REDUCING PILOT / ATC COMMUNICATION ERRORS USING VOICE RECOGNITION

CLAUDIU – MIHAI GEACĂR
Faculty of Aerospace Engineering
Politehnica University

Keywords: *voice, recognition, ATC, Link2000+*

Abstract

Traditionally, pilots and air traffic controllers use radios to communicate with each other. As air traffic increases, routes get more and more diverse and light and ultra light aircraft are becoming more and more popular, the classic system is beginning to show its weaknesses. The two main issues regarding classic radio communications are channel congestion and the language – related problems.
Statistics have shown that almost 80% of all pilot radio communications contain one or more errors. Furthermore, in the case of the air traffic controllers, the same statistics show that 30% of all incidents are caused, among others, by communication errors and 23% of flight level intrusions are caused by communications errors (40% in the case of runway incursions).
Channel congestion issues are approached today by CPDL (Controller – Pilot Data Link), which uses an alternate system for the routine tasks that take up most of the available channel. One of the best known applications of CPDL is Eurocontrol's Link2000+, which is set to become operational in all EU countries by 2016.
CPDL uses a data link to exchange preset messages between pilots and air traffic controllers. Message types include approval/denial, request and information. Air traffic controllers can assign flight levels, approve/reject route changes, assign radio frequencies and request information. Pilots can respond to messages, request approvals and information, declare and dismiss an emergency. Text messages can be sent, in case there is no standard message type applicable for the information being exchanged.
The purpose of this paper is to provide error reduction methods for pilot-air traffic controller communications using voice recognition and data links.
Voice recognition converts spoken words into machine-usable data. Its uses include communications applications, medical applications and military applications.
The output of the voice recognition system can be used in two ways: text broadcast, where the system acts as a backup to the classic voice radio system or it can be used to control CPDL systems, such as Eurocontrol's Link2000+.
In the first case, the system is designed to reduce communication errors between non-native English speakers or between native and non-native English speakers and also to provide a solution to channel congestion.
In the second case, the system is designed as an improvement to the Link2000+ system, reducing pilot / traffic controller workload and providing a more natural control interface for Link2000+. In this case, the system acts as a "vocal control" interface, eliminating the need to use the Control Display Unit (CDU) or the keyboard and mouse to input commands and allowing more time for important activities.
The main benefit of this system is the increase in safety, as communications were one of the few aircraft systems without redundancy. Other benefits are workload reduction and more accurate communications.

1 Introduction

Pilot/ATC communication is normally done using voice radios. Initially, the capacity and capabilities offered by this system were sufficient. As air traffic and routes increased and small, private aircraft became more available, this system began to show its weaknesses. The most important ones are channel congestion and language issues.

Channel congestion is caused by the increasing number of aircraft using the same radio frequency while in the same ATC sector. As this number increased, the amount of information became more and more difficult to transfer using that single frequency.

Channel congestion can cause aircraft delays and, most importantly, can affect the safety of the passengers and flight crew. One of the most dangerous situations is encountered when two people try to talk on the same frequency at the same time, when everybody listening to that frequency will only hear a squeal.

Probably the most tragic accidents caused, among other factors, by channel congestion, was the one in Tenerife, on March 27, 1977, when a KLM Boeing 747 crashed while taking off into the top of a Pan Am Boeing 747 that was on the same runway and resulting in the loss of 583 lives (making it the deadliest accident in aviation history).

Language issues are becoming ever more common today. As the word becomes more and more "global", language becomes a key factor in the efficiency of pilot/ATC communication. The two main types of language issues are related to similar phonetics of words or numbers and to different levels of proficiency in using the English language. Other types of issues are the use of non-standard phraseology or the use of other languages than English.

Language issues are very important because they can lead to dangerous situations without any of the parts involved being aware.

Studies have shown that about 80 percent of all pilot radio messages contain at least one error. For air traffic controllers, the same studies have shown that:

- 30% of all incidents are caused (among other factors) by communication errors (in TMA areas, the percentage rises up to 50%);
- 23% of all flight level incursions are caused by communication errors;
- 40% of all runway incursions are caused by communication errors.

These results have urged the need for ways to increase the reliability of radio voice communications. Efforts have been made in order to improve crew English knowledge, to ensure proper use of ICAO standard phraseology and to avoid confusing situations.

The channel congestion issue was initially solved by dividing the crowded sector into two smaller sectors, each with its own frequency. This proved to be a temporary solution as the dividing could not keep up with the rate of traffic growth in some of the sectors.

The latest approach in channel congestion is the development of a system that handles the routine operations that normally imply the use of voice radio communications and is not affected by traffic volume. This system is known as Controller-Pilot Data Link or CPDL. One implementation of CPDL is Eurocontrol's Link2000+, which will become operational in all European Union countries by 2016.

As for the language issues, efforts have been made in order to increase the level of personnel training, use of standard ICAO phraseology, as well as the development of procedures for the preventions of incidents and accidents related to language issues.

Furthermore, rules and recommendations have been issued in order to avoid confusing situations, such as:

- clearly stating if three-digit numbers ending in a "zero" refer to altitude or heading;
- use of ICAO letter pronunciation in order to avoid confusion between B and G or C, D and the number "three";
- the avoidance of confusing statements, such as "made a …" and "mayday" or "hold in position" and "holding position";

- grouping of similar words ("climb to two thousand" can be mistaken for "climb two two thousand").

The language issue remains open, with a growing concern caused by the increasing traffic and availability of affordable light aircraft.

This paper presents a new approach for voce radio communication issues, using state-of-the-art technologies like CPDL and voice recognition. The key benefits of this approach are the workload reduction for both pilots and air traffic controllers, as well as the increase in air traffic safety, by eliminating language and congestion-related communication errors.

2 Voice recognition

Voice (or speech) recognition is the process of converting spoken words into machine-usable information (like binary codes of a character string). Voice recognition was initially developed in the United States as a substitute for the filling of medical transcripts. Today, voice recognition is widely used, some of its most important applications being:

- medical: people with physical disabilities can use voice commands to control various equipment;
- military aviation: voice-commanded cockpit (radio tuning, autopilot setting, display control), battle management;
- civil aviation: air traffic controller training.

2.1 Voice recognition performance

The performance of voice recognition systems is usually given by accuracy and speed.

Voice recognition systems can be speaker-dependent (requiring training data and accurate only for a given speaker) and speaker-independent (less accurate, slower, but usable by any speaker without prior training).

Accuracy can be measured using word error rate (WER), single word error rate (SWER) and command success rate (CSR).

Most commercially available systems have, according to manufacturers, 98-99% accuracy, in optimum conditions (training data, quiet environment). This is why, in spite of high factory accuracy levels, most systems cannot manage real-life conditions (voices with accent, ambient noise etc.) satisfactorily.

2.2 Types of voice recognition systems

Voice recognition systems can be described by many parameters, the most important ones being shown in Table 1.

Parameter	Variation
Speech mode	Isolated words to continuous speech
Speech style	Reading to spontaneous speech
Training	User-dependent to user-independent
Vocabulary	Small (less than 20 words) to large (more than 20000 words)
Language model	Finite to context-dependent
Perplexity	Low (below 10) to high (above 100)
Signal-noise ratio	High (above 30 dB) to low (less than 10 dB)
Input device	Noise-canceling microphone to telephone

Table 1, Main parameters of speech recognition

Usually, the difficulty of speech recognition increases as the vocabulary gets larger or the words have similar phonetics. In the case of word sequences, language models or artificial grammar are used in order to limit the number of possible word combinations.

The simplest language model can be defined as a finite network, where the words that can be used after a specific word are explicitly defined. More general models, that can approximate natural speech, use context-dependent grammar.

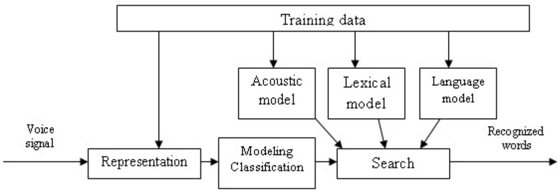

Fig. 1, Speech recognition process

The leading standard in voice recognition is Hidden Markov Models (HMM). HMM is a stochastic model, in which the generation of phoneme strings is represented probabilistically as Markov processes. HMM is used together with neural networks (used to determine the sound-word correspondence probabilities) in so-called hybrid systems.

2.3 Language modeling

The error rate of any given speech recognition system is smaller than the percentage of words not included in its vocabulary. Thus, the most important part of building a language model is selecting a vocabulary that will cover as much as possible the words that need to be recognized.

Some of the newest types of language models are class-based models and dynamic models.

Class-based models use word classes instead of words. Classes are based on common properties of certain words, properties like morphological analysis or semantic information.

Dynamic models use the recurrence of a word based on the history of its appearances during use. Dynamic language models can be built using a list of recently recognized words, as well as estimating the probability of appearance for related words.

3 Data Link

CPDL (**C**ontroller **P**ilot **D**ata **L**ink) was developed as an upgrade of the classic ATC-pilot communication systems.

Normally, ATC-pilot communication uses VHF or HF voice radios, depending on the distance. The main drawback of this system is that all pilots in one sector use the same frequency. As the number of flights increases, so does the number of users of one given frequency and the probability of simultaneous transmissions, which require the repeating of messages. Furthermore, each message requires a specific amount of time for its transmission. All of these aspects lead to a point of saturation, beyond which the frequency is so congested that communication becomes impossible.

Channel congestion is usually handled by the division of the congested sector into two or more smaller sectors, each with its own frequency. This solution poses its own problems, like the increase of workload for pilots and traffic controllers.

CPDL uses a data link for ATC-pilot communication. This includes a set of pre-defined request/approval/information messages, in accordance with ATC phraseology. The air traffic controller can assign flight levels, airspeeds, radio frequencies, approve or request route changes and request information. The pilot can respond to messages, request approvals and information, report information, declare or cancel a state of emergency. Furthermore, the pilot can request approvals and information from a traffic controller further ahead on its flight plan.

There is a possibility for sending text messages, when the message exceeds the pre-defined set as well as a possibility for sending information from one ATC station to another using CPDL.

The main functions of CPDL are:

- exchange of messages between traffic controller and pilot;
- transfer of authority from one controller to another;
- receipt of approvals from controllers further away on the route.

All CPDL implementations have to be based on a safety study that shows the achievement of all safety-related objectives. For continental airspace, the performance and safety requirements are given by EUROCAE ED-120 (RTCA DO-290).

Among the safety objectives is the need for the messages not to be corrupted and to be delivered to the intended recipient. Other important aspects are the correct timestamp and the rejection of expired messages. This leads to the necessity of very precise clocks both onboard aircraft and on ground. One solution is the use of the GPS system clock.

Currently, there are two implementations of CPDL:

- FANS-1/A, developed by Boeing, used on trans-oceanic routes. FANS-1/A is based on the ACARS (Aircraft Communication Addressing and Reporting System) system and uses Inmarsat Data-2 (Classic Aero) satellite communication.
- ATN/CPDLC, extended through the Link2000+ program in the EU countries. The system uses VHF Digital Link (VDL) Mode 2 networks, operated by ARINC and SITA.

3.1 LINK2000+

Link2000+ is a set of CPDL communication services, implemented in the European airspace using ATN (Aeronautical Telecommunication Network) and VDL (VHF Data Link) Mode 2 networks.

The system has three main functions, used to automate routine operations that take up to 50% of air traffic controllers' time:

- ATC Communications Management (ACM), used to manage radio frequencies;
- ATC Clearances (ACL), used for standard clearances (such as, for example, "Climb to level 350");
- ATC Microphone Check (AMC), which provides a backup communication solution in case of standard communication failure.

These services do not substitute voice communication; they provide an alternate solution in case of failure, increasing safety.

4 Speech recognition implementation

The block diagram of the proposed solution is presented in Figure 2.

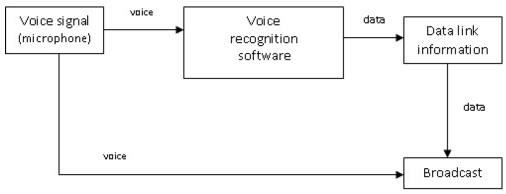

Fig. 2 Block diagram of speech recognition implementation

When the PTT (Push To Talk) button is pressed by the pilot or air traffic controller, the system begins recording the voice signal from the microphone and initiates speech recognition. Once the button is released, the recording stops and the result of the speech recognition process is broadcasted, in digital format, using the data link.

Because speech recognition is performed at the source, the system can be trained specifically for each user, giving it high accuracy. The high accuracy is also a benefit of the use of standard phraseology, because it allows the use of more efficient predictive recognition algorithms and requires a shorter training time for the system. Nevertheless, the use of non-standard phraseology can dramatically reduce the accuracy of the system and can even lead to confusing situations.

The system can be fitted with an interface for uploading pre-defined language models tailored for each crew member, for example through the integration of a removable memory module (memory card, USB slot). The system can therefore easily adapt to any user change.

Each crew member station or ATC station must also have its own customized noise model (microphone noise, ambient noise, interference etc.), that the speech recognition algorithm can use in order to reduce the impact of noise on the accuracy.

Speech recognition can be implemented in pilot-ATC communication in two ways: text broadcast or a voice control interface to an existing system, in this case Link2000+. Both implementations allow for voice and text transmission, providing a backup in case of malfunction.

4.1 Text broadcasting

In this case, the speech recognition encodes the recognized speech in ASCII format and broadcasts it to the ATC center and/or all aircraft in the sector using a data link. Before broadcasting, the message is given a unique identifier (such as the S-mode transponder address, aircraft callsign or ATC station ID), so that the source of the message can be identified. The text is displayed in the same format both at the source and at the receivers.

For example, the message “Speedbird 09, request climb to flight level 130” will be displayed to all receivers, and also to the senders, in a format similar to:

Speedbird 09: Speedbird 09, request climb to flight level 130.

The main benefit of this solution is that it allows the monitoring of all radio communications in that sector so that, in the event of a misunderstanding, the text message can be used to solve any problem that may arise.

A recording function can also be implemented in order to log all messages for the duration of the flight, making them accessible in case of need. Even more, the recorded messages can offer clues in order to facilitate investigations in case of incidents or accidents. Because the system uses digital format, the memory modules can be smaller, easier to protect from heat, impact or water.

Another benefit is that air traffic controllers can review instructions sent to aircraft in their sector, their responses and also the source of a given message, allowing for quick response to conflicting messages or instructions.

Other problems that can be solved through the implementation of text broadcasting are:

- Misunderstanding of messages (usually, confusion regarding numbers);
- Language issues;
- Message callback (time consuming and workload increasing).

In the event that the voice message is misunderstood (partially or totally), the text offers the same information in an easier to understand form.

For safety reasons, a text-messaging function is implemented, so that communication can continue in case of failure of standard voice systems. In the event of microphone or radio failure, messages can still be sent using a keyboard (for ATC) or a scratchpad (for pilots).

The data link for this system is VDL Mode 4, which allows broadcasting messages.

This implementation requires the installation of additional equipment onboard aircraft (screens for message display and one or more control panels).

4.2 Voice control interface for LINK2000+

This implementation uses the results of speech recognition to control the Link2000+ system.

Link2000+ is usually controlled manually both by ATC using a keyboard and mouse and by pilots, using the FMS. This makes controlling the system a very workload-inducing task, especially for the pilots who have to follow specific routines in order to send specific messages, each time having to press buttons, enter text using the FMS keyboard and other time-consuming tasks.

By using a voice control interface, these workload-increasing tasks can be replaced by the more intuitive and familiar voice messages. Because standard ICAO phraseology and the Link2000+ messages are very similar, the language model for the system is easy to build and the accuracy of the speech recognition process is very high.

The main advantage is the time and workload economy, time that can be used by the crew and ATC for other important tasks.

Another advantage is an easier way to control the Link2000+ system, the only remaining manual task to be done being the selection of the ATC center to connect to.

Similar to the text broadcasting system, this one has a recording system, allowing the review of previous messages, if required.

This implementation requires the development of an interface between the speech recognition system and the Link2000+ system. This interface must select the proper message type (request for divert, information, clearances, instructions and so on) depending on the recognized words. This selection system is similar to the one used in context-dependant speech recognition and can be embedded in the speech recognition system itself.

Because this solution does not require the installation of major equipment onboard aircraft and the Link2000+ will be mandatory for all EU countries by 2016, it can be implemented easier than the text broadcasting system.

5 Conclusion

Communication errors are a key factor in flight safety, most of the times being one of the triggering (even though sometimes not the most important) factors in accidents and incidents.

The key benefits of the implementation of systems designed to reduce communication errors are the increase in flight safety, which in turn leads to increasing confidence in air transport.

Another benefit of speech recognition integration is the cost reduction of ATC services by eliminating the need for sector division.

6 Contact Author Email Address

claudiugeacar@gmail.com

7 Acknowledgement

Grateful acknowledgement is made for financial support by Bucharest Politehnica University through POSDRU Grant 7713.

Copyright Statement