DIVERSITY AND CONTRASTIVITY IN PROSODIC AND SYLLABIC DEVELOPMENT

D. Kimbrough Oller[1], Suneeti Nathani Iyer[2], Eugene H. Buder[1],
Kyounghwa Kwon[1,] Lesya Chorna[1], and Kelly Conway[2]

[1]The University of Memphis, USA, [2]The University of Georgia, USA
E-mail: koller@memphis.edu

ABSTRACT

The development of prosodic capabilities in infants has been studied extensively. Yet generalizations about particular capabilities emerging across the first year of life are surprisingly hard to tie down. Results have often been contradictory, in part because approaches to the study have been divergent. A traditional approach with an innatist bias has sought to demonstrate very early mastery of adult-like speech characteristics of the prosodic system. More recently, research has sought to illustrate that infants actively build a phonological system, including both prosodic and syllabic features. In our approach, adult-like features of speech are not expected to emerge fully formed early in life, but to unfold in infrastructural stages. This unfolding suggests that infant vocal categories evolve toward adult-like, language-specific features. Here, we preview a longitudinal project on infant vocal development, and argue that two widely-studied prosodic phenomena (final-syllable lengthening and pitch control) show complexities that require recognition of the infant's active participation in development of prosody. This exploratory participation leads to nuances that contradict simplistic generalizations about early emergence of adult-like prosodic structures.

Keywords: infant vocalization, final-syllable lengthening, pitch control, vocal development.

1. PROSODY AND SYLLABICITY

Prosodic (rhythmic) development has been claimed to be the starting point in the emergence of vocal language [6,16]. This conclusion has been questioned since prosodic development is so complex that it is hard to know what such a conclusion might mean [36]. We are similarly inclined to be circumspect about prosody preceding syllabicity in development, because the claim suggests that prosodic abilities must be contrasted with syllabic/segmental abilities. In fact, the earliest developments in the vocal domain in humans appear to pertain to both prosody and syllabicity.

To illustrate how prosody and syllabicity are intertwined in early vocal development, consider that descriptions of infant vocalizations often provide an account of basic acoustic parameters: duration, amplitude, fundamental frequency, noise, and resonance patterns. Such descriptions indicate that infants show major variations in how they vocalize in all these acoustic domains even in the first three months of life. But such a description says nothing about degree of development of prosody as opposed to syllabicity. All the basic acoustic parameters provide cues that differentiate both prosodic and syllabic contrasts. Therefore, the attempt to compare rate of development in prosody vs. syllabicity is subject to confusion.

If we are to observe the emergence of vocal capabilities, including prosodic capabilities, in a way that illustrates the real developments of the child, we must avoid naïve shoe-horning of infant vocalizations into categories to which they do not properly pertain. An infraphonological framework of analysis [27] explicitly avoids comparisons that shoehorn infant vocal patterns into adult linguistic categories such as those implied by the distinction between prosodic and syllabic domains. The confusion engendered by shoe-horning is well-known with regard to segmental phonetic transcription in infant vocalizations [20,28]. Since there exist no well-formed (mature) segmental contrasts in precanonical vocalization, it is unjustified to characterize precanonical sounds in terms of segmental transcription. Similarly there appear to exist no well-formed (mature) prosodic contrasts in infant vocalizations, e.g., no interrogative or declarative intonation (because they do not exist) and no final-syllable lengthening

(because there are no well-formed syllables or well-formed rhythmic sequences to instantiate syllable lengthening systematically). So categorizing infant sounds in terms of such contrasts constitutes shoe-horning, and in the absence of substantial qualifications and empirical clarifications may yield similar confusion to that found with segmental transcription of infant sounds.

Infants in the first months of life develop *their own contrast systems*. We interpret these systems as precursors to both syllabic and prosodic development. For example, among the earliest salient categories (leaving cry aside) are vowel-like sounds and vocal types that are called squeals and growls (see e.g., [33]. Using mean, maximum, minimum and variability of f0, the three groups of vocalizations can be segregated, and the outcome concords moderately well with auditory categorization – squeals have high f0, growls have low f0, and vowel-like sounds fall in the middle [13]. However, the categories are more complex than mere f0 groupings. Vibratory regimes of phonation and related vocal quality factors also play roles in auditory categorization of these sounds [3]. Infants also contrast yells and whispers with vocalizations of medium amplitude [26], and produce sounds that vary dramatically in duration during the first months [2].

The categories thus formed (squeals, growls, vowel-like sounds in modal phonation, yells, whispers, long and short patterns) are among the earliest recognizable and repetitive sounds of human infancy, yet none of them constitutes a speech category per se. Across the first year of life, these categories diversify and are elaborated to progressively resemble mature speech categories. The immature categories are products of infraphonological development, and they constitute a system of simple contrasts for the infant, functioning to express primarily emotional content and simple illocutions. The categories represent a sort of embryological form manifesting an unfolding capability to control parameters required by speech.

One important generalization we have observed is that phonatory control precedes supraglottal articulatory control in infancy [26,27]. Early emerging contrasts such as those among squeals and growls, or yells and whispers are managed primarily at the larynx, requiring no supraglottal articulation. The supraglottal articulations that do occur frequently in the first months (those pertaining to "gooing"), are diffuse and relatively uncoordinated, yielding no discernible system of articulatory contrasts [26,34,39]. Categorical emergence of

vowel-like sounds differing by articulatory posturing or emergence of syllabic contrasts such as those found in canonical babbling appear months later. This generalization (phonatory control precedes articulatory control) runs counter to a widely publicized view emphasizing reduplicated babbling as the key first step laying foundations for vocal language [21]. In fact longitudinal research beginning early in life shows that laryngeally controlled contrasts *always* precede reduplicated babbling (which requires coordinated laryngeal control *and* supraglottal articulation) [12,34,39]. The contrasitivity produced primarily by laryngeal modulations in the first months of life lays a foundation for fully well-formed syllabicity, just as it does for mature prosody. Consequently to say that phonatory control precedes articulatory control is not equivalent to saying that prosody precedes syllabicity. Phonatory control is required in both prosody and in syllabicity.

2. DIVERSITY AND EXPLORATION IN HUMAN VOCAL DEVELOPMENT

It is important not only to emphasize the salient development of infrastructural vocal contrasts in infants during the first months of life, but also to point out that the contrasting units that emerge are composed of greatly variable raw vocal material. Within any category of early infant vocalizations there is so much variation that reliable categorization by any means is demanding. The variability reflects the exploratory nature of human vocalization in infancy, a characteristic that has been noted to be unique among the primates [9,30]. Exploration appears to be a critical characteristic of early motoric development in general, not just of vocalization [35]. Out of exploration, categories of action (including reaching, and vocalizing) appear to be self-organized. It is sensible to suggest that vocal language could never emerge in the absence of vocal exploration in the early months of life, and indeed vocal language in non-human primates may be impossible because they show insufficient vocal exploration to support vocal category development.

One of the consequences of the exploratory character of development in humans is that the vocal categories that arise in infants are not crisply distinct, but allow overlap. Furthermore, individual infant utterances often include distinct time periods showing different vocal regimes (distinct modes of vibration, such as modal, loft, pulse, etc., as specified in non-linear dynamic theories of phonation) and consequently the individual utterances may not be easily distinguished as squeals, growls or vowel-like sounds. This natural fuzziness of categories owing to

exploration by the infant results in a natural limit on reliability of categorization – thus lack of high reliability in either auditory or acoustic categorization should not always be taken as a failure of training or procedure, but as the natural consequence of learning through exploration.

3. METHODS

This brief report emphasizes that diversity of vocalization owing to vocal exploration and infrastructural category formation produces complexity that confounds simplistic coding in terms of adult-like categories of vocal language. We argue for a procedure aimed at accounting for, rather than ignoring, the diversity of infant vocal development. Only in the context of such a diversified approach will we be able to observe the emergence of adult-like patterns from the background of complex and exploratory infant vocal action. We will consider two examples briefly, in each of which we explore prosodic phenomena in ways that illustrate the changeability of infant action. The two phenomena to be evaluated concern the rhythmic phenomenon of "final syllable lengthening" (FSL), and pitch control, a presumable foundation for intonation.

Three normally developing full-term girls from a longitudinal study provided the data to be discussed here. Recording sessions used here were at roughly 4, 7 and 11 months of age. The sound-treated recording suite included four cameras, two of which could be selected to record simultaneously at any point. Each session was 20 minutes long and two sessions were recorded for each of the three infants at each of the ages. Cameras were controlled and audio was monitored from an adjacent room with one-way glass.

Infants wore a custom-designed vest [4] housing a wireless microphone. Mouth-to-microphone distance was approximately 5-7 cm. During the recording sessions, the mother and often an experimenter were present in the recording area. Audio was recorded at 48 kHz.

Infant vocalizations were elicited in normal caregiver play. Cry or fuss vocalizations were rare because sessions were terminated and restarted if infants expressed distress. Often infants were in a "separated" condition, in which they were in the room with the parent, but the parent was engaged in a conversation with the experimenter, usually several feet away, often in a play station or high chair. Coding was based on the AACT system (Action Analysis Coding and Training, from Intelligent Hearing Systems, Corp., Rafael Delgado, software engineer) that coordinates video and acoustic/spectrographic (TF32, [23] displays. In AACT coders select dimensions (fields) in which to code (including facial affect, gaze direction, protophone type, vibratory regime, illocutionary force, and others). Audio/visual events are located in time using the TF32 cursors and those events can be coded in any of the dimensions. Acoustic data (f0, amplitude, duration) are automatically stored. Acoustic analysts can make additional quantitative modifications within the TF32 module and data are automatically linked to the codes. Because the actions of infants in both vocal and other modalities include much exploration, it is unreasonable to

expect very high levels of intercoder reliability, and indeed we typically find Cohen's Kappas ranging from 0.5 to 0.7 in auditory/visual coding of infant actions. Kappa values for visual coding of acoustic data on vocal regimes are typically better, 0.7 or higher [3].

For the present report we used the TF32 module within AACT to segment syllables within identified utterances (defined according to a breath-group criterion) for the three infants at the three ages. Syllable durations were determined according to principles that have been used in prior research by the current authors [25]. In addition we analyzed f0 for the same utterance set utilizing specially updated TF32 facilities that provide flexible options for assessing f0 in the wide pitch range that has been so widely reported to occur in infant vocalizations [10,14,24]. The utterances were also coded based on the video displays for facial affect and gaze direction. Based on auditory judgments, each utterance was coded as primarily consisting of vowel-like, squeal-like or growl-like phonatory characteristics.

4. RESULTS

4.1 Position-in-utterance effects

In prior studies by the current authors and collaborators [19,25,31], it has been found that the phenomenon of FSL [5], has a complex manifestation in infant vocalizations. Nathani et al. found that FSL appeared to decrease across the period from precanonical vocalization to fully canonical vocalization in normally hearing English-learning infants, but with much intersubject variability. The report attempted to synthesize many contradictory findings of the literature on FSL in infants (see e.g., [7,8,17,18,32]. FSL was extremely variable depending on at least three key factors: Age of infant, type of vocalization (canonical/noncanonical, reduplicated/variegated), and number of syllables per utterance. Methodological problems make the existing literature hard to interpret. A key problem is that some reports have confounded true FSL effects with possible isochrony effects [15], by treating monosyllabic utterances as final syllables.

We hasten to acknowledge that the procedures we used depend upon syllable identification, though the infant's vocalizations are often ambiguous even with regard to the number of syllables in an utterance. For the current results, a reliability check showed two coders agreed on the exact number of syllables for 73% of utterances over the infants. 90% of the disagreements consisted of a one-syllable discrepancy. Thus results reported here should be interpreted with the understanding that the data are noisy.

The results can be summarized best by emphasizing their complexity. Syllable durations in

the present study show high variance levels at every utterance position – the average standard deviation for syllable duration was typically 60% or more for both final and non-final syllables. Sometimes FSL occurs, and sometimes it does not, depending on a variety of both situational and methodological variables. The three infants in the present study did not show strong FSL at any age under the "whole-syllable" segmentation procedure that we utilized as a primary method. In this procedure, syllables are segmented so that both phonated and non-phonated elements are counted: Periods of apparent vocal tract closure (e.g., as seen with consonant-like elements) are treated as pertaining to the following syllable nucleus unless the closure occurs in final position, in which case the closure (if it includes significant sound) is treated as pertaining to the final syllable. Thus every vocalized period or interphonation silent period within each utterance is counted.

For the three infants at the three ages evaluated, not only did FSL not occur significantly, but in fact final syllable shortening (FSS) was the rule, especially at 7 months. These three infants may thus represent a somewhat unique cohort compared with prior infants that have been studied. It may also be that the results are more a product of session effects that would disappear with broader sampling (which will be possible in our continuing research based on additional recordings that are available for the same infants). Under a different procedure where only durations of syllable nuclei are counted and closures are eliminated, results may vary substantially. Substantial FSL sometimes occurred for utterances with the syllable nucleus procedure even though the whole-syllable procedure suggested shortening of the final syllable. The method of computation plays a role here as well – the ratio of mean final to non-final durations for a sample tends to be lower than the mean ratio of final to non-final durations of individual utterances. The tendency toward FSS is thus weakened (and FSL strengthened) in the present data if the latter computational procedure is used.

The most interesting observations in the position-in-utterance duration effects concern shifts in the effects that seem to be controlled by the infant's own systematic actions. For example, notably contrasting patterns of lengthening (one of FSL and one of FSS) were sometimes observed when data were analyzed separately for two-syllable utterances as opposed to longer ones. Further, in a case where utterances were segregated in terms of whether they were primarily reduplicated or not, FSL occurred to a reliably greater extent for the non-reduplicated than for the reduplicated utterances.

In yet another case, FSL occurred for utterances with normal phonation but not for utterances with squeal quality. Session effects seem to predominate in these cases of variation in FSL. In two 20-minute sessions where the second started only a few minutes after termination of the first, we observed dramatic shifts in FSL. In light of the apparent sensitivity of FSL to individual infant style of vocal activity from session to session, we are suspicious that apparent age effects (including large shifts that we have observed in duration of all syllables across age) in small samples, may really be session effects, the product of active infant exploration moment by moment in the vocal domain.

That reduplicated utterances failed to show FSL in the way non-reduplicated utterances did in one of the sessions is perhaps predictable given that prior research on reduplicated utterances only failed to find significant FSL [31]. Although FSL will surely emerge in English-learning infants at some point, it is hard to imagine why consistency of FSL should occur in all circumstances of development given that infants are establishing control of a complex system that, even in the adult case, sometimes shows FSL and sometimes does not [29]. When we seek to find FSL in infants early in life, failing to consider the infant's tendency to explore, we run the risk of obscuring the complexity of vocal action by shoe-horning observations into analyses focused too narrowly on the adult pattern of FSL. Variations in position-in-utterance duration effects that occur systematically under infant control, need to be evaluated on an equal footing with overall tendencies related to possible adult models of vocal rhythm.

4.2 F0 effects

Our story regarding f0 and its role in intonational development is similar. The literature tends to emphasize early control of falling intonations [11,32,38], with rising intonations developing as a means of eliciting a response from adults [22], and with intonation patterns of infants interpreted as attempts to communicate much like adults do [1]. Our own view of such matters emphasizes complexity and context dependency of how infants produce f0. Prior research by this group [37] found infants producing a wide range of intonational types when judged either auditorily or acoustically. Similarly the current research shows dramatic swings in f0 at all ages, a degree of variability that greatly exceeds that seen in mature speech and again calls to mind the exploratory nature of vocal development. Attempts to categorize infant intonation in terms of anticipated adult categories

seem ill-designed given the diversity of raw material found in infant vocalization. In our attempts to categorize the f0 contours of vocalizations in the sample under study, we were tempted to introduce a large number of additional categories – because so many utterances of infants seemed unique, not fitting well into any of the traditional groupings.

In addition, vocal quality as indicated by vibratory regimes plays a major role in two ways in any attempt to categorize intonation of infant vocalizations. The tendency in intonational research has been to treat all f0 contours along a continuum of frequency, and yet there are dramatic discontinuities in intonational phenomena within utterances, corresponding to regime shifts that can be characterized within the framework of non-linear dynamics [3]. Discontinuities in pitch and in the perception of contours that result from vibratory regime shifts sometimes occur even in mature speech, but they run rampant in infant vocalizations. We take this highly variable pattern of vocal quality and f0 to represent once again an aspect of the exploratory nature of infant development.

Another fact revealed by our analysis of the dataset is that control of f0 as a factor in the infant's creation of the categories of squeal, growl and vowel-like sounds, varies systematically in remarkable ways. When f0 was analyzed in terms of the parameters mean, maximum, minimum and log standard deviation per utterance, we found that the three infants differed reliably in *how* they implemented the vocal categories (which were determined by auditory judgment for each utterance) in terms of those parameters. But the systematic variation goes much further. The infants also differed by age in how the parameters were implemented and the changes by age were significantly different for the different parameters of f0. To put all this another way, the three infants composed squeals, growls and vowels in different ways (in terms of parameters), and this tendency varied by age. These complex and statistically reliable patterns suggests child-specific, time-varying exploration of the ability to produce and use pitch in a functional way – by manipulating these variables infants indicate that they can create vocal categories and the very fact that they do so in such differentiated ways provides justification for our suggestion that infants are actively involved in the process. The study of intonation development in the future will need to take fuller account of the fact that manipulation of the parameters of pitch in infancy appears to differ across vocal types, across ages, and across individuals, with many interactions among these variables.

5.   CONCLUSIONS

In the context of infant exploration, diversity of acoustic outcome is to be expected. Infants begin life with little fine control of vocalization, and discontinuities are to be expected given inherently non-linear mechanical properties of the vocal system, and given the need of infants to exercise capabilities through exploration. We conclude that any attempt to shoe-horn description of infant vocalization into adult categories runs the risk of failing to highlight the variability and raw material that provide the stuff from which phonological systems are built.

6. REFERENCES

[1]   Beaumont, S.L., Bloom, K. 1993. Adults' attributions of intentionality to vocalizing infants. *First Language* 13: 235-247.

[2]   Boliek, C.A., Hixon, T.J., Watson, P.J., Morgan, W.J. 1996. Vocalization and breathing during the first year of life. *Journal of Voice* 10: 1-22.

[3]   Buder, E.H., Chorna, L., Oller, D.K., Robinson, R. (in press) Vibratory Regime Classification of Infant Phonation. *Journal of Voice*.

[4]   Buder, E.H., Stoel-Gammon, C. 2002. Young children's acquisition of vowel duration as influenced by language: Tense/lax and final stop consonant voicing effects. *Journal of the Acoustical Society of America* 111: 1854-1864.

[5]   Delattre, P. 1966. A comparison of syllable length conditioning among languages. International Review of *Applied Linguistics* 4: 183-198.

[6]   Fry, D.B. 1966. The development of the phonological system in the normal and deaf child. In F. Smith & G. Miller (Eds.), *The Genesis of Language*, MIT Press, Cambridge, 187-206.

[7]   Goldfield, E.C. 1999. Prosody during disyllable production of full-term and preterm infants. *Ecological Psychology*, 11.

[8]   Hallé, P.A., de Boysson-Bardies, B., Vihman, M.M.. 1991. Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by Japanese and French infants. *Language & Speech* 34: 299-318.

[9]   Hauser, M. 1996. *The evolution of communication.* MIT.

[10] Kent, R., Bauer, H.R. 1985. Vocalizations of one-year-olds. *Journal of Child Language* 12: 491-526.

[11] Kent, R.D., Murray, A.D. 1982. Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *Journal of the Acoustical Society of America* 72: 353-365.

[12] Koopmans-van Beinum, F.J., van der Stelt, J.M. 1986. Early stages in the development of speech movements. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of early speech*. New York: Stockton Press. 37-50.

[13] Kwon, K., Buder, E.H., Oller, D.K., Chorna, L.B. 2006. The role of fundamental frequency characteristics in classifying prebabbling vocalizations. Annual meeting of the American Speech-Language-Hearing Association, Miami, FL.

[14] Laufer, M.Z., Horii, Y. 1977. Fundamental frequency characteristics of infant non-distress vocalizations during the first twenty-four weeks. *Journal of Child Language* 4: 171-184.

[15] Lehiste, I. 1977. Isochrony reconsidered. *Journal of Phonetics* 5: 253-263.

[16] Lenneberg, E. 1967 *Biological foundations of language*, John Wiley & Sons, Inc.

[17] Levitt, A.G., Utman, J.A. 1992 From babbling towards the sound systems of English and French: A longitudinal two-case study. *Journal of Child Language* 19: 19-49.

[18] Levitt, A.G., Wang, Q. (1991). Evidence for language-specific rhythmic influences in the reduplicative babbling of French- and English-learning infants. *Language & Speech* 34: 235-249.

[19] Lynch, M.P., Oller, D.K., Steffens, M.L., Buder, E.H.. 1995. Phrasing in prelinguistic vocalizations. *Developmental Psychobiology* 28: 3-23.

[20] Lynip, A. 1951. The use of magnetic devices in the collection and analyses of the preverbal utterances of an infant. *Genetic Psychology Monographs*, 44: 221-262.

[21] MacNeilage, P.F. (1998). The frame/content theory of evolution of speech production. *Behavioral & Brain Sciences* 21: 499-546.

[22] Masataka, N. 1993 Relation between pitch contour of prelinguistic vocalizations and communicative functions in Japanese infants. *Infant Behavior and Development* 16: 397-401.

[23] Milenkovic, P.. 2001.TF32, University of Wisconsin-Madison, Madison, WI.

[24] Murry, T., Amundson, P., Hollein, H. 1977. Acoustical characteristics of infant cries: fundamental frequency. *Journal of Child Language* 4: 321-328.

[25] Nathani, S., Oller, D.K., Cobo-Lewis, A.B. 2003. Final Syllable Lengthening (FSL) in Infant Vocalizations. *J. of Child Language* 30: 3-25.

[26] Oller, D.K. 1980. The emergence of the sounds of speech in infancy. In G.Yeni-Komshian, J. Kavanagh &

C. Ferguson (Eds.), *Child phonology, Volume 1, Production*. New York: Academic Press. 93-112.

[27] Oller, D.K. 2000. *The emergence of the speech capacity*, Lawrence Erlbaum and Associates.

[28] Oller, D.K. 1986. Metaphonology and infant vocalization*s*. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of early speech*.New York: Stockton Press. 21-35.

[29] Oller, D.K. 1979. Syllable timing in English, Spanish and Finnish. In H. Hollien & P. Hollien (Eds.), *Current issues in the phonetic sciences, Vol. 9*. Amsterdam: John Benjamin's. 331-343.

[30] Oller, D.K., Griebel, U. 2005. Contextual freedom in human infant vocalization and the evolution of language. In R. Burgess & K. MacDonald (Eds.), *Evolutionary Perspectives on Human Development*. Thousand Oaks, CA: Sage Publications,. 135-166.

[31] Oller, D.K., Smith, B. 1977 The effect of final syllable position on vowel duration in infant babbling. *Journal of the Acoustical Society of America* 62: 994-997.

[32] Snow, D. 1994. Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech & Hearing Research* 37: 831-840.

[33] Stark, R. 1981 Infant vocalization: A comprehensive view. *Infant Medical Health Journal* 2:118-128.

[34] Stark, R.E. 1980. Stages of speech development in the first year of life. In G.Y. Komshian, J. Kavanagh & C. Ferguson (Eds.), *Child phonology, vol. 1*, New York: Academic Press. 73-90.

[35] Thelen, E. 1995. Motor development: A new synthesis. *American Psychologist* 50: 79-95.

[36] Vihman, M.M.. 1996. *Phonological development: The origins of language in the child*, Blackwell Publishers.

[37] Weed, J., Nathani, S. 2002. Pitch and Intonation in Prelinguistic Vocalizations and Deafness,.Annual convention of the American Speech-language-hearing Association, Miami.

[38] Zlatin-Laufer, M.A., Horii, Y. 1977. Fundamental frequency characteristics of infant non-distress vocalization during the first 24 weeks. *Journal of Child Language* 4: 171-184.

[39] Zlatin, M. A. 1975 Explorative mapping of the vocal tract and primitive syllabification in infancy: The first six months. Paper presented at the Am. Speech. and Hearing Association Convention, Washington, D.C.