Speech technology for language tutoring

Helmer Strik, Ambra Neri, Catia Cucchiarini

Department of Language and Speech, Radboud University Nijmegen, The Netherlands

h.strik@let.ru.nl, a.neri@let.ru.nl, c.cucchiarini@let.ru.nl

1. Introduction

- Increased mobility
- Emphasis on plurilingualism
=> increasing demand for language training

Optimal training: 1 learner – 1 teacher, esp. for feedback on oral proficiency.
But many teachers needed - time-consuming and costly
In classroom: individual, corrective feedback is not possible

Computer Assisted Language Learning (CALL) systems with **Automatic Speech Recognition** (ASR) offer new perspectives for language tutoring.

2. CALL applications of ASR

Some researchers are skeptical. Why?

1. Language tutoring & ASR

Standard (ASR) dictation packages have often been used for language tutoring.

Results were not convincing:

> "voice recognition technology is still at an early stage of development in terms of accuracy and single-speaker dependency"

> ASR "cannot be considered to be of benefit to ESL speakers"

However, standard dictation packages are mainly intended
- for native speakers (not non-natives!)
- for recognizing words (not language tutoring!)
i.e. not suitable for language tutoring/CALL

CALL requires dedicated ASR

2. Software design (feedback)

Design: Technology push or ...?

Waveforms & spectrograms
They look flashy and impressive,
but are not easily interpretable for students.
Furthermore: should one imitate the model?
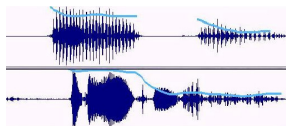
Detailed feedback on various aspects
could be useful, but not if often erroneous

- False accepts: errors that are not detected by the system
- False rejects: correct realizations that are flagged as errors by the system
False rejects can be confusing and discouraging to language learners
=> try to minimize the number of false rejects

Shortcomings in the design of the ASR-based CALL programs contribute to creating the impression that speech technology is to blame

3. Technological challenges

- Try to improve the technology (gradual, slow)
- Try to make optimal use of current technology, taking into account what is possible and what isn't possible with current technology

1. ASR/Speech technology

Non-native speech is problematic for ASR
Possible solutions – try to make use of non-native speech
- lexica with non-native pronunciation networks
- language models based on words and word orders as spoken by non-natives
- acoustic models that represent the way non-natives pronounce sounds

For the acoustic models there are several possibilities:
- simply train them on L2 speech
- use acoustic models of L1 and L2 in parallel
- use a combination of L1 and L2 models
- include intermediate phones

Improve the ASR -or- make the task less difficult:
- constrained lexicon and language model => better performance
- limited number of possible answers => utterance verification techniques

Challenge:
 develop engaging items, for which the possible answers can be predicted

2. Assessment

Discrepancies between the incoming speech and the native speech model
- recognition - be tolerant of these discrepancies
- assessment – look exactly for those discrepancies

Assessment:
- pronunciation grading - global score, at level of utterance(s)
 using more global measures, e.g. temporal measures
- error detection - score at a local (e.g. phoneme) level
 possible approaches:
  1. focus on frequent errors, 'hard-wire' them
  2. ASR-based metrics, e.g. confidence measures
  3. acoustic phonetic classifiers, better performance but …
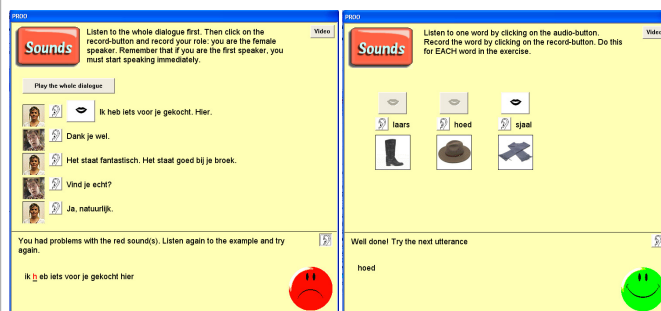
3. Projects

Dutch-CAPT: Computer Assisted Pronunciation Training for learning Dutch pronunciation (see Figures above)
- Design: based on a thorough study of existing CALL systems and analyses of 3 databases of 116 learners with different mother tongues.
- Focus: 11 problematic Dutch phonemes
- Results on pedagogical effectiveness: language learners who used this system only four times for about 30 to 60 minutes improved more than a control group that did not use the system

DISCO: 'Development and Integration of Speech technology into COurseware for language learning'
3 year project, started on 1-02-2008
Develop and test a prototype system for training various aspects of oral proficiency: pronunciation, morphology & syntax.

4 Conclusions

Speech technology holds great potential for language tutoring.

- Developing good applications requires mixed expertise: knowledge of speech technology, education/pedagogy, language acquisition/learning, software design and development.
- Developing good products therefore requires that the right people work together: speech technologists, teaching professionals, software designers and industrial partners (e.g. publishers).
- Standard dictation packages are not suitable for CALL, CALL requires dedicated ASR

ASR-based CALL systems can offer
- extra learning time and material
- specific feedback on individual errors
- the possibility to simulate realistic settings

ASR-based CALL could be employed to
- interaction in a private and stress-free environment
- develop new methods for teaching language aspects
 e.g. literacy, reading, oral proficiency, speaking fluency, and vocabulary