
Feedback management in the pronunciation training system ARTUR

Olov Engwall

Centre for Speech Technology
KTH
SE-100 44 Stockholm, Sweden
olov@speech.kth.se

Anne-Marie Öster

Centre for Speech Technology
KTH
SE-100 44 Stockholm, Sweden
annemarie@speech.kth.se

Olle Bälter

Interaction and Presentation
Laboratory
KTH
SE-100 44 Stockholm, Sweden
balter@kth.se

Hedvig Kjellström

Interaction and Presentation
Laboratory
KTH
SE-100 44 Stockholm, Sweden
hedvig@kth.se

Abstract

This extended abstract discusses the feedback given to the user of a computer-assisted pronunciation training system, depending on what level of feedback management that has been implemented.

Keywords

Computer-assisted pronunciation training, virtual tutor, feedback, Wizard of Oz, usability.

ACM Classification Keywords

H5.2. Information interfaces and presentation (e.g., HCI): User interfaces.

Introduction

Learning a language may be very rewarding, but also frustrating, if a hearing-impairment makes it difficult to discern important distinctions, or when it is a foreign language with speech sounds (i.e. phonemes) that are unfamiliar from the mother tongue. In order to master the new phonemes, the learner is required to first become perceptually aware of the distinction between the target phoneme and familiar sounds. The learner must then understand how to realize the distinction in the own production. The final step is to achieve automaticity, i.e. to be able to produce the sound without conscious planning.

The task of speech therapists and language teachers is to support this process by detecting pronunciation errors, diagnose the cause, give feedback on how to improve the pronunciation and to stimulate the student to reach automaticity by repeated training.

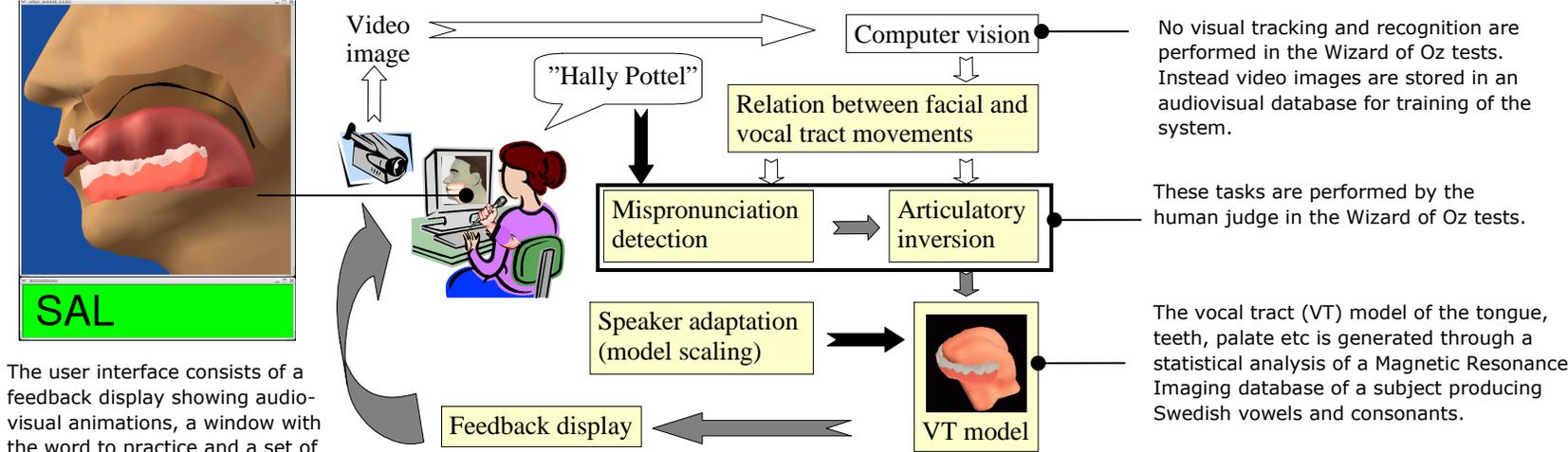
Human teachers are very apt at detecting errors and can often give pedagogical explanations on how to improve the pronunciation. Class-room teaching does however not permit the large amounts of repeated training needed to achieve automaticity.

Computer-assisted pronunciation training (CAPT) has the benefit that the student may get unlimited amounts of practice, at any time. All the existing commercial or research systems are however still vastly inferior to human teachers. One reason is that their detection and diagnosis of pronunciation errors is not good – and especially not robust – enough. The major problem is

however that the pedagogy for giving feedback to the student has fallen behind [4]. Instead of basing the feedback on the pedagogical needs of the student, it has been defined by what is technologically easy to present, such as a pronunciation score or a graphical representation of the acoustic difference between the user’s pronunciation and a correct model. Such feedback is profoundly non-intuitive and difficult to interpret for untrained users. As a consequence, the major breakthrough for CAPT has yet to come.

ARTUR - the ARTiculation TutoR

In order to address the issue of non-intuitive feedback in existing CAPT systems, we are currently developing ARTUR, the ARTiculation TutoR [2], a virtual tutor who uses three-dimensional animations of the face and internal parts of the mouth to give his students feedback in pronunciation training. The structure and components of this virtual tutor is outlined in figure 1.



The user interface consists of a feedback display showing audio-visual animations, a window with the word to practice and a set of interaction buttons.

Figure 1. Overview of the ARTUR system in automatic and Wizard of Oz set-ups.

An overview of ARTUR is also given in the introduction video [1] recorded for CHI 2006.

To test the usability of the system and involve end users at an early stage of the development process, we are conducting Wizard of Oz studies [2], in which a human, phonetically trained judge replaced the automatic detection of mispronunciations and the diagnosis of the cause. The human wizard chooses the feedback given to the student from a set of pre-generated audiovisual instructions on how to improve the articulation. We are now investigating feedback management in the virtual tutor, as one conclusion in [2] was that the set of pre-generated feedback was not optimal for all the mispronunciations that occurred during the training.

Feedback management in CAPT

In the most basic form of pronunciation feedback in CAPT, the user will only get information on if the pronunciation was correct enough or not or which part of the utterance that was most incorrect, without any cues about the problem. In ARTUR the feedback has been increased to include both detailed instructions on how a pronunciation should be corrected and general encouragement. In a previous user study [2], the wizard found that the detailed instructions were inadequate when

- The student repeated the same error several times. It would then be pedagogically unsound to repeat exactly the same feedback.
- The error fell between the defined categories – the pronunciation was not correct, but it was better than in the predefined prototypic mispronunciations.

- The wizard was unable to clearly diagnose which articulation mistake had caused the error.
- The student started to lose motivation, because the virtual tutor's feedback was too long and detailed.

To solve these issues, we have begun investigating feedback strategies used by human language teachers (e.g. [3]), when they are faced with repeated pronunciation errors or cannot pinpoint what the error was, in order to evaluate which of the strategies that could be automated in a CAPT system.

In parallel, we are implementing a multi-level feedback strategy in ARTUR, to be able to give better feedback for more varied pronunciation errors. At CHI 2006 Interactivity we will illustrate the different levels of feedback given by ARTUR in a training task focused on two of the Swedish fricatives: "s" and "sj" (for the latter the constriction is made with the tongue body at the velum, which is uncommon in other languages). Figure 2 shows the feedback loop of a training word. If the wizard deems that corrections are needed, the amount and detail of feedback is adapted to the user's previous performance, progress and mood, in order to maximize the efficiency of the feedback instructions and avoid demoralizing the student. Examples of such feedback are given in Table 1.

Successful management of feedback is even more important in a fully automatic system, as adequate feedback solutions are needed in cases where the mispronunciation detection or the articulatory inversion fails, which will happen more often with current state-of-the-art speech technology components than with a human judge.

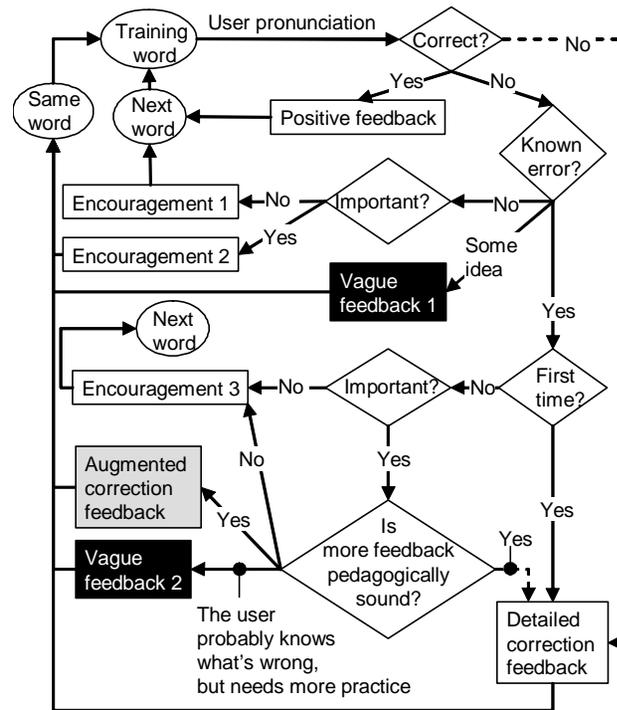


Figure 2. Flowchart over the feedback management in the ARTUR system. The grey and black boxes show feedback at higher levels. The dashed arrows indicate feedback solutions that can be avoided with the multi-level feedback system.

Acknowledgements

The ARTUR project is funded by the Swedish Research Council and the Centre for Speech Technology is supported by VINNOVA (The Swedish Agency for Innovation Systems), KTH and participating Swedish companies and organizations. The ARTUR information video was sponsored by the Christian Benoît award received by the first author.

Table 1. Examples of feedback responses given in the different categories, for the training word "sjal" (scarf).

Type of feedback	Example
Positive	"Yes, that was really good!"
Detailed, first time	"That sounded more like 'shal'; try to retract the tongue to get the narrow passage further back."
Augmented, second time	"The constriction is still too forward. Remember to let the back of the tongue touch the palate."
Vague 1	"Not quite. Think about where you place the tongue tip."
Vague 2	"Yes, almost. Say it once more: 'sjal'."
Encourage 1	"Not bad at all. Let's try the next word."
Encourage 2	"Good try! Could you say it again?"
Encourage 3	"It sounds much better now!"

References

- [1] ARTUR information video. Available at <http://www.speech.kth.se/multimodal/ARTUR>
- [2] Bälter, O., Engwall, O., Kjellström, H., Öster, A-M., Wizard-of-Oz Test of ARTUR - a Computer-Based Speech Training System with Articulation Correction. *Proc ASSETS 2005*
- [3] Morley, J., The Pronunciation Component in Teaching English to Speakers of Other Languages. *TESOL QUARTERLY*, 25 (1991), 481-520.
- [4] Neri, A., Cucchiarini, C., Strik, H. and Boves, L., The pedagogy-technology interface in Computer Assisted Pronunciation Training. *Computer Assisted Language Learning* 15 (2002), 441-467.