

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.

THE IMPACT OF PROSODY IN FOREIGN ACCENT DETECTION. A PERCEPTION STUDY OF ITALIAN ACCENT IN ENGLISH

Luca Rognoni

Università degli Studi di Padova
luca.rognoni@studenti.unipd.it

ABSTRACT

The purpose of this study is to investigate the role played by prosody in the perception of foreign accent. The hypotheses are the following: English native listeners (ENL) can still detect foreign accent when most of the segmental information is removed; intonation will be the strongest prosodic cue; ENL will still detect foreign accent using prosodic temporal cues even when segmental information is removed and intonation is monotonized; ENL will encounter difficulties in recognizing native speakers when segmental information is preserved and intonation monotonized.

The hypotheses were tested in a perceptual experiment conducted on 10 ENL. The stimuli were selected from samples of read speech elicited from two groups: Italian and English native speakers. Three digital manipulations were applied: delexicalization with PURR filtering (Sonntag & Portele 1998); a combination of PURR filtering and f0 monotonization, and f0 monotonization alone.

The results show that listeners can detect foreign accent even when presented with stimuli stripped from most of their segmental information. In addition intonation proves to be the strongest prosodic cue, but the prosodic cues of a temporal nature (i.e. rhythm and speech rate) also play an important role in accent detection. The results also show that the ability to detect foreign accent is hindered when intonation is monotonized. This finding can be seen as perceptual evidence in support of the possible indication coming from production studies reporting that English L2 intonation tends to be flatter than the native one.

Keywords: foreign accent, signal manipulation, phonetic cues, English L2, L2 prosody.

1. INTRODUCTION

The majority of studies on foreign accent have focused on the segmental aspects (e.g., acquisition of vowels and consonants), and the contribution of prosody has only recently started to be studied. Intonation patterns, pitch range, rhythm, and speech rate are the prosodic aspects that are thought to be more influent in foreign accent detection, although so far no study has managed to define a hierarchy in their importance and to quantify their relative role in foreign accent detection ([8, 9, 12, 13]

The main problem when testing the impact of single prosodic aspects lies in the fact that, in natural speech, prosody cannot be disentangled from the segmental dimension. One way to deal with these problems is to manipulate the speech signal and degrade or remove some parts of the information while preserving others, in order to separate the different streams of information and to evaluate their relative contribution in speech perception. For this purpose, a variety of signal manipulation techniques have been used in experimental studies, such as low-pass filtering ([8, 13, 19, 20]), pitch monotonization ([14, 20]),

speech synthesis and resynthesis ([4, 17]), prosody transplantation ([1, 4, 5]), reverse speech [14] and splicing [14]. These techniques have been often combined with each other in order to define the impact of the single prosodic elements in foreign accent detection.

From the point of view of prosody, Italian and English present a wide range of differences. First of all, the two languages differ in rhythm and syllable length: while in English unstressed vowels are constantly reduced, in Italian the reduction is limited [2]. The inventories of intonation patterns in the two languages are also different. For example, while in English *wh*-questions are produced in a falling intonation, in Italian they are realized with a rising tone [5]. There is also a great difference in the way information status and prominence are conveyed: while in English prominence can be marked by stressing any word in a sentence, Italian tends to rely on word order strategies [10]. In addition, recent studies have also found differences in pitch range: Italians speaking English seem to have a narrower pitch range [2], resulting in a flatter intonation than the native one. This has been reported as a typical prosodic error made by nonnative speakers of English [12].

The present study is aimed to investigate the relative contribution of prosodic aspects in the perception of foreign accent using a combination of signal manipulation techniques. In particular, four hypotheses were tested:

- H1: Native listeners will detect foreign accent even when most of the segmental information is degraded;
- H2: Native listeners will detect foreign accent more accurately when segmental information is degraded and the intonation patterns are maintained than when segmental information is degraded and intonation is monotonized to a constant value;
- H3: Native listeners will be able to detect foreign accent even when segmental information is degraded and intonation is monotonized;
- H4: Native listeners will encounter difficulties in recognizing native speakers when segmental information is preserved and intonation is monotonized.

2. METHODOLOGY

2.1. Data set

2.1.1. Elicitation of speech samples

Five Italian native speakers (from the North-East Veneto area) and five British English native speakers read a version of Aesop's fable "The Fox and the Crow" adapted by the author. The speakers of both groups were all female, their average age was 21. The speech samples were recorded on DAT with a Shure SM58 microphone in a silent room. The following four sentences from each speaker were selected, presenting a variety of intonation patterns and syntactic structures:

A: Hi, Crow, how are you?

B: Will you sing a song for me?

C: He saw a piece of cheese sitting on the shelf of a window

D: I want to hear your beautiful voice!

The resulting set of speech samples consisted in 40 utterances (4 sentences x 10 speakers): 20 sentences uttered by Italian native speakers and 20 by English native speakers.

2.1.2. Digital manipulation

The 40 natural sentences were digitally manipulated using *Praat* scripts [1]. Two signal transformation techniques were applied, namely a

content-masking method and the monotonization of f0. A third version of the sentences was obtained applying a combination of the two techniques.

Low-pass filtering has been frequently adopted in studies on the impact of prosody in foreign accent detection (see section 1 above). With this technique, the higher frequencies of the speech signal are cut off and the lower frequencies are preserved. As a result, most of the segmental information is removed, while suprasegmental aspects are preserved. Although this method preserves many prosodic aspects (pitch, speaking rate and rhythm) and therefore cannot capture the impact of the single prosodic elements [19], it can still be considered a useful tool for content-masking of speech signal with the aim of testing the impact of prosody.

The content-masking technique that was adopted in this study was the PURR (Prosody Unveiling through Restricted Representation) method developed by [18]. This method, originally meant for the evaluation of prosody in text-to-speech software, was chosen because of the smoothness of the resulting filtered speech, which is easier and less tedious to be evaluated in the extended sessions required by perception tests.

The second signal manipulation technique consisted in the monotonization of f0. This technique consists in leveling pitch to a constant value: as a result, the segmental information is preserved and the intonation patterns are monotonized to a constant value, in this case the mean f0 value of each speech sample, resulting in a totally flat line. Both techniques were applied using *Praat* scripts.

The combination of monotonization and PURR filtering yielded speech samples with dramatically reduced segmental information and flat intonation. The only cues available to the listener were the temporal aspects of prosody (rhythm and speaking rate). This combination of delexicalization and monotonization was used to test the impact of the mentioned temporal factors in English L2 produced by native speakers of German [8] and dialect identification [9], and it is replicated here to test if the same results can be observed for the English of Italian native speakers.

2.2. Perception experiment

2.2.1. Stimuli

4 sets of stimuli were used for the experiment. 3 were obtained with signal manipulation, 1 consisted of natural speech samples. These were used to create a total of 160 utterances, that is:

- 40 PURR-filtered sentences, 20 spoken by English native speakers (NS) and 20 by nonnative speakers (NNS);
- 40 monotonized PURR-filtered sentences, 20 NS and 20 NNS;
- 20 monotonized-only sentences, NS only;
- 40 unfiltered sentences, 20 NS and 20 NNS.

In the case of the monotonized-only sentences, an auditory evaluation of the stimuli revealed that the presence of segmental information could easily betray the status of the speaker, in agreement to what found in [20]. It was so decided to use only the 20 monotonized sentences uttered by the five native English speakers and to discard the ones spoken by Italians (see section 2.2.3 below).

2.2.2. Subjects

The subjects of the perception experiment were 10 English native speakers, all female undergraduate exchange students at the University of Padova, in Italy. They had all a certain familiarity with the Italian accent and language, but at the time of the experiment none of them had received any specific phonetic instruction.

2.2.3. Experiment design and procedure

The aim of the experiment was to understand the relative contribution of segmental and suprasegmental information manipulating both dimensions using signal manipulation. The perceptual test was conducted using the OpenSesame stimuli presentation program [14].

The stimuli were grouped in blocks by sentence type (A, B, C, D, see section 2.1.1 above). Within each block, the sentences were presented by condition, i.e., PURR-filtered, monotonized and PURR-filtered, monotonized-only, and unfiltered. While the former two conditions were presented in random order, the monotonized and the unfiltered ones always occupied the third and forth position in the order of presentation. This is because of the well-known difficulty in evaluating manipulated speech in which segmental information is degraded ([13, 15]). For this reason, it was decided to present the less informative content-masked versions before the more redundant monotonized-only and unfiltered conditions, which both included a great deal of segmental information (see [9]).

Within all groups three repetitions of each stimulus were presented randomly to the listeners.. As a result, for each of the three mentioned conditions, the tokens to be evaluated were 120 (40x3). As for the monotonized-only sentences,

being only 20 (see section 2.2.1 above), it was decided to repeat them six times in order to maintain a total of 120 repetitions (20x6) also for this part of the experiment. The subjects did not receive any special instructions for this part and were led to think that the monotonized sections would include both native and foreign-accented stimuli, although in fact there were only stimuli produced by native speakers.

The orthographic transcription of each sentence to be evaluated was always available to the listeners on screen to make their task less demanding. This solution has been frequently adopted in similar experimental studies about prosody and foreign accent when the interest is not in the actual intelligibility of the sentences but in their prosodic aspects (e.g., [14], [20]). This choice is also motivated by the difficulty of evaluating non-natural speech in which segmental information is removed or degraded. For this reason, the actual test was also preceded by a brief training session, presenting two voices not included in the actual test uttering the same sentence, again not included in the test. The subjects were asked to give their answers in a forced-choice between “English native speaker” and “Italian native speaker”.

3. RESULTS

The listeners’ scores were well above chance level for both the PURR-filtered and the monotonized PURR-filtered sentences, confirming the hypothesis that prosody is an effective clue to detect foreign accent. These results also show that the temporal aspects of prosody preserved in the samples (rhythm and speaking rate) allowed accent detection even in the absence of any intonational pattern.

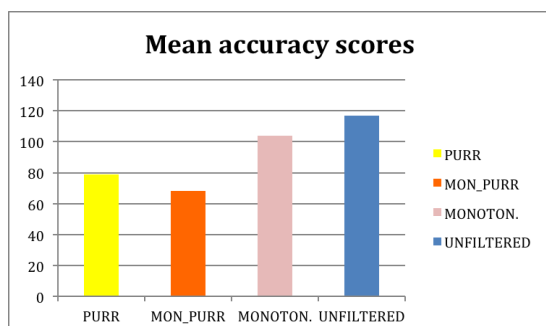
A two-tail unpaired t-test showed that there is a significant difference between the mean scores in PURR-filtered and monotonized PURR-filtered sentences ($t=2.8122$, $p=0.0115$) (see Table 1 and Fig. 1). This finding is in line with the results obtained by [8] and [9] and shows that intonation is a stronger phonetic cue to detect foreign accent when compared to temporal prosodic aspects (see section 4 below).

Another two-tail unpaired t-test showed that there is a significant difference between the scores in monotonized-only English L1 sentences vs. unfiltered sentences ($t=26481$, $p=0.0164$). This finding shows that listeners have a significant tendency to fail to identify native speakers when segmental information is preserved and intonation is monotonized.

Table 1: Mean accuracy scores (with standard deviation) per condition (n=120).

SUBJECT	PURR	MON_PURR	MONOTON.	UNFILTERED
MEAN	79	68	104	117
SD	12,5914433	3,25917507	15,8145503	4,03457277

Figure 1: Mean accuracy scores per condition (n=120).



4. DISCUSSION

The results confirm the tested hypotheses. The accuracy scores in the PURR-filtered sentences and in the monotonized PURR-filtered sentence were both well above chance level, showing that prosody indeed plays a crucial role in the detection of foreign accent.

Among the prosodic cues, intonation seems to have the greatest impact: the significant difference between the scores obtained in the PURR-filtered and in the monotonized PURR-filtered sentences shows that the presence of discernible intonation patterns significantly improves foreign accent detection, as found out by [8] and [9] in similar tests.

However, the fact that listeners performed well above chance level even when intonation patterns and most of the segmental information were removed, show that temporal prosodic aspects (rhythm and speaking rate) are also effective cues in detecting foreign accent.

Finally, the difference between the scores obtained in the monotonized and in the natural conditions was also statistically significant. This result shows that a flat intonation hinders the correct identification of a native speaker.

5. CONCLUSIONS

The results of the experiment show that the nonnative status of a speaker can be successfully detected by listening to speech samples where most of the segmental information is degraded and only prosodic aspects are preserved (intonation,

rhythm, speaking rate). The hypothesis that prosody is a strong cue to detect foreign accent is therefore confirmed.

Although intonation seems to be the strongest prosodic aspect in cueing foreign accent detection (see [8]), native listeners are still able to successfully identify nonnative speakers using temporal cues (rhythm and speaking rate) when intonation patterns are removed.

If intonation patterns in native productions are monotonized, listeners tend to make more mistakes in identifying the native/nonnative status of a speaker. This finding can be seen as perceptual evidence in support of the findings that nonnative English intonation is flatter than the one produced by native speakers ([14, 12]).

Although the results of this study confirmed the hypotheses, a word of caution in interpreting the results is in order. First, only 10 native listeners were tested. It would be interesting to expand the number of subjects to test if the results of this study are really generalizable. Second, the use of read speech should be accompanied by spontaneous speech, in order to see if there is a difference in perceiving foreign accent between read and freer speech. The literature has produced contrasting evidence in this regard [16], and it would be interesting to shed more light on this aspect too. Besides, all the results of perception experiments with speech manipulation techniques should be considered with caution because the listeners' judgments might not correspond to the perceptions in face-to-face interaction [13]. Finally, one cannot completely rule out the possibility that subjects' relative familiarity with Italian could have been played a role as a facilitating factor in accent detection [7].

To conclude, this study has confirmed the importance of prosody in the detection of foreign accent in a language combination (Italian L1 – English L2) that has only recently started to be tackled by studies of prosody and foreign accent (e.g., [14], [2], [3]). The methodology adopted proved to be suitable to test the experimental questions and will be enhanced in future studies in order to confirm the trends reported in this work.

6. REFERENCES

- [1] Boersma, Paul & Weenink, David. 2012. Praat: doing phonetics by computer [Computer program]. Version 5.3.19, retrieved 24 June 2012 from <http://www.praat.org/>
- [2] Busà, M.G., 1995. L'inglese degli italiani. L'acquisizione delle vocali. Padova: Unipress.

- [3] Busà, M.G. 2009. Effects of L1 on L2 pronunciation: Italian prosody in English. In A. Gagliardi, A. Maley (ed.), *ILS, ELF, Global English: Teaching and Learning Processes, Linguistic Insights: Studies in Language and Communication*. Bern: Peter Lang, 200-228.
- [4] Boula de Mareüil, P., Vieru-Dimulescu, B. 2006. The contribution of prosody to the perception of foreign accent. *Phonetica*, 63(4), 247-67.
- [5] De Meo, A., Pettorino, M., Vitale, M. (in press) "Transplanting Credibility into a Foreign Voice. An Experiment on Synthesized L2 Italian". *Proc. GSCP 2012 International Conference, Belo Horizonte*.
- [6] Duguid, E. 2001. Italian Speakers. In Swan, M. (Ed.), *Learner English: A Teacher's Guide to Interference and Other Problems*. Second Edition. Cambridge: Cambridge University Press, 73-89.
- [7] Gass, S. and Varonis, E. 1984 'The effect of familiarity on the comprehensibility of nonnative speech' *Language Learning* 34, 65-89.
- [8] Jilka, M. 2000. Testing the Contribution of Prosody to the Perception of Foreign Accent. *Proc. New Sounds*, Amsterdam, pp. 199 - 207.
- [9] Leyden, K., van & Heuven, V.J.J.P., van (2006). On the prosody of Orkney and Shetland dialects. *Phonetica*, 63, 149-174.
- [10] Ladd, D.R., 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
- [11] Mathôt, S., Schreij, D., Theeuwes, J. (in press). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*.
- [12] Mennen, I. 2007. Phonological and phonetic influences in non-native intonation. In J. Trouvain & U. Gut (eds.), *Non-Native Prosody. Phonetic Description and Teaching Practice*. Berlin: Mouton de Gruyter.
- [13] Munro, M.J. 1995. Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition*, 17(1), 17-4.
- [14] Munro, M.J., Derwing, T.M., Burgess, C.S. 2010. Detection of nonnative speaker status from content-masked speech. *Speech Communication*, 52(7-8), 626-637.
- [15] Ohala, J., Gilbert, J.B. 1981. Listener's Ability to Identify Languages by their Prosody. *Studia Phonetica*, 19, 123-131.
- [16] Piske, T., Mackay, I.R.A., Flege, J.E. 2001. Factors affecting degree of foreign accent in an L2 : a review. *Journal of Phonetics*, 29(2), 191-215.
- [17] 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.
- [18] Sonntag, G.P., Portele, T. 1998. Comparative evaluation of synthetic prosody with the PURR method, *Proc. ICSLP 98 Sydney*, 3-6.
- [19] Trofimovich, P., Baker, W. 2006. Learning Second Language Suprasegmentals: Effect of L2 Experience on Prosody and Fluency Characteristics of L2 Speech. *Studies in Second Language Acquisition*, 28(01), 1-30.
- [20] Van Els, T., De Bot, K. 1987. The role of intonation in foreign accent. *The Modern Language Journal*, 71(2), 147-155. Blackwell Publishing; National Federation of Modern Language Teachers Associations.