‘Intelligent Listener’: Experimentation in a Live Headquarters

Dr. Dominique Estival[a], Dr. Marilyn Cross [b] and Dr. Ahmad Hashemi-Sakhtsari[a]
[a] Human Systems Integration Group
[b] HQAST Group
Command and Control Division, Defence Science & Technology Organisation
PO Box 1500, Edinburgh SA 5111, AUSTRALIA

ABSTRACT

In this paper we report on a feasibility experiment we conducted during EX TENDI WALK at HQAST. The purpose of the experiment was to assess the potential use of current speech recognition technology in the context of operation planning and to identify the limitations that will need to be addressed before such technology can be installed and used operationally, e.g. in the new Collocated Headquarters. The technical limitations of the current set-up did not prevent us from collecting data which can now be used for an analysis of the interactions during planning sessions. The experiment also helped us ascertain that users are willing to use the technology, even with the relative inconvenience of having to wear a headset, and that they are already convinced of the benefits of automatically recording planning sessions: auditing and information retrieval, re-use of deliberate and immediate planning products, induction and training for incoming staff. We have identified improvements to the system set up which can be immediately implemented and we have narrowed the areas for longer term research.

1. INTRODUCTION

Rarely do scientists have the opportunity to experiment in a live headquarters. The many reasons precluding experimentation include a high operational tempo that means staff are concentrating on the current operations and have little time for experimenting on ‘futures’. However, scientific work usually focuses on the observation and analysis of ongoing processes, which lay the groundwork for experimentation, and one of the major opportunities for experimentation is during the conduct of exercises.

In this paper we report on a feasibility experiment we conducted during EX TENDI WALK at HQAST, from 02 to 06 February 2004. The purpose of the experiment was to assess the potential use of current speech recognition technology in the context of operation planning and to identify the limitations that will need to be addressed before such technology can be installed and used operationally, e.g. in the future Collocated Headquarters due to be completed in 2007. Another aim of the project was to collect real conversational data from planning sessions to perform further research on spoken language understanding, information presentation and pervasive computing. Although technical difficulties did not allow us to record all the data from all the planning sessions during the exercise, we were still able to collect enough data to allow us to start working on an analysis of the interactions during planning sessions. Even more important than the technical information we gathered or the data we were able to collect during this exercise, was the information about users and users' acceptance of the technology in their environment. Performing the experiment helped us ascertain that users are willing to use the technology, even with the relative inconvenience of having to wear a headset, and that they are already convinced of the benefits of automatically recording planning sessions: keeping an audit trail, improving training and induction and enabling the re-use of planning products.[1]

After giving the background to our research project in Sections 2 and 3, we discuss the issues related to planning such an experiment in Section 4 and then describe the actual experiment in Section 5. In Section 6, we outline the outcomes and benefits we expect to derive from this research project and what has already been achieved by conducting this feasibility experiment and in Section 7 we present the lessons we have learned and some remaining issues. We conclude in Section 8 by sketching where our next experiments will take us.

2. PREVIOUS EXPERIMENT

In 1996, Cross and Bopping (Cross and Bopping, 1998) were able to conduct an initial observation of the planning processes of an operational headquarters. In that study, as part of an analysis of Commander's Planning Group (CPG) and Theatre Planning Group (TPG) meetings, observers attempted to capture and categorise the elements of discourse within the planning process 'on the fly'. More specifically, the scientists sought to gain a sense of how theatre level decision making and problem solving processes could be understood in terms of the discourse and dialogue of which they are composed. The analysis focussed on patterns of discourse structure as it had the potential for elucidating the cycles of decision making that observers believed typified the planning process. It was anticipated that the analysis of patterns of discourse structure would not only 'standardise' the data and assure higher levels of confidence in any results obtained, but would also provide a valuable opportunity to investigate the suitability of this type of method for future use in the analyses of military planning and decision making. While the findings were suggestive, the study was marred by the necessity to capture data manually.

3. AUTM AND THE INTELLIGENT LISTENER

Since 1996, automatic speech recognition (ASR) and automatic transcription technology have progressed to the point where it is possible to envisage their use in group settings such as planning over the next few years. One such system is the *Automatic speech-to-text Transcriber for Meetings and Interviews* (AuTM) which was developed at DSTO (Zschorn et al, 2003). AuTM is a client-server application operating over a TCP/IP network to record and transcribe meetings and interviews. At the time of the experiment, AuTM had been extended to handle up to sixteen meeting participants.

Using a commercial ASR, AuTM automatically collects both textual and audio records. The records are organised according to the structure of the meeting and the audio and textual segments are time-stamped, aligned and linked. The output is a Word or HTML document containing the transcription of each speaker turn, which is linked to the audio file for that segment. Since the text and audio files are already matched, it is very easy to listen to the original audio input and correct the transcript as necessary. The AuTM output can be organised according to an agenda set by the meeting moderator; this facility can be customised to give any transcript a structure more relevant to the particular kind of interaction being recorded. In the case of a collaborative planning session, this could for instance follow the stages of the JMAP process.

---

[1] This was later reinforced by comments made by the ADFWC instructors who had attended EX TENDI WALK, during the "Introduction to Joint Warfare course" attended by one of the authors.

The work on AuTM is part of a task entitled *"Human Computer Interaction Research, including Speech and Natural Language Processing"*, referred to as the *Speech task*. In parallel to the development and improvement of the technology conducted within the HSI group as part of the *Speech task* (Sladek et al, 2003) we have embarked on a project, the "Intelligent Listener". The aims are to provide:

1) the ability to recognise and understand communicative intents and speech acts in planning sessions or meetings;
2) the ability to extract and retrieve information from recorded interactions during such sessions or meetings; and
3) the ability to organise this information to present it in a useful fashion.

As part of this work, we planned to conduct experiments in the live headquarters, in order to assess the feasibility of providing an "Intelligent Listener" capability in the new Collocated Headquarters. The experiments are designed to provide:

1) analytical assessment;
2) assessment and management of security;
3) assessment and management of ethical concerns;
4) assessment of technology for capture; and
5) assessment of processing requirements for captured information.

The first experiment was conducted during EX TENDI WALK at HQAST. For this experiment the aim was primarily to concentrate on assessing the security aspects, the technology for capture and to a lesser degree, the processing requirements for the captured data.

4. PLANNING THE FEASIBILITY EXPERIMENT AT EX TENDI WALK

Experimenting in a live headquarters relies on a number of factors which may be categorised as technical, people, organisational and financial. Technical factors turn out to be typically minor and are not usually prime indicators of success or failure. People and organisational facts represent a continuum from the individual user through to the organisation as user and sponsor.

The goodwill and support of the user population is critical to getting the experiment off the ground. In this case, the user group have been engaged at a number of levels. The initial group engaged was the Theatre Collocation Working Group in Canberra which comprises representatives from both what is now the Theatre Headquarters Project headed by a one star and users from Headquarters Australian Theatre. The reason for engaging both the Theatre Collocation and HQAST users was that speech recognition and language technology tools are now coming of age and are under constant commercial and research pressure to improve. DSTO will be testing and experimenting with the integration of such technology in HQAST over the next few years as it evolves incrementally to the new Collocated Headquarters.

The next level of engagement of the user group was more specific discussions with staff from HQAST Plans Branch. This was followed by an invited presentation to the whole of the Branch in which there was lively feedback and pledged support from the Branch. The next step in gaining support for the experiment was discussions with the senior level at HQAST and a presentation at the HQAST task reviews. There is never one point at which the scientist can assume that all the people issues are solved. Identifying and engaging a primary sponsor is fundamental. However, continuous engagement at many levels of the organization is also essential and it is at this point that engagement of people merges into engagement of the organization. In the military, security is a sine qua non of the organization and a number of alternative approaches to the experiment were considered but dismissed, such as running on the more highly classified network. We report on the user level of acceptance and support during the experiment in Section 7, but it is important to note that engaging the support of the actual users during the experiment was contingent on the level of support that had been obtained at the organisational level during the preceding months.

Finally, financial factors had to be considered. The feasibility experiment during EX TENDI WALK essentially involved setting up a network of fifteen high processing laptop computers with the appropriate peripherals, i.e. fourteen headsets and fourteen USB sound cards, and with the software licences for the commercial ASR to be used. The required hardware was not immediately available, either within the task or within HQAST, and it was necessary to resort to hiring thirteen of the machines from a commercial source with some task funding, while collegiate support provided us with two other machines on a loan basis.[2] For the ASR software licenses, we approached *VoicePerfect*, the Australian vendor for ScanSoft, who market Dragon NaturallySpeaking. They were very generous in their support to the experiment and graciously provided fourteen temporary licenses to use Dragon NaturallySpeaking v.7 for the two-week period. *VoicePerfect* also lent us fourteen headsets and fourteen USB sound cards for that same period.

5. CONDUCTING THE FEASIBILITY EXPERIMENT

5.1. Methodology

After the initial planning and organisation of the preceding weeks, the actual feasibility experiment involved setting up the hardware and software within the environment in which it would be used, i.e. the planning room at HQAST, and organising how the recordings would be conducted. We had allocated the week preceding EX TENDI WALK to set up the equipment and train the users on the ASR. We describe the technical set-up in detail in Section 5.2.

Dragon NaturallySpeaking is a speaker-dependent ASR, which means it needs to be trained for each user's voice and manner of speaking. This ensures better recognition rates and, although this is not yet quite satisfactory, allows the possibility of recording and transcribing conversational speech. In the current state of the technology, the alternative of a speaker-independent ASR does not allow the transcribing of conversational speech in meetings. Thus, to ensure the best possible results, we also had to organise user training with the ASR, and as the week preceding EX TENDI WALK was induction week at HQAST, we expected the EX TENDI WALK participants would be available during that time. We describe the user training phase in Section 5.4.

From the point of view of conducting the experiment, it was decided that the transcripts produced by AuTM would be saved at the end of each planning session and the data transferred to CD-ROM at the end of each day. We also decided that, during the capture of the planning sessions, at least two DSTO observers (the first two authors) would be present during each planning session to observe the interactions: this was intended to facilitate later correction of the transcripts and to enable the conversation analysis of the planning sessions. We report in Section 6 on what we were actually able to do during the recording sessions.

---

[2] We thank the *Speech Task* for allowing us to use their two demo laptops for several weeks.

5.2. Technical Set-up

From the technical point of view, it was first of all necessary to install the AuTM software and the ASR on the machines that had been hired. This was done during the week before the actual exercise by a member of the *Speech Task*, Jason Littlefield.[3] Then the machines had to be cabled and networked in the planning room. Again this was done during the week before the exercise, with the help of a member of the HQAST group, Peter Astenstorfer.

The hardware equipment originally consisted of fifteen networked Pentium IV laptops, one of which was used as the server. The server was a Compaq P4 1.7Mhz 1GB Ram 20GB HDD Win XP Pro Laptop with CD writer and the user machines were twelve Compaq P4 1.7Mhz 512 Mb Ram 20GB HDD Win XP Pro Laptops. The operating and standard software consisted of Microsoft Windows XP Professional and Microsoft Windows 2000. The speech recognition and meeting processing software consisted of DSTO's AuTM and the commercial ASR Dragon NaturallySpeaking v.7. The AuTM server was installed on the laptop selected as the server machine and the AuTM clients on the laptops provided to each member of the planning team around the table. The client machines would record, process and transcribe all the utterances of the planners during the planning sessions and then send both audio and text files to the server for insertion in the AuTM output transcript. The fourteen user laptops, each with its headset and USB sound card, were placed on the planning room table, one at each user position, and all the machines were networked together.

We already knew that the environment of the planning room at HQAST would present a number of challenges for the experiment: acoustically, because of the high level of acoustic noise (air conditioner, slide projector, number of participants); and logistically, because of the lack of space on the meeting table for all our machines alongside the laptops used by the planners-- and under the table for all the cables connecting all these machines on two different networks.

It was necessary to use two different networks because the AuTM system had not been accredited to run on the higher security network used by the planners. Otherwise, we would have been able to install our software on the laptops used by the planners and take advantage of the network already set up in the room. However, while installation on the higher security network would have alleviated the problem of space on and under the planning table, it would have introduced the more serious problem of making the users deal with "on/off" switches on the headsets dynamically: in the current accreditation for that network, only headsets with "push-to-talk" switches are permitted. In effect with a default setting of "off", the users would have had to press the switch every time they spoke in order to capture their voices.

Another challenge was the use of headsets to record speech. Although we could expect that some users would be accustomed to wearing headsets, we had already heard during the preparatory discussions that we could also expect a certain amount of reluctance on the part of other users. However, it is widey recognised that speech recognition requires a very high acoustic quality for the input and that in a noisy environment, such as a meeting room, table or ceiling microphones will be useless for ASR, therefore individual close-talking and noise-cancelling microphones are absolutely mandatory with the current technology. Background or environment noise and reverberation drastically affect the ASR performance. Experiments with lapel microphones in that kind of environment have also yielded disappointing results, so for now, the only

[3] One lesson to be learned from that phase of the experiment is that preparation will always take longer than expected. Jason Littlefield was scheduled to spend two days in Sydney, but had to delay going back to Edinburgh several times and ended up staying at HQAST for four days.

alternative is for users to wear headsets with built-in microphones.[4] The microphones which were used in the experiment were all *VXI TalkPro* headsets.

5.3. Surprises and changes

An important lesson to be learned is to "expect the unexpected". As mentioned above, there were originally thirteen laptops hired from *Hire Intelligence* (twelve for the AuTM clients and one for the AuTM server) and two laptops from DSTO (both used for AuTM clients). This gave a total of fifteen machines networked together, with fourteen laptops on the briefing table for the participants (labelled #1-14) and 1 server (labelled #15) on a separate table at the back of the room.

On the next to last day of preparation, all the machines had been installed, labelled and networked together, and we were going through the user training phase when we found out during one of the user training sessions that there were actually sixteen positions around the planning table, instead of the fourteen we had expected. We then had to hurriedly arrange for the hiring of two more machines from *Hire Intelligence*, obtain two more temporary licenses for Dragon and borrow two more headsets and USB sound cards from *VoicePerfect*. The AuTM system and the Dragon ASR software then had to be installed on these additional machines, and these machines had to be added to the network. The final set-up was finally completed late on the Friday before the exercise.

5.4. User Training

As mentioned above, to ensure as good results as possible with a speaker-dependant ASR, it is necessary to capture the users' speech profiles and train the ASR to their manner of speaking. We tried to conduct the training of the ASR and the capture of the users' speech profile during the HQAST induction in the week prior to the exercise. However, it proved rather more difficult than expected to identify who the actual participants to the exercise were going to be in time for them to go through the ASR training, because some of these decisions had not yet been made by the end of the week. Identifying who the participants were and scheduling times that would be convenient for them turned out to require extensive use of personal contacts within HQAST and with the other units involved.

An additional complication was the fact that the planning room was being used during our preparation week and was often not available to us. This delayed our installation of the machines and the network, but also prevented us from conducting training sessions at times that would have been convenient for the users. For best results, training must be conducted in the same acoustic environment in which speech recognition is expected to occur; the acoustic properties of the planning room (see Section 5.2) were sufficiently challenging that we could not hope to get away with training in a different environment, such as an office. In fact, we even made sure we trained each user in the position they were expected to occupy during the planning sessions, and on the machines which they would be using. This helps with the problem of reverberation and ambient noise.

---

[4] Note that for this kind of application, it is not necessary for the users to actually put the earpiece part of the headset on their ear because they do not need to, or indeed would want to, hear what is being recorded. This fact helped some users better accept wearing their headset, in particular some people from Lan who said they had difficulty hearing.

By the Friday before the exercise, we had trained fourteen potential users with Dragon. Since there were sixteen positions around the table, and several individuals would occupy those positions during the exercise, it was clear we would need to use some generic profiles, so we created generic profiles for the positions for which we had no specific individual profiles. We were able to train some additional users during the first few days of the exercise but, by the second day of the exercise, we realised that we would in fact need to load generic profiles, both male and female, on all the machines. This was necessary to cater for the situations when new users came in and took the place of one of the established users. We also realised early on that we needed to have all the individual user profiles available on all the machines, because people would sometimes come in and sit, not at their assigned position, but at the one next to it. Ensuring that all machines had all the individual user profiles plus generic profiles available was a time-consuming task, as it required uploading each profile from the machine on which the ASR had been trained up to the server, and then downloading each profile onto each of the client machines. One lesson to be drawn is that the user profile management needs to be simplified and automated as much possible.

Most of the users who went through the training were quite enthusiastic about the project. Some of the users who were not able to do the ASR training, either before or during the week of the exercise, were also quite supportive and made encouraging comments.[5] The level of support for the project did not appear to correlate with whether the users had the opportunity to do the training.

5.5 Recording the planning sessions

As expected, a number of things did not go according to plan once the exercise was under way and we started recording during the planning sessions. Some difficulties were technical and will be remedied in the next version of the AuTM software (e.g. network architecture, interface, user profile management). Some difficulties were due to the quality of speech recognition at this point in the development of the technology and will require different solutions in the future (e.g. different types of microphones for better quality audio input, better noise cancellation and reverberation cancellation, improved acoustic design for meeting rooms). Other difficulties were due to human factors and should be resolved when the technology becomes an accepted part of the environment (e.g. remembering to put on the headset and/or turning the microphone on).

The first difficulty was that, although the users had been willing to put their headsets on at the beginning of the first session, when the leader of that session announced that there would first be a presentation by himself and some of the instructors, they all took their headsets off and put them back on the table. The room was packed, with not only the sixteen participants around the table, but about thirty more people sitting behind them. With about fifty people in the room, it was impossible for the two DSTO experimenters sitting at the back next to the AuTM server to intervene. Besides, it was an interesting situation: how would the system handle this? It was not long before the answer became obvious: the microphones were all picking up speech signals from the presenters and all the AuTM clients were sending files to the AuTM server. Soon, the network was unable to cope, some clients failed and eventually the server failed. No transcript was produced for that session. A similar problem recurred throughout the week when sessions were broken up in smaller

[5] It is interesting to note that one of the users who did not manage to find the time to go through the training turned out to be one of the most active and important participants in the planning sessions, while many of the users who did find the time to do the training were actually very quiet during the exercise. This points to another issue: the trade-off between the seniority and experience of the participants, which correlates with their level of participation, and the time they have available to ensure the system will work for them, which inversely correlates with their level of participation.

groups, or when people got up and left their headsets on the table. In a number of those cases, no transcript was produced for those sessions. However, we discovered that even when the AuTM server had declared itself unable to cope and had failed, the audio files continued to be uploaded to the server machine and were saved in the data directories. Thus we still have the recordings for most of the planning sessions, even for the sessions for which we do not have a full transcript.

The issue of networking machines for the AuTM server and the AuTM clients and the related problems of network overload should be solved in the new AuTM version which is being implemented with Elvin as the messaging system.

Another issue which became obvious on the first day of the experiment was that of users trying to either look at the transcript being displayed on the screen of the laptops used for recording and speech recognition, or trying to use that machine for other purposes (e.g. to access their email). There were some amusing instances, especially during the first session when, as mentioned earlier, the experimenters could not intervene and had to watch helplessly as the session leader tried to use the DSTO laptop while a presentation was being given until he gave up and finally noticed the "DSTO Restricted" sticker on the cover of the laptop. We had labelled all the experiment laptops and tried to keep the covers closed but obviously this was not enough when equipment got moved around on the crowded table.

We did not want the users to have access to the transcript during the experiment because we knew the accuracy of the speech recognition would be low. Testing the accuracy of the speech recogniser was not our aim in this experiment, the goal was to assess the useability of the system and to identify the issues that have to be addressed before such a system can be used in operations. However, we did not want our users to be distracted by the output, and maybe then modify their behaviour, nor did we want them to form an opinion about the system based on the recognition accuracy. It would be a simple matter to blank the screens or to install a password-protected screen saver. However, this would have required modifications to the AuTM interface and we were not willing to risk making such modifications once the system was installed. This is a good example of a useability problem which only actual experimentation will uncover. The ultimate solution is to have all the input microphones leading to one single computer where speech recognition is performed; this computer should not be visible or accessible to the users. The lesson is that only the meeting moderator, or the experimenters, should be able to access the transcript being produced.

The related issue of users wanting to use the speech recognition laptops for doing other work will also be addressed by not having separate machines for the AuTM clients but only one computer, removed from the users.

6. OUTCOMES AND BENEFITS

In the short term, we will have demonstrated that the ADF can be provided with the capability to produce and retain electronic recording of planning sessions and to have access to drafts of transcriptions for those sessions. These draft transcriptions are automatically segmented and aligned with the recordings. This capability will ensure that in the short and mid-term, the headquarters will have access to easily searchable archives, which are aligned with the products from the planning sessions. This will permit more rigorous and accurate auditing, and will allow information retrieval from those archives, both for re-use of deliberate and immediate planning products and for the induction and training of incoming staff. The experience gained during this experimentation will also greatly help the ADO establish the skills and expertise base needed to implement and use video and audio mining technologies in information and intelligence applications and in the extraction of information from spoken data. It will provide a basis for further research on spoken language understanding, information presentation and pervasive computing. The data gathered during the exercise can be subjected to a more detailed linguistic analysis than is possible with notes and human observation and will inform further development in spoken dialogue processing.

7. ISSUES AND LESSONS LEARNT

7.1. Technical aspects

The main technical issues identified were the following:

- architecture of the system at the time of development: network communication between clients and server not robust enough for the amount of data that was actually being sent to the server;
- need for a better user profile management;
- do not make the experiment machine available to the participants.

We already knew that it would be a lot better to have only one computer for the whole system and not use individual laptops for each AuTM clients, and we are working towards that solution.

In addition the following minor requirements were noted:

- need laptop number in transcript;
- need User Profile ID on AuTM Client GUI;
- need to be able to control the AuTM Clients microphones from the Server (or the Moderator);
- need to be able to cope with clients taking their headsets off when the microphone is still recording;
- need to blank the screen so the transcript is not available to the users.

7.2. People Aspects: User acceptance and users' needs

As mentioned previously, the number of users was not constant. In addition, users were not always willing to keep their headsets on. User acceptance of the training regime, the set up and set down of the individual laptops for each session and the necessity to wear headsets correctly and continuously required the following:

- Initial and explicit endorsement of the experiment by leader of the exercise and whenever the group composition changed substantially.
- A reminder and re-endorsement of the experiment by the session leader of the necessity to follow the experimental protocols, for example, wearing the headsets.
- Session leader’s example in following the protocols.

A system of rewards for users who wore their headsets during a session and gentle personal reminders to users on a one-to one basis proved most effective. We considered and rejected punitive shaming on a group basis.

7.3. Organisational Aspects: Perceived User Benefit

Conducting an experiment in a live headquarters represents a substantial undertaking by the organisation. While it is possible to argue that substantial benefits will ensue (see Section 6) and that the utility of these benefits does emerge through the experiment, there is a requirement to demonstrate a user benefit that is meaningful and particular. We identified such a benefit in discussion with one of the primary sponsors immediately following the exercise. During the second day of the exercise, we were asked whether we had indeed captured all the discussion from the preceding day, because it was already felt that not enough of the discussion had been captured by the traditional means of producing PowerPoint slides and that intellectual work carried out in the earlier phases of the planning had then to be reworked without the advantage of the earlier work. As we analyse the recordings and transcripts from the experiment, we will document that incident (and many of a similar nature).

7.4. Financial Aspects

An experiment in a live environment with the dynamic "surprises and changes" noted in section 5.3, requires sufficient and flexible funding in order to bring immediate changes. The funding itself is a given requirement, but it also requires efficient and flexible processing systems, e.g. purchasing, to direct the funding in a timely manner to achieve the required outcome, such as, obtaining extra machines and extra effort from personnel. Of course, for the latter, much more than funding is required, primary of which is a dedicated and committed workforce.

8. CONCLUSION

Speech recognition technology works adequately for individuals. It is getting to the point of being useful for groups. There will be a time when it is pervasive – a part of every well equipped planning room – perhaps even with the discretion of an intelligent human listener. Eventually such tools and the accompanying human processes will be embedded organically into C2 environments of the future.

In this paper we have reported on the feasibility of experimenting in a live headquarters using speech technology. We have described the sustained effort to set up and conduct the experiment. By the time the symposium takes place we will have concrete results to report, and we will be able to draw more lessons from the experiment. For now, we have discussed the necessity to organise this kind of activity very much in advance, to engage the customers at as many levels of the organisation as possible and to actively seek collaboration between DSTO, commercial partners and the Military.

REFERENCES

1. Cross M. and Bopping D. (1998). "Collaborative Planning Processes in Command and Control". *Proceedings of ICCRTS, 1998, Sweden.*
2. Sladek, J., Zschorn, A. and Hashemi-Sakhtsari, A. (2002). "Speech to Text Transcription in Support of Pervasive Computing", *Ubiquitous Computing: Trends, Tricks and Traps, and the Inaugural Asia Pacific Forum on Pervasive Computing,* October/November 2002, Adelaide Convention Centre, Australia. This paper was published in *Volume 25 - Advances in Computer Science Selected Papers from the ACS State Conferences*, edited by Brenda Ainsley.
3. Zschorn A., Littlefield J. S., Broughton M., Dwyer B. and Hashemi-Sakhtsari A. (2003). *Transcription of Multiple Speakers Using Speaker Dependent Speech Recognition*. DSTO Technical Report, DSTO_TR_1498.