557848 p10
N90-20923

Automatic Speech Recognition in Air Traffic Control

Joakim Karlsson
Flight Transportation Laboratory
Massachusetts Institute of Technology
Cambridge, MA

Automatic Speech Recognition (ASR) technology and its application to the Air Traffic Control system are described. The advantages of applying ASR to Air Traffic Control, as well as criteria for choosing a suitable ASR system, are presented. Results from previous research and directions for future work at the Flight Transportation Laboratory are outlined.

Introduction

M.I.T.'s Flight Transportation Laboratory (FTL) is renewing its research on the application of Automatic Speech Recognition (ASR) technology to Air Traffic Control (ATC). This report presents an overview of the available technology and its potential use within the ATC system. ATC is a suitable candidate for the application of speech input/output technology due to the well-defined syntax and existing reliance on voice communication. Other motivations for introducing ASR into the Air Traffic Control environment are listed within the body of this report. Furthermore, past research efforts are described, with emphasis on work already completed by the Flight Transportation Laboratory. Finally, directions for future research are outlined.

- **Just what is Automatic Speech Recognition (ASR) anyway?**
- **ASR in Air Traffic Control.**
- **Some motivations for using ASR in Air Trafic Control.**
- **Previous work.**
- **Conclusions from Trikas' work.**
- **Work to be done at the Flight Transportation Laboratory.**

Automatic Speech Recognition

ASR systems consist of hardware and software that convert verbal input into machine-useable form (i.e., "text"). These systems can be categorized by three basic parameters: Speaker dependence/independence describes whether the system has to be trained by the user before operational use (speaker dependent), or whether it can be used by any user without specific training (speaker independent). Discrete/connected/continuous speech recognition describes the extent to which naturally spoken speech can be recognized. Single-utterance (isolated-speech) recognizers impose severe constraints on the user's manner of speech, but are relatively easy to implement. Connected speech recognizers allow the user to speak at a normal rate, but finite pauses must be inserted between each word. A continuous speech system recognizes input spoken at a natural rate, with no artificial pauses. Finally, the number of words that the system can recognize at any time (active vocabulary size) is a critical application and performance parameter.

An Automatic Speech Recognition (ASR) system is a system that recognizes verbal input and translates it into text. There are three basic factors that categorize an ASR system:

- **Speaker dependence/independence.**
- **Discrete, connected, or continuous speech recognition.**
- **Vocabulary size.**

ASR in Air Traffic Control

Today, the Air Traffic Control system relies on verbal communication between the air traffic controllers and the pilots of the aircraft in the controlled airspace. Although a computer system exists that processes radar and other information regarding the aircraft, the information contained within the verbal communications is not retained. The introduction of ASR technology would allow this information to be captured. The purpose of this research effort is to demonstrate the feasibility of using ASR technology within the ATC environment and to address the problems involved, especially the relevant human factors issues. Off-the-shelf ASR technology will be used in conjunction with FTL's real-time ATC simulator running on the laboratory's TI-Explorer Lisp machines.

We want the "computer" to capture the information given by the controller to aircraft, so that it can be processed. In this particular project, we want to start by using ASR to drive the Flight Transportation Laboratory's real-time ATC simulator.

Why Use ASR in ATC

There are several strong motivations for introducing speech input/output technology into the Air Traffic Control system. Communications are already in the verbal form, and the syntax used is clearly defined by the FAA and has to some degree been designed to reduce the possibility of communication errors. The use of voice as an input modality allows for a high information throughput capacity and allows the controllers to keep their eyes and hands busy controlling traffic. Once the verbal information has been captured, it can be transferred to the aircraft via Mode S, conformance monitoring can be improved, and routine clearances can be pre-stored during less busy periods.

- **ATC communication is verbal.**
- **ATC syntax is clearly defined.**
- **ATC training can be automated.**
- **High information throughput.**
- **ASR allows controller to use hands and eyes where they belong.**
- **Captured information can be transmitted to aircraft via Mode S.**
- **Conflict alert can be improved.**
- **Clearances can be pre-stored.**

Previous Research

ASR technology can be used in many aviation and non-aviation applications, and as a result, much research has been conducted on the use of speech input/output in general. However, relatively little research has been dedicated toward the application of ASR to Air Traffic Control. The research to be undertaken within the framework of this project will be a continuation of the initial work presented in Thanassis Trikas' S.M. thesis, *Automated Speech Recognition in Air Traffic Control* (FTL report R87-2).

A lot of research has been done on ASR, but not much in conjunction with ATC:

- **FTL: Thanassis Trikas S.M. work.**
- **Arthur Gerstenfeld (Worcester Polytechnic Institute/UFA, Inc.): Emphasis on ATC training.**
- **ITT Defense Communications Division VRS 1280 demonstration.**

Trikas' Conclusions

Trikas' thesis demonstrated the feasibility of using ASR technology in conjunction with an ATC simulator, utilizing a relatively small vocabulary. An initial error correction strategy based on verbal correction commands alone proved to be unacceptable. Also, problems related to speech articulation variations were encountered. In the process of evaluating his experiment, Trikas implicitly set forth a set of criteria for selecting a suitable ASR system.

Trikas' S.M. thesis was essentially a proof of concept of using ASR in ATC:

- **ASR <u>can</u> be used with the ATC simulator (with an active vocabulary of only 64 words).**
- **Correction of recognition errors using voice alone is not feasible.**
- **Problems with sensitivity to variations in articulation.**
- **Developed criteria for choosing a suitable ASR system.**

Selecting the Right ASR System

The first step in renewing FTL's ASR research effort will be to select a suitable hardware system. For this purpose, performance criteria specific to ATC applications of speech input/output technology have been defined.

Our particular application calls for the following ASR requirements:

- **Speaker independence <u>not</u> required.**
- **Continuous speech recognition.**
- **Vocabulary size 200-300 words.**
- **95% baseline recognition accuracy.**
- **Well-designed training procedure.**
- **Open architecture.**
- **Reduced sensitivity to variations.**
- **Short recognition delays (1-4 s).**

Future Work

The future research to be conducted at FTL will be based on previous work completed by Trikas. Hence, his system setup must be reactivated. In order to improve the simulation and the overall performance of the system, new hardware will be acquired. The actual research will concentrate on the introduction of multi-modal input, improved error correction and recognition accuracy, the evaluation of Mode S usage, and the application of ASR to secondary functions.

- **Reassemble Trikas' system.**
- **Evaluate current ASR technology.**
- **Acquire a new ASR system.**
- **Introduce multi-modal input.**
- **Increase number of commands and responses to improve simulation.**
- **Improve error checking/correction, as well as recognition accuracy.**
- **Evaluate Mode S usage.**
- **Use ASR for functions other than ATC commands.**