The DISCO ASR-based CALL system: practicing L2 oral skills and beyond

Helmer Strik [a], Jozef Colpaert [b], Joost van Doremalen [a], Catia Cucchiarini [a]

[a] CLST, Department of Linguistics, Radboud University, Nijmegen, The Netherlands
[b] Linguapolis, Institute for Education and Information Sciences, University of Antwerp, Antwerp, Belgium
E-mail: H.Strik | J.vanDoremalen | C.Cucchiarini@let.ru.nl; Jozef.Colpaert@ua.ac.be

Abstract

In this paper we describe the research that was carried out and the resources that were developed within the DISCO (Development and Integration of Speech technology into COurseware for language learning) project. This project aimed at developing an ASR-based CALL system that automatically detects pronunciation and grammar errors in Dutch L2 speaking and generates appropriate, detailed feedback on the errors detected. We briefly introduce the DISCO system and present its design, architecture and speech recognition modules. We then describe a first evaluation of the complete DISCO system and present some results. The resources generated through DISCO are subsequently described together with possible ways of efficiently generating additional resources in the future.

Keywords: Computer Assisted Language Learning, Automatic Speech Recognition, Second Language Oral Skills.

1. Introduction

Second language (L2) learners tend to make different morphologic and syntactic errors when they speak than when they write. It is generally acknowledged that the fact that L2 learners are aware of certain grammatical rules (i.e. those concerning subject-verb concord of number, tenses for strong and weak verbs, and plural formation) does not automatically entail that they also manage to marshal this knowledge on line while speaking. In other words, in order to learn to speak properly, L2 learners need to practice speaking and to receive corrective feedback on their performance on line, not only for pronunciation, but also for morphology and syntax.

However, intensive practice and feedback on speaking performance is costly and therefore not feasible for the majority of language learners. In the classroom, providing individual corrective feedback on oral skills is not always possible, mainly due to lack of time.

Computer Assisted Language Learning (CALL) systems that make use of Automatic Speech Recognition (ASR) provide new opportunities for practicing oral proficiency. These systems can potentially offer extra learning time and material, specific feedback on individual errors and the possibility to simulate realistic interaction in a private and stress-free environment. For pronunciation training, systems have been developed that either provide overall scores of pronunciation performance or try to diagnose specific pronunciation errors (Eskenazi, 1999; Kim, Franco & Neumeyer, 1997; Mak, Siu, Tam, Chan & Chan, 2003; Menzel, Herron, Bonaventura & Morton, R., 2000; Precoda, Halverson & Franco, H., 2000). Commercial systems are e.g., marketed by Digital Publishing (http://www.digitalpublishing.de), Auralog (http://www.tellmemore.com/), and Rosetta Stone (http://www.rosettastone.com/). Systems for practicing grammar skills in general do not feature spoken interaction and feedback on spoken utterances (Bodnar, Cucchiarini & Strik, 2011).

The CALL system to be developed in the DISCO project (http://lands.let.ru.nl/~strik/research/DISCO/) was conceived to make this possible. In this paper we first briefly introduce the DISCO project (Section 2), and then go on to describe its design (Section 3) and architecture and speech recognition modules (section 4). We then describe a first evaluation of the complete DISCO system and present some results (Section 5). The resources generated by DISCO and possible future resources to be developed through DISCO are described in Section 6. Finally, In Section 7 we draw some conclusions and consider possible future developments.

2. The DISCO ASR-based CALL system

The DISCO project was carried out within the framework of the Dutch-Flemish STEVIN programme (Spyns, D'Halleweyn & Cucchiarini, 2008) and was aimed at developing a prototype of an ASR-based CALL application for Dutch as a second language (DL2). The application allows practice in DL2 speaking through interaction in realistic communication situations, detects errors made by DL2 learners at the level of pronunciation, morphology, and syntax, and provides intelligent feedback on the errors detected.

With respect to pronunciation, we aimed at the achievement of intelligibility, rather than accent-free pronunciation. As a consequence, the system was intended to target primarily those aspects that appear to be most problematic. In previous research (Neri, Cucchiarini & Strik, 2006) we gathered relevant information in this respect.

It is well-known that recognition of non-native speech is problematic. In previous research on pronunciation error detection (Cucchiarini, Neri & Strik, 2009) we severely restricted the exercises and thus the possible answers by the learners. Since DISCO also addresses morphology and syntax, the exercises have to be designed in such a way that L2 learners have some freedom in formulating their answers in order to show whether they are able to produce correct forms. So, the challenge in developing an ASR-based system for practicing oral proficiency consists in designing exercises that allow some freedom to the learners in producing answers, but that are predictable enough to be handled automatically by the speech technology modules.

In morphology and syntax we wanted to address errors that are known to cause problems in communication and that are known to be made at the low proficiency level (A1/A2 of the CEFR) that is required in national language citizenship examinations in the Netherlands. For morphology this concerns (irregular) verb forms, noun plural formation; and for syntax it concerns word order, finite verb position, pronominal subject omission, and verb number and tense agreement.

The DISCO project was carried out by a Dutch-Flemish team consisting of two partners from the Radboud University in Nijmegen (CLST and Radboud in'to Languages), the University of Antwerp (Linguapolis), and the company Knowledge Concepts.

3.   The design of DISCO

After an initial design based on a concept where the user was expected to make choices (communicative situation, pronunciation / morphology / syntax), we eventually decided to limit our general design space to closed response conversation simulation courseware and interactive participatory drama, a genre in which learners play an active role in a pre-programmed scenario by interacting with computerized characters or "agents".

The simulation of real-world conversation is closed and receptive in nature: students read prompts from the screen. However, at every turn, students pick the prompt of their choice, which grants them some amount of conversational freedom. a) it "reduces inhibition, increases spontaneity, and enhances motivation, self-esteem and empathy" (Hubbard, 2002), b) it casts language in a social context and c) its notion implies a form of planning, scenario-writing and fixed roles, which is consistent with limitations for the role of speech technology in DISCO.

This framework allows us to create an engaging and communicative CALL application that stimulates Dutch L2 (DL2) learners to produce speech and experience the social context of DL2. On the other hand, these choices are safe from a development perspective, and are appropriate for successfully deploying ASR while taking into account its limitations (Strik et al., 2009; Van Doremalen, Strik & Cucchiarini, 2011). In order to make optimal choices with respect to important features of the system design, a number of preparatory studies was carried out in order to gain more insight into important features of system design such as feedback strategies, pedagogical and personal goals.

To gain more insight into appropriate feedback strategies, pedagogical goals and personal goals (Colpaert, 2010) preparatory studies were conducted: exploratory in-depth interviews with DL2 teachers, focus group discussions and pilot studies. The results were taken into account in finalizing the DISCO design. The learning process starts with a relatively free conversation simulation that takes account of what is (not) possible with speech technology: learners can choose from a number of prompts at every turn. Based on their errors they are offered remedial exercises, which are very specific and constrained exercises. Feedback depends on individual preferences: the default strategy is immediate corrective feedback visually implemented through highlighting, which puts the conversation on hold and focuses on the errors. Learners that wish to have more conversational freedom can choose to receive communicative recasts as feedback, which let the conversation go on while highlighting mistakes for a short period of time.

The methodological design of the system led to a relatively simple software architecture that is sustainable and scalable, a straightforward interface that appeals to – and is accepted by – the users (by responding to their subconscious personal goals), a sophisticated linguistic-didactic functionality in terms of interaction sequences, feedback and monitoring, and an open database for further development of conversation trees.

Figure 1. Learners can choose to practice different skills: pronunciation (*uitspraak*), morphology (*woordvormen*) and syntax (*zinsbouw*).

3.1 Dialogues and remedial exercises

Based on the results of the preparatory studies different approaches were chosen for the various exercises.

Pronunciation exercises: we decided that simply reading aloud sentences is a good modality for reliably detecting and correcting errors in pronunciation.

Morphology exercises: a multiple choice approach was recommended. For example, for personal and possessive pronouns: "Hoe gaat het met (jij/jou/jouw)?" ("How are (you/you/your)?") and for verb inflections: "Hoe (ga/gaat/gaan) het met jou?" ("How (are/is/to be) you?").

Figure 2 . Morphology exercise

Syntax exercises: constituents (e.g. max. four) are presented in separate blocks in a randomized order. Some of these blocks are fixed, such as the beginning and the end of the sentence. This is made clear by using differently colored blocks.
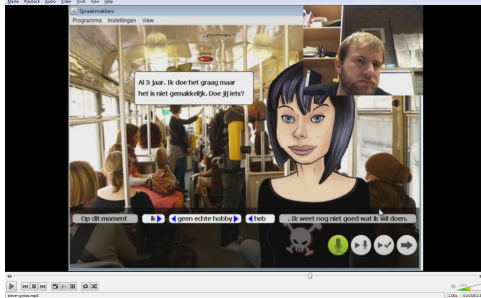
Figure 3. Syntax exercise, with in the upper-right corner a language learner using the system

After each conversation, the learner receives a report containing all mistakes that were made during the conversation, no matter what the feedback type was during the conversation. This report also contains links to remedial exercises for highly frequent mistakes. When the learner clicks on any of these links, she is directed to exercises on specific phonetic or grammatical topics in the remediation environment. The learner can also enter the remediation environment from the start/home page of the programme by clicking a link to go directly to the remediation environment rather than to the conversations.

3.2 Feedback on speaking performance

In DISCO feedback on speaking performance is provided in two contexts, during the dialogues and in the remedial exercises. Two linguistic-didactic feedback strategies are adopted. At start-up or during the use of the programme, the learner can choose either a very explicit or a more implicit, communicative feedback strategy.

The first strategy highlights pronunciation or grammatical errors and allows learners to immediately correct themselves. In the conversation environment, this puts the conversation temporarily on hold, until the learner has produced a recognizable and correct utterance.

The second strategy, which is only available in the conversation environment, is mainly communicative in nature: it repeats the student's response without the errors highlighting the erroneous graphemes, morphemes or words. This kind of feedback is known as a recast. It is corrective because it removes errors and stresses the correct forms through highlighting. It is communicative because it does not interrupt the conversation flow.

There are no fundamental differences between the feedback loops in the conversations and the feedback loops in the remedial exercises, except for the fact that in the remediation environment, only explicit feedback can be given when a learner makes errors. Thus, the software for the feedback is basically the same code, but it is activated with different parameters depending on the environment (conversation/remediation).

4. The speech technology modules

For developing the speech technology components the DISCO project was able to profit from two previous STEVIN projects: SPRAAK (Demuynck, Roelens, Van Compernolle, & Wambacq, 2008), which provided the speech recognition engine employed in DISCO, and JASMIN-CGN (Cucchiarini, Driesen, Van Hamme, & Sanders, 2008), whose adult non-native speech data were

used for different experiments within the DISCO project.

4.1 System architecture

Based on the exercises described in the previous section, we designed a system architecture which in principle is able to fulfill all the requirements stated during the courseware design phase.

The system consists of three main components: the client, the server and the courseware database. The client will handle all the interaction with the user, such as recording the audio, showing the current exercise and appropriate feedback, as well as keeping track of the user's progress. The content of the courseware is stored in the courseware database. The server is the component which processes the spoken utterances and detects errors.
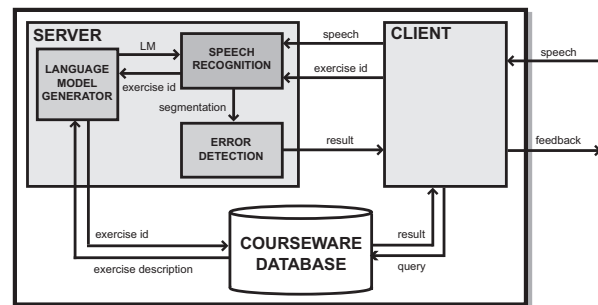

Figure 4. System architecture

In DISCO, the students' utterances have to be handled by the speech technology. For this purpose we employ a two-step procedure which is performed by the server: first it is determined what was said (content), and second how it was said (form). On the basis of the current exercise, the server generates a language model (language model generator) which is used by the speech recognition module to determine the sequence of words uttered by the student. If the speech recognition manages to do this, possible errors in the utterance are detected by the error detection module. Finally, a representation of the spoken utterance, together with detected errors, is sent back to the client. The client then provides feedback to the learner.

4.2 Speech recognition

During speech recognition, which is needed to establish whether the learner produced an appropriate answer, the system should tolerate deviant realizations of utterances. We call this step utterance selection. Exercises are designed such as to elicit constrained responses from the learner. For each exercise there is a specific list of predicted correct and incorrect responses. Incorrect responses are automatically generated using language technology tools based on the correct target responses.

4.2.1. Syntax Exercises

In syntax exercises, three or four groups of words are presented on the screen. The task of the student is to speak these word groups in a syntactically correct order. For these exercises, language models are automatically generated by including all permutations of the word groups as paths in a finite state grammar (FSG). The task of the speech recognizer is to determine which of these paths in the FSG is the most likely one given the speech input from the student.

4.2.2. Morphology Exercises

In morphology exercises, a whole sentence is presented on the screen, but for one word a multiple choice list containing alternatives for that word, typically around two to four, is presented. Here, the language models are generated in a similar fashion as in the syntax exercises. For the word that has to be chosen by the student, alternative paths are included in the FSG.

4.2.3. Pronunciation Exercises

In pronunciation exercises, language models contain only one path: the target utterance. The reason for doing this recognition is explained below.

The sequence of words that is now selected does not always correspond exactly to what was actually spoken: the spoken utterance might not be present in the FSG, or even if it is present it might not be the one that is actually recognized. Since providing feedback on the wrong utterance is confusing, we try to avoid this as much as possible. To this end we automatically verify whether the recognized utterance was spoken using a so called confidence measure, which indicates how well the recognized word sequence reflects the spoken utterance. The confidence measure is compared to a predefined threshold to determine whether the utterance has to be accepted (confidence measure above the threshold) or rejected (below the threshold). This step is called utterance verification. When the utterance is accepted the learner gets feedback on the utterance, if it is rejected the learner might be asked to try again.

4.3 Error detection

In the current system design syntactical and some morphological errors can already be detected after speech recognition, so no additional analysis is needed for these kinds of errors. However, for pronunciation errors such an analysis is required because these errors often concern substitutions of acoustically similar sounds. The canonical phone string (target pronunciation) is encoded in a weighted FSG, together with frequently observed pronunciation errors which are represented in parallel arcs. The arcs carrying pronunciation errors have a certain transition cost assigned to them, in order to keep the number of false alarms at an acceptable level.

5. Evaluation of the DISCO system

5.1 Evaluation of the subcomponents

Evaluation was carried out at several times and at several levels during the project. Pilot experiments were conducted to test the various subcomponents of the DISCO system in isolation: the exercises, the speech recognition module, and the error detection module. In Van Doremalen et al. (2010) we presented and evaluated different methods for calculating confidence measures that are employed for the utterance verification step, which is needed after the selection of the best matching utterance to verify whether the selected response was actually uttered by the learner.

To improve the detection of pronunciation errors we first conducted an experiment with artificial pronunciation errors in native speech (Van Doremalen et al. 2009a). We then studied whether the pronunciation error patterns in non-native read and spontaneous speech differ in terms of phoneme errors (Van Doremalen et al. 2010a). We used this knowledge in the development of a new type of pronunciation error classifier, which is designed to automatically capture specific error patterns using logistic regression models (Van Doremalen et al. submitted).

5.2 First evaluation of the complete system

5.2.1. Method

The criterion we defined for the evaluation of the whole system is that it should operate in a manner that is similar to what a competent teacher would do. Therefore, we chose a design in which different groups of students of Dutch as a second language (DL2) at Linguapolis, the language centre of the University of Antwerp, Flanders, and Radboud in'to Languages, the language centre of the Radboud University, Nijmegen, The Netherlands, used the system and filled in a questionnaire with which we could measure the students' satisfaction in working with DISCO. Recordings were made of the students working with DISCO and completing the dialogues and the remedial exercises. The sets of system prompts, student responses and system feedback were to be evaluated by DL2 teachers to assess the quality of the feedback provided by the system on the level of pronunciation, morphology and syntax.

5.2.2. Participants

Fourteen DL2 students at Linguapolis and nine DL2 students at Radboud in'to Languages participated in evaluation tests in which they worked with the DISCO program and subsequently evaluated it on a number of points. The students were assigned dialogues and exercises from different components of the system: syntax (7 students), morphology (8 students) and pronunciation (8 students). The age of the students varied between 20 and 40. The highest level of education was mostly university level and, in one case, secondary education. The students had different first languages (Farsi, Armenian, Russian, Portuguese, Italian, Spanish, Arabic, Polish, English) and they could all speak one or more foreign languages, most often English, followed by French. The length of stay in the Netherlands or Flanders varied between 4 months and 13 years. There was also a large difference in the amount of time spent learning Dutch, from 4 months to 7 years, but their proficiency level was at or just above CEFR level A2.

5.2.3. Results

On the whole, the students were positive about the program. In general they did not have problems using the buttons, the mouse, the keyboard, or the microphone. One student commented that the mouse did not work properly and two students found it annoying to speak into the microphone. Comments on the program's speed were largely neutral. All the students found that the program looked good; two students commented on the

attractiveness of the background and the colours. Other student comments were mainly positive (a very good program, interesting), although one student commented that the program made him/her nervous.

The students were also positive about the dialogues: they thought they were fun to do and realistic. They were also satisfied with the feedback; they understood it and thought they learnt from it. One student found the dialogues too difficult and another one observed that the system had problems in processing the answer if this was produced too quickly.

After finishing a dialogue the students were presented with a summary of their errors and were given the opportunity to practice the areas that needed improving. In general the students indicated that they learnt from the remedial exercises and thought they were fun to do.

For syntax the level of difficulty of these exercises was assessed to be "just right". For morphology, on the other hand, the practice exercises were considered to be too easy. For pronunciation, the difficulty level of the practice exercises varied from easy to difficult. One student commented that a more detailed introduction would have been useful, while another student found the response time too short.

Extra help was provided in different forms. Students could listen to a recording of their own utterance, listen to an example utterance (as they should have said it) and first see the correct utterance on the screen by choosing the correct utterance from a number of alternatives. The students found it useful to listen to their own recording and the example utterance, as well as to click on the correct answer.

All students said they would use the program themselves. The average mark assigned to the program varied from 9.0 to 7.2 in Belgium and from 8.5 to 5.0 in the Netherlands. This had to do with the defective connection with the speech processor. As one student commented, 'It's good when it works.' Several of the extra comments referred to problems with the interface and 'bugs'. One student found it annoying to have to click so much. Suggestions were also made: fun to have different levels and to have dialogues with other themes. Another student liked the fact that the sentences were first short then got longer.

To assess the quality of the feedback provided by the DISCO system DL2 teachers listened to recordings of the evaluation experiment and made annotations of the error made by the learners. For logging and data collection purposes the system automatically generates Praat (Boersma & Weenink) textgrids containing word alignments, phone alignments and pronunciation errors for each utterance. DL2 teachers in Nijmegen and Antwerp listened to sets of responses using these textgrids. For syntax and morphology the teachers transcribed the response as they perceived it. For pronunciation the experts listened to the students' responses in the test and transcribed the errors they heard. Comparisons between the feedback provided by the system and the teachers' annotations are now being conducted.

6.   Resources

6.1. Present resources

Within the framework of the DISCO project various resources have been developed. First of all a blue-print of the design and the speech technology modules for recognition (i.e. for selecting an utterance from the predicted list, and verifying the selected utterance) and for error detection (errors in pronunciation, morphology, and syntax). Furthermore, an inventory of errors at all these three levels, a prototype of the DISCO system with content, specifications for exercises and feedback strategies, and a list of predicted correct and incorrect utterances.

The fact that DISCO was carried out within the STEVIN programme implies that its results, all the resources mentioned above, become available for research and development through the Dutch Flemish Human Language Technology (HLT) Agency (TST-Centrale; www.inl.nl/tst-centrale). This makes it possible to reuse these resources for conducting research and for developing specific applications for ASR-based language learning.

6.2. Future resources

In addition to the resources that were developed during the DISCO project, there are resources that can be generated in the future by using the DISCO system.

First of all, an ASR-based CALL system like DISCO can be used for acquiring additional non-native speech data, for extending already existing corpora like JASMIN, or for creating new ones. This can be done within the framework of already ongoing research without necessarily having to start corpus collection projects. The advantage is that in DISCO all speech data become available with the relevant information for further processing such as alignments and confidence measures.

By way of illustration, the evaluation tests described above already provided new speech data that can be used to optimize the system.

Second, DISCO has been designed and developed in such a way that it is possible to log details regarding the interactions with the users. This logbook can contain, e.g., the following information: what appeared on the screen, how the user responded, how long the user waited, what was done (speak an utterance, move the mouse and click on an item, use the keyboard, etc.), the feedback provided by the system, how the user reacted on this feedback (listen to example (or not), try again, ask for additional, e.g. meta-linguistic, feedback, etc.). So when language learners use DISCO to practice oral skills all their utterances can be recorded in such a way that it is possible to know exactly in which context the utterance was spoken, i.e. it can be related to all the information in the logbook mentioned above.

Such a corpus and the corresponding log-files can be useful for various purposes: for research on language acquisition and second language learning, for studying the effect of various types of feedback, for research on

various aspects of man-machine interaction, and of course for developing new, improved CALL systems. For instance, we intend to use the additional data collected to optimize the DISCO system.

An ASR-based CALL system like DISCO also makes it possible to create research conditions that were hitherto impossible to create, thus opening up possibilities for new lines of research. For instance, at the moment a project is being carried out at the Radboud University of Nijmegen, which is aimed at studying the impact of corrective feedback on the acquisition of syntax in oral proficiency (http://lands.let.kun.nl/~strik/research/FASOP). Within this project the availability of an ASR-based CALL system makes it possible to study how corrective feedback on oral skills is processed on-line, whether it leads to uptake in the short term and to actual acquisition in the long term. This has several advantages compared to other studies that were necessarily limited to investigating interaction in the written modality: the learner's oral production can be assessed on line, corrective feedback can be provided immediately under near-optimal conditions, all interactions between learner and system can be logged so that data on input, output and feedback are readily available to be studied from different perspectives.

7. Conclusions

We have presented the research and the resources produced by the STEVIN DISCO project, which was aimed at developing an ASR-based system for Dutch L2 speaking practice and feedback. This presentation has made clear that DISCO provided us with useful resources for research and development. In addition, we can conclude that through its future use DISCO has the potential to efficiently generate new, valuable resources for innovative research and applications.

8. Acknowledgements

9. References

Bodnar, S., Cucchiarini, C. and Strik, H. (2011). Computer-assisted grammar practice for oral communication. In *Proceedings of the 3rd International Conference on Computer Supported Education (CSEDU)*, 2011.

Boersma, P. and Weenink D. (2009). Praat: doing phonetics by computer (Version 5.1.10) [Computer program]. Retrieved July 8, 2009, fromhttp://www.praat.org/.

Colpaert, J. (2010). Elicitation of language learners' personal goals as design concepts. *Innovation in Language Learning and Teaching,* 4 (3), pp. 259--274.

Cucchiarini, C., Neri, A. and Strik, H. (2009). Oral Proficiency Training in Dutch L2: the Contribution of ASR-based Corrective Feedback. *Speech Communication*, 51 (10), pp. 853--863.

Cucchiarini, C., Driesen, J., Van Hamme, H. and Sanders, E., (2008). Recording speech of children, non-natives and elderly people for HLT applications: the JASMIN-CGN corpus. In *Proceedings of LREC08*.

Demuynck, K., Roelens, J., Van Compernolle, D., and Wambacq, P., (2008) SPRAAK: An Open Source SPeech Recognition and Automatic Annotation Ki. In *Proceedings of Interspeech*, pp. 495--498.

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

Hubbard, P. (2002). Interactive Participatory Dramas for Language Learning. *Simulation and Gaming*, 33, pp. 210--216.

Kim, Y., Franco, H. and Neumeyer, L. (1997). Automatic pronunciation scoring of specific phone segments for language instruction. In *Proceedings of Eurospeech, Rhodes, Greece*, pp. 645--648.

Mak, B. Siu, M., Ng, M., Tam, Y.-C., Chan, Y.-C., Chan, K.-W. (2003). PLASER: pronunciation learning via automatic speech recognition. In *Proceedings of the HLT-NAACL 2003 Workshop on Building Educational Applications using Natural Language Processing, Edmonton, Canada*, pp. 2329.

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

Neri, A., Cucchiarini, C. and Strik, H. (2006). Selecting segmental errors in L2 Dutch for optimal pronunciation training. *International Review of Applied Linguistics*, (44) pp. 357-- 404.

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

Spyns, P., D'Halleweyn, E. & Cucchiarini, C., (2008). The Dutch-Flemish comprehensive approach to HLT stimulation and innovation: STEVIN, HLT Agency and beyond. In *Proceedings of LREC'08*, pp 1511--1517.

Strik, H., Cornillie, F., Colpaert, J., van Doremalen, J. and . Cucchiarini, C. (2009.) Developing a CALL System for Practicing Oral Proficiency: How to Design for Speech Technology, Pedagogy and Learners. Proceedings of the SLaTE-2009 workshop, Warwickshire, England.

Van Doremalen, J., Cucchiarini, C. and Strik, H. (2011). Automatic Speech Recognition in CALL systems: The essential role of adaptation. *Communications in Computer and Information Science*, Springer Verlag, 126, pp. 56--69.