Speech-like vocalizations in infancy: an evaluation of potential risk factors*

D. KIMBROUGH OLLER, REBECCA E. EILERS, MICHELE L. STEFFENS

University of Miami

MICHAEL P. LYNCH

Purdue University

AND

RICHARD URBANO

University of Miami

ABSTRACT

This work reports longitudinal evaluation of the speech-like vocal development of infants born at risk due to prematurity or low socio-economic status (SES) and infants not subject to such risk. Twenty infants were preterm (10 of low SES) and 33 were full term (16 of low SES), and all were studied from 0;4 through 1;6. The study provides the indication that at-risk infants are not generally delayed in the ability to produce well-formed speech-like sounds as indicated in taperecorded vocal samples. At the same time, premature infants show a tendency to produce well-formed syllables less consistently than full terms after the point at which parents and laboratory personnel note the onset of the canonical babbling stage (the point after which well-formed syllables are well established in the infant vocal repertoires). Further, even though low SES infants produce well-formed speech-like structures on schedule, they show a reliably lower tendency to vocalize in general, as reflected by fewer utterances per minute in recorded samples.

INTRODUCTION

The robustness of speech-like vocal development in humans

Vocal development through the first year of life proceeds along a course that appears strongly influenced by the biological endowment of the species. This conclusion is suggested by the consistency in infant vocal development across

[*] This work was supported by NIH/NIDCD grant No. R01 RC00484 to D. Kimbrough Oller. Address for correspondence: D. Kimbrough Oller, Mailman Center for Child Development, P.O. Box 016820, Miami, FL 33101 USA.
OLLER ET AL.

different environmental and developmental conditions. For example, the
onset of the canonical babbling stage (i.e. the stage wherein infants show
controlled production of well-formed syllables or syllable sequences such as
\textit{baba} or \textit{dada}) appears to occur not later than ten months of age in infants who
are (a) either full term or preterm (Eilers, Oller, Levine, Basinger, Lynch &
Urbano, 1993); (b) of middle or low socioeconomic status (Eilers et al.
1993), and (c) from Spanish or English speaking, or bilingual families (Oiler,
Lewedag, Umbel & Basinger, 1992). Even Down syndrome infants usually
begin canonical babbling within the typical period (Smith & Oiler, 1981;
Lynch, Oller, Eilers & Basinger, 1990). The robustness of canonical babbling
is also reflected in the fact that infants tracheostomized at birth, and
consequently unable to vocalize until decannulation during the second year
of life or later, commonly show canonical babbling, and/or speech utilizing
canonical syllables within a few weeks of decannulation (Locke & Pearson,

Although considerable stability of infant vocal development in the face of
developmental and environmental variation is by now well-recognized, it is
important to document the conditions that do affect the topography of infant
speech-like sounds. This importance is both theoretical because our under-
standing of the nature of the human linguistic endowment may be enhanced,
and practical because screening for developmental disorders through evalu-
ation of vocalizations requires broad awareness of conditions under which
vocal development may vary.

Among the most common risk factors generally associated with abnormal
development are premature birth and low socioeconomic status (SES), and
both provide special opportunities to assess the robustness of infant vocal
development. Although recent work in our laboratories has suggested that
there is no delay in the onset of canonical babbling in either low SES or
preterm infants, there has emerged a surprising trend in which preterms (at
ages corrected for prematurity) actually appear to show somewhat earlier
onset than full term infants in two motoric developments, rhythmic hand
banging and canonical babbling (Eilers et al. 1993). The tentative in-
terpretation of these findings depends on recognition of the fact that preterm
infants have a different history of experience from full terms. Being born
early, they witness more events ‘out in the world’, more sound, more sight,
more self-generated movement than do full term infants matched for
corrected age. Preterm infants may profit from their precocious sensory and
motoric experience, and, consequently, certain developmental events that
may require experience for activation (among them canonical babbling and
hand banging), may occur earlier in development than expected.

Early experience appears to stimulate accelerated vocal development,
while inadequate early experience can, in extreme cases, have a retarding
effect. Notwithstanding widely publicized claims to the contrary (see Lenne-
berg, Rebelsky & Nichols, 1965; Lenneberg, 1967), normal human vocal development is markedly hampered by lack of normal auditory experience. Profound hearing impairment has been shown to have a major retarding effect on the onset of canonical babbling (Oller, Eilers, Bull & Carney, 1985; Stoel-Gammon & Otomo, 1986; Kent, Osberger, Netsell & Hustedde, 1987; Oller & Eilers, 1988). The sharp difference between deaf and hearing infants suggests that audition is a requirement of normal vocal maturation, whether by virtue of listening experience or by virtue of self-generated feedback.

Given that hearing is fundamental in infant vocalizations, it is important to explain why deaf infants ever begin canonical babbling. One possibility is that deaf infants acquire auditory experience gradually (because they usually have some residual hearing), and that their delayed onsets are predictable based on the degree of their hearing losses. Although this possibility is still in the process of evaluation, Lynch, Oller & Steffens (1989) have demonstrated that canonical babbling can begin (albeit many months late) in an infant with bilateral cochlear aplasia, a condition which precludes audition of any degree. In this case, the onset of well-formed syllables occurred following intensive speech therapy and experience with a vibrotactile stimulation system, a simple artificial hearing device. Thus, although audition (whether of others’ voices or of one’s own) seems to be the most effective form of stimulation for the onset of canonical babbling, it appears that other kinds of stimulation (from another sensory modality), in combination with social interaction can be substituted to initiate the process.

Interpreting the body of existing results on the onset of babbling is thus complicated. The following account represents one possible interpretation of the data. Canonical babbling usually begins in response to a certain amount of auditory experience (otherwise deaf infants would not begin late), but the human infant may be responsive even in unusual circumstances to conditions that may stimulate babbling development. Even if experience occurs prematurely (as in the preterm infant), it appears to have a stimulating effect; even if the environment is less vocally stimulating (as may be the case with the low SES infant), the process of development appears to be initiated on schedule; even if there is no auditory experience at all (as in the acochlear infant), other sensory experiences can, with sufficient time, activate the process; and even if the infant is prevented from exercising a canonical babbling ability (as in the case of the tracheostomized infant), there is a broad time window within which completion of required motoric developments is possible. These empirically-based suggestions are consistent with the possibility that our evolutionary histories have provided a capability for vocal development, the robustness of which is reflected in its flexibility. The evolutionary success of our distant ancestors may have depended on the emergence and growth of linguistic abilities, and a key to the power of human language may have been its resilience in the face of changing circumstances.
Goals of the present work: effects of prematurity and low socioeconomic status on vocal development

The infant born at risk for developmental delay because of premature birth (Greenberg & Crnic, 1988) represents an especially interesting case where the natural timing relationships of sensory/motoric experience and maturational events may be disrupted. Yet despite such potential asynchronies, most preterm infants seem to develop normally, and, in certain domains, accelerating effects of early experience have been noted. Similarly, the evaluation of vocal development in infants of low SES has produced the suggestion that the onset of babbling is broadly similar to that of more socially and economically advantaged infants. Our previous reports based on a longitudinal study of preterm and full term as well as low and mid SES infants (Eilers et al. 1993) focused exclusively on phenomena of vocal stage onset. The present paper focuses instead on the growth of babbling measured quantitatively during the first 18 months of life as well as on the details of the development of other vocal categories in infancy during the emergence of the speech capacity.

The goal of the present study is to evaluate quantitatively a variety of speech-like vocalizations in preterm and full term as well as low and mid SES infants. The work may help to clarify the sense in which human vocal development is robust, and may provide the basis for further understanding of linguistic handicaps that sometimes are presaged by babbling anomalies.

METHODS

Subjects

Twenty preterm infants (13 female) were evaluated in a longitudinal study of infant vocal development and early speech. They were found through a combination of active recruitment and referrals from local hospitals. Their mean birth weight was 1820 g (range = 1446–2070 g), and average gestational age was 33 weeks (range = 35–11.5 weeks premature) based on estimation procedures consistent with the recommendations of DiPietro & Allen (1991). The procedure includes evaluation of a combination of data from prenatal examinations and appears to be ‘conservative’ in that empirical data suggest that estimates of prematurity so obtained tend to be slightly less than the degree of prematurity that would be designated based on procedures that are commonly used with non-at-risk births (in particular, the common procedure based on the mother’s indication of date of the last menstrual period). Additional description of the evaluation of prematurity can be found in Eilers et al. (1993).

Because a primary goal of the work was to assess the potential effects of premature extra-uterine stimulation, infants with significant health problems
were not included in the study. In particular, infants manifesting intraventricular haemorrhage, respiratory distress syndrome, severe hyperbilirubinemia, or any other important complications requiring perinatal medical intervention, were excluded. In addition, five-minute Apgar scores were required to be seven or greater, indicating substantially normal response in the first minutes after birth. All the preterm infants were housed in minimal care nurseries during postnatal hospitalization.

Thirty-three full term infants (11 female) were selected from a pool identified through health department birth records. When the infants were one month old, parents received a letter inviting them to participate in a three-year study of infant vocal development and young child speech. Parents who returned a postcard were then contacted by phone and interviewed regarding the infant's health and the family's employment and educational history. Because of the longitudinal design of the research, a key issue evaluated in the initial phone interview was the likelihood of the family moving out of the metropolitan area within the time frame of the study.

Families whose infants met the study criteria (either preterm or full term) and who planned to stay in the metropolitan area were invited to come in for an initial appointment. During this appointment, the first laboratory recording of infant vocalizations was made, and parents filled out an extensive questionnaire about their infant's health and development and the family's educational and social background. The questionnaire data provided the basis for socioeconomic status evaluation. The families were categorized as middle or low SES based on educational level and employment in white-collar or blue-collar occupations. The families were distributed evenly across the middle and lower SES groups (10 vs. 10 in the preterm group, and 17 vs. 16 in the full term group).

Assignment to socioeconomic status. The SES categorizations were based on a synthesis of methods (especially Hollingshead, 1978; Nam & Powers, 1983). The present approach made adaptations to account for the information that was reliably and practically obtainable from the families in the study. Three basic dimensions of SES were taken into account: parental educational background, source of family income and stability of family structure. Each family was assigned to one of five SES levels. No family in the present study was categorized at the fifth or lowest level (no completion of high school, unskilled workers, single parent and highly unstable families). Low SES families in the study were mostly from level 4 (no college, one parent with high school diploma, blue-collar employment), and a few were from level 3 (some college, but no completed degrees, transitional white-collar, non-management employment). Most mid SES families in the study were from level 2 (at least one parent having completed college, white collar, middle-management, teachers, nurses, mid-scale proprietors, two-parent homes),
but some were from level 1 (both parents having completed college, professional or high level management employment, stable two-parent homes), and a few from level 3. The mean SES level of mid SES families was 2.04 (s.d. = 0.71) and of low SES families was 3.73 (s.d. = 0.45).

Hearing. In order to document normal hearing sensitivity, both the preterm and full term infants received at least one hearing evaluation during the second half year of life using visual reinforcement audiometry. None of the infants demonstrated a significant sensorineural hearing impairment. Infants who demonstrated a mild conductive hearing impairment with accompanying middle-ear effusion were referred for treatment by primary health care providers.

Procedure
The primary data to be discussed in the present paper, spanning the period from 0.4 to 1.6, were based upon tape recordings made at each visit by the infants and their families to the laboratory. All infants were seen at least monthly in the first year and at least bimonthly thereafter. During the sessions, conducted in a single-walled IAC chamber with a Marantz PMD-221 cassette tape recorder and a Bose condenser microphone mounted on a boom stand, a research assistant and a care-giver (usually a parent) interacted with the infant using quiet toys. Both parent and research assistant made an effort to be silent whenever the infant vocalized and to elicit speech-like vocalization through eye contact, verbal encouragement and playful interaction. Some infants appeared to vocalize most effectively when left to play in a noninteractive fashion, especially after the onset of mobility. The parents’ and experimenters’ knowledge of the individual children was drawn upon in deciding how best to encourage vocalization in each child at each age. Recording sessions were half-an-hour long.

Data analysis
Selection of samples to be evaluated. In order to keep the data at a manageable size, research assistants selected recorded vocalization samples from each child to represent the ages 0.4, 0.6, 0.8, 1.0, 1.2, 1.4 and 1.6. Each selected recording was judged to be typical of the infant’s vocal behaviour at the designated age. More than half of the samples chosen fell within two weeks of the designated ages, and samples were not accepted if they did not fall within four weeks of the age for the interval. The average deviation of actual sampling ages from the designated ages was about one week. More than 90%
SPEECH-LIKE VOCALIZATIONS

of the samples chosen included 70 utterances (the maximum number categorized). Samples that contained fewer than 30 utterances were not selected.

**Infraphonological categorization.** Each selected tape recording was categorized according to the infraphonological model of Oller (1986) and Oller & Lynch (1992). The intent of the model is to provide a basis for assessment of the degree to which vocalizations approximate the characteristics of mature speech. The term 'infraphonological' is intended to invoke the idea that infrastructural properties of speech can be characterized and that pre-linguistic vocalizations of infants can be assessed in terms of the extent to which they incorporate those properties. The framework of description that has resulted from such work specifically notes the extent to which infant utterances possess well-formed nuclei (or vowel-like elements) and well-formed transitions between nuclei and margins (or consonant-like elements).

The definitions of well-formedness that the infraphonological framework entails represent an attempt to provide an explicit characterization of the tacit knowledge of well-formedness possessed by all mature human listeners. It is certain that mature humans have such knowledge because, without it, they would not know the difference between speech and nonspeech. Furthermore, they would be unable to recognize the fact that a person speaking an unfamiliar foreign language is speaking, as opposed to performing some other vocal act.

The goal of infraphonological training is merely to bring the tacit awareness of what is and what is not speech to a conscious level and to provide the trainee with a set of labels by which to characterize varying degrees of speech-sound well-formedness. In keeping with the goals of infraphonological training, research assistants in the project were taught to categorize a nucleus (or vowel-like sound) as well-formed if it included normal phonation (the kind of phonation that typifies speech and is, consequently, not whispered, dysphonated, creaky, falsetto, or characterized by tremor) and a resonance pattern typical of speech, implying a special (not at rest) posturing of the vocal tract. An 'at rest' tract (either closed or in a muscularly lax slightly open state associated with quiet breathing) produces a recognizable resonance pattern in vocalization, and such a pattern is not considered to be well-formed for speech, because the vowels of natural languages (with rare exceptions) are produced in specialized postures. The at rest posture is characterized by primarily low frequency resonances and usually by nasal anti-resonances (see Oiler et al. 1985). In our experience with research on infraphonological development, adult listeners quickly learn (within a single session) to designate ill-formed nuclei (or quasi-vowels) as being less speech-like than well-formed ones, and, in particular, they easily acquire the ability to designate reliably that a nucleus is a fully resonant
vowel (a ‘postured’ nucleus) or a quasi-vowel (an ‘at rest’ nucleus). This proves to be an important distinction because infants commonly produce an abundance of quasi-vowels, especially in the first year of life.

Similarly, our experience indicates that trainees quickly learn to bring to conscious awareness their tacit ability to judge well-formedness of transitions between consonants and vowels. Natural speech systems include syllables that are constrained in time, and the pattern of constraint is similar across languages. A formant transition (from a closed vocal tract associated with a consonant to a full vowel posturing) that is from 25 to 120 ms and has no phonation breaks is usually judged by listeners to be appropriate for speech. This time frame encompasses syllables that tend to have relatively short transitions (e.g. stop-vowel sequences such as [ba]) and syllables that tend to have relatively long transitions (e.g. glide-vowel syllables such as [wa]). Even longer transitions (> 120 ms) occur commonly in infant protosyllables (called ‘marginal syllables’) and are usually judged by listeners to be less speech-like than otherwise analogous syllables with shorter transition durations.

Categories of judgement and measures of infant vocalizations. Vegetative sounds (crying, laughter, sneezes, coughs, hiccoughs, etc.) are taken within the infraphonological model to be special cases of nonspeech vocalization. Consequently, such utterances were not considered in the categorizations to be reported here. Instead, the focus of the study was on those vocalizations that appear to be more indicative of the emerging speech capacity. In particular, the study focused on the following categories: full vowels, quasi-vowels, marginal syllables (which include a consonant and vowel with no well-formed transitions), and canonical syllables (which include at least one well-formed consonant-vowel transition). These categories are all specified directly by the infraphonological model, because they exemplify syllabic units with or without key features of well-formedness.

In addition, in the present work we took note of the occurrence of a series of additional, commonly occurring speech-like vocalizations of infants: squealing (vocalization at high pitch, usually in falsetto, distinctly out of the range that would be typical in speech); growling (vocalization at low pitch, often in creaky voice, distinctly out of the range that would be typical in speech); glottal fricative and glottal stop sequences (wherein vowels or quasi-vowels occur in syllable-like sequences with glottal consonants, which require no formant transition because they require no supraglottal articulation); and raspberries (labial trills or vibrants). These vocal types (and a few other nonvegetative sounds not reported on here) are sometimes referred to as ‘precursor’ vocalizations because they occur prior to the onset of fully well-formed babbling. Squealing and growling are particularly interesting precursor sounds because they appear to represent an exploration
of the pitch parameter by the infant. Glottal stop and fricative sequences are early attempts at primitive syllables (with consonant-like elements, but lacking formant transitions). These utterance types are widely documented in infants during the first few months of life (Zlatin, 1975; Roug, Landberg & Lundberg, 1989; Oiler, in press). Glottal stop sequences have been specifically noted as occurring with inordinate frequency in deaf infants (Oiler et al. 1985). Raspberries are one of the most salient utterance types of the middle of the first year, and preliminary evidence has indicated an inordinate proportion of occurrence in at least one linguistically handicapped group, Down syndrome (Smith & Oiler, 1981).

Each vocalization type was assessed as a proportion of the total speech-like (nonvegetative) syllables of the infant. The value is computed as a ratio where the numerator is the total number of syllables, in a given sample of utterances, that fit a given category, and the denominator is the total number of syllables of all categories in the sample. For example, the canonical babbling ratio is the number of canonical syllables divided by the total number of syllables of any kind. Three of the ratio variables so constituted (canonical babbling ratio, full vowel ratio and quasi-vowel ratio) were used to provide an indication of the degree of well-formedness of infant productions along a given dimension. Precursor vocalizations were also evaluated as a proportion of all syllables in the samples in order to provide an indication of the extent of occurrence of the sounds in the general vocal repertoire of the infants. The reader may note that the ratio used here with a ‘syllables’ denominator differs from the ratio presented in Oiler & Eilers (1988) where the denominator was ‘utterances’. The change to a syllables denominator began with Carney (1991) and was followed up by Steffens, Oiler, Lynch & Urbano (1992) in response to the concern that, as far as possible, numerator and denominator should have comparable potential ranges. Of course the shift of definition produces lower ratio values, but all the trends to be discussed below are similar whether the denominator is syllables or utterances.

In addition to evaluating well-formedness and proportional occurrence of sound types, the present work also examined the ‘volubility’ of infants, as measured by the rate of vocalization per unit time in the vocalization samples. By considering volubility along with the category measures, it was hoped that a more general perspective could be established on the occurrence of speech-like vocalizations in at-risk infants.

Observer preparations and reliability. As research on infraphonological development proceeds, it is necessary to establish formal methods of categorization and ways to ensure inter-rater reliability. The present research is part of a long-term programme of work to develop the infraphonological framework and its required methodological underpinnings. Consequently, prepar-
ations for the research reported here have been extensive. The seven members of the categorization team worked together with six additional trainees in a four-month seminar on infant vocalization, in which the goal was to establish clear definitions and criteria for judgement of infraphonological categories. The topics for the seminar included definitions of the infraphonological framework as specified by Oiler (1986), evaluation of infant vocalization examples from a previous longitudinal study (Oiler & Eilers, 1982) dubbed to a training tape to illustrate infraphonological categories occurring at various ages of development, and synthesized (utilizing Klatt, 1980) continua of vowels to quasi-vowels and canonical to marginal syllables. A number of exercises during the seminar required participants to categorize samples of utterances from infants and later to discuss the discrepancies among the various observers' judgements.

Through the course of these sessions, categorizational criteria were set for the group on the quasi-vowel/full vowel distinction, the canonical/marginal babbling distinction, and the utterance count. In order to attain high interobserver reliability on these judgements, it was found useful to establish a group awareness of how to parse the relevant continua. For example, along an eleven-step continuum from an unambiguous quasi-vowel to a mid-central schwa vowel, the listeners were encouraged to categorize synthesized syllables based on a specific boundary (between steps three and four) that appeared to correspond best to the infraphonological definition. Group variance was reduced by adoption of the common criterion. The boundary for individual subjects differed by as much as 7 steps before training but differed by 2 steps or less after one hour of instruction. Later, working with real samples of speech-like vocalizations from infants, the group compared the proportion of quasi-vowel and full vowel judgements obtained from independent categorizations. An effort was made again to adopt a common criterion, and individual observers whose judgements differed most from the mean outcomes were instructed to adjust their criteria appropriately and to recategorize the sample in question.

The same sorts of training were pursued for the canonical/marginal babbling distinction and for the utterance count issue. It appears to have been of particular importance that the group found best reliability (to within 15% agreement across all observers) on the utterance count factor when they were instructed to count even the least audible nonvegetative sounds. The decision to adopt this sort of utterance count criterion had the advantage of increasing reliability, but it should be remembered that such a procedure may affect the values of measures used to assess the extent of production of individual sound types—the measures used in our studies are based on ratios of particular syllable types to the total syllable count. For example, it appears likely that the utterance count procedure used in the present work has the effect of yielding relatively low canonical babbling ratios (number of
canonical syllables divided by total number of syllables) compared to ratios reported in previous studies (e.g. Oller & Eilers, 1988), because the least audible utterances (or syllables) tend to be noncanonical.

Using this sort of definitional training followed by group study of taperecorded continua, then categorization of real vocal samples, followed by recategorization to reduce group variance through adoption of criteria based on mean performance, it was possible to improve reliability and confidence in categorizations. The procedure culminated in formal tests of reliability prior to the beginning of categorization of the data for the study in question. In essence, as described in Steffens et al. (1992), 70-utterance samples of vocalizations were transcribed independently by each transcriber on three occasions: just prior to the beginning of categorization for the study, about one-third of the way through the study, and about two-thirds of the way through the study. It was intended that all transcribers should give values on all the key measures that met a preset standard of agreement – no score would differ from the mean score on the measure for the group by more than 10% of a predesignated maximum expected value on the given measure. The maximum expected values were determined empirically for each measure based on data from earlier longitudinal studies on infants ranging in age from 0;2 to 1;2 (Smith & Oller, 1981; Oller & Eilers, 1982). By referencing the reliability score to an expected maximum, it was possible to avoid inflated error percentages in cases where obtained ratios were near zero. It should be noted that these standards were strictly enforced prior to the beginning of categorization data-gathering with regard to canonical babbling ratio, vowel ratio and quasi-vowel ratio. With the marginal babbling ratio, utterance and syllable count, the criterion was relaxed to 15% error. At the reliability check points, it was verified that all transcribers had scores within the designated ranges for the key variables. The remaining measures that were monitored (glottal stop and glottal fricative sequences, squeals, growls, and raspberries) were considered less likely to be of major importance and were categorized without extensive reliability evaluation.

Coding and preliminary analysis software. Entry of data by the transcribers was facilitated by software developed in our laboratories and in use for over five years. The system speeds data entry and includes a variety of error traps to prevent observers from keying in illegal data, and to alert observers to possible errors when they enter especially atypical values.

The preliminary analysis of the data is conducted entirely by the computer. The output of that analysis yields all the necessary ratios and the utterances-per-minute values.
RESULTS

Chronological age analyses

Four mixed ANOVAs for four dependent variables, canonical babbling ratio (CBR), full vowel ratio (FVR), quasi-vowel ratio (QVR) and utterances per minute (U/M) were conducted. Each design had two between-subject factors (SES, term) and one within-subject factor (age). The analyses were conducted with seven samples of data collected from each infant at chronological ages of 0;6 through 1;6 at two-month intervals. An additional sample collected at age 0;4 was not included in this analysis but was used in the corrected-age-matched analyses reported below.

The preterm infants' chronological ages are comparable to those of the full term infants, as indicated in Table 1a, where the groups are presented collapsed over SES. The number of infants included in each group varies from sample to sample owing to factors such as illness on the part of infants, scheduling difficulties, and subject attrition across the longitudinal study. In Table 1b the mean corrected ages of the preterm infants evaluated at each sample are provided.

Results confirm the anticipated main effects for age for the three key well-formedness variables, CBR (F[6, 259] = 24.87, p < 0.0001); FVR (F[6, 259] = 8.89, p < 0.0001); and QVR (F[6, 259] = 9.53, p < 0.0001). These statistically reliable effects support the idea that the measures effectively provide a basis for monitoring infraphonological growth. Post hoc analyses of these main effects (Duncan's Range test) are generally consistent
with the expectation that CBR and FVR should increase as a function of development over the study period while QVR should decrease significantly over the same period of time. The means and standard deviations for each age and ratio variable are shown in Table 2. For CBR, samples at ages 0;6 and 0;8 differ significantly from one another and from samples at ages 0;10 to 1;2, and all of these differ from the samples at ages 1;4 and 1;6. For both FVR and QVR, the 0;6 and 0;8 samples differ significantly from all of the remaining samples.

In the chronological age analyses, main effects are also found for term status for each dependent ratio variable. In each case, full term infants exhibit the more mature behaviour to a reliably greater extent than the preterm infants. Thus, term infants have reliably higher CBRs ($F[1, 39] = 4.28, p < .05$) and FVRs ($F[1, 39] = 4.95, p < .05$) and reliably lower QVRs ($F[1, 39] = 5.00, p < .05$). Means and standard deviations as a function of term status are displayed in Fig. 1a–c.

In addition to the main effects, reliable interactions between term and age are found for FVR ($F[6, 259] = 2.23, p < .05$) and QVR ($F[6, 259] = 2.56, p < .05$). These interactions are reflected in Fig. 1b and c by the differences between the groups at younger ages and convergence of the groups thereafter. A series of one way ANOVAs were computed to evaluate the interactions. Results of these analyses indicate that the term by age interaction for both FVR and QVR is carried by the reliable differences in the samples from ages 0;8 and 0;10 between preterm and full term infants. After the samples from age 0;10, the ratios of full term and preterm infants converge. It appears that from age 1;0 on, the infants show relatively stable and comparable production of both full vowels and quasi-vowels.

As far as the SES variable is concerned, the three well-formedness variables do not show reliable differences between low SES and mid SES infants. However, the volubility variable, utterances per minute (U/M) does show a main effect for SES ($F[2, 39] = 7.05, p < .01$). The low SES infants...
Fig. 1. (a) Canonical babbling ratio as a function of age in months in preterm and full term infants for the chronologically matched sample. (b) Full vowel ratio for the chronologically matched sample. (c) Quasi-vowel ratio for the chronologically matched sample. ■—■, preterm; ▲—▲, full term.
TABLE 3. Corrected age sample: mean age, standard deviation and number of subjects

<table>
<thead>
<tr>
<th>Target age:</th>
<th>0;4</th>
<th>0;6</th>
<th>0;8</th>
<th>0;10</th>
<th>1;0</th>
<th>1;2</th>
<th>1;4</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>14</td>
<td>17</td>
<td>15</td>
<td>15</td>
<td>18</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>Full term</td>
<td>19:60</td>
<td>28:00</td>
<td>37:63</td>
<td>46:28</td>
<td>55:00</td>
<td>64:88</td>
<td>75:00</td>
</tr>
<tr>
<td>N</td>
<td>25</td>
<td>32</td>
<td>30</td>
<td>28</td>
<td>24</td>
<td>25</td>
<td>22</td>
</tr>
</tbody>
</table>

(a) Corrected age (in weeks)

(b) Chronological age (in weeks)

have reliably fewer utterances per minute (mean = 6:0, s.d. = 3:2) than their mid SES counterparts (mean = 7:5, s.d. = 3:7), while preterm and full term infants show similar volubility (preterm mean = 6:7, s.d. = 3:5, full term mean = 6:8, s.d. = 3:6).

Analysis of U/M also shows an age effect ($F[6, 259] = 4:28, p < 0.001$), reflective of the comparatively high values of the two SES and term groups on U/M in the samples from ages 1;4 and 1;6 (means 7:8 and 8:3). These values are confirmed through post hoc tests to be reliably greater than those seen in samples from the younger ages (means of 6:0, 6:8, 5:9, 6:4, and 6:4 for samples from ages 0;6 through 1;2, respectively).

**Corrected age analyses**

In order to determine the extent to which the reliable differences found above are a function of the differing gestational ages of the infant subjects at each sampling period, an additional set of parallel ANOVAs was conducted with infants matched for gestational age (corrected age). To obtain the matched samples, new groups of infants were constituted from the same pool so that chronologically younger full term infants could be compared with their preterm counterparts of similar gestational age. The construction of the corrected-age-matched data corpus included restructuring of relationships at every sample, and was done in such a way as to maximize the size of the available data set without violating the principles upon which construction of the chronologically-age-matched samples were based – namely, all samples accepted for the database were required to have been taken within one month of each designated age, and no samples less than 30 utterances were accepted.
OLLER ET AL.

(the great majority of samples were of 70 utterances). Because the preterm infants were born at variable degrees of prematurity and because individual samples were taken at variable points (though within one month) with respect to the targeted ages, the optimal restructuring yielded centres of age that were somewhat variable. For simplicity, they will be indicated in the following discussion and figures by the approximate values: ages 0;4, 0;6, 0;8, 0;10, 1;0, 1;2 and 1;4. Table 3a presents the means and standard deviations for age in weeks for corrected ages of preterm infants included in the samples for the corrected-age-matched analyses. Table 3b includes the chronological ages of the preterm and full term infants at each sample point.

Like the chronological age analyses, the corrected age analyses yield main effects of age for each of the ratio measures, CBR ($F[2, 216] = 30.46$, $p < 0.0001$), FVR ($F[6, 216] = 11.85$, $p < 0.0001$), and QVR ($F[6, 216] = 11.76$, $p < 0.0001$), indicating again that the infraphonological analysis system provides useful developmental data through which infant speech-like vocalizations can be monitored across the early months of life. The means and standard deviations for the age variables are presented in Table 4 for each ratio measure. Post hoc analyses for the CBR variable indicate a complex pattern of differences that is generally consistent with the expectation that CBR should increase with age. The age 0;4 and 0;6 samples differ reliably from one another and from all other samples; the following pairs also differ reliably from all other samples: (a) ages 0;8 and 1;0, (b) ages 0;10 and 1;0, (c) ages 0;10 and 1;2, and (d) ages 1;2 and 1;4.

For FVR, the samples from ages 0;4 and 0;6 differ from all other samples. In addition, the samples from age 1;0 to 1;4 differ from the sample from age 0;10. For QVR, the samples from ages 0;4 and 0;6 differ from all other samples. No other main effects or interactions were found.

The reliable differences between full term and preterm infants on well-formedness variables as seen in the chronological age analyses are not present in the corrected age analyses. Fig. 2a–c display the data graphically and

<table>
<thead>
<tr>
<th>Target age:</th>
<th>0;4</th>
<th>0;6</th>
<th>0;8</th>
<th>0;10</th>
<th>1;0</th>
<th>1;2</th>
<th>1;4</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBR</td>
<td>0;04</td>
<td>0;12</td>
<td>0;27</td>
<td>0;36</td>
<td>0;32</td>
<td>0;39</td>
<td>0;42</td>
</tr>
<tr>
<td>S.D.</td>
<td>0;06</td>
<td>0;14</td>
<td>0;19</td>
<td>0;20</td>
<td>0;20</td>
<td>0;20</td>
<td>0;20</td>
</tr>
<tr>
<td>FVR</td>
<td>0;58</td>
<td>0;64</td>
<td>0;75</td>
<td>0;82</td>
<td>0;80</td>
<td>0;80</td>
<td>0;80</td>
</tr>
<tr>
<td>S.D.</td>
<td>0;23</td>
<td>0;22</td>
<td>0;17</td>
<td>0;15</td>
<td>0;14</td>
<td>0;20</td>
<td>0;16</td>
</tr>
<tr>
<td>QVR</td>
<td>0;41</td>
<td>0;34</td>
<td>0;23</td>
<td>0;16</td>
<td>0;18</td>
<td>0;18</td>
<td>0;19</td>
</tr>
<tr>
<td>S.D.</td>
<td>0;24</td>
<td>0;22</td>
<td>0;17</td>
<td>0;13</td>
<td>0;14</td>
<td>0;20</td>
<td>0;16</td>
</tr>
</tbody>
</table>

48
Fig. 2. (a) CBR as a function of age in months in preterm and full term infants for the corrected age matched sample. (b) FVR for the corrected age matched sample. (c) QVR for the corrected age matched sample. ■—■, preterm; ▲—▲, full term.
OLLER ET AL.

indicate a notable reduction of the differences between the two groups seen in the chronologically based data (Fig. 1a–c).

A view of the similarity of infants of differing socioeconomic status on vocal development is provided in Fig. 3a, which shows the data on CBR for

![Diagram](image.png)

**Fig. 3.** (a) CBR for infants of low and mid SES for the corrected age matched sample. (b) Volubility (or utterances per minute) for infants of low and mid SES for the corrected age matched sample. □—□, low SES; ▲—▲, mid SES.

the two SES groups in the corrected age analysis. As in the chronological age analysis, the SES factor proves not to be reliable, as should be expected given the substantial similarity of the two groups on the babbling measure. Furthermore, that the outcome should be the same as in the chronological age analysis should not be surprising, because it is merely the result of reanalysing the corpus of data after shifting subjects in accord with gestational age—such shifts have little effect because gestational age is more or less balanced across the SES factor. It is similarly expected and confirmed also
that both FVR and QVR do not show reliable differences across the SES
groups in the corrected age analysis, and that the main effects of SES
\((F[1, 27] = 7.61, p < .001)\) and age \((F[6, 216] = 3.49, p < .01)\) for volubility
\((U/M)\) are maintained in the corrected age analysis. Again, the low SES
group produces fewer utterances per minute (mean = 5.7, s.d. = 3.1) than
their mid SES counterparts (mean = 7.3, s.d. = 3.6) as indicated in Fig. 3b.
And again, the age effect owes to the greater volubility of infants in both SES
groups and both term groups in the oldest two age samples (mean = 7.6 and
8.0 utterances per minute) as opposed to the five youngest samples
(mean = 6.4, 6.1, 6.3, 5.8, 5.9).

**Within-group and session-to-session variability**

Examination of the various figures reveals that for both chronological and
corrected age analyses, within group variances for the variables evaluated in
the study are high – in some cases, one standard deviation at a particular age
exceeds the difference between the highest and lowest means for individual
age groups. For example, in Fig. 1b, FVR is seen to vary across ages for the
full term group from a mean of 0.68 to a mean of 0.83, a difference of 0.15;
one standard deviation for either the full term or the preterm group at both
ages of 0.4 and 0.6 exceeds this difference. High within-group variances
with respect to age or group effects are important to note because they
suggest that different infants are developing at different rates; although
group summaries of data present the appearance of smooth patterns, in fact,
across ages and individual children, the picture is quite complex, and
individual subjects may differ greatly from any group mean. For example,
the group patterns suggest stabilization of the FVR by about 0.10 corrected
age at a value of about 0.80; yet past this age, individual samples continue to
occur with FVR's as high as 1.0 or as low as 0.06 (a very extreme case), and
many samples are below 0.50. For CBR, it might be imagined that, because
all the infants are designated to be in the canonical stage by the 0.10
(corrected age) sample and because the means for the two term groups range
from 0.30 to 0.44, all the infants would be producing considerable proportions
of canonical syllables in all samples after age 0.10. In fact, however, samples
with very few canonical syllables continue to occur until the age 1.6 samples,
and the range of individual CBRs at age 1.4 is from 0.90 to 0.02.

The extent of variability on production of well-formed syllables raises
questions about the nature of stages of vocal development. It seems possible
that different groups of infants may show differences in stage attainment
pattern, even if they do not show significant group differences in quantitative
values such as the CBR. A previous study from our laboratory on Down
syndrome infants (Lynch et al. 1990) has indicated that for several months
after the designated onset of canonical babbling (based on parent report and
verified in the next occurring laboratory evaluation), Down syndrome infants
are less likely than their typically developing peers to produce samples of canonical syllables exceeding a criterion of 0.15 CBR, a value that laboratory personnel set early in the study as corresponding to the minimal impressionistically 'fully canonical' sample. Other work (Oller & Eilers, 1988) has also suggested that the canonical stage may begin in deaf infants more gradually and with more fits and starts than in the case of hearing infants.

These suggestions raise the possibility that even though preterm infants at corrected ages do not differ statistically from their full term peers on CBR, they might, like the Down syndrome or deaf infants, show a somewhat different pattern of onset. To evaluate this possibility, infants in the full term and preterm groups were evaluated to determine how many of the four consecutive samples after the designated onset of canonical babbling (as reported in Eilers et al. 1993) exceed the criterion of 0.15 CBR. Seventy-five per cent of the samples from the full term group exceed the criterion after the designated onset, but only 57% of the preterm samples do. A \( \chi^2 \) analysis of these data reveals that the full term infants are significantly more likely (\( \chi^2 = 8.6, \text{d.f.} = 3, p < 0.05 \)) than the preterm infants to exceed the criterion of 0.15 CBR. No such difference obtains between the low SES and mid SES groups, where the comparisons are 73% and 71% respectively exceeding the criterion.

Additional vocal categories

The evaluation of additional categories of vocalization was conducted descriptively, with no statistical analysis of possible group differences, for two reasons: (1) many samples had no vocalizations at all in the particular precanonical categories and (2) no observational reliability data were available.

The descriptive information of primary interest is summarized in Table 5, where the average ratio for each of five precanonical vocal types is presented, broken down by full term and preterm infants for the 0;4 to 1;6 chronological age samples.

The table indicates that there were vast differences between the rates of production of the five sound types, glottal fricative sequences occurring with an average ratio (number of glottal fricative sequences in the sample divided by the total number of syllables in the sample) between 0.15 and 0.18 and raspberries occurring with an average ratio of less than 0.01. The occurrence of no vocalizations in particular categories during particular sessions is of interest because it further illustrates the ubiquitous pattern of variability in vocalizations of infants. The data suggest that infants 'play' with sounds, producing particular elements repetitively for a time, then leaving them and going on to others. Infants in the two term groups and SES groups often tend to focus on a particular precanonical vocal type within a session, but then may not produce that same type again in subsequent sessions. Note for
example that about 50% of samples show no squeals at all, but that the highest sample ratios (0.4 for the full terms, 0.34 for the preterms) are indicative of repetitive production of squeals in individual sessions. Even glottal fricative sequences, which are rarely absent in individual sessions, sometimes show extremely high values, as high as 0.65 for the highest full term sample, compared to the mean ratio of less than 0.2; raspberries, which are absent in nearly three-quarters of the sessions, show individual sessions where ratios shoot up, with a highest sample ratio of 0.22. Also notable is the fact that while all the precanonical ratios reduce to low levels in the latest recording sessions, there persist occasional occurrences of each of these precanonical vocal types through the samples from age 1;6.

The pattern of the production of particular sound types in individual sessions results in extreme variability within child from session to session. Consequently, even if there are reliable differences between full term and preterm infants, or between low SES and mid SES infants on rate of production of precanonical sounds, it would be very difficult given current methods of observation to obtain solid evidence of such differences.

**DISCUSSION**

The longitudinal research reported here indicates that the measures of infraphonological development employed in the study are sensitive to growth in the production of well-formed speech-like units across the first year of life. Furthermore, the measures of infraphonological well-formedness are sensitive to differences in vocal development of chronologically age-matched infants born at full term as opposed to those born prematurely. The premature infants, especially early in the first year, show reliably less mature vocal patterns than their full term age mates do. However, when premature
OLLER ET AL.

infants are compared on infraphonological development with full term infants matched for gestational age, the statistically reliable differences in rate of production of well-formed units disappear. Infants from middle and low SES backgrounds also prove statistically indistinguishable in ratios of speech-like vocalizations.

This pattern of results bolsters an emerging perspective on the biological foundations of speech. It appears that infants come to produce well-formed speech-like units in a manner that is notably flexible under changing circumstances. If the infant is born early, the vocal system is adaptable and emerges basically intact. Similarly, if the infant is born into a family of low SES where evidence suggests environmental support for language development may be weak (Hart & Risley, 1989), well-formed vocalizations develop, apparently fully on schedule. The robustness of the development of infraphonological capabilities may reflect the fundamental importance of being able to produce the sounds of speech throughout human history.

At the same time, it is important not to overstate the similarities in at-risk infants and infants born free of such risk. Infants with more ominous risk conditions (e.g. deafness or Down syndrome) clearly show aberrations of vocal development, and evaluation based on adjustments of method may reveal differences that are missed in initial tests even with infants beset by milder risks. The present work also provides evidence of a difference in the nature of canonical babbling in premature infants as opposed to their term counterparts. It has been traditionally assumed (Koopmans-van Beinum & van der Stelt, 1986; Oller, 1986) that the canonical stage begins suddenly and that the ability to produce well-formed syllables is never thereafter lost. This switching-on of the ability to produce speech-like syllables has been presumed to be reflected in substantial consistency of production of canonical syllables, day after day, whenever the canonical-stage infant vocalizes. The present research suggests that the onset of canonical babbling in premature infants may not be as stable as in full term infants. Thus, in four consecutive samples after the onset of canonical babbling, premature infants were reliably less likely to reach a minimum criterion of canonical syllable production than their full term counterparts. How could this be so, given that gestationally age-matched premature and full term infants produced statistically indistinguishable amounts of canonical babbling? We have noted two potential sources of the difference. First, the premature infants actually began the canonical stage (as well as handbanging) somewhat earlier (by about three weeks on the average, as reported in Eilers et al. 1993); and second, at gestationally matched ages they produced fewer (though not significantly so) canonical syllables than the full terms. This combination may account for the significant differences between the full term and premature infants on consistency of well-formed syllable productions after the canonical stage was in force.
SPEECH-LIKE VOCALIZATIONS

That the stage of canonical babbling should emerge with some differences in infants of differing risk has been suggested by previous studies as well. Work with hearing-impaired infants (Oller & Eilers, 1988) and infants with Down syndrome (Lynch et al., 1990) has suggested that extreme risk of speech disorder may be accompanied by instability in the production of well-formed syllables after the apparent onset of the canonical stage. Such results, of course, raise questions about the meaning of the 'stage' construct. In our view, the notion of stages of vocal development has been intended primarily as a heuristic device, and no strong theoretical claims have been intended. But now that it has been seen that canonical babbling emerges more gradually in some infants than in others, and that gradual emergence may be associated with certain disorders, it becomes more appealing to explore the idea of stages in the hope of establishing a more generally useful perspective.

To expand the perspective on vocal stages it will be important to seek a characterization of the mechanism that might underlie the apparent differences in onset of canonical babbling among at-risk infants and those not at-risk. There are several possible explanations for inconsistency in babbling after the onset of the canonical stage, possibilities that revolve around potential asymmetries in motoric development of premature infants, and which suggest that the production of canonical syllables involves a complex motoric structure, different components of which can come into play at different points of maturation. Premature infants may progress faster than full term infants matched for gestational age, owing perhaps to their greater motoric experience, but only in some subset of the complex motoric structure that drives the production of well-formed syllables. The part of the structure that premature infants develop ahead of schedule may make it possible for them to produce canonical syllables but may not facilitate the productions under as many circumstances as occurs with the full term infant.

Another possible asymmetry that could account for the inconsistency of babbling in premature infants depends on the possibility that the requisite motoric capability develops ahead of the social factors that motivate much vocal activity. Previous research indicates that premature infants function somewhat differently in social interaction than their full term counterparts (less smiling, Goldberg, 1978; less activity, alertness, responsiveness, Bake- man & Brown, 1980; more crying, Friedman, Jacobs & Werthmann, 1982), and that parents respond differently to premature infants (Barnard & Bee, 1982; Stern & Hildebrandt, 1986). These facts encourage the suggestion that preterm infants may show social development that is unique, and perhaps disarticulated from the normally expected pattern of motor development.

That social factors play a role in the performance of vocal acts in infants has been well-known for some time (see in particular Bloom, 1977; Papoušek & Papoušek, 1982; Poulson, 1983; Papoušek, Papoušek & Bornstein, 1984). Although the present study indicates that low SES infants display a rich
repertoire of vocalizations, and although well-formed speech-like units seem to make their appearance on schedule and in normal proportions in children born into low SES families, it is notable that low SES conditions appear to produce a depressive effect on the amount of vocalizing of infants. The lower rate of vocalization in the infants of lower SES families may be a reflection of differences in the amount of vocal stimulation through interaction with care-givers. Recent studies of vocal interaction among parents and children in ‘disadvantaged’ families suggest markedly low rates of conversation (Hart & Risley, 1989), and related studies suggest some impoverishment of the vocal repertoire of children in low SES families, an impoverishment that is reflected not in structural (grammatical) deficiencies, but in vocabulary usage (Hart & Risley, 1981).

The present study has noted differences in the amount of vocalization among infants being reared in differing social circumstances, but has emphasized similarities in the development of infraphonology. To note that infraphonological patterns of production are notably stable does not, of course, deny the possibility of subtle differences in other domains of speech-related abilities. For example, whatever the many global similarities in vocal patterning of infants reared in differing language environments (see, for example, Oller & Eilers, 1982), there may be quantitative differences in articulation of certain phonetic segments (see de Boysson-Bardies, Hallé, Sagart & Durand, 1989), and there appear to be differences in speech perception abilities by 0;6 to 0;9, dependent upon the specific language in the home (Eilers, Gavin & Wilson, 1979; Werker, Gilbert, Humphrey & Tees, 1981; Eilers, Gavin & Oller, 1982; Werker & Tees, 1983). Such results remind us that the emergence of the speech capacity may be multifaceted and may reflect complexities of both biological determination and environmental sensitivity.

REFERENCES


OLLER ET AL.


