Alignment of head nods in French focus: an EMA study

C. Carignan[a], N. Esteve-Gibert[b], H. Lœvenbruck[c], M. Dohen[d], M. D'Imperio[e]

[a]Univ. College London, [b]Univ. Oberta de Catalunya, [c]Univ. Grenoble Alpes, [d]GIPSA-lab, [e]Rutgers Univ.

Rhythmic, co-verbal movement of the head always accompanies speech (Munhall et al., 1994). Previous work on rhythmic head gestures (head-nods and eyebrow movements) during speech specifically focused on timing and motor organization. These studies suggested that co-verbal head movements are linked to the production of prosodic features of speech such as stress, prominence, and intonation (Hadar et al., 1984; House et al., 2001; Esteve-Gibert et al., 2017a, *inter alia).* Recent cross-linguistic work in Japanese and English has on the other hand shown that eyebrow movement tend to occur in correspondence with phrase boundaries and not prominent syllables (de la Cruz-Pavía et al., 2019). The current study examines head movement correlates of contrastive and corrective focus in French interactive speech (e.g., 'Take the ORANGE dress [not the blue dress]'). Previous work showed that, in a similar task, French preschoolers produce contrastive focus only through head movement (but not through prosodic strategies), by marking contrastive and corrective focus words with more frequent head gestures than broad focus productions (Esteve-Gibert et al., 2017b).

Head nod gestures were captured using a Carstens AG500 electromagnetic articulometry (EMA) system, with sensors on the left and right mastoids and the nasion. Data were collected from 16 native French speakers. Speakers participated in a game that elicited spontaneous production of sentences in 3 conditions (no-focus; contrastive focus; corrective focus) and 2 target focus positions (on the noun; on the adjective). In order to quantify head nod movement, the vector extending from the inter-mastoid point (the midpoint of the line connecting the two mastoid sensors) to the nasion was calculated, and the unit vector x-y-z components were transformed to spherical coordinates. The resultant polar angle, ϕ, captures upward-downward angle of head movement within the spherical space defined by the inter-mastoid point as the origin and the nasion as the zenith (1, θ, ϕ). The ϕ signal (henceforth "head nod signal") was down-sampled to 50 Hz via spline fitting. Peaks in the filtered head nod signal were located within each phrase, with a 500 ms buffer at the end of the phrase to allow for the possibility of gesture peaks occurring after the offset of a phrase-final syllable. The most prominent peak (defined as the widest peak) was logged,[1] and the time point of the gesture apex was used to calculate the distance (in ms) between the head nod apex and the end of the phrase.

The results are shown in Figure 1, which displays box plots of the temporal distance between the apex of the head nod gesture and the end of the phrase. Gray bars behind the box plots display the average temporal interval of the accented syllable within the phrase: dark gray bars denote the interval of the accented syllable in noun-focused conditions, while light gray bars denote the interval of the accented syllable in adjective-focused conditions. In both focus conditions, the apex of the head nod gesture is aligned with the accented syllable in nouns, but before the accented syllable in adjectives (i.e. on the preceding noun in the noun-adjective sequence of French). Specifically, for nouns, the apex aligns with the onset of the accented syllable in contrastive focus ("fn") vs. the nucleus of the accented syllable in corrective focus ("cn") items; for adjectives, the apex moves from $\approx$ 250 ms before the onset of the accented syllable in contrastive focus ("fa") to just prior to the onset of the accented syllable for corrective focus ("ca"). A linear mixed effects (LME) model tested the effect of focus condition on apex timing, with random intercepts for both speaker and word; the results for pairwise contrasts in the LME model are shown in Table 1. The results reveal that all inter-condition differences suggested by the box plots in Figure 1 are significant, with no intra-condition differences between nouns and adjectives. Note though a visible trend differentiating syllable alignment in nouns and adjectives, which might be due to phrase edge proximity, in line with Esteve-Gibert et al. (2017a). In other words, prominent speech intervals attract head gesture intervals, though proximity to prosodic edges might determine precise gesture apex alignment. Although our findings do suggest that the alignment of head nod gestures in French is conditioned systematically by focus type (contrastive

[1]Both maxima (nod upward) and minima (nod downward) were located and considered prominent gesture candidates.

vs. corrective), these results are exploratory in nature. We are currently refining our analyses to further determine the kinematic properties of these nodding gestures, to investigate differences in comparison to gestures that occur under broad focus, to compare our kinematic measurements with expert annotation of ELAN video data, and to correlate our refined measures with F0 patterns.

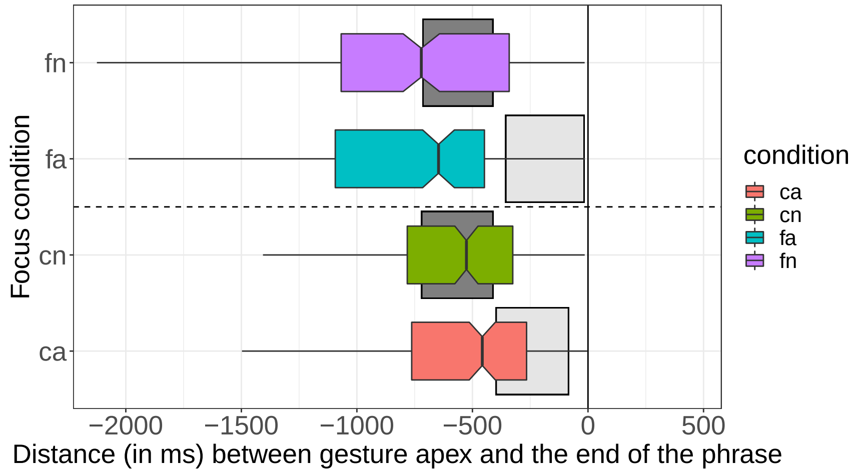

Figure 1: Box plots of the lag (ms) between the apex of the head nod gesture and the end of the phrase (x-axis = 0). Dark gray boxes denote the average temporal interval of the target nouns and light gray boxes denote the average temporal interval of the target adjectives. The horizontal dashed line separates the contrastive focus conditions (fn, fa) from the corrective focus conditions (cn, ca).

Table 1: Tukey contrasts in the LME model with phrase offset lag of the gesture apex as the DV. Cells highlighted in gray indicate significant differences, after p-value adjustment for multiple comparisons.

Linear hypothesis	β	*SE*	z-statistic	$\Pr(>\|z\|)$
cn − ca == 0	-65.35	61.47	-1.063	$= 0.701$
fa − ca == 0	-289.74	43.55	-6.653	< 0.001
fn − ca == 0	-243.35	60.75	-4.006	< 0.001
fa − cn == 0	-224.39	61.22	-3.665	< 0.01
fn − cn == 0	-178.01	41.74	-4.265	< 0.001
fn − fa == 0	46.38	60.53	0.766	$= 0.864$

References

de la Cruz-Pavía, I., Gervain, J., Vatikiotis-Bateson, E., and Werker, J. F. (2019). Coverbal speech gestures signal phrase boundaries: A production study of Japanese and English infant- and adult-directed speech. *Language Acquisition*, pages 1–27.

Esteve-Gibert, N., Borràs-Comes, J., Asor, E., Swerts, M., and Prieto, P. (2017a). The timing of head movements: the role of prosodic heads and edges. *Journal of the Acoustical Society of America*, 141(6):4727–4739.

Esteve-Gibert, N., Lœvenbruck, H., Dohen, M., and D'Imperio, M. (2017b). The use of prosody and gestures for the production of contrastive focus in French-speaking 4 and 5 year olds. Oral presentation at the IPS Workshop *'Abstraction, Diversity and Speech Dynamics'*, Munich (Germany), 3-5 May, 2017.

Hadar, U., Steiner, T. J., Grant, E. C., and Rose, F. C. (1984). The timing of shifts in head posture during conversation. *Human Movement Science*, 3:237–245.

House, D., Beskow, J., and Granström, B. (2001). Timing and interaction of visual cues for prominence in audiovisual speech perception. In *Proceedings of Eurospeech 2001*, pages 387–390, Aalborg, Denmark.

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., and Vatikiotis-Bateson, E. (1994). Visual prosody and speech intelligibility: head movement improves auditory speech perception. *Psychological Science*, 15(2):133–137.