Acoustic Evidence for the Development of Speech

RLE Technical Report No. 548

October 1989

Corine Anna Bickley

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

ACOUSTIC EVIDENCE FOR THE DEVELOPMENT OF SPEECH

by

Corine Anna Bickley

Submitted to the Department of Electrical Engineering and Computer Science on May 22, 1987, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

This thesis develops models of relationships between the physical attributes of young children and the acoustic characteristics of their vocalizations. Two questions are addressed. What are the acoustic characteristics of children's speech? How can acoustic data be used to describe the development of speech? Children's vocalizations exhibit a range of characteristics, becoming more adult-like as a child grows and develops abilities in motor control and cognitive functioning. Developmental changes which take place during childhood result from changes in the child's anatomy and in motor-control and cognitive abilities.

The physical attributes of a child's vocal system and the constraints on motor control must be known in order to construct models of the characteristics of children's speech. These attributes include the dimensions of the lungs, trachea, glottis, vocal tract and head and the tissue properties of the glottal structure and vocal-tract walls. Data on physical attributes of the speech production systems of young children are summarized in the form of model parameters in this thesis.

A new model of the fundamental frequency of vocal-fold vibration is developed; this model is based on the theory of bending beams. An important parameter of the model is the ratio of vocal-fold thickness to length. This ratio affects the fundamental frequencies which are calculated from the model.

Predicted durations rely on assumptions of the volume of air used in speech and the airflow through the glottis and vocal tract. The predictions imply that children's voiced utterances can be longer than adults' in spite of significantly smaller lung volumes, due to differences in airflow through the glottis. On the other hand, children are predicted to produce fewer consonant-vowel syllables in an utterance because the durations of individual segments are usually longer.

Formant frequencies and bandwidths are predicted using parameter values appropriate for child-sized vocal systems and models which are similar to those for adult speech. Values of frequency and bandwidth are not related to adult values by one simple scale factor.

Measurements of young children's vocalizations are interpreted with reference to the models describing the relationships between acoustic characteristics and articulatory parameters. Developmental changes in vocal-tract characteristics, in respiratory and articulatory control, and in phonological representation are inferred from acoustic characteristics.

Thesis Supervisor: Louis D. Braida, H.E. Warren Professor of Electrical Engineering

Thesis Supervisor: Kenneth N. Stevens, C.J. LeBel Professor of Electrical Engineering

Acknowledgments

The motivation for my research comes from talking with children, particularly the ones for whom speaking is a constant battle against a wide array of difficulties. I thank them for their persistence; I derive inspiration from their struggles.

I am grateful to my thesis advisor Lou Braida for offering me the opportunity to participate in a project which is so closely tied to my goals for graduate work. His support over many years allowed me to continue my involvement in the project. I thank my thesis advisor Ken Stevens for his encouragement throughout my graduate career. He taught me many of the techniques which were used in pursuing this research. More importantly, though, he has been an example to me of a thoughtful researcher and skilled scientist. My readers, Victor Zue and John Locke, have been most helpful in various aspects of this research effort. I thank Victor for making room for the "LIDS kids" in the SPIRE workstation environment, a facility which proved to be enormously valuable for measuring children's speech. I thank John for helping me remember the many facets of children's development.

Several other persons have participated in this work through discussion of key issues in the development of the ideas presented in this thesis. I thank Jim Melcher for his untiring encouragement in applying the analysis techniques of continuum mechanics to a structure of a rather different sort – a vocal fold. I am most grateful to Bjorn Lindblom for the opportunity to participate with him in analyses of the vocalizations of Swedish children. I am deeply indebted to my other readers Patrick Mock, Dave Feldmeier, Kevin Brown, Dan Sobek, Stefanie Shattuck-Hufnagel, Abeer Alwan, Victor Reid and Bonnie Walters for many stimulating discussions and insightful suggestions during the course of the work and the preparation of this document. An extra thank you is in order to Kevin Brown for all the drawings in this thesis. I thank Keith North for his extraordinary commitment to maintaining an environment in which research can be accomplished with relative ease. Rob Kassel has given generously of his time in order to find solutions to many text-processing problems. The preparation of this document was greatly facilitated through his efforts.

Any research effort the magnitude of the LIDS or Jollerutveckling project can only be accomplished through the concerted efforts of many persons. I thank the researchers of the LIDS project, especially Paula Menyuk, Jackie Liebergott, Martin Schultz, Linda Ferrier and Marie Chesnick, for all their work which made the recordings available. I am indebted to Bob Pearsall, Kevin White, Eric Levine, Buzzy Dale, Adele Proctor, Barbara Hubbard, Becky Gray and Hilary Sheldon for their assistance

with the many measurements, and to Barbara Hubbard and Katy Abramson for their thorough library research. I also thank the researchers of the Jollerutveckling project, and Bjorn Lindblom, Karin Holmgren and Liselotte Roug in particular, for many hours spent documenting and interpreting the vocalizations of Swedish children.

The understanding of friends and family during a project such as a thesis is a blessing. My thanks especially to Pat, Dave, Lori and Linda for freeing up my time so that I was able to concentrate solely on this thesis during the final stages. I thank Rob for his friendship, confidence and constant assurance throughout the many phases of my graduate career. I thank my near and extended family for their continual encouragement, including most specially my mother, friend and teacher, Jeanette, whose constant thoughtfulness and support has enabled me to pursue this work.

This research was partially supported by the National Institute of Neurological and Communicative Disorders and Stroke in the form of an NIH traineeship, by a Whitaker Health Sciences fellowship, and by a C.J. LeBel fellowship. The experimental work used facilities supported in part by the Research Laboratory for Electronics, the Defense Advanced Research Projects Agency, and the Institute for Linguistics at Stockholm University.

Biographical Note

Corine Bickley was born in St. Louis, Missouri, on April 13, 1949. She received B.S. degrees in Mathematics and in Computer Science from Iowa State University in 1970. In 1972 she completed the M.S. degree in Computer Science from Washington State University. The Master's thesis is titled "Language Typologies: a Comparative Study." She joined the faculty of the California Polytechnic State University-San Luis Obispo in the fall of 1972 and taught courses in mathematics and computer science. She worked at Gesco Corp. in Fresno, California, from 1974 to 1976 as a programmer/analyst. Between 1976 and 1978, she studied speech pathology at the California State University–Fresno and taught computer programming courses in the Department of Quantitative Studies at the same university. During 1978 she worked as a speech therapist for the Fresno Unified School District. In 1979 she began graduate studies at M.I.T. in order to learn skills in engineering that she plans to apply to research concerning communication disorders. She has authored several technical journal articles, and has presented her research at many conferences.

Contents

List of Figures

List of Tables

Chapter 1

Children's speech

Children produce a wide variety of sounds. During the early weeks of a child's life, crying and the sounds of breathing and eating are prevalent. Within the first year, babbling and other vocalizations begin to dominate. Children's vocalizations exhibit a range of characteristics, becoming more adult-like as a child grows and develops abilities in motor control and cognitive functioning.

The common thread of change over time is apparent in nearly every report of children's utterances. Children begin to produce sounds at a very early age. Normal children communicate with speech and language skills which approximate those of adults by the age of two or three years (Wood, 1976). The developmental changes which take place between these times result from changes in the child's anatomy and in motor-control and cognitive abilities. Each of these components constrains the sounds produced by a child. Some of the acoustic characteristics of these sounds may be a consequence of the size of the structures involved in speech production: the lungs, the larynx, the vocal tract. Other characteristics may be influenced most by the motor-control skills of a young child. The cognitive ability to form and manipulate mental representations of words clearly has a significant influence on the sound sequences produced by a child.

The problem of analyzing the development of speech is discussed in the following paragraphs. A method of solution is outlined; arguments are presented in support of a method which compares model predictions to acoustic measurements. The applicability of previously reported models and the need for additional data are described. A summary of the research plan concludes the chapter.

1.1 The research problem

There are many questions that could be asked about the process of early development of speech. During the first few years of life, a child's speech changes from the cries and babbles of an infant to adult-like words and phrases of a young child. We could ask in what ways children's speech is similar to adults' speech. At what age does a child begin to produce adult-like sounds? How does children's speech change over time? How does a child's anatomy constrain speech? The ways in which vocalizing relates to other behaviors could be studied, as could the ways in which adults respond to children's vocalizations. Questions about linguistic representations of words spoken by children could be asked. These and other questions have been addressed by many researchers from a variety of perspectives. However, the process of speech development is still only partially understood.

One aspect of an examination of speech development is a description and a modeling of the acoustic characteristics of children's speech. Many studies which describe the acoustics of speech have focused on the speech of adults; fewer have analyzed children's speech. Much of the literature on children's speech describes sounds produced by one or a few children. The completeness of this body of literature is limited by the difficulties of collecting, analyzing and interpreting children's vocalizations. Models which predict the acoustic characteristics of adult speech are common in the literature, but few models of children's vocal systems can be found (see, however, Goldstein, 1980). Data are available describing the anatomy, motor-control skills and cognitive abilities of children, but the effects of each of these constraints on a child's speech have not been analyzed comprehensively. Analyses of acoustic models can provide insights into the constraints imposed by a child's anatomy and motor-control abilities on speech and the physical bases for the development of speech. Currently, there is a need for a comprehensive examination of children's speech with respect to the applicability of adult models and, if needed, the development of new models which more accurately predict the acoustic characteristics of children's speech.

This thesis develops quantitative analyses of the acoustic characteristics of children's speech and provides examples of using acoustic evidence to describe the development of speech. The following questions are addressed:

1. What are the acoustic characteristics of children's speech?

2. How can acoustic data be used to describe the development of speech?

Developmental changes in vocal-tract characteristics, in respiratory and articulatory control, and in phonological representation can be inferred from acoustic characteristics. Advantages of using acoustic data are the non-invasive nature of the measurement procedure and the greater quantification provided by acoustic measurements than by auditory evaluation. In addition, models of the speech of normal children would provide a basis for the evaluation of delayed or disordered speech.

1.2 Modeling children's speech

The questions asked about children's speech influence the choice of methods used to determine answers. In order to answer the above questions concerning the acoustic characteristics of children's speech, models of the speech production capabilities of a child have been formulated. The models, which are based on acoustic theory (Fant, 1960/70), incorporate known sources of constraint on speech production in terms of model parameters, such as dimensions of the vocal tract and volume of the lungs. The models predict temporal and spectral characteristics of sound segments as well as of sequences of these segments. The predicted characteristics are compared with data collected from children's speech; these data consist primarily of acoustic measurements. The choice of acoustic measurements as a representation of children's sounds is discussed below. In order to describe the development of speech, the acoustic data are analyzed with reference to the models for evidence of the developmental process. The physical growth of a child, both in size and articulatory control, affects the acoustic characteristics which are apparent at different ages. The changes which cannot be predicted from physical parameters are likely to be due to linguistic or cognitive development.

1.2.1 Determination of model parameters

The parameters of the models are the dimensions and tissue properties of the vocal structures and the rates of articulator movement of children. The models do not include constraints imposed by cognitive and language skills, or differences in perceptual abilities between children and adults. All of the model parameters are determined from data reported in the medical and speech literature.

The anatomical parameters include the dimensions of the structures involved in speech production: the lung volume, the dimensions of the larynx and vocal tract, the area of the mouth and the size of the head. Tissue properties of the laryngeal structures and vocal-tract walls complete the set of anatomical parameters.

The motor-control parameters describe respiratory and articulatory control. The ability to maintain subglottal pressure and to control airflow influences respiratory control. Articulatory control is measured in terms of rates and patterns of movement of the articulatory structures.

1.2.2 Choice of a representation of children's sounds

All methods of studying speech represent the sounds generated by a talker. The most common representation is a transcription of speech as a series of phonetic symbols. Transcription techniques assume that at some level speech consists of a sequence of phonemes. This level may be a relatively low level of motor control or a higher level of mental representation. Fant (1960/70) remarks that "the most common approach [to speech analysis] has been to start with the linguistic criteria in terms of a phonemic transcription and to impose this as a basis for divisions." Descriptions in terms of acoustic measurements are also common. Analyses of children's sounds often use the same acoustic measures as the ones used in studies of adult speech. In those cases in which adult measures or representations appear inadequate because of the acoustic differences between children's and adults' speech, other measures have been proposed (e.g., Ferguson and Farwell, 1975; Velleman, 1983).

Broad labels for general characteristics of utterances are also used to describe children's vocalizations. Vocal stages and coding schemes are examples of broad descriptors. Some researchers specify vocal stages to highlight abrupt changes in the child's ability to vocalize, such as the onset of the production of sequences of similar syllables. Coding schemes are used in a few studies to describe characteristics of children's vocalizations and to relate vocalizing to other activities. Coding schemes do not specify details of sound production, but instead classify entire utterances as "requesting" or "vegetative," for example.

Transcriptions, although useful for describing sounds as they are perceived by adult listeners, are not sufficient as a representation of children's sounds. They do not adequately describe details of early development of speech for the following reasons:

1. Transcription systems tend to exclude the sounds that do not correspond to sounds in some language. Thus some sounds which could provide data on child/adult speech differences may be eliminated from analysis.

2. A transcriber's perceptual abilities influence phonetic transcriptions. Even a narrow transcription maps children's sounds onto symbols for sounds in languages known to the transcriber.

3. Conversion from sounds to phonetic symbols and diacritics collapses some details of the variations within and among productions. The variability in children's productions can be evidence of the development of articulatory control.

Objections to the use of transcriptions are not new. In criticizing Irwin and his colleagues, Lynip (1951) took the extreme position that *no* useful data could be obtained from transcriptions. Winitz (1960) countered with the argument that the contributions of "phonetic [transcriptions] and physical [spectrograms] were distinct" and that a "certain degree of subjective judgment is present in all interpretations placed on spectrogram records." As he reminds us, "accuracy depends not only on ... smallness of error ... but on correctness of interpretation."

The results of studies of child language are consistent with the hypothesis that children's representations of words are in terms of phonemes and representations of phonemes are in terms of distinctive features. Berko (1958) asked children to add inflections such as plural endings and past tense to nonsense words. Children as young as four years used the plural forms [-s], [-z] and [-ɪz] appropriately, depending on the features [voiced], [strident] and [coronal] of the last phoneme of the word. The age at which a child develops phoneme-based or distinctive-feature-based representations is unclear.

Transcriptions of young children's speech differ among transcribers (Stockman *et al.*, 1981; Holmgren *et al.*, 1983). The variability complicates the comparison of data from various studies. Comparison of transcriptions in terms of distinctive features reduces the differences among perceptual analyses (Holmgren *et al.*, 1983).

Acoustic measurements are not constrained by the categorization inherent in a transcription system. The acoustic signal can be analyzed directly from recordings of a child's speech and can show details of the structure of the sound in terms of time-varying acoustic parameters. An acoustic analysis can provide a quantitative description of the changes over time in the child's productions of sounds.

The choice of a particular set of acoustic measures influences the process of speech analysis. Using measures which describe differences which are linguistically distinctive in the ambient adult language can impose on children's speech a description which captures only those characteristics which are most adult-like and can miss other characteristics. A child's productions may differ from an adult's in ways which are not distinctive (e.g., the amplitude of noise in vowels). However, the measurement of non-distinctive attributes could provide useful information for the analysis of children's speech. For instance, the amount of noise generated by a child during productions of vowels could be used to determine the relative cross-sectional areas of the vocal-tract configurations during the vowel productions. Also, in some cases children make an acoustic distinction which is not the same as adults make (e.g., voice onset time for the voiced/voiceless distinction (Kewley-Port and Preston, 1974)).

Abrupt changes in particular characteristics of a child's vocal productions, such as the change from a predominance of glottal closures to supraglottal closures, delineate stages of speech development. The mastery of a speech-like skill is a note-worthy event in the development of speech. It can be argued, though, that any division of the developmental process into times before and after a specific event, such as the emergence of recognizable forms of adult words, is arbitrary. Evidence in support of continuity from vocalizations produced early in life through babbling and into speech (Oller *et al.*, 1975) is widely accepted. Vihman *et al.* (1985) review arguments concerning the relationship between babbling and early speech. They report a "striking parallelism between babbling and words within each child, across time and within time period." Locke (1983) found support from studies of babbling in various languages to support this position. As Lieberman (1980) comments, "the hypothetical abrupt change in behavior (Jakobson, 1941/68) that is supposed to occur between the babbling and phonological stages of language acquisition thus may be a consequence of a different frame of interpretation that the phonetician brings to bear."

Transcriptions and acoustic measurements are more explicit than coding schemes or stage identification and therefore are used in this thesis to represent sounds. For the most part, the analyses utilize acoustic measurements. Particular measurement procedures are chosen so as to be sufficiently detailed to reveal the mastery of distinct skills, yet sufficiently broad to group together productions which share particular speech-like attributes. Transcriptions are used in a minor way, as labels of speech-like characteristics.

1.2.3 Data base of measurements of children's sounds

The accumulation of previous reports of sounds produced by children constitutes a valuable source of data on children's speech, which includes frequencies of occurrence of vowels and consonants and acoustic measurements of fundamental frequency, formant frequencies and segment durations. The present study includes measurements of these characteristics plus formant bandwidths of vowels and durations of syllables and utterances. Analyses of recordings of a group of several children who have been followed over a period of years are sparse. In the present study, measurements were made of vocalizations of a relatively large population (44 children); each child was recorded for approximately two years.

The period of a few months of age through two and one-half years was chosen for analysis because during this time period a child's utterances change from apparently unstructured vocalizations to adult-like speech. By the age of approximately two and one-half years, normal children have progressed through babbling and one-word utterances to the production of phrases and sentences.

1.2.4 Interpretation of evidence of speech development

The goal of this research is to understand the changes which occur in the sounds produced by young children during early speech development. Models based on anatomical and motor-control parameters attempt to account for the changes in a child's speech which are the direct result of growth and the development of control of respiration and articulation. The identification of the changes in children's speech which are due to these constraints simplifies the analysis of development by restricting the set of characteristics which must be accounted for in terms of the development of cognitive and language skills.

The method used in this thesis is to compare the acoustic characteristics predicted from models to acoustic measurements. Measurements of the speech of a child at various ages provide data on developmental patterns. The changes in temporal and spectral characteristics of a child's vocalizations which can be predicted from changes in the parameter values of the models are interpreted as evidence of physical growth or increased motor skill.

The changes in the sounds produced by children show various patterns. For instance, in the course of development there might be a change from a random distribution of formant frequencies of vowel-like sounds over all of the vowel space to clusters of formant frequency values. In contrast, development might be inferred from evidence of a change from a single gesture being produced in all contexts to a variety of gestures, dependent on context. Another pattern that has been observed is the development of the ability to replicate a sound with relatively consistent spectral and temporal characteristics on repeated attempts. Some of the patterns observed in the acoustic measurements can be interpreted in terms of physical changes. If there does not appear to be an explanation based on physical changes, then hypotheses need to be formed in terms of linguistic development to account for the observed patterns. An example of an argument for linguistic development is given in Chapter 4.

1.3 Related research

Other researchers have reported results which are useful in an analysis of speech development. Of particular relevance are modeling studies of child-sized vocal tracts (Nordström, 1975; Goldstein, 1980; Pearsall, 1985), discussions of the accuracy of acoustic measurement of high-pitched speech (Lindblom, 1962; Monsen and Engebretson, 1983; Glass, 1983) and reports of acoustic measurements of children's speech (e.g., Irwin and Chen, 1946; Irwin, 1948; Peterson and Barney, 1952; Eguchi and Hirsh, 1969; Keating and Buhr, 1978; Golub, 1980; Lieberman, 1980; Buhr, 1980; Stark, 1980; Oller, 1980; Kent and Murray, 1982; de Boysson-Bardies *et al.*, 1981). Modeling studies demonstrate the effects on formant frequencies of scaling down vocal-tract configurations; dimensions of children's vocal structures can be inferred from formant frequencies. Because children speak with a higher pitch than do adults, the effect of higher pitch on the accuracy of acoustic measurements should be understood in order to assess the significance of differences among measurements reported in one or a group of studies. Reports of acoustic measurements of children's sounds consist primarily of measurements of fundamental and formant frequencies and temporal measurements of segment durations. The combination of existing data and new measurements represents a wide variety of children. Developmental patterns in these data are likely to be independent of any particular data collection procedure.

1.3.1 Modeling studies of children's vocal tracts

Computer models of vocal tracts have been used to predict and interpret formant frequency patterns of children's speech. Nordström (1975) investigated methods of estimating formant frequencies from a scaled model of an adult-sized vocal tract. In one experiment, he compared the effect of a uniform scaling of vocal-tract length with the effect of scaling the lengths of the mouth cavity and the pharynx by different factors. In a second experiment, he calculated formant frequencies from a vocal-tract configuration in which both the length and cross dimensions were scaled by the same factor. The values calculated in each experiment are in partial agreement with formant frequencies reported in the literature, although neither experiment completely accounts for the differences between formant frequencies of men, women and children.

Goldstein (1980) implemented a model of vocal-tract shape based on anatomical data on children. The sizes of the articulatory structures at various ages affect vocal-tract length and cross-sectional area. Formant frequencies were calculated from the model by a computer program (Henke, 1966). Goldstein demonstrated that the vocal tract of a child as young as one year old is capable of forming articulatory configurations which correspond to the production of the vowels /i ɑ u/.

Pearsall (1985) estimated vocal-tract configurations from formant frequencies of vowels produced by young children. He attempted to determine configurations for which the formant frequencies approximated those measured from children's productions of /i/ and /ɑ/. Pearsall's algorithm created a configuration by perturbing a uniform vocal tract at one or two locations. Formant frequencies were calculated from the vocal-tract area function using Henke's program. The algorithm then selected a new configuration based on the match between the calculated and measured formant frequencies of children's vowel productions. Pearsall was able to determine a vocal-tract configuration for /i/; lack of convergence of his iterative procedure prevented the determination of a configuration corresponding to the production of /ɑ/.

1.3.2 Accuracy of measurement of children's speech

In some ways, children's speech is more difficult to measure reliably than adults' and in other ways, equally difficult. Measuring formant frequencies and bandwidths of vowels is particularly troublesome in children's speech. Measuring fundamental frequency, rates of abrupt change, and segment and syllable durations is no less difficult than in adults' speech.

The attribute of children's speech which most seriously complicates the measurement of acoustic characteristics is the high fundamental frequency present in most children's vocalizations. Procedures for measuring formants of vowels often rely on the acoustic signal's having a low fundamental frequency – low enough so that each formant envelope covers several harmonics. The accuracy of formant frequency estimation has been explored over a range of fundamental frequencies (Lindblom, 1962; Monsen and Engebretson, 1983; Glass, 1983); this range overlaps the lower end of the range of fundamental frequencies found in children's speech (Keating and Buhr, 1978). While none of these studies identifies an optimal method of formant estimation, the results provide some constraints on the accuracy of formant estimation by commonly-used techniques.

An early study of accuracy of formant measurements (Lindblom, 1962) discusses variability in formant estimation. Lindblom reports that the mean error in formant estimation from wideband spectrograms and wide- and narrowband spectra was approximately 40 Hz for vowels synthesized with a fundamental frequency typical of an adult male. For vowels synthesized with a higher fundamental frequency, the mean error ranged from about 40 Hz to one-fourth of the fundamental frequency.

Monsen and Engebretson (1983) examined utterances with fundamentals in the range of 100 to 500 Hz and determined the measurement accuracy to be within 60 Hz for fundamental frequencies in the range of 100 to 300 Hz. For higher fundamentals, they report that neither spectrographic estimation nor linear prediction is accurate.

Glass (1983) compared formant frequencies estimated from spectral representations to the known formant frequencies of a synthesized neutral vowel. He estimated formants from narrowband discrete Fourier transforms (DFTs), wideband DFTs, cepstrally-smoothed spectra and 19-coefficient linear prediction (LPC) envelopes. The fundamental frequency of the vowel ranged from 50 to 500 Hz. He concluded that the wideband DFT peaks correspond the most closely to actual formants for vowels with fundamental frequencies between 280 and 500 Hz. The accuracy of formant estimation depends on the relative values of the fundamental and formant frequencies. For instance, for fundamental frequencies in the range of 400 to 500 Hz, the formant frequency was estimated to be approximately equal to the frequency of the harmonic nearest the actual formant.

1.3.3 Previous acoustic analyses of children's sounds

Children's sounds have been analyzed by many researchers using a variety of techniques. The sounds children produce from the earliest of ages are commonly thought to provide evidence concerning speech development. A few of the more comprehensive studies of children's sounds are described below; these studies provide data on the acoustic characteristics of children's speech.

Analyses of the sounds produced by children typically focus on either cries or vocalizations. In most reports, a vocalization is defined as an utterance which is not a cry or vegetative sound. Measurements of acoustic characteristics of children's vocalizations have included both spectral measurements, such as fundamental and formant frequencies, and temporal measurements, usually segment durations.

1.3.3.1 Cries

Cries of newborn children have been analyzed in detail (Wasz-Höckert *et al.*, 1968; see also the review in Golub, 1980). The cry of a newborn is thought to be primarily a reflexive activity, and therefore is a forerunner of speech in only a general sense. Measurements of fundamental and formant frequencies of newborn cries are approximately equal to the frequencies measured in early vocalizations. For instance, Golub reports an average fundamental frequency of 460 Hz during cry phonation, and first and second formant frequencies of 1113 Hz and 3715 Hz, respectively. He defines the phonation mode of crying as characterized by a fundamental frequency in the range of 200 to 600 Hz. Other modes of crying are less similar to vocalizing due to their very high fundamental frequencies or lack of periodic excitation.

1.3.3.2 Vocalizations

The reports of Irwin and his colleagues (Irwin and Chen, 1946; Irwin, 1948) are typical of studies of children's vocalizations which rely primarily on transcriptions. Irwin's reports document the frequency of occurrence in children's speech of each of the vowels and consonants. In these studies, adult listeners transcribed utterances produced by children of age one through thirty months. Many studies of children's speech have followed similar analysis procedures and provide a wealth of information on the sounds and sound sequences which adults perceive in the vocal productions of children.

Reports of acoustic measurements of children's sounds are of more direct relevance to this thesis than are analyses based on transcriptions. In a pioneering study of vowels, Peterson and Barney (1952) included measurements of the formant frequencies of vowels produced by children and perceptually identified by adult listeners. The children in their study were able to read the list of test words and thus are assumed to be older than three years. From graphs of the first formant frequency (F1) vs. the second formant frequency (F2), they observed that the children's vowel space was shifted to higher frequencies (34 percent on the average) in comparison to the adults'.

The formant frequency range over all vowels produced by men is 270 – 730 Hz for F1 and 840 – 2290 Hz for F2. For children, the range of F1 is 370 – 1030 Hz and of F2, 1060 – 3200 Hz.

Measurements of formant frequencies reported by other researchers are consistent with the pattern of higher formant frequencies in children's speech than in adults'. Eguchi and Hirsh (1969) measured formants of vowels produced by children in an imitation task. Their data show that imitations of the adult vowels /i æ u ɛ ɑ ɔ/ were well separated in the F1 vs. F2 vowel space by age three years. These data document children's ability to produce recognizable vowels by at least three years of age. In related studies, Lieberman (1980) and Buhr (1980) report acoustic measurements of young children's vowel-like and consonant-like utterances. Buhr's study focuses on one child from age 14 to 64 weeks. Formants of vowels were measured and plotted on an F1 vs. F2 graph. Each vowel-like utterance was classified by the researcher and verified by listeners as an English vowel or as an unrecognizable vowel-like sound. Lieberman reports similar data on the productions of five children aged sixteen weeks to five years. Kent and Murray (1982) analyzed utterances produced by three-, six- and nine-month-old children during one-hour sessions. They observed that the range of formant frequencies of all of the vowels produced by each age group increases somewhat with age, although the center of the F1 vs. F2 space remains constant. For three-month-olds, the first formant frequency falls between 500 and 1300 Hz, and the second formant frequency falls between 1900 and 3800 Hz. At age nine months, the range of formant frequencies used by the children has increased to 500 to 1600 Hz for F1, and 1500 to 4000 Hz for F2. Bond *et al.* (1982) present data consistent with the graphs of formant frequencies of Kent and Murray, Buhr, and Lieberman. They observed overlap in transcribed vowel groups in the F1 vs. F2 space during months 17 through 26 and separation of these groups by month 29. Pearsall (1985) analyzed the formant frequencies estimated from narrowband spectra of the vowels /i ɑ u/ and concluded that certain resonances are produced with more variability in Bark frequency than others. In particular, resonances which require a Helmholtz-type vocal-tract configuration (e.g., F1 of /i/ and /u/) are produced less precisely than those which do not (e.g., F2 of /i/, F1 and F2 of /ɑ/).

Fundamental frequencies of children's vocalizations have also been measured. Kent and Murray (1982) found evidence of instability of laryngeal control in vocalizations produced by children of ages three, six and nine months. They observed tremor and abrupt changes in fundamental frequency (including doubling and halving). Keating and Buhr (1978) observed a wide range of fundamental frequencies and various phonation types in a study of fundamental frequencies of utterances produced by infants and by children of ages eight months to three years. The range of fundamental frequencies used by the children during modal phonation is 150 to 700 Hz. Although most of the utterances were produced in modal register, phonations in fry and high register were also observed.

Some studies have concentrated primarily on the temporal characteristics of children's babbling. Kent and Murray (1982) observed approximately the same durations of children's babbles as of syllables in adult speech. Stark (1980) observed that in reduplicated babbling "sufficient control over phonation is needed to coordinate it with repeated opening and closing of the vocal tract above the glottis." Others have defined acoustic properties of babbles (de Boysson-Bardies *et al.*, 1981) and have applied the label "canonical babble" to sequences of repeated, similar syllables (Oller, 1985).

1.3.4 Developmental interpretations

Many researchers of children's speech have documented changes which occur over time. A few have focused on the interpretation of these changes in terms of the development of a child's anatomy, motor-control abilities, or phonological representation. Buhr (1980) observed the differential development of the vowel space with age, with front vowels appearing before back vowels. He notes that "there exists considerable development of the vowel triangle during babbling" and relates his observations to increased articulatory skill and general physiological changes such as lengthening of the vocal tract. He remarks that it is surprising not to find a general lowering of the formants with age due to the increase in length of the vocal tract of the child. Lieberman (1980) notes that children produce vowels with "equivalent formant frequency patterns [to those of adults]...scaled to shorter vocal tract lengths." His formant data show increasing separation by vowel group in the F1 vs. F2 vowel space until age three years.

Some researchers focus on the onset of canonical babbling and regard this change as a milestone in the development of speech. Holmgren *et al.* (1985) have hypothesized that the regular timing component of canonical babbling might appear earlier in children's vocalizations than the onset of canonical babbling. This timing regularity might appear in children at approximately the same age as other rhythmic behaviors (Kent and Murray, 1982; Thelen, 1981).

Locke (1983) presents a comprehensive review of evidence of phonological development in children from many different language backgrounds. Phonological development can be assessed from evidence of the existence of a pattern in the mastery of phonemes (Menyuk, 1972; Ferguson and Farwell, 1975). Certain features are preserved in early word attempts, while others are not. Thus, one process in the development of speech in children might be the acquisition of mental representations in terms of distinctive features. Holmgren *et al.* (1983) report that transcriber agreement is higher when evaluated in terms of distinctive features than in terms of phonemes. These data are consistent with the view that a child's speech develops by mastering the production of the acoustic correlates of distinctive features rather than whole segments.

1.4 Outline of the modeling process

The goal of this research is to understand the relationship between the physical attributes of young children and the acoustic characteristics of their vocalizations. This goal is pursued by constructing models of the relationship between articulation and acoustic characteristics and by analyzing measurements of young children's vocalizations. The emphasis is the development of speech. Patterns in the measured acoustic characteristics are interpreted in terms of growth of physical structures, changes in articulatory control, and development of linguistic abilities.

In order to construct a model of the characteristics of children's speech, the physical attributes of a child's vocal system and the constraints on motor control must be known. These attributes include the dimensions of the lungs, trachea, glottis, vocal tract and head and the tissue properties of the glottal structure and vocal-tract walls. The ability of the child to control respiration and movement of articulatory structures must also be known. Data on physical attributes and constraints which are available in the literature on children's anatomy and motor-control abilities are summarized in Chapter 2. In the cases for which data on children are unavailable, parameter values are inferred from adult data. Models which predict fundamental frequencies, durations, and formant frequencies and bandwidths from the physical parameters are presented in Chapter 3.

A selection of vocalizations, i.e., sounds which are not cries or vegetative sounds, is analyzed in this research. The vocalizations are selected to exemplify the variety of sounds produced by children but not to form an exhaustive inventory. Data consisting of acoustic measurements reported by other researchers are used if available; in addition, selected recordings of a group of forty young American and four young Swedish children are analyzed. Oller (1985) has proposed guidelines for establishing a useful relationship between transcriptions and acoustic analyses. These guidelines are followed in this research.

The acoustic analyses are based on data measured from spectrograms and computer-generated time waveforms and spectral displays (Zue *et al.*, 1986). The relatively high fundamental frequency of children's speech influences the choice of analysis methods. The acoustic characteristics of children's speech which are similar to those of adults' speech are used as bases for comparison of sounds produced at different stages of development because the endpoint of the process is adult-like speech. Therefore, acoustic measures which are commonly applied to the analysis of adult speech are used in this research. The measurements are fundamental frequencies, formant frequencies and bandwidths, and durations of segments, syllables and utterances.

In Chapter 4, the measurements of young children's vocalizations are interpreted with reference to the models of the relationships between acoustic characteristics and articulatory parameters. The adequacy of the models is assessed; possible refinements of the models in terms of parameter selection and determination are proposed. The results of the comparison of measured data to predicted acoustic characteristics are discussed in terms of growth effects, individual variability, and phonological development. The relationships between the various acoustic characteristics are analyzed in Chapter 5. Areas in which further work is needed are outlined at the end of the chapter.

Chapter 2

Model parameters

The purpose of this chapter is to describe physical characteristics of the speech production systems of young children. The models of acoustic characteristics of children's speech, which are described in Chapter 3, use parameters based on some of these physical data. Some data not used for the calculations of model parameters are included in this chapter in order to provide a more complete description of physical characteristics with the hope that future analyses will find these data useful.

The models reflect the differences between the speech production systems of children and adults. The parameters of the models include physical attributes, such as dimensions and tissue properties, and dynamic constraints, such as rates of respiration and articulator movement. Parameter values are determined from relevant medical and speech literature. The collection of parameter values contained in the following sections is not exhaustive. The listing simply serves to outline differences between children and adults for the purpose of forming comparisons between the acoustic characteristics of the speech of children and the speech of adults.

Vocal systems of children are described in Section 2.1. Dimensions of the lungs, the subglottal airways and the vocal tracts of children and adults are summarized. Data on tissue properties of the vocal system and subglottal pressure are presented. The constraints imposed by children's motor-control abilities are outlined in Section 2.2. Rates of respiration and articulator movement are described, as are constraints on the precision of movement of articulatory structures. Values are assigned to parameters corresponding to the ages one, two and three years and adult based on reported measurements. In the cases in which no measurements are available at these ages, parameter values are estimated. The parameter values are tabulated at the end of this chapter.

2.1 Physical attributes

An obvious difference between children and adults is size. The dimensions of the structures of the vocal system significantly influence the acoustic characteristics of sounds. An understanding of the dimensional differences between the vocal structures of children and adults aids in the development of a model of the acoustic characteristics of children's speech. In addition to differences in size, children differ from adults in mechanical properties of vocal structures. Differences in subglottal pressure also affect some of the characteristics of the speech of children and adults.

There is a great deal of variation in experimental procedures for measuring physical attributes of individuals. Even within a single study, a wide range of values is often observed in measurements of different subjects. Difficulties in comparing the results of various studies arise due to differences in the ages of subjects and the scarcity of measurements for children. Many values for adults and some for newborns are reported. For children between the ages of a few months and a few years, data on dimensions, tissue properties, and motor-control abilities are sparse. When the results of a variety of studies are compared, monotonic increases or decreases with age are usually observed in these measurements. Apparently, humans grow steadily, changing relatively slowly in their physical characteristics. In light of this process of slow change, attribute values at some ages were interpolated from experimental values which were measured at other ages.

2.1.1 Dimensions

Data on the dimensions of vocal structures, from the lungs through the airways to the mouth, are summarized below. For each structure, information is given which describes the dimensions which most directly influence the acoustic characteristics of speech sounds.

2.1.1.1 Lungs: volumes

The function of the lungs in speech is to provide a reservoir of air. The lungs consist of a large number of small air sacs, or alveoli, connected by a root-like system of airways. The structure of the lungs can be neglected in calculations involving airflow; only the volume of air which is available for speech and the overall compliance of the lungs are important. The amount of air which is exhaled in a single breath during normal breathing is called tidal volume. Another quantity of air is vital capacity, or the maximum amount of air which can be exhaled after a maximum inspiration. Some researchers have also measured phonation volume, the amount of air exhaled during a single sustained phonation.

Lung volumes can be measured directly by a spirometer or indirectly by a body plethysmograph. A spirometer measures the volume of air rebreathed into a water-sealed container. A body plethysmograph consists of an air-tight box and a pressure gauge. A subject's body is placed in the box with only the head exposed, and the box is sealed around the base of the subject's head. As the subject breathes, changes in lung volume cause changes in body volume and therefore in the pressure in the sealed box. Changes in lung volume are determined from changes in pressure based on a constant product of *pressure* × *volume* (at constant temperature).

Tidal volume has been estimated by a number of researchers. Altman and Dittmer (1971) report tidal volumes by age and by weight. Tidal volumes of newborns are usually measured by a body plethysmograph. For children and adults, volumes are measured by a spirometer. Average tidal volumes for children of ages one, two and three years were converted from values reported by weight by using weight-by-age standards (Broadbent *et al.*, 1975). Altman and Dittmer's average values and corresponding ages are shown in Fig. 2.1. Hoshiko (1965) reports somewhat higher values of tidal volume for adults than do Altman and Dittmer: 680 cm^3 for females and 770 cm^3 for males. The subjects in Hoshiko's study were young adults. Standard textbook values for tidal volume of adult males range from about 500 cm^3 (West, 1974) to 750 cm^3 (Zemlin, 1968). The ranges of tidal volumes reported in the studies discussed above are listed in Table 2.1 for children and adults.

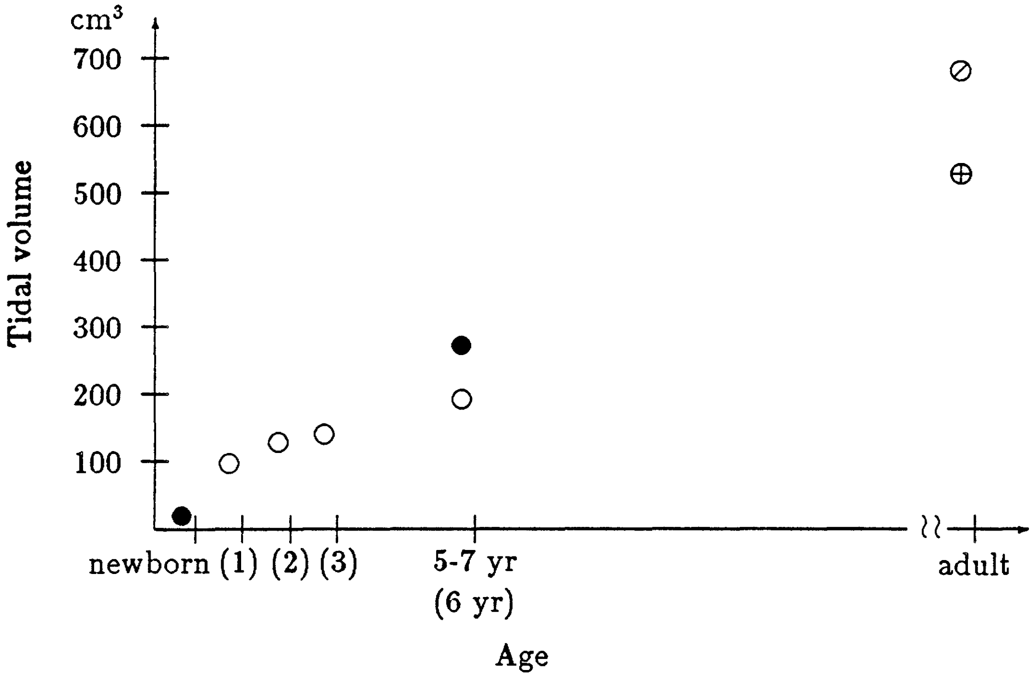

Figure 2.1: Tidal volumes of children and adults (from Altman and Dittmer, 1971). Children's values are shown by •'s (reported by age) and ∘'s (reported by weight); adult males', by ⊘; and adult females', by ⊕.

Vital capacity has been measured for both children and adults. An average value of 1500 cm^3 for ten-year-old children is listed by Altman and Dittmer. This value is based on data of different authors who used comparable methods. Beckett *et al.* (1971) measured a value of 1447 cm^3 for seven-year-old children (ten boys and ten girls). A spirometer equipped with a child's breathing mask was used by Beckett and his colleagues. Values of vital capacity ranging from 3520 cm^3 for adult females to 4800 cm^3 (overall adult average) are found in Altman and Dittmer's handbook. Hoshiko reports somewhat smaller values for vital capacity of 2850 cm^3 for young adult females and 4410 cm^3 for young adult males. These measurements were made with a spirometer; thirty males and thirty females participated as subjects. Zemlin's textbook cites vital capacities in the range of 3500 to 5000 cm^3 in adult males. Vital capacities of children of ages one, two and three years were estimated by multiplying the average vital capacity of an adult male by the ratios of tidal volume for young children to that of an adult male. These values are shown in the parameter table at the end of this chapter.

Table 2.1: Reported ranges of lung volumes of newborns, children, and adults. (See text for sources.) Ages estimated from weight-by-age tables are shown in parentheses. The letters ND indicate that no data are available.

Age	Lung volumes (in cm^3)	
	tidal volume	vital capacity
newborn	16 – 22	ND
(1 yr)	95	ND
(2 yr)	126	ND
(3 yr)	140	ND
5-7 yr	193	ND
7 yr	225	1447
adult (female)	380 – 680	2850 – 3520
adult (male)	590 – 770	4410 – 5000

Estimates of phonation volume vary considerably between studies and appear to depend on the measurement method. Beckett *et al.* measured phonation volume in seven-year-olds who had been instructed to phonate the vowel /ɑ/ for as long as possible. Hoshiko, on the other hand, instructed his young adult subjects to phonate a vowel whenever they felt ready. It is not surprising that the values measured by Beckett and his colleagues are considerably greater in relation to vital capacity than are Hoshiko's values. Beckett *et al.* report phonation volumes of 1069 cm^3 for girls and 1310 cm^3 for boys, or on average 82% of vital capacity. Hoshiko found a much smaller percentage of vital capacity to be used in phonation. He reports that on average phonation volume is 20% of vital capacity. The results of these two studies suggest upper and lower bounds on phonation volume with respect to vital capacity.

2.1.1.2 Subglottal airways: lengths and areas

The system of airways connecting the air sacs in the lungs to the vocal tract consists of a series of branching tubes. The largest of these tubes is the trachea, which divides into right and left main bronchi. Each bronchus divides into shorter, narrower tubes, which in turn further divide into even smaller tubes. The air in the trachea and the bronchi is part of the air tract for speech while the glottis is open. Thus, the movement of air in the subglottal airways can influence the acoustic characteristics of the speech waveform.

Lengths and cross-sectional areas of the trachea and bronchi of children and adults as reported by Altman and Dittmer (1971) are shown in Table 2.2. Ratios of children's dimensions to adults' are shown in Table 2.3. The acoustic mass of an airway is proportional to the ratio of length to area. The length-to-area ratio of the trachea is approximately 6.5 cm^{-1}, regardless of age. The length-to-area ratio of each main bronchus ranges from approximately 12.6 cm^{-1} for one-year-olds to 4.5 cm^{-1} for adults. A pair of bronchi has the same acoustic mass as one tube with a length-to-area ratio of 6.3 cm^{-1} in the case of one-year-olds, and 2.25 cm^{-1} for adults. Thus for one-year-old children, the acoustic mass of the bronchi is approximately the same as the acoustic mass of the trachea. For adults, though, the acoustic mass of the trachea is approximately three times as great as that of the bronchi.

Dimensional data on the smaller airways are scarce. Adult data indicate that the ratio of length to area of each of the smaller bronchi is approximately the same as the ratio for the main bronchi. In light of the large numbers of smaller bronchi at each level of branching, the acoustic mass of the system of these bronchi is negligible in comparison to the acoustic mass of the trachea and the main bronchi.

Table 2.2: Average lengths and areas of the trachea and the main bronchi of children and adults (from Altman and Dittmer, 1971).

Age	Trachea		Bronchi	
	length (in cm)	area (in cm^2)	length (in cm)	area (in cm^2)
1-2 yr	4.5	0.67	2.9	0.23
2-3 yr	5.0	0.80	2.9	0.31
3-4 yr	5.3	0.82	3.1	0.37
adult	12.0	1.91	5.4	1.20

Table 2.3: Scale factors for the trachea and the main bronchi of children with respect to adults.

Age	Tracheal scale factors		Bronchial scale factors	
	length	area	length	area
1-2 yr	0.38	0.35	0.54	0.19
2-3 yr	0.42	0.42	0.54	0.26
3-4 yr	0.44	0.43	0.57	0.31

2.1.1.3 Vocal folds: length, height and thickness

The dimensions of the vocal folds affect the pattern of vibration of the vocal folds. The variation in glottal opening with respect to time is called the glottal area function. During the production of some speech sounds, the glottis is open; for other sounds, the vocal folds are vibrating and the glottal area function shows a pattern of open and closed intervals. A few measurements of glottal area functions have been reported for adults (for example, Childers *et al.*, 1985); similar data for children are not available. A glottal area function for a child can be estimated from data on the length and tissue properties of children's larynges in conjunction with a model of the vibration of children's vocal folds.

The length of the vocal folds has been measured for newborns, children and adults. Several values are available for newborns and adults; fewer are reported for children between the ages of one and seven years. Hirano *et al.* (1980) report measurements of vocal-fold length, including both the membranous and cartilaginous portions, for males of various ages. Their measurements of the length of the membranous portion for newborns, one-, two-, three-, four- and ten-year-olds are listed in Table 2.4. Lengths for children and adults reported by Gedgoud (1900), Negus (1929), and Kahane (1975) are summarized by Goldstein (1980). In each case, the measurements were made from cadavers. Goldstein's figure shows lengths of 4 to 7.5 mm for the total vocal-fold length of children under the age of one year; these lengths most likely include both the membranous and cartilaginous portions of the vocal folds.

The thickness of the mucosa of the vocal folds has been measured by Hirano and his colleagues for newborns, children and adults. The vocal fold thickens somewhat with age, but the change in thickness is not as great as the change in length. No direct measurements of vocal-fold height are reported. The change in height is assumed to be comparable to the change in thickness. The effective thickness and area of the vocal folds were estimated, assuming a square cross-sectional area, from measurements of the vibrating mass of the vocal folds (see Section 2.1.2.3), tissue density (1.1 g/cm^3) and vocal-fold length.

2.1.1.4 Vocal tract: lengths and areas

Adjustments to the shape of the vocal tract alter the sound which is produced by airflow in the glottis or in the vocal tract. The properties of the walls of the vocal tract also modify the acoustic characteristics of sounds, but to only a minor degree compared to the effect of the shape of the vocal tract. Certain articulatory configurations, such

Table 2.4: Lengths of the membranous portion of the vocal folds of children and adults (from Hirano *et al.*, 1980).

Age	Vocal-fold length (in cm)
newborn	0.30
1 yr	0.35
2 yr	0.40
3 yr	0.45
4 yr	0.50
10 yr	0.80
adult (female)	1.50
adult (male)	1.80

as a narrow constriction, cause sounds to be generated in the vocal tract. It is not surprising that the acoustic characteristics of children's speech differ from those of adults' speech given the smaller size of children compared to adults.

Data on vocal-tract lengths and cross dimensions are available from a variety of sources, primarily the medical and dental literature. Measurements of the dimensions of the vocal tract during speech are sparse. The cross-sectional area of the vocal tract can be estimated from the positions of vocal structures during production of speech sounds. Articulator positions of adults are commonly determined from X-ray tracings. Due to the risks and difficulties involved in collecting data of these sorts, direct measurements of the areas of children's vocal tracts are unavailable. Areas can be inferred from anatomical data and from modeled configurations.

Goldstein (1980) compiled data describing vocal-tract lengths of children and adults. The data include oral-cavity, pharyngeal, and total vocal-tract lengths for a neutral configuration. Table 2.5 shows lengths for children of ages one, two and three years, together with values for adult males. Differences in lengths due to sex at age three years or younger are less than 3.1 mm, or 4% of the adult male vocal-tract length. In light of the small differences between boys' and girls' lengths, an average length was computed for each age as the mean of the girls' and boys' lengths.

Length scale factors were calculated as the ratio of child length to adult length. Scale factors for ages one, two, and three years are shown in Table 2.6. The length of the oral cavity of a young child more nearly approximates that of an adult than does the length of the pharynx. The pharyngeal cavity of a three-year-old is less than half as long as that of an adult; in contrast, the oral cavity is approximately three-quarters as long as that of an adult.

Table 2.5: Lengths of vocal tracts of girls and boys at ages one, two, and three years and of adult males (from Goldstein, 1980).

Age	Vocal-tract lengths (in cm)		
	pharynx	oral cavity	total
1 yr	3.6	5.7	9.3
2 yr	4.0	5.9	10.0
3 yr	4.3	6.1	10.4
adult	8.9	8.1	16.9

Table 2.6: Length scale factors, or ratios of vocal-tract lengths of a child to those of an adult.

Age	Scale factors for vocal-tract length		
	pharynx	oral cavity	total tract
1 yr	0.40	0.70	0.55
2 yr	0.45	0.73	0.59
3 yr	0.48	0.75	0.62

Goldstein also reports data describing various cross dimensions of children's vocal tracts. From these data, she calculated vocal-tract configurations corresponding to the productions of a neutral vowel by a newborn infant and by children of ages one and three years. Cross-sectional areas of the oral and pharyngeal regions estimated from Goldstein's model of a vocal tract in the configuration of a neutral vowel are shown in Table 2.7. The area of the mouth opening is assumed to be the same as the area of the oral cavity.

Area scale factors, or ratios of the area of a child's vocal tract to the area of an adult's, are shown in Table 2.8. It can be seen from these data that the cross-sectional area of the oral cavity of a one-year-old child is approximately one-third as large as that of an adult, while the pharyngeal area is about two-thirds as large. The average cross-sectional areas of children's vocal tracts are larger in comparison to adults' than would be predicted from simply squaring the longitudinal scale factor. Average heights and widths of children's heads (see discussion in the next section) are also larger relative to adult dimensions than are vocal-tract lengths. The short pharynx of a child accounts for most of the difference between scale factors for vocal-tract length and cross dimensions.

Table 2.7: Cross-sectional areas of the pharynx and oral cavity of children and adults for an assumed neutral-vowel configuration (from Goldstein, 1980).

Age	Vocal-tract areas (in cm^2)	
	pharynx	oral cavity
1 yr	2.3	1.5
3 yr	2.8	1.8
adult	3.5	4.2

Surface areas and volumes of the vocal tract can be estimated from data on lengths and cross-sectional areas. For a neutral-vowel configuration, the vocal tract can be approximated as a circular cylinder with a fixed cross-sectional area. The surface

Table 2.8: Scale factors for vocal-tract areas of children with respect to adults.

Age	Scale factors for vocal-tract areas	
	pharynx	oral cavity
1 yr	0.66	0.36
3 yr	0.80	0.43

area of an adult-sized tube is approximately 126 cm^2 and the volume is 74 cm^3. The approximate surface area and volume of the vocal tract of a one-year-old child are 36 cm^2 and 14 cm^3, respectively. The surface area of the vocal tract of a one-year-old is about one-fourth as large as the surface area of an adult tract, and the volume is nearly one-fifth as large.

2.1.1.5 Head: height, width and depth

The relative sizes of the mouth and the head of a speaker influence the radiation of the speech waveform. Head sizes can be estimated from outlines of dentofacial structures, such as the Bolton Standards (Broadbent *et al.*, 1975). Outlines of the head of a three-year-old child are overlaid on outlines of an adult head in Fig. 2.2. The maximum dimensions of the head were measured: the height, the width, and the distance from the forehead to the back of the skull. These dimensions are listed in Table 2.9. The width of the head for one-year-old children was estimated by multiplying the adult head width by the ratio of head depth for a one-year-old to that of an adult. The width for two-year-olds was calculated in a similar manner. These estimated widths are shown in the parameter table at the end of this chapter. The ratio of linear dimension, on average, of a three-year-old child's head to an adult's is about 0.9.

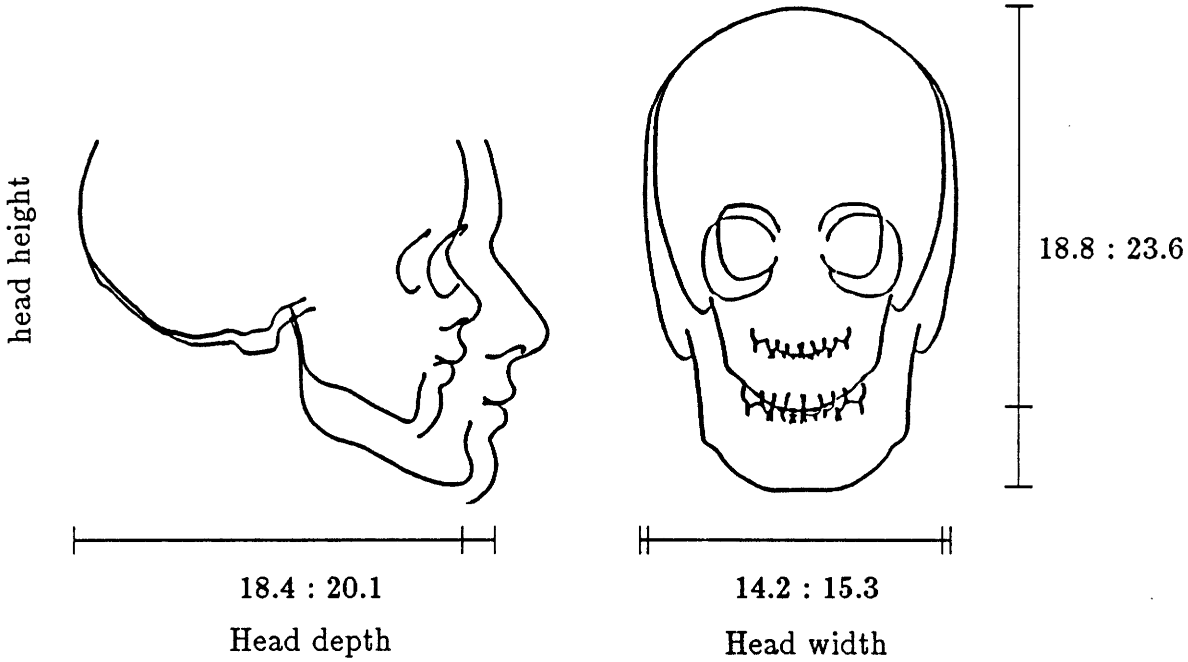

Figure 2.2: Outlines of a three-year-old's head overlaid on outlines of an adult's (sketches are from the Bolton standards, Broadbent *et al.*, 1975). The dimensions are in cm.

2.1.2 Tissue properties

The mechanical properties of the structures of the vocal system can influence the production of speech sounds. Although these properties play a lesser role compared to the effects of the dimensions of vocal structures and the pressures and airflows present in the vocal system, some tissue properties have an appreciable influence on the acoustic characteristics of certain sounds. Properties for which data exist include the compliances of the lungs, the vocal folds and the sub- and supraglottal-tract walls, and the masses of the vocal folds and the vocal-tract walls; these data are limited.

2.1.2.1 Lungs: compliance

The pressure in the lungs and the airflow from the lungs depends in part on the compliance of the lung tissues as well as the dimensions of the airways above the lungs. Lung compliance is defined as the ratio of the change in volume to the change in pressure, $\Delta volume/\Delta pressure$. In order to compare lung compliances of individuals of

Table 2.9: Standard dimensions of the head of children and adults (from Broadbent *et al.*, 1975) (ND: no data).

Age	Head dimensions (in cm)		
	height	width	depth
1 yr	17.8	ND	17.0
2 yr	18.3	ND	17.9
3 yr	18.8	14.2	18.4
adult	23.6	15.3	20.1

different sizes, a normalized lung compliance is often used. One definition of normalized compliance is the ratio of the change in volume divided by body weight to the change in pressure, or $(\Delta volume/weight)/\Delta pressure$. Lung compliance can also be normalized with respect to tidal volume.

Lung compliance has been measured in newborns and in adults. An average value of 2 $cm^3/kg/cm$ H_2O for newborns is reported by Tooley (1975). He measured lung volume using a body plethysmograph and calculated transpulmonary pressure as the difference between airway pressure (measured at the mouth) and esophageal pressure. Altman and Dittmer (1971) report compliances of 5.2 cm^3/cm H_2O for newborns, 45 cm^3/cm H_2O for five-year-olds, and 262 cm^3/cm H_2O for adults. For average weights of 3 kg for newborns, 21 kg for five-year-old children, and 70 kg for adults (Broadbent *et al.*, 1975), Altman and Dittmer's values can be normalized to approximately 1.7, 2.1, and 3.7 $cm^3/kg/cm$ H_2O for newborns, five-year-olds, and adults, respectively.

Tidal volume can also be used as a normalization factor for lung compliances. The compliance measurements were normalized with respect to tidal volumes of 19, 185 and 680 cm^3 for newborns, five-year-olds and adults, respectively. Lung compliances normalized for volume were interpolated for ages one, two and three years from the values for newborns and five-year-olds and are shown in the parameter table at the end of this chapter.

Lung compliances, normalized with respect to both body weight and tidal volume, are shown in Table 2.10. These values indicate that children's lung tissue is stiffer than the tissue of adults' lungs.

Table 2.10: Average values of lung compliance and values normalized with respect to body weight and with respect to tidal volume of children and adults. (See text for sources.)

Age	Lung compliance	Normalized lung compliance	
		by weight	by tidal volume
	(in $cm^3/cm\ H_2O$)	(in $cm^3/kg/cm\ H_2O$)	(in $(cm\ H_2O)^{-1}$)
newborn	5.2	1.7	0.27
5 yr	45.0	2.1	0.24
adult	262.0	3.7	0.39

2.1.2.2 Subglottal airways: mass of walls

The mass of the walls of the subglottal airways influences the frequencies of the subglottal resonances, particularly the lowest resonance. When the glottis is relatively open, the sub- and supraglottal tracts are coupled and the subglottal resonances are weakly apparent in the speech spectrum. The mass of the walls of the subglottal airways can be inferred from the input impedance of the subglottal system. Measurements of the input impedance of adults' subglottal systems have been reported in the literature; similar data for children are not available.

Ishizaka *et al.* (1976) measured the input impedance of the subglottal system in five laryngectomized subjects. A value for the effective mass per unit area of the walls of the trachea and main bronchi can be calculated from the measurements of the resonances of the subglottal system and from estimates of the acoustic compliance of the air in the lungs and the subglottal airways. Ishizaka and his colleagues report a value of 0.3 g/cm^2 for the effective mass per unit area of the subglottal airways. Fant *et al.* (1972) used the Ishizaka data in analyses of spectra of aspirated sounds. They modeled the effective mass of the subglottal system with a lumped acoustic mass of 0.0037 g/cm^4 and computed spectra which include extra formants due to coupling to the subglottal cavities. These spectra are similar to spectra of aspirated intervals of natural speech.

2.1.2.3 Vocal folds: mass and stiffness

The mass and stiffness of the vocal folds influence the pattern of vibration of the folds. These physical parameters, as well as the dimensions of the vocal folds and the pressure drop across the glottis, determine the shape of the glottal pulse and the fundamental frequency of vocal-fold vibration.

The mass of the vocal folds depends on the density of the tissue and the dimensions of the vocal folds. The composition of various tissues has been determined for adult males (Allen *et al.*, 1959). Muscle, tendon, ligament and epithelium have approximately the same density, 1.1 g/cm^3. The structure of children's vocal folds is different from the structure of the vocal folds of adults. Hirano (1980) observes that no vocal ligament is found in infants' vocal folds. An immature vocal ligament is apparent in children of age four years; the vocal ligament does not fully develop until after puberty. The tissue composition of children's vocal folds differs from that of adults', but the densities of both the muscle and epithelial layer of children's vocal folds are approximately equal to the densities of the muscle, ligament and epithelial layer of the vocal folds of adults. Thus, the density of the vocal folds of children can be assumed to be approximately the same as the density of adults' folds.

It is reasonable to assume that the vibrating mass of the vocal folds is proportional to the combined mass of the thyroarytenoid (TA) and lateral cricoarytenoid muscles (LCA) and the vocal ligament. Kahane and Kahn (1984) report the mass of these muscles as 0.874 g for adults and 0.076 g for infants. These mass values are shown in Table 2.11. Kaneko and his colleagues (1987) estimated a vibrating mass of 0.141 g for adult vocal folds. Based on the ratio of the vibrating mass to the measured mass of the TA and LCA muscles for adults and the measured mass of the TA and LCA for infants, the mass of the vibrating portion of the vocal fold was calculated for infants. Values for one-, two- and three-year-old children were interpolated from the vibrating masses of infants and adults and are shown in the parameter table at the end of this chapter.

Table 2.11: Values of the mass of the vocal folds of infants and adults (ND: no data). (See text for sources.)

Age	Vocal-fold mass (in grams)	
	Muscle mass	Vibrating mass
infant	0.076	ND
adult (male)	0.874	0.141

The stiffness K of a particular volume of material is related to Young's modulus, E, and the area over which the force is applied and the length of the material in the direction of the force. Young's modulus is the ratio of stress to strain. Measurements of longitudinal stress versus strain for vocal-fold tissue of young dogs and adult dogs were performed by Perlman and Titze (1983). They found that the vocal-fold tissue of young dogs is stiffer than the tissue of adult dogs. Using Perlman and Titze's graphs of stress vs. strain, values of 6.4×10^5 dynes/cm^2 for old tissue and 8.1×10^5 dynes/cm^2 for young tissue were calculated. Young's moduli are shown in Table 2.12. For young tissue compared to old tissue, the ratio of Young's moduli is 1.27.

During vocal-fold vibration, force is applied laterally to the vocal folds. For a lateral force on the vocal folds, the stiffness K is given by

$$K = E \; \frac{height \times length}{thickness} \quad , \qquad (2.1)$$

where *height*, *length* and *thickness* are dimensions of the vocal folds. An effective vocal-fold stiffness of 7.4×10^4 dynes/cm has been reported for adult humans by Kaneko and his colleagues. The Young's modulus which corresponds to this stiffness for adult vocal-fold dimensions is 4.1 × 10^4 dynes/cm^2. The modulus for children is computed by multiplying the modulus for adult tissue by the moduli scale factor calculated from the data of Perlman and Titze. The moduli assumed for children and adults are listed in the parameter table at the end of this chapter. These values are consistent with the range of transverse moduli of vocal-fold tissue reported by Kakita *et al.* (1980). The effective stiffness of young vocal folds is the adult effective stiffness multiplied by the scale factor for Young's moduli and the dimensions of the vocal fold.

Table 2.12: Values of the Young's moduli of the vocal folds of young and adult dogs (from Perlman and Titze, 1983).

Age	Young's moduli (in dynes/cm^2) of vocal-fold tissue
young dog	8.1 × 10^5
adult dog	6.4 × 10^5

2.1.2.4 Vocal-tract walls: mass and compliance

The impedance of the walls of the vocal tract influences the frequencies of the resonances of-the supraglottal system. Direct measurements of the impedance of various tissues at the surface of the body have been made. Franke (1951) measured the properties of the tissues of the thigh by means of a vibrating piston. Ishizaka *et al.* (1975) used a similar procedure to measure the mass and compliance of the neck, cheek and forearm. The mass of the walls of the vocal tract has also been determined indirectly from measurements of the impedance of the vocal tract and from estimates of the compliance of the air in the oral cavity (Fant *et al.*, 1976). In all of these experiments, data were collected for adult males only.

The direct and indirect measurements of the mass of the tissues surrounding the vocal tract are in general agreement. Ishizaka and his colleagues determined values for the mass of the neck and relaxed cheek of 2.4 and 2.1 g/cm^2, respectively. Fant *et* *al.* estimated a value of 2 g/cm^2 for the mass of the tissue of the vocal-tract walls. An average value of 2.2 g/cm^2 is listed in the parameter table at the end of this chapter. Values for children are calculated by multiplying the adult value by a tissue-thickness scale factor, or ratio of vocal-fold thicknesses of children and adults.

The only values reported for the stiffness of the vocal-tract walls are those of Ishizaka *et al.* (1975). They report values of 8.5×10^4 dynes/cm^3 for the cheek, and 4.9×10^4 dynes/cm^3 for the neck.

2.1.3 Subglottal pressures

The pressure below the glottal opening is a factor in determining the airflow through the glottis. Subglottal pressure can be measured most accurately in adults by puncture of the throat. Direct measurements of subglottal pressure have been made only infrequently because of the risks involved in throat puncture.

Measurements have been made of the pressures in the adult trachea and esophagus. Kunze (1964) measured both of these pressures and found that intratracheal pressure approximately equals esophageal pressure for moderate lung volumes. He measured intratracheal pressures directly while subjects phonated the vowels /æ ɑ o ʊ i/. Lieberman (1968) reiterates the argument of Bouhuys *et al.* (1966) that the difference between esophageal and subglottal pressure is due to the elastic recoil of the lungs. His data show that at the beginning of an utterance, just after an inspiration, the esophageal pressure is approximately the same as subglottal pressure. Lieberman's pressure recordings show values of 8 to 10 cm H_2O for the subglottal pressure used during speech. He measured an average intratracheal pressure of 6.1 cm H_2O in ten male subjects. Muta (1983) reports values of subglottal pressure ranging from approximately 5 to 11 cm H_2O, with an average of 8 cm H_2O, during a 3-second interval of phonation of a vowel. The range of subglottal pressures reported for adult speech is shown in Table 2.13.

No measurements of subglottal pressure during vocalization are available for children. Tooley (1975) reports values of subglottal pressure during lung inflation for infants. The range of subglottal pressures measured by Tooley is shown in Table 2.13.

Table 2.13: Ranges of subglottal pressures in infants and adults. (See text for sources.)

Age	Subglottal pressure (in cm H_2O)
newborn (during lung inflation)	2 – 20
adult (during speech)	5 – 11

Subglottal pressure can be estimated from values for lung compliance and the dimensions of the lungs. Measurements of the compliance of the lungs of children show that children's lungs are stiffer than adults' lungs (see discussion in Section 2.1.2.1). Therefore it is reasonable to expect that the pressure in children's lungs is greater than the pressure in the lungs of adults for the same conditions of lung inflation. Lung compliance is the ratio of the change in volume to the change in pressure. Normalized lung compliance is the ratio of the change in volume divided by the tidal volume to the change in pressure. The scale factor for lung compliance (normalized by tidal volume) can be used to estimate the subglottal pressure which children use during vocalizing from the subglottal pressure which adults use during speech. Lung compliances normalized by tidal volume for one-, two- and three-year-olds were interpolated from the values for newborns and five-year-olds, which are listed in Table 2.10. Scale factors, or ratios of lung compliance of a child to that of an adult, multiplied by an average value of 8 cm H_2O for the subglottal pressure of adults give subglottal pressures of 11.8, 12.1 and 12.4 cm H_2O for one-, two- and three-year-olds, respectively.

2.1.4 Sound pressures

The sound pressure at a distance from a speaker is proportional to the total acoustic volume velocity at the speaker's mouth, nose and vibrating surfaces of the cheeks and throat. Sound pressure levels due to various sources are reported by Beranek (1954).

For an adult speaking in a normal voice, the overall level is approximately 63 dB *re* 0.0002 dyne/cm^2 at 1 m. Peak levels are as high as 73 dB, when read with a fast C scale. Comparable data for young children are currently not available.

2.2 Motor-control constraints

Speech is in part a physical activity and as such is constrained by the limitations of the motor-control system. Linguistic and cognitive factors impose further constraints on the production of speech. Speech motor skills include control of timing, such as the rate of respiration and of movement, and precision of movement, as in the positioning of articulators. Neural and motor control in infants and young children appears to develop gradually. This development can be related in general, descriptive terms to improvements in skills needed for speech production (Netsell, 1981).

2.2.1 Rates of respiration

Breathing patterns of children and adults have been reported in terms of the number of breaths per minute and the volume flow of air per minute (Altman and Dittmer, 1971). The volume of air breathed in one minute by a newborn is approximately six times less than the volume breathed by an adult. Tidal volumes of infants and adults are shown in Table 2.1. A newborn takes approximately four times as many breaths per minute as does an adult. By age five years, a child's breathing rate has slowed to about twice as fast as an adult's. Typical rates of respiration for children and adults are shown in Table 2.14.

2.2.2 Rates of crying

Cries and speech are both produced by controlling the respiratory system and the laryngeal and oral structures, and are both subject to constraints of the motor-control system. Much research has documented the characteristics of infant cries (see

Table 2.14: Average rates of respiration and volume flow of children and adults (from Altman and Dittmer, 1971) (ND:no data).

Age	Rates of respiration	
	breaths/min	liters/min
newborn	48	1
5-7 yr	24	ND
adult	12	6

summaries in Stark and Nathanson, 1973; Sirviö and Michelsson, 1976; Michelsson *et al.* (1982)). Golub (1980) reports measurements showing expiration intervals in the range of 1.1 to 3.6 seconds for the cries of normal infants. From these data, the average repetition rate of cry intervals is about 0.4 Hz or 24 per minute.

2.2.3 Rates of movement

Rates of movement can be measured in terms of the number of repetitions during an interval of time or the velocity of the movement. Investigators of infant motor behavior have observed a general tendency for movements to form repetitive patterns from an early age. Patterns of limb movements involve relatively gross control; tasks such as tapping or vocalizing involve finer control. Diadochokinesis, or syllable repetition, is often used as an indicator of the ability to move oral structures. Dickson (1962), though, argues that fast diadochokinetic rates do not necessarily imply articulatory proficiency. However, the maximum rate of repetition at which a child can tap or repeat a syllable relative to an adult rate is related in part to limitations of the motor-control system.

Diadochokinetic rates, or rates of repetition of the syllables [pə], [tə] or [kə] or the string [pətəkə], are frequently reported in the literature. Prins (1962) measured diadochokinetic rates for a group of children of ages three to seven years: 3.5 Hz on average for strings of [pə],[tə] or [kə] and 3.3 Hz for repetitions of the syllables in the string [pətəkə]. Irwin and Becklund (1953) report a rate of 3.4 Hz on average for strings of [pə], [tə] or [kə] for children of age six years. A rate of 2.9 Hz was found by Weiner (1972) for children repeating strings of syllables with initial consonant sequences of /p,t/, /k,t/ and /p,t,k/. Fletcher (1972) also reports data for children of age six years: 3.9 Hz for [pə], [tə] or [kə], and 2.9 Hz for [pətəkə]. Older children are able to repeat syllables at a faster rate. Dworkin and Culatta (1985) tested a group of children whose average age was 7.9 years and found an average rate of 4.7 Hz for strings of [pə], [tə] or [kə]. Adults can repeat these syllables somewhat faster. For [pə], [tə] or [kə], Lass and Sandusky (1971) report rates of 6.0 Hz. Fletcher (1972) measured 6.5 Hz; and Tiffany (1980), 6.8 Hz. For repetitions of [pətəkə], rates of 7.5 Hz (Fletcher, 1972; Tiffany, 1980) have been measured. All of these diadochokinetic rates are shown in Fig. 2.3.

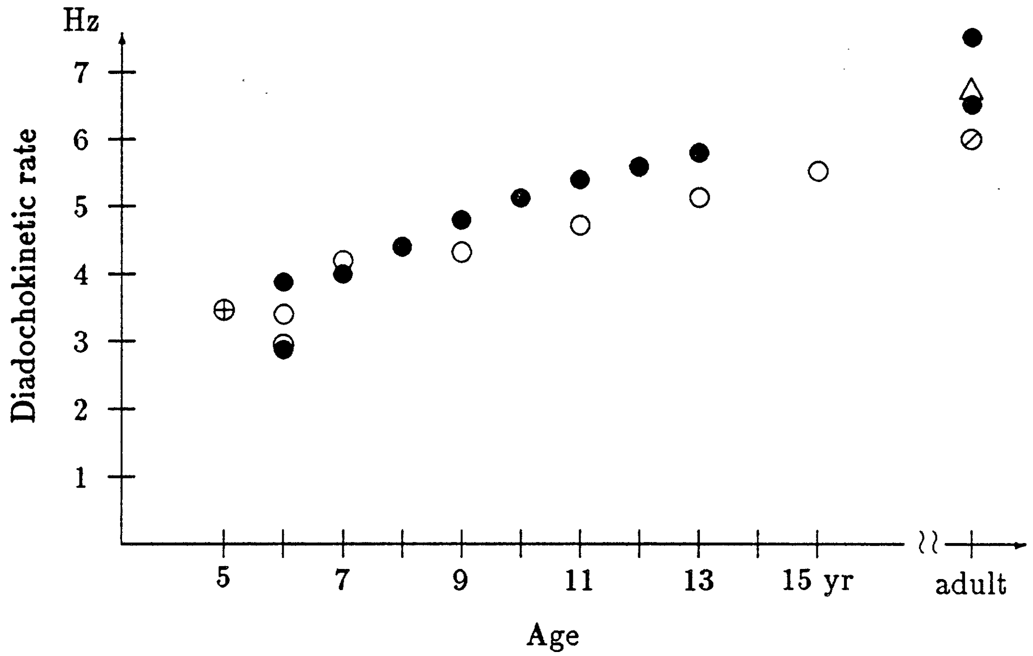

Figure 2.3: Diadochokinetic rates of children and adults reported by Irwin and Becklund (1953) are shown by ∘; Prins (1962), by ⊕; Weiner (1972), by ⊖; Lass and Sandusky (1971), by ⊘; Fletcher (1972), by •; and Tiffany (1980), by △.

The development of articulatory control can also be inferred from differences between children and adults in movement rates for other structures. Kent (in preparation) reports Fenn's (1938) observation that the maximum repetition rate "for speech is similar to the maximum rate of cyclic movements for finger wagging or forearm swinging." Finger tapping skills have been measured by Irwin and Becklund (1953) and Tingley and Allen (1975). Irwin and Becklund report a rate of 3.5 Hz at age six years for tapping as fast as possible. Children are able to tap faster with increased age, as shown in Figure 2.4. Tingley and Allen instructed children between the ages of five and eleven years and adults to tap at a comfortable rate. They note that at five years of age the average rate of comfortable tapping is slower than the average rate chosen by adults. At ages seven and nine years, children chose progressively faster tapping rates. By eleven years, children tapped at a rate which is within the normal adult range. The speed of self-paced tapping changes with age. Five-year-olds tap slower than older children and adults (Schellekens *et al.*, 1984). Todor and Kyprie (1980) found that adults could tap as fast as 6.6 Hz when tapping with the dominant hand. Diadochokinetic rates for children of ages one, two and three years were estimated by linear extrapolation from the rates for older children.

The rhythmic movements of the limbs that have been measured for children and adults show no difference based on age. Rates of arm movements, calculated for the task of placing pellets in a bottle as fast as possible, are reported by Prins (1962). He found the average (both arms) rate of movement to be 0.6 Hz. The children in the Prins study were between the ages of three and seven years. Thelen (1981) reports movement rates for children as young as six weeks. Rhythmic movements such as rocking and waving occur at rates in the range of 1–4 Hz for ages between 20 weeks and one year. Activities such as kicking are performed at the rate of approximately 0.6 Hz.

The movement patterns of articulators and fingers all show increases in speed with increasing age. From these data, it appears that young children are not able to move their articulators as quickly as are adults. Average rates of repetition are listed in Table 2.15. Oral repetition rates of 2.7, 3.0 and 3.2 Hz for one-, two- and three-year-olds, respectively, were interpolated from the values for older children. These values are listed in the parameter table at the end of this chapter.

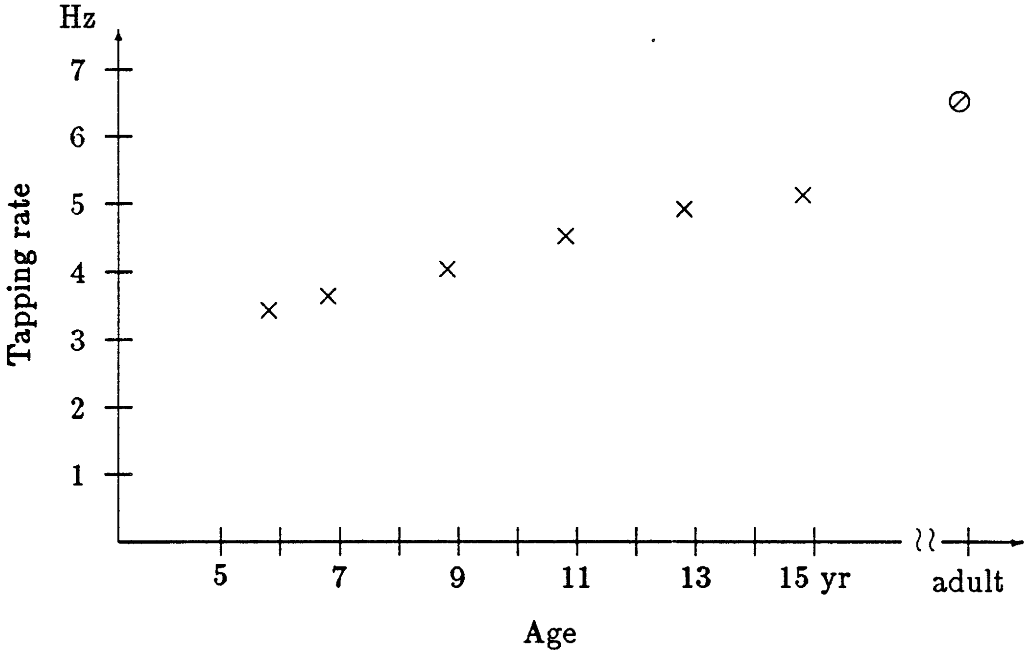

Figure 2.4: Tapping rates of children reported by Irwin and Becklund (1953) are shown by ×'s; rates of adults reported by Todor and Kyprie (1980), by ⊘.

Movements of the jaw, lips and tongue have been documented during production of various sounds. Fujimura (1961) reports average values in the range of 40 to 90 cm/sec over the first five milliseconds of opening of the lips for the initial consonants /p,b,m/. For the tongue body, Ostry and colleagues report 7.5 cm/sec (1983). Folkins (1981) reports a maximum of 8 cm/sec for jaw closing. Folkins and Linville (1983) show movement profiles from which velocities can be estimated: 18 cm/sec for the lower lip. Gay's (1977) movement tracks imply somewhat greater velocities: 25 cm/sec for the lower lip and 31 cm/sec for the tongue tip and body. Differences in experimental procedure might explain the wide variation in these measured rates.

Sharkey and Folkins (1985) report that the variability of the duration of lip and jaw opening is greater for children than for adults. The variability of lip displacement also decreases with age, but the jaw movement was consistent across the age groups.

Table 2.15: Average repetition rates of oral structures, fingers and limbs of children and adults (ND: no data). (See text for sources.)

Age	Rates of repetition (in Hz)		
	oral structures	fingers	limbs
1 yr	ND	ND	0.6 – 4.0
5 yr	3.4	ND	0.6
6 yr	ND	3.5	1.0 – 4.0
adult	6.5 – 7.5	6.6	ND

2.2.4 Precision of movement

Precision of movement is a skill which children must develop in order to articulate sounds. Arm and hand movements which require accuracy of movement have been measured in children. Fruhling and Basmajian (1969) found that four- to five-year-olds could be trained to consciously control individual motor units. Their training procedure did not produce evidence of learning in children of ages two to three years. Tingley and Allen (1975) found that consistency of finger tapping improves through the age range of five to eleven years. Hay (1979) found that children of age five years could point to within 5% of the distance to a moving target; this precision is close to adult performance. This ability decreased to a 15% error by age seven years and then improved to near an adult level of accuracy by age eleven years. Von Hofsten (1983) tested children of age 35 weeks in a reaching task. The children in his study timed their reaches within 50 msec of the arrival time of the target. Hofsten concluded that the development of timing control needed to catch an object is developed early in life. Apparently, the development of precision of execution of motor skills varies with type of skill. Some abilities are mastered by age one year, while others continue to develop throughout childhood.

2.3 Summary of parameter values

Values for the dimensions and properties of the vocal system are shown in Table 2.16 for children of ages of one, two and three years and for adults. In the cases for which data are not available at these ages, values were interpolated or extrapolated by the methods described in the corresponding sections of this chapter.

Table 2.16: Summary of parameter values of newborns, children, and adult males. Estimated values are shown in parentheses. (See text for sources.)

Parameter	Age				
	1 yr	2 yr	3 yr	adult male	
Dimensions					
Lungs: vital capacity	(640)	(840)	(940)	4550	cm^3
tidal volume	95	126	140	680	cm^3
Vocal folds: length (membranous)	0.35	0.40	0.45	1.80	cm
cross-sectional area	(0.048)	(0.058)	(0.063)	(0.073)	cm^2
thickness	(0.22)	(0.24)	(0.25)	(0.27)	cm
Vocal tract: length	9.3	10.0	10.4	16.9	cm
cross-sectional area	(1.9)	(2.1)	(2.3)	(3.9)	cm^2
surface area	(45)	(51)	(56)	(118)	cm^2
volume	(18)	(21)	(24)	(66)	cm^3
Head: height	17.8	18.3	18.8	23.6	cm
width	(12.9)	(13.6)	14.2	15.3	cm
depth	17.0	17.9	18.4	20.1	cm
Tissue characteristics					
Lungs: normalized compliance	(0.27)	(0.26)	(0.26)	0.39	$(cm\ H_2O)^{-1}$
Vocal folds: vibrating mass	(0.018)	(0.025)	(0.031)	0.141	g
Young's modulus	(5.2)	(5.2)	(5.2)	(4.1)	$\times 10^4$ dyne/cm^2
effective stiffness	(1.8)	(2.1)	(2.3)	7.4	$\times 10^4$ dyne/cm
Vocal-tract walls: mass	(1.80)	(1.94)	(2.05)	2.20	g/cm^2
Subglottal pressure	(12)	(12)	(12)	8	cm H_2O
Articulator repetition rate	(2.6)	(2.8)	(3.0)	6.0	Hz

Chapter 3

Model predictions

Some differences between children's vocalizations and adults' speech are due simply to differences in the size of the articulatory structures and control of articulation and respiration. Others are due to the stage of development of the child's linguistic and cognitive systems, or to a combination of these factors. Differences in physical characteristics of the speech production system between children and adults are discussed in Chapter 2. In this chapter, acoustic characteristics of children's utterances are predicted from parameters describing their speech production systems – the sizes of children's lungs, larynges and vocal tracts and their ability to control respiration and articulator movement.

The frequency of vocal-fold vibration influences various characteristics of children's utterances. Airflow from the lungs is modulated by the glottal vibration, and thus the maximum duration of airflow is determined in part by the pattern of glottal opening and closing. The periodic vibration of the vocal folds is reflected in the harmonic structure of the spectra of vowel-like and of some consonant-like sounds. Predictions of children's fundamental frequencies from models and physical parameters are developed in Section 3.1.

Temporal characteristics of children's vocalizations are influenced in part by the ability of children to control and to coordinate the muscles of respiration and of the laryngeal and supralaryngeal structures. By varying the degree of openness of the vocal tract in conjunction with appropriate control of respiration, a child can produce speech-like sequences of relatively open or vowel-like vocalizations alternating with lower amplitude, consonant-like sounds. The durations of a child's utterances are constrained by physical characteristics of the child's speech production system and by the linguistic and cognitive abilities of the child. In Section 3.2 durations of children's utterances are predicted from lung sizes and airflow rates, and consonant-like and vowel-like intervals are predicted from rates of articulator movement.

Spectra of vocalizations can be predicted from acoustic theory given data on properties of the larynx and vocal tract. The dimensions of the vocal tract, including lengths and cross-sectional areas, the attributes of the walls, and the characteristics of the terminations of the vocal tract determine the spectral shape of speech sounds. Section 3.3 develops predictions of formant frequencies and bandwidths of vowel-like sounds of young children.

3.1 Fundamental frequency

Various models, including the vibrating string and spring-mass (for example, Ishizaka and Matsudaira, 1968, and Flanagan and Landgraf, 1968), have been put forth to account for the fundamental frequencies of speech. Actual values measured during phonation are slightly higher than the natural frequencies of the vocal-fold structure (Kaneko *et al.*, 1987).

The models have been useful in predicting the natural frequencies of the vocal folds of adults. The vibrating string and spring-mass models differ in their predictions of the frequency of a child's vocal folds compared to that of an adult's. The vibrating string model is one-dimensional; its parameters are vocal-fold length and tension. The model predicts the general trend of higher fundamental frequencies of children's speech than adults', due primarily to the differences in the lengths of children's and adults' vocal folds. For the vibrating string model the scale factor SF_{string}, or ratio of fundamental frequency of a child to that of an adult male, depends on vocal-fold length (L) and tension (T):

$$SF_{string} = \frac{F0_c}{F0_a} = \frac{L_a}{L_c}\sqrt{\frac{T_c}{T_a}} \quad . \tag{3.1}$$

The subscripts a and c refer to adult and child values, respectively. Assuming that the tensions T_c and T_a of the vocal folds of children and adults are approximately the same, $SF_{string} \approx 5.1$ for a one-year-old child.

Spring-mass models represent the vocal folds in terms of lumped elements for the mass, stiffness and losses of the vocal-fold structure. These models have been successful in predicting values for the fundamental frequencies of adult speech and for airflow through the glottis during vocal-fold vibration. The effective mass and stiffness of the vocal folds are the parameters used in spring-mass models. The scale factor for the fundamental frequency predicted by a spring-mass model is

$$SF_{spring-mass} = \sqrt{\frac{K_c}{K_a}\ \frac{M_a}{M_c}} \quad , \tag{3.2}$$

where K is the effective stiffness of the vocal-fold tissue and M represents the vibrating mass of the vocal fold. Solving for K in terms of the Young's modulus E and the dimensions of the vocal folds gives

$$SF_{spring-mass} = \frac{F0_c}{F0_a} = \sqrt{\frac{E_c}{E_a}\ \frac{h_c L_c}{b_c}\ \frac{b_a}{h_a L_a}\ \frac{M_a}{M_c}} \quad , \tag{3.3}$$

where h and b are the vocal-fold height and thickness, respectively. Assuming the same ratio of child to adult value for both cross dimensions h and b and substituting ρLhb for M, the scale factor for the spring-mass model reduces to

$$SF_{spring-mass} = \frac{b_a}{b_c}\sqrt{\frac{E_c}{E_a}} \quad . \tag{3.4}$$

For the values listed in Table 2.16, $SF_{spring-mass} \approx 1.4$ for the fundamental frequency of a one-year-old child compared to an adult.

The laryngeal structure of a young child is different from an adult larynx in several characteristics. The vibrating string and spring-mass models capture important aspects of vocal-fold vibration, but fail to adequately model some aspects of the vocal-fold anatomy. These models do not adequately reflect the structure of the larynx of a child and are not as appropriate for children as they are for adults. Children's vocal folds are relatively shorter and thicker than adults', as shown in Fig. 3.1a. The attachments of the vocal folds to the arytenoid and thyroid cartilages can be expected to affect the stiffness of the vocal folds of children more than for the adult case. The vibrating string model does not take into account the effect of the cross dimensions of the vocal folds on the stiffness of the structure. Another shortcoming of this model is that discontinuities in slope at the juncture of the cartilages and the vocal-fold tissue are allowed. The spring-mass model allows for discontinuities in both position and slope at the endpoints of the vocal folds. The specification of boundary conditions is important in analyses of the vibration of children's vocal folds.

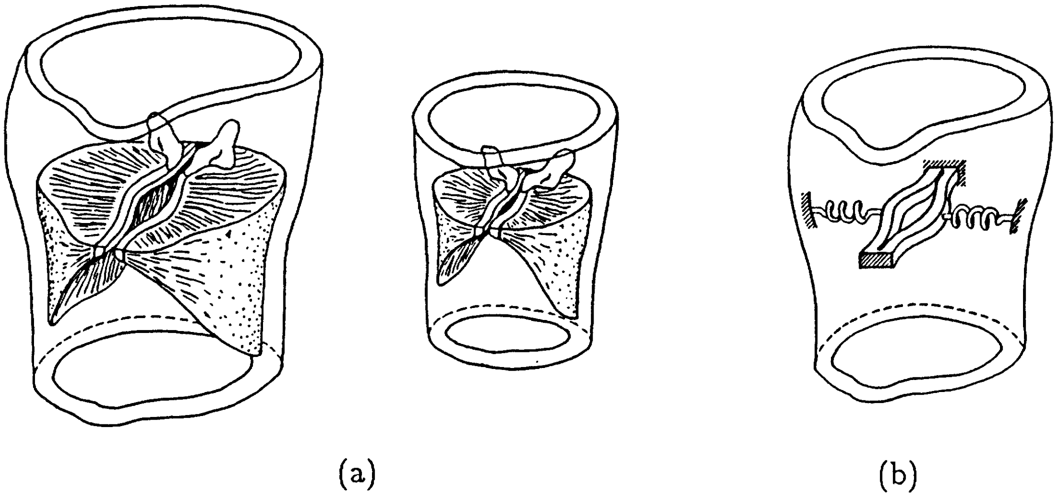

Figure 3.1: (a) Adult's and child's vocal-fold structures (not drawn to scale). (b) Bending beam model of vocal folds.

A model of vocal-fold vibration which reflects the anatomical structure of children's vocal folds is the bending beam model (Bickley, in press). This model has been useful for predicting the vibratory motion of relatively stiff structures which are attached rigidly at their ends and vibrate a small amount in the transverse direction (Woodson and Melcher, 1968). In order to model children's vocal folds, the standard bending beam model has been augmented by the addition of a distributed stiffness along one side. Figure 3.1b shows a bending beam which is fixed at both ends and which is coupled to material on one side by means of a spring. The fixed ends model the attachment of the vocal fold to the arytenoid and thyroid cartilages. The spring models the lateral stiffness of the vocal-fold tissue.

The fundamental frequency ω of the bending beam model results from solving the equation for transverse motion of the vocal fold,

$$\left(\frac{Eb^2}{12\rho}\right)\frac{d^4\xi}{dx^4} - \left(\omega^2 - \frac{K}{\rho Lbh}\right)\xi = 0 \quad , \tag{3.5}$$

where E represents the Young's modulus of the vocal fold, K models the stiffness of the vocal-fold tissue, b and h are the thickness and height of the vocal fold, ρ is the density of the tissue, and ξ is the transverse displacement of the fold. Four boundary conditions are imposed: continuity of displacement and of slope at both ends of the vocal fold. These conditions model the attachment of the vocal-fold tissue to the arytenoid, thyroid and cricoid cartilages. The homogeneous solution is a linear combination of trigonometric and hyperbolic functions. The values of ω for which equation (3.5) has a solution are given by

$$\omega^2 = \frac{\beta^4}{12}\,\frac{Eb^3h}{L^3M} + \frac{K}{M} \quad , \tag{3.6}$$

where β satisfies

$$\cos\beta - \frac{1}{\cosh\beta} = 0 \quad . \tag{3.7}$$

The variable β takes on discrete values which are found by graphical solution; the lowest non-zero value of β is approximately 4.73. A derivation of these equations is given in Appendix A. Numerical values for ω can be found by substitution of the dimensions and tissue properties of the vocal folds into equation (3.6).

Equation (3.6) can be written to show the combined contributions of the beam and the spring character of the vocal fold structure:

$$\omega = \sqrt{\omega_b^2 + \omega_s^2} \quad . \tag{3.8}$$

The first term under the square root in equation (3.8) is the square of the frequency of vibration due to the characteristics of the beam (assuming no lateral stiffness), or

$$\omega_b = \sqrt{\frac{\beta^4}{12}\ \frac{Eb^3h}{L^3M}} \quad . \tag{3.9}$$

The second term is the square of the frequency of the spring-mass model of the vocal fold, or

$$\omega_s = \sqrt{\frac{K}{M}} \quad . \tag{3.10}$$

For small L, as in the case of a child's vocal folds, the ω_b term dominates the expression for ω, and $\omega \approx \omega_b$. The vibration of a child's vocal folds is like that of a bending beam. For a child-sized vocal fold with length 0.35 cm, height and thickness 0.22 cm, mass 0.018 g, and stiffness 1.8×10^4 dynes/cm, we find $\omega_b = 2543$ and $\omega_s = 1000$. The child's fundamental frequency is thus 435 Hz.

In the adult male case, or for large L, the ω_s term dominates, or $\omega \approx \omega_s$. The vibration of adult-sized vocal folds is similar to the vibration of a mass coupled to a spring. For an adult whose vocal folds are of length 1.8 cm, height and thickness 0.27 cm, mass 0.141 g, and stiffness 7.4×10^4 dynes/cm, $\omega_b = 103$ while $\omega_s = 724$. The corresponding fundamental frequency is 116 Hz.

The fundamental frequency of the bending beam model depends on the mass, length, cross dimensions and Young's modulus of the vocal-fold tissue and the boundary conditions at the ends of the folds. The scale factor for the bending beam model is

$$SF_{beam} = \frac{b_a}{b_c}\sqrt{\frac{E_c}{E_a}\left(\frac{\beta^4 b_c^4}{12L_c^4}+1\right)\left(\frac{\beta^4 b_a^4}{12L_a^4}+1\right)^{-1}} \quad . \tag{3.11}$$

The scale factor reflects differences in anatomical parameters between children and adults. For adult vocal folds of length 1.8 cm and of thickness 0.27 cm, the term $\left(\frac{\beta^4 b_a^4}{12L_a^4}+1\right)^{-1} = 0.986 \approx 1$. Thus the scale factor reduces to

$$SF_{beam} = \frac{b_a}{b_c}\sqrt{\frac{E_c}{E_a}\left(\frac{\beta^4 b_c^4}{12L_c^4}+1\right)} \quad . \tag{3.12}$$

As a child grows, the term $\left(\frac{\beta^4 b_c^4}{12L_c^4}+1\right) \rightarrow 1$ and SF_{beam} approximates $SF_{spring-mass}$. For the dimensions and tissue properties of a one-year-old child and an adult male (see Table 2.16), $SF_{beam} = 3.7$. By age three years, SF_{beam} is 2.6.

Scale factors for fundamental frequency are listed in Table 3.1 and compared in Fig. 3.2. Each model predicts higher fundamental frequencies for a child's vocal folds than for the vocal folds of an adult. The models differ in the values of the scale factors which are predicted. In the next chapter, these predictions are compared to values of fundamental frequency measured from the speech of young children between the ages of one and three years.

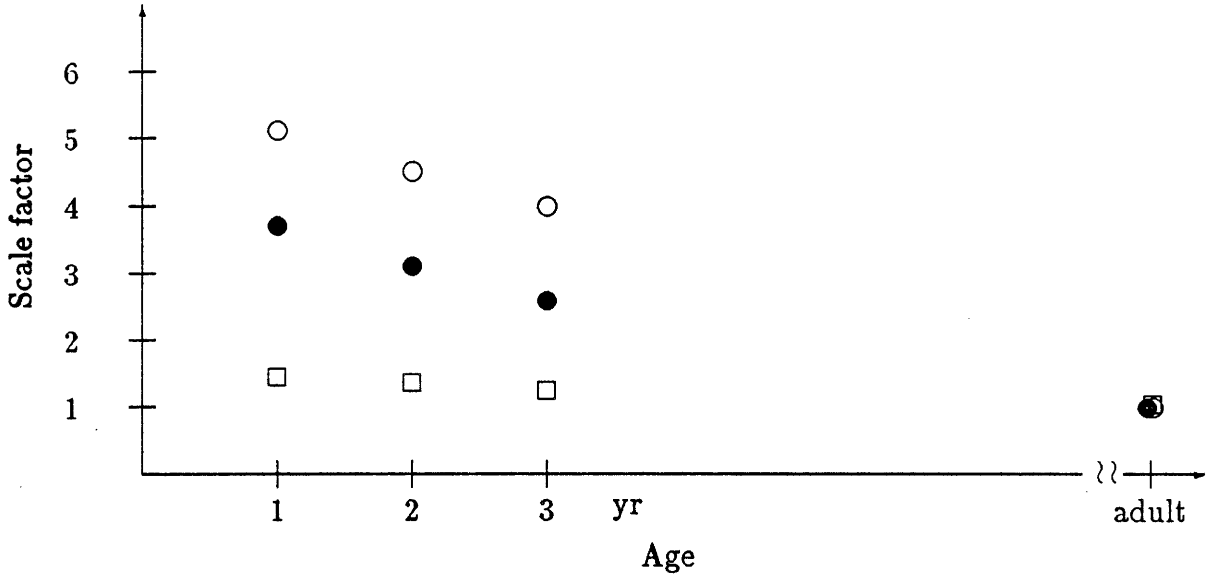

Figure 3.2: Scale factors for $F0$. Values calculated from the vibrating string model are shown by ○'s. Scale factors predicted by the bending beam model are shown by •'s. Values computed from a spring-mass model of the vocal folds are shown by □'s.

3.2 Durations

The duration of a vocalization (an interval of sound bounded by inspiration or by silence) is constrained by the amount of air in the lungs and by the control of airflow through the glottis and vocal tract. The duration of a vowel-like or consonant-like interval within a vocalization is influenced by a speaker's ability to maintain a vocal-tract posture and to move from one articulatory configuration to another. The

Table 3.1: Predicted scale factors for the fundamental frequency of voicing of children with respect to an adult male.

Age	Scale factor		
	Vibrating string	Spring-mass	Bending beam
1 yr	5.1	1.4	3.7
2 yr	4.5	1.3	3.1
3 yr	4.0	1.2	2.6

number of articulatory targets in an utterance is determined by the interplay between the total amount of airflow available for speech and the speed with which a talker changes configurations.

Voiced sounds require less airflow than voiceless sounds (Klatt *et al.*,1968). The waveform of the glottal area versus time, the frequency of vibration of the vocal folds, and the subglottal pressure influence the average airflow which passes from the lungs through the glottis during the production of voiced sounds. The airflow used in producing voiceless consonants depends on the cross-sectional area of the constriction and the subglottal pressure. These factors and the amount of air in the lungs determine the maximum duration of an utterance. A lesser amount of air in children's lungs does not necessarily imply that children's utterances are shorter than adults' utterances.

As shown in Section 2.1.1.1, the volume of air in children's lungs is considerably less than the volume of air in adults' lungs. The percentage of lung capacity which children use during vocalizing is unknown. It is possible (but unlikely) that children use their entire vital capacity (the maximum amount of air which can be exhaled) when producing sounds. A child might use this maximum amount of air during production of a sound such as a cry, but probably uses somewhat less air during the production of vocalizations such as babbling. Direct measurements of vital capacity for young children are not available. Values for the vital capacity of children were estimated from measured tidal volumes of children and the ratio of tidal volume to vital capacity in adults. Estimated vital capacities for one-, two- and three-year-olds of 640, 840 and 940 cm^3, respectively, are listed in Table 2.16. These values represent upper bounds on the volume of air which young children have available for sound production.

In an early attempt at sound production, a child might use only the amount of air used in breathing. Data on tidal volume in newborns, young children, and adults are presented in Section 2.1.1.1 and are listed by age in Table 2.16. At an older age, a child might learn to inhale a greater amount of air in order to produce sounds, as do adults when speaking. Phonation volumes ranging between approximately 20% and 80% of vital capacity have been reported for older children and for adults (see discussion of these data in Section 2.1.1.1). These percentages correspond to a range of 910 – 3640 cm^3 for the phonation volume of an adult male. Direct measurements of phonation volume for young children are not available. Values for young children can be estimated from the ratio of phonation volume to vital capacity for adults in conjunction with the estimated vital capacities of young children. Ranges of volumes which represent from 20% to 80% of the vital capacities of one-, two- and three-year-olds are shown in Table 3.2, along with tidal volumes.

Table 3.2: Tidal volumes and estimated ranges of phonation volumes for young children and adults. (See text for sources.)

Age	Tidal volumes (in cm^3)	Estimated ranges of phonation volumes (in cm^3)
1 yr	95	128 – 512
2 yr	126	168 – 672
3 yr	140	188 – 752
adult	680	910 – 3640

During the production of voiced sounds, the average airflow through the glottis is determined by the peak value of airflow, the shape of the glottal area during a cycle of vibration, and the frequency of vibration of the vocal folds. The peak airflow U_{peak} (in cm^3/s) is given approximately by

$$U_{peak} \approx A_g\sqrt{\frac{2P_{sg}}{\rho}} \quad , \tag{3.13}$$

where A_g is the maximum area of the glottis, P_{sg} is the subglottal pressure, and ρ is the density of air. No direct measurements of glottal areas during a cycle of vocal-fold vibration are available for children. A great deal of variation has been observed in the shape of glottal areas for adults (Holmberg *et al.*, in preparation). Holmberg and her colleagues found that most speakers produced voiced sounds with a constant amount of leakage flow through the glottis. The leakage flow might be due to flow through the cartilaginous portion of the vocal-fold structure or to a laryngeal configuration in which the arytenoid cartilages are held apart. Estimates can be made for glottal areas for children which are similar to those observed for adults, but are scaled in size. Sketches of estimated waveforms of a glottal area for a child and an adult are shown in Fig. 3.3. The waveforms in Fig. 3.3a represent cases of minimum airflow per cycle. The glottis is closed for as long as it is open and there is no leakage of air during the closed interval. A schematized version of an estimated glottal cycle (shown in Fig. 3.3b) represents a maximum-flow case. The glottis does not close during a cycle of vibration and there is some amount of airflow even when the glottis is the least open. A pattern in which the duration of the least-open interval is very short is more common for women than for men. In this example, the area A_{min} during minimum flow is equal to half the maximum area in the no leakage case shown in Fig. 3.3a. The maximum area of the waveform with leakage flow is equal to three times the minimum area. It appears reasonable to assume that a glottal configuration similar to that shown in Fig. 3.3b is appropriate for children because the relative dimensions of the laryngeal structure of a child are more similar to those of an adult female than to an adult male. The amount of leakage flow which is appropriate for children is unknown.

A peak value for the glottal area of a child can be estimated from the vocal-fold length, the maximum displacement of a fold from a rest position and the cross-sectional shape of the glottis. For adults, the displacement δ_a of a vocal fold from a rest position

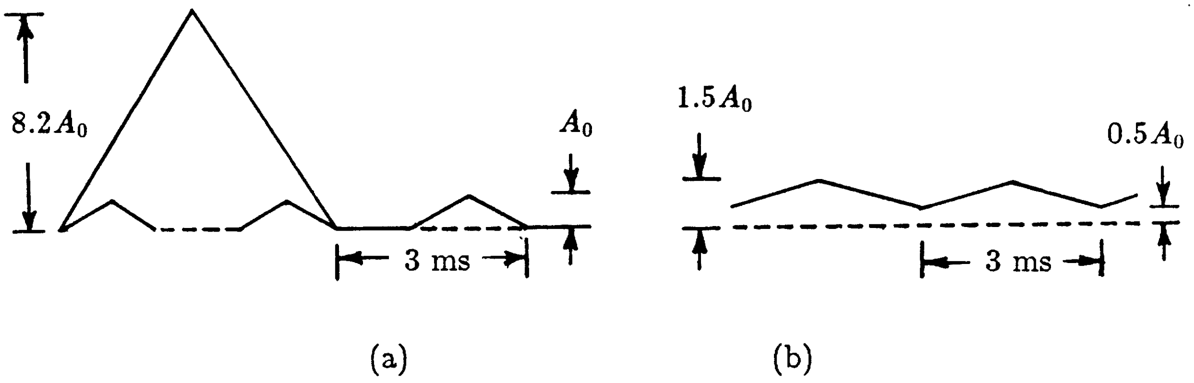

Figure 3.3: Sketches of glottal cycles (area of glottal opening with respect to time) for an adult overlaid on those for a child of age two years. Examples illustrate (a) no leakage and (b) leakage flow. (See text for details.)

is the ratio of the applied force, $P_{sg-a}h_aL_a$, to the stiffness of the vocal-fold tissue, K_a:

$$\delta_a = P_{sg-a}\frac{h_aL_a}{K_a} \quad , \tag{3.14}$$

or 0.051 cm for the values given in Table 2.16. For children, the displacement δ_c is that of a beam of length L_c, thickness b_c, height h_c and Young's modulus E_c which is subjected to a force $P_{sg-c}L_ch_c$. The expression for the displacement reduces to

$$\delta_c = P_{sg-c}\frac{L_c^4}{16E_cb_ch_c^2} \quad , \tag{3.15}$$

and gives 0.020, 0.028 and 0.039 cm for one-, two- and three-year-old children, respectively, for the values listed in Table 2.16. The derivation leading to equation (3.15) is available in standard textbooks (for example, see Cook, 1984). The area of the peak glottal opening A_g is assumed to be approximately $1.5\delta L + A_{min}$, where A_{min} is the minimum area during a vibratory cycle. If there is no leakage through the glottis, $A_{min} = 0$. Values for glottal areas of children and adults calculated from the length and displacement of the vocal folds and the subglottal pressure are shown in Table 3.3 for the leakage-flow and no leakage cases corresponding to the waveforms shown in Fig. 3.3b and a, respectively.

Table 3.3: Estimated glottal areas during voicing of children and adult males, assuming no leakage (middle column) and leakage flow (right column). (See text for details.)

Age	Maximum glottal area (in cm^2)	
	no leakage	with leakage
1 yr	0.011	0.017
2 yr	0.017	0.026
3 yr	0.026	0.039
adult	0.140	0.210

Given values for P_{sg} and the peak area A_g, the airflow U_{peak} can be calculated using equation (3.13). For a glottal cycle with no leakage and closed and open intervals of equal duration (as in Fig. 3.3a), the average flow is one-quarter the peak flow. For the waveform shown in Fig. 3.3b, the average flow is four times greater due to the leakage flow and the longer open interval. Values for the average flow during voicing (assuming glottal cycles as shown in Fig. 3.3a and 3.3b) are shown in Table 3.4. The average adult value assuming no leakage is within the range of 100 – 150 cm^3/sec reported by Klatt *et al.* (1968). Without leakage flow, the average airflow through a child's glottis is much smaller than the airflow through an adult's, due primarily to the differences in vocal-fold length between children and adults. Assuming a leakage flow as in Fig. 3.3b for a child and no leakage for an adult, the average airflow for a two-year-old child (76 cm^3/s) is approximately equal to one-half the average airflow for an adult (138 cm^3/s).

For an utterance which is produced entirely with a voiced source at the glottis, the duration of the utterance is equal to the volume of air in the lungs divided by the average glottal airflow. Table 3.5 lists ranges of the durations of voiced utterances of children and adults based on both tidal volumes and phonation volume. The lower limit of each phonation volume range in Table 3.2 and the average glottal airflows shown in

Table 3.4: Estimated glottal airflows during production of voiced sounds by children and adult males, assuming no leakage (middle column) and leakage flow (right column). (See text for details.)

Age	Average glottal airflow (in cm^3/s)	
	no leakage	with leakage
1 yr	12	48
2 yr	19	76
3 yr	30	120
adult	138	552

Table 3.4 were used in the calculations. For adults, the upper limit (corresponding to no leakage) of the range based on phonation volume is probably the most applicable. For children, the lower limit (corresponding to glottal leakage as shown in Fig. 3.3b) of the tidal volume range is likely to be most appropriate. Predicted durations of voiced utterances based on these assumptions are shown in Table 3.6. The durations for children are based on tidal volume and assume a leakage flow; the adult values are calculated from phonation volumes and assume no leakage flow.

The number of segments in an utterance can be predicted from the volume of air available for speech and the amount required for individual consonant-like and vowel-like sounds. In adult speech, a sequence of syllables consisting of voiceless consonants followed by vowels uses approximately 100 cm^3 per consonant and 30 cm^3 per vowel (Klatt *et al.*, 1968). For a glottal flow of 138 cm^3/s, the duration of a vowel would be approximately 200 ms, within the range of the lengths of vowels produced in read or conversational speech. House (1961) reports an average vowel duration of 100 – 400 ms. A similar range is reported by Crystal and House (1982); they calculated a mean duration of 102 ms for vowels in read sentences. For these volumes for consonants and vowels and for phonation volumes of 910 – 3640 cm^3, between seven and thirty syllables per breath are predicted for an adult male. In a sample of read sentences (MIT, 1985),

Table 3.5: Ranges of durations of voiced utterances for children and adults based on phonation volume and on tidal volume. The lower limit of each range corresponds to a glottal waveform with leakage flow. The upper limit is based on no leakage through the glottis.

Age	Ranges of voicing duration (in sec)	
	(phonation volume)	(tidal volume)
1 yr	3 – 11	2 – 8
2 yr	2 – 9	1 – 7
3 yr	2 – 6	1 – 5
adult	2 – 7	1 – 5

sentence durations ranged from 1.7 sec (seven syllables) to 3.1 sec (sixteen syllables). In these sentences, average syllable durations ranged between 200 and 250 ms.

Alternations between open and closed vocal-tract configurations are required during a diadochokinetic task, which consists of repetition of syllables such as [pə], [tə] and [kə]. Children as young as five years of age have been tested on diadochokinetic tasks. Rates for children and adults are discussed in Section 2.2.3. Table 2.16 lists values of 2.7, 3.0 and 3.2 Hz for one-, two- and three-year-old children, respectively, and 6.0 Hz for adults. The diadochokinetic rate for adults is somewhat faster than the syllable rate of 4 – 5 Hz for the MIT read sentences. These data predict that children younger than three years of age are able to produce syllables only about half as fast as are adults. That is, children's syllables are predicted to be on average about twice as long as are syllables produced by adults, or approximately 400 – 500 ms in duration.

The volume of air needed to produce a vowel can be determined by estimates of average airflow through the glottis and segment duration. For a vowel duration of 400 msec and average flow during voicing of 76 cm^3/sec, a two-year-old child would use 30 cm^3 of air per vowel, about the same amount used by adults.

The airflow used by a child in the production of a voiceless consonant can be

Table 3.6: Predicted durations of voiced utterances for children (based on tidal volumes and glottal leakage), and for adults (based on phonation volumes and no leakage).

Age	Predicted durations of voiced utterances (in sec)
1 yr	2.0
2 yr	1.7
3 yr	1.2
adult	6.6

predicted from the subglottal pressure and area of the constrictions formed by the child. In the simple case of only a vocal-tract constriction (wide-open glottis) or only a glottal constriction (no vocal-tract constriction), the airflow U can be calculated from

$$U \approx A_{con}\sqrt{\frac{2P_{sg}}{\rho}} \quad , \tag{3.16}$$

where A_{con} is the area of the constriction.

In the case of a fricative, a constriction is formed in the vocal tract. Children might not be able to maintain a constriction in the vocal tract which is as narrow in comparison to the maximum area as is the constriction formed by an adult. This hypothesis is consistent with the data describing consistency of finger tapping (Tingley and Allen, 1975). A child might only be able to form a constriction of about the same area as an adult's constriction. For an adult, the average volume increment for a voiceless fricative is about 90 cm^2. For a fricative of duration 200 ms, the airflow would be 450 cm^2 and the corresponding A_{con}, 0.12 cm^2. The airflow used by a child differs from the airflow used by an adult due to the difference in subglottal pressure (if the same A_{con} is used). Using a constriction of area 0.12 cm^2, a child would produce an airflow of about 550 cm^3/s. In order to produce a fricative about twice as long as an adult's, a two-year-old child would use as much as 220 cm^3, or somewhat more than the lower limit on phonation volume. A child would need to use a greater volume of air from the lungs in order to produce consonant-vowel sequences in which the consonants are voiceless fricatives.

During the production of some voiceless sounds, the area of the glottis determines the airflow. The glottis of a child is considerably smaller than the glottis of an adult. For an aspirated consonant (/h/) produced by an adult, the average volume increment is 130 cm^3 (Klatt *et al.*, 1968). A duration of 200 ms implies a flow of 650 cm^3/sec. Equation 3.16 gives a corresponding glottal opening of approximately 0.18 cm^2, or slightly greater than the peak glottal opening during voicing. It is possible that children would also use a larger glottal opening in order to produce aspiration – an opening which was comparably larger than the maximum glottal opening used during voicing. For a two-year-old child, a peak glottal opening during voicing of 0.026 cm^2 is listed in Table 3.3. Assuming the same ratio of aspirated opening to peak-voicing opening for children as for adults, a glottal opening of 0.033 cm^2 is predicted during aspiration for the child. The airflow would then be 150 cm^3/sec. To produce an /h/ of duration 400 ms, the child would use 60 cm^3 of air. According these predictions, two-year-old children use about one-half the amount of air during aspiration as adults use.

For an average volume increment of 140 cm^3 per voiceless consonant (averaged over fricatives and /h/) and 30 cm^3 per vowel, two-year-olds could produce only about one syllable which consists of a voiceless consonant and a vowel by using tidal volume. By using the maximum estimate of phonation volume, a child could produce about four syllables with voiceless consonants on one breath. In contrast, a child could produce about twice as many syllables consisting of voiced consonants and vowels.

The estimates of duration of voicing imply that children's voiced utterances can be longer than adults' in spite of significantly smaller lung volumes, due to differences in airflow through the glottis. On the other hand, children are predicted to produce fewer consonant-vowel syllables in an utterance because the durations of individual segments are probably longer. The number of syllables consisting of voiceless consonants is predicted to be even fewer since the airflow for voiceless consonants is relatively much greater than the airflow during voicing. In the next chapter, these values of duration are compared to measurements of vocalizations of young children.

3.3 Formants

Spectra of children's vocalizations can be described in terms of formant frequencies and bandwidths. A strength of this approach is its explanatory power – a spectrum can be predicted from a description of a vocal-tract configuration and the impedances of the walls, glottis and mouth opening, and, inversely, vocal-tract configurations can be inferred from spectra. The range of formant frequencies dictated by the dimensions and tissue properties of a child-sized vocal tract is explored in Section 3.3.1. The differences in formant bandwidths between children and adults are analyzed in Section 3.3.2.

3.3.1 Formant frequencies

Formant frequencies depend on the vocal-tract shape and the characteristics of the vocal-tract walls and terminations. Calculation of the resonant frequencies of a child-sized vocal tract of an idealized shape is straightforward. The reactances due to radiation, flexible walls, and glottal opening cause shifts in the peak frequencies from those for the hard-wall, closed-glottis configuration with negligible radiation impedance (unperturbed case). Formulas for calculating the contributions due to these reactive components are available in the literature (Fant, 1960/70; Stevens, in preparation).

3.3.1.1 Uniform vocal tract

The vocal tract is unconstricted during the production of a neutral vowel-likesound. A sketch of an idealized vocal tract with a uniform cross-sectional area is shown in Fig. 3.4. For a uniform vocal tract of length 10 cm, corresponding to that of a two-year-old child, the lowest three formant frequencies ($F1$, $F2$ and $F3$) are 885, 2655 and 4425 Hz, respectively. These values were calculated assuming a closed glottis, hard vocal-tract walls and no radiation. The effects of a glottal opening, flexible walls and radiation from the mouth perturb these frequencies and are analyzed independently below.

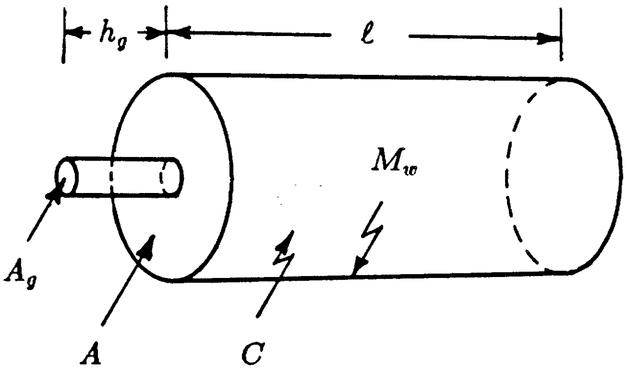

Figure 3.4: Uniform vocal tract.

Radiation from the mouth causes a shift in the resonant frequencies to lower frequencies than those calculated for the negligible-radiation case. The resonances ω which include the effects of radiation are solutions of

$$-\frac{\rho c}{A}\cot\frac{\omega\ell}{c}+\omega\frac{\rho\,0.8a}{A}=0 \quad , \tag{3.17}$$

where ρ is the density of air, c is the speed of sound in air, ℓ and a are the length and radius of the vocal tract, and A is the corresponding cross-sectional area of the vocal tract. The lowest three values of ω correspond to formant frequencies (in Hz) of approximately $F1=\frac{c}{4\ell}(1-\frac{0.8a}{\ell+0.8a})$, $F2=\frac{3c}{4\ell}(1-\frac{0.8a}{\ell+0.8a})$, and $F3=\frac{5c}{4\ell}(1-\frac{0.8a}{\ell+0.8a})$. For a two-year-old's vocal tract of length 10 cm and radius 0.82 cm, the radiation reactance causes the first formant to be shifted down by 54 Hz; the second formant, by 163 Hz; and the third formant, by 272 Hz. The formant frequencies including the effects of radiation are therefore 831, 2492 and 4153 Hz. These shifts represent approximately 6% of each formant frequency. For an adult-sized vocal tract of 16.9 cm in length and 1.1 cm in radius, the frequency shift due to radiation is approximately 5%. The shifts in formant frequency due to radiation change little with age during the years one to three, due to the relatively constant relationship between length and radius of the vocal tract during early childhood.

The reactance of the glottal opening causes the resonant frequencies to be higher than for a vocal tract with a closed glottis. The frequencies for a uniform vocal tract with an open glottis are given by

$$\frac{A}{\rho c}\cot\frac{\omega\ell}{c}+\frac{\omega M_g}{\omega^2 M_g^2+R_g^2}=0 \quad , \tag{3.18}$$

where $M_g = \frac{\rho h_g}{A_g}$ g/cm^4 and $R_g = \frac{\rho U_g}{A_g^2}$ dyne-s/cm^5. In these expressions, the variable h_g represents the height of the glottal opening; A_g, the average glottal area; and U_g, the average airflow through the glottis. For an average glottal flow of 76 cm^3/s, average glottal area of 0.017 cm^2 and the glottal dimensions of a two-year-old child, $R_g^2 \approx 9\times10^4$. For $\omega = 2\pi(885)$, $\omega^2 M_g^2 \approx 8\times10^3$. For these values, $R_g^2 \gg \omega^2 M_g^2$ and equation (3.18) becomes

$$\frac{A}{\rho c}\cot\frac{\omega\ell}{c}+\frac{\omega M_g}{R_g^2}=0 \quad . \tag{3.19}$$

The solution to equation (3.19) for low frequencies is $F1 = \frac{c}{4\ell}(1+\frac{\rho c^2 M_g}{A\ell R_g^2})$. Equation (3.18) has solutions for ω which correspond to a second formant of $F2 = \frac{3c}{4\ell}(1+\frac{cA_g^2}{3\pi AU_g})$ and a third formant of $F3 = \frac{5c}{4\ell}(1+\frac{cA_g^2}{5\pi AU_g})$. For a two-year-old child's vocal tract, the shifts in the formant frequencies due to glottal opening are negligible (at most 11 Hz, or 1.2% of the first formant frequency of the unperturbed vocal tract). For vowel configurations of an adult, the shifts are also negligible.

Flexible vocal-tract walls also act to increase the formant frequency. The effect of the walls is distributed along the length of the vocal tract and is due primarily to the mass of the walls, which acts in parallel with the compliance of the air in the vocal tract. Thus, the resonant frequencies of a uniform vocal tract with flexible walls are solutions of

$$\omega^2-\left(\frac{1}{CM_w}+\frac{\pi^2c^2}{4\ell^2}\right)=0 \quad , \tag{3.20}$$

where C is the acoustic compliance per unit length of the air in the vocal tract and M_w represents the acoustic mass per unit length of the vocal-tract tissue. The low frequency solution to equation (3.20) gives $F1 = \frac{c}{4\ell}\left(1+\frac{2\ell^2}{\pi^2c^2CM_w}\right)$. The frequency of the second formant is given by $F2 = \frac{3c}{4\ell}\left(1+\frac{2\ell^2}{9\pi^2c^2CM_w}\right)$ and the third, by $F3 = \frac{5c}{4\ell}\left(1+\frac{2\ell^2}{25\pi^2c^2CM_w}\right)$. For the vocal tracts of both children and adults, the effect of flexible walls on the formant frequencies of a uniform vocal tract is negligible. For a labial stop consonant (closed vocal tract), the effect of the walls would be more significant.

Formant frequencies of a vocal tract with uniform cross-sectional area, such as that shown in Fig. 3.4, are shown in Fig. 3.5 as a function of vocal-tract length. The lengths correspond to the vocal tracts of one-, two and three-year-old children (see Section 2.1.1.4). Frequencies for a vocal tract with hard walls, closed glottis and no radiation are shown by o's. It can be seen that the lowest formant is at approximately 850 to 950 Hz, and that formant frequency spacing is approximately 1770 Hz for vocal-tract lengths of young children. Formant values corresponding to an adult-sized vocal tract are shown at the right for reference. Including the effects of radiation results in lower frequencies; these frequencies are shown by •'s. A glottal opening causes the formant frequencies to be slightly higher compared to the unperturbed case. The reactive effect for both child- and adult-sized glottal openings is negligible.

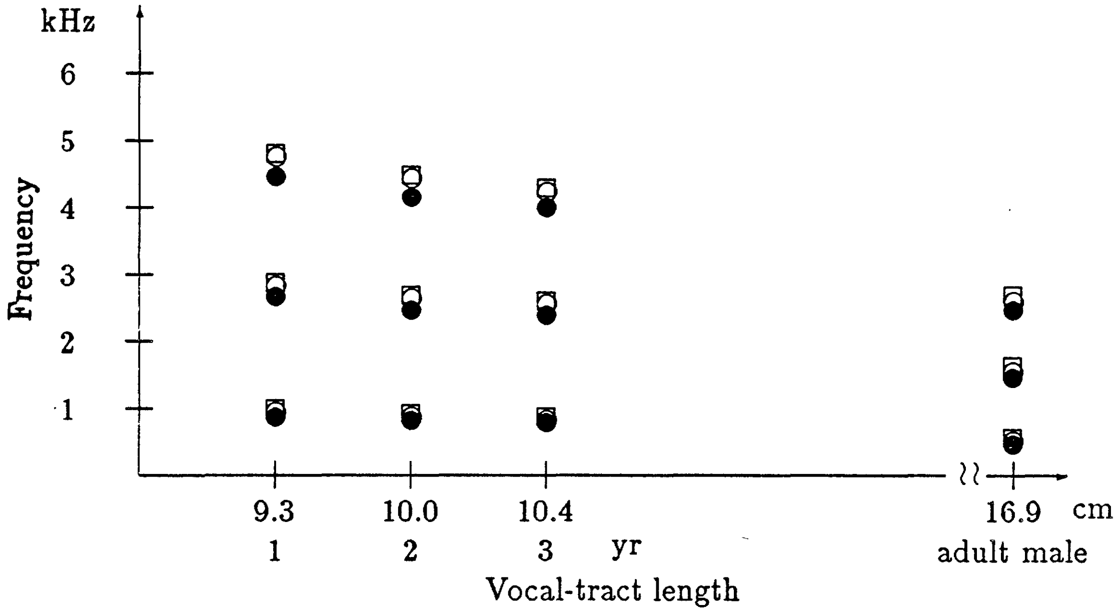

Figure 3.5: Formant frequencies as a function of vocal-tract length. The values for the uniform tube are shown by o's. The frequencies due to the effects of radiation are shown by •'s. Those due to glottal reactances are shown by □'s.

3.3.1.2 Moderately constricted vocal tract

During the production of most vowel-like sounds, the vocal tract is moderately constricted in one or two places along its length. Sketches of several idealized configurations are shown in Fig. 3.6. These configurations correspond to vocal tracts for a low back vowel such as /ɑ/ (Fig. 3.6a), a high front vowel /i/ (Fig. 3.6b), and a high back vowel such as /u/ (Fig. 3.6d). Figure 3.6c corresponds to an unrounded high back vowel.

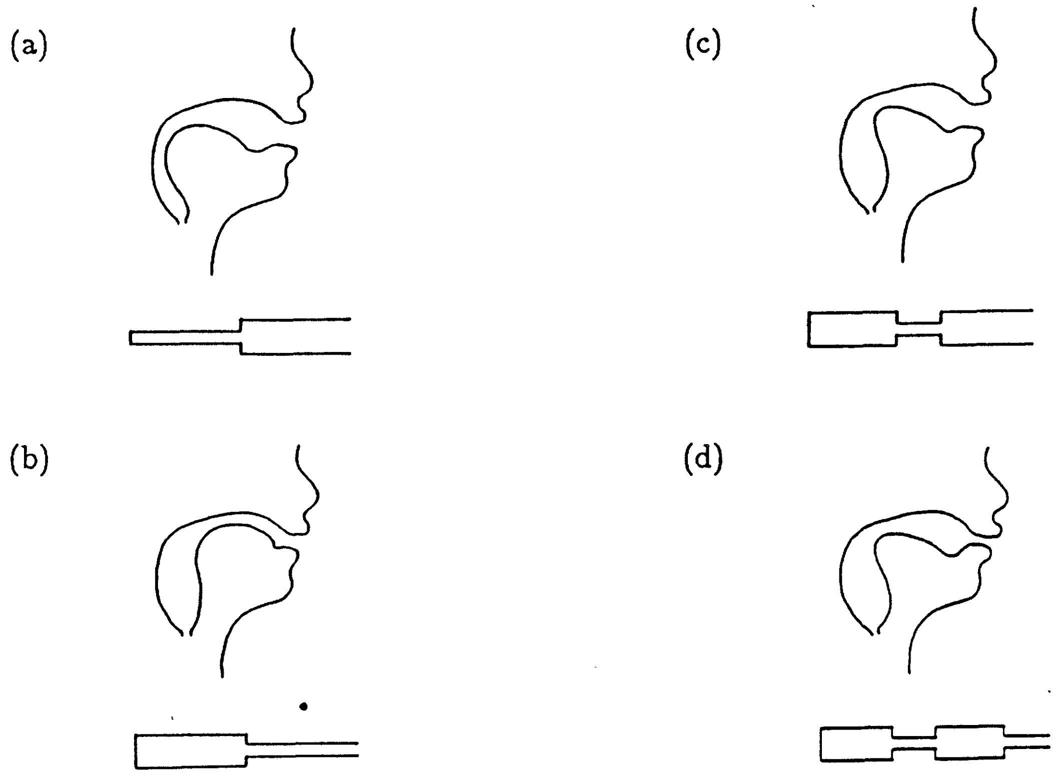

Figure 3.6: Idealized vocal-tract configurations with various constrictions: posterior (as in the vowel /ɑ/), anterior (/i/), middle (ɯ), and middle and lips (/u/).

The formants of vowels produced with a vocal-tract configuration similar to the idealized case shown in Figs. 3.6a or 3.6b can be found by treating the vocal tract as coupled sections of uniform, but different, cross-sectional area. Acoustic coupling causes the resonances of the configuration to differ from those calculated for the individual sections separately. The minimum value of formant separation ΔF for cavities of equal length is found by solution of

$$-\frac{\rho c}{A_b}\cot\frac{\omega\ell}{c}+\frac{\rho c}{A_f}\tan\frac{\omega\ell}{c}=0 \quad , \tag{3.21}$$

where A_b is the cross-sectional area of the back cavity and A_f is the area of the front cavity. Both cavities are assumed to be of the same length ℓ. Solution of equation (3.21) gives $\Delta F \approx \frac{c}{\ell\pi}\sqrt{\frac{A_f}{A_b}}$ as the minimum separation (in Hz) between formants for configurations in which $A_b \gg A_f$. For a configuration such as that shown in Fig. 3.6a ($A_f \gg A_b$), $\Delta F \approx \frac{c}{\ell\pi}\sqrt{\frac{A_b}{A_f}}$.

/ɑ/-like configuration The formant frequencies of the configuration shown in Fig. 3.6a are given approximately by the solution of

$$-\frac{A_f}{A_b}\cot\frac{\omega\ell_b}{c} + \tan\frac{\omega(\ell_f + 0.8a_f)}{c} = 0 \quad , \tag{3.22}$$

where ℓ_f is the length of the front cavity, A_f and a_f are the area and radius, respectively, and the back cavity length and area are ℓ_b and A_b. Equation 3.22 includes the effects of radiation and of coupling between the sections. The shifts in frequency due to the reactances of the glottal opening and flexible walls are negligible. Values of the lowest three formants for a front cavity of cross-sectional area 2.1 cm^2 and back cavity area of 0.3 cm^2 are graphed as a function of front-cavity length in Fig. 3.7. The values were calculated for the vocal tract of a two-year-old child. The • represents frequencies affiliated primarily with the front cavity; the ∘'s show frequencies associated with the back cavity. Due to the effects of coupling, $F1$ and $F2$ are separated by at least 800 Hz in all cases, as predicted by the expression for ΔF shown above. A stable region in which the values for $F1$ and $F2$ do not vary appreciably regardless of cavity length is apparent in configurations in which the front-cavity length is approximately equal to the back-cavity length. Due to the coupling of sections of a child's vocal tract, this stable region exists for over a variation in cavity length of approximately 2 cm. For an adult vocal tract, the effect of coupling is probably less, due to the assumed ability of an adult to maintain a narrower constriction in relation to the maximum cross-sectional area and therefore a smaller minimum spacing between formants. The most stable region correspondingly extends over a smaller range of cavity lengths.

/i/-like configuration Figure 3.6b shows a vocal tract with a large back cavity and narrow front cavity. The formants of this configuration are given by equation 3.22,

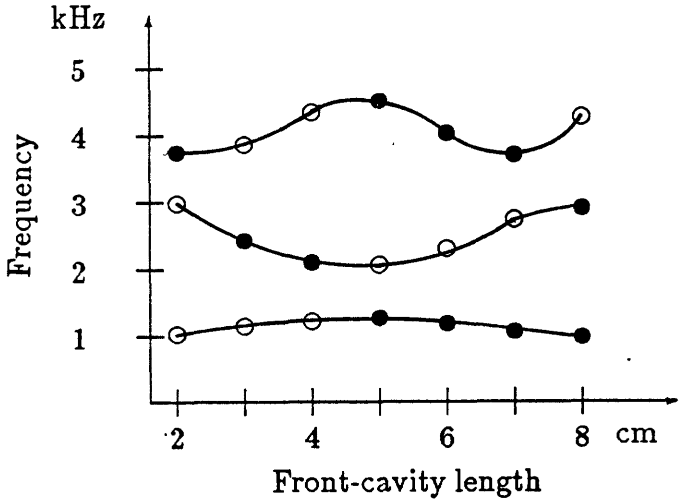

Figure 3.7: Resonances of an /ɑ/-like configuration. Front and back cavity resonances are identified by •'s and o's, respectively.

where A_b is 2.1 cm^2 and A_f is 0.3 cm^2. For the first formant, however, the effects of the walls are not negligible and a correction must be made, as indicated below. The effect of the walls can be included in the calculation of the first formant of the configuration shown in fig. 3.6b by estimating the first formant from the low-frequency circuit model shown in Fig. 3.8. The element C_v models the acoustic compliance of the air in the back cavity; M_w is the acoustic mass of the vocal-tract walls; R_g dominates the impedance of the glottal slit; and M_c is the mass of the air in the narrow front cavity. The first formant is given by $F1 = \frac{1}{2\pi\sqrt{(M_c||M_w)C_v}}$. For a young child, the first formant of this /i/-like configuration is greater than the $F1$ for an adult, due primarily to the differences in the volume of the back cavity. For a front-cavity length equal to the back-cavity length, $F1$ = 364 Hz for a two-year-old child, which is close to the average fundamental frequency for two-year-old children. Without including the effect of the walls, the lowest value of $F1$ would be approximately 397 Hz. For a similar configuration for an adult, the frequency is 249 Hz. In the case of a very narrow front cavity, $M_w \gg M_c$, and $F1 \approx \frac{1}{2\pi\sqrt{M_w C_v}}$, resulting in a lower limit for the first formant of approximately 213 Hz for vocal tract of a two-year-old child. For an adult, the lower limit is approximately 171 Hz for the dimensions given in Table 2.16. The $F1$ for a child can be lower than the fundamental frequency, while the $F1$ for an adult male usually falls well above the fundamental frequency.

Values for the lowest three formants of an /i/-like configuration are plotted with respect to the length of the front cavity in Fig. 3.9. Dimensions and tissue properties are typical of a two-year-old child. Cavity affiliations are indicated as in Fig. 3.7. The effects of coupling cause the separation between $F2$ and $F3$ to be greater than 800 Hz.

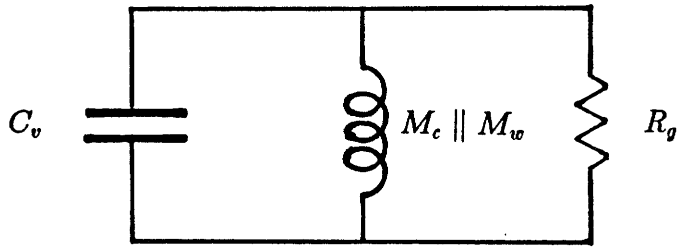

Figure 3.8: Circuit model of a back cavity, including the masses of the air in the constriction and of the walls, the resistance of the glottal opening, and the compliance of the air in the cavity.

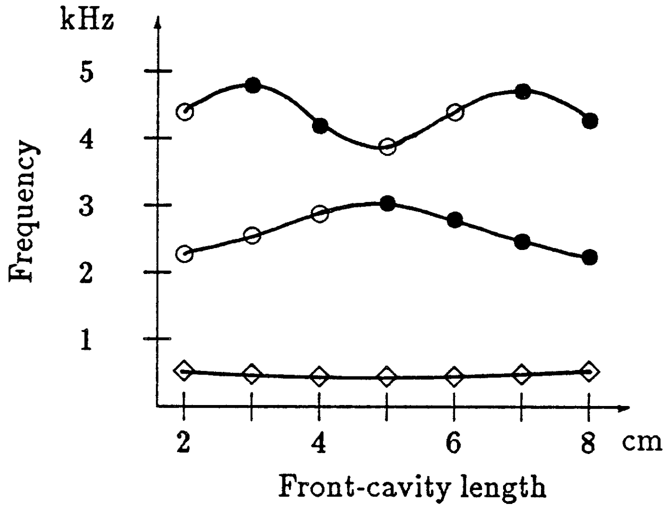

Figure 3.9: Resonances of an /i/-like configuration. Front and back cavity resonances are identified by •'s and o's, respectively. Helmholtz resonances are shown by ◇'s.

/ɯ/-like configuration The formant frequencies of a vocal tract with front and back cavities of similar cross-sectional area separated by a constriction are given by

$$-\frac{c}{A_b}\cot\frac{\omega \ell_b}{c}+\frac{\omega \ell_c}{A_c}+\frac{c}{A_f}\tan\frac{\omega(\ell_f+0.8a_f)}{c}=0 \quad . \tag{3.23}$$

This configuration models an unrounded configuration. Figure 3.10 shows formant frequencies of the configuration shown in Fig. 3.6c with cavities of area 2.1 cm^2 separated by a constriction of length 2 cm and area 0.3 cm^2. The length of the back cavity is $8-\ell_f$ cm, where ℓ_f is the length of the front cavity.

For a very narrow constriction, the lower limit on $F1$ is the same as that found for the case of an /i/-like configuration. The only dimension which affects this lower limit on $F1$ is the radius of the back cavity.

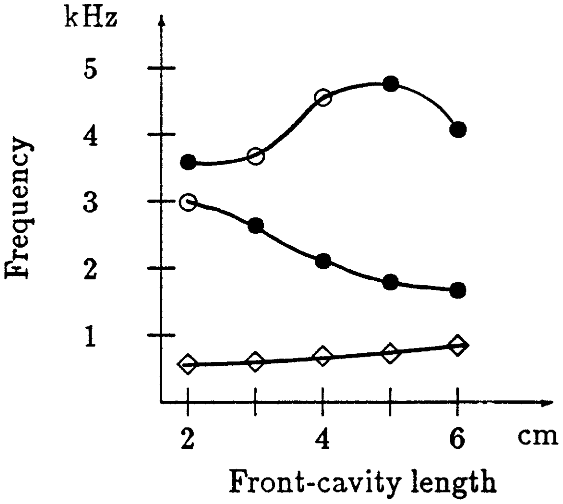

Figure 3.10: Resonances of an /ɯ/-like configuration. Front and back cavity resonances are identified by •'s and ∘'s, respectively. Helmholtz resonances are shown by ◇'s.

/u/-like configuration For the /u/-like configuration shown in Fig. 3.6d, the lowest two natural frequencies can be estimated from the circuit shown in Fig. 3.11. The elements C_b and C_f represent the compliance of the air in the back and front cavities, respectively, and $M_b \parallel M_{wb}$ and $M_f \parallel M_{wf}$ are the parallel combinations of the acoustic masses of the walls and the air in the constrictions. Higher resonances are

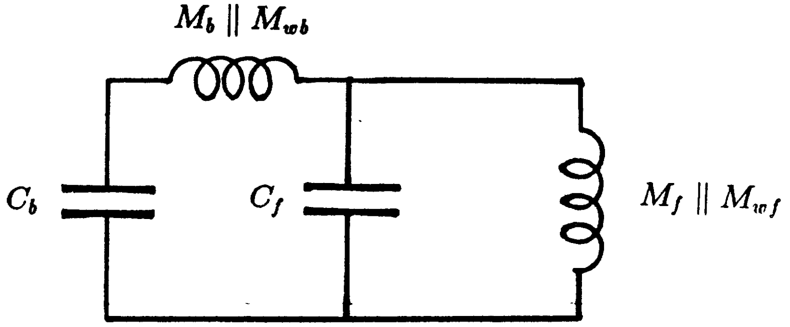

Figure 3.11: Circuit model of the lowest resonances of a coupled configuration of front and back cavities and constrictions in the middle and at the lips. The compliances of the air in the cavities are represented by C_b and C_f; the masses of the air in the constrictions, by M_b and M_f; and the masses of the walls, by M_{wb} and M_{wf}.

approximately equal to integral multiples of $\frac{c}{2\ell_b}$ and $\frac{c}{2\ell_f}$, where ℓ_b and ℓ_f are the lengths of the back and front cavities, respectively. The lowest three formant frequencies for a configuration with constrictions of length 2 cm are shown in Fig. 3.12. The cavities of the configuration have areas of 2.1 cm^2; the constrictions are of length 2 cm and area 0.3 cm^2. The length of the back cavity is $6 - \ell_f$ cm.

3.3.2 Formant bandwidths

The major contributions to the bandwidths of formants are due to radiation, wall resistance, viscosity, heat conduction, and glottal resistance. Formulas for calculating each of these bandwidth contributions are shown in the left column of Table 3.7. These formulas are those given in Stevens (in preparation) and are equivalent to the ones in Liljencrants (1985) and, except for glottal losses, Fant (1972). In the right column are ratios of frequency and dimensions; each ratio is proportional to the corresponding bandwidth.

Bandwidth contributions due to radiation (B_r), wall resistance (B_w), viscosity (B_v), heat conduction (B_h), and glottal resistance (B_g) are shown as a function of frequency

Table 3.7: Theoretical bandwidth contributions for a uniform tube of cross-sectional area A and length ℓ are shown in the left-hand column (from Stevens, in preparation). At the right are proportions of each bandwidth to frequency (F) and dimensions.

$B_r = \frac{K_s(\omega) f^2 A}{lc}$	$\propto F^2$ *radius*^{2} *length*$^{-1}$
$B_w = \left(\frac{R_w}{X_w^2}\right) \frac{\rho c^2}{\pi a}$	$\propto F^{-2}$ *radius*$^{-1}$
$B_v = \frac{1}{a}\sqrt{\frac{f\mu}{\pi\rho}}$	$\propto \sqrt{F}$ *radius*$^{-1}$
$B_h = \frac{0.4}{\pi a}\sqrt{\frac{\pi\lambda f}{c_p \rho}}$	$\propto \sqrt{F}$ *radius*$^{-1}$
$B_g = \frac{\rho c^2}{\pi \ell A R_g (1+\omega^2 M_g^2/R_g^2)}$	$\propto$ *area*$_{glottis}$ *length*$^{-1}$ *radius*$^{-2}$

for $R_w = 1000$ g/s-cm^2, $X_w = 2\omega$ g/s-cm^2,

$\mu = 1.86 \times 10^{-4}$ dynes-s/cm^2, $\lambda = 5.5 \times 10^{-5}$ cal/cm-s-deg,

$c_p = 0.24$ cal/g-deg,

$R_g = \frac{\rho U}{A_g^2}$ g/s-cm^4, and $M_g = \frac{\rho h_g}{A_g}$ g/cm^4.

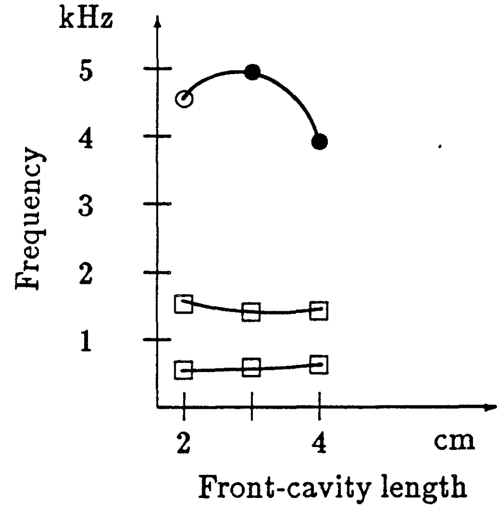

Figure 3.12: Resonances of an /u/-like configuration. Front and back cavity resonances are identified by • and ∘, respectively. Resonances due to the effects of coupling, based on Fig. 3.11, are shown by □'s.

in Figs. 3.13 and 3.14. Figure 3.13 shows radiation bandwidths for three different mouth openings. The largest opening, $A_m = 2.0$ cm^2, corresponds to an open mouth of a two-year-old child. Figure 3.14 shows the bandwidth contributions due to heat, viscosity, glottal opening and flexible walls.

The ratio of vocal-tract cross dimensions of a child relative to an adult is not the same as the ratio of lengths. The scale factor for the radius of a cross section of the vocal tract of a two-year-old child to an adult's is approximately 0.73; the scale factor for vocal-tract lengths is 0.59. The glottal dimensions scale differently from either of these vocal-tract dimensions. For the dimensions of a two-year-old child, the bandwidth contributions due to the effects of radiation are greater in comparison to adults. Part of this increase is due to higher formant frequencies for the child (B_r is proportional to $frequency^2$) and part is due to differences in vocal-tract dimensions (an additional factor of $radius^2$). For a uniform vocal tract of a two-year-old child, the bandwidths of the first three formants are $B1_r = 8$, $B2_r = 71$ and $B3_r = 197$ Hz. Typical adult values are $B1_r = 3$, $B2_r = 27$ and $B3_r = 74$ Hz.

Losses due to glottal opening are greater for children than for adults if the glottal configurations discussed in Section 3.2 are assumed. The bandwidths due to glottal

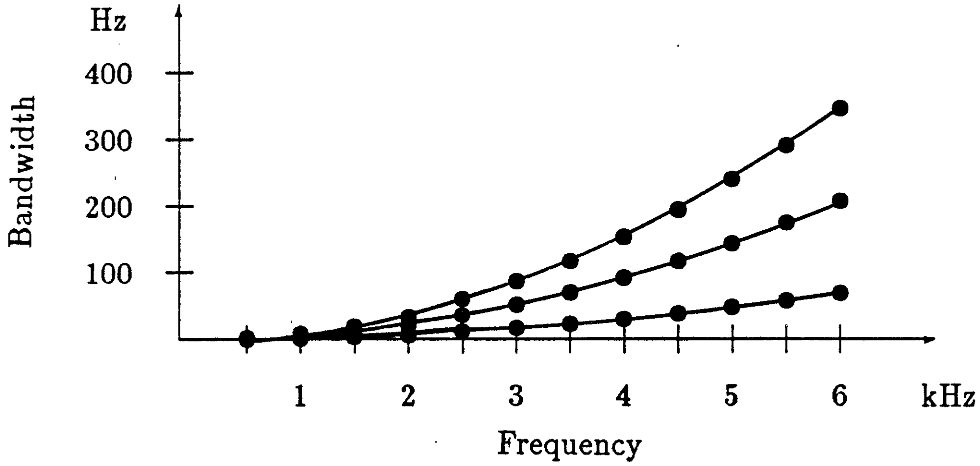

Figure 3.13: Bandwidth contributions of radiation effects for three mouth areas, $A_m = 2.0$, 1.2, and 0.4 cm^2, assuming a uniform tube of length 10 cm. The largest bandwidth corresponds to the largest mouth opening.

losses (assuming an average glottal opening of 0.017 cm^2 for a child and 0.035 cm^2 for an adult) are $B1_g = 66$ Hz for two-year-old children and 53 Hz for an adult. The effects of flexible walls, viscosity and heat conduction on formant bandwidth are similar for children and adults. Table 3.8 lists estimated bandwidths of child- and adult-sized uniform tubes due to the combined effects of radiation, flexible walls, viscosity, heat conduction and glottal opening.

Table 3.8: Predicted bandwidths for children and for adults. (See text for details.)

Age	Predicted bandwidths of a uniform tube		
	B1	B2	B3
2 yr	90	132	246
adult	72	78	115

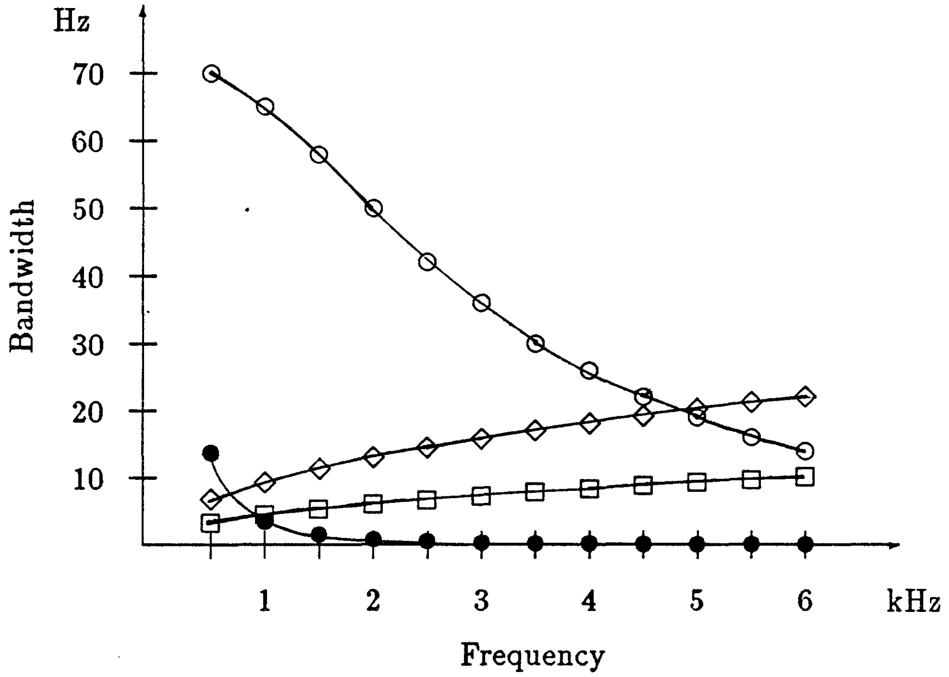

Figure 3.14: Bandwidth contributions for a uniform tube of length 10 cm due to wall resistance (●), viscosity (◇), heat conduction (□), and glottal resistance (○).

3.4 Summary of predictions

In this chapter, a new model for vocal-fold vibration has been presented and existing models of duration and of formant frequency and bandwidth have been examined using parameter values appropriate for child-sized vocal systems. The applicability of the bending beam model can be judged by comparing the predicted fundamental frequencies of the model with values of $F0$ measured from the speech of young children. The predicted ranges of durations of utterances, syllables and segments rely on assumptions of the volume of air used in speech and the airflow through the glottis and vocal tract. The different rates of growth of the lungs and the larynx and the development of articulatory coordination are reflected in the predicted changes in duration throughout early childhood. The existing model of formant frequencies and bandwidths, which has proven useful in analyses of adult speech, predicts values of frequency and bandwidth which are not related to adult values by one simple scale factor. Here again, differences are the result of non-uniform scaling of vocal-tract dimensions between children and adults. The predictions in each of these domains of fundamental frequency, duration, and formant frequency and bandwidth are compared to values measured from the speech of young children in Chapter 4.

Chapter 4

Acoustic characteristics of children's speech

Measurements of recordings provided by the Language Indices Development (LIDS) researchers and by the Jollerutveckling research team are analyzed in this chapter in reference to the model predictions of Chapter 3. Children raised in (American) English-speaking homes were recorded by the LIDS research team of Children's Hospital in Boston, Massachusetts. A description of the children and of the recording procedure has been presented by LIDS researchers (Chesnick *et al.*, 1983; Murphy *et al.*, 1983). Recordings of Swedish children were prepared by the Jollerutveckling research team of St. Goran's Children's Hospital in Stockholm, Sweden. The procedures for recording and transcribing the Swedish children's utterances are described by Holmgren *et al.* (1985).

The recording situations and procedures used by the two research teams are similar. Recordings of the American children were made by speech clinicians in home settings at two- to three-month intervals from age one month to three years. The children's vocalizations were transcribed by a speech clinician. Word attempts were glossed during each recording session and later were confirmed by another clinician listening to the session tapes. The word gloss was based on the therapists' perception, the context of the child's utterance, and the mother's reaction. Several physiological and speech/language scores were used by the LIDS team to describe the children. The descriptors include premature/normal term, sex, physiological risk including otitis media, Developmental Sentence score, Peabody Picture Vocabulary Test score, morphology score at 29 months, mean length of utterance, phonology score at 29 months, and lexical acquisition score. The recordings of forty of the children who tested normal in speech and language skills at age three years and for whom transcriptions were prepared are analyzed in this chapter.

Recordings of the Swedish children were made in their homes at two-week intervals between the ages of six weeks and eighteen months. Transcriptions were prepared by graduate students of linguistics for most of the recording sessions (Holmgren *et al.*, 1985). Early words were glossed from notebooks written by the recording team during each visit. All of these children appear to have developed normal speech. Recordings of four of the Swedish children are also analyzed in this chapter.

4.1 Measurement techniques

The techniques used for measuring acoustic characteristics of the LIDS and Joller-utveckling recordings are described in this section. The utterances were digitized at 16 kHz. Spectrograms and computer-generated spectral and temporal displays of the speech signal were used to measure the fundamental frequency, durations (utterance, syllable, segment), and formants (frequency and bandwidth) of the children's speech. Computer displays were produced by an interactive speech analysis program, SPIRE (Speech and Phonetics Interactive Research Environment) (Zue *et al.*, 1986) for the analyses of the utterances of all of the children. In addition, spectrograms were prepared using a Voiceprint spectrograph of the vowels produced by the American children. Durations of the Swedish children's utterances were measured using the computer facilities of the Institute of Linguistics of Stockholm University.

4.1.1 Fundamental frequencies

Fundamental frequencies were calculated from spectral displays by measuring the frequency of a particular harmonic and dividing by the harmonic number. The harmonic frequencies were determined from narrowband (78 Hz) discrete Fourier transforms (DFT's). The fundamental frequency of each vowel was measured at one point in the middle of the vowel. Figure 4.1 shows an example of a narrowband spectrum of the vowel /i/. In this example, the frequency of the tenth harmonic is divided by ten to give a fundamental frequency of 362 Hz.

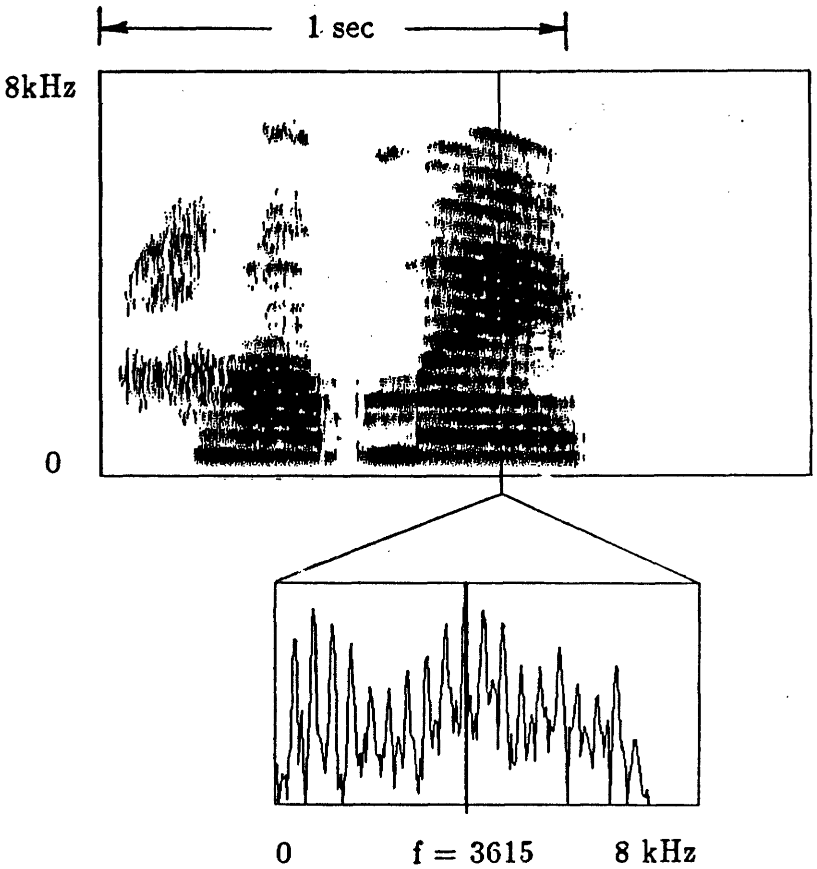

Figure 4.1: Measurement of fundamental frequency of a vowel from a narrowband DFT computed in the middle of the vowel.

The fundamental frequencies of approximately 1600 vowels produced by the American children were measured. Vowels were selected for analysis from words produced by children of ages 14, 19, 25 and 29 months. The number of words available at each age varied due to the stage of language development of the individual children. For 14-month-old children, approximately 315 vowels were analyzed. At the ages of 19, 25 and 29 months, approximately 580, 540 and 165 vowels were measured. Fundamental frequencies were averaged over each age group.

4.1.2 Durations

Durations were measured from spectrograms and waveform envelopes. In order to eliminate the low-frequency background noise which was evident in many of the recordings (typical of home environments), the utterances were high-pass filtered at approximately 200 Hz. The waveform envelope was then computed for each filtered waveform and smoothed with a low-pass filter (20 Hz). All durations were measured to the nearest 5 msec.

Vowel durations were measured primarily from spectrograms. The duration of a vowel was measured as the interval of relatively high amplitude and stable or slowly-changing formant frequencies. An example of a measurement of vowel duration from a spectrogram is shown in Fig. 4.2. The spectrogram shows the utterance "help me" spoken by a two-year-old child.

Both spectrograms and waveform envelopes were used in the measurements of syllable and utterance duration. The smoothed envelopes were analyzed for evidence of vowel-like and consonant-like intervals of high and low amplitude. The interval of time between vowel onsets was measured as the duration of a syllable. The duration of an utterance was measured as the interval during which a child produced sound on one breath. Silence or inspiration marked the endpoints of an utterance.

Durations of approximately 1140 utterances produced by 44 children between the ages of six and 29 months were measured. Approximately half of these measurements were of utterances produced at months 19, 22 and 25 months. Fewer recordings were available at the younger and older ages. Durations were averaged by month.

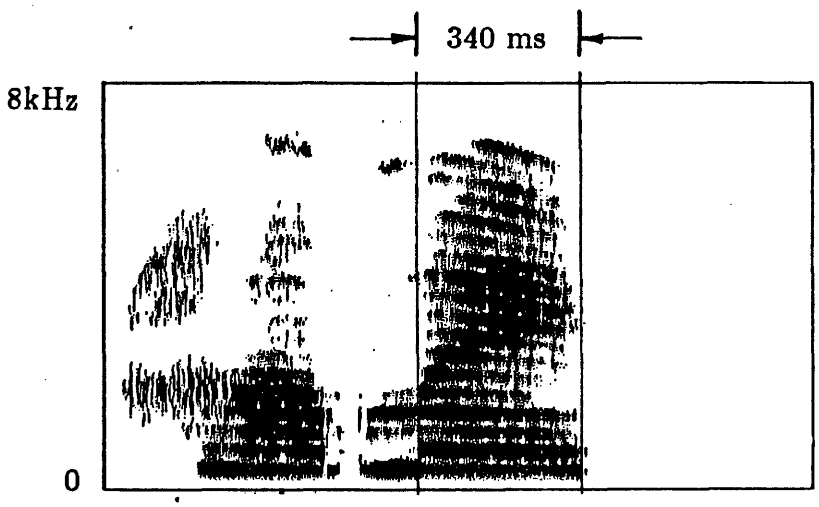

Figure 4.2: Example of a vowel-duration measurement from a wide-band spectrogram of the vowel /i/, spoken a two-year-old child.

4.1.3 Formant frequencies

Formant estimation is problematic, even for speech with a low fundamental frequency. Separating formants that are close in frequency, determining the frequency of a broad-bandwidth formant, and identifying spectral peaks which do not correspond to vocal-tract resonances present difficulties. Speech with a high fundamental frequency further complicates formant estimation.

The accuracy and the limitations of formant estimation of young children's speech were examined through an analysis-by-synthesis approach. The analyses are described in detail in Appendix II. Formant estimation is influenced by the following factors:

1. The accuracy of formant estimation depends on the closeness of formants to each other and the closeness of $F1$ to $F0$; thus no one method of formant estimation is best for all vowel configurations.

2. Accuracy also depends on $F0$ relative to each formant frequency; thus no one method is best for all $F0$ ranges.

In light of these interactions, formant frequencies were estimated using several representations of the speech signal. It appears that formant frequencies can only be measured accurately to within a range of approximately half the fundamental frequency above or below the formant frequency. Narrowband (78 Hz) discrete Fourier transforms (DFT's), spectrograms and 13-coefficient LPC envelopes were computed at a few points near the middle of the vowel and displayed. Formant frequencies were determined from LPC envelopes and the relative amplitudes of harmonics near a spectral peak.

Formant frequencies of approximately 2000 vowels were measured. Vowels were selected for analysis from productions covering an age range of approximately one year for each child. The initial month was chosen as the recording session in which the child produced at least five word attempts which contained one of the vowels /i ɪ u ʊ ɑ ɔ/. The measurements cover the ages of 12 through 29 months for the group of American children. Formant values were averaged over three vowels groups (/i ɪ/, /u ʊ/, and /ɑ ɔ/) at each month. Each average value represents between 30 and 170 measurements.

4.1.4 Formant bandwidths

Formant bandwidths were measured from steady-state portions of vowels produced by one child. Little variation in bandwidth is expected between children because children are probably similar to each other in terms of tissue properties. The vowels /i ɛ æ ʌ u ɑ/ were analyzed. Two tokens of /i/ and one of each of the others were measured. Each vowel was filtered to isolate oscillations in the waveforms due to each of the lowest three formants. The center frequencies of the filters were chosen to correspond to the formant frequencies for each vowel. Formant frequencies were estimated from narrowband (78 Hz) discrete Fourier transform and linear prediction spectra. The filter bandwidths were in most cases 800 Hz, but were narrower (400 – 600 Hz) in the case of a low first-formant frequency or closely-spaced formants. Relative amplitudes (compared to the highest amplitude oscillation) were calculated for three or more periods at two or three locations in each vowel (depending on the steady-state duration). The interval between peaks was noted. Assuming an exponential decay in amplitude of oscillation, a value $\alpha = -\ln \frac{y_{i+1}}{y_i} / \Delta t$ was calculated. The values y_i and y_{i+1} are amplitudes of consecutive peaks in the filtered waveform and Δt is the duration of the interval between peaks. Sample filtered waveforms of the vowel /i/ are shown in Fig. 4.3. An average α for each vowel was calculated to give an average bandwidth $\frac{\alpha}{\pi}$ for each of the lowest three formants of the vowels /i ɛ æ ʌ u ɑ/.

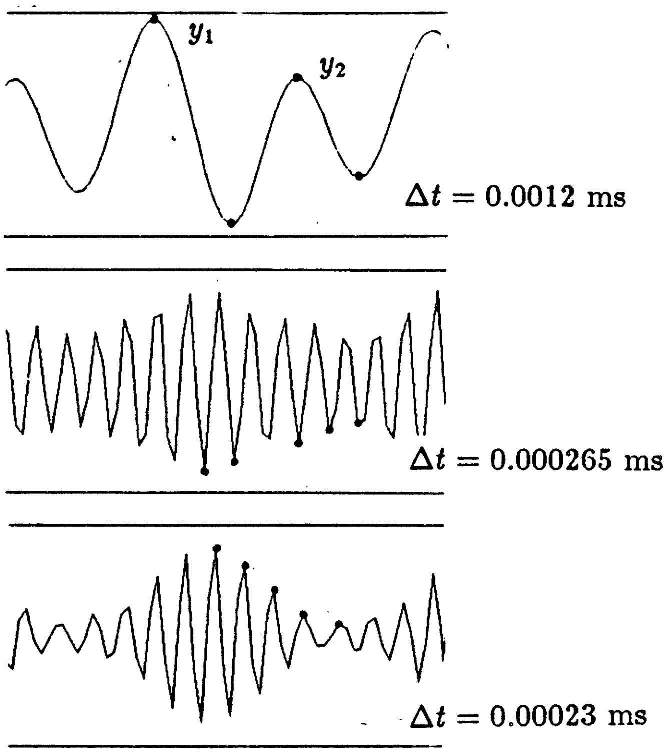

Figure 4.3: Bandwidth measurements from filtered waveform sections for the vowel /i/.

4.2 Fundamental frequencies

The fundamental frequencies of vocalizations of young children have been measured by several researchers. Eguchi and Hirsh (1969) report an average $F0$ of 298 Hz for three-year-olds. For children between the ages of 8 months to approximately 3 years, values have been measured by Keating and Buhr (1978). They report average values of 440 Hz at 67 weeks, 400 Hz at 109 weeks, and 410 Hz at 126 weeks. Robb and

Saxman (1985) present average values of fundamental frequency for groups of children within three-month age groups, covering the age range of approximately one to two years. Their measurements show a trend of decreasing $F0$ with age: 400 Hz at about 12 months, 363 Hz near 18 months, and 314 Hz near 24 months. These values are shown in Fig. 4.4. Also shown are average values of fundamental frequencies of the vowels /i ɪ ɑ ɔ u ʊ/ spoken by the children in the LIDS study. In all reports of fundamental frequencies of children's vocalizations, there is a large range of values. The measurements of the LIDS children are no exception. Fundamental frequencies range between 100 and 1170 Hz, but very few measurements are near these extreme values.

Overlaid on the measurements are values predicted by the string, spring-mass and bending beam models for ages one, two and three years. Each fundamental frequency for the string and spring-mass models is the product of an adult male fundamental frequency of 116 Hz and a scale factor (see equations 3.1 and 3.2). The values shown by •'s are the results of applying equation 3.8 of the bending beam model. It can be seen that the predictions of the bending beam model most closely approximate the average fundamental frequencies of these one- to three-year-old children.

4.3 Durations

Measurements of durations of children between the ages of six and 29 months are presented in this section. The measured durations of utterances, syllables and segments are compared to predictions based on lung size, airflow and motor-control skills of children.

Average and maximum utterance durations are shown in Fig. 4.5. The longest utterance produced by any one child within each age group is represented by the symbol ∘. Some examples of long utterances include one of 5.5 sec at age 6 months, 3.3 sec at 8 months, and 3.0 sec at 10 months. The average utterance duration (shown by •'s) decreases during months 6 through 12 and then slowly increases from 16 months

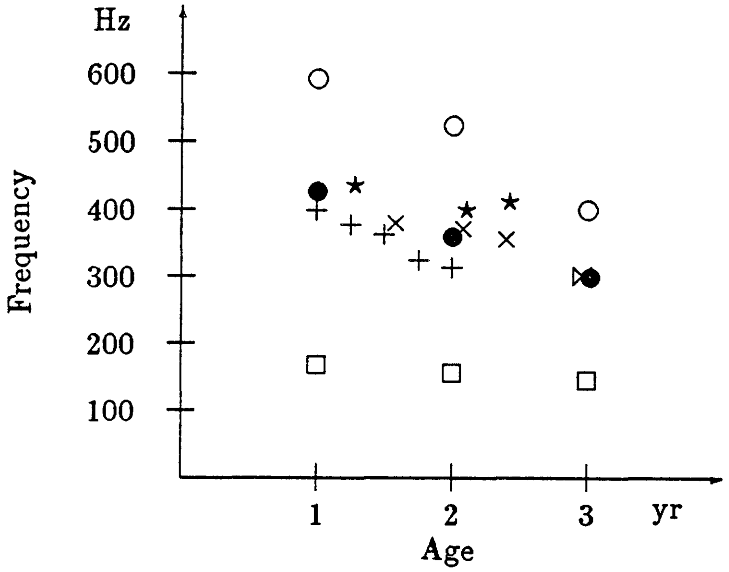

Figure 4.4: Predicted (string: ○; beam: •; spring-mass: □) and measured values of $F0$. Averages of values reported by Keating and Buhr are shown by ⋆'s; the values for three-year-olds of Eguchi and Hirsh, by ⋈; the values of Robb and Saxman, by +'s; and those of the LIDS study, by ×'s.

through 25 months. Utterances of children between the ages of 10 and 29 months are about one second in duration. During younger months, utterance duration varies considerably within the measurements for any one month. The measurements for children older than about one year show a smaller range in duration.

Predicted utterance durations are shown in Fig. 4.5 by □'s and a solid line (see also Table 3.6). The duration predictions were based on tidal volume and a glottal airflow which includes a leakage flow (the corresponding glottal area is assumed to be similar to that shown in Fig. 3.3b). Between the ages of one and two years, all of the measured utterances are shorter than the predicted maximum utterance duration. After age 16 months, a trend of increasing utterance duration is seen. This trend might be evidence of increased control over respiration. At months 25, 27 and 29, children produced utterances which are longer than the predicted durations. These older children might have used a larger lung volume or a glottal area with less leakage flow.

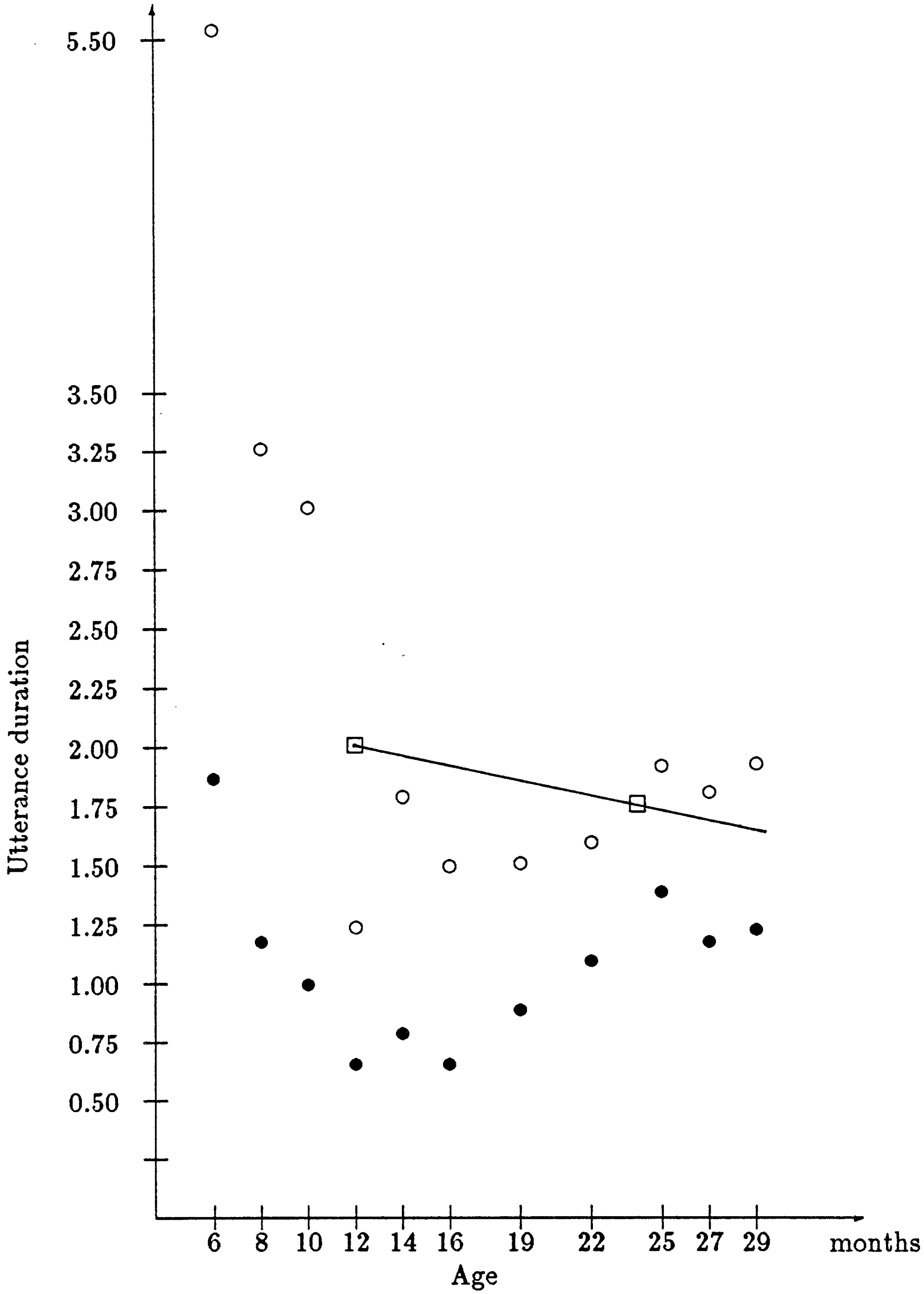

Figure 4.5: Measured average (•) and maximum (∘) utterance durations by age. Predicted maximum utterance durations are shown by □'s.

The changes in syllable duration (see Fig. 4.6) mirror those observed for utterance duration between age six months and two years. As with utterance duration, a large amount of variation in syllable duration is seen at younger ages. The durations of syllables produced by children younger than approximately eight months range from approximately 100 ms to more than 1 second. For ages older than approximately eight months, syllable duration is more uniform. Syllables produced by children between the ages of 10 and 19 months are approximately 300 – 400 ms long. Around two years of age, syllable duration is seen to increase to over 500 ms. Shorter syllables dominate again at months 27 and 29.

Children older than two years produce longer utterances consisting of more syllables than younger children. The measurements of syllable duration correspond to a syllable rate of approximately 2 – 3 Hz. This rate remains relatively constant between the ages of 10 and 29 months. Children of these ages speak about half as fast as adults. This relationship is consistent with that seen for other rates of movement (diadochokinesis and tapping) of children compared to adults.

The ability to produce sequences of syllables of approximately equal duration might be evidence of the child's developing awareness of the syllable-structure of speech. Holmgren *et al.* (1985) hypothesized that this regular timing component might appear shortly before the onset of canonical babbling (which typically appears around eight months of age). The relationship between acoustic measures of syllable timing and the onset of canonical babbling was investigated by Bickley *et al.* (1986). The appearance of syllable-like units in children's vocalizations might also be evidence of the development of control of the tongue and lips as articulators. Holmgren *et al.* (1985) documented the appearance of acoustic evidence for supraglottal constrictions in young children's articulations at the age of approximately six months.

Most of the children's utterances consisted of only a few syllables. The children of ages six to 22 months produced approximately two syllables per utterance. Older children (25 to 29 months) produced utterances containing on average three to four syllables. These values are shown in Table 4.1 and are in general agreement with

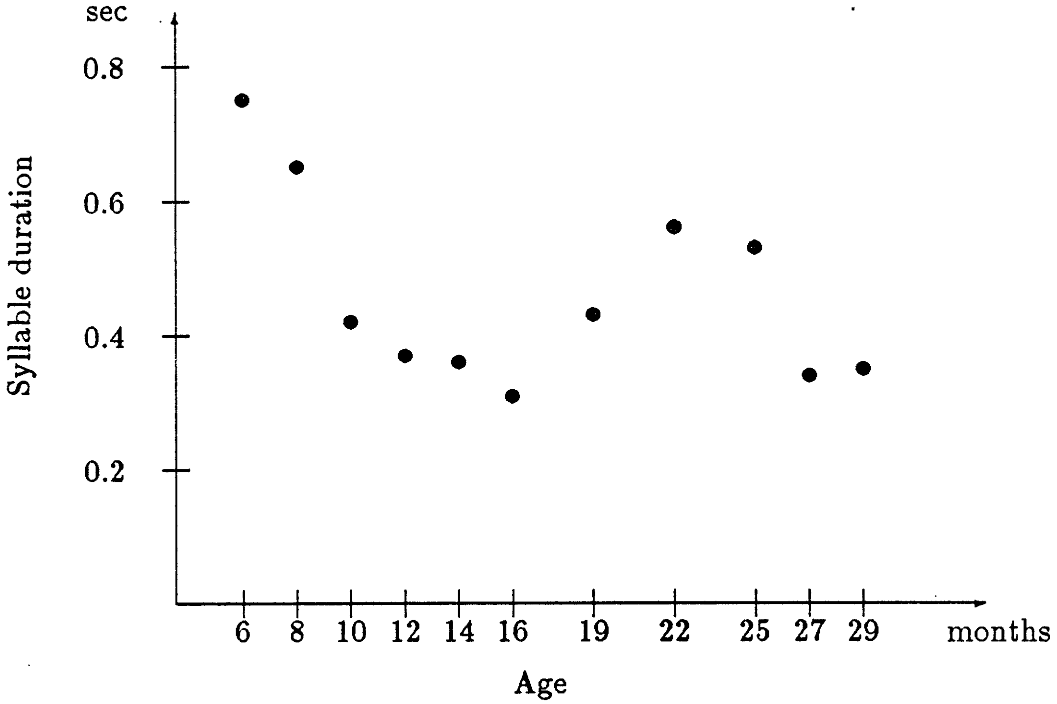

Figure 4.6: Measured average durations of syllables (•) by age.

the predictions of Section 3.2. The typical number of consonant-vowel syllables in an utterance was predicted to be between two and four for a two-year-old child.

Average durations of vowels produced by children between the ages of six and 29 months are shown in Fig. 4.7. The vowels produced by younger children were longer than those produced by older children. Average vowel duration was relatively constant at approximately 230 ms for children older than about one year. This value of average duration is approximately twice as long as the average vowel duration measured by Crystal and House (1982) for adults. In Section 3.2, vowel durations were predicted to be approximately twice as long for children as for adults due to differences in the ability to perform repetitive openings and closings of the vocal tract, as in a diadochokinetic task.

In summary, the measurements of duration are in general agreement with the predictions. Children's utterances are typically shorter than adults', but utterance duration in general does not appear to be limited by respiratory or laryngeal control.

Table 4.1: Number of syllable-like units in children's vocalizations. Ages are shown in months.

Age	Average number of syllables
6	2.4
8	2.0
10	2.1
12	1.9
14	2.2
16	2.1
19	2.0
22	2.0
25	2.6
27	3.5
29	3.5

The segments and syllables of children's utterances are about twice as long as those of adults.

4.4 Formants

4.4.1 Formant frequencies

In this section formant frequencies of the vowels /i ɑ u/ produced by 25-month-old children are compared with the formant frequencies predicted from the simple models of Section 3.3. Frequencies measured from words containing the vowel /ɑ/ are compared to the frequencies plotted in Fig. 3.7. A match between the measured and predicted frequencies means that during vowel production a two-year-old child could have used a vocal-tract configuration similar to that shown in Fig. 3.6a. Frequencies

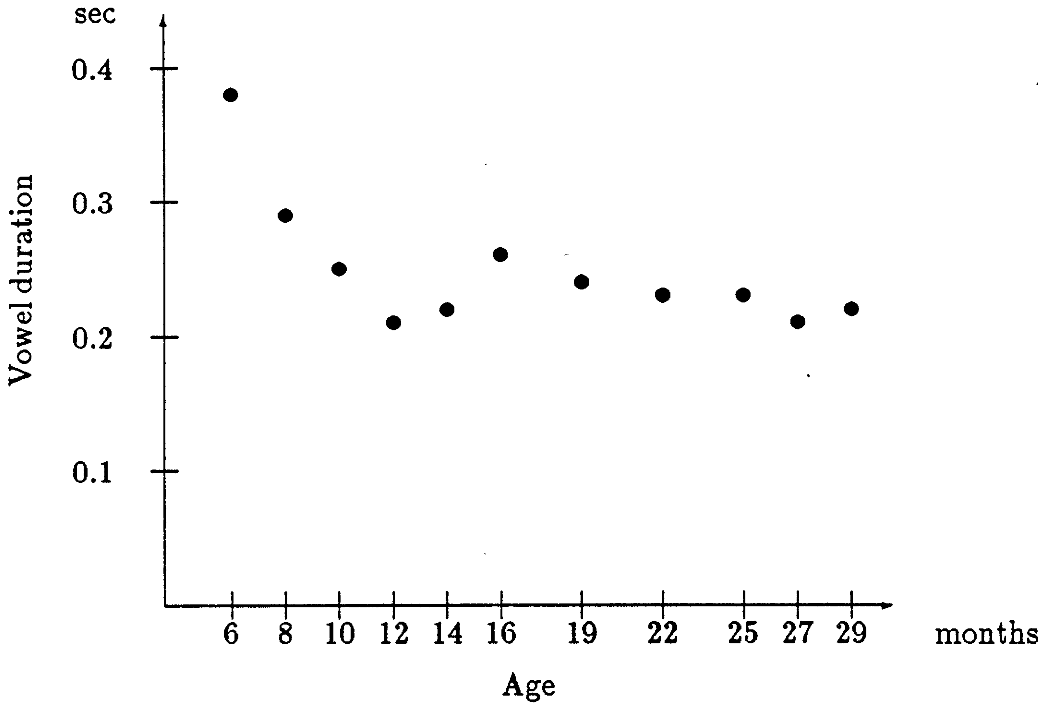

Figure 4.7: Measured average durations of vowels (●) by age.

measured from /i/-words are compared to those corresponding to Fig. 3.6b. The frequencies of the /u/'s are compared to those of Figs. 3.6c and 3.6d. From each of these comparisons, a match in frequency is taken as an indication of the vocal-tract configuration of a two-year-old.

For /ɑ/-words, average first and second formant frequencies of 1100 and 1950 Hz, respectively, were measured for 25-month olds. The lines in Fig. 4.8 represent these average values of $F1$ and $F2$; also shown are the predicted formant frequencies. For a front-cavity length of between 4 and 5 cm and the /ɑ/-like configuration of Fig. 3.6a, the predicted first and second formant frequencies are approximately 1260 and 2090 Hz. A possible explanation of the differences in frequency is that the vocal tracts of the 25-month-old children are slightly longer than the 10 cm used in the calculations of Section 3.3. The match is the closest in the region of minimum spacing between $F1$ and $F2$. In this region, the length of the front cavity can vary as much as 2 cm with minimal effect of the acoustic signal.

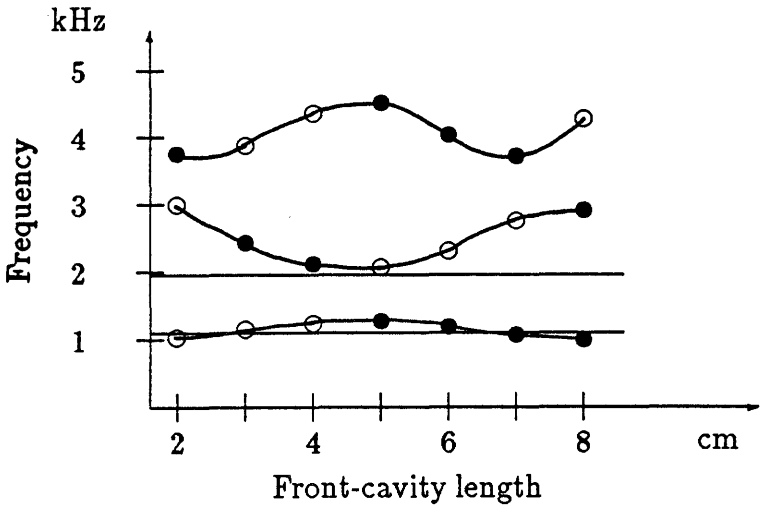

Figure 4.8: Average values (shown by the lines) of measured first and second formant frequencies of /ɑ/ vowels overlaid on predicted resonances of an /ɑ/-like configuration.

For the vowel /i/, the average $F1$ and $F2$ are 580 and 3090 Hz, respectively, for the 25-month-old children. Measured values are plotted against the predicted values of Section 3.3 in Fig. 4.9. The predicted $F2$ of 3050 Hz for a 5 cm front cavity is about the same as the measured frequency. For the corresponding $F1$, the prediction of 410 Hz is not close to the measured average value. A front-cavity of twice the area used in the predictions would result in a better match to the measured value of $F1$, but a poorer match to $F2$. In this case, the simple configuration of Fig. 3.6b is not as adequate a model as is the /ɑ/-like configuration.

The measured formants of /u/ are compared to both the rounded and unrounded configurations of Figs. 3.6c and 3.6d in Figs. 4.10 and 4.11. The measurements of 650 and 1850 Hz for the first and second formant frequency match the values of 690 and 1820 Hz of the unrounded high front vowel-like configuration with a front-cavity length of about 5 cm (see Fig. 4.10). The predicted formants for the rounded /u/-like configuration do not provide a match to the measured data.

For each of the vowels, there is variation in the measured formant frequencies. Some of the individual measurements agree quite closely with the predicted values. In other cases, the measured and predicted values are dissimilar. At least two explanations are

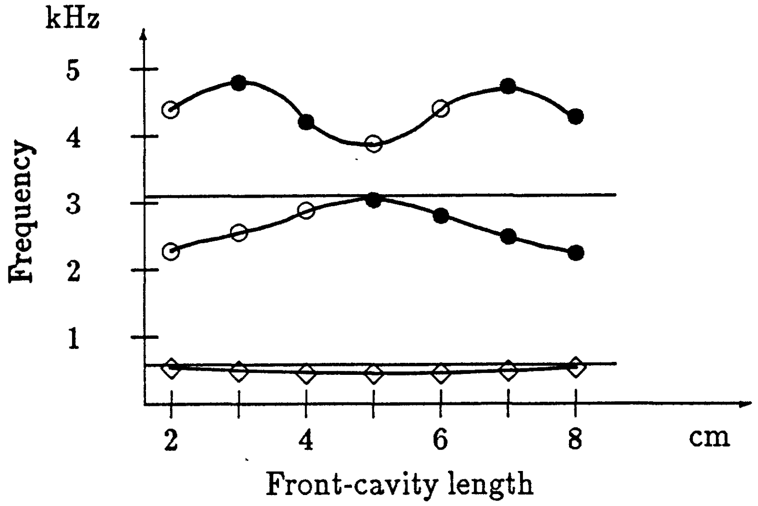

Figure 4.9: Average values (shown by the lines) of measured first and second formant frequencies of /i/ vowels overlaid on predicted resonances of an /i/-like configuration.

possible:

1. The simple models do not adequately represent the vocal-tract configurations used by the children during production of the vowels /i ɑ u/.
2. The measured formant frequencies differ from the actual values.

For some of the vowels, the second explanation seems likely. The difference between the measured and predicted frequencies is often less than one-half the fundamental frequency. This range of error is expected for the fundamental frequencies of the speech of two-year-old children.

4.4.2 Formant bandwidths

Sample bandwidth measurements were made of vowels produced by a two-year-old child. Bandwidths for each of the vowels are shown in Table 4.2; predicted bandwidths for a uniform vocal tract are also shown for comparison with the measured values.

Other researchers have measured bandwidths of the vowels /i ɛ æ ʌ u ɑ/. Fant's (1972) measurements for /i ɛ æ u ɑ/ produced by males and measurements of House and Stevens (1958) for /i ɛ æ ʌ ɑ/ are shown in Table 4.3. It can be seen that in

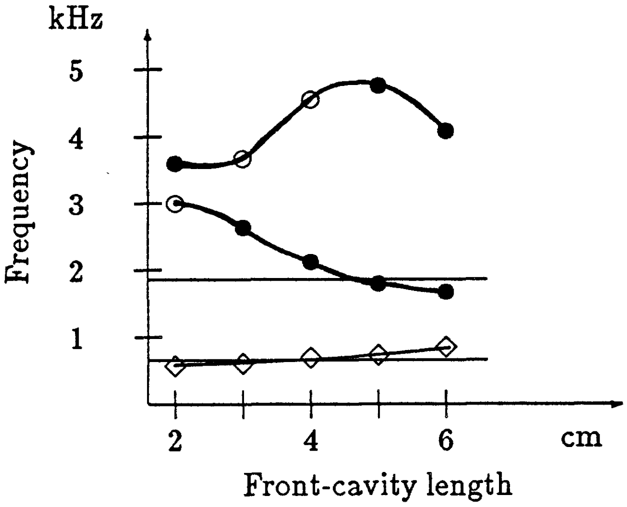

Figure 4.10: Average values (shown by the lines) of measured first and second formant frequencies of /u/ vowels overlaid on predicted resonances of an /ɯ/-like configuration.

general these previously reported bandwidth values are quite a bit lower than those listed for children's speech. This difference was theoretically predicted in Section 3.3. Fant's bandwidths of /ɑ u ɛ/ are approximately one-quarter to one-fifth as wide as those shown in Table 4.2. The second and third bandwidths of /æ/ follow this same pattern. The House and Stevens bandwidth measurements are more uniform than those reported by Fant or those shown in Table 4.2. For the vowels /i ɛ æ ʌ ɑ/, bandwidth measurements of House and Stevens fall within the range 40 – 103 Hz.

According to the predictions of Section 3.3, radiation resistance and glottal losses contribute the most to the bandwidths of children's speech. As the proportions in Table 3.7 show, the bandwidth component due to radiation (B_r) is influenced by the dimensions of the front cavity. An increase in the area of the mouth opening would cause B_r to increase proportionally. A decrease in the length of the front cavity would also cause an increase in B_r. For the back-cavity resonances, the bandwidth component due to glottal losses (B_g) is related to the average area of the glottal opening. An increase in the average area would result in an increase in B_g. For the first formant, the measured bandwidths would be consistent with a larger average glottal opening.

Table 4.2: Measured bandwidths of a child's vowels (ND: no data).

Vowel	B1	B2	B3
/i/	71	668	284
/i/	180	180	520
/ε/	119	270	426
/æ/	342	290	364
/ʌ/	48	267	197
/u/	239	265	ND
/ɑ/	163	159	384
predicted	90	132	246

Table 4.3: Previously reported bandwidth measurements (from Fant (1972) and House and Stevens (1958)).

Vowel	B1	B2	B3
From Fant:			
/i/	72	35	105
/ε/	41	58	119
/æ/	37	61	89
/u/	63	43	42
/ɑ/	45	42	100
From House and Stevens:			
/i/	66	75	76
/ε/	67	96	88
/æ/	73	92	103
/ʌ/	40	47	64
/ɑ/	53	60	66

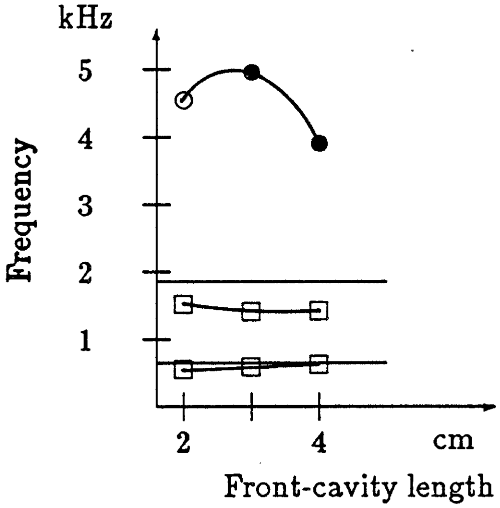

Figure 4.11: Average values (shown by the lines) of measured first and second formant frequencies of /u/ vowels overlaid on predicted resonances of an /u/-like configuration.

4.5 An example of using acoustic measurements in analyses of the development of speech

Acoustic data can provide evidence of developmental patterns. In this section, measurements of formant frequencies of vowels produced by a group of 14 children are examined for patterns of change during a period in which the children began to produce recognizable words. The measurements discussed in Section 4.4 indicate that 25-month-old children produce vowels which are acoustically distinct in terms of $F1$ and $F2$ patterns. Their productions can be interpreted to correspond to articulatory configurations which differ in tongue placement in terms of height and backing. In this section, measurements of the acoustic characteristics of the speech of some of these children at younger ages are examined in terms of $F1$ and $F2$ patterns. A pattern in the acoustic measurements could be interpreted as evidence for growth of the articulatory structures or for development of more precise articulatory skills. The measurements could also reflect a child's acquisition of a phonological system.

In this example, the analysis focuses on acoustic evidence for the acquisition of the phonological distinctions of vowel height and backing (see Bickley, 1984, for further detail). The first formant frequency depends most directly on degree of mouth opening and therefore corresponds to the height distinction for vowels. Low vowels are produced with a relatively open mouth position; high vowels, with a more closed mouth. Acoustic theory predicts a high value for the first formant of vowels produced with an open mouth (the low vowels) and a low value for the first formant of high vowels. Control of the second formant frequency corresponds to control of the body of the tongue in the front-back dimension. The theory also predicts a relatively low value for the second formant for the back vowels, resulting in $F2$ being close to $F1$, and a relatively high value for front vowels, or a large $F2 - F1$.

Vowels were grouped into three pairs: /i/ and /ɪ/ formed one pair, as did /u/ and /ʊ/; the low vowels /ɑ/ and /ɔ/ formed the third pair. Utterances were grouped by target vowel according to the word gloss, not by transcribed vowel. For example, an utterance which was glossed as "ball" was grouped with the /ɑ ɔ/ utterances, regardless of whether it sounded like [bɔ] or [bi] or [bʌ]. For each child, the average formant value was calculated for each vowel group at each month; the standard deviation was calculated for groups of three or more as a measure of dispersion of formant value. The $F1$ versus $F2$ vowel space was graphed for each child at each month (Peterson and Barney, 1952).

Graphs of $F1$ versus $F2$ for child BA are shown for months 14, 16, and 19 in Fig. 4.12. In month 14 there is overlap in both the $F1$ and $F2$ dimensions (although one of the vowel pairs did not occur in that month). At month 16, the vowels are separating along the $F1$ axis. That is, /i/, /ɪ/, /u/, and /ʊ/ are almost separate from /ɑ/ and /ɔ/ (high vowels are separating from low vowels), but the front vowels still overlap the back vowels. By month 19 there is good separation of high from low vowels and of front from back vowels; that is, /i/ and /ɪ/ have separated from /u/ and /ʊ/.

The trends of vowel differentiation along the $F1$ and the $F2$ axes can be viewed in another manner by averaging the formant values for each vowel pair for a child by session. For this comparison, the first-formant frequencies and second-formant frequencies were converted to a critical band or Bark scale. For each child, the average

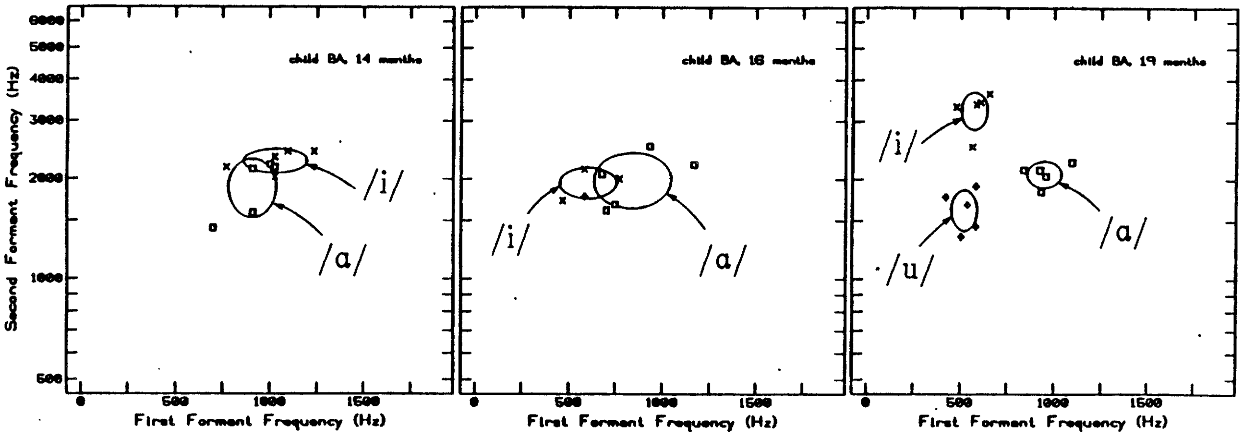

Figure 4.12: Frequency of $F1$ plotted relative to $F2$ for child BA. The ×'s represent /i,ɪ/; the +'s, /u,ʊ/; and the squares, /ɑ,ɔ/. Two standard deviations about the average are shown by an elliptical outline.

first-formant and second-formant frequencies in Bark were calculated for each vowel group at each selected month. The average value represents a "center value" for each vowel group. The standard deviations were calculated for groups of three or more as a measure of "tightness of clustering" along a dimension. Figure 4.13 shows the first formant data for child BA represented in terms of averages and standard deviations at each of five consecutive sessions (months 14, 16, 19, 22 and 25). The productions of high and low vowels overlap in the $F1$ dimension at month 14; this overlap was also seen in Fig. 4.12 above as overlap of the vowel groups in the $F1$ versus $F2$ plane. Separation of high from low vowels is seen from 16 through 19 months. This separation was seen in Figure 4.12 as separation along the $F1$ axis in the $F1$ versus $F2$ plane.

Figure 4.14 shows the development of backing for child BA as measured by the difference between the second and first formant frequencies. The average values of $F2 - F1$ and standard deviations are shown for each vowel group. The productions of front vowels and back vowels are similar in $F2 - F1$ value at months 14 and 16. Front vowels have separated from back vowels at month 19. Front and back vowels are maximally separate at month 22.

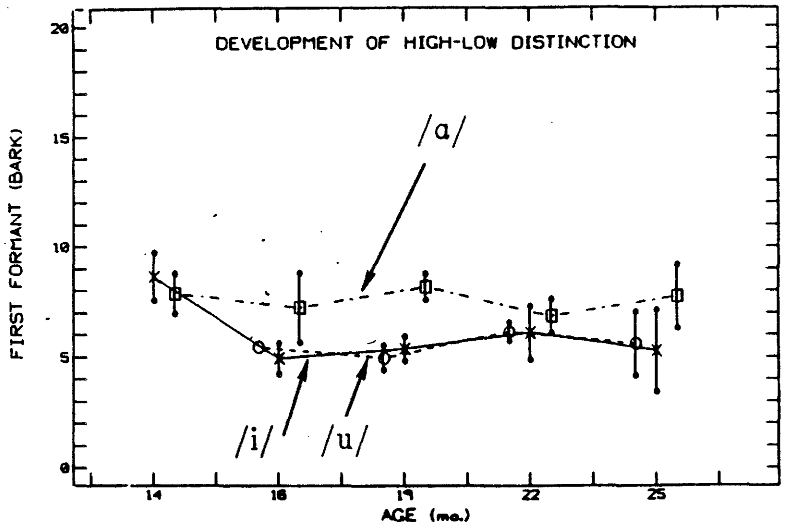

Figure 4.13: Averages and standard deviations of $F1$ in Bark for child BA.

For each child, utterances with target high vowels /i/, /ɪ/, /u/, or /ʊ/ were compared to those with target low vowels /ɑ/ or /ɔ/ by examining patterns in the data represented in the $F1$ versus $F2$ plane and the averages and standard deviations of $F1$ of the high group and of the low group. Development of the front vowels was compared to that of the back vowels by examining the $F1$ versus $F2$ plane and the averages and standard deviations of $F2 - F1$. The data on acquisition of the height and backing distinctions are displayed in the histogram in Fig. 4.15. This graph shows the number of sessions between acquisition of the height and backing distinctions. Zero indicates that the height and backing distinctions were acquired within too short a time interval to be apparent from the data. It can be seen that in most cases in which it was possible to observe a lag in time between acquisition of the two types of distinction the height distinction preceded backing. Although the children differed in the age at which the height and backing distinctions were acquired and the lag in time between distinction acquisition, the order of acquisition was height before backing in seven of the fourteen children and backing before height in only one child.

The data suggest that there does exist a pattern in the acoustic data which could be interpreted as evidence for the development of vowel height preceding development

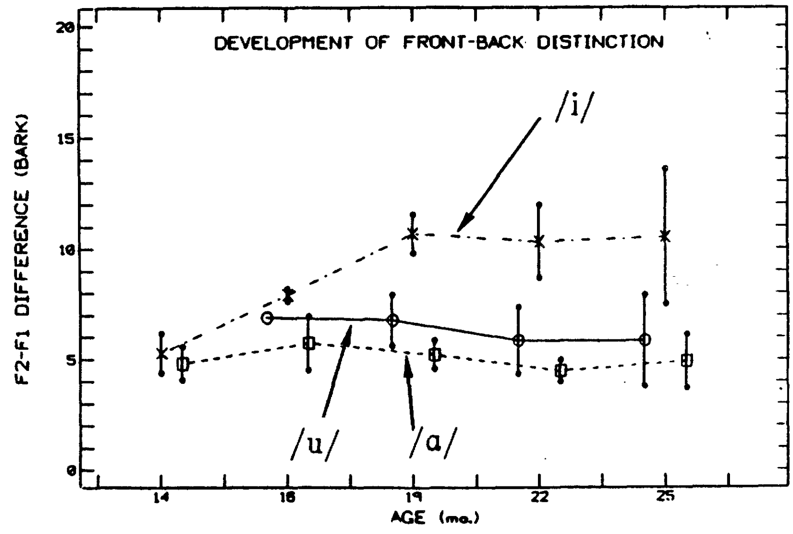

Figure 4.14: Averages and standard deviations of the difference $F2 - F1$ in Bark for child BA.

of vowel backing. These data can be interpreted as evidence of increased articulatory skills, acquisition of a distinctive feature based phonological system, or discovery of a set of phonological rules, or as some combination of these.

It is likely that articulatory development consists of progressive mastery of a series of subskills. The data shown above argue for articulatory development in terms of the ability to control the degree of mouth opening (control of $F1$) prior to the development of the ability to control the anterior-posterior positioning of the tongue body (control of $F2$). In some sense, it seems that vowel height is more salient than vowel backing. Perhaps the apparent control of vowel height is simply the result of young children being able to control mouth opening. A partial explanation might also be that visual information as well as auditory information is available to the child during the perception of vowels that differ in height (Kuhl and Meltzoff, 1982), as is kinesthetic information. Perhaps acoustic differences in $F1$ values are perceptually easier to recognize, or articulatory control of vowel height might be in some sense simpler.

An argument for the acquisition of a distinctive-feature-based phonological system is also reasonable in light of the acoustic data. We assume that children at the age of

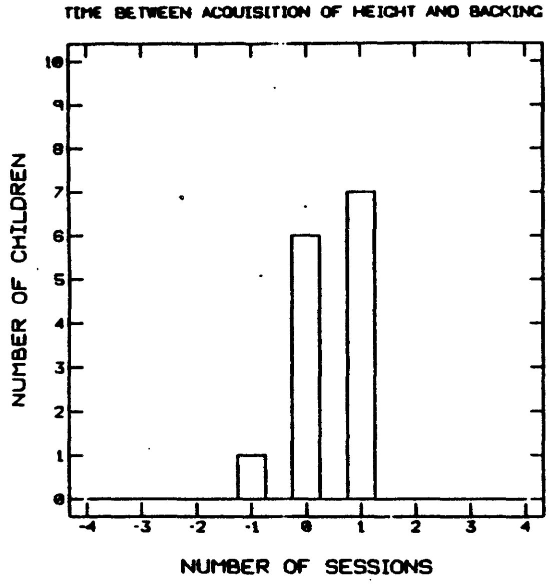

Figure 4.15: Histogram showing the number of sessions between the acquisition of the height and backing distinctions.

those in this study (one year of age and older) are sensitive to the acoustic correlates of the vowel features of height, backing, and tensing (Trehub, 1973; Kuhl, 1979; Kuhl, 1977) and that children's internal representations of morphemes exist in terms of some kind of distinctive features (Chomsky and Halle, 1968). It is possible that children's feature markings are incomplete at this stage of development. A child may have the articulatory skill but not the representation of appropriate phonological features. Thus, the child would not produce distinctions corresponding to those features. An incompletely marked vowel could be filled in across-the-board by a fixed value (for instance, all vowels could be realized as [-back] during early development) or by a random value (some vowel productions could be realized as [+back], others as [-back], with no apparent pattern). Acquisition of a feature distinction is said to occur when acoustic measurements of vowels indicate that a feature contrast is being controlled with some degree of precision (for instance, the high vowels /i ɪ/ and /u ʊ/ separate from the low vowels /ɑ ɔ/ along the height dimension). It was seen that most of the children controlled the height dimension before the backing dimension.

The changes in formant patterns discussed above cannot be predicted based solely on growth of the articulatory structures. Goldstein's (1980) results indicate that the dimensions of the vocal tracts of infants do not prohibit a child from forming /ɑ/-like, /i/-like and /u/-like configurations. The differential development of the ability to produce utterances with distinct acoustic characteristics can be interpreted as evidence of the development of a child's phonological system.

4.6 Summary of measurements

In this chapter, measurements of the acoustic characteristics of children's utterances were compared to values predicted by models of fundamental frequency, duration, and formant frequency and bandwidth.

The bending beam model of the vocal folds reflects the anatomical differences between the larynges of children and adults. Of the three models discussed in Section 3.1, the bending beam model most closely matches the measurements reported in this thesis and by other researchers. The beam model is appropriate because a child's vocal fold is stiff due to its structure which is short and thick in comparison to an adult vocal fold.

The durations of utterances, syllables, and vowels produced by young children were compared to predicted values. The predicted and measured values of maximum utterance duration are in good agreement, given the assumptions for children of a glottal airflow with leakage and of an available lung volume equal to tidal volume. As predicted, syllables and vowels of young children are approximately twice as long as those of adults.

Formant frequencies of the vowels /i ɑ u/ produced by 25-month-old children were compared to resonances of simple configurations (similar in shape to those used to model adult vowel productions). The match for /ɑ/ is fairly close; the larger spacing between $F1$ and $F2$ for children is consistent with a cross-sectional area of a vocal-tract constriction which for children is wider in comparison to vocal-tract length than for adults . For /i/, the predicted first formant differs from the measured average value by more than 100 Hz. An unrounded /ɯ/-like configuration provides a close match to the measured /u/ values. Formant bandwidths of vowels produced by a two-year-old child agree with theoretical predictions of broader bandwidths for children. A mouth opening and average glottal area which are larger in relation to vocal-tract length for the child than for an adult would account for some of the broadening. Both duration and bandwidth data support the hypothesis about average glottal area.

Developmental changes in acoustic characteristics of young children's utterances which cannot predicted from physical parameters alone motivate explanations in terms of other factors, such as phonological development; an example of such an argument is presented in Section 4.5.

Chapter 5

Conclusions

Models which are based on physical parameters explain the anatomical and physiological bases for some of the changes which occur in speech production during the first few years of life. Components of the speech production system mature at different rates throughout childhood – a child is not simply a small adult. For instance, the vocal tract grows more in length than in radius during early childhood. A similar relationship holds for vocal-fold length relative to vocal-fold thickness. Dimensions, tissue properties and motor-control skills each influence the acoustic characteristics of children's speech in various ways. The data summarized in Chapter 2 document aspects of the growth of the lungs, larynx, and vocal tract and the development of speech motor-control skills.

Acoustic characteristics of children's speech are predicted from models of children's speech production systems in Chapter 3. Parameter values for the models are selected from the data describing physical characteristics. Predictions of the models are compared in Chapter 4 to measurements of the acoustic characteristics of fundamental frequency, duration, and formant frequency and bandwidth. Changes in these acoustic characteristics can be interpreted as evidence of the development of speech. Some of the changes are the direct result of physical growth. Other changes cannot be predicted from the physical parameters alone and are therefore likely to be caused by development of linguistic or cognitive skills.

5.1 Summary of models

In Section 3.1, a new model of vocal-fold vibration is described. The fundamental frequencies of the new model are in closer agreement with those measured for young children than are the fundamental frequencies predicted by existing models. The values of a spring-mass model are too low, and values of a vibrating string model are too high. The new model is an improvement because it captures aspects of the structure of the vocal folds and the attachments of the vocal-fold tissue to the arytenoid and thyroid cartilages, which were previously ignored. For example, a one-year-old's vocal folds are approximately one-fifth as long as an adult's vocal folds, but only 10% thinner. Such structural dissimilarities between children and adults call for a new model. The model presented in Section 3.1 is based on the theory of bending beams. An important parameter of the model is the ratio of vocal-fold thickness to length. This ratio affects the fundamental frequencies which are calculated from the model.

Durations of utterances, syllables, and segments and the number of syllables in an utterance are predicted in Section 3.2. If the durations of utterances are limited by the total volume of air which a child has available for speech, then changes in utterance duration could be explained in terms of lung volumes and airflows. The airflows are in turn determined by the child's subglottal pressure, glottal waveform and vocal-tract configuration. These factors conspire to cause a child to use proportionally more airflow in relation to lung volume than an adult uses. Therefore shorter utterances are predicted for children. The average measurements of utterance duration are much shorter than the predicted maximum duration, though. Apparently physical parameters of the speech production system are not the limiting factor for the durations of children's utterances. In contrast, durations of syllables and vowels produced by children are likely to be constrained by the child's physical abilities. Longer syllables and vowels for children than for adults are predicted from data on diadochokinetic rates.

The duration analyses predict that children can produce only a few syllables per utterance. The number of syllables which a young child can produce on one breath is limited by the amount of air available for speech, the airflow through the glottis and vocal tract, and the number of vocal-tract openings and closings per second. Measurements of some speech-like tasks indicate that children cannot change vocal-tract configurations as quickly as can adults. If a child's glottal airflow is characterized by a relatively large amount of leakage flow, then the child would use about half the amount of airflow during voicing as an adult uses. The airflow used by a child in the production of voiceless fricatives is possibly greater than the airflow used by adults, due to the child's less well-developed motor-control abilities and to the differences in subglottal pressure. A smaller lung volume, a relatively larger airflow and slower articulatory movements cause children to produce fewer syllables per utterance than adults. Utterances consisting of short sequences of syllables are commonly observed in the speech of young children.

Section 3.3 includes calculations of formant frequencies and bandwidths from acoustic theory applied to simplified vocal-tract models with child-sized dimensions. Simplified models of concatenated sections of uniform cross-sectional areas can be used to account for acoustic characteristics of adults' speech. Formant frequencies calculated from these sorts of models with children's dimensions are also in good agreement with measured formants. Formant frequencies are predicted to be higher for children approximately in proportion to the ratio of vocal-tract lengths of adults and children. The quantitative analyses of formant frequencies can be used to infer vocal-tract configurations from acoustic measurements. Average formants measured from productions of /i/ and /ɑ/ by two-year-olds match the frequencies of configurations which are adult-like in general shape (the lengths of the back and front cavities are approximately equal). It appears that children, like adults, produce the vowels /i/ and /ɑ/ using configurations which are the least sensitive to changes in location of constriction. In the case of /u/, the frequencies corresponding to an unrounded configuration are closer to the measured formant frequencies than those for the rounded case.

The minimum spacing of formant frequencies is also correctly predicted for children if certain assumptions are made. A less narrow constriction in comparison to the widest area increases the minimum spacing between formants in configurations which consist of constricted and unconstricted sections of the vocal tract. Researchers have observed that young children tend to produce only front vowels. This perception might be caused by a constraint for children on the minimum spacing between $F1$ and $F2$ in /ɑ, ɔ/ productions. The measurements of $F1$ and $F2$ of two-year-olds show an average spacing which is wider than the spacing which would be predicted if the ratio of a child's constriction area to an adult's were the same as the ratio of vocal-tract cross-sectional areas. The average difference between $F1$ and $F2$ in /ɑ/ productions by two-year-old children was measured to be 830 Hz. For the frequencies of $F1$ and $F2$ of these children, this difference corresponds to about 4 Bark. Productions with this amount of spacing between the first and second formants are likely to be perceived as front or mid vowels, regardless of the degree of backing of the tongue.

Formant bandwidths of children's speech are predicted to be broader than the bandwidths of adults' speech. The larger ratio of vocal-tract area to length in children compared to adults increases the contribution of the radiation component of formant bandwidth. The broadness of the bandwidths of formants influences the measurement of formant frequencies. For instance, a broad bandwidth of the lowest front-cavity resonance in conjunction with a high fundamental frequency would cause the spectrum of an /i/ to appear to have a single peak in the region of the second and third formants.

5.2 Interpretation

5.2.1 Acoustic characteristics of children's utterances

The question "What are the acoustic characteristics of children's speech?" was posed in Chapter 1. One of the most obvious acoustic differences between the speech of children and adults is the higher fundamental frequency of children. Measurements of $F0$ are consistent with fundamental frequencies of the bending beam model described in Section 3.1. The measurements of $F0$ of speech productions by the forty LIDS children are similar to those reported by others, and serve to augment the set of measurements of acoustic characteristics of young children.

Children's utterances are typically shorter than adults'. The individual syllables and segments are about twice as long as adults', and children's utterances usually consist of only a few syllables. The durations of syllables and segments and the number of syllables in an utterance are in line with the predictions of Chapter 4. The overall utterance duration in not accounted for by the predictions.

The formant patterns of children's vowels are similar to those of adults, but are scaled to higher frequencies. In Section 3.3 a difference between the formant patterns of children and adults in the minimum spacing is predicted. The data reported in Chapter 4 show that children do produce formant patterns with less closely spaced formants than do adults. A ratio of cross-sectional areas which is closer to unity for children than for adults would explain the measured minimum spacing.

Another difference is that the bandwidths of formants tend to be much broader in children's speech than in adults'. The bandwidth contributions due to radiation and glottal opening which are predicted in Chapter 3 account for much of the difference between children and adults. Examples of a child's vowels with broad bandwidths are reported in Chapter 4.

5.2.2 Acoustic evidence for the development of speech: mechanisms and processes

Patterns in acoustic measurements over time can be interpreted as evidence for the growth of a child, the development of articulatory skill, the acquisition of a phonological system, or a combination of these factors. The change in fundamental frequency during the first few years of life reflects the increase in length of the vocal folds in relation to the increase in thickness. The bending beam frequency is the dominant component of the fundamental frequency for young children; the spring-mass frequency dominates for adults. For older children, the fundamental frequency is influenced by the effects of both the beam and spring-mass frequencies. The age at which the spring-mass term becomes dominant might correspond to a time of abrupt change in fundamental frequency.

An increase in the duration of utterances which a child produces is evidence of the growth of the lungs and an increase in the ability to control the airflow through the glottal and supraglottal constrictions. Development of the ability to change articulatory configurations more quickly would act to increase the number of syllables per utterance which a child can produce and to decrease the durations of syllables and segments.

The growth of the vocal tract in length is reflected in lower formant frequencies. Narrower bandwidths in conjunction with lower fundamental frequencies create more distinct formant peaks in the spectra of vowels. An increase in motor-control skills results would result in the child being able to create a vocal-tract configuration with a greater difference in cross-sectional areas between cavities. The growth of the oral cavity in relation to the size of the tongue could also result in a greater difference in cavity areas. Configurations which differ more in area correspond to formant patterns with more closely spaced formant frequencies. The development of the ability to form narrow constrictions thus results in the child's ability to use more of the $F1$ vs. $F2$ vowel space.

5.3 Further work

Although some questions concerning the speech of young children have been answered in this thesis, many areas of uncertainty remain. Additional anatomical and physiological measurements are needed in most areas for young children. The measurements which would most effectively augment the analysis of vocal-fold vibration are the dimensions of the vibrating cross-sectional area of the vocal folds. The fundamental frequency of the bending beam model is highly dependent on the ratio of vocal-fold thickness to length. Estimates of the glottal area during a cycle would be useful in refining the model of vocal-fold vibration and in calculating airflows. Understanding the role of tension in controlling $F0$ would be an important next step in the process of modeling vocal-fold vibration.

Measurements of phonation volumes of children are needed in order to permit better estimates of typical and maximum durations of utterances. Current estimates of utterance duration depend on assumptions of the shape of the glottal area through a cycle of vibration, including the duration of the open and closed phases. Children's vocal folds might close for a very short period of time, or might not close at all during a cycle of vibration. In either of these cases, the airflow through the glottis would be increased and the predictions of utterance duration would be correspondingly decreased.

Data are lacking for both children and adults on rates of movement of the articulators. In lieu of these data, diadochokinetic rates for younger than age five years would be useful. The production of stop consonants is influenced by the rate at which articulators can be moved.

For the predictions of formant spacing and bandwidths, the cross-sectional areas of the vocal tract during speech production are needed. As more data describing the physical parameters of the speech production system of children become available, the analyses of Chapter 3 will need to be reexamined. In particular, the mechanisms of loss and the resulting prominence of spectral peaks deserve further attention.

References

Allen, T.H., Krzywicki, H.J. and Roberts, J.E. (1959) Density, fat, water and solids in freshly isolated tissues. J. Appl. Physiology **14**:1005-1008.

Altman, P.L. and Dittmer, D.S. (1971) *Respiration and Circulation.* Bethesda: Federation of Am. Soc. for Exper. Bio.

Beckett, R.L., Thoelke, W. and Cowan, L. (1971) A normative study of airflow in children. Brit. J. Dis. Com. 13-16.

Beranek, L.L. (1954) *Acoustics.* New York: McGraw-Hill.

Berko, J. (1958). The child's learning of English morphology, Word **XIV**:150-77.

Bickley, C.A. (1984) Acoustic evidence for phonological development of vowels in young children. Working Papers **IV**:111-124. Cambridge, MA: Speech Communication Group, Massachusetts Institute of Technology.

Bickley, C.A. (in press) Modeling the acoustic characteristics of children's speech: fundamental frequency. To appear in *Proceedings of the Eleventh International Congress of Phonetic Sciences*, Tallinn, Estonia.

Bickley, C.A., Lindblom, B. and Roug, L. (1986) Acoustic measures of rhythm in infants' babbling, or "All God's Children Got Rhythm." In *Proceedings of the 12th International Congress on Acoustics*, A6-4, Toronto.

Bond, Z., Petrosino, L., and Dean, R. (1982). The emergence of vowels: 17 to 26 months. J. Phonetics **10**:417-22.

Bouhuys, A., Proctor, D.F. and Mead, J. (1966) Kinetic aspects of singing. J. Appl. Physiol. 21:483-496. Referenced in Lieberman, P. (1968) Direct comparison of subglottal and esophageal pressure during speech. J. Acoust. Soc. Am. **43**(5):1157-1164.

Broadbent, Sr., B.H., Broadbent, Jr., B.H., and Golden, W.H. (1975) *Bolton Standards of Dentofacial Developmental Growth.* St. Louis: C.V. Mosby.

Buhr, R. (1980). The emergence of vowels in an infant. J. Speech Hear. Res. **23**:73-94.

Chesnick, M., Menyuk, P., Liebergott, J., Ferrier, L. and Strand, K. (1983) Who leads whom in language development? *Abstracts of the biennial meeting of the Society for Research in Child Development.*

Childers, D., Naik, J., Larar, J., Krishnamurthy, A., and Moore, G.P. (1985) Electroglottography, speech, and ultra-high speed cinematography. In I.R. Titze and R.C. Scherer (Eds.) *Vocal Fold Physiology.* Denver: The Denver Center for the Performing Arts.

Cook, N.H. (1984) *Mechanics and Materials for Design.* New York: McGraw-Hill.

Crystal, T.H. and House, A.S. (1982) Segmental durations in connected speech signals: preliminary results. J. Acoust. Soc Am. **72**(3):705-716.

de Boysson-Bardies, B., Bacri, N., Sagart, L. and Poizat, M. (1981) Timing in late babbling. J. Child Lang. **8**:525-539.

Dickson, S. (1962) Differences between children who spontaneously outgrow and children who retain functional articulation errors. J. Sp. Hear. Res. **5**:(3)263-271.

Dworkin, J.P. and Culatta, R.A. (1985) Oral structural and neuromuscular characteristics in children with normal and disordered articulation. J. Sp. Hear. Dis. **50**:150-156.

Eguchi, S. and Hirsh, I. (1969). Development of speech sounds in children. Acta Oto-Laryngologica. **S257**:5-51.

Fant, G. (1960/70). *The Acoustic Theory of Speech Production.* The Hague: Mouton.

Fant, G. (1972) Vocal tract wall effects, losses, and resonance bandwidths. Quarterly Progress Status Reports **2-3**:28-52. Stockholm: Speech Transmission Laboratory, Royal Technical Institute.

Fant, G., Ishizaka, K., Lindqvist, J., and Sundberg, J. (1972) Subglottal formants. Quarterly Progress Status Reports **1**:1-12. Stockholm: Speech Transmission Laboratory, Royal Technical Institute.

Fant, G., Nord, L., and Branderud, P. (1976) A note on the vocal tract wall impedance. Quarterly Progress Status Reports **4**:13-20. Stockholm: Speech Transmission Laboratory, Royal Technical Institute.

Fenn, W.D. (1938) The mechanics of muscular contraction in man. J. Appl. Physics. **9**:165-177. Referenced in Kent, R.D., Kent, J.F., and Rosenbek, J.C. (in preparation) Maximum performance tests of speech production.

Ferguson, C.A. and Farwell, C.B. (1975) Words and sounds in early language acquisition. Language **51**:419-439.

Flanagan, J.L. and Landgraf, L.L. (1967/68) Self-oscillating source for vocal-tract synthesizers. *Proceedings of the IEEE-AFCRL Symp. Speech Commun. Process.*, Boston (1967); IEEE Trans. Audio Electroacoust. **AU-16**:57-64 (1968).

Fletcher, S.G. (1972) Time-by-count measurement of diadochokinesis syllable rate. J. Sp. Hear. Res. **15**:763-770.

Folkins, J.W. (1981) Muscle activity for jaw closing during speech. J. Sp. Hear. Res. **24**:601-615.

Folkins, J.W. and Linville, R.N. (1983) The effect of varying lower lip displacement on upper lip movements: implications for the coordination of speech movements. J. Sp. Hear. Res. **26**:209-217.

Franke, E.K. (1951) Mechanical impedance of the surface of the human body. J. Appl. Physiology **3**:582-590.

Fruhling, M., Basmajian, J.V., and Simard, T.G. (1969) A note on the conscious controls of motor units by children under six. J. Motor Behavior **1**:65-68.

Fujimura, O. (1961) Bilabial stop and nasal consonants: a motion picture study and its acoustical implications. J. Sp. Hear. Res. **4**:233-246.

Gay, T. (1977) Articulatory movements in VCV sequences. J. Acoust. Soc. Am. **62**(1):183-193.

Gedgoud, V.A. (1900) Anatomical peculiarities of the respiratory organs in children. Translated by S. Pelvoy, 1957. St. Petersburg: thesis. Referenced in U.G. Goldstein (1980) *Articulatory Model for the Vocal Tracts of Growing Children.* Cambridge, MA: unpublished Ph.D. dissertation, Massachusetts Institute of Technology.

Goldstein, U.G. (1980) *Articulatory Model for the Vocal Tracts of Growing Children.* Cambridge, MA: unpublished Ph.D. dissertation, Massachusetts Institute of Technology.

Golub, H.L. (1980) *A Physioacoustic Model of the Infant Cry and its Use for Medical Diagnosis and Prognosis.* Cambridge, MA: unpublished Ph.D. dissertation, Massachusetts Institute of Technology.

Glass, J. (1983). A comparison of spectra used for formant tracking. Unpublished paper. Cambridge: MIT.

Hay, L. (1979) Spatial-temporal analysis of movements in children: motor programs versus feedback in the development of reaching. J. Motor Behavior **11**:189-200.

Henke, W. (1966). Dynamic articulatory model of speech production using computer simulation. Cambridge, MA: unpublished Ph.D. dissertation, Massachusetts Institute of Technology.

Hirano, M., Kurita, S. and Nakashima, T. (1980) The structure of the vocal folds. In M. Hirano and K.N. Stevens (Eds.), *Vocal Fold Physiology.* Tokyo: University of Tokyo Press.

Holmberg, E.B., Hillman, R.E. and Perkell, J.S. (in press) Glottal airflow and pressure measurements. M.I.T. Speech Communication Group Working Papers.

Holmgren, K., Lindblom, B., Aurelius, G., and Jalling, B. (1983). On the problem of transcribing pre-linguistic vocalization in terms of linguistic categories. Presented at the Tenth International Congress of Phonetic Sciences, Utrecht, Holland.

Holmgren, K., Lindblom, B., Aurelius, G., Jalling, B. and Zetterström, R. (1985). On the phonetics of infant vocalization. In B. Lindblom and R. Zetterström (Eds.), *Precursors of early speech.* London: MacMillan.

Hoshiko, M.S. (1965) Lung volume for initiation of phonation. J. Appl. Physiology **20**:480-482.

House, A.S. (1961) On vowel duration in English. J. Acoust. Soc. Am. **33**(9):1174-1178.

House, A.S. and Stevens, K.N. (1958) Estimation of formant bandwidths from measurements of transient response of the vocal tract. J. Sp. Hear. Res. **1**(4): 309-315.

Irwin, O.C. (1948). Infant speech: development of vowel sounds. J. Speech Hear. Dis. **13**:31-34.

Irwin, J.V. and Becklund, O. (1953) Norms for maximum repetitive rates for certain sounds established with the sylrater. J. Sp. Hear. Dis. **18**(2):149-160.

Irwin, O.C. and Chen, H.P. (1946). Infant speech: vowel and consonant frequency. J. Speech Hear. Dis. **11**:123-5.

Ishizaka, K., French, J.C. and Flanagan, J.L. (1975) Direct determination of vocal tract wall impedance. I.E.E.E. Trans. Acoust. Sp. Signal Proc. **23**:370-373.

Ishizaka, K. and Matsudaira, M. (1968) What makes the vocal cords vibrate. In Y. Kohasi (Ed.), *Reports of the 6^{th} International Congress on Acoustics.* Tokyo: Maruzen and Amsterdam: Elsevier.

Ishizaka, K., Matsudaira, M. and Kaneko, T. (1976) Input acoustic-impedance measurement of the subglottal system. J. Acoust. Soc. Am. **60**:190-197.

Jakobson, R. (1941/68). *Child Language, Aphasia and Phonological Universals*, (Trans. A. Keiler, 1968). The Hague: Mouton.

Kahane, J.C. (1975) The developmental anatomy of the prepubertal and pubertal larynx. Pittsburgh: unpublished Ph.D. dissertation, University of Pittsburgh. Referenced in U.G. Goldstein (1980) *Articulatory Model for the Vocal Tracts of Growing Children.* Cambridge, MA: unpublished Ph.D. dissertation, Massachusetts Institute of Technology.

Kahane, J.C. and Kahn, A.R. (1984) Weight measurements of infant and adult intrinsic laryngeal muscles. Folia phoniat. **36**:129-133.

Kakita, Y., Hirano, M. and Ohmaru, K. (1980) Physical properties of the vocal fold tissue: measurements on excised larynges. In M. Hirano and K.N. Stevens (Eds.), *Vocal Fold Physiology.* Tokyo: University of Tokyo Press.

Kaneko, T., Komatsu, K., Suzuki, H., Kanesaka, T., Masuda, T., Numata, T., and Naito, J. (1983) Mechanical properties of the human vocal fold – resonance characteristics in living humans and in excised larynxes. In I.R. Titze and R.C. Scherer (Eds.), *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control.* Denver: The Denver Center for the Performing Arts.

Kaneko, T., Masuda, T., Shimade, A., Suzuki, H., Hayasaki, K. and Komatsu, K. (1987) Resonance characteristics of the human vocal fold in vivo and in vitro by an impulse excitation. In T. Baer, C. Sasaki and K. Harris (Eds.), *Laryngeal Function in Phonation and Respiration.* Boston: College-Hill Press.

Keating, P. and Buhr, R. (1978). Fundamental frequency in the speech of infants and children. J. Acoust. Soc. Am. **63**(2):567-71.

Kent, R.D., Kent, J.F., and Rosenbek, J.C. (in preparation) Maximum performance tests of speech production.

Kent, R., and Murray, A. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. J. Acoust. Soc. Am. **72**(2):353-65.

Kewley-Port, D. and Preston, M.S. (1974) Early apical stop production: a voice onset time analysis. J. Phonetics **2**:195-210.

Klatt, D.H., Stevens, K.N. and Mead, J. Studies of articulatory activity and airflow during speech. Annals N.Y. Aca. Sci. **155(1)**:42-55.

Kunze, L.H. (1964) Evaluation of methods of estimating sub-glottal air pressure. J. Sp. Hear. Res. **7**:141-150.

Lass, N.J. and Sandusky, J.C. (1971) A study of the relationship between diadochokinetic rate, speaking rate and reading rate. Today's Speech **19**:49-54.

Lieberman, P. (1968) Direct comparison of subglottal and esophageal pressure during speech. J. Acoust. Soc. Am. **43**(5):1157-1164.

Lieberman, P. (1980). On the development of vowel production in young children. In G. Yeni-Komshian, J. Kavanagh, and C. Ferguson (Eds.), *Child Phonology.* New York: Academic Press.

Liljencrants, J. (1985) Speech synthesis with a reflection-type line analog. Stockholm: unpublished Ph.D. dissertation, Royal Institute of Technology.

Lindblom, B. (1962). Accuracy and limitations of sona-graph measurements. In A. Sovijarvi and P. Aalto (Eds.), International Congress of Phonetic Sciences, 4th, Helsingfors, 1961. The Hague: Mouton.

Locke, J. (1983). *Phonological Acquisitions and Change.* New York: Academic Press.

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**:221-62.

Menyuk, P. (1972). *The Development of Speech.* Indianapolis:Bobbs-Merrill.

Michelsson, K., Raes, J., Thorden, C-J., and Wasz-Höckert, O. (1982) Sound spectrographic cry analysis in neonatal diagnostics. An evaluative study. J. Phon. **10**:79-88.

MIT (1985) Course notes on speech spectrogram reading. Cambridge, MA: Massachusetts Institute of Technology.

Monsen, R.B. and Engrebretson, A.M. (1983) The accuracy of formant frequency measurements: a comparison of spectrographic analysis and linear prediction. J. Speech Hear. Res. **26**:89-97.

Murphy, R., Menyuk, P., Liebergott, J. and Schultz, M. (1983) Predicting rate of lexical acquisition. In *Abstracts of the biennial meeting of the Society for Research in Child Development.*

Muta, H., Fukuda, H., Machino, K. and Kokawa, N. (1983) Basic study on manifestation of phoniatric function of the larynx. In *Proceedings of the 19th Congress of International Association of Logopaedics and Phoniatrics.* Edinburgh, Scotland.

Negus, V.E. (1929) *The mechanics of the larynx.* London: Heinemann Medical Books. Referenced in U.G. Goldstein (1980) *Articulatory Model for the Vocal Tracts of Growing Children.* Cambridge, MA: unpublished Ph.D. dissertation, Massachusetts Institute of Technology.

Netsell, R. (1981) The acquisition of speech motor control: a perspective with directions for research. In R.E. Stark (Ed.), *Language Behavior in Infancy and Early Childhood.* New York: Elsevier Science Publishers.

Nordström, P.-E. (1975). Attempts to simulate female and infant vocal tracts from male area functions. Speech Transmission Laboratory Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm, Sweden. **2-3**:20-33.

Oller, D.K. (1980). The emergence of the sounds of speech in infancy. In G. Yeni-Komshian, J. Kavanagh, and C. Ferguson (Eds.), *Child Phonology.* New York: Academic Press.

Oller, D.K. (1985). Metaphonology and infant vocalizations. In B. Lindblom and R. Zetterström (Eds.), *Precursors of early speech.* London: MacMillan.

Oller, D.K., Wieman, L.A., Doyle, W.J. and Ross, C. (1975) Infant babbling and speech. J. Child Lang. **3**:1-11.

Ostry, D.J., Keller, E. and Parush, A. (1983) Similarities in the control of the speech articulators and the limbs: kinematics of tongue dorsum movement in speech. J. Exper. Physiology **9**(4):622-636.

Pearsall, C.R. (1985). Acoustic modelling of infant vocal tracts. Cambridge, MA: unpublished S.B. thesis, Massachusetts Institute of Technology.

Perlman, A.L. and Titze, I.R. (1983) Measurements of viscoelastic properties in live tissue. In I.R. Titze and R.C. Scherer (Eds.), *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control.* Denver: Denver Center for Performing Arts.

Peterson, G. and Barney, H. (1952). Control methods used in a study of the vowels. J. Acoust. Soc. Am.. **24**:175-84.

Prins, T.D. (1962) Motor and auditory abilities in different groups of children with articulatory deviations. J. Sp. Hear. Res. **5**(2):161-167.

Robb, M.P. and Saxman, J.H. (1985) Developmental trends in vocal fundamental frequency of young children. J. Sp. Hear. Res. **28**:421-427.

Schellekens, J.M.H., Kalverboer, A.F., and Scholten, C.A. (1984) The micro-structure of tapping movements in children. J. Motor Beh. **16**(1):20-39.

Sharkey, S.G. and Folkins, J.W. (1985) Variables of lip and jaw movements of children and adults: implications for the development of speech motor control. J. Sp. Hear. Res. **28**:8-15.

Sirviö, P. and Michelsson, K. (1976) Sound-spectrographic analysis of normal and abnormal newborn infants. Folia phoniat. **28**:161-173.

Stark, R. (1980). Stages of speech development in the first year of life. In G. Yeni-Komshian, J. Kavanagh, and C. Ferguson (Eds.), *Child Phonology.* New York: Academic Press.

Stark, R.E. and Nathanson, S.N. (1973) Spontaneous cry in the newborn infant: sounds and facial gestures. In J. Bosma (Ed.), *Fourth Symposium on Oral Sensation and Perception.* Bethesda: U.S. Department of Health, Education, and Welfare.

Stevens, K.N. (in preparation) Book on acoustic phonetics.

Stockman, I.J., Woods, D.R. and Tishman, A. (1981) Listener agreement on phonetic segments in early infant vocalizations. J. Psycholing. Res. **10**(6):593-617.

Thelen, E. (1981) Rhythmical behavior in infancy: an ethological perspective. Dev. Psych. **17**:237-257.

Tiffany, W.R. (1980) The effects of syllable structure on diadochokinetic and reading rates. J. Sp. Hear. Res. **23**:894-908.

Tingley, B.M. and Allen, G.D. (1975) Development of speech timing control in children. Child Dev. **46**:186-194.

Todor, J.I. and Kyprie, P.M. (1980) Hand differences in the rate and variability of rapid tapping. J. Motor Behavior **12**:57-62.

Tooley, W.H. (1975) Development of the control of respiration. In J.F. Bosma and J. Showacre (Eds.), *Development of Upper Respiratory Anatomy and Function.* U.S. Government Printing Office.

Velleman, S.L. (1983) Flatness and spread in children's fricatives. In *Abstracts of the Tenth International Congress of Phonetic Sciences.* Dordrecht: Foris.

Vihman, M., Macken, M., Miller, R., Simmons, H. and Miller, J. (1985). From babbling to speech: a re-assessment of the continuity issue. Language. **61**(2):397-445.

von Hofsten, C. (1983) Catching skills in infancy. J. Exp. Psych. **9**(1):75-85.

Wasz-Höckert, O., Koivisto, M., Vuorenkoski, V., Partanen, T. and Lind, J. (1968) *The infant cry. A spectrographic and auditory analysis.* London: Heinemann.

Weiner, P.S. (1972) The perceptual level functioning of dysphasic children: a follow-up study. J. Sp. Hear. Res. **15**:423-438.

West, J.B. (1974) *Respiratory Physiology.* Baltimore: Williams and Wilkins.

Winitz, H. (1960). Spectrographic investigation of infant vowels. J. Genetic Psychology. **96**:171-81.

Woodson, H.H. and Melcher, J.R. (1968) *Electromechanical Dynamics.* New York: John Wiley and Sons.

Zemlin, W.R. (1968) *Speech and Hearing Science.* Englewood Cliffs, NJ: Prentice-Hall.

Zue, V.W., Cyphers, D.S., Kassel, R.H., Kaufman, D.H., Leung, H.C., Randolph, M., Seneff, S., Unverferth, J.E., III and Wilson, T. (1986) The development of the MIT Lisp-machine based speech research workstation. In *Proceedings of IEEE-IECEJ-ASJ International Conference on Acoustics, Speech, and Signal Processing*, 7.6.1-7.6.4, Tokyo.

Appendix A

Derivation of the equation of transverse motion of the vocal fold

A side view of a vocal fold displaced a small distance ξ in the transverse direction is shown in Fig. A.1. The vocal fold has thickness b cm.

Locally, there is no deformation of the vocal fold (see Fig. A.2). Therefore, the displacement is the longitudinal direction δ_1 can be written in terms of the change in displacement in the transverse direction as

$$\delta_1 = -x_2 \frac{\partial \xi}{\partial x_1} \quad . \tag{A.1}$$

By definition, the normal strain e_{11} is

$$e_{11} = \frac{\partial \delta_1}{\partial x_1} \quad . \tag{A.2}$$

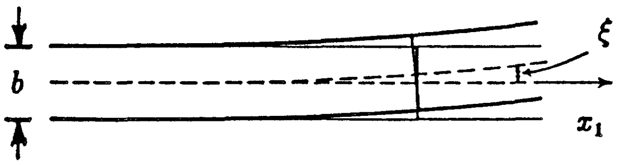

Figure A.1: Displacement of a vocal fold. The longitudinal direction is denoted by the variable x_1 (from Woodson and Melcher, 1968).

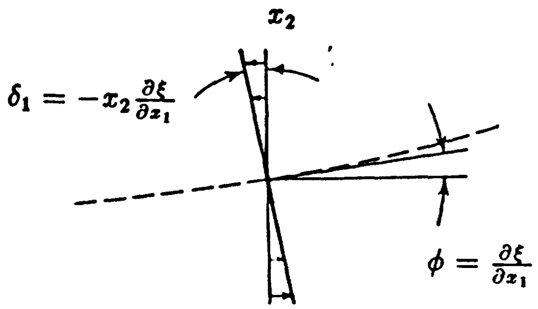
x_2
$\delta_1 = -x_2 \frac{\partial \xi}{\partial x_1}$
$\phi = \frac{\partial \xi}{\partial x_1}$

Figure A.2: Geometry of a vocal fold displaced a small amount in the transverse. It is assumed that the vocal fold is not deformed in the region of interest (from Woodson and Melcher).

Assuming that the normal stresses T_{22} and T_{33} are negligible, the elastic properties of the material can be described in terms of the empirically-determined Young's modulus and the normal stress T_{11} and strain e_{11} as

$$T_{11} = E\, e_{11} \quad . \tag{A.3}$$

Substituting for e_{11} gives

$$T_{11} = E \frac{\partial \delta_1}{\partial x_1} \quad . \tag{A.4}$$

Substituting for δ_1 gives a relationship between the transverse displace δ_1 and the elastic properties in terms of E:

$$T_{11} = -E x_2 \frac{\partial^2 \xi}{\partial x_1^2} \quad . \tag{A.5}$$

Summing all forces gives

$$\rho \frac{\partial^2 \delta_m}{\partial t^2} = \frac{\partial T_{mn}}{\partial x_n} \quad . \tag{A.6}$$

Equation A.6 represents three equations by use of the summation convention; the right-hand side of each equation consists of a sum of three terms.

In the longitudinal direction, acceleration is ignored. The force equation A.6 reduces to

$$\frac{\partial T_{11}}{\partial x_1} + \frac{\partial T_{12}}{\partial x_2} = 0 \quad . \tag{A.7}$$

Substituting for T_{11} gives

$$x_2 E \frac{\partial^3 \xi}{\partial x_1^3} = \frac{\partial T_{12}}{\partial x_2} \quad . \tag{A.8}$$

Integrating over the transverse direction (dx_2) gives

$$\frac{x_2^2}{2} E \frac{\partial^3 \xi}{\partial x_1^3} + g(x_1, t) = T_{12} \quad , \tag{A.9}$$

where $g(x_1, t)$ is specified by boundary conditions.

At the surface, there is no shear stress:

$$T_{12}|_{\pm\frac{b}{2}} = 0 \quad . \tag{A.10}$$

The shear stress in the longitudinal direction which meets these boundary conditions is

$$T_{12} = \frac{x_2^2 - (\frac{b}{2})^2}{2} E \frac{\partial^3 \xi}{\partial x_1^3} \quad . \tag{A.11}$$

Applying the force equation A.6 in the transverse direction gives

$$\rho \frac{\partial^2 \delta_2}{\partial t^2} = \frac{\partial T_{21}}{\partial x_1} + \frac{\partial T_{22}}{\partial x_2} \quad . \tag{A.12}$$

Once more integrating over dx_2 gives

$$\rho \frac{\partial^2}{\partial t^2} \int_{-\frac{b}{2}}^{\frac{b}{2}} \delta_2 dx_2 = \frac{\partial^4 \xi}{\partial x_1^4} E \int_{-\frac{b}{2}}^{\frac{b}{2}} \frac{x_2^2 - (\frac{b}{2})^2}{2} dx_2 + T_2 \quad . \tag{A.13}$$

Assuming that locally there is no deformation,

$$\int_{-\frac{b}{2}}^{\frac{b}{2}} \delta_2 dx_2 = b\xi \quad . \tag{A.14}$$

Substituting $b\xi$ into the transverse force balance equation gives

$$\rho \frac{\partial^2 (b\xi)}{\partial t^2} = \frac{\partial^4 \xi}{\partial x_1^4} \frac{E}{2} \left(\frac{x_2^3}{3} - \frac{b^2}{4} x_2 \right) \Big|_{-\frac{b}{2}}^{\frac{b}{2}} - \frac{K\xi}{Lh} \quad . \tag{A.15}$$

Rewriting equation A.15 as a function of ξ gives

$$\frac{\partial^2 \xi}{\partial t^2} + \frac{Eb^2}{12\rho} \frac{\partial^2 \xi}{\partial x_1^4} = -\frac{K}{\rho b L h} \xi \quad , \tag{A.16}$$

or

$$\frac{Eb^2}{12\rho} \frac{\partial^4 \xi}{\partial x_1^4} - \left(\omega^2 - \frac{K}{M} \right) \xi = 0 \quad , \tag{A.17}$$

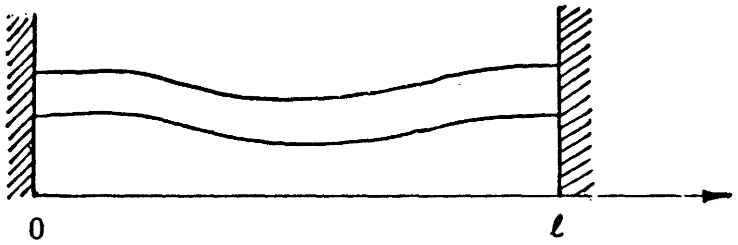

Figure A.3: Vocal fold attached at the ends to cartilage and displaced in the transverse direction.

where $M = \rho bLh$ is the effective mass. Equation **A.17** is the equation for transverse motion of the vocal fold.

The boundary conditions describe the attachment of the vocal fold at the arytenoid and thyroid cartilages, as shown in Fig. A.3. The displacement and the slope of the displacement at the endpoints are zero:

$$\xi(0) = 0 \tag{A.18}$$

$$\xi(\ell) = 0 \tag{A.19}$$

$$\frac{\partial \xi(0)}{\partial x_1} = 0 \tag{A.20}$$

$$\frac{\partial \xi(\ell)}{\partial x_1} = 0 \tag{A.21}$$

Assume a solution of the form

$$\xi = A \sin \alpha x_1 + B \cos \alpha x_1 + C \sinh \alpha x_1 + D \cosh \alpha x_1 \quad . \tag{A.22}$$

Applying the boundary conditions specifies the constants:

$$C = -A \tag{A.23}$$

$$D = -B \tag{A.24}$$

$$A = B \frac{(\sin \alpha\ell + \sinh \alpha\ell)}{(\cos \alpha\ell - \cosh \alpha\ell)} \tag{A.25}$$

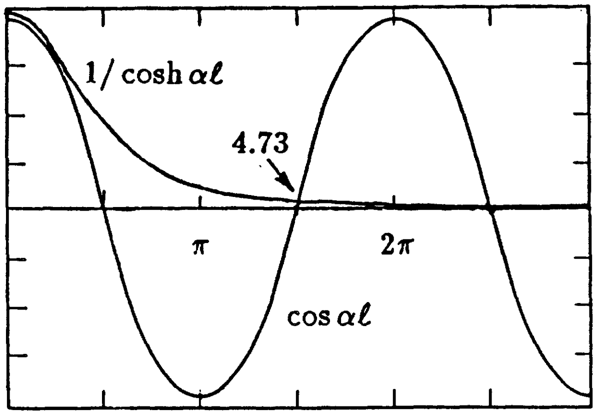

Figure A.4: Graph of solution for $\alpha\ell$.

The product $\alpha\ell$ is specified by

$$\cos\alpha\ell\;\cosh\alpha\ell = 1 \quad . \tag{A.26}$$

Equation A.26 can be found by numerical methods or by graphical solution (see Fig. A.4).

Substituting the expression for ξ into the equation of transverse motion gives

$$\frac{Eb^2}{12\rho}\alpha^4\xi - \left(\omega^2 - \frac{K}{M}\right)\xi = 0 \quad . \tag{A.27}$$

The frequencies of vibration are independent of ξ and are given by

$$\omega^2 = \frac{(\alpha\ell)^4 Eb^2}{\ell^4 12\rho} + \frac{K}{\rho b\ell h} \quad . \tag{A.28}$$

Rewriting in terms of the effective mass M gives

$$\omega^2 = \frac{(\alpha\ell)^4 Eb^3 h}{12\ell^3 M} + \frac{K}{M} \quad . \tag{A.29}$$

Appendix B

Analysis-by-synthesis approach to formant estimation

A set of synthesized utterances representing typical examples of children's speech were generated; the utterances exhibited fundamental frequencies of around 400 Hz. The formants of the synthesized utterances were measured by several formant estimation methods. Measured formants were compared to formant parameters used in the synthesis in order to assess estimation accuracy by the different methods.

Words were synthesized which included vowels with spectra representing a range of problematic formant patterns. These patterns are:

$F1$ close to $F0$ (/i,u/)

$F2$ close to $F1$ (/ɑ,u/)

$F3$ close to $F2$ (/i/)

In the first case ($F1$ near $F0$), the spectrum usually exhibits a fundamental component of high amplitude. The difficulty rests with determining whether $F1$ is below or above $F0$, and by how much. In the other cases, two formants appear as one peak in the harmonic structure, sometimes a broad peak. Locating the individual formants within the broad peak is troublesome. An example of a vowel with a spectrum exhibiting well-separated formants (/æ/) was also examined.

The synthesized utterances are based on words spoken by a two-year-old child. Figure B.1 shows a wideband spectrogram (top panel) of the word "piece." Also shown (bottom panels) are narrowband spectral slices computed (using a 25.6 ms window) near the beginning, middle and end of the vowel /i/. Three additional monosyllabic words were selected as models for synthesis: "blue," "box" and "glass."

The lowest four formant frequencies, the fundamental frequency, and waveform amplitude were measured in each natural utterance to make an initial determination of synthesis parameters. A first approximation was synthesized (using the Klatt (1980) formant synthesizer in cascade mode) using these values and a standard set of default values for formant bandwidths and glottal spectrum. In the case of /i/, $F0$ was determined accurately from narrowband spectra at several points throughout the utterance. The lowest four formants were traced on wideband spectrograms. The amplitude of voicing was estimated from the shape of the waveform envelope. The product of the radiation characteristic and the glottal spectrum was represented by appropriate real-axis and low-frequency poles and zeroes. At each of several points throughout the utterance, a narrowband spectral slice of the synthesized utterance was compared to a corresponding slice of the natural utterance. Synthesis parameters were adjusted and the synthesis-to-natural comparison repeated in an iterative manner until a 'best-match' synthesis was obtained. A 'match' was defined as a set of synthesis parameters which resulted in narrowband spectral slices in which harmonic amplitudes matched those of the natural to within ± 2 dB at each harmonic near a spectral peak (from 0-6 kHz). A 'best match' was selected based on smoothness of formant frequency, formant bandwidth and glottal parameter tracks. An example of an utterance synthesized from a 'best-match' set of parameters is shown in Fig. B.2. The wideband spectrogram and narrowband spectral slices computed near the beginning, middle and end of the synthesized word "piece" are shown.

Figure B.3 shows additional comparisons of the natural and synthetic words. The linear prediction (LPC) envelopes were computed with 13 coefficients (the sampling rate was 16 kHz in all cases).