

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/101719>

Please be advised that this information was generated on 2017-10-20 and may be subject to change.

ASR-based systems for language learning and therapy

Helmer Strik

Centre for Language and Speech Technology (CLST)
Radboud University Nijmegen, the Netherlands

Abstract — ASR-based CALL seems to offer many possibilities for language learning and therapy. However, in both domains the speech of the users generally differs substantially from standard speech. ASR of such atypical speech is complex and challenging. Furthermore, developing successful CALL systems requires a mix of expertise. This combination of factors has led to misconceptions and pessimism on the use of speech technology in CALL. In the current paper, we provide an overview of our research in this area, which shows that speech technology can be applied in developing useful CALL systems.

Keywords - ASR-based CALL, language learning, therapy, atypical speech

I. INTRODUCTION

Computer Assisted Language Learning is a relatively young discipline that has already produced a considerable body of research and applications, which is attested by the numerous dedicated journals, conferences, workshops, proceedings, applications and commercial products. A general observation about this impressive output is that, overall, there are relatively fewer applications and products that address language learners' production in speaking and writing. Very often CALL systems offer practice and testing in such skills indirectly, for instance by asking learners to check different sentences and indicate which one is correct. This is obviously related to the complexity of processing the learners' output, i.e. the input to the CALL system. While it is fairly easy to process input from a mouse or touch screen, e.g. clicks and drag-and-drop, or restricted text, e.g. typed with a keyboard, which usually contains short utterances to be compared to lists of possible answers, processing unrestricted nonnative text or nonnative speech is far more complex.

However, research in second language (L2) learning has indicated the importance of skill-specific output, practice and feedback [11; 23] for learning to speak and write in the L2. In other words, learners should get the chance to extensively practice speaking and writing to try to at least approximate near-native performance. Considering that the majority of CALL systems are usually employed as supplements to traditional classroom-based teaching, it seems that especially CALL systems that address speaking proficiency would offer added value with respect to teacher-fronted classes. Because of its on-line nature, speaking practice is relatively demanding in terms of teacher time: teachers need to listen to individual learners, interact with them and provide individual feedback synchronously. For writing, on the other hand, practice can take place off line, without the teacher being present and feedback can be provided asynchronously. Against this background it is understandable that researchers have been

looking for ways of providing speaking practice in CALL within the limitations of the available technology [17].

In many CALL systems that address speaking proficiency learners are encouraged to speak, but their speech output is not further processed. The rationale behind this approach is that for L2 learners producing spoken output in the L2 is in itself a worthwhile activity. Although this is probably true, as is supported by research on the importance of language output [38; 9] language learners in general prefer to check whether the speech they produced was correct or not. To make this possible CALL systems have been developed in which learners are asked to imitate examples played by the system and are invited to compare their own production with the example. This latter approach can be useful to a certain extent, but self-assessment has its limitations [13], if only because learners have difficulties in perceiving certain target language contrasts [19].

For this reason researchers and commercial companies have tried to produce CALL systems that in different ways provide some form of feedback on L2 learners' speech production. Some of the earlier systems provided visual feedback in the form of intonation contours, waveforms and spectrograms, as these could be easily achieved by employing a speech analyzer. Such feedback is still used in current systems, in spite of the fact that it remains questionable whether it is useful and effective [27; 28]. An important drawback of CALL systems that do not make use of automatic speech recognition (ASR) technology is that it cannot be verified whether the learners indeed produced the intended, target utterance. In other words, if the learner says something different from the prompted utterance, the system is not able to check that and will provide feedback on the learner's production as if it was indeed the target utterance. This may affect the credibility of the system and its pedagogical value. This is a complaint often heard from users of such systems. Therefore, researchers have been exploring how speech technology could be employed to the benefit of language learning, and in particular, speaking proficiency training.

ASR-based CALL not only offers many possibilities for language learning, but also for therapy, for people with so-called communicative disabilities. However, in both domains the speech of the users generally differs substantially from standard speech. ASR of such atypical speech is complex and challenging. Furthermore, developing successful CALL systems requires a mix of expertise. This combination of factors has led to misconceptions and pessimism on the use of speech technology in CALL (see section III).

In the current paper, we focus on ASR-based CALL. In the sections below we first present a brief history (section II), then look at ASR of non-native speech (section III), present an overview of our research in this field (section IV), and we end with a discussion (section V).

In the current paper we mainly use the term ASR-based CALL, as many others do. However, it should be noted that the methods (techniques, algorithms, tools, etc.) used 'for speech' in CALL are not restricted to ASR in the narrow sense, and that ASR is used more as a term covering speech technology in general including phonetically-based methods.

II. HISTORY OF ASR-BASED CALL

ASR-based CALL received increasing attention in the late nineties. At the CALICO conference in 1996 the CALICO - EUROCALL 'Special Interest Group' (SIG) 'Computer Assisted Pronunciation Investigation Teaching And Learning' (CAPITAL) was started, which in 1999 became the CALICO - EUROCALL - ISCA SIG 'Integration of Speech Technology in (Language) Learning' (InSTIL). Furthermore, in 1998 the 'Speech Technology in Language Learning' (STiLL) workshop was organized in Marholmen (Sweden) [www.speech.kth.se/still/]. This was the starting point of a number of STiLL and InSTiL related activities.

In 1999 a special issue of CALICO appeared, entitled 'Tutors that Listen', which focused on ASR [21]. It concerned mainly so-called 'discrete ASR', i.e. the recognition of individual words that are uttered with pauses between the words. Obviously, this is not the preferred way of communicating when learning a language. Therefore attention shifted towards continuous speech. InSTIL organized an 18 poster exhibition called 'An Illustrated History of Speech Technology in Language Learning', which was shown at EUROCALL 2001 in Nijmegen (Netherlands) and EUROSPEECH 2001 in Aalborg (Denmark). For more information on the history in this field see e.g. Delcloque [10] and Eskenazi [18].

At the Interspeech 2006 conference of the International Speech Communication Association (ISCA) in Pittsburgh there was a special session on 'Speech and Language in Education'. This was the starting point of the ISCA SIG on 'Speech and Language Technology in Education' (SLaTE) [www.sigslate.org]. SLaTE has organized several workshops since then. In addition, at the 'Innovative Use of NLP for Building Educational Applications' (BEA) workshops of the 'Association for Computational Linguistics' (ACL), now (i.e. 2012) in its seventh edition, the role of speech technology has gradually increased [www.cs.rochester.edu/~tetreaul/academic.html]. The workshops and other activities of the SIGs mentioned above (e.g. conference special sessions), have led to many publications, see e.g. the proceedings of these events.

Speech recognition technology also gradually found its way into commercial CALL systems by companies. Well-known are 'Tell me More' by Auralog [www.tellmemore.com], 'Rosetta Stone' [www.rosettastone.com], and 'IntelliSpeech' by 'digital publishing' [www.digitalpublishing.de].

III. ASR OF NON-NATIVE SPEECH

As the quality of speech technology improved, more and more researchers tried to apply it to language learning, sometimes with disappointing results. Some researchers were skeptical about the usefulness and effectiveness of ASR-based CALL programs: evidence gathered in different lines of research seemed to confirm that either speech technology was

not mature enough, or ASR-based CALL programs were not effective in improving second language (L2) skills [e.g., 3; 12]. For the sake of our own research, we studied this literature thoroughly and gradually acquired the impression that, while it is undeniable that speech technology still presents a number of limitations, especially when applied to non-native speech, part of this pessimism is in fact due to misconceptions about this technology and CALL in general.

For instance, in some studies unsatisfactory results were obtained when standard dictation systems were used for CALL [3; 12]. But such dictation systems are not suitable for L2 training, as CALL requires dedicated speech technology. Apart from the fact that the majority of dictation packages are developed for native speakers, the major problem in using this technology for CALL has to do with the different goals of dictation and CALL which require different approaches in ASR. The aim of a dictation package is to convert an acoustic signal into a string of words and not to identify L2 errors, which requires a different, more complex procedure. Consequently, the negative conclusions related to the use of dictation packages should be related to those specific cases and not to ASR technology in general.

ASR of native speech is already complex because of many well-known problems such as background sounds, (low) signal-to-noise ratio (SNR), end-point detection, pronunciation variation, and disfluencies. However, ASR of atypical speech is even more complex, since the grammar, the words used, and the pronunciation can deviate considerably, thus affecting all three 'knowledge sources' of the ASR system (language model, lexicon, and acoustic models, resp.). In the ASR community, it has long been known that the differences between native and non-native speech are so extensive as to degrade ASR performance considerably [14; 16; 22; 26; 39; 40]. Furthermore, native and non-native speech can differ in many (sometimes unexpected) ways, e.g. for non-native we found more broken words for (cold) reading, and many more filled pauses in spontaneous speech [4].

IV. ASR-BASED CALL RESEARCH

The current section presents a brief overview of our research in this field to give an idea of what can be achieved with current technology. In 1997, we started the project 'Automatic testing of oral proficiency' [41] in which we aimed at developing a system for automatic assessment of foreign speakers' oral proficiency in Dutch by using ASR technology. The results showed that automatic testing of certain aspects of oral proficiency was feasible: the scores obtained by means of ASR technology were strongly correlated with human judgments of oral proficiency. Especially oral fluency appeared to be easily predictable based on automatically calculated temporal measures [7; 8].

Pronunciation grading, as in the ATOP project, is used to calculate a score at a rather global level (e.g. for a couple of utterances), which might be sufficient for testing purposes, but in general it is not detailed enough for training purposes. For training error detection is required, that is the procedure by which a score at a local (e.g. phoneme) level is calculated. For grading more global measures can be used, such as temporal measures [7; 8]. In general, the relation between human and automatic grading improves if longer stretches of speech are

used, i.e. complete utterances or a couple of utterances [see e.g., 20]. Such cumulative measures can also be adopted for error detection, for instance by combining the scores of several utterances. This can be useful to assess the problems of a specific speaker, to obtain an overview and suggest remedial exercises for the problematic cases. However, for remedial exercises immediate feedback based on local calculations is to be preferred. For pronunciation error detection, some approaches can be used: (1) focus on frequent errors, (2) use ASR-based metrics or (3) acoustic phonetic classifiers.

In the first approach, errors frequently made by language learners are explicitly taken into account [25]. For instance, in DL2, if the sound /h/ is often deleted (e.g. 'elmer' instead of 'helmer'), /g/ is often pronounced as /k/, and long and short vowels are interchanged, then these frequent errors can be included in the pronunciation models. The ASR then has to find the best path in these pronunciation networks, and can thus determine whether a pronunciation error was made.

In the second approach, ASR-based metrics are used, such as posterior probabilities and (log) likelihood ratios [20; 22; 26]. Research has shown that these confidence measures can be used for detecting pronunciation errors [20; 22; 26; 40]. A special case concerns the so-called goodness of pronunciation (GOP) algorithm [40], which has been used in several studies. We have conducted detailed studies of the GOP algorithm [15; 24; 36; 37]. If properly trained, the GOP algorithm works satisfactorily; e.g. in Dutch-CAPT system (see below) 80-95% of the sounds were classified correctly. However, there are large variations between individuals and sounds. If specific settings (thresholds) could be used for each person sound combination, better results could be achieved [24]; but in practice this is not possible. And since the GOP algorithm has some other limitations, we have been studying possible alternative measures [see e.g., 15].

The third approach, based on acoustic phonetic classifiers, is not often used in CALL applications; still it can be useful [36; 37]. We compared the results of acoustic phonetic classifiers to those obtained with the GOP algorithm, and found that results for acoustic phonetic classifiers were generally better [36; 37]. As can be expected, a combination of approaches probably yields the best results. Therefore, the challenge here is to find the proper combination of approaches and settings to achieve the best results.

Most approaches, such as the often applied (supervised) machine learning approach, require large amounts of annotated data in order to train the classifiers. Since obtaining annotated data is laborious, we have been studying other ways to carry out pronunciation detection. The acoustic-phonetic approach mentioned above is already a first step in that direction. Another approach we studied, is to use artificial errors [24]. We first obtained an overview of frequently made errors, then artificially introduced these errors into native training material, used this material to train error detectors, which were subsequently employed in the Dutch-CAPT system. Language learners then used the Dutch-CAPT system, their interactions were recorded and annotated afterwards. Analysis of these annotations showed that the performance of these error detectors in real use was comparable to the performance during development. This is remarkable, since with speech technology performance during real use is often lower than during

development, and this is especially the case when there is a training-test mismatch, which was the case here (training: artificial errors in native speech, test: real errors in non-native speech). Probably, this is because we carefully introduced artificial errors that were based on analyses of actually occurring errors [24].

In the 'Dutch Computer-Assisted Pronunciation Training' (Dutch-CAPT) project [42] a pronunciation training program was developed to provide automatic feedback on segmental (phoneme) errors (see Figure 1). We evaluated this system by comparing production data by an experimental group that used the Dutch-CAPT system, with those of a control group that did similar exercises but did not get feedback on pronunciation errors. The learners in the two groups had been living in the Netherlands and had followed DL2 lessons. Already after two short sessions of about 30-60 minutes, we could observe that the decrease in the number of pronunciation errors was substantially larger for the experimental group compared to the control group that did not receive feedback) [6; 30].

Before developing a CALL system, we generally try to obtain an overview of frequent errors made by language learners by combining information found in the literature, expertise of language teachers, and analysis of data. Even if the artificial error procedure described above is used, such an overview is essential to carefully introduce the right errors in the right way. We have already derived overviews of frequent segmental errors for different combinations of first (L1) and target (L2) languages: many L1s - Dutch [27; 29], Spanish - Dutch [2], Dutch - English [5]; and also for grammatical errors in DL2 [34; 35], and segmental errors in dysarthric speech (see PEDDS project below). Deriving information on segmental errors from data was achieved through well-known procedures, while to derive information on grammatical errors from data we developed a novel procedure [34; 35].

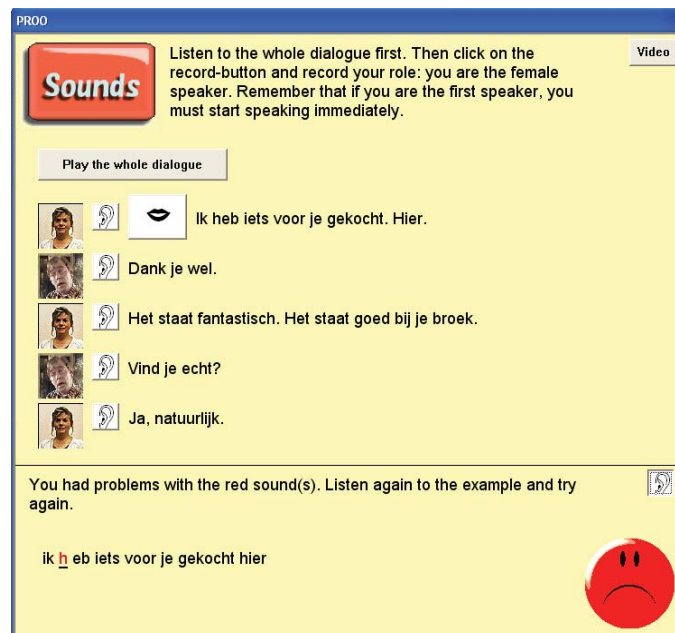


Figure 1. Screenshot of the Dutch-CAPT system. The user first watches a video, then plays a role, and gets feedback on pronunciation errors.



Figure 2. A screenshot of the DISCO system. The user can choose an interlocutor ('spraakmakker' – 'speech buddy') to speak to. The topics vary: a train journey, choosing a course, and going to the shop with a broken DVD player, respectively.

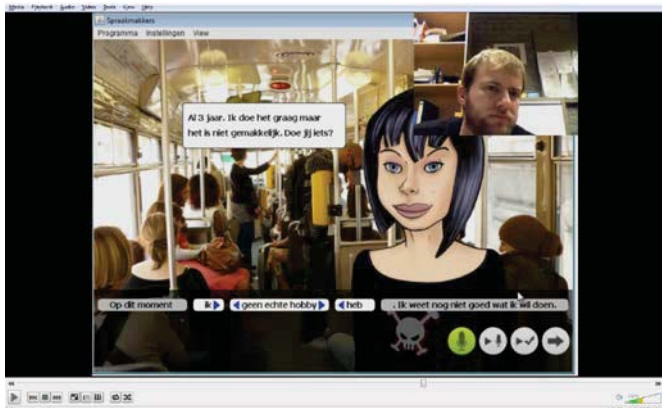


Figure 3. A screenshot of the DISCO system with a user in the upper-right corner. It concerns a syntax exercise: the user has to speak the words in the correct order.

Since good results were obtained with the Dutch-CAPT system on pronunciation, we decided to go develop a system for training not only pronunciation, but also grammar (morphology and syntax) in spoken language. To this end, we employed the overviews of pronunciation and grammatical errors mentioned above. This work was carried out in the 'Development and Integration of Speech technology into Courseware for language learning' (DISCO) project (see Figures 2 and 3), which is now almost finished [32; 33; 43]. The first user tests are encouraging, students are very positive about the system and additional evaluations will be performed soon.

In our research we develop CALL systems. In turn, we also use these CALL systems to carry out research, and the results of this research can in turn be used to improve CALL systems, and the way they are employed. We thus hope to create an upward spiral. For instance, an important issue in CALL systems for training oral proficiency is how to provide feedback. This issue is studied in the project 'Feedback and the Acquisition of Syntax in Oral Proficiency' (FASOP) [44], in

which a modified version of the DISCO system is employed to conduct experiments on oral syntax practice and acquisition (see Figure 4). Dutch L2 learners are pre-tested before undergoing specific training in L2 syntax through different versions of the CALL system that provide different forms of feedback. Post-tests are then administered to determine the effects of the feedback (see Figure 5). The first results are encouraging [1].

Besides research on ASR-based CALL systems for DL2, we recently started a project on English pronunciation training for Dutch learners ('My Pronunciation Coach', MPC) [45], and the 'Lifelong Learning Programme' (LLP) project 'Games Online for Basic Language learning' (GOBL) [46] in which mini-games for language learning will be developed.



Figure 4. A screenshot of the FASOP system. Learners first watch a video clip and then answer questions. In this example, the tutor is asking 'What does it say on the box that Melvin has packed his things in?'. To answer, learners compose an utterance using the prompt and word groups presented on the screen. All ('Allemaal') the word groups in the blue box have to be used, and only one ('Eentje') from the box in green.

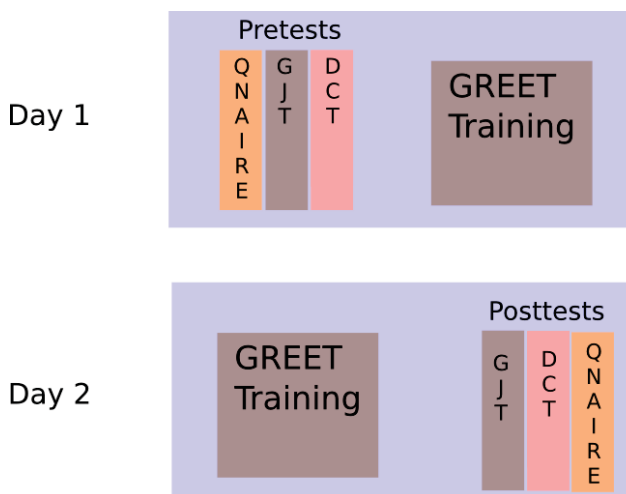


Figure 5. Overview of the FASOP experiment: 'QNAIRE' – questionnaire, GJT – grammatical judgment test, DCT – discourse completion test.

As explained above, non-native speech deviates from native speech in different respects. Another type of atypical speech is that produced by people with communicative disabilities (a 'speech handicap'). Similar techniques as those used for language learning can be applied in this clinical setting. For instance, in a pilot study we studied ASR of dysarthric speech. Dysarthria exists in different forms and can vary from mild to severe. If one is not familiar with the specific kind of dysarthric speech, it is usually difficult to understand the speaker in question. In our pilot study it was shown that also for dysarthric speech the performance of ASR degrades, but can be substantially improved by optimizing the ASR system for dysarthric speech [31]. The challenge here is to capture the patterns for this type of atypical speech in the models of the ASR system. This becomes more problematic if the speech (and its patterns) is not constant, e.g. in the case of progressive dysarthria. In any case, it is advisable (esp. in clinical applications, but probably also in CALL applications) to regularly update the models of the ASR system.

In the 'Pronunciation Error Detection for Dysarthric Speech' (PEDDS) project (see Figure 6), we developed technology for detecting pronunciation errors in dysarthric speech [47]. We also made a video demo to show what the possibilities are of using such technology for pronunciation training [47]. In this demo the user first watches a video (in this case an old news broadcast), then produces some utterances, gets immediate feedback on the pronunciation errors made, optionally can listen to the example utterances, and can try to pronounce the utterances again.

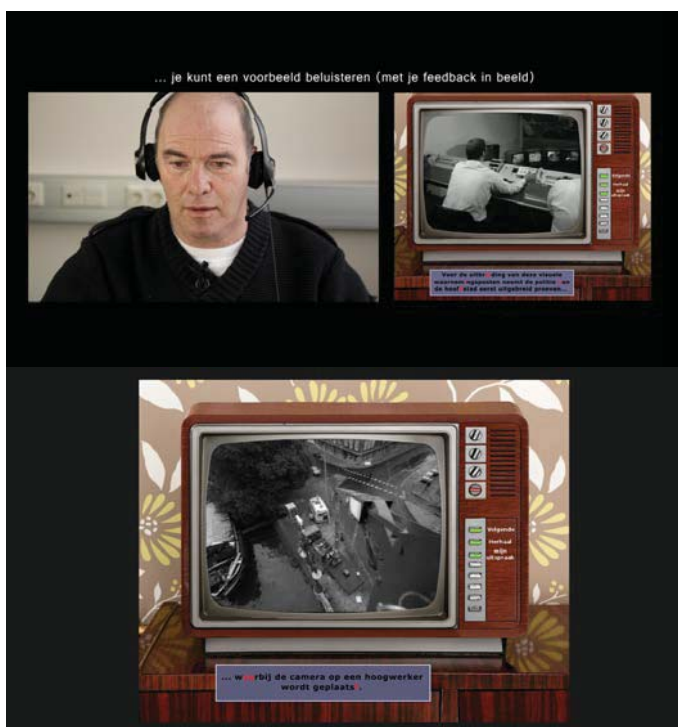


Figure 6. Two screenshots of the PEDDS system. The user speaks utterances and gets immediate feedback on errors in its dysarthric speech.

Finally, in the 'Communication & Revalidation DigiPoli' (ComPoli) project we are developing technology to assist users in communicating with e-Health websites [48]. Nowadays, more and more people have to use websites (so called 'digipolies'), to look for information, communicate with other patients and/or experts, make appointments, etc. However, for people with communicative disabilities, this can cause problems. We will use different technologies (such as ASR, text-to-speech synthesis, and word prediction) to enhance their possibilities of communication with these websites. A first version of the website is finished, and soon we will start user trials.

V. DISCUSSION

Above we already mentioned some reasons why developing high quality applications for atypical speech is complex, and therefore challenging; some additional issues are briefly discussed here.

To develop sound ASR-based CALL systems a mix of expertise is needed, expertise on technology for atypical speech, but also on, e.g., language acquisition, language learning, pedagogy, language course and software design, when it concerns foreign or second language learning; and similar expertise for clinical applications. In such applications eliciting speech is also challenging. It should be done in a way that is does not feel unnatural, is motivating, and, of course, effective. However, since automatic handling of spontaneous speech is not feasible yet, it should also be constrained, and the technology should be optimized for the (constrained) target speech in such a way that the system works properly. The challenge here is to develop the appropriate algorithms, and optimize them while finding the right balance between all these, often conflicting, preconditions.

Evaluation can be carried out in different ways. It is possible to evaluate the individual system components, off-line, using suitable data (speech corpora). This is generally done during development of the system. A problem is that often large amounts of suitable training material are not available. This is especially the case for detection of less frequent errors. If the interactions of the users with the system are recorded, and annotated afterwards, the same system components can also be evaluated in a more realistic context, i.e. during real use. Another possibility is to ask the system users to fill in questionnaires. More challenging is to evaluate whether the system is effective, e.g. by comparing the results of pre- and post-tests. This is what we have already done in the Dutch-CAPT, DISCO and FASOP projects (see above), in which we were able to show that such systems can be effective for pronunciation and grammar training.

Therefore, although developing CALL systems for atypical speech is complex and challenging, the overview of the projects presented above, and the positive results we obtained in these projects, makes it clear that with current state-of-the-art technology it is possible to develop useful applications for language learners (for testing and training), and persons with communicative disabilities (for diagnosis, therapy, monitoring, and AAC: augmentative and alternative communication). Such research is interesting from a scientific point of view, but obviously the resulting technology and CALL systems can be very useful for the target groups. Valorization of research and

transfer of knowledge from academia to industry are becoming more and more important, and the topics described above offer numerous opportunities in this direction.

REFERENCES

References are listed in alphabetical order, URLs at the end.

- [1] S. Bodnar, B. Penning de Vries, C. Cucchiari, H. Strik, R. Van Hout (2011) Feedback in an ASR-based CALL system for L2 syntax: A feasibility study. Proc. of the SLATE-2011 workshop, Venice, Italy.
- [2] J.M. Burgos Guillén, C. Cucchiari, R. van Hout, H. Strik (2012) Spanish Learners of Dutch: Directions to Improve Phonology Acquisition. Submitted to Interspeech 2012, Portland, USA.
- [3] D. Coniam (1999) Voice recognition software accuracy with second language speakers of English. *System*, 27, 49-64.
- [4] C. Cucchiari, J. van Doremalen, H. Strik (2010) Fluency in non-native read and spontaneous speech. Proc. of DiSS-LPSS Joint Workshop 2010, Tokyo, Japan.
- [5] C. Cucchiari, H. van den Heuvel, E. Sanders, H. Strik (2011) Error selection for ASR-based English pronunciation training in 'My Pronunciation Coach'. Proceedings of Interspeech 2011, Florence, Italy.
- [6] C. Cucchiari, Neri, A., Strik, H. (2009) Oral Proficiency Training in Dutch L2: the Contribution of ASR-based Corrective Feedback. *Speech Communication*, Volume 51, Issue 10, October 2009, Pages 853-863.
- [7] C. Cucchiari, H. Strik, L. Boves (2000) Different aspects of expert pronunciation quality ratings and their relation to scores produced by speech recognition algorithms *Speech Communication* 30 (2-3), pp. 109-119.
- [8] C. Cucchiari, Strik, H., & Boves, L. (2002) Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America*, 111(6), 2862-2873.
- [9] K. De Bot (1996) The psycholinguistics of the Output Hypothesis, *Language Learning* 46, 529-555.
- [10] P. Delcloque (2000), History of CALL, see www.ict4lt.org/en/History_of_CALL.pdf, and www.eurocall-languages.org/resources/history_of_call.pdf
- [11] R. M. DeKeyser, Sokalski, K. J. (1996). The differential role of comprehension and production practice. *Language Learning*, 46 (4):613-642 (29).
- [12] T. M. Derwing, Munro, M. J., & Carbonaro, M. (2000). Does popular speech recognition software work with ESL speech? *TESOL Quarterly*, 34, 592-603.
- [13] A. Dlaska and C. Krekeler, "Self-assessment of pronunciation," *System*, 36, pp. 506-516, 2008.
- [14] J. van Doremalen, C. Cucchiari, H. Strik (2010) Optimizing automatic speech recognition for low-proficient non-native speakers. *EURASIP Journal on Audio, Speech, and Music Processing*, volume 2010 (2010), Article ID 973954, 13 pages.
- [15] J. van Doremalen, C. Cucchiari, H. Strik (2010) Using Non-Native Error Patterns to Improve Pronunciation Verification. Proc. of Interspeech 2010, Tokyo, Japan.
- [16] J. Van Doremalen, C. Cucchiari, H. Strik (2011) Speech Technology in CALL: The Essential Role of Adaptation. *Interdisciplinary approaches to adaptive learning; Communications in Computer and Information Science series*, 2011, Volume 26, pp. 56-69.
- [17] F. Ehsani, & Knodt, E. (1998) Speech technology in computer-aided learning: Strengths and limitations of a new CALL paradigm. *Language Learning & Technology*, 2, 45-60.
- [18] M. Eskenazi (2009) An overview of Spoken Language Technology for Education. *Speech Communication*, 2009.
- [19] J. Flege, "Second-language speech learning: Findings and problems," In *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, Winifred Strange (ed.), Timonium, MD: York Press Inc, pp. 233-273, 1995.
- [20] H. Franco, Neumeyer, L., Digaikis, V., & Ronen, O. (2000) Combination of machine scores for automatic grading of pronunciation quality. *Speech Communication*, 30, 121-130.
- [21] V. M. Holland (Ed.). (1999). Tutors that listen: Speech recognition for language learning [Special issue]. *CALICO Journal*, 16(3).
- [22] ISLE 1.4 (1999) Pronunciation training: Requirements and solutions, ISLE Deliverable 1.4. Retrieved February 27, 2002, from <http://nats-www.informatik.uni-hamburg.de/~isle/public/D14/D14.html>.
- [23] S. Izumi (2002). Output, input enhancement, and the noticing hypothesis. *Studies in Second Language Acquisition*, 24:541 - 577 (36).
- [24] S. Kanters, C. Cucchiari, H. Strik (2009) The Goodness of Pronunciation Algorithm: a Detailed Performance Study. Proc. of the SLATE-2009 workshop, Warwickshire, England.
- [25] G. Kawai, Hirose, K. (1998) A method for measuring the intelligibility and nonnativeness of phone quality in foreign language pronunciation training, Proceedings of ICSLP, Sydney, Australia, 1823-1826.
- [26] W. Menzel, Herron, D., Bonaventura, P., & Morton, R. (2000) Automatic detection and correction of non-native English pronunciations, Proceedings of InStiL, Dundee, Scotland, 49-56.
- [27] A. Neri. (2007) The pedagogical effectiveness of ASR-based computer assisted pronunciation training. PhD thesis, University Nijmegen.
- [28] A. Neri, C. Cucchiari, H. Strik (2002) Feedback in computer assisted pronunciation training: Technology push or demand pull? Proc. of ICSLP-2002, Denver, USA, pp. 1209-1212.
- [29] A. Neri, C. Cucchiari & H. Strik (2006) Selecting segmental errors in L2 Dutch for optimal pronunciation training. *IRAL - International Review of Applied Linguistics in Language Teaching*, 44, pp. 357-404.
- [30] A. Neri, C. Cucchiari, H. Strik (2008) The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch. *ReCALL*, Volume 20, Issue 02, May 2008, pp. 225-243.
- [31] E. Sanders, M. Ruiter, L. Beijer, H. Strik (2002) Automatic recognition of Dutch dysarthric speech: A pilot study Proc. of ICSLP-2002, September 16-20, 2002, Denver, USA, pp. 661-664.
- [32] H. Strik, J. Colpaert, J. van Doremalen, C. Cucchiari (2012) The DISCO ASR-based CALL system: practicing L2 oral skills and beyond. Proc. of the Conference on International Language Resources and Evaluation (LREC 2012)
- [33] H. Strik, F. Cornillie, J. Colpaert, J. van Doremalen, C. Cucchiari (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.
- [34] H. Strik, J. van Doremalen, J. van de Loo, C. Cucchiari (2011) Improving ASR processing of ungrammatical utterances through grammatical error modeling. Proceedings of the SLATE-2011 workshop, Venice, Italy.
- [35] H. Strik, J. van de Loo, J. van Doremalen, & C. Cucchiari (2010) Practicing Syntax in Spoken Interaction: Automatic Detection of Syntactic Errors in Non-Native Utterances. Proc. of the L2WS, SLATE-2010 workshop, Tokyo, Japan.
- [36] H. Strik, K. Truong, F. de Wet, C. Cucchiari (2007). Comparing classifiers for pronunciation error detection. Proceedings of Interspeech 2007, Antwerp.
- [37] H. Strik, K. Truong, F. de Wet, C. Cucchiari (2009) Comparing different approaches for automatic pronunciation error detection. *Speech Communication*, Volume 51, Issue 10, October 2009, Pages 845-852.
- [38] M. Swain (1985) Communicative competence: some roles of comprehensible input and comprehensible output in its development, in Gass, M.A., Madden, C.G. (eds.) *Input in Second Language Acquisition*, Rowley MA: Newbury House, 235-253.
- [39] D. Van Compernelle (2001) Recognizing speech of goats, wolves, sheep and ... non-natives. *Speech Communication*, 35, 71-79.
- [40] S. Witt (1999) Use of Speech Recognition in Computer Assisted Language Learning. PhD thesis, University of Cambridge.
- [41] lands.let.ru.nl/~strik/research/ATOP.html
- [42] lands.let.ru.nl/~strik/research/Dutch-CAPT/
- [43] lands.let.ru.nl/~strik/research/DISCO/
- [44] lands.let.ru.nl/~strik/research/FASOP.html
- [45] lands.let.ru.nl/~strik/research/MPC-STW-VG2.html
- [46] lands.let.ru.nl/~strik/research/GOBL
- [47] lands.let.ru.nl/~strik/research/PEDDS
- [48] lands.let.ru.nl/~strik/research/CompPoli