GESTURES-PROSODY (A)SYNCHRONY: TEMPORAL INTERACTION BETWEEN HEAD AND EYEBROW PEAKS IN STATEMENTS AND YES-NO QUESTIONS

Marisa Cruz & Sónia Frota

Center of Linguistics, School of Arts and Humanities, University of Lisbon
marisac@campus.ul.pt, sfrota@edu.ulisboa.pt

ABSTRACT

Gestures have been shown to be temporally linked to prosody, as strokes are synchronous to prominent syllables and pitch accents. However, their precise temporal coordination is still under debate. Exploring kinematics using a 2D motion analysis software, we inspected the temporal interaction between head, eyebrows and prosody in statements and yes-no questions in European Portuguese. We found that head peaks mainly align with the nuclear prosodic word, with different anchoring sites depending on sentence type: before the stressed syllable in statements; within the stressed syllable in yes-no questions. Eyebrow movement, relevant in conveying interrogativity, mainly precedes head peaks, and occurs before the nuclear word. In line with Loehr's suggestion [1], we conclude that *synchrony* cannot be interpreted as strict co-occurrence (not even between concurrent gestures). Instead, eyebrow and head peaks are motorically coordinated with intonation, and with each other, on a regular (but not concurrent) timing distribution.

Keywords: kinematics, temporal coordination, eyebrow peaks, head peaks, intonation.

1. INTRODUCTION

The term synchrony is generally used to refer to the temporal interaction between gestures and speech, from a phonological, semantic, and pragmatic point of view (see McNeill's rules in [2], providing evidence for a cognitive interdependence between the two modalities). However, several studies suggest that the time coordination between gestures and speech does not necessarily mean a strict synchrony or simultaneity in production [1, 3, 4]. Additionally, it has been shown that misaligned gestures up to 600ms before or after their original anchoring timepoint in the speech chain are still semantically integrated from a perception viewpoint [5]. It has also been suggested that the temporal coordination between gestures and speech is constrained by gesture type [6] and function [7]. For instance, eyebrow movements were shown to align with prominent syllables [8, 9, 10, 11, 12, 13], thus leading to a multimodal increase in prosodic prominence. Similar results were found from a developmental perspective, as deictic (pointing) gestures were shown to be timely coordinated with prominent syllables and pitch accents [14]. However, the way the two modalities relate to each other, i.e., in parallel *(hand-in-hand hypothesis*) or complementarily (*trade-off relation*) [15, 16 *versus* 17, 18, *inter alia*], and how they temporally interact is still under discussion.

Our main goal is to contribute to this debate, by exploring the kinematics of head and eyebrow movements along the production of statements and yes-no questions in European Portuguese (EP). Specifically, we analysed the temporal alignment of head and eyebrow peaks relative to the nuclear prosodic word (NPW) in both sentence types. When peaks occurred within the NPW time-window, we further observed their exact location relative to the stressed syllable. In addition, peak amplitude was also inspected in order to establish whether it differed across sentence types and depending on the time-alignment to the NPW.

2. METHODOLOGY

2.1. Materials

We used the audiovisual data of three female native speakers of Standard EP, aged between 20-45 years-old, videotaped in a sound attenuated laboratory with a JVC video camera (model GY-HM11E), while performing a Discourse Completion Task [19, 20], adapted for EP within the project *InAPoP* [21]. For the analysis, 11 neutral statements and 34 neutral yes-no questions were selected, which respectively exhibit the falling H+L* L% and falling-rising H+L* LH% nuclear contours [e.g., 22, 23]), as well as different visual cues: head falling movement in statements *versus* head falling and eyebrow raising movements in yes-no questions [e.g., 24]). Additionally, we also inspected a small set of less frequent yes-no questions produced by the same speakers: 10 yes-no questions with the same nuclear contour but involving head movement only (without eyebrow raising), and 7 yes-no questions exhibiting a different melodic pattern (H*+L L%, hereafter the *other contour*), but involving the two visual cues [25].

This small set allowed us to inspect whether the alignment of head and eyebrow peaks varied depending on the amount of visual cues in the signal and/or on the type of intonational cues for the same sentence type and pragmatic meaning. Thus, in total, the kinematics of 62 utterances was analysed, following the procedures detailed below.

2.2. Kinematic analysis

The kinematic analysis was previously performed in [25], using *Kinovea*, a free 2D motion analysis software [26]. For each video a marker was added to one of the eyebrows and to the chin of the speaker to track the vertical displacement of the two visual cues in pixels (px), along the absolute time series (ms). These data were extracted into .csv files. After the inspection of the kinematic curves, 5 sentences (2 yes-no questions with eyebrow, 1 yes-no question without eyebrow, and 2 yes-no questions *other contour*) were excluded from the analysis because, unlike all the others, they exhibited a rising head curve instead of a falling one.

We then identified in the time axis of each sentence the start and end points of the NPW, as well as the start and end points of the respective stressed syllable (Fig. 1).

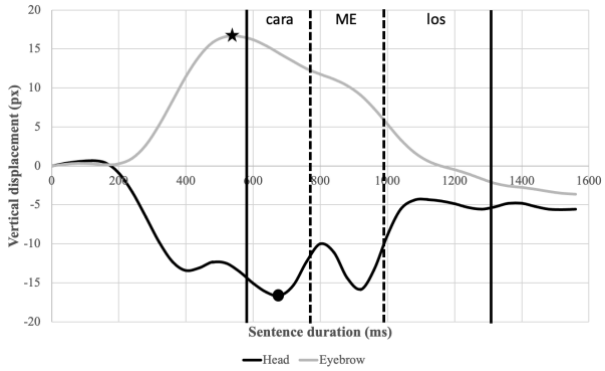

Figure 1: Vertical displacement (px) of the eyebrows (grey line) and head (black line) in the yes-no question *Querem caramelos?* (‘Do you want candies?’). The NPW is marked by vertical solid lines; the stressed syllable is marked by dashed lines. The star marks the eyebrow peak, and the dot marks the head peak.

The amplitude peaks of both visual cues were identified and manually coded as occurring before, within or after the NPW. When within the NPW, their occurrence relative to the stressed syllable was coded. For the eyebrow peaks, we also annotated their placement considering the head peak – before, aligned with, and after. Results on the head and eyebrow peaks temporal alignment are described in section 3.1.

Since the vertical displacement (px) of eyebrows may be biased by the head falling movement, namely in utterances without the eyebrow raising cue, we visually inspected the kinematics of each production and applied a correction in the cases in which the eyebrow curve mirrors the head curve – the eyebrow peak amplitude minus the head peak amplitude at the same timepoint – to set the corrected amplitude for this visual cue (Fig. 2). When the eyebrow curve reflects a rising curve, opposite to the head curve (Fig. 1), the eyebrow peak amplitude is kept unchanged.

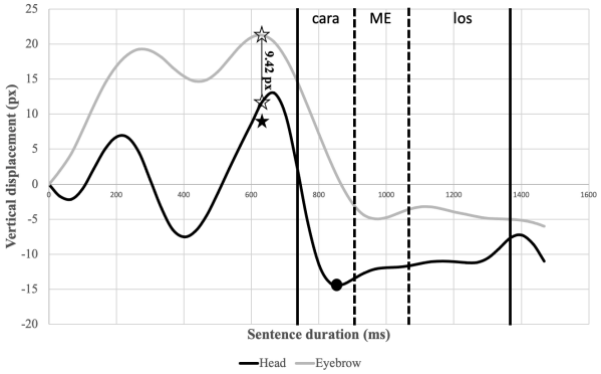

Figure 2: Vertical displacement (px) of the eyebrows (grey line) and head (black line) in the statement *Querem caramelos.* (‘(They) want candies.’). The NPW is marked by vertical solid lines; the stressed syllable is marked by dashed lines. The dot marks the head peak, and the stars the corrected eyebrow peak, reset from 21.24px to 9.42px.

Finally, we also examined whether sentence type and peak alignment relative to the NPW had any effect on peak amplitude (see section 3.2 for the results).

2.3. Statistical analysis

Two generalized linear mixed models (GLMM) – one per visual cue – were run, with the amplitude peak (px) as the dependent variable. Fixed effects were sentence type (statement, yes-no questions with eyebrow, yes-no questions without eyebrow, and yes-no questions *other contour*), peak alignment relative to the NPW (before, within, after), and the interaction peak alignment*sentence type. We applied the Satterthwaite method to control for unbalanced data, and a robust estimation was used to test fixed effects and coefficients in order to handle violations of model assumptions. Statistical procedures were performed in SPSS (IBM SPSS Statistics, version 26.0).

3. RESULTS AND DISCUSSION

3.1. Temporal alignment

3.1.1. Head movement

The head peak mainly occurs within the NPW, independently of sentence type (Fig. 3).

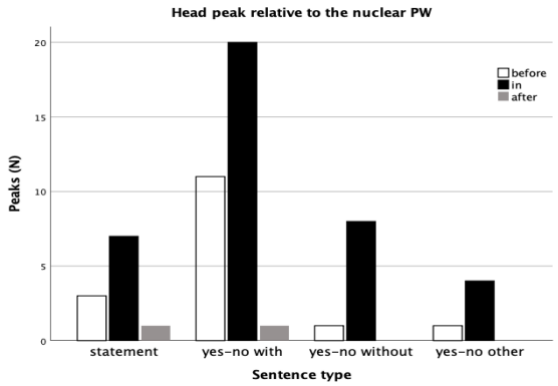

Figure 3: Temporal alignment of the head peak relative to the NPW.

However, the time-alignment of the head peak in relation to the stressed syllable of the NPW varied with sentence type. It precedes the stressed syllable in statements only; in yes-no questions, the peak mainly aligns with the stressed syllable (Fig. 4).

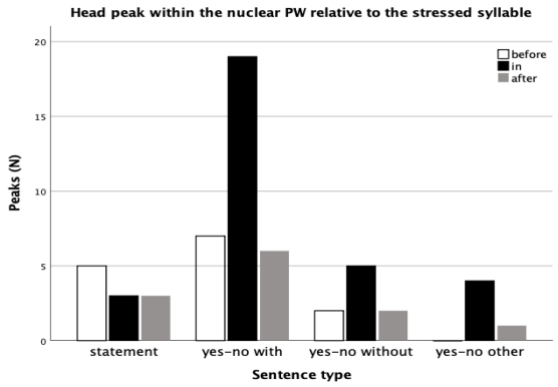

Figure 4: Temporal alignment of the head peak within the NPW relative to the stressed syllable.

3.1.2. Eyebrow movement

In contrast with the head, eyebrow peaks mainly occur before the NPW (Fig. 5).

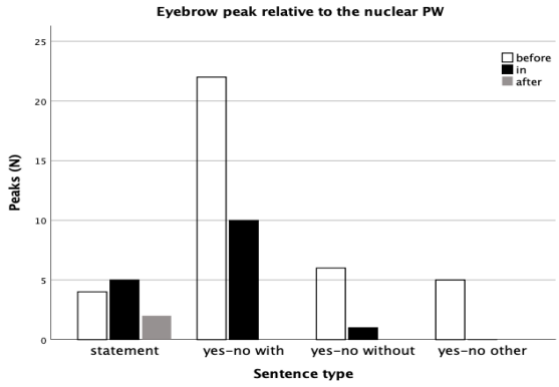

Figure 5: Temporal alignment of the eyebrow peak relative to the NPW.

In statements, eyebrow peaks predominantly occur within the NPW. However, the difference is residual (of one case only). Moreover, the alignment of eyebrow peaks relative to head peaks shows a pattern whereby the eyebrow peak mainly precedes the head peak, even in statements (Fig. 6).

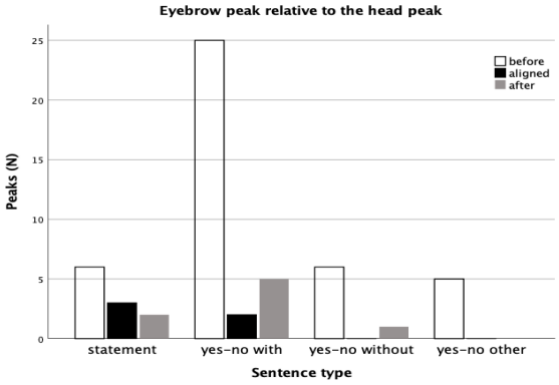

Figure 6: Temporal alignment of the eyebrow peak relative to the head peak.

We also observed the time lapse between the eyebrow and head peaks (excluding the very few cases in which the head peaks occur after the NPW). We concluded that this time-interval is not constant, not even within sentence type: in yes-no questions with eyebrow, it precedes the head peak in 451.5ms (av.), whereas in yes-no *other contour*, also produced with eyebrow raising, the peak is further away from the head peak (648ms, av.). In statements the eyebrow peak is closer to the head peak (222.8ms, av.).

3.2. Peak amplitude

3.2.1. Head peak

The GLMM analysis showed a significant effect of sentence type ($F(3, 47)=13.71$, $p<.001$), of the head peak alignment relative to the NPW ($F(2, 47)= 8.73$, $p<.01$), and of the interaction peak alignment*sentence type ($F(4, 47)=6.76$, $p<.001$).

Pairwise contrasts for the sentence type showed that the head peak amplitude significantly differs between statements and yes-no questions without eyebrow (M=6.03px, -27.89px, respectively; $ß$=33.92, SE=5.94, t=5.71, $p<.001$), thus pointing to the relevance of the head peak amplitude for the distinction between these sentence types. When the eyebrow cue is present, the head peak amplitude does not differ between statements and yes-no questions with eyebrow (M=1.39px) ($ß$=4.64, SE=6.46, t=.72, $p>.05$), but it significantly differs between statements and yes-no questions *other contour* (M=-22.42px) ($ß$=28.45, SE=6.07, t=4.69, $p<.001$). This suggests that the eyebrow peak might be the crucial visual cue for the first sentence type contrast, but not for the latter. Additionally, the head peak amplitude significantly differs between yes-no questions with eyebrow and yes-no questions *other contour* ($ß$=23.81, SE=3.11, t=7.66, $p<.001$), which means

that besides the auditory (intonational) cue, there is also a visual cue distinguishing between these productions. Finally, the head peak amplitude, together with the presence/absence of the eyebrow cue, also plays a relevant role to distinguish between yes-no questions with and without eyebrow movement (yes-no with-without: ß=29.28, SE=2.86, t=10.24, p<.001; yes-no without-*other contour*: ß=23.81, SE=3.11, t=7.66, p<.001).

Although the head peak mainly aligns with the NPW, the main effect of peak alignment is due to its amplitude being higher when the peak precedes the NPW (M_{before}=-17.35px; M_{in}=-10.86px). However, this amplitude difference was not significant (ß=-6.49, SE=4.93, t=-1.32, p>.05).

As for the interaction peak alignment*sentence type, pairwise contrasts for the most frequent head peak alignment position (within the NPW) showed that the head peak amplitude did not differ across yes-no questions (with-without: ß=6.68, SE=4.13, t=1.62, p>.05; with-*other contour*: ß=8.08, SE=4.79, t=1.69, p>.05; without-*other contour*: ß=1.40, SE=3.60, t=.39, p>.05), but differed between statements and yes-no questions without eyebrow (ß=25.12, SE=11.03, t=2.28, p<.05), and between statements and yes-no questions *other contour* (ß=26.52, SE=11.30, t=2.35, p<.05). The fact that the head peak amplitude within the NPW did not differ between statements and yes-no questions with eyebrow (ß=18.44, SE=11.48, t=1.61, p>.05) further supports the claim that the eyebrow peak might be enough for the visual distinction between these sentence types.

3.2.2. Eyebrow peak

The GLMM analysis showed a significant effect of sentence type (F(3, 47)=8.86, p<.001), and of the interaction peak alignment*sentence type (F(2, 47)=3.53, p<.05). The eyebrow peak alignment relative to the NPW, by itself, did not have a significant effect on the eyebrow peak amplitude (F(2, 47)=2.04, p>.05).

Pairwise contrasts for the sentence type showed that the eyebrow peak amplitude did not differ between statements and yes-no questions without eyebrow (M=8.17px, 6.64px, respectively; ß=1.54, SE=1.53, t=1.00, p>.05), thus strengthening the relevance of the head peak amplitude for the distinction between these sentence types, as suggested in section 3.2.1. Statements and yes-no questions *other contour* also did not differ (M=17.05px) (ß=-8.88, SE=6.08, t=-1.46, p>.05). However, the eyebrow peak amplitude is relevant to distinguish between statements and yes-no questions with eyebrow (M=14.47px) (ß=-6.30, SE=2.10, t=-3.01, p<.01). It thus seems that head and eyebrow peaks play a complementary role in marking a given sentence type as different from statements.

Additionally, the eyebrow peak amplitude did not differ between yes-no questions with eyebrow and yes-no questions *other contour* (ß=-2.58, SE=6.11, t=-.423, p>.05), which strengthens the relevance of the head peak amplitude for the distinction between these productions, as noted in section 3.2.1. Finally, although the eyebrow (together with the head) peak amplitude plays a relevant role to distinguish between yes-no questions with and without eyebrow (ß=7.84, SE=1.62, t=4.84, p<.001), the same does not apply to the contrast between yes-no questions without eyebrow and yes-no questions *other contour* (ß=-10.42, SE=5.94, t=-1.75, p>.05), for which the head peak amplitude was more relevant. However, the interaction peak alignment*sentence type showed that eyebrow peak amplitude in the most frequent eyebrow peak alignment position (before the NPW) significantly differed between yes-no questions without eyebrow and yes-no questions *other contour* (ß=-14.01, SE=6.01, t=-2.33, p<.05).

4. CONCLUSION

It was found that head and eyebrow peaks are motorically coordinated with intonation, and with each other, on a regular (but not concurrent) timing distribution: head peaks mainly co-occur with the stressed syllable of the NPW, regardless of sentence type; although exhibiting a varying time lapse, eyebrow peaks precede the head peaks and mainly occur before the NPW. This supports Loehr's suggestion [1] that *synchrony* cannot be interpreted as strict co-occurrence (not even between concurrent gestures). Beyond the alignment of these gesture peaks, peak amplitude was found to vary across sentence type for both head and eyebrow, and as a function of the peak position relative to the NPW for head peaks. Moreover, an interaction between peak alignment and sentence type was found for both visual cues. The results point to the complementary role between head and eyebrow, and between gestures and intonation. These results thus provide support to the theory of a *trade-off relation* between gesture and speech production, in which head, eyebrow peaks and intonation convey prominence and meaning in an additive way.

5. ACKNOWLEDGEMENTS

This research was supported by the UIDB/00214/2020 Project, funded by Fundação para a Ciência e a Tecnologia. We also thank all participants for their precious collaboration.

6. REFERENCES

[1] Loehr, D. P. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Lab. Phonol*. Special Issue Gesture as Language, Gesture and Language 2, 3(1), 71–89. doi: 10.1515/lp-2012-0006

[2] McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press.

[3] de Ruiter, J.P. 2000. The production of gesture and speech. In: McNeill, D. (eds), *Language and Gesture*. Cambridge University Press, 248–311.

[4] McNeill, D. 2005. *Gesture and Thought*. University of Chicago Press.

[5] Kirchhof, C., de Ruiter, J.P. 2012. "On the audiovisual integration of speech and gesture", presented at the Fifth Conference of the International Society for Gesture Studies, Lund, Sweden, July 24-27, 2012.

[6] Leonard, T., Cummins, F. 2010. The temporal relation between beat gestures and speech. *Lang. Cognitive Proc.* 26(10), 1457–1471. doi: 10.1080/01690965.2010.500218

[7] Wagner, P., Malisz, Z., Kopp, S. 2014. Gesture and speech in interaction: An overview. *Speech Commun.* 57, 209–232. doi: 10.1016/j.specom.2013.09.008

[8] Ekman, P., 1979. About brows: Emotional and conversational signals. In: von Cranach, M., Foppa, K., Lepenies, W., Ploog, D. (eds), *Human Ethology: Claims and Limits of a New Discipline*. Cambridge University Press, 169–202.

[9] Krahmer, E., Swerts, M. 2007. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *J. Mem. Lang.* 57(3), 396–414. doi: 10.1016/j.jml.2007.06.005

[10] Granström, B., House, D. 2007. Inside out – Acoustic and visual aspects of verbal and non-verbal communication. In: Trouvain, J, Barry, W. (eds) *Proc. of the XVIth International Congress of the Phonetic Sciences*, Saarbrücken, Germany, 11–18.

[11] Swerts, M., Krahmer, E. 2008. Facial expression and prosodic prominence: Effects of modality and facial area. *J. Phonetics*, 36(2), 219–238. doi: 10.1016/j.wocn.2007.05.001

[12] al Moubayed, S., Beskow, J., Granström, B., 2010. Auditory visual prominence: From intelligence to behavior. *J. Multimodal User In.* 3(4), 299–309. doi: 10.1007/s12193-010-0054-0

[13] Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., Vatikiotis-Bateson, E. 2004. Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychol. Sci.* 15(2), 133–137. doi: 10.1111/j.0963-7214.2004.01502010.x

[14] Esteve-Gibert, N., Prieto, P. 2014. Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Commun*. 57, 301–316. doi: 10.1016/j.specom.2013.06.006

[15] So, W.C., Kita, S., Goldin-Meadow, S., 2009. Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand. *Cognitive Sci.* 33, 115–125. doi:10.1111/j.1551-6709.2008.01006.x

[16]Fernández-Baena, A., Montaño, R., Antonijoan, M., Roversi, A., Miralles, D., Alías, F. 2014. Gesture synthesis adapted to speech emphasis. *Speech Commun*. 57, 331–350. doi: http://dx.doi.org/10.1016/j.specom.2013.06.005

[17] de Ruiter, J.P., Bangerter, A., Dings, P. 2012. Interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis. *Top. Cogn. Sci.* 4(2), 232–248. doi: 10.1111/j.1756-8765.2012.01183.x

[18] Ferré, G. 2014. A multimodal approach to markedness in spoken French. *Speech Commun*. 57, 268–282. doi: 10.1016/j.specom.2013.06.002

[19] Félix-Brasdefer, J. C. 2009. Data collection methods in speech act performance: DCTs, role plays, and verbal reports. In: Juán, E. U., Martinéz-Flor, A. (eds), *Speech act performance: Theoretical, Empirical, and Methodological Issues*. John Benjamins Publishing, 41–56.

[20] Billmyer, K., Varghese, M. 2000. Investigating instrument-based pragmatic variability: Effects of enhancing discourse completion tests. *Appl. Linguist.*, 21(4), 517–552. doi: https://doi.org/10.1093/applin/21.4.517

[21] Frota, S. (Coord.). 2012-2015. *Interactive Atlas of the Prosody of Portuguese Project* (PTDC/CLE-LIN/119787/2010), funded by FCT http://labfon.letras.ulisboa.pt/InAPoP/.

[22] Frota, S. 2002. Nuclear falls and rises in European Portuguese: A phonological analysis of declarative and question intonation. *Probus* 14(1) (Special Issue on Intonation in Romance, ed. by Hualde, J. I.), 113–146. doi: 10.1515/prbs.2002.001

[23] 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 University Press, 235–283.

[24] Cruz, M., Swerts, M., Frota, S. 2015. Variation in tone and gesture within language. In: The Scottish Consortium for ICPhS 2015 (ed), *Proc. of the 18th International Congress of Phonetic Sciences*, University of Glasgow. Paper 452 retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0452.pdf

[25] Cruz, M., Frota, S. Submitted. "Talking heads" in Portuguese. *Language and Cognition.*

[26] *Kinovea 0.9.4. https://www.kinovea.org/*