Chapter 4

AN ACOUSTIC PHONETIC CATALOG OF PRESPEECH VOCALIZATIONS FROM A DEVELOPMENTAL PERSPECTIVE

Eugene H. Buder, Anne S. Warlaumont and D. Kimbrough Oller

INTRODUCTION

From a newborn's cry to a child's first word at approximately 12 months of age, a wide variety of sounds is produced. Many of these have much in common with sounds produced by other mammals, for example vegetative sounds such as coughing and burping, or "fixed signals" (Lorenz, 1951) such as crying due to hunger or other discomfort, or laughter in response to being tickled. Other sounds are clearly more volitional in nature, and are often termed babbling or, following Oller (2000), "**protophones**."

There are many reasons to believe the protophones are precursors to speech: because only humans produce them, because all normally developing infants produce them before speaking, because they occur in a rough developmental sequence, becoming more speech-like across the first year, and because disruption of the normal pattern for age at onset is associated with a variety of developmental disorders (for review see Oller, 2000). Protophones that emerge earliest differ greatly from speech in terms of phonatory, articulatory, and acoustic characteristics. This fact compels us to utilize a special methodology capable of illustrating the relations between protophones and speech. Instead of describing the protophones *as* speech sounds, we address them on their own terms as special pre-speech categories and features. This chapter offers a catalog of protophones from the first months of life to the onset of meaningful speech. State-of-the-art spectrograms and accompanying sound files illustrate each of these protophones, along with a rough outline of orders of occurrence (empirically verified or presumed based on the infrastructural theory) of different protophone groups.

HISTORY OF THE PROTOPHONE APPROACH TO STUDYING INFANT VOCALIZATIONS

The linguist Jakobson asserted that infant babble was random and purposeless behavior unrelated to the development of speech sounds. His "discontinuity theory" (Jakobson, 1941, 1968) was later shown to be incorrect. For instance, the phonological characteristics of canonical babble (e.g., [ba], [ma]. [baba], [dada]) and early words are very similar with respect to syllable types and shapes (Nakazima, 1962; Oller, Wieman, Doyle, and Ross, 1975; Stoel-Gammon and Cooper, 1984; Stoel-Gammon and Dunn, 1985; Vihman, Macken, Miller, Simmons, and Miller, 1985). Moreover, most children produce babbling for several months after the emergence of first words, at least up until age 18 to 20 months (Stoel-Gammon and Dunn, 1985).

Prior to the late 1970s, scientists interested in studying infant vocalizations took one of two approaches. They either transcribed infant vocalizations phonetically, using the same phonetic categories as would be used for transcribing adult speech (Irwin, 1947, 1948), or they measured the acoustic properties of infant vocalizations, using acoustic measurements designed for adult speech (Lynip, 1951). While both methods have their place and continue to sometimes be used by scientists and clinicians alike, in the late 1970s and early 1980s a new way of understanding and measuring infant vocalizations was developed (Koopmans-van Beinum and van der Stelt, 1986; Oller, 1980, 1986; Stark, 1980). This approach focuses on the precanonical protophones, recognizing that these early prelinguistic sounds are more primitive than, and quite distinct from, later sounds that can appropriately be characterized within the International Phonetic Alphabet (IPA). The protophone categories are generally recognizable intuitively by adult listeners (where caregivers are deemed to be the key listeners). Within each infant, the protophone categories that are produced regularly can be thought of as the infant's infraphonological repertoire, representing the infant's capabilities to produce contrastive sounds volitionally.

Taking a protophone-based approach to the study of infant vocalizations, rather than a phonetic transcription approach or pure acoustic measurement approach, has allowed researchers to discover what appear to be fairly universal developmental trends in infant vocalizations across the first year. Researchers have tended to describe these trends in terms of stages. Similar stages have been independently identified by a variety of research groups. A typical characterization is provided below (see "Comments on order of development and ages of onset in protophones").

It is interesting that while there seems to be no dispute about the idea that there are consistent universal patterns, there has been considerable effort to try to determine possible effects differing ambient languages may have on very early vocalizations. Although no one disputes that there are also major similarities in babbling across infants from all language groups that have been studied, there are a number of reports that contend that by the end of the first year there are specific language effects (de Boysson-Bardies, 1999; Levitt and Wang, 1991; Vihman and de Boysson-Bardies, 1994). At the same time there have been a number of failures to determine ambient language effects in the first year (Atkinson, MacWhinney, and Stoel, 1970; Eady, 1980; Navarro, Pearson, Cobo-Lewis, and Oller, 1998). Because the methodology of such research is difficult at best, the matter of whether or not there are ambient language effects on prelinguistic vocalizations remains, in our opinion, unresolved.

The one issue that seems clear is that much of infant vocal development, and presumably the basic scheme of protophone stages, is universal.

Other research on protophone use has focused on infants with hearing impairment or autism (Ertmer and Iyer, 2010; Paul, Fuerst, Ramsay, Chawarska, and Klin, 2011; Sheinkopf, Mundy, Oller, and Steffens, 2000), on patterns of repetition and coordination of vocalizations between child and caregiver (Buder, Warlaumont, Oller, and Chorna, 2010; Warlaumont, Oller, Dale, Richards, Gilkerson, and Xu, 2010), on the influence of social reinforcement on protophone productions (Goldstein and Schwade, 2008), and on automatic classification of protophones (Warlaumont, Oller, Buder, Dale, and Kozma, 2010).

In the glossary below, we describe the major protophone categories currently used in our infant vocalization research. They constitute the kinds of sounds that, in accord with our findings, most infants produce during the first year or so of life as they progress from simple, primitive sounds in the months right after birth to sophisticated sequences of sounds that resemble adult speech to a much greater extent. We focus on the most commonly occurring features of the most commonly occurring protophone categories in infancy.

GUIDELINES FOR CODING PROTOPHONE CATEGORIES

Protophone categories are those primitive categories of sounds that are thought to be speech-related, or precursors to mature speech. Thus we exclude fixed, species-specific signals such as crying or laughter, and vegetative sounds, produced as byproducts of purely physiological events such as feeding, moving the torso, lifting, or sneezing. As speech-related sounds, protophones can generally be divided into two categories: those related to phonation and those related to syllabification and/or articulation.

Three issues regarding our system of coding (and corresponding guidelines) need to be taken into account in order to understand how we intend the term protophone to be understood and how we go about categorization. These concern:

1. Coding at the utterance-level
2. Forced-choice coding
3. The focus on volitionally produced sounds

First, we generally code at the "utterance" level, where utterances are taken to be "breath-groups" (Lynch, Oller, Steffens, and Buder, 1995); that is, an utterance consists of the vocalization occurring on a single egress of the breathing cycle. We have of course coded and analyzed infant sounds at other levels as well (from sub-segmental to super-clusters of utterances), but the protophone terminology applies primarily to the utterance level.

The second issue concerns the fact that in keeping with utterance-level coding, we generally impose a forced choice on coders for both phonatory and syllabification/articulation judgments of protophones—only one code (though sometimes it consists of the combination of two features as in for example squeal-yells or growl-yells) may be used to describe the phonation and one code to describe the syllabification/articulation for each utterance. This approach forces the listener to ignore many characteristics of utterances and to focus instead on only the most salient features of utterances, despite the fact that it is often clear that

multiple features are recognized by the listener. Especially for long utterances the forced choice is often difficult, because, for example, infants may vary their phonation, as well as types of syllabification/articulation, considerably within the utterance. The justification for this coding approach is partly theoretical: parents and other caregivers, the most important listeners in children's lives, tend to focus on global features of infant utterances and to ignore details within utterances. The justification is also partly practical: the time cost of coding is often prohibitive if details are taken into account. As children get older we consider much more syllabic or even phonetic detail for each utterance, because (among various reasons) it is clear that by that time syllabic and phonetic details play important roles in parent-infant communication (see Chapter 5 by Anna Sosa for a description of early speech).

To exemplify this whole-utterance categorization approach, consider that an utterance can contain, in sequence, both a squeal portion and a vowel-like or "**vocant**" portion (in the phonatory realm), but coders are required to indicate one or the other. Similarly an utterance can consist of both a glottal stop syllabification and a canonical syllabification (in the articulation/syllabification realm), but coders are required to choose one of the two. The coder's choice (in accord with our protocol) is guided by a combination of intuition about saliency of the competing features of the utterance, infraphonological theory, and lessons learned from evaluating inter-observer reliability of coding for protophones. For example, if squeal and vocant features co-occur within an utterance, the listener chooses squeal, which is often very salient because it represents a substantial departure from the infant's most common phonatory mode. However, if the squeal portion is deemed by the listener to be very low in saliency, it can be ignored and the utterance can be treated as a vocant. If a glottal stop syllable co-occurs in the same utterance with a canonical one, the listener chooses canonical because canonical syllables are deemed, theoretically, to be more advanced developmentally.

However, this sort of mutual exclusivity in using the coding categories is not an irreversible requirement for the future—it is certainly possible to envision protophone coding procedures where multiple protophone codes are involved (perhaps weighted by their saliency). Such complex coding seems ultimately desirable especially if we wish to characterize fully the relations between protophones and the more advanced forms of real speech. As greater control over the protophones develops, their characteristics gradually become integrated as features of a more elaborate scheme of vocalization. For example, within a single utterance, phonatory and pitch-related features (first seen as squeals, growls and vocants) can come to serve as prosodic accompaniments of richly differentiated syllabic sequences—in adult speech, well-formed sentences or parts of sentences are sometimes produced in a squeal or growl register.

The final issue is that protophone coding focuses on volitionally produced sounds, and because humans can produce any kind of sound volitionally on some occasions (even vegetative sounds such as coughing!), the coder is forced also to make an intuitive judgment about the volitional nature of infant vocalization. This requirement forces the coder to simulate the natural judgment style of caregivers, whose task it is to nurture the infant's well-being and who interpret infant sounds as indicators of state, need and fitness. Caregivers' judgments are taken to be the most important reference point in coding because it is the caregivers who negotiate with infants over the functions and ultimately the meanings of their vocalizations. The volitional nature of vocalization is clearly an important indicator of the progression toward speech in and of itself, and caregivers appear to have a sense of the extent to which infants vocalize intentionally. They naturally recognize the difference between

playful, exploratory sounds and sounds that are motivated by pain or fear, for example, and they understand that the former are more volitional. Further, no other primate shows volitional vocalization to nearly the extent that the human infant does (Cheney and Seyfarth, 1999; Hauser, 1996; Owren and Goldstein, 2008). The extent of volitional control appears, however, to vary continuously across utterances rather than representing a binary (on-off) characteristic of utterances. For example, it is often observed that a single infant utterance may contain both fuss and vowel-like elements. Again, a forced choice is required (in accord with our protocol) in coding to differentiate among the more reflexive sounds of infants (the ones that resemble more the sounds of other primates' warning calls, vocal threats or vegetative sounds) and the more volitional ones, the ones that form more important foundations for speech.

GLOSSARY OF PROTOPHONE CATEGORIES

In this section, we introduce the key protophone categories for coding infant vocalizations. We begin by describing reflexive sounds not included among the protophones. Then we describe protophones in two realms: phonation and syllabification/articulation.

Reflexive Sounds

As noted above, the protophone coding scheme focuses on volitional sounds and excludes reflexive sounds. The latter include cries, fusses, laughs, and vegetative sounds as described below.

Cries and **fusses**. Cry and fuss sounds are present as reflexive, species-specific vocalizations from birth. They are always associated with negative emotional valence and are not considered speech-related. Sound file 4S21 included with Suggested Activities below is an example of the cry of a one-week-old infant.

Laughs. Although they do not emerge until about four months of age, laughs, like cries, are species-specific vocalizations that are strongly associated with a fixed set of social and emotional contexts, in this case positive contexts (Sroufe and Wunsch, 1972). When they are produced in an apparently reflexive way they are considered to be similar to sounds of other mammals and thus to be less speech-related than the protophones.

Vegetative sounds. These sounds include burps, sneezes, coughs, hiccups, grunts, and yawns. Vegetative sounds are unintentional side-effects of various behaviors such as feeding or postural control, and they are not specialized as (that is, naturally selected to be) communicative signals (Hockett and Altmann, 1968). It is worth noting that grunts often require visual (as opposed to just audio) information in order to distinguish them from certain protophones, especially quasivowels or growls. A grunt (as we use the term) is deemed to be produced reflexively (as an accompaniment of movement or physical straining), while the protophones are deemed to be produced volitionally.

Phonation-Related Protophones

Phonation-related protophones are defined in terms of how they are produced at the glottis. These include **eggresive** vocants, squeals and growls, whispers and yells, and ingressive or ingressive-egressive vocalizations. These phonation-related protophones are described below.

Quasivowels and **full vowels**. Both of these are types of vocants. Quasivowels are the more primitive vocants. The characteristic differentiating them from full vowels is the auditory impression that they are produced with the vocal tract in a neutral or "unpostured" configuration. They tend to be acoustically short and quiet, and often include the muffled quality associated with nasality, though they do not necessarily always have all these characteristics. Full vowels tend to be longer and louder. A defining characteristic of full vowel quality is the auditory impression that the vocal tract *is* postured, with deliberate positioning of the mouth and tongue in a speech-like way, yielding a vowel quality distinct from that corresponding to an at rest position of the tract.

The vocants, both quasivowels and full vowels, are also defined by the auditory impression of normal speech-like phonation and pitch (in contrast with the phonation and pitch characteristics of squeals and growls). The phonatory and pitch characteristics of vocants must be within the infant's "habitual" range.

Squeals and **growls**. Squeals differ from vocants in that they have extreme high pitch, beyond the range of the infant's habitual voice (typically twice as high or higher than the mean habitual pitch) and may also be associated with falsetto or "loft" voice. Growls differ from vocants primarily in that they either have very low pitch, below that of the infant's habitual voice, or that they are within the habitual pitch range but have a harsh or noisy quality. The growls that are not harsh or noisy are often associated with pulse register, in phonatory patterns usually dubbed "fry" or "creak," and these tend to be well below the habitual pitch range. It is critical to emphasize that growls are sometimes *within* the habitual pitch range, and can even exceed the perceived mean pitch if they have a harsh, rough or noisy phonatory quality (but do not reach the pitch range of squeals). Squeals can also have vocal qualities departing from normal phonation but (in our definition) always have above-typical pitch and usually quite a bit of pitch variability. Thus, squeals and growls are protophones that are marked by differences from vocants with regard to phonation in terms of pitch and phonatory quality.

If a vocalization contains both a vocant and either a squeal or a growl (occurring in sequence within the vocalization), the listener is encouraged to make a judgment as to whether the squeal or growl characteristic is "salient." If the auditory impressionistic answer is yes, the vocalization is categorized as squeal or growl, not vocant. This judgment depends on the overall auditory impression rather than exclusively on either the relative amplitude or the relative duration of the squeal or growl portion of the sound. Similarly, if a vocalization includes both squeal and growl characteristics, the observer is encouraged to categorize the sound in terms of which characteristic seems more salient, squeal or growl. Some utterances present a relative balance of the features and can thus be difficult to code with high interjudge agreement.

Whispers and **yells**. Whispers and yells are also phonation-related protophones. Whispers are quiet and not voiced, though the air can be heard passing through the laryngeal cavity. Importantly a sound is not judged to be a whisper if it is deemed to be the mere

byproduct of breathing. It must instead be perceived as a volitional sound. Yells are notably loud sounds, and usually have a pressed quality, but the defining quality with yells is that they are produced with amplitude that is beyond the habitual range of the producer's voice.

Ingressive vocalizations and ingressive-egressive sequences. Phonatory protophones with a sound source that is produced somewhere within the larynx (or possibly trachea) by air flowing inward toward the lungs, as opposed to the typical case of phonation in speech which occurs as air is flowing out of the lungs, are termed ingressive. Ingressive vocalizations can occur in isolation, but some infants produce regular coherent sequences of ingressive and egressive vocalization in rapid alternation, where the rate of each ingress and egress is near that of syllabic timing, and where there are scarcely any silences between ingresses and egresses—these sequences are not categorized as multiple utterances (though we use a breath-group criterion to define utterance boundaries in all other cases), but as ingressive-egressive sequences.

Syllabification and Articulation-Related Protophones

Some infant protophones include significant supraglottal activity. When the articulatory tract is constricted sufficiently, a consonant-like sound ("**closant**") emerges. In some cases the supraglottal activity constitutes a sound source that may occur in the absence of phonation or completely concurrent with it. In other cases supraglottal activity can break utterances up into syllable-like chunks. In addition the glottis itself can be used to break up phonatory sequences into syllables. These syllabification/articulation-related protophones include raspberries, clicks, goos, glottal sequences, and both marginal and canonical babbling and are described below.

Raspberries. Raspberries are trills or vibrants formed most often with the lips or the tongue and lips, and occasionally by the tongue body against the toothless alveolar ridge. Such non-phonatory sound sources are associated with certain consonants in a very few languages, but in infancy they are often also produced as extended sounds, with or without concurrent voicing.

Clicks. Clicks are sounds that involve creating a supraglottal sub-cavity that has negative pressure. When the pressure is released, it forms a non-phonatory sound source that is considered in our scheme to be an articulation-related protophone.

Goos. Goos are vocalizations with very primitive articulation where a tongue closure articulation in the back of the oral cavity is superimposed upon phonation, usually normal phonation. While goos do include articulation, the product usually does not yield clear syllabification, but rather seems only to hint at the infant's emerging potential for syllabification by articulation. Typically goos are variable in timing, and thus usually seem disorganized or erratic and inconsistent with the features of well-formed syllabification. However, occasionally, as if by chance, they do show well-timed movements consistent with more mature syllabification. We will return to goos in the following sections on marginal and canonical babbling to provide a few additional tips on how to distinguish goos from those other protophones.

Glottal stop sequences. Interruption of phonation at the larynx can be thought of as the most primitive mechanism of syllabification, because it requires no supraglottal movement. Glottal stop sequences in our coding system always consist of at least two perceived syllables

broken up by one or more glottal interrupts. The coding of glottal stop sequences does not take account of glottal stop onsets or offsets to utterances. Repetitive glottal stop sequences occur when phonation and glottal stops alternate within an utterance and when the phonatory periods are long enough to be deemed "syllables" and the glottal stops are deemed long enough to be syllable margins (that is to have at least the duration that stop consonants have in mature speech). On many occasions the auditorily perceived glottal interruptions actually consist of phonation that merely dips into pulse register (yielding the vocal quality often termed creaky voice or fry), without an extended stop *per se*. The ear seems to interpret this pulse interrupt in much the same way as a full glottal stop.

Marginal babble and **canonical babble**. These are the most advanced of the protophones. They are sounds with syllable-like patterns generated by supraglottal articulation. Marginal babble is distinguished from canonical babble in one of two ways. The first is timing: canonical babble has speech-like speed of timing of the transitions (nominally <120 ms) between consonant-like and vowel-like portions of the syllable, whereas marginal babble does not include such speech-like timing of C to V or V to C. The key factor here is that the listener should judge a syllable as a marginal babble if the transition itself can be directly perceived, that is to say "auditorily tracked," but the listener should judge a syllable as a canonical babble if the transition is so quick that it becomes an integral part of the syllable whole, not auditorily recognizable *as* a transition. Even syllables with glide onsets ([wa], for example) can be (and in real speech usually are) pronounced with rapid transitions (i.e., <120 ms), and such onsets are by this criterion deemed canonical. However, if transitions are drawn out (regardless of the type of consonant-like element) to the point of being slow enough to be perceived as transitions, the syllable is classified as a marginal babble. For example, it is possible to "auditorily track" a 200 ms formant transition starting with lip approximation and rounding and concluding in a low vowel-like sound—such a transition can be variously interpreted auditorily as a sequence of a vowel-like sound resembling [u] followed gradually by a vowel-like sound resembling [a], or as a single unified transition between an initiating glide-like element and a low vowel. Either way, if the transition itself is perceived, we categorize the sound as a marginal babble. The second way that a syllable can be "marginal" is by possessing no full vowel as a nucleus and possibly thus having no substantial transition to track—a quasivowel nucleus adjacent to most consonant-like gestures produces relatively little formant movement and is thus unlike distinctively well-formed CV or VC syllables occurring in natural languages. It should be added however, that quasivowel nuclei can appear in reduced, unstressed syllables in adult speech. Canonical syllables must, then, have a full vowel as a nucleus along with a rapid transition from margin to nucleus.

A note of clarification is in order regarding gooing. In essence gooing is an early developing form of marginal babbling (very often with quasivowel nuclei) and with consonant-like articulations that are predominantly produced in a relatively disorganized way in the back of the vocal tract, probably with the dorsum of the tongue against the soft palate, although this has not been physiologically verified. As infants mature in the first year, the back articulation predominance gives way to more frequent use of anterior articulations and full vowel nuclei become more common. Consequently marginal babbling in 5 and 6 month olds often seems much more well-organized from the standpoint of syllabicity than gooing, which is very common in two and three month olds but becomes rare after four months.

Within canonical babbling, there are subtypes of protophones: isolated syllables, where the infant produces a CV or VC alone in an utterance; **reduplicated babble**, where syllables repeat as in [mama]; and **variegated babble**, where perceived syllables change notably across the sequence, as in [mami] or [mana]. Importantly, we require that reduplicated babbling includes at least one clearly canonical syllable. Usually there are more, but as long as the remaining syllables are perceived to be of similar form, they may in fact be marginal syllables, and the utterance as a whole will still qualify as a reduplicated babble. Again this is a case of utterance-level forced-choice coding. The most advanced (canonical) syllables in the sequence are used as the basis for the overall judgment. Variegated babble requires at least two canonical syllables, and they must be deemed different by the listener. This is another case of required intuitive judgment by the listener, who is encouraged to react as a caregiver might. A sequence of similar syllables is deemed reduplicated and minor details of articulation are ignored as long as the overall impression is that the infant intends the syllables (at least one of which is canonical) to be the same category. A sequence of syllables is deemed variegated if the overall impression of the listener is that at least *two different* canonical syllables were intentionally produced by the infant. Reasoning supporting this intuitive approach to differentiation of syllable types in canonical babbling has been presented in Oller and Griebel (2008) and Ramsdell (2009).

A question that often arises in the context of research on canonical babbling is, "What kinds of phonetic elements occur?" We prefer to focus this discussion at the syllabic level, because we find little reason to believe that infants in the first year possess a segmented syllabic structure, that is, we presume they possess no distinction between consonants and vowels, but rather produce syllabic wholes (MacNeilage, 1998). Further, it appears that the most accurate way to portray early syllables is in underspecified phonetic terms focusing at the syllabic level. Using this approach, it can be said that common canonical syllables of the end of the first year of life, the syllables that appear to be intentionally produced by infants and consistently recognized by caregivers, are very few in number. They virtually always include a CV-like structure where the offset portion is usually perceived as a relatively low or central vowel-like element, and where the onset is typically a labial or alveolar stop or nasal, or a glide-like onset, either labial or lingual. Less frequently velar stop onsets or fricative-like onsets are seen. Thus the number of syllabic contrasts that commonly occur in infants in the first year represents a mere handful of the global patterns that appear to occur across a wide variety of ambient languages.

Comments on Order of Development and Ages of Onset in Protophones

There have been numerous publications addressing stages of development of protophones (e.g., Elbers, 1982; Holmgren, Lindblom, Aurelius, Jalling, and Zetterstrom, 1986; Koopmans-van Beinum and van der Stelt, 1986; Stark, 1981). While the stage models that have been proposed are far from identical, they tend to agree on a series of key points. Here is a summary of stages consistent with a review by Oller (2000), who sought to encapsulate key points of agreement among the published models:

1. Crying and vegetative sounds are present from the first day of life, but laughter (even though it is considered a fixed signal) shows an onset at around four months.
2. The earliest protophones, including quasivowels and occasional glottal stop sequences, occur right after birth (the phonation stage).
3. At one to four months of age gooing occurs, usually in face to face interaction with caregivers (the primitive articulation stage).
4. By three months many new protophones begin to appear, especially full vowels, raspberries, squeals, growls, yells and whispers (the expansion stage). During this period marginal babbling also emerges.
5. Usually in the second half year of life canonical babbling appears (the canonical stage). It has not been empirically proven that reduplicated canonical babbling precedes variegated babbling (Mitchell and Kent, 1990; Smith, Brown-Sweeney, and Stoel-Gammon, 1989), even though other authors have presumed that it does (Elbers, 1982; Oller, 1980). It may be that the lack of confirmation thus far regarding the presumed ordering is due to reliance on traditional phonetic transcription as a means of trying to determine whether infant utterances should be deemed reduplicated or variegated. That method may obscure global patterns because of focus on phonetic details that may be of little relevance to the infant's intended syllabic categories (or the caregiver's perception of the infant's intended categories). In other words, phonetic transcription in such cases may cause a failure to see the forest for the trees. Our expectation is that as more infrastructurally sensitive methods of coding come into play—methods that emphasize global patterns presumably intended by infants and recognized by caregivers—it may be possible to confirm the presumed ordering of reduplicated babbling preceding variegated (see again (Oller and Griebel, 2008; Oller and Ramsdell, 2008, July; Ramsdell, 2009)).

Infant vocalization research has shown clearly that late onset of canonical babbling (after 10 months of age) is grounds for clinical concern (Eilers and Oller, 1994). Infants with severe or profound hearing impairment show such late onset, and hearing infants with similar delays are at risk for late onset of talking. Considerable research is focused on a variety of clinical groups (including autism, Down syndrome, and William's syndrome) to try to determine other characteristics of early protophone development that may be of clinical significance (Lynch, Oller, Steffens, Levine, et al., 1995; Masataka, 2001; Paul et al., 2011).

## AN ACOUSTIC PHONETIC CATALOG OF SELECTED PROTOPHONE EXAMPLES

The following pages review 20 examples of protophones, displaying waveforms, spectrograms, and occasionally amplitude-by-frequency spectra of the sound files associated with this chapter. To duplicate a typical coding experience, the reader may wish to first listen to the sound files, and then consult the displays and commentaries for guidance. The emphasis is on audible and visible acoustic correlates of the defining characteristics of primary protophone types as reviewed in the Glossary section of this chapter above. Note that protophone coding is essentially based on auditory impressions and does not require

spectrography in our standard approach. However, consultation of acoustic visualizations, such as waveforms and spectrograms, can be helpful in learning and making protophone judgments, can be useful for expository purposes, and may lead to more detailed evidence for the mechanisms of production (and may even lead to a systematically "infraphonetic" level of classification). In our examples, we have focused on the characteristics that we believe require additional guided practice; thus, we have omitted examples of the non-protophone categories and protophones whose characteristics are relatively obvious, such as whisper and click. We include only a few examples of canonical babbling, because canonical syllables are relatively easily recognized (Oller, Eilers, and Basinger, 2001). We also include only one episode of reduplicated babbling, and do not present examples differentiating reduplicated babbling, variegated babbling, and **gibberish**, because the definitional issues related to these categories are less well resolved than in other cases. The examples are arranged roughly in order of the ages at which they tend to first appear, but exceptions are made to emphasize contrasts and to point out that there is extensive overlap amongst the stages of vocal development, especially as protophone characteristics developing earlier are typically available, and utilized, by older infants (or even adults).

It is hoped that simply listening and reading the labels and basic definitions will give readers the core experience needed to build skills and to achieve a basic level of reliability in coding infant vocalizations. Many technical terms used to describe the spectrogram and waveform have been introduced in Chapter 2. There is also much technical "subtext" in the discussion of these examples, which may presume familiarity with more advanced speech-scientific theories and methods; we encourage motivated readers to consult references listed with the examples below to achieve a deeper understanding of the technical aspects of the displays and source-filter mechanisms underlying vocal production. For additional general support the reader may benefit from consultation of textbooks in acoustic phonetic theory and instrumentation (e.g., Baken and Orlikoff, 2000; Johnson, 2003; Kent and Read, 2002; Stevens, 1998). Note, finally, that we do not restrict ourselves to focusing purely on the features that would determine the forced-choice coding of the utterances in our formal coding procedure for the phonatory and syllabification/articulation realms; we provide considerable additional detail about the vocalizations *per se*.

4S1 through 4S20 Protophone Examples 1 through 20.
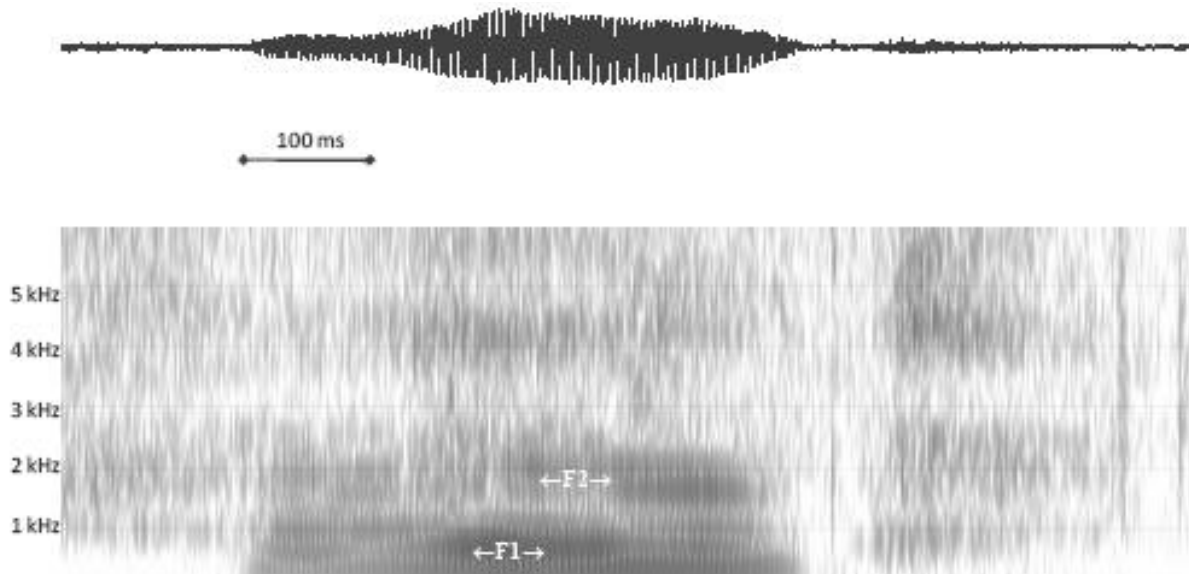
Example 1. A quasivowel, produced by an infant aged 3½ months. Quasivowels are produced with a neutral vocal tract configuration and are typically quiet and short. In this and all other spectrogram displays to follow, horizontal grid lines mark 1 kHz frequency intervals, frequency scales will be provided, and a 100 ms duration reference line will be placed at utterance onset to indicate the time scale.

What we can learn from careful listening: This quasivowel is fairly brief (at approximately 450 ms), it is audibly quiet, and the vowel quality is relatively neutral and stationary.

What we can learn from the acoustic analysis: First of all some technical notes for orientation to formant observations—the bandwidth of analysis in this particular spectrogram is 600 Hz. Normally such a wide analyzing filter renders formant frequencies clearly and obscures harmonics. However, because the fundamental frequency of this infant is quite high compared to an adult (i.e., 400 Hz), this interacts with the analyzing filter to produce a horizontal texture that obscures the formant frequencies somewhat (Buder, 1996)—see also some notes at the end of this paragraph regarding the 'muffled' appearance of formant structure. Nonetheless, F1 appears to be around 800 Hz (especially salient in the middle) and F2 appears to be around 1800 Hz (especially salient at the end). The faint band of energy around 4 kHz is probably F3. It is important to recognize that the "neutral" vocal tract posture characteristic of quasivowels is not the central vowel schwa, which would produce more even formant frequency spacing, and the relatively low F2 of this production is consistent with a slightly back tongue position that seems likely for an infant 'resting' oral anatomy. Several other common characteristics of the quasivowel are visible here: softness and breathiness of phonation accounts for the dominance of low energy and an absence of any harmonics above 2.5 kHz, and it is likely that a slightly open velo-pharyngeal port introduces nasality. Nasality damps (i.e., muffles or reduces) the overall sound energy, broadening the formant bandwidths and making them difficult to discern both spectrographically and audibly, and coupling to the nasal cavities also can introduce anti-resonances that disrupt the clarity of more basic vowel formant structure (Buder, 2005). In addition to overall quietness, evidenced by the sound energy being not much above the background noise level, these phonatory source and upper vocal tract filter characteristics combine to create the overall muffled energy distribution (wider band resonances) and low frequency emphasis in this sound.
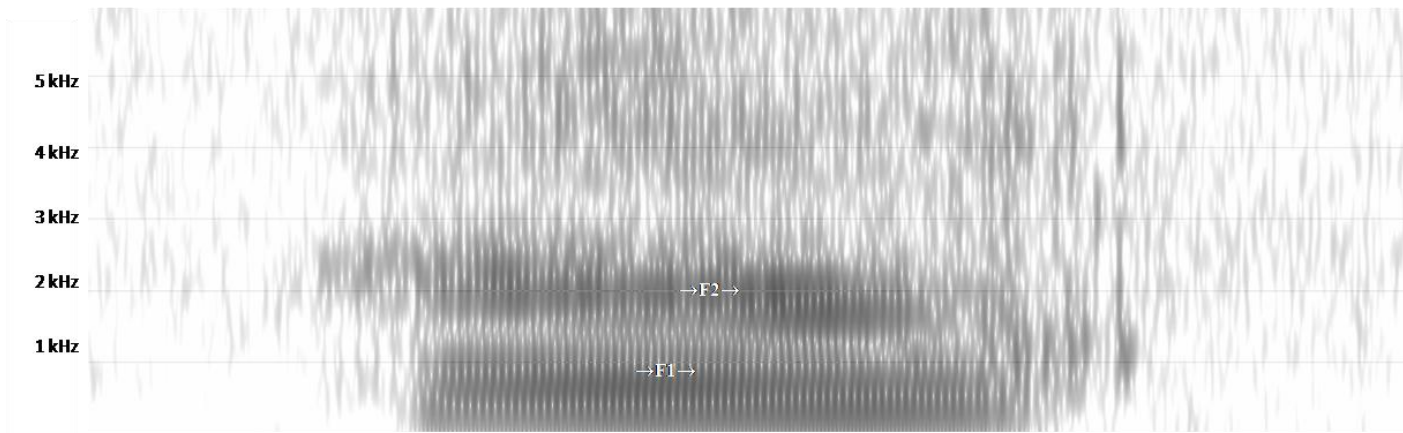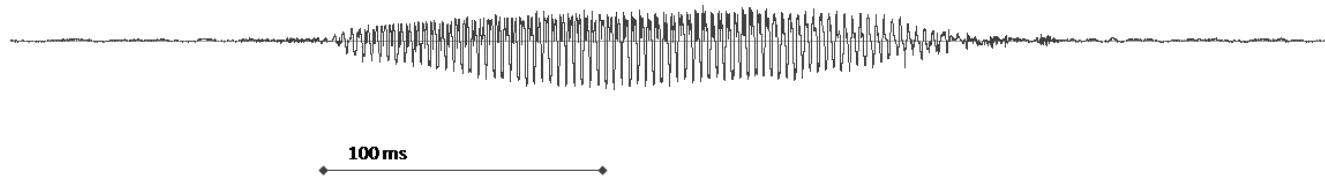
Example 2. Another quasivowel, produced by the same infant at the same age (3½ months) as Example 1.

What we can learn from careful listening: This vocalization was produced with just slightly more phonatory energy and perhaps less nasalization than Example 1. It may be heard to have a slightly different resonance quality but it is still impressionistically unarticulated (unpostured) as a vowel and thus has the neutral sound characteristic of quasivowels.

What we can learn from the acoustic analysis: The same waveform scale and spectrogram range and bandwidth settings from Example 1 have been used. Formants 1 and 2 are more visible in part because there is less damping due to nasal resonance. F1 and the F2 do not shift during the production—note the complete absence of any significant formant movement in the nucleus of this sound. When no obstruent turbulence source energies occur at the onsets or offsets of vocalizations, they would be coded as quasivowels only. In this example, it is possible to hear some turbulence at the end possibly resulting from frication. Thus, this particular example might also be coded as a goo if it is perceived that the frication is based on constriction of the tongue dorsum. The vertical striations following Example 2 are likely glottal in origin, but some audible and visible impression of slight formant movement at the very end of this example suggests the possibility of vocal tract movement towards closure.
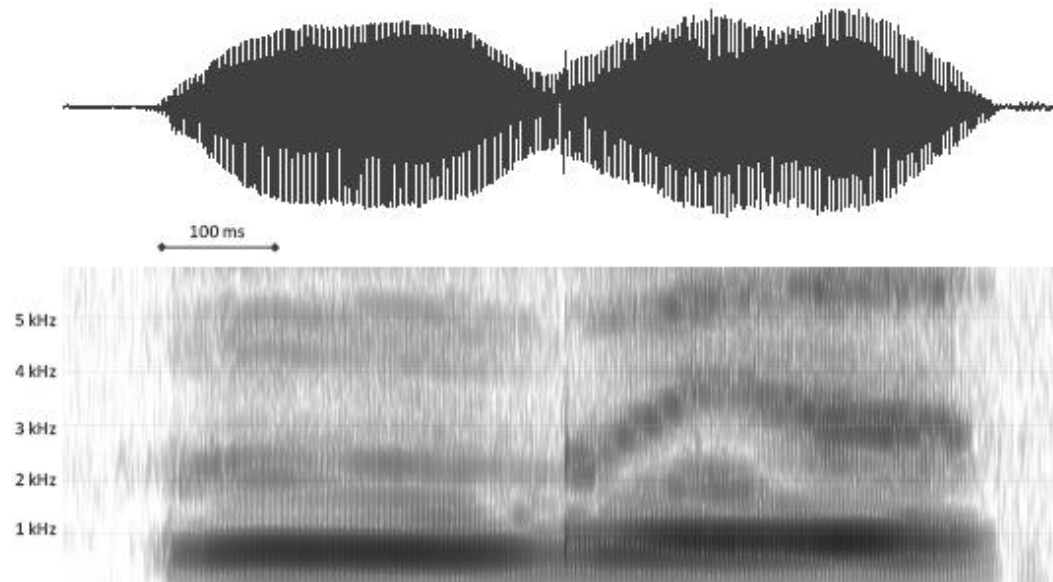
Example 3. A goo, preceded by a quasivowel, and followed by a full vowel, produced by the same infant at the same age as the two preceding examples.

What we can learn from careful listening: We hear stronger phonation and most importantly articulation. Characteristic of the typical goo, this articulation can be heard as involving dorsal tongue contact probably to the soft palate.

What we can learn from the acoustic analysis: The same display settings as Examples 1 and 2 are again used here. We see stronger phonation evidenced by the greater amplitude as well as the greater excitation of the higher frequency resonances and clearer formant structure. We also see articulation, which can be heard as involving dorsal tongue contact probably to the soft palate. It is likely that the velo-pharyngeal port is acoustically and aerodynamically closed for at least the second half of this production. Evidence for this closure can be found in the sharp closing and release-like event that is visible about half-way through the vocalization, caused by some other oral valve. This stop-gap and pressure release would only manifest itself if the VP port is closed; the closure is also evident in the attenuation of the waveform and reduced higher frequency energy in the spectrogram during the closure, and in the faint appearance of a plosive-like burst following this evidence of closing. Aerodynamically no such events can occur if significant air is leaking through the nasal cavity. We can also hear and see evidence of this closure in the relative clarity of the formants which would result from relatively more sharply-tuned resonances that should occur when there is reduced acoustic damping by the nasal cavity, but there is also stronger phonation and a somewhat lower fundamental frequency, which also helps to render those formants more visible. The clear formants also help to display formant movement quite clearly in the latter half of this vocalization, characterizing the full vowel as having been produced with a non-neutral vocal tract position. As we will see in other examples, movement is not necessary for this characterization, but here it audibly and visibly involves an anterior gesturing with the tongue (seen in F2 movement) and some lowering of the jaw (seen as F1 elevation). Goos can have only slow transitions and thus do not include canonical syllables. The slow transitions imply that goos can be thought of as a type of marginal babble that tends to occur particularly early in life and generally includes back articulation.
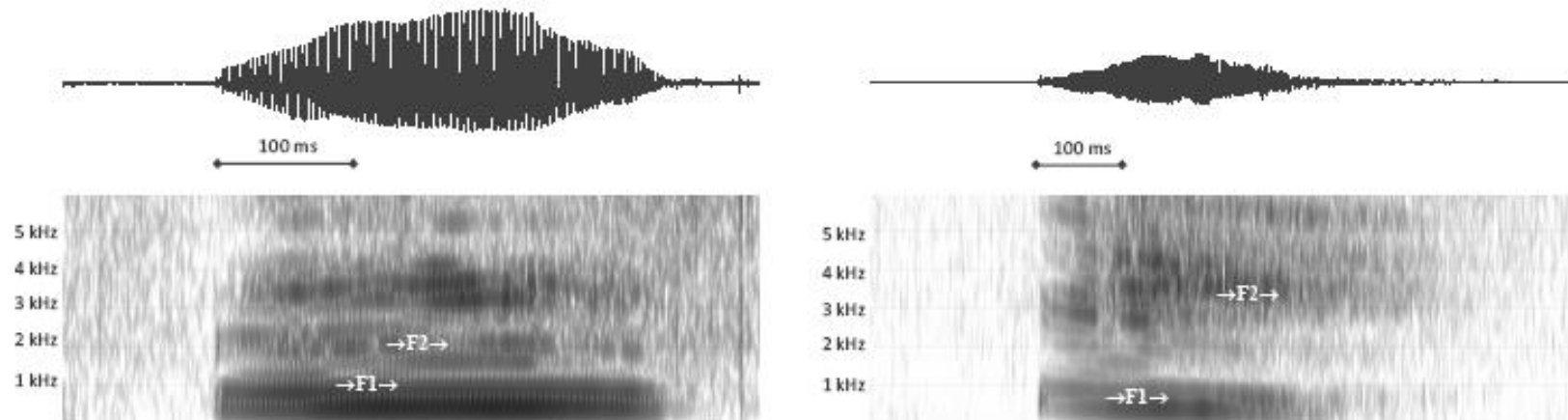
We categorized Example 4 on the left as a quasivowel, and Example 5 on the right as a full vowel. These examples are displayed here side by side to reinforce some additional principles and distinguishing characteristics regarding quasivowels and full vowels as isolated types of vocalizations (i.e., as nucleus-only vocalizations lacking syllabic margins).

While Example 5 has been coded as a full vowel, rather than as a quasivowel, it is not the most unambiguous of cases. Examples 4 and 5 were produced by a much older infant, aged approximately 11 months, which reinforces the fact that early-emerging protophones may still occur later in development even though infants this age will normally also be producing fully articulated canonical syllables.

What we can learn from careful listening: The pitch of the two examples is lower than in previous examples. The quasivowel of Example 4 has no clearly non-neutral vowel, but in Example 5 it is possible to hear a sudden onset (hard glottal attack) and a relatively clear non-central vowel articulation.

What we can learn from the acoustic analysis: The formants are generally a bit lower than in previous examples, consistent with elongation of the vocal tract that will have occurred by this age. Also note that the full vowel on the right, unlike that previously seen in Example 3, is produced without marginal articulations (although it audibly and visibly does begin with a hard glottal attack), or with much change in vocal tract shape, and therefore it has mostly steady formant frequencies. Example 4 on the left, classified as a quasivowel, illustrates a more neutral formant pattern and can therefore be viewed as a baseline against which the full vowel example can be seen and heard as having a postured vocal tract. It is also noteworthy in these examples that the full vowel is in fact produced more quietly, with soft and breathy phonation, and is shorter in duration than the quasi-vowel on the left. Quasivowels are often produced as relatively quiet and short protophones, occurring most often in younger infants whose respiratory and phonatory mechanisms are less well developed, but Example 4 is an exception. Finally, note that there is also less evidence of nasality in these older infant's productions.
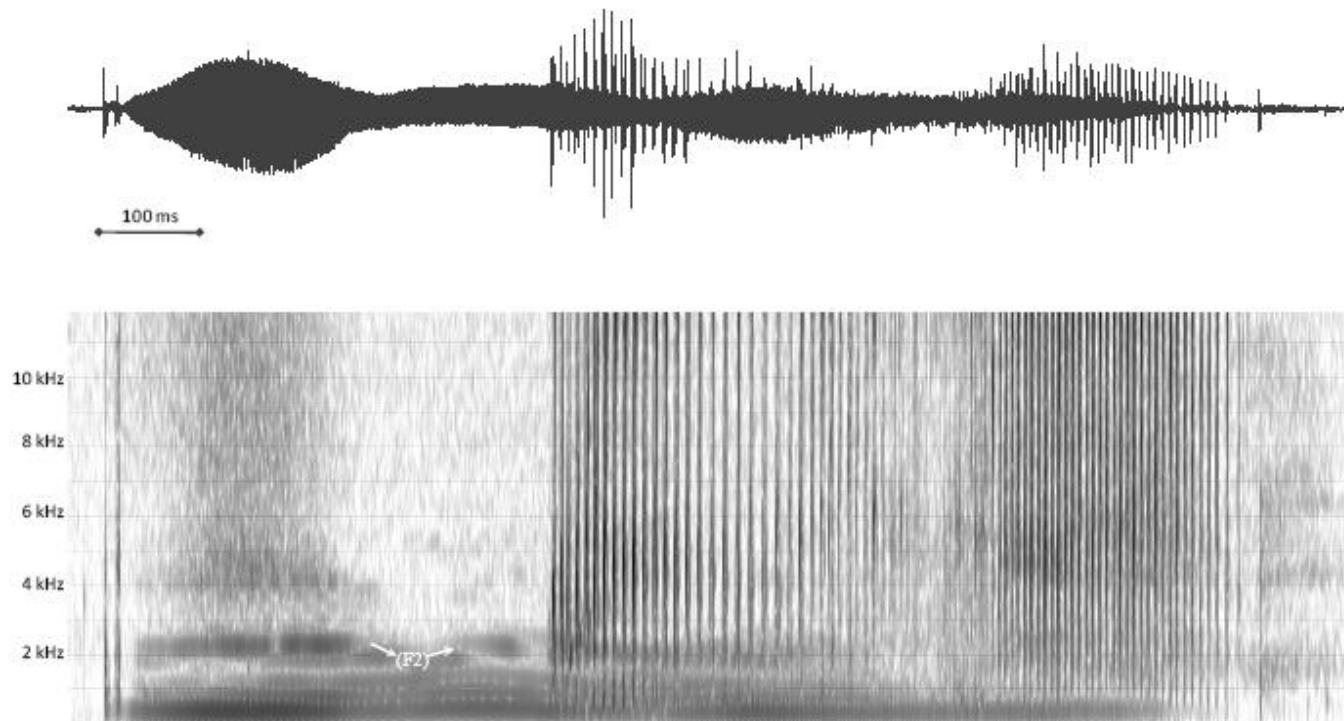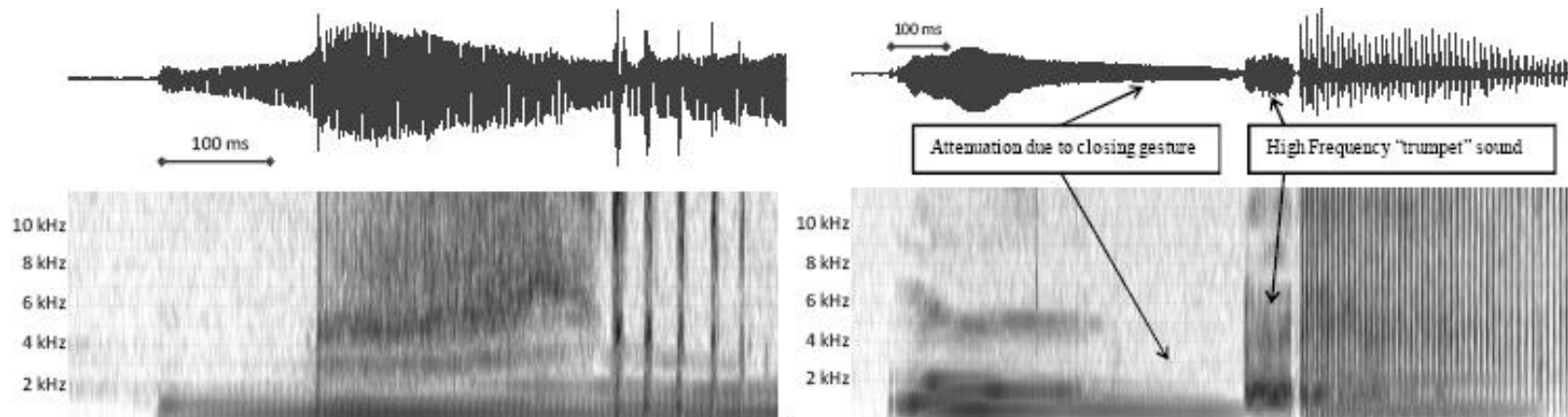
Example 6 returns to the younger infant from previous examples (aged 3½ months), and illustrates a more complex sequence that includes a new protophone type, the raspberry.

What we can learn from careful listening: The first portion begins with a burst at the onset which sounds labial followed by some vowel articulation. The second portion has rapid vibration that is characteristic of a raspberry. It is interesting to note that although most infants produce raspberries, as noted in the glossary above they correspond to a phoneme that is rare in adult languages.

What we can learn from the acoustic analysis: For this display the spectrogram frequency range has been expanded to illustrate some of the higher frequency components that occur with "closant" types of sources like trilling/vibrating at the lips, as seen in the latter 3/5 of this vocalization. The first portion of this vocalization is a goo, with bursting at the utterance onset (that sounds velar), followed by a vowel nucleus, and then by a quieter and somewhat higher pitch segment with evidence of formant movement in the resonance near 2kHz, which is probably F2 (see label in spectrogram). The latter portion of this vocalization is a raspberry; acoustic manifestations of rapid vibrating of the lips are clearly evident in the spectrogram. Phonation continues during this raspberry, and while phonation often occurs in raspberries, it is not a defining characteristic of this protophone type.

100 ms

100 ms

100 ms

Attenuation due to closing gesture

High Frequency "trumpet" sound

10 kHz
8 kHz
6 kHz
4 kHz
2 kHz

10 kHz
8 kHz
6 kHz
4 kHz
2 kHz

Examples 7 and 8 are presented here side by side to compare varieties of raspberry. All the raspberries presented in examples 6-8 were produced by the same 3½ month old infant during one 20 minute recording session. One possible interpretation of this phenomenon is that the infant was exploring her ability to produce sounds by controlling pressures and airflows with a valving mechanism other than at the glottis (i.e., at the lips). Depending on the speed of labial vibration a raspberry may be a *trill*, a *vibrant* with salient roughness (as in Example 6), or even a very high frequency vibrant that produces a tonal characteristic, like the lip action used in playing a trumpet.

What we can learn from careful listening: In Example 7 there is first voicing during bilabial closure, then bilabial frication concurrent with phonation, then bilabial trilling with phonation (this is especially visible during the trilling in this example). In Example 8 the full vowel phonation is followed by a raspberry characterized by rough vibration, i.e., a vibrant type of raspberry.

What we can learn from acoustic analysis: The spectrograms displayed on this page were obtained with an extremely broad analyzing bandwidth of 1000 Hz, which helps to resolve the fine temporal details of these productions (see basic acoustic phonetic resources for more on the topic of "time/frequency" tradeoffs in spectrography). Measurements of the periodicity in the labial vibrations may help provide acoustic phonetic bases for auditory classifications of trill vs. vibrant type. The raspberry segment at the end of Example 7 is classified as a trill, with a frequency of about 30 Hz. There are actually at least two raspberry-type phenomena at the end of Example 8: a fast vibrant type is evident through the last third, with vibrations occurring at approximately 100 Hz. Prior to this, after the initial phonated full vowel segment becomes fully attenuated, there seems to be an extremely high frequency vibrant lasting about 85 ms. Note that in Example 7 there is first voicing during bilabial closure, then bilabial frication concurrent with phonation, then bilabial trilling with phonation (this is especially visible during the trilling in Example 7). In Example 8 the full vowel phonation is marked by falling formant frequencies (beginning about midway through the full vowel) associated with attenuation of the output, which are both clear correlates of closing at the lips, followed by the tonal vibrant which must be produced with very high pressure behind a tight closing, which then yields to a rough vibrant probably associated with a slight reduction in the resistance of the bilabial valve (with a very brief complete closing in between).
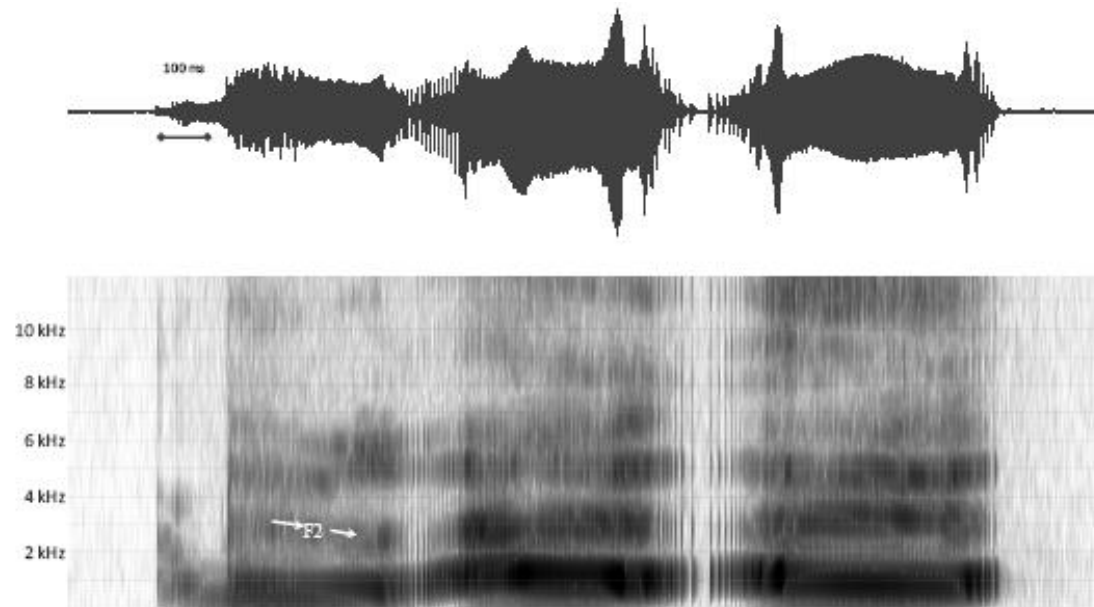
Example 9 introduces a *glottal* sequence. This vocalization is one of many produced in a single 20 minute session by an infant nearly 4 months of age who, similarly to the previous infant, seemed to be practicing with valving operations, but here, glottaly. Because the beginning of the utterance includes a presumably velar closure, the utterance would best be characterized within the forced choice system as a goo (the most advanced protophone category to which the utterance pertains), but here we focus on the characteristics of glottal stop sequencing as another kind of articulatory feature.

What we can learn from careful listening: This example begins with an oral closure, then a full vowel that is interrupted by short silences heard auditorily as glottal gestures that rapidly suppress phonation.

What we can learn from the acoustic analysis: The spectrogram is again very wideband to emphasize temporal structure. The two interruptions of phonation dividing this vocalization into three segments are produced by glottal gestures that decelerate phonation, clearly visible as glottal pulses between the segments. The first glottal interruption is essentially a continuous yet brief dip of fundamental frequency in and out of pulse register, while the second effectively stops phonation for an instant and so might be considered a full glottal stop—see Buder, Oller, Chorna and Robinson (2008) for further details on these and other types of vibratory regimes found in infant phonation. The initial segment of this sequence is a goo, with a presumably velar onset transitioning slowly into a full vowel; some articulatory movement continues during this vowel, visible as a slowly falling F2 (labeled). The second and third segments which are broken up by the glottals are also full vowels: even though the formants appear to be evenly spaced, the audible vowel qualities that justify the full vowel categorization in both cases reinforce the fact that even spacing does not characterize the neutral posture in infant vocants. These spectrographic images also lack the lower frequency emphasis and broadband hallmarks of typical quasivowels. The second and third "syllables" here provide the most unambiguous full vowels (as opposed to quasivowels) to this point in our examples.
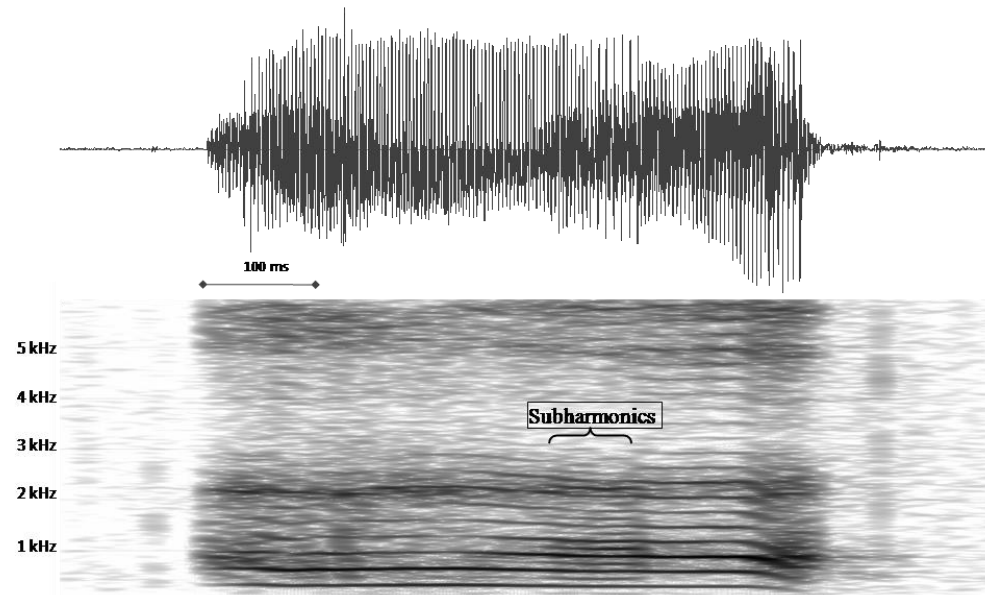
Example 10 is a *growl*, produced by a different infant than example 9 but again at about 4 months of age.

What we can learn from careful listening: Fairly low pitch and rough phonation is maintained throughout the utterance. It is not necessary for growl characteristics to last for an entire utterance to mark it as a growl vocalization; the requirement is substantial salience of the growl characteristic.

What we can learn from acoustic analyses: This display is our first example of an extremely narrowband spectrogram (10 Hz bandwidth) which is useful for discerning the frequency detail of harmonic structure, and therefore the types of vibratory dynamics that characterize distinct *regimes* used by Buder et al. (2008) to classify types of infant phonation. Fairly low frequency and rough phonation is maintained throughout the utterance and three additional features characterize the current example as a clear growl: A. There is interharmonic noise throughout, giving the sound a harsh characteristic. B. Due to this interharmonic noise and/or variations in periodicity affecting the harmonics themselves, harmonic structure is not always clear—note the instability especially at the onset and the lack of patterning above the 7th harmonic in most of the first half. C. There is an episode of subharmonics about 2/3 of the way through the utterance, marked by the appearance of distinct but relatively lower intensity harmonics appearing in between the main harmonics. During this episode there is resemblance to the pulsing waveform to the previous example of glottal interrupts, which results from the appearance of a dramatically low reduction in fundamental frequency in both cases, and both productions can be heard to involve medial compression at the glottis. Unlike pulsed phonation due to glottal interruption, however, in which the voice drops more or less smoothly to just one very low fundamental frequency, subharmonics in phonation can be perceptually noted as a distinct kind of roughness created by the concurrent presence of two sets of harmonics that appear abruptly (at octave relationships in this "period-2" type of subharmonic regime); this leads to some ambiguity regarding the 'true' fundamental frequency.
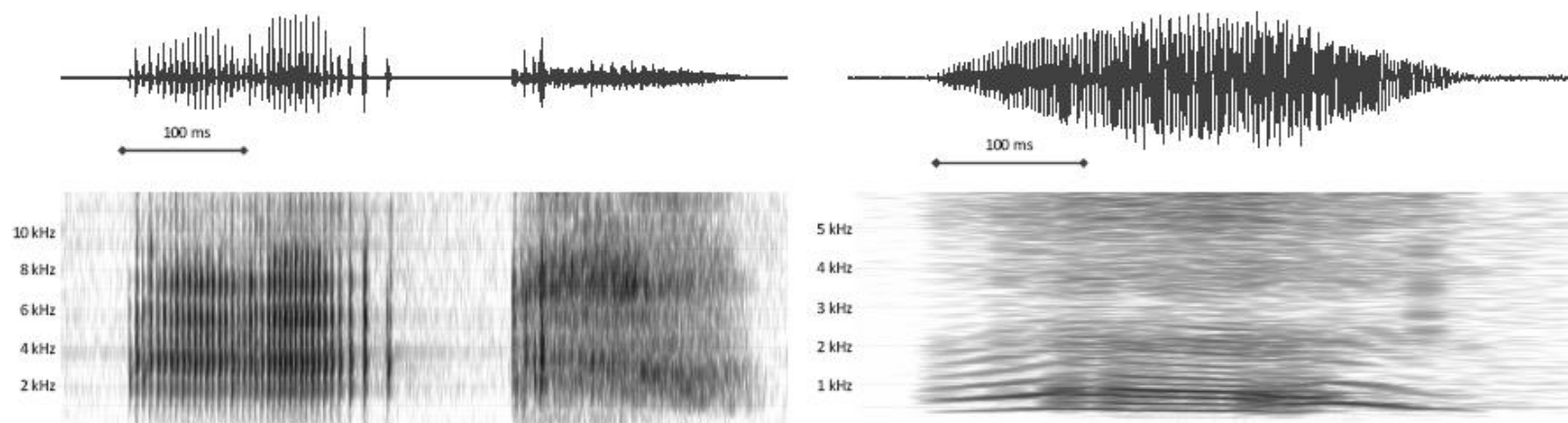
Examples 11 and 12 above are two more examples of growls and help illustrate that the term encompasses quite a range of possible variations from normal phonation. The growl on the left was produced by an older infant at about 10 months of age, and the growl on the right by a different infant at about 4 months of age.

What we can learn from careful listening: Both examples have rough vocal quality that characterizes them as growls. Example 11 can be compared with the glottal sequence of Example 9 to illustrate the principle that pulse phonation can be sustained and salient enough in an utterance to justify the growl categorization (there is a also a full glottal stop in this utterance, followed by an expiratory release, like a "glottal plosive").

What we can learn from acoustic analysis: Example 11's spectrogram is wideband for temporal resolution of pulses, but Example 12 on the right is again displayed with 10 Hz bandwidth analysis to reveal the complex harmonic patterning that dominates the central portion of the utterance. The utterance begins with modal phonation and ends with a higher frequency loft sound as the amplitude is dropping, and these portions include relatively little complexity or interharmonic noise. In contrast, the central portion of example 12 is extremely unstable, with alternating patterns of period-2 and period-3 subharmonics and ending with virtually chaotic vocal fold vibration just before jumping into the loft register. This central portion is the most salient auditorily and thus marks the vocalization as a growl. Note that there is an audible click near the end of the phonation (marked on the spectrogram), which causes some irregular bands of energy to appear in the higher frequencies (2.5 kHz and higher) of the spectrogram; this is almost certainly an environmental sound not produced by the infant.
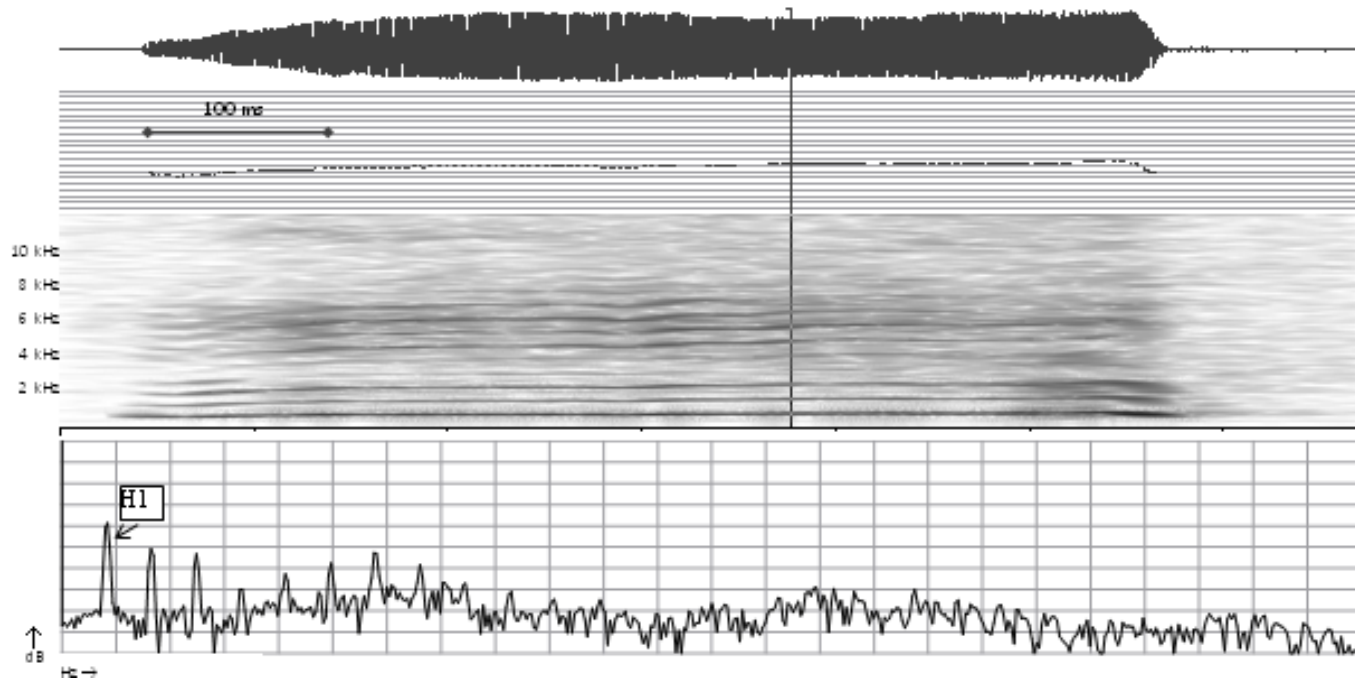
Example 13 is a squeal, produced by the same 4-monther who produced the growl of Example 12 (and recorded in the same session).

What we can learn from careful listening: Most definitively, this squeal is produced loudly with a very high pitch. This high pitch clearly marks the example as a squeal.

What we can learn from acoustic analysis: We have provided a fundamental frequency (f_0) trace between the waveform and narrowband spectrogram; the grid marks in the f_0 panel are at 100 Hz intervals, and at the point where the vertical cursor has been placed it can be seen that the f_0 is above 800 Hz. From this trace, it is clear that the f_0 is far above this infant's (or any typical infant's) modal f_0, a hallmark of squeal. Acoustic analysis of f_0 in infant squeals may be difficult for some analyzing software's algorithms, especially if they are not designed to accommodate such high frequencies through adaptive algorithms or parameter controls that allow the algorithm to be tuned to the appropriate range. Narrowband spectrographic or spectral displays shown above can be used to further validate the f_0. In the bottom spectrum panel of this figure (with intensity in dB on the vertical axis (10 dB grid) and frequency in Hz on the horizontal axis (1 kHz grid)), the first harmonic is f_0 and this can be seen to occur just below 1 kHz. Examination of its higher harmonics help to confirm that the f_0 is about 800 Hz (in voiced sounds, harmonics appear at both odd and even integer multiples of the first harmonic). Another reason that the spectrum is displayed in this example is to illustrate a typical aspect of the "loft" regime, which is an audible characteristic of this particular squeal: the first harmonic is much stronger in amplitude than any of the successive harmonics. In modal voice the first harmonic may be equal to or even weaker than the second (but upper harmonics should show an overall general decline in amplitudes for all voices).
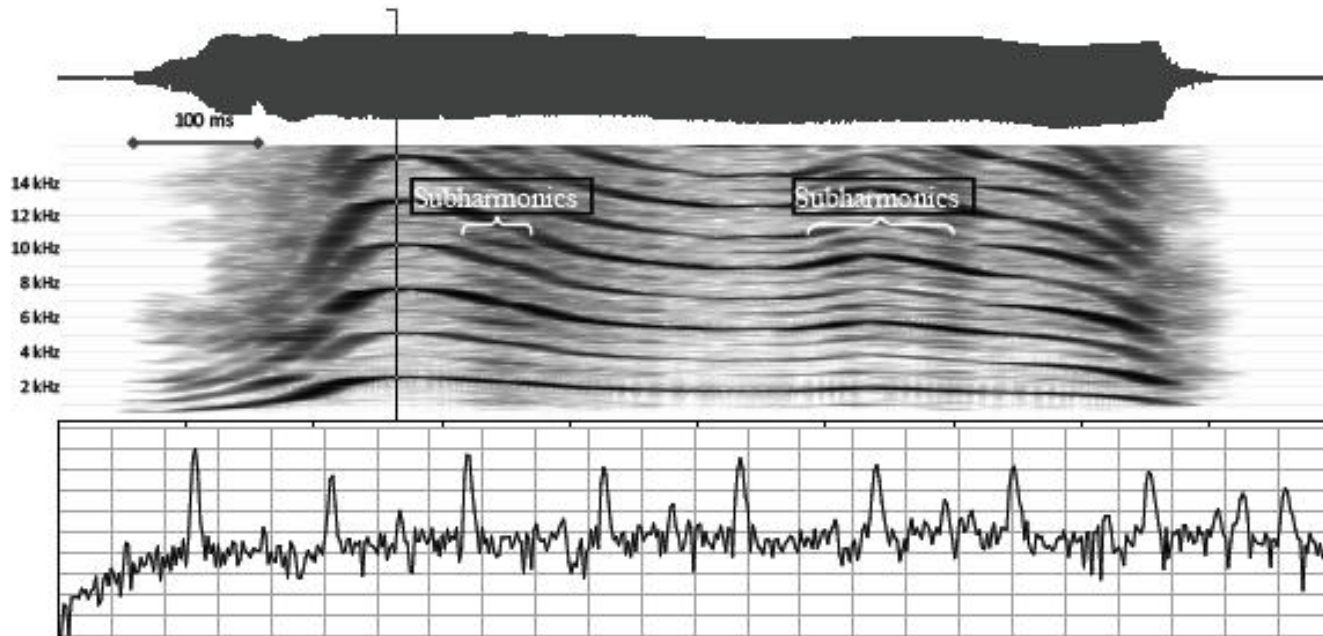
Example 14 is another squeal, produced by the same infant in Example 13, but at 6 months.

What we can learn from careful listening: Watch out! This impressive vocalization is very loud, even after scaling it down from its original recorded level. It is also very high-pitched. Thus, this vocalization may be considered to be both a "squeal" and a "yell."

What we can learn from acoustic analysis: The default algorithm of the commercially available TF32 program could not track this squeal, which is why there is no f_0 contour displayed here. Instead we have presented the spectrum taken at the cursor, where the harmonic time contours in the spectrogram peak. The spectrum reveals that the harmonics are more than 2500 Hz apart and the seventh harmonic at 18 kHz, so the f_0 is approximately 2600 Hz (18000/7 = 2570) at its peak. Still, we have observed even much higher f_0 produced by infants of this age. In comparison to Example 13, it can also be seen that the energies of the harmonics decline only slightly; even though this is an extremely high-pitched phonation it is also quite pressed-sounding, and is definitely not in loft register. It is primarily the high f_0, typically accompanied by extensive f_0 movement, that characterizes squeals. Inspection of the spectrogram reveals incipient subharmonics in this phonation of period-2 just past the first peak and apparently even period-4 during the second pitch peak. However, they are not audible as the double-pitched kind of roughness associated with fully developed subharmonics. These patterns might even be related to the fact that the intensity of this phonation exceeded the linear range of the microphone used to record it, causing analog amplitude clipping which accounts for the strange flatness of the waveform contour; the concomitant distortion may have exacerbated the appearance of subharmonics. This is a known issue in obtaining and interpreting infant vocalizations: in order to pick up reasonably salient quiet vocalizations (cf. Quasivowel in Examples 1 and 2), one must allow for distortions at the higher end of infants' amazing range.
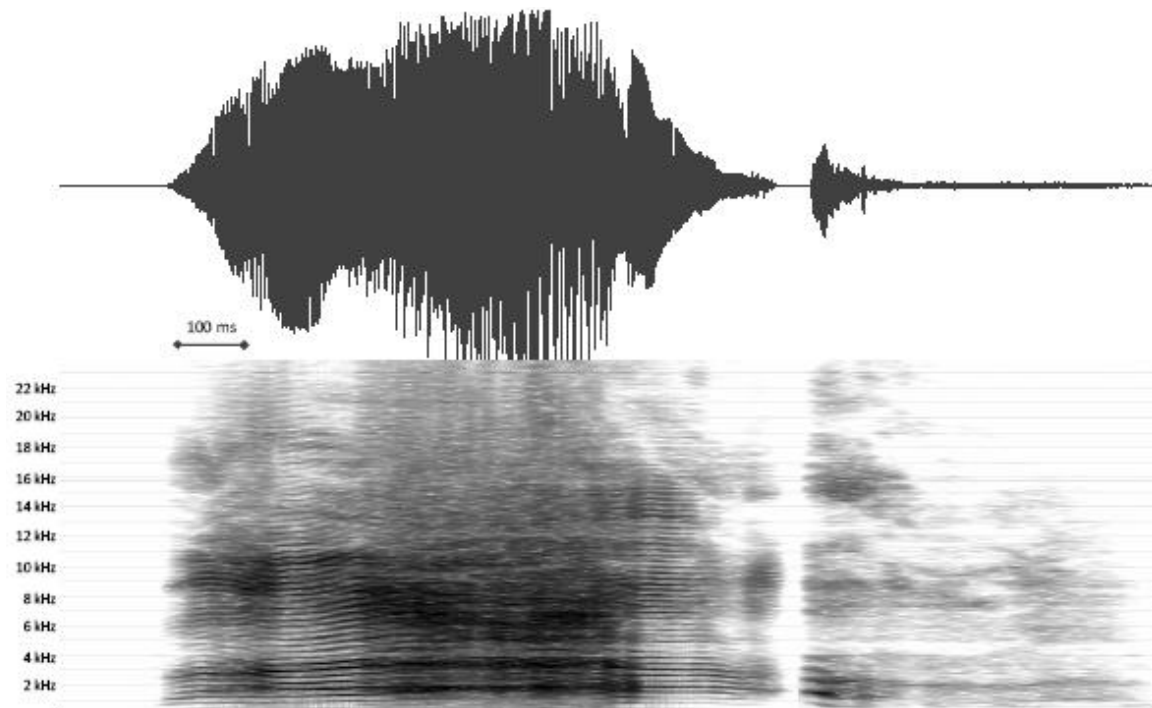
Example 15 is a *yell*, produced by a 10 month old infant.

What we can learn from careful listening: Similar to Example 14, there is a pressed quality to this loud phonation, but its pitch is only slightly higher than the infant's normal pitch and it is not perceived as a squeal. It is loud and has rough qualities that also justify a "growl" classification.

What we can learn from the acoustic analysis: Note that in this spectrogram, as in the display of Example 14, a full 16 kHz is displayed, so the harmonics are much more closely spaced than in many of the previous examples. A very salient aspect of this vocalization is its loudness but also its vocal quality—again very pressed, with relatively flat harmonic energies extending many multiples up through the spectrum. The utterance should also be categorized as a growl due to the rough qualities in the middle and highest amplitude portion. Closer inspection of this segment reveals subharmonics which so deeply modulate the basic rate of vibration that they develop a primary pulse-like form that can be seen and heard to be as low as 100 Hz in fundamental frequency (those details are not perspicuous in the image above). This sample also has some clipping, and in this case the amplitude of the recorded signal also exceeded the range of the digitizing hardware; the waveform peaks touch the edges of the displayed range. This may again somewhat exacerbate the rough perception of the loudest portion, but it does not fully account for the phonatory roughness yielding the *growl-yell* categorization.
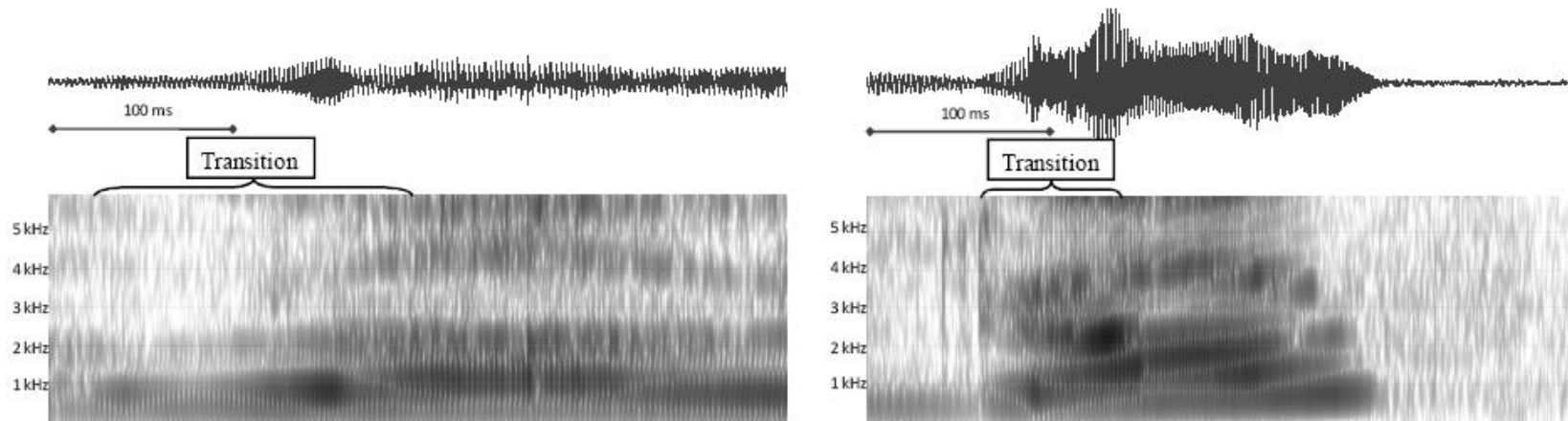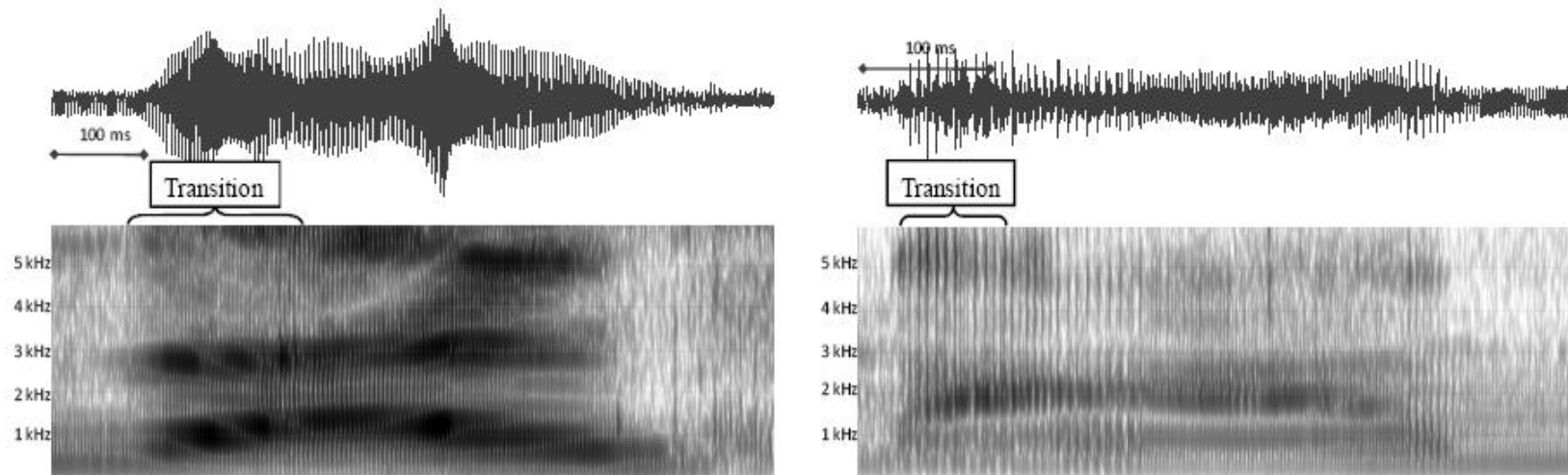
Example 16 on the left is a *marginal babble* produced by a 12 month old infant, and Example 17 on the right is a *canonical babble* produced by the same infant in the same recording session.

What we can learn from careful listening: Both examples have an audible syllable structure, with full vowels and articulated onsets. The transition from the onset into the vowel in Example 16 is slow, and, thus, this syllable is classified as "marginal babble." In contrast, Example 17 is a well-formed syllable with a rapid transition from the consonant to the vowel (it is also produced in squeal register).

What we can learn from the acoustic analysis: The spectrograms are 800 Hz bandwidth analyses to highlight the formant structure (although the rise in pitch at the end of Example 17 causes some visual resolution of harmonic patterning even at this large bandwidth). It is also important to know that the time intervals displayed here are exactly the same—just under 400 ms—to help compare some important temporal aspects of the formant and amplitude dynamics. Most importantly, visual and segment-specific auditory examinations suggest that the onset of the marginal babble develops over an interval of 150 ms or more before reaching a stable vowel nucleus, while the onset of the canonical babble is complete within 60-90 ms. These onsets are defined primarily in terms of formant movement: in theory, when the oral cavity is closed and then opens, F1 always rises (Stevens, 1998). The marginal babble in Example 16 does not seem to begin with a full closure, so this principle does not specify a necessary rise in F1. Further, the full vowel in the syllable may be relatively high, resulting in a low F1 with relatively little rise in F1 across the utterance. In accord with speech acoustic theory, F2 transitions will be more or less evident depending on place of consonant closure—none seems to occur here—and/or place (front/back tongue position) of vowel. These formant-frequency-determining factors are complex and interact in ways that do not always yield straightforward measurability of formant transitions, especially in infant utterances. Thus we often cannot distinguish easily between canonical and marginal syllables based on formant transition measurements alone. However, the rapid formant movements of Example 17 are evident enough, and the transition patterns of this example are visibly different than the much more gradual rise in overall amplitude and very slowly developing F2 amplitude discernible in the marginal babble of Example 16. These factors, especially with the slow amplitude rise in Example 16, support the auditory judgment that 16 is a marginal syllable while 17 is canonical. Example 17 can also be categorized as a squeal in the phonatory domain, given its very high frequency ending.

Example 18 on the left is another marginal babble produced by a 10 month old infant, and Example 19 on the right is a canonical babble produced by the same child in the same recording session.

What we can learn from careful listening: Again, both examples could be said to have audible syllable structure, with articulated onsets and an open nucleus, but only Example 19 is deemed a well-formed syllable, with an onset that is articulated rapidly enough into the nucleus to have canonical status. Example 19 is also audibly a growl.

What we can learn from the acoustic analysis: Unlike the previous marginal babble in Example 16, Example 18 has very clear formant visibility and a strong voice, but again the transition is slow enough to be "audibly tracked" (lasting at least 130 ms). In this particular example the overall difference between the amplitude at the beginning of the opening gesture and at the vowel nucleus is also small over the transition interval, and there is barely any F2 frequency change. The canonical babble in Example 19 may seem a bit ambiguous in at least two respects: (1) There is no turbulence event associated with an obstruent release, but note that this is not necessary for a well-formed adult canonical syllable. (2) The utterance is also a growl, which illustrates again the fact that growls and squeals can occur as the phonatory qualities of articulated syllables (also possible in adult connected speech!). Notable positive features characterizing Example 19 as a canonical syllable include a rapid transition from opening into a stable vowel nucleus (c. 65 ms), a rapid rise in amplitude over this interval, and an F2 which (though faint) falls rapidly in frequency.
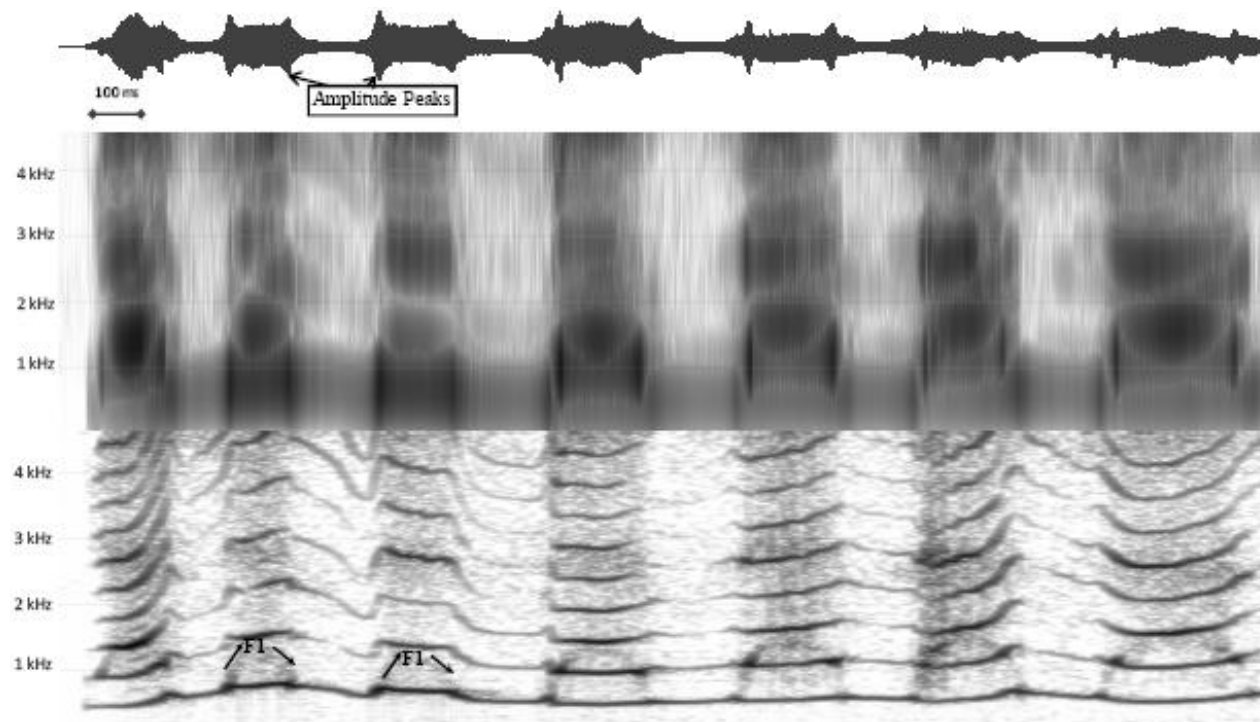
Example 20 is a single breath group sequence of reduplicated canonical syllables, a kind of celebration of babbling that adult listeners recognize as truly speech-like and to which they are usually therefore tempted to ascribe meaning. The infant producing this example was 10½ months old.

What we can learn from careful listening: In this example, we can hear seven clearly produced CV syllables with full vowels (or squeals) in each nucleus.

What we can learn from the acoustic analysis: Two spectrograms are presented, a very wideband (900 Hz) spectrogram is displayed on the top and a very narrowband (10 Hz) spectrogram on the bottom. The rate of production of these syllables is just under 3 per second (average duration 370 ms per syllable); adults generally produce about 4 or 5 syllables per second in conversation. Some acoustic curiosities are worth explaining here as babbling often includes such features: the odd peaks in amplitude at syllable margins are caused by the passing of F1 through the first and second harmonics, as can be seen quite clearly in the narrowband spectrogram. This concurrence of specific source energies with vocal tract resonance may occur in adults as well but is far more likely in infant vocalizations due to especially high f_0 in proportion to vocal tract resonances. More to the point of characterizing canonical syllabicity, the clear and relatively rapid traversal of F1 from low to high frequencies and back verifies that these sounds are produced by rapid closing-opening gestures. Finally, note that at least some of the higher pitch syllables in this sequence, like those for which F1 movement has been marked, would also be squeals in their phonatory feature coding.

CONNECTIONS

The preceding chapter describes anatomical substrates for sound production that are essentially in place at birth. In this chapter, we considered the great variety of pre-speech vocalizations typically produced by infants as they progress towards first words. We show how these vocalizations attain increasing similarity with speech in terms of timing, articulation, and complexity.

Chapter 5 describes how, around their first birthdays, young children produce first words, how first words not only resemble characteristics of babbling but also coexist with it for a few months, and how word productions become less variable and also become increasingly sophisticated in terms of segmental content and word shape. As infants mature, their pre-speech vocal development relies on sensory feedback. Chapter 19 describes pre-speech development of infants who have hearing impairments or other developmental disabilities.

CONCLUDING REMARKS

The focus of this chapter was on the speech-related sounds infants produce starting at birth. These sounds were described in terms of protophone categories and features, including quasivowels, full vowels, squeals, growls, whispers, yells, ingressives, raspberries, clicks, glottal stops, goos, marginal syllables, canonical syllables, reduplicated babble, and variegated babble.

The historical and theoretical underpinnings of the categories as well as guidelines for applying the protophone categorizations were discussed. As infants mature, the categories build on each other, leading to increasing phonatory and articulatory sophistication and increasing resemblance to speech. This direct correspondence to observed stages of prelinguistic vocal development is at the heart of the protophone-based approach.

CHAPTER REVIEW QUESTIONS

1. The discontinuity theory held that early vocalizations were unrelated to speech. We now know the theory to be false. What are some ways that babbling and speech are related?
2. How are protophones different from vegetative sounds and fixed signals such as cry and laughter?
3. What are the differences between marginal and reduplicated or variegated canonical babbling?

SUGGESTIONS FOR FURTHER READING

Oller, D. K. (2000). The emergence of the speech capacity. Mahwah, N. J.: Laurence Erlbaum.

SUGGESTIONS FOR STUDENT ACTIVITIES

1. Listen to sound file 4S22. Alternatively, make your own recording of an infant, preferably one between 2 and 12 months old, during a time when he/she is vocalizing. Identify the protophones from this chapter that are present in the recording, excluding any reflexive sounds. Also note the frequency with which each protophone occurs.

4S22 Recording of a 9-week-old infant (this example, and 4S21 below, were provided by Beate Peter).

2. Compare and contrast the cries in sound file 4S21 with those from 4S22. Do you sense a difference in the volitional nature of the vocalizations in the two sound files? Do you hear any evidence of supraglottal articulations in 4S21?

4S21 Cries of a one-week-old infant.

3. Interview a person who is a primary caregiver for an infant between the ages of two and twelve months of age. Describe the protophone categories to them and ask them which ones their infant currently produces, and how frequently.
4. Observe a caregiver/infant interaction. What protophones does the infant produce, and how does the caregiver respond to them?
5. Conduct a literature review on the differences in protophone development across different populations, such as infants with hearing impairment, those with Down syndrome, and those of varying socioeconomic status. Address both differences and similarities in protophone emergence across these populations.

ANSWERS TO CHAPTER REVIEW QUESTIONS

1. The discontinuity theory held that early vocalizations were unrelated to speech. We now know the theory to be false. What are some ways that babbling and speech are related?

While protophones are mostly not speech-like in the sense of having identifiable phonemic characteristics, they are infraphonological in the sense that they are contrastive. For example, squeals and growls are quite distinctive in their phonatory characteristics. Another way in which babbling is related to speech is that it is clearly volitional. Together, these characteristics strongly support the notion that infants are actively exploring the ways in which sounds can be structured. Additionally, children's vocalizations come to have more speech-like characteristics as they progress from birth to first words. For example, they come to have fully resonant nuclei and syllabic timing (as in canonical babbling).

2. How are protophones different from vegetative sounds and fixed signals such as cry and laughter?

Protophones are produced volitionally, whereas vegetative sounds and fixed signals are produced reflexively. Sometimes, however, the difference can be a little difficult to discern as the boundaries may be a bit fuzzy. For example, fuss or whimper may be perceived as somewhat volitional compared to full-blown cry, and chuckling can also be somewhat volitional in comparison to reflexive laughter.

3. What are the differences between marginal and and reduplicated or variegated canonical babbling?

The essential feature distinguishing canonical babbles is the closant-vocant transition. In a fully developed canonical syllable, the transition between the consonant-like and vowel-like portions of the speech are typically less than 120 ms. In a marginal syllable, this transition is typically longer than 120 ms. Another way to distinguish canonical and marginal syllables is by listening. If the transition between consonant and vowel can be directly tracked as a gradual movement, then the syllable is marginal. If the transition is fast enough that the transition between consonant and vowel sounds immediate, it is canonical. Canonical syllables are the prototypical syllables of adult speech. Canonical babble may consist of a single syllable or of sequences of syllables. In a sequence, if at least one syllable in the sequence is clearly canonical then the sequence is said to be canonical.

REFERENCES

Atkinson, K., MacWhinney, B., and Stoel, C. (1970). An experiment in the recognition of babbling. *Papers and Reports on Child Language Development*, No. 1. Stanford, CA.

Baken, R. J., and Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*. San Diego: Singular Publishing Group.

Buder, E. H. (1996). Experimental phonology with acoustic phonetic methods: Formant measures from child speech. In D. Ingram (Ed.), *Proceedings of the UBC international conference on phonological acquisition* (pp. 254-265).

Buder, E. H. (2005). The acoustics of nasality: Steps towards a bridge to source literature. *Perspectives on Speech Science and Orofacial Disorders*, 15(1), 4-8.

Buder, E. H., Chorna, L., Oller, D. K., and Robinson, R. (2008). Vibratory regime classification of infant phonation. *Journal of Voice,* 22, 553-564.

Buder, E. H., Warlaumont, A. S., Oller, D. K., and Chorna, L. B. (2010). Dynamic indicators of mother-infant prosodic and illocutionary coordination. In *Proceedings of Speech Prosody 2010.* Chicago, IL.

Cheney, D. L., and Seyfarth, R. M. (1999). Mechanisms underlying vocalizations of primates. In M. D. Hauser and M. Konishi (Eds.), *The design of animal communication* (pp. 629-644). Cambridge, MA: MIT. Press.

de Boysson-Bardies, B. (1999). *How language comes to children: From birth to two years.* Cambridge, MA: MIT Press.

Eady, S. J. (1980). *The onset of language-specific patterning in infant vocalization.* (Master's thesis), University of Ottawa, Ottawa, Canada.

Eilers, R. E., and Oller, D. K. (1994). Infant vocalizations and the early diagnosis of severe hearing impairment. *Journal of Pediatrics*, 124, 199-203.

Elbers, L. (1982). Operating principles in repetitive babbling: A cognitive continuity approach. *Cognition,* 12, 45-63.

Ertmer, D. J., and Nathani Iyer, S. (2010). Prelinguistic vocalizations in infants and toddlers with hearing loss: Identifying and stimulating auditory-guided speech development. In M. Marschark and P. E. Spencer (Eds.), *The Oxford handbook of deaf studies, language, and education, Vol. 2* (pp. 360–375). Oxford: Oxford University Press.

Goldstein, M. H., and Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science,* 19(5), 515-523.

Hauser, M. (1996). *The evolution of communication*. Cambridge, MA: MIT Press.

Hockett, C. F., and Altmann, S. A. (1968). A note on design features. In T. A. Sebeok (Ed.), *Animal communication: Techniques of study and results of research* (pp. 61-72). Bloomington, IN: Indiana University Press.

Holmgren, K., Lindblom, B., Aurelius, G., Jalling, B., and Zetterstrom, R. (1986). On the phonetics of infant vocalization. In R. Zetterstrom (Ed.), *Precursors of early speech* (pp. 51-63). New York: Stockton Press.

Irwin, O. C. (1947). Infant speech: consonant sounds according to manner of articulation. *Journal of Speech and Hearing Disorders*, 12, 402-404.

Irwin, O. C. (1948). Infant speech: development of vowel sounds. *Journal of Speech and Hearing Disorders,* 13, 31-34.

Jakobson, R. (1941). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almquist and Wiksell.

Jakobson, R. (1968). *Child language, aphasia and phonological universals* (A. R. Keiler, Trans.). Den Haag: Mouton.

Johnson, K. (2003). *Acoustic and auditory phonetics* (2nd ed.). Malden, MA: Blackwell Publishing.

Kent, R. D., and Read, W. C. (2002). *The acoustic analysis of speech* (2nd ed.). Albany, N. Y.: Delmar.

Koopmans-van Beinum, F. J., and van der Stelt, J. M. (1986). Early stages in the development of speech movements. In R. Zetterstrom (Ed.), *Precursors of early speech* (pp. 37-50). New York: Stockton Press.

Levitt, A. G., and Wang, Q. (1991). Evidence for language-specific rhythmic influences in the reduplicative babbling of French- and English-learning infants. *Language and Speech*, 34(3), 235-249.

Lorenz, K. (1951). Ausdrucksbewegungen höherer Tiere. *Naturwissenschaften*, 38, 113-116.

Lynch, M. P., Oller, D. K., Steffens, M. L., and Buder, E. H. (1995). Phrasing in prelinguistic vocalizations. *Developmental Psychobiology*, 28, 3-23.

Lynch, M. P., Oller, D. K., Steffens, M. L., Levine, S. L., Basinger, D. L., and Umbel, V. M. (1995). Development of speech-like vocalizations in infants with Down syndrome. *American Journal of Mental Retardation*, 100(1), 68-86.

Lynip, A. W. (1951). The use of magnetic devices in the collection and analysis of the preverbal utterances of an infant. *Genet Psychol Monogr,* 44(2), 221-262.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21(4), 499-546.

Masataka, N. (2001). Why early linguistic milestones are delayed in children with Williams syndrome: Late onset of hand banging as a possible rate-limiting constraint on the emergence of canonical babbling. *Developmental Science*, 4(2), 158-164.

Mitchell, P. R., and Kent, R. D. (1990). Phonetic variation in multisyllabic babbling. *Journal of Child Language,* 17, 247-265.

Nakazima, S. (1962). A comparative study of the speech development of Japanese and American English in childhood. *Studia Phonologica*, 2, 27-39.

Navarro, A., Pearson, B. Z., Cobo-Lewis, A. B., and Oller, D. K. (1998). Identifying the language spoken by 26-month-old monolingual- and bilingual-learning babies in a no-context situation. In H. Walsh (Ed.), *Proceedings of the 22nd Annual Proceedings of the 22nd Boston University Conference on Language Development* (Vol. 2, pp. 557-568). Somerville, MA: Cascadilla Press.

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. H. Yeni-Komshian, J. F. Kavanagh and C. A. Ferguson (Eds.), *Child phonology* (Vol. 1: Production, pp. 93-112). New York: Academic Press.

Oller, D. K. (1986). Metaphonology and infant vocalizations. In R. Zetterstrom (Ed.), *Precursors of early speech* (pp. 21-35). New York: Stockton Press.

Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, N. J.: Lawrence Erlbaum Associates.

Oller, D. K., Eilers, R. E., and Basinger, D. (2001). Intuitive identification of infant vocal sounds by parents. *Developmental Science*, 4(1), 49-60.

Oller, D. K., and Griebel, U. (2008). The origins of syllabification in human infancy and in human evolution. In B. Davis and K. Zajdó (Eds.), *The syllable in speech production* (pp. 29-62). New York: Lawrence Erlbaum Associates.

Oller, D. K., and Ramsdell, H. L. (2008, July). *The creation of phonological categories and the negotiation of word meanings in early lexical development*. Paper presented at the XI International Congress for the Study of Child Language, Edinburgh.

Oller, D. K., Wieman, L., Doyle, W., and Ross, C. (1975). Infant babbling and speech. *Journal of Child Language*, 3, 1-11.

Owren, M. J., and Goldstein, M. H. (2008). Scaffolds for babbling: Innateness and learning in the emergence of contextually flexible vocal production in human infants. In D. K. Oller and U. Griebel (Eds.), *Evolution of communicative flexibility: Complexity, creativity and*

adaptability in human and animal communication (pp. 169-192). Cambridge, MA: MIT Press.

Paul, R., Fuerst, Y., Ramsay, G., Chawarska, K., and Klin, A. (2011). Out of the mouths of babes: Vocal production in infant siblings of children with ASD. *Journal of Child Psychology and Psychiatry*, 52, 588-598.

Ramsdell, H. (2009). *The emergence of phonological categories in human infancy: A study of methodologies.* (Doctoral dissertation), The University of Memphis. ProQuest database.

Sheinkopf, S. J., Mundy, P., Oller, D. K., and Steffens, M. (2000). Vocal atypicalities of preverbal autistic children. *Journal of Autism and Developmental Disorders*, 30(4), 345-353.

Smith, B. L., Brown-Sweeney, S., and Stoel-Gammon, C. (1989). A quantitative analysis of reduplicated and variegated babbling. *First Language*, 9, 175-190.

Sroufe, L., and Wunsch, J. (1972). The development of laughter in the first year of life. *Child Development*, 43, 1326-1344.

Stark, R. E. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komshian, J. F. Kavanagh and C. A. Ferguson (Eds.), *Child phonology* (Vol. 1: Production, pp. 73-92). New York: Academic Press.

Stark, R. E. (1981). Infant vocalization: A comprehensive view. *Infant Medical Health Journal*, 2(2), 118-128.

Stevens, K. N. (1998). *Acoustic Phonetics*. Cambridge, MA: The MIT Press.

Stoel-Gammon, C., and Cooper, J. A. (1984). Patterns of early lexical and phonological development. *Journal of Child Language*, 11(2), 247-271.

Stoel-Gammon, C., and Dunn, C. (1985). *Normal and disordered phonology in children*. Austin, TX: Pro. Ed.

Vihman, M. M., and de Boysson-Bardies, B. (1994). The nature and origin of ambient language influence on infant vocal production and early words. *Phonetica*, 51, 159-169.

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., and Miller, J. (1985). From Babbling to Speech - a Re-Assessment of the Continuity Issue. *Language*, 61(2), 397-445.

Warlaumont, A. S., Oller, D. K., Buder, E. H., Dale, R., and Kozma, R. (2010). Data-driven automated acoustic analysis of human infant vocalizations using neural network tools. *The Journal of the Acoustical Society of America*, 127, 2563-2577.

Warlaumont, A. S., Oller, D. K., Dale, R., Richards, J. A., Gilkerson, J., and Xu, D. (2010). Vocal interaction dynamics of children with and without autism. In S. Ohlsson and R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, Austin, TX: Cognitive Science Society, 121-126.