AUTOMATIC SPEECH RECOGNITION IN AIR TRAFFIC CONTROL:
A HUMAN FACTORS PERSPECTIVE*

Joakim Karlsson
Research Assistant
Flight Transportation Laboratory, M.I.T.
Cambridge, MA

ABSTRACT

The introduction of Automatic Speech Recognition (ASR) technology into the Air Traffic Control (ATC) system has the potential to improve overall system safety and efficiency. However, because ASR technology is inherently a part of the man-machine interface between the user and the system, the human factors issues involved must be addressed. This paper identifies some of the relevant human factors problems and presents related methods of investigation. Research at M.I.T.'s Flight Transportation Laboratory is being conducted from a human factors perspective, focusing on intelligent parser design, presentation of feedback, error correction strategy design, and optimal choice of input modalities.

INTRODUCTION

In today's ATC system, communication between controllers and aircraft is almost exclusively verbal. This is especially true for such critical tasks as the issuing of clearances and vectors, to achieve traffic separation. Although a digital datalink is in development (Mode S), there is no reason to believe that voice communication between ATC and aircraft will disappear in the near future. As a result, most of the information transferred within the system is never captured in machine readable form. Herein lies the promise of introducing ASR technology into the ATC system: it would permit processing of ATC clearances, to ensure conformance to safety and separation criteria. It would allow the ATC computer system to predict the future state of the airspace. The controller could prestore routine clearances during periods of little activity. Mode S equipped aircraft could be provided with a machine readable copy of verbal clearances for confirmation purposes.

Thus, introduction of ASR technology could result in the reduction of human errors, resulting in increased system safety. However, the dilemma of ASR is that its purported advantages are not automatically realized by simply making the technology available. Careful human factors design is necessary to capitalize on its potential [Berman, 1984]. This is especially true in the case of ATC, which is plagued by human factors problems such as intense levels of workload during traffic peaks intermixed with controller boredom during low demand periods. Furthermore, the high probability of loss of lives in the case of errors makes it imperative that the human factors problems created by introducing ASR into the Air Traffic Control system are properly addressed and solved.

The speech recognition devices available today are not sufficiently capable to be used operationally within the ATC environment. However, there are units available that are useful for the required preliminary human factors research. In order to minimize human factors problems, it is necessary to implement an iterative design cycle that should be continued until the needs of the system users are met [Cooper, 1987]. The research presented within this paper should be considered as one step in that cycle.

---

*Paper presented at Military and Government Speech Technology 1989, Nov. 13–15, 1989, Arlington, VA.

MODELING HUMAN FACTORS

In order to approach human factors in an analytic way, a conceptual model of the system resources available can be used. The system resources include hardware, software (rules and regulations), liveware (users), and the environment. The SHEL model, named by the initial letters of these resources, can be used to represent the components and their links [Edwards, 1988]. Figure 1 contains a graphical representation of the SHEL model. The connecting lines between the system components represent the interfaces between the respective resources. It is at the interfaces to the liveware component that most human factors issues occur.

Figure 1: The SHEL Model.
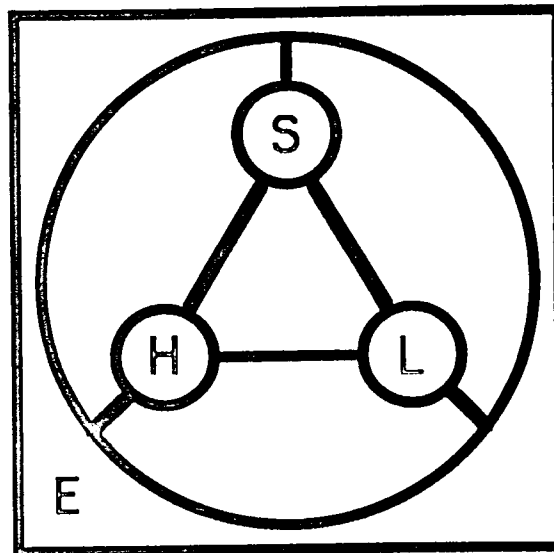
In terms of the SHEL model, examples of human factors problems include microphone placement and characteristics (hardware-liveware interface), speech variations due to background noise (environment-liveware interface), and design of error correction strategies (software-liveware interface). Note that not all human factors issues are strictly related to one single interface to the liveware component. Examples include fatigue, stress, boredom, and user acceptance of ASR technology. It should also be emphasized that ATC is a multi-user system. Thus, there are also liveware-liveware interfaces that must be considered.

IDENTIFYING HUMAN FACTOR ISSUES

Identifying human factors issues related to ASR technology is a topic that has been covered adequately and extensively [Constantine, 1984]. However, ATC is fundamentally different from other ASR applications in several ways:

- In ATC, voice is the primary communication channel, and microphones are already used.
- The ATC vocabulary and syntax are already defined and cannot be easily altered.
- Human errors in the ATC environment can lead to fatal results.
- The background noise consists of distinct voices, not random noise.

Hence, we can consider three categories of human factors issues: common issues that are mutual to both ATC and other ASR applications, unique issues that are typically not encountered in other applications, and non-issues – problems that may be significant in other applications, but that do not play a major role in ATC.

The last group, non-issues, is of course the most trivial to consider: a good example is the hardware-liveware issue of microphone characteristics and placement. Headset mounted noise-cancelling microphones are already in use in the current ATC environment, and hence it is an issue that has been addressed extensively before ASR technology has come under consideration. Also, the software-liveware problem of vocabulary and syntax definition, normally an important human factors issue, has also been completed: these definitions are controlled by the Federal Aviation Administration (FAA). Another common problem, communication with other people while in recognition mode, is not likely to occur in ATC, as controllers are already using Push-To-Talk (PTT) switches on their headsets.

Issues that are common to both ATC and other ASR applications are abundant, and must not be neglected, although they have already been addressed extensively. These include:

- Speech variations due to stress, fatigue, or background noise.
- Spurious recognition due to background noise.
- Inter-speaker variations (the "sheep and goats" issue).
- User acceptance of the technology.
- User motivation.
- Presentation of feedback to the user.
- Error recognition, presentation, and correction.
- User training.
- Selection of proper hardware.
- Optimal use of mixed input modalities.
- Recognition accuracy and use of higher levels of knowledge.
- Failure to adhere to the syntax.

Although several of these problems remain unsolved, most have been addressed previously. The research being conducted at the Flight Transportation Laboratory covers some of these issues, since much of the previous work has not been applied specifically to ATC.

The final group, unique issues, includes problems that are either specific to ATC, or that are more significant in the ATC environment than elsewhere. A typical example is stress induced reduction of recognition accuracy, mentioned above as an issue common to other ASR applications. This is a much more critical issue in Air Traffic Control, since the cases where the introduction of ASR has the greatest potential of improving safety, are likely to be stressful situations. The possibility of automating conformance monitoring would greatly benefit the controller during scenarios where a large number of aircraft are being controlled – a stressful period for the controller. It is exactly in the conditions where ASR technology is needed most, that it performs worst. To the human factors researcher this points out the importance of high baseline recognition accuracy, high levels of robustness in the presence of speech variations, introduction of functional automatic error correction techniques, and the design and implementation of parsers that make use of higher levels of knowledge such as prosodic, syntactic, semantic, and pragmatic information.

Another major issue facing the introduction of ASR technology into the ATC environment is that of cognitive workload. The controllers are already presented with a wealth of information, and if any new technology is to be introduced it must reduce workload, not increase it. There exists a need to ensure that the information captured through the use of speech input technology is the same as the information transmitted to the aircraft. Hence, the controller must monitor what is understood by the machine, in order to be able to correct it. However, this would introduce another task for the controller, and possibly distract from the visual attention that the radar display demands. This dilemma underscores the importance of designing adequate feedback and error correction strategies.

RESEARCH AT THE M.I.T. FLIGHT TRANSPORTATION LABORATORY

It is within the framework presented above that the ASR research effort at M.I.T.'s Flight Transportation Laboratory has been conducted. Only a brief description of this research can be presented in this paper -- more detailed descriptions are available elsewhere [Karlsson, 1990]. Preliminary results include a study to choose the ASR hardware most suited for ATC human factors research, purchase and evaluation of the Votan VPC 2000 and Verbex Series 5000 voice I/O systems, and design and implementation of a low-cost portable research station using the Verbex Series 5000 and a PC based ATC simulator. An extensive annotated bibliography of related papers has also been compiled.

Future work will concentrate on means to improve recognition accuracy while maintaining a low workload level. Techniques will include the use of semantic and pragmatic information, adaptive (on-the-fly) training, introduction of confusability matrices and other automatic error correction techniques [Loken-Kim, 1985], mouse and menu input for error correction, and need-to-know type feedback that ensures that the use of ASR technology remains mostly transparent to the controller. Furthermore, a receiver station has been established to monitor ATC communication in the greater Boston area, to study real life use of the ATC language and provide data for issues such as syntax deviation.

CONCLUSIONS

The importance of the human factors aspects of introducing ASR technology into the ATC environment cannot be underestimated. In particular, it must be realized that ATC applications are uniquely different from other applications where voice input may be of benefit. As a result, much greater emphasis must be placed on issues such as mental workload, user feedback, mixed use of input modalities, intelligent parser design, and improved robustness with respect to speech variations. The ASR research being conducted at the M.I.T. Flight Transportation Laboratory has resulted in a set of tools that can be used to identify, quantify, and provide preliminary solutions to the human factors issues described within this paper. The results can then be used as a step in an iterative design cycle to obtain a system acceptable to the user.

REFERENCES

Berman, J. V. F., "Speech Technology in a High Workload Environment", *Proceedings of the 1st International Conference on Speech Technology*, IFS Publications Ltd, Bedford, U.K. (1984), pp. 69-76.

Constantine, Betsy J., "Human Factors Considerations in the Design of Voice Input/Output Applications", *The Official Proceedings of Speech Tech '84*, Media Dimensions, Inc., New York, N.Y. (1984), pp. 219-224.

Cooper, Martin, "Human Factors Aspects of Voice Input/Output", *Speech Technology*, Vol. 3, No. 4 (March/April 1987), pp. 82-86.

Edwards, Elwyn, "Introductory Overview", *Human Factors in Aviation*, eds Wiener, Earl L. and Nagel, David C., Academic Press, Inc., San Diego, CA (1988), pp. 3-25.

Karlsson, Joakim, "Automatic Speech Recognition in Air Traffic Control", *Joint University Program for Air Transportation Research – 1988,1989* NASA Conference Publication 3063, NASA Langley Research Center, Hampton, VA (1990), pp. 3-15.

Karlsson, Joakim, *The Introduction of Automatic Speech Recognition Technology into the Air Traffic Control System*, FTL Report 90-1,Flight Transportation Laboratory, M.I.T., Cambridge, MA (Jan. 1990).

Loken-Kim, K.H., "An Investigation of Automatic Error Detection and Correction in Speech Recognition Systems", *The Official Proceedings of Speech Tech '85*, Media Dimensions, Inc., New York, N.Y. (1985), pp. 72-74.