Influence of semantics on the perception of corrective focus in spoken Italian

Sonia Cenceschi, Licia Sbattella, Roberto Tedesco
Dept of Electronics, Information and Bioengineering, Politecnico di Milano, Italy
https://doi.org/10.36505/ExLing-2018/09/0006/000339

Abstract

This study is a web-based, psychoacoustic test for adult, Italian native-speakers, investigating detection of different prosodic phenomena in Standard Italian utterances. The purpose was to investigate the influence of semantics on human ability to recognise different prosodic aspects, in order to understand the basic pieces of information involved into the psychoacoustic process of verbal comprehension. In particular, one section of the test regarded the ability to recognize the presence of a Corrective Focus, which is a spoken constituent that is a direct rejection of an alternative. Results show Corrective Focus seems difficult to detect into isolated audio utterances. Semantics seems to improve detection accuracy; phonotactics, instead, seems not to add useful information; finally, our test confirms correlation with prominent syllables.
Key words: contrastive focus, prosody, perception, semantics, psychoacoustics

Introduction

Usually, speaker express emotions or introduce a new topic/concept into the dialog by adding acoustic stress or pronouncing more clearly one or more syllables of their speech. In the context of a dialog, the Corrective Focus (CF) is a particular kind of stress, where the current speaker's intention is to correct a concept introduced by the other speaker in the previous dialog turn (Gussenhoven 2008). The acoustic realization of CF depends on culture and language of the speaker (Bosch and van der Sandt 2009).

The Standard Italian language is strongly syllable-timed: syllables take approximately an equal amount of time to be pronounced, and they are temporally stretched by speakers when they intend to underline a word. Prominence is also characterized by changes in fundamental frequency excursion and intensity-related parameters, with respect to their average values. Finally, listener's expectations (i.e., how the speaker believes her/his interlocutor will react) affects how prominence is perceived (Tamburini, Bertini and Bertinetto 2014).

The aim of this study is to investigate how much semantics – conveyed by syntax and lexicon – and phonotactics affect the human ability to detect CF in real sentences. This experiment focuses on

isolated sentences, so the listener's expectations are not considered because they can only be detected in dialogs.

The experiment

The subjects were presented with simple questions; in particular, for the test on CF, the question was: *In which one of these audios do you perceive the wish to correct the interlocutor? – Example: "I want BREAD, not meat"* (the uppercased word was the one with CF). Then, the subjects started hearing three audio fragments – selected at random in two sets: one containing utterances (more precisely, Units of Intonation - UIs) with CF, and one with all other UIs – and checked audios that she/he recognized as carrying a CF. The interface allowed the subject to hear each audio fragment several times.

The UIs were taken by the SI-Calliope corpus (Cenceschi, Sbattella and Tedesco 2018), recorded by professional speakers (i.e., the corpus contains recited speech), 7 women and 7 men. Each test was conducted for three UI typologies: *real words* (where the audio fragments contained regular Italian words, and thus preserved all the syntax, lexicon, phonotactics, and acoustic information), *pseudo-words* (where the audio fragments contained invented words, with a "sound" similar to the one of real Italian words, and thus only preserved phonotactics and acoustic features, while removing syntax, and lexicon), and *pitch envelopes only* (where the audio was restricted to a pitch contour, reducing the acoustic features to a minimum, and removing all other pieces of information). We also collected subjects' age, region, and gender.

To generate pseudo-word UIs, we started from the CoLFIS corpus of Italian words (Bertinetto 2005), where we removed every word containing characters in the {w, y, j, k, x} set, and every word containing characters with diacritical signs different from acute and grave accents. Then, the remaining words where split into syllables by means of Hyphenator 0.5.1 (Berendsen 2013), a Python module that leverages the OpenOffice hyphenation dictionary. Finally, we trained a trigram of syllables that thus encoded an approximation of the Italian phonotactics. Given a real-word UI, the algorithm leverages the trigram to generate, for each word, a random pseudo-word composed of the same number of syllables. As an example, from the real-word UI "*Domani è bel tempo!*" we obtained the pseudo-word UI "*Selèzio è bel àmmi!*".

Pitch envelopes were computed with the Praat (Boersma 2002) to Pitch command, smoothed with a value of 5Hz, and then used to generate a sound by means of the `hum` command.

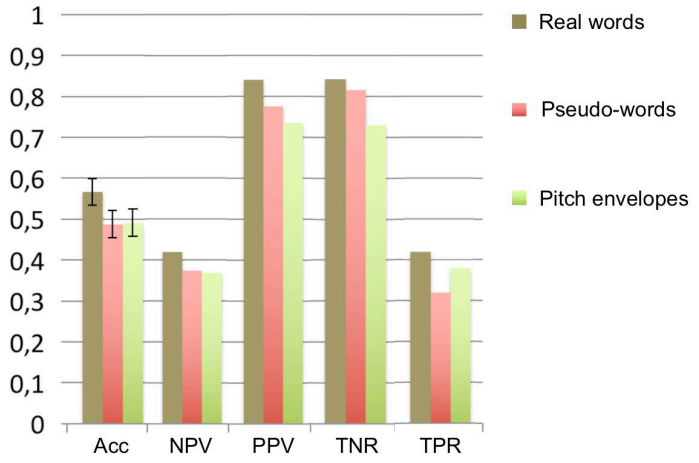

Figure 1. Accuracy (error bars: binomial at 95%), Negative Predictive Value, Positive Predictive Value, Specificity (TNR), and Sensitivity (TPR).

Results

We collected 306 tests. We found a possible correlation between accuracy in recognizing CF and the subject's origin; data, however, were not conclusive. We did not find relevant correlations with subjects' age and gender.

Figure 1 shows statistical results about perception of CF. The overall results showed that recognising CF required a combination of semantics and vocal clues. In particular, the *Accuracy* was 0.566 for real words, 0.488 for pseudo-words and 0.491 for pitch envelopes. The t-test confirmed with $p<0.001$ that UIs with real words were simpler to understand; accuracies of pseudo-words and pitch envelopes, instead, were not significantly different. This result highlights how important the "prediction" process – allowed by semantics – is in perceiving CF; our test also confirms that the fundamental frequency envelope affects the perception of CF (Terken 1991), while other acoustic features and the phonotactics do not add further information.

The *Negative Predictive Value* (the fraction of UIs recognized as not carrying CF, which are actually not carrying CF) is much lower than the *Positive Predictive Value* (the fraction of UIs recognized as carrying CF, which are actually carrying CF). Moreover, the *Specificity* (the fraction of UIs not carrying CF, which are correctly identified as such) is much higher than the *Sensitivity* (the fraction of UIs carrying CF, which are correctly identified as such). This behaviour is found in all UI typologies.

These results suggest subjects were very selective in perceiving CF: they tend not to asses its presence unless it was clearly perceivable. In this way they often missed a CF but were rarely wrong in recognizing it.

Conclusion and future works

CF seems very difficult to detect into isolated UIs; this could be justified by the fact that CF exists because there is a *dialogue*, and so it is probably better perceived if contextualized (Kakouros and Räsänen 2016). Thus, in a future experiment we could provide the subjects with a dialogue where the last UI could carry the CF. Moreover, CF recognition showed similar results for pseudo-words and pitch envelopes; this result confirms that CF is related to syllable's prominence and thus to the F0 contour and duration, while other acoustic clues and phonotactics do not add useful information. Anyway, semantics seems to play a crucial role, as real-word UIs reached a (slightly but measurable) better accuracy. Finally, we did not find relevant correlations with subjects' age and gender, while there were hints of a possible correlation with subjects' geographical origin.

Reference

Berendsen, W. 2013. Available on: https://pypi.python. org/pypi/hyphenator/0.5.1/.

Bertinetto, P.M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., Thorton, A.M. 2005. CoLFIS. Corpus and frequency lexicon of written Italian http://linguistica.sns.it/CoLFIS/Home.htm.

Boersma, P. 2002. Praat, A system for doing phonetics by computer. Glot international 5, 41-345.

Bosch, P., van der Sandt, R. (Eds.). 1999. Focus: Linguistic, cognitive, and computational perspectives. Cambridge University Press, 30-31.

Cenceschi, S., Sbattella, L., Tedesco, R. 2018. Towards Automatic Recognition of Prosody. Proc. 9th International Conference on Speech Prosody, 319-323, Poznań, Poland.

Gussenhoven, C. 2008. Types of focus in English. In Topic and focus, 83-100. Springer.

Kakouros, S., Räsänen, O. 2016. Perception of sentence stress in speech correlates with the temporal unpredictability of prosodic features. Cognitive science 40, 1739-1774. Wiley Online Library.

Liberman, M., Pierrehumbert, J. 1984. Intonational invariance under changes in pitch range and length. In Aronoff, M, Oehrle, R. (eds.) 1984, Language and Sound Structure, 157-233. Cambridge: MIT Press.

Tamburini, F., Bertini, C., Bertinetto, P.M. 2014. Prosodic prominence detection in Italian continuous speech using probabilistic graphical models. Proc. Speech Prosody, 285-289, Dublin, Ireland.

Terken, J. 1991. Fundamental frequency and perceived prominence of accented syllables. The Journal of the Acoustical Society of America 89(4), 1768-1776.