Automatic measures of fluency in L2 speech: OASIS Summary of De Jong, Pacilly, & Heeren,(2021) in Assessment in Education: Principles, Policy & Practice https://oasis-database.org
Jong, N.H. de

Automatic measures of fluency in L2 speech

What this research was about and why it is important

Fluency reflects automaticity and speed of speech production and is a key construct in communicative speaking ability. Aspects of fluency therefore need to be apparent in speaking assessment criteria. However, research linking different aspects of fluency to levels of speaking proficiency is lacking. One reason for this lack in research is the fact that measuring the separate aspects of fluency by hand is very time-consuming. To do more detailed research into the relation between specific aspects of fluency and second language (L2) proficiency, as well as to assess fluency automatically, the authors created automated scripts to measure fluency automatically, without use of transcriptions. The new script gathered information on filled pauses ("uh" and "uhm") and was added to an existing script that already measured silent pausing and speed of speech. The script to measure filled pauses was built and tested on existing Dutch and English L2 corpora. To validate the script further, it was subsequently tested on another corpus, including first language (L1) and L2 Dutch and English speech. Finally, the validity of the automatic measures of fluency for the purpose of language assessment was gauged, by relating the outcomes of the script to judgements on fluency. It showed that filled pauses can be detected in both Dutch and English L2 speech by the new script, to a certain extent. The largest mismatch between manual detection and automatic detection is that too many 'normal' syllables are indicated as being filled pauses by the script.

What the researchers did

- An existing script to measure silent pauses and speed of speech was updated.
- An algorithm to detect filled pauses (e.g., "uh" and "uhm") in a new script was created using subsets of L2 Dutch and English corpora, in which filled pauses were manually annotated.
- The algorithm detects filled pauses on the basis of their acoustics.
- The algorithm was tested on new subsets of the corpora that were not used in creating the algorithm, by comparing automatic detection of filled pauses to manual detection of filled pauses.
- The script was further validated by comparing manual annotations of filled pauses to automatic measures of filled pauses on a new corpus of L1 and L2 English and Dutch.
- Correlations between judgements on fluency and measures of manual counts of filled pauses were compared to correlations between judgements on fluency and measures of automatic counts.

What the researchers found

- The algorithms for L2 Dutch and L2 English were slightly different, but in both cases, it turned out that a filled pause would be detected when, compared to a 'normal' syllable, the syllable is
    - long in duration,
    - like an "uh"-sound
    - pronounced with a relatively low pitch,
    - pronounced without change, stable
    - pronounced lazily
- The accuracy of the script is promising, especially because the accuracy was similar for the new corpus, which was not used in the creation of the algorithms.
- For both Dutch and English, too many normal syllables are detected as being filled pauses (false positives).
- Automatic measures and manual measures similarly predicted judgements on fluency.

Things to consider

- The many false positives may be syllables that are pronounced hesitantly (such as lengthened syllables).
- The corpora in the study were ecologically valid choices (e.g., real recorded assessment data), which also means that sound quality was sometimes below the high standards needed for precise acoustic measurements.
- Future research is needed to find out whether filled pauses can be detected for more languages than English and Dutch, and to find out to what extent L2 proficiency plays a role in the accuracy of the algorithms.
- For use in classroom (self-)assessment, the scripts would need to be rebuilt in an easy-to-use app, because the current implementation requires experience with the software PRAAT (https://www.fon.hum.uva.nl/praat/).