

Methodological Perspectives on Second Language Prosody

Papers from ML2P 2012

edited by
Maria Grazia Busà, Antonio Stella

cleup

Prima edizione: novembre 2012

ISBN 978 88 6129 937 5

© CLEUP Sc
Coop. Libreria Editrice Università di Padova
Via Belzoni – 118/3 – Padova (Tel. 049/650261)
www.cleup.it

Tutti i diritti di traduzione, riproduzione e adattamento,
totale o parziale, con qualsiasi mezzo (comprese
le copie fotostatiche e i microfilm) sono riservati.

TRANSPLANTING NATIVE PROSODY INTO SECOND LANGUAGE SPEECH

Massimo Pettorino, Marilisa Vitale

Università degli Studi di Napoli “L’Orientale”

mpettorino@unior.it, vitale@unior.it

ABSTRACT

In this study we will describe the technique for superimposing the prosodic features of a native speaker’s utterance (donor) onto the same utterance produced by a non-native speaker (receiver). Particular attention will be given to the specific problems of L2 speech and operational suggestions on the procedure to follow will be provided. Suprasegmental features cannot be straightly transferred from one voice to another, but a neutralization process must occur before any transposing operation in order to avoid voice mismatch phenomena, which could alter listeners’ perception. Two different kinds of transplanting procedure, global and selective, will be described in detail. The transplantation technique can be very useful in many fields of research and can be a practical tool for the assessment of the perceptual correlates of acoustic features.

Keywords: Prosody, speech synthesis, foreign accent, language acquisition.

1. INTRODUCTION

Foreign accented speech, generally characterized by segmental and suprasegmental deviations with respect to the standard pronunciation, may induce native speakers to discriminate and to develop negative stereotypes against non-native speakers [1, 2, 22, 32, 33].

The elimination of foreign accent, or at least its reduction, therefore, would facilitate communication in today’s multicultural society, and it would promote greater social inclusion among its various components. Although many studies have been conducted on the factors that can influence the degree of foreign accentedness, there is not full agreement among researchers on the role that segmental and suprasegmental features have in the perception of foreign accent. Over the years a great deal of research has focused mainly on segments deviating from those typical of the pronunciation of native speakers [15, 25, 35]. By contrast, there are few studies that have dealt with the role of prosody in the perception of foreign accented speech. As a matter of fact, prosody has also played a minor role in recent theoretical models developed to explain speech perception and production in L2. The Speech Learning Model [16], the Perceptual Assimilation Model [3] and the Ontogeny and Phylogeny Model [23] have all investigated the notion of interference between L1 and L2 primarily at the segmental level.

However, in the last years the role of suprasegmental aspects of speech is being revalued

with regards to the perception of non-native speech and to the recognition of foreign speakers’ mother tongue. In this regard, the studies on second language acquisition [20, 26], on speech synthesis [15, 21, 24, 31, 7] and on automatic approaches for foreign accent identification [28] seem relevant. Additionally, some recent studies based on a pragmatic and acquisitional approach, have tried to qualitatively and quantitatively assess the role played by individual rhythmic and prosodic parameters in the development of effective communication. In particular, they have dealt with the relationship among foreign accent, communication effectiveness, credibility, reliability and persuasiveness [26, 13, 34]. In this research field, the employment of the prosodic transplantation technique has proved to be very useful. This technique, carried out by means of Praat, a software for speech analysis, makes 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.

The transplantation of native intonation on non-native speech involves a substantial improvement in communicative effectiveness and the reduction, if not the disappearance, of the foreign accent. These results are in agreement with what asserted by Munro [24] and Jilka [21], who emphasize the

role of rhythm and intonation in the perception of foreign accented speech.

Similarly, several studies on Italian accent in English have primarily focused their attention on the segmental aspects [8, 17, 18, 29], even if some of them have also dealt with the prosodic level. According to Busà [9, 10, 11], indeed, it seems that the Italian speakers of English are not able to achieve an effective communication, since they do not use the prosodic features that the English listeners expect.

2. TRANSPLANTATION TECHNIQUE

The rhythmic-prosodic transplantation technique, which is based on the PSOLA (Pitch-Synchronous Overlap and Add) algorithm [12], implemented in Praat [5], was illustrated in a 2007 study by Yoon [36] with regard to the English productions of Korean speakers. This extremely interesting study, which explains the procedure for the prosodic transposing, does not sufficiently focus on the possible obstacles and difficulties that the technician could face, especially when dealing with the L2 speech. As a matter of fact, non-native speech has very special characteristics, both segmental and suprasegmental, which require appropriate operational choices. For this reason, in this paper we will try to provide practical guidance and suggestions that can facilitate the fulfillment of this technique when applied to L2 Italian. To this end, we will discuss the whole procedure, with particular attention to the methods of intervention relating to the most frequent problems in spoken L2 and we will highlight the importance of each step in the perspective of language acquisition. It is also worth noting that, compared to the instructions given by Yoon, we made some procedural changes. These modifications have been determined by the outcome of acoustic and perceptual investigations on transplanted L2 Italian speech produced by foreign speakers with different backgrounds and different levels of L2 competence.

2.1. Global transplantation

The prosodic features that can be transplanted from one voice to another are essentially four: the length of the segments, the pitch contour of the utterance, the energy contour and the silent pauses. Transplantation may be selective or global. In the first case, one or more features can be transposed, while in the other the procedure involves the cloning of all the prosodic aspects taken into consideration.

In this paper we will illustrate a transplantation in which the prosody of the native voice (donor) is

superimposed to that of non-native (receiver). We take for granted that the reader is familiar with Praat and its functionalities.

2.2. Global transplantation

The procedure of global transplantation involves a fixed sequence of steps, since each phase can be performed only after having accomplished a series of necessary operations. This sequence, as also suggested by a series of tests, is the most effective procedural path to obtain the best result in terms of quality.

The five phases are: treatment of anomalies, segmentation, duration transplantation, intensity transplantation, pitch contour superimposition.

2.2.1. Treatment of anomalies

Different levels of competence in L2 may give rise to more or less marked speech anomalies, such as repetitions, corrections, false starts, nasalizations, extra-phones, intrawords silences and so on. In the first phase of the transplantation all these disturbing elements will be removed, except for the vocalic lengthenings, which will be discussed and treated in the phase of segmentation.

The removal of the anomalies must be simultaneously carried out on the two audio files, i.e. the donor's and the receiver's utterances. The realtime comparison between the two tracings will allow the identification of the parts to be removed from the receiver's utterance.

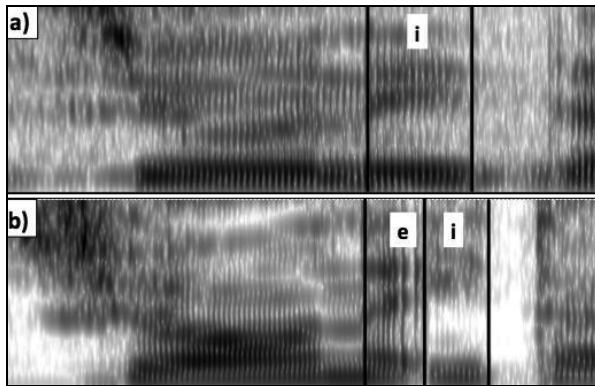
A particular attention is to be paid to the joining points, in order to avoid any extra spikes that could damage the final result. Care should also be taken in the treatment of silent pauses: the silences not corresponding to the silent pauses of the donor have to be eliminated from the utterance produced by the receiver. At the same time, intervals of silence must be added inside the receiver's utterance in correspondence with the remaining donor's silent pauses. It is better that these pauses do not consist of total absence of signal, as the final result of the transplantation would be perceptually unnatural. Conversely, it is appropriate to copy a silent interval extracted from the same utterance.

It should be pointed out that the higher the number of interventions, the worse the quality of the resulting transplantation. Therefore, it is advisable to work with read speech, in order to reduce the number of errors, interruptions and disfluencies.

2.2.2. Segmentation

Segmentation is the most important step of all, because the quality of the final result will depend on the accuracy of this operation. It can be performed manually or automatically, if an *ad hoc* tool¹ is available. However, even in the latter case it is appropriate to manually check the segmentation, since the unpredictability of speech, especially for an L2, may hinder the correct interpretation of the signal. For this reason the operator should have the necessary skills to be able to associate spectro-acoustic patterns to articulatory traits, in order to convert the continuum of the speech chain in a series of discrete elements, corresponding to the phones that constitute it. Even the segmentation, as well as the removal of the anomalies, must be performed for both voices simultaneously, detecting in real time any mismatch. The purpose of this operation is to obtain two TextGrid² files containing the same number of elements. This correspondence is necessary in order to proceed, in the subsequent steps, with the transfer of the various prosodic components, from one segment to the corresponding one in the other utterance. Due to the peculiarities of the L2 speech, the operator will have to act differently according to the situation. Here we will focus our attention on some of the most frequent problems.

Figure 1: An example of crasis: a) Italian voice; b) Chinese voice.



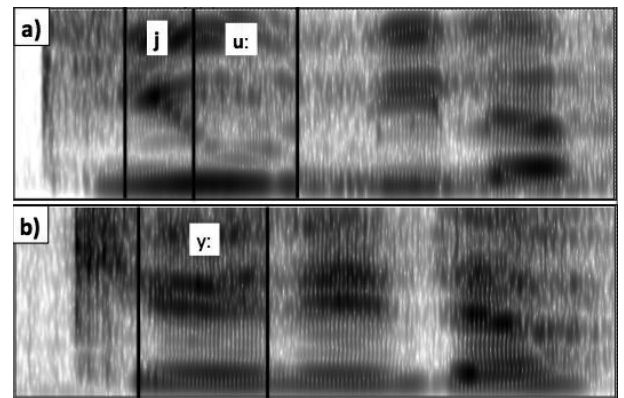
¹ On the web some products are available dedicated to the automatic segmentation of some languages, as EasyAlign, prepared by Jean-Philippe Goldman [19] and already in operation for the French and Spanish, and SPASS by Brigitte Bigi and Daniel Hirst [4].

² “A TextGrid object consists of a number of tiers. There are two kinds of tiers: an *interval tier* is a connected sequence of labelled intervals, with boundaries in between. A *point tier* is a sequence of labelled points” [6]. In this study we refer to interval tiers.

Crasis is one of the phenomena which is advisable to pay attention to, especially in the segmentation phase. In L1 Italian speech, adjacent vowels of two words in succession tend to merge into a single phone. On the contrary, foreign speakers tend to reach the correct articulatory target for each uttered segment and, as a consequence, the two vowels are kept separate (Fig. 1).

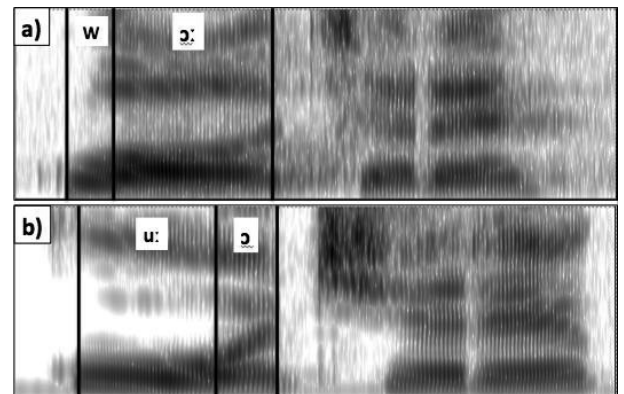
Therefore, in this case, the operator should consider as a whole the two adjacent vowels of the receiver and, in the same way, as a whole the element resulting from the crasis in the utterance of the donor.

Figure 2: An example of missing diphthong: a) Italian voice; b) Japanese voice.



As for the diphthongs, it is possible to face three different types of error in the receiver's utterance: 1) the presence of a diphthong when not required, 2) the absence of a diphthong when needed (Fig. 2); 3) wrong realization in terms of duration ratio between the first and the second vocalic element (Fig. 3). In the first and in the second case, the diphthong has always to be considered as a single element, whether it is present in the L1 or in the L2, in order to preserve the segmental characteristics of the L2 utterance.

Figure 3: An example of wrong ratio of vowel durations: a) Italian voice; b) Russian voice.



In the third case, however, if the utterance of the non-native speaker is far from the L1 model in terms of ratio of durations between the two vowels of the diphthong (e.g. [io:] > [i:o]), the operator must necessarily subdivide the diphthong into two segments. This will make possible to restore the correct ratios in the next steps.

2.2.3. Cloning durations

The purpose of this transplantation phase is to expand or reduce the duration of each segment of the receiver according to the length of the corresponding segment present in the utterance of the donor. As a result the two utterances will have the same rhythmic structure, including the location of phrase breaks. Since the most frequent characteristics of L2 speech are a lower articulation rate and a more complex strategy in the use of silent pauses, it is evident that the segmental duration cloning completely distorts the temporal organization of the L2, making it more similar to the native model. From the technical point of view, a new duration tier has to be created, in order to remodel the receiver's segmental durations, on the basis of the ratio between the model and the to-be-modeled temporal extents. The new duration tier thus obtained will replace the original one through the manipulation Praat object.

2.2.4. Remodeling intensity

On the basis of the instructions provided by Yoon [36], in the third step of the prosodic transplantation we proceeded to change manually the intensity levels. However, unlike what Yoon suggested, we decided to perform first the manipulation of the intensity and then the reshaping of the pitch contour. According to tests carried out on L2 Italian, we have verified that this procedural sequence gives rise to better results. Transplantation of the intensity contour is particularly important on pragmatic level, because different locations of energy peaks, helping to emphasize different portions of the utterance, determines different interpretations of the message.

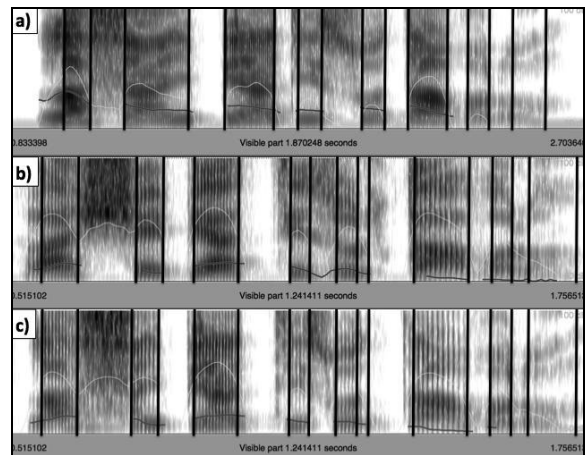
2.2.5. Reshaping pitch contour

In the last step of the prosodic transplantation the fundamental frequency contour of the non-native utterance is replaced by that of the donor. This procedure does not present any special difficulties for the operator since, in previous stages, the two utterances have already been prepared for this final step: the two utterances have the same number of segments, the same durations, the same dislocation of phrase breaks and energy peaks. Therefore, the

pitch tier of the receiver is ready to be replaced by that of the donor, without any possibility of misalignment: the distribution of the donor's tonal peaks throughout the text is kept unchanged in the remodeled receiver's utterance.

The simplicity of the execution of this phase must not lead us to underestimate the importance that the pitch contour has on the perceptual level. Pettorino et al. [27] have shown that variations in F0 affect the perception of foreign accent and the communication effectiveness more than other suprasegmental parameters (intensity and duration) or segmental errors.

Figure 4: Global transplantation: a) receiver; b) donor; c) receiver's utterance with donor's prosody.



2.3. Selective transplantation

Up to this point we have examined the procedure to make the global prosodic transplantation. However, it is possible to transfer from one utterance to another even a single prosodic parameter. The selective transplantation may be useful for research purposes, as it allows the operator to evaluate the different roles that specific prosodic parameters play on the perceptive level.

The selective transplantation involves the preliminary treatment of anomalies and the segmentation of the two utterances according to the already described procedure (§ 2.2.1, 2.2.2). We will now examine the various possible types of selective transplantation:

a) transplantation of durations: the procedure applied to this purpose is exactly the same performed for the global transplantation. After the treatment of the anomalies, the segmentation and the cloning of the duration, the audio file will have the required features and so the procedure can be considered completed;

b) transplantation of intensity: in order to proceed to this step it is necessary to have previously manipulated segmental durations, in

order to avoid misalignment between peaks of energy and structure of the utterance. In this case, however, the manipulation of the durations will follow an inverse pathway, in order to preserve the receiver's durations: the segments of the donor will be extended or reduced according to the receiver's segment durations;

c) transplantation of pitch contour: in order to align the two utterances on the time axis, the donor's and the receiver's voices have to undergo the manipulation of duration and intensity just described in sections a) and b). Once completed these procedural steps, the donor's intonation movement can be superimposed to the receiver's voice by replacing, in the manipulation object, the original PitchTier with the native one.

3. CONCLUSIONS

The prosodic transplantation appears to be a useful tool and a very effective survey technique for the study of the L2 speech. The global or selective manipulation of an utterance allows to evaluate experimentally the role played by each acoustic parameter on the pragmatic-communicative level, and to develop tools and actions to solve L2 learning problems. Moreover, this technique may find useful application in other research fields, such as speaker's identification, man-machine interaction, production of artificial voice, etc.

With regards to the language acquisition field, Probst et al. [30] have highlighted the importance of the similarity between the voice of the teacher and that of the learner, in terms of F0 and articulation rate, for learning the L2 suprasegmental features. In addition, Felps et al. [14] state that for the learner it would be ideal to listen to her/his own voice with a native-like accent. In this perspective, it is clear that the technique of prosodic transplantation could be extremely useful because, if applied to learners' voice, it would allow switching from the traditional imitation exercises to more effective self-imitation activities.

4. REFERENCES

- [1] Anisfeld, M., Bogo, N., Lambert, W.E. 1962. Evaluation reactions to accented English speech. *Journal of Abnormal Psychology* 65, 223-231.
- [2] Arthur, B., Farrar, D., Bradford, G. 1974. Evaluation reactions of college students to dialect differences in the English of Mexican-Americans. *Language Speech* 17 (3), 255-270.
- [3] Best, C. T. 1995. A direct realist view of cross-language speech perception. In: *Speech perception and linguistic experience: issues in cross - language research*, Strange, W. (ed). Baltimore: York Press, 171-203.
- [4] Bigi, B., Hirst, D. 2012. SPeech Phonetization Alignment and Syllabification (SPPAS): a tool for the automatic analysis of speech prosody. In: *Proceedings of the 6th International Conference on Speech Prosody*, Tongji University Press, 1, 19-22.
- [5] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5:9/10, 341-345.
- [6] Boersma, P., Weenink, D. 2011. *Introductory tutorial to Praat*. <http://www.fon.hum.uva.nl/praat/manual/TextGrid.html>
- [7] Boula de Mareüil, P., Vieru-Dimulescu, B. 2006. The Contribution of Prosody to the Perception of Foreign Accent. *Phonetica, Int. Journal of Phonetic Science*, 63 (4), 247-267.
- [8] Busà, M. G. 1995. *L'inglese degli italiani: L'acquisizione delle vocali*. Padova: Unipress.
- [9] Busà, M. G. 2007. New Perspectives in Teaching Pronunciation. In Baldry, A., Pavesi, M., Taylor Torsello, C., Taylor, C. (ed), *From Didactas to Ecolingua: an ongoing research project on translation & corpus linguistics*. Trieste: Edizioni Università di Trieste, 171-188.
- [10] Busà, M. G. 2008. Teaching prosody to Italian learners of English: working towards a new approach. In Taylor, C., (ed.), *Ecolingua: The Role of e-corpora in Translation, Language Learning & Testing*. Trieste: Edizioni Università di Trieste, 113-126.
- [11] Busà, M. G. 2010. Effects of L1 on L2 pronunciation: Italian prosody in English. In Gagliardi, A., Maley, A. (ed), *ILS, ELF, Global English: Teaching & Learning Processes, Linguistic Insights: Studies in Language & Communication*. Bern: Peter Lang, 200-228.
- [12] Charpentier, F., Moulines, E. 1989. *Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones*. First European Conference on Speech Communication and Technology, Eurospeech. Paris, 2013-2019.
- [13] De Meo, A., Vitale, M., Pettorino, M., Martin, P. 2011. Acoustic-Perceptual Credibility Correlates of News Reading by Native and Chinese Speakers of Italian. *Proc. 17th ICPHS Hong Kong*, 1366-1369.
- [14] Felps, D., Bortfeld, H., Gutierrez-Osuna, R. 2009. Foreign accent conversion in computer assisted pronunciation training. *Speech Communication* 51, 920-932.
- [15] Flege, J. E., Bohn, O-S., Jang, S. 1997. Effects of experience on non-native speaker's production and perception of English vowels. *Journal of Phonetics* 25, 437-470.
- [16] Flege, J. E., 2003. Assessing constraints on second-language segmental production and Perception. In: Meyer, A. Schiller, N., (ed), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*. Berlin: Mouton de Gruyter, 319-355.
- [17] Flege, J. E., MacKay, I. R. A., Meador, D. 1999. Native Italian Speakers' Production & Perception of English Vowels. *Journal of Acoustical Soc. of America* 106, 2973-2987.
- [18] Flege, J. E., Schirru, C., MacKay, I. R. 2003. Interaction between the native and second language phonetic subsystems, *Speech Communication* 40, 467-491.
- [19] Goldman, J.-Ph. 2011. EasyAlign: an automatic phonetic alignment tool under Praat. I: *Proceedings of Interspeech XI*, Ses1-S3:2, Florence, Italy.
- [20] Horgues, C. 2005. Contribution à l'étude de l'accent français en anglais. Quelques caractéristiques prosodiques de l'anglais parlé par des apprenants francophones et leur évaluation perceptive par des juges natifs. In: *Actes des VIIIèmes RJC ED268 'Langage et langues'*, Paris, 79-83.

- [21] Jilka, M. 2000. The contribution of intonation to the perception of foreign accent. Doctoral Dissertation, Arbeiten des Instituts für Maschinelle Sprachverarbeitung (AIMS) Vol. 6(3), University of Stuttgart.
- [22] Lippi-Green, R. 1997. *English With an Accent: Language, Ideology, and Discrimination in the United States*. New York: Routledge.
- [23] Major, R. C. 2001. *Foreign Accent: The Ontogeny and Philogeny of Second Language Phonology*. Mahwah, NJ: Erlbaum.
- [24] Munro, M. J. 1995. Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition* 17, 17-34.
- [25] Pallier, C., Bosch, L., Sébastien-Galles, N., 1997. A limit on behavioral plasticity in speech perception. *Cognition* 64, 9-17.
- [26] Pettorino, M., De Meo, A., Pellegrino, E., Salvati, L., Vitale, M. 2011. Accento straniero e credibilità del messaggio: un'analisi acustico-percettiva. In: Gili Fivela, B., Stella, A., Garrapa, L., Grimaldi, M. (ed), *Contesto comunicativo e variabilità nella produzione e percezione della lingua*. Roma: Bulzoni, 284-292.
- [27] Pettorino, M., De Meo, A., Vitale, M. (in press). Transplanting vowels: towards the acoustic correlates of foreign accent. In *Actas del V Congreso en Fonética Experimental*. Cáceres 2011, October 2011 .
- [28] Piat, M., Fohr, D., Illina, I. 2008. Foreign accent identification based on prosodic parameters. *Proceedings INTERSPEECH*. Brisbane, 759-762.
- [29] Piske, T., Flege, J. E., MacKay, I. R. A., Meador, D. 2002. The Production of English Vowels by Fluent Early and Late Italian-English Bilinguals. *Phonetica* 59, 49-71.
- [30] Probst, K., Ke, Y., Eskenazi, M. 2002. Enhancing foreign language tutors - in search of the golden speaker. In: *Speech Communication*, 37/3-4, 161-173.
- [31] Ramus, F., Mehler, J. 1999. Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America*, 105(1), 512-521.
- [32] Ryan, E. B., Carranza, M. A., 1975. Evaluative reactions of adolescents toward speakers of standard English and Mexican American accented English. *J. of Personality and Social Psychology* 31, 855-863.
- [33] Schairer, K. E. 1992. Native speaker reaction to non-native speech. *Mod.Language J.* 76 (3), 309-319.
- [34] Vitale, M., De Meo, A., Pettorino, M. 2012. Foreign accent and persuasion. Native and non native voices in a radio spot. In: De Meo, A., Pettorino, M., (ed), *Prosodic and Rhythmic Aspects of L2 Acquisition. The case of Italian*. Cambridge Scholars Publishing.
- [35] Walley, A. C, Flege, J. E. 1999. Effect of lexical status on children's and adults' perception of native and non-native vowels. *Journal of Phonetics* 27, 307-332.
- [36] 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.