


SEPTEMBER 16 2024

Co-speech head nods are used to enhance prosodic prominence at different levels of narrow focus in French

Christopher Carignan ; Núria Esteve-Gibert; H el ene L evenbruck; Marion Dohen; Mariapaola D'Imperio



J. Acoust. Soc. Am. 156, 1720–1733 (2024)

<https://doi.org/10.1121/10.0028585>



LEARN MORE

Advance your science and career as a member of the
Acoustical Society of America

Co-speech head nods are used to enhance prosodic prominence at different levels of narrow focus in French

Christopher Carignan,^{1,a)}  Núria Esteve-Gibert,² H  l  ne L  venbruck,³ Marion Dohen,⁴ and Mariapaola D'Imperio⁵

¹University College London, London, United Kingdom

²Universitat Oberta de Catalunya, Barcelona, Spain

³Universit   Grenoble Alpes, Universit   Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France

⁴Universit   Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

⁵Aix-Marseille Universit   and Laboratoire Parole et Langage, CNRS, 13100 Aix-en-Provence, France

ABSTRACT:

Previous research has shown that prosodic structure can regulate the relationship between co-speech gestures and speech itself. Most co-speech studies have focused on manual gestures, but head movements have also been observed to accompany speech events by Munhall, Jones, Callan, Kuratate, and Vatikiotis-Bateson [(2004). *Psychol. Sci.* **15**(2), 133–137], and these co-verbal gestures may be linked to prosodic prominence, as shown by Esteve-Gibert, Borr  s-Comes, Asor, Swerts, and Prieto [(2017). *J. Acoust. Soc. Am.* **141**(6), 4727–4739], Hadar, Steiner, Grant, and Rose [(1984). *Hum. Mov. Sci.* **3**, 237–245], and House, Beskow, and Granstr  m [(2001). *Lang. Speech* **26**(2), 117–129]. This study examines how the timing and magnitude of head nods may be related to degrees of prosodic prominence connected to different focus conditions. Using electromagnetic articulometry, a time-varying signal of vertical head movement for 12 native French speakers was generated to examine the relationship between head nod gestures and F0 peaks. The results suggest that speakers use two different alignment strategies, which integrate both temporal and magnitudinal aspects of the gesture. Some evidence of inter-speaker preferences in the use of the two strategies was observed, although the inter-speaker variability is not categorical. Importantly, prosodic prominence itself is not the cause of the difference between the two strategies, but instead magnifies their inherent differences. In this way, the use of co-speech head nod gestures under French focus conditions can be considered as a method of prosodic enhancement.    2024 Acoustical Society of America. <https://doi.org/10.1121/10.0028585>

(Received 1 February 2024; revised 28 June 2024; accepted 21 August 2024; published online 16 September 2024)

[Editor: Susanne Fuchs]

Pages: 1720–1733

I. INTRODUCTION

A. Co-speech gesture

Despite the fact that speech, by its very definition, involves the gestural coordination of vocal tract articulators as its primary source of delivery, there is little doubt that body gestures also play a fundamental role in human communication. Such gestures are closely related to speech itself and are present in practically all spoken communicative acts, even when an interlocutor is not physically present (Goldin-Meadow, 1999; Rime, 1982). Although the manual modality (i.e., hand-related gestures) of such co-speech articulation has often been the focus of research into this gesture-speech interaction (Kita, 2000; McNeill, 1992), movements of other parts of the body (torso, shoulders, head, mouth, eyebrows) also co-occur with speech production in patterned ways (Condon, 1976; Wagner *et al.*, 2014). In recordings of conversational English, for example, Hadar *et al.* (1983) observed a general correlation between the occurrence of head movements and peaks of acoustic speech intensity. This close relationship between body gestures and

speech gestures can be assumed to occur at a linguistically functional level, since co-speech gestures have been observed to complement or supplement the linguistic meaning expressed via speech.¹

1. Timing of co-speech gestures

Body gestures have been found to be temporally coupled with speech in a multitude of ways. Speech and gestures are synchronized at relatively large time scales, e.g., pacing and speech rate: Manual gestures slow down if speech is slowed down (Kelso *et al.*, 1983; Pouw and Dixon, 2019; Stoltmann and Fuchs, 2017), speech slows down if a manual gesture is interrupted (Chu and Hagoort, 2014), and greater entrainment between jaw (speech) and head (co-speech) movement has been observed as speech rate increases (Tiede *et al.*, 2019). However, there is evidence to suggest that speech gestures and body gestures are also temporally intertwined at a more fine-grained level, as well, i.e., within a phrase, a word, or even a syllable. In his seminal work on co-speech gesticulation, Kendon (1980) proposed that the timing of the most “effortful” part of the movement (i.e., the *stroke*) of a co-speech gesture either precedes or is synchronized with, but does not follow, the most

^{a)}Email: c.carignan@ucl.ac.uk

prosodically prominent syllable of a speech utterance. Subsequently, McNeill (1992) formalized a phonological synchrony rule by which speakers temporally align the stroke of a co-speech gesture with the interval of phonological (or prosodic) prominence in a speech utterance: The other phases of the gestural movement—the *onset* (i.e., the beginning of the gesture) and the *apex* (i.e., the end of the gesture, the point of maximum extension of the movement itself)—are thereby determined by this synchrony between the (co-speech) gestural stroke and the phonological (speech) prominence. For the purposes of the current study, rather than referring to a temporal *interval* of the gesture, we will henceforth use the word “stroke” to refer to a primary temporal *moment* of the gesture stroke, i.e., the point of most rapid movement, in order to differentiate three key successive temporal moments of the gesture: the onset, the stroke, and the apex.

Further studies have since revealed how the timing of prosodic events influences the temporal alignment of body movements with speech (Esteve-Gibert *et al.*, 2017; Esteve-Gibert and Guellà, 2018; Hadar *et al.*, 1984; Krivokapić, 2014; Leonard and Cummins, 2011; Loehr, 2007; Renwick *et al.*, 2004; Roustan and Dohen, 2010, *inter alia*), including the effect of prosodic structure on the alignment of speech with manual gestures involving the fingers (Rochet-Capellan *et al.*, 2008) or the entire hand (Krivokapić *et al.*, 2017). In the current study, we extend this body of work by exploring the nature of prosodic and co-speech gestural alignment of head movements, as well as the role of prominence of the prosodic events in this alignment.

2. Magnitude of co-speech gestures

Prosodically prominent syllables may be correlated with larger co-speech gestural movements (Ambrazaitis and House, 2022; Parrell *et al.*, 2014), suggesting that synchrony can occur between gestures and speech not only in time but also in magnitude. Such a magnitudinal relationship has been observed for both upper and lower limbs, indicating a potential biomechanical link to the torso, where the power-generating organs of speech (e.g., the lungs and diaphragm) are found (Pouw and Fuchs, 2022): More rapid manual gestures during pointing lead to higher amplitude envelope peaks and higher fundamental frequency (F0) values in the speech signal (Kadavá *et al.*, 2023), and higher acceleration of leg movements is correlated with greater speech amplitude envelope (Serré *et al.*, 2022). Pouw *et al.* (2020) found that the moments of physical impulses in arm or wrist movement coincide with peaks of speech intensity and F0 frequency, suggesting that the gesture-speech relationships of both time and magnitude may in fact be interrelated.

The temporal-magnitudinal relationship between gesture and speech may potentially serve a communicative purpose and can be integrated into the processing of speech by a listener. For example, Pouw *et al.* (2022) found that listeners spontaneously synchronized their arm movements with those of an interlocutor, whom they could hear but not see.

This suggests that the listeners used acoustic cues alone to identify the changing velocity and magnitude of the interlocutor’s hand movements, thus relying upon the temporal and magnitudinal relationship between gesture and speech in their processing of the speaker’s speech-gesture production. Ultimately, research into co-speech gestures should investigate not only the timing of the gesture-speech relationship, but also the kinematics of the gesture itself. Therefore, in the exploratory analysis presented in the current study, we investigate both timing and kinematics of co-speech gesture in its relation to prosodic prominence.

3. Sources of speech-gesture interaction

The alignment of co-speech body gestures with vocal tract articulatory gestures can be influenced by cognitive (i.e., linguistic) and/or non-cognitive (i.e., biomechanical) factors. Biomechanical forces have been found to guide the coupling of respiration, body movements, and speech (e.g., Fuchs and Rochet-Capellan, 2021; Pouw *et al.*, 2020; Schmid *et al.*, 2004; Serré *et al.*, 2022, *inter alia*), due to biomechanical interactions between respiration and the vocal system (Klein and Codd, 2010), between the skeletal and respiratory systems (Levin, 1997, 2006), and—of particular interest to the current study—between the head and vocal system (Anegawa *et al.*, 2008; Miller *et al.*, 2012; Moisk *et al.*, 2019).²

Above and beyond these biomechanical forces, linguistic factors can influence the alignment of body and speech gestures. In languages with pitch accents, speakers spontaneously align the stroke and apex of manual co-speech gestures with the temporal boundaries of pitch accents and prosodic breaks (Alexanderson *et al.*, 2013a; Ambrazaitis and House, 2022; Esteve-Gibert *et al.*, 2017; Esteve-Gibert and Prieto, 2013; Jannedy and Mendoza-Denton, 2005; Leonard and Cummins, 2011; Renwick *et al.*, 2004; Rohrer, 2022; Shattuck-Hufnagel and Ren, 2018; Türk and Calhoun, 2023). In languages without pitch accents, gesture apexes are prone to align within the edges of the prosodic word and are not aligned as closely with F0 peaks (Fung and Mok, 2018). In a correlation analysis between head movements and F0 values, the occurrence of outlying observations with very low correlation between these speech and co-speech phenomena led Yehia *et al.* (2002) to posit that the coupling between head motion and F0 is indeed functional, rather than mechanical.

The pragmatic demands of the interlocutor are another important linguistic factor in the alignment of co-speech gesticulation. Gestures that provide redundant information (compared to gestures that complement spoken information) and those that are more relevant for the interlocutor (compared to those that are more predictable or less important for the message) have been observed to show tighter temporal alignment with speech (Bergmann *et al.*, 2011). There is also evidence that new discourse referents (which themselves are more prosodically prominent than referents that have previously been a part of the discourse) are more often

accompanied by co-speech gesturing compared to given referents (Ambrazaitis and House, 2017; Debrelioska *et al.*, 2013; Ferré, 2014; Loehr, 2007; Rohrer, 2022). Most pertinent to the current study are the findings that co-speech gestures that accompany narrowly focused speech elements have shown closer temporal alignment with speech than gestures that accompany broadly focused elements (Kim *et al.*, 2014) and that the distinction between narrowly focused head motion and broadly focused head motion may be regulated by speaking style (Pagel *et al.*, 2023). Speakers thus coordinate gesture and speech in a way that seems to be modulated by the communicative purposes they are intending to express, and this coordination may be used to enhance paralinguistic effect.

B. The current study: Aims and hypotheses

The research outlined above suggests that speakers may coordinate the timing and magnitude of co-speech gestures with linguistic phenomena of spoken language and that speakers' co-speech gesturing can help listeners process linguistic units of the speech chain, potentially even for high-level linguistic domains. The presence of co-speech body movements has been shown to increase syllable recognition (Munhall *et al.*, 2004) and the perception of prosodic prominence (House *et al.*, 2001; Krahmer and Swerts, 2007) and lexical stress (Bosker and Peeters, 2021) and can be linked to the production of pitch-accented words (Swerts and Krahmer, 2010). In the current study, we investigate how the timing and kinematics of co-speech head nods may (or may not) be entangled with different levels of prosodic prominence for two different levels of narrow focus in French.

To this aim, we examine head movement correlates— with regard to both gesture timing and magnitude—of two degrees of narrow focus in French interactive speech, i.e., contrastive (mild) focus and corrective (strong) focus. Previous work has shown that, in a similar task to the one carried out here, French preschoolers mark focus only through head movement (but not through prosodic strategies), by accompanying contrastive and corrective focus words with more frequent head gestures than broad focus productions (Esteve-Gibert *et al.*, 2022). In this study, we investigate whether adult speakers (who *do* use prosodic strategies) align head nods with F0 peaks, whether the alignment is dependent on focus type (contrastive vs corrective), and how the alignment may be realized with regard to the dimensions of both time and magnitude.

As is the case for other languages, most studies of prosody-gesture alignment in French have focused on manual co-speech gestures. For instance, Roustan and Dohen (2010) have shown that prosodic emphasis in speech attracts prominent landmarks in manual gesture movements. Rohrer (2022) found that when both an initial phrase accent and a final phrase accent are present, gesture apexes align more often with the initial one. Data on head movements are scarce, however, and descriptions of which features of head

movement align with prosodic anchors are lacking. This study aims at filling this gap by motion-tracking head movements of French speakers while they convey two degrees of focus, by examining whether and how head-gesture events align with F0 peaks, and by investigating whether this alignment differs by focus strength (i.e., the degree of prosodic prominence).

Our hypothesis is that greater prominence (i.e., corrective focus) should show greater head nod alignment with F0, either temporally or with regard to some aspect of the kinematics of the head nod gesture. As explained previously, various manual events have been observed to synchronise with prosodic events: gesture onset, gesture stroke, and gesture apex (e.g., Esteve-Gibert and Prieto, 2013; Krivokapić *et al.*, 2017; Pouw *et al.*, 2023). Given this variability, different parameters of the head nod gesture are first considered in order to understand the temporal and kinematic nature of potential alignment with prosodic focus events. These parameters are used in a clustering analysis to identify—using a combination of knowledge-based (top-down) and data-driven (bottom-up) approaches—potential strategies in conveying various degrees of prosodic prominence through both speech and head gesture. In this way, we impose few constraints or assumptions on how different co-speech alignment strategies might be realized (if they indeed are realized at all).

II. METHODS

A. Data collection

1. Speakers and material

A total of 19 French native speakers originally participated in the production task. From these, four participants had to be excluded due to technical problems in recording their head movements (i.e., the video recording stopped inexplicably after the first few minutes), two participants were excluded due to lack of engagement in the task (i.e., they did not move during data collection), and one participant was excluded because they did not follow the task instructions, resulting in a final retention of 12 participants for analysis. These were all undergraduate or graduate students at the University Aix-Marseille and received monetary compensation for their participation. The study was approved by the ethics committee of the University Aix-Marseille.

Participants took part in a task designed to elicit the semi-spontaneous production of utterances in distinct degrees of focus (see details of the task and materials in Esteve-Gibert *et al.*, 2022). Participants saw a visual display in a POWERPOINT presentation and had to interact with a character that had to pick certain objects from a big bag. The elicited target phrases had the structure of [Verb +] Article + Noun + Adjective, e.g., [*Prends*] *la valise orange* (“[Take] the orange suitcase”). Focused words could be either be the noun or the adjective, although we do not differentiate between these word types for the purposes of the current

study (see Sec. II B 1). The distinct focus conditions were elicited by manipulating the nature of the target objects to be picked and the potential alternatives available in the visual display. Although the original task was designed to also elicit broad focus productions, only utterances under narrow (contrastive and corrective) focus contexts were selected for the purpose of the current study, in order to compare head nod movements under these two different degrees of contrast, for a focus context that has previously been shown to be associated with closer co-speech alignment (Kim *et al.*, 2014). Repp (2016) suggests that discourse relations determine the degree of contrast that is expressed. In this regard, whenever the character in our visual game had picked the wrong object in a previous attempt, a more prominent focus (i.e., corrective focus) was expected to be produced by the participant as a response. We therefore follow in this study the notion that contrast may be a gradable phenomenon (Calhoun, 2010; Molnár, 2006; Paoli, 2009; Repp, 2016), in which we consider corrective focus to be *stronger* than contrastive focus.

2. Electromagnetic articulometry

Kinematic data were collected at a sampling rate of 200 Hz using a Carstens AG501 electromagnetic articulograph (Carstens Medizinelektronik, Bovenden, Germany), using sensors placed at the following locations: the left and right mastoid processes, the nasion, the peak of the left and right eyebrows, the chin, and the vermilion border of the upper and lower lips. In the current study, we are interested in downward vertical head nod gestures; as such, we focus here on the three sensors which track rigid body movement of the head: the left and right mastoids and the nasion. The participants remained seated throughout the data collection session.

B. Data analysis

In the following sections, we describe a combination of knowledge-based (top-down) and data-driven (bottom-up) approaches used to analyze how multiple dimensions of co-speech head nod gestures might be coordinated with pitch prominence, as well as how focus strength may interact with this coordination. The methodological pipeline can be summarized as outlined below, with the numbering of the methodological steps matching the numbers of Secs. II B 1–II B 7 in the text:

- (1) Visual identification of head nod gestures in video data
- (2) Auditory identification of F0 peaks in audio data
- (3) Quantification of vertical head movement in electromagnetic articulographic data
- (4) Creation of kinematic and temporal metrics to capture potential aspects of coordination between the F0 peaks (i.e. speech events) and gestures (i.e. co-speech events)
- (5) Feature integration and de-correlation using principal component analysis (PCA)
- (6) Identification of latent clusters in the retained principal component (PC) scores (i.e., head nod “strategies”)

- (7) Investigation of the interaction between these strategies and the degree of focus strength using linear mixed-effects (LME) models and estimated marginal means (MMEs)

1. Visual identification of head nods

Occurrences of head and body movements were first manually identified from video data by the second author and labeled using the EUDICO Linguistic Annotator (ELAN) (Language Archive, 2023). The annotator only had access to video data and the time intervals of the segmented speech: Audio data were not used for identifying movement events, and the annotator was blind to the intended focus condition. The annotator visually identified gestures whose interval (or part of it) was produced during the elicited target phrase and then labelled these occurrences. Any movement for which a gesture interval (either whole or partial) was produced during the elicited target phrase was included in the annotation.

All gesture types were labeled, including torso movements, eyebrow raising, chin pointing, lateral head movement (i.e., head “shakes”), head tilts, and vertical head nods. Table I displays the relative proportions of these movements for the 12 speakers included in the current study. The majority of utterances (59.2%) were not produced with any co-speech gesturing, 24% of the utterances were produced with a single vertical head nod, and 16.8% of the utterances were produced with some other type of co-speech head/body movement. The word bearing the gesture (i.e., the word that the speaker was producing when the gesture occurred in the video) was also labeled in the ELAN transcription by the annotator; this word was either the noun or the adjective in all cases.

Among the 24% of movements that were annotated as single vertical head nods, only those that were produced by the participant as canonical Noun + Adjective readings were retained. This manual process resulted in a total of 116 head nod gestures identified across the 12 speakers [5–20 occurrences per speaker, mean = 9.67, standard deviation (SD) = 5.23], 50 of which (43.1%) were produced with contrastive focus and 66 of which (56.9%) were produced with corrective focus.³ In each relevant phrase, the word bearing the nod (i.e., the word that the speaker co-produced with the gesture, as explained above) was used for subsequent analysis, which in all cases was either the noun or the adjective of the target phrase.

2. Auditory identification of F0 peaks

F0 estimations were made in PRAAT (Boersma and Weenink, 2021) for the phrases associated with the 116

TABLE I. Relative proportions of co-speech movements identified in the video recordings.

Movement type	Proportion (%)
Vertical head nod	24.0
Other movement	16.8
No gesture	59.2

visually identified vertical head nods. Similar to the manual identification of head nods in the video data, the third and fourth authors (both native speakers of French) visually and auditorily identified F0 peaks in the audio data that they considered to be perceptually prominent. The annotators only had access to the audio data and the time intervals of the segmented speech; video data were not used for identifying F0 peaks. For each of the 116 phrases that were identified as containing an occurrence of a head nod, the time of the F0 peak that was closest to the ELAN-annotated word bearing the head nod gesture was logged.

3. Head nod metric

For each speaker, the center of the head was estimated as a reference for calculating vertical head nod movement, in the following way (a schematic of the process is shown in Fig. 1). First, the line extending between the two mastoid sensors was used as the axis of vertical rotation, i.e., *pitch angle*. Second, the unit vector extending from the inter-mastoid midpoint to the nasion sensor was calculated in Cartesian space, and the three-vector components ($\hat{x}, \hat{y}, \hat{z}$) were subsequently converted to spherical coordinates (ρ, θ, ϕ). Finally, the polar angle of the spherical coordinate system, ϕ , was normalized for each speaker via z-score transformation and used as a metric of vertical head movement, relative to the center of the head (i.e., the inter-mastoid midpoint). In this way, the final metric is interpreted as speaker-normalized movement in the *x-z* (i.e., sagittal) plane of the face (i.e., the nasion) relative to the center of the head (i.e., the inter-mastoid midpoint).

4. Gesture interval quantification

For the purposes of this study, we are interested in downward head nod movements (Alexanderson *et al.*, 2013a,b;

House *et al.*, 2001; Ishi *et al.*, 2007; Maynard, 1989). As such, time points of minimum velocity of the head nod signal (i.e., the most rapid downward movement) were used to locate head nod gestures. For each phrase that was identified as containing an occurrence of a head nod (Sec. II B 1), the time point of minimum velocity (i.e., maximum negative/downward velocity) nearest in time to the ELAN-annotated word was used to automatically identify the relevant prominent head nod. Subsequently, 20% velocity thresholds were used to demarcate the onset and the offset/apex of the head nod gesture: The time point preceding the point of minimum velocity that crossed 20% of the velocity value was taken as the head nod onset, and the time point following the point of minimum velocity that crossed 20% of the velocity value was taken as the head nod offset, i.e., the gesture apex.

An example of this head nod gesture quantification is shown in Fig. 2. In this example, the F0 values are denoted by red circles, and the head nod signal is denoted by the solid black line. The acoustic boundaries of the words *robe rouge* (“red dress”) are denoted by the vertical yellow line (*robe*) and blue line (*rouge*), with the shared boundary between the two words denoted by the dashed yellow/blue line. The three key kinematic time points—onset, peak velocity (i.e., the stroke), and apex—of the downward head nod gesture are denoted by the dotted vertical gray lines, while the vertical gray rectangle denotes the entire gesture, from the onset to the apex. In this example, the head nod gesture begins toward the end of the word *robe* and ends near the middle of the word *rouge*, the point of minimum velocity (i.e., peak downward velocity, the stroke) is aligned with the acoustic boundary between the two words, and the apex of the gesture is roughly aligned with the F0 peak that occurs in the middle of *rouge*. In this example utterance, the word “rouge” was the word that was visually identified as bearing the head nod gesture (see Sec. II B 1).

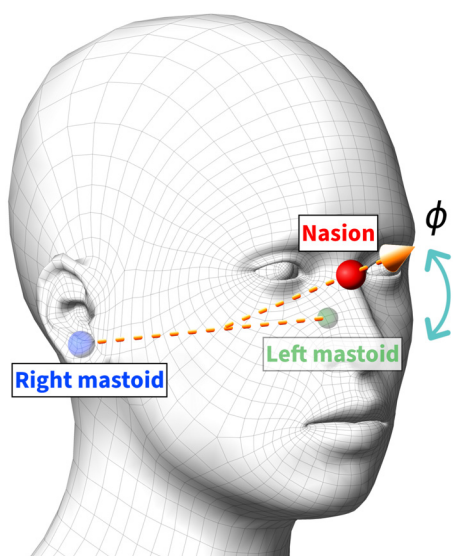


FIG. 1. (Color online) Schematic representation of the generation of the head nod metric (ϕ) used in this study: the spherical polar angle related to the Cartesian components of the vector extending from the midpoint of the two mastoid EMA sensors to the nasion EMA sensor.

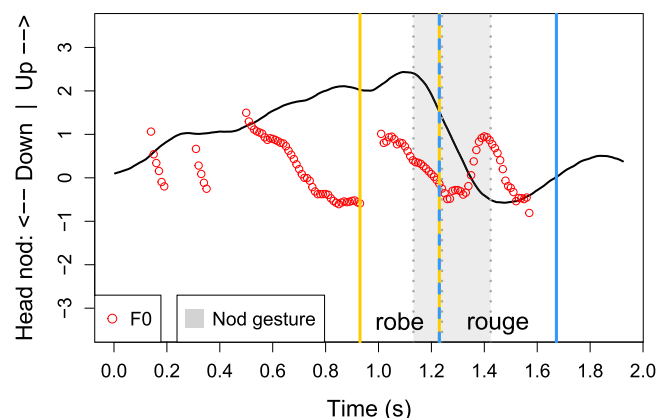


FIG. 2. (Color online) An example of the correspondence between the head nod metric (solid black line) and F0 measurements (red circles) for an utterance containing the noun phrase *robe rouge* (“red dress”). The acoustic boundaries of the two words are denoted by solid/dashed vertical colored lines, the time points of the three key kinematic moments of the head nod gesture (onset, peak velocity, apex) are denoted by dotted vertical gray lines, and the kinematic boundary of the head nod gesture (i.e., from onset to apex) is denoted by the gray rectangle.

In order to observe patterns of temporal alignment of the three key time points of the head nod gesture (onset, maximum downward velocity, apex) with the time point of the F0 peak, each of these time points was normalized as a relative percentage of the word duration. Figure 3 displays the temporal alignment of the resulting word-normalized time points for the three parts of the head nod gesture (onset, maximum downward velocity, apex) and the F0 peak, represented as probability densities of the respective time points. As a percentage of the interval of the word bearing the nod (see Sec. II B 1), the onset of the gesture occurs on average at 0.7% (median, 2.1%) of the word interval, the gesture stroke (i.e., the point of maximum downward velocity) occurs on average at 32.6% (median, 30.5%) of the word interval, the gesture apex occurs on average at 62.6% (median, 59.0%) of the word interval, and the F0 peak occurs on average at 28.1% (median, 37.2%) of the word interval. In comparison with the timing of the F0 peak, the gesture onset occurs on average 28.8% earlier (median, 31.9%) than the F0 peak, the gesture stroke occurs on average only 4.5% later (median, 0.9% earlier) than the F0 peak, and the gesture apex occurs on average 34.5% later (median, 20.2%) than the F0 peak. In summary, these results suggest that the onset of the head nod gesture is generally aligned with the onset of the focused word, whereas the gesture stroke is generally aligned with the F0 peak.

As can be seen in Fig. 3, whereas the F0 distribution generally displays a unimodal characteristic (i.e., there is a single primary peak of values just prior to the midpoint of the word interval), the timing of the head nod gesture displays multimodal characteristics (i.e., there are both a primary peak of values and an earlier secondary peak of values, for each of the three key points of the head nod gesture). This suggests that there are multiple strategies of temporal alignment between the head nod gesture and the F0 peak present across the 116 items. In the following sections (Secs. II B 5–II B 7), we investigate the possibility of multiple strategies of temporal and magnitudinal coordination using unsupervised, non-parametric Gaussian mixture clustering of underlying orthogonal

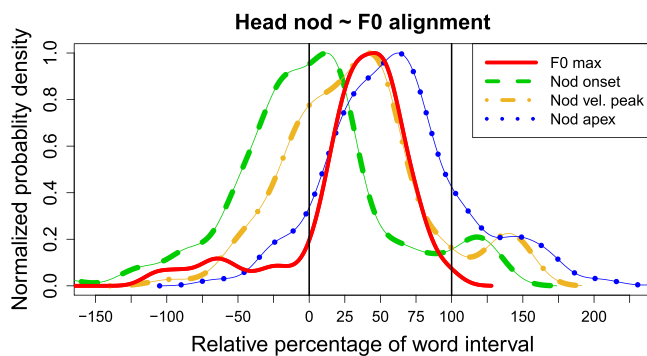


FIG. 3. (Color online) Probability densities of the distributions of time points related to the three key points of the head nod gesture (onset, velocity peak, apex) and the F0 peak. The time points are normalized as a relative percentage of the interval of the word on which the head nod occurs. The relative distributions of the four time points are denoted by line type and color.

dimensions shared by the relevant temporal, acoustic, and kinematic features that characterize these items.

5. PCA

Given that our primary research question is to determine if there are any differences in speech/co-speech coordination between focus types when co-speech head nod gestures occur, it is reasonable to assume that these differences might not necessarily be limited to the temporal domain: Such coordination can be realized through any dimension (or combination of dimensions) that might capture change in head nod and/or pitch movement. Moreover, within such a combination of gestures, we are most interested in their shared attributes, i.e., the covariance structures among these dimensions. For these reasons, in order to investigate the nature of the multiple strategies suggested above (Sec. II B 4), we focus on orthogonal dimensionality-reduction of multiple features related to kinematics and timing of both the head nod gesture and F0 peaks.

The following nine features were constructed to capture various potential aspects of speech/co-speech coordination:

- “onset.lag” = the temporal lag between the word-normalized time point of the F0 peak and the word-normalized time point of the head nod onset. Positive values indicate that the gesture time point occurs after the F0 peak, and negative values indicate the opposite.
- “velmin.lag” = the temporal lag between the word-normalized time point of the F0 peak and the word-normalized time point of maximum downward velocity (i.e., the gesture stroke). Positive values indicate that the gesture time point occurs after the F0 peak, and negative values indicate the opposite.
- “apex.lag” = the temporal lag between the word-normalized time point of the F0 peak and the word-normalized time point of the head nod apex. Positive values indicate that the gesture time point occurs after the F0 peak, and negative values indicate the opposite.
- “velmin” = the value of the maximum downward/negative velocity
- “gest.range” = the ϕ range of the head nod gesture
- “stiffness” = the kinematic stiffness of the gesture, operationalized as $|\text{velmin}|/\text{gest.range}$
- “log.dur” = the log-transformed duration (ms) of the head nod gesture
- “f0max” = the value of the F0 peak (Hz)
- “f0max.norm” = the word-normalized time point of the F0 peak

The data from these nine features were submitted to a PCA model. PCs with eigenvalues greater than 1 were retained (i.e., the Kaiser criterion) and verified visually using a scree plot.

6. Gaussian mixture model clustering

The retained PC score data were submitted to non-parametric Gaussian mixture modeling using the MCLUST R package (Fraleigh *et al.*, 2022), in order to verify if there are

indeed underlying structures within the data and, if so, how many. Here, a benefit of the PCA transformation is the orthogonal nature of the resulting features, which avoids the finite mixture model biasing the results toward clusters of multiple correlated features. Instead, clustering based on the uncorrelated PC scores allows observation of subgroups of the data set based on its underlying dimensions of covariance.

7. Statistical modeling of results

After identification of potential underlying clusters in the retained PC score data, separate LME models were constructed for each of the nine original features as a response variable, with fixed effects for cluster group and focus type (as well as their interaction), along with by-speaker random intercepts; all models were constructed using the LME4 R package (Bates *et al.*, 2023). In order to investigate the nature of significant interactions, EMMs were computed using the EMMEANS R package (Lenth *et al.*, 2023) with the default alpha adjustment method [Tukey’s honestly significant difference (HSD)].

III. RESULTS

A. PCA results

The first three PCs (PCs 1–3) were retained according to the selection criteria (Sec. II B 5), which cumulatively explain 82.9% of the total variance among the original nine features. Two-dimensional biplots of PCs 1–3 are shown in Figs. 4 and 5, with the focus type denoted by point shape and color. PC1, which explains 37.4% of the total variance, is interpreted as a dimension related to the timing of the F0 peak and the nod gesture: Positive scores are associated with an earlier F0 peak and delayed head nod gesture (i.e., all three moments of the gesture). PC2, which explains 28.4% of the total variance, is interpreted as a dimension related to the kinematics of the head nod gesture: Positive scores are associated with increased stiffness, reduced range,

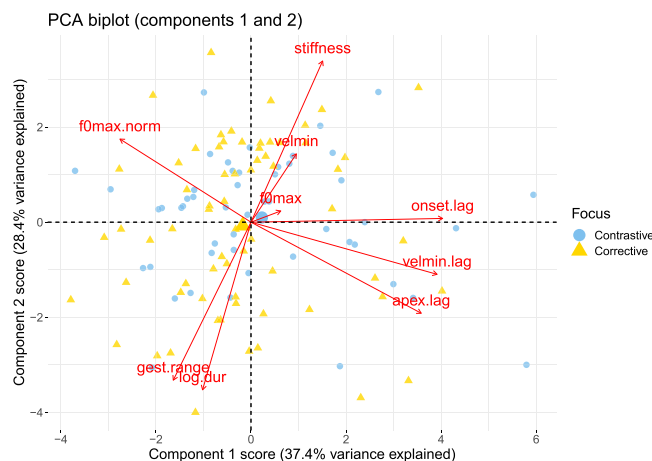


FIG. 4. (Color online) A biplot of scores and loadings related to principal components 1 and 2. The two focus types are denoted by point shape and color.

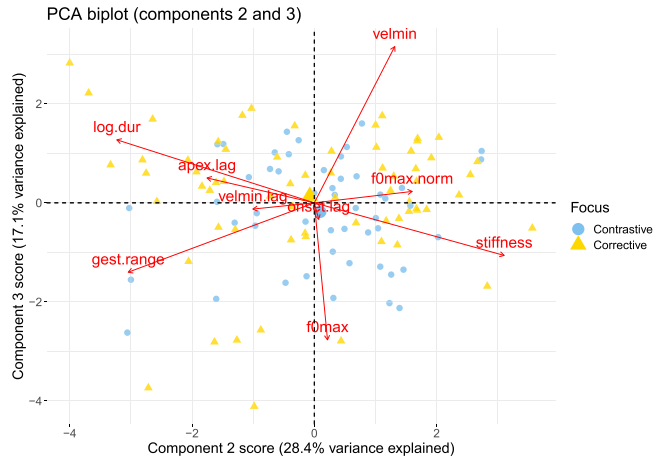


FIG. 5. (Color online) A biplot of scores and loadings related to principal components 2 and 3. The two focus types are denoted by point shape and color.

and shorter duration. PC3, which explains 17.1% of the total variance, is interpreted as a more complex interaction between primarily the velocity of the head nod and the value of the F0 peak: Positive scores are associated with a slower nod gesture (i.e., negative velocity increasing toward 0) and lower F0 peak value.

B. PC score clustering results

The Gaussian finite mixture model revealed that the retained PC score data are represented best by a mixture of two underlying Gaussian distributions of varying volume and equal shape, whose axes are parallel to the coordinate axes (a “VEI” model): Henceforth, the terms “clusters” and “head nod strategies” will be used interchangeably to refer to these two underlying Gaussian distributions. Of the 116 total items, 65 were classified in cluster 1 and 51 in cluster 2. The relative proportions of the items identified in these clusters were the same for both focus types: 56% in cluster 1 for both contrastive focus (28/50 items) and corrective focus (37/66 items), and 44% in cluster 2 for both contrastive focus (22/50 items) and corrective focus (29/66 items). There were, however, slight differences in how these clusters were distributed across the type/class of the word bearing the nod. In total, 78 of the 116 head nods were produced on adjectives and 38 were produced on nouns. Sixty-three percent of adjectives (49/78 items) and 42% of nouns (16/38 items) were produced with cluster 1 strategies, while 37% of adjectives (29/78 items) and 58% of nouns (22/38 items) were produced with cluster 2 strategies.

There is a large degree of variability in how these clusters are used across the 12 speakers; a table of the counts and summary statistics for all speakers can be found in the Appendix. Figure 6 shows the relative proportions of total head nod gestures produced by each speaker, with lines connecting each speaker’s values for the two clusters. Here, lines with flatter slopes indicate speakers who use a relatively equal proportion of nod gestures from both cluster groups, while lines with steeper slopes (either positive or

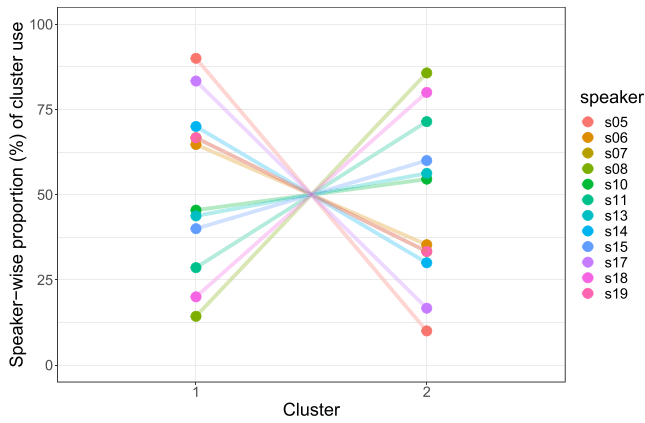


FIG. 6. (Color online) Relative proportions of total head nod gestures produced by each speaker, with speaker-specific lines connecting the proportion of nods produced with the strategy identified in cluster 1 to the proportion of nods produced with the strategy identified in cluster 2.

negative) indicate speakers who primarily used one cluster group over another. There are no discernable patterns in these proportions: The proportions are spread fairly evenly across the speakers, rather than clear groupings of speakers who use one head nod strategy over another. Indeed, non-parametric Gaussian mixture modeling revealed that the proportions are best represented by a univariate normal distribution, suggesting that there are no clusters present within the speaker-wise proportions; i.e., there are not groups of speakers who use one cluster over another. However, the large range of values—from speaker s05, who uses predominantly (90%) head nods of cluster 1, to speaker s08, who uses predominantly (86%) head nods of cluster 2—suggest that there may be speaker-specific tendencies for preferring one head nod strategy over the other, although all speakers produced head nods from both cluster groups.

A pairs plot of the classified observations (Fig. 7) suggests that the clustering is based primarily on PC2 score values, with little to no discriminability provided by PC1 or PC3 scores. This was verified via dimension reduction using the *MclusDR* function of the *MCLUS* package, which revealed that the magnitude of the contribution of PC2

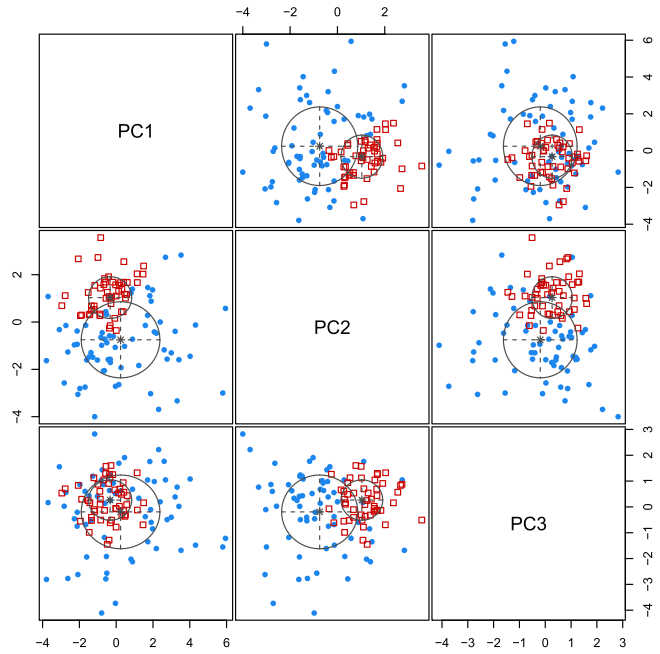


FIG. 7. (Color online) A pairs plot of the three principal component scores for all head nod gestures, with the point color and shape denoting the classification identified by non-parametric Gaussian mixture modeling as either cluster 1 (blue circles) or cluster 2 (red squares).

(0.90) to the basis vector of the dimension that separates the two clusters is more than double the contributions of PC1 (−0.22) or PC3 (0.38). In other words, the mixed-Gaussian clustering is determined primarily by the kinematics of the head nod gesture: stiffness, range, and duration (i.e., the very features that form the basis of the PC2 dimensionality; see Sec. II B 5).

C. Characteristics of head nod strategies

The results from the separate LME models constructed for each of the nine original features as a response variable, with fixed effects for cluster group and focus type (along with their interaction), are shown in Table II.

TABLE II. Estimates and standard errors for LME models constructed for the nine original kinematic and acoustic features as a response variable.^a

LME model	Estimate (SE) for model with feature:								
	onset.lag	velmin.lag	apex.lag	velmin	gest.range	Stiffness	log.dur	f0max	f0max.norm
(Intercept)	5.03 (12.52)	37.83*** (10.52)	68.50*** (11.03)	−0.03*** (0.00)	1.36*** (0.13)	0.03*** (0.00)	4.03*** (0.09)	216.37*** (18.80)	10.80 (9.62)
Cluster 2	−54.77** (18.09)	−63.57*** (15.79)	−70.17*** (16.63)	0.01 (0.00)	−0.52* (0.20)	0.00 (0.00)	−0.21 (0.13)	−4.95 (7.42)	35.98** (13.02)
Focuscorr	−35.38* (15.96)	−21.53 (13.89)	−9.64 (14.62)	−0.00 (0.00)	0.47** (0.17)	−0.01* (0.00)	0.30* (0.11)	−2.83 (6.65)	6.12 (11.53)
Cluster 2: focuscorr	45.08 (23.97)	27.67 (20.94)	9.28 (22.04)	0.01 (0.01)	−0.70** (0.26)	0.01** (0.00)	−0.51** (0.17)	−7.73 (9.95)	−7.06 (17.27)

^aSE, standard error; Cluster 2: focuscorr, relationship between cluster 2 and corrective focus. Asterisks indicate the relative level of significance: ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$.

With regard to the main effect of cluster group, in comparison to cluster 1, cluster 2 is characterized by an earlier head nod gesture (i.e., all three moments of the gesture), a reduced gesture range, and delayed F0 peak. With regard to the main effect of focus type, in comparison to contrastive focus, corrective focus is characterized by an earlier head nod onset, an increased gesture range, decreased kinematic stiffness, and increased gesture duration. With regard to the interaction between cluster group and focus type, there are significant effects for gesture range, stiffness, and duration. Intriguingly, these are the same three features which contribute most strongly to the PC2 dimension (Sec. II B 5), which itself was observed to be the primary basis for discrimination in the Gaussian mixture model (Sec. II B).

D. Interaction of clustering and focus

The EMMs used to investigate the nature of the significant interactions in the LME models are displayed in Figs. 8–10. For each of these EMM plots, the blue bars denote the confidence intervals and the red arrows denote the direction of each pairwise comparison. *Ad hoc* interpretations of the effect of a given pairwise comparison should be considered with reference to the lengths of the arrows, which represent the amounts by which corresponding confidence intervals for the comparison cover the value zero: Where any given pair of arrows do not overlap along the x axis, the corresponding pairwise difference is estimated to be significant at the Tukey’s HSD-adjusted level.

With regard to the range of the head nod gesture (Fig. 8), in comparison to cluster 2, the cluster 1 gesture is characterized by greater range for both focus types (i.e., for both contrastive focus and corrective focus—in other words, the significant main effect described above). However, the significant interaction between cluster group and focus type is evidenced in the fact that the effect of cluster group is large for corrective focus, but small for contrastive focus (indeed, not surpassing the threshold for significance). In other words, the kinematic differences between the two head nod strategies are similar for both focus types, but the difference is stronger under corrective (i.e., strong) focus compared to contrastive (i.e., mild) focus.

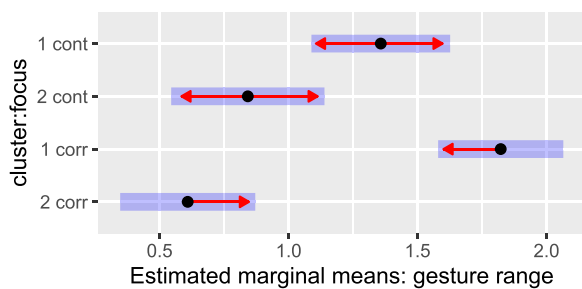


FIG. 8. (Color online) Estimated marginal means of gesture range computed with Tukey’s HSD alpha adjustment, focusing on the significant interaction between cluster group and focus type. Arrows that do not overlap along the x axis denote pairwise comparisons that are identified as significantly different at the alpha-adjusted level.

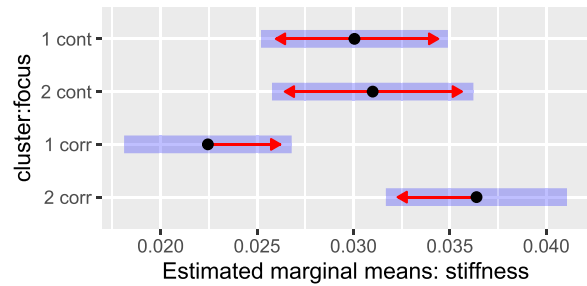


FIG. 9. (Color online) Estimated marginal means of gesture stiffness computed with Tukey’s HSD alpha adjustment, focusing on the significant interaction between cluster group and focus type. Arrows that do not overlap along the x axis denote pairwise comparisons that are identified as significantly different at the alpha-adjusted level.

With regard to the kinematic stiffness of the head nod gesture (Fig. 9), in comparison to cluster 2, the cluster 1 gesture is characterized by lesser stiffness for both focus types. However, whereas the effect of cluster group is large for corrective focus, it is essentially non-existent for contrastive focus (although the difference trends in the same direction as for corrective focus). Although the difference is evident for corrective focus, the overlapping values for contrastive focus result in the lack of a significant main effect for cluster group with all data combined. Here, again, the kinematic differences between the two head nod strategies are similar for both focus types, but the difference is stronger under corrective (i.e., strong) focus compared to contrastive (i.e., mild) focus.

With regard to the duration of the head nod gesture (Fig. 10), in comparison to cluster 2, the cluster 1 gesture is characterized by longer duration for both focus types. However, like for gesture range and kinematic stiffness, the difference is only significant for corrective focus, resulting in the significant interaction observed above. Again, the overlapping values for contrastive focus result in the lack of a significant main effect for cluster group with all data combined. In other words, as was also the case for gesture range and stiffness, the kinematic differences between the two head nod strategies are similar for both focus types, but the difference is stronger under corrective (i.e., strong) focus compared to contrastive (i.e., mild) focus.

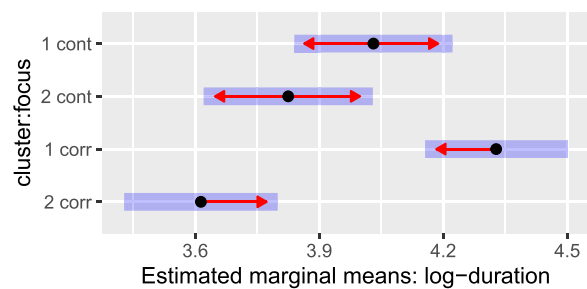


FIG. 10. (Color online) Estimated marginal means of gesture log-duration computed with Tukey’s HSD alpha adjustment, focusing on the significant interaction between cluster group and focus type. Arrows that do not overlap along the x axis denote pairwise comparisons that are identified as significantly different at the alpha-adjusted level.

IV. DISCUSSION

A. Summary of findings

We observed that all speakers used both gestural strategies to some degree. There is nonetheless a wide range of inter-speaker variability with regard to the preference for the use of one strategy over the other, suggesting that there may be speaker-specific tendencies in how co-speech head nod gestures are implemented in French. However, the nature of this inter-speaker variability is as yet unclear: Rather than clustering into distinct groups of gesture use, the variability is normally distributed across the 12 speakers, and non-parametric Gaussian mixture modeling revealed that there are indeed no clusters present within the cross-speaker variation, suggesting that the realization of these speaker preferences is gradient rather than categorical. Future research into speaker-specific preferences in co-speech gesturing may benefit from focusing on similar variability in a more targeted way, especially with regard to the causal factors involved in generating the variability.

With regard to the temporal alignment between the head nod gestures and speech, we observed in Sec. II B 4 that the onset of the head nod gesture is generally aligned with the onset of the word, whereas the stroke of the gesture (i.e., the point of maximum downward velocity of the head nod) is generally aligned with the F0 peak at a point near (but prior to) the midpoint of the word. Moreover, the PCA revealed that the underlying factor that explains the largest amount of variance across the various acoustic and kinematic metrics included in this study is related to the temporal alignment between the F0 peak and the three moments of the head nod gesture (Sec. II B 5). In this way, *co-speech* head nod gestures align with tone targets in a manner similar to gestures of *speech* articulators: D'Imperio *et al.* (2007) previously found evidence for synchrony between the maximum velocity of labial gestures and F0 peaks in French. These results also lend support to the phonological synchrony rule proposed by McNeill (1992) in which the stroke (i.e., the *moment of most rapid motion* in the current study) of a co-speech gesture is temporally aligned with the interval of phonological/prosodic prominence in a speech utterance.

We have also observed differences in head nod gestures between the two focus types (Sec. III C), especially with regard to the kinematics of the gestures: Corrective focus is characterized by an earlier head nod gesture onset, but also a greater gesture range, decreased kinematic stiffness, and longer gesture duration. This suggests that prosodic prominence in French has an effect on the kinematics as well as the temporal alignment of co-speech head nod gesturing and corroborates previous research that has shown that prosodically prominent syllables are correlated with larger co-speech gestural movements, e.g., in Swedish (Ambrazaitis and House, 2022) and English (Parrell *et al.*, 2014).

B. Prosodic enhancement

Somewhat surprisingly, it is neither the temporal alignment of the head nod gesture with prosodic events nor is it

the gestural differences between the focus types themselves that are independently responsible for the two underlying head nod strategies identified by the unsupervised, non-parametric Gaussian mixture clustering (Sec. III B). One strategy (cluster 1) involves a head nod gesture that occurs later, has a greater range, and is associated with an earlier F0 peak than the other (cluster 2). However, it is only in the way that these different strategies are used *between* the different focus conditions that we can observe a more nuanced explanation for how prosodic prominence influences the use of co-speech head gesture nods in French. The LME models (Sec. III C) revealed significant interactions for only three features: the range, stiffness, and (log) duration of the head nod gesture. Importantly, these were the same three features that contributed most strongly to the second PC of the combined acoustic/kinematic data (Sec. II B 5), and it was predominantly this component that was found to be the basis for the two gesture strategies identified by the unsupervised clustering (Sec. II B).

In examining these three significant interactions (Sec. III D), we observed that *both* gestural strategies were utilized under *both* focus conditions, but the magnitude of the differences between the two strategies was only large enough to reach a statistically significant level for the corrective focus condition. The statistical effect of prosodic prominence in our models can therefore be interpreted as a matter of *the intensity of the effect*, rather than the two prosodic conditions showing completely different strategies of temporal/magnitudinal alignment with tonal events. One head nod strategy (cluster 1) is produced with a greater range, is longer in duration, and has less kinematic stiffness than the other strategy (cluster 2), and the intensity of these effects is where we find an entanglement and alignment with speech prosody: Both co-speech head nod strategies are used for both focus conditions, but the difference between the two strategies is more prominent when they are produced under more prominent focus (i.e., corrective focus). In other words, prosodic prominence *magnifies the difference* between the co-speech gestural strategies observed here, rather than *motivating the difference*. In this way, the use of co-speech head nod gestures can be considered as a method of prosodic enhancement (Cho, 2005; Cho *et al.*, 2019) under French focus conditions, similar to the use of co-speech head nods as a method to enhance speaking style in German (Pagel *et al.*, 2023). The differences in the range, stiffness, and duration of the two head nod strategies are amplified when the level of prosodic prominence increases, thus serving as a way to reinforce and strengthen the higher level of prosodic prominence.

C. Study limitations

There are several limitations of the current study that should be considered when interpreting its results. First, the work presented here is largely exploratory in nature, and, like any exploratory study, the results and interpretations that arise from this preliminary work should be scrutinized

TABLE III. Counts, means, and SDs for the nine features used in the study.^a

Part.	Foc.	Clus.	Count	Mean (SD) for feature									
				onset.lag	velmin.lag	apex.lag	velmin	gest.range	Stiffness	log.dur	f0max	f0max.norm	
s05	cont	1	7	46.2 (66.7)	72.4 (70.9)	93.9 (68.3)	-0.04 (0.01)	1.26 (0.78)	0.035 (0.011)	3.80 (0.40)	295 (31)	-59.7 (107.9)	
	cont	2	1	14.0	28.6	37.0	-0.02	0.46	0.049	3.14	254	14.5	
	corr	1	2	-63.9 (53.3)	-39.5 (47.4)	-10.7 (36.4)	-0.07 (0.00)	2.65 (0.41)	0.026 (0.004)	3.95 (0.32)	283 (24)	28.2 (28.3)	
	corr	2	0										
s06	cont	1	5	37.3 (106.2)	67.5 (95.7)	109.0 (81.2)	-0.03 (0.01)	1.30 (0.66)	0.032 (0.021)	4.07 (0.75)	291 (37)	-3.6 (75.7)	
	cont	2	3	-46.9 (16.7)	-27.4 (22.9)	-8.1 (9.0)	-0.03 (0.01)	0.99 (0.10)	0.031 (0.008)	3.65 (0.20)	293 (21)	60.0 (5.3)	
	corr	1	6	-11.9 (57.0)	30.7 (44.9)	53.7 (50.2)	-0.04 (0.03)	1.82 (1.08)	0.025 (0.010)	4.11 (0.43)	276 (11)	18.0 (56.4)	
	corr	2	3	-40.1 (22.0)	0.1 (16.3)	22.9 (13.2)	-0.03 (0.01)	1.12 (0.30)	0.023 (0.004)	4.14 (0.14)	256 (8)	50.8 (18.3)	
s07	cont	1	3	-19.8 (46.8)	3.2 (37.7)	56.6 (29.3)	-0.03 (0.01)	1.64 (0.48)	0.019 (0.013)	4.24 (0.57)	166 (13)	11.5 (41.9)	
	cont	2	2	-17.4 (20.7)	-4.8 (33.6)	29.7 (31.7)	-0.02 (0.01)	0.62 (0.11)	0.026 (0.005)	3.84 (0.24)	151 (29)	38.7 (25.8)	
	corr	1	1	31.7	41.3	73.8	-0.01	0.52	0.029	3.74	168	-22.0	
	corr	2	0										
s08	cont	1	1	56.4	90.0	129.2	-0.02	0.67	0.029	4.29	189	40.5	
	cont	2	1	-71.0	-57.6	-12.1	-0.01	0.65	0.020	4.08	205	43.5	
	corr	1	0										
	corr	2	5	-29.8 (21.4)	-13.8 (14.6)	4.9 (15.9)	-0.02 (0.00)	0.40 (0.12)	0.040 (0.006)	3.53 (0.22)	179 (9)	49.1 (14.5)	
s10	cont	1	2	-67.1 (36.8)	-17.2 (29.4)	18.0 (44.3)	-0.03 (0.01)	1.62 (0.76)	0.018 (0.002)	4.44 (0.09)	228 (10)	66.4 (13.2)	
	cont	2	4	-88.6 (22.7)	-50.8 (11.7)	-26.8 (16.0)	-0.02 (0.02)	0.87 (0.56)	0.032 (0.011)	4.05 (0.46)	237 (39)	62.1 (11.7)	
	corr	1	3	-46.5 (115.2)	15.6 (69.4)	60.5 (117.9)	-0.04 (0.03)	1.60 (1.42)	0.029 (0.012)	4.49 (0.83)	250 (79)	-16.1 (83.4)	
	corr	2	2	-59.5 (48.0)	-45.8 (49.2)	-32.4 (44.6)	-0.02 (0.00)	0.42 (0.28)	0.050 (0.024)	3.29 (0.13)	236 (35)	28.0 (11.7)	
s11	cont	1	0										
	cont	2	2	-55.5 (72.1)	-36.9 (63.1)	-3.1 (46.7)	-0.02 (0.00)	0.91 (0.27)	0.027 (0.007)	3.89 (0.50)	296 (50)	42.1 (7.7)	
	corr	1	2	30.8 (79.3)	50.4 (63.5)	63.2 (62.8)	-0.06 (0.05)	1.44 (1.68)	0.053 (0.024)	3.41 (0.53)	324 (32)	39.9 (10.5)	
	corr	2	3	-38.6 (18.3)	-11.5 (16.2)	4.1 (15.5)	-0.01 (0.00)	0.33 (0.04)	0.040 (0.005)	3.75 (0.08)	267 (45)	56.9 (24.9)	
s13	cont	1	3	11.3 (93.4)	61.8 (50.8)	76.3 (55.2)	-0.05 (0.03)	1.76 (1.75)	0.031 (0.009)	4.04 (0.64)	229 (11)	46.0 (6.7)	
	cont	2	4	-40.6 (27.2)	-17.3 (22.4)	-0.8 (21.9)	-0.03 (0.01)	0.87 (0.30)	0.033 (0.006)	3.68 (0.14)	227 (26)	54.3 (22.1)	
	corr	1	4	-47.2 (101.7)	6.6 (92.7)	31.4 (83.7)	-0.03 (0.01)	1.62 (0.75)	0.019 (0.007)	4.30 (0.40)	224 (12)	45.9 (25.4)	
	corr	2	5	-61.3 (24.8)	-38.0 (26.2)	-18.4 (28.0)	-0.03 (0.01)	0.94 (0.22)	0.032 (0.005)	3.73 (0.24)	228 (21)	52.5 (21.1)	
s14	cont	1	4	-54.0 (75.0)	-26.8 (80.8)	11.0 (78.0)	-0.03 (0.01)	1.38 (0.24)	0.024 (0.010)	4.17 (0.07)	153 (20)	32.1 (28.8)	
	cont	2	2	-83.9 (87.1)	-49.2 (77.7)	-25.1 (76.1)	-0.03 (0.00)	1.06 (0.06)	0.028 (0.001)	4.07 (0.19)	159 (8)	45.9 (13.7)	
	corr	1	10	-39.8 (72.5)	26.0 (48.6)	75.2 (53.5)	-0.03 (0.01)	2.00 (0.59)	0.015 (0.003)	4.68 (0.39)	163 (8)	-6.1 (41.2)	
	corr	2	4	-62.1 (71.6)	-45.9 (63.0)	-26.5 (60.3)	-0.02 (0.01)	0.61 (0.27)	0.038 (0.012)	3.49 (0.48)	139 (17)	58.0 (24.0)	
s15	cont	1	0										
	cont	2	1	-18.1	9.8	34.9	-0.02	0.65	0.026	3.97	240	37.4	
	corr	1	2	-113.7 (36.9)	-82.3 (21.4)	-54.2 (39.3)	-0.05 (0.02)	2.90 (1.36)	0.018 (0.001)	4.09 (0.04)	273 (16)	58.4 (7.5)	
	corr	2	2	5.5 (33.6)	25.7 (29.3)	39.3 (32.0)	-0.01 (0.00)	0.44 (0.14)	0.033 (0.003)	3.52 (0.05)	298 (3)	63.4 (51.2)	
s17	cont	1	1	-63.6	5.5	18.7	-0.02	1.07	0.021	4.41	135	46.7	
	cont	2	1	-49.4	1.2	27.9	-0.02	0.91	0.025	4.35	145	28.8	
	corr	1	4	-22.7 (97.7)	12.4 (89.3)	80.1 (121.3)	-0.02 (0.02)	1.48 (0.97)	0.021 (0.016)	4.54 (0.51)	134 (9)	35.9 (21.3)	
	corr	2	0										
s18	cont	1	1	-25.1	-11.5	-2.0	-0.07	1.91	0.036	3.14	160	19.5	
	cont	2	0										
	corr	1	0										
	corr	2	4	-25.1 (31.3)	-9.0 (30.6)	5.1 (28.6)	-0.02 (0.01)	0.46 (0.28)	0.040 (0.006)	3.40 (0.13)	132 (10)	28.6 (14.7)	
s19	cont	1	1	60.4	80.6	91.5	0.00	0.07	0.069	3.44	145	53.3	
	cont	2	1	-26.3	-13.4	6.9	-0.03	0.71	0.045	3.50	171	42.7	

01 October 2024 12:06:22

TABLE III. (Continued)

Part.	Foc.	Clus.	Count	Mean (SD) for feature									
				onset.lag	velmin.lag	apex.lag	velmin	gest.range	Stiffness	log.dur	f0max	f0max.norm	
	corr	1	3	-14.2 (103.8)	44.4 (67.9)	135.4 (39.4)	-0.01 (0.01)	1.59 (1.27)	0.013 (0.011)	4.87 (0.63)	134 (9)	28.2 (24)	
	corr	2	1	-39.4	-16.6	-0.7	-0.02	0.60	0.028	3.65	149	35.6	
Overall				-28.8 (66.1)	4.5 (60.3)	34.5 (66.1)	-0.03 (0.02)	1.22 (0.84)	0.029 (0.013)	3.99 (0.54)	216 (63)	28.1 (49.3)	

^aCounts, means (where applicable), and SDs (where applicable) for the nine features used in the study are separated row-wise by the unique combination of participant (Part.), focus condition (Foc.), and cluster (Clus.) identified by the non-parametric Gaussian mixture model. corr, corrective; cont, contrastive. SDs are shown in parentheses for counts greater than 1. The overall means and standard SDs across all 116 observations are displayed in the bottom row.

with an appropriate degree of circumspection. It is our hope that, at the very least, the results presented here may serve as important catalysts to help guide future research questions and goals. Second, the data presented in this paper come from only 63% of the participants that were originally recorded (12 of 19), and, therefore, the resulting 116 head nod gestures that were included in the final analysis are far fewer than we originally intended, thus reducing the statistical power and, potentially, the generalizability of our results. Future research would benefit from appropriate efforts to avoid the issues that necessitated the exclusion of so many of the subjects that were originally recorded. Finally, the head nod metric used in the current study characterizes the angle of the head *in relation to objective space*. This means that any factor that can modulate this head angle will necessarily affect the measure. For the results and interpretations in the current study, we have assumed that the only relevant factor is the independent movement of the head. However, we acknowledge that other factors, such as movement of the body torso/trunk, will also affect the angle of the head and, thus, the resulting value of our head nod metric. Without independent data relating to rigid body kinematics of the torso, we cannot tease apart these two factors. Future research on co-speech head movements would benefit from tracking torso movement in addition to head movement, in order to address this problem.

V. CONCLUSION

In this exploratory study, we examined how the production of vertical head nods may be temporally and kinematically related to different levels of pitch prominence connected to difference focus conditions in French, using a time-varying head nod metric generated from electromagnetic articulometry sensors placed at three stable locations on the head. Across all of the data collected from the 12 speakers included in this study, a total of 116 co-speech downward head nod gestures were identified. Since previous research has shown that co-speech gesturing may be aligned with prosodic events in both time and magnitude, we used a combination of knowledge-based (top-down) and data-driven (bottom-up) approaches to determine whether and how alignment might occur between co-speech head nods and pitch prominence and whether and how this alignment

might be modulated by different levels of prosodic prominence.

Our results show that two different gesture strategies underlie the distribution of the nine kinematic and acoustic measurements of our data. However, these two gesture strategies are not differentiated explicitly by the two levels of prosodic prominence investigated here (i.e., contrastive and corrective focus). Rather than motivating the distinction between the two gestural strategies, prosodic prominence serves to magnify their inherent difference: The kinematic characteristics of the two strategies are consistent under both focus conditions, but the difference between the two strategies is enhanced when the prosodic prominence is also enhanced (i.e., under corrective focus), thereby suggesting that co-speech head nods may be used as a means of prosodic enhancement in French. Although we observed some evidence of speaker-specific preferences for the use of the two strategies, the variation is spread evenly across speakers, rather than clustering into two groups of speakers (i.e., those who predominantly use the first strategy and those who predominantly use the second). These results suggest that the alignment of co-speech gestures with speech itself, and the intra-language variability of this alignment across speakers, may be more nuanced than has previously been assumed.

ACKNOWLEDGMENTS

The authors would like to thank the associate editor and two anonymous reviewers for their comments and suggestions.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to declare.

Ethics Approval

The study was approved by the ethics committee of the University Aix-Marseille.

DATA AVAILABILITY

The data that support the findings of this study (i.e., the pre-processed data and R code used to recreate the results of this study) are openly available in Open Science Framework (OSF) at <https://osf.io/nj3rv/>.

APPENDIX

The counts, means, and standard deviations for the nine features examined in this study are shown in Table III.

¹Although this relationship may often be taken for granted, the linguistic function of co-speech gesturing is not necessarily uncontroversial. Krauss *et al.* (1991), for example, argue that these gestures are not informative in and of themselves and that any meaning they may happen to convey is largely redundant with speech, since semantic judgments of co-speech gestures are determined principally by the speech that the gestures accompany rather than the gestures themselves.

²See Pouw and Fuchs (2022) for a more complete overview of the biomechanical interactions that may be involved in co-speech gesticulation, referred to by the authors as “vocal-entangled gesture.”

³Two additional head nods were identified in broad focus productions, equating to 1.7% of the total number of identified head nods (118). However, given the research goal of comparing head nods produced under two different levels of prosodic prominence, only the 116 head nods produced under the two focus conditions are considered in this study, as mentioned in Sec. II A 1.

Alexanderson, S., House, D., and Beskow, J. (2013a). “Aspects of co-occurring syllables and head nods in spontaneous dialogue,” in *Proceedings of 12th International Conference on Auditory-Visual Speech Processing (AVSP2013)*, Annecy, France (ISCA, Stockholm), pp. 169–172.

Alexanderson, S., House, D., and Beskow, J. (2013b). “Extracting and analyzing head movements accompanying spontaneous dialogue,” in *Conference Proceedings of TiGeR: Tilburg Gesture Research Meeting: 10th International Gesture Workshop (GW) and 3rd Gesture and Speech in Interaction (GESPIN) Conference*, Tilburg University, Tilburg, Netherlands.

Ambrazaitis, G., and House, D. (2017). “Multimodal prominences: Exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings,” *Speech Commun.* **95**, 100–113.

Ambrazaitis, G., and House, D. (2022). “Probing effects of lexical prosody on speech-gesture integration in prominence production by Swedish news presenters,” *Lab. Phonol.* **24**(1).

Anegawa, E., Tsuyama, H., and Kusukawa, J. (2008). “Lateral cephalometric analysis of the pharyngeal airway space affected by head posture,” *Int. J. Oral Maxillofacial Surg.* **37**(9), 805–809.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., and Krivitsky, P. N. (2023). “lme4: Linear mixed-effects models using ‘Eigen’ and S4 [computer program],” <https://cran.r-project.org/package=lme4> (Last viewed June 7, 2023).

Bergmann, K., Aksu, V., and Kopp, S. (2011). “The relation of speech and gestures: Temporal synchrony follows semantic synchrony,” in *Proceedings of the 2nd Workshop on Gesture and Speech in Interaction (GeSpIn)*, Bielefeld, Germany, pp. 1–6.

Boersma, P., and Weenink, D. (2021). “Praat: Doing phonetics by computer [computer program],” <http://www.praat.org/> (Last viewed January 23, 2023).

Bosker, H. R., and Peeters, D. (2021). “Beat gestures influence which speech sounds you hear,” *Proc. R. Soc. B* **288**(1943), 20202419.

Calhoun, S. (2010). “The centrality of metrical structure in signaling information structure,” *Language* **86**(1), 1–42.

Cho, T. (2005). “Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a,i/ in English,” *J. Acoust. Soc. Am.* **117**(6), 3867–3878.

Cho, T., Kim, D. J., and Kim, S. (2019). “Prosodic strengthening in reference to the lexical pitch accent system in South Kyungsang Korean,” *Linguist. Rev.* **36**(1), 85–115.

Chu, M., and Hagoort, P. (2014). “Synchronization of speech and gesture: Evidence for interaction in action,” *J. Exp. Psychol. Gen.* **143**(4), 1726–1741.

Condon, W. S. (1976). “An analysis of behavioral organization,” *Sign Lang. Stud.* **13**, 285–318.

Debrelioska, S., Özyürek, A., Gullberg, M., and Perniss, P. (2013). “Gestural viewpoint signals referent accessibility,” *Discourse Process.* **50**(7), 431–456.

D’Imperio, M., Espesser, R., Loevenbruck, H., Menezes, C., Nguyen, N., and Welby, P. (2007). “Are tones aligned with articulatory events? Evidence from Italian and French,” *Pap. Lab. Phonol.* **9**, 577–608.

Esteve-Gibert, N., Borrás-Comes, J., Asor, E., Swerts, M., and Prieto, P. (2017). “The timing of head movements: The role of prosodic heads and edges,” *J. Acoust. Soc. Am.* **141**(6), 4727–4739.

Esteve-Gibert, N., and Guellai, B. (2018). “Prosody in the auditory and visual domains: A developmental perspective,” *Front. Psychol.* **9**, 338.

Esteve-Gibert, N., Loevenbruck, H., Dohen, M., and D’Imperio, M. (2022). “Pre-schoolers use head gestures rather than prosodic cues to highlight important information in speech,” *Dev. Sci.* **25**(1), e13154.

Esteve-Gibert, N., and Prieto, P. (2013). “Prosodic structure shapes the temporal realization of intonation and manual gesture movements,” *J. Speech Lang. Hear. Res.* **56**(3), 850–864.

Ferré, G. (2014). “A multimodal approach to markedness in spoken French,” *Speech Commun.* **57**, 268–282.

Fraley, C., Raftery, A. E., Scrucca, L., Murphy, T. B., and Fop, M. (2022). “mclust: Gaussian mixture modelling for model-based clustering, classification, and density estimation [computer program],” available from <https://cran.r-project.org/package=mclust> (Last viewed January 15, 2023).

Fuchs, S., and Rochet-Capellan, A. (2021). “The respiratory foundations of spoken language,” *Annu. Rev. Linguist.* **7**, 13–30.

Fung, H. S. H., and Mok, P. P. K. (2018). “Temporal coordination between focus prosody and pointing gestures in Cantonese,” *J. Phon.* **71**, 113–125.

Goldin-Meadow, S. (1999). “The role of gesture in communication and thinking,” *Trends Cogn. Sci.* **3**(11), 419–429.

Hadar, U., Steiner, T. J., Grant, E. C., and Rose, F. C. (1983). “Head movement correlates of juncture and stress at sentence level,” *Lang. Speech* **26**(2), 117–129.

Hadar, U., Steiner, T. J., Grant, E. C., and Rose, F. C. (1984). “The timing of shifts in head posture during conversation,” *Hum. Mov. Sci.* **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 the 7th European Conference on Speech Communication and Technology (Eurospeech 2001)*, Aalborg, Denmark (ISCA, Stockholm), pp. 387–390.

Ishi, C. T., Haas, J., Wilbers, F. P., Ishiguro, H., and Hagita, N. (2007). “Analysis of head motions and speech, and head motion control in an android,” in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Diego, CA (IEEE, New York), pp. 548–553.

Jannedy, S., and Mendoza-Denton, N. (2005). “Structuring information through gesture and intonation,” *Interdiscip. Stud. Inf. Struct.* **3**, 199–244.

Kadává, S., Cwiek, A., Stoltmann, K., Fuchs, S., and Pouw, W. (2023). “Is gesture-speech physics at work in rhythmic pointing? Evidence from Polish counting-out rhymes,” in *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS 2023)*, Prague, Czech Republic, edited by R. Skarnitzl and J. Volín (GUARANT International, Prague, Czech Republic), pp. 4190–4194.

Kelso, J., Tuller, B., and Harris, K. (1983). “A ‘dynamic pattern’ perspective on the control and coordination of movement,” in *The Production of Speech*, edited by P. MacNeilage (Springer-Verlag, Berlin), pp. 137–173.

Kendon, A. (1980). “Gesticulation and speech: Two aspects of the process of utterance,” in *The Relationship of Verbal and Nonverbal Communication* (De Gruyter Mouton, Berlin), pp. 207–228.

Kim, J., Cvejic, E., and Davis, C. (2014). “Tracking eyebrows and head gestures associated with spoken prosody,” *Speech Commun.* **57**, 317–330.

Kita, S. (2000). “How representational gestures help speaking,” in *Language and Gesture*, edited by D. McNeill (Cambridge University Press, Cambridge, UK), pp. 162–185.

Klein, W., and Codd, J. (2010). “Breathing and locomotion: Comparative anatomy, morphology and function,” *Respir. Physiol. Neurobiol.* **173**, S26–S32.

Krahmer, E., and 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.

- Krauss, R., Morrel-Samuels, P., and Colasante, C. (1991). "Do conversational hand gestures communicate?," *J. Pers. Social Psychol.* **61**, 743–754.
- Krivokapić, J. (2014). "Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes," *Phil. Trans. R. Soc. B* **369**(1658), 20130397.
- Krivokapić, J., Tiede, M. K., and Tyrone, M. E. (2017). "A kinematic study of prosodic structure in articulatory and manual gestures: Results from a novel method of data collection," *Lab. Phonol.* **8**(1), 3.
- Language Archive (2023). "ELAN (Version 6.7) [computer program]," <https://archive.mpi.nl/elan> (Last viewed January 3, 2023).
- Lenth, R. V., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., and Singmann, H. (2023). "emmeans: Estimated marginal means, aka least-squares means [computer program]," <https://cran.r-project.org/package=emmeans> (Last viewed June 7, 2023).
- Leonard, T., and Cummins, F. (2011). "The temporal relation between beat gestures and speech," *Lang. Cogn. Process.* **26**(10), 1457–1471.
- Levin, S. (1997). "Putting the shoulder to the wheel: A new biomechanical model for the shoulder girdle," *Biomed. Sci. Instrum.* **33**, 412–417.
- Levin, S. (2006). "Tensegrity: The new biomechanics," in *Textbook of Musculoskeletal Medicine*, edited by M. Hutson and R. Ellis (Oxford University Press, Oxford, UK), pp. 69–80.
- Loehr, D. (2007). "Aspects of rhythm in gesture and speech," *Gesture* **7**(2), 179–214.
- Maynard, S. K. (1989). *Japanese Conversation: Self-Contextualization through Structure and Interactional Management* (Ablex Publishing, Norwood, NJ).
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought* (University of Chicago Press, Chicago).
- Miller, N., Gregory, J., Semple, S., Aspden, R., Stollery, P., and Gilbert, F. (2012). "Relationships between vocal structures, the airway, and cranio-cervical posture investigated using magnetic resonance imaging," *J. Voice* **26**(1), 102–109.
- Moisik, S. R., Yun, D. P. Z., and Dediu, D. (2019). "Active adjustment of the cervical spine during pitch production compensates for shape: The ArtiVarK study," in *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS 2019)*, Canberra, Australia, edited by S. Calhoun, P. Escudero, M. Tabain, and P. Warren (Australasian Speech Science and Technology Association Inc.), pp. 864–868.
- Molnár, V. (2006). "On different kinds of contrast," in *The Architecture of Focus*, edited by V. Molnár and S. Winkler (Mouton de Gruyter, Berlin), pp. 197–234.
- Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., and Vatikiotis-Bateson, E. (2004). "Visual prosody and speech intelligibility: Head movement improves auditory speech perception," *Psychol. Sci.* **15**(2), 133–137.
- Pagel, L., Sóskuthy, M., Roessig, S., and Mücke, D. (2023). "A kinematic analysis of visual prosody: Head movements in habitual and loud speech," in *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS 2023)*, Prague, Czech Republic, edited by R. Skarnitzl and J. Volín (GUARANT International, Prague, Czech Republic).
- Paoli, S. (2009). "Contrastiveness and new information: A new view on focus," *Riv. Grammatica Generativa* **34**, 137–161.
- Parrell, B., Goldstein, L., Lee, S., and Byrd, D. (2014). "Spatiotemporal coupling between speech and manual motor actions," *J. Phon.* **42**, 1–11.
- Pouw, W., Burchardt, L. S., and Selen, L. (2023). "Postural and muscular effects of upper-limb movements on voicing," bioRxiv, <https://www.biorxiv.org/content/10.1101/2023.03.08.531710v1>.
- Pouw, W., and Dixon, J. (2019). "Entrainment and modulation of gesture–speech synchrony under delayed auditory feedback," *Cogn. Sci.* **43**(3), e12721.
- Pouw, W., and Fuchs, S. (2022). "Origins of vocal-entangled gesture," *Neurosci. Biobehav. Rev.* **141**, 104836.
- Pouw, W., Harrison, S., and Dixon, J. (2020). "Gesture-speech physics: The biomechanical basis for the emergence of gesture-speech synchrony," *J. Exp. Psychol. Gen.* **149**(2), 391–404.
- Pouw, W., Harrison, S., and Dixon, J. A. (2022). "The importance of visual control and biomechanics in the regulation of gesture-speech synchrony for an individual deprived of proprioceptive feedback of body position," *Sci. Rep.* **12**(1), 14775.
- Renwick, M. E. L., Shattuck-Hufnagel, S., and Yasinnik, Y. (2004). "The timing of speech-accompanying gestures with respect to prosody," *J. Acoust. Soc. Am.* **115**, 2397.
- Repp, S. (2016). "Contrast: Dissecting an elusive information-structural notion and its role in grammar," in *The Oxford Handbook of Information Structure* (Oxford University Press, Oxford, UK).
- Rime, B. (1982). "The elimination of visible behavior from social interactions: Effects on verbal, nonverbal and interpersonal variables," *Euro. J. Soc. Psychol.* **12**, 113–129.
- Rochet-Capellan, A., Laboissière, R., Galván, A., and Schwartz, J.-L. (2008). "The speech focus position effect on jaw–finger coordination in a pointing task," *J. Speech Lang. Hear. Res.* **51**(6), 1507–1521.
- Rohrer, P. (2022). "A temporal and pragmatic analysis of gesture-speech association: A corpus-based approach using the novel multimodal multi-dimensional (M3D) labeling system," Ph.D. thesis, Universitat Pompeu Fabra, Barcelona, Spain.
- Roustan, B., and Dohen, M. (2010). "Gesture and speech coordination: The influence of the relationship between manual gesture and speech," in *Proceedings of Interspeech 2010—11th Annual Conference of the International Speech Communication Association*, Chiba, Japan (International Speech Communication Association).
- Schmid, M., Conforto, S., Bibbo, D., and D'Alessio, T. (2004). "Respiration and postural sway: Detection of phase synchronizations and interactions," *Hum. Mov. Sci.* **23**(2), 105–119.
- Serré, H., Dohen, M., Fuchs, S., Gerber, S., and Rochet-Capellan, A. (2022). "Leg movements affect speech intensity," *J. Neurophysiol.* **128**(5), 1106–1116.
- Shattuck-Hufnagel, S., and Ren, A. (2018). "The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech," *Front. Psychol.* **9**, 1514.
- Stoltmann, K., and Fuchs, S. (2017). "Syllable-pointing gesture coordination in Polish counting out rhymes: The effect of speech rate," *J. Multimodal Commun. Stud.* **4**(1–2), 63–68.
- Swerts, M., and Kraemer, E. (2010). "Visual prosody of newscasters: Effects of information structure, emotional content and intended audience on facial expressions," *J. Phon.* **38**(2), 197–206.
- Tiede, M., Mooshammer, C., and Goldstein, L. (2019). "Noggin nodding: Head movement correlates with increased effort in accelerating speech production tasks," *Front. Psychol.* **10**, 2459.
- Türk, O., and Calhoun, S. (2023). "Phrasal synchronization of gesture with prosody and information structure," *Lang. Speech* **67**(3), 702–743.
- Wagner, P., Malisz, Z., and Kopp, S. (2014). "Gesture and speech in interaction: An overview," *Speech Commun.* **57**, 209–232.
- Yehia, H. C., Kuratate, T., and Vatikiotis-Bateson, E. (2002). "Linking facial animation, head motion and speech acoustics," *J. Phon.* **30**(3), 555–568.