Kraal, Ben J. (2009) ***Contrasting scenarios : embracing speech recognition***. In: Maurtua, Inaki (Ed) Human-Computer Interaction. INTECH, Austria, pp. 497-516.

1

Contrasting Scenarios: Embracing Speech Recognition

Ben Kraal
Queensland University of Technology
Australia

1. Introduction

The purpose of this chapter is to describe the use of caricatured contrasting scenarios (Bødker, 2000) and how they can be used to consider potential designs for disruptive technologies. The disruptive technology in this case is Automatic Speech Recognition (ASR) software in workplace settings. The particular workplace is the Magistrates Court of the Australian Capital Territory.

Caricatured contrasting scenarios are ideally suited to exploring how ASR might be implemented in a particular setting because they allow potential implementations to be "sketched" quickly and with little effort. This sketching of potential interactions and the emphasis of both positive and negative outcomes allows the benefits and pitfalls of design decisions to become apparent.

A brief description of the Court is given, describing the reasons for choosing the Court for this case study. The work of the Court is framed as taking place in two modes: Front of house, where the courtroom itself is, and backstage, where documents are processed and the business of the court is recorded and encoded into various systems.

Caricatured contrasting scenarios describing the introduction of ASR to the front of house are presented and then analysed. These scenarios show that the introduction of ASR to the court would be highly problematic.

The final section describes how ASR could be re-imagined in order to make it useful for the court. A final scenario is presented that describes how this re-imagined ASR could be integrated into both the front of house and backstage of the court in a way that could strengthen both processes.

1.1 Speech Recognition

Speech recognition, specifically, large vocabulary desktop-based speech recognition, or dictation software, hereafter referred to as automatic speech recognition (ASR), is a disruptive technology because it requires the adoption of new ways of working in order to

be useful. For example, as (Kraal, 2008) showed, speech recognition cannot be adopted without also making changes to the work that a person does. Some ways of working, and some kinds of work, are not compatible with the ways that speech recognition must be used.

Understanding how adopting speech recognition will change a particular situation is difficult because the implications are not just limited to one person or one interaction. Adopting speech recognition has wide-ranging implications not just for one person but potentially for whole organisations.

1.2 Field Work with Speech Recognition Users

In order to understand how a disruptive technology, in this instance ASR, affects how people work, field work was done with ASR users. The people studied all worked in the Federal Public Service in Canberra, the national capital of Australia. There were two groups of ASR users who were studied. One group of users had occupational over-use injuries, often called Repetitive Strain Injury (RSI), though RSI is a form of occupational over-use injury. The occupational over-use injury group were unable to type on a standard computer keyboard without pain.

The second group of users did not have an occupational over-use injury and worked for the Parliamentary Reporting Service in Parliament House. Their job was to create the document known as Hansard which is the record of what is said in Parliament's upper and lower houses and various committees. The Hansard group of users used ASR as a way to rapidly moved from audio recordings of what was said in the houses of parliament to a text representation. The Hansard users workflow did not involve feeding the audio record into a speech recognition system as that process is not yet accurate enough. Instead, the Hansard users listened to the audio recordings and re-spoke them. This was significantly more accurate which lead to significantly faster turnaround. Fast turnaround is important for the Hansard users as they have a 24 hour turnaround time on the Hansard document — what is said in parliament needs to be available the next day.

The field work research of speech recognition users showed that using speech recognition systems in a workplace required the user to maintain a delicate socio-technical system of software, hardware and office politics. If any one element of the socio-technical system should fail, the whole assemblage fails, resulting in their inability to use speech recognition for work. Often, users of speech recognition found themselves adapting their ways of working to suit the ASR system, rather than the system adapting to them.

2. The Magistrates Court

A team of researchers I was studying with were approached by the Chief Magistrate of the ACT Magistrates Court to investigate the introduction of ASR technology to the courtroom for use by him in the process of communicating an outcome.

In meetings with the Chief Magistrate we learned that communicating an outcome is a highly charged moment in the Court when the magistrate speaks an outcome for the case that he or she is hearing. An outcome may be a sentence, for example a fine or jail term, or it

may be the decision to set a case over until another time to allow all the parties to the case to gather relevant information. An outcome may also be a procedural decision specific to the Court such as a request by the magistrate for any number of specialised reports that are used to inform the actual sentence when it is finally delivered. Calling these varied events "outcomes" allowed us to define them as end points in a courtroom process. The Chief Magistrate would not have referred to these acts formally as outcomes, however this terminology is useful in the context of designing a system to support such actions.

At the time this study was conducted, the magistrate speaking his decision on an outcome aloud to the court. The decision also had to be written down on the "bench sheet" and incorporated into the "defendant's folder". The bench sheet was a piece of paper used by the sitting magistrate for note-taking during a hearing. The bench sheet was also the place where the magistrate would record his or her decision on an outcome. Because many decisions made by the magistrate were repetitive and procedural in nature, a set of large rubber stamps were available for the magistrate's use. The stamps were templates of boilerplate text allowing the magistrate to tick boxes or fill in only limited details in order to record an outcome on the bench sheet. The defendant's folder was a collection of bench sheets and associated documents passed to the magistrate by the prosecution and defence during the series of court appearances that typically make up any particular case.

The Chief Magistrate asked for an ASR system that could replace his existing manual system of handwriting and rubber stamps. 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 sentences on paper. His main reason for wanting an ASR system was so that he could save time. 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. A magistrate will often decide to waive many of the individual charges and sentence a defendant on a small selection of the total number. The waived charges still require a stamp and some writing and so still take up some of the magistrate's time that could otherwise be used to hear cases.

After some preliminary ethnographic 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. A great deal of "back room" work was initiated from one spoken outcome in the courtroom. This contrasted with the Chief Magistrate's view of the process as one that was enacted by him and contained within the courtroom (Dugdale & Kraal, 2006).

The courtroom and the "back room" of the Magistrates Court can be considered as front of house and backstage, much as in a theatre. The front of house refers the courtroom itself and the activity that takes place there. The backstage contains all the unseen workers who perform the mundane "bulk work" of the Court, processing documents and ensuring that the activities on the front of house can take place. The distinction between front of house and backstage is useful because it aligns with the aspects of the court that are more fixed and the aspects that are more able to be changed to suit ASR. Aspects of the Court that are

front of house are typically resistant to change and aspects that are backstage are less resistant.
The elements in the heterogeneous assemblage of the court that are resistant to change are:

- The social world of the Court;
- The Court room layout, as it influences the social world of the Court;
- The work process in the courtroom and all public-facing areas of the Court; and,
- The requirement to record decisions on outcomes made by the magistrate during court

The elements that are more easily able to be changed are:

- The detail of the work process of the "back room", particularly the after-court section; and,
- Details of the defendant's folder but not its use or existence.

These fixed front of house and less fixed backstage aspects of the Court are described in the following subsections.

2.1 Front of House

On entering the courtroom, it is apparent that the physical layout of the Court is something that cannot be significantly changed to accommodate the needs of ASR.

In the courtroom, the magistrate sits at the bench which is raised above the main floor level. The magistrate's associate sits to one end of the bench and below it. The lawyers sit facing the magistrate. The defendant sits next to his or her lawyer. The witness box is toward the other end of the bench with the witness facing the lawyers. A public gallery of varying size sits behind a barrier behind the lawyers. In the Magistrates Court, because of the relatively fast pace, there are often other defendants waiting in the court. Some of these waiting defendants are in the public gallery and some are queued up in front of the barrier, but still behind the lawyers, having been brought up from the lock-up by bailiffs.

The social world of the Court is necessarily bound up in the spatial layout of the participants. The magistrate's authority is symbolised by their elevated position at the bench. The interaction of associate and magistrate is enabled and constrained by the associate's position to the side and below the magistrate. To have a discussion with the associate the magistrate must slide his or her chair over to the associate's side of the bench. Being able to move along the length of the bench requires that the magistrate not be tethered by any cords to microphones that may be required for ASR. This is just one example of how the physical and social aspects of the courtroom have the potential to influence the design of an ASR system for the court.

Slightly less fixed in theory is the architecture of the courtroom. In practice, though, the architecture of the Court is fixed as it is expensive to make significant changes. Court Room One in the Magistrates Court of the Australian Capital Territory is a modern design with pale wood in place of the more traditional dark wood paneling. The ceiling is stepped and indented in various places which leads to areas in the public gallery with very poor

acoustics. The design of the Court does not seem to affect the ability of the active players in the Court's process to hear each other.

Various technologies are employed in the courtroom. Microphones are placed in front of the magistrate, the lawyers and the witness box, not to broadcast the speech of the players to the gallery but to allow it to be recorded. The microphones currently used are basically of "lectern" style with a small bud at the end of a semi-flexible stalk. The players in the court do not interact with the microphones due to their unobtrusive placement and because they do not hear any "foldbac (That is, they cannot hear what they sound like through the microphone, as one can when the amplified output of the microphone is broadcast to the room in which one is talking) from the microphones which might inform them that their speech is not being fully captured. The use of microphones in the Court is therefore not a part of the work in the Court nor is it part of the embodied social world of the court, leaving the use and placement of microphones open to change and re-enrollment in a new network that will use ASR.

In communicating a decision, the magistrate's speech must be captured and the lawyers', defendant's and witnesses' speech is of no importance. This means that a method of communicating sentence need not be concerned with the microphones in front of anyone save the magistrate. By extension, the lawyers, defendant and witnesses do not need to be enrolled in a new network.

Depending on the design of the ASR system proposed for the Court, magistrates and associates may have to be enrolled in the new actor-network. From my examination of the work process of the Court, and particularly the work done during court sessions, the associate has a lot of work to do and it would not be possible to add to the workload of the associate without taking some parts away. Because much of the associate's work in court has to do with helping the magistrate manage the work process of the court, it is not desirable for the associate to perform work extraneous to that management. Similarly, the magistrate is concerned, while court is in session, with managing the process of the court and imposing new work on the magistrate, particularly when that work could involve errorful ASR, is extremely undesirable.

As has already been said, the social world of the court is embodied, at least in part, in the spatial layout of the courtroom. This is also true of the work process of the Court as a whole if that work process is seen as the administration of cases that appear before the magistrate. Defendants enter and leave with their lawyers and only those called to appear at any one time can interact with the magistrate and the public prosecutor. Other defendants and lawyers must wait in the spaces assigned to them. Similarly, witnesses may only interact with the Court, the lawyers and the magistrate when called and must otherwise wait. The spaces for waiting are important to the Court's work process as a whole but do not need to be enrolled in a ASR system for communicating decisions.

2.2 Backstage

In backstage of the Court documents are paramount. The most important document for the Court is the Defendant’s Folder, which is really a collection of documents. The Defendant’s

Folder is made up of many documents. Reports from external agencies are used to inform decisions. Bench sheets with notes from a defendant's previous appearances remind magistrates about their previous decisions.

In much the same way as Air-Traffic Controllers' work is embodied in paper strips (Hughes, Randall, & Shapiro, 1992) the work of the Court is embodied in the defendant's folder. As the folder moves through time in the Court, it is used at various stations to reconstruct what occurred when the defendant it describes appeared before a magistrate. Staff in the back room use the folder as part of their work in recording magistrates' decisions in the Court's computer system. They also insert documents that they generate into the folder for future reference. Magistrate's associates use the folders to prepare for court sessions and will re-order the documents in the folders to assist the magistrate to work more efficiently during court. Magistrates refer to bench sheets and reports in a particular defendant's folder during court to reacquaint themselves with their previous decisions. Because the bench sheet and folder embodies the work of the Court it is quite obdurate in the existing work process, and therefore in the existing actor-network of the court. It would be very difficult to replace it or significantly change it to accommodate a ASR system.

A tension exists within the use of the defendant's folder. The Chief Magistrate wanted to do away with the time-consuming act of writing down decisions but, as is apparent from the descriptions above of the activity in the courtroom, writing down decisions is important for the future reference of the magistrate. Finding a solution to this tension is the main task of designing a ASR system for the Court. Doing away with the act of writing down decisions on outcomes may actually make the work of a future magistrate harder as there will be no record of the past.

The diverse actors who make up the parts of the Court's work process for communicating decisions that take place outside the courtroom will also need to be interested and enrolled in the new ASR actor-network. From a system design point-of-view, the easier it is to enroll the "back room" actors, the easier it will be to introduce the new system. The human actors in the back room are the monitor, the after-court person and the list clerk, all of whom have an interest in what the magistrate says and how decisions are communicated.

The monitor uses a computer system to annotate and mark-up the recording of what is said by everyone in court. Making the monitor part of the new design would be desirable because they already deal with the magistrate's speech. Enrolling the monitor's computer system and the audio feed it relies on may also be necessary.

The after-court person is someone who relies almost totally on the defendant's folder to perform their job. Any change to the folder changes the work of the after-court person because of their reliance on being able to reconstruct the decisions made in court. The work of the after-court person depends on the work of the Court being embodied in the defendant's folder.

2.3 Summary

In Actor-Network Theory, a “point of passage” (Callon, 1986) is an actor who directs, or tries to direct, the network to their interests. In the case of the new network at the ACT Magistrates Court, the "point of passage" whose interest is of importance to a technologist is ASR itself. Other actors in the new network may attempt to assert their power as points of passage but the power of the Chief Magistrate in the Court would seem to be so much greater than that held by the other individuals in the Court's work process that any objections by them would be ignored or cast aside by the Chief Magistrate who is seemingly in favour of the ASR system.

A "trial of strength" is when all the heterogeneous elements of an actor-network must perform their roles. Discussing the trial of strength of an imaginary system is largely moot, but by speculating about such an event points of weakness can be identified and later strengthened. The most obvious point of technical failure is the ASR system itself, not just within the recognition software but the wider system of vocabulary-models, acoustic models, microphones, cabling, user-interfaces and so on. Careful system design involving the users of the system in a user-centred process would lessen the severity of this point of failure, particularly if an iterative user-centred software development process was followed.

The non-technical points of possible failure are more difficult to "design out" of concern because they are less predictable. The best way to "design out" the non-technical points of possible failure is to "design them in" by respecting and valuing the existing work of the human actors in the system and using a new design to aid them in their skilled work. This "designing in" (ideally) takes place when the system is being designed and is necessarily a process of negotiation.

3. Caricatured Scenarios for the Magistrates Court

In order consider the implications of designing an ASR system for the Magisrates Court, I needed to create rough prototypes of the system, both for considering various implications of such a system and for communicating aspects of any potential future system with the Magistrate himself and the Magistrates Court organisation.

Creating even a simple speech recognition system requires creating software which is “heavyweight”. Instead, scenarios (Carroll, 1995) were used as a lightweight way to consider different potential implementations of ASR for the Magistrates Court. Technical constraints are incorporated into the scenarios by basing them on pre-existing technical research.

The scenarios used for the description of potential implementations are inspired by those described by Bødker (2000) who suggested paired positive and negative scenarios that were caricatures of action. The positive and negative scenarios are caricatures because they are “over the top”. The positive scenarios are wildly optimistic, the negative scenarios pessimistic. Instead of a single scenario that describes imagined future action, these paired scenarios allow description of action as well as showing the imagined positive and negative effects of the design upon the situation. These paired scenarios stimulate ideas and make

clear the potentials and problems of ASR in the Court (Kraal, et al., 2006). They serve as "sketches" of potential implementations of a user interface that has no physical representation.

ASR presents unique challenges when producing prototype designs, particularly early prototypes that are intended to stimulate further design thinking rather than be seen as design proposals. A graphical user interface (GUI) can be prototyped using various low-fidelity methods, even to the extent of prototyping in paper (Snyder, 2003), however in ASR interfaces, much of the interaction is invisible and impermanent and therefore difficult to represent on paper using methods derived from GUI design. Other aspects of the speech user interface, for example vocabulary and grammar, are too detailed to specify at the early prototyping stage.

The detailed, caricatured, scenarios that follow describe the use of an ASR system for the Magistrates Court. The scenarios describe a potential implementation that is entirely front of house and is a near-direct replacement for the magistrate writing outcome decisions on a bench sheet. Technical constraints are incorporated into the scenarios by considering pre-existing technical data and fieldwork with users of contemporary ASR software (Kraal, 2008).

3.1 Utopian Scenario

It's 9.30am on a Tuesday as Rob, Chief Magistrate of the ACT, enters the courtroom. He sits down at the bench and court begins. On the bench are several objects: Rob's favourite coffee-mug, a carafe of water and a glass, a few pens, an array of tiny microphones embedded into the small shelf above the surface of the bench and a touch-screen that's about as big as a hand-held computer game. The microphones work together, canceling noises from the Court and capturing Rob's speech when necessary and the touch-screen allows Rob to trigger various modes and actions of the ASR system.

The first few cases that appear are dealt with very perfunctorily and are all set over to another date. Rob does this in concert with the List Clerk who advises him when the next available dates are for the particular sort of cases that appear. Rob's Associate, Claire, organises the cases in this way as it suits Rob's way of working. Once Rob and the List Clerk have found a suitable date, Rob uses the touch-screen to trigger a recognition event that allows him to speak the date for the next part of the case to the Court. Speaking the outcome records it.

The next cases involve people who have been in the lock-up overnight. Rob usually makes a judgment on these cases, often just a bail arrangement but if someone pleads guilty he will sentence them on their first appearance if the sentence is simple and not severe.

The first difficult appearance today is a Mr Tailor who was in a street brawl last night and has been in the lock-up since about 2am. The public prosecutor hands Claire a police report on the incident that Claire hands to Rob for him to read. Mr Tailor's lawyer says that the fight was uncharacteristic and that Mr Tailor is a member of society in good standing who has been employed as a carpenter since he left school at 16. Rob says that the report

indicates that Mr Tailor hit three people, including a woman, and that he swore at a police officer. Rob says that these are fairly serious charges and that he will have to sentence Mr Tailor.

Mr Tailor's lawyer and Rob have an exchange that results in Rob postponing sentencing to a date in three week's time. To make this decision official, Rob touches a button on a small touch-screen mounted on the bench. The button is labeled speak decision. The button changes colour from grey to green, showing Rob that the system is ready. Rob says, "Decision in case 54897," and then says the words of the bail agreement, "the defendant is released on bail, recognisance self in the amount of $1000 to reappear three weeks hence" . An indicator next to the button turns yellow and then green, indicating that the decision has been recognised. Rob taps another button labeled print decision. A small laser printer in the bench produces a piece of paper with the decision printed on it. Rob checks that he is happy with the wording, signs it and places it in the bench sheet folder. He taps the next button in the touch-screen, confirm decision. Next to Claire, a laser printer comes to life and produces three identical pages. Claire hands one to each lawyer and one to Mr Tailor. These pages contain the text of the decision and the date of Mr Tailor's next court date. Pressing the confirm decision button has also added the decision to the Court's computer system. The touch screen goes back to its initial state, ready for the next case, as Claire calls for the next defendant.

3.2 Dystopian Scenario

It's 9.32am on a Monday as Rob, Chief Magistrate of the ACT, enters the courtroom. He sits down at the bench and court begins. As Claire, Rob's Associate is calling the first case, Rob plugs himself in to the speech recognition system. A lapel microphone is sewn into the black gown that Rob wears and it needs to be connected to the system.

The first case today is a Mr Jones who caused a car accident last night while he was drunk and has been in the lock-up since about 2am. Mr Jones is pleading guilty on all charges. The public prosecutor hands Claire a police report on the incident. Claire hands the report to Rob. Mr Jones's lawyer says that the drunkenness and accident were uncharacteristic and that Mr Jones is normally home looking after his four children by 9pm. Last night Mr Jones had attended a party at a local club and made a mistake in driving home intoxicated. Rob says that the report indicates that Mr Jones hit two cars and resisted arrest and that these are fairly serious charges, so he will have to sentence Mr Jones.

The defence counsel assents to Rob passing sentence immediately. To make the sentence official, Rob touches a button an a small touch-screen mounted on the bench. The button is labelled speak decision. Nothing happens. Rob taps the touch-screen again and this time it changes colour from grey to green, indicating that the system is ready. Rob says, "Sentence in case 86572," and then says the words of the sentence, "the defendant is found guilty on all charges and is sentenced to three months imprisonment to be suspended forthwith and is released on a good behaviour bond of $1000" . An indicator next to the button turns yellow… and stays yellow, indicating that the decision parser has not been able to correctly determine the sentence. This usually means that the recognition engine has misrecognised a word so that the spoken sentence is not in a form that makes legal sense. Rob hates

repeating sentences when the system gets them wrong because he thinks it makes him look foolish. Rob taps the yellow speak decision button again and repeats the sentence. Just as he's finishing, someone in court sneezes! At least half the time, a sneeze or cough from the gallery will ruin the speech recognition of the decision. This time, though, the button turns green so Rob taps the print decision button.

A small laser printer in the bench produces a piece of paper with the decision printed on it. Rob checks the wording, but the system has misrecognised the length of the sentence and the amount of the bond. Why the decision parser can't check these things, Rob doesn't know. He supposes that different amounts are equally legal, even if they are wrong in this instance. It's often the case that when Rob gets a yellow from the speak decision button that the system has also got something else wrong. Rob slides his chair closer to Claire's desk to ask her to try to fix what's gone wrong but he feels the microphone cord tension as he reaches its full length, still not quite close enough to have a quiet word with Claire. So instead he glances down at Claire and lifts his eyebrows significantly. Claire taps a few keys, giving her access to the transcript of what Rob's just said, and begins editing the transcript. The system allows Claire to edit the transcript of the spoken sentence only after it's been parsed correctly.

When Claire's done she nods at Rob and he taps the print decision button again. The decision comes out of the printer and Rob signs it and places it in the bench sheet folder. He taps the next button in the touch-screen labelled confirm decision. Next to Claire, a laser printer comes to life but produces no output. Claire leans over it and sighs. Paper jam. She flips covers and latches and pulls out a mangled piece of paper. She gives Rob a small nod again and he taps the confirm decision button. This time the printer produces three identical pages. Identically faulty. The toner cartridge in the laser printer has run out.

Claire whispers to Rob that they have a problem and Rob says to the court at large, "let's have a ten minute recess while we get someone up here to deal with some small problems we're having". Most people in the court sigh — it's clearly going to be a long day.

3.3 Summary

The scenarios above show how the same technology, implemented in basically the same way, can have radically different outcomes in use. In the utopian scenario, everything is perfect, the interaction is virtually seamlessly integrated into the business of the court. In the dystopian scenario everything breaks down, including the magistrate's sense of control and prestige in the court.

Contrasting the scenarios shows that the introduction of ASR to the court in a near-direct replacement for the magistrate writing on a bench sheet 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 utopian 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 Rob from having a private word with Claire, reducing him to facial gestures.

Aspects of the use of ASR in the court are also problematic because of the properties of the court itself. These properties are related to the established work process of the court, the physical arrangement of the space, how the required technologies relate (or do not relate) to one another and so on.

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.

4. Re-imagining Speech Recognition for the Magistrates Court

To use ASR in the Magistrates Court necessitates that ASR, as a technology, be re-imagined. Often, ASR applications are seen as a way to replace the act of typing by one user, that is, 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.

One possible re-imagined form of ASR that might work for the Court is a model where the users and computers are distributed in space and time, just 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, the magistrate, is not necessarily the person working with the transcript generated by the ASR system, a back room worker. Distributing the computers involved allows separation of work tasks and recognition tasks as well as allowing multi-pass ASR (Whittaker et al., 2002) which can improve the accuracy of hard-to-recognise speech by allowing a recognition engine to refine a transcription.

As stated previoisly, 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; and,
- Details of the defendant's folder but not its use or existence.

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. Analysing the work of the Court has shown that these elements are the most flexible to change.

Re-imagined ASR for the Court incorporates a model of the Court's workflow. In the existing work process the magistrate speaks a decision and writes it down and other people perform the coding of that decision into something that allows the machinery of the Court to keep working. A dictation paradigm of ASR can't perform that task because the translation

of the magistrate's decision into codes is too nuanced, too detailed, too specialized and too reliant on intelligence and experience. The goal of a re-imagined ASR is to make it easier for the back-room workers to do the parts of their job where intelligence is required and minimise the parts of their jobs that are repetitive.

The Chief Magistrate's request was that any new system remove the need for him to write down every decision. It is quite simple, technically, to record the speech of the Chief Magistrate when he makes a decision. In fact, it is done already and annotated by the monitor. However, using those recordings as a resource to replace or augment the bench sheet is considerably more problematic. Adding to the possible problems of moving to a speech record are the stamps which are currently used and which may act as a prompt to the magistrate would no longer be available. Minimising the amount of writing on a bench sheet by a magistrate requires that the information previously contained on the bench sheet is available elsewhere. Given that the magistrate's speech is recorded it is reasonable to attempt to provide access to that recorded speech.

The problem with accessing speech recordings for the Court, and the back room in particular, is that using the bench sheets and the entire folder is a fact finding exercise where users scan and browse the documents in the folder looking for the specific information that the case-management software requires. Replacing the bench sheet, which at least in part embodies the process of communicating decisions, with an ASR system is a similar problem to that faced by the designers of the air traffic control system described by Hughes, et al. (1992) where a design for replacing paper that was an embodiment of work was proposed. As with the air traffic control work by Hughes et al. (1992) the proposed design for an ASR system for the Court attempts to retain the communicative aspects of the embodied work while introducing new possibilities for interaction to the process.

Using a speech record instead of paper to communicate decisions necessitates scanning and browsing a recording of the magistrate's speech in court. Scanning and browsing a recording of speech is time consuming because speech is one-dimensional and ephemeral. One way of making speech persistent and two-dimensional to support scanning and browsing behaviours is to turn speech into text. A group at AT&T Research looked at voicemail and speech in general and developed the Spoken Content-based Audio Navigation (SCAN) interface. SCAN was developed to be used as a way to access transcripts of broadcast news (Whittaker et al., 1999) and later voicemail (Whittaker et al., 2002). The AT&T researchers had no illusions about the errorful nature of ASR, however they showed that the errorful transcripts allowed users to obtain an overview of audio recordings that was previously impossible. The errorful transcripts turned unusable speech recordings into something useful.

It is important to note that the interface described in the scenario below is not a proposed solution, but is a way of exploring and building on the ideas presented until this point in this thesis. A solution would need extensive testing with proposed users and would have to undergo several iterations before it would be ready to be used "live". The design for the interface is speculative and arises from the fieldwork described elsewhere in this thesis. Presenting this design here is not an attempt to say "here is an ideal design for ASR in the

Court" but rather a way to show how the fieldwork and contrasting caricatured scenarios have led to a potential outcome.

The interface described in the next section has some similarities and some differences with the SCAN interface (Whittaker et al., 2002; Whittaker et al., 1999). SCAN was the implementation of a new paradigm in accessing speech records, What You See Is (Almost) What You Hear or WYSIAWYH. The primary goal of WYSIAWYH was to present a visual analogue to recordings of speech. SCAN used transcripts of speech generated by ASR software to facilitate the visual analogue. To create the transcripts the speech was broken into "paratones" and then passed through an ASR engine several times, allowing the recogniser to improve on the transcript. The results of the transcription for each paratone was then combined into the errorful transcript of the particular audio recording. The terms in the transcripts were then indexed for later retrieval. Users could enter natural language queries into the SCAN interface and the system would return ranked transcripts that the user could select to view and, if required, listen.

SCAN had an "overview" feature that displayed the incidence of keywords in the paratones of the transcript and the transcript itself. By providing a visual overview of the keywords in various paratones, SCAN allowed the user to skim the document more quickly than if they had to scan the entire transcript, which could be the textual representation of 25 minutes of speech. After using the overview section to jump to a potentially relevant paratone, the user could read the (errorful) transcript. If the transcript contained too many errors to be sensible the user could click the paragraph to play the audio it represented.

The SCAN interface was empirically tested and found to be more effective than just listening to the recordings in fact-finding tasks. Additionally, subjects in the experiments found the SCAN interface easier to use than just listening to the audio and listened to a shorter amount of audio to complete the tasks set them. The researchers found that increases in transcript accuracy had an influence on the perception of the difficulty of the task and on the actual quality of the answers the subjects gave. The mean accuracy of the transcripts in the tests was 67% with a maximum of 88% and a minimum of 35%. SCAN was found to be particularly useful for fact-finding tasks using the broadcast news corpus.

The researchers noted that there are disadvantages to the SCAN approach. The chief disadvantage is over-reliance by users on transcripts. Because the transcripts are inherently errorful relying on them can introduce errors into information extracted from the transcripts. They suggest introducing a representation of the ASR confidence measure, that is word probability, to the transcript. Words that the system was more confident about could be presented in darker type and words that had a lower confidence could be presented in a lighter type allowing the user to judge for themselves the accuracy of the text. Some users in the studies of the SCANMail interface suggested that the transcripts could be editable to allow for correction of errors should they be found. (Whittaker & Amento, 2004) built and tested an editable version of the SCANMail interface and found it to be usable and useful.

For clarity, it must be said that the interface described here has no relationship to the caricatured interfaces described in previous sections. Where caricatured scenarios are useful for sketching ideas and making clear benefits and traps of an implementation, a single scenario is still best for communicating a final solution.

I will refer to the interface described in this section as the Interface for Court Audio Access (ICAA). The main difference between SCAN and ICAA is that SCAN was intended to provide open-ended search capabilities over a large corpus of speech, either broadcast news (Whittaker et al., 1999) or voicemail (Whittaker et al., 2002) where ICAA would not require the ability to search over all speech recorded by the system but would instead be directed at searches of a single transcript or group of transcripts relevant to a particular case. 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.

4.1 The ICAA Scenario

The Interface for Court Audio Access (ICAA) scenario is partly inspired by the work of (Kraal, et al. (2002). In keeping with previous work that has used ethnographic methods (Hughes et al., 1992) and scenarios (Satchell, 2003) to describe future designs, the technical details of the scenario that follows are not described. The purpose of the scenario is to describe the system in use. The system does not exist – the scenario is an example only. The scenario has two parts, front of house (section 4.1.1) and back of house (section 4.1.2).

4.1.1 Front of House

It's Monday morning, always the busiest time for the A-list with all of the weekend arrests to deal with, and Court has just resumed at 11.07am, Magistrate Rob Cowley presiding. They're up to the drink-driving charges. First up, Henry Webb, representing himself. Claire hands up Mr Webb's folder. As it crosses the boundary from Claire's desk to the Bench, the touch-screen on the bench shows the charge numbers for the case in the folder---Mr Webb's driving under the influence charge---there's only one number. Mr Webb pleads guilty but states that this is his first charge for driving under the influence in 38 years of driving and indeed his first criminal charge ever.

Rob asks the public prosecutor what Mr Webb's blood-alcohol content was. "Zero point zero six, your worship". Barely over the legal limit and fairly obviously a lapse of judgment on Mr Webb's part. Rob notes it down on a blank sheet of paper in the folder in front of him. He's obviously contrite and just appearing in court seems to have scared him so much he'll be catching cabs from now on. Rob decides to give Mr Webb a good behaviour bond and a stern lecture.

Use of ICAA begins in the courtroom when no actual "interface" is visible. ICAA's intrusion into the courtroom itself is limited to a microphone, a few RFID sensors, a small touch screen on the bench and a small, fast printer on the associate's desk.

As court progresses, ICAA makes no intrusion into proceedings until a case comes to a point where the magistrate would previously have written a decision on the bench sheet.

The microphone and touch-screen are directly related to ASR. The magistrate uses the touch-screen as a way to start and stop the speech recognition when he's speaking a decision.

The printer on the associate's desk produces dockets that show a decision, or series of decisions, have been made relating to the case at hand. The RFID sensors sense RFID tags embedded in the folder. As the folder is passed between associate and magistrate sensors in the bench record the passing, allowing the system to dip into a database for pertinent information, for example charge numbers and various details relating to the defendant, if known, such as address and employer. The touch screen can then display these details.

Stern lecture over, it's time to sentence Mr Webb to good behaviour. Rob taps the touch-screen to start the decision-recording process. The gesture is so subtle that no-one in court really notices it. The screen shows “ready for decision” and still shows the charge numbers.

An audio recording has been going on since Rob sat down and court began. When Rob taps the screen to tell it he is about to speak a decision, the system tags the recording, allowing a future listener, or the ASR system, to jump to the sentence.

"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."

Rob taps the screen again, ending the recording. The screen shows “recording finished”. Rob 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.

So far, most aspects of the court's work process are much the same as they are currently. Handwritten decisions have been done away with, as was the purpose of this design, and replaced with what is from the magistrate and associate's perspective technology that is unobtrusive. The technology introduced into the court is strong and simple, in keeping with the findings that the ASR system introduced to the Court and does not significantly disrupt the work process in the courtroom or impinge on the theatre of the court.

While Mr Webb has been getting his lecture, and indeed since court has started, Molly has been in the monitor's booth watching and listening to everything. Molly has a computer in front of her with special software that can annotate the audio recording of what's going on in court. Since this is the A-list, Molly's job is just to record which lawyers are appearing when. Molly also has a paper master charge sheet listing every charge that's appearing in

court today. She uses the charge sheet to record which charge numbers are dismissed and which charge numbers the magistrate decides to deal with. In theory, with the ASR system in place, the monitor's job is unnecessary, however, ICAA keeps the monitor's job and makes the annotations work as part of the ASR system.

The monitor still uses the paper charge sheet to cross off dismissed charges so that the ASR system has a backup in case something goes very wrong. The charge sheet also helps the person doing the after-court processing, the process of which is described later in this scenario.

4.1.2 Backstage

The defendants' folders and the monitor's master charge sheet make their way to the back room and become the responsibility of Julie. Julie works in the after court section, processing folders from the day in court and entering details of the magistrates' decisions into the Court's case management software. The ICAA and the case management software (CMS) work together to help Julie do her job.

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. This works much better than the way things were about a month ago when they installed sensors in Julie's desk to automatically detect which folder Julie had selected. The sensors worked fine but they meant that Julie couldn't place the folders on her desk the way that she used to. Julie had the I.T. guys remove the sensors – she's happy to type a number if it means she can put the folder she's working on wherever she likes.

The folders and the RFID sensors work in the courtroom because there is a very clear demarcation between the bench, where the folders are "in play" and the associate's desk where the folders are "waiting". On Julie's desk the distance between the "in play" area and the "waiting" area is too fine and too variable for the sensors to work reliably.

Julie had the ICAA thrust upon her by the court. The ICAA is intended to help the magistrates in court and allow them to not write down decisions on outcomes but it still needs to communicate those outcomes to the people who need to know them. Julie's uses the ICAA because it is her job to process what is said in court. The ICAA is designed to make it as easy as possible for Julie to work with the audio and the transcript that is generated from it using ASR. Julie's previous experience with her job before the ICAA was introduced allows her to work with the transcript. Finally, Julie understands that the transcript is pretty close to the old handwritten bench sheet. She could either fight the new system or see it as a different skill to learn.

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 scanning 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 spoken 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 ICAA is so good at recognising charge numbers because the touch-screen shows the magistrate the charge numbers for the case at hand. This serves two purposes. It prompts the magistrate so that the charge numbers are easy to view and it primes the ASR software so that when it "hears" a charge number it will only recognise it if the number is from the list of charges in the case at hand. The RFID tags in the folder allow the ASR system to narrow the possibilities of what charge numbers the magistrate could say, leading to better recognition accuracy.

After taking care of the dismissed charges, Julie is able to get the longer part of Mr Cowley's decision where the case is set over for a date in three weeks time. The system has jumped through the transcript to the next part of the decision. Mr Cowley said that he'll hear the case on the 23rd of this month. The system understood that really well as it's in black text. He gave a few other orders that the system isn't that confident it's understood---they're in varying shades of gray---though they make enough sense as Julie reads through the transcript.

Using different shades to display the confidence of a recognised word or phrase has been used successfully in other transcript-based interfaces to underlying audio (Whittaker & Amento, 2004).

Julie is able to select the part of the transcript that has the date in it and drag it to the field in the CMS that accepts dates. The ICAA knows enough about spoken dates to convert the spoken "23rd of January, 2006" to 23/01/06. Julie makes sure the conversion is correct. Now she switches her attention to the CMS pane and fills in the rest of the required information. Mr Cowley has neglected to say which charges he'll be hearing on the 23rd, which isn't a problem in court as it's fairly obvious when he's dismissed a lot of charges, but the CMS needs to know exactly which ones he'll be hearing. The CMS assumes that unless charges are dismissed they're still current, so Julie confirms that with the CMS and checks quickly with the master charge sheet from the monitor. Before this folder is done, Julie has to print the CMS's summary of the outcomes so far and some letters to send to the various parties involved in the case. These letters are just proforma and are generated by the CMS. A letter for the public prosecutor's office; one for Mr Smith; one for Mr Smith's lawyer. They're printed in duplicate; one copy for the folder and one copy for Julie's outbox. While the printer takes its time, Julie pulls out the next folder, Ms Barker.

4.1.3 Summary

The front of house scenario above shows how ASR can be incorporated into the work process of the Court without disrupting the theatre of the courtroom. The back stage scenario shows how ASR could work with the back room workers, relieving them of some of the more repetitive aspects of their work and leveraging their expertise in interpreting and working with magistrates decisions.

The caricatured scenarios contributed to these final scenarios by making clear where the weak points in ASR for the court lay. These new scenarios aim to overcome those weak points by working with the strengths of each of the actors, human and non-human, in the entire Court work process and by also working with the strengths of ASR to produce a fast, if errorful, good enough transcript of a magistrates spoken decision.

5. Conclusion

This chapter has shown how caricatured scenarios (Bødker, 2000) can be used to "sketch" interactions that have no real physical or graphical manifestation by examining a potential use of speech recognition software (ASR) in the Magistrates Court of the Australian Capital Territory. The caricatures scenarios are used to show the benefits and limitations of potential implementations of ASR, allowing the potentials and pitfalls to be readily apparent. Of course, caricatured scenarios are of use in areas other than ASR and could be of great use in fields where it is difficult to create a manifestation of an interface, for example ubiquitous computing. Caricatured scenarios could also be used to "sketch" interactions where a combination of products mediate interaction with a service, for example in the way that the iPod and iTunes desktop software work together as an interface to the iTunes store.

This chapter has also shown how ASR can be re-considered as a productive technology that has benefits and limitations. Methods to mitigate the limitations are presented which can be of benefit to ASR designers. Productive use of ASR requires bringing together many aspects of a situation including technology, work process, spatial layout and acoustic considerations. Using caricatured scenarios allows initial ideas for ASR systems to be tested conceptually before committing to implementation and can also be used to direct further fieldwork which can be used to obtain a deeper understanding of a particular situation.

Caricatured scenarios are a tool that designers and researchers can use to explore the use of disruptive technologies and communicate the implications of introducing disruptive technologies to existing work practices.

6. 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 the fishermen of St Brieuc Bay. In J. Law (Ed.), Power, Action and Belief (pp. 196--233). London, Boston and Henley: Routledge & Kegan Paul.

Carroll, J. M. (Ed.). (1995). Scenario-based design. Envisioning work and technology in System Development. New York: Wiley.

Dugdale, A., & Kraal, B. (2006). Magistrates and voice recognition: reconceptualising agency. Paper presented at the OZCHI 2006, Sydney, Australia.

Hughes, J., Randall, D., & Shapiro, D. (1992). Designing with Ethnography: Making work visible. Interacting with Computers, 5(2).

Hughes, J. A., Randall, D., & Shapiro, D. (1992). Faltering from ethnography to design. Paper presented at the CSCW '92: Proceedings of the 1992 ACM conference on Computer-supported cooperative work.

Kraal, B. (2008). Considering use and design of speech recognition systems : investigating users of complex socio-technical systems. : Verlag, VDM.

Kraal, B., Dugdale, A., & Collings, P. (2006). Scenarios for Embracing Errorful Automatic Speech Recognition. Paper presented at the OZCHI 2006, Sydney, Australia.

Kraal, B., Wagner, M., & Collings, P. (2002, November). Improving the design of dictation software. Paper presented at the Australian Speech Science and Technology Conference, University of Melbourne, Victoria, Australia.

Read, J. C., MacFarlane, S., & Casey, C. (2002). Oops! Silly me! Errors in a handwriting recognition-based text entry interface for children. Paper presented at the NordiCHI '02: Proceedings of the second Nordic conference on Human-computer interaction, New York, NY, USA.

Satchell, C. (2003, November 26-28). The Swarm: Facilitating fluidity and control in young people's use of mobile phones. Paper presented at the OZCHI 2003, Brisbane, Australia.

Snape, L., Casey, C., MacFarlane, S. J., & Robertson, L. (1997). Using Speech in Multimedia Applications. Paper presented at the TCS Conference, Bangor, Wales.

Snyder, C. (2003). Paper prototyping : the fast and easy way to design and refine user interfaces. San Francisco, Calif.: Morgan Kaufmann.

Whittaker, S., & Amento, B. (2004). Semantic speech editing. Paper presented at the CHI 2004: SIGCHI conference on Human factors in computing systems, New York, NY, USA.

Whittaker, S., Hirschberg, J., Amento, B., Stark, L., Bacchiani, M., Isenhour, P., et al. (2002). SCANMail: a voicemail interface that makes speech browsable, readable and searchable. Paper presented at the CHI '02: SIGCHI conference on Human factors in computing systems.

Whittaker, S., Hirschberg, J., Choi, J., Hindle, D., Pereira, F., & Singhal, A. (1999). SCAN: designing and evaluating user interfaces to support retrieval from speech archives. Paper presented at the SIGIR '99: 22nd annual international ACM SIGIR conference on Research and development in information retrieval.