

## RESEARCH ARTICLE

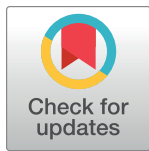
# Infant vocal category exploration as a foundation for speech development

Hyunjoo Yoo<sup>1</sup>\*, Pumpki Lei Su<sup>2</sup>, Gordon Ramsay<sup>3,4</sup>, Helen L. Long<sup>5</sup>, Edina R. Bene<sup>6</sup>, D. Kimbrough Oller<sup>6,7,8</sup>

**1** Department of Communicative Disorders, College of Arts & Sciences, The University of Alabama, Tuscaloosa, Alabama, United States of America, **2** Department of Speech, Language and Hearing, School of Behavioral and Brain Sciences, The University of Texas at Dallas, Richardson, Texas, United States of America, **3** Spoken Communication Laboratory, Marcus Autism Center, Children's Healthcare of Atlanta, Atlanta, Georgia, United States of America, **4** Department of Pediatrics, Emory School of Medicine, Atlanta, Georgia, United States of America, **5** Waisman Center, University of Wisconsin–Madison, Madison, Wisconsin, United States of America, **6** Origins of Language Laboratory, School of Communication Sciences and Disorders, University of Memphis, Memphis, Tennessee, United States of America, **7** Institute for Intelligent Systems, University of Memphis, Memphis, Tennessee, United States of America, **8** Konrad Lorenz Institute for Evolution and Cognition Research, Klosterneuburg, Austria

\* These authors contributed equally to this work.

\* [hyoo3@ua.edu](mailto:hyoo3@ua.edu)



## OPEN ACCESS

**Citation:** Yoo H, Su PL, Ramsay G, Long HL, Bene ER, Oller DK (2024) Infant vocal category exploration as a foundation for speech development. PLoS ONE 19(5): e0299140. <https://doi.org/10.1371/journal.pone.0299140>

**Editor:** Michael Döllinger, University Hospital Erlangen at Friedrich-Alexander-University Erlangen-Nürnberg, GERMANY

**Received:** July 18, 2023

**Accepted:** February 6, 2024

**Published:** May 29, 2024

**Copyright:** © 2024 Yoo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data is available in Mendeley Data (repository): [10.17632/xy7phf6jh.1](https://doi.org/10.17632/xy7phf6jh.1)

**Funding:** The research for this manuscript was funded to DKO and GR by the National Institutes of Health grants (R01 DC015108) from the National Institute on Deafness and Other Communication Disorders (<https://www.nidcd.nih.gov/>). The research for this manuscript was funded to GR by the National Institutes of Health grants (P50 MH100029) from the National Institute of Mental

## Abstract

Non-random exploration of infant speech-like vocalizations (e.g., squeals, growls, and vowel-like sounds or “vocants”) is pivotal in speech development. This type of vocal exploration, often noticed when infants produce particular vocal types in clusters, serves two crucial purposes: it establishes a foundation for speech because speech requires formation of new vocal categories, and it serves as a basis for vocal signaling of wellness and interaction with caregivers. Despite the significance of clustering, existing research has largely relied on subjective descriptions and anecdotal observations regarding early vocal category formation. In this study, we aim to address this gap by presenting the first large-scale empirical evidence of vocal category exploration and clustering throughout the first year of life. We observed infant vocalizations longitudinally using all-day home recordings from 130 typically developing infants across the entire first year of life. To identify clustering patterns, we conducted Fisher's exact tests to compare the occurrence of squeals versus vocants, as well as growls versus vocants. We found that across the first year, infants demonstrated clear clustering patterns of squeals and growls, indicating that these categories were not randomly produced, but rather, it seemed, infants actively engaged in practice of these specific categories. The findings lend support to the concept of infants as manifesting active vocal exploration and category formation, a key foundation for vocal language.

Health (<https://www.nimh.nih.gov>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Training grant information: Helen Long NICHD T32HD007489 (Role: Trainee) NICHD U54HD090256 (Role: Trainee)

**Competing interests:** D. K. Oller is an unpaid member of the LENA Scientific advisory board. However, this does not alter our adherence to PLOS ONE policies on sharing data and materials.

## Introduction

### Clustering of vocal types in early human development

To uncover mechanisms underlying language acquisition and the distant origins of language, various learning mechanisms have been proposed, including imitation and social interaction [1,2]. Infants' endogenous vocal production and exploration of vocal types has received comparatively little attention in the literature until recently [3]. The present study provides empirical evidence of the early emergence of vocal development in infants by systematically quantifying the occurrence of infant vocal categories, suggesting that active vocal exploration is fundamental to subsequent speech development and provides a foundation for interaction with caregivers.

The work is inspired by the fact that various vocal types of early human vocalizations known to be precursors to speech ("protophones") have been observed to occur in clusters [4,5] of particular phonatory types. There are three predominant types: 1) vocants are vowel-like sounds produced with normal phonation in the mid-pitch range of the individual infant; 2) squeals are produced at very high pitch, often in falsetto phonation; and 3) growls are low pitched or harshly phonated sounds often in fry. Clustering can be observed in that a baby may produce several squeals in a short period of time, and then for the next few minutes no squeals at all, while other vocal types continue to be produced. Also, from day to day there appear to occur wide variations in the number of squeals babies produce. Similar clustering has been observed for growls and other less frequently occurring sounds such as raspberries or whispers [6]. In the present paper we will greatly expand the quantitative exploration of this presumable non-randomness of infant vocal types by reporting on a sample of 130 infants recorded longitudinally all-day in their homes. In the present paper we define clustering as the non-random occurrence of particular protophones across sessions, that is, across specified time periods. The paper does not address immediate "repetition" of particular protophone types, one after another within sessions.

The pattern leads us to consider why a tendency for clustering exists. The activity often seems playful, and one cannot avoid wondering if it might constitute practice. But practice for what? If for language, then we are forced to ask how being capable of producing such sounds as squeals and growls could form foundations for language, because such utterances are not in and of themselves elements of language. They are not well-formed syllables, and they do not, indeed they cannot become words. We are driven to question why there could be any advantage to practicing such sounds. Our paper is based on the supposition that clustering of infant vocal types 1) helps to establish the principle of vocal category formation, the mastery of which is required for vocal language [6] and 2) provides information to caregivers about infant wellness.

### Foundations for mature vocal communication in infancy

In the first year of life, human infants engage in remarkably extensive vocal activity with protophones, a term that includes both precanonical and canonical speech-like sounds, while excluding cries, laughs and vegetative sounds. This vocal activity has been estimated to involve an average of 4–5 protophones per minute every waking hour from the first month of life and continuing throughout the first year [3]. Caregivers respond to protophones in face-to-face sustained interactions called "protoconversations" [7,8] that have no precedent in the animal kingdom as far as we know.

Surprisingly, even though these salient interactive periods between human caregivers and infants have generated considerable interest and speculation about foundations for language, the great bulk of the protophone production of human infants does not actually occur during

social interaction. Instead, examination of extensive laboratory recordings along with all-day recordings in infant homes has led to the conclusion that more than 90% of protophones are produced in endogenous, non-socially-directed activity [9], much of which might be termed infant “vocal exploration” or “vocal play” [5]. In an attempt to understand the mechanisms of language development, researchers have explored cross-species “babbling” in non-human primates and songbirds (for more information, see [10–14]). However, active vocal exploration seems to be uniquely observed in human infants, distinguishing them from other species. In all the cases where birds, bats, or humans engage in seemingly playful infant vocal activity, it is presumed that the babbling is a precursor to the mature form of vocal communication, and that the activity can be thought of as a kind of exploratory practice laying groundwork for the elaborate forms of song and speech.

### Categories of early vocalization and the idea of infant vocal practice

Widely accepted descriptions of early human vocal development designate categories of phonation that appear to occur universally in the first months of life [15–18]. Although they can be broken down into further subcategories, the three most common phonatory categories include: 1) vocants (also called vowel-like sounds) produced in normal phonation, the kind that dominates the utterances of natural languages, and produced at the mid-pitch range of the individual, 2) squeals, which are saliently high-pitched sounds, produced with typically at least twice the fundamental frequency ( $f_0$ ) of the infant’s usual voice, and 3) growls, produced typically with salient harsh, noisy phonation, often at low pitch, or produced in vocal fry (“pulse”) at very low pitch with respect to the infant’s typical  $f_0$  [19]; generally, no sound that is deemed to occur at higher than the mid-pitch range for the individual is categorized as a growl.

These three vocal types have been reported to occur not only by ethologically oriented observers who have tracked infants with longitudinal recordings [5,16,20], but also by parents who have often responded to open-ended questions about their infants’ vocal sounds in the first half year of life with terms drawn from the common parlance to describe them, in English, vowel-like sounds (which in our technical terminology are called vocants), squeals and growls. Other languages also often have terms to designate the primary phonatory protophones [6].

Of course, additional infant sounds do occur (raspberries, ingressive sounds, whispers, and so on) though far less frequently, and many utterances combine phonatory characteristics of the three primary protophones. As the first year progresses, infant utterances begin to show more speech-like characteristics, and by the second half year, more complex syllables occur, culminating in canonical babbling, with well-formed syllables such as “ba”, “da”, or “na”, and often reduplicated sequences of these sounds (“baba”, “nanana”, and so on) [21–23]. The phonatory characteristics of vocants, squeals and growls persist throughout the first year, and even the most speech-like fully canonical sequences are produced with phonatory properties that allow them to be categorized as vocants, squeals and growls, just as in the case of precanonical protophones.

Research having tracked the occurrence of protophones has long posited that vocal play, in particular with the three primary phonatory types, involves clustered production of the individual types, a non-random occurrence that has suggested infants may be engaged in vocal practice, perhaps in an attempt to consolidate the categories themselves [16,22]. This reasoning is founded in the idea that the categories emerge from exploration, rather than being innate vocal givens or learned from parents. The three principal protophones are thought to be creations of the infant, and the fact that they seem to fall into particular categories is thought of as the result of exploration in a non-random landscape of phonatory possibilities, where Waddingtonian wells of attraction [24] tend to draw the exploratory activity into the three

phonatory types. The pattern of development is reminiscent of the idea of quantal categorization in speech perception [25].

The fact that parents seem to notice the distinctions among the three has been proposed to be attributable at least in part to their non-randomness of occurrence across time [6]. An infant may be heard producing a train of squeals, for example, and although the great majority of protophones are vocants, the parent may notice the squeals as salient departures from the more typical vocants especially if they tend to be repeated or to occur in concentrated clusters. Parents engaging in face-to-face vocal interaction with their infants often attempt to elicit one of the three phonatory protophones by producing imitated versions of the sounds they believe to be in the infant repertoire. It is as if parents seek to confirm their infants' emerging vocal competence in a game of mutual imitation, where parents most often initiate imitation by trying to draw infants into producing one of the protophone types already under their command [26].

So the non-random occurrence of the protophone types is seen as an important feature of infant vocal development in two ways. First, clustering may constitute a type of practice, where infants seek to firm up their control of vocal categories through intentional manipulation of each type. This manipulation and learning of new types appears to constitute a critical foundation for language, because one of the necessary requirements of language is the ability to adapt to using new vocal categories (intonational types as well as syllables and phonemes) that are language specific and to provide the basis for learning to use indefinitely large inventories of possible words and sentences [6]. Second, if the reasoning is correct, clustering provides a basis for caregivers to recognize infant progress toward language in the sense that they can thereby discern infant category control. This parental discernment would seem to begin with simple recognition of the primary phonatory protophones, a recognition forming a basis for elaborate vocal interchanges with infants. Later in the first year, parents recognize the appearance of repetitive canonical syllables, which results in an intuitive reaction whereby parents begin to "negotiate" with infants [27] over the possible meanings of particular syllable sequences (yes, you said ba, it's a ball, say ba. . .) [28].

### **Why has there been so little research on vocal practice by clustering in human infancy?**

Given the seeming importance of clustered production of vocal types in infants, it may be seen as odd that very little research has been devoted to quantitative determination of the extent of such practice or its course of development. As far as we know, only we and some of the collaborators of the University of Memphis Origin of Language Laboratories (OLL) have addressed the issue empirically and in each case at small scale. 1) In the Supplementary Material to one of our papers [4], we presented a demonstration using lag sequential analysis [29,30] of a tendency of 9 infants in brief laboratory recordings to produce the three primary phonatory types in non-random sequences where the identity of an infant utterance (for example, a growl) predicted the likelihood that the next utterance in sequence would be of the same type at a higher than chance level. 2) In a separate demonstration reported in the same Supplementary Material, recurrence quantification analysis [31] was used to illustrate the clustering of squeals in a sequence of utterances from an infant, that is, a tendency for few squeals to be present at the beginning of a particular recording session, but a large number of squeals to cluster at the end of the session. 3) In a separate paper [32], we reported "session effects", whereby 3 infants showed a tendency to produce durational patterns in babbling that were notably different from one period (or session) of recording to another for the same infant, even within the same day.

Such session effects have been recognized throughout decades of longitudinal research on vocal development in infancy by the OLL, having required us to conclude that small-

scale sampling by recording of infant vocalizations could not be expected to yield representative samples of any aspect of infant vocal patterns (see also [33]). From one period to another, a wakeful and alert infant can change from being completely silent to being very voluble, and during periods of volubility [34], from producing a very limited repertoire of vocal types to producing a wide array of types. With a comfortable baby, the most common protophones by far are vocants [4]. But we have long noted that periods occur where squeals or growls can abruptly take center stage, and when they do, an observant individual cannot help but take notice.

There are several reasons that vocal exploration, particularly the quantification of vocal clustering, has received limited attention in the study of infant vocal development. First, researchers have historically focused heavily on the anatomic and physiological aspects of infant vocal development, viewing the vocal products as showing gradual progression from immature and unstructured forms to mature speech [35–39]. While this perspective has merit, it tends to overlook the critical role of infants as exploratory agents in their own vocal development. This perspective has been influenced by longstanding views, from the concept of the "tabula rasa" in Western philosophy to Jakobson's claim that prelinguistic vocalizations are random byproducts of biological inclinations [40]. These views are outdated, yet the tendency to underplay infant active involvement in vocal development persists.

Second, while some scholars have recognized and described vocal exploration in infancy, the precanonical stage of phonatory control development has received considerably less attention than the canonical stage and higher-level language development (e.g., words and sentences) in childhood (see review in [41]). Yet without the fundamental capability of phonatory control, articulatory development as manifest in canonical babbling would seem largely superfluous, since syllables overwhelmingly require phonation in their nuclei, and articulatory movements without phonation are largely soundless.

Third, in terms of the mechanisms of language development, social interaction and imitation have been the primary focus in the literature, with the seeming assumption that infants learn vocal categories by listening to parents, interacting with them vocally and imitating the existing baby sounds [42–44]. But in fact, only a small proportion of infant vocalizations are produced during vocal interaction, and imitation of parental sounds is extremely rare in early life [26,45]. In fact, it appears that the vast majority of apparent imitation by infants may actually be based on parental elicitation of the sounds parents know the infants already have in their repertoires, for example, squeals, growls, and vocants [28]. It is difficult to find evidence that infants introduce any new sounds into their exploratorily developed vocal repertoires as a result of listening to parental talk until late in the first year, when the first words begin to appear.

Lastly, most prior studies on infant vocal development have been conducted in a laboratory, for only a few minutes to about an hour [2,46,47]. Therefore, vocal exploration may often not have been salient enough to capture researchers' attention during such constrained time periods, and especially since parents are often instructed to elicit vocalization from their infants during such recordings. However, with the widespread availability of all-day home recordings in recent time [48–50] more representative data are now being obtained from naturalistic environments.

### Purposes of the present study

Our goal is to provide a large-sample quantitative assessment of the tendency of typically developing human infants to produce the three primary protophone categories in clusters, rather than simply randomly distributing the categories across segments of time. This

assessment, taking advantage of a massive database of recordings described in Methods, will also offer perspective on the degree to which infants across the first year give evidence of vocal category development, which is a crucial capacity in language. The work should also provide perspective on vocal practice in infancy.

The present study targets two primary questions: 1) to what extent do infants tend to produce phonatory categories in clusters rather than randomly across segments in recordings, and 2) to what extent do clustering patterns change with age during the first year of life.

## Method

### Participants

A total of 130 typically developing, English-learning infants participated in the present study (55% were male and 63% were designated as living in homes of relatively high SES as indicated by maternal education level). Participants whose data were analyzed in this study were recruited from September 2012 to October 2020. These infants were recruited in the Atlanta area as part of a much larger longitudinal study conducted by the Marcus Autism Center, Children's Healthcare of Atlanta and Emory University School of Medicine (NIH P50 MH100029) investigating vocal development in infants who 1) had no family history of autism or other developmental disorders or 2) had an older sibling with confirmed diagnosis of autism and thus were at elevated likelihood for autism. All 130 infants included in the present study were classified as having no clinical features at 2 and 3 years after thorough evaluations by Marcus Autism Center expert clinicians. Among the 130 infants, 103 were classified at enrollment as being at low likelihood for autism diagnosis. We report data on all 130 after having verified that the 27 at elevated likelihood showed very similar proportions of recordings with significant growl and squeal clustering to the low likelihood group of 103.

The OLL contributed human coding of the recordings of the infants as part of a collaborative NIDCD longitudinal study between the University of Memphis and the Marcus Autism Center/Emory University (NIH R01 DC015108). All the recording/coding protocols and consent documents signed by all the parents were approved by the Institutional Review Boards (IRBs) of Emory University and the University of Memphis. The third author at Emory University had access to identifying information for all participants in the present study. The data for the present work are not secondary. The recordings were collected under the supervision of the third author in Atlanta, and the coding was conducted under the supervision of the last author in Memphis.

### Recordings and the selection of 5-minute segments for coding

Language ENvironment Analysis (LENA) all-day recorders [51] were used to collect the audio data. The recorders are small and light enough (about the size of an iPod) for infants to carry in a pocket of a vest without any disturbance. Because LENA allows recording up to 16 hours a day with high audio quality (16 kHz sampling rate), it enables researchers to obtain representative data on infant vocalization and the auditory environment in the home. The recorder has been used routinely in a wide variety of research on early language development, starting with Zimmerman et al. [48].

Caregivers were instructed as in a variety of our prior studies on how to use a LENA device, including how to start, stop and pause recording. All recordings were carried out remotely in the home, once every month from birth to two years of age, and were scheduled as far as possible on the same calendar day of each month to ensure rotation of weekdays. Recordings were mailed out and retrieved via USPS priority mail. Each completed recording was uploaded through the LENA software to save the data. For further recording details see [52] for a

description of the procedures that were used at the Marcus Autism Center, where the recordings were conducted.

A total of 1154 all-day recordings were collected, an average of 8.9 recordings per infant (range 4–12). The recordings were always available for coding in the OLL and have been accessed repeatedly from September 1<sup>st</sup>, 2015 to the present. The human coding was conducted on 21 randomly selected 5-minute segments from each of the recordings. At the point of analysis, each available recording was assigned to one of six age groups as follows: 0–2 months, 3–4 months, 5–6 months, 7–8 months, 9–10 months, and 11–13 months.

### Coders, training, and coding environment

Thirty-six English-speaking graduate students in the University of Memphis School of Communication Sciences and Disorders were trained and coded the recordings. They were trained in phonetic transcription in their program of study but were more intensively trained for infant vocalization research by the fifth and the last author regarding all the coding parameters, including the primary phonatory types (i.e., squeals, growls and vocants). The six- to eight-week training for the coding of infant vocalizations is described in greater detail in some of our prior work [3,52]. The coders met periodically across the training period for weekly one-to-two-hour lectures and presentation of audio examples drawn from the OLL archives of infant vocalizations. After each lecture session, they conducted practice coding on recording segments. At the end of the week, these practice sessions were reviewed with the trainers during the lecture sessions and as needed individual coders had individual sessions with the trainers to address discrepancies between the categorizations that had been made by the trainees and also by key codes that are available for many of the segments based on coding by the last author, who is the longest term researcher on vocal development among any of the OLL participants. Coders were required to meet standards of agreement with the key codes (to be within 10% of the counts on protophones) by the 6<sup>th</sup> week of training or else additional training would ensue.

As a part of the training and providing a basis for coder agreement research on the three phonatory protophone categories, all the coders completed some practice coding on several of 9 full recordings where they coded all 21 randomly selected 5-minute segments independently of any other coder. These were not the research recordings on the 130 infants but were drawn instead from the same archives of LENA recordings at the Marcus Autism Center that produced the data on the 130 infants. These recordings are part of a set of recordings from the Marcus Autism Center utilized in the OLL for training of coders and about which agreement data are presented below.

Once trainees had completed training, they were admitted to the “coding team”, which typically consisted of about 12 individuals year-by-year, with a new group of trainees being recruited and trained in the fall semester of each year, replacing individuals having graduated that spring or summer.

A total of 24,234 5-minute segments were coded. The coders were assigned to groups of four infants, being blind to the age, risk status, and diagnostic outcome of the infants. They coded each recording completely (all 21 randomly selected segments in chronological order as occurring in the recordings from morning to night time) before proceeding to the next recording, which was randomly selected from among all the recordings from four infants assigned to each coder. Many of the coders were assigned to additional groups of four infants depending on their availability during the period of the research. The order of infants whose recordings were to be coded by each coder was also randomly assigned. Each individual coded all the recordings of an average of 4 infants (range 1 to 9 infants per coder).

Action Analysis, Coding, and Training (AACT [53]) was used for identification of each vocal type. AACT is a software environment enabling coders to label any kind of action in audio or video or both (see [4] Supplementary Material for details). Coders used keystroke or mouse-selected coding for each of the three protophone categories (squeals, vocants, growls), along with a few additional very infrequently occurring protophone categories (ingresses, whispers, raspberries, and non-phonated frication sounds) plus cries, whimpers, and laughs, while listening to each of the 5-minute segments. Coding was conducted in real time, so that each 5-minute segment took 5 minutes to code. Each coder responded to a brief questionnaire at the end of each coding segment, adding about another minute to coding time per segment. They answered the questions on a Likert-like scale from one to five in order to provide additional relevant information for each segment. For the present study, the key question concerned infant sleep. If the coder determined the infant to be asleep during the whole 5-minute segment, a value of 5 was assigned to that question, and the segment was then not included in the clustering evaluation for that recording.

### Coding categories

The coding yielded counts of squeals, vocants, and growls, leaving aside many other possible vocalizations of infants that occur far less frequently and are not considered to be precursors to speech. All coding was conducted at the utterance (that is, the breath-group) level. Extensive additional coding category information as well as details on how squeals, vocants, and growls were defined, along with spectrographic illustrations, are provided in Supporting Information, in the initial section called Coding categories, definitions and coding criteria.

### Data processing and statistical analysis of clustering

**Hypotheses tested.** The 21 randomly selected 5-minute segments coded for each available recording for each infant were tested for clustering of squeals with respect to vocants in one case and growls with respect to vocants in the other. In addition, the data were binned into six age groups to evaluate the possibility that clustering patterns might change with age.

**Preliminaries to data analysis.** We applied several exclusion criteria for segments prior to conducting Fisher's exact tests to determine evidence of clustering. For example, segments with high cry or whimper rates or segments where the infant was asleep were assumed to be incompatible with vocal play. Thus, we excluded segments where 1) 5 or more negative (cry and whimper) utterances occurred or 2) the infant was asleep throughout the 5 minutes according to the coder questionnaires. We also designated as not analyzable (NA) for the squeal vs. vocant comparison any recording where the sum for squeals over all segments was 0 (i.e., no squeals were coded), and similarly, we designated as NA for the growl comparison any recording where the sum for growls was 0 (i.e., no growls were coded). Segments remaining after these exclusions were termed "surviving" segments.

Finally, we designated as NA any recording that had only 1 surviving segment because at least 2 segments are required to conduct Fisher's exact tests. Those requirements having been met, a significant Fisher's exact test for the available comparisons within a recording would indicate that the protophone types were distributed in a non-random way, indicating clustering. In other words, either squeals or growls or both were being produced significantly more often in particular segments than in other segments of the same recording (see Supporting Information for details).

**Statistical analysis.** Fisher's exact test was selected as an appropriate test of whether the proportions for one variable (types of vocal categories) were different with respect to values of the other variable (5-minute segments) across each recording. This test is particularly suitable

for our data given that the raw count of each vocal category was often small and was usually uneven across cells (i.e., segments). In our case, we sought to determine whether the proportions for squeals and growls were different across 5-minute segments in a recording with respect to the proportion of vocants. The null hypothesis was that the proportion of squeals/growls or vocants was the same with respect to vocants across segments. A significant result indicated there was a significant difference in squeals/growls and vocants across segments, and thus that squeals/growls did not distribute randomly with respect to vocants.

### Coder agreement

To establish a foundation for assessing the reliability of coding across three protophone types, we conducted a comprehensive evaluation of coder agreement through various analyses (for details, refer to Supporting Information, Coder agreement). Our evaluation focused on a very large coding agreement set, comprising 21 segments from each of the 9 recordings, coded during the training period. The agreement set enabled us to obtain a very large number of inter-coder agreement pairings because, on average, about 32 coders independently coded the same 21 segments for each recording. A second, though smaller, agreement study involved 9 coders, which was conducted toward the end of the coding by those individuals on 523 segments semi-randomly selected from the data reported in Results. It makes sense to assume coder agreement would be lower during the training period and higher after actual data collection. The higher agreement obtained during the second, though smaller evaluation, tends to confirm this assumption.

These evaluations including coefficients of variation and Spearman rank order correlations demonstrated that coder agreement was highly statistically significant (see Supporting Information, Coder agreement). Permutation tests conducted on the 9 recording agreement set further confirmed that coder agreement was more than adequate to justify the analyses of clustering reported below.

### Results

We cleaned the data to ensure they met our exclusion criteria in accord with the specifications above, before conducting Fisher's exact tests. Thus, at the outset of the analysis, we identified the segments where infants were asleep according to the coders. Among all the recordings, 18,399 segments (76%) were identified in which the infants did not sleep, according to the coders' questionnaire responses, throughout the entire 5-minute period. We then located the number of segments that included too many negative vocalizations, defined as the sum of cries and whimpers being greater than or equal to 5. Fourteen percent of the non-sleep segments were thus eliminated from consideration in the Fisher's tests. These eliminations left 15,774 surviving segments for potential evaluation. Finally, we evaluated all the recordings for the possibility that they had 0 squeals or 0 growls. There were 125 recordings with no coded squeals and 123 with no coded growls. These recordings were designated as NA and were not evaluated by Fisher's exact tests on squeals vs. vocants or growls vs. vocants respectively.

### Percentage of recordings determined to show clustering

There was a considerable tendency for the infants to show significant clustering patterns. For squeals, 40% of the recordings showed significant clustering ( $p < .05$ ) by the Fisher's test, and for growls the value was 39% ( $p < .05$ ). These percentages of recordings that met the significant clustering criterion for each infant at each age were computed to include the NA recordings (which of course were not analyzed by Fisher's test) in their denominators, and consequently they supply an indication of squeal and growl clustering across *all* the recordings

available for each infant. Since we only selected 21 segments from each all-day recording for coding (only about 15% of the available recording time), these values surely underestimate the percentage of days the infants actually engaged in some clustering of squeals and/or growls. The data thus estimate a lower bound on the amount of clustering of squeals and growls at around 40% of all-day recordings.

Furthermore, evaluating whether individual infants showed *either* significant squeal clustering *or* significant growl clustering on each recording, we found that 61% of recordings showed a significant amount of clustering for one or the other. So it can be concluded that even though we sampled randomly from a relatively small proportion of the day, infants *usually* showed some discernible clustering activity of protophones.

At the infant level, we evaluated all the recordings available for each infant and found that 87% of infants showed at least one age at which their recordings showed significant squeal clustering *and* at least one age with significant growl clustering. There was not a single infant who, on evaluation of all the available recordings for the infant, showed neither a significant case of squeal clustering nor of growl clustering. Only 8 infants (6%) had 3 or fewer recordings with significant squeal or growl clustering. In contrast, 35 infants showed 6 or more recordings with significant growl clustering and 30 showed 6 or more with significant squeal clustering. Some infants seemed clearly to cluster growls more than squeals (about 10% of infants showed at least 3 more recordings with significant growl clustering than squeal clustering) while some (about 8%) clustered squeals more than growls.

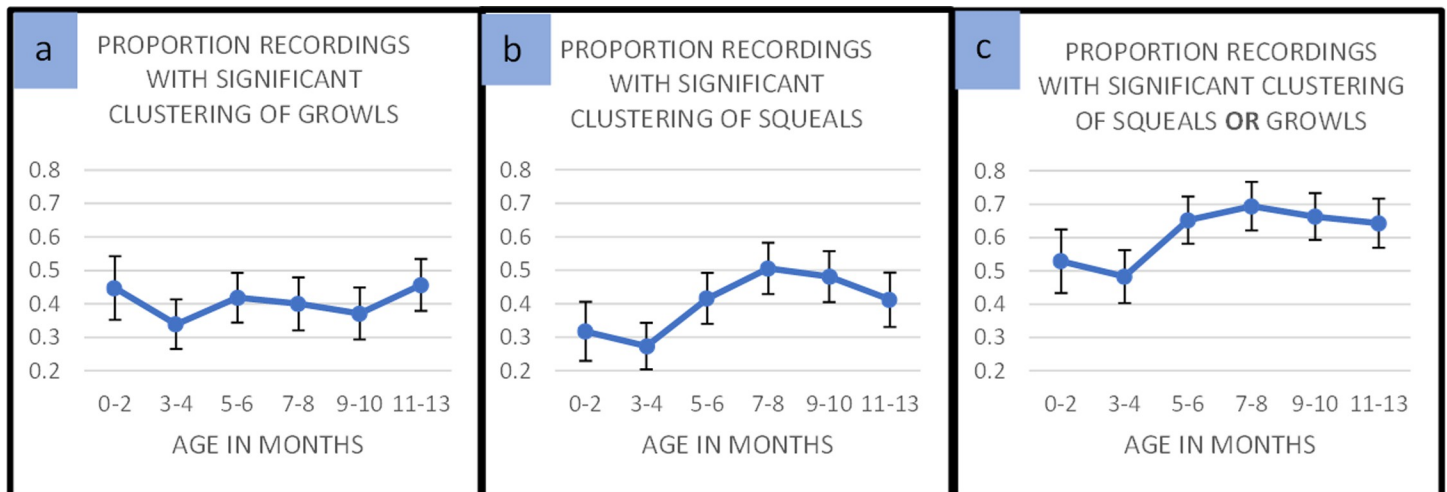
### Percentage of recordings with significant clustering across ages

When the data were analyzed by age for each vocal type comparison, we found significant clustering patterns of vocal types across all age groups. Fig 1A–1C supply a summary of the data, indicating that clustering occurred at all the ages. The error bars represent 95% bootstrapped confidence intervals computed around the means for all infants at each age level. Infant data were averaged for all the recordings of each infant within the age interval prior to computing means and confidence intervals.

In panels 1a and 1b the data show considerable occurrence of squeal and growl clustering in the very first months of life. Interestingly, the highest amount of clustering did not fall within the 3–4 month range, traditionally thought of as the period for vocal play in stage models of vocal development [16]. In fact, the 3–4-month interval showed the *lowest* mean values for both squeal and growl clustering. Squeal clustering showed a tendency to increase toward the middle of the year, peaking at 7–8 months. The squeal clustering pattern appeared to vary more across age than the growl clustering pattern, but in both cases, the final interval of the year, like the first interval, showed substantial clustering. Thus, a pattern of clustering occurred at all ages.

In Fig 1C, the data are presented in such a way that either squeal clustering or growl clustering was counted. Thus, each data point represents the proportion of infants whose recordings at the age in question showed either significant growl or squeal clustering. Consequently, the means are higher at each age than for either Fig 1A or 1B, illustrating the fact that infants often tended to produce recordings with clustering of squeals or growls but not both.

The confidence intervals supply information about possible significant differences across the ages. The apparent differences suggest, for example, a possible tendency for clustering overall (panel 1c) or clustering of squeals (panel 1b) to be more common beyond 5 months of age than before 5 months. Yet there are important provisos to offer about these apparent age differences, as we explain below.



**Fig 1. Clustering across ages:** The figure shows means of statistically significant clustering for recordings from infants at each of 6 age intervals. Values were computed at the infant level, such that all recordings for any infant within the age intervals indicated were averaged first, and 95% bootstrapped confidence intervals displayed as error bars were computed based on the averages for the number of infants who had recordings within those age intervals. **a.** Growls showed significant clustering for more than 30% of the infants at all ages. **b.** Squeals showed significant clustering for more than 30% of infants at all ages except the 3–4-month interval, where 27% of infants had significant clustering of squeals. **c.** In the final panel we display the proportion of infants who showed either significant growl or squeal clustering at each of the age intervals. From 48% to 69% of infants showed significant clustering of either growls or squeals at the various age intervals, and all the age intervals beyond 5 months revealed more than 60% of infants had significant clustering.

<https://doi.org/10.1371/journal.pone.0299140.g001>

## Discussion

It has long been observed, mostly without quantification, that human infants produce vocalizations in systematic patterns where some protophone categories occur in clusters across time, and it has been thought that these patterns of non-random occurrence might represent practice of categories emerging through infant vocal exploration [6,16]. To the best of our knowledge, the present paper presents the first large-scale empirical study systematically investigating the non-random occurrence of the three most salient phonatory protophone types of infancy, a study intended to provide perspective on the formation of vocal categories across the first year of life. Our analysis focused on vocants, squeals, and growls. Since vocants are the seeming default category (representing 80% or more of protophones), and since they manifest normal phonation, the type of phonation occurring in the vast majority of utterances in languages all over the world, our analysis focused on the clustering of protophones with non-normal phonation (squeals and growls) with respect to vocants. As seen in Fig 1C, more than 60% of infants exhibited significant clustering patterns on average across the ages for either squeals or growls. Moreover, every one of the 130 infants showed significant clustering of either squeals or growls in at least one recording. Across the six age groups, we observed significant clustering at all of them. Our findings offer robust empirical evidence supporting the idea that infants engage in systematic production of three vocal categories from the first months of life.

It was not unexpected that infants would show strong signs of clustering of squeals and growls in the age range from 3–4 months, because the idea of vocal play in infancy has been associated with stage models of vocal development that emphasize seeming vocal practice and vocal category repertoire “expansion” typically occurring during those months [6,16]. But our empirical investigation did not provide support for the idea that 3–4 months of age is the predominant period of clustering. On the contrary, that age range showed the *lowest* proportion of significant clustering for both squeals and growls. The data in fact suggest that clustering

was common at every age, with a tendency for the highest rates to occur from 5 months forward.

We are not, however, inclined to view this pattern of *age results* as offering the final word about vocal category clustering, if for no other reason, because there are many possible ways of coding phonatory types in infancy, and different approaches to coding could produce different age results. We adopted an efficient, simplifying approach, implementing real-time coding at the “utterance” level, restricting coders to three primary phonatory possibilities for each protophone utterance (vocant, squeal and growl). This approach made it possible to collect a large sample of data (assessing 1154 all-day recordings) with the resources available, but it clearly glossed over massive complexities. For example, *within* protophone utterances, more than one phonatory type often occurred, and as a result, there is no absolute standard of correctness for categorization of such utterances either on auditory or acoustic grounds. Furthermore, even the 3 phonatory types could have been subcategorized, if there had been time and money to implement more extensive coding. The squeal type is manifest in at least two high-pitch vocal types (loft/falsetto and harsh) and the growl type in at least two non-high-pitch vocal types (pulse/fry and harsh). Vocants also are manifest in subcategories that include at least modal, tense and breathy types. Emphasis should be applied to the “at least” phrase in these depictions, because there are yet additional vocal regimes at stake—for example, “harsh” growls can include at least three vocal regimes (subharmonic, biphonation, and chaotic, see [54]). So while our coding system is systematic and yields reliable data, it may have oversimplified the vocal categories of human infants. The age pattern in Fig 1C that suggests more clustering beyond 5 months than before 5 months is thus subject to question because of the limits of our coding system.

These provisos in mind, we offer a few additional thoughts about the apparent age patterns reported above and their relations with speculations derived from stage models (especially [16]) that have influenced our expectations about when vocal play should be expected to occur. The fact that significant clustering was seen even in the youngest group in the present study was something of a surprise (although other research has also shown surprisingly early flexibility of protophone production [55]). We anticipated finding at least a significant increase in clustering from the first to the second age interval. The occurrence of significant clustering beginning at the very first age interval suggests that some mechanism of vocal exploration is present from the beginning of life, even before cortical control is thought to fully emerge. Prior to the descent of the larynx and expansion of the oral cavity, practice of vocal types may be focused on phonatory control categories. Whether additional protophone categories, including primitive supraglottal articulations (e.g., gooing) as well as raspberries, whispers, and yells, which become prominent at later stages are also show clustering practice is a matter for future research. Our simplified real-time coding scheme limits our ability to clarify all the possible patterns of vocal practice at present.

Another issue of concern is the difficulty of consistent coding in accord with our method. Identifying vocal types in protophones produced by newborns clearly poses challenges beyond those requiring differentiation *among* the three phonatory types. Additional categories of infant sound making complicate the coding. For example, differentiating effort grunts (which often occur as a byproduct of movement, and which we exclude from the protophones by definition) and growls can be difficult due to their often shared characteristic of phonatory harshness. As infants get older, it appears most growls come to have longer durations, while effort grunts remain short, making it easier to recognize growls as distinct from grunts. During the newborn period, distinguishing between grunts and growls might then be viewed as resulting in an overestimation of growl quantities and thus tending to yield considerable apparent growl clustering at the earliest ages. The use of the term “grunts” in the vocal development literature

sometimes to refer not to effort grunts but to quasivowels used to communicate assent toward the end of the first year [56] or simply to continue vocal conversation also suggests a complication with the counting of vocants (quasivowels being among them in our interpretation). These coding issues, along with others, clearly complicate workable interpretations of the early age patterns.

Another unexpected finding was the relatively stable significant proportions of growl clustering across the 6 age groups (Fig 1A), compared to higher proportions of squeal clustering at ages beyond 5 months (Fig 1A). The apparent differing developmental trajectories for growls and squeals might be attributed to some special characteristic of maturation of vocal fold control that might affect squeal and growl production differently across ages. Squeals, by definition, are produced in high pitch, whereas growls are produced with low pitch, and raising vs. lowering of pitch requires complex coordination of very different combinations of intrinsic and extrinsic muscles of the larynx [54] with different aerodynamic requirements, the development of which is not yet well understood. While the present coding results suggest infants do produce squeals from birth, as they grow and gain greater control of laryngeal muscles, perhaps they become able to exhibit more consistent production of squeal register, especially for the falsetto/loft regime. On the other hand, growls typically occur at lower pitch and, it seems, often show harsh phonation that may require less fine glottal control than in falsetto/loft squealing. Perhaps, then, it is not particularly challenging for infants to produce harsh growls even from the earliest months, leading to somewhat similar proportions of clustering of growls across different ages. Squealing might in contrast grow in clustering across ages because the glottal control required for falsetto/loft may require maturation. These are of course speculations. Growling can also occur in non-harsh forms, with a pulse/fry regime. Does the pulse/fry regime require maturation similar to that for falsetto/loft? We do not know, and this supplies yet further reasons that our interpretation of the apparent age effects must remain uncertain.

Yet another issue worth considering is that inter-coder agreement was considerably higher for squeals than for growls. The segment-by-segment correlations in the agreement data for numbers of squeals within recordings showed 88% of pairwise comparisons were statistically significant, while 49% of pairwise comparisons were statistically significant for growls. Thus it seems possible that coders may not have been able to accurately capture developmental changes in growl production to as great an extent as for squeal production.

Even in the context of these difficulties of interpretation, our simplified approach to research on clustering of infant vocal types appears to reliably reflect the existence of three perceptually common protophone types that are often reported spontaneously by parents as forming features of infant repertoires. Squeals and growls represent salient departures from the default form of vocants that parents typically call vowel-like sounds. In the common parlance of many natural languages, squeals and growls are often represented with special terms [6]. Indeed, researchers of infant vocal development in the English-speaking world have adopted the terms squeal and growl directly *from* the common parlance, precisely because parents use them to describe salient infant vocal categories. Our inclination consequently is to take the data from the present real-time coding study seriously, at least to the extent that the data confirm the common parental report of the existence of these vocal categories in human infancy and help substantiate the impression that infants indeed practice these categories. We are also inclined to emphasize that the perceptions of human parents (as well as adults who do not have children, but often serve as caregivers and as member of our coding teams) are the natural gold standard of judgment about the nature and importance of infant vocalizations.

Thinking biologically, it is sensible to assume that human adults serve as the primary selection force on infant vocalization both in type and in quantity. If we are at some point to succeed in developing an automated system for categorization of human infant vocalizations, the

standard of judgment for the success of that automated system will be the extent to which it can simulate the judgments of adult human listeners. So when we ponder on why there is so much infant vocal activity and so much clustering of vocal types within that activity, the biological perspective harkens back to the selective role of caregivers. The fitness signaling theory [57–59] suggests that ancient hominin caregivers noticed (and modern human caregivers notice) the occurrence of infant protophones and judge them in terms of the extent to which they indicate the wellness of infant vocalizers, providing a basis for selective investment in infants whose protophones are most indicative of fitness. Infants are not required to direct their vocalizations to caregivers (although it is important that they sometimes do so), because caregivers can notice protophones as indicating an infant is well and progressing normally toward vocal communication even if the infant is simply playing with sounds.

Perhaps one of the primary ways infants can reveal their wellness to caregivers is by producing protophone types in clusters that suggest practice with their spontaneously developed vocal categories. One might ask how competent parents could possibly fail to notice vocal clustering. Surely recognition of clustering of vocal types suggests infants are acquiring a system, admittedly a prelinguistic system, but a system of vocal categories under infant autonomous control. The seeming practice may even serve the purpose of confirming to infants themselves that they are on a path of increasing voluntary vocal control. Whether the infants realize it or not, that path reveals their possession of a capacity and an inclination not seen in any other ape, a capacity and inclination for voluntary manipulation of vocal categories, a capacity and inclination without which, it has been argued [59], language would be impossible.

## Supporting information

### S1 File.

(ZIP)

### S2 File.

(ZIP)

**S3 File. Supporting information includes details on coding categories and coding agreement.** It also includes waveforms (S1–S5 Waveforms) and spectrograms of each vocal category (S1–S10 File) along with relevant details.

(PDF)

## Acknowledgments

We would like to express our gratitude to the families whose infants participated in this research and to the graduate student coders in Memphis.

## Author Contributions

**Conceptualization:** Hyunjoo Yoo, Pumpki Lei Su, Gordon Ramsay, Helen L. Long, Edina R. Bene, D. Kimbrough Oller.

**Data curation:** Hyunjoo Yoo, Pumpki Lei Su, Gordon Ramsay, Helen L. Long, Edina R. Bene, D. Kimbrough Oller.

**Formal analysis:** Hyunjoo Yoo, Pumpki Lei Su, Gordon Ramsay, D. Kimbrough Oller.

**Funding acquisition:** Gordon Ramsay, D. Kimbrough Oller.

**Investigation:** Hyunjoo Yoo, Pumpki Lei Su, Gordon Ramsay, Helen L. Long, D. Kimbrough Oller.

**Methodology:** Hyunjoo Yoo, Pumpki Lei Su, Gordon Ramsay, Edina R. Bene, D. Kimbrough Oller.

**Project administration:** Gordon Ramsay, Edina R. Bene, D. Kimbrough Oller.

**Resources:** Gordon Ramsay, Edina R. Bene, D. Kimbrough Oller.

**Software:** Hyunjoo Yoo, Pumpki Lei Su, Gordon Ramsay, D. Kimbrough Oller.

**Supervision:** Gordon Ramsay, Edina R. Bene, D. Kimbrough Oller.

**Validation:** Hyunjoo Yoo, Pumpki Lei Su, Gordon Ramsay, Helen L. Long, Edina R. Bene, D. Kimbrough Oller.

**Visualization:** Hyunjoo Yoo, Pumpki Lei Su, Gordon Ramsay, Helen L. Long, Edina R. Bene, D. Kimbrough Oller.

**Writing – original draft:** Hyunjoo Yoo, Pumpki Lei Su, D. Kimbrough Oller.

**Writing – review & editing:** Hyunjoo Yoo, Pumpki Lei Su, Gordon Ramsay, Helen L. Long, Edina R. Bene, D. Kimbrough Oller.

## References

1. Trevarthen C. Infant Intersubjectivity: Research, Theory, and Clinical Applications. *Journal of Child Psychology and Psychiatry*. 2001(42):3–48. PMID: [11205623](https://pubmed.ncbi.nlm.nih.gov/11205623/)
2. Beebe B, Stern D, Jaffe J. The kinesic rhythms of mother-infant interactions. In: Aw S, S F, editors. *Of Speech and Time*. Hillsdale NJ: Erlbaum; 1979.
3. Oller DK, Caskey M, Yoo H, Bene ER, Jhang Y, Lee C-C, et al. Preterm and full term infant vocalization and the origin of language. *Scientific Reports*. 2019;9.
4. Oller DK, Buder EH, Ramsdell HL, Warlaumont AS, Chorna L, Bakeman R. Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences*. 2013; 110(16):6318–632. <https://doi.org/10.1073/pnas.1300337110> PMID: [23550164](https://pubmed.ncbi.nlm.nih.gov/23550164/)
5. Stark RE. Infant vocalization: A comprehensive view. *Infant Medical Health Journal*. 1981; 2(2):118–28.
6. Oller DK. *The Emergence of the Speech Capacity*. Mahwah, NJ: Lawrence Erlbaum Associates; 2000. 428 p.
7. Gratier M, Devouche E, Guellai B, Infanti R, EbruYilmaz, Parlato-Oliveira E. Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology*. 2015; 6:1–10.
8. Rochat P, Querido JG, Striano T. Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Developmental Psychology*. 1999; 35(4):950. <https://doi.org/10.1037//0012-1649.35.4.950> PMID: [10442864](https://pubmed.ncbi.nlm.nih.gov/10442864/)
9. Long HL, Ramsay G, Bene ER, Burkhardt-Reed MM, Oller DK. Perspectives on the origin of language: Infants vocalize most during independent vocal play but produce their most speech-like vocalizations during turn taking. *PLoS One*. 2022. <https://doi.org/10.1371/journal.pone.0279395> PMID: [36584126](https://pubmed.ncbi.nlm.nih.gov/36584126/)
10. Elowson AM, Snowdon CT, Lazaro-Perea C. 'Babbling' and social context in infant monkeys: parallels to human infants. *Trends in Cognitive Sciences*. 1998; 2(1):31–7. [https://doi.org/10.1016/s1364-6613\(97\)01115-7](https://doi.org/10.1016/s1364-6613(97)01115-7) PMID: [21244960](https://pubmed.ncbi.nlm.nih.gov/21244960/)
11. Knörnschild M, Behr O, von Helversen O. Babbling behavior in the sac-winged bat (*Saccopteryx bilineata*). *Naturwissenschaften*. 2006; 93:451–4. <https://doi.org/10.1007/s00114-006-0127-9> PMID: [16736178](https://pubmed.ncbi.nlm.nih.gov/16736178/)
12. Nooteboom SG. Anatomy and timing of vocal learning in birds. In: Hauser MD, Konishi M, editors. *The design of animal communication*. Cambridge, MA: MIT Press; 1999. p. 63–110.
13. Oller DK, Griebel U, Iyer SN, Jhang Y, Warlaumont AS, Dale R, et al. Language origin seen in spontaneous and interactive vocal rate of human and bonobo infants. *Frontiers in Psychology*. 2019; 10(729).
14. Snowdon CT, Elowson AM, Rousch RS. Social influences on vocal development in New World primates. In: Snowdon CT, Hausberger M, editors. *Social influences on vocal development*. New York, NY: Cambridge University Press; 1997. p. 234–48.
15. Holmgren K, Lindblom B, Aurelius G, Jalling B, Zetterstrom R. On the phonetics of infant vocalization. In: Lindblom B, Zetterstrom R, editors. *Precursors of Early Speech*. New York: Stockton Press.; 1986. p. 51–63.

16. Stark RE. Stages of speech development in the first year of life. In: Yeni-Komshian G, Kavanagh J, Ferguson C, editors. *Child Phonology*, vol 1. New York: Academic Press; 1980. p. 73–90.
17. Kent RD. The maturational gradient of infant vocalizations: Developmental stages and functional modules. *Infant Behavior and Development*. 2022; 66:101682. <https://doi.org/10.1016/j.infbeh.2021.101682> PMID: 34920296
18. Yoo H, Oller DK, Ha S. Early emergence and development of proto-phonemes in the first year of life. *Communication Sciences & Disorders*. 2021; 26(1):1–12.
19. Buder EH, Chorna L, Oller DK, Robinson R. Vibratory regime classification of infant phonation. *Journal of Voice*. 2008; 22:553–64. <https://doi.org/10.1016/j.jvoice.2006.12.009> PMID: 17509829
20. Koopmans-van Beinum FJ, van der Stelt JM. Early stages in the development of speech movements. In: Lindblom B, Zetterstrom R, editors. *Precursors of Early Speech*. New York: Stockton Press.; 1986. p. 37–50.
21. Lohmander A, Holm K, Eriksson S, Lieberman M. Observation method identifies that a lack of canonical babbling can indicate future speech and language problems. *Acta Paediatrica*. 2017; 106(6):935–43. <https://doi.org/10.1111/apa.13816> PMID: 28271541
22. Oller DK. The emergence of the sounds of speech in infancy. In: Yeni-Komshian G, Kavanagh J, Ferguson C, editors. *Child phonology*, Vol 1: Production. New York: Academic Press; 1980. p. 93–112.
23. Stoel-Gammon C. Prespeech and early speech development of two late talkers. *First Language*. 1989; 9:207–24.
24. Waddington CH. *The strategy of genes: a discussion of some aspects of theoretical biology*. London: George Allen & Unwin; 1957.
25. Stevens KN, Keyser SJ. Quantal theory, enhancement and overlap. *Journal of Phonetics*, 38(1), 2010; 38(1):10–9.
26. Long HL, Oller DK, Bowman DD. Reliability of Listener Judgments of Infant Vocal Imitation. *Frontiers in Psychology*. 2019; 10(1340). <https://doi.org/10.3389/fpsyg.2019.01340> PMID: 31244735
27. Ramsdell HL, Oller DK, Buder EH, Ethington CA, Chorna L. Identification of prelinguistic phonological categories. *Journal of Speech Language and Hearing Research*. 2012; 55:1626–9. [https://doi.org/10.1044/1092-4388\(2012/11-0250\)](https://doi.org/10.1044/1092-4388(2012/11-0250)) PMID: 22490623
28. Papoušek M. *Vom ersten Schrei zum ersten Wort: Anfänge der Sprachentwicklung in der vorsprachlichen Kommunikation*. Bern: Verlag Hans Huber; 1994.
29. Bakeman R, Brownlee JR. The strategic use of parallel play: A sequential analysis. *Child Development*. 1980; 51(3):873–8.
30. Bakeman R, Quera V. Log-linear approaches to lag-sequential analysis when consecutive codes may and cannot repeat. *Psychological Bulletin*. 1995; 118(2):272–84.
31. Webber CL, Zbilut JP. Recurrence quantification analysis of nonlinear dynamical systems. In: Riley MA, Van Orden GC, editors. *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*. Web Book: National Science Foundation Program in Perception, Action, and Cognition; 2005.
32. Oller DK, Nathani Iyer S, Buder EH, Kwon K, Chorna L, Conway K. Diversity and contrastivity in prosodic and syllabic development. In: Trouvain J, Barry W, editors. *Proceedings of the International Congress of Phonetic Sciences*. Saarbrücken, Germany: International Phonetics Society; 2007. p. 303–8.
33. Molemans I, Van den Berg R, Van Severen L, Gillis S. How to measure the onset of babbling reliably. *Journal of Child Language*. 2012; 39(3):523–52. <https://doi.org/10.1017/S0305000911000171> PMID: 21892989
34. Abney DH, Warlaumont AS, Oller DK, Wallot S, Kello CT. Multiple coordination patterns in infant and adult vocalizations. *Infancy*. 2017; 22(4):514–39. <https://doi.org/10.1111/infa.12165> PMID: 29375276
35. Kent RD. Psychobiology of speech development: coemergence of language and a movement system. *American Journal of Physiology*. 1984; 246:R888–R94. <https://doi.org/10.1152/ajpregu.1984.246.6.R888> PMID: 6742163
36. Bosma JF. Anatomic and physiologic development of speech apparatus. In: D.B.Tower, editor. *Human Communication and its Disorders* (vol 3). New York: Raven Press; 1975.
37. Kent RD. The biology of phonological development. In: Ferguson C, Menn L., Gammon CS, editors. *Phonological development: Models, Research, Implications*. Parkton, MD: York Press, Inc; 1992. p. 65–89.
38. Laitman JT, Reidenberg JS. Specializations of the human upper respiratory and upper digestive systems as seen through comparative and developmental anatomy. *Dysphagia*. 1993; 8(4):318–25. <https://doi.org/10.1007/BF01321770> PMID: 8269722
39. Harold MP, Barlow SM. Effects of environmental stimulation on infant vocalizations and orofacial dynamics at the onset of canonical babbling. *Infant Behavior and Development*. 2013; 36(1):84–93. <https://doi.org/10.1016/j.infbeh.2012.10.001> PMID: 23261792

40. Jakobson R. *Kindersprache, Aphasie, und allgemeine Lautgesetze*. Uppsala: Almqvist and Wiksell; 1941.
41. Vihman MM. *Phonological Development: The First Two Years*. Malden, MA: Wiley-Blackwell; 2014.
42. Gratier M, Devouche E. Imitation and repetition of prosodic contour in vocal interaction at 3 months. *Developmental Psychology*. 2011; 47(1):67–76. <https://doi.org/10.1037/a0020722> PMID: 21244150
43. Kugiumutzakis G. Genesis and development of early infant mimesis to facial and vocal models. In: Nadel J, Butterworth G, editors. *Imitation in Infancy*. Cambridge Studies in Cognitive Perceptual Development. New York, NY, USA: Cambridge University Press; 1999. p. 36–59.
44. Legerstee M. Infants use multimodal information to imitate speech sounds. *Infant Behavior & Development*. 1990; 13(3):343–54.
45. Athari P, Dey R, Rvachew S. Vocal imitation between mothers and infants. *Infant Behavior and Development*. 2021; 63:101531. <https://doi.org/10.1016/j.infbeh.2021.101531> PMID: 33582572
46. Hsu HC, Fogel A. Social regulatory effects of infant non-distress vocalization on maternal behavior. *Developmental Psychology*. 2003; 39(6):97–991.
47. Jaffe J, Beebe B, Feldstein S, Crown CL, Jasnow MD. *Rhythms of dialogue in infancy: Coordinated timing in development*. Chicago: Univ of Chicago Press; 2001.
48. Zimmerman F, Gilkerson J, Richards J, Christakis D, Xu D, Gray S, et al. Teaching By listening: The importance of adult-child conversations to language development. *Pediatrics*. 2009; 124:342–9. <https://doi.org/10.1542/peds.2008-2267> PMID: 19564318
49. Johnson K, Caskey M, Rand K, Tucker R, Vohr B. Gender differences in adult-infant communication in the first months of life. *Pediatrics*. 2014; 134(6):e1603–10. <https://doi.org/10.1542/peds.2013-4289> PMID: 25367542
50. Richards JA, Xu D, Gilkerson J, Yapanel U, Gray S, Paul T. Automated assessment of child vocalization development using LENA. *Journal of Speech, Language, and Hearing Research*. 2017; 60(7):2047–63.
51. Ford M, Baer CT, Xu D, Yapanel U, Gray SS. *The LENA language environment analysis system: Audio specifications*. Boulder, CO; 2007.
52. Oller DK, Griebel U, Bowman DD, Bene ER, Long HL, Yoo H, et al. Infant boys found to be more vocal than infant girls. *Current Biology*. 2020;30(10): PR426-R7.
53. Delgado RE, Buder EH, Oller DK. *AACT (Action Analysis Coding and Training)*. Miami, FL: Intelligent Hearing Systems; 2010.
54. Buder EH, McDaniel VF, Bene ER, Ladmirault J, Oller DK. Registers in infant phonation. *Journal of Voice*. 2018. <https://doi.org/10.1016/j.jvoice.2017.12.013> PMID: 29650330
55. Jhang Y, Oller DK. Emergence of functional flexibility in infant vocalizations of the first 3 months. *Frontiers in Psychology*. 2017; 8(300). <https://doi.org/10.3389/fpsyg.2017.00300> PMID: 28392770
56. McCune L, Vihman MM, Roug-Hellichius L, Delery DB, Gogate L. Grunt communication in human infants (*Homo sapiens*). *J Comp Psychol*. 1996; 110(1):27–36. <https://doi.org/10.1037/0735-7036.110.1.27> PMID: 8851550
57. Locke JL. Parental selection of vocal behavior: Crying, cooing, babbling, and the evolution of language. *Human Nature*. 2006; 17:155–68. <https://doi.org/10.1007/s12110-006-1015-x> PMID: 26181412
58. Locke JL. Evolutionary developmental linguistics: Naturalization of the faculty of language. *Language Sciences*. 2009; 31:33–59.
59. Oller DK, Griebel U. Functionally flexible signaling and the origin of language. *Frontiers in Psychology*. 2021; 11(4092). <https://doi.org/10.3389/fpsyg.2020.626138> PMID: 33574785