Graz University of Technology

Signal Processing and Speech Communication Laboratory

Inffeldgasse 12, 8010 Graz, Austria

http://www.spsc.tugraz.at

ATCOSIM Air Traffic Control Simulation Speech Corpus

Konrad Hofbauer

Stefan Petrik

Technical Report

May 2008

TR TUG-SPSC-2007-11

Under contract of:

EUROCONTROL Experimental Centre

Graz University of Technology
Signal Processing and Speech Communication Laboratory
Inffeldgasse 12, 8010 Graz, Austria
http://www.spsc.tugraz.at

ATCOSIM Air Traffic Control Simulation Speech Corpus Technical Report

May 2008
TR TUG-SPSC-2007-11
Version 1.0

Authors:
Konrad Hofbauer
(Email: konrad.hofbauer@tugraz.at)
Stefan Petrik
(Email: stefan.petrik@tugraz.at)

Under contract of:
EUROCONTROL Experimental Centre

Version History

Version 0.8, October 2007: First pre-release for validation.
Version 0.9, November 2007: Recommendations of validation report incorporated.
- License terms updated.
- File list added.
- Pointers to ATC phraseology resources added.
- Pointers to PDF software added.
- Additional information on speakers added.
- Validation report added.
- Translation of foreign words added.
- Some transcriptions updated.

Version 1.0, May 2008: Official release.
- License terms updated.
- Conference article added.

Contents

List of Tables

List of Figures

Abstract

The ATCOSIM Air Traffic Control Simulation Speech corpus is a speech database of air traffic control (ATC) operator speech. It consists of ten hours of speech data, which were recorded during ATC real-time simulations. The database includes orthographic transcriptions and additional information on speakers and recording sessions. The corpus is publicly available and provided free of charge. This report describes the production process of the corpus and gives a thorough description of the final corpus. Possible applications of the corpus are, among others, ATC language studies, speech recognition and speaker identification, as well as listening tests within the ATC domain.

1 Introduction

A *speech corpus* is a set of digital recordings of speech together with annotation, meta data, and documentation.

Speech corpora provide the basic data for research and development in different fields such as [1, 2]:

- spoken language communication
- spoken language processing (SLR)
- automatic speech recognition (ASR)
- text-to-speech systems (TTS)
- speech synthesis
- spoken language interfaces
- spoken language understanding
- speaker verification
- spoken language modelling

Appendix A lists a selection of publicly available speech corpora and provides web resources which contain extensive listings of existing corpora. There are many different aspects that characterise a corpus, among those are (from [1]):

- speaker profiles
- number of speakers
- vocabulary
- domain
- task
- phonological distribution
- speaking style
- recording setup
- annotation
- technical aspects
- structure
- validation
- meta data
- documentation

The corpus presented in this document is within the domain of civil *air traffic control (ATC)*. The aim of air traffic control is to maintain a safe separation between all aircraft in the air in order to avoid collisions, and to maximise the number of aircraft

that can fly at the same time. Besides a set of fixed flight rules and a number of navigational systems, air traffic control relies on human air traffic control operators (ATCO, or *controllers*). The controller monitors air traffic within a so-called *sector* (a geographic region or airspace volume) based on radar pictures and gives flight instructions to the aircraft pilots in order to maintain a safe separation between the aircraft.

Although digital data communication links between controllers and aircraft are slowly emerging, most of the communication between controllers and pilots is verbal and by means of analogue voice radios. The communication occurs on a party-line channel, which means that all aircraft within a sector as well as the corresponding controller can hear all messages that are transmitted on that channel frequency.

The phraseology that is used for this communication is strictly formalised by the International Civil Aviation Organization (ICAO) [3]. In practise however, both controllers and pilots deviate from this standard phraseology. The international standard language for ATC communication is English. In certain countries also the use of the local language among local controllers and local pilots is permitted.

A description of the characteristic ATC phraseology would be beyond the scope of this document. The manual of radiotelephony [3] contains an overview and many illustrative examples of the recommended phraseology. An updated list of airline radio telephony designators (call-signs) is published in regular intervals by ICAO [4]. Lists with names of navigational aids and location names can be found in the national aeronautical information publications (AIP) [5] and can also be obtained from specialised publishers such as Jeppesen or governmental agencies [6]. Also starting points for building a grammar for the ATC communication already exist [7, 8, 9].

1.1 Existing Corpora Related to Air Traffic Control

The development of spoken language technologies for air traffic control requires corpora that are tailored to the task and the domain. This is due to the specific language and phraseology used in air traffic control. Air traffic control corpora also facilitate the study and modelling of the actual controller language in use. Despite the large number of existing corpora, only a few corpora are in the air traffic control domain. The following sections give a short description of the ATC-related corporas known to the authors, of which some might or might not be available to the public.

1.1.1 NIST Air Traffic Control Complete Corpus

The NIST Air Traffic Control Complete Corpus consists of recordings of approach control radio transmissions in the United States [10]. The documentation states:

> "The audio data on the discs is composed of voice communication traffic between various controllers and pilots. The audio files are 8 KHz, 16-bit linear sampled data, representing continuous monitoring, without squelch

> or silence elimination, of a single FAA frequency for one to two hours. There are also files which indicate the amplitude of the received AM carrier signal at 10 ms intervals. Full transcripts, including the start and end times of each transmission, are provided for each audio file. Each flight is identified by its flight number.
>
> ATC0 consists of three sub-corpora, one for each airport in which the transmissions were collected—'Dallas Fort Worth (DFW), Logan International (BOS) and Washington National (DCA). The complete set contains approximately 70 hours of controller and pilot transmissions collected via antennas and radio receivers which were located in the vicinity of the respective airports. The ATC0 Corpus was collected by Texas Instruments under contract to DARPA. It was produced on CD-ROM by the National Institute of Standards and Technology for distribution by the Linguistic Data Consortium."

The database is commercially available.

1.1.2 HIWIRE

The HIWIRE database is a collection of read or prompted words and sentences from the area of military air traffic control. The recordings were made in a studio setting, and cockpit noise was artificially added afterwards [11]. The documentation states:

> "The database contains 8100 English utterances pronounced by non-native speakers (31 French, 20 Greek, 20 Italian, and 10 Spanish speakers). The collected utterances correspond to human input in a command and control aeronautics application. The data was recorded in studio with a close-talking microphone and real noise recorded in an aeroplane cockpit was artificially added to the data. The signals are provided in clean (studio recordings with close talking microphone), low, mid and high noise conditions. The three noise levels correspond approximately to signal-to-noise ratios of 10dB, 5dB and -5dB respectively."

No usage restrictions were found for this database. The corpus seems to be available on request.

1.1.3 nnMATC

The non-native Military Air Traffic Control (nnMATC) database is a collection of military ATC radio speech. The recordings were made in a military air traffic control centre, wire-tapping the actual radio communication during military exercises [12]. The documentation states:

> "The nnMATC database combines the adverse effects of non-native speech and noisy environment, through realistic air traffic control communications recorded from an operational military Air Traffic Control (ATC) centre.

> The nnMATC database consists of 24+ hours of ATC communications. All recordings were tapped from the ATC centre itself, implying a different speech quality depending on the speakers location: on the controller-side, speech is mostly clean; on the pilot side, recordings suffer from a combination of background noise (cockpit) and communication interferences.
>
> The non-Native English accents covered at controller side are mainly Belgian Dutch and Belgian French. At the pilot side, the variety is much wider with—among others—Dutch, Belgian Dutch, French, Belgian French, German, Italian, and Spanish accents. Few Native American, British and Canadian English speakers are represented among the pilots as well. Although most speakers are males, there are few female speakers as well, mainly among the controllers."

The use of this database is restricted to the NATO/RTO/IST-031 working group and its affiliates. Use has been extended to participants of the Interspeech 2007 Special Session entitled "Novel techniques for the NATO non-native Air Traffic Control and HIWIRE cockpit databases". Its commercial use is strictly prohibited.

1.1.4 VOCALISE

The VOCALISE project recorded and analysed a large amount of operational air traffic control voice radio communication in France [13, 14]. Besides a transcription the database also includes additional information such as the corresponding radar images and flight plans.

The recordings in the database reflect the three main categories of civil air traffic control:

En route control 60 hours at different en-route control centres during heavy traffic

Approach control 50 hours at three large approach centres (Roissy, Nice and Lyon)

Tower control 40 hours at three large airports (Roissy, Nice and Lyon)

The database and its documentation seem not to be available for the public. According to the website, the use is restricted to research groups affiliated with the French 'Centre d'Études de la Navigation Aérienne' (CENA)—'now part of the 'Direction des Services de la Navigation Aérienne' (DSNA).

1.2 Purpose of the ATCOSIM Corpus

The aforementioned corpora vary significantly among each other with respect to e.g. scope, technical conditions or public availability. The aim of the ATCOSIM corpus is to fill the gap that is left by the above corpora (Tbl. 1.1): ATCOSIM provides publicly available direct-microphone recordings of operational air traffic

controller speech in realistic en-route control scenarios. The corpus is meant to be versatile and as such is not tailored to any specific speech technology application. It provides an utterance segmentation and an orthographic transcription. Depending on the application, a further annotation of the corpus might be necessary.

1.3 History

The audio recordings used to create this corpus were initially made in 1997 for a different purpose. The recordings were at that time also orthographically transcribed and used for a study on the language used by the controllers [7]. Unfortunately the transcriptions have since been lost, and it was decided in 2006 to create a new corpus based on the original recordings still existing on digital audio tapes (DAT). The creation of this corpus based on these recordings is the topic of this document. Beside the information given in the language study [7], no other meta data such as recording protocols or detailed speaker profiles is available. It is also not guaranteed that the set of recordings used for the language study fully matches the set of recordings used for this corpus. There is however a significant overlap between the two sets.

Table 1.1: Feature comparison of ATC-related speech corpora

	NIST	**HIWIRE**	**nnMATC**	**VOCALISE**	**ATCOSIM**
Recording Situation					
- Recording content	civil ATCO & PLT	N/A	military ATCO & PLT	civil ATCO & PLT	**civil ATCO**
- Control position	approach	N/A	military	mixed	**en-route**
- Geographic region	USA	N/A	Europe (BE)	Europe (FR)	**Europe (DE/CH/FR)**
- Speaking style (context)	operational	prompted text	operational	operational	**operational** [(1)]
Recording Setup					
- Speech bandwidth	narrowband	wideband	mostly narrowband	unknown	**wideband**
- Transmission channel	radio	none	none / radio	none / radio	**none**
- Radio transmission noise	high	none	mixed	mixed	**none**
- Acoustical noise	CO & CR	CO (artificial)	CO & CR	CO & CR	**CR**
- Signal source	VHF radio	direct microphone	mixed	unknown	**direct microphone**
Speaker Properties					
- Level of English	mostly native	non-native	mostly non-native	mixed	**mostly non-native**
- Gender	mixed	mixed	mostly male	mixed	**mixed**
- Operational	yes	no (!)	yes	yes	**yes**
- Field of prof. operation	civil	N/A	military	civil	**civil**
- Number of speakers	unknown (large)	81	unknown (large)	unknown (large)	**10**
Publicly Available	yes	yes (?)	no	no	**yes**

• CO: Cockpit • CR: Control Room • ATCO: Controller • PLT: Pilot

[(1)] Large-scale real-time simulation

2 Data Collection

As the recordings which this corpus is based on were made ten years prior in a different context, the information available is not complete. However, a large part of the information could be reconstructed from various sources.

2.1 Recording Situation

The voice recordings were made in the air traffic control room of the EUROCONTROL Experimental Centre (EEC) in Brétigny-sur-Orge, France (Fig. 2.1). The room and its controller working positions closely resemble an operational control centre room and are used for large-scale air traffic control simulations. The EEC has a long history of performing air traffic control simulations for the evaluation of alternative or modified air traffic control concepts before their implementation in real-world operations. The simulations aim to provide realistic air traffic scenarios and working conditions for the air traffic controller. The controller communicates via a headset with pseudo-pilots which are located in a different room and control the simulated aircraft. Several controllers operate at the same time, in order to simulate also the inter-dependencies between different control sectors.

The simulations during which the recordings were made studied the impact of reduced vertical separation minima (RVSM) between aircraft, a concept that is by now in operation in core Europe. A detailed report on this '3rd Continental RVSM Real-Time Simulation' is available [15]. The simulation implemented real sector layouts and air traffic samples. The amount of air traffic was changed to simulate different controller work loads, and also the vertical separation of the aircraft was modified. During the simulations only the controllers' voice, but not the pilots', was recorded. The recordings cover simulations of airspace sectors in Germany (Söllingen, controlled by the Karlsruhe centre) and Switzerland (Zürich and Geneva).

2.2 Speaker Profiles

The participating controllers were all professional and actively employed air traffic controllers (Tbl. 2.1). In the simulations, each controller was assigned to a sector that actually belonged to the ATC centre they were from. The controllers were therefore familiar with the situations presented to them. All controllers were of either German or Swiss nationality (Tbl. 2.2). The controllers had agreed to the recording of their voice for the purpose of language analysis as well as for research and development in speech technologies, and were asked to show their normal working behaviour.

Figure 2.1: Control room and controller working position at the EUROCONTROL Experimental Centre (recording site)

Table 2.1: Distribution of speakers over control centres (sectors), native tongue, sex, mean age and mean ATC experience

Sector	**Native Tongue**	**Speakers**	**Female**	**Male**	**Mean Age**	**ATC Exp.**	**Sector Exp.**
Söllingen	German	4	0	4	~38 a	~15 a	~11 a
Geneva	Swiss French	3	1	2	~28 a	~4 a	~4 a
Zürich	Swiss German	3	3	0	~27 a	~6 a	~4 a
Total		10	4	6	~31 a	~8 a	~6 a

Table 2.2: Complete list of speakers in the corpus

Spk. ID	**Nationality**	**Native Tongue**	**Gender**	**ATC Sector**	**Sessions**	**Utterances**	**Int. ID**
sm1	German	German	Male	Söllingen	7	1167	a
sm2	German	German	Male	Söllingen	9	1848	b
sm3	German	German	Male	Söllingen	5	808	c
sm4	German	German	Male	Söllingen	6	1162	d
gf1	Swiss	Swiss French	Female	Geneva	1	238	e
gm1	Swiss	Swiss French	Male	Geneva	2	384	f
gm2	Swiss	Swiss French	Male	Geneva	2	378	g
zf1	Swiss	Swiss German	Female	Zürich	8	1716	i
zf2	Swiss	Swiss German	Female	Zürich	7	1739	j
zf3	Swiss	Swiss German	Female	Zürich	3	638	k

2.3 Recording Setup

The controller's speech was picked up by the microphone of a Sennheiser HME 45-KA headset. The microphone signal and a push-to-talk (PTT) switch status signal were recorded with a Sony DTC-60ES digital audio tape (DAT) recorder in LongPlay (LP) mode, which results in the data format given in Tbl. 2.3. The DAT recorded was integrated into the existing voice communication system using an STIF interface [16]. The push-to-talk switch is the push-button that the controller has to press and hold in order to transmit the voice signal on the real-world radio. The microphone signal is muted during the recording by the STIF interface when the PTT button is not pressed.

Table 2.3: Data format of raw DAT recordings

Sampling Frequency	32000 Hz
Resolution	12 bit (provided as zero-padded 16 bit on digital S/PDIF output)
Tracks	2 [PTT (left channel) and speech (right channel)]

3 Data Processing

The basis for the production of this corpus was a set of eleven digital audio tapes (DAT). The tapes were labelled “AS08-xx" with xx being a two-digit number in the range of 01, 02, ..., 09, 10, 11.

3.1 Transfer of Audio Data to Computer

The recordings were transferred from the provided DAT tapes onto a hard disk using the same Sony DTC-60ES DAT recorder that had been used for the recording of the tapes. For the transfer the recorder was connected to a personal computer via an optical S/PDIF (TOSLINK) connection. The PC was equipped with a professional RME Hammerfall HDSP 9632 sound-card. The setup was verified to provide bit-true digital transfer without any modification to the digital audio data. It showed necessary to assure that in no step of the transfer process dithering is applied to the data, as this would change the data in its least significant bits. No resampling or other signal processing operations were applied.

In order to circumvent possible file-size limitations that might exist in certain file systems, the size of the recorded files on the PC was limited to 2 GB. As a consequence two separate audio files (parts) exist for some of the tapes. One other tape resulted in two parts as the tape was not continuously recorded (formatted) and in this case the DAT recorder aborts playing. The tape-to-filename assignment is given in Tbl. 3.1. The first two digits represent the number of the tape, the third digit whether it is the first or second part of the tape. The same filenames are used as identifiers for the 'takes' of the corpus (one 'take' being all the utterances included in one part). For reference, the original dates of recording as well as the positions of the so-called Start-IDs as provided by the subcode channel of the DAT tapes are included in Tbl. 3.1.

3.2 Utterance Segmentation

In general practise, the controller presses the push-to-talk (PTT) button in order to activate his microphone, the radio transmitter, and to transmit his speech on air. The button is released again at the end of the utterance in order to free the channel for other users. The same functionality is provided in the simulation so that the (phantom) pilot can hear the controller only while the PTT button is pressed. The original DAT tapes contain the recorded controller speech on the right channel and a PTT status signal on the left channel. The recorded speech signal was gated by

Table 3.1: Name, recording dates and Start-ID positions of the original tapes, as well as name of the saved audio files

Tape	**Recording Dates**	**Start IDs**	**File Names**
AS08-01	21.01.97 22.01.97	00:00:08, 00:16:35, 00:24:04, 00:25:19, 00:43:21, 01:10:36, 01:14:19, 02:25:02, 03:24:56	011, 012
AS08-02	22.01.97 23.01.97	00:00:09, 01:00:19, 02:02:09, 03:00:11, 03:01:11, 03:49:01	021
AS08-03	23.01.97 24.01.97	00:00:08, 01:01:08, 02:00:10, 03:00:01, 03:00:11	031
AS08-04	27.01.97 28.01.97	00:00:08, 00:48:47, 00:50:12, 01:52:12, 02:52:13	041
AS08-05	28.01.97 29.01.97	00:00:10, 01:04:10, 02:05:11, 03:05:10, 03:06:12	051
AS08-06	29.01.97 30.01.97	00:00:09, 00:58:09, 01:51:10, 02:37:11, 03:24:54	061
AS08-07	31.01.97 04.02.97 05.02.97	00:00:06, 00:47:45, 00:48:07, 01:48:08, 02:48:09, 03:48:09, 04:36:10	071, 072
AS08-08	05.02.97 06.02.97 07.02.97	00:00:08, 00:37:29, 00:39:32, 00:46:09, 01:48:09, 03:17:10, 04:17:11	081, 082
AS08-09	07.02.97 10.02.97 11.02.97	00:00:10, 01:02:11, 02:05:12, 03:11:11, 04:11:03, 04:11:14, 05:04:22, 05:12:15	091,092
AS08-10	11.02.97 12.02.97	00:00:06, 01:00:26, 01:01:06, 01:59:09, 02:33:09, 03:33:10	101
AS08-11	13.02.97	00:00:10, 00:49:10, 01:23:54, 01:43:11, 02:49:45	111

the PTT button, so that the microphone signal was only recorded while the PTT button was pressed. The status of the PTT button inherently marks the beginning and ending of each controller utterance. One such utterance can consist of one or several ATC instructions, can contain noise or off-talk, or can be empty altogether.

3.2.1 PTT Signal Processing

The left channel of the DAT tapes provides the status of the PTT button. It contains a high-frequency tone which is *muted* while the PTT button is pressed. Therefore the tone is present in the regions where the PTT button was *not* pressed. Due to the analogue circuitry used to create and mute the tone, peaks occur at the transition regions and the tone is not instantaneously and completely muted. Figure 3.1 shows an example of a short utterance.

The beginning and the end of the utterances were marked in a semi-automatic way, as it provides a more accurate and consistent result compared to manual segmentation alone. The accurate detection of the instances of switching to 'PTT ON' and switching to 'PTT OFF' required the pre-processing of the recorded PTT status signal.

The signal was at first high-pass filtered with a cutoff frequency of 9 kHz in order to eliminate undesired low frequency components of the signal. In a second step, the envelope of the filtered signal was computed by low-pass filtering the absolute value of the signal, using a cutoff frequency of 10 Hz. All segments where this envelope was below a threshold of -55 dB (relative to digital full scale) were considered as segments of 'PTT ON' at first.

The PTT status signal as recorded on the DAT tape was not always stable but contained distortion, drop-outs, etc., so that a further processing of the detected segments showed necessary. First all 'PTT OFF' segments shorter than 20 ms were transformed to 'PTT ON', thus forming together with their neighbouring segments longer 'PTT ON' sequences. Vice versa, in a second step 'PTT ON' segments shorter than 20 ms, surrounded by 'PTT OFF' segments longer than 20 ms, were transformed to 'PTT OFF'. This order of steps introduced a slight bias towards 'PTT ON', as a sequence of segments that are shorter than 20 ms results in 'PTT ON'.

The analogue switch in the recording setup caused an audible noise burst or click where the PTT went on and where the PTT went off. This was in turn used to improve the timing accuracy of the detection algorithm. A local search was performed in a window of 5 ms before and 2 ms after the previously estimated instances of the PTT status changing from 'PTT OFF' to 'PTT ON'. The position where the absolute value of the original, unfiltered PTT status signal is maximum was considered as the new correct transition position. Similarly, a more accurate position was identified in a window of 2 ms before and 5 ms after the previously estimated change from 'PTT ON' to 'PTT OFF'.

Since the tapes contain blanks at the beginning and at the end, which would have been interpreted as muted and therefore as 'PTT ON', all takes were set to start and finish with 'PTT OFF'. In order to eliminate blanks in the middle of the tape as well as drop-outs of the PTT signal, which would all have been interpreted as 'PTT

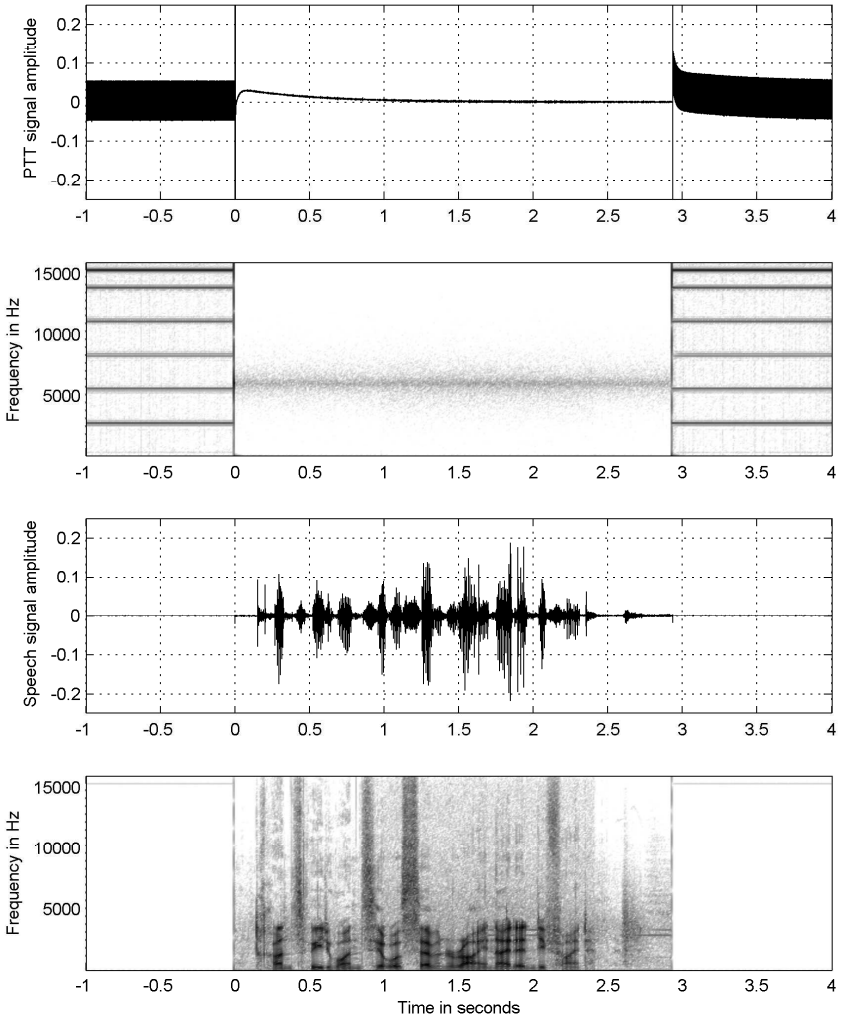

Figure 3.1: A short speech segment (`transwede one zero seven rhein identified`) with push-to-talk (PTT) signal. Time domain signal and spectrogram of the PTT signal (top two) and time-domain signal and spectrogram of the speech signal (bottom two)

ON', all 'PTT ON' segments where the speech signal did not contain a single sample above a threshold of -55 dB (relative to digital full scale) were eliminated. This very low threshold preserved all those segments in which the controller pressed the PTT button but did not say anything, as the recorded background noise was well above this threshold.

In order to assure that no segment was falsely classified as 'PTT OFF', a further control step was undertaken. The maximum instantaneous power for each segment classified as 'PTT OFF' (minus transition regions of 200 ms at the beginning and the end) was computed. All segments with a *maximum peak* power below -60 dB can be safely considered as empty and were treated as such. All other segments containing a peak with a speech power level above -60 dB were manually inspected. The inspection confirmed that all of these segments were correctly classified as 'PTT OFF' and that the encountered signal energy resulted from technical issues in the original recording setup, such as clicks or a remaining direct current (DC) signal after the muting of the microphone.

The segmentation process resulted in 10,078 segments of 'PTT ON', which are further on referred to as 'utterances'. The total length of all utterances is approximately 10.7 hours, taken from a total of 51.4 hours of recordings.

3.2.2 Export to Text and Audio Files

The detected 'PTT ON' and 'PTT OFF' region boundaries were exported into plain text files in order to be used within the transcription tool described in Section 4.1. For every 'PTT ON' segment a separate audio file was created containing the voice recording of the corresponding segment or controller utterance. Additionally, for every 'PTT ON' segment a time-stretched version of the voice recording was created using the PRAAT implementation of the Pitch-Synchronous Overlap and Add (PSOLA) method [17]. The duration of each utterance was stretched by a factor of 1.7. The transcriptionist used these slowed-down versions of the recordings only when dealing with utterances that were difficult to understand.

4 Speech Transcription

The speech corpus includes an orthographic transcription of the controller utterances. The orthographic transcriptions are aligned with each utterance.

4.1 Transcription Environment

A large number of transcription tools, both commercial and open-source, are available on the market, albeit most of them being tailored to specific tasks. An extensive but still far from complete overview is available online [18, 19].

The open-source tool TableTrans was chosen for our application. This program was created in large part by the Linguistic Data Consortium (LDC), which is also a major provider of speech databases. TableTrans is written in Python and Tcl/Tk and based on the Annotation Graph Toolkit AGTK, a library of annotation-related functions, and Snack/WaveSurfer, a toolkit for handling, displaying and analysing sound data. [20, 21, 22]

TableTrans was selected for its table-based input structure as well as for its capability to readily import our automatic segmentation. Fig. 4.1 shows a screen-shot of the application as used by the transcriptionist. The transcriptionist fills out a table in the upper half of the window. Each row of the table represents one utterance. In the lower half of the window the waveform of the utterance that is currently selected or edited in the table is automatically displayed. The transcriptionist can play, pause and replay the currently active utterance by a single key stroke or as well select and play a certain segment in the waveform display. TableTrans would also support the display of many other parameters of the speech signal such as spectrogram or pitch, but this was not deemed necessary for our application.

A small number of minor modifications to the TableTrans source code was necessary. First and foremost, the given segmentation was locked in order to prevent the transcriptionist from accidentally changing it. Second, a number of unnecessary and possibly harmful menu items were disabled and some key bindings changed for convenience. The automatic import was configured to work with our text-based annotation files and non-editable protected columns were created in the table. Finally, the replay capabilities of the software were extended.

One keystroke was designated to pause replay. Another keystroke restarted replay half a second before where replay was paused. This allowed the replay of the last word or word segment that was said before the pause. This proved to be very helpful to the transcriptionist.

The audio files were provided as Microsoft WAV files to TableTrans. Data import and export to and from TableTrans was done via plain text comma-separated value

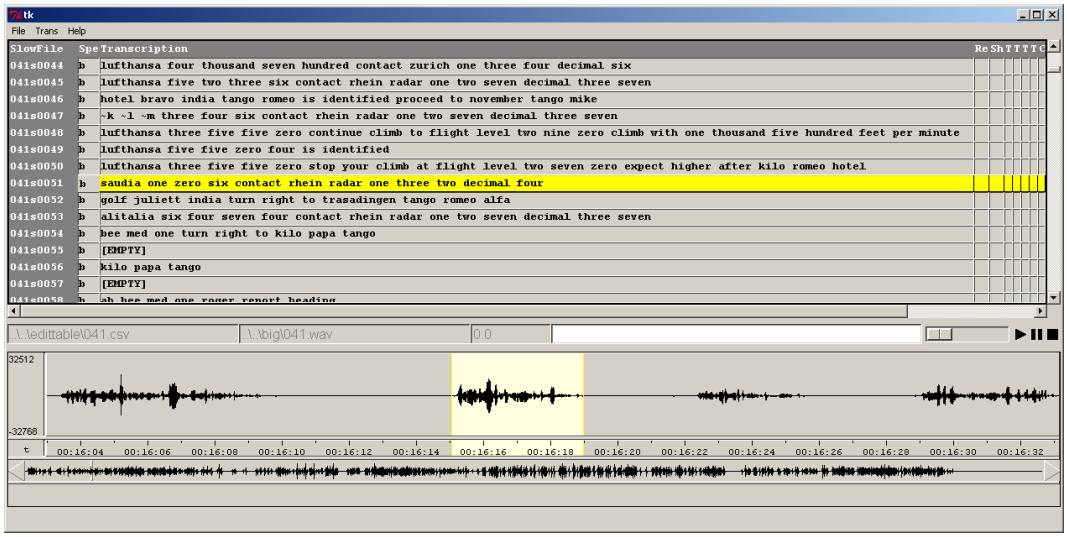

Figure 4.1: Screen-shot of the transcription tool TableTrans

(CSV) files, which can be easily created and also processed by a large number of programs or converted into various different formats (including XML) using readily available open source tools.

A number of key combinations (keyboard shortcuts) were provided to the transcriptionist using the open-source tool AutoHotKey [23]. These were used for conveniently accessing the time-stretched audio files and for entering frequent character or word sequences such as pre-defined keywords, ICAO alphabet spellings and frequent commands, both for convenience and in order to avoid typing mistakes.

The system was installed on a personal computer running Windows XP and the audio monitored using Sennheiser HD 560 ovation II headphones. The system was also tested on Mac OS X using Parallels. An installation of the setup on other computer platform should in principle be possible. The TableTrans software is also available for Unix/Linux/OS X but requires a recompilation of the source code and of the underlying libraries. For the Windows-only AutoHotKey program an appropriate replacement on the corresponding operating system would have to be found, but is often already part of the operating system itself.

4.2 Transcription Format

The orthographic transcription follows a strict set of rules which is presented hereafter. In general, all utterances are transcribed word-for-word in standard British English. All standard text is written in lower-case. Punctuation marks including periods, commas and hyphens are omitted. Apostrophes are used only for possessives (e.g.

pilot's radio)[1] and for standard English contractions (e.g. it's, don't).

Technical noises as well as speech and noises in the background—produced by speakers other than the one recorded—are not transcribed. Silent pauses both between and within words are not transcribed either. Numbers, letters, navigational aids and radio call signs are transcribed as follows.

In terms of notation, stand-alone technical mark-up tags are written in upper case letters with enclosing squared brackets (e.g. [HNOISE]). Regular lower-case letters and words are preceded or followed by special characters to mark truncations (=), individually pronounced letters (~) or unconfirmed airline names (@). Groups of words are embraced by opening and closing XML-style tags to mark off-talk (<OT> ... </OT>), which is also transcribed, and foreign language (<FL> </FL>), for which currently no transcription is given.

4.2.1 ICAO Phonetic Spelling

Definition: Letter sequences that are spelt phonetically.

Notation: As defined in the reference.

Example: india sierra alfa report your heading

Word List: alfa bravo charlie delta echo foxtrot golf hotel india juliett kilo lima mike november oscar papa quebec romeo sierra tango uniform victor whiskey xray yankee zulu

Supplement: fox (short form for foxtrot)

Reference: [3]

4.2.2 Acronyms

Definition: Acronyms that are pronounced as a sequence of separate letters in standard English.

Notation: Individual lower-case letters with each letter being preceded by a tilde (~). The tilde itself is preceded by a space.

Example: ~k ~l ~m

Exception: Standard acronyms which are pronounced as a single word. These are transcribed without any special markup.

Exception Example: NATO, OPEC, ICAO are transcribed as nato, opec and icao respectively.

4.2.3 Numbers

Definition: All digits, connected digits, and the keywords 'hundred', 'thousand' and 'decimal'.

[1] The mono-spaced typewriter type represents words or items that are part of the corpus transcription.

Notation: Standard dictionary spelling without hyphens. It should be noted that controllers are supposed to use non-standard pronunciations for certain digits, such as 'tree' instead of 'three', 'niner' instead of 'nine', or 'tousand' instead of 'thousand' [3]. This is however applied inconsistently, and in any case transcribed with the standard dictionary spelling of the digit.

Example: `three hundred`, `one forty four`, `four seven eight`, `one oh nine`

Word List: `zero oh one two three four five six seven eight nine ten hundred thousand decimal`

4.2.4 Airline Telephony Designators

Definition: The official airline radio call sign.

Notation: Spelling exactly as given in the references.

Examples: `air berlin`, `britannia`, `hapag lloyd`

Exceptions: Airline designators given letter-by-letter using ICAO phonetic spelling as well as airline designators articulated as acronyms.

Exceptions Examples: `foxtrot sierra india`, `~k ~l ~m`

References: [4, 24, 25]

4.2.5 Navigational Aids and Airports

Definition: Airports and navigational aids (navaids) corresponding to geographic locations.

Notation: Geographical locations (navaids) are transcribed as given in the references using lower-case letters. The words used can be names of real places (ex. `hochwald`) or artificial five-letter navaid or waypoint designators (ex. `corna`, `gotil`).[2] Airports and control centres are transcribed directly as said and in lower-case spelling.

Examples:
```
contact rhein on one two seven
alitalia two nine two turn left to gotil
alitalia two nine two proceed direct to corna charlie oscar romeo november alfa
```

References: [15, Annex A: Maps of Simulation Airspace], [6, 26]

4.2.6 Human Noises

Definition: Human noises such as coughing, laughing and sighs produced by the speaker. Also breathing noises that were considered by the transcriptionist as exceptionally loud were marked using this tag.

Notation: `[HNOISE]` (in upper-case letters)

[2] In some occasions less popular five-letter designators are also spelt out using ICAO phonetic spelling.

Example: `sabena [HNOISE] four one report your heading`

4.2.7 Non-verbal Articulations

Definition: Non-verbal articulations such as confirmative, surprise or hesitation sounds.

Notation: Limited set of expressions written in lower-case letters.

Example: `malaysian ah four is identified`

Word List: `ah hm ahm yeah aha nah ohh`[3]

4.2.8 Truncated Words

Definition: Words which are cut off either at the beginning or the end of the word due to stutters, full stops, or where the controller pressed the push-to-talk (PTT) button too late. This also applies to words that are interrupted by human noises (`[HNOISE]`). This notation is used when the word-part is understandable. Empty pauses within words are not marked.

Notation: The missing part of the word is replaced by an equals sign (=).

Examples: `good mor= good afternoon` (correction), `luf= lufthansa three two five` (stutter), `=bena four one` (PTT pressed too late), `sa= [HNOISE] =bena` (interruption by cough)

Exception: Words which are cut off either at the beginning or the end of the word due to fast speech or sloppy pronunciation are recorded according to standard spelling and not marked. If the word-part is too short to be identified, another notation is used (see below).

Exception Example: "goo'day" is transcribed as `good day`.

4.2.9 Word Fragments

Definition: Fragments of words that are too short so that no clear spelling of the fragment can be determined.

Notation: `[FRAGMENT]`

Example: `[FRAGMENT]`

4.2.10 Empty Utterances

Definition: Instances where the controller pressed the PTT button, said nothing at all, and released the button again.

Notation: `[EMPTY]`

Exception: If the utterance contains human noises produced by the speaker, the [HNOISE] tag is used.

[3] In contrast to `ohh` as an expression of surprise, the notation `oh` is used for the meaning 'zero', as in `one oh one`.

4.2.11 Off-Talk

Definition: Speech that is neither addressed to the pilot nor part of the air traffic control communication.

Notation: Off-talk speech is transcribed and marked with opening and closing XML-style tags: `<OT> ... </OT>`

Example: `speedbird five nine zero <OT> ohh we are finished now </OT>`

4.2.12 Nonsensical Words

Definition: Clearly articulated word or word part that is not part of a dictionary and that also does not make any sense. This is usually a slip of the tongue and the speaker corrects the mistake.

Notation: `[NONSENSE]`

Example: `[NONSENSE] futura nine three three identified`

4.2.13 Foreign Language

Definition: Complete utterances, or parts thereof, given in a foreign language.

Notation: The foreign language part is not transcribed but is in its entirety replaced by adjacent XML-style tags: `<FL> </FL>`

Example: `<FL> </FL> break alitalia three seven zero report mach number`

Exception: Certain foreign language terms, such as greetings, are transcribed according to the spelling of that language, and are not tagged in any special way. A full list is given below.

Exception Examples `bonjour, tag, ciao`

Exception Word List: See Section 4.3.4.

4.2.14 Unknown Words

Definition: Word or group of words that could not be understood or identified.

Notation: `[UNKNOWN]`

Example: `[UNKNOWN] five zero one bonjour cleared st prex`

4.3 Additions to the Transcription Format

The actual language use in the recordings required the following additions to the above transcription format definitions.

4.3.1 Airline Telephony Designators

The following airline telephony designators cannot be found in the references cited above, but are nonetheless clearly identified.

4.3.1.1 Military Radio Call Signs

There was no special list for military aircraft call signs available. The following call signs were confirmed by an operational controller:

- `~i ~f ~o`
- `mission`
- `nato`
- `spar`
- `steel`

4.3.1.2 General Aviation Call Signs

In certain cases general aviation aircraft are addressed using the aircraft manufacturer and type number (e.g. `fokker twenty eight`). The following manufacturer names occurred:

- `fokker`
- `~b ~a` (Short form for British Aerospace.)

4.3.1.3 Deviated Call Signs

In certain cases the controller uses a deviated or truncated version of the official call sign. The following uses occurred:

- `bafair` (Short form for `belgian airforce`.)
- `netherlands` (Short form for `netherlands air force`.)
- `netherlands air` (Short form for `netherlands air force`.)
- `german air` (Short form for `german air force`.)
- `french air force` (The official radio call sign is `france air force`.)
- `israeli` (Short form for `israeli air force`.)
- `israeli air` (Short form for `israeli air force`.)
- `turkish` (Short form for `turkish airforce`.)
- `hapag` (Short form for `hapag lloyd`.)
- `french line` (Short form for `french lines`.)
- `british midland` (This is the airline name. The radio call sign is `midland`.)
- `berlin` (Short form for `air berlin`.)
- `algerie` (Short form for `air algerie`.)
- `hansa` (Short form for `lufthansa`.)
- `lufty` (Short form for `lufthansa`.)
- `luha` (Short form for `lufthansa`.)
- `france` (Short form for `airfrans`.)
- `meridiana` (This is the airline name, which also used to be the radio call sign. The official call sign was changed to `merair` at some point in the past.)
- `tunis air` (This is the airline name. The radio call sign is `tunair`.)

- malta (Short form for air malta.)
- lauda (Short form for lauda air.)

4.3.1.4 Additional Verified Call Signs

The following call sign occurred and is also verified:

- london airtours (This call sign is listed only in the simulation manual [27].)

4.3.1.5 Additional Unverified Radio Call Signs

The following airline telephony designators could not be verified through any of the available resources. They are transcribed as understood by the transcriptionist on a best-guess basis and preceded by an at symbol (@).

@aerovic	@cheeseburger	@indialook	@period
@alpha	@color	@ingishire	@roystar
@aviva	@devec	@jose	@sunwing
@bama	@foxy	@metavec	@taitian
@cheena	@hanseli	@nafamens	@tele

4.3.2 Navigational Aids

4.3.2.1 Deviated Navaids

In certain cases the controller uses a deviated version of the official navaid name. The following uses occurred:

- milano (Local Italian version for milan.)
- trasa (Short form for trasadingen.)

4.3.2.2 Additional Verified Navaids

The following additional navaids occurred and are verified as they were occasionally spelt out by the controllers:

- corna
- gotil

4.3.3 Special Vocabulary and Abbreviations

The following ATC specific vocabulary and abbreviations occurred. This listing is most likely incomplete.

- masp (Minimum Aviation System Performance standards, pronounced as one word)
- ~r ~v ~s ~m (Reduced Vertical Separation Minimum)

- `~c ~v ~s ~m` (Conventional Vertical Separation Minimum)
- `~i ~f runway` (Initial Fix runway)
- `sec` (sector)
- `freq` (frequency)

4.3.4 Foreign Language Greetings and Words

Due to their frequent occurrence the following foreign language greetings and words were transcribed, using a simplified spelling which avoids special characters:

- `hallo` (German for 'hello')
- `auf wiederhoren` (German for 'goodbye')
- `gruss gott` (German for 'hello')
- `servus` (German for 'hi')
- `guten morgen` (German for 'good morning')
- `guten tag` (German for 'hello')
- `adieu` (German for 'goodbye')
- `tschuss` (German for 'goodbye')
- `tschu` (German for 'goodbye')
- `danke` (German for 'thank you')
- `bonjour` (French for 'hello')
- `au revoir` (French for 'goodbye')
- `merci` (French for 'thank you')
- `hoi` (Dutch for 'hello')
- `dag` (Dutch for 'goodbye')
- `buongiorno` (Italian for 'hello')
- `arrivederci` (Italian for 'goodbye')
- `hejda` (Swedish for 'goodbye')
- `adios` (Spanish for 'goodbye')

4.4 Transcriptionist

The entire corpus was transcribed by a single person, which promises high consistency of the transcription across the entire database. The transcriptionist is a native English speaker with a lot of experience in understanding non-native speakers of various countries through teaching English language courses. The actual raw transcription was carried out in a limited time period of less than three months. The transcriptionist was introduced to the basic ATC phraseology [3] and given lists covering country-specific toponyms and radio call signs [15, 4, 24, 25, 6, 26, 27]. Together with the transcriptionist clear guidelines were established and new cases that were not yet covered by the guidelines immediately discussed.

4.5 Transcription Process and Quality Assurance

Roughly three percent of all utterances were randomly selected across all speakers and used for a pre-training of the transcriptionist. This pre-transcription was also used to validate the applicability of the transcription format definition. Minor changes were applied, yielding to the format as given in Section 4.2. The transcriptions collected during the training phase were discarded and the material re-transcribed in the course of the final transcription.

In the initial transcription of the full corpus special tags were used to mark words or utterances which were not clearly understood. A very large part of these unclear cases was resolved by the transcriptionist on first review thanks to the gained listening experience. Often the same words or expressions occurred again later on, with the same or a different speaker. In a second review the transcriptionist listened to all utterances again, verified the transcriptions and applied corrections where necessary.

In a third review the remaining unclear cases were shown and played to an operational air traffic controller, which resolved further cases, such as the military call signs, call signs that sounded different to the transcriptionist but were indeed all identical, and not understood particular expressions that are used in ATC communication.

Due to the frequent occurrence of special location names and radio call signs an automatic spell check was not performed. Instead of this, a list of all occurring words was created, which includes a count of occurrence and examples of the context in which the word occurs. Due to the limited vocabulary used in ATC, this list consists of less than one thousand entries including location names, call signs, truncated words, and special mark-up codes. Every item of the list was manually checked and thus typing errors eliminated.

A number of other errors and ambiguities could be resolved through manual consistency checks within the transcriptions. For example, airline designators that were considered unknown due to an altered pronunciation could be identified through cross-checks with other utterances containing the same three-number part of the call sign.

4.6 Speaker Identification

The initial transcription was performed in the same order as the original recordings were made. In this first run every speaker change was marked by the transcriptionist, occurring roughly every two hundred utterances. After all recordings were transcribed and the transcriptionist was very familiar with the differences between the occurring voices, the segments were manually grouped by speaker and every utterance labelled with a speaker identification. Ten speakers were identified.

Based on a previous report [7] it was expected to find eleven speakers, which was not the case. The original author of that report confirmed that the underlying recordings do not necessarily fully coincide with the recordings on which this corpus is based on.

4.7 Session Identification

The utterances were grouped into different sessions. In principle, a session represents one continuous ATC simulation exercise, which lasts approximately one hour. Since no recording protocols were available, the primary criterion for session boundaries was the occurrence of a speaker change. Every change of speaker was considered as the start of a new session. This happens to coincide with the Start ID track markers found on the DAT tapes (Tbl. 3.1). A track marker had been set at every occurring speaker change. Besides a number of spurious track markers there were three occurrences where the speaker did not change but nevertheless a new recording sessions started. These cases were determined and confirmed based on the length of the sessions, an existing track marker, as well as based on the content of the session which indicated that a new simulation exercise started (e.g. change of greetings from 'good morning' to 'good afternoon').

5 Corpus Structure, Format and Distribution

5.1 Distribution

The corpus is publicly available and provided free of charge, except for potential shipping and handling costs. The entire corpus including the recordings and all meta data has a size of approximately 2.5 gigabyte and is available in digital form on a single DVD, or as an electronic ISO disk image at `http://www.spsc.tugraz.at/ATCOSIM`.

Both the DVD and the disk image are based on an ISO9660 Level 2 file system with Rock Ridge extensions. ISO9660 is the standard cross-platform interchange format for CDs and some DVDs, and is understood by virtually all operating systems. Also the Joliet extensions to ISO9660 and a UDF file system are present on the disk image and the DVD.

To obtain a DVD copy of the corpus please contact:

EUROCONTROL Experimental Centre
Horst Hering
Centre du Bois des Bordes
B.P. 15
F-91222 Brétigny-sur-Orge CEDEX
France

5.2 License, Copyright and Disclaimer

content and suitability of this corpus for any purpose. It is provided 'as is' without express or implied warranty. The publishers disclaim all warranties with regard to this database, including all implied warranties or merchantability and fitness. In no event shall the authors be liable for any special indirect or consequential damages or any damages whatsoever resulting from loss of income or profits, whether in an action of contract, negligence or other tortious action, arising in connection with the use or performance of this database.

In particular, it shall be noted that due to the age of the recordings and the given simulation context the language contained in this corpus may not strictly adhere to any official or legal air traffic control phraseology. The corpus is thus not suitable for the training or appraisal of personnel involved in operational air traffic control, such as for air traffic control operator or pilot training.

5.3 Validation

The validation of the ATCOSIM corpus was carried out in October 2007 at the Signal Processing and Speech Communication Laboratory (SPSC) of Graz University of Technology. The examiner and author of the validation report has not been involved in the production of the ATCOSIM corpus, but exclusively carried out an informal pre-validation and the formal final validation of the corpus.

The validation procedure followed the guidelines of the Bavarian Archive for Speech Signals (BAS) [1]. It included a number of automatic tests concerning completeness, readability, and parsability of data, which were successfully performed without revealing errors. Furthermore, manual inspections of documentation, meta-data, transcriptions, and the lexicon were done, which showed minor shortcomings that were fixed before the public release of the corpus. Finally, a re-transcription of 1 % of the corpus data was made, showing a transcription accuracy on word level of 99.4 %, proving the transcriptions to be accurate. The ATCOSIM corpus was therefore considered to be in a usable state for speech technology applications. The validation report is included in the corpus documentation.

5.4 Data Format

The corpus data is provided in four directories, namely WAVdata containing the recordings, TXTdata containing the annotations, HTMLdata containing the annotations in a browsable form and DOC containing documentation.

5.4.1 WAVdata

The WAVdata directory contains the recorded speech signal data. Each file corresponds to one controller utterance. The file format is single-channel Microsoft WAVE with a sample rate of 32 kHz and a resolution of 16 bits per sample. The 10,078 files are

located in a sub-directory structure with a separate directory for each of the ten speakers and sub-directories thereof for each session of the speaker.

The speaker directories are named according to the speaker ID given in Tbl. 2.2, where the first lower-case letter stands for the controller's control centre (geneva, söllingen or zürich), the second letter for the gender of the controller (female or male), and the digit on the third position being a consecutive numbering for controllers with identical gender and control centre.

The session directories are sub-directories of the speaker directories and are named by the speaker ID, followed by underscore, followed by a consecutive two-digit number that identifies the session within that speaker.

The utterance files within the session directories are named by the speaker ID, followed by an underscore, followed by the two-digit session number, followed by a three-digit utterance number within this session. We refer to this sequence as the 'full utterance ID'. The full file name is therefore the full utterance ID, followed by the file extension '.wav'.

For example, the file 'zf2_04_010.wav' is the tenth utterance in the fourth session of the second female Zürich speaker.

5.4.2 TXTdata

All files described herein are text files in plain-text 7-bit ASCII encoding. They are thus also compliant to e.g. ISO/IEC 8859-1 (ISO Latin-1) and Unicode (UTF-8) encoding, as no special characters outside the 7-bit ASCII range are used.

*.txt files

The TXTdata directory contains the same directory structure as the WAVdata directory. It contains a plain-text file for each utterance which consists of the orthographic transcription of the utterance. The file name is the full utterance ID as described above, followed by the file extension '.txt'.

fulldata.csv file

In the root of the TXTdata directory, the file fulldata.csv contains the complete annotation and meta data for all utterances and should be the primary data source when using the corpus. The file is a comma-separated value (CSV) file according to RFC 4180, and should therefore be simple to import in a large number of database programs, spreadsheet programs and programming languages. The CSV file represents the annotation data in a table-like manner. Each line in the file corresponds to one utterance. Lines are terminated by a Unix-style LF (Line feed, 0x0A) newline character.[1] Each line consists of several data fields, which are separated by commas. The data fields itself do not contain any commas, double quotes or newline characters.

[1] If the file occurs as one long line on Microsoft Windows systems, a conversion of the newline character from LF to CR+LF (carriage return followed by a line feed) is necessary.

The first line is a header line which, also separated by commas, briefly describes the meaning of each field (column). The following fields are given:

1. directory
   The directory within TXTdata or WAVdata to which this utterance belongs. Identical to speaker_id.

2. subdirectory
   The session directory within the speaker directory to which this utterance belongs.
   Identical to session_id, but zero-padded at the left to form a two-digit number.

3. filename
   The filename of the utterance file without the .txt or .wav file extension. Identical to the 'full utterance ID' as described above.

4. speaker_id
   The speaker ID.

5. session_id
   The session ID within the speaker.

6. utterance_id
   The utterance ID within the session.

7. transcription
   Orthographic transcription. This is the same data as given in the .txt file of each utterance.

8. recording_corrupt
   Boolean field (0 for false, 1 for true) that indicates that the original recording is technically corrupt and contains audible artifacts. 0 means recording is O.K., 1 means recording is corrupt.

9. comment_transcriptionist
   Text comments provided by the transcriptionist. There is no guarantee that the comments are used in a consistent manner.

10. length_sec
    The length of the utterance in seconds, based on the instances when the PTT button was pressed and released.
    Identical to the length of the utterance's .wav file.

11. recording_id
    The original ID of the recording, indicating the two-digit tape number and one-digit transfer file number (resulting in the three digit file name as given in Tbl. 3.1), followed by an underscore and a four-digit utterance number within the transferred file. For most uses of this corpus this information is irrelevant.

12. recording_startpos_sec
    The original position of the utterance in the transferred file relative to the beginning of the file, in seconds. For most uses of this corpus this information is irrelevant.

wordlist.txt

The wordlist.txt file contains an alphabetically sorted list of all occurring words, including location names, airline radio call-signs, truncated words and special mark-up characters, codes and symbols.

5.4.3 HTMLdata

The files in the HTMLdata directory are HTML files which present the data in a table form so that they can be displayed in a standard HTML web browser. These files are provided purely for convenience and should not be used for further processing, as certain special characters are escaped in the HTML code and also conversion errors might have occurred. The fulldata.csv file should be the primary source of information.

The functionality provided by these files may vary depending on the operating system and web browser used, and also depending on the configuration thereof. The files were tested using Mozilla Firefox 2.0 (with installed Quicktime Plug-in) on Microsoft Windows XP (SP2) and Apple Mac OS X 10.4.10.

As the tables are comparably large, they might take a long time to load. The column headers of the tables use abbreviated titles, with the full titles being shown as tool-tips. A click on the 'Play' field next to each utterance may start a JavaScript which replays the audio of the corresponding utterance in a separate browser window or tab. In those tables that are dynamic, a single-click on one of the column headers sorts the entire table according to this column. On a current state-of-the-art desktop computer[2] the JavaScript-based sorting of the dynamic tables takes approximately one minute.[3]

fulldata_static.htm The file contains the same information as included in the fulldata.csv file, but presented as a static HTML table.

fulldata_dynamic.htm The file contains the same information as included in the fulldata.csv file, but presented as a dynamically sortable HTML table.

[2] System configuration : 2.2 GHz Intel Core 2 Duo, 4 GB RAM, Mac OS X, Firefox

[3] If Mozilla Firefox warns about an unresponsive script and this warning should be disabled, setting the 'dom.max_script_run_time' preference in 'about:config' to 0 will allow the sorttable script to run for as long as it needs.

overview_sortedby_*.htm The files show only the most relevant data fields. This provides a better overview and more space to display the actual transcription. The data is presented in a static HTML table, which is pre-sorted according to the criterion indicated by the filename.

overview_dynamic.htm The file shows only the most relevant data fields, presented as a dynamically sortable HTML table.

wordlist.htm The file provides a list of all occurring words including location names, airline radio call-signs, truncated words and special mark-up characters, codes and symbols. It also includes the number how often each word occurred in the corpus and provides an exemplary list of utterances that contains the corresponding word. The same functionality as above to replay the utterance is included.

5.4.4 DOC

The DOC directory contains all documentation related to the corpus. Documents in the Adobe PDF format can be opened using for example Ghostscript[4] or Adobe Reader[5].

readme.pdf An introductory document.

license.txt License terms and disclaimer in plain text format.

atcosim_report.pdf Complete documentation of the corpus (this document).

filelist.txt Full listing of all files and directories in the ATCOSIM distribution in plain text format.

eec_simulation Directory containing documents about the simulation during which ATCOSIM was recorded.

papers Directory containing publications about the ATCOSIM corpus.

validation Directory containing the validation report.

artwork Directory containing a DVD jewel case inlet.

[4] Available online at `http://pages.cs.wisc.edu/~ghost/`
[5] Available online at `http://www.adobe.com/products/acrobat/readstep2.html`

6 Conclusion

The previous chapters introduced the ATCOSIM Air Traffic Control Simulation Speech Corpus and described in detail the production process of the corpus.

Application Examples

The corpus can be utilised within different fields of speech-related research and development in air traffic control.

ATC Language Study Analysis can be undertaken on the language used by controllers and on the instructions given.

Speech Recognition The corpus can be used for the development and training of speech recognition systems. Such systems might become useful in future ATC environments and for example be used to automatically input controller instructions given to pilots into the ATC system.

Listening Tests The speech recordings can be used as a basis for ATC-related listening tests and provide real-world ATC language samples.

Speaker Identification The corpus can also be used as testing material for speaker identification and speaker segmentation applications in the context of ATC.

Extension of Corpus Annotation

Depending on the application, further annotation layers might be useful and could be added to the corpus.

Phonetic Transcription A phonetic transcription is beneficial for speech recognition purposes. For accurate results it requires transcriptionists with a certain amount of experience in phonetic transcription or at least a solid background in phonetics and appropriate training.

Word and Phoneme Segmentation A more fine-grained segmentation is also beneficial for speech recognition purposes. It can often be performed semi-automatically, but still requires manual corrections and substantial effort.

Semantic Transcription A semantic transcription would describe in a formal way the actual meaning of each utterance in terms of its functionality in air traffic

control, such as clearance and type of clearance, read-back, request, ... This would support speech recognition and language interface design tasks, as well as ATC language studies. The production of such a transcription layer requires good background knowledge in air traffic control. Due to lack of contextual information, such as the pilots' utterances, certain utterances might appear ambiguous.

Call Sign Segmentation and Transcription. This transcription layer marks the signal segment in which the call sign is extracts the corresponding part of the (already existing) orthographic transcription. This can be considered as a sub-part of the semantic transcription which can be achieved with significantly less effort and requires little ATC related expertise. Nevertheless this might be beneficial for certain applications such as language interface development.

Extension of Corpus Size and Coverage

With an effective size of ten hours of control speech the corpus might be too small for certain applications. There are two reasons that would support the collection and transcription of more recordings. The first reason is the pure amount of data that is required by the training algorithms in modern speech recognition systems. The second reason is the need to extend the coverage of the corpus in terms of speakers, phonological distribution and speaking style, as well control task and controlled area.

A Selection of Publicly Available Speech Corpora

- BNC World
  - http://www.natcorp.ox.ac.uk/corpus/index.xml.ID=products
  - BNC World is a revised version of the original British National Corpus which contains 100 million words: 90% written, 10% orthographically transcribed spoken text.
- CSR WSJ Corpora (Wall Street Journal)
  - http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC93S6A
  - The first two CSR Corpora consist primarily of read speech with texts drawn from a machine-readable corpus of Wall Street Journal news text and are thus often known as WSJ0 and WSJ1. The texts to be read were selected to fall within either a 5,000-word or a 20,000-word subset of the WSJ text corpus. Some spontaneous dictation is included in addition to the read speech. The dictation portion was collected using journalists who dictated hypothetical news articles.
- TIMIT Acoustic-Phonetic Continuous Speech Corpus
  - http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC93S1
  - The TIMIT corpus of read speech is designed to provide speech data for acoustic-phonetic studies and for the development and evaluation of automatic speech recognition systems. TIMIT contains broadband recordings of 630 speakers of eight major dialects of American English, each reading ten phonetically rich sentences.
- NTIMIT
  - http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC93S2
  - The NTIMIT corpus was developed by the NYNEX Science and Technology Speech Communication Group to provide a telephone bandwidth adjunct to TIMIT.
- SpeechDat
  - http://www.speechdat.org/
  - The aim of the SpeechDat data collections is to establish speech databases for the development of voice operated teleservices and speech interfaces.

- TIDIGITS
  - http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC93S10
  - The TIDIGITS corpus was originally designed for the purpose of designing and evaluating algorithms for speaker-independent recognition of connected digit sequences. There are 326 speakers (111 men, 114 women, 50 boys and 51 girls) each pronouncing 77 digit sequences.
- The Numbers Corpus
  - http://cslu.cse.ogi.edu/corpora/corpCurrent.html
  - A collection of naturally produced numbers taken from other CSLU telephone speech data collections, and include isolated digit strings, continuous digit strings, and ordinal/cardinal numbers.
- Alphadigit
  - http://cslu.cse.ogi.edu/corpora/corpCurrent.html
  - A collection of 78,044 examples from 3,025 speakers saying 6 digit strings of letters and digits over the telephone.
- AURORA Project Database 2.0
  - http://catalog.elra.info/product_info.php?products_id=693&osCsid=f1c9f301975cea7150b6a728b7211394
  - This revised version of the Noisy TI digits database is intended for the evaluation of algorithms for front-end feature extraction algorithms in background noise but may also be used to evaluate and compare the performance of noise robust speech recognition algorithms.

The following websites provide extensive lists of further corpora.[1]

- ELDA - Evaluations and Language resources Distribution Agency http://www.elda.org/
- LDC - Linguistic Data Consortium http://www.ldc.upenn.edu/
- OGI CSLU Corpora Group http://cslu.cse.ogi.edu/corpora/
- W3-Corpora List of Corpora http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/corpora/list/index2.html
- Corpus Resources http://leo.meikai.ac.jp/~tono/resources.html
- Texts and corpora list http://torvald.aksis.uib.no/corpora/sites.html
- David Lee's Bookmarks for Corpus-based Linguists http://devoted.to/corpora

[1] The availability of these sites was last checked in September 2007. Archived copies of these listings are usually available at the Internet Archive http://www.archive.org/

Bibliography

[1] Florian Schiel and Christoph Draxler. *Production and Validation of Speech Corpora*. Bastard Verlag München, 2003.

[2] Xuedong Huang, Alex Acero, and Hsiao-Wuen Hon. *Spoken Language Processing (A Guide to Theory, Algorithm, and System Development)*. Prentice Hall, 2001.

[3] *Manual of Radiotelephony*. Number 9432 in AN/925. International Civil Aviation Organization (ICAO), 3 edition, 2006.

[4] *Designators for Aircraft Operating Agencies, Aeronautical Authorities and Services*. Number 8585. International Civil Aviation Organization (ICAO), 138 edition, 2006.

[5] The european ais database (ead) [online, cited September 2007]. Available from World Wide Web: `http://www.ead.eurocontrol.int/`.

[6] *Digital Aeronautical Flight Information File (DAFIF)*. Number 0610. National Geospatial-Intelligence Agency, 6 edition, October 2006. Electronic database.

[7] Horst Hering. Technical analysis of ATC controller to pilot voice communication with regard to automatic speech recognition systems. EEC Note 01/2001, Eurocontrol Experimental Centre, 2001.

[8] Olivier Grisvard. SCOPE: Safety of controller-pilot dialogue. Project deliverables WP1–WP4, Eurocontrol Experimental Centre, 2003. Available from World Wide Web: `http://www.eurocontrol.int/care-innov/public/standard_page/innov2_scope.html`.

[9] Olivier Grisvard. ESCALE: Enhanced speech tracking of air traffic control communications. Project deliverables WP1–WP3, Eurocontrol Experimental Centre, 2005. Available from World Wide Web: `http://www.eurocontrol.int/care-innov/public/standard_page/innov2_escale.html`.

[10] Linguistic Data Consortium. NIST air traffic control complete [online, cited September 2007]. Available from World Wide Web: `http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC94S14A`.

[11] J.C. Segura, T. Ehrette, A. Potamianos, D. Fohr, I. Illina, P-A. Breton, V. Clot, R. Gemello, M. Matassoni, and P. Maragos. The HIWIRE database, a noisy and non-native english speech corpus for cockpit communication [online, cited September 2007]. Available from World Wide Web: `http://www.hiwire.org/`.

[12] Stephane Pigeon, Wade Shen, and David van Leeuwen. Design and characterization of the non-native military air traffic communications database (nnMATC). In *Proceedings of the International Conference on Spoken Language Processing (INTERSPEECH)*, Antwerp, Belgium, September 2007.

[13] L. Graglia, B. Favennec, and A. Arnoux. Vocalise: assessing the impact of data link technology on the r/t channel. In *Digital Avionics Systems Conference, 2005. DASC 2005. The 24th*, volume 1, October 2005.

[14] VOCALISE - the today use of VHF as a media for pilots/controllers communications [online, cited September 2007]. Available from World Wide Web: `http://www.cena.aviation-civile.gouv.fr/divisions/ICS/projets/vocalise/index_en.html`.

[15] Roger Lane, Robin Deransy, and Diena Seeger. 3rd continental RVSM real-time simulation. EEC Report 315, Eurocontrol Experimental Centre, 1997.

[16] Horst Hering. Stif interface (speech techniques for simulation facilities). EEC Note 25/96, Eurocontrol Experimental Centre, 1996.

[17] Paul Boersma and David Weenink. PRAAT: doing phonetics by computer [computer program] [online]. 2007 [cited September 2007]. Available from World Wide Web: `http://www.praat.org/`.

[18] David Lee. Bookmarks for corpus-based linguists [online, cited September 2007]. Available from World Wide Web: `http://devoted.to/corpora`.

[19] Joaquim Llisterri. Speech analysis and transcription software [online, cited September 2007]. Available from World Wide Web: `http://liceu.uab.es/~joaquim/phonetics/fon_anal_acus/herram_anal_acus.html`.

[20] Annotation graph toolkit [online, cited September 2007]. Available from World Wide Web: `http://sourceforge.net/projects/agtk`.

[21] Kazuaki Maeda, Steven Bird, Xiaoyi Ma, and Haejoong Lee. Creating annotation tools with the annotation graph toolkit. In *Proceedings of the Third International Conference on Language Resources and Evaluation*, Paris, 2002. European Language Resources Association.

[22] Wavesurfer [online, cited September 2007]. Available from World Wide Web: `http://www.speech.kth.se/wavesurfer/`.

[23] Autohotkey - free mouse and keyboard macro program with hotkeys and autotext [online, cited September 2007]. Available from World Wide Web: `http://www.autohotkey.com/`.

[24] *Designators for Aircraft Operating Agencies, Aeronautical Authorities and Services*. Number 8585. International Civil Aviation Organization (ICAO), 107 edition, 1998.

[25] *Designators for Aircraft Operating Agencies, Aeronautical Authorities and Services.* Number 8585. International Civil Aviation Organization (ICAO), 93 edition, 1994.

[26] *Location Indicators.* Number 7910. International Civil Aviation Organization (ICAO), 122 edition, 2006.

[27] Diena Seeger and Hugh O'Connor. S08 ANT-RVSM 3rd continental real-time simulation pilot handbook. Eurocontrol Internal Document, December 1996.