Published in final edited form as:

Infant Behav Dev. 2021 November ; 65: 101648. doi:10.1016/j.infbeh.2021.101648.

The Origin of Language and Relative Roles of Voice and Gesture in Early Communication Development

Megan M. Burkhardt-Reed[1,2], **Helen L. Long**[5], **Dale D. Bowman**[1,2,4], **Edina R. Bene**[1], **D. Kimbrough Oller**[1,3,4]

[1]School of Communication Sciences and Disorders, University of Memphis, Memphis, TN, USA

[2]Department of Mathematics, University of Memphis, Memphis, TN, USA

[3]Konrad Lorenz Institute for Evolution and Cognition Research, Klosterneuburg, Austria

[4]Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA

[5]Waisman Center, University of Wisconsin, Madison, WI, USA

Abstract

Both vocalization and gesture are universal modes of communication and fundamental features of language development. The gestural origins theory proposes that language evolved out of early gestural use. However, evidence reported here suggests vocalization is much more prominent in early human communication than gesture is. To our knowledge no prior research has investigated the rates of emergence of both gesture and vocalization across the first year in human infants. We evaluated the rates of gestures and speech-like vocalizations (protophones) in 10 infants at 4, 7, and 11 months of age using parent-infant laboratory recordings. We found that infant protophones outnumbered gestures substantially at all three ages, ranging from >35 times more protophones than gestures at 3 months, to >2.5 times more protophones than gestures at 11 months. The results

Correspondence: Megan M. Burkhardt-Reed, 4055 N. Park Loop, Memphis, TN 38152, mreed@memphis.edu.

Author Contributions

Megan Burkhardt-Reed: Conceptualization, Methodology, Investigation, Formal analysis Writing-Original draft preparation, Reviewing and Editing, Visualization; **Helen Long:** Conceptualization, Methodology, Writing-Reviewing and Editing; **Dale Bowman:** Formal analysis; **Edina Bene:** Project administration; **D. Kimbrough Oller:** Supervision, Writing-Reviewing and Editing.

Author Note

Megan Burkhardt-Reed is PhD student in the School of Communication Sciences and Disorders at the University of Memphis in Memphis, TN, USA. Helen Long, PhD is a Postdoctoral Research Associate at the Waisman Center of the University of Wisconsin-Madison in Madison, WI, USA. Dale Bowman, PhD is an Associate Professor in the Department of Mathematical Sciences and an affiliate of the Institute for Intelligent Systems at the University of Memphis in Memphis, TN, USA. Edina Bene, PhD is a Research Project Coordinator at the School of Communication Sciences and Disorders at the University of Memphis in Memphis, TN, USA. D. Kimbrough Oller, PhD is a Full Professor and Plough Chair of Excellence for the School of Communication Sciences and Disorders, an affiliate of the Institute for Intelligent Systems at the University of Memphis in Memphis, TN, USA, and an External Faculty Member of the Konrad Lorenz Institute for Evolution and Cognition Research, Klosterneuburg, Austria.

Disclosures

None of the authors have any financial or non-financial interests to declare

Ethics Statement

This study was conducted in accordance with the recommendations of the Institutional Review Board (IRB) guidelines for the University of Memphis (https://www.memphis.edu/research/researchers/compliance/irb/index.php) with written consent from all parents of infants in the study in accord with the Delcaration of Helsinki.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

suggest vocalization, not gesture, is the predominant mode of communication in human infants in the first year.

Keywords

Infant gesture; infant vocalization; prelinguistic communication; language development; origins of language

1. Introduction

Considerable controversy about the origin of language has centered on whether vocalization or gesture played a more fundamental role (Armstrong & Wilcox, 2007; Liszkowski et al., 2011; Oller et al., 2019; Sterelny, 2012). A gestural origins theory seems to have the upper hand currently (Caselli et al. 2012; Arbib et al. 2008), invoking a key argument that since great apes communicate more flexibly in the visual-gestural than the vocal modality, our hominin ancestors must have been gestural communicators (Call & Tomasello, 2007). Another argument cites reports claiming gestural communication begins earlier than vocal communication in human infants (Caselli et al,. 2012).

In support of vocal origins of language, some suggest that vocal capabilities of great apes have been underestimated (Cheney & Seyfarth, 2005; Lameira, 2017) and that vocalization has the advantage of communicating in darkness or when receivers are not looking (Kendon, 2017). The massive amount of early infant vocal behavior and communication is also consistent with a primarily vocal foundation (Oller, Caskey, et al., 2019).

The present paper aims for the first time to quantify rates of speech-like vocalization and gesture across the first year to gain insight into which modality may play the more fundamental role. We also address the extent to which gestures and vocalizations are directed to potential receivers by gaze direction.

1.1. Gesture as a Foundation for Language

Although infants vocalize from birth, many believe gesture provides the first communicative opportunity (Bates, 1976; Iverson & Goldin-Meadow, 2005; Silva Lima & Cruz-Santos, 2012) and the primary driving force for the development of symbols (Bates et al., 1979; Gillespie-Lynch et al., 2013; Orr, 2018). Furthermore, research on early communicative behaviors has emphasized gestures as the first means to convey and structure communicative intent (Bates et al., 1979). Evidence has been presented to suggest that children use meaningful gestures several months before they use words (Caselli et al. 2012).

Pointing is viewed as constituting primitive deixis, a foundation for word learning (Iverson & Wozniak, 2016; Volterra et al., 2005; Tomasello et al., 2007). Caregivers have the opportunity to label objects of shared interest as infants begin to understand and to use pointing (Wu & Gros-Louis, 2015). However, pointing does not constitute naming, but instead designates entities that can be named.

Some theorists have supported the idea that language evolved from a primarily gestural mode (Arbib et al., 2008; Corballis, 2010; Hewes, 1973; Tomasello, 2010); great apes in captivity have shown both deliberate and voluntary gestures with distinctive functions (Byrne et al., 2017). Gestural symbols in human infants have been found to become less frequent than vocal symbols with age, but age-matched ape infants appear to use gestural symbols increasingly frequently across development (Gillespie-Lynch et al., 2013).

Research suggests captive apes communicate primarily through gesture (Pika et al., 2005; Pollick & De Waal, 2007; Tomasello & Zuberbühler, 2002). But unlike humans, our ape relatives do not normally assemble vocalizations into complex utterances composed of syllables, words, and sentences (Riede et al., 2005), with perhaps rare exceptions (see e.g., Clay & Zuberbühler, 2009). Gestural flexibility as opposed to vocal flexibility in great apes is supported by relatively successful sign language learning in great apes raised by humans, but almost total failure to learn spoken language in the same circumstances (Bonvillian & Patterson, 1999; Gardner & Gardner, 1969; Rivas, 2005).

1.2 Vocalization as a Foundation for Language

A variety of non-cry, speech-like vocalizations, called “protophones”, are a primary means by which infants engage in communicative interaction soon after birth (Gratier et al., 2015). By ~7 months protophones come to include well-formed (or “canonical”) syllables (Oller, 1980). Longitudinal investigations indicate newborn infants produce protophones from the first weeks (Koopsmans-van Beinum & Van der Stelt, 1986; Oller, 2000; Stark, 1980; Nathani et al., 2006), at much higher rates than cries in both preterm and full-term infants, with preterms producing protophones in neonatal intensive care from as soon as they can breathe on their own (Oller, Caskey, et al., 2019). Human infants produce protophones at least ten times more frequently than chimpanzees and bonobos produce sounds viewed as analogous to protophones (Oller, Griebel, et al., 2019). Spoken words begin at the end of the first year, but human infants use protophones to communicate needs and states of being from the first month (Gratier et al., 2015; Jhang & Oller, 2017), long before gestural communication has been reported. The importance of vocal communication is emphasized by evidence that gaze-coordinated vocalizations are stronger predictors of later language outcomes than either gestures alone or gesture-vocal combinations (Donnellan et al., 2020).

Long et al. (2020) emphasized the endogenous nature of infant vocalization; in the first year infants produced three times as many protophones independently as during social engagement. From 9 to 18 months, babbling practice alone, unaffected by social environment, has been found to be a strong determining factor for word onset (McGillion et al., 2017).

1.3 An Evolutionary-Developmental Perspective

Evolutionary-developmental (evo-devo) biology emphasizes the widespread tendency for new structural features or capabilities to evolve by modification of developmental patterns (Müller & Newman, 2003). In evo-devo theory, conservation of foundational structures is expected, and natural selection is seen to build upon the foundational structures (West-Eberhard, 2003). Thus, the order of appearance of structures or capabilities/activities in development is expected to emerge following evolutionary orders (Carroll, 2005; Newman, 2016). In accord with this line of thought, one should predict that if gesture forms the primary foundation of language, then gestural communication should predominate in early communication in humans, and conversely if vocalization forms the primary foundation, then vocalization should predominate.

1.4 Aims

Although there exist well-established procedures for judging both the structure and communicativeness of protophones from birth and across the whole first year, we know of no descriptive framework identifying gestures and their potential communicative roles that is applicable to the whole first year. In the present work, we propose a framework to allow comparable counting of communicative and/or potentially communicative events of both infant gesture and vocalization. Thus, we aim to provide new perspectives on the relative roles of gesture and vocalization in modern human communicative development and indirectly in the evolutionary origin of language. We expect that empirical data will contradict the expectation, based on the gestural origins theory, that gesture should occur more frequently than vocalization in the first year. We hypothesize the opposite:

1. speech-like vocal events (protophones) will occur at a higher rate than gestures in early human development and
2. protophones, more often than gestures, will show signs of constituting intentional communications, being accompanied more often by gaze directed toward another person.

2. Methods

2.1 Selection of Participants

Data were acquired from archived longitudinal audio-video recordings from the from the University of Memphis Origin of Language Lab-oratories. Approval for the research was obtained the University of Memphis Institutional Review Board for the Protection of Human Subjects (IRB). All participants resided in or around Memphis, Tennessee. Recruitment was conducted in child-birth education classes and by word of mouth. All infants' parents completed a written consent approved by the IRB prior to recordings. An inclusion criterion for participation was normal pregnancy. Typical development (i.e., lack of hearing, vision, language, or other developmental disorders) was confirmed throughout participation in the study via parent report using information such as passed hearing screenings and mastery of developmental milestones at expected ages. None of the infants was born prematurely.

The archived recording sessions included 21 parent-infant dyads from two waves of longitudinal study (see e.g., Oller et al., 2013; Oller, Caskey, et al., 2019). Based on funding available for coding and analysis, we were able to select only a subset for the present research, with recordings from 10 infants (5 male, 5 female) balanced for age, gender, recording session type, and recording length. The recordings from which the data were drawn usually included three sessions, each approximately 20-minutes in length, often based on a single continuous ~60-minute recording. We analyzed only the sessions termed “interactive” at approximately 4, 7, and 11 months—thus we coded 30 sessions, one for each infant at each age.

In the interactive sessions, parents were instructed to engage their infants as they normally would, whereas the other two recording sessions during the ~60 minutes required the infant to be present with the parent reading (the no-talk-to-baby circumstance) or playing separately while the parent was interviewed by another adult. We selected the interactive sessions for our study, assuming gestures would more likely occur in those sessions since parents were more likely to look at infants during those sessions than during the no-talk-to-baby or interview sessions. Thus, selection of the interactive sessions served our intention to maximize the occurrence of gestures.

We selected infants with recordings fitting the age criteria to the extent possible while taking advantage of existing vocalization and gaze coding from prior work. All but one of the infants were White (Infant 4). All were learning English as their native language except for Infant 2, who was exposed to German, Spanish, and English. All infants were of middle to low-middle socio-economic status. Demographics and recording ages are provided in Table 1. Selections were made without regard to rate of occurrence of gestural or vocal activity.

2.2 Data Collection

One recording was selected for each infant (total: 10) at each age (total: 3), yielding 30 recordings. The actual length of the planned 20-minute sessions was ~19 minutes (range: 16 – 20 minutes). All parent-infant dyads were recorded in a quiet laboratory/playroom with toys and books appropriate for the infant age. Both infants and parents wore high fidelity wireless microphones. The playroom was equipped with cameras in the corners, either four cameras or eight (one high camera and one low camera in each corner). The differing number of cameras corresponded to three phases of recording in three different laboratories. Laboratory staff operated the cameras for zoom and tilt from an adjacent control room. Two of the four or eight channels were selected to record the interaction at each moment. Selection and zooming afforded close-up views of the infant face, torso, and actions on one of the selected channels and a broader view of the interaction (including the parent) on the other channel. Details regarding laboratory equipment and procedures can be found in previous work from this laboratory (Buder et al., 2008; Oller et al., 2013). Instructions to parents for the interactive segments emphasized playing with and interacting with infants in a natural way, allowing for vocal, gestural, and tactile interaction at any time (additional information on recordings in SM, 2.1).

2.3 Coding Approach and Rationale

2.3.1 Coding Rationale—Gesture and vocalization are not easy to compare given that the two modalities have different advantages and disadvantages. Still, it is both appropriate and necessary to quantitatively compare gesture vs. vocalization across development in order to address the relative roles of gesture and vocalization in language origins (see SM, 1.1). We have striven to construct an approach allowing well-motivated quantitative comparison. We sought to ensure that comparisons would not bias outcomes in favor of vocal acts; on the contrary, we sought to ensure that any bias would favor gestures (see SM, 2.3). We selected an event-based analysis, not a duration based-analysis. This choice is motivated by the fact that *events* of communication or potential communication are the optimal points of comparison, since each gestural or vocal event is a potential communicative act (for duration comparisons see SM, 3.2).

2.3.2 Software Environment—

Coding was conducted in AACT (Action Analysis Coding and Training software, Delgado et al., 2010), a software environment facilitating simultaneous coding of video/audio. AACT presents two channels of video synchronized with audio at frame-level accuracy, with audio displayed spectrographically and in waveform through a version of TF32 (Milenkovic, 2010) designed for AACT. The coding fields of interest for the present research were protophone types, gestural act types, gestural illocutionary functions, and gaze direction accompanying both protophones and gestures (see SM, 1.3, 2.4 and 3.4). Regarding facial affect coding, see SM, 3.3.

Aside from gesture coding, data collection was completed in a way that was identical to previous infant vocalization studies conducted in the in the Origin of Language Laboratories (Jhang & Oller 2017; Long et al., 2020; Oller et al., 2013). During coding, the coder places a start and an end cursor on the TF32 spectrographic display and can play the selected sequence repeatedly in an AACT “loop”. The cursor can be dragged on the spectrographic display or shifted with keyboard controls producing frame accurate shifting in the video from both cameras to enable selection of onset and offset points for events in any field.

2.4 Vocalization Coding

Vocalizations had been coded in prior research (e.g., Oller et al. 2013, Oller, Caskey, et al., 2019). The primary focus of this previous work was speech-like vocalizations, focusing on phonatory properties. This approach resulted in three primary types: vowel-like (vocant), growl-like, or squeal-like sounds. Oller (2000) refers to these types as “protophones” (including both canonical and non-canonical sounds). Thus, syllables or syllable sequences such as “dada” or real words were, like precanonical protophones, categorized in terms of phonation as vocant, growl or squeal. Cries, laughs, and whimpers as well as a variety of additional infrequently occurring types (whispers, voiceless friction sounds, ingressive sounds) were also coded but not included for analysis.

Coding was conducted with repeat-observation, assigning boundaries in TF32 at the onset and offset of each protophone. Cursor placements for each utterance were recorded in AACT, specifying duration in ms. Boundaries were assigned using a “breath-group” criterion (Lynch et al., 1995); thus an utterance was determined to begin with phonation during exhalation (i.e., egress) and end with termination of phonation, often accompanied by inhalation (i.e., ingress). Thus, a new utterance could begin only after an observed breathing pause. A protophone type was then selected from the coding panel, and in a subsequent coding pass, a gaze direction category was assigned to the time period of each vocalization. If any gaze was directed to a person during the utterance, even if only briefly, the utterance was coded as having been directed to a person. Details on vocalization and gaze direction coding and associated coder agreement can be found in prior publications, especially in their supplementary materials (e.g., Jhang et al., 2017; Oller et al. 2013).

2.5 Gesture Coding

2.5.1 Global Categories for Gesture Coding and their Relation to Global Vocalization Categories—

The coding scheme was intended to maximize comparability between events in the gestural and vocal domains. We sought to categorize gestural acts during the first year that could be considered precursors to signs (i.e., precursors to gestural symbols) just as we categorized vocal acts that could be considered precursors to words. The coding consisted of four global movement categories: 1) Utilitarian Acts, 2) Non-social Gestures, 3) Universal Social Gestures, and 4) Conventional Gestures. We defined *Utilitarian* Acts as those actions that are simply world-exploratory or manipulative, without any inherently social communicative function. For example, if a child reached for and obtained any object, the event was coded as a Utilitarian Act. These acts were not counted as gestures even though they were coded in the initial real-time coding pass. Vocalizations also include Utilitarian Acts in the form of vegetative sounds (coughing, burping, clearing one's throat, blowing out a candle…) that perform functions related to respiration and digestion—vegetative sounds were not counted as protophones.

The remaining three global gestural categories were intended to include all the acts that could conceivably be interpreted as communicative or those that could be expected to be brought into the service of communication at some later point in development. *Non-social Gestures* are gestural actions that are not merely Utilitarian and are also not inherently communicative, although they have *the potential for being utilized communicatively*. For instance, rhythmic hand banging and foot tapping are examples of Non-social Gestures, akin to babbling/protophones in the vocal domain, actions that can eventually be brought to the service of communication, but that are not yet intentionally communicative. Non-social Vocalizations included all the protophones—the very infrequently occurring words were categorized as both words and protophones, but words were treated as Social, while non-word protophones were treated as Non-social.

Universal Social Gestures include acts with an inherently social communicative intent but with no reason to presume they are learned from specific cultural experience. For example, an extended flat hand to indicate refusal is a Universal Social Gesture. Also, reaching upward with both hands when an infant wishes to be picked up is a Universal Social Gesture. Pointing and other clearly deictic gestural acts are also Universal Social Gestures.

Universal Social Gestures have fixed functions in infancy; e.g., deictic gestures designate entities but cannot name them or perform other affectively-valenced functions such as expressing distress or delight (see SM, 1.2). Other Universal Social Gestures, including reaching toward someone to request being picked up or a flat hand outstretched to indicate refusal are also fixed in function, in that they cannot acquire a different function through learning in infancy. Similarly, there are *vocal* acts in infancy that have fixed functions (the Universal Social Vocalizations), crying and laughter, which express negative or positive affect respectively, but cannot designate objects, as deictic gestures can. While we counted Universal Social Gestures in the quantitative comparison of gestures and protophones, Universal Social Vocalizations were not included (see SM, 2.3).

Conventional Gestures are those that are culturally transmitted, acts with a discernible communicative function, such as waving to convey "hello" or "bye-bye", clapping in celebration, or thumbs up to indicate approval or agreement. These gestures are learned and can be viewed as analogous to primitive words, i.e., Conventional Vocalizations. Non-word protophones (including both non-canonical and canonical babbling) can, in accord with our scheme, be subcategorized as Non-social, while words (treated here as a subclass of protophones) are subcategorized as Conventional Vocalizations (speech). The four global categories, then, facilitate comparison across vocalization and gesture since both the gestural and vocalization coding schemes presume the same four.

It is important to recognize comparable aspects of the two modalities while also taking account of inherent differences in how the two modalities can function (see SM, 2.2). Universal Social acts, both of gesture and vocalization are communicative but require no associative learning, and their intended functions are interpretable to potentially everyone around the world. Universal Social Gestures are capable of transmitting certain critically important communicative functions such as refusal, request, and designation (pointing), while in the exclusively vocal domain, these functions seem impossible to transmit without symbols (words), and even then, vocal transmission is more complicated. But vocalization has the advantage of being able to transmit affective valence (Jhang et al., 2017), which is difficult if not impossible to transmit by gesture alone. Thus, a fundamental difference between the modalities is that vocalization is well-suited to transmission of emotional valence, while gesture is well-suited to transmission of deixis.

Consider designation, which is possible with words ("look to your left and notice an orange object"), but not with protophones. Even in young infants, a simple act of pointing can serve as a designation (directing attention to the orange object without a single word). Similarly, other functions that can be transmitted with gesture universally even in infancy (refusal or request, for example) cannot be uniquely transmitted in vocalization without words ("I don't want it", "stop", "give me the book"). Still, prosodic features of vocalization and/or facial affect can emphasize or modulate the flavoring of such functions in either modality and regardless of whether the communication is prelinguistic or based on signed or vocal symbols.

There is thus a gap in the potential for communication transmitted with protophones, a missing deictic function, a gap that can be at least partially filled by Universal Social Gestures, which consequently provide a scaffold for early communication and especially for learning of labels. This special capability of Universal Social Gestures may account for the widespread opinion that human language is founded in gesture. Yet, vocal communication begins in the first week of life, while gestural deixis does not appear until late in the second half year. The gestural-origin opinion also neglects the fact that gesture has a gap in transmission of affective valence, a gap which is partly filled with Universal Vocal Acts such as crying, whimpering, laughter, and whining, which may also supply support for the emergence of language. Furthermore, deictic gestures, while supplying a support system for learning symbols in either the gestural or vocal domain, are not themselves symbols (SM, 1.2).

2.5.2 Gesture Coding Procedures—The primary coding was conducted by the first author. We drew a distinction between gestural *actions* and the potential *intent*, or communicative function, associated with each action. This distinction was also drawn between *vocal* actions and their communicative functions. For example, "reaching" is a code for an action, but that code does not characterize the communicative function of the action. An infant could reach to show an object, to request an object, to try to get the attention of another person, or to offer or accept something. Our coding scheme included two fields, in one case for each individual gestural action and in the other for the communicative function of each action.

The actions and functions were coded for each ~19-minute segment in three separate passes using both audio and video. The first pass, using real-time observation, allowed the coder to acquire an overview of the recorded segment and to mark the approximate location of infant gestures, labeled by keystrokes in real-time, each corresponding to one of the global codes: Utilitarian, Non-social, Universal Social, or Conventional. In the second pass, the coder ignored Utilitarian Acts but used repeat-observation to designate boundaries for individual Non-social, Universal Social, or Conventional gestural actions at the onset and offset of each action, specifically designating the point of the movement beginning the event through the point where the event reached its full extension. For example, the onset of an index-finger point begins when the arm, hand, or finger moves into motion from rest until the moment where it reaches its full extension. In cases of rhythmic movements such as hand banging, boundaries were determined using a criterion similar to the breath-group criterion for vocalizations. Thus, duration of any cluster of actions (such as the individual strokes of rhythmic hand banging), deemed to constitute a gesture occurring before a gestural pause (treated similarly to pauses in speech), specified the duration of the gesture.

In the second pass of gesture coding, the cursors in TF32 were adjusted to make each boundary decision, which sometimes required the cursors first to be set as much as 1000 ms before the onset and after the offset of the gesture to allow extended viewing of the event surroundings. Then the cursors could be moved (by dragging them in the acoustic display with corresponding frame accurate changes in the video or by keystrokes that could move the video display one frame at a time with a corresponding shift in the audio display) to home in on the actual boundaries of the gesture specifying the onset and offset points. A keystroke indicating a particular gestural act was then recorded in AACT, indicating both onset and offset times.

Once onset and offset boundaries had been determined, we used the bounded time frames of the gestural actions to automatically create placeholders in a new coding panel (the gestural function panel). The placeholders were recorded sequentially in the gestural function panel (without showing the gestural action codes) and allowed the coder to categorize all the bounded events in a third pass, where a gestural function was designated for the precise time frame of each gestural action, the entire period of each action being taken into account in coding the function. The second and third coding passes for actions and functions used the detailed gestural category labels as listed in Table 2 and Table 3, respectively.

2.6 Gaze Coding

2.6.1 Rationale for gaze coding—Previous studies have argued that gaze coordination can be used as an indicator of intentionality in prelinguistic vocalizations and gestures (Bates, 1976; Donnellan et al., 2020; Harding & Golinkoff, 1979; Iverson, 2010; Iverson et al., 2000; Thal & Tobias, 1992; Wu & Gros-Louis, 2015). Gaze direction is an excellent predictor of judgments of social directivity based on the conjunct of factors (timing, prior social context, etc.) that suggest an infant as trying to communicate (Long et al. 2020).

2.6.2 Gaze coding procedure—In a fourth pass, gaze direction was coded during each vocalization and during each gesture where at least one of the two video views allowed such judgement. The coding determined whether infants produced a protophone or a gesture while looking at another person (i.e., a socially-directed communication) or while not looking at another person (i.e., a non-socially-directed communication).

As indicated above, the gaze direction for protophones had been previously coded during other studies from our laboratory, and those judgments were used in the current study. Following the same coding procedure as in the third pass of gesture coding, coders clicked on each individual placeholder event indicating a previously coded gesture or vocalization and then coded the direction of the infant's gaze during the bounded action plus 50 ms on either side. Playback thus started automatically 50 ms before the event onset and continued through 50 ms after the offset. This procedure was designed to ensure that all video frames of each event would be viewed before determining the directedness of the infants' gaze. The categories for coding were 1) directed toward another person, 2) not directed toward another person, or 3) can't see (sometimes the infant's gaze direction could not be judged based on either camera view). If there was any moment of gaze direction to a person observed during the gesture or vocalization interval (the whole event plus the 50 ms additions), the event was coded as person directed.

2.7 Coder Agreement Training and Agreement Outcomes

Two graduate students were trained as agreement coders for gestural actions, gestural functions, and gaze direction. The training began with a lecture by the first author on the gesture coding scheme, during which coders were presented with examples of video-recorded infant gestural actions previously coded by the first author and confirmed by the last author, all of which either met a consensus standard for one of the gesture categories or displayed ambiguities of possible judgements deemed instructive for training. Gaze direction training was similar.

Once training was completed, coders followed the same coding procedure as the first author, using the criteria outlined in the gestural coding scheme and the gaze direction scheme. The first author selected 12 recordings semi-randomly from the 30 recordings for the agreement study, with one five-minute segment selected from within each of the 12 recordings. Four samples came from each of the three ages and all ten infants were represented in the agreement samples. Neither of the agreement coders knew the hypotheses for the study and were blinded to coding of the first author.

There was high agreement for the number of gestures identified in the five-minute segments for both agreement coders with respect to the primary coder ($r = 0.92$, $r = 0.81$, $N = 12$), and for the two with respect to each other ($r = 0.80$). On proportion of gestures with gaze directed to a person, the agreement coders also showed good agreement with respect to the primary coder ($r = 0.92$, $r = 0.73$) and each other ($r = 0.72$).

Across the five-minute segments in the agreement samples, all three coders showed at least a doubling of the rate of gesture from the first to the second age, with an increase ranging from ~50% to 60%. The three coders also showed an increase ranging from 22% to 65% in the amount of gesture from the first to the third age. Also, the three coders showed very similar proportions (27%, 30% and 30%) of person-directed (by gaze) gestural actions. For all three coders, 13% to 14% of gestures were deemed to be directed to a person, whereas degree of directivity for protophones was > 27% for all three. The agreement data suggest that if the entire data set had been coded by either of the agreement coders instead of the primary coder, none of the conclusions associated with the results reported below would have changed.

3. Results

3.1 The Hypotheses

3.1.1 Hypothesis 1, Distribution of Gestures and Protophones—

The data on protophone and gesture rates (Figure 1) across all 30 recordings revealed vastly more protophones (3903) than gestures (752), with gestures occurring infrequently (only about one every 4 minutes) at the earliest age, increasing to 1.7 – 2 per minute at the later ages. In contrast, protophones occurred approximately 10 times per minute at the youngest age, and nearly 6 per minute at the oldest. Thus, there were more protophones than gestures, and this was especially true at the youngest age. These results support a more vocal than gestural origin for communication.

Shapiro Wilks tests for normality on the gesture and protophone rates were run separately. Based on the tests, protophone rates and protophone proportions of social directivity can be assumed to be normally distributed, but gesture rates and proportions of social directivity were far from normal, primarily because of zeros at the earliest age. Consequently, a non-parametric approach to repeated-measures analysis was necessary.

Data were analyzed with Generalized Estimating Equations, a nonparametric alternative to generalized linear mixed models, producing unbiased regression estimates for use in longitudinal or repeated-measures research designs with non-normal response variables (Liang & Zeger, 1986; Ballinger, 2004). Indeed, the data in question were nonnormal especially because of the very low and unequal numbers of gestures at the early Age. We determined GEE was appropriate also because of the unequal amounts of data in the two Modalities across Age and lack of precise Age matching across infants. The model included two Modalities (gestures vs. protophones) and a factor with three Ages (allowing comparison of early to middle and early to late), with 10 infants at each Age. The dependent variable was number of events produced, either gestures or protophones. The covariance matrix was exchangeable. With an unstructured matrix, the GEE failed to converge. The robust sandwich estimator was used to obtain standard errors of estimates.

GEE revealed a significant early to middle Age by Modality interaction ($p < .02$) as well as a stronger early to late Age by Modality interaction ($p < .0005$), reflecting the dramatic increase in the number of gestures across Age, in contrast to protophones, which showed no such increase, but in fact fell in frequency across Age, though not so dramatically as gestures rose. A significant main effect of Modality ($p < .00001$) reflected the much larger number of protophones than gestures. Main effects of early vs. middle Age ($p < .00001$), and early vs. late Age ($p < .00001$) reflected a mean increase in the combination of gesture and protophone rates across age.

Because of lack of uniformity of ages within the groups (see Table 1), we also ran GEE with Age as a continuous variable. The results indicated a significant Age by Modality interaction ($p < .005$), reflecting the fact that gestures increased dramatically with Age, while protophones fell. There was also a significant main effect for Modality ($p < .00001$), as in the analysis with Age as a three-level factor, and a significant effect for Age ($p < .00001$), again reflecting growth in the combination of gesture and vocalization across Age. GEE details are in SM, 3.1, Table SM1 and SM2.

Follow-up Mann Whitney U tests (non-parametric) also showed a significant difference reflecting more protophones than gestures (z score = 3.74, $p < .0002$). In addition, Mann Whitney U tests confirmed that gesture rate was lower at the early than at the middle (z score = 3.10, $p < .002$) or late Ages (z score = 3.36, $p < .001$). The difference between the middle and late Ages was not significant. For protophones, rates fell across Age, but only the comparison between early and late Ages was significant (z score = 1.97, $p < .05$).

3.1.2 Hypothesis 2, Directivity of Gestures and Protophones—

The data on directivity of gestures and protophones indicated, consistent with Hypothesis 2’s prediction, that protophones were more likely than gestures to be socially directed as indicated by gaze direction toward a person. Averaged across ages, the grand total of protophones was nearly twice as likely to include person-directed gaze as gestures (34.6% to 17.7%). The tendency was greatest at the youngest age (37.6% to 21.7%) and least at the latest age (30.2% to 17.7%). Figure 2 presents averages at the infant level, illustrating a tendency for the directivity proportion to be highest at the latest age for gestures, but lowest at the latest age for protophones.

To compare proportion of directed protophones vs. gestures statistically, we selected a similar GEE model to the one used for Hypothesis 1, with Age as a three-level factor. The dependent variable was proportion of events in each Modality directed by gaze to a person (Figure 2). GEE revealed a significant interaction of early to late Age with Modality ($p < .05$), reflecting conflicting tendencies; gestures showed greater directivity at the late than the early Age while protophones showed the opposite. There was no significant difference for the interaction of early to middle Age with Modality. As with the test for gesture and protophone counts, GEE revealed a very strong main effect for Modality (p < .00001), indicating that protophones were more often socially directed than gestures. The result again supports the prediction that protophones are more socially-directed than gestures, diametrically contradicting the gestural origins theory prediction that early gestures should be more socially directed than vocalizations. The analysis also revealed a main effect for early vs late Age ($p < .05$), reflecting an overall tendency for directivity to be higher at the late Age for gestures and protophones combined.

As with the analysis of gesture and protophone counts, we conducted a separate GEE analysis of directivity proportions with Age as a continuous variable rather than as a three-level factor. The results indicated a significant interaction of Age by Modality ($p < .00001$), a significant main effect of Modality ($p < .00001$), and a weaker significant effect of Age ($p < .02$). Details of these GEE analyses are in the SM, 3.1, Tables SM3 and SM4.

Follow-up Mann Whitney U tests also showed the proportion of socially-directed protophones was greater than that of gestures (z score = 2.91, $p < .001$). The same pattern occurred for early (z score = 2.78, $p < .006$) and middle Ages (z score = 2.38, $p < .02$) but was not significant at the late Age. Follow-up tests also showed gesture had higher directivity at the late than the early Age (z score = 2.15, $p < .05$), but other Age comparisons were not significant. No Age comparison by Mann Whitney U test was significant for proportion of protophone directivity. The great majority of all events were not directed toward another person in *either* Modality.

3.2 Distribution of Gesture and Protophone Types

Table 4 shows the distribution of three global gesture categories. Numbers of Non-Social and Universal Social Gestures at the early and middle ages were similar and were by far the most frequent gesture types at the early and middle ages. At the late age, ~60% of gestures were Universal Social Gestures. Pointing only occurred 12 times in the ~600 minutes of coded recording, 9 times from one infant at 11 months, and three times from another infant. Both at the middle and late ages, the great bulk of Universal Social Gestures involved reaching as if to request an object, offering an object to another person by reaching, or accepting an object by reaching.

Conventional Gestures were not frequent at any age (42 total gestural events, 26 of which occurred at 11 months). Those occurring at 11 months included 4 types: no-no (head or finger shaking), hand clapping, bye-bye (waving), and face covering. All Conventional Gestures at the middle age were associated with a peekaboo game with one infant only.

Although the available coding of protophones did not specify them for the three global categories (social and non-social usage of vocalizations had not been coded in the earlier studies), we evaluated tokens of Conventional Vocalizations across the three ages. We found 50 conventional words (including mama, dada, bye-bye, no-no, yum-yum, mmm (tastes good), and yeah), compared with the 42 Conventional Gestures.

4. Discussion

4.1 Outcome Summary

The present study provides the first direct comparison of gesture and protophone rates across the first year. The infants produced more than five times as many protophones (3903) as gestures (752), with a substantial imbalance favoring protophones at all three ages. The vocal predominance applied whether the communication was directed or not directed by gaze toward another person. There were 1,349 cases where a protophone was directed to a person by gaze, compared with 133 cases of directed gaze for gesture. Furthermore, protophones were nearly twice as likely to be person-directed as gestures (34.6% to 17.7%). If then, we take gaze direction as an indicator of communicative intent, we can conclude that not only did protophones occur much more frequently than gestures, but protophones were far more likely to be intentionally communicative than gestures. The results diametrically contradict the gestural origins theory: early communicative development was not dominated by gestures, but by protophones.

The interactions between Age and Modality indicate that gesture not only became more frequent across the first year, but also became more socially directed, while the opposite occurred for protophones. We are doubtful about the replicability of the apparent fall in protophone rate across Age, partly because the effect was only statistically significant from the early to the late Age, and also because other research has not clearly found such a pattern (Gilkerson et al., 2017; Oller, Caskey, et al., 2019; Iyer & Oller, 2008; Iyer et al., 2016). But the significant interaction suggesting a rise in directivity of the gestures and a seeming fall in directivity of the protophones (patterns that have to our knowledge never before been addressed) inspires us to consider an interpretation based on the possibility that infants learn about the communicativeness of their gestures and protophones across time and that they adjust their gaze direction accordingly. Namely, the results suggest the possibility that infants learn across the first year that someone needs to be looking at them for their gestures to be communicatively effective, while also learning across the same period that their vocalizations can often be communicatively effective even without anyone looking at them.

4.2 Inherent Differences Between the Gestural and Vocal Modalities

At the late age, most gestures were Universal Social Gestures, highlighting a particular communicative feature of the gestural modality that is difficult to implement vocally (see SM, 1.2). Universal Social Gestures communicate intents, e.g., indicating something the infant wants or may want another person to look at. In contrast, vocalizations must first become symbolic in order to serve such functions. An infant can gesture by extending his or her arm(s) to signal the desire to be picked up, but there is no equivalent in the vocal domain without words (e.g., “pick me up”).

On the other hand, vocalization can transmit emotional valence, and thus can be used universally to modulate the affective tone of communication. It is universal for caregivers to recognize cry and fussy protophones; each protophone type can be flavored by intonation or other acoustic modulations to convey affect. Approximately 15% of protophones in laboratory recordings labeled as fussy can be judged as negative from sound alone (Jhang & Oller, 2017). Thus, protophone prosody can assist in flavoring affect of communication in a way gestures cannot, but prosody cannot supplant a gesture's deictic function. Facial affect can be utilized to modulate the emotional tone of communication in either gesture or vocalization.

Vocalization can be used to assist in gestural communication by supporting attention seeking (Franco et al., 2008; Gros-Louis & Wu, 2012). Pointing or reaching are often accompanied by vocalizations to attract listener attention. But even in these cases, the vocalizations (unless they are words) cannot serve the deictic functions that are natural to the gestural domain.

An important feature of protophones is functional flexibility (Oller et al. 2013). This feature contrasts sharply with the functions associated with Universal Social Vocalizations or Universal Social Gestures. Crying, for example, is naturally associated with a function of distress expression. Pointing is naturally associated with a designative function. All universal communicative acts have a specifiable function or class of functions. In contrast, no protophone type has a universal function—all protophones can be produced to serve multiple functions. It is a critical feature of protophones that they must be free to serve functions with all possible valences, because if they were not free in that way, they could not form a foundation for symbolic words, which are by definition free of any particular illocutionary function.

Of course, both gesture and vocalization are capable of developing into full-fledged language with full functional flexibility. Both modalities include actions that are not inherently tied to particular universal functions. With both Non-Social Gestures and protophones, infants explore actions free of particular function; only with this freedom can they be adapted at a later point by learning to form Conventional acts.

The flexibility of protophones is emphasized by recent results from coding of laboratory recordings, where it was found that ~75% of infant protophones are produced without social directivity (Long et al., 2020). This fact suggests protophone production is largely endogenous. It has been reasoned that protophone production offers caregivers information about an infant's well-being (Locke, 2006 and see SM, 1.3). The same kind of reasoning may be thought to apply to gestural babbling, i.e., to Non-Social Gestures, but if gesture formed the primary foundation for language, we would not expect infants to look towards caregivers only about half as often during gestures as during protophones.

4.3 Interpretations of Unanticipated Findings

The great majority of gestures observed were not the ones expected based on the common suggestion that language is founded in gesture. Pointing, known to form a foundation for word learning, occurred far less frequently than expected. Only 12 pointing events were observed (<2% of observed gestures). An even lower frequency of pointing was found in recent work on communicative vocalizations and gestures at the end of the first year in a study of 134 prelinguistic infants (Donnellan et al. 2020), where only 28 cases of indexical pointing were observed, and many of those were not gaze-coordinated.

In our own data, reaching accounted for 48% of all gestures, suggesting that the “declarative” function of pointing, thought to be so important as a foundation for language (Bates et al. 1979), occurs far less frequently than the instrumental functions associated with reaching through 11 months of age. Only one other category of gesture occurred frequently. Rhythmic hand shaking, which we interpret as a Non-social Gestural act similar to reduplicated babbling, accounted for 35% of gestures overall and 26% at 11 months. Together, hand shaking and reaching accounted for 83% of gestures.

Both Conventional Gestures and Conventional Vocalizations were infrequent in the first year, with slightly more Conventional Vocalizations (words) than Conventional Gestures, and both the Conventional Gestures and the words were overwhelmingly performatives, that is, they constituted illocutionary acts such as greeting (waving, saying hello), celebrating (clapping, saying hooray), or refusing (head shaking, or saying no), rather than semantic acts of reference, such as naming an object or describing an event. Thus, the Conventional acts in both modalities were overwhelmingly dyadic, constituting communications between two parties with respect to each other, rather than being triadic communications where the two parties jointly referred to a third entity through a conventional symbolic act.

4.4 Evaluating Events of Communication, Individual Differences, and Evolutionary Implications

Although the average duration of a protophone was only about half as great as that of a gesture, protophones in the recordings accounted for considerably more time than gestures (SM 3.2). Comparison at the event-level is more useful than duration comparison, because each event in either modality constitutes a possible communicative act, and it is the number of such possible communications that matters most in assessing the relative importance of gesture and vocalization.

There was notable variation among the 10 infants in protophone rate, perhaps due largely to natural day-to-day variability, but the variation is intriguing nonetheless. Even in the vocal domain, individual variation was salient, with a low of 35 protophones at the middle age from one infant to a high of 281 from a different infant at the early age. Of special interest was the fact that only two infants accounted for >70% of all gestures at the early age, and those two were the only infants who were >3 months at the time. The largest number of gestures among the other 8 infants at the early age was 4, and five infants produced either 0 or 1 gesture at that age. Consequently, it is tempting to speculate that human infants do not significantly engage in gesture until after 3 months.

In spite of substantial individual variation, the key differences were robust. All 10 infants produced more protophones than gestures; the infant producing the fewest protophones produced 1.9 times more protophones than the gestures produced by the infant producing the most gestures. 9 of the 10 infants produced a higher proportion of person-directed protophones than person-directed gestures; even the infant who produced a higher proportion of person-directed gestures than protophones overall, did so only at the late age, while all the other infants showed a higher proportion of person-directed protophones than gestures at all three ages.

Gestures and vocalizations did not tend to co-occur. Only 17% of gestures overlapped with a protophone (see SM, 3.5). Thus, the data did not suggest extensive coordination of gesture with protophones across the first year.

One might suggest that motoric development is simply faster in the vocal domain than in the domain of hand and arm movement. But this suggestion does not undercut the conclusions of the present work, because the earlier development of vocal capacities, which are known to be among humankind’s most complex motoric capacities (see SM, 2.2), would require explanation of its own. In fact, motoric development of vocal capacities may have been naturally selected to occur early, precisely because of the importance of hominin infant vocalization as a signal of wellness across human evolution, a signal that could have been noticed even when caregivers were not looking, while gesture would have had no such advantage.

The relative tendency to communicate in the vocal domain compared to the gestural domain in early life is not only important in informing our understanding of the emergence of the speech capacity in modern human development, but it also offers insights into the likelihood that vocal communication predominated in the evolution of language. We reason, consistent with the evo-devo perspective of modern theoretical biology, that if language indeed originated from gestural use, gestural activity should have occurred to a far greater extent than we saw.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The research for this manuscript was funded by NIH Grants DC006099 and DC0011027 from the National Institute on Deafness and Other Communication Disorders and by the Plough Foundation, which supports D. Kimbrough Oller’s Chair of Excellence.

5. References

Arbib MA, Liebal K, & Pika S (2008). Primate Vocalization, Gesture, and the Evolution of Human Language. Current Anthropology, 1053–1076. [PubMed: 19391445]

Armstrong DF, & Wilcox SE (2007). The gestural origin of language. Oxford University Press. 10.1093/acprof:oso/9780195163483.001.0001

Bates E (1976). Language in context. New York: Academic Press.

Bates E, Benigni L, Bretherton 1, Camaioni L, & Volterra V (1979). The emergence of symbols: Cognition and communication in infancy. Academic Press.

Bonvillian JD, & Patterson FG (1999). Early sign-language acquisition: Comparisons between children and gorillas. In Parker ST, Mitchell RW, & Miles HL (Eds), The mentalities of gorillas and orangutans: Comparative perspectives (p. 240–264). Cambridge University Press. 10.1017/CBO9780511542305

Buder EH, Chorna LB, Oller DK, & Robinson RB (2008). Vibratory regime classification of infant phonation. Journal of Voice, 22(5), 553–564. 10.1016/j.jvoice.2006.12.009 [PubMed: 17509829]

Byrne RW, Cartmill E, Genty E, Graham KE, Hobaiter C, & Tanner J (2017). Great ape gestures: Intentional communication with a rich set of innate signals. Animal Cognition, 20(4), 755–769. 10.1007/s10071-017-1096-4 [PubMed: 28502063]

Call J, & Tomasello M (Eds.). (2007). The gestural communication of apes and monkeys. Taylor &Francis Group/Lawrence Erlbaum Associates.

Carroll SB (2005). Evolution at two levels: On genes and form. PLoS Biol, 3(7), e245. 10.1371/journal.pbio.0030245 [PubMed: 16000021]

Caselli MC, Rinaldi P, Stefanini S, & Volterra V (2012). Early action and gesture “vocabulary” and its relation with word comprehension and production. Child Dev, 83(2), 526–542. doi:10.1111/j.1467-8624.2011.01727.x [PubMed: 22304431]

Cheney DL, & Seyfarth RM (2005). Constraints and preadaptations in the earliest stages of language evolution. The Linguistic Review, 22(2–4), 135–159. 10.1515/tlir.2005.22.2-4.135

Clay Z, & Zuberbühler K (2009). Food-associated calling sequences in bonobos. Animal Behaviour, 77, 1387–1396. 10.1016/j.anbehav.2009.02.016

Corballis MC (2010). The gestural origins of language. Wiley Interdisciplinary Reviews: Cognitive Science, 1(1), 2–7. 10.1002/wcs.2 [PubMed: 26272832]

Delgado RE, Buder EH, & Oller DK (2010). AACT (Action Analysis Coding and Training). Intelligent Hearing Systems, Miami, FL.

Donnellan E, Bannard C, McGillion ML, Slocombe KE, & Matthews D (2020). Infants’ intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language. Developmental Science, 23(1), e12843. 10.1111/desc.12843 [PubMed: 31045301]

Fogel A, & Hannan TE (1985). Manual actions of nine-to fifteen-week-old human infants during face-to-face interaction with their mothers. Child Development, 1271–1279. [PubMed: 4053742]

Franco F, Perucchini P, & March B (2009). Is infant initiation of joint attention by pointing affected by type of interaction? Social Development, 18(1), 51–76. 10.1111/j.1467-9507.2008.00464.x

Gardner RA, & Gardner BT (1969). Teaching sign language to a chimpanzee. Science, 165(3894), 664–672. 10.1126/science.165.3894.664 [PubMed: 5793972]

Gillespie-Lynch K, Greenfield PM, Feng Y, Savage-Rumbaugh S, & Lyn H (2013). A cross-species study of gesture and its role in symbolic development: Implications for the gestural theory of language evolution. Frontiers in Psychology, 4, 160. 10.3389/fpsyg.2013.00160 [PubMed: 23750140]

Gros-Louis J, & Wu Z (2012). Twelve-month-olds’ vocal production during pointing in naturalistic interactions: Sensitivity to parents’ attention and responses. Infant Behavior and Development, 35(4), 773. 10.1016/j.infbeh.2012.07.016 [PubMed: 22982278]

Hewes GW (1973). Primate communication and the gestural origin of language. Current Anthropology, 14(1–2), 5–24. 10.1086/201401

Iverson JM, Tencer HL, Lany J, & Goldin-Meadow S (2000). The relation between gesture and speech in congenitally blind and sighted language-learners. Journal of Nonverbal Behavior, 24(2), 105–130.

Iverson JM, & Goldin-Meadow S (2005). Gesture paves the way for language development. Psychological Science, 16(5), 367–371. 10.1111/j.0956-7976.2005.01542.x [PubMed: 15869695]

Iverson JM, & Wozniak RH (2016). Transitions to intentional and symbolic communication in typical development and in autism spectrum disorder. In Keen D, Meadan H, Brady NC, & Halle JW (Eds.), Prelinguistic and minimally verbal communicators on the autism spectrum (p. 51–72). Springer Science + Business Media. 10.1007/978-981-10-0713-2_4

Iyer SN, & Oller DK (2008). Prelinguistic vocal development in infants with typical hearing and infants with severe-to-profound hearing loss. The Volta Review, 108(2), 115. [PubMed: 21499444]

Iyer SN, Denson H, Lazar N, & Oller DK (2016). Volubility of the human infant: Effects of parental interaction (or lack of it). Clinical Linguistics & Phonetics, 30(6), 470–488. 10.3109/02699206.2016.1147082 [PubMed: 27002533]

Jhang Y, & Oller DK (2017). Emergence of functional flexibility in infant vocalizations of the first 3 months. Frontiers in Psychology, 8, 300. 10.3389/fpsyg.2017.00300 [PubMed: 28392770]

Jhang Y, Franklin B, Ramsdell-Hudock HL, Oller DK (2017). Differing roles of the face and voice in early human communication: Roots of language in multimodal expression. Frontiers in Communication, 2(10), 1–12. 10.3389/fcomm.2017.00010

Kendon A (2017). Reflections on the “gesture-first” hypothesis of language origins. Psychonomic Bulletin & Review, 24(1), 163–170. 10.3758/s13423-016-1117-3 [PubMed: 27439503]

Koopsman-van Beinum FJ, & Van der Stelt JM (1986). Early stages in the development of Speech movements. In Lindblom B, & Zetterstrom R (Eds.), Precursors of early speech (p. 37–50). Wenner-Gren Center International Symposium Series.

Lameira AR (2017). Bidding evidence for primate vocal learning and the cultural substrates for speech evolution. Neuroscience and Biobehavioral Reviews, 83, 429–439. 10.1016/j.neubiorev.2017.09.021 [PubMed: 28947156]

Lynch MP, Oller DK, Steffens ML, Levine SL, Basinger DL, Umbel V (1995). Onset of speech-like vocalizations in infants with down syndrome. American Journal on Mental Retardation, 100(1), 68–86. [PubMed: 7546639]

Liszkowski U, Brown P, Callaghan T, Takada A, & De Vos C (2012). A prelinguistic gestural universal of human communication. Cognitive Science, 36(4), 698–713. 10.1111/j.1551-6709.2011.01228.x. [PubMed: 22303868]

Locke JL (2006). Parental selection of vocal behavior: Crying, cooing, babbling, and the evolution of language. Human Nature, 17, 155–168. [PubMed: 26181412]

Long HL, Bowman DD, Yoo H, Burkhardt-Reed MM, Bene ER, & Oller DK (2020). Social and endogenous infant vocalizations. PloS One, 15(8), e0224956. 10.1371/journal.pone.0224956 [PubMed: 32756591]

McGillion M, Herbert JS, Pine J, Vihman M, DePaolis R, Keren-Portnoy T, & Matthews D (2017). What paves the way to conventional language? The predictive value of babble, pointing, and socioeconomic status. Child development, 88(1), 156–166. 10.1111/cdev.12671 [PubMed: 27859008]

Milenkovic P (2010). TF32 [Computer Program]. Department of Electrical and Computer Engineering, University of Wisconsin, Madison.

Müller GB, & Newman SA (2003). Origination of organismal form: Beyond the gene in developmental and evolutionary biology. MIT Press.

Nathani S, Ertmer D, & Stark R (2006). Assessing vocal development in infants and toddlers. Clinical Linguistics and Phonetics, 20(5), 351–369. 10.1080/02699200500211451 [PubMed: 16728333]

Newman SA (2016). Origination, variation, and conservation of animal body plan development. Cell Biology and Molecular Medicine Reviews, 2, 130–162. 10.1002/3527600906.mcb.200400164.pub2

Oller DK (2000). The emergence of the capacity for speech. Lawrence Erlbaum Associates.

Oller DK, Buder EH, Ramsdell HL, Warlaumont AS, Chorna LB, & Bakeman R (2013). Functional flexibility of infant vocalization and the emergence of language. Proceedings of the National Academy of Sciences, 110(16),6318–6323. 10.1073/pnas.1300337110

Oller DK, Griebel U, & Warlaumont AS (2016). Vocal development as a guide to modeling the evolution of language Topics in Cognitive Science (topiCS), Special Issue: New Frontiers in Language Evolution and Development, Editor, Wayne D. Gray, Special Issue Editors, D. Kimbrough Oller, Rick Dale, and Ulrike Griebel, 8(2), 382–392.

Oller DK, Griebel U, Iyer SN, Jhang Y, Warlaumont AS, Dale R, & Call J (2019). Language origins viewed in spontaneous and interactive vocal rates of human and bonobo infants. Frontiers in Psychology, 10, 729. 10.3389/fpsyg.2019.00729 [PubMed: 31001176]

Oller DK, Caskey M, Yoo H, Bene ER, Jhang Y, Lee CC, Bowman DD, Long HL, Buder EH, & Vohr B (2019). Preterm and full term infant vocalization and the origin of language. Scientific Reports, 9(1), 1–10. 10.1038/s41598-019-51352-0 [PubMed: 30626917]

Orr E (2018). Beyond the pre-communicative medium: A cross-behavioral prospective study on the role of gesture in language and play development. Infant Behavior and Development, 52, 66–75. 10.1016/j.infbeh.2018.05.007 [PubMed: 29864605]

Pika S, Liebal K, Call J, & Tomasello M (2005). Gestural communication of apes. Gesture, 5(1–2), 41–56. 10.1075/gest.5.1.05pik

Pollick AS, & De Waal FB (2007). Ape gestures and language evolution. Proceedings of the National Academy of Sciences, 104(19), 8184–8189. 10.1073/pnas.0702624104

Riede T, Bronson E, Hatzikirou H, Zuberbühler K (2005) Vocal production mechanisms in a non-human primate: morphological data and a model. Journal of Human Evolution, 48, 85–96. 10.1016/j.jhevol.2004.10.002 [PubMed: 15656937]

Rivas E (2005). Recent use of signs by chimpanzees (pan troglodytes) in interactions with humans. Journal of Comparative Psychology, 119(4), 404. 10.1037/0735-7036.119.4.404 [PubMed: 16366774]

Silva Lima ED, & Cruz-Santos A (2012). Acquisition of gestures in prelinguistic communication: A theoretical approach. Revista da Sociedade Brasileira de Fonoaudiologia, 17(4). 10.1590/S1516-80342012000400022

Sterelny K (2012). Language, gesture, skill: the co-evolutionary foundations of language. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1599), 2141–2151. 10.1098/rstb.2012.0116

Stark RE (1980). Stages of speech development in the first year of life. In Yeni-Komshian G, Kavanaugh J, & Ferguson C (Eds.), Child Phonology (p. 73–90). Academic Press.

Thal DJ, & Tobias S (1992). Communicative gestures in children with delayed onset of oral expressive vocabulary. Journal of Speech and Hearing Research, 35, 1281–1289. [PubMed: 1494275]

Tomasello M, & Zuberbühler K (2002). Primate vocal and gestural communication. In Bekoff M, Allen C, & Burghardt GM (Eds.), The cognitive animal: Empirical and theoretical perspectives on animal cognition (p. 293–299). MIT Press.

Tomasello M, Carpenter M, & Liszkowski U (2007). A new look at infant pointing. Child Development, 78(3), 705–722. 10.1111/j.1467-8624.2007.01025.x [PubMed: 17516997]

Tomasello M (2010). Origins of human communication. MIT press.

Volterra V, Caselli MC, Capirci O, & Pizzuto E (2005). Gesture and the emergence and development of language. In Tomasello M & Slobin DI (Eds.), Beyond nature-nurture: Essays in honor of Elizabeth Bates (p. 3–40). Lawrence Erlbaum Associates.

West-Eberhard MJ (2003). Developmental plasticity and evolution. New York: Oxford University Press.

Wu Z, & Gros-Louis J (2015). Caregivers provide more labeling responses to infants' pointing than to infants' object-directed vocalizations. Journal of Child Language, 42(3), 538–561. 10.1017/S030500091400022 [PubMed: 24923871]

Highlights

- Compared rates of gesture and vocalization across the first year.
- Gesture and vocalization were observed at 4, 7, and 11 months in a recording playroom where parents were asked to interact naturally with their infants.
- The research suggests vocalization is overwhelming predominant over gesture in communication of the first year, especially early in the first year.

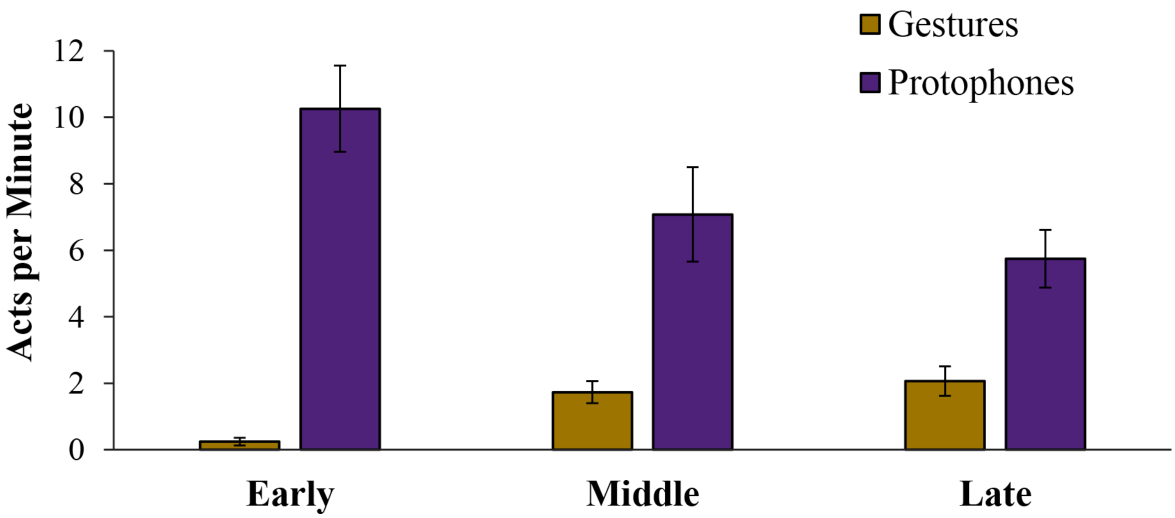

Figure 1.

Gestures and Protophones per Minute across the First Year

Note. This figure shows protophones and gestures per minute across the three ages in our sample (error bars show standard errors).

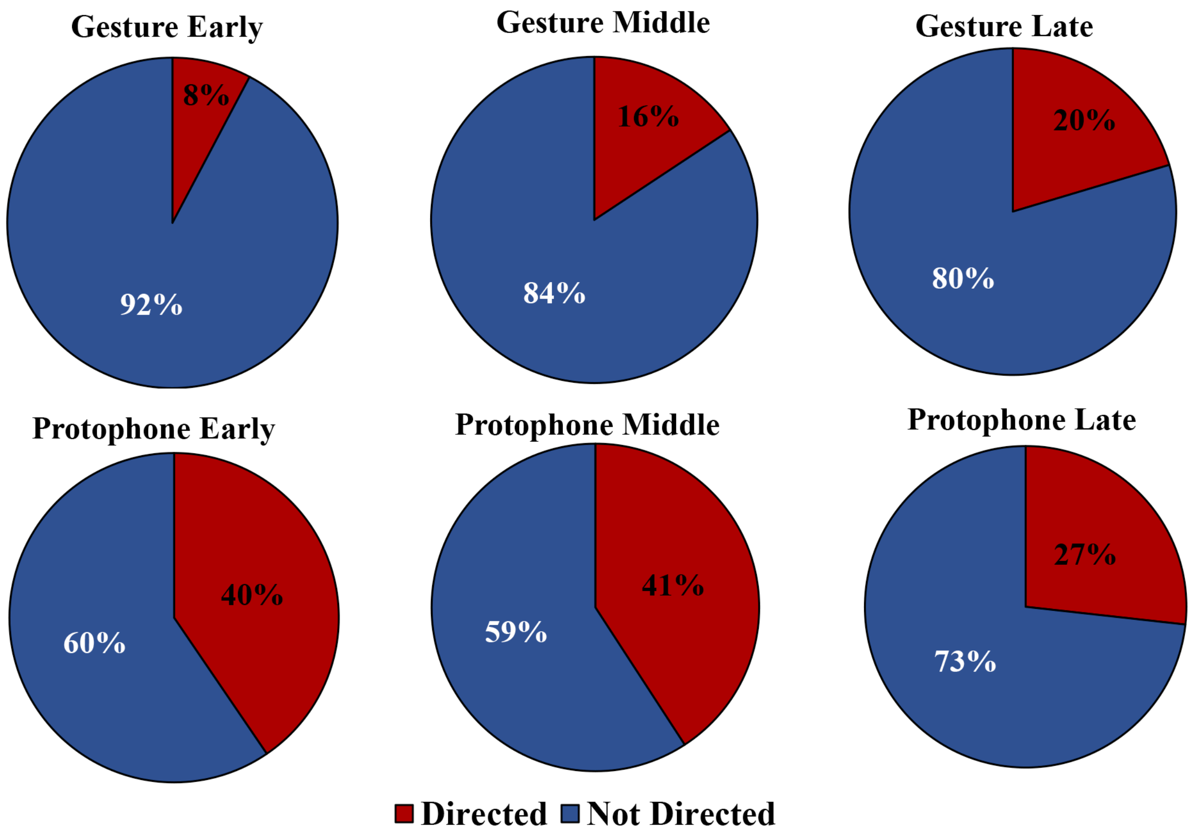

Figure 2.

Proportion of Directed Protophones and Gestures

Note. This figure shows proportions of protophones and gestures that were socially directed as determined by gaze direction at three ages.

Table 1

Infant Demographics and Ages at Recordings

Infant	Gender	Race	Age at Recordings (months; weeks)		
			Early	Middle	Late
1	F	White	3;0	5;0	10;1
2	F	White	4;2	6;0	11;2
3	M	White	5;1	7;2	11;1
4	F	Black	3;1	7;1	12;0
5	M	White	3;3	6;3	10;1
6	F	White	3;2	7;1	10;2
7	M	White	3;3	7;1	11;3
8	F	White	3;3	7;0	12;2
9	M	White	3;3	7;1	11;3
10	M	White	3;3	7;0	11;3
Average age in months;weeks *M* (*SD*)			4;0 (0;3)	7;1 (0;3)	11;2 (0;3)

Note. This table displays demographics and recording ages in months and weeks for each infant at each session.

Table 2

Gestural Action Categories Used in the Present Study

Global Category	Gestural Action	Definition
Non-Social	Hand Shake	Shaking or flapping of hands with no explicit social intent
	Hand Position	Posturing of hand intentionally, such as creating a “d” or an “f” hand (as in American Sign Language)
	Body Rock	Rocking body back-and-forth or bouncing in an upward and downward motion
	Foot Shake	Shaking or flapping of feet with no explicit social intent
Universal Social	Reach	Extension of arm away from body, stretching toward an object, person, or surface
	Point	Extension of index finger while remaining fingers flex into the palm (“d” hand)
	Arm up	Extension of arms and hands in an upward motion toward another person
	Throw	Throwing object to someone
	Push	Pushing object to someone
	Block	Extension of hand or arm to prevent any contact with one’s person by another individual or object
	Touch	Extension of hand and/or arm to gently make contact with another person
Conventional	Clap	Striking palms together repeatedly
	Cover Face	Hand(s) or object (e.g., a blanket or cloth) placed over face to obstruct another’s view of face
	Hand Wave	Movement of hand(s) or entire arm back-and-forth with palms facing away from body
	Head Shake	Shaking head from side to side in a continuous motion

Note. The list is intended to include actions that could conceivably be interpreted as communicative gestures. The list is not complete, but includes all the gestural types actually observed. The terms are drawn partly from literature on ape and human infant gesture.

Table 3

Gestural Function Categories

Function Category	Gestural Function	Definition
Non-communicative	Non-social act	Any non-utilitarian act, not conveying communicative intent (e.g., rhythmic hand banging, body rocking)
Communicative	Request object	Show desire to obtain something
	Accept	Receive something offered
	Social playful	Engage in interactive game, usually involving an object (e.g., playing catch)
	Offer	Present something to someone to accept/reject
	Request up	Show desire to be held or picked up
	Designate	Indicate person or object of interest
	Exult	Show happiness or excitement, celebrate
	Bye-bye	Wave “bye-bye”
	Refuse	Indicate unwillingness to do something
	Show	Make something visible to be perceived by another
	Seek attention	Show desire to engage with another
	Assent	Express approval or agreement

Note. All Non-social actions from Table 2 Were treated as expressing Non-communicative functions. Universal Social and Conventional actions were categorized as expressing one of the Communicative functions. Depending on the apparent intent of the infant, Reach could be categorized as expressing Request Object, Offer, or Accept. As in Table 2, the Communicative function list is not complete, but includes all the gestural functions actually observed.

Table 4

Distribution of Gesture Types

Gesture Type	**Infant Age Group**			
	Early	**Middle**	**Late**	**Total**
Non-Social	25	164	142	**331**
Universal Social	21	149	206	**376**
Conventional	0	16	26	**42**
Total	**46**	**329**	**374**	**752** *

Note. This table shows the distribution of three global gesture categories.

*Total gestures across the ~600 minutes of recording.