

Content of the BulPhonC Corpus Version 1

The BulPhonC Version 1 contains the following:

1. Directory *speakers*

The directory *speakers* contains a subdirectory for each speaker. The name of the subdirectory for a speaker coincides with the speaker's identifier. The subdirectory for each speaker contains the speaker's utterances and their corresponding phonetic annotations. The file name of each utterance has the form *number_of_sentence.wav* and the file name of the corresponding phonetic annotation has the form *number_of_sentence.TextGrid* where *number_of_sentence* is the identifier of the sentence that has been read by the speaker.

2. Directory *sentences*

In the directory *sentences* for each sentence there is a file with name *number_of_sentence.txt* where *number_of_sentence* is the identifier of the sentence. The encoding of *number_of_sentence.txt* is UTF-8.

3. A file *speakers-info.txt*

This file is in tab-separated format and contains for each speaker the speaker's gender (M or F) and the speaker's age.

4. A file *phoneticSystem.pdf*

This file describes the phonetic system that has been used for the phonetic annotations.