Effective feedback on L2 pronunciation in ASR-based CALL

Ambra Neri, Catia Cucchiarini, Helmer Strik

A²RT, Department of Language and Speech
University of Nijmegen, The Netherlands
A.Neri@let.kun.nl, C.Cucchiarini@let.kun.nl, W.Strik@let.kun.nl

Abstract

Computer Assisted Language Learning (CALL) has now established itself as a prolific area whose advantages are well-known to educators. Yet, many authors lament the lack of a reliable integrated conceptual framework linking technology advances and second language acquisition research within which effective materials can be designed [1],[2].

The CALL world has recently witnessed a flourishing of software applications among which Automatic Speech Recognition (ASR) is gaining growing importance. The reasons for this popularity lie in the opportunities this technology offers for practising oral skills and addressing pronunciation problems, two areas that are hard to improve within traditional class-based settings.

ASR-based CALL systems appear to be particularly suited for pronunciation teaching as they allow evaluation of the learner's speech and provide appropriate, individual feedback in real-time. However, given the lack of guidelines for CALL design, most courseware products often do not provide adequate guidance to the learner [3],[4]. Moreover, owing to the limitations in the state-of-the-art technology, all ASR systems will at times generate errors [4].

The main objective of our research is to study how the frequency and seriousness of feedback errors affect learning. The domain in which we work is the acquisition of pronunciation in Dutch as L2 by adults with different language backgrounds. In our study we will use a Dutch language multimedia course, *Nieuwe Buren*, to which we will add a speech recognition module previously developed at our department, which is able to recognize and score disfluent non-native speech [5]. Automatic pronunciation evaluation and feedback will focus on segmental and supra-segmental aspects.

The system's functionality will be first tested on a group of experts and subsequently on two experimental groups of different proficiency levels. After three months the pronunciation performance of these groups will be compared to that of two control groups who used the original version of *Nieuwe Buren* without immediate feedback. This will allow us to establish whether automatic immediate feedback does indeed lead to global improvement in L2 pronunciation. A more detailed analysis of individual results within the experimental groups will then provide insight into the specific effects of different feedback errors generated by the system and their impact on learning. Students and teachers will be finally required to complete a questionnaire meant to establish the feedback system user-friendliness, comprehensibility and adequacy (the control group will, by contrast, report on learning without real-time feedback). The experiment and the tests will be subsequently repeated using an improved version of the system which also takes into account newly emerged needs.

In this paper, we first provide an overview of the literature on second language pronunciation teaching and learning in order to derive some general guidelines for effective teaching. Second, we present an appraisal of available CALL systems for L2 pronunciation training with a view to identifying pros and cons. Finally, we describe the choices made for our research on the effects of erroneous automatic feedback on pronunciation.

1. Introduction

The last decades in second language teaching and learning have been dominated by communicative language teaching methods and interactionist theories, respectively. Since the adoption of these approaches, interaction has become paramount in courseware products as the best tool to develop the learner's L2 communicative skills. Although it seems obvious that oral proficiency should play an important role within such a language curriculum, training of oral skills is still often neglected in traditional classroom instruction. The main reason for this is that it requires prolonged practice and the provision of feedback on individual problems, two tasks which are extremely time-consuming and therefore difficult to implement with class-based instruction [3].

Computer assisted language learning (CALL) systems seem to offer a viable solution to this problem as they allow students to access aural input in the L2 and practise individually as often as they wish. The integration of multimedia courseware with automatic speech recognition (ASR) technology opens up new possibilities for the training of conversational skills, thus adding extra value to these teaching environments. Software designers have devised different methods to provide automatic instruction and pronunciation scoring through ASR. Recently, some systems have been deployed that are also able to provide immediate feedback in written, aural and visual formats. Yet, given the current limitations in the technology and the lack of guidelines for CALL design, most systems do not provide adequate guidance to the learner [3],[4],[6]. For instance, highly sophisticated error diagnosis still shows erroneous performance that results in confusing feedback [7].

The aim of our research is to determine to what extent mistakes generated by a system can be tolerated in the sense that they do not impede learning. Our study will focus on the acquisition of pronunciation in Dutch by adults with different language backgrounds. Pronunciation training, scoring and feedback will be provided via a widely used comprehensive multimedia language course to which we will add an ASR module.

In this paper, we will analyse available literature on classroom pronunciation training in order to establish which feedback forms are most desirable. We will subsequently evaluate various CALL systems in which pronunciation is addressed. While examining the limitations of today's ASR technology, we will consider possible ways to deal with these shortcomings. Finally, we will combine the information thus gathered to present and substantiate the choices made in our research.

2. L2 Pronunciation: guidelines for optimal training

Guidelines for CALL practitioners that might ensure more effective material design are generally lacking. Likewise, when it comes to pronunciation training, teachers trying to devise an optimal program for their students seem to be faced with many questions, such as, 'can training (durably) affect pronunciation problems?', 'how much and what kind of attention needs to be devoted to pronunciation?', 'are there areas that need to be specially tackled and in which ways?' Given the paucity of experimental studies - not to mention the variety in teaching contexts and needs - it is difficult to offer a straightforward answer to these questions.

One of the reasons for the scarcity of relevant research is that, until recently, many educators were convinced that pronunciation did not deserve as much attention as other linguistic aspects, such as grammar, and that accent-free pronunciation was a myth [8],[9]. While it is true, as Abercrombie (in [10]) puts it, that eradicating the finest traces of foreign accent might only be necessary for the training of future spies, it is now generally agreed that intelligible pronunciation is an essential component of communicative competence [11],[12]. Moreover, a general intolerance for foreign accents has been evidenced which might place learners at a professional or social disadvantage [11],[13]. Consequently, it appears that some degree of pronunciation training needs to be provided, and that the ultimate goal in this respect should be the attainment of a reasonably intelligible pronunciation. The importance of "comprehensibility" over "correct pronunciation" emerged also from the user requirement study carried out within the European project ISLE, which aimed at developing a pronunciation training system for Italian and German learners of English [3].

In this connection, it is fundamental that a distinction be drawn between accentedness and intelligibility. These are two different, albeit related, dimensions of nonnative 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 [14]. With the exception of a few (see [14]), the studies carried out so far have not drawn such a borderline, 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 significantly and to derive some general guidelines for the teaching of pronunciation. Various studies have revealed that pronunciation learning is affected by a number of variables such as L1, level of education, age on arrival (for naturalistic settings), amount of use of L1 and L2, motivation for learning L2, etc. [12],[13]. These are all factors that can vary from person to person and that cannot be manipulated 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 blended so as to obtain better results. These are input, output and feedback. These elements 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 [15]), there are reasons to believe that input can benefit pronunciation learning. As pointed out by Leather and James in [16], 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 [17]. Furthermore, input can be modified to facilitate perception of articulatory and prosodic information, so that the target models can be correctly identified and formed, for instance, through techniques such as cue-enhancement [18][19]. Finally, specific instruction on different pronunciation aspects can lead to improvement [20].

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 [11]. 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 [21], as 'comprehensible input' alone may not be sufficient [22]. Output, by contrast, enables learners to test their hypotheses by comparing their own production with the correct input and to consequently form correct L2 representations. Through production, speakers receive a first, proprioceptive feedback on their own performance: auditory and tactile feedback are available from air-and bone-conducted pressure changes and from surfaces of articulators, while feedback from the joints, tendons, and muscles provide 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 [16]. Furthermore, output can bring learners to proceduralize target language knowledge already internalized in declarative form, it allows elicitation of more input and feedback from peers, and engages self-monitoring skills [23][24].

It therefore seems that special care should be taken to create meaningful, engaging and stress-free environments that encourage speech production even from the least talkative students – such as adults - and promote learning [11][21]. Selecting varied material that meets different individual cognitive styles and that also allows self-monitoring and planning on the part of the student should help stimulate student motivation and participation [11][25].

2.3. Feedback

Although the issue of feedback is still controversial [26],[15], research on adult second language acquisition indicates that corrective feedback makes adult learners notice the discrepancies between their output and the L2 [27], an awareness which mere exposure to the L2 does not guarantee. The importance of feedback appears even more obvious for learning L2 pronunciation, as many errors produced by L2 learners can be attributed to interference phenomena from the L1 built-in phonological representations [13],[28]. The L1 influence can be so overwhelming that simple comparison of input with output may not lead to the perception of the deviations in the learner's interlanguage from L2 standards [13],[28]. Feedback must then come into play, more specifically, "a type of feedback that does not rely on the student's own perceptions" (p. 9 in [3]) Through the provision of feedback, teachers can bring the students to focus on specific individual problems, which hopefully stimulates

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.

One of the most recent studies of corrective feedback in second language research was conducted by Lyster and Ranta (in [29]) in French immersion classrooms in Montreal. These authors studied feedback strategies in student-teacher interactions, and found that the corrective feedback given to the students could be classified into six types: Explicit Correction, Recast, Elicitation, Metalinguistic Feedback, Clarification Requests, and Repetition. They further evaluated the effectiveness of these strategies on the basis of the student's uptake, i.e., the immediate response to the feedback. This was done by looking at what happened in the turns immediately following the provision of feedback: did the students try to correct their previous utterance?, was that reformulation correct?, and was the correct form initially suggested by the student concerned, or by the teacher or another student? They found that Recast - the most commonly used technique, a reformulation of the student's utterance by the teacher - had the lowest rate of uptake, with few attempts at self correction and even fewer correct reformulations. Explicit Correction had a good rate of correct student reformulations, but these were always generated by the teacher, rather than by the student. The use of Elicitation always required the students to attempt to generate the correct form themselves, and as such produced the highest rate of correct student-generated repairs. Although Metalinguistic Feedback had a lower rate of uptake overall, a similar proportion of the attempted reformulations were correct.

Although this study did not specifically address pronunciation, it may be interesting to consider how these results can be brought to bear on pronunciation training. They seem to suggest, for example, that the most effective feedback forms are those that not only indicate the correct form but that stimulate the students to produce the correct forms themselves. If we then consider that in the case of pronunciation these strategies will also stimulate the kind of proprioceptive feedback and self-monitoring mentioned above, there are reasons to believe that elicitation will prove to be the most effective form of feedback for pronunciation too.

It goes without saying that teachers do not need to provide feedback on each of the student's mistakes: such a course of action would be discouraging for the student and extremely lengthy for the teachers themselves. 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.

A number of studies have addressed the issue of pronunciation error gravity hierarchies, but several suffer from methodological limitations because no distinction was drawn between the two dimensions of pronunciation mentioned above, accentedness and intelligibility. Although clear indications are still lacking, it appears that both segmental and supra-segmental factors are important (see [20] for an overview). Segmental errors can sometimes preclude full intelligibility of speech [14], [30]. On the other hand, intonation is important too [12]: it is "the glue that holds a message together" (p.64 in [31]) as it helps listeners to process the segmental content by indicating which meanings are important. 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 recommendations for the 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 and are stimulated to actively practise oral skills. Pertinent feedback should be provided individually and in real-time and should focus on those segmental and supra-segmental aspects which affect intelligibility most.

3.   Available L2 pronunciation training systems in CALL environments

3.1. Advantages of CALL systems for pronunciation training

CALL systems particularly seem to meet the requirements of pronunciation training and offer a number of advantages. First, they make it possible to address individual problems. Second, they allow students to train as long as they want and at self-paced speed. Third, the privacy and the self-directed kind of learning offered by these systems may lead to a reduction of foreign language classroom anxiety and thus indirectly favour learning [32]. Fourth, student profiles can be stored by the system and kept in a log-file so that both the teacher and the student can monitor problems and improvements. On account of

these advantages, there have been various attempts to develop CALL systems that specifically address pronunciation.

3.2. Non-ASR-based CALL systems for pronunciation training

Some of the systems that are currently available provide information on the way speech sounds should be produced [33] or attempt to prevent frequently occurring mistakes by explaining how the articulators should be positioned for the target sound as opposed to similar sounds in the mother tongue. Despite the advantage they offer by displaying visual cues, which have been shown to improve speech perception and production [34], these systems are remarkably limited. They merely train receptive abilities and do not prompt the student to produce an utterance, while we have shown in section 2.1 that speaking is crucial for improving pronunciation. For this reason systems have also been developed in which the learner is stimulated to produce speech that can subsequently be recorded and evaluated by a teacher or used for comparison with a native utterance by the students themselves [35][36]. The problem with such systems is that it is up to the students to determine whether and how their own utterances differ from the native ones, while numerous studies have revealed that L2 learners often fail to perceive phonetic differences between their L1 and the L2 [13]. Systems in which the recorded speech has to be evaluated by a teacher suffer from the unfavourable teacher-student ratios, just like language classes in schools and universities.

To circumvent this problem, the web-based company ViaSpeech has recently presented a distance-learning method for modifying pronunciation by combining its pronunciation training CD-Rom *SpeechWorks* with *The Internet Way* learning system. Students can first practise on their own with *SpeechWorks*; they can subsequently download a specific diagnostic test, pronounce the test items, and up-load them to their personal web page on the ViaSpeech website. Licensed trainers listen to the test items using their own computers, score them and finally send the score back to each student [37]. The obvious limitation of such a system is that it does not provide real-time feedback, which could benefit learners to a greater degree. Some researchers even believe that immediate intervention can prevent the repetition of errors that would otherwise become hard-to-break habits [31].

Some other CALL systems make use of tools that perform acoustic analyses of amplitude, intonation, duration and spectrum of the student's speech and show the results on a spectrographic display [38]. The effectiveness of these systems is also questionable, as students will have a hard time deciphering these displays and even expert phoneticians may find it difficult to extract the information needed to improve pronunciation.

3.3. ASR-based CALL systems for pronunciation training

In view of the problems mentioned so far, ASR technology seems to provide the optimal solution to pronunciation learning. The systems mentioned above only offer generic instruction that can be relevant for many different learners. But each learner is unique and ideally deserves undivided attention, therefore optimal training should envisage a one-to-one learner to tutor relationship. Systems incorporating ASR modules can provide this type of interaction, making it possible to detect individual errors and to provide immediate feedback. Nowadays there are various commercial products that make use of ASR technology to teach L2 pronunciation (see also [39],[40]). However, owing to the limitations of this technology, most of the systems available are far from ideal.

Various studies on automatic pronunciation scoring of nonnative read and extemporaneous speech indicate that some machine scores are strongly correlated with human ratings of the same speech [5],[41]. The usefulness of automatic scoring is evident as this technology gives the learner immediate information on overall output quality. Besides, anecdotic evidence of positive student appreciation of global automatic pronunciation scoring has been reported [6]. However, computer measures of nonnativeness that appear to be suitable for pronunciation testing do not necessarily constitute an appropriate basis for providing feedback on pronunciation. For example, temporal measures appear to be strongly correlated with human ratings of pronunciation and fluency, and may be suitable for pronunciation testing, but no useful feedback could be provided on the basis of these measures. *FreshTalk* exemplifies the sort of system in which nonnativeness measures are used as a basis for providing feedback, and indeed, the feedback related to speech rate did not prove to be effective [42].

Some systems, like *TriplePlayPlus* by Syracuse Language Systems (see [31]) or *Auralang* by Auralog [43] analyse the student's utterance and detect errors, but do not provide appropriate, interpretable feedback, so that the student is 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 [31]. In order to be effective, feedback should be first of all

comprehensible. Many visual displays such as oscillograms and spectrograms may look very impressive, but there is little chance that they will provide useful information on the pronunciation errors the student made [3][44].

Kommissarchik and Kommissarchik (in [44]) have discussed the shortcomings of these forms of feedback and have developed a system for teaching American English prosody to nonnative speakers of English, *BetterAccentTutor*, in which comprehensible 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 have to match. The system offers two major, easy-to-interpret visualization modes: intonation - visualized as a pitch graph on vowels and semivowels - and intensity/rhythm - visualized as steps (syllables) of various length (duration) and height (vowel's energy). This program, however, does not address segmental errors.

A serious attempt at diagnosing segmental errors and providing feedback on them has been made in the ISLE project [7]. 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 lexical-stress errors. The knowledge-based character of this system implies that this approach can only be adopted when the L1 background of the user is known, when the number of L1s is limited, and when knowledge on typical errors is available. The danger of such systems is that they are not able to detect individual intra-learner idiosyncrasies, which may also be detrimental to comprehension.

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" (p. 54 in [7]).

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 (in [39]) recommend using implicit rather than explicit, judgmental feedback.

3.4. Conclusions

To summarize, this overview of available CALL systems for pronunciation training has identified a number of pros and cons of these systems, which should be taken into consideration when developing new prototypes. We have seen that systems which incorporate ASR technology are to be preferred because they make it possible to evaluate the students' own speech and to provide feedback in real-time. However, when designing CALL systems that make use of ASR technology we will have to reckon with the limitations of this technology, which, among other things, imply that the speaker's utterance has to be predictable and that error diagnosis is only possible with a limited degree of detail. As to the type of feedback to be provided, it appears that, ideally, feedback should address both segmental and supra-segmental aspects of speech production. In addition, the form in which feedback is provided is very important: feedback should be pertinent and easy to interpret. Finally, although detailed diagnosis may be desirable, this is definitely not feasible with state-of-the-art ASR technology, because the performance levels attained are too poor. It therefore seems that we will have to settle for something which is less ambitious but that can guarantee correct feedback at least in the majority of the cases.

4. The present study

4.1. Aim

Although ASR technology appears to offer numerous advantages for pronunciation training, at the present state of development it cannot be guaranteed that the automatic evaluation of the responses will always be correct. As a consequence, erroneous automatic feedback might impede learning rather than support it and may also have negative effects on learner motivation.

The major objective of our research is to determine to what extent feedback errors generated by an ASR-based pronunciation training system can be tolerated in the sense they do not disrupt learning.

Since the generation of erroneous feedback is not a peculiarity of ASR-based systems for pronunciation learning, but is a problem that afflicts all CAL systems that go beyond the simple evaluation of response in multiple choice format, we expect that the results of this research will be relevant to a wider context by advancing our understanding of the effects on learning of the frequency and the seriousness of feedback errors in advanced self-tutoring environments.

4.2. Method

4.2.1. The context

Our research on the effects of erroneous feedback on pronunciation learning will be carried out within the context of the multimedia package for teaching Dutch as a second language *Nieuwe Buren (New Neighbours)*. This is a comprehensive course for lower and higher educated adult learners with different L1s but already familiar with the Roman alphabet. Since its release in 1995, this piece of courseware has been successfully used by a growing number of schools in the Netherlands. The multimedia program it contains, consists of 40 video episodes. Each episode, which is further divided into 10 sub-segments with 15 exercises each, presents real-life situations that are particularly likely to be experienced by the learners. In accordance with communicative language teaching guidelines, this type of material ensures that the students are exposed to authentic and meaningful language and that they actively practise even outside the classroom context. The teacher should in principle act as a coach and supervisor whose main task is to stimulate communication and interaction [46].

4.2.2. ASR-based pronunciation teaching

In line with the communicative approach in *Nieuwe Buren*, in our study we aim at attainment of speech intelligibility, following Abercrombie's claim that most language learners "need no more than a comfortably intelligible pronunciation [...] which can be understood with little or no conscious effort on the part of the listener" (p.93 in [10]). Bearing in mind this goal, and following the recommendations stemming from research on pronunciation and from our analysis of the pros and cons of available CALL applications, we will deploy an ASR-based system that enables students to actively practise oral skills and receive scores and feedback on their mistakes. For this purpose we will use an ASR module that has been developed at our department, which is able to recognize and score nonnative speech [5]. This ASR module will be included in the Dutch language course *Nieuwe Buren*.

Pronunciation training will be done, just as is currently the case, partly during group classes with the teacher, and partly individually in sessions devoted to work on the computer. During the sessions, the students' interactions will be stored in a log-file. The course requires the teacher to supervise these sessions so that help can be provided whenever a student encounters problems. The presence of the teacher should allow for a reduction of anxiety in 'technophobic' learners [44] and for provision of a certain degree of extrinsic motivation [25]. Exercises will be designed so as to prompt the students to produce oral utterances that they will be able to compare with model utterances. Audiovisual material already developed for use in the course will ensure speaker variability in the oral input.

The pronunciation errors to be addressed will be selected on the basis of the following criteria:
1) frequent, 2) persistent, 3) perceptually important, 4) reliably detectable with automatic techniques.
The rationale behind each of these criteria is explained below.
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 in this project: improving the learners' communication capacities. It follows that priority should be given to those errors that slow down and even hamper communication. Fourth, as explained above, not all pronunciation errors can be detected automatically with a sufficient degree of reliability. Reliability is crucial in language learning. Nothing could be more confusing for a learner than the system reacting in different ways to separate realizations of the same mistake.

In assessing pronunciation performance both segmental and supra-segmental aspects will be considered such as temporal and spectral quality of speech sounds, word-stress, and sentence-stress.

4.2.3. Feedback on pronunciation

Automatic feedback on the students' responses will be given in real-time at two levels: first a graded-bar will be used to score overall comprehensibility, subsequently a description of the errors will be provided by highlighting mispronounced phones or syllables by means of visual and aural feedback so that students can also compare their own output with the correct form. In order not to overwhelm the students with excessive, overly verbose and discouraging information, we will set a maximum number of errors to be pinpointed per utterance.

4.2.4. Procedure

The system will first be tested on a group of experts consisting of phoneticians and speech therapists. Once the system has shown good functionality, it will be tested on two experimental groups for three months: one group will consist of beginners while the other one will consist of advanced students. At the beginning pre-tests will be run on the experimental groups and on two control groups who will use the original version of *Nieuwe Buren* without immediate feedback. Variables such as L1, age, sex, age on arrival, level of education and reason for studying Dutch will be taken into account. After the training, the pronunciation performance of the experimental groups will be measured against that of the control groups. This evaluation will tell us whether automatic immediate feedback does indeed lead to global improvement in pronunciation. Just as human tutors sometimes make mistakes, we can already predict that the system will at times generate errors due to limitations of state-of-the-art speech technology. Therefore we will perform a more detailed analysis of individual results within the experimental groups. This will provide better insight into the specific effects of correct and erroneous feedback at different proficiency levels. Finally, students and teachers will be required to complete a questionnaire meant to evaluate the feedback system's user-friendliness, comprehensibility and adequacy (the control group will, by contrast, report on learning without real-time feedback). The experiment and the tests will subsequently be repeated using an improved version of the system that will take into account needs that emerged from the first phase of the study.

5. Acknowledgments

6. References

[1] C.A. Chapelle, CALL in the year 2000: Still in search of research paradigms? *Language Learning and Technology* 1, 1997, 19-43.

[2] M. Levy, *Computer-assisted language learning: context and conceptualization*, Oxford: Clarendon Press, 1997.

[3] F. Ehsani, E. Knodt, Speech technology in computer-aided learning: Strengths and limitations of a new CALL paradigm, *Language Learning and Technology* 2, 1998, 45-60.

[4] K.B. Egan, S. LaRocca, Speech recognition in language learning: A must, *Proceedings of InSTILL 2000,* 4-9.

[5] C. Cucchiarini, H. Strik, L. Boves, Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology, *Journal of the Acoustical Society of America* 107, 2000, 989-999.

[6] ISLE Deliverable 1.4. Pronunciation training: Requirements and solutions, 1999.

[7] W. Menzel et al., Automatic detection and correction of non-native English pronunciations, *Proceedings of InSTILL 2000,* 49-56.

[8] J. Hill, Foreign accents, language acquisition, and cerebral dominance revisited, *Language Learning* 20, 1970, 237-248.

[9] V.J. Cook, Multicompetence and the effects of age. In: D. Singleton and Z. Lengyel (eds*.), The Age Factor in Second Language Acquisition.* Clevedon: Multilingual Matters, 1995, 51-66.

[10] D. Abercrombie, Teaching pronunciation. In: A. Brown (ed.), *Teaching English Pronunciation: A Book of readings*. Routledge: London, 1991, 87-95.

[11] J. Morley, The pronunciation component in teaching English to speakers of other languages, *TESOL Quarterly* 25, 1991, 481-519.

[12] M. Celce-Murcia, D.M. Brinton, J.M. Goodwin, *Teaching Pronunciation.* Cambridge: Cambridge University Press, 1996.

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

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

[15] J. Schachter, Recent research in language learning studies: Promises and problems, *Language Learning* 48, 1998, 557-583.

[16] J. Leather, A. James, Second language speech. In: W.C. Ritchie, T.K. Bhatia, (eds.) *Handbook of Second Language Acquisition.* San Diego, CA: Academic Press, 1996, 269-316.

[17] J.S. Logan, S.E. Lively, D.B. Pisoni, 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, 1991, 874-886.

[18] A. Protopapas, B. Calhoun, Adaptive phonetic training for second language learners, *Proceedings of InSTILL 2000,* 31-38.

[19] V. Hazan, A. Simpson, The effect of cue-enhancement on consonant intelligibility in noise: Speaker and listener effects, *Language and Speech* 43, 2000, 273-294.

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

[21] H. Hendrik, Keep them talking! A project for improving students' L2 pronunciation, *System* 25, 1997, 545-560.

[22] M. Swain, S. Lapkin, Problems in output and the cognitive process they generate: A step towards second language learning, *Applied Linguistics* 16, 1995, 371-391.

[23] W.J.M. Levelt, *Speaking. From Intention to Articulation.* Cambridge, MA: MIT Press, 1989.

[24] K. De Bot, The psycholinguistics of the output hypothesis, *Language Learning* 46, 1996, 529-555.

[25] G.L. Murray, Autonomy in language learning in a simulated environment, *System* 27, 1999, 295-308.

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

[27] M.H. Long, The role of the linguistic environment in second language acquisition. In: W.C. Ritchie, T.K. Bhatia, (eds.) *Handbook of Second Language Acquisition.* San Diego, CA: Academic Press, 1996, 413-468.

[28] W. Strange, Cross-language studies of speech perception and production in second-language learning. In: W. Strange (ed.) *Speech Perception and Linguistic Experience.* Timonium, MD: York Press, 1995.

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

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

[31] M. Eskenazi, Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype, *Language Learning & Technology 2*, 1999, 62-76.

[32] D.J. Young, An investigation of students' perspectives on anxiety and speaking, *Foreign Language Annals,* 23, 1990, 539-553.

[33] http://www.Glearner.com.

[34] J.E. Flege, Using visual information to train foreign-language vowel production, *Language Learning* 38, 1988, 365-407.

[35] http://www.nieuweburen.nl.

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

[37] L. Ferrier, L. Reid, Accent modification training in *The Internet Way®*, *Proceedings of InSTILL 2000,* 69-72.

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

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

[40] J. Dalby, D. Kewley-Port, Explicit pronunciation training using automatic speech recognition, *CALICO Journal* 16, 1999, 425-446.

[41] H. Franco et al., Combination of machine scores for automatic grading of pronunciation quality, *Speech Communication* 30, 2000, 121-130.

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

[43] N. Sioufi, Lessons from Auralog, Auralang Lessons, *Proceedings of InSTILL 2000,* 117-118.

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

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

[46] M. Van de Voort, *Gluren naar de buren*, Nijmegen: UTN, 1999.