

A self-imitation technique for the improvement of prosody in L2 Italian

Debora Vigliano¹, Kei Yoshimoto², Elisa Pellegrino³

¹⁻² Tohoku University, Graduate School of International Cultural Studies

³ L'Orientale University of Naples, Department of Literary, Linguistic and Comparative Studies

debora.vigliano.q7@dc.tohoku.ac.jp, kei.yoshimoto.d2@tohoku.ac.jp, epellegrino@unior.it

1. Introduction

When L2 learners train their pronunciation they usually listen to native speakers' voices and try to imitate their performance. However, in this process they have to disregard the idiosyncratic characteristics of the native voice (e.g. voice quality, F0, etc.) in order to focus on the segmental and suprasegmental characteristics of the target language (L2). According to Probst, Ke & Eskenazi (2002) the acquisition of second language pronunciation would be facilitated if the learners imitated an L1 speaker with similar prosodic features (such as speed of articulation and F0). Felps, Bortfeld & Gutierrez-Osuna (2009) suggested that there is no speaker who matches these requirements but the L2 learner himself with native accent, after undergoing a process of accent conversion. Thanks to voice conversion techniques, such as the prosodic transplantation, the learners' voice is modified to match the prosody of a reference native speaker. The result of this process is a synthetic speech having the same voice characteristics of the L2 learner but with the prosodic patterns of the native speaker. Thus, the learner would train their pronunciation skills through self-imitation, i.e. imitating his voice with native accent. Comparing the performance produced through the imitation of L1 speaker with those produced with the self-imitation technique, L2 learners would be enabled to increase the awareness of the mistakes in their speech productions and, consequently, to accommodate to the native pronunciation (Nagano & Ozawa, 1990).

During the last two decades, a number of studies has proven this technique to be pedagogically effective. Nagano and Ozawa (1990) tested the use of prosodic modification to teach English to Japanese speakers; Bissiri et al. (2006) evaluated a prosodic-conversion method to the purpose of teaching German to Italian learners; Peabody and Seneff (2006) investigated a similar technique to train English learners of Mandarin Chinese. As for Italian L2, recent experiments based on the prosodic transplantation technique were conducted on Chinese learners with different levels of linguistic competence by De Meo et alii (2013) and by De Meo, Vitale & Pellegrino (in press).

2. The study

2.1. Objectives

The purpose of this study is twofold: on the one hand it is intended to explore the pedagogical effectiveness of the self-imitation training method on another group of Italian learners differing for L1, quantity and quality of language exposure from the learners involved in the above mentioned studies. The subjects are in fact seven Japanese speakers, students at Tokyo University of Foreign Studies with an upper intermediate level of linguistic competence.

On the other hand, given the difference in rhythm between Italian and Japanese (Ramus, Nespor & Mehler, 1999), further aim of the study was to assess whether segment durations, among the various phonetic features contributing to prosodic variation, were susceptible of variations after self-imitation.

2.2. Technique and Subjects

This study is based on the rhythmic-prosodic transplantation technique involving the transfer of some acoustic parameters such as pitch contour and articulation rate, from a native speaker ("donor") to a non-native speaker ("receiver"). Seven Japanese learners of Italian, two males and five females, aged between 21 and 28, took part to the research as "receivers". The Non Native Speakers (NNSs) had been studying Italian in a formal educational environment for 5 to 6 years and had no hearing or language impairments.

Donors of their acoustic parameters were two native Italian speakers (NSs, henceforth), one male and one female aged 27 and 25, respectively. Both had been living in Japan for 6 months at the time the research was being conducted.

3. Procedure

The study was carried out into two training sessions:

- Pre self-imitation session, where students were asked to read sentences according to different communicative intentions relying only on their prosodic skills in Italian (cf. 3.1)
- Self-imitation training session, where they trained their pronunciation skills, imitating their voice previously shaped on the basis of the native prosodic model (cf. 3.2)

3.1. Pre-training session

NNSs and NSs were involved in a read speech activity. The stimuli were two Italian sentences (1. Accendi la radio/ eng. You turn on the radio; 2. Chiudi la finestra/ eng. You close the window). The learners were asked to read the sentences on a computer screen, according to three different communicative intentions: request (R), order (O) and granting (G).

Sentence 1: Accendi la radio

- (Request) Accendi la radio? / eng. Can you turn on the radio?
- (Order) Accendi la radio! /eng. Turn on the radio!
- (Granting) Accendi la radio / eng. Ok, you can turn on the radio.

Sentence 2: Chiudi la finestra

- (Request) Chiudi la finestra? / eng. Can you close the window?
- (Order) Chiudi la finestra! / eng. Close the window!
- (Granting) Chiudi la finestra / eng. Ok, you can close the window.

Native Japanese speakers specialized in Italian language and linguistics translated the sentences into their L1 for the learners, as to prevent misunderstandings, especially of the intended pragmatic function.

Sentence 1	
REQUEST	Accendi la radio?
質問	ラジオつけてくれない？
ORDER	Accendi la radio!
命令	ラジオつけて！
GRANTING	Accendi la radio.
譲歩	ラジオつけていいよ
Sentence 2	
REQUEST	Chiudi la finestra?
質問	窓閉めてくれない？
ORDER	Chiudi la finestra!
命令	窓閉めなさい！
GRANTING	Chiudi la finestra.
譲歩	窓閉めていいよ

In this phase of the study (pre-training session), the seven subjects had to perform the task, neither listening to a native model, nor receiving any clue about how to differentiate the three tunes. Recordings were performed after individual periods of separated training. The recordings were taken in single sessions, in the silent room of Tokyo University, at 44.100 Hz sampling rate. The same recording protocol was used with the two native Italian speakers. The corpus of read speech collected in the pre-training session (pre-training productions, henceforth) consisted of:

- 42 utterances in L2 Italian (7 NNSs * 2 sentences * 3 communicative intentions)
- 14 requests, 14 orders, 14 grantings
- 12 utterances in L1 Italian (2 NSs * 2 sentences * 3 communicative intentions)
- 4 requests, 4 orders, 4 grantings.

This kind of task was chosen because it was supposed to be challenging for Japanese learners. Although also in Japanese, a different modulation of the fundamental frequency (f0) allow a sentence such as いい会社じゃない (ii kaisha janai) to be interpreted as a statement (It isn't a good company) or a question (it's a good company, isn't it?), three pragmatic meanings mentioned above, are cued by a series of syntactical and lexical devices in addition to prosodic ones (Abe, 1998). In Italian, on the contrary, the lack of morphological and syntactical means for distinguishing sentence modality, make pitch contours the most important clues to shape the pragmatic function of an utterance (D'Imperio, 2002). As a consequence, even if the syntactic structure of the two sentence is kept unchanged, learners have to transmit the three modalities only suprasegmental features.

3.2. Self-imitation prosodic training

The prosodic transplantation procedure, based on the PSOLA (PitchSynchronous Overlap and Add) algorithm (Charpentier & Moulines, 1989) and implemented in Praat (Boersma, 2001), represents a preliminary step for the self-imitation prosodic training.

The transplantation procedure involved a fixed sequence of steps, divided into four phases:

- manual segmentation of the utterances of the native and non-native speakers in consonantal and vocalic portions,
- treatment of anomalies in the case of L2 utterances,

- transplantation of segment duration and pitch contour superimposition from the donors' utterances (Italian native speakers) to the corresponding utterances produced by the receivers (Japanese learners)

The last two operations were automatized through a Praat script and then applied to the voices selected for this study.

For the transplantation procedure the criterion of donor-to-receiver gender match was followed. The voice of the male native speaker was paired with the voices of male non native speakers, whereas the voice of female native speaker served as a 'model' for the utterances produced by the female Japanese learners.

After the prosodic transplantation procedure, a new corpus of 42 synthesized utterances was built. These utterances underwent self-imitation treatment. During this session, each learner trained to mimic their utterances with native accent as many times as they needed to approximate the model. When they felt confident, they recorded the new performance. Consequently, a new corpus of 42 post-training productions (14 requests, 14 commands, 14 grantings) was collected.

4. The perception test

In order to assess the improvement of prosodic skills in Italian after the self-imitation prosodic training a perception test was carried out. The 42 pre- and 42 post-training productions were presented to 17 native Italian listeners, who possessed a general familiarity with foreign accents but no knowledge of Japanese. The test was administered online through the software SurveyGizmo.

The listeners listened to the 84 utterances split in three groups, each consisting of 28 items. Ten minutes of interval between one session and the other was provided to avoid an overload of information. After each stimulus, they had to :

- identify the conveyed pragmatic functions choosing among five given options, three expected ('request', 'order', 'granting') and two distractors ('statement' and 'other');
- rate the degree of foreign accentedness on a five-point scale (1 = native accent; 5 = strong foreign accent).

4.1. Results

The first point of this analysis will be the examination of the relationship between intended and perceived pragmatic functions of pre- and post-training productions, thus inferring which prosodic contours are most often confused by L2 students. As shown by the confusion matrix depicting the pre-training phase in table 1, granting is the most difficult speech act to perform. Generally confused with orders (47.68%), less than 10% of Italian listeners recognized this modalities in the utterances read by the learners. The recognition threshold rises to about 40% with orders, mostly confused with requests (32.35%). Our data reveals that the request performed by Japanese learners was the most successfully recognized intention (52.75%).

Post-training results (table 2), however, indicate a more accurate identification of all the speech acts, presenting a great decrease in the confusion between intended and perceived pragmatic functions, especially as regards grantings and orders.

A comparison between the percentage of correct answer obtained before and after the prosodic training (table 3) will enable a better assessment of the validity of self-imitation. Furthermore, in order to gain insight on the speech acts for

which self-imitation was most effective, we will contrast the percentage of correct match between intended and perceived pragmatic meanings by training phase and speech act (table 4).

As it is shown in table 3, the average percentage of correct match between intended and perceived pragmatic functions in the post-training phase exceeds the one obtained in the pre-training phase of about 26 points. The results of statistical analysis (ANOVA with repeated measure) indicate that there is a significant main effect of training [$F(1,32) = 65.18, p < .001$].

Mean scores of correct match were also calculated for the single speech acts (requests, orders and grantings) (table 4).

		Perceived pragmatic functions				
		O	G	R	S	Other
Intended pragmatic functions	O	39.92%	10.92%	32.35%	13.87%	2.94%
	G	47.68%	8.44%	20.25%	18.57%	5.06%
	R	16.88%	5.06%	52.74%	11.81%	13.50%

Table 1: Confusion matrix between intended and perceived pragmatic functions in pre-training phase.

		Perceived pragmatic functions				
		O	G	R	S	Other
Intended pragmatic functions	O	57.98%	11.34%	14.29%	14.71%	1.68%
	G	11.34%	47.06%	17.23%	17.23%	7.14%
	R	12.61%	4.20%	75.21%	5.88%	2.10%

Table 2: Confusion matrix between intended and perceived pragmatic functions in post-training phase.

	Pre-training (A)	Post-training (B)	Difference (B - A)
Average	33.61%	60.04%	+ 26.43

Table 3: Mean percentage of correct match between intended and perceived pragmatic functions by training phase.

	Pre-training (A)	Post-training (B)	Differences (B - A)
Requests	52.52%	75.21%	22.69
Orders	39.92%	57.98%	18.06
Grantings	8.40%	47.06%	38.66

Table 4: Mean percentage of correct match between intended and perceived pragmatic functions by speech act and training phase.

The differences between pre- and post-training phases were statistically significant [$F(2; 32) = 32.13, p < 0.001$] for

the three speech acts under study. The results show that the self-imitation prosodic training can assist L2 learners in improving their rhythmic and prosodic skills, leading to a better understanding by native speakers. However, this training technique exerts a different influence depending on the speech act, as indicated by the fourth column in table 4 (Differences B-A). With a percentage of correct identification shifting from 8.40% in the pre-training phase to nearly 50% in the post-training phase, grantings obtained the highest improvement. Indeed, the statistic analysis of data reveals significant interactions between training and speech act [$F(2;32) = 3.51, p < 0.005$].

On the contrary, the self-imitation prosodic training does not seem to produce any meaningful effect on the degree of foreign accent. Our data show that the average rate of accentedness does not change significantly before and after training (Pre: 3.43; Post: 3.53).

5. Spectro-acoustic analysis

Contrastive spectro-acoustic analysis of the pre- and post-training productions was carried out in order to highlight the acoustic features most susceptible of variations after self-imitation.

Given the difference in rhythm between Italian and Japanese languages, among the various phonetic features contributing to prosodic variation, in this study the attention was focused on segment duration. The acoustic analysis was performed by means of Praat. Each sentence produced by the NSs and by the NNSs pre- and post- training was measured by total duration as well as by vocalic portion and consonant cluster. The following differences were then calculated:

- duration of pre-training utterance - duration of the corresponding L1 utterance (L2 pre-L1)
- duration of post-training utterance - duration of the corresponding L1 utterance (L2 post-L1)
- comparison of the differences
- comparison of vowel and consonant duration



Figure 1: difference between L2 and L1 mean duration of utterances by speech act and training phase (0 = same duration as the model).

As shown in Figure 1, the results of analysis indicate that the utterance duration varies after self-imitation. Indeed, the mean duration of all speech acts approximates the native model after the prosody training. In contrast with the tendency of foreign speech to be characterized by lower articulation rate our data show that the pre-training productions were too fast compared to the native model. The best improvement obtained by Grantings on the perceptual level finds its counterpart on the acoustic one, since this is the speech act which presents a major improvement in sentence duration as a result of training.

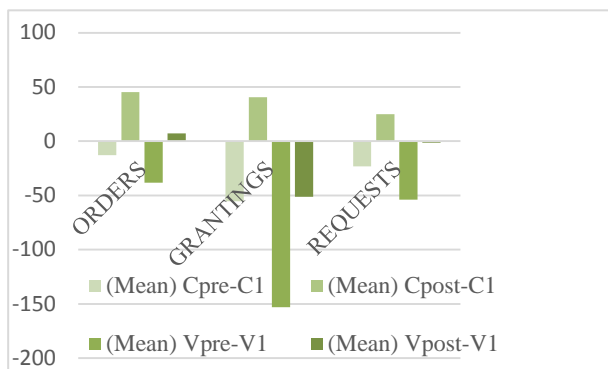


Figure 2: Comparison between vocalic and consonantal portions

As regards the contribution given by vowels and consonants to the match of the target duration, Figure 2 shows that the major role in this sense was played by the vowel portion. Indeed, the better match between vowel duration in L2 and L1 indicates that a clear improvement in vowel production occurred after self-imitation training. Consonant production, on the other hand, worsen considerably, resulting in an over lengthening.

6. Conclusions

This study shows that prosody training based on self-imitation is effective in terms of pronunciation improvement for Japanese learners of L2 Italian. As a result, also communication effectiveness was enhanced. As demonstrated by the increase in the post-training match between intended and perceived pragmatic function, self-imitation technique boosts learners' ability to reproduce intonation patterns corresponding to the native listeners' expectations. Indeed, all the speech acts under study improve, especially Grantings.

Being rarely presented even in advanced level language courses, Granting is the most difficult speech act to produce among the three intentions proposed. Therefore, they follow requests and orders in the recognition threshold both before and after the training. Directives indeed are the most frequently used speech acts in classroom interaction (Searle, 1969) and for this reason they are more accurately conveyed by learners.

The results of spectro-acoustic analysis carried out on pre- and post-training productions showed that segment duration underwent a major variation after self-imitation, resulting in a better match with target duration. Particularly, the general improvement is entirely due to an improvement in vowel production. On the contrary, consonants are on average longer than before, thus negatively affecting the target duration. However, since the training led to an improvement in speech

act recognition on the perceptual level, it is possible to assume that vowel duration plays an important role in correct identification.

According to De Meo et al. (2013) self-imitation generally achieves more satisfactory results than traditional imitation exercise. For this reason, further steps of this research will involve a comparison between the results achieved by Japanese learners of L2 Italian with self-imitation strategy and those obtained by imitating a native speaker. On the acoustic level, further investigations will involve the comparison of pitch variations and peak dislocations before and after the prosodic training.

7. References

- Abe, I. 1998. "Intonation in Japanese," *Intonation Systems- A survey of twenty languages*, Hirst D., Di Cristo, A. (ed.), Cambridge: Cambridge University Press, 363-378.
- 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," *Proceedings of the 11th Australian International Conference on Speech Science & Technology*, New Zealand: University of Auckland, 24-29.
- Boersma, P. 2001. Praat, a system for doing phonetics by computer, *Glott International* 5, 341-345.
- Charpentier, F., Moulines, E. 1989 "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Proceedings of the First European Conference on Speech Communication and Technology - Eurospeech*, Paris: European Speech Communication Association, 2013-2019.
- D'Imperio, M. 2002. "Italian Intonation: an overview and some questions". *Probus* 14, Issue 1, 37-69.
- 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." *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*, Levis J., LeVelle K. (eds.) Ago. 2012, Ames, IA: Iowa State University, 90-100.
- De Meo, A., Vitale, M., Pellegrino, E. (in press), "Tecnologia della voce e miglioramento della pronuncia in una L2: imitazione e autoimitazione a confronto. Uno studio su cinesi apprendenti di italiano L2," *Atti del XV Convegno Nazionale dell'Associazione Italiana di Linguistica Applicata (AItLA)*, «Linguaggio e apprendimento linguistico:metodi e strumenti tecnologici», Università del Salento, Lecce.
- Felps, D., Bortfeld, H., Gutierrez-Osuna R. 2009. "Foreign accent conversion in computer assisted pronunciation training", *Speech Communication* 51, 920-932.
- Jilka, M., Möhler, G. 1998. "Intonational foreign accent: speech technology and foreign language teaching", *Proc. ESCA Workshop on Speech Technology in Language Learning*, 115-118
- Nagano, K., Ozawa, K. 1990. "English speech training using voice conversion", *1st Internat. Conf. on Spoken Language Processing (ICSLP 90)*, Kobe, Japan, 295-308.
- Peabody, M., Seneff, S. 2006. "Towards automatic tone correction in nonnative mandarin", *Chinese Spoken Language Processing: 5th International Symposium, ISCSLP 2006*, 602-613.
- Probst, K., Ke, Y., Eskenazi, M. 2002. "Enhancing foreign language tutors - In search of the golden speaker", *Speech Communication* 37, 161-173.
- Ramus, F. Nespors, M. and Mehler J. 1999. Correlates of linguistic rhythm in the speech signal, *Cognition* 73(3), 265-292.
- Searle, J.R. 1969. *Speech Acts. An Essay in the Philosophy of Language*, Cambridge: Cambridge University press.
- Sundström, A. 1998. "Automatic prosody modification as a means for foreign language pronunciation training", *Proc. ISCA Workshop on Speech Technology in Language Learning (STILL 98)*, Marholmen, Sweden, 49-52.