Feedback strategies of human and virtual tutors in pronunciation training

Engwall, O.

KTH Computer Science and Communication

Feedback strategies of human and virtual tutors in pronunciation training

Olov Engwall

Centre for Speech Technology, KTH, Stockholm, Sweden
`olov@speech.kth.se`

Abstract

This paper presents a survey of language teachers' and their students' attitudes and practice concerning the use of corrective feedback in pronunciation training. The aim of the study is to identify feedback strategies that can be used successfully in a computer assisted pronunciation training system with a virtual tutor giving articulatory instructions and feedback. The study was carried out using focus group meetings, individual semi-structured interviews and classroom observations. Implications for computer assisted pronunciation training are presented and some have been tested with 37 users in a short practice session with a virtual teacher.

Introduction

Computer assisted pronunciation training (CAPT) may be a very important tool in second language learning, as it gives the students access to private training sessions that are both unlimited in time and "confidence safe", i.e., learners may practice on their own without having to worry about the judgment of a teacher or their peers.

The success of CAPT technology is nevertheless still limited. One reason is that necessary system components are not advanced or robust enough, e.g., that the automatic speech recognition (ASR) is error-prone when it comes to classifying or judging deviant pronunciations and that this leads to confusing feedback (Menzel et al., 2000). However, Neri et al. (2002b) argue that successful CAPT systems are already possible, despite the technological shortcomings, and that the main flaw instead lies in the pedagogy, or rather the lack of pedagogy, of existing CAPT software. They conclude that if only the learners' needs, rather than technological possibilities, are put into focus during system development, pedagogically sound CAPT systems could be created with available technology.

One attempt to answer this pedagogical need is to create virtual tutors (Massaro, 2003; Massaro et al., 2003; Engwall et al., 2004), computer programs where talking head models should act and interact as human language teachers. One major benefit of using a virtual teacher is that the social interaction using both verbal and non-verbal gestures promotes the development of the language competence, as argued by Mead (1934).

In such a virtual tutor system it becomes important not only to improve the pedagogy of the given feedback, but to do it in such a way that it resembles *human* feedback, in order to take benefit of the social process of learning (Mead, 1934).

The aim of this study is hence to investigate feedback strategies of human language teachers and assess which of them that could be used in a virtual tutor. A literature survey, interviews with language teachers and students and classroom observations were used to explore pronunciation training in the second language classroom. Current state-of-the-art of CAPT systems are first described in order to give a background to features, problems and future development of computerized pronunciation training.

Feedback in CAPT

The combination of speech technology and computers is potentially a great asset. Ayres (2002) used a questionnaire to probe language learners' opinions about computer assisted language learning (CALL) and found that 80% of the respondents answered that CALL is relevant to their needs, 70% that the computer tasks provide useful information and 60% that CALL should be used more. A large majority of students hence found that CALL addresses the right issues, a slightly lower proportion that they received useful instructions, but only a little more than half of the students actually wanted to use the systems. It could thus be suspected that it is not the content of the training that is not to par with expectations, but *how* the information is presented – existing CALL

systems are not stimulating enough.

An important reason for this is that that the pedagogy of feedback is not yet an integral part of software development (Neri et al., 2002a). This article focuses on the pedagogy of giving feedback on pronunciation and articulatory features, such as place and manner of articulation in particular. The long-term goal of developing and using computers in second language learning should certainly be to practice communication, rather than the pronunciation of isolated words or phrases, but current state-of-the-art technology is mostly only applicable to pronunciation training, if feedback is to be provided. Even so, as Hincks (2002) points out, "in the year 2000, many commercial software programs that claimed to train pronunciation used the computer as no more than a recording device", where the student's utterance is played back, without any feedback at all.

Correctly designed, computer programs may provide large amounts of repeated training that are impossible in class-room teaching due to time limits. They may further be significantly more interactive than the traditional, listen-to, repeat-after-me practice of audio cassettes, as a program equipped with automatic speech recognition (ASR) may respond to the student's utterances and give advise on the pronunciation.

As early as in the 1970s speech analysis was used to give feedback on intonation, in that the student's pitch was analyzed and presented visually together with a teacher model. Among others, Bot & Mailfert (1982) and Öster (1999) have shown that such feedback is beneficial for the learners. Feedback showing a pitch curve is quite easily interpreted, since a rise in intonation, e.g., when ending a question, is shown as a rise in the pitch curve. As this is the only variable displayed, the user will quite easily see how a change of intonation is reflected in a change in the display.

For articulatory features, which are of main interest here, displaying interpretable feedback is less straight-forward. An example, IBM SpeechViewer, described in Öster (1999), displays the waveform of the utterance. As the waveform will differ depending on the speech volume, it is more difficult to see that voicing changes are directly represented in the display and the waveform itself does not indicate to the student how to improve the pronunciation. SpeechViewer has consequently mostly been used as an illustrative tool when the teacher and student are working together, whereas the large demand for CAPT is for unassisted student practice.

Such systems, described in e.g., Tutsui et al. (1998) and in the CALICO Review (CALICO, 2004–05), use speech recognition to judge the learner's pronunciation based on a statistical comparison with that of native speakers'. The output from the ASR, a score of how close the learner is to the target phoneme or word, does not give any information about what the learner should do to improve the pronunciation. Moreover, the score does not always represent an absolute level comparable between learners, as some speakers are more difficult to recognize than others and language background, speaking style, background noise and recording equipment influence the recognition. Even more importantly, improvements in the pronunciation as judged by human listeners are not guaranteed to be reflected in a higher score from the ASR.

The output from the ASR may be represented graphically instead. In the "Pron Gun" by Machovikov et al. (2002) the pronunciation score is plotted on a target with circles representing one, two and three standard deviations for a native speaker model. A second display shows the pronunciation level over the word as a line in the range bad to good. The European Orto-Logo-Paedia project (Öster et al., 2003) used visual maps, on which different target phonemes may be placed and ASR is used to show how close to the different phonemes the user is. In Box-of-tricks (Vicsi et al., 2000), which is primarily intended for children with speech disorders, spectrogram features and targets are represented in cartoon pictures. The student's task may, e.g., be to say words containing [s] and cover the clouds, but not the sun (different parts of the picture correspond to different frequency intervals). These visualization methods may be a pedagogically more stimulating interface than a numerical score, but still does not provide any clues on *how* to modify an incorrect pronunciation and the user has to resort to a trial-and-error scheme.

In order to improve both the recognition and the feedback, some research groups, e.g., Deroo et al. (2000) and Asakawa et al. (2005), have created databases that are specific for the pair of a target language (L2) and the users' mother tongue (L1). Not only native L2 speaker's correct pronunciations are stored, but also the non-native L1 speakers' incorrect pronunciations of the L2. In this way, knowledge about typical errors that the L1 speakers are prone to make are built into the system, and pre-generated feedback messages can be delivered for these expected errors.

Using such statistical knowledge and phonetic a priori hypothesis about the errors that L1 speakers of the L2 make, it becomes possible to provide the user with much more relevant feedback that pinpoints the error and may suggest how to modify the pronunciation.

Even when the CAPT system is able to detect, identify and provide feedback on a set of errors that the learner is prone to make, the questions of which, when and how still remain. Neri et al. (2002c) suggest that the errors that the system should give feedback on should depend on the frequency of the error, the persistence of the error, the perceptual relevance and the robustness of the error detection. The three first criteria can be considered to coincide with overall feedback decision for human teachers, while the last is more specifically related to computer systems.

The pedagogy of the feedback in existing CAPT software can be questioned, as the feedback given is limited, needs to be interpreted by the user, and is not guaranteed to highlight relevant differences compared to the target as opposed to natural variation. Hincks (2005) found that beginners using a commercial software for practicing English improved, whereas those at the intermediate level did not. The student who spent the most time with the software (50 hours) actually got slightly worse, most probably due to self-consciousness. This illustrates the importance of giving feedback that is relevant, interpretable and does not choke the student with too much information.

ARTUR - the ARticulaton TUtoR

Neri et al. (2002c) suggest that "detailed study of articulatory movements could be catered for by means of 3D computer animations of the lips and oral cavity". We believe that not only should instructions be given with such computer animations, as already made by Massaro & Light (2003), but feedback as well, instructing the learners how to change their pronunciation. We are therefore currently developing ARTUR, the ARticulation TUtoR (Bälter et al., 2005), as a part of a general virtual language teacher (Engwall et al., 2004). This private tutor Artur gives detailed instructions and feedback on articulatory positions and movements with a computer animated model of the face and internal parts of the mouth, c.f. Figure 1. The model is based on a parametrically controlled wireframe mesh of the face (Beskow, 2003) and tongue (Engwall, 2003), adapted to measurements of real speech (Beskow et al., 2003; Engwall, 2002).

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, in which a human, phonetically trained judge performs the detection of mispronunciations and the diagnosis of the cause (Bälter et al., 2005). The human wizard then chooses the feedback that Artur gives 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 the set of pre-generated feedback was not optimal for all the mispronunciations that occurred during the training with a first user group. The wizard found that the detailed instructions were inadequate when the student repeated the same error several times or when the error was of the same type as in the previous repetition, but the pronunciation had been improved. In both these cases, it would be pedagogically unsound to repeat exactly the same feedback. In the first case, either more or less feedback should be given, depending on the importance of the error. In the second, the user must be told that an improvement was made, even if a correct pronunciation was not reached. Another case when the feedback needed improvement was when the user's error was not as prototypic as in the predefined mispronunciation types, i.e., that the error fell between the defined categories or the wizard was unable to clearly diagnose which articulation mistake had caused the mispronunciation. Finally, the students sometimes started to loose motivation, because the virtual tutor's feedback was too long and detailed.

We are therefore now implementing and testing a multi-level feedback strategy in ARTUR (Engwall et al., 2006), to be able to give better feedback for more varied pronunciation errors. The amount and detail of feedback need to be adapted to the user's previous performance, progress and motivation, in order to maximize the efficiency of the feedback instructions and avoid demoralizing the student. The present study of human language teachers and their students is one path towards defining a feedback management that is more efficient and closer to that of human teachers.
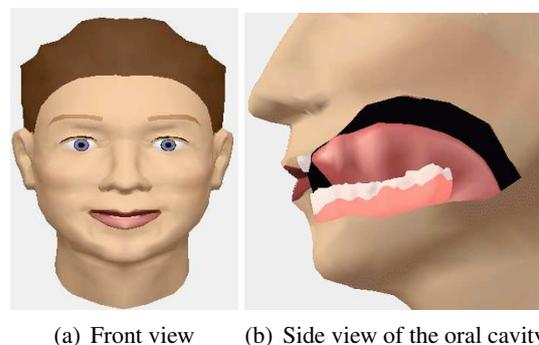

(a) Front view   (b) Side view of the oral cavity

Figure 1. Artur, the articulation tutor, gives instructions and feedback with computer animations of the face, tongue and jaw.

Pronunciation training

In order to master the phonemes (speech sounds) of a second language, the learner is required to first become perceptually aware of the distinction between the target phoneme and familiar sounds from the mother tongue, then 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 acquisition of a pronunciation is hence an illustration of Piaget's constructivist view of learning, in particular the alternative outcomes described by Bourgeois (2002). Awareness of a new pronunciation may be followed by assimilation when the pronunciation fits in the learner's phonetic map (cognitive balance) or by homeostas (i.e., a severe foreign accent) when it does not. The cognitive imbalance may also result in accommodation, in which the new phonetic sound is learned after repeated practice.

Simply put, the task of 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. The teachers' role is in particular of great importance when accommodation is to be achieved, but the role of as a supervisor of correct pronunciation is not uncontroversial.

In the decades between 1940 and 1960, pronunciation exercises were considered as an important part of teaching a language, but in the beginning of the 1960s the use of pronunciation training in the classroom was questioned. A main reason was that pronunciation training traditionally exhorted students to strive for "perfect pronunciation", which is an unattainable goal for most learners. Among others, Scovel (1969) argued that adult L2 learners may not achieve a native-like pronunciation, due to phonological fossilization and L1 transfer. Pronunciation training was eliminated or reduced from the curriculum, as it was increasingly considered that the goal of reaching a near-native pronunciation was an unnecessary goal, that it was ineffective in classroom training and perhaps that it was even impossible to teach and learn correct pronunciation. *Learn* here designates the active, instructed attainment of pronunciation, as opposed to gradual, unsupervised *acquisition*.

This negative view on pronunciation training prevailed throughout the 1970s and early 1980s, but studies in the late 1970s started to claim that the problem was perhaps not pronunciation training as such, but the way feedback was given. El Tatawy (2002) summarizes several studies from this period that argued that the inefficiency of corrective feedback was due to the ambiguous and unsystematic approaches of the teachers, i.e., that they sometimes accepted faulty utterances in order not to disrupt communication and at other occasions corrected the same errors.

From the mid-1980s and onwards, there has been a renewed interest in research on how pronunciation is taught and learned and how to give feedback (c.f. Morley, 1991 for a survey).

The statement by Morley (1991) that "it has become increasingly clear in recent years that ignoring students' pronunciation needs is an abrogation of professional responsibility" is currently widely accepted by language teachers, but the question on how to support these needs is still unresolved, even for the classroom setting.

Goals of pronunciation training

Morley (1991) argued that the learners' goal should be to achieve a "reasonably intelligible pronunciation" rather than becoming "perfect pronouncers". Instead the focus should be on *functional intelligibility*, on *functional communicability*, *increased self-confidence* and on *speech monitoring abilities and speech modification strategies for use beyond the classroom*.

These guidelines have been influential for pronunciation training in the second language classroom and are equally important for a computerized language tutor.

The first point states that the pronunciation does not have to perfect, which is of practical importance for an automated pronunciation training system, as this puts less constraints on the ability to exactly classify the pronunciation as native or not. As automatic speech recognition can never match human detection, this makes the system's task more feasible.

The second point argues that the training should be centered on communication, rather than repeat and practice drills. To some extent one could claim that one could allow the virtual tutor to take over pronunciation drills as a support for the communicative training provided by human teachers, but in the longer run, the repertoire of the virtual tutor needs to be expanded to communication exercises, rather than pronunciation training of single sounds, words and phrases.

The third point addresses an issue that has long been neglected in CAPT, and incidentally in some language classrooms, in that it is at least as important to encourage the student to speak and communicate as it is that what is said is correct.

The final point, that the feedback must support the student in his or her efforts to become aware of the pronunciation and repair it when

there is need, could be used as a severe criticism against most existing CAPT systems, where no cues are given regarding how to modify the pronunciation or which the most important features of a phoneme are.

How to practice pronunciation

The way feedback is used will largely depend on how pronunciation is trained. Before addressing the issue of feedback in the next section, it is hence important to summarize the different methods used for practicing pronunciation.

Murphy (1991) argues that improvement in pronunciation foremost depends on the commitment in time and energy from the students themselves. Elliot (1995) even found that the subjects' attitude (see further the Section on Student interviews on the Pronunciation Attitude Inventory) regarding the goal of attaining a near-native pronunciation was the most significant factor for achievement, i.e., students who were more concerned about their L2 pronunciation tended to have a better mastery of L2 allophones.

Echoing this opinion, Morley (1991) stresses the importance of involving the learner, by promoting self-responsibility with clear guidelines, carefully designed tasks and sharply focused cues for self-monitoring. The role of the teacher is to be a "pronunciation coach" for the student (the "speech performer"), shifting from dependent (teacher led and assessed) to independent (self-monitoring) practice. With this method, the teacher should still provide diagnostic analysis, choose and prioritize features with maximal impact, provide pronunciation models, develop a set of instruction formats and exercises and give cues and suggestions for modifications, on the one hand. On the other, the teacher-coach should promote the independence of the students by encouraging them towards self-awareness and self-monitoring. The teacher should give suggestions on how the students can observe one or two features in their own speech at the time, develop self-rehearsal (listening to oneself), help them set short- and long-term goals, and encourage and support the learners in their efforts.

This credo is well in line with the support that pedagogically well-designed CAPT could provide. Some feedback promoted by Morley (1991) and Murphy (1991) can quite easily be used in CAPT, such as the recommendation to supply the students with cues on the relationship between standard orthography and phonological references and to use written text, as well as oral utterances to elicit learner production. Speech-oriented listening tasks that may help the learners to develop their auditory perception are already part of some systems.

Both Morley (1991) and Murphy (1991) stress that pronunciation training must focus both on macro- and microlevel perspectives, i.e., that the pronunciation training can not focus solely on the importance of vowel and consonant sounds, but must also focus on suprasegmentals (stress, intonation etc.). As it is easier both to automatically analyze and give interpretable feedback on suprasegmentals (c.f. the Section on Computer Assisted Pronunciation Training), the tendency is that suprasegmentals get more attention in CAPT. It is thus rather the microlevel perspective that needs to be addressed further.

It is clear that current CAPT does not meet all recommendations. Morley (1991) argues that pronunciation training should be an integral part of communication, and that imitation practice should be used only as necessary and as a short-term component. Instead, rehearsed speaking practice, e.g., oral reading of a script, can be used to work toward stabilization, and extemporaneous speech practice, such as small-group discussions, should be used to integrate the modified speech patterns into naturally occurring speech. The classroom activities for oral production that Murphy (1991) outline are closely related to the ones proposed by Morley (1991) above. Such exercises are seldom used and are more difficult, but not impossible, to incorporate into CAPT.

Another recommendation, that exercises should be adaptable to each student, as every student will have different background and motivation, is almost never met in current CAPT systems. This article will mainly focus on the adaptation of the *feedback*, but exercises should certainly also be tailored for different users. The most realistic method to adapt the training to the user is to provide a framework and a database of exercises and let language teachers or the learners themselves adapt the training to become suitable. Some semi-automatic adaptation should be performed based on initial information on the user's age and mother tongue, and the choice of exercises may be based on the learner's performance on previous exercises, but users should be given the power to design their own training.

One possibility is to present exercises and information that are suitable for many different learning styles, as the multimodal methodology used by Elliot (1995) when teaching Spanish to English speaking students. The methodology focused on articulatory instructions. When a Spanish allophone was studied, the students were prompted to describe point, place and manner of articulation in their own words and compare it to the English counterpart. The teacher then showed

a transparency with a facial diagram for the students to compare with their own description.

Working in small groups of two or three, the students listened to each other and provided feedback and correction if needed. The benefit of the exercise would be that the students' own production could be improved by practicing the perception of others' errors.

Two concerns may be raised. Elliot (1995, p. 539) himself admits that "it is possible that frequent exposure to incorrect language usage [...] may reinforce errors in pronunciation". Secondly, the control of the feedback given is lost, and students may be uneasy about giving or receiving correction to/from their peers.

In CAPT, Elliot's method may well be used to give multimedia instructions on the articulation, to let the user describe articulations by typing or mouse clicking and to train the user's perception by presenting native as well as non-native productions of the target and let the user provide feedback on these.

Feedback in pronunciation training

The current concord is that pronunciation training has a role to fill in language teaching, but views are parted on how and even if feedback should be given. This section will attempt to define what feedback is, review different influential opinions on giving feedback and summarize proposed feedback methods.

What is feedback?

Lightbown & Spada (1999, p. 171) state that corrective feedback is "any indication to the learners that their use of the target language is incorrect". Lyster & Ranta (1997) analyzed feedback strategies used by four language teachers and classified them as explicit correction, recasts, repetition, clarification requests, metalinguistic feedback and elicitation.

Explicit correction signifies that the teacher gives the correct form and clearly indicates that what the student said was incorrect.

In *recasts*, the teacher reformulates the student's utterance, removing the error.

Repetition differs from recast in that the teacher repeats the student utterance *with* the error using the intonation to indicate that the utterance contained an error. Repetitions may also be used as positive feedback on a correct utterance.

Clarification requests urge the student to repeat and/or reformulate the utterance, because the meaning was unclear.

Metalinguistic feedback concerns comments or questions about an error used to make the student reflect upon and find the error themselves with the information given by the teacher.

Elicitation encourages students to provide the correct pronunciation, word or form, by open-ended questions or fill-in-the-gap utterances.

To give feedback or not?

Traditionally every error would be corrected immediately, in methods based on behavioural theories of learning, such as audio-lingualism. This strategy followed the general pedagogical view that human beings were purely reactive creatures, whose actions were determined by internal or external forces to reach success in the interaction with the environment, or to gain a reward or avoid a punishment.

When this notion of human motivation had changed in general, so did pronunciation training in the beginning of the 1970s. The active individual with an own responsibility and motivation to control the interaction and goals became the predominant view. For pronunciation training, this led to a new methodology stating that not all errors should be corrected, and those that are should not be done so immediately, e.g Krashen (1981).

One reason for not giving immediate corrective feedback is that not all deviations are errors, in the meaning that the student is unaware of the correct pronunciation. They may instead be mistakes or *slips*, which occur naturally in the mother tongue as well, without any need for corrective feedback. The difference is that the speaker may self-correct a slip, but not an error.

A second reason is that, according to nativist theories, negative evidence is of little importance in acquiring a language (e.g., Chomsky, 1975). This view is disputed by e.g., Schachter (1991) who argue that it is the corrective feedback that make the student abandon incorrect hypotheses and formulate new ones.

The third reason is the affective factor, that students may loose self-confidence if being corrected all the time. Murphy (1991) stressed that teachers must be tactful when deciding on how and when to give feedback about student errors.

Stevick (1978, p. 146) argues strongly against giving too much explicit feedback, because "all to often, self-consciousness leads to tension, tension leads to poor performance, poor performance leads to frustration, frustration leads to added tension, and so on around a downward spiral". He further insists that a student in such a downward spiral will demand more and more physical details about how to produce the sound. Stevick thus advises teachers to give non-evaluative feedback (i.e., refrain from explicit feedback) and instead produce the student utterance in the correct way,

regardless of if the student's pronunciation was accurate (i.e., either a recast of an erroneous utterance or a reinforcing repetition of a correct one).

As pointed out by Ancker (2000, p.20) "many students still expect, even request, the teacher to correct all their errors". In the survey by Ancker (2000), 802 teachers, 126 teacher trainees and 143 students were asked if teachers should correct every error the students make when speaking English. The vast majority of the teachers (75%) and the majority of teacher trainees (64%) agreed that not every error should be corrected, while the students, to the same degree (76%), thought it should!

Ancker did however neither investigate the proportion of the errors that the teachers did indeed correct, nor the students' estimation of the proportion of the errors that was corrected. It may well be the case that the students thought that the teachers corrected many more errors than they did, and would actually not wish to have more corrective feedback than already given. Ancker's results should hence not be taken as evidence that a CAPT system should correct every error in order to satisfy the user.

Ancker summarized the reasons for *not* correcting every error as *affective* (the students' self-confidence may suffer), *classroom management* (that it would take too much time) and *pedagogical* (that not all corrections can be processed at the same time and will only confuse the students).

The reasons for correcting every error concerned *learning* (the students must know what is correct, or else they will be confused), *fossilization* (if errors are not corrected, the students will repeat the same errors that will end up as being perceived as correct in their own view) and *professional* (that students expect the teacher to correct every error).

Finally, it is worth considering the statement of Ancker (2000, p. 24) that "not correcting an error is not the same as teaching incorrect forms", which is true for CAPT as well.

What kind of feedback should be given?

Carroll & Swain (1993) investigated different types of corrective feedback in second language learning and found that all groups receiving feedback, explicit or implicit, performed significantly better at the end of the study than the control group. Any kind of feedback was hence found to be beneficial, but the group receiving *explicit* feedback outperformed the other groups.

One explanation for this may be how the feedback is processed by the learners. Lyster & Ranta (1997) found that recasts was by far the most common type used by human language teachers

and that the recasts did not lead to any repair from the students. Their explanation was that recasts already provided the learners with the correct form and did hence not elicit active correction. In another study, Mackey et al. (2000) found that the learners perceived recasts *as another way to say the same thing, rather than a correction.*

Several studies, e.g., Lyster (1998); Lyster & Ranta (1997); Hendrickson (1978), have found that feedback techniques that require that the students reformulate the utterance themselves and those where metalinguistic clues about the error were used were more likely to improve learners results than simply recasting. The students should hence be active in the correction, relying on their own resources, by using corrective techniques that elicit them to modify and self-repair their utterances. Clarification requests, elicitation and confirmation checks are hence better when it comes to improving the students' development and awareness of the own output.

There is unfortunately no simple measure of the efficiency of different types of corrective feedback, as the efficiency is related to both uptake (i.e., that the student has understood the feedback) and successful repair (i.e., that the student actually changes the pronunciation). The first difficulty is how the internal process of the students' uptake should be measured and the second that uptake and successful repair are related, but not in direct correspondence.

Lyster & Ranta (1997, p. 49) define uptake as "a student's utterance that immediately follows the teacher's feedback" and Chaudron (1977, p. 440) argues that feedback effectiveness could be measured as a "frequency-of-count of the students' correct responses following each type of feedback". These measures seem to be quite conservative, as the student may well both have processed the feedback and be able to use it later without having to make an immediate repair. The converse is also true, that students may well do an immediate repair, but may still not retain the correct form for future utterances. Several researchers, among them Lyster & Ranta (1997), Oliver (1995) and Mackey & Philip (1998), have hence pointed out that "uptake, whether reformulation or repetition, is not necessarily an indication that feedback had a positive effect [. . .], nor is its existence the sole evidence for learning" (El Tatawy, 2002, p. 15).

The type of feedback that would be appropriate also depends on the type of exercise. Oliver (1995) found that implicit negative feedback was frequently used in interactions between native speaking and nonnative speaking children. Depending on the type of error that the nonnative

speaker made, the native speaker responded differently. For multiple errors, negotiations were used to clarify meaning, whereas recasts were used for single errors to correct form.

Another important point is learner motivation and the importance of creating a comfortable classroom atmosphere, meaning that "negative feelings about corrections [...] need to be eliminated" (Morley, 1991, p. 504). This should be achieved by promoting the recognition of self-accomplishment, so that learners are aware of small successes in modifying features and become proud of their own accomplishments. The assessment of achievement should be based on degrees of change rather than on absolutes, a finding replicated in user interviews by Eriksson et al. (2005). Critique should be constructive with emphasis on positive features as well as those in need of correction.

Summary of criteria for good feedback

El Tatawy (2002) summarizes the prerequisites for successful corrective feedback as:

1. The feedback should be provided systematically and consistently.

2. The feedback should be clear enough to be perceived as feedback.

3. Time and opportunity for self-repair and modified output should be given.

4. The feedback should be fine-tuned, i.e., the teacher's intent, the target error and the learner's perception of the feedback should match closely.

5. The feedback should focus on one error at a time and over a period of time in order to be consistent and intensive.

6. The learner's readiness to process the feedback should be accounted for.

With the above survey of feedback and language acquisition in mind, the following prerequisites should be added:

7. The feedback should be adapted to the situation, i.e., different feedback is suitable for different exercises.

8. The learner should be actively involved in pronunciation monitoring and correction.

9. The feedback should promote the learners' communicative self-confidence.

Data collection

The data in this study was collected through three different channels. The first was in-depth discussions in a teacher group about feedback in pronunciation training, the second was through classroom observation of feedback given and the third through a survey of student opinions, using a questionnaire and interviews.

One reason for including both teachers and students was the results in Ancker (2000) that teachers and students may have quite different perspectives on the adequate amount of feedback. Most previous studies have concentrated on the teachers' perception of feedback and it is hence worthwhile to probe the learners' views about the feedback that they wish to receive, as they will be the primary users of a CAPT system.

All interviews were recorded on a DAT player or directly on computer disc, while notes were taken during the classroom observations. Some notes were also taken during the interviews as a backup in case of recording problems, but the sound files were the main data source for the analysis below.

Teacher interviews

Six language teachers participated in the study, four in a focus group setting and two were interviewed individually using a semi-structured protocol (Rubin, 1994) with open-ended questions and the possibility to probe the respondent further if needed.

The teacher group was intentionally heterogeneous with respect to target language and student level, in order to capture general pedagogical strategies, rather than language or level specific solutions. As pronunciation training is mostly included at beginner level, there was a preponderance for teachers active at beginner and intermediate level, but with some participation from advanced level teachers as well. The respondents teach Swedish, English, German, Spanish, Japanese and Chinese to engineering students at the university level. Five out of the six teachers are native speakers of the language they teach.

The discussions with the teachers were centered around the framing questions put forward by Hendrickson (1978):

1. Should learners' errors be corrected?

2. When should learners' errors be corrected?

3. Which errors should be corrected?

4. How should errors be corrected?

5. Who should do the correcting?

with the focus on questions 2-4. In addition, in this study, more focus was placed on repeated errors and the balance between correcting and motivating the students was discussed in particular.

The focus group meeting

The focus group convened during a one hour meeting. The format was a round-table discussion where the investigator raised open questions for discussion and elicited responses from all participants. The discussions did involve all participants addressing their colleagues, while the interviewer monitored the discussions and the recording.

Individual interviews

The individual interviews were 20-40 minutes long and were focused on the same topics as the focus group meeting. The interviewer took a more active role than in the focus group meetings, but was careful not to venture personal beliefs in the questions given to the teachers.

Classroom observations

The classroom observations were made in beginner level courses, in order to maximize the amount of pronunciation training and feedback given during the lessons.

Class A was a four hour class with 18 beginner level students at the end of the course in a L2 that was moderately different from the L1. The teacher mainly spoke in the L1 language with inserted L2 phrases and words.

Class B was the two first hours of a four hour session with 26 beginner level students at the third session of the course in a L2 that was close to the L1. The teacher alternatively spoke in the L1 and L2 language, with the communicative utterances in the L2, but with practical, grammatical or important information in the L1 language.

Class C was the two first hours of a four hour session with 25 students at the fourth session of the course in a L2 that was very different from the L1. The teacher almost exclusively spoke in the L1, except when giving the pronunciation of words and phrases of the L2.

Student interviews

Five students took part in the interviews, three of them in a focus group meeting and two were interviewed individually. The students were first given a questionnaire, c.f. Table 1, based on the Pronunciation Attitude Inventory (PAI) by Elliot (1995), in order to probe their feelings about pronunciation training and their personal aims. The idea was to investigate whether the PAI index was

related to the student's opinions about when and how pronunciation feedback should be given.

The same questionnaire was also given to the teachers, but with statements 1, 2, 3, 5, 6 and 9 changed so that the teachers should comment upon them relating to their students, e.g., "It is important to me that the students achieve a good pronunciation in the second language I teach". In addition, statements 1, 5 and 9 were doubled, in order to let the teachers comment on their attitude towards their *own* pronunciation. The teachers were asked to answer the PAI questionnaire in order to investigate if their attitude towards their own pronunciation and that of their students affected how they distributed feedback.

The interviews with the students were focused on how much feedback they wished the teacher to provide, when, how and which errors should be corrected, how the motivation is affected by different type of feedback, and how they would like a CAPT program to function.

Table 1. Pronunciation Attitude Inventory questionnaire, adapted from Elliot (1995).

1. It is important to me to achieve a good pronunciation in the second language

2. I will never be able to speak the second language with a good accent

3. I believe that I can improve my pronunciation skills in the second language

4. I believe more emphasis should be given to proper pronunciation in class.

5. I'd like to sound as native as possible when speaking the second language

6. I try to imitate native speakers as much as possible

7. Communicating is much more important than sounding like a native speaker

8. Good pronunciation skills are not as important as learning vocabulary and grammar.

9. I'm concerned with my progress in my pronunciation of the second language

Answers were given by ticking a square corresponding to the statements *5 = Always or almost always true of me*, *4 = Usually true of me*, *3 = Somewhat true of me*, *2 = Usually not true of me* and *1 = Never or almost never true of me.*

The focus group meeting

The focus group meeting lasted one hour. These students were now studying beginner or intermediate level courses and had previous experience of foreign language training as well. The interviewer asked open-ended questions to the group, or sometimes addressed one particular student, in order to get clarifications on what the student had said or to elicit responses from all students. The students mostly addressed the interviewer rather than the group to a larger extent than the teachers had, but they did also react to and comment upon the responses given by the fellow students.

Individual interviews

The two students who were interviewed separately had a more diverse language training experience, including private-lessons, small class lessons and intensive courses. The duration of these interviews was 20-30 minutes. One of the students was very enthusiastic of the feedback given by the teacher in the private lessons, but at the same time deemed that the feedback in a small-group conversation course of another teaching center was totally counter-productive. The interview with this student was hence expanded to pinpoint the differences in feedback strategies between the two teachers' use of feedback.

Data analysis

The focus group meeting and the interviews were analyzed by the investigator listening to the recordings and simultaneously noting down and sorting relevant utterances into the different response categories *when*, *which*, *how* and *motivation*. Due to the large amount of statements, they are not presented in this section, but the important findings are instead summarized in the next section.

The pronunciation attitude index was calculated for each student and teacher, as the numerical average of the responses. For negative statements, the scale was reversed before averaging.

The score for the students and the overall score for the teachers were based on the nine questions in Table 1, where the teachers answered the corresponding questions, but relating to their students. For the teachers, a second score was calculated regarding their attitude towards their own pronunciation, based on the questions relevant to this. For comparison, the students' score on the same questions is also given.

Results

The results of the interviews and classroom observations are first analyzed separately, before the implications for CAPT are discussed. The citations in this section of the respondents' oral utterances have been transcribed as well-formed sentences and translated to English to facilitate reading.

Table 2 lists the PAI index calculated for the students and the teachers. The number of respondents is too low to make any quantitative statistical analysis of the score, but some qualitative tendencies may be found. Teachers at beginner level had higher PAI index than those teaching at higher levels, with a range in index from 4.3 to 5.0. During the interviews, the beginner level teachers were also more positive about the importance of correcting the students' pronunciation errors. This may be due to the proficiency of the students; that beginner level students make more mistakes and more serious ones, which may hence be more important to correct. There was also a strong correlation between the teachers' attitude towards their own pronunciation and that of their students.

Worth noting, even though it is impossible to draw statistical conclusions from the small populations in the study, is that the PAI score for the students is substantially lower than for the teachers. According to the PAI index, the students in this study hence feel that pronunciation as such is less important, with the lower student score being mostly due to questions 3, 7 and 8, as the student rank communication and vocabulary and grammar skills and practice as more important than the teachers do.

Table 2. Mean scores on the Pronunciation Attitude Inventory index by students and teachers. For the teachers, the overall score relates to their attitude towards their students' pronunciation, while "own" to their own. n signifies the number of questions from Table 1 that were used to calculate the index, and Questions which ones.

	Teach.	Stud.	n	Questions
overall	4.4	3.7	9	1-9
own	4.4	3.4	5	1, 5, 7-9

Teachers

Due to the different levels at which the teachers are active, pronunciation is taught and monitored differently, ranging from repeating after the teacher, over reading out loud, teacher–student dialogs, and small group discussions to oral presen-

tations for the whole class. Despite these differences, the similarities between the teachers' feedback strategies prevailed.

The teachers deemed that very little feedback on pronunciation was given in their classes, due to time limits, but also because there tend to be a progress in pronunciation throughout the course even without feedback, and they were reluctant to interfere with the students' communication. As could expected there was a difference between different student proficiency level, with more feedback given in beginner level courses.

When?

There was a large consensus among the teachers about the importance of providing feedback at the right moment, as incorrectly timed feedback is counterproductive. Feedback should not interrupt discussions, the speaker's train of thoughts or fluent stride. Even if it means that errors are left uncorrected, the teachers prefer to provide feedback when it does not affect the communication.

Several teachers stressed the importance of the students daring to speak and that the amount of feedback needs to be limited in order not to intimidate them.

Which?

The teachers ventured several criteria for which errors should be corrected, most of which converge with the goals or criteria cited above, while others added new dimensions.

The first concerned comprehensibility, i.e., if the utterance could not be correctly understood.

The second intelligibility, i.e., if the utterance could not be understood without effort.

The third frequency, i.e., if the student repeats the same (type of) error several times (note that what the teacher is doing is thus actually monitoring slips *vs.* errors that the student can not self-correct, as discussed above).

The fourth social impact, i.e., if the error is grave in the sense that the listener gets a negative impression of the speaker.

The fifth generality, i.e., if the error is one that is often made in the L2 by either foreign speakers in general or the L1 speakers in particular. This criteria hence addresses the efficiency of the feedback with respect to the entire student group, i.e., if more students benefit from the feedback, but also that such general errors may be more difficult to overcome without teacher assistance.

The sixth proficiency, i.e., that a student with a better overall pronunciation would get corrective feedback on an error for which a student with a less good pronunciation does not get one (which conforms to criteria 6. in El Tatawy,

2002). Every student will thus get approximatively the same amount of feedback, no matter their absolute level, but this criteria was not universal. Some teachers used an absolute criteria, so that better pronouncers would get significantly less negative feedback. The reason for this was both said to be to encourage the students who perform well, and that spending less time on strong students signify that more time could be spent helping weaker pronouncers to progress.

The seventh personality, i.e., that a student who seems to appreciate corrective feedback receives more than one who seems uncomfortable with it. In other words, the teacher who stated this is implicitly using and adapting the suggestions by Stevick (1978) in order to avoid self-consciousness in weaker students.

The eighth was commonality, i.e., that an error that is common among native speakers of the L2 language is regarded as less grave than such errors that a native speaker would never make. These "native-like" errors would either be left uncorrected or be pointed out with less emphasis.

The ninth exercise focus, i.e., that feedback would primarily be given on the feature that was targeted by the exercise (corresponding to the fifth criteria for good feedback of El Tatawy, 2002).

How?

One teacher described a strategy going from minimal implicit feedback towards more explicit, when it was required. In the most minimal form, the teacher indicates that an error was produced by a questioning look, if needed accompanied by an attention-catching sound. This gives the students the opportunity to identify and correct the error themselves and is hence a form of implicit elicitation (matching criteria three in El Tatawy, 2002). If the student was unable to correct the error, the next step would be a recast, uttering the mispronounced word correctly and encouraging the student to repeat it. If the student was still unable to correct the pronunciation, the recast would be repeated once or twice, which signifies that the recast turns into an explicit correction. Should the recast approach be unsuccessful, only then would an explicit explanation of the difference between the correct and erroneous pronunciation be used, and only if the error was important and relevant for the student and the whole class. The latter judgment was a consequence of avoiding too much focus on a particular student's pronunciation both for motivational reasons and due to the time limit.

Several teachers stated that explicit, phonetic, explanations were avoided, because the students did not understand how to achieve different fea-

tures from a phonetic description. Other teachers agreed that formal descriptions and sketches on place of articulation were of little use, because the students were unaccustomed to actively thinking about how they produce different sounds.

One of the teachers *did* use articulatory instructions, opening the mouth widely, pointing with his finger to where the sound should be produced and encouraging the students to actively explore different places of articulation (*"Now advance the tongue to make the sound further front."*).

Others encouraged students to use sensory feedback, by placing their hands on their neck to feel the vibration of voiced sounds or in front of the mouth to feel the puff of air in aspiration.

Instead of using formal descriptions, unfamiliar pronunciations were compared to similar sounds in Swedish or other known languages as an approximation. Swedish examples of sounds or words with minimal difference in voicing or in accent (the difference between *anden* – the spirit – and *anden* – the duck) were also used to illustrate the importance of making certain distinctions.

Metalinguistic explanations were also used to enforce the feedback, e.g., *"this is a grave error. It disturbs the listener"* or *"This changes the meaning of the word. It is as different as if you change the vowel in a Swedish word"*). Such explanations were also used to motivate why it is important to get a particular pronunciation right, e.g., because there is a risk of word confusion or the mispronunciation becomes impolite. It should be noted that the teachers teach adult, engineering students who may be more interested in knowing why it is important to achieve a particular pronunciation.

Several teachers stated that they sometimes employ general recommendations rather than feedback on particular errors, e.g., *"You should try reading aloud by yourself at home"*. What the teachers are aiming for is thus for the students to reach automaticity by repeating the articulatory movements required when reading the text out loud. When doing this, the student receives no feedback, but the perception of the own production is trained.

Motivation

Several, but not all, teachers expressed concerns about correcting students in front of the whole class and preferred to give most of the feedback during small-group exercises, separate exercise sessions with only the teacher present, after class or as a written recommendation after a student presentation. One teacher has introduced short (15 minutes) private training sessions to be able to give feedback that the students might be intimidated by if it was given in the whole-class setting.

One teacher did not feel that there was any problem correcting students in front of the class and one adapted to the students, in that students with self-confidence, who did not seem to mind getting corrected publicly got more detailed correction in whole-class setting, while those who did seem to mind did not.

Another adaptation to the whole-class situation used by several teachers is to expand the feedback to concern a general error that the whole class may have difficulties with. The reason is partly not to single out one particular speaker, but mainly that the entire class should benefit from the instructions. If the error on the other hand is an individual problem non-representative for the group, correction would be given only in small group exercises or personally.

The risk of negative self-consciousness was mentioned as well as a reason for limiting the feedback, as some students are ashamed of their own faulty pronunciation. One teacher pointed out that explicit explanations of an error should be impersonal, i.e., *"When one says. . . "* should always be used rather than *"When you say. . . "*.

One teacher particularly mentioned that exercises with non-problematic pronunciation was inserted among the more difficult ones in order to get the students to feel that they do make progress and that it is not impossible to learn the language.

Who?

A couple of the teachers spontaneously touched upon the fifth of the questions from Hendrickson (1978), i.e., if someone else than the teacher could be doing the correction. These teachers encouraged either all students or strong pronouncers to help and correct their peers in small group exercises. The advocated form of correction was recasts, that the correcting student should repeat the mispronounced word in the correct form.

Classroom observations

The practice in pronunciation the students got during the session in class A was in repeating words together after the teacher, reading prepared answers out loud, reading a paragraph of a text out loud, talking in small groups with one or two other students and a dictation exercise.

The students in class B were quite active producers of the L2 during the class, responding individually in simple communication exercises, reading dialogs in whole-class or pairwise and repeating words and sentences after the teacher.

Class C was more focused on practicing isolated words and short utterances, and student par-

ticipation was more guided. The training consisted in repeating after the teacher, practicing specific details oneself, pair-wise practice of short dialogs, either read or following an example given orally by the teacher, and short dialog interactions between one student and the teacher.

(A), (B) and (C) preceding an observation indicate the class it was made in.

When?

(A) The timing of the feedback differed somewhat between the different situations, but in general, the key feature was that the feedback was as non-intrusive as possible. When reading out loud, corrections came *after* the student had finished reading, unless the student stumbled over the word and hesitated. If so, the teacher provided the correct pronunciation – but as a help, rather than as a correction. The feedback after reading consisted of overall encouragement, combined with an explicit correction or recast, or in the case when a recast had been given during the reading, a repetition of this recast, as a reminder.

(B & C) As for class A, the feedback never interrupted the students, but more pronunciation feedback was given, in whole-class setting, in pair-wise reading and by summing general difficulties in whole class after pair-wise exercises.

Which?

(A) Overall, very little feedback was given, and when it was, it was most often because the students indicated that they were unsure about the pronunciation, by an interrogative prosody and/or a questioning look at the teacher.

(B) Feedback was given mainly on one specific target at the time, such as intonation or difficult sounds. The latter was most clearly illustrated by the use of tongue twisters, for which the focus was entirely on the problematic sound.

(B) If the mispronunciation was caused by not following a certain pattern the general pronunciation rule was repeated.

(C) The feedback was mostly concentrated on aspects that are of little or no importance in the L1, but carry information in the L2. The reason was probably both that mispronunciation of these features may cause the words to be given another meaning and that the students need to be reminded of features that they are normally not aware of.

How?

(A & C) Recasts and repetitions of correct utterances were the most common feedback given. The repetitions were used to confirm a correct answer given by the student and at the same time

provide a clearer (or louder; some student responses were quite low) target for the class. Recast were used in the same way and situation, with the difference that recasts had more stress and longer duration.

(A) The dictation exercise practiced the relation between the perception and spelling, and when the exercise was corrected, some differences in the spelling and pronunciation relation between the L1 and the L2 were discussed.

(B) Both implicit (*"Sorry?"*) and explicit (*"Could you repeat that?"*) elicitation for the student to self-correct was used frequently. Different levels of feedback or correction was then employed. The first was a repetition for a successful self-repair and a recast for an unsuccessful, accompanied by either a pause or an exaggerated intonation/stress indicating that the pronunciation had not been correct. With the finding by Mackey et al. (2000) in mind that recasts were not perceived as a correction, it is interesting to note that the students did repeat the recast, suggesting that they were aware of the correction. The repair was then often followed by the teacher repeating the word twice in order to give a clear target.

(B) For a general error or difficulty, the utterance was repeated by the entire class together. Worth noting is that the teacher then often said some version of a sentence with the content *"This is very difficult"*. The reason for doing so was probably two-fold, both affective and attention catching. By stating that the pronunciation is difficult in general, the student who made the mispronunciation becomes less personally targeted, and at the same time the other students are made aware that the feedback may be relevant to them as well.

(B) At some occasions, explicit explanations were used (*"it should be more rounded"*) when the articulation differed from the closest Swedish sound.

(C) The main feedback strategy was that the teacher provided a good acoustic target, repeating the word or utterance many times, both alone and together with the entire class. The feedback was more explicit than in class A and B, indicating exactly which feature that was incorrect and how it should be changed. Since the sounds of the target language were quite unfamiliar to the students, both similarities with Swedish sounds and multimodal feedback strategies were used to a larger extent. The students were instructed to use sensory input as feedback (putting the hand in front of the mouth to feel aspiration), and the teacher illustrated intonation with hand movements, almost like a conductor.

(C) An exaggerated pronunciation was en-

couraged, in order to make the students aware of the salient features of the sound, e.g., *"That was correct, but do it even more"*.

Motivation

(A) During the small group discussions the teacher circulated between the groups, but with the main task of facilitating and encouraging the communication by providing missing words and asking questions. Only in the cases when the communication was flowing well was any corrective feedback given, either as recasts or explicit corrections.

(B) Feedback was extensively given in front of the whole class, but corrections were always balanced by positive encouragements for improvements. The author's subjective impression was that students reacted as if the correction was indeed a help, rather than a rebuke. It could seem as a trivial observation that the students feel that the teacher intends to help, but as discussed in the Section on Feedback in pronunciation training this is not self-evident, nor is it always the case.

(B) A failed student correction after a recast was almost always dropped, either for affective reasons or time constraints. Instead the difficulty was acknowledged (*"Yes, this is a difficult sound."*) and a suggestion could be provided (*"Try to practice it by saying it out loud at home."*)

(C) Corrections and feedback were given in whole class settings, but often after letting the students practice the pronunciation separately first. Students were very insecure and self-conscious about the pronunciation, often stating that they could not pronounce the word, even when they made an attempt. In the pair-wise read dialogs, the communication flow was a little higher, and the teacher circulated, listened to each pair and made one or two remarks on the most important observation.

The students' self-confidence in the three different classes was clearly influenced by their previous familiarity and exposure to the L2 language, i.e., even if they did not speak the target language beforehand, they had a general feel for the language when the language was closer to the mother tongue or when they had heard the L2 language on television etc. The factor of how familiar the L2 language is to the students does hence not only influence the types of errors that they make, but also their readiness to attempt speaking. This should be taken into account in CAPT in order to give feedback on a level that promotes student production.

Students

When questioned on how to best improve your pronunciation several students answered that the key is to both hear much of the target language and getting many opportunities to speak, i.e., the importance of establishing the perception–production relationship.

One of them also stressed the importance of hearing different persons saying the same word, in order to be able to figure out what the important features of the sound are.

When?

The most important criteria according to the students should be to not interrupt, and give feedback *after* the student utterances. This opinion, expressed spontaneously by several students, does hence clearly oppose the behavioristic notion of the importance of immediate feedback. The reasons stated were that this allows the student's flow and train of thought to be maintained, that it disturbs the exercise less and that more general errors are addressed, which means that the feedback is less personal (*"It is not: 'You said something wrong', but more like: 'This is a general problem that you all should think about'."*).

Some of the other students stressed that corrections should come as early as possible, both in the sense that it should come early in the course and that it should follow directly after the mispronounced utterance (but not interrupting it!).

One student, who had a negative experience with a teacher (none of those participating in this study) who interrupted every utterance three or four times with corrections and long explanations, saw a clear conflict between feedback and communication, *"If the goal is to reach perfect pronunciation, then the teacher may interrupt whenever there is an error, but if it is that the students should be able to speak the second language, they must be allowed to finish the utterances"*.

Which?

Interestingly, when compared to the results of Ancker (2000), none of the students thought that all errors should be corrected, but only the worst. The significance of "worst" differed somewhat between the students, but the general opinion was that mispronunciations that changed the meaning of the word were the most important.

Notable is also that while the teachers had quite clear ideas of the different types of mispronunciations and some implicit rank of their importance, several of the students did not seem to think actively about the impact of different types

of mispronunciations, but rather regarded the pronunciation as correct or not, and placed the responsibility of defining the threshold for correctness entirely on the teacher (i.e., adhering to the dependent practice described by Morley, 1991).

One student wanted the teacher to correct recurring errors, showing that the student has not grasped a particular type of pronunciation, but avoid correcting slips (*"That is why I was so frustrated; I knew the correct form already, but I made a mistake, since I was concentrating on a more complex content"*).

Generally, this student primarily wanted correction on errors that lead to misunderstandings or deteriorated communication and, secondly, on errors that affect the listener's view of the speaker negatively. The above criteria for feedback hence correspond to those of comprehensibility, intelligibility, frequency and social impact given by the teachers.

Another student thought that mainly errors that deteriorated the understandability should be corrected, and that it apart from this depended on the student's ambition. This student argued that vocabulary was the most crucial, and that even if pronunciation is important, a perfect pronunciation should not be the goal too early, as the level of all parts of the language should match (*"You must expand the vocabulary at the same time as the pronunciation."*)

This opinion is hence somewhat different from that given by several teachers, that pronunciation should be the most important in the early training, in order to avoid learning wrong pronunciation patterns in the vocabulary. As stated above, the teachers' view was shared by some of the other students. They in particular thought that it was important to correct phoneme errors, as the error would otherwise be repeated in many different words.

How?

The question of how feedback should be given is, naturally, situation dependent with respect to the type of exercise and the number of students involved. The students wanted more detailed feedback at the introduction of new words and expressions, in order to try to get the pronunciation correct from the start. In case of word confusion between minimal pairs (e.g., "here" *vs.* "heir"), both words should be repeated in order to illustrate the consequences of the mispronunciation.

One student firmly thought that not too much feedback should be given at a time. Instead, there should be a focus of a session, and the feedback should primarily be given on this feature, one aspect at a time (corresponding to criteria 5 in

the list by El Tatawy, 2002). If other pronunciation problems are discovered, then the teacher should adapt another session, to elicit practice of the problematic pronunciation. If immediate correction is given, it should be in the form of recasts, in order not to interrupt the communication with long explanations. Such explanations of general rules should come at a separate occasion.

The most commonly advocated feedback strategy was that the teacher should recast the utterance (*"In most cases, you can tell the difference yourself."*) and let the student retry. The next step would be an exaggerated recast, stressing the important feature. A third feedback turn would be the combination of an explicit repeat–recast, where the teacher would repeat the student utterance (*"What you are saying sounds like this:"*) and recast it (*"It should sound like this:"*).

One of the students requested articulatory feedback instead and argued that the best was to observe a person saying the sound and getting an explanation of what the person is doing (*"You want to know how this sound should be made – physically. How the mouth shape should be and where the tongue is."*). Such articulatory feedback, together with simplified phonetic descriptions, had also been given in the language courses he had taken. This student thus clearly differs from the teachers' opinion that articulatory and phonetic instructions were of little use. One explanation for this may be that he was a quite advanced language learner in that he is now learning his fifth foreign language and may hence be accustomed to more detailed feedback.

Another feedback preferred by the students was to compare to resembling Swedish sounds or words, in particular since this was considered an efficient method to remember the pronunciation.

Motivation

The students did mention the embarrassment of being corrected in front of the whole class, and that not too much feedback should be given in whole class settings (*"When you have been corrected five or six times in front of the whole class you may not be so eager to speak anymore"*).

One intermediate level student remarked that it is not being corrected by the teacher in front of the class as such that is frustrating, but that the teacher may not be the principal addressee of the utterance. If the student responds to what another student has said or addresses the fellow students, it can feel inappropriate if the teacher intervenes to correct, comparable to the social faux pas of interrupting a normal conversation by answering a question addressed to someone else.

The same respondent argued that the students'

self-confidence is the most important and the teacher should primarily promote the students' willingness to speak and communicate. Note the difference between promoting the willingness and the actual amount of speaking, as an increased willingness need not be reflected in more student utterances (in the short term), but rather in a more important urge to speak. One way to achieve this is by letting the student feel that the teacher is interested in what the student has to say (i.e., the content) and not only by how the student says it (the form), which has consequences for which exercises that are suitable.

One student mentioned that positive feedback should be given when the student has made an effort or when a progress is made (i.e., corresponding to the assessment based on degree of change in Morley, 1991).

The impression that too much corrective feedback is given arises primarily from how intrusive the feedback is and, secondly, on how frequent. One student expressed frustration about corrections that were perceived to be at detail that were irrelevant for the comprehension (*"Any native speaker would have understood what I was trying to say. It was not correct, but they would have understood it."*). According to the questionnaire, this student would like to and tried to sound like a native speaker as much as possible, but this should hence probably be interpreted as a native-like pronunciation in fluency and communication skills rather than exact phoneme (or grammatical) correctness.

Another student also saw a conflict between the goal for a perfect pronunciation and communication (*"It depends on what you yourself think is the most important, if it is to have a perfect pronunciation or to be able to make oneself understood."*), and in general, the students were firmer than the teachers in their opinion that communication and vocabulary and grammar are more important than a native-like pronunciation, according to the PAI index.

Other students did not see the balance between feedback and communication as problematic, but rather that it will depend on the aim of learning the language (*"If you should speak before the Congress or just need it for two weeks traveling"*).

A conclusion to be drawn from the interviews is that the first of the criteria for good feedback in El Tatawy (2002) should not be imposed strictly; the students are clearly in favor of accepting more pronunciation errors in exercises where communication is perceived as the most important.

Further, the second of El Tatawy's criteria does not imply that all feedback must be explicit,

as the students believe that implicit feedback is enough. What it does imply is that the teacher should systematically test whether or not implicit feedback is enough and augment it if it is not.

Computer Assisted Pronunciation Training

When asked what the most important feature of a CAPT program should be, in order to be used, one student responded *"I want to feel that I am good. That I am making progress."*

When probed further, this signified in particular getting feedback at a meta level, regarding how many exercises and which feature sets that have been completed successfully (*"You have now made 314 exercises. You have completed the training on: Nasality, Voicing, Duration."*).

This student wanted the tutor to offer a smorgasbord of exercises grouped into different categories and the choice of category and when to switch to practicing another feature should be left to the student. The program should hence not predefine that one set of exercises must be completed before another is started. It may suggest a suitable set of exercises or provide recommendations regarding the level of difficulty of each set, but it should be possible to overrule these.

The user should have the option to choose different sets of exercises for different features or common pronunciation errors and the main feedback within this group of exercises should be on the feature that is practiced. Other errors should not be corrected, but suggestions on which other exercise sets that may be suitable should be provided depending on these other errors.

The training material could contain minimal word-pairs, where errors may occur, but in order to be more than temporarily interesting, the exercises should quickly advance to sentences, dialogs and simple role plays. Other students suggested pronunciation training on word lists, saying what is on a presented picture (i.e., without the written text as support), reading sentences and replying to questions.

In order for the exercises to run smoothly, the feedback given should be minimal as default and the user should click a button to get more details or explanations concerning the feedback. As the objective of the program is to practice pronunciation, more detailed feedback is adequate, but only when the student actively seeks it.

Regarding scoring of the exercises, getting points is a compelling feature, but it may be that the score should be given on the number of successfully completed exercises, rather than utterance by utterance. It was suggested that the final score should be given as a communicative impression, ranging from, e.g., unintelligible, acceptable

to correct and native-like.

The most important feature of a pronunciation training should be that you can have confidence in the feedback, that the feedback is relevant (the most important problem is highlighted) and correct (the system must not indicate an error where there is none). Further that you do get informative feedback, e.g. *"It would not be good if you only got an error message saying that something was wrong"*. Note that this is more or less the feedback given by many CAPT systems currently.

One student was at first somewhat skeptical about CAPT (*"How would feedback be given?"*), but then became more enthusiastic when another student proposed that you could have a talking face to speak with (*"Yes, that would be good. It would be like having a friend to practice with. Well, not a friend, but..."*). This student had earlier stated that it was a pity that so little one-to-one practice was given and the teaching was almost exclusively whole class, suggesting that she could indeed appreciate the virtual tutor training.

Another student proposed a set-up corresponding to a popular karaoke game. In the Playstation game in question, the players get the text of the song, together with bar icons indicating the key and target duration. Color coded bars, remarks such as "awful", "OK", "cool", and a numerical score give feedback on to which extent the key and duration was on spot. For pronunciation training, the student would get a text to read on the screen with the possibility to listen to it, then try reading and mispronounced words would turn red. The mispronounced words would then be practiced until corrected. This far, the description corresponds more or less to some existing commercial CAPT systems, but with much more feedback than is currently given in state-of-the-art CAPT systems. This respondent wanted to *"see how you shape your mouth and compare this to the correct, e.g., by placing the two over each other"*, but also acknowledged that the detail of feedback should be different for different students or different situations and that it should be possible to choose the amount and type of feedback at start-up from a simple menu, containing e.g., five options for different types of training.

One student commented on the differences between practicing with a real teacher and a virtual tutor. Firstly, when asked how he would like a CAPT system to be, he replied: *"What I really would like is a human."*, not meaning that communication would be more stimulating with a human, but that it is easier to see how the pronunciation is made on a human teacher. The second best would be a physical model of the head, but *"a computer screen visualization is quite OK as*

well.". He secondly thought that more feedback could be given by a CAPT program, as the user would be more focused on practicing pronunciation, like when studying individually. Thirdly, a virtual tutor was considered to have some advantages, mainly that *"You can take the time you want. You do not have to feel stressed. You can stop and take a break and continue later"*.

Implications for feedback management in CAPT

Neri et al. (2002a) conclude that effective feedback should be comprehensible, should not rely solely on the user's own perception, should allow the response correctness to be verified, should pinpoint errors and ideally suggest a remedy.

This conclusion is in accordance with the above interviews and observations, but these also indicate that other criteria are relevant, depending on the type of exercise. The two main categories of CAPT exercises are either feature-based or communicative practice. In the first case, the objective is to practice a certain articulation or feature on the word or sentence level, while in the second it is to solve a communicative task using dialog or role-play (i.e., a rehearsed dialog). As the two categories are different, so should the feedback be, as outlined below.

Another finding, from the teacher interviews and from the classroom observations in particular, was that very little feedback was actually given. This may, on the one hand, be taken as evidence for the need of CAPT, as the students do not get much feedback on the pronunciation in the classroom. On the other, it indicates that the amount of corrective feedback should be limited, to maintain a positive atmosphere and a communicative flow.

When?

In feature-based exercises, feedback should come after each student utterance, never interrupting. In communicative feedback, explicit feedback on pronunciation details should only be given after the dialog has ended. During the dialog, feedback should be on the communication level, i.e., clarification requests when the understandability or intelligibility is low (meaning that the ASR fails to recognize the utterance), or recasts of recurring student errors.

Which?

In feature-based exercises, corrective feedback should only be given on the practiced feature, but other errors should be monitored in order to be able to suggest suitable follow-up exercises. Pos-

itive feedback may, on the other hand, be given on the general production.

In communicative practice, the focus should be on understandability, which means that only utterances that could not be understood should lead to a correction or a clarification request. Feedback on pronunciation features should be provided *after* the dialog exercise. The decision on which pronunciation errors to give feedback on may be partly statistical (frequency of a certain type of error, score from the ASR for understandability), but a knowledge-based approach is most probably required (relating to social impact, intelligibility issues and generality). It is suitable that the basis is two-tiered.

At one level, the dialog exercise may have been constructed to elicit certain problematic pronunciations or words, and it is then natural to give feedback only on these targets – provided that the student knows that the pronunciation focus of the exercise is on these features or set of words.

The second level concerns more general errors, if the exercise has no specifically stated pronunciation focus. In that case a database is needed, where potential errors are stored and ranked with respect to graveness (social impact, comprehensibility, intelligibility) and whether it is a generally occurring error (so as to avoid correcting slips).

The feedback presented to the student should be limited to the, e.g., three most important, but all discovered errors or features should be stored in a student database, in order to be able to provide positive feedback on the student's progress.

How?

The type of increasing feedback suggested by one of the teachers, in which the student is first given the chance to self-correct, then to adjust to a recast and is explicitly corrected only as a last resort, seems appropriate regardless of if the exercise is feature-based or communicative. Within this strategy the use of implicit elicitation to give the student the opportunity to self-correct is potentially very fruitful. As the mispronunciation detection component cannot be assumed to be faultless, there might be a risk that a request for a repetition when the system judges that a mispronunciation has been found is confused with one where a repetition is needed because the system was unable to judge the utterance (due to a misrecognition, background noise etc.). These two requests for repetition should hence be managed differently, in order not to fool the student to change an already correct pronunciation.

In general, detailed feedback should be given upon request from the user, e.g., by pressing a but-

ton. As Bälter et al. (2005) found that the children testing the ARTUR system seldom used the provided buttons spontaneously, the tutor should indicate when important additional information may be received if requested.

Ideally, the user should then be able to choose from a set of different feedback presentations (an enhanced acoustic target; explicit identification of the error; articulatory animations or metalinguistic explanations) in order to suit different users' learning styles. The classroom observations indicated the importance of multimodal feedback, especially when the L2 was so different that acoustic targets did not suffice. While some common feedback strategies may be defined for all pronunciation training, the efficiency of different types of feedback is dependent on the L2 language and need to be adapted to it.

Motivation

Due to the varied wishes by different students, a user set feedback level may be a fruitful option, i.e., incorporating a slider with which the user can control individually how much feedback that is given. The threshold should be adapted to the L2, the exercise and the student performance (i.e., the student's proficiency), but in addition it should be possible for the student to adjust the amount of feedback for two reasons. The first is the affective, that students should be able to set the amount and detail of feedback to a level that they are comfortable with. The second is that this does put the responsibility and initiative with the student, who can decide how much advice he or she requires from the tutor, in conjunction with the teacher-coach methodology proposed by Morley (1991). This promotes self-monitoring and the feeling of owning the training situation, which should be beneficial for motivation and self-confidence.

Another strategy, observed in the classroom, could be to introduce a third classification category *Satisfactory for the time being* in addition to the standard *Correct* (no corrective feedback needed) and *Incorrect* (feedback needed). The reason for introducing the new category would be to avoid getting stuck on a problematic pronunciation, but at the same time not having to classify the last student utterance as correct. Depending on the history of the training session, the classification into the satisfactory category could result in responses similar to *"Yes, that was better!"* or *"Yes, this is difficult. Let's practice it more later.".* The first reply acknowledges that an improvement has been made, even if there is still a mispronunciation, while the second indicates that the problem still remains, but that it does not seem fruitful to continue with more corrective feedback.

While more feedback may be given in CAPT than in a classroom setting, as the situation is different regarding the type of exercise, the lack of potential embarrassment and the possibility to give layered feedback, the statement by one student *"I want to feel that I am good. That I am making progress."* should serve as a guiding star. The CAPT program should aim at improving the student's pronunciation, but it should also aim at improving the student's confidence in it.

Multilayered feedback in ARTUR

Some aspects of the feedback strategies proposed above have been implemented in a Wizard-of-Oz version of ARTUR that was tested by experts in human-computer interaction at the International Conference on Human Factors in Computing Systems, CHI 2006, in Montréal, April 24-27. The conference attendees got the opportunity to practice their pronunciation of the Swedish sound "sj" [ɕ] with an English speaking virtual teacher.

The graphical user interface consisted of two frames showing the progress within the exercise sentence, in Swedish and in English; one window showing Artur and articulatory animations; one transcription frame where all Artur's utterances were transcribed and one interaction frame.

The interaction controls let the user choose the amount of feedback given ("Little", "Medium" or "Much"), to have the pronunciation and the animation repeated at slow rate ("Repeat Slowly"), to have the difference between the target and the own pronunciation illustrated in an articulatory animation ("Show Difference"), to listen to the own previous try, to jump to the next word, to request an attempt at the whole training sentence, and to end the practice. Contrary to a previous user test with speech impaired children (Bälter et al., 2005), Artur did not explain the use of each button in the introduction, but merely encouraged the user to try them.

The practice sentence was the tongue twister "Sju själviska sjuksköterskor stjäl schyst champagne" ("Seven selfish nurses steal quite nice champagne"). The phoneme [ɕ] was chosen since it is unique for Swedish and would hence be unfamiliar to most conference attendees. It is further produced at the velum, which means that the articulation is difficult to infer from the speaker's face, while it can be illustrated with an articulatory animation of the tongue. The sentence also exemplifies different spellings of the same sound, and the first three words become progressively more complex, from the one-syllable "sju" to the

quite difficult "sjuksköterskor". The difficulty of the latter word was intentional, to provoke users to make use of the interaction buttons. It further illustrated how the practice on a compound word can be facilitated by trying its two parts separately and then retry the whole word (this divide-and-conquer strategy proved highly efficient).

As the training session was short and focused on one phoneme only, all parts of the above recommended feedback strategy were not tested. The training session was also too short to study longitudinal use of any kind, but the actual testing of the feedback in a virtual tutor by experts in the human-computer interaction field is nevertheless an important step in finding which types of feedback that can be directly transferred from human to virtual tutors.

Two slightly different practice openings were used. In both, Artur gave some instructions, explained the goal of the exercise (to practice the sound "sj") and then produced the tongue twister. In the first, the user was given the opportunity to try the whole sentence directly, and if the attempt was unsuccessful, the practice continued word by word. In the second, the practice began with the separate words directly, only prompting the user to try the sentence at the end of the practice. The first opening was tested with 17 of the respondents and the second with 20. The reason for trying the two openings was to see if it was beneficial to make the users more aware of their progress during the practice by letting them try the difficulty first, or if it was merely frightening.

Artur then explained and showed the articulatory difference between an English "sh" [ʃ] and the Swedish [ɕ]. In the following practice, the user was assumed to make primarily two types of error, either a pronunciation closer to the English [ʃ] or to a German [x], and detailed corrective feedback could hence be given for these two. In addition six other types of feedback were also given: positive (for a correct pronunciation); minimal positive (no explicit positive feedback was given, either in order to create a flow in the training or because the user had requested little feedback to be given); satisfactory (the pronunciation was not entirely correct, but it is pedagogically sounder to accept it and move ahead); augmented (a repeated error on the same word, more detailed feedback given); vague (a general hint is given, rather than an explicit feedback) and encouragement (encouraging the student and asking for a new try). The last two types of feedback could be used either when the error made did not correspond to the predefined types of errors, meaning that the detailed feedback would be inaccurate, or because more explicit feedback would af-

Table 3. Examples of Artur's utterances corresponding to different types of feedback, P=positive, P0=minimal positive, S=satisfactory, C=corrective for English type errors, CG=corrective for German type errors, A=Augmented for a repeated error, V=vague, and E=Encouragement.

Type	Example utterance
P0	Head nod
P	*Yes, correct!*
S	*Not bad, let's try the next word.*
C	*Think of how you make the word 'cue' and lower the tongue just a little bit in the back.*
CG	*The Swedish 'sj' is made further front than the German 'ch'.*
A	*Remember to round the lips and use the same part of the tongue as for the beginning of 'kelp'.*
V	*Think of which part of the tongue touches the palate.*
E	*Good, but try it once more.*

fect the user's motivation negatively. Examples of the different types of feedback utterances are given in Table 3. The feedback was different for each word, in order to minimize repetitions. Each feedback utterance was accompanied by either the face speaking or a computer animation of the articulation or the articulatory difference illustrated. In this set up, contrary to the one used in Bälter et al. (2005), Artur did not explicitly state how the user had produced the word, only giving instructions on how to improve (i.e., *"Start with the tongue in the position for 'shell', lower the tongue tip and raise the back part of the tongue instead."* rather than *"Now you said 'shellviska', lower…"*)

The users' pronunciation did indeed improve during the practice. The improvement depended to some extent on 1) the user's familiarity with Swedish (a few users did know some already), 2) the similarity of L1 sounds to [ɧ] (users who had more similar fricatives in the mother tongue found it easier to follow the articulatory instructions to adjust the place of articulation), 3) the similarity of the L1 to Swedish (probably because the other phonemes in the words were in that case more familiar and the user could concentrate more on the production of [ɧ]), and 4) the number of foreign languages the user spoke already.

When asked if they thought that their pronunciation improved by the practice, the users' answers can be grouped into three categories: those who 1) spoke some Swedish before and thought that they did improve; 2) did not speak any Swedish before and thought they learnt the pronunciation, at least temporarily, and 3) did not know if they improved, because it was the first time they tried Swedish.

The tendency that students speak more quietly when they are uncertain about a pronunciation, observed by the real language teachers, was apparent in the practice with the virtual tutor as well. A feedback to ask the user to speak louder is hence essential in order to get the best possible acoustic input to the system.

Users' feedback on the feedback

The users who completed a training session (of approximatively 5 minutes) were asked to fill in a questionnaire on how they perceived the feedback given by the tutor, in order to guide the future development of the feedback management in the system. The questionnaire, shown in Table 4 was answered by 37 users.

The main reason for choosing a questionnaire rather than an interview was the time constraint, that users could fill in the questionnaire individually while a new user tested the system. A secondary reason was to improve the chances for objective anonymous responses, since the test set-up was not neutral: 1) It was apparent to the users that the persons presenting the system were researchers involved in the project, rather than objective evaluators. 2) The users were asked to provide a peer review at an ongoing conference, which may induce more positive responses. 3) The set-up was such that the Wizard was visible. Most users did know that they were testing a Wizard-of-Oz system, either because they had been told when introduced to the test, or because they noticed the Wizard. Not all did, however. Several users thought the Wizard was doing something else, e.g., "checking his email".

All users but one were quite positive about the practice, as indicated by the mean opinion scores in Fig. 2. The users thought that the practice was clear, easy to follow and interesting. The articulatory instructions and animations were perceived as quite clear and the users judged the usefulness of an automatic ARTUR system as high.

It should be acknowledged that the responses may be biased as users were attendees who stopped at the ARTUR booth in the fair hall. It can hence be assumed that they were interested in or positive to CAPT before trying the system or that the social atmosphere at a conference venue led to more positive responses. The user comments given in the questionnaire, which are sum-

Table 4. The questionnaire given to the users who tried the training session with Artur. Each question could be answered on a scale from 1 to 9, where the extreme values were indicated as shown below, or with NA (no answer). For each question, comments could be given.

Question	Extremes		
1. What was your main impression of practicing with Artur?	1=confusing	9=clear	(A)
1. What was your main impression of practicing with Artur?	1=boring	9=interesting	(B)
2. Do you think the training improved your pronunciation of the Swedish sound practiced?	1=not at all	9=improved greatly	
3. What is your opinion about the articulatory explanations?	1=confusing	9=clear	
4. What is your opinion about the articulatory animations?	1=confusing	9=clear	
5. Do you prefer the text or the animation instructions?	1=text	9=animation	
6. Different types and levels of feedback were used in the training. Should the mix be	1=less explicit	9=more explicit	
7. What potential do you see for a fully automatic virtual pronunciation teacher Artur?	1=none	9=enormous	

marized in Table 5, do suggest that the users did try to respond objectively and provide constructive comments for improvements.

The one negative user seemed, on the other hand, overly negative, commenting e.g., that Artur's "synthetic" voice was unnatural and that the Wizard had a pronunciation that was a much better target than Artur's (Artur's utterances were actually recordings of the Wizard's natural voice!) and that the questionnaire was bad.

The mean opinion scores in Fig. 2 indicate that there were no statistical differences between the user group who was asked to make an attempt at the sentence to start with and the one that was not. From the comments, summarized in Table 5, it was nevertheless clear that many users in the first group felt intimidated by the task when first presented with an entire, and difficult, sentence. There was a tendency in the comments, and a slight – but non-significant – difference in the score, that this first group thought that their pronunciation had improved more. Making the students aware of their performance before and after the practice may hence be beneficial, provided that the goal is less difficult. Alternatively, the tutor can set up a goal (*"When we finish you will be able to say:"*), but begin the practice with a part at an adequate level.

As shown in Fig. 2b, there was a slight preference for the animations over the verbal instructions (*"The text was useful, but animations clearer.", "animation! No question. Didn't even notice the text!"*), but many users commented that they needed or liked the combination of instruction and animation, rather than any alternative separately (*"Matches together", "I liked the combination of both – sometimes one was better than the other, depending on the word.", "both!", "Animation + text = 9"*).

Some users found it difficult to relate the animations to their own pronunciation (*"Sometimes it's hard to follow the animations because it is hard to visualize the inside of your own mouth."*), but many commented that they actually became more aware of their own tongue movements by looking at the animations. Others commented that it would be useful to be shown the difference between the own articulation and the target (*"Show my own articulation overlaid on model's"*). A similar feature was actually provided on request from the user (the "Show difference" button), but most users did not try this button during the short training session.

Many users indicated that the amount and level of feedback was about correct as it was (question 6, Fig. 2b), but some wished more detailed feedback. Noteworthy is that few users tried to adjust the amount of feedback by changing the radiobutton from the default "Medium".

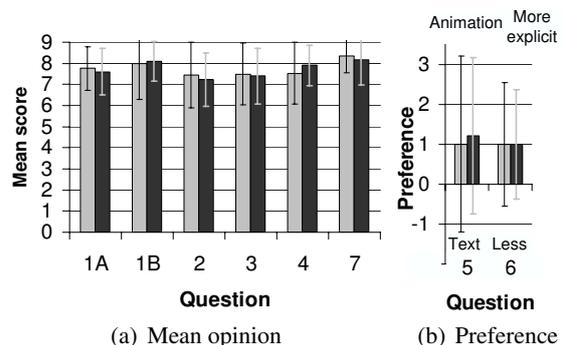(a) Mean opinion

(b) Preference

Figure 2. The user's mean opinions according to the questionnaire, with standard deviation indicated. The answers to questions 5-6 were transformed to preference scores, with 0 being neutral. Grey bars correspond to the group that started with the whole sentence, black to the one that went directly to separate words.

Table 5. Representative examples of user comments about the practice. **+** *indicate positive comments,* **-** *negative comments or requests for changes.*

Overall
    + A very impressive and intuitive method of learning.
    + It was fun - very engaging.
    + Helpful.
    + It can be helpful in adult learning situations. It is not patronizing/offensive, you don't "feel"
      foolish as for a real person. It is very patient.
    - I was overwhelmed when asked to pronounce the whole sentence, but
      breaking down by word or syllable was very helpful.
    - I got scared at the beginning when I got presented the whole sentence.
    - He speaks very fast.
Animations
    + The 3D model with the movement of tongue & lips is really useful.
    + Excellent. Very, very useful and practical. I would absolutely buy a language product
      that utilized the training method.
    + Visualization of the tongue is very nice. Absolutely necessary.
    + This is very useful and is the point that a real teacher cannot really do.
    - Most of the time the visual explanation of the pronunciation went too fast.
    - I like the option of a slower pronunciation but it could have been slightly faster than it was.
      Maybe allowing the participant to adjust the speed would be a nice addition.
    - Didn't help as much as I thought at first.
    - Would help to see the tongue from different angles.
Instructions & Feedback
    + Very easy to follow, I liked how it explains the position of [the] tongue.
    + Relating to other words I know is really helpful to me.
    + English comparisons were very helpful as were animations.
    - Too forgiving.
    - Using a sound that I had just mastered as an example later on made me less confident.
      [Referring to the feedback that "stjäl" is pronounced similarly to the beginning of "självviska"]
Suggestions for improvement
    More engaging base animation? "Artur" isn't that appealing...
    Maybe video instead of an avatar.
    Add some button to repeat the word again.

This might be explained by the fact that users did not have the time to explore the possibilities of the user controls during the short training session. A couple of users also remarked that it would have been good to be able to listen to the own attempt, *"It could be good for the user to be able to listen to his/her words as another way of feedback (comparing his pronunciation with Artur's pronunciation)"*, apparently missing that there was such a button on the interaction panel.

In general, there was a clear tendency that the subjects used the interaction buttons more if Artur had suggested it (some of the vague feedback referred to the buttons, e.g., *"You can see and hear an exaggerated pronunciation by pressing the "Repeat slowly" button."*). This was not only true directly following the suggestion, but later in the training as well. For subsequent words, users who had received such a suggestion started to use other buttons to some extent, but mainly continued using the ones that Artur had suggested.

This indicates that, at least in the short term, the user seldom voluntarily takes the initiative to go outside the teacher-driven practice and that the tutor can influence the amount and type of user interaction by suggestions. If specific features of the system are to be used it is hence important that the tutor encourages the student to use these explicitly during the first part of a training session.

A few users indicated that it would be beneficial to see the tongue from different angles, e.g., diagonally from the front so that the upper tongue surface is visible. As the 3D model may be rotated freely, it would indeed be possible for the users to adapt the view to their preference.

Several users requested the possibility to hear the target word again at normal speed (thinking that the "Repeat slowly" option was too slow). Such a button was presented in a previous user test (Bälter et al., 2005), but as it was never used then, it had been removed in order to keep the number of interaction buttons low for clarity.

The two latter suggestions indicate that an important feature to maximize usability is that the interface should be adaptable, by the user or a teacher, as suggested in Eriksson et al. (2005).

When questioned about the most important improvements needed in the interaction with the user, several respondents requested a more appealing tutor, either a video of a real teacher or a more video-realistic animation. With the fast improvement of realism in computer games and cinematic animations, user expectations are increasing, and the importance of the cosmetic appearance of the virtual tutor should not be neglected.

Conclusions

This study is quite small, and can hence only give qualitative indications of good feedback. These indications are nevertheless relevant, as they suggest both a slightly different view from that presented in earlier studies' recommendations for good feedback, and that the wishes for feedback are clearly individual. A virtual tutor has better potentials to meet such individual wishes than a teacher in a whole class setting and this potential should be exploited by letting the student decide the amount and type of feedback that is given.

In general, this study has shown that teachers and students do feel a need for computer assisted training and while there are individual differences there is a general consensus on the criteria for how feedback should be given efficiently. Many of the criteria are applicable to CAPT.

Very few of these criteria are yet implemented in state-of-the-art systems and while some of them can be quite easily implemented, others are further away. Nevertheless, some simple improvements in feedback management and presentation could drastically increase the benefit of already existing speech technology based pronunciation training. We have tried some of these improvements in an implementation of a virtual teacher and the feedback from the users indicate that articulatory instructions, and articulatory animations in particular, are perceived as useful. Varying the feedback, combining explicit instructions with vaguer suggestions and encouragements, was also perceived as positive by the respondents. Both interviews and the test suggest that slightly more detailed feedback can be given in CAPT. As the interaction buttons were seldom used spontaneously, a virtual tutor needs to drive the interaction more at the onset of the practice, until the user has found out how to use different features and can take a more active role in controlling how much and how feedback is given.

Finally, it should be emphasized that CAPT systems should give feedback that is relevant and detailed enough to allow the users to understand how to change a deviant pronunciation, but the details must be given in such a way that the practice is stimulating and that the students become proud about the progress already made and increasingly confident in their own production.

Acknowledgments

References

Ancker W (2000). Errors and corrective feedback: Updated theory and classroom practice. *FORUM*, 38:20–26.

Asakawa S, Minematsu N, Isei-Jaakkola T & Hirose K (2005). Structural representation of the non-native pronunciations. In *Proc of Interspeech*, 165–168.

Ayres R (2002). Learner attitudes towards the use of CALL. *Computer Assisted Language Learning*, 15:241–249.

Bälter O, Engwall O, Öster A & Kjellström H (2005). Wizard-of-Oz test of ARTUR - a computer-based speech training system with articulation correction. In *Proc of the Seventh Int ACM SIGACCESS Conf on Computers and Accessibility*, 36–43.

Beskow J (2003). *Talking Heads – Models and Applications for Multimodal Speech Synthesis*. Ph.D. thesis, KTH, Stockholm, Sweden.

Beskow J, Engwall O & Granström B (2003). Resynthesis of facial and intraoral motion from simultaneous measurements. In *Proc of ICPhS*, 431–434.

Bot K & Mailfert K (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly*, 16:71–77.

Bourgeois E (2002). A constructivist approach to adult learning. In A Bron & M Schemmann, eds., *Social science theories in adult education research*, 130–153. Münster, LIT Verlag.

CALICO (2004–05). The CALICO review. Http://calico.org/CALICO_Review/, last visited Nov 4th, 2005.

Carroll S & Swain M (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition*, 15:357–386.

Chaudron C (1977). A descriptive model of discourse in corrective treatment of learners' errors. *Language Learning*, 27:29–46.

Chomsky N (1975). *Reflections on language*. New York: Pantheon.

Deroo O, Ris C, Gielen S & Vanparis J (2000). Automatic detection of mispronounced phonemes for

language learning tools. In *Proc of the 6th Int Conf on Spoken Language Processing*, 681–684.

El Tatawy M (2002). Corrective feedback in second language acquisition. *Working Papers in TESOL & Applied Linguistics Teacher's College, Columbia University*. As retrieved on Oct 21, 2005 from http://www.tc.columbia.edu/academic/tesol/Webjournal/El-Tatawy.pdf.

Elliot R (1995). Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *The Modern Language Journal*, 75:530–542.

Engwall O (2002). *Tongue Talking – Studies in Intraoral Visual Speech Synthesis*. Ph.D. thesis, KTH, Stockholm, Sweden.

Engwall O (2003). Combining MRI, EMA & EPG in a three-dimensional tongue model. *Speech Communication*, 41/2-3:303–329.

Engwall O, Bälter O, Öster A & Kjellström H (2006). Feedback management in the pronunciation training system ARTUR. In *Proc of the Int Conf on Human Factors in Computing Systems, CHI*, 231–234.

Engwall O, Wik P, Beskow J & Granström B (2004). Design strategies for a virtual language tutor. In *Proc of the 8th Int Conf on Spoken Language Processing*, 1693–1696.

Eriksson E, Bälter O, Engwall O, Öster A & Kjellström H (2005). Design recommendations for a computer-based speech training system based on end-user interviews. In *Proc of the 10th Int Conf Speech and Computers*, 483–486.

Hendrickson J (1978). Error correction in foreign language teaching: Recent theory, research, and practice. *The Modern Language Journal*, 62:387–398.

Hincks R (2002). Speech recognition for language teaching and evaluating: a study of existing commercial products. In *Proc of the 7th Int Conf on Spoken Language Processing*, 733–736.

Hincks R (2005). *Computer Support for Learners of Spoken English*. Ph.D. thesis, KTH, Stockholm, Sweden.

Krashen S, ed. (1981). *Principles and Practise in Second Language Acquisition*. English Language Teaching Series. London: Prentice-Hall.

Lightbown P & Spada N, eds. (1999). *How Languages are Learned*. London: University Press.

Lyster R (1998). Recasts, repetition and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition*, 20:51–81.

Lyster R & Ranta L (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition*, 20:37–66.

Machovikov A, Stolyarov K, Chernov M, Sinclair I & Machovikova I (2002). Computer-based training system for Russian word pronunciation. *Computer Assisted Language Learning*, 15:201–214.

Mackey A, Gass S & McDonough K (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, 22:471–497.

Mackey A & Philip J (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal*, 82:338–356.

Massaro D (2003). A computer-animated tutor for spoken and written language learning. In *Proc of the 5th Int Conf on Multimodal Interfaces*, 172–175.

Massaro D, Bosseler A & Light J (2003). Development and evaluation of a computer-animated tutor for language and vocabulary learning. In *Proc of the 15th Int Congress of Phonetic Sciences*, 143–146.

Massaro D & Light J (2003). Read my tongue movements: Bimodal learning to perceive and produce non-native speech /r/ and /l/. In *Proc of the 8th Eurospeech*, 2249–2252.

Mead G, ed. (1934). *Mind, Self and Society*. London: University of Chicago Press.

Menzel W, Herron D, Bonaventura P & Morton R (2000). Automatic detection and correction of nonnative English pronunciation. In *Proc of Workshop Intergrating Speech Technology in the (Language) Learning and Assistive Interface, InStil*, 49–56.

Morley J (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, 25:481–520.

Murphy J (1991). Oral communication in TESOL: Integrating speaking, listening and pronunciation. *TESOL Quarterly*, 25:51–75.

Neri A, Cucchiarini C & Strik H (2002a). Feedback in computer assisted pronunciation training: technology push or demand pull? In *Proc of the 7th Int Conf on Spoken Language Processing*, 1209–1212.

Neri A, Cucchiarini C & Strik H (2002b). Feedback in computer assisted pronunciation training: When technology meets pedagogy. In *Proc of CALL Professionals and the Future of CALL Research*, 179–188.

Neri A, Cucchiarini C, Strik H & Boves L (2002c). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, 15:441–467.

Oliver R (1995). Negative feedback in child NS-NNS conversation. *Studies in Second Language Acquisition*, 17:459–481.

Öster A (1999). Strategies and results from spoken L2 teaching with audio-visual feedback. *TMH Quarterly Progress & Status Report, KTH*, 1-2:1–7.

Öster A, House D & Green P (2003). Testing a new method for training fricatives using visual maps in the Ortho-Logo-Paedia project (OLP). In *Proc of Fonetik, Umeå University, PHONUM 9*, 89–92.

Rubin J, ed. (1994). *Handbook of Usability Testing*. New York: John Wiley & Sons Inc.

Schachter J (1991). Corrective feedback in historical perspective. *Second Language Research*, 7:89–102.

Scovel T (1969). Foreign accents, language acquisition and cerebral dominance. *Language Learning*, 19:245–254.

Stevick E W (1978). Toward a practical philosophy of pronunciation: Another view. *TESOL Quarterly*, 12:145–150.

Tutsui M, Masashi K & Mohr B (1998). Closing the gap between practise environments and reality: An interactive multimedia program for oral communication training in Japanese. *Computer Assisted Language Learning*, 11:125–151.

Vicsi K, Roach P, Öster A, Kacic Z, Barczikay A, Tantoa A, Csatári F, Bakcsi Z & Sfakianaki A (2000). A multilingual teaching and training system for children with speech disorders. *Int Journal of Speech Technology*, 3:289–300.