SCIENTIFIC REPORTS

natureresearch

OPEN

Preterm and full term infant vocalization and the origin of language

D. Kimbrough Oller[1,2,3], Melinda Caskey[4], Hyunjoo Yoo[5], Edina R. Bene[1], Yuna Jhang[6], Chia-Cheng Lee[7], Dale D. Bowman[1,2], Helen L. Long[1], Eugene H. Buder[1,2] & Betty Vohr[8]

How did vocal language originate? Before trying to determine how referential vocabulary or syntax may have arisen, it is critical to explain how ancient hominins began to produce vocalization flexibly, without binding to emotions or functions. A crucial factor in the vocal communicative split of hominins from the ape background may thus have been copious, functionally flexible vocalization, starting in infancy and continuing throughout life, long before there were more advanced linguistic features such as referential vocabulary. 2–3 month-old modern human infants produce "protophones", including at least three types of functionally flexible non-cry precursors to speech rarely reported in other ape infants. But how early in life do protophones actually appear? We report that the most common protophone types emerge abundantly as early as vocalization can be observed in infancy, in preterm infants still in neonatal intensive care. Contrary to the expectation that cries are the predominant vocalizations of infancy, our all-day recordings showed that protophones occurred far more frequently than cries in both preterm and full-term infants. Protophones were not limited to interactive circumstances, but also occurred at high rates when infants were alone, indicating an endogenous inclination to vocalize exploratorily, perhaps the most fundamental capacity underlying vocal language.

In the origin of language, it appears increasingly probable that a crucial factor was copious, functionally flexible vocalization, starting in early infancy[1,2]. Without such vocal raw material, it seems vocal interaction leading toward speech and language could never have gotten off the ground. It has long been known that very young human infants produce "protophones", including at least three types of functionally flexible non-cry precursors to speech[3,4], reported to occur rarely in other ape infants[5].

But how early in life do protophones actually appear? We report here on research with both preterm infants still in neonatal intensive care and full-term infants monitored at home from the first month after birth. Longitudinal all-day audio recordings were reviewed by trained human listeners for both groups of infants. The goal was to determine the extent to which protophones occurred in preterm infants along with cries and to provide a comparison of relative cry and protophone rates in both groups. It has often been thought that cries are the predominant vocalizations of early infancy and that protophones emerge from cries[6–8]. However, recent research has shown that protophones occur from the first days of life in full term infants and far more frequently than cries from at least 3 months of age[2,3,9]. The present work provides new perspectives on very early protophone and cry rates and even addresses the possibility that preterm infants, shortly after extubation, when they can first breathe independently, may also produce protophones and thus manifest a flexible vocalization capability required as a foundation for vocal language. Our method also allows evaluation of the extent to which protophones occur in both interactive and non-interactive circumstances and especially when infants are alone. That infants vocalize when they are not interacting with caregivers[10] suggests an endogenous inclination to vocalize exploratorily, perhaps the most fundamental capacity underlying vocal language.

[1]University of Memphis, Memphis, Tennessee, USA. [2]Institute for Intelligent Systems, University of Memphis, Memphis, Tennessee, USA. [3]Konrad Lorenz Institute for Evolution and Cognition Research, Klosterneuburg, Austria. [4]Kaiser Permanente, Oregon, USA. [5]University of Alabama, Tuscaloosa, Alabama, USA. [6]Chung Shan Medical University, Taichung, Taiwan. [7]Portland State University, Portland, Oregon, USA. [8]Alpert Medical School of Brown University, Women and Infants Hospital, Providence, RI, USA. Correspondence and requests for materials should be addressed to D.K.O. (email: koller@memphis.edu)

Background on the Origins of Language

Roots of human speech and language are being sought in a variety of domains. Cross-species comparative evaluations have illustrated that many species have vocal capabilities that, while far less elaborate than those of humans, do display notable flexibility[11–14]. Apes have been shown to have important gestural communication abilities[15,16], even in the first years of life[17,18], and learning of considerable language-like behavior has been shown to be possible in animals trained by humans from an early age[19,20]. Computational modeling and robotics research has illustrated that aspects of language learning and evolution can be mutually informative and has encouraged the speculation that language evolution is driven by endogenous factors in infants, especially by curiosity[21–25].

Human development research is central in the search for roots of language[26,27], and our group has argued that longitudinal observation of human infants[2,28] is especially important in providing evidence relevant to possible early changes in the hominin line that formed foundations for the evolution of language. This reasoning is based on the fact that human infants begin to talk only after crucial earlier steps of development that have been reasoned to have also been steps that ancient hominins must have taken on the path toward language. For example, active vocal interchange is widely recognized as foundational for vocal language, developing by the first three months of human life[29–32]; vocal interchange is also actively being evaluated in non-human primates[7,33], although the results suggest considerably more restricted interactivity than in humans. Canonical babbling, developing usually by 7 months[34], is recognized as a critical step in human vocal development since words in natural languages are overwhelmingly built from canonical syllables (e.g., baba and mama), and command over the use of canonical syllables has been proposed to have represented a critical step in human evolution, predating language[35–37]. Production of canonical syllables has never been reported to occur in non-human primates. Joint attention, often supported by vocalization and developing by 9–12 months[38,39] is recognized as a critical foundation for vocabulary learning[40,41] and is further seen as having been a necessary step in hominin evolution prior to the appearance of referential vocabulary[42,43]. Joint attention has been documented for apes trained by humans[44] but has never been reported for non-human primates in the wild. Vocal imitation of well-formed speech-like sounds by very young infants is quite rare[45], but becomes more common by the end of the first year[10]. Failures to demonstrate vocal imitation in non-human primates[46,47], along with the obvious importance for word learning of the ability to imitate, has led to extensive speculations that vocal imitation capability must have been a key factor in human language evolution[36,47,48].

All these developments (vocal interchange, canonical babbling, joint attention, and vocal imitation) correspond to stages in development of speech and ultimately vocal language in human infants[49]. And yet systematic protophone production precedes all of them, apparently beginning within the first days of life. We are pursuing research on the earliest protophones because the ability to produce vocalization flexibly appears to form a foundation without which none of the later stages considered above would be possible, a point argued extensively in prior publications from our laboratory[28,36,37,50].

Protophones in the second half year, including canonical syllables, have long been recognized as precursors to speech[51–54]. But earlier protophones such as the primary vocal types of early infancy (vocants, squeals, and growls) that do not show well-formed consonant-vowel-like form, have largely been ignored in the discussion of language origins[55,56], a discussion that has also failed to exploit the implications of findings indicating coordination of parent and infant vocalization in the first months of life[57,58].

Granted, non-cry speech-precursor vocalizations have been reported in newborns[9], and automated analysis of all-day recordings has suggested protophones may even occur in preterm infants in neonatal intensive care[59]. However, a closer look at vocalizations of preterm infants before their due dates is in order since the automated analysis that was used to assess vocal behavior of the preterm infants in the prior work of Caskey *et al.* was only modeled for older infants and suggested, at best, vastly lower rates of protophones than in full-term infants. Furthermore, the prior work was unable to assess occurrence in preterms in neonatal intensive care of the three primary types of protophones of the first month after full term birth[4], and could not reliably compare protophone and cry rates.

It has been claimed that "cry is the primary means of communication for very young infants[60] (p. 265)". Moreover it has been thought that human vocalization *begins* with cry, and that speech-like vocalization emerges from the cry root[8], with all newborn infant vocalizations being treated as some form of cry[6,61]. This expectation appears to be founded on the idea that humans begin with pan-primate vocal capabilities that only diverge from the primate pattern months after birth. Indeed, evidence suggests the "phee" call in the common marmoset may indeed develop *from* its cry[7]. Thus, it might be expected that prematurely-born human infants produce *only* cries before their due dates.

Goals and rationale. Our intention is to evaluate the extent to which protophones, the earliest vocal precursors to speech, occur even in infants who are born prematurely and are still in neonatal intensive care, and in full-term infants starting in the first month after birth and continuing through the first year. Comparison of rates of protophone production with rates of cry can help put in perspective the origin of protophones, which have been thought by many to be based on cry. Yet protophones do not usually express distress (although they sometimes do[2]), and they are often produced with no obvious intention for anyone to hear them. Speculations about why they occur at all can best be made in the context of quantitative information about the extent and contexts of protophone usage. The comparison of protophone and cry rates may be expected to yield surprises, because prior research with infants beyond two months suggests protophones occur considerably more frequently than cries[2,3].

Cries are often very salient when they occur, being usually both long and loud. This saliency may account for the traditional impression that cries are the predominant sounds of early infancy. Yet quantitative comparison of rates of occurrence is necessary in order to place the protophones in proper perspective and to form a better foundation for speculations about the importance of protophones in the origin of language.
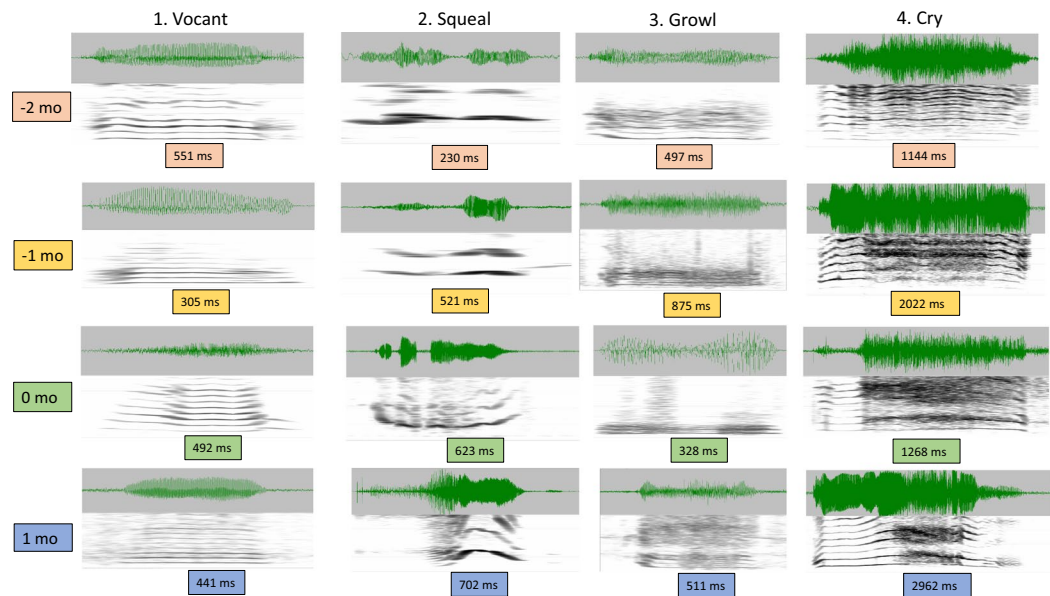
Figure 1. Spectrographic examples of vocalizations in preterm and full-term humans. Protophones (vocants, squeals, and growls) are displayed in the first three columns and cries in the fourth. Nearly all infants (all preterms and 11 of 12 full terms), produced all three protophone types (>40,000 observed). 18 of 20 preterms produced cries, whereas cries were represented for all the full term infants (>6,000 observed). A time domain display is provided at the top of each of the 16 examples, and spectrographic displays are provided at the bottom of each (range 4–5 kHz, 30 Hz bandwidth). The examples were produced by four different infants at each age in each of the four columns. Human infant sounds are extremely variable; every utterance of any of these infant vocal categories is distinct, yet the four categories of sounds represented in the figure's columns correspond to auditorily identifiable exemplars of the categories as labeled. Wave files for each of the examples are available in Supplementary Information. (1) Vocants (or vowel-like sounds): By far the most frequent protophone category (>70% of all protophones at all ages), vocants show little interharmonic noise throughout all four exemplars. Vocants are usually produced in comfort. (2) Squeals: Produced with very high pitch, indicated by periods of widely-spaced harmonics; squeals represented ~7% of protophones, occurring in all but 3 infants. (3) Growls: these exemplars include considerable interharmonic noise and pitch lower than squeals. Some growls also include periods of very low pitch in a phonatory pattern called vocal fry (or pulse). Growls represented ~15% of protophones, and occurred in all the infants. (4) Cry: All four exemplars show human-typical, evenly-spaced harmonics (normal phonation) at the beginning of each of the cries, followed by significant and auditorily salient periods of dysphonation (sometimes called harsh phonation) in each of the four, corresponding to chaotic and/or subharmonic phonatory regimes. The pattern of normal phonation followed by dysphonation and then normal phonation again has been argued to constitute the most prototypical form of intense human infant cry[83].

In order to pursue this research, 40 all-day recordings were made using the LENA device[62,63] from 20 preterm infants in neonatal intensive care 8 and 4 weeks prior to due date (-2 and -1 months of age). In addition we obtained all-day recordings from 9 full-term infants at 0 months and from 12 (including the 9) full-term infants at 1, 3, 6, 9, and 12 months, yielding 69 all-day recordings of full-terms. We randomly-selected 24 five-minute segments from each recording and conducted human coding to estimate cry and protophone volubility. A questionnaire completed by coders at the end of each five-minute segment indicated the extent to which there was infant-directed speech (IDS), adult-directed speech (ADS), and the extent to which the infant was alone or asleep. For preterms and full terms together, we thus examined human coding of >2600 five-minute segments, yielding >13,000 minutes of vocalization data.

Results

Contradicting the expectation that speech precursors emerge from cry, we found plentiful protophones even in awake −2 month-olds. Segments coded from the preterms included >11,000 protophones. The human coders indicated whether each protophone was a squeal, a vocant, or a growl (see Methods and Supplementary Information for details). Even for the youngest preterm infants (−2 mo), this coding yielded hundreds of exemplars in each of the three categories. Figure 1 provides spectrographic examples (wave files in Supplementary Information) of unambiguous cases for each category, selected to illustrate that preterm infants produced all three protophone types as well as cries that resembled the same kinds of sounds in full-term infants.

Figure 2 illustrates that, contrary to the common expectation, protophones in the randomly-selected segments occurred far *more* frequently than cries at both preterm ages ($p < 0.0001$ by t-test). For the full-term infants, the protophones also outnumbered cries dramatically, again at every age ($p < 0.0001$). Even the lowest rate found, 1.4 protophones/min for - 2 month-olds, corresponded to >80 protophones per waking hour. For both preterms and full terms, protophones outnumbered cries by a factor >5. Using Generalized Estimating Equations[64] (see
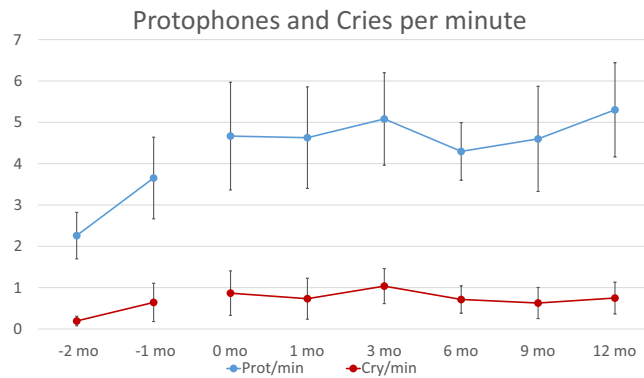
Figure 2. Vocalization rates in preterm and full-term humans. Protophones are infant sounds produced mostly in exploration or play rather than in social interaction. The figure illustrates that protophones/minute generally exceeded cries/minute by a factor of >5 (error bars represent 95% confidence intervals, computed based on the raw data from each infant and independent of the modeled data from the GEE analyses) in randomly-selected five-minute infant-awake segments from all-day audio recordings. Even in the preterm infants, protophones were plentiful, more than two per minute even at two months before due date, suggesting deep natural selection for the tendency to produce endogenous vocalization in humans. GEE analysis indicated a significant increase in the protophone rate for preterms from - 2 to -1 month of age. The differences between protophone and cry rates were massively significant by t-tests. The groups of infants recorded longitudinally were 20 preterms at -2 and -1 month of age (that is, 32 and 36 weeks gestational age), 9 full term infants at 0 months (that is, during the first month after full term birth) and the same 9 plus 3 additional full-term infants through 12 months. Each infant at each age provided an all-day recording from which 24 five-minute segments were randomly selected for the human coding that yielded the estimated rates of vocalization.

Methods and Supplementary Information for rationale), implemented in R, it was found that the full-term infants showed no significant effects of Age for either protophones or cries, but the preterms showed a significant effect for protophones ($p < 0.0005$) and a near significant effect for cries ($p = 0.07$), reflecting the increase in overall vocalization rate that presumably accompanied increasing respiratory sufficiency and maturational changes across age in the preterms.

Figure 3 provides illustrations of the copious occurrence of protophones regardless of circumstances. In these analyses, five-min segments were assigned to a high and a low condition in each case, splitting the data into two groupings with similar numbers of segments. Figure 3A shows that there were high protophone rates when awake infants were in a room with other persons but also when infants were alone, with an average of >3 protophones/minute in the alone circumstance (full terms: 220 alone segments, >4300 protophones; preterms 244 alone segments, >3100 protophones). The great majority of the protophones produced when infants were alone (full terms: 74%; preterms, 78%) came from five-minute segments with little or no crying (<5 cries/segment). GEE showed that full terms produced significantly fewer protophones when not alone ($p < 0.05$), a difference that was particularly obvious at younger ages. However, caregivers may tend to approach when infants are vocalizing, and thus caregivers may have driven the effect suggesting more vocalization when infants were not alone.

There has been considerable emphasis in speculation about the origin of language on parent-infant vocal interaction starting at 2–3 months[65]. Figure 3B shows that infants produced large numbers of protophones either with or without IDS. Still, full-term infants vocalized at nearly twice the rate during segments with infant-directed speech (IDS) than without it (GEE, $p < 0.0001$). Segments for preterm infants did not show significantly higher volubility with IDS, but parents visited their infants relatively infrequently in the hospital, and nurses presumably produced less IDS than their parents did[66]. It cannot be concluded with certainty that higher IDS *caused* higher infant volubility in young full term infants since the effect could have been influenced by parents' choosing times to speak to infants when the infants were already engaged in a period of vocalizing.

The rate of IDS in the study was notable, suggesting a parental tendency to engage infants conversationally very early in life, even at home, a tendency that has been documented in recent research[67]. For full-term infants, caregivers used IDS in >65% of segments with infants awake. Even in the hospital, preterm infants heard IDS from caregivers and hospital staff in >33% of awake segments. While there appear to exist notable cultural differences in amount of IDS used by humans[68], no quantitative study has indicated total absence of IDS in any culture. On the other hand our recent work found no caregiver vocalizations at all directed to three bonobo infants observed across the first year[5].

Not just IDS, but also "overheard" speech may provide important input[69], perhaps especially in cultures where parents speak to infants little in the first year. Infants produced high rates of protophones with and without ADS. Figure 3C illustrates that adult-directed speech (ADS) produced in the vicinity of full-term infants did correspond to higher protophone rates (GEE, $p < 0.05$), but the effect size was <1/3 as high as for IDS.

A final GEE analysis for full-terms was fit to the protophones with predictors Age, Alone, IDS, and ADS, where Alone, IDS, and ADS were treated as parameters with values from 1–5, corresponding to the questionnaire responses. A statistically significant effect was found for IDS ($p < 0.0001$); no interaction terms were significant. GEE analysis for the preterms yielded no significant main effects for protophones or cry on Alone, IDS, or ADS, but a single significant interaction of Age and ADS ($p < 0.05$) for protophones.
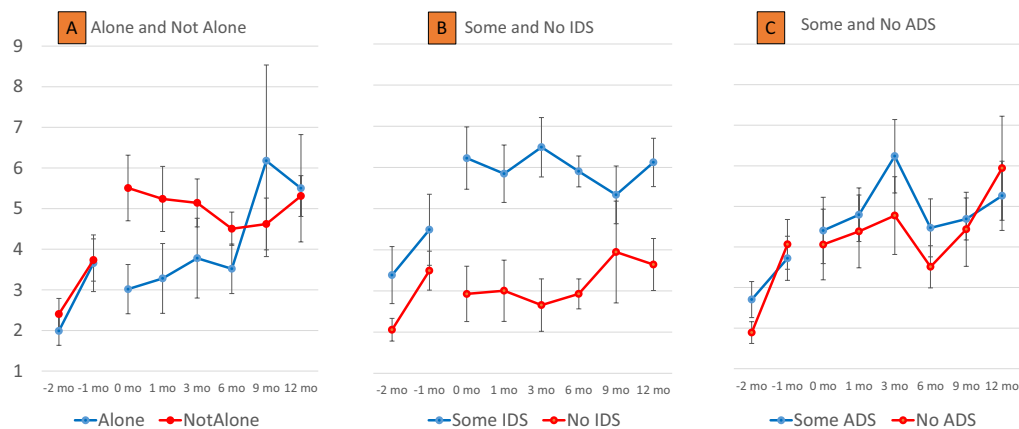
Figure 3. Protophone rates in preterm and full-term humans in differing circumstances. (**A**) Even when infants were alone (no one else was deemed present in the room by the coders), they produced large numbers of protophones (≥2/minute even in the -2 month old preterms), a strong indication of the endogenous nature of infant vocal exploration. The full-term infants, especially at younger ages, produced significantly more protophones when they were not alone, but adult speakers may have driven this effect by approaching very young full terms when they were producing protophones. Error bars are standard errors of the mean, computed based on the raw data from each infant and independent of the modeled data from the GEE analyses. (**B**) GEE analysis showed that full-term infants produced significantly more protophones when there was infant-directed speech (IDS), but once again parents may have driven the effect especially at the youngest ages, by engaging infants vocally when they heard the infants begin to vocalize. (**C**) Higher rates of adult-directed speech (ADS) were reliably related to higher protophone rates for the full-term infants, but the effect was not observed for all ages, and the effect size was much smaller than in the case of IDS.

Discussion

The roots of language appear to run deep in human infancy according to our data since as soon as infants were capable of vocalizing, even when born prematurely by more than two months, as soon as they could breathe independently, they began producing substantial numbers of protophones, the early precursors to speech. Full-term infants from the first month also produced protophones abundantly, and we observed no social circumstances under which protophone production was lower than 2.5 per minute for full-term infants. The *typical* rate for full-term infants was 4–5 per minute and not much lower for preterm infants at -1 month. Even at -2 months preterm infants produced more than 6 times more protophones/min than protophone-like sounds reported for three bonobo infants in the first year[5].

We have reasoned that without the ability and inclination to produce such sounds, there would be only a much reduced basis for vocal interaction with infants, since parents tend to engage infants by responding to protophones and seeking to elicit them[36]. Considerable prior work has proposed that vocal interaction is fundamental in launching other developments required for language[58,70,71].

It seems necessary to ask how the tendency of the human infant to produce protophones is sustained and to posit selection pressures that could have produced the ability and inclination to produce abundant sounds that express no obvious immediate needs. The proposal we have advocated in parallel with work of Locke is that such flexible, playful vocalization has long served to advertise fitness of the altricial (i.e., born helpless) human infant, who has a long path ahead of need for parental care[28,31,72–74]. As the reasoning goes, both modern and ancient hominins were altricial, presumably because bipedality had narrowed the human pelvis and thus required the fetal brain case to be smaller at birth in order to negotiate the passage—hence hominin infants *had* to be born altricial. As hominin evolution progressed, the mature head became progressively larger, requiring continuing limitations on the size of the infant head at birth and progressively longer periods of infancy and childhood. As hominin infants required longer periods of parental care, the selection pressure to advertise fitness also increased.

According to the reasoning, helpless infants can improve their prospects for survival and reproduction by convincing caregivers to invest extensively in them. Infant vocalization produced freely and comfortably can signal well-being, advertising the likelihood that investment in that particular infant is worth the effort. That vocalization can evolve as a fitness signal is well-documented across thousands of species of songbirds, hummingbirds, parrots, pinnipeds, cetaceans, and bats[20,75–82]. In these species, song-like vocalization advertises fitness to potential mates and to competitors. The selection principle is similar to that proposed for the human infant protophones; vocalization in circumstances that appear to involve little or no distress serve, in all these cases, to display well-being by e.g., advertising a healthy respiratory system, an ability to modulate the phonatory system, sometimes intricately, and by advertising the very fact that the organism is not in distress.

The current results suggest the selection pressure on fitness signaling has been sufficiently general to launch protophone production even when infants are born well ahead of schedule. The results further suggest that the

roots of protophone production run so deep as to call into question the idea that speech-like vocalization is grounded in cry. Human language appears to have required special pressures to afford the production of sounds that are not bound to the expression of emotion, opening a path to a vastly more powerful communication system.

Methods

Approvals. The work described here was approved by the Institutional Review Boards of the University of Memphis, Memphis, TN and Women and Infants Hospital, Providence, RI. All methods were performed in accordance with the relevant guidelines and regulations of the IRBs. Informed consent was signed by parents of all the infants recorded.

Participants and recordings. 12 full-term infants of mid to high-mid SES were recruited in Memphis. All-day recordings of ~12 hours in duration were made throughout the first year. 20 low risk preterm infants without congenital abnormalities, born at 30-weeks gestational age or younger were recruited through Women and Infants Hospital, Brown University. These infants were in neonatal intensive care through 36 weeks gestational age. All the preterm infants had been extubated by 32 weeks and were able at least minimally to vocalize by that time. Seven of the mothers of the preterms were high school graduates and the remainder completed at least partial college. The preterm infants were recorded for 16 hours each, at both 32 and 36 weeks gestational age (-2 and -1 months of age respectively).

The recordings were made with the battery-powered LENA system[62] worn in infant vests by full-term infants and placed in the isolette or open crib near the preterm infants' heads, minimizing mouth-to-microphone distance in both cases. We selected 24 segments of five minutes each for human coding and analysis from each recording, at equal time intervals beginning with a semi-randomly-selected five-minute segment. A questionnaire item to determine if the infant had been asleep was administered to coders at the end of coding for each five-minute segment. Segments where the infant was asleep were excluded from analyses. >46,000 infant utterances were identified by the coders in these randomly-selected samples. Table S1 (Supplementary Information, SI) characterizes the sample, and the text of the SI elaborates on participants and recording procedures (sections S1.1.1-S1.1.3).

Coding categories. Consistent with the goal of the research, the only categories of sounds quantified for the analyses were Protophones and Cries. Coders, however, classified utterances in more detail, using the categories: squeal, vocant, growl, whisper, ingress, wail, whimper, laugh, and other. Vegetative utterances such as burps, hiccoughs and sneezes were not coded. Squeals, vocants, and growls (which accounted for the vast majority of all assigned codes) were collapsed into the single category Protophone for analysis and the two distress types (wail and whimper, which accounted for the vast majority of the remaining codes) were collapsed to Cry for analysis. ~1% of coded utterances pertained to any category other than Protophone or Cry. The SI (sections S1.1.4-S.1.1.5) details the coding system and its rationale.

The distinction between the collapsed categories of Protophone and Cry is based on the fact that wails and whimpers include characteristics in both vocalization and facial affect that mark them as obligatorily negative—wails and whimpers are assumed to have been evolved as inherently negative emotional expressions. Sound alone is sufficient to identify both wails and whimpers reliably as distinct from other infant sounds, in particular from protophones[83], which usually do not express negativity although they can do so on occasion[2,4]. Acoustically, wails consist of intense (loud) nuclei with variable phonatory regimes of substantial duration usually including substantial harsh quality[83]. Typical whimpers are less intense, have shorter nuclei, and are required by definition to include at least one glottal burst (see SI for illustrations). Wails can also include glottal bursts and/or catch breaths[83].

The three primary protophones can be characterized acoustically[2,84] as follows: vocants (often called vowel-like sounds) typically consist of nuclei with normal phonation at variable lengths and typically at moderate amplitudes; squeals include salient periods where pitch is typically double that of vocants, usually in loft (falsetto) register; and growls are usually lower in pitch than vocants, though growls are primarily characterized by a salient period of either harsh phonatory quality (manifest in subharmonic, biphonation, or chaotic vocal regimes) or pulse (vocal fry) register.

While the categories have these acoustic characterizations, coding is based first and foremost on intuitive judgments. In fact, the acoustic characterizations available in our own and related literature describe findings for auditorily-specified categories. The coding system is founded on the assumption that human caregivers have been naturally selected to recognize categories of infant sound because that recognition puts them at an advantage in nurturing their own offspring. Thus distress sounds must be identifiable without training, and by implication, other sounds must be differentiable from distress sounds. Laughter, similarly, must be identifiable and must be auditorily distinct from other sounds.

In accord with our reasoning, parents must be able to identify the protophones as the flexible product of infant vocal exploration, sounds that are not bound to any particular emotion (although of course each protophone type can accompany any emotion on particular occasions). The three principal subcategories of protophones appear to be the self-organized products of infant vocal exploration in the non-random space provided by the infant phonatory apparatus[28]. The explorations presumably yield certain categories preferentially, and infants tend to produce these favored categories repetitively. Following the same reasoning, it makes sense that parents must recognize the categories that infants favor.

The three primary protophone types appear to represent natural kinds[85]. They were reported spontaneously by parents as being "sounds the baby makes" in our first longitudinal research more than 40 years ago[86,87]. In ethologically-oriented coding of recordings from that first study of infants under 6 months of age, the same three types were proposed by coders. Subsequent research has supported the conclusion that indeed the vast majority of early infant protophones can be coded reliably and sensibly as having phonatory characteristics that pertain to one of these three protophones types.

Coding procedure and questionnaire.   Both Cries and Protophones were counted within this coding system as "breath groups"[88], where each voiced period produced on a single egress was counted as one utterance. Coding was conducted in real-time. After coding each five-minute segment, coders responded to a questionnaire to determine for that segment (on a five-point scale) the extent to which 1) there was infant-directed speech, 2) there was other-directed speech, 3) the infant was alone in the room, and 4) the infant was asleep (see S1.1.7).

Coders and training.   Eighteen female master's students in Speech-Language-Pathology served as coders. They were trained extensively as described in S1.1.8 before coding the recordings that were analyzed. Eleven individuals coded preterm recordings and eight coded full-terms (one coder worked on some recordings from both groups). Because coders worked in the project ten hours per week for a period of not less than two years and usually more, they could not be blinded in general to the research interests, and they were fully aware of whether they were coding full-terms or preterms. On the other hand, they were given no information about the particular infants they were assigned to, and their assignments with regard to infant age were ordered randomly. Still it was possible for them to discern much about the infants and families while they were listening to the recorded segments.

The primary task of coding training is not to teach coders to recognize the categories used in the coding system (distinguishing wails and whimpers from protophones is a natural human capacity), so much as to teach them to systematically use the labels the coding system assigns to those natural categories and to teach them to count utterances using the breath-group criterion.

Coder agreement.   As reported in S1.1.9, there were several types of coder agreement studies conducted both within the largely disjunct coder groups (preterm and full-term) and between them. In all cases the agreement as measured by correlations of numbers of Protophones counted exceeded 0.8, and in most cases the same was true of Cry counts. More important, however, was the fact that coder differences on counts as indicated by coefficients of variation (COVs) showed that the massive differences between Protophone and Cry rates illustrated in Fig. 2 were more than 6 times larger than estimated coder differences.

Statistical procedures.   The analyses regarding protophone and cry usage (Fig. 2) were conducted with paired comparisons t-tests for each age independently. Generalized Estimating Equations (GEE)[64] were used for additional analyses. GEE is a modeling approach that can account for fixed and random effects but is preferable to more traditional mixed-models frameworks for semi-longitudinal research, especially when there are correlations among data from participants across conditions, and when the number of observations varies for participants within or across conditions. In essence the approach offers an assessment that estimates on a principled basis the means and standard deviations relevant for the analysis while taking into account intragroup correlations and variations in numbers of observations. The GEE method requires an assumption of a link function between means and a linear predictor (we chose a linear link) and the specification of a covariance structure (we chose an exchangeable covariance structure). Other advantages of GEE over traditional mixed models are that it requires no normality assumption, and it will result in consistent estimates of means even if the correlation structure is misspecified.

The GEE analyses were conducted on the preterm samples separately from the full-term samples, because the groups of participants were disjunct in the two cases, and because of the fundamental differences in circumstances of recording (in the hospital vs at home). The statistical approaches are discussed in more detail in the SI (section S1.1.10).

Data Availability

The authors will supply the spreadsheets from which the data for this article were computed on request and these will also be deposited in the Open Science Framework. These spreadsheet data will allow readers to recompute the values or analyze the data in ways that differ from those reported here. In addition the R code for the GEE analyses will be deposited in the Open Science Framework.

References
1. Locke, J. L. *The child's path to spoken language*. First edn, (Harvard University Press, 1993).
2. Oller, D. K. *et al.* Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences* **110**, 6318–6632, 0.1073/pnas.1300337110 (2013).
3. Iyer, S. N., Ertmer, D. J. & Stark, R. E. Assessing vocal development in infants and toddlers. *Clinical Linguistics & Phonetics* **20**, 351–369 (2006).
4. Jhang, Y. & Oller, D. K. Emergence of Functional Flexibility in Infant Vocalizations of the First 3 Months. *Frontiers in Psychology* **8**, https://doi.org/10.3389/fpsyg.2017.00300 (2017).
5. Oller, D. K. *et al.* Language origin seen in spontaneous and interactive vocal rate of human and bonobo infants. *Frontiers Psychology* **10**, https://doi.org/10.3389/fpsyg.2019.00729 (2019).
6. Wasz-Höckert, O., Lind, J., Vuorenkoski, V., Partanen, T. & Valanne, E. *The Infant Cry: A Spectrographic and Auditory Analysis*. (Heinemann, 1968).
7. Takahashi, D. Y. *et al.* The developmental dynamics of marmoset monkey vocal production. *Science* **349**, 734–738 (2015).
8. Lester, B. M. & Boukydis, C. F. Z. In *Nonverbal vocal communication* (eds Papoušek, H., Jürgens, U. & Papoušek, M.) 145–173 (Cambridge University Press, 1992).
9. Dominguez, S., Devouche, E., Apter, G. & Gratier, M. The Roots of Turn-Taking in the Neonatal Period. *Infant and Child Development*, https://doi.org/10.1002/icd.1976 (2016).
10. Long, H. L., Oller, D. K. & Bowman, D. Reliability of Listener Judgments of Infant Vocal Imitation. *Frontiers in Psychology* **10**, https://doi.org/10.3389/fpsyg.2019.01340 (2019).
11. Crockford, C., Herbinger, I., Vigilant, L. & Boesch, C. Wild chimpanzees produce group-specific calls: a case for vocal learning? *Ethology* **110**, 221–243 (2004).

12. Goldstein, M. H., King, A. P. & West, M. J. Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences* **100**, 8030–8035 (2003).
13. Lipkind, D. *et al*. Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature* **498**, 104–108, https://doi.org/10.1038/nature12173 (2013).
14. Laporte, M. N. C. & Zuberbühler, K. The development of a greeting signal in wild chimpanzees. *Developmental Science* **14**, 1220–1234 (2011).
15. Call, J. In *Evolution of Communicative Flexibility: Complexity*, *Creativity and Adaptability in Human and Animal Communication* (eds Oller, D. K. & Griebel, U.) 235–252 (MIT Press, 2008).
16. Call, J. & Tomasello, M. *The Gestural Communication of Apes and Monkeys*. (Taylor & Francis Group/Lawrence Erlbaum Associates, 2007).
17. Heesen, R. *et al*. Linguistic laws in chimpanzee gestural communication. *Proc. R. Soc. B* **286**, 1–9, https://doi.org/10.1098/rspb.2018.2900 (2019).
18. Kersken, V., Gómez, J.-C., Liszkowski, U., Soldati, A. & Hobaiter, C. A gestural repertoire of 1- to 2-year-old human children: in search of the ape gestures. *Animal Cognition* **22**, 577–595, https://doi.org/10.1007/s10071-018-1213-z (2019).
19. Pepperberg, I. M. Vocal learning in Grey parrots: A brief review of perception, production, and cross-species comparisons. *Brain & Language* **115**, 81–91 (2010).
20. Griebel, U., Pepperberg, I. M. & Oller, D. K. Developmental plasticity and language: A comparative perspective. *Topics in Cognitive Science (topiCS)* **8**, 435–445 (2016).
21. McMurray, B., Aslin, R. N. & Toscano, J. C. Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science* **12**, 369–379 (2009).
22. Moulin-Frier, C., Nguyen, S. M. & Oudeyer, P.-Y. Self-organization of early vocal development in infants and machines: the role of intrinsic motivation. *Frontiers in Psychology* **4**, 1–20, https://doi.org/10.3389/fpsyg.2013.01006 (2014).
23. Breazeal, C. *et al*. How children treat robots as informants. *Topics in Cognitive Science (topiCS)*, *Special Issue: New Frontiers in Language Evolution and Development*, *Editor*, Wayne D. Gray, *Special Issue Editors*, Oller, D. K., Dale, R., and Griebel, U. **8**, 481–491, 10:1111/tops.12192 (2016).
24. Oudeyer, P.-Y. & Smith, L. How evolution may work through curiosity-driven developmental process. *Topics in Cognitive Science (topiCS)*, *Special Issue: New Frontiers in Language Evolution and Development*, *Editor*, *Wayne D. Gray*, *Special Issue Editors*, *D. K.Oller*, *R. Dale, and U. Griebel* **8**, 492–502, 10:1111/tops.12196 (2016).
25. Oudeyer, P.-Y. & Kaplan, F. Language Evolution as a Darwinian Process: Computational Studies. *Cognitive Processing* **8**, 21–35 (2007).
26. Tomasello, M. *Origins of Human Communication*. (MIT Press, 2008).
27. Bergelson, E. & Swingley, D. At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences* **109**, 3253–3258, https://doi.org/10.1073/pnas.1113380109 (2012).
28. Oller, D. K., Griebel, U. & Warlaumont, A. S. 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. K.Oller*, *R. Dale*, *and U. Griebel* **8**, 382–392 (2016).
29. Brazelton, T. B., Tronick, E., Adamson, L., Als, H. & Wise, S. Early mother-infant reciprocity. *Ciba Found Symp.* **1**(33), 137–154 (1975).
30. Feldman, R. Parent–Infant Synchrony: Biological Foundations and Developmental Outcomes. *Current directions in psychological science* **16**, 340–345 (2007).
31. Locke, J. L. Evolutionary developmental linguistics: Naturalization of the faculty of language. *Language Sciences* **31**, 33–59 (2009).
32. Northrup, J. B. & Iverson, J. M. Vocal Coordination During Early Parent–Infant Interactions Predicts Language Outcome in Infant Siblings of Children with Autism Spectrum Disorder. *Infancy*, https://doi.org/10.1111/infa.12090 (2015).
33. Hage, S. R., Jurgens, U. & Ehret, G. Audio–vocal interaction in the pontine brainstem during self-initiated vocalization in the squirrel monkey. *European Journal of Neuroscience* **23**, 3297–3308, https://doi.org/10.1111/j.1460-9568.2006.04835.x (2006).
34. Oller, D. K. In *Child phonology*, Vol *1: Production* (eds Yeni-Komshian, G., Kavanagh, J., & Ferguson, C.) 93–112 (Academic Press, 1980).
35. MacNeilage, P. F. The frame/content theory of evolution of speech production. *Behavioral & Brain Sciences* **21**, 499–546 (1998).
36. Oller, D. K. *The Emergence of the Speech Capacity*. (Lawrence Erlbaum Associates, 2000).
37. Griebel, U. & Oller, D. K. In *Evolutionary Science of Human Behavior:* An Interdisciplinary *Approach* (eds Lafreniere, P. J. & Weisfeld, G.) 257–280 (Linus Learning, 2014).
38. Butterworth, G. Species typical aspects of manual pointing and the emergence of language in human infancy. (Waseda University International Conference Center, Tokyo, 1996).
39. Bakeman, R. & Adamson, L. B. Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development* **55**, 1278–1289 (1984).
40. Mundy, P. A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *European Journal of Neuroscience*, 1-18, https://doi.org/10.1111/ejn.13720 (2017).
41. Tomasello, M. & Farrar, J. Joint attention and early language. *Child Development* **57**, 1454–1463 (1986).
42. Kwisthouta, J., Vogt, P., Haselager, P. & Dijkstra, T. Joint attention and language evolution. *Connection Science: Special Issue on Social Learning in Embodied Agents* **20**, 155–171, https://doi.org/10.1080/09540090802091958 (2008).
43. Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral & Brain Sciences* **28**, 675–735 (2005).
44. Leavens, D. A., Hopkins, W. D. & Bard, K. A. Indexical and referential pointing in Chimpanzees (Pan Troglodytes). *Journal of Comparative Psychology* **110**, 346–353 (1996).
45. Papoušek, M. & Papoušek, H. Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language* **9**, 1989 (1989).
46. Hage, S. R., Gavrilov, N. & Nieder, A. Developmental changes of cognitive vocal control in monkeys. *Journal of Experimental Biology* **219**, 1744–1749, https://doi.org/10.1242/jeb.137653 (2016).
47. Hauser, M. The evolution of the language faculty: The essential role of interfaces. (Plenary address to The XI Conference of the International Association for the Study of Child Language, Edinburgh, UK, 2008).
48. Davila-Ross, M., Allcock, B. & Bard, K. A. Aping expressions? Chimpanzees produce distinct laugh types when responding to laughter of others. *Emotion* **11**, 1113–1120 (2011).
49. Hoff, E. *Language Development* 5edn, (Wadsworth, Cengage Learning, 2009).
50. Griebel, U. & Oller, D. K. In *Evolution of Communicative Flexibility: Complexity*, *Creativity and Adaptability in Human and Animal Communication* (eds Oller, D. K. & Griebel, U.) 9–40 (MIT Press, 2008).
51. Cameron, J., Livson, N. & Bayley, N. Infant vocalizations and their relationship to mature intelligence. *Science* **157**, 331–333 (1967).
52. Leopold, W. F. In *Child language: A book of readings*, *published in 1971.* (eds Aaron Bar-Adon & Werner F. Leopold) (Prentice-Hall, Inc., 1953).
53. Lewis, M. M. *Infant Speech*. (Harcourt Brace, 1936).
54. Stark, R. E. In *Child Phonology*, vol. *1* (eds Yeni-Komshian, G., Kavanagh, J., & Ferguson, C.) 73–90 (Academic Press, 1980).
55. Christiansen, M. H. & Kirby, S. Language evolution: Consensus and controversies. *Trends in Cognitive Sciences* **7**, 300–307 (2003).

56. Pinker, S. & Bloom, P. Natural language and natural selection. *Behavioral & Brain Sciences* **13**, 707–784 (1990).
57. Stern, D. N., Jaffe, J., Beebe, B. & Bennett, S. L. Vocalizing in unison and in alternation: two modes of communication within the mother-infant dyad. *Annals of the New York Academy of Sciences* **263**, 89–100 (1975).
58. Trevarthen, C. Infant Intersubjectivity: Research, Theory, and Clinical Applications. *Journal of Child Psychology and Psychiatry*, 3–48 (2001).
59. Caskey, M., Stephens, B., Tucker, R. & Vohr, B. R. Importance of Parent Talk on the Development of Preterm Infant Vocalizations. *Pediatrics* **128**, 910–916 (2011).
60. Zeifman, D. M. An ethological analysis of human infant crying: Answering Tinbergen's four questions. *Developmental Psychobiology* **39**, 265–285 (2001).
61. Soltis, J. The signal functions of early infant crying. *Behavioral and Brain Sciences* **27**, 443–490 (2004).
62. Gilkerson, J. *et al*. Mapping the early language environment using all-day recordings and automated analysis. *American Journal of Speech-Language Pathology* **26**, 248–265, https://doi.org/10.1044/2016_AJSLP-15-0169 (2017).
63. Zimmerman, F. *et al*. Teaching By Listening: The Importance of Adult-Child Conversations to Language Development. *Pediatrics* **124**, 342–349 (2009).
64. Liang, K.-Y. & Zeger, S. Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22 (1986).
65. Feldman, R. Parent–infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions. *Journal of Child Psychology and Psychiatry* **48**, 329–354, https://doi.org/10.1111/j.1469-7610.2006.01701.x (2007).
66. Caskey, M., Stephens, B., Tucker, R. & Vohr, B. R. Adult Talk in the NICU With Preterm Infants and Developmental Outcomes. *Pediatrics* **133**, e578–e584 (2014).
67. Yoo, H., Bowman, D. & Oller, D. K. The origin of protoconversation: An examination of caregiver responses to cry and speech-like vocalizations. *Frontiers in Psychology*, 1–15, https://doi.org/10.3389/fpsyg.2018.01510 (2018).
68. Lieven, E. V. M. In *Input and interaction in language acquisition* (eds Clare Gallaway & Brian J. Richards) 56–73 (Cambridge University Press, 1994).
69. Akhtar, N. The robustness of learning through overhearing. *Developmental Science* **8**, 199–209 (2005).
70. Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L. & Jasnow, M. D. *Rhythms of dialogue in infancy: Coordinated timing in development*. Vol. 66(2) (Univ of Chicago Press, 2001).
71. Stern, D. N. In *The effect of the infant on its caregiver* (eds Lewis, M. & Rosenblum, L. A.) 187–213 (Wiley, 1974).
72. Locke, J. L. Parental selection of vocal behavior: Crying, cooing, babbling, and the evolution of language. *Human Nature* **17**, 155–168 (2006).
73. Locke, J. L. & Bogin, B. Language and life history: A new perspective on the evolution and development of linguistic communication. *Behavioral & Brain Sciences* **29**, 259–325 (2006).
74. Oller, D. K. & Griebel, U. In *Evolutionary Perspectives on Human Development* (eds Burgess, R. & MacDonald, K.) 135–166 (Sage Publications, 2005).
75. Searcy, W. A. & Nowicki, S. In *The design of animal communication* (eds Hauser, M. D. & Konishi, M.) 575–595 (MIT Press, 1999).
76. Krebs, J. R. The significance of song repertoires: The Beau Geste hypothesis. *Animal Behavior* **25**, 475–478 (1977).
77. Kroodsma, D. E. In *The design of animal communication* (eds Hauser, M. D. & Konishi, M.) 319–342 (MIT Press, 1999).
78. Nottebohm, F. A brain for all seasons: cyclical anatomical changes in song control nuclei of the canary brain. *Science* **214**, 1368–1370 (1981).
79. Tyack, P. L. & Sayigh, L. In *Social influences on vocal development* (eds Snowdon, C. T. & Hausberger, M.) 208–233 (Cambridge University Press, 1997).
80. West, M. J., King, A. P. & Freeberg, T. M. In *Social influences on vocal development* (eds Charles T. Snowdon & Martine Hausberger) 41–56 (Cambridge University Press, 1997).
81. Winn, H. E. & Winn, L. K. The songs of the humpback whale Megaptera novaeangliae in the West Indies. *Marine Biology* **47**, 97–114 (1978).
82. Helweg, D. A., Frankel, A. S., Mobley, J. R. & Herman, L. M. In *Marine Mammal* Sensory Systems (eds J. A. Thomas, R. Kastelein, & A. Ya Supin) 459–483 (Plenum, 1992).
83. Yoo, H., Buder, E. H., Bowman, D. D., Bidelman, G. M. & Oller, D. K. Acoustic Correlates and Adult Perceptions of Distress in Infant Speech-Like Vocalizations and Cries. *Frontiers in Psychology* **10**, https://doi.org/10.3389/fpsyg.2019.01154 (2019).
84. Buder, E. H., Chorna, L., Oller, D. K. & Robinson, R. Vibratory Regime Classification of Infant Phonation. *Journal of Voice* **22**, 553–564 (2008).
85. Quine, W. V. O. In *Ontological Relativity and Other* Essays (ed W. V. O. Quine) 114–138 (Columbia Univ. Press, 1969).
86. Oller, D. K., Wieman, L., Doyle, W. & Ross, C. Infant babbling and speech. *Journal of Child Language* **3**, 1–11 (1976).
87. Oller, D. K. Infant vocalization and the development of speech. *Allied Health and Behavioral Sciences* **1**(523), 549 (1978).
88. Lynch, M. P., Oller, D. K., Steffens, M. L. & Buder, E. H. Phrasing in prelinguistic vocalizations. *Developmental Psychobiology* **28**, 3–23 (1995).

Acknowledgements

Author Contributions

D.K.O. wrote the main text and Supplementary Information and prepared Figs 1–3. D.K.O., M.C, E.H.B, and B.V. designed the research. D.D.B. supervised statistical analyses and conducted the GEE analyses. E.R.B. supervised and coordinated recordings in Memphis (the full-term infants). M.C. and B.V. supervised and coordinated recordings in Providence, RI (the preterm infants). D.K.O., E.R.B., Y.J., H.Y., C-C.L, and H.L. trained coders and supervised the coding process. All authors reviewed the manuscript and helped refine the writing.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-51352-0.

Competing Interests: The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.