



Published in final edited form as:

J Voice. 2008 September ; 22(5): 553–564. doi:10.1016/j.jvoice.2006.12.009.

Vibratory Regime Classification of Infant Phonation

Eugene H. Buder^{*}, Lesya B. Chorna^{*}, D. Kimbrough Oller^{*}, and Rebecca B. Robinson^{*}

^{*} School of Audiology and Speech-Language Pathology, The University of Memphis, Memphis, TN

Abstract

Infant phonation is highly variable in many respects, including the basic vibratory patterns by which the vocal tissues create acoustic signals. Previous studies have identified the regular occurrence of non-modal phonation types in normal infant phonation. The glottis is like many oscillating systems that, because of non-linear relationships among the elements, may vibrate in ways representing the deterministic patterns classified theoretically within the mathematical framework of non-linear dynamics. The infant's pre-verbal vocal explorations present such a variety of phonations that it may be possible to find effectively all the classes of vibration predicted by non-linear dynamic theory. The current report defines acoustic criteria for an important subset of such vibratory regimes, and demonstrates that analysts can be trained to reliably use these criteria for a classification that includes all instances of infant phonation in the recorded corpora. The method is thus internally comprehensive in the sense that all phonations are classified, but it is not exhaustive in the sense that all vocal qualities are thereby represented. Using the methods thus developed, this study also demonstrates that the distributions of these phonation types vary significantly across sessions of recording in the first year of life, suggesting developmental changes. The method of regime classification is thus capable of tracking changes that may be indicative of maturation of the mechanism, the learning of categories of phonatory control, and the possibly varying use of vocalizations across social contexts.

Introduction

Background

One required aspect of spoken language development is an ability and preference for modal voice, as heard in typical adult speech. Oller¹ has listed normal phonation as a first step towards mastery of canonical syllable production, commonly known as “babbling,” which occurs typically around the 7th month of life. Caregivers, researchers, and others primarily interested in tracking incipient language understandably attend to productions spoken with modal voice, as being indicative of emerging linguistic control, while treating squealy or growly voices as pertaining to more paralinguistic communication indicating emotion, attitude, or overall fitness. It may be that both modal and non-modal voice uses are important in infants' development of vocal control. Consequently, there should be considerable explanatory value in categorizing phonatory patterns found in infancy, and in their developmental course leading to fine control of the laryngeal source mechanisms, as a key foundation for speech.

Studies of infant phonation *per se* have focused almost exclusively on crying as an indicator of health status.^{2–14} Even in some of the earliest acoustic work in this area, spectrographic

Corresponding Author: Eugene H. Buder, School of Audiology & Speech-Language Pathology, The University of Memphis, 807 Jefferson Ave., Memphis, TN 38105, Tel: 901-678-5853, Fax: 901-525-1282, ehbuder@memphis.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

inspection of harmonic structure revealed categorically distinct non-modal phonation types such as pitch breaks into loft register and sudden appearances of subharmonics.¹⁵ However, it was anticipated by such early researchers that these phonation types would serve as differential indicators of neurological or structural pathologies. It has meanwhile become apparent that these phonation types are prevalent in the cries of quite normally developing infants.¹⁶

Employing spectrographic inspection, classic work by Stark in the 1970s identified subharmonic phonation as a regular feature of cry and discomfort vocalizations.¹⁷ Since then, other researchers have sporadically noted the regular presence and significant frequencies of many non-modal vibration patterns in the comfort vocalizations of normally developing infants. Keating¹⁸ was among the first researchers to reveal the variety of non-modal types in infant phonation (generally classifying “Fry” and “High” phonations but observing many features within these registers similar to those observed in the present work, such as pulse, subharmonics, and loft), and to interpret this variety in terms of laryngeal configurations. In the context of a comprehensive acoustic characterization of infant non-cry vocalizations, Kent and Murray¹⁹ noted a high incidence of alternative phonations in 3, 6, and 9 month-old infants (including “biphonation” and “fry,” and again observing many examples of harmonic doubling). Somewhat later, Robb and Saxman²⁰ inventoried occurrences of non-modal phonation within a large sample of non-cry vocalizations by young children aged 11–25 months (specifically, “biphonation,” “harmonic doubling,” and “fundamental frequency shifts”), and thereby established the normality of these phonations even in older children who were babbling and producing early words. A recent study by Rvachew and colleagues focusing on methodological issues in the acoustic classification of infants’ syllable inventories also acknowledged and included “abnormal” phonations (including a subset of the regimes explored here, such as “biphonation” and “harmonic doubling”), and found that they could be regularly and reliably identified.²¹

Concurrent with these developmentally-oriented research efforts, theoretical work applying non-linear dynamics to voice was also identifying vibratory regimes such as harmonic doubling and biphonation,^{22,23} and one of the earliest such reports focused on newborn infant cries.²⁴ As has been overviewed by several useful tutorial pieces,^{25,26} non-linear dynamics provides an organizing framework for vocal vibratory regimes. This is because the array of diverse vibration types can be mathematically understood to result from a single dynamic system, usually as a function of a small set of control parameters (such as sub-glottal pressure). As a result of variations in such a control parameter, the system can be observed to jump suddenly, or “bifurcate,” from one vibratory regime to another. Crossing of the phonatory threshold, by which the vocal folds held static by medial compression suddenly begin to vibrate when sub-glottal pressure is increased to a certain critical value, is perhaps the simplest example of such a bifurcation. Indeed, the very suddenness by which some vocal fold vibration types appear and disappear in phonation, as will be seen below in subharmonics for example, is a hallmark of non-linear dynamic systems undergoing bifurcations. This discreteness is also a methodological boon to the demands of a classification scheme.

The concept of chaos as a vibratory regime offers another conceptual and methodological advantage to the understanding of vocal fold vibration via non-linear dynamics; while appearing to be as noisy as purely stochastic turbulence, chaos can be treated as just another type of vibration that occurs within systems that are low-dimensional, as governed by the small set of parameters comprising typical models of phonation. While some authors may use the term “chaos” to refer to all the oscillatory possibilities of non-linear dynamic systems, inclusive of periodic behaviors, we will reserve the term in our coding scheme described below to refer to only those behaviors that appear to be dominated by an aperiodic vibration of the vocal folds.

See Jiang et al.²⁷ for a recent review of the usefulness of low-dimensional modeling and algorithmic measurement of chaos in the understanding of pathological phonation.

Rationale

The central rationale of the current project is as follows: Infants, via the wide variety of phonatory conditions experienced in the first year of development, may be manipulating a nonlinear dynamic system through its range of possible vibratory regimes. The mathematical theory of non-linear dynamics specifies that these vibratory regimes can be unified and classified under a single framework. It should therefore be possible to inspect, discretely bound, and classify all infant phonations under this framework, selecting among a reasonably small set of possible regimes.

It may also be possible to apply signal processing algorithms to the measurement of at least some such regimes.^{27,28} However, these algorithms assume certain signal conditions that are not likely to be satisfied by freely produced infant vocalizations that often also include non-phonatory sources, articulations, modulations, etc. For such conditions, the most valid and reliable classifications may be achieved through auditory and visual inspection by human analysts. A specific proof of concept is thereby motivated: Can analysts be trained to exhaustively analyze a continuous record of the phonation occurring within infants' spontaneous vocalizations into a small set of vibratory regimes that can be identified with the possibilities expected under the theory of non-linear dynamics? The purpose of this report is to demonstrate just this possibility, and to furthermore demonstrate that the resulting classifications can help to document infants' developing phonatory control.

The objective of applying regime analysis to infant phonation, however, is motivated not directly by the theory of non-linear dynamics, but by the goal of understanding phonatory and speech development in infants. The analysis of regimes described in this report is not oriented therefore to the detection of every theoretically discernable vibration type, but rather to the classification of vocal behaviors appearing to have interpretive significance for a developmental analysis. Only those bifurcations that could be efficiently and meaningfully tracked were therefore targeted in the training and analysis protocols reported below. In particular, sudden pitch breaks or shifts clearly fit the paradigm of non-linear dynamics as bifurcations, but the regimes that are delineated by such shifts may all be modal in perceived quality and apparent dynamics. Although these regimes might therefore be considered theoretically distinct because a dynamic break occurs between them, the regimes themselves do not necessarily carry the same interpretive significance as a break from, say, modal to loft.

By the same token, it should be acknowledged that a non-linear dynamic classification scheme by no means encompasses all vocal qualities of interest. Many variations in quality may occur within a class, most notably within the infant modal voice in which varying degrees of breathiness, harshness, pressed, and other qualities occur. These were not targeted as distinct classes from the current perspective, but may be revisited using other tools such as perturbation analysis, estimates of glottal turbulence noise,²⁹ relative harmonic amplitudes,³⁰ and other tools oriented to general voice quality analysis.³¹

The research on these phonation types may also contribute to explication of the widely reported tendency of parents to recognize "categories" of vocalization produced by infants during the first half year of life, categories that are described impressionistically in a way that suggests that phonation type is the primary factor determining the categorization.³²⁻³⁴ In particular many observers have indicated at least three widely recognizable categories occurring in most infants: A mid pitch category (vowel-like sounds, full vowels, or quasivowels), a high pitch category (squeals or squeaks), and a category that can be either low in pitch or mid pitch with very harsh vocal quality (growls).³⁵⁻³⁷ These apparent categories are recognized by their

repetitive and systematically alternating occurrence within sessions of recording.³⁸ Observers appear to assign the vocalizations to squeal, vowel, or growl categories based on some “predominant” or “most salient” characteristic of the utterance.

The importance of these apparent categories of vocalization in the first months of life has been argued to be very fundamental, because they appear to represent the first contrastive vocal categories that are created by the infant. This ability to form new vocal categories, unknown in other primates, has been argued to form a necessary basis for the creation and learning of further contrastive vocal categories required for speech.³⁹ The systematic study of vocal regimes and their perception is, we reason, the appropriate method to begin to unravel the nature of these categories, including how they are physically composed and how they are perceived.

Overview

The list of distinct regimes that were targeted in this application is presented in Table 1, and the distinguishing characteristics of each regime are presented in the following Method section. After developing basic methods and operational characteristics for regime classification, the report describes aspects of the protocols by which analysts were trained to perform classifications. Two types of results are then presented: (1) aspects of the reliability with which analysts performed classification on utterance sets selected to represent all types are examined; and (2) an application of the classification to all the phonatory events in recordings from a female infant at three different developmental stages in the first year of life is presented to examine the apparent developmental significance of the regime classifications.

Method

Materials

Materials for training and reliability of regime coding were selected from three female infants aged between four and 11 months. These were normally developing infants recruited for a study of spontaneous vocalizations across varying social contexts, and no laryngeal or other clinical examinations were included in this protocol. Materials for the developmental analysis of regimes comprised all the non-distress vocalization from three 20-minute sessions recorded from one of these female infants at three ages: four, six and one half, and eleven months. (In fact, the six and one half month age was sampled in sessions that were separated by 10 days, so the age by week specifications in figure captions below vary, depending on whether an age or a session is represented). For all materials, infants were video- and audio-recorded during free play with their caregivers in a sound-treated room furnished with soft mats and toys. The infants were fitted with custom-built vests that housed a wireless microphone system (Samson Airline UHF AL1 transmitter, equipped with a Countryman Associates low-profile low-friction flat frequency response MEMWF0WNC capsule, sending to a Samson UHF AM1 receiver). The vest configuration followed an original design developed by Buder and Stoel-Gammon,⁴⁰ with the microphone capsule housed within a velcro patch and oriented to maintain the mouth-to-microphone distance at approximately 5 cm. TF32 software⁴¹ operating a DT312 acquisition card (Data Translation, Inc.) was used to digitize the infant signals at 48 kHz after low-pass filtering at 20 kHz via an AAF-3 anti-aliasing board.

Regime Analysis and Definitions

TF32 software was used for all regime classifications. Settings for spectrographic display were: bandwidth = 10 Hz, displayed frequency range = 0–6 kHz, dB floor = 95 dB, and dynamic range = 64 dB. Analysts were advised always to consider 1) harmonic structure in the lowest frequency ranges of the narrowband spectrograms, 2) glottal cycle patterns in waveforms, and 3) the auditory quality of selected segments. Upon full consideration of these resources,

analysts placed cursors around successive regimes and used the labeling function of the software to assign a regime label to the selected segment.

Only laryngeal tissue vibrations, and not supra-laryngeal (e.g., uvula, tongue) tissue vibrations, were considered for regime coding, but analysts were not asked to discriminate among vibrations of laryngeal structures other than the true vocal folds (e.g. false vocal folds, epiglottis, and possibly soft tissues of the trachea). Purely “reflexive” types of sounds (coughs, hiccups, burps, etc.) were not coded. Complaining or whining sounds were included and coded, but full cries (during which it seemed the infant had ‘lost control’ and could therefore be considered in a reflexive manner of vocalizing) were excluded from regime coding. Finally, vocalizations or distinct regime segments that were less than 50 ms were considered to be too short to constitute instances of clearly intentional vocalization and were disregarded for coding purposes.

A regime was defined for the analysts as a consistent vibratory pattern as heard and as seen in harmonic and glottal pulse structures, disregarding any pitch breaks that did not change register (i.e., to loft or pulse), modulations such as tremors, flutters, and changes in vocal quality such as breathiness or harshness. All qualifying phonations were classified into one and only one of six phonatory regime codes—“Modal,” “Loft,” “Pulse,” “Subharmonics,” “Biphonation,” “Chaos,”—or to one of two stop codes used to mark phonation breaks during utterances that were perceived to be otherwise intended as glottally active: “Closed Stop,” (i.e., non-phonation due to excessive adduction) and “Open Stop” (i.e., non-phonation due to insufficient adduction). The following subsections discuss distinguishing characteristics of the eight regime codes.

Modal

The distinctive spectrographic feature of modal regime is the presence of harmonics at regular multiples of f_0 . Figure 1 shows a clear example of a regime coded as modal. For the purposes of classification in highly variable infant vocalizations, the category is large and inclusive of many voice qualities including very pressed or breathy voices. In very weak or breathy vocalizations the harmonics may be unclear, with visibility of only the lowest few. In harsh or loud vocalizations, inter-harmonic intervals may be noise filled. Sudden pitch changes could also cause changes in harmonic structure, and caution would be applied (via inspection of the waveform and/or segmental listening) to compensate for the temporal smearing effects of narrowband spectrograms that could, by spurious overlap, sometimes cause a false appearance of complex harmonic patterns. In all such cases, as long as the harmonic structure was indicative of a fundamental tone with simple odd and even harmonics, and the combination of listening and inspection did not indicate a loft register, the Modal code was used to indicate regular vocal fold vibration, with relatively full involvement of both lamina propria and muscular layers.

Loft

The concept of a “register” of infant phonation corresponding to the distinctive adult register of loft (or, in the context of singing, “falsetto” or “head voice”) is certainly difficult, if not inevitably vague. Nonetheless, in order to test the loft concept by assessing its operational merit as a vibratory regime, analysts were asked to examine instances of “widely spaced harmonics, corresponding to a higher pitched phonation.” For the labeling of this regime, the analysts were furthermore asked to compare such segments and consider whether the infants were phonating with a distinctive high-pitched non-cry “squeal” that can be heard to have the asthenic “reedy” quality of loft registers.

The contrast with Modal was especially encouraged when the higher regime could be seen in contrast to lower-pitched regimes, not just as a jump in frequency but also as a change in quality.

Before or after lower frequency regimes (e.g. Modal, Pulse, Subharmonics), a prospective Loft regime was considered most valid when a visible upward “break” in the harmonics could also be heard as a distinctive break in voice quality in the direction of the thinner quality of adult loft. Figure 2 shows an example of this situation. These situations were sometimes ambiguous—most notably when a very large upward pitch break yielded a squeal that was nonetheless quite strong sounding and exhibited the shallow spectral slope of stronger upper harmonics typical of that infant’s modal voice. To allow for such difficulties, we used the more loosely conceived label “High Modal” to allow for the possibility that some regimes so labeled would merit further refinement using additional acoustic analysis tools or refined psychometrics. We nonetheless continue to investigate and report on the operational classification success of this labeling approach at present as Loft, while planning to develop and adopt the progressively more objective tools in future investigations to distinguish loft from high-pitched modal.

Pulse

The Pulse regime was defined by the appearance of very closely spaced harmonics often resulting in temporal resolution of individual glottal pulses in the waveform and sometimes also the spectrogram, and a clear perception of a low “zipper-like” quality. A guideline of 200 Hz for f_0 was suggested to analysts as a rule-of-thumb for this regime, but it was noted that lower fundamental frequencies could still be heard as modal in older infants (and certainly for more mature phonators), and that for very young infants (e.g., four months or younger) fundamental frequencies higher than 200 Hz might still be classified as pulse according to the waveform appearance and perceptual quality. In this context, it is useful again to distinguish between the concept of a regime and a “pulse register”—the latter may imply classification according to a pitch range while here the emphasis is on the analysis of harmonic and glottal waveform structures. Figure 3 demonstrates an example of Pulse surrounded by modal. In some cases, especially at the ends of phonations, analysts were encouraged to overlook episodes of glottal pulsing for 50 ms or more if it was felt that the episode was simply secondary to laryngeal abductory or adductory gestures of offset or onset (respectively), or otherwise heard as inadvertent glottalization and not as prolonged “intentional” segments.

Subharmonics

The Subharmonics regime was defined primarily by the abrupt appearance in the narrowband spectrogram of intervening harmonics, doubling, tripling, or even higher integer multiples in relation to the surrounding set. However, when such patterns sometimes appeared so weakly as to not be auditorally discernable from modal voice, or sometimes so strongly as to look like a simple octave drop to pulse, additional criteria were adopted: 1) the distinctive “two-tone” roughness of subharmonics had to be audible, and 2) it was preferable also to observe period-doubling in the waveform. This waveform aspect was not defined as critical: Period-doubling in waveforms is readily visible in the typical cases in which the paired cycles differ in amplitude or waveshape, but not in atypical cases in which the paired cycles differ only in cycle length yet still present the hallmark rough quality and weak intervening harmonic amplitudes relative to the Modal pattern. See Figure 4 for an example of Subharmonics. Intriguingly, subharmonics could sometimes spectrographically resemble Chaos as well, as alternations between modal periods, period-doubling, period-tripling, etc., would yield an effectively aperiodic harmonic texture. In such cases, a predominance of regular glottal pulses and an overall tonal quality would distinguish the phonation from chaos.

Biphonation

Biphonation was defined by the appearance of extra harmonics at non-integer relationships, moving in non-parallel directions in relation to the accompanying fundamental set. An example of biphonation is shown in Figure 5. Generally, biphonation seemed to occur very rarely. In

some cases, a low frequency extra harmonic could appear in apparent association with the periodic vibrations of supra-laryngeal structures such as trilling tongue or vibrant pressed-lip sound, but when heard as such these were not coded as biphonation. While instances of non-parallel harmonics with a distinctive double-tone percept not ascribed to extra-laryngeal sources were included in this category, it was allowable that false vocal folds, epiglottis, or other laryngeal structures might be heard as potential sources.

Chaos

Chaos was defined by the appearance of a non-harmonic spectral structure created by glottal pulses with highly variable effectively random periods. An example of Chaos is shown on Figure 6. Chaos was a rare occurrence in which glottal pulse aperiodicity, deteriorated spectral structure, and an absence of tonality were the hallmarks, and many samples that could be considered “chaotic” but which still exhibited harmonic structure were coded as modal or another related regime (such as pulse). It could also be difficult to distinguish between chaotic tissue vibrations and turbulent aerodynamics as in a loud breathy glottal source. In such situations, analysts were instructed to take into consideration that signals from true tissue vibration should be seen mostly at the lowest frequencies inspect for a predominance of noise at higher frequencies (e.g. 2 kHz or more), excluding such samples from coding. As an episode containing chaos might typically also be interspersed with brief episodes of periodicity, such episodes were disregarded when they typically lasted no longer than 40–50 ms.

Closed Stop

The Closed Stop (abbreviated C-Stop) was defined by the sudden cessation of voicing followed by sudden resumption with no evidence of airflow during the gap. The perceptual and spectrographic cues allowing analysts to ascribe this cessation to excessive adduction could be difficult to classify, but the manners in which phonation ceased and resumed were usually essential cues, especially an abrupt onset. Surrounding phonation would typically be energetic and in pulse or modal register, and some perceptual cues to glottal articulation at the margins (as in the glottal stopping of “uh-oh”) were requisite. The presence of low intensity turbulence in or sometimes even a bit of a “squeaking” during the stop gap did not disqualify classification as a Closed Stop, as the medial compression of the infant’s glottal adduction might not always have completely overcome a high sub-glottal pressure. The coding definition allowed that one or two glottal pulses could occur during the stop. An archetypal example is shown on Figure 7.

Open Stop

The Open Stop (abbreviated O-Stop) was defined by the cessation of voicing followed by resumption, usually with evidence of airflow in the gap. Even more than with C-Stops, the classification had to be somewhat intuitive, relying on the percept that there was a voice break occurring during a time that the infant’s laryngeal and/or neurological settings were otherwise for phonation. This most typically occurred in a quiet “loft” phonation at the top of the pitch range, or at the sound pressure threshold. Acoustic observations indicated that the infant’s glottal conditions (primarily either insufficient adduction or subglottal pressure) were temporarily below the phonation threshold, including the observation of some air leak during the cessation and offsets and onsets that were more gradual than in Closed Stops. An example of an Open Stop is shown in Figure 8.

Summary of Regime Classification Scheme

The classification scheme described here is avowedly subjective. As summarized in the decision-making tree of Figure 9, however, the scheme is quite systematic, combining spectrographic inspection, auditory impressions, and waveform inspection with varying

priorities depending on the regime under consideration. The flowchart layout implies that some automatization could conceivably be developed. Nonetheless, the current approach maximizes the value of human analysis by exploiting intuitive impressions (as in the multi-faceted judgment that a brief episode of non-phonation was in fact a more or less inadvertent glottal stop that would otherwise have been intended as continuous phonation) and auditory judgments (as in the impressionistic but apparently categorical classifications of pulse or subharmonics), while eschewing arbitrary acoustic rules (such as simple but fallible fundamental frequency cutoffs for delimiting modal from either pulse or loft regimes).

Analyst Training

The work of four analysts is represented in results below. All analysts had at least one year's graduate-level training in speech-language pathology and had passed Masters-level coursework in speech science and in anatomy and physiology of the speech mechanism. Specialized training for infant vocal coding proceeded in several rounds all under the direct supervision of the first author with the assistance of co-authors and one doctoral student with a Masters in speech-language pathology. The training team also developed practice and assessment materials and formed regime label keys by consensus. First, several tutorial meetings were held to orient the team to the classification scheme by visual and auditory inspection of examples for each type and to introduce the mechanics of the TF32 labeling facility. This training also attended to problems with discerning infant phonation from other sound sources (mother overlay, environmental sounds, infant's non-laryngeal sound sources, etc.). Second, the analysts completed the labeling of a round of 20 samples with one or more utterances, each containing one or more phonations, selected to represent all the vibratory types of interests. During this round trainees were free to consult with trainers regarding interpretive and technical issues. Third, the results of the first round were reviewed and discussed in group sessions until all analysts' discrepancies had been addressed. Fourth, the analysts completed a practice round of another 12 wave files again containing all vibratory types of interest but with an emphasis on difficult decision scenarios. During this round, analysts were allowed to ask questions but discouraged from discussing the particular examples under inspection. Fifth, the results from the practice round were assessed according to the scheme described in Results to follow, and analysts received customized feedback designed to correct all individual tendencies for missing or confusing regimes in comparison to the key. Finally, analysts completed a set of 24 files for the reliability assessments presented below: these files had been keyed with 99 regime labels in the following distribution: 44 Modal, 25 Loft, 10 Pulse, 9 Subharmonics, 4 Biphonation, 4 O-Stop, 2 C-Stop, and 1 Chaos.

Results

Reliability of Regime Location and Identification

Classification of infant phonation episodes for vibratory regimes actually involves three conceptually distinct decisions: 1) determining whether or not the sound should be considered to involve glottal behavior, or "intended phonation," eligible for vibratory regime classification, 2) deciding on the classification, and 3) marking the onsets and offset times of a distinct class. Each of these analysis decisions is evaluated separately in the following subsections. Preparatory to performing the analyses described below, trainers collated the results of the analysts' labels against the key labels. In doing so, some patterns of errors performed differently across the analysts altered the total number of comparisons with the key: mis-determining phonations would result in either too many or too few regime labels, missing a regime change would result in too few labels, and splitting a single regime into two or more regimes would result in too many labels. To compare all decisions across all individuals against the key, the number of reliability decisions evaluated in the following sections therefore exceeds the core set of 99 regimes originally presented.

Determining Phonation

Within this decision set, there were in fact two classes of errors: either the analyst placed a regime label where the key had none (a regime class was given where the key was “null”), or the analyst missed events where the key codes indicated eligible phonation (a “null” answer was given for a given regime class). Table 2 summarizes the number of times among the four analysts that vibratory regimes were ‘imagined,’ where no phonation should have been coded. By far the largest number of such errors occurred when analysts misjudged a simple silence to be a glottal stop, either adductory (“C-Stop”) or abductory. Errors attributing Chaos to non-phonation also occurred, in which cases turbulent air flow was most likely misconstrued as chaotic vocal fold motion. Table 3 summarizes the number of times among the four analysts that vibratory regimes were nullified. Here the errors were more broadly distributed among types; while four were missed stops, other times analysts neglected to code audible phonations (these phonation segments may have been very low in sound pressure level, very brief, or both).

Classification Accuracy

Table 4 displays all the mis-classifications of regimes made after accounting for all errors in determining phonation (listed in Tables 2 and 3). A Cohen’s Kappa of 0.76 was obtained for this table overall, indicating good agreement rates even accounting for the marginal distributions (e.g., favoring modal).⁴² The measure is statistically quite significant (divided by a standard error of 0.0274, it is equivalent to a t of 27.68 with $df = 49$ and $p < .001$), confirming that this overall agreement substantially exceeded chance levels. Individually, the four analysts produced Cohen’s Kappas ranging from 0.74 to 0.78, which were also quite satisfactory. However, as discussed by Bakeman and Gottman,⁴² good Kappa reliabilities with high significance levels may or may not indicate adequate coding depending on the validation demands of a given protocol or application.

It can be seen by closer inspection of Table 4 that some regimes were more difficult to classify accurately than others. Modal, Loft, and Pulse regimes were all recognized with accuracies approaching or exceeding 90%. It might appear that C-Stops were recognized most accurately, at 100% of 7 cases, but in 1 case a C-Stop was overlooked (Table 3), and in 11 more cases C-Stops were seen where none occurred (Table 2), so actual accuracy on this regime was arguably below 50%. O-Stops fared somewhat better, but Chaos and Subharmonics were only recognized with 50% accuracy and Biphonation was particularly difficult to recognize with poor agreement and accuracy of only 25%. Even though the Kappa statistics therefore indicated overall good performance, it was clear that additional training and vigilance could help to obtain acceptable data.

In ongoing work, we have adopted procedures that select specific materials for retraining analysts to learn better than 50% accurate discrimination of regimes that had proved difficult in their reliability assessments. We also continue to systematically double-check results to establish consensus validation of all such regimes, including Subharmonics. The data presented below were obtained only from those analysts who produced better than 50% accuracy on difficult regimes such as Subharmonics, and all difficult or exotic regime codes such as Biphonation and Chaos were also inspected and validated by the trainers (first, second and fourth authors).

Timing Accuracy

Figure 10 presents histograms of the onset and offset timing differences between all four analysts and the regimes as keyed by trainers. The mean difference in beginning times was +4.5 ms and the mean difference in ending times was -0.6 ms. Among the nearly 400 comparisons, only 9 exceeded 100 ms, 99% were under 100 ms, and more than 80% were under 30 ms in difference.

Age-related Differences in Vibratory Regime Use by One Child

Using the methods (and skills criteria) outlined above, the most highly skilled analysts assessed all non-distress utterance phonations from six 20 minute sessions of a girl interacting freely with her mother: two on the same day at 4 months of age, two sessions approximately 2 weeks apart at 6 to 7 months of age, and two on the same day at 11 months of age. Figure 11 depicts the percents with which various regimes were counted. As Stops, Chaos, and Biphonation occurred overall very infrequently in comparison to the more common regimes, they are lumped as “other” in this figure. With that grouping (which also helped populate a chi-squared distribution), there is a significant association of regime with age ($\chi^2 = 48.75$ [$n=1486$], $p < .001$). Graphically this appears to be driven by the increase in observations of subharmonics at the 6 ½ month sample, and the decrease with age of occurrence of “other.” Variability is seen again in the micro-longitudinal sampling represented in Figure 12, which breaks apart the 6 ½ month sessions into the two 20 minute samples on separate days. The two days in the 6 to 7 month age range exhibited significantly different regime usages, especially in Modal versus Subharmonics. The two sessions comprising the 4 and 11 month samples, on the other hand, did not differ in the percentages of regime occurrence.

Another way of looking at regime occurrence is by durations. Figure 13 depicts variations in the raw amount of time the infant occupied a single-regime phonation interval across the age samples. Significance levels are those associated with four simple ANOVAs, one for each of the four major regime types examined, with three levels of age as the independent factor. For statistical analyses, the duration data had first been log-transformed to correct for positive skewing. While neither within-subjects design elements nor Bonferroni corrections were applied, the results seem clear, especially for age related trends in Subharmonics and Pulse. However, as seen in Figure 14 (with text notes summarizing the results of an ANOVA of log-transformed units), continuous phonations also varied across ages, peaking at 6 ½ months. This could be related to the increase of canonically babbling or articulated syllable forms (and some words) at 11 months that was observed in separate analyses for this infant. The trend lines and analyses of Figure 15 (with ANOVA performed as in Figure 14) take into account the age variations in total phonation time by examining the percent of total utterance phonation time accounted for by a given regime type. These results further strengthen certain trends like the increase in controlled use of Pulse with age, and the special exploration of Subharmonics at 6 ½ months.

Discussion

In summary, this manuscript has demonstrated that infant phonations can be exhaustively classified by vibratory regime types. The specification of those types in accord with the framework of non-linear dynamics provides additional external validation that the typology may ultimately be associated with physical mechanisms of production. It was also demonstrated, in at least one normally developing infant, that the regime classifications may be developmentally significant, with the somewhat surprising result that non-modal regimes can grow in usage during the first year of life. The latter result generally extends previous findings.^{19,20} There need be no expectation that the specific pattern of results, such as the child’s tendency to explore subharmonic vocalizations, will replicate in other children. In fact, the observation of significant session-to-session variation found in this child’s sixth month of life (Figure 12) suggests that a great deal of variability is to be expected within child as well as across children. Auditory impressions furthermore suggest that this child preferred growling vocalizations, whereas at least one of the other children who provided the training materials for this study displayed a tendency for exploring squeals at around this age. The latter child would therefore be expected to utilize a larger proportion of loft regimes, and further analysis

of age-related and child-to-child variation is underway in our lab utilizing the methods reported here.

With the availability of a proven method for regime classification and the accumulation of more such data from children phonating under a variety of conditions, evidence may be acquired for better understanding of the “control parameter” question: If vocal regimes, like those of other non-linear dynamic systems, can be organized along a continuum (along the lines of a bifurcation diagram), then the variable or variables underlying this continuum may be revealed by systematic observation of infant phonations. Ultimately, as has been suggested by other speech development researchers,⁴³ the variables associated with an infant’s growing control (regulation, modulation, etc.) of sub-glottal pressure may be critical for the understanding of phonatory development. Ongoing efforts by our group have sought to add other acoustic dimensions to the basic task of describing speech development in these early stages. While serving larger theoretical goals, these acoustic dimensions are also being explored in relation to the occurrence of distinct regimes. One study found some tendency for an association of regimes with vocal intensity,⁴⁴ while subsequent investigations have revealed an even stronger tendency for the regimes to organize in association with f_0 ,⁴⁵ as was also suggested in findings by Robb et al.²⁰

To support such systematic research goals, the current report documents a technique for the comprehensive classification of 100% of a given child’s recorded phonation to produce a distributional result. The combination of acceptable reliabilities with a theoretical framework, both reinforcing the assumption that all phonation should be classifiable (at least in infants without gross structural anomalies), supports new agendas for developmental research in voice. It should now be possible to systematically delineate the development of vocal quality and the emergence of control over modal voice in normally developing infants. In addition to extended longitudinal research with a larger set of infants at many more ages, we are exploring how vibratory regimes may contribute to the perception of global categories of sound that infants appear to create in the first 6 months of life as classified by previous research.^{35,46}

It must be acknowledged that the acoustic analyses presented here allow only very limited inference regarding the structures that produced such signals. The infant’s developing ability to reproduce and control acoustic categories of output is nonetheless interesting on its own merits as a basis for speech and language abilities, even while the anatomical substrate for such categories may remain under investigation. Fundamental questions in speech acquisition concern the development of repeated vocalization types that become perceptibly categorical in nature, and the classification scheme developed here provides adequate documentation of vocal control for addressing such questions. Nonetheless, numerous opportunities pertaining to disorders of communication are also now enabled, for example, research on early vocal quality in autistic individuals, previously indicated to be anomalous on the basis of perceptual categorizations.⁴⁷ Similar efforts are easy to imagine in the area of deafness to assess long-standing reports of anomalous vocal qualities produced by deaf children.⁴⁸ With the accrual of a normative database, and in combination with other clinical assessments of the vocal mechanism, the classification techniques presented here should also be useful for an understanding of early voice disorders in the very young.

Acknowledgements

This work has been supported by grants from the National Institutes of Deafness and other Communication Disorders (R01DC006099 to D. K. Oller PI and Eugene H. Buder Co-PI). The authors are also grateful to the caregivers of our participants for their time and generosity, to the graduate student research assistants who contributed to this study as analysts, and to Jamie L. Edrington for her assistance with training and reliability assessments.

References

1. Oller, DK. The emergence of the speech capacity. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.
2. Barr RG, Chen S, Hopkins B, Westra T. Crying patterns in preterm infants. *Developmental Medicine & Child Neurology* 1996;38(4):345–355. [PubMed: 8641539]
3. Furlow FB. Human neonatal cry quality as an honest signal of fitness. *Evolution and Human Behavior* 1997;18:175–193.
4. Gilbert HR, Robb MP. Vocal fundamental frequency characteristics of infant hunger cries: Birth to 12 months. *International Journal of Pediatric Otorhinolaryngology* 1996;34:237–243. [PubMed: 8839074]
5. Goberman AM, Robb MP. Acoustic examination of preterm and full-term infant cries: The long-time average spectrum. *Journal of Speech, Language, and Hearing Research* 1999;42(4):850–861.
6. Grauel EL, Höck S, Rothgänger H. Jitter-index of the fundamental frequency of infant cry as a possible diagnostic tool to predict future developmental problems: Part 2: Clinical considerations. *Early Child Development and Care* 1990;65:23–29.
7. Irwin JR. The cry as a multiply specified signal of distress. *Dissertation Abstracts International: Section B: The Sciences and Engineering* 1999;59(9B):5138.
8. Lester, BM.; Zeskind, PS. A biobehavioral perspective on crying in early infancy. In: Yogman, MW., editor. *Theory and research in behavioral pediatrics*. 1. New York: Plenum; 1982. p. 133-180.
9. Möller S, Schönweiler R. Analysis of infant cries for the early detection of hearing impairment. *Speech Communication* 1999;28(3):175–193.
10. Thodén, CJ.; Järvenpää, AL.; Michelsson, K. Sound spectrographic cry analysis of pain cry in prematures. In: Boukydis, CFZ., editor. *Infant crying: Theoretical and research perspectives*. New York: Plenum; 1985. p. 105-117.
11. Wasz-Hockert O, Lind J, Vuorenkoski V, Valanne E. The identification of some specific meanings in the newborn and infant vocalization. *Experientia* 1964;20:154. [PubMed: 4159066]
12. Wolff, PH. The natural history of crying and other vocalizations in early infancy. In: Foss, BM., editor. *Determinants of infant behavior*. 4. London: Methuen; 1969. p. 81-109.
13. Zeskind PS, Barr RG. Acoustic characteristics of naturally occurring cries of infants with “colic”. *Child Development* 1997;68(3):394–403. [PubMed: 9249956]
14. Robb MP, Goberman AM, Cacace AT. An acoustic template of newborn infant crying. *Folia Phoniatica et Logopaedica* 1997;49:35–41. [PubMed: 9097493]
15. Sirvio P, Michelsson K. Sound spectrographic cry analysis of normal and abnormal newborn infants. *Folia Phoniatica et Logopaedica* 1976;28:161–173.
16. Robb MP. Bifurcations and chaos in the cries of full-term and preterm infants. *Folia Phoniatica et Logopaedica* 2003;55:233–240. [PubMed: 12931057]
17. Stark RE, Rose SN, McLagen M. Features of infant sounds: The first eight weeks of life. *Journal of Child Language* 1975;2:205–222.
18. Keating, P. Patterns of fundamental frequency and vocal registers. In: Murry, T.; Murry, J., editors. *Infant communication: Cry and early speech*. Houston: College-Hill; 1980.
19. Kent RD, Murray AD. Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *Journal of the Acoustical Society of America* 1982;72:353–365. [PubMed: 7119278]
20. Robb MP, Saxman JH. Acoustic observations in young children’s non-cry vocalizations. *Journal of the Acoustical Society of America* 1988;83:1876–1882. [PubMed: 3403803]
21. Rvachew S, Creighton D, Feldman N, Sauve R. Acoustic-phonetic description of infant speech samples: Coding reliability and related methodological issues. *Acoustics Research Letters Online* 2002;3(1):24–28.
22. Awrejcewicz J. Bifurcation portrait of the human vocal cord oscillation. *Journal of Sound Vibrations* 1990;136:151–156.
23. Wong D, Ito MR, Cox NB, Titze IR. Observation of perturbations in a lumped-element model of the vocal folds with application to some pathological cases. *Journal of the Acoustical Society of America* 1991;89:383–394. [PubMed: 2002176]

24. Mende W, Herzel H, Wermke K. Bifurcations and chaos in newborn infant cries. *Physics Letters A* 1990;418–424.
25. Herzel H, Berry D, Titze IR, Saleh M. Analysis of vocal disorders with methods from nonlinear dynamics. *Journal of Speech and Hearing Research* 1994;37(5):1008–1019. [PubMed: 7823547]
26. Titze, IR.; Baken, RJ.; Herzel, H. Evidence of chaos in vocal fold vibration. In: Titze, IR., editor. *Vocal fold physiology: Frontiers in basic science*. San Diego: Singular Publishing Group, Inc; 1993. p. 143-188.
27. Jiang JJ, Zhang Y, McGilligan C. Chaos in voice, from modeling to measurement. *J Voice*. in press
28. Baken RJ. Irregularity of vocal period and amplitude: A first approach to the fractal analysis of voice. *Journal of Voice* 1990;4:185–197.
29. Michaelis D, Gramss T, Strube HW. Glottal to noise excitation ratio - a new measure for describing pathological voices. *Acustica* 1997;83:700–706.
30. Hanson HM. Glottal characteristics of female speakers: Acoustic correlates. *Journal of the Acoustical Society of America* 1997;101(1):466–481. [PubMed: 9000737]
31. Buder, EH. Acoustic analysis of voice quality: A tabulation of algorithms 1902–1990. In: Kent, RD.; Ball, MJ., editors. *Voice quality measurement*. San Diego: Singular Publishing Group; 2000. p. 119-244.
32. Oller, DK. Infant vocalizations: Exploration and reflexivity. In: Stark, RE., editor. *Language behavior in infancy and early childhood*. New York: Elsevier North Holland; 1981. p. 85-104.
33. Stark, RE. Phonatory development in young normally hearing and hearing impaired children. In: Hochberg, I.; Levitt, H.; Osberger, MJ., editors. *Speech of the hearing impaired: Research, training, and personnel preparation*. Baltimore: University Park Press; 1983. p. 251-266.
34. Zlatin, M. Preliminary descriptive model of infant vocalization during the first 24 weeks: Primitive syllabification and phonetic exploratory behavior. *National Institutes of Health Research Grants*; 1975.
35. Stark, RE. Stages of speech development in the first year of life. In: Ferguson, C., editor. *Child phonology*. 1. New York: Academic Press; 1980. p. 73-90.
36. van der Stelt, JM. Finally a Word: A Sensori-Motor Approach of the Mother-Infant System in its Development toward Speech [dissertation]. Amsterdam: IFOTT Studies in language and language use, University of Amsterdam; 1993.
37. Zlatin-Laufer MA, Horii Y. Fundamental frequency characteristics of infant non-distress vocalization during the first 24 weeks. *Journal of Child Language* 1977;4:171–184.
38. Oller, DK.; Buder, EH. Origins of speech: How infant vocalizations provide a foundation. Paper presented at the Annual meeting of the American Speech-Language-Hearing Association; 2003; Chicago, IL.
39. Oller, DK.; Griebel, U. Contextual freedom in human infant vocalization and the evolution of language. In: Burgess, R.; MacDonald, K., editors. *Evolutionary perspectives on human development*. Thousand Oaks, CA: Sage Publications; 2005. p. 135-166.
40. Buder EH, Stoel-Gammon C. American and Swedish children's acquisition of vowel duration: Effects of vowel identity and final stop voicing. *Journal of the Acoustical Society of America* 2002;111:1854–1864. [PubMed: 12002869]
41. *TF32* [computer program]. Madison, WI: University of Wisconsin- Madison; 2001.
42. Bakeman, R.; Gottman, JM. *Observing interaction: An introduction to sequential analysis*. 2. Cambridge: Cambridge University Press; 1997.
43. Scheiner E, Hammerschmidt K, Jurgens U, Zwirner P. Acoustic analyses of developmental changes and emotional expression in the preverbal vocalizations of infants. *Journal of Voice* 2002;16(4):509. [PubMed: 12512639]
44. Buder, EH.; Oller, DK.; Magoon, JC. Vocal intensity and phonatory regimes in the development of infant protophones. In: Romero, J., editor. *Proceedings of the XVth International Congress of Phonetic Sciences*. Adelaide, Australia: Causal Productions; 2003.
45. Buder, EH.; Oller, DK. Parametric representations for the acoustic study of infant vocalization. Paper presented at the International Conference on Infant Studies; 2004; Chicago, IL.

46. Oller, DK. The emergence of the sounds of speech in infancy. In: Ferguson, C., editor. *Child phonology, Volume 1, Production*. New York: Academic Press; 1980. p. 93-112.
47. Sheinkopf SJ, Mundy P, Oller DK, Steffens M. Vocal atypicalities of preverbal autistic children. *Journal of Autism and Developmental Disorders* 2000;30(4):345–353. [PubMed: 11039860]
48. Hudgins CV, Numbers FC. An investigation of the intelligibility of the speech of the deaf. *Genetic Psychology Monographs* 1942;25:289–392.

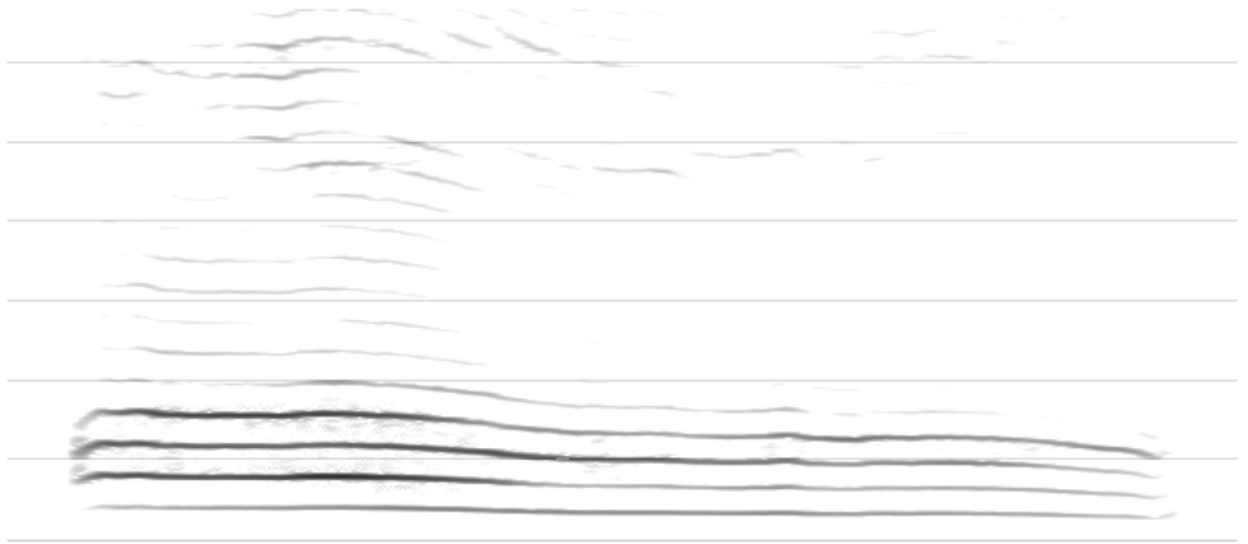


Figure 1.
Spectrographic example of utterance with single modal regime. Total frequency range is approximately 5.6 kHz



Figure 2.

Spectrographic example of utterance with loft regime preceded and followed by modal. Total frequency range is approximately 3.5 kHz. Although some evidence of more complex harmonic structure is visible in the transition from loft back to modal, the episode is shorter than 50 ms and so is not classified as a distinct regime (spurious appearances of transitional overlap are also the unavoidable result of the temporal smearing in extreme narrowband spectrograms).

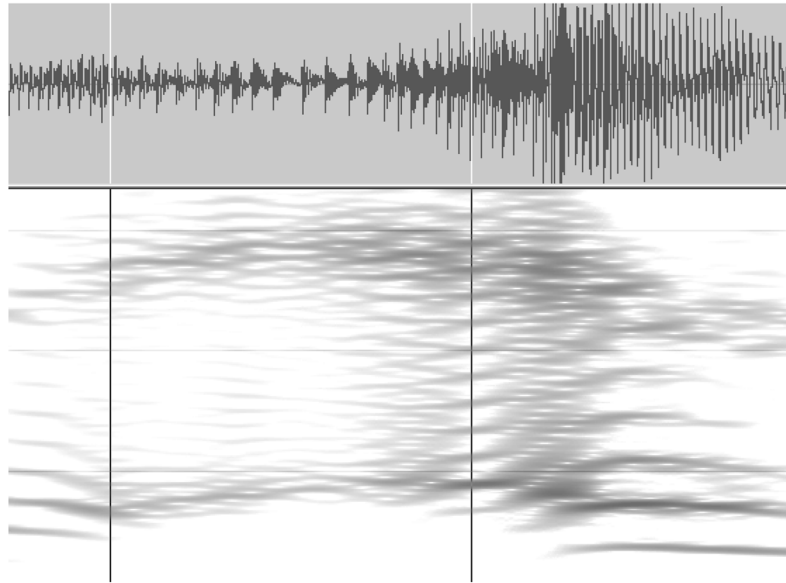


Figure 3. Example of pulse regime preceded and followed by modal, including both a spectrogram with a range of 3.5 kHz and a waveform. The displayed sample is approximately 350 ms in duration, with the segment classified as pulse delimited by cursors. A spurious “cross-hatching” pattern is visible during the pulse phonation caused by the fact that the distance between harmonics begins to approach the effective frequency resolution of the spectrogram.

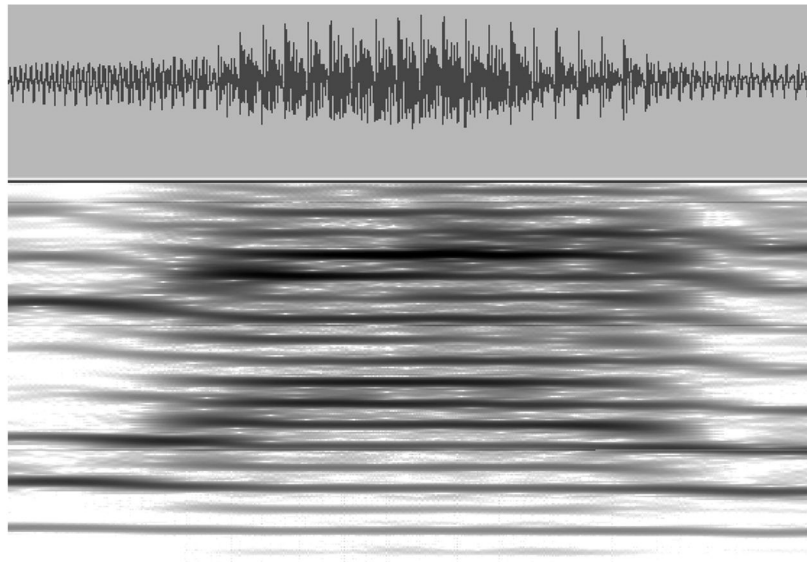


Figure 4. Example of subharmonic regime preceded and followed by modal, including both a spectrogram with a range of 3.5 kHz and a waveform. The displayed sample is approximately 250 ms in duration.

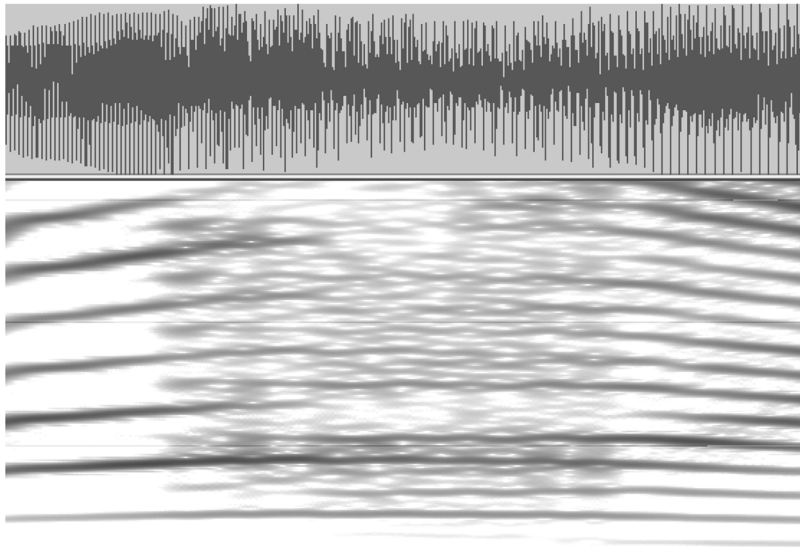


Figure 5.

An example of biphonation that begins with a high modal phonation and ends with a low modal phonation. While some low frequency harmonics appear to be continuous throughout this example this may be an artifact of temporal smearing; it is clear approximately one third through the approximately 500 ms sample that some harmonics are moving in non-parallel directions. The middle segment of the waveform also exhibits a deeply modulated nearly chaotic appearance typical of two overlapping waveforms whose periods are in changing non-integer relationships.

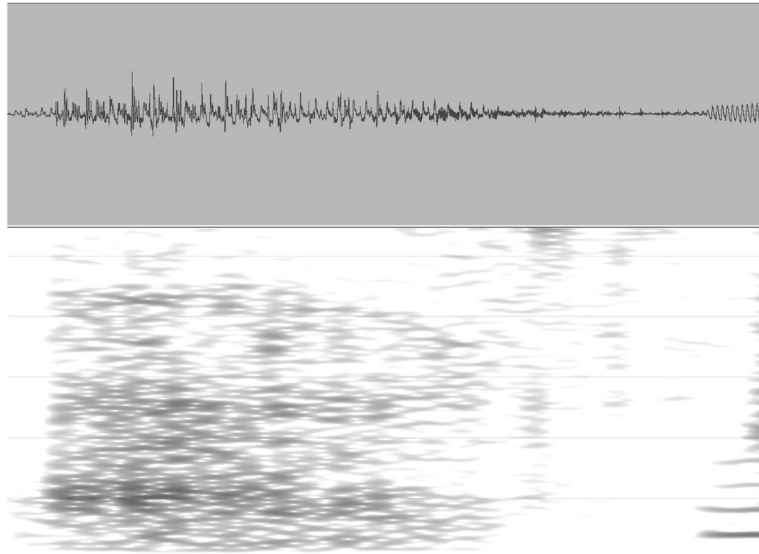


Figure 6. An example of approximately 200 ms of chaos, followed after a pause by some high modal phonation. Note the absence of harmonic structure but a predominance of low frequency energy in the spectrogram, and corresponding waveshapes that appear glottal in form but lacking in periodicity.

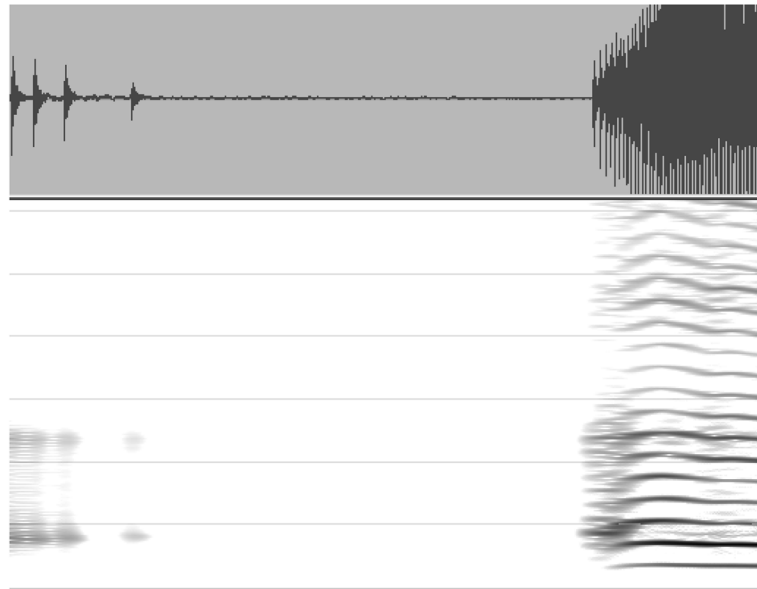


Figure 7.

An example of a closed stop lasting approximately 400 ms. The complete silence during the gap, preceded by a pulse phonation of rapidly decreasing f_0 and followed by a highly abrupt onset of modal phonation (as seen most clearly in the waveform) are features that support this as an especially clear instance of stopping by glottal adduction.

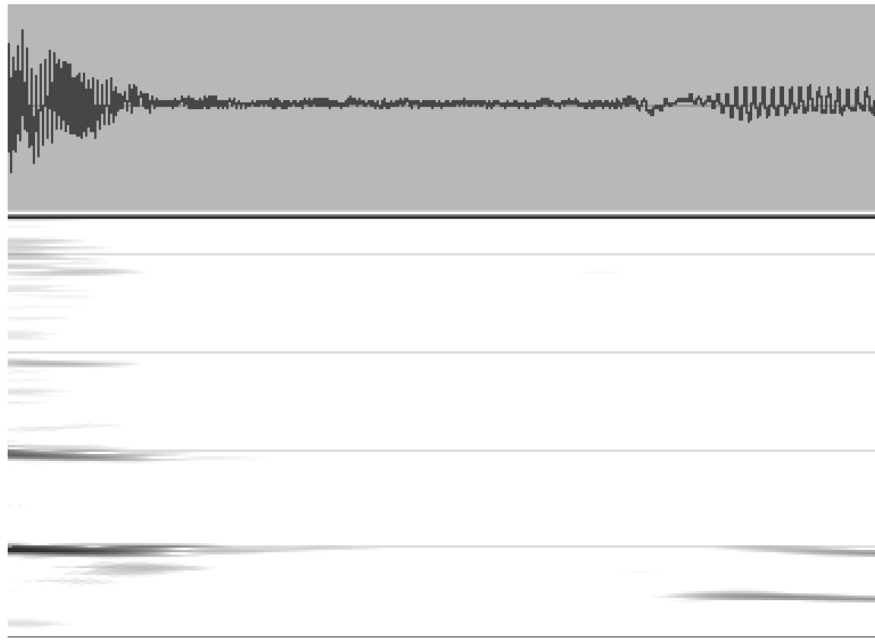


Figure 8. An example of an open stop lasting approximately 100 ms. The waveform exhibits gradual cessation and resumption of voicing, along with some high frequency noise consistent with air turbulence through an open glottis.

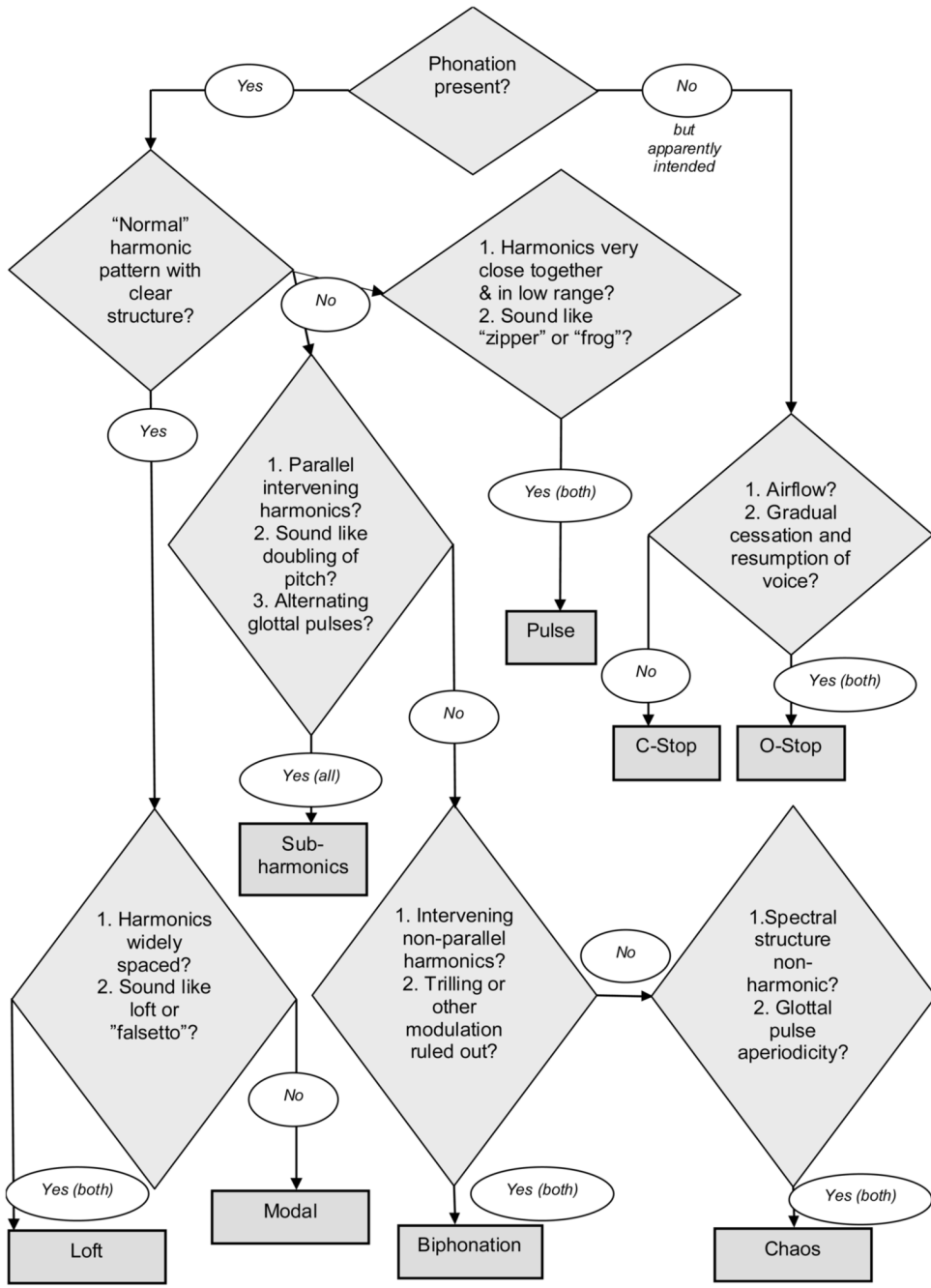


Figure 9. Decision-making tree for regime classification.

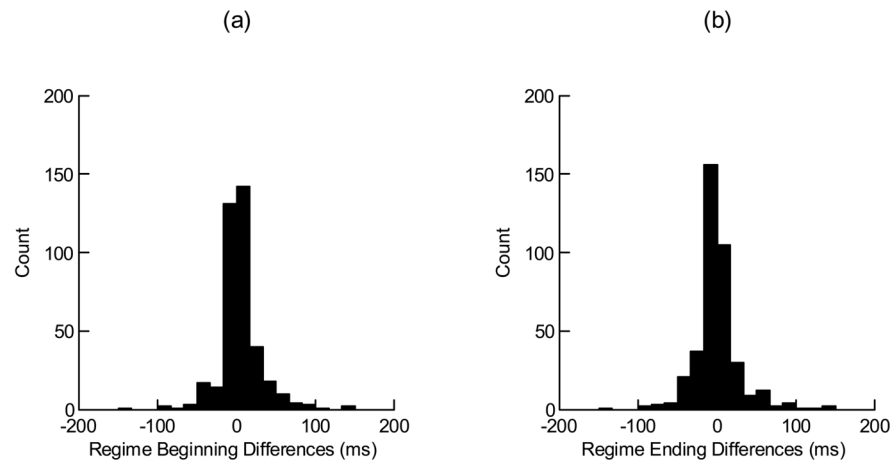


Figure 10. Histograms of differences in timing between analysts' and key's placements of regime beginning boundaries (a) and ending boundaries (b).

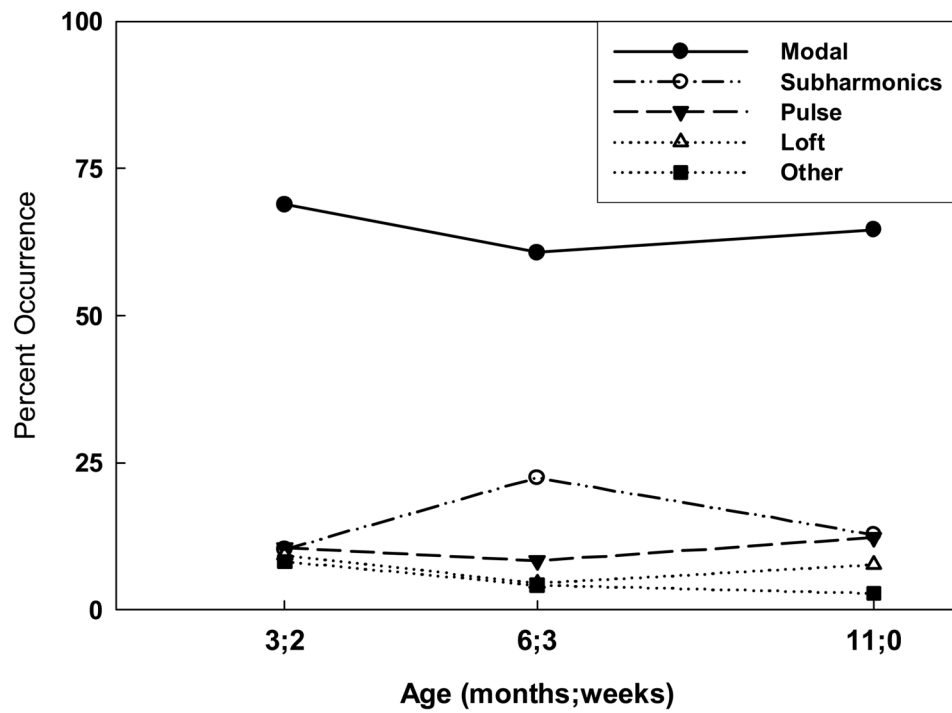


Figure 11. Percentages at which distinct regime types were observed at 3 ages in a girl's first year of life (40 minutes of interaction with mother at each age).

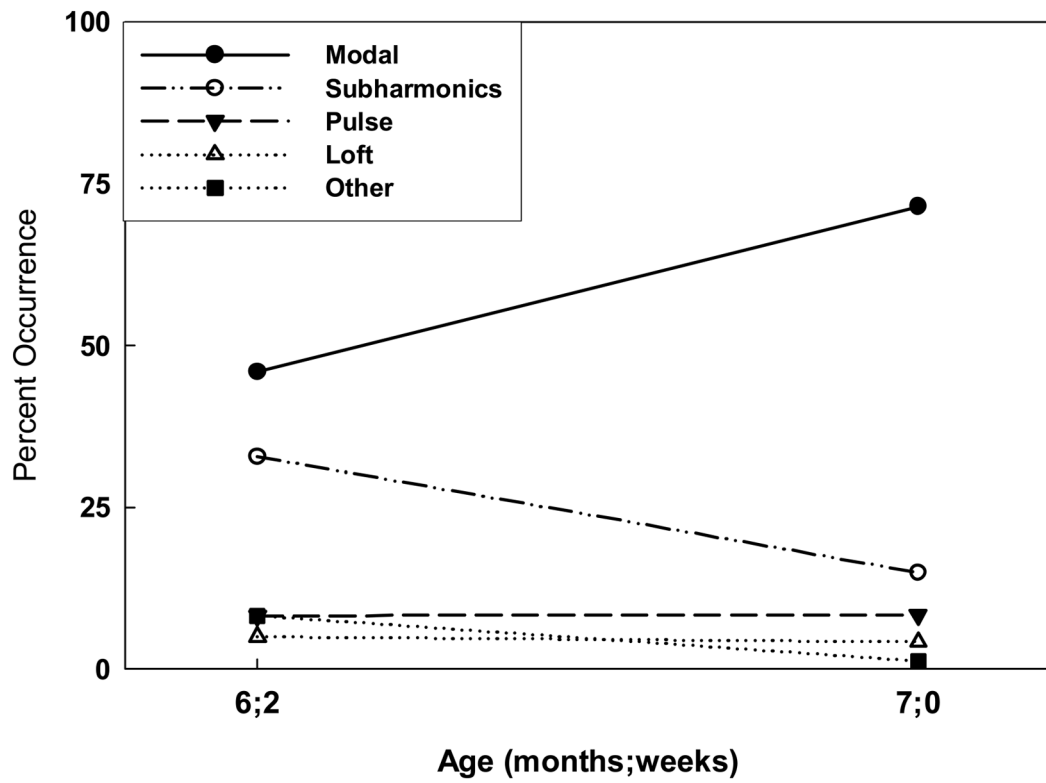


Figure 12. Percentages at which distinct regime types were observed across two longitudinal samples at approximately 7 months of age.

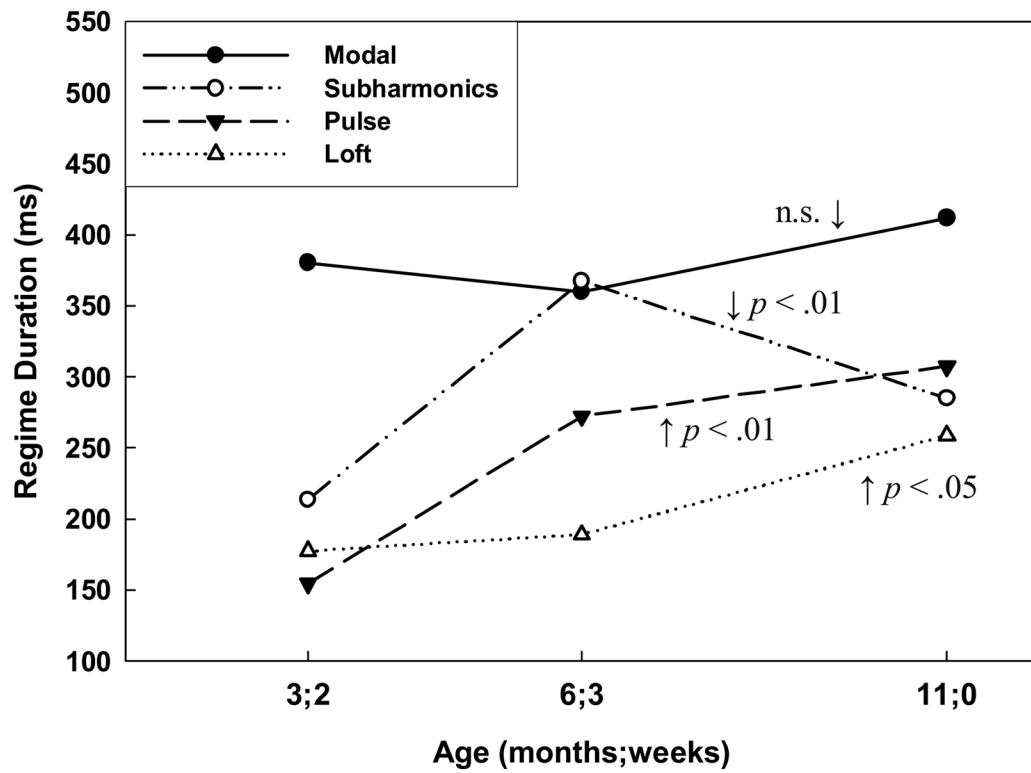


Figure 13.
Mean durations of distinct regime types at three ages (see text for statistical results).

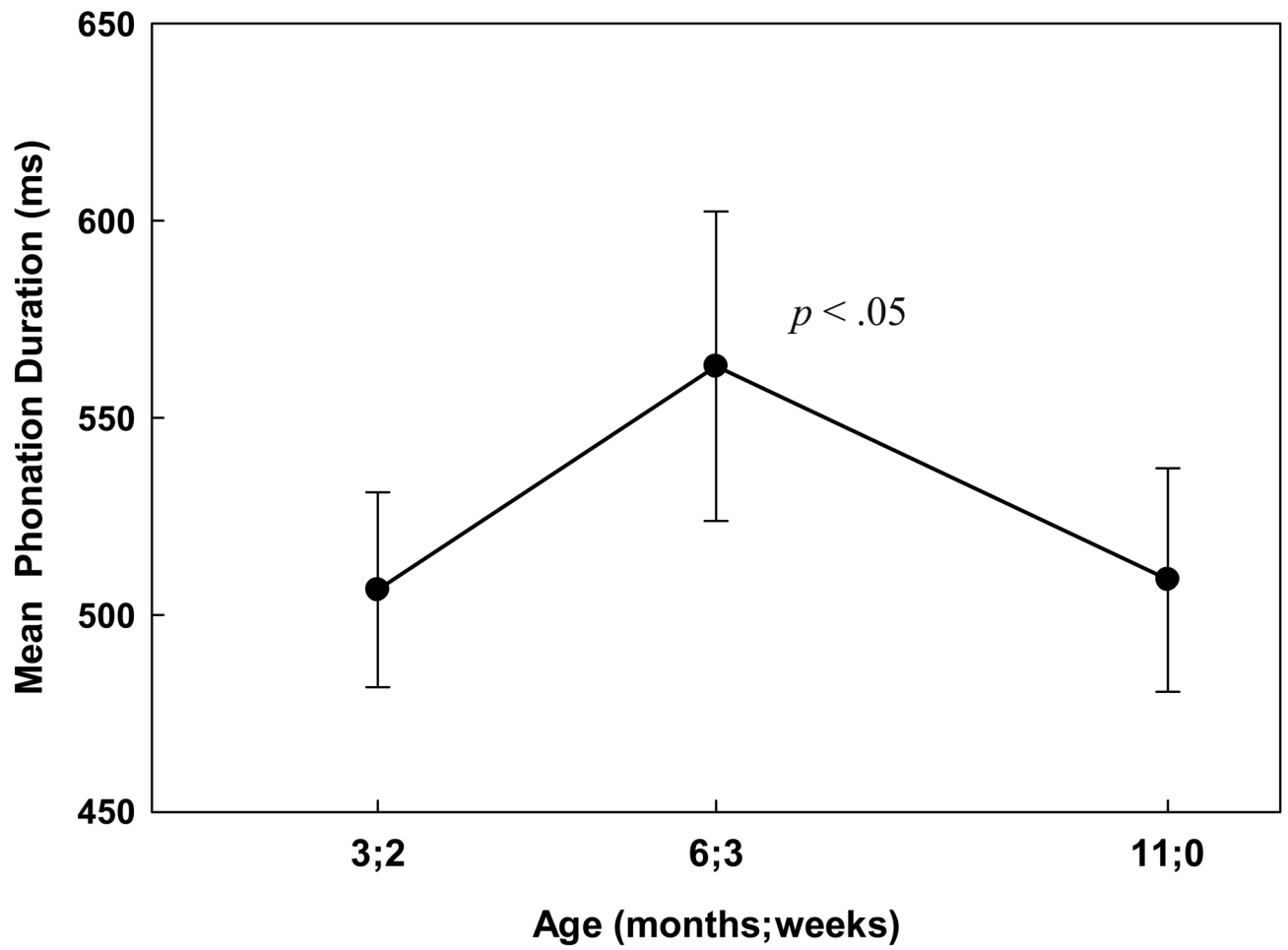


Figure 14.
Mean continuous phonation durations at three ages (see text for statistical results).

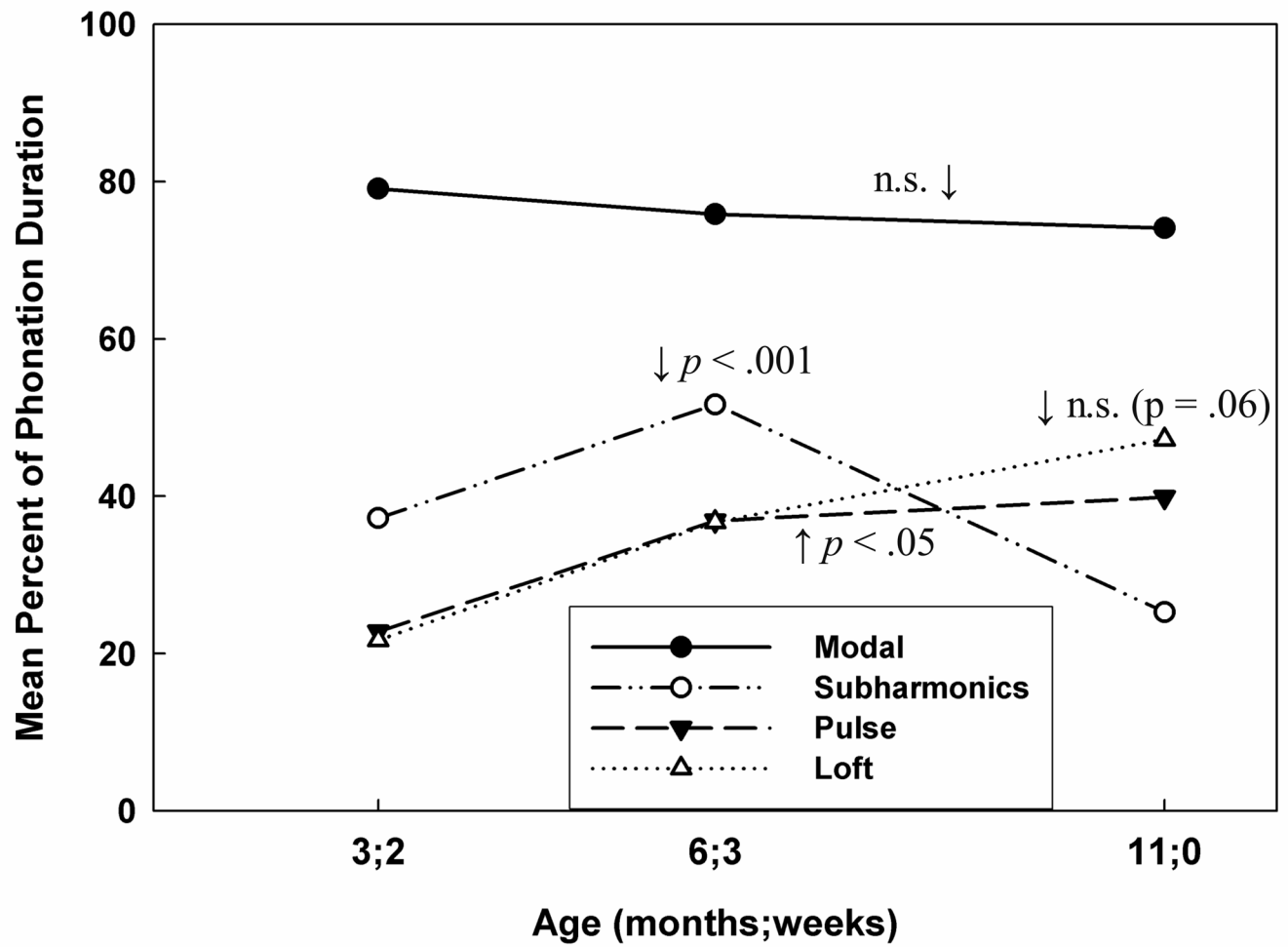


Figure 15.
Mean percentage of continuous phonations accounted for distinct regime types at three ages
(see text for statistical results).

Table 1
List of Regimes

Modal
Loft
Pulse
Subharmonics
Closed-Stop
Open-Stop
Biphonation
Chaos

Table 2
 Errors Made by Analysts' Finding Regime Classifications Where No Eligible Phonation Occurred

Analyst	Regime			
	C-Stop	Chaos	O-Stop	Pulse
A	2	2	2	
B	3	1		
C	1			1
D	7		1	
Total	11	3	3	1

Table 3
 Errors Made by Analysts' Finding No Eligible Phonation Where Regime Classifications Actually Occurred

Analyst	Regime			
	Loft	O-Stop	Pulse	Modal
A	1	1		
B		1		
C	2		1	2
D		1	1	
Total	3	3	2	2
				1

Table 4
 Confusion Matrix Identifying Four Analysts' Codes in Rows and Key Codes in Columns: Absolute Numbers Precede Slash and Percent of Total Key Code Comparisons Follow Slash

	Modal	Loft	Pulse	Subharm.	O-Stop	Biphon.	C-Stop	Chaos
Modal	149/87%							
Loft	6/3%	4/4%	4/11%	10/31%	2/15%	2/17%		1/25%
Pulse	5/3%	84/94%	34/89%			5/42%		1/25%
Subharm.	7/4%	1/1%		16/50%		2/17%		
O-Stop	3/2%			3/9%	10/77%	3/25%		
Biphon.					1/8%		7/100%	
C-Stop								
Chaos	2/1%			3/9%				2/50%
Total	172	89	38	32	13	12	7	4