Nasjonalbiblioteket

The NST text-to-speech dataset

The Norwegian Language Bank

At the turn of the millennium, the firm Nordisk språkteknologi (NST) developed a dataset for text-to-speech in Norwegian Bokmål. This dataset consists of close to 8 hours of recordings of a male speaker of Eastern Norwegian reading from a manuscript. Most of the segments are sentences, but there are also some numbers and other non-sentences utterances. Everything is in Bokmål The Language Bank has distributed this dataset since 2011: https://www.nb.no/sprakbanken/en/resource-catalogue/oai-nb-no-sbr-15/

However, the formats of the audio files and metadata files and the lack of documentation made it difficult to make use of the dataset. We have therefore made a new release with updated audio and metadata files.

This release consists of a metadata file *nst_tts_dataset.jsonl* and two directories *channel_1* and *channel_2* containing 5363 audio files each with the audio from the two channels of the original recordings.[1] There is one audio file per segment. The metadata file contains one line per segment with the duration of the audio file in seconds, the transcription, and the relative path to the channel one and channel two recordings. The audio format is 22kHz wav.[2]

[1] According to the original documentation, one channel contains audio from a microphone and the other contains signals from a laryngograph, but from manual inspection of selected audio files, both channels appear to contain audio from a normal microphone.

[2] The original documentation claims that the audio has a sampling frequency of 44kHz, but this is not correct.