Feedback Management in the Pronunciation Training System ARTUR

Olov Engwall
Centre for Speech Technology
KTH
SE-100 44 Stockholm, Sweden
engwall@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 development of a computer-assisted pronunciation training system that gives articulatory feedback, and in particular the management of feedback given to the user.

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 is very rewarding, but may also be frustrating, if a hearing-impairment or the language background makes it difficult to discern important distinctions between different speech sounds (i.e. phonemes). 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, then to understand how to realize the distinction in the own production, and finally 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, diagnosing the cause, giving feedback on how to improve the pronunciation and stimulating 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 [6]. 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 ARticulaton 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 teacher who uses three-dimensional animations of the face and internal parts of the mouth to give his students feedback in pronunciation training. An overview of the goal of ARTUR is given in the introduction video [1] and the structure is outlined in figure 1. All components, except for the previously developed vocal tract model [3], represent work in progress. We have implemented algorithms for tracking and classifying the user's lip shapes in video images and audiovisual-to-
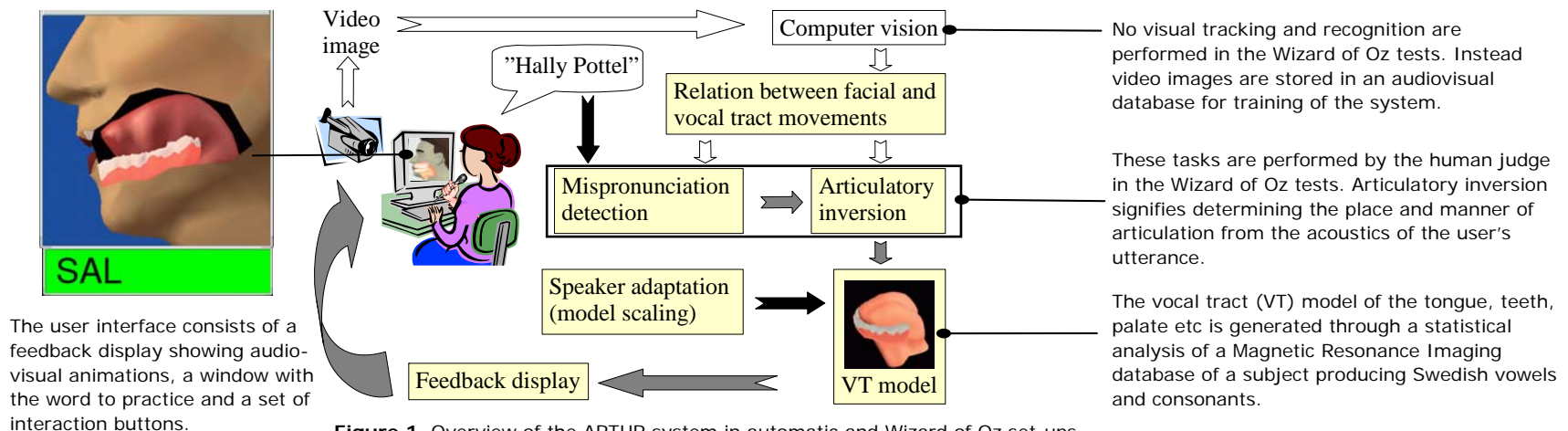

The user interface consists of a feedback display showing audio-visual animations, a window with the word to practice and a set of interaction buttons.

No visual tracking and recognition are performed in the Wizard of Oz tests. Instead video images are stored in an audiovisual database for training of the system.

These tasks are performed by the human judge in the Wizard of Oz tests. Articulatory inversion signifies determining the place and manner of articulation from the acoustics of the user's utterance.

The vocal tract (VT) model of the tongue, teeth, palate etc is generated through a statistical analysis of a Magnetic Resonance Imaging database of a subject producing Swedish vowels and consonants.

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

articulatory inversion (i.e. determining what the user did with the tongue, jaw, lips etc based on the acoustics and images of the speaker's face) [4] in offline experiments with single subjects. To make the ARTUR system automatic, these components have to be able to perform in real time on any user. The mispronunciation detection is also under development, using audiovisual data to train the detector to handle the variability of different users' speech.

To test the usability of the system and involve end users at an early stage of the development process, before all the above components are functional, we are conducting Wizard of Oz studies [2], in which a human, phonetically trained judge replace 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. The users in [2], language impaired children, were in general very positive about practising with the system, as they found it fun and that new insights were gained. As outlined in the next section, the set of pre-generated feedback was however not optimal for all the mispronunciations that occurred during the training. We are therefore now investigating feedback management.

Feedback management

In most CAPT systems, 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 information about the problem. In ARTUR the feedback instead includes detailed instructions on how to correct the pronunciation and general encouragement. In the previous user study [2], we found that these instructions were inadequate

when: 1) The student repeated the same error several times. It would then be pedagogically unsound to repeat exactly the same feedback. 2) The error was of the same type as in the previous repetition, but the pronunciation had been improved. 3) The error fell between the defined categories – the pronunciation was not correct, but it was better than in the predefined prototypic mispronunciations. 4) The wizard was unable to clearly diagnose which articulation mistake had caused the error. 5) The student started to loose 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. [5]), 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 a Wizard of Oz version of ARTUR in a training task focused on the Swedish fricative "sj" [ɧ] (made with a constriction between the tongue body and 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 in order to maximize the efficiency of the instructions and avoid demoralizing the student. Successful management of feedback will be even more important in a fully automatic system, as adequate fallback solutions are needed in cases when
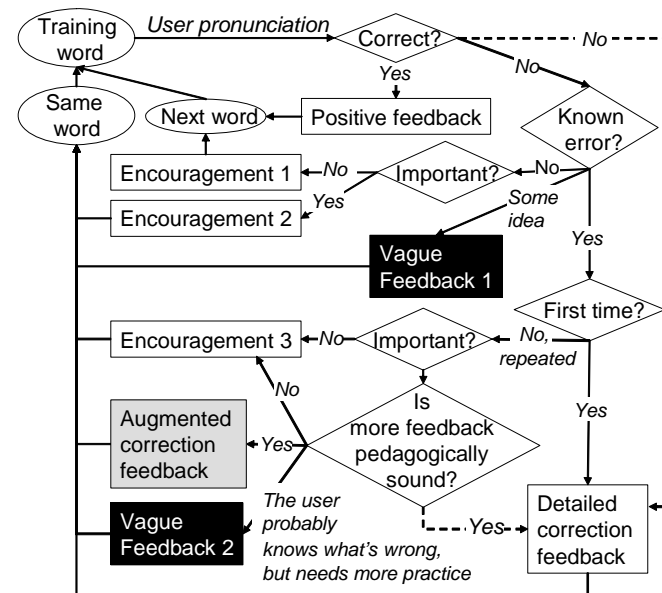
Type	Example (for *"sjal"/scarf*).
Positive	"Yes, that was really good!"
Detailed, first time	"It sounded more like 'shal'; try to retract the tongue to get the narrow passage further back."
Augmented, second time	"The constriction is still too forward. The back of the tongue should almost touch the palate."
Vague 1	"Not quite. Think about where you place the tongue tip."
Vague 2	"Yes, almost. Try once more: 'sjal'.
Encourage 1	"Not bad! Let's take the next word."
Encourage 2	"Good try! Could you say it again?"
Encourage 3	"It sounds much better now!"

Figure 2. Flowchart over the feedback management in the ARTUR system and response examples. 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.

the mispronunciation detector or the articulatory inversion fails.

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*, 36-43.

[3]   Engwall, O., Combining MRI, EMA and EPG measurements in a three-dimensional tongue model, *Speech Communication*, 41 (2003), 303-329.

[4]   Engwall, O., Introducing visual cues in acoustic-to-articulatory inversion. *Proc Interspeech 2005*.

[5]   Morley, J., The Pronunciation Component in Teaching English to Speakers of Other Languages. *TESOL QUARTERLY*, 25 (1991), 481-520.

[6]   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.