Air Traffic Control Communication (ATCC) Speech Corpus

Luboš Šmídl and Pavel Ircing
Department of Cybernetics, Faculty of Applied Sciences, University of West Bohemia
Univerzitní 8, 306 14 Plzeň, Czech Republic
{smidl, ircing}@kky.zcu.cz

Abstract content

1. Introduction

This extended abstract introduces the motivation for creating and a basic overview of the structure of the speech corpus Air Traffic Control Communication (ATCC) that was designed, collected and annotated within the research project *Intelligent technologies for improving air traffic security (IT-BLP)*[1] and is available in the LINDAT-Clarin repository.

2. Motivation

Air Traffic Control (ATC) arguably constitutes the most critical part of the whole air traffic industry. Every civil (and, in many cases, also military) aircraft that enters the airspace of a particular country is immediately contacted by an air traffic controller (we will use just the term *controller* from now on) who communicates with the pilot(s) in order to prevent collisions, organize and expedite the flow of traffic, and also to provide further information and support. This communication takes place until landing or until leaving the country's airspace.

It is evident that the job of a controller is extremely demanding and requires (besides necessary personality traits such as the resistance to stress) an intensive training. This training is predominantly focused on teaching and reinforcing the communication skills of the aspiring controller. In order to minimize the probability of misunderstanding, the ATC authorities have specified detailed communication protocols and designed a special phraseology which aims to reduce the acoustic confusability between individual phrases. The observance of the protocol and phraseology is strictly enforced and it takes many hours for the controller in training to acquire it reliably.

The current state-of-the-art training procedure uses air traffic simulators that involve so-called *pseudopilots*. The pseudopilot is usually a retired pilot who prepares the training scenario and then acts as a pilot of several virtual planes, communicates with the controller (*trainee*) and enters the trainee's instructions into the software that simulates the plane movements on the radar screen.

The length of the controller training (approx. 2 years on average) and the relatively high salaries of the pseudopilots make the whole process very expensive. This sparked the idea of developing an automatic training simulator based on the intelligent spoken dialogue system. Such a system needs (besides other modules) a reliable automatic speech recognition engine that would recognize the trainee's utterances. Since the domain of the ATC task is very specific (see the details in Section 3.), we have decided not to rely on existing ASR speech corpora but we have designed and collected a new one, described in a nutshell in Section 4.

3. Specifics of the ATC communication

Non-native speakers

Given the international character of the air traffic industry, it is only natural that most English[2] utterances are pronounced by the pilots or controllers who are not the native speakers of the language (it is especially true in our scenario where the data was collected in the Czech Republic). This complicates the creation of the ASR pronunciation lexicon, as it is not possible to extract the pronunciation variants directly from existing English ASR resources. Moreover, the communication frequently contains fragments pronounced in the "domestic" language (greetings, side comments, etc.), causing problems not only in the phonetic baseform generation but even in the definition of the phonetic alphabet itself.

High level of noise

The communication over the VHF radio inherently has a low signal-to-noise ratio. The whole ATC communication protocol is designed to compensate for this but it is well-known that the ASR engine is even less robust to the presence of noise in the speech signal than the human auditory system. Therefore it will be necessary to pay special attention to noise canceling techniques in the ASR system development (this problem, however, is outside the scope of the paper).

Special phraseology

This is actually the property that works in our advantage. The rather rigid structure of the utterances used in the ATC communication allows us to employ semantic entity detection techniques based on context-free grammar on the higher level of the spoken dialogue system, the natural language understanding (NLU) module. On the other hand, the ATC communication protocols sometimes prescribe also a special pronunciation of some words (especially digits - e.g. "nine" should be pronounced as "niner") which further adds to the above described problems with the pronunciation lexicon creation.

[1] Project of the Technology Agency of the Czech Republic No. TA01030476, 2011-2015

[2] Although the ATC communication can be conducted in native language in regional air traffic, the use of English is naturally indispensable in international ATC.

Rarity of the data
The ATC data are only very rarely being collected and transcribed. In fact, the only comparable corpus that we are aware of is the Air Traffic Control Complete (Godfrey, 1994) distributed by LDC. It contains approximately 70 hours of data from Dallas Fort Worth (DFW), Logan International (BOS) and Washington National (DCA) airports. This corpus has several drawbacks – less than 50% of the recorded signal actually contains speech and, moreover, the recording is of poor quality with a lot of noise (even for the ATC standards) as the most of it was obtained by digitizing the analog tapes.

4. Corpus design

Recordings
We took advantage of the fact that our project partner in *IT-BLP*, the CS SOFT company, develops complex IT solutions for several ATC authorities and airports and, as such, has an access to the ATC communication recordings. It was able to secure the following data (predominantly from the Air Navigation Services of the Czech Republic in Jeneč, but also some communication from the Lithuania and Philippines airspace):

- GRP (ground control) – communication before takeoff and after landing – 19.2 hours of data
- TWR (tower control) – communication during takeoff, landing and landing standby – 22.5 hours
- APP (approach control) – communication during landing approach – 25.5 hours
- ACC (area control) – communication during overflights and cruises – 71.3 hours

Those data were first automatically preprocessed and then manually transcribed, as described in the following paragraphs.

Data preprocessing
We wanted to make the manual transcription as efficient as possible and thus we used automatic methods for preprocessing of the raw recordings. First of all, the data were segmented and the segments were classified as speech/non-speech using our proprietary Voice Activity Detector (Prcín et al., 2002). The speech segments were then imported to our own annotation tool Webtransc that serves for online annotation of multimedia data. It allows to play the segments, transcribe their content (including the marking of various non-speech events) and add several types of metadata (such as the speaker's communication role - pilot or controller, in this case). The screenshot of Webtransc is shown on Fig. 1, more details will be provided in the full version of the paper.

Transcription
Our annotators are quite experienced, as the majority of them has already participated in the creating of other speech corpora. We have nevertheless prepared a detailed transcription manuals, paying special attention to instructions

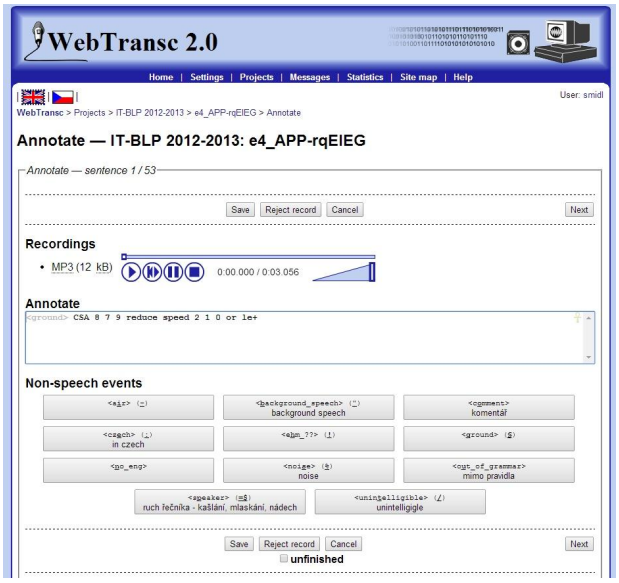

Figure 1: The WebTransc tool screenshot

that concern handling of non-standard pronunciations, special ATC terminology, spelling alphabet and other issues peculiar to ATC (details will be given in the full version of the paper).
The first portion of the transcribed data (20 hours of speech) was released via LINDAT-Clarin as *Air Traffic Control Communication* (Šmídl, 2011) speech corpus, the pronunciation lexicon and n-gram counts (unigrams, bigrams and trigrams) are distributed also via LINDAT-Clarin as additional resource *ATCC: Pronunciation lexicon and n-gram counts for ASR module* (Šmídl, 2012).

Conformation to CLARIN-recommended standards
The formats of the data in the ATCC speech corpus have actually arisen from our long-term research practices, the resources that we have been using previously and the annotation scenarios used in the previous and current projects. However, it turned out that they, for the most part, do conform to the CLARIN-recommended standards:

- The **acoustic data** are enconded in PCM format (8kHz, 16bit PCM, mono) which is, according to (Kemps-Snijders, 2009), a standard recommended for CLARIN data.
- The **transcriptions** are stored in the TRS file which is an XML-based text format native to the Transcriber tool (Barras, 2001). As far as I know, it is not directly supported in CLARIN but has been widely used within the speech research community in the last 15 years. The text of the transcription itself is encoded using Unicode UTF-8 – a general standard for text encoding, also mentioned as supported standard in (Kemps-Snijders, 2009).
- The **phonetic transcription** in the pronunciation lexicon employs the Arpabet transcription code (Arpabet,) that is (most notably) used in the CMU Pronouncing Dictionary. There is a one-to-one mapping between

Arpabet and the International Phonetic Alphabet (IPA) symbols. I was not able to find whether either of these notational systems is supported in CLARIN but I believe that there are not many other options.

The metadata for both the mentioned corpora conform to the CMDI, using a profile `clarin.eu:cr1:p_1349361150622`.

5. References

Arpabet. `http://en.wikipedia.org/wiki/Arpabet`

Barras, C., Geoffrois, E., Wu, Z., Liberman, M. (2001). Transcriber: Development and use of a tool for assisting speech corpora production. *Speech Communication - special issue on Speech Annotation and Corpus Tools*, 33(1-2):5-22

Godfrey, J. (1994). Air traffic control complete LDC94S14A. `https://catalog.ldc.upenn.edu/LDC94S14A`

Kemps-Snijders, M., et al. (2009). Standards for LRT. `http://www.clarin.eu/content/standard-recommendations`

Prcín, M., Müller, L., and Šmídl, L. (2002). Statistical based speech/non-speech detector with heuristic feature set. In *6th World Multi-Conference on Systemics, Cybernetics and Informatics (SCI 2002) / 8th International Conference on Information Systems Analysis and Synthesis (ISAS 2002)*, pages 264–269, Orlando, FL.

Šmídl, L. (2011). Air traffic control communication. `http://hdl.handle.net/11858/00-097C-0000-0001-CCA1-0`

Šmídl, L. (2012). ATCC: Pronunciation lexicon and n-gram counts for ASR module. `http://hdl.handle.net/11858/00-097C-0000-000D-EC92-F`