Published in final edited form as:

Clin Linguist Phon. 2006 July ; 20(5): 351–369. doi:10.1080/02699200500211451.

Assessing Vocal Development in Infants and Toddlers

Suneeti Nathani,

Communication Sciences and Disorders, The University of Georgia, (706) 542-4602, (706) 542-5348 fax

David J. Ertmer, and

Speech, Language and Hearing Sciences, Purdue University, (765) 496-2249, (765) 494-0771 fax

Rachel E. Stark

Suneeti Nathani: snathani@uga.edu; David J. Ertmer: dertmer@purdue.edu

Abstract

The purpose of this study was to examine changes in prelinguistic vocal productions during the first 20 months of life. Vocalizations were classified into 23 mutually exclusive and exhaustive types, and grouped into five ascending levels using the Stark Assessment of Early Vocal Development-Revised (SAEVD-R). Data from 30 typically developing infants, aged 0–20 months, show that older infants attained higher developmental levels on the SAEVD-R than younger infants. Infants 0–2, 3–5, and 6–8 months of age primarily produced vocalizations from Levels 1 (Reflexive), 2 (Control of Phonation), and 3 (Expansion). Infants 9–20 months of age also produced vocalizations from Level 4 (Basic Canonical Syllables). Only infants from 16–20 months of age produced Level 5 (Advanced Forms) vocalizations in significant quantities. The outcomes indicate that the SAEVD-R is a valuable instrument for evaluating prelinguistic vocal development.

Keywords

PRELINGUISTIC SPEECH DEVELOPMENT; VOCAL DEVELOPMENT; SPEECH ASSESSMENT; INFANTS; TODDLERS

The nature of prelinguistic vocalizations has long been the subject of speculation and model building (Jakobson, 1968). Most recent theoretical explanations have suggested that ‘self-organization’ and principles of ‘natural logic’ result in advances in vocal production from vowel-like sounds to closant-vocant combinations to mature adult-like syllables. This development is facilitated by progressive integration of subcomponents of communication (e.g., perceptual biases, cognitive, anatomical, and physiological substrates). In addition, it has been proposed that the social context or the communicative function of these vocalizations is critical in linking vocal production with meaning, thereby facilitating the development of mature language (Oller, 2000, see Vihman, 1996 for a review).

Stark, Bernstein, and Demorest (1993) speculated that several possible relationships could exist between social communication and vocal production as they become integrated. For example, advances in one subsystem might immediately coincide with advances in the other subsystem or, in contrast, there might be a progressive integration of the various subsystems.

They suggested that in order to examine these different hypotheses, it is necessary to evaluate social communication and sound production independent of each other and then examine potential relationships across the subsystems.

Independent assessment of various domains of prelinguistic communication development may also enable us to pinpoint the locus of a communication disorder, if one exists. For example, it may allow us to make preliminary decisions about whether the locus of the communication disorder is at the sound production level or at the social level of communication. The present study is the second in a series of studies designed to assess the relationship between social communication and sound production in infants and toddlers. The first study conducted by Stark et al. (1993) assessed the adequacy of a scale designed to exclusively address the social aspects of communication. This second study is designed to assess the vocal production abilities of infants and toddlers from approximately 0–20 months of age using a research-based instrument, the Stark Assessment of Early Vocal Development-Revised (SAEVD-R),

Commercially available clinical tools for assessing early vocal development, including the Bayley Scales of Infant Development - II (BSID-II; Bayley, 1993), the Communication and Symbolic Behaviour Scales (CSBS; Wetherby & Prizant, 1993), the Early Language Milestones Scale (ELM; Coplan, 1986), the Infant-Toddler-Meaningful Auditory Integration Scale (IT-MAIS; Zimmerman-Phillips, Robbins, & Osberger, 1997), the Rossetti Infant-Toddler Language Scale (Rossetti, 1990), and the Preschool Language Scale-4 (PLS-4; Zimmerman, Steiner, & Pond, 2002) vary widely in the range of prelinguistic vocal behaviours examined and definitions of vocalization types. For example, ‘babbling’ is subdivided into ‘monosyllabic’ babbling (e.g., [ba]) and ‘polysyllabic’ babbling (e.g., [babababa]) on the ELM whereas on the Rossetti scale, babbling refers only to multiple productions of a syllable. The ELM and the Rossetti scale examine the occurrence of cooing and babbling whereas the CSBS, Bayley, and IT-MAIS assess only the production of different consonant, vowel, and syllable types. No specific mention of cooing or babbling is made.

Similarly, the ages of emergence of assessed vocalization types vary across instruments. The earliest age of emergence of babbling (defined as disyllabic combinations) as noted by the PLS-4 is 12–17 months whereas the earliest age of babbling (defined as multisyllabic combinations) on the Rossetti is 3–6 months in contrast to the widely-accepted ages of emergence between 6 and 12 months by most researchers (see Vihman, 1996). Taken together, these disparities highlight the need for a comprehensive assessment tool that is research-based.

Describing the phases that infants and toddlers must pass through on their way to becoming talkers has been the primary goal of vocal development research for almost 30 years. The first models of vocal development were proposed in the 1970s and 1980s by Koopmans-van Beinum and van der Stelt (1986), Oller (1980), Roug, Landberg, and Lundberg (1989), Stark (1980), and Zlatin (1975). These models identified four to five levels of vocal development that were characterized by the emergence of increasingly complex vocalization types. The models also provided estimates of the ages at which these vocalization types were expected to emerge (see Vihman, 1996 for review). For example, Stark (1980) proposed a model with five levels: Reflexive Sounds (0–2 months); Cooing and Laughter (2–4 months); Vocal Play (4–8 months); Reduplicated Babbling (8–10 months); and Nonreduplicated Babbling & First Words (10–14 months).

Although all these models included similar vocalization types, ages of emergence, and numbers of levels, and are widely accepted as benchmarks for assessing progress in infant

vocal development, there are differences in terminology and operational definitions across them. For example, both Oller and Stark include quasi-resonant nuclei (Q) in the first stage of vocal development. Oller defines Qs as vowel-like sounds with normal phonation but with limited resonance, hence the name quasi-resonant. If sounds have normal resonance, they are classified as fully resonant nuclei, or Fs, and are presumed to occur in the second stage of vocal development. Stark (1986) uses Oller's basic definitions of Qs and Fs. However, Stark acknowledged that she also used length as a criterion to decide whether to categorize a vocalization a Q or an F. Thus, short vowel-like sounds irrespective of resonance get categorized as Qs under her scheme whereas short vowels with full resonance would be categorized as Fs under Oller's categorization scheme (personal communication, Oller & Stark, 1999). These differences across laboratories, often unspecified, hamper the portability of vocalization types and make comparisons across studies difficult. Correspondingly, despite refinements of these models over the years, (e.g., Oller, 2000; Stark, Ansel, & Bond, 1988) a need for widely accepted, replicable, and comprehensive operational definitions for a full range of early vocalizations persists (Mitchell & Kent, 1990).

Age-of-emergence estimates are also somewhat different across models. For example, Oller (1980) noted the emergence of canonical syllables around 5–10 months of age whereas Stark (1980) and Roug et al. (1989) noted the occurrence of these syllables starting only around 8 months of age at the earliest. It may be that the operational definitions of these vocalization types and/or criteria used to determine emergence of vocal types varied across investigations, or they may reflect individual variability (Roug et al., 1989). Detailed evaluations of specific vocalization types have also provided some revised estimates of the original age of emergence estimates. For example, early models suggested that reduplicated (RB) and nonreduplicated babbling (NB) develop sequentially. Intensive examinations by Mitchell and Kent (1990) and Smith, Brown-Sweeney, and Stoel-Gammon (1989) revealed that RB and NB develop concurrently rather than sequentially. These findings suggest that an intensive examination of a full range of vocalization types is needed to verify and expand the age-of-emergence findings from these pioneering studies.

In addition, vocalization types included in existing models were defined using a combination of perceptual, motor, sound production, and social communication characteristics. For example, 'cooing' is described as 'comfort' sounds occurring in the context of turn-taking (Stark, 1980). Infants, however, often produce similar sounds across a variety of contexts (Goldstein, King, & West, 2003; Oller & Buder, 2004). Vocalization types, therefore, need to be characterized irrespective of the context in which they occur. Such an independent appraisal of vocal production would also allow researchers to examine potential relationships across subsystems of communication - a primary goal of this series of studies. The SAEVD-R was developed to capture infant vocal production by exclusively relying on perceptual and articulatory characteristics of vocalizations.

Finally, many studies of infant vocal development have focused on the emergence of particular milestones. For example, Oller and colleagues have primarily focused on the onset of canonical syllables (Oller, 1980; Oller & Eilers, 1988; Oller & Lynch, 1992). Two case studies that intensively examined early- and late-emerging types of vocalizations have shed new light on typical and disordered prelinguistic vocal development. Observations of the vocalizations of a child born without any cochleas (i.e., total deafness) indicated that quasi-resonant nuclei predominated this child's vocalization repertoire prior to the onset of canonical babbling; fully resonant nuclei were rarely produced (Lynch, 1989). This finding is in contrast to children with normal hearing who typically produce both quasi- and fully-resonant nuclei prior to the onset of canonical babbling. Similarly, a preliminary study of two deaf children with multichannel cochlear implants by Ertmer and colleagues (2002)

revealed that besides the establishment of canonical syllable production, advances in vocal development were signalled by increases in consonant, vowel, and diphthong diversity and the emergence of postcanonical vocalizations, such as jargon (see definition in Appendix). The study by Lynch highlights the role of audition for even the most basic types of vocalizations. The findings of Ertmer et al. suggest that examination of complex prelinguistic vocalization types is also necessary to adequately document progress in children who acquire speech at later-than-typical ages. The microanalysis required for these detailed investigations, however, is practically feasible only with small sample sizes. The present study evaluated vocalization characteristics in 30 typically developing infants and toddlers.

Researchers have also varied in their treatment of speech-like and non speech-like vocalizations when analyzing prelinguistic vocalizations. Prelinguistic vocalizations are broadly classified into: (a) non speech-like vocalizations, and (b) speech-like vocalizations, or protophones (Oller, 2000). Non speech-like sounds, as the name suggests, are not considered precursors to speech. They have fixed links to biological functions and are typically subdivided into two categories: fixed signals (such as crying, laughter, and so on) and vegetative sounds (such as burps, hiccups, and so on). Protophones, on the other hand, are presumed to be precursors to speech. They comprise several different vocalization types and progress along a continuum becoming more speech-like toward the end of the first year of life. For example, protophones in the first few months of life primarily consist of vowel-like vocalizations whereas vocalizations toward the end of the first year of life include various consonant and vowel combinations.

Oller (1980) as well as Koopmans-van Beinum and van der Stelt (1986) advocate using only protophones to assess advances in prelinguistic vocal development whereas Stark (1980) included both protophones and non speech-like vocalizations. Stark recommends inclusion of non speech-like sounds because she claims these might provide foundational elements recruited later for speech. In the present study, the ratio of non speech-like sounds to protophones across ages was tracked. It was expected that non speech-like sounds would reduce in frequency over time thereby providing a relatively simple measure to assess the child's progress toward speech-like vocalizations.

The purpose of the present study was to track progression in vocal development during the first 20 months of life using a hierarchical five-level instrument, the SAEVD-R. A strong association between age and SAEVD-R levels should be evident if the instrument is to be considered a valid measure of vocal development. Infants and toddlers at older ages would be expected to produce more complex vocalization types than younger infants. Furthermore, if the levels on the SAEVD-R are reliable, then judges should assign vocalizations to corresponding levels consistently.

METHOD

Participants

Participants for the present study were randomly selected from an archival pool of 51 children whose social communication behaviours were examined in the Stark et al. (1993) study and whose sound production skills had undergone preliminary, descriptive investigation in Stark et al. (1988). Because archived data were used, only brief details about the sample are provided here. Please refer to Stark et al. (1993) for more complete details about the participants, recording procedures, and social communication characteristics.

In the present study, 30 infants at six age ranges were studied with a mixed cross-sectional and longitudinal design. There were five infants in each of six age groups, 2–10 weeks or 0–

2 months, 12–22 weeks or 3–5 months, 26–36 weeks or 6–8 months, 40–55 weeks or 9–12 months, 56–62 weeks or 13–15 months, 72–88 weeks or 16–20 months of age. Because there are 52 weeks in a year, age in months was calculated by dividing the age in weeks by 4.25. For convenience, the age groups will be referenced by months from now on. It should also be noted that the upper age for the third group (9–13 months) was reduced by one month to 12 months to avoid the appearance of overlap with the fourth age group (13 – 15 months). Five infants in the 13–15 months age group were included that were not previously analysed by Stark et al. (1988).

Socioeconomic status (SES) and race were proportionally sampled by Stark et al. (1993). Information about these parameters was not available for the specific subsample included in the present study. No qualitative differences in early vocal development have, however, been demonstrated for SES (Oller, Eilers, Basinger, Steffens, & Urbano, 1995). Any quantitative differences were minimized because we attempted to obtain comparable numbers of utterances from each infant. Gender was balanced across all age groups except for the 6 – 8 month age group due to sampling difficulties (e.g., failure of hearing testing), as described in Stark et al (1993). Again, gender differences in qualitative aspects of early vocal development have not been demonstrated (e.g., Lynch, Steffens, Oller, & Buder, 1995a). All infants had normal hearing and developmental histories.

Data Collection

Infants were audio and video recorded following a standard protocol that included free play with parent and/or toys, a feeding period, and solitary play. The objective of these activities was to elicit a range of representative sound production behaviours. Each infant was recorded five times within the time span for his or her age group. Thus, an infant in the 0–2 month age group was first recorded close to 2 weeks of age and the fifth and last recording was made closer to 2 months of age. Audiotapes from the first and the last recording for each infant were analysed because it was expected that some progress on the SAEVD-R might be seen across the two sessions even within each age group. Session therefore, served as a longitudinal dimension and age group served as a cross-sectional dimension in the design of this study. Only audio recordings were analysed because the video recordings could provide observers with information about social communication and thereby influence judgments of vocalization type. Approximately 40–50 consecutive and clearly audible utterances were analysed for each session. It should be noted that the duration of the sessions might have varied somewhat in order to obtain 40–50 utterances. Our laboratory experiences indicate that a 20-minute representative sample is usually sufficient to elicit 40–50 utterances in typically developing babies.

SAEVD-R Description

The vocalization types defined in the original SAEVD (Nathani, Ertmer, & Stark, 2000) were derived from the infraphonological and developmental stage models of infant vocalizations proposed by Oller (1980) and Stark (1980). In the SAEVD-R, operational definitions of these vocalization types, including fixed signals, vegetative sounds, and protophones, were added, elaborated and/or consolidated based on findings from our pilot investigations on typically developing children and on children fitted with cochlear implants (Ertmer & Mellon, 2001; Ertmer et al., 2002; Nathani & Stark, 1996). For example, we originally assigned single productions of a vocalization type (e.g., V or a single vowel, Level 2) to a lower level than multiple productions of the same vocalization type (e.g., V2 or a series of vowels, Level 3) because Stark (1980) hypothesized that repeated productions of the same vocalization type might represent greater complexity than single vocalizations. Pilot observations revealed that infants produced single and multiple productions of a

vocalization type simultaneously. Consequently, single and multiple productions were scored at the same level (e.g., both V and V2 were assigned to Level 3).

Following these modifications, the SAEVD-R (see Appendix) consisted of operational definitions for 23 vocalization types. Each vocalization type was nested within five developmental levels according to the average age of expected emergence in typically developing children. Prior versions of the scale (Nathani & Stark, 1996; Stark et al., 1988) had as many as 31 vocalization types and seven levels. This revised version of the SAEVD was therefore, much easier to use than its predecessors. The 23 vocalization types on the SAEVD-R were defined using acoustic and articulatory characteristics as observed through auditory perceptual coding. Wide-band spectrograms were used to support auditory perceptual coding decisions as needed. Vocalization types and levels were developed to be mutually exclusive. In addition, the SAEVD-R definitions were developed to be comprehensive enough to capture all vocalizations produced by infants across the first 20 months of life. Levels were also intended to be hierarchical such that operational definitions within higher levels described more complex vocalizations than definitions within lower levels.

Levels of the SAEVD-R

The first level on the SAEVD-R, namely the Reflexive Level, includes primarily fussing, crying (CR), and vegetative (VEG) sounds. In addition to these non speech sounds, Level 1 in the SAEVD-R also included a protophone, namely quasi-resonant nuclei, produced either singly (Q) or in a series (Q2). All sounds under Level 1 are produced by infants 0–2 months of age (Oller, 1980; Stark, 1980). Please see Appendix for precise definitions of these vocalization types.

The second level, Control of Phonation, includes sounds expected to be produced by infants 1–4 months of age (Oller, 1980; Stark, 1980). As the term ‘control’ suggests, sounds at this level are no longer reflexive in nature, i.e., they are not produced in response to distress or physiological needs alone. The infant appears to be instead producing sounds that require some voluntary control of the vocal tract. Controlled and systematic production of specific vowel and consonant types, however, does not occur until later. Vocalization types within this level include fully resonant nuclei (F and F2), i.e., vowel-like sounds, or vocants, with full resonance but that cannot be transcribed as adult vowels; consonant-like sounds (closants such as raspberries or clicks) or primitive closant-vocant combinations (ev and ev2); and chuckle or laughter (CH). Glottal stops interrupting a series of vowel-like productions are also frequently seen at this stage. Except for chuckle or laughter, all other vocalization types in this category were considered protophones. Perception of voice quality changes, e.g., harshness, high pitch, were noted but not analyzed.

The third level, the Expansion phase, is comprised of protophones exclusively. Vocalizations from this level are expected to emerge between 3 and 8 months of age. Infants at this age produce mature vowels with full resonance that could potentially be transcribed by listeners as adult vowels (V, V2 in Appendix). Infants also appear to experiment with sound-making in this period (e.g., squeals [SQ], vowel glides [Vg]). Another important development that occurs at this level is the production of ‘marginal’ babbling (MB). As defined by Oller (1980, 2000), marginal syllables contain both consonant and vowel-like elements but the formant transition between the consonant and the vowel is slower than expected in mature syllables.

The fourth level is the Basic Canonical Syllables level. Again, all vocalizations at this level represent protophones. This level has been given special status in the literature because its emergence is striking even to naïve observers and it represents a significant advance toward

the production of mature, adult speech (Oller & Lynch, 1992). In addition, all typically developing infants are expected to produce canonical syllables some time between 5–10 months of age (Oller & Eilers, 1988). Canonical syllables refer to consonant-vowel combinations with adult-like formant transitions between the consonant and the vowel. In addition, whispered vocalizations were included at this level because pilot data indicated emergence at this level rather than the Expansion level.

Protophones considered as ‘Advanced Forms’ on the fifth level of the SAEVD-R occur around 9–18 months of age and either just precede or coincide with the onset of word production. Vocalizations at this level are canonical but in addition, represent considerable articulatory and phonatory complexity, e.g., closed syllables (CMPX), and imply more refined control over vocal tract parameters than in any previous level (see Appendix for specific vocalization types).

Procedures for Classifying Vocalizations

Infant vocalizations were analysed in the following way. First, audio tracks of the videorecorded sessions were digitized and separated into utterances by Stark et al. (1993). Researchers have used either breath-group definitions or time-based units to separate utterances. Stark et al. used a two-second rule to parse the utterances (i.e., a single vocalization or a series of vocalizations separated from all others by 2 sec was considered one utterance). Individual vocalizations within the digitized utterances were then categorized using the vocalization types of the SAEVD-R.

Four students in Communication Disorders were trained to categorize the data. Instruction on infant vocal development was provided weekly for 6–8 weeks. Audio examples were used to illustrate SAEVD-R vocalization types during training. Waveforms and spectrograms generated using the CSpeech SP (Milenkovic & Read, 1997) software were available during training and data analysis to supplement perceptual analysis whenever auditory judgments were ambiguous and spectrographic evidence was salient. Adequacy of training was assessed by comparing trainee judgments with those of experienced judges (SN & DE) on 10 training files containing approximately 400 utterances. When agreement between naïve and expert judgments reached 80%, the naïve judges were considered ‘trained’ and began actual data analysis.

Judges were blind to the age of the infant when categorizing the data in order to avoid listener bias toward associating specific vocalization types with age. Twenty percent of all vocalization types were reviewed by an experienced judge on a weekly basis to ensure that listeners used the definitions as intended and to trouble-shoot any classification difficulties. Definitions were clarified, as needed, and any ambiguous judgments were resolved. For example, the definition of cry (CR) was expanded to include procedures to follow when dealing with fussy utterances that were not full-blown cries. Thus, judges felt that the cry-like elements within such utterances were merely fussy or that these utterances also had enough non-cry elements that it could not be classified as “Cry”. It was decided that such fussy utterances would be classified according to the vocalization types contained within that utterance (e.g., fussy CV syllables within would be categorized at Level 4) and ‘fussiness’ would be noted (see Nathani & Oller, 2001, for a discussion of difficulties in coding these utterances).

Procedures for Assigning Level of Vocal Development

An utterance could contain more than one vocalization type because it could be made up of several vocalizations. For example, an utterance might contain a squeal followed by a canonical syllable. Such an utterance would therefore, be classified as SQ, CV. Following

this preliminary classification of vocalization types, levels were assigned to the entire utterance. The level of an utterance was equal to the highest level associated with an individual vocalization type within the utterance. For example, in the previous utterance, SQ would be classified at Level 3 and CV would be at Level 4. So, the entire utterance would receive a Level score of 4. Although cry (CR), laughter (CH), and vegetative sounds (VEG), were noted, these three vocalization types were not included when calculating level of an utterance because these vocalizations did not represent protophones.

Level ratios for each infant were then obtained by dividing the number of utterances at each level with the total number of utterances produced by that infant. The highest level that attained a ratio of 0.1 or greater was considered the infant's most advanced level of production skill. This procedure was a modification of the one previously used by Oller and Eilers (1988) and Nathani and Stark (1996). In those two studies, when a vocal behaviour (e.g., canonical babbling) reached a proportion of 0.2 of the entire vocal repertoire of the infant, it was considered well established. This 0.2 criterion was set because Oller and Eilers found that at this rate of production, parents were able to reliably identify the onset of canonical babbling, as confirmed by laboratory observation. In the present study, a lower 10% or 0.1 emergence criterion was used because it was expected that trained listeners would identify the emergence of new vocalization types earlier than parents. In addition, we had several more vocalization types and levels than the Oller and Eilers study. Correspondingly, we used a relatively lower threshold to capture the emergence of a vocalization type.

Reliability

Three judges categorized the actual data, namely 2581 utterances and 3292 syllables. The fourth judge classified utterances from a randomly selected subset (10%) of all utterances to determine reliability of the primary judges' utterance level classification of vocalization types. Cohen's Kappa (Cohen, 1960) for interjudge agreement on SAEVD-R level assignments was found to be 0.64. Kappas between 0.6 and 0.74 are considered 'good' agreement between independent judges (Fleiss, 1981).

RESULTS

Statistical Analysis

An average of 43 (SD=4.05) utterances per session were analysed for each infant. The distribution of non speech-like and speech-like utterances (or protophones) across age is shown in figure 1 (The age group of 13–15 months is excluded from this analysis because the frequency of utterances containing predominantly non speech-like vocalizations was inadvertently not tabulated for this age group and judges were no longer available to recode the utterances. Based on known developmental trends, no significant additional information was expected for this analysis if data from the 13–15 month age group had been available). Data were averaged across the two sessions for each age group in figure 1. There was a steady increase in the proportion of utterances that contained predominantly speech-like vocalizations, or protophones, with age from 65% in the youngest age group (0–2 months) to 94% utterances in the oldest age group (16–20 months). The reduced proportion of protophones in the youngest age group occurred because many utterances produced by infants in this age group were cries.

A simple one-way ANOVA with age group serving as a between-subjects factor and number of protophones serving as the dependent variable was then conducted. Data were averaged across the two sessions for each infant. The number of protophones was found to increase significantly with age, $F(4, 20) = 10.9$, $p < 0.01$. Conversely, the number of non speech-like

utterances decreased with age. Post-hoc Tukey tests revealed that the youngest age group of 0–2 months produced significantly fewer utterances with protophones than all other age groups; infants 3–5 months of age produced significantly fewer protophones than infants 9–20 months of age; and infants from 6–20 months of age did not significantly differ from each other in the number of utterances produced with protophones.

Utterances containing protophones were further analysed using a repeated measures multivariate analysis of variance (MANOVA). Age group (0–2, 3–5, 6–8, 9–12, 13–15, and 16–20 months) was considered a between-subjects factor. Session (first and last within the infant's age range) and level of utterance (1, 2, 3, 4, and 5) were included as within-subjects variables. The number of utterances at each combination of age, session, and level of utterance was the dependent variable. Analyses were also run using proportion of utterances at each combination of age, session, and level of utterance as the dependent variable to determine whether differential frequencies of speech-like utterances in the various age groups would influence results. Results for both frequencies and proportions were similar. Results below are therefore, only reported for analyses using frequency of utterances as the dependent variable.

Because the younger age groups (0–2, 3–5, and 6–8 months of age) rarely produced any Level 5 utterances, there was essentially no variance on this level for these age groups. Consequently, we ran two sets of analyses. One set of analyses included all age groups but only included data from the first four levels. The second set of analyses included data from all levels of the SAEVD-R but only included the oldest three age groups, 9–12, 13–15, and 16–20 months of age.

Results using all age groups and the first four levels on the SAEVD-R revealed significant interaction effects for Level X Session, $F(3, 22) = 3.304$, $p < 0.05$, $\eta^2_p = 0.31$, and Level X Age, $F(15, 61) = 6.793$, $p < 0.05$, $\eta^2_p = 0.59$, and significant main effects for Level, $F(3, 22) = 49.099$, $p < 0.05$, $\eta^2_p = 0.87$, and Age, $F(5, 24) = 9.265$, $p < 0.05$, $\eta^2_p = 0.66$. These interaction effects indicate that the level of speech-like vocalizations changed with age and with session. Other interaction and main effects were not significant.

A main effect analysis was conducted to determine the source of the Level X Age interaction. It revealed that Level was significant at four of the six age groups ($p < 0.05$). Results did not reach significance at 3–5 months and at 13–15 months. Multiple comparisons of the four age groups (0–2, 6–8, 9–12, and 16–20 months of age) showed significant differences in level using the Bonferroni procedure. Infants in the 0–2 and 6–8 months age group primarily produced utterances from Levels 1–3 whereas infants in the 9–12 and 16–20 months age group produced utterances primarily from Levels 3 and 4. The 3–5 month age group may not have reached significance because a single infant in this group produced an unusually large proportion of canonical syllables during the last session of the 3–5-month age range (61%). Level 3 vocalizations predominated in the 13–15 month age group and thereby may have obscured potential differences with other age groups for the remaining levels.

A main effect analysis was also carried out to determine the source of the Level X Session interaction. Level 4 was the only level to show a significant difference across sessions, $F(1, 29) = 4.84$, $p < 0.05$, $\eta^2_p = 0.143$ suggesting that Levels 1, 2, and 3 are stable across small intervals of time of (i.e., two to three months) whereas Level 4 is sensitive to small differences in age. Thus, sessions recorded two to three months later in life for an infant had a greater number of Level 4 utterances than sessions recorded at the beginning of that time period for the same infant.

Results of the MANOVA using the three older age groups (9–12, 13–15, and 16–20 months of age) and all five levels of the SAEVD-R revealed a significant interaction effect for Level X Age, $F(8, 18) = 3.648$, $p < 0.05$, $\eta^2_p = 0.618$, and significant main effects for Level, $F(4, 9) = 66.167$, $p < 0.05$, $\eta^2_p = 0.96$, and Age, $F(2, 12) = 5.551$, $p < 0.05$, $\eta^2_p = 0.48$. No other main or interaction effects were significant. The Level X Age interaction is highly relevant because it indicates that infants from 9–20 months of age attained progressively higher levels on the SAEVD-R with age. A univariate main effect analysis of the Level X Age interaction showed that Levels 3 and 5 significantly differed across age groups, $p < 0.05$. Level 4 approached significance, ($p = 0.05$. Levels) but Levels 1 and 2 were not significantly different across the three age groups. Multiple comparisons using the Bonferroni statistic revealed that infants 9–12 months and 13–15 months of age primarily produced Level 3 and few Level 5 utterances, whereas infants 16–20 months of age produced few Level 3 and significantly greater Level 5 utterances.

Descriptive Analysis

Figure 2 displays the proportion of protophones at each level across age groups. Data were averaged across the two sessions within each age group for ease of display. As seen in this figure, Levels 1 and 2 decreased across age whereas Levels 4 and 5 increased across age. Level 3 remained constant and was the most predominant form of vocalizations across age except for the oldest group. Thus, protophones increased in complexity with age. When non speech-like vocalizations were included, the only age group that changed, as would be expected, was the youngest age group of 0–2 months. Instead of Level 3 being the predominant form of vocalizations, Level 1 comprised the majority of vocalizations at that age. Recall that Level 1 contains mainly cries and vegetative sounds.

Data from the two sessions within each age group also show that the SAEVD-R level that accounted for at least 0.1 or 10% of all utterances increased with age. Up to 8 months of age, the highest level that met this criterion was level 3; from 9 – 15 months it was level 4; and from 16 –20 months it was level 5. Thus, the 10% criterion for estimating the child's level of vocal development also showed progressive increases in level with age.

Frequencies of individual protophones were then examined. The most frequent protophone produced by each age group is shown in table 1. Data were averaged across the two sessions within each age group. Infants from 0–8 months of age most frequently produced quasi-resonant nuclei; those from 9–15 months of age, most frequently produced vowels; and those from 16–20 months, most frequently produced canonical syllables. These trends for the individual vocalization type follow those for levels, i.e., they become more complex with age. Whisper, $\epsilon\epsilon\epsilon$, jargon, CV-C, marginal and canonical babbling series were rarely produced; each type constituted less than 5% of all vocalizations in the study.

DISCUSSION

The primary purpose of the study was to identify prelinguistic vocalization types and to track developmental changes in early vocal production with chronological age. Our results indicated that trained judges reliably classified prelinguistic vocalization types using SAEVD-R operational definitions. As evidenced in Figure 2, the following age-related changes in vocal production were noted: Levels 1 (Reflexive) and 2 (Control of Phonation) decreased with age. In contrast, Level 4 (Basic Canonical Syllables) was rarely produced by infants in the first nine months of life. Older infants from 9–20 months of age primarily produced this level. Level 4 was also found to be sensitive to small differences in age. Not only did infants show progress on this level across age groups, but infants in the same age group also showed progress within a two- to three-month age span on this level. Only infants from 16–20 months of age produced Level 5 (Advanced Forms) vocalizations in any

significant quantities. Proportions of these Level 5 vocalizations jumped from 0–10% across 0–15 months of age to approximately 20% from 16–20 months of age. These vocalizations might, therefore, represent the pinnacle of prelinguistic vocal development and, as such, signal the infant's readiness for transition into consistent production of meaningful speech. Further discussion of the results will be presented according to developmental order.

The frequency of non-speech vocalization types, such as cries and vegetative sounds, reduced significantly with age whereas protophones increased in frequency (approximately 66% at 0–2 months of age to 99% by 16–20 months of age, see Figure 1). Often, classification schemes (e.g., Oller & Lynch, 1992) exclude these sound types even before coding an utterance because the sounds themselves are not precursors to speech. As demonstrated in the present study, this exclusion can inadvertently result in the neglect of an important dimension of human vocal development. Decreases in the proportions of non speech-like utterances to protophones might, therefore, serve as a simple benchmark to evaluate vocal progress, especially in the earliest stages of infants' vocal development. Impressively, because two-thirds of all vocalizations produced by the youngest infants were speech-like, it suggests that even infants in the first couple months of life possess basic building blocks for later speech development.

Another striking finding was that vocalizations from Level 3 (Expansion), primarily vowels, remained predominant across all ages until 16–20 months. Studies by Mitchell and Kent (1990) on 7–11 month old infants and Kent and Bauer (1985) on 13-month old infants confirm the predominance of vowels at these ages. Because vowels are produced for such a long period, they may be overlooked—especially when the focus of an assessment is on emergence of new vocalization types, such as babbling. Examination of vowel productions may, in fact, be particularly important for some special populations. Studies of deaf children (e.g., Lynch, 1989; Ertmer et al., 2002) show that attainment of vowels is often difficult for these children and that increases in vowel diversity signal early benefits of cochlear implantation. Acquisition of an abundance of vowel-like vocalizations might therefore indicate progress in early speech development and be a useful intervention goal for children experiencing speech delays.

Vowels also showed a spike in production at the age of 13–15 months almost doubling in frequency from approximately 20% below 12 months of age to 40% of all protophones produced at 13–15 months. Production of vowels again tapered off in the oldest age group of 16–20 months to 15%. One explanation for this spike may be that vowels, already a well-established aspect of prelinguistic vocalizations prior to 13 months of age, were recruited for communicative purposes just before the onset of first words resulting in an unusually high frequency of occurrence. This explanation is underscored by findings of McCune et al. (1996) that 'grunts' are often the first vocalizations to be used communicatively. It may, therefore, be useful to monitor increases in vowel frequency as a potential cue to the onset of word production in future investigations.

Canonical syllables also increased with age such that they constituted 10–20% of the speech-like repertoire of most infants from 9–20 months of age. Two infants, one in the 9–12 month and one in the 13–15 month age groups, did not achieve the 0.1 criterion production of canonical syllables. Examination of the data from Stark et al. (1988) reveals that the 9–12 month old infant was indeed identified as having reading difficulties later in life. Thus, the SAEVD-R appears to show promise in identifying delays in vocal development that may impact language-based skills. The consequences of observed delays in vocal development for the 13–15 month old infant are essentially unknown.

Conversely, with the above-mentioned exception of one infant at 5 months of age, infants under 8 months did not produce canonical syllables in any measurable quantities. Substantial increases in the proportion of Level 4 and 5 vocalizations occurred after 9 months and confirm the observation made by several researchers that the abrupt production of canonical syllables is a key indicator of advancement in prelinguistic vocal development (Oller, 1980, 2000; Stark, 1980). Oller, however, indicated that canonical syllables are produced by typically developing infants sometime between 5–10 months of age. Although still within the ranges proposed by Oller, the limited occurrence of canonical babbling in our 6–8 month age group suggests that this event may be consolidated somewhat later than first thought. In fact, Stark (1980) and Roug et al. (1989) reported the emergence of canonical syllables around 8 months of age. Again, such differences may reflect individual variability, the small number of infants studied, differences in operational definitions, and/or different criterion for emergence of canonical syllables across studies.

Finally, it should be noted that infants recorded in the present study were English-learning infants. Broad protophone categories within Levels 1–4 are fairly universal across infants from different language backgrounds (Koopmans-van Beinum and van der Stelt, 1986; Roug et al., 1989). Any language-related differences are typically found in the phonetic inventories of infants from different language backgrounds rather than the broad protophone categories included on the SAEVD-R. Some caution is, however, needed when generalizing results regarding Level 5 vocalization types to infants from non English-learning backgrounds. Presumably, infants at this level are producing words from their ambient language and vocalization types at this level may therefore, be more susceptible to ambient language influences. For example, Vihman (1993) noted that closed syllables predominated in the vocal repertoires of English-learning infants by the time infants acquired a 25-word vocabulary but this was not the case for infants learning other languages such as French, Swedish, and Japanese. Investigation of the SAEVD-R with infants from different language backgrounds is therefore, warranted.

Clinical Implications

Recent advances in technology (i.e. newborn hearing screening) and mandated early intervention programmes have increased the number of infants being referred for communication assessments and amplified the need for appropriate assessment tools. Although the SAEVD-R provides a comprehensive way to address this need, the protocol used in this study--especially the extensive training and supplemental use of spectrographic information--is likely to be impractical in some clinical settings. A simpler tool with fewer vocalization categories and levels would increase the clinical usefulness of the SAEVD-R.

Ertmer and colleagues (Ertmer & Mellon, 2001; Ertmer et al, 2002) have used a 3-level classification system to track vocal development in deaf children who receive cochlear implants. Their approach groups SAEVD-R Levels 1–3 under the heading of precanonical vocalizations; Level 4 as canonical vocalizations; and Level 5 as Advanced Forms to simplify clinical decision-making. This consolidation of Levels 1–3 into precanonical vocalizations and separation from Levels 4 and 5 is supported by findings from the present study. Infants below nine months of age produced comparable amounts of speech-like vocalizations across Levels 1–3. The use of Level 4 and 5 vocalizations varied across the age groups. Clinicians who wish to familiarize themselves with this system can access audio recordings of precanonical, basic canonical syllables, and advanced forms vocalizations at the Web site vocaldevelopment.com (Ertmer & Galster, 2001).

Other modifications might also be considered based on the data in Figure 2. For example, Levels 1 and 2 could be collapsed because these did not provide any differentiation across ages. Vocalizations from Levels 3, 4, and 5 could be classified separately because Level 3

was the predominant form of vocalizations from 3–15 months of age and because Levels 4 and 5 provided differentiation in vocal maturity across ages. Age-of-emergence estimates on the SAEVD-R might also need revision because Level 3 vocalizations emerged earlier than expected at 0–2 months of age whereas Level 4 and 5 vocalizations emerged later than expected at 9–12 months and 16–20 months. Application of the SAEVD-R to a large and diverse sample of typically developing and disordered infants is, however, needed to verify these observations.

These modified versions of the SAEVD-R may be useful in assessing early speech progress in children at risk for developmental delays due to cleft lip and palate, Down syndrome, Pervasive Developmental Disorders, and other disorders that impact oral communication. For example, in some children with cleft lip and palate, canonical syllable production may be achieved prior to surgery (Chapman, Hardin-Jones, Schulte, & Halter, 2001). In these children, similar to that observed in infants with cochlear implants (Ertmer et al., 2002), documentation of Level 5 vocalizations might be more useful to assess progress with surgery and readiness for speech rather than increases in canonical syllable production alone. Similarly, not all children with Down Syndrome are delayed in their production of canonical syllables (Lynch et al., 1995b). However, pronounced language delays at later ages in these children are well documented. It may be that these children do not progress rapidly beyond the production of basic canonical syllables and Level 5 may be a more useful starting point for assessment and intervention.

Conclusions

The present investigation successfully examined one aspect of prelinguistic communication in infants and toddlers, namely vocal production. Changes in vocal production skills with age were reliably detected using the vocalization types and levels contained in the SAEVD-R. Future investigations of the relationships between vocal production and social communication using the Stark et al. (1993) data on this same subset of children are also planned. It is hoped that these comparisons will shed more light on the mechanisms responsible for the remarkable progress children make in their communication abilities during the first 20 months of life.

Acknowledgments

We gratefully acknowledge the invaluable assistance of Kelli Howard, Melissa Price, Jessica Weed, and Claire Johnson in data analysis. We would also like to thank Kim Oller for his comments, Veena Dhankher and Steve Olejnik for their assistance with statistical analysis, and Melanie Yahner for her assistance with manuscript preparation. Finally, we dedicate this article in memoriam to Dr. Rachel E. Stark, our mentor and the inspiration behind this project. Portions of this work were presented at the 2002 American Speech-Language-Hearing Association Conference in Atlanta, GA.

References

- Bayley, N. The Bayley Scales of Infant Development. 2. San Antonio, TX: The Psychological Corporation; 1993.
- Chapman KL, Hardin-Jones M, Schulte J, Halter KA. Vocal development of 9-month-old babies with cleft palate. *Journal of Speech, Language, and Hearing Research*. 2001; 44:1268–1283.
- Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960; 20:37–46.
- Coplan, J. Early Language Milestone Scale. Tulsa, OK: Modern Education Corporation; 1986.
- Ertmer, DJ.; Galster, J. Vocaldevelopment.com. 2001. [On-line]. Available: <http://www.vocaldevelopment.com>
- Ertmer DJ, Mellon JE. Beginning to talk at 20 months: Early vocal development in a young cochlear implant recipient. *Journal of Speech, Language, and Hearing Research*. 2001; 44:192–206.

- Ertmer DJ, Young N, Grohne K, Mellon JA, Johnson C, Corbett K, et al. Vocal development in young children with cochlear implants: Profiles and implications for intervention. *Language, Speech, and Hearing Services in Schools*. 2002; 33:184–195.
- Fleiss, JL. *Statistical methods for rates and proportions*. New York: Wiley; 1981.
- Goldstein MH, King AP, West MJ. Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*. 2003; 100:8030–35.
- Jacobson, R. *Child language, aphasia, and phonological universals*. The Hague: Mouton; 1968.
- Kent RD. Psychobiology of speech development: Coemergence of language and a movement system. *American Journal of Physiology*. 1984; 246:R888–R894. [PubMed: 6742163]
- Kent RD, Bauer HR. Vocalizations of one-year olds. *Journal of Child Language*. 1985; 12:491–526.
- 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*. Basingstoke, Hampshire: Macmillan Press; 1986. p. 37–50.
- Lynch MP, Oller DK, Steffens M. Development of speech-like vocalizations in a child with congenital absence of cochleas: The case of total deafness. *Applied Psycholinguistics*. 1989; 10:315–333.
- Lynch MP, Oller DK, Steffens ML, Buder EH. Phrasing in prelinguistic vocalizations. *Developmental Psychology*. 1995a; 28(1):3–25.
- Lynch MP, Oller DK, Steffens ML, Levine SL, Basinger DL, Umbel V. Onset of speech-like vocalizations in infants with Down Syndrome. *American Journal on Mental Retardation*. 1995b; 100(1):68–86. [PubMed: 7546639]
- McCune L, Vihman MM, Roug-Hellichius L, Borden, Delery DB, Gogate L. Grunt communication in human infants. *Journal of Comparative Psychology*. 1996; 110:27–37. [PubMed: 8851550]
- Milenkovic, PH.; Read, C. *CSpeech* (Version SP) [Computer software]. Department of Electrical Engineering, University of Wisconsin; Madison: 1997.
- Mitchell PR, Kent R. Phonetic variation in multisyllable babbling. *Journal of Child and Language*. 1990; 17:247–265.
- Nathani, S.; Ertmer, DJ.; Stark, RE. Unpublished manuscript. Purdue University; West Lafayette, IN: 2000. Stark Assessment of Early Vocal Development.
- Nathani S, Oller DK. Beyond ba-ba and gu-gu: Challenges and potential strategies in coding infant vocalizations. *Behavior Research, Methods, Instrumentation, and Computers*. 2001; 33(3):321–330.
- Nathani S, Stark RE. Can conditioning procedures yield representative infant vocalizations in the laboratory? *First Language*. 1996; 16(3):365–368.
- Oller, DK. The emergence of the sounds of speech in infancy. In: Yeni-Komshian, G.; Kavanagh, J.; Ferguson, CA., editors. *Child phonology, Volume 1, Production*. New York: Academic Press; 1980. p. 93–112.
- Oller, DK. The emergence of the speech capacity. Mahwah, NJ: Erlbaum; 2000.
- Oller, DK.; Buder, EH. Vocal contextual freedom in the first months of life. Symposium conducted at the International Conference on Infant Studies; Chicago, IL. 2004 May.
- Oller DK, Eilers RE. The role of audition in infant babbling. *Child Development*. 1988; 59:441–449. [PubMed: 3359864]
- Oller DK, Eilers RE, Basinger D, Steffens ML, Urbano R. Extreme poverty and the development of precursors to the speech capacity. *First Language*. 1995; 15:167–187.
- Oller, DK.; Lynch, MP. Infant vocalizations and innovations in infraphonology. In: Ferguson, CA.; Menn, L.; Stoel-Gammon, C., editors. *Phonological development: Models, research, implications*. Timonium, MD: York Press; 1992. p. 509–536.
- Rossetti, L. *The Rossetti Infant-Toddler Language Scale: A measure of communication and interaction*. East Moline, IL: LinguiSystems; 1990.
- Roug L, Landberg I, Lundberg LJ. Phonetic development in early infancy: A study of four Swedish children during the first 18 months of life. *Journal of Child Language*. 1989; 16:19–40. [PubMed: 2925811]
- Smith BL, Brown-Sweeney S, Stoel-Gammon C. A quantitative analysis of reduplicated and variegated babbling. *First Language*. 1989; 9:175–189.

- Stark, RE. Stages of speech development in the first year of life. In: Yeni-Komshian, G.; Kavanagh, J.; Ferguson, CA., editors. *Child phonology, Volume 1, Production*. New York: Academic Press; 1980. p. 73-90.
- Stark, RE. Prespeech segmental feature development. In: Fletcher, P.; Garman, M., editors. *Language Acquisition. 2*. Cambridge, NY: Cambridge University Press; 1986. p. 149-173.
- Stark, RE.; Ansel, BM.; Bond, J. Are prelinguistic abilities predictive of learning disability? A follow-up study. In: Masland, RL.; Masland, MW., editors. *Preschool prevention of reading failure*. Parkton, MD: York Press; 1988. p. 3-18.
- Stark RE, Bernstein LE, Demorest ME. Vocal communication in the first 18 months of life. *Journal of Speech and Hearing Research*. 1993; 36:548–558. [PubMed: 8331912]
- Vihman MM. Variable paths to early word production. *Journal of Phonetics*. 1993; 21:61–82.
- Vihman, MM. *Phonological development: The origins of language in the child*. Cambridge, MA: Blackwell Publishers; 1996.
- Wetherby, AM.; Prizant, BM. *Communication and Behavioural Scales Manual*. Chicago, IL: The Riverside Publishing Company; 1993. (Normed ed.)
- Zimmerman, IL.; Steiner, VG.; Pond, RE. *The Preschool Language Scale. 4*. San Antonio, TX: The Psychological Corporation; 2002.
- Zimmerman-Phillips, S.; Osberger, M. *Meaningful Auditory Integration Scale (IT-MAIS) for Infants and Toddlers*. Sylmar, CA: Advanced Bionics Corporation; 1997.
- Zlatin, M. Final Report, Project No 3-4014, NE-G-00-3-0077. National Institutes of Health Research Grants; 1975. Preliminary descriptive model of infant vocalization during the first 24 weeks: Primitive syllabification and phonetic exploratory behaviour.

APPENDIX. The Stark Assessment of Early Vocal Development-Revised (SAEVD-R) Copyright 2003 by Nathani, Ertmer, & Stark

The SAEVD-R employs 23 vocalization types to characterize and classify the vocalizations of infants and toddlers. The vocalization types are grouped into five ascending developmental levels. Operational definitions for the vocalization types are presented below. Vocalization types, levels, and expected ages of emergence are partly derived from Oller (1980), Stark (1980), and Stark, Ansel, and Bond (1988).

Level 1. Reflexive (0–2 months)

- VEG Vegetative sounds, e.g., burp, cough, sneeze, etc.
- CR Sustained crying/fussing or series of brief discomfort sounds. Ingressive sounds, squeals, and vegetative sounds in cry are not classified as separate items. Cries that contain syllables (e.g., mamama) are classified as CR and the term 'Fussy Syllables' is noted as a comment. Conversely, utterances that are fussy (not full-blown cries) or utterances that contain non-fussy elements + cry are assigned applicable vocalization types (e.g., CVCV, V) and 'fussiness' and/or 'cry' are noted.
- Q Quasi-Resonant Nuclei. Faint, low-pitched grunt-like sounds with muffled resonance. Characterized by a lack of energy above 2000Hz. If there is energy across mid and higher frequencies, the vocalization may be classified as Q if the sound is brief (<100 ms). Q sounds cannot be transcribed as an adult vowel.
- Q2 Two or more Qs in a series or row.

Level 2. Control of Phonation (1–4 months)

- F Fully-Resonant Nuclei. Vowel-like sounds that are longer than Qs but cannot be readily transcribed as adult vowels. These vocalizations have energy across a wide range of frequencies (i.e., not restricted to low frequencies like Q). They may have poor vocal quality (harshness, high pitch, etc.). If glottal stops or [h] interrupt a Q or F, they are counted as a single syllable and glottal stops are noted.
- F2 Two or more Fs in a series or row.
- ev Vocalization in which a vocant (vowel-like segment) or an F are combined with a superimposed closant (consonant-like segment). Also includes an isolated closant (e.g., 'raspberry', trill, click) or an isolated consonant (e.g. m, n, sh). Glottal stops and [h] are not considered closants.
- ev2 Two closant-vocant combinations, or two or more closants in a series.

- CH At least two brief chuckles or sustained laughter. Frequently, a [h]-like closant is perceived before the vowel. Ingressive sounds during laughter are not classified separately.

Level 3: Expansion (3–8 months)

- V An isolated vowel. Vowels can be distinguished from Q/F because Vs are longer and more fully resonant than Qs and are of better quality and more easily recognized as vowels than Fs. They may contain some harshness, high pitch, etc. but are transcribable as adult vowels. Note any aberrant voice quality features.
- V2 Two or more vowels in a series or row.
- Vg Vowel Glide. Vocant in which a change in vowel quality is present. No audible gap is present between the two segments. No closure can be identified, (e.g., [pa], [da]). The formant transition is characteristically slow; greater than 200 ms. If formant transition duration is less than or equal to 200 ms, classify as CV if closure perceived (e.g., w, j), or judge as diphthong if no closure is perceived (e.g. oi, ai).
- IN Ingressive Sound. Single long (>200 ms) ingressive sound or series of short ingressive sounds.
- SQ Squeal. High-pitched sound or series of squeals.
- MB Marginal babbling. Series of closant and vocant segments or series of Vgs. Irrespective of the nature of the closant/vocant, the key characteristic of MB is that formant transitions between the closant and vocant are prolonged. Therefore, even sequences of real consonants and vowels would be considered MB if they had long (> 120 ms) formant transitions. Elements in the sequence need not always contain a closant and vocant; occasional isolated vocants and closants might also be present. Well-formed, rapid glide and other semivowel sequences (e.g., wa, ja) would not be included under MB; they would be included under CV.

Level 4: Basic Canonical Syllables (5–10 months)

- CV Single consonant-vowel syllable. Does not include syllables with /h/ or a glottal stop as a consonant.
- CB Repeated canonical syllable production. More than two CV syllables in sequence are required for this category. Because the consonants and vowels in the sequence can be same or different, this category includes reduplicated babbling (repeated productions of the same consonant-vowel sequence) and nonreduplicated babbling (sequence of different consonant-vowel combinations). If squeals, ingressive sounds, etc. occur during CB, corresponding vocalization types are merely noted.
- WH Whispered productions. V1, V2, Vg, MB, CB, or CV vocalizations produced without voice
- CV-C A consonant-vowel combination followed by an isolated consonant. A silent gap between CV and C should be observed.
- CVCV Disyllables. Two adjacent CV syllables or series of two CV syllables with an audible gap separating the CVs.

Level 5: Advanced Forms (9–18 months)

- CMPX Complex syllables. 1) Single syllable types other than CV (e.g., VC, CCV, CCVC, etc.), or 2) Complex Disyllables (e.g., VCV, VCVC), or 3) Multisyllabic strings with complex syllables and without variable stress or intonation patterns (e.g., VCVCV, VCVCVCV), or 4) Multisyllabic utterances with varied stress and/or intonation patterns in which the consonants and vowels remains unchanging. The latter are designated as Canonical Jargon (CBJN).
- JN Jargon. A series of syllables with at least two different Cs and Vs with a changing stress and/or varied intonation pattern within the series. The series must contain more than 2 syllables.
- DIP Diphthongs, e.g. /oi/, /ai/, /au/, or other forms with rapid formant transitions. Diphthong is characterized by formant transition that is less than 200 ms and overall syllable duration of less than 500 ms.

Figure 1.

Mean proportion (and standard error) of speech-like and non speech-like utterances across age. Data are averaged across the two sessions within each age group.

Figure 2.
Mean proportion (and standard error) of speech-like utterances across each level and age group. Data are averaged across the two sessions within each age group.

Table 1

The most frequent protophone produced by each age group. Data are averaged across the two sessions for each age group.

Age Group (in months)	Most frequent protophone	Proportion of occurrence
0–2	Q	0.24
3–5	Q	0.17
6–8	Q	0.24
9–12	V	0.23
13–15	V	0.38
16–20	CV	0.25