ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

# Multimodal Development in Children's Narrative Speech: Evidence for Tight Gesture–Speech Temporal Alignment Patterns as Early as 5 Years Old

**Júlia Florit-Pons,[a] Ingrid Vilà-Giménez,[a,b] Patrick Louis Rohrer,[a,c] and Pilar Prieto[d,a]**

[a] Grup d'Estudis de Prosòdia, Department of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain [b] Department of Subject-Specific Education, Universitat de Girona, Catalonia, Spain [c] Laboratoire de Linguistique de Nantes, Nantes Université, France [d] Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain

ARTICLE INFO

Article History:
Received August 1, 2022
Revision received October 13, 2022
Accepted December 1, 2022

Editor-in-Chief: Cara E. Stepp
Editor: Elizabeth Salmon Heller Murray

https://doi.org/10.1044/2022_JSLHR-22-00451

ABSTRACT

**Purpose:** This study aims to analyze the development of gesture–speech temporal alignment patterns in children's narrative speech from a longitudinal perspective and, specifically, the potential differences between different gesture types, namely, gestures that imagistically portray or refer to semantic content in speech (i.e., referential gestures) and those that lack semantic content (i.e., non-referential gestures).
**Method:** This study uses an audiovisual corpus of narrative productions ($n$ = 332) from 83 children (43 girls, 40 boys) who participated in a narrative retelling task at two time points in development (at 5–6 and 7–9 years of age). The 332 narratives were coded for both manual co-speech gesture types and prosody. Gestural annotations included gesture phasing (i.e., preparation, stroke, hold, and recovery) and gesture types (in terms of referentiality, i.e., referential and non-referential), whereas prosodic annotations included pitch-accented syllables.
**Results:** Results revealed that by ages 5–6 years, children already temporally aligned the stroke of both referential and non-referential gestures with pitch-accented syllables, showing no significant differences between these two gesture types.
**Conclusions:** The results of the present study contribute to the view that both referential and non-referential gestures are aligned with pitch accentuation, and therefore, this is not only a characteristic of non-referential gestures. Our results also add support to McNeill's phonological synchronization rule from a developmental perspective and indirectly back up recent theories about the biomechanics of gesture–speech alignment, suggesting that this is an inherent ability of oral communication.

In the last decades, studies on language development have revealed that both speech and gesture develop hand in hand, supporting the claim that gesture and speech constitute a single multimodal system of communication (McNeill, 1992). While research has shown that the adult multimodal communication system reveals a high level of temporal alignment between co-speech gesture production and pitch accentuation in speech (i.e., the most prosodically prominent syllables in speech; see, e.g., Shattuck-Hufnagel & Ren, 2018, for a review), less is known from a developmental perspective and specifically about the gesture–speech temporal alignment patterns at the age when children start performing more complex discourses, such as narratives. Therefore, the main objective of this current longitudinal study is to examine the patterns of temporal alignment of both referential (i.e., visually representing speech semantic content, such as *iconic*, *metaphoric*, and

Correspondence to Júlia Florit-Pons: julia.florit@upf.edu. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

*deictic* or *pointing* gestures; see McNeill, 1992) and non-referential (i.e., lacking semantic content; see Rohrer et al., 2021; Shattuck-Hufnagel & Ren, 2018) gestures produced in children's narrative discourses at two time points in development: at the age of 5–6 years and 2 years later.

## Temporal Alignment Between Gesture and Speech in Adult Speech

McNeill (1992) claimed that gesture and speech modalities form a well-integrated communicative system from a phonological perspective, mainly through the so-called *phonological synchronization rule*. This rule argues that gesture strokes (i.e., the most prominent phase of a gesture; Kendon, 1980; McNeill, 1992) are produced co-occurring with or just before stressed syllables; in other words, that prominence in speech and prominence in gesture occur in close temporal synchronization. Previous empirical research has assessed gesture–speech temporal alignment, in terms of the overlap between the gesture stroke or gesture apex[1] and regions of prosodic prominence, such as pitch-accented words and pitch-accented syllables, lending support to McNeill's (1992) phonological synchronization rule. For instance, studies focusing on academic-lecture-style discourses have demonstrated that the stroke of co-speech gestures overlapped with pitch accentuation in 80%–90% of the cases (Im & Baumann, 2020; Shattuck-Hufnagel & Ren, 2018; Shattuck-Hufnagel et al., 2016; Yasinnik et al., 2004).

Interestingly, previous investigations have also considered the potential differences in the temporal alignment between different gesture types (in terms of referentiality, i.e., referential and non-referential gestures). While *referential* gestures bear a close relationship to the semantic content of the speech they accompany by imagistically representing the properties of a referent (i.e., iconicity, metaphoricity) or by locating entities in space (i.e., deixis; McNeill, 2005), *non-referential* gestures do not visually represent speech content (Prieto et al., 2018; Rohrer et al., 2021; Shattuck-Hufnagel & Ren, 2018). Our definition of non-referential gestures does not only limit to McNeill's (1992) beat gestures, traditionally described as up-and-down or in-and-out flicks that act as rhythmic markers, such that "the hand moves along with the rhythmical pulsation of speech" (p. 15). In fact, in this study, we adopt a recent and inclusive view of non-referential gestures, such that they can have a more complex phasing structure and be produced with different hand shapes (Prieto et al., 2018; Shattuck-Hufnagel & Ren, 2018; Shattuck-Hufnagel & Prieto, 2016).

Despite the claim that non-referential (beat) gestures are typically associated with prosodic prominence, research on gesture–speech association has argued that both referential and non-referential gestures occur in synchrony with speech prominence across languages (for non-referential gestures, see Krahmer & Swerts, 2007; Leonard & Cummins, 2011; Loehr, 2007; Shattuck-Hufnagel et al., 2007, 2016; for both referential and non-referential gestures, see Kendon, 2004; Pouw & Dixon, 2019a, 2019b; Shattuck-Hufnagel & Ren, 2018). In one of the few studies that report the temporal alignment patterns of both gestures separately, Shattuck-Hufnagel and Ren (2018) pointed out that in an academic-lecture-style discourse, strokes of both referential and non-referential gestures co-occurred with pitch-accented syllables in around 83% of the cases (82.85% for referential gestures and 83.13% for non-referential gestures). Also, in line with these results, Pouw and Dixon (2019b) conducted a kinematic analysis of gesture–speech synchrony (specifically taking into account pitch peaks and various kinematic aspects of the manual gesture, such as onset, maximum velocity, maximum acceleration, and maximum deceleration). The study revealed that maximum velocity and deceleration points were closely associated with pitch peaks and found no significant differences by gesture type in terms of how they were associated with pitch peaks. Despite these findings, a little amount of research has been devoted to comparing the differences in temporal alignment patterns between referential and non-referential gestures from a developmental perspective and, specifically, in more elaborated and complex discourses such as narrative productions.

## Temporal Synchronization Between Gesture and Speech in Development

Gesture–speech synchronization has been assessed in infants' early speech. At around 11 and 12 months of age, infants start to produce pointing gestures in spontaneous interactions (Butcher & Goldin-Meadow, 2000; Goldin-Meadow & Butcher, 2003; Murillo & Belinchón, 2012), which have been shown to be temporally aligned with speech, both at the word/vocalization level (Butcher & Goldin-Meadow, 2000; Goldin-Meadow & Butcher, 2003) and at the prominent vocalization/pitch accentuation level (Esteve-Gibert & Prieto, 2014; Murillo & Capilla, 2016; Murillo et al., 2018).

More specifically, these previous studies have found that during the one-word period (i.e., between 12 and 18 months of age), children started producing referential gestures that were semantically synchronized with meaningful

[1]The gestural apex (or "hit") has adopted different definitions according to its kinematic characteristics, for example, the peak or abrupt stop of the stroke (Loehr, 2007; Yasinnik et al., 2004) or the equilibrium of the gestural movement (Jannedy & Mendoza-Denton, 2005). However, from a kinematic perspective, these definitions lack clarity, since they are not based on kinematic measures such as maximum velocity, maximum acceleration, or maximum deceleration.

words, an aspect that was further developed during the two-word period (i.e., between 18 and 26 months of age; Butcher & Goldin-Meadow, 2000; Goldin-Meadow & Butcher, 2003). The abovementioned results are comparable to Esteve-Gibert and Prieto's (2014), who examined how pointing gestures and speech prominence co-occurred in development in a longitudinal study (age frame: between 11 months and 1 year and 7 months). Results showed that 11-month-old infants were able to coordinate the stroke of a pointing gesture with pitch-accented syllables and that the number of gesture–speech combinations increased at 15 months of age. These results are consistent with Murillo and Capilla's (2016) study, which showed that infants aged 9 months to 1 year and 3 months produced gestures accompanied by vocalizations in spontaneous speech during a play situation with a caregiver and that these gestures had specific declarative and imperative functions. Moreover, their results indicated that the duration and the fundamental frequency ($f_o$) patterns of these vocalizations were similar to the syllables produced during mature speech (i.e., canonical syllables). Importantly, this similarity was only found when the vocalizations were produced together with gestures with pointing and declarative functions, whereas no differences were found when these vocalizations were produced without gesture or with other gesture functions (e.g., imperative function). Similar findings were found in Murillo et al.'s (2018) study, which showed that combinations of gesture strokes and prominent vocalizations (i.e., maximum $f_o$) increased between the ages of 9 and 15 months (see Murillo & Belinchón, 2012; Murillo et al., 2021, for similar results on gesture–speech synchronization).

Despite the evidence provided for early patterns of gesture–speech temporal association in infancy, less is known about those patterns in children's narratives. So far, to our knowledge, only one study has examined the temporal alignment patterns between gesture and pitch accentuation in children's narrative discourse during the school-age period. Mathew et al. (2018) explored the temporal and lexical association of non-referential gestures (what they describe as "stroke-defined beats," adapting McNeill's proposal in 1992) in the discourses of twelve 6-year-old children by assessing whether the gesture apex falls within pitch-accented words. Findings from nine children who produced stroke-defined beats revealed that across both narrative and exposition tasks, 63% of the apexes of stroke-defined beats produced fell within a pitch-accented word. Since the overall results showed no systematic alignment between non-referential gestures and pitch accents, the authors highlight the high variability among children and conclude that "children might not yet have established a close link between the use of stroke-defined beats (i.e., manual channel) and pitch accent (i.e., verbal channel)" at this time period (Mathew et al., 2018, p. 125).

All in all, studies assessing the development of gesture–speech temporal alignment patterns have shown that while the link between gesture (specifically pointing gestures) and pitch accentuation seems to emerge early in development, that is, from the start of vocalization and word productions (e.g., Esteve-Gibert & Prieto, 2014; Murillo & Capilla, 2016; Murillo et al., 2018), the temporal alignment between non-referential gestures and pitch accentuation is not clear in childhood. In our view, a good point in time to assess the temporal alignment properties of both referential and non-referential gestures in development is the start of children's early complex narratives. It has been shown that at around 26 months of age, there is a spurt in iconic gesture production (Özçalişkan & Goldin-Meadow, 2011) and that the production of the first spontaneous non-referential gestures at around 2–3 years of age is related to more complex utterances (Nicoladis et al., 1999). However, non-referential gestures generally start to appear in complex narrative discourses around the ages of 4–6 years (e.g., Colletta et al., 2010, 2015; Graziano, 2009; McNeill, 1992), serving important linguistic pragmatic and structuring functions in discourse (Graziano, 2009; McNeill, 1992; Rohrer et al., 2022; see Vilà-Giménez & Prieto, 2021, for a systematic review). Specifically, a recent study found that non-referential gestures play a key role in marking information structure in children's narrative speech and that this relationship intensifies at a period in development that coincides with a spurt in non-referential gesture production, at around 7–9 years of age (Rohrer et al., 2022).

Therefore, in general, more research is needed to further assess the patterns of gesture–speech temporal alignment in children's discourse within a wider developmental perspective and by taking into account the potential differences between referential and non-referential gestures. Given the evidence that the production of non-referential gestures starts to increase as more complex language skills develop, for example, complex narratives, this study will take this opportunity and focus on systematically assessing the temporal gesture–speech alignment patterns in children's narratives.

## Goals of the Study

Following a longitudinal approach, this study aims to assess how children temporally align referential and non-referential gesture strokes with pitch-accented syllables in narrative discourse from a developmental view. Specifically, the study aims to assess potential differences between these two gesture types by using the longitudinal database, "Audiovisual Corpus of Catalan Children's Narrative Discourse Development" (Vilà-Giménez, Florit-Pons, et al., 2021). Specifically, we will assess whether any differences in these gesture–speech temporal association patterns emerge

when comparing both referential and non-referential gestures. The ages under scrutiny are 5–6 and 7–9 years, which is a key age period when children start producing linguistically complex narratives that actively involve both referential and non-referential gestures (e.g., Colletta et al., 2015; Graziano, 2014a, 2014b; McNeill, 1992; Rohrer et al., 2022).

The following hypotheses will be tested. First, we hypothesize that at 5–6 years of age, children will align gestures with pitch accentuation in their narrative productions. Our hypothesis is based on evidence about early gesture–speech temporal alignment (e.g., Esteve-Gibert & Prieto, 2014; or the studies by Murillo & Capilla, 2016; Murillo et al., 2018) and on the general predictions of McNeill's (1992) synchronization rule. We also predict that the percentage of gesture and pitch accentuation alignment patterns will increase as complex language develops over time (e.g., Colletta et al., 2010, 2015). Finally, we hypothesize that not only will non-referential gestures tend to align with prosodic prominence—since they have been historically defined by their rhythmic properties (e.g., McNeill, 1992)—but so will referential gestures, based on the findings from most recent research on adult speech (e.g., Shattuck-Hufnagel & Ren, 2018).

## Method

### Corpus

Given that the goal of this study is to analyze the developmental patterns of temporal alignment between both referential and non-referential gestures and pitch accentuation in children's narrative speech, a longitudinal design was used. The "Audiovisual Corpus of Catalan Children's Narrative Discourse Development" (Vilà-Giménez, Florit-Pons, et al., 2021) was employed, which is composed of narratives performed by 83 children (43 girls and 40 boys) at two time points in development: at the ages of 5–6 years (Time 1: $M_{age}$ = 5.9 years, $SD$ = 0.55) and 7–9 years (Time 2: $M_{age}$ = 7.98 years, $SD$ = 0.60).

At both time points, children performed a narrative retelling task. They were asked to watch two wordless cartoons (*Die Sendung mit der Maus*, accessible at https://www.wdrmaus.de, approximate length: 41–50 s) and to retell them to the experimenter, who was unfamiliar to the children. The task was presented as a game, such that the experimenter pretended not to have watched the story or know the plot, and she had to guess the story the child had retold based on a set of pictures. Each participant was randomly assigned two stories to retell at Time 1 and the same two stories at Time 2, with the only constraint being that at each time, the first story they had to retell had only one character—the mouse—and the second story had two characters—the mouse and the elephant. Two years later, the same children did the same task with the same cartoons they had watched at Time 1. For further details on the experimental procedure, see the description of the corpus by Vilà-Giménez, Florit-Pons, et al. (2021).

All in all, the corpus contains narratives produced by 83 children, corresponding to two narratives per child at each time point. Nevertheless, one narrative was excluded given that the retelling was not video-recorded due to technical issues. For this study, the total number of narratives analyzed was 331. Narratives had an average length of 27.1 s ($SD$ = 10.45) at Time 1 and 28.4 s ($SD$ = 9.49) at Time 2.

### Data Coding

Each narrative was first transcribed orthographically and coded for manual gestures in ELAN (Sloetjes, 2017). That is, the phases of all manual gestures and their referential/non-referential status (i.e., gesture type) were annotated. After that, each narrative was prosodically coded in terms of pitch accentuation in Praat (Boersma & Weenink, 2019) and was then imported into the previous ELAN file that contained the previously mentioned annotations. The following subsections first describe how children's gestures and pitch accentuation patterns were annotated. Second, a description of their subsequent gesture-speech temporal analysis follows.

#### Gesture Coding

All manual gestures in the database were annotated for gesture phases and referentiality. As Kendon (1980) and McNeill (1992) claim, manual gestures are communicative, complex, and extensive hand movements that constitute part of the speaker's communicative acts. All manual gestures were annotated in two steps.

First, an annotation of the video file without audio was performed to identify the phases of each gesture. For each gesture unit, the first and second authors of this study, together with an external annotator,[2] coded the different gesture phases (Kendon, 1980; McNeill, 1992). The only obligatory phase is the *stroke*, which is the most prominent movement of the gesture and contains the gesture apex. Other gesture phases include *preparation* (i.e., movement found before the stroke in which the hand leaves the rest position and reaches the position to start the stroke), *holds* (i.e., minimal movements in which the hand position and form stay intact, either before or after the stroke), and *recovery* (i.e., phase in which the hand moves back to the rest position).

[2]Please refer to Vilà-Giménez, Florit-Pons, et al. (2021) to see a detailed description of the annotation process of the corpus.

**Figure 1.** Example of the phasing of a referential iconic gesture produced while saying, "Va agafar una poma" ("He grabbed an apple").


In a second step, an annotation of the video file with audio was performed to identify the referentiality of each gesture stroke, as being either referential or non-referential in nature (Kendon, 1980; McNeill, 1992). On the one hand, referential gestures were identified as those that have a close relationship with what is being described in discourse (e.g., an action or entity). We particularly included *referential deictic* gestures, which indicate concrete or spatial positions, and *referential iconic* gestures, which pictorially represent the semantic content in speech. *Referential metaphoric* gestures were not included in our study, given that only one was produced in the whole database. *Non-referential* gestures, on the other hand, were identified as not representing any propositional content in discourse, following an inclusive definition, such that they can have different hand shapes and can be produced coordinated with other articulators (Prieto et al., 2018; Rohrer et al., 2021; Shattuck-Hufnagel & Prieto, 2019; Shattuck-Hufnagel & Ren, 2018). Figures 1 and 2 illustrate the phases of a referential iconic gesture and a referential deictic gesture, respectively, whereas Figure 3 illustrates a non-referential gesture.

## Pitch Accentuation Coding

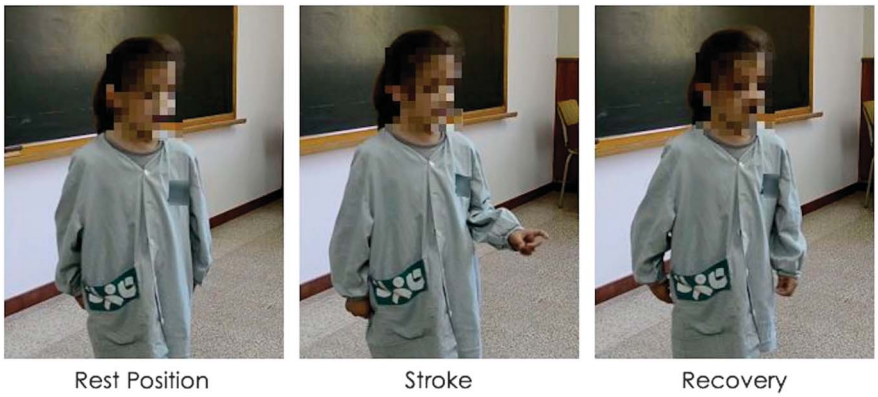
All narratives were prosodically coded using Praat (Boersma & Weenink, 2019) by one external annotator.[3] The prosodic annotations were conducted on the audio files independently from the gestural annotations, so that the annotators were blind to the potential presence of a gesture, as gestures have been shown to influence the perception of prominence (Krahmer & Swerts, 2007). The tier with the orthographic transcription in ELAN was imported into Praat. For pitch accentuation, the Cat_ToBI labeling system, which is the labeling system for Catalan prosody, was followed (Prieto et al., 2015). Particularly, for all utterances, words that received phrasal prominence were identified and segmented into syllables. Then, a pitch accent was annotated within the most prominent syllable of the word. Once the coding was performed, all prosodic annotations were imported from Praat into the corresponding ELAN file.

## Gesture–Speech Temporal Alignment Criteria

For this study, we have used a binary distinction (aligned vs. non-aligned), following McNeill's (1992) phonological synchronization rule as well as Shattuck-Hufnagel and Ren's (2018) definition of gesture–speech alignment. This definition postulates that gesture and speech align when there is "any degree of overlap between the temporal region labeled as an accented syllable and the region labeled as the gestural stroke" (Shattuck-Hufnagel & Ren, 2018, p. 6). The main reasons to use a discrete measure have been (a) that it has often been used by previous research on this topic (e.g., Im & Baumann, 2020; Mathew et al., 2018; Yasinnik et al., 2004) and (b) because this binomial measure helps us assess in a straightforward way whether gestures and pitch-accented syllables are temporally aligned and, thus, whether pitch-accented syllables can be considered "prosodic attractors" for gesture production. All in all, we assessed gesture–speech alignment by looking at the overlap between the stroke and a pitch-accented syllable. Thus, if there was a temporal overlap between a gesture stroke and pitch-accented syllables, the gesture was coded as "aligned," and if there was no temporal alignment, the gesture was coded as "non-aligned." Previous studies on gesture–speech temporal alignment have also used non-discrete measurements to assess the alignment between gestural occurrences and prosodic prominence, such as the temporal distance between the

[3]See Vilà-Giménez, Florit-Pons, et al. (2021).

**Figure 2.** Example of the phasing of a referential deictic gesture produced while saying, "Hi havia un animal" ("There was an animal").




gesture apex and prosodic prominence annotations (though with variable criteria, such as adopting an arbitrary time window, e.g., 275 ms in Loehr, 2007) or by assessing the temporal distances between various kinematic points in gesture (e.g., onset, peak velocity, point of maximum extension or peak deceleration) and landmarks in speech, such as vowel onsets or pitch peaks (e.g., Leonard & Cummins, 2011; Pouw & Dixon, 2019b). Despite the fact that these measures are precise in terms of assessing specific distances between landmarks, it is more difficult to use them to assess whether gestures are aligned with pitch-accented syllables or not. Also, these measures typically report a good amount of variability; for instance, Loehr (2007) found a standard deviation of more than 300 ms when looking at the temporal distances between apex and pitch-accented syllable annotations, which corresponds to the duration of the average word in his database.

Before conducting any analysis, we applied the following three specific exclusion criteria. First, gestures produced together with disfluent speech (e.g., repetitions, breaks, or speech accompanied by non-lexical words such as interjections) were excluded from the analysis. Second, since the goal is to assess how gesture is temporally aligned with pitch accentuation in discourse, gestures that did not co-occur with stretches of speech production were excluded. Moreover, complex gestures involving multiple apexes (i.e., gestures that have a multidirectional stroke and each of these apexes marks a change of velocity and direction, as per Loehr, 2007) were excluded, as the strokes of these gestures tend to be longer in duration and can even align with multiple pitch-accented syllables, thus biasing the alignment results. The three exclusion criteria resulted in the elimination of 35% of the gestures (a total of 314 gestures, 108 at Time 1 and 206 at Time 2) out of the total number of 896 gestures produced (296 at Time 1 and 600 at Time 2). Of this 35%, 68.47% correspond to gestures accompanied by disfluent speech ($n$ = 215), 24.74% correspond to complex gestures ($n$ = 78), and 6.69% correspond to gestures produced without speech ($n$ = 21). All in all, 65% of the gestures from the initial

**Figure 3.** Example of the phasing of a non-referential gesture produced while saying, "Hi havia un ratolí" ("There was a mouse").

896 were included in the final analysis (i.e., a total of 582 gestures, 188 at Time 1 and 394 at Time 2).

## Inter-Annotator Reliability

### Inter-Annotator Reliability for Gesture Types

Inter-annotator reliability was calculated for gesture classification with 64 narratives from the database (32 from Time 1 and 32 from Time 2), which represented around 20% of the corpus. These 64 narratives were annotated by the third co-author who was already familiar with the coding system used. The 64 narratives included a total of 147 gestures in which the experienced annotator had to assign the referentiality for each gesture stroke (i.e., referential or non-referential).

For the reliability analysis, we calculated percent agreement and Gwet's agreement coefficient 1 (AC1; Gwet, 2008) using the *irrCAC* package in R (Gwet, 2019). We used Gwet's AC1 for reliability measures in order to resist the kappa paradox (where the kappa statistic can be reduced despite high agreement due to unbalanced totals of the categories; see, e.g., Cicchetti & Feinstein, 1990; Feinstein & Cicchetti, 1990). Gwet's agreement coefficient can be interpreted similarly to kappa. The agreement analysis showed that agreement for gesture type was good, percent agreement: 78.2%; AC1 = .731 (95% CI [.648, .815]), $p < .001$.

### Inter-Annotator Reliability for Pitch Accentuation

Inter-annotator reliability was also calculated for pitch accentuation using 64 different narratives (32 from Time 1 and 32 from Time 2). The pitch accentuation of these 64 narratives was coded by two different annotators: the first author and an external annotator (see the Pitch Accentuation Coding section). The two of them participated in a 2-hr training session, in which they covered the Cat_ToBI labeling system: Pitch accentuation was annotated within the syllables of words that received phrasal prominence. The reliability analysis was run considering the presence or absence of pitch accent. Results indicated that agreement was almost perfect, percent agreement: 94.7%; AC1 = .942 (95% CI [.903, .976]), $p < .001$.

## Statistical Analyses

To assess the aim of this study, that is, how do both referential and non-referential gestures align with pitch accentuation in the children's narratives and how these patterns evolve over time, a generalized linear mixed model (GLMM) with Poisson distribution and controlling for zero-inflated count data was run in R (R Core Team, 2021), using the *glmmTMB* package (Brooks et al., 2017). The analysis assessed the patterns of temporal alignment between gesture and pitch accentuation across times and gesture types using the total number of gestures as the dependent variable. Time (two levels: Time 1 and Time 2), gesture type (two levels: referential and non-referential), and gesture alignment (two levels: aligned and non-aligned) were set as fixed factors, along with their two-way and three-way interactions. The random-effects structure included by-participant varying intercepts, and the model was adjusted using Bonferroni correction for multiple comparisons. We used a zero-inflated model because the database was distributed with a high number of zero-valued observations. A total of five participants were excluded from the analysis given that they did not produce any gesture at Time 1 or at Time 2. Therefore, the analysis consisted of the gestures produced by 78 participants. Finally, using the *emmeans* package (Lenth, 2021), we carried out post hoc pairwise comparisons with Bonferroni correction.

A visual inspection of the data revealed that a large majority of the gestures were aligned. So as to better understand and interpret the interactions between gesture alignment, gesture type, and time, the model was rerun offset for the total number of aligned or non-aligned gestures. By doing so, the model takes the relative frequency of occurrence into account, as opposed to the raw counts.

In order to answer our research question about how gesture–speech alignment develops from Time 1 to Time 2 and to assess potential differences between gesture types, we conducted two equivalence tests to correctly interpret the potential null findings that may come up from the inferential statistical analyses. For the equivalence tests, we used the *equivalence* package (Robinson, 2016).

# Results

In this study, a total of 188 gestures at Time 1 (116 referential and 72 non-referential) and 394 gestures at Time 2 (173 referential and 221 non-referential) were analyzed. More specifically, children produced a mean of 1.21 gestures per narration at Time 1 ($SD = 2.00$; range: 0–12) and 2.53 at Time 2 ($SD = 2.76$; range: 0–15).

The results of the GLMM analysis for assessing gesture–speech temporal alignment showed a significant main effect of time, $\chi^2(1) = 64.741$, $p < .0001$, which suggested that there were more gestures at Time 2 than at Time 1, $t(614) = -4.608$, $SE = 0.172$, $p < .0001$, and also a main effect of gesture alignment, $\chi^2(1) = 61.975$, $p < .0001$, indicating that the number of aligned gestures was significantly larger than the number of non-aligned gestures, $t(614) = -7.447$, $SE = 0.234$, $p < .0001$. No main effect was observed for gesture type ($p = .76$).

The two-way interaction between time and gesture type was found to be significant, $\chi^2(1) = 17.2835$, $p < .0001$. This interaction showed that there was a significant increase from Time 1 to Time 2 in the use of non-referential gestures,

$t(614) = -5.383$, $SE = 0.215$, $p < .0001$. The two-way interactions between time and gesture alignment ($p = .488$) and between gesture type and gesture alignment ($p = .0835$) were not found to be significant. Furthermore, the three-way interaction between time, gesture type, and gesture alignment was not found to be significant ($p = .8024$).

The non-significant results from the interactions involving gesture alignment suggest that gesture alignment patterns are not modulated by gesture type, nor does it change across time. To further assess these non-significant results, we ran two complementary equivalence tests to correctly interpret these null findings: one to compare the potential differences between gesture types and another one for the comparison between the two time points (i.e., Time 1 and Time 2). First, the equivalence test for comparing gesture type differences was run using alignment data from Time 1 and Time 2 together. The test was found to be significant ($p = .036$; 95% CI [−.41, .938]), which meant that we can reject the null hypothesis and assume that there are no differences in how referential and non-referential align with pitch accentuation. Second, the equivalence test for assessing the null findings for time was run with all gestures, considering the two time points separately. The test was also found to be significant ($p < .001$; 95% CI [−.167, .538]), which indicated that there are no differences between the two time points in development in how gestures align with pitch accentuation.

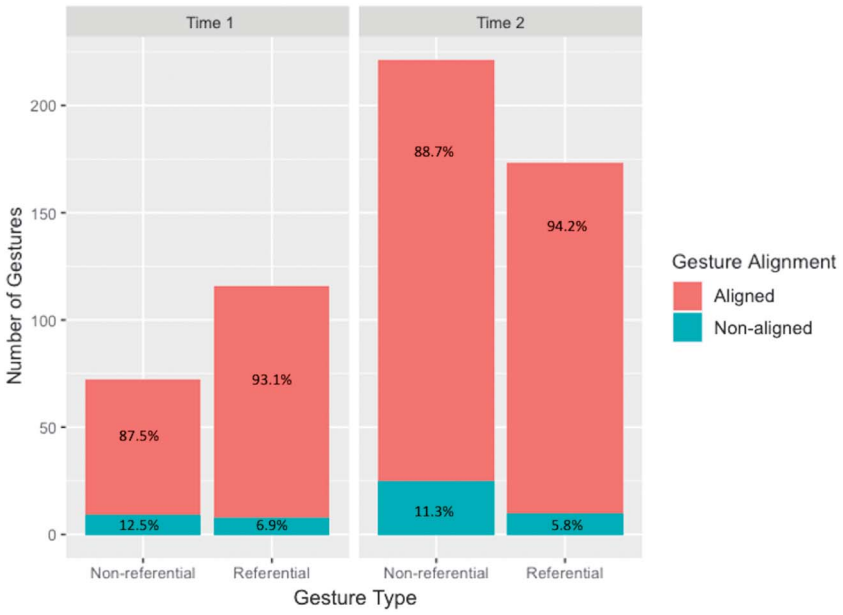
Figure 4 shows the percentages of alignment for referential and non-referential gestures across times. The graph visually displays the systematicity in the alignment patterns across gesture types already at Time 1, which, importantly, are remaining stable at Time 2.

## Discussion

This article reports on the results of a longitudinal investigation that has the goal of assessing the gesture–speech temporal alignment patterns found in children's multimodal narrative productions. The database analyzed contained four narratives uttered by a total of 83 children at two time points in development, namely, at the ages of 5–6 years and, 2 years later, 7–9 years. The main objective of the study was to assess the development of the alignment patterns found between gestures and pitch accentuation, specifically focusing on the differences between referential and non-referential gestures. To our knowledge, this study is the first to investigate this issue from a longitudinal perspective that involves a time span where children generally start producing more complex narratives and that considers and compares the temporal alignment behavior of both referential and non-referential gestures. In general, the results revealed a stable temporal association between the stroke of gestures and pitch-accented syllables across these two time points in development (ages 5–6 years and 7–9 years), showing no differences between gesture types (i.e., referential and non-referential gestures).

The findings of the study help refine our knowledge about gesture–speech temporal alignment patterns in children's narrative speech. Our results revealed that by the age

**Figure 4.** Number of gestures produced at Time 1 and Time 2 by gesture referentiality and alignment. The relative frequency (%) of alignment is indicated for each gesture type at each time.



of 5–6 years (i.e., when children start to produce more elaborated narratives), they produce gestures that are highly aligned with pitch-accented syllables, regardless of gesture type (in our data, 87.5% of non-referential gestures and 93.1% of referential gestures were temporally aligned). Given that the results of previous studies on gesture–speech temporal alignment across languages have reported high levels of overlap between strokes and pitch accents (between 80% and 90% of alignment in English, as reported by Shattuck-Hufnagel & Ren, 2018, and Shattuck-Hufnagel et al., 2016), our results suggest that when 5- to 6-year-old children start producing complex narratives, they have already established an adultlike tight temporal alignment between pitch accentuation and manual gestures.

Importantly, our findings contrast with the results reported in Mathew et al.'s (2018) study, which assessed children's gesture–speech temporal alignment patterns in narrative speech in the same developmental window (5–6 years of age). The authors suggested that 6-year-old children had not yet established a strong alignment between gestural and prosodic prominence. By contrast, in this study, we found greater rates of alignment in non-referential gestures (87.5%) than in Mathew et al.'s study (63%). The contrast between these findings and the gesture–speech temporal alignment patterns reported in our study might have been due to two main reasons. First, the sample sizes of both studies might have had an influence on the results. While the study by Mathew et al. consisted of 12 participants, out of which only nine produced non-referential gestures, in this study, we analyzed the gestures of 78 participants, out of which 38 produced stories with non-referential gestures at Time 1 and 67 produced narrations with non-referential gestures at Time 2. Second, the treatment of data variability within the databases is another potential reason for differences between the two studies. While, in our study, by-participant variability was controlled for (e.g., the fact that some children barely gesture while other children produce more than 10 gestures during a 30-s narrative retelling) in the statistical model's random-effects structure, this was not the case in Mathew et al.'s study, which used a Wilcoxon test. At this point, we would like to point out that a direct comparison between our database and Mathew et al.'s database is justifiable, given that our exclusion criteria for analyses were similar, namely, the exclusion of gestures produced with unintelligible or disfluent speech. The only exclusion criterion that did not coincide in the two studies was that related to complex gestures. While we decided to exclude these gestures, Mathew et al. do not mention directly complex gestures, but mention that they excluded gestures that did not have a well-defined stroke, such as circular movements. Along this line of reasoning, in order to assess whether including such complex gestures would trigger a change in our results, we ran a second analysis including both simple and complex gestures (still excluding disfluent gestures and gestures produced without speech). The results showed that the alignment percentages were similar, namely, 88% for non-referentials and 94.2% for referentials at Time 1 and 89.2% for non-referentials and 95.3% for referentials at Time 2. All in all, we believe that the differences between our findings and those of Mathew et al. are largely due to the number of participants and the treatment of individual variability.

An important finding of this study is that both referential and non-referential gestures have been shown to behave similarly with respect to gesture–speech temporal alignment patterns. This finding has implications for the characterization of gesture types. Although non-referential (beat) gestures have been traditionally described to be temporally aligned with pitch accentuation (McNeill, 1992), evidence from recent adult studies has also reported that referential gestures are linked to prosodic prominence in similar ways as non-referential gestures (Pouw & Dixon, 2019a, 2019b; Shattuck-Hufnagel & Ren, 2018). Therefore, the results of the present investigation are in line with these findings in adult speech, as no significant differences in temporal alignment were found between these two gesture types in children's narrative productions.

From a developmental perspective, the result that 5- and 6-year-old children systematically align the gesture stroke with pitch accentuation in their early complex narratives is coherent with reports of initial gesture–speech temporal alignment in infants' spontaneous speech (e.g., Esteve-Gibert & Prieto, 2014; Murillo & Capilla, 2016; Murillo et al., 2018). Furthermore, in our view, the findings of this study about the tight temporal alignment patterns reported for referential and non-referential gestures in children's early complex narratives seem to adhere to this theoretical view of gesture–speech alignment as a biomechanical ability of human communication. Our results seem to reinforce the idea postulated by previous research that there might be an inherent motor and physical capacity already present from early infancy that allows gesture–speech alignment to happen so stably in childhood (e.g., Pouw & Fuchs, 2022; Pouw et al., 2019; Rusiewicz, 2011). Interestingly, recent investigations have initiated a new debate on the physical and biomechanical foundations of gesture–speech synchronization patterns that are based on a set of empirical arguments, such as that babies' first motor and oral babbling are also temporally synchronized, as well as babbling and limb movements (Pouw et al., 2019). Also, Ejiri and Masataka (2001) showed that the vocalizations of 6- to 11-month-old babies were temporally synchronized with their rhythmic movements (e.g., swinging the arms up and down). Moreover, recent work has shown that the temporal production of gestures is coupled with respiratory dynamics, specifically the tight coordination patterns between the peak of vertical up-and-down wrist and arm movements and the peaks in the

amplitude envelope expressing greater respiration activity (Pouw, Harrison, et al., 2020; see Pouw & Fuchs, 2022, for a review). All these results are in line with motor-based gesture theories, such as the dynamic systems theory (e.g., Iverson & Thelen, 1999) or the Theory of Entrained Manual and Speech Systems (see Rusiewicz, 2011), which argue that the underlying mechanism of coupling and/or synchronization of speech and gesture is due to internal temporal entrainment of the pulses planned and produced by the two motor systems, that is, speech and gesture. All in all, while we believe that part of the reason behind the tight alignment rates found in early narratives can be due to the strong biomechanical basis of speech, at the same, time, children need to learn language-specific patterns of multimodal language production, for example, how to use both gesture and pitch accentuation in a pragmatically adequate manner, such as using gestures to mark new information in discourse (e.g., Rohrer et al., 2022) or using a different pitch accent type to show contrast (Chen, 2010).

In addition, it is important to mention that when analyzing the temporal alignment patterns between gesture and pitch accentuation in children's narratives, we also observed an important increase in gesture production. Therefore, we believe that the results of our study also help expand our knowledge of children's use of referential and non-referential gestures in narrative speech across development. Our findings confirm the results of previous research that has reported that by the age of 4–6 years, children already produce non-referential gestures in narrative discourse and that they show a strong increase in production over the upcoming years (e.g., Colletta et al., 2010, 2015; Graziano, 2009). Specifically, our database revealed that the production of non-referential gestures increases significantly from ages 5–6 years to 7–9 years, suggesting a spurt of non-referential gesture production that occurs during this period of development (see also Rohrer et al., 2022, for a similar conclusion from a pragmatic analysis of this same database).

This study has left some open questions for future work. First of all, although we have compared our findings with the results on English-speaking adults, it is important to bear in mind that, to our knowledge, there are no data about gesture–speech temporal alignment in Catalan-speaking adults that we can compare to. Therefore, future work should assess how Catalan-speaking adults align gestures with pitch accentuation in narrative speech. Moreover, future research is also needed to analyze speakers of different languages to control for potential crosslinguistic differences in gesture production (see, e.g., Pouw, Jaramillo, et al., 2020). Second, even though the results of this study showed that gestures are overwhelmingly aligned with pitch-accented syllables, we believe that these findings could be complemented with more precise assessments of temporal distances between various landmarks in speech and gesture to gain a fuller picture of the temporal alignment between gesture and speech (e.g., Loehr, 2007; Pouw & Dixon, 2019b). Third, a further assessment about individual differences across children would be needed. Also, we believe that it would be interesting to assess whether gesture use and gesture–speech temporal alignment patterns in school-age children have a direct relationship with narrative skills. For instance, Vilà-Giménez, Dowling, et al. (2021) documented that the frequency of use of non-referential gestures from younger children (14–58 months of age) during spontaneous naturalistic interactions with their caregivers predicted better narrative ability later in development (see also Vilà-Giménez et al., 2020, for similar predictive results with referential iconic gestures). Furthermore, the study by Rohrer et al. (2022), who used the same longitudinal audiovisual corpus as in this study, found that the children's non-referential gestures served important discourse-pragmatic functions in narrative speech, such that they were found to mark information that updated speakers' common ground (in terms of information structure marking, namely, focus, predication, and new referents).

In conclusion, this study adds important insights into our knowledge about children's multimodal development in narrative speech. Our investigation reinforces McNeill's (1992) phonological synchronization rule, by showing evidence that temporal patterns of alignment between gesture (both referential and non-referential types) and pitch accentuation are overwhelmingly present in early complex narratives produced by 5-year-old children.

## Data Availability Statement

The data sets generated and analyzed during the current study are available in the Open Science Framework repository, https://osf.io/y7n3g/. This project has been conducted using data from the "Audiovisual Corpus of Catalan Children's Narrative Discourse Development" (Vilà-Giménez, Florit-Pons, et al., 2021). The annotated files generated in ELAN (Sloetjes, 2017) as well as a detailed description of the longitudinal audiovisual corpus are also available in the Open Science Framework repository, https://osf.io/npz3w/. Due to ethical issues related to audiovisual sharing of children's data, raw audiovisual recordings used in this study can only be made available upon specific and reasonable request to the first and second authors of this study.

## Acknowledgments

The authors would like to acknowledge the financial support awarded by the Spanish Ministry of Science, Innovation and Universities, Agencia Estatal de Investigación,

and Fondo Europeo de Desarrollo Regional (PGC2018-097007-B-100: "Multimodal Language Learning: Prosodic and Gestural Integration in Pragmatic and Phonological Development," and PID2021-123823NB-I00: "Multimodal Communication: The Integration of Prosody and Gesture in Human Communication and in Language Learning"); by the Generalitat de Catalunya (2017 SGR_971 and 2021 SGR_922); and by the GEstures and Head Movements in language research network, funded by the Independent Research Fund Denmark (9055-00004B). The first author would also like to acknowledge a PhD grant awarded by the Agency for Management of University and Research Grants from the Generalitat de Catalunya, (Award no. 2021 FI_B 00778), and the third author would like to acknowledge a joint PhD grant awarded by the Department of Translation and Language Sciences, Universitat Pompeu Fabra, and SGR AGAUR Grant, Generalitat de Catalunya (Award no. 2017 SGR_971). Also, the second author would like to acknowledge a postdoctoral fellowship funded by European Union-NextGenerationEU, Ministry of Universities and Recovery, Transformation and Resilience Plan, through a call from Universitat Pompeu Fabra (Barcelona). The authors would like to express their gratitude to Glenda Gurrado (Università degli Studi di Bari Aldo Moro) for her help in the gestural coding and to Sara Coego (Universitat Pompeu Fabra) for her help with the annotation of pitch accentuation. Many thanks are also extended to Stefanie Shattuck-Hufnagel (Massachusetts Institute of Technology), Ada Ren (Massachusetts Institute of Technology), Núria Esteve-Gibert (Universitat Oberta de Catalunya), Maria Graziano (Lund University), Mili Mathew (St. Cloud State University and Emerson College), and Alfonso Igualada (Universitat Oberta de Catalunya) for their feedback provided on early versions of this study.

## References

**Boersma, P., & Weenink, D.** (2019). *Praat: Doing phonetics by computer* (Version 6.1.08) [Computer program]. http://www.praat.org/

**Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M.** (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal, 9*(2), 378–400. https://doi.org/10.32614/RJ-2017-066

**Butcher, C., & Goldin-Meadow, S.** (2000). Gesture and the transition from one- to two-word speech: When hand and mouth come together. In D. McNeill (Ed.), *Language and gesture* (pp. 235–258). Cambridge University Press. https://doi.org/10.1017/CBO9780511620850.015

**Chen, A.** (2010). Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping? *Lingua, 120*(8), 1926–1939. https://doi.org/10.1016/j.lingua.2010.02.012

**Cicchetti, D. V., & Feinstein, A. R.** (1990). High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology, 43*(6), 551–558. https://doi.org/10.1016/0895-4356(90)90159-M

**Colletta, J. M., Guidetti, M., Capirci, O., Cristilli, C., Demir-Lira, Ö. E., Kunene-Nicolas, R. N., & Levine, S.** (2015). Effects of age and language on co-speech gesture production: An investigation of French, American, and Italian children's narratives. *Journal of Child Language, 42*(1), 122–145. https://doi.org/10.1017/S0305000913000585

**Colletta, J. M., Pellenq, C., & Guidetti, M.** (2010). Age-related changes in co-speech gesture and narrative: Evidence from French children and adults. *Speech Communication, 52*(6), 565–576. https://doi.org/10.1016/j.specom.2010.02.009

**Ejiri, K., & Masataka, N.** (2001). Co-occurrences of preverbal vocal behavior and motor action in early infancy. *Developmental Science, 4*(1), 40–48. https://doi.org/10.1111/1467-7687.00147

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

**Feinstein, A. R., & Cicchetti, D. V.** (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology, 43*(6), 543–549. https://doi.org/10.1016/0895-4356(90)90158-L

**Goldin-Meadow, S., & Butcher, C.** (2003). Pointing toward two-word speech in young children. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 85–107). Erlbaum.

**Graziano, M.** (2009). *Rapporto fra lo sviluppo della competenza verbale e gestuale nella costruzione di un testo narrativo in bambini dai 4 ai 10 anni* [Relationship between the development of verbal and gestural competence in the construction of a narrative text in children aged 4 to 10 years] [Unpublished doctoral dissertation]. SESA – Scuola Europea di Studi Avanzati – Università degli Studi "Suor Orsola Benincasa," Napoli, Italy, and Université Stendhal.

**Graziano, M.** (2014a). The development of two pragmatic gestures of the so-called open hand supine family in Italian children. In M. Seyfeddinipur & M. Gullberg (Eds.), *From gesture in conversation to visible action as utterance: Essays in honor of Adam Kendon* (pp. 311–330). John Benjamins.

**Graziano, M.** (2014b). Gestures in Southern Europe: Children's pragmatic gestures in Italy. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body – Language – Communication: An international handbook on multimodality in human interaction* (Vol. 2, pp. 1253–1258).

**Gwet, K. L.** (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology, 61*(1), 29–48. https://doi.org/10.1348/000711006X126600

**Gwet, K. L.** (2019). *irrCAC: Computing chance-corrected agreement coefficients* (*CAC*) (R Package Version 1.0). https://cran.r-project.org/web/packages/irrCAC/index.html

**Im, S., & Baumann, S.** (2020). Probabilistic relation between co-speech gestures, pitch accents and information status. *Proceedings of the Linguistic Society of America, 5*(1), 685–697. https://doi.org/10.3765/plsa.v5i1.4755

**Iverson, J. M., & Thelen, E.** (1999). Hand, mouth and brain: The dynamic emergence of speech and gesture. *Journal of Consciousness Studies, 6*(11–12), 19–40.

**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). De Gruyter. https://doi.org/10.1515/9783110813098.207

**Kendon, A.** (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

**Krahmer, E., & Swerts, M.** (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language, 57*(3), 396–414. https://doi.org/10.1016/j.jml.2007.06.005

**Lenth, R. V.** (2021). *emmeans: Estimated marginal means, aka least-squares means* (R Package Version 1.5.5-1). https://cran.r-project.org/web/packages/emmeans/emmeans.pdf

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

**Loehr, D.** (2007). Aspects of rhythm in gesture and speech. *Gesture, 7*(2), 179–214. https://doi.org/10.1075/gest.7.2.04loe

**Mathew, M., Yuen, I., & Demuth, K.** (2018). Talking to the beat: Six-year-olds' use of stroke-defined non-referential gestures. *First Language, 38*(2), 111–128. https://doi.org/10.1177/0142723717734949

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

**McNeill, D.** (2005). *Gesture and thought*. University of Chicago Press. https://doi.org/10.7208/Chicago/9780226514642.001.0001

**Murillo, E., & Belinchón, M.** (2012). Gestural–vocal coordination. *Gesture, 12*(1), 16–39. https://doi.org/10.1075/gest.12.1.02mur

**Murillo, E., & Capilla, A.** (2016). Properties of vocalization- and gesture-combinations in the transition to first words. *Journal of Child Language, 43*(4), 890–913. https://doi.org/10.1017/S0305000915000343

**Murillo, E., Montero, I., & Casla, M.** (2021). On the multimodal path to language: The relationship between rhythmic movements and deictic gestures at the end of the first year. *Frontiers in Psychology, 12,* 616812. https://doi.org/10.3389/fpsyg.2021.616812

**Murillo, E., Ortega, C., Otones, A., Rujas, I., & Casla, M.** (2018). Changes in the synchrony of multimodal communication in early language development. *Journal of Speech, Language, and Hearing Research, 61*(9), 2235–2245. https://doi.org/10.1044/2018_JSLHR-L-17-0402

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

**Özçalişkan, Ş., & Goldin-Meadow, S.** (2011). Is there an iconic gesture spurt at 26 months? In G. Stam & M. Ishino (Eds.), *Gesture studies: Vol. 4. Integrating gestures: The interdisciplinary nature of gesture* (pp. 163–174). John Benjamins. https://doi.org/10.1075/gs.4.14ozc

**Pouw, W., & Dixon, J. A.** (2019a). Entrainment and modulation of gesture–speech synchrony under delayed auditory feedback. *Cognitive Science, 43*(3), Article e12721. https://doi.org/10.1111/cogs.12721

**Pouw, W., & Dixon, J. A.** (2019b). Quantifying gesture-speech synchrony. In A. Grimminger (Ed.), *Proceedings of the 6th Gesture and Speech in Interaction (GESPIN) Conference* (pp. 75–80). Universitätsbibliothek Paderborn.

**Pouw, W., & Fuchs, S.** (2022). Origins of vocal-entangled gesture. *Neuroscience & Biobehavioral Reviews, 141,* 104836. https://doi.org/10.1016/j.neubiorev.2022.104836

**Pouw, W., Harrison, S. J., & Dixon, J. A.** (2019). Gesture–speech physics: The biomechanical basis for the emergence of gesture–speech synchrony. *Journal of Experimental Psychology: General, 149*(2), 391–404. https://doi.org/10.1037/xge0000646

**Pouw, W., Harrison, S. J., Esteve-Gibert, N., & Dixon, J. A.** (2020). Energy flows in gesture-speech physics: The respiratory-vocal system and its coupling with hand gestures. *The Journal of the Acoustical Society of America, 148*(3), 1231–1247. https://doi.org/10.1121/10.0001730

**Pouw, W., Jaramillo, J., Özyürek, A., & Dixon, J.** (2020). Quasi-rhythmic features of hand gestures show unique modulations within languages: Evidence from bilingual speakers. In *Proceedings of the 7th Gesture and Speech in Interaction Conference*. KTH Royal Institute of Technology.

**Prieto, P., Borràs-Comes, J., Cabré, T., Crespo-Sendra, V., Mascaró, I., Roseano, P., Sichel-Bazin, R., & Vanrell, M. M.** (2015). Intonational phonology of Catalan and its dialectal varieties. In S. Frota & P. Prieto (Eds.), *Intonation in Romance* (pp. 9–62). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199685332.003.0002

**Prieto, P., Cravotta, A., Kushch, O., Rohrer, P., & Vilà-Giménez, I.** (2018). Deconstructing beat gestures: A labelling proposal. In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), *Proceedings of the 9th International Conference on Speech Prosody* (pp. 201–205). Poznań. https://doi.org/10.21437/SpeechProsody.2018-41

**R Core Team.** (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

**Robinson, A.** (2016). *equivalence: Provides tests and graphics for assessing tests of equivalence* (R Package Version 0.7.2). https://cran.r-project.org/package=equivalence

**Rohrer, P. L., Florit-Pons, J., Vilà-Giménez, I., & Prieto, P.** (2022). Children use non-referential gestures in narrative speech to mark discourse elements which update common ground. *Frontiers in Psychology, 12*(661339), 6094. https://doi.org/10.3389/fpsyg.2021.661339

**Rohrer, P. L., Vilà-Giménez, I., Florit-Pons, J., Gurrado, G., Gibert, N. E., Ren, P., Shattuck-Hufnagel, S., & Prieto, P.** (2021). *The MultiModal MultiDimensional (M3D) labeling system*. https://doi.org/10.17605/OSF.IO/ANKDX

**Rusiewicz, H. L.** (2011). Synchronization of prosodic stress and gesture: A dynamic systems perspective. In *Proceedings of the 2nd Gesture and Speech in Interaction (GESPIN) Conference* (pp. 109–114). Universität Bielefeld.

**Shattuck-Hufnagel, S., & Prieto, P.** (2019). Dimensionalizing co-speech gestures. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne* (pp. 1490–1494). Australasian Speech Science and Technology Association.

**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*(1514), 1–13. https://doi.org/10.3389/fpsyg.2018.01514

**Shattuck-Hufnagel, S., Ren, A., Mathew, M., Yuen, I., & Demuth, K.** (2016). Non-referential gestures in adult and child speech: Are they prosodic? In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), *Speech prosody 2016* (pp. 836–839). International Speech Communication Association. https://doi.org/10.21437/SpeechProsody.2016-171

**Shattuck-Hufnagel, S., Yasinnik, Y., Veilleux, N., & Renwick, M.** (2007). A method for studying the time alignment of gestures and prosody in American English: "Hits" and pitch accents in academic-lecture-style speech. In A. Esposito, M. Bratanic, E. Keller, & M. Marinaro (Eds.), *Fundamentals of verbal and nonverbal communication and the biometric issue* (pp. 34–44). NATO.

**Sloetjes, H.** (2017). *ELAN* (Version 5.8.0) [Computer program]. https://archive.mpi.nl/tla/elan

**Vilà-Giménez, I., Demir-Lira, Ö. E., & Prieto, P.** (2020). The role of referential iconic and non-referential beat gestures in

children's narrative production: Iconics signal oncoming changes in speech. In *Proceedings of the 7th Gesture and Speech in Interaction (GESPIN) Conference*. KTH Speech, Music & Hearing and Språkbanken Tal, Stockholm.

**Vilà-Giménez, I., Dowling, N., Demir-Lira, Ö. E., Prieto, P., & Goldin-Meadow, S.** (2021). The predictive value of non-referential beat gestures: Early use in parent–child interactions predicts narrative abilities at 5 years of age. *Child Development, 92*(6), 2335–2355. https://doi.org/10.1111/cdev.13583

**Vilà-Giménez, I., Florit-Pons, J., Rohrer, P. L., Muñoz-Coego, S., & Prieto, P.** (2021). *Audiovisual corpus of Catalan children's narrative discourse development*. https://doi.org/10.17605/osf.io/npz3w

**Vilà-Giménez, I., & Prieto, P.** (2021). The value of non-referential gestures: A systematic review of their cognitive and linguistic effects in children's language development. *Children, 8*(2), 148. https://doi.org/10.3390/children8020148

**Yasinnik, Y., Renwick, M., & Shattuck-Hufnagel, S.** (2004). The timing of speech-accompanying gestures with respect to prosody. In *Proceedings of the International Conference: From Sound to Sense: 50+ Years of Discoveries in Speech Communication* (pp. C97–C102). MIT.