- We use the SEAME corpus, which is a conversational Mandarin-English code-switching speech corpus

- We cleaned data, removed and corrected words with ambiguity and retained words with proper available pronunciations (removing non-speech utterances)

- A total of $\cong$ 15% data was removed in the cleaning process

Data Description

Mandarin-English code-switching speech corpus

- We use the SEAME corpus, which is a conversational ambiguity and retained words with proper available

- We cleaned data, removed and corrected words with

- A total of ≅ 15% data was removed in the cleaning pronunciations (removing non-speech utterances) process

Data Description