# Transcription Documentation

The audio recordings in Part 1 and Part 2 have been transcribed orthographically, as accurately as possible. Transcriptions can be found in the \SCRIPT folders of every channel. Transcripts follow a 6-digit format, where the first number denote the channel number. The next four numbers correspond to the speaker ID that mark each audio recording folder (in \WAVE), and the last number denotes the session number.

Transcriptions have been done verbatim and are generally in lowercase. Only proper nouns (such as names, addresses, months, etc.) and abbreviations have been capitalized. Additionally, initialisms are also separated by spaces (e.g. *G S T* and *E R P*).

Number sequences are spelled out according to what has been said, and punctuations are generally omitted unless it is a necessary part of a written form, such as the apostrophe in *Mary’s*. Where punctuations have been read out, transcriptions were done corresponding to what is being read (e.g. *ham + cheese* would be transcribed as *ham plus cheese* if that was what was read).

Unintelligible speech is marked by two asterisks (\*\*).

Distinguishable non-speech acoustic events have also been marked by the following tags.

|  |  |
| --- | --- |
| **Tags** | **Definition** |
| <FIL/> | Fillers (e.g. *oh, ah, uh, um*) |
| <SPK/> | Noises from the speaker that are not intended in the prompted text, such as coughing, sighing or loud breathing. |
| <STA/> | Constant background noises, such as rain or traffic noise. |
| <NON/> | Intermittent non-human noises such as door slams or mouse clicks. |
| <NPS/> | Noises that were made by someone other than the speaker. |