De Meo, A., Vitale, M., Pettorino, M., Cutugno, F., & Origlia, A. (2013). Imitation/self-imitation in computer-assisted prosody training for Chinese learners of L2 Italian. In. J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*, Aug. 2012. (pp. 90-100). Ames, IA: Iowa State University.

IMITATION/SELF-IMITATION IN COMPUTER-ASSISTED PROSODY TRAINING FOR CHINESE LEARNERS OF L2 ITALIAN

Anna De Meo, University of Naples "L'Orientale"
Marilisa Vitale, University of Naples "L'Orientale"
Massimo Pettorino, University of Naples "L'Orientale"
Francesco Cutugno, University of Naples "Federico II"
Antonio Origlia, University of Naples "Federico II"

Recent studies on L2 acquisition, speech synthesis and automatic identification of foreign accents argue for a major role of prosody in the perception of non-native speech. Research on the relationship between pronunciation improvement and student/teachers' voice similarities has also shown that the better the match between the learners' and native speakers' voices in terms of f_0 and articulation rate, the more positive the impact on pronunciation training. This study investigates the effects of imitation and self-imitation on the acquisition of L2 suprasegmental patterns. Degree of foreign accent, improvements in intelligibility, and effectiveness of communication were measured to determine the success of each technique. For this purpose, a prosodic transplantation technique and a computer-assisted learning methodology were used.

The study was conducted with 26 Chinese speakers of L2 Italian. The stimuli used for the pronunciation training consisted of four speech acts (granting, order, request and threat) uttered by the 26 Chinese and two Italian native speakers, and the items necessary for the self-imitation training were obtained through prosodic transplantation, i.e. transferring suprasegmental features from native speakers' voices to the L2 ones. Chinese students divided into two different groups practiced imitation and self-imitation and the self-imitation impact was evaluated by comparing pre- and post-training performances of both groups. Both teaching strategies promoted an improvement in learners' performances; however, the self-imitation training proved to result in more accurate prosodic realizations.

INTRODUCTION

Pronunciation improvement is a relevant issue in the area of spoken language technology for language learning (Chun, 2013; Eskenazi, 2009; Levis, 2007; Martin, 2012). However, during the last two decades, only a few studies, carried out on learners with different L1s and focusing on different target languages, have investigated the relationship between the student/teacher voice similarity and pronunciation improvement (Bissiri, Pfitzinger, & Tilmann, 2006; Jilka & Möhler, 1998; Nagano & Ozawa, 1990; Peabody & Seneff, 2006; Sundström, 1998; Tang, Wang & Seneff, 2001). Results from these studies have shown that the better the match between the learners' and native speakers' voices in terms of f_0 and articulation rate, the more positive the impact on pronunciation achievement, suggesting the existence of a user-dependent golden speaker (Probst, Ke, & Eskenazi, 2002). As claimed by Felps, Bortfeld, & Guiterrez-Osuna (2009), it would be beneficial for L2 students to be able to listen to their own voices producing utterances in a native accent. As a consequence, the most effective golden speaker to learn

segmental and suprasegmental features of a second language is the learner's own voice with a native accent.

This study investigates the effects of the self-imitation strategy, i.e. the speaker's imitation of his/her own voice properly modified according to the target native model on the acquisition of L2 suprasegmental patterns, comparing the results with those achieved with traditional imitation exercises. For this purpose, a prosodic transplantation technique and a computer-assisted learning methodology were used. To determine the success of each technique, four different variables were considered: speech act identification, communication effectiveness, degree of foreign accent and improvements in intelligibility.

MATERIALS AND METHOD

Technique

This study is based on the use of the rhythmic-prosodic transplantation technique (Pettorino & Vitale, 2012; Yoon, 2007), which makes it possible to transfer one or more acoustic parameters (pitch, intensity, articulation rate, frequency and duration of silent pauses) from a native speaker (the "donor") to a non-native speaker (the "receiver"), without altering the segmental sequence and the identity of the synthesized voice. This technique is based on the PSOLA (Pitch-Synchronous Overlap and Add) algorithm (Charpentier & Moulines, 1989), implemented in Praat (Boersma, 2001).

The transplantation procedure involves a fixed sequence of steps, divided into five phases: treatment of anomalies, segmentation, transplantation of duration, transplantation of intensity, and pitch contour superimposition. All these operations have been automatized through a Praat script and then applied to the voices selected for this study.

Stimuli

For the purpose of this research, two Italian sentences were chosen. The meaning of these two sentences can vary by using different pitch contours, even if the syntactic structure is kept unchanged. Human languages, indeed, generally allow the speaker to express the modal meaning of the sentence, i.e., attitude towards the message content, by using different pitch contours (Sorianello, 2006). For example, a sentence such as "We were sharing a hamburger" can be uttered and interpreted as a question or a statement, depending on the different modulation of the fundamental frequency (f_0) contours.

For this study, four speech acts were considered (*request*, *order*, *granting* and *threat*). The first three were already tested in previous studies (De Meo & Pettorino 2011; De Meo, Pettorino & Vitale, 2012, Pettorino, De Meo & Vitale, 2012), while *threat* was newly introduced in order to create a balanced set of stimuli, composed of the same number of pragmatically basic and complex utterances. In fact, the four considered speech acts are characterized by different degrees of familiarity for NNSs. Requests and orders are always introduced early in the learning process and, at the same time, are more frequent in the language input. Grantings and threats, by contrast, are rarely presented in advanced level language courses.

The sentences and the different meanings considered are the following:

Sentence 1 - "Lascia i piatti sul tavolo"
granting ("Ok, you can leave the dishes on the table.") (Figure 1)
order ("Leave the dishes on the table!") (Figure 2)

Sentence 2 - “Ne parliamo stasera a casa”
request (“Shall we talk about it at home tonight?”) (Figure 3)
threat (“We will talk about it at home tonight.”) (Figure 4)

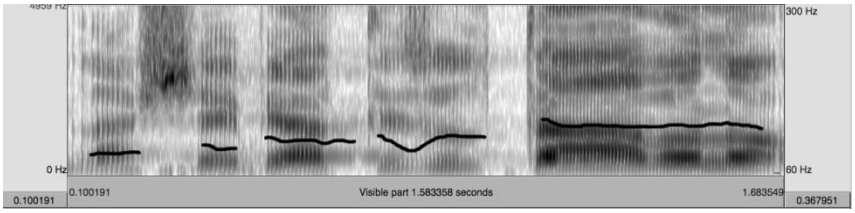


Figure 1. Granting, male voice.

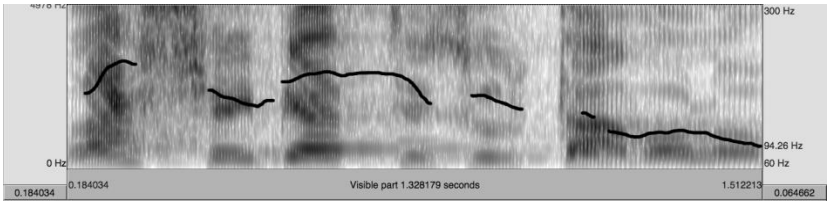


Figure 2. Order, male voice.

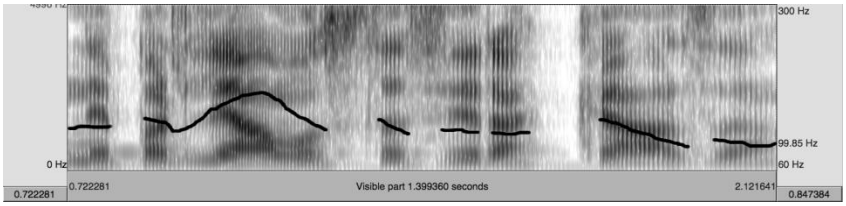


Figure 3. Request, male voice.

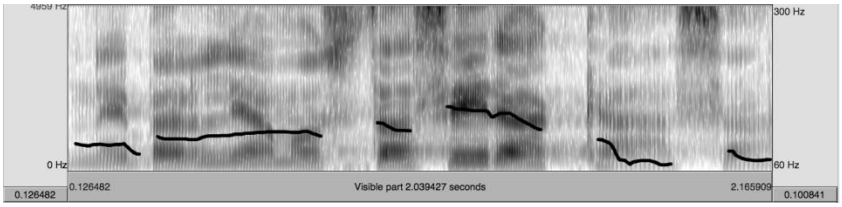


Figure 4. Threat, male voice.

The four sentences were recorded by the NSs and the NNSs involved in this study in an anechoic chamber. With the help of a professional translator, it was ensured that the non-native subjects had understood the true meaning associated with each speech act before the recording phase.

Subjects

Speakers

The subjects involved in this study were 6 native Italians (3 males and 3 females) and 30 L1 Chinese speakers (25 females and 5 males), all having a high-intermediate level of competence in Italian (B2 of the Common European Framework of Reference). Both groups were university students living in the Campania Region, Southern Italy.

Listeners

Two different groups of native Italian listeners were involved in this study: 30 subjects (Group 1) and 52 subjects (Group 2). All the listeners, male and female with an average age of 24, lived in

Naples and were accustomed to the same diatopic variety of Italian as the six native Italian speakers.

PROCEDURE

The entire experimental procedure can be divided into five main steps:

1. two pre-tests to select the native speakers (NSs) to be involved in the study and to confirm the presence, already detected by researchers, of a strong foreign accent in the non-native voices;

2. a perceptual test to select the non-native speakers (NNSs) to be trained;

3. a rhythmic-prosodic transplantation of the NNSs' utterances;

4. imitation and self-imitation prosodic training;

5. a final perceptual test to compare imitation and self-imitation post-training performances.

Step 1

In order to select the most suitable native and non-native subjects for this study, two different pre-tests were carried out.

Since the prosodic model to be used for both the imitation and the self-imitation trainings was offered by the native Italian speakers, it was important to choose the most communicatively accurate native male and female voices. Furthermore, we decided to select only Chinese learners having a strongly accented L2 speech, in order to easily observe the post-training effects.
To this end, the four utterances for the stimuli were recorded by the six NSs and administered in random order to the listeners of Group 1, who were asked to:

- identify the speech act (multiple choice task);
- evaluate the communication effectiveness on a five-point-scale (1=min, 5= max);
- evaluate the degree of foreign accent on a three-point-scale (native accent, mild foreign accent, strong foreign accent);
- assess intelligibility on a three-point-scale (poor, sufficient, good).

Regarding the Chinese voices, all NNSs were instructed to read a short text in Italian, then assessed by the listeners of Group 1 in a second test session. Hence, for each non-native voice the degree of foreign accent (native accent; mild foreign accent; strong foreign accent) was evaluated.

As far as the L1 Italian subjects are concerned, the pre-test was used to choose the best male and female performances for each of the four provided utterances. As was to be expected, although all the NSs received a positive evaluation (the average correct speech act identification was of 72%), two of them appeared to be more communicatively accurate than the others (correct recognition: male voice 82%; female voice 90%). These two speakers were the only voices to be finally involved in the study, since there were no other significant differences among all the NSs, in terms of both degree of foreign accent and intelligibility.

With respect to the second pre-test, results showed that 26 out of 30 NNSs were judged as prevalently strongly foreign accented (73%). The remaining 4 NNSs were excluded from the experiment.

Step 2

As the 26 Chinese learners had a high-intermediate level of competence in Italian, they were already able to produce some acceptable utterances. It was important to exclude from the training sessions utterances that were properly uttered and therefore perfectly acceptable from the start. This was accomplished by administering a perceptual test to Group 2 listeners.

For this second step, the four sentences recorded by the 26 NNSs and by the 2 NSs were randomly arranged. The L1 Italian utterances were used as control elements to set the threshold of acceptability for the L2 productions. A total of 112 stimuli (28 x 4) was administered in a perceptual test, specifically devoted to the evaluation of all the involved subjects on each of the following aspects:

- speech act identification (multiple choice task);
- communication effectiveness (five-point scale rating: 1=min, 5=max)
- intelligibility (three-point scale assessment: poor, sufficient, good).

The minimum percentage of correct speech act identification obtained by the NSs, which was approximately 60%, represented the threshold for the selection of the NNSs productions to be accepted for the training phase. In other words, all the speech acts that gained a percentage of correct identification higher than 60% were considered acceptable and not in need of training. On the basis of the set threshold, only 68 NNSs' utterances were selected (26 grantings, 21 threats, 10 orders and 11 requests). Figure 5 shows the mean percentage values of correct speech act identification concerning only the 8 NSs' and the 68 NNSs' utterances selected for the training.

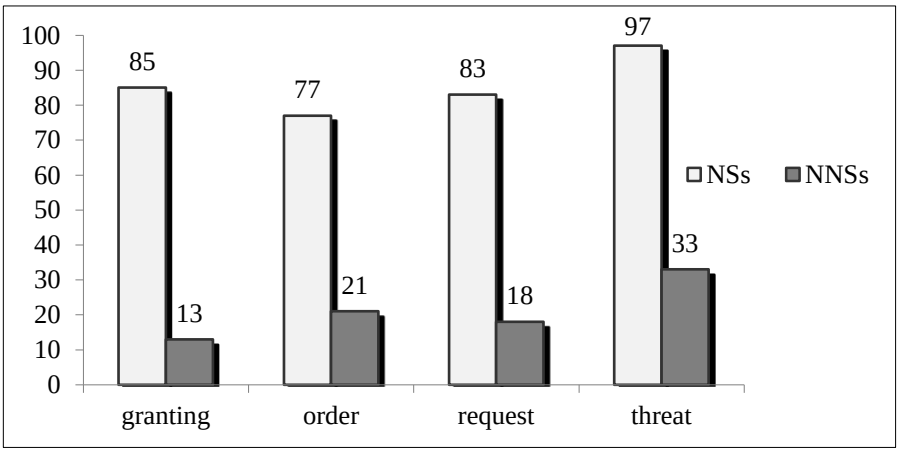


Figure 5. Correct speech act identification (mean percentage values). NSs and NNSs to be trained in comparison.

It should be noted that for the 55% of the errors produced in the speech acts recognition assessment, the judging audience gave "order" as their answer, probably because the prevailing flat pitch of the Chinese L2 Italian speech coincides with the pitch profile of the order in L1 Italian. The remaining 45% of the errors were distributed fairly randomly among the other speech acts (granting 14%, request 19%, threat 12%).

The gap existing between native and non-native speakers' achievements is evident also in terms of communication effectiveness, since Italians were given an average score of 4.7, while the Chinese learners only reached a 2.5 level. No clear variations were observed between the

different speech acts (Figure 6). As for intelligibility, the NSs, as expected, were fully understood by all the listeners ("good" intelligibility: 98%), while the NNSs, although they also got 61% "good" evaluations, were judged as just sufficiently intelligible by the 33% of the evaluators and not understandable at all by the remaining 6%.

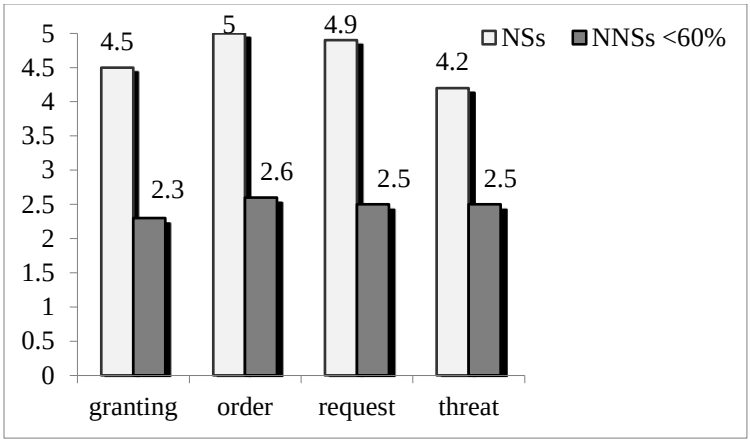
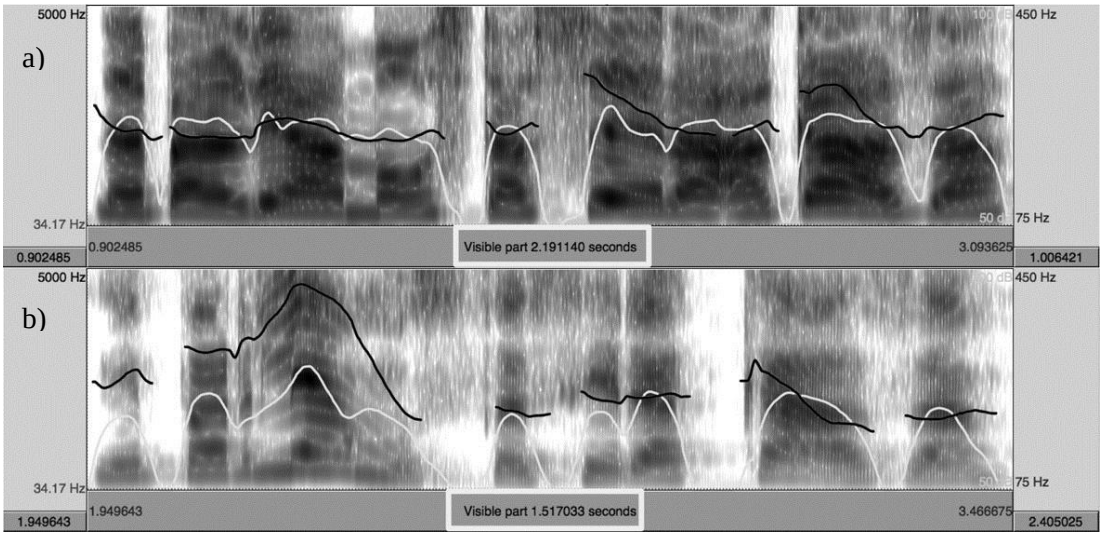


Figure 6. Communication effectiveness per speech act of the NSs and the selected NNSs (1=min, 5=max).

Step 3

All the NNSs correctly produced at least one of the proposed speech acts. Of the 68 NNSs' items finally selected, half underwent the imitation treatment and the other half the self-imitation one. Although a partially random subdivision was carried out, an attempt was made to maintain a balanced distribution of the two treatments in terms of speech acts and speakers. The prosodic transplantation was performed only on the 34 items to be used for the self-imitation training.



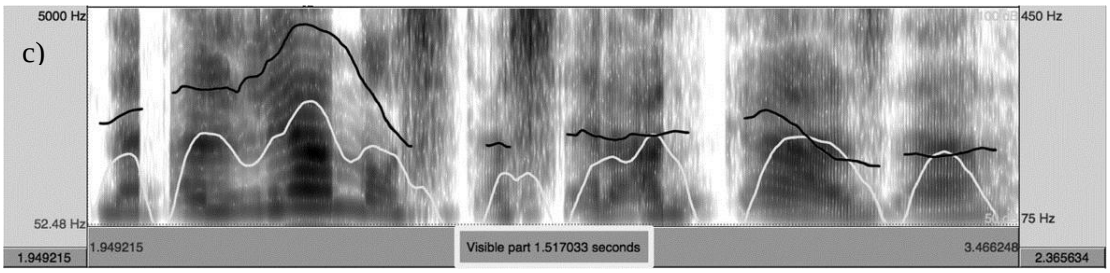


Figure 7. Request "Ne parliamo stasera a casa?", female voices. a) Chinese original voice; b) Italian model voice; c) Chinese synthesized voice. Black line: pitch contour – White line: intensity.

As can be seen in Figure 7, the Chinese original voice (Figure 7a) receives the same segmental duration, intensity contour and pitch movement of the Italian "donor's" voice (7b) resulting from the transferring of the rhythmic-prosodic features (7c). With the exception of micro variations that are not perceptually relevant, the temporal extension of the whole sentence (as indicated in the box "visible part" below each spectrogram) and, more specifically, the duration of each segment of the manipulated utterance are comparable to those of the Italian model voice. With regards to the pitch contour, as a result of the transplantation procedure, the flattened movement of the original Chinese voice (7a) was transformed into a much more varied intonation, with a marked peak on the third syllable and slighter pitch shifts on the second part of the utterance (Figure 7c). As shown by the continuous grey line on the spectrogram, with the transplantation even the energy peaks are perfectly repositioned according to the model voice.

Step 4

The audio files to be used for the imitation and self-imitation trainings were arranged into 26 exercise packages, each containing a different series of audio files to be used by one of the 26 NNSs. Each NNS was asked to listen to the utterances produced by a native speaker's voice or by his/her own synthesized voice contained in his/her own package and to exercise in imitating the input. After 5 minutes of training, learners were instructed to record the output sentences.

Step 5

The collected post-training performances were used to arrange a final perceptual test to compare the improvement induced by the two treatments. To this end, both the pre- and post- training productions were assembled in random order and administered to the Group 2 listeners, who were asked to perform the same task as in the Step 2 perceptual test. In this case the foreign accent assessment was added (three-point scale evaluation: native accent, mild foreign accent, strong foreign accent) in order to evaluate the impact of the exercises that were carried out.

RESULTS

As it can be seen in Table 1, both trainings improved the NNSs' rhythmic-prosodic performances, although the results obtained by the self-imitation exercises seem to be more relevant for the order and the request. Results are statistically significant ($p < 0.001$, ANOVA).

Table 1
Speech Act Identification Improvement (Δ values: post training % – pre training %)

	Imitation	Self-imitation
Granting	+47%	+50%
Order	+22%	+48%
Request	+47%	+71%
Threat	+26%	+33%

Communication effectiveness also underwent a slight general improvement, but even in this case the self-imitation treatment was more effective, especially for the granting and the request functions (Table 2). Results are statistically significant ($p < 0.001$, ANOVA).

Table 2
Communication Effectiveness Improvement (five-point scale evaluation, Δ values: post training – pre training)

	Imitation	Self-imitation
Granting	+0.9	+1.3
Order	+0.5	+0.3
Request	+0.3	+1.5
Threat	+0.3	+0.4

The two trainings under investigation produced very similar improvements in terms of intelligibility (Table 3), while in the foreign accent reduction the results obtained by the self-imitation are slightly more positive: the post-imitation utterances have produced a decrease of the "strong foreign accent" in favour of the "mild foreign accent", whereas the post self-imitation productions were judged as being of a native accent by a small percentage of the listeners (Table 4). Even in this case results prove to be statistically significant ($p < 0.001$, ANOVA).

Table 3
Intelligibility Improvement (Δ values: post training % – pre training %)

	Imitation	Self-imitation
Poor	+4%	+4%
Sufficient	-21%	-22%
Good	+17%	+19%

Table 4
Foreign Accent Improvement (Δ values: post training % – pre training %)

	Imitation	Self-imitation
Native accent	0	+6%
Mild foreign accent	+27%	+19%
Strong foreign accent	-27%	-25%

CONCLUSIONS

This study shows that computer-assisted prosody training based on both imitation and self-imitation produces good results in terms of pronunciation improvement, providing a spin-off for

prosody learning, communication effectiveness and intelligibility improvement, and foreign accent reduction. However, self-imitation, made possible by the use of the rhythmic-prosodic transplantation technique, generally achieves more satisfactory results.

These results led us to develop a project that aims at creating a software program, ProsoTrainer, devoted to the prosodic pronunciation improvement for learners of L2 Italian, favouring at the same time foreign accent reduction and the improvement of communication effectiveness. By means of the prosodic-intonational transplantation procedure, the suprasegmental features of the native speaker (pitch, intensity, articulation and speech rate, frequency and duration of pauses) would be cloned and transferred in real time to the L2 learner's voice, without altering the perception of the L2 speaker's identity. The learner's voice thus becomes the "native" model to imitate. However, speakers with different L1s and different levels of L2 competence, and a greater number of speech acts have to be considered in order to get sufficient data to support the development of a technological tool that makes teaching and/or autonomous learning of the L2 suprasegmental features easier.

ABOUT THE AUTHORS

Anna De Meo is Professor at the University of Naples “L’Orientale,” Italy. She teaches Multiculturality and Language Acquisition and Specialized Translation. Her main fields of interest are L2 acquisition (with special reference to Italian), L2 speech perception and production, interlanguage pragmatics, and translation. Email: ademeo@unior.it +39-0816909031 University of Naples “L’Orientale” - via Nuova Marina 59 - 80133 Naples - Italy

Marilisa Vitale is a PhD student in Linguistics at the University of Naples “L’Orientale,” Italy, and currently trainee at the LIMSI-TLP (Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur - Groupe Traitement du Langage Parlé) of the CNRS in Paris. Her main fields of interest are rhythmic-prosodic aspects of L2 Italian, foreign accent and credibility correlates, speech manipulation, and teaching L2 Italian prosody. Email: marilisavitale@hotmail.it +39-0816909031 University of Naples “L’Orientale” - via Nuova Marina 59 - 80133 Naples - Italy

Massimo Pettorino is Professor of Experimental Phonetics at the University of Naples “L’Orientale,” Italy, and Director of the GSCP, a special interest group on spoken communication of the Italian Linguistic Society (SLI). He works in the field of acoustic and prosodic analysis of speech. Email: mpettorino@unior.it +39-0816909912 University of Naples “L’Orientale” - via Duomo 219 - 80139 Naples - Italy

Francesco Cutugno is Assistant Professor at the University of Naples “Federico II.” His main research interests concern speech technologies and information retrieval in spoken language corpora. He is also interested in automatic prosodic analysis, syllable segmentation and rhythm. He is responsible for the Language Understanding and Speech Interfaces (LUSI) Lab in the Department of Physics. Email: cutugno@unina.it +39-081676850 Department of Physics – University of Naples “Federico II” - via Cinthia - 80126 Naples - Italy

Antonio Origlia obtained a Master degree in Computer Science at the University of Naples, “Federico II” in 2009 and is currently undertaking a PhD in mathematical and computational sciences at the Language Understanding and Speech Interfaces (LUSI) Lab. His area of interest mainly covers affective computing with focus on emotional speech. He is also interested in prosodic analysis tools development, automatic syllable segmentation, and prominence detection

and rhythm. Email: antonio.origlia@unina.it +39-081679961 Department of Physics – University of Naples "Federico II" - via Cinthia - 80126 Naples – Italy

REFERENCES

Bissiri, M. P., Pfitzinger, H.R., & Tillmann, H.G. (2006). Lexical stress training of German compounds for Italian speakers by means of resynthesis and emphasis. In *Proceedings of the 11th Australian International Conference on Speech Science & Technology* (pp. 24–29). University of Auckland, New Zealand.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International 5*, 341-345.

Charpentier, F., & Moulines, E. (1989). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. In *Proceedings of the First European Conference on Speech Communication and Technology - Eurospeech* (pp. 2013-2019). Paris: European Speech Communication Association.

Chun, D.M. (2013). Computer-assisted pronunciation teaching. In C.A. Chapelle (ed.) *The Encyclopedia of Applied Linguistics* (pp. 823-834). Oxford, UK: Wiley-Blackwell.

De Meo, A. & Pettorino, M. (2011). L'acquisizione della competenza prosodica in italiano L2 da parte di studenti sinofoni. In E. Bonvino & S. Rastelli (eds.), *La didattica dell'italiano a studenti cinesi e il progetto Marco Polo*. Proceedings of the 15th Workshop AICLU (pp. 67-78). Pavia: Pavia University Press.

De Meo, A., Pettorino, M., & Vitale, M. (2012). Comunicare in una lingua seconda. Il ruolo dell'intonazione nella percezione dell'interlingua di apprendenti cinesi di italiano. In M. Falcone & A. Paoloni (eds.), *La voce nelle applicazioni*. Proceedings of the 7th Congress of Italian Association of Speech Sciences AISV (pp. 117-129). Roma: Bulzoni editore.

Eskenazi, M. (2009). An overview of spoken language technology for education. *Speech Communication, 51*(10), 832-844.

Felps, D., Bortfeld, H., & Gutierrez-Osuna, R. (2009). Foreign accent conversion in computer assisted pronunciation training. *Speech Communication, 51*(10), 920-932.

Jilka, M., & Möhler, G. (1998). Intonational foreign accent: speech technology and foreign language teaching. In P*roceedings of ESCA Workshop on Speech Technology in Language Learning* (pp. 115–118).

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics, 27*, 184-202.

Martin, P. (2012). Automatic prosodic comparison between model and imitation sentences in a second language teaching computerized environment. In A. De Meo & M. Pettorino (eds.), *Prosodic and rhythmic aspects of L2 acquisition. The case of Italian* (pp. 263-276). Newcastle upon Tyne: Cambridge Scholars Publishing.

Nagano, K., & Ozawa, K. (1990). English speech training using voice conversion. In *1st International Conference on Spoken Language Processing* (ICSLP 90) (pp. 1169-1172). Kobe, Japan.

Peabody, M., & Seneff, S. (2006). Towards automatic tone correction in non-native mandarin. In *Proceedings of the 5th international conference on Chinese spoken language processing* (pp. 602-603). Kent Ridge, Singapore.

Pettorino, M., De Meo, A., & Vitale, M. (2012). La competenza prosodico-intonativa nell'italiano L2. Analisi e sintesi del segnale fonico di cinesi, vietnamiti e giapponesi. In S. Ferreri (ed.), *La linguistica educativa*. *Proceedings of the 44th International Congress of Italian Linguistics Society SLI* (pp. 329-342). Roma: Bulzoni.

Pettorino, M., & Vitale, M. (2012). Transplanting native prosody into second language speech. In M.G. Busà Maria Grazia, A. Stella (eds.), *Methodological Perspectives on Second Language Prosody. Papers from ML2P 2012* (pp. 11-16). Padova: CLEUP.

Probst, K., Ke, Y., & Eskenazi, M. (2002). Enhancing foreign language tutors - in search of the golden speaker. *Speech Communication 37*(3-4), 161-173.

Sorianello, P. (2006). *Prosodia*. Roma: Carocci.

Sundström, A. (1998). Automatic prosody modification as a means for foreign language pronunciation training. In *Proceedings of ISCA Workshop on Speech Technology in Language Learning* (STILL 98) (pp. 49-52). Marholmen, Sweden.

Tang, M., Wang, C., & Seneff, S. (2001). Voice transformations: From speech synthesis to mammalian vocalizations. In *Proceedings of Eurospeech 2001*, Aalborg, Denmark.

Yoon, K. (2007). Imposing native speakers' prosody on non-native speakers' utterances: The technique of cloning prosody. *Journal of the Modern British & American Language & Literature 25*(4), 197-215.