USE OF SPEECH TECHNOLOGY IN LEARNING TO SPEAK A FOREIGN LANGUAGE

Term Paper, Speech Technology, Autumn 2005

Natalia Zinovjeva

1. Introduction

Speech-enabled systems for computer-assisted language learning (CALL) offer a number of advantages for language learners. These systems make it possible to address individual problems of the learners, allow practising at self-paced speed, and the opportunity to practise is not limited to the time the teacher is available. They offer a possibility to store student profiles in log-files, and both the students and the teachers can monitor the improvements and the problems. CALL systems can also help students who find it difficult to practise speaking in public to improve their language skills.[1] In this paper we discuss the use of speech technology in CALL applications designed to improve spoken language skills of L2 learners.

The paper is organized as follows. In Section 2 we explain and exemplify different types of errors that may occur in the learners' speech as seen from the learner's and the teacher's point of view. Section 3 gives an overview of some publications dedicated to the use of speech technology in systems for pronunciation training and testing the student's language skills. In Section 4 we discuss the design of new systems based on speech technology that can help learners to improve their spoken language skills in an optimal way. Section 5 summarizes and concludes.

2. Errors in spoken language

2.1 Mispronounced phones

One of the first problems that learners usually encounter is the pronunciation of "difficult" phones and phone sequences. Learners of languages like Swedish, Finnish, Estonian may experience difficulties trying to pronounce some vowels. Learners of Russian are sometimes unable to produce palatalized consonants, such as /l'/, /p'/, /t'/, and make the difference between palatalized and non-palatalized consonants. Native speakers of Russian tend to pronounce consonants followed by vowels like *ä* and *ö* with a strong palatalization when speaking Swedish. Learners of French and Polish may have problems with the pronunciation of nasal vowels.

A mispronounced phone may impede comprehensibility, especially if the error results in saying a word with a different meaning. In Swedish, *ö* pronounced like [o] in the noun *mönster* ("pattern") would produce another word, *monster* ("monster"). In Russian, if the learner mispronounces the last consonant, the palatalized /t'/ in the noun *мать* (*mat'*, "mother") so that it sounds like the non-palatalized /t/, the resulting word may be interpreted by the listener like the noun *мат* (*mat*, "swearing, obscene language").

2.2 Spelling and pronunciation

The difference between the spelling and the pronunciation in many languages tends to cause errors, especially at the beginner level: problems may occur in pronunciation of words that are familiar to the learner only in written form. Words and morphs in which the spelling differs from the pronunciation may be pronounced as they are spelled. Another common mistake is interpreting some combinations of characters in accordance with the spelling rules of another language. For instance, beginner learners of Italian who know English may pronounce *ch* like [tʃ], when the

[1] Wachowicz and Scott (1999), Eskenazi (1999), Neri et al. (2001), Hincks (2005).

correct pronunciation is [k]. In Russian, the last consonant in an utterance is usually devoiced, and native speakers of Russian tend to erroneously devoice the last consonant in the utterances when speaking other languages, for example English and Swedish.

Native speakers of languages which use the Latin alphabet may experience reading difficulties when beginning to learn languages like Russian, Greek, Arabic, Chinese or Japanese, and the confusion caused by different characters may cause pronunciation errors.

2.3 Lexical stress

Lexical stress in polysyllabic words that are familiar to students mostly in written form can sometimes be placed incorrectly. Hincks (2005) gives some examples of English words that are often mispronounced by Swedish learners: *access*, *capacity*, *component* and *contribute* realized as /ˈækˈsɛs/, /ˈkəˈpæsɪti/, /ˈkʌmpəˌnənt/, and /ˈkəntrɪbjut/.

In some languages, the accent indicates the morphological features of the word. In Russian the only difference between the genitive singular and the nominative or accusative plural of the noun *река* (*reka*, "river"), which are spelled exactly in the same way, is the stress. In the genitive singular *реки* (*rekì*) the second syllable is stressed. The first syllable is stressed in the nominative/accusative plural of the same noun, *реки* (*rèki*). In this case, incorrectly placed stress results in a grammar error.

2.4 Intonation

Intonation is often important for comprehensibility of the learner's utterances. In Russian interrogative clauses the intonation often helps to interpret the meaning of the questions and to distinguish them from affirmative clauses. Other examples of intonational patterns that may need practising are the polar question intonation in English and the distinction between acute and grave word accent in Swedish.

2.5 Other errors

Speaking about the errors in the spoken language we should not forget about the grammar and the vocabulary. When speaking, the learner usually does not have time for thinking, looking up words in a dictionary, checking every sentence or using a grammar book. As a consequence many learners tend to make more grammar errors and lexical errors in speaking than in writing.

2.6 Comprehension

Problems in comprehension may be caused by fast speech tempo or by differences between the spelling and the pronunciation in some words. Beginning learners of languages like French and Danish, in which the spelling differs significantly from the pronunciation, may experience difficulties in understanding spoken language. Many Russian words might not be recognized by the learners because of the difference between stressed and unstressed vowels, the devoicing of the last consonant in the utterance and some assimilations that are not shown in the spelling.

3. Technologies used in teaching spoken language

3.1 Speech analysis

Speech analysis has been used for teaching intonational patterns to second language learners since 1970s. The main principle is that the sound waveform or pitch contour of a student's utterance are visually displayed alongside those of the model utterance. Studies have shown that audio-visual feedback improves both perception of target language

intonation and prosody as well as segmental accuracy. Many commercial software packages incorporate a speech analysis element. However, the guidance in interpreting the feedback is often insufficient.[2]

3.2 Speech synthesis

Speech synthesis is not widely used in computer assisted language learning; at present most developers seem to prefer recordings of natural voices, because speech synthesis can often sound quite artificial. However, some work has been done in investigating the possibilities of using synthetic stimuli in language teaching. Hincks (2005) discusses the use of speech synthesis as an interactive tool for teaching English to young adults who speak Swedish as their first language. In another study mentioned by Hincks formant synthesis has been successfully used to teach Cantonese and Mandarin learners distinctions in English vowel quality. According to Hincks, the potential for speech synthesis in CALL applications lies in using text-to-speech synthesis, and especially in the possibilities of integrating speech synthesis with visual models of the face, mouth and vocal tract.

Handley and Hamel (2004) investigate the requirements of speech synthesis for CALL and describe an experiment in which utterances produced by a speech synthesizer have been presented to a group of teachers and CALL researchers. The participants of the experiment have been asked to rate the comprehensibility and the acceptability of the utterances for three different functions of speech synthesis in CALL: *reading machine*, *pronunciation tutor*, and *conversational partner*. The results show that the ratings of comprehensibility, acceptability and overall appropriateness differ for the three different functions. The utterances were found to be most comprehensible in the context of use as a conversational partner, and least comprehensible in the context of use as a pronunciation tutor. Similarly, the output of the speech synthesizer was found to be least acceptable and appropriate for use as a pronunciation tutor, and most acceptable and appropriate for use as a conversational partner. Handley and Hamel discuss the problem of evaluation of speech synthesis for CALL and argue that evidence is needed to show the developers that speech synthesis is suitable for use in CALL applications.

3.3 Automatic speech recognition

Use of automatic speech recognition creates the possibility to let the learners "converse" with the computer in a spoken dialogue system, which is useful, although the conversations can only be carried out within limited domains. The research shows clearly that the systems for speech recognition that are designed for native speakers are worse at recognizing foreign-accented speech. However, some researchers have created non-native speech databases for developing speech recognizers specific to learners' mother tongue. Neri et al. (2001) come to the conclusion that speech recognition provides an optimal solution to pronunciation learning. A study presented in Neri et al. (2003) aims to examine various reviews on the usability of speech recognition in pronunciation training. In some publications that have appeared in the language teaching community, criticism has been expressed with regard to the ability of the speech recognition systems to recognize accented or mispronounced speech and to provide meaningful evaluation of the pronunciation quality. The study shows that part of this criticism is not entirely justified, being rather the result of limited familiarity with automatic speech recognition. Wachowicz and Scott (1999) suggest that the effectiveness of CALL systems based on speech recognition is not only determined by the capabilities of the speech recognizer, but also by (a) the design of the language learning activity and feedback and (b) the inclusion of repair strategies to safeguard against recognizer error.

Unfortunately, speech recognition systems at present are poor at handling information contained in the speaker's prosody. The limitations of the technology imply that the learner's utterances have to be predictable and that detecting of errors is only possible with a limited degree of detail, which makes it difficult to give the learner corrective feedback.[3] However, speech recognition systems can be used to measure the speed at which the learners speak, and rate of speech has been shown to correlate with speaker proficiency.[4]

[2] Hincks 2005.
[3] Neri et al. (2002a)
[4] Hincks (2005)

3.4 Non-native speech databases and projects using speech technology

Speech databases which consist of utterances from non-native speakers have been created and used, and some researchers have investigated the possibilities to improve the performance of speech recognizers on non-native speech. Bonaventura et al. (2000) report on collecting and annotating a corpus of English sentences recorded by students with German and Italian as their mother tongue. Work on the *ISLE* project aimed at modeling German- and Italian-accented English. The ISLE corpus of non-native spoken English consists of almost 18 hours of annotated speech signals spoken by Italian and German learners of English (Atwell et al. 2000, Menzel et al. 2000). Bratt et al. (1998) describe the methodologies for collection and detailed transcription of a Latin-American Spanish speech database which consists of utterances from native and non-native speakers. Hincks (2005) describes the collection of a database of recordings and transcripts of oral presentations made by Swedish natives studying English. Mayfield Tomokiyo (2000), Mayfield Tomokiyo and Jones (2001), Mayfield Tomokiyo and Waibel (2001) focus on differences between native and non-native speech and adaptation methods for processing non-native utterances. Gerosa and Giuliani (2004a and 2004b) exploit two children databases, one consisting of speech collected from native English children, the other one consisting of English sentences read by Italian learners of the same age, for training a speech recognizer and investigating the possibilities to improve recognition performance on non-native speech. Wang et al. (2003) investigate how different acoustic model adaptation techniques can help to improve the speech recognizer performance on non-native speech with small amount of non-native data available. Morgan and LaRocca (2004) use Gaussian mixture merging for improving the performance of an Arabic speech recognizer on non-native data.

Automatic pronunciation error detection, pronunciation grading and spoken language tests are discussed in a number of publications. Ronen et al. (1997) investigate techniques for detecting mispronunciations as a part of a language instruction system. Kim et al. (1997) try various probabilistic models to produce pronunciation scores for phone utterances from the phonetic alignments generated by a HMM-based speech recognition system. The speech database used in their experiments consists of speech from native speakers of Parisian French and American students speaking French. Langlais et al. (1998) discuss automatic detection of mispronunciation in non-native Swedish speech. The database used in their research consists of speech from 21 non-native speakers from different countries. Witt and Young (1997, 1998) present methods of assessing non-native speech and providing detailed information about the pronunciation quality of the learners. The methods are based on automatic speech recognition techniques. Jo et al. (1998) present a system for detecting pronunciation errors and providing diagnostic feedback to learners of Japanese through speech processing and recognition methods. Automatic pronunciation grading for Dutch is discussed in Cucchiarini et al. (1998a and 1998b). Franco and Neumeyer (1998) propose a paradigm for automatic assessment of pronunciation quality using HMMs to generate phonetic segmentations of the learner's speech. Spectral match and duration scores are obtained from these segmentations. Franco and Neumeyer focus on the task of calibrating different machine scores to obtain the best results. Franco et al. (1998) have observed that beginner language learners often pause within words while reading. They propose a method for modeling intra-word pauses for producing more robust segmental scores.

Teixeira et al. (2000) address the task of predicting the degree of nativeness of the learner utterances. To achieve the best results they do not focus only on on the segmental assessment of the speech signal, but consider also other aspects of speech, such as prosody. Cucchiarini et al. (2000a, 2000b) show that expert fluency ratings of read speech can be predicted on the basis of automatically calculated temporal measures of speech quality. Rate of speech appears to be the best predictor; two other important determinants of reading fluency are the rate at which the speakers articulate the sounds and the number of pauses they make. Imoto et al. (2002) propose a method to evaluate sentence stress in English spoken by Japanese students and achieve accuracy of 95.1% for native and 84.1% for non-native speakers. Raux and Kawahara (2002a and 2002b) introduce a method to diagnose pronunciation errors that are most critical to the intelligibility. They use a probabilistic algorithm to derive intelligibility from error rates computed by a speech recognition-based system and define an error priority function that indicates which errors are most critical to intelligibility. Ishida (2004) proposes a method for classification of Mandarin Chinese bisyllabic words based on the appropriateness of their lexical tones. This method can be used to help non-native learners acquiring tone pronunciation slkills. Truong et al. (2004) use the DL2N1 corpus (Dutch as L2, Nijmegen Corpus 1) which contains speech from native and non-native speakers of Dutch to develop classifiers for three sounds that are frequently pronounced incorrectly by L2-learners of Dutch: /A/, /Y/ and /x/. Tsubota et al. (2002, 2004) present a method for detecting and diagnosing pronunciation errors in Japanese-speaking learners of

English. Neri et al. (2004) report on a study that was carried out to obtain an inventory of segmental errors in the speech of adult learners of Dutch and discuss establishing of priorities for pronunciation training. Bernstein et al. (2004) describe the construction of a spoken Spanish test which is automatically scored using speech recognition technology. Hincks (2005)[5] mention the commercially successfull PhonePass test which uses speech recognition to assess the correctness of students responces and gives scores in pronunciation and fluency.

Some works have been dedicated to *teaching intonation*. Two discourse-level uses of intonation, the use of intonational paragraph markers and the distribution of tonal patterns, and teaching intonation in discourse by means of speech visualisation technology, are discussed in Levis and Pickering (2004). Taniguchi and Abberton (1999) find that interactive visual feedback of the voice fundamental frequency can help Japanese learners to improve their English intonation. Hardison (2004) discusses computer-assisted prosody training and two experiments that show the effective pedagogical application of speech technology. The experiments show that audio-visual training can help learners of French to improve not only their prosody but also their segmental accuracy. Prosody training and its effects are also discussed in Delmonte et al. (1997), Delmonte (1999), and Herry and Hirst (2002).

WebGrader, a multilingual pronunciation grading tool based on speech recognition and pronunciation grading technologies, which is designed for practising pronunciation in a second language, is described in Neumeyer et al. (1998). The *Subarashii* system described in Bernstein et al. (1999) is designed for beginning students of Japanese. The system analyses the students' utterances and responds in a meaningful way in spoken Japanese. Holland et al. (1999) describe a speech-interactive graphics microworld in which learners speak to an animated agent. Dalby and Kewley-Port (1999) present a pronunciation training program for adult learners which uses automatic speech recognition. The *Voice Interactive Training System (VILTS)*, a language-training prototype developed to improve the comprehension and speaking skills, described by Rypa and Price (1999) incorporates speech recognition and pronunciation scoring technology. Kirschning and Aguas (2000) use speech recognition technology for verification of the correct pronunciation of spoken words in Mexican Spanish. In a study presented by Mayfield Tomokiyo et al. (2000) a speech-recognition based system has been successfully used for learning to pronounce the voiced and voiceless interdental fricatives in English. *PLASER*, a multimedia tool with instant feedback presented in Mak et al. (2003), is aimed to teach English pronunciation to students whose mother tongue is Cantonese Chinese. Levi et al. (2004) describe the use of speech recognition in *German Express*, an interactive program for learning German. Their results are encouraging end the learner feedback is positive. Seneff et al. (2004) discuss the use of multilingual spoken dialogue systems as an aid to second language acquisition. Speech recognition is used in the *Let's Go Spoken Dialogue System* described by Raux and Eskenazi (2004), in *Parling*, a CALL system for children, presented by Mich et al. (2004), in the *SCILL* (Spoken Conversational Interaction for Language Learning) project presented by Ye and Young (2005), and in the system described by Bianchi et al. (2004). Some commercially available pronunciation training systems, such as *TriplePlayPlus* and *Accent Coach* developed by Syracuse Language Systems, *Talk to Me* and *Tell Me More* by the French company Auralog, and *Rosetta Stone*, make use of speech recognition.[6]

Badin et al. (2000) investigate the possibility to use a *virtual talking head* and the *Speech mapping* tools for pronunciation training with audio-visual speech stimuli. Granström (2004) presents some work aimed at creating a *virtual language tutor* that can be engaged in many different aspects of language learning, such as pronunciation training and conversational practice.

In Section 2 we mentioned some of the most common errors that occur in the learner's speech. As we can see, speech technology can help the learner with many of these problems. It allows practising the pronunciation of phones that are often mispronounced, teaching lexical stress and intonation. Efforts have been made to detect errors, find out which errors affect comprehensibility most, and evaluate the pronunciation quality. Interactive CALL systems that create the possibility of "conversing" with the computer can also help the learner to improve the comprehension and the conversational skills.

[5] pp. 22-23
[6] Hincks (2005), pp. 26-27

4. Future applications

Learning a foreign language means learning to *use* it. To achieve good results we must not only practise the pronunciation of separate vowels, consonants, words and sentences. We must also be able to *communicate*, making use of our knowledge of grammar and vocabulary and our pronunciation skills. This is something that we have to keep in mind when trying to learn or teach a language, or when designing CALL systems.

We cannot take it for granted that any kind of oral practice will help all students to improve their spoken language skills, or that receiving detailed feedback will always give the best results. Hincks (2005) describes an experiment in which a group of students was given opportunity to practise English pronunciation using *Talk to me*, a system mentioned in Section 2. The evaluation of the results show that all students did not benefit from the training: some of them were found to either not improve or worsen in their oral ability. A possible explanation is that the fluency of intermediate students was impeded by negative feedback which was not constructive enough to be useful for these students.[7] Precoda et al. (2000) study the effects of speech-recognition based pronunciation feedback on the learners' pronunciation ability and come to the conclusion that the design of the user interface deserves as serious attention as the underlying technology.

From an overview of publications dedicated to the use of speech technology in CALL we can outline some recommendations for the design of systems for effective oral training. Learners should have an opportunity to hear large quantities of speech from different native speakers. They should be stimulated to actively practise oral skills and produce large quantities of utterances. The feedback should be comprehensible and pertinent, it should be provided individually and in real-time, and it should focus on those segmental and suprasegmental aspects which affect intelligibility most.[8] The learner's mother tongue, sex and age should be taken into consideration if the CALL application makes use of the speech recognition technology. Use of audio-visual speech stimuli in prosody training, integrating speech synthesis with visual models of the face, mouth and vocal tract and use of animated agents can make the training more efficient.[9]

It is important to take into consideration the pedagogical aspects of language learning and the real needs of the students when developing speech technology-based CALL software. Language teachers can contribute to the research and development of new speech-enabled CALL applications by helping to design and evaluate systems and to create non-native speech databases. To be able to participate in projects and give feedback to the developers, and to use the software in the most efficient way, the language teachers need to get more familiar with speech technology, computer-assisted language learning, ICALL (intelligent computer-assisted language learning) and the current research in these areas.

5. Summary

We have looked at different aspects of the spoken language and examples of errors that may occur in a learner's speech. We have described some experiments and projects aimed at creating speech technology-based CALL systems. We have shown that speech technology can be useful for teaching *intonation*, *lexical stress*, the pronunciation of *phones that are often mispronounced*, *detecting errors*, *evaluating the pronunciation quality*, and improving the *comprehension* and the *conversational skills*. We have also discussed the requirements for designing speech-enabled CALL systems for efficient language learning and underlined the importance of considering the pedagogical aspects of the task and evaluating the effects of training by means of CALL systems based on speech technology.

[7] pp. 38-52
[8] Eskenazi (1999), Neri et al. (2002a, 2002b)
[9] Taniguchi and Abberton 1999, Badin et al. 2000, Granström 2004, Hardison 2004, Hincks 2005

References

Atwell E., Baldo P., Bisiani R., Bonaventura P., Herron D., Howarth P., Menzel W., Morton R., Souter C., Wick H., 2000. *User-Guided System Development in Interactive Spoken Language Education*. Natural Language Engineering 1(1): 1-12, 2000.

Badin P., Bailly G. and Boë L., 1998. *Towards the use of a Virtual Talking Head and of Speech Mapping tools for pronunciation training*. In Proc. of STiLL 98, May 25-27 1998, Marholmen, Sweden

Bernstein J., Najmi A. and Ehsani F., 1999. *Subarashii: Encounters in Japanese Spoken Language Education*. CALICO Journal 1999, Volume 16 Number 3.

Bernstein J., Barbier I., Rosenfeld E. and De Jong J., 2004. *Development and validation of an Automatic Spoken Spanish Test.* In Proc. of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning, Venice, Italy.

Bianchi D., Mordonini M. and Poggi A., 2004. *Spoken dialog for e-learning supported by domain ontologies*. In Proc. of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning, Venice, Italy.

Bonaventura P., Howarth P. and Menzel W., 2000. *Phonetic annotation of a non-native speech corpus.* In Proc. of Workshop Integrating Speech Technology in the Language Learning and Assistive Interface, InStil 2000, Dundee (UK), 2000.

Bratt H., Neumeyer L., Shriberg E. and Franco H., 1998. *Collection and Detailed Transcription of a Speech Database for Development of Language Learning Technologies*. in ICSLP 98, Proc. of the 5th International Conference on Spoken Language Processing.

Cucchiarini C., Strik H. and Boves L., 1998a. *Automatic Pronunciation Grading For Dutch*. In Proc. of the ESCA workshop 'Speech Technology in Language Learning', Marholmen, Sweden, May 1998

Cucchiarini C., de Wet F., Strik H. and Boves L., 1998b. *Assessment of Dutch Pronunciation by Means of Automatic Speech Recognition Technology*. In Proc. of the fifth Int. Conference on Spoken Language Processing (ICSLP'98), Sydney, Australia, Vol. 5

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

Cucchiarini C., Strik H. and Boves L., 2000b. *Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology*. Acoustical Society of America, 107 (2), February 2000

Dalby J. and Kewley-Port D., 1999. *Explicit Pronunciation Training Using Automatic Speech Recognition Technology*. CALICO Journal, 1999, Volume 16 Number 3

Delmonte R. 1999. *A Prosodic Module for Self-Learning Activities*, Proc. MATISSE, London, 129-132.

Delmonte R., Petrea M. and Bacalu C., 1997. *SLIM, Prosodic Module for Learning Activities in a Foreign Language*. In Proc. ESCA , EuroSpeech'97, Rhodes, Vol.2, pagg.669-672

Eskenazi M., 1999. Using a Computer in Foreign Language Pronunciation Training: What Advantages? CALICO Journal, Volume 16 Number 3, 1999.

Franco H. and Neumeyer L., 1998. *Calibration of Machine Scores for Pronunciation Grading*. In ICSLP 98, Proc. of the 5th International Conference on Spoken Language Processing. Sydney Convention Centre, Sydney, Australia. 30th November - 4th December 1998.

Franco H., Neumeyer L. and Bratt H., 1998. *Modeling Intra-Word Pauses in Pronunciation Scoring.* In Proc. of ESCA Workshop on Speech Technology in Language Learning (STiLL 98), Marholmen, Sweden, May 24-17, 1998

Gerosa M. and Giuliani D., 2004a. *Investigating Automatic Recognition of Non-native Children's Speech*. In Proceedings of ICSLP 2004

Gerosa M. and Giuliani D., 2004b. *Preliminary investigations in automatic recognition of English sentences uttered by Italian children.* In Proc. of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning, Venice, Italy.

Granström B. 2004. *Towards a virtual language tutor*. In Proc. of InSTIL/ICALL2004 – NLP and Speech Technologies in Advanced Language Learning Systems – Venice 17-19 June, 2004

Handley Z. and Hamel M., *Investigating the Requirements of Speech Synthesis for CALL with a View to Developing a Benchmark*. In Proc. of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning, Venice, Italy.

Hardison D. *Generalization of Computer-Assisted Prosody Training: Quantitative and Qualitative Findings.* In Language Learning and Technology, January 2004, Vol. 8, Number 1, pp. 34-52

Herry N., and Hirst D., 2002. *Subjective and Objective Evaluation of the Prosody of English Spoken by French Speakers: the Contribution of Computer Assisted Learning*. In Proc. of Speech Prosody 2002 (Aix-en-Provence April 2002)

Hincks R., 2005. *Computer Support for Learners of Spoken English*. Doctoral Thesis in Speech and Music Communication. KTH, Stockholm, 2005

Holland V. M., Kaplan J. D., and Sabol M. A., 1999. *Preliminary Tests of Language Learning in a Speech-Interactive Graphics Microworld*. CALICO Journal 1999, Volume 16, Number 3

Imoto K., Tsubota Y., Raux A., Kawahara, T. and Dantsuji M., 2002. *Modeling and Automatic Detection of English Sentence Stress for Computer-Assisted English Prosody Learning System*.

Ishida A., 2004. *Computational method to determine the appropriateness of lexical tones*. In Proc. of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning, Venice, Italy.

Jo C., Kawahara T., Doshita S. and Dantsujiy M., 1998. *Automatic Pronunciation Error Detection and Guidance for Foreign Language Learning.* In Proc. of ICSLP 1998

Kim Y., Franco, H. and Neumeyer L., 1997. *Automatic Pronunciation Scoring of Specific Phone Segments for Language Instruction*. In Proc. of EUROSPEECH 97

Kirschning, I. and Aguas N., 2000. *Verification of Correct Pronunciation of Mexican Spanish using Speech Technology.* In Proc. of the Mexican International Conference on Artificial Intelligence: Advances in Artificial Intelligence

Langlais P., Öster A. and Granström B., 1998. *Automatic Detection of Mispronunciation in non-native Swedish Speech.* In Proc. of ESCA workshop STILL98, Marholmen, Sweden, 1998.

Levi T., Stokowski S., Koster N. and Ryschka A., 2004. *Voice Recognition Feature of the German Express Courseware: Conceptualization, Specification and Prototyping – Model Elaboration through Phases*. In Proc. of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning, Venice, Italy.

Levis J. and Pickering L., 2004. *Teaching intonation in discourse using speech visualization technology*. System, 32 (4): 505-524.

Mak B., Siu M., Ng M., Tam Y., Chany Y., Chan K., Leung K., Ho S., Chong F., Wong J. and Lo J., 2003. *PLASER: Pronunciation Learning via Automatic Speech Recognition.* In Proc. of HLT-NAACL, May, 2003, Edmonton, Canada

Mayfield Tomokiyo L., Wang L. and Eskenazi M., 2000. *An Empirical Study of the Effectiveness of Speech-Recognition-Based Pronunciation Training.* In Proc. of ICSLP2000

Mayfield Tomokiyo L. and Jones R., 2001. *You're Not From 'Round Here, Are You? Naïve Bayes Detection of Non-native Utterance Text.* In Proc. of the 2nd Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL-01)

Mayfield Tomokiyo L., 2000. *Lexical and Acoustic Modeling of Non-Native Speech in LVCSR*. in Proc. of ICSLP. '00, Beijing, China, 2000.

Mayfield Tomokiyo L., and Waibel A., 2001. *Adaptation Methods for Non-native Speech.* In Proc. of Multilinguality in Spoken Language Processing, Aalborg, September, 2001.

Menzel W., Atwell E., Bonaventura P., Herron D., Howarth P., Morton R. and Souter C., 2000. *The ISLE corpus of non-native spoken English.* In Proc. of 2nd International Conference on Language Resources and Evaluation (LREC 2000), vol.2, 31 May-2 June 2000, Athens

Mich O., Giuliani D. and Gerosa M., 2004. *Parling, a CALL system for children*. In Proc. of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning, Venice, Italy.

Morgan J. and LaRocca S., 2004. *Making a Speech Recognizer Tolerate Non-native Speech through Gaussian Mixture Merging*. In Proc. of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning, Venice, Italy.

Neri A., Cucchiarini, C. and Strik H., 2001. *Effective feedback on L2 pronunciation in ASR-based CALL.* In Proc. of the workshop on Computer Assisted Language Learning, Artificial Intelligence in Education Conference, San Antonio, Texas

Neri A., Cucchiarini C. and Strik H., 2002a. *Feedback in Computer Assisted Pronunciation Training: When technology meets pedagogy.* Proc. of CALL Conference "CALL professionals and the future of CALL research", Antwerp, Belgium

Neri A., Cucchiarini C. and Strik H., 2002b. *Feedback in Computer Assisted Pronunciation Training: Technology Push or demand Pull?* In Proc. of ICSLP 2002, Denver, USA

Neri A., Cucchiarini C. and Strik W., 2003. *Automatic Speech Recognition for second language learning: How and why it actually works.* In Proc. of 15th ICPhS, Barcelona, Spain

Neri A., Cucchiarini C. and Strik W., 2004. *Segmental errors in Dutch as a second language: how to establish priorities for CAPT.* In Proc. of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning, Venice, Italy.

Neumeyer L., Franco H., Abrash V., Luc J., Ronen O., Bratt H., Bing J., Digalakis V. and Rypa M., 1998. *WebGrader: A Multilingual Pronunciation Practice Tool.* In Proc. of Speech Technology in Language Learning Workshop, Stockholm, Sweden.

Precoda K., Halverson C. A. and Franco H., 2000. *Effects of Speech Recognition-based Pronunciation Feedback on Second-Language Pronunciation Ability.* In Proc. of InSTILL 2000.

Raux A. and Kawahara T., 2002a. *Automatic Intelligibility Assessment and Diagnosis of Critical Pronunciation Errors for Computer-Assisted Pronunciation Learning.* ICSLP 2002, pp.737--740, Denver.

Raux A. and Kawahara T., 2002b. *Optimizing Computer-assisted Pronunciation Instruction by Selecting Relevant Training Topics*. InSTIL 2002 Advanced Workshop, Davis, CA.

Raux A. and Eskenazi M., 2004. *Using Task-Oriented Spoken Dialogue Systems for Language learning: Potential, Practical Application and Challenges.* In Proc. of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning, Venice, Italy.

Ronen O., Neumeyer L. and Franco H., 1997. *Automatic Detection of Mispronunciation for Language Instruction*. In Proc. of Eurospeech, Vol. 2, Rhodes, Greece

Rypa M. E. and Price P., 1999. *VILTS: A Tale of Two Technologies*. CALICO Journal 1999, Volume 16, Number 3.

Seneff S. and Wang C., 2004. *Spoken Conversational Interaction for Language Learning*. In Proc. of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning, Venice, Italy.

Taniguchi M. and Abberton E., 1999. *Effect of interactive visual feedback on the improvement of English intonation of Japanese EFL learners.* Speech, Hearing and Language: Work in Progress No. 11 (pp.76-89) Department of Phonetics and Linguistics, University College London (1999)

Teixeira C., Franco H., Shriberg E., Precoda K. and Sönmez K., 2000. *Prosodic Features for Automatic Text-Independent Evaluation of Degree of Nativeness for Language Learners.* ICSLP - 6th International Conference on Spoken Language Processing, Beijing, China, October 2000

Truong K., Neri A., Cucchiarini C. and Strik H., 2004. *Automatic pronunciation error detection: an acoustic-phonetic approach.* In Proc. of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning, Venice, Italy.

Tsubota Y., Kawahara T. and Dantsuji M., 2002. *CALL System for Japanese Students of English Using Pronunciation Error Prediction and Formant Structure Estimation*. In InSTIL 2002 Advanced Workshop

Tsubota Y., Dantsuji M. and Kawahara T., 2004. *Practical Use of Autonomous English Pronunciation Learning System for Japanese Students*. In Proc. of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning, Venice, Italy.

Wachowicz K. A. and Scott B., 1999. *Software That Listens: It's Not a Question of Whether, It's a Question of How.* CALICO Journal 1999, Volume 16, Number 3.

Wang Z., Schultz T. and Waibel A., 2003. *Comparison of Acoustic Model Adaptation Techniques on Non-native Speech*. In Proceedings of ICASSP 2003.

Witt S. and Young S., 1997. *Language learning based on non-native speech recognition.* In Proc. of EUROSPEECH '97, Rhodes, Greece, 1997

Witt S. and Young S., 1998. *Performance Measures for Phone-Level Pronunciation Teaching in CALL*. In Proc. STiLL 1998

Ye H. and Young S., 2005. *Improving the Speech Recognition Performance of Beginners in Spoken Conversational Interaction for Language Learning*. In Proc. of Interspeech 2005, Lisbon, Portugal.