

Speech Systems for Autonomous Unmanned Aircraft

Enabling autonomous unmanned aircraft to communicate in civil airspace

Chris R. Burger
CSIR Meraka Institute
P O Box 395, 0001 Pretoria,
South Africa
crburger@csir.co.za

Etienne Barnard
Multilingual Speech Technologies
Northwest University
1900 Vanderbijlpark
South Africa
etienne.barnard@gmail.com

Thomas Jones
Department of Electrical and
Electronic Engineering
Stellenbosch University
Private Bag X1, 7602 Matieland,
South Africa
jones@sun.ac.za

Abstract— Airspace control is currently based largely on the exchange of speech between aircraft and Air Traffic Service Units, or between aircraft themselves. ICAO regulatory guidelines make no distinction between unmanned and manned aircraft, implying that unmanned aircraft will have to comply with requirements for radio communication in certain airspaces. The availability of speech capability is therefore imperative for autonomous operations in civil airspace. The paper assesses the feasibility of automated speech in unmanned aircraft given the current state of the art.

Keywords- *unmanned aircraft; speech systems, air traffic control.*

I. INTRODUCTION

Unmanned aircraft (UA) operations are attractive for a number of reasons, including cost, expendability and sustained vigilance in long-duration missions. UA operations can therefore be expected to become ubiquitous over the next few years.

Current UA operations are normally conducted in restricted airspace, and subject to a pre-approval process. Practical applications are hampered by such restrictions. The goal should be to achieve unmanned operations subject only to prior notification, such as a flight plan.

The International Civil Aviation Organisation (ICAO) regulatory guidelines make no distinction between manned and unmanned aircraft. Any UA operations are therefore subject to existing traffic separation arrangements. The primary separation mechanism is to see and avoid traffic, supported by an airspace system that provides lateral, vertical or temporal separation. The airspace system makes extensive use of clearances and traffic information transmitted via voice channels in the VHF band.

In this paper, the focus is mostly on autonomous aircraft. Unlike remote-controlled aircraft, which have a ground-based human pilot, autonomous aircraft function without direct human intervention. In practice, most autonomous aircraft would have at least some intervention capability to address unforeseen malfunctions, but routine operations would take place unassisted.

Remote-controlled UAs use a remote voice link, through which the remote pilot exchanges radio messages much like

an on-board pilot would. In contrast, autonomous aircraft must have on-board communications capability. This capability must include both synthesis and speech recognition to support two-way communications.

The paper proposes the implementation of a speech system for autonomous aircraft. It defines a number of sub-tasks which are required to interface existing speech building blocks with the other subsystems in a UA. Finally, a test protocol is described. This protocol involves implementation on ground-based platforms, simulating flight activity in a controlled airspace with the cooperation of the controlling Air Traffic Service Unit. The test protocol will both identify problem areas so that they can be rectified and provide reliability statistics to evaluate the reliability of the system. The reliability statistics can form the basis of a safety case on which new regulations and certification requirements for automated voice systems can be based.

The tasks proposed for a full development project are:

- Define the speech tasks in a range of operational situations, including different flight rules and airspace classes.
- Define interfacing with other UA subsystems, including the autopilot and the navigation system.
- Develop speech recognition and synthesis subsystems that can handle the tasks defined above.
- Implement the subsystem on a platform with suitable dimensions, mass, processing capacity and current consumption for on-board use.
- Evaluate the performance of the system against a range of targets, from baseline capabilities to more advanced real-life scenarios, and compare this performance with that of a human pilot.
- Implement a virtual aircraft that can conduct simulated flights through a controlled airspace, to evaluate its ability to integrate into civil airspace and to identify and rectify any possible shortcomings.

The paper also provides an estimate of the development effort required to implement the system to the point where a safety case can be constructed.

II. CURRENT PRACTICE IN UA OPERATIONS

Currently, virtually all Unmanned Aircraft (UA) operations are remote-controlled through radio links. UAs operate under two dispensations¹:

1. Small models can be operated under regulations intended for recreational flying, subject to constraints on size, mass, speed and line of sight operations.
2. Larger models can be operated with a Special Flight Permit (SFP) issued on a case-by-case basis by the Civil Aviation Authority (CAA) for every operation. SFPs are subject to advance notice and a well-developed case-specific safety model, and are most often only issued to government departments with a proven need. For law enforcement and surveillance missions, the prior notification and the publication of all permits (through the Notam system) removes any possible element of surprise.

Regulations pertaining to aircraft operations place great emphasis on the responsibilities of the Pilot in Command (PIC). Many examples are found in CAR91¹ Subparts 1 and 2. These regulations generally do not distinguish between an on-board pilot and a remote-control pilot. However, no provision is made for an aircraft that does not have a PIC. Suitable regulations will have to be formulated, based on a considered decision on a suitable liability regimen.

Autonomous aircraft have advantages compared to remote-controlled aircraft. Examples of these advantages include:

- **Geographic limits:** Autonomous aircraft can respond to requirements beyond line of the sight of the origin without long lead times (as required to reposition a ground control station). Some of these constraints can be relieved through long-distance links such as satellites, but such operations are outside the realm of current regulations.
- **Costs:** Pilots are expensive and involve considerable logistics, human resource management and supply planning.
- **Endurance:** Humans have a limited attention span. Long endurance missions are not well suited to human operators, including pilots. Some applications, such as ubiquitous surveillance, require endurance of many days.

III. AIRSPACE ARRANGEMENTS

Internationally, airspaces are divided into seven classes, labeled A to G. A is the most restrictive, and G the least restrictive².

Airspaces are distinguished by flight rules and by level of control.

Flight rules could be Visual Flight Rules (VFR) and Instrument Flight Rules (IFR). Typical local surveillance missions are done under VFR, but almost all passenger-carrying operations are done under IFR. Because many urban surveillance missions take place in close proximity to airline hubs, IFR capability is required. Rural and peri-urban operations require VFR capability.

The levels of control are:

- **Controlled airspace:** In controlled airspace, a controller provides binding instructions to ensure that traffic separation is maintained. In some cases, the controller also assumes responsibility for terrain separation. Often, primary or secondary radar is used to monitor separation between traffic.
- **Advisory airspace:** Although there are differences in the services provided and the way in which traffic separation is provided, this model is identical to controlled airspace for the purposes of this discussion. Advisory airspace is not used in all countries.
- **Information airspace:** Air Traffic Services Units provide information about weather and other traffic to the extent that they are known, but no separation from other traffic or terrain is provided.

South Africa uses only four of the seven ICAO airspace classes. Class A airspace is controlled and IFR only. Class C is controlled, both VFR and IFR. Classes F and G are information airspace, both VFR and IFR.

The generic term describing a facility that provides services to pilots is an Air Traffic Services Unit (ATSU). A specific type of ATSU is Air Traffic Control (ATC).

Also of importance to our discussion is the concept of Traffic Information Broadcast to Aircraft (TIBA). In certain information airspaces, pilots maintain traffic separation by broadcasting their positions and intentions and listening out for other pilots doing the same. The broadcast is structured in such a way that unaffected pilots can determine early in the transmission that they are unaffected and do not need to listen to the entire transmission.

Controlled airspaces therefore require communications only with a single entity (the ATSU), while information airspace may require either a similar capability for interfacing with an ATSU or the ability to broadcast and to interpret the transmissions of many other aircraft.

IV. ABILITIES REQUIRED OF THE SYSTEM

The capabilities required of a speech system differ slightly depending on whether the aircraft is operating in communication with an ATSU or under TIBA conditions. The required abilities for each situation are therefore described separately below³.

1. Tasks when communicating with an ATSU:
 - a. **Establish communications**, using a standardised protocol.
 - b. **Make position reports**. Under IFR, radio beacons and formally-defined intersections with five-letter names are used. Under VFR, normal prominent landmarks are used.
 - c. **State requests**, such as a request for taxi instructions on the airport, takeoff clearance, clearance to enter an airspace, landing clearance or information.
 - d. **Understand** responses and instructions from the ATSU.

- e. **Read back clearances**, normally more or less verbatim.
 - f. **Make emergency transmissions:** If the aircraft is in trouble, the system must be able to advise others of the situation and solicit help via radio.
2. Tasks when operating in Peer-to-Peer (TIBA) conditions:
- a. **Make traffic broadcasts**, including position reports and intentions.
 - b. **Understand** transmissions from other pilots.
 - c. **Interpret** the geographic position and intentions of the other aircraft, using a database of known landmarks.
 - d. **Make emergency transmissions:** If the aircraft is in trouble, the system must be able to advise others of the situation and solicit help via radio. If reduced separation is experienced, the aircraft must also be able to advise conflicting traffic of the threat.

3. Performance requirements

An outsider may assume that error-free operation is required. Such performance levels are not necessary in practice. Aviation systems are designed with the knowledge that human error occurs frequently, and includes multiple redundancy to alleviate the effects of such errors.

Human pilots do not operate at a zero error rate, and frequent requests for clarification or correction are exchanged. The goal when designing an aircraft speech system therefore only needs to be to match human performance levels.

The readback mechanism mentioned before provides a means of correcting misunderstandings before actions are taken based on erroneous information.

Readback is not dissimilar from strategies employed by commercial speech systems, where the user is questioned to confirm or repudiate the system's understanding from previous speech recognition efforts.

The John A. Volpe National Transportation Systems Center has done much work on measuring error rates in pilot-ATSU communications. The work is funded by the US Federal Aviation Administration (FAA). Typical error rates are between 1 and 3%⁴, with one-quarter to two-thirds of all readback errors being caught^{4,5}. With typical movement rates in busy airspaces, these figures translate into well over one communications error per hour on each frequency.

This work will have to be extended to provide a local measure of error rate to provide a benchmark that can be used as a design and acceptance standard for a system tailored to a local environment. The methodologies are well established and could be applied to existing recordings of different ATSUs to establish target error rates for different airspaces and environments.

V. SUITABILITY OF EXISTING SPEECH TECHNOLOGIES

No evidence of existing use of automatic speech in this application has been found. However, other applications exist in aviation:

- o **Voice control of aircraft systems:** Advanced aircraft use voice technology to allow pilots to control aircraft systems. Known developers of such systems are QinetiQ⁶ in the UK and SRI International⁷ in the USA.
- o **Flight simulators with speech capability:** Recreational simulators such as Microsoft's FlightSim X provide speech interfaces to simulate interactions in the cockpit and ATC communications. Other companies provide add-ons to improve capabilities or realism. Examples include Cockpit Chatter⁸ by Flight1 Software and VoiceBuddy⁹ by eDimensional Software.
- o **Training of pilots and air traffic controllers:** Several vendors supply building blocks to facilitate the training of personnel to adhere to predefined voice protocols. Examples include DynEd's training systems¹⁰ and Adacel's ATC simulators¹¹.
- o **Control of UAs:** Like with manned aircraft, UA pilots are assisted by voice systems. Examples include Adacel's use of SRI International technology¹² and Sytronic's MAGE¹³, which facilitates operation of several UAs together.

In general, Text-to-Speech System (TTS) technology is well-established¹⁴. Even consumer-market synthesisers offer adequate performance when used with a well-defined and limited vocabulary. Although existing systems are criticised for monotony in extended speech, the short transmissions in aviation are mainly judged by intelligibility, and are short enough not to pose any challenge in this regard.

TTS systems also do not do well with unknown vocabulary¹⁴. However, in aviation, a comprehensive lexicon with predefined pronunciation can certainly be compiled.

The emphasis in the following system will therefore be on Automatic Speech Recognition (ASR) systems.

Factors that improve ASR performance include¹⁴:

- o A constrained vocabulary.
- o Fixed syntax.
- o Fixed channel characteristics.
- o Speaker-specific training.
- o Low noise.

The first two factors are typical of aviation radio telephony, especially when participants adhere to standard phraseology^{15,16}. The standard phrase guidelines published by ICAO and adhered to by most member nations contain hundreds of words and phrases, as opposed to the tens of thousands of words in typical ASR systems.

Channel characteristics can be calibrated for a specific aircraft communicating with a specific ATSU. Speaker-specific training is also feasible with a single ATSU. Although the use of a single ATSU is restrictive, it is not

altogether useless. UAs on local surveillance duties may only interact with one or two ATSU's in their sphere of activity. As a first step, consideration can definitely be given to speaker-specific training using a calibrated, known channel. Such constraints can considerably enhance the probability of accurate communications.

A number of South African institutions are active in the field of human language technologies. The CSIR's Meraka Institute has an active Human Language Technology Group. Universities known to be active in the field include the University of Pretoria and both Stellenbosch¹⁷ and Northwest Universities.

VI. AN IMPLEMENTATION STRATEGY

The following tasks and sub-tasks have been identified as a means to implementing a UA speech system:

- Design the speech tasks in text form
- Design the architecture
- Develop TTS and ASR systems on a suitable platform
- Evaluate performance
- Build a safety case

Each of these tasks is described in somewhat more detail below.

1. Design the speech tasks in text form

ICAO guidance documents^{15,16} provide relatively detailed guidance on terminology and phrases to be used. For a given environment, some phrases may prove to be superfluous, while other phrases and vocabulary may have to be added to take account of local airspace geometries and landmarks and other operational requirements.

The text-form speech tasks will have to be broken down into standard and non-standard vocabulary. An obvious example of aviation speech that does not conform to common usage is the transmission of numbers, where both "five" and "nine" have been modified to make them less similar (to "fife" and "niner" respectively), and where the decimal separator is neither "comma" nor "point", but "decimal".

2. Design the architecture

The speech system will have to interface with the autopilot, the navigation system, the navigation database (for landmark recognition) and with the communications equipment (for the audio link to and from the ground). At least the following tasks will have to be defined to a sufficient level of detail to allow the aircraft to perform its tasks:

Operational

- Select frequencies appropriate to the route being flown and the airspaces being penetrated.

Receiving

- Extract sufficient information from incoming speech to recognise landmarks, instructions,

altitudes and actions, as well as traffic information.

- Interpret instructions and traffic information to determine a route vector for other traffic and determine likely conflicts. For this purpose, interfacing with the navigation subsystem will be required to facilitate identification of landmarks.
- Provide information to the autopilot to execute routine remarks and possible avoiding action.
- Revise the existing model of reality when readback errors are pointed out by an ATSU. This step would involve understanding corrections transmitted in response to an incorrect readback, constructing an alternative model of what needs to be accomplished, comparing that model to the previous model constructed during the original clearance and then modifying the plans accordingly.

Transmitting

- Transmit routine position reports without interfering with other users of the same frequency.
- Transmit timeous requests to facilitate airspace penetration, takeoff, landing etc.
- Transmit warnings associated with breaches of separation.
- Read back all clearances.
- Acknowledge all received information.
- Declare an emergency, using prescribed phrases, if the aircraft encounters an abnormal situation.

3. Develop TTS and ASR systems on a suitable platform

As an initial design goal, a platform occupying 1 dm³ (with a cubic form factor), drawing 1 A at a supply voltage of approximately 14 V and having a mass of 1 kg has been proposed. Such platforms are compatible with the capabilities of even small UAs with a wing span of less than 3,5 m, and are available with sufficient processing power to accommodate commercial TTS and ASR systems.

Considerable bespoke development will be required to ensure that the reliability levels are high enough for the requirements of aviation. As mentioned in a previous section, considerable advantage can be gained from initially operating in a known airspace with known channel characteristics, and even with a limited number of known voices rather than trying to interpret a generic voice. If this strategy is chosen, a procedure must also be developed to enrol new staff at the relevant ATSU's and to obtain sufficient samples of their speech with known content to allow the ASR system to be trained. The procedure could be as simple as asking them to read a pre-defined text into a microphone.

4. Evaluate performance

Several key components have to be completed before performance can be fully evaluated.

- a. **Establish a benchmark for satisfactory performance.** Although an American benchmark is available in the form of a series of studies by the Volpe Center^{4,5}, a local benchmark must be derived using local data. Deviations can be expected from the US data because of South Africa's multilingual environment relative to New England's relative homogeneity, as well as the fact that their studies appeared to concentrate on busy airspaces where there is a very high incidence of professional pilots. In the South African context, there will be a requirement for urban surveillance in relatively quiet airspace, with a high incidence of much less experienced pilots. While the presence of less-experienced pilots is likely to adversely affect the error rate, the lower speech rates and lower workload for both pilots and controllers should serve to improve error rates.
- b. **Evaluate the system's performance under laboratory conditions.** Evaluation would start against large quantities of recorded ATC communications and TIBA. Unfortunately, the gathering of such data has become much harder in South Africa with the August 2011 conviction of plane spotter Julian Swift¹⁸ for "intercepting communication" while practicing his hobby at O R Tambo International Airport. The legal situation will have to be clarified before serious recording of such airband communication can commence.
- c. **Evaluate the system's performance in flight.** Although the laboratory evaluation will have provided a reasonable level of confidence by the time flight testing is introduced, strict measures will have to be taken to ensure that dangerous or obstructive situations are not created. The proposal is to operate the system from the CSIR campus in eastern Pretoria, virtually flying routes within the surrounding controlled airspace.

The cooperation of the two resident ATC units will be required, along with their parent bodies. As one airspace is military and the other civilian, the cooperation of Air Traffic and Navigation Services (ATNS) and the South African Air Force (SAAF) will also be required. Representatives of both organisations have indicated in informal discussions that they would favourably consider such an application.

During the trial, any deviations from safe practice will have to be logged for further investigation and fault finding. This phase could provide not only opportunities for thorough fault finding but also a statistical basis that can be used for the final phase.

It is also possible that the performance of the system is too good, impacting negatively on safety. Specifically, if the performance of the TTS is so good that human pilots and ATSU personnel have

trouble distinguishing between it and other human pilots, the UA may be confronted with non-standard phraseology that may erode its accuracy to the point of becoming a safety hazard. In this case, a means of identifying UAs will have to be introduced to ensure that ATSUs and pilots comply with standard phraseology when communicating with the UA. A precedent exists, as student pilots are already required to prefix their callsigns with the word "Student" to ensure that they are not bombarded with hard-to-interpret instructions, possibly causing them to become flustered.

5. Develop a safety case

Using the reliability statistics derived from the previous phase, a safety case can be constructed to prove the ability of autonomous UAs to operate safely in controlled or information airspace. The safety case may initially be confined to a specific situation (such as the Pretoria area), but can form the basis for approval of autonomous operations and a suitable regulatory framework on which future implementations can be based.

VII. SCOPE OF WORK

The second author's research group has estimated the effort required to develop the system as follows (in man-days):

Table 1: Estimated resources required to complete the project as proposed.

Phase	Junior staff	Senior staff	Total
Task definition	25	7	32
Protocol definition	20	16	36
ASR development	350	40	390
TTS development	200	40	240
Hardware	180	30	210
Evaluation	40	35	75
Project overhead		10	10
Total	815	178	993

Total project duration was estimated at 24 months from commencement of the work.

The estimates do not include provision for the establishment of the test criteria against which the system will be qualified or for the compilation of the safety case and subsequent regulation effort.

VIII. FUTURE WORK

With the necessary funding, the following opportunities are known to exist:

- o Implement a system to support Government's stated intention to facilitate UA operations to meet the requirements of various government departments, using the strategy outlined above.

- Pursue the creation of suitable regulations to facilitate the operation of UAs in South African civil airspace.
- Investigate the implementation of a computer-based speech training system to support current pilot and controller training syllabi. The proposed system could also provide a very valuable objective assessment tool in the context of ICAO-mandated Language Proficiency Rating testing.
- Implement a generic framework within which other authorities could evaluate and regulate similar UA activities.

IX. CONCLUSIONS

The work being proposed is regarded as an essential component of future autonomous aircraft operations. No similar systems are known to exist. The absence of speech capabilities for unmanned aircraft can hamper the regulation and operation of autonomous aircraft once their technology becomes mature enough for useful civil operations.

References

- [1] *South African Civil Aviation Regulations*, Parts 61 (Flight Crew Licencing), 91 (Operating and Flight Rules) and 121 (Commercial Operations of Large Aeroplanes), accessible at <http://www.caa.co.za>.
- [2] *Aeronautical Information Manual, Chapter 3*, US FAA, available at http://www.faa.gov/air_traffic/publications/atpubs/aim
- [3] *CAP413 Radiotelephony Manual, Edition 20 (November 2011)*, a Civil Air Publication of the UK Civil Aviation Authority, available at <http://www.caa.co.uk>.
- [4] Cardosi, K., Falzarano, P., and S. Han, (1998) *Pilot-Controller Communication Errors: An Analysis of Aviation Safety Reporting System (ASRS) Reports*. DOT/FAA/AR-98/17.
- [5] Cardosi, Kim M., *Human Factors for Air Traffic Control Specialists: A User's Manual for your Brain*, November 1999, US DOT/FAA/AR-99/39.
- [6] http://www.qinetiq.com/home/newsroom/news_releases_homepage/2007/2nd_quarter/QinetiQ_speech_recognition_technology_allows_voice_control_of_aircraft_systems.html
- [7] <http://www.aviation.com/technology/071016-afpn-f-35-voice-recognition.html>
- [8] <http://www.flight1.com>
- [9] <http://www.VoiceBuddy.eDimensional.com>
- [10] <http://www.dyned.com/products/ae>
- [11] http://www.adacel.com/solutions_services/air_traffic_control.htm
- [12] http://www.adacel.com/solutions_services/speech_unmanned_air_vehicles.htm
- [13] <http://www.sytronics.com/intell/programs/mage.html>
- [14] Huang X D, A Acero, Hon HW: *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*, Prentice-Hall, ISBN 978-0130226167.
- [15] *Aeronautical Telecommunications Volume II (Communication Procedures including those with PANS Status), Chapter 5: Aeronautical Mobile Service—Voice Communications*, ICAO Annex 10.
- [16] *Manual of Radiotelephony*, Doc 9432-AN/925, ICAO
- [17] Scholtz, Pieter E., Justus C. Roux, Jacques P. du Toit, "Speech Synthesis in the Mobile User Interface", Proceedings of the 7th ISCA Tutorial and Research Workshop on Speech Synthesis (SSW7), 2010, pp. 328-331.
- [18] <http://www.iol.co.za/news/crime-courts/swift-found-guilty-1.1112980>