Pre-schoolers use head gestures rather than prosodic cues to highlight important information in speech

Núria Esteve-gibert, Hélène Loevenbruck, Marion Dohen, Mariapaola d'Imperio

▶ **To cite this version:**

Núria Esteve-gibert, Hélène Loevenbruck, Marion Dohen, Mariapaola d'Imperio. Pre-schoolers use head gestures rather than prosodic cues to highlight important information in speech. Developmental Science, 2022, 25 (1), pp.e13154. 10.1111/desc.13154 . hal-03348546

HAL Id: hal-03348546
https://hal.univ-grenoble-alpes.fr/hal-03348546

Submitted on 20 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Check for updates

Received: 24 July 2020 | Revised: 21 June 2021 | Accepted: 22 June 2021

DOI: 10.1111/desc.13154

Developmental Science WILEY

PAPER

Pre-schoolers use head gestures rather than prosodic cues to highlight important information in speech

Núria Esteve-Gibert[1,2] | Hélène Lœvenbruck[3] | Marion Dohen[4] | Mariapaola D'Imperio[5,2]

[1] Faculty of Psychology and Educational Sciences, Universitat Oberta de Catalunya, Barcelona, Spain

[2] Laboratoire Parole et Langage, UMR CNRS 7309, Aix Marseille Université, Marseille, France

[3] Laboratoire de Psychologie et Neurocognition, UMR CNRS 5105, Université Grenoble Alpes, Grenoble, France

[4] GIPSA-lab, UMR CNRS 5216, Université Grenoble Alpes, Grenoble, France

[5] Department of Linguistics, Rutgers University, New Jersey, USA

Correspondence
Núria Esteve-Gibert, Universitat Oberta de Catalunya, Rambla del Poblenou, 156, Barcelona, 08018 Spain.
Email: nesteveg@uoc.edu

Abstract

Previous evidence suggests that children's mastery of prosodic modulations to signal the informational status of discourse referents emerges quite late in development. In the present study, we investigate the children's use of head gestures as it compares to prosodic cues to signal a referent as being contrastive relative to a set of possible alternatives. A group of French-speaking pre-schoolers were audio-visually recorded while playing in a semi-spontaneous but controlled production task, to elicit target words in the context of broad focus, contrastive focus, or corrective focus utterances. We analysed the acoustic features of the target words (syllable duration and word-level pitch range), as well as the head gesture features accompanying these target words (head gesture type, alignment patterns with speech). We found that children's production of head gestures, but not their use of either syllable duration or word-level pitch range, was affected by focus condition. Children mostly aligned head gestures with relevant speech units, especially when the target word was in phrase-final position. Moreover, the presence of a head gesture was linked to greater syllable duration patterns in all focus conditions. Our results show that (a) 4- and 5-year-old French-speaking children use head gestures rather than prosodic cues to mark the informational status of discourse referents, (b) the use of head gestures may gradually entrain the production of adult-like prosodic features, and that (c) head gestures with no referential relation with speech may serve a linguistic structuring function in communication, at least during language development.

KEYWORDS

contrastive focus, French, head gestures, information structure, language acquisition, non-referential gestures, prosody development

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2021 The Authors. *Developmental Science* published by John Wiley & Sons Ltd

1 | INTRODUCTION

Imagine you and your friend go shopping and are deciding which clothes to buy. If your friend says to you 'Buy the BLUE jacket', with prosodic prominence on the word 'blue', you will understand 'buy the blue jacket, but not the pink one'. Putting an emphatic accent on the color adjective signals that there is a choice among other possible colors for that item. Speakers can signal to listeners which discourse referent is new (or focused), and hence needs to be added to the previously shared common ground, among a set of contrastive alternatives in that context (Rooth, 1992; Vallduví, 1991). In the present study, we will investigate young children's use of prosodic and gesture strategies to mark focus information in speech.

Information in speech can be focused with different degrees of prominence. When no specific element is emphasized because the entire utterance needs to be added to the common ground, we talk about all-focus or broad focus condition. When only one specific element in the utterance is emphasized because it needs to be chosen among a set of alternatives in order to be added to the common ground, we talk about a contrastive focus situation. Finally, when one specific element might be even more strongly emphasized because it has to replace a preceding element that was wrongly added to the shared common ground, we then talk about a corrective focus situation (Krifka, 2008). Speakers use distinct prosodic strategies to signal different degrees of prominence. In general, in broad focus situations speakers use prosodic cues that are 'unmarked' or less salient; instead, in contrastive focus situations speakers use 'marked' or more salient (i.e., prominent) prosodic cues.

Prosodic strategies to mark focus can have a phonetic or a phonological nature, and are language-specific. Although the distinction between phonetics and phonology is not always clear-cut, it is generally assumed that phonetic modulations of prosody refer to gradient variations in the acoustic features of speech that do not imply categorical changes in mental representations, whereas phonological modulations refer to prosodic variations that induce categorical shifts. In English or German, pitch accents are phonological devices marking focus (Baumann & Grice, 2006; Pierrehumbert & Hirshberg, 1990), whereas in Mandarin Chinese, for instance, speakers use phonetic cues such as duration or pitch range. French is an interesting case because acoustic cues can be employed phonologically to mark the focused element: speakers can use longer syllable durations and wider pitch expansions (Dohen & Lœvenbruck, 2004; Féry, 2001), the insertion of a break between the focused element and the preceding sequence (Jun & Fougeron, 2000; Michelas & D'Imperio, 2015), and even an initial accent or 'initial (intonation) rise' (German & D'Imperio, 2015).

Previous developmental research suggests that young children aged 2 to 5 can use phonetic (but not phonological) prosodic modulations to signal contrastive focus (see Chen, 2018, for a review). Young 4- and 5-year-old Mandarin-speaking children can successfully use the expected adult-like phonetic cues to mark focus (Yang & Chen, 2018); Dutch-speaking 4- and 5-year-olds can only use phonetic (non-adult-like) cues (Romoren & Chen, 2015), and when they turn 7–8 year of age children start using the expected adult-like phonological cues (Chen, 2018).

RESEARCH HIGHLIGHTS

- Young children use non-referential head gestures, rather than acoustic prosodic cues, to signal the informational status of discourse referents
- Children produce more head gestures in contrastive focus than in broad focus conditions, and even more in corrective focus conditions
- Young children timely align head gestures with prosodic landmarks, with some unexpected exceptions
- When children produce a head gesture, this seems to entrain distinct prosodic (duration and pitch range) modifications of the accompanying speech, independent of focus condition

To our knowledge, French-speaking children's acquisition of prosodic focus has been mainly studied at the perception and comprehension levels, but rarely in terms of production skills. Rapin and Ménard (2019) showed that 8- to 10-year-old children are able to detect focus, but (a) their performance is lower than that of adults, and (b) they do not use formant or visual articulatory cues as much as adults do. Szendroi et al. (2018) studied focus comprehension in 3-, 4-, and 5-year-old French children (comparing them to German and English), and found that French pre-schoolers show adult-like comprehension of subject and object contrastive focus. On the production side, Ménard et al. (2006) showed that 4-year-old French children use variations in intensity, formant, and articulation values to mark focus, and only 8-year-old children adopt adult-like acoustic and articulatory labial strategies. Altogether these results suggest that, given that adult French speakers use phonetic cues in a phonological way to mark focus, a late acquisition of adult-like prosodic cues to focus can be expected in French.

Body movements are another important strategy that speakers use to highlight specific discourse elements. We move our hands (manual gestures), our heads (head gestures), and our facial expressions in temporal and functional synchrony with our speech. From a timing point of view, prosodic events in speech serve as anchoring landmarks for gestural alignment, as points of maximal displacement in gestures usually occur within the temporal limits of prominent words or syllables in speech (e.g., Carignan et al., 2020; Esteve-Gibert & Prieto, 2013; Esteve-Gibert et al., 2017; Krivokapic et al., 2017; Leonard & Cummins, 2011). From a functional point of view, gestures convey meanings that can be referential (representing an entity or event deictically, iconically, or metaphorically) and non-referential (signaling information structure, modal information, or discourse cohesion) (Kendon, 1980; McNeill, 2000; and many others thereafter). In the context of focus marking, head gestures are a specific type of body movements that serve a non-referential meaning and can indicate focus quite consistently (Ambrazaitis & House, 2017; Esteve-Gibert et al., 2017; Ishi et al., 2014), together with other movements such as eyebrow raising

(Cavé et al., 1996; Dohen et al., 2006; Moubayed & Beskow, 2011) and manual beats (Roustan & Dohen, 2010).

Body movements are relevant at all stages of language acquisition. The acquisition of co-speech body movements with a non-referential meaning occurs much later in development and has been much less studied that the development of gestures with a referential meaning. Non-referential body movements are produced along with speech, marking rhythmic prominence in speech, structuring discourse information, or signalling socio-pragmatic intent, while not semantically referring to any entity, action, nor any object (McNeill, 1992). They have been typically referred to as 'beat' gestures in the gesture literature because they hold a close rhythmic relation with speech (Kendon, 1980; McNeill, 1992; and many others thereafter), a category that has typically only included manual gestures. However, there is a growing body of evidence showing that gestures with a non-referential value may be produced using different body articulators such as the eyebrows or the head (see discussions in Shattuck-Hufnagel & Prieto, 2019; Shattuck-Hufnagel & Ren, 2018), and in the present study we explore this broad definition of non-referential gestures.

Non-referential body movements marking discourse structure seem to emerge around 4–6 years of age, though only manual gestures have been extensively studied (Colletta et al., 2014; Mathew et al., 2017; Nicoladis et al., 1999; Vilà-Giménez & Prieto, 2020). Mathew et al. (2017) identified manual non-referential gestures co-occurring with pitch accents in narrative productions of 5- to 7-year-old Australian English-speaking children. Since pitch accents are by definition prosodically prominent, it is plausible to infer that their coding was tapping into manual gestures that were intended to signal focus. Their results, though, indicated that English-speaking children start producing these manual gestures at age 6 and that only some (but not all) of them co-occurred with a pitch accent. This is intriguing because (a) pitch accents are expected to serve as the anchoring landmarks for the temporal alignment of body movements (as revealed by the adult literature) and (b) even young infants can time-align pointing gestures with prosodically prominent syllables in speech (Esteve-Gibert & Prieto, 2014). The authors hence suggest that, at this age, children might be good narrators even if they do not yet master the temporal alignment of non-referential manual gestures with prosodic cues to express the informational status of discourse referents. However, most of these previous studies do not indicate which proportion of children's non-referential gestures mark focus, and they have not taken into account non-referential gestures produced with the head.

Our main goal here is to investigate the interplay between non-referential head gestures and prosody to signal the informational status of discourse referents in pre-school children. More specifically, we aim to examine (1) whether French-acquiring children at age 4–5 can use head gestures and prosody (syllable duration, F0 range) to signal that a referent is contrastively focused, and, additionally, to signal the degree of contrast (i.e., broad focus vs. contrastive focus vs. corrective focus), (2) whether head gestures and prosody interact with each other for the purpose of implementing different degrees of focus, and (3) whether age and linguistic (expressive and receptive) abilities influence children's use of head gestures and prosodic cues for focus marking. Most previous studies on the early production of contrastive focus have used repetition (Chen, 2009, 2011; Romoren & Chen, 2015) or narrative tasks (Colletta et al., 2014; Mathew et al., 2017). Instead, we used an elicitation procedure in a controlled experimental setting to obtain productions that were spontaneous and still controlled in terms of the size of the focused element (number of syllables) and its position within the phrase. These two factors (size and position of the focused element) are crucial in French because they influence how focused elements are prosodically marked in the utterance: (a) longer phrases in French are more likely to elicit left-edge tonal movements (German & D'Imperio, 2015), and (b) pre-boundary lengthening occurs on the last syllable of phrase-final words in French independent of information structure and focus status (Jun & Fougeron, 2000).

First, if young children use non-referential gestures to signal the informational status of discourse referents, we expect more head gestures to mark higher degrees of prominence (i.e., more head gestures in the corrective focus condition than in the contrastive focus condition, and more head gestures in the contrastive focus condition than in the broad focus condition). Second, if young children also use prosodic strategies to mark focus, we expect them to use wider F0 range values on the focused elements, and to lengthen both the initial and final syllable of the focused elements (the stronger the contrast, the longer the syllables). Given that in French syllables immediately preceding a boundary are usually lengthened, we expect final syllables of the focused elements to be even longer than those preceding a boundary without focus marking. Third, if gesture and prosodic strategies interact to signal the information status of discourse referents, we expect that the presence of a gesture accompanying an element within the target phrase would imply a variation in the prosodic features of that element: elements marked with a head gesture will also be marked with prosodic cues to focus such as longer syllables and wider F0 range. Finally, we expect older children to produce more gestural and prosodic cues to implement focus relative to younger children, and that higher linguistic abilities would imply greater use of gestural and prosodic cues to mark focus.

2 | METHODS

2.1 | Participants

A total of 24 4- and 5-year-old children participated in our study (mean age: 56 months; age range: 50–67 months; nine boys). Two additional children were tested but excluded from the final sample (one due to colour-blindness issues that could affect the results of the task, and the other one due to fussiness). All children spoke French as their primary language at home and at school, and all parents reported that their child had no hearing problems. They were recruited from a child database at the Babylab at Grenoble University and tested in a lab room at the GIPSA-lab of the same university. Parents gave previous written consent for the participation of their children and received either a book or a 10€ voucher as compensation. The procedure was also orally explained to the children and they gave a spoken consent.

TABLE 1 Examples of sentences in each experimental condition

Focus condition	Position of focalised element	Example
Broad-focus	None	*Prends la valise orange* Take the suitcase orange
Contrastive focus	Non-phrase-final (Noun)	*Prends la VALISE orange* Take the SUITCASE orange
	Phrase-final (Adjective)	*Prends la valise ORANGE* Take the suitcase ORANGE
Corrective focus	Non-phrase-final (Noun)	*Prends la **VALISE** orange* Take the **SUITCASE** orange
	Phrase-final (Adjective)	*Prends la valise **ORANGE*** Take the suitcase **ORANGE**

Capital letters indicate contrastive focus, bold letters indicate corrective focus.

FIGURE 1 Example of a visual display in one experimental trial. Children were prompted to tell the little girl, Claire, that she had to take the object depicted on the top right corner of the screen from inside the bag in order for her to get to see the turtle's neck. On the left, visual display before Claire was instructed by the child to act. On the right, visual display after Claire understood the instruction

2.2 | Materials

Children were prompted to produce sentences containing Noun Phrases (NP) with the following structure: Article + Noun + Adjective (e.g. *Prends [la valise orange]*$_{NP}$ 'Take [the orange suitcase]$_{NP}$'). Disyllabic Nouns and Adjectives were prompted because previous research with adults had shown that when adult French speakers have to mark focus constituents in longer phrases, the probability of using an initial intonation rise increases (e.g. German & D'Imperio, 2015). All target Nouns and Adjectives were highly frequent words belonging to the children's vocabulary, and they all belonged to the clothing and adornment semantic fields (see Appendix A for a complete list of expected sentences).

Two factors were manipulated and fully crossed: the information status of the elements within the NP (broad focus, contrastive focus, or corrective focus), and the position of the focused element within the NP (non-phrase final position –i.e. the noun– or phrase-final position –i.e. the adjective). This resulted in five experimental conditions, summarized in Table 1. In total, 60 trials were prompted, 12 trials per experimental condition, in a within-subject design. Each combination of disyllabic Noun and disyllabic Adjective was elicited in all five experimental conditions to rule out potential effects of segmental and syllabic structure in our data.

The visual display depicted a girl named Claire (on the bottom left corner) whom children were asked to interact with. The girl had her eyes covered and was thus unable to see the details of the scene. At the centre of the screen there was a bag containing different coloured objects. On the top right corner of the screen an action was depicted (e.g. 'getting to see the turtle's neck'), together with one of the objects contained in the bag. The top-right object was the target element that the character would need to take from the big bag in order for the action to take place (see Figure 1). Children were instructed to help her by naming the target object, and thus producing the target sentences (see section 2.3 for further details on the procedure).

The number and nature of the objects inside the big bag were manipulated to elicit the distinct experimental conditions. In the broad focus condition only one coloured object was shown in the bag, so there was no need to contrastively or correctively focus either the noun or the adjective to name the target element. In the contrastive focus condition, two or more items were displayed inside the bag. Some items were of the same type (e.g. two suitcases) but contrasted in colour (e.g. one item would be orange and the other purple, such as in Figure 1), and sometimes items only differed in type (e.g. a suitcase vs. a shoe) but not in colour (e.g. they would both be purple). If they only contrasted in colour, we expected to elicit contrastive focus on the phrase-final element (the adjective in French); if they only contrasted in type, we

expected to elicit contrastive focus on the non-phrase final element (the noun in French). To elicit corrective focus, the game was configured in such a way that Claire sometimes would accidentally select the wrong object from the bag, so that, as a consequence, children would have to repeat the instruction in order for her to correct her action. Corrective sentences always followed contrastive sentences, since they were expected to be the consequence of Claire's wrong selection of an item from the bag.

2.3 | Procedure

The study was approved by the local ethics committee at Grenoble University (France, IRB00010290-2016-07-05-06). Before the experimental task, children went through a pure-tone audiometric screening test using an audiometer (Robé médical 9910, testing 125 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz and 8000 Hz with 20 dB intensity). They were also tested for their expressive and receptive language abilities using the *Évaluation du Langage Oral* (ELO) test (subscales on receptive lexicon and on sentence production; Khomsi, 2001).

Children were individually tested in a quiet room, sitting in front of a computer and next to the experimenter. A video-camera was placed behind the laptop, focusing on the children's bust, to monitor the children's head movements, and a sound recorder (Zoom H4nPro digital audio recorder) was placed on the table next to the child to record their speech productions. The video stimuli were displayed on a laptop computer using a PowerPoint presentation. The experimenter controlled the presentation and launched successive trials using a wireless mouse.

The game unfolded as follows: children were told that in order for Claire (the little girl) to launch the different actions (e.g. disclosing the turtle's neck), she had to pick the right object from inside the bag. Children were shown Claire's covered eyes and were told that they had to help her by giving her instructions about which target object she had to pick from the bag. Sometimes the game was programmed in such a way that Claire took the wrong object (to elicit a corrective focus context). In these cases, the experimenter prompted the child to re-tell her which was the right object to take. No instructions were given to children as to which lexical items they had to produce or how to order them within the sentence. If children produced sentences that did not have the expected target NP structure, the experimenter did not correct the child nor asked him/her to repeat, and instead simply initiated the following trial.

2.4 | Analysis

2.4.1 | Prosodic analysis

All utterances were first orthographically annotated using Praat (Boersma & Weenink, 2012), and then automatically segmented into words and syllables using SPPAS (Bigi, 2015). Because SPPAS is trained with adult speech material, all segmentations were also manually checked after the automatic segmentation to correct for potential computation errors. After the orthographic segmentation was completed, acoustic-prosodic measures were automatically extracted, such as Fundamental Frequency (F0) and syllable duration. As for F0, both F0 maximum and F0 minimum values were extracted within each target word of the NP, i.e. the Noun and the Adjective, which resulted in four F0 values: Noun F0 max, Noun F0 min, Adjective F0 max, and Adjective F0 min. Pitch range values (in semitones) were then calculated for each target element of the NP by subtracting F0 min from F0 max for each word. As for syllable duration, we automatically extracted values resulting from the syllabic segmentation. Given that children produced connected speech, resyllabification strategies were common (e.g. in an utterance such as *Prends [une valise orange]*$_{NP}$ 'Take an orange suitcase' children would produce the target Noun Phrase like UNE.VA.LI.SO.RANGE [yn.va.li.zo.ʁɑ̃ʒ] instead of UNE.VA.LIS.O.RANGE [yn.va.liz.o.ʁɑ̃ʒ]). In such cases, the duration of Adjective-initial syllable also included the Noun-final segment [z] as syllabic onset (see Figure 2 for an illustration of the acoustic analysis).

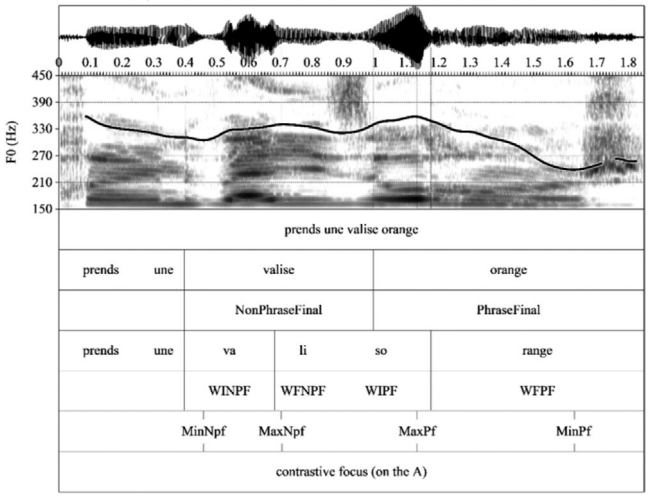

FIGURE 2 Example of a waveform and spectrogram with superimposed F0 curve of a child's utterance, and its annotation in Praat in terms of orthographic transcription (Tier 1), word by word segmentation (Tier 2), word position within the target NP (Tier 3), syllable by syllable segmentation (Tier 4), syllable position within the target NP (Tier 5; WINPF: word-initial non-phrase-final; WFNPF: word-final non-phrase-final; WIPF: word-initial phrase-final; WFPF: word-final phrase-final), F0 targets (Tier 6; MinNpf: F0min in non-phrase-final word; MaxNpf: F0max in non-phrase-final word; MaxPf: F0max in phrase-final word; MinPf: F0min in phrase-final word), and focus condition that was being prompted (Tier 7)

2.4.2 | Gesture analysis

The ELAN annotation tool (Lausberg & Sloetjes, 2009) was used to code for (1) the presence or absence of a head gesture within the limits of the target NPs (i.e. article + Noun + Adjective) that had a non-referential value, i.e., not bearing a semantic link with speech but informing about the structure of the utterance (McNeill, 1992); (2) the type of head gesture (if any), with the following set of possible

categories: head nod, head tilt, head-and-body moving forward, chin-forward pointing, or eyebrow raising (see Supplementary materials for a video sketch exemplifying each type of head gesture), and (3) the temporal alignment of the head gesture (if any) with respect to the speech stream (2 possibilities: 'correctly' aligned or 'incorrectly' aligned). We decided to use a broad definition of head gesture in order to pinpoint any movement affecting the head that children could use to mark the informational status of discourse referents. Thus, we included not only the more canonical head nods and head tilts, but also facial movements such as eyebrow-raising and protruded postures such as head-and-body moving forward or chin-forward movements (see Wagner et al., 2014, for a complete description of types of head gestures). No specific head gesture categories were pre-defined, and instead the distinct categories emerged as the coding unfolded. As for the temporal alignment, because all head gestures were bi-phasic (with a preparation phase, a gesture apex at the turning point where the head movement changes its direction, and a retraction phrase), a head gesture was considered to be 'correctly' aligned with a certain word if its apex occurred within the temporal limits of that word, and 'incorrectly' aligned if this temporal pattern was not identified (based on previous studies in adults by Esteve-Gibert & Prieto, 2013; Esteve-Gibert et al., 2017; Krivokapic et al., 2017; Leonard & Cummins, 2011). Only head gestures that co-occurred either with the target noun or the target adjective within the NP were annotated. This means that if a child produced a head gesture during the production of the word *Prends* 'Take' in the sentence *Prends la valise orange* 'Take the orange suitcase', that gesture was not coded. If a child produced a head gesture both on the target noun and on the target adjective, we took into consideration the head gesture that was more salient or prominent (i.e., that implied a wider displacement).

Because we did not use any motion tracking system in our data collection, only the coder's perception was used for gesture transcription purposes. In order to avoid perception biases, two actions were conducted. First, annotation of (a) the presence/absence of head gestures, (b) the type of head gesture, and (c) the position of the gesture apex within the limits of the head gesture were done in a 'muted' mode in ELAN and also blind to the focus condition. Second, (informal) intra-rater reliability checks were performed during the gesture coding: after all participants were coded in a first round, the coder went back to the initial participant and inspected all annotations again to include modifications if needed. We believe that a muted and blind coding of gestures is essential to avoid biases in studies that explore body movements in relation to speech. Likewise, a second round of coding is necessary when various speakers need to be analysed, to get used to inter-speaker variability and to avoid biases related to speaker style.

3 | RESULTS

A total of 1,418 utterances were included in the analyses. The total amount of utterances expected was 1,540 utterances (60 trials per participant x 24 participants) but some children did not produce some trials due to fatigue (especially younger children). Because the experimental procedure elicited spontaneous productions, we expected some variation in the nature of the Noun Phrase (NP) structures we would obtain. In 85.9% (N = 1,218) of the cases the children's productions had the expected canonical NP structure (Noun + Adjective; e.g. *Prends [la valise orange]*$_{NP}$ 'Take [the orange suitcase]'). In 12.2% (N = 187) of the cases children only produced the Noun (especially in the broad focus condition and when the focused element was the Noun; e.g. *Prends [la valise]*$_{NP}$ 'Take [the suitcase]$_{NP}$'), and in 0.7% (N = 10) of the cases they only produced the phrase-final Adjective (e.g. [*L'orange!*]$_{NP}$ '[The orange one!]$_{NP}$'). Additionally, in 0.2% (N = 3) of the cases children produced an Adjective in an ungrammatical non-phrase-final position (especially when the focused element was the Adjective; e.g. *Orange, la valise* 'Orange, the suitcase').

3.1 | Prosody for focus marking

Utterances that were produced with the ungrammatical Adjective + Noun NP structure were excluded from subsequent prosodic analyses. All the other NP structures were included in subsequent analyses.

3.1.1 | Syllable duration

We conducted a linear mixed-effects model using the *lmer* function of the lme4 package in R (Bates et al., 2011). The dependent variable was 'Syllable duration', and fixed factors were 'Focus type' (broad focus, contrastive focus, corrective focus), Age (in months), Receptive language skills, and Expressive language skills. 'Syllable position' (word-initial phrase-final, word-final phrase-final, word-initial non-phrase-final, and word-final non-phrase-final) was also included as a fixed factor because previous findings on French adults revealed that the syllable position influenced how speakers used duration for focus marking (German & D'Imperio, 2015; Jun & Fougeron, 2000). Random variables included a by-Participant random slope for Syllable position and a by-Item random slope for focus condition.

Results of this model revealed that Syllable duration was significantly affected by Syllable position ($\chi^2(3) = 13.662$, $p < .01$), and coefficients revealed that word-final phrase-final syllables (Intercept: $\beta = -63.402$, SE = 100.712, t = −0.063) were significantly longer than word-initial phrase-final syllables ($\beta = -108.728$, SE = 49.432, t = −2.2), all the other levels not differing significantly (word-final non-phrase final: $\beta = -59.988$, SE = 46.299, t = −1.296; word-initial non-phrase final: $\beta = -41.972$, SE = 71.847, t = −0.584). The model also revealed a main effect of Age by which older children produced longer syllable durations independent of the condition ($\beta = 5.510$, SE = 1.632, t = 3.376). No main effect of Focus type was observed (Broad focus/Intercept: $\beta = 430.467$, SE = 28.159, t = 15.287; Contrastive focus: $\beta = -14.062$, SE = 15.809, t = −0.889; Corrective focus: $\beta = -8.462$, SE = 16.776, t = −0.504), nor of Receptive linguistic skills ($\beta = 6.235$, SE = 4.524, t = 1.378) or of Expressive linguistic skills ($\beta = 4.693$, SE = 4.757, t = 0.986). Similarly, we did not find any interaction between Focus type and Syllable position ($\chi^2(6) = 4.5941$, $p = .59$), and no 3-way interaction (Focus type x Syllable position x

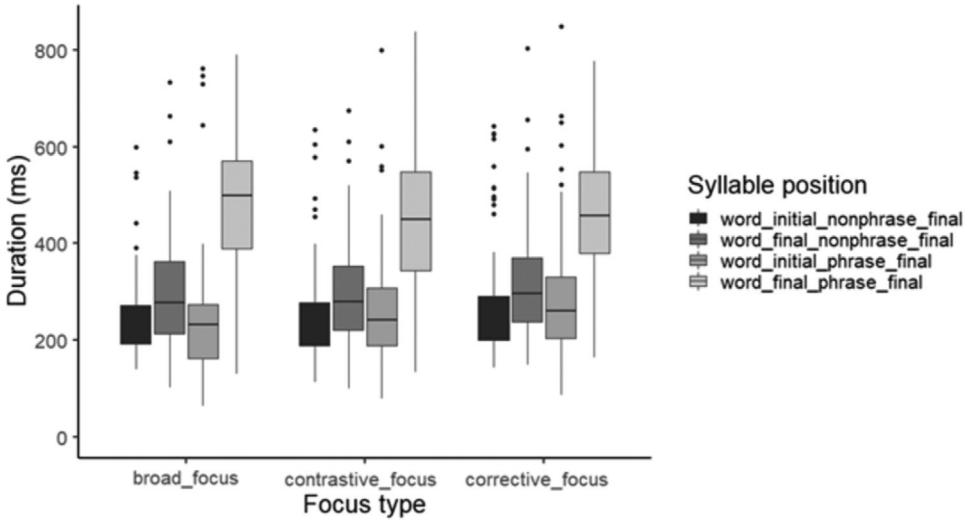

FIGURE 3 Syllable duration values (in milliseconds) across Focus conditions and Syllable position within the NP

Age: $\chi^2(17) = 24.229$, $p = .11$; Focus type x Syllable position x Receptive linguistic skills: $\chi^2(17) = 17.87$, $p = .39$; Focus type x Syllable position x Expressive linguistic skills: $\chi^2(17) = 13.758$, p = .68). Figure 3 illustrates the duration patterns across Focus types and Syllable positions.

3.1.2 | Word-level pitch range

A linear mixed-effects model was run with the *lmer* function in R. The dependent variable was Pitch range (in Hz), and fixed factors were Focus type (broad focus, contrastive focus, corrective focus), Focus position (non-phrase final, phrase-final), Age (in months), Receptive linguistic skills, and Expressive linguistic skills. Random variables included a by-Participant random slope for Focus type and a random intercept for Item.

Results of the model revealed a main effect of Focus position ($\chi^2(1) = 133.31$, p < .001), by which target words in a non-phrase final position (Intercept: $\beta = 3.513$, SE = 0.3226, t = 10.89) had a significantly lower F0 range than target words in a phrase final position ($\beta = 1.3177$, SE = 0.117, t = 11.80), but no main effect of Focus type ($\chi^2(2) = 1.1607$, $p = .559$), Age ($\chi^2(1) = 1.9684$, $p = .161$), or Receptive language ($\chi^2(1) = .0.283$, $p = .594$). Expressive language skills did influence F0 range values ($\chi^2(1) = 5.951$, $p < .05$): children with higher scores in this language test produced narrower pitch ranges ($\beta = -0.373$, SE = 0.134, t = −2.776) than children with lower scores in the test (Intercept: $\beta = 7.408$, SE = 1.483, t = 4.994). The fixed factor Focus position did not interact with Focus type ($\chi^2(2) = 1.7796$, $p = .411$), and no 3-way interactions were observed (Focus position x Focus type x Age: $\chi^2(7) = 12.071$, $p = .098$; Focus position x Focus type x Receptive language: $\chi^2(7) = 11.484$, $p = .118$: Focus position x Focus type x Expressive language: $\chi^2(7) = 13.204$, $p = .07$). Figure 4 shows that pitch range values varied across Focus positions but not across Focus types.

An additional analysis was run to examine whether children, instead of expanding the F0 range of the focused elements, compressed it in the accompanying non-focused elements. Dohen and Lœvenbruck (2004) had shown that French speakers can deaccent post-focal elements in an utterance to highlight the contrast between focused and non-focused items (e.g., deaccenting the word 'suitcase' in 'Take the ORANGE suitcase' instead of only expanding the F0 range in the word 'orange'). We thus ran a new analysis in which the factor Focus type included 2 extra categories: elements accompanying contrastively focused words in the NP (e.g. the word 'orange' in 'Take the orange SUITCASE' or the word 'suitcase' in 'Take the ORANGE suitcase', where capital letters signal contrastive focus), and elements accompanying correctively focused words in the NP (e.g. the word 'orange' in 'Take the orange **SUITCASE'** or the word 'suitcase' in 'Take the **ORANGE** suitcase', where capital bold letters signal corrective focus). A linear mixed-effects model was run with the *lmer* function in R, with Pitch range (in semitones) as the dependent variable, and the fixed factor Focus type with five levels: broad focus, contrastive focus, corrective focus, accompanying contrastive, accompanying corrective. Participant and Item were set as random factors. We found a main effect of Focus type ($\chi^2(4) = 27.326$, $p < .001$), by which elements accompanying a correctively focused word had significantly higher F0 range ($\beta = 0.367$, SE = 0.148, t = 2.485) than any other Focus type (Intercept/broad focus: $\beta = 4.194$, SE = .333, t = 12.584; accompanying contrastive: $\beta = -.226$, SE = .150, t = −1.508; contrastive focus: $\beta = -0.005$, SE = .147, t = -.037; corrective focus: $\beta = .225$, SE = .146, t = 1.542).

3.2 | Head gestures for focus marking

Children produced a total of 533 head gestures accompanying the target NP. The most frequent gesture type accompanying the target NP was chin-forward movement (35.3%, *N* = 188), followed by head-and-body-forward movement (26.3%, *N* = 140), head nod (19.3%, N = 103),

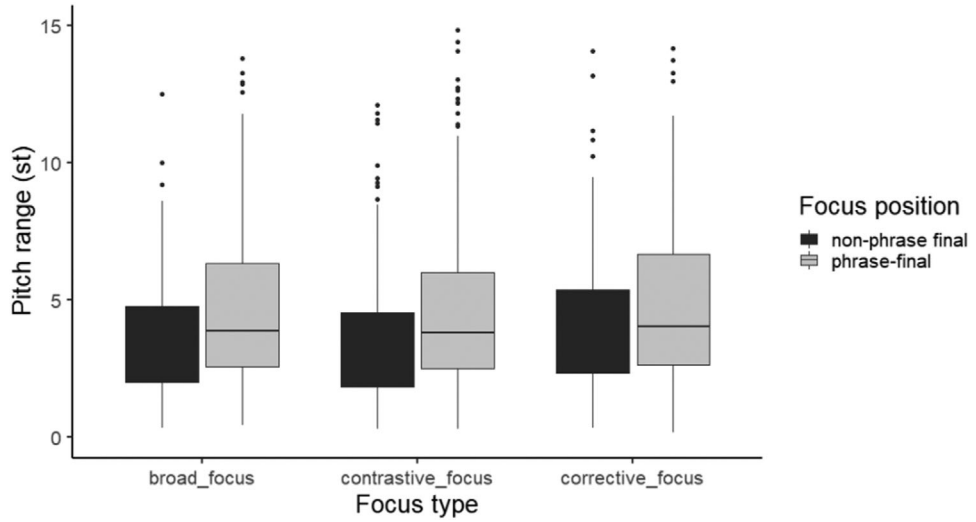

FIGURE 4 Box plots depicting word-level pitch range values (in semitones) across Focus types as a function of the position of the focused element within the NP

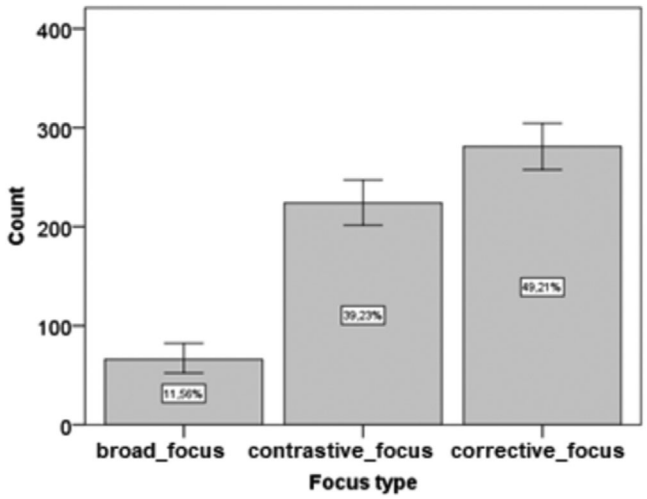

FIGURE 5 Absolute count and proportion of utterances accompanied by a head gesture in each experimental condition. Error bars: 95% CI

eyebrow-raising (14.8%, $N = 79$), or other gestures such as head tilt (3.2%, $N = 17$). All these gesture types were included in the subsequent analyses.

3.2.1 | Children's production of head gestures

Figure 5 illustrates the proportion of utterances accompanied by a head gesture that children produced in each focus condition. A logistic regression analysis with the *glmer* function in R was applied to the data, with 'Presence of gesture' as dependent variable, 'Focus type', 'Age (in months)', 'Expressive language abilities' and 'Receptive language abilities' as fixed factors, and with Participant and Item as random variables (a more complex random effect structure failed to converge). The model revealed a significant main effect of Focus type: children produced significantly more head gestures in the contrastive focus condition ($\beta = 0.890$, SE = .180, z = 4.936, $p < .001$) than in the broad focus condition (which was the Intercept category: $\beta = -0.183$, SE = .960, z = −0.191, $p = .848$), and the corrective focus condition elicited even more head gestures than the other two levels ($\beta = 1.282$, SE = .179, z = 7.141, $p < .001$). Children's age (in months) was not significant in the model ($\beta = .008$, SE = .037, z = .219, $p = .826$), and neither were children's expressive ($\beta = -0.112$, SE = .083, z = −1.34, $p = .18$) nor receptive linguistic abilities ($\beta = -0.142$, SE = .085, z = −1.662, $p = .09$). Focus type interacted with children's receptive linguistic abilities: children with higher receptive abilities produced more gestures in the contrastive focus condition ($\beta = .211$, SE = .093, z = .252, $p < .05$) than those with lower receptive abilities. All the other comparisons and interactions were non-significant.

3.2.2 | Alignment of the head gestures with the focused words

Most of the head gestures that children produced co-occurred in time with the target focused element (81.8%), while a minority was incorrectly aligned (18.2%). A logistic regression model was run with 'Gesture alignment (correct, incorrect)' as the dependent binomial variable, and fixed factors were 'Focus type' (contrastive focus, corrective focus) and 'Focus position' (phrase-final, non-phrase final). Note that the 'broad focus' category was not included in the 'Focus type' factor because in this condition there was no focused element within the NP, so any head gesture produced by children in that category would be superfluous. Participant and Item were set as random intercepts (a more complex random effects structure failed to converge).

Results of the model showed a main effect of Focus position ($\chi^2(1) = 24.048$, $p < .001$): there were significantly more correctly aligned head gestures when the focused element was phrase final

(β = 1.218, SE = .379, z = 3.21, $p < .01$) than when it was non-phrase final (Intercept category: β = .895, SE = .272, z = 3.291, $p < .001$). No main effect of Focus type ($\chi^2(1) = .761$, $p = .382$), Age ($\chi^2(1) = .448$, $p = .503$), nor of Receptive linguistic skills ($\chi^2(1) = 1.055$, $p = .304$), or of Expressive linguistic skills were found ($\chi^2(1) = .361$, $p = .544$). Similarly, no 2-way interaction between Focus position and Focus type was observed ($\chi^2(1) = .478$, $p = .49$), nor any 3-way interaction between these factors and either Age ($\chi^2(4) = 7.187$, $p = .13$), Receptive linguistic skills ($\chi^2(4) = 4.727$, $p = .316$), or Expressive linguistic skills ($\chi^2(4) = 2.527$, $p = .64$). These findings indicate that when children produced head gestures to mark focus, they did so in a temporally appropriate manner when the element to be focused was in phrase-final position (in French, the adjective). Instead, when the element to be focused was in a non-phrase final position (in French, the noun) children made more alignment mistakes and produced more head gestures in the inappropriate position (in that case, aligned with the phrase-final element).

3.3 | Integration of prosody and gestures in children's focus marking

Two *lmer* models were run to investigate if children modified the prosodic pattern of a word when a head gesture was produced on that word. In a first model, the dependent variable was Syllable duration (in milliseconds), and the fixed factors were 'Presence of gesture on that syllable' (yes, no) and Focus type (broad focus, contrastive focus, corrective focus). We included random intercepts by Participant and a by-Item random slope for Focus type. In a second model, the dependent variable was word-level Pitch range (in semitones), and the fixed factors and random effects structure were the same as in the first model.

The first model revealed that Syllable duration was significantly affected by the presence of a gesture on that syllable ($\chi^2(1) = 5.361$, $p < .05$), while not interacting with Focus type ($\chi^2(2) = .7293$, $p = .694$). The coefficients showed that syllables that were not accompanied by a gesture (β = 351.703, SE = 19.706, t = 17.84) were shorter than syllables that were accompanied by a gesture (β = 24.554, SE = 9.907, t = 2.47). Results of the second model showed that Pitch range was also affected by the presence of a gesture on that word ($\chi^2(1) = 9.006$, $p < .01$), while not interacting again with Focus type ($\chi^2(2) = 0.1132$, $p = .945$). The coefficients showed that words that were accompanied by a head gesture had a wider pitch excursion (β = 21.183, SE = 2.597, t = 8.155) than words not accompanied by a head gesture (β = 74.139, SE = 7.431, t = 9.977).

4 | DISCUSSION

Our study investigated whether French pre-schoolers (aged 4 and 5 years) use prosodic acoustic features (pitch range and syllable duration) and speech-accompanying head gestures to signal the informational status of discourse elements. We designed an experimental task to elicit spontaneous productions in three distinct focus conditions: one in which no element within a target phrase had specific emphasis (broad focus), one in which a specific element within the phrase contrasted with a set of possible alternatives (contrastive focus), and one in which an element within the phrase was particularly emphasised because it corrected a previously presented alternative (corrective focus). Our main finding is that 4- and 5-year-old French-speaking children use head gestures for the purpose of marking focus type, while not using syllable duration nor F0 range expansion for the same effect (while adults do). Crucially though, even though children seem not to be able to independently control prosody to the purpose of focusing lexical elements, we observed that the production of head gestures does entrain prosodic variation within the accompanying speech, as target words marked by head gestures showed increased duration and pitch range.

Specifically, children produced more head gestures in a contrastive focus context than in a broad focus context, and even more so in a corrective focus context than in the other conditions. By means of head gestures, 4- and 5-year-old children can convey if what they say needs to be added to the shared common ground, or if it contrasts with (or corrects) a set of possible alternatives. This is the first study showing that young children use head gestures in such a precise way to indicate focus marking. Previous research had shown that when 5-year-old children spontaneously narrate a story, they move their hands to signal discourse cohesion and interactivity (Colletta et al., 2014; Nicoladis et al, 1999). Here we show that children can also signal fine linguistic and communicative differences by means of head gestures even if they still do not master other adult-like linguistic strategies to do so, and thus we reveal that head gestures may serve a clear linguistic function in communication, at least in language development.

We also found that children with higher linguistic (receptive) abilities were more inclined to use head gestures to signal contrastive focus. The relation between gesture use and language development has been previously established in the literature, with evidence coming from production studies. Young infants' early use of manual pointing gestures is related to their future lexical and syntactic production abilities (e.g. Rowe & Goldin-Meadow, 2009), and pre-schoolers' discursive abilities increase after being trained with the observation of a narrative task containing manual beat gestures (Vilà-Gimenez et al., 2020). Our findings add to this body of work by showing that receptive linguistic abilities at the preschool stage might also be related to the children's gesture use (at least when gestures have a non-referential meaning), and that receptive linguistic abilities may be the driving force for these gestural abilities to develop (see also Griffiths et al., 2020, for more results in this direction), as in a sort of complex dynamic system in which distinct components interact with each other at distinct moments in time. Further studies taking into account receptive linguistic skills should investigate their relationship with infants' and children's gesture use, in order to confirm if our result also holds for other types of gestures and in other stages of language development.

Moreover, we observed that children at all ages tested, and irrespectively of their linguistic abilities, aligned head gestures and speech in an appropriate manner. Adult research has revealed that prosodic events such as pitch accents and phrasal edge tones serve as

anchoring landmarks for the temporal alignment of body movements (Esteve-Gibert et al., 2017; Krivokapic et al., 2017; Leonard & Cummins, 2011), and some proposals even advocate for a biological motoric pulse governing this gesture-speech coordination (e.g. Iverson & Thelen, 1999; Pouw, Harrison, & Dixon, 2020; Rusiewicz, 2011). In the development literature, most of the (few) previous studies suggest that the fine-grained temporal coordination of gesture and speech, and the possible biological motoric pulses, are observed very early on in referential gestures (Esteve-Gibert & Prieto, 2014; Murillo & Capilla, 2015). Our study shows that this might also be the case for head gestures with a non-referential value, which share a close rhythmic relation with speech.

Our findings, however, also indicate some exceptions to this general alignment trend, given that a small proportion of head gestures was not appropriately aligned with speech. The misalignment of some gesture-speech combinations seems to be a repeated finding in developmental (Mathew et al., 2017) and adult literature (Bergmann et al., 2011; Rohrer, Prieto, & Delais-Roussarie, 2019; Shattuck-Hufnagel & Ren, 2018), and it deserves future investigations. The direction of the effect in our data (i.e., most of the incorrectly aligned head gestures were anchored to the phrase-final element) could be a consequence of the French prosodic structure, which places metrical prominence at the end of the Accentual Phrase (Jun & Fougeron, 2000). Alternatively, it could be that children produced these phrase-final gestures with a pragmatic function that is not related to focus marking, such as signalling the end of the utterance.

Young children's systematic use of head gestures for marking focus contrast is at odds with their failure to use prosodic cues for that same purpose. In our data, French children were not able to modify neither syllable duration nor pitch range to indicate the informational status of the target element. In contrast, adult French speakers do use both accentual and phrasing strategies to mark focused elements (Dohen & Lœvenbruck, 2004; Féry, 2001; German & D'Imperio, 2015; Jun & Fougeron, 2000; Michelas & D'Imperio, 2015). This has acoustic consequences, in that the initial and final syllables of the re-phrased element are usually lengthened, while the pitch range values are increased. Our analysis, instead, reveals that children use the same acoustic patterns in all focus conditions. Moreover, independent of the focus status, they used syllable duration to mark the end of the phrasal constituent (by lengthening syllables in a word-final phrase-final position and widening pitch span on phrase-final words). Chen (2018) proposes that children acquiring languages with phonological strategies (as opposed to phonetic strategies) for the purpose of signalling information structure show a later development of acoustic-prosodic cues to focus. Our results seem to support this hypothesis.

We also obtained an unexpected result regarding children's use of pitch range: children expanded the pitch range of elements immediately following correctively focused words, i.e. the word 'hat' in the sentence 'Take the [**purple**]$_{\text{corrective focus}}$ hat'. Even if adult French speakers can use pitch compensation strategies in contrastive focus context (Dohen & Lœvenbruck, 2004), they compress (but never expand) the pitch range of focus-accompanying elements, in both following and preceding positions. Our findings might indicate that French children at age 4–5 might have difficulties with precisely controlling the timing of pitch expansion.

Last but not least, an interesting finding in our study is that when head gestures were used, the prosodic features of the accompanying speech varied (syllables were lengthened and words were accompanied by a wider pitch range), even if these children could not yet use prosodic strategies to signal focus. We speculate that the initial use of the gesture modality might be entraining the emergence of the prosodic modality, at least in the context of focus marking. Although this hypothesis would need to be confirmed by further research on the dynamics of the gesture-prosody interaction in the developing child, our results suggest that this could be the case. Language is a complex system with many components developing at the same time, and the way that gesture, prosodic, and meaning components are dynamically intertwined in this process is still underexplored (Iverson & Thelen, 1999; Smith & Thelen, 2003). Previous research suggests that the gesture modality precedes the speech modality in expressing semantic or pragmatic meaning (Hübscher et al., 2019; Rowe & Goldin-Meadow, 2009); other results point at the exact opposite pattern (Özçaliskan et al., 2013, for the relation between iconic gestures and verbs). In the present study we found that the children's use of the gesture modality for focus marking precedes their use of the prosodic modality, and that it is precisely gesture use which might entrain the emergence of the prosodic modality in that context. This is highly interesting because it suggests that gesture not only precedes (and predicts) the emergence of other linguistic abilities, but it might also actively contribute to their emergence.

A limitation of the study is that gesture coding was manually conducted. Contrary to when automatic tools are used, our data could potentially include coder's biases and/or misperceptions. However, we think that biases and misperceptions were minimized because gestures were coded in a muted mode (without listening to the audio) and blind to the focus condition. Also, a second round of coding once all participants were already annotated enabled to correct potential misperceptions and biases due to inter-speaker variability and their distinct gesturing styles.

We know that children's ability in using prosody to signal the informational status of discourse referents emerges quite late in development, especially in those languages that use prosodic features at a phonological level. We showed that this does not result in children not being able to signal information structure, given that they trade prosodic means with the use of head gestures for that linguistic purpose. What is more, those early head gestures might entrain the development of prosodic cues to focus. More studies are needed to confirm the entrainment patterns of gesture and prosody in language development, to examine whether populations with difficulties in acquiring prosody and meaning can make use of head gestures (and body movements, in general) to improve their linguistic abilities. Future research studies using motion tracking techniques to automatically capture accurate alignment patterns would contribute to sketching a more comprehensive model of how gesture and speech interact to structure information in the discourse. Human communication is multimodal, and the study of how infants and children develop the ability

to communicate with others cannot and should not leave out the visual components of language.

ACKNOWLEDGMENTS

We thank Tim Mahrt, Paolo Roseano and Wendy E. Garcia for their help with Python and Praat scripts. We also thank Marie Sarremejeanne for data collection. We would like to thank Cristel Portes, Pilar Prieto, Stefanie Shattuck-Hufnagel for discussions on data annotation and results.

FUNDING

Labex Brain and Language Research Institute (BLRI) to first author and IUF (Institut Universitaire de France) funding to the last author.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to ethical restrictions.

REFERENCES

Ambrazaitis, G., & 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 Communication*, *95*, 100–113. http://doi.org/10.1016/j.specom.2017.08.008

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly of Behavior and Development*, *21*(3), 205–226.

Baumann, S., & Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics*, *38*(10), 1636–1657. http://doi.org/10.1016/j.pragma.2005.03.017

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

Bigi, B. (2015). SPPAS - Multi–lingual Approaches to the Automatic Annotation of Speech. *International Society of Phonetic Sciences*, *111/112* 54–69.

Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. Version 5.3.19.

Carignan, C., Esteve-Gibert, N., Loevenbruck, H., Dohen, M., & D'Imperio, M. (2020). Strategies of head nod alignment with pitch prominence in French focus. *12th International Seminar on Speech Production (ISSP)*, Yale University (USA). December, 14–18.

Cavé, C., Guaïtella, I., Bertrand, R., Santi, S., Harlay, F., Espesser, R., Parole, L., Provence, U. D., Phonétique, L. D., Franche-comté, U. D., Fonctionnelle, L. D. N., & Provence, U. D.. (1996). About the relationship between eyebrow movements and F0 variations. *Proceedings of the ICSLP*, *4*, 2175–2179.

Chen, A. (2009). The phonetics of sentence-initial topic and focus in adult and child Dutch. In M. C. Vigario, S. Frota & M. J. Freitas (Eds.), *Phonetics and phonology: Interactions and interrelations* (pp. 91–106). Amsterdam: John Benjamins.

Chen, A. (2011). Tuning information packaging: Intonational realization of topic and focus in child Dutch. *Journal of Child Language*, *38*(5) 1055–1083.

Chen, A. (2018). Get the focus right: Acquisition of prosodic focus-marking across languages. In P. Prieto & N. Esteve-Gibert (Eds.), *The Development of Prosody in First Language Acquisition* (pp. 295–313). John Benjamins.

Colletta, J.-M., Guidetti, M., Capirci, O., Cristilli, C., Demir, O. E., Kunene-Nicolas, R. N., & Levine, S. (2014). Effects of age and language on co-speech gesture production: An investigation of French, American, and Italian children's narratives. *Journal of Child Language*, 1–24.

Dohen, M., & Lœvenbruck, H. (2004). Pre-focal rephrasing, focal enhancement and post-focal deaccentuation in French. *Proceedings of the 8th International Conference on Spoken Language Processing*, 2–5.

Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, *56*, 850–864. http://doi.org/10.1044/1092-4388(2012/12-0049)

Esteve-Gibert, N., Borràs-Comes, J., Asor, E., Swerts, M., & Prieto, P. (2017). The timing of head movements: The role of prosodic heads and edges. *The Journal of the Acoustical Society of America*, *141*(6), 4727–4739. http://doi.org/10.1121/1.4986649

Esteve-Gibert, N., & Prieto, P. (2014). Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Communication*, *57*, 301–316. http://doi.org/10.1016/j.specom.2013.06.006

Féry, C. (2001). Focus and phrasing in French. In C. Féry & W. Sternefeld (Eds.), *Audiatur Vox Sapientes: A Festschrift for Arnim von Stechow* (pp. 153–181). Akademie Verlag.

Frota, S., & Prieto, P. (2015). *Intonation in Romance*. Oxford University Press.

German, J. S., & D'Imperio, M. (2015). The status of the initial rise as a marker of focus in French. *Language and Speech*, *59*(2), 165–195. http://doi.org/10.1177/0023830915583082

Griffiths, S., Goh, S. K. Y., & Norbury, C. F. (2020). Early language competence, but not general cognitive ability, predicts children's recognition of emotion from facial and vocal cues. *PeerJ*, *8*, e9118. http://doi.org/10.7717/peerj.9118

Hübscher, I., Vincze, L., & Prieto, P. (2019). Children's signaling of their uncertain knowledge state: Prosody, face, and body cues come first. *Language Learning and Development*, *15*(4), 366–389. http://doi.org/10.1080/15475441.2019.1645669

Ishi, C. T., Ishiguro, H., & Hagita, N. (2014). Analysis of relationship between head motion events and speech in dialogue conversations. *Speech Communication*, *57*, 233–243. http://doi.org/10.1016/j.specom.2013.06.008

Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, *6*, 19–40.

Jun, S., & Fougeron, C. (2000). A phonological model of French intonation. In A. Botinis (Ed.), *Intonation: Analysis, Modeling and Technology* (pp. 209–242). Kluwer Academic Publishers.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The Relationship of Verbal and Nonverbal Communication* (pp. 207–227). Mouton.

Khomsi, A. (2001). *Evaluation du Langage Oral (ELO)*. ECPA.

Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, *55*, 243–76. http://doi.org/10.1556/ALing.55.2008.3-4.2

Krivokapic, J., Tiede, M. K., & Tyrone, M. E. (2017). A kinematic study of prosodic structure in articulatory and manual gestures: Results from a novel method of data collection. *Laboratory Phonology*, *8*(1), 1–26. http://doi.org/10.5334/labphon.75

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.

Lausberg, H. Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES–ELAN system. *Behavior research methods*, *41*(3) 841–849.

Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, *26*(10), 1457–1471. http://doi.org/10.1080/01690965.2010.500218

Mathew, M., Yuen, I., & Demuth, K. (2017). Talking to the beat: Six-year-olds' use of stroke-defined non-referential gestures. *First Language*, *38*(2), 111–128. http://doi.org/10.1177/0142723717734949

McNeill, D. (1992). Hand and mind: What gestures reveal about thought. University of Chicago Press.

McNeill, D. (Ed.). (2000). Language and Gesture. Cambridge: Cambridge University Press.

Ménard, L., Lœvenbruck, H., & Savariaux, C. (2006). Articulatory and acoustic correlates of contrastive focus in French children and adults.

In J. Harrington & M. Tabain (Eds.), *Speech Production: Models, Phonetic Processes and Techniques*, (pp. 227–251). Psychology Press.

Michelas, A., & D'Imperio, M. (2015). Prosodic boundary strength guides syntactic parsing of French utterances. *Laboratory Phonology*, *6*(1), 119–146. http://doi.org/10.1515/lp-2015-0003

Moubayed, S. A.l, & Beskow, J. (2011). Audio-visual prosody: Perception, detection, and synthesis of prominence. *Lecture Notes in Computer Science*, *6456*, 55–71. http://doi.org/10.1007/978-3-642-18184-9_6

Murillo, E., & Capilla, A. (2015). Properties of vocalization- and gesture-combinations in the transition to first words. *Journal of Child Language*, *43*(4), 890–913. http://doi.org/10.1017/S0305000915000343

Nicoladis, E., Mayberry, R. I., & Genesee, F. (1999). Gesture and early bilingual development. *Developmental Psychology*, *35*(2), 514–526. http://doi.org/10.1037/0012-1649.35.2.514

Özçaliskan, S., Gentner, D., & Goldin-meadow, S. (2013). Do iconic gestures pave the way for children's early verbs? *Applied Psycholinguistics*, *35*(6), 1143–1162. https://doi.org/10.1017/S0142716412000720

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in Communication*. MIT Press.

Pouw, W., Harrison, S. J., & Dixon, J. A. (2020). Gesture-speech physics: The biomechanical basis for the emergence of gesture-speech synchrony. *Journal of Experimental Psychology: General*, *149*(2), 391–404.

Rapin, L., & Ménard, L. (2019). The multimodal perception of contrastive focus in French: A developmental study. *Frontiers in Communication*, *3*, 60. http://doi.org/10.3389/fcomm.2018.00060

Romøren, A. S. H. Chen, A. (2015). Quiet is the New Loud: Pausing and Focus in Child and Adult Dutch. *Language and Speech*, *58*(1), 8–23.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*, 75-116 http://doi.org/10.1007/BF02342617

Roustan, B., & Dohen, M. (2010). Co-production of contrastive prosodic focus and manual gestures: Temporal coordination and effects on the acoustic and articulatory correlates of focus. *Proceedings of Speech Prosody 2010*, 11–14 May.

Rowe, M. L., & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science*, *12*(1), 182–187. http://doi.org/10.1111/j.1467-7687.2008.00764.x

Rusiewicz, H. L. (2011). Synchronization of speech and gesture: A dynamic systems perspective. *Paper presented at the 2nd Gesture and Speech in Interaction (GESPIN)*, Bielefeld, Germany.

Sauer, E., Levine, S. C., & Goldin-Meadow, S. (2010). Early gesture predicts language delay in children with pre- or perinatal brain lesions. *Child Development*, *81*(2), 528–539. http://doi.org/10.1111/j.1467-8624.2009.01413.x

Shattuck-Hufnagel, S., & Prieto, P. (2019). Dimensionalizing co-speech gestures. *Proceedings of the International Conference on Phonetic Sciences (ICPhS)*, Melbourne, Australia.

Shattuck-Hufnagel, S., & Ren, A. (2018). The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in Psychology*, *9*(SEP), 1–13.

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, *7*(8), 343–348. http://doi.org/10.1016/S1364-6613(03)00156-6

Szendroi, K., Bernard, C., Berger, F., & Gervain, J. (2018). Acquisition of prosodic focus marking by English, French, and German three-, four- five- and six-year-olds. *Journal of Child Language*, *45*, 219–241. http://doi.org/10.1017/S0305000917000071

Vallduví, E. (1991). *The informational component*. Garland.

Vilà-Giménez, I., & Prieto, P. (2020). Encouraging kids to beat: Children's beat gesture production boosts their narrative performance. *Developmental Science*. First view online.

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, *57*, 209–232. http://doi.org/10.1016/j.specom.2013.09.008

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Esteve-Gibert, N., Lœvenbruck, H., Dohen, M., & D'Imperio, M. (2021). Pre-schoolers use head gestures rather than prosodic cues to highlight important information in speech. *Developmental Science*, e13154. https://doi.org/10.1111/desc.13154