Computer Assisted Language Learning
2002, Vol. 00, No. 0, pp. 000–000

0958-8221/02/0000–000$16.00

The Pedagogy-Technology Interface in Computer Assisted Pronunciation Training

A. Neri, C. Cucchiarini, H. Strik, and L. Boves
A^{2}RT, Department of Language and Speech, University of Nijmegen, The Netherlands

ABSTRACT

In this paper, we examine the relationship between pedagogy and technology in Computer Assisted Pronunciation Training (CAPT) courseware. First, we will analyse available literature on second language pronunciation teaching and learning in order to derive some general guidelines for effective training. Second, we will present an appraisal of various CAPT systems with a view to establishing whether they meet pedagogical requirements. In this respect, we will show that many commercial systems tend to prefer technological novelties to the detriment of pedagogical criteria that could benefit the learner more. While examining the limitations of today's technology, we will consider possible ways to deal with these shortcomings. Finally, we will combine the information thus gathered to suggest some recommendations for future CAPT.

1. INTRODUCTION

The advantages that Computer Assisted Language Learning (CALL) can offer are nowadays well-known to educators struggling with traditional language classroom constraints. Computer Assisted Pronunciation Training (CAPT), in particular, can be beneficial to second language learning as it provides a private, stress-free environment in which students can access virtually unlimited input, practise at their own pace and, through the integration of Automatic Speech Recognition (ASR), receive individualised, instantaneous feedback. It is not surprising, then, that a wealth of CAPT systems have been developed, many of which are available on the market for the language teacher or the individual learner.

When examined carefully, however, the display of products may not look entirely satisfactory. Many authors describe commercially available programs

Address correspondence to: A. Neri, A^{2}RT, Department of Language and Speech, University of Nijmegen, The Netherlands. E-mail: A.Neri@let.kun.nl or {C.Cucchiarini, H.Strik, L.Boves}@let.kun.nl

as fancy-looking systems that may at first impress student and teacher alike, but eventually fail to meet sound pedagogical requirements (Murray & Barnes, 1998; Pennington, 1999; Price, 1998; Warschauer & Healey, 1998; Watts, 1997). These systems, which do not fully exploit the potentialities of CAPT, look more like the result of a technology push, rather than of a demand pull. This may not necessarily be due to a lack of willingness, on the part of the developers, to include pedagogical guidelines in the design. It may simply be due to a failure to adopt a multidisciplinary approach involving speech technologists, linguists and language teachers (Cole et al., 1998; Price, 1998), or more fundamentally, to the absence of clear pedagogical guidelines that suit these types of environments.

What are, then, the guidelines that should be considered when developing a pedagogically sound CAPT system? We believe that research on second language acquisition and teaching can already provide us with some indications on *which* ingredients are needed for effective pronunciation training. Although much work still needs to be done, especially with respect to the issue of feedback, we feel that it is possible to suggest ways to *blend* these ingredients in order to obtain the optimal outcome. However, incorporating this knowledge within state-of-the-art technology may not be as straightforward as educators hope. Current ASR technology, for instance, 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 will first analyse available literature on traditional pronunciation training in order to identify the basic pedagogical criteria that a system should ideally meet. Second, we will provide a critical evaluation of those CAPT systems that more closely fulfill those demands, with a view to establishing which pedagogical aims can be achieved with state-of-the-art technology. In doing so, we will focus in particular on the issue of feedback. Finally, we will combine the information thus gathered in an attempt to provide some recommendations for the development of CAPT systems that employ state-of-the-art technology in order to meet pedagogical requirements.

2. SECOND LANGUAGE PRONUNCIATION TRAINING: IDEAL PEDAGOGICAL REQUIREMENTS

According to many researchers, the biggest problem in CALL is a lack of guidelines from the second language acquisition research field that could be

used to better employ the enormous progress recently made in technology, to design better courseware (Chapelle, 1997, 2001; Levy, 1997; Pennington, 1999). Of all CALL tools, this problem particularly affects CAPT systems (Pennington, 1999). Although valuable criteria have been outlined in the past few years to evaluate CALL, these are either of a general nature (as in Chapelle, 2001) or they mainly concern computer assisted learning of *vocabulary* or *grammar*, while *pronunciation* is hardly mentioned. This scarcity of indications makes it hard for CAPT practitioners to develop effective courseware. Similarly, within traditional teaching environments, conscientious teachers trying to devise optimal pronunciation training for their students are faced with many questions.

Why has research as yet been unable to offer a straightforward answer to these questions? First of all, the considerable variety in teaching contexts and learning aims makes it difficult to set hard-and-fast rules that can be applied across different learning settings. Besides, until recently, many educators were convinced that teaching pronunciation was pointless because accent-free pronunciation of the second language (L2) was considered a myth (Scovel, 1988) and because training would either have no impact or, even worse, would hinder the natural, unconscious process needed for the acquisition of pronunciation (Krashen, 1981; Krashen & Terrell, 1983). This view led to a general tendency to neglect pronunciation in favour of grammar and vocabulary in research on second language acquisition. As a result, little information is available on how pronunciation can best be taught.

Some of these beliefs have been contradicted by recent studies indicating that tailor-made training can improve a learner's pronunciation in the L2 to such a degree that – to human judges – she/he can sound indistinguishable from a native speaker (Bongaerts, 1999). Other studies have evidenced a general intolerance for strong foreign accents that might place learners in the L2 country at a professional or social disadvantage (Brennan & Brennan, 1981; Morley, 1991). Furthermore, the number of professionals who regularly communicate in a foreign language for their work has increased with globalisation. In order to ensure that these learners are able to efficiently communicate in the L2, it is imperative that language teaching methods include pronunciation training.

With respect to the ultimate goal of pronunciation training, many researchers now agree that, while eradicating the finest traces of foreign accent might only be necessary for the training of future spies (Abercrombie, 1991), a reasonably intelligible pronunciation is an essential component of

communicative competence (Abercrombie, 1991; Celce-Murcia et al., 1996; Morley, 1991; Munro & Derwing, 1995). In this respect, it is important to draw a distinction between intelligibility or comprehensibility on the one hand, and accentedness of L2 speech on the other hand. These are two different, albeit related, dimensions of non-native pronunciation. A strong foreign accent does not always hinder intelligibility of speech and specific types of instruction do not necessarily lead to improvement of both these aspects (Derwing & Munro, 1997). The importance of ''comprehensibility'' over ''correct pronunciation'' also emerged from a recent study on user requirements that was carried out within the framework of the European project *ISLE*, which aimed at developing an automatic pronunciation training system for Italian and German learners of English (ISLE 1.4, 1999). Many of the studies carried out so far have not drawn that distinction, thus obtaining blurred results that make it difficult to make comparisons and draw significant, generalizable conclusions.

A close examination of recent research can nevertheless help to identify some of the factors that affect L2 pronunciation most severely and to derive general guidelines for the teaching of pronunciation. Various studies have revealed that pronunciation learning is affected by a number of variables such as first language (L1), level of education, age on arrival (for naturalistic settings), amount of use of L1 and L2, motivation for learning L2, and so forth (see Piske et al., 2001 for an overview). These are all factors that can vary from person to person and that cannot be controlled directly by the teacher to produce the desired learning outcomes. However, there are other variables that are also known to affect pronunciation learning and that can be manipulated so as to obtain better results. These are input, output and feedback. These factors will be analysed in more detail in the following three sections.

2.1. Input

According to interactionist theories, the basic ingredient for successful language acquisition is input. Students must be able to access large quantities of input, so that target models become available. Although the majority of the studies on the impact of different types of input have addressed the acquisition of linguistic aspects other than pronunciation (see Schachter, 1998), there are reasons to believe that input can benefit pronunciation learning. As pointed out by Leather and James (1996), the initial production of new speech patterns, whether in L1 or L2, implies some phonetic representation in auditory-perceptual space that must have been previously derived from exemplars

available in the community or explicitly presented during training. Just like for the acquisition of L1 sounds, multiple-talker models seem to be particularly effective to improve perception of novel contrasts as the inherent variability allows for induction of general phonetic categories (Logan, Lively, & Pisoni, 1991). To this end, it may be important that lip movements be visible for the students, as both seeing and hearing a sound that is being articulated has been shown to improve production and perception (Jones, 1997; Massaro, 1987).

It has also been suggested that specific instruction on different pronunciation aspects can lead to improvement of those aspects (Bongaerts, 2001; Derwing, Munro, Wiebe, 1998; Flege, 1999). This may be taken as an indication that metalinguistic awareness is conducive to learning gains in pronunciation. With regard to the way input should be presented, teachers should try to contextualize input, as meaningful learning, that is, learning through associations, generally facilitates long-term retention (Ausubel, 1968). Furthermore, input that is meaningful to a learner is perceived by the learner as relevant to his/her needs, a factor that can stimulate intrinsic motivation and thus indirectly favour learning (Dörnyei, 1998; Keller, 1983). Another way to stimulate learner motivation is to present the student with engaging input that also accommodates different learning styles (Crookes & Schmidt, 1991; Oxford & Anderson, 1995). For instance, input could be presented in written, aural and audio-visual form (e.g., a radio interview and a short film episode).

2.2. Output

Although essential, mere exposure to the L2 does not appear to be a sufficient condition for pronunciation improvement, as is exemplified by long-term foreign residents who retain a strong accent and are hardly intelligible in the L2 (Morley, 1991). As a matter of fact, it is now generally accepted in second language acquisition research that, if the learners' aim is to speak the foreign language fluently and accurately, it is necessary for them to practise speaking it (Hendrik, 1997; Swain, 1985; Swain & Lapkin, 1995). By producing speech, learners can test their hypotheses on the L2 sounds. Learners can compare their own output with the input model and consequently form correct L2 representations. Through production, speakers receive a first, proprioceptive feedback on their own performance: auditory and tactile feedback is available from air- and bone-conducted pressure changes and from contact surfaces of articulators, while feedback from the joints, tendons, and muscles provides a sense of articulatory positions and movements; motor programs are

then gradually adjusted until a satisfactory match is made between feedback signals and target model (De Bot, 1983; Leather & James, 1996). Furthermore, through practice, knowledge about the L2 already internalised can become more automatic and thus enhance fluency (De Bot, 1996). Output also allows elicitation of more input and feedback from peers (Swain & Lapkin, 1995) and engages self-monitoring skills – aspects that appear to be linked to good L2 performance (O'Malley & Chamot, 1990; Rubin, 1987; Wharton 2000).

However, the activities aimed at developing the students' productive skills should be designed carefully. Contrary to past practice, they should not be limited to 'listen and repeat' drills of isolated, decontextualised sounds as in the case of minimal pairs: exclusively attending to the sheer mechanical, articulatory aspects of pronunciation and achieving accuracy and dexterity in controlled practice does not necessarily lead to transfer those same skills to actual conversation (Jones, 1997). Aspects should also be considered that are typical of connected speech, therefore sentences and dialogues should also be taken into account – especially those that are more likely to occur in everyday communication. Moreover, varied practice material should be chosen that meets different individual cognitive styles, in order to stimulate student motivation and participation, two aspects that go hand in hand with good performance (Morley, 1991; O'Malley & Chamot, 1990; Rubin, 1987; Wharton, 2000). The exercises, for instance, could follow a recurring pattern made up of different formats (e.g., 'listen and repeat' exercises, 'build a sentence' exercises, dialogues and role-plays, etc.)

Finally, in order to encourage even the most reticent students to engage in talk, special care should be taken to create a stress-free environment, as communicative tasks in the L2 have been shown to generate the highest levels of anxiety of all learning tasks (Young, 1990). This need is particularly acute when training adults, who are generally more inhibited than children and reluctant to produce speech in a foreign language for fear of losing face – or even their linguistic identity (Guiora, Brannon, & Dull, 1972).

2.3. Feedback

The issue of feedback is still controversial: There appears to be a general disagreement on the definition of corrective, implicit, explicit, or metalinguistic feedback, on whether different types of feedback should be considered as a form of positive or negative evidence, and on what constitutes evidence for the effectiveness of this factor, especially where pronunciation is

concerned. On the whole, however, research on adult second language acquisition indicates that corrective feedback from teachers, peers or native speakers makes adult learners notice the discrepancies between their output and the L2 (Long, 1996), an awareness which mere exposure to the L2 does not guarantee. According to the 'noticing hypothesis' formulated by Schmidt (1990), it is only this awareness that can lead to the acquisition of a specific linguistic item. The importance of feedback appears even more obvious for learning L2 pronunciation, because many errors produced by L2 learners can be attributed to unconscious interference phenomena from the L1 built-in phonological representations (Flege, 1995). The L1 influence can be so overwhelming that the learner may not perceive the deviations in his/her interlanguage from L2 standards. Feedback must then come into play, and more specifically, "a type of feedback that does not rely on the student's own perceptions" (Ehsani & Knodt, 1998, p. 9). Through the provision of feedback, teachers can bring the students to focus on specific individual problems and (indirectly) stimulate them to attempt self-improvement. It is obvious that it is only once this awareness has been raised that the individual can take remedial steps.

In spite of the crucial role of this factor, very little research has been carried out on the effectiveness of different types of feedback. Recent studies seem to indicate that *recast*, that is, a "repetition with change" (and possibly with emphasis) of the student's incorrect utterance (Chaudron, 1977, p. 39), is the most common type of feedback adopted by teachers (see Nicholas et al., 2001 for an overview). Recasts seem to be effective because they are unobtrusive and thus do not interrupt the conversational flow, and because they are immediate and thus allow for comparing and noticing of the new item to be learned (Nicholas, Lightbown, & Spada, 2001). With regard to pronunciation, it is worthwhile mentioning a study conducted by Lyster (1998) in French immersion classrooms, which investigated teachers' feedback strategies in student-teacher interactions and attendant learner uptake – that is, the immediate repair that students adopt on the basis of feedback. In this study, Lyster found that recast had the highest rate of uptake for phonological errors, while it yielded the lowest rates of uptake for grammatical and lexical errors. For these errors, the use of *elicitation feedback*, which always required the students to attempt to generate the correct form themselves, produced the highest rates of correct student-generated repairs. Contrary to research on complex grammatical tasks (Crompton & Rodrigues, 2001; Lyster & Ranta, 1997; Nagata, 1993), which indicates that feedback should generate

self-repairs by stimulating higher order cognitive processes in the learner, Lyster's study suggests that a simple reformulation of the mispronounced utterance immediately following the student's turn might be sufficient to successfully correct it. Other studies have also indicated the effectiveness of recasts at least for errors for which the learner has already *acquired* the specific linguistic alternative offered and thus only needs to activate lower order functions (Nagata, 1993; Nicholas et al., 2001). Recasts have also been indicated as a good form of feedback for beginner learners, because these learners are not proficient enough to discover the correct version themselves if an error is merely signalled (Lightbown, 2001). However, in this case recasts only seem to lead to a temporary repair, rather than to the long-term retention of the correct item (Nicholas et al., 2001).

It should be noted that most studies, including Lyster's (1998), have only investigated the *short-term* effects of corrective feedback, because of the difficulty in isolating this factor in a real learning environment. Moreover, contradictory results have been obtained because of inconsistent operasionalisations of different types of feedback among different studies. What seems uncontroversial is that feedback should not be limited to classifying a response as correct or wrong, but should pinpoint specific errors and possibly suggest a remedy (Chun, 1998; Crompton & Rodrigues, 2001; Warschauer & Healey, 1998). In other words, besides receiving a score, the student should *comprehend* why she/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, the student's level of proficiency in the L2, and so forth.

A number of studies have addressed the issue of pronunciation error gravity hierarchies in an attempt to establish which errors should be given priority in a pronunciation training programme (Anderson-Hsieh, Johnson, & Koehler, 1992; Derwing & Munro, 1997; Van Heuven, Kruyt, & de Vries, 1981). The problem with many of these studies is that they suffer from methodological limitations, because no distinction was drawn between accentedness and intelligibility (see Derwing & Munro, 1997). As a result, the findings of the various studies may sometimes seem contradictory. Although clear

indications are still lacking, it appears that both segmental and supra-segmental factors are important (see Derwing et al., 1998 for an overview). Segmental errors can preclude full intelligibility of speech (Derwing & Munro, 1997; Rogers & Dalby, 1996). On the other hand, lexical stress and intonation are important too, as they help listeners to process the segmental content by adding structure to the complex and continuously varying speech signals (Celce-Murcia et al., 1996). Furthermore, both levels are so tightly interwoven that, while they can be separated and measured instrumentally, in reality they influence each other, as the case of stress placement well illustrates.

2.4. Conclusions

On the basis of this brief synopsis, we can outline some basic recommendations for the ideal design of effective pronunciation teaching and learning. Learning must take place in a stress-free environment in which students can be exposed to considerable and meaningful input, are stimulated to actively practise oral skills and can receive immediate feedback on individual errors. Input should pertain to real-world language situations, it should include multiple-speaker models and it should allow the learner to get a sense of the articulatory movements involved in the production of L2 speech. Oral production should be elicited with realistic material and exercises catering for different learning styles, and should include pronunciation of full sentences. Pertinent and comprehensible feedback should be provided individually and with minimum delay and should focus on those segmental and supra-segmental aspects that affect intelligibility most.

3. CAPT SYSTEMS

If we assume that traditional class-based pronunciation teaching should be shaped according to the recommendations that we just outlined, CAPT systems should be able to follow the same recommendations. Moreover, if implemented adequately, CAPT can even offer a number of advantages compared to classroom instruction. First, by allowing the student to freely roam through the system, these programs make it possible to address individual problems. In addition, on the basis of priorities set by the user, CAPT is nowadays able to present the student only with certain tasks aimed at developing specific skills. Second, CAPT systems allow the students to train

as long as they want and at self-paced speed. Third, as some studies suggest (Murray, 1999), the privacy and the self-directed kind of learning offered by these environments may lead to a reduction of foreign language anxiety – a phenomenon strongly linked to social-judgement factors (Young, 1990) – and thus indirectly favour learning. Furthermore, student profiles can be stored by the system in a log-file so that the students themselves can monitor problems and improvements, which in turn might result in increased motivation. Alternatively, the teacher can refer to the logs and suggest appropriate remedial steps. Finally, the student might in certain cases receive feedback on oral performance from the program itself, in real-time. On account of these advantages, there have been various attempts to develop CAPT systems. However, the ideal requirements that we sketched in the previous section are not often met by existing CAPT systems.

3.1. Input and Output in CAPT

Needless to say, all the systems that are currently available provide abundant oral input. Some systems – presumably in an attempt to save disk space and compact the package in a single CD-Rom – make use of stills to accompany the information provided orally, sometimes adding text in balloons (Auralog, 2000; ILT, 1997). Several systems also provide information on the way the target speech sounds are to be produced by explaining how the articulators should be positioned. This is often done by means of a 3D representation of a mouth producing a sound, sometimes accompanied by a written explanation (Auralog, 2000; Glearner, 2001; Pro-nunciation, 2002), or by a video of a native speaker pronouncing the targeted sound (see for instance the *Advanced* series, Eurotalk, 2002; Glearner, 2001; Nieuwe Buren, 2002). Animations and videos are obviously to be preferred: while the mouth animations provide precise and realistic visual cues of single phones, the film fragments also include information on facial expressions and gestures that accompany L2 speech acts and thus provide information on pragmatic function too. Moreover, research indicates that the use of digital multimedia materials can foster language learning because it looks authentic and appealing, it promotes proactive involvement and engages various learning processes (Liontas, 2002; Wachowicz & Scott, 1999). Despite the pedagogical usefulness of such functionalities, we have seen that speaking is crucial for improving pronunciation. Therefore, a system that only provides input and merely trains receptive abilities will appear remarkably limited from a second language learning perspective.

For this reason, most current CAPT systems are designed so as to stimulate the user to produce speech that can subsequently be recorded and played back. The student can thus study his/her own output and attempt to improve it by comparing it with a model, pre-recorded utterance. Examples of these systems are described in Tutsui, Masashi, & Mohr (1999) and Van de Voort (1995); the reader is also invited to consult the CALICO Reviews (2002), the CALL Product Reviews (2002), and the LLT Software Reviews (2002), which critically describe many systems featuring these functionalities. The main problem with such systems is that it is up to the students to determine whether and how their utterances differ from the native ones, while they may lack the criteria and the awareness required to perform such an evaluation. As we have already pointed out, numerous studies have revealed that L2 learners often fail to perceive phonetic differences between their L1 and the L2 and that therefore external feedback is needed (see 2.3). On the other hand, those systems that require a teacher to listen to the recordings and to evaluate them suffer from unfavourable teacher-student ratios, just like language classes in schools and universities (e.g., Nieuwe Buren, 2002). Moreover, the functionalities offered by these systems are not innovative compared to those employed in the traditional language lab. Finally, there are systems for distance learning that resort to external feedback. These systems require the students to first practise and record themselves and then either up-load the audio-files to a web page or send the files via e-mail in near real-time. Licensed trainers listen to the files, evaluate and score them, and finally send them back to each student (Ferrier & Reid,2000; Ross, 2001). This time, the problem is due to the fact that the student has to rely on a third party, and the feedback arrives with a substantial delay.

On account of these shortcomings, we will now consider those systems that provide input, opportunity for student's output and *automatic* feedback that the student can retrieve and study when and as long as she/he wishes. More specifically, we will provide an appraisal of the most representative CAPT systems featuring these options by concentrating on the issue of feedback, which represents the biggest challenge for systems that claim to aim at providing a one-to-one tutor-student interaction, and by looking at how the pedagogical indications we have outlined are practically implemented with presently available technology.

3.2. Feedback in CAPT

The exact notion of external, corrective feedback is far from clear: in second language acquisition, the term generally refers to information provided by native speakers or teachers on a non-targeted utterance – often called *negative evidence* – but a more detailed definition is lacking, as is a classification of different types of feedback and of their respective effectiveness for learning. In CALL systems, the term is mainly used to refer to information on errors or on performance on a task in general, as a form of assessment of success, thus including scores as well. Sometimes the term is even used to refer to instructions, explanations or clues in help facilities (see for instance Pujolà, 2001). It is only too natural, then, that current computer-generated feedback on pronunciation exploits different techniques and graphical displays, targets different aspects of pronunciation, and is more or less informative and explicit. In the following section we will examine various approaches to feedback, in an attempt to establish which forms are more effective for learning.

3.2.1. Visual Displays

Some CAPT systems provide instantaneous feedback in the form of graphic displays such as 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, which are generally authoring tools, make use of programs that perform acoustic analyses of amplitude, pitch, duration and spectrum of the students' speech. Some of these systems are *WinPitchLTL* (Germain-Rutherford & Martin, 2000; WinPitch, 2002) developed by Pitch Instruments Inc., *VisiPitch* by Kay-Elemetrics (Kay, 2002; Molholt, 1988, 2001) and *VICK* (Nouza, 1998). Akahane-Yamada & McDermott (1998) and Lambacher (1999) also describe similar systems, which they used to teach English consonants to Japanese native speakers. The signal representations used in these CAPT systems were not originally conceived as a means to support pronunciation training. On the contrary, they were all designed to support phoneticians and speech scientists in specialized scientific research. Nevertheless, research on pronunciation has generally shown that these types of visual displays, if paired to auditory feedback, can contribute to improve pronunciation, especially with respect to intonation (Akahane-Yamada & McDermott, 1998; Anderson-Hsieh, 1992; De Bot, 1983). The effectiveness of these types of displays is nonetheless questionable for a number of reasons.

First of all, while attesting the usefulness of visual displays, some researchers also hypothesize that the improvements noticed after training with this type of systems might simply be the result of the fact that the student has devoted extra time to practice (De Bot, 1983). Second, 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. Third, the fact that the system shows two comparable displays, one corresponding to the incoming utterance and one corresponding to the model utterance, wrongly suggests that the ultimate aim of pronunciation training is to produce an utterance whose spectrogram or waveform 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, while capturing waveforms and computing spectrograms is relatively easy, 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 when using systems of this kind might be to train the students to autonomously read the spectrograms and the waveforms. However, even students who have received some specific training are likely to have a hard time deciphering these displays and extracting, from these raw data, the information needed to improve pronunciation: correcting articulatory behaviour on the basis of spectrograms and waveforms is particularly difficult because there is no simple correspondence between the articulatory gesture and the acoustic structure in the properties displayed. In other words, as many authors regret, this type of feedback is not in line with the requirement that feedback should first of all be easy to comprehend (Ehsani & Knodt, 1998; Eskenazi, 1999; Kommissarchik & Kommissarchik, 2000; Menzel, Herron, Bonaventura, & Morton, 2000). Finally, since spectrograms and waveforms cannot tell the average learner much about his/her errors and the specific causes of those errors, 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 (Eskenazi, 1999).

Pro-nunciation (Brown, 2001; Pro-nunciation, 2002) is a prototypical system that aims at teaching pronunciation of words by providing limericks, tongue twisters, 3D animated mouth representations of phonemes, and the possibility to display waveforms of the student's utterance for comparison

with the model one. The criticism of these kinds of displays is all the more appropriate in the case of waveforms, since these are even more variable and less informative than spectrograms. Other systems, like the *Talk To me* (TTM, 2002) and the more comprehensive *Tell me More* series (Auralog, 2000), are not exclusively based on waveforms as a form of feedback, in that a global score is also provided and words that are incorrectly pronounced within a sentence are colour-coded. However, the graphical importance the waveforms have on the screen suggests that they are presented because of their flashy look, to impress the users – that is, the buyers.

A much-praised system, *WinPitchLTL* (Germain-Rutherford & Martin, 2000; WinPitch, 2002), has been developed by two phoneticians working on speech technology and pedagogy, as an authoring tool for different learning environments. This system is able to analyse recorded speech of a maximum duration of 12 min and display the pitch curve, the intensity curve and the 'speech signal' (in the form of a waveform or of a spectrogram). The main advantage of this system is that it features 'word-processing' facilities: the teacher can easily segment the speech signal displayed, label it by adding text on the display, highlight with different colours relevant segments in the melodic curve or significant cues on a spectrogram, thereby making important information easily visible and retraceable for the student. These are operations that the system cannot perform automatically as the technology that underlies it cannot segment a complex speech signal. *WinPitchLTL* also contains a synthesis feature that allows the teacher to modify the prosodic parameters of a student's utterance and redesign its acoustic properties within a given range on the basis of the target model. In this way, the student can hear the correct prosodic contours with his/her own voice, which has been shown to help the student to better perceive important deviations (Nagano & Ozawa, 1990). However, the effectiveness of this system totally relies on the teacher: a teacher must be available who previously received sufficient training in phonetics and acoustics and who is able to pass on that information to the students by editing the speech signal, while this, of course, is not the common rule (Price, 1998). While this system offers the stated advantage to help teachers clearly indicate what a pronunciation problem was and how it can be improved, it is unlikely that a teacher will be able to edit a large number of utterances in such a detailed way. In other words, feedback will be subject, once again, to time constraints and unfavourable teacher-to-student ratios.

Sometimes graphic displays of pitch contours, without the addition of the oscillogram or spectrogram, are used to give feedback on intonational patterns

(see Chun, 1998). Like other systems using displays, these programs presuppose some degree of training in interpreting the displays. However, pitch contours are easier to interpret than spectrograms or oscillograms. In addition, while it is doubtful whether attempting to match a spectrogram or an oscillogram is a meaningful exercise, trying to approach a pitch contour does certainly make sense. Kommissarchik and Kommissarchik (2000) have discussed the shortcomings of various forms of supra-segmental feedback and have developed a system for teaching American English prosody to non-native speakers of English, *BetterAccentTutor*, in which readily accessible feedback is provided. Visual feedback is provided on all three components of prosody: 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 are displayed on the screen so that the students can compare them and notice the most relevant features they should match. The system offers two major, easy-to-interpret visualisation modes: intonation – visualised as a pitch graph on vowels and semivowels – and intensity/rhythm – visualised as steps (syllables) of various length (duration) and height (vowel's energy). This program, however, does not address segmental errors. The rationale behind the system is based on the assumption that "the three factors that have the biggest impact on intelligibility of speech are intonation, stress and rhythm" (Betteraccent, 2002), but no hierarchy order for speech intelligibility has yet been established and research has evidenced that segmental errors can be detrimental for comprehension too.

3.2.2. Automatic Assessment

With the exception of *BetterAccentTutor*, the systems described above have in common that the computer produces some kind of direct visual representation of the speech signals, and all interpretation or manipulation is left to the student and/or the teacher. Let us now take a look at some programs that do not require constant support of a teacher and that let the *computer* compare model and student's utterances with a view to producing a pronunciation quality score. In this case, the feedback usually consists of a numerical or symbolic score, for example, an icon such as a smiley, an oral comment such as 'well-done', or a graded-bar indicating the degree of 'nativeness' – which is automatically generated by the system. The usefulness of automatic scoring is evident as this technology gives the learner immediate, comprehensible information on output quality. 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 ratings of pronunciation quality 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 the case of 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 (Cucchiarini, Strik, & Boves, 2000; Franco, Neumeyer, Digalakis, & Ronen, 2000). 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. SRI's *FreshTalk* exemplifies the sort of system in which measures of non-nativeness such as temporal speech properties are used as a basis for providing feedback, and indeed, the feedback related to speech rate did not prove to be effective for improving the students' pronunciation skills (Precoda, Halverson, & Franco, 2000). Given the limited usefulness of scores, programs should not solely rely on this type of feedback. Rather, they might use it to integrate 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. The *Tell me More* and the *Talk to Me* series by Auralog (Auralog, 2000; TTM, 2002) allow the students to train communicative skills through interactive dialogues with the computer. The student hears an oral question that is simultaneously displayed on screen, and replies with an answer that she/he chooses from three written responses that are phonetically different. Through ASR that has been specifically trained for non-native speech, the computer recognizes the student's utterance and accordingly moves on to the following conversational turn. If the program does not understand the student, it will prompt him/her to repeat the response. As each choice leads the dialogue along a different path, the program ensures a certain degree of realism. Additionally, the student can choose to check his/her oral performance on a page displaying the score she/he received, the waveform and the sentence she/he produced, with the mispronounced words coloured red. Another system that simulates a real-world, game-like learning setting is the *Microworld* contained in the *Military Language Trainer* (MILT, the version being used at the U.S. Military Academy – MITAS is its

commercially available sibling; Holland, Kaplan, & Sabol, 1999; LaRocca, Morgan, & Bellinger, 2001; MITAS, 2002). In this case, the student orally asks the computer to perform a simple action in a room with several objects, such as 'put the book on the table'. If the computer understands the utterance, it will perform the command given by the student. A similar method is used in CPI's *TraciTalk* (see TraciTalk, 2002; Wachowicz & Scott, 1999) even though this system was conceived as a more generic CALL environment, rather than a CAPT system: the student interacts with an animated agent whose task is to help the student to solve a mystery using the target language. The type of feedback these systems provide is undoubtedly very effective to reinforce correct pronunciation behaviour, as it realistically resembles the type of interaction that would take place with a human interlocutor. Moreover, it exploits the advantages that involvement in games has for learning (Wachowicz & Scott, 1999) and it allows the student to acquire knowledge through task-based learning, that is, while achieving non-linguistic goals (Nunan, 1989). However, neither *Microworld* nor *TraciTalk* are able to offer any help if a student cannot make him/herself intelligible because, for instance, she/he cannot correctly pronounce a certain sound.

A serious attempt at diagnosing segmental errors and providing feedback on them has been made in the *ISLE* project (*Interactive Spoken Language Education*; Menzel et al., 2000; ISLE 4.5, 2001). This system targets German and Italian learners of English, and aims at providing feedback on pronunciation errors, focusing in particular on the word level, for which it checks mispronunciations of specific sounds and word-stress errors. Limiting the system to a (few) known language pair(s), allows for good recognition performance by the ASR: not only is the system specifically trained to recognize non-native, deviant speech in the given L2, it is also trained to recognize typical errors due to interference from (a) specific L1(s). However, this approach can only be adopted for specific L1-L2 pairs for which sufficient knowledge of typical pronunciation errors is available, as in the case of the languages addressed in the *ISLE* system. It follows that these systems are not able to handle unexpected, idiosyncratic errors that may be frequently made by some learners and that may be detrimental to intelligibility. Another limitation is that phonemically different sounds (such as /æ/ and /ɛ/ in English) overlap in acoustic space if the models of the sounds must be trained independently of a specific speaker (which is necessary if it must be possible for arbitrary learners to use the system). This makes it very difficult for the system to reliably decide whether the English words *land* and *end* were

pronounced with the correct vowel (Hillenbrand, Getty, Clark, & Wheeler, 1995).

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 report that only 25% of the errors are detected by the system and that over 5% of correct phones are incorrectly classified as errors. As the authors comment, with such a performance "students will more frequently be given erroneous discouraging feedback than they will be given helpful diagnoses" (Menzel et al., 2000, p. 54). Thus, future CAPT systems that use ASR to detect pronunciation errors should focus on errors that can be detected with a high degree of robustness. In addition, it should help if more, carefully transcribed non-native speech in different L2s became available: this could be used to train an ASR system for the specific task of detecting typical pronunciation errors. Nevertheless, even if the performance of an ASR system is optimised, it will never be perfect, and, consequently erroneous feedback will occasionally be provided.

Erroneous feedback is a common problem in CAPT systems using ASR technology (see for instance the evaluation of *TriplePlayPlus* and *Learn German Now*! in the CALICO Software Reviews, 2002; LLT Software Reviews, 2002). Patently wrong error detection can be so frustrating for the student that Wachowicz and Scott (1999) recommend using implicit rather than explicit, judgemental feedback. For example, a system that only indicates the part of a word or utterance that was mispronounced, without indicating exactly which erroneous sounds it recognised, is likely to make fewer errors than the *ISLE* system, simply because it makes only half the number of decisions. And, as some suggested with regard to recasts, telling the student that some areas in his/her utterance were incorrect and offering him/her the possibility to listen to the correct version – without attempting to also play a version of the confusable counterpart – might just be sufficient feedback. Still, it would be necessary to focus on pronunciation problems that are robust to detect. It goes without saying that those are errors where the distance between the wrong and correct pronunciations is relatively large. Even if these errors do not cause confusions between words, they are so conspicuous for a listener that they are likely to affect intelligibility.

3.3. Conclusions

To summarize, this overview of available CAPT systems has identified a number of pros and cons of these systems, which should be taken into consideration when developing new prototypes. On the whole, we have seen that an ideal system should provide input, output and feedback, and should incorporate ASR technology.

With regard to input and output, we have observed that presently available technology is sufficiently advanced to match the pedagogical requirements sketched in Section 2. The technology can now even offer possibilities that are not available in traditional classroom learning. The limitations of those systems that make use of outdated or less effective multimedia are only attributable to economic constraints or choices made by the developers, and not to problems inherent in the technology.

What still remains problematic is the issue of feedback: its implementation in CAPT systems needs to be studied carefully. We have seen that it is only through the integration of ASR technology and pedagogical guidelines that we can design programs providing real-time, pertinent and easy-to-interpret feedback both on segmental and supra-segmental aspects. However, the limitations in current ASR technology imply that error diagnosis will only be possible with a limited degree of detail. Even if pedagogically desirable, detailed diagnosis is simply not feasible because the performance levels attained are too poor. Reliability is crucial in language learning: nothing could be more confusing for a learner than a system reacting in different ways to successive realizations of the same mistake. It therefore seems that, if we want to reach an ideal compromise between technology and demand, we will have to settle for something that is less ambitious, but that can guarantee correct feedback at least in the majority of the cases.

4. RECOMMENDATIONS FOR FUTURE CAPT SYSTEMS

In this analysis of available CAPT systems, we have observed that various devices are sometimes used without an underlying pedagogical criterion, simply to make a fancy-looking product. In other cases, displays that are easy to produce are used, while they either have little pedagogical value, or are not transparent for the student and thus require support from an expert. We therefore suggest that developers first focus on the learner's needs and accordingly select functionalities and technology that meet those needs.

A promising way to do this is by incorporating the indications that are available from research on second language acquisition and teaching. In this way, it is possible to suggest ways to design CAPT systems that make use of advanced technologies to achieve pedagogical effectiveness.

As a general rule, that is, whenever the learner does not have special needs, it seems advisable to include pronunciation training within a comprehensive programme based on a communicative approach, as such an environment is more likely to lead to meaningful, contextualised learning. Similarly, pronunciation training should aim at attainment of speech intelligibility, rather than 'nativeness' or accent-free pronunciation. As to the basic technical requirements an ideal system should fulfill, it is mandatory to include multimedia while keeping the navigation through the system intuitive and easy, in order to make the learning setting as 'human' and realistic as possible, and to prevent 'technophobia' in the students (Murray & Barnes, 1998).

The system should contain a considerable amount of input in the L2. Input should ideally be presented in interactive audio-visual material produced by different native speakers, such as film fragments and radio interviews. Detailed study of articulatory movements could be catered for by means of 3D computer animations of the lips and oral cavity. Simulations of real-life situations that are particularly likely to be experienced by the learners should serve as learning context, because the relevance and the authenticity of this type of input can boost the learner's motivation.

The system should also include tasks that stimulate the student to practise what she/he has learnt by interacting with the system. Exercises should be realistic, varied and engaging and should not be limited to listen-and-repeat drills with isolated sounds or words. Role-plays with the characters in the system or interactive dialogues with the computer as those used in some of the systems we have presented are, for instance, a good method to let the student practise. Obviously, this type of exercises is only possible when ASR technology specially tuned for non-native speech recognition is used. Moreover, this implies that the speaker's utterance has to be predictable because state-of-the-art ASR is not able to recognize free output with a satisfactory degree of reliability. Having to reckon with this constraint nevertheless means that the students will always be able to compare their output with a model utterance. Another means to allow the students to self-monitor their problems and progress is to automatically store each utterance in a log-file.

Ideal systems should always include an option to provide feedback by means of ASR technology, so that the user can receive immediate, individualised information on his/her performance. Because of the limitations of this technology, error diagnosis will only be possible with a limited degree of detail. Automatic feedback on the students' responses could for instance be given in real-time at two levels: a graded-bar could be used to score overall comprehensibility, while presumably incorrect areas are highlighted. A description of mispronounced phones or syllables could then be offered by means of visual and aural feedback, with an option for comparing input with output. In assessing pronunciation performance, both segmental and supra-segmental aspects should be considered, such as temporal and spectral quality of speech sounds, word-stress, and sentence-accent.

Furthermore, in order not to discourage the students, a maximum number of errors to be pinpointed should be set for each utterance. The feedback system could also be programmed to select and address only certain pronunciation errors. Instead of trying to match an entire waveform, for example, the student's attention might be directed only to a few specific deviations. A question that could arise at this point is according to which criteria the errors should be selected. We would like to suggest at least four important criteria that could be used to select errors with a view to increasing the efficiency and the effectiveness of CAPT systems: (1) error frequency; (2) error persistence; (3) perceptual relevance; and (4) robustness of error detection.

First, the importance of error frequency is obvious: addressing errors that are infrequent will have little impact on pronunciation performance and will therefore not significantly contribute to improving communication. Second, concentrating on persistent pronunciation errors is a question of efficiency. Why should we put effort in errors that simply disappear through exposure to the L2? Third, focusing on errors that are perceptually relevant is a direct consequence of the ultimate aim of pronunciation training as we see it: improving learners' intelligibility. It follows that priority should be given to those errors that slow down and even hamper communication. Finally, as explained above, not all pronunciation errors can be detected automatically with a sufficient degree of robustness. As mentioned above, since reliability is crucial in language learning, only errors that can be reliably detected should be addressed.

To conclude, in this paper we have outlined some pedagogical requirements that CAPT should ideally meet, and we have looked at how those requirements are technologically implemented in available CAPT systems.

We hope that the suggestions we have given for future work can contribute to ameliorating CAPT design. However, further research is needed to establish the effectiveness of specific systems that employ the functionalities suggested here.

ACKNOWLEDGEMENTS

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

REFERENCES

Abercrombie, D. (1991). Teaching pronunciation. In A. Brown (Ed.), *Teaching English pronunciation: A book of readings* (pp. 87–95) London. Routledge: (Original work published 1956).

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

Anderson-Hsieh, J. (1992). Using electronic visual feedback to teach suprasegmentals. *System*, *20*, 51–62.

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

Auralog (2000). *Tell me more*. *User's manual*. Montigny-le-Bretonneux, France.

Ausubel, D. (1968). *Educational psychology: A cognitive view.* New York: Holt, Rinehart & Winston.

Betteraccent. (2002). *Our methodology* [Last consulted 25/02/2002]. Available: http://www.betteraccent.com/baapproach.htm

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late learners. In D. Birdsong, (Ed.), *The critical period hypothesis and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Bongaerts, T. (2001). Age-Related Differences in the Acquisition of L2 Pronunciation: The Critical period. *Hypothesis revisited*. Paper presented at the EUROSLA pre-conference workshop on the Age Factor in L2 acquisition. Paderborn, Germany.

Brennan, E.M., & Brennan, J.S. (1981). Accent scaling and language attitudes: Reactions to Mexican American English speech. *Language and Speech*, *24*, 207–221.

Brown, I. (2001). Pro-Nunciation: The English Communication Toolkit. CALICO Journal Software Reviews. *CALICO Journal*, *19*, 205–217.

CALICO Reviews. (2002). The CALICO Review [Last consulted 26/02/2002]. Available: http://astro.temple.edu/∼jburston/CALICO/index.htm

CALL Product Reviews. (2002). Computer Assisted Language Learning @ Chorus [Last consulted 26/02/2002]. Available: http://www.writing.berkeley.edu/chorus/call/index.html

Celce-Murcia, M., Brinton, D.M., & Goodwin, J.M. (1996). *Teaching pronunciation*. Cambridge: CUP.

Chapelle, C.A. (1997). CALL in the year 2000: Still in search of research paradigms? *Language Learning and Technology*, *1*, 19–43 [On-line] [Last consulted 27/02/2002]. Available: http://llt.msu.edu/vol1num1/chapelle/default.html

Chapelle, C.A. (2001). Innovative language learning: Achieving the vision. *ReCALL*, *13*, 3–14.

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

Chun, D.M. (1998). Signal analysis software for teaching discourse intonation. *Language Learning and Technology*, *2*, 61–77 [On-line] [Last consulted 27/02/2002]. Available: http://llt.msu.edu/vol2num1/article4/index.html

Cole, R., Carmell, T., Connors, P., Macon, M., Wouters, J., De Villiers, J., Tarachow, A., Massaro, D., Cohen, M., Beskow, J., Yang, J., Meier, U., Waibel, A., Stone, P., Fortier, G., Davies, A., & Soland, C. (1998). Intelligent animated agents for interactive language learning. *Proceedings of InSTILL* (pp. 163–166). Marholmen, Sweden.

Crompton, P., & Rodrigues, S. (2001). 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 workshop on Computer Assisted Language Learning, Artificial Intelligence in Education Conference* (pp. 70–82). San Antonio, TX.

Crookes, G., & Schmidt, R.W. (1991). Motivation: Reopening the research agenda. *Language Learning*, *41*, 469–512.

Cucchiarini, C., Strik, H., & Boves, L. (2000). Different aspects of pronunciation quality ratings and their relation to scores produced by speech recognition algorithms. *Speech Communication*, *30*, 109–119.

De Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech*, *26*, 331–350.

De Bot, K. (1996). The psycholinguistics of the Output Hypothesis. *Language Learning*, *46*, 529–555.

Derwing, T.M., & Munro, M.J. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition*, *20*, 1–16.

Derwing, T.M., Munro, M.J., & Wiebe, G. (1998). Evidence in favour of a broad framework for pronunciation instruction. *Language Learning*, *48*, 393–410.

Dörnyei, Z. (1998). Motivation in second and foreign language learning. *Language Teaching*, *31*, 117–135.

Ehsani, F., & Knodt, E. (1998). Speech technology in computer-aided learning: Strengths and limitations of a new CALL paradigm. *Language Learning and Technology*, *2*, 45–60 [On-line] [Last consulted 27/02/2002]. Available: http://llt.msu.edu/vol2num1/article3/index.html

Eskenazi, M. (1999). Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype. *Language Learning and Technology*, *2*, 62–76 [On-line] [Last consulted 27/02/2002]. Available: http://llt.msu.edu/vol2num2/article3/index.html

Eurotalk (2002). [Last consulted 27/02/2002]. http://www.eurotalk.co.uk/ETWebPages/Products/DVDF.html

Ferrier, L., & Reid, L. (2000). Accent modification training in The Internet Way®. *Proceedings of InSTILL* (pp. 69–72). Dundee, Scotland.

Flege, J.E. (1995). Second-language speech learning: Findings and problems. In W. Strange. (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 233–273). Timonium, MD: York Press.

Flege, J.E. (1999). Age of learning and second language speech. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 101–131). Mahwah, NJ: Lawrence Erlbaum.

Franco, H., Neumeyer, L., Digalakis, V., & Ronen, O. (2000). Combination of machine scores for automatic grading of pronunciation quality. *Speech Communication*, *30*, 121–130.

Germain-Rutherford, A., & Martin, P. (2000). Présentation d'un logiciel de visualisation pour l'apprentissage de l'oral en langue seconde. *ALSIC*, *3*, 61–76 [On-line] [Last consulted 25/02/2002]. Available: http://alsic.u-strasbg.fr/Menus/frameder.htm

Glearner (2001). [Last consulted 10/05/2001]. http://www.glearner.com

Guiora, A.Z., Brannon, R.C., & Dull, C.Y. (1972). Empathy and second language learning. *Language Learning*, *22*, 111–130.

Hendrik, H. (1997). Keep them talking! A project for improving students' L2 pronunciation. *System*, *25*, 545–560.

Hillenbrand, J., Getty, L.A., Clark, M.J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*, 3099–3111.

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

ILT (1997). *Interactive language tour.* München: Digital Publishing.

ISLE 1.4. (1999). Pronunciation training: Requirements and solutions [On-line] [Last consulted 27/02/2002]. *ISLE Deliverable* 1.4. Available: http://nats-www.informatik.uni-hamburg.de/∼isle/public/D14/D14.html

ISLE 4.5. (2001). Error diagnosis for spoken language, *ISLE Deliverable* 4.5 [On-line] [Last consulted 27/02/2002]. Available at http://nats-www.informatik.uni-hamburg.de/∼isle/public/D45/D45.html

Jones, R.H. (1997). Beyond 'Listen and Repeat': Pronunciation teaching materials and theories of Second Language Acquisition. *System*, *25*, 103–112.

Kay. (2002). Kay [On-line] [Last consulted 26/02/2002]. Available: http://www.kayelemetrics.com.

Keller, J.M. (1983). Motivational design of instruction. In C.M. Reigelruth (Ed.), *Instructional design theories and models: An overview of their current status* (pp. 383–434). Hillsdale, NJ: Lawrence Erlbaum.

Kommissarchik, J., & Komissarchik, E. (2000). Better accent tutor – Analysis and visualization of speech prosody. *Proceedings of InSTILL 2000* (pp. 86–89).Dundee, Scotland.

Krashen, S.D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.

Krashen, S.D., & Terrell, T.D. (1983). *The natural approach: Language acquisition in the classroom*. Oxford: Pergamon Press.

Lambacher, S. (1999). A CALL tool for improving second language acquisition of English consonants by Japanese learners. *Computer Assisted Language Learning*, *12*, 137–156.

LaRocca, S., Morgan, J., & Bellinger, S. (2001). Optimizing speech recognition for use by learners of less commonly taught languages. *Show and Tell presentation, EuroCALL*. Nijmegen, The Netherlands.

Leather, J., & James, A. (1996). Second language speech. In W.C. Ritchie & T.K. Bhatia (Eds.). *Handbook of second language acquisition* (pp. 269–316). San Diego, CA: Academic Press.

Levy, M. (1997). *Computer-assisted Language Learning: Context and conceptualization*. Oxford: Clarendon Press.

Lightbown, P.M. (2001). Input filters in second language acquisition. *EUROSLA Yearbook 1* (pp. 79–97). Amsterdam: John Benjamins.

Liontas, J. (2002). CALLMedia digital technology: Whither in the new millennium. *CALICO Journal*, *19*, 315–330.

LLT Software Reviews. (2002). LLT Archives – Software Reviews [Last consulted 27/02/2002]. Available: http://llt.msu.edu/archives/software.html.

Logan, J.S., Lively, S.E., & Pisoni, D.B. (1991). Training Japanese listeners to identify English /r/ and /l/ III: Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, *89*, 874–886.

Long, M.H. (1996). The role of the linguistic environment in second language acquisition. In W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). San Diego, CA: Academic Press.

Lyster, R. (1998). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning*, *48*, 183–218.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition*, *19*, 37–66.

Massaro, D.W. (1987). *Speech perception by ear and eye: A Paradigm for psychological enquiry.* Hillsdale, NJ: Lawrence Erlbaum.

Menzel, W., Herron, D., Bonaventura, P., & Morton, R. (2000). Automatic detection and correction of non-native English pronunciations. *Proceedings of InSTILL* (pp. 49–56). Dundee, Scotland.

MITAS. (2002). Multimedia Instructional Tutoring and Authoring System with 3D [Last consulted 25/06/2002]. Available: http://www.maad.com/MaadWeb/products/mitas/mitasma.htm

Molholt, G. (1988). Computer-assisted instruction in pronunciation for Chinese speakers of American English. *TESOL Quarterly*, *22*, 91–111.

Molholt, G. (2001). *Three Modes of Visualization*. Paper presented at InSTILL. EuroCALL, Nijmegen, The Netherlands.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, *25*, 481–519.

Munro, M.J., & Derwing, T.M. (1995). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning*, *45*, 73–97.

Murray, G.L. (1999). Autonomy in language learning in a simulated environment. *System*, *27*, 295–308.

Murray, L., & Barnes, A. (1998). Beyond the 'wow' factor – Evaluating multimedia language learning software from a pedagogical point of view. *System*, *26*, 249–259.

Nagano, K., & Ozawa, K. (1990). English speech training using voice conversion. *Proceedings of ICSLP*, Kobe, 1169–1172.

Nagata, N. (1993). Intelligent computer feedback for second language instruction. *The Modern Language Journal*, *77*, 330–339.

Nicholas, H., Lightbown, P.M., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, *51*, 719–758.

Nieuwe Buren. (2002). *Nieuwe Buren* [Last consulted 26/02/2002]. Available: http://www.nieuweburen.nl

Nouza, J. (1998). Training speech through visual feedback patterns. *Proceedings of ICSLP*, Sydney, Australia.

Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge, UK: CUP.

O'Malley, J.M., & Chamot, A.U. (1990). *Learning strategies in second language acquisition*. Cambridge: CUP.

Oxford, R.L., & Anderson, N.J. (1995). A crosscultural view of learning styles. *Language Teaching*, *28*, 201–215.

Pennington, M.C. (1999). Computer-aided pronunciation pedagogy: Promise, limitations, directions. *Computer Assisted Language Learning*, *12*, 427–440.

Piske, T., MacKay, I.R.A., & Flege, J.A. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*, 191–215.

Precoda, K., Halverson, C.A., & Franco, H. (2000). Effects of speech recognition-based pronunciation feedback on second-language pronunciation ability. *Proceedings of InSTILL* (pp. 102–105). Dundee, Scotland.

Price, P. (1998). How can speech technology replicate and complement skills of good language teachers in ways that help people to learn language? *Proceedings of InSTILL* (pp. 81–86). Marholmen, Sweden.

Pro-nunciation. (2002). Products [Last consulted 26/02/2002]. Available: http://users.zipworld.com.au/∼pronunce/products.htm

Pujolà, J.-T. (2001). Did CALL feedback feed back? Researching learners' use of feedback. *ReCALL*, *13*, 79–98.

Rogers, C., & Dalby, J. (1996). Prediction of foreign-accented speech intelligibility from segmental contrast measures. *Journal of the Acoustical Society of America*, *100* (Pt. 2), 2725 (A).

Ross, K. (2001). *Teaching Languages With Asynchronous Voice Over the Internet*. Paper presented at InSTILL, EuroCALL, Nijmegen, The Netherlands.

Rubin, J. (1987). Learning strategies: Theoretical assumptions, research history and typology. In A.L. Wenden & J. Rubin (Eds.), *Learner strategies in language learning* (pp. 15–30). Englewood Cliffs, NJ: Prentice Hall.

Schachter, J. (1998). Recent research in language learning studies: Promises and problems. *Language Learning*, *48*, 557–583.

Schmidt, R.W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158.

Scovel, T. (1988). *A time to speak. A psycholinguistic inquiry into the critical period for human speech*. Rowley, MA: Newbury House.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In M.A. Gass & C.G. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Rowley, MA: Newbury House.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive process they generate: A step towards second language learning. *Applied Linguistics*, *16*, 371–391.

TraciTalk (2002). Traci Talk, The Mystery [Last consulted 24/06/2002]. Available: http://www.encomium.com/CPI/CPITTTM.html

TTM. (2002). Talk to Me, the Conversation Method [Last consulted 26/02/2002]. Available: http://www.auralog.com/en/talktome.html

Tutsui, M., Masashi, K., & Mohr, B. (1999). Closing the gap between practice environments and reality: An interactive multimedia program for oral communication training in Japanese. *Computer Assisted Language Learning*, *11*, 125–151.

Van de Voort, M. (1999). *Gluren naar de buren*. Nijmegen: UTN.

Van Heuven, V.J.J.P., Kruyt, J.G., & de Vries, J.W. (1981). Buitenlandsheid en begrijpelijkheid in het Nederlands van buitenlandse arbeiders; een verkennende studie. *Forum der Letteren*, *22*, 171–178.

Wachowicz, K., & Scott, B. (1999). Software that listens: It's not a question of whether, it's a question of how. *CALICO Journal*, *16*, 253–276.

Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview, *Language Teaching*, *31*, 57–71.

Watts, N. (1997). A learner-based design model for interactive multimedia language learning packages. *System*, *25*, 1–8.

Wharton, G. (2000). Language learning strategy use of bilingual foreign language learners in Singapore. *Language Learning*, *50*, 203–243.

WinPitch (2002). Pitch Instruments Inc. [Last consulted 26/02/2002]. Available: http://www.winpitch.com.

Young, D.J. (1990). An investigation of students' perspectives on anxiety and speaking. *Foreign Language Annals*, *23*, 539–553.