interspeech 2008 incorporating SST 08

DISCO: Development and Integration of Speech technology into Courseware for language learning

Catia Cucchiarini, Joost van Doremalen, and Helmer Strik

Department of Linguistics, Radboud University Nijmegen, The Netherlands

[C.Cucchiarini | J.vanDoremalen | H.Strik]@let.ru.nl

Abstract

Recent research has shown that a properly designed ASR-based CALL system (Dutch-CAPT) was capable of detecting pronunciation errors and of providing comprehensible feedback on pronunciation. Since pronunciation is not the only skill required for speaking a second language, we explored the possibility of extending the Dutch-CAPT approach to other aspects of speaking proficiency like morphology and syntax. In this paper we explain how a number of errors in morphology and syntax that are common in spoken Dutch L2 could be addressed in an ASR-based CALL system. Finally, we present our new project in which corrective feedback will be provided on all three aspects of spoken proficiency: pronunciation, morphology and syntax.

Index Terms: pronunciation training, CALL, ASR, error detection.

1. Introduction

One-on-one interactive learning with corrective feedback (CF) is known to be optimal for language learners. The two sigma benefit demonstrated by Bloom [1] has provided further support for the advantages of one-on-one tutoring relative to classroom instruction. However, one-on-one tutoring by trained language instructors is costly and therefore not feasible for the majority of language learners. In the classroom, providing individual CF is not always possible, mainly due to lack of time. This particularly applies to oral proficiency, where CF has to be provided immediately after the utterance has been spoken, thus making it even more difficult to provide sufficient practice in the classroom.

The emergence of Computer Assisted Language Learning (CALL) systems that make use of Automatic Speech Recognition (ASR) seems to offer new perspectives for training oral proficiency. These systems can 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. However, existing CALL systems hardly begin to fulfill these requirements. We believe that this is due to the lack of proper detection of performance problems, coupled to feedback that is embedded in a realistic communicative setting and helps the learners to effectively improve their performance.

Recent research has shown that a properly designed ASR-based CALL system is capable of detecting pronunciation errors and of providing comprehensible CF on pronunciation [2]. This system, called Dutch-CAPT, was designed to provide CF on a selected number of speech sounds that had appeared to be problematic for learners of Dutch from various first language (L1) backgrounds [3]. The results showed that for the experimental group that had been using the CALL system for four weeks the reduction in the pronunciation errors addressed in the training system was significantly larger than in the control group [2].

These results are promising and show that it is possible to use speech technology in CALL applications to improve speaking proficiency. In the Netherlands speaking proficiency plays an important role within the framework of civic integration examinations. Foreigners who wish to acquire Dutch citizenship have to show that they are able to get by in Dutch society and that they speak the Dutch language at the Common European Framework (CEF) A2 level, which means that they can make themselves understood in Dutch and that others understand what they say. For instance, they must be able to pay for their purchases in the supermarket or buy a train ticket.

However, pronunciation is only one of the skills required for speaking a second language. There are also other aspects of spoken language that are important and that have to be mastered in order to be comprehensible and proficient in a second language. For instance, morphology and syntax also play an important role in language comprehension and language learning. It is known that learners tend to make different morphologic and syntactic mistakes when they speak than when they write. It is generally acknowledged in the second language (L2) literature 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 in a second language, L2 learners need to practice speaking and need to receive CF on their performance on line, not only on pronunciation, but also on morphology and syntax.

A CALL system that is able to detect errors in speaking performance, point them out to the learners and give them the opportunity to try again until they manage to produce the correct form would be very useful because in L2 classes there is not enough time for this type of practice and feedback. We therefore decided to explore the possibility of extending the approach adopted in Dutch-CAPT to other aspects of speaking proficiency like morphology and syntax, and the results are presented in this paper. It turned out that there are a number of errors in morphology and syntax that are common in spoken Dutch L2 and that could be addressed in an ASR-based CALL system. In this paper we first describe these errors, then we explain how these problematic aspects could be addressed in a CALL system and finally we present our new project in which CF will be provided on all three aspects of spoken proficiency: pronunciation, morphology and syntax.

Accepted after peer review of full paper

2. Morphological and syntactic errors in spoken Dutch L2

2.1. Morphological errors in spoken Dutch L2

Problems with morphology are persistent in L2 learning [4] and phonetic-phonological properties play a prominent role in this learning process. As stated in [4]: ``The meaning of morphemes and the distribution of their allomorphs cannot be acquired without the phonological capacity to extricate them from the flood of sounds in every sentence". To develop this capacity learners first have to notice the contrast between their own erroneous realization (output) and the target form (input), as explained in Schmidt's Noticing Hypothesis [5]. Difficulties in learning Dutch verbal morphology are related to perception and production of L2 phonemes such as schwa and /t/. As to perception, it is crucial to perceive the differences in (1) in order to understand the Dutch agreement paradigm, and in (2) in order to understand the tense system (present vs. past tense).

(1) /maak/, /maakt/, /make(n)/

(2) /maakt/, /maakte/

On the production side, difficulties in pronouncing certain sound combinations may lead a Moroccan learner to say (3) when trying to pronounce /loopt/.

(3) /lopet/, /loopte/

2.2. Syntactic errors in spoken Dutch L2

In syntax problems have been observed with word order, finite verb position, and pronominal subject omission. Owing to L1 transfer, Turkish learners are known to produce sentence-final verbs as in (4) instead of the correct (5).

(4) * *Jong mandarijn sneeuwman neus maakte.*
Boy tangerine snowman nose made.
(intended form is: 'maakt')

(5) *De jongen maakt met een mandarijn de neus*
The boy makes with a tangerine the nose
van de sneeuwman.
of the snowman.

A second difficult but basic syntactic phenomenon to acquire is the obligatory presence of the subject in Dutch. Pronominal subject omission (or subject pro-drop) is allowed in the L1 of many learners of Dutch and is frequently produced in early L2 developmental stages, as in (6a) and (6b). The subject in sentence-final position (6c) is another manifestation of the same pro-drop phenomenon. The correct form is given in (7).

(6)
a. * *loop(t) naar huis* (typically Moroccan)
walk(s) home
b. * *naar huis lopen* (typically Turkish)
home walk
c. * *loopt naar huis de jongen*
walks home the boy

(7) *de jongen loopt naar huis*
the boy walks home

Another syntactic phenomenon known to be problematic for learners of Dutch L2 is Verb Second following an adverbial adjunct. Dutch is a verb-second language that requires subject inversion following an adverbial in initial position, as in (8b), but many learners construct an SVO clause, as in (8a).

(8)
a. * *dan hij gaat tv kijken*
then he goes tv watch
b. *dan gaat hij tv kijken*
then goes he tv watch

3. Extending Dutch-CAPT to morphology and syntax

It is well-known that recognition of non-native speech is problematic. In the Dutch-CAPT system recognition of the utterances was successful because we severely restricted the exercises and thus the possible answers of the learners. Confidence measures were then used to determine which of the utterances was spoken. In order to extend ASR-based feedback to morphology and syntax it is necessary to design exercises that are appropriate for practicing these aspects of spoken proficiency on the one hand, but that are controlled enough to be handled by ASR. For pronunciation it is possible to use imitation and reading exercises and these can be handled by ASR because the vocabulary is known in advance. For morphology and syntax such exercises cannot be used because learners then have no freedom to show whether they are able to produce correct forms. So, the exercises that are required have to be such that they allow some freedom to the learners in formulating answers, but that are predictable enough to be handled by ASR. To this end we went on to explore whether it would be possible to design exercises that comply with these requirements. We found that suitable exercises can be designed by stimulating students to produce utterances containing the required morphological and syntactic forms by showing them words on the screen, without declensions, or pictograms, possibly in combination with figures representing scenes (e.g. a girl reading a book). In addition, as in Dutch-CAPT, use can be made of dialogues and scenarios illustrating so-called "crucial practice situations" (in Dutch cruciale praktijksituaties or CPS), which correspond to realistic situations in which learners might find themselves in Dutch society and in which they have to interact with other citizens. These CPSs form the basis of the various civic integration examinations. The students can be asked to play a certain dialogue by using simple prompts concerning the vocabulary to be used and they have to formulate the correct sentences themselves.

In these exercises realistic communicative situations can be presented and the learners have the opportunity of performing realistic tasks. They receive prompts as to the words they have to use, so that vocabulary can be anticipated for ASR, but the learners have to produce the grammatically correct forms themselves, so that morphology or syntax can be tested and practiced. For morphology: a picture is shown on the screen of a person performing a certain task/action, the student receives prompts as to the words (i.e. verbs in infinitive form) to be used and he/she has to speak a complete sentence with the correct forms of verbs and nouns. Such an exercise can also be used for syntax to check whether the pronominal subject is being used appropriately or whether words are being used in the right order. For the latter aspect, cloze exercises can be designed in which an incomplete

utterance is shown on the screen; one word is missing, that word is displayed somewhere else on the screen, and the learner has to speak up the complete utterance with the word inserted on the right place.

3.1. Error detection for morphology and syntax

For detecting morphological and syntactic errors, response expansion software can be used. This software takes appropriate responses as input and expands them to form pools of correct and incorrect responses. The software is based on modules for sentence and word expansion, which have been developed by Polderland: the Polderland lemmatizer, the Polderland Part-of-Speech tagger and Lexpand, a product for morphologic expansion of lemma's to all possible word forms, and KLiP Thesaurus, a product for semantic expansion of tokens resulting from another STEVIN project "Rechtsorde" .

3.1.1. Detecting syntactic errors

For detecting syntactic errors it is sufficient to know which words were spoken in which order. The speech recognition module determines which utterance was spoken, the exercises database contains the syntactic errors for the utterance (generated by response expansion module, e.g. pronominal subject omission, incorrect word order etc.). Depending on which of the possible utterances has been recognized, the system can determine whether errors have been made with respect to e.g. word order and/or pronominal subject omission.

3.1.2. Detecting morphological errors

For detecting morphological errors, the system should be able to distinguish, e.g., /maak/, /maken/, /maakte/ and /maakt/. Some of these variants are included in the list of possible responses, i.e. the ones related to frequent errors, which can be detected with sufficient reliability by means of confidence measures at utterance - or word – level, and for which inclusion improves the performance of the speech recognition module. This already provides information on some morphological errors. However, our previous research made clear that for many of these pronunciation related errors a more detailed analysis at segmental level is needed.

To this end, an automatic segmentation at phone level is made [2], [6], followed by a calculation of confidence measures for the individual phones. Criteria similar to those described in [2] can be used to select the phones / errors to be addressed. In short, the focus has to be on errors that are frequent, salient, persistent and that can be detected with sufficient reliability. In the Dutch-CAPT system we have employed the Goodness-Of-Pronunciation (GOP) score [7]. A GOP score is a log-likelihood ratio that can be calculated with the same algorithm for every phone, and this score then has to be compared with a phone specific threshold to determine whether the pronunciation was correct or not [2], [7]. We also experimented with acoustic-phonetic classifiers for error detection [8]. In [8] we compared the two techniques and found that the performance of the acoustic-phonetic classifiers was better. We now intend to study what works best: GOP, acoustic-phonetic classifiers, or a combination of the two. Since classifiers have been developed for only a small number of phonetic contrasts, additional classifiers need to be developed for those phonetic contrasts that are relevant in this context.

4. Development and Integration of Speech technology into COurseware for language learning (DISCO)

The idea of extending the Dutch-CAPT approach to morphology and syntax by using the exercises and the detection techniques described above was elaborated in a research proposal named DISCO, which was eventually financed within the framework of the Dutch Flemish stimulation programme for HLT called STEVIN. The aim of the DISCO project is to develop a prototype of an ASR-based CALL application for Dutch as a second language (DL2). The application optimizes learning through interaction in realistic communication situations and provides intelligent feedback on important aspects of DL2 speaking, viz. pronunciation, morphology, and syntax. The application should be able to detect and give feedback on errors that are made by learners of Dutch as a second language.

With respect to pronunciation, we aim at the achievement of intelligibility, rather than accent-free pronunciation. As a consequence, the system will target primarily those aspects that appear to be most problematic. In previous research [3] we have gathered relevant information in this respect. The pronunciation exercises will address the sounds that were trained in [2] and some additional problematic sounds.

4.1. Design

A general framework for implementing and testing communicative CALL exercises is being developed. The client-server architecture integrates an ASR module, and several modules for further processing of the ASR output in an environment in which media content can be re-used to develop exercises. The system also supports a simple mechanism for the generation of feedback and it comes with a tool that supports the implementation of new exercises on the basis of existing media content. As in Dutch-CAPT use will be made of media content from the Nieuwe Buren program [2], which will be adapted to suit the aims of DISCO. In DISCO the ASR module will be based on SPRAAK, the result of another STEVIN project.

All courseware is stored in a database. It consists of the course structure, course material to be presented to the user (consisting of moving images, pictures, texts and sounds), and exercise details: content, expected responses, and feedback information. Tools are provided to fill the courseware database and automatically expand expected responses. User performance and progress information are stored in a second database.

The courseware application is realized as a client/server application which enables realization as a stand-alone as well as a web-based version. All logic functionality is located in the server; the "thin" client contains GUI representation software and user-server communication functions.

The server contains a module to handle interaction with the client. A second module, the course and exercise logic handling module, guides the user through the course and presents course material – including exercises – to the user. It collects user responses to exercises, has them processed, and tracks user progress.

User responses to exercises are forwarded to the speech recognition module, which uses the courseware and exercise database to check for matches with expected correct or incorrect responses. When a response has been identified, the diagnostic modules are activated to validate the speech

realization quality and the morpho-syntactic quality of the user's response. Depending on the results of the validation, a proper feedback form is selected and passed on to the course and exercise logic handling module. Following an update to the user performance database, feedback is forwarded to the client. When the speech recognition software fails to identify one of the expected responses, an appropriate message is passed on to the user, and the user is asked to retry.

4.2. Feedback

Feedback is provided on two levels: (1) on the utterance level, and (2) on the error level. Regarding the former, the speech recognition module determines which utterance was spoken, and before proceeding to the error detection module the learner is given feedback on the recognized utterance. After all, it would be highly confusing if the learner gets feedback on (parts of) an utterance that was not spoken at all by the learner. Only after the learner has indicated that the utterance has been recognized correctly, does the system proceed to error detection. On the other hand, if the utterance cannot be recognized, the learner will get a message that the system cannot process the utterance.

For providing CF we adopt a user interface which is based on the one developed for Dutch-CAPT, which is extended to provide CF on morphological and syntactic errors. The exact form of feedback on this latter type of errors will be chosen on the basis of pilot experiments in which different formats will be tested.

In the preliminary research we carried out while preparing the research proposal, a limited number of experienced teachers were asked to indicate how they provide feedback in specific situations. One method that appears to be very effective for providing feedback on syntax concerns the use of gestures that refer to specific syntactic errors. In the DISCO project the effectiveness of this type of CF will be tested by using pictograms that refer to such gestures. In addition, more experienced teachers will be asked to indicate how feedback could best be provided in specific situations. On the basis of their input rules will be defined and implemented in the system. Feedback can consist of textual and graphical information rendered on the screen. For example, if a morpho-syntactic error is detected, DISCO will display the correct form with the errors highlighted. In all cases the student will eventually have the opportunity to listen to a correct version of the response.

For each utterance, feedback will be provided on a limited number of errors, for instance, maximally three or four, and these errors will be selected on a number of selection criteria. In any case feedback will be provided only on those errors that can be detected with an acceptable degree of reliability. In this respect it is important to mention that in this project we will follow the approach adopted in Dutch-CAPT with respect to false detections. As is well known, there is a trade-off between false accepts (FAs, accepting an error as correct) and false rejects (FRs, rejecting something that was actually correct). In Dutch-CAPT we decided to minimize FRs and tolerate some FAs on the grounds that for learners erroneously rejecting correct realizations would be more detrimental than erroneously accepting incorrect ones. This will enable learners to concentrate only on the most serious errors and to gain self-confidence, while minimizing the number of times an error is incorrectly given feedback on.

4.3. Evaluation

Evaluation will take place at several times and at several levels. Four pilot experiments will be carried out which are aimed at testing the exercises, the speech recognition module, the error detection module, and the whole system, respectively. The latter is a preparation of the final evaluation of the whole system.

A system that gives meaningful feedback must operate in a manner that is similar to what a competent teacher would do. Therefore, for the final evaluation of the whole system we propose a design in which different groups of students of DL2 use the system and fill in a questionnaire with which we can measure the students' satisfaction in working with the system. Teachers of DL2 will then assess all sets of system prompt, student response and system feedback for the quality of the feedback on the level of pronunciation, morphology and syntax. For this purpose, recordings will be made of students who complete the exercises developed to test the DISCO system.

Given the evaluation design sketched above, we consider the project successful from a scientific point of view if the DL2 teachers agree that the system behaves in a way that makes it as useful for the students as a teacher is, and if the students rate the system positively on its most important aspects. From a valorization point of view we consider the project successful if the results of this project are taken up to develop applications.

5. Acknowledgements

Partners in the DISCO project are J. Colpaert (Linguapolis, University of Antwerp), J. Bakx (Universitair Taal- en Communicatiecentrum Nijmegen), and I. de Mönnink (Polderland Language & Speech Technology). The DISCO project is carried out within the STEVIN programme which is funded by the Dutch and Flemish Governments (http://taalunieversum.org/taal/technologie/stevin/).

6. References

[1] Bloom, B. S. "The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring". Educational Researcher, 13, 4-16, 1984.

[2] Neri, A. Cucchiarini, C. and Strik, H. "The effectiveness of computer-based corrective feedback for improving segmental quality in L2-Dutch", ReCALL, Vol 20, No. 2, May 2008.

[3] Neri, A., Cucchiarini, C. and Strik, H. "Selecting segmental errors in non-native Dutch for optimal pronunciation training", International Review of Applied Linguistics, 44, 2006.

[4] DeKeyser, R. "What Makes Learning Second-Language Grammar Difficult? A Review of Issues", Language Learning, 55, S1, 1-25, 2005.

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

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

[7] Witt, S.M. & Young, S. "Phone-level Pronunciation Scoring and Assessment for Interactive Language Learning", Speech Communication, 30(2), 95-108, 2000.

[8] Strik, H., Truong, K., de Wet, F. and Cucchiarini, C. "Comparing classifiers for pronunciation error detection", Proceedings of Interspeech-2007, Antwerp, Belgium, 1837-1840, 2007.

[9] Cucchiarini, C., Neri, A. de Wet, F. and Strik, H. "ASR-based pronunciation training: scoring accuracy and pedagogical effectiveness of a system for Dutch L2 learners", Proceedings Interspeech 2007, Antwerp, Belgium, 2007.