

An Automatic Training Tool for Air Traffic Control Training

Petr Stanislav, Luboš Šmídl, Jan Švec

Department of Cybernetics,
University of West Bohemia, Pilsen, Czech Republic
{pstanisl, smidl, honzas}@kky.zcu.cz

Abstract

In this paper we present an automatic training tool (ATT) for air traffic control officer (ATCO) trainees. It was developed using our cloud-based speech recognition and text-to-speech systems and allows dynamically generate the content. Our system significantly expands the available training materials, allowing ATCOs practice the basics of communication and phraseology. Furthermore, the automatic training tool is designed generally to be used for teaching in various areas, from specialized skills to a simple general knowledge.

Index Terms: speech recognition, human-computer interaction, computational paralinguistics

1. Introduction

Our training system is focusing on interactive training of phraseology, speaking and understanding of the air traffic communication. During the training of the ATCO is important to learn how to quickly understand the point of the air-to-ground utterance and react properly [1]. In most cases students have only books with rules and examples of the communication. Not until the later stages of training they can really practice on the simulator. The ATT significantly expands and improves the available learning materials and helping in early stages of study when the ATCO cannot practice on the simulator.

During the development, we were focusing on creating user-friendly and easy-to-use tool¹. For this reason, we decided to build it on web technologies, which among others, reduce compatibility issues and provides access to interactive learning materials anytime, anywhere.

Teaching takes place through simple interactions between the server and the student when the ATT ask the various questions and the user answers them (Fig. 1). A series of interactions focusing on a certain area (eg. *flight levels*, *communication frequencies*, *headings*, etc.) are part of the exercises. The exercises are then organized into thematic courses.

Questions are presented via text, prepared or generated audio (with our own text-to-speech framework), image or eventually video. The student can respond using pure text or speech input (using ASR), choose one or more options from prepared set of possibilities, or supplement the missing information into a text (Fig. 1).

The main benefit and advantage of the ATT is dynamically generated content of exercises and ability to extract semantic meaning from the answer and evaluate it. This allows us to provide an almost unlimited number of educational materials and supports the ability to quickly and accurately determine the nature of utterances and the right formulation of answers.

¹Examples are available at <https://youtu.be/GThdJRje7eA> and <https://youtu.be/1XygiZebb-s>.

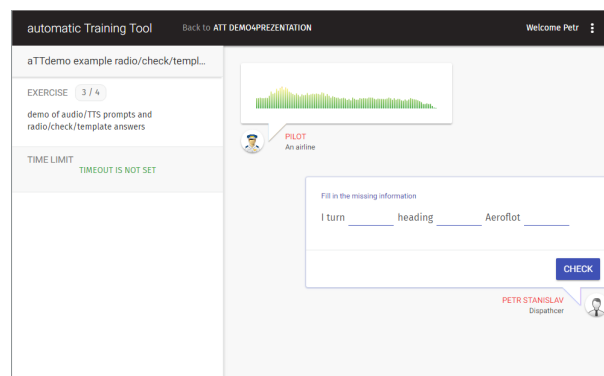


Figure 1: The graphical user interface presented during ATCO training.

The ATT also contains an interactive editor of content and a tool for evaluating the ATCO progress with various statistics related to the advance in an exercise.

2. Automatic Training Tool

The automatic training tool is a complete set of tools for teaching, content creation and evaluating. It uses ASR and TTS and can dynamically generate the content.

2.1. Training tool

Every user of the training tool has access to a set of courses, which combine exercises on the same topic. Passing the exercises can be done in two modes, namely teaching and testing. In the teaching mode is the results of evaluating revealed after the every answer. The user has instant insight into whether the answer is correct or where he made a mistake. If necessary, the lecturer can set the number of attempts to answer correctly before the correct answer is revealed. In contrast, in test mode, the user receives no feedback and the results are evaluated after completion of the entire exercise in the evaluation tool.

In both modes it is possible to set a timeout to complete the exercise. Validation of the answer is performed on the server and can work in two modes:

- Exact match: an answer must exactly match the expected answers (most suitable for interactions with options).
- Semantic match: an answer may not be exactly the same as the reference, but it must contain all semantically relevant information.

2.1.1. Semantic validation

The answer is sent to the server where the semantic entities are detected (e.g. *{callsign: CSA024, command: descent, flight level: 220}*) and on that basis the correctness of the answers is determined. The reference entities are defined by an expert. This significantly increases the impact of education, since it puts more emphasis on information contained in answers than on the formulation. It also enhances the usability of ASR in teaching.

2.2. Editor of learning scenarios

Even the best learning tool is worthless without quality content. For ease of creation of it serves interactive editor of learning scenarios. The main task is to create interaction in visual editor (Fig. 2). Here it is possible to easily create content of the exercises and immediately see and test what then see a student. Other features include creating and managing courses, semantic entities, multimedia files and so on.

2.3. Dynamically generated content

Any learning tool using a limited static content over time loses its teaching effectiveness. In such cases, the students often tend to remember the answers and begin to respond mechanically without thinking. To eliminate these influences, we created a mechanism for dynamic generation of exercises. This requires the exercise has information about where and how the content should be generated. Since most of the content (of interactions and answers) is defined by the text, we established for this purpose special keywords. These words begin with '\$' and are called flags and in most cases correspond to the defined semantic entities. These flags are replaced with specific values during generation.

Complete flag syntax looks like this *\$type.groups(value):id*, where *type* represents a type of flag (e.g. *CS* - *callsigns*, *FL* - *flight levels*, etc.). *Groups* are used to specify values, such as callsigns of only certain airlines or certain flight levels belonging to a group. The generated value may be specified by more than one group at a time (e.g. *\$FL.IFR.even*, *\$FL.odd.IFR*) and *value* is used for the choice of specific value. During generation, every flag is replaced with the specific value corresponding to the flag content. The value is applied each time the flag is used. It allows easy use of questions and answers. The problem arises when it is necessary to generate multiple values for the same flags. In this case, it is possible to specify the *ID* of a flag to distinguish one from the other.

Dynamic generation of a content significantly expands the possibilities of training tools, especially in connection with the semantic entities and TTS.

2.4. Automatic speech recognition

The ATT uses our own cloud-based ASR and TTS technology called SpeechCloud. It is based on LVCSR decoder with semantic entity detection [2] and uses SIP, WebRTC and WebSockets for communication and transferring the data. The ASR outputs both the best word hypothesis and the word lattice.

The ATT with SpeechCloud can seamlessly switch between different ASR models (e.g. generic, spelling, etc.) based on the current interaction.

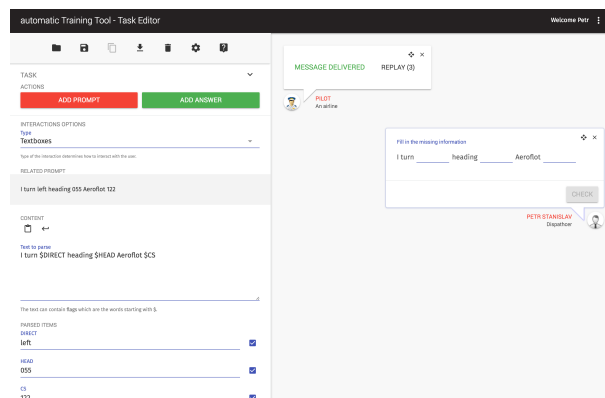


Figure 2: ATT editor of exercises.

2.5. Text to speech

We used TTS to support the main advantage of the ATT – the ability to automatically generate the new content. It uses a speech corpus of several voices created specifically for this domain. Source texts for the corpus were generated by an algorithm described in [3] from the air traffic control phraseology and the list of air carriers, waypoints and ICAO spelling alphabet. To support real conditions additional noises can be superimposed on to signal to simulate conditions in the real air-to-ground communication.

3. Discussion

The tool brings the greatest benefit in the early stages of ATCO training procedure because significantly expands the amount of available materials. It allows the student to interactively practice communication and phraseology in the phases when there is no simulator available. Another advantage is the possibility to learn anytime and anywhere.

Although the ATT is presented as the training tool for the ATCO is possible to use it as a general training tool, e.g. to learn languages, driving rules, etc.

Another interesting use of the training tool with ASR support could be learning of English pronunciation. The student will repeat sentences after the system and his aim is to speak so naturally that the utterances are correctly recognized by the ASR.

4. Acknowledgements

This work was supported by the project LO1506 of the Czech Ministry of Education, Youth and Sports.

5. References

- [1] R. E. Kidd, "Training air traffic controllers," *Marine Corps Gazette*, vol. 96, no. 3, p. 46, 2012.
- [2] J. Švec and L. Šmídl, "Semantic Entity Detection in the Spoken Air Traffic Control Data," in *Proceedings of 16th International Conference, SPECOM 2014*. Novi Sad, Serbia: Springer, 2014, pp. 394–401.
- [3] M. Jůzová and D. Tihelka, "Minimum text corpus selection for limited domain speech synthesis," *Text, Speech and Dialogue*, pp. 398–407, 2014.