frontiers in Psychology

REVIEW
published: 19 March 2018
doi: 10.3389/fpsyg.2018.00338

Prosody in the Auditory and Visual Domains: A Developmental Perspective

Núria Esteve-Gibert[1] and Bahia Guellaï[2]**

[1] Departament de Llengües i Literatures Modernes i d'Estudis Anglesos, Universitat de Barcelona (UB), Barcelona, Spain,
[2] Laboratoire Ethologie, Cognition, Développement, Université Paris Nanterre, Nanterre, France

The development of body movements such as hand or head gestures, or facial expressions, seems to go hand-in-hand with the development of speech abilities. We know that very young infants rely on the movements of their caregivers' mouth to segment the speech stream, that infants' canonical babbling is temporally related to rhythmic hand movements, that narrative abilities emerge at a similar time in speech and gestures, and that children make use of both modalities to access complex pragmatic intentions. Prosody has emerged as a key linguistic component in this speech-gesture relationship, yet its exact role in the development of multimodal communication is still not well understood. For example, it is not clear what the relative weights of speech prosody and body gestures are in language acquisition, or whether both modalities develop at the same time or whether one modality needs to be in place for the other to emerge. The present paper reviews existing literature on the interactions between speech prosody and body movements from a developmental perspective in order to shed some light on these issues.

Keywords: speech, gestures, prosody, development, multimodality

OPEN ACCESS

Edited by:
Marianne Gullberg,
Lund University, Sweden

Reviewed by:
Mili Mathew,
St. Cloud State University, United States
Francesca Marina Bosco,
Università degli Studi di Torino, Italy

****Correspondence:***
Núria Esteve-Gibert
nuria0esteve0gibert@gmail.com
Bahia Guellaï
bahia.guellai@gmail.com

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *15 December 2017*
Accepted: *27 February 2018*
Published: *19 March 2018*

Citation:
Esteve-Gibert N and Guellaï B (2018) Prosody in the Auditory and Visual Domains: A Developmental Perspective. Front. Psychol. 9:338. doi: 10.3389/fpsyg.2018.00338

INTRODUCTION

Human language is an interesting input as it can be perceived through both ears and eyes. For example, adults' comprehension of speech in noisy and quiet environments is enhanced when they have access to the visual cues conveyed by the speaker's face (Sumby and Pollack, 1954). In face-to-face interactions, the whole body is involved and may serve informative purposes (Kelly and Barr, 1999 for a review; Kendon, 2004). People around the world produce spontaneous gestures while talking. These gestures accompanying speech, called 'co-speech gestures,' are so connected with speech that people use their hands even when nobody sees them (Corballis, 2002), and congenitally blind people gesture when interacting with each other (Iverson and Goldin-Meadow, 1998 and Goldin-Meadow, 1998). Gestures can be defined on the basis of the articulator that is being used to produce them (the head, as in head nods or head tilts; the hand, as in manual pointing, manual beats or iconic gestures; the face, as in oral gestures or in facial expressions such as eyebrow movements), on the basis of whether or not they are accompanied by speech (co-speech gestures), or based on whether the gesture movement is continuous or discrete (see Wagner et al., 2014 for a review). Another order of things is the function for which they are used in language and communication. Gestures can serve a deictic or highlighting function, they can depict and represent semantic meanings, and they can structure information in the discourse and be an indicator of pragmatic

implicatures to be driven for a successful communication to take place. Because all these levels have parallels with the prosodic properties of speech, these gestures are also called visual correlates of prosody.

It is now clear that co-speech gestures fulfill multiple cognitive functions. Some studies focused on speaker-directed functions suggesting that gestures may ease the speaker's cognitive load (Cook and Goldin-Meadow, 2006; Chu and Kita, 2011), promote learning (Ping and Goldin-Meadow, 2010), help in the conceptual planning of information and discourse (Alibali et al., 2000; Cutica and Bucciarelli, 2008), and facilitate lexical access (Rauscher et al., 1996; Alibali et al., 2000). Others stress that gestures enhance the transfer of information by providing it cross-modally, thereby facilitating uptake for addressees (De Ruiter et al., 2012; Guellaï et al., 2014). These proposals account for the adults' use of co-speech gestures and focus on gestures with a referential value in communication (deictic and iconic hand movements). Yet, they are less effective for explaining developmental patterns as well as the role of gestures with a non-referential value in communication (such as facial expressions and rhythmic 'beats').

In the following sections we propose to explore the developmental links between speech and body movements (i.e., hand and head gestures, and facial expressions), focusing on one specific linguistic aspect, namely prosody. Prosodic properties of speech encode prominence, phrasal organization, speech act types, emotions, attitudes, and beliefs (e.g., Pierrehumbert and Hirschberg, 1990; Ladd, 1996; Byrd and Saltzman, 2003; Jun, 2005). There is a growing body of research showing that prosody is not only expressed through the tonal and temporal properties of speech, but also by means of body movements produced with the hand, head, or face (e.g., Krahmer and Swerts, 2007; Cvejic et al., 2012; Guellaï et al., 2014). The speech and gesture dimensions of prosody are found to be tightly intertwined at the temporal, semantic, and pragmatic levels, and this is true not only in adult speech but also in language development.

Speakers' body movements are temporally coordinated with the prosodic structure in speech, pitch accents and boundary tones serving as anchoring points for prominent phases in body movements (Hadar et al., 1983; De Ruiter, 1998; Leonard and Cummins, 2011; Esteve-Gibert and Prieto, 2013; Ishi et al., 2014; Ambrazaitis and House, 2017; Esteve-Gibert et al., 2017a). At the semantic and pragmatic levels, prosody and gestures can both have a deictic component through which speakers highlight certain elements in speech (Levelt et al., 1985; Roustan and Dohen, 2010), they can disambiguate syntactic constituents (Guellaï et al., 2014; Krivokapic et al., 2016), and mutually influence the processing of speaker's emotions, beliefs, and attitudes (Ekman, 1979; Kendon, 2004; Poggi et al., 2013). In the multimodal expression of prosody, the gesture dimension can consist of movements of the hand or head, facial expressions, or body postures. Traditionally, different types of body movements have been studied independently (for instance, facial expressions have received more attention in the literature on emotions, while hand movements have been the focus of studies on the referential value of gestures in language). In the present paper we will refer to these different types of movements as 'gestures,' as we propose that it is more interesting to take them as a whole to have a complete picture of the speech-gesture relationship in language and communication development.

TEMPORAL ASPECTS OF THE AUDIO–VISUAL SPEECH INTEGRATION IN INFANCY

Infants need to make sense of the rich multisensory stimulations present in their everyday experiences. From the earliest stages of development, infants are found to relate phonetic information from the lips and the voice (Kuhl and Meltzoff, 1984; Aldridge et al., 1999; Patterson and Werker, 2003). In these studies, infants were presented with videos, side-by-side, of two faces articulating two vowels (i.e., /i/ vs. /a/), while hearing only one vowel (i.e., either /i/ or /a/). Infants are considered to be able to detect audio–visual congruency if they look longer at the matching stimulus. Remarkably, there is evidence that from birth, infants detect equivalent phonetic information in the lips and voice (Aldridge et al., 1999). Auditory-visual phonetic matching is also shown at 2 months (Patterson and Werker, 2003), at 4 months and a half (Patterson and Werker, 1999), and at 8 months based on the gender of the talker (Patterson and Werker, 2002). When the vowels are reduced to sine-wave analogs or simple tones, infants do not detect the congruent video anymore (Kuhl et al., 1991). Taken together, these studies, focusing on perioral and facial cues, suggest that infants already have the primitives of lip reading for single speech sounds.

On the production side, newborns bring their hands and objects to their mouth, and explore them orally, these behaviors being considered to be the earliest signs of the oral-manual link in language development (Iverson and Thelen, 1999). Around 6–7 months of age infants start to babble, a rhythmic close–open movement of the jaw that results in the production of syllables (Oller, 2000; Vihman et al., 2009). At the same age infants start producing rhythmic arm movements that are temporally aligned with the vocal babbling (Ejiri, 1998; Iverson and Fagan, 2004). Interestingly, the acoustic quality of the infants' babbles improves when infants combine these vocalizations with rhythmic arm movements, as syllables become shorter and display shorter formant-frequency transitions (Ejiri and Masataka, 2001).

The time-aligned coordination of gesture and speech is also present at later stages of language development. At the onset of word production infants start combining vocalizations with pointing gestures signaling referents in space, and these gestural and speech dimensions are timely aligned in an adult-like way: the accented syllable in speech coincides with the apex of the pointing gesture (Butcher and Goldin-Meadow, 2000; Esteve-Gibert and Prieto, 2014). Later, at 4–5 years of age we observe the emergence of bi-phasic body movements that have no referential meaning and that are timed with pitch accents that children use to emphasize specific information in the sentence (Nicoladis et al., 1999; Capone and McGregor, 2004; Esteve-Gibert et al., 2017b; Mathew et al., 2017). These movements are typically produced with the hand, arm, or head, and are called beats in the gesture literature (Kendon, 2004; McNeill, 2005; Wagner et al., 2014).

Beats provide clear evidence of the rhythmic entrainment between the acoustic and visual dimensions of language, because speakers are found to necessarily modify the acoustic properties of speech when they produce these body movements (Krahmer and Swerts, 2007). Thus, prosodic structure seems to be observed at the speech and at the gestural levels, both dimensions being temporally aligned in a precise way from early stages of language development.

IMPLICATIONS OF THE AUDIO–VISUAL INTEGRATION FOR WORD LEARNING

When addressing infants, adults usually use a speech register which is commonly called Infant-Directed Speech (IDS). This speech register has been the focus of numerous studies as it presents particularities in the auditory domain. It is characterized by slower speech rate and exaggerated pitch excursions compared to Adult-Directed Speech (ADS) (e.g., Fernald and Simon, 1984; Grieser and Kuhl, 1988; Fisher and Tokura, 1995). Vowel and consonant contrasts are more clearly produced in IDS, and this acoustic difference helps infants to build their phoneme inventories (Kuhl et al., 1991; Werker et al., 2007; Cristia, 2011). Also, the slower speaking rate and vowel properties help 21-month-olds learn and remember new words better (Song et al., 2010; Ma et al., 2011).

It has also been observed that IDS is associated with exaggerated facial cues: when addressing infants, caregivers usually exaggerate facial expressions and articulatory lip gestures for corner vowels (Chong et al., 2003; Green et al., 2010). It has been argued that visual IDS attracts infants' attention to the speaker and helps them to parse the speech stream (Kitamura and Burnham, 2003). Some authors have examined sensitivity to the temporal synchrony of visual prosody using continuous IDS (Blossom and Morgan, 2006). They found that infants aged 10–11 months use visual prosody to extract information about the structure of language as they matched synchronous faces and voices. More recently, it has been shown that 8-month-old infants reliably detect congruence between matching auditory and visual displays of a talking face based on prosodic motion (Kitamura et al., 2014), and that 9-month-olds can detect whether a manual deictic gesture is congruently aligned with the corresponding speech segment (Esteve-Gibert et al., 2015). Using an intermodal matching paradigm, Kitamura et al. (2014) presented 8-months-old infants with two visual displays of talking faces (i.e., only moving dots) and one utterance that matched one of the two facial configurations. Results showed that infants reliably detect auditory and visual congruencies in the displays. It seems that this ability emerges early in development as newborns are already able to match a facial display to the corresponding speech stream (Guellaï et al., 2016).

Another dimension of IDS is found in the body gestures of caregivers, which trigger and enhance speech processing. Indeed, caregivers accompany speech with deictic and iconic gestures when talking about objects and actions to infants (Clark and Estigarribia, 2011; Esteve-Gibert et al., 2016), and highlight referential communication by labeling objects while moving them in synchrony with speech (Gogate et al., 2000; Jesse and Johnson, 2016). The caregivers' use of co-speech gestures seems to boost infants' receptive vocabulary and memory skills (Goodwyn et al., 2000; O'Neill et al., 2005; Zammit and Schafer, 2011; Igualada et al., 2017). Igualada et al. (2017) tested preschoolers in a word learning task in which certain words in the list were accompanied by a beat gesture, and results indicated that words co-occurring with gestures were better remembered than gesturally unmarked words.

Yet the impact of Infant-Directed Gestures (or 'gesturese') on language development is an unresolved issue. Some studies have found that toddlers learn words better if adults accompany object labels with deictic and symbolic gestures, and direct their gaze toward the object (Booth et al., 2008; McGregor et al., 2009). However, other findings do not support this hypothesis, some results showing an absence or very small effect of parental use of deictic and symbolic gestures on infants' word learning abilities (Zammit and Schafer, 2011; Puccini and Liszkowski, 2012).

MULTIMODAL DEVELOPMENT OF DISCOURSE AND NARRATIVE SKILLS

An interesting aspect of prosody is that it can also convey information about syntax (Nespor and Vogel, 1986, 2007; Langus et al., 2012). For example, one can manipulate prosodic cues to influence how listeners interpret syntactically ambiguous sentences (Lehiste, 1973; Cooper and Paccia-Cooper, 1980; Price et al., 1991; Carlson et al., 2001). These effects emerge very quickly during sentence comprehension (Marslen-Wilson et al., 1992; Warren et al., 1995; Nagel et al., 1996; Kjelgaard and Speer, 1999; Weber et al., 2006). In the visual domain, the so-called beat gestures seem to be also used to process the structure of the speech signal. In languages such as Italian, English, Dutch, or Catalan, beat gestures are temporally aligned with pitch accents and boundary tones (Yasinnik et al., 2004; Krahmer and Swerts, 2007; Esteve-Gibert et al., 2017a; Krivokapic et al., 2017). Guellaï et al. (2014) showed that spontaneous gestures accompanying speech can be perceived as prosodic markers by adults. This evidence goes in the same direction as a model based on Israeli Signed Language (ISL) showing that body positions align with rhythmic manual features of the signing stream to mark prosodic boundaries (Nespor and Sandler, 1999; Sandler, 1999, 2005, 2011, 2012).

Speakers use prosodic means to emphasize new and important information in ongoing discourse, and for signaling the conceptual structure of the utterances in narrations (Swerts and Geluykens, 1994; Gussenhoven, 2004; Baumann and Grice, 2006; Ladd, 2008). Likewise, visual strategies are found to serve similar functions. Articulatory and head gestures enhance the perception of contrastive focus (Dohen and Loevenbruck, 2009; Swerts and Krahmer, 2010; Kim et al., 2014; Prieto et al., 2015), and body gestures such as eyebrow and head movements are produced less often as a marker of the theme than as a rheme marker (Ambrazaitis and House, 2017).

Children develop discourse and narrative skills relatively late. At around 5 years of age, children use adult-like discourse

markers, dependent clauses and sentential focus to narrate actions with a coherent structure, and these abilities continue to develop over the next years (Hudson and Shapiro, 1991; Berman and Slobin, 1994; Diessel and Tomasello, 2005; Kallay and Redford, 2016). The question is whether gesture and prosodic markers emerge together with the development of syntactic and lexical markers of conceptual structure. On the gesture side, at ages four to five children use beat gestures to emphasize specific information in the sentence (Nicoladis et al., 1999; Capone and McGregor, 2004; Esteve-Gibert et al., 2017b; Mathew et al., 2017). In narrations, children seem to gesture more when they produce longer sentences with more connectives (Nicoladis et al., 1999; Graziano, 2011, 2014; Colletta et al., 2014), and they use different gesture types depending on the age and the type of discourse they produce (Alamillo et al., 2013). Also, they display better narrative skills in a story retelling game if they have had access to manual beat gestures marking information focus and event boundaries (Vilà-Giménez et al., 2017). On the speech prosody side, children at age five and six are found to use the appropriate pitch accents with the right alignment to signal new information in the discourse (see Chen, 2018 for a review), and in narratives they mark event boundaries through pitch direction and linearity (Kallay and Redford, 2016). While results from the gesture literature seem to suggest that gesture marking of discourse structure is directly correlated with the development of linguistic skills, results are less conclusive from the speech prosody side. Kallay and Redford (2016) propose that the correlation between the development of linguistic skills and the development of discourse structure might occur at the level of local pitch features, while more global aspects of discourse prosody such as slope steepness, pitch resets, or pause duration might be mediated by non-linguistic factors such as breathing.

MULTIMODAL CUES IN DEVELOPING EMOTION PERCEPTION AND PRODUCTION

Perceptual skills related to emotion develop very early in infancy. It has been found that 5-month-old infants are able to distinguish between two different emotions on the basis of the speaker's facial expressions and the acoustic properties of speech (Fernald, 1993; Grossmann et al., 2006; Vaillant-Molina et al., 2013). Evidence using continuous speech typically shows that young infants rely on the congruence between auditory emotions (happy, angry) and the appropriate facial expressions (Soken and Pick, 1992; Walker-Andrews, 1997). Production-wise, young infants at 4–5 months of age express emotions such as sadness or enjoyment through facial expressions, and at 12 months of age their facial expressions can signal fear, pain, surprise, or interest (Sullivan and Lewis, 2003). At similar ages, vocal cues are also found to reflect their emotional states (Scheiner et al., 2002; Oller et al., 2013; Lindová et al., 2015).

It is not until much later, however, that children use this early sensitivity to visual and acoustic features of emotion to understand their interlocutor's affective state (Nelson and Russell, 2011; Quam and Swingley, 2012; Berman et al., 2016). Berman et al. (2016) designed a task in which 3- and 5-year-old children had to match pictures of happy-looking and sad-looking faces to happy-sounding and sad-sounding speech, while explicit (pointing) and implicit (eye gaze) responses were measured. Results indicated that only 5 years old children were able to explicitly match the appropriate acoustic and visual cues of emotion, and that at 3 years of age they could only do it implicitly for the negative valence pair.

Even more challenging for children are stimuli in which the speaker intentionally mismatches the audiovisual cues of emotion from the contextual and lexical information, with the purpose of being ironic. In such cases, children at 5–6 years of age tend to interpret the utterance literally even if prosodic cues of emotion signal the speaker's irony (Nakassis and Snedeker, 2002; Laval and Bert-Erboul, 2005; Aguert et al., 2013; Bosco et al., 2013), and only if the utterance is produced together with visual cues of emotion can children infer non-literal meaning (Gil et al., 2014; González-Fuente, 2017). Taken together, all these findings indicate that vocal and visual cues of emotion are recognized and used very early in infancy, and that children use these early skills to process other people's emotions once more complex cognitive abilities are in place.

ACOUSTIC AND VISUAL MARKERS OF INTENTIONS, ATTITUDES, AND BELIEFS

Infants recognize and express their social intentions and communicative goals very early in development, and they use prosodic and gestural means to do so. Twelve-month-old infants rely on pitch, duration, and the shape of the gesture (open-palm pointing, index-finger pointing, etc.) to understand whether the interlocutor is communicating in order to request an object, to inform the caregiver about its presence, or to share interest about it (Behne et al., 2012; Sakkalou and Gattis, 2012; Esteve-Gibert et al., 2017c; Rohlfing et al., 2017). For example, 12-month-old infants use the shape of a pointing gesture and the information from the context to understand that their interlocutor is referring to a certain object in space with a specific social intention (Behne et al., 2012). Interestingly, when contextual cues are ambiguous or uninformative, 12-month-old infants use the shape of the pointing gesture in combination with the prosodic features of speech to infer the speakers' pragmatic intentions (Esteve-Gibert et al., 2017c). Some months later, at around 15 months of age, infants distinguish an action as being accidental or intentional only through the prosodic features of the interlocutor's speech (Sakkalou and Gattis, 2012).

At these pre-lexical stages of language development, prosody and gesture also enable infants to express their intentions toward their interlocutor. We know that 12-month-old infants produce pointing gestures toward referents in space with the purpose of requesting or declaring information, interest, attitudes, or actions (Tomasello et al., 2007; Kovács et al., 2014). It seems that not only pointing gestures but also the prosodic cues of the vocalizations accompanying them indicate the infants' intention (Grünloh and Liszkowski, 2015; Aureli et al., 2017). Aureli et al.

(2017), for instance, found that when Italian-learning 12- to 18-month-olds intend to produce points with a declarative function, the intonation of the vocalization accompanying these points is mostly falling, while it rises to accompany points aimed at asking objects from the interlocutor (thus paralleling what happens in adult speech).

The speaker's beliefs and attitudes about the content of the message are also signaled through vocal and visual strategies. Prosodic cues such as speech rate, pitch level and direction, or voice quality, and gestures such as eyebrow furrowing, head tilt, or shoulder shrugging, are reliably markers of the speaker being uncertain, incredulous, or polite (Krahmer and Swerts, 2005; Dijkstra et al., 2006; Crespo Sendra et al., 2013). Children need complex cognitive mental abilities (the so-called 'Theory of Mind') to understand and express these meanings in language (Wellman, 1990; Perner, 1991; Gopnik, 1993). A large body of research has dealt with the question of when these abilities emerge. Some researchers propose that only at ages four to five do children have fully developed mind-reading abilities, since it is at this age that they succeed in false-belief tasks (Wimmer and Perner, 1983; Baron-Cohen et al., 1985). Yet others claim that younger infants show early cognitive abilities of this kind when less cognitively demanding tasks are used (Onishi and Baillargeon, 2005; Baillargeon et al., 2010; Kovács et al., 2010). Studies exploring the development of prosodic and gesture cues to interpret the other's beliefs and attitudes suggest that children's belief comprehension increases significantly during the preschool years. For example, at 3–5 years of age children detect at above chance level the speaker's beliefs about what she/he is saying thanks to the speaker's facial expressions and, interestingly, those that are more accurate are those with more sophisticated belief-reasoning skills (Armstrong et al., 2014). Visual information is found to be a stronger cue for preschoolers than prosodic cues of uncertainty, even if prosody is a stronger indicator still than lexical information (Moore et al., 1993; Hübscher et al., 2017). On the production side, children first use prosody than lexical cues to mark uncertainty in speech (Hübscher et al., 2016), and at 7–8 year of age they signal uncertainty through facial expressions such as eyebrow raising or furrowing or funny faces, and with prosodic cues such as fillers, delays, and high intonation (Krahmer and Swerts, 2005; Visser et al., 2014). All together, these studies suggest that children use the acoustic and visual components of prosody before lexical markers to understand and produce beliefs and attitudes in language. Yet, more studies are required to disentangle which of these prosodic dimensions (visual or acoustic) comes first, and whether this developmental path depends on the child's cognitive abilities and/or on the specific linguistic meaning that is investigated.

DISCUSSION

The present review is aimed at highlighting recent discoveries on the developmental integration of speech in the auditory and visual domains, focusing on the prosodic level. Although there are more and more evidence of links between speech and gestures, we do not fully understand the relative weight of each modality in language comprehension, and we need to clarify whether prosody has parallel forms and functions in the acoustic and visual domains. Adopting a developmental approach could help in answering these questions.

Developmental research can help disentangle whether gestures are part of the speakers' linguistic system. There is consistent evidence that infants and children first use the gesture modality to refer to objects in space before they use words and word-gesture combinations to do so (Bates et al., 1979; Butcher and Goldin-Meadow, 2000; Esteve-Gibert and Prieto, 2014). In fact, the rate of gesturally pointed referents is a reliable sign of the infants' vocabulary skills at later stages (Rowe and Goldin-Meadow, 2009; Igualada et al., 2015), and the rate of pointing-speech combinations at 18 months of age (when pointing and speech provide complementary meanings) is a reliable predictor of sentence complexity at 42 months of age (Rowe and Goldin-Meadow, 2009). Mathew et al. (2017) observed that 6-year-olds produce 'beat' gestures with an emphasizing function, but surprisingly the gesture-accompanying words did not always bear a pitch accent, suggesting that children are still learning to use the speech modality to emphasize discourse elements, while they seem to already master the gesture. Although not all language functions emerge first in the visual modality (note, for instance, that toddlers first express actions with verbs and only later are able to represent that same action with iconic gestures depicting that action; Özçaliskan et al., 2003), the abovementioned results indicate that infants and children do use gestures for linguistic purposes, and that speech and gestures might be part of the same linguistic and communicative system (Kendon, 1980; McNeill, 1992; Goldin-Meadow, 1998).

It is still an open question the reason why certain linguistic functions are first expressed through gestures and some others are first observed in the acoustic dimension. Parladé and Iverson (2011) propose a dynamic systems approach to cope with the fact that infants prefer to use one modality over the other for a given linguistic function at certain stages in language development. According to these authors, in periods where infants increase their skills in one communicative behavior, there might be a temporary regression in an alternative communicative behavior. For instance, the authors find that when infants' vocabulary increases, their production of multimodal communicative behaviors (i.e., combination of vocal, gestural, and affect behaviors) is reduced. Later, once vocabulary skills are stabilized, the rate of multimodal communicative behaviors increases again. It remains unclear, however, why certain linguistic functions emerge first through gesture rather than through speech, and vice-versa, as well as what motor, cognitive, or communicational factors might influence this behavior.

Studies in brain imagery could also help tease apart the possibility of a gesture/speech linkage in language. Indeed, in adult populations, it has been shown that listening to speech evokes neural responses in the motor cortex. This has been controversially interpreted as evidence that speech sounds are processed as articulatory movements (Pulvermüller and Fadiga, 2010). Recently, Biau et al. (2016) evaluated beat synchrony against arbitrary visual cues bearing equivalent

rhythmic and spatial properties as the gestures. Their results revealed that left Middle Temporal Gyrus and Inferior Frontal Gyrus were specifically sensitive to speech synchronized with beats, compared to the arbitrary vision–speech pairing. Hence, it seems that co-speech gestures and speech perception are instantiated through a specialized brain network sensitive to the communicative intent conveyed by the speaker's whole body.

There are very few studies investigating the developmental signs of the vocal-motor linkages at the neural level, and most evidence comes from populations with developmental disorders and brain injuries. For instance, children with perinatal brain lesions are found to have both lower rates of gesture production and smaller vocabularies (Sauer et al., 2010). Another way to specify the links between gestures and speech would be to explore how sensorimotor feedback influences auditory-visual speech processing, for instance by investigating whether the production of gestures influences infants' speech fluency. If more evidence is obtained showing that gesture and speech mutually influence each other in language production, perception, and comprehension, this would suggest that they are part of the linguistic system and not only communicative means, especially in development.

Among the linguistic aspects revealing the gesture/speech link more clearly, we have shown that prosody has a prominent status. Prosodic targets are anchoring points for manual gestures and facial expressions to align, pitch accents attracting prominent gestural phases and prosodic phrase boundaries framing the scope of gesture movements. This is true in adults (Hadar et al., 1983; De Ruiter, 1998; Leonard and Cummins, 2011; Esteve-Gibert and Prieto, 2013; Ferré, 2014; Ishi et al., 2014; Ambrazaitis and House, 2017; Esteve-Gibert et al., 2017a), and it also seems to hold for infants and children (Butcher and Goldin-Meadow, 2000; Esteve-Gibert and Prieto, 2014; Mathew et al., 2017). While more research is needed to examine the patterns of this temporal linkage in infants' productions (especially in stages when these prosodic targets become adult-like), perception studies show that infants are sensitive to the alignment of prosodic and visual cues as early as 8–9 months of age (Kitamura et al., 2014; Esteve-Gibert et al., 2015). It has been proposed that the driving force of this temporal linkage is a bi-directional influence between gesture and speech 'pulses' (i.e., peaks in an ongoing rhythm) (McNeill, 1992; Tuite, 1993; Iverson and Thelen, 1999; Port, 2003; Rusiewicz and Esteve-Gibert, 2018).

Prosody and gestures also overlap in terms of which linguistic functions they are used for. Infants use visual correlates of prosody to segment the speech stream (e.g., Kitamura et al., 2014; Guellaï et al., 2016), to organize information at the discourse level (e.g., Nicoladis et al., 1999; Capone and McGregor, 2004; Mathew et al., 2017), and to express emotions, intentions, and beliefs (Sullivan and Lewis, 2003; Esteve-Gibert and Prieto, 2014; Berman et al., 2016; Aureli et al., 2017; González-Fuente, 2017). Children are sensitive to the fact that visual cues convey relevant linguistic meaning, and experimental evidence shows that gestures are processed earlier and more accurately than prosodic or lexical cues (Armstrong et al., 2014; Esteve-Gibert et al., 2017c; Hübscher et al., 2017). If future studies confirm that infants and children first process through visual cues what they later learn to process acoustically, this would mean that gestures are key in the development of linguistic categories, and that they not only precede but also scaffold language development (see a proposal on this regard in Hübscher et al., 2017). Furthermore, by examining in more detail how visual and acoustic cues of prosody emerge, evolve, and interact across development, we will be able to develop models that can predict and guide intervention in the case of atypical language development. The studies reviewed here have shown that gestures are tightly linked to prosody at the formal and functional levels and across different stages of language development. Still, further studies are needed to fully clarify the origin of these links and their implications for language acquisition.

AUTHOR CONTRIBUTIONS

All authors have equally participated to the discussion and writing of the manuscript. BG has had a leading role in section 1 (Introduction), section 3 (Word Learning), and section 7 (Discussion), while NE-G has had a leading role in section 2 (Temporal Aspects), section 4 (Narrative Skills), section 5 (Emotions), and section 6 (Intentions, Attitudes, and Beliefs).

FUNDING

This research was funded by the FJCI-2015-26845 postdoctoral grant (Spanish Ministry of Economy, Industry, and Competitiveness) to NE-G, and by the Fyssen Foundation for BG.

ACKNOWLEDGMENTS

We thank Pilar Prieto, Maya Gratier, and Alan Langus for their insights and discussion of the research presented in this article.

REFERENCES

Aguert, M., Laval, V., Lacroix, A., Gil, S., and Le Bigot, L. (2013). Inferring emotions from speech prosody: not so easy at age five. *PLoS One* 8:e83657. doi: 10.1371/journal.pone.0083657

Alamillo, A. R., Colletta, J. M., and Guidetti, M. (2013). Gesture and language in narratives and explanations: the effects of age and communicative activity on late multimodal discourse development. *J. Child Lang.* 40, 511–538. doi: 10.1017/S0305000912000062

Aldridge, M. A., Braga, E. S., Walton, G. E., and Bower, T. G. R. (1999). The intermodal representation of speech in newborns. *Dev. Sci.* 2, 42–46. doi: 10.1111/1467-7687.00052

Alibali, M. W., Kita, S., and Young, A. J. (2000). Gesture and the process of speech production: we think, therefore we gesture. *Lang. Cogn. Process.* 15, 593–613. doi: 10.1080/016909600750040571

Ambrazaitis, G., and House, D. (2017). Multimodal prominences: exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings. *Speech Commun.* 95, 100–113. doi: 10.1016/j.specom.2017.08.008

Armstrong, M., Esteve-Gibert, N., and Prieto, P. (2014). "The acquisition of multimodal cues to disbelief," in *Proceedings of the 7th International Conference on Speech Prosody*, Dublin.

Aureli, T., Spinelli, M., Fasolo, M., Garito, M. C., Perucchini, P., and D'Odorico, L. (2017). The pointing-vocal coupling progression in the first half of the second year of life. *Infancy* 22, 801–818. doi: 10.1111/infa.12181

Baillargeon, R., Scott, R. M., and He, Z. (2010). False-belief understanding in infants. *Trends Cogn. Sci.* 14, 110–118. doi: 10.1016/j.tics.2009.12.006

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition* 21, 37–46.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., and Volterra, V. (1979). *The Emergence of Symbols: Cognition and Communication in Infancy*. New York, NY: Academic Press.

Baumann, S., and Grice, M. (2006). The intonation of accessibility. *J. Pragmat.* 38, 1636–1657. doi: 10.1016/j.pragma.2005.03.017

Behne, T., Liszkowski, U., Carpenter, M., and Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *Br. J. Dev. Psychol.* 30, 359–375. doi: 10.1111/j.2044-835X.2011.02043.x

Berman, J. M. J., Chambers, C. G., and Graham, S. A. (2016). Preschoolers' real-time coordination of vocal and facial emotional information. *J. Exp. Child Psychol.* 142, 391–399. doi: 10.1016/j.jecp.2015.09.014

Berman, R. A., and Slobin, D. I. (1994). *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillsdale, NJ: Erlbaum.

Biau, E., Fernández, L. M., Holle, H., Avila, C., and Soto-Faraco, S. (2016). Hand gestures as visual prosody: BOLD responses to audio–visual alignment are modulated by the communicative nature of the stimuli. *Neuroimage* 132, 129–137. doi: 10.1016/j.neuroimage.2016.02.018

Blossom, M., and Morgan, J. L. (2006). "Does the face say what the mouth says? A study of infants' sensitivity to visual prosody," in *Proceedings of the 30th Annual Boston University Conference on Language Development*, Somerville, MA.

Booth, A. E., McGregor, K. K., and Rohlfing, K. L. (2008). Socio-pragmatics and attention: contributions to gesturally guided word learning in toddlers. *Lang. Learn. Dev.* 4, 179–202. doi: 10.1080/15475440802143091

Bosco, F. M., Angeleri, R., Colle, L., Sacco, K., and Bara, B. G. (2013). Communicative abilities in children: an assessment through different phenomena and expressive means. *J. Child Lang.* 40, 741–778. doi: 10.1017/S0305000913000081

Butcher, C., and Goldin-Meadow, S. (2000). "Gesture and the transition from one- to two-word speech: when hand and mouth come together," in *Language and Gesture*, ed. D. McNeill (Chicago, IL: Cambridge University Press), 235–257.

Byrd, D., and Saltzman, E. (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *J. Phon.* 31, 149–180. doi: 10.1016/S0095-4470(02)00085-2

Capone, N. C., and McGregor, K. K. (2004). Gesture development: a review for clinical and research practices. *J. Speech Lang. Hear. Res.* 47, 173–186. doi: 10.1044/1092-4388(2004/015)

Carlson, K., Clifton, C., and Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *J. Mem. Lang.* 45, 58–81. doi: 10.1006/jmla.2000.2762

Chen, A. (2018). "Get the focus right across languages: acquisition of prosodic focus-marking in production," in *The Development of Prosody in First Language Acquisition*, eds P. Prieto and N. Esteve-Gibert (Amsterdam: John Benjamins).

Chong, S. C. F., Werker, J. F., Russell, J. A., and Carroll, J. M. (2003). Three facial expressions mothers direct to their infants. *Infant Child Dev.* 12, 211–232. doi: 10.1002/icd.286

Chu, M., and Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *J. Exp. Psychol. Gen.* 140, 102–115. doi: 10.1037/a0021790

Clark, E. V., and Estigarribia, B. (2011). Using speech and gesture to introduce new objects to young children. *Gesture* 11, 1–23. doi: 10.1075/gest.11.1.01cla

Colletta, J.-M., Guidetti, M., Capirci, O., Cristilli, C., Demir, O. E., Kunene-Nicolas, R. N., et al. (2014). Effects of age and language on co-speech gesture production: an investigation of French, American, and Italian children's narratives. *J. Child Lang.* 42, 122–145. doi: 10.1017/S0305000913000585

Cook, S. M., and Goldin-Meadow, S. (2006). The role of gesture in learning: Do children use their hands to change their minds? *J. Cogn. Dev.* 7, 211–232. doi: 10.1207/s15327647jcd0702_4

Cooper, W. E., and Paccia-Cooper, J. (1980). *Syntax and Speech*. Cambridge, MA: Harvard University Press. doi: 10.4159/harvard.9780674283947

Corballis, M. C. (2002). *From Hand to Mouth: The Origins of Language.* Princeton, NJ: Princeton University Press.

Crespo Sendra, V., Kaland, C., Swerts, M., and Prieto, P. (2013). Perceiving incredulity: the role of intonation and facial gestures. *J. Pragmat.* 47, 1–13. doi: 10.1016/j.pragma.2012.08.008

Cristia, A. (2011). Fine-grained variation in caregivers'/s/predicts their infants'/s/category a. *J. Acoust. Soc. Am.* 129, 3271–3280. doi: 10.1121/1.3562562

Cutica, I., and Bucciarelli, M. (2008). The deep versus the shallow: effects of co-speech gestures in learning from discourse. *Cogn. Sci.* 32, 921–935. doi: 10.1080/03640210802222039

Cvejic, E., Kim, J., and Davis, C. (2012). Recognizing prosody across modalities, face areas and speakers: examining perceivers' sensitivity to variable realizations of visual prosody. *Cognition* 122, 442–453. doi: 10.1016/j.cognition.2011.11.013

De Ruiter, J. P. (1998). *Gesture and Speech Production.* Doctoral dissertation, Katholieke Universiteit, Nijmegen.

De Ruiter, J. P., Bangerter, A., and Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: investigating the tradeoff hypothesis. *Top. Cogn. Sci.* 4, 232–248. doi: 10.1111/j.1756-8765.2012.01183.x

Diessel, H., and Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language* 81, 1–25. doi: 10.1353/lan.2005.0169

Dijkstra, C., Krahmer, E., and Swerts, M. (2006). "Manipulating uncertainty: the contribution of different audiovisual prosodic cues to the perception of confidence," in *Proceedings of the International Conference on Speech Prosody*, Dresden.

Dohen, M., and Loevenbruck, H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Lang. Speech* 52, 177–206. doi: 10.1177/0023830909103166

Ejiri, K. (1998). Relationship between rhythmic behavior and canonical babbling in infant vocal development. *Phonetica* 55, 226–237. doi: 10.1159/000028434

Ejiri, K., and Masataka, N. (2001). Co-occurrence of preverbal vocal behavior and motor action in early infancy. *Dev. Sci.* 4, 6–11. doi: 10.1111/1467-7687.00147

Ekman, P. (1979). "About brows: emotional and conversational signals," in *Human Ethology: Claims and Limits of a New Discipline*, eds M. von Cranach, K. Foppa, W. Lepenies, and D. Ploog (Cambridge: Cambridge University Press), 169–202.

Esteve-Gibert, N., Borràs-Comes, J., Asor, E., Swerts, M., and Prieto, P. (2017a). The timing of head movements: the role of prosodic heads and edges. *J. Acoust. Soc. Am.* 141, 4727–4739. doi: 10.1121/1.4986649

Esteve-Gibert, N., Loevenbruck, H., Dohen, M., and D'Imperio, M. (2017b). "The use of prosody and gestures for the production of contrastive focus in French-speaking 4- and 5-year-old children," in *Proceedings of the Workshop on Abstraction, Diversity and Speech Dynamics*, Munich.

Esteve-Gibert, N., Liszkowski, U., and Prieto, P. (2016). "Prosodic and gestural features distinguish the intention of pointing gestures in child-directed communication," in *Interdisciplinary Approaches to Intonational Grammar in Ibero-Romance*, eds M. E. Armstrong, N. Henriksen, and M. D. M. Vanrell (Amsterdam: John Benjamins), 251–275.

Esteve-Gibert, N., and Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *J. Speech Lang. Hear. Res.* 56, 850–864. doi: 10.1044/1092-4388(2012/12-0049)

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

Esteve-Gibert, N., Prieto, P., and Liszkowski, U. (2017c). Twelve-month-olds understand social intentions based on prosody and gesture shape. *Infancy* 22, 108–129. doi: 10.1111/infa.12146

Esteve-Gibert, N., Prieto, P., and Pons, F. (2015). Nine-month-old infants are sensitive to the temporal alignment of prosodic and gesture prominences. *Infant Behav. Dev.* 38, 126–129. doi: 10.1016/j.infbeh.2014.12.016

Fernald, A. (1993). Approval and disapproval: infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Dev.* 64, 657–674. doi: 10.2307/1131209

Fernald, A., and Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Dev. Psychol.* 20, 104–113. doi: 10.1037/0012-1649.20.1.104

Ferré, G. (2014). A multimodal approach to markedness in spoken French. *Speech Commun.* 57, 268–282. doi: 10.1016/j.specom.2013.06.002

Fisher, C., and Tokura, H. (1995). The given-new contract in speech to infants. *J. Mem. Lang.* 34, 287–310. doi: 10.1006/jmla.1995.1013

Gil, S., Aguert, M., Le Bigot, L., Lacroix, A., and Laval, V. (2014). Children's understanding of others' emotional states. *Int. J. Behav. Dev.* 38, 539–549. doi: 10.1177/0165025414535123

Gogate, L. J., Bahrick, L. E., and Watson, J. D. (2000). A study of multimodal motherese: the role of temporal synchrony between verbal labels and gestures. *Child Dev.* 71, 878–894. doi: 10.1111/1467-8624.00197

Goldin-Meadow, S. (1998). *The Development of Gesture and Speech as an Integrated System*. San Francisco, CA: Jossey- Bass.

González-Fuente, S. (2017). *Audiovisual Prosody and Verbal Irony*. Ph.D. dissertation, Universitat Pompeu Fabra, Barcelona.

Goodwyn, S. W., Acredolo, L. P., and Brown, C. A. (2000). Impact of symbolic gesturing on early language development. *J. Nonverbal Behav.* 24, 81–103. doi: 10.1023/A:1006653828895

Gopnik, A. (1993). How we know our minds: the illusion of first-person knowledge of intentionality. *Behav. Brain Sci.* 16, 1–15. doi: 10.1017/S0140525X00028636

Graziano, M. (2011). "The development of two pragmatic gestures of the so-called "Open Hand Supine family" in Italian children," in *From Gesture in Conversation to Visible Action as Utterance: Essays in Honor of Adam Kendon*, eds M. Seyfeddinipur and M. Gullberg (Amsterdam: John Benjamins Publishing Company), 311–330.

Graziano, M. (2014). "Gestures in Southern Europe: children's pragmatic gestures in Italy," in *Body-Language-Communication*, eds C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and S. Tessendorf (Berlin: De Gruyter), 1253–1258.

Green, J. R., Nip, I. S., Wilson, E. M., Mefferd, A. S., and Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. *J. Speech Lang. Hear. Res.* 53, 1529–1542. doi: 10.1044/1092-4388(2010/09-0005)

Grieser, D. L., and Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: support for universal prosodic features in motherese. *Dev. Psychol.* 24, 14–20. doi: 10.1037/0012-1649.24.1.14

Grossmann, T., Striano, T., and Friederici, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Dev. Sci.* 9, 309–315. doi: 10.1111/j.1467-7687.2006.00494.x

Grünloh, T., and Liszkowski, U. (2015). Prelinguistic vocalizations distinguish pointing acts. *J. Child Lang.* 42, 1312–1336. doi: 10.1017/S0305000914000816

Guellaï, B., Langus, A., and Nespor, M. (2014). Prosody in the hands of the speaker. *Front. Psychol.* 5:700. doi: 10.3389/fpsyg.2014.00700

Guellaï, B., Streri, A., Chopin, A., Rider, D., and Kitamura, C. (2016). Newborns' sensitivity to the visual aspects of infant-directed speech: evidence from point-line displays of talking faces. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 1275–1281. doi: 10.1037/xhp0000208

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511616983

Hadar, U., Steiner, T. J., Grant, E. C., and Rose, F. C. (1983). Kinematics of head movements accompanying speech during conversation. *Hum. Mov. Sci.* 2, 35–46. doi: 10.1016/0167-9457(83)90004-0

Hübscher, I., Esteve-Gibert, N., Igualada, A., and Prieto, P. (2017). Intonation and gesture as bootstrapping devices in speaker uncertainty. *First Lang.* 37, 24–41.

Hübscher, I., Vinzce, L., and Prieto, P. (2016). "Epistemic meaning is first communicated through gesture, face and prosody," in *Poster at the Workshop on Audiovisual Speech Processing and Language Learning*, Universitat Pompeu Fabra, Barcelona.

Hudson, J., and Shapiro, L. (1991). "From knowing to telling: the development of children's scripts, stories, and personal narratives," in *Developing Narrative Structure*, eds A. McCabe and C. Peterson (Hillsdale, NJ: Lawrence Erlbaum Associates), 89–136.

Igualada, A., Bosch, L., and Prieto, P. (2015). Language development at 18 months is related to communicative strategies at 12 months. *Infant Behav. Dev.* 39, 42–52. doi: 10.1016/j.infbeh.2015.02.004

Igualada, A., Esteve-Gibert, N., and Prieto, P. (2017). Beat gestures improve word recall in 3- to 5-year-old children. *J. Exp. Child Psychol.* 156, 99–112. doi: 10.1016/j.jecp.2016.11.017

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

Iverson, J. M., and Fagan, M. K. (2004). Infant vocal-motor coordination: precursor to the gesture-speech system? *Child Dev.* 75, 1053–1066. doi: 10.1111/j.1467-8624.2004.00725.x

Iverson, J. M., and Goldin-Meadow, S. (1998). Why do people gesture as they speak? *Nature* 396:228. doi: 10.1038/24300

Iverson, J. M., and Thelen, E. (1999). Hand, mouth and brain. *J. Conscious. Stud.* 6, 19–40.

Jesse, A., and Johnson, E. K. (2016). Audiovisual alignment of co-speech gestures to speech supports word learning in 2-year-olds. *J. Exp. Child Psychol.* 145, 1–10. doi: 10.1016/j.jecp.2015.12.002

Jun, S.-A. (2005). "Prosodic typolog," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, ed. S.-A. Jun (Oxford: Oxford University Press), 430–458. doi: 10.1093/acprof:oso/9780199249633.003.0016

Kallay, J., and Redford, M. A. (2016). "A longitudinal study of children's intonation in narrative speech," in *Proceedings of the 17th Annual Conference of the International Speech Communication Association*, San Francisco, CA. doi: 10.21437/Interspeech.2016-1396

Kelly, S., and Barr, D. (1999). Offering a hand to pragmatic understanding: the role of speech and gesture in comprehension and memory. *J. Mem. Lang.* 40, 577–592. doi: 10.1006/jmla.1999.2634

Kendon, A. (1980). "Gesticulation and speech: two aspects of the process of utterance," in *The Relationship of Verbal and Nonverbal Communication*, ed. M. R. Key (The Hague: Mouton De Gruyter), 207–227.

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511807572

Kim, J., Cvejic, E., and Davis, C. (2014). Tracking eyebrows and head gestures associated with spoken prosody. *Speech Commun.* 57, 317–330. doi: 10.1016/j.specom.2013.06.003

Kitamura, C., and Burnham, D. (2003). Pitch and communicative intent in mother's speech: adjustments for age and sex in the first year. *Infancy* 4, 85–110. doi: 10.1207/S15327078IN0401_5

Kitamura, C., Guellaï, B., and Kim, J. (2014). Motherese by eye and ear: infants perceive visual prosody in point-line displays of talking heads. *PLoS One* 9:e111467. doi: 10.1371/journal.pone.0111467

Kjelgaard, M. M., and Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *J. Mem. Lang.* 40, 153–194. doi: 10.1006/jmla.1998.2620

Kovács, A. M., Tauzin, T., Teglas, E., Gergely, G., and Csibra, G. (2014). Pointing as epistemic request: 12-month-olds point to receive new information. *Infancy* 19, 543–557. doi: 10.1111/infa.12060

Kovács, Á. M., Téglás, E., and Endress, A. D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science* 330, 1830–1834. doi: 10.1126/science.1190792

Krahmer, E., and Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Lang. Speech* 48, 29–53. doi: 10.1177/00238309050480010201

Krahmer, E., and Swerts, M. (2007). The effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. *J. Mem. Lang.* 57, 396–414. doi: 10.1016/j.jml.2007.06.005

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

Krivokapic, J., Tiede, M. K., Tyrone, M. E., and Goldenberg, D. (2016). "Speech and manual gesture coordination in a pointing task," in *Proceedings of the 8th International Conference on Speech Prosody*, Boston, 1240–1244. doi: 10.21437/SpeechProsody.2016-255

Kuhl, P. K., and Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behav. Dev.* 7, 361–381. doi: 10.1016/S0163-6383(84)80050-8

Kuhl, P. K., Williams, K. A., and Meltzoff, A. N. (1991). Cross- modal speech perception in adults and infants using non-speech auditory stimuli. *J. Exp. Psychol. Hum. Percept. Perform.* 17, 829–840. doi: 10.1037/0096-1523.17.3.829

Ladd, R. (1996). *Intonational Phonology, Cambridge Studies in Linguistics*, Vol. 79. Cambridge: Cambridge University Press.

Ladd, R. (2008). *Intonational Phonology*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511808814

Langus, A., Marchetto, E., Bion, R. A., and Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *J. Mem. Lang.* 66, 285–306. doi: 10.1016/j.jml.2011.09.004

Laval, V., and Bert-Erboul, A. (2005). French-speaking children's understanding of sarcasm: the role of intonation and context. *J. Speech Lang. Hear. Res.* 48, 610–620. doi: 10.1044/1092-4388(2005/042)

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa* 7, 102–122. doi: 10.1121/1.1982702

Leonard, T., and Cummins, F. (2011). The temporal relation between beat gestures and speech. *Lang. Cogn. Process.* 26, 1457–1471. doi: 10.1080/01690965.2010.500218

Levelt, W. J. M., Richardson, G., and La Heij, W. (1985). Pointing and voicing in deictic expressions. *J. Mem. Lang.* 24, 133–164. doi: 10.1016/0749-596X(85)90021-X

Lindová, J., Špinka, M., and Nováková, L. (2015). Decoding of baby calls: can adult humans identify the eliciting situation from emotional vocalizations of preverbal infants? *PLoS One* 10:e0124317. doi: 10.1371/journal.pone.0124317

Ma, W., Golinkoff, R. M., Houston, D. M., and Hirsh-Pasek, K. (2011). Word learning in infant-and adult-directed speech. *Lang. Learn. Dev.* 7, 185–201. doi: 10.1080/15475441.2011.579839

Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., and Lee, C. S. (1992). Prosodic effects in minimal attachment. *Q. J. Exp. Psychol.* 45, 73–87. doi: 10.1080/14640749208401316

Mathew, M., Yuen, I., and Demuth, K. (2017). Talking to the beat: six-year-olds' use of stroke-defined non-referential gestures. *First Lang.* (in press). doi: 10.1177/0142723717734949

McGregor, K. K., Rohlfing, K., Bean, A., and Marschner, E. (2009). Gesture as a support for word learning: the case of under. *J. Child Lang.* 36, 807–828. doi: 10.1017/S0305000908009173

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago.

McNeill, D. (2005). *Gesture and Thought*. Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226514642.001.0001

Moore, C., Harris, L., and Patriquin, M. (1993). Lexical and prosodic cues in the comprehension of relative certainty. *J. Child Lang.* 20, 153–167. doi: 10.1017/S030500090000917X

Nagel, H. N., Shapiro, L. P., Tuller, B., and Nawy, R. (1996). Prosodic influences on the resolution of temporary ambiguity during on-line sentence processing. *J. Psycholinguist. Res.* 25, 319–344. doi: 10.1007/BF01708576

Nakassis, C., and Snedeker, J. (2002). "Beyond sarcasm: intonation and context as relational cues in children's recognition of irony," in *Proceedings of the Annual Boston-University Conference on Language Development,* Vol. 26, eds A. Greenhill, M. Hughs, H. Littlefield, and H. Walsh (Somerville, MA: Cascadilla Press), 429–440.

Nelson, N. L., and Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *J. Exp. Child Psychol.* 110, 52–61. doi: 10.1016/j.jecp.2011.03.014

Nespor, M., and Sandler, W. (1999). Prosody in Israeli sign language. *Lang. Speech* 42, 143–176. doi: 10.1177/00238309990420020201

Nespor, M., and Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris, 327.

Nespor, M., and Vogel, I. (2007). *Prosodic Phonology*, 1st Edn. Berlin: Mouton De Gruyter. doi: 10.1515/9783110977790

Nicoladis, E., Mayberry, R. I., and Genesee, F. (1999). Gesture and early bilingual development. *Dev. Psychol.* 35, 514–526. doi: 10.1037/0012-1649.35.2.514

Oller, D. K. (2000). *The Emergence of the Speech Capacity*. Mahwah, NJ: Lawrence Erlbaum.

Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., and Chorna, L. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6318–6323. doi: 10.1073/pnas.1300337110

O'Neill, M., Bard, K. A., Linnell, M., and Fluck, M. (2005). Maternal gestures with 20-month-old infants in two contexts. *Dev. Sci.* 8, 352–359. doi: 10.1111/j.1467-7687.2005.00423.x

Onishi, K. H., and Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science* 308, 255–258.

Özçaliskan, S., Gentner, D., and Goldin-meadow, S. (2013). Do iconic gestures pave the way for children's early verbs? *Appl. Psycholinguist.* 35, 1143–1162.

Parladé, M. V., and Iverson, J. M. (2011). The interplay between language, gesture, and affect during communicative transition: a dynamic systems approach. *Dev. Psychol.* 47, 820–833. doi: 10.1037/a0021811

Patterson, M. L., and Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behav. Dev.* 22, 237–247. doi: 10.1016/j.cognition.2008.05.009

Patterson, M. L., and Werker, J. F. (2002). Infants' ability to match dynamic phonetic and gender information in the face and voice. *J. Exp. Child Psychol.* 81, 93–115. doi: 10.1006/jecp.2001.2644

Patterson, M. L., and Werker, J. F. (2003). Two-month-old infants match phonemic information in lips and voice. *Dev. Sci.* 6, 191–196. doi: 10.1016/j.cognition.2008.05.009

Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, MA: MIT Press.

Pierrehumbert, J., and Hirschberg, J. (1990). "The meaning of intonational contours in the interpretation of discourse," in *Intentions in Communication*, eds P. R. Cohen, J. Morgan, and M. E. Pollack (Cambridge, MA: MIT Press), 270–311.

Ping, R., and Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cogn. Sci.* 34, 602–619. doi: 10.1111/j.1551-6709.2010.01102.x

Poggi, I., D'Errico, F., and Vincze, L. (2013). Comments by words, face and body. *J. Multimodal User Interfaces* 7, 67–78. doi: 10.1007/s12193-012-0102-z

Port, R. F. (2003). Meter and speech. *J. Phon.* 31, 599–611. doi: 10.1016/j.wocn.2003.08.001

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. (1991). The use of prosody in syntactic disambiguation. *J. Acoust. Soc. Am.* 90, 2956–2970. doi: 10.1121/1.401770

Prieto, P., Puglesi, C., Borràs-Comes, J., Arroyo, E., and Blat, J. (2015). Exploring the contribution of prosody and gesture to the perception of focus using an animated agent. *J. Phon.* 49, 41–54. doi: 10.1016/j.wocn.2014.10.005

Puccini, D., and Liszkowski, U. (2012). 15-month-old infants fast map words but not representational gestures of multimodal labels. *Front. Psychol.* 3:101. doi: 10.3389/fpsyg.2012.00101

Pulvermüller, F., and Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* 11, 351–360. doi: 10.1038/nrn2811

Quam, C., and Swingley, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Dev.* 83, 236–250. doi: 10.1111/j.1467-8624.2011.01700.x

Rauscher, F. H., Krauss, R. M., and Chen, Y. (1996). Gesture, speech and lexical access: the role of lexical movements in speech production. *Psychol. Sci.* 7, 226–231. doi: 10.1002/wcs.1211

Rohlfing, K. J., Grimminger, A., and Lüke, C. (2017). An interactive view on the development of deictic pointing in infancy. *Front. Psychol.* 8:1319. doi: 10.3389/fpsyg.2017.01319

Roustan, B., and Dohen, M. (2010). "Co-production of contrastive focus and manual gestures: temporal coordination and effects on the acoustic and articulatory correlates of focus," in *Proceedings of the International Conference on Speech Prosody*, Chicago, IL.

Rowe, M. L., and Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Dev. Sci.* 12, 182–187. doi: 10.1111/j.1467-7687.2008.00764.x

Rusiewicz, H. L., and Esteve-Gibert, N. (2018). "Set in time: temporal coordination of prosody and gesture in the development of spoken language production," in *The Development of Prosody in First Language Acquisition*, eds P. Prieto and N. Esteve-Gibert (Amsterdam: John Benjamins).

Sakkalou, E., and Gattis, M. (2012). Infants infer intentions from prosody. *Cogn. Dev.* 27, 1–16. doi: 10.1016/j.cogdev.2011.08.003

Sandler, W. (1999). Prosody in two natural language modalities. *Lang. Speech* 42, 127–142. doi: 10.1177/00238309990420020101

Sandler, W. (2005). Prosodic constituency and intonation in sign language. *Linguist. Ber.* 13, 59–86.

Sandler, W. (2011). "The phonology of movement in sign language," in *Blackwell Companion to Phonology*, eds M. van Oostendorp, C. Ewen, K. Rice, and E. Hume (Oxford: Wiley-Blackwell), 557–603.

Sandler, W. (2012). Dedicated gestures the emergence of sign language. *Gesture* 12, 265–307. doi: 10.1075/gest.12.3.01san

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

Scheiner, E., Hammerschmidt, K., Jürgens, U., and Zwirner, P. (2002). Acoustic analyses of developmental changes and emotional expression in the preverbal vocalizations of infants. *J. Voice* 16, 509–529. doi: 10.1016/S0892-1997(02)00127-3

Soken, N. H., and Pick, A. D. (1992). Intermodal perception of happy and angry expressive behaviors by seven-month-old infants. *Child Dev.* 63, 787–795. doi: 10.2307/1131233

Song, J. Y., Demuth, K., and Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *J. Acoust. Soc. Am.* 128, 389–400. doi: 10.1121/1.3419786

Sullivan, M. W., and Lewis, M. (2003). Contextual determinants of anger and other negative expressions in young infants. *Dev. Psychol.* 39, 693–705. doi: 10.1037/0012-1649.39.4.693

Sumby, W. H., and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215. doi: 10.1121/1.1907309

Swerts, M., and Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Lang. Speech* 37, 21–43. doi: 10.1177/002383099403700102

Swerts, M., and Krahmer, E. (2010). Visual prosody of newsreaders: effects of information structure, emotional content and intended audience on facial expressions. *J. Phon.* 38, 197–206. doi: 10.1016/j.wocn.2009.10.002

Tomasello, M., Carpenter, M., and Liszkowski, U. (2007). A new look at infant pointing. *Child Dev.* 78, 705–722. doi: 10.1111/j.1467-8624.2007.01025.x

Tuite, K. (1993). The production of gesture. *Semiotica* 93, 83–105. doi: 10.1515/semi.1993.93.1-2.83

Vaillant-Molina, M., Bahrick, L. E., and Flom, R. (2013). Young infants match facial and vocal emotional expressions of other infants. *Infancy* 18, 1–15. doi: 10.1111/infa.12017

Vihman, M. M., DePaolis, R. A., and Keren-portnoy, T. (2009). "A dynamic systems approach to babbling and words," in *The Cambridge Handbook of Child Language*, ed. E. L. Bavin (Cambridge: Cambridge University Press), 163–182.

Vilà-Giménez, I., Igualada, A., and Prieto, P. (2017). *The Positive Effect of Observing and Producing Beat Gestures on Children's Narrative Abilities. Architectures and Mechanisms of Language Processing (AMLaP)*. Lancaster: Lancaster University.

Visser, M., Krahmer, E., and Swerts, M. (2014). Children's expression of uncertainty in collaborative and competitive contexts. *Lang. Speech* 57, 86–107. doi: 10.1177/0023830913479117

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

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviours: differentiation of multimodal information. *Psychol. Bull.* 121, 437–456. doi: 10.1037/0033-2909.121.3.437

Warren, P., Grabe, E., and Nolan, F. (1995). Prosody, phonology and parsing in closure ambiguities. *Lang. Cogn. Process.* 10, 457–486. doi: 10.1080/01690969508407112

Weber, A., Braun, B., and Crocker, M. W. (2006). Finding referents in time: eye-tracking evidence for the role of contrastive accents. *Lang. Speech* 49, 367–392. doi: 10.1177/00238309060490030301

Wellman, H. M. (1990). *The Child's Theory of Mind*. Cambridge, MA: MIT Press.

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., and Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition* 103, 147–162. doi: 10.1016/j.cognition.2006.03.006

Wimmer, H., and Perner, J. (1983). Belief about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128. doi: 10.1016/0010-0277(83)90004-5

Yasinnik, Y., Renwick, M., and Shattuck-Hufnagel, S. (2004). "The timing of speech-accompanied gestures with respect to prosody," in *Proceedings of the From Sound to Sense Conference* (Cambridge, MA: MIT), 97–102.

Zammit, M., and Schafer, G. (2011). Maternal label and gesture use affects acquisition of specific object names. *J. Child Lang.* 38, 201–221. doi: 10.1017/S0305000909990328

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.