



COVER SHEET

This is the author version of article published as:

Kraal, Ben J. (2006) Embracing errorfulness speech recognition for the ACT Magistrates Court. In Gomez, Rafael E. and Gaddum, Nicholas, Eds. *Proceedings Faculty of Built Environment and Engineering Design Theme Conference 2006*, Gardens Point Campus, QUT, Brisbane Australia.

Copyright 2006 The Author

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

Embracing Errorfulness Speech Recognition for the ACT Magistrates Court

Ben Kraal
School of Design, QUT
b.kraal@qut.edu.au

Introduction and Background

A group that I worked with while completing my PhD was approached by the Chief Magistrate of the ACT 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. The process of communicating outcomes is a highly charged moment in the Court when the magistrate speaks an outcome for the case that he or she is hearing. Each case may have more than one outcome. An outcome may be a sentence, for example a fine or jail term, or it may be the decision to set a case over to allow all the parties to the case more time 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. The magistrate's speech act changes the world. It determines whether a defendant can leave the courtroom or is returned to the cells.

The Chief Magistrate asked for an ASR system that could replace his existing manual system of handwriting and rubber stamps. When the time comes to speak an outcome, a magistrate has the option of using a one or a combination of large rubber stamps (see Figure 1, Figure 2 and Figure 3) 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 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.

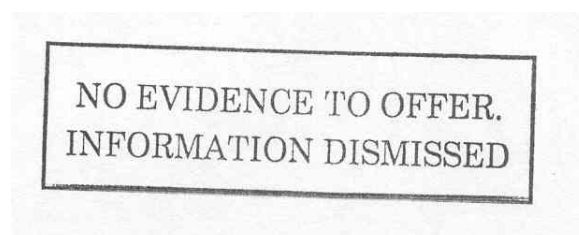


Figure 1: A magistrate's stamp dismissing a charge.

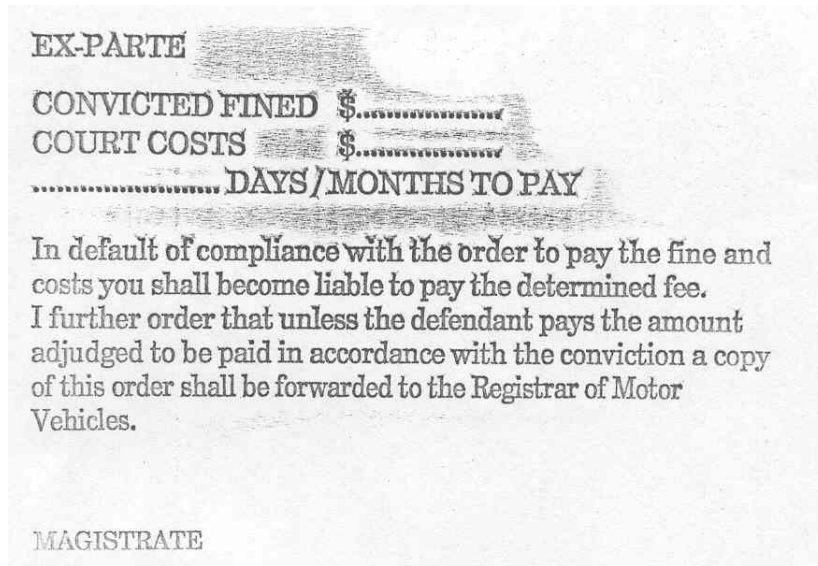


Figure 2: A magistrates stamp imposing a fine.

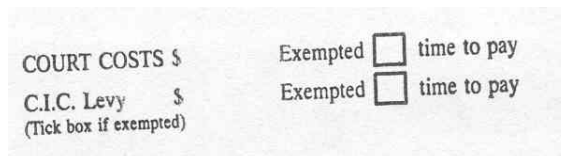


Figure 3: A magistrates stamp imposing court costs and CIC levy and other details.

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.

In the next sections, I describe the “fantasy” scenario that prompted the Chief Magistrate to contact my group regarding ASR. While there was never a time when he explicitly said to me “this is my dream for speech recognition” it became apparent to me that the following scenario is very much what he had in mind. Following the fantasy scenario is a worst-case scenario that shows how the same basic implementation could be disruptive to the Court. These scenarios are not indicative of my design for ASR for the Court. Instead they show a positive and negative view of ASR from the magistrate’s perspective to illustrate the demands of a future application (Bødker 2000). These scenarios are not drawn from real examples but are constructed. The scenarios act as “means to hold on to situations and how they may be changed because of a design” (Bødker 2000). In both the positive and negative case, the scenarios are extreme, very good and very bad, to show the “full-blown consequences” of an ASR system.

The Techno-utopia 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 Taylor 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 Taylor's lawyer says that the fight was uncharacteristic and that Mr Taylor 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 Taylor 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 Taylor.

Mr Taylor'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 (see figure 7.1) mounted on the bench. The button is labelled 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"². An indicator next to the button turns yellow and then green, indicating that the decision has been recognised. Rob taps another button labelled 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, labelled, 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 Taylor. These pages contain the text of the decision and the date of Mr Taylor'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.

The Dysfunctional Dystopia 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 that Claire hands to Rob for him to read. 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 a small touch-screen mounted on the bench (see figure 7.1). 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 which is not a good way for a magistrate to look. 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 when 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 and... nothing.

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.

Analysing the Scenarios

The scenarios above show how the same technology, implemented in basically the same way, can have radically different outcomes in use. In the techno-utopia scenario, everything is perfect, the interaction is virtually seamlessly integrated into the business of the court. In the dystopic 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 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 dystopic 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.

Considering ASR for the Court

Using ASR productively in the ACT 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 I use the language of Actor-Network Theory (Callon 1986; Latour 1987; Law 2003) to describe the difficulties involved in introducing ASR to the Court.

The court and the process of communicating outcomes is a largely stable 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 I have previously argued that ASR is flawed, suggesting, as I do here, that an ASR system become indispensable to the Court does not sit easily with me. 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, I have reconciled myself with these apparently contradictory view points.

Using an automatic ASR system at the ACT 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 interested and 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 aspects of the Court will have to change in order to preserve the main elements of the Court. How these translations will begin to be achieved is described in the next section.

To use ASR in the ACT Magistrates Court necessitates that ASR, as a technology, be re-imagined. Often, ASR applications are seen as being a replacement for typing to be used 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.

The re-imagined form of ASR that would 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; 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 and that is where the design work is concentrated. 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 interface described in the next section has some similarities and some differences with the SCAN interface (Whittaker, Hirschberg et al. 1999 ; Whittaker, Hirschberg et al. 2002). The first use of SCAN was as an interface to archived broadcast news recordings and was intended to solve what the researchers termed the under-utilisation of growing archives of speech collected from radio programmes, Congressional Debates and private archives of audio conferences. 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 were 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 next section describes an ASR interface for the Court inspired by the work of Whittaker et al. For clarity, it must be said that the interface described here has no relationship to the caricatured interfaces described in scenarios above. 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, Hirschberg et al. 1999) or voicemail (Whittaker, Hirschberg 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.

The ICAA Scenario

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 better judgment in the future, won't you." "Yes, your worship." 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.

"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.

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.

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.

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 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.

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.

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 that the CMS wants dates in a YYYYMMDD format and can convert "23 January" into 20060123 on the fly. 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.

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. 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 counselling 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.

Two down. So many more to go.

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. 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.

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: domesticaion of the scallopes and fishermen of St Brieuc Bay. Power, Action and Belief. J. Law, Routledge and Kegan Paul: 196-233.
- 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
<http://doi.acm.org/10.1145/503376.503426> 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
<http://doi.acm.org/10.1145/312624.312639> Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval Berkeley, California, United States ACM Press: 26-33