

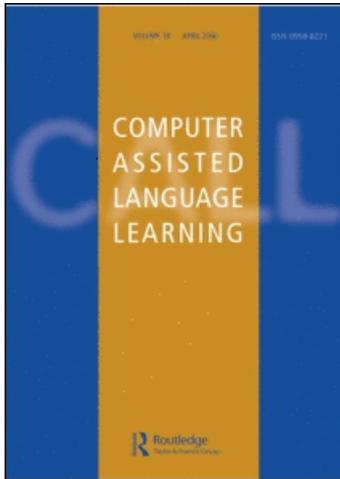
This article was downloaded by: [Radboud Universiteit Nijmegen]

On: 12 May 2009

Access details: Access Details: [subscription number 908165048]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Computer Assisted Language Learning

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t716100697>

Pronunciation feedback from real and virtual language teachers

Olov Engwall; Olle Bälter

Online Publication Date: 01 July 2007

To cite this Article Engwall, Olov and Bälter, Olle(2007)'Pronunciation feedback from real and virtual language teachers',Computer Assisted Language Learning,20:3,235 — 262

To link to this Article: DOI: 10.1080/09588220701489507

URL: <http://dx.doi.org/10.1080/09588220701489507>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Pronunciation Feedback from Real and Virtual Language Teachers

Olov Engwall* and Olle Bälter

KTH, Sweden

The aim of this paper is to summarise how pronunciation feedback on the phoneme level should be given in computer-assisted pronunciation training (CAPT) in order to be effective. The study contains a literature survey of feedback in the language classroom, interviews with language teachers and their students about their attitudes towards pronunciation feedback, and observations of how feedback is given in their classrooms. The study was carried out using focus group meetings, individual semi-structured interviews and classroom observations. The feedback strategies that were advocated and observed in the study on pronunciation feedback from human teachers were implemented in a computer-animated language tutor giving articulation feedback. The virtual tutor was subsequently tested in a user trial and evaluated with a questionnaire. The article proposes several feedback strategies that would improve the pedagogical soundness of CAPT systems.

Introduction

The success of computer-assisted pronunciation training (CAPT) technology is still limited, despite the great potentials that it holds. The major advantage of CAPT compared to self-studies using audio cassettes is that speech technology may be used to identify problems in the student's pronunciation. Hincks (2002), however, points out that: "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. Systems that do present visual feedback on the student's utterance, such as the IBM SpeechViewer (Adams, Crepy, Jameson, & Thatcher, 1990) or Box of Tricks (Vicsi *et al.*, 2000), can be used successfully when the teacher and the student are working together. In an assisted training session, the teacher indicates which features the student should look

*Corresponding author. Centre for Speech Technology, KTH, Lindstedtsv. 24, SE-100 44 Stockholm, Sweden. Email: engwall@kth.se

for in the visual representation and advises the student how the target pronunciation should be reached. Without this teacher support, the systems are difficult to use.

Unsupervised CAPT systems (refer to Tutsui, Masashi, & Mohr (1998) or the CALICO Review (CALICO, 2004–05) for an overview of such systems) must try to replace the feedback that the teacher would give in a supervised session by an automatically generated equivalent. One possibility is to use automatic speech recognition (ASR) to judge the learner's pronunciation based on a statistical comparison with that of native speakers. One problem with using ASR is that it is still error-prone, but an even more critical point is that the student often does not get any indication on how the pronunciation should be changed. Instead a score is presented, indicating the quality of the pronunciation, as judged by the ASR. The score can be presented numerically or graphically, as suggested by Machovikov, Stolyarov, Chernov, Sinclair, & Machovikova (2002). They plot the pronunciation score on a target with circles representing standard deviations for a native speaker model and a second display showing the pronunciation level within the word as a line in the range from bad to good. This graphical representation may be more appealing, since it relates the learner's pronunciation more clearly to that of native speakers, but the fact remains that the training is still purely trial and error.

Pedagogically more effective CAPT software should help the learner by indicating what the error is, and how the pronunciation could be improved. One possible way to achieve this is to adapt the ASR to learners of a specific mother tongue (L1) by recording their non-native pronunciation of the target language (the L2) as well as native L2 speakers' correct pronunciation (Deroo, Ris, Gielen, & Vanparis, 2000; Asakawa, Minematsu, Isei-Jaakkola, & Hirose, 2005). Instead of judging how close to the native pronunciation the learner is, the ASR may then classify the pronunciation into accepted and various unaccepted categories. For each of the unaccepted categories of typical errors that the L1 speakers are prone to make, feedback may be pre-generated to indicate how that particular error should be addressed by the learner.

It should be acknowledged that such L1-specific systems are even more vulnerable to ASR errors, since recognition errors may lead to confusing feedback (Menzel, Herron, Bonaventura, & Morton, 2000). Technological improvements of the ASR are hence vital in order to achieve a major break-through for CAPT. Neri, Cucchiarini, Strik, & Boves (2002) nevertheless argue that it is possible to build successful CAPT systems with available technology, if the learners' needs are carefully considered during system development. They maintain that the main problem of existing CAPT software lies in the lack of strategies for how feedback should be given in a pedagogical manner.

One attempt to answer this pedagogical need is to create virtual tutors (Massaro & Light, 2003; Engwall, Bälter, Öster, & Kjellström, 2006a), where talking head models interact as language teachers with the user. There are several reasons for using a computer-animated talking head as a virtual teacher. One of the most important is that previous studies (Walker, Sproull, & Subramani, 1994; Koda & Maes, 1996; Lester *et al.*, 1997; van Mulken, André, & Müller, 1998) have shown that users who

interact with an animated agent spend more time with the system, think that it performed better and enjoy the interaction more. Since the time spent on the training and the user's motivation are key issues for learning, much may be gained by adding a virtual tutor. A second reason is that non-verbal signals are very important in communication, and the feedback may hence be more effective if it is given by an expressive talking head that underlines the important parts of the instructions. A third reason is that the interaction with a virtual tutor is more similar to human-human communication, and one may hence take advantage of the learner's experience of human teachers. If the feedback from the system resembles that given in the classroom, then understanding it is straightforward. A fourth reason is that a virtual teacher can use computer animations to illustrate how the pronunciation should be made, rather than just explaining it verbally. An animated tutor may even show how the student should position the tongue for different phonemes. Such a focus on the learner's articulation may however lead to a counter-productive self-consciousness if too much feedback is provided. It is thus of particular interest to study how human teachers assure that the feedback is constructive and promotes the learner's motivation.

The aim of this study is hence to investigate feedback strategies that human language teachers use to help students correct phoneme errors and assess how they could be used in a virtual tutor. The focus on the phoneme level signifies that the feedback is related to the pronunciation of the L2 sounds, rather than to higher level features, such as prosody.

We first report on an introductory literature survey of studies on pronunciation feedback in the language classroom. The investigation then continues with our interviews with language teachers and students, which is used to explore their opinions on pronunciation feedback. They are followed by classroom observations focused on how the advocated feedback strategies are employed in practice. We conclude the part on feedback from human teachers with a summary of feedback methods that could be used in a virtual pronunciation tutor. We subsequently describe a small-scale user test and evaluation of a CAPT implementation that incorporates the listed feedback guidelines.

Feedback in Pronunciation Training at the Phoneme Level

The task of language teachers in pronunciation training at the phoneme level is to make the students perceptually aware of the distinction between features in the second language and those in the mother tongue, to help them understand how to realize the distinction, and to use repeated training to stimulate the student to be able to produce the phoneme without conscious planning.

There is still an on-going debate as to whether or not corrective feedback on pronunciation errors should be provided. In this article, we take the standpoint that feedback is beneficial and investigate how it may be provided as effectively as possible. In the type of pronunciation training that is envisaged here, the teacher detects pronunciation errors, diagnoses the cause, gives feedback on how to improve the

pronunciation and provides opportunities for practise. Even though the teacher has this responsibility to judge pronunciation errors, it is a key issue to involve the students in the monitoring of their own speech. Morley (1991) stressed that the role of the teacher is to be a “pronunciation coach” for the student (the “speech performer”), shifting from teacher-led and teacher-assessed to independent, self-monitored practise. A pronunciation coach should, in addition to providing diagnostic analysis and prioritizing exercises, help the students develop self-rehearsal and self-monitoring techniques.

The remainder of this article will analyse and discuss feedback in terms of issues related to a set of questions on pronunciation feedback: when errors should be corrected, which errors should be corrected and what kind of feedback should be used. These questions are based on and adapted from Hendrickson (1978). In addition, we introduce the issue of how student motivation is maintained. Since this may be achieved by adapting either the feedback or the practise as such, the sections on motivation will to some extent discuss exercises as well as feedback.

When Should Errors be Corrected?

Until the beginning of the 1970s, the predominant way of teaching pronunciation was to point out every error immediately (Ancker, 2000). A new methodology then appeared stating that not all errors should be corrected by the teacher. Further, if an error was corrected, the corrective feedback should not come immediately (Krashen, 1981). There are two main reasons for not giving immediate corrective feedback. One is that not all deviations are errors, in the sense that the student is unaware of the correct pronunciation. They may instead be mistakes that occur naturally in the mother tongue as well. A second reason is the affective factor, that students may lose self-confidence if they are corrected immediately all the time. The dominating opinion in the current literature on language learning is that some, but not all, errors should be corrected and that the feedback should be provided in close relation to the mispronunciation, but without affecting the students’ self-confidence and their own monitoring of the pronunciation.

Which Errors Should be Corrected?

Morley (1991) argued that the learners’ focus should be on functional intelligibility, functional communicability and increased self-confidence. Speech monitoring and modification strategies for use beyond the classroom should also be developed. The decision on which errors to correct and the feedback given should suit these aims. It could be noted that the teacher’s decision on which errors to correct may not necessarily coincide with the students’ expectations. While the vast majority (75%) of the teachers in a study by Ancker (2000) agreed that not every error should be corrected, the students to the same degree (76%) thought they should. Even if the decision on correcting an error or not should primarily be based on the

effectiveness of the feedback, it hence needs to take student expectations into account.

What Kind of Feedback Should be Used?

Lightbown and Spada (1999, p. 171) define corrective feedback as “any indication to the learners that their use of the target language is incorrect” and feedback may hence be given in many different forms. Lyster and Ranta (1997) classified the feedback types used by four language teachers as:

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

Recasts—the teacher reformulates the student’s utterance, removing the error.

Repetition—the teacher repeats the student utterance *with* the error, using intonation to indicate where the error occurred. Repetitions may also be used as positive feedback on a correct utterance.

Clarification requests—the teacher urges the student to reformulate the utterance, because the meaning was unclear.

Metalinguistic feedback—the teacher comments or asks questions to make the students find the error themselves with the information given by the teacher.

Elicitation—the teacher encourages the student to provide the correct pronunciation by open-ended questions or fill-in-the-gap utterances.

The many studies that have been performed on feedback in pronunciation training have come to slightly different conclusions on what kind of feedback is the most effective. Stevick (1978) strongly advises the teachers to only use recasts (for an erroneous student utterance) and repetitions (to reinforce a correct one), since too much explicit feedback would lead to tension. Carroll and Swain (1993) conclude that any kind of corrective feedback was beneficial, but *explicit* feedback gave the best results. One explanation may be that students often do not get the opportunity to repair after recasts, due to topic continuation (Sheen, 2004). Another reason is that recasts are perceived as a communicative confirmation rather than a correction (Lyster, 1998).

Several studies, e.g. Hendrickson (1978) and Lyster and Ranta (1997), have found that the students’ performance is more likely to improve if they are encouraged to try to correct the mispronunciation themselves (after metalinguistic cues, a clarification request or an elicitation) than if a recast is provided. The students should hence be made active in the correction, by using corrective techniques that elicit them to modify and self-repair their utterances.

El Tatawy (2002) argues that feedback should be provided systematically and consistently; that it should be clear enough to be perceived as feedback; that time and opportunity for self-repair should be given; that the feedback should be fine-tuned, in the sense that the error and the learner’s perception of the feedback should match closely; that one error should be in focus at a time and over a period of time in order to be consistent and intensive; and that the learner’s readiness to process the feedback should be accounted for.

How is Student Motivation and Self-confidence Maintained?

The problem of not discouraging the student with too much corrective feedback has been discussed in many studies, including, e.g. Ancker (2000), Murphy (1991) and, in particular, Stevick (1978). The teacher must find strategies to balance the students' need for corrective help against the task of promoting students' self-confidence. Morley (1991) proposed assessment of individual improvements rather than of the absolute level and emphasis of positive features. In this way, the students' recognition of self-accomplishment will be promoted, and they will become aware of even small successes in modifying features in the pronunciation.

Conclusions: Implications for CAPT feedback based on the literature review

Based on the literature survey, the following guidelines could improve the effectiveness of CAPT.

When should feedback be given?

- Directly following the user's utterance, but allowing time for self-correction.

Which errors should be corrected?

- Errors that disrupt the intelligibility and communicability.
- Errors related to the focus of the exercise.
- Errors that the student is able to correct given the proficiency level.
- Errors that are at a proficiency level for which the learner wants feedback, as indicated by a user set level of ambition.

What kind of feedback should be given?

- Feedback that is varied and of different types (recasts, repetitions, elicitations, and metalinguistic explanations).
- Feedback that is suitable for the exercise and the error.
- Feedback that involves the learner actively in pronunciation monitoring and correction.
- Feedback that is explicit enough to be perceived, but limited enough to not intrude on the exercise more than necessary.

How is student motivation and self-confidence maintained?

- Limit the amount of corrective feedback and let the student request more feedback if wanted.
- Balance corrective feedback with positive feedback and encouragements.

- Promote the student's feelings of self-accomplishment by highlighting relative improvements.
- Adjust and vary the type of feedback to avoid negative self-consciousness.

Data Collection

The data in this study were collected using teacher and student interviews and classroom observations. Most previous studies have concentrated on the language 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. It is, however, also fully acknowledged that students are not always aware of what is most effective for their own acquisition and the student opinions need hence to be balanced with the teachers' professional view and observations.

Interviews

Six language teachers, who teach adult engineering students at KTH (the Royal Institute of Technology), Stockholm, Sweden, and five students were interviewed. Four of the teachers were gathered in one one-hour focus group meeting and three students in another. In addition, two teachers and two students 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 individual interviews were 20–40 minutes long. The reason for using both focus group meetings and individual interviews was that a focus group meeting is suitable for collecting consensus views and gathering opinions on issues that the interviewer has not foreseen, whereas the individual interviews allow for deeper and more detailed analysis of personal strategies and preferences. The combination of the two was hence made to cover both levels. The two students who were interviewed separately had further a more diverse language training experience, including private lessons, small class lessons and intensive courses, which made it more suitable to interview them individually.

The teacher and student groups were intentionally heterogeneous with respect to target language (Swedish, English, German, Spanish, French, Polish, Japanese and Chinese) and student level, in order to capture general pedagogical strategies, rather than language or level specific solutions. As pronunciation training is most important at beginner level, there was a preponderance of teachers and students active at the beginner and intermediate levels. Five out of the six teachers are native speakers of the language they teach. The mother tongue of the majority of the students is Swedish.

The discussions were centred on the framing questions introduced above. In the interviews more focus was placed on repeated errors and the balance between correcting and motivating the students.

The interviews with the students also included a closing discussion on how they would like a CAPT system to function, in general and concerning feedback in

particular. None of the students had any previous experience of practising pronunciation with computers, but had used software for practising vocabulary and grammar in which their input was made through the keyboard.

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. Three different classes were observed: one with 26 students at the third session (each session consisting of two 2-hour lessons) of a course in German, one with 18 students at the end of a 56-hour course in Spanish, and one with 25 students at the fourth session of a course in Chinese. The second languages that the students were learning were hence from the same specific language group (the Germanic) in the first class, belonging to the same larger family (the Indo-European) in the second, and from a different one (the Sino-Tibetan) in the third. These relationships between the L1 and the L2 are manifested in the similarities of the phoneme inventories and the phonology.

The students in the first class were quite active in speaking the L2 during the class, responding individually in simple communication exercises, reading dialogues in whole-class or pair-wise and repeating words and sentences after the teacher. The pronunciation practise that the students received in the second class was repeating words together after the teacher, reading prepared answers out loud, reading a paragraph of a text for the class, talking in small groups with one or two other students, and a dictation exercise. The practise in the third class was focused on practising isolated words and short utterances, and student participation was more guided. The training consisted of: repeating after the teacher; practising specific details oneself; pair-wise exercises with short dialogues, either read or based on an example given orally by the teacher; and short dialogue interactions between one student and the teacher.

Results

The results in this section are divided into opinions on feedback given during the interviews and the observations of how feedback was distributed in class. Both the interviews and classroom observations are first analysed under the four different categories: when feedback should be given, for which errors, what kind of feedback should be used, and how to promote student motivation. The implications for CAPT are then discussed in a separate section.

Interviews

The teachers deemed that very little feedback on pronunciation is in general given in their classes, mainly due to time limits and because they are reluctant to interfere with the students' communication. The teachers, however, also argued that the students

tend to make progress in the pronunciation throughout the course even without much corrective feedback. This opinion was supported by the students, who argued that hearing the target language frequently and getting many opportunities to speak are more important factors for pronunciation improvement than receiving actual feedback. Both teachers and students hence favoured self-monitoring from the students based on a perception–production relationship.

As expected, there was a difference between student proficiency levels, with more feedback being given in beginner-level courses. There was also a quite natural influence of the phonetic similarities between the L1 and the L2. For languages that are more phonetically and phonologically different from the L1, the teachers indicated that they give more feedback and the students also requested more.

When should feedback be given?

There was a large consensus among the teachers and students about the importance of the timing when providing feedback. Feedback should not interrupt communication or the speaker's train of thought. Even if it means that errors are left uncorrected, the teachers prefer to provide feedback when it does not affect the communication. The students in addition thought that more general errors could be addressed if the feedback is not immediate, which means both that it is less personal and that it is easier to generalise to other words.

The students, however, also stressed that corrections should come as early as possible, both in the sense that it should follow directly after the mispronounced utterance and that it should come early in the course to avoid errors at the phoneme level spreading to other words in the vocabulary.

Which errors should be corrected?

The teachers suggested several criteria for the choice of errors to be corrected, depending on both the error and the student.

The error criteria could be summarised as follows.

- Understandability: if the meaning of the utterance could not be understood.
- Intelligibility: if the words of the utterance were not perceived correctly.
- Comprehensibility: if the utterance required much effort to be understood.
- Frequency: if the student repeats the same (type of) error several times.
- Social impact: if the listener gets a negative impression of the speaker making the error.
- Generality: if the error is one that is often made in the L2 either by foreign speakers in general or the L1 speakers in particular. The criterion addresses both the effectiveness for the group, i.e. whether more students benefit from the feedback, and the likelihood that such general errors may be more difficult to overcome without teacher assistance.

- **Commonality:** an error that occurs among native speakers of the L2 language is regarded as less serious 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.

Student-related criteria identified the following.

- **Proficiency:** a student with a better overall pronunciation may get corrective feedback on an error for which a student with a lower proficiency level does not. This criterion was not universal, as some teachers gave students with a better pronunciation significantly less corrective feedback, both in order to increase their self-confidence and to be able to spend more time on weaker students.
- **Personality:** a student who appreciates corrections receives more than one who does not.

In addition, another criterion put forward by the teachers was quite naturally:

- **exercise focus:** feedback is primarily given on the feature targeted by the exercise.

Notable is that while the teachers had quite clear ideas of the different types of mispronunciations and some implicit rank of their importance, the beginner level students did not seem to think actively about the effect of different types of errors, but rather regarded the pronunciation as correct or not, and placed the responsibility of defining the threshold for correctness entirely on the teacher. Contrary to the results of Ancker (2000), all of the students thought that not all errors should be corrected, 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. The beginner-level students argued that it was otherwise only the student’s ambition that should influence which errors should be corrected.

The more experienced language learners who were interviewed individually had more specific opinions about which type of errors they wanted corrected. One of them asked for feedback in line with the first six criteria given by the teachers. In addition, feedback should be given on the focus of the session only, one aspect at a time. If other pronunciation problems are discovered, then the teacher should make use of another session to elicit practise of the problematic pronunciations, rather than correcting them immediately.

What kind of feedback should be used?

Recasts and repetitions were the most advocated by the students, as they considered that it was often enough to hear the correct pronunciation. The teachers added that recasts and repetitions were used as much to provide a correct and clear target pronunciation for the entire class, as to give feedback to the individual student.

Increasing feedback was proposed by both teachers and students. The students suggested that the teacher should first recast, then recast with emphasis on the important feature, and finally both repeat the erroneous student utterance and recast it. Among the teachers' suggestions for increasing feedback, it was proposed to first use minimal, non-verbal, feedback to give the student the opportunity to self-correct before recasting. The teachers also proposed to use an explicit explanation of the difference between the correct and erroneous pronunciations in the last step, but only if the error was important.

Comparisons to familiar phonemes (in the L1, the L2 or another L2) together with the recast were preferred by the students, who in particular considered it an effective method to remember the pronunciation. The teachers also stated that they use such comparisons instead of phonetic descriptions or articulatory instructions.

Articulatory instructions, phonetic descriptions or sketches on place of articulation are not used by many of the teachers, as they thought that the students are unaccustomed to actively thinking about how they produce different sounds. Some teachers do use such instructions, even opening the own mouth wide, pointing to where the sound should be produced and encouraging the students to actively explore different places of articulation. One of the more experienced students particularly requested articulatory instructions and argued that the most effective technique is to observe a person saying the sound and receive an explanation of what the speaker is doing.

Sensory feedback was advocated by the teachers to remind the students of features that are clearly different in the L2 compared to the L1 in a particular aspect, e.g. voicing or aspiration.

The three last types of feedback above (comparisons, articulatory and sensory) can be considered to be different forms of *explicit feedback*, since they require that the teacher first states that the student utterance was mispronounced and then provides guidance on how to improve it. It should be noted, however, that at least the articulatory and sensory types are explicit feedback that actively involves the students' cognition or perception, rather than simply providing the correct form. They may hence also be considered as *metalinguistic explanations*.

Contrastive feedback is another type of metalinguistic feedback that was proposed by both teachers and students. If word confusion occurred because the phonetic feature that distinguishes between two words was mispronounced, then the two words should be repeated, to establish the difference in terms of perception and production.

Metalinguistic explanations are also used by the teachers to reinforce the feedback and to show why it is important that this particular pronunciation is correct, to avoid word confusion or a negative social impact. The importance of making certain distinctions may be illustrated in the mother tongue by confusions that would result between L1 words if a phonetic feature is mispronounced.

General recommendations, such as "You should try reading aloud by yourself at home", are used when the teachers do not consider that additional feedback was timely or motivationally constructive. What the teachers are aiming for is that the

students should practise both the perception of their own production and the articulatory movements.

How is student motivation and self-confidence maintained?

The teachers stressed that the amount of feedback needs to be limited in order not to create negative self-consciousness. One suggestion on how to avoid this was to make explicit explanations impersonal and expand the feedback to a general error that the whole class may have difficulties with. Another suggestion was to insert exercises with non-problematic pronunciation among the more difficult ones in order to make the students feel that they are making progress.

The students agreed that motivation is maintained by avoiding too frequent, too detailed and too lengthy corrective feedback. Efforts and improvements should also be praised by means of positive feedback and communication should be stimulated.

Classroom observations

The classroom observations confirmed that many of the feedback strategies that were proposed in the interviews were actively used in the classroom. Some new information was however also obtained, as outlined below.

When should feedback be given?

The feedback management differed somewhat between the groups. In one, little corrective feedback was given in the whole class, whereas it was more common in the other two, but with corrections always being balanced by positive encouragements for improvements.

The teachers did adhere to the strategy of letting the students finish before providing feedback, unless the student clearly appeared to need the teacher's help. In that case, a recast was provided by the teacher. This recast was then repeated, as a reminder, when the student had finished.

A failed student correction after a recast was almost always dropped, either for affective reasons or time constraints. Instead the difficulty was acknowledged, with statements such as "Yes, this is a difficult sound". The reason for doing so was probably both affective and attention catching. By stating that the pronunciation is difficult in general, the student who made the mispronunciation becomes less personally targeted. At the same time, the other students are made aware that the feedback may be relevant to them as well.

Which errors should be corrected?

Feedback was most commonly provided for errors made on the phonetic feature that was in focus of the specific exercise or on features that are distinctive in the L2, but not in the L1, such as aspiration, fricative voicing or tone. Such errors were corrected

both because they affect the intelligibility or comprehensibility and because the students need to be reminded of features that they are normally not aware of.

What kind of feedback should be used?

Recasts and repetitions were the most common feedback in the classroom (constituting well over half of the total number of feedback utterances from the teachers), similarly to the findings of, for example, Lyster and Ranta (1997), Nicholas, Lightbrown, and Spada (2001) and Sheen (2004). The repetitions in the classrooms 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 softly uttered) target for the class. Recasts were used in the same way and situation, with the difference that recasts had more stress and longer duration. Noteworthy is that the students tried to repair after recasts, contrary to claims in, for example, Lyster (1998) of the inefficiency of recasts.

Implicit and explicit elicitations (such as “Sorry?” and “Could you repeat that?”) were the second most common reply to erroneous pronunciations, hence encouraging self-correction. A repetition was subsequently used for a successful self-repair and a recast for an unsuccessful, accompanied by either a pause or an exaggerated intonation indicating that the pronunciation had not been correct.

Explicit corrections were indeed more common than in, for example, the study by Sheen (2004), if the three types of comparative auditory feedback, sensory and articulatory are grouped into this category (then accounting for about one-fifth of the feedback utterances). They were explicit only in the sense that they provided information on the student’s mispronunciation and on the correct target. The teacher did most often not explicitly state that the student had made an error. Of the explicit corrections, *Comparisons to familiar phonemes* were the most common, both comparing to words in the L1 that sound similarly and to phonemes that had already been practised in the L2. *Sensory feedback* and *articulatory instructions* were much more infrequent, occurring once or twice in each of the observed lessons. The sensory instructions consisted of asking all students to place 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. The articulatory instructions concerned features that could be easily illustrated, such as lip rounding.

Metalinguistic feedback was used when repeating the pronunciation rule that had been broken or in *contrastive* repetition of confounded word pairs.

The *general recommendation* to practise reading out loud at home was given after an unsuccessful repair following a recast.

How is student motivation and self-confidence maintained?

The students’ willingness to make pronunciation attempts varied in the three different classes and it seems that it was influenced by their previous familiarity and exposure to the L2 language. In the class with the L2 that had the least phonetic similarities with the L1, the students were insecure about their pronunciation skills

and therefore more reluctant to practise pronunciation. The factor of how familiar the L2 language is to the students should therefore be taken into account when determining the appropriate type, frequency and detail of feedback.

As already mentioned in the section on when feedback should be corrected, the teacher most often avoided repeating corrective feedback after one or perhaps two failed repairs, and also acknowledged that the pronunciation could be difficult, which can be interpreted as a strategy to avoid discouraging the student.

Finally, and almost self-evidently, the teachers aimed at maintaining a positive atmosphere where the corrective feedback should be perceived as a help, not a rebuke, by the students.

Feedback in Computer-assisted Pronunciation Training

This section discusses the implications that the above findings in the teacher and student interviews and the classroom observations may have for CAPT. It also relates the students' opinions about practising pronunciation with a computer. A summary list, indicating the points that we consider are directly applicable to CAPT is then presented. It should be noted that all strategies used by human teachers would not necessarily be successful in CAPT, just as some features that are clearly useful in CAPT would be of little use in the language classroom. The discussion and list below should be applicable in most respects to feedback for CAPT in general, but we in particular focus on CAPT using a virtual teacher.

Neri *et al.* (2002) conclude that effective feedback in CAPT 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. Our interviews and observations suggest that some other criteria are also relevant.

When should feedback be given?

One finding from the present study was that very little feedback was actually given in the classrooms. This may be taken as evidence for the need of CAPT, as the students do not get much corrective feedback in class. However, it also indicates that the amount of feedback should be limited (as already suggested by Neri *et al.*, 2002), to maintain a positive atmosphere and a communicative flow. In exercises on specific features, feedback should come after each student utterance. If the student's utterance is displayed in real time (such as when the student's pronunciation of a phoneme is plotted relative to a target in an acoustic space), the plot should be presented as an exploratory tool that the student may use to find out how the distance from the target varies with variations made in the articulation, and not as corrective feedback. The students should themselves request that formal judgement is made and feedback given (for example, by pressing a button) when they are satisfied with their attempt.

In communicative training, explicit feedback on pronunciation details should only be given after the dialogue has ended. During the dialogue, feedback should be on the

communication level, i.e. clarification requests when the intelligibility is low, or recasts of recurring student errors.

Which errors should be corrected?

The students proposed that the corrective feedback should be provided only on the feature that is practised. For CAPT, this means that other errors should not be corrected, but suggestions on which other suitable exercise sets should be provided depending on these other errors.

The students also argued that a key issue is that the feedback from the CAPT software is correct and adequate—‘correct’ in the sense that the system must not fail a student pronunciation that would be accepted by a human listener and ‘adequate’ in the sense that the feedback should correspond to the error that the student has made. If the student has made several errors, feedback should be given on the most important.

The ‘correct’ criterion has implications for which errors are appropriate for feedback. As the speech technology components of a CAPT system cannot be assumed to be error free, a confidence measure is required in order to judge whether or not a possible error should be indicated to the student. The formulation of the feedback should then be adapted to the certainty of the error identification, so that the system will emphasize the feedback more for an error identified with high confidence than for one with a low.

The ‘adequate’ criterion signifies that the software should be able to identify different types of errors, rank them according to importance, and provide feedback that is relevant to help correcting the most important error. This decision may be partly statistical, by counting the frequency of a certain type of error, or based on the score from the ASR for intelligibility, but a knowledge-based approach is also required to relate to social impact and generality. The feedback presented to the student should be limited to the 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.

What kind of feedback should be used?

Human language teachers use a vastly more varied repertoire of feedback than currently given in CAPT. To some extent, different teachers prefer different feedback techniques, but the choice of feedback mainly depends on what is suitable for the exercise, the error or the student. More important than using ‘the’ most effective type of feedback is hence to vary the feedback more and introduce adaptability to the situation and the user.

The student interviews indicated that not all feedback must be explicit, since the students believe that implicit feedback is enough and the classroom observations mostly supported this opinion. The software should instead systematically test whether or not implicit feedback is enough and augment it if not.

In order for the exercises to run smoothly, the feedback given should be minimal as default and detailed feedback and explanations should be given upon request from the user. Ideally, the user should be able to choose from a set of different feedback presentations (e.g. an enhanced acoustic target; explicit identification of the error; articulatory animations or metalinguistic explanations) to suit different learning styles.

The classroom observations indicated the importance of multimodal feedback, especially when the L2 was so phonetically different that acoustic targets did not suffice. CAPT should hence take advantage of the possibility that computers have to create multimedia feedback, combining for example acoustic recasts, spoken and written instructions, and computer animations of the articulation.

How is student motivation and self-confidence maintained?

On the one hand, a CAPT system should give feedback that is relevant and detailed enough to allow the users to understand how to change a deviant pronunciation, but on the other hand the details must be given in such a way that the practise is stimulating and that the students become proud of the progress made and increasingly confident in their own production. Due to the varied needs and wishes of different students concerning the amount of feedback, a good option might be to allow users to set the desired amount of feedback individually. The threshold should be adapted to the L2, the exercise and the student performance, but in addition it should be possible for the student to adjust the amount of feedback, for two reasons. The first is the affective factor, 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 puts the responsibility and initiative with the student, who can decide how much advice he or she requires from the tutor, in conjunction with the student-performer/teacher-coach paradigm proposed by Morley (1991). This promotes self-monitoring and the feeling of owning the training situation, which should be beneficial for motivation.

Another strategy, observed in the classroom, is 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 is to avoid getting stuck on a problematic pronunciation, but at the same time not having to accept the last student utterance as correct. Depending on the history of the training session, the classification into the new category could result in either encouraging feedback or an acknowledgement of the difficulty before moving on to the next exercise. The encouragement would be used when an improvement from previous attempts has been made, even if there is still a mispronunciation. The acknowledgement of the difficulty would be used if the same problem remains, but it does not seem fruitful to continue with corrective feedback at the moment.

The differences between CAPT and classroom settings, such as different types of exercises, the lack of potential embarrassment and the functionality to give layered

feedback, make it possible to give more feedback in CAPT. It is nevertheless very important that the CAPT program also aims at improving the student's self-confidence by balancing corrective with positive feedback and by promoting communication.

The students indicated that scoring was a compelling motivational feature in CAPT, but that the score should be a verbal statement concerning the intelligibility and fluency, rather than numerical. Another wish was for information on the accomplishment on a meta-level, regarding how many exercises and which feature sets that have been completed successfully. It was also suggested that a practise set-up that was similar to existing karaoke games, with acoustical and visual targets and immediate visual feedback, would make repetitive practise more stimulating. In the pronunciation training, the student would try to repeat a text presented on the screen. Mispronounced words would turn red and each of them would then be practised until corrected. As a support, computer animations could show how the pronunciation should be made.

Conclusions: Implications for CAPT feedback

Based on the interviews and observations we conclude that at least the following strategies are useful for CAPT:

When should feedback be given?

- Provide feedback when the student asks for it and has finished speaking.

Which errors should be corrected?

- Limit the feedback to the features that are targeted by the exercise, if there is such a focus.
- Let the decision on whether an error should be corrected or not be influenced by a user-set choice concerning how strict the system should be.
- Provide feedback on errors that have a social impact.
- Use statistical measures and a knowledge-based approach to identify the most important errors.

What kind of feedback should be given?

- Use various types of feedback (i.e. recasts and repetitions, elicitations, articulatory instructions, comparison to familiar phonemes, and metalinguistic explanations) both to accommodate different learning styles and to increase variation.
- Adapt the feedback to the exercise and the error.
- Identify the error and provide suggestions for improvement.
- Start with less detailed feedback and increase it for repeated errors.
- Involve the learner actively in pronunciation monitoring and correction.

- Use a computer screen animation of the articulation to help the student achieve unfamiliar articulatory features.
- Adapt the feedback to the degree of certainty of the identification of the error.

How is student motivation and self-confidence maintained?

- Let the user indicate the amount of feedback wanted.
- Limit the amount of feedback and let the user request more feedback by using control buttons.
- Introduce a judgment category *Satisfactory for the time being* in order not to get stuck on one exercise.
- Formulate the feedback as relating to general pronunciation difficulties rather than the user's personal shortcomings.
- Provide the user with information on the proportion of the lesson that has been fulfilled.
- Mix easier pronunciations, or those that the student already masters, with the new and difficult, to keep student motivation high.
- Balance corrective feedback and encouragements.
- Provide a verbal judgment about the student's proficiency, rather than a numerical score only, when a set of exercises has been completed.
- Take the phonological difference between the L1 and L2 into account in order to give feedback on a level that promotes student production.

In order to test the above features of feedback, we have integrated them in our implementation of a pronunciation coach, described in the next section.

Feedback from the Virtual Teacher, Artur

Neri *et al.* (2002) suggest that “detailed study of articulatory movements could be catered for [in CAPT software] by means of 3D computer animations of the lips and oral cavity”. Such animated instructions on how to achieve a target sound are technically quite straightforward to generate and have already been provided by Massaro and Light (2003). We believe that not only should instructions be given with such animations, but also feedback that helps the learners change their pronunciation. Instructions that relate to what the students are already doing will be more effective because they involve the users' cognitive awareness of their own pronunciation. Providing feedback is a decidedly more difficult task than generating instructions, since it requires that the student's current articulation is estimated from the acoustic signal, possibly supplemented by video images of the speaker's face (Engwall, 2005; Kjellström, Engwall, & Bälter, 2006).

Because of the potential benefits of using articulatory feedback, rather than instructions only, we are currently developing Artur, the ARticulation TUtoR (Engwall *et al.*, 2006a). This tutor should be able to give detailed instructions and feedback on articulatory positions and movements using 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) that have been adapted to measurements of real speech.

A very important issue for Artur is how feedback is presented to the user, as mainly two potential problems need to be addressed. The first is that users may be unable to relate the articulatory animation and instructions to their own tongue movements. A group of children who participated in a previous user study (Engwall *et al.*, 2006a), however, stated that they could transfer the instructions and animations to their own articulation. A second problem might be that users who interact with a computer-animated human-like character tend to anthropomorphize the character and hence expect human-like behaviour from it (Walker *et al.*, 1994; Koda & Maes, 1996). For Artur this means that users expect to receive feedback that is more similar to that of a real teacher than to that of an impersonal computer interface. We have therefore implemented the feedback strategies listed at the end of the previous section using a decision tree for the management of feedback (Engwall, Bälter, Öster, & Kjellström, 2006b).

User Test of the Feedback in Artur

An English-speaking version of the virtual tutor was tested by participants of the International Conference on Human Factors in Computing Systems, CHI 2006. The test subjects practised their pronunciation of the palato-velar fricative [ʃ] in a Swedish

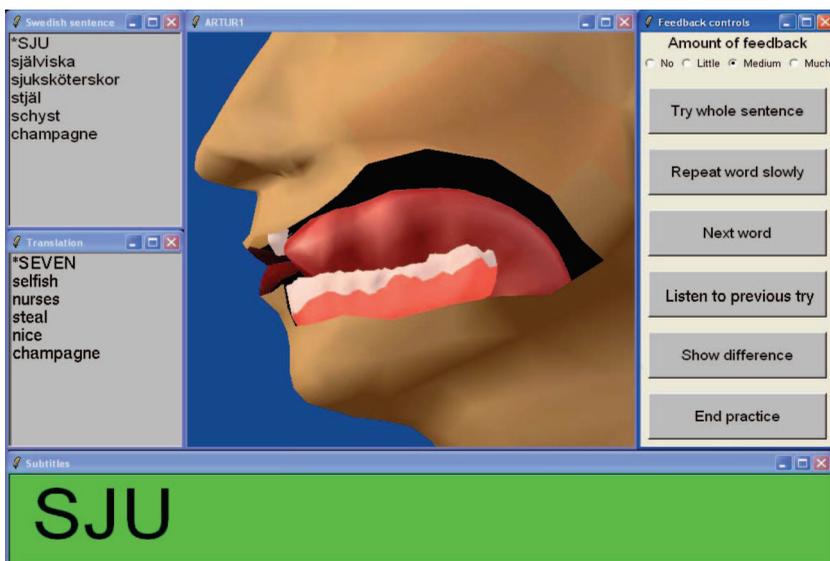


Figure 1. The user interface of the virtual pronunciation tutor Artur. Progress through the sentence is shown in the windows, computer animations in the centre, subtitles and instructions in the bottom window. The functionality of the user interaction buttons in the right panel is described in the text

sentence. A Wizard-of-Oz set-up was used, where a human, phonetically trained judge detected mispronunciations and diagnosed the cause. The wizard then decided what feedback the virtual teacher gave the user from a set of pre-generated audiovisual instructions on how to improve the articulation.

The graphical user interface (c.f. Figure 1) had one main window showing a front view of Artur speaking or a side view with articulatory animations, one transcription window with instructions, frames indicating the progress within the exercise sentence, and one interaction panel. The interaction controls made it possible for the users to choose the amount of feedback, to request an attempt at the whole training sentence, to hear and see the pronunciation of the word repeated at slow rate, to jump to the next word even if the pronunciation had not been accepted, to listen to their own previous try, to see the difference between the target and their own pronunciation illustrated in an articulatory animation (“Show difference”) and to end the practise. Artur did not explain the use of each button in the introduction, but encouraged the user to try them. Since the button labels were explicit in terms of the use of each button, we assume that the users knew already at the onset what most of the buttons could be used for. As the action behind the “Show difference” button was less evident, Artur explained its use as part of a feedback utterance some time into the training.

The practise sentence was “Sju själviska sjuksköterskor stjal schyst champagne” (“Seven selfish nurses steal nice champagne”) where each word starts with [ʃ]. This phoneme was chosen since it is unique to Swedish and it would hence be unfamiliar to most users. 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. The difficulty of the sentence was intentional, to increase the number of feedback turns needed and to provoke users to make use of the interaction buttons.

The training session lasted approximately five minutes, which is too short to study longitudinal use of any kind, but the actual testing of the feedback in a virtual tutor by experts on human–computer interaction may nevertheless indicate which types of feedback can be transferred from human to virtual tutors. Two slightly different practise openings were used. In both, Artur gave some instructions, explained the goal of the exercise and then produced the sentence. In the first opening, the user was given the opportunity to try the whole sentence directly, and when the attempt was unsuccessful, the practise continued word by word. In the second, the practise began with the separate words directly, prompting the user to try the whole sentence only at the end of the practise. 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 practise by letting them try the difficulty first, or if it was merely discouraging.

In order to evaluate the feedback given by Artur, the 37 users were asked to fill in a questionnaire, shown in Table 1, after completing the training session. In addition, the presenters also discussed the training with many of the users, which provided

Table 1. 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 practising with Artur?	1 = confusing 1 = boring	9 = clear (A) 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

background knowledge that was sometimes necessary to interpret the comments in the questionnaire.

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. It was apparent to the users that the persons presenting the system were researchers involved in the project, rather than objective evaluators. Most users further knew 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 who was visible in the set-up.

The questionnaire does not explicitly address the questions on feedback discussed with the teachers and students, since these issues were considered too complex for a questionnaire with the current setting of the test. Questions 2 and 7 are more general and not directly related to the feedback. The answers and comments to questions 1, 3–6 are, on the other hand, relevant, since the feedback that the users received was based on the outcomes of the interviews and classroom observations.

The mean opinion scores in Figure 2 indicate that the users were quite positive about the practise. The comments stated that the practise was intuitive, easy to follow and interesting. The articulatory instructions and animations were perceived as quite clear and the users judged the usefulness of the system as high. Many users commented (15 written) that the method was effective.

The question of *when* feedback should be given was not addressed by this user test. Since it was a teacher-initiated repetition exercise, feedback was provided after each student utterance.

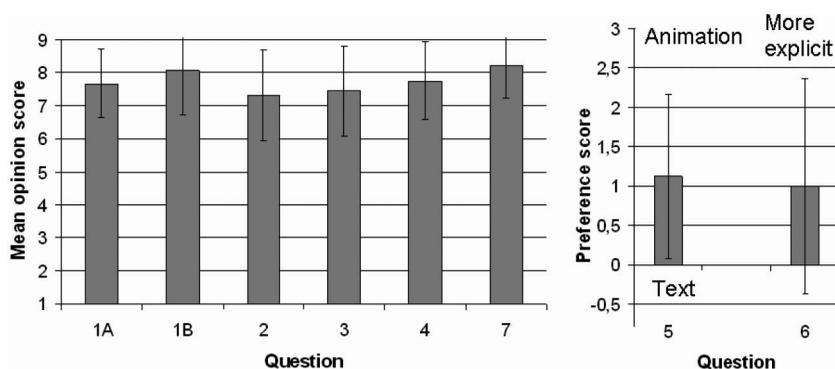


Figure 2. The mean opinion score on questions 1–4 and 7 and the preference score on question 5 and 6. The score on questions 5–6 were transformed so that 0 indicates no preference. Bars indicate one standard deviation from the mean

Which errors should be corrected?

Artur stated in the beginning of the exercise that the task was to practise on the “sj” sound and showed the articulatory difference between the English [ʃ] and the Swedish [ɧ]. Feedback was given on this phoneme only. The user was assumed to make primarily two types of error; replacing the Swedish phoneme with either the English [ʃ] or the German [x], and detailed corrective feedback was given for these two.

The users did concentrate on the [ɧ], rather than the vowels during the practise. The comments about future development suggested exercises for other sounds, rather than feedback on additional sounds, suggesting that the scheme of focusing on one target phoneme at a time appealed to the users.

The amount and level of feedback was perceived as adequate by most users (indicated by their preference score on question 6 and by eight written comments), but three explicitly commented that they would have preferred more corrections, and there was a preponderance for more feedback in the preference score. Despite this, few users tried to adjust the amount of feedback by changing the radio button from the default ‘Medium’.

What kind of feedback should be given?

Artur’s feedback repertoire included:

- (1) Articulatory instructions in text and animation.
- (2) A recast of the target word for each attempt.
- (3) A recast plus a repetition, if requested by the user with the “Show difference” button.
- (4) Comparisons to an English word.
- (5) Sensory feedback on an articulatory feature (e.g. the touch between tongue tip and palate or the air stream).

- (6) Elicitations to make the user try to self-repair, if it was judged possible.
- (7) Metalinguistic explanations on the spelling.

Several users stated that the explanations and animations on how to place the tongue (eight comments) and the comparisons to English words were very useful (four comments). There was a slight preference for the animations over the text instructions, but many users indicated that the combination of the two was important (12 comments), as the animations were better for some words and the instructions for others or because they complemented each other. Four users commented that they did find it difficult to relate the animations to their own pronunciation, but four others wrote that they actually became more aware of their own tongue by looking at the animations. Of the users who preferred the instructions, two commented that it was because the animations went too fast. Two users on the other hand stated that the “Repeat slowly” animation was too slow, which indicates the need for a user-controlled animation speed, as suggested by two users.

Three users indicated that it would be beneficial to have different views of the tongue. Since the 3D model may be rotated freely, it would indeed be possible for the users to adapt the viewpoint to their personal preference. Two other users requested the possibility to hear the target word again at normal speed. Such a button was presented in a previous user test (Engwall *et al.*, 2006a), 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 user suggestions indicate that an important feature to maximize usability is that the interface should be adaptable, by the user or a teacher, as already suggested by user interviews in Eriksson, Bälter, Engwall, Öster, and Kjellström (2005).

How is student motivation and self-confidence maintained?

Features related to promoting the user’s motivation included information on the progress within the practise sentence, the potential for the user to adjust the amount of feedback and the variation of the feedback response. The response depended on both the user’s current attempt and the previous ones. *Positive* feedback was given for a correct pronunciation. *Corrective* feedback was given the first time a word was mispronounced, and an *augmented* version with more or other details if the error was repeated on the same word. In both cases, the feedback was formulated as impersonal instructions, rather than explicit statements about the error made by the user. As in the classroom, the pronunciation could also be judged as *satisfactory*, if it was not entirely correct, but it seemed pedagogically sounder to accept it and move ahead. Two other responses were also used to avoid too many explicit corrections. The *vague* feedback encouraged the user to try to self-monitor a specific articulatory feature and *encouragements* were used as an elicitation for self-correction. The last two categories of teacher reactions could also be used when the error did not correspond to the predefined types, to avoid giving inaccurate detailed feedback. Examples of the different feedback responses are given in Table 2. The feedback utterances were different for each word, in order to avoid repetitions.

Table 2. Examples of different feedback responses. P = positive, C = corrective for English type errors, CG = corrective for German type errors, A = augmented for a repeated error, S = satisfactory, V = vague and E = encouragement. Example utterances marked with (*) were accompanied by an articulatory animation illustrating the most important articulatory feature of the feedback. For the others utterances a front face view of Artur was shown

Type	Example utterance
P	Yes, correct!
C	Think of how you make the word 'cue' and lower the tongue just a little bit in the back. (*)
CG	Try to make the sound in the back of the mouth rather than in the throat. (*)
A	Pull the tongue back as far as you can and then make the constriction a little bit further front. (*)
S	I think we could try the next word now.
V	Think of which part of the tongue touches the palate. (*)
E	Good, but try it once more.

Several users spontaneously commented (four in writing) that the practise was fun or engaging, hence indicating that they judged that Artur possessed one of the most important qualities for successful CAPT, namely that it should stimulate the user to spend time on the task. One may argue that the real challenge is to make CAPT interesting for longer periods of time than the short practise that was tested, but many more of the users gave other comments in writing or personally that indicate that they believed that the software may have potential for more continuous practise.

An unpaired *t*-test showed no statistical difference in scores 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, it was, however, clear that some users felt intimidated by the task when first presented with an entire, and difficult, sentence (four comments). In line with the recommendation that practise should include easier tasks (which was not followed in this implementation in order to maximize the amount of feedback that had to be given during the short session), two users suggested that it would be good to start with a less difficult sound.

One motivational issue that we intended to test was the combination of giving a limited amount of feedback and letting the user request more with interaction buttons. In order to promote the use of the interaction buttons, Artur sometimes suggested that the user should try one. In general, the subjects who received such advice used the interaction buttons more, not only directly following the suggestion, but later in the training as well. This indicates that, at least in the short term, the user seldom takes the initiative to go outside the teacher-driven practise and that the tutor can influence the amount and type of user interaction.

Discussion and Conclusions

Our interviews with teachers and students showed that there is a general consensus on the criteria for how feedback on pronunciation errors should be given. In general,

our interviews and observations are well in line with previous studies reported in the literature. We used these results to provide a list of criteria that we believe are applicable to CAPT (c.f. the section Feedback in computer-assisted pronunciation training) and we implemented them and made a short-term user test in the virtual pronunciation teacher Artur.

The responses from the user test indicated that the subjects, as opposed to the teachers in the interviews, perceived articulatory instructions as useful. This may partly be due to the difference in setting, since the individual tutor can use accompanying visualisations of the articulations. CAPT software may use other functionality for feedback than is applicable in the language classroom, thanks to the augmented reality provided by the computer animations.

Another reason for the difference between the teachers' view and the user test may be due to the difference in learner expertise. While the teachers argued that their students are unaccustomed to phonetic instructions, the subjects in the user test with Artur were expert users in terms of human-computer interfaces and they might have been both more accustomed to computer visualisations and more demanding regarding the interaction than naïve users. Similarly, we found in our interviews that learners who had more previous experience of learning languages requested more detailed and more complex feedback when practising a new language. The amount and type of feedback that is preferred is hence not only dependent on the student's level in the language that is practised, but competencies in other language and in human-computer interaction as well. This has two implications. The first is that while tests with an expert group have the benefit that the subjects can be expected to provide well-founded suggestions on the interaction with the computer, additional tests with naïve users will be needed, since they may have slightly different needs. The second is that the user should be able to influence not only the amount, but also the type of feedback. Since it is an established fact that students are not always the best judges as to what kind of feedback is the most effective for uptake, the decision should however not rely solely on the user's preference.

One of our recommendations for CAPT is to limit the feedback in order to promote student motivation, and give more feedback if requested by the user. Varying the feedback, combining explicit instructions of different types with vaguer suggestions and encouragements, was indeed perceived as positive by the users. Some, however, indicated that they thought Artur had been too forgiving, illustrating two issues.

The first is that learners are ready to accept more corrective feedback in the one-to-one lesson with the virtual tutor than is normal in a classroom setting. The strategy of classifying pronunciations as *Satisfactory for the time being* should still be valid, but additional corrective turns may be provided in CAPT compared to classroom lessons. It is nevertheless important to balance the articulatory instructions with exercises and feedback on the communicative level, or else negative self-monitoring risks leading to a situation where the functional communicability is impeded.

The second is that the CAPT software needs to steer the interaction more at the onset of the practise, before the student is ready to take a more active role. Few users

tried to adjust the amount of feedback in the menu and the users tended to use interaction buttons almost exclusively after suggestions from the teacher. Those who did receive such a suggestion did however continue to use the buttons afterward. It is hence important that the most important information is provided without user request and that students are made aware of the additional feedback that may be received by request.

In conclusion, since human teacher–learner interaction is vastly more effective than current CAPT pedagogy, it is our conviction that pronunciation training software may be improved by studying how feedback is distributed in the language classroom. Particular focus should be paid to how classroom feedback may be transferred and adapted to individualised pronunciation training with computers. Virtual tutors have additional advantages over impersonal software interfaces, since they have the potential of being more engaging, more communication-oriented and more effective in combining various type of feedback. The introduction of virtual teachers in CAPT may hence improve the practise, even when the performance of the other speech technology components remains the same. Since computer-animated characters tend to be personalised by the users who interact with them, it is important to make the virtual teachers behave in a manner that establishes a constructive pedagogical relationship between the tutor and the learner. We believe that this is achieved by enabling the tutor with strategies that are based on the best practices in traditional teaching and we see the guidelines provided in this study as a contribution to that effort.

Acknowledgements

This work is carried out within the ARTUR project, funded by the Swedish Research Council. The Wizard-of-Oz study was supported by an award to the first author from the Christian Benoît Association. Pronunciation Feedback from Real and Virtual Language Teachers.

The authors would like to thank the participating teachers, students and test subjects. We would also like to thank the anonymous reviewer who provided very valuable comments on how to improve this article.

References

- Adams, F. R., Crepy, H., Jameson, D., & Thatcher, J. W. (1990). IBM products for persons with disabilities. *IEEE GlobeCom'89*, 496–502.
- 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 *Proceedings of Interspeech*, 165–168.
- Beskow, J. (2003). *Talking heads—Models and applications for multimodal speech synthesis*. Ph.D. thesis, KTH, Stockholm, Sweden.
- CALICO (2004–05). The CALICO review. Retrieved November 4, 2005 from [http://calico.org/CALICO Review/](http://calico.org/CALICO%20Review/)

- 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.
- Deroo, O., Ris, C., Gielen, S., & Vanparis, J. (2000). Automatic detection of mispronounced phonemes for language learning tools. In Proceedings of the 6th International Conference 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*. Retrieved October 20, 2006 from <http://www.tc.columbia.edu/academic/a&hdept/tesol/Webjournal/El-Tatawy.pdf>
- Engwall, O. (2003). Combining MRI, EMA & EPG in a three-dimensional tongue model. *Speech Communication*, 41(2–3), 303–329.
- Engwall, O. (2005). Introducing visual cues in acoustic-to-articulatory inversion. *Proceedings of Interspeech 2005*, 3205–3208.
- Engwall, O., Bälter, O., Öster, A.-M., & Kjellström, H. (2006a). Designing the user interface of the computer-based speech training system ARTUR based on early user tests. *Journal of Behavioural and Information Technology*, 25(4), 353–365.
- Engwall, O., Bälter, O., Öster, A.-M., & Kjellström, H. (2006b). Feedback management in the pronunciation training system ARTUR. Proceedings of the International Conference on Human Factors in Computing Systems, CHI, 231–234.
- Eriksson, E., Bälter, O., Engwall, O., Öster, A.-M., & Kjellström, H. (2005). Design recommendations for a computer-based speech training system based on end-user interviews. Proceedings of the 10th International Conference on Speech and Computers, 483–486.
- Hendrickson, J. M. (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 Proceedings of the 7th International Conference on Spoken Language Processing, 733–736.
- Kjellström, H., Engwall, O., & Bälter, O. (2006). Reconstructing tongue movements from audio and video. *Proceedings of Interspeech 2006*, 2238–2241.
- Koda, T., & Maes, P. (1996). Agents with faces: The effects of personification of agents. *Proceedings of Human-Computer Interaction*, 239–245.
- Krashen, S. ed. (1981). *Principles and practise in second language acquisition*. English Language Teaching Series. London: Prentice-Hall.
- Lester, J., Converse, S., Kahler, S., Barlow, S., Stone, B., & Bhogal, R. (1997). The persona effect: Affective impact of animated pedagogical agents. *Proceedings on Human Factors in Computing Systems*, 359–366.
- Lightbown, P. & Spada, N. (Eds.) (1999). *How languages are learned*. Oxford: Oxford 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*, 19, 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.
- Massaro, D., & Light, J. (2003). Read my tongue movements: Bimodal learning to perceive and produce non-native speech /r/ and /l/. In Proceedings of the 8th Eurospeech, 2249–2252.
- Menzel, W., Herron, D., Bonaventura, P., & Morton, R. (2000). Automatic detection and correction of nonnative English pronunciation. In Proceedings of Workshop Integrating 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., & Boves, L. (2002). The pedagogy-technology interface in computer-assisted pronunciation training. *Computer Assisted Language Learning*, 15, 441–467.
- Nicholas, H., Lightbown, P. M., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, 51, 719–758.
- Rubin, J. (Ed.) (1994). *Handbook of usability testing*. New York: John Wiley & Sons.
- Schachter, J. (1991). Corrective feedback in historical perspective. *Second Language Research*, 7, 89–102.
- Sheen, Y. H. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research*, 8, 263–300.
- 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.
- van Mulken, S., André, E., & Müller, J. (1998). The persona effect: How substantial is it? *Proceedings of HCI'98*, 53–66.
- Vicsi, K., Roach, P., Öster, A., Kacic, Z., Barczikay, A., Tanta, A., Csatári, F., Bakcsi, Z., & Sfakianaki, A. (2000). A multilingual teaching and training system for children with speech disorders. *International Journal of Speech Technology*, 3, 289–300.
- Walker, J., Sproull, L., & Subramani, R. (1994). Using a human face in an interface. Proceedings of the CHI '94 Conference Companion on Human Factors in Computing Systems, pp. 85–91.