New Perspectives in Teaching Pronunciation

MARIA GRAZIA BUSÀ

1. Introduction

Computer Assisted Language Learning (CALL) applications are a useful aid for both language teachers and individual learners. CALL applications offer individualized environments where learners learn at their own pace making autonomous decisions on the order of study topics, lesson reviews, lesson repeats, etc. In other words, CALL applications both promote language learning objectives and overcome traditional language classroom constraints. Some of these applications for the enhancement and practice of oral skills consist in applications for pronunciation teaching. The goal of Computer Assisted Pronunciation Training (CAPT) systems is to provide learners with private, stress-free practice with individualized and instantaneous feedback on pronunciation. The introduction of CALL applications has stimulated a debate on the relationship between pedagogy and technology, and the role of the language teacher in the classroom. Some applications, particularly commercial applications, seem to drive technological advances to the detriment of pedagogical criteria which would be more beneficial to the learner (Neri, Cucchiarini, Strik and Boves 2002). However, these applications are unquestionably making an important contribution to linguistic research and language-teaching practice.

This paper reviews recent technology for teaching pronunciation, and the trends emerging in this field. One particular method for teaching prosody, in particular intonation and pitch patterns, is reviewed in considerable detail. This method uses speech analysis software to provide students with visual data in addition to audio data, and give them feedback on their L2 production. It thus of-

fers visual and audio information on exactly where students' production differs from native speakers'. The last section reports on a pilot study conducted at the University of Padova designed to implement this pronunciation teaching method in future courses by testing the effectiveness of using these computer-based visual feedback systems to modify non-native speakers' intonation patterns.

2. Foreign accent, intelligibility and pronunciation teaching

Individuals who start acquiring a second language after early childhood rarely develop native-like speech patterns, even after considerable exposure to a second language. In fact, in most cases, second-language learners speak with a *foreign accent*. But what exactly is a foreign accent? A foreign accent can span from a barely perceptible accent to strongly accented, unintelligible speech. However, what contributes to native speakers' judgments of a foreign accent is still not fully understood, nor is the boundary between a foreign accent and unintelligibility well defined. "Heavily-accented" speech does not necessarily correspond to unintelligible speech, and it is possibly the *type* more than the *number* of learner's mistakes that affects L2 speech intelligibility (Munro, forthcoming, Munro and Derwing, 2001). In recent years, the study of foreign accents has attracted the interest of scholars from a variety of fields, from first- and second-language acquisition to speech perception and production, from sociolinguistics to applied linguistics. Studies on foreign accents have *inter alia* investigated what factors contribute to our perception of *foreign accent* vs. *unintelligible speech* and why human beings have difficulties acquiring L2 speech articulatory patterns while no other limitation in their motor-control system is reported that would prevent them from learning any other articulatory behavior.

In applied linguistics, the issue of foreign accents is connected to pronunciation teaching. In L2 instruction, the amount of attention that has been given to pronunciation teaching has changed considerably over the past fifty years as have the opinions regarding the extent to which non-native pronunciation errors should be corrected. In the 50's and 60's, during the heyday of the audio-lingual approach, the goal of L2 pronunciation instruction was the attainment of a native-like accent, as modeled by the language teacher. In this period, pronunciation teaching instruction focused on the discrimination and articulation of sounds as a way of improving the perception and production of L2 non-native sounds. This approach brought phonetics and phonology into the language classroom, as it was believed that correct articulation of L2 sounds required a basic understanding of the mechanisms for L2 and L1 sound production (Lambacher, 1996a). In the late 60's and 70's, when the cognitive approach was dominant, the belief that native-like pronunciation was impossible to attain for an adult second language learner had the overall effect of decreasing the attention given to pronunciation, as well as the amount of knowledge about L2 phonetics and phonological systems that were deemed necessary for the language learner. In the early 1980's, communication-oriented approaches to language teaching recognized the key role of pronunciation in improving the learner's oral skills, and in contributing to ensuring the success of oral communication. Today, language

teachers and researchers generally agree that the ultimate goal of pronunciation teaching should not be to eradicate a foreign accent, but rather to promote pronunciation which is reasonably intelligible, as intelligible pronunciation is considered an essential component of communicative competence (Celce-Murcia, 1987; Anderson-Hsieh, 1989; Morley, 1991; Lambacher, 1996a, 1996b; Stibbard, 1996). The attainment of intelligible pronunciation is considered essential for the learner to increase self-confidence and promote social interactions outside the classroom (Morley, 1991; Cunningham Florez, 1998). Pronunciation accuracy may also help improve a person's social acceptance, since a foreign accent may be socially stigmatized and contribute to negative stereotyping of some second-language learners, and thus result in social or professional discrimination (Munro, forthcoming; Derwing, Rossiter, and Munro, 2002). Finally, it is believed that, because «the number of professionals who regularly communicate in a foreign language for their work has increased with globalization[, i]n order to ensure that these learners are able to efficiently communicate in the L2, it is imperative that language teaching methods include pronunciation training» (Neri, Cucchiarini, Strik, and Boves, 2002: 442).

3. Prosodic features of second-language speech

Research has investigated what components of a foreign accent play a role in the intelligibility of spoken language. The critical role of prosody in the production and perception of L2 speech has been ascertained, and prosody is believed to have an effect on judgments about foreign accents. For example, the perception of L2 fluency and speech has been found to be affected by differences in speech rate (Munro and Derwing, 2001; Derwing and Munro, 2001; Kormos and Dénes, 2004), pitch prominence, pitch range, length and location of pauses (Pickering, 2002; 2004), intonation contours (Wennerstrom, 2000; 2001), prosodic stress, as characterized by acoustic parameters such as amplitude and duration (Chang, 2002; Silipo and Greenberg, 2000). As for Italian learners of English, the perception of an Italian accent in English decreases significantly, and intelligibility increases significantly, as Italians learn English timing strategies for vowel and syllable production (Busà, 1995).

The finding that prosodic features affect the production and perception of L2 speech comes as no surprise given the fundamental role prosody plays in first language acquisition and, in general, in speech communication. Research into infant speech development has shown that, even before they are born, infants are finely tuned to perceiving prosodic aspects of speech (i.e., variations in duration, tempo, pitch, and intonation patterns), and it is, in fact, through timing and intonation that they learn to understand their caregivers' emotions and to express their own (Lieberman, 1986, Gerken and Aslin, 2005). Even in adult communication, prosody is what glues sounds in words and words in utterances, and it is through prosody that speakers prioritize information, signal emphasis, disambiguate sentences, make meaning in context, etc. Prosodic features such as stress and intonation contribute an essential part of the linguistic interpretation of an utterance, as they provide overt and, especially, covert information on

the message transmitted, and/or the emotions and attitudes conveyed with it (Wilson and Wharton, forthcoming).

Though suprasegmentals (prosody) represent the basic step in first language acquisition, they seem to be extremely hard for second language learners to acquire. There may be many reasons for this. In general, while speakers are usually able to use and interpret prosody successfully in their everyday communication, they may have no awareness of prosodic patterns in speech, and may have difficulties hearing, recognizing or labeling different prosodic patterns, such as segmental durations, rhythmic or intonation patterns (i.e., rising vs. falling intonation, rising-falling vs. falling-rising intonation, etc.). Prosodic phenomena are difficult even for native speakers to describe and analyze (Bradford, 1992: 1) and to agree on (Brazil, 1994: 6). Moreover, as discussed in Section 4 below, learners may have a hard time acquiring L2 prosodic patterns because traditional in-class explanations, methods and materials may not always be adequate, as they may not enhance comprehension of the differences between the L1 and L2 prosodic systems. For example, learners are more likely to practice with drills for discriminating minimally contrasting *word* pairs than with drills for discriminating minimally contrasting *prosodic* pairs. So, suprasegmentals may be difficult to acquire because L2 prosodic patterns may be hard for the learner to perceive and make sense of. It is reasonable to assume that pronunciation instruction could benefit from new methods or applications enhancing students' comprehension and perception of L2 prosodic features.

4. Approaches to teaching L2 pronunciation and prosody

As seen in Section 2, today's approaches to second-language teaching generally recognize pronunciation as having a key role in the achievement of successful communication (Cunningham Florez, 1998). However, as reported in the literature, the extent to which pronunciation is taught in the language classroom varies, and «the amount of time and effort devoted to it seems to depend, to a large degree, on the individual teacher. This means that it may or may not form part of regular classroom activities or student self-study» (Macdonald, 2003: 1). Several factors seem to contribute to language teachers' tendency to avoid teaching pronunciation. Teachers often feel that they are inadequately prepared to teach it. Also, pronunciation instruction is not appropriately emphasized in curricula. And finally, suitable materials for teaching pronunciation are often unavailable (Fraser, 2000; Yates, 2001; Macdonald, 2003 and referenced works). The methods used for teaching pronunciation also vary widely, ranging from «drilling sounds, words, and dialogues, [...] instruction in the phonological rules of English, including stress placement, spelling-to-sound rules, intonation patterns, [...], listening to authentic materials, [or] a mix of these methods» (Fraser, 2000: 29).

Pronunciation instruction also differs widely as regards which aspects of pronunciation get emphasized in the classroom. Many traditional approaches tend to focus more on language segments than on suprasegmentals (Spaai and Hermes, 1993; Chun, 1998; Cunningham Florez, 1998; Yates, 2001). In other words, more emphasis is placed on the production and discrimination of indi-

vidual sounds than on how sounds are modified when produced in stretches of speech, i.e., words and utterances, due to the combined effect of stress, language rhythm, connected speech processes, prominence and intonation patterns. The focus on segmentals in pronunciation teaching both reflects and benefits from previous trends in phonetic and phonological research, which have contributed significant descriptions and explanations on sound articulation and acoustics. In addition, a large number of contrastive studies conducted on L1 and L2 phonetic and phonological systems have privileged speech segments over suprasegmentals. Overall, then, teachers may be more prepared to teach segmental as opposed to suprasegmental properties of L2 speech, and may be aware of differences between the L1 and L2 at segmental but not suprasegmental level.

Language pronunciation teaching approaches emphasizing the acquisition of L2 segments over L2 prosody are typically characterized by the extensive use of drills on word minimal pair discrimination and repetition (e.g., /bead/ - /bid/). These drills are based on the idea that perceiving a difference that does not exist in L1 is an essential prerequisite for good pronunciation. For example, Italian speakers' inability to produce English vowel contrasts such as those existing in words like "bid" and "bead" may be rooted in the inability to discriminate the two vowel sounds perceptually. Repeated listening to the sounds in contrast is considered an effective method to help learners discriminate the sounds auditorily, and consequently help them produce the sounds contrastively.

On the other hand, the fact that, overall, pronunciation instruction has placed less emphasis on suprasegmentals than on segmentals may be due to the fact that, to date, we still have only a partial understanding of language prosody, and that what we know «is split up into a large number of competing approaches» used for different languages (Mixdorff, 2002: 31). Research on suprasegmentals is complex requiring investigations of physical (i.e., acoustic, articulatory and perceptual) properties, as well as communicative functions; prosodic meaning depends on individual, social and contextual factors. Thus, because of its inherent complexity, attempts at describing prosody in ways amenable to instruction have proved elusive, particularly with regard to context-related variation, interdialect and interpersonal variation. If learners are not provided with clear explanations of the rules governing L2 prosodic patterns, they may not be able to make useful generalizations or comparisons with patterns in the native language. As Spaai and Hermes (1993) report, if prosody is taught implicitly, and with no clear explanations, i.e., by means of the "listen and repeat" method, it cannot really be learned.

However, the growing interest in the study of suprasegmentals generated by the recognition of the role of prosody in first and second language speech communication is causing a shift in emphasis in foreign language pronunciation teaching. The new approaches to pronunciation teaching are more balanced in focus, and more emphasis is placed on pitch, stress, rhythm coarticulation and intonation, and how they are used to communicate meaning, the general goal being to achieve comprehensible speech for better overall speech performance (Lambacher, 1996a).

In the past ten years or so, a new impulse to teaching L2 prosody has come from technology, and particularly from speech technology. At the present stage,

the use of technology for pronunciation teaching is still largely experimental in nature, but there are indications that new methods and frameworks may be developing that will be beneficial to the study and acquisition of L2 suprasegmentals.
In the following sections, this paper will review some of the new tools for speech and communication research (*Section 5*), how they have impacted pronunciation teaching technology (*Section 6*), and, in particular, how they have affected the teaching of prosody (*Section 7*).

5. New tools in speech and communication research

In the past fifteen years, scientific research has been enhanced by the greater accessibility and lower costs of computer hardware and software, as well as by the huge increase in computers' data storage capabilities. For speech research, tools have been developed that allow the recording and digitalization of authentic spontaneous speech for storage and analysis in ways that only a few decades ago were not conceivable. These tools have contributed to the advancement of speech and natural language research and have provided the input for pronunciation teaching applications. The speech research tools relevant to the present paper are tools for speech analysis, including prosodic analysis, and multimodal analysis.

5.1 *Tools for speech analysis*

Hardware and software systems have been used for many decades in experimental phonetics as an aid to study the physical properties of speech sounds, whether acoustic, articulatory, aerodynamic or perceptual. Computerized speech signal analysis and processing have long been the basis for speech technology (speech synthesis and speech perception) applications. However, while in the past highly specialized hardware and software systems for speech analysis were confined to university computer labs, today, reduced costs and the availability of freely accessible or relatively inexpensive software has made it possible to store and analyze speech data from any home or portable computer. There is also a wide variety of signal analysis software with features for quick and accurate extraction of frequency, pitch contours, intensity levels, as well as the on-screen display of speech sound waves and spectrograms, filtering signals and so on. This software may also include tests for listening and discrimination of various types of signals, for signal processing, etc. A widely-used freeware program for speech analysis which is gaining increasing acceptance is *Praat*, developed by Paul Boersma and David Weenink at the Institute of Phonetic Sciences of the University of Amsterdam (NL) and available at http://www.praat.org. *Section 8* of this paper reports on a pilot project for teaching prosody to Italian learners of English using *Praat*.

For prosody, software for computerized signal analysis has favored the creation of conventional systems for transcribing intonation and prosodic structures of spoken utterances. One of the most widely used systems is *ToBI*, developed at the Ohio State University Department of Linguistics by Mary E. Beckman and her co-workers (e.g., Beckman and Elam, 1997; see also the *ToBI* website at: http://

www.ling.ohio-state.edu/~tobi/). Conventional systems like *ToBI* have made an important contribution to the study and description of speech prosody. However, we are far from achieving a "standard model" for prosodic representation, because these systems still require language-specific (or even dialect-specific) adaptation, as well as adaptations to the specific research of different research teams.

5.2 *Tools for multimodal analysis*

For discourse analysis, conversational analysis, text analysis, pragmatics, anthropology, human-computer interaction, computer animation and many other fields, systems are being developed that allow researchers to integrate linguistic with non-linguistic information. These systems allow simultaneous recording, notation and analysis of visual and audio information to study how meaning is conveyed through language and through other resources as well. The characteristics of each system depend on the individual research team as well as the specific purposes of analysis for which these systems have been developed. For example, in Italy, a research project involving three Italian universities (i.e., Pavia, Trieste and Padova) has developed *MCA* (*Multimodal Corpus Authoring System*), a system which allows users to analyze film texts and to study the meaning-making processes and meaning-making structures characterizing them (e.g., Baldry 2004, Baldry and Taylor, 2004; Baldry, 2005; see also the *MCA website at: http://mca.unipv.it/*). Because it adopts a comparative, corpus-based approach to film analysis and transcription, *MCA* can, for example, focus on different soundtracks comparing the ways in which similar communicative functions are realized and connected to others in different film texts (Ackerley and Coccetta *in press*; Dalziel and Metelli, *in press*); this includes comparative analysis/transcription of native/non-native phonetic and prosodic meaning oppositions in film soundtracks (see Baldry and Thibault, 2006: 51-54). For more general purposes, multimodal systems have been created which serve both theoretical and applied language research. ANVIL is a free video annotation tool, used at research institutes worldwide, providing hierarchical multi-layered annotation driven by user-defined schemes (Kipp, 2004; see also the *ANVIL* website at: http://www.dfki.de/-kipp/anvil/). Because it can import data from common signal analysis software such as *Praat* and *Xwaves*, and can display waveforms and pitch contours, it is also used by teams conducting research on speech. *SignStream*, developed by the American Sign Language Research Project at Boston University, is a database tool for analysis of data captured on video. Although this system was designed to work with data from American Sign Language, it may be applied to any kind of data captured on video and is useful when studying the gestural component of oral interaction (see the *SignStream* website at: http://www.bu.edu/asllrp/SignStream/). Another commonly used system is *MacVisSTA*, designed to perform analyses of multimodal human communication through video, audio, speech transcriptions, and gesture, head, posture, facial expression and gaze orientation data, and is particularly focused on the analysis of the co-temporality of behavior modes (Rose, Quek and Shi, 2004).

6. Technological applications in speech communication and their impact on pronunciation teaching

As reviewed in Section 5, greater accessibility and lower costs of computer hardware and software, increase in computer data storage capabilities, new tools for speech and interaction analysis have favored the collection of extensive corpora containing natural language data that can be studied at many different levels, i.e., from the acoustic to the discourse level, using a variety of techniques for transcription, labeling and examination. The bulk of language data collected is at the same time furthering our understanding of human communication, and contributing to the creation of new technological applications involving speech. Research on language teaching has investigated ways to make beneficial use of technological advances for improving language learning. Here too, a wide variety of applications are being developed for different purposes and many are yielding favorable results. This section will briefly review the major speech technology applications that have a bearing on pronunciation teaching.

Speech synthesis and automatic speech recognition, two of the most common applications in speech technology, provide the basic technology for the development of applications for pronunciation teaching. Speech synthesis, i.e., computer-generated speech production, is typically used «for rudimentary listening comprehension and for learning sound-symbol (phoneme-grapheme) correspondences» (Chun, 2006: 279). For example, a commercial software package, *RealSpeak™ Word* by Nuance, uses speech synthesis to convert the words and idioms from a dictionary into speech output, to allow learners to hear how words should be pronounced (see the product website at: http://www.nuance.com/realspeak/word/).

Automatic speech recognition constitutes the basis for a large number of applications for pronunciation improvement, in spite of the fact that speech recognition has not reached the same high level of performance as speech synthesis (speech recognition applications work better when either the number of users or vocabulary items is restricted). Typically, in automatic speech recognition systems, L2 learners' pronunciation is compared against native speaker models and learners are told their errors and/or corrected accordingly. Speech recognizers are used in the development of automatic (phone) pronunciation error detection as an aid in pronunciation teaching classes or for individual learners (Kim, Franco and Neumeyer, 1997; Truong, Neri, Cucchiarini, Strik, 2004; Truong, Neri, de Wet, Cucchiarini, Strik, 2005). In these applications, learners listen to native language samples, repeat and record their productions, and compare these to native speaker models (Wachowicz and Scott, 1999). *Fluency*, for example, developed by Eskenazi (e.g., 1999a) at the Language Technology Institute at Carnegie Mellon University, in Pennsylvania, is a system that gives the user visual and audio suggestions on how to detect and correct his/her pronunciation mistakes (both segmental and suprasegmentals). Additionally, such systems can also give feedback on the correctness of some learners' limited reading tasks (Mostow and Aist, 1999).

Paradigms based on speech recognizers are also used for automatic assessment of pronunciation quality and are increasingly used in educational systems

to assess students' oral discourse proficiency levels (Bernstein, 1997; Levow and Broman Olsen, 1999; Neumeyer, Franco, Digalakis and Weintraub, 2000; Bernstein, Balogh, Lennig, Rosenfeld, 2005). When assessing learners' overall proficiency, these systems require complex design architectures, which, in addition to learners' pronunciation skills, must take other aspects of learners' language, such as grammar and vocabulary, into account.

Closed-response systems are applications that check the correctness of learners' vocabulary or spoken conversational skills in certain virtual interactions requiring a limited set of user responses. They also make use of speech synthesis and speech recognition technology (e.g., Egan, 1999; Harless, Zier and Duncan, 1999). Open-response systems check the correctness of learners' vocabulary or spoken conversational skills without restricting a learner's utterances. They have more complex architectures and require higher processing capabilities. The development of open-response systems is based on the expansion of the capabilities of so-called *Spoken Dialogue Systems*, originally developed to support access to online information sources. Open-response systems are meant to provide students with the possibility of practicing spoken dialogue interactions (with a computer), and to give them feedback on the quality of their utterances during the dialogue exchange (Seneff, Wand and Zhang, 2004; Rauz and Eskenazi, 2004).

Finally, an aid to pronunciation teaching classes may come from the development of so-called *Talking Heads*, i.e., computer-animated heads (conversational agents) which combine speech technology with studies on gestures and head and face movements. These talking heads are designed to appear on the learner's computer monitor and function as the learner's virtual tutors to be involved in many aspects of his/her language learning process, from reading to pronunciation to conversation practice. Talking heads are also being developed for children learning their first language and disabled people, the deaf in particular. Researchers believe that, because of their realistic speech and expressions, and their convincing emotions, talking heads will become patient and fun-giving interactive tutors for learners to learn languages with (Massaro, 2006a). Massaro and his team have developed probably the most well-known talking heads: *Baldi* and his sister *Baldette* (e.g. Massaro, 2006b), *Timo* (the interactive children's tutor, see http://animatedspeech.com), *Baldini*, the Italian version of Baldi (e.g., Cosi, Cohen and Massaro 2002). Other teams conducting research into talking heads are working with Kalberer and Müller at the Department of Information Technology and Electrical Engineering (Computer Vision Laboratory) in Zürich (http://www.vision.ee.ethz.ch). Granström, at the Centre for Speech Technology in Stockholm, Sweden, is also developing a virtual language tutor (Granström, 2004).

Before concluding this section a word of caution on new technological advances is in order. Even though new technological advances offer exciting perspectives for second-language pronunciation teaching, people are still the best teachers, evaluators and correctors of learners' performances in L2. Much research and careful evaluation is needed before the new technological advances can offer a valid, unquestionable aid to the pronunciation teacher. While scholars agree that ASR (automatic speech recognition)-enhanced materials can effectively increase students' learning potential over conventional materials, care needs to be taken to prevent the exploitation of these systems, especially

if their development is driven by commercial rather than pedagogical purposes (Wachowicz and Scott, 1999; Derwing, Munro and Carbonaro, 2000; Delmonte, 2000; Neri, Cucchiarini, Strik and Boves, 2002).

7. Using signal analysis software for teaching intonation in discourse

As seen in Sections 3 and 4, prosody *should* be taught in the language classroom from the beginning and with effective methods. New inputs for new methods for teaching L2 intonation and prosody are emerging from phonetic research, aided by speech analysis software now available. This section will briefly review how speech analysis software has been used in pronunciation teaching classes and what insights into teaching prosody can be gained from this method.

7.1 *Existing software for speech analysis and uses in foreign language pronunciation teaching*

Section 4 discussed how teaching and learning prosody is more difficult to implement than teaching and learning single sound production. As seen in Section 4, this difficulty may be partly due to the fact that teachers themselves cannot rely on explanations, methods and materials on how to teach intonation effectively. However, a further problem may also be that speakers vary in their ability to hear prosodic patterns in their L1 and in their L2, and thus find themselves at a loss when asked to discriminate or reproduce prosodic patterns in the L2. As suggested by a few researchers (e.g., Spaai and Hermes, 1993; Lambacher, 1996b; Stibbard, 1996; Chun, 1998; Eskenazi, 1999; Wennerstrom, 2000), a combination of audio and visual feedback may have a major impact on learners and enhance their ability to learn both segmental and suprasegmental aspects of pronunciation. On these grounds, speech analysis software has started to be introduced experimentally in L2 pronunciation classes as a source of feedback for students' productions. The use of speech analysis software allows learners to record and visualize their speech output on their computer monitors to obtain real-time information about the acoustic properties of this output. These visualizations can be used by both learners and teachers to compare and evaluate learners' productions with those of native speakers. Through these visualizations, learners have an objective measure of the distance or closeness of their pronunciation with respect to the target pronunciation. This method is considered to be highly effective by the researchers who have used it. Visualization of intonation curves would appear to be particularly effective. So, for example, Eskenazi (1999) maintains that the visual display of L2 prosodic patterns may be crucial for correcting students' inaccurate prosody, because it allows them to visualize where exactly their prosodic patterns differ from native speakers'. Similarly, Wennerstrom (2000), argues that the visualization of pitch ranges in speech makes it easier for the learner to increase pitch to signal topic shift, and this has a bearing on learners' overall intelligibility in L2. Reports of successful teaching experiences using systems developed for phonetic and speech research and on the effectiveness of visual displays for teaching prosody and intonation are also found in De Bot (1983), Spaai and Hermes (1993), Lambacher (1996a, 1996b), Stibbard (1996), Chun (1998).

7.2 *How do speech analysis systems work and how can they be used in the language classroom?*

Typically, speech analysis systems allow users to record, visualize and analyze speech on their computer screen. Students using these programs can obtain an accurate visualization of their production at both the segmental and the suprasegmental level. In order for students to make sense of the on-screen visualizations of their speech productions, they need to be given a theoretical background on how to read and interpret speech signals, spectrograms and prosodic patterns of pitch, intonation and loudness. Hence, for L2 instruction, it is crucial for students to acquire some notion of L1 and L2 phonology, as well as some elements of acoustic phonetics before they start experimenting with visualizing sound waves and pitch contours. Students should also be made aware of the great variability that may differentiate inter- and intra-speaker productions. However, visualizing speech and comparing students' own productions with native speakers' is a rather simple task and does not require much technical or theoretical expertise. The pros of this method largely outweigh its cons. As reported by Lambacher (1996b: 32), «The function of this computerized training system is very appealing and effective as a learning and teaching tool in pronunciation since it allows students to visualize their pronunciation as they learn to associate the patterns on the display with the sounds. The sound analyzer is also very motivating to students because it provides them with a deeper sense of their own articulation by allowing them to visually compare their own pronunciation with their teacher's [or with the native speaker's]. Students visualize their pronunciation and learn to interpret the different patterns of sound segmentals and suprasegmentals, by associating the patterns on the screen with the sounds they are producing».

Intonation contours and pitch levels can be easily visualized and analyzed by students who do not have much training in phonetics or speech analysis. The visualization of intonation and pitch patterns enhances the comprehension of intonation contours (e.g., falling, rising intonation, etc.) and pitch levels (i.e., high, medium, low pitch). On the other hand, the visualization and analysis of production details of speech segments requires more practice in acoustic analysis for the identification of vowel and consonant sounds, and for the measurement of phonetic details such as duration, frequency and intensity. Hence, teachers should decide whether visualization of phonetic details of vowels and consonants is worth pursuing given the well-known constraints in instruction times. However, the fact that timing factors (duration of individual sounds, words, and sentences) in L2 can greatly affect a speaker's intelligibility should not be underestimated and time should be spent emphasizing differences in language timing between the L1 and the L2.

As a result of the findings that speech visualization helps language learning, both commercial and university research teams have shown growing interest in exploring the applications and potentials of speech analysis software for language pronunciation instruction. Various systems have been developed for this purpose, including: *WinPitch LTL II* by Pitch Instruments Inc. (http://www.winpitch.com), *VisiPitch* by Kay-Elemetrics developed by Molholt (1998) and *VICK*, developed at the Speech Lab, Department of Electronics and Signal Processing,

University of Liberec, Czeck Republic (Nouza, 1999). They all provide computer-assisted pronunciation instruction with automatic audio and visual feedback. In addition to regular features such as prosodic real time display, variable speed playback, etc., *WinPitch LTL II* also has capabilities for processing multimedia files and for the automatic alignment of the learner's imitation of the teachers' model (Martin, 2004). *BetterAccent Tutor* (http://www.betteraccent.com/) provides audio-visual feedback of intonation, stress and rhythm in American English based on the assumption that these three factors have the biggest effect on intelligibility. Another product by Kay-Elemetrics, *Sona-Match*, provides real time representation of the learner's vowel space with different productions in the vowel space using different fonts. Carey (2004) gives a review of *Sona-Match* as well as a report of successful results using this system. In addition to the systems mentioned above, many others, both commercial and non commercial, are available, which try to integrate pedagogy and technology for pronunciation teaching. The reader should refer to Neri, Cucchiarini, Strik, and Boves (2002) and Chun (2006) for more detailed information on this issue.

8. Using signal analysis software to teach intonation in discourse: A pilot study conducted at the University of Padova

8.1 *Course and students*

In the 2005-06 academic year, a pilot study was conducted to investigate the feasibility and benefits of using speech analysis software as an aid in a pronunciation teaching class. The pilot study was conducted during an English Linguistics course for students of Foreign Languages for International Communication and of Modern EuroAmerican Languages, Literatures and Cultures at the University of Padova. Part of the English Linguistics course was devoted to the study of English phonetics and phonology for the improvement of basic English pronunciation skills. The module on English phonetics and phonology consisted of 20 hours of lessons in the language lab. Each lesson was divided into theory and practice. In the theoretical part, explanations were given of basic sound articulation and acoustics, and of the main differences between the Italian and English phonetic and phonological systems. In the practical part, the students were given exercises – mainly web-based – to improve their discrimination and production of English sounds or non-existent sound patterns in Italian.

About 30 students attended the class regularly, all with little or no previous formal instruction in English pronunciation or English phonetics and phonology. Based on a questionnaire that the students had to fill out at the beginning of the course, all the students considered English pronunciation to be extremely important for their future jobs and for successful communication in general.

8.2 *Using a visual display for learning English intonation contours*

The role of intonation seems to be particularly critical for L2 speech interpretation (Chun, 1998, 2002; Wennerstrom, 2000; Pickering, 2002, 2004). Italian speakers of English have major difficulties with English rhythm, particularly as it relates to vowel duration and vowel reduction patterns (Busà, 1995), which

are largely determined by the position of the syllable in relation to word and sentence stress, and emphatic stress. Hence, learning how to modulate intonation and assign stress in English could help Italians overcome their problems with English rhythm and thus improve their pronunciation.

The pilot study conducted during the phonetics and phonology module was designed to test whether Italian students can benefit from the visualization of their own productions of English sentences as compared to native speakers' productions of the same utterances. The public domain software *Praat* (available from http://www.praat.org) was used for this study. The grammatical functions of intonation in English were first explained to the students, both theoretically and with the aid of visual displays of intonation contours and speech waveforms. The students were asked to practice saying and recording a few English utterances exemplifying different intonation contours with different grammatical functions. With minor modifications, the utterances were the same as those reported in Chun (1998). The students were also given examples of the same utterances as spoken by two native speakers so that they could compare their own productions with those of the native speakers. They were instructed that each speaker may present individual variations but that there is usually a recognizable pattern that they should aim to pursue. For example, in English, falling and rising intonation patterns correspond with virtually no exceptions to statements and *yes-no* questions respectively. The students were also instructed to pay particular attention to the part of speech which was given prominence by the native speakers, and try to reproduce a similar prominence pattern.

Figures 1 and 2 are some examples of the visualizations obtained during the pilot study. Figure 1 shows the native speaker's sound wave (*upper box*) and pitch contour (*lower box*) of the question "Are you going?" The figure shows that the intonation is rising from the beginning of the utterance to the peak of prominence, corresponding to the vowel /o/, and then has a falling pattern before rising again after the vowel /i/ for the *yes-no* question.

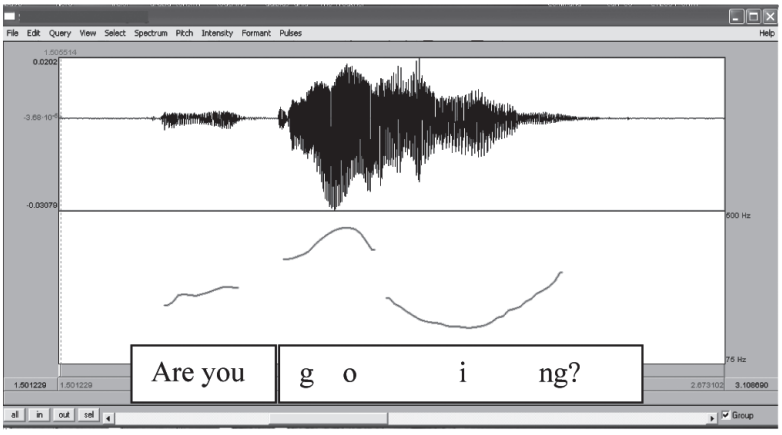

Fig. 1. Waveform (upper box) and intonation contour (lower box) of a native speaker's production of the question "Are you going?"

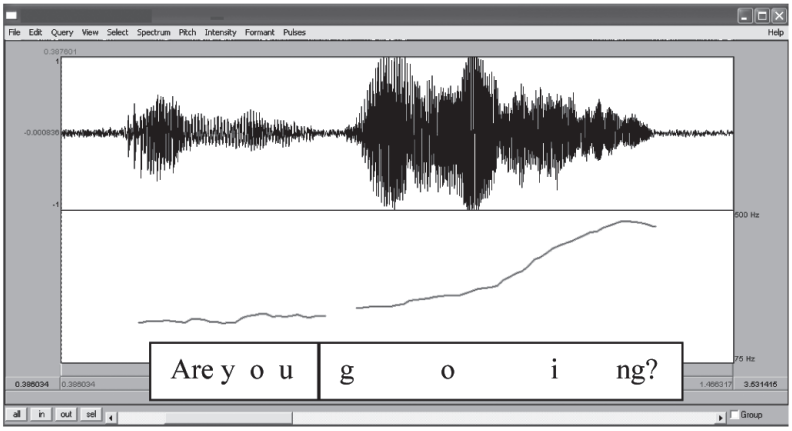

Fig. 2. Waveform (upper box) and intonation contour (lower box) of an Italian speaker's production of the question "Are you going?"

Figure 2 shows the sound wave and pitch contour of an Italian speaker's utterance of the question "Are you going?". Comparing these patterns with the native speaker's patterns in Figure 1, the differences are clearly visible even to a non-expert speech analyzer. The waveform (upper box) in Figure 2 shows, most noticeably, that the Italian produces the word "are"with much longer duration than the native speaker. As for the intonation contour, the Italian utterance does not have a prominent pitch in the word "going"; intonation is characterized by a single rising contour peaking on the final sound. The comparison of the two visualizations can easily provide students with enough detail to improve their productions and get closer to the target utterance.

Practice with the visualization of students' own speech utterances and comparison with the native speaker's was received favorably. The students considered this approach to be valuable and effective for improving their pronunciation in English and claimed that after several repetitions their intonation patterns tended to resemble those of the native speakers more closely. From the instructor's point of view, the overall experience was positive and will be repeated after adding more structure to the students' practice drills so as to set up ways to monitor and control the results of the students' practice. Crucial to this kind of approach is understanding whether the effects of visualizing and comparing speech utterances extend beyond the classroom and actually enhance students' understanding beyond a superficial level.

9. Conclusions

Many technological tools are being developed that assist learners in achieving communicative competence in L2. This paper has reviewed ways in which computer-assisted instruction can be used to enhance L2 pronunciation teaching and learning. Thanks to advances in research into language and speech, increased computer capabilities and lower computer costs, the number of ap-

plications available to both the teacher and the individual learner is increasing rapidly. Thus, these are exciting times for second language instruction. However, more research is needed to find teaching methods compatible with the new technology, as well as ways of improving and implementing classroom activities which can effectively and appropriately benefit from the use of technological tools. The last section in this paper has reported on a pilot experiment using a public domain speech analysis tool to help Italian students raise their awareness of English intonation and prosodic patterns. The method does not require students to have an in-depth knowledge of the phonetics and phonology of English and Italian, but does allow them to gain important insights into the differences between Italian and English prosody. Overall, the experience was viewed positively and worth a more thorough investigation to assess the effectiveness of the method proposed.

10. References

Ackerley, Katherine / Coccetta, Francesca (*in press.*), "Multimodality in an Online English course", in: Baldry, Anthony / Montagna, Elena (eds.), *Interdisciplinary Perspectives on Multimodality: Theory and Practice. Proceedings of the Third International Conference on Multimodality.* Campobasso, Palladino.

Anderson-Hsieh, Janet (1989), *Approaches Toward Teaching Pronunciation: A Brief History*, in: "Cross Currents" 16/2, 73-78.

Baldry, Anthony (2004), "Concordancing Film Corpora: The Role of MCA in Multimodal Discourse Analysis", in: Taylor, Carol / Busà, Maria Grazia / Gesuato, Sara (eds.), *Lingua Inglese e Mediazione Iinterlinguistica. Ricerca e Didattica con Supporto Telematico.* Padova, Unipress, 19-32.

Baldry, Anthony (2005), *A Multimodal Approach to Text Studies in English. The Role of MCA in Multimodal Concordancing and Multimodal Corpus Linguistics*. Campobasso, Palladino.

Baldry, Anthony / Taylor, Christopher (2004), "Multimodal concordancing and subtitles with MCA", in: Partington, Alan/Morley, John/ Haarman, Louann (eds.) *Corpora and Discourse.* Bern, Peter Lang, 57-70.

Baldry, Anthony / Thibault, Paul J. (2006), *Multimodal Transcription and Text Analysis. A Multimodal Toolkit and Coursebook with Associated Online Course.* London and New York, Equinox.

Beckman, Mary E. / Elam, Gayle A. (1997), "Guidelines for ToBI Labelling". (Version 3.0 March 1997). Retrieved at: http://www.ling.ohio-state.edu/-tobi/ame_tobi/labelling_guide_v3.pdf.

Bernstein, Jared (1997), *Automatic Spoken Language Assessment by Telephone.* "Technical Report" N. 5-97. Menlo Park, CA, Entropic Inc.

Bernstein, Jared / Balogh, Jennifer/ Lennig, Matt / Rosenfeld, Elizabeth (2005), "Workable Models of Standard Performance in English & Spanish". Paper presented at the EALTA Conference, Voss, Norway, 2nd - 5th June, 2005.

Bradford, Barbara (1992), *Intonation in Context.* Cambridge, Cambridge University Press.

Brazil, David (1994), *Pronunciation for Advances Learners of English (Teacher's Book).* Cambridge, Cambridge University Press.

Busà, Maria Grazia (1995), *L'Inglese degli Italiani. L'acquisizione delle Vocali.* Padova, Unipress.

Carey, Michael (2004), *CALL Visual Feedback for Pronunciation of Vowels: Kay Sona-Match*, in: "CALICO" 21/3, 571-601.

Celce-Murcia, Marianne (1987), "Teaching Pronunciation as Communication", in: Morley, Joan (ed.), *Current Perspectives on Pronunciation.* Washington, DC, TESOL, 1-12.

Celce-Murcia, Marianne / Brinton, Donna / Goodwin, Janet (1996), *Teaching Pronunciation: A Reference for Teachers of English as a Second or Foreign Language.* Cambridge, MA, Cambridge University Press.

Chang, Stephen (2002), *A Syllable, Articulatory-Feature, and Stress-Accent Model of Speech Recognition.* Doctoral Dissertation, University of California at Berkeley.

Chun, Dorothy M. (1998), *Signal Analysis Software for Teaching Discourse Intonation*, in: "Language Learning and Technology" 2/1, 61-77.

Chun, Dorothy, M. (2002), *Discourse Intonation in L2: From Theory and Research to Practice.* Amsterdam, John Benjamins.

Chun, Dorothy M. (2006), "Technological Advances in Researching and Teaching Phonology", in: Pennington, Martha C. (ed.), *Phonology in Context. Advances in Linguistics Series 1.* Palgrave, Macmillan. 274-299.

Cosi, Piero / Cohen, Michael M./ Massaro, Dominic W. (2002), "Baldini: Baldi speaks Italian". *Procee-*

dings ICSLP 2002, 7th International Conference on Spoken Language Processing. September 16-20, Denver Colorado.

Cunningham Florez, MaryAnn (1998), *Improving Adult ESL Learners' Pronunciation Skills*. Center for Adult English Language Acquisition, http://www.cal.org/caela/esl_resources/digests/Pronun.html.

Dalziel, Fiona / Metelli, Marianna (*in press*.), "Teaching about Genre with a Multimodal Concordancer", in: Baldry, Anthony/Montagna, Elena (eds.), *Interdisciplinary Perspectives on Multimodality: Theory and practice. Proceedings of the Third International Conference on Multimodality*. Campobasso, Palladino.

De Bot, Kees (1983), *Visual Feedback of Intonation 1: Effectiveness and Induced Practice Behavior*, in: "Language and Speech" 26/4, 331-350.

Delmonte, Rodolfo (2000), *SLIM Prosodic Automatic Tools for Self-Learning Instruction*, in: "Speech Communication" 30, 145-166.

Derwing, Tracey / Munro, Murray/ Carbonaro, Michael (2000), *Does Popular Speech Recognition Software Work with ESL Speech?*, in: "TESOL Quarterly" 34, 592-603.

Derwing, Tracey / Munro, Murray J. (2001), *What speaking rates do nonnative listeners prefer?*, in: "Applied Linguistics"22, 324-337.

Derwing, Tracey / Rossiter, Marian / Munro, Murray J. (2002), *Teaching Native Speakers to Listen to Foreign-Accented Speech*., in: "Journal of Multilingual and Multicultural Development" 23/4, 245-259.

Egan, Kathleen, B. (1999), *Speaking: A Critical Skill and a Challenge*, in: "CALICO Journal" 16/3, 277-293.

Eskenazi, Maxine (1999a), *Using a Computer in Foreign Language Pronunciation Training: What Advantages?*, in: "CALICO Journal" 16/3, 447-469.

Eskenazi, Maxine (1999b), *Issues in the Use of Speech Recognition for Foreign Language Tutors*, in: "Language Learning and Technology Journal" 2/2, 62-76. Retrieved at: http://llt.msu.edu/vol2num2/article3/index.html.

Eskenazi, Maxine / Hansma, Scott (1998), "The Fluency Pronunciation Trainer". *Proceedings STiLL Workshop on Speech Technology in Language Learning*, Marhollmen, May 1998.

Fraser, Helen (2000), *Literacy vs. Oral Communication Skills for ESL Learners*, in: "Literacy Link: Newsletter of the Australian Council for Adult Literacy" 19/3, 4-6.

Gerken, LouAnn / Aslin, Richard N. (2005), *Thirty Years of Research on Infant Speech Perception: The Legacy of Peter W. Jusczyk*, in: "Language Learning and Development" 1/1, 5-21.

Granström, Björn (2004), "Towards a Virtual Language Tutor". *Proc. of InSTIL/ICALL Symposium: NLP and Speech Technologies in Advanced Language Learning Systems*, Venice, Italy 17-19 june 2004, 1-8.

Hardison, Debra M. (2004), *Generalization of Computer-Assisted Prosody Training: Quantitative and Qualitative Findings*, in: "Language Learning & Technology" 8/1, January 2004, 34-52.

Harless, William, G. / Zier, Marcia, A. / Duncan, Robert C. (1999), *Virtual Dialogues with Native Speakers: The Evaluation of an Interactive Multimedia Method*, in: "CALICO Journal" 16/3, 313-337.

Kim, Yoon / Franco, Horacio / Neumeyer, Leonardo (1997), "Automatic Pronunciation Scoring of Specific Phone Segments for Language Instruction", *Proc. Eurospeech 1997*, Rhodes, Greece, 645-648.

Kipp, Michael (2004), *Gesture Generation by Imitation - From Human Behavior to Computer Character Animation*. Boca Raton, Florida, Dissertation.com.

Kormos, Judit / Dénes, Mariann (2004), *Exploring Measures and Perceptions of Fluency in the Speech of Second Language Learners*, in: "System" 32, 145-164.

Lambacher, Stephen (1996a), "Taching English Pronunciation Using a Computer Visual Display". Paper presented at the IATEFL 29th International Annual Conference, York, England, 1995. Retrieved at: http://www.u-aizu.ac.jp/-steeve/york95.html.

Lambacher, Stephen (1996b), *Spectrogram Analysis as a Tool in Developing L2 Pronunciation Skills*, in: "Speak Out!", London, IATEFL, 32-35.

Levow, Gina-Anne / Mari Broman Olsen, (1999), "Modeling the Language Assessment Process and Result: Proposed Architecture for Automatic Oral Proficiency Assessment". *Proceedings of Computer-mediated Language Assessment and Evaluation in Natural Language Processing Workshop*, ACL '99, 24-31.

Lieberman, Philip (1986), "The Acquisition of Intonation by Infants: Psychology and Neural control", in: C. John-Lewis (ed.), *Intonation in Discourse*. London, Croom Helm, 239-257.

Macdonald, Shem (2003), *Pronunciation – Views and Practices of Reluctant Teachers*, in: "Prospect. An Australian Journal of TESOL" 17/3. Retrieved at: http://www.nceltr.mq.edu.au/prospect/17/pros17_3.html.

Martin, Philippe (2004), "WinPitch LTL II, a Multimodal Pronunciation Software". *Proc. of InSTIL/ICALL Symposium: NLP and Speech Technologies in Advanced Language Learning Systems*, Venice, Italy 17-19 June 2004, paper 042.

Massaro, Dominic W. (2006a), "The Psychology and Technology of Talking Heads: Applications in Language Learning", in: O. Bernsen / L. Dybkjaer / J. van Kuppevelt (eds.), *Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*. Dordrecht, The Netherlands, Kluver Academic Publishers, 183-214.

Massaro, Dominic W. (2006b),

"A computer-Animated Tutor for Language Learning: Research and Applications", in: P.E. Spencer / M. Marshark (eds.), *Advances in the Spoken Language Development of Deaf and Hard-of-Hearing Children*. New York, NY, Oxford University Press.

Mixdorff, Hansjörg (2002), "Speech Technology, ToBI and Making Sense of Prosody". *Proceedings Speech Prosody 2002*, Aix-en-Provence, France, 11-13 April 2002, 31-38.

Molholt, Garry (1998), *Accent Reduction Via Acoustic Analysis: ESL Applications for the Visi-Pitch II, Model 3300*. Lincoln Park, New Jersey, Kay Elemetrics Corporation.

Morley, Joan (1991), *The Pronunciation Component in Teaching English to Speakers of Other Languages*, in: "TESOL Quarterly" 25/3, 491-520.

Mostow, Jack / Aist, Gregory (1999), *Giving Help and Praise in a Reading Tutor with Imperfect Listening–Because Automated Speech Recognition Means Never Being Able to Say You're Certain*, in: "CALICO Journal" 16/3, 407-424.

Munro, Murray J. (in press), "Foreign Accent and Speech Intelligibility". To appear in: J. Hansen /M. Zampini (eds.), *State-of-the-Art Issues in Second-Language Phonology*. Cambridge, CUP.

Munro, Murray J. /Derwing, Tracey (2001), *Modelling Perceptions of the Comprehensibility and Accentedness of L2 Speech: The Role of Speaking Rate*, in: "Studies in Second Language Acquisition" 23, 451-468.

Neri, Ambra / Cucchiarini, Catia / Strik, Helmer / Boves, Lou (2002), *The Pedagogy-Technology Interface in Computer Assisted Pronunciation Training*, in: "CALL Journal" 15/5, 441-467.

Neumeyer, Leonardo / Franco, Horacio / Digalakis, Vassilios / Weintraub, Mitchel (2000), *Automatic Scoring of Pronunciation Quality*, in: "Speech Communication" 30, 83-93.

Nouza, Jan (1999), "Computer-aided Spoken-language Training with Enhanced Visual and Auditory Feedback". *Proceedings Eurospeech '99*, 6th European Conference on Speech Communication and Technology, September 5-9, 1999 Budapest, Hungary, 183-186.

Pickering, Lucy (2002), "Patterns of Intonation in Cross-Cultural Communication Exchange Structure in NS TA & ITA Classroom Discourse". *Proceedings of the Seventh Annual Conference on Language, Interaction and Culture*. University of California, Santa Barbara, 4. 1-17.

Pickering, Lucy (2004), *The Structure and Function of Intonational Paragraphs in Native and Nonnative Instructional Discourse*, in: "English for Specific Purposes" 23, 19-43.

Raux, Antoine / Eskenazi, Maxine (2004), "Using Task-Oriented Spoken Dialogue Systems for Language Learning: Potential, Practical Applications and Challenges". *Proc. of InSTIL/ICALL Symposium: NLP and Speech Technologies in Advanced Language Learning Systems*, Venice, Italy 17-19 June 2004, paper 035.

Seneff, Stephanie / Wang, Chao / Zhang, Julia (2004), "Spoken Conversational Interaction for Language Learning". *Proc. of InSTIL/ICALL Symposium: NLP and Speech Technologies in Advanced Language Learning Systems*, Venice, Italy 17-19 June 2004, paper 035.

Silipo, Rosaria / Greenberg, Steven (2000), "Prosodic Stress Revisited: Reassessing the Role of Fundamental Frequency". *Proceedings of the NIST Speech Transcription Workshop*. College Park, MD, May 16-19, 2000, 1-6.

Spaai, Gerard W.G. / Hermes, Dik J. (1993), *A Visual Display for the Teaching of Intonation*, in: "CALICO Journal" 10, 19-30.

Stibbard, Richard (1996), *Teaching English Intonation with a Visual Display of Fundamental Frequency*, in: "The Internet TESL Journal" 2/8, August 1996, http://iteslj.org/Articles/Stibbard-Intonation/.

Truong, Khiet / Neri, Ambra / Cucchiarini, Catia / Strik, Helmer (2004), "Automatic Pronunciation Error Detection: an Acoustic-Phonetic Approach". *Proceedings of the InSTIL/ICALL Symposium*, Venice 2004, 135-138.

Truong, Khiet / Neri, Ambra / de Wet, Febe / Cucchiarini, Catia / Strik, Helmer (2005), "Automatic Detection of Frequent Pronunciation Errors Made by L2-Learners". *Proceedings of InterSpeech 2005 (IS2005)*, Lisbon, 4-8 Sept. 2005, 1345-1348.

Yates, Lynda (2001), "Teaching Pronunciation in the AMEP: Current Practice and Professional Development". *AMEP Research Centre*. Retrieved from http://www.nceltr.mq.edu.au/conference2001/index.html.

Wachowicz, Krystyna, A. / Scott, Brian (1999), *Software That Listens: It's Not a Question of Whether, It's a Question of How*, in: *Tutors that Listen: Speech Recognition for Language Learning*, "CALICO Journal" 16/3, 253-276.

Wennerstrom, Ann (2000), "The Role of Intonation in Second Language Fluency", in: H. Riggenbach (ed.), *Perspectives on Fluency*. Ann Arbor, MI, University of Michigan Press.

Wennerstrom, Ann (2001), *The Music of Everyday Speech*. Oxford, Oxford University Press.

Wilson, Deirdre / Wharton, Tim (in press), *Relevance and Prosody*, in: "UCL Working Papers in Linguistics" 17.