FEEDBACK IN COMPUTER ASSISTED PRONUNCIATION TRAINING: TECHNOLOGY PUSH OR DEMAND PULL?

Ambra Neri, Catia Cucchiarini, Helmer Strik

A^{2}RT, Dept. of Language and Speech,
University of Nijmegen, The Netherlands
{A.Neri, C.Cucchiarini, H.Strik}@let.kun.nl

ABSTRACT

In this paper, we examine the type of feedback that currently available Computer Assisted Pronunciation Training (CAPT) systems provide, with a view to establishing whether this meets pedagogically sound requirements. We show that many commercial systems tend to prefer technological novelties that do not always comply with pedagogical criteria and that despite the limitations of today's technology, it is possible to design CAPT systems that are more in line with learners' needs.

1. INTRODUCTION

The advantages that Computer Assisted Language Learning can offer are nowadays well known to educators struggling with traditional language classroom constraints. CAPT can be particularly beneficial for second language (L2) learning: not only does it provide a private, stress-free environment with virtually unlimited access to input and self-paced practice, it can also provide individualized, instantaneous feedback. It is not surprising, then, that a wealth of CAPT systems have been developed, most of which are already available on the market.

An alert purchaser, however, might find the display of products disappointing. Many authors describe available programs as fancy-looking systems that may at first impress student and teacher alike, but that eventually fail to meet sound pedagogical requirements [1,2]. These systems, which do not fully exploit the potentialities of CAPT, seem to be the result of a technology push, rather than of a demand pull. This may be due to a failure to adopt a multidisciplinary approach involving speech technologists, linguists and language educators [3], or more fundamentally, to the absence of clear pedagogical guidelines that suit these types of environments. This problem is especially serious with respect to feedback, a crucial factor in learning pronunciation, for which little research is available.

What are, then, the guidelines that should be considered when designing feedback for pedagogically sound CAPT? In spite of the scarcity of studies on this issue, we believe that research on second language acquisition (SLA) and teaching can already provide us with some indications. However, incorporating this knowledge within state-of-the-art technology may not be as straightforward as language educators hope. Current ASR technology well illustrates this problem, because it still suffers from several limitations that pose constraints on the design of CAPT, as is exemplified by the occasional provision of erroneous feedback.

In this paper, we first analyse literature on feedback in conventional pronunciation training to identify the basic pedagogical criteria for CAPT systems. Then we give a critical overview of various products. Finally, we sketch some recommendations for developing automatic feedback in pedagogically sound CAPT systems that employ reliable state-of-the-art technology.

2. THE NOTION OF FEEDBACK

The exact notion of feedback in SLA is far from clear. 'Feedback' is used as an umbrella term to refer to different types of information on the learner's performance on a given task in the L2. The term commonly refers to *external* and *explicit* information that is provided by a teacher, a peer or a native speaker on the learner's production in the L2. For instance, feedback can be provided on correct production, to encourage the student, e.g. by saying 'You are on the right track'. More often, the term is used to refer to *corrective* information on an ill-formed utterance. However, teachers and interlocutors can also resort to a more implicit type of feedback that does not contain metalinguistic information and is unobtrusive for the task at hand, as in the case of expansions during a conversational exchange. In CALL, the term is sometimes found as a synonym for help-hints that the user can retrieve to correctly complete a task [4].

In this paper, we use the term 'feedback' to refer to external information on the student's pronunciation. In the following section we will examine different types of feedback in an attempt to establish which forms are more effective and feasible in CAPT.

3. EFFECTIVE FEEDBACK ON L2 PRONUNCIATION

The issue of feedback in SLA is still controversial. Although L2 research has identified several types of feedback, the difficulty in homogenously labelling and operationalizing them makes it extremely hard to draw general conclusions, for instance on their effectiveness for learning. Some authors believe that the provision of metalinguistic knowledge through feedback might hinder the unconscious, natural process of acquiring an L2 and reduce it to mere learning - never resulting in automated knowledge [5]. Yet, others believe that adult L2 learning require some degree of awareness [6]. When it comes to learning L2 pronunciation, feedback appears to be crucial because the L1 influence can be so overwhelming that the learner is not able to notice the discrepancies between the sounds (s)he produces and the correct target sounds [7]. For these errors, feedback should be provided "that does not rely on the student's own perceptions" [8: p.51]. It is obvious that it is only once an awareness of the problems has been raised in the learner, that the individual can take remedial steps.

In spite of the crucial role of feedback, very little research has been carried out on its effectiveness for the acquisition of L2 pronunciation. A recent study on corrective feedback in

teacher-to-student classroom interactions indicated that 'recast' – a reformulation of the student's utterance by the teacher – was the most commonly used technique and had the highest rate of uptake for phonological aspects, while it yielded the lowest rates of uptake for grammatical and lexical aspects [9]. These results may suggest that a simple reformulation of the mispronounced utterance immediately following the error might be sufficient to successfully correct it, contrary to research on feedback on grammatical errors, which should stimulate self-repairs through higher-order cognitive processes in the learner [10, 11, 12]. Moreover, the results seem indirectly corroborated by studies indicating that this type of feedback can be effective for learners that are already relatively proficient, and for learning aspects that involve lower-order mental processes [10, 13]. However, it should be noted that the majority of these studies only investigated the *short-term* effects of corrective feedback.

What research seems to indicate consistently is that feedback should allow verification of response correctness (e.g. by telling the student whether output was good or bad), but also pinpoint specific errors and possibly suggest a remedy [12, 14, 15]. In other words, besides receiving a score, the student should comprehend why (s)he got that score.

It goes without saying that teachers do not need to provide feedback on each of the student's mistakes: such a course of action might be discouraging for the student and extremely lengthy for the teachers. The pronunciation errors to be addressed could be selected on the basis of different criteria, such as the ultimate aim of the training - be it accent-free pronunciation or intelligible pronunciation - the specific L1-L2 combination, the degree of hindrance to comprehensibility and the degree of persistence of the various errors, etc.

Various studies have addressed the issue of pronunciation error gravity hierarchies, to establish priorities in pronunciation training [16, 17]. Despite some apparent contradictions due to methodological limitations, it appears that both segmental and supra-segmental factors are important (see [18, 19]). .

4. FEEDBACK IN AVAILABLE CAPT SYSTEMS

In this section we examine various approaches to providing feedback in CAPT systems, in an attempt to establish which forms are more effective for learning. Some CAPT systems provide instantaneous feedback in the form of spectrograms and waveforms which are often accompanied – for comparison - by previously stored displays of a model utterance pronounced by the teacher or by a native speaker. These systems make use of tools that perform acoustic analyses of amplitude, pitch, duration and spectrum of the students' speech [20,21,22,23,24]. The effectiveness of these systems, however, is questionable for a number of reasons.

First of all, most of these systems perform an analysis of the incoming speech signal without first 'recognizing' the utterance. This implies that there is no guarantee that the student's utterance does indeed correspond to the intended one. Second, simultaneously displaying the incoming utterance and the model utterance wrongly suggests that the student should ultimately aim at producing an utterance whose acoustic representation closely corresponds to that of the model utterance. In fact, this is not necessary at all: two utterances with the same content may both be very well pronounced and still have waveforms or spectrograms that are very different from each other. Moreover, these kinds of displays are not easily interpretable for students. Actually, they are representations of raw data that require the presence of a teacher to interpret them. Another option might be to train the students to autonomously read these displays. However, even students who have received specific training are likely to have a hard time deciphering these data and extracting the information needed to improve pronunciation, as there is no simple correspondence between the articulatory gesture and the acoustic structure in the properties displayed. In other words, as many authors lament, this type of feedback is not in line with the requirement that feedback should first of all be comprehensible [8,25,26,27]. As a consequence, students are likely to make random attempts at correcting the presumed errors - which, instead of improving pronunciation, may have the effect of reinforcing poor pronunciation and eventually result in fossilization [25].

Pro-nunciation [28] is a prototypical system that provides 3D animated mouth representations of phonemes, limericks, tongue twisters and the possibility to display waveforms of the students' utterance for comparison with the model one. The criticism that we expressed above is all the more appropriate in the case of waveforms, since these are even more variable and less informative than spectrograms. In other systems, like the *Talk to Me* and the *Tell me More* series [29], the graphical importance the waveforms have on the screen suggests that, even if other forms of feedback are provided, waveforms are presented because of their flashy look, to impress the users.

A much praised system, *WinPitchLTL* [24], has been developed by two phoneticians as an authoring tool for different learning environments. This system can display a signal's pitch curve, intensity curve, waveform, and spectrogram. It also features 'word-processing' and editing facilities that allow the teacher to add text and highlight relevant segments or cues, thereby making important information easily visible and retraceable for the student. Moreover, through a synthesis feature, the prosodic parameters of a student's utterance can be modified so that the correct version can be played back with the student's own voice. However, the effectiveness of this system totally relies on the presence of a teacher who received training in phonetics and acoustics and is able to pass on that information to the students, while this, of course, is not the common rule [3].

Sometimes graphic displays of pitch contours are used to give feedback on intonational patterns (see [14]). Although training is needed to interpret these displays too, matching a pitch contour rather than an oscillogram or spectrogram is intuitive and meaningful. Kommissarchik and Kommissarchik (2000) have developed a system for teaching American English prosody to non-native speakers of English, *BetterAccentTutor*, in which readily accessible feedback is provided on intonation, stress and rhythm. The students listen to a native speaker's recording studying its intonation, stress and rhythm patterns, utter a phrase and receive immediate audio-visual feedback from the system. Both the students' and the natives' patterns can be displayed on the screen for comparison, with two major visualization modes: intonation is visualized as a pitch graph, whereas syllable intensity/rhythm is visualized as steps of various length (duration) and height (energy). This program, however, does not address segmental errors.

Some programs let the *computer* compare model and student's utterances, with a view to producing a pronunciation quality score. The feedback, in this case, consists of a numerical or symbolic score – e.g. a smiley – that is automatically generated by the system. The usefulness of automatic scoring is evident as it gives the learner immediate, comprehensible evaluation on output quality, a type of feedback that is appreciated by students [30]. However, the great challenge in developing systems of this kind is to define the appropriate automatic measures the computer has to calculate, where appropriate means 1) strongly correlated with human pronunciation ratings and 2) suitable to be used as a basis for providing feedback. The importance of the relation to human ratings is obvious: in the end the students will have to talk to people and not to machines, so the quality of the pronunciation has to be determined on the basis of what people deem acceptable. The second point can best be illustrated by referring to temporal measures of speech quality: these measures appear to be strongly correlated with human ratings of pronunciation quality and fluency, and are therefore suitable for pronunciation testing [31, 32]. However, they do not constitute an appropriate basis for providing feedback on pronunciation: telling students to speak faster is unlikely to lead to an improvement in the quality of their pronunciation. *FreshTalk* exemplifies the sort of system in which measures of non-nativeness such as temporal speech properties are used as a basis for feedback, and indeed, the feedback related to speech rate did not prove to be effective [33]. Given the limited usefulness of such scores, programs should integrate this type of feedback with more meaningful and detailed information on the student's oral performance.

Other CAPT systems provide a similar, albeit more implicit and more realistic type of feedback. Auralog's courseware [29] allows the students to train communicative skills through interactive dialogues with the computer. The student reacts to an oral question by choosing and producing one of three written responses that are phonetically different. Through ASR, the computer recognizes the student's utterance and moves on to an appropriate conversational exchange. In this way, the program ensures a certain degree of realism. A similar method is being used by U.S. Army researchers and by the developers of *TraciTalk* to develop game-like programs to teach L2s [8], [34]. In this case, the student orally asks the computer to perform a task in a simulated 'microworld', such as 'put the book on the table'. If the computer understands the utterance, it will perform the action required by the student. This type of feedback is undoubtedly very effective to reinforce correct pronunciation behaviour, as it simulates the type of interaction that would take place with a human interlocutor and it exploits the advantages that involvement in games has for learning [35]. However, both these programs are unable to offer any help if a student cannot make him/herself intelligible because, for instance, he or she cannot correctly pronounce a certain sound.

A serious attempt at diagnosing segmental errors has been made in the EU-funded ISLE project [26, 36]. This system targets German and Italian learners of English, and aims at providing feedback, focusing in particular on word level errors, for which it checks mispronunciations of specific sounds and lexical-stress errors. The knowledge-based character of this system allows for good recognition performance by the ASR, which is trained to recognize typical, predictable errors due to interference from specific L1s. However, this approach can only be adopted when the L1 background of the user is known, and when knowledge on typical errors for specific L1-L2 pairs is available. It follows that such a system is not able to handle unexpected, idiosyncratic errors that may be frequently made by some learners and that may be detrimental to intelligibility. The ISLE system provides feedback by highlighting the locus of the error in the word. In addition, example words are shown and can be listened to which contain, highlighted, the correct sound to imitate and the one corresponding to the mispronounced version. While this feedback design seems satisfactory, the system yields poor performance results. The authors comment that "students will more frequently be given erroneous discouraging feedback than they will be given helpful diagnoses" [26: 54].

The generation of erroneous feedback is such a common problem for CALL pronunciation training systems, and patently wrong error detection can be so frustrating for the student, that Wachowicz and Scott [35] recommend using implicit rather than explicit, judgmental feedback. For instance, a system that does not have the ambition of telling the student to which sound his/her mispronounced version corresponded is likely to make fewer errors than the ISLE system. It will also provide less detailed, but more frequently correct information to the student. Moreover, this level of detail in feedback may just be sufficient for the student: (s)he is told that his/her pronunciation was not completely correct, (s)he receives information on which areas were incorrect and has the possibility of listening again to the model utterance, this time paying special attention to those aspects of the utterance which (s)he did not get right the first time.

5. CONCLUSIONS

Our overview of literature on L2 pronunciation feedback has revealed that effective feedback should first of all be comprehensible, should not rely solely on the learner' s own perception, should allow verification of response correctness, pinpoint specific errors and possibly suggest a remedy. In our overview of available CAPT systems, we have seen that the feedback provided often makes use of technological features that do not always comply with these requirements. In other words, the choices made in these systems seem to result from a technology push, rather than from a demand pull. However, this need not be so. On the basis of our survey, we are convinced that new technologies hold great potential for effective feedback in CAPT. For instance, ASR can be extremely useful, even though it still has some limitations, which imply, among other things, that the student's utterance has to be predictable and that error diagnosis is only possible with a limited degree of detail. Nevertheless, ASR should be used because it allows verification of response correctness, real-time evaluation and comprehensible feedback. However, it is important that in employing these techniques, developers first of all focus on the learner's needs and accordingly select functionalities that meet those needs.

6. ACKNOWLEDGMENTS

This research was supported by the Netherlands Organization for Scientific Research (NWO). .

7. REFERENCES

[1] Murray, L., Barnes, A., Beyond the 'wow' factor – evaluating multimedia language learning software from a pedagogical point of view, *System 26*, 249-259, 1998.

[2] Pennington, M.C., Computer-aided pronunciation pedagogy: Promise, limitations, directions, *Computer Assisted Language Learning 12*, 427-440, 1999.

[3] Price, P., How can speech technology replicate and complement skills of good language teachers in ways that help people to learn language? *Proceedings of StiLL 98,* Marholmen, Sweden, 81-86, 1998.

[4] Pujolà, J.-T., Did CALL feedback feed back? Researching learners' use of feedback, *ReCALL 13*, 79-98, 2001.

[5] Krashen, S.D. *Second language Acquisition and Second Language Learning,* Oxford: Pergamon Press, 1981.

[6] Schmidt, R., The role of consciousness in second language learning, *Applied Linguistics 11*, 129-158, 1990.

[7] Flege, J.E., Second-language speech learning: Findings and problems, in Strange, W. (ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues,* Timonium, MD: York Press, 233-273, 1995.

[8] Ehsani, F., Knodt, E., Speech technology in computer-aided learning: Strengths and limitations of a new CALL paradigm, *Language Learning and Technology 2*, 45-60, 1998, http://llt.msu.edu/vol2num1/article3/index.html.

[9] Lyster, R., Negotiation of Form, Recasts, and Explicit Correction in relation to error types and learner repair in immersion classrooms, *Language Learning 48*, 183-218, 1998.

[10] Nagata, N., Intelligent computer feedback for second language instruction, *The Modern Language Journal 77*, 330-339, 1993.

[11] Lyster, R. Ranta, L., Corrective feedback and learner uptake, *Studies in Second Language Acquisition* 19, 37-66, 1997.

[12] Crompton, P., Rodrigues, S., The role and nature of feedback on students learning grammar: A small scale study on the use of feedback in CALL in language learning, *Proceedings of the CALL workshop, AIED Conference*, San Antonio, Texas, 70-82, 2001.

[13] Nicholas, H., Ligthbown, P.M., Spada, N., Recasts as feedback to language learners, *Language Learning 51*, 719-758, 2001.

[14] Chun, D.M., Signal analysis software for teaching discourse intonation*, Language Learning and Technology 2*, 61-77, 1998, http://llt.msu.edu/vol2num1/ article4/index.html.

[15] Warschauer, M., Healey, D., Computers and language learning: An overview*, Language Teaching 31*, 57-71, 1998.

[16] Anderson-Hsieh, J., Johnson, R., Koehler, K., The relationship between native speaker judgements of nonnative pronunciation and deviance in segmentals, prosody and syllable structure, *Language Learning 42*, 529-555, 1992.

[17] Derwing, T.M., Munro, M.J., Accent, intelligibility, and comprehensibility, *Studies in Second Language Acquisition 20*, 1-16, 1997.

[18] Derwing, T.M., Munro, M.J., Wiebe, G., Evidence in favour of a broad framework for pronunciation instruction, *Language Learning 48*, 393-410, 1998.

[19] Celce-Murcia, M., Brinton, D.M., Goodwin, J.M., *Teaching Pronunciation,* Cambridge: CUP, 1996.

[20] Molholt, G., Computer-assisted instruction in pronunciation for Chinese speakers of American English, *TESOL Quarterly 22*, 91-111, 1988.

[21] Akahane-Yamada, R., McDermott, E., Computer-based second language production training by using spectrographic representation and HMM-based speech recognition scores, *Proceedings of ICSLP*, Sydney, Australia, 1998.

[22] Nouza, J., Training speech through visual feedback patterns, *Proceedings of ICSLP*, Sydney, Australia, 1998.

[23] Lambacher, S. A CALL tool for improving second language acquisition of English consonants by Japanese learners, *Computer Assisted Language Learning 12,* 137-156, 1999.

[24] Germain-Rutherford, A., Martin, P., Présentation d'un logiciel de visualisation pour l'apprentissage de l'oral en langue seconde, *ALSIC 3*, 61-76, 2000, http://alsic.u-strasbg.fr/Menus/frameder.htm.

[25] Eskenazi, M., Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype, *Language Learning and Technology 2*, 62-76, 1999, http://llt.msu.edu/vol2num2/article3/index.html.

[26] Menzel, W., Herron, D., Bonaventura, P., Morton, R., Automatic detection and correction of non-native English pronunciations, *Proceedings of InSTILL 2000,* Dundee, Scotland, 49-56, 2000.

[27] Kommissarchik, J., Komissarchik, E., Better Accent Tutor –Analysis and visualization of speech prosody, *Proceedings of InSTILL 2000,*Dundee, Scotland, 86-89, 2000.

[28] Pro-nunciation, *Products*, 2002. http://users.zipworld.com.au/~pronunce/products.htm.

[29] Auralog, *Innovations for language learning*, 2002 http://www.auralog.com/.

[30] ISLE Delivarable 1.4, Pronunciation training: Requirements and solutions, 1999, http://nats-www.informatik.uni-hamburg.de/~isle/public/D14/D14.html.

[31] Cucchiarini, C., Strik, H., Boves, L., Different aspects of pronunciation quality ratings and their relation to scores produced by speech recognition algorithms, *Speech Communication 30*, 109-119, 2000.

[32] Franco, H., Neumeyer, L., Digalakis, V., Ronen, O., Combination of machine scores for automatic grading of pronunciation quality, *Speech Communication 30,* 121-130, 2000.

[33] Precoda, K., Halverson, C.A., Franco, H., Effects of speech recognition-based pronunciation feedback on second-language pronunciation ability, *Proceedings of InSTILL 2000,* Dundee, Scotland, 102-105, 2000.

[34] Holland, V.M., Kaplan, J.D., Sabol, M.A., Preliminary tests of language learning in a speech-interactive graphics microworld, *CALICO Journal 16*, 339-359, 1999.

[35] Wachowicz, K., Scott, B., Software that listens: It's not a question of whether, it's a question of how, *CALICO Journal 16*, 253-276, 1999.

[36] ISLE Deliverable 4.5, Error diagnosis for spoken language, 2001, http://nats-www.informatik.uni-hamburg.de/~isle/public/D45/D45.html.