

Self-imitation in prosody training: A study on Japanese learners of Italian

Elisa Pellegrino¹, Debora Vigliano²

¹⁻² University of Naples “L’Orientale”, Italy
epellegrino@unior.it, dvigliano@yahoo.it

Abstract

The proficiency in a second language is fully attained only if students have learnt to modulate the rhythmic and prosodic parameters equivalent to those of the native speakers. This study is aimed to test the pedagogical effectiveness of the self-imitation technique for the purpose of developing a native-like prosodic competence. Seven intermediate Japanese learners of Italian (NNSs) and 2 native Italian speakers (NSs) were involved in a read speech activity. NSs and NNSs were asked to read and record two Italian sentences conveying three different pragmatic functions (granting, order, request). NNSs performed the task twice, before and after the self-imitation prosodic training. The items used for the training were obtained by transferring the suprasegmental features of the native speakers, used as donors, to the Japanese learners, considered as the receivers. During the training phase, Japanese learners mimic their utterances previously modified to match the prosody of the reference native speaker, and then recorded the new performance. Seventeen native Italian listeners rated pre- and post-training productions for pragmatic function and accentedness. The results indicate that self-imitation promoted an improvement in learners' performances in terms of communicative effectiveness. Conversely, average rate of accentedness does not change significantly before and after training.

Index Terms: L2 prosody teaching and learning, foreign accent, prosodic transplantation technique

1. Introduction

Adult second language acquisition is a very complex process that is usually characterized by general failure and variation in success [1]. The proficiency in a second language is not fully attained if students have just interiorized the phonological, morpho-syntactic and lexical rules of the target language. Even advanced L2 learners, indeed, may commit pragmatic failures if they are not aware of the cross-linguistic differences in speech act realization rules [2]. For example, members of one culture are likely to perform requests more or less directly than members of another culture [3]. Requests addressed to social inferiors might tend to be phrased more directly than requests addressed to superiors, or vice versa [4].

Additionally, advanced learners can be perceived as foreigner to the social or ethnic group which native listeners think to be part of if their spoken productions present phonetic and prosodic deviations from standard pronunciation. Although the suprasegmental features of speech play a crucial role in everyday spoken communication and in foreign accent detection [5]-[7], for long time they have been relegated to a purely expressive function [8]. Moreover, the near-exclusive attention paid by foreign language pedagogy on segmental accuracy overshadowed the importance of prosody and

intonation in second language acquisition [9]-[14]. Nevertheless, the end state of L2 learners matches the competence of an L1 speaker only if students have learnt to modulate the rhythmic and prosodic parameters equivalent to those of the native speakers [15]-[18].

The improvement of the prosodic competence in an L2 is a current issue in the area of spoken language technology for education and language learning [19]-[21]. During the last decades, studies on computer assisted pronunciation training (CAPT) have emphasized the importance of the student/teacher voice similarity for the enhancement of pronunciation skills. Particularly interesting are the outcomes presented by Probst et al. [22]. They found that the better the match between the learners' and native speakers' voices, the more positive the impact on pronunciation training. L2 English students who mimicked a well-matched native speaker in terms of articulation rate and f0 were more accurate than students who imitated a poor match. These results suggested the existence of a user-dependent “golden speaker” and, additionally, led Felps et al. [23] to assume that the optimal ‘golden speaker’ for learners is ‘their native-accented selves’. It would be a great advantage for L2 learners to listen their own voices previously modified to match the prosody of the reference native speaker. Pedagogically speaking, the most effective technique to achieve native-like prosodic competence would be self-imitation. In other words, learners should imitate their own voice producing native accented utterances. According to [23], the process of foreign accent conversion will also enable students to understand more easily the differences between their foreign accented utterances and their ideal native counterparts. In this way, it would be possible to overcome one of the major limitations suffered by CAPT software - the lack of sufficient correct feedback [24].

The effectiveness of prosodic-conversion methods has already been tested on Japanese learners of L2 English [25], on Italian learners of L2 German [26] and on English learners of Mandarin Chinese [27]. However, the use of prosodic modification to teach Italian prosody was only recently investigated [28], [29].

Preliminary researches were conducted on Chinese learners with an intermediate and elementary level of linguistic competence. These studies were based on the prosodic transplantation technique [30], [31], based on the PSOLA (Pitch-Synchronous Overlap and Add) algorithm [32]), implemented in Praat [33]. Through this technique, the acoustic parameters (pitch, intensity, articulation rate, duration of pauses) of the native Italian speakers (the “donors”) were transferred to L2 speakers (the “receivers”), without altering the segmental sequence. Results of these studies have shown the effectiveness of the prosodic transplantation technique in developing a native-like prosodic competence. Chinese students trained to mimic utterances of their own voices with native accent were rated more communicatively accurate than

those who imitated utterances from a reference Italian speaker. The self-imitation technique had a positive impact also on accentedness for intermediate learners. Post training productions were rated more native-like than pre-training performance.

2. The study

2.1. Objectives and participants

The purpose of this study is to extend the investigations on pedagogical effectiveness of self imitation prosodic training on another group of students: Japanese learners whose first language (L1) is distant typologically, phonetically and rhythmically from Italian.

To the purpose, the research was conducted at the Tokyo University of Foreign Studies, in Japan. Seven Japanese learners of Italian were involved into the study. The Non-Native Speakers of Italian (NNSs, henceforth), were 2 males and 5 females, were aged between 21 and 28 and had an upper intermediate level of linguistic competence (B2 of the Common European Framework of Reference). They had been studying Italian in their country, in a formal learning environment for 5-6 years. Moreover they had studied Italian language and linguistics for one year in some Italian universities. They had no hearing or language impairments.

Two native Italian speakers (NSs, henceforth), one male and one female, aged 27 and 25 respectively, took part to the research as ‘donors’ of their prosodic parameters to the Japanese learners, considered as ‘receivers’. NSs had been living in Japan for 6 months when the research was being conducted.

2.2. 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). Due to the lack of morphological and syntactical means for distinguishing sentence modality, in Italian intonation plays a crucial role in shaping the pragmatic function of an utterance [34]. In other words, a sentence like *Leggi il giornale* (eng. You read the newspaper) could be uttered as a question, a statement or as an order by exclusively manipulating its melodic contour.

Basing on this specific characteristics of Italian language, the two sentences had to be read with 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 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.

It is important to underline that this kind of task is supposed to be challenging for Japanese learners, whose L1 is constrained by syntactical, lexical and prosodic devices to vary

the pragmatic meaning of an utterance [35]. In order to prevent NNSs from misunderstanding the meaning of the Italian sentences, and particularly the pragmatic function to convey, the sentences and the intended communicative intentions were translated to their L1. The translations were made by a native Japanese speaker specialized in Italian language and linguistics.

Participants were instructed to read aloud the sentences from a computer screen, modulating the pitch contour to perform the three different speech acts. The utterances were displayed as follows:

| | |
|----------------|---------------------|
| Frase 1 | |
| RICHIESTA | Accendi la radio? |
| 質問 | ラジオつけてくれない? |
| COMANDO | Accendi la radio! |
| 命令 | ラジオつけて! |
| CONCESSIONE | Accendi la radio |
| 譲歩 | ラジオつけていいよ |
| Frase 2 | |
| RICHIESTA | Chiudi la finestra? |
| 質問 | 窓閉めてくれない? |
| COMANDO | Chiudi la finestra! |
| 命令 | 窓閉めなさい! |
| CONCESSIONE | Chiudi la finestra. |
| 譲歩 | 窓閉めていいよ |

In this phase of the research (Pre-training phase, henceforth) they had to read the sentences according to the three pragmatic meanings, neither listening to a native model, nor receiving any clue about how to differentiate the three tunes. They were only allowed to train separately and then, when they felt confident, the recordings were performed. 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 production, henceforth) consisted of:

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

2.3. Self-imitation prosodic training

The self imitation prosodic training requires the execution of the prosodic transplantation procedure as a preliminary step. In order to transfer the acoustic parameters from the utterances produced by the native Italian speakers (the “donors”) to the corresponding ones performed by the Japanese learners (the “receivers”), a series of operations were realized:

- manual segmentation of the utterances produced by NSs and NNSs in consonantal and vocalic portions, by means of the software Praat,

- treatment of anomalies so that the segments of the utterances produced by the ‘donors’ (NSs) can be aligned to those produced by the ‘receivers’ (NNSs),
- transplantation of duration,
- pitch contour superimposition.

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 NS was paired with the voices of male NNSs. The voice of female NS, instead, served as a ‘model’ for the utterances produced by the female Japanese learners.

After the manipulation 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 need to approximate the model. When they felt confident, they recorded the new performance. Consequently, a new corpus of 42 post-training performances (14 requests, 14 commands, 14 grantings) was collected. We will refer to these performance as post-training productions.

2.4. The perception test

The effectiveness of the self-imitation prosodic training for the improvement of the prosodic competence in Italian was tested by means of a perception test. The 42 pre-training productions and the 42 post-training productions were randomly arranged and divided into three groups of 28 items each, interspersed with a break of 10 minutes in order to avoid information overload.

Seventeen native Italian listeners, aged between 23 and 30, familiar with different foreign accents, but with no prior knowledge of Japanese, listened to the three groups of items. For each of them they were asked:

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

The test was administered online through the software SurveyGizmo.

2.5. Results

In the analysis of data, we will start to examine the relationship between expected and perceived pragmatic functions for the pre- and post- training productions. This will allow to infer the prosodic contours that are mostly confused by L2 learners. As for the data regarding the pre-training phase, the confusion matrix in table 1 shows that for Japanese learners, the request is the easiest speech act to perform. This was correctly recognized by 52.74% of Italian listeners. The percentage of correct identification falls below the 40% with orders, that is mostly confused with requests (32.35%). The recognition threshold falls below the 10% with granting, generally confused with order (47.68%).

After the self-imitation prosodic training (tab. 2) all speech acts were more neatly recognized. Even more, the confusion between intended and perceived pragmatic meanings decreases considerably for orders and grantings. In order to better assess the validity of self-imitation, we will compare the percentage of correct answer obtained before and after the prosodic training (table 3). Then, we will contrast the percentage of correct match between intended and perceived

pragmatic meanings for training phase and speech act (table 4). These data enable to gain insight on the speech acts for which self imitation was mostly effective.

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 (repeated measure ANOVA) 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).

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 | 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 2. *Confusion matrix between intended and perceived pragmatic functions in post-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 3. *Mean percentage of correct match between intended and perceived pragmatic functions by training phase.*

| | Pre-training (A) | Post training (B) | Differences (B – A) |
|---------|------------------|-------------------|---------------------|
| Average | 33.61% | 60.04% | 26.43 |

Table 4. *Mean percentage of correct match between intended and perceived pragmatic functions by speech act and 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 |

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. These results thus suggests that self-imitation prosodic training improves the ability of L2

learners to modulate the rhythmic and prosodic parameters as it was expected by native listeners. However this training technique exerts a different influence depending on the speech act. The statistic analysis of data also reveals significant interactions between training and speech act [$F(2;32) = 3.51, p < 0.005$]. Indeed, as shown by the fourth column of table 4 (Differences B-A), the best improvement is obtained by Grantings. The percentage of correct identification indeed shifts from 8.4% in the Pre-training phase to 47.06% of the Post-training phase.

As regards the validity of self-imitation prosodic training to weaken the strength of foreign accent, our data seem to indicate that this technique does not seem to produce any meaningful effect. The average rate of accentedness does not change significantly before and after training (Pre: 3.43; Post: 3.53)

3. Conclusions

This work aimed to assess the validity of self-imitation technique for the improvement of pronunciation and communication effectiveness in L2 Italian. The study involved Japanese learners of Italian with an upper intermediate level of linguistic competence.

The results showed that self-imitation prosodic training helps learners memorize and reproduce intonation patterns corresponding to the native listeners' expectations. In the pragmatic function identification task, the percentage of correct match between intended and perceived communicative intentions increases significantly after the training session. The improvement regards all the three speech acts under study, especially grantings.

In line with previous studies on Italian [28],[29], request is the easiest speech act to perform by L2 Italian learners. It is immediately followed by order, that is unambiguously recognized by more than half of native listeners. Japanese learners' ability to better convey requests and orders than granting is not surprising. In fact, language learners do not have the same degree of difficulty about the three speech acts. Directives (requests and orders) are the most frequently used speech act in classroom interaction [36] and, thus, they are present in the input since the early stage of interlanguage development. On the contrary, granting is rarely presented in advanced level language courses.

As for the effectiveness of self-imitation for foreign accent reduction, in this study the training session has not played any relevant role. Similarly to the outcomes found with elementary Chinese learners of Italian [29], no differences were found between the pre- and post-training phases. Even though the subjects examined in this study have an upper intermediate level of morpho-syntactic competence in the target language, their productions were not so accurate on segmental level. It is important to remember that the Japanese subjects had a limited exposure to native input, since they had been studying Italian above all in classroom setting and with Japanese teachers. Therefore, a training specifically focused on the suprasegmental features of speech does not ensure the reduction of foreign accentedness.

In order to study in depth the effectiveness of self imitation prosodic training, further steps of this research will involve learners with different mother tongues and level of linguistic competence. Additionally, contrastive spectro-acoustic analysis of the pre- and post-training productions will be carried out in order to highlight the acoustic features most susceptible of variations after self imitation.

4. References

- [1] R. Bley-Vroman, "What is the logical problem of foreign language learning?," *Linguistic Perspectives on Second language Acquisition*, S. Gass, J. Schachter (eds.), Cambridge New York: Cambridge University Press, 1989.
- [2] L. M. Kurisack, "The effects of individual-level variables on speech act performance", *Speech Act Performance. Theoretical, empirical and methodological issues*, A. Martínez-Flor and E. Usó-Juan ed., Amsterdam/Philadelphia: John Benjamins Publishing Company, 2010.
- [3] S. Blum-Kulka, "Learning to say what you mean in a second language: a study of the speech act performance of Hebrew second language learners", *Applied Linguistics* HI/1: 29-59, 1982.
- [4] S. Blum-Kulka and E. Olshtain, "Requests and apologies: A cross-cultural study of speech act realization patterns" (CCSARP), *Applied Linguistics*, vol.5, no.3, pp. 196-213, 1984.
- [5] M. Piat, D. Fohr, and I. Illina, "Foreign accent identification based on prosodic parameters", *Interspeech*, 2008, pp. 759-762.
- [6] T. Piske, I. R. A. MacKay, J. E. Flege, "Factors affecting degree of foreign accent in an L2: a review," *Journal of Phonetics*, vol. 29, pp. 191-215, 2001.
- [7] P. Boula de Mareüil and B. Vieru-Dimulescu, "The Contribution of Prosody to the Perception of Foreign Accent," *Phonetica. Int. Journal of Phonetic Science*, vol.63, no.4, December, pp. 247-267, 2006.
- [8] A. De Meo and M. Pettorino (eds.), *Prosodic and Rhythmic Aspects of L2 Acquisition. The case of Italian*, Newcastle upon Tyne: Cambridge Scholars Publishing, 2012.
- [9] J. E. Flege, O. S. Bohn, and S. Jang, "Effects of experience on non-native speaker's production and perception of English vowels," *Journal of Phonetics*, vol. 25, 1997.
- [10] J. E. Flege, E., Frieda, and T. Nozawa, "Amount of native-language (L1) use affects the pronunciation of an L2," *Journal of Phonetics*, vol. 25, pp. 169-186, 1997.
- [11] J. E. Flege, I. R. A. MacKay, and D. Meador, "Native Italian speakers' perception and production of English vowels," *Journal of the Acoustical Society of America*, vol. 106, pp. 2973-2988, 1999.
- [12] J. E. Flege, "Assessing constraints on second-language segmental production and Perception", *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, A. Meyer and N. Schiller, (eds.), Berlin: Mouton de Gruyter, pp. 319-355, 2003.
- [13] R. C. Major, *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*, Mahwah, NJ: Erlbaum, 2001.
- [14] C. T. Best, "A direct realistic view of cross-language speech perception", *Speech perception and linguistic experience*, W. Strange, (eds.) Baltimore (MD): York Press, pp. 171-206, 1995.
- [15] D. H. Hymes, *On communicative competence*, Philadelphia: University of Pennsylvania Press, 1971.
- [16] G., Kasper and K. R. Rose, *Pragmatic development in a second language*, Malden, Oxford: Blackwell, 2002.
- [17] C. Kramersch, "From language proficiency to interactional competence," *The Modern Language Journal*, vol. 70, no.4, pp. 366-372, 1986.
- [18] Y. A. Lee, "Towards respecification of communicative competence: Condition of L2 instruction or its objective?," *Applied Linguistics*, vol. 27, no.3, pp. 349-376, 2006.
- [19] L. Neumeyer, H. Franco, M. Weintraub, and P. Price, "Automatic text-independent pronunciation scoring of foreign language student speech" in *Proc. of the 1996 International Conference on Spoken Language Processing*, 1457-1460, 1996.
- [20] M. Rypa, "VILTS: The Voice Interactive Language Training," *Proc. of CALICO*, 1996.
- [21] M. Eskenazi, M. Ke, J. Albornoz and K. Probst, "Update on the Fluency Pronunciation Trainer," *Proceedings of InSTIL 2000*, Dundee, pp. 73-76, 2000.
- [22] K. Probst, Y. Ke and M. Eskenazi, "Enhancing foreign language tutors - In search of the golden speaker" in *Speech Communication*, vol. 37, pp. 161-173, 2002.

- [23] D. Felps, H. Bortfeld, and R. Gutierrez-Osuna, “Foreign accent conversion in computer assisted pronunciation training,” *Speech Communication*, vol.51, no.10, pp.920-932, 2009.
- [24] M. Hismanoglu, “Computer Assisted Pronunciation Teaching: From the Past to the Present with its Limitations and Pedagogical Implications,” in *Frontiers of Language and Teaching, Proceedings of the 2011 IOLC*, vol.2, 2011, pp.193-202.
- [25] K. Nagano, K. Ozawa, “English speech training using voice conversion,” *1st Internat. Conf. on Spoken Language Processing (ICSLP 90)*, Kobe, Japan, pp. 1169–1172, 1990.
- [26] M. P. Bissiri, H. R. Pfitzinger, and H.G. Tillmann, “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, 2006. pp. 24–29.
- [27] M. Peabody and S. Seneff, “Towards automatic tone correction in nonnative mandarin. Chinese. Spoken Language Process”, pp. 602–613, 2006
- [28] A. De Meo, M. Vitale, M. Pettorino, F. Cutugno, and A. Origlia, “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*, J. Levis and K. LeVelle (Eds.) Aug. 2012, Ames, IA: Iowa State University, pp. 90-100, 2013.
- [29] A. De Meo, M. Vitale and E. Pellegrino, (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.
- [30] M. Pettorino and M. Vitale, “Transplanting native prosody into second language speech,” *Methodological Perspectives on Second Language Prosody. Papers from ML2P 2012*, M.G. Busà, A. Stella, (eds.), Padova: CLEUP, pp. 11-16, 2012.
- [31] K. Yoon, “Imposing native speakers’ prosody on non-native speakers’ utterances: The technique of cloning prosody,” *Journal of the Modern British & American Language & Literature*, vol. 25, no.4, pp. 197-215, 2007.
- [32] F. Charpentier and E. Moulines, “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, pp. 2013-2019, 1989.
- [33] P. Boersma, Praat, a system for doing phonetics by computer, *Glott International 5*, pp. 341-345, 2001.
- [34] M. D’Imperio, “Italian intonation: an overview and some questions,” *Probus*, vol. 14, pp. 37-69, 2002
- [35] I. Abe, “Intonation in Japanese,” *Intonation Systems- A survey of twenty languages*, D. Hirst and A. Di Cristo (ed.), Cambridge: Cambridge University Press, pp.363-378, 1998.
- [36] J. R. Searle, *Speech Acts*, Cambridge: Cambridge University Press, 1969.