Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

An automatic prosodic transcriber for the P-ToBI system

Wendy Elvira-García[1], Marisa Cruz[2], Marina Vigário[2], Sónia Frota[2]

[1]Universitat de Barcelona
[2]Universidade de Lisboa

wendyelviragarcia@ub.edu, {marisac, mvigario, sfrota}@edu.ulisboa.pt

Abstract

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

The system was trained on a 96-sentence corpus comprising all Nuclear Pitch Accents (NPA) and Boundary Tones (BT) in European Portuguese (EP). Evaluation was conducted on a separate corpus of 141 sentences showing a success rate of 73.8% (k=0.6) for NPA and 78.7% for BT (k=0.6). The qualitative analysis of errors, excluding those stemming from the pitch tracking algorithm, exposes challenges in accurately identifying falling NPAs, particularly instances of L*, H*+L, and H+L* followed by a low BT (although they can be accurately distinguished using an additive model). The performance of the system contrasts with results obtained with similar procedures for other Romance languages that get to 90% of success.

We argue that the performance difference stems from principles underlying different ToBI systems (with P-ToBI being more phonological), and specificities of the phonological system of EP, namely word-final vowel reduction and deletion. This suggests that a rule-based approach relying solely on the acoustic signal may not be the most suitable for European Portuguese.

Index Terms: P-ToBI, automatic transcription, prosody

1. Introduction

This study introduces a rule-based Praat script for transcribing the prosody of European Portuguese. The system is designed to generate P-ToBI labels based on the pitch contour given a tier with by-syllable intervals and stress marks.

The field of automatic prosodic transcription has seen significant development and research efforts, particularly in the context of non-linguistic features detection related to states of mind and deception. Since the early works on speech recognition and synthesis in 1980 there has been a continuous interest in automating prosodic labeling. This research direction was initially driven by the advantages of avoiding subjective errors, distractions, and typographical mistakes that can occur with human transcribers as well as by the need to reduce time used for annotation and allowing the annotation of larger amounts of data.

The tasks available include prosodic boundary detection, stress detection and pitch contours classification. Among these, the labeling of contours is the most difficult, and also the one that is more language-dependent [1].

Automatic transcribers for prosody can be broadly categorized based on the data they use for creating the transcription—linguistic knowledge alone (rule-based systems) or a combination of linguistic knowledge and statistical predictions. The latter category often involves the use of predictive statistical models, such as regression and machine learning algorithms trained on large language-specific oral corpora.

Predictive statistical models include systems based on linear regression [2], decision trees [3], [4] or Markov models [5], support vector machine [1] or principal component analysis [6], among others. Other systems focus more on the stylization of the F0 curve and less on the labels used which are often reduced to marking peaks and valleys [7] [8].

Rule-based systems, on the other hand, rely solely on linguistic knowledge. In these systems, phonetically trained experts design a set of conditions (rules) based on the intonational phonology of the language to label utterances. Examples include IPO approaches for Spanish [9], [10] and French [11], [12].

The tool presented in this study falls into the rule-based category, emphasizing the use of linguistic knowledge for prosodic labelling in EP.

Out of the possibilities for analyzing EP prosody we have chosen to use P-ToBI [13]. P-ToBI's systematic categorization and detailed labelling of prosodic features make it particularly advantageous for linguistic description. ToBI systems in general are widely adopted but they are also highly time-consuming, therefore, the development of transcription tools that speed up the transcription has been ongoing since the creation of ToBI systems [14].

The language-specific nature of ToBI systems proved by the difficulty of adapting the first ToBI automatic transcriber, AuToBI (designed for English) [1] to other languages [15] has led to the creation of transcribers tailored for various languages, including, Korean [3], [4], Chinese [16] , Japanese [17], Dutch [6] or Spanish and Catalan [18].

In the case of Portuguese, there are limited systems available for transcribing prosody, such as one for the automatic detection of prosodic boundaries in Brazilian Portuguese spontaneous speech [19] and one that uses musical notation [20].

There has been also attempts to adapt AuToBI but only for prominence detection (not pitch contours classification) getting to a 74% of success [21]. Therefore, and given that prominence and boundary detection can be performed using systems already available, at least for the Brazilian variety of Portuguese, we

10.21437/SpeechProsody.2024-223

focus on presenting a system for the classification of pitch events in EP and its transcription using P-ToBI labels.

2. Methods

2.1. Materials

In order to create and evaluate the tool, two corpora have been used. The first one is a training corpus that has been used to create the rules and adapt them to the specific pitch alignment and range in Portuguese.

The second corpus has been used for evaluation purposes to test the reliability of the tool with new utterances (i.e., different from the ones used in the training corpus). The following two subsections specify the characteristics of both corpora.

2.1.1. Training corpus

Hand-labeled corpus that contains 96 sentences specially selected for training intonational labels. The sentences were selected from [22]–[24].

These data were selected as an "easy data-set" where contours were prototypical and without practically F0 errors apart from those derived from final vowel unvoicing.

The corpus contains samples coming from read and spontaneous speech as well as sentences elicited with a Discourse Completion Task questionnaire [25]. The types of sentences included were declarative sentences, yes-no questions (neutral and focused), calling contours (greeting and insistent), requests and commands. The sentences were produced by 6 speakers.

Table 1: *Frequency of nuclear pitch accents and boundary tones in training data.*

	Contour	Frequency
	L*	23
NPA	H*	15
	H*+L	8
	H+L*	39
	L*+H	11
	L%	46
BT	H%	3
	!H%	7
	LH%	33
	HL%	7

As a result, we obtained a corpus that contained all Nuclear Pitch Accents (NPA) and Boundary Tones (BT) identified for EP in [22], [23] (Table 1). Pitch accents with more phonetic variation, thus more prone to a difficult characterization, are more represented in the database. For example, a H% BT is easily characterized, from a phonetic point of view, as a rising F0 in a domain final syllable and thus only has 3 instances in the training corpus. But falling NPA, such as H+L* or L* are more difficult to distinguish and therefore more instances were needed in the training phase. Moreover, these falling accents are especially abundant in EP since they are used to signal statements and yes/no questions. The stylizations of the pitch accents included in the corpus are illustrated in Figure 1.

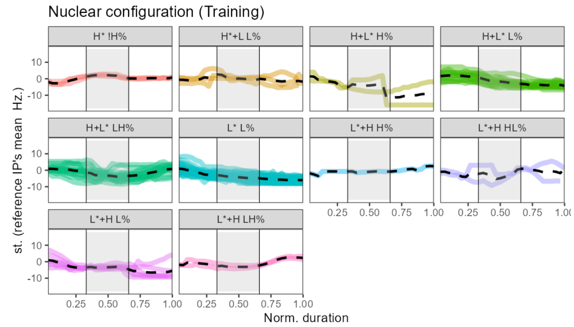

Figure 1: *Nuclear configurations colored by intonational pattern included in the training phase (human transcription).*

2.1.2. Evaluation corpus

The evaluation was conducted on a separate corpus of 141 sentences. This second corpus comes also from [23] and it also contained read and spontaneous speech as well as sentences elicited with the Discourse Completion Task questionnaire. The corpus contained the same sentence types included in the training phase, that is, declarative sentences, yes-no questions (neutral and focused), calling contours (greeting and insistent), requests and commands, resulting in the pitch contours in Table 2.

Table 2: *Frequency of nuclear pitch accents and boundary tones in evaluation data.*

	Contour	Frequency
	L*	30
NPA	H*	44
	H*+L	3
	H+L*	52
	L*+H	16
	L%	93
BT	H%	4
	!H%	7
	LH%	31
	HL%	6

For the evaluation corpus, we did not control for devoicing cases or pitch-tracking errors. As can be seen in Figure 2, this led to some instances that were part unvoiced (seen as broken lines in the figure) or octave jumps due to fricatives' noise or bursts.

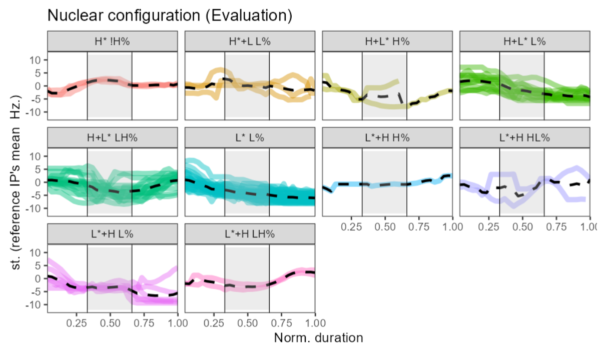

Figure 2: *Nuclear configurations colored by intonational pattern included in the evaluation phase (human transcription).*

2.2. Procedure

The system consists of a Praat script that takes as input a .wav file and a matching Textgrid, extracts and stylizes the F0 contour and uses a set of rules to annotate its intonation. After that, it saves the TextGrid and, optionally, a high-resolution picture. The phases of the analysis pipeline are detailed in the following subsections.

2.2.1. Input

The input of the system consists of a sound in wav format and a TextGrid. The TextGrid must have a syllable segmentation tier and a stress mark (usually a [ˈ] within the stressed syllable). The TextGrid can optionally have a Break Indices (BI) tier. If BI are annotated, the system would look for each 4 level and consider the Intonational Phrase (IP) independently for its transcription. The system will also annotate intermediate boundary tones (such as H- or L-) if the respective BI are annotated. If no BI tier is provided the whole sound file will be considered as an IP.

2.2.2. F0 detection and rules

As a Praat script, the tool uses the Praat autocorrelation algorithm to extract a Pitch object. To minimize F0 tracking errors due to incorrect pitch range settings we used a two-pass method [26], using the parameters for "expressive speech" [27] because Portuguese pitch movement within the syllable can be an octave.

In the case of prenuclear pitch accents, the system measures pitch at 3 points at the prestressed syllable (start, mid and end) stressed (mid and end) and post-stressed (mid and end). Note that the end point of the prestressed syllable is also the start point of the stressed syllable making the total account of points 7 and not 9. In the case of nuclear pitch accents, it measures 3 points within the prestressed syllable, 6 in the stressed and 3 within the post-stressed. In NPA accents, the ending point (end of IP) is measured at 0.05 seconds before the boundary to avoid the last glottal pulses that usually lack F0. Therefore, we are using a modellization of the pitch track with a sample of 7 points for prenuclear accents and of 13 points for NPA. This choice of points is motivated by peak and valley alignments described for P-ToBI [28], [29].

The set of rules applied was designed by building upon prior knowledge derived from analogous transcribers and the acoustic analysis of the training corpus. For each of the sentence included in the training corpus, ad-hoc rules were created to generate transcriptions that matched the transcriptions given by human transcribers.

The rules that the system applies are divided into rules for prenuclear pitch accents, rules for nuclear pitch accents and rules for boundary tones. Duration is also taken into account for some pitch events, mainly for the calling contour.

Differences in pitch level (from L to H or vice-versa) are set as movements greater than 1.5 semitones (st) [30], [31]. With that, the system computes the pitch mean and the IP uses that value as a reference, for the computation of high and low levels for the first tone in the sentence. The rest of the contour differences are computed locally, using differences in st between two points in the contour (1).

$$stDifference = \frac{12}{\log_{10}(2)} \cdot \log_{10}\left(\frac{f0_1}{f0_2}\right) \qquad (1)$$

This way, the system establishes that, if there is a difference greater than 1.5 st between the prestressed and the stressed syllable, there is a rise, and thus assigns a H* label.

Most of the rules rely on the intonation solely but some are only applied if the syllable has certain duration. For instance, the H+L* pattern is associated to a stressed syllable duration shorter than 0.20 seconds; when this condition is not met, the label is modified to H*+L.

Following these rules, the system can recreate the transcription done by humans for 81.2% of the cases in NPA and 83.3% in BT. The cases where the system fails are due to pitch tracking errors, final devoicing or contour transcriptions that cannot be estimated using only rules because there is a phonetical overlap between categories. For example, a rise in the stressed syllable can be H* but it can also correspond to L*+H when the peak is in the post-stressed; if a H* NPA has a slightly delayed peak, the system will consider that is a L*+H**.** In total, the transcriber uses more than 100 rules, many of them nested. The specific if-then conditions used can be checked at the transcriber code [32].

3. Results and assessment of the tool

The resulting tool is available as a Praat script at https://github.com/wendyelviragarcia/P_ToBI_transcriber. The output is the same TextGrid provided in the input with a new tier with the P-ToBI transcription and optionally a figure file.

In order to assess the P-ToBI prosodic transcriber and check if the rules could be generalized to new data we performed a quantitative and qualitative analysis of the resulting transcriptions. The quantitative evaluation of the transcriber was done by comparing the automatic transcriptions with those provided by human annotators. To do that we used the assessment corpus detailed in section 2.1.2 containing 141 sentences.

3.1. Evaluation of the tool

In this section, we present the evaluation results and inter-rater agreement (kappa, k). The evaluation metrics presented here are average values between the five categories of Boundary Tone (BT) and five Nuclear Pitch Accent (NPA) categories.

The NPA evaluation showed an accuracy of 73.8%, with a kappa value of 0.6 (Sensitivity = 0.67, Specificity= 0.93, F-score= 0.63), indicating moderate agreement. Similarly, the BT evaluation yielded 78.7% of accuracy, also with a kappa value of 0.6 (Sensitivity = 0.67, Specificity= 0.95, F-score= 0.58).

These results are similar to the ones achieved by [21] for the detection of prominence in Portuguese. But are lower than the performance for pitch contours classification in other languages, e.g. ToDI (PA 75.4%, BT 84.7%) [6], Sp_ToBI (NPA 88.11% and BT 81.28%) or Cat_ToBI that, depending on the transcriber, can get to a coincidence of 85.71% for NPA and 93.88% for BT [18].

3.2. Qualitative assessment

To further explore the performance of the P-ToBI transcriber, we examined the agreement between nuclear configurations. This analysis aimed to identify specific NPA categories that posed greater challenges for the system. It has been trained on a relatively small sample (93 sentences); nevertheless, the rules have proven to be effective in labeling Portuguese intonation.

Table 3: *Confusion matrix for NPA, correctly identified PA in italics.*

	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*

The analysis revealed that the transcriber encountered difficulties in accurately identifying falling NPAs (Table 3), particularly instances of "L*" and "H+L*" followed by a low BT.

Out of the 20 instances of H+L* L% included in the assessment corpus, 8 were correctly labelled and 10 were mistakenly labelled as L* L%. The other 2 were labelled as H* !H% and H*+L L%.

To gain further insights into the challenges observed in the P-ToBI transcriber's performance for falling tunes, we conducted a visual inspection of the pitch contours. The visual inspection suggested that the acoustic differences between "H+L* L%" and "L* L%" might be subtle and challenging to distinguish based solely on F0 (see Figure 2).

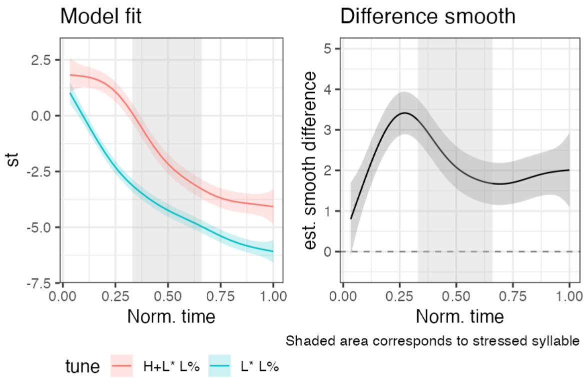

Figure 3: *GAMM modellization of H+L* L% and L* L% intonational contours (left) and its differences (right).*

Subsequently, we applied Generalized Additive Mixed Models (GAMM) to check the differences between the two pitch accents. We extracted 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. The results (Figure 3) show that the H+L* L% and L* L% exhibit noticeable differences, particularly in the prestressed area. The observed difference can be as significant as 3.5 st (st reference sentence mean).

4. Discussion and conclusions

The system presented makes a series of assumptions that have proven problematic.

The first one is that by selecting and computing differences between several points of the pitch contour we are treating the F0 curve as a series of straight lines. This is also a common practice in the depiction of pitch contours. Schematic ToBI figures usually depict a three-window table with a shadowed square for stressed syllables where the pitch contour is drawn as a series of straight lines from target to target. In contrast to this common practice, P-ToBI has many times depicted those contours as curves in its figures [13]. In addition, it has been proven that prosodic contours in Brazilian Portuguese are better modelled as a curve (as we have done with falling contours), so also other varieties could benefit from a curve representation.

The second is related to the number of F0 points considered in the prestressed, stressed and post-stressed regions. Following previous research [32] for Catalan and Spanish, this paper considered 6 points for the post-stressed area in NPA while it considered 3 for the pre-stressed area. However, as seen in Figure 3 the prestressed area is of utmost importance in P-ToBI and therefore, including more points within the pre-stressed area could help to classify the PA (as seen for the difference between L* L% and H+L* L%). The necessity of more pre-stressed points may be associated with the EP rise-fall patterns being off-ramp [33] in contrast with Catalan and Spanish systems. Since changes in the prestressed area are more subtle, the system needs more points to detect them.

Finally, the detection of F0 in Portuguese poses a significant challenge due to word-final vowel reduction and deletion, leading to instances of failures and missing data in the pitch track.

In conclusion, this paper has presented a P-ToBI prosodic transcriber available as a Praat script under a GNU license. Its evaluation revealed a success rate of 73.8% for NPA and 78.7% for BT, and a moderate agreement with human-made transcriptions. These results, although comparable to similar studies for other languages, fall short in comparison, for instance, to those achieved for Cat_ToBI. This gives information about Portuguese prosodic system highlighting the necessity of non-linear approaches to its representation and the importance of the prestressed region for its phonology. 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.

5. Acknowledgements

This research benefits from previous research by the project InAPoP (PTDC/CLE-LIN/119787/2010) funded by Fundação para a Ciência e a Tecnologia. The contributors to this work have received funding from project UIDB/00214/2020 funded by Fundação para a Ciência e a Tecnologia (Portugal) and project PID2022-139004OA-I00 funded by MCIN/AEI/10.13039/501100011033 and co-funded by ERDF/ESF "A way to make Europe".

6. References

[1] A. Rosenberg, "AutoBI - a tool for automatic ToBI annotation.," in INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, Mihama, Japan, 2010, pp. 146–149. Accessed: Aug. 26, 2014. [Online]. Available: http://eniac.cs.qc.cuny.edu/andrew/papers/autobi-is10.pdf

[2] A. W. Black and A. J. Hunt, "Generating F0 contours from ToBI labels using linear regression," in ICSLP 96. Fourth International Conference on Spoken Language Proceedings, Philadelphia: IEEE, 1996, pp. 1385–1388. doi: 10.1109/ICSLP.1996.607872.

[3] J. Lee, B. Kim, and G. Lee, "Automatic corpus-based tone and break-index prediction using K-ToBI representation," ACM

Transactions on Asian Language Information Processing (TALIP), vol. 1, no. 3, pp. 207–224, 2002, doi: 10.1145/772755.772757.
[4] B. Kim, J. Lee, and G. Lee, “Corpus-based Pitch Prediction based on K-TOBI Representation,” in ACM Transactions on Asian Language Information Processing (TALIP), ACM New York, NY, USA, 2002, pp. 207–224. doi: 10.1145/772755.772757.
[5] C. Wightman and M. Ostendorf, “Automatic labeling of prosodic patterns,” in IEEE Transactions on Audio, Speech, and Language Processing, 1994, pp. 469–481. Accessed: Nov. 17, 2014. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=326607
[6] N. Hu, B. Janssen, J. Hanssen, C. Gussenhoven, and A. Chen, “Automatic analysis of speech prosody in Dutch,” in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2020. doi: 10.21437/Interspeech.2020-2142.
[7] D. Hirst and R. Espesser, “Automatic Modelling of Fundamental Frequency Using a Quadratic Spline Function,” Travaux de l’Institut de Phonétique d’Aix-en-Provence, pp. 75–85, 1993.
[8] A. C. M. Rietveld, “Syllaben, klemtonen en de automatische detectie van beklemtoonde syllaben in het Nederlands,” PhD thesis, Université de Nijmegen, 1984.
[9] J. M. Garrido Almiñana, “A Tool for Automatic F0 Stylisation, Annotation and Modelling of Large Corpora,” in Speech Prosody, Chicago, 2010. Accessed: Jul. 03, 2014. [Online]. Available: http://speechprosody2010.illinois.edu/papers/100041.pdf
[10] J. M. Garrido Almiñana, “SegProso: A Praat-Based Tool for the Automatic Detection and Annotation of Prosodic Boundaries,” in Proceedings of TRASP (Interspeech satellite), B. Bigi and D. Hirst, Eds., Aix-en-Provence, 2013, pp. 74–77.
[11] P. Mertens, “The Prosogram : Semi-Automatic Transcription of Prosody based on a Tonal Perception Model,” in Proceedings of Speech Prosody 2004, 23-26 March, B. Bel and I. Marlien, Eds., Nara (Japan), 2004. Accessed: Jul. 13, 2014. [Online]. Available: http://bach.arts.kuleuven.be/pmertens/papers/sp2004.pdf
[12] C. Alessandro and P. Mertens, “Automatic pitch contour stylization using a model of tonal perception,” Computer Speech and Language 9(3, vol. 9, no. 3, pp. 257–288, 1995.
[13] S. Frota, P. Oliveira, M. Cruz, and M. Vigário, P-ToBI: tools for the transcription of Portuguese prosody. Lisboa: Laboratório de Fonética, CLUL/FLUL. [http://labfon.letras.ulisboa.pt/InAPoP/P-ToBI/], 2015.
[14] K. Silverman et al., “ToBI: A Standard for Labeling English Prosody,” in ICSLP 92 Proceedings 1992 International Conference on Spoken Language Processing. Volume 2, M. M. H. and G. E. W. J. J. Ohala, T. M. Nearey, B. L. Derwing, Ed., Department of Linguistics, University of Alberta., 1992, pp. 867-870.
[15] A. Rosenberg, E. Cooper, R. Levitan, and J. Hirschberg, “Cross-language prominence detection,” in Proceedings of the 6th International Conference on Speech Prosody, SP 2012, 2012.
[16] N. Zhi, D. Hirst, and P. M. Bertinetto, “Automatic analysis of the intonation of a tone language. Applying the Momel algorithm to spontaneous Standard Chinese (Beijing).,” Interspeech, 2010, Accessed: Sep. 02, 2014. [Online]. Available: http://20.210-193-52.unknown.qala.com.sg/archive/archive_papers/interspeech_2010/i10_1788.pdf
[17] N. Campbell, “Autolabelling Japanese ToBI,” in ICSLP 96. Fourth International Congress on Conference on Language Processing Proceedings, Philadelphia: IEEE, 1996, pp. 2399–2402. Accessed: Sep. 03, 2014. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=607292
[18] W. Elvira-García, P. Roseano, A. M. Fernández-Planas, and E. Martínez-Celdrán, “A tool for automatic transcription of intonation: Eti_ToBI a ToBI transcriber for Spanish and Catalan,” Lang Resour Eval, vol. 50, no. 4, pp. 767–792, Dec. 2016, doi: 10.1007/s10579-015-9320-9.
[19] T. Raso, B. Teixeira, and P. Barbosa, “Modelling automatic detection of prosodic boundaries for brazilian portuguese spontaneous speech,” Journal of Speech Sciences, vol. 9, 2020, doi: 10.20396/joss.v9i00.14957.
[20] A. R. M. Simões and A. R. Meireles, “Speech prosody in musical notation: Spanish, POrtuguese and English,” in Proceedings of the International Conference on Speech Prosody, 2016. doi: 10.21437/speechprosody.2016-44.
[21] H. Moniz, A. I. Mata, J. Hirschberg, F. Batista, A. Rosenberg, and I. Trancoso, “Extending AuToBI to prominence detection in European Portuguese,” in Proceedings of the International Conference on Speech Prosody, 2014. doi: 10.21437/speechprosody.2014-43.
[22] S. Frota, “The Intonational Phonology of European Portuguese,” in Prosodic Typology 2, S.-A. Jun, Ed., Oxford: Oxford University Press, 2014, pp. 6–42.
[23] S. Frota and M. Cruz, “Interactive Atlas of the Prosody of Portuguese,” http://labfon.letras.ulisboa.pt/InAPoP/.
[24] S. Frota, “(Coord.). (2012-2015) InAPoP - Interactive Atlas of the Prosody of Portuguese. Project funded by Fundação para a Ciência e a Tecnologia (PTDC/CLE-LIN/119787/2010).”
[25] J. C. Félix-Brasdefer, “Data collection methods in speech act performance,” 2010. doi: 10.1075/lllt.26.03fel.
[26] D. J. Hirst and C. de Looze, “Measuring Speech. Fundamental frequency and pitch.,” in Cambridge Handbook of Phonetics, vol. 1, 2021.
[27] C. De Looze, “Analyse et interprétation de l’empan temporel des variations prosodiques en français et en anglais,” Aix-en-Provence, 2010. Accessed: Jul. 19, 2014. [Online]. Available: http://halshs.archives-ouvertes.fr/tel-00470641/
[28] S. Frota, “Surface and Structure: Transcribing Intonation within and across Languages,” Lab Phonol, vol. 7, no. 1, p. 7, Jun. 2016, doi: 10.5334/labphon.10.
[29] S. Frota and P. Prieto, “Intonation in Romance,” in Intonation in Romance, 2015. doi: 10.1093/acprof:oso/9780199685332.003.0011.
[30] A. Pamies, A. M. Fernández Planas, E. Martínez Celdrán, A. Ortega-Escandell, and M. C. Amorós Cespedes, “Umbrales tonales en español peninsular,” in Actas del II Congreso de Fonética Experimental, 2002, pp. 272–278.
[31] A. C. M. Rietveld and C. Gussenhoven, “On the relation between pitch excursion size and prominence,” J Phon, vol. 13, no. 3, pp. 299–308, Jul. 1985, doi: 10.1016/S0095-4470(19)30761-2.
[32] L. Miranda, J. Moraes, and A. Rilliard, “Statistical modeling of prosodic contours of four speech acts in Brazilian Portuguese,” in Speech Prosody 2020, ISCA: ISCA, May 2020, pp. 404–408. doi: 10.21437/SpeechProsody.2020-83.
[33] S. Frota, “Surface and Structure: Transcribing Intonation within and across Languages,” Lab Phonol, vol. 7, no. 1, p. 7, Jun. 2016, doi: 10.5334/labphon.10.