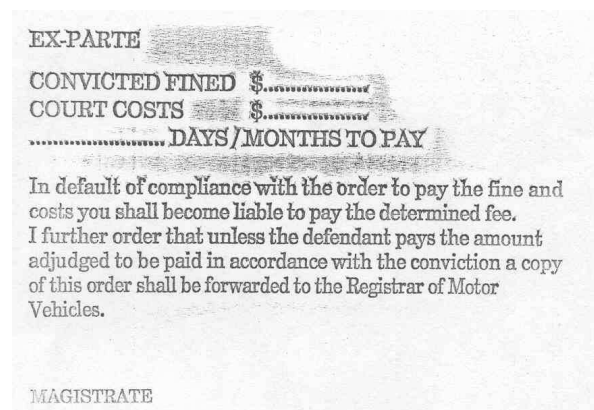


Figure 1: A magistrates stamp imposing a fine.

After some preliminary field work at the Court it emerged that the magistrate's act of speaking an outcome was not an event that was self-contained but was the beginning of a process distributed in space and time throughout the Court and led to the recording of an outcome in many different places and for many different purposes. This contrasted with the Chief Magistrate's view of the process as one which was enacted by him and contained within the courtroom.

COMMUNICATING INSIGHTS FROM FIELD WORK

Having completed field work at the Court, it became necessary to be able to communicate what had been seen to software designers, to ASR designers and engineers and to the Court itself. Presenting field notes, tape recordings and transcripts of interviews, photos and so on is not practical for many reasons, and is likely to overwhelm anyone who is not interested in the full thick detail of the situation. More importantly, it is not necessary to present the complete detail of the field work to a designer or any other interested party. Creating a software specification directly from the field work was quickly dismissed as being too focused on one possible interpretation of the situation as well as being too generic.

OzCHI'06, November 22-24, 2006, Sydney, Australia.
Copyright the author(s) and CHISIG
Additional copies are available at the ACM Digital Library
(<http://portal.acm.org/dl.cfm>) or ordered from the CHISIG secretary
(secretary@chisig.org)

OZCHI 2006 Proceedings ISBN: x-xxxxx-xxx-x

It is difficult to capture the sociality of a situation in a dry specification document.

The work of Bødker (2000) was influential in inspiring the use of scenarios to capture the detail of a social situation in such a way that the specifics of technology could also be described.

These scenarios are set in the Magistrates Court during the “A-list”. The A-list occurs every morning that court is in session and is the first appearance of anyone who has to appear in court, whether they were arrested the previous evening or they are responding to a summons. The most complicated cases in the A-list are set over to a future date and easy or quick cases are dealt with as they are called. The A-list is always pressed for time and the magistrate presiding over an A list session moves through cases at quite a pace.

The Techno-utopia Scenario

In this scenario, the use of an ASR application is presented as perfect. Everything works and does so very simply. Additionally, this scenario views the process of recording and communicating sentences as being the sole preserve of the magistrate. The various duties of the magistrate, List Clerk and Magistrates Associate are derived from field work at the Court. The purpose of this scenario was to show that we understood the Chief Magistrate’s point of view.

The Dysfunctional Dystopia Scenario

In this scenario, the same basic technology is considered as in the fantasy scenario. The difference here is that the technology is shown as it breaks. By making the potential pitfalls clear it is possible to initiate a discussion with stakeholders about whether the disadvantages of the proposed technology are overcome by the advantages.

ANALYSING THE SCENARIOS

Contrasting the scenarios shows that the introduction of ASR to the court does not just require a computer, but a microphone or system of microphones, a printer, a means to engage the ASR system when necessary and contingency plans when some or all of the interconnected technologies fail. Where the techno-utopia scenario shows how simple the system could be, the dystopian scenario shows that the same technologies could be tremendously disruptive not just to the large-scale running of the courtroom but also the small-scale interpersonal interactions between the magistrate and the associate, as illustrated when the too-short microphone cord prevents Mr Cowley from having a private word with Claire, reducing him to facial gestures.

Neither the specifically technical nor the specifically non-technical aspects of introducing an ASR system to the court are responsible for the difficulties involved in such an introduction. Solving the problems in the technical sphere but ignoring the non-technical problems does not make a future system useful or usable. Both the technical and non-technical must be considered together in order for the design of a future ASR system to take into account the complex environment of the court.

What the scenarios do not show is the work of people behind the scenes. The scenarios only show the courtroom itself and not the work done after court, which, while very important for the smooth running of the court, was deliberately left out of the fantasy and dystopian scenarios so that they could focus on the magistrate’s interaction with the imagined ASR system.

CONSIDERING SPEECH RECOGNITION FOR THE COURT

Using ASR productively in the Magistrates Court courtroom is fraught with difficulty. The courtroom environment is complex, both from work process and social perspectives. Automatic speech recognition technology is currently errorful in nature and its use in the courtroom will require the assemblage of a body of associated technologies in order to make it useful. In this section the language of Actor-Network Theory (Callon 1986; Latour 1987; Law 2003) is used to describe the difficulties involved in introducing ASR to the Court. In actor-network theory, assemblages of heterogeneous objects and people are termed networks of actors. Technical and non-technical elements are given equal weight in analysis.

The court and the process of communicating outcomes can be considered a stable (actor) network. The proposed introduction of an ASR system will necessarily destabilise the existing network and successful use of ASR will require that a new network be established and stabilised. By problematising the existing way of communicating outcomes and proposing ASR as the solution, I am suggesting that an ASR system become the obligatory passage point (Callon 1986) for the system, that is, to make the ASR system indispensable.

Because we have previously said that ASR is flawed (Kraal, Collings et al. 2004), suggesting, as we do here, that an ASR system become indispensable to the Court may seem counter-intuitive. However, by viewing the introduction of ASR to the Court as a design exercise to solve the problem of introducing ASR to the Court in such a way as to make it useful, these apparently contradictory points of view can be reconciled.

Using an automatic ASR system at the Magistrates Court will involve translating the ASR system and the Court. The Court’s interests are the administration of the law and the accurate recording of decisions. The ASR system’s “interest” is recognising speech. To allow the ASR system to pursue its interest without interference, many actors will need to be enrolled in the new network. Allowing the Court to continue to pursue its interests with as little interference as possible from the ASR system means that some aspects of the Court’s existing work process will have to change. As the courtroom itself and the procedures established for working there are steeped in tradition and rich with meaning, it would be difficult and even dangerous to drastically change them to accommodate an ASR system. Our field work in the courtroom has shown that much of the work done in the courtroom established and maintains the authority of the magistrate. Introducing an errorful system for use, live,

by a person in authority could set them up for ridicule or embarrassment if that system should fail. Conversely, our field work has also shown that much of the work of communicating sentences at the Court is performed behind the scenes in the “back room”. These back room workers use the bench sheet and other documents to interpret what was said in court and record the outcomes from various court sessions in ways that result in the magistrates’ orders being carried out. Where the work done in the courtroom is relatively resistant to change, work done in the back room is more malleable and therefore more open to translation to accommodate a system that uses ASR to replace an existing work practice. How these translations (or changes) will begin to be achieved is described in the next section.

Re-imagining Speech Recognition

To use ASR in the Magistrates Court necessitates that ASR, as a technology, be re-imagined. In actor-network terms it can be said that ASR needs to be translated in order to work at the Court. Often, ASR applications are seen as being a replacement for typing to be used by one user—the dictation paradigm. In the dictation paradigm, an ASR application is used to replace a secretary who takes dictation as the user speaks. However, this is not the only paradigm for the use of ASR.

The re-imagined form of ASR that could work for the Court would not use the one-user-to-one-computer model of dictation but a model where the users and computers are distributed in space and time as the work process of the Court is distributed in space and time. Inherent in this distributed model is the fact that the person whose speech is recognised is not necessarily the person working with the transcript generated by the ASR system. Distributing the computers involved allows separation of work tasks and recognition tasks as well as allowing multi-pass ASR (Whittaker, Hirschberg et al. 1999; Whittaker, Hirschberg et al. 2002) which can improve the accuracy of hard-to-recognise speech by allowing a recogniser to refine a transcription.

As stated in the previous section, the elements of the Court that are most plastic, and therefore easiest to change, are the detail of the work process of the “back room”, particularly the after-court section. This is not to say that these elements will be easy to change, just that they are easier to change than, say, the physical layout of the courtroom.

Having identified the following elements that are particularly resistant to change this design does not attempt to encroach on their existence, though it will necessarily have follow-on effects that cannot be predicted. These resistant elements are the social world of the Court, the “theatre” of the courtroom, the Court room layout, as it influences the social world of the Court, the work process in courtroom and all public-facing areas of the Court; and the requirement to record decisions on outcomes made by the magistrate during court.

The next section describes, in scenario form, an ASR interface for the Court inspired by the work of Whittaker et al. (2002). For clarity, it must be said that the interface

described here has no relationship to the caricatured interfaces described in scenarios above. The utopian and dystopian scenarios were designed to present an argument for not using ASR in the courtroom while the following scenario presents a vision for the use of ASR in the court’s “back room”. The interface described in this section is called the Interface for Court Audio Access (ICAA). ICAA would replace bench sheets or augment a greatly simplified version of the existing bench sheets, allowing the magistrates freedom from writing large amounts by hand while still allowing workers in the back room access to the information they require to perform their work.

The ICAA Scenario

This scenario goes into a lot more detail about the court and in particular about the work process after court though it still takes place during the A-list. As with the previous scenarios, the technology described is plausible, if not completely possible given the current state of the art.

This scenario introduces a new character to the Court –the After Court Officer, Julie.

Scenario extract

“In the matter of charge number HW39674, Henry Webb is hereby released on recognisance self in the amount of \$1000 on the condition that he be of good behaviour for twelve months.” Mr Cowley taps the screen again, ending the recording. The screen shows *recording finished*. Mr Cowley hands Mr Webb’s folder back to Claire and as it crosses the boundary from the bench to her desk the touch screen shows *next case*. At the same time, a small printer on Claire’s desk produces a docket with a ten-digit number and a few details relating to the case. She puts it in the folder and puts the folder on her “done” pile. Mr Webb’s day in court is over and he’s free to go. [...]

The defendants’ folders and the monitor’s master charge sheet make their way to the back room and become the responsibility of Julie. Julie takes the first folder, which belongs to a Mr Smith, from the big pile next to her desk, opens it and types the code on the docket at the top of the documents in the file into the ICAA.

After entering the code from the docket, the ICAA case window appears with the most recent transcript from Mr Smith’s trial already open in the transcript pane. If there were other transcripts from previous appearances, they’d be in the archive pane, but this is Mr Smith’s first time in court. By reading the transcript, Julie is able to assess what has happened in court and what decisions the magistrate has made. In this case, Mr Cowley has dismissed a bunch of charges and set aside hearing the remaining charges for a later date. Clearly this person has pleaded not guilty. The ICAA is really good at recognising charge numbers so Julie quickly scans the transcript to make sure that nothing is really wrong and tells the ICAA to tell the CMS to record that the charges were dismissed. All this takes is a few mouse clicks. [...]

The next folder is quite thick. Ms Barker has generated a lot of paperwork and has obviously been in court many

times. Since this is the A-list pile she has probably re-offended while on bail. Julie quickly types in the code number from the docket from the top of the folder. She sees that the system has not managed to make a very good transcription. Bad transcripts are always different and this one starts, "butler company on does enter..." all in black indicating that the system is very confident that this is exactly what was said in court. It's weird how sometimes the speech recognition can be confident about gibberish and not confident when the transcript makes perfect sense.

Scrolling down shows that the rest of the transcript is not much better. Selecting the first paratone in the transcript, Julie plays the audio, "But her companion doesn't..." - ah that explains it. The magistrate has woken up ICAA in the middle of speaking which always seems to confuse it. No matter as the audio is good, so Julie can listen to the judgment. This time it is an order to undergo counseling and drug rehabilitation at a facility 300km to the east. The system invariably gets the name of that facility wrong in a transcript anyway, so Julie resigns herself to the fact that she would have had to listen in even if the transcript was good. While she listens to the rest of the audio, Julie picks up the letters from the printer and files them appropriately, distributing them between Mr Smith's folder and her outbox. Switching her attention to the case management software, Julie checks that she is looking at the relevant case and charge (there's only one) and enters the information by hand. This requires more letters be printed. While the printer whirs away at these, Julie picks up the next folder.

Considering the ICAA Scenario

The ICAA Scenario presents one view of how ASR could be implemented at the Magistrates Court. It is not presented as definitive but instead as a tool for inspiring designers and those who would build future automatic speech recognition systems. What the ICAA scenario shows is that, unlike much work in the field of ASR, successful use depends on tight integration of the system with the work process of the organization and integration with technologies outside of what may be considered to be the bounds of ASR. By building on the work of Whittaker et al (1999; Whittaker, Hirschberg et al. 2002) the ICAA scenario shows the value of an errorful speech recognition application by leveraging the expertise of those using the transcripts of the system.

CONCLUSION

The interface described in the scenario above is not intended to be produced. Indeed, it is beyond the state-of-the-art by several years and would require a great deal more field work at the Magistrates Court to more fully understand the work, work process and procedures that would need to be embodied in such a system. Instead, by describing a system that might work in the Magistrates Court and showing how significantly such a system impacts on the work of many people in the court, this paper shows how non-trivial the introduction of an Automatic Speech Recognition system is, even when the

situation of proposed use seems, at first glance, to be ideally suited.

This paper has also demonstrated the application of a novel use for scenarios to package and communicate field work data in a way that is accessible to a wide range of stakeholders. By using a caricatured approach to stimulate discussion of the positives and negatives of making a change to a situation the "full blown consequences" of a change are revealed and more nuanced details can also be seen. These caricatured scenarios are also useful to elicit further details from stakeholders about their preconceptions of a particular technology, in this case automatic speech recognition.

Using a scenario to imagine a future automatic speech recognition application is particularly useful as it is difficult to represent what such an application can do without building it and imagining automatic speech recognition is difficult for humans who are so familiar with "normal" speech recognition.

ACKNOWLEDGEMENTS

This work was supported by a University of Canberra Research Grant. Thanks to the contribution of Prof Michael Wagner in this research project.

REFERENCES

- Bødker, S. (2000). "Scenarios in user-centred design - setting the stage for reflection and action." Interacting with Computers **13**: 61-75.
- Callon, M. (1986). Some elements of a sociology of translation: domestication of the scallopes and fishermen of St Briec Bay. Power, Action and Belief. J. Law, Routledge and Kegan Paul: 196-233.
- Kraal, B. (2006). Considering Design for Automatic Speech Recognition in Use. School of Information Sciences and Engineering. Canberra, University of Canberra.
- Kraal, B., P. Collings, et al. (2004). An Ethnography of Speech Recognition. Proceedings of OZCHI 2004. Wollongong, Australia.
- Latour, B. (1987). Science in action: how to follow scientists and engineers through society. Cambridge, Mass, Harvard University Press.
- Law, J. (2003). Traduction/trahision: Notes on ANT. **2005**.
- Whittaker, S., J. Hirschberg, et al. (2002). SCANMail: a voicemail interface that makes speech browsable, readable and searchable. Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves. Minneapolis, Minnesota, USA, ACM Press: 275-282.
- Whittaker, S., J. Hirschberg, et al. (1999). SCAN: designing and evaluating user interfaces to support retrieval from speech archives. Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. Berkeley, California, United States, ACM Press: 26-33.