Using Voice Recognition for Assessment and Pedagogy in Second Language Learning (SLA)

Jesse Elam

Speaking evaluations in Second Language Acquisition (SLA) can prove to be problematic and time consuming for many educators. However, thanks to the ubiquitous nature of digital devices, voice recognition software, and artificial intelligence, we are feasibly living in an age where teachers have the ability to evaluate their students' speaking ability automatically with custom software (e.g. Christensen, Hendrickson, & Lonsdale, 2010; De Wet, Muller, Van der Walt, & Niesler, 2011; Lonsdale & Christiansen, 2011). To this end, Google Speech API has been a key interest for developing Elicited Imitation (EI) tests and other speaking activities to promote pedagogical benefits.

EI tests have a long history in SLA and draw on the cognitive information processing theory. Based in terms of cognitive psychology, this theory attempts to explain how people process and store information (Driscoll, 2005). One aspect that links the cognitive information processing theory to EI tests is the notion of chunking (Abney, 1991). That is, EI tests require second language (L2) learning students to listen to an audio prompt and repeat what was heard under a restricted amount of time, which forces them to use their working memory in unison with long-term grammatical knowledge prevalent in their interlanguage system (DeKeyser, 2001). From a semantic standpoint, then, how we chunk linguistic items together can be used as a way to test if we have mastered a certain grammatical function. In this regard, "When an L2 learner has difficulty reproducing a grammatical feature contained in a stimulus sentence this is believed to be due to the feature still not being fully automatized as part of the learner's interlanguage knowledge" (Ashwell & Elam, 2017, p. 63). Hence, through a simple listen and repeat activity which focuses on specific grammatical items multiple times, an EI test is theorized to have the capacity of evaluating students' comprehension and productive skills, as well as their spontaneous language capacity in their L2 (Purpura, 2004).

Although EI tests are attractive for the assessment of students' grammatical skills in oral production, analyzing the data through analogous means brings into account issues of rater bias and requires administrators to utilize a large amount of time in order to analyze the results; hence, many in the field of SLA have been looking for a way to automate the process. In this respect, many researchers have turned to Sphinx Speech Recognition software to build EI tests (see Christensen, Hendrickson, & Lonsdale, 2010; De Wet, Muller, Van der Walt, & Niesler, 2011; Lonsdale & Christiansen, 2011). Nevertheless, Sphinx requires a deep understanding of programming as well as technical knowledge of how to train the system with acoustic models for speech recognition (Twiefel et al., 2014). In this way, Sphinx makes it quite difficult for novice EI researchers to approach the field as well as limiting the accuracy of the students' transcribed inputs due to the constraints inherent to storing the acoustic model data on the host computer. "[However,] systems, like Google's ASR, no longer need to be reliant on the data stored locally on the computer, as it had the

ability to transcribe speech-to-text in real-time, making the number of identifiable words seemingly limitless" (Ashwell & Elam, 2017, p. 61).

Using a custom program, Ashwell and Elam (2017) investigated the accuracy of Google Speech API to assess whether or not it would be suitable to build EI tests or pedagogical tasks using Google Speech ASR in the future. They found that the Google Speech API was almost 90% accurate for transcribing native speakers' oral production, while only being around 66% accurate for that of Japanese students' (2017). Nonetheless, a number of the issues are thought to be a result of the limitations of the data gathering, system design, and EI test items. Hence, in the next iteration of the research project, the authors intend to make adjustments to v3 of the software in order to execute EI tests using Google Speech API to its fullest capability. These changes will include: a) using inter-rater reliability tests to correlate the data with the results from the Google Speech API, b) refining the user interface to limit mistaken inputs, and c) screening all items with native speakers before administration of L2 speakers to ensure reliability.

References

Abney, S. P. (1991). Parsing by chunks. In R. C. Berwick, S. P. Abney, & C. Tenny (Eds.), *Principle-based parsing: Computation and psycholinguistics* (pp. 257–278). Dordrecht, NL: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-011-3474-3_10

Ashwell, T., & Elam, J. R. (2017). How accurately can the Google Web Speech API recognize and transcribe Japanese L2 English learners ' oral production ? *JaltCall Journal, 13*(1), 59–76.

Christensen, C., Hendrickson, R., & Lonsdale, D. (2010). Principled construction of elicited imitation tests. *Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC'10),* 233–238.

De Wet, F., Muller, P., Van der Walt, C., & Niesler, T. (2011). Readability index as a design criterion for elicited imitation tasks in automatic oral proficiency assessment. In *ISCA International Workshop on Speech and Language Technology in Education (SLaTE 2011)* (pp. 24–26). Venice, Italy.

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125–151). New York: Cambridge University Press.

Driscoll, M. P. (2005). Psychology of learning for instruction. Boston, MA: Pearson.

Lonsdale, D., & Christiansen, C. (2011). Automating the scoring of elicited imitation tests. *Symposium on Machine Learning in Speech and Language Processing.*

Purpura, J. E. (2004). *Assessing grammar.* New York: Cambridge University Press.

Twiefel, J., Baumann, T., Heinrich, S., & Wermter, S. (2014). Improving domain-independent cloud-

based speech recognition with domain-dependent phonetic post-processing. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI-14)* (pp. 1–7).