4-23-2014

Acoustic and Respiratory Characteristics of Infant Vocalization

Maki Doiuchi

Recommended Citation

Doiuchi, Maki, "Acoustic and Respiratory Characteristics of Infant Vocalization" (2014). *Electronic Theses and Dissertations*. 912.
https://digitalcommons.memphis.edu/etd/912

ACOUSTIC AND PHYSIOLOGICAL CHARACTERISTICS OF INFANT VOCALIZATION

by

Maki Doiuchi

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Communication Sciences and Disorders

The University of Memphis

May 2014

Acknowledgements

Over the course of past several years, I have received support from many individuals. Dr. Eugene Buder has been a mentor, teacher, and friend. His course on acoustic phonetics opened my eyes to science. I am greatly appreciative of his insight, patience, and encouragement. I am also indebted to my dissertation committee consisting of Dr. Kimbrough Oller, Dr. Michael Cannito, and Dr. Joel Kahane. Their advice and expertise have been invaluable in the conception and completion of this dissertation. In particular, Dr. Oller's work on infant vocal development is inspiring, and it has been an honor to have the opportunity to work with him. I am also grateful to Dr. Nicholas Duran of Arizona State University for generously providing me with technical expertise.

I would like to thank Dr. Douglas Parham for providing me with his dataset and guidance in respiratory kinematics. I also thank Anna Kharlamova, Laura Green, and Mary Dickinson for completing reliability coding for this study.

My deepest appreciation goes to my family and my parents. I thank my husband, Nick, for believing in me and giving me the tough love I needed to push through. I would also thank my daughter, Olivia, for giving me unlimited happiness and putting up with my ups and downs. And finally, I am eternally grateful to my wonderful parents for their support. My appreciation to them can never be put into words.

Abstract

Doiuchi, Maki. Ph.D. The University of Memphis. May/2014. Acoustic and Respiratory Characteristics of Infant Vocalization. Major Professor: Eugene H. Buder.

The purpose of this dissertation was to explore vibratory regime of infant phonation. The first study examined 1) differences in overall levels of acoustic and respiratory variables between different regimes and 2) differences in relationships between the acoustic and respiratory variables among regimes. The second study examined 3) the acoustic and respiratory ranges of modal phonation with respect to other regimes and 4) the range of modal phonation among infants of different ages.

Two datasets were used in the study. Dataset I was acquired from eight infants of ages 8-18 months, and Dataset II from one infant of ages 4-6 months. Their vocalizations and respiratory movements were recorded during adult-interaction. Phonated segments were identified through waveform, spectrogram, and auditory inspection, and categorized into six mutually exclusive regimes (modal, pulse, loft, subharmonics, biphonation, and chaos). For each regime segment, the following measurements were made: fundamental frequency (F_0), sound pressure level (SPL), expiratory slope, and relative lung volume at regime initiation.

A series of linear mixed-effects model analysis and analysis of variance revealed differences in mean F_0 between regimes, mean SPL, and mean. Correlations between the acoustic and respiratory variables differed among regimes, indicating their relationships were regime-dependent.

The most revealing findings were that regime categories readily distributed into different regions of the intensity-frequency space, and that F_0 ranges of modal regime tended to decrease with increasing age. In addition to modal, pulse, and loft distributing around the mid, low, and high intensity-frequency regions, respectively, biphonation and subharmonics were found between modal and loft ranges. The upper end of F_0 range for pulse was much higher in infants compared to adults, however, biphonation and subharmonics rarely occurred between pulse and modal ranges. A range of modal F_0 was about 500 Hz for the young infant in the vocal expansion stage, and about 200 Hz for older infants in the (post-)canonical stage. Although the results are tentative, this finding suggests that F_0 variability decreases with age and phonation becomes more restricted to a lower end of an F_0 range.

Table of Contents

CHAPTER PAGE

List of Tables

List of Figures

FIGURES PAGE

Chapter 1: Introduction

After discovering vocal sounds early in life, infants produce a variety of sounds to signal physical states, express emotions, and facilitate social interaction. Their vocalizations change systematically during the first year to culminate in their "first words," which is the meaningful use of a recognizable speech form. Research shows that even though young infants lack referential language, their first year is important in forming foundations for both social communication and sound production skills for later speech (Koopmans-van Beinum & van der Stelt, 1986; Oller, 1980, 1986, 2000; Oller, Wieman, Doyle, & Ross, 1976; Stark, 1980; Stark, Bernstein, & Demorest, 1993). In particular, the relevance of babbling to speech is well established, and articulatory development in late infancy has been studied extensively. On the other hand, details of phonatory development remain relatively unexplored despite the fact that phonation is a primary means of communication during infancy.

Modal phonation, which results from quasi-periodic excitation of the vocal folds, is the standard type of voice in adult speech. It forms a starting point of development in many infant vocalization studies, serving as the first defining trait of emerging speech-like quality (Oller, 2000). With modal phonation being the most powerful sound composing a syllable, it is logical that its acquisition precedes the production of more advanced forms of vocalization. However, studies have reported that acoustic parameters of phonation are variable throughout infancy and young childhood. Young children are known to produce non-modal phonation that is typically rare in adult speech. Infant voice is acoustically rich as illustrated by observations of a wide range of fundamental frequency (F_0), sound pressure level (SPL), and phonation types (Buder, Chorna, Oller, &

Robinson, 2008; Keating & Buhr, 1978; Kent & Murray, 1982; Robb & Saxman, 1988), which interact to form perceptually distinct categories such as squeals and growls (Buder & Oller, 2004; Buder, Oller, & Magoon, 2003; Oller, 2000; Oller et al., 2013).

Acoustic studies of phonation have so far most frequently focused on adult and disordered populations, and relatively little work has been done on infants. By far, the most studied parameter of infant voice is a mean F_0. Although informative, one measure of central tendency of a single variable can only provide incomplete information about developing phonatory ability. Physiological studies on infant vocalization are even scarcer. The first few years of life are a period of marked physical, motor, and cognitive development. Different aspects of speech production are managed more or less separately by different organs, each of which follows its own developmental schedule to achieve physical growth while fulfilling its required function. It may be that phonatory control has a developmental course of its own that is independent of the articulatory development, however, it is not clear at this time how modal phonation observed in early months might differ from mature phonation, how infants regulate their larynx and respiratory apparatus to produce voice, or how developing anatomy and physiology may contribute to phonatory variability.

Long-term goals of the research along these lines include developing measures of phonatory control and understanding the developmental course of phonation. More immediate questions addressed here are related to modal phonation and its acoustic and respiratory characteristics. This dissertation consists of two studies examining F_0, SPL, relative airflow, and relative lung volume level obtained from audio and respiratory kinematic signals during vocalization. Two datasets were used in the studies. Dataset I was an existing dataset collected for a previous study (Parham, 2008) from eight infants of ages between 8-18 months. Dataset II was collected for the purpose of this dissertation from a single infant between ages of 4-6 months.

Using Dataset I, Study 1 examined overall levels and relationships of the acoustic and respiratory variables between different vocal fold vibratory regimes. Because human voice results from the modulation of airflow by the vocal folds, how input respiratory drive and output acoustics are related is one of the key aspects in understanding vocal production. Findings of Study 1 demonstrated the way they interact is dependent on vibratory regime, highlighting the need to incorporate the concept of vibratory regime in vocal analysis.

Study 2 examined ranges of F_0, SPL, and relative airflow associated with modal phonation with respect to non-modal phonations as well as with respect to age. Unlike most previous studies on the same topic, the current study focused on a range rather than a mean of F_0 because mean values obscure where modal phonation is sustained and where it breaks down. Acoustic and respiratory ranges for non-modal regimes were examined to determine how they were distributed around modal phonation and if certain regimes appeared in systematic regions. Variable phonation of younger stage transitions to modal phonation of well-formed syllables as the infant matures. However, it is uncertain how this preference toward modal may be reflected in ranges of modal phonation. Although results would be tenuous because data on early infancy are obtained from a single participant, developmental comparisons were made between younger (Dataset II) and older (Dataset II) infants? The modal ranges may 1) be unchanged across age and non-modal phonations may tend to disappear, 2) expand, possibly indicating that the maturing vocal folds (e.g., development of vocal ligament) are able to oscillate more steadily over wider ranges without breaking into other regimes, or 3) diminish, possibly indicating a learning effect, meaning that the infant is seeking an optimal range of vocal fold vibration where modal phonation can be more readily sustained. Examining changes in distribution patterns of modal phonation at different ages will help formulate more concrete hypotheses in future studies.

Chapter 2: Literature Review

Previous studies on topics related to infant vocalization are reviewed in the subsequent section. Each topic and its subtopics describe findings in the literature that are important in formulating specific research questions of this study. First, models of infant vocalization are outlined to highlight important stages of vocal development. Second, vocal fold vibration is described in the framework of nonlinear dynamic systems principles as they provide an organizational framework for conceptualizing and evaluating phonatory abilities in early infancy. Third, published data on acoustic characteristics of infant vocalization are summarized. Lastly, structural as well as functional changes of the larynx and respiratory system are reviewed as they are presumed to play an important role in shaping output vocalization.

Vocal Production in the First Year of Life

Stages of vocal development. One of the developmental milestones in an infant's first year is the production of their first word around 12 months of age. Research has shown that pre-linguistic vocalizations are related to early speech, bearing similar infraphonology (Blake & Boysson-Bardies, 1992; Boysson-Bardies, Sagart, & Bacri, 1981; Oller et al., 1976; Vihman, Ferguson, & Elbert, 1986; Vihman, Macken, Miller, Simmons, & Miller, 1985). Ranging from birth cries to remarkably speech-like gibberish, infants produce diverse sounds during their first year. By studying salient types of vocalizations at different age, researchers have uncovered step-by-step progression in which infants acquired their ability for articulated speech in the pre-linguistic period (Koopmans-van Beinum & van der Stelt, 1986; Oller, 1980, 1986; Stark, 1980).

Definitions and timing vary from author to author but they align on general characteristics. It is important to note that the developmental sequence described below is a common trend seen in most typically developing children but not without inter- and intra-individual variability in types of vocalization and frequency of their occurrences. Vocalization types that are identified as relevant of a particular stage may not necessarily be the most frequent ones for all individuals, nor within the same child over a course of a day, week, or month. Moreover, certain vocalization types can be frequent during a particular period of time and infrequent but likely not completely absent elsewhere. Therefore, isolated productions of a certain vocal behavior do not necessarily mark onset of a particular stage of development. Stages of pre-linguistic vocal development are summarized in Table 1, and briefly described as the following:

> Stage 1 (0-2 months): Many vocalizations are associated with reflexive reactions. However, among vegetative and distress sounds, infants also make vowel-like sounds with limited resonance.
>
> Stage 2 (1-4 months): Infants produce comfort sounds. They often show crude articulation in the form of gooing or cooing, which is a simple alternation of opening and closing of the vocal tract.
>
> Stage 3 (3-8 months): Infants' vocal output shows rich diversity in phonatory parameters.
>
> Stage 4 (5-12 months): Infants produce reduplicated and variegated babbling, which is a series of consonant and vowel elements.

Table 1

Stages of Vocal Development during the First Year

	Age in months											
Author	1	2	3	4	5	6	7	8	9	10	11	12
Oller (2000)	Phonation stage Normal phonation											
		Primitive articulation stage Limited articulation										
				Expansion stage Full resonance and marginal articulation								
						Canonical stage Well-timed articulation						
Koopmans-van Beinum and van der Stelt (1986)	Stage 1 Uninterrupted phonation											
		Stage 2 Interrupted phonation										
			Stage 3 Single articulation									
						Stage 4 Phonatory variations						
								Stage 5 Reduplicated articulatory movements				
Stark (1981)	Stage 1 Reflexive vocalizations											
		Stage 2 Cooing, laughter										
				Stage 3 Vocal play								
								Stage 4 Reduplicated babbling				
										Stage 5 Proto-words		

This general developmental pattern described above has been studied in various populations and found consistent irrespective of ambient language (Levitt & Aydelott Utman, 1992; Nakazima, 1962; Oller & Eilers, 1982; Tanaka, 1968). Researchers have also looked at possible effects on early vocalization of various physical, neurological, and environmental factors that are known to affect language development later in life. Interestingly, it was discovered that early vocal development tends to be resistant with respect to conditions such as motor or cognitive deficits of Down syndrome (Dodd, 1972; Oller & Seibert, 1988; Smith & Oller, 1981; Stoel-Gammon, 2001); bilingualism (Nakazima, 1962; Oller, Eilers, Urbano, & Cobo-Lewis, 1997); prematurity, socio-economic status, and sibling effect (Eilers et al., 1993; Irwin, 1952; Oller, Eilers, Steffens, Lynch, & Urbano, 1994).

On the other hand, the presence of various structural or cognitive deficits have been shown to adversely affect babbling. Delays and atypical vocal patterns have been reported in cases of unrepaired and repaired cleft palate (Chapman, Hardin-Jones, Schulte, & Halter, 2001; F. Murai, Arai, & Kimura, 2005; Willadsen & Albrechtsen, 2006), cerebral palsy (Levin, 1999), focal brain injury (Marchman, Miller, & Bates, 1991), mental retardation of various etiologies (Oller & Seibert, 1988), autism (Sheinkopf, Mundy, Oller, & Steffens, 2000), cri du chat syndrome (Sohner & Mitchell, 1991), and hearing impairment (Eilers & Oller, 1994; Lynch, Oller, & Steffens, 1989; Oller & Eilers, 1988; Smith, 1982; Stark, 1983; Stoel-Gammon & Otomo, 1986). These studies have highlighted the importance of structural and neurological intactness, social and auditory feedback, and child's desire to communicate in normal vocal development, however, the overwhelming universality and robustness of babbling suggested that at least some degree of biological propensity was at work for very early vocal behavior.

The similarity of the developmental trends across studies is clear, especially towards the end of the first year. Around 6-8 months of age, infants begin to produce canonical syllables (Oller, 1980). As canonical syllables come into infants' vocal repertoire, their utterances dramatically resemble speech. This is because consonant-vowel (CV) syllables are basic building blocks of spoken language (Oller, Eilers, Neal, & Cobo-Lewis, 1998; Oller, Eilers, Neal, & Schwartz, 1999). Therefore, from a standpoint of development of speech form, canonical babbling is one of the key steps of pre-linguistic vocal development. Once it is established that vocal behavior progresses systematically toward the production of canonical babble before the onset of speech, it becomes imperative to address issues such as where speech originates and what drives the progression from one stage to another. Researchers agree that vocal development is a process of successive integration of developing subsystems: structural, neuro-motor, auditory perceptual, and cognitive. Anatomical growth and physiological maturation obviously impact vocal output but also, there is little doubt that self-monitoring, through auditory and social feedback, is important in linking phonatory and articulatory gestures to their acoustic consequences (Fry, 1966), and communicative contexts in establishing sound-meaning correspondence (Oller, 2000). How maturation in one of the subsystems may contribute to qualitatively distinct categories of vocalizations at each stage is a highly complex issue that needs to be studied from various perspectives. To better understand the nature of developmental changes, works by three groups of researchers are reviewed below.

Models of vocal development. Early researchers of infant vocalization often adopted a transcription method in describing infant sounds (Chen & Irwin, 1946; Irwin, 1946, 1947a, 1947b, 1947c, 1948; Irwin & Chen, 1943, 1946). Although occasionally very young infants do produce sounds that resemble vowels and consonants of adult speech, transcription at early ages is problematic in several ways.

First of all, many infant vocalizations are exotic in qualities and dissimilar to those found in natural language. By limiting analysis only to phoneme-sounding segments, many other vocalizations are treated as uncategorizable and therefore neglected. Secondly, even when some sounds are classified as certain speech sounds, adult phoneme perception is often fundamentally affected by one's native language, making it difficult to identify unfamiliar sounds reliably. Skilled phoneticians are able to overcome this difficulty to some extent, but the amount of training or experience that the transcribers had in many studies is unspecified. Thirdly, a phoneme encompasses a range of physical variations that are considered identical in the mind of a language user. Frequently, subtle differences that are not easily discernible to non-native speakers are made clearly perceivable to native speakers because of systematic contrasts among phonemes that exist in a particular language system.

In short, it is inappropriate to impose phonemes on infant sounds because how sounds are contrasted in the pre-linguistic stage does not abide by rules of mature language. To circumvent methodological issues in using transcription, many researchers took a more instrumental approach and shifted the level of analysis to measurements of various acoustic features. However, before acoustic analysis can be useful, a novel approach was needed to guide interpretation of quantitative analysis and study infant vocalization.

Oller (1980, 1986, 2000) explains the developmental sequence as progressive acquisition of infrastructural capabilities underlying speech. Some components such as normal phonation, oral resonance, and syllabification are fundamental bases of spoken languages of the world (Greenberg, 1965; Ladefoged, 2006; Ladefoged & Maddieson, 1996). In Oller's view, control over these phenomena can yield *speechiness* and thus such control constitutes skills of speech production that develop in the pre-linguistic period. His analyses focus on non-reflexive, non-distress sounds, emphasizing infant's tendency to vocalize in the absence of stimuli as a unique ability relevant to speech. As infants advance across stages, infant vocalization increases in its speechiness. In order to lay a foundation for later speech, infants are actively involved in learning to approximate their vocal output to that of mature speech.

More specifically, Oller's theory of infraphonology centers on principles that define well-formedness of a syllable. Characteristics of a well-formed CV syllable are: 1) a nucleus produced by normal phonation, 2) vocal tract opening and closing during phonation, 3) modulations of F_0 and SPL, 4) oral resonance, 5) rapid and smooth formant transitions between a consonant and a vowel, and 6) a series of CV syllables produced within 100 to 500 milliseconds (Oller, 1986, 2000).

According to Oller (1980, 1986, 2000), normal phonation is present in the first months of life in the form of quasivowels, which are vowel-like utterances produced with normal phonation but without full oral resonance because the vocal tract is at rest. Next, a primitive articulatory gesture is evidenced in gooing. Articulation at this point, however, preferentially involves a point of constriction with the posterior portion of the oral cavity, thus lacking the flexibility of mature articulation, and being seemingly disorganized and unpredictable in terms of duration and manner, and even whether or not a vowel-like nucleus can be identified. The next stage is referred to as the expansion stage in which vocalizations are distributed in wide ranges of F_0 and SPL. Oller interprets this variability as exploration of vocal potential. Another notable achievement of this stage is increased oral resonance. An open vocal tract accompanies normal phonation to result in a fully resonant nucleus or full vowel. A rapid formant movement associated with vocal tract opening and closing marks the beginning of the last stage. The canonical stage is developmentally important for it reflects a controlled production of well-formed syllables that are building blocks of spoken words (Oller et al., 1998). In Oller's model of infraphonology, the early vocal development is an active process, in which the infant systematically exploits his vocal capabilities at a different stage to produce vocal sounds with certain speech-like properties.

Koopmans-van Beinum and van der Stelt (1986) take a similar perspective. They relate description of infant vocalizations to perceivable physiological events. That is, without taking actual physiological measurements, they attempted to describe infant sounds in terms of underlying vocal production mechanism. They created a simple non-segmental transcription system based on the source-filter model of speech production so that infant vocalizations can be judged for their syllabicity. Their analysis of non-crying, non-vegetative sounds showed continuous shifts in phonatory and articulatory characteristics as phonological development progressed in the first year.

The first stage in Koopmans-van Beinum and van der Stelt (1986) is a period of continuous phonation, without articulatory movement. In their model, utterances with interrupted phonation series (i.e., with intervening glottal stops) are regarded as stage 2 for they are deemed to be more syllable-like and also showing essential ability of turning voicing on and off. Utterances in stage 3 contain one articulatory movement. Increased syllabicity is achieved by a supraglottal constriction rather than a glottal stop. In stage 4, previously acquired articulatory movements become infrequently used, and instead uninterrupted phonation becomes prevalent. Utterances in stage 4 are different from those observed in stage 1 in that they show great variability in F_0, SPL, and duration. Finally in stage 5, the focus of development shifts back to articulatory movements and repetitive CV sequences become observable.

In Stark's model (1978, 1980, 1981), stage 1, the few months following birth, is predominated by reflexive sound making of discomfort sounds (i.e., both crying and fussing) and vegetative sounds. The former sound type is characterized by features of voicing, egressive flow, vowel-like, open vocal tract, while the latter by voicelessness, short duration and consonant-like quality. Soon cries become less prominent as comfort sounds emerge in infants older than 3 months, often in association with caregiver interaction. Cooing and laughter in stage 2 are, in Stark's terms, a combination of consonant-likeness of vegetative sounds and vowel-likeness of cries. Various sound types, such as squeals, nasals, and noises, occur frequently in stage 3, the vocal play stage. Distinct perceptual qualities often reported with these types of sounds are taken as isolated portions of laughter. Segments of laugh sound are disassembled, elaborated, and manipulated along the dimensions of duration, F_0, and SPL. Some of these vowel-like and consonant-like elements may be produced one after the other towards the end of stage 3, however, the sequences lack the timing of coordination required for adequate syllabification. More rapid, regularly timed vocal tract opening and closing are achieved in stage 4 when the infant begins to produce canonical babbling.

Stark (Stark, 1978, 1980, 1981; Stark, Rose, & McLagen, 1975) takes an approach slightly different from the other two. She examined a wide variety of sounds, including vegetative, discomfort, and laugh sounds in addition to comfort sounds, as relevant to later speech sounds. She theorizes that all relevant features of speech sounds, such as breath direction, voice quality, pitch, loudness, vocal tract openness, and manners of articulation to be present in earliest reflexive sounds and the vocal development to be repetitive restructuring of elements already found in those primitive reflexive sounds. Stark's view of vocal development as recombination of existing sound properties is somewhat in contrast to Oller's view of accumulation of new speech skills. Nonetheless, many of her features overlap Oller's infraphonological properties, and the similar and lawful developmental patterns are maintained.

Phonatory and articulatory development. The common thread in all three accounts is developing syllabicity. The vocal development is organized into a single hierarchical model, in which properties of a well-formed syllable emerge in steps. When examining the developmental sequence, it is noticed that some steps are in a conditional relationship, whereas others are less so. It is obvious that production of a single articulatory gesture should precede multiple articulatory gestures within one utterance. If normal phonation and single articulatory gesturing are already present in the first few months of life, multiple articulatory movements should follow. However, the development does not simply move from single to multiple articulatory gestures. There is an intermediate step where focus of development appears to shift from articulation to phonation.

The stage is referred to as "expansion stage" (Oller, 2000) "vocal play" (Stark, 1980), or "phonatory variations" (Koopmans-van Beinum & van der Stelt, 1986). The main infraphonological achievement of this stage is oral resonance characterized by a fully open vocal tract, but arguably more notable is the great variability in phonatory parameters such as F_0, SPL, and duration. Koopmans-van Beinum and van der Stelt (1986) describe this stage as a period of *seeming relapse,* suggesting, although without elaborating on why it may be, that the developmental target derails from articulatory improvement. Oller (2000) and Stark (1980) offer an interpretation in which infants are actively exploring their phonatory space or practicing different aspects of their voice. Oller (2000) further argues to claim that systematic use of distinctive regions of F_0 or SPL creates primitive sound categories that are contrastive (e.g., squealing versus growling, and yelling versus whispering).

At first glance, this phenomenon seems somewhat inconsistent with the step model because if normal phonation is an ability that underlies speech and is already established in the earliest stage of development, then the use of extreme ranges of F_0 and SPL is something that should be avoided to maintain modal phonation. It can be interpreted that because infants have control over phonation, glottal parameters are the first means of categorization at infants' disposal. Or, it can be alternatively argued that the expansion stage represents ongoing development of phonatory ability.

As will be reviewed in subsequent sections, there is acoustic evidence that support that phonatory variability observed in 3 to 8 months persists to some extent beyond the canonical stage. In addition, there is a great amount of structural independence among vocal subsystems. Since phonation is a conversion of airstream into sound, it is primarily contributed by the respiratory system and the larynx, whereas resonance and articulation are mainly supraglottal events regulated by the vocal tract. The structural separation seems to suggest that the integration of the two in speech production does not necessitate normal phonation to be a prerequisite to later full vowel or canonical babble productions. There are findings in developmental anatomy and physiology of the respiratory system and vocal fold that may provide insight into why infant's voice may differ from modal phonation in mature speech. Then it becomes a valid question to ask if phonation too has a logical developmental pattern of its own. Figure 1 presents conceptualization of vocal development hierarchy where phonation and articulation have separate developmental lines that interact. Addressing vocal development in terms of both phonatory and articulatory abilities, as well as how they are integrated at a particular stage of development may help further understand vocal development.

Summary. Over 30 years of research have revealed an orderly course of vocal development in the first year of life before the appearance of meaningful word productions. It has been established in the literature that pre-linguistic vocal development proceeds in stages, each of which reflecting emerging ability for articulated speech. Although phonation is integrated in current developmental models, the bulk of vocal development research has focused on canonical babbling as it is thought to facilitate later speech. In particular, the level of mastery of normal phonation in infancy has not been studied in detail, and there is little empirical investigation on how this preferred source of speech develops as the infant matures.

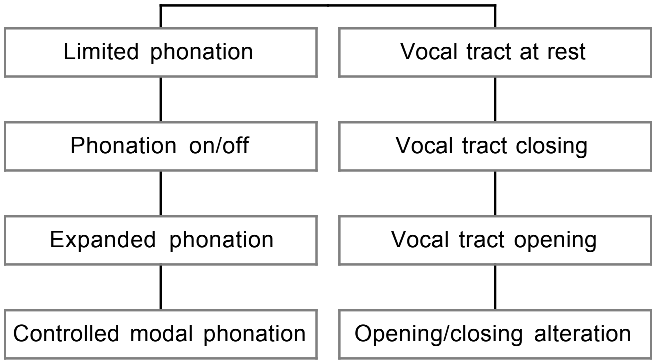


Figure 1. A schematic of vocal development sequence with distinct lines for phonation and articulation.

Vocal Fold Vibration

The vibrating vocal folds constitute the major source of sound used in human communication. Voice is not only present for many speech sounds, but also essential to singing, cries, and laughter. The act of simply producing voice has an importance in signaling and releasing emotions. In speech, speakers show a great degree of control over fine aspects of voice to express attitudes, individuality, moods, and affects. Voice encompasses a family of laryngeal sounds that arise from various patterns of vocal fold vibration. How vocal sounds are produced by the vocal folds is an important aspect of the study of infant vocalization.

Basic mechanism of phonation. The basic mechanism of phonation has been understood for sometime now (Titze, 2000; Van den Berg, 1958). Many models have been developed to explain certain key aspects of the mechanism. In particular, a one-mass model, multi-mass models, and a mucosal wave model are helpful in explaining pressure asymmetry, the driving force of self-sustaining oscillation (Ishizaka & Flanagan, 1972; Titze, 1988). A basic process of phonation is described as follows according to a two-mass model (Titze, 2000). To initiate, the vocal folds must be taut and approximated to the midline. This creates resistance to airflow rising from the lungs, causing pressure to build up underneath. At some point, this subglottal pressure overcomes the glottal resistance and blows the vocal folds apart from bottom to top. During this opening phase of a cycle, the vertical shape of the glottis causes the passing airflow to converge, resulting in an increase in intraglottal pressure as well as lateral force exerted onto the vocal folds. The tissue moves laterally until its elastic force overcomes the lateral force to bring the vocal folds back toward the midline. During this closing phase, the bottoms of the vocal folds again lead the tops. This time, the shape of the vocal folds creates a divergent airflow, resulting in a decrease in intraglottal pressure because air is not coming from below as fast as it is leaving above. Consequently the lateral force too decreases. The vocal folds close, pressure builds again underneath the adducted vocal folds, and a cycle repeats. Alternating asymmetry in driving force and tissue elasticity is essential to self-sustained oscillation.

During a closing phase, negative pressure in the glottis causes the vocal folds to shut abruptly, stopping airflow momentarily. This sudden cessation of airflow creates pressure disturbance above the glottis. Rapid succession of regularly timed glottal pulses make up voice. Each glottal cycle has an asymmetric flow pattern, that is, an increase and decrease of airflow through the glottis is not the same. This asymmetry gives rise to a spectrum of frequencies that are harmonically related to the fundamental glottal cycle.

There are three aspects of phonation that are particularly important in characterizing voice: the rate of the vocal fold vibration, intensity of the sound, and vocal quality. Although phonation is a complex process, some of the physiological control mechanisms have been understood.

The natural resonance of the vocal folds is determined largely by overall mass and length of the vocal folds. Normative values of speaking F_0 are roughly 115 Hz for men, 215 Hz for women, 240 Hz for boys, and 243 Hz for girls (Awan & Mueller, 1996; Hollien & Shipp, 1972; Saxman & Burk, 1967), reflecting age and sex differences. In general, men have thicker and longer vocal folds than do women or children. Such vocal folds are expected to vibrate at a slower rate. F_0 can be varied more locally by changing the stiffness and mass of the vocal fold. Actions of two antagonistic muscles, the cricothyroid and the thyroarytenoids, are important in controlling these parameters. Unopposed contraction of the cricothyroid muscle lengthens the vocal fold and increases F_0 because it increases the stiffness and decreases mass per unit length. On the other hand, unopposed contraction of the thyroarytenoid muscle shortens the vocal fold and decreases F_0 because it slackens the cover. Under some phonatory conditions, both the cover and body are involved in oscillation. In such cases, the activation of the thyroarytenoid muscle increases stiffness of the vibrating tissue, which results in an increase of F_0 (Titze, Luschei, & Hirano, 1989). Therefore, the net effect of F_0 change is determined by the relative contribution of these muscles, which changes dynamically during phonation. Yet another contributing factor is subglottal pressure. Larger amplitude of vibration caused by increased subglottal pressure stretches, or increases strain of, the vocal fold. This raises F_0 because there is a greater amount of elastic recoil force (Titze, 1989). Although the effects of subglottal pressure on F_0 are not as pronounced as those of the muscles, it is an important factor to consider because it serves as a source of covariation of F_0 and SPL.

Vocal intensity is usually measured in sound pressure level (SPL). It is primarily determined by subglottal pressure besides vocal tract resonances that affect levels of harmonics. A typical value of conversational SPL is reported to be around 55-70 dB depending on a mouth-to-microphone distance (Gelfer & Young, 1997; Pearsons, Bennett, Fidell, & Bolt, 1977). Sex and age differences are less marked compared to those of F_0. It is somewhat intuitive that the greater the subglottal pressure, with everything else being equal, the greater the SPL because a greater amount of air is moved through the glottis. There are two distinct ways to regulate subglottal pressure. One is by way of glottal closure and resistance. An increase in the duration, speed, and degree of closure is associated with higher SPL (Isshiki, 1964; Stathopoulos & Sapienza, 1993). However, a later study reports that subglottal pressure is more effectively increased by raising lung pressure than by glottal adjustments (Finnegan, Luschei, & Hoffman, 2000). The same study also confirms the aspect of vocal fold that affects lung pressure control of SPL is the looseness of the cover (i.e., rather than medial compression, both of which are related to tension), which determines the amplitude of mucosal wave.

The third dimension of voice is its quality. It is a perceptual thus subjective phenomenon that describes voice beyond its F_0 and SPL characteristics. Unlike F_0 or SPL, it is hard to quantify voice quality. Various adjectives are commonly used, for example, breathy, pressed, rough, and harsh. Excessively aberrant qualities may indicate laryngeal pathology, but normal voice also ranges in perceived quality. Physiologically, the important aspect concerned with voice quality is glottal adduction, including variables such as glottal configuration and abruptness of vocal fold closure. Differences in these parameters are acoustically manifested in the source spectrum. For example, incomplete glottal closure during a closed phase of vibration introduces noise in the spectrum in addition to tonal energy, giving a percept of breathiness. Similarly, the abruptness of closure is associated with spectral tilt because it is directly related to the degree of asymmetry of glottal waveform (Hanson, 1997; Stevens, 1980). More unusual voice qualities may arise if vibratory patterns themselves become disturbed. Abnormal modes of vibration produce irregularity in mucosal wave, resulting in a rough voice.

The process of phonation is a complex interaction of airflow, subglottal pressure, and vocal fold tissue. Changes in these parameters affect various aspects of voice. F_0 and SPL may be adjusted independently, but often, a change in one causes a change in another. For example, raising subglottal pressure to increase SPL raises F_0. Multiple regulatory mechanisms are available for F_0 and SPL, indicating certain vibratory conditions under which it is more effective. For example, during soft or high-pitched phonation, oscillation is mainly contained in the cover of the vocal fold. In such cases, the vocal fold tension is more successfully controlled through its length rather than active increase in tension of the vocalis muscle. Similarly, the effect of glottal resistance on SPL is much more pronounced in lower frequencies than in higher frequencies. This is because that at high frequencies, the vocal folds are under considerable tension, required for high rates of vibration, and resistance cannot be further increased (Isshiki, 1964).

Modes of phonation and vocal registers. Control of F_0, SPL, and voice quality demonstrate how the ways in which the vocal folds move affects output sounds. The vocal folds, despite high degrees of freedom, tend to vibrate in certain modes (Titze, 2000). Normal modes of vibration in phonation are the sinusoidal motions of the vocal folds that produce regular patterns of F_0. Each of the normal modes is characterized by a distinct way that tissue oscillates along the horizontal or vertical axis. The most dominant modes of normal vocal fold vibration are ones where, during each cycle, one half-wavelength flexion appears either in the horizontal dimension or in both horizontal and vertical dimensions of the vocal fold (Titze, 2000).

Normal modes of vibration include different registers. A vocal register is a distinct vocal quality that can be maintained for a range of F_0s (Hollien & Michel, 1968). The number of registers differs from one to seven depending on the author and area of study, but it is commonly accepted that there are three registers for speaking: pulse, modal, and loft (Colton & Hollien, 1972; Hollien, 1974). Hollien and Michel (1968) suggested that there was little overlap of F_0 between adjacent registers but more recent studies have found evidence of hysteresis effects at register transitions (Švec, Schutte, & Miller, 1999; Tokuda, Zemke, Kob, & Herzel, 2010). The modal register is the standard type of voice for speech. Acoustically, the modal register is associated with F_0 ranges of 75-500 Hz for men and 130-750 Hz for women and with the greatest dynamic range of intensity compared to the other registers. Correspondingly, it is perceived as mid-pitched, loud, and unremarkable in quality (Hollien, 1974). The pulse register is associated with lower F_0 range, up to about70 Hz for both men and women, which corresponds to the cutoff frequency for perception of individual glottal pulses (Titze, 2000). The intensity range tends to be small and softer (Hollien & Michel, 1968). Slow rates of vocal fold vibration in pulse register are associated with shorted and thickened vocal folds (Hollien, 1960, 1962), a prolonged closed phase (Whitehead, Metz, & Whitehead, 1984), and decreased airflow and subglottal pressure (McGlone & Shipp, 1971a). The loft register is interchangeable with falsetto, and often characterized by a thin quality. Associated F_0 ranges are high, 150-750 Hz for men and 220-1700 Hz for women, and intensity range is greater than that is seen for pulse register (Hollien, 1974). At high rates of vibrations, vocal folds are generally long, thin, and tensed (Hirano, 1974).

In all three registers, the vocal folds vibrate regularly. However, as F_0 increases, perceived quality of voice changes between registers partly because of specific F_0 values, but also because other aspects of voice change as a result of changing biomechanical characteristics of vocal fold vibration. These differences are categorically perceived. With continuous changes in underlying physiologic or acoustic variables, quantal changes in voice quality occur only at register transitions. Titze (1988) refers to a pulse-to-modal transition a periodicity transition, where the perception of glottal pulse train shifts to the perception of uninterrupted signal. A modal-to-loft transition is referred to as a timber transition and a change in perception is attributed to changes in spectral slopes (Titze, 1988).

Irregularity is as important as regularity in vocal fold vibration. Perturbations and fluctuations are always expected in a vocal signal for there are certain levels of inherent noise present in all living systems. Irregularity is observed at multiple levels of the phonation process. At the neuro-physiological level, small fluctuations in vocal fold tension can introduce jitter and shimmer. Possible sources include variable firing rates of neurons that affect the steadiness of muscle contractions (Titze, 1991), and pulsation of blood vessels that periodically alter the shape of the vocal fold (Orlikoff & Baken, 1989), but a larger-scale variation in the neurological activity can also affect steadiness of F_0 and

SPL. At the structural and biomechanical level, various asymmetries between the right and left vocal folds, such as unbalanced geometry, mass, and tension, are major sources of irregularity in vibration (Berry, Herzel, Titze, & Story, 1996; Mergell & Herzel, 1997; Švec, Schutte, & Miller, 1996; Titze, 2006). At the level of acoustics, the interaction between the vocal tract and vocal folds becomes important. The vocal folds are physically coupled to the subglottal and supraglottal vocal tracts. Resonances occurring in these tubes feed energy back into the vocal folds and interfere with their vibration (Mergell & Herzel, 1997; Titze, 2006; Titze, Baken, & Herzel, 1993; Tokuda et al., 2010).

The net sum of these irregularities and fluctuations is manifested in three major forms of variability in the emitted signal: perturbation, modulation, and bifurcation. A perturbation is small, random cycle-to-cycle variability, for example, jitter and shimmer. During a normal mode of vibration, F_0 and SPL vary slightly between cycles. No two cycles are exactly the same, and each cycle deviates slightly from an expected value. When the expected value is thought of as a mean of vocal fold vibration over a period of time, then a standard deviation from the mean represents one measure of perturbation. Other measures of perturbation may take a linear trend into account in calculating deviations (i.e., no single reference value). The pattern of perturbations can also be periodic. F_0 or SPL of a voice can be systematically altered in a way so that it varies over a period of several cycles of vocal fold vibration. A modulation in voice may result from involuntary rhythmic muscle contractions as in vocal tremor, or can be voluntarily produced as in vibrato. Perturbations and modulations are examples of variability in voice. The variability may be random or non-random, but these types of variability do not change the overall vibratory behavior. F_0 should be relatively clear and voice quality should be normal.

Other types of irregularity can lead to bifurcations and chaos, causing qualitative changes in vibratory behavior. Phonation onset, the appearance of period doubling, biphonation, and register shifts are manifestations of bifurcations in vocal fold vibration (Berry, 2001; Herzel, Berry, Titze, & Saleh, 1994; Švec et al., 1999; Tokuda, Horáček, Švec, & Herzel, 2007, 2008). Complex patterns of vocal fold vibration can be effectively explained as different manifestations of a single oscillating system using principles of the theory of non-linear dynamics (Berry, 2001; Titze et al., 1993; Wilden, Herzel, Peters, & Tembrock, 1998). If a system is nonlinear, continuous changes in control parameters can cause an abrupt change in output, but at the same time, complex output vibratory patterns often reduces to a small set of fundamental modes of vibration. Since phonation is a process of interaction between airflow and vocal fold tissue, some of known control parameters include subglottal pressure, asymmetry between the right and left vocal folds, and vocal fold tension (Berry et al., 1996; Jiang, Zhang, & Stern, 2001; Mergell, Herzel, & Titze, 2000; Švec et al., 1996, 1999; Titze, 2000, 2006; Tokuda et al., 2007, 2008).

The number of modes of vibration is determined by the number of degree of freedom of the vocal folds. Even though continuous tissue of the vocal fold has infinite degrees of freedom, only a small set of modes is excited in most phonation types (Titze, 2000). That is to say, even complex patterns of vibration can be generated by different superpositions of these fundamental modes. Unlike a system like the vocal tract, where multiple resonances can be retained, the vocal folds produce single resonance because different modes become entrained because of coupling of the system via vocal fold collision and aerodynamic force. In baseline modal phonation, different modes are entrained at 1:1 to oscillate at the same frequency (Berry, 2001). However, changes in control parameters can shift entrainment patters and alter vibratory behaviors. Interesting phonatory events occur when there are competing modes in vocal fold vibration. For example, severe asymmetry between the two vocal folds reduces the degree of coupling between them, and lead to entrainment patterns of N:M or even disentrainment (Giovanni et al., 1999; Mergell & Herzel, 1997; Švec et al., 1996). An integer relationship between N and M (e.g., 1:2 or 1:3) captures a common type of non-modal phonation called subharmonics, where periodicity is achieved every second or third cycle of the original frequency. When the ratio of N and M is irrational, the there is no integer relationship between the original and additional frequencies, resulting in biphonation. Subharmonics and bifurcations are often antecedents to chaos. A progressive change in a control parameter may lead to chaotic vibration with further deterioration of periodicity (Berry et al., 1996; Titze, 2000).

Pitch breaks and non-modal phonations have been reported in infant cries (Michelsson, 1986; Michelsson & Wasz-Höckert, 1980; Sirviö & Michelsson, 1976), infant non-cry vocalizations (Buder et al., 2008; Kent & Murray, 1982; Robb & Saxman, 1988), excised larynx experiments (Berry et al., 1996; Jiang, Zhang, & Ford, 2003; Švec et al., 1999; Tokuda et al., 2008), as well as in disordered voice (Herzel et al., 1994), while these irregularities are typically avoided in adult voice (Fitch, Neubauer, & Herzel, 2002). Wide observation of these phonation types is of interest, as it seems to relate to the developing control of phonation.

Summary. The production and control of voice is a complex interplay between aerodynamic forces and biomechanical properties of the vocal folds. Three major dimensions of voice are F_0, SPL, and quality. Independent control of these parameters can be afforded to some extent by changing the driving pressure, glottal geometry, and vocal fold tension individually. Often times, however, a change in one parameter results in a change in another aspect of voice. A manner of vocal fold vibration is also important in how airflow and the tissue may interact. A non-linear dynamics framework parsimoniously explains various types of phonation as different manifestations of a single system. Various nonlinearities in the vocal fold vibration facilitate mode entrainment to result in modal phonation. But changes in parameters, such as vocal fold asymmetry, subglottal pressure, and vocal fold tension, can also cause sudden qualitative changes in vibratory patterns. Instances of subharmonics and bifurcations have been widely reported in infant vocalization, suggesting their relevance in phonatory development.

Acoustic Characteristics of Infant Phonation

Some aspects of anatomical characteristics and motor control are evidenced in acoustic properties of output sounds. Although we should be cautious about over-extrapolation, certain significant structural characteristics and developmental changes should be kept in mind when describing sound patterns. The following review will concentrate on acoustic characteristics of phonation in infancy. Since their advent, recording devices and sound analysis techniques have been applied in speech and voice acoustics. By inspecting spectrograms, earlier researchers have noted qualitative differences in infant vocalizations from mature speech sounds. The vocal development has been described as a series of enlargement, defining, and refining process. Very young infants around the age of 1 month tend to produce undifferentiated utterances with a simple pitch pattern. In the following several months, their utterances increase complexity in acoustic structures, reflecting experimentation with the vocal organs. In the last few months of the first year, their vocalizations become more restricted in frequency and temporal domains as infants begin to produce babbling, presumably suggesting greater control (Lynip, 1951; J. Murai, 1963).

The conceptual models of infant vocalization described in a preceding section are in agreement with these observations. Quantitative evidence, however, seems to be incomplete so far. Among the most frequently studied acoustic parameters of infant phonation include utterance duration, F_0, and harmonic structure of the vocal fold vibration. The number of acoustic studies in infant sounds is limited compared to those of speech sounds, especially those on emotional vocalizations are scarce in general but virtually nonexistent in infants (Bachorowski, Smoski, & Owren, 2001; Scheiner, Hammerschmidt, Jürgens, & Zwirner, 2002; Stark et al., 1975). Laughter does not occur till around 4 months (Sroufe & Waters, 1976), and comparable acoustic data are yet to be published. Perhaps the only exception is cry studies where alterations in cry characteristics have been noted in infants with various congenital abnormalities (Hirschberg, Szende, Koltai, & Illényi, 2009; Michelsson, 1971, 1986; Michelsson, Järvenpää, & Rinne, 1983; Michelsson, Sirviö, & Wasz-Höckert, 1977). While the relationship between acoustic features of cry sounds and neurological integrity is a fascinating topic, a review of the literature is out of scope of this paper. Most studies have focused on either cry or comfort sounds. Very few studies have examined both types of signals at the same time to contrast their acoustic characteristics. Previous acoustic studies of infant vocalization are summarized in Table 2. All measurements were taken using a narrow band spectrogram unless otherwise indicated.

Table 1

Summary of Acoustic Studies of Infant Vocalization

Study	N_1[a]	N_2[b]	Age	Signal types	Acoustic measures
Faiebanks (1942)	1	54[c]	1-9mo	Cry (hunger)	F_0 mean, range, SD
Ringel & Kluppel (1964)	10	30	4-40hr	Cry (pain)	F_0 mean; SPL mean; duration
Truby & Lind (1965)	30	NA	1-12d	Cry (pain)	RGM
Sheppard & Lane (1968)	2	108[c]	1-5mo	Cry (spontaneous) Comfort sounds	F_0 mean, variability; SPL; duration
Wasz-Höckert et al (1968)	192	212	1-30d	Cry (birth, hunger, pain)	F_0 mean, minimum, maximum; duration; RGM
	207	159	2-7mo	Cry (hunger, pain) Comfort sounds	
Lieberman et al (1971)	20	NA	1-4d	Cry (birth, pain, spontaneous)	RGM
Tenold et al (1974)	9	9	48hr	Cry (pain)	F_0 median
Prescott (1975)	4	150	1-10d; 6-9mo	Cry (spontaneous)	F_0 mean and variability; duration
	10	30	1-2mo		
	10	30	6-8mo		
Laufer & Horii (1977)	4	787	1-6mo	Comfort sounds	F_0 mean, minimum, maximum
Murry et al (1977)	8	24	3-6mo	Cry (hunger, pain, startle)	F_0 mean
Keating & Buhr (1978)	6	300	8-40mo	Comfort sounds	F_0 range, RGM
Lester (1978)	12	NA	2d	Cry (pain)	F_0 mean
Zeskind & Lester (1978)	24	24[c]	2d	Cry (pain)	F_0 mean; duration
Keating (1980)	4	500	4-16mo	Comfort sounds	F_0 range, RGM
Gardosik et al (1980)	103	103	Newborns	Cry (birth)	F_0 mean
Thodén & Koivisto (1980)	38	401	1d, 5d, 3mo, 6mo	Cry (pain)	F_0 minimum, maximum; duration; RGM

Table 2

(*Continued*)

Author	N_1[a]	N_2[b]	Age	Signal types	Acoustic measures
Kent & Murray (1982)	21	NA	3, 6, 9mo	Comfort sounds	F_0 mean, duration, RGM
Michelsson et al (1983)	27	54	1-5d	Cry (pain)	F_0 minimum, maximum; duration; RGM
Robb & Saxman (1985)	14	980+	11-25mo	Comfort sounds	F_0 mean; duration
Robb & Saxman (1988)	14	1,200	11-25mo	Comfort sounds	RGM
Robb et al (1989)	7	4,428	8-14mo	Comfort sounds	F_0 mean, variability
Wermke & Mende (1993)	50	NA	1-12mo	Cry (spontaneous) Comfort sounds	F_0 variability
Mowrer (1994)	1	24	4mo	Laugh	Duration, RGM
Gilbert & Robb (1996)	4	48	1-12mo	Cry (hunger)	F_0 median
Lind & Wermke (2002)	1	658	1-3mo	Cry (spontaneous)	F_0 mean
Michelsson et al (2002)	172	1,836	1-7d	Cry (pain)	F_0 mean, minimum, maximum; duration
Rothgänger (2003)	15	583	0-12mo	Cry (hunger, pain, spontaneous)	F_0 mean; duration
	15	201		Comfort sounds	
Nathani Iyer & Oller (2008)	8	2168	3-12mo	Comfort sounds	F_0 mean, variability

Note. Ages in weeks in some studies were converted to ages in months by dividing the original number by 4.25

[a]Number of infants studied. [b]Number of utterances analyzed. [c]Number of samples of particular duration analyzed. NA indicates non-availability.

Utterance duration and sound pressure level. Utterance duration and intensity are some of the basic attributes of vocalization. Utterance duration is generally defined as the total time of vocalization within a single expiratory cycle, but may exclude a very short isolated segment either at the onset or offset. SPL is a physical measure of loudness and most measured in decibels. When manipulated appropriately, these measures can provide information about underlying respiratory drive. For example, an increase in mean length of utterance is correlated with an increase in vital capacity, indicating the effect of respiratory capacity on utterance duration (Lalonde & Boucher, 2009). However, SPL is rarely measured in infant vocalization studies and the measure of utterance duration has limited utility in the study of infant vocalization because the length changes arbitrarily from utterance to utterance for reasons unrelated to the respiratory capacity. The average value of an utterance can vary greatly depending on signal types, making comparisons across studies difficult.

Information about utterance durations is found in Table 3. The average duration varies from 1.1 to 10.1 seconds, but there are two fairly clear findings. First, utterance duration has a high dependence on signal types. Cry sounds tend to be much longer than comfort sounds, with pain cries consistently being the longest (Wasz-Höckert, Lind, Vurenkoski, Partanen, & Valanne, 1968). When mean durations are measured in comfort sounds, the distribution tends to be highly skewed, with the great majority of utterances being shorter than 400 ms, which roughly equals the duration of a typical syllable (Kent & Murray, 1982). Second, when examining vocalizations in cry bouts, the first phonation is often the longest in duration (Sirviö & Michelsson, 1976; Thodén & Koivisto, 1980). No explanation has been offered for this tendency. One possibility is that the child does not take a full inhalation during a cry bout (i.e., catch breath), but the reviewed studies did not include physiological measures. A similar effect is certainly possible in comfort sound productions, especially in later infancy when the child is producing chains of syllables over multiple breath cycles, however no comparable data are available so far.

Table 3

Summary of Durations of Utterances Reported in Previous Studies

Study	Signal type		Age	Mean	SD	Range
Ringel & Kluppel (1964)	Cry (pain)		4-40hr	1.5	0.6	0.6 – 4.0
Wasz-Höckert et al (1968)	Cry (birth)			1.1	0.4	-
	Cry (pain)		1mo	2.6	1.5	-
	Cry (hunger)			1.3	0.6	-
	Cry (pain)			2.7	1.1	-
	Cry (hunger)		2-7mo	1.2	0.6	-
	Comfort sounds			1.1	0.5	-
Michelsson et al (1971)[a]	Cry (pain)		1-10d	2.0	-	1.6 – 3.6
Prescott (1975)	Cry (spontaneous)		1-10d	1.2	0.6	-
			4-6wk	10.1	0.7	-
			6-8mo	1.2	0.4	-
			6-9mo	2.0	1.0	-
Lester (1978)	Cry (pain)		2d	4.9	2.8	-
Thodén & Koivisto (1980)	Cry (pain)	First in bout	1d	5.0	2.8	0.5 – 12.2
			5d	5.2	2.3	0.9 – 10.4
			3mo	4.2	2.5	0.3 – 9.9
			6mo	4.1	2.2	1.5 – 10.4
		Second in bout	1d	1.8	0.9	0.6 – 4.0
			5d	1.6	0.7	0.5 – 3.6
			3mo	1.7	0.8	0.5 – 3.6
			6mo	2.0	1.1	0.4 – 5.1
		Third in bout	1d	1.2	0.8	0.3 – 4.3
			5d	1.3	0.8	0.5 – 3.7
			3mo	1.6	1.0	0.5 – 4.4
			6mo	1.5	0.8	0.7 – 3.6
Michelsson et al (1983)	Cry (pain)		1-5d	4.3	1.9	-
Michelsson et al (2002)	Cry (pain)		1-7d	1.4	0.6	0.7 – 3.9

[a] Median and Q_1–Q_3 were reported instead of mean and range.

There seems to be no systematic mean difference due to individual (Ringel & Kluppel, 1964) or gender (Michelsson, Eklund, Leppänen, & Lyytinen, 2002) differences. A possible effect of age on utterance duration is inconclusive so far due to conflicting evidence (Rothgänger, 2003; Thodén & Koivisto, 1980). The duration variation between utterances has a tendency to decrease with increasing age (Sheppard & Lane, 1968), however, the utterance variability remains very high throughout infancy (Prescott, 1975).

Data on SPL are extremely limited in published studies of infant vocalization probably because of difficulties in calibrating a microphone and ensuring a constant mouth to microphone distance during recording. Overall intensity levels of infant cry sounds tend to be high, reaching over 80 dB SPL (re: 20 μPa) at 10 inches from the mouth, which is about 20 dB greater than ordinary conversation at 3 feet (Ostwald, 1963). A mean SPL, measured at 12 inched from the mouth, reported in Ringel and Kluppel (1964) had a similar value of 82 dB, raging from 65 to 94 dB. They also found that a significant difference between neonates in mean SPL, indicating possible individual differences in overall SPL of cries. Sheppard and Lane (1968) examined relative amplitude of cry and non-cry utterances. They found a high positive correlation between the mean utterance amplitude and duration, indicating the tendency for longer utterances to have greater amplitude. No developmental trend or other relationship especially with F_0 has been reported.

Fundamental frequency. By far, F_0 is the most frequently used parameter of infant vocalization. F_0 and perceived pitch are thought of as an important feature of voice that reflects neural and structural maturation of the phonatory apparatus (Bosma, 1975; Lind & Wermke, 2002) as well as information about approximate age, identity, and physical condition of a child (Prescott, 1975). Various measures of F_0, therefore, are useful in evaluating the phonatory performance of infants as a function of time. Measures from previous studies include mean, maximum, minimum, range, and standard deviation (SD). F_0-related statistics are available in both cry and non-cry sounds. Numerical data from previous studies are summarized in Tables 4 and 5. Mean values of F_0 from previous studies are also plotted in Figure 2.

Mean F_0. Average values of F_0 are most frequently reported in newborn and neonatal cries. Ringel and Kluppel (Ringel & Kluppel, 1964) studied pain-induced cries produced by ten newborns of age 4-40 hours. Three cry utterances were obtained for each newborn, and a mean F_0 was found to be 423 Hz for the group. In their normative study, Wasz-Höckert et al. (1968) examined a dominant pitch for birth, pain, and hunger cry utterances obtained from192 neonates of 30 days and younger. Following the same method as in Ringel and Kluppel (1964), they reported a mean general F_0 ranging from 500, 530, and 470 Hz for birth, pain, and hunger cries, respectively. In another study of pain cries, nine newborns of age approximately 48 hours were examined (Tenold et al., 1974). By means of cepstral analysis, they reported a median F_0 of 518 Hz. A few more studies provide data on mean F_0 in newborn pain cries. A mean F_0 was 467 Hz for a group of twelve newborns (Lester, 1978); and 468 Hz for a group of 24 (Zeskind & Lester, 1978).

Table 1

Summary of F_0 Means Reported in Previous Studies

	Age in months											
Study	1	2	3	4	5	6	7	8	9	10	11	12
Fairbanks (1942)	373	415	485	585	814	607	600	708	640	-	-	-
Ringel & Kluppel (1964)	413	-	-	-	-	-	-	-	-	-	-	-
Sheppard & Lane (1968)	407	418	450	434	436	-	-	-	-	-	-	-
Tenold et al (1974)	518	-	-	-	-	-	-	-	-	-	-	-
Prescott (1975-1)	384	-	-	-	-		415		-	-	-	-
Laufer & Horii (1977)	317	338	338	339	337	342	-	-	-	-	-	-
Murry et al (1977)	-	-		434					-	-	-	-
Lester (1978)	467	-	-	-	-	-	-	-	-	-	-	-
Zeskind & Lester (1978)	468	-	-	-	-	-	-	-	-	-	-	-
Gardosik et al (1980)	462	-	-	-	-	-	-	-	-	-	-	-
Kent & Murray (1982)	-	-	445	-	-	450	-	-	415	-	-	-
Robb & Saxman (1985)	-	-	-	-	-	-	-	-	-	-	366	435
Robb et al (1989)	-	-	-	-	-	-	-	430	374	396	400	435
Gilbert & Robb (1996)	429	448	447	451	472	413	466	447	440	468	477	527
Lind & Wermke (2002)	417	410	410	-	-	-	-	-	-	-	-	-
Michelsson et al (2002)	496	-	-	-	-	-	-	-	-	-	-	-
Rothgänger (2003, cry)	452	-	476	-	-	483	-	-	479	-	-	503
Rothgänger (2003, noncry)	389	-	358	-	-	343	-	-	311	-	-	337
Nathani Iyer & Oller (2008)				339			344			359		

Table 5

Summary of F_0 Ranges Reported in Previous Studies

Age (mo)	Fairbanks (1942)	Keating & Buhr (1978)	Thodén & Koivisto (1980)	Michelsson et al (2002)
1	153-888	-	130-3420	262-1565
2	63-947	-	-	-
3	89-2120	-	170-2500	-
4	214-1495	-	-	-
5	229-2387	-	-	-
6	206-1487	-	150-2750	-
7	150-2348	40-850	-	-
8	134-2329	-	-	-
9	207-2631	-	-	-
10	-	-	-	-
11	-	-	-	-
12	-	-	-	-

Other studies report additional data for a large sample of newborn birth cries. A mean F_0 of 467 Hz was found for 103 newborns (Gardosik, Ross, & Singh, 1980); and 496 Hz for 172 newborns (Michelsson et al., 2002). Mean F_0 data on older infants are available from Murry, Amundson, and Hollien (Murry, Amundson, & Hollien, 1977). They studied pain, hunger and startle cries produced by eight 3-6 month old infants. By using visicorder tracings, they reported a mean F_0 of 434 Hz.

The number of studies that looked at cry sounds beyond the neonatal period is smaller. As seen in Table 4 and Figure 2, three patterns of developmental changes have been proposed: steady increase, increase then plateau, and no change. In a study of hunger cries collected monthly from 4 infants during their first year of life, the median F_0 values for the group were correlated with increasing age over the 12-month period, showing a steady increase from 429 Hz at one month to 527 Hz at 12 month (Gilbert & Robb, 1996). A similar tendency was reported in another longitudinal study involving 15 infants in their first year. Rothgänger (2003) observed an increase in F_0 for discomfort cries from 452 Hz at one month to 503 Hz at 12 months, which was significantly positively correlated with chronological age.

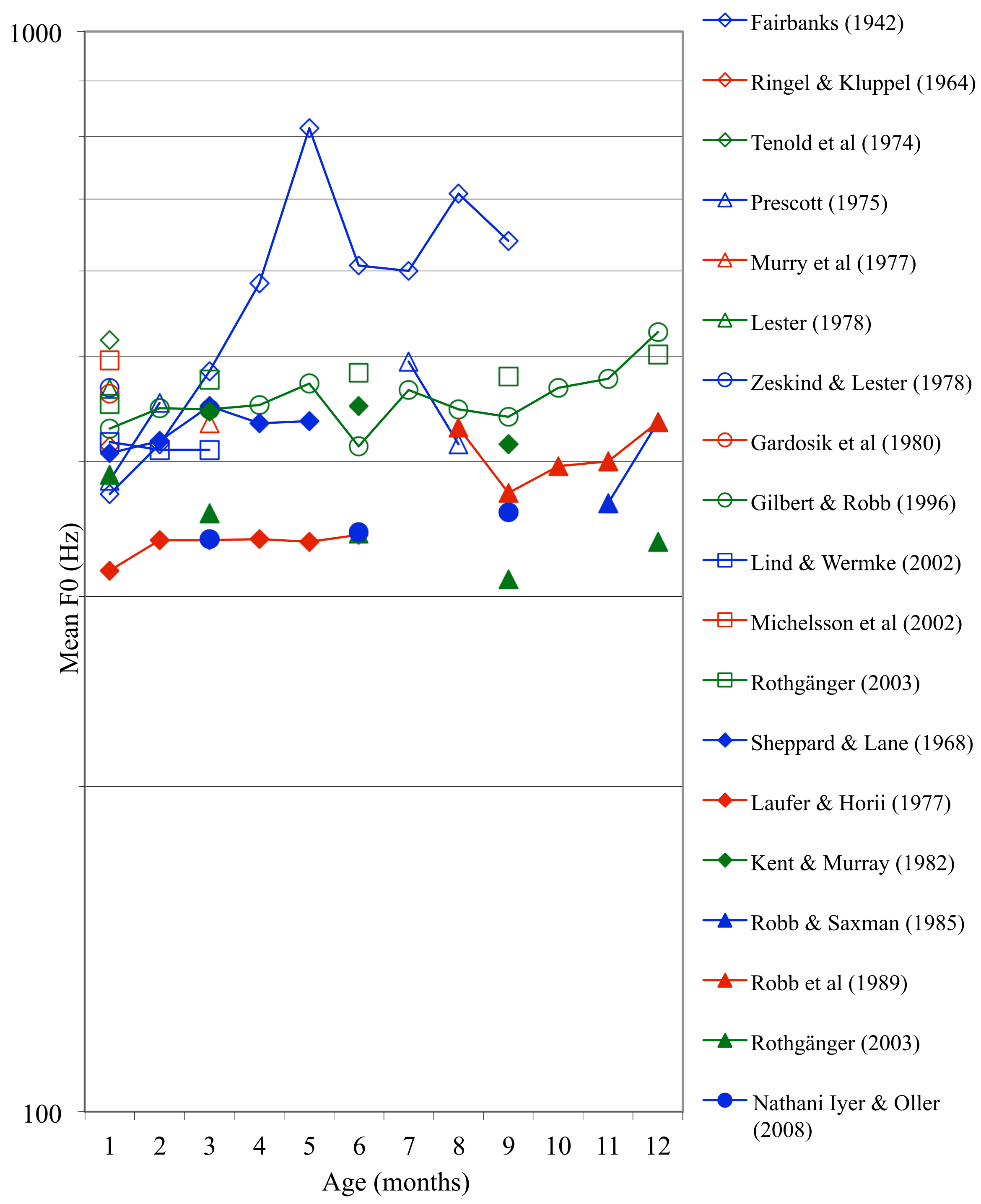


Figure 2. Mean F_0 values of cries (open symbols) and comfort vocalizations (filled symbols) over the first year of life.

Fairbanks (1942) studied hunger cries of his infant son during 1-9 months of age. He obtained monthly recordings from which F_0 information was extracted by means of phonophotography. Mean F_0s varied from 373 Hz to 814 Hz. An increasing tendency was observed from 373 Hz at 1 month to 814 Hz at 5 months. The mean F0 drops to plateau at a lower value thereafter. Prescott (1975) studied spontaneous cries from three groups of children. The first group had four participants who were recorded first when they were 1- to 10-day newborns and again when they were 6- to 9- month infants. She also provided additional two groups for comparison: one with ten neonates of 4-6 weeks and another with infants of 6-8 months. Mean F_0s for the groups were, from younger to older, 384 Hz, 453 Hz, 495 Hz, and 415 Hz, respectively. The trend is less pronounced but is consistent with the findings in Fairbanks (1942). It is possible that the second trend, increase with plateau, represents the same type of developmental change as the first pattern, steady increase, with a difference in degree to which F_0 means depart from the general trend in middle months.

On the other hand, some authors have reported negligible changes in mean F_0. In acoustic and spectrographic attributes of pain and hunger cries studied in Wasz-Höckert's group (1968), little changes were observed after the newborn period. They noted changes in some acoustic attributes (i.e., melody type, the occurrence of glottal plosives, F_0 shift, and vocal fry) during the first week of life and no changes throughout the remaining 7 months that they collected the cries. They collapsed data across 2-7 months, and reported a general F_0 of 530 Hz for pain cries and 500 Hz for hunger cries. Also, Lind and Wermke (2002) studied the development of F_0 in daily recordings of cries in one infant during 1-3 months of age. They did not find any change in mean F_0 during this time.

Instead, interestingly they observed an interaction between mean F_0 and utterance duration. That is, when they divided utterances into short (< 0.8 sec) and long (> 0.8 sec) groups, shorter utterances had a mild positive correlation with age, while longer utterances did not show any trend.

Comparable data on comfort sounds are available from a few studies. The first observation of non-crying vocalization varies from study to study. This is partly because of a difference in sampling intervals among studies, but may also be because of a difference in what is counted as a valid vocalization at a very young age. Lynip (1951) was the first to use spectrogram in the study of infant non-crying vocalizations. He followed one infant from birth to 56 weeks. After observing a first non-crying utterance at age of 6 weeks, he found F_0 of the child to be 360-420 Hz between eighth and tenth weeks (Lynip, 1951). Sheppard and Lane (1968) recorded all vocalizations from two infants during the first 141 days of life. A mean F_0 was calculated to be 407 Hz in the first month, steadily rising to stabilize at around 440 Hz after the third month. Laufer and Horii (1977) studied non-distress vocalizations collected from four infants during the first 6 months of life. Monthly means of F_0 ranged little through the duration of the study, ranging from 317-342 Hz. Kent and Murray (1982) studied comfort-state utterances of 21 infants at 3, 6, and 9 months of age, and reported mean F_0 values of 445, 450, 415 Hz, respectively. Rothgänger (2003) observed a decrease in F_0, a trend opposite to that found for discomfort cries, from 389 Hz at one month to 337 Hz at 12 months. Nathani Iyer and Oller (2008) studied 8 infants at three different stages of vocal development. A mean F_0 for the 8 infants was 339, 344 and 359 Hz for 3 months pre-canonical, canonical, and 3 months post-canonical stages, respectively, and no significant change in mean F_0 was found with vocal development.

Studies of F_0 in early childhood also provide a relevant comparison. A mean F_0 was 357 Hz in babbling and word productions among children of age 11-25 months (Robb & Saxman, 1985), 396 Hz in monosyllable and 399 Hz in disyllable productions among children of age 8-26 months (Robb, Saxman, & Grant, 1989), and 298 Hz in sentence productions in children of age 3-6 years was (Eguchi & Hirsh, 1969). When a comparison is made across a wide range of ages, the global change observed is a decrease in habitual F_0. With all being equal, such change is assumed to reflect an increase in the overall size of the laryngeal structures. Extrapolating such estimate into infancy is problematic because infants do not have a habitual F_0 and conditions of phonation are likely to be not uniform across children and vocalizations. Therefore, the use of mean F_0 to track maturation may be inappropriate in the population in which speech is not established. Often times, especially in cries, an opposite trend was reported, to which greater ability to exert more tension on the vocal folds was attributed (Fairbanks, 1942). A general pattern of mean F_0 development is so far unclear, and a relationship between cries and non-cry vocalizations is inconclusive. Although there seems to be a tendency of increasing F_0 for cries and decreasing F_0 for comfort vocalizations, definitive conclusions cannot be made until other factors influencing F_0, such as SPL, are accounted for.

F_0 variability: range and deviation. A measure of central tendency alone provides an incomplete picture of infants' phonatory abilities. One of the reasons why a relationship between average F_0 and chronological age does not reach statistical significance is because of a wide range of mean F_0 values found in infant vocalization.

The following studies have demonstrated how much F_0, and not F_0 means, fluctuates in infant vocalization. Average values of the lowest F_0 in cry utterances varied from 330 to 450 Hz, and the highest F_0 from 583 to 1080 Hz (Michelsson, 1971; Michelsson et al., 2002; Thodén & Koivisto, 1980; Wasz-Höckert et al., 1968). These are mean values of minimum and maximum F_0s, so actual F_0 ranges were larger, for example, the entire range of F_0 observed in Michelsson (2002) was 262-1565 Hz. As represented in Table 5, similar ranges were reported in Fairbanks (1942) and Thodén and Koivisto (1980). The end values reported for 1-, 3-, and 6-month-olds in Thodén and Koivisto (1980) showed little change in association with increasing age. Similar observations have been made in non-cry vocalizations. Keating and Buhr (1978) studied F_0 in babbling, words, or sentences produced by six infants ranging in age from 7 to 39 months. They too found a wide range of F_0 values, 30-2500 Hz, across infants of different ages, with the smallest range being 60-750 Hz, and the largest 250-2500 Hz. They also identified, through spectrographic and perceptual evaluation, three registers of fry, modal, and high. In general, the fry register was associated with 30-250 Hz, modal, 150-700 Hz and high 380-2500 Hz, with no apparent developmental change in the age range studied.

Three important points emerge from these studies. First, continuous distributions of F_0 values reject the possibility that a wide F_0 range is due to outliers. Second, evident bimodal distributions in 5-8 months (Fairbanks, 1942) support regular occurrences of extreme F_0 values at least in infant cry utterances. Third, findings in Keating and Buhr (1978) confirm that both high and low F_0 values are common part of vocal output throughout infancy and early childhood.

Van Oordt and Drost (1963) presented two competing hypotheses about the development of F_0 range: gradual expansion of F_0 range from birth into childhood, and establishment of at least 2-octave range in the first year of life. They examined physiological and musical ranges of F_0 in 126 children of 0-16 years of age. They found that the physiological range of F_0 was 2.5-3 octaves regardless of age. This range is similar to the range of modal phonation for adult males (71-561 Hz) and adult females (122-798 Hz), which are both about three octaves (Hollien & Michel, 1968). Furthermore, within this rather stable range, a musical F_0 range showed an increase with age and training, indicating greater control of voice in older age. Unfortunately, a musical range for children under the age of six is not available for the obvious reason that small children cannot easily follow instructions. However, F_0 ranges for different vocal types, such as comfort sounds, may serve as a reasonable alternative so that the interpretation can be extended to early childhood and infancy.

F_0 values, unlike duration, change continuously within an utterance as well as between utterances. As a result, two kinds of deviation can be derived: within-utterance and between-utterance variability. Little change in between-utterance variability with age has been reported in Sheppard and Lane (1964), Prescott (1975), Laufer and Horii (1977), Robb et al. (1989), and Nathani Iyer and Oller (2008), indicating that utterances have similar amount of fluctuation in F_0 during infancy. On the other hand, within-utterance variability was reported to be constant in Sheppard and Lane (1964) and Nathani Iyer and Oller (2008) but increase in Prescott (1975) and Laufer and Horii (1977). Several methodological and technological differences are thought to result in conflicting findings on developmental changes in F_0 variability. Short duration of observation and small sample sizes in both the number of children and utterances are obvious sources of error, but computation of F_0 variability and types of vocalizations also seem to be influencing factors.

Harmonic structure and vibratory patterns. The use of F_0-related measures assumes periodicity in the signal. One problem is that irregular F_0 and F_0 shifts are normal part of the vocal repertoire in infants (Buder et al., 2008; Keating & Buhr, 1978; Robb & Saxman, 1988; Truby & Lind, 1965). Various patterns of phonation have been explained by the nonlinear dynamical nature of vocal fold vibration (Mergell et al., 2000; Švec et al., 1996, 1999; Titze et al., 1993). Buder et al. (2008) provided acoustic and perceptual criteria in identifying various phonation types in infant vocalizations, which are modal, loft, pulse, subharmonics, biohonation, chaos, and glottal stops.

Truby and Lind (1965) were among the first to identify three types of phonation in pain cries of infants: phonation, dysphonation, and hyperphonation. Phonation is the basic performance of the vocal folds where vibrations are regular. Dysphonation refers to the turbulent noise produced by irregular movements of the vocal folds. Hyperphonation is the phonation at very high F_0 levels, and often called shifts among cry researchers (Truby & Lind, 1965). Occurrences of other, more complex source variations have also been reported in both cry and non-cry vocalizations. Michelsson and her associates have described occurrences of F_0 shift, vibrato, biphonation, subharmonics, and furcation in cries of healthy and sick neonates (Michelsson, 1986; Michelsson & Wasz-Höckert, 1980). Kent and Murray (1982) also reported occurrences of source variations such as F_0 shift, aperiodic noise, vocal tremor, subharmonics, and biphonation, in infants of various ages. Robb and Saxman (1988) studied non-cry vocalizations of children between the ages of 11 and 25 months. They verified the presence of subharmonics, F_0 shift, and biphonation. They further measured mean, range, and standard deviation of F_0 for these segments, which is valuable because most studies exclude these segments for having ambiguous F_0.

Non-modal phonations are often linked to laryngeal pathology and only infrequently observed in normal adult speech (Kent & Murray, 1982). However, they occur regularly in infants and small children, presumably reflecting their unique laryngeal functioning. Lieberman and his colleagues (1971), who studied various types of cries in newborns, observed a sudden shift in phonation from modal vibration to aperiodic noise. They reasoned that a sudden alteration in the laryngeal excitation is likely due to the condition in which the infant fails to increase medial compression in the presence of increasing subglottal pressure. Therefore, occurrences of similar phenomena can be thought of as a lack of control in the maturing vocal system. They additionally speculated the possibility that properties of the infantile vocal folds might not be suited to withstand high subglottal pressure. This is in line with another speculation in Kent and Murray (1982), in which they considered greater susceptibility of infant vocal folds to nonlinearities that speakers gradually control to avoid with learning. In addition, vocal fold stiffness and medial compression are also believed to influence vocal fold vibration. A high F_0 often associated with F_0 shift may be interpreted as indirect evidence of elevated stiffness or medial compression (Robb & Saxman, 1988).

Mechanisms of various patterns of vocal fold vibration have been well explained in studies using mathematical models (Jiang et al., 2001; Mergell et al., 2000; Titze et al., 1993) and experiments with excised larynges (Berry et al., 1996; Jiang et al., 2003;

Tokuda et al., 2008), and probable causes of nonlinear phenomena in vocal fold vibration have been formulated based on these observations (Titze, 2006). Abrupt changes in harmonic structure of voice reflect changes in the movement pattern of the vocal folds, but explaining why and how they occur in infant phonation must be attempted with caution due to a vast difference in age, physical characteristics, and vocal tasks involved in different studies.

Summary. Acoustic studies of infant phonation mostly examined F_0-related measures and harmonic organization. Measures of duration tend to vary throughout infancy without showing any developmental trend. Information on SPL is extremely limited, and covariation of SPL is not accounted for when evaluating F_0. A mean F_0 seems to lie between 400 and 600 Hz, but F_0 varies widely. No clear changes in F_0-related measures have emerged as a function of age. Findings are often conflicting and general statements about acoustic characteristics of infant phonation are difficult to make. Part of the problem is due to unrepresentative sampling of utterances, infants, or both in each study. Most studies on infant comfort sound productions have a very small sample size, making findings easily biased by idiosyncrasies of individuals. Most cry studies have adequate sample sizes and well-controlled environment of data collection, but are limited in an age range of observation. But more importantly, the difficulty is also due to the complex nature of vocal production in infancy. The presence of non-modal phonations in infant vocalization immediately questions the traditional way of studying vocal development mainly through F_0-related measures. The exclusive analysis of modal phonation may overlook important information about developing phonatory abilities. So far, little is know about when modal and non-modal phonations occur. Describing their distributions along the ranges of F_0 and SPL can be a helpful first step in understanding how infants use and manage their voice.

Developmental Anatomy and Physiology of Infant Phonation

Since the moment of birth, infants are capable of producing voice on expiratory airflow as evidenced by birth cry. Types of utterances produced change from more involuntary to more voluntary as the infant grows, and motor skills underlying these different utterance types are fundamentally related to the maturation of both vocal organs and neural control (Kent, 1976, 1992). The main organs of speech are divided into three subsystems: respiratory system, hyo-laryngeal complex, and upper aero-digestive tract, corresponding to three basic processes of speech: airstream, phonation, and resonance processes. Smooth execution of mature speech requires coordination of all three processes. When they speak, human adults draw a quick deep breath in, and then exhale at a rate appropriate for length and intensity of a planned utterance while making glottal adjustments to regulate voicing and configuring the vocal tract for target speech sounds. Before a series of such highly coordinated actions can be established through orderly relationships among component structures, not only a considerable amount of learning must be achieved but some physical and neuromuscular development must also be attained. Various anatomical and physiological observations are reviewed below that may be in part contribute to early vocal patterns.

The respiratory system. Breathing has two main functions in humans besides defensive reflexes for airway protection. The primary function is gas exchange for life support and the secondary function is airstream regulation in speech. Each of the two respiratory tasks is characterized by a distinct pattern of respiratory movements that capitalizes on different aspects of the breathing mechanism. During tidal breathing, muscles of inspiration contract to increase the thoracic volume and lower the lung pressure. This results in ingress of air into the lungs because air moves from the area of higher pressure to the area of lower pressure. The opposite occurs in exhalation; the elasticity of the lung tissue and ribcage decreases the lung volume to increase the lung pressure, resulting in egress of air. A typical volume involved in a tidal breathing cycle is relatively small. Inspiratory and expiratory phases alternate in an approximately symmetric manner. A typical value of inspiratory duration is about 1.3 s and expiratory duration 2.0 s, which makes the ratio of inspiratory duration to expiratory duration (IE ratio) about 0.4 and the respiration rate about 18 breaths per minute (BPM) (Bunn & Mead, 1971; Langlois, Baken, & Wilder, 1980).

In speech, breathing is also a source of aerodynamic energy. Especially the lungs provide airstream to the larynx that is needed to generate voice. Among other mechanisms of sound source generation in natural languages, quasi-periodic oscillation of the vocal folds almost always employs pulmonic egressive airflow (Ladefoged, 2006). However, tidal breathing is not suitable for sustaining speech in two main aspects. First, the amount of air taken in during an inspiratory phase is usually too small to support utterances of sufficient length or intensity. Second, symmetry of inspiratory and expiratory phases makes utterances to break too often for too long. Therefore, a typical breathing pattern becomes modified in at least three main aspects: inspiratory volume, inspiratory and expiratory durations, and use of active forces during expiration. When breathing for speech, inspiration is deeper. An increase in the volume of air intake provides a greater amount of raw material available for speech. Inspiration is also quicker, whereas exhalation is drawn out. In addition, active muscle forces complement passive forces during expiration in order to achieve a rate of airflow intended for a given length and intensity of an utterance. A typical value of inspiratory duration decreases to less than 0.5 s and expiratory duration increases to over 3.5 s, which makes the IE ratio about 0.1 and the respiration rate about 14 BPM (Bunn & Mead, 1971; Langlois et al., 1980). Although these parameters vary according to specific vocal and linguistic demands, inspiratory duration is found to be a reliable marker distinguishing speech from tidal breathing (McFarland, 2001). By studying ribcage movements during quiet breathing and various speech tasks in adults, McFarland (2001) found a significantly shorter mean inspiratory duration speech tasks.

Newborns are clearly capable of adequate breathing for life support since the moment of birth, however, the respiratory system continues to develop post-natally. Acquisition of respiratory control for speech occurs concurrently with growth of the breathing apparatus. The respiratory system continues to develop after birth in terms of physical size, biomechanical changes, and maturation of control (Hoit, Hixon, Watson, & Morgan, 1990; Reilly & Moore, 2009). Obviously, the musculoskeletal framework of the thorax increases in size as the infant grows. But in addition, mechanical characteristics of the thorax undergo changes as well. Young infants have more horizontal orientation of ribs, flatter diaphragm, and a smaller zone of apposition, reducing inspiratory action of infant's chest wall by making the chest wall more compliant (Reilly & Moore, 2009). Although specifics remain largely unknown, the distinct shape, control, and biomechanical characteristics of chest wall suggest the infant regulates his breathing apparatus differently than do adults. Hoit and colleagues (Hoit et al., 1990) studied changes in anteroposterior diameter of the chest wall during speech tasks in 20 children from four age groups of 7, 10, 13, and 16 years, then compared to adult data. They found that the youngest group had initiations and terminations of breath groups at significantly larger lung, ribcage, and abdominal volumes, and larger lung volume excursions per breath group compared to the older groups, which lead them to conclude that basic patterns of speech breathing is adult-like by the age of 10 (Hoit et al., 1990).

Table 6

Temporal Characteristics of Infant Respiration

	Tidal	Cry				Vocalization	
	Langlois et al (1980)	Wilder & Baken (1978)		Langlois & Baken (1976)		Langlois et al (1980)	
Age (mo)	BPM	BPM	IE ratio	BPM	IE ratio	Cycle duration (BPM)	IE ratio
1	87	50	0.19	-	-	-	-
2	84	40	0.18	-	-	1.45 (41)	0.21
3	69	37	0.16	-	-	1.29 (46)	0.37
4	82	35	0.15	-	-	1.38 (43)	0.27
5	74	34	0.15	-	-	1.67 (35)	0.27
6	61	31	0.14	-	-	1.60 (37)	0.41
7	62	29	0.12	30	0.16	1.48 (40)	0.38
8	58	23	0.11	25	0.16	1.55 (38)	0.38
9	43	-	-	26	0.19	1.45 (41)	0.42
10	48	-	-	27	0.15	1.62 (37)	0.38
11	37	-	-	23	0.13	1.61 (37)	0.34
12	42	-	-	23	0.14	1.75 (34)	0.36
13	43	-	-	19	0.15	1.74 (34)	0.39

One aspect of breathing that is of interest is its temporal characteristics. A few studies have investigated respiratory activity in infancy. Temporal characteristics of breathing patterns from these studies along with respiratory rates during tidal breathing are summarized in Table 6.

Mean respiratory rates decrease with increasing age during quiet breathing. In general, respiratory rates are higher and tidal volumes smaller than those of adults, the breathing pattern is regular and the proportion of tidal volume to total lung capacity is similar to that of adult's (Langlois et al., 1980). During cry, mean respiratory rates are much slower compared to tidal breathing, and they too decrease with increasing age. In fact, the respiratory rate during cry at 12 months is comparable to the respiratory rate typically found in adult speech. It is important to note that although BPM decrease with age in cry, IE ratio tends to be stable. That is, the relationship of inspiratory and expiratory phases in cry is similar to speech breathing and is present in the first year. Although Wilder and Baken (1978) interpreted IE ratios to have a decreasing tendency, it is uncertain if the trend exists due to paucity of data and sample, especially, when their results are interpreted in light of results from Langlois and Baken (1976). The BPM and IE ratio underlying non-cry, non-reflexive vocalization, on the other hand, occupy ranges between tidal and cry breathing patterns. Mean duration of respiratory cycles during vocalization tends to be shorter than cry but longer than quiet breathing, which in terms of BPM is on the order of 34 to 46. Similarly, the IE ratios for vocalization occupies a range between cry and quiet breathing, and there is no apparent developmental trend in the first year. Therefore, speech breathing, a pattern maximized for long phonation, is not yet employed in vocalization in the first year.

Apart from temporal aspects of breathing, patterns of breathing movements are another important aspect to consider. During quiet breathing, high compliance of the ribcage and their supine postures make young infants obligate belly breathers (Hixon, Goldman, & Mead, 1973; Reilly & Moore, 2009), which is in contrast to adults where the ribcage and abdomen move in synchrony (Hixon, Weismer, & Hoit, 2008). As the infant grows, chest wall compliance decreases, resulting in more participation of the ribcage to tidal breathing (Hershenson, Colin, Wohl, & Stark, 1990). In speech breathing among adults, they tend to use the midrange of their vital capacity (Hixon et al., 1973). Both older children and adults tend to accomplish changes in lung volume predominantly by the ribcage during utterances (Hoit et al., 1990). On the other hand, infants in the first year of life initiate and terminate vocalization at variable levels of their vital capacity, use a wide range of lung volume excursions for utterances, and use a variety of ribcage and abdomen contributions to achieve lung volume excursions (Boliek, Hixon, Watson, & Morgan, 1996). In a longitudinal study where infants of age 9 to 48 months were observed, researchers have found distinct patterns of development between tidal and vocalization breathing (Connaghan, Moore, & Higashakawa, 2004). By studying the movements of the ribcage and abdomen, they observed an increase in decoupling of the ribcage and abdomen during vocalization with increasing age, while the two components continued to show tight coupling during tidal breathing throughout the experimental period (Connaghan et al., 2004).

Previous findings in infant respiration indicate that very young infants are predominantly belly breathers, however, the ribcage begins to participate in vocalization breathing at fairly young age. As the infant grows, the ribcage contributes more and more to the breathing activity, and musculoskeletal growth, biomechanical changes, and postural change seem to play an important role in this change in respiratory pattern. A task-specific division of respiratory patterns is already evident in infancy. However, more research is needed to understand the relationship between breathing patterns that underlie cry and vocalization, and how development of speech breathing is related to mastery of other speech motor skills and voice control.

The larynx. The larynx is a cartilaginous framework located in the neck between the hyoid bone and trachea. It serves three functions: airway protection, airway patency, and voice generation. The larynx is closed to prevent accidental aspiration, open for respiration, and approximated for self-sustaining oscillation in phonation (Negus, 1929; Sasaki & Isaacson, 1988). The larynx is composed of three unpaired and three paired cartilages with muscles and soft tissues connecting them. Two of the three paired cartilages, cuneiform and corniculate cartilages, are vestigial in humans, but the remaining four cartilages are integral part of the basic functions. Muscles of the larynx are divided into intrinsic and extrinsic groups depending on their points of attachment. Intrinsic laryngeal muscles control the position, length, and tension of the vocal folds to produce voice of different F_0 and intensity. There are two vocal fold adductors, the lateral cricoarytenoid and interarytenoid muscles; one abductor, the posterior cricoarytenoid muscle; and two tensors, the thyroarytnoid and cricothyroid muscles. Extrinsic laryngeal muscles are also important in phonation in that they stabilize the larynx within the neck to provide a steady foundation on which the intrinsic laryngeal muscles work. The sternothyroid and sternohyoid muscles anchor the thyroid cartilage to the sternum, while the thyrohyoid muscle fixes the thyroid from above (Zemlin, 1998).

The infant larynx is not simply a miniature of the adult larynx, although it is much smaller in size. It is different from the adult larynx in terms of its dimensions and location. Even though the dimensions of the adult larynx are not established till puberty, considerable changes are reported to occur during the first two years of life (Crelin, 1973;

Tucker & Tucker, 1979). The infantile larynx is short and round in appearance compared to adults. It is funnel-shaped and more pliable due to underdeveloped cartilages (Kahane, 1996). It measures approximately 2 cm in both height and width, which is about one third of the size of the adult larynx (Crelin, 1973; Tucker & Tucker, 1979). The thyroid cartilage is located immediately below the hyoid bone with very little space between them. The space within the larynx becomes greater due to the size increase in the thyroid and cricoid cartilages (Bosma, 1975; Isaacson, 2003). A change seen in the arytenoid cartilages is more of form than size. The arytenoid cartilages are disproportionately large in infants, making the intercartilaginous glottis long relative to the membranous glottis. This arrangement is sometimes called "respiratory glottis" and favors the functions of ventilation and airway protection (Bosma, 1975; Hirano, Kurita, & Nakashima, 1983). Another implication of this arrangement is that a shorter length of the vocal fold is available for vibration during phonation, one acoustic consequence of which is a raised fundamental frequency (Kahane, 1996).

An increasing laryngeal skeletal size is also reflected in an increase in a size of the glottis. The length and thickness of the vocal fold increase dramatically in the first two decades of life. The length of the vocal fold is 3-7 mm in infants (Crelin, 1973; Hirano et al., 1983; Tucker & Tucker, 1979); they grow up to 11-17 mm in adult females and 15-25 in adult males (Hirano et al., 1983; Kahane, 1982). Within the vocal fold, Hirano el al. (1983) observed a rapid increase especially in the membranous vocal fold in the first decade of life. They noted an increase in membranous-cartilaginous length ratios from 1.1-1.8 in newborns to 3.3-6.2 in adults, and indicated a superior ability of vocal control in adults than in infants since mechanical properties of the vocal fold are mostly manipulated in the membranous portion where activities of the laryngeal muscles are maximally translated.

Intrinsic laryngeal muscles have also been studied (Kahane & Kahn, 1984). These investigators took weight measurements of an abductor, posterioir cricoarytenoid muscle, tensor, cricothyroid muscle, and adductors, lateral cricoarytenoid, thyroarytenoid, and interarytenoid muscles, and compared infants with adults. They found relative muscle weights to be similar between infants and adults, but the cricothyroid muscle to be the most massive muscle in infants. Since short-term variations in F_0 are regulated greatly by the vocal fold tension, the fairly well developed cricothyroid muscle may be related to the importance of pitch in infant vocal behaviors.

With regard to the position, the larynx is located relatively high in the neck at birth. During the first few years of life, there is considerable remodeling of the pharynx that accompanies the descent of the larynx. The pharynx, extending from the base of the skull to the lower airway, is short and straight in infants because their cranial base is flat and is aligned with the mandible. In this arrangement, the epiglottis and velum are in contact to create separate channels for food and air (Bosma, 1975, 1976; King, 1949). This is a typical arrangement widely found in mammals as well as human infants, and it is thought to be favorable in that it enables respiration while eating and also because it offers better protection from choking (Negus, 1949). However, in humans, the pharynx elongates postnatally and bends to attain an angular shape as the growing brain curves the basicranium (Laitman & Crelin, 1976). Consequently, the hyo-laryngeal complex descends from about the level of the second cervical vertebra to the fifth cervical vertebra during this pharyngeal growth (Lieberman, Crelin, & Klatt, 1972). Despite the obvious disadvantage in airway protection, the laryngeal descent results in two advantages in speech production.

First the pharyngeal restructuring gives rise to a “two-tube” resonator system in humans. As the pharynx elongates and curves, the posterior one third of the tongue moves with it and drops to form the anterior wall of the pharynx by 4-5 years of age (Crelin, 1987). In the infant, the entire tongue is found within the oral cavity, forming a single main resonator. On the other hand, with the new configuration, the anterior two thirds of the tongue resides in the oral cavity and the remaining posterior portion comes to be housed in the oropharynx. The two parts of the tongue can move separately to control cross-dimensions of the two tubes independently. This configuration is reported to be unique in humans and has an advantage in increasing the number of resonances to be produced (Laitman & Reidenberg, 1997; Lieberman et al., 1972).

Second, the descent of the larynx also offers separation of the larynx from the hyoid bone and supraglottal structures (Nishimura, 2003). In the infant, the hyoid bone is located roughly at the level of the mandible. The two bones line up horizontally and jointly support the tongue in its antero-posterior movement important in sucking and suckling. That is, when the hyo-laryngeal complex assumes a high position in the neck, it primarily serves swallowing function. According to Nishimura (2003), the laryngeal descent happened as a two-part sequence in evolutionary history. First, the hyoid bone descends relative to the mandible, then the thyroid cartilage descends further with respect to the hyoid bone. The first part creates the oropharynx, and the second partially frees the larynx from the deglutitive function. With a loose ligamentous connection between the larynx and hyoid bone, the two structures can function as a unit in deglutition, while preventing oral activities from being translated to the larynx other times. The fact that the hyoid bone is a point of attachment for many muscles of mastication, facial expression, as well as phonation, is not favorable in speech production, as the source should supply steady energy regardless of differences in articulatory gestures. Because some acoustic features are controlled at the larynx (e.g, vocal intensity and pitch), while others are controlled by the shape of the vocal tract (e.g., formant frequencies), some degree of structural independence between the vocal organ and resonator is desirable.

The vocal folds. The adult vocal fold is said to be a sophisticated oscillator that is adapted for the function of phonation (Hirano, 1974). The fine structure of the vocal fold is described to have different histological organizations that divide into distinct layers of the surface epithelium; superficial, intermediate, and deep layers of lamina propria; and the vocalis muscle. These layers mechanically form three sections that are the cover, consisting of the epithelium and the superficial layer of the lamina propria; the vocal ligament, consisting of the intermediate and deep layers of the lamina propria; and the vocalis muscle. Each section is characterized by different mechanical properties that can vibrate somewhat independently. The pliable cover is the main vibrating portion, where the mucosal wave takes place. There are no muscle fibers but extracellular matrices are distributed uniformly, giving the cover little internal stiffness. The vocal ligament or intermediate layer is high in elastic fibers and the deep layer is mostly collagenous. The gradual increase in elasticity of the tissue acts as an impedance matching mechanism and makes the adult’s vocal fold a superb vibrator that can oscillate in a wide range of amplitudes and fundamental frequencies (Kurita, Nagata, & Hirano, 1983).

A loose arrangement of reticular fibers is found in spaces of the superficial and intermediate layers, especially around the free edge of the vocal fold. Reticular fibrils do not form bundles like typical connective tissue bundles, instead they spread in three dimensions and form delicate networks of fibrils with spaces are filled with hyaluronic acid as well as elastic fibers. This structure elegantly explains the viscoelasticity of the vibrating portion of the vocal fold, which had been difficult to explain by means of elastic and collagenous fibers alone. Hyaluronic acid contributes to tissue viscosity, while elastic fibers and reticulate fibers provide elasticity and structural integrity in the vocal fold cover (Sato, 1998). Viscoelasticity is important in determining phonation threshold as well as shock absorption (Chan, Gray, & Titze, 2001; Titze, 1988; Ward, Thibeault, & Gray, 2002).

Significant differences in the internal structure of the vocal fold have been observed in infancy. Hirano and his colleagues demonstrated that the infant larynx lacks a layered structure (Hirano et al., 1983; Hirano & Sato, 1993). In the infant, instead of layered sections of tissue, there is an undifferentiated layer of tissue immediately below the surface epithelium that closely resembles the cover of the adult vocal fold. A ligamentous layer begins to form during the first few years of life, but it is slow growing and takes years before enough elastic and collagen fibers are produced, aligned, and differentiated to form distinct layers. The adult-like layers complete around puberty when sexual dimorphism of voice is manifested.

Even though the infant is capable of phonating, the indistinct structure of their vocal fold is suggestive of mechanical inferiority for this function (Hirano et al., 1983; Kurita et al., 1983). Two examples of such inferiority are increased phonation threshold and susceptibility to tissue injury. Maculae flavae are reported to be responsible for synthesis of extracellular matrices that later migrate to the lamina propria to provide optimal distribution of fibers and other extracellular matrices and ground substance. In the infant vocal fold, hyalurinic acid and elastic fibers are only scarcely present in the lamina propria, making the infant vocal fold high in viscosity and low in elasticity. Since hyalurinic acid decreases viscosity of the vocal fold, a low concentration of hyalurinic acid raises phonation threshold and causes high energy loss during phonation, making the infant vocal fold energy inefficient (Chan et al., 2001; Sato, Hirano, & Nakashima, 2001). On the other hand, low elasticity affects the tissue's ability to control F_0 (Chan et al., 2001).

The importance of maculae flavae is further illustrated in their involvement in the development of the vocal ligament. The anterior and posterior maculae flavae contain a large number of fibroblasts and are considered to be growth centers for collagenous and elastic fibers that are later migrate into the lamina propria to form a vocal ligament. In electromicroscopic studies of developing histology, it was revealed that collagenous and reticular fibers form and migrate prior to an increase in elastic fibers to provide structural scaffolding to which elastic fibers can be deposited (Hirano, Sato, & Nakashima, 1999; Sato, Hirano, & Nakashima, 2000). This process takes years to complete and the maculae flavae continuously produce extracellular matrices for the vocal fold mucosa (Sato, 1998). An innovative study of unphonated vocal folds provides a further insight into the unique nature of the human vocal fold (Sato, Nakashima, Nonaka, & Harabuchi, 2008). In the study, unphonated vocal folds were obtained from young adults who had structurally normal vocal folds but had been unable to phonate due to severe cerebral palsy. The histological structures of the unphonated and normal vocal folds were compared, and the unphonated folds were found to be hypoplastic and atrophic. A layered structure was absent, including a vocal ligament. The maculae flavae were much smaller and the vocal fold mucosa thinner. In light of these findings, they hypothesized that vocal fold oscillation of phonation causes the stress that stimulates the maculae flavae to accelerate production of extracellular matrices, which in turn leads to a formation of vocal ligament. Therefore, the lack of proper stimulation in unphonated vocal folds resulted in a reduced size of the macula flavae and also underdeveloped lamina propria. More evidence is needed to support the hypothesis, but these findings support a link between phonation and unique morphology of the human vocal fold.

Summary. Numerous developmental changes are observable in the anatomy and physiology of the vocal organs. Changes of great interest in relation to infant voice include an increase size and greater muscular control of the respiratory system, an increase in vocal fold length and thickness, and the histological makeup of the fine structure of the vocal fold. These critical changes are likely to alter some aspects of vocalization, but acoustic manifestations are not likely to be straightforward. Vocal output only reflects a net result of all underlying processes, and determining an exact contribution of each will be difficult as some changes are complementary, some are antagonistic, and interactions are hard to predict. However, understanding the state of development of the vocal anatomy should help guide acoustic analysis.

Chapter 3: Statement of Purpose

Research in infant vocalization has identified a systematic development for articulated speech in the first year. Modal phonation, the standard voice source for speech, is evidenced at the earliest stage, but studies have found non-modal phonations to be common in child's vocal repertoire throughout infancy and early childhood (Buder et al., 2008; Herzel & Reuter, 1997; Keating & Buhr, 1978; Kent & Murray, 1982; Robb & Saxman, 1988). Current literature on vocal fold vibration explains that changes in underlying aspects of phonation, such as subglottal pressure and vocal fold tension, can cause abrupt shifts in vibratory patterns of the vocal folds, allowing them to oscillate in various alternative modes (Berry et al., 1996; Švec et al., 1999). So far, few studies have examined phonatory characteristics of pre-linguistic vocalizations in terms of their vibratory patterns and none has incorporated respiratory behaviors. As a result, although phonation is integrated into current models of infant vocalization, a developmental course specific to modal phonation remains yet to be empirically established. In order to contribute to the body of knowledge of early vocal behavior, this dissertation aims to answer the following questions by examining acoustic and physiological aspects of phonation during the first 4-18 months of life.

Study 1 was designed to investigate the effect of vibratory regime on relationships between respiratory and acoustic measures of voice. If a mode of vocal fold vibration is conceived of as a distinct valving mechanism that converts respiratory airstream to acoustic signal, then the input-output relationship is expected to differ depending on the vibratory mode that accomplishes this conversion.

The preceding literature review indicated that respiratory drive is a major contributor of SPL (Isshiki, 1964; McGlone, 1970; Titze, 2000), and that there is also evidence that respiratory adjustments are related to changes in F_0 (Lieberman, 1967; Reilly, 2004; Strik & Boves, 1995). Such relationships become complex when vibratory regime is factored in. Some studies have reported differences in average levels of acoustic or aerodynamic measures between different regimes (Hollien, 1974; Keating & Buhr, 1978; McGlone, 1967; McGlone & Shipp, 1971b; Murry, 1971; Shipp & McGlone, 1971); others have reported differences in relationships between aerodynamic and acoustic measures when comparing across regimes (Isshiki, 1964; Titze, 1989). However, these findings have not been evaluated using chest wall movements recorded with respiratory inductance plethysmography, much less in spontaneous infant vocalizations.

Study 1 examined differences in means of F_0, intensity, relative airflow, and relative lung volume level between regimes, as well as relationships between these acoustic and respiratory measures. The two respiratory variables "airflow" and "relative volume" were not measured directly in this study. They are, instead, inferred from a slope of expiratory excursion and a point of excursion at regime initiation relative to local rest cycles. Specifically, sets of research questions and possible outcome based on adult findings are formulated as follows:

1. Is there a difference in means of F_0, intensity, airflow, or relative lung volume between the different regimes as seen in adult phonation?
    a. The pulse, modal, and loft regimes are by definition related to low, mid, and high pitch, respectively, but their associated F_0 values can vary among genders, age, and individuals (Hollien, 1974; Keating &

Buhr, 1978). Furthermore, mean F_0 means associated with subharmonics and biphonations remain unclear.

    b. Although intensity comparisons are difficult to make across different frequencies, the pulse regime is often reported to be lower in intensity than the modal or loft regimes (Hollien, 1974), but whether levels of intensity systematically differ between other regimes (i.e., subharmonics, biphonations, and chaos) is unclear (Buder, Oller, & Magoon, 2003).
    c. Similarly, airflow has been found in adults to be lower in the pulse regime than in modal (Blomgren, Chen, Ng, & Gilbert, 1998; McGlone, 1967; Murry, 1971). Whether this finding may extend to the expiratory slope measure or other regimes in infancy is unknown.
    d. There has been no report on relative lung volume level (i.e., what level relative to average tidal cycle did a particular regime segment occur?) and vocal regimes. However, it may be that relative lung volume at regime initiation can provide supplemental information on airflow because lung volume is related to the amount of elastic recoil force that the respiratory system is under.
2. Is there a relationship between acoustic and respiratory measures within each regime in infant phonation? If so, do the relationships differ between regimes?
    a. Subglottal pressure is the main regulator of intensity in adults, which is the product of glottal resistance and airflow. The contribution of airflow to intensity becomes more dominant with higher F_0 (Isshiki,

1964). On this basis, it is predicted that there should be a moderate to high positive correlation between intensity and both expiratory slope and relative lung volume in modal and loft, whereas there should be minimal correlation between them for pulse. No such relationships have been studied for other regimes nor has the question ever been investigated in infancy.

b. Although the relationship between airflow and F_0 is less robust, it is known that subglottal pressure tends to rise with increasing frequency (Titze, 1989), especially at high frequencies where the stiffness within the larynx is maximized (Shipp & McGlone, 1971), indicating an increase in F_0 at high frequencies is achieved by an increase in airflow rather than by an increase in stiffness of the vocal folds. On the other hand, F_0 declination at the end of a breath group demonstrates association between F_0 lowering and diminishing air supply (Lieberman, 1967; Strik & Boves, 1995), suggesting reduced airflow and pressure producing a decrease in F_0. The most direct evidence in infants comes from Reilly (2004), in which a correlation was found between expiratory velocities and F_0. Following these, it is predicted that there is a positive correlation between F_0 and both expiratory slope and relative lung volume for modal and loft but not in pulse. No such relationships have been studied for other regimes.

Following the assessment of these questions in Study 1 which demonstrated the need to incorporate the concept of vibratory regime, main research questions will focus on the production of modal phonation produced by an infant at an earlier stage of vocal development. Based on the observation that the expansion stage is known for highly variable patterns of phonation (Koopmans-van Beinum & van der Stelt, 1986; Oller, 2000; Stark, 1980), which is in contrast with steady modal voice of well-formed syllables found in the following canonical stage, vocal samples were collected from an infant of ages 4-6 months.

3. What are the acoustic (i.e., F_0 and SPL) and respiratory (i.e., expiratory slope) ranges of modal phonation with respect to other phonation types?
    a. Do pulse, modal, and loft regimes occupy low, mid, and high frequency ranges, respectively as presumed?
    b. The range of F_0 where modal and loft overlap is known to create a region of instabilities susceptible to bifurcations in experiments (Berry et al., 1996). Do locations for observed bifurcations (i.e., subharmonics, biphonation, and chaos) correspond to this region in Dataset II?
4. Is there a difference in ranges of modal phonation between the different ages? Much anatomical and physiological development is underway in the first year of life, both in intrinsic and extrinsic morphology and histology of the larynx and in mechanics of the respiratory apparatus. It may be that such maturation of speech motor system may be reflected in changes observed in vocal behaviors during this period. Furthermore the infant's experience in vocal exploration across that range may be the source of significant modification.

a. The two datasets are divided into three ages: early (4-6 months), middle (8-11 months), and late (14-18 months) infancy to compare ranges of F_0, SPL, and expiratory slope of modal phonation between the early and the two older ages.

b. The tendency for infants to use modal phonation in conjunction with canonical babbling suggests proportionally higher occurrence of modal phonation in older children, but it is unknown if this is due to:

    i. a decrease in occurrences of non-modal phonation types

    ii. reduction of modal ranges, suggesting a learning effect to avoid unstable regions to stay within certain limits of phonation

    iii. expansion of modal ranges, suggesting vocal fold maturation that enables the vocal folds to oscillate more steadily to sustain modal phonation over wider input conditions

Chapter 4: Methods

Two sets of data were involved in this dissertation: Dataset I and Dataset II. Dataset I provided data for Study 1 examining the effects of vibratory regime type on acoustic and respiratory variables. Dataset II was used to address the first question of Study 2 regarding ranges of modal phonation with respect to other phonation types, and both datasets were used for age comparisons of modal range. The two datasets employed the same instrumentation in obtaining signals. Recording procedures and subsequent data analysis were also identical except for minor details. The methods outlined below, therefore, focus on the collection of the new dataset. Details of data collection procedures for Dataset I are found in Parham (2008).

The study protocols were approved by the Institutional Review Board of University of Memphis (IRB numbers H07-84, Parham, 2008; IRB# H10-86, Appendix A). The purpose and procedures of the study as well as the safety of equipment were discussed with parents of infant participants prior to their enrollment. The parent was given an opportunity to ask questions before signing informed consent (Appendix B). Financial compensation or equivalent accommodation was provided for their participation.

Participants

Dataset I was acquired from eight infants who participated in a previous study (Parham, 2008). One to six recording sessions were obtained from each infant. Four of the eight infants had multiple recording visits at different dates, which resulted in 24 total sessions with a collective age range of 8-18 months (Table 7).

Dataset II was acquired from one infant (MW) between ages of 4-6 months. The intent behind targeting a younger infant was to sample the wider ranges of F_0, intensity, and voice quality commonly observed among infants of age 3-8 months (Koopmans-van Beinum & van der Stelt, 1986; Oller, 2000; Stark, 1981), and to do so with more intensive sampling over time. Increasing observations of non-modal phonations and allowing for possible age comparisons against Dataset I were additional benefits of obtaining new data.

Table 7

Ages of Participants and the Number of Recording Sessions Obtained

	Participant	Age in months	Number of sessions	Total sessions
Parham, 2008	1	8	4	4
	2	9	2	2
	3	14	4	6
		15	2	
	4	17	2	4
		18	2	
	5	10	1	1
	6	10	1	3
		11	2	
	7	15	2	2
	8	9	2	2
Current	MW	4	13	29
		5	9	
		6	7	

Recordings were made daily unless there were holidays, illness, or scheduling conflicts. MW's participation was terminated due to family relocation after two months. MW was mainly producing cooing, marginal babbles and vowel-like sounds of various qualities during his participation. The production of canonical babbles began to appear in parent report after 5 months and 1 week of age, likely marking an onset of the canonical stage, however marginal babbles and other vocalic sounds remained predominant during recording sessions. A vocal development questionnaire (Appendix D) was administered weekly to gauge whether vocal samples obtained in the laboratory were representative of his general vocal ability and tendency.

Prior to the initial recording, the parent filled out a demographic and health history questionnaire and vocalization survey (Appendix C). All participants were typically developing infants without significant pre- or postnatal histories including respiratory or cardiac problems. All had hearing sensitivity within normal limits and no signs of speech or language delay. No selection was made based on gender or ethnicity as there is no known effects of these factors on early vocal development.

Recording Procedures

All recordings were made in a sound-treated recording suite consisting of two rooms. Recording sessions were held in the afternoon usually after a nap. Infants were engaged in interaction in the playroom side of the suite furnished with quiet toys. A parent, the author, or an assistant remained in the room with the infant to elicit vocalization. Next to the playroom was the recording room where all equipment was located for signal monitoring and data acquisition.

At the beginning of each session, the infant was first fitted with two respiratory bands, over a thin garment to avoid skin irritation, and a microphone before placed in position. In Parham (2008), infants wore a microphone over the middle of their right clavicle on their clothing, and they were seated in a high chair with a feeding tray. In the new data collection, the microphone was located over the sternum, and the infant was positioned on his back either in a stroller or in a playpen. After equipment fitting, a sound level was calibrated by generating a tone adjacent to the infant's mouth and taking a sound pressure level (dB SPL) reading with a sound level meter near the microphone. Each recording session lasted up to 30 minutes. Each recording visit was consisted of one to two sessions in Parham (2008). The infant was monitored throughout a session and breaks were taken as needed.

Instrumentation

Three types of signals were collected for the study: audio, respiratory, and video signals. For the acquisition of acoustic signals, MW wore a custom vest that housed a microphone. The construction of the vest followed a design developed by Buder and Stoel-Gammon (2002), in which a Velcro patch securely held the microphone in place to maintain a constant mouth-to-microphone distance of about 5 cm. A small, low-friction, flat response microphone (EMW Omni Classic Lavalier, Countryman Associates, Inc., Menlo Park, CA) was connected to a wireless transmitter that sent the audio signal to a receiver (Airline UHF AL1/AM1 system, Samson Technologies Corp., Hauppauge, NY) located in the recording room.

The infant's respiratory movements were transduced using a respiratory inductive plethysmograph (RIP; Inductotrace®, Model 10.9000, Ambulatory Monitoring, Inc., Ardsley, NY). RIP exploits the self-inductance of a coil and measures changes in electrical current that occur in response to changes in a circumferential area surrounded by the coil (Martinot-Lagarde, Sartene, Mathieu, & Durand, 1988). That is, when a coil expands and contracts as the infant breathes in and out, the concurrent expansion and contraction of an enclosed magnetic field produces a current in the coil. The infant wore two elastic bands with coiled wires sewn into them, one around the ribcage under the axillae and one around the abdomen between the lowest vertebral rib and iliac crest. Two cables connected each of the transducer coils to a pediatric oscillator then to a single extended cable to an Inductotrace interface in the recording room. The RIP system generated three signals: ribcage, abdomen, and sum, which is a weighted sum of the other two. Gain levels of the RIP system were calibrated prior to the first recording session by following the procedure provided with the system. The gains for ribcage and abdomen channels were adjusted to achieve equal deflection for equal input voltages. Settings remained unchanged throughout the duration of data collection.

Two channels of video recording were also obtained from four adjustable cameras (Spectra III, Pelco, Clovis, CA) located in each of four corners of the playroom. The direction, focus, and zoom of selected two cameras were controlled from the recording room throughout a session to capture full-body view of the infant. Images were used for identification of movement artifact during respiratory signal screening prior to data analysis.

Signal Acquisition and Conditioning

One channel of audio and three channels of respiratory signals were captured by a DT322 acquisition card (Data Translation, Inc., Marlboro, MA), and digitized using the Time-Frequency Analysis for 32-bit Windows Lab Automation Level program (TF32; Milenkovic, 2001). The audio signal was digitized at 40 kHz, and the respiratory signals at 10 kHz, after being low-pass filtered at 18 kHz and 200 Hz, respectively using AAF-3 anti-aliasing card (Alligator Technologies, Inc., Costa Mesa, CA).

Following signal acquisition, respiratory waveforms were calibrated to determine relative contributions of the ribcage and abdomen to the overall breathing movement. During each recording session, small isovolume maneuvers were elicited by tickling, which usually caused the infant to hold his breath momentarily and shift a small volume of air between the ribcage and abdomen (Boliek et al., 1996). These movements produced waveforms where the ribcage waveform (RC) and abdomen waveform (AB) were opposite in direction with the slope of the sum signal equaling zero. Using MATLAB (The MathWorks, Inc., Natick, MA), respiratory waveforms were first downsampled to 100 Hz then isovolume segments were located. After adjusting an initial point to zero, values for the ribcage and abdomen for these segments were plotted on x-y coordinates with the x-axis representing the abdomen and y-axis the ribcage then a best-fit line was calculated. A new sum signal represented a measure of relative lung excursion (RLE), reflecting relative changes (i.e., no volume calibration) in a net movement of the chest wall, thus relating to lung volume. RLE was the arithmetic sum of a ribcage waveform and an abdomen waveform multiplied by the negative of the slope, m, of the best-fit line: $RLE = RC + (-m \times AB)$. The resulting sum signal replaced the one that was automatically computed by the RIP system during recording.

Variables and Data Coding

Once an audio signal was paired with its time-locked respiratory sum signal, all sessions were screened for valid utterances and adequate respiratory waveform. First, a human coder listened to each recording and identified all protophone utterances produced by an infant. A protophone utterance was defined as a non-cry, non-vegetative sound or a group of those sounds produced on a single expiratory phase. Respiratory signals were then inspected, and only those utterances with 1) relatively little movement artifact and 2) one or two referent rest cycles in the vicinity (Boliek et al., 1996; Parham. 2008) were further selected.

Vibratory regime. Vibratory regime refers to a distinct pattern of vocal fold vibration as revealed by auditory and visual inspection of a vocal signal. Six categories of mutually exclusive vibratory regime included in this study were: modal, pulse, loft/ high modal, subharmonics, biphonation, and chaos. Following Buder et al. (2008), defining characteristics of these regimes are summarized as follows.

Modal is defined as steady, periodic phonation that results in a spectrogram with evenly spaced harmonics that are integer multiples of the fundamental. Perceptually, it has a single pitch but may range in voice quality from pressed to breathy or even harsh. Segments with such qualities were coded as modal as long as enough tonality was detected by listening.

Pulse is low-pitched phonation that results in perception of individual glottal pulses that is heard as creaky, popping, or zipper-like. Harmonics are closely spaced in a narrow-band spectrogram. The pulse waveform often shows heavy damping between glottal excitations (Hollien, 1968). Although the concept of vocal register associates pulse, modal, and loft with low, mid, and high pitches, respectively, a coder was encouraged to listen for specific auditory qualities for each regime.

Loft is high-pitched phonation that is also known as falsetto. Harmonics are widely spaced in a narrow-band spectrogram. Loft is often associated with a sudden pitch jump from modal. Such shift is marked not only with a change in perceived pitch but also with a change in voice quality to a thinner, lighter, or squeakier quality. As it was pointed out in Buder et al. (2008), loft can be easily confused with high-pitched modal phonation (high modal = HM). Additional steps (described below) were implemented in order to help distinguish between the two categories.

Subharmonics (SH) is a set of extra harmonics that usually has twice or trice the original period. SH frequencies appear abruptly, accompanying a sudden lowering of pitch and “two-toned” roughness. Another characteristic is a unique waveform with regularly occurring glottal pulses with alternating stronger (original frequency) and weaker (subharmonic frequency) peaks, peaks of different shapes, or both.

Biphonation (BIPH) is defined as the presence of an extra set of harmonics that are, unlike subharmonics, unrelated to (i.e., not integer multiple of) the original set. A rapid modulation is often discernable in the waveform. The additional pitch must be clearly perceptible as a tone and not as modulation of the original pitch.

Chaos is characterized by irregular phonation. Harmonic structure is unclear and low frequency noise predominates in the spectrogram. Glottal pulses in the waveform show little periodicity and the tonality is audibly degraded.

Coders were graduate students majoring in speech-language pathology or audiology. Prior to coding, training was conducted using the protocol developed in Buder et al. (2008), in which coders learned to discern infant phonation from other sounds (e.g., background noise, supralaryngeal sounds, etc), to identify salient auditory and acoustic features of each regime, as well as to navigate the mechanics of the TF32 program. Vocal samples used in training were not part of the current datasets but produced by infants of similar age. Regime classification was performed using TF32 with the following settings:

bandwidth = 10 Hz, smallest displayed frequency range ≈ 0-6 kHz, dynamic range = 64 dB, amplitude floor = 90-100 dB.

Each utterance was examined for the presence of voicing (i.e., sound of laryngeal origin) that was longer than 50 milliseconds (ms). A portion of waveform that was uniform in 1) harmonic organization in a narrowband spectrogram, 2) waveform pattern, and 3) auditory quality were considered to form a single regime segment corresponding to one of the six vibratory patterns defined above. Segments shorter than 50 ms were subsumed into longer segments and their regime characteristics were disregarded. Once a regime segment was identified, a coder assigned a regime category.

Regime coding of Dataset I was done in two steps. In the first step, two independently working coders identified segments and assigned regime categories within each session. In the next step, two sets of regime codes were merged to produce a single regime category per identified segment. Disagreements in coding were resolved by a third coder who had been involved in the development of the original training scheme (Buder et al., 2008). Resulting sets of codes were the material used in subsequent analyses. In Dataset II, primary data were produced by a single coder. All utterances were coded by the author in consultation with an additional coder who was familiar with the task. One additional coder coded 20% of Dataset II for the purpose of reliability assessment only.

As mentioned earlier, the distinction between loft and HM can be difficult to make on the basis of perception alone. The acoustic measure of amplitude difference of the first (H1) and second (H2) harmonics (H1-H2) was used to aid this categorization in the present study. This follows the observation that the "true" loft is associated with more sinusoidal vocal fold oscillation which produces weak overtones (Titze, 2000), which in turn tends to produce greater H1-H2 for loft than for HM despite comparable F_0.

The procedures were as follows. All coders initially used HM to mark any regime segments that were perceived either loft or high-pitched modal. After two coders' responses were merged in Dataset I and all utterances in Dataset II were coded once, the author reevaluated all HM segments. The segments were assigned either to loft (i.e., judged with confidence) or to HM (i.e., uncertain if they were modal or loft). In all 53 recording sessions, there were 1739 modal, 158 loft, and 95 HM segments. Values of H1 and H2 were obtained from a power spectrum in TF32. H1-H2 was calculated by subtracting H2 from H1 for all three categories. Since no published data was available that specified what value of H1-H2 should separate modal and loft, discriminant analysis was used for the re-categorization of HM. Using H1-H2 values of modal and loft, a discriminant function was computed that classified all HM into either modal ($n = 69$) or loft ($n = 26$). Box's M indicated that the assumption of equality of covariance matrix of H1-H2 for the modal and loft groups was supported ($F = 0.004$, $p = 0.95$). The canonical correlation was 0.56, which suggested that the model explained 31.16% of the variance in the dependent variable (i.e., vocal category). Wilks' lambda of 0.69 indicated the statistical significance of the discriminant function ($p < 0.001$).

Acoustic variables. Two acoustic variables of interest were infant's vocal F_0 and sound pressure level (SPL). Changing values of both F_0 and SPL were calculated from audio signals using TF32 within the Action Analysis Coding and Training software (Delgado, 2008).

The program has a built-in pitch detection algorithm with adjustable parameters that allow for flexible control of voiced/unvoiced decision frequency, inverse filter detuning bandwidth, double pulse detection, energy floor, and minimum correlation coefficients and intervals. This algorithm has been developed as a collaboration of Paul Milenkovic and Eugene Buder, implemented in the AACT version of TF32. Pitch analysis was done in two steps. In the first path, the algorithm ran with voiced/unvoiced decision frequency set high (~ 8k Hz), correlation coefficients and intervals set at relatively liberal levels (~ 0.4 for coefficient and ~ 7 ms for interval), and the remaining parameters set at default. Recordings used in the present study contained cross-talk, background noise, and ambiguous F_0 of infant voice (i.e., subharmonics and biphonation) that caused the pitch detection algorithm to fail. Therefore, in the second path, local editing was applied to ensure proper tracking of F_0. Each regime segment was reviewed individually, and errors were overridden by one of the two methods: parameter adjustment or hand correction. Fine-tuning of the parameters worked effectively when the cause of error was a result of low amplitude, high F_0, or presence of subharmonics. An appropriate parameter was adjusted and re-applied locally to accommodate such idiosyncrasies. Manual correction was most often performed in cases of cross-talk and background noise. F_0 traces were created to match a first harmonic that was visible in the corresponding spectrogram. Segments were rejected when the spectrogram lacked F_0 or when it was unable to determine infant's F_0 from the spectrogram.

F_0 traces were computed for all but one of the six regimes. The three normal phonation types of modal, pulse, and loft are associated with a single F_0, while F_0 is ambiguous in other regimes. SH and BIPH have conflicting F_0s and chaos has no clear F_0.

The original, higher F_0 was tracked for SH and the dominant F_0 was tracked for BIPH. No F_0 was obtained for chaos.

SPL was obtained regardless of the periodicity of the signal. Relative amplitude of the signal was tracked by a root mean square (RMS) amplitude trace. Similar to pitch analysis, RMS analysis was a two-step process. AACT TF32 first computed RMS amplitude from the audio waveform. Next, each segment was inspected and hand edited to remove unwanted energy that was overlaid on infant's RMS. Segments were rejected from RMS analysis when there was continuous noise affecting overall level of recording throughout the duration of a segment. RMS values were translated into dB SPL by referencing them against a calibration tone. The formula used was: $SPL = CT + 20log_{10}\frac{RMS_{obs}}{RMS_{ref}}$, where CT is the calibration tone in dB SPL for a session, RMS_{obs} is an observed RMS value, and RMS_{ref} is the reference RMS value for a session.

Actual data for analyses were obtained from these tracings in two ways. Values used for Study 1 were means of F_0 and RMS tracings that were automatically calculated for each regime segment by AACT. For Study 2, F_0 and RMS tracings were exported from AACT at the sampling rate of 1000 Hz. All cases with missing F_0 values were then excluded from further analysis.

Respiratory variables. Two respiratory variables were obtained from an adjusted respiratory sum signal. The sum signal represented overall excursion of the chest wall (i.e., net effect of both ribcage and abdomen), thus its voltage levels were analogous to lung volume. The two variables used in the current study were relative lung volume level at regime initiation (RLV) and expiratory slope representing relative airflow. RLV was defined as the ratio of regime-initial voltage level (Vol_{RGMi}) to average tidal peak level

(Vol_{PEAK}), both relative to the tidal floor level (Vol_{FLOOR}). It was calculated as:

$$\boldsymbol{RLV = (Vol_{RGMi} - Vol_{FLOOR}/Vol_{PEAK} - Vol_{FLOOR}) \times 100.}$$

Expiratory slope was the time rate of change of chest wall excursion, and was calculated as: $Slope = (Vol_{RGMi} - Vol_{RGMe})/t$, where Vol_{RGMe} is regime-end voltage level, and t a regime duration in seconds.

Values of Vol_{RGMi} and Vol_{RGMe} were extracted from adjusted respiratory sum signals at start and end regime boundaries. Tidal peak and floor levels were calculated from a segment of the adjusted sum signal that contained five to six consecutive tidal cycles identified within each session. Vol_{PEAK} was calculated as the average of maxima of those cycles and Vol_{FLOOR} was the average of minima.

Reliability

Different numbers of utterances were repeated for different reliability assessments. An additional coder recoded all utterances for regime in Dataset I, while an additional coder repeated 20% of Dataset II for inter-rater reliability. The author recoded 20% of both datasets for intra-rater reliability of regime and 10% for intra-rater reliability of acoustic and respiratory measurements.

Reliability for regime coding needed to be evaluated for identification of 1) regime segments, 2) regime categories, and 3) segment onset and offset (Buder et al., 2008). Because two coders may differ in their identification of voiced segments, proper alignment of segments was ensured by inspection of boundaries. Figure 3 illustrates the decision making process involved in the judgment of regime paring. Examples of missing boundaries and agreement patterns are shown in Figure 4. Normally, there were corresponding segments (open double arrow in Figure 4) and boundaries (double arrow) that were marked by both coders, within which actual regime categories assigned may be the same or different (cases 1 & 2 in Figures 3 and 4). When the number of identified segments disagreed between two coders within a single utterance, "did not code (DNC)" (gray) and missing boundaries (dashed arrow) were indicated based on boundary locations (cases 3 & 4 in Figures 3 and 4). When one coder identified more segments than the other those boundaries that were closest in time were paired up first, then a missing boundary or a missing boundary and DNC were filled in to mark the presence of an extra segment in the other coder.

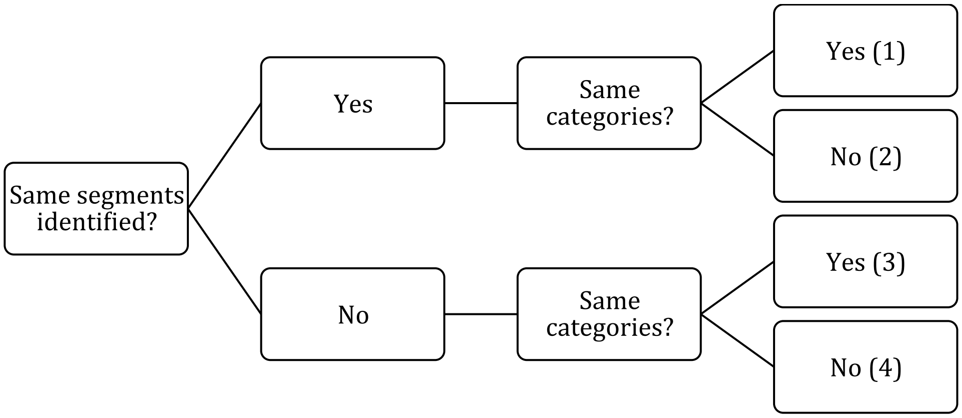
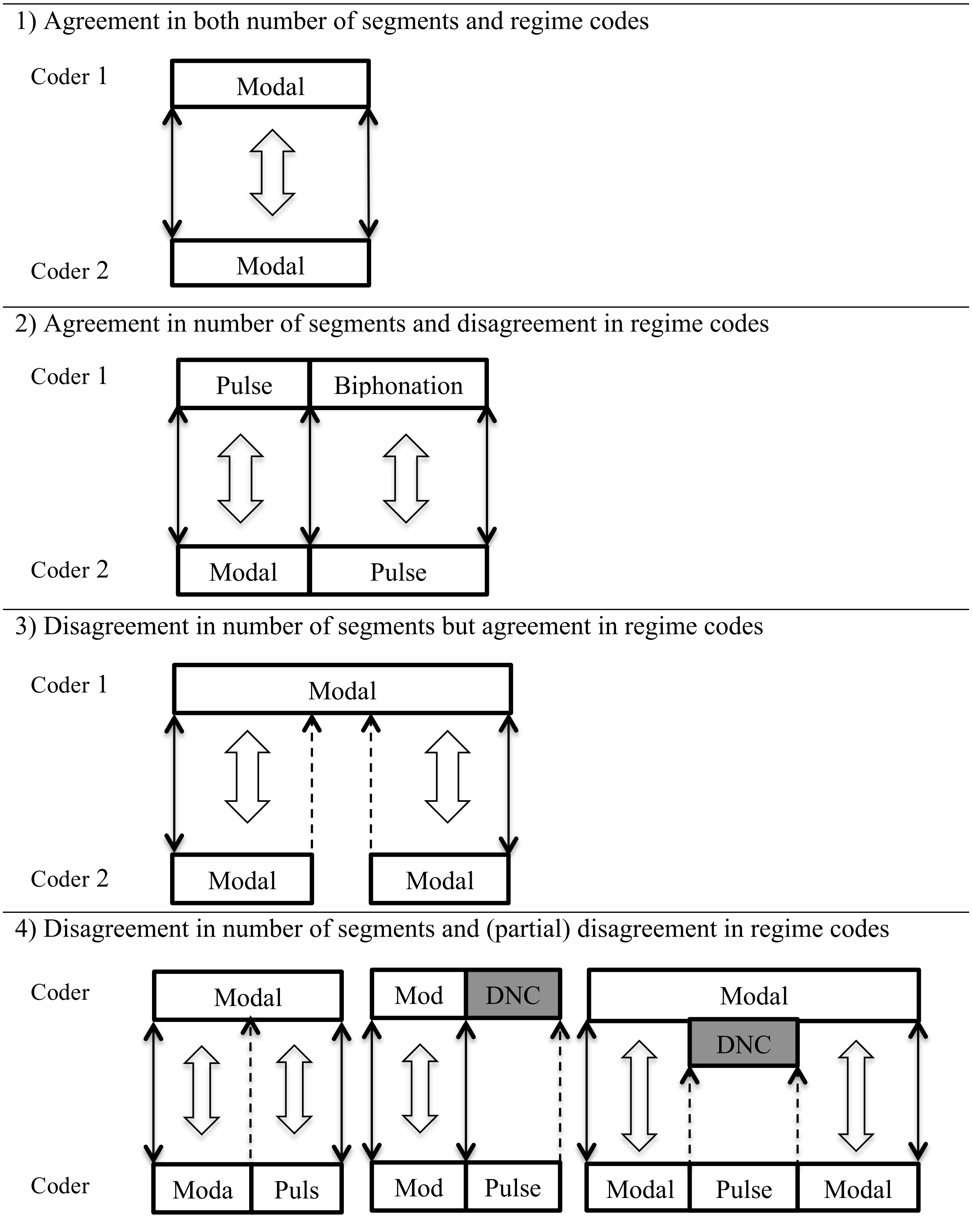


Figure 3. Decision-making tree for regime agreement.


1) Agreement in both number of segments and regime codes

2) Agreement in number of segments and disagreement in regime codes

3) Disagreement in number of segments but agreement in regime codes

4) Disagreement in number of segments and (partial) disagreement in regime codes

Figure 4. Schematic examples of missing boundaries and codes.

For the reliability of segment identification, the presence/absence of a segment was evaluated by percent agreement, which is $= \frac{segments\ in\ agreement}{total\ segments\ identified\ by\ rater\ pair} \times 100$, and also qualitatively by comparing regime categories where discrepancy occurred. Using actual regime labels assigned by two coders (i.e., regime pairs with DNC were excluded), the reliability of regime categories was assessed via Cohen's Kappa and percent agreement. They are both common measures of reliability involving categorical data. Cohen's Kappa factors in chance agreement levels due to observed cell frequencies into calculation of above-chance agreement. In general, Kappa values greater than 0.6 are considered to be good agreement (Di Eugenio, 2000), while Buder et al. (2008) reported 0.74-0.78 for regime classification among four coders trained using the same protocol.

All corresponding boundary pairs were used for the assessment of timing accuracy between two raters. The reliability was assessed quantitatively by correlations and *t*-tests of boundary locations between two raters, and also qualitatively by comparing differences in initial and final boundaries of each segment to ensure absolute differences among raters are within the margin of error.

Reliability of acoustic and respiratory measurements was evaluated by repeating approximately 10% of utterances as these semi-automated procedures involved less subjectivity. Using the regime boundaries identified within repeated utterances, the author reproduced F_0 and RMS contours from which descriptive statistics (mean, minimum, maximum, and standard deviation) were derived. Respiratory voltage levels were extracted based on regime-initial and regime-final positions. Intra-reliability was then assessed via Pearson Product Moment Correlation Coefficients (Pearson's *r*) and difference data and standard deviations of the differences between repeated measurements of these variables.

Statistical Analysis

Study 1: Effect of regime. The purpose of Study 1 was to determine the effect of vocal fold vibratory regime on 1) means of acoustic and respiratory variables and 2) relationships between the acoustic and respiratory variables. The first question was studied using a linear mixed-effects model (LMM) analysis. The independent variable was vibratory regime of voice; the dependent variables were two acoustic measures, mean F_0 and mean SPL, and two respiratory measures, expiratory slope which is a proxy for airflow, and regime-initial volume level relative to average tidal peal which is a proxy for lung volume level. A series of LMMs was performed to test differences in means of these acoustic and respiratory variables between six regimes, except for F_0 which was not obtained for Chaos. Regime was treated as a categorical fixed effect, and inherent variability in infants as a random effect. LMM considers observed values as the sum of both fixed and random effects. In addition, unlike the analysis of variance method used in the traditional general linear model, the restricted maximum likelihood method used in LMM is capable of handling unbalanced designs such as found in the current study.

The second question was studied via a series of correlation matrices between the two acoustic and two respiratory variables computed for different regime groups. Patterns of correlations were expected to differ among regime groups to reflect different relationships between input respiratory and output acoustic variables. Between-group differences in correlation coefficients were tested using Fisher's z-transformation to normalize correlation distributions.

Study 2: Ranges of modal phonation. The main question of this thesis focuses on the production of modal phonation in infants. This conceptual question is framed in terms of the acoustic and respiratory ranges associated with modal phonation. The operational research questions are straightforward: 1) what are the acoustic and respiratory ranges of modal phonation with respect to other phonation types? 2) How may those ranges differ with respect to age?

Data used for the first analysis were F_0, SPL, and expiratory slope values obtained from all regime types in Dataset II. Values of F_0 and SPL originated in respective tracings created in AACT that were exported at the sampling rate of 1000 Hz. The data were then downsampled to 100 Hz to ensure the ease of computation while maintaining enough sampling points per segment (i.e., 5 samples for a 50 ms segment, which was the shortest possible). On the other hand, a single data point in expiratory slope represented a mean value of expiratory slope for a given regime segment. Medians and middle 90% ranges were used to establish ranges of F_0, SPL, and expiratory slope for each regime type. A middle 90% range is the difference between 5th and 95th percentiles. It was chosen over a range (i.e., difference between a maximum and minimum) because it is less susceptible to extreme outliers that may be erroneous. Analysis of variance (ANOVA) was performed to test an effect of regime on means of F_0, SPL, and slope. Then, a Tukey's post-hoc test was performed to detect differences between modal phonation and other phonation types.

The second analysis employed all modal segments found in both Dataset I and II. The two datasets were divided into three age groups: early (1 participant of age 4-6 months), middle (5 participants of age 8-11 months), and late (3 participants of age 14-18 months) infancy groups. There were 1373 modal segments found in the early infancy group, 175 in the middle infancy group, and 261 in the late infancy group. In an attempt to correct for significant cell imbalance, 10,000 data points of F_0 and SPL were randomly selected from the original sets of F_0 and SPL values. ANOVA was performed to test an effect of age on means of F_0, SPL, and slope. Then, a Tukey's post-hoc test was performed to detect differences between the early and middle as well as between the early and late infancy groups.

Chapter 5: Results

Reliability

Inter-rater reliability of regime coding. Three coder pairs found 1008 regime segment pairs among 565 utterances used in the reliability assessment. Table 8 summarizes the number of segments found in the analysis by each coder pair. Coder pair 1 (LG and MD) analyzed 43 utterances from Dataset I and found a total of 89 regime pairs. Coder pair 2 (AK and MD) analyzed 181 utterances from Dataset I and found a total of 237 regime pairs. Coder pair 3 (MA and MD) analyzed 137 utterances from Dataset I and 204 from Dataset II, and found a total of 682 regime pairs. A regime-finding accuracy is the percent of regime pairs that were identified by two coders. Given these total numbers of regime pairs identified by each coder pair, the percent agreement was 95.5%, 99.5%, and 97.5% for coder pair 1, 2, and 3, respectively. This resulted in an overall regime-finding agreement of 97.8%.

Table 8

The Number of Segments Identified by Each Coder and Coder Pair

N	Pair	Coder	Segments found	Segments missed	Total segments found	Total segments missed	Percent agreement
43	1	LG	89	0	89	4	95.5%
		MD	85	4			
181	2	AK	236	1	237	1	99.5%
		MD	237	0			
137	3	MA	215	0	682	17	97.5%
		MD	210	5			
204	3	MA	463	4			
		MD	459	8			
Total					1008	22	97.8%

In total, 22 segments (2%) were marked by one coder and missed by the other. As summarized in Table 9, such disagreements occurred in all six regime categories, but more so in modal, pulse and chaos. A close examination of missing regime pairs suggested that many of these segments were short (average of 101 ms), had low amplitude, had noise overlay, or their combination. There were also cases, especially chaos, where two coders disagreed on the sound source (i.e., laryngeal vs. supra-laryngeal). Thirteen of the 22 "missed" segments occurred at either onset or offset of phonation, suggesting that the disagreements were due to discrepancies in boundary placement. One coder may have decided to code a segment where the other coder decided to extend a preceding or succeeding segment.

Table 9

The Number of Disagreement in Segment Identification

Coder pair	Missed regime						Total
	Modal	Pulse	HM[a]	SH[b]	BIPH[c]	Chaos	
1	1		1		2		4
2				1			1
3	5	8				4	17
Total	6	8	1	1	2	4	22

[a] high modal, [b] subharmonic, [c] biphonation

Regime categorization agreement was evaluated after accounting for regime-finding disagreements. Of 986 regime pairs, 820 pairs (83%) were in agreement. Percent agreement was 86%, 91%, and 80% for coder pair 1, 2, and 3, respectively. An overall Cohen's Kappa across all three coder pairs was 0.66. With an observed standard error of 0.022, this value is equivalent to a t of 30 with $df = 36$ ($p < 0.001$). Cohen's Kappas for individual coder pairs were 0.78, 0.76, and 0.63, for coder pair 1, 2, and 3, respectively, indicating good agreement well over chance levels.

Table 10 shows all the regime classifications after accounting for all the disagreements in segment identification, with off-diagonal cells representing codes in disagreement. Although there are no “correct” answers, some regimes are more consistently categorized than others. Coders agreed more on the classification of modal, pulse, and HM than they did on SH, BIPH, and chaos. Using MD’s codes as reference, accuracies were 88% for modal and pulse, followed by 87% for HM, 62% for BIPH, 53% for SH, and 12% for chaos. When the three non-normal regimes were combined, the agreement was 48%.

Some individual differences in application of assessment criteria are observed. LG, AK, and MA coded more pulse and high modal than MD, who tended to classify many of them as modal. In particular, 43% of what LG, AK, and MA coded as HM was considered modal by MD, illustrating the difficulty distinguishing the two regimes as reported in Buder et al. (2008). On the other hand, MD tended to code more chaos than the other coders as a group. What MD coded chaos was most often coded as biphonation. However, given the fact that chaos-biphonation confusion occurred between MD and one particular coder indicates that additional training could have helped calibrate their approaches to discern certain qualities in judging these regimes. Many cases of confusions occurred between modal and SH or BIPH, reflecting individual differences in deciding whether a segment presented enough qualities of an alternative regime to diverge from default modal phonation. Distinction between SH and BIPH was also difficult to make reliably. Coders may benefit from additional training emphasizing these regimes and also from the use of other acoustic tools such as power spectrum to judge the relationship of original and secondary sets of harmonics.

Table 10

Confusion Matrix Showing Codes by MD in Columns and Codes by LG, AK, and MA in Rows

		MD						Total
		Modal	Pulse	HM	SH	BIPH	Chaos	
LG, AK, MA	Modal	631	6	6	5	7	3	658
	Pulse	13	67	0	6	1	3	90
	HM	57	0	61	1	4	2	125
	SH	1	0	0	26	7	0	34
	BIPH	11	0	3	10	32	14	70
	Chaos	1	3	0	1	1	3	9
Total		714	76	70	49	52	25	986

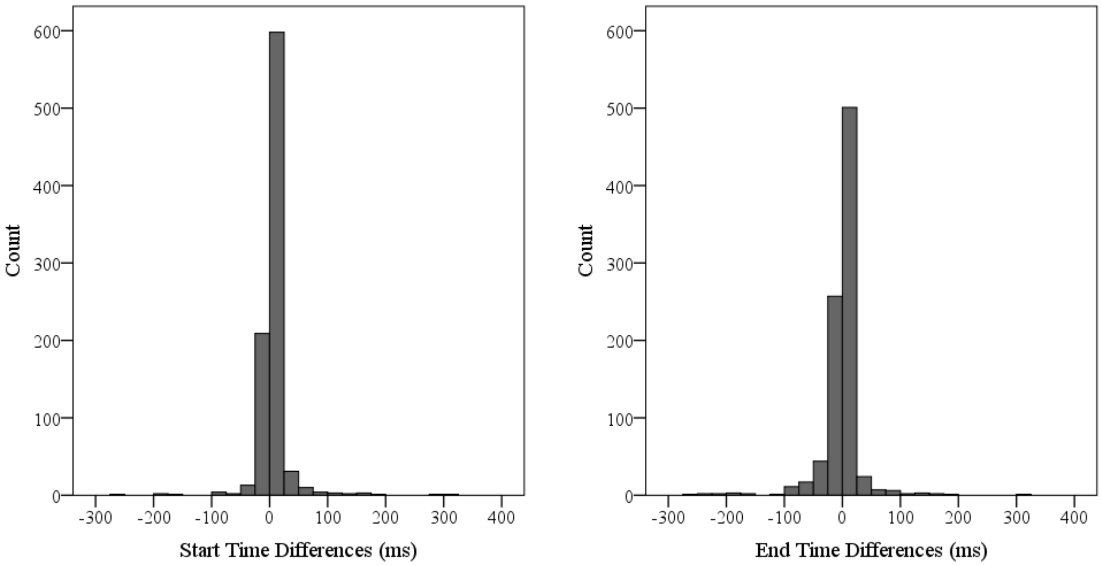


Figure 5. Histograms of start and end time differences between two coders.

There were 886 regime initial boundary pairs and 887 regime final boundary pairs. The mean difference was 3.55 ms [t (885) = 3.55; $p < 0.001$] for initial boundaries, and 2.34 ms [t (886) = 2.34; $p < 0.001$] for final boundaries, both with Pearson correlation coefficient of 1 ($p < 0.001$). The mean differences reached statistical significance due to large sample sizes, however, extremely small effect sizes (*Cohen's d* < 0.001) highlight that absolute differences in boundary locations between two coders were slight. Figure 5 displays histograms of boundary time differences between two coders. Among 886 comparisons made for initial boundaries, 98% were under 100 ms, 91% were under 25 ms, and only 15 cases (2%) exceeded 100 ms. For final boundaries 887 comparisons were made, among which 98% were under 100 ms, 86 % were under 25 ms, and 20 cases (2%) exceeded 100 ms.

Intra-rater reliability of regime coding. The author recoded regime in 279 utterances for intra-rater reliability about six months after the initial round of coding. A total of 545 pairs of regime segments were found. Five regime segments (1%) were only identified in one of the two codings, resulting in the regime finding accuracy of 99%. Within 540 pairs of segments that were identified in both codings, 470 pairs were in agreement (87%). Modal, BIPH and chaos were missed once, and pulse was missed twice. Table 11 displays all agreements and disagreements in regime classification. A Cohen's Kappa was 0.76. The accuracy was highest for modal at 95%, followed by HM (83%), SH (80%), pulse (70%), chaos (65%), and BIPH (54%). Greater categorization accuracies were observed within a coder than between coders especially for SH and chaos. Although inra-rater reliability was not tested for other coders, suggesting that regime categorization accuracy can be improved for these difficult regimes, and that different coders rely on different attributes of regimes in coding. Inconsistencies were also detected between pulse and modal. Given that the average F_0 of those pulse segments that were recoded as modal was 194 Hz, high frequency pulse can be ambiguous.

Table 11

Confusion Matrix Showing Codes in First coding in Columns and Codes in Second Coding in Rows

		First coding						Total
		Modal	Pulse	HM	SH	BIPH	Chaos	
Recoding	Modal	332	17	8	5	4	0	366
	Pulse	3	39	0	0	0	0	42
	HM	9	0	45	1	2	1	58
	SH	3	0	0	24	6	0	33
	BIPH	1	0	1	0	19	5	26
	Chaos	0	0	0	0	4	11	15
Total		348	56	54	30	35	17	540

There were 484 segment pairs found for both initial and final boundaries. The mean difference was 0.9 ms [t (483) = 0.91; p = 0.928] for initial boundaries, and 0.4 ms [t (483) = 0.39; p = 0.695] for final boundaries, both with Pearson correlation coefficient of 1 ($p < 0.001$). Non-significant t-tests and extremely small effect sizes show that differences in boundary locations between two codings were slight. Comparisons of 484 initial boundaries found that 99% were under 100 ms, 95% were under 25 ms, and only 7 cases (1%) exceeded 100 ms. Comparisons of 484 final boundaries found that 99% were under 100 ms, 93% were under 25 ms, and only 6 cases (1%) exceeded 100 ms.

Intra-rater reliability of acoustic coding. The number of utterances reanalyzed was 138, which contained varying numbers of regime segments for different variables. F_0 traces were found in 264 segments. Correlations of F_0 means, minima, maxima, and standard deviations between first and second traces were 0.985, 0.940, 0.960, and 0.911, respectively. RMS traces were found in 247 segments. Correlations of RMS means, minima, maxima, and standard deviations between first and second traces were all 1. Respiratory voltage levels were reanalyzed in all 275 segments. Correlations of regime-initial and regime-final voltage levels were 1 for both. The perfect correlations for the RMS statistics and voltage levels were as expected since the variables required minimal to no manipulation. Especially for the extraction of voltage levels, the only source of human error was correct detection of segment boundaries.

Difference data are summarized in Table 12. The examination of difference data revealed small to no mean differences between repeated measurements, in keeping with correlations. There was no difference between the two measurements for voltage levels. Differences of RMS-related measurements were miniscule. The number of segments where the two measurements disagreed was 8 for RMS means, 1 for RMS minima, 6 for RMS maxima, and 8 for RMS standard deviations, indicating that manual editing of RMS traces were kept to a minimum and scarcely affected the RMS statistics. Greater discrepancies were observed for F_0-related measures, though they were small. Of 264 comparisons made for F_0 means, 79% had differences smaller than 10 Hz and 2% exceeded 100 Hz. For F_0 minima, 61% had differences smaller than 10 Hz and 8% exceeded 100 Hz. For F_0 maxima, 69% had differences smaller than 10 Hz and 7% exceeded 100 Hz. For F_0 standard deviation, 85% had differences smaller than 10 Hz and 0.4% exceeded 100 Hz. The numbers of segment pairs with differences greater than 100 Hz were small, ranging from 1 to 22. Large differences in F_0 means often occurred when a secondary F_0 was mistakenly traced in cases of subharmonics and biphonation. On the other hand, large differences in F_0 minima and maxima occurred due to outliers. It was noticed during training that some of the incorrect extreme values computed by the pitch algorithm could be overlooked during local and hand editing when they are isolated. When this happened, affected F_0 minima and maxima produced large differences from those reflecting correct values. Nonetheless, the occurrence of such values seemed too infrequent to have affected F_0 means.

Table 12

Differences between Two Measurements of Acoustic and Respiratory Variables

Variable		*N*	*Mean*	Maximum difference	*SD*
F_0	Mean	264	2.0	457.3	38.8
	Minimum		5.0	491.5	72.0
	Maximum		0.7	505.8	66.9
	Std. Dev.		0.8	103.9	13.7
RMS	Mean	245	0.00006	.006	.0009081
	Minimum		0.00014	0.008	.0022361
	Maximum		0.00089	.1220	.0147137
	Std. Dev.		0.00114	.0180	.0174159
Initial voltage level		275	0	0	0
End voltage level			0	0	0

Study 1: Effect of Regime

Dataset I consisted of 348 utterances, in which 528 regime segments were identified. Segments ranged in duration from 50 ms, which was the minimum duration requirement, to 3141 ms, with a mean of 364 ms. The number of regime segments identified within a single utterance varied from 1 to 9, with a mean of 1.5. There were 436 modal, 27 pulse, 18 loft, 25 SH, 13 BIPH, and 9 chaos segments. Modal phonation was by far the most frequent, making up 82.6% of the data. Nevertheless, non-normal phonation modes (subharmonics, biphonation, and chaos) contributed 8.9% of the data.

This is similar to the frequencies reported by Kent and Murray (1982) and Robb and Saxman (1988), and supports regular occurrences of non-normal phonation modes in 9-18 month-olds.

The first question examined means of F_0, SPL, expiratory slope, and relative volume level at regime initiation (RVL) among different regimes using LMM. Overall *F* tests revealed significant differences among regime categories in three of the four outcome variables: F_0, SPL, and RI (Table 13). There was no statistically significant difference in RLV between regime categories. For the three variables where regime differences were found significant, means, standard errors (SEs) and differences from modal are represented in Table 14. For multiple pair-wise comparisons, a Bonferroni-corrected alpha level of 0.013 (= 0.05/4) was used for mean F_0 and 0.01 (= 0.05/5) for mean SPL and slope.

Table 13

Overall F-test Results for the Four Outcome Measures

Measure	Numerator *df*	Denominator *df*	*F*	*p*
F_0	4	512	117.56	< 0.001
SPL	5	473	10.16	< 0.001
Slope	5	522	5.99	< 0.001
RVL	5	522	1.67	0.14

As expected, F_0 was found to be statistically significantly higher for loft and lower for pulse than modal. Mean F_0 values for pulse and loft are similar to the means reported for 66-69 week olds in Keating and Buhr (1978), and the mean for modal is also similar to what has been reported in the previous studies, especially the values in Nathani

Iyer and Oller (2008) and Rothgänger (2003). Furthermore, mean F_0 for SH and BIPH were also found to be statistically significantly higher than that for modal.

In comparison of mean SPL values, the mean SPL for pulse was statistically significantly lower than that for modal. In addition, BIPH was found to have a significantly higher mean SPL than modal. No significant difference was found in mean SPL were found between modal and the other three regimes.

For means of expiratory slope, the only statistically significant difference was found between modal and BIPH. Similarity expected between SPL and airflow was observed only in BIPH but not in pulse. No significant difference was found in mean expiratory slopes between modal and the other four regimes.

Table 14

Means and Mean Differences of F_0, SPL, and Expiratory Slope for Each Regime

	Regime	*Mean* (*SE*s)	Difference from Modal	*p*
F_0	Modal	321 (5.12)		
	Pulse	134 (20.52)	- 187	< 0.001
	Loft	780 (25.13)	459	< 0.001
	SH	417 (21.32)	96	< 0.001
	BIPH	520 (29.57)	199	< 0.001
	Regime	*Mean* (*SE*s)	Difference from Modal	*p*
SPL	Modal	61 (0.45)		
	Pulse	53 (1.75)	- 8	< 0.001
	Loft	65 (2.22)	4	1.000
	SH	64 (1.82)	3	1.000
	BIPH	73 (2.47)	11	< 0.001
	Chaos	58 (2.97)	- 3	1.000
	Regime	*Mean* (*SE*s)	Difference from Modal	*p*
Slope	Modal	1.72 (0.10)		
	Pulse	1.06 (0.41)	- 0.66	1.000
	Loft	2.85 (0.50)	1.13	0.422
	SH	1.80 (0.42)	0.08	1.000
	BIPH	4.20 (0.59)	2.48	0.001
	Chaos	3.38 (0.71)	1.66	0.311

The second question examined regime-dependent relationships between acoustic and respiratory variables via correlation analysis. Table 15 displays Pearson correlation coefficients that were computed for each regime. Using an alpha level of 0.01 for multiple comparisons, a statistically significant correlation was found between mean SPL and RVL for modal and between mean SPL and slope for SH as well as between mean F_0 and slope for modal and between mean F_0 and RVL for loft. Differences in observed correlations among regimes illustrate that regime needs to be taken into account when examining relationships between acoustic and respiratory variables.

Table 15

Correlations between Acoustic and Respiratory Variables for Different Regime Types

		n	Slope	Relative volume level
Mean SPL	Modal	391	0.056	0.153*
	Pulse	26	-0.292	-0.043
	Loft	16	0.152	-0.257
	SH	24	**0.536***	0.239
	BIPH	13	-0.054	0.204
	Chaos	9	0.714	-0.191
Mean F_0	Modal	434	0.334*	0.090
	Pulse	27	-0.145	-0.123
	Loft	18	-0.446	**-0.737***
	SH	25	0.198	-0.430
	BIPH	13	0.005	-0.310

* $p < 0.01$

Note. Bold face indicates a significant difference from modal ($p < 0.05$)

A mild positive correlation indicated that a higher mean SPL was associated with a greater RLV for modal, however, considering the size of correlation and the high representation of modal in the dataset, this correlation is negligible. A moderate positive correlation between mean SPL and expiratory slope for SH indicated that a greater mean

SPL was associated with a steeper time rate change of chest wall movement during subharmonic segments. The Fisher r-to-z transformation revealed a statistically significant difference of this correlation coefficient from that found for modal ($z = -2.42$, $p = 0.016$). Contrary to the prediction, no significant correlation was observed between mean SPL and respiratory variables for loft.

A moderate positive correlation indicated that a higher mean F_0 was associated with a steeper expiratory slope during modal segments. A moderate correlation was found between mean F_0 and RLV for loft, which was significantly different from the correlation found for modal ($z = 3.94$, $p < 0.001$). However, the direction of correlation was opposite of what was predicted. The negative correlation indicated that a greater mean F_0 was associated with a lower relative volume level at regime initiation.

Study 2: Ranges of Modal Phonation

Dataset II consisted of 1022 utterances, in which a total of 2084 regime segments were identified. Segments ranged in duration from 51 ms to 5607 ms, with a mean of 375 ms. The number of regime segments identified within a single utterance varied from 1 to 13, with a mean of 2. There were 1373 modal, 245 pulse, 166 loft, 87 SH, 164 BIPH, and 49 chaos segments. The proportion of modal in this dataset was 65.9%, while SH, BIPH, and Chaos contributed 14.5% of the data. The higher proportion of non-normal phonation types in the younger dataset (Table 16) is in agreement with the characteristic of the expansion stage.

Table 16

Frequency Table for the Number of Observations in Each Regime

	Dataset I		Dataset II	
Regime	Frequency	%	Frequency	%
Modal	436	82.6	1373	65.9
Pulse	27	5.1	245	11.8
Loft	18	3.4	166	8.0
SH	25	4.7	87	4.7
BIPH	13	2.5	164	6.3
Chaos	9	1.7	49	2.4
Total	528	100	2084	100

Medians, middle 90% ranges, numbers of observations (*n*), and results of post-hoc comparisons are summarized in Tables 17-19. Medians and ranges are also depicted in Figures 6, 7 and 9 in which upper and lower ends of each gray line correspond to 95th percentile and 5th percentile, respectively, and medians are marked with cross signs along the lines.

Median values of F_0 for pulse, modal, loft were as expected, ranging from low to high, respectively. In addition to loft, median values of F_0 for SH and BIPH were also higher than that for modal. A one-way ANOVA revealed a statistically significant effect of regime types on F_0, $F\ (4, 71726) = 15926$, $p < 0.001$. Post-hoc comparisons revealed statistically significant differences between modal and all other regime types (Table 17). The medians of each regime found in this dataset were similar to means of each regime found in Study 1, and also their ordering was identical.

The 90% range was smallest for SH and greatest for loft. A large range for loft is consistent with the high frequency nature of this regime. On the contrary, the F_0 range was not smallest for pulse, which had the lowest median F_0. In fact, in terms of octaves, the F_0 range was greatest for pulse (5 octaves), whereas the range of loft was 2.5 octaves.

Most of pulse frequencies occupied a region of F_0 that is below 290 Hz, whereas most of loft frequencies were found above 450 Hz. Modal occupied a mid-frequency range between 250-500 Hz, whereas SH and BIPH appear commonly between 350-500 Hz.

Table 17

Medians and Ranges of F_0 and Results of Post-hoc Comparisons

	Regime (*n*)				
	Modal (58270)	Pulse (5443)	Loft (3362)	SH (1868)	BIPH (2788)
95th percentile	525	290	1074	540	706
5th percentile	258	58	436	342	364
90% range	267	232	638	198	342
Median	392	153	586	441	524
Tukey's *p* value	-	< 0.001	< 0.001	< 0.001	< 0.001

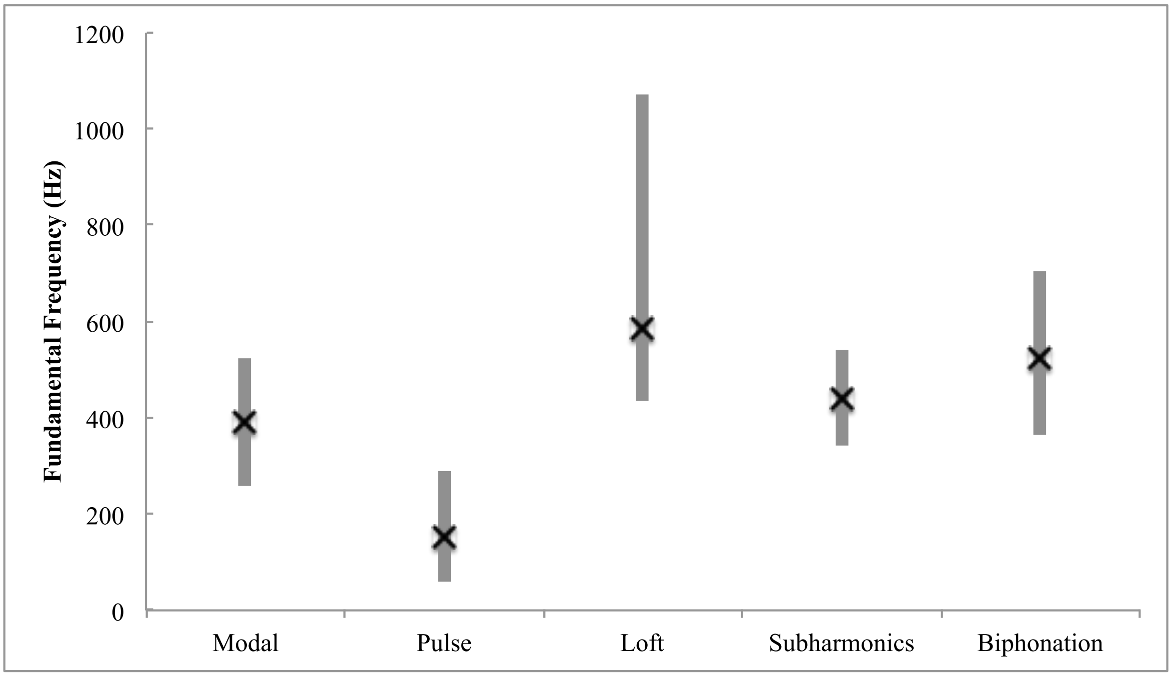


Figure 6. Middle 90% ranges of F_0 for each regime.

A median of SPL was lowest for pulse, followed by modal. Medians for loft, SH, BIPH, and chaos were somewhat similar, only ranging from 79 to 86. A one-way ANOVA revealed a statistically significant effect of regime types on SPL, F (5, 72645) = 3242, $p < 0.001$. Post-hoc comparisons revealed statistically significant differences between modal and all other regime types (Table 18). The directions of significant mean differences between modal and pulse as well as between modal and BIPH were consistent with the findings in Study 1. In Dataset II, mean differences were found additionally between modal and loft, SH, and chaos. Ranges of SPL were similar for all regime types, indicating that 90% of values were contained in a 31-33 dB range about their medians except for BIPH, whose range was somewhat smaller.

Table 18

Medians and Ranges of SPL and Results of Post-hoc Comparisons

	Regime (*n*)					
	Modal (582611)	Pulse (54475)	Loft (33660)	SH (18728)	BIPH (27800)	Chaos (9236)
95th percentile	89	78	94	98	97	98
5th percentile	56	47	63	67	73	65
90% range	33	31	31	31	24	33
Median	74	59	79	83	84	86
Tukey's *p* value	-	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

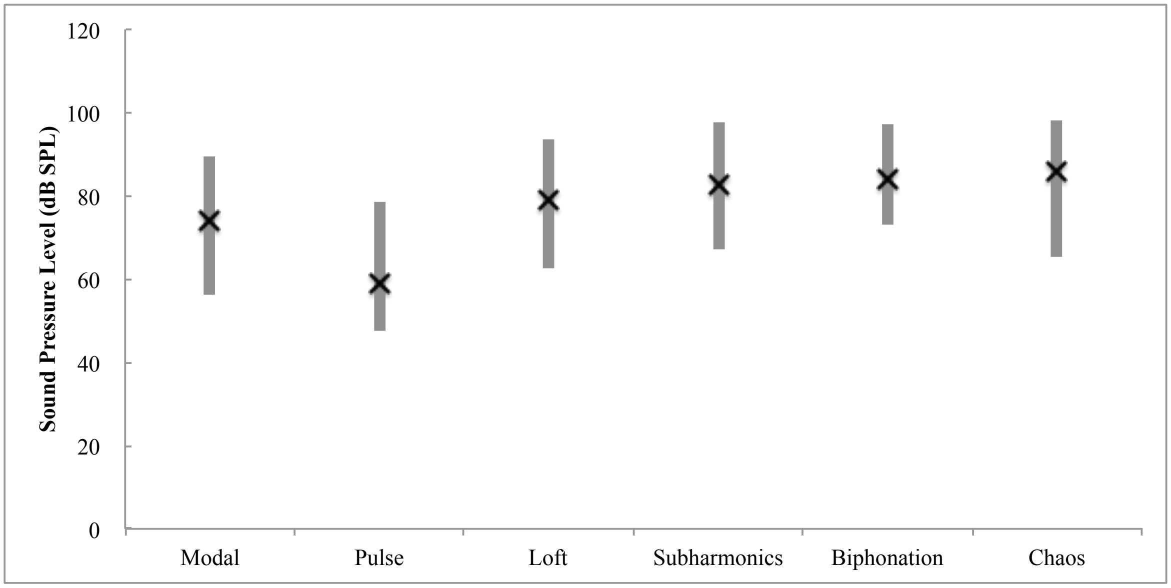


Figure 7. Middle 90% ranges of SPL for each regime.

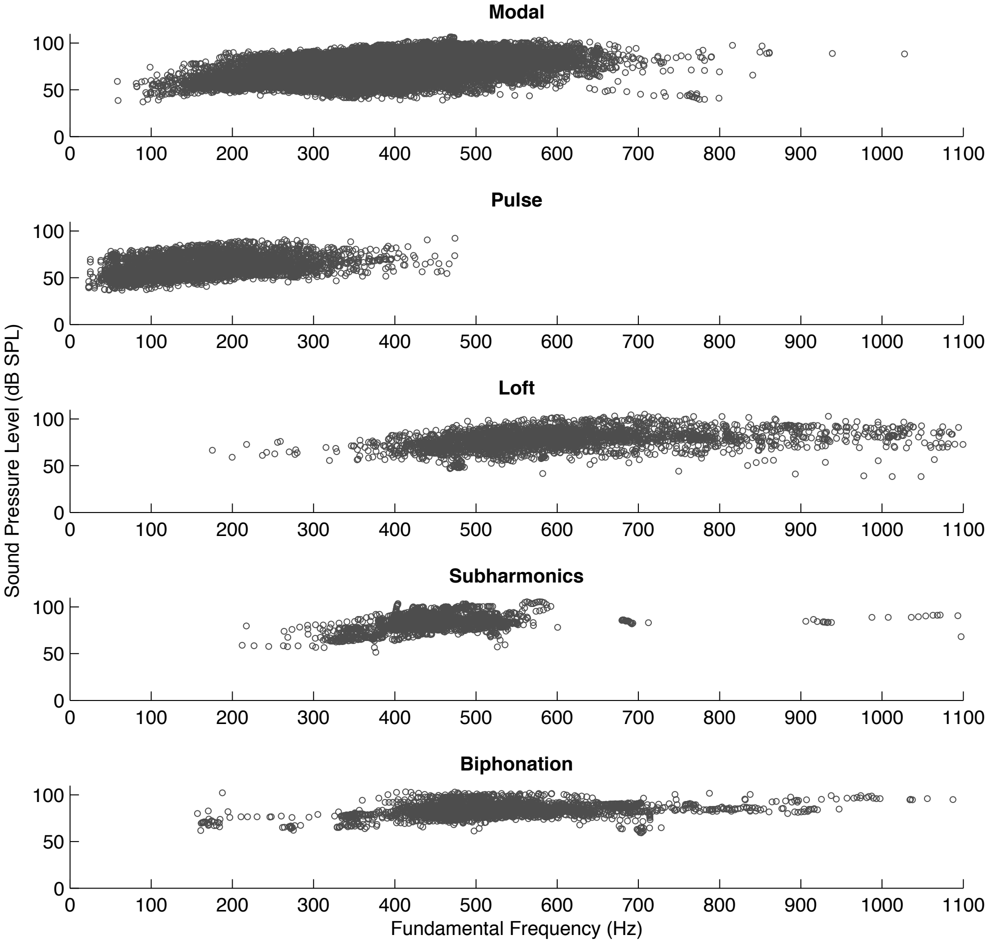


Figure 8. Intensity-frequency plot for five regimes. The frequency is truncated at 1100 Hz to include 90% ranges for all regimes.

Figure 8 is a series of intensity-frequency plots constructed for five regimes with F_0 truncated at 1000 Hz in order to show a close-up view of distributions where the vast majority of data points were found. Several observations can be made from the figure about ranges of each regime. Modal spreads fairly widely along the mid-range of both F_0 and SPL. Pulse covers the region that is lower both in F_0 and SPL than modal. Loft tends to show the opposite trend, covering the region that is higher in both F_0 and SPL than modal. SH has a restricted distribution along F_0 that is concentrated between 400 and 500 Hz. Less so than SH, but the distribution of BIPH is also tight, spreading over 400-700 Hz. The SPL ranges of both SH and BIPH are slightly elevated than modal, and they are rarely found in low levels of SPL. The majority of data points for SH and BIPH are present in the area that is the upper end of modal range and lower end of loft in terms of F_0 and higher in SPL.

Expiratory slopes in this study can be treated as a proxy for airflow. Though these uncalibrated values do not correspond to a meaningful physical quantity, comparisons of relative magnitudes and sizes of ranges could still be made. The order of medians of expiratory slope was, from smallest to greatest: pulse, SH, modal, BIPH, chaos, and loft. A one-way ANOVA revealed a statistically significant effect of regime types on expiratory slope, $F\ (5, 2078) = 49.09$, $p < 0.001$. Post-hoc comparisons revealed statistically significant differences between modal and pulse, loft and BIPH (Table 19). That is, the mean expiratory slope was statistically significantly flatter in pulse than in modal, which is in agreement with the prediction made in Study 1 based on the literature. The mean slope was steeper in loft and BIPH than in modal. The significant mean difference between modal and BIPH was consistent with the finding in Study 1. The significance levels for SH and chaos did not reach a Bonferroni-corrected alpha level of 0.01 (= 0.05/5).

Table 19

Medians and Ranges of Expiratory Slope and Results of Post-hoc Comparisons

	Modal	Pulse	Loft	SH	BIPH	Chaos
95th percentile	2.951	2.503	5.801	2.599	3.801	4.805
5th percentile	-0.215	-0.540	0.014	-0.808	-0.169	-0.690
90% range	3.166	3.043	5.787	3.407	3.970	5.496
Median	0.686	0.225	1.908	0.567	1.161	1.240
Tukey's *p* value	-	< 0.001	< 0.001	0.923	< 0.001	0.042

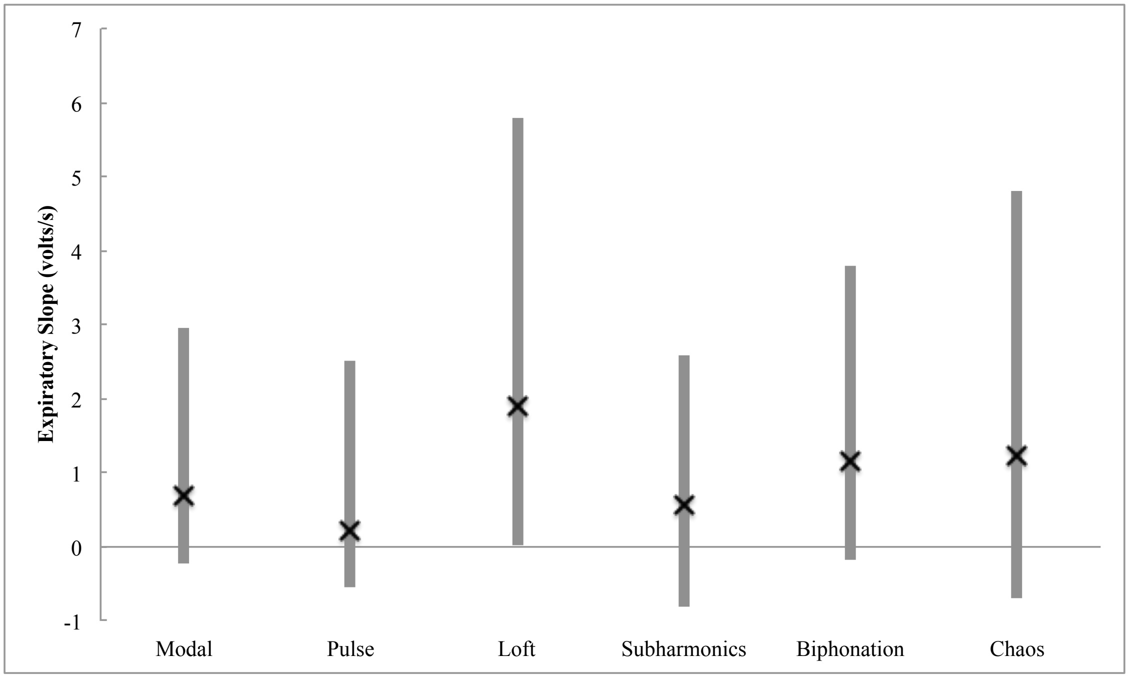


Figure 9. Middle 90% ranges of expiratory slope for each regime.

The 90% ranges of slope were smaller for pulse, modal, and SH (< 4 volts/s), and restricted to a lower range. On the contrary, the ranges were larger for chaos and loft (> 5 volts/s), encompassing nearly twice the range of modal. This indicated that relative airflow rates used to produce loft and chaos were highly variable. Both range and median slope of BIPH lay between the two groups.

It was also noticed that negative values of slope were not uncommon among the regime segments studied. These negative slopes are associated with local maxima in a respiratory signal and not caused by gross bodily movement or utterances made on ingressive airflow. Although they may be due to instrumental error, further exploration of these segments in conjunction with perceived glottal resistance (e.g., squeezed voice) may reveal additional laryngeal behavior of interest.

For the exploration of an overall age effect, all modal segments from both Dataset I and II were considered. Dataset I formed an early infancy group, although it was a group of sessions obtained from a single child. Dataset II was divided into two older age groups: middle (8-11 months) and late (14-18 months) infancy. Figure 10 represents a set of intensity-frequency plots showing distributions of 10,000 F_0 and SPL values randomly selected from each age group, and Figure 12 is a plot showing ranges of expiratory slope for three age groups. Associating measures of spread are presented in Tables 20-22.

Medians of F_0 and SPL were greatest for the 4-6 month old. A one-way ANOVA revealed a statistically significant effect of age on both F_0, $F\ (2, 29997) = 4423$, $p < 0.001$, and SPL, $F\ (2, 29997) = 5098$, $p < 0.001$. Post-hoc comparisons revealed statistically significant differences between the young and the two older groups for both variables (Tables 20 and 21), suggesting a decreasing trend of mean F_0 and SPL with increasing age.

Table 20

Medians and Ranges of Modal F_0 and Results of Post-hoc Comparisons

	Age		
	Early 1 participant 4-6 months	Middle 5 participants 8-11 months	Late 3 participants 14-18 months
95th percentile	527	484	406
5th percentile	258	239	216
90% range	270	245	190
Median	394	304	290
Tukey's *p* value	-	< 0.001	< 0.001

Table 21

Medians and Ranges of Modal SPL and Results of Post-hoc Comparisons

	Age		
	Early	Middle	Late
95th percentile	90	80	77
5th percentile	56	48	47
90% range	34	33	30
Median	74	63	61
Tukey's *p* value	-	< 0.001	< 0.001

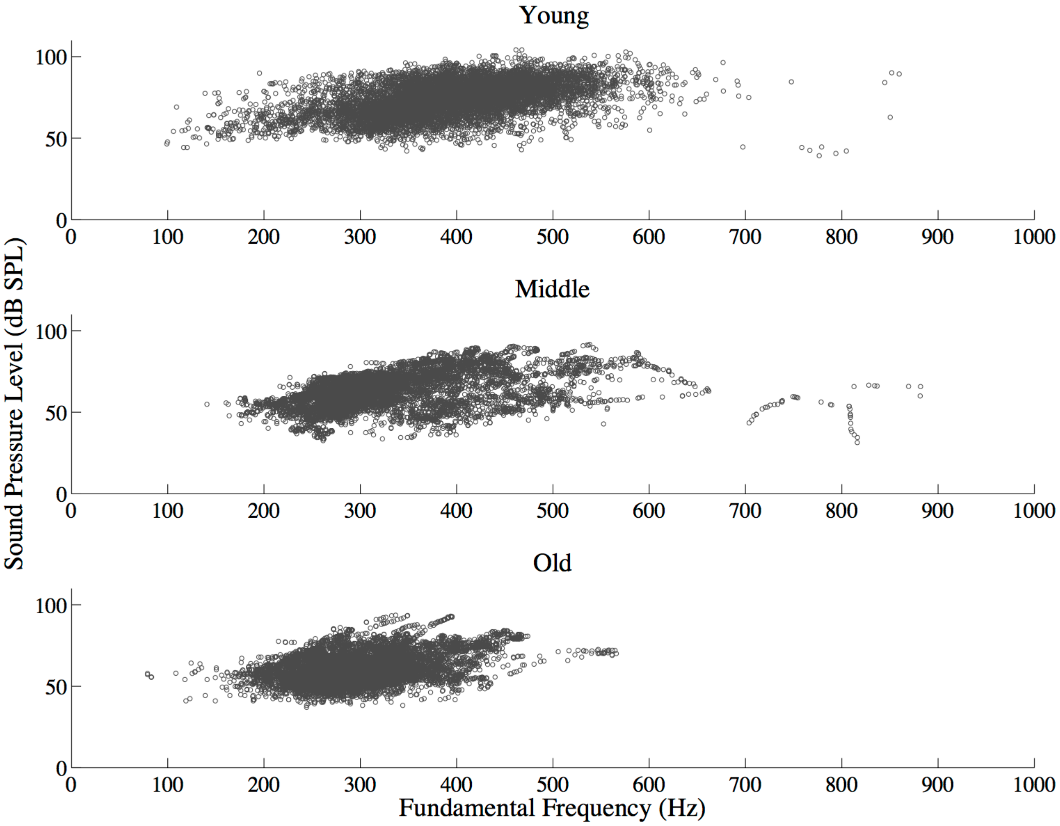


Figure 10. Distributions of 10,000 cases of F_0 and SPL values of modal segments randomly selected from age group.

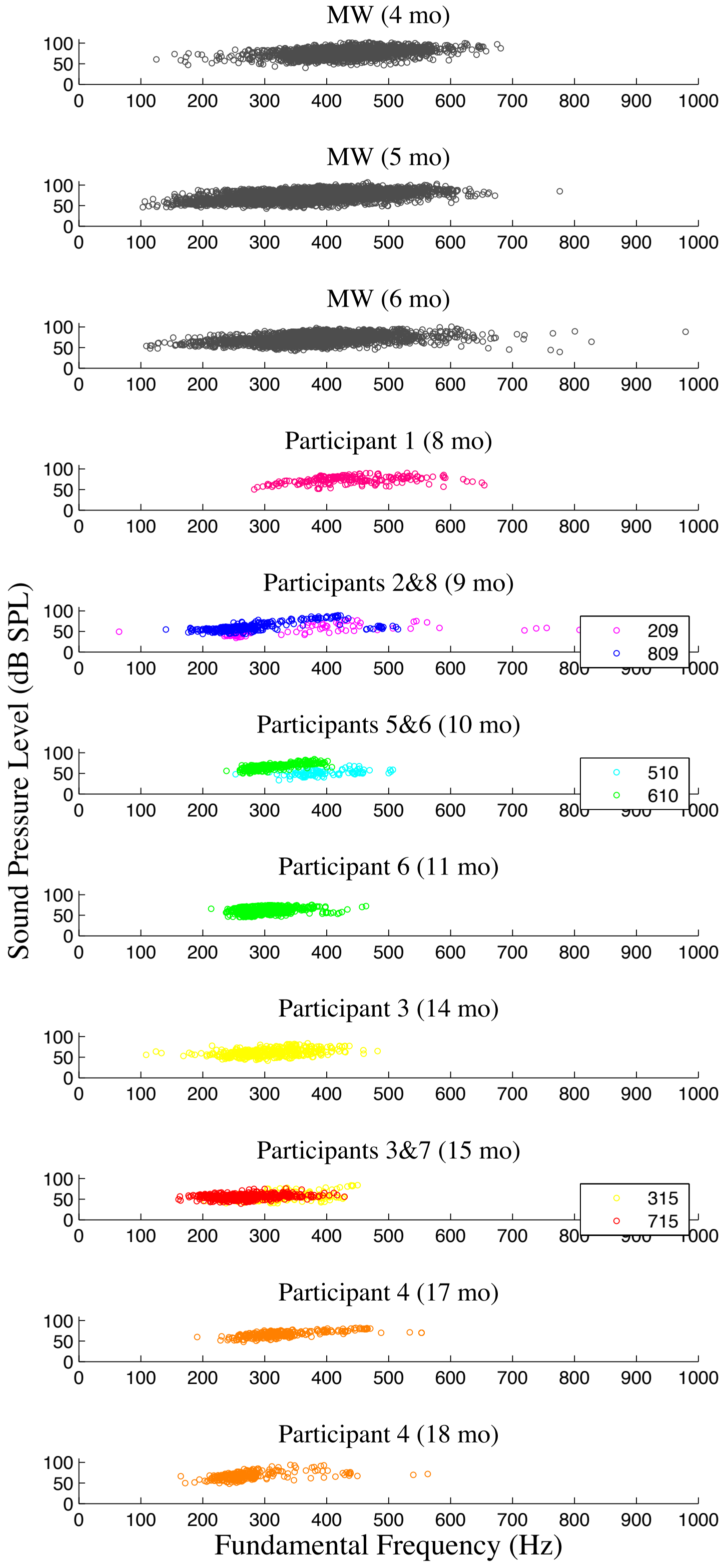


Figure 11. Intensity-frequency plots of modal segments for all participants by months. A different color is assigned to each participant.

Table 22

The Number and Mean Duration of Modal Segments Represented in Each Panel of Scatterplot in Figure 11

Age (mo)	Participant	Number of modal segments	Mean duration (ms)
4	MW	377	342
5	MW	493	485
6	MW	357	537
8	1	32	271
9	2	14	305
	8	21	698
10	5	12	325
	6	27	567
11	6	54	540
14	3	80	270
15	3	41	273
	7	56	402
17	4	26	405
18	4	28	423

Both F_0 and SPL ranges were greatest in the early infancy group and smallest in the late infancy group. Both of these are reflected in the intensity-frequency plots in which distributions become shifted lower in both F_0 and SPL and more tightly clustered with increasing age. The difference in F_0 ranges between the early and middle infancy groups was not as prominent as the difference between the early and late infancy group. It was suspected that individual variability contributed to a wider range of distribution in the middle infancy group. Intensity-frequency plots of modal were computed for the all participants by months (Figure 11). Figure 11 illustrates individual differences in intensity-frequency distributions among the participants, especially participants 2 and 9. Multiple participants in the middle infancy group formed a collective range that is slightly smaller than the range produced by MW alone. Although the number of modal segments varies between participants, none of the older participants produced a range as great as MW.

A median expiratory slope was smallest for the early infancy group, almost half the size of medians for the two older groups. A one-way ANOVA revealed a statistically significant effect of age on slope, F (2, 1806) = 58.36, $p < 0.001$. Post-hoc comparisons revealed statistically significant differences between the young and the two older groups (Tables 22), indicating that the mean expiratory slope was statistically significantly flatter in the early infancy group during modal phonation.

Table 23

Medians and Ranges of Modal Slope and Results of Post-hoc Comparisons

	Age		
	Early 4-6 months	Middle 8-11 months	Late 14-18 months
95th percentile	2.951	7.710	4.337
5th percentile	-0.215	0.151	-0.072
90% range	3.166	7.559	4.408
Median	0.686	1.122	1.246
Tukey's *p* value	-	< 0.001	< 0.001

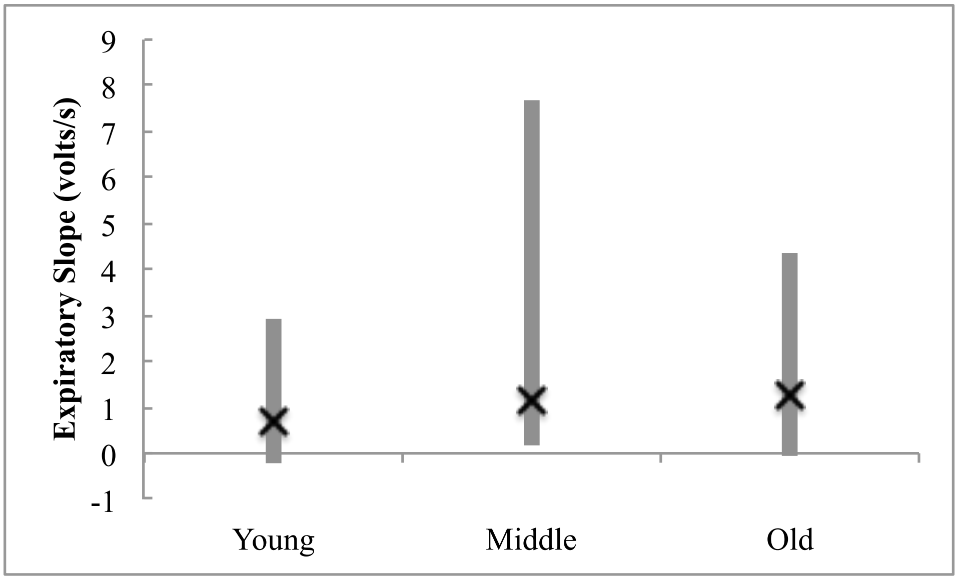


Figure 12. Middle 90% ranges of expiratory slope for each age group.

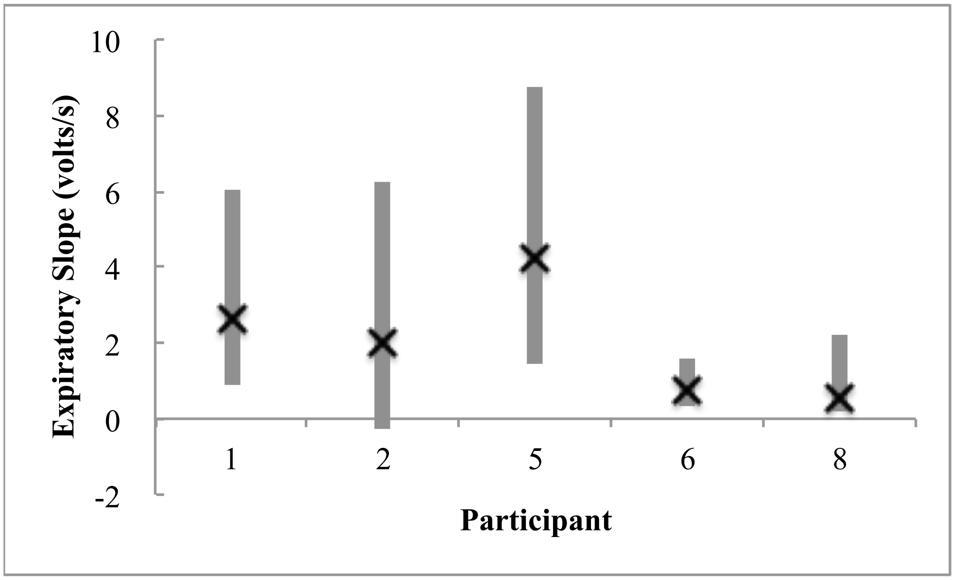


Figure 13. Middle 90% ranges of expiratory slope for each participant in the middle infancy group.

A middle 90% range was greatest for the middle age group, followed by the late infancy group, indicating that slope values used by the older groups were more variable compared to the early infancy group. As was the case for the F_0 ranges, individual variability was explored by plotting slope ranges for each participant in the middle infancy group (Figure 13). Three participants (1, 2, & 5) used a relatively large range, while two participants (6 & 8) used a small range that was also small in magnitude, showing individual differences in the use of expiratory slope values in the production of modal phonation. This suggests that collapsing slope ranges of multiple participants into a single age group may not be appropriate.

Chapter 6: Discussion

Range of Modal Phonation with Respect to Other Phonation Types

Regime categories in Dataset II readily distributed into different regions of the intensity-frequency space. By definition, the three basic registers of pulse, modal, and loft correspond to low, mid, and high pitch, respectively. In the present study, modal occupied a wide area along the mid-range of both frequency (200-600 Hz) and SPL (50-90 dB SPL). Pulse clustered in an area that was lower in both frequency (below 300 Hz) and SPL (below 80 dB SPL); loft in an area that was higher in both frequency (above 400 Hz) and SPL (above 60 dB SPL).

The upper end of pulse range is much higher in MW compared to pulse ranges reported for adult males (7-78 Hz) and females (2-78 Hz), who showed no gender difference (Hollien & Michel, 1968). In contrast to the wide ranges of modal and loft regimes adults can produce (about three octaves), their pulse ranges are small. By noting a fairly stable crossover frequency for the pulse register, Titze (2000) proposed a temporal gap transition for the pulse-modal transition, explaining that the auditory system perceives glottal pulses individually below a certain frequency (around 70 Hz) and continuously above it. The present findings indicate that a crossover frequency can be higher in infant phonation, suggesting that the pulse-modal transition is not entirely perceptual and production-based.

The ability of infants to produce high-F_0 pulse may be related to the unique histology of their vocal folds. In the pulse regime, activities of both the vocalis and cricothyroid muscles are minimum, making the vocal folds short and all layers slack (Hirano, 1982). This situation is similar to the infant vocal folds that lack a layered structure. Although cellular differentiation initiates in the end of the first year, the maturation is not complete till adolescence, leaving the entire lamina propria of infant a uniform structure (Hartnick, Rehbar, & Prasad, 2005; Sato et al., 2001). A single undifferentiated layer of infant vocal folds resembles lax layers of adult vocal folds in pulse. In adults, however, an increase in stiffness and tension with rising F_0 may force the vocal folds to transition to a different vibratory regime, while it can be speculated that the absence of the vocal ligament enables infants to stay in pulse for a wider range of $F_{0.}$

SH and BIPH were found in the region where the upper end of modal and the lower end of loft meet, but the distribution of SH was more restricted than BIPH. The difference in sizes of SH and BIPH ranges may simply reflect a fewer number of observations for SH or indicate more strict conditions for SH to occur. It has been reported in the literature that bifurcations frequently occur in the regions of regime overlap. Voice scientists have explained that because vocal folds are a nonlinear system, they exhibit regions where different vibratory regimes overlap due to hysteresis. The system is instable in the region of regime overlap because there is more than one patterns of output vocal fold vibration for a single input configuration. In such a region, a spontaneous change in vocal fold vibration can occur without changes in laryngeal muscle adjustments.

A slightly different mechanism for SH and BIPH is due to small changes in parameters such as subglottal pressure, vocal fold symmetry, and vocal fold tension (Berry et al., 1996; Švec et al., 1999). The F_0 and SPL ranges obtained for SH and BIPH in this study confirm their occurrence in the modal-loft transition. However, time information was disregarded in this study. The present findings can be strengthened by additional sequence analysis that can show exact progression of regime sequences as well as occurrences of BIPH and SH during regime transition.

The idea of “region of instabilities raises the question of whether bifurcations occur similarly in the frequency region where pulse and modal overlap. The presence of SH was reported at the pulse-modal transition in experiments with excised canine larynges (Berry et al., 1996). The current findings suggested rare occurrences of SH and BIPH in such a region in human infants. Whether this is due to the scarcity of observations or bifurcations are not typically associated in this region remains to be seen.

Range of Modal Phonation with Respect to Age

Both SPL and F_0 ranges of modal phonation shifted lower with age. For SPL, even though the upper and lower end values of SPL ranges decreased by about 10 dB SPL, ranges remained at about 30 dB for all age groups. F_0 ranges of modal showed a tendency to both decrease and shift lower with age. End values of F_0 range decreased by 50-100 Hz for late infancy group compared to MW, but such change was not observed between MW and the middle infancy group. When modal ranges were examined for each of all participants in both studies, individual variations in distributions were observed. Especially for the middle infancy groups, the upper and lower ends of the collective range were contributed by different participants. Despite differences in F_0 minima and maxima across infants, the majority of modal F_0 occurred in a 200 Hz range for all of the older infants. MW’s distribution, spanning about 500 Hz, was wider than any of the older infants.

Shifting of F_0 and SPL ranges is consistent with a lower means of F_0 and SPL for the late infancy groups. Medians of modal F_0 for the three ages were 391 Hz, 324 Hz, and

292 Hz for the early, middle, and late infancy group, respectively. Medians of modal SPL for the three ages were 74 dB SPL, 63 dB SPL, and 61 dB SPL for the early, middle, and late infancy group, respectively. Statistically significant mean difference between MW and the two older groups found in this study indicate a decreasing trend in both F_0 and SPL with age.

The current findings also tentatively suggest a decreasing trend in modal range of F_0 with age. This finding does not agree with Keating and Buhr (1978), who found no trend in F_0 ranges with age between 7-39 months. One reason for the disagreement may be because Keating and Buhr (1978) used a total range instead of 90% range. Some of the high F_0 values found in upper limits of their ranges only occurred once or twice in the sample. Outliers can skew ranges and distort the true nature of distributions. Because human errors in calculating F_0 is always a possibility, and also unusually high F_0s, "momentary slips of the larynx" (Baken & Orlikoff, 2000), can occur that are unrepresentative of habitual F_0, outliers must be carefully screened or measures need to be resistant to these values.

In speculating about possible explanations for this observed change in F_0 range, one possibility is that infants may be learning to avoid certain regions of phonatory instabilities. By producing different types of phonation that vary in F_0, SPL, and voice quality, infants explore different regions of the intensity-frequency space. By doing so, they may be testing their phonatory apparatus. As far as a form is concerned, early vocal development culminates in a well-formed syllable that is a basic building block of speech. Syllable nuclei are supported by controlled production of modal phonation that can flexibly vary in F_0 and SPL. The current findings suggest an intensity-frequency distribution narrows with time, perhaps as infants begin participating more in other-oriented communication in the late babbling or early word production stage.

Middle 90% ranges of F_0 obtained for the infants were about 1 octave (i.e., 10-13 semitones) in this study. This is smaller than maximum phonational ranges for adults or children of ages 0-16 years, but greater than normal variation found in adults. The maximum phonational ranges for adults (Hollien & Michel, 1968) or children of ages 0-16 years (van Oordt & Drost, 1963), were reported to be 2.5-3 octaves regardless of age. On the other hand, a 90% range of speaking fundamental frequency for adults was reported to be 6 semitones (Coleman & Markham, 1991). Using a modal F_0 range of 262-1565 Hz in cries (Michelsson, 2002), which is 2.6 octaves or 31 semitones, is can be speculated that the maximum phonation ranges remain rather constant throughout development, whereas a habitual range of F_0 is much greater in infants than in adults. An immature larynx of infants can produce a wide range of F_0, which may be related to the anatomical finding that the cricothyroid muscle is the most massive muscle in the intrinsic laryngeal muscles in infants (Kahane & Kahn, 1984) since short-term variations in F_0 are regulated greatly by the vocal fold tension.

Relationship between Regime and Mean Values of Acoustic and Respiratory Variables

There were statistically significant differences in mean F_0 between regimes in both Dataset I and II. Compared to that for modal, a mean F_0 was lower for pulse but higher for loft, SH, and BIPH. The relationship found between pulse, modal, and loft is consistent with what previous studies found among adults and children (Hollien, 1974; Keating & Buhr, 1978). The current study also found that values of mean F_0 for SH and

BIPH are comparable to what were reported in Robb and Saxman (1988), and that those mean F_0 values distinguish themselves from other regimes.

The effect of regime on other variables was less clear in the literature. A mean SPL was found to be smaller in pulse than in modal in both datasets, as expected from Hollien's (1974) findings with adult data. A mean expiratory slope was found to be flatter for pulse than in modal in Dataset II, as expected from adult findings in Blomgren et al. (1998) and Murry (1971). Additional differences in mean SPL were also observed between modal and BIPH in both datasets, whose numerical data have not been reported in the literature.

The current finding of flatter slopes for pulse is in agreement with Murry's (1971) interpretation that relates lower airflow to a shorter open time that results from the vocal fold vibration associated with pulse (Murry, 1971). Similarly, a particular pattern of vocal fold vibration involved in BIPH can be responsible for a higher relative airflow. Mergell and Herzel (1997) specifies conditions of biphonation to be large airflow, incomplete glottal closure, and coincident of F_0 with a formant frequency. This observation is important in explaining the need for large airflow in BIPH as well as its association with high F_0 found above. Following Mergell and Herzel (1997), it can be interpreted that large airflow is required to drive independently vibrating vocal folds. Because the two frequencies are not integer-related, the vocal folds do not meet, creating incomplete closure, which also allows for greater airflow.

The same analogy can be applied to the higher expiratory slope ranges associated with loft and chaos in Dataset II. Slope ranges for loft and chaos involved higher values compared to slope ranges for pulse, modal, and SH, which were concentrated in lower ranges. Higher values of slope can be explained by the presence of incomplete closure associated with the loft regime (Berry et al., 1996), and irregular vibration associated with chaotic phonation.

Effect of Regime on Relationships between Acoustic and Respiratory Variables

The pattern of correlations between the acoustic (F_0 and SPL) and respiratory (slope and lung volume) was complex. Nonetheless, the fact that no uniform relationships between the acoustic and respiratory variables were observed illustrates that their relationships are in fact regime specific.

In adult data, airflow is known to be a regulator of intensity at high frequencies where the glottal resistance cannot be increased easily. However, a positive correlation was not observed between SPL and either slope or lung volume for loft. The only correlation that was consistent with predictions was a positive correlation between F_0 and slope for modal. The mild correlation between them is similar to what was reported in Reilly (2004). He reasoned that changes in subglottal pressure may contribute to F_0 variability in infants who lack vocal ligament which is important in transmitting muscular tensions to the vibrating surface of the vocal folds. This finding is somewhat puzzling because no correlation was found between SPL and slope for the modal regime, which should have been the case if this relationship results from frequency-amplitude dependence via amplitude-dependence of tension in the vocal fold explained by Titze (1989).

Some of the significant correlations found were inconsistent and others were opposite of what was expected. The moderate positive correlation found between SPL and slope for SH indicated an increase in mean SPL was associated with a higher airflow.

However, the absence of similar relationships for other regimes, especially for modal that has a sufficient sample size, suggest that mean values are too global a measure and some of the important details in these variables may have been lost by taking an average value of SPL and slope for the whole segment.

A significant correlation was found between F_0 and relative lung volume level (RLV) for loft. A negative correlation suggested an increase in mean F_0 was related to a lower level of lung volume at regime initiation. The direction of correlation was opposite of the prediction. It is possible that, especially given a small sample size for loft ($n = 18$), a few outliers were driving the effect. There may be different types of high-F_0 phonation that were collapsed into loft. For example, squeak phonation characterized with a high F_0 and a low SPL may need to be excluded from loft. The fact that this variable did not yield significant results in mean comparisons, and that the pattern of correlations was dissimilar to the pattern of correlations found for slope, suggest that RLV and slope variables may not be converging on the same construct as proposed. This variable was conceptually designed to represent a point along inspiratory reserve volume, with the assumption that the rate of airflow may be greater when lung pressure is high. However, RLV findings were inconsistent with both slope findings and predictions. The slope findings converged on a pattern that was interpretable against SPL findings and also the literature on airflow. The validity of the relative volume level variable may need to be called into question.

Reliability

In general, reliability was high. Both inter-rater and intra-rater reliability of regime coding showed that coders were able to locate voiced segments with similar precision. There was no particular regime that was consistently missed. Rather, segments were missed due to a degraded signal quality or due to individual differences in marking an onset or offset segment. Segments were categorized into six mutually exclusive regime categories. Three of them belonged to normal regimes characterized with tonal phonation with a single pitch (modal, pulse, and loft), and three to non-normal characterized with either ambiguous pitches (SH and BIPH) or aperiodicity (chaos). Regime assignment showed varying degrees of agreement that resulted from difficulty in identifying certain regimes or individual differences in regime judgments. In other words, the three normal phonation types had much higher agreement than the non-normal phonation types. Regime confusions were also observed frequently between SH and BIPH, chaos and BIPH, and modal and loft, which occurred in both between coders and within a coder. Limited reliability found for the non-normal regimes suggest that results involving these regimes need to be interpreted with caution.

Difficulty in identifying the non-normal phonation types explains the lower regime assignment accuracy found between coders who coded Dataset II that contained a higher number of the non-modal phonation types. Although judgments of some regimes were found to be more difficult to mutually agree on, higher agreement in regime assignment found within a coder indicates that coders can be trained to discern certain phonatory characteristics reliably. Certain regime confusions also indicate a need for coder training that emphasizes the contrasts between these regimes. Such training may need to be tailored to a specific coder pair when a particular disagreement pattern emerges. In addition, the use of other tools such as a power spectrum may help categorizing certain regimes. For example, relative amplitudes of first two harmonics may help distinguish loft from modal, or an integer relationship of F_0s for SH may be more readily visible on a spectrum than on a spectrogram.

Reliability of acoustic and respiratory analysis was assessed only within a coder. No difference between repeated measurements of initial- and end-voltage levels ensures automaticity of the procedure. F_0 and RMS statistics derived from repeated traces had high positive correlations and minute mean differences overall. However, large differences (> 100 Hz) in values were observed in F_0 statistics in 7 % of the repeated tokens, which is not negligible.

Large differences in F_0 statistics result typically from tracking of improper F_0 in cases of SH and BIPH, or from a few incorrect extreme values that went undetected during editing. Although great care needs to be taken when inspecting F_0 traces, it is impossible and too labor intense to examine every value of F_0. Since human errors are unavoidable, the use of median and 90% range are encouraged in future studies involving the same method of F_0 extraction because they are less susceptible to outliers than mean and range.

Limitations and Future Directions

Modal ranges showed a decreasing trend with age. A younger infant in the expansion stage of vocal development tended to have a wider habitual range of F_0 than infants who were in the canonical stage. However, developmental comparisons conducted with ANOVA lacks generalizability because of a case study format. Given individual and session variability found in distribution of modal phonation, especially in $F_{0,}$ future studies will benefit from longitudinal data collection from multiple participants, desirably from the expansion stage through the post-canonical stage.

The younger infants produced more non-modal phonation types that had distinct F_0-SPL ranges. In particular, two types of bifurcation regimes, BIPH and SH, were found to occur in the region where modal and loft overlapped. Ranges of non-modal regimes could not be compared across age groups due to small numbers of observation of these regimes in older infants, but developmental information on such ranges will help further understand the development of modal phonation.

The effect of regime was demonstrated most clearly in differences in mean values of F_0. Other findings were more robust in Dataset II which contained greater numbers of non-modal phonation types. Failure for mean differences of SPL to reach significance in Dataset I may be due to small sample sizes in non-modal segments. Alternatively, it may be that significant mean SPL differences between modal and all other regimes in Dataset II may be driven by large sample sizes.

A possible source of discrepancy in slope findings between the two datasets is the difference in body positions. Older participants in Dataset I were seated upright. The younger participant in Dataset II was supine. A difference in body position can alter the respiratory mechanics because of different ways gravity acts on the body. According to Hixon, Weismer, and Hoit (2008), in the upright position, the effect of gravity is expiratory on the rib cage but inspiratory on the abdominal wall. In the supine position, on the other hand, the effect of gravity is expiratory on both, which raises the relaxation pressure higher and lowers the resting level of the respiratory system. This implication is that rate of airflow can be greater without compensatory inspiratory effort. Such difference may be a confounding factor affecting the results.

Another consideration regarding a possible difference in respiratory behaviors between younger and older infants is that the difference in relative contribution of ribcage and abdomen. Young infants mostly use their abdomen while the ribcage participate more in breathing as the infant grows (Hershenson et al., 1990). Since the abdominal cavity is a space filled with organs and tissues other than the diaphragm, there is likely a complex relationship between the traceable abdominal movement and the actual breathing movement effected by the diaphragm, whereas the relationship between the ribcage movement and resulting breathing movement is more direct. This could mean that respiratory movements in younger infants may contain more noise in the signal compared to older infants who use their ribcage more.

Although some differences were found in patterns of correlations among different types of phonation, no clear relationships between respiratory drive and output acoustics emerged in the present study. Contributing factors include small sample sizes in non-modal phonation types and the use of average values. The problem in using average values is that they do not faithfully represent changing values of F_0, SPL, and slope within a segment. Further analysis can be made by deriving instantaneous velocity of respiratory excursion, instead of slope, from the current datasets in addition to time-locked values of F_0 and SPL in order to capture more details of these variables within a segment.

The two datasets contained different sets of data points. In Dataset I, each segment was associated with a single mean value per variable. In Dataset II, each segment had a set of finely sampled F_0 and SPL values. While the former suffered from small sample sizes, the latter had too many data points that were redundant and could detect significance in small effects. In the current study, data points were simply downsampled to ensure five samples in the shortest segment, but other data reduction methods should be explored.

Finally, in order to further explore development of phonation and strengthen current findings, a more balanced design across pre- through post-canonical stages is warranted. More observations of modal regimes from multiple infants are needed either to improve the generalizability of a decreasing trend, or to understand patterns of individual variability in development of F_0 and SPL ranges. More observations of non-modal phonation types are also needed from various ages. In order to clarify how preference toward modal phonation emerges, developmental trajectories of non-modal regimes must be incorporated.

References

Awan, S. N., & Mueller, P. B. (1996). Speaking fundamental frequency characteristics of White, African American, and Hispanic kindergartners. *Journal of Speech and Hearing Research*, *39*, 573–577.

Bachorowski, J. A., Smoski, M. J., & Owren, M. J. (2001). The acoustic features of human laughter. *Journal of the Acoustical Society of America*, *110*, 1581.

Baken, R. J., & Orlikoff, R. F. (2000). In *Clinical measurement of speech and voice* (p. 170). Clifton Park, NY: Thompson Learning.

Berry, D. A. (2001). Mechanisms of modal and nonmodal phonation. *Journal of Phonetics*, *29*, 431–450.

Berry, D. A., Herzel, H., Titze, I. R., & Story, B. H. (1996). Bifurcations in excised larynx experiments. *Journal of Voice*, *10*, 129–138.

Blake, J., & Boysson-Bardies, B. D. (1992). Patterns in babbling: a cross-linguistic study. *Journal of Child Language*, *19*, 51–74.

Blomgren, M., Chen, Y., Ng, M. L., & Gilbert, H. R. (1998). Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *Journal of the Acoustical Society of America*, *103*, 2649–2658.

Boliek, C. A., Hixon, T. J., Watson, P. J., & Morgan, W. J. (1996). Vocalization and breathing during the first year of life. *Journal of Voice*, *10*, 1–22.

Bosma, J. F. (1975). Anatomic and physiologic development of the speech apparatus. In D. B. Towers (Ed.), *Human communication and its disorders* (Vol. 3, pp. 469–481). New York: Raven.

Bosma, J. F. (1976). Discussion of the paper, “Postnatal development of the basicranium and vocal tract in man” by J. T. Laitman and E. S. Crelin. In J. F. Bosma (Ed.), *Symposium on Development of the Basicranium* (pp. 219–220). Bethesda, MD: DHEW Publication.

Boysson-Bardies, B. D., Sagart, L., & Bacri, N. (1981). Phonetic analysis of late babbling: a case study of a French child. *Journal of Child Language*, *8*, 511–524.

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

Buder, E. H., & Oller, D. K. (2004). Parametric representations for the acoustic study of infant vocalization. Presented at the International Conference on Infant Studies, Chicago, IL.

Buder, E. H., Oller, D. K., & Magoon, J. C. (2003). Vocal intensity and phonatory regimes in the development of infant protophones. In *Proceedings of the XVth International Congress of Phonetic Sciences, Causal Productions, Adelaide, Australia*.

Bunn, J. C., & Mead, J. (1971). Control of ventilation during speech. *Journal of Applied Physiology*, *31*, 870–872.

Chan, R. W., Gray, S. D., & Titze, I. R. (2001). The importance of hyaluronic acid in vocal fold biomechanics. *Otolaryngology-Head and Neck Surgery*, *124*, 607–614.

Chapman, K. L., Hardin-Jones, M., Schulte, J., & Halter, K. A. (2001). Vocal development of 9-month-old babies with cleft palate. *Journal of Speech, Language, and Hearing Research*, *44*, 1268–1283.

Chen, H. P., & Irwin, O. C. (1946). Infant speech vowel and consonant types. *Journal of Speech and Hearing Disorders*, *11*, 27–29.

Coleman, R. F., & Markham, I. W. (1991). Normal variations in habitual pitch. *Journal of Voice*, *5*(2), 173–177.

Colton, R. H., & Hollien, H. (1972). Phonational range in the modal and falsetto registers. *Journal of Speech and Hearing Research*, *15*, 708–713.

Connaghan, K. P., Moore, C. A., & Higashakawa, M. (2004). Respiratory kinematics during vocalization and nonspeech respiration in children from 9 to 48 months. *Journal of Speech, Language, and Hearing Research*, *47*, 70.

Crelin, E. S. (1973). *Functional anatomy of the newborn*. New Haven, NJ: Yale University Press.

Crelin, E. S. (1987). *The human vocal tract: Anatomy, function, development, and evolution*. Vantage Press.

Delgado, R. E. (2008). *AACT - Action Analysis Coding and Training Software*. Miami, FL: Intelligent Hearing Systems Corp.

Di Eugenio, B. (2000). On the usage of Kappa to evaluate agreement on coding tasks. In *Proceedings of LREC* (Vol. 1, pp. 441–444). Retrieved from http://www.nlp.cs.uic.edu/PS-papers/lrec00.pdf

Dodd, B. J. (1972). Comparison of babbling patterns in normal and Down-syndrome infants. *Journal of Intellectual Disability Research*, *16*, 35–40.

Eguchi, S., & Hirsh, I. J. (1969). Development of speech sounds in children. *Acta Oto-Laryngologica. Supplementum*, *257*, 1–51.

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

Eilers, R. E., Oller, D. K., Levine, S., Basinger, D., Lynch, M. P., & Urbano, R. (1993). The role of prematurity and socioeconomic status in the onset of canonical babbling in infants. *Infant Behavior and Development*, *16*, 297–315.

Fairbanks, G. (1942). An acoustical study of the pitch of infant hunger wails. *Child Development*, *13*, 227–232.

Finnegan, E. M., Luschei, E. S., & Hoffman, H. T. (2000). Modulations in respiratory and laryngeal activity associated with changes in vocal intensity during speech. *Journal of Speech, Language, and Hearing Research*, *43*, 934–950.

Fitch, W., Neubauer, J., & Herzel, H. (2002). Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour*, *63*, 407–418.

Fry, D. B. (1966). The development of the phonological system in the normal and the deaf child. In F. Miller & G. A. Miller (Eds.), *The Genesis of language; a psycholinguistic approach: Proceedings of a conference on" language development in children,"* (pp. 187–207). Cambridge, MA: The MIT Press.

Gardosik, T. A., Ross, P. J., & Singh, S. (1980). Acoustic characteristics of the first cries of infants. In T. Murry & J. Murry (Eds.), *Infant Communication: Cry and Early Speech* (pp. 106–123). Houston, TX: College-Hill Press.

Gelfer, M. P., & Young, S. R. (1997). Comparisons of intensity measures and their stability in male and female speakers. *Journal of Voice*, *11*, 178–186.

Gilbert, H. R., & Robb, M. P. (1996). Vocal fundamental frequency characteristics of infant hunger cries: birth to 12 months. *International Journal of Pediatric Otorhinolaryngology*, *34*, 237–243.

Giovanni, A., Ouaknine, M., Guelfucci, B., Yu, P., Zanaret, M., & Triglia, J. M. (1999). Nonlinear behavior of vocal fold vibration: the role of coupling between the vocal folds. *Journal of Voice*, *13*, 465–476.

Greenberg, J. H. (1965). Some generalizations concerning initial and final consonant sequences. *Linguistics*, *3*, 5–34.

Hanson, H. M. (1997). Glottal characteristics of female speakers. *Journal of the Acoustical Society of America*, *101*, 466–481.

Hartnick, C. J., Rehbar, R., & Prasad, V. (2005). Development and maturation of the pediatric human vocal fold lamina propria. *The Laryngoscope*, *115*, 4–15.

Hershenson, M. B., Colin, A. A., Wohl, M. E., & Stark, A. R. (1990). Changes in the contribution of the rib cage to tidal breathing during infancy. *The American Review of Respiratory Disease*, *141*, 922–925.

Herzel, H., Berry, D. A., Titze, I. R., & Saleh, M. (1994). Analysis of vocal disorders with methods from nonlinear dynamics. *Journal of Speech and Hearing Research*, *37*, 1008–1019.

Herzel, H., & Reuter, R. (1997). Whistle register and biphonation in a child's voice. *Folia Phoniatrica et Logopaedica: Official Organ of the International Association of Logopedics and Phoniatrics (IALP)*, *49*, 216–224.

Hirano, M. (1974). Morphological structure of the vocal cord as a vibrator and its variations. *Folia Phoniatrica et Logopaedica*, *26*, 89–94.

Hirano, M. (1982). The role of the layer structure of the vocal fold in register control. In P. Hurme (Ed.), *Vox Humana: Studies Presented to Aato Sonninen on the Occasion of his Sixtieth Birthday* (pp. 50–62). Jyväskylä, Finland: University of Jyväskylä.

Hirano, M., Kurita, S., & Nakashima, T. (1983). Growth, development and aging of human vocal folds. In D. M. Bless & J. H. Abbs (Eds.), *Vocal fold physiology: Contemporary research and clinical issues* (pp. 22–43). San Diego, CA: College-Hill Press.

Hirano, M., & Sato, K. (1993). *Histological color atlas of the human larynx*. San Diego, CA: Singular.

Hirano, M., Sato, K., & Nakashima, T. (1999). Fibroblasts in human vocal fold mucosa. *Acta Oto-Laryngologica*, *119*, 271–276.

Hirschberg, J., Szende, T., Koltai, P. J., & Illényi, A. (Eds.). (2009). *Pediatric Airway: Cry, Stridor, and Cough*. San Diego, CA: Plural Publishing.

Hixon, T. J., Goldman, M. D., & Mead, J. (1973). Kinematics of the chest wall during speech production: Volume displacements of the rib cage, abdomen, and lung. *Journal of Speech and Hearing Research*, *16*, 78–115.

Hixon, T. J., Weismer, G., & Hoit, J. D. (2008). *Preclinical speech science: anatomy, physiology, acoustics, and perception*. San Diego, CA: Plural Publishing.

Hoit, J. D., Hixon, T. J., Watson, P. J., & Morgan, W. J. (1990). Speech breathing in children and adolescents. *Journal of Speech and Hearing Research*, *33*, 51–69.

Hollien, H. (1960). Vocal pitch variation related to changes in vocal fold length. *Journal of Speech and Hearing Research*, *3*, 150–156.

Hollien, H. (1962). Vocal fold thickness and fundamental frequency of phonation. *Journal of Speech and Hearing Research*, *5*, 237–243.

Hollien, H. (1974). On Vocal Registers. *Journal of Phonetics*, *2*, 125–143.

Hollien, H., & Michel, J. F. (1968). Vocal fry as a phonational register. *Journal of Speech and Hearing Research*, *11*, 600–604.

Hollien, H., & Shipp, T. (1972). Speaking fundamental frequency and chronologic age in males. *Journal of Speech and Hearing Research*, *15*, 155–159.

Irwin, O. C. (1946). Infant Speech Equations For Consonant-Vowel Ratios. *Journal of Speech and Hearing Disorders*, *11*, 177–180.

Irwin, O. C. (1947a). Development of speech during infancy: curve of phonemic frequencies. *Journal of Experimental Psychology*, *37*, 187–193.

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

Irwin, O. C. (1947c). Infant speech: Consonantal sounds according to place of articulation. *Journal of Speech and Hearing Disorders*, *12*, 397–401.

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

Irwin, O. C. (1952). Speech Development In The Young Child: 2. Some Factors Related To The Speech Development Of The Infant and Young Child. *Journal of Speech and Hearing Disorders*, *17*, 269–279.

Irwin, O. C., & Chen, H. P. (1943). Speech sound elements during the first year of life: A review of the literature. *Journal of Speech and Hearing Disorders*, *8*, 109–122.

Irwin, O. C., & Chen, H. P. (1946). Development of speech during infancy: curve of phonemic types. *Journal of Experimental Psychology*, *36*, 431–436.

Isaacson, G. (2003). The larynx, trachea, bronchi, lungs, and esophagus. In C. D. Bluestone (Ed.), *Pediatric otolaryngology* (Vol. 2, pp. 1361–1370). Philadelphia, PA: Elsevier Health Sciences.

Ishizaka, K., & Flanagan, J. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell System Technology Journal*, *51*, 1233–1268.

Isshiki, N. (1964). Regulatory mechanism of voice intensity variation. *Journal of Speech and Hearing Research*, *7*, 17–29.

Jiang, J. J., Zhang, Y., & Ford, C. N. (2003). Nonlinear dynamics of phonations in excised larynx experiments. *Journal of the Acoustical Society of America*, *114*, 2198–2205.

Jiang, J. J., Zhang, Y., & Stern, J. (2001). Modeling of chaotic vibrations in symmetric vocal folds. *Journal of the Acoustical Society of America*, *110*, 2120–2128.

Kahane, J. C. (1982). Growth of the human prepubertal and pubertal larynx. *Journal of Speech and Hearing Research*, *25*, 446–455.

Kahane, J. C. (1996). Lifespan changes in the larynx: an anatomical perspective. In W. S. Brown, B. P. Vinson, & M. A. Crary (Eds.), *Organic voice disorders: assessment and treatment* (pp. 89–110). San Diego, CA: Singular.

Kahane, J. C., & Kahn, A. R. (1984). Weight measurements of infant and adult intrinsic laryngeal muscles. *Folia Phoniatrica et Logopaedica*, *36*, 129–133.

Keating, P. (1980). Patterns of fundamental frequency and vocal registers. In T. Murry & J. Murry (Eds.), *Infant Communication: Cry and Early Speech*. Houston, TX: College-Hill Press.

Keating, P., & Buhr, R. (1978). Fundamental frequency in the speech of infants and children. *Journal of the Acoustical Society of America*, *63*, 567–571.

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research*, *19*, 421–445.

Kent, R. D. (1992). The biology of phonological development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: models, research, implications* (pp. 65–90). Timonium, MD: York Press.

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *Journal of the Acoustical Society of America*, *72*, 353–365.

King, E. W. (1949). *A roentgenographic study of pharyngeal growth*. University of Illinois.

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

Kurita, S., Nagata, K., & Hirano, M. (1983). A comparative study of the layer structure of the vocal fold. In D. M. Bless & J. H. Abbs (Eds.), *Vocal fold physiology: Contemporary research and clinical issues* (pp. 3–21). San Diego, CA: College-Hill Press.

Ladefoged, P. (2006). *A course in phonetics* (5th ed.). Boston, MA: Wadsworth.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Malden, MA: Blackwell.

Laitman, J. T., & Crelin, E. S. (1976). Postnatal development of the basicranium and vocal tract region in man. In J. F. Bosma (Ed.), *Symposium on Development of the Basicranium* (pp. 206–219). Bethesda, MD: U.S. Dept. of Health, Education, and Welfare, Public Health Service, National Institutes of Health.

Laitman, J. T., & Reidenberg, J. S. (1997). The Human Aerodigestive Tract and Gastroesophageal Reflux: An Evolutionary Perspective. *The American Journal of Medicine*, *103*(Suppl 5A), 2–8.

Lalonde, B., & Boucher, V. J. (2009). *Breathing constraints on mean utterance length throughout life*. Poster Session presented at the American Speech and Hearing Association Convention, New Orleans, LA.

Langlois, A., Baken, R. J., & Wilder, D. N. (1980). Pre-speech respiratory behaviour during the first year of life. In T. Murry & J. Murry (Eds.), *Infant Communication Cry and Early Speech* (pp. 56–84). San Diego, CA: College-Hill.

Laufer, M. Z., & Horii, Y. (1977). Fundamental frequency characteristics of infant non-distress vocalization during the first twenty-four weeks. *Journal of Child Language*, *4*, 171–184.

Lester, B. M. (1978). The organization of crying in the neonate. *Journal of Pediatric Psychology*, *3*, 122–130.

Levin, K. (1999). Babbling in infants with cerebral palsy. *Clinical Linguistics & Phonetics*, *13*, 249–267.

Levitt, A. G., & Aydelott Utman, J. G. (1992). From babbling towards the sound systems of English and French: a longitudinal two-case study. *Journal of Child Language*, *19*, 19–49.

Lieberman, P. (1967). Intonation, perception, and language. *MIT Research Monograph*, *38*. Retrieved from http://psycnet.apa.org/psycinfo/1967-16665-001

Lieberman, P., Crelin, E. S., & Klatt, D. H. (1972). Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee. *American Anthropologist*, *74*, 287–307.

Lind, K., & Wermke, K. (2002). Development of the vocal fundamental frequency of spontaneous cries during the first 3 months. *International Journal of Pediatric Otorhinolaryngology*, *64*, 97–104.

Lynch, M. P., Oller, D. K., & Steffens, M. (1989). Development of speech-like vocalizations in a child with congenital absence of cochleas: The case of total deafness. *Applied Psycholinguistics*, *10*, 315–333.

Lynip, A. W. (1951). The use of magnetic devices in the collection and analysis of the preverbal utterances of an infant. *Genetic Psychology Monographs*, *44*, 221–262.

Marchman, V. A., Miller, R., & Bates, E. A. (1991). Babble and first words in children with focal brain injury. *Applied Psycholinguistics*, *12*, 1–22.

Martinot-Lagarde, P., Sartene, R., Mathieu, M., & Durand, G. (1988). What does inductance plethysmography really measure? *Journal of Applied Physiology*, *64*, 1749–1756.

McFarland, D. H. (2001). Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research*, *44*, 128–143.

McGlone, R. E. (1967). Air flow during vocal fry phonation. *Journal of Speech, Language and Hearing Research*, *10*, 299–304.

McGlone, R. E. (1970). Air flow in the upper register. *Folia Phoniatrica et Logopaedica*, *22*, 231–238.

McGlone, R. E., & Shipp, T. (1971a). Some physiologic correlates of vocal-fry phonation. *Journal of Speech, Language and Hearing Research*, *14*(4), 769.

McGlone, R. E., & Shipp, T. (1971b). Some physiologic correlates of vocal-fry phonation. *Journal of Speech and Hearing Research*, *14*, 769–775.

Mergell, P., & Herzel, H. (1997). Modelling biphonation–The role of the vocal tract. *Speech Communication*, *22*, 141–154.

Mergell, P., Herzel, H., & Titze, I. R. (2000). Irregular vocal-fold vibration—High-speed observation and modeling. *Journal of the Acoustical Society of America*, *108*, 2996–3002.

Michelsson, K. (1971). Cry analyses of symptomless low birth weight neonates and of asphyxiated newborn infants. *Acta Paediatrica Scandinavica*, *60*(S216), 9–45.

Michelsson, K. (1986). Cry analysis in clinical neonatal diagnosis. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of early speech* (pp. 67–77). New York: Stockton.

Michelsson, K., Eklund, K., Leppänen, P., & Lyytinen, H. (2002). Cry characteristics of 172 healthy 1-to 7-day-old infants. *Folia Phoniatrica et Logopaedica*, *54*, 190–200.

Michelsson, K., Järvenpää, A. L., & Rinne, A. (1983). Sound spectrographic analysis of pain cry in preterm infants. *Early Human Development*, *8*, 141–149.

Michelsson, K., Sirviö, P., & Wasz-Höckert, O. (1977). Pain cry in full-term asphyxiated newborn infants correlated with late findings. *Acta Paediatrica Scandinavica*, *66*, 611–616.

Michelsson, K., & Wasz-Höckert, O. (1980). The value of cry analysis in neonatology and early infancy. In T. Murry & J. Murry (Eds.), *Infant Communication: Cry and Early Speech* (pp. 152–182). Houston, TX: College-Hill Press.

Murai, F., Arai, T., & Kimura, T. (2005). 口蓋裂児の発声行動について: 口蓋形成術前と術後での発声行動の変化について [The Utterances of the Cleft Palate Children: the Comparison between before and after Palatal Surgery]. *Institute of Electronics, Information and Communication Engineers Technical Report*, *104*, 41–46.

Murai, J. (1963). The Sounds of Infants: Their Phonemization and Symbolization. *Studia Phonologica*, *3*, 17–34.

Murry, T. (1971). Subglottal pressure and air flow measures during vocal fry phonation. *Journal of Speech and Hearing Research*, *14*, 544–551.

Murry, T., Amundson, P., & Hollien, H. (1977). Acoustical characteristics of infant cries: Fundamental frequency. *Journal of Child Language*, *4*, 321–328.

Nakazima, S. (1962). A comparative study of the speech developments of Japanese and American English in childhood (1): A comparison of the developments of voices at the prelinguistic period. *Studia Phonologica*, *2*, 27–46.

Negus, V. E. (1929). *The mechanism of the larynx*. St. Louis, MO: Mosby.

Negus, V. E. (1949). *The comparative anatomy and physiology of the larynx*. London: Heinemann.

Nishimura, T. (2003). Comparative morphology of the hyo-laryngeal complex in anthropoids: two steps in the evolution of the descent of the larynx. *Primates*, *44*, 41–49.

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. *Child Phonology*, *1*, 93–112.

Oller, D. K. (1986). Metaphonology and infant vocalizations. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of early speech* (pp. 21–35). New York: Stockton.

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

Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences*, 6318–6323.

Oller, D. K., & Eilers, R. E. (1982). Similarity of babbling in Spanish-and English-learning babies. *Journal of Child Language*, *9*, 565–577.

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, *59*, 441–449.

Oller, D. K., Eilers, R. E., Neal, A. R., & Cobo-Lewis, A. B. (1998). Late onset canonical babbling: a possible early marker of abnormal development. *Journal Information*, *103*, 249–263.

Oller, D. K., Eilers, R. E., Neal, A. R., & Schwartz, H. K. (1999). Precursors to speech in infancy: the prediction of speech and language disorders. *Journal of Communication Disorders*, *32*, 223–246.

Oller, D. K., Eilers, R. E., Steffens, M. L., Lynch, M. P., & Urbano, R. (1994). Speech-like vocalizations in infancy: an evaluation of potential risk factors. *Journal of Child Language*, *21*, 33–58.

Oller, D. K., Eilers, R. E., Urbano, R., & Cobo-Lewis, A. B. (1997). Development of precursors to speech in infants exposed to two languages. *Journal of Child Language*, *24*, 407–425.

Oller, D. K., & Seibert, J. M. (1988). Babbling of Prelinguistic Mentally Retarded Children. *American Journal of Mental Retardation*, *92*, 369–375.

Oller, D. K., Wieman, L. A., Doyle, W. J., & Ross, C. (1976). Infant babbling and speech. *Journal of Child Language*, *3*, 1–11.

Orlikoff, R. F., & Baken, R. J. (1989). The effect of the heartbeat on vocal fundamental frequency perturbation. *Journal of Speech and Hearing Research*, *32*, 576–582.

Ostwald, P. F. (1963). *Soundmaking: the acoustic communication of emotion*. Springfield, IL: Charles C Thomas.

Parham, D. F. (2008). *A comparison of respiratory behavior for articulated and unarticulated utterances in infants* (Doctoral dissertaion). University of Memphis.

Pearsons, K. S., Bennett, R. L., Fidell, S., & Bolt, B. (1977). *Speech levels in various noise environments*. Washington, DC: Office of Health and Ecological Effects, Office of Research and Development, US EPA.

Prescott, R. (1975). Infant cry sound; developmental features. *Journal of the Acoustical Society of America*, *57*, 1186–1191.

Reilly, K. J. (2004). *Characteristics of Respiratory and Laryngeal Function During Vocal Productions in the First Year of Life*. University of Washington.

Reilly, K. J., & Moore, C. A. (2009). Respiratory Movement Patterns During Vocalizations at 7 and 11 Months of Age. *Journal of Speech, Language, and Hearing Research*, *52*, 223–239.

Ringel, R. L., & Kluppel, D. D. (1964). Neonatal crying: A normative study. *Folia Phoniatrica et Logopaedica*, *16*, 1–9.

Robb, M. P., & Saxman, J. H. (1985). Developmental trends in vocal fundamental frequency of young children. *Journal of Speech and Hearing Research*, *28*, 421–427.

Robb, M. P., & Saxman, J. H. (1988). Acoustic observations in young children's non-cry vocalizations. *Journal of the Acoustical Society of America*, *83*, 1876–1882.

Robb, M. P., Saxman, J. H., & Grant, A. A. (1989). Vocal fundamental frequency characteristics during the first two years of life. *Journal of the Acoustical Society of America*, *85*, 1708–1717.

Rothgänger, H. (2003). Analysis of the sounds of the child in the first year of age and a comparison to the language. *Early Human Development*, *75*, 55–69.

Sasaki, C. T., & Isaacson, G. (1988). Functional anatomy of the larynx. *Otolaryngologic Clinics of North America*, *21*, 595–612.

Sato, K. (1998). Reticular fibers in the vocal fold mucosa. *The Annals of Otology, Rhinology & Laryngology*, *107*, 1023–1028.

Sato, K., Hirano, M., & Nakashima, T. (2000). Comparative histology of the maculae flavae of the vocal folds. *The Annals of Otology, Rhinology & Laryngology*, *109*, 136–140.

Sato, K., Hirano, M., & Nakashima, T. (2001). Fine structure of the human newborn and infant vocal fold mucosae. *The Annals of Otology, Rhinology & Laryngology*, *110*, 417–424.

Sato, K., Nakashima, T., Nonaka, S., & Harabuchi, Y. (2008). Histopathologic investigations of the unphonated human vocal fold mucosa. *Acta Oto-Laryngologica*, *128*, 694–701.

Saxman, J. H., & Burk, K. W. (1967). Speaking fundamental frequency characteristics of middle-aged females. *Folia Phoniatrica et Logopaedica*, *19*, 167–172.

Scheiner, E., Hammerschmidt, K., Jürgens, U., & Zwirner, P. (2002). Acoustic analyses of developmental changes and emotional expression in the preverbal vocalizations of infants. *Journal of Voice*, *16*, 509–529.

Sheinkopf, S. J., Mundy, P., Oller, D. K., & Steffens, M. (2000). Vocal atypicalities of preverbal autistic children. *Journal of Autism and Developmental Disorders*, *30*, 345–354.

Sheppard, W. C., & Lane, H. L. (1968). Development of the prosodic features of infant vocalizing. *Journal of Speech and Hearing Research*, *11*, 94–108.

Shipp, T., & McGlone, R. E. (1971). Laryngeal dynamics associated with voice frequency change. *Journal of Speech and Hearing Research*, *14*, 761–768.

Sirviö, P., & Michelsson, K. (1976). Sound-spectrographic cry analysis of normal and abnormal newborn infants. *Folia Phoniatrica et Logopaedica*, *28*, 161–173.

Smith, B. L. (1982). Some observations concerning premeaningful vocalizations of hearing-impaired infants. *Journal of Speech and Hearing Disorders*, *47*, 439–414.

Smith, B. L., & Oller, D. K. (1981). A comparative study of pre-meaningful vocalizations produced by normally developing and Down's syndrome infants. *Journal of Speech and Hearing Disorders*, *46*, 46–51.

Sohner, L., & Mitchell, P. (1991). Phonatory and phonetic characteristics of prelinguistic vocal development in cri du chat syndrome. *Journal of Communication Disorders*, *24*, 13–20.

Sroufe, L. A., & Waters, E. (1976). The Ontogenesis of Smiling and Laughter: A Perspective on the Organization of. *Psychological Review*, *83*, 173–189.

Stark, R. E. (1978). Features of infant sounds: The emergence of cooing. *Journal of Child Language*, *5*, 379–390.

Stark, R. E. (1980). Stages of speech development in the first year of life. *Child Phonology*, *1*, 73–92.

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

Stark, R. E. (1983). Phonatory development in young normally hearing and hearing impaired children. In I. Hochberg, H. Levitt, & M. J. Osberger (Eds.), *Speech of the hearing-impaired: Research, training, and personnel preparation* (pp. 251–266). Baltimore, MD: University Park Press.

Stark, R. E., Bernstein, L. E., & Demorest, M. E. (1993). Vocal communication in the first 18 months of life. *Journal of Speech and Hearing Research*, *36*, 548–558.

Stark, R. E., Rose, S. N., & McLagen, M. (1975). Features of infant sounds: The first eight weeks of life. *Journal of Child Language*, *2*, 205–221.

Stathopoulos, E. T., & Sapienza, C. (1993). Respiratory and laryngeal function of women and men during vocal intensity variation. *Journal of Speech and Hearing Research*, *36*, 64–75.

Stevens, K. N. (1980). Acoustic correlates of some phonetic categories. *Journal of the Acoustical Society of America*, *68*, 836–842.

Stoel-Gammon, C. (2001). Down syndrome phonology: Developmental patterns and intervention strategies. *Down Syndrome Research and Practice*, *7*, 93–100.

Stoel-Gammon, C., & Otomo, K. (1986). Babbling development of hearing-impaired and normally hearing subjects. *Journal of Speech and Hearing Disorders*, *51*, 33–41.

Strik, H., & Boves, L. (1995). Downtrend in F0 and Psb. *Journal of Phonetics*, *23*, 203–220.

Švec, J. G., Schutte, H. K., & Miller, D. G. (1996). A subharmonic vibratory pattern in normal vocal folds. *Journal of Speech and Hearing Research*, *39*, 135–143.

Švec, J. G., Schutte, H. K., & Miller, D. G. (1999). On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges. *Journal of the Acoustical Society of America*, *106*, 1523–1531.

Tanaka, Y. (1968). 幼児の言語発達 [Speech development in infancy]. *Nippon Jibiinkoka Gakkai Kaiho*, *71*, 1454–1470.

Tenold, J. L., Crowell, D. H., Jones, R. H., Daniel, T. H., McPherson, D. F., & Popper, A. N. (1974). Cepstral and stationarity analyses of full-term and premature infants' cries. *Journal of the Acoustical Society of America*, *56*, 975–980.

Thodén, C. J., & Koivisto, M. (1980). Acoustic analysis of the normal pain cry. In T. Murry & J. Murry (Eds.), *Infant communication: cry and early speech* (pp. 124–151). Houston, TX: College Hill Press Houston.

Titze, I. R. (1988). The physics of small-amplitude oscillation of the vocal folds. *Journal of the Acoustical Society of America*, *83*, 1536–1552.

Titze, I. R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *Journal of the Acoustical Society of America*, *85*, 901–906.

Titze, I. R. (1991). A model for neurologic sources of aperiodicity in vocal fold vibration. *Journal of Speech and Hearing Research*, *34*, 460–472.

Titze, I. R. (2000). *Principles of voice production* (2nd ed.). Iowa City, IA: National Center for Voice and Speech.

Titze, I. R. (2006). *The myoelastic aerodynamic theory of phonation*. Iowa City, IA: National Center for Voice and Speech.

Titze, I. R., Baken, R. J., & Herzel, H. (1993). Evidence of chaos in vocal fold vibration. In I. R. Titze (Ed.), *Vocal fold physiology: frontiers in basic science* (pp. 143–188). San Diego, CA: Singular Publishing Group.

Titze, I. R., Luschei, E. S., & Hirano, M. (1989). Role of the thyroarytenoid muscle in regulation of fundamental frequency. *Journal of Voice*, *3*, 213–224.

Tokuda, I. T., Horáček, J., Švec, J. G., & Herzel, H. (2007). Comparison of biomechanical modeling of register transitions and voice instabilities with excised larynx experiments. *Journal of the Acoustical Society of America*, *122*, 519–531.

Tokuda, I. T., Horáček, J., Švec, J. G., & Herzel, H. (2008). Bifurcations and chaos in register transitions of excised larynx experiments. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *18*, 013102.

Tokuda, I. T., Zemke, M., Kob, M., & Herzel, H. (2010). Biomechanical modeling of register transitions and the role of vocal tract resonators. *Journal of the Acoustical Society of America*, *127*, 1528–1536.

Truby, H. M., & Lind, J. (1965). Cry sounds of the newborn infant. *Acta Paediatrica Scandinavica*, *163*, 7–59.

Tucker, J. A., & Tucker, G. F. (1979). A clinical perspective on the development and anatomical aspects of the infant larynx and trachea. In G. B. Healy & J. McGill (Eds.), *Laryngotracheal problems in the pediatric patient* (pp. 3–8). Springfield, II: Charles C Thomas.

Van den Berg, J. (1958). Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research*, *1*, 227–244.

Van Oordt, H. W. A., & Drost, H. A. (1963). Development of the frequency range of the voice in children. *Folia Phoniatrica et Logopaedica*, *15*, 289–298.

Vihman, M. M., Ferguson, C. A., & Elbert, M. (1986). Phonological development from babbling to speech: Common tendencies and individual differences. *Applied Psycholinguistics*, *7*, 3–40.

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A re-assessment of the continuity issue. *Language*, *61*, 397–445.

Ward, P. D., Thibeault, S. L., & Gray, S. D. (2002). Hyaluronic Acid:: Its Role in Voice. *Journal of Voice*, *16*, 303–309.

Wasz-Höckert, O., Lind, J., Vurenkoski, V., Partanen, T., & Valanne, E. (1968). *The infant cry: a spectrographic and auditory analysis* (Vol. 29). London: Heinemann.

Whitehead, R. L., Metz, D. E., & Whitehead, B. H. (1984). Vibratory patterns of the vocal folds during pulse register phonation. *Journal of the Acoustical Society of America*, *75*, 1293–1297.

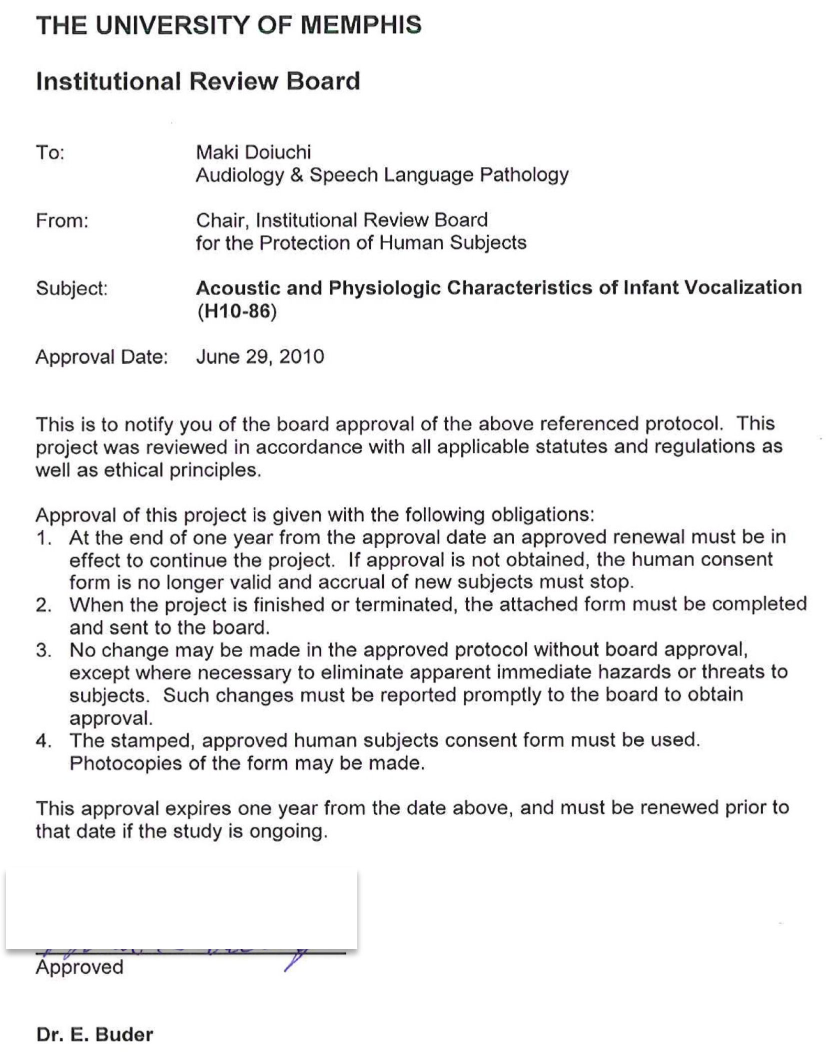
Wilden, I., Herzel, H., Peters, G., & Tembrock, G. (1998). Subharmonics, biphonation, and deterministic chaos in mammal vocalization. *Bioacoustics*, *9*, 171–196.

Willadsen, E., & Albrechtsen, H. (2006). Phonetic description of babbling in Danish toddlers born with and without unilateral cleft lip and palate. *The Cleft Palate-Craniofacial Journal*, *43*, 189–200.

Zemlin, W. R. (1998). *Speech and Hearing Science: Anatomy and Physiology* (4th ed.). Needham Heights, MA: Allyn & Bacon.

Zeskind, P. S., & Lester, B. M. (1978). Acoustic features and auditory perceptions of the cries of newborns with prenatal and perinatal complications. *Child Development*, *49*, 580–589.

Appendix A: IRB Approval form



THE UNIVERSITY OF MEMPHIS

Institutional Review Board

To: Maki Doiuchi
Audiology & Speech Language Pathology

From: Chair, Institutional Review Board
for the Protection of Human Subjects

Subject: **Acoustic and Physiologic Characteristics of Infant Vocalization (H10-86)**

Approval Date: June 29, 2010

This is to notify you of the board approval of the above referenced protocol. This project was reviewed in accordance with all applicable statutes and regulations as well as ethical principles.

Approval of this project is given with the following obligations:

1. At the end of one year from the approval date an approved renewal must be in effect to continue the project. If approval is not obtained, the human consent form is no longer valid and accrual of new subjects must stop.
2. When the project is finished or terminated, the attached form must be completed and sent to the board.
3. No change may be made in the approved protocol without board approval, except where necessary to eliminate apparent immediate hazards or threats to subjects. Such changes must be reported promptly to the board to obtain approval.
4. The stamped, approved human subjects consent form must be used. Photocopies of the form may be made.

This approval expires one year from the date above, and must be renewed prior to that date if the study is ongoing.

[redacted]

Approved

Dr. E. Buder

THE UNIVERSITY OF MEMPHIS
Institutional Review Board for the Protection of Human Subjects

Request for Modification/Addendum

Name Maki Doiuchi Phone 9014841355 Fax

Department Audiology and Speech-Language Pathology E-mail mdoiuchi@memphis.edu

Faculty advisor (if student) Eugene Buder

IRB# H10-86 Date of most recent approval 06/29/2010

Please indicate your modification/addendum to your previously approved protocol and provide brief justification. Forward request to IRB@memphis.edu. (Submit Button below)

- ☐ Additional instrument (attach copy)
- ☐ Modification of instrument (attach copy)
- ☑ Change in subject pool
  - ☐ Number of subjects
  - ☑ Age range of subjects
  - ☐ Other
- ☐ Change/addition of data collection site
- ☐ Change to compensation
- ☐ Change in recruitment methods
- ☐ Advertisement/flyers (attach copy)
- ☐ Other

Do the changes impact the consent form in any way? If yes, please attach the revised consent form. ☐ Yes, consent attached ☑ No

Details and Justification:

I would like to request a change in the age range of subjects from 4-12 months to 2-12 months. Unanticipated availability of a younger child enables us to conduct a longitudinal study and allows us to examine developmental changes. The change in age range does not affect any other parts of the approved study protocol or scientific objectives.

Approved by IRB [redacted] Date 9/30/2010

Submit to IRB

Institutional Review Board

To:	Maki Doiuchi Audiology and Speech Pathology
From:	Chair, Institutional Review Board For the Protection of Human Subjects irb@memphis.edu
Subject:	Acoustic and Physiologic Characteristics of Infant Vocalization (#2412)
Approval Date:	October 19, 2012

This is to notify you of the board approval of the above referenced protocol. This project was reviewed in accordance with all applicable statuses and regulations as well as ethical principles.

Approval of this project is given with the following obligations:

1. At the end of one year from the approval date, an approved renewal must be in effect to continue the project. If approval is not obtained, the human consent form is no longer valid and accrual of new subjects must stop.
2. When the project is finished or terminated, the attached form must be completed and sent to the board.
3. No change may be made in the approved protocol without board approval, except where necessary to eliminate apparent immediate hazards or threats to subjects. Such changes must be reported promptly to the board to obtain approval.
4. The stamped, approved human subjects consent form must be used unless your consent is electronic. Electronic consents may not be used after the approval expires. Photocopies of the form may be made.

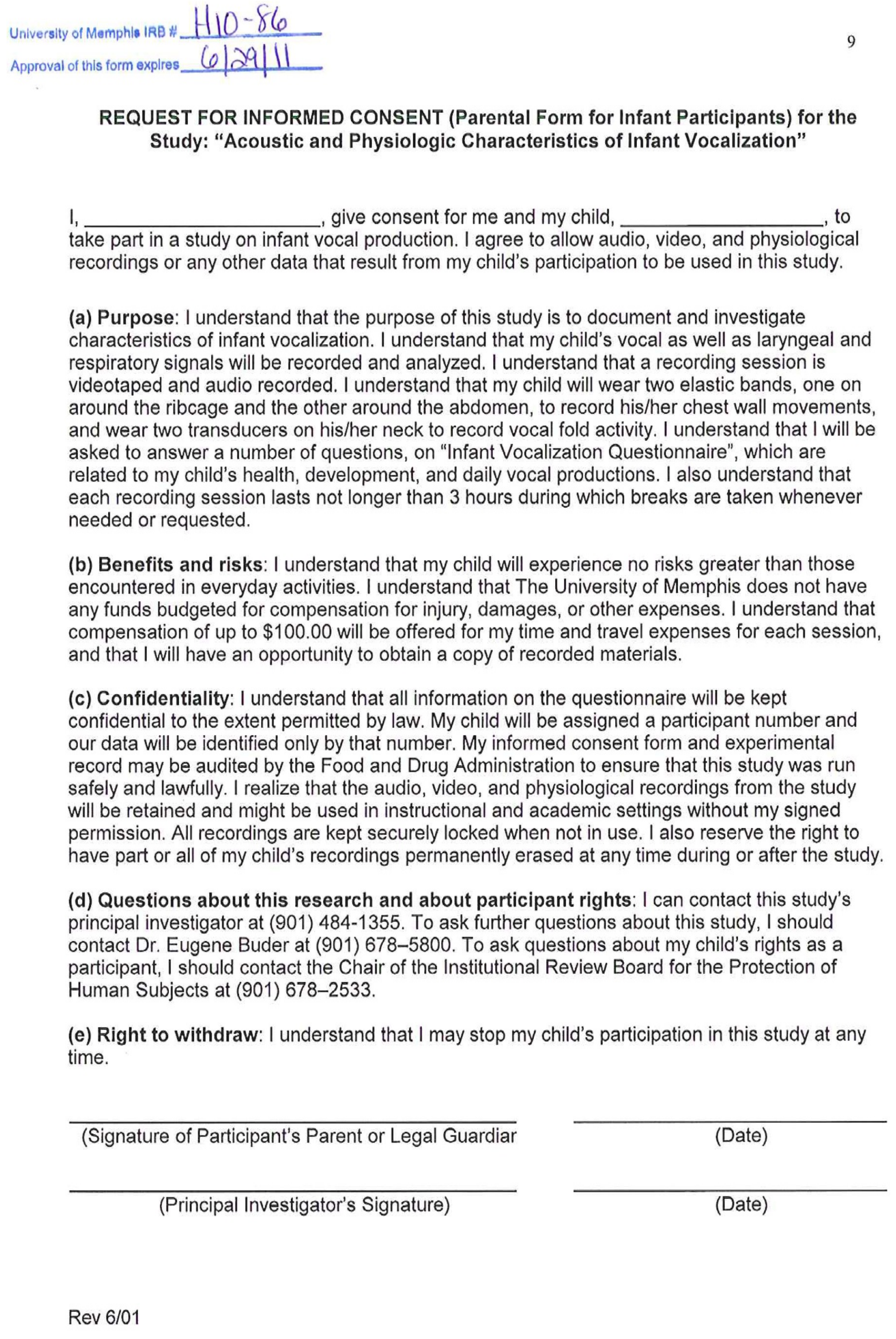
This approval expires one year from the date above, and must be renewed prior to that date if the study is ongoing.

____

Chair, Institutional Review Board
The University of Memphis

Cc: Dr. Eugene Buder

Appendix B: Parental Consent Form



University of Memphis IRB # H10-86

Approval of this form expires 6/29/11

REQUEST FOR INFORMED CONSENT (Parental Form for Infant Participants) for the Study: "Acoustic and Physiologic Characteristics of Infant Vocalization"

I, ______________________, give consent for me and my child, ___________________, to take part in a study on infant vocal production. I agree to allow audio, video, and physiological recordings or any other data that result from my child's participation to be used in this study.

(a) Purpose: I understand that the purpose of this study is to document and investigate characteristics of infant vocalization. I understand that my child's vocal as well as laryngeal and respiratory signals will be recorded and analyzed. I understand that a recording session is videotaped and audio recorded. I understand that my child will wear two elastic bands, one on around the ribcage and the other around the abdomen, to record his/her chest wall movements, and wear two transducers on his/her neck to record vocal fold activity. I understand that I will be asked to answer a number of questions, on "Infant Vocalization Questionnaire", which are related to my child's health, development, and daily vocal productions. I also understand that each recording session lasts not longer than 3 hours during which breaks are taken whenever needed or requested.

(b) Benefits and risks: I understand that my child will experience no risks greater than those encountered in everyday activities. I understand that The University of Memphis does not have any funds budgeted for compensation for injury, damages, or other expenses. I understand that compensation of up to $100.00 will be offered for my time and travel expenses for each session, and that I will have an opportunity to obtain a copy of recorded materials.

(c) Confidentiality: I understand that all information on the questionnaire will be kept confidential to the extent permitted by law. My child will be assigned a participant number and our data will be identified only by that number. My informed consent form and experimental record may be audited by the Food and Drug Administration to ensure that this study was run safely and lawfully. I realize that the audio, video, and physiological recordings from the study will be retained and might be used in instructional and academic settings without my signed permission. All recordings are kept securely locked when not in use. I also reserve the right to have part or all of my child's recordings permanently erased at any time during or after the study.

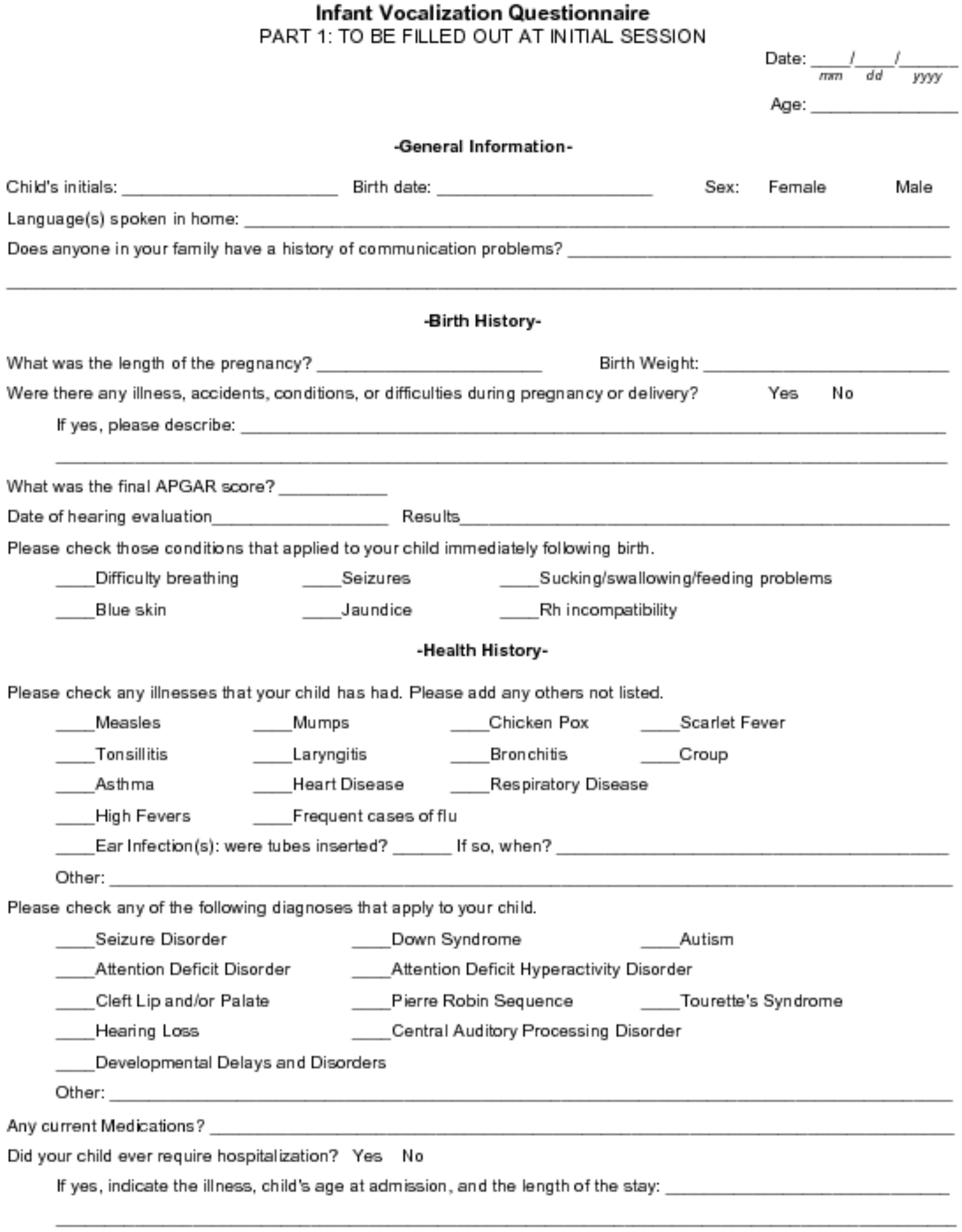
(d) Questions about this research and about participant rights: I can contact this study's principal investigator at (901) 484-1355. To ask further questions about this study, I should contact Dr. Eugene Buder at (901) 678–5800. To ask questions about my child's rights as a participant, I should contact the Chair of the Institutional Review Board for the Protection of Human Subjects at (901) 678–2533.

(e) Right to withdraw: I understand that I may stop my child's participation in this study at any time.

______________________________	______________________
(Signature of Participant's Parent or Legal Guardiar	(Date)
______________________________	______________________
(Principal Investigator's Signature)	(Date)

Rev 6/01

Appendix C: Infant Vocalization Questionnaire Part 1



Infant Vocalization Questionnaire

PART 1: TO BE FILLED OUT AT INITIAL SESSION

Date: ____/____/______
mm dd yyyy

Age: ______________

-General Information-

Child's initials: ____________________ Birth date: ____________________ Sex: Female Male

Language(s) spoken in home: ______________________________

Does anyone in your family have a history of communication problems? ______________________________

____

-Birth History-

What was the length of the pregnancy? ____________________ Birth Weight: ____________________

Were there any illness, accidents, conditions, or difficulties during pregnancy or delivery? Yes No

If yes, please describe: ______________________________

____

What was the final APGAR score? __________

Date of hearing evaluation________________ Results______________________________

Please check those conditions that applied to your child immediately following birth.

____Difficulty breathing
____Blue skin
____Seizures
____Jaundice
____Sucking/swallowing/feeding problems
____Rh incompatibility

-Health History-

Please check any illnesses that your child has had. Please add any others not listed.

____Measles
____Tonsillitis
____Asthma
____High Fevers
____Mumps
____Laryngitis
____Heart Disease
____Frequent cases of flu
____Chicken Pox
____Bronchitis
____Respiratory Disease
____Scarlet Fever
____Croup

____Ear Infection(s): were tubes inserted? ______ If so, when? ______________________________

Other: ______________________________

Please check any of the following diagnoses that apply to your child.

____Seizure Disorder
____Attention Deficit Disorder
____Cleft Lip and/or Palate
____Hearing Loss
____Developmental Delays and Disorders
____Down Syndrome
____Attention Deficit Hyperactivity Disorder
____Pierre Robin Sequence
____Central Auditory Processing Disorder
____Autism
____Tourette's Syndrome

Other: ______________________________

Any current Medications? ______________________________

Did your child ever require hospitalization? Yes No

If yes, indicate the illness, child's age at admission, and the length of the stay: ______________________________

____

Infant Vocalization Questionnaire

PART 1: TO BE FILLED OUT AT INITIAL SESSION

Date: ____/____/______
mm dd yyyy

Age: _______________

-Developmental History-

During the first year, other than crying, would you say that your child was:

Very quiet _____ Quiet _____ "Average" quiet _____ Noisy _____ Very noisy _____

At what age did the following occur? (If you can't remember specific time, please indicate if it occurred at the expected time or was delayed)

Held up his/her head _____ Sat up alone _____ Pulled up to a standing position _____

Crawled _____ Walked _____ Toilet trained_____

Compared to other children your child's age, describe how your child is able to sit, stand, run, use his/her hands:
____

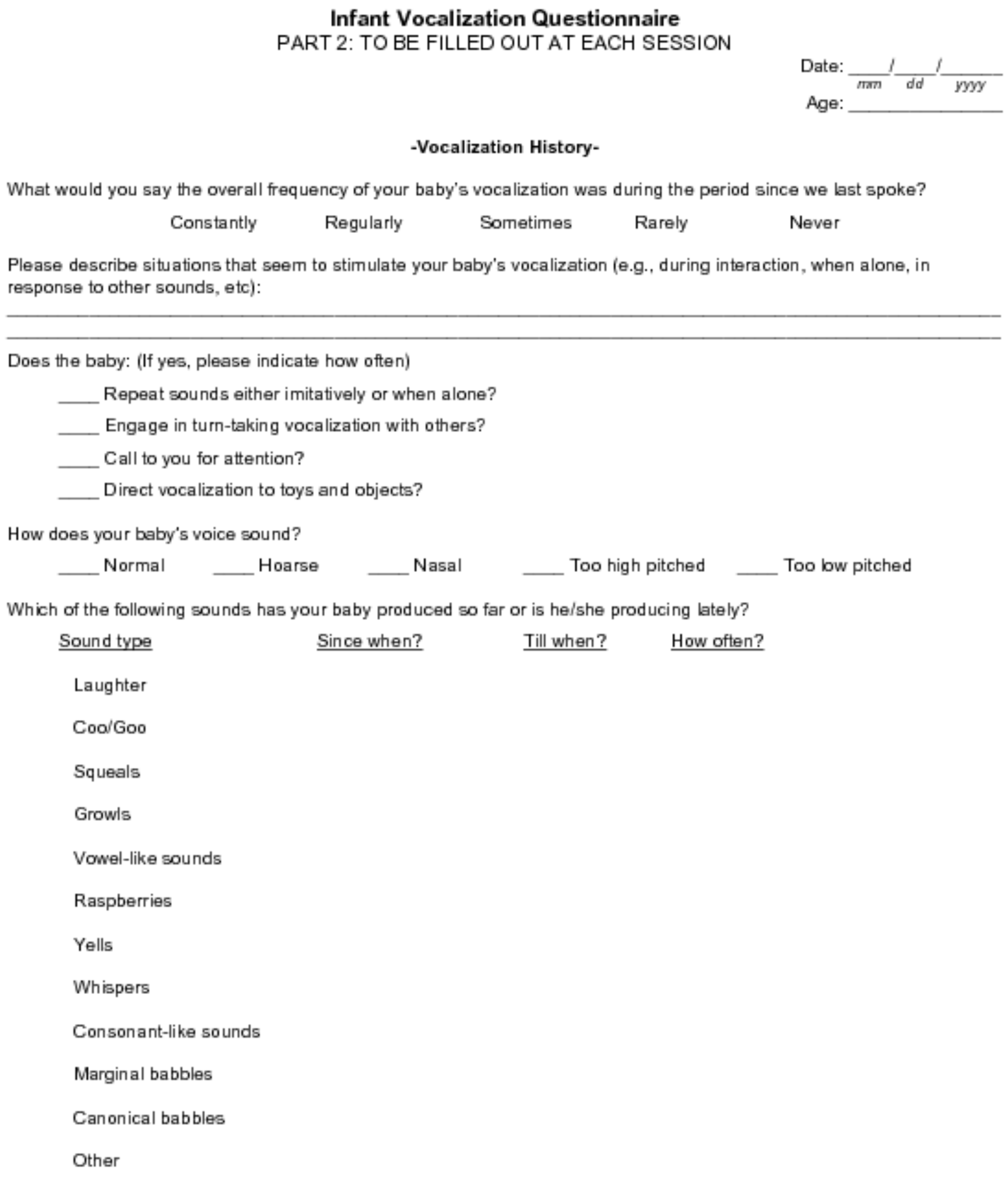
Does your child respond to: Light? __________ Sound? __________ People? __________

Does your child: Cry appropriately? __________ Laugh? __________ Smile? __________

How does your child make wants known? ____

Does your child show unusual behavior (please explain)? ____
____

Appendix D: Infant Vocalization Questionnaire Part 2



Infant Vocalization Questionnaire

PART 2: TO BE FILLED OUT AT EACH SESSION

Date: ____/____/______
mm dd yyyy

Age: ______________

-Vocalization History-

What would you say the overall frequency of your baby's vocalization was during the period since we last spoke?

Constantly Regularly Sometimes Rarely Never

Please describe situations that seem to stimulate your baby's vocalization (e.g., during interaction, when alone, in response to other sounds, etc):

____

____

Does the baby: (If yes, please indicate how often)

____ Repeat sounds either imitatively or when alone?

____ Engage in turn-taking vocalization with others?

____ Call to you for attention?

____ Direct vocalization to toys and objects?

How does your baby's voice sound?

____ Normal ____ Hoarse ____ Nasal ____ Too high pitched ____ Too low pitched

Which of the following sounds has your baby produced so far or is he/she producing lately?

Sound type	Since when?	Till when?	How often?
Laughter			
Coo/Goo			
Squeals			
Growls			
Vowel-like sounds			
Raspberries			
Yells			
Whispers			
Consonant-like sounds			
Marginal babbles			
Canonical babbles			
Other			