"Don't drop the plane to fly the mic!" – Designing for Modern Radiotelephony Education in General Aviation

Peter Heinrich[1], Dagmar Hollerer[2], Christoph Karthaus[3] and Mario Gellrich[4]

Abstract: Learning to fly an aeroplane is a challenging endeavour; mastering the radio, i.e. talking to air traffic control (ATC) is considered by many students as an especially daunting activity. Using the correct voice procedures and phraseology is paramount for a smooth and safe aircraft operation in busy airspaces. However, training new general aviation pilots in radiotelephony is still a largely manual activity, where students and teachers engage in role-play to mimic the different ends of communication. The design of an aeronautical chatbot, capable of simulating ATC would provide students with plenty of additional training opportunities that they can conduct on their own and at their own pace. A detailed analysis of the voice input would provide feedback on the learning process. This paper reports on the preliminary proof-of-feasibility prototype design as well as synthetic language model training data generation.

Keywords: Aviation English, Computer Based Language Learning, Automatic Speech Recognition

1   Introduction

Pilot: "Tower, please call me a fuel truck." – Tower: "Roger! Sir, you are a fuel truck." [LU23]. Obviously, this is a fun joke, but it highlights the general problem of ambiguous communication with the aim to establish shared mental models. To avoid such ambiguities in aviation, a special (restricted) variant of the English language is used for communication over the radio specified by the International Civil Aviation Organization (ICAO). Although it is based on English, learning this language is comparable to learning a new language altogether as it has its own, restricted vocabulary, some special words as well as a precise grammar and order of information items given in each statement. Today, flight schools offer courses in radiotelephony, or RTF for short. These courses are synchronous interactions between a teacher and flight students that pretend to be flying. They can be held on premises or distant via some form of remote audio/video call. However, gaining the required practice outside of this training environment is difficult as no communication partner is available. Therefore, an overarching research question to answer would be: "*How can a radiotelephony learning environment for GA (student) pilots be designed to be an effective learning aid for building and maintaining RTF proficiency?*". We answer

---
[1] ZHAW School of Management and Law, Theaterstrasse 17, Winterthur, 8401, Switzerland, heip@zhaw.ch
[2] fly-T GmbH, Rotackerstrasse 15, Jona SG, 8645, Switzerland, hello@fly-T.ch
[3] fly-T GmbH, Rotackerstrasse 15, Jona SG, 8645, Switzerland, hello@fly-T.ch
[4] ZHAW School of Management and Law, Theaterstrasse 17, Winterthur, 8401, Switzerland, gell@zhaw.ch

this question by demonstrating an extensible prototype design and implementation to train one specific RTF procedure based on automatic speech recognition (ASR) with chatbot-like system interaction. These initial results are relevant to practices, first and foremost to companies building educational tools on radiotelephony and in the second line to private pilot students that prepare for the exams as well as for flight-schools and instructors as such technology can free them from repetitive tasks of re-training standard situation. Relevance for the scientific community is given by the identification and description of the research gap using literature review techniques as well as by the call for action and research agenda presented in the outlook.

2   Background

Correct mutual understanding between pilots and air traffic controllers is clearly a critical element of aircraft operation. The most prominent example probably is the 1977 Tenerife disaster that killed 583 people due to an uncleared take-off accompanied by a misunderstanding of the situation [WE90]. But even outside of controlled airspaces and airports, communication between pilots (e.g., local aerodrome traffic) needs to be concise and as short as possible as only one radio participant can talk at any given time. Official documents state that "The information and instructions transmitted are of vital importance in the safe and expeditious operation of aircraft" [IC07, p.2–1] and define the structure of communication from individual wording, pronunciation, and complete procedures (like departing and arriving at an airport). Although standard RTF voice is in English language, there are many special words and abbreviations like "WILLCO" (will comply) or "CAVOK" (ceiling and visibility ok), deliberate mispronunciation like "TREE" instead of "THREE" (because the "th" makes an unpleasant noise on the radio) or "NINER" instead of just "NINE" for the sake of clarity. Special formats for frequencies, time, altitudes etc. are specified. Additionally, although the rest is in English, local name (e.g.: cities, lakes and mountains) are pronounced in the local language.

Today, RTF is commonly trained in local flight schools in small group settings, sometimes using distance learning techniques but still most commonly in synchronous, on-site classroom settings. In Switzerland, there are 84 training organizations offering courses for obtaining a private pilot license (PPL for aeroplane or helicopter) and thus also on RTF training [FO23]. The Swiss Federal Office of Civil Aviation (FOCA) issued 431 light aircraft and private pilot licenses in 2022 [FE23]. Therefore, it is safe to assume that approximately 400 RTF exams are passed per year as well. Specific numbers are not publicly available but according to field experts, initial failure rates in RTF exams are in the range of 20 to 30% in Switzerland. This also raises the question on how private pilots maintain long-term RTF proficiency.

Thus, designing a virtual assistant based on ASR seems helpful for both initial training and for long term proficiency maintenance, but interestingly hardly any prior research on ATC simulation and only very few commercial virtual ATC products for private pilot

training are available. Some of these products are even not covering ICAO language procedures sufficiently. Plenty of related work can be found on the training of new ATC controllers (where pilots need to be simulated) and also on operational support of ATC in the tower (e.g., [HE21, HE16]). Thus, there seems to be a research gap in the whole area of supporting pilot education on RTF procedures. The training of air traffic controllers is even more involved and expensive. While a student pilot just needs a single, simulated air traffic controller, training a controller requires several "pseudo-pilots" simulating airplanes in the relevant airspace [JAR20]. More general, this type of training is called human-in-the-loop-simulations and first studies have been conducted on the question of the potential of virtual agents replacing human actors [SDB19]. However, the potential of ASR in ATC application is not restricted to training situations but could also be used to assist active controllers. Use cases for ASR in the field of ATC range from call-sign-highlighting on radar screen to pilot readback error detection. One example is the large scale HAAWAII-EU Horizon 2020 project, exploring the applicability of ASR in real world settings with the aforementioned application areas [HE21]. Another example is the MALLORCA project aiming for automatic digitization of analogue, verbal communication to increase controllers' performance by extracting commands from the transcripts to check for plausibility and to predict next commands [HE20].

3   An Architecture and Prototype Implementation for Low-Cost ASR Training Tools

3.1   Prototype Training Scenario and General Requirements

To prototype, we selected one procedure, namely a CTR crossing. A CTR is a controlled region of airspace where a pilot must ask for permission for entering. Typically, such CTRs are established around controlled airports to ensure separation between all planes operating in this region. For this specific operation, the pilot would call the tower to make initial contact, state the current plane's position, altitude, and the intended route through the CTR. The tower would then either reject the request, grant the request or state further preparatory tasks (like selecting a specific transponder code). Instead of using one specific, existing CTR we opted to abstract from the real world and provide an artificial airspace for the students built around a fictional station called "Lilyfield Tower".

```
Pilot: Lilyfield Tower, HB-ABC, C172, Robinhill, 5500 ft, Request
to Cross CTR direct E, 5500 ft.
ATC: HB-ABC, Lilyfield Tower, Crossing Approved direct E,
Report Entering
Pilot: Crossing Approved direct E, WILLCO, HB-ABC
Pilot: HB-ABC, 5500 ft, Entering CTR
ATC: HB-ABC, Maintain 5500 ft, Report Overhead
Pilot: Maintain 5500 ft, next Overhead, HB-ABC
```

Figure 1: Partial transcript of the prototype's CTR crossing scenario

Figure 1 depicts the initial call of the pilot and parts of the further communication transcript. This conversation follows the typical request → response → readback structure. The readback is especially important as it aids as a verification step to the controller that the pilot has correctly understood his or her commands. Further, all message elements must appear in order, the numbers and letters ought to be correct. Additionally, the pilot must answer timely (no more than a couple of seconds delay) and at appropriate speed (expressed by words per minute). Hence, the training goal is to enable the student pilot to engage in radio communication timely and according to the standard procedure. Although these procedures are rigid in structure, the controller can request additional things anytime that require appropriate reaction, again.
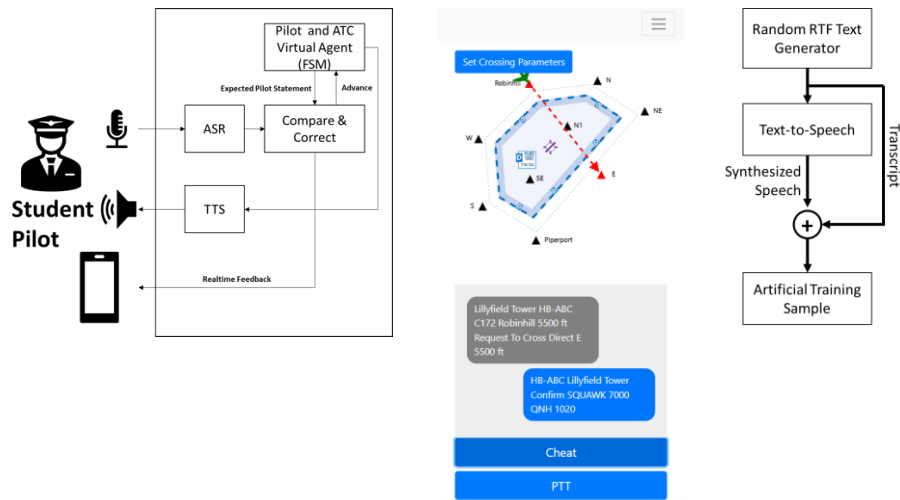
Figure 2: High level PoC backend system components (left) / Screenshot of prototype frontend (center) / Automated sample generation for custom langue model training flow (right)

3.2   Implementing a Pragmatic Conversational Learning Environment

As depicted in Figure 2 (left) we opted to simulate both Pilot and ATC using virtual agents (implemented as a finite state machine). A small, domain specific language expresses the dialog flow and its possible variations for both communication parties. As the XML-snippet (Listing 1) demonstrates, the structure of a message is encoded as a sequence of different message elements, expected in a specific radio transmission. Many of these "messages" together form the dialog of a complete procedure. Using a deterministic state machine, there always exists a predefined expected transmission from the student pilot. More specifically we applied an even stricter variant than the "Single Intent – Multi Turn"

[ZA20, S.203] chatbot interaction pattern that seems well suited for this kind of deterministic multi turn conversation. This strictness enables the system to give specific feedback on where the mistakes (missing elements, incorrect sequence, incorrect numbers or names) were made within a message statement. This 'shortcut' is possible as one single correct variant of a radio statement can be specified during pilot education – in contrast to practice where people sometimes deviate from the defined structures. Such situations are currently also not trained in traditional settings but leave options for later improvement.

```
<message role="pilot">
    <element type="Station"/>          <!-- Lilyfield Tower        -->
    <element type="CallSign"/>         <!-- HB-ABC                 -->
    <element type="Aircraft"/>         <!-- C172                   -->
    <element type="Location2D"/>       <!-- Robinhill              -->
    <element type="Altitude"/>         <!-- 4500 ft                -->
    <element type="Phrase" value="Request to Cross Direct" />
    <element type="Location2D"/>       <!-- E                      -->
    <element type="Altitude"/>         <!-- 4500 ft                -->
</message>
```

Listing 1: Procedure dialog excerpt matching the first statement listed in Figure 1

3.3   Generating Training Samples

In contrast to the deterministic nature of the dialog, automatic speech recognition needs supervised machine learning in order to achieve an acceptable recognition rate, given the number of deviations from standard English. The prototype uses Microsoft's Cognitive Speech services [MI23] for ASR and text-to-speech (TTS) synthesis. While language specialities can be implemented straight forward programmatically in TTS (encode "tree" vs. "three"), ASR needs training to create a custom speech model, tailored towards RTF dialect. However, large, annotated datasets of ATC communication are scarce and often also contain environment noises from cockpits or from the tower. From an ASR perspective however, these noises are not beneficial for language model training as students will typically train in quieter areas or in areas with noise sources different from cockpits. Such noises on the other hand could later be introduced in the TTS phase to make the training experience more realistic. Further, real-world data also include all miss-spellings and wording errors real pilots make. Therefore, special training datasets for ASR model training are required. Using the prototype system, we can create synthetic transcripts of random instances of the CTR crossing procedure. In this case, a random number generator sets all flight parameters (i.e. call sign, start, destination, crossing altitude) and by advancing the pilot-ATC state machine, a complete, syntactically and semantically valid transcript of a procedure is generated. The output of the automaton is plain text but by applying TTS synthesis, audio-data matching the transcript to 100% is generated. Using this approach, we synthesized 1400 text messages and generated the corresponding audio representation (.wav files; pilot, ATC using two different voices) to train a custom speech model (see Figure 2, right). This custom speech model is instantiated using Microsoft Cognitive Speech Services and then used for ASR within the prototype.

4   Outlook

Although a rigorous evaluation of the selected approach is still in planning, we could already demonstrate the feasibility of the concept with a first research prototype that implements a single procedure with a pragmatic approach. A next step would be to explore the effectiveness of different design options in real world settings with real student pilots as test subjects. Further research is additionally required on the synthetic learning dataset generation approach. Exploratory research using experimental techniques could be applied to determine the effects of varying the number of different synthetic voices, artificial noise addition and training data set size on ASR performance.

Bibliography

[FE23]   Federal Office of Civil Aviation FOCA: Statistics of Flight Personnel 2022, 2023

[FO23]   Federal Office of Civil Aviation FOCA: Training organisations (ATO/DTO): https://app02.bazl.admin.ch/web/bazl/en/#/facilities/search, accessed: 2023-03-30

[HE16]   Helmke, H. et.al.: Reducing controller workload with automatic speech recognition, In Proc. 35th Digital Avionics Systems Conference (DASC), 2016

[HE20]   Helmke, H. et.al.: Machine learning of air traffic controller command extraction models for speech recognition applications, In Proc. 39th Digital Avionics Systems Conference (DASC), 2020

[HE21]   Helmke, H. et.al.: Readback error detection by automatic speech recognition to increase ATM safety, In Proc. Fourteenth USA/Europe Air Traffic Management Research and Development Seminar, 2021

[IC07]   International Civil Aviation Organization ICAO: Manual of Radiotelephony, 2007

[JAR20]   Juričić, B.; Antulov-Fantulin, B.; Rogošić, T.: Project ATCOSIMA–Air traffic Control Simulations at the Faculty of Transport and Traffic Sciences, Engineering Power: Bulletin of the Croatian Academy of Engineering 2/20, 2-9, 2020

[LU23]   Lufthansa on Twitter: https://twitter.com/lufthansa/status/263668108944891904, accessed: 2023-04-01

[MI23]   Microsoft Speech Services: https://azure.microsoft.com/en-us/products/cognitive-services/speech-services/, accessed: 2023-06-17

[SDB19]   Schier, S.; Duensing, A.; Britze, S.: Sparring partners for human-in-the-loop simulations: The potential of virtual agents in air traffic simulations, CEAS Aeronautical Journal, 10, pp. 553–564, 2019

[WE90]   Weick, K.: The vulnerable system: An analysis of the Tenerife air disaster, Journal of management 3/1990, pp. 571-593, 1990

[ZA20]   Zamanirad, S. et.al.: State machine based human-bot conversation model and services, In Proc. 32nd International Conference on Advanced Information Systems Engineering, 2020