In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Curriculum Studies

The University of British Columbia
Vancouver, Canada

Date Oct 7, 1999
ABSTRACT

The present study is a qualitative and quantitative interdisciplinary investigation of young children’s vocal development. It investigated how vocalizations of young children mutate in relation to the children’s linguistic and musical development, and the contexts in which these developments take place.

Eight girls age eighteen to thirty-eight months participated in this study. Four spoke Chinese and four spoke English as their first language. Each child was visited every four to six months over a 42 month period. Acoustic analyses were performed on recorded vocal responses, and three judges classified the vocalizations and provided perceptual evaluation.

It appears that young children have established communicative pitches that are associated with different forms of vocalizations by age two. All children consistently sang with higher fundamental frequencies than they used for speaking, while other forms of vocalization appear to be positioned consistently between singing and speaking.

Both the mean fundamental frequency data and the qualitative data suggest some possible differences in vocal pitch behaviours across language. Chinese bilingual children made comparatively less but stable distinction between their speech and song; in their acoustic intermediate vocalizations however, the boundary between speech and song was "fuzzy". English monolingual children made increasingly clearer and wider acoustical distinctions between their speech and songs; their contextual intermediate vocalizations were made up of intermittent singing and speaking.

The intermediate vocalizations observed in the present study appear to confirm that singing and speaking are two vocal phenomena that exist along a continuum. They call into question the entire concept of differences between singing and speaking both acoustically and contextually. These intermediate vocalizations offer a rich account of the linguistic and musical development of a child; they suggest that while first spoken language appears to affect vocal development, a child’s non-speech auditory environment is also crucial to the understanding of her vocal behaviours.
# TABLE OF CONTENTS

Abstract ..................................................................... ii  
Table of Contents .......................................................... iii  
List of Tables ................................................................. vi  
List of Figures ............................................................... vii  
Acknowledgments .......................................................... ix  

## CHAPTER I  EXPLANATIONS OF THE STUDY

Introduction .................................................................. 1  
Outline of the Dissertation ........................................... 2  
Psychoacoustics of Sung and Spoken Sounds ................. 3  
   The Organization of Pitch in Music and Speech .......... 3  
   The Acquisition of Linguistic Pitch ......................... 4  
   The Perceptual Processing of Musical Pitch .............. 5  
   Pitch Processing in Music and Speech ..................... 6  
   Acculturation and Music Processing ....................... 8  
A Cross-cultural, Comparative Conception of Melody ...... 9  
   The Phenomenology of Melody .............................. 10  
The Relationship Between Speaking and Singing .......... 13  
   Musical Features in Infants' Vocalizations ............... 13  
   The Relationship of Speaking and Singing in Children's Vocalizations .......................... 15  

## CHAPTER II  RELATED STUDIES OF CHILDREN'S SPEECH, SONG, AND VOCALIZATIONS

The Papouseks Study (1981) ........................................ 17  
   Implications of the Papouseks' Study ..................... 22  
The Moog Study (1976) ............................................. 23  
   Age 0 .................................................................... 23  
   Age 1 .................................................................... 24  
   Age 2 .................................................................... 25  
   Age 3 .................................................................... 26  
   Age 4 .................................................................... 27  
   Implications of the Moog Study ............................. 27  
      The issue of first spoken language ....................... 27  
      Issues concerning the criterion tasks ................. 28  
The Chen-Hafteck Study (1996) ................................... 29  
   Issues of linguistic tone and its possible effects on sung melodies .................................. 30  
   Issues of the integration hypothesis on speaking and singing .................................. 33  
   Issues associated with correlating pitch movement in speech to pitch movement in melody .................................. 34  
   Issues concerning speech melody perception ........... 37  
   Implications of the Chen-Hafteck Study .................. 41  
Statement of the Problem ........................................... 43  
Research Questions .................................................... 43
CHAPTER III FIELWORK METHODOLOGY ........................................... 45
  Concerns and Analytical Premises ........................................... 45
  Mediating the Qualitative and Quantitative Paradigms .............. 46
  Method ........................................................................ 46
    The Sample and Subjects ............................................. 46
    The Role of the Researcher ........................................... 48
  Home Visits .................................................................. 49
  Setting for the Site of Data Collection .............................. 50
  Data Collection ........................................................... 50
    Recording Children's Vocalizations ............................... 50
      Singing .................................................................... 50
      Speaking ................................................................... 51
    Observations and field notes ........................................ 52
    Informal conversation .................................................. 52
  Analysis of Data ................................................................ 53
    Acoustic Analyses: Method ............................................ 53
    Pitch Analyses ............................................................ 53
    Spectrographic Analyses .............................................. 56
    Perceptual Evaluations ................................................ 57
      Qualifications of Judges ........................................... 57
      Perceptual Categorization of Vocalizations .................. 58
      Assessment of Musical Maturity ................................. 59
  Summary of Sources and Strategies for Data Analyses .......... 59

CHAPTER IV DESCRIPTION, CATEGORIZATION, AND OBSERVATION
  OF CHILDREN'S VOCALIZATIONS .............................................. 61
  Subjects ........................................................................ 61
  Classification of Vocal Forms ........................................... 62
  Results of Acoustic Analyses ............................................ 63
  Discussion ...................................................................... 66
  Analysis of Vocal Development: All Subjects ....................... 78
    One- and-a-Half Level .................................................. 78
      General Speech Development of the Child ..................... 78
      Singing ...................................................................... 83
      Intermediate Vocalizations ........................................ 85
    Two to Three Level ...................................................... 86
      General Speech Development of the Children ................. 86
      Singing ...................................................................... 92
        Learned songs .......................................................... 93
        Spontaneous songs .................................................. 98
        Intermediate Vocalizations ...................................... 101
    Three to Four Level ..................................................... 107
      General Speech Development of the Children ................. 107
      Singing ...................................................................... 109
        Learned songs .......................................................... 109
        Spontaneous songs .................................................. 111
        Intermediate Vocalizations ...................................... 115
    Four to Five Level ......................................................... 122
      General Speech Development of the Children ................. 122
      Singing ...................................................................... 123
      Intermediate Vocalizations ........................................ 131
LIST OF TABLES

TABLES

2.1 Development of Melodic Contour in Tanya's Vocalizations ..................19
2.2 Tanya's Vocal Imitation and Learning to Sing a Song .......................20
4.1 Some Basic Information About the Subjects ..................................61
4.2 Mean $F_0$ for Children's Vocalizations Across Age Levels
   and Languages ...........................................................................64
4.3 Mean $F_0$ for Singing and Speaking for Individual Subjects ...............65
4.4 A Summary of Selected Studies of $F_0$ Characteristics of Preschool
   Children .................................................................................72
4.5 A Summary of Features Observed in the Children's Vocalizations ........146
5.1 Some Developmental Traits Observed in Intermediate
   Vocalizations of Children ..........................................................153
5.2 Some Parallels Observed in the Children's Developmental
   Milestones ................................................................................154
LIST OF FIGURES

FIGURES

3.1 A CSL Display of the Fundamental Frequencies of a Voiced Data ..... 55
3.2 A CSL Spectrogram of a Singing Segment .................................. 56
4.1 Mean $F_0$ for Chinese Utterances .................................. 68
4.2 Mean $F_0$ for English Utterances .................................. 68
4.3 A Comparison of Mean $F_0$ for Speaking Between Selected Studies and the Present Study .................................. 71
4.4 Mean $F_0$ for Speech Utterances Across Languages ................. 74
4.5 Mean $F_0$ for Speaking and Singing Across Languages ............. 74
4.6 Mean $F_0$ for Singing of Each Subject .................................. 75
4.7 Differences Between Mean $F_0$ for Speaking and Singing Across Languages .................................. 77
4.8 A Spectrogram of a Vocal Sequence of Heidi Collected When She Was Reluctant to Communicate .................. 80
4.9 Heidi Uttering “aah-ap-ple” .................................. 81
4.10 Heidi Saying Good-bye to Her Grandmother .............................. 82
4.11 Heidi “Singing” a Taiwanese Nursery Rhyme ........................... 85
4.12 Amber’s Spontaneous Singing When Bouncing Around in Circles .......... 89
4.13 Amber’s Learned Song Performed While Sitting on a Couch .......... 90
4.14 A Speech Sequence of Vicki Reading a Storybook ...................... 91
4.15 A Pitch Extraction of Vicki’s Cantonese Speech ........................ 92
4.16 Wendy Substituting Nonsense Syllables for Forgotten Words .......... 95
4.17 A Transcription of Vicki’s Performance of Twinkle, Twinkle Little Star .................................. 100
4.18 A “Prosody of Poem” By Heidi .................................. 102
4.19 Amber’s “a la-dy Oh!” Motive .................................. 114
4.20 A Spectrogram of Heidi’s “Drawing Narrative” ........................... 119
4.21 A Modified Melodic Pattern in Jenny’s Song ........................................ 125
4.22 A Modified of a Repetitious Melodic Pattern in Polly’s Song .............. 125
4.23 A Modified Teddy Bear Song Performed by Amber and Polly ............ 126
4.24 A Segment of “Story-Singing” Versus “Story-Speaking” .................... 138
4.25 A Transcription of the Recurring Melodic Motive in Polly’s “Story-
Singing” ........................................................................................................ 142
4.26 A Sequential Treatment on the Word “Grandma” in Polly’s “Story-
Singing” ........................................................................................................ 143
4.27 Polly Sang an Additional Embellished “ah” Sound to Round Off a Verse ........................................................................................................ 144
5.1 A Graphic Depiction of Intermediate Vocalizations in the Vocal
Development of Children ............................................................................. 152
5.2 Melodic Features Observed in the Children’s Learned Songs and
Spontaneous Songs .................................................................................... 158
6.1 List’s Chart for Classifying Vocal Forms Intermediate to Speech and Song .................................................................................................. 209
ACKNOWLEDGMENTS

I am indebted to members of my supervisory committee, Dr. John Gilbert, Dr. Janet Werker, and Dr. Linda Rammage, for their immense support in putting this interdisciplinary study together. Their scholarly insights and stimulating perspectives are critical to the present thesis. Deep appreciation also goes to them for stretching their workload to walk me through a difficult phase of my program. In many respects, it had been a truly privileged learning experience for me.

I am grateful to Dr. Robert Walker, who in the past four years nurtured this research project and consistently assured me of its worth. Bob also provided me training on acoustic analysis, and delivered enlightening inspirations through seminars and our weekly discussions.

My sincere thanks go to Dr. Guy Carden, who provided extensive and invaluable technical support on acoustic analysis, and assisted me to interpret the acoustic data. I would like to thank Dr. Carolyn Johnson for her meticulous editorial suggestions, and Dr. Kathleen Pichora-Fuller, who generously commented on the auditory aspects of the study. Many of my ideas have also been clarified in respond to valuable comments from Dr. Sandra Trehub, Dr. Laural Trainor, Dr. Peta White, Dr. Elvi Whittaker, and Dr. Rushen Shi; I thank them for their enthusiasm and interest in my study. My appreciation is also expressed to Mrs. Sandra Davies and Sr. Fleurette Sweeney, who patiently read early drafts of the thesis.

I must thank in particular the families who allowed me to collect data at their homes. My observations were greatly enriched by the informative conversations I had with the parents. I also owe a special debt to the children who shared with me some most precious moments of their childhood. Their voices still reverberate within me.

And to my husband, Tak-shun, whose enduring faith and abiding love carried me through an otherwise lonely journey, I thank you.
CHAPTER ONE
EXPLANATIONS OF THE STUDY

Introduction

This study is concerned with the vocal development of preschool children. My interest was to learn more about what happens during the preschool years when a child learns to speak and sing mostly relying on the natural home environment. It is remarkable that every normal child acquires a first language without special training or carefully sequenced linguistic input. However, the capacity to sing is not as easily acquired. The population of children who can carry a tune\(^1\) is significantly less than those who can engage in a daily conversation. Also, a young child's singing may appear to be private compared to that of an adult. Very often parents, especially mothers, serve as interpreters when their young child's vocalization is not readily understood by a wide range of audiences. What constitutes a child's vocalization? When does a child "sing"? Why are listeners unable to decode a child's vocalization as accurately as a parent? An investigation of the culture\(^2\) of young children's vocalizations may provide some answers to these questions.

In this qualitative and quantitative interdisciplinary study, I hypothesize that the boundaries between speaking and singing are often "fuzzy" in preschool children's vocalizations. I argue that an investigation of young children's singing should embrace their production of speech, songs, and other vocalizations which may not be readily

\(^1\) Technically speaking, it is the ability to maintain a sense of tonality and to perform the melodic intervals relatively accurately.

\(^2\) I use the term "culture" here in the specific context of qualitative inquiry, where it informs us of the phenomenon at hand. "In ethnography, culture is used as an analytic not a descriptive term. In other words, the term does not describe a set of traits of a group but refers to a form or pattern abstracted from observed behaviour" (Schwandt, 1997, p. 26). Hence, framed in terms of meaning and discourse drawing respectively on behaviours observed, ethnography is "the process and product of describing and interpreting cultural [sic] behaviour" (p. 44).
classified into either speech or song. The following topics serve as foundations for the present study:

- psychoacoustics of sung and spoken sounds;
- a cross-cultural, comparative conception of melody; and,
- the relationship between speaking and singing.

This foundation guides my investigation of young children's vocalizations and provides a rationale for including perspectives from adjacent disciplines. The following outline serves as an overview for the dissertation. Operational definitions for terms used in the present study may be found in Appendix A.

Outline of the Dissertation

In this chapter, *Explanations of the Study*, I discuss pertinent issues that have shaped the development of the research problems: psychoacoustics of sung and spoken sounds; a cross-cultural, comparative concept of melody; and, the relationships between speaking and singing.


Chapter three, *Fieldwork Methodology*, discusses conceptual frameworks and how they relate to the choice of methodology. The design of the study is presented along with details of the ethnographic observations and acoustic analyses used.

Chapter four, *Description, Categorization, and Observation of Children's Vocalizations*, presents results from acoustic analyses and the ethnographic observations. The chapter concludes with a summary of the key features observed in children's vocalizations collected at each age level.
Chapter five, *Sing Me a Song of a Preschool Child*, revisits the research questions and presents a discussion of the findings within the context of prior research. Pedagogical suggestions are made based on the findings of this study.

**Psychoacoustics of Sung and Spoken Sounds**

In studies of children's singing, investigators have traditionally equated singing ability with achievement of good intonation. Early investigators characterized children who failed to sing in tune as singing-disabled, and even assigned them labels such as "droners", "grunters", "growlers", "backward singers", "tone dumb", "tone deaf", "monotones" (Bentley, 1969). In existing research literature a concern for pitch accuracy has outweighed that for other observable singing behaviours such as the quality and phonation of the singing voice. I begin therefore by examining the concept of pitch in music and speech.

**The Organization of Pitch in Music and Speech**

Pitch is a psychoacoustic concept that has distinctive functions in speech and music. In music, western melodies are constructed on systems of organized pitches where the pitch of one tone (A = 440 Hz) is used as a point of reference. Musical interval, the perceived relations between two or more pitches within a musical context, are manipulated to establish a sense of a tonal centre around which melodies and harmonies are constructed.\(^3\) Musical interval, rather than discrete pitch, is therefore, a more practical structure used in discussions of issues pertinent to Western music. (See Shepard, 1982, for various structural representations of musical pitch.)

In speech, pitch has been used to describe certain characteristics of morphemes, syllables, and of sentences (linguistic pitch). A distinctive function of pitch (the fundamental frequency, \(F_0\)) is used to contrast two linguistic systems--tonal and nontonal.

---

\(^3\) Melody is created when two pitches are sounded in succession (melodic interval), and harmony is created when two or more pitches are sounded simultaneously (harmonic interval).
Tonal language is defined as "a language characterized by variations in pitch or tone which distinguish the meanings of words of the same or very similar written form or of the same vowel sound or sounds" (Pei & Gaynor, 1954, p. 218). Tonal language speakers use relative pitch to distinguish one lexical item from another (Van Lancker & Fromkin, 1973), hence, the scale of the pitch contour for a phoneme has lexical function. Mandarin and Cantonese constitute the largest group of tonal languages (Anderson, 1978): there are four relative pitch levels in Mandarin and nine relative pitch levels in Cantonese.  

Nontonal language speakers use variations in the shape of the pitch contour over a sentence to convey grammatical (the formation, nature or function of words in the structure of the sentence) and paralinguistic (such as emotional tones and pragmatics) but not lexical (the definition and description of the various meanings of words) information. For instance, an English declarative sentence has a falling contour and some English question have a rising contour. Bolinger (1955) describes such intonation contour in languages as "the melodic line of speech, the rising and falling of the 'fundamental' or the singing pitch of the voice" (p. 20). Speakers of tonal languages use pitch in this way in addition to its phonemic role.

The Acquisition of Linguistic Pitch

Evidence suggests that both acquisition and perception of language specific intonation patterns emerge early in a child's development. It has been claimed that the first speech utterances of a child exhibit language specific patterns of intonation (Halliday, 1975), and in the first year of life, these intonational differences correspond to attributes of the target language (Whalen, Levitt, & Wang, 1991). By age two, variations in prosodic patterns between children of different languages are observed (Bolinger, 1964), and

---

Chao (1956) explains the function of pitch in Chinese as follows: The actual pitch movement of Chinese speech is the algebraic sum of tone and intonation. If an upswing of a sentence coincides with a rising tone, the result will be higher than usual. If a falling intonation coincides with a rising tone, the rising tone will rise less or be made at a lower register than the first part of the sentence. (p. 53)
children are capable of perceiving and producing the characteristic intonation patterns of their native language (de Villiers & de Villiers, 1978; Nakazima, 1962). Cantonese-speaking children are reported to have acquired all vowels and tonal contrasts by age two (So & Dodd, 1995; J.K.-P. Tse, 1978; S. M. Tse, 1982).

The Perceptual Processing of Musical Pitch

Musical experiences are phenomenological and therefore, the processing of musical pitch is closely related to the subjective perception of an individual. Moore (1997) explains the perception of musical pitch as the "attribute of auditory sensation in terms of which sounds may be ordered on a musical scale. . . . A sound that evokes a pitch is often called a 'tone', especially when the pitch has a clear musical quality" (p. 3). Hence, the perception of a musical pitch may be considered an interpretation of a sound that has musical qualities. Jourdain (1997) describes this perception as an act of modeling musical relations in the sound we hear:

When sometimes we descry a snip of a melody in the song of a sparrow, or a wisp of harmony in the chanting of whales, it is our brains, not theirs, that have discovered an opportunity to be musical . . . One reason we hear music when animals don't is that our brains are able to manipulate patterns of sound far more complex than those the brains of other animals can manage. . . As our brain encodes these relations, the sensations of sound arise. It's not that our brains assemble a web of relations into music and then "hear" it. Rather, hearing is the act of modeling such relations. (p. 4)

According to Jourdain's perspective, the perception of a pitch, therefore, requires the brain to conceive a musical interpretation and assign its acoustic information to a specific category, whether it be an acoustical or a musical one. The former is a form of

---

5 Perceptual differences cause a number of phenomena labelled as subjective pitch, residue pitch, or the missing fundamental. (See Appendix B for a summary of various pitch perception theories. For more detailed discussion, see de Boer, 1976.)

5
innate sensory capability (acoustical pitch), while the latter has been the result of a categorizing pitched sound from our environment (musical pitch).

**Pitch Processing in Music and Speech**

Evidence suggests that within the first year of life, infants already possess sophisticated schemas for the perception of melody (see Trehub, Unyk & Trainor, 1993). They are sensitive to changes in the temporal properties of musical sequences (Demany, McKenzie, & Vurpillot, 1977; Chang & Trehub, 1977a; Morrongiello, 1984) and are able to discriminate different pitches and their changes in temporal order (Chang & Trehub, 1977b; Morrongiello, 1986). Infants of 9 to 11 months are reported to be capable of discriminating sets of melodies having the same contour but composed of different component frequencies and melodic intervals (Trehub, Thorpe, & Morrongiello, 1987). Infants are also capable of imposing temporal structure in organizing musical sequences into patterns (Thorpe, Trehub, Morrongiello, & Bull, 1988). Trehub, Bull, and Thorpe (1984) report that infants responded differentially to changes in temporal order depending on whether these violate or preserve the pattern or direction of successive pitch changes of the melody. Given the important role of intonation in early vocalizations and young infants’ exceptional performance in pitch-contour processing of musical and speech patterns, it is of interest to investigate a possible link between first spoken language and musical pitch perception. Such a link has been investigated in adults.

A series of studies was conducted on the perception of the “tritone paradox”, a phenomenon in which the tones of a melodic pattern are well defined with pitch class, but are vaguely defined with pitch height (Deutsch, 1987, 1991; Deutsch, Kuyper, & Fisher, 1987; Deutsch, North, & Ray, 1990). Subjects were asked to judge whether tone pairs related by exactly a half-octave appeared to form an ascending or descending pattern. The pitch class may be defined as “[t]he category into which a tone falls according to its tone chroma, or placement among the twelve equal-tempered divisions of the octave, but which does not distinguish enharmonic spelling or octave of placement.” (Butler, 1992, p. 233)
investigators reported a significant and consistent correlation between the subjects’ differing judgment of the criterion patterns with the linguistic communities in which they grew up, and with the fundamental frequency of the subjects’ spontaneous speaking voice. While a difference in speaking pitch range was found between Californians and speakers who grew up in Southern England, their judgment of the tritone paradox also differed. The Californians tended to perceive the tritones as ascending but subjects who grew up in Southern England tended to perceive the tritones as descending (Deutsch, 1991; 1994). Based on their findings, Deutsch et al. suggested that subjects in their studies had acquired an internal representation of the pitch classes according to the prevailing pitch range imparted by their linguistic communities. They suggested that the presence of an internal representation of pitch classes might be responsible for their differing perception of the “tritone paradox”. Because it is assumed that this is a developmental process, a relationship between childhood speech processing and musical pitch processing is hypothesized.

Is Deutsch’s proposal credible? Abundant evidence shows that even with differences in body size accounted for, the mean speaking fundamental frequency varies significantly across communities of different ethnicity and language. (See Dolson, 1994 for a review of related studies and this position.) In addition, extensive findings on experiential influences on infant speech processing suggest that infants are already attuned to language-specific properties of their native language by the end of the first year of life (Werker, Gilbert, Humphrey, & Tee, 1981; Werker & Tees, 1984; 1999). Considering this salience in the production and perception by a native speaker of a language, hypothesis of different acquired internal representation of pitch classes based on the linguistic environment in which one grew up appears persuasive. Neurophysiological studies confirm that repeated experience increases the strength of interconnections in the brain, and
neural network models confirm the plausibility of a developmental process in keeping with the Deutsch et al. proposal (Churchland & Sejnowski, 1992).

**Acculturation and Music Processing**

Acculturation is a form of learning through repeated experience and its effects on music processing are well documented. Trehub, Cohen, Thorpe, and Morrongiello (1986) examined infants' and preschool children's performance in a task involving detection of a semitone change in melodies developed according to Western music conventions. Whereas 6-month-old infants were reported to be able to perceive mistuning equally well in both culturally familiar and unfamiliar melodic patterns in both western and nonwestern melodies, 12-month-old infants demonstrated better perception in culturally familiar melodic patterns. Preschoolers were better able to discriminate music of their culture, whereas adolescents and adults were found to be even more affected by acculturation. Thus, it appears that the effects of becoming adapted to the musical patterns of a culture are commensurate with age.

Krumhansl and Keil (1982) found that older school-age children show increased awareness of tonal structure in musical sequences. Such children are reported to have better retention of diatonic than nondiatonic melodies, increased sensitivity to key information, and preferences for diatonic over nondiatonic notes in appropriate contexts. Similarly, Lynch, Eilers, Oller, and Ubano (1990) studied the ability of 6-month-old English-language infants to notice mistuning of less than a semitone in melodies on either the major, minor, or Javanese pelog scales. These infants noticed mistuning in all scales (p < 0.05), while adults affected by acculturation performed better in perceiving mistuning in major and minor scales than in the Javanese pelog scale. In a follow-up study, Lynch and Eilers (1992) concluded that their 6-month-old subjects had not yet developed sufficient

---

7 Similar findings are also reported by Werker, Gilbert, Humphrey, and Tee (1981) in performance of speech perception by infants.
perceptual knowledge of the interval patterns of the major scale to enhance discrimination performance in that scale.

Divenyi (1979) provided empirical support for the role of acculturation in the acquisition of musical pitch concepts. He pointed out that pitch concepts are noticeably different among people with congenital hearing loss and those who lose their hearing after auditory gestalts are formed. First, Divenyi reported that those with congenital hearing loss have difficulties in performing frequency discrimination tasks, while those who lost their hearing after they have had early experience perform much better. He attributed the difference between these two groups to the significance of auditory experience needed for frequency discrimination. Second, those who lost their hearing only after they had more advanced pitch perception, demonstrated much better control over vocal pitch. Walker (1990) accounts for this finding in effects of musical acculturation, whereby those with congenital hearing loss did not have the opportunity to be exposed to culturally appropriate musical structures. Putting the findings of Divenyi together with earlier discussions, we can hypothesize that first spoken language and acculturation are both crucial in young children’s processing of music.

A Cross-cultural, Comparative Conception of Melody

The inherent connection between language and music warrants attention in investigations of vocalizations of young children. Levis (1963) suspects that “the difference that exists in the natural selection of speech sounds in various languages implies a difference in the inherent melody of each” (p. 7). Given the number of distinctive languages spoken in the world, it is not surprising that different linguistic communities reflect differences in their selection of acoustic properties in their singing, as reflected in the use of different musical scales by different cultures. I attempt to investigate the phenomenology of melody with reference to two broad linguistic communities, namely tonal language (Chinese) and nontonal language (English) speakers.
The Phenomenology of Melody

Chinese music uses a tuning system analogous to that used in the West, having made numerous attempts to solve the mathematical challenge of creating 12 semitones within an octave (Yang, 1985). Archeological and written records indicate that a complete chromatic scale and a pentatonic scale were used in China from at least 522 B.C. (Randel, 1986). In traditional Chinese music however, just tuning rather than tempered tuning was used. Also, since Chinese music is structurally pentatonic, deriving 12 semitones within an octave was mainly to facilitate transposition among the modal systems that were built upon the basic pentatonic scale. Nevertheless, because the Chinese and the West have very different concepts of harmony, their approaches of constructing melodies are distinctive.

In Western tonal music, the concept of implied harmony is central to a melody, which is either derived from or representative of harmony (Sadie, 1980). The emphasis on harmony is often found in the underlying harmonic structure of a melody, which “may serve to enhance the shape of a melody that they support; may stabilize, confirm, or prolong a key; or may bring about a change of key (termed modulation). Such purposes can be accomplished with many changes of harmony or only a few, with regular or irregular harmonic rhythm, or by means of a sequence” (Randel, 1986, p. 367). Western melody may therefore be perceived as “fundamentally inseparable from harmony, since melodies in this system clearly imply simultaneous combinations of sounds such as major or minor triads” (Randel, 1986, p. 482). This strong relationship between the perceived

8 Since the Zhou dynasty (1122-770 B.C.), Chinese used frequency ratios (circle of successive fifths) to derive pitches tuned according to a fundamental pitch handed down by the imperial court called huangzhong (黃鐘, "yellow bell"). Systematic mathematical approaches to equal tempered tuning dates back to He Chengtian (370-447), who derived Xin lu (新律, "new tone"). His calculation may be considered an early version of the tempered scale because all his 12 semitones deviate less than 10 cents from the equal tempered pitches. A successful calculation of the tempered scale was created by Zhu Zaiyu (1536-1610) sometime before 1595.
melodic and harmonic structure in Western music is so persistent that it is found in both trained and untrained listeners (see Butler, 1992, p. 122).

A study by Thompson and Cuddy (1987) demonstrated that listeners established as clearly a sense of key from melodies presented without harmonic accompaniment as they did when the melodies were presented in a four-part chorale setting. These findings suggest that Western listeners have learned to seek tonal references in melodies even without being given chords to convey harmonic messages. Butler contends this phenomenon to be “an important aspect of the sound of the Western tonal idiom” (p. 125), but the same is not evident in atonal music. From the earlier discussions on pitch processing, it might be inferred that the behaviour of seeking tonal reference by subtracting harmonic information from a melody might be an effect of acculturation.

Compared with the Western musical tradition, the ancient Chinese approached the concept of harmony from very different perspectives. Contrary to the importance of harmonic reference in Western melodies, Chinese melodies are more concerned with tonal movement from the perspective of a horizontal plane (i.e., the succession of one pitch moving to another over time), while treating harmony as being of secondary importance (Levis, 1963). Harmonized forms of Chinese melodies only appeared after World War I as a result of the educated elite’s attempts to integrate Western musical idioms into Chinese music.

Among these pioneer composers are Chao Yuen-ren (1892-1982), a renowned linguist and composer who advocates composing melody to reflect the Chinese linguistic tones and contour; Huang Tzu (1904-1938), a composer and music educator, who adopted impressionist harmony for his choral works; and Ma Ssu-ts’ung (b.1913), a pioneer composer, who added harmony to traditional Chinese folk tunes (Randel, 1986, p. 249).
has three meanings: (1) vocal composition; (2) vocal and instrumental performance; and (3) lyrics for a vocal composition (p. 114). This description calls into question the assumed universality of the phenomenology of melody. While Westerners understand lyric and melody to be two dichotic components in a song, the Chinese bond them together, and use the same word to signify both components. Yang cited an example found in *A General Biography of the Lay People* (逸民列傳) of *Hou Han Shu* (後漢書, “History of the later Han”, dated 947-950 A.D.) According to this source, Liang Hung (梁鴻) was said to have “created” or “made” (the Chinese word used here is zuo [作], which signifies “compose” in this context) *Wu Yi Zhi Ge* (五噫之歌 “Song of the Five Sighs”). However, Liang Hung was in fact, the librettist of the song (Yang, p. 114). It can be seen that ancient Chinese may sometimes attribute the authorship of a song to the librettist, while in comparison, librettist of Western vocal composition most often occupies a secondary role to that of the composer. For example, the young Schubert is famed for his musical setting of the *Erlkonig*, while an audience often gives secondary consideration to Goethe’s accomplishment for his vivid depiction in this narrative poem.

What makes the difference? The nature of the controversy lies in the musical element “tone” (or “sheng”, 音), which is used both in the Chinese language and Chinese melodies as *melodic movements* to provide the basis of directional tonal movement (Levis, p. 20). Tonal movement is of primary importance in Chinese melodies because it manifests a central philosophical thought, *chi* (氣). In the domain of song, *chi* is conducive to the movement and balance between the contour of the melody and the

---

11 *Chi* is used as the fundamental investigative tool for a broad range of inquiry including Chinese medicine, cosmology, and aesthetics. The literal translation for *chi* may look simple because it signifies air, breath, or energy. However, the term could also imply air in motion, transforming energy, and an animated force which transmits influence. In a Chinese philosophical context, *chi* therefore, concerns balancing the inner qualities of the mind, body, and spirit. The notion of *chi* as an aesthetic sentiment is found in the studies of Chinese poetry. Like music, it is also the pursuit of movement and balance of *chi* that leads the ancient Chinese scholars to develop a highly systematic scheme of phonetic elements in poetry.
linguistic tones of Chinese. Given that ancient Chinese has a more encompassing cosmic conception about harmony, the ideal state of perfect harmony between the inner self and the outer world is therefore consummated in a balance achieved through the tonal movement of melody and speech. Nevertheless, depending on genre and regional practice, varying degrees of interaction may be found between the linguistic tones of the lyrics and the Chinese melodies.  

(See Appendix C for a discussion of the effects of Chinese linguistic tones on some of its vocal genres.) Given the differences in the phenomenology of Chinese and Western melodies, would the singing of Chinese-speaking children also exhibit a stronger affiliation with their spoken language?

The Relationship Between Speaking and Singing

Despite on-going debates about the similarities and differences between music and language, some interesting relations between musical and language development are evidenced in infants' vocalizations and in the early singing of young children.

Musical Features in Infants' Vocalizations

Researchers have studied the presence of musical elements in parental vocalizations to preverbal infants and argue that the melody is the message (H. Papousek, 1996; M. Papousek, 1996; Fernald, 1994). Parents typically incorporate rich musical properties when interacting with infants: they speak with higher pitch, wider pitch range, longer pauses, at a slower rate, and use smooth, simple but highly modulated intonation contours (Fernald & Simon, 1984; Papousek, Papousek & Bornstein, 1985; Fernald, 1984; Stern, Spieker & MacKain, 1982; Stern, Spieker, MacKain, & Barnett, 1983). Maternal vocalizations are therefore sometimes regarded as melodies rather than suprasegmental or paralinguistic elements of speech sounds (Papousek & Papousek, 1981). The notion of melody as message is proposed because it conveys meaning (Fernald, 1994).

---

12 The interaction between the tone of the lyrics and the tones in a melody is also pointed out by List (1971). For this reason, he recommends adding a third dimension of linguistic tone to his pitch chart for classifying vocalizations of different linguistic groups.
Correspondingly, these melody-like prosodic features are found to be compatible with an infant's perceptual capacity and the musical features of nursery rhymes (Trehub, 1990; Trehub & Unyk, 1991).

Similar melody-like prosodic features are also exemplified in the vocalizations of young infants, where the boundaries between speaking and singing are hard to draw. Thus, a reciprocal effect in infants' vocalization has been observed—such infant vocalization triggers the caretakers to imitate the infant, which in turn provides essential acoustic information for language acquisition (Fernald, 1989, 1994). An infant's vocalization may then be seen as being mediated by auditory perception and her world experience. For these reasons, intonation and characteristic prosodic patterns in infants' vocalization are hypothesized to serve a prelinguistic function, that is, they provide communicative cues. The following model proposed by Fernald (1994) summarizes hypothesized developmental functions of intonation in speech to infants over the first year of life.

It is hypothesized that an infant is biologically predisposed to respond differently to certain prosodic characteristics of infant-directed speech. Thus, maternal vocalizations serve initially to elicit infant attention. According to Fernald's model, the "melodies" or prosodic patterns of maternal vocalizations then gradually "become increasingly effective in directing infant attention and modulating infant arousal and emotion" (1994, p. 65). Later, along with facial expressions, recurring prosodic patterns in specific affective contexts "come to function as the first regular sound-meaning correspondences for the infant" (1994, p. 65). It is hypothesized that toward the end of first year, the prosodies of mothers' speech begin to facilitate speech processing and comprehension because "prosodic marking of focused words helps the infant to identify linguistic units within the stream of speech" (1994, p. 65). Thus, Fernald proposes that "the exaggerated melodies
of mothers' speech are highly salient to young infants, influencing infant attention, arousal, emotion, and language comprehension” (Fernald, 1994, p. 66).

The Relationship of Speaking and Singing in Children’s Vocalizations

Western investigators concur that the initial stage of a child learning to sing requires using a singing voice and differentiating it from that used for speaking (Buckton, 1983; Welch, 1986). In the earliest stage of learning to sing, a young child’s vocalizations are found to be chant-like and remain within the range of the speaking voice. This may be explained by the fact that the same vocal mechanism is used for both speaking and singing. Evidence also suggests that language production and musical performance may be two integrally related modalities (Handel, 1989). I argue that for young children, speaking and singing are two vocal phenomena that occupy the same acoustic continuum. The transition from a “speaking” to a “singing mode” demonstrates that a child has acquired proficiency to use the vocal mechanism for singing and is able to differentiate these two distinct forms in her culture. While findings on children’s singing are largely based on criterion referenced assessment, a methodology for qualitative discrimination between speech and singing by young children remains elusive. Hence, qualitative data on when and how a child learns to use her singing voice remains scarce.

Some young children’s vocalizations are reported to be “in between” speaking and singing (Fujita, 1990; Minami & Umezawa, 1990). Fujita observed a kind of “intermediate performance” between speaking and singing when “the phrase or words which were uttered metrically, or melodically with certain fixed pitch” (p. 146). To differentiate young children’s vocal behaviour into either a “speech mode” or a “singing mode” is therefore an

---

13 Source-filter theory (Fant, 1960) is widely used to explain the sound production mechanism in human vocal utterances (Sundberg, 1987; Zemlin, 1988). In light of this theory, it is likely that during singing, a certain extent of filter functions performed for the phonation of words may have imposed timbral characteristic on the sung sound.

14 Evaluation of performance based on a set of predetermined criteria or standards, rather than in comparison to the performance of others.
ambitious task. Welch (1994) cautions investigators that "the distinction between speech and song, and our adult perception of such vocal acts, is often blurred by a relative linguistic and musical immaturity and by our own (adult) desire to make sense of the vocalizations" (p. 3). He recommended that investigators observe situations when a young child sings to maximize contextual understanding. Minami (1994) identified two such situations: the "energy conservation phenomenon" and "poor vocal shift", both of which may have impeded a young girl's capacity to demonstrate her singing competence. Given a child's relative linguistic and musical immaturity (which occasions misunderstanding of the child's intended vocal act by an adult), detailed documentation of scenarios when vocalizations occur is desirable. Such documentation could provide investigators a better understanding of the context of the vocal acts, and hence, a differentiation between the development of language and singing in young children.
CHAPTER TWO

RELATED STUDIES OF CHILDREN'S SPEECH, SONG, AND VOCALIZATIONS

This chapter reviews related studies of children's vocal development and provides information about the theoretical background and specific traditions from which the present study is launched. Whereas the Papousek and Papousek study (1981) provides a framework for a methodology to investigate young children's vocalizations, findings from Moog's (1976) study are used as a benchmark for qualitative analysis. A recent study by Chen-Hafteck (1996) is also reviewed for a discussion of the interesting perspectives on singing of tonal language speaking children.

The Papouseks Study (1981)

The Papouseks studied the emergence of musicality in preverbal communication of their daughter, Tanya, from birth to 1;4 (1981). The fieldwork methodology employed by the Papouseks included a variety of strategies which provided the extensive context desirable in ethnographic research. Vocalizations by Tanya were recorded with a portable tape recorder, with contextual descriptions and comments written by the investigators at the time of data collection. Four procedures provided data for the following analyses: musical transcription; phonetic transcription; spectrographic analyses; and analyses of intensity, temporal structures, and fundamental frequency. The investigators identified "sound groups—"cry, consonant-like sounds, vowel-like sounds, and syllable-like sounds—in Tanya's vocalizations. Each of these sound groups is described using the following criteria:

1. acoustic features (such as frequency, intonation, and duration);
2. facial and other motor activities that accompanied the sounds;
3. the situation or presence of stimulus in which the sounds were produced;
4. reaction from parents in response to the infant's vocalization;
5. spectral information of the sound illustrated in spectrograms; and
Comparisons between the initial observation and any developmental changes found at later stages of the study.

Besides classifying Tanya's vocalizations, the Papouseks also analyzed the musical characteristics of those vocalizations. The Papouseks explained that “[u]nlike pitch, the differences in musical quality of voice sounds are difficult to quantify” (p. 184). Hence, spectrographic illustrations were used to supplement their descriptions. Since their analyses were extensive, each aspect of their investigations is briefly summarized, and implications then discussed.

1. Pitch range

In Tanya's vocalizations, pitch range varied with age and type of vocalization. For instance, Tanya's cheerful squeals were at around 880.0 Hz; her spontaneous singing of melodic pattern at 11 months was reported to be between 349.2 and 698.5 Hz; while her single-word utterances at 1;2 were within the range of 220.0 and 440.0 Hz, and within the range 195.9 to 880.0 Hz when she sang her favorite song at 1;3 (p. 186). The relation between pitch range and the type of vocalization suggests that pitch may be used as a criterion to classify different vocal forms (Papousek and Papousek use the term “sound groups”).

2. Musical intervals

The Papouseks reported that unisons and seconds, as well as glides and “undulating” pitch contours dominated in all age levels during their study of Tanya. These characteristics were explained observing a consequence of a young infant's limited control of the vocal tract. In addition, some intervals such as unisons, seconds, minor thirds, and fourths were frequently found in Tanya's spontaneous songs. These findings were however, not attributed to any biological predisposition for singing those intervals, but rather to the fact that those intervals are commonly found in traditional nursery rhymes and are frequently sung in other cultures (p. 188).
3. Melodic contour

Papousek and Papousek gave a detailed account of the different characteristics observed in Tanya’s melodic contour at each age level. Their analyses (pp. 189-190) are summarized in Table 2.1. (A comprehensive account of this same period in a large population of normal infants is given in Vihman, 1996, chap. 5.)

Table 2.1

Development of Melodic Contour in Tanya’s Vocalizations

<table>
<thead>
<tr>
<th>Age</th>
<th>Characteristics of Melodic Contour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neonatal</td>
<td>Distinct rising-falling contours</td>
</tr>
<tr>
<td>0;2-0;3</td>
<td>Vowel-like sounds contained falling terminal glides with voice fading. Quite regular fundamental voicing, with sequences of repeated sounds consisted of melodic and rhythmic structures. Later, rising contours emerged together with squealing sounds of steep glides over an octave.</td>
</tr>
<tr>
<td>0;4</td>
<td>Quiet melodious patterns were heard during vocal play. They may be in rising-falling contour covering more than 2 octaves. These patterns later became smoother and more continuous.</td>
</tr>
<tr>
<td>0;7</td>
<td>Pitches were sung in discrete steps. Hummed first original short melodies made up of sequences of musical tones.</td>
</tr>
<tr>
<td>After 0;7</td>
<td>“Reduplicated babbling” (p. 190)--single syllables such as “da da da da”--sung to a variety of melodic patterns (0;11).</td>
</tr>
<tr>
<td></td>
<td>Further development in two parallel lines--“the speech-like intonation contours accompanying the linguistic differentiation, and the musical melodic contour in imitated, or invented songs” (p.190).</td>
</tr>
</tbody>
</table>

Based on Papousek and Papousek (1981).

Analyses of Tanya’s melodic contour demonstrated how she advanced from exploring her own voice (0;2-0;3), to the babbling of discrete speech and song sequence (around 0;7). This finding implies that an infant begins to engage in a rudimentary form of speaking and singing within the first year after birth, which for speaking, is in accord with the large body of extant data (Vihman, 1996).
4. Imitation and learning to sing a song

To analyze this aspect of musical development along with evidence of cognitive operations, contextual information such as Tanya's facial expression, her emotional expressions, and the development of her linguistic abilities were documented. (See Table 2.2.)

Table 2.2
Tanya's Vocal Imitation and Learning to Sing a Song

<table>
<thead>
<tr>
<th>Age</th>
<th>Characteristics of Vocalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>0;2</td>
<td>First signs of attempting to match a pitch or short intonation contour in babytalk.</td>
</tr>
<tr>
<td>0;3</td>
<td>Tanya matched single tone stimuli (sung or played on a piano).</td>
</tr>
<tr>
<td>0;4</td>
<td>Turn-taking imitation of long utterances in a global mode.</td>
</tr>
<tr>
<td>0;8</td>
<td>Tanya imitated “ta-ti”, the family name of her father, with dramatic manipulation of musical elements (i.e., exaggerated intensity, intonation contour, and varied duration of the syllables).</td>
</tr>
<tr>
<td>0;10</td>
<td>Proto-words and speech-like intonation contour correctly imitated. Tanya also imitated tonal sequences and short melodies correctly.</td>
</tr>
<tr>
<td>0;11</td>
<td>The first phrases of songs were correctly hummed. Often, the lyrics and temporal aspects of the song were correctly imitated, but the more difficult passages were omitted. Tanya enjoyed listening to new songs.</td>
</tr>
<tr>
<td>1;1</td>
<td>Tanya appeared to enjoy learning new versions of lyrics in her favorite song. She was able to complement the parent’s singing, or at times echo part of the song. The Papouseks therefore concluded that “Tanya had the entire concept of the song at her disposal” (p. 196).</td>
</tr>
<tr>
<td>1;3</td>
<td>Tanya was able to sing a lullaby melody together with her mother. She later incorporated the song into her monologues.</td>
</tr>
<tr>
<td>1;4</td>
<td>Tanya sang fragments of songs during spontaneous play.</td>
</tr>
</tbody>
</table>

Based on Papousek and Papousek (1981).
Some interesting observations were made when Tanya was 1;3:

- When Tanya sang a song whose lyrics were beyond her linguistic competence at that time, she managed to hum the song or borrowed corresponding words from a different strophe;
- Tanya was able to freely transpose the melody into different keys;
- When Tanya detected herself singing a wrong note, she would either return to the initial start up key or use the wrong note as a new starting point to reestablish the tonality of the melody;
- When singing in a sad mood, Tanya tended to flatten the key; whereas when her mood improved as she sang, she restored the melody back to the original, higher key;
- Phonation of the lyrics during singing may have affected Tanya’s singing accuracy. Her singing was often found to be more accurate when she hummed melodies alone;
- Some inaccuracy in Tanya’s singing was probably due to a confusion of the structural design of the song (mixing up similar phrases) or ambiguity inherent in the melodic interval; and
- While maintaining the correct strong-weak pulse of the nursery song, she sometimes adapted typical babytalk rhythm into her rendition.

5. Development of speech-like intonation contours

The Papouseks maintain that the prosodic patterns in infant vocalizations have a crucial communicative function:

The development of speech-like contours parallels that of melodic contours in singing, and cannot be distinguished from it convincingly prior to the infant’s transition from babbling to the first lexical items and one-word sentences. However, the speech-like contours must not necessarily be restricted to the lexical content or social context of vocalization. (p. 199)
In the case of 7-month-old Tanya, a melodic high pitch rising contour was an invitation to mutual play; a horizontal contour with vibrations was a nagging request; and a slowly and fluently falling contour was a soothing utterance. Until 16 months, Tanya’s vocalizations still embodied exaggerated pitch contour of speech and songs.

Implications of the Papouseks’ Study

The Papouseks’ study corroborates research on the vocalizations of young children (see Vihman, 1996 for an extensive review); their contribution to the literature is in respect to singing. Their analytical strategies include musical transcription, spectrographic and digital acoustic analyses of fundamental frequency, amplitude time, and harmonic spectrum. These strategies offer a range of analyses of the acoustic features of vocalizations, and allow the reader to visually scan the results of the analyses represented both numerically and graphically. Furthermore, the Papouseks also provided qualitative description such as Tanya’s cognitive, linguistic, and emotional aspects to enrich the content and context of the vocalization.

The analyses of prosodic features as musical elements in the Papouseks’ study strongly support research on early childhood vocal development. As the Papouseks put it, “musical elements do not develop as a solitary phenomenon, but as a part of global behaviors in the context of social interaction, and, therefore, may be viewed from different aspects” (p. 216). Thus, singing could not be reliably analyzed in isolation from speaking. M. Papousek (1996) contends that “the preverbal origins of musical skills cannot easily be differentiated from the prelinguistic stages of speech acquisition and from the basic alphabet of emotional communication” (p. 92-93). Hence, in the context of an infant/child learning to sing, it is imperative to study the vocalizations of young children in the context where the vocal communications occur.
The Moog Study (1976)

A longitudinal study conducted by Helmut Moog on musical development in children aged 0;6 to 5;5 was ambitious in scale. Moog administered over 8,000 individual tests using acoustic stimuli to nearly 500 children, and evaluated observations of about 1,000 parents. The same set of recorded acoustic stimuli was used for all children regardless of age. The resultant responses were then evaluated within the context of general development of the child. The yearly vocal responses reported by Moog are summarized here to serve as a benchmark for the current investigation.

1.0 Age 0;6 to 0;11

Moog labelled the earliest forms of vocal responses as vocalization and musical babbling. For Moog, vocalization is the vocal sound expressing pleasure upon hearing a piece of music; musical babbling includes song-like vocal sounds that are varied in pitch produced on one or a few syllables (p. 60). Moog differentiated speech from musical babbling: “speech babbling is produced in the six or seven month old infant by talking to him; musical babbling only occurs if music is sung or played to him” (p. 60). It is interesting that Moog explicitly discriminated between the different types of stimuli that initiated vocal responses. Some crucial findings reported for this age level:

- vocalization showed no reference to a diatonic system nor any rhythmic regularity;
- more descending than ascending melodic contours were found;
- although the pitch range could be more than an octave in some children, the babbling monologues often moved in microintervals (a quarter of a tone);
- “speech babbling begins before musical babbling, but the child sings his earliest songs before he can say his first word” (p. 62);
- vocalization to music began at around 6 months, which was about 2 to 4 months following the stage at which speech babbling began; and
• children’s vocal responses bore no musical or linguistic resemblance to the stimuli that were played to them.

2.0 Age 1;0 to 1;11

A marked advancement in singing was reported for all children in this study between ages 1;0 to 1;11. Moog identified two types of singing at this age level. Babbling song: a type of spontaneous singing found in all the children by age two, and imitative singing: a phenomenon just beginning to emerge in some of the children at this age.

2.1 In babbling songs:
• children managed to sing longer songs and establish habits of singing regularly in their daily routine;
• movements sometimes accompanied their singing, but words were often omitted. When words were used, they typically appeared in single words, parts of words, or in a string of nonsense syllables;
• children began to establish a clearer sense of the diatonic system at the end of the second year;
• the rhythm of babbling songs was simple. Most were made up of notes in equal duration, some of them employed notes of two lengths, of which the longer note was approximately twice the length of the shorter note; and
• no regularity was found in the length of the rest and the distribution of accent.

2.2 In imitative singing
at 1;0:
• only the sounds of some words (such as the vowels) but not the rhythm and pitch were imitated;
• children often selected words with a distinctive sound pattern and repeated them for babbling; and
• the imitated song was usually fragmented.
at 1;6:
• children’s imitative singing followed the learning sequence of matching the words, the rhythm, the pitch, then extending the length of the imitated fragment to a phrase, and to the entire song.

3.0 Age 2;0 to 2;11

A marked increase in the amount of singing was reported for this age level, the songs were also reported to be much longer. Two types of singing were found: spontaneous songs were melodies that could not be traced back to something that had been presented to the children; and songs that imitated a model.

3.1 Spontaneous songs:
• the lengths of these songs varied but a child at this age could hum or sing with one vowel or syllable for about 4 minutes;
• these songs gave an impression of a narrative, whereby a child sang to herself with text that made little sense to others;
• spontaneous songs were characterized by single notes organized with some loose reference to structural phrasing, and separated by short pauses for breath; and
• simple rhythmic patterns were used repetitively throughout the singing.

3.2 Songs that imitated a sung model:
• all children at this age were able to imitate at least a part of the song previously presented to them. Imitation of pitches was still not very accurate, but there was a good sense of melodic contour; and
• Moog used the terms “word singer” and “word and rhythm singer” respectively, to label the children who could imitate the words in a song, and those who could imitate the word and the rhythm of the song.
3.3 “Pot-pourri” songs

A type of half-learned and half-improvised singing emerged. Moog described it as follows:

[In this mixed form the child begins to sing spontaneously but lets fragments of the words or melodies of songs he has learnt come into his singing as he goes along. So bits of the words of a learnt song may be mixed with an “original melody”. This must not be put down to inability to imitate: it is simply that the child feels like putting into his own song bits of the tunes he has learnt--making up his own words for them. He is not in the least concerned with singing songs he knows already but is including his existing musical experience in his vocal improvisation. (p. 100)

4.0 Age 3;0 to 3;11

Spontaneous singing dominated the musical activities of children at this age. Moog identified three different forms-- “imaginative”, “narrative”, and “pot-pourri” songs.

*Imaginative songs* form only a small repertoire of *spontaneous songs*. They bore no resemblance to a known song, and were often hummed or sung to a single syllable.

*Narrative songs* seemed not to be intended for an audience, although they often told stories. They were sung monologues composed mainly of nonsense snatches of words and tunes, and occasionally included excerpts of learned songs. *Pot-pourri songs* were combinations of *learned songs* with mixed up words and melodies, and they might contain original improvisations.

Children were observed to have adapted the same formal design from their *learned songs* into songs that they created. Interestingly, for children who did not demonstrate singing any learned songs in the study, no original song was observed. Moog also observed correlations between home musical environment and children’s song repertoire--children who were reported to have a large repertoire always had parents who sang often to them. Children were reported to particularly enjoy singing games.
5.0 Age 4;0 to 6;0

A marked decrease in singing response was reported in the four- and five-year-old children but a rise occurred in the six-year-olds. Moog accounted for these findings in a number of ways: four- and five-year old children had a more restricted repertoire because they only responded to songs they had already learned. They tended to be more conscious of their singing performance, and might be more inhibited by the presence of strangers.

Half of the four- and five-year-old children were able to sing songs with several verses correctly, but only 15% could sing learned songs in tune. Six-year-old children sang more imaginative songs but fewer “pot-pourri” songs. They often made up words and tunes even though they appeared to be able to correctly reproduce a learned song. Their improvised singing were often found before or after the learned song as an introduction or a coda to the learned song itself. It was also observed that when children altered a song, they could not pick up the original song without making a few more mistakes before they resumed singing the proper song.

In summary, Moog reported that pitch accuracy seemed to be the most obvious problem for children attempting to imitate singing. Occasional problems using an unsuitable tempo were also reported, but all other mistakes made in rhythm are related to the problem of singing out of tune. The fewest mistakes were found in imitating the words of a song.

Implications of the Moog Study

The issue of first spoken language

Moog’s study was administered to German speaking children using German songs, so the generalizability of results to tonal speaking children is uncertain. Based on his discussion of the differences between European and Asian languages (pp. 39-41), Moog states that “[m]usic and language are man made and are peculiar to man and his fellow men” (p. 39).
Issues concerning the criterion tasks

Moog’s use of recorded musical stimuli in a carefully planned sequence served to control a number of variables. However, administrating the same sequence of musical stimuli over several years of a longitudinal study could have induced test-retest effects.

Moog developed a number of labels to classify the different types of emerging vocal behaviours observed. Terms such as “musical babbling”, “pot-pourri songs”, and “imaginative songs” were used freely to label the behaviour he observed. Moog had been careful to avoid undefined labels by providing clear description and examples for the type of behaviour the label was intended to capture. The use of such labels assists the reader in clarifying the subtleties in the nature of the vocal responses. The present study borrows some of the singing labels from Moog in order to facilitate comparisons.

Moog also reported that children often mastered the words of a song before they mastered rhythm and the pitch. Although other investigators (Davidson, McKernon, & Gardner, 1981) have accounted for this phenomenon, Moog argued that it is the richness of the acoustical characteristics of speech which attracts a child’s attention. This argument is worth examining, especially since Moog reported evidence demonstrating that children at this age already possessed the ability to differentiate between musical and linguistic discourses (p. 85). This evidence appears to suggest that although children age one and two are able to differentiate the roles played by speech and song, an explicit discrimination between the acoustical features of these vocal phenomena may be redundant. Following this line of argument, we can hypothesize that in a given song, children may be attracted to different aspects of the words or musical elements according to whichever provides them with maximum stimulation--this hypothesis is explored in the present investigation.
The Chen-Hafteck Study (1996)

Although the specific research questions addressed in Chen-Hafteck’s (C-H) study are quite different from those explored in the current study, both studies address the relationship between speaking and singing, and their effects on preschool children’s melodic singing performance. The C-H study examined the effects of text-melody relationships in Cantonese children’s songs on the singing performance of children living in Hong Kong.

The C-H study was guided by two propositions: first, that music and language are processed in an integrated manner; and second, that a causal relationship exists between text-melody relation and melodic singing accuracy in children (p.70). The C-H study was constructed as follows: three songs were taught to kindergarten children in two quasi-experiments—12 children participated in a pilot investigation, and 192 children participated in the main experiment. Three criterion songs were taught to the children:

(a) a “tone-mismatched” song in which melodic contour mismatched the pattern of the linguistic tones in the text;

(b) a “tone-matched” song where the melodic contour matched the pattern of the linguistic tones in the text; and

(c) a traditional Cantonese nursery rhyme, which was to be spoken rather than sung.

No statistically significant differences in the overall pitch accuracy were found between the performance of the tone-matched and tone-mismatched song. Some differences were reported however, in the singing of individual musical intervals—the perfect fourth and major sixth intervals were sung less accurately by the Cantonese-speaking children when compared to English-speaking children.

Previous research has not addressed the possible challenge presented to tonal speaking children when they are learning to sing. Selected pertinent issues in the C-H
study are therefore examined, and contrasted with related literature and some personal perspectives.

1. Issues of linguistic tone and its possible effects on sung melodies

According to the Chen-Hafteck hypothesis, tone-mismatched songs require children to focus their attention because the “musical intonation deviates from linguistic tonal models” (p. 71). C-H appears to suggest that Cantonese linguistic tones could serve appropriately as a model for the organization of musical pitches in a song. Hence, melodic pattern that fails to comply with linguistic tones is considered a “deviation” from a target model. In both her experiments, however, C-H found that “the effects of singing tone-matched and tone-mismatched songs on children’s pitch accuracy in singing are shown to be limited” (p. 160). These findings may illustrate some false assumptions about the effects of linguistic tones on sung melodies.

It is debatable whether Cantonese linguistic tones might serve as an appropriate model for the organization of musical pitches in a song. A simple organization of discrete tones in music and in speech is neither as meaningful nor as aesthetically satisfactory as is perceived tonal movement. As Levis (1963) has suggested, “[t]he real nature of these musical elements in the Chinese language both in linguistic and musical usage is that not of tone but of melodic movement.” (p. 21). This suggestion implies that there should be a focus on a holistic sense of melodic movement in a musical phrase and in a spoken sentence, rather than on the different degrees of high and low pitch spoken or sung to individual phonemes. We may therefore infer that an excessive focus on the manipulation of discrete tones (both linguistic and musical) could hinder appreciation of the overall musical flow in speech and in music. This important issue will be addressed in greater detail below.

Levis advocates the “inherent musical potentialities” in the tonal movement of the Chinese language:
The elements of tonal movement are in speech for the most part merely gliding movements of the voice or a "continual gliding change in pitch" as Barton puts it in his *Textbook on Sound* (latest edition, 1932). It is in song that these elements of melodic movement assume their great significance. For this reason it is possible for one who is tone deaf to be able to pronounce the so-called "tones" in Chinese correctly. And this is because in speech or in recited poems no exact pitch sense is necessary but only a perception of tonal movement. (p. 18)

Although both C-H and Levis advocate the observation of linguistic tones in sung melodies, their conceptual understandings of the function of pitch in speech and song are not compatible. Levis differentiates between the ability to master linguistic tones and the ability to master musical pitch; C-H assumes that linguistic tones should ideally be used as a model for "musical intonation".

To delineate the explicit nature of linguistic tone and its possible effects on sung melodies, I propose to integrate Levis’s analogy of a so called “tone deaf” person (someone who consistently sings out-of-tune) into the C-H model of learning to sing “tone-matched” songs. Hence, hypothetically speaking, two different results might be derived from the same scenario because the conceptual understandings of Levis and C-H are incompatible. This incompatibility is illustrated as follows.

According to the C-H model, a person with a weak sense of musical pitch but a good sense of linguistic tones would be more likely to sing in tune with the help of a tone-matched song. C-H suggested that when the tones of the text are diligently matched with the organization of musical tones in a song, then a speaker-singer would benefit from the linguistic tones, as reference markers for singing the melody in tune.

According to the Levis model, although the “tone deaf” person may speak with correct linguistic tones, s/he could still fail to sing a tone-matched song in tune because of his/her lack of a sense of precise musical pitch. In this case, the transferability of skill in
producing linguistic tones to musical pitch is doubtful. Both models remain to be supported with empirical data.

In the case of nontonal language speakers, tonal movement in their speech must be explained in a broader context, because the overall contour in a spoken sentence has bearing on its phonological load. This observation gives rise to several questions. Is it possible to match the contour of nontonal speech with the melodic contour in a song? To what extent could we speak also of a contour-matched and contour-mismatched dichotomy in nontonal language? Since very little effort has been made to match the melodic contour to the linguistic contour of a nontonal language, could we then speculate that it is, therefore, much more difficult to sing in tune with a nontonal language? Does the Cantonese language help children to sing more accurately, as suggested by C-H?

The problems posed in seeking to answer these questions lead us back to the core question: what is the function of pitch in speaking and in singing? The Levis model seems to be more advanced in its identification of two types of pitch perception and production—(a) the relatively subtle (nondiatonic and microtonal), language-bounded ability of discrimination between different types of linguistic tones, and the production of such fine distinctions in speech; and (b) an acute sense of musical pitch relevant to the musical practice (or more specifically, the scalar structure) of the culture as observed through singing performance. The relations between speaking and singing are complicated because pitch is not a sole consideration in singing. Many other cognitive operations such as emotional and kinesthetic processes are involved. So, to what extent can the relations between the processes be limited, and how important is the totality of auditory experience necessary to singing? The questions posed in the foregoing discussion are now addressed.
2. Issues of the integration hypothesis on speaking and singing, and its possible effects on learning songs.

Some studies on the recall of melody use an “integrated whole”\(^\text{15}\) to explain how text and melody may be stored, and processed, in a melody recall task (Serafine, Crowder, & Repp, 1984; Serafine, Davidson, Crowder, & Repp, 1986). Using this hypothesis, C-H explained the effects of tone-matched and tone-mismatched songs on children’s singing performance. Although she argued that “recognition ability does not imply ability to reproduce but reproduction can imply recognition” (p. 177), she also pointed out the problem of applying the “integrated whole” hypothesis to her study:

Studying children’s singing involves more complicated processes than song recognition. In this way, investigating whether texts and melodies are integrated in children’s cognitive processes by analyzing singing may pose problems of validity in such research. (p. 177)

Although C-H contended that text and melody are integrated in children’s song-learning, research suggests that text and melody are acquired at different rates and at different stages of development during childhood. In particular, there is a gap between the chronological age for mastering linguistic tones and learning to sing in tune. For tonal speakers, linguistic tones are acquired before segmental phonology. Evidence suggests that most Cantonese-speaking children complete acquisition of all vowels and tonal contrasts by age two (So & Dodd, 1995; J.K.-P. Tse, 1978; S. M. Tse, 1982). On the other hand, reports suggest that children often begin to sing in tune between ages five to six (Welch, 1986). It can therefore be seen that Cantonese linguistic tones are mastered much earlier in childhood while singing skills defined as acceptable by the Western musical and scientific traditions are acquired comparatively later in life. Based on the research literature, a Cantonese-speaking child with normal cognitive abilities could be expected to speak with

\(^{15}\) Also - the integration hypothesis.
correct linguistic tones around age two, but could not be expected to sing accurately in tune until about age five. That is not to say that linguistic tones would always have dominance over musical pitch in the process of learning songs; it suggests that the acquisition of linguistic tones and musical intervals seem to be two separate skills that mature at different ages.

The problem concerning the acquisition of linguistic tones and musical intervals is multifaceted: Levis's example of a so called "tone deaf" person who could not sing in tune but could speak with accurate linguistic tones is a good illustration. Another aspect of the problem has to do with auditory processing versus vocal production, as illustrated by reports of musicians who develop a sophisticated concept of musical pitch but cannot sing reliably in tune. Thus, the explicit nature of the problem does not seem to lie simply in whether text and melody are integrated in the processing of melody recall or not; or whether text dominates melody, or vice versa, during the process of learning a song.

The integration hypothesis could be useful as a tool for investigating possible cognitive processes involved in melody recall. How other aspects of speaking and singing may be related, however, remains uncertain. What is also needed to understand the problem is further investigation of the relationships of vocal production in speech and song, and how speaking and singing are used in different communication contexts.

3. Issues associated with correlating pitch movement in speech to pitch movement in melody

The terms "tone-matched" song and "tone-mismatched" song used in the C-H study deserve further explanation. C-H defined them as follows:

*Tone-mismatched melody*, in which the melodic contour does not match the pattern of linguistic tones in the text. In other words, when the pitch pattern of the tones in the text ascends, the melodic direction descends, and vice versa.
Tone-matched melody, in which the melodic contour closely matches the pattern of linguistic tones in the text. In other words, when the pitch pattern of the tones in the text ascends, the melodic direction also ascends, and vice versa. (p. 50)

In her analyses of the relationship between linguistic tones and musical intervals, C-H also alerted us to the relations between pitch movement in Cantonese speech and in sung melody:

[M]usical intervals implied by the movement of the Cantonese tones are relative in nature. In other words, the tones only imply a larger or smaller upward or downward movement in pitch, and the degree of such movement may vary as long as the relative relationship among the various tones is maintained. The closer the linguistic tones, the less there exists the possibilities of variations in the degree of pitch movement occurring; and the larger the linguistic tonal movement, the more variation it allows. (p. 102)

It can be seen that the nature of a relation between linguistic tones and pitch in a song could only be expressed in an implied perception of a parallel between linguistic tonal movement and the general directional contour of a sung melody.16 Following this line of argument, an attempt to equate the function of pitch in speech to the function of pitch in song could be due to some misconception. This may be illustrated with a hypothetical question such as “Is the English language diatonic?” One may ask why English and other

---

16 Levis (1963) provides a more detailed analysis of the implied relationship. In his book, he uses the term “tone” for musical pitch, and “neumes” for linguistic tone. The following basic principles of correlating pitch movement in speech to pitch movement in melody are proposed by Levis:

(a) The neumes are merely forms of tonal movements, which, when used for music, define the basic elements of melodic line.
(b) They are not limited to any definite pitch or key.
(c) They are not limited to any interval.
(d) They are in principle made of single units, either of melodic movement or rhythm (quantitative or dynamic or both) and maintain a unity between melodic and rhythmic elements. (pp. 39-40)
European languages are not diatonic, but that Western European music is built on a diatonic system? Such a question may also demonstrate a confusion over a number of pitch concepts (such as musical pitch, spectral pitch, pitch contour of the fundamental frequency, fixed pitch, relative pitch, and tonal movement). Although most traditional English melodies are diatonic, it would be inaccurate to infer that spoken English is also diatonic.

Misconceptions concerning pitch concepts may be illustrated further. C-H compared the singing performance between English- and Cantonese-speaking children. Chen-Hafteck’s rationale for including a group of English-speaking children in her study may be seen as a challenge to construct validity:

Diatonic melody is Western and it was argued that it matches English texts better than Cantonese texts. Thus, the second hypothesis is that English diatonic songs will be sung more accurately than those with Cantonese texts. (p. 71)

As a result of such a challenge to construct validity, C-H reported the following: The results turned out to be contradictory to what had been hypothesized. The Cantonese-speaking children performed significantly higher in pitch accuracy than the English-speaking children. (p. 170)

This finding is further illustrated in her analyses of pitch accuracy on individual musical intervals:

The hypothesis that certain intervals will be sung more accurately in Cantonese text and others more accurately in English cannot be upheld. This is because a higher accuracy of Cantonese children over English children was shown in all intervals studied. (p. 171)

That her hypotheses were not supported could have been due to some misconceptions concerning pitch concepts and uncontrolled experimental variables, which are illustrated in the following arguments:
According to Levis (1963), the significance of musical qualities in both speech and language lies in *tonal movement* rather than discrete level of fundamental frequencies of individual tones. Hence, it is the perceptual similarity of the tonal movement in speech and in the melody of a song that may be seen as complementary, not the individual musical intervals and the variation of pitch levels among phonemes. This accuracy in singing musical intervals cannot be directly attributed to the language spoken.

To illustrate my second argument, I shall use the hypothetical question “Is the English language diatonic?” If the answer to that question is negative because English is not organized according to a diatonic scalar structure, then singing a diatonic song in English would not be of particular help. Similarly, singing a diatonic song in Cantonese is not particularly problematic, at least for pitch accuracy. It is debatable whether a language could be found that would be accepted as diatonic in the Western musical sense.

Whether the Cantonese- and the English-speaking children in the C-H study came from two comparable samples of the Hong Kong population is uncertain. C-H did not establish that her two experimental groups of children were comparable at the outset of the study, and no attempt was made to control variables such as socioeconomic status and type of schooling. In her discussion of research findings, C-H reports that Cantonese- and English-speaking children had very different attention spans and “learning attitudes” (pp. 157-158). She concludes that these differences “reflect the differences in educational approach at children’s schools and home” (p. 158). Since her subjects were drawn from an “international” kindergarten, it is likely that their families were quite different from those who attended local kindergartens for Cantonese-speaking children.

4. Issues concerning speech melody perception

The differences between musical pitch in a melody and the function of pitch in linguistic tone and speech contour, deserve careful scrutiny. The following discussion is guided by three aspects related to speech melody:
(a) sensitivity threshold required in the perception of speech melody;
(b) requirements of scalar structure and pitch stability in the perception of melodies; and
(c) phonological loading of linguistic tones in tonal languages.

Psychoacoustically speaking, perception of the prosody of speech is often related to the perception of melodies, especially in the absence of semantic consequences (e.g., hearing a conversation in a foreign language). From a physical perspective, melodic patterning of speech will depend on how fundamental frequency changes in the speech flow might be perceived as musical pitch in a melody. Some Russian linguists, such as Baryshev (1989), cautioned against attempts to equate all elements of speech to musical events:

Despite the musical quality of human speech, human perception of its prosodic properties is conditioned by a series of factors entirely uncharacteristic of music. Thus musical sounds are characterized by periodic vibrations and represented in a system of discrete units, while speech sounds are represented by quasi-periodic vibrations whose frequency does not change discretely, and also by aperiodic vibrations. It is precisely the indiscreteness of changes in the fundamental pitch of vowels that constitutes the main difficulty in correlating speech melody with discrete units of the musical notation system. (p. 87)

Not only is it erroneous to strictly correlate the “melody” of speech to the melody in a song, but Baryshev (1989) also demonstrates that even for musically trained ears, speech melody perception is a challenging task. Several of his findings (pp. 94-95) are pertinent to C-H’s investigation. Thus, Baryshev showed that:

- Perception of speech melody is different from the perception of melody in music;
- Speech melody constitutes a sequence of randomly changed frequencies linked to amplitude modulation; different segments of speech are perceived with different degrees of accuracy; and
• The accuracy of the perception of melodic intervals diminishes, on the one hand, in the transition from musical to speech intervals, and on the other, within speech intervals in the transition from suprasegmental speech intervals to segmental ones.

According to Baryshev (1989), a “musical ear” that can discriminate musical intervals of half a semitone, performs poorly when discriminating speech melodic intervals. For suprasegmental melodic intervals, the “musical ear” could only discriminate to an accuracy of a semitone in 38 percent of the trials; for segmental melodic intervals, only one-third were discriminated (p. 95). Indeed, extraction of musical pitch from speech may require explicit training. In a discussion of the relation between linguistic tones and music learning, Wang (1991) explains that tones can be decomposed into phonological features, and like consonants and vowels, they are perceived in linguistic categories (Wang 1976). This parallels C-H’s comment that “fitting words to music while respecting the pitch relationship between text and melody, is unconscious among the lyric-writers or composers” (p. 53). What she described as “unconscious” is that speech melody perception cannot be made as explicit as the discrimination of discrete tones in the melody of a song, even though native speakers of a tonal language are capable of discriminating the fine variations at the tonal level for semantic consequences. Hence, threshold of sensitivity required for the perception of speech melody may make it difficult for explicit adaptation to the melody of a song.

In order for some combination of fundamental frequencies to be perceived as melody, pitch stability and some complexity of scalar structure are crucial. Scalar structure is the essential building block in constructing melodies and in order to extract scalar structure from a speech sequence, a reasonable level of stability in pitch is needed. On the other hand, in a complex system of intonational usage in speech, such as in the case of a tonal language, scalar structure might be negated because variations in fundamental frequencies of speech have semantic consequences. Hence, tonal contrasts in speech might
affect melody in speech. This argument may be further supported by Ross, Edmondson, and Seibert (1986):

*F₀* Variation is basically a measure of the bandwidth of the intonational contour. If affect causes the bandwidth to broaden or narrow, then significant magnitude changes between local pitch constituents of the intonational contour may occur. In non-tone languages such alterations will not disturb linguistic information. However, in tone languages, in which the relative height between and among tones may be contrastive, such alterations could disrupt linguistic information. (p. 298)

It can therefore be seen that a requirement for preserving the integrity of local pitch relationships in a tonal language is in conflict with the requirement of exhibiting complex scalar structure and pitch stability in a melody. Hence, to produce a speech melody that may resemble a musical melody in the Western tradition, speech intonation accuracy, especially of linguistic tone, may be undermined.

This brings us to the effects of phonological loading of linguistic tones on speech melody. Ross et al. (1986) reported that tonal language is inflexible when it comes to varying *F₀* for signaling affect. They explained that the inflexibility may partly be related to linguistic demands that pitch contrasts in tonal languages be very precise (p. 300). It is therefore, reasonable to infer that tonal languages may also be inflexible in varying the *F₀* to match the exact pitches in a melody. In such a case, to observe pitch accuracy of the melody, certain words may have to be sung incorrectly, that is, without observing their corresponding linguistic tones. Thus observation leads to the question: will the mispronounced linguistic tone bother a young child in the way it may annoy an adult?

For adults, linguistic tone is crucial for conveying the meaning of the song text. If adults choose to maintain pitch accuracy in singing a melody, they must be ready to bear
with the distorted meaning conveyed by the song text being sung.\textsuperscript{17} When confronted with such linguistic and musical choice, it may be difficult to assess what young children would do and whether they are capable of demonstrating their choice of action. Two things could happen in children's vocalizations. First, according to Moog (1976), the symbolic function of speech is less attractive to young children compared to speech sounds, and therefore they began imitating the words of a song they do not understand. This argument then implies that children would be more likely to focus on the acoustical and musical qualities of a melody, and hence, would disregard any mispronounced linguistic tones since the meaning of the text does not appeal to children as much as it does for adults. Second, a further conflicting perspective comes from studies on the acquisition of phonology by Cantonese-speaking children. So and Dodd (1995) reported that tonal speaking children acquired tones comparatively earlier than English-speaking children acquired phonemes because this aspect of Cantonese phonology makes important contributions to the burden of carrying meaning (p. 490). Also, linguistic tones were reported to be mastered by age two, which is much earlier than the ability to sing accurately in tune. Based on these findings, one may suspect that young children, between ages two to five, may pay more attention to the linguistic tones to maintain the meaning of the song text when they sing. On the other hand, children may not be able to maintain pitch accuracy because of musical immaturity but they already possess the linguistic maturity to handle the linguistic tones in the song text.

**Implications of the Chen-Hafteck Study**

Chen-Hafteck found no significant differences between children's performance on tone-matched and tone-mismatched songs. She accounted for this by suggesting that in

\textsuperscript{17} Chan (1987) studied the interaction of linguistic tone and melody in contemporary Cantonese and Mandarin songs. She reported that Mandarin melodies tend to dominate over the lexical tones, while Cantonese songs tend to preserve the contour and pitch height of the lexical tones.
Hong Kong, children are acculturated to sing diatonic songs; therefore, the mismatched linguistic tones of Cantonese in children’s songs did not appear to have hindered her subjects’ singing accuracy. These findings support Levis’s (1963) differentiation between two separate types of pitch concepts—tonal movement in speech prosody versus musical pitch.

In earlier discussion, I speculated that a person without an established sense of musical pitch would only derive minimal help from singing a tone-matched song. Similarly, a tone-mismatched song may provide only minimal distraction to a child singing diatonic songs if the child has already mastered the concept of Western musical pitch and the rules for Western scalar structures. This is speculation is in accord with C-H, who stated that “[t]he tunes of Western nursery rhymes have become popular for Hong Kong children and diatonic melodies are imprinted on their minds” (p. 165). It should be noted, however, that this is speculation and the notion of “imprinting” diatonic melodies is one that at this time lacks empirical support.

One final question: Is it possible to find a “pure” example of a tone-matched song in the contemporary Chinese repertoire? In considering the interactions between the intricate functions of pitch in a song, the idea of a “tone-matched song” may be elusive. Citing a number of composers and their works, Chao (1956) commented that most contemporary Chinese song writers pay no attention to tone, and that “most of contemporary Chinese song composition is independent of tone and only slightly influenced by the composer’s dialectal and literary background” (p. 58). Even as a linguist and an advocate of observing Chinese linguistic tones in vocal composition, Chao admitted that not all of his songs are consistently well matched for linguistic tones and melodic contour. For example, for one of his songs, he commented that it was “frankly foreign in style and paid no attention to tone” (footnotes, p. 58). In spite of his linguistic training and musical background, Chao admitted that “[i]t is not even possible to separate sharply the
functions the pitch of the voice plays in the various phases of speaking and singing” (p. 58). This statement suggests that there are many difficulties inherent in speech melody perception, and that tone perception in speech may be more subtle than that required for perceiving pitch relations in a musical context.

Not only is observation of linguistic tones technically demanding, as C-H admitted, but observation of linguistic tones in melody is also constrained.

Due to the restricted relative pitch levels of the language, there are certain limitations to the melodic contour of songs created by Cantonese text. Writing Cantonese songs in which the melodies fit the characteristics of the language is not as free as writing songs in other nontonal languages. (p. 49)

Writing songs using a nontonal language would be highly restricted if the writing of such songs was unnecessarily concerned with matching the pitch characteristics of speech to a song. It is the relationship between speech and song explored in this chapter that is the focus of the present investigation. I now turn to a discussion of the research problem and the questions addressed.

Statement of the Problem

A comparison of vocal development between children of tonal and nontonal languages requires an interdisciplinary examination of speech related vocal behaviours because acoustical and perceptual definitions for the different behaviours of singing and speaking have not yet been reliably established, particularly in young children. Research on the development of singing in young children should therefore focus on the contexts that constitute singing and speaking.

Research Questions

(1) How do young children’s vocal fundamental frequencies change across time, and how do such changes inform us about their singing and speaking?
(2) In what contexts do young children sing?
(3) What forms of vocalizations are observed in young children and what are their roles in vocal development?

(4) Is there evidence suggesting that first spoken language may affect vocalizations? That is, is vocal development different in tonal and nontonal languages?
CHAPTER THREE
FIELDWORK METHODOLOGY
Concerns and Analytical Premises

Music may be viewed as a construct whereby sound is organized according to social conventions (Walker, 1990, 1996). Blacking (1984/1995) states that “[m]usical styles are based on what people have chosen to select from nature as a part of their cultural expression rather on what nature has imposed on them” (p. 33). Contextualization as an approach to understanding the construction of meaning, and subjectivity as an approach to interpreting meaning are discussed below.

It has been maintained that context is indispensable to music. For example, Nadel (1951) asserts that music is never “action autonomous” (pp. 87-90). Blacking points out that “every musical performance is a patterned event in a system of social interaction, whose meaning cannot be understood or analyzed in isolation from other events in the system” (1984/1995, p. 227). Likewise, Eisner (1996) argues that “[c]ontext applied to qualities makes interpretation possible” (p. 9). There is, therefore, a need to understand the context in which the product (music) and the action (music making) exist. Hence, context is an integral part of the present investigation of children’s vocalizations.

Subjectivity is also indispensable to the interpretation of meaning in music. Jansen and Peshkin (1992) point out that “[o]bjectivists study objects (following Durkheim and Comte), while subjectivists study processes, people, or events” (p. 686). The focus of this research is to understand young children’s vocal development through examining the nature and context of vocal events but not vocalizations per se. In view of the concern over verbal interpretations of musical behaviour, Blacking (1984/1995, p. 226) points out the importance of establishing intersubjectivity. I interpreted children’s vocalizations using various data to establish intersubjectivity. These data included: field observations and notes, parents’ accounts, audio recordings, computerized acoustic analyses, descriptive
categorization, and perceptual evaluation by three human judges. More comprehensive interpretations of data were made possible by integrating these quantitative and qualitative strategies.

Mediating the Qualitative and Quantitative Paradigms

Qualitative research is used to understand the social context of many phenomena as perceived by the participants. Quantitative research is based on objectivity and quantification of phenomena. There is increasing interest in employing a combination of quantitative and qualitative approaches in studies of human behaviour. Observations made in this study were informed by approaches developed by ethnographers and both qualitative and quantitative methods were employed in the evaluation of sung and spoken vocalizations.

Method

The Sample and Subjects

Subjects were recruited through already established personal contacts, including friends and colleagues. Various aspects of the study were communicated to parents who had young children in the appropriate age groups and who expressed interest in learning more about their child’s vocal development. I met and communicated with the parents and children personally before the study, to ensure that: (1) subjects met the age and language criteria for inclusion in the study; (2) subjects were comparable on the basis of social status; (3) all parents expressed similar interest in the musical development of their children; and (4) all subjects were in good health. There were no indications of auditory deficit, either observed directly or reported by the family. Throughout the forty-two months of the study, contact was maintained with the parents to update them on progress made by each child and to obtain more in-depth views about the child’s vocal development. Confirmation and synthesis of data were enhanced by this strategy.
The eight girls, age two to four years, were from an urban, middle-class population in Canada. Six of the eight subjects lived in a family housing area associated with a university where their parents either worked or studied. Parents ranged in age from 30 to 45 years of age and had completed post-secondary education; some parents had graduate degrees. Each subject was a singleton child in a two-parent family. I purposely included a sample of children who spoke tonal and nontonal languages: two spoke Cantonese, two spoke Mandarin, and four spoke English as their first language.

All subjects attended full-time daycare centres or preschools. Six of the eight subjects attended daycare and preschool facilities at the University residence. Two subjects attended a Montessori private preschool. The daycare centres and preschools were assumed to be comparable, because they are governed by provincial legislation.

McMillan and Schumacher (1993) state that “internal validity of qualitative design is the degree to which the interpretations and concepts have mutual meaning [sic] between the participants and researcher” (p. 391). Due to the limited verbal and social skills of children at the ages studied, it was not possible to verify my interpretations with them.

Data for the present study were collected longitudinally over forty-two months, and corroborated by parent report. During this period, I got to know parents and children personally and was then able to observe consistency and changes in their behaviours. Ongoing data analyses allowed me to compare and corroborate observations on each child during different phases of vocal development.

To obtain information about the type of songs the children had acquired, I asked the children to sing their favorite songs and to teach me the games they learned in school. It seemed that all children had acquired traditional nursery rhymes and popular North American children’s songs such as: *Eency Weency Spider; The Wheels on the Bus; Bingo;*

18 A nursery rhyme is a short, rhymed, traditional poem for children sometimes sung to a simple melody.
It's Raining, It's Pouring; Pop Goes the Weasel; and This Old Man. It was not clear whether children had acquired these songs at their daycare centres or elsewhere; nevertheless, their song performances provided me some idea of their repertoires. I observed that some songs were favourites of all the children: The Wheels on the Bus for 3-year-olds, and Bingo for 4- and 5-year-olds. While most of the songs were simply sung in the daycare centres, sometimes finger plays and simple actions were taught along with the songs (e.g., Eency Weency Spider). To the best of the parents' knowledge, no formal or structured music program was incorporated into the daycare curricula but it should be noted that daycares do routinely encourage group singing of the songs noted above. Six of the subjects had some informal musical training with me before the study commenced.

The Role of the Researcher

Let me describe my relationship with six of the subjects before the study commenced. I had been offering music programs at the University residence on Saturday mornings for six years. Most of the subjects in the present study had attended my music program for eight or 16 forty-five minute sessions before the study commenced. Those children typically joined my music program at the age of 18 months to 20 months when they had little social experience, and no concept about schooling and teacher. I therefore introduced myself by my first name so the children could accept me as their friend. I always addressed the music sessions I offered as music games, and we met on Saturday mornings to play. I contacted the parents to invite them to participate in the study only after their child left the music program. The following summary of the music activities I offered to two-year-old children indicates the types of musical experiences those children had before the study commenced.

1. Singing songs--I would introduce a song, then parents modelled and joined in to sing as a group. Songs were usually accompanied by finger play or parent-toddler activities. At a later stage, singing games in a group setting were introduced along with the songs.
2. Movement—children responded to recorded music, matching body movement to fundamental music concepts (high/low, loud/soft, fast/slow).
3. Listening—quiet time with “sleeping music”, passing balls while listening attentively.
4. Playing musical instrument—explored and shared new ways of making sound, using instruments to accompany songs and recorded music.
5. Connecting auditory and visual experience—children pointed at pictures while singing; fingers move to the subjects of the song. Later, they traced the graphical representations of musical properties in the song (phrasing, melodic contour) with fingers.

For the two other subjects who did not attend my music program, I was a family friend and had established a rapport relationship with them through previous home visits and social events. For the present study, I assumed the role of an observer-participant during the home visits. Before my visits, the parents would tell their child that I would come to play. Since I knew all children before the study commenced, I was treated as a visiting family friend.

Home Visits

I visited each child every four to six months over forty-two months. A typical home visit consisted of approximately forty-five minutes to tape the child speaking and singing, and at least fifteen minutes of an informal conversation interview with the parents. Field notes were also written after each home visit to describe the setting and the activities of the children. To ensure the accuracy of these notes, the written records of the observations were completed on the same day immediately after the home visit. (See Appendix F for a sample of field notes.)
Setting for the Site of Data Collection

All data collection took place at each subject's home. Whenever possible, the children were asked to sit on the carpet in their bedrooms where they were surrounded by their own toys, books, and other familiar things to ensure that interaction could take place in a comfortable manner. Sometimes, data were collected in the living room, where the child would be asked to bring along some favourite toys and books. I felt that the atmosphere was usually more formal when we met in the living room because I was seen to be a guest of the family, whereas when the child interacted with me in the bedroom, I was treated as the child's friend. When the data were collected in the living room, more time was needed to allow the child to feel comfortable when interacting with me. In this setting, the child usually appeared to be more self-conscious as though the expectation was proper behaviour in front of a guest.

The parents sometimes sat beside or within several feet of the child to show support. Some parents deliberately stayed away from the area for about 15 minutes to provide an opportunity for the child to become independent when interacting with me. Usually the parents were curious, however, and returned to the scene after a short time. Some of the parents, especially the mothers, typically assumed the role of an observer at the beginning of the recording session, but would soon join as a participant. Other parents participated actively throughout the recording session, encouraging their children to sing and speak. Since the children were already familiar with me through personal contact before the study commenced, their responses to the recording procedures were enthusiastic and all recording sessions were completed without difficulty.

Data Collection

1. Recording children's vocalizations

Speech and singing responses were recorded during child-adult interactions and in various play settings. All speech and singing performances were recorded with a portable
professional standard cassette recorder (Marantz PMD 201) using a unidirectional professional microphone (Shure, SM48-LC). “Digital accuracy” cassette tapes (TDK SA-X) were used to store the vocal signals.

Singing: During each visit, the children were asked to sing favourite songs, regardless of the language they preferred. Most children sang English songs since English was the language of instruction at school, and because I spoke English when I visited them. Some younger children included songs in their first language, but their song repertoire in their first language was typically reduced after they began attending English-speaking full-time preschools. If they showed no preference for a particular song when asked to sing, the parents would remind them of their favourite songs. Seldom would a child refuse to sing, but would at times find it difficult to choose a song without assistance. It appeared more natural to first engage the child in a conversation, then have the parents gradually encourage the child to “teach me” her favourite song by performing it. Frequently, a child would be more interested in singing a song when the mother had prompted her by singing the beginning of the song.

Speaking: To record the children’s English utterances, I used two preschool picture books with objects illustrated in colour photos. Recordings of the children speaking in their native language were obtained during their conversations with their parents and with me. Whenever appropriate, I also spoke the child’s native language (Mandarin or Cantonese) to make them feel less inhibited. Some children shifted swiftly between the two languages they spoke, and were able to respond according to the language spoken by either the parent or myself. Speech utterances were also collected by inviting them to tell stories. A typical session included playing with toys, allowing the child to read picture books that I brought to the session, and interacting with the child through her favourite storybooks and songs. To maintain a relaxed atmosphere during data collection, no specific sequence was adopted.
2. Observation and field notes

Descriptive notes were written at the end of each visit to record the setting and the observed activities. I tried to include as much information as possible to give me a more elaborated context of the children’s vocalizations. Some general questions adapted from Cohen, Stern, and Balaban (1983, p. 175) were used to guide writing the field notes. The following observations were made:

(a) the source of stimulus (a storybook, an invitation to sing, vocal matching, etc.) versus the type of vocal response;

(b) other corresponding behaviours (body movements, use of toys, etc.) observed when the vocalizations occurred;

(c) the child’s communication intent as inferred from the scenario when the vocalization occurred (items a and b) or verification with the child; and

(d) the parents’ interpretation of the vocal responses.

3. Informal conversation

I usually had a ten-minute chat with the parents after each recording session to obtain information to facilitate data interpretation. These included the types of songs sung and listened to by the child; musical games played by the child; the child’s accessibility to musical instruments, musical activities, and audio visual resources. For field notes and informal conversations, I tried to keep the verbal descriptions brief. I was aware that a more literal description would give me a relatively precise account of the observations. Therefore, some of the conversations were also taped to enable me to cite exact wording used by parents in response to my questions. Due to the informal nature of the conversations, those general questions frequently led to further discussion focusing on some of the interesting vocal responses collected during that particular recording session.
Analyses of Data

The vocal data in the present study were analyzed by means of: (1) computerized acoustic analysis, and (2) human judges’ perceptual evaluation of the vocalizations. The functions of the judges’ anecdotal descriptions were (a) to assist the perceptual categorization of vocalizations as vocal forms (i.e., speech, song, intermediate vocalization and to provide perceptual judgment needed for acoustic analyses; and (b) to assess the musical maturity of the children. These analytical measures are now described.

Acoustic Analyses

Method

Taped vocal signals were initially captured digitally at a sampling rate of 40,000 Hz and subsequently sampled to 20,000 Hz. The sampling rate of 40,000 Hz meant that the original analog signal was sampled 40,000 times per second. Based on the Nyquist theorem, the number of samples needed to represent a signal is twice the highest frequency of interest of the signal. Vocal signals captured at a sampling rate of 40,000 Hz would yield an analysis limit of up to 20,000 Hz, which provides enough information of the original analog signal with minimum loss. (See Kent & Read, 1992, pp. 48-55 for more information on digital sampling of speech.) Digitized speech and singing responses were analyzed by computer software (Computerized Speech Laboratory [CSL], developed by Kay Elemetrics, NJ, USA, specifically for speech and singing analyses).

1. Pitch Analyses

The fundamental component frequency ($F_0$) defines the pitch of a tone, but the human brain can assign pitch and even recognize other musical relationships based on a limited number of harmonically related partials. Attempting to quantify a sensation is the principle problem associated with studying the relationship between physical properties of a tone and attributes of human perception. (See de Boer, 1976; Moore, 1989; Bernade, 1990, for a discussion of theories of pitch perception.)
Data on the frequency of the first harmonic ($F_0$) provides a quantitative evaluation which normally corresponds to perceived pitch. Findings from acoustic analyses may therefore be used to assist musical transcription. $F_0$ may also be used to make indirect estimates of the physiological state of the vocal mechanism. The present longitudinal analyses of $F_0$ illustrates some of the general trends observed in children's vocal development.

To perform pitch extraction on a waveform using CSL, standard procedures of acoustic and auditory phonetics were used to locate that portion of the signal corresponding to the phonation. Vocal segments were identified and marked off into smaller segments corresponding to individual vowels. This was accomplished by locating a stable section of each vowel in the signal through the graphic display of the waveform and the spectrogram, while simultaneously playing the vocal signal. $F_0$ was measured for all stable sections of each vowel for all selected utterances. The mean $F_0$ of all vowels was derived for each type of utterance for each child. Figure 3.1 illustrates the fundamental frequency of a selected portion of a singing segment which is computed and displayed over the resulting time domain. The $F_0$ results were checked with a Fast Fourier Transform (FFT) by averaging several well-defined upper harmonic peaks (procedure recommended by Dr. Guy Carden, personal communication, April 1998).
Pitch extraction was performed on "o-ver and pull me out", sung by a five-year old girl. Pitch calculation (statistics for the entire selected segment and frame by frame analyses) is shown in separate windows.
2. Spectrographic Analyses

Time and frequency distributions of energy in sound may be represented graphically with a spectrogram. With appropriate filter settings, individual harmonics of vocal sounds can be visually distinguished. Variations in voice pitch are represented as curves (the spectrum envelope). The peaks and valleys in the spectrum envelope are determined by the frequencies of the characteristic vibration patterns of the corresponding vocal tract shapes (the resonant or formant frequencies). (See figure 3.2 for a spectrogram display from the CSL.) Thus a spectrogram can illustrate changes in resonant frequencies of speech sounds for each utterance. Variations in energy intensity in the overtone content of the voice source are represented in corresponding gradations of the colour displayed.

Since speech sounds are related to interaction between particular characteristic frequencies of the sound source (vocal folds) and vocal tract, the spectrogram illustrate the production not of perception, thus provides a quasi-quantitative method of examining the acoustic structure of vocal sounds.

Figure 3.2

A CSL Spectrogram of a Singing Segment

A spectrogram display of the same singing segment shown in Figure 3.1. Using narrow band filter, the relative intensity of the harmonics is evident.
The present study attempted to provide physical descriptions of the vocalizations of the children who formed the subject pool. Nevertheless, since comprehensive studies of young children's sung and spoken sounds using these physical descriptions are scarce, acoustic analysis is used in the present study as a descriptive device to supplement perceptual judgments (Kent & Murray, 1982/1991).

Perceptual Evaluations

In the present study, I used three judges to provide perceptual analyses and interpretation of the recorded vocalizations. Transcriptions of vocalizations were used for illustrative purpose when needed, and results from acoustic analyses were then used in conjunction with the judges' perceptual evaluation.

Qualifications of Judges

For the present study, judges were selected on the basis of their vocal training, and experience working with children; that is, (1) all judges had received vocal training both during their childhood and during their university studies; (2) judges all had received an undergraduate degree in music. Both judges A and B had post graduate degrees in music education, and judge C majored in music education in her undergraduate studies. All judges had experience working in music education programmes. Both judges A and B had more than five years experience teaching in high schools. Judge A also had four years of experience working with preschool music programs. Judge B had extensive training in the Orff and Dalcroze approaches and taught in-service elementary school teachers in music education methods courses. Although judge C had never worked in a formal school setting, she had over ten years experience working in a private music school for children, where she taught and managed choral programs. Her experience with children learning music might therefore be considered comparable to that of judges A and B.

Kent and Murray (1982/1991) called such an approach "acoustic documentation", which is the "acoustic confirmation of perceptually determined categories, in order that linguistic or general perceptual bias of the observer may be minimized" (p. 411).
1. Perceptual Categorization of Vocalizations

I attempted to provide a descriptive, yet perceptual account of the children’s vocalizations. Three predetermined categories, *speech utterances*, *singing responses*, and *intermediate vocalizations* were used. Whereas the characteristics attributed to speech and songs were determined according to Western tradition, intermediate vocalizations denote vocalizations that could not be confidently classified as either speech or song. In identifying intermediate vocalization, both the acoustical characteristics and the communication intent were considered.

Guideline for Assessment

To ensure that all judges were using the same assessment criteria, a list of questions was formulated to provide a guideline for writing anecdotal comments. (See Appendix G for an assessment sheet.) Open-ended questions were used to encourage descriptive responses. For example, instead of posing questions of the type that encourage dichotomous response (i.e., yes-no answer) such as “Did the child sing when invited by the adult?”, the questions were phrased as, for example, “Describe how the child responded to the adult’s invitation to sing a song”. Questions were also designed to encourage more complex and elaborate responses; for example: “What did you observe in the child’s vocalizations to convey your sense of her achievement in singing?”. This type of query requires a more detailed narrative, and therefore provides a richer interpretation.

To assist the judges, three assessment sheets were provided (one sheet for each question). Questions were formulated on the basis of *a priori* issues discussed in earlier chapters. Copies of the recorded vocalizations were given to each judge independently, allowing the tapes to be replayed as many times as needed to provide elaborated comments. Each judge was instructed to answer any one part from questions 1 and 2, and both parts of question 3. Question 1 required descriptive comments on the context of the vocalization. Question 2 assessed the child’s vocal competency according to the judges’ presupposition
of the attributes attached to the vocal behaviour she identified. Question 3 required the judges to analyze vocalizations for their perceived aspects on musical pitch and stylized vocal behaviour appropriate to the culture.

2. Assessment of Musical Maturity

$F_0$’s were used for musical transcriptions into Western music notation to aid interpretation and comparisons with established literature on children’s vocal development (Papousek & Papousek, 1981; Moog, 1976; Welch, 1986, 1994; Welch, White, & Sergeant, 1996). The following parameters of Western music were used to evaluate the children’s level of musical maturity: scalar structure; tonality; melodic contour; melodic pattern; treatment of motive; repetition and contrast; cadence; dynamics; articulations (including rhythmic structures); interpretation of the lyrics by manipulating musical elements/controls; and mannerism of singing styles (e.g., operatic “vibrato”, head tones versus chest voice and melodic and rhythmic embellishment, etc.).

Summary of Sources and Strategies for Data Analyses

To investigate young children’s vocal development through observations of their vocalizations, the present study employed a range of eclectic approaches. The following summary outlines the multiple strategies used in data collection and analyses, and serves as an introduction to results presented in the next chapter.

- Acoustic Analyses  
  (1) $F_0$ Mean and range were analyzed using the pitch extraction function of CSL.
  (2) Sound spectrograms of the vocal data were used to illustrate variation of phonation frequency and the comparative intensity of energy at different harmonic peaks.

- Ethnographic Analyses  
  (1) Field notes were used to document the context of the vocalizations.
(2) Singing performances were transcribed into Western musical notation.

(3) Vocal utterances were classified into speech, song, and intermediate vocalizations.

(4) Independent judges provided perceptual evaluation of the vocal data according to an assessment guideline.

(5) Parents commented on the child's vocal behaviour during home visit and provided information of the child's language and musical experience.
CHAPTER FOUR
DESCRIPTION, CATEGORIZATION, AND OBSERVATION OF
CHILDREN'S VOCALIZATIONS

Subjects

The subjects have been given pseudonyms to make it possible for the reader to follow the discussion in this chapter. Basic information about the children is presented in Table 4.1 to facilitate comparisons amongst subjects. Additional information about individual children is addressed in the discussion of results.

Table 4.1
Some Basic Information About the Subjects

<table>
<thead>
<tr>
<th>Name of Subject</th>
<th>Age (at start of the study)</th>
<th>First spoken Language</th>
<th>Place of Birth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heidi</td>
<td>18 months</td>
<td>Mandarin&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Taiwan</td>
</tr>
<tr>
<td>Clare</td>
<td>25 months</td>
<td>English</td>
<td>Canada</td>
</tr>
<tr>
<td>Vicki</td>
<td>27 months</td>
<td>Cantonese&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Hong Kong</td>
</tr>
<tr>
<td>Jodi</td>
<td>29 months</td>
<td>Mandarin (English)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Canada</td>
</tr>
<tr>
<td>Wendy</td>
<td>31 months</td>
<td>Cantonese&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Hong Kong</td>
</tr>
<tr>
<td>Amber</td>
<td>33 months</td>
<td>English</td>
<td>Canada</td>
</tr>
<tr>
<td>Gina</td>
<td>34 months</td>
<td>English</td>
<td>Canada</td>
</tr>
<tr>
<td>Polly</td>
<td>38 months</td>
<td>English (Spanish)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Peru</td>
</tr>
</tbody>
</table>

Note. The language given in parentheses is the secondary language spoken by a child at home.

<sup>a</sup>The child gradually became bilingual (Mandarin/English) at around 1;3.
<sup>b</sup>The child gradually became bilingual (Cantonese/English) at around 2;1.
<sup>c</sup>The child gradually became bilingual (Mandarin/English) at around 0;8.
<sup>d</sup>The child gradually became bilingual (Cantonese/English) at around 2;3.
<sup>e</sup>Both of her parents are native speakers of Spanish from Peru. English is the primary language spoken at home, as the mother explained, because the child became more reluctant to communicate in Spanish after they moved to Canada when she was 2;0.

<sup>20</sup>Throughout this report, "subjects" and "children" are used as equivalents in both ethnographic and acoustic analyses.
Classification of Vocal Forms

Three major types of vocalization, *speech utterances*, *singing responses*, and *intermediate vocalizations*, were collected. *Speech utterances* included *single words* elicited from the subjects by showing them a colour picture book, giving a relatively similar English vocabulary across subjects, although the same picture occasionally elicited two different words, e.g., “pants” and “jeans”. *Sentences* were collected during conversations when young children tried to express themselves in multiple words. These sentences were usually not grammatically accurate but were, nevertheless, effective for communication. It was not until about age 4-5 that *sentences* became more complex.\(^{21}\)

*Singing responses* included both *learned song* and *spontaneous song*. Mostly the children sang popular children’s songs. I called them *learned songs* after Moog (1976) because the children were probably recalling these songs according to a previously introduced model as they sang. In other words, the children had been taught to sing the song through a vocal model. *Spontaneous song* includes a variety of *singing responses* that were initiated by the child. Adapting the labelling from Moog, I included “pot-pourri” songs, “imaginative”, and “narrative” songs under this category.\(^{22}\) Some of the *spontaneous song* could also have been attempts to model songs that had been presented to the children rather than originally improvised singing. However, it was classified as a *spontaneous song* when the judges could not identify a specific song the child might be attempting.

Children were also asked to tell stories as they read or made up the stories using their favourite books. Since speech responses elicited during story-telling were more emotional and spoken with exaggerated inflections, I decided to compile them into a category named *story-telling*. Since *story-telling* possessed some pitch characteristics quite

\(^{21}\) A complete description of the language of children of this age may be found, for example, in texts on children’s language (Fletcher & MacWhinney, 1996).

\(^{22}\) More detailed description of these three song types may be found in chapter 2.
distinctive from *single word* utterances and *sentences*, results from *story-telling* were excluded in later discussion of *speech utterances*.

Vocalizations such as *story-telling*, were best described as intermediate to speech and song. These vocalizations may be seen as *intermediate* because their functional characteristics appeared to be somewhat ambiguous according to their daily usage. Two major criteria were used for the present classification: (1) the acoustical characteristics of the vocalization—when there was a mixture of sung and spoken utterances in the vocalization, and (2) the communicative intent—when vocal utterances did not correspond to the anticipated stimulus. Vocalizations classified as "intermediate" are discussed later.

**Results of Acoustic Analyses**

The $F_0$ data presented in this chapter are integral to an understanding of the children's vocal development because they provide quantitative documentation to support the qualitative perceptual evaluation. Results from spectrographic analyses are subsequently presented along with qualitative data in a later section. Since the present study is an ethnographic investigation, in-depth statistical comparisons are not given.

The *Statistical Package for the Social Sciences for Windows (SPSS Windows)* was used to calculate the means and standard deviations of vocal fundamental frequencies ($F_0$). Analyses were derived from two main data sets: (1) $F_0$ results for all types of vocalizations arranged according to four age levels (Table 4.2), and (2) $F_0$ results for each subject's *speech utterances* and *singing responses* (Table 4.3).
Table 4.2
Mean $F_0$ for Children’s Vocalizations Across Age Levels and Languages

<table>
<thead>
<tr>
<th>English Utterances</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age Group</td>
<td>$n^a$</td>
<td>2 - 3 ($n = 7$)$^b$</td>
<td>$n$ 3 - 4 ($n = 8$)</td>
<td>$n$ 4 - 5 ($n = 7$)</td>
<td>$n$ 5 - 6 ($n = 4$)</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>Mean</td>
<td>SD</td>
<td>Range</td>
<td>Mean</td>
</tr>
<tr>
<td>Single words</td>
<td>7</td>
<td>269-392</td>
<td>319 30</td>
<td>8</td>
<td>266-455</td>
</tr>
<tr>
<td>Sentences</td>
<td>7</td>
<td>260-466</td>
<td>310 42</td>
<td>8</td>
<td>252-493</td>
</tr>
<tr>
<td>Story-telling</td>
<td>4</td>
<td>225-417</td>
<td>336 36</td>
<td>6</td>
<td>220-539</td>
</tr>
<tr>
<td>Learned Songs</td>
<td>7</td>
<td>231-493</td>
<td>368 39</td>
<td>8</td>
<td>228-503</td>
</tr>
<tr>
<td>Spontaneous Singing</td>
<td>3</td>
<td>261-488</td>
<td>333 36</td>
<td>3</td>
<td>235-487</td>
</tr>
<tr>
<td>Intermediate Forms</td>
<td>3</td>
<td>202-491</td>
<td>345 54</td>
<td>2</td>
<td>255-488</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chinese Utterances</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age Group</td>
<td>$n$</td>
<td>2 - 3 ($n = 4$)</td>
<td>$n$ 3 - 4 ($n = 4$)</td>
<td>$n$ 4 - 5 ($n = 3$)</td>
<td>$n$ 5 - 6 ($n = 2$)</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>Mean</td>
<td>SD</td>
<td>Range</td>
<td>Mean</td>
</tr>
<tr>
<td>Single words</td>
<td>4</td>
<td>242-466</td>
<td>331 40</td>
<td>3</td>
<td>226-432</td>
</tr>
<tr>
<td>Sentences</td>
<td>4</td>
<td>237-490</td>
<td>324 56</td>
<td>3</td>
<td>190-454</td>
</tr>
<tr>
<td>Story-telling</td>
<td>3</td>
<td>261-509</td>
<td>355 53</td>
<td>2</td>
<td>250-510</td>
</tr>
<tr>
<td>Learned Songs</td>
<td>2</td>
<td>261-490</td>
<td>380 48</td>
<td>1</td>
<td>292-489</td>
</tr>
<tr>
<td>Spontaneous Singing</td>
<td>3</td>
<td>251-497</td>
<td>339 49</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Intermediate Forms</td>
<td>4</td>
<td>248-500</td>
<td>352 66</td>
<td>3</td>
<td>199-602</td>
</tr>
<tr>
<td>All types Group</td>
<td>246-475</td>
<td>341 46</td>
<td>238-496</td>
<td>326 49</td>
<td>198-475</td>
</tr>
</tbody>
</table>

Note: The range, means and standard deviations of the fundamental frequency are in Hertz.

$^a$Number of children in each language group who responded in the particular type of vocalization.

$^b$The total number of child participants in that language group.
Table 4.3
Mean F0 for Singing and Speaking for Individual Subjects

<table>
<thead>
<tr>
<th>Age Group</th>
<th>1-2</th>
<th>2-3</th>
<th>3-4</th>
<th>4-5</th>
<th>5-6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Range</td>
<td>Mean</td>
<td>SD</td>
<td>Range</td>
<td>Mean</td>
</tr>
<tr>
<td><strong>Singing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heidi</td>
<td>259-428</td>
<td>342</td>
<td>36</td>
<td>208-486</td>
<td>323</td>
</tr>
<tr>
<td>Clare</td>
<td>244-471</td>
<td>341</td>
<td>44</td>
<td>221-472</td>
<td>341</td>
</tr>
<tr>
<td>Vicki</td>
<td>258-626</td>
<td>420</td>
<td>56</td>
<td>324-615</td>
<td>398</td>
</tr>
<tr>
<td>Jodi</td>
<td>226-420</td>
<td>299</td>
<td>33</td>
<td>201-396</td>
<td>284</td>
</tr>
<tr>
<td>Wendy</td>
<td>258-520</td>
<td>377</td>
<td>51</td>
<td>200-499</td>
<td>357</td>
</tr>
<tr>
<td>Amber</td>
<td>259-480</td>
<td>344</td>
<td>41</td>
<td>317-501</td>
<td>330</td>
</tr>
<tr>
<td>Gina</td>
<td>259-501</td>
<td>364</td>
<td>46</td>
<td>271-480</td>
<td>351</td>
</tr>
<tr>
<td>Polly</td>
<td>251-501</td>
<td>374</td>
<td>41</td>
<td>255-475</td>
<td>366</td>
</tr>
<tr>
<td><strong>Speaking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heidi</td>
<td>290-347</td>
<td>326</td>
<td>28</td>
<td>289-393</td>
<td>306</td>
</tr>
<tr>
<td>Clare</td>
<td>225-407</td>
<td>346</td>
<td>55</td>
<td>230-401</td>
<td>321</td>
</tr>
<tr>
<td>Vicki</td>
<td>278-512</td>
<td>341</td>
<td>48</td>
<td>252-519</td>
<td>339</td>
</tr>
<tr>
<td>Jodi</td>
<td>236-401</td>
<td>284</td>
<td>34</td>
<td>202-392</td>
<td>267</td>
</tr>
<tr>
<td>Wendy</td>
<td>259-487</td>
<td>319</td>
<td>38</td>
<td>219-489</td>
<td>306</td>
</tr>
<tr>
<td>Amber</td>
<td>247-478</td>
<td>337</td>
<td>47</td>
<td>252-528</td>
<td>319</td>
</tr>
<tr>
<td>Gina</td>
<td>231-498</td>
<td>316</td>
<td>46</td>
<td>213-470</td>
<td>291</td>
</tr>
<tr>
<td>Polly</td>
<td>247-478</td>
<td>321</td>
<td>33</td>
<td>196-431</td>
<td>309</td>
</tr>
<tr>
<td><strong>Group (Speaking)</strong></td>
<td>252-453</td>
<td>321</td>
<td>42</td>
<td>233-458</td>
<td>306</td>
</tr>
<tr>
<td><strong>Group (Singing)</strong></td>
<td>251-492</td>
<td>355</td>
<td>43</td>
<td>251-493</td>
<td>344</td>
</tr>
<tr>
<td><strong>Group (Speaking and Singing)</strong></td>
<td>252-473</td>
<td>338</td>
<td>43</td>
<td>242-475</td>
<td>325</td>
</tr>
</tbody>
</table>
1. Fundamental Frequency of Vocalizations Across Age Levels

Fundamental frequencies for different types of vocalizations at four age levels are summarized in Table 4.2. For instance, the 2-3 level included children whose ages ranged from 24 to 35 months; the 3-4 level included children whose ages ranged from 36 to 47 months. Results from English and Chinese utterances are reported separately in the table. Also, only Heidi falls into the 1.5-2 level, I have therefore, excluded her data from the summary table and discuss her vocal behaviours along with the qualitative data.

2. Speaking and Singing Fundamental Frequency of Individual Subjects

Mean $F_0$ for speech utterances and singing responses collected from each of the subjects are reported in Table 4.3. For Chinese bilingual children, the mean $F_0$ calculation includes both English and Chinese utterances.

Results for Speech utterances were based on single word utterances and sentences. The singing data included learned songs and spontaneous songs. Mean $F_0$ results for singing were calculated using five learned songs and five spontaneous songs sung by each child at each age level. Mean $F_0$ results for speaking were calculated using 30 single word utterances and 10 sentences.

Discussion

1. Language shift in the Chinese bilingual children

Although the four tonal language speakers communicated mainly in Chinese at home, a gradual shift in their first language of communication over the period of data collection was observed. Whereas the Chinese girls had increased the size of their English vocabulary and gained competence in communicating in English, their Chinese utterances decreased over time. This was particularly evident by the third year of data collection when the Chinese bilingual children relied heavily on their parents’ prompting to tell stories in

---

23 Throughout this chapter, the plural form “mean fundamental frequencies” is written “mean $F_0$”.

66
Chinese. As a result, none of these Chinese bilingual children succeeded in telling a complete story in her first spoken language. The English song repertoire of these Chinese bilingual children, however, increased at the same time as their Chinese song repertoire decreased. This difference became more apparent after the subjects began attending preschools where they were exposed to English nursery rhymes and children’s songs. When invited to sing Chinese songs, those children would say that they had forgotten those songs. When she failed to perform any of her favourite Chinese songs at age 5 years 3 months, one of the subjects (Vicki), explained to me that she “had a particularly bad memory for Chinese songs”. It was clear that although the Chinese bilingual children still communicated in their first spoken language at home, an overall language shift was occurring. It can be seen in Table 4.2 that fewer Chinese responses were collected as age increased, and no responses were collected on some of the vocalization categories between the ages 4 to 6 years.

2. Mean F0 Results for All Vocalization Types

To facilitate comparison, Figures 4.1 and 4.2 illustrate the mean F0 for different types of vocalizations produced in English and Chinese respectively. The mean F0 across different types of vocalizations suggest some differences in the communicative use of pitch in different vocalizations.
Figure 4.1
Mean Fo for Chinese Utterances

- Sentences
- Single words
- Story-telling
- Learned Songs
- Spontaneous Singing
- Intermediate Forms

Figure 4.2
Mean Fo for English Utterances by All Subjects

- Sentences
- Single words
- Story-telling
- Learned Songs
- Spontaneous Singing
- Intermediate Forms

* The mean Fo of this intermediate vocalization is based on vocal data collected from one child only.

68
A very similar pattern for mean $F_0$ emerges in vocalizations across languages (see Figure 4.1 and 4.2). *Sentences* have the lowest mean $F_0$, with *single word* utterances having slightly higher mean $F_0$. *Learned songs* have the highest mean $F_0$, which is especially evident in English (Figure 4.2). The mean $F_0$ for *story-telling*, *spontaneous songs*, and *intermediate vocalizations* lie somewhere between the mean $F_0$ for *learned songs* and *speech utterances* (*sentences* and *single words*).

Results show that when children sang *spontaneously*, they tended to use a singing tessitura closer to their speaking $F_0$ (i.e. lower) than when they sang *learned songs* (see Figure 4.2). The results replicate findings of Flowers & Dunne-Sousa (1988), and support studies which report that preferred singing tessitura for children is lower than the pitch range of music published for children (Wilson, 1970; Welch, 1979; Atterbury, 1984).

The mean $F_0$ for *story-telling*, *spontaneous songs*, and *intermediate vocalizations* tend to form a cluster lying between the mean $F_0$ for *learned songs* and *speech utterances*. *Learned songs* were consistently performed at a higher mean $F_0$ than *speech utterances*, while *story-telling*, *spontaneous songs*, and *intermediate vocalizations* have less distinctively defined relative height for pitch centres. These results appear to suggest that the relative height of the pitch centres for different forms of vocalizations may be already established in young children by age two.

3. Developmental Trend in Vocal Fundamental Frequency

When all vocalizations were considered, a longitudinal decrease in mean $F_0$ is found across languages (Figures 4.1 and 4.2). Since $F_0$ is considered to be inversely proportional to vocal fold length and laryngeal size, growth in children's body size over time may generally explain the developmental changes in $F_0$.\(^{24}\) The lowering of $F_0$ across

\(^{24}\) An exception was found in the mean $F_0$ of English *intermediate vocalizations* at age level 5-6 which appeared to have been raised dramatically. This exceptional case was derived from only one child, who "sang" a story in a vocal range way above her normal singing range. I called this interesting vocal response *story-singing*.
time is consistent with established literature (Table 4.4 and Figure 4.3). The data also support previous reports that a relatively more gradual decrease in mean $F_0$ is found between the age of 3 to 6 in comparison to the rapid decrease reported in infancy (Titze, 1994; Kent, 1976). Based on a simple model that predicts a lowering of $F_0$ across time as a function of increase in membranous length, Titze (1994) points out that since a relatively constant rapid increase in membranous length of the vocal fold is found at this age, the gradual decrease in the mean $F_0$ between 3-6 years is unanticipated. The major deviations from the model occur at ages 3 to 10. It is not clear why the fundamental frequency drops less in this age range. Length appears to increase in a rather constant manner . . . . There may be compensating dimensional changes in the vocal fold tissue layers, or there may be changes in tissue elasticity. It is reported, for example, that development of the vocal ligament and the thyroarytenoid muscle begins at ages 3 to 4 (Hirano, Kurita, & Nakashima, 1981). This could well stiffen the tissue, on the average, counteracting the drop in $F_0$ due to increasing length. It is also possible, though less likely, that children maintain a hypertense musculature in the larynx while producing fundamental frequency in speech during those years. (p. 180)

Nevertheless, the developing structural properties of a young child’s vocal fold tissue layers can cause changes in the intrinsic stress and in turn affect phonation frequencies (Kent & Vorperian, 1995; Morrison & Rammage, 1994).

When compared with research literature, speech $F_0$ results found in the present study also appeared to be slightly higher but within the range reported in previous studies. Although more detailed analysis is needed to explain the differences, they may be attributed

---

25 Titze (1994) reports that membranous length of a female grows by 0.4 mm per year from the infant length of 2 mm (p. 179).
to discrepancies in the type of speech analyzed, the sample of children, and other experimental differences such as cross-sectional data versus longitudinal data.

Figure 4.3
A Comparison of Mean $F_0$ for Speaking Between Selected Studies and the Present Study

- Both sexes
- Girls
- Boys
- Present study (all speech across languages)
- Present study (Chinese speech by Chinese bilingual children)
- Present study (English speech by English monolingual children)
Table 4.4

A Summary of Selected Studies of $F_0$ Characteristics of Preschool Children

<table>
<thead>
<tr>
<th>Study</th>
<th>Vocalization</th>
<th>Age</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>McGlone (1966)</td>
<td>1-3 word utterances</td>
<td>13-23 mos.</td>
<td>443</td>
<td>--</td>
</tr>
<tr>
<td>Robb &amp; Saxman (1985)</td>
<td>spontaneous speech</td>
<td>17-19 mos.</td>
<td>363</td>
<td>68</td>
</tr>
<tr>
<td>Robb &amp; Saxman (1985)</td>
<td>spontaneous speech</td>
<td>20-22 mos.</td>
<td>328</td>
<td>50</td>
</tr>
<tr>
<td>Robb &amp; Saxman (1985)</td>
<td>spontaneous speech</td>
<td>23-25 mos.</td>
<td>314</td>
<td>60</td>
</tr>
<tr>
<td>Eguchi &amp; Hirch (1969)</td>
<td>sentences</td>
<td>3 yrs.</td>
<td>298</td>
<td>31</td>
</tr>
<tr>
<td>Ferrand &amp; Bloom (1996)</td>
<td>spontaneous speech</td>
<td>3-4 yrs.</td>
<td>249$^a$</td>
<td>257$^b$</td>
</tr>
<tr>
<td>Eguchi &amp; Hirch (1969)</td>
<td>sentences</td>
<td>4 yrs.</td>
<td>286</td>
<td>21</td>
</tr>
<tr>
<td>Baken (1987)$^{26}$</td>
<td>spontaneous speech</td>
<td>4;6</td>
<td>252$^b$</td>
<td>--</td>
</tr>
<tr>
<td>Eguchi &amp; Hirch (1969)</td>
<td>sentences</td>
<td>5 yrs.</td>
<td>289</td>
<td>46</td>
</tr>
<tr>
<td>Hasek et al. (1980)</td>
<td>/a/</td>
<td>5 yrs.</td>
<td>258$^a$</td>
<td>248$^b$</td>
</tr>
<tr>
<td>Weinberg &amp; Bennett (1971)</td>
<td>spontaneous speech</td>
<td>5 yrs.</td>
<td>248$^a$</td>
<td>252$^b$</td>
</tr>
<tr>
<td>Baken (1987)</td>
<td>spontaneous speech</td>
<td>5;5</td>
<td>248$^a$</td>
<td>--</td>
</tr>
<tr>
<td>Ferrand &amp; Bloom (1996)</td>
<td>spontaneous speech</td>
<td>5-6 yrs.</td>
<td>243$^a$</td>
<td>256$^b$</td>
</tr>
<tr>
<td>Sorenson (1989)</td>
<td>spontaneous speech</td>
<td>6 yrs.</td>
<td>296$^a$</td>
<td>251$^b$</td>
</tr>
<tr>
<td>Eguchi &amp; Hirch (1969)</td>
<td>sentences</td>
<td>6 yrs.</td>
<td>271</td>
<td>28</td>
</tr>
<tr>
<td>Hasek et al. (1980)</td>
<td>/a/</td>
<td>6 yrs.</td>
<td>254$^a$</td>
<td>263$^b$</td>
</tr>
<tr>
<td>Vuorenkoski et al. (1978)</td>
<td>long vowel</td>
<td>6 yrs.</td>
<td>256$^a$</td>
<td>230$^b$</td>
</tr>
<tr>
<td>Baken (1987)</td>
<td>spontaneous speech</td>
<td>6;4/6;5</td>
<td>247$^a$</td>
<td>247$^b$</td>
</tr>
</tbody>
</table>

$^a$Utterances collected from girls.

$^b$Utterances collected from boys.

$^{26}$ Data cited in Baken (1987) compiling results from Weinberg and Zlatin (1970) and Weinberg and Bennett (1971).
4. Mean $F_0$ Results for Speech Utterances

Figure 4.4 illustrates the mean $F_0$ for speech utterances across the two languages. Chinese was consistently spoken with a higher mean $F_0$ than was English by the Chinese bilingual children. This appears to suggest a language difference in the pitch of speech and in basic auditory development; however, because of the small sample size, more information on individual differences including linguistic and physical development is needed to confirm this finding.

A further comparison between two different compilations of group mean $F_0$ for speech utterances (see Figures 4.4 and 4.5 yielded some interesting results. Since English speech utterances were collected from all subjects, a closer look at the individual children's $F_0$ results point to some unique vocal behaviours for the four bilingual children. Figure 4.5 compares the mean $F_0$ for speech uttered by English monolingual and Chinese bilingual children. It shows that Chinese bilingual children had an overall lower speaking $F_0$ when both English and Chinese speech data were compared. However, when English and Chinese speech utterances were compared separately (see Figure 4.4), higher group mean $F_0$ were reported for Chinese utterances. This suggests that although the Chinese bilingual children have an overall lower mean $F_0$ for speaking, their Chinese-speaking fundamental frequencies were higher than that of their English-speaking fundamental frequencies.

When the results from Figures 4.4 and 4.5 are considered together, some language differences for speaking $F_0$ are suggested. While the mean $F_0$ results for English speech are consistent with previous findings (refer to Table 4.4), English utterances were spoken with a lower mean $F_0$ by the Chinese bilingual children. It is hypothesized that the Chinese bilingual children in the study had acquired some form of learned behaviour in the selection of speaking $F_0$. Evidence of a cultural effect on speaking $F_0$ is reported by Moore and Kemp (1991), who found that English, American and Australian children show different means and ranges for $F_0$ when speaking and singing in English.
Figure 4.4
Mean Fo for Speech Utterances Across Languages

- Chinese speech by Chinese bilingual children
- English speech by Chinese bilingual children
- English speech by English monolingual children

Figure 4.5
Mean Fo for Speaking and Singing Across Languages

- Chinese bilingual children (singing in both languages)
- Chinese bilingual children (speaking in both languages)
- English monolingual children (speaking)
- English monolingual children (singing)
Figure 4.6
Mean Fo for Singing of Each Subject

- Heidi (Mandarin/English)
- Jodi (Mandarin/English)
- Wendy (Cantonese/English)
- Vicki (Cantonese/English)
- Amber (English)
- Polly (English)
- Gina (English)
- Clare (English)
5. \(F_0\) Results for the Singing Responses

Because Chinese singing data collected during the third year decreased in amount, comparisons between the \(F_0\) characteristics of Chinese and English singing was limited (see Figure 4.1). Figure 4.5 shows that the English-speaking children had higher mean \(F_0\) for singing than did the bilingual children, except at the 2-3 level. However, differences in the means of the singing \(F_0\) between English and bilingual children became wider over time. The differences grew from 9 Hertz at the 2-3 level to 17 Hertz at the 5-6 level. Nevertheless, it is difficult to draw any reliable inferences because of the small sample size and the ethnographic design of the present study. Since no conclusive pattern emerged from the means of the singing \(F_0\) among subjects (Figure 4.6), further analysis of their singing responses using qualitative evaluations is warranted.

6. Differences Between Speaking and Singing Mean \(F_0\)

To investigate possible language differences in using \(F_0\) characteristics to differentiate between speaking and singing, mean \(F_0\) results were compared between English and Chinese-speaking subjects (see Figure 4.7). An increase in the difference between speaking and singing mean \(F_0\) across time was found in the English-speaking subjects. The difference between speaking and singing mean \(F_0\) increased from 16.6 Hertz at the 2-3 level to 74 Hertz at the 5-6 level (or 5\% of differences increasing to 29\%). This increased difference between speaking and singing mean \(F_0\) might suggest that the English-speaking subjects selected a different pitch centre for speaking and singing. It could imply that as the English-speaking subjects matured, they differentiated speaking and singing as two distinctive types of vocalizations with different pitch characteristics. Hence, the mean \(F_0\) for both speaking and singing became more clearly marked as these English-speaking subjects grew older.

For Chinese-speaking subjects, a relatively small decrease in the differences between speaking and singing mean \(F_0\) was observed over time. The difference between
their speaking and singing mean $F_0$ for the subjects at the 2-3 level was 47 Hertz, gradually decreasing to 31 Hertz at the 5-6 level (or 15% of differences decreasing to 11%). Evidence suggests that the vocal behaviours of these Chinese-speaking subjects might be different from the English-speaking subjects.

This language-specific use of $F_0$ for musical and linguistic communication found in subjects in the present study supports research findings concerning the effects of ambient language (the linguistic environment) on infant vocalizations. de Boysson-Bardies, Sagart, and Durand (1984) report that prosodic features of infant vocalizations display influences from the adult language. Whalen, Levitt, and Wang (1991) report that reduplicated babbling of French and English children display differences in intonation pattern corresponding to adult models. The suspected language differences in vocal behaviour, among English and Chinese-speaking subjects in the present study, warrant further investigation.

Figure 4.7
Differences Between Mean $F_0$ for Speaking and Singing Across Languages

![Graph showing differences between mean $F_0$ for speaking and singing across languages for Chinese bilingual and English monolingual children at different age levels. The graph shows a decrease in $F_0$ from 47 Hz at 2-3 years to 31 Hz at 5-6 years for Chinese bilingual children, and a similar trend for English monolingual children.]
Analyses of Vocal Development: All Subjects

One and a Half Level

General Speech Development of the Child

Among the children who participated in the present study, Heidi (Mandarin/English)\textsuperscript{27} was the youngest, and incidentally, the only child at this age level. (Refer to Table 4.1 for the age of the subjects at the beginning of the data collection period.) All analyses discussed at this age level were based on observations of Heidi.

Heidi came to Canada from Taiwan when she was 14 months. Since both of her parents spoke Mandarin at home, she had no knowledge of English when she began to attend daycare at 15 months. I first visited Heidi when she was 18 months. Although Heidi appeared to be a quiet girl in the presence of strangers, her mother described her as a curious and talkative girl who frequently asked questions and made subjective comments. At 18 months, Heidi was able to express herself quite effectively by chaining several Mandarin words together. Like the *speech utterances* of most children at this age, some of Heidi’s Mandarin utterances resembled private speech, which were intelligible only to her parents.

During my first visit, Heidi was reluctant to communicate with me directly. Since Heidi knew me from a previous music program that I offered, I speculate that my role as a researcher, as well as the recording instrument I brought with me affected Heidi’s initial response. Hoping to establish rapport with Heidi, I spoke Mandarin during my visit. Nevertheless, Heidi primarily responded to me indirectly through speaking and singing to her parents. Based on parental input and previous contact I had with Heidi, the timbre of her voice during that visit was considered peculiar. For most of the time, it may be described as a breathy, guttural sound. Although the parents tried to encourage Heidi to

\textsuperscript{27} A bilingual (Mandarin/English) speaking child.
open her mouth wider by uttering the sound “ah”, Heidi was reluctant to respond. From my perspective, my visit might have caused some anxiety for Heidi and impeded her vocal production. Curran and Cratty (1978) stated the following:

Any kind of intricate muscular coordination is impeded by an emotional environment marked by periodic or sustained stress. Problems in rhythm (stuttering), articulation (speaking clearly), and vocabulary development may be caused by the adverse emotional content of the situation in which the child finds himself. (p. 26)

Both parents reported that they had never heard Heidi use her voice in that way. Since they were not aware that Heidi had any abnormalities with her throat during that time, they both speculated that Heidi’s strange vocal behaviour was a deliberate act reflecting her reluctance to cooperate with me. From her parents’ perspective, Heidi was playing tricks on me because she understood that I came to collect samples of her voice, but she chose not to respond with socially appropriate speech.

Figure 4.8 is a spectrogram of the vocal sequence collected at the beginning of the recording session when Heidi was reluctant to talk to me. Heidi was shown a colour photo of a baby. Her father asked, “Who is this?” Heidi responded with some brief vocal utterances. The most notable feature is the noise component seen as dark patches at 1500 - 5000 Hz caused by irregular vocal fold vibrations, which produced aperiodic energy. A pitch extraction performed on this voice sample yielded a $F_0$ with a range of 289-347 Hz, with a mean at 313 Hz and a standard deviation of 20 Hz. The formant and harmonic structures of this waveform are much more difficult to detect visually when compared to a typical vocalic pattern. (The most obvious harmonic structure is the third sound, an /i/ with $F_0$ around 300 Hz and the second harmonic at around 600 Hz.) The lack of detectable formant and harmonic structures made Heidi’s speech unintelligible. At times, even her parents experienced difficulty understanding her speech. An FFT shown above the
Spectrogram in Figure 4.8 illustrates that most of the energy lies within the range of 2800 to 6000 Hz, with very little energy distributed at the $F_0$.

Figure 4.8

A Spectrogram of a Vocal Sequence of Heidi Collected When She Was Reluctant to Communicate

(An unintelligible sentence)

(It's me!)
Figure 4.9 shows Heidi uttering “aah-ap-ple” when she saw a photo of apples. (Her mother confirmed that Heidi was attempting to say the word “apple”.) The acoustic features are similar to those in Figure 4.7. Evidence suggests that at 18 months, Heidi appeared to be in the process of fine tuning her vocal control.

Contrary to the vocal data collected earlier at the same home visit, some normal speech utterances in Mandarin were much more intelligible. This may indicate that Heidi had “warmed up”, and appeared to be less inhibited by my presence. The most interesting speech utterances were collected from Heidi when she was playing with a toy phone.

Heidi had been brought up by her grandmother whom she missed her very much following her family’s move to Canada. This speech sequence was emotional, since Heidi was pretending to call her grandmother in Taiwan to tell her how much she missed her. Heidi’s Mandarin speech in this excerpt was very clear and the earlier roughness in her voice disappeared (Figure 4.10). Compared to Heidi’s other speech utterances, the FFT
of this speech sequence shows clear energy peaks at the first 3 harmonics, and the
spectrogram shows clear harmonic structures up to the 9th harmonic. Figure 4.10 is a
highly emotional speech sequence which came at the end of Heidi’s phone conversation to
her grandmother. Acoustically, these perceptual qualities are reflected in the higher $F_0$
mean (387 Hz), and the number of harmonics detected under 5000 Hz (13 harmonics).

Figure 4.10

Heidi Saying Good-bye to Her Grandmother

(Hua Hua miss - ed grand - ma)

(Grand-ma ah! Grand-ma ah! Bye - bye!)
Singing

Heidi produced fewer singing responses at eighteen months than in later ages. During my visit, Heidi did not respond with enthusiasm to my request for her to sing, although according to her mother, Heidi had acquired a repertoire of English-speaking children's song from the daycare and enjoyed singing spontaneously when she was playing alone. Heidi's mother also sang with her daughter at home, and taught her Taiwanese nursery rhymes. During my visit, Heidi occasionally sang along with her mother but she typically sang softly and briefly. If her mother paused or sang softer, Heidi would stop. As a result, singing data available for analysis were limited.

The typical singing utterances collected from Heidi at this age consisted of short melodic fragments of a few pitches with speech syllables uttered in a "babtalk" manner (Curran & Cratty, 1978). Heidi would omit or substitute sounds, especially consonants that are often considered more difficult for young children to articulate. Although melodic fragments sung by Heidi consisted of short and repetitive sounds, they carried a clear melodic contour. From observing her mother's prompting, it would appear that Heidi was singing some learned songs. However, the judges were not able to identify the learned songs because Heidi's singing consisted of only short snatches of unintelligible words in which the rhythmic and melodic patterns were too short for identification.\footnote{28}

Another type of response was recorded when Heidi sang while playing the piano. Since Heidi did not have any training, she spontaneously pressed her fingers on the keyboard to pretend that she was playing piano. With a song book in front of her, Heidi sang as she played. For most of the time, the singing was not synchronized to the pitches produced by the piano. This spontaneous song was made up of repeated fragments of

\footnote{28 The judges' ability to identify familiar songs affected the classification of vocalizations into either learned songs or spontaneous songs, notwithstanding the possibility that a child was in fact attempting to model a song unfamiliar to all judges. In case of doubt, the parents' perspectives were used as a more insightful judgment.}
“Mommy, mommy, . . . I love you” sung in Mandarin, which were separated by long pauses of different duration. It appeared that at times, Heidi focused more on playing piano, while at other times, she focused more on singing. During the singing and piano playing, however, Heidi always looked at the piano and appeared to be listening attentively while she was searching for the notes on the keyboard and was busy coordinating her finger movements.

Heidi’s singing consisted mostly of snatches of two to four long tones and were performed briefly along with the piano playing. The piano playing, on the other hand, conveyed better sense of continuity because it consisted of successions of clustered notes played simultaneously by both hands. Since those clusters of notes and the words sung were the only two recurring elements in Heidi’s spontaneous tune, they were perceived by one of the judges as melodic ideas. However, it is questionable that these elements were evidence of a deliberate creation of a “musical” idea on Heidi’s part. The perception of some structural design may well be a construct of the judge due to her training in Western music.

Heidi demonstrated the use of a variety of high and low pitches, as well as long and short notes in this song. Her spontaneous music making was free from the preconception of tonality or musical phrasing that the judges were looking for. Hence, the spontaneous explorations were speculated by the judges to be experiences modelled after her mother playing songs on the piano. Heidi’s piano song could be considered as a form of musical babbling (Moog, 1976), in which she used her voice to practice creating a rich variety of pitches in different length, texture, and timbre. Evidence from Heidi’s piano song also suggests that at 18 months, she might possess fine auditory discrimination skills, which allowed her to produce distinctive tones having contrasting acoustic properties.
Intermediate vocalizations

A type of vocal interplay was collected when Heidi’s mother asked her to “sing” Mandarin “songs”. On the one hand, Heidi’s mother called them songs, but on the other hand, the judges concurred that this type of vocal response performed by Heidi and her mother resembled chanting in a call and response (or turn-taking) style. Hence, combining those two different descriptors used by Heidi’s mother and the judges, I called it a “chanting song”.

The performance of this “chanting song” sequence resembled a common practice in which a mother fills in the important interlocking clause or words in a sentence while assisting the child to practice producing a speech sequence. Figure 4.11 shows Heidi “singing” a Taiwanese nursery rhyme. Not only the mean $F_0$ of the vocal response corresponds to Heidi’s speaking voice (312 Hz, SD = 19 Hz), it has a narrow range (289 to 345 Hz). Hence, I conclude that speech elements predominated in this type of vocal interplay.

Figure 4.11
Heidi “Singing” a Taiwanese Nursery Rhyme

小皮球; 香蕉油
Lit - tle ball; bana - na oil
The discrepancy between the mother’s and the judges’ conception of singing was interesting: what was repeatedly labelled as “singing” by Heidi’s mother was identified as a *speech utterance* by the judges. Although contextually, both Heidi and her mother were “singing” because they both agreed that this type of vocal behaviour was “singing”, the vocal responses exhibited general pitch characteristics of speech, and were therefore, aurally perceived by two of the judges as speaking. Nevertheless, judges described it as a form of speaking with “a degree of musical interest, animated with contrasts in pitch level and duration”; “a form of chanting... there is noticeable patterning in the sounds which resembles a poetic form... For instance, the duration of the sounds follows the basic design of short-short-long”. In other words, “singing” and “chanting” appeared to be used synonymously by different individuals to signify the same vocal phenomenon. I shall discuss how “chanting song”, as a form of intermediate vocalization, developed into what I called “prosody of poem” when Heidi was 30 months.

**Two to Three Level**

**General Speech Development of the Children**

Data were collected from seven children. Among these children, four were bilingual: Heidi and Jodi (Mandarin/English); Vicki and Wendy (Cantonese/English). Three other children, Amber, Gina, and Clare, spoke English. Compared with other age levels, vocal responses collected at the 2-3 level exhibited the widest disparity. Of the data collected, there were noticeable differences between the children’s English verbal competence, as well as the amount of English utterances and songs collected. This can be explained by the language and cultural challenges faced by the four Chinese bilingual children. I shall explain the nature of this challenge and how it has been considered in the interpretation of results.

Although English utterances were collected from all seven children, their ability to speak English varied. Even among the four bilingual children, their level of English
competence varied a great deal. Wendy (C/E) and Vicki (C/E) had been in Canada for a few months when the present study commenced; they had just begun to learn English when I first visited their homes to collect data. Their spoken English was limited to a small collection of single words and some occasional 2 to 3 word sentences. Most of the speech utterances collected from Wendy and Vicki were, therefore, in Chinese.

Two other Chinese bilingual children, Heidi (M/E) and Jodi (M/E), had more advanced English skills at this age level when compared with Wendy (C/E) and Vicki (C/E). However, Heidi and Jodi also had to face the challenge of growing up in a bilingual environment because they spoke English during the day at daycare, but spoke Mandarin at home with their parents. Heidi’s mother reported that when her daughter was 24-30 months, she did not view Mandarin as an effective means of communication, and was therefore reluctant to speak Mandarin at home. On the other hand, Jodi’s parents tried to encourage their daughter to speak both English and Mandarin at home to provide a bilingual environment, her father spoke English and the mother spoke Mandarin. However, I observed that Mandarin appeared to be the language of major use at Jodi’s home because the parents communicated in Mandarin between themselves.

Three children at this age level, Clare, Amber, and Gina, were native English speakers. Clare was considered by her parents and daycare workers to be more advanced in verbal skills than other children of her age. During my home visits, Clare avoided establishing direct eye contact with me, and clung to her mother most of the time. Although she appeared to be shy and quiet, she could express herself very clearly in well-structured sentences with correct grammar. Clare was also the child who correctly labelled the greatest number of objects in the two picture books I used to elicit single word utterances.

Amber (E) was a lively girl who expressed curiosity about people and things in her environment. During my visits, she asked questions about me, showed me her toys, and
initiated games and activities. At the 2-3 level, I noticed that Amber still used immature articulations. As a result, omissions and substitutions of specific sounds were common in Amber’s speech. However, children’s sound systems are typically immature and incomplete at this stage of language development (Stoel-Gammon & Dunn, 1985). I noticed that Amber’s mother attempted to assist her daughter to master appropriate articulations by repeating the word immediately in her conversations with Amber, thus providing Amber with a model articulation without pinpointing her mistakes. Hence, Amber enjoyed excellent emotional support from her mother during the period that her speech articulation was maturing.

Gina (E) was a playful girl who responded with enthusiasm to the various stimuli initiated by her mother and me. During my visits while she was at the 2-3 level, Gina appeared to be more content to respond to questions and activities initiated by adults rather than to take the lead in conversations and activities. Gina enjoyed counting, labelling, and sorting objects. However, she needed a lot of help when she was trying to recite the letters of the alphabet in their correct order.

In terms of the quality of children’s speech production, none of the seven children at the 2-3 level exhibited the interesting vocal behaviour I reported for Heidi (M/E) at 18 months. Instances of diplophonia (subharmonics) were found in some of the children’s speech and singing utterances, a phonatory characteristic commonly found in infants and young children, which is described by Kent and Murray (1982/1991) as follows:

This characteristic is an abrupt change in harmonic structure, sometimes appearing as a doubling of the harmonic pattern (that is, the insertion of a 1/2 $F_0$ harmonic series) and sometimes appearing as an altered harmonic pattern bearing no simple relationship (especially halving or doubling) with the neighboring harmonic pattern(s). (p. 408)

The appearance of diplophonia is often associated with a perception of vocal roughness or harshness, which arises from a modulation of the vibratory pattern of the
vocal folds (Kent and Murray, 1982/1991; Omori et al., 1997). Figures 4.12 and 4.13 are narrow-band spectrograms of Amber's singing. They illustrate abrupt changes in harmonic pattern (filled arrowheads). Figure 4.12 shows some *spontaneous songs* collected when Amber was bouncing around in circles. The unstable subglottic pressure could have resulted from her body movements, which caused some notch-like indentation in the source spectrum display. Figure 4.13 shows the utterances collected when Amber was singing a *learned song* while she was sitting still on a couch. Hence, the diplophonia found in this singing sequence could not have been caused by body movement. It suggests that some children at this age level still exhibit immature control of their vocal folds.

Figure 4.12

Amber's Spontaneous Singing When Bouncing Around in Circles

```
All a - round, all a - round
```
Of special interest were the Cantonese speech utterances collected from Vicki at age 30 months. She had a high, bright voice and was able to express herself clearly in Cantonese at the time data were collected. A speech sequence collected when Vicki was reading a storybook, is shown in Figure 4.14. In this sequence she described how a naughty baby took his shoes off. A complex contour of the speech melody and a comparatively wide $F_0$ range is noted in this speech sequence. Another pitch extraction (Figure 4.15) performed on a conversation with her mother showed that Vicki's Cantonese speech exhibited some characteristics of singing: distinctive melodic contour was evident over a sentence, resembling a speech melody, and her choice of speech intervals was wide, resembling those usually found only in singing utterances. During our conversation, she noticed a mosquito and nervously said, "(the mosquito) flew in from the window". This
speech sequence had an overall descending contour, that reflected Vicki’s emotional state: she was afraid the mosquito might bite her.

Although there were noticeable features in both speech sequences that might be said to resemble characteristics of singing, all judges perceived these sequences as speech. Since all judges understand Mandarin and Cantonese, it is possible that they were following the context of the speech. Although I used judges who could speak English, Mandarin, and Cantonese in order that they might better understand the context of the vocal communication, it would be of interest to determine whether non-Chinese-speaking judges would perceive these same vocalizations as speaking or singing.

Figure 4.14
A Speech Sequence of Vicki Reading a Storybook

(Took his shoes off, ba - by took his shoes off.)
Singing

English *singing responses* were collected from all seven children, and Chinese *singing responses* were collected from all four bilingual children. *Singing responses* collected at this age level were in a variety of configurations. For example, there were many examples of *learned songs, spontaneous songs,* “vocal babbling”, and “pot-pourri” songs. Whereas *learned songs* were performed by all children at this age level, differences in the extent of their song repertoires were observed. Although Wendy (C/E) and Vicki (C/E) had limited English-speaking skills, both had acquired some simple children’s songs in English. Heidi (M/E) and Vicki had a larger Chinese song repertoire because their mothers had made singing part of their children’s daily routine. (Both Heidi’s and Vicki’s mothers possessed elementary piano skills. Heidi had an old, untuned piano in her bedroom which she enjoyed playing spontaneously.)
1. *Learned songs*

At the 2-3 level, a large part of the children's singing was modelled after songs previously introduced to them. In most cases, parents explicitly taught these songs to their children. Sometimes, songs were also acquired by a child through repeated listening to recordings of children songs.

Although parents were able to identify the *learned songs* sung by their children, judges often experienced difficulties in correctly identifying songs sung by a child. This difficulty as identified may be explained by the characteristics of *learned songs* performed by these children; that is, the songs were seldom rendered in their complete form--omission of parts were frequent, as if only excerpts of the *learned songs* were being sung. The words of the songs were also simplified; some key words of the song remained intact but other words were freely replaced with nonsense syllables or words that appeared to fit the sound. Frequently, only short segments of melody were intact, while other parts of the song might be difficult to recognized because the melodic contour, the intervals, and the rhythmic patterns were changed. Although children often said that they had forgotten the song, it appeared that they had forgotten the words but not the overall musical structure of the song.

The relationships between performance of words and melody were particularly interesting. When a child said that she had forgotten how to sing a song, she was often able to sing the *learned song* accurately after I prompted her with the words. However, a child often failed to “remember” a song even after I prompted her with the melody alone with monosyllables such as “do”, or simply hummed the melody. This suggests that the words (or specifically, the sounds that made up the words because they could understand very little about the meaning of the words), appeared to operate as an “anchor” that secured the musical elements of a song. Hence, when words were forgotten, some children considered the song as a whole was forgotten. Nevertheless, some children at the 2-3 level
were able to render a *learned song* without perfect recall of all the words. For example, when Wendy (C/E) performed a *learned song*, she managed to accurately render all the Cantonese words in the first verse, but failed to sing the melody accurately (see Figure ). After singing the first verse, she paused briefly before continuing with the second verse; that is, she appeared to have forgotten the latter part of the words in the second verse. However, without noticeable interruption, she substituted nonsense syllables, “da-dee-da-dee” for the forgotten portion, and continued to perform the song. Evidence suggests that although a child at the 2-3 level possesses the ability to perform a *learned song* without accurate recall of all the words, she may sometimes consider the entire song forgotten when only some of the words are forgotten.
Wendy Substituting Nonsense Syllables for Forgotten Words

Verse one of this *learned song* was sung to its original Cantonese words.

(走了大街; 穿过小巷)

(Walked through main streets; passed through narrow allies)
Verse two of this learned song was sung to "da dee da dee" to maintain the original melody. Notice the agreement in pitch accuracy between the two verses.
The word meanings of a *learned song* often appeared to be more attractive to a child when sung with body movement. For example, when Clare (E) was invited to sing an action song, *Teddy Bear*, she acted out all the body movements corresponding to the correct sequence of the song but did not attempt to sing the song at any time. Since the whole song was performed as if it was a "silent movie", the body movements were acted out in their correct sequence, possibly because Clare was mentally rehearsing the song from the beginning to the end. When Clare performed *Teddy Bear*, I also noticed that her movements were acted out in accordance with the appropriate tempo of the song. Hence, Clare did not simply remember the content of the song, she acted it out with body movements; that is, she "expressed" the song without actually singing it aloud.\(^{29}\)

The *learned song* at the 2-3 level was sometimes expressed through tactile experience. For example, just shortly after Clare began singing *Mary had a Little Lamb*, she stopped abruptly and walked away to get her toys. She appeared to be anxious and asked her mother to help her find “Mary” and “Mary’s lamb”. With the mother’s assistance, Clare found these two characters and returned to the sofa with her toys. She completed singing the first verse of the song accurately and then attempted to sing the second verse. However, she was soon absorbed in manipulating “Mary” and her “lamb” to act out the rest of the song and never completed singing the second verse. It therefore appears possible that for a young child, singing may not be an isolated form of experience that involves only vocal production; singing is sometimes integrated with, or even contrasted from other forms of expression such as body movement and imaginative play. This topic deserves further attention.

\(^{29}\) This is a complex example of dissociation, the ramifications of which are not explored in this dissertation.
2. *Spontaneous songs*

Examples of *spontaneous songs* by children at the 2-3 level were abundant. According to the parents, spontaneous singing occurred even more frequently than did *learned songs*. Parents reported that their children often sang spontaneously while in solitary play (i.e., playing alone without any companion). For example, Jodi’s mother reported that Jodi (M/E) might sing for up to one hour when sitting in the shopping cart. Heidi’s (M/E) mother reported that while she was working in the kitchen, her daughter frequently sang loudly for a long time, “as if she was performing to an audience”. Both Amber’s (E) and Clare’s (E) mothers reported that their children sang spontaneously to their toys while in solitary play.

*Spontaneous songs* by children at the 2-3 level took on a range of different configurations such as “pot-pourri songs” and “vocal babbling” as defined by Moog (1976). The children appeared to take pleasure in using their voices to manipulate learned materials and explore new resources for sounds. For a “pot-pourri” song, it was often difficult to differentiate whether a child was intentionally manipulating the musical materials from more than one song or if it was simply a case of confusion over two songs. This difficulty is illustrated in the following scenarios.

**Scenario 1:** At age 2 years; 3 months, Clare (E) sang the first phrase of *Bingo* but shifted to sing the first phrase of *Old McDonald* without a pause. After that, she hesitated for a while as though she had forgotten something or made a mistake. Clare paused for a moment but did not resume singing either of the two songs. It is possible that there was some confusion over the two songs, i.e., she ambiguously connected the final word “Oh”, from the first phrase of *Bingo* to the first word “Old”, as in *Old McDonald*. This interesting scenario may also support my earlier suggestion that words often serve as an anchor securing the musical elements of a *learned song*.
Scenario 2: At age 2 years; 8 months, Wendy (C/E) sang the first three phrases of *Happy Birthday* in English. She was later distracted from the words when pretending to blow out candles, consequently she did not sing the last phrase of the song. However, since all judges agreed that the first three phrases were sung in tune, it appeared that Wendy knew the song well. Wendy then continued to recite a Cantonese nursery rhyme and sang two Cantonese songs. While Wendy sang the second song, *It's a Small World* (a version with words translated into Cantonese), she paused after singing the first phrase when she noticed her father had just returned home. Wendy subsequently continued to sing but modified the words to fit into the melody of *Happy Birthday*. She then gradually returned to the melody of *It's a small World* and continued to sing the refrain. Thereafter, although the tonality was successfully maintained, fragments of the two songs were chained together. It is not clear whether Wendy was confused over the melodies of the two songs, whether she was intentionally mixing the two *learned songs*, or whether she had forgotten some of the words of both songs.

This “pot-pourri” song demonstrates how a child might combine melodic material from two *learned songs*. Since earlier in the same recording session, Wendy sang almost the entire *Happy Birthday* song accurately and in tune, it is unlikely that she was confusing the two songs. Because the two songs were sung in two different languages (*Happy Birthday* in English and *It's a small World* in Cantonese), I speculate that it is unlikely that Wendy confused the two songs as one.

In summary, inspirations for most of the *spontaneous songs* recorded from children at the 2-3 level appeared to stem from fragments of *learned songs*. Children usually began singing a *learned song* with repetition of a single syllable or nonsense words before they sang the rest of the song. Reference to the diatonic system and metrical rhythm would often loosen when *spontaneous song* began. The following scenario was recorded when Vicki (C/E) was 30 months:
Vicki sang the first two phrases of *Twinkle, Twinkle Little Star* in English. Although the singing was melodically accurate, Vicki sang it in a “joking mood”; some laughter could be heard along with her singing. This resulted in slightly staccato singing, which moved along in a spirited manner. At the third phrase, nonsense syllables emerged. These syllables gradually took over the fourth phrase, which was sung to the original melody in a manner similar to an augmentation (the time values of the original notes were doubled). (see Figure 4.17 for a transcription).

![A Transcription of Vicki’s Performance of *Twinkle, Twinkle Little Star*](image)

1. *Twin-kle Twin-kle lit-tle star, how I won-der what you are.*

3. *Up a-bout dee world so high, like a dia-mond i ah hoi,*

5. *Chee-co, chee-co li oo ah, oh I wa won-der what you are!*
Some of the spontaneous songs collected from children at the 2-3 level resembled speech babbling. At 30 months, Vicki's (C/E) spontaneous song began with a descending melodic line sung in repetitive single-syllable utterances, for example, “mew, mew”. Later, more single-syllable nonsense words such as “da da, yeah, yeah”, were used. Although Vicki spoke mainly Cantonese during the time that data were collected, her singing included unintelligible speech sounds that resembled babble. At the end of the spontaneous song, some clearer utterances resembling Cantonese speech emerged. Two of the judges transcribed it as “Baby, Baby ah!” spoken in Cantonese. Unlike most of the spontaneous song collected at this age level, the judges did not identify any reference to a learned song in this sequence. It is noteworthy that this form of vocalization was also collected from Vicki at age 3;6. A more detailed analysis can be found in the discussion of “babbling vocalise”.

Intermediate Vocalization

An interesting form of vocalization, “chanting song”, was reported earlier for Heidi (M/E) when she was 1;6. Although labelled by her mother as singing, it was perceived by the judges as chanting, a form of speech utterance. A similar vocal genre was collected again when Heidi was 25 months. As before, Heidi’s mother prompted her child to sing, but this time, Heidi performed two ancient Chinese poems, “Evening Thoughts” by Li Po (“夜思”李白), and “Spring Dreams” by Tao Fu (“春晓”杜甫).

Analyses of these two performances showed some characteristics of songs, such as a wide pitch range and metrical rhythmic patterns. However, the words were not sung to a typical Western European diatonic melody. The judges described the intervals produced by Heidi as “microtonal, speechlike, unstable, and not pertaining to the Western diatonic system”. The pitch ranges of these two performances were much wider, and the syllables were uttered much longer duration than when spoken in normal speech (see Figure 4.18). An acoustic analysis confirmed the perceptual evaluation of the judges.
yielded a vocal range of 450 Hz. The utterances were relatively longer than that found in a typical speech sequence and they had a distinctive wave-like contour that resembled characteristics of a melody. I called this kind of performance “prosody of poem” because the prosody of the words in those two performances took on a leading role and resulted in a form of heightened speech that displayed certain characteristics of singing.

Figure 4.18

A “Prosody of Poem” By Heidi
疑是地上霜  
(Looks like frost on ground.)
Head raised, musing at the bright moon.
低頭思念故鄉

(Head lowered, engrossed in homesickness!)
Since Heidi’s mother referred to these performances as “songs”, I also investigated whether the poems were the lyrics for a melody. To do so, I recorded Heidi’s mother performing the poems because she was Heidi’s vocal model. I asked her to produce the poems in the same manner she performed them for Heidi (i.e., a simulated version of the maternal song). The mother explained that she intentionally made adjustments to the poems so they would be more appealing to Heidi. Since Heidi was sitting next to her, I suggested that she ignored me and “sing” directly to Heidi. After recording her performance of poems for Heidi, she also allowed me to record her performing the poems in the way she would for an adult audience. A comparison between the two performances was therefore possible.

The two performances showed marked differences in pitch and syllable length. The version performed for Heidi was much more animated with exaggerated pitch contour and variations in syllable length. This form of vocalization was in accord with the characteristics of “maternal speech” or “infant-directed speech” reported by Papousek (1995) and Fernald (1994) (see discussion in Chapters one and two). When Heidi was between two- and two-and-a-half years, her mother attempted to encourage her to speak Mandarin by teaching her Chinese poems and nursery rhymes. Since Mandarin was spoken only at home, Heidi was reluctant to learn the language. “Prosody of poem” was a device used to assist Heidi to learn Mandarin, and it illustrates some of the inherent relations between speech and songs.

---

30 A series of studies on maternal singing were conducted by Trehub and her team of researchers. Their cross-cultural data suggested that when parents sang to their infants, they sang in a higher range and in a slower tempo (Trehub, Unyk, & Trainor, 1993b). Differences in pitch and tempo were also found between the simulated and the infant-directed versions (Trehub, Unyk, Kamenetsky, Hill, Trainor, Henderson, & Saraza, 1997).
Three to Four Level

General Speech Development of the Children

Data were collected from all eight children. Since all Chinese bilingual children attended daycare, English gradually became their language of major use. The English language environment outside of the Chinese bilingual children’s homes thus became a factor that influenced their language development at this age level. According to the Chinese parents, they attempted to provide stimulation at home to encourage their children to speak their native language. However, they all observed that their children were reluctant to communicate in Chinese at home.

Use of English at this age level also varied a great deal. When asked to name objects from two picture books, the English monolingual children successfully labelled almost all objects; Chinese bilingual children were capable of labelling only some objects in English. Certain social and environmental factors might have affected the Chinese bilingual children’s performance on this task. For example, fewer Chinese bilingual children successfully identified “cheese” when shown a photo. Because cheese is not found in traditional Chinese cuisine, it is probable that the Chinese bilingual children were not familiar with what a slice of cheese might look like, and the word “cheese” might not have been used in their homes.

Although the Chinese bilingual children could comprehend English when it was spoken to them within an appropriate accompanying context, in conversation their English production was limited to short sentences consisting of several English words. In addition, much of the Chinese bilingual children’s English speech was not well understood by adults other than their parents. The English monolingual children, on the other hand, also used short sentences but employed a wider range of words (such as nouns and pronouns), much more sophisticated sentence structure, and a variety of appropriate morphological endings. As a result, the English monolingual children expressed ideas more clearly. I observed that
the English and the Chinese parents reacted quite differently to the English-speaking ability of their children. Thus Chinese parents were more eager to assist their children to speak English. At times, my visits to collect data became opportunities for the Chinese parents to teach their children English. Occasionally, they would even interrupt a conversation to correct any words mispronounced by their children.

When asked to tell stories, Chinese bilingual children mostly labelled objects and characters with simple descriptors. They needed to rely on their parents to prompt them with the story line. The English monolingual children also needed assistance from their parents but were able to express the details of the story using comparatively richer contextual referents such as “here and now”, “and then”, and “after that”. The English monolingual children also demonstrated greater use of function in their immature sentences, such as demand and comment. Chinese bilingual children often used Chinese sentences to supplement their English story telling. Interestingly, when they used Chinese to supplement details in an English story, they either turned to their parents or spoke in a softer tone. This suggests that they were aware of their use of Chinese language to supplement their English, and perhaps reflected their attempt to seek confirmation for accuracy, from their parents.

Although the Chinese bilingual children spoke English at daycare, they all appeared to be relatively more capable of expressing their thoughts in Chinese, rather than English, at this age level. At that time, two of the Chinese bilingual children, Vicki (C/E) and Wendy (C/E), who had already lived in Canada for about a year, used relatively complete Chinese sentences to interact with adults. However, all Chinese bilingual children spoke mostly English when they were playing and when they were speaking to themselves. It was particularly noteworthy that the Chinese bilingual children always responded to English addressed to them in English; sometimes they also responded in English, when addressed in Chinese. There is then a complex picture of Chinese bilingual children’s
language development at this age. This issue is addressed in a number of studies of bilingual children (see, for example, a review of language contact phenomena in young bilingual children by de Houwer, 1995, pp. 244-248).

**Singing**

1. *Learned Songs*

   There was a marked increase in the number of *learned songs* collected from children at the 3-4 level. Some of the favourite *learned songs* collected from all children in this age level were: *Bingo; Teddy Bear; Twinkle, Twinkle, Little Star; Old McDonald; I Love You; Jingle Bells;* and *The Alphabet Song.* (See Appendix H for music scores for some of these songs.) Because English had become the language of major use for the Chinese bilingual children at this age, their *learned song* repertoire was obviously affected. Although the Chinese bilingual children’s English song repertoire was comparable to that of the English monolingual children at this age, there was a marked decrease in the singing of Chinese songs.

   When compared with data collected earlier, most children at this age level were able to sing phrases of *learned songs* with greater accuracy in melodic and rhythmic patterns. The general contour of the repetitive patterns in the songs, as well as most of the narrower melodic intervals were often sung correctly. Judges were therefore able to identify successfully almost all *learned songs* sung by the children, even when they were incomplete.

   Maintaining tonality of a song was a problem for all children. Even when able to demonstrate a sense of tonality at the beginning of a song they often changed the tonality several times within a song. In traditional Western music, most melodies contain a clear reference to a defined tonal centre. That children modulated during the course of singing a song even after they had established a tonal centre, suggests that they had yet to learn how to retain the primacy of an established key as a function of musical memory.
A comparison of the development of singing between the children was possible because they all sang a similar song repertoire. Among the children who participated in the study, Jodi (M/E) made least reference to the melodic contour of *learned songs* she sang. Almost all Jodi’s songs were sung in a chant-like manner, although the words and the rhythmic patterns were sung accurately. At the 3-4 level, Jodi did not yet demonstrate the same level of melodic accuracy in her singing as observed in the other children. Her speech was described by the judges as “whisper-like, breathy, not projecting very well, and without much variation in intonation”. Whether there were interactions between Jodi’s melodic contour in singing and her pitch variations in speaking is beyond the scope of the present study. Nonetheless, the vocal data showed some correspondence between the pitch characteristics of Jodi’s speech and her singing.

The words of *learned songs* appeared to be a challenge for children at the 3-4 level. They sometimes expressed frustration about forgotten words which made it difficult for them to complete a song. Even when a child could successfully recall a melody but not the words, she would frequently give up singing the song. At this age level, only occasionally would a child substitute forgotten words with nonsense words or syllables, and continue singing. Children at this age level often failed to produce the precise structure of *learned songs*, thus the order of verse and chorus was frequently confused. The present study did not explicitly investigate the cause of this confusion.

At age 3;4, Gina (E) wanted to sing *The Whale Song*. She began by humming some melodic fragments of the song. Unfortunately, she failed to recall any of the words and soon gave up singing the song. Similarly, at age 3;6, Wendy (C/E) sang the first phrase of a *learned song* but forgot the rest of the words. She then repeated the first phrase three times before she eventually gave up singing.

When a child claimed that she forgot the words of a *learned song*, melodic accuracy suffered. At 3;10, Gina (E) successfully sang the first verse of *Baby Beluga* in tune.
However, she appeared to have forgotten some of the words in the second verse, from which she sang fragments, then hummed the melody as she tried to recall the words she was missing. The humming later faded away as she tried harder to recall the words. Eventually, she chanted fragments of the words without attempting to sing the melody. It appeared that Gina had shifted her attention to recalling the words to such an extent that the melody could no longer be recalled accurately. This issue is addressed in studies on text-melody integration in melody recall (see, for example, Serafine, Crowder, & Repp, 1984; Serafine, Crowder, Davidson, & Repp, 1986).

2. Spontaneous Song

"Pot-pourri" songs were abundant among children's spontaneous singing. For example, at age 3;5, Heidi (M/E) sang *Jingle Bells* spontaneously but also attempted to fit in fragments of *Old McDonald*. Nevertheless, she managed to make her way back to *Jingle Bells* and rendered the song in an arrangement of verse-chorus-verse. In common practice, *Jingle Bells* is sung with the chorus section as the conclusion of the song. It appeared that Heidi deliberately ended the song with a verse section because she added a sprightly "Hey!" to conclude the song. As mentioned earlier in the discussion of "pot-pourri" song, it was often difficult to discriminate a "pot-pourri" song from a mix of two melodies. Heidi clearly demonstrated a thorough understanding of the structure of *Jingle Bells* because she was able to perform both the verse and the chorus sections and achieve a sense of unity and contrast. Hence, her attempt to integrate *Old McDonald* into *Jingle Bells* appeared to be her experiment with manipulating learned song materials.

*Spontaneous songs* at the 3-4 level showed a strong emphasis on the manipulation of musical elements such as rhythmic patterns, scalar structure, and melodic contour (ascending versus descending, and parallel verses contrary motion). *Spontaneous songs* at this age level might be seen as musical exercises, which appear to play an important role in assimilating culturally relevant musical practices.
Some game-like spontaneous songs were collected from Amber (E) when she was 3;1. About 20 minutes into the recording session, she showed me her toy phone and then shifted her attention to a toy xylophone. Using a mallet, she first played several notes and then sang "do-re-mi-fa-sol" while playing the xylophone. Although she sang the scale pattern one beat per syllable, her singing did not follow the pitches she played on the instrument; that is, the pitch movement of Amber's singing moved in a different direction from the scalic passage she was playing on the xylophone. From an adult's perspective, the choice of a parallel motion would appear to be the initial step in an attempt to match the singing with an instrument, but to play and sing scalic passages in contrary motion would appear to require much more sophisticated perceptual and performance skills.

After that initial attempt to sing while playing the toy xylophone, Amber then moved on to develop more rhythmic variety in her spontaneous song. At this point, the xylophone playing and the singing were no longer synchronized. The song was largely sung with nonsense words which included a repetitive phrase, "a la-dy Oh!"31, while the earlier ascending and descending scalic pattern recurred only in the xylophone part.

There was a brief diversion when Amber suddenly went to pick up her toy phone for an imaginative play. She talked to her "sister" on the phone, and said, "Hello sister, . . . I'll meet you at the ball game." Immediately after she hung up the phone, Amber returned to the spontaneous song she created just a short moment ago. The "a la-dy Oh!" motive was performed in a chant-like manner with strong accent and a clearly metrical pattern. Although the spontaneous song retained the "a la-dy Oh!" motive, it was no longer accompanied by the xylophone. The song then took on a playful character. Amber took my hands and led me in a dance while we swung our arms and skipped around in circular motion. At this point, she extended the repetitive "a la-dy Oh!" sequence by adding "and

31 Since *The Sound of Music* was one of Amber's favourite movies, it is speculated that the source of this "a la-dy Oh!" motive was from the song *The Lonely Goatherd*. In other words, Amber was possibly yodelling.
all fall down”. This game reminded me of the singing game *Ring Around the Rosy.*
(Amber was familiar with this singing game which she learned at daycare.) The chanting soon became faster and faster, and Amber even pretended to fall down onto the floor from time to time. Amber became more excited as she sang, and the game went on until Amber was exhausted. (Refer to Figure 4.19 for a spectrogram with musical transcription of Amber’s motive in her singing game.)

Two other children at this age level also performed body movements when they sang *learned songs.* However, no movement was recorded from these two children’s *spontaneous songs*; they performed only the specific actions they were taught to accompany the *learned songs.* Such types of song with corresponding actions are commonly called “action songs”. Hence, the above excerpt demonstrated Amber’s ability for assimilating different forms of musical experience into her own *spontaneous song,* such as improvising movement and playing a musical instrument.
Figure 4.19

Amber’s “a la-dy Oh!” Motive

A la-dy Oh, a a la-dy Oh!

A lay, Oh lay, a la-dy Oh, a la-dy Oh!
Intermediate Vocalizations

1. "Babbling vocalise"[^32]

When Vicki (C/E) was age 3;5, she had already spent almost a year in daycare, where she learned to communicate in English while speaking Cantonese at home. Data collected during this period saw a marked increase in spontaneous singing and speaking in both languages. At this time, it was interesting to note that Vicki’s vocalizations sometimes shifted freely between speech and singing modes. For example, when asked to sing *Twinkle, Twinkle, Little Star*, she first responded by reciting the words of the song in a heightened speech manner according to the rhythmic pattern of the song. She then switched to sing a short melodic fragment, then resumed in the heightened speech mode. Since her mother attempted to prompt her to recall the original melody, Vicki later gave up her rhythmic recitation and switched back to singing the rest of the original song.

In Vicki’s spontaneous vocalizations, short phrases of learned song were often sung in the middle of spontaneous speech. These were intermittent with passages that resembled vocal babbling of speech sounds. Due to the variety of vocal sounds observed, I called it “babbling vocalise”. Vicki’s “babbling vocalise” consisted mainly of: spontaneous singing of some learned song fragments and diatonic pitch sequences; spontaneous speech; and speech babbling of nonsense syllables and articulations of consonants. The following scenario documents this type of intermediate vocalization.

To initiate English speech, Vicki’s mother asked her daughter some questions in English. Vicki replied to the questions and said her name, address, and telephone number in English. While her production of vowels was mostly accurate, she made many mistakes in producing consonants. Therefore, although her speech was intelligible, the immature sounds resembled babytalk. Vicki then began a monologue, which is best described as a

[^32]: In the context of Western music, a vocalise is a composition for voice without words. See Randel, 1986, p. 925.
kaleidoscope of vocal forms using mainly nonsense words. The following utterances were identified by the judges according to their order of being heard: fragments of spontaneous singing using unintelligible words; a counting sequence in English; and Cantonese speech utterances that were intelligible but nonsense in content.

In an attempt to interrupt her daughter's monologue, Vicki's mother tried to initiate conversation in Cantonese by asking, "What is your name? What is your father's name?" However, Vicki was still absorbed in her monologue and responded with spontaneous singing. The singing sequence contained recognizable fragments of a learned song, Edelweiss, but was cut short by a short descending scalar pattern sung in monosyllables. Vicki then switched to counting in English. From then on, the vocal sequence became freer in style and form. Fragments of singing and speaking were interspersed with whispers and explosive vocal sounds. One of the judges described the vocal sounds as "a type of speech articulatory sounds". The other two judges described them as "vocal noise-making" and "an exploration of vocal sound".

Vicki's mother again attempted to interrupt Vicki's monologue and asked her to name her relatives. This time Vicki replied to her mother's questions by uttering only the first sound of several names, then sang nonsense melodic fragments using combinations of the sounds from her relative's names. Some Cantonese words as well as nonsense syllables were incorporated into these sung fragments. Since Vicki's mother was persistent in her attempt to elicit meaningful speech, Vicki eventually ended her monologue and answered her mother's questions in Cantonese.

Despite her mother's repeated attempts to interrupt, Vicki engaged in what I called "babbling vocalise" for over ten minutes. Constant efforts to stimulate meaningful word production appeared to have minimal effect on the child. In fact, the interruptions appeared to inspire Vicki to greater vocal babbling. Vicki appeared to be attracted by the acoustics of sounds: English or Chinese speech phonemes, diatonic pitch movement, learned song.
fragments, non-speech sounds, and words that were uttered in a regular, rhythmic pattern (e.g., counting). “Babbling vocalise” might be considered a comprehensive vocal exercise that captures the vocal, auditory, and cognitive skills of a child. Interestingly, some spontaneous vocalizations collected from Vicki at 30 months also displayed features similar to this “babbling vocalise” sequence.

2. “Drawing narrative”

At age 41 months, Heidi’s English pronunciations were still considered imprecise by the judges. Using immature sentences, however, she could express herself effectively in English. Heidi (M/E) had also acquired an expanded English song repertoire at preschool. Recognizing the challenge for Heidi to learn both Mandarin and English at the same time, her parents were less concerned about providing her with abundant Mandarin speech input. As a result, Heidi also spoke English at home, especially when she was playing. At age 3-4 level, Heidi’s favourite activity was drawing. She had an easel, newsprint paper, and crayons in her room, so that she could draw whenever she wanted.

After Heidi’s mother showed me some of her daughter’s drawings displayed on the walls, Heidi was enthusiastic and said she could draw some eggs for me. She walked over to the easel, picked up a crayon, and announced a title for her drawing. With a loud and clearly projected voice, Heidi enunciated “EGGS. There is a nest...” (See Figure 4.20 for a spectrogram of Heidi’s “drawing narrative”). She then drew an egg in the centre of the newsprint paper, and sang, “and inside the nest, there is an egg.” Heidi chose another crayon and drew a nest around the egg. She then sang the final two phrases of a learned song, The Green Grass Grew All Around.

Heidi moved on to draw a second egg, using another colour. As she drew, she said, “There was (were) two eggs, sitting in a nest”. Repeating what she had done before, she then sang the final two phrases of The Green Grass Grew All Around. The “drawing narrative” now took on a strophic form exemplified by a recurring pattern of organization in
Heidi's speaking and singing. Although Heidi spoke and sang as she drew, she appeared to be fully absorbed in the drawing activity. Hence, speaking and singing were observed to be fully integrated during the "drawing narrative".

More detailed descriptions of the activities observed in Heidi's "drawing narrative" may further illustrate how speaking and singing were integrated. (1) Heidi was constantly making choices about what she wanted to draw, and how she would achieve her objectives. Not only did she pause for a while in between the verses to painstakingly choose different colours for her eggs, she expressed frustration when she could not find the specific colour she wanted, and asked her mother for assistance. She complained that she could not find the right hue among several crayons within the same colour range, and even asked for a felt tip pen in order to get the exact colour she had in mind. (2) Heidi was telling a story as she drew. She paused occasionally to look at her mother and me, as if she was trying to establish some eye contact with her audience. Heidi also raised the volume of her voice and spoke slowly in a clear manner. Heidi's mother explained that her daughter frequently made up stories for her drawings. Even when Heidi was drawing alone in the room, she would raise her voice and tell a story as she drew. (3) Heidi was singing as she drew. It is interesting to see how a 41 month child adapted a learned song, *The Green Grass Grew All Around*, as the basis for her "drawing narrative". Heidi used only the beginning and the ending phrases of the song, omitting the more complicated middle section. To simplify even further, Heidi spoke the first phrase but sang the ending. Each verse was given a title sentence, "There was (were) ___ eggs", and followed the same recurring pattern. Heidi ended the "drawing narrative" after she drew five eggs and could not find any room for more eggs in the nest.
Figure 4.20

A Spectrogram of Heidi’s “Drawing Narrative”

EGG - S. There is a nest.
(Spoken)

And inside the nest there is an egg.
And the green grass grow all around and around,

(Spoken) green grass grow all around and around,
and the green grass grow all around.

There was (were) FIVE EGGS sitting in a nest.

(Spoken)
Four to Five Level

General Speech Development of the Children

Data were collected from all children, except for Heidi (M/E) as noted earlier. All five children at this age level used English as their language of communication, but Wendy (C/E), Vicki (C/E), and Jodi (M/E) also spoke Chinese at home. As expected, English-speaking children were able to communicate in grammatically correct sentences, whereas the Chinese bilingual children still used immature sentences. The differences in ability to use English were particularly evident when a child was telling a story. English monolingual children were able to elaborate on details in the stories, whereas (except for Vicki) the Chinese bilingual children conveyed only the general plots of the stories using simple sentences. Vicki was able to provide astonishing details of stories in English with the aid of pictures. Her parents were unable to explain why Vicki could recall all details of the stories even from books that were borrowed from the public library. Although the parents had only read these stories to Vicki a few times, she was able to retell the stories with almost the exact wordings in which they were written by simply looking at the pictures. Since Vicki could not yet read English, the parents were proud but amazed at her ability.

The Chinese bilingual children showed greater ability in communicating in Chinese than in English. They were able to express complex thoughts in Chinese using grammatically correct sentences. However, at this time, a language shift began to emerge. The Chinese bilingual children now had a smaller repertoire of Chinese songs, and their spontaneous songs were almost always sung in English. When asked to sing Chinese songs, they typically replied that they had forgotten the songs, including songs that had been their favourites when they were younger.

Interestingly, although the Chinese bilingual children were able to communicate effectively in their first spoken language, according to their parents their children’s English
reading skills were more advanced than their Chinese reading skills, probably the result of exposure to English in the preschool environment. The Chinese parents complained about a lack of Chinese reading resources for their children. Although they searched the local public library and book stores, Chinese books for children were not in abundance. Thus, although the Chinese bilingual children used Chinese to communicate at home, they made much less progress in learning written Chinese. Since the written form of Chinese is different from spoken Cantonese (Cantonese is a dialect), I found that Wendy and Vicki (whose first language is Cantonese) were unable to tell a story in Chinese by reading a picture book. They tried to make adjustments in Cantonese in order to align their speech with written Chinese, but they then confused grammatical rules and word usage. As a result, they were unable to reproduce stories orally in the same manner as read by their parents.

**Singing**

1. **Song Presentation**

   The children in this age level had a large repertoire of English *learned songs*. In particular, three children, Vicki (C/E), Amber (E), and Polly (E), could perform song after song for me with minimum stimulation. They frequently told me what they wanted to sing, whereas when they were younger, the parents and I had to prompt them to sing. In some of the recording sessions, I was able to collect more than 20 songs from these children. Also, most children at this age were able to sing the *learned songs* in tune without prompting from their parents. In some instances, a child might even refuse her parent’s prompting—she would prefer to start over, insisting on singing the song on her own. At the 4-5 level, a child was also more likely to sing a complete song. Should a song have more than one verse, a child often wanted to sing all verses, as long as she was not distracted.
2. Melodic Singing Accuracy

There was a marked improvement in the accuracy of each child's melodic singing as they grew older. Not only could each child preserve the general melodic contours, she was also beginning to demonstrate a better sense of tonality. For example, Vicki (C/E) demonstrated the ability to transpose the melody of a learned song. At 51 months, she could perform *Row, Row, Row Your Boat* in tune. However, after she sang the first verse, her father teased her, commenting that it sounded very high in pitch. Vicki then sang the song in a lower pitch range but still maintained correct intonation. It was only when she reached the final phrase that she gradually modulated her voice back to the high pitch range used at the beginning of the song.

Although most children in the present study were able to maintain tonality, the judges reported that some children modulated their tonality from verse to verse, or even from phrase to phrase. It was also observed that when both melodic contour and tonality were maintained, the main problem in melodic singing was that pitches were sung lower than their target frequencies, in other words, they were "flat". This problem was particularly salient when Jenny (M/E) sang pitches in the upper range because she had a comparatively lower speaking and singing voice. For other children in this age level, the final phrase of a learned song was also more likely to be sung flat.

A child often made fine adjustments to the melodies of learned songs, which appears to demonstrate that she could modify fragments of an original song into a more idiomatic melody. For example, when Jenny (M/E) was age 4;5, she substituted the melodic pattern "r-r-d" into "r-s₁-d" in two different learned songs (*The Alphabet Song* and *Baa Baa Black Sheep*). The latter pattern, "r-s₁-d", appears to require a more advanced

---

33 That is, instead of singing two repeated pitches and then moving a step downward to the tonic, she sang two wide leaps in a row, descending a perfect fifth downward, then up a perfect fourth to the tonic. (See Figure 4.21 for a comparison of the two melodic patterns.)
singing skill, and is more idiomatic because it has more interesting melodic turns and more sophisticated implied harmonies.

Figure 4.21

A Modified Melodic Pattern in Jenny's Song

The process of modifying repetitive melodic patterns into more musically sophisticated patterns was also observed in Polly (E) at 4;6 when she sang *Rain, Rain, Go Away*. Instead of singing the repetitive melodic pattern, she sang the second half of the song a major second below, hence enriching the rather repetitive melody with a sequential treatment. (See Figure 4.22 for a transcription of Polly's modified melodic pattern.)

Figure 4.22

A Modified Repetitious Melodic Pattern in Polly's Song

(First phrase sung as original melody.)

(Rather than repeating the first phrase, the second phrase is sung in sequence a 2nd below.)
It was equally interesting to find that two children produced the same modification to *Teddy Bear* which has the same repetitive melodic pattern for the beginning of all four phrases. Amber at 4;4 and Polly at 5;2 both modified “s-s-m” to “f-f-r” in the second and the fourth phrases. The sequential treatment of a major second down from the original melody enhances the melodic interest of *Teddy Bear*, giving it the needed contrast but retaining the overall symmetrical design of the song. (See Figure 4.23 for a transcription of *Teddy Bear* performed by Amber and Polly.)

Figure 4.23

A Modified *Teddy Bear* Song Performed by Amber and Polly

![First phrase, sung as original melody.](image)

Ted- dy bear, ted- dy bear, turn a-
round,

![Second phrase, modified.](image)

Ted- dy bear, ted- dy bear, touch the ground.

(Rather than repeating the first phrase, the second phrase is sung in sequence a 2nd below.)

Since I introduced all four songs (*The Alphabet Song*; *Baa Baa Black Sheep*; *Rain, Rain, Go Away*; and *Teddy Bear*) to the children at age 2-3 level before the present study began, I have first hand information on the exact melody that I taught to them. The present findings pose an interesting question: What kind of musical material could be used to maximize young children’s learning? It is commonly suggested by music educators that songs containing repetitive melodic patterns are more appropriate for young children because of their limited cognitive abilities. It is possible, however, that repetitive melodies
impede young children’s exposure to quality songs that are musically stimulating. Observations from the present study suggest that the 4- and 5-year-olds have the capacity to learn songs with more melodic contrasts because they spontaneously enrich the melodies according to culturally appropriate musical practices. Analyses of the learned songs sung by children in the present study suggest that they out-perform the level of musical skills commonly prescribed in books on music methods.

3. Words and Melody Recall

Children at the 4-5 level still had difficulties in remembering all of the words in songs. However, this problem did not appear to have much effect on the accuracy with which they could recall melodies.34 For example, at age 4;0, Gina (E) hummed the melody of a learned song when she forgot the words. The melody was in tune and the song was presented in its complete form. Wendy (C/E) at 4;7 also hummed a complete song because she claimed to have forgotten the words. Occasionally, a child would make up new words that fit well with the original meaning of the song. It is suggested that for children at this age, words and melody were not clearly delineated. Thus, most children could continue singing even when some of the words were forgotten; that is, children at this age level recalled a learned song in a more holistic manner.

4. Musical Interpretations

Some children demonstrated an awareness of tempo and dynamics by spontaneously manipulating these two musical elements in their singing. Such musical interpretations were more often found in learned songs that were in strophic form. Children at this age level often sang the first verse using a moderate tempo with little variations in dynamics. However, noticeable variations in tempo and dynamics were found in subsequent verses. Not only could a child sing with contrasts in tempo (fast/slow) and

34 As noted at the 2-3 level, a child often had difficulties completing a learned song when words were forgotten, and when nonsense syllables were used in a learned song, it often emerged into a spontaneous song.
dynamics (loud/soft), she also demonstrated refined control in singing these contrasts in a gradual fashion (i.e., gradually becoming louder or softer, and gradually becoming faster or slower). It appeared that children at the 4-5 level were able to manipulate different musical properties in order to achieve variety in recurring musical patterns. These changes could have also come about because of a child’s ability to interpret words.

Examples of musical interpretation in a strophic song were found in Polly’s singing at age 4;6. Polly (E) sang the first verse of Bingo accurately in a moderate tempo. In subsequent verses, she always sang the first phrase in its original tempo, but experimented with the tempo and dynamic levels in the remainder of each verse. It is noteworthy that Polly was able to clap or remain vocally silent according to the rhythmic patterns of the syllables for the middle section of the song. In the other instances, she was able to manipulate tempo and dynamic levels within a song.

The data suggest that at this age level, children begin to learn that a song has a beginning and an end. A child sometimes gave the first note of each verse greater length, but then sang the rest of the song in an appropriate tempo. The final notes of a song were often sung with a ritardando (gradually becoming slower) thus conveying a sense of closure. For example, Amber (E) demonstrated these techniques when she was 4;4. She was able to provide interesting interpretations to highlight the song structure of The Farmer in the Dell and The Wheels on the Bus, which, with their multiple verses, might otherwise be rather repetitious.

Polly (E) also used changes in vocal timbre to suggest different characters in a song. At 4;6, she sang The Ducklings using her normal singing voice in the first section which describes the little ducklings swimming. In the second section, she switched to a much deeper voice when the song described a turtle swimming slowly, and wanting to catch the ducklings. The performance of the two different timbres in Polly’s singing also
enhanced the contrasting tonalities of the two sections (the first section of The Ducklings is in major tonality, the second section is in its relative minor key).

5. Manipulation of Song Materials

Fewer “pot-pourri” songs were observed at this age. Rather than combining two or more songs into one, children now grouped songs and performed them as a series under a common theme or story. The children appeared to see a connection between a number of songs because they would often sing a second song immediately on completing the first. For example, children often sang several animal songs in a row. They would frequently interrupt a song not because they wanted to be finished with it, but because they wanted to sing another one. The idea of the second song would come to them while they were singing the first. As a result, learned songs collected at this age level were almost always sung in their complete form, thus corroborating my earlier observation that children were beginning to understand a song as a whole, since their singing demonstrated a clearer sense of beginning and end, and connected words with melody.

6. Complexity of Singing

Some children in this age level demonstrated remarkable skill at singing songs that could be considered challenging for much older children or even for adults. Songs such as My Favourite Things (from the musical The Sound of Music) consist of difficult melodic leaps and modulations often considered to be too demanding for a young child. So, for example, the judges agreed that Amber (E) at age 4;11 performed this song exceptionally well. She was able to maintain the tonality, and even sang accurately, intervals that are often considered to be difficult. Vicki (C/E) at 4;6 also sang Once Upon a Time, a song with a disjunct melodic line made up of wide leaps.

Amber had begun piano lessons several months prior to my visit at age 4;11. When she performed some songs she had learned at piano lessons, she switched back and forth—singing the words and singing the melody in solfege (or tonal syllables such as dol, re, mi,
fa, sol, la, ti). She sang *Three Blind Mice* in tune, with correct rhythm (especially the long and short notes), whilst at the same time patting the book. The patting was synchronized with the melodic rhythm of her singing. When she reached the final two phrases, she sang the solfege names (m-r-d, m-r-d) instead of the words “Three Blind Mice”. Singing in solfege appeared to be a newly acquired skill which Amber enjoyed. She sang the first phrase of *Hot Cross Buns* but abruptly interrupted herself, said “No!”, and then switched immediately to sing “m-r-d, m-r-d”. Amber subsequently sang the rest of the song in solfege while practicing the fingering as though she was playing the piano.

7. Singing While Playing a Musical Instrument

I also recorded Amber (E) singing while she was playing a guitar. At age 4;4 Amber first sang *My Favourite Things* unaccompanied. The pitches were sung in a somewhat detached manner but were melodically accurate. Before continuing to sing the chorus section, Amber repeated the first verse and increased her speed of singing. Amber’s father then suggested that she play the guitar and sing. Since Amber had no training on the guitar, she strummed some random tones on every strong beat of her singing, maintaining the tonality of the song despite the random strumming. This clearly demonstrated Amber’s secure sense of tonality of the song. Interestingly, Amber also sang the verse twice while she played the guitar: during the first iteration, she sang in detached notes; for the second iteration, she increased the speed of the tempo. She also increased the speed of strumming to synchronize with her singing. It appeared that Amber had an overall sense of unity and contrast for her singing performance because both the accompanied and the unaccompanied versions were performed with the same style; that is, the first rendition was sung in detached notes, and the second was at a faster pace.

Towards the end of a recording session when she was 4;11, Amber said that she wanted to show me a beautiful tune. She went to a piano and started playing the theme from the movie *Titanic*. Although she did not use all fingers when playing, the pitches and
rhythms were correctly reproduced. The judges were particularly impressed by the accurate execution of the rhythm for the repeated notes at the beginning of the song. By the middle section of the song, she attempted to match her voice with the piano, but Amber later stopped the singing and concentrated on playing the piano. According to her mother, Amber had not been taught to play the melody, she simply recalled the tune after listening to the soundtrack. Amber’s father explained that Amber enjoyed listening to recorded music and would sometimes dance to the music for nearly an hour.

**Intermediate vocalizations**

Some *intermediate vocalizations* were collected from Amber (E) when she was 4;11. Since the inspiration for these vocalizations appears to have been drawn from nursery rhymes and children’s songs, I called them “nursery musicals”. I categorized these vocalizations as an intermediate form because they were made up of intermittent singing and speaking as determined by the drama of the story; that is, there was a constant but unpredictable shift between speaking and singing.

One “nursery musical” began with Amber telling the story of “Peter Rabbit” as she was reading a book. The story was delivered in an animated voice with speech inflections. Amber used a voice noticeably different from the one she used in her normal conversations. At the point in the story where Peter Rabbit met “Winnie the Pooh”, Amber interrupted her story and shifted abruptly to sing a *learned song*, *Winnie the Pooh*. After singing several phrases, Amber continued to tell the story. To build up climax and anticipation, Amber slowed down her speech and whispered “a . . big . . black . . wolf” in long sounds; to keep the intensity, Amber whispered as she described how the big black wolf chased Peter Rabbit. When she described how Peter Rabbit managed to climb a tree to escape, she raised her voice gradually to the peak of her vocal register: when she described how the wolf cried because he could not climb the tree, her speech slowed down to convey a
Amber rounded off the story by stating “The End”, and switched to look for songs in her folder.

The second segment of the “nursery musical” began with Amber singing *The Wheels on the Bus*, as she was going through her music scores collection. She sang two verses of the song accurately. In the second verse, Amber imitated baby cries by modifying her voice into a high, piercing sound. However, without a pause, Amber suddenly switched back to the “story-telling” mode. She continued with the earlier story except that rather than describing how the big black wolf could eat up animals and even a tree, she introduced the character of a bumble bee into her story. Again without a pause, she inserted a *learned song* fragment about bees into the story. The *spontaneous song* then emerged into a “pot-pourri” song, including some fragments of a song about fishing. The singing sequence ended with an improvised section, which was sung slower and slower until it gradually integrated seamlessly with another section of story-telling. Amber again returned to the excerpt describing how the big black wolf ate up other animals. This and the earlier segment on a big black wolf were not from the storybook which Amber was reading. Since Amber was familiar with Prokofiev’s *Peter and the Wolf*, it appeared that she was trying to incorporate the story of this instrumental work into her “nursery musical”.

The whole “nursery musical” lasted for about ten minutes. It is noteworthy that a storybook was the stimulus in the first segment, while a folder of music scores was the stimulus in the second segment of the vocalization. When Amber was reading the storybook, I assumed that she intended to tell me a story. Similarly, when she was sifting through her music score collection, I assumed that she was looking for songs. It was interesting to find that in reading a story, Amber inserted songs; in going through her music scores, she made up stories that connected her songs. The intermittent singing and

---

35 An instrumental work that introduces different musical instruments to young children.
speaking appeared to be the most suitable mode of discourse for Amber to communicate her story. The theatrical element found in Amber's "nursery musical" was not found in other children who participated in the present study. I speculate that Amber might have become familiar with such a manner of communication by watching children's entertainment programs, which often incorporate story-telling and singing songs. The "nursery musical" may well be an enactment of this experience.

Five to Six Level

General Speech Development of the Child

At the time of data collection, Polly (E) was the only child in this age level. To assist the reader to follow the interpretations of data, a more detailed discussion of her language and music background is warranted.

Polly was born in Peru and came to Canada just before age two. The parents spoke Spanish at home, but since Polly had no opportunity to speak Spanish outside of home, she was reluctant to speak Spanish. According to Polly's mother, her daughter's lack of Spanish speaking ability became more apparent when Polly's grandmother came from Peru to visit them. There were occasions when Polly's mother had to translate grandmother's speech into English to help Polly understand. It can therefore be seen that English had become Polly's dominant language. Polly was also considered by her teachers and neighbours to be advanced in English language development. A number of factors may have contributed to such a view: Polly was a verbal child who communicated effectively with adults because she was able to express complicated thoughts and ask questions that reflected independent thinking, she also enjoyed reading. She had a large collection of bedtime storybooks; her parents spent long periods of time (up to 1.5 hours every evening) reading stories to her before she went to bed; and she listened to recordings of children's stories whenever she travelled in the family car.
General Musical Development of the Child

Polly had developed an unusual passion for music and dance at a young age. According to her father, Polly was exposed to Spanish ballads, Latin American music, and African folk and pop songs at home. Since the parents noticed her interest in singing, they enrolled her in weekly music classes beginning at age two-and-a-half. She became enthusiastic about dancing at age three, often danced with her father, and was therefore enrolled in a pre-ballet program for four months. At about age three, Polly enjoyed movie productions of Broadway musicals such as *Mary Poppins*, and *The Sound of Music*. According to her father, Polly could sing the songs and model after some of the dance sequences in those musicals. At age four, she even enjoyed watching opera and ballet productions. Polly attended live performances of Tchaikovsky’s *Nutcracker Suite* at age three and at age five. When she watched the video recording of *Sleeping Beauty*, her father noticed that she recognized the main themes and danced along to the music. The many experiences Polly had in music and dancing appeared to be integrated with her fondness for story-reading. Hence, dramatic works combined with music and dance had always captured her fascination.

Polly’s unusual fondness of music was also evident in her participation at the annual Vancouver Children’s Festival. She first attended the Festival at age three, and attended again at age four. The parents described to me how Polly paid exceptional attention to the live musical performances at the festival. Polly attended different types of musical events, and even stayed after the performances to talk to the performers. During informal conversations I had with Polly’s parents, her father commented that there appeared to be a “bonding” between Polly and the performers.

At age five, Polly enjoyed operas. At Polly’s request, the family frequently rented opera videos. The parents would give Polly a general outline of the plots before they watched the operas. Polly had just seen Bizet’s *Carmen* at the time I visited her. Polly’s
father commented that Polly was able to sit through the whole production, despite details that were too complicated for her to fully comprehend. It was surprising to learn from Polly’s parents that she did not enjoy popular children’s movies produced by Disney, even though these also featured songs and dances. According to her parents, Polly criticized cartoon movies for their superficial story lines and “uninteresting” songs. She repeatedly commented that Disney movies were “boring; for babies only”, but she did not explain what she meant by this comment.

Singing

At age 5;1, Polly sang 14 songs in a row during my visit. All 14 songs were either learned at school or from the music program she attended about two years earlier. Little efforts were needed to elicit singing. She thought of songs she would like to sing as she went along. For songs that had several verses, she often sang all verses; for songs that had actions, she performed the corresponding movement as she sang. Although she occasionally forgot the words of a song, she could make up words that fit the general meaning of the phrase. When she failed to find words to substitute, she would carry the melody along by singing monosyllables.

Polly demonstrated an impeccable sense of tonality. For learned songs she performed at this time, she maintained accurate melodic contour and sang most of the melodic intervals correctly. The following scenarios illustrate these points.

Scenario 1: Polly demonstrated a fine discrimination of appropriate singing range and tempo. She first sang Ten Little Angels in a low vocal range and a rather slow tempo. As a result, she had difficulty singing the lower pitches of the song accurately. Since the song is considerably longer than other songs, the slow tempo demanded more breath control and appeared to be tedious. Nevertheless, Polly maintained the tonality and tempo she established at the beginning of the song and completed the first verse. Then spontaneously, she sang the entire song again, only this time she chose a higher starting
pitch. The whole song then appeared to be more comfortable for her vocal range after she transposed it to a higher tonality. Polly also sang the song in a faster tempo this time, thus making the song livelier and the breath control less demanding.

Scenario 2: Polly spontaneously sang *Bow, Bow, Belinda*, which was one of her favourite action songs. Slight mistakes were made in the first verse but otherwise it was sung accurately. For the second verse, Polly sang it in staccato (the notes were sung sharply detached), and for the third verse, she built up excitement with an accelerated tempo (the verse was sung while the pace gradually speed up). Polly had inadvertently omitted the fourth verse and skipped to her favourite final verse. When she realized this, she immediately interrupted her singing and went back to sing the fourth verse. Yet before singing the final verse of the song, she paused and stepped to the middle of the living room to perform the actions. Unlike the melodious singing for all previous four verses, the final verse was sung in a chant-like manner. Polly appeared to enjoy performing the corresponding body movements, but as she became more preoccupied with the movements, she sang cheerfully and acted silly. The final portion of this song was sung jokingly in a vibrato voice an octave above the original melody.

*Intermediate vocalizations*

A type of intermediate vocalization was observed from Polly at 5;1. The recording session began with conversation and reading picture books to elicit English speech utterances. Polly then sang songs for me, and by the end of the 14th song said that she felt tired. However, she still wanted to “do” a book, and she went upstairs to bring along a “touch and smell activity book”.

I was expecting Polly to tell me the story when she opened up the book. However, she sang the title of the book, and for approximately ten minutes, she continued to sing

---

This type of book is designed to enhance reading experience. To stimulate a child’s imagination during reading, the book has straps and patches made with different materials for a child to pull, scratch, touch, and smell.
through every page. From my perspective, her vocalization was rather obscure; it took on an ambiguous mannerism suggesting an uncertainty of communicative intent. I called it “story-singing” because she told the story using an original melody that she improvised on the spur of the moment. Interestingly, the improvised melody did not bear much resemblance to musical characteristics commonly found in children’s songs; rather, the “story-singing” reminded me of operatic singing. A more elaborate discussion based on the analyses reported by the judges further illustrates this observation.

The story was consistently sung in an extraordinarily high vocal range. The tonality centered around 640 to 700 Hz (E5 to F5). Polly appeared to have a melody in mind because even though she was already singing within a high range, she still made the effort to sing to a the target pitch, as if this was how the melody was “supposed” to be sung. As a result, the melody exceeded 880 Hz (A5) at its highest point. (See Figure 4.24 for a segment of “story-singing” versus “story-speaking”.)
This is Tom and Sara. Tom pats the puppy. ("Story-singing")
This is Tom and Sa-ra. Tom pats the pu- py.  
("Story-speaking")
They both walked over to grand-ma and grand-pa's house.

("Story-singing")
They both walked over to grand-pa and grand-ma's house.
(“Story-speaking”)
Polly sang with a moderately slow tempo, and the note values for pitch were relatively long. Phrases were occasionally repeated, and there were pauses in-between phrases. At the end of each page, Polly sang the words to a melodic pattern that implied cadential structure. Each page would typically end with a step-wise motion to the tonal centre established in previous phrases, thus produced an effect which “marked off” individual musical sections of the storybook.

Polly’s singing also conveyed an overall strophic design, with each verse (every two pages) beginning with a similar melodic motive, “s-m-s-m-s-l”. Hence, motivic (i.e., a small unit of musical pattern) unity for the whole “story-singing” was achieved with this recurring melodic pattern. (See Figure 4.25 for a transcription of the recurring melodic motive of Polly’s “story-singing”.)

Figure 4.25
A Transcription of the Recurring Melodic Motive in Polly’s “Story-Singing”

This is Tom and Sara,
Tom pats the puppy.

They found a frog and made it jump,
They smelled something good,

Sara and Tom push the rocking chair,
Grand-ma walks, she walks them home.
“Story-singing” demonstrated Western European melodic techniques such as the use of melodic sequence and embellishment. For the word “Grandma”, one of the main characters in the story, the words were sung three times in a descending melodic sequence (see Figure 4.26). Interestingly, this descending sequence appeared after an ascending melody which had climbed in leaps to a climax. Hence, the melody achieved a broad arching shape which is typical of Western musical tradition. An example of melodic embellishment was found at the end of the second “verse”, when Polly added an extra “ah” to the final sentence. It was sung in a long note embellished with a trill (see Figure 4.27). The next verse then began with the same melodic pattern “s-m-s-s-l” in which she adapted the tones to fit with the words.

Figure 4.26
A Sequential Treatment on the Word “Grandma” in Polly’s “Story-Singing”
Throughout the entire “story-singing”, Polly appeared to be relaxed; she smelled, touched, pulled, and showed me activities on each page and expected me to follow what she did. Although her song was spontaneous, her timing was well controlled. There was an easy flow in her singing which allowed sufficient time for playing and interacting with me. The “story-singing” was delivered at ease.
To facilitate a comparison between "story-singing" and normal speaking of the story, I asked Polly to "tell" me the story instead of "singing" it. Polly then repeated the story in normal speech. In this second rendition, which might be called "story-speaking", speech inflections were the most prominent feature observed. Polly's voice animated the actions of the characters in the story through enunciation of key descriptors and variations in pitch and duration. According to her mother, although Polly was not yet capable of reading the words, she told the story precisely according to the way it was written in the book. Polly also paused between pages to find time to manipulate the fixtures in the book and to interact with me. She was as relaxed in this activity as she had been in "story-singing". Some features identified in Polly's "story-singing" were not, however, found in her speaking. For instance, sequential treatment on the word "Grandma" (see Figure 4.26) did not occur in this "story-speaking" version, and neither was the "ah" embellishment found. Noticeable difference in the length of the vowels was also found between the sung and read versions; the sounds were much shorter in the "story-speaking" version than in the "story-singing" version.

Summary

From the preceding analyses, young children's vocalizations provided insights into important aspects of their vocal and musical development. A compilation of features observed in these vocalizations suggests three broad stages in a developmental continuum:
Stage 1--Young children learn to fine tune their vocal control in both singing and speaking.
Stage 2--Young children begin to sing in ways that more closely resemble an adult model. Musical idioms to which they are exposed emerge first in learned songs and then gradually became assimilated into spontaneous songs.
Stage 3--As children mature, they not only master the skills needed to perform songs accurately, but may express novel forms of vocalizations that purposefully alternate easily between singing and speaking.
Table 4.5 summarizes the key features observed in vocalizations collected at each age level. Since the proposed stages may be considered as “tendencies”, I have avoided marking off stages with reference to age. A more extensive discussion is presented in the next chapter.

### Table 4.5

**A Summary of Features Observed in the Children’s Vocalizations**

<table>
<thead>
<tr>
<th>Age</th>
<th>Subjects</th>
<th>Features Observed in the Children’s Vocalizations</th>
</tr>
</thead>
</table>
| 1.5 level | Heidi                     | • Immature sentences in Mandarin, and some breathy, guttural, unintelligible utterances.  
                     |                                           | • Short melodic fragments using repetitive words.  
                     |                                           | • Spontaneous singing while playing piano - snatches of nondiatomic and unmetered sung sounds with a variety of high/low and long/short tones separated by long pauses.  
                     |                                           | • “Chanting song” was an animated vocalization which resembled maternal speech. It was identified by the mother as song but perceived by the judges as speech. |
| 2-3 level | Heidi, Vicki, Jodi, Wendy, Amber, Gina, and Clare | • Chinese bilingual children spoke mainly immature Chinese sentences with some single English word utterances.  
                     |                                           | • English monolingual children spoke immature English sentences.  
                     |                                           | • Cases of diplonphonia were found in both speech and singing utterances across language.  
                     |                                           | • Incomplete rendition of learned songs with words freely substituted with nonsense syllables were abundant. When the words of a learned song were forgotten, a child failed to recall the melody.  
                     |                                           | • Only fragments of a melody were intact, other parts of the song had melodic and rhythmic patterns incorrectly sung.  
                     |                                           | • Some children recalled a song through body movements or imaginative play rather than singing it aloud.  
                     |                                           | • Abundant spontaneous songs, especially “pot-pourri” songs combining fragments of learned songs and improvised songs, sometimes regardless of different languages.  
                     |                                           | • Children made less references to diatonic system and metrical rhythm in their spontaneous songs than in their learned songs.  
<pre><code>                 |                                           | • “Prosody of poem” was a speech-like vocalization with sustained tones and distinctive melodic contour. |
</code></pre>
<table>
<thead>
<tr>
<th>Age</th>
<th>Subjects</th>
<th>Features Observed in the Children’s Vocalizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-4</td>
<td>All subjects</td>
<td>• English appeared to be the language of major use for all subjects.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The Chinese bilingual children became increasingly reluctant to speak first language at home and their Chinese song repertoires decreased.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• A marked increase of learned songs was found in all children; the melodic and rhythmic patterns were sung mostly correctly.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• A growing sense of tonality with some modulations was evident.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The entire song was often considered forgotten when only the words were forgotten. Nonsense syllables were used to fill in missing words.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The precise structure of learned songs was not recalled accurately even when all other aspects of the song were sung accurately.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Pot-pourri” songs were abundant; they were sometimes improvised with movements and musical instruments, or might resemble singing games.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Babbling vocalise” was a form of monologue consisting of fragments of learned song integrated with improvised singing and intermittent vocal babbling of speech sounds.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Drawing narrative” was a vocalization that integrated narration and adaptation of learned song into a drawing activity.</td>
</tr>
<tr>
<td>4-5</td>
<td>Vicki, Jodi, Wendy, Clare, Amber, Gina, and Polly</td>
<td>• Evidence from children’s story-telling suggested that the English monolingual children were verbally more advanced.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Although the Chinese bilingual children spoke fluent Chinese, they sang almost entirely in English.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The Cantonese children had difficulty speaking with a Chinese book.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• All children had an increased English song repertoire and sang complete version of learned songs accurately with minimum stimuli.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Forgotten words now appeared to have less effect on melodic singing accuracy. Most children maintained the tonality; some of them were even able to transpose or embellish original melodies.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Evidence of an awareness of tempo and dynamics was found in a child’s manipulation of strophic songs.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Songs were often performed as a set unified by a common theme.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Some children were able to sing musically demanding songs that consisted of wide intervals and angular contour.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• One child sang songs in solfege, accompanied singing with a guitar, and recalled a movie theme on the piano.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• &quot;Nursery musical” was a highly dramatized vocalization for story-telling that integrated speaking and singing intermittently.</td>
</tr>
<tr>
<td>Age</td>
<td>Subjects</td>
<td>Features Observed in the Children’s Vocalizations</td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>--------------------------------------------------</td>
</tr>
</tbody>
</table>
| 5-6 level | Polly   | • Sang abundant *learned songs* fluently with all verses and accompanying actions.  
• Tonality was securely maintained, and the intervals were sung mostly correctly even when some words were forgotten.  
• Demonstrated fine discrimination of an appropriate vocal range.  
• Various fine controls such as articulations and variations in tempo were used to provide contrasts for repeated verses.  
• “Story-singing” was a story presented spontaneously in a pseudo-operatic singing manner. Many features demonstrated implementation of Western musical conventions, including melodic embellishment, achievement of motivic unity, and the use of cadence to “mark off” musical sections. |
CHAPTER FIVE
SING ME A SONG OF A PRESCHOOL CHILD

Statement of the Problem

A comparison of vocal development between children of tonal and nontonal languages requires an examination of speech-related vocal behaviours because acoustical and perceptual definitions for the different behaviours of singing and speaking have not yet been reliably established, particularly in young children. Research on the development of singing in young children should therefore focus on the contexts that constitute singing and speaking. The following discussion addresses findings for research questions initially posed in Chapter 2.

Research Problems Revisited

(1) *How do young children's vocal fundamental frequencies change across time, and how do such changes inform us about their singing and speaking?*

An overall longitudinal decreasing trend in mean fundamental frequencies ($F_0$) was observed in the vocalizations of children. This finding is consistent with previous studies and is generally considered to be an effect of growth in vocal tract and body size. Between ages three to six, however, decrease in mean $F_0$ was found to be relatively small compared to the relatively constant rapid increase in membranous length of the vocal fold. This interesting finding warrants further research.

Adults in Western culture have developed distinct vocal behaviours for singing and speaking. The Papouseks (1981) propose that by about age one, singing and speaking may gradually become more disparate to better serve distinctive social and communicative functions in later life. Throughout the 42 months of this investigation, all children consistently sang with higher $F_0$ than they used for speaking, while other forms of vocalization appear to be positioned consistently between singing and speaking. It
therefore appears that by age two, young children have established communicative pitches that are associated with different forms of vocalizations.

(2) In what contexts do young children sing?

A song appears to be a more dynamic concept for a child, which appears not to be true for adults. In the present study, singing was observed in a number of different contexts, including being given an invitation to sing; an invitation to tell a story; and when engaged in a game or an imaginative play. The type of singing seemed to vary across these contexts: when invited to sing, a child was most likely to perform a learned song; when invited to tell a story, a child often included fragments of learned songs with improvised singing, or sometimes created original songs on the spur of the moment; and, in a game setting, a child often spontaneously improvised songs.

After age two, a child appears to have a clear understanding of what constitutes a song. In Western culture, when an adult engages in singing, he or she is more likely to render a composed song rather than to improvise a song. When invited to sing, a child often responds by performing a learned song, probably a result of modelling adult productions. On the other hand, observations suggest that children may not be limited to the circumscribed vocal behaviour practiced by adults; they may aptly integrate singing into a range of musical discourse such as vocalizations in a game or a story-telling context. Hence, it can be seen that young children’s singing behaviours are dynamic, that is, young children are flexible in adapting to different contexts to facilitate self-expression.

37 For example, a child responded to her mother’s invitation to sing a song by reciting a Chinese poem (Heidi’s “chanting song” at 1;6 months). Another child responded to an invitation to tell a story by singing in a pseudo-operatic vocal style (Polly’s “story-singing” at 5;2).
38 For example, Amber’s “nursery musical at 4;11 and Polly’s “story-singing” at 5;2.
39 For example, Amber’s spontaneous singing game at 3;1.
(3) What forms of vocalization are observed in young children and what are their roles in vocal development?

In the present study, most of the children’s vocalizations were easily classified by judges as corresponding to vocal responses, which accord with adult conceptions of song and speech. Some other vocalizations could not be easily classified into either singing or speech because they appeared to be acoustically or contextually ambiguous. Further analyses point to two broad forms of intermediate vocalization.

(1) A considerable number of vocalizations were considered intermediate because their acoustic characteristics were between those of speaking and singing, making it difficult to draw clear perceptual distinctions.

(2) Older children sometimes alternated between speaking and singing although they were able to make clear distinctions between their song and speech; these vocalizations were therefore considered intermediate.

These two forms of intermediate vocalization appear to play unique roles in the vocal development of a young child (see Figure 5.1). Many examples of acoustic intermediate vocalizations were found in the initial stage of vocal development when children were acquiring vocal control. With maturation of the vocal mechanisms, clearer forms of singing and speaking, resembling that of the adult model, emerged. Finally, it was observed that after a child had mastered the appropriate ways to sing and speak, she might even purposefully alternate easily between singing and speaking to communicate in novel forms of contextual intermediate vocalization.

---

40 Dowling (1984), for example, distinguishes singing from speaking by the presence of sustained vowels on steady and discrete pitch levels, and the more complex rhythmic structures with an underlying steady beat.
Figure 5.1

A Graphic Depiction of Intermediate Vocalizations in the Vocal Development of Children

<table>
<thead>
<tr>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children learn to fine tune their vocal control in both singing and speaking. (Approximate age: 1;6-2;0)</td>
<td>Children begin to make productive distinctions between singing and speaking in ways that more closely resemble that of the adult model. (Approximate age: 2;0-3;6)</td>
<td>Children master the skill needed to perform songs accurately and to communicate effectively in speech. They may also express novel forms of vocalizations that purposefully alternate easily between singing and speaking. (Approximate age: 3;6-6;0)</td>
</tr>
</tbody>
</table>

Table 5.1 is a compilation of some developmental traits observed in intermediate vocalizations. In these intermediate vocalizations, linguistic and musical traits were observed, providing interesting information on the acquisition of speech, and of song production.

Because the preschool years encompass a number of milestones in the acquisition of linguistic and musical behaviours, observations of vocalizations at this age level point to possible interactions or “temporal coincidence” (Imberty, 1996) between linguistic and musical development. Table 5.2 compares some of the milestones in a child’s linguistic and musical development as observed through intermediate vocalizations. Although the behaviours described were observed in all children in the study, the table illustrates approximations rather than absolutes because the precise time for the emergence of a particular behaviour varied between or among children. Results appear to suggest, however, that acquisition of musical and linguistic behaviours by preschool children follow...
a similar developmental trend. Further research is needed however, to provide more conclusive evidence about the parallels illustrated from present observations.

Table 5.1
Some Developmental Traits Observed in Intermediate Vocalizations of Children

<table>
<thead>
<tr>
<th>Vocalizations</th>
<th>Ages</th>
<th>Linguistic Traits</th>
<th>Musical Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stage 1</strong> Children learned to fine tune their vocal control.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Chanting song&quot;</td>
<td>1;6</td>
<td>The vocal interplay between a mother and her child resembled maternal speech.</td>
<td>Contrasts in pitch level and duration were evident. Patterning of sounds suggested schematic organization salient to musical behaviour.</td>
</tr>
<tr>
<td>&quot;Prosody of poem&quot;</td>
<td>2;1</td>
<td>In reciting poems, pitch contour and syllable length were exaggerated.</td>
<td>In response to a mother’s invitation to “sing”, a child produced some sustained vocal utterances with distinctive melodic contour.</td>
</tr>
<tr>
<td><strong>Stage 2</strong> Children began to sing and speak in ways resembling that of the adult model.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Pot-pourri&quot; songs</td>
<td>2;3</td>
<td>Two otherwise unrelated songs were ambiguously connected with two words that have the same sound.</td>
<td>Words served as the anchor for a song. Children often failed to render a melody when words were forgotten.</td>
</tr>
<tr>
<td>Spontaneous singing game</td>
<td>3;1</td>
<td>Performed an imaginative play that integrated conversations, recited repetitive phrases and a movement game.</td>
<td>Intermittent singing consisted of scalic passages accompanied by a toy instrument. It resembled a singing game adapted from a learned song.</td>
</tr>
<tr>
<td><strong>Stage 3</strong> Children mastered the skills to sing and speak effectively but they sometimes expressed themselves through novel forms of vocalizations.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Drawing narrative&quot;</td>
<td>3;5</td>
<td>Used a narrative to inform the audience while drawing a picture.</td>
<td>Adapted a learned song to accompany a drawing activity. Together with the narration, it assumed a strophic form.</td>
</tr>
<tr>
<td>&quot;Babbling vocalise&quot;</td>
<td>3;6</td>
<td>Speech monologue made up of phoneme babbling, counting, reciting words of a learned song, and articulatory sounds.</td>
<td>Sung utterances intermittent with spontaneous singing of diatonic melodic patterns and learned song fragments.</td>
</tr>
<tr>
<td>&quot;Nursery musical&quot;</td>
<td>4;11</td>
<td>Story-telling in animated speech along with spontaneous singing. The words of the singing fit nicely with the story.</td>
<td>Integrated learned songs and improvised melodies to enliven story-telling.</td>
</tr>
<tr>
<td>&quot;Story-singing&quot;</td>
<td>5;1</td>
<td>Presented a story using an unconventional form of vocalization while actively engaged in the tactile activities provided by a book.</td>
<td>Pseudo-operatic singing on an improvised song that exhibited rich Western European melodic idioms.</td>
</tr>
</tbody>
</table>
Table 5.2

Some Parallels Observed in Children's Developmental Milestones

<table>
<thead>
<tr>
<th>Developmental Milestones</th>
<th>Speech Acquisition</th>
<th>Song Acquisition</th>
<th>Behaviours Observed in Intermediate Vocalizations</th>
</tr>
</thead>
</table>
| 1. Reacted to speech and musical stimuli in vocal response. | Responded to vocal sounds through infant-directed speech and monologue babbling. | Demonstrated discrete pitches and contrast of high/low and long/short tones in sung sounds. | • In "chanting song" and "prosody of poem", speech was heightened with pitch and duration contrasts.  
• In "babbling vocalise", speech phonemes and other articulatory sounds were practiced along with sung fragments as in a vocalise. |
| 2. Assimilated recurring sound structures from linguistic and musical environment into vocal production. | Acquired vocabulary and other language structures and gradually combined units to form immature sentences. | Scalic structure, tonality, and other melodic features emerged in learned song fragments and spontaneous songs. | • In various spontaneous singing games and "pot-pourri" songs, learned musical and speech structures were manipulated into new configurations.  
• In "drawing narrative", learned song material was adapted into a narrative performed during a drawing activity. |
| 3. Refined and implement learned structures to communicate thoughts in vocal production. | As a child became fluent in speaking a language, messages and feelings were communicated in more sophisticated speech. | As a child became fluent in a range of musical elements such as tempo and dynamics, songs were sung to facilitate expression of emotions. | • In "nursery musical", learned and improvised songs were integrated to dramatize story-telling.  
• In "story-singing", a story was sung in a pseudo-operatic style. A child also shared the tactile stimuli in the book with the audience while she spontaneously sang the story. |
(4) Is there evidence suggesting that first spoken language may affect vocalizations? That is, is vocal development different in tonal and nontonal languages?

In the present study I investigated the extent to which a child's vocal development might be susceptible to effects of first spoken language. While the mean $F_0$ data suggest some possible differences in vocal pitch behaviours between English-speaking and Chinese bilingual children, the qualitative data illustrate an interesting divergence in the way these children use vocalization as a form of discourse. Hence, results of the present study tend to support a position of language-specific vocal pitch behaviours (Dolson, 1994; Deutsch, 1987; 1991; 1994).

Evidence of language-specific vocal pitch behaviour was found in a comparison of ways in which English-speaking and Chinese bilingual children differentiate between their speech and song production. As an English child matured, she made greater differentiation between speaking and singing by selecting a more distinctive pitch centre for these two types of vocalization. In other words, an English child's speech occupied a distinct vocal range from her song. On the contrary, it appeared that a Chinese child made comparatively less distinction between her speaking and singing as she matured. Note that the present findings of vocal pitch behaviour cannot be attributed to differences in pitch characteristics of the language sung because only English songs were collected from the Chinese bilingual children at the age 4-6 level. In other words, comparisons between the two groups of children were based on a similar English song repertoire. In this way, any inherent interactions due to the language used in singing were eliminated. Thus, it may be inferred that differences in the vocal behaviours of these children might be predominantly language specific.

Language differences appear particularly in intermediate vocalizations. Across time, a Chinese bilingual child was found to make comparatively fewer, but nonetheless more stable acoustical distinction between her speaking and singing. Some of her
intermediate vocalizations even had acoustical characteristics of vocal babbling of speech sounds. For example, Heidi (Mandarin/English) made no explicit distinction between reciting a poem and singing a song.\textsuperscript{41} Another example was found in Vicki (Cantonese/English).\textsuperscript{42} At age 3;6 Vicki spoke fluently and had surpassed the initial stage of speech development when vocal babbling is most evident. However, it appeared that acoustic features of speech were more attractive to Vicki than word meanings. Hence, the sounds of words became resources for vocal improvisation where they were combined with some learned musical structures such as diatonic melodic patterns and learned song fragments. In other words, Vicki’s “babbling vocalise” might be viewed as a kaleidoscope of spoken and sung sounds. The boundary between speech and song was therefore “fuzzy” in this form of acoustic intermediate vocalization. This form of acoustic intermediate vocalization was also observed in other Chinese bilingual children.

Older English-speaking children displayed another form of intermediate vocalization. As English-speaking children got older they made increasingly clearer and wider acoustical distinctions between their speech and songs, their intermediate vocalizations were made up of intermittent singing and speaking in which they easily switched between the two. Analyses of the contexts of these intermediate vocalizations suggest that novel vocal forms were used to dramatize speech and to elevate the level of excitement in a story. For example, the swift transactions between speech and song used by Amber were reminiscent of a typical children’s television program in which a skillful television personality engaged the audience in a succession of activities such as storytelling, song singing, and puppetry.\textsuperscript{43} In this intermediate vocalization, not only did Amber demonstrate a clear understanding of the contextual implications of speech and song, she

\textsuperscript{41} See analyses of “chanting song” (pp. 85-86) and “prosody of poems” (pp. 101-106) by Heidi at 1;6 and 2;1.
\textsuperscript{42} See analyses of “babbling vocalise” (pp. 115-117) by Vicki at 3;6.
\textsuperscript{43} See analyses of “nursery musical” (pp. 131-133) by Amber at 4;11.
was also able to integrate speaking and singing to extend the spectrum of her discourse. It may be that this *intermediate vocalization* results from an exposure to a style of intermittent speech and songs used in television programming for young children.

*Intermediate vocalizations* might also show effects of the auditory environment. Polly's "story-singing" was classified as an *intermediate vocalization* because rather than telling a story as she opened a book, she spontaneously sang the whole story in a pseudo-operatic vocal style. What motivated a five-year-old to communicate a story in such a creative way? A detailed analysis of Polly's childhood illustrates the extraordinarily rich linguistic and musical experiences to which she was exposed. Polly was even drawn towards operas at the time data were collected. Effects of auditory environment are therefore a pertinent consideration in the investigation of children's vocal development. Evidence from the present study suggests that while the first spoken language appears to affect vocal development, the child's non-speech auditory environment is also crucial to the understanding of her vocal behaviours.

**Implications for Vocal Development of Young Children**

(1) Auditory Perception and Vocal Production

The perceptual ability of young children is important to consider in the investigation of vocal development. Studies of fetal sound environment suggest that fundamental frequencies of musical stimuli can be heard by a fetus with a maturing inner ear during the last trimester of pregnancy (Abrams & Gerhardt, 1997), and learning in response to sound is therefore possible in utero (Kolata, 1984; Birnholtz & Benacerraf, 1983). Hence, studies involving pre- and postnatal exposure to music demonstrate that the newborn infant's mind is not a *tabula rasa*. The infant is sensitive to musical properties, including melodic contour, octaves, simple frequency ratios, and certain properties of harmony (Trehub, Schellenberg, & Hill, 1997).
Young children's vocal production capabilities however, are reported to be much less sophisticated than the auditory perceptual capabilities of an infant. Observations of children's vocalizations show that some melodic features were produced more abundantly in songs learned from adult models, whereas others were produced more abundantly in spontaneous songs. Figure 5.2 illustrates some melodic features observed in children's vocalizations.

Figure 5.2

Melodic FeaturesObserved in Children's *Learned Songs* and *Spontaneous Songs*

<table>
<thead>
<tr>
<th>Melodic Features Observed in Children's Singing</th>
<th>Learned Songs</th>
<th>Spontaneous Songs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age Levels (in Years)</td>
<td></td>
</tr>
<tr>
<td>Repetitive melodic patterns</td>
<td>1.5</td>
<td>2-3</td>
</tr>
<tr>
<td>Melodic contour</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Descending contour</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Melodic phrasing</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Diatonic scale structure</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Maintainance of tonality</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Accurate Intervals</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Transposition of melody</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

- Abilities not observed in any of the children
- Abilities demonstrated by some of the children
- Abilities demonstrated by all children
- N/A Not applicable (learned songs contain essentially both ascending and descending contour).

Note: Only one child was available in the 1.5 and 5-6 levels for observation.

For the purpose of the present investigation, distinctions were made between observations of singing *learned songs* versus *spontaneous songs* because these forms of song appeared to require different cognitive skills (see "Classification of the Vocal Forms" in chapter 4). Singing a *learned song* requires a child to recall musical material according to a previously introduced model and such learned vocal behaviours might be bounded by cultural conventions. *Spontaneous song* on the other hand, are initiated by a child and the
child's aesthetic judgment is therefore an important component when she is not required to reproduce according to a model, when singing spontaneously. It is therefore inferred that there may be some idiosyncrasies in melodic features which could cause them to emerge differently in *learned songs* and *spontaneous songs*.

Results reported in the literature suggest that, although young infants are reported to be sensitive to a variety of fundamental elements of melody (Trehub, Schellenberg, & Hill, 1997), the ability to vocally produce these melodic features is observed at a much later stage in life. Hence, there is a large gap between what a child can perceive and what she can produce. The successful production of these melodic features in specific singing contexts is worthy of further examination. More explicit investigations of these singing contexts will advance our understanding of the process of converting perceptual capabilities into the successful production of melodic features. In addition, a more complete understanding of singing abilities will need to take account of children's intellectual capacity as well as the physical development of their vocal mechanisms.

(2) Physical Development of Children's Vocal Mechanisms

The physical maturation of a young infant's vocal apparatus occurs at a much slower rate than her perceptual capabilities. Not only are her vocal mechanisms significantly smaller than those of an adult's, but her vocal tract anatomy is also very different, resulting in differences more complex than simple proportional relationships. The following is a discussion of pertinent literature as it related to the findings of the present study.\(^\text{44}\)

At birth, the respiratory system is much smaller than the adult's. According to Eichorn (1970), the trachea of a new born is about 1/3 of the adult size, the bronchi are about 1/2 adult size, and the bronchioles, about 1/4 adult size. Hence, for maintenance

\(^{44}\) The primary sources of the information discussed are from: Thurman and Klitzke (1994); Morrison and Rammage (1994); Sundberg (1987); Titze (1994); and, Kent and Vorperian (1995).
breathing, infants are reported to use mixtures of diaphragmatic and thoracic movement to facilitate breathing (Bouhuy, 1977; Cotes, 1979; Dunnhil, 1962), and breathing rate of one-year-old’s is found to be at least double the adult’s. Since the nasal airway was reported to be smallest at age four (Fujioka, Young & Girdany, 1979), Kent and Vorperian (1995) predict that “velopharyngeal or nasal resistance to air flow would be increased in young children relative to older children and adults” (p. 163). Considering that larger air volume and significantly longer expirations than inspirations are required for speech and singing, a young child has to make substantive effort in breath management to vocalize like an adult.

Since the vital capacity of the respiratory system is not reached until after puberty, breath management is a continual challenge for young children. Muscle relaxation times are reported to be twice as long in young children than those of a ten-year-old, thus resulting in a slower speaking rate (Kent & Vorperian, 1995). Because older children have larger respiration capacity and stronger respiratory muscles, they show an improved sense of phrasing, as observed in children between ages 3;0 - 4;0 (see Figure 5.2 on various melodic features observed in their singing).

It appears that young children before age 3;0 did not yet possess adequate subglottic pressure and glottis resistance, their utterances were therefore typically fragmented and not clearly projected (for example, Heidi at age 1;6). Sundberg (1994) points out that “[l]oudness of phonation is primarily controlled by subglottic pressure. . . . Airflow (which, of course, determines air consumption) is entirely determined by subglottic pressure and glottal resistance, the latter of which, in turn, reflects the degree of adduction activity” (p. 41). Thus when a young child’s glottal resistance is low, her vocal folds may fail to make contact and therefore her vocalizations become “breathy” and not projected.

Because the larynx is responsible for creating most of the sound energy used in speech, and phonation is produced by the narrowing or closing of the glottis, to produce
fine variations in pitch and resonance effects, coordinated movements of numerous muscles of the larynx and pharynx are also needed. As Sundberg explains:

By what means do we control phonation frequency? Two factors affect it: the overpressure of air in the lungs, which is called the subglottic pressure, and the laryngeal musculature which determines the length, tension, and vibrating mass of the vocal folds. . . . if subglottic pressure is raised, phonation frequency rises, although not very much. . . . Instead, one has to give the main responsibility to the muscles manipulating the properties of the vocal folds. What happens mechanically is simply that the vocal folds are stretched to different degrees, depending on the intended phonation frequency. The longer, the thinner, and the more tense they are, the higher the phonation frequency becomes. The vocal fold tension is achieved by increasing the distance between their terminating structures, or in other words, between the thyroid (anteriorly) and the arytenoid cartilages (posteriorly). Normally, this is the result of a contraction of the cricothyroid muscles. (p.16)

Although the vocal ligament in the vocal folds begins to appear between ages one to four (Hirano, Kurita, & Nakashima, 1983), up to age six, the lamina propria of the vocal folds still lacks the distinctive layered structure found in adults (Kent & Vorperian, 1995). Hence, it might be expected that fine control of pitch production would be difficult for preschool children. Observations made in the present study suggest that fine control of phonation frequency was not mastered by a child before age four, thus it is to be expected that the maintenance of tonality and the production of accurate melodic intervals would be very demanding tasks for younger children, even though they may already possess good perceptual capabilities.
Since $F_0$ is considered to be inversely proportional to vocal fold length and laryngeal size,\textsuperscript{45} growth in children’s body size over time may generally explain the developmental changes in $F_0$. The vocal folds also undergo changes in size and mass during childhood. Thus, the vocal folds have been reported to increase by about 6.5 mm in length (Kahane, 1983), and the dimension of the laryngeal cartilages also increases in size and firmness. Observations of a general lowering of speaking and singing fundamental frequencies in the present study is in accord with this physical growth in the larynx which anticipates a steadily lowering and deepening of a child’s voice. Because the vocal ligament and mucosal tissues will not reach maturity until after puberty, (although children sing and speak with a higher register than adults) their vocalizations tend to be more breathy and produced in the speaking-voice range.\textsuperscript{46} These characteristics were most evident, for example, in Jodi’s utterances before she gradually changed at the 4-5 level.

The infant’s vocal tract is shorter and has a different configuration when compared with that of an adult’s. According to Thurman and Klitzke (1994), the infant has a short and slightly curved vocal tract to enable simultaneous breathing and nursing. At about 1;6 the base of the tongue begins a gradual inferior progression into the pharynx, until the posterior one third of the tongue is located in the pharynx at age 4. It is at this age when the vocal tract becomes longer and more curved to enable a richer variety of vocal resonation. It is reported that a child’s vocal tract reaches the basic adult configuration at about age 5. Observations made in the present study suggest that children at about age 4-5 level appear to have reached a milestone in their singing development--not only did their

\textsuperscript{45} Titze (1994) points out that “$F_0$ is inversely proportional to the vocal fold length and directly proportional to the square root of tissue stress” (p.172). He also writes that “it seems reasonable to assume that laryngeal size affects the fundamental frequency more closely” (p.171).

\textsuperscript{46} Morrison and Rammage (1994) also point out that “breathiness” can be “due to a lowered resistance to airflow, and subsequent longer open phase of the phonation cycle”. They explain that the speaking-voice register is produced by the “contraction of the vocalis/thyro-arytenoid muscle group, without a proportional increase in opposing lengthening forces (cricothyroid muscle) (p. 180).
song repertoires increase, there was also a marked improvement in their melodic singing accuracy; some children even began to demonstrate contrasts of tempo and dynamics in their singing to facilitate musical interpretations. Thus, physical growth and configuring of vocal tract structures appear to be correlated with (in a non-statistical sense) many of the abilities I observed in children's vocal production. Interpretations of children's vocal behaviours must therefore take account of the physical conditions of their vocal mechanisms.

Pedagogical Implications

The preceding discussion is a summary of findings taking into consideration the physical and perceptual capabilities of young children. In making the following pedagogical suggestions, my intent is to translate the findings of this study into educational strategies to enhance the vocal development of children.

1. General development of vocal facility

There are a number of reasons behind the incentive to encourage children to experiment with their vocal facility. Young infants are biologically predisposed to engage in vocal exercise and vocal play in childhood is reported to be an integral aspect of human development (H. Papouseks, 1996). In addition, the development of vocal facility is a prerequisite for speech and singing. That young children's vocal production matures much later than their auditory capacity makes it apparent that experimentation with vocal production is useful to provide auditory feedback for learning.

(a) Breath management

Breath control is an essential skill for singing. The ability to sustain long tones (extending the time for expiration) while maintaining pitch steadily requires careful breath management. Appropriate physical exercises and exercises to refine muscle control for expiration can increase children's lung capacity. Some suggestions on exercises for breath and vocal production are provided by Bennett & Bartholomew (1997, pp. 73-74).
(b) Exploration in phonation

Phonation is directly linked to the regulation of the rate of vocal fold vibrations and the use of articulators such as the tongue, lips and the velum for modifying the air stream coming from the lungs through the vocal folds. Exercises that focus on discovering different voice qualities might promote vocal coordination and experimentation with the voice. Vocal glides without excessive stress allow children to “feel” kinesthetically how their voice sounds over an extended range not commonly used on a daily basis. The experience can help minimize inhibition of singing in an unfamiliar vocal range. It can also enrich young children’s auditory experience in producing melodic contour before more precise pitch movements are mastered. Some strategies are provided by Bennett and Bartholomew (1997, pp. 81-82)\(^47\). Because preschool children are attracted to imaginative play and singing songs in nonsense syllables, vocal production employing a variety of articulations can be integrated into these activities. Later, children can be guided to fine tune vocal control to articulate speech phonemes and to sustain vowels for singing.

2. Singing development

The following pedagogical suggestions are derived from findings in the present study. Since singing is regarded as a continuum of behaviours acquired over time (Welch, 1986; Welch, White, & Sergeant, 1996), the following suggestions are arranged hierarchically according to the developmental challenges identified in young children.

(a) Vocal model

Because children assimilated familiar acoustic structures from their auditory environment, a good vocal model may facilitate the development of singing (Mang, 1997; Goetze, Cooper, & Brown, 1990; Green, 1990). It is hypothesized specifically, that vocal

\(^47\) Their model outlining six phases for vocal awareness, is designed for school age children, and therefore requires a more sophisticated level of language ability.
models that demonstrate proper vocal control over such parameters as pitch, volume, timbre, breath, and articulations are advantageous for children learning to sing.

(b) Singing in monosyllables

Although the words of a song often appear to attract the initial interest of younger children, evidence from research findings suggests that young children sing more accurately in monosyllables (Levinowitz, 1989; Goetze, Cooper, & Brown, 1990). To assist older children to acquire melodic features, songs could sometimes be presented in monosyllables (or “chinning” [Creighton, 1962]). Some music educators advocate the use of “chinning” to heighten the musical characteristics of a song and to facilitate learning musical elements while singing (Richards, 1980; Richards & Langness, 1982; Bennett & Bartholomew, 1997).

(c) Preserving tonality

The maintenance of tonality in singing is mastered by children at a relatively later age. In fact, it was reported to be a problem even in school age children (Goetze, Cooper, & Brown, 1990; Green, 1990). At the initial stage of learning to establish a sense of tonality, children often demonstrate the ability in familiar songs but not in unfamiliar songs. For this reason, experience with singing a familiar song using different tonal centres may assist children to conserve the relative tonal relationships in a song. It may also provide children with abundant stimuli generated from the same archetype of melodic patterns in order that they may advance progressively and learn to maintain tonality in less familiar songs. Singing a familiar song in different tonal centres also provides opportunities to extend the vocal registers of young children. Children in the present study demonstrated that they were able to sing in a much higher range than is often considered suitable. Hence, singing in a comfortable, yet “unconventional” vocal register would be a commendable experience.
(d) Complexity of songs

Evidence from children's vocalizations demonstrates a complexity level not adequately acknowledged in current music teaching resources. Bennett and Bartholomew (1997) propose that choosing a song solely on the basis of technical considerations may not be appropriate because songs lacking some degree of complexity are often not of interest to young children (p. 46). My observations suggest that although repetitive patterns in a song may appear to make it easier for a child, variety and contrast in a song appeal to her. Hence, songs with an appropriate level of complexity provide perceptual challenges and hence, may help to give incentive for a child to repeat singing the songs.

(e) Musical interpretations

When a child masters a preliminary level of singing skills (i.e., when she is able to perform a song with relative accuracy), she often begins to incorporate various musical features in subsequent performances. Children at the 4-5 level often consolidated musical features such as dynamics and tempo to provide contrast for different verses of a strophic song. Not only did a young child demonstrate abilities for musical interpretations, she often appeared to enjoy this new found accomplishment. Thus, experimentation with musical features in singing can be fun and exciting for children perhaps because variations in tempo and dynamics add to the complexity of songs that are already familiar to children, and manipulation of these musical elements might be cognitively stimulating. By varying the musical features of a familiar song, children appear to develop a sense of ownership because they can create and improvise according to their imaginations. Hence, to promote personal involvement in singing, it is important that young children be provided with singing experiences to strengthen their awareness of expressive possibilities inherent in songs.
3. Singing as a dynamic experience

Evidence suggests that young children do not necessarily view singing as an isolated activity. In other words, the total experience of singing may be integrated with other facets of childhood experience.

(a) Singing as a language experience

My observations of children's intermediate vocalizations suggest that it is sometimes difficult to delineate the explicit relations between speaking and singing (White, 1998). In fact, a consensus for the perceptual and acoustical description of "singing" and "speaking" remains elusive. Bennett and Bartholomew (1997) maintain that singing a song is an experience in language (p. 18-23). The parallels observed between some of the milestones in music and linguistic development may be supporting evidence for this position. The present study also points towards the presence of language-specific vocal behaviours in children, suggesting that English and Chinese bilingual speaking children may have different conceptions about singing--this needs further study. Considering the relationships between speaking and singing, however, it would be natural to integrate singing with language experience.

Experience in phonation could potentially benefit the acquisition of both speaking and singing skills because acoustic properties in speech may be considered musical (Papousek, 1981; Moog, 1976), and the patterning of sounds found in songs requires cognitive processing of both musical and linguistic patterns. Songs that have linguistic interest, i.e., that use interesting sounds, or those that have concrete meanings are desirable repertoires for young children. (See Bennett and Bartholomew, 1997, p. 18, for some possible language experience in learning a song.)

(b) Singing as a movement experience

Children were sometimes observed to voluntarily integrate singing with body movements. A child usually begins by singing a song, gradually adding body movements.
Since movements appear to elevate a child’s involvement in sensory experience, she may become too absorbed in moving to continue singing. In other words, my data suggest body movements may become the performance mode of a song, and singing the melody and the words may assume a secondary role. Bruner (1960; 1966) describes the initial mode of knowledge representation as “enactive”. Such a mode is found when a child demonstrates concept acquisition by engaging in overt behaviours in response to a stimulus. Here, externalizing a song in body movement depicts how musical experience for a child may be in a “movement” or “enactive” mode, which holds the musical elements of a song together even when it is not sung aloud (Bennett and Bartholomew, 1997). Connecting movements with singing can therefore be beneficial for young children, especially during the developmental stage when they connect more easily with tactile experience.

(c) Singing as playing

Playing is an important component in a child’s singing activities. Children in the present study displayed a rich spectrum of musical and linguistic behaviours when they played. In fact, all intermediate vocalizations were collected in play contexts (see Table 5.1). In contrast, children’s singing responses to requests of adults displayed comparatively less creativity and personal involvement. Observations of children’s intermediate vocalizations suggest that they produced much more interesting vocalizations when they were immersed in playing because the high level of social interaction in a play context may have induced the need to express and communicate.

---

48 Some music teaching approaches advocate movement activities. For example, the Orff-Schulwerk approach encourages “translating body movement into percussion instrument” and the Kodaly approach combines vocal and movement activities. Both approaches are widely used in elementary classrooms and in music therapy settings.
How can a singing activity be initiated as playing? Singing games are particularly effective in engaging children and thus provide excellent settings for learning. Observations of children's intermediate vocalizations demonstrated the diversity of vocal behaviours in games. For example, Amber's spontaneous singing game at 3;1 consisted of vocal activities including imaginative play with conversations on a toy phone; a repetitive speech sequence; some movement reminiscent of a singing game Ring Around the Rosy; and, playing ascending and descending scalar passage on a toy xylophone while singing tonal syllables. It is, therefore, important to recognize that each of these vocal activities could be integral to the total experience of Amber's game. Thus, singing games can be employed as a foundation for musical experience in childhood.

Suggestions for Future Studies

The present investigation is a small group study in which data were collected in a semistructured fashion. It combined acoustic analyses and ethnographic observations to answer a set of questions related to the vocal development of young children. The following methodological suggestions are discussed because of their merit in enhancing careful documentation of procedures for generating and interpreting data. Caution should however be taken because implementing these suggestions may transform the present investigation into a more experimentally oriented study. It is not clear that sacrificing the ecological validity of an ethnography for a more experimental design may better answer the original research questions I sought to explore.

49 A body of research has been accumulated on the singing games of children in different cultures (Marsh, 1997; Addo, 1995; Blacking, 1967, 1995; Riddell, 1990; Prim, 1989).
50 Lincoln and Guba (1985) called these measures as strategies for establishing "dependability".
1. The language environment of children

That all subjects in the present study attended English-speaking daycare and preschool may have affected their vocal development. It was reported that effects of ambient language are already evident in prelinguistic vocalizations by ten months of age (de Boysson-Bardies et al., 1989; 1992). Although the Chinese-speaking children spoke Chinese at home, their immediate environment outside the family is English. Because the Chinese subjects in the present study had continued exposure to both English and Chinese, an explicit comparison with the English-speaking subjects are somewhat difficult to interpret. Indeed, early childhood bilingualism is a wide-spread phenomenon\(^{51}\) (de Houwer, 1995; Garcia, 1983) adding considerably to this problem. However, employing two groups of children in different countries growing up separately in two entirely different language environments may not necessarily facilitate comparison because many variables would be difficult to control (e.g., cultural and socio-economic status of the families under observation). The present study placed a priority in selecting subjects on the basis of their compatibility in family background (parent’s educational level and socioeconomic status) and the environment outside the family (type of daycare and preschool the subjects attended), possible differences due to interlinguistic influence were not determined by any formal assessment.

2. Drawing benchmarks with standardized norms

Some estimates of the children’s sensory-perceptual communication, auditory development, and cognitive abilities at the outset of the study may be useful. Published norm-referenced measures for evaluating children’s language behaviour may be considered in future studies. Measures of general language performance and particularly measures of phonological development (both production and perception) could be helpful should the

\(^{51}\) It was estimated that nearly half of the world’s population is bilingual (Garcia, 1983).
investigation require more in-depth analyses. Standardized tests such as the *Bayley Scale of Infant Development* (Bayley, 1969) would provide some means for comparison between subjects. Measures of weight and height at different intervals could also be beneficial in controlling physical growth as a variable that affect vocal development, especially if physical growth predicts the development of fundamental frequency.

**Time Line**

Children ages 1;6 to 5;11 appear to be within appropriate age level for investigation of vocal development. However, a larger group of children in the same age level at the start of the study could have facilitated cross-sectional comparisons both within and between age levels.

**Data Collection**

1. Systematize collection procedures

   More systematic procedures should be considered for data collection. A list of standardized questions might be used to initiate vocal responses from children, to assist comparisons. This strategy, however, places restriction on the researcher, who must be able to adapt to the varied and emerging interactions between herself, the parents, and the child during home visits.

2. Vocal tasks

   The singing and speaking stimuli could be more systematically selected to assist analyses of vocal data. A fixed repertoire of songs would enable more detailed comparisons both within and between age levels. Caution should be exercised, however, because such a strategy may discourage *intermediate vocalizations* and diminish the number of idiosyncratic approaches to singing that were observed.

3. Developmental inventory of children

   An in depth inventory of the children's developmental behaviours would be useful. Parents could be asked to provide information, such as: observable milestones of the
child’s language and music development; documentation of the child’s language and music environment; and, a collection of the child’s favorite songs and musical activities throughout the data collection period. Although such an inventory is desirable, it is technically rather difficult to obtain. To expect such a high level of commitment from the parents may be unrealistic, although feasible.

Data Analyses

If a larger sample were collected, sufficient power would be available to use more statistical analyses to explore the differences observed between speaking and singing fundamental frequencies. Measures could then also be used to estimate correlations between variables such as physical development, language environment, and the differences in children’s vocal pitch behaviours. Such statistical analyses could another dimension to the ethnographic observations.

Conclusion

In my investigation of young children’s vocal development, I examined the hypothesis that singing and speaking are two vocal phenomena that exist in parallel along a continuum. The “fuzzy” vocal behaviour exhibited by some children in this study appears to confirm such an hypothesis. This fuzziness calls into question the entire concept of differences between singing and speaking both acoustically and contextually. Investigators have pointed out that a clear distinction between speaking and singing could be rather difficult to establish because the definition of singing behaviour is problematic (Welch, 1994; Welch, White, & Sergeant, 1996). My findings suggest that while a child demonstrates her capacity to assimilate culturally specific vocal behaviours, she does not necessarily initially subscribe to an adult’s assumptions of what constitutes singing and speaking behaviours.

As concerns the development of children’s singing, the level of agreement between the present findings and established research literature is convincing. The intermediate
vocalizations observed in the present study offer a rich account of some trends observed in the linguistic and musical development of children. Although these trends also provide us with glimpses of the young child’s auditory abilities, many research questions remain unanswered. In addition, more investigations into children’s vocalizations are needed to provide a comprehensive understanding of how the first spoken language may affect vocal development of young children.

Drawing upon the observed developmental vocal behaviours of children, some pedagogical implications are proposed. Among suggestions for fostering a rich learning context for children to experience singing, I have emphasized the importance of integrating singing activities into other early childhood experiences. The following quotation cited in DeWoskin (1982) from ancient Chinese literature, Great Preface ("Ta-hsu" 太始), sums up how speaking, singing, moving, feeling, and expressing may be seen as a spectrum of complementary human experience.

Poetry is where the intention goes. Within the mind, it is intention; issued forth in words, it is poetry. Feelings are stirred within. Their external form is in words. When words are not enough, feelings are expressed with exclamations. When exclamations are not enough, feelings are extended in song. When extending them in song is not enough, feelings are expressed unconsciously with waving [lit., “dancing”] hands and bouncing feet (p. 20).

Were children in the present study representative of others at their age level? This question can perhaps, be more aptly answered through further experimental studies which exercise physical or statistical control of variables and measurement to maximize the validity of findings. My investigation sought to document and interpret the vocal development of preschool children based on observations of a broad pattern of similarities in their vocalizations, the complexity of which could not be captured through unmediated knowledge about the contexts.
Locke (1995) explains the requirement for studying language production of young children:

To understand the child's development of spoken language we must come to grips with a great many things. We must figure out how children develop the emotional, social, perceptual, motoric, neural, cognitive, and linguistic capabilities required for the efficient use of language (p. 278).

Following Locke, it may be seen that the complexity involved in understanding vocal development of young children is enormous. Thus, it may be more practical to select a small group of subjects for more in depth investigation. Qualitative inquirers maintain that ethnography aims to study specific cases rather than to develop theoretical knowledge. Hence, ethnography is a training of the capacity for perception to enhance practical wisdom (Schwandt, 1997, pp. 57-60). The merit of ethnography is depicted by Van Maanen (1988) as follows:

[O]ne culture is portrayed in terms of another in an ethnography. It rests on the peculiar practice of representing the social reality of others through the analysis of one's own experience in the world of these others. Ethnography is therefore highly particular and hauntingly personal (p. ix).

The present qualitative and quantitative interdisciplinary study described the rich variety of vocal behaviours of eight Chinese and English-speaking preschool children, how I saw these vocalizations mutate in relation to children's linguistic and musical development, and the contexts in which these developments took place.
BIBLIOGRAPHY


184


Appendix A

Definition of Terms

Musical discourse is a form of communication whereby some agreed upon musical constructs are used. Similar to other forms of discourse, musical discourse is rule governed and internally structured, and therefore, reflects the musical experience of the individual and it constitutes an important part of that individual's musical experience.

Vocalization includes all vocal sounds produced by the human voice.

Speech utterances include "single words" and "sentences". "Single words" are elicited by showing the child a colour picture book and "sentences" are multiple words spoken by the child during conversations.

Singing responses include both learned songs and spontaneous songs.

Learned songs denote songs that are recalled by the child according to a previously introduced vocal model rather than songs that are originally created by that child.

Spontaneous songs include a variety of singing responses that are initiated by the child. Adapting the labelling from Moog (1976), "pot-pourri" song, "imaginative", and "narrative" songs are included in this category. Some of the spontaneous singing may have been attempts to model learned songs rather than originally improvised singing. However, when the judges fail to identify a specific song the child might be attempting to model, it is classified as spontaneous singing.

Story-telling is a type of vocal response elicited by using the child's storybooks as the stimulus. Story-telling has a higher level of emotional intensity than speech utterances, and is characterized by exaggerated inflections and pitch characteristics.

Intermediate vocalizations are found when the vocal responses appear to be somewhat ambiguous according to the daily usage of speech and song. Two main criteria are used for the present classification: (1) the acoustical characteristics of the vocalization--when there
is a mixture of sung and spoken utterances in the vocalizations, and (2) the communication intent—when the vocal utterances do not correspond to the stimulus as anticipated.

**Vocal babbling** is unintelligible vocalizations that resemble babytalk. It may consist of repeatedly uttered speech phonemes, sung melodic fragments, and a variety of vocal articulations.

**Pot-pourri songs** consist of learned song fragments chained together. During spontaneous singing, a child often manipulates musical and linguistic resources by freely combining different songs.

**Pitch** is the perceived quality of a periodic sound chiefly as a function of the fundamental frequency. However, when the fundamental is not present, any two or more adjacent harmonics below the tenth partial may be effective conveyors of the fundamental pitch. This is possible because a pitch may be extracted on the basis of the spectral resolution of individual components in a complex tone stimulus. Since pitch perception may be affected by a number of conditions, pitch has also been used as a psychoacoustic term denoting phenomena such as "residue pitch" and "virtual pitch" (Pierce, 1992). In music, pitch is defined as an "attribute of auditory sensation in terms of which sounds are ordered on a musical scale" (Moore, 1997, p.3). Western music adopts a standardized association of some particular frequency to a pitch name (e.g., A1=440 Hz).

**Fundamental frequency** ($F_0$) is the lowest sinusoidal component of a complex tone. In music, the fundamental is the lowest tone in a harmonic series (Moore, 1997; Truax, 1978).

**Tonal language** (also tone language) is a language characterized by variations in relative in pitch or tone which distinguish the meanings of words of the same or very similar written form or the same vowel sound or sounds (Pei and Gaynor, 1954, p. 218). "In a tone language it is not the absolute pitch of the syllables that is important but the relations among the pitches of different syllables" (Fromkin, Rodman, Hultin, & Logan, 1997, p. 186).
Nontonal language (also intonation language) is a language characterized by intonation contour of pitch pattern over the sentence or phrase which may be grammatically distinctive. “In English, syntactic differences may be shown by different intonation contours” (Fromkin, Rodman, Hultin, & Logan, 1997, p. 215).
Appendix B
The Ancient Chinese’s Concept of Harmony

Several meanings are attributed to He ( 和 ), the word for Harmony in Chinese. All of them are concerned with either theories of tuning or extra-musical philosophies: (1) just intonation—the beatless tuning of a tone according to huangzhong ( 黃鐘 ), which is the ancient Chinese measurement of the fundamental, or sometimes known as the root pitch; (2) the achievement of “purity” in tuning the 12 lu ( 律 , the 12 tones used in Chinese scalar structure) derived from the natural resonance of the overtone series; and, (3) the achievement of order between man, and his family, society and the universe as manifested in the acoustical phenomenon of just intonation of pitches. Because of these meanings attributed to he, ancient Chinese allied their philosophical conception of cosmic harmony to the discussion of music and theorizing acoustical principles in sound. This interest reached its height during the Han dynasty, which was exemplified in the Confucian doctrine of music edification, and the establishment of yuefu ( 音府 , “Music Bureau”, established around 112 B.C., closed at 6 A.D.) to administrate musical practices under the auspices of the court.

The ancient Chinese concept of cosmic harmony may even be reckoned to be in a form of silence. Liang Ming-yue (1985) discusses the idea of “soundlessness” to be the aesthetic object of harmony in Chinese musical aesthetics. Soundlessness was argued to be the “ultimate music” termed zhìyue ( 至樂 ). According to Wu and Liu (1983), this idealized form of music, zhìyue, was originally proposed by Ji Kang, 晉康 (223-63 A.D.) in his Sheng-wu -ai -le Lun ( 音無哀樂論 , “Discourse on the Nonemotional Nature of Sound”). Ji Kang differentiates zhìyue from yinzheng. Yinzheng is the “ordinary” music we come into contact on a daily basis; it exists in physical form and therefore we can hear it. Zhìyue, on the other hand, is soundless. It could only be achieved when virtue is attained through he ( 和 , harmony), both internally within the person, and externally
among a virtue ruler and its society. In fact, the soundless state is a supersensory realm alluded to in Chinese poems, stories and music. The following quote from The Inner Chapters of Chuang Tsu (周易) may explain why zhiyue is soundless.

Duke Ai asked, “What do you mean by achieving full harmony?” Confucius said, “Life and death, profit and loss, failure and success, poverty and wealth, value and worthlessness, praise and blame, hunger and thirst, cold and heat—these are natural changes in the order of things. They alternate with one another like day and night. No one knows where one ends and the other begins. Therefore, they should not disturb our peace or enter our souls. Live so that you are at ease, in harmony with the world, and full of joy . . .” “And what is this the lack of outward manifestation of virtue?” Confucius said, “Balance is the perfect state of still water. Let that be our model. It remains quiet within and is not disturbed on the surface. Virtue is the attainment of perfect harmony. Because virtue has no outward form, nothing can escape from it.” (Trans. Feng & English, 1974, p. 105)

Zhiyue is therefore, an idealized, metaphysical form of music which achieved the objective of he. Like still water, it is at ease and self-contained, and therefore has no outward manifestation. Discussion of this “inner presence” realm of music (as opposed to the outward expression of music in acoustic medium) may be found in Yueji (乐记, “Annotations on Music”) in the Liji (礼记, “Book of Rites”):

Music is formed in the heart. Tones [pitch, timbre] are the shape in which music is expressed. Elegance [ornamentation] and rhythm are the decoration of the tones. The Superior Man takes the feelings in his heart as basis; he gives them shape in music; and then he gives this music its final form. (Trans. Van Gulik, 1941, p. 24)

The presence of this metaphysical form of music is said to be “soundless”. Liang (1985) explains that zhiyue is sounds that “outwardly ‘appear soundless’ because they exist in extended (supersensory) perception” (p. 173). Since ancient Chinese stated that
“soundless tone is the greatest” and that “soundlessness gives birth to the five tones.”

(Huainanzi edited by Liu An of the 2nd century B.C.), According to Liang, “soundlessness” could be defined as “a sound idea existing in the mind that could potentially give rise to a sound-filled composition” (p. 173). From the above discussion, we can infer that the supersensory perception or sound idea existing in the mind is a psycho-aesthetic concept of harmony.

Harmony as achieved by the Western’s sense of simultaneous sounding of tones is therefore, of secondary importance to ancient Chinese. Understanding how harmonious music is realized by ancient Chinese, we can appreciate that even in a Chinese ensemble, the instruments typically play in unison. The achievement of any perceived harmonic texture in Chinese ensemble in a Westerner’s sense may only be incidental. It is often what Liang called a “by-product” or a “polychronomic texture” resulting from the practice of instrumentalists doubling the melodic line with improvised ornaments.
Appendix C

The Effects of Chinese Linguistic Tones on Some of its Vocal Genres

To provide a succinct review on the literature of vocal music for a civilization with 5,000 years of history is beyond the scope of the present study. (See Thrasher, 1993 for bibliographical references.) I am, therefore, limiting the discussion to briefly address some interesting aspects that point to the integral relationship between speaking and singing. I shall focus on the relations between the tones of the Chinese language and some Chinese vocal genres. My discussion is guided by Chao (1956) and Yung (1983a; 1983b; 1983c; 1989). While Chao (1956) discussed the function of the pitch of the voice in speech and various forms of Chinese vocal genre, Yung (1983a; 1983b; 1983c; 1989; 1991) researched the role of linguistic tones in the creative process in Cantonese operas. Some of their research findings provided valuable insights and clarification of crucial aspects for my present investigation of young children's vocalization.

Several vocal genres were identified by Chao: (1) Singsong, (2) Chant, (3) Recitative, and (4) Songs--Tonal and Atonal Composition. It is important to note that the definitions of these terms are largely operational because Chao made limited reference to their general usage in musical context. Citing directly as Chao defined these terms is therefore more appropriate in clarifying the present discussion. A discussion of Yung's findings on Cantonese operas can be found in (5) Melodies in Cantonese opera.

(1) Singsong

It is a style of vocalization closer to speech than to song. It is a term invented by Chao.

---

52 Chao was not only a professor of Linguistic but a prolific composer who pioneered composed music in modern China through composing melodies that preserve the unique tones and contour of the Chinese language (Yang, 1984, p. 498-499). (Refer to Wang, 1983, on the accomplishments of Chao.)
There is a style of use of the language which is neither the algebraic addition of tone and intonation, resulting in ordinary speech, nor singing of a musical melody, but something intermediate, which is based largely on the phonemic tones of the words, spoken in a stereotyped manner. . . . In the singsong style of speech, each tone receives its usual full value practically without intonational variation. It has a monotonous singsong effect. The popular impression of such speech or reading is that it is "intoned", whereas the fact is that it is speech minus the element of intonation. (p. 53)

According to Chao's description, singsong is a way of speaking without attending to the ordinary use of speech contour. Since this is an intermediate vocal form not found in ordinary speaking or singing, one may question why Chao would name it "singsong", which, on the surface, may appear to be a vocal form more closely related to singing. The key to this question lies in the intervals used in substitution of the speech contour. Since the contour of the speech is absent, the phonemes are spoken "all on one pitch until the second half of the last syllable in each line, where the voice slides to a minor third below" (p. 54). It is the uniform use of the minor third interval that conjures up a musical impression in the listener's mind. The descending minor third becomes a recurring motive that suggests a sense of melodic patterning. However, whether the speaker of this style of speech has any musical intention is questionable. Chao cited two examples of the singsong style: vendors of Tiger Balm (a traditional Chinese ointment) praising its virtues on river boats in Southern China, and a form of speech used in rote-learning where Beijing schoolchildren recite aloud in unison (p. 54).

Reporting as an informant, those experiences mentioned by Chao are all familiar to me. I remember I used to hear street vendors using singsong to sell all sorts of household products in a quiet school and residential area in Hong Kong. That they "sang" everything to a minor third interval amazed me, and directed much of my attention away from my
teacher’s lecturing during my elementary and secondary schooling. During that time while I was a schoolchild, I also practiced the singsong style of reciting on a daily basis. Whenever our whole class was instructed to read aloud, the singsong style would emerge automatically. As students, we enjoyed it very much as we could relate the style of speech to singing in unison. It was fun because we could apply it to both Chinese and English (since the school subjects were taught bilingually in Hong Kong), and sometimes to tease the teacher, we singsonged in an accelerated mode only to restart again in a more acceptable speed. Functionally speaking, our singsongs were analogous to speeches because we recited the content of the textbooks and the class notes. From the aspect of enjoyment, we liked it for its musical qualities— that is, as a class of 40, we could perform singsong in unison, and manipulate the tempo as if it was a piece of music. It was, however, to the dismay of some teachers who expressed concern because they believed that a lack of speech contour was not conducive to understanding the content we were to master. The interesting point here, is that as schoolchildren, we were not formally taught or even encouraged to recite in the singsong style. This style of speech was probably picked up from the auditory experience we had accumulated from being exposed to it under many different situations. In southern China and in some of the Chinese restaurants in North America, one may still hear waitresses calling out the names of the dim sum as they roll the dim sum cart around. Occasionally, they may singsong the dim sum names to draw the customer’s attention.

However, Chao pointed out that the singsong style of speech is not unique to tonal language. George List (1971) called this form of heightened speech “intonational recitation”. His examples are the jumping rope rhymes of American children and the recitation of poetry by small children (p. 261-262). List described intonational recitation as revolving “around a central speech monotone which is ornamented with one or two auxiliary tones” (p. 262). On the surface, Chao and List were describing the same type of
vocal phenomenon found in different cultures. However, a subtle difference lies in the fact that the Chinese singsong could not achieve the same degree of monotone as in List's example of intonational recitation by the English-speaking children. Even when Chao writes that in the singsong style, the phonemes are produced "all on one pitch", what he really means is that the phonemes are recited to a relatively stable pitch level while the corresponding tones of the phonemes are still maintained. Otherwise, the meaning in the singsong would be obscured without proper linguistic tones to convey the lexical. It is only in English and other nontonal language where linguistic tone is absent, that a true monotone recitation may be found. Chao concluded that since "singsong is completely determined by tone, each dialect will of course have a different system of singsong" (p. 54).

(2) Chanting

Contrary to singsong is chanting in which Chao argued linguistic tone plays an important role but not as the sole determinant of the pitch level executed. As Chao explained:

Chanting is largely, but not unambiguously, determined by tone. Given the words, their tones allow a limited range of variations within which the reader or chanter, after a sufficient amount of listening and practice, can improvise and does not, usually cannot, repeat the same tune each time he renders the same text. (p. 54)

Chao's definition of chanting in Chinese as improvised tunes is different from the ordinary usage of the word "chant", such as when it is used to signify the type of vocal recitation with fixed melodies performed by Buddhist monks. Chanting as signified here, is a more advanced and sophisticated form of recitation, often applied to Chinese poems and classic literature of antiquity. Two criteria determine the "tune" of a chant: (a) the different systems of rhyming and metrical construction of the literature; and (b) the systems of tones in different Chinese dialects according to geographical areas. Hence, Chao
explained that for the same poem, different versions of chant "tune" exist. Chao used the word tune to signify the musical quality in this type of chanting because of the flexible use of syllable duration and the alteration of linguistic tone for expressive purposes. Although recited with a general reference to the natural inflection of speech, this type of chanting exaggerates contrasts of dynamics, syllable lengths, recitation speed, and pitch levels not found in conversational speaking style. The chanter assumes the role of a storyteller similar to that of an instrumental performer of a piece of Western music. She is expected to deliver the poem or literature in an appropriate chanting style with much the same level of sophistication as we would expect from a piano performer of a prelude by Chopin. To be able to exercise taste in the manipulation of dynamics, rhythm, tempo, and pitch in chanting would require insights into the literature much the same way as training and taste is required to play a Chopin prelude in a stylistically appropriate manner. Chinese chanting as a vocal genre (whether it be classified as speech or song) is therefore, a more artistic and musical genre when comparing to that of speech in conversation style and recitation in singsong style.

With Chinese chant (as defined by Chao) emphasizing the manipulation of acoustic properties similar to the treatment of musical elements in a song, I speculate that chanting is an ancestor, or at least, a close relative of Chinese quyi. Quyi, or narrative, dated back to the Warring States period (475-221 B.C.). According to The Dictionary of Chinese Music (Yang, 1985), about 300 styles of quyi were reported, among which some are narratives that consist of speaking only (e.g., Pinghua 聽話, Xiangsheng 相聲, Shuoshu 說書, Shuohua 說話, and Pingshu 聽書, all of these are story-telling narrative genres). The majority of the quyi genres are singing narratives, consisting of both spoken prose and sung verse with instrumental accompaniment (p. 322). The most outstanding feature of quyi is the integration of music and the spoken language, and hence regional dialects shape the stylistic differences among narrative genres. Liang (1985) pointed out that in Suzhou
tanci (蘇州 弹词), and especially in the jingyun (京韵) style of dagu (大鼓),
the speech tonal patterns dominate the song melody (p. 144, 147). According to The
Dictionary of Chinese Music (Yang, 1985), the close relation between Jingyun dagu and the
Beijing dialect is evident in a number of ways. (a) In the fast movements: the syllables are
short with relatively fewer changes of pitch, and therefore resemble chanting. (b) Speaking
and singing are intertwined in the performance; unpredictable shifts between song and
speech during climatic passages are common. (c) The melody is constructed according to
the four basic linguistic tones of the Beijing dialect and the rhythmic inflections of the
Chinese language (translated from p. 201). As Liang (1985) stated, “the jingyun
melodization is truly extended speech where each word is distinctly audible” (p. 147). In
fact, the name of this narrative genre tells it all: Jingyun (京韵) literally means “the
inflections of the Beijing dialect”.

(3) Recitatives

They are found in Chinese operas. Chao (1956) provided the following account:
In the traditional form of Chinese drama, which is always of the aria-recitative type, the
recitative part follows largely the tones of the words, but tones of a special artificial
dialect known as the Chung-chou pronunciation or Central pronunciation. It differs
from singing in being more flexible and allowing more than one way of rendering a
given text and from chanting in that it is not intoned into easily located notes in the
musical scale, but makes frequent use of gliding pitch, as in spoken tones. This
artificial dialect is used only in dramatic dialogue and never actually spoken in real life
in any locality. It differs from Mandarin in the tone values and from normal speech in
any dialect in great uniformity of syllabic length, a slower overall tempo, and slower
changes in volume (with sweeping changes over phrases rather than quick changes on
individual words), all of which are characteristic of the declamatory as against the
intimate (conversational) style of dramatic presentation. (p. 57)
Based on this description, we can infer this type of Chinese operatic recitative to be an intermediate form of speech and singing. On the one hand, although it is not found in daily speech, one could not preclude it as a mode of speaking because of the close observation of the linguistic tones. On the other hand, its range of musical treatment (uniformity of syllabic length, a slower overall tempo, and sweeping change of volume over phrase) resembles features commonly found in songs. However, this type of recitative may not fit easily into a song model because of its lack of definite scalar structure. Hence, I would contend that it is a good example with which to demonstrate how the boundaries of speech and song may be “fuzzy” in the vocalizations of some linguistic communities.

(4) Songs--Tonal and Atonal Composition

Chao (1956) discussed songs under the titles “Tonal Composition” and “Atonal Composition”. He operationally defined “tonal composition” as song composed according to the linguistic tones of the Chinese language (p. 57). “Tonal” in this context implies a compliance with the linguistic tone of the Chinese language, not tonality as used in the Western musical context. “Tonal composition” was identified by Chao as composed song where performers were expected to sing according to the notation but with limited freedom to embellish the melody. As a result to the conformity of linguistic tone in writing melodies, somewhat “stereotyped” forms of melodies were found in Kunqu (昆曲), or Kun operas (p. 57). Liang (1985) provided a detailed explanation about the relationship to speech and song in Kun operas:

The linguistic importance associated with Kun opera is significant. The same “labelled tune” used as a melodic source can have altered appearances because the correlating text is different, which would then impose its own linguistic-musical synchronization. Often, in the event there is a difference between a melodic tone and a linguistically-designated tone choice, the linguistic tone justification would prevail. That is, the
musical melody follows the linguistic tone direction such as the inflected rising tone, falling tone, straight tone and variations. A beginning Kun opera singer often makes the comment that the tune is easy, but correlating it with the words is never easy enough. (p. 240)

It can be seen that in some Chinese vocal genre such as kunqu, linguistic tones are proposed as the foundation of melodic movement. This concept of manipulating melodic movement is parallel to Levis's (1963) hypothesis that tonal movements in the Chinese language are closely observed in musical compositions. In Levis's wording, "these elements of tonal movement in speech find conscious musical expression in terms of the tones of the scale" (p. 23). That linguistic tones form an integral component in some Chinese melodies is also consistent with the concept of chi (⋵), which symbolizes an animated force that harmonizes the words and the melody in a song.

It is important to point out that not all Chinese songs are composed with meticulous consideration to linguistic tones. Chao labelled those types of songs as "atonal composition". By "atonal" he implied songs that are composed with little attention to the linguistic tones. These types of compositions are mainly contemporary, where the agreement between melodic movement and linguistic tone is scarce. Chao asserted that it is impossible to separate the functions of pitch in speaking and singing and prompted the revival of composing melodies that correspond to the linguistic tone of the Chinese language.

(5) Melodies in Cantonese Opera

The reciprocal connections between melody and the prosody of speech may be illustrated with Cantonese opera. To understand the interactions between the prosody of Cantonese speech and the sung melodies, a preliminary grasp of the nature of some specific features in Chinese operas would be helpful to interpret the findings of an ingenious study by Yung (1989).
It was reported in a survey conducted during the 1950s that China has over 300 different styles of regional operas, each with its own repertory using preexistent tunes (Yung, 1991). Drawing an analogy with Western music, such a practice of using preexistent melodies is somewhat similar to the use of Medieval plainsong as the *cantus firmus* in Renaissance motets and masses. Since only a limited number of preexistent tunes are shared in Chinese operas within a regional style, and these tunes were transmitted orally through generations, they are known by their literary titles rather than by proper notation. A considerable proportion of Chinese regional operas are, therefore, without musical notation, and they are frequently performed without rehearsal. Hence, part of the creative process is carried out by the opera performers during performance, and a certain degree of improvisation is expected from these artists. In this way, a particular preexistent tune may be sung in many different operas, and the same tune may also be used several times set to different text settings within a single opera. The linguistic tone of the dialect sung, therefore, plays a distinguishing role in determining the choice of tune and the appropriate use of improvised embellishment during performance.

Using an operatic tune sung to nine different text settings transcribed from performances of several operas by different singers, Yung (1989) illustrated a consistent direct relationship between the melodic contour of a Cantonese operatic tune and the linguistic tonal contour of the Cantonese dialect. In each of the nine slightly modified versions of the preexistent tune, the linguistic tones of the lyrics are matched with the melodic patterns of the tune (p. 413). Since the choice of pitches sung to syllables belonging to the same linguistic tone is maintained throughout the nine versions of performance, it can be seen that the relative pitches in Cantonese dialect are reliably paralleled with considerably well defined musical pitches in operatic singing.

While the organization of pitch and tonal patterns in speech may be viewed as a communication tool implicit to a particular linguistic community, the choice of melodic
patterns used in Cantonese operas is evidently language based. These musical compositions and performances may be viewed as implicit to their own linguistic community because the linguistic tones play a vital and central role in the creative process. However, as Chao (1956) pointed out, the practice of matching melodic patterns to the corresponding linguistic tones is not found in all forms of Chinese vocal music. On this point, A. R. Thrasher (personal interview, December 1997) commented that the linguistic tones are often loosely observed in Chinese music composed after the 1920s.
Appendix D
Vocal Forms in List's Hemispheric Chart

Classification of vocal forms may be made according to a variety of perspectives. List (1971) identified three major characteristics common to both speech and song: (1) vocally produced, (2) linguistically meaningful, and (3) melodic. Although List acknowledges problems involved in the first two characteristics, his classification system of vocal forms is based solely on melodic characteristics. Using this criterion, List sets speech and song at the end of the north and south poles of his classification chart, while other intermediate forms are positioned according to their relative emphasis on intonation or melodic structure. Vocal forms that place relatively more emphasis on melodic structure are positioned closer to the song pole; forms that place relatively more emphasis on speech intonation are positioned closer to the speech pole. (See Figure 6.1 for the chart List used to classify forms intermediate to speech and song.)

Some explanations of the terms used in the chart may be helpful. The following discussion is based on List's (1971) article and *The New Harvard Dictionary of Music* (Randel, 1986). Although these definitions had some influence on my approach to data analyses, I also realized the need to supplement List's categorization with other forms of data analysis.
Figure 6.1
List's Chart for Classifying Vocal Forms Intermediate to Speech and Song
**Speech.** In the general sense, is the act of speaking, an oral form of linguistic discourse involving spoken communication in different forms such as words and conversation. For the purpose of making comparisons with other forms of vocalizations, List defines it in a narrow sense, as “casual utterance, as in conversation” (p. 255). He identifies the “melody” of speech as produced by: (a) either the intonation or inflections common to the particular language; (b) the tonal structure if the language possesses phonemic tones; and (c) a combination of the intonation or inflections and the tonal structure when both are present in the particular language (p. 254). List differentiates between tones and intonation in language: tones are meaningful at phonemic level—he called it meaningful “in the dictionary sense” while intonation is also used to imply meaning but it is “less susceptible to exact analysis than phonemes or tones” (p. 254).

**Song.** In the general sense, is a generic term for “a piece of music sung or composed for singing” (Randel, p. 1278). In Western cultures, the term often signifies a vocal composition with a melody and a verse text. List defined it “as a form exhibiting relative stable pitches, possessing a scalar structure at least as elaborate as the heptatonic, and showing little, if any, influence melodically of speech intonation” (p. 256). I can see at least two problems when adapting these characteristics of song as defined by List to categorize the vocalizations of children in my study. First, young children have a limited singing range. Their singing range is often reported to lie within an octave, or as narrow as a perfect fourth at the age of two (Jersild & Bienstock, 1934). Hence, it could be difficult to find children in this study singing songs that exhibit scalar structure of a heptatonic, which is a collection of seven tones as found in a diatonic scale. Second, since some children in the present study are tonal language speakers, it is possible that the tonal movement of phonemes may have influenced the melody sung. In this sense, whether the

---

53 There is a very extensive linguistic literature on melody/intonation, which is not referenced here (see, for example, Bolinger [1955]). Linguistic interpretations do not coincide with musical interpretations.
pitches in their songs could be considered as stable or not might be a matter of subjective judgment based on language and cultural practice. Those two problems imply that classification of vocalizations solely as a function of pitch may not be satisfactory.

Two characteristics in the use of pitch in speech and song, *Sprechstimme* and *Monotone*, are labelled respectively at the west and east poles in List’s classification chart. *Sprechstimme* may be defined as “[German., speaking voice, speech-song]. A use of the voice midway between speech and song. In general, it calls for only the approximate reproduction of pitches and in any case avoids the sustaining of any pitch” (Randel, 1986, p. 804). In List’s description, it is “the type of elevated or heightened speech characteristic of the melodrama” (List, 1971, p. 263). List uses an example of *haka* by the Maori which has a range of approximately one octave and a perfect fifth (p. 264). *Monotone* may be defined as “a single tone on which a liturgical text . . . is recited. An inflected monotone is a formula for recitation in which a single tone predominates but in which divisions in the text may be marked by brief deviations from the principal tone” (Randel, 1986, p. 507). List cites a *Thai* Buddhist chant as an example of a pure monotone (List, 1971, p. 263).

**Intermediate forms.** In its general sense, speech includes a wide variety of forms. Nevertheless, List categorizes forms other than conversation as “heightened speech”. List defines it as a form “somewhat intermediate to speech and song”. According to List, examples of intermediate form are dramatic representation, the delivery of a sermon, or the telling of jokes and tales (List, 1971, p. 255-256). Based on a comparison of (a) negation of intonation, (b) expansion of scalar structure, (c) stability of pitch, and (d) expansion of intonation, four main intermediate forms are identified by List. These are:

1. **Recitation.** In Western culture, is a form of speech read aloud in public, usually in a more formal declamation such as an announcement. List places it between speech and monotone to indicate a lesser use of speech inflection melodic contour to convey meaning. List cites the example of a jumping rope rhyme for this form (p. 262).
2. Intonational recitation. List places this between speech and Sprechstimme to indicate an expansion of intonation. This may be found in dramatic speech delivered in a play, in story telling (especially to a child audience) or to a piece of prose or verse. The use of intonation and speech inflections is emphasized in a very dramatic manner to convey emotions. List uses the singing of the women of Palau, one of the Caroline Islands in Micronesia as an example. Its range is approximately a major third (List, 1971, p. 263).

3. Chant. This may be defined as "a monotonous tone of voice; singsong mode of speaking" and "a simple liturgical song in which a string of syllables or words is sung to each tone" (Neufeldt, 1991, p. 234). List describes a monotonic chant as one that "display a few auxiliary tones and, in particular, descending vocal glides" (List, 1971, p. 257). A common example is what List called "auctioneering," the form of verbal communication used by the auctioneer, which often resembles the characteristics of a monotonic chant with few auxiliary tones and some types of melodic cadences (List, 1971, p. 259). More "song-like" chant can be found towards the south pole of the classification chart. List cites the diatonic songs of the Vedda, tritonic taunts of American children, and tetratonic Bulgarian folk songs as examples of chants that show increasing complexity of scalar structure (p. 264).

4. Intonational chant. According to List, intonational chant has less intonational complexity but exhibits an increased level of pitch stability. List cited the Hopi "announcement" as an example, which uses two unstable pitches approximately a fourth apart, with vocal glides and the sharpening of the upper chant tone (p. 256).

Limitations of List's Classification System

The rationale for List's examination of pitch as the sole criterion to classify vocal forms may be traced in research literature. Pitch has always been a dominant concept in Western music culture and has therefore, been of central concern in research on singing. Since there is no incentive to refrain from what we have been exposed to and trained to
evaluate as significant features in western music, pitch accuracy is an important criterion in vocal performance. Young children, however, often engage in both speech and song babbling activities when they explore and exercise their voice. Classification such as chanting, rhyming, speaking, and singing emerges only when adults attempt to assign meaning to the vocal acts, while a number of other factors may have been overlooked. For example: A child may not have developed the fine vocal skills to make explicit differentiation between linguistic and musical discourse to an adult's satisfaction.

Furthermore, a child often has very different perceptions of the various stimuli compared to those of an adult. Hence, it may sometimes be difficult for adult observers to fully understand a child's communicative intent. Since the intent of any form of discourse is to communicate and express, classification of vocal forms would need to consider its meaningfulness for linguistic or musical implications.

I did not find it adequate to simply analyze vocal forms, and then categorize them according to the pitch construct implemented by the performer. Therefore, for the purposes of the present study, I collected several forms of data to establish intersubjectivity in the process of classifying and analyzing vocalizations.
Appendix E

The Daycare Centres

The documentation on The University of British Columbia daycare system provided me with the information on its philosophy and details of its program. I found many parallels between its philosophy and the Montessori approach: "We believe that children learn through their senses, through their feelings, through their intellect and through interaction with other children and adults. . . . we believe that the child's self image is the most important part of the learning process. Children need to repeatedly perceive themselves as worthwhile, capable and significant, and therefore their experiences at school should be positive and reinforcing. . . . The educators work closely together to individualize the programs so that the needs and strengths of each child can be maximized" (Parent Handbook: Preschool, p. 1). All daycare centres attended by children in the present study offer dynamic play facilities, such as, puzzles, blocks, water play, and books. They all schedule outdoor play on a daily basis, and organize field trips.

I had the opportunity to visit one of the university daycare centres the year before my study commenced. While the physical set up of the centre can be described as "Montessorian", with child-sized furniture and rugs to define activity areas, some of the facilities, in particular, its outdoor space, exceed the province's legislated standard. I was most impressed with the unique ambiance when I taught in one of the centres. Since the facilities are centrally administered, the general layout of each daycare centre is identical: each centre is a detached house with windows that allow plenty of natural sunlight; the heart of the house is the central play area with a high ceiling (more than 20 feet); the backyard is not just a play area, its set up resembles a home garden with inviting details that suites a child's fancy (e.g., some fireman's gear, and a fire hydrant at a corner podium).

Although I have never visited the Montessori preschool which two of the children in my study attended, I have seen photographs of the institute. The abundance of toys and
materials for manipulation is impressive. The classroom atmosphere is also inviting to a child; it is colorful, with playhouse and other designated areas for dramatic play.

Despite the similarity between the university daycare centres and the Montessorian daycare centre, two main distinctive features are found in the University daycare system: (1) a parent advisory committee is in place to maintain a philosophy consistent with the needs of the families served, and (2) the University daycare system is sometimes used as a research site by graduate students and faculty members. The parents may be invited to participate in some research projects, and observations of the children with parental consent are anticipated.
### NAME: Clare
### Date: May, 98
### Age: 4;1

<table>
<thead>
<tr>
<th>Event/Task</th>
<th>Remarks/Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naming objects from picture books</td>
<td>The words she uttered were relatively short but she could label practically all the objects and colours.</td>
</tr>
<tr>
<td>Reading a story book</td>
<td>Needed help from mother because she had not been reading that book of nursery rhymes for some time. Shy, I have to look away and pretended that I was not paying any attention to her. Clare needed constant emotional support from mother; mother gave lots of encouragement and positive responses.</td>
</tr>
<tr>
<td>Conversations</td>
<td>Clare had a birthday party at home the day before. She also talked about her trip to fair. She responded well to my questions and expressed excitement, then became shy (self-conscious?) and was reluctant to talk (she buried herself in her mother’s clothes).</td>
</tr>
<tr>
<td>Baby Beluga</td>
<td>Sang one of her favourite song briefly (about 2 phrases) upon mother’s prompting. (Still staying physically close to her mother.)</td>
</tr>
<tr>
<td>There is Music</td>
<td>Sang with a sense of tonality and contour but some melodic intervals were inaccurate. Her singing got softer and softer in volume as she became self-conscious again.</td>
</tr>
<tr>
<td>Conversations</td>
<td>Mother discussed with Clare the gifts she received from birthday party. (Clare smiled throughout the conversations)</td>
</tr>
<tr>
<td>Lazy Daisy</td>
<td>Sang with modulations and some inaccurate melodic intervals but the song was recognizable.</td>
</tr>
<tr>
<td>Rig-a-Jig-Jig</td>
<td>Accurately sung, both melodically and rhythmically.</td>
</tr>
<tr>
<td>Away in a Manger</td>
<td>Sung the general melodic outline of the song but the phrases were quite fragmented. She was reluctant to sing aloud (as if she was trying to prevent me from hearing her voice).</td>
</tr>
<tr>
<td>Playing</td>
<td>Played with a toy phone briefly and soon lost interest.</td>
</tr>
<tr>
<td>Question &amp; Answer</td>
<td>Mother initiated a game-like activity to elicit conversations. Discussed about Pets.</td>
</tr>
</tbody>
</table>
Appendix G

Assessment Sheet for Judges

The following questions are from the assessment sheets:

(1a) Describe how the child responds to the adult’s invitation to sing a song.

(1b) Describe how the child responds to the adult’s invitation to engage in a conversation.

(1c) Describe how the child responds to the adult’s invitation to read a book.

(2a) What did you observe in the child’s vocalizations to convey your sense of her achievement in singing? Give reasons to justify your answers.

(2b) What did you observe in the child’s vocalizations to convey your sense of her achievement in speaking? Give reasons to justify your answers.

(2c) What did you observe in the child’s vocal production to convey your sense of her achievement in vocalization? (Here, “vocalization” is considered to be the form of vocal production which you think neither the terms “speaking” nor “singing” can best describe the phenomenon you heard.) Give reasons to justify your answers.

(3) Please provide a detailed description and perceptual analysis on ALL of the following:

(a) aspects of musical pitch, including but not limited to: contour, tonality/scalar structure, melody, intervals, and range.

(b) culturally appropriate singing or speaking style, such as: tessituras, intensity, timbre, articulation, and any other stylized singing or speaking behaviours such as phrasing and dynamics.
Appendix H

Music Scores for a Selection of Children’s Songs

Eency Weency Spider

1. The een- cy ween- spi- der went up the wa- ter spout.

6. Down came the rain and washed the spi- der out.

10. Out came the sun and dried up all the rain, and the
een- cy wee- cy spi- der went up the spout a- gain.

Ring Around the Rosy

Ring a- round the ro- sy, pock- et full of po- sies.

Ash- es, ash- es, we all fall down.
Teddy Bear

1 Teddy bear, teddy bear, turn around,

5 Teddy bear, teddy bear, touch the ground.

9 Teddy bear, teddy bear, show your shoes,

13 Teddy bear, teddy bear, that will do.

The Farmer in the Dell

1 The farmer in the dell, the farmer in the dell,

6 Heigh-ho the der-ry-o, the farmer in the dell.
The Ducklings

1. See the little ducklings swimming here and there,

Heads are in the water tails up in the air.

9. There's a little turtle swimming down below,

Tries to catch the ducklings but he is too slow.

Bow, Bow, O Belinda

1. Bow, Bow, O Belinda, Bow, Bow, O Belinda,

5. Bow, Bow, O Belinda, won't you be my darling?
Muffin Man

1 Oh, do you know the muffin man,

6 the muffin man, the muffin man?

11 Oh, do you know the muffin man

16 that lives in Drury Lane?
The Wheels on the Bus

1 The wheels on the bus go round and round,

6 round and round, round and round. The

10 wheels on the bus go round and round

14 All o-ver town!
As I was walking down the street, down the street, down the street, a pretty girl I chanced to meet Hi-O! Hi-O! Hi-O!

Rig-a-jig-jig and away we go, away we go, away we go,

Rig-a-jig-jig and away we go, Hi-O! Hi-O! Hi-O!