

Beyond ba-ba and gu-gu: Challenges and strategies in coding infant vocalizations

SUNEETI NATHANI and D. KIMBROUGH OLLER
University of Maine, Orono, Maine

Infant vocal behaviors are extremely complex. Consequently, coding these behaviors is difficult and is typically associated with low reliability across observers. Various difficulties that arise when dealing with prelinguistic vocalizations, especially in the first 6 months of life, are outlined here. A proposed database of digitized infant vocalizations that illustrates strategies used to deal with these difficulties is then described. These strategies are based on theoretical infraphonological constructs, empirical observations, and information about the nature of mature phonological systems. Furthermore, the strategies are open-ended and can be modified as new information becomes available regarding infant vocal behaviors. At present, a preliminary database is available on the Web that illustrates some of these strategies. As the database is expanded, it is expected to provide a general framework for observers to categorize infant vocalizations and thereby enhance observer reliability.

In recent years, considerable attention has been paid to infant vocalizations. This increased interest has occurred primarily because, contrary to previous beliefs (Jakobson, 1941), a host of studies conducted in the 1970s and 1980s across Europe and the United States (e.g., Holmgren, Lindblom, Aurelius, Jalling, & Zetterstrom, 1986; Koopmans-van Beinum & van der Stelt, 1986; Locke, 1989; Oller, 1980; Stark, 1980; Stoel-Gammon & Cooper, 1984; Zlatin Laufer, 1980) convincingly established that prelinguistic vocal development is related to later speech and language development (see "Brief History" section below for details). There are, however, a number of enormously complex problems that arise when one attempts to achieve an appropriate and reliable system of coding of infant vocalizations, because vocal behaviors are remarkably complex, even in the first half year of life (Oller, 2000; Oller & Lynch, 1992).

The intricacy of vocalizations requires a coding scheme and methods of monitoring interobserver reliability that are much more complex than the schemes and methods of coding that have been utilized in typical prior research on parent–infant interaction. Generally, such research has included a small number (two to five) of coding categories for characterizing infant vocalizations (Anderson, Vietze, & Doeckki, 1977; Bloom & Esposito,

1975; Ginsburg & Kilbourne, 1988; Masataka, 1993; Stern, Jaffe, Beebe, & Bennett, 1975). This limitation has offered a frame in which interobserver reliability can be relatively easily monitored and in which high reliability values can often be achieved. The research has been successful in yielding insights about infant affective control and parent–infant intersubjectivity (Trevarthen, 1979). However, the cost of the limitation in coding categories is high from the standpoint of potential insights regarding infant vocal control and the roots of speech capability. Many categories (e.g., resonance, formant transitions, pitch, etc.) appear to be manipulated in infant sounds, and, consequently, research on the roots of linguistic capabilities in infancy requires much larger numbers of coding units (anywhere from a dozen or so to hundreds) than have been used in most parent–infant research. These increases in the numbers of categories entail geometric increases in the complexity of coding, training, and reliability assessment.

In addition, unique symbology must be developed and standardized to reflect infant vocal behaviors, which are often very different from mature, adult-like vocalizations (Oller, 1980; Stark, 1980). The problem of coding has recently become even more daunting since infant vocalizations have been shown to be structured in rhythmic hierarchies rather than in mere linear strings for infants developing normally and for infants with disordered development (Lynch, Oller, Steffens, & Buder, 1995; Nathani, 1998).

A BRIEF HISTORY OF THE INVESTIGATION OF PRELINGUISTIC VOCALIZATIONS

Studies in the 1970s and 1980s (references above), which looked at linkages between early vocal develop-

This work was supported in part by Grant R01 DC00484 from the National Institutes of Deafness and Other Communication Disorders to D.K.O. and by a faculty research grant from the University of Maine to S.N. The authors thank Jennifer Brill for her assistance with manuscript preparation. Portions of this work were presented at the Third International Conference on Methods and Techniques in Behavioral Research and at the 2000 American Speech-Language-Hearing Association Convention. Correspondence should be addressed to S. Nathani, Department of Communication Sciences and Disorders, 5724 Dunn Hall, University of Maine, Orono, ME 04469 (e-mail: snathani@maine.edu).

ment and later speech and language, partly circumvented these issues by focusing only on the well-formed or “canonical” syllables in infancy. Such syllables appear relatively late, usually being limited to the second half year of life, and are highly salient and relatively easy to analogize to mature phonetic categories. Listeners find it relatively easy to identify such syllables as “ba” or “ma,” for example, while finding enormous difficulty in reliably identifying precanonical sounds such as squeals, growls, or gooing sounds of the first 3 months. These precanonical sounds do not fit the definitions of canonical sounds; they do not abide by the constraints that apply to well-formed speech sounds and so cannot be made to fit into the canonical categories of speech syllables and segments. To force them into such categories is “shoehorning” and yields a description that can only be characterized as nonsensical.

By restricting attention in research to canonical sounds, prior researchers have been able to show an undeniable link between babbling and early speech. For instance, they have observed that babbling and early speech share similar sound types; stop consonants (e.g., [b], [d]) are frequent, and consonant clusters (e.g., [st], [pl], [kr]) rarely occur in both babbling and early speech. The researchers have documented that typical infants progress through a series of increasingly complex stages of vocal development in the first year of life. By restricting their attention to speech-like syllables, these investigators have avoided problems that are inherently associated with the identification and coding of primitive, or non-speech-like vocalizations.

In fact, one reason early empirical researchers beginning in the 1940s (e.g., Irwin, 1947) failed to find a clear relationship between early vocal development and later speech is that they collapsed speech-like and non-speech-like vocalizations together by using the International Phonetic Alphabet (IPA), which is designed to transcribe only mature sounds. Anyone who has listened to the gurgles and squeals of infancy can surely gauge the daunting task researchers faced when they attempted to shoehorn primitive sounds into IPA categories. In their attempt to categorize infant sounds in terms of the categories of mature speech, they presupposed that infants were already producing well-formed sounds and syllables. In fact, early infant sounds are precursors to speech that must be categorized in terms of their own structure and must be related to speech through an infrastructural (infraphonological) model that characterizes the fundamental constraints on well-formed mature speech sounds (Oller, 1995).

By the early 1950s, the IPA transcription of infant sounds was under attack, and there were attempts to classify all infant vocalizations through acoustic analysis alone (Lynip, 1951). Such methods did not prosper any more than those based on IPA transcription because they did not provide insight into the relationship between infant sounds and speech. The acoustic method, like its transcriptional predecessor, clearly needed to be augmented by an infraphonological interpretive scheme in-

voking and taking account of the acoustic properties of mature speech in a systematic way.

The focus on the production and onset of canonical babbling in research on roots of the speech capacity has yielded important successes that could not have been achieved with the numerically limited coding schemes that have characterized research in parent–infant interaction. Canonical syllables come in many forms, and, consequently, a rich coding scheme (with potentially hundreds of categories) was invoked in pursuing research on relationships between babbling and speech. The research was successful in showing the importance of canonical syllable production and has led to new ways of predicting such handicapping conditions as deafness, which is accompanied by a delay in onset of canonical babbling (Eilers & Oller, 1994; Koopmans-van Beinum, Clement, and van den Dikkenberg-Pot, 2001; Oller & Eilers, 1988).

Still, the thrust toward early identification of disorders related to speech and language requires us to look beyond canonical sounds, which are not normally in place until 5 months of age or later in normally developing children and occur much later in children with hearing or linguistic handicaps (Oller & Eilers, 1988). Babies are being referred for speech and language services at ever-earlier ages primarily due to sophisticated diagnostic techniques, and, so, tools for handling precanonical sounds are sorely needed.

Research findings have hinted that precanonical sounds may indeed be a fruitful area for further research. For instance, Lynch (1996) noted that quasiresonant nuclei (Q), precanonical sounds that occur often from the first month of life, were predominant in an acochlear child’s production prior to the onset of canonical babbling, whereas, in typical infants, a combination of sound types were present in precanonical stages, including Qs, fully resonant nuclei (F; a more mature “vowel-like” sound type than Q), and marginal babbles (primitive syllables with slow or erratic formant transitions). Fs became prominent in the acochlear child’s production only after the onset of canonical babbling. Anecdotal evidence suggests that the presence of “gooing” in precanonical stages is readily recognized even to naïve observers and may eventually be identified as an important milestone in early vocal development, a milestone that may yield predictive power comparable to that which has been found with the onset of canonical babbling. Given these potential implications for precanonical vocalizations, we need coding schemes and reliability analysis methods to handle *all* infant vocalizations.

Newer investigations, starting in the 1980s (e.g., Koopmans-van Beinum & van der Stelt, 1986; Oller, 1980; Oller & Lynch, 1992; Stark, 1980), have attempted to deal with both canonical and precanonical (less mature) sounds of infancy in order to fully understand the true nature of prelinguistic vocal development. The schemes that have been developed are, however, unique and highly specific to the purposes of the authors that developed them. For instance, Oller and colleagues have identified the basic building blocks of mature speech and

evaluate precanonical infant vocalizations according to how well they conform to these general parameters of speech production. Stark and colleagues have adopted a more motoric perspective and examine all infant vocalizations according to the articulatory complexity of the sounds interpreted in the context of acoustic analysis. On the other side of the Atlantic, Koopmans-van Beinum and colleagues have used a sensorimotor approach and described precanonical infant sounds according to the phonatory and articulatory characteristics of these sounds.

In each case, the sounds being categorized are themselves enormously complex. As a result, the coding schemes that have been proposed have much in common at general levels but have many differences at the level of details. These differences make comparisons across laboratories, as well as the validity and reliability of transcription of infant vocalizations, highly problematical. Furthermore, as described earlier, coding of sounds of infants, especially precanonical sounds, is not an easy task, even when theoretical constructs are available to guide coding decisions. Consequently, standard criteria for coding of prelinguistic vocalizations need to be developed in order to promote validity and reliability of linguistically oriented prelinguistic vocal coding schemes.

PROBLEMS ENCOUNTERED IN CODING INFANT VOCALIZATIONS

Because we need to account for their richness and complexity, coding of infant vocalizations is difficult at a variety of points. Potential confusions range all the way from the most basic decisions (e.g., which vocalizations should be included) to coding of rhythmic hierarchies and acoustic analysis. What follows is an outline of key areas of coding difficulty, but it should be emphasized that the outline is not comprehensive. Special difficulties associated with the coding of higher order rhythmic units are not treated (because the decisions about how to handle them are in flux), and problems of interpreting acoustic analyses are also left largely aside. Furthermore, the emphasis here is on difficulties of coding for precanonical sounds as opposed to segmental characterization of canonical syllables. The focus is on auditory-based coding of infant sounds at the level of syllables and utterances up to the beginning of the canonical stage. Our goal here is not to set coding criteria or to advocate a fixed categorization scheme but rather to outline a few of the areas of inherent difficulty in coding for infant vocalizations and to propose a coding and training framework in which criteria can be set and standardized across laboratories with systematic reliability evaluation and plenty of room for modification of coding schemes as researchers find it necessary to elaborate or otherwise change them.

Inclusion of Vocalizations

Audibility. Prelinguistic vocalizations, especially at the very early ages of 0–3 months, are often faint, and, therefore, a criterion needs to be set about which utter-

ances to include in coding on the basis of their audibility. The criteria that have been used in the past have varied from laboratory to laboratory, but it is hard to know what the criteria have been; the notion “audibility” is vague and unquantified. Whatever criterion is chosen, it can always be called into question when faint utterances occur due to variations in gain levels of recording and/or playback.

The problem of audibility is not merely one of amplitude of the infant sounds, since utterances can also be partially obscured by noise in the recording room, often in the form of another person’s speech or sounds of toys or other objects. Obviously, coding decisions have to be made regarding exclusion of the utterances with low signal-to-noise ratio or inclusion of parts of utterances that are not obscured. Such decisions imply that coders need to be trained to include vocalizations that exceed some criterion level of audibility and to exclude others. As will be seen below, we propose that such training to achieve reliable perceptual coding be done on the basis of recorded examples that are precategorized to be used in training according to theoretically derived standards.

Differentiating protophones from vegetative sounds and fixed signals. To the extent that researchers wish to focus on speech-like sounds or presumed precursors to speech in infant vocalizations, they usually exclude from consideration vegetative sounds (burps, coughs, etc.) and fixed vocal signals (reflexive crying, laughter, shrieking, etc.). The focus instead is on “babbling” sounds or protophones, the vocal types that appear to be relevant to speech, partly because the protophones appear to be produced in circumstances that do not involve obvious stimulus control, either external or internal.

But the distinction between protophones and other infant sounds is not always clear. A particular problem arises when coding utterances that are fussy/whiny. Clear crying episodes are typically discarded in describing the infant’s speech-like sound repertoire. However, there is considerable ambiguity about the coding of fussy vocalizations, and there is good reason for attending to some sounds that have fussy quality. Stark (1989), for instance, expressed the opinion that cry-like, fussy vocalizations serve as a playground for experimentation with the vocal track and, consequently, should be included in determining or classifying prelinguistic vocal behaviors. On the other hand, to the extent that that fussing is merely physiological and involuntary, like prototypical infant cry, it is usually discarded in coding, according to the prevailing wisdom of infant vocalizations research.

There are, however, perceived degrees of fussiness and apparent volition of fussy sounds; different listeners may parse these continua differently. Furthermore, vocalizations with alternating periods of fussiness are problematical. Again, different coding approaches are possible; individuals may decide to code only the nonfussy part or to exclude the whole utterance, and the decision may depend on the extent of the fussiness that occurs. Fussy vocalizations that have substantial speech-like quality (e.g., as in the case of fussy canonical babbles) are typically

included (because it seems unlikely that they could be produced entirely involuntarily), but, again, the matter of degree of speech-like quality must be considered. Coders clearly need help in making consistent decisions about parsing of continua of fussiness and speech-like quality. Our proposal is to approach the problem with empirically based training designed to set perceptual criteria consistently across listeners. The approach can apply to the problem of inclusion of utterances not only with fussy quality but also with laugh-like quality, shriek-like quality, and so on.

Lower Order Rhythmic Unit Determinations

Utterance determination and utterance boundaries. Once we determine which vocalizations to include, the next problem that arises is one of segmentation of the speech-like stream into “utterances.” Recently, the notion of “breath group” has been invoked to designate groupings of vocalizations into utterances. Utterances are generally defined as vocalizations or groups of vocalizations separated from all others by audible ingressive breaths or separable in accord with adult judges’ intuitions that vocalizations are separated by pauses that potentially could include ingressive breaths (Lynch, Oller, & Steffens, 1989; Oller, Eilers, Bull, & Carney, 1985; Stark, 1980). Reliability across observers in the same training laboratory using these breath group criteria has been highly satisfactory and substantially better than reliability obtained using previous criteria for utterance segmentation based on fixed pause duration (Oller & Lynch, 1992).

Recent studies have, however, found differences in utterance boundary determinations across even closely related laboratories, such as those at Purdue University and the University of Miami (the previous home institutions of the first author and the second author, respectively), even when similar breath group criteria were used to delineate utterances. Thus, when the Lynch et al. (1995) and Nathani (1998) data were compared by the present authors, it became apparent that some utterances from the Nathani study had been classified as “phrases” (a level of rhythmic organization just higher than utterances) in the Lynch (1995) study, even though Nathani intended to apply the Lynch criteria as described in the published research. These measurement differences highlight the difficulties encountered when transporting definitions across laboratories, based merely on written descriptions of coding criteria. Utterance boundary judgments appear to be based on continua that must be parsed by the listener; again, we propose that recorded training exemplars be utilized to instill consistency across laboratories.

Syllable determination. The next measurement issue has to do with determining the number of syllables, or syllable-like elements, within utterances. Syllables are the minimal rhythmic units in speech, and all rhythmic phenomena in language are reckoned with respect to syllables. Every syllable must have a nuclear element or vowel,

and consonants can be thought of as transitional elements between syllabic nuclei. Consequently, the coding of infant vocalizations in a way that will allow comparison with the fundamental infrastructure of speech requires a determination of the minimal rhythmic units in infant sounds. Counting of canonical syllables produced by infants is fairly straightforward (although we will highlight some problems later). However, counting of pre-canonical syllables is inherently problematical because, by definition, these elements do not have the timing characteristics of mature speech. Articulation in pre-canonical sounds is also rarely orthodox, further complicating counting. Training is, therefore, essential in order to achieve consistency across observers, and such training needs to be able to draw on a well-categorized database of recorded infant sounds.

Quality of Phonation

Superimposed on all the levels of categorization considered above is the issue of voice quality. Infant vocalizations are frequently accompanied by aberrant phonation, breaks in voicing, rapid pitch variations, and so on. When such features impinge on the production of syllables, counting, even of canonical syllables, becomes very difficult. Variations in voice quality also significantly influence the coding of the syllables themselves.

Of course, voice quality variations are of interest in and of themselves. Infants use changes in vocal quality to mark particular protophone categories, such as squeals and growls, and these changes can be implemented in varying degrees. Consequently, training is needed to help listeners determine which utterances are, for example, sufficiently squeal-like to be deemed “squeals” and which should be deemed to have normal phonation (even if they possess some minimal squeal-like quality).

Classification of Proto-phoneme Types

Once syllables are counted, segments or syllable types need to be classified. Recent theoretical advances—notably, the infraphonological approach developed by Oller and colleagues (Oller, 2000)—provide a logical basis for classifying infant sounds. A key focus of the classification is the proto-phonemes, sound types that appear to have a functional status for infants (because, if for no other reason, they are each produced repetitively and in systematic alternation with other sound types) and have no fixed links to biological functions, as in the cases of vegetative sounds or fixed signals. Characteristics of proto-phoneme syllables are examined to determine the extent to which they abide by universal principles that are necessary in order to generate mature speech. In other words, proto-phonemes are classified differently on the basis of their resemblance to the characteristics of mature, canonical syllables.

Four major types of proto-phonemes include quasi-resonant nuclei, fully resonant nuclei, marginal syllables, and canonical syllables. Canonical syllables are mature, adult-like syllables that contain a vowel-like nucleus and

consonant-like margin, have rapid formant transitions between the margin and the nucleus, and are produced with normal phonation. Marginal syllables are like canonical syllables, except they have slow formant transitions. Fully resonant nuclei have only vowel-like elements that sound like mature, adult vowels. Quasi-resonant nuclei are like fully resonant nuclei, except the vowel-like elements are produced by a vocal tract that is at rest, and, therefore, they lack full resonance and do not sound like adult vowels. Other infant vocalizations (e.g., squeals, raspberries, etc.) that represent other dimensions of control in mature speech (e.g., squeals are steps toward control of the pitch parameter) are coded as features that can be superimposed on the four base protophone categories and any subcategories of them that may be invoked by particular investigators.

The infraphonological categories are, however, not entirely discrete. The sounds are differentiated in accord with acoustic and articulatory continua, and training is necessary to set consistent boundaries along the criteria to be applied across observers. To some extent, boundary determinations are arbitrary (since one could shift quantitative values of boundaries without changing fundamental principles of parsing), but the general criteria are set in accord with principled constructs specified by the infraphonological model. Ambiguity is, however, inherent in prelinguistic sounds, making the task of achieving validity and reliability highly difficult despite the availability of principled constructs. Training on how to parse continua can be a very effective means of enhancing coding reliability.

Additional Measurement Issues

An additional limitation in the measurement of inter-observer reliability in infant vocalizations has to do with equipment limitations. Counter numbers on tape recorders for analog recordings are difficult to match up to original numbers on subsequent listenings, even when the same playback device is used. Consequently, reliability across observers for some utterances is difficult to obtain, merely because it may be difficult to locate specific utterances. Recent trends toward digitized recordings

with time-locked digital codes should go a long way in alleviating this problem.

A STRATEGY FOR IMPROVEMENT IN INFANT VOCALIZATIONS CODING

As can be seen from the preceding discussion, problems in measuring infant vocal behavior are rampant. Does this imply that achieving reliability and validity for infant vocal measurements is a pipedream? We think not. Technologies can be developed to improve measurement and to monitor reliability more effectively. In order to ensure workable category definitions, to establish standardization, and to ensure reliability of coding, we need standardized tools for training of observers. Our recent experiences with training students in research seminars suggest that it is possible, with sufficient training, to develop consistent coding criteria for classifying infant vocalizations.

Design of the Strategy

We are in the process of developing a digitized interactive database for coding infant vocalizations at the University of Maine. This proposed and emerging database will include digitized examples of utterances and phrases from typical and disordered infants at various ages and stages of vocal development. These examples will illustrate the criteria we (representing two of the primary schools in modern vocal development research) use for inclusion of vocalizations, utterance boundary determinations, syllable counts, protophone classification, vocal quality judgments, and establishment of rhythmic hierarchies. Users can train on each of these aspects separately. Because examples will be digitized, users need not rely on fickle counter numbers from analog recordings, thereby ensuring that they are listening to the exact vocalization for which the code is intended. In addition, users will be able to randomly access the desired vocalization on multiple occasions.

Each example utterance or phrase in the database (see Figure 1) will be linked to a standardized transcription within a flexible, computerized system (Logical Interna-

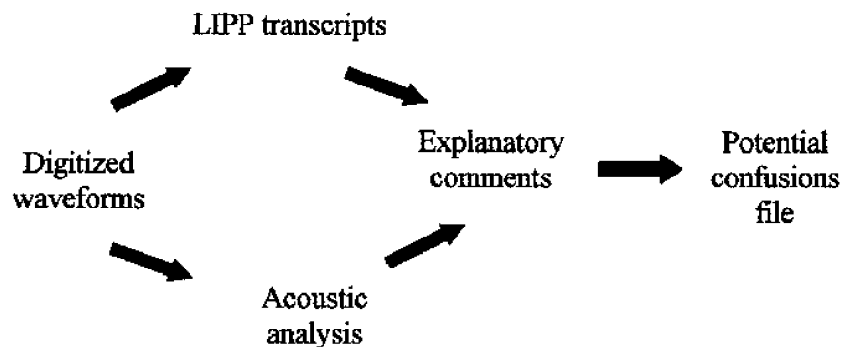


Figure 1. Training database design.

tional Phonetics Programs, LIPP; Oller, 1991) that facilitates user design and definition of symbols and keyboard mappings. LIPP allows the user to define categories along 16 dimensions (20 values per dimension) with hundreds of potential base categories and/or features (each coded event or segment can represent multiple features). Hierarchical rhythmic units can be categorized using specific boundary symbols. The LIPP Analysis Language (LAL) provides a simple programming syntax for assessment of complex LIPP codings, thereby providing convenient reliability assessment of infant codings. LIPP allows user design of symbols, definitions, diacritic combinations, and keyboard mapping such that, as new decisions are developed on the basis of theoretical or empirical considerations, symbols can be modified and new symbols can be introduced.

Another important feature of the strategy will be the use of acoustic analysis. Each example in the database will be linked to a sound wave file that presents a convenient spectrographic or oscillographic display of the digitized information. A Windows version of CSpeech (TF32, 2000, available as freeware) will be used in the database and can be confidently recommended (although LIPP can be interfaced with many acoustic analysis options), since it offers crisp acoustic evaluation of many definitional features (e.g., formant transitions) included within each infraphonological category. It is hoped that a combination of LIPP and CSpeech will greatly enhance reliability and promote consistency in criterion setting for coding infant vocalizations.

The acoustic analysis and transcription will ultimately be linked to a commentary explaining why the standard transcription is recommended. Finally, the example will be linked to a file that indicates possible confusions of identification regarding the particular example and points to other example utterances or phrases in the database that can be utilized to help clarify how and why coding

decisions are recommended by the standard coders. In particular, the confusions file will be designed to facilitate location of appropriate foils for training modules. Since the coding scheme is multidimensional and very complex, there are often many possible confusion types. The database and its linkages can be utilized to create training modules in LIPP. The user would simply open a LIPP file and, line by line, would be presented with digitized examples with all the appropriate linkages available.

In the future, we anticipate the development of adaptive, expert systems modules for training that will present each successive item to trainees based on the results of their attempted categorizations. Using such methods, it should be possible to train observers to standard criteria of coding at a variety of locations at low cost. In this way, we expect to see substantial improvements in interlaboratory communication and in interobserver reliability in coding of infant vocalizations.

At present, a preliminary version of this database (Ivoc) is available at <http://www.umaine.edu/comscidis/ivoc/ivoc.htm>. Instructions for using this database are available at the Web site. This preliminary database contains digitized infant vocalizations that illustrate some strategies for coding infant vocalizations (e.g., criteria we use for inclusion of vocalizations, utterance boundary determinations, syllable counts, and protophone classification). Reference will be made to these examples, whenever available, as we discuss our strategies for dealing with problems in coding infant vocalizations. The digitized infant vocalizations, or TF32 sound wave files, can be accessed via a ThinLIPP (freeware version of LIPP) transcript file, *ivocdat.lip*. The alphabet used in *ivocdat.lip* is a customized alphabet, *Newlip.alp*, created specifically to enable coding of infant vocalizations. Figures 2 and 3 provide an example of an utterance within this training file—the associated LIPP transcript for that utterance and the corresponding acoustic analysis, respec-

File: IVOCDAT.LIP Line: 39 Pos: 1

Play Select Spec.Ana. Wave File: VNOTGR1.WAV

Comment: First Syl, MB or CV?

**
 P C*U U ? U: P
 ~
 Cr
 # U ? U: P
 # P

First Line Last Line Up Dwn

V → Vowel
 C*V → Marginal Syllable
 ? → Glottal stop

Figure 2. LIPP screen showing the infraphonological codes for a sample utterance from the training files.

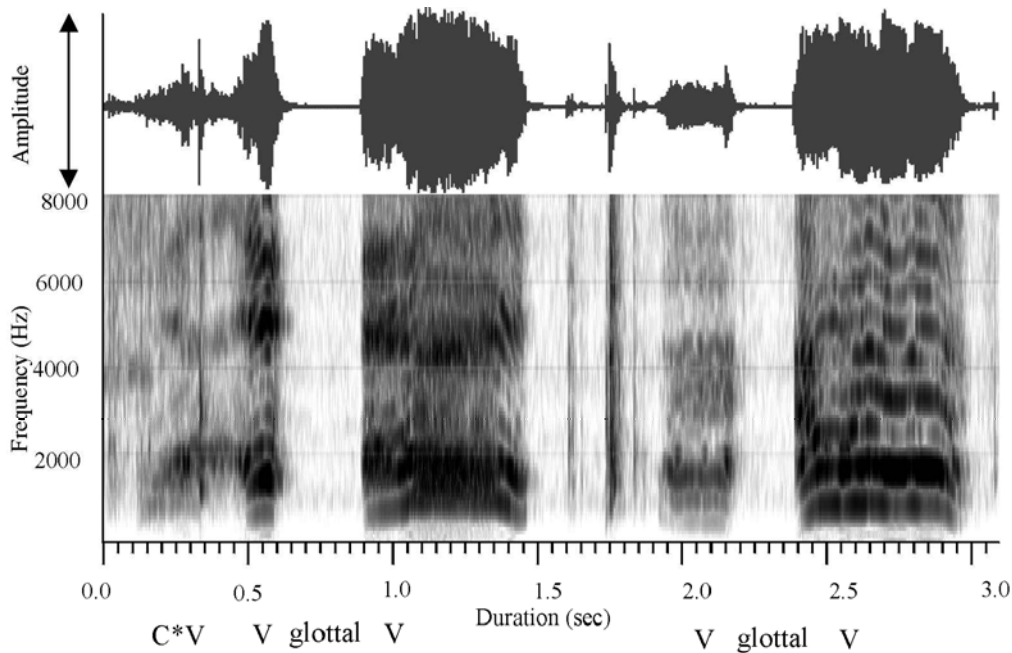


Figure 3. Acoustic analysis (waveform and spectrogram) of the sample utterance that corresponds to Figure 2.

tively. This example refers to Line 39 in *ivocdat.lip*, and the associated *tf32* wave file is *VNOTGRLDF.wav*.

The *ivocdat.lip* training file and others of its kind (in development) are presently being used to illustrate criteria for coding infant vocalizations to students at the University of Maine. Reliability studies on use of these criteria are also being conducted. It should be recognized that, at present, these criteria are flexible and will undergo change as more data on their use become available. It is hoped that users across other laboratories will make use of this and other preliminary training files as they become available, provide input on our criteria, and thereby assist in the establishment of reliability for these criteria.

Specific Strategies Used

For parsing of continua, standard boundaries have been set on the basis of our experience with infant vocalizations and, in some cases, on the basis of empirical information about well-formedness of phonological elements in mature natural languages. The general training approach is to attempt to teach observers to parse infra-phonological continua according to these standard criteria. The criteria are set provisionally for training but are expected to be adjusted as new information emerges when users implement these criteria.

Inclusion of Vocalizations

Audibility. With regard to each type of coding difficulty discussed above, the database will be designed to provide digitized standard examples that can be used to teach observers to code according to the criteria below. On the issue of inclusion of utterances in coding, for example, the completed database will provide examples of

utterances that are to be coded and examples of others that are to be excluded from coding on the basis of various criteria. These examples are not available in the preliminary *Ivoc* database.

In dealing with the issue of inclusion of utterances based on audibility, the standard coders (at this point, the authors of this paper) attempt to maximize confidence in coding judgments at the level of syllable counts and protophone classification. Consequently, the system of training is designed to teach observers to include utterances that are audible to the extent that syllable coding decisions can be made with relative confidence. When utterances are faint merely due to recording/playback levels, we adjust the gain to enable sufficient audibility for coding decisions, and we require that trainees utilize the same levels. If coding decisions cannot be made with confidence, the utterances are designated as insufficiently audible and therefore would not be coded further. However, examples of utterances that are insufficiently audible to meet the standard criterion will be included in the database for the purposes of training observers. We recommend that coders utilize "global symbols" (symbols defined within the standard alphabet to indicate the presence of utterances without characterizing their content; e.g., "U") to designate utterances that are obscured by noise or extraneous voices.

Inaudible or noise-obscured utterances may be appropriate to include given the designs of particular research. For instance, noise-obscured utterances may be important to include in studies of volubility in order to ensure that samples taken from different children or taken at different points in time (and, consequently, including different amounts of noise or extraneous voice) are fairly com-

pared in terms of numbers of utterances occurring (Nathani, Lynch, & Oller, 1999). In the work cited, any portion of a noise-obscured utterance that was reliably identifiable was coded with specific protophone symbols rather than global symbols. On the other hand, these utterances were not included in other aspects of the research—for example, when the investigators were evaluating the extent of final-syllable lengthening in utterances where onsets or offsets of acoustic information in the infant utterances were not always identifiable in the context of the noise (Nathani, Oller, & Cobo-Lewis, 2001).

Inclusion criteria for protophones as opposed to fixed vocal signals and vegetative sounds. As with audibility, examples will be provided in the database of utterances that are to be excluded from coding because they are deemed to be reflexive fixed signals or vegetative sounds. The example 3FUL1.wav (ivocdat.lip line 7) in the preliminary Ivoc database is presented as an example of when breaths are excluded from coding. In addition, examples will be provided that indicate how the standard coders treat utterances that are somewhat fussy, but not so fussy as to be judged to be reflexive. In general, we include and code partly fussy vocalizations with notations to indicate that fussiness was present. Examples are not presently available in Ivoc. Examples will also be eventually provided to differentiate reflexive laughter (which is excluded from infraphonological coding) from vocalizations that are produced intentionally to sound like laughter (e.g., infants sometimes use a syllable that sounds like [ha] as an indicator of assent) and, consequently, must be coded.

Lower Order Rhythmic Unit Determinations

Utterance judgments. Examples will be provided within the database to indicate the standard coders' criteria for designation of individual utterances and boundaries between utterances. We use the notion of "breath groups" to identify utterance boundaries. As mentioned earlier, despite this working definition of utterance boundary, differences can exist across observers in the setting of this criterion. The use of digitized samples has greatly alleviated this problem as trainees can listen to the examples and shift their criterion, if necessary, to match our consensus coding rather than relying on their own intuitions about breath groups. Often, two syllables occurring in close succession are treated as separate utterances if there is an audible breath between them (and acoustic displays are often especially useful in confirming auditory impressions on this point), but, on other occasions, the assignment of utterance boundaries does not require an intervening breath. Example waveforms to illustrate when utterance boundaries are and are not invoked are found to be extremely useful in training. In general, the method conforms to intuitions of most listeners prior to training, but the examples make it possible to tighten the distribution of listener responses quickly. Examples that illustrate the assignment of utterance boundaries in Ivoc include 3FUL1.wav, CVP4.wav, and VNOTGRLDF.wav (ivocdat.lip lines 5, 8, and 36).

Syllable determination. We use the construct of countable "beats" in order to identify number of syllables within infant utterances. Listeners use their intuition to judge the number of beats in an utterance, which corresponds to the number of elements that will eventually be coded as to type. Acoustic analysis can be helpful here: Syllables are identified as occurrences of marked acoustic energy, generally less than 500 msec in duration as indicated on spectrograms and amplitude displays (Lynch et al., 1995). As in the case of utterance boundary training, the use of standard examples can tighten the distribution of responses from listeners, although much remains to be learned about optimal syllabification strategies in infant utterances. The database should go a long way toward improving the state of the art, where it has been found that interobserver reliability on syllable judgments is often quite low. Within Ivoc, there are several examples that illustrate the criteria for syllable identification, including QST1.wav, CVP4.wav, CVS2.wav, MBQ3.wav, MM01U37.wav, MM01U31.wav, MM01U32.wav, MM01U40.wav, MBSLDF1.wav, MLTSYDF1.wav, and VNOTGRLDF.wav (ivocdat.lip lines 1, 10, 13, 16, 19, 21, 24, 29, 32, 34, and 37).

Voice Quality

Similar criteria are used when utterances are produced with aberrant phonation. Utterances for which coding decisions can be made with confidence despite aberrant phonation are included with a note to indicate the type of aberration in voice quality (e.g., falsetto, etc.). One example of this in Ivoc is MBQ3.wav (ivocdat.lip line 18).

Classification of Protophone Types

The infraphonological approach has been primarily used to develop the constructs in coding syllable types. Because syllable types, however, exist along a continuum, access to digitized materials allows trainees to access our decisions regarding boundaries for the various syllable types. One critical continuum is the one along which marginal babbles (with slow formant transitions) and canonical babbles (with rapid transitions) are differentiated. Intuitive judgments by ear are usually relatively good in achieving reliability in parsing of this continuum, but some improvements have been found with training to standard criteria with auditory stimuli alone. Since there are empirical reasons to set a boundary for slow formant transitions at about 120 msec (see Oller, 1986), acoustic analysis is often helpful in confirming and illustrating standard decisions in training with the database. Specific examples in Ivoc of marginal and canonical syllables include QST1.wav, CVP4.wav, CVS2.wav, MBQ3.wav, MM01U31.wav, MM01U32.wav, MM01U36.wav, MM01U40.wav, MBSLDF1.wav, and VNOTGRLDF.wav (ivocdat.lip lines 4, 12, 15, 17, 23, 26, 28, 31, 33, and 39).

The continuum between quasivowels and full vowels may also admit ultimately an acoustically oriented criterion based on some complex notion of spectral tilt perhaps combined with measures of antiresonance. However, at

present, the distinction is somewhat vague acoustically and is best treated on the basis of the auditory perception of posturing of the vocal tract. If an utterance is produced with the vocal tract at rest (unpostured), it is deemed to be a quasivowel, but if it is produced with discernible posturing (e.g., jaw lowering, lip spreading, etc.), it is deemed to be a full vowel. Ivoc presents examples QST1.wav, 3FUL1.wav, CVP4.wav, CVS2.wav, MM01U37.wav, MM01U31.wav, MM01U32.wav, MM01U34.wav, MM01U40.wav, MLTSYDF1.wav, and VNOTGRLDF.wav (ivocdat.lip lines 2, 3, 6, 11, 14, 20, 22, 25, 27, 30, 35, and 38) to help learners identify differences between quasivowels and full vowels.

Additional Features of the Overall Training Strategy

Users will eventually be able to conduct searches using parameters suited to their interests and needs. Each entry in the database will provide information about the infant (age, typical or disordered, level of vocal development, gender, language background, etc.). A user could therefore search, for instance, for all utterances produced by infants with severe-to-profound hearing impairment or for all precanonical vocalizations produced by a particular group of children. Training could then be tailored to coding problems that may be group specific. Trainees can listen to the selected utterances conveniently, code them, and view the descriptions, the infraphonological coding scheme, and consensus codings either before or after their own coding and can evaluate coding decisions in the light of flexible acoustic analyses. Reliability analyses can be easily formulated in LIPP to evaluate trainee codings against standard consensus codings. A record can be kept of vocalizations that pose particular challenges for the trainee.

CONCLUSION

Clearly, infant vocal behaviors represent more than just ba-ba and gu-gu, even though the lay public often conceives of infant vocalizations as rather simple. In fact, infant vocal behaviors represent a rich and complex repertoire that holds immense promise for early identification and interpretation. This richness and complexity, however, render the task of coding infant vocalizations challenging.

At the University of Maine, we are in the process of developing a database of typical and disordered infant vocalizations that have been primarily coded using infraphonological principles. This database will allow the user instant, random, and multiple access to codes in LIPP and acoustic analyses of these vocalizations. A preliminary version of this database, Ivoc, is presently available. It is hoped that this database will provide the user with an initial basis for the principled measurement of infant vocal behaviors. At the same time, it is recognized that the criteria for coding infant vocalizations are not absolute, and the database will, therefore, be open-ended.

As further knowledge about infant vocal development becomes available—particularly when more sophisticated coding of typical and disordered vocalizations is conducted—criteria will be updated and refined. Reliability studies also need to be performed before these criteria are used routinely to code infant vocalizations.

REFERENCES

- ANDERSON, B. J., VIETZE, P., & DOKECKI, P. R. (1977). Reciprocity in vocal interactions of mothers and infants. *Child Development*, **48**, 1676-1681.
- BLOOM, K., & ESPOSITO, A. (1975). Social conditioning and its proper control procedures. *Journal of Experimental Child Psychology*, **19**, 209-222.
- EILERS, R. E., & OLLER, D. K. (1994). Infant vocalizations and the early diagnosis of severe hearing impairment. *Journal of Pediatrics*, **124**, 199-203.
- GINSBURG, G. P., & KILBOURNE, B. K. (1988). Emergence of vocal alternation in mother-infant interchanges. *Journal of Child Language*, **15**, 221-235.
- HOLMGREN, K., LINDBLOM, B., AURELIUS, G., JALLING, B., & ZETTERSTROM, R. (1986). On the phonetics of infant vocalization. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of early speech* (pp. 51-63). New York: Stockton.
- IRWIN, O. C. (1947). Infant speech: Consonant sounds according to manner of articulation. *Journal of Speech & Hearing Disorders*, **12**, 402-404.
- JAKOBSON, R. (1941). *Kindersprache, Aphasie, und allgemeine Lautgesetze* [Child language, aphasia, and phonological universals]. Uppsala: Almqvist & Wiksell.
- KOOPMANS-VAN BEINUM, F. J., CLEMENT, C. J., & VAN DEN DIKENBERGPOT, I. (2001). Babbling and the lack of auditory speech perception: A matter of coordination? *Developmental Science*, **4**, 61-70.
- 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* (pp. 37-50). New York: Stockton.
- LOCKE, J. L. (1989). Babbling and early speech: Continuity and individual differences. *First Language*, **9**, 191-206.
- LYNCH, M. P. (1996). The case of total deafness: II. Phrasing in the prelinguistic vocalizations of a child born with congenital absence of cochleas. *Applied Psycholinguistics*, **17**, 293-312.
- LYNCH, M. P., OLLER, D. K., & STEFFENS, M. (1989). Development of speech-like vocalizations in a child with congenital absence of cochleas: The case of total deafness. *Applied Psycholinguistics*, **10**, 315-333.
- LYNCH, M. P., OLLER, D. K., STEFFENS, M. L., & BUDER, E. H. (1995). Phrasing in prelinguistic vocalizations. *Developmental Psychobiology*, **28**, 3-23.
- 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.
- MASATAKA, N. (1993). Effects of contingent and noncontingent maternal stimulation on the vocal behaviour of three- and four-month-old Japanese infants. *Journal of Child Language*, **20**, 303-312.
- NATHANI, S. (1998). *Phrases, prelinguistic vocalizations, and hearing impairment* (Doctoral dissertation, Purdue University). *Dissertation Abstracts International*, **60**, 3245B.
- NATHANI, S., LYNCH, M. P., & OLLER, D. K. (1999, November). *Amount of vocalization in typical and deaf infants*. Paper presented at the American Speech-Language-Hearing Association Convention, San Francisco.
- NATHANI, S., OLLER, D. K., & COBO-LEWIS, A. (2001). *Final syllable lengthening (FSL) in infant vocalizations*. Manuscript submitted for publication.
- OLLER, D. K. (1980). The emergence of the sounds of speech in infancy. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child phonology: Vol. 1. Production* (pp. 93-112). New York: Academic Press.

- OLLER, D. K. (1986). Metaphonology and infant vocalizations. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of early speech* (pp. 21-35). New York: Stockton.
- OLLER, D. K. (1991). Computational approaches to transcription and analysis in child phonology. *Journal for Computer Users in Speech & Hearing*, *7*, 44-59.
- OLLER, D. K. (1995). Development of vocalizations in infancy. In H. Winitz (Ed.), *Human communication and its disorders: A review* (Vol. 4, pp. 1-30). Timonium, MD: York.
- OLLER, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Erlbaum.
- OLLER, D. K., & EILERS, R. E. (1988). The role of audition in infant babbling. *Child Development*, *59*, 441-449.
- OLLER, D. K., EILERS, R. E., BULL, D. H., & CARNEY, A. E. (1985). Prespeech vocalizations of a deaf infant: A comparison with normal metaphonological development. *Journal of Speech & Hearing Research*, *28*, 57-63.
- OLLER, D. K., & LYNCH, M. P. (1992). Infant vocalizations and innovations in infraphonology: Toward a broader theory of development and disorders. In C. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 509-536). Parkton, MD: York.
- STARK, R. E. (1980). Stages of speech development in the first year of life. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child phonology: Vol. 1. Production* (pp. 73-90). New York: Academic Press.
- STARK, R. E. (1989). Temporal patterning of cry and non-cry sounds in the first eight months of life. *First Language*, *9*, 107-136.
- STERN, D. N., JAFFE, J., BEEBE, B., & BENNETT, S. L. (1975). Vocalizing in unison and in alternation: Two modes of communication within the mother-infant dyad. In D. Aaronson & R. W. Rieber (Eds.), *Developmental psycholinguistics and communication disorders* (Annals of the New York Academy of Sciences, Vol. 263, pp. 89-100). New York: New York Academy of Sciences.
- STOEL-GAMMON, C., & COOPER, J. (1984). Patterns of early lexical and phonological development. *Journal of Child Language*, *11*, 247-271.
- TREVARTHEN, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech: The beginnings of human communication* (pp. 321-347). Cambridge: Cambridge University Press.
- ZLATIN LAUFER, M. (1980). Temporal regularity in prespeech. In T. Murry & J. Murry (Eds.), *Infant communication: Cry and early speech* (pp. 284-309). Houston: College Hill.

(Manuscript received November 16, 2000;
revision accepted for publication May 10, 2001.)