

VARIATION IN TONE AND GESTURE WITHIN LANGUAGE

Marisa Cruz¹, Marc Swerts², Sónia Frota¹

¹University of Lisbon, ²Tilburg University
marisasousacruz@gmail.com, m.g.j.swerts@uvt.nl, sonia.frota@mail.telepac.pt

ABSTRACT

The present research focuses on the relation between tone and gesture across varieties of the same language, European Portuguese (EP). Three questions are addressed: (i) whether EP varieties use different visual cues while producing different sentence types/pragmatic meanings, (ii) if there is a relation between intonational variation and variability (if any) of visual cues, and (iii) if each linguistic factor involved can predict the type of visual cues used.

Two sentence types (statements/yes-no questions) and pragmatic meanings (broad/narrow focus) were examined in four varieties of EP. Results show that visual cues, like intonation, may vary across varieties and sentence types/pragmatic meanings. Furthermore, sentence type and pragmatic meaning are good predictors of how visual cues are time-aligned with intonation, in contrast with language variety. Consequently, we hypothesize that visual cues might play an important role in discriminating sentence types/pragmatic meanings, especially in the absence of tonal contrasts.

Keywords: visual prosody, intonation, language varieties, sentence type, pragmatic meaning.

1. INTRODUCTION

Languages are known to differ in several aspects of their intonational systems [18, 19, 20, 25], and language varieties have been shown to exhibit similar variation in intonation [1, 14, 29]. However, to our knowledge, the role of visual prosody in the distinction between varieties of a given language has not been investigated, and little is known about the association between pitch accent/boundary tone types and gesture types.

Previous studies on facial gestures as a complement of spoken language revealed that visual signals are organized into a system sharing several features with the prosody of spoken language [24, 26, 27, 28]. Thus, it is reasonable to assume that visual cues also vary across languages and language varieties. Moreover, variation in gesture could mirror intonational variation, or compensate lack of tonal contrast. However, the variability of gestures

(if any) and its relation with intonational variation remains largely understudied.

There has been extensive research on variation in EP spoken prosody, with studies on prosodic structure [9, 11, 32], phrasing patterns [4, 8], intonational contours [3, 10, 13], pitch accent distribution [4, 16, 33], and rhythmic patterns [3, 5, 15, 17]. Visual prosody, however, is a research field still to be explored in this language.

The present paper taps into the tone/gesture relation in EP by addressing three main questions. First, we aim to observe whether EP varieties use different facial gestures to convey specific sentence types and pragmatic meanings. Then, we explore the relation (if any) between variation in the visual domain and in intonation. Finally, we try to find out what are the best predictors for facial gestures: sentence type, pragmatic meaning, intonation, or language variety.

2. METHODOLOGY

The audiovisual database of prosodic variation in Portuguese from the *Interactive Atlas of the Prosody of Portuguese* (<http://labfon.letras.ulisboa.pt/InAPoP/>) was used [12]. Semi-spontaneous data was used, elicited by means of a Discourse Completion Task (DCT) [21] in four EP varieties already described for intonation [2, 3, 9, 11, 13]: the standard variety (SEP), two central-southern varieties (Ale and Alg), and one insular variety (Azores - PtD). The DCT was performed twice by three speakers per variety, aged between 20 and 45 years old.

For the analysis, we selected two sentence types (statements and yes-no questions) and two pragmatic meanings (broad and narrow focused statements). A total of 197 utterances was considered for the analysis of nuclear contours, using the P_ToBI system [11], and for the inspection of potential visual parallels for the following intonational features: (i) pitch accent (type), (ii) boundary tone (type), and (iii) configuration of the nuclear contour. For the analysis of visual cues, we considered three visual elements: head, eyebrows and eyes. Based on the Facial Action Coding System (FACS) [6], we annotated the following movements for each visual element: (i) up-down, down-up, head nod up-down, head nod left-right, as head movements, (ii) raising and lowering, as eyebrow movements, and (iii) non-

physiological eye blinks and eyes closed, for the eye movements. Data were annotated in ELAN 4.6.2 [7], where three tiers were created (Gestures, Tones, Sentences), in order to synchronize visual cues with the waveform and the previous intonational analysis independently carried out.

3. RESULTS

For the four EP varieties analysed, first we present the most frequent intonational contour and its time-aligned most frequent visual cue, across and within speakers. This allows us to observe whether EP varieties use different facial gestures to convey the same sentence type (section 4.1) and pragmatic meaning (section 4.2). Then, we explore the relation (if any) between variation in the visual domain and in intonation, trying to find out which are the best predictors for facial gestures (section 4.3).

3.1. Visual cues per sentence type

The analysis per sentence type shows that neutral statements are predominantly produced with the same basic visual cue (head up-down) across varieties, which is time-aligned with the same basic contour type: a falling pitch movement (H+L* L%) in SEP and Alg and a falling or low pitch movement ((H+)L* L%) in Ale and PtD (Table 1).

Table 1: Visual cues aligned with pitch accent/boundary tone types in broad focused statements across EP varieties (SEP, Ale, Alg, and PtD). Dominant tone and gesture patterns across and within speakers are represented.

EP	Broad focused statements	
	Tonal	Visual
SEP	H+L*	head up-down
	L%	neutral position
Ale	(H+)L*	head up-down
	L%	neutral position
Alg	H+L*	head up-down
	L%	neutral position
PtD	(H+)L*	head up-down
	L%	neutral position

Thus, in all varieties considered, the nuclear contour of broad focused statements is accompanied by the parallel up-down visual movement. Interestingly, nuclear pitch accent (NPA) types and boundary tone (BT) types seem to be related with specific visual cues: H+L*/L* are associated with the up-down head movement, and the low boundary tone (L%), that signals the end of the verbal production, is visually aligned with the return to the neutral position.

However, visual cues related to the nuclear contours of broad focused yes-no questions appear to be less tightly connected to intonational patterns.

If similarly to broad focused statements, H+L* (in SEP, Ale and PtD) and L* (also frequent in the last two varieties) are aligned with the same head movement (up-down), an additional visual cue is present (the eyebrow), and it exhibits the same movement across varieties (raising). Although the head movement follows pitch directionality (falling - H+L*/L*), the additional visual cue given by eyebrow movement goes in the opposite direction (raising), thus showing that the movement of facial gestures does not necessarily reinforce the intonational configuration. Like for French [30], eyebrow raising seems to be a question marker in EP.

Contrary to the observed relation between NPA and BT types and visual cues types in broad focused statements, in yes-no questions this parallelism is not present. For example, the return to the neutral position that is aligned to the low BT (L%) in broad focused statements across all varieties is aligned here with a complex rising BT (LH%) in broad focused yes-no questions in SEP. This also suggests that intonational complexity and visual complexity do not necessarily correlate.

Table 2: Visual cues aligned with pitch accent/boundary tone types in broad focused yes-no questions across EP varieties (SEP, Ale, Alg, and PtD). Dominant tone and gesture patterns across and within speakers are represented.

EP	Broad focused yes-no questions	
	Tonal	Visual
SEP	H+L*	head up-down +eyebrow raising
	LH%	neutral position
Ale	(H+)L*	head up-down +eyebrow raising
	H%	head back-forward +eyebrow raising
Alg	L*+H	eyebrow raising
	H%	head back-forward (+eyebrow raising)
PtD	(H+)L*	head up-down +eyebrow raising
	L%	neutral position

The absence of correlation between intonational and visual cues is further illustrated by PtD. In this variety, statements and yes-no questions are produced with the same falling nuclear configuration ((H+)L* L%), but visual cues differ. Besides head up-down followed by the return to the nuclear position, in yes-no questions the additional eyebrow raising conveys interrogativity. This suggests that visual cues can be crucial, even within a given variety, to distinguish between sentence types. However, perception experiments need to be conducted in order to confirm this hypothesis.

3.2. Visual cues per pragmatic meaning

Narrow focused statements are conveyed by the same nuclear contour (H*+L L%) across varieties (Table 3), and the same basic visual cue is also present: head up-down movement followed by the return to neutral position. Thus, as in broad focus (Table 1), in narrow focused statements the nuclear configuration (H+L) is accompanied by the same basic visual cue type (head movement) and directionality (falling). However, in contrast with broad focus statements, but similarly to yes-no questions (Table 2), an additional visual cue is associated with narrow focus, as in the case of Dutch [22, 23, 31]. The most frequent additional cue across EP varieties is the eyebrow movement, but in Alg closing the eyes is the mostly used visual strategy to convey focus. Although similar in type (except for Alg), the visual cue conveying focus in statements is not necessarily the same in form as in yes-no questions: in SEP and PtD, the same eyebrow raising is observed, but in Ale narrow focus is conveyed by eyebrow lowering.

Table 3: Visual cues aligned with pitch accent/boundary tone types in narrow focused statements across EP varieties (SEP, Ale, Alg, and PtD). Dominant tone and gesture patterns across and within speakers are represented.

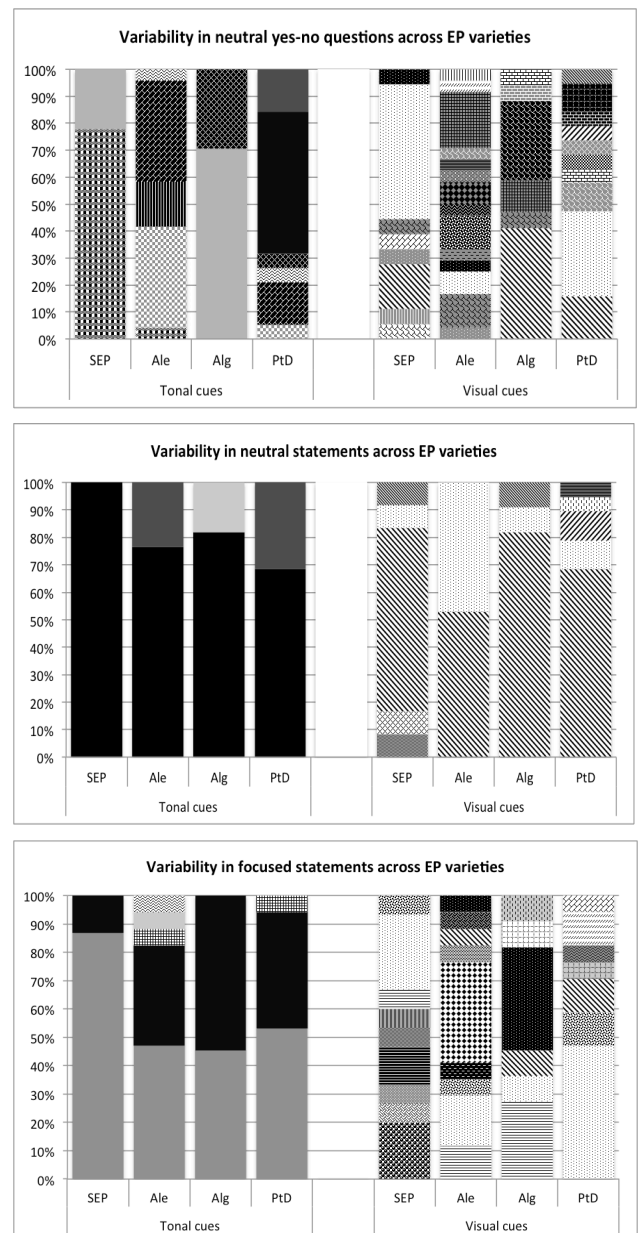
EP	Narrow focused statements	
	Tonal	Visual
SEP	H*+L	head up-down +eyebrow raising
	L%	neutral position
Ale	H*+L	head up-down +eyebrow lowering
	L%	neutral position
Alg	H*+L	head up-down +eyes closed
	L%	neutral position
PtD	H*+L	head up-down +eyebrow raising
	L%	neutral position

Visual cues seem to play an important role to distinguish between sentence types within a given variety, as it was observed for PtD, where statements and yes-no questions are produced with the same nuclear contour, but with different facial gestures. However, the analysis of visual cues in SEP reveals that intonation is also crucial to distinguish between sentence types or pragmatic meanings. Indeed, in SEP the same visual cues (head up-down and eyebrow raising, followed by neutral position) are used in broad focused yes-no questions (Table 2) and narrow focused statements (Table 3).

3.3. Visual and intonational variation across EP varieties

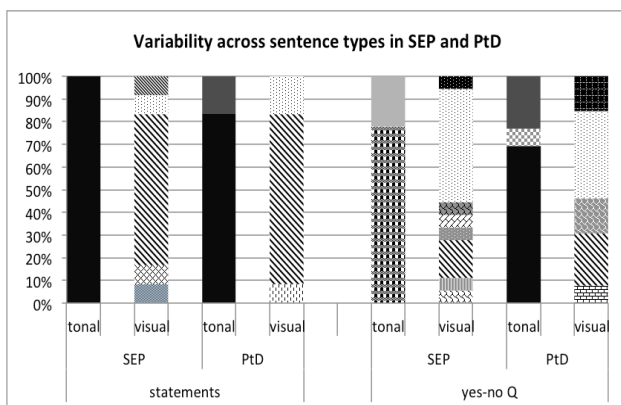
To further examine visual and intonational variation across EP varieties, all the possible intonational contours and visual cues per sentence type and pragmatic meaning in the four varieties were considered (Figure 1). Overall, yes-no questions (top) present greater variability (within and across varieties) of both tonal and visual cues than statements (mid). Narrow focused statements (bottom) exhibit more variability than broad focus statements, especially in the visual domain, but less variability than yes-no questions.

Figure 1: Variability of tonal and visual cues in broad focus statements (top), yes-no questions (mid), and narrow focused statements (bottom) across EP varieties. Within and across speaker variation is represented.



Moreover, not all varieties display similar patterns of variability. Zooming in on the main tonal and visual patterns across sentence types (Figure 2) shows that higher visual variability in SEP relates to the presence of a tonal contrast between statements and yes-no questions, and lower visual variability in PtD relates to the absence of tonal contrast between these sentence types. These results suggest that sentence type, pragmatic meaning, intonation, and language variety, all have an impact on facial gestures.

Figure 2: Variability of tonal and visual cues across sentence types in SEP and PtD (zooming-in main tonal patterns).



To determine which factors (sentence type, pragmatic meaning, intonation, or language variety) contribute the most to the variability found in gestures, two multinomial logistic regressions following a step-wise model were run: (i) facial gestures time-aligned with nuclear pitch accent (NPA), and (ii) facial gestures time-aligned with boundary tone (BT). Results show that language variety together with sentence type have a significant effect on facial gestures timed with NPA type ($\chi^2(30)=49.16, p=.015$) and with BT type ($\chi^2(18)=36.27, p=.007$). Sentence type together with pragmatic meaning were also good predictors of gestures (NPA: $\chi^2(5)=43.07, p=.000$; BT: $\chi^2(3)=11.79, p=.008$). However, language variety alone is not a good predictor (NPA: $\chi^2(15)=18.68, p=.229$; BT: $\chi^2(9)=10.05, p=.346$), unlike sentence type (NPA: $\chi^2(5)=28.24, p=.000$; BT: $\chi^2(3)=10.51, p=.015$) or pragmatic meaning (NPA: $\chi^2(5)=13.66, p=.018$; BT: $\chi^2(3)=12.30, p=.006$). Overall, only two of six possible interactions are good predictors of facial gestures: sentence type*pragmatic meaning and language variety*sentence type.

4. DISCUSSION AND CONCLUSION

We conclude that visual cues, similarly to intonational cues, may vary across varieties of the same language, and across sentence type and

pragmatic meaning. Although a relation between pitch accent types and gesture types was observed, sentence type or pragmatic meaning constrain this relation: visual cues time-aligned with the (H+)L* NPA differ in neutral statements and neutral yes-no questions (as in SEP, or in PtD), and the same facial gesture within a variety (e.g., head up-down and eyebrow raising in SEP) may be associated with different NPAs, conveying different sentence types/pragmatic meanings (e.g., narrow focused statements vs. broad focused yes-no questions in SEP). In short, our findings show a complex picture where intonation, sentence type, pragmatic meaning and language variety are all relevant factors. A logistic regression revealed that the interactions between (i) sentence type and pragmatic meaning and between (ii) sentence type and language variety were good predictors of visual cues time-aligned with NPA and BT types. Furthermore, language variety alone was shown not to be a good predictor, in contrast with sentence type and pragmatic meaning.

Overall, these results suggest that facial gestures may, like intonation, display some degree of grammaticalization across language varieties. However, audiovisual perception experiments are needed in order to further examine the role of facial gestures in the expression of intonational variation in EP. Since in production facial gestures are affected by sentence type, pragmatic meaning, and language variety, we hypothesize that speakers across varieties will be sensitive to visual information, especially in the absence of tonal contrast (e.g. SEP perceiving PtD questions, see also Figure 2), and in the presence of audiovisual mismatches. These predictions will be addressed in the near future.

5. ACKNOWLEDGMENTS

This research was supported by the Post-Doctoral grant BPD/94695/2013 to the first author, and the project grant *InAPoP – Interactive Atlas of the Prosody of Portuguese* (PTDC/CLE-LIN/119787/2010), both funded by Fundação para a Ciência e a Tecnologia, Portugal. We thank all the participants involved, as well as the research assistants Pedro Oliveira and Nuno Paulino for their help in data collection.

6. REFERENCES

- [1] Bruce, G. 2005. Intonational prominence in varieties of Swedish revisited. In: Jun, S.-A. (ed.), *Prosodic Typology I. The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 441-429.
- [2] Crespo-Sendra, V., Cruz, M., Silva, J., Frota, S. 2014. Asking questions across Portuguese varieties. *Talk*

- given at the *6th International Conference on Tone and Intonation in Europe (TIE)*, September 10-12, University of Utrecht, Netherlands.
- [3] Cruz, M. 2013. *Prosodic variation in European Portuguese: phrasing, intonation and rhythm in central-southern varieties*. Unpublished PhD Dissertation, University of Lisbon.
- [4] Cruz, M., Frota, S. 2013. On the relation between intonational phrasing and pitch accent distribution. Evidence from European Portuguese varieties. *Proceedings of the 14th Annual Conference of the International Speech Communication Association (ISCA) - Interspeech 2013*, 300-304. [ISSN 2308-457X].
- [5] Cruz, M., Frota, S. 2014. Rhythm in central-southern varieties of European Portuguese: production and perception. In: Moreno, A., Silva, F., Falé, I., Pereira, I., Veloso, J. (org.), *Textos Seleccionados do XXIX Encontro Nacional da Associação Portuguesa de Linguística*. Porto, APL, 213-231.
- [6] Ekman, P., Friesen, W. V., Hager, J. C. 2002. *Facial Action Coding System*. Salt Lake City, UT: A Human Face.
- [7] *ELAN – ELAN Linguistic Annotator*. Version 4.6.2. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. [<http://tla.mpi.nl/tools/tla-tools/elan/>].
- [8] Elordieta, G., Frota, S., Vigário, M. 2005. Subjects, objects and intonational phrasing in Spanish and Portuguese. *Studia Linguistica*, vol. 59 (2/3), 110-143.
- [9] Frota, S. 2000. *Prosody and focus in European Portuguese. Phonological phrasing and intonation*. New York: Garland Publishing.
- [10] Frota, S. 2002. Nuclear falls and rises in European Portuguese: a phonological analysis of declarative and question intonation. *Probus* 14, 113-146.
- [11] Frota, S. 2014. The intonational phonology of European Portuguese. In: Jun, S.-A. (ed.), *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 6-42.
- [12] Frota, S., Cruz, M. (coords). 2012-2015. *Interactive Atlas of the Prosody of Portuguese Webplatform*. [<http://labfon.letras.ulisboa.pt/InAPoP/>].
- [13] Frota, S., Cruz, M., Fernandes-Svartman, F., Collischonn, G., Fonseca, A., Serra, C., Oliveira, P., Vigário, M. 2015. Intonational variation in Portuguese: European and Brazilian Varieties. In: Frota, S., Prieto, P. (eds.), *Intonation in Romance*. Oxford: Oxford University Press, 235-283.
- [14] Frota, S., Prieto, P. (eds.). 2015. *Intonation in Romance*. Oxford: Oxford University Press.
- [15] Frota, S., Vigário, M. 2001. On the correlates of rhythmic distinctions: the European/Brazilian Portuguese case. *Probus* 13, 247-273.
- [16] Frota, S., Vigário, M. 2007. Intonational phrasing in two varieties of European Portuguese. In: Riad, T., Gussenhoven, C. (eds.), *Tones and Tunes*, Vol. 1. Berlin: Mouton de Gruyter, 263-289.
- [17] Frota, S., Vigário, M., Martins, F. 2002. Language Discrimination and Rhythm Classes: Evidence from Portuguese. *Speech Prosody Proceedings*. Aix-en-Provence, 315-318.
- [18] Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- [19] Jun, S.-A. (ed.). 2005. *Prosodic Typology I: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.
- [20] Jun, S.-A. (ed.). 2014. *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.
- [21] Kasper, G., Dahl, M. 1991. Research Methods in Interlanguage Pragmatics. *Studies in Second Language Acquisition* 13, 215-247.
- [22] Krahmer, E., Ruttkay, Z., Swerts, M., Wesselink, W. 2002. Pitch, eyebrows and the perception of focus. In: Bel, B. (ed.), *Proceedings of the Speech Prosody 2002*, 443-446.
- [23] Krahmer, E., Swerts, M. 2007. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language* 57 (3), 396-414.
- [24] Krahmer, E., Swerts, M. (eds.). 2009. Audiovisual prosody (special issue). *Language and Speech*, 52(2-3), 129-386.
- [25] Ladd, D. R. 2008. *Intonational Phonology*. 2nd ed. Cambridge: Cambridge University Press.
- [26] Loehr, D. P. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology. Special Issue Gesture as Language, Gesture and Language* 2, 3(1), 71-89.
- [27] Mol, L., Krahmer, E., Maes, A., Swerts, M. 2011. Seeing and Being Seen: The effects on gesture production. *Journal of Computer-Mediated Communication* 17(1), 77-100.
- [28] Mol, L., Krahmer, E., Maes, A., Swerts, M. 2012. Adaptation in gesture: Converging hands or converging minds? *Journal of Memory and Language* 66, 249-264.
- [29] Prieto, P., Roseano, P. (eds.). 2010. *Transcription of Intonation of the Spanish Language*. Lincom Europa: München.
- [30] Purson, A., Santi, S., Bertrand, R., Guaitella, I., Boyer, J., Cavé, C. 1999. The relationships between voice and gesture: Eyebrow movements and questioning. *Proceedings of European Conference of Speech Communication and Technology* 6, Budapest, Hungary, 1735-1739.
- [31] Swerts, M., Krahmer, E. 2008. Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics* 36 (2), 219-238.
- [32] Vigário, M. 2003. *The Prosodic Word in European Portuguese*. Berlin/New York: Mouton de Gruyter.
- [33] Vigário, M., Frota, S. 2003. The intonation of Standard and Northern European Portuguese: a comparative intonational phonology approach. *Journal of Portuguese Linguistics* 2-2 (Special issue on Portuguese Phonology edited by Wetzels), 115-137.