Transplanting credibility into a foreign voice: an experiment on synthesized L2 Italian

Massimo PETTORINO, Anna DE MEO, Marilisa VITALE
University of Naples "L'Orientale"
Naples, Italy
mpettorino@unior.it, ademeo@unior.it, vitalem@unior.it

Abstract

This study intends to verify through perceptual tests conducted on original and artificially modified speech whether a relationship exists among the degree of comprehensibility of an utterance, the foreign accent and the credibility of the message. Four bizarre-but-true news read in Italian by four non-native speakers were artificially modified with Praat and WaveSurfer. Each piece of news was transplanted, so that segmental and prosodic features of a text read by a native speaker were transferred onto the same text uttered by a non-native speaker. The corpus was administered to 265 native Italian listeners, who were requested to indicate the degree of comprehensibility, the level of foreign accent and the truthfulness of each item. The results point out the existence of a close inverse relationship between comprehensibility and credibility. The presence of foreign accent, providing an impediment to the understanding of the message, tends to create an attitude of distrust in the listener. The most important features for the foreign accent reduction are the suprasegmental ones and, in particular, the durations of the phones and the pitch movement.

Keywords: foreign accent; comprehensibility; credibility; L2 Italian; prosody.

1. Introduction

Our recent study on socio-cultural effects of foreign accent on communication effectiveness (De Meo *et al*., 2012) revealed the relevance of comprehensibility factors - such as disfluency, frequency of silences, pitch range variation, silent pauses, segmental errors - on message credibility. A hundred seventy-five native Italian listeners, after hearing a set of 10 news uttered in Italian by one native speaker of Italian and four non-native speakers of L1 Chinese, Vietnamese, Arabic and Japanese, were asked to assess the comprehensibility, i. e. listener's estimation of difficulty in understanding an utterance (Munro & Derwing, 1999), and the truthfulness of each news item. The four non-native speakers, all late bilinguals with a basic (A2) and a mid (B1) level of competence as laid out in the Common European Framework of Reference, and an average stay time in Italy of 6 months, were chosen for the study after a global foreign accentedness rating test which was administered to 70 male and female native Italian listeners. Listeners rated the degree of foreign accentedness of a short read text on a 4-point scale (0 = native speaker; 3 = strong foreign accent). The results allowed to select four L2 speakers of Italian with a strong foreign accent.

Ten bizarre-but-true news from around the world read by the native speaker and the four non native speakers were presented to native listeners in form of radio news magazines, each combining the four voices reading different news, same news sequences but random voice order, pretending to administer a survey on media reliability, in order to avoid to focus the attention on foreign voices.

Obviously each single piece of news revealed to have its own degree of credibility, in accordance with the textual content of the message. However results showed that, within the same text, ratings were significantly different depending on the auditory comprehensibility level.

The study showed that when there are no comprehensibility problems the assessment of real/false is maintained around 50%, so in a range of randomness. On the contrary, when the level of comprehensibility lowers, due to various acoustic factors (disfluencies, errors, percentage of silence, tonal variation, etc.), the judgments of "false" increase rapidly, reaching 90% when the statement proves to be poorly understandable for the 40% of listeners. Therefore, there seems to be a threshold of comprehension tolerance, i.e. a level of difficulty in understanding an utterance at which the listener's effort to understand the message leads him to believe that what he has just heard is not credible.

Following this line of research, our current study intends to carry out a perceptual test on artificially modified speech, in order to evaluate the role played by both segmental and suprasegmental features in the achievement of an L2 comprehensible communication and find out if there is a relationship between the perceived degree of foreign accent and credibility.

2. Methods and materials

The corpus used for this study, taken from the one used in De Meo *et al*. (2012), consists of 4 news artificially modified with Praat (Boersma & Weenink, 2012). Each single piece of news was manipulated, so that disfluencies and errors were removed, and the prosodic features of the native speaker's utterances were transferred onto the same utterances produced by the non-native speakers (prosodic transplantation technique).

2.1 Corpus and Informants

1) Informants: 5 female voices
- 1 Italian speaker (L1)
- 4 L2 Italian speakers (Chinese, Vietnamese, Japanese, Arabic L1s)

2) Corpus: 18 audio files (bizarre-but-true news)
- 8 original news (4 L1, 4 L2)

Heliana Mello, Massimo Pettorino, Tommaso Raso (edited by), *Proceedings of the VIIth GSCP International Conference : Speech and Corpora*
ISBN 978-88-6655-351-9 (online)

- 10 L2 artificially modified (4 with removed disfluencies and cloned pauses, 2 with removed errors, 4 with cloned durations and pitch contour).

2.2 The transplantation technique

The rhythmic-prosodic transplantation technique is based on the algorithm PSOLA (pith-synchonous overlap-add, Moulines & Charpentier, 1990), implemented in Praat and illustrated in Yoon (2007) with regard to the English productions of Korean speakers. 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 intensity contour and the silent pauses.

The procedure of transplantation must follow a well-defined sequence of steps, since each of them is preparatory for the subsequent ones. The five phases are: anomalies treatment (disfluencies removal, pause cloning, errors elimination), segmentation and labelling, transplantation of the duration, intensity transplantation, pitch contour superimposition.

This technique seems to be a rather effective tool for the study of the spoken L2, since the manipulation of an utterance allows to evaluate the role played by individual acoustic parameters at the pragmatic-communicative level.

2.3 Perceptive test

The whole corpus was administered in a randomized order to 265 native listeners (male and female, mean age 21, university students) organized into 5 groups, so that nobody could listen to the same news more than once. As the purpose of the survey was to assess the credibility, the repeated exposure to a same input would have affected the reliability of the test results.

For each utterance, listeners were asked to evaluate the comprehensibility (poor, sufficient, good), assess the degree of perceived foreign accent (native accent, mild foreign accent, strong foreign accent) and judge on its truthfulness (true/false).

3. Results and discussion

In this section we will examine the results of the abovementioned test, in order to evaluate the relevance of each manipulated factor on the perceptual level. The discussion will be organized into three parts corresponding to the different steps of the synthesis procedure. For the data analysis the One-Way ANOVA was performed.

3.1 First step: Removing disfluencies and cloning native silences

Figures 1, 2, 3 show the average percentage values of the judgements given to the utterances, both original and modified, produced by the native (NS) and the non-native speakers (NNS), with respect to comprehensibility, degree of foreign accent and credibility.

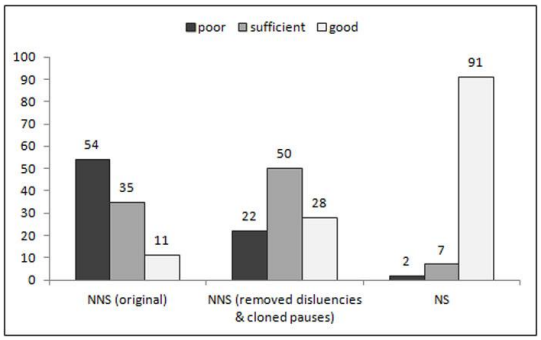

Figure 1: Average comprehensibility values (%) of the NS and the NNS

As for the foreign accent, both the NS and the NNSs were correctly recognized by almost all the listeners (Figure 2). The modifications carried out on NNSs' utterances produced a decrease of about 20% of the judgments of "strong foreign accent" (from 79% to 60%). In addition, it is worth noting that the 5% of the listeners assumed to have heard a native voice. Data are statistically significant ($p<0.001$).

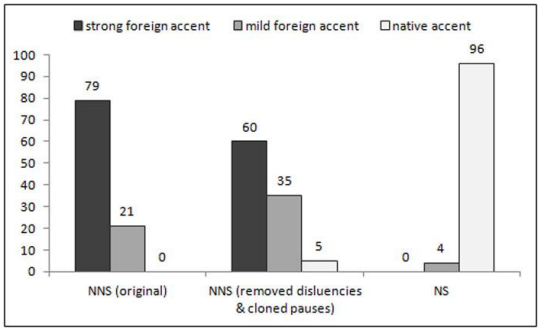

Figure 2: NS and NNSs' average percentage values of the foreign accent ratings

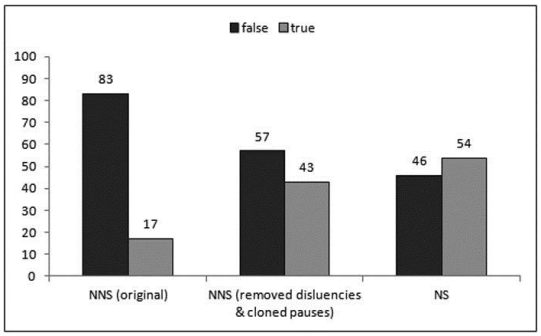

Figure 3: NS and NNSs' average percentage values of credibility

The removal of disfluencies and the repositioning of the silences determine a statistically significant improvement ($p<0.001$) of the NNSs' utterances comprehensibility (Figure 1). As a result of the manipulation, the majority of the listeners (78%) judged the non-native productions at least sufficiently comprehensible. Obviously, the NS proved to be highly comprehensible.

Removal of disfluencies and changes of silences (Figure 3) determined a significant increase ($p<0.005$) in the level of news credibility (+26%), taking the NNSs' values to levels very similar to those obtained by the native speaker.

3.2 Second step: Errors removal

For the second step of the study data are limited to the A2 level NNSs, since in the other speakers' productions there were no particular segmental irregularities. Using WaveSurfer, phones perceptually detected as wrong by three native trained phoneticians were artificially modified or substituted through a self-transplantation procedure, i.e. using adeguate micro-segments produced by the speaker within the same utterance.

Because of the large variability and unpredictability of the errors, this phase is technically the most problematic. The typology and frequency of errors require operations that may damage the quality of the synthesized audio file and interfere with the perceptive evaluation.

Our data show that the segmental modifications give anyhow rise to a slight but significant improvement in terms of foreign accent assessment ($p<0.005$; "strong" from 77% to 66% and "mild" from 20% to 31%). No significant variations were observed for the comprehensibility and the credibility ($p>0.05$).

3.3 Third step: Duration and pitch transplantation

The final step of the transplantation procedure involved the cloning of the duration of each segment and, subsequently, the superimposition of the intonation contour from the NS's utterances to the NNSs' ones. The perceptive test outcomes are generally satisfactory. Figure 4, concerning comprehensibility, shows that, if compared to the first step of the procedure (disfluencies removal and silences cloning), the negative judgments decreased by 10% (from 23% to 12%) in favour of the "sufficient" ratings, while the "good comprehensibility" values did not change ($p<0.05$). It should be noted that the results of the overall transplantation process, when compared to the original utterances, reveal a remarkable improvement: the "poor comprehensibility" lowers by 42% and the "good comprehensibility" rises by 16% ($p<0.001$).

The most evident effects of this last step are those related to the degree of foreign accent (Figure 5), with a gain of about 30% for the judgment of "native" and a 60% reduction with regards to the judgment of "strong foreign accent" ($p<0.001$).

Finally, the values of credibility (Figure 6) do not undergo further significant variations ($p>0.05$).

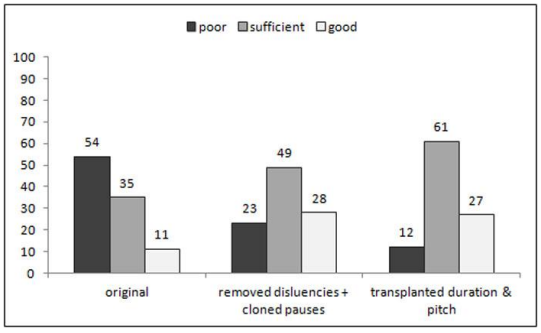

Figure 4: Average comprehensibility values (%) of the NNS

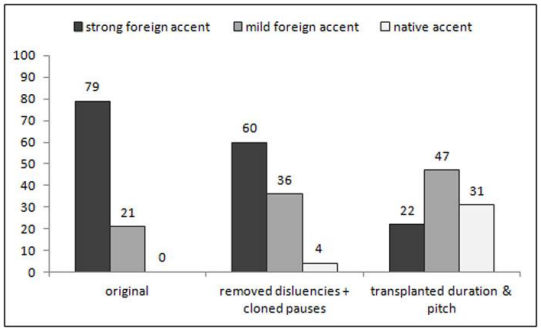

Figure 5: NNSs' average percentage values of the foreign accent ratings

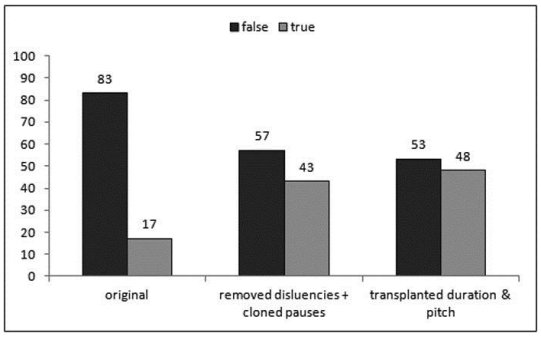

Figure 6: NNSs' average percentage values of credibility

4. Conclusion

In conclusion, the study confirms the existence of a close relationship between comprehensibility and credibility, both for original and manipulated audio files ($p<0.001$). The more the utterance is easy to understand the more the listener is led to believe true what he/she has just heard. In this perspective, for our informants, beginners (A2) and low-intermediate (B1) speakers of L2 Italian, whose speech is characterized by disfluencies, anomalous silences, segmental errors, and inappropriate pitch contour, foreign accent provides an impediment to the understanding of the message and, consequently, tends to create an attitude of distrust in the listeners. However, it is worth to emphasize that it is not the "foreigness" as such to cause a lowering of credibility, but it is rather the difficulty of decoding the message determined by the presence of anomalies typical of an early L2 speech.

5. References

Boersma, P., Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. Version 5.3.23, retrieved 7 August 2012 from <http://www.praat.org/>.

De Meo, A., Pettorino, M. and Vitale, M. (2012). Non ti credo: i correlati acustici della credibilità in italiano L2. In G.. Bernini, C. Lavinio, A. Valentini and M. Voghera (Eds.), *Atti dell'XI Congresso dell'Associazione Italiana di Linguistica Applicata. Competenze e formazione linguistiche. In memoria di Monica Berretta*. Guerra Edizioni, Perugia, pp. 229--248.

De Meo, A., Vitale, M., Pettorino, M. and Martin, P. (2011). Acoustic-perceptual credibility correlates of news reading by native and Chinese speakers of Italian. In *Proceedings of the 17th International Congress of Phonetic Sciences.* Hong Kong, PRC, pp. 1366--1369.

Moulines, E., Charpentier, F. (1990). Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones. *SpeechCommunication*, 9, pp. 453--467.

Munro, M.J., Derwing, T.M. (1999). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. In *Language Learning*, 49(s1), pp. 285--310.

WaveSurfer 1.8.8. Available at : <www.spee.kth.se/wavesurfer/>.

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), pp.197--215.