INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

L1 perceptions of L2 prosody: The interplay between intonation, rhythm, and speech rate and their contribution to accentedness and comprehensibility

Lieke van Maastricht[1], *Tim Zee*[2], *Emiel Krahmer*[1], *Marc Swerts*[1]

[1] Tilburg Center for Communication and Cognition, Tilburg University, The Netherlands
[2]Radboud University, Nijmegen, The Netherlands

l.j.vanmaastricht@uvt.nl, t.zee@student.ru.nl, e.j.krahmer@uvt.nl, m.g.j.swerts@uvt.nl

Abstract

This study investigates the cumulative effect of (non-)native intonation, rhythm, and speech rate in utterances produced by Spanish learners of Dutch on Dutch native listeners' perceptions. In order to assess the relative contribution of these language-specific properties to perceived accentedness and comprehensibility, speech produced by Spanish learners of Dutch was manipulated using transplantation and resynthesis techniques. Thus, eight manipulation conditions reflecting all possible combinations of L1 and L2 intonation, rhythm, and speech rate were created, resulting in 320 utterances that were rated by 50 Dutch natives on their degree of foreign accent and ease of comprehensibility.

Our analyses show that all manipulations result in lower accentedness and higher comprehensibility ratings. Moreover, both measures are not affected in the same way by different combinations of prosodic features: For accentedness, Dutch listeners appear most influenced by intonation, and intonation combined with speech rate. This holds for comprehensibility ratings as well, but here the combination of all three properties, including rhythm, also significantly affects ratings by native speakers. Thus, our study reaffirms the importance of differentiating between different aspects of perception and provides insight into those features that are most likely to affect how native speakers perceive second language learners.

Index Terms: intonation, rhythm, speech rate, accentedness, comprehensibility.

1. Introduction

When listeners are confronted with speech in their native language (L1), they generally are able to distinguish between L1 speakers and non-native speakers that have acquired the language after childhood, i.e., L2 speakers. The perceivable difference between speech produced in the L1 and in the L2 has been attributed to various phonetic properties; Prior studies have shown that both the pronunciation of L2-specific phonemes, and the production of target-like speech rate, intonation, and rhythm, affects how L1 listeners perceive L2 speakers. They generally distinguished between different types of L1 perceptions of L2 speech, focusing on the degree of perceived foreign accent (accentedness), the reported ease of understanding of L2 speech (comprehensibility), or measures reflecting actual processing of L2 speech by L1 listeners, such as reaction times (intelligibility).

Nevertheless, to the best of our knowledge, previous work on the perception of L2 speech has usually been limited to one or two of the aforementioned language-specific properties at a time. Furthermore, most previous studies only investigated one aspect of perception, usually accentedness. However, as shown by Van Maastricht, Krahmer and Swerts ([1]), the manipulation of prosodic features (in their case pitch accent distributions used to mark focus) affects accentedness, comprehensibility, and intelligibility in different ways and to varying degrees. As both the perception (e.g., [2-3]) and the production (e.g., [4-5) of L2 speech are known to result from the interplay between various language-specific features, the current study aims to determine what the cumulative effect is of three prosodic properties to L1 perceptions of accentedness and comprehensibility. Before turning to the design of our experiment, we briefly review the most relevant studies that examined these factors on a more individual basis, using accentedness and/or comprehensibility as dependent variables.

Derwing and Munro defined accentedness as the extent to which "an L2 accent differs from the variety of speech commonly spoken in the community" ([6], p.385). In this sense, it can be interpreted as the opposite of nativeness, i.e., "the degree to which a speaker sounds like a native speaker of a particular language" ([7], p. 2). Derwing and Munro defined comprehensibility as "the listener's perception of the degree of difficulty encountered when trying to understand an utterance" ([6], p.385). In comparison to accentedness, there are relatively few studies focusing on this perception measure when it comes to the relative contribution of different prosodic cues. A notable exception are Saito, Trofimovich and Isaacs ([8]), who performed a multiple regression analysis to examine the contribution of segmental errors, word stress errors, intonation, and speech rate (and other, non-phonetic factors) to comprehensibility and accentedness ratings. Their research on utterances produced by Japanese learners of English showed that all areas of pronunciation correlated significantly with both measures, with no differences in strength of the association for comprehensibility, while accentedness was more strongly associated with segmental errors than with intonation and speech rate.

These findings are in line with studies examining these prosodic cues in isolation: e.g., Munro and Derwing ([9]) examined the effect of speech rate changes on foreign accent and comprehensibility ratings. They reported that L1 English listeners generally considered typical L2 English with a relatively low speech rate as more accented and less easy to comprehend than L2 English that was somewhat faster than typical L2 speech. However, both very fast and very slow speech were rated as more accented and more difficult to understand. In order to determine the effect of intonation on foreign accent perception, Van Els and De Bot [10] flattened the pitch contour in speech by L1 and L2 speakers of Dutch while retaining most of its segmental properties. This manipulation affected the degree of success with which L1 Dutch judges were able to determine whether they listened to

http://dx.doi.org/10.21437/Interspeech.2017-1150

an L1 or L2 speaker of Dutch, suggesting that intonational features also contribute to the perception of a foreign accent.

While there are studies that investigated the contribution of durational (i.e., rhythmic) properties[1] to intelligibility (e.g., [11-12]), to our knowledge there are no studies that examined the isolated effect of rhythmic deviance on accentedness or comprehensibility. Conversely, there are studies that looked at the combined effect of rhythmic and intonational properties on perceived foreign accent: Boula de Mareüil and Vieru Dimulescu ([13]) showed that the transplantation of both melodic and durational features of one language onto the segmental string of a related language (in this case Spanish and Italian) resulted in speech that was more often classified as the language whose prosody was used than as the language from which the segments were taken. Similarly, Ramus and Mehler ([14]) used speech resynthesis techniques to create four conditions in which they preserved (1) the intonation, rhythm, and segments, (2) the rhythm and intonation, (3) only the intonation, or (4) only the rhythm of English and Japanese utterances. Their results revealed that rhythm was a "necessary and sufficient cue" for French adults when discriminating between English and Japanese utterances ([14], p.1).

Presently, only one study combined all three prosodic cues in one design (though using a less specific L2 intonation manipulation than the current study) to determine how rhythm, intonation, and speech rate contribute to foreign accent perceptions: Polyanskaya, Ordin and Busa ([15)] manipulated utterances by French learners of English in such a way that (1) the original L2 rhythm was preserved, while controlling for speech rate, (2) the L2 speech rate was maintained, while rhythm was controlled for, and (3) both rhythm and speech rate were preserved. All stimuli were created with an imposed intonational contour and with monotonized pitch. The results showed that both speech rate and rhythm affected accentedness, but that rhythm does so more strongly than speech rate. Additionally, intonation was shown to boost the perception of minimal rhythmic differences, but to reduce the salience of small differences in speech rate.

In sum, while prior research investigated the individual effects of intonation, rhythm, and speech rate on either accentedness or comprehensibility, no study ever combined all three prosodic features in one design, while also measuring their contribution to different aspects of L1 perception. As accentedness and comprehensibility ratings are known to differ with respect to their range (Van Maastricht et al. reported mean accentedness ratings on a 9-point scale ranging from 2.4 to 8.7 depending on the proficiency of the speaker, while mean comprehensibility ratings only varied from 5.9 to 8.5, [1], p. 26), we argue that it is important to include both perception measures in one study, using identical stimuli and manipulations for both tasks. Finally, the results of our study are not only theoretically and didactically relevant to L2 learners and teachers; they are also applicable to other scientific fields that rely on speaker comprehensibility, e.g., speech pathology and automatic speech recognition and production. Moreover, they are applicable forensic linguistics in which it is relevant to link phonemic properties, including intonation, speech rate and rhythm, to specific speakers.

[1] Rhythm arguably incorporates more than just timing differences. In the current study, this is taken into account by the fact that the rhythm manipulation reflects the accentual and final lengthening patterns that are typical of Dutch (and not of Spanish). In this sense, our manipulation reflects the durational variation that exists in Dutch due to the marking of prominence and boundaries, without combining it with intonation, as has been done in previous studies.

2. Method

2.1. Participants[2]

50 monolingual Dutch adults participated in the current study: 34 women (*M* age = 20.56 years, *SD* = 2.28) and 16 men (*M* age = 21.13 years, *SD* = 2.25). Since French is obligatory at the higher levels of the Dutch secondary educational system, most participants had basic knowledge of this language. However, data by participants reporting knowledge of other Romance languages were excluded from analysis. While other metalinguistic factors could not be controlled for, all participants had at least completed higher secondary education, as they were students or PhDs at Tilburg University participating voluntarily or for course credit. Participants were randomly divided between the rating tasks.

2.2. Materials

The experiment was presented as an online survey in Qualtrics ([16]). All utterances used in the study were previously elicited and recorded during a production study ([17]) in which L1 speakers of Dutch and Spanish and learners of both languages were asked to read aloud thirty sentences. Ten of these were used in the current perception experiment, based on a selection of speakers who had produced syllabic structures that were identical to the L1 speaker that was used as a donor in this study. This selection was narrowed down by only using female speakers to guarantee comparability across stimuli and by discarding recordings with poor sound quality. The fluency with which speakers had produced the sentences in the production study was the final selection criterion. The sentences were typical of the Dutch language regarding syllable structure (using a mix of open and closed syllables) and word frequency, as exemplified in (1).

(1) Delegaties uit meer dan twintig landen komen naar dit congres.

('Delegations from more than twenty countries are coming to this congress').

2.3. Prosodic manipulation

In the current experiment, three prosodic parameters were manipulated: speech rate, rhythm, and intonation. These features were extracted from a donor utterance and transplanted onto a receiver utterance of the same sentence. An L2 speaker always produced the receiver utterance, while the donor utterances were produced by an L1 speaker for the seven experimental conditions, and by an L2 speaker for the control condition. The utterances were prosodically manipulated by modifying the pitch and duration of the speech signal using the PSOLA algorithm ([18]) implemented in Praat ([19]). As previous papers have provided in-depth explanations of this process and its use in L2 perception research (e.g., [[13], and [20-22]), this section focuses on the specific application to the present experiment, starting with the preparation of the speech material.

Successful transplantation of prosody requires that the donor and receiver versions of an utterance have a similar

[2] At the moment of submission, data collection was being concluded. At the conference, analyses on the complete data set using more advanced statistical tests will be presented.

syllabic structure. As such, by using manual word- and syllable-level annotations, only utterances that matched in the number of syllables that were used for each word were selected. Subsequently, if either version of the remaining sentences did not contain a silent pause that was present in the other version, a very small corresponding pause was inserted in which the amplitude was set to close to zero. Another requirement for successful transplantation is an accurate F0 analysis of both the receiver and the donor utterance. For that reason, the parameters of the pitch analyses for each utterance of each sentence were set manually. The result of these steps served as the input to the transplantation process, as represented in the upper part of Figure 1.

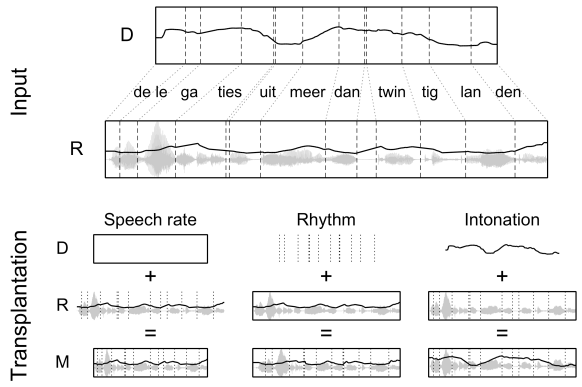

Figure 1: *Schematic overview of the transplantation process for part of the stimulus presented in (1). The length of the rectangular outlines represents total duration of the utterance, the vertical dashed lines represent rhythm, and intonation is represented by the F0 contour. D, R, and M refer to donor, receiver, and manipulated utterance.*

Depending on the experimental condition, the transplantation step consisted of resynthesizing the receiver utterance using either the speech rate, rhythm, intonation, or any combination of those parameters from the donor utterance (see Table 1 for the resulting conditions). The transplantation of speech rate was implemented as follows: The total duration of the receiver speech signal was either compressed or stretched to match the duration of the donor signal. The leftmost column of the lower part of Figure 1 shows how the duration of a donor L1 utterance was used to compress the receiving L2 utterance, which was typically more slowly articulated. For the transplantation of rhythm, each receiver syllable was compressed or stretched so that its duration relative to the total duration of receiver utterance was identical to the proportional duration of the corresponding donor syllable. By adapting the proportional duration of all syllables in this way, the resulting utterance reflected the final and accentual lengthening patterns typical of L1 Dutch. This is exemplified by the middle column in the bottom part of Figure 1. The transplantation of intonation was also done on a syllable-by-syllable basis. The pitch contour of each donor syllable was made to fit the corresponding receiver syllable, see the right-most column in the lower part of Figure 1. To retain the receiving speaker's pitch register, this process also involved a shift in ERB units ([23]) that centered the donor pitch contour around the mean of the receiver pitch contour. After the transplantation step, the sound pressure level of each resynthesized utterance was normalized to 64 dB. Silent pauses in the utterances were excluded from this process.

Table 1: *Manipulation conditions used in the accentedness and comprehensibility rating tasks.*

Cond.	Donor speaker	Transplanted features
C1	L1 Spanish	Intonation, rhythm, speech rate
C2	L1 Dutch	Intonation
C3	L1 Dutch	Rhythm
C4	L1 Dutch	Speech rate
C5	L1 Dutch	Intonation, rhythm
C6	L1 Dutch	Intonation, speech rate
C7	L1 Dutch	Rhythm, speech rate
C8	L1 Dutch	Intonation, rhythm, speech rate

2.4. Procedure

Although the experiment was performed online, experimental sessions took place in a quiet computer room, to make sure that all participants performed the experiment in equal conditions and were not distracted during the task. Sessions were performed in a group setting and took approximately 60 minutes. In order to prevent unreliable answers due to fatigue or boredom, participants performed the task in three blocks of roughly 10 minutes each, separated by 10-minute breaks.

During the task, participants were instructed to carefully listen to the utterances through headphones, as the differences between them would be subtle, and were encouraged to use the complete range of the 7-point scale on which they were to judge the accentedness or comprehensibility of the items. The following instructions and scales for accentedness (2) and comprehensibility (3) were presented to the participants (English translations of the original Dutch sentences).

(2) Indicate to which extent the speaker you heard has a foreign accent

No foreign accent – Very strong foreign accent

(3) Indicate to which degree the speaker you heard is easy/difficult to understand

Incomprehensible – Very easy to understand

Before starting the rating task, the participants answered a block of questions about their age, nationality, L1 and possible L2s, and the existence of audio and/or visual impairments to ensure that they met the requirements explained above.

3. Results

First, the accentedness ratings were transformed to reflect the same direction of effect as expected for comprehensibility using the following formula. In what follows, higher ratings hence indicate higher comprehensibility and lower accentedness.

$$[\text{NewRating} = -1 * \text{Rating} + 8] \quad (1)$$

Subsequently, two repeated measures analyses were performed with Manipulation Condition (8 levels) as a within-subjects factor and the accentedness and comprehensibility ratings as dependent variables. A Greenhouse-Geisser correction was used on the degrees of freedom of an analysis if the sphericity assumption was violated. Figure 2 summarizes the results. The analysis revealed a significant main effect of Manipulation Condition on the ratings for accentedness ($F(2.66, 24) = 6.57$, $p < .001$, $\eta_p^2 = .215$), as well as comprehensibility ($F(2.24, 24) = 7.47$, $p < .001$, $\eta_p^2 = .237$). Pairwise comparisons using the Bonferroni method were performed to further investigate which manipulation conditions differed significantly from each other (relevant *p*-

values corresponding to these pairwise comparisons are reported in Table 2).

Table 2: p-*values of pairwise comparisons between relevant Manipulation Conditions for accentedness and comprehensibility.*

Pairwise comparison	Accentedness	Comprehensibility
1 - 2	p = .014*	p = .000***
1 - 3	p = 1.000	p = .903
1 - 4	p = 1.000	p = .024*
1 - 5	p = 1.000	p = .207
1 - 6	p = .015*	p = .034*
1 - 7	p = .321	p = .004**
1 - 8	p = .002**	p = .003**
8 - 2	p = 1.000	p = 1.000
8 - 3	p = .015*	p = .057
8 - 4	p = .073	p = 1.000
8 - 5	p = .088	p = .556
8 - 6	p = 1.000	p = 1.000
8 – 7	p = .011*	p = 1.000

ACCENTEDNESS All experimental conditions received higher mean ratings than the control condition, which indicates that using L1 donor intonation, rhythm and speech rate all contribute to a lower perception of accentedness by L1 speakers, see Figure 2. Statistically however, only the 'intonation', the 'intonation + speech rate', and the 'intonation + rhythm + speech rate' conditions differ significantly from the baseline condition in which the intonational, rhythmic, and speech rate properties of an L2 speaker were manipulated using donor material from another L2 speaker of the same proficiency level. This suggests that utterances in those conditions especially were considered more native-like than all others. This is also reflected in the pairwise comparisons with the 'intonation + rhythm + speech rate' condition, which received the highest mean accentedness rating: the 'intonation' and the 'intonation + speech rate' conditions are the only two conditions that have p-values that are unquestionably insignificant (all other comparisons either result in significant differences or differences that might be defined as 'trends').

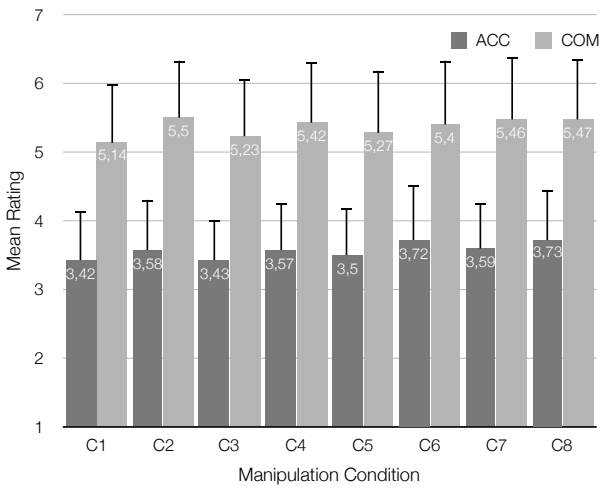

Figure 2: *Mean accentedness and comprehensibility ratings (standard deviations) by L1 speakers of Dutch for all manipulation conditions.*

COMPREHENSIBILITY Again, as shown in Figure 2, all experimental conditions received higher mean ratings than the control condition, showing that manipulating speech by transplanting L1 donor intonation, rhythm and/or speech rate onto L2 segments contributes to a higher comprehensibility of the utterance. Several conditions differ significantly from the baseline condition, i.e., pairwise comparisons between the baseline condition and the 'intonation', 'speech rate', 'intonation + speech rate', 'rhythm + speech rate', and the 'intonation + rhythm + speech rate' conditions all produce significant p-values. As was the case for accentedness, performing pairwise comparisons with the condition that received the highest mean comprehensibility ratings, i.e., the condition in which all three prosodic features are manipulated, reveal that those conditions have p-values of 1.000, while all other conditions have lower p-values.

4. Conclusion and Discussion

Our study shows that transplanting the intonation, speech rate, and rhythmic properties of an L1 speaker of Dutch onto identical utterances produced by Spanish L2 learners of Dutch positively affects L1 listeners' perceptions of foreign accent and ease of comprehension. Specifically, our results show that accentedness and comprehensibility are not equally affected by identical prosodic manipulations: while only manipulations including intonation (either alone, combined with speech rate, or in the full manipulation) significantly affect accentedness, comprehensibility ratings are significantly influenced by manipulations of intonation, as well as speech rate.

These results are congruent with most of the studies mentioned in the introduction, i.e., [8-10, and 13]. Crucially, our results differ from previous findings pertaining to the effect of multiple cues on accentedness. For Ramus and Mehler ([14]), this might be due to the fact that all meaning was extracted from their stimuli: depending on the manipulation, they transformed all segments into different variants of *sasasa* replacing all consonants with /s/ and all vowels with /a/. Conversely, the present study used utterances with meaning. As intonation is highly context-dependent, it might have been difficult for their participants to distinguish between two languages based on this cue without any context at all. Regarding Polyanskaya et al. ([15]), the nature of the pitch manipulation might explain the contradictory findings: While they used utterances with a flat F0 as a baseline condition, the current study relied on utterances with the original L2 intonation. Arguably, the difference between monotonized F0 and L1 intonation is less salient than the difference between L2 and L1 intonation, as L2 speech typically contains noticeable lexical and phrasal stress errors. Hence, our study also has didactic consequences: while speech rate is known to increase with proficiency level ([9]), requiring little extra coaching, specific training to improve intonation might be a useful addition to an L2 curriculum.

5. References

[1] L. van Maastricht, E. Krahmer, and M. Swerts, "Native speaker perceptions of (non-)native prominence patterns: Effects of deviance in pitch accent distributions on accentedness, comprehensibility, intelligibility, and nativeness," *Speech Communication*, vol. 83, pp. 21–33, 2016.

[2] J. Anderson-Hsieh, R. Johnson, and K. Koehler, "The relationship between Native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure," *Language Learning*, vol. 42, pp. 529–555, 1992.

[3] J. Caspers, and K. Horłoza, "Intelligibility of non-natively produced Dutch words: interaction between segmental and suprasegmental errors," *Phonetica*, vol. 69, pp. 94– 107, 2012.
[4] L. van Maastricht, E. Krahmer, M. Swerts, and P. Prieto, "Learning direction matters: A study on L2 rhythm acquisition by Dutch learners of Spanish and Spanish learners of Dutch," in preparation, 2017.
[5] A. Li, and B. Post, "L2 acquisition of prosodic properties of speech rhythm," *Studies in Second Language Acquisition*, vol. 36, no. 2, pp. 223-255, 2014.
[6] T.M. Derwing, and M.J. Munro, "Second language accent and pronunciation teaching: a research-based approach," *Tesol Quarterly*, vol. 39, pp. 379–397, 2005.
[7] P. Edmunds, "ESL Speakers' Production of English Lexical Stress: The Effect of Variation in Acoustic Correlates on Perceived Intelligibility and Nativeness," Unpublished doctoral dissertation, The University of New Mexico, 2010.
[8] K. Saito, P. Trofimovich, and T. Isaacs, "Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels," *Applied Psycholinguistics*, vol. 37, no. 2, pp. 217-240, 2016.
[9] M.J. Munro, and T.M. Derwing, "Modelling perceptions of the accentedness and comprehensibility of L2 speech the role of speaking rate," *Studies in Second Language Acquisition*, vol. 23, pp. 451–468, 2001.
[10] T. van Els, and K. de Bot, "The role of intonation in foreign accent," *Modern Language Journal*, vol. 71, pp. 147–155, 1987.
[11] H. Quené, and L.E. van Delft, "Non-native durational patterns decrease speech intelligibility," *Speech Communication*, vol. 52, no. 11, pp. 911-918, 2010.
[12] K. Tajima, R. Port, and J. Dalby, "Effects of temporal correction on intelligibility of foreign-accented English," *Journal of Phonetics*, vol. 25, no. 1, pp. 1-24, 1997.
[13] P. Boula de Mareüil, and B. Vieru-Dimulescu, "The contribution of prosody to the perception of foreign accent," *Phonetica*, vol. 63, no. 4, pp. 247-267, 2006.
[14] F. Ramus, and J. Mehler, "Language identification with suprasegmental cues: A study based on speech resynthesis," *The Journal of the Acoustical Society of America*, vol. 105, no. 1, pp. 512-521, 1999.
[15] L. Polyanskaya, M. Ordin, and M.G. Busa, "Relative salience of speech rhythm and speech rate on perceived foreign accent in a second language," *Language and Speech*, pp. 1-23, 2016.
[16] "Qualtrics," [Computer software]. Retrieved from http://www.qualtrics.com/, 2016.
[17] L. van Maastricht, E. Krahmer, M. Swerts, and P. Prieto, "Learning L2 Rhythm: Does the direction of acquisition matter?," *Proceedings of Speech Prosody*, pp. 974–978, 2016.
[18] E. Moulines, and F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech communication*, vol. 9, no. 5-6, pp. 453-467, 1990.
[19] P. Boersma, and D. Weenink, "Praat: Doing phonetics by computer," [Computer software]. Version 6.0.21, retrieved from http://www.fon.hum.uva.nl/praat/, 2016.
[20] M. Jilka, "The contribution of intonation to the perception of foreign accent: Identifying intonational deviations by means of F0 generation and resynthesis," Unpublished doctoral dissertation, Universität Stuttgart, 2000.
[21] M. Pettorino, and M. Vitale, "Transplanting native prosody into second language speech," In M. G. Busà, and A. Stella (eds.), *Methodological Perspectives on Second Language Prosody: Papers from ML2P 2012*, pp. 11-16. Padova: CLEUP, 2012.
[22] K. Yoon, "Imposing native speakers' prosody on non-native speakers' utterances: The technique of cloning prosody," *Journal of the Modern British and American Language & Literature*, vol. 25, no. 4, pp. 197-215, 2007.
[23] D. D. Greenwood, "A cochlear frequency-position function for several species—29 years later," *The Journal of the Acoustical Society of America*, vol. 87, no. 6, pp. 2592-2605, 1990.