# *Author:* Jindřich Matoušek, Daniel Tihelka

# *Title:* English TTS speech corpus of air traffic (pilot) messages - Serbian accent

*Identifier:* ZCU\_TTS\_ITBLP\_EN\_SR\_PM

*Subject keywords:* speech corpus; text-to-speech (TTS), pitch-marks

## Description:

The corpus contains recordings of male speaker, native in Serbian, talking in English. The sentences that were read by the speaker originate in the domain of air traffic control (ATC), specifically the messages used by plane pilots during routine flight. The text in the corpus originates from the transcripts of the real recordings, part of which has been released in LINDAT/CLARIN (<http://hdl.handle.net/11858/00-097C-0000-0001-CCA1-0>), and individual phrases were selected by special algorithm described in Jůzová, M. and Tihelka, D.: *Minimum Text Corpus Selection for Limited Domain Speech Synthesis* (DOI 10.1007/978-3-319-10816-2\_48). The corpus was used to create a limited domain speech synthesis system capable of simulating a pilot communication with an ATC officer.

## Technical information:

number of files: 3000

sampling frequency: 16kHz

bit depth: 16bit

## Content:

*ZCU\_TTS\_ITBLP\_EN\_SR\_PM\_speech.zip*

... speech signal (wav files)

*ZCU\_TTS\_ITBLP\_EN\_SR\_PM\_pitch-marks.zip*

... glottal closure instants (corresponding to peaks in the speech signal) marked by:

* V ... in voiced parts of speech
* U ... in unvoiced parts of speech (placed artificially with period approx. 4ms)
* T ... transient (artificial) pitch-marks place in the middle between V-U or U-V pitch-marks

*annot.snt* ... world-level annotations of speech

*alignment.mlf*

... phoneme-level segmentation in HTK's MLF (master label file)

format

*dict* ... phonetic dictionary (contrary to *annot.snt*, all words are lowercased in the dictionary)

*phone\_set* ... phonetic alphabet.

The used phone set consists of 40 phones, based on Arpabet symbols (<http://en.wikipedia.org/wiki/Arpabet>). Lexical stress is not counted in. Moreover, long pause (typically utterance leading and trailing silences) is marked as "$" and short pause (typically cross-word pauses) is marked as "#". Optionally, loud breath is marked as "%".

## Special licenses:

## Some parts of the corpus (namely dict) are derived from works released under their own copyrights, see copyright\_CMU and copyright\_BEEP for more details.

## References:

Detection of pitch-marks:

* Legát, M. and Matoušek, J. and Tihelka, D. : On the detection of pitch marks using a robust multi-phase algorithm. Speech Communication, p. 552-566, Elsevier, 2011.

Automatic phone-level segmentation of speech:

* Matoušek, J. and Romportl J. : Automatic Pitch-Synchronous Phonetic Segmentation. INTERSPEECH 2008, proceedings of 9th Annual Conference of International Speech Communication Association, p. 1626-1629, ISCA, Brisbane, Australia, 2008.
* Matoušek, J. and Tihelka, D. and Psutka, J. : Automatic segmentation for Czech concatenative speech synthesis using statistical approach with boundary-specific correction . Eurospeech 2003 - Interspeech, proceedings of the 8th European Conference on Speech Communication and Technology, p. 301-304, ISCA, Geneva, Switzerland, 2003.

Selection of phrases to the corpus:

* Jůzová, M. and Tihelka, D.: **Minimum Text Corpus Selection for Limited Domain Speech Synthesis.** Text, Speech, and Dialogue, 17th International Conference, TSD 2014, Lecture Notes in Artificial Intelligence, vol. 8655, p. 398-407, Springer, 2014.

Limited-domain speech synthesis using the corpus:

* Jůzová, M. and Tihelka, D.: **Tuning Limited Domain Speech Synthesis Using General Text-to-Speech System .** Text, Speech, and Dialogue, 17th International Conference, TSD 2014, Lecture Notes in Artificial Intelligence, vol. 8655, p. 408-415, Springer, 2014.