Cross-linguistic analysis of two speaking styles in Brazilian Portuguese and German by using the quantitative Target Approximation model

Plinio BARBOSA (UNICAMP)
Hansjörg MIXDORFF (Beuth University of Applied Sciences)
Sandra MADUREIRA (PUC-SP)

The Parallel ENcoding and Target Approximation (PENTA) model was proposed by Xu (Xu, and Wang, 2001) to associate tone and intonation into a single generative approach. Its main principle relies on the assumption that the communicative functions control F0 contours via specific (and parallel) encoding schemes. These encoding schemes specify the values of the melodic primitives, which include domain-restricted pitch target, pitch range, articulatory strength and duration. The values of the melodic primitives can be specified both symbolically and numerically. Since the encoding schemes are hypothesised to be language-specific, the goal of this paper is to study the differences between Brazilian Portuguese (BP) and German in implementing the functions of signalling prominence and boundary in two speaking styles, read speech and narration.

At the end of the PENTA model generation, the F0 contour is implemented by the quantitative Target Approximation model (qTA) (Prom-on, and Xu, 2010). The quantitative value of the melodic primitives are used together to specify an asymptotic contour given by the following equation: F0(t) = (c1 + c2.t + c3.t^2).exp(−lambda.t) + m.t + b, where lambda, m, and b are respectively target strength, target slope, the target height. The coefficientes c1, c2 and c3 depend on the initial conditions. Thus, given these conditions, three parameters completely specify the modelled F0 contour for a given linguistic segment.

These model parameters were obtained from an analysis-by-synthesis learning process (Xu, and Prom-on, 2010), which searches the optimal values for target slope, height and strength inside a linguistic segment (here, the word) in order to fit the original F0 contour. The algorithm runs onto Praat (Boersma, and Weenink, 2010). For investigating the intonational differences for implementing prominence and boundary in BP and German we evaluated the patterns of the three qTA parameters according to four labels associated to each word of the corpora in the two languages: prominent (p), non-prominent (n), terminal boundary (t), and non-terminal boundary (c). Each word was labelled with one among these four labels by two experts in each language after hearing the utterances selected for analysis. Although is possible for a word to signal both prominence and boundary (usually in different acoustic segments of the uttered word), we labelled prominent, preboundary words with the boundary labels only (t or c).

The corpora consist of parallel productions in BP and German. Two native female and four native male speakers in both languages read a 1,500-word text on the origin of the pastries pastéis de Belém (reading style, RE) in their own language. The BP text is an adaptation of a text in European Portuguese, whereas the German text was translated from the BP text by the second author. The translation was carried out sentencewise to make cross-linguistic comparisons easier. After the reading, all subjects told what the text was about (storytelling style, ST). All speakers were aged 30 to 45 years, and were students with a Linguistics or a Computer Science background. The analyses shown here are based on excerpts from 150 to 200 words in each language and style.

As for the results, two measures of performance of the lerning algorithm were used: Root Mean Square Error (RMSE) and correlation medians, both computed for the first thirteen sentences of the read material (188 words) and ten utterances (from 150 to 200 words) in the ST style for the two languages, showed a fair fit: for BP, RMSE was between 1 (RE) and 1.2

semitones (ST), and for German, between 1.3 (RE) and 1.4 semitones (ST). As for correlations, they were 99% for BP in both styles and of 90% (RE) and 92% (ST) in German. As for the qTA model parameters, it is important to check the behaviour of non prominent words, in order to see whether target values need to be specified for these words, as suggested by Prom-on and Xu (2010) when studying Mandarin. Both languages behave exactly the same way as for target slope, showing a concentration of values around zero, and additional peaks at the ends of the target slope distribution, related to movements of F0 just before or after realizing a peak associated to prominence. This seems to indicate a transitional contour between F0 peaks.

As for target height, there are at least three important differences between the two languages. For the narratives, F0 within non prominent words tend to be higher in BP than in German in the ST style. The main reason for that is the fact that, when narrating, German speakers lower very quickly their F0 values after marking a word as prominent, whereas the BP speakers lower much lesser, often maintaining a relative high level of F0 between two prominences. As for the target values in words asssociated with a non terminal function, German speakers have mean heights 5 (RE) and 3 (ST) semitones re 1 Hz above the BP speakers, which is related with the difference in F0 contour shapes for signalling non terminality: in German this is realized with a rising contour, whereas in BP, with a rising, short falling contour. In reading, terminality in BP is signalled with a mean height 3 semitones lower than in German. However, this last result could be related to differences in F0 floor in both groups of speakers in the two countries.

References

Boersma, P., Weenink, D. Praat: doing phonetics by computer (Version 5.1.37) [Computer program]. Online: http://www.praat.org. 2010.

Fujisaki, H. and Hirose, K. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. Journal of the Acoustical Society of Japan (E), 5(4): 233–241, 1984.

Prom-on, S., Xu, Y. Articulatory-Functional Modeling of Speech Prosody: A Review. Proc. Interspeech 2010, Makuhari, 46–49, 2010.

Xu, Y., Prom-on, S., PENTAtrainer.praat. Online: http://www.phon.ucl.ac.uk/home/yi/PENTAtrainer/, 2010.

Xu, Y., Wang, Q.E. Pitch targets and their realization: Evidence from Mandarin Chinese. Speech Communication, 33: 319–337, 2001.