Speech Transcription Guidelines

Revision v0.1.5

General

1. Transcript files should be plain text
2. The reference string for an audio file should be typed as a single line with no embedded line feeds
3. Word strings should be all lower-case
4. Do not add punctuation
5. Do use apostrophes (‘) for contractions and possessives

Disfluency Markers

1. **Filled Pauses** – Use + to delimit filled pauses. The set we currently use is: **+um+, +er+, +uh+, +ah+, +hm+**. You can make up additional ones if none of these fits.

2. **Non-speech events** – Delimit non-speech events with <>. The set we currently use is:
    a. **<breath>** indicates a relatively loud breath event like big inhale or sigh, not just normal breathing
    b. <laugh>
    c. <cough>
    d. **<noise>** is used for a relatively loud punctual signal, like a door slam, not for steady state background noise
    e. **<side_speech>** is for soft speech spoken to or by another person, rather than the subject speaking to the system
    f. **<no_signal>** indicates an empty audio file
    g. **<silence>** used only if there is an audio signal, but it is only silence or background noise for the whole file
3. Truncated words – Indicate truncated words by typing the part of the word that was spoken followed by a hyphen, as in **wh-**
4. Unintelligible or severely mispronounced – indicate unintelligible speech with and asterisk in parentheses **(*)**

Example Transcriptions

1. Utterance with all variations
   `<breath>` when he came (*) he s- he said +uh+ why don’t we call him <laugh>
2. For an empty file
   `<no_signal>`
3. For just background noise
   `<silence>`