Derwing, T. M., Munro, M. J., & Carbonaro, M. (2000). Does popular speech recognition software work with ESL speech? *TESOL Quarterly, 34(3),* 592–603. https://doi.org/10.2307/3587748

The suitability of consumer dictation software for providing feedback to L2 speakers

What this research was about and why it is important

Some researchers have advocated the use of popular dictation software as a tool to assist L2 learners with their pronunciation. The argument is that a learner could speak into a computer and obtain transcribed output showing where pronunciation errors occurred. The authors assessed the accuracy of Dragon Naturally Speaking Preferred (1997) for high proficiency adult English speakers whose native languages were Cantonese and Spanish. Their goal was to determine whether the software could identify pronunciation problems that affect human listeners' understanding of nonnative speech. If a learner is to benefit from computer feedback on errors, that feedback must be as "human-like" as possible. The software should misunderstand the learner's speech whenever a human listener would and should do so in the same way that a human would. To achieve their goal, the investigators compared the output of the computer software with human transcriptions of the same spoken utterances from a set of speakers, both native English and L2 learners. The software's recognition of the nonnative speech was much less accurate than human transcriptions. Furthermore, the recognition errors from the software bore no useful resemblance to errors made by the human transcribers.

What the researchers did

- Speech samples were provided by 10 native speakers of Spanish and 10 native speakers of Cantonese, all of whom had immigrated to Canada after age 18. Their TOEFL scores ranged from 543 to 650. Ten speakers of Canadian English also participated to provide comparison sentences. The mean age of all speakers was 37.5 years.
- The listeners, who had a mean age of 27.5 years, were 41 native speakers of Canadian English.
- The software was trained on the speakers' voices using instructions in the commercial software package.
- The researchers recorded the speakers reading 60 true and false sentences, which were simultaneously presented to the computer. One true and one false sentence were selected from each speaker for presentation to the human listeners. They transcribed the sentences and rated them for comprehensibility (how easy they were to understand) and accentedness (how foreign-sounding they were).
- For each sentence, the computer's score was calculated by determining the percentage of words it correctly recognized. The percentage of words correctly transcribed by the listeners was the humans' intelligibility score.
- The accuracy of the vowels and consonants transcribed by the computer and the humans was also compared.

What the researchers found

- The software recognized native-produced sentences at the advertised accuracy level: 90% of the words. The human listeners transcribed the same sentences 99.7% correctly. The computer's intelligibility score for the Cantonese speakers was 72%; for the Spanish speakers it was 71%. Corresponding rates for the humans were 95% and 96%.
- An analysis of the computer and human intelligibility scores revealed no meaningful relationship between them.
- Furthermore, the computer score bore no relationship to the listeners' comprehensibility and accentedness ratings.
- There was also no significant relationship between the vowel and consonant errors made by the speakers and the computers' output. In contrast, the L2 speakers' error rates were significantly correlated with the listeners' intelligibility scores and their comprehensibility ratings.

Things to consider

- The authors concluded that the software did not recognize the oral language of the Cantonese and Spanish speakers at an acceptable level. Furthermore, this version of Dragon Naturally Speaking Preferred (1997) was unsuitable as a pedagogical tool for pronunciation purposes given its failure to provide human-like feedback.
- To inform language learners' pronunciation, recognition software should pinpoint pronunciation difficulties that cause problems for humans.
- Pronunciation learners are likely to benefit most from software that is specifically designed to meet their needs rather than "off-the-shelf" products.
- Since this study was conducted, advances have been made in ASR directed specifically at ESL speakers.

How to cite this summary: Derwing, T. M., Munro, M. J., & Carbonaro, M. (2018). Can speech recognition software be used for evaluating ESL speech? *OASIS Summary* of Derwing, Munro, & Carbonaro (2000) in *TESOL Quarterly*. https://oasis-database.org