An automatic prosodic transcriber for the P-ToBI system

Wendy Elvira-García[1], Marisa Cruz[2], Marina Vigário[2], Sónia Frota [2]
[1] Laboratori de Fonètica & CLiC, Universitat de Barcelona
[2] Laboratório de Fonética e Fonologia, CLUL, Faculdade de Letras da Universidade de Lisboa

Background

- Prosodic annotation is time-consuming and subjective
- For decades, automatic systems have been designed (Black & Hunt, 1996; Hu et al, 2020, among others):
  - Based on linguistic knowledge, explicit rules
  - Based on predictive statistical models
- For European Portuguese there is no P-ToBI transcriber

Objective

Create a **Praat script** designed to generate **P-ToBI labels** based on the pitch contour given a tier with by-syllable intervals and stress marks.

Apply rule-based techniques available for Spanish and Catalan (Elvira-García et al, 2016) to discover potential differences between intonational phonology implementation in these languages.

Methodology

Materials

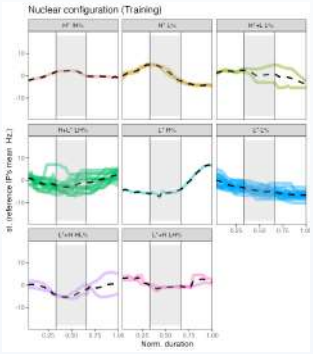
Interactive Atlas of the Prosody of Portuguese (Frota (coord.), 2012-15)

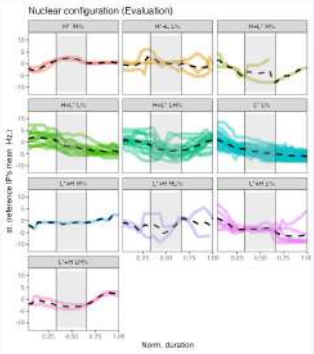
Training corpus

- 96-sentence corpus comprising all Nuclear Pitch Accents (NPA) and Boundary Tones (BT) in European Portuguese (EP).

Evaluation corpus

- 141-sentence corpus comprising all Nuclear Pitch Accents (NPA) and Boundary Tones (BT) in European Portuguese (EP).





Procedure

Input

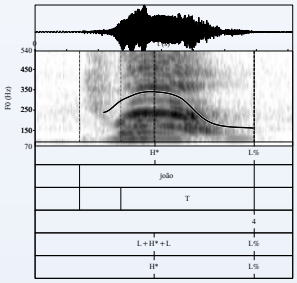
Sound file and Textgrid segmented by-syllable and with a stress mark.

Rules

Differences in pitch level (from L to H or vice-versa) are set as movements greater than 1.5 st. The system looks for differences based on AM theory (around the stressed syllable and at boundaries).

Output

The transcription in a new tier. Figure shows human transcription in tier 1 and automatic (last tier).



Results

Script available at the following address, scan QR:

NPA evaluation

Accuracy 73.8%, with a kappa value of 0.6 (Sensitivity = 0.67, Specificity= 0.93, F-score= 0.63), moderate agreement.

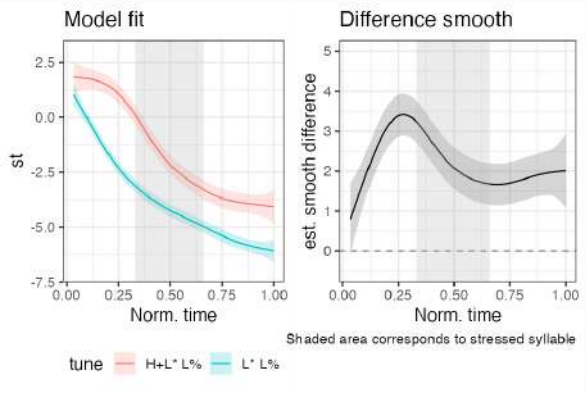
BT evaluation

Accuracy 78.7% , kappa 0.6 (Sensitivity = 0.67, Specificity= 0.95, F-score= 0.58), moderate agreement.

Most mistakes are in falling tunes , specially, H+L* L% and L* L% .

	H*	H*+L	H+L*	L*	L*+H
H*	37	1	3	0	3
H*+L	0	2	1	0	0
H+L*	1	1	36	10	0
L*	0	1	3	25	1
L*+H	2	1	3	6	4

GAMM using 30 pitch points within from the prestressed syllable to the end of the IP (10 points by region) mimicking the data and pitch extraction parameters used by the transcriber. Main differences between these PA are in prestressed area (3.5 st)



Discussion & conclusions

The system performs like other available systems for Portuguese (Moniz et al 2014), but its performance does not reach the level of other languages (Hu, 2020; Elvira-García, 2016).

So, Portuguese prosodic system compared to Cat & Sp ones:

- Needs more points to model the curve (non-linear approaches to the representation are more accurate)
- Has more information encoded in the prestressed region.

The findings of this paper will be taken into account for future enhancements to the system and may potentially contribute to modeling Portuguese prosody more effectively in general.

References available in the paper

Speech Prosody 2024 (SP2024)

PTDC/CLE-LIN/119787/201, UIDB/00214/2020 funded by Fundação para a Ciência e a Tecnologia (Portugal) PID2022-139004OA-I00 funded by MCIN/AEI/10.13039/501100011033 and co-funded by ERDF/ESF "A way to make Europe".