

Developing a comprehensive pedagogical framework for pronunciation training based on blended learning and adapted Automatic Speech Recognition systems

Saandia Ali, LIDILE EA3874, Université Rennes 2

Submission Type: Research and Development Paper (30 minutes)

Status: reviewed and accepted for *Eurocall 2016* (In Limassol, Cyprus, August 24-27)

Topic(s):

- Materials and modes of CALL
- Designs and pedagogy for CALL
- Autonomous and lifelong learning

Keywords: Computer assisted pronunciation training, speech recognition, blended learning, phonetic accuracy

Abstract

This paper reports on the early stages of a locally funded research and development project taking place in a French university. It aims at developing a comprehensive pedagogical framework for pronunciation training for adult learners of English. This framework will combine a direct approach to pronunciation training (Face-to-Face teaching) with online instruction using and adapting existing ASR systems.

The sample of learners chosen for the study, are university students majoring in Arts, Literature or Communication at graduate and undergraduate level. These students have generally been learning English for 7 years prior to entering university and might show an advanced mastery of grammar and syntax, but their spoken English remain heavily accented and may hinder effective communication. Furthermore, the classes are overcrowded (up to 40 students per group) and the emphasis is placed on fluency and communication skills rather than phonetic accuracy. In addition to that, most teachers don't feel confident with teaching pronunciation as they often haven't received any training themselves.

Under these circumstances, students experience performance anxiety, and they only have a limited amount of time for teacher-student interaction and individualised feedback. Language learning appears most efficient when the teacher constantly monitors progress to guide remediation or advancement.

Computer Assisted Pronunciation Training programs (CAPT, Abuseleek 2007) could help realising these goals by offering individual practice and feedback in a safe environment.

A considerable body of research has already shown the efficacy of ASR systems for pronunciation training (Hincks 2002, Kim 2006 or Elimat 2014). Recent ASR based CAPT programs include Subarashii (Entropic HTK recognizer), VILTS (SRI recognizer), FLUENCY (Carnegie Mellon University SPHINX recognizer), Naturally Speaking (Dragon Systems), and FluSpeak (IBM ViaVoice recognizer).

We intend to build on these existing programs and on previous research to develop a set of tools to address bad pronunciation habits among French learners of English.

This approach is based on intensive hybrid tuition, starting from teaching English phonology, pronunciation rules and contrastive analysis in the classroom and then developing online courses with embedded ASR systems for autonomous learning with automatic corrective feedback at both segmental (phone) and suprasegmental (intonation) level. Pronunciation is more cognitive than articulatory and we believe that such an approach could provide both a cognitive input so as to help students become more aware of their pronunciation habits and opportunities for practice and feedback.

The ideal adapted ASR system should include error detection, scoring of pronunciation, full diagnosis (error visualisation and analysis) and remediation tools. It should also accommodate different levels of proficiency (priorities being set for each level and for each learner profile). Reading tasks and virtual conversations based on elicited language will be used for the online course. Finally, most CAPT systems present part of the feedback through the recording and visual representation of learners' performance which is compared to the recording of a native speaker. Commonly used spectrograms will be enriched with modelled representations that are closer to learners' perception using pitch modelling algorithms such as Momel (Modelling Melody, Espesser & Hirst 1993).

References

- AbuSeileek, A. (2007). Computer-based pronunciation instruction as an effective means for teaching stress. *The jalt call Journal*, 3 (1-2) 3-14.
- Elimat, A. K., & Abuseileek, A. F. (2014). Automatic speech recognition technology as an effective means for teaching pronunciation. *The JALT CALL Journal*, 10 (1), 21-47.
- Eskenazi, M. (1999). Using a computer in foreign language pronunciation training: What advantages? *CALICO Journal*, 16(3), 447-470.
- Eskenazi, M., Ke, Y., Albornoz, J., & Probst, K. (2000). The Fluency Pronunciation Trainer: Update and user issues. In *Proceedings of InSTIL 2000, Dundee*.
- Hincks, R. (2002). "Speech recognition for language teaching and evaluating: A study of existing commercial products". In *Proceedings of ICSLP'02*, 733-736, 2002.
- Hirst, D.J., Espesser, R. (1993). Automatic Modelling of Fundamental Frequency using a quadratic spline function, *Travaux de l'Institut de Phonétique d'Aix*, 15, p. 75-85.
- Kim, I. S. (2006). Automatic speech recognition: Reliability and pedagogical implications for teaching pronunciation. *Educational Technology and Society*, 9(1), 322-334