Head movements highlight important information in speech: an EMA study with French speakers

Núria Esteve-Gibert[1,2], Thierry Legou[1], Hélène Loevenbruck[3], Marion Dohen[4], Mariapaola D'Imperio[1]

[1]*Aix Marseille Université, CNRS, LPL UMR 7309, 13100, Aix-en-Provence, France*
[2]*University of Barcelona, Barcelona, Spain*
[3]*University of Grenoble-Alpes, LPNC, CNRS UMR 5105, Grenoble, France*
[4]*Speech and Cognition Department, University of Grenoble-Alpes, Grenoble, France*

Speakers use different strategies in different languages to signal which element in the discourse is particularly informative for the interlocutor, a phenomenon called 'information focus'. One of the most frequent strategies to achieve this across languages is by prosodic means. In French, for instance, speakers are found to compress the pitch range of the elements surrounding the highlighted constituent to make it become more salient [1], [2], and to probabilistically use an intonation rise at the left-edge of the focused constituent [3], [4]. But human communication is multimodal, and a growing body of research is showing that visual cues (such as orofacial gestures, head movements, and manual gestures) are also used to express and identify focus [5]–[8].

The present study investigates how French speakers use visual cues (and specifically, head movements) to highlight new and contrastive information in semi-spontaneous situations, and how these cues interact with the audio signal. While previous studies have mostly used motion capture with controlled speech data or manual annotation of video-recording (which can be subjective and unprecise), in the present study we elicited semi-spontaneous utterances recorded with 3D Articulography to allow for precise measurements.

To do so, 19 French-speaking adults participated in a task in which they had to tell to a virtual character which object had to be taken out of a bag in order to trigger a playful activity. Spontaneous sentences like *Prends le bonnet violet* 'Take the purple hat' were elicited. We manipulated the number and type of objects inside the bag to elicit 3 focus-type conditions (broad-focus; contrastive focus; corrective focus) and 2 focus-position conditions (focused-noun; focused-adjective), full crossing the five levels. A microphone was used to record the audio signal, head movements were captured using eight EMA sensors attached to the participants' face (EMA AG501, sampling frequency of 250 Hz) (see Figure 1 for further details on the position of the sensors), and a video-camera was used for control.

Before the coding, two trained native labelers blindly screened all sentences for appropriate prosodic production. Only those sentences in which the coders' classification and the testing condition matched were further used for the analyses. We then used the EMA data to automatically analyze head nods (alpha angle from the inter-mastoid points to the top part of the nose), chin-forward movements (difference between the top part of the nose and the chin position along x), and eyebrow movements (difference between top part of the nose and the eyebrows' position). Word boundaries and F0 peaks were coded using PRAAT and then imported into a Matlab script. A parallel manual annotation of video-recordings was conducted using ELAN software to code the presence and timing of head movements and to compare it with the EMA data.

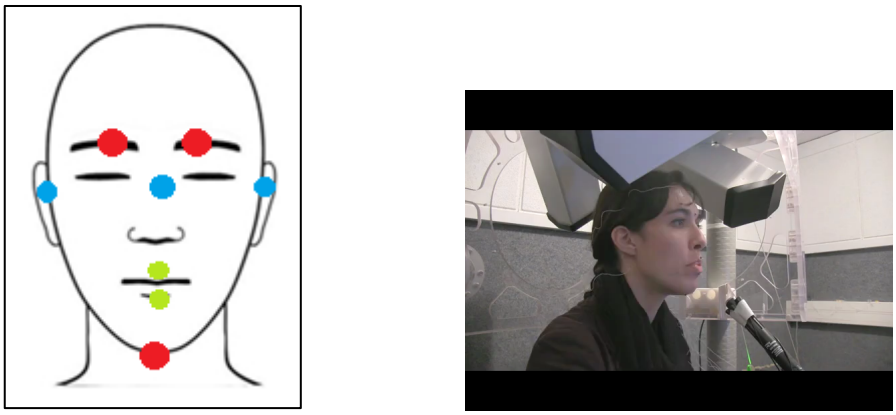

Figure 1. *Left panel, position of the EMA sensors on the participants' face. The red sensors were used to calculate eyebrow and head nod movements; the blue sensors were used as reference sensors; the green ones were not used for the current study. Right panel, image depicting the experimental scene.*

Table 1 summarizes our variables of interest for each target word. First, we expect higher Nod-Displ and Eyebrow-Displ values in the focus conditions (corrective focus > contrastive focus) than in the broad focus condition. Second, we expect Nod/Eyebrow-PeakDispl and Nod/Eyebrow-PeakVel to occur more often during focused words (noun or adjective) than during non-focused words. Third, we expect the F0Peak-PeakDispl-Dist values to be closer to 0 (more synchronization) during focused words than during non-focused words. Finally, we expect EMA results to refine the results of the manual annotation from video data, providing more accurate spatial and temporal information.

Variable	**Description**
Nod-Displ	Amount of displacement of the head nod
Nod-PeakDispl	Maximum displacement of the head nod
Nod-PeakVel	Peak velocity of the head nod movement
Eyebrow-Displ	Amount of displacement of the eyebrow
Eyebrow-PeakDispl	Maximum displacement of the eyebrow
Eyebrow-PeakVel	Peak velocity of the eyebrow movement
F0Peak-PeakDispl-Dist	Distance between F0 peak and the maximum displacement of the head nod or eyebrow

References

[1] M. Dohen and H. Loevenbruck, "Pre-focal Rephrasing , Focal Enhancement and Post-focal Deaccentuation in French," in *Proceedings of the 8th International Conference on Spoken Language Processing*, 2004, pp. 2–5.

[2] S. Jun and C. Fougeron, "A Phonological Model of French Intonation," In Antonis Botinis, ed., Intonation: Analysis, Modeling and Technology, 209–242. Kluwer Academic Publishers.

[3] J. S. German and M. D'Imperio, "The Status of the Initial Rise as a Marker of Focus in French," *Lang. Speech*, 2015.

[4] C. Astésano, E. G. Bard, and A. Turk, "Structural influences on initial accent placement in French.," *Lang. Speech*, vol. 50, no. 3, pp. 423–446, 2007.

[5] N. Esteve-Gibert, J. Borràs-Comes, E. Asor, M. Swerts, and P. Prieto, "The timing of head movements: The role of prosodic heads and edges," *J. Acoust. Soc. Am.*, vol. 141, no. 6, pp. 4727–4739, 2017.

[6] G. Ambrazaitis and D. House, "Multimodal prominences: Exploring the

patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings," *Speech Commun.*, vol. online ver, 2017.

[7] M. Dohen and H. Loevenbruck, "Interaction of audition and vision for the perception of prosodic contrastive focus.," *Lang. Speech*, vol. 52, pp. 177–206, 2009.

[8] J. Kim, E. Cvejic, and C. Davis, "Tracking eyebrows and head gestures associated with spoken prosody," *Speech Commun.*, vol. 57, pp. 317–330, Feb. 2014.