THE EFFECTS OF TRAINING WITH MALE AND FEMALE VOCAL MODELING ON MELODIC SINGING ACHIEVEMENT OF GRADE ONE CHILDREN

by

ESTHER HO SHUN MANG

B.A., Dalhousie University, 1984
Dip.Ed., The University of British Columbia, 1992

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE OF GRADUATE STUDIES

(Department of Curriculum studies)

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April 1995

Department of Curriculum Studies

The University of British Columbia
Vancouver, Canada

Date April 25, 1995

DE-6 (2/88)

Abstract

The purpose of this study was to investigate the effects of training with two different types of vocal modeling, the male and female, on the melodic singing achievement of grade one students who possess high or low tonal aptitude. Eighty-three children from two schools participated as subjects.

After the *Tonal* subtest of the *Primary Measures of Music Audiation* was administered to all children, two criterion songs, one each in major and minor tonality, were taught by rote for five weeks. Forty-three children received music instruction with a female vocal model and 40 children received music instruction with a male vocal model. Each child's singing performance was recorded individually and evaluated by three independent judges using a five-point rating scale. To strengthen the reliability of the singing scores, acoustic analysis was performed on 41 selected subjects using an audio processing computer software. Interjudge correlation coefficients ranged from .89 to.94, and reliabilities between human rating and the acoustic analysis were .90 and .80 for the major and minor song respectively.

A t-test on the *PMMA Tonal* scores indicates that the two groups of children were comparable at the outset of the study. Analysis of the singing scores yields the following findings:

1. A significant difference was found in the major song scores between the two types of vocal modeling.

2. An interaction effect of vocal model by tonal aptitude was found in the minor song scores.

3. Children with high tonal aptitude and trained by the female vocal model performed the best in both songs.

4. A positive relationship between tonal aptitude levels and singing scores existed consistently in children trained by the female vocal model.

5. Tonal aptitude levels seemed to have little effect on the singing scores for children trained by the male vocal model.

6. All children scored higher in the minor song than the major song regardless of the vocal model used in training.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT

I am deeply grateful to my advisor, Dr. Robert Walker, for his exceptional guidance and for making available the use of equipment for data analysis. My appreciation is also expressed to Dr. Nand Kishor for his recommendations on statistical analysis and Mr. Alex McLeod for his guidance in writing.

To my husband, Tak-shun, for his continued confidence and encouragement which have provided me the privilege to pursue the degree and see this project to fruition, I shall always be grateful.

CHAPTER ONE

PURPOSE OF THE STUDY

Introduction and the Nature of the Problem

Singing has traditionally been regarded as the central activity in the elementary school music curriculum and as a primary means through which all musical concepts and elements may be introduced. Many educators propose that the cultivation of good vocal habits and establishment of vocal accuracy in early childhood should be an important objective of music teachers (Beall, 1981; Gould, 1969; Greenberg, 1979).

It has been generally accepted that pitch concepts and singing accuracy are very important aspects of music teaching. Music teachers frequently devote considerable attention to the training of intonation because it is believed to be the primary objective through which singing accuracy could be achieved. Among the approaches of training intonation, pitch matching and learning songs by rote are most commonly used in classrooms. Pitch matching exercises require children to listen to a stimulus composed of short melodic patterns provided by a vocal model, and children either imitate or complement it with a singing response. Similarly, songs are usually introduced to young children by rote which is usually based on the practice of echoing the stimulus presented by a vocal model. Although the exact technique of presentation depends on the length and complexity of the song, all rote teaching approaches require children to echo different phrases of the song alternately with the teacher. The expectation is that children are able to imitate the model voice and learn the song through reproducing the melody. However, when direct imitation is the primary teaching focus, a child is required to aurally perceive the acoustic information and process it to reproduce the stimulus. Such a requirement to aurally apprehend the acoustic stimulus presented by adult vocal models could be challenging to young children because it assumes all young children would perceive the sound as the vocal model anticipated and encode the acoustic information accurately. Even more controversial is the presumption that all vocal models are equally effective presenters of auditory stimuli for young children.

The Child as an Auditory Receptor

The human brain is a compelling mechanism which is capable of performing incredibly complicated tasks to analyze and interpret signals transmitted from our senses. The ability to recognize a person's voice transmitted over the telephone even with a considerable amount of distortions is indeed cultivated through experience: the better we know the persons's voice, the more recognizable it would be over the telephone. Previous conversations with a person we know provided us with exposures to the same set of acoustic information upon which we encode into our schemata. Acoustic information gradually accumulates to form a system of reference points from which new auditory signals could be deciphered and evaluated. Experience is crucial to the process of encoding information, and the apprehension of experience converts the unfamiliar to become familiar.

The human auditory mechanism is very sensitive to learning experience. Auditory experience determines how an acoustic signal is received by a listener and hence influences how the information is being processed. Preisler (1993) examined the influence of the spectral composition of complex tones and of musical experience on virtual pitch perception which she defines as a phenomenon "whereby the fundamental pitch assigned to a complex tone is more strongly determined by higher harmonics than by the fundamental frequency itself" (p. 589). She suggested the more superior performance of the musicians in her study was due to their greater experience with different kinds of complex tones, which she believed has allowed a more exact identification of the stimulus material on the basis of its spectral composition. Preisler proposed that final pitch percept is determined by previous experience which provided strategies for subjective evaluation of the acoustical stimulus. Similarly, one might deduce that the more listening experience a child has with a particular vocal model, the more efficiently the child would process the vocal stimulus and a more accurate singing response might be made.

Auditory Perception of Children

The term auditory perception refers to a process whereby the listener organizes acoustic information. Although there are objective evaluations on the physical attributes of a sung pitch such as the measurement of a waveform's repetition rate, intensity and duration, these measures are hardly used in daily situations. For example, concert A expressed in the form of a frequency of 440 hertz and 40 decibel for 20 milliseconds means little to both musicians and non musicians alike, and therefore, its implication is limited and has very little practical significance in musical situations. Instead, ambiguous and subjective descriptive terms such as high and low are commonly used to convey and communicate the pitch perceived. Similarly, colours and light intensities are often described by terms like red, white, dark and light which we all seem to understand well but have merely accepted by faith. These vague descriptive terms can be generalized within a specific culture which shares a similar set of evaluative criteria. What might be regarded for example as a "white" color or a "dark" room are as subjective and ambiguous as how "high" a pitch should sound to be qualified as a "high pitch. These labels or subjective descriptions indicate that a type of learned concepts and categories are in effect, otherwise, they would fail to communicate the visual and audio perspectives they refer to.

Musical pitch perception is a trained response where learning effects on the operations of environmental acculturation are evident. Musicians have been reported as perceiving pitch categorically in the same way that language phonemes are perceived. Musicians were observed to employ boundaries to categorize pitch stimuli varying in small steps, while no such identification boundaries were reported in non musicians (Burns and Ward, 1978; Zatorre and Halpern, 1979). Musicians were found to have better discrimination between stimuli in different categories than in the same category (Siegel and Siegel, 1977a, b). How a listener codes the acoustic stimulus with respect to his or her established categories is a vital process which would extensively determine the evaluation of pitch. Likewise, since musical perception is based on a predisposition for perceiving the auditory stimulus, young children would employ their existing categories which are operated under the influence of acculturation and auditory experience, to evaluate melodic stimulus provided by a particular vocal model. More extensive discussion on the issues of auditory perception and experience of children will be addressed in the literature review. However, in order to draw inferences on possible strategies required for children to perceive auditory stimuli presented by adults, the differences between the male and female voices must be examined .

<u>Male Versus Female Vocal Models</u>

Of the many types of vocal models used in the elementary classroom, we may broadly classify them into two categories: the male and female. The two categories are generally distinguished by their differences in formant frequencies and tessituras.

Formant frequencies are determined by the distribution of resonant peaks within the frequency spectrum of a complex sound pressure wave and they establish the quality of a particular voice and the vowels produced. The uniqueness of formant frequencies in human beings is discussed in *The New Grove Dictionary of Music and Musician* (1980):

> The vocal cords produce a basic tone that can be varied . . . in envelope, but the tone can also have many different formants imposed on it by the amplification and resonances of all the various cavities of the nose, throat and mouth. Some of these are not variable and impose several of the characteristics that distinguish one speaker from another, male or female, youth from age and so on. Others are variable and allow the vowel sound to be produced. (p.557)

Although individual voices are distinguished by their formant frequencies, there are certain common acoustic features within the male and female voices which allow us to differentiate the timbres of the two categories. The perceived deeper quality of the male speaking voice is explained again in *The New Grove Dictionary of Music and Musician*:

> The two lowest formant frequencies are generally decisive in the vowel quality perceived. . . . Females have shorter vocal tracts and therefore higher formant frequencies. On average for vowels, the three lowest formant frequencies of female voices are 12, 17 and 18% higher, respectively, than those of male voices. (p. 84)

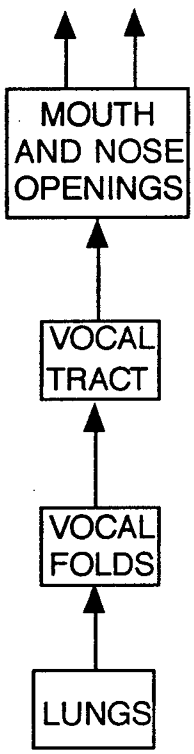
The male and female voices also vary in their singing tessituras. Although very often the male voice is felt or perceived to be singing the same pitch as a female voice, it is actually producing pitches an octave lower. The deception has to do with how our brain processes signals from the sensory organs. Two tones with an octave apart are often being perceived as "the same note" because even though they are separated in tone height (an octave apart), they both have the same chroma (a quality shared by all notes described by a particular letter). Nonetheless, to further explain those acoustical differences found between the male and female voices, the human voice organ must be examined.

Campbell and Greated (1987) outlined the operation of the voice organ by a schematic diagram (Fig 1).

FIGURE 1

A SCHEMATIC DIAGRAM ILLUSTRATING

THE OPERATION OF THE VOICE ORGAN (p. 472)



Openings radiate sound waves

Resonances modify source spectrum, creating formant regions

Vibration modulates air flow, giving pulsed waveform and source spectrum

Pressure creates steady flow of air through trachea and larynx

When producing speech or a sung pitch, the singer first creates a steady flow of air from the lung through the trachea and larynx, then to the vocal folds which modulate air flow to become pulsed waveform. A series of muscles vary the length and thickness of the folds by manipulating the position of the laryngeal cartilages. Since the sung pitch is determined by the vibration frequency of the vocal folds, its mass and tension would control the frequencies of the pitches produced. The thinner vocal folds of female execute a higher pitch range while the larger and more massive feature of the male vocal folds contribute to the difference in compass between male and female voices.

The length of the vocal tract is a also a primary determinant of the range of formant frequencies a voice produces because it is the resonator of the voice organ. The inverse proportion between formant frequencies and the length of the vocal tract implies the longer the vocal tract, the higher the frequencies produced. Whereas the female vocal tract is on average 17% shorter than the male, the range of formant frequencies produced by adult male voice is therefore, lower than female. The difference in the length of the vocal tract is even more apparent between very young children which is on average 8.75 cm, compared to 17.5 cm for the average adult males. As a result of such differences, children possess formant frequencies that are 35 to 40% higher than those of males and hence, a significant difference in their singing range would be expected.

Fortunately, the female voice has a wider singing range than the child's voice, and therefore, most female specialists have little problem in presenting a vocal stimulus in the treble range where a child's vocal range lies. On the other hand, since the male voice lies an octave below the vocal range of a child, the child is expected to transpose what is heard an octave higher in order to match a stimulus provided by the male vocal model.

Some generalizations based on teachers' experience have speculated that the use of an adult male vocal stimulus may confuse young children and suggested male teachers to use music instruments (for example, recorder) to provide the proper pitch, or to ask a child to provide the vocal stimulus (Smith, 1963; Greenburg and MacGregor, 1972; Nye and Nye, 1985). Some studies on the effects of male and female vocal modeling on children's singing accuracy concluded that the female vocal model was more desirable than the male because it was reported to be easier for children to match pitch patterns (Petzold, 1969; Hermanson, 1972; Green, 1990; Yarbrough, Green, Benson and Bowers, 1991). These researchers proposed that the need to transpose stimuli provided by a male vocal model to be especially troublesome for inexperienced young singers (Kramer, 1986). Another study reported children matched pitches more accurately when the vocal stimuli were presented in their own singing range (Petzold, 1963). These studies concluded that pitch-matching with a male model may be more difficult and its use may result in the child taking longer to achieve singing skills. They suggested male teachers either use their countertenor voice, recorded a female model, or ask a child to provide the vocal stimulus for the class (Goetze, Cooper and Brown, 1990).

However, other conflicting evidence showed that the requirement of octave transposition in male vocal modeling may not necessarily present unique pitch-matching problems to children (Small and McCachern, 1983; Montgomery, 1988). Those findings either suggested octave transposition may be possible in some young children or there are other generic problems with the design in some of the studies.

Need for the Study

The inconclusive results generated from these studies on the effects of vocal models on children's singing may have been caused by the use of a wide variety of methodologies, measurement instruments and techniques of analysis which inhibits drawing generalizations. They may also be due to the effects from a number of variables that are not comparable across studies. Variables such as the differences in the sample's ages, the choice of singing tasks as the criterion measure, the identification of the vocal model used in training period as distinct from the vocal model use as a stimulus for children to echo melodic patterns could all have contributed to the inconclusive results. Moreover, most of these studies have not explicitly investigated the tonal aptitude levels of the samples as a possible dimension (Petzold; Hermanson; Small and McCachern; Montgomery; Green; Yarbrough et al.). In fact, some studies conducted were, nevertheless, descriptive research without any account on whether the children were comparable, in terms of either their musical aptitudes or achievements (Green; Yarbrough et al.; Sims, Moore and Kuhn, 1982). It is, therefore, difficult to account for their differences in performance of encoding melodic stimuli at the outset of the study which might be consequential to their singing performance.

The inconclusive evidence with regard to the effects of vocal modeling may also be due to differences in the characteristics of singing tasks. Studies on the effects of vocal modeling have primarily focused on pitch-matching accuracy with very few practical findings on children's singing achievement of songs. Researchers who compared young children's performance on various singing tasks concluded that pitch-matching (echoing single note, interval and short pitch patterns) and vocal accuracy (the ability to accurately sing a song) are separate skills that are not highly related (Davies and Roberts, 1975; Welch, 1985; Flowers and Dunne-Sousa, 1990).

Similarly, research findings also indicated a low correlation between pitch-matching performance and accuracy of song performance (Flowers and Dunne-Sousa; Roberts and Davies). The author, therefore, believes that findings from studies which used pitch-matching as the criterion task may be of little value for generalization to children's singing achievement.

In order to assist elementary music teachers to identify and implement an effective approach to vocal modeling in the classroom, there is an imperative to conduct additional research. Hence, the purpose of this study is to to learn more about the effects of vocal modeling on children's song singing skills. The specific problem is to investigate the effects of training with male baritone and female vocal modeling on melodic singing achievement of two criterion songs (one each in major and minor tonalities) for grade one children who possess high and low tonal aptitudes.

CHAPTER TWO

RELATED RESEARCH

Introduction

Studies describing children's singing ability have been the focus of early researchers (Jersild and Bienstock, 1931; 1934; Drexler, 1938). During the 60's and 70's, more comprehensive findings were reported from several extensive studies and hence led to an interest to identify factors affecting children's vocal accuracy (Gould, 1969; Petzold, 1969; Davies and Roberts, 1975; Welch, 1979; Wassum, 1979). Variables such as text, the characteristics of singing tasks, differences in individual and group singing, accompaniment, and special or remedial training programs have been investigated. Among those variables, vocal modeling is an area in which a considerable amount of research has been conducted in the last decade.

This chapter reviews five pertinent studies on the effects of vocal models on singing achievement of young children. On account of recent reports on the relations of musical aptitudes to children's musical performance, two studies of musical aptitudes and children's singing achievement will also be reviewed. In view of the inherent impact of the auditory perception and experience of children on their singing achievement, a discussion on related research findings will also be included. A summary of the literature review with implications for the present study be then be addressed.

Studies on the Effects of Vocal Modeling on Singing Achievement

It is commonly accepted that a vocal stimulus provided by the female vocal model is easier for young children to match. Investigators have concluded that children sing more accurately when they echo vocal models heard in the register in which their own voices lie (Sims, Moore, and Kuhn, 1982; Green, 1987). A recent study by Yarbrough, Green, Benson and Bower (1991) reported that the female vocal model have been responded significantly more accurately than the male vocal model. Similarly, Small and McCachern (1983) also reported the female model to be slightly easier for children to match as compared with a male vocal model.

In view of the general assumption that the male vocal model would be more difficult for young children to pitch-match because of the requirement of octave transposition, Montgomery (1988) investigated the effects of the male baritone and the male countertenor (falsetto) vocal model. Interestingly, he reported that no significant differences were found in children's performance on pitch patterns after the administration of twelve weeks training by a baritone vocal model and a countertenor vocal model.

The Sims, Moore and Kuhn Study

An *ex post facto* design was used by Sims, Moore and Kuhn (1982) to investigate the effects of female and male vocal stimuli, tonal pattern length, and age on vocal pitch-matching abilities of grade one children. One hundred and twenty children were drawn from two populations--a middle class infant school in England (n=60) and a middle class public school in the United States (n=60). Whereas the children from England received twenty minutes of music instruction five times a week from a male instructor, the children from the United States received thirty minutes of music instruction once a week from a female specialist.

A pitch-matching test consisted of twenty tonal patterns sang once by a female mezzo soprano vocal model and once by a male baritone vocal model was administrated to all the children. The patterns were constructed by combinations of up to six notes in the C major scale to form intervals ranged from minor seconds to an octave wide. In consideration of external validity, half of the children echoed to the female model first, and the other half of the children echoed to the male model first. Also, for each pattern, half were sung with words and the other half were sung with a neutral syllable.

The researchers reported significant differences existed between countries, pattern length, and responses to the vocal model. All children were found to have responded more correctly to the female vocal model, and hence, the researchers concluded that it might be difficult for young children to echo a vocal model singing in a register different from their own.

A threat to generalization might exist because the study was based on availability sample and generalization is therefore limited to the characteristics of the samples. The investigators could have chosen two groups that are as homogeneous as possible and are different only with respect to the male or female vocal stimuli received. The threat of selection bias was addressed by the researchers, but regarded as beyond the scope of the study. They believe that "although the results may be due to country/cultural differences, the extensive data needed to isolate those differences were not considered in this study" (p.106). History is also a threat when both quality and quantity of music training received by the two groups of children have not been controlled. Although the fact that the English children had more frequent music instruction might have contributed to their better performance in this study, many other plausible causes such as the vocal model used in training and unmeasured musical aptitudes of the children might have been overlooked. Since the findings of this study have many possible rival hypotheses, the generalizability of the results are quite limited.

<u>The Small and McCachern Study</u>

Small and McCachern (1983) investigated the effect of male and female vocal modeling on the pitch-matching accuracy of grade one children. Fifty-five children were given a pretest in which they were expected to match three melodic fragments (*do, re,* and *mi* arranged in six different orders) sung by a female vocal model, and similar fragments (using the same three tones but arranged differently) sung by a male vocal model. All children received two trials on each of the three melodic fragments. After the pretest, eight children who performed the patterns perfectly were eliminated from further participation in the study. All remaining children were randomly assigned to experiment groups that received training with either a female vocal model, a male vocal model, or to a control group which received no vocal training. After five days of tonal pattern training, 44 children were posttested with the same pretest material presented in the same order. Small and McCachern reported that no statistically significant differences were found between the pretest and posttest pitch singing accuracy among the control and the treatment groups. They concluded that practice with either male or female vocal models did not produce significant differences in pitch-matching accuracy among grade one children, and that male vocal modeling may not necessarily present unique pitch-matching problems to those children.

On a content validity level, Small and McCachern did not provide a solid rationale to support the criterion used in the construction of the test patterns. Findings from a related study suggest that children in grade one are able to sing tonal patterns well beyond combinations of *do, re,* and *mi* (Jarjisian, 1983). Criterion patterns which represent contrasts of melodic characteristics such as contour and intervals should be systematically constructed to assess the pitch-matching achievement of children. Such criterion could generate a wider variability among test results and hence more valid conclusions may be drawn.

Small and McCachern's use of a pretest-posttest control group design is superior to control selection and maturation threats. However, several design flaws may be considered as threats to the internal validity. The experimental treatment of only five sessions (each thirty minutes in length over the period of five days, including one shortened session due to an unexpected school assembly), may not have been long enough to affect children's pitch-matching accuracy. In fact, the repeated use of the same test material on the pretest and post tests within such a short time span may have produced a practice effect on the children.

In terms of external validity, ecological threats are minimal. All testing and treatment material with instructions were tape-recorded by a female and a male vocal model. However, a type II error is suspected as Small and McCachern acknowledged that the differences found between the effects of male and female vocal modeling could have been statistically more significant for a larger sample size. The inferences and generalization made from this study are, therefore, threatened by statistical conclusion.

Based on the results of this study, Small and McCachern suggested that some grade one children have little difficulty matching pitch whether the vocal model was male or female, while some of them have pitch-matching difficulty with both models. It was unfortunate that this study failed to account for any possible differences in auditory perception and experience of the children nor provided any indication of their musical aptitude levels to enhance the interpretation of results. As a result, the findings of this study can only generate limited practical implications for music teachers.

The Green Study

Green (1987) studied the effect of vocal modeling on the pitch-matching accuracy of 282 elementary students in grades one through six. Three pitch-matching tests were administered to all children at seven-day intervals in the same order--the female soprano vocal model, the male tenor vocal model, and then the nine-year-old child vocal model. The descending minor third interval (G to E above middle C) was used in each test. Green reported the three vocal models had a significant effect on pitch-matching accuracy of children: responses to the child vocal model were most accurate, followed by the female vocal model, and the least accurate responses were to the male vocal model.

In terms of internal validity, Green's use of a time series design with equal time intervals between the three pitch-matching tests of different vocal models reduced the threat of instrumentation. However, the validity of the instruments in this study is challenged by Goetze, Cooper, and Brown (1986). They pointed out that the female and male vocal models used by Green were opera singers who probably possessed very different timbres and vocal techniques than most music teachers. The use of female and male classroom music teachers as the vocal models would have strengthened the external validity.

Green reported the only experimental group that showed a consistent increase in singing accuracy from grades one to five received training with the female vocal model. He accounted for this phenomenon by the fact that all those children had been taught by a female specialist since enrolment in kindergarten. He concluded that the effect of training with a female vocal model increased singing accuracy and suggests further investigation on the effect of training with a particular vocal model, in particular, a male countertenor or vocal models using the different ranges of the male voice.

In terms of selection, the inclusion of all children in grades one through six enables observation across different grade levels. The results indicated that children in grades one and six have the most difficulty in pitch-matching with all three vocal models. Other studies (Welch, 1979) have also concurred that grade one children are still in the process of developing their singing skills, and therefore, an investigation using grade one children is more liable to assess the effects of different types of vocal models.

The Montgomery Study

Montgomery (1988) investigated children's performance in echoing pitch patterns provided by a male baritone and a male falsetto model. Forty students in two intact grade three classes were randomly assigned to receive biweekly training for twelve weeks with one of the two vocal models. Both classes were taught by the investigator, and all songs were taught by rote.

Six pitch patterns were administered to the children in both pretest and posttest. Results from the pretest indicated that the two classes were not significantly different in their ability to replicate pitch patterns. The analysis of posttest means of the two classes revealed no statistical difference. Montgomery, therefore, concluded that training with a particular vocal model did not make a significant difference in singing achievement and that instruction with a falsetto vocal model did not significantly improve vocal accuracy. In addition, all children were reported to have responded more accurately to a falsetto vocal model regardless of the type of male vocal model used in the training period.

Some questionable procedures were employed in Montgomery's study which may result, in the reviewer's opinion, in susceptibility to observer and experimenter bias. For example, the investigator compared the effectiveness of two vocal models both taught by himself and was the only judge who scored the taped singing tests. In addition, only 20% of the scores was verified by another independent judge. These procedures impede the reliability of the singing scores, and therefore validity of the results may be also be suspected.

The external validity of this study was relatively weak as well. The rationale to select the particular grade level for investigation is unclear, and hence generalization is limited to children in that particular grade. Furthermore, without any assessment of the children's musical aptitudes, the interpretation of findings is questionable because individual differences have not been accounted for.

The Yarbrough, Green, Benson and Bower Study

Yarbrough, Green, Benson and Bower (1991) investigated the effects of male and female vocal modeling on pitch-matching accuracy of 163 children in kindergarten through grades three and grades seven and eight, who were identified as inaccurate singers in a university laboratory school. A pitch-matching test that used a descending minor third criterion pattern sung in neutral syllables by a female and a male vocal model was administered after 8 weeks of instruction with the

Kodaly approach. To assess the effects of three response modes: Curwen hand signals; "sol - mi" syllables; and to sing "la - la", all children were randomly assigned to use one of the three modes in the pitch-matching test. Children were tested twice, one each of the male and female vocal models, with two weeks in between the tests.

The taped pitch-matching responses were converted to digital data and input into a sequencer application software (Performer) for analysis. For each sung response, the duration and pitch as recognized by the application software were collected, and the percentage of correct response became the child's score. A pilot test that used 40 children indicated the instrument had a .983 correlation coefficient between two analyses, and an interjudge reliability of .97. A comparison was made between the music teacher's classification of certain and uncertain singers and the scores obtained through the computer analysis to establish criterion-related validity. A 77% agreement was reported among the two methods of classification.

Yarbrough et al. reported no significant differences among the three response modes (Curwen hand signals; "sol - mi" syllables; and to sing "la - la"), and no significant response modes by models interaction effect. However, there was a significant difference between responses to the male and female vocal models. Moreover, the mean percentage of correct response to the male vocal model was

34.19%, while 51.12% correct response was reported for the female vocal model.

This study distinguishes itself from other related studies by the use of digitizing taped singing response for data analysis. However, in terms of internal validity, the appropriateness of using a sequencing software application to analyse singing response is questionable. First, the researchers did not provide a rationale for the choice of the sequencing software application. The justification for such analysis procedures is doubtful because the program is not designed to analyse vocal inputs. Second, the results obtained from the sequencing software depended entirely on how the software program processed the frequencies of the vocal signals into discrete pitches. The criteria are, therefore, predominantly determined by sensitivity of the instrument and the preset allowable deviations within the program. Although the instrument may have been chosen by the researchers in attempt to obtain more objective data, the instrument is actually a major threat to internal, hence external validity.

Another threat to the reliability of the design is that the vocal model variable has never been explicitly controlled by the researchers. It is not known whether all children have received regular music instruction from a female teacher prior to the study and what vocal model was used during the eight weeks of instruction. Also, the children are reported to have responded much more accurately to a female vocal model than a male model at the outset of the study (the mean percentages of time in correct sung response to the female vocal model ranged from 31.56% to 52.67, but correct response to the male vocal model only ranged from 8.59% to 26.51%). Nevertheless, the researchers neither built in any leveling device such as a musical aptitude test to equate that wide disparity, nor administered any experimental treatment using different vocal models. It is, therefore, not surprising to find a significant difference among the correct singing response made to the two vocal models because results from the pretest has already indicated such differences existed. Nevertheless, the construct validity of this study is relatively weak probably due to the fact that the focus of the investigation seemed to be on the effects of three different response modes (Curwen hand signals; "sol - mi" syllables; and to sing "la - la") rather than the vocal models .

Studies of Musical Aptitudes and Children's Singing Achievement

The Jones Study

Jones (1993) investigated the music aptitudes of accurate and inaccurate singers in grades one through three. One hundred and forty-four children participated the study were classified as either accurate or inaccurate singers by their music teacher. Twenty-four accurate singers and 24 inaccurate singers were selected from each grade, with 12 boys and 12 girls in each group. The *PMMA*

Tonal and *Rhythm* subtests were administered to all children within a two week period.

Jones reported the accurate singers in all three grade levels gained a consistently higher mean scores for both *Tonal* and *Rhythm* subtests of *PMMA* as well as the composite scores for the two subtests. Moreover, the differences between the composite scores of accurate and inaccurate singers were significant, and even more significant differences were found in the *Tonal* subtest. Jones concluded that accurate singers in her study scored higher than inaccurate singers on the *PMMA*, and suggested that pitch discrimination and vocal accuracy may be related.

In terms of the design, Jones used equal cell size as an equating device to minimize the threat to selection. However, she did not provide the reader with the explicit criteria used in the classification of accurate and inaccurate singers, and the selection was subjectively determined by the music teacher. Replication possibilities would have been strengthened if the procedures and the chant used to classify children's singing achievement were explained in greater detail. Also, the classification of accurate and inaccurate singers could have been confirmed by another independent judge to establish reliability. In light of the lack of detailed description on the procedures used in the classification of accurate and inaccurate singers, additional studies are warranted to confirm the finding claiming that the Tonal subtest of *PMMA* to be a significant predictor of children's singing achievement.

The Martin Study

Martin (1991) investigated the effects of three instructional approaches--tonal syllables, hand signs, and letter representations--on the development of verbal and syllabic tonal skills of grade one children. Secondary problems investigated by the researcher include the dependence on other instructional elements in use at the same time during the acquisition of tonal syllable skill, and the comparative effectiveness of teaching students to echo melodic patterns by using tonal syllables with and without visual aids.

Sixty five children were randomly selected from an elementary school in suburban Oklahoma City. Martin used a stratified randomization procedure to assign the children to one of three groups, each instructed by herself. Each group was taught with a different teaching approach: group one echoed patterns only (T1), group two echoed patterns with hand signals (T2), and group three echoed with hand signals and written letter representations of patterns (T3).

To assess the children before the study began, the *PMMA Tonal* subtest was administered to obtain levels of tonal aptitude, and the *Metropolitan Readiness Test*

(*MRT*), Level II, Form P, was administered to obtain levels of school readiness. The researcher devised three singing tests to assess the effectiveness of the three different teaching approaches. Tests one and three assessed children's performance in echoing tonal patterns without visual stimuli; while test two assessed children's performance on the sight-singing of patterns written on cards. Three independent judges scored the tests, and interjudge reliability coefficients for the three tests was reported to be .9945, .9811, and .9944 respectively.

Martin reported tonal aptitude to be a significant predictor of both tonal syllable and pitch accuracy. On the other hand, school readiness scores were reported to be a "less significant predictor" and hence were dropped by the researcher as a variable of interest in the remainder of the statistical analysis. Martin reported significant gains in tonal syllable scores for both the group who echoed patterns with hand signals (T2) and the group who echoed with hand signals and written letter representations of patterns (T3). A slight regression in the pitch scores of the group who echoed patterns only (T1) between test one and three was reported, while (T2) gained by the same degree on test three, and (T3) gained the most on test three among the three groups. However, the gains in (T2) and (T3) were not statistically significant and tonal aptitude was the only significant main effect on the performance of all three tests.

Based on the results, Martin concluded that tonal aptitude, rather than any one of the three teaching approaches, determined the fluency of children's verbal and symbolic generalization of tonal syllables. Although high-aptitude students were found to have made the greatest gains, Martin found all training and tests sessions in her study to be "disappointing and essentially the same" (p. 167). She accounted for such phenomenon based on the consideration that the criterion tasks were too difficult for grade one students and that the children's performances were affected when asked to consider too many things at one time.

Martin further concluded that hand signals, letter representations, and note heads may not increase children's performance on either pitch or tonal syllable singing. In her observations, the children in her study have not grasped the concept of pitch relationships presumably implied by tonal syllables that were used during singing. The children might have focused most attention on singing the tonal syllable rather than the actual pitches in the patterns and therefore, were unable to give sufficient attention to pitch accuracy and melodic contour.

Martin's study is interesting with inspiring findings. All stated problems were thoroughly investigated by the researcher and the procedures were logically derived from the research purpose. To negate selection effects, the sample in this study was randomly selected. Martin's use of a stratified randomization procedure to assign children to the three training groups strengthened the generalizability of results. For possible replication of this study, the reader is also provided with a detailed description on the children and the criterion of the stratification procedure.

It is unfortunate that the criterion level for each point was never described in the article. Although three qualified independent judges scored the three singing tests, without a clear description of the criteria developed by the experimenter to assess children's pitch accuracy, the report of mean scores is not a very meaningful indication of the children's singing achievement. However, considering the superior design and procedures of this study, the findings which indicated *PMMA* as a significant predictor of children's singing achievement was an important issue.

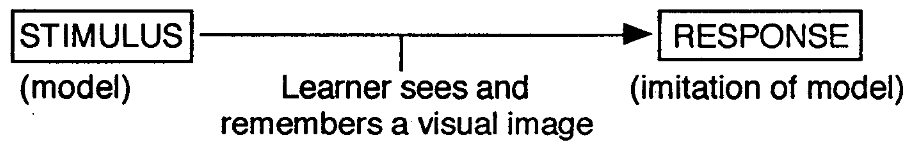
The Auditory Perception and Experience of Children

When a child is expected to produce a singing response to a stimulus presented by a vocal model, how the child perceives the auditory stimulus becomes an inherent determinant to his or her singing performance. Similarly, the effects of training with a particular vocal model is also depend on how the stimulus is being processed by the child. Apfelstadt (1983) suggested a model of singing in terms of a stimulus and response scenario (p.3) (see figure 2-1).

FIGURE 2-1

A STIMULUS/RESPONSE MODEL OF SINGING

Most psycho-motor skill-learning is acquired by imitation:



In singing, the stimulus is aural:

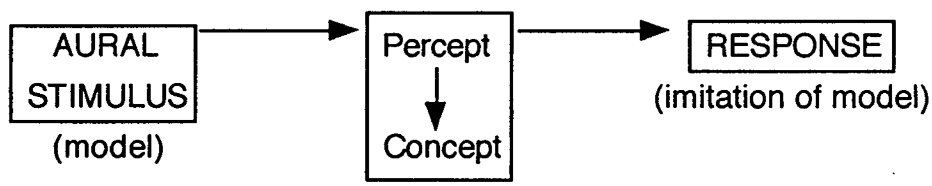


Image (later internalized) helps the learner to focus
and remember the sound; visual and kinesthetic mediators
may assist perception and concept formation

The process of internalizing a percept into concepts, as described by Apfelstadt, is when auditory experiences are formed. The auditory experience in turns provides a system of reference for perception to take place and the whole process of percept to concept may then repeat themselves over again in another cycle. Considering the magnitude of influence a percept or concept of the vocal stimulus might have on the singing response, a closer look at issues related to auditory perception and auditory experience of children is warranted.

<u>Auditory Perception</u>

In order to understand the auditory perception of children, it is essential to examine their capabilities and the strategies they employed to evaluate music stimulus. There are evidence indicate that infants within the first year of life possess very refined discrimination ability. They are sensitive to changes in the temporal properties of musical sequences (Demany, McKenzie and Vurpillot, 1977; Chang and Trehub, 1977a; Morrongiello, 1984) and are able to discriminate different pitches and their changes in temporal order (Chang and 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 and Morrongiello, 1987).

Young infants also possess enormously sensitive music perception skills. Researchers demonstrated infants can use their fine discrimination ability to process auditory stimulus and are capable to impose temporal structure in organizing musical sequences into patterns (Thorpe, Trehub, Morrongiello and Bull, 1988). More surprisingly, Trehub, Bull and Thorpe (1984) reported 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.

With such sophisticated discrimination and perception abilities already in possession within the first year of life, it demonstrates the drastic difference in innate potential and individual achievement because most children and adults have much inferior performance in auditory perception tasks compared to young infants. However, recent research on speech perception in infants has found that although infants are born with the abilities to discriminate both native and non-native speech contrasts, they have become increasingly sensitive and fine-tuned for their native language over the course of the first year (Werker and Tees, 1984). Such change in the perception in the auditory system become more evident when older infants and adults were found to be less able to discriminate non-native speech contrast compared to young infants (Werker, Gilbert, Humprey and Tee, 1981; Eilers and Oller, 1985; Eilers, Wilson, and Moore, 1977). Based on these findings, researchers have identified the important role of linguistic experience on the development of perception of one's native language.

Whether the change in language perception in the auditory system can be generalized to music perception is still uncertain but there are numerous studies showing that acculturation may have played a part on the changes in perception of musical properties occur throughout childhood. Six months old infants were reported to be able to perceive mistuning equally well in both culturally familiar and unfamiliar melodic patterns. Whereas 12 months old infants have already demonstrated better perception in culturally familiar melodic patterns, the perceptions of adolescents and adults were found to be much more affected by acculturation.

Trehub, Cohen, Thorpe, and Morrongiello (1986) compared infants and preschool children on their discrimination skill on a semitone change in melodies in terms of Western music conventions. Preschoolers were reported to have better discrimination ability on the music of their culture, whereas infants responded equivalently on western and non-western melodies. On the other hand, Krumhansl and Keil (1982) found that older school-age children show increased awareness of tonal structure in musical sequences. They are reported to have better retention of diatonic than non-diatonic melodies, increasing sensitivity to key information, and preferences for diatonic over non-diatonic notes in appropriate contexts. Similarly, Lynch, Eilers, Oller and Ubano (1990) studied the ability of 6-month-old infants to notice mistuning in less than a semitone in melodies on either the major, minor, or Javanese pelog scales. The infants noticed mistuning reliably better than chances in all scales because they were not yet affected by acculturation, while adults affected by acculturation performed better in perceiving mistuning in major and minor scales than in the Javanese pelog scale. Lynch and Eilers (1992) also concluded after a follow-up study that the 6-month-old infants had not yet developed sufficient perceptual knowledge of the interval patterns of the major scale to enhance discrimination performance.

Many questions regarding children's auditory perception of music still remain. Although there is evidence demonstrating the development of perception of acoustic frequency begins from a broad, innate discrimination ability towards a more refined, musical acculturated categorization, future research is needed before we can understand how and when children begin to categorize acoustic frequency according to categories such as octaves, the western pitch classes of 12 semitones in an octave, tonalities, and key relationships. Based on the finding in speech perception, we can hypothesize auditory perception is formulated through auditory experience. It therefore, seems possible to learn more about children's auditory perception from the aspect of their auditory experience.

<u>Auditory Experience</u>

Auditory experience should be considered one of the major determinants of musical development and performance. Michel (1972) suggested that a child's "active dialogue with the musical phenomena of his environment" (p.14) has a great significance in the process of musical development.

Some researchers suggested auditory experience actually begins in prenatal stage, possible at twenty to twenty five weeks when the auditory mechanism is comparable to the adult's (Thurman, Chase, and Langness, 1987; Woodward, 1992). They suggested auditory stimulation during the prenatal and post natal periods lays foundation for future language and musical abilities.

Infants can discriminate and process melodic contours and derived pleasure from the sound of their own voices and through imitation of the acoustic impressions of their environment (Fernald, 1984; Trehub, 1990). They demonstrated enormous ability to engage in nonverbal vocalizations with a surprisingly rich melodic resources. As early as three to four months, infants exhibit a variety of experimentation with pitch and sounds. Their pitch range can be remarkably wide and descending melodic contours, especially glissandi are common. Papousek and Papousek (1989) suggested that infants in the first months of life can discriminate and process melodic contours from the vocal communication of their caretakers, to conceptualize and anticipate the contexts of the interaction. Infants respond affectively to the vocal melodies of their mothers' speech which provide cues to her feelings and intentions. In fact, the comprehension through melodic cues in vocalization is considered a prerequisite cognitive skill for linguistic functions (Fernald, 1992).

Young infants seems to have an universal predisposition to engage in active dialogue and experimentation with their auditory environment to gain more auditory experience. Vocal matching is the typical form of vocalization in infants and it plays a significant role in human preverbal communication and seems to be specific to humans. It is an important avenue of environmental influence on early vocal communication and is important to speech learning and speech evolution. Recent studies on vocalizations of mothers and infants have observed a typical mode of turn-taking or alternating exchange found in as early as 3 month old infants (Stern, Jaffe, Beebe, and Bennet, 1975; Anderson, Vietze, and Dokecki, 1977; Rosenthal, 1982). Vocal matching between infants and care givers offer frequent displays of imitative models and corrective feed-back to infant vocal practice (Papousek and Bornstein, 1989; Kaye, 1982). Such reciprocal vocal matching has been reported a regular component which involves 34% to 56% of infant sounds and considered important to speech learning and to speech evolution (Papousek & Bornstein, 1992). However, a large individual differences in frequencies of maternal vocalization was found in the vocal dialogues between 3 month old infants and their mother (Rosenthal). It ranged from mother who did not vocalize even once during the observation to the one who vocalized 98% of the observation time. Since infants were reported to be more likely to vocalize in response to the presence of maternal vocalization, a considerable difference in auditory experience could be expected even in these young infants.

There are evidence that young infants have not yet possessed the mature form of auditory perception ability found in adults. Adults possess the ability to extract a particular pitch even when some of the spectral components are missing because the association of spectral patterns with pitch develops through experience. Since young infants have lesser auditory experience than adults, they failed to extract a pitch when the spectral characteristics vary from their limited auditory experience (Bundy, Colombo, and Singer, 1982).

Although music perception skills develop throughout childhood without formal training, Morrongiello (1992) suggested that training in children serves to enhance perception of more detailed aspects of melodies, to facilitate the speed at which the information is encoded, and to result in better memory for the material. Morrongiello found that contour was a particular salient melodic feature for untrained children, whereas discrimination of pitch and interval information posed greater difficulties for them. Training seems to accelerate the emergence of sensitivity to more precise features of melodies and the ease of encoding, and hence, application of this knowledge to music perception tasks.

The importance of home music environment was considered a crucial variable in children's performance on pitch matching. Perceptual proclivity is an interesting perspective that influences young children's singing performance. Under a longitudinal study, Wendrich (1981) reported a lost in propensity to match pitch in infants after three years of musically inactive environment, while infants from musically rich environment continued to demonstrate such abilities. Wendrich suggested that children's singing achievement reflect what they had perceived until that point in time. This notion was supported by other researchers who believed the proclivity to categorize is innate, develops early and is susceptible to training and experiences (Lynch and Eilers; Lynch, et al.)

The environment plays a substantial role in shaping each individual's auditory perception. Neisser (1976) suggested "Perception and cognition are usually not just operations in the head, but transactions with the world." (p.11). What we heard actually becomes a part of our experience which in turn affect how we encode the audio signal to form a perception of the sound heard. Cognitive and affective processes continuously affect one another and likewise make impact on the vocal mechanism and hence manifest in singing performance.

<u>Summary and Implications for the Present Study</u>

The review of literature indicates that findings with regard to the problem of vocal modeling are still inconclusive. In addition, inconsistencies were found in the methodologies and choice of variables which may have inhibited drawing generalizations. The following discussion will focus on aspects that were taken into consideration for the design of the present study.

Grade Levels of the Sample

Age has been regarded as a crucial variable that affects children's singing achievement, and researchers have reported that children's vocal accuracy improves with age (Jersild and Bienstock, 1934; Drexler, 1938; Geringer, 1983; Ramsey, 1983; Goetze, 1986; Flower and Dunne-Sousa, 1990). However, there is a wide disparity regarding to the grade level of the sample used in these studies which ranged from kindergarten to grade eight. Among the studies reviewed, grade one students were reported to have most difficulty with vocal modeling (Green, 1990; Yarbrough, Green, Benson and Bower, 1991). Other researchers have also concurred that grade one children are at the beginning stage of a hierarchy or continuum in the development of singing (Welch, 1979), and the highest number of inaccurate singers was also reported in this grade level (Gould, 1969; Davies and Roberts, 1975; Ramsey, 1983). Based on these findings, the author proposes that children in grade one would be expected to be affected more significantly by the type of vocal model, and hence, investigations should focus on the this grade level.

Auditory Perception and Experience of the Sample

Although the review of literature suggests a relationship exists between children's auditory perception, auditory experience and their singing achievement, the five studies reviewed in this chapter and many other studies on vocal modeling have not explicitly investigated whether their samples were comparable or not as the outset of the investigation (Petzold, 1969; Hermanson, 1972; Small and McCachern; Sims, Moore and Kuhn; Montgomery; Green; Yarbrough et al.). Furthermore, the extent to which the relationships between auditory perception and auditory experience could be appropriately assessed and stipulated is still inconclusive.

A host of studies on children's musical performances have employed level of musical aptitudes, measured by *PMMA,* as a dependent variable. Some researchers maintained that the extent to which a child performs in various musical skills is dependent upon his or her musical aptitude (Gordon, 1979a; 1979b; 1989; Schleuter, 1991), other studies on musical aptitudes in relation to vocal accuracy (Martin; Jones) also claimed that the *PMMA* to be a predictor of children's singing achievement. The author, therefore, considers an assessment of the tonal aptitude levels of the samples is warranted to confirm the research findings. In addition, results from the tonal aptitude test may be used as an equating device for the possible disparity among the samples' auditory perception and experience.

Vocal Model Used During the Training Period

Studies that explicitly identified the vocal model used in the training period as a variable have generated interesting results (Small and McCachern; Montgomery). Interestingly, both studies reported that the vocal models used in the training period appeared to have no effect on the pitch-matching accuracy of children when that particular model provided the stimulus. All the children in Montgomery's study responded more accurately to the falsetto model regardless of the vocal model used in the training period. Similarly, Small and McCachern reported that posttest result on pitch-matching accuracy after practice with a particular vocal model did not produce significant differences. However, Green accounted for the higher accuracy reported in response to the female model was possibly due to the fact that all his samples from grade one through six have received training with a female teacher since kindergarten. The inconclusive results warranted additional investigations that focus on the effects of vocal modeling used in the training period as a separate variable, and should be distinguished from the effects of immediate singing response to the stimulus. The findings could provide more practical implications for music teachers from the perspective of long term training effects in contrast to immediate singing response.

Criterion of the Singing Task

The studies reviewed have primarily focused on pitch-matching accuracy with very few practical findings on children's singing performance of songs. Pitch-matching task requires a subject to echo individual pitches or short melodic patterns, usually composed by several pitches, provided by an auditory stimulus. The subject is expected to vocally reproduce the melodic pattern immediately after the stimulus has been provided. On the other hand, it may seem that a more complicated process of assimilation is required to perform a song: a child need to organize the lyrics with all the music elements (rhythm, pitch and timbre), as well as to establish a perception of the underlying formal structure of the song as a whole. To simultaneously replicate individual pitches or short melodic patterns after an auditory stimulus involves less synthesis skills and requires less self-monitoring over time.

Researchers who compared young children's performance on various singing tasks reported that pitch-matching and song performance are separate skills and not highly related (Davies and Robert; Geringer; Welch; Flower and Dunne-Sousa). Flower and Dunne-Sousa reported a low correlation between pitch-matching skills and the maintenance of a tonal centre in song performance. Remedial vocal training was reported to have improved significantly children's vocal range, single pitch and interval production, but no improvement was found in children's accuracy in song performance (Davies and Robert). The author, therefore, suspects that findings from studies which used pitch-matching as the criterion task may be of little value for generalization to children's singing performance of songs. It should also be noted that singing, not pitch-matching, comprises the central activities of elementary school music curriculum, and that song performance can provide much more musical expressive experience for children. Since song performance also resembles contemporary classroom practice, the effects of training with a particular vocal model as observed in song performance may provide more practical implications for music teachers.

Measurement and Evaluation of Singing Achievement

The review of literature indicates three major methods --rating scales, transcriptions, and stroboscopic and oscilloscopic devices, were used in current practice. Most studies on children's singing performance were scored according to rating scales which usually have five to seven points with specific descriptions assigned to each level of performance. Although specific descriptors were assigned to each scale level of a rating scale, the evaluation is, nonetheless, rely on the subjective perceptions of the judges.

Transcription of the singing response may provide a general view of the performance in terms of western musical notation. However, this assessment procedure was also liable to the auditory evaluations of the transcriber because under the operation of categorical perception, frequencies are assigned to the nearest half step at the discretion of the transcriber.

Measurement obtained through stroboscopic and oscilloscopic devices provide the degree of deviation from the given frequency measured in cents. With the recent development in electronic computer technology, more advanced software application have been specifically designed for acoustic analysis and may provide more precise and objective assessment. However, it should also be noted that such measurement may be more exact than necessary for evaluation of singing performance compared to the capacity of judgment made by music teacher in the classroom.

In order to provide more reliable and objective assessment of singing performance and to resolve the differences in reliability and precision of instruments and techniques of analysis, the assessment procedures of present study would include both a rating scale scored by three independent judges and an objective acoustic analysis performed by a computer software application program specifically designed for the analysis of speech and singing. The correlations of the two procedures would then enhance the reliability and the generalizability of results.

CHAPTER THREE

PROCEDURE

The purpose of this study was to investigate the effects of training with two different types of vocal modeling, the male and female, on the singing achievement of grade one children who possess high or low tonal aptitudes. This chapter provides a detailed description of the subjects and the two vocal models followed by the design and procedures of the study.

Selection of Subjects

Site for Data Collection

The study was conducted in a suburban community in British Columbia, Canada. According to Statistics Canada, 86.9% of the population in that community have English as their first language and the largest single occupation (33.3%) among the population is in managerial and administrative professions.

Random selection of subjects or schools was not possible because the study required that all subjects receive similar music instruction from one of two music specialists who were both experienced and used comparable teaching approaches and music material. However, efforts were made to control the selection of socioeconomic status of the families from which the subjects came. The neighbourhoods of the two schools selected for the study, "School 1" and "School

2", have a number of similarities. Table 3-1 reports some data from the 1993 census by Statistics Canada. They suggested homogeneity in terms of the socioeconomic levels of the two neighbourhoods where the subjects came from.

TABLE 3-1

SELECTED DATA OF THE NEIGHBOURHOODS OF SCHOOLS

Neigh-bourhood	Dwelling Units % Detached	Tenure % Owned	%Husband/ Wife families	Average household size	Average family size
School 1	98	95	92	2.9	3.0
School 2	74	95	90	3.0	3.1

Efforts were also made to control the type of school and the school's musical environment. The musical environment of the two schools were both very rich. Music could be considered an integral part of both schools because they offered a range of musical activities for their students, with performance opportunities at school concerts and other events. Both schools had a junior (grade 5) and a senior (grades 6 and 7) band. School 1 had two choirs, the junior (grade 3 and 4) and the senior (grades 6 and 7); School 2 had three choirs (grades 1 and 2; 3 and 4; 5 and 6). Both schools had music rooms reserved specifically for music lessons and musical functions. The music rooms were spacious, well equipped, and abundant musical instruments were available in abundant for the students in both schools.

Subjects

Eighty-three grade one children participated in the study. All children received music instructions from music specialists twice per week and more than 90% of the children had music instruction from the same music specialist since kindergarten. The children from School 1 received 40 minutes of music instruction from a female specialist during kindergarten and the instruction time was increased to a total of 80 minutes per week in grade one. Children from School 2 received music instruction with a male music specialist for a total of 100 minutes per week in both kindergarten and grade one.

A brief survey was administered to estimate how often and how much time the students sang with a female vocal model in the classroom other than during the music periods (see Appendix A) because all children participating in the study had female classroom teachers since kindergarten. Response from the classroom teachers showed that singing had been an integral part of the classroom activities because the children did engage in singing activities with the classroom teachers on a regular basis for a range of one to three hours per week. Nevertheless, the investigation on the amount of time children spent with different vocal models outside the music classes was beyond the scope of this study and hence, no further effort was made to obtain information pertinent to this factor.

The Vocal Models

To account for the similarities and differences between the two teachers who served as the vocal model in this study, additional information on the teacher's academic background and teaching experience were obtained (see Appendix B). Both teachers were experienced music specialists with a similar training in the Orff teaching approach. Although the male teacher was also trained in the Kodaly approach, he acknowledged that his teaching approach was primarily based on the Orff approaches. In addition, both teachers have teaching experience at local Universities and Colleges where they were sessional instructors of music education courses for elementary teachers.

The teachers were also interviewed with respect to their teaching approaches, the types of musical activities used in their music classes and the teaching objectives and activities they emphasized. Since almost all children received music instruction from the same teacher since kindergarten, information from both the kindergarten and grade one years were considered. Both teachers offered similar musical activities in their music lessons: singing, body movement, playing classroom instruments, listening and music reading and the piano was used considerably during their music classes. (See Appendix C for a comparison between the emphasis for kindergarten and grade one music classes as described by the two music specialists, and Appendix D for the complete questionnaire on the two music specialists).

Design of the Study

Introduction

The study consisted of five phases: 1) administration of the *Tonal* subtest of the *Primary Measures of Music Audiation* (*PMMA*), 2) a five weeks treatment period, 3) a singing test of two criterion songs, 4) the evaluation of the taped singing performance by three independent judges, and 5) an acoustic analysis on selected singing performance.

Procedure

Prior to experimental treatments, the *Tonal* subtest of *PMMA* was administered to all children. *PMMA* is designed to evaluate the developmental music aptitudes of children in kindergarten through grade three. It includes two subtests, *Tonal* and *Rhythm*. For standardization sample, *PMMA* has a composite reliability of .90 (Gordon, 1979); the *Tonal* subtest has a reliability of .85 (Gordon, 1979). The *Tonal* subtest measures the child's ability to perceive, audiate and compare two tonal patterns played by an electronic synthesizer. It consists of forty recorded patterns which requires approximately 25 minutes to administrate and standardized verbal instructions are provided in the test manual. Children are asked to circle the pair of faces which look the same if the two tonal patterns heard on the tape sound the same and to circle the pair of faces which look different if the two tonal patterns heard on the tape sound different. Since a pictorial answer sheet is used, no formal training in music, numbers, or language is needed in order to do this test.

For children in School 1, the *Tonal* test was administered during their music classes whereas children in School 2 took the test in their regular classroom setting. The difference was due only to school administration and all children completed the test according to the standardized procedures as suggested in the test manual. All the tests were administered by the researcher, with help from the music teacher or classroom teachers to invigilate and distribute stationary and worksheets.

<u>Experimental Treatment</u>

The training period lasted for five weeks in School 1 and four weeks in School 2 due to the different amounts of time allocated for music lessons in the two schools. All children received a total of 400 minutes of music lessons over the training period. Since children in School 1 had two 40 minutes music lessons per week, the training period was five weeks. For School 2, the training period was four weeks because the children had two 50 minutes music lessons per week.

Besides regular musical activities which included movements, listening and playing instruments, two criterion songs of comparable phrase length, vocal range, and singing difficulty were taught to all the children during the training period. Criterion song one was in major tonality and criterion song two was in minor tonality (see Appendix E). For each music lesson during the training period, an average of half the time was spend on rote song instruction to learn each of the criterion songs. The songs were learned solely by aural imitation without harmonic accompaniment so the children might focus on the melodic contour. The teachers were instructed to present the two criterion songs in the same key every time they perform the song, to guarantee that children were instructed in similar approaches with the same material. The training period was non-intrusive and the investigator visited all the classes once a week only to observe and participate in the music activities to establish rapport with the children. At no point did the investigator presented the song to the children during the visits.

Both music teachers possessed a trained voice which projected clearly and accurately without any excessive use of vibrato. They both have strong choral background and have been actively involved in local choral organisations. The only difference in experimental treatment for children in the two schools during the training period was the type of vocal models. The children in School 1 received music instruction from a female music specialist with a mezzo soprano voice, while the children in School 2 received music instruction from a male teacher with a baritone voice.

Singing Test

After the training period, a singing test on the two criterion songs was administered. Children were taken to a quiet room in small groups of four and all children were tested in the morning. Children sat on chairs with a small table in front of them. Drawings which depicted the objects and actions in the criterion songs were used to initiate singing. They were awarded to all children after they completed the singing performance. All singing performance was taped by a portable cassette recorder (Marantz PMD201) using a PZM microphone (Realistic).

Since the children were told that they had to perform the criterion songs individually to the investigator at the beginning of the study and they have been acquainted with the investigator over the two months prior to the singing test, all children were enthusiastic and eager to sing to the investigator. In order to allow the child to sing in a range he or she was comfortable in and to avoid adding a pitch matching criterion to the singing test, no starting pitch was used to prompt the child to sing. The child began to sing only when he or she was ready. At no time the investigator sang or gave any pitch to prompt the child. During the testing process, the investigator kept her comments to a minimum, speaking only when necessary to encourage the children and to help maintain their concentration. All 83 children completed the singing test without any difficulties.

Evaluation of the Singing Performance

The taped singing performance was subsequently evaluated by a five-point continuous rating scale. Three independent judges with choral background and solid experience in teaching music, evaluated all the taped singing performance. Two of the judges have a graduate degree in music education and one has a bachelor of music degree with a major in voice. The judges listened to the taped singing performances and scored each of them. They were given a copy of the criterion songs in music notation with the description of the rating scale printed below the song (see Appendix F and G). Precise performance of melodic rhythm were not judged since only tonal aptitude was measured in the *Tonal* subtest of the *PMMA*.

Acoustic Analysis

To obtain a more objective assessment of the children's singing performance and hence, to strengthen the reliability of the singing scores, acoustic analysis was performed to selected subjects after they have been evaluated by the three judges. The analysis was also based on the same five point rating scale and therefore a total of four singing scores on each of the two criterion songs were obtained from those selected subject .

To establish reliability of the two types of assessment approaches--human rating and acoustic analysis--subjects who had composite scores (the combined scores of the three judges on the two criterion songs) of 30, 18 and below 13 (high, medium, low) were selected for further acoustic analysis. A composite score of 30 indicates the subject scored an average of five points (full marks) on each criterion song; 18 represents the subject has an average score of three, and composite score below 13 represent an average score of 2 on each criterion song. The investigator found the need to delimit the scale of the acoustic analysis because a comparison of the two assessment approaches was not the main focus of the present study. The acoustic analysis was used to establish reliability and the results from a significant portion of the sample would be enough to draw inference. A stratification procedure was, therefore, employed to ensure the subjects selected could satisfactorily represent all the different achievement levels of the sample as evaluated by human judges. Random selection, on the other hand, might provide a subgroup that was less reliable in terms of representation unless a much larger sample was available.

Forty-one subjects were selected based on their composite scores awarded by the three judges. The size of this subgroup was considered a sufficient proportion of the sample to establish reliability in the singing scores and hence, to draw inference based on the results. In addition, the stratifying procedure has included three levels of singing scores and hence, comparison between the acoustic analysis and the human rating approach was possible.

Data Analysis

The Rating Scale

A five-point continuous rating scale (Table 3-2) was used to evaluate the singing performance of the children in all treatment groups. Both criterion songs were evaluated with the same rating scale by three independent judges. The score for each criterion song ranged from 1 to 5 for each judge. The composite score for the two criterion songs ranged from 6 to 30.

TABLE 3-2

RATING SCALE FOR THE EVALUATION OF SINGING PERFORMANCE

Scores	Description of singing achievement
5	The child maintained a tonal centre without modulation and sang most of the melodic patterns accurately.
4	The child was able to maintain a tonal centre and the general melodic contour but sang some incorrect melodic patterns.
3	The child exhibited a sense of the melodic contour but was unable to maintain a tonal centre or modulated the tonal centre established at the beginning of the song.
2	The child consistently uses a singing voice but exhibited no sense of the general melodic contour.
1	The child occasionally uses a singing voice.

Acoustic Analysis

After the three judges evaluated all the taped singing performances, acoustic analysis was performed on 41 subjects. The taped singing performances were sampled using an IBM compatible 486DX66 Desk top computer fitted with a digital signalling processing board and subsequently analysed with the *Computerized Speech Laboratory* (*CSL*) software.

CSL is an audio processing software developed by Speech Technology Research Ltd., University of Victoria, B. C. , Canada. It provides real time graphic and numerical display, audio output, signal editing, and a variety of analysis functions. Each taped singing performance was captured as sampled data, and the graphic format of the waveform was displayed in a view screen. The waveform could then be edited and a pitch extraction function computed and displayed the fundamental frequency of the waveform. A pitch analysis function then enabled the experimenter to obtain the average hertz value of each pitch sung by the subject.

After all pitches were obtained in Hertz, the experimenter computed the size of each interval performed using the difference between the logarithm of these frequencies, expressed in cent values (see Appendix H). The cent value served better as an objective evaluation of the interval sung because small tuning discrepancies expressed as pitch intervals in cents are easier to identify than with frequency ratios. The formula to calculate the cent value for a given interval is:

$$C = \log_{10} I \times \left(\frac{1200}{\log 2} \right)$$

where I = the ratio of the two frequencies f_2 and f_1
C = the number of cents in the interval
log = logarithm in any base

Accuracy was determined according to the equal tempered scale of which an octave is divided in 12 equal units, hence, a semitone equals 100 cents and one octave equals 1200 cents. Tolerance level was set at 50 cents whereby all frequencies within 100 cents are in essence assigned to the nearest half step which allows a deviation range of 50 cents. Such a tolerance level is realistic and essential due to the operation of categorical perception reported in trained musicians (Siegel and Siegel, 1977 a, b). The categorical perception of musical intervals suggests that pitch information was coded according to discrete pitch categories (Krumhansl, 1991) and a deviation of 50 cents was used in current related studies on children's singing performance (Goetze, 1986; Flowers and Dunne-Sousa, 1988).

After the logarithm of frequencies were obtained from all the 41 selected taped singing performances, acoustic data evaluation sheets were used to assist the evaluation of the singing performance according to the same criteria in the five point rating scale (see Appendix I). A more detailed explanation on the scoring procedures used for the data derived from acoustic analysis may be found in Appendix J.

Experimental Design and Analysis

Figure 3-3 illustrates the nonequivalent group quasi-experimental design used in this study.

FIGURE 3-3

NONEQUIVALENT GROUP QUASI-EXPERIMENTAL DESIGN

Group		Pretest		Treatment		Posttest
School 1	→	0	→	F	→	0
School 2	→	0	→	M	→	0

The treatment variables--training with a female vocal model (F) and training with a male vocal model (M)--was labelled factor A.

The aptitude variable--high and low tonal audiation abilities--was labelled as factor B. *PMMA Tonal* scores was used to organize the children into high and low levels of tonal audiation ability. The standardized mean score was used to separate the two levels of tonal aptitude. This design may be illustrated by figure 3-4.

To determine the effects of training with two different types of vocal models on the melodic singing achievement of children, three judges assessed the taped singing performance independently. The dependent variables are composed of the average of the judges' ratings for the major criterion song, the minor criterion song, and the composite scores for the two criterion songs. The score for each criterion song ranges from 1 to 5; the composite score for the two criterion songs ranges from 6 to 30. Means and standard deviations were calculated. The interjudge reliabilities of the rating scale for each criterion song were calculated using Pearson product-moment coefficient. The reliability between the average of the three judges and the acoustic analysis was also calculated.

FIGURE 3-4

AN ILLUSTRATION OF THE INDEPENDENT VARIABLES

		Factor A Vocal Models F	 M
Factor B	Low		
Tonal Aptitudes	High		
		n =43	n = 40

A 2 x 2 MANOVA was used to analyze the differences in singing achievement on the three dependent variables (average score of the major criterion song, average score of minor criterion song and composite singing score). Interaction effects, main effects, and simple effects were evaluated at the level of .05 level of significance.

CHAPTER FOUR

ANALYSIS AND RESULTS

The purpose of the study is to investigate the effects of training with male and female vocal modeling on the singing achievement of grade one children with high or low tonal aptitudes. The study was designed to examine the following null hypothesis ($p = .05$):

> There is no difference between the means for grade one children who possess high and low tonal aptitudes, who learn to sing criterion songs in major and minor tonalities with male and female vocal models.

This first section of this chapter is an analysis of the pretest data (*PMMA*), and in order to answer the research hypothesis, the second section reports statistical analyses of data obtained from the singing performance after the training period.

Pretest Data Analysis

To determine the tonal aptitude levels of the subjects before the training period, the *Tonal* subtest of Primary *Measure of Music Audiation (PMMA)* was administered as a pretest to all subjects. The mean *Tonal* score for children of School 1 is 31.77 and the standard deviation is 5.43. The mean *Tonal* score for children of School 2 is 32.88 and the standard deviation is 6.04. Since the *t*-value is less than 1.99 (the critical value for 81 degrees of freedom at the .05 alpha level), it can be concluded that the group means for School 1 and School 2 are not significantly different. This indicates that the two groups were not significantly different in terms of their tonal aptitudes at the outset of the study. A comparison of means for the *PMMA Tonal* subtest scores are presented in Figure 4-1.

FIGURE 4-1

COMPARISON OF MEANS FOR THE *PMMA TONAL* SCORES

Vocal models	n	Mean	SD	*t*	*p*
Female (School 1)	43	31.77	5.43		
Male (School 2)	40	32.88	6.02	-.88	.783

<u>Analysis of Tonal Aptitude Scores</u>

Two crosstabulations tables are used to examine the *PMMA Tonal* test results. The number of subjects who scored high or low at the test are plotted on the vertical axis and the number of subjects who received training with either a female or a male vocal model on the horizontal axis (see Figure 4-2 and Figure 4-3). In this way it was possible to determine the number of subjects with high and low tonal aptitudes who received training with a particular vocal model.

To determine the tonal aptitudes of the subjects, both a median split (33) and a standardized means split (29.8) have been considered. The standardized mean score was considered unsuitable because the sample in the present study has yielded a significantly higher mean score compares to the standardized sample in the same grade level. Firstly, a mean score of 32 is, in fact, the mean score for the standardized sample in grade two which places the children who participated this study a grade above their chronological age. A split using the standardized mean, would exacerbate the negative skew on the distribution and therefore, could not account for the differences found in the tonal aptitudes among the two groups of children at the outset of the study. Secondly, to generate reliable results with the analysis of variance, equal numbers of observations in each cell of the factorial is preferred. It can be observed that more than 70% of the children participated the present study have high tonal aptitudes when compared to the standardized sample. A split using the standardized mean would place more than 70% of the children in the high tonal aptitude row (see Figure 4-2), while a median split would naturally place half of the sample in high and low levels of tonal aptitude (see Figure 4-3). Since equal numbers of observations in each cell could not be possible due to the design of the study, the median split is advantageous in providing more equally distributed numbers of observations in both row and column (see Figure 4-3). Hence, the *PMMA Tonal* median score was used to separate the two levels of tonal aptitude. (See Appendix K for a histogram on the frequencies distribution of *PMMA Tonal* scores).

FIGURE 4-2

CROSSTABULATIONS SHOWING TONAL APTITUDE LEVELS WITH A STANDARDIZED MEAN SPLIT AT 29

		Vocal Models: Female	Male	Row Total (%)
Tonal Aptitude Levels (Tonal scores, *PMMA*)	Low	14	10	24 28.9
	High	29	30	59 71.1
	Column Total (%)	43 51.8	40 48.2	83 100.0

FIGURE 4-3

CROSSTABULATIONS SHOWING TONAL APTITUDE LEVELS WITH A MEDIAN SPLIT AT 33

		Vocal Models: Female	Male	Row Total (%)
Tonal Aptitude Levels (Tonal scores, *PMMA*)	Low	27	14	41 49.4
	High	16	26	42 50.6
	Column Total (%)	43 51.8	40 48.2	83 100.0

Singing Tests Reliabilities

Interjudge reliability coefficients for the rating scale used for both the major and minor criterion songs are presented in Figure 4-4. The resultant correlations for both criterion songs are high. Reliabilities between the judges means range from .89 to .94. The judges mean and the acoustic analysis performed on the 43 subjects are also highly correlated. The major criterion song yields a .90 correlation coefficient and the minor song is .80 (see Figure 4-5).

FIGURE 4-4
INTERJUDGE RELIABILITY

	Major criterion Song	
	Judge 1	Judge 2
Judge 2	.90	
Judge 3	.94	.94

	Minor criterion Song	
	Judge 1	Judge 2
Judge 2	.89	
Judge 3	.90	.91

FIGURE 4-5
RELIABILITIES BETWEEN THE JUDGES MEAN AND THE ACOUSTIC ANALYSIS

	Major song	Minor song
	Acoustic Analysis	
Judges' mean	.90	.80

Analysis of Singing Performance

Having established the reliabilities between the three judges and the acoustic analysis, three sets of singing scores were used for analysis: the major song means (an average of the major song scores awarded by the three judges, ranges from 1 to 5), the minor song means (an average of the minor song scores awarded by the three judges, ranges from 1 to 5), and the composite singing score (the combined score awarded by all three judges on the two criterion songs, ranges from 6 to 30).

Multivariate analysis of variance (MANOVA) procedure was carried out with the *Statistical Package for the Social Sciences for Windows* (*SPSS Windows*) to determine the effects of the vocal models, tonal aptitude levels and possible interaction between vocal models and tonal aptitude levels, on the major, minor and composite singing scores (see Figure 4-6 for MANOVA summary table).

A significant difference was found in the overall effect ($p > .05$). The main effect of vocal model was statistically significant in the major criterion song scores ($p < .05$). It indicates there is a significant difference in the achievement of the major criterion song between the two groups of children taught by the male and female vocal models. A statistically significant interaction effect of vocal model by tonal aptitude was found in the minor criterion song scores ($p < .05$). It indicates a significant difference in the minor criterion song scores are due to a combined effect of vocal model and tonal aptitude levels on the subjects. The null hypothesis is rejected.

FIGURE 4-6

SUMMARY TABLE FOR MANOVA OF THE SINGING SCORES

Effects	Wilks Sig. of F	Variables	SS	DF	MS	F	Sig. of F
		Major song	9.95	1	9.95	6.62	.012*
Vocal Model	.018*	Minor song	.05	1	.05	.04	.836
		Composite	98.49	1	98.49	2.64	.108
		Major song	4.39	1	4.39	2.92	.092
Tonal Aptitude	.284	Minor song	.79	1	.79	.74	.391
		Composite	76.85	1	76.85	2.06	.155
Vocal Model		Major song	.28	1	.28	.19	.665
by	.099	Minor song	5.31	1	5.31	5.01	.028*
Tonal Aptitude		Composite	72.29	1	72.29	1.94	.168

* significant at .05 level

The Major Criterion Song

Means and standard deviations for the major song are presented in crosstabulations table (Figure 4-7). It can be seen that those children who received training with the female vocal model and had high tonal aptitudes, scored the highest in the major song, while children who received training with the male vocal model and had low tonal aptitudes, scored the lowest in the major song. For the major song, the combined means shows that children with high tonal aptitudes scored higher than children with low tonal aptitudes, whereas children who received training with a female vocal model scored higher than children who received training with a male vocal model.

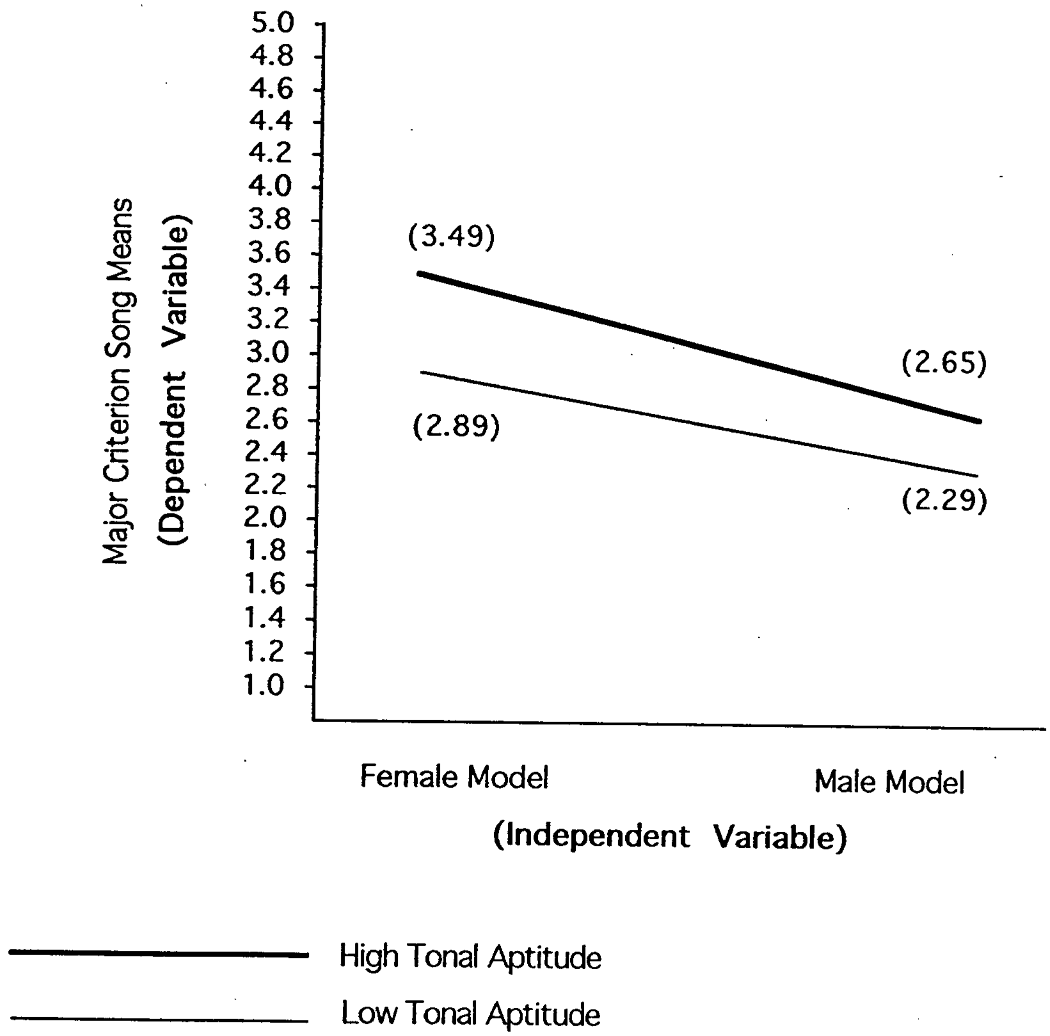
To examine possible main and interaction effects on the major criterion song, the cell means are plotted graphically (see Figure 4-8). The scale of the major song mean is placed on the vertical (*Y*) axis; the levels of tonal aptitude is placed on the horizontal (*X*) axis. The graph indicates no interaction between vocal model and tonal aptitude because the lines connecting the cell means are nearly parallel. However, the graph shows that children who received training with the female vocal model scored significantly higher than children who received training with the male vocal model.

FIGURE 4-7

CROSSTABULATIONS TABLE OF MAJOR SONG MEANS

		Vocal Models: Female	Vocal Models: Male	Combined Means
Tonal Aptitudes levels (Tonal scores, PMMA)	Low	2.89 (.92)	2.29 (1.33)	2.69 (1.06)
	High	3.49 (1.46)	2.65 (1.29)	2.97 (1.35)
Combined Means		3.11 (1.12)	2.52 (1.30)	

FIGURE 4-8
GRAPHIC REPRESENTATION OF ANOVA RESULTS
FOR THE MAJOR SONG



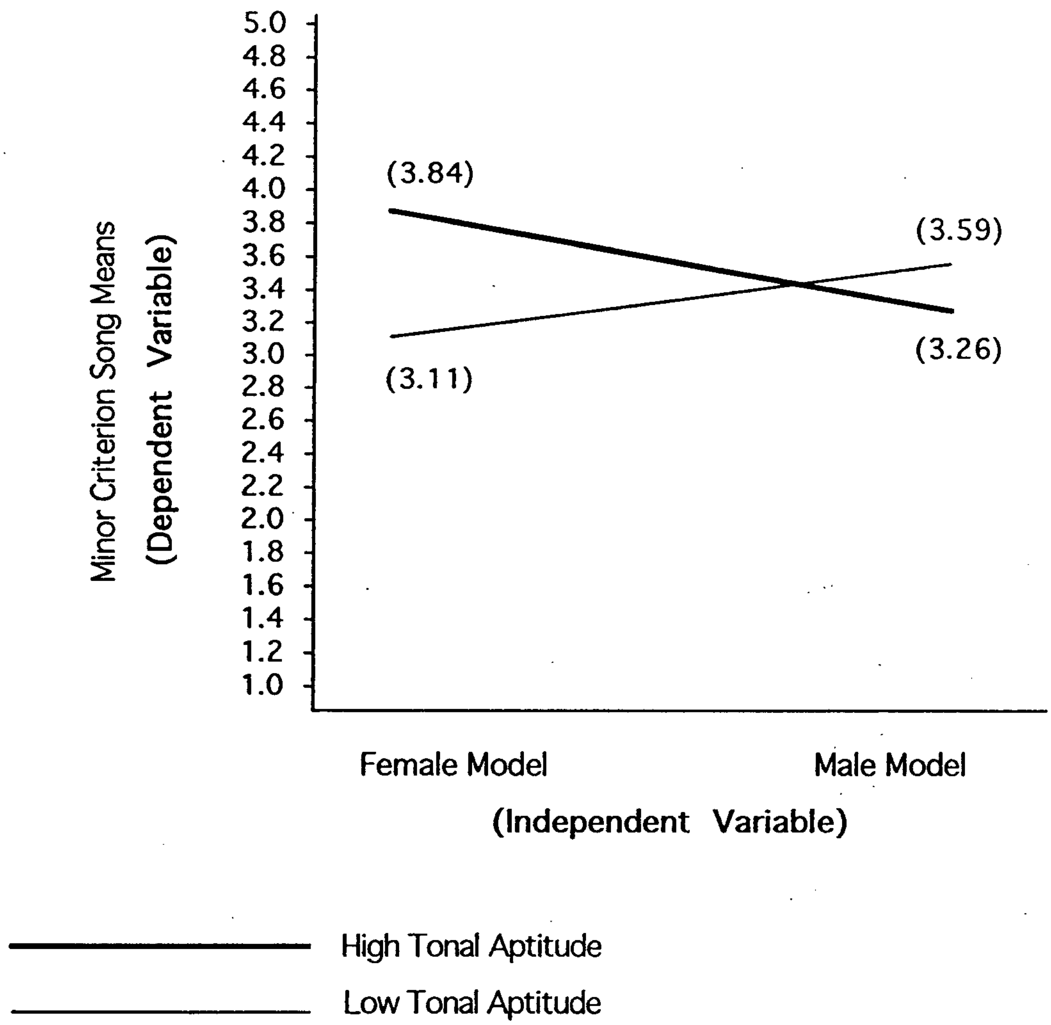
<u>The Minor Criterion Song</u>

Means and standard deviations for the minor song are presented in crosstabulations table (Figure 4-9). It can be seen that children who received training with the female vocal model and had high tonal aptitude, scored the highest in the minor song among all the children. However, for children who received training with the male vocal model, high tonal aptitude children scored lower than low tonal aptitude children. For the minor song, the combined means shows that children with high tonal aptitudes scored slightly higher than children with low tonal aptitudes, whereas children who received training with either a female vocal model or a male vocal model scored equally.

FIGURE 4-9
CROSSTABULATIONS TABLE OF MINOR SONG MEANS

Tonal Aptitudes levels (Tonal scores, PMMA)		Vocal Models: Female	Vocal Models: Male	Combined Means
	Low	3.11 (.84)	3.59 (1.0)	3.27 (.89)
	High	3.84 (1.07)	3.26 (1.19)	3.48 (1.14)
Combined Means		3.38 (.93)	3.38 (1.12)	

FIGURE 4-10
GRAPHIC REPRESENTATION OF ANOVA RESULTS
FOR THE MINOR SONG



A significant interaction effect between the independent variables (training with either male or female vocal model and tonal aptitude levels) is found by graphically plotting the cell means (see Figure 4-10). A disordinal interaction is found because the lines intersect within the plot. This indicates the difference between the minor song means for high and low tonal aptitude levels is not the same across the two type of vocal modeling. For the female vocal model, the high tonal aptitude children scored higher than the low tonal aptitude children; while for the male vocal model, the low tonal aptitude children scored higher than the high tonal aptitude children. However, no simple effects was found (see Figure 4-6).

The Composite Singing Scores

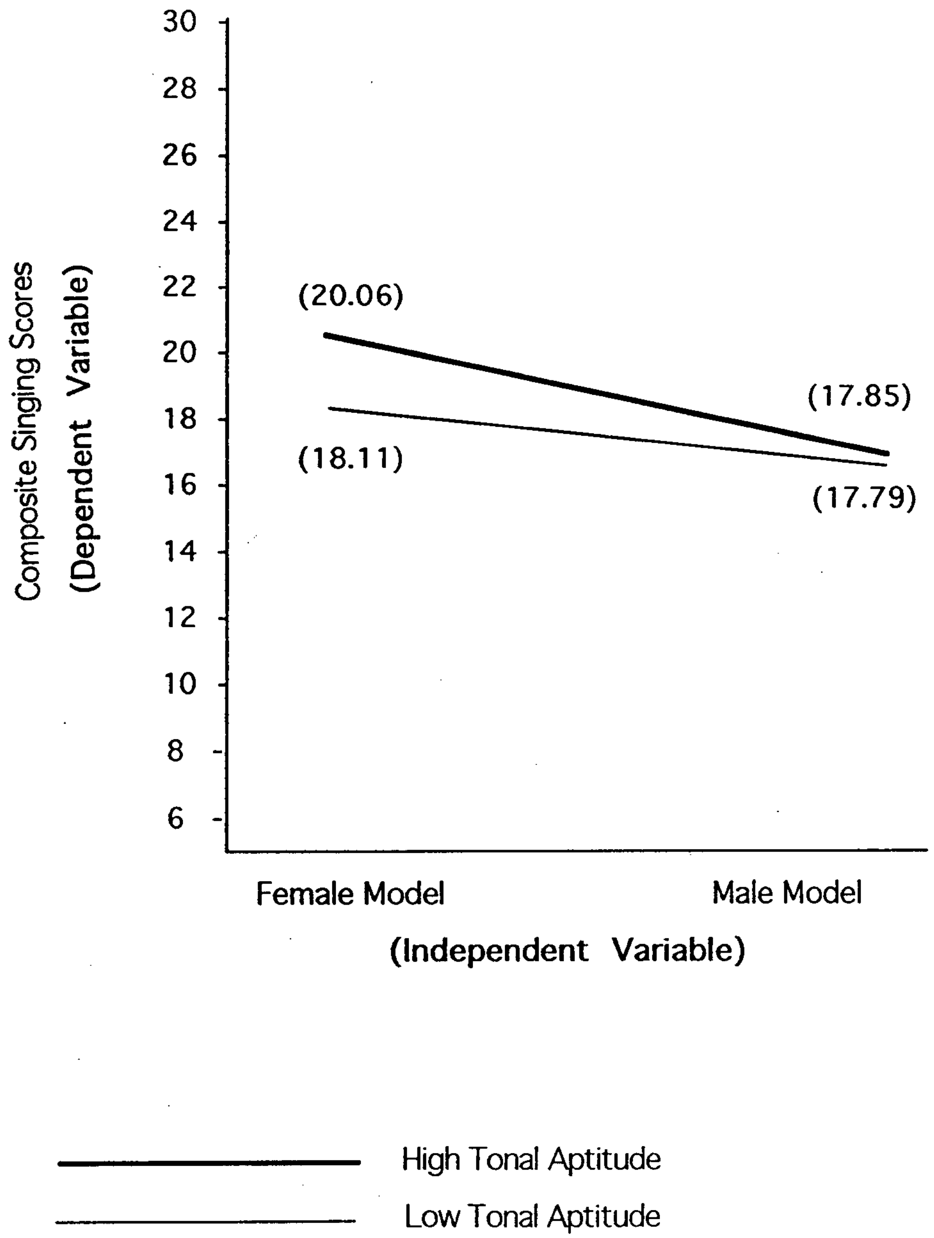
Means and standard deviations for the composite singing score are presented in crosstabulations table (Figure 4-11). It can be seen that children who received training with the female vocal model and had high tonal aptitude scored the highest among all other children. For children who received training with the male vocal model, the composite singing score means were very similar regardless of their tonal aptitude levels. For the composite singing score, the combined means shows that children with high tonal aptitudes scored slightly higher than children with low tonal aptitudes, whereas children who received training with a female vocal model scored higher than children who received training with a male vocal model.

Although no significant main and interaction effect is found by graphically plotting the cell means (see Figure 4-12), the line segments connecting the cell means are not parallel. This suggests a nonsignificant interaction because if there is no interaction, the lines connecting the cell means will be parallel or nearly parallel.

FIGURE 4-11
CROSSTABULATIONS TABLE
OF COMPOSITE SINGING SCORES MEANS

Tonal Aptitudes levels (Tonal scores, PMMA)		Vocal Models: Female	Vocal Models: Male	Combined Means
	Low	18.11 (5.07)	17.79 (5.45)	18.0 (5.20)
	High	20.06 (6.95)	17.85 (6.84)	18.69 (6.88)
Combined Means		18.84 (5.77)	17.83 (6.35)	

FIGURE 4-12
GRAPHIC REPRESENTATION OF ANOVA RESULTS
FOR THE COMPOSITE SINGING SCORE MEANS



Summary of the Findings

Vocal Models

1. Children who received training with the female vocal model have slightly higher composite singing scores than children who received training with the male vocal model.

2. Children who received training with the female vocal model scored significantly higher in the major criterion song than children who received training with the male vocal model.

3. Children with high tonal aptitudes and received training with the female vocal model performed the best in both criterion songs.

4. Vocal modeling has minimal effect on the minor criterion song scores.

Tonal Aptitude

1. Children with high tonal aptitudes scored slightly higher in all three sets of singing scores than children with low tonal aptitudes.

2. A positive relationship between the tonal aptitude levels and singing scores was found in children who received training with the female model, especially on the major criterion song.

3. Tonal aptitude levels seemed to have little effect on all the singing scores for children who received training with the male vocal model. In fact, a slight inverse

relationship is found in the minor criterion song means.

Major Versus Minor Criterion song

1. For the major criterion song, children who received training with the female vocal model performed significantly better than children who received training with the male vocal model.
2. Vocal modeling has no effect on the means scores of the minor criterion song.
3. All children scored higher in the minor criterion song regardless of the vocal model used in the training period.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Summary

The purpose of this study was to investigate the effects of training with vocal modeling on the singing achievement of grade one children. Two sets of contrasting factors were examined: 1) the effects of training with either a male or female vocal model, and b) the effects of high and low tonal aptitude.

Eighty-three grade one students from two elementary schools in British Columbia, Canada, participated the study. Random selection of subjects was not possible but the two groups were comparable in terms of socioeconomic levels, racial balance, the musical environment of the schools, and the quality of music instructions received.

The study consisted of five phases: the administration of the *Tonal* subtest of the *Primary Measures of Music Audiation* (*PMMA*), a five weeks treatment period, a singing test of two criterion songs, an evaluation of the taped singing performance by three independent judges, and an acoustic analysis on 41 selected singing performances.

The *Tonal* subtest of *PMMA* was administered to all children to determine their tonal aptitude levels. The training period spanned five weeks in School 1 and four weeks in School 2, during which children were taught two criterion songs by either a male or female vocal model for a total of 400 minutes of music instruction time. The two songs, one each in major and minor tonality, were comparable in phrase length, vocal range and singing difficulty. The training period was non-intrusive and the two music specialists taught the criterion songs by rote and without accompaniment. After the training period, children were asked to perform the two criterion songs individually and their singing performance were taped. No starting pitch was used to prompt the child to avoid the addition of a pitch matching criterion to the singing task and to allow the child to sing in a comfortable range.

Three independent judges evaluated the taped singing performance according to a five-point continuous rating scale. Each child yielded three sets of singing scores (an average of the major song scores, an average of the minor song scores, and a composite singing score which is the combined rating of the three judges). The interjudge reliabilities ranged from .89 to .94. To strengthen the reliability of these singing scores and to obtain a more objective assessment of the children's singing performance, acoustic analysis was performed on 41 selected subjects. A stratification procedure was used to ensure the subjects selected could satisfactorily represent all the different achievement levels of the sample and therefore, random selection was not appropriate. The taped singing performance of the 41 subjects were captured as sampled data and analysed with the *Computerized Speech Laboratory* (*CSL*) software. The logarithm of frequencies were obtained and the data were then evaluated according to the same five-point rating scale used by the judges. The correlation coefficient of the judges and the acoustic analysis was .90 for the major song and .80 for the minor song.

A *t*-test on the *PMMA Tonal* score revealed the two groups were not significantly different at the outset of the study. MANOVA was used to analyze the three sets of singing scores--average score of the major song, average score of the minor song, and the composite scores for the two songs. Analysis of the scores yielded the following findings:

1. Children who received training with the female vocal model have slightly higher composite singing scores than children who received training with the male vocal model.

2. Children who received training with the female vocal model scored significantly higher in the major criterion song than children who received training with the male vocal model.

3. Children with high tonal aptitudes and received training with the female vocal model performed the best in both criterion songs.

4. Vocal modeling has minimal effect on the minor criterion song scores.

5. Children with high tonal aptitudes scored slightly higher in all three sets of singing scores than children with low tonal aptitudes.

6. A positive relationship between the tonal aptitude levels and singing scores was found in children who received training with the female model, especially on the

major criterion song.

7. Tonal aptitude levels seemed to have little effect on all the singing scores for children who received training with the male vocal model. In fact, a slight inverse relationship is found in the minor criterion song means.

8. For the major criterion song, children who received training with the female vocal model performed significantly better than children who received training with the male vocal model.

9. Vocal modeling has no effect on the means scores of the minor criterion song.

10. All children scored higher in the minor criterion song regardless of the vocal model used in the training period.

Conclusions

The effects of training with male and female vocal model on children's singing achievement has received minimal attention among the host of studies on vocal modeling. Although a number of researchers (Petzold, 1969; Hermanson, 1972; Sims, Moore and Kuhn, 1982; Small and McCachern,1983; Green, 1987; Yarbrough, Green, Benson and Bower, 1991) suggested that the male model may be more difficult to pitch-match, the auditory perception and experience of young children have not been considered as a possible dimension in their investigations. The present study included tonal aptitude levels as a possible factor that affects children's singing achievement and it is also used as an equating device to account for the possible differences in student characteristics. The aptitude-treatment interaction design enables a more detailed examination of the different effects of the male and female vocal models on young children who have high or low levels of tonal aptitude.

Vocal Models

In general, children who received training with the female model scored better in both criterion songs when compared to children who received training with the male model. The results of this study confirmed that the female vocal model is superior for young children (Figure 4-12) regardless of the comparatively lower tonal aptitude scores for this group of children at the outset of the study (see Figure 4-3 and Appendix K).

The female model was found to be advantageous for young children (Petzold; Hermanson; Green; Yarbrough et al.; Sims, Moore and Kuhn; Small and McCachern) young children have more auditory experience with this particular vocal model. The fact that all children in this study also engaged in singing activities with their female classroom teachers should be considered. It seems that children performed better when trained by a female vocal model because they already have more experience with the female model and hence, strengthened their auditory perception for stimuli presented by females.

The process of internalizing an aural percept into aural concept as suggested by Apfelstadt (1983) could have begun in prenatal period (Thurman, Chase and

Langness, 1987; Woodward, 1992) and developed in full force during the first year of life when infants were frequently engaged in vocal matching between their caregivers (Stern, Jaffe, Beebe, and Bennet, 1975; Anderson, Vietze, and Dokecki, 1977; Rosenthal, 1982; Papousek and Bornstein, 1989; Kaye, 1982). During these early exposures with vocal models, the female model definitely stands out from the rest of the models because in a typical western society, females often spend more time, comparatively with infants. For the children in this study, their singing achievement would have manifested not only the effects of training with their particular music teacher but all the auditory experience they possessed at that time when the singing evaluation was conducted.

Although the female voice has been considered as a more effective model for young children, all three sets of singing scores indicates a considerable disparity between children with high and low tonal aptitude. The results suggests that while the female model was superior for children with high tonal aptitude, learning to sing by rote can still be considered challenging for low tonal aptitude children. On the other hand, children who received training with the male vocal model seemed to be less affected by their tonal aptitude levels but with slightly lower attainment. This may suggest that the male vocal model is equally challenging for young children regardless of their tonal aptitude levels.

Tonal Aptitude

For children who received training with the female vocal model, tonal aptitude was a more reliable predictor of singing achievement in both criterion songs. It confirmed the recent findings of Martin (1991) and Jones (1993) who reported tonal aptitude to be a significant predictor of tonal accuracy.

For children who received training with the male vocal model, tonal aptitude levels have minimal effect on their overall singing achievement. The results seemed to indicate that the effects of training with a male vocal model was a more significant factor that affected children's singing achievement. This indicates that acoustic stimuli presented by the male voice was equally challenging for children who had either high or low tonal aptitude, while acoustic stimuli presented by a female voice was challenging only to children who had low tonal aptitude.

In terms of the *PMMA Tonal* test results, it is interesting to find out that children taught by the male vocal model did not score better in the singing test than the children taught by the female vocal model, even though they have comparatively higher tonal aptitude scores (see Figure 4-3 and Appendix K). It may suggest that acoustic stimuli presented by the male voice is more difficult to perceive and discriminate than the electronic synthesizer used in the *PMMA*. An electronic synthesizer generates pure tones which is the simplest type of wave form (sine wave), whereas there are much more acoustic information to be processed in a male vocal stimulus because it produces complex waveforms (a combination of several sine waves that differ in wavelength). Preisler (1993) suggested that a greater experience with different kinds of complex tones would allow a more exact identification of the stimulus material on the basis of its spectral composition. Also, since pure tone exists literally in theory but hardly found in daily situations, the generalizability of *Tonal* test results on the effects of male vocal modeling could be limited.

For the present study, the most crucial auditory skills required to process a male vocal stimulus was 1) the discrimination of tones that are an octave apart but perceived as "the same note" because they are separated in tone height (an octave apart), but have the same chroma (a quality shared by all notes described by a particular letter name); and 2) the auditory skills of transposing the adult male voice up an octave to the treble vocal range of the child. For children who received training with the male model, the fact that *PMMA Tonal* score was a less reliable predictor for singing achievement is possibility due to the more advanced auditory skills required for young children to process the male vocal stimuli.

The predictability of tonal aptitudes on the singing achievement of children who received training with the male vocal model could have been diminished by the method used to separate the tonal scores into the two levels. As discussed in chapter four, two methods of splitting the tonal aptitude levels using the standardized mean (29) and the median (33), have been considered (see Figure 4-2 and Figure 4-3). The median split was favoured in view of 1) obtaining more equally distributed numbers of observations in both row and column for factorial analysis procedure; 2) to avoid the negative skew on the distribution of tonal aptitude scores; and, 3) to account for the differences found in the tonal aptitudes among the two groups of children at the outset of the study. However, the children who participated this study have significantly higher tonal aptitudes than the standardized sample and their mean score (32) places them a grade above their chronological age. The use of a median split for this particular sample have, therefore, included some medium tonal aptitude children into the low tonal aptitude group (see Appendix K for histograms). I

For the present study, separating the tonal scores into three levels--low, medium and high--would have provided a more accurate observation on the predictability of tonal aptitude levels on children's singing achievement. Moreover, with children classified into three levels of tonal aptitudes, the significant interaction effect found in the minor criterion scores would be more reliably explained and the results could generalize more valid implications.

Training

The effects of training with the male vocal model on the two criterion songs were different from the female model. As addressed before, for children trained by the male vocal model, there was virtually no difference in the composite singing scores between high and low tonal aptitude levels. In particular, a slight inverse relationship was found between the two levels of tonal aptitude on the minor criterion song. Despite of the compatibility of the two vocal models, a more detailed comparison between the music instructions received by the two groups of children during kindergarten and grade one yields some interesting findings.

In terms of instructional content, there were some incongruity among the emphasis of the two vocal models. The male vocal model had more emphasis on cultivating vocal control and improving the use of the singing voice, while the female vocal model had slightly less emphasis on singing in comparison to her other teaching objectives, such as social interaction and dramatic play. For children who received training with the male vocal model, the effects of a stronger emphasis on vocal development were exemplified in the similar overall singing achievement of both high and low tonal aptitude children. Whereas for children who received training with the female vocal model, the effects of high and low tonal aptitudes were stronger and more consistent in both criterion songs.

In terms of the amount of music training received by the two groups of children since kindergarten, children with the male vocal model was certainly superior because their music instruction time was more than double the other group of children during kindergarten. A even more prominent factor about this group of children was that 24 out of the 40 children participated in the Junior choir for three months prior to the study during which they rehearsed twice for a total of 45 minutes each week. The researcher also found out that more than 70% of the low tonal aptitude children in this group participated the choir. Training is suggested to be crucial to young children's development of music perception skills because it enhances the perception of more detailed aspects of melodies and facilitates the speed to process acoustic information (Morrongiello, 1992). It could then be suspected that the significantly more music instruction and vocal training received by the children with the male model might have weaken the effects tonal aptitude levels on this group of children. Moreover, the significantly more music training received by these children may have contributed the interaction effect in the minor criterion song if they have been taught more songs in the minor tonality. It is, however, beyond the scope of the present study to investigate the repertoire of the children since kindergarten.

In addition, the slight difference in *PMMA Tonal* scores among the two groups of children at the outset of the study might confirmed the developmental aspect of music aptitudes (Gordon, 1989a, b). According to Gordon, for children up to age nine, the average level of music aptitude would increase as the quality and quantity of of both formal and informal music instruction increases. The more music instruction received by the group of children with the male vocal model might have contributed to their higher tonal aptitude scores when compared to the other group of children.

Major Versus Minor Criterion Songs

In general, the children in this study performed the minor criterion song much better than the major song, especially for children who received training with the male vocal model. Two possible factors were considered by the investigator: 1) whether melodic material in minor tonalities (or songs in the La pentatonic scale) were more frequently used by the male teacher; and 2) the inherent differences among the major and the minor criterion song.

Although the male teacher received pedagogical training in the Kodaly approach and was a sessional instructor for the approach at local colleges and universities, he neither assigned any emphasis on minor melodic patterns nor taught more songs in the La pentatonic scale. In fact, he introduced melodic material in major tonalities to kindergarten children before minor tonalities was presented. Nevertheless, the investigator noticed another minor song was being rehearsed by the children in that group for a school event during the period the study was conducted. In terms of pitch-matching materials, the descending minor third interval (sol - mi) was used in kindergarten, and in grade one, the ascending perfect fifth (dol - sol), descending perfect fourth (dol - sol), and the descending pentachord (sol - fa - mi - re - dol) was used. Whether melodic material in minor tonalities were used more by the male teacher was beyond the scope of the present study and therefore, a closer examination of the compatibility of the two criterion song is warranted.

Initial efforts were made during the selection process to ensure the two criterion songs were compatible in terms of vocal range, phrase length, and singing difficulty. It was later observed by the two teachers and the investigator that the formal structure of the major song requires more advanced analytical skills than the minor song. The major song consists of four phrases A1, A2, B and A3 which in essence, requires the young singer to realize three variations of phrase A, in that the same melodic fragment is used in the first portion of three phrases but resolves with contrasting melodic material (see Appendix E). It was concluded by the two music teachers and the investigator that these salient variations confused most of the young children in the study because they have not yet possessed the cognitive skills to process such subtle melodic and structural features. The minor song, on the other hand, is much simpler in formal structure and therefore, less cognitive skills are required.

The minor song may be more appealing to young children because more variety of the tonal patterns are employed. The first phrase of the minor song has an overall arch shape, the second phrase is made up of a descending perfect fourth and an ascending perfect fifth interval, and the last phrase is a descending tonic triad. It should be pointed out that all children in the study have already been introduced to the concept of high and low in melody and they would have possessed the skill to perceive these melodic features in the minor song. Also, the group of children taught by the male teacher had, in particular, experience in singing the ascending perfect fifth and the descending perfect fourth interval which would enable them to performed better on the minor criterion song.

For the present study, the maintenance of a tonal centre and accuracy of intervals were the primary criteria upon which the evaluation of singing performance was based. Initial selection of criterion songs was based on the location of the tonic triad. The fact that the tonic triad is located at the beginning of the two selected songs was considered compatible by the investigator in terms of facilitating the recognition of the tonal centre and hence, the intervallic structure of the whole song (Edworthy, 1985; Croonen and Kop, 1989). In terms of the underlying harmonic structure, the whole major song is nevertheless, built on the tonic triad, whereas a distinctive tonic and dominant relationship is evident in the minor song. The results of the singing scores for the two criterion songs confirmed that the tonal triad does not provide enough information about a particular tonality, while a series of triads, preferably with more explicit harmonic relation such as the tonic-dominant relationship, would be adequate (Schoenberg, 1954; Simonton, 1984).

Recommendations

In view of the results and findings of this study, several recommendations are appropriate. While some suggestions have immediate implication for the classroom teachers, others are imperative for future research.

The analysis of the singing scores did not produce statistically significant results except for the difference in the major song means between the two types of vocal model and an interaction effect in the minor song. The limited generalizability of results were due to a relatively small sample. To generate results with more significant differences, either a much larger sample or equal cell size is necessary. Interactions between vocal models and the tonal aptitude levels could have been more significant should the the power be bigger. Also, the sample in the study were from two homogeneous middle class schools. To provide findings on the effects of male and female vocal modeling, a larger number of samples drawn from a wide variety of school settings, socioeconomic levels, racial and ethnic groups would be more favourable.

The effects of training with different types of vocal modeling have received minimal attention in research. The amount of formal musical training and informal experience children had with a particular vocal model should receive more attention. Regardless of the procedures used to control the compatibility of the music environment of the schools where the subjects came from and the quality of music instruction they received, the two groups of children seemed to have received very different amounts of music training and experience with the male and female vocal models. Either a longitudinal study or a study with much younger subjects should be conducted to verify the findings. Additional time would allow more significant training effects to occur, while the use of younger children may make it easier to identify the amount of formal and informal music experience they have with a particular vocal model. Where none of the above is possible, a qualitative aspect should be incorporated to the study to account for the more subtle conditions of the subjects informal and formal experience with vocal modeling. Questionnaires on vocal modeling at home, in the general classroom, and anecdotal notes taken during the children's music classes could assist the interpretation of results.

Replication of the study may include several grade levels to determine the effects of vocal modeling at different age levels. It would then allow researchers to collect evidence on when the ability of octave transposition is possessed by children or when the deception between pitch height and chroma could be readily discriminated by children.

As demonstrated in the results, the selection of criterion songs is crucial to children's singing performance. A replication of the study may include more than one song for each tonality to provide more explicit result. Comparisons of children's singing achievement on songs or melodic patterns in different tonalities have been documented and future study is needed to investigate the effects of training with different vocal modeling on criterion songs in different tonalities.

In view of the different effects the vocal models had on children with different tonal aptitude levels, a replication of this study is needed to determine if more consistent and significant effects could be found with a larger sample conducted longitudinally over several years. Where that is not possible, the *Tonal* subtest of the *PMMA* could be administered several times during a school year, with subsequent singing tests to investigate the possible effects between the two factors. Classifying children into low, medium and high tonal aptitude levels would enable more explicit statistic analysis and the results generated would provide more vigorous implications for music educators.

Finally, the use of an acoustic analysis on the taped singing performance proved to be successful and advantageous in providing a more objective evaluation. The high correlation coefficient between the judges' scores and the scores obtained from acoustic analysis strengthened the reliability of results. The acoustic analysis has been demonstrated to be a pertinent measurement tool for vocal production and more extensive use should be encouraged. Further study is needed to investigate and compare the objectivity and reliability of different measurement such as acoustic analysis, human judges, transcriptions, and the use of stroboscopic and oscilloscopic devices.

BIBLIOGRAPHY

Apfelstadt, H. (1984). Effects of melodic perception instruction on pitch discrimination and vocal accuracy of kindergarten children. Journal of research in Music Education, 32, 15-24.

Anderson, B. A., Vietze, P. M., & Dokecki, P. R. (1977). Reciprocity in vocal interactions of mothers and infants. Child Development, 1977, 48, 1676-1681.

Beall, G. (1981). Music as Experience. Dubuque, 1A: William C. Brown.

Bundy, R. S., Colombo, J., & Singer, J. (1982). Pitch perception in young infants. Developmental Psychology, 18(1), 10-14.

Burns, E. M., & Ward, W. D. (1978). Categorical perception--phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. Journal of Acoustical Society of America, 63, 456-68.

Campbell, M., & Greated, C. The Musician's Guide to Acoustics. London: J. M. Dent & Sons Ltd., 1987.

Chang, H. & Trehub, S. (1977a). Infants' perception of temporal grouping in auditory patterns. Child Development, 48, 1666-1670.

Chang, H. & Trehub, S. (1977b). Auditory processing of relational information. Journal of Experimental Child Psychology, 24, 324-331.

Croonen, W. L. M., & Kop, P. F. M. (1989). Tonality, tonal scheme and contour in delayed recognition of tone sequences. Music Perception, 7, 49-68.

Demany, L., McKenzie, B. & Vurpillot, E. (1977). Rhythm perception in early infancy. Nature, 266, 718-719.

Davies, A. D. M., & Roberts, E. (1975). Poor pitch singing: A survey of its incidence in school children. Psychology of Music, 3(2), 24-36.

Drexler, E. M. (1938). A study of the development of the ability to carry a melody at the preschool level. Child Development, 9, 319-322.

Edworthy, J. Melodic contour and musical structure. In P. Howell, & I. Cross (Eds.), Musical Structure and Cognition. New York: Academic Press, 1985.

Eilers, R. E., & Oller, D. K. (1985). Developmental aspects of infants speech discrimination: the role of linguistic experience. Trends in Neuroscience, 453-456.

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental changes in speech discrimination in infants. Journal of Speech and Hearing Research, 20, 766-780.

Fernald, A. (1984). The perceptual and affective salience of mothers' speech to infants. In L. Feagans, C. Garvey, & R. Golinkoff (Eds.), The Origins and Growth of Communication (pp. 5-29). Norwood, N. J.: Ablex.

Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In Papousek, M., Jurgens, U., & Papousek, H. (Eds.), Nonverbal Vocal Communication (pp. 262-282). Cambridge University Press.

Flowers, P. J. & Dunne-Sousa, D. (1988). Pitch-pattern accuracy, Tonality, and vocal range in preschool children's singing. Journal of Research in Music Education, 38(2), 102-114.

Geringer, J. M. (1983). The relationship of pitch-matching and pitch-discrimination abilities of preschool and fourth-grade students. Journal of Research in Music Education,31, 93-99.

Goetze, M. (1986). Factors affecting accuracy in children's singing. (Doctoral Dissertation, University of Colorado, 1985). Dissertation Abstracts International, 46, 2955A.

Goetze, M., Cooper, N., & Brown, C. J. (1990). Recent research on singing in the general music classroom. Bulletin of the Council for Research in Music Education, 104(4), 17-37.

Gordon, E. (1979a). Primary Measures of Music Audiation. Chicago: G. I. A. Publications.

Gordon, E. (1979b). Developmental music aptitude as measured by the *Primary Measures of Music Audiation.* Psychology of Music, 7(1), 42-49.

Gordon, E. (1989). Learning Sequences in Music: Skill, Content, and Patterns. Chicago: G. I. A. Publications.

Gould, A. O. (1969). Developing specialized programs for singing in the elementary school. Bulletin of the Council for Research in Music Education, 17, 9-22.

Green, G. A. (1990). The effect of vocal modeling on pitch-matching accuracy of elementary schoolchildren. Journal of Research in Music Education, 38(3), 225-231.

Greenberg, M. (1979). Your Children Need Music. Englewood Cliffs, NJ: Prentice Hall.

Greenburg, M., & MacGregor, B. (1972). Music Handbook for the Elementary School. West Nyack, NY: Parker.

Hermanson, L. W. (1972). An investigation of the effects of timbre on simultaneous vocal pitch acuity of young children (Doctoral Dissertation, Teachers College, Columbia University, New York, 1971). Dissertation Abstracts International, 32, 3558A.

Jarjisian, C. S. (1983). Pitch pattern instruction and the singing achievement of young children. Psychology of Music, 11(1), 19-25.

Jersild, A. T. & Bienstock, S. F. (1931). The influence of training on the vocal ability of three-year-old children. Child Development, 4, 272-292.

Jersild, A. T. & Bienstock, S. F. (1934). A study of the development of children's ability to sing. The Journal of Educational Psychology, 25(7), 481-503.

Jones, M. (1971). A pilot study in the use of a vertically-arranged keyboard instrument with the uncertain singers. Journal of Research in Music Education, 19(2), 183-194.

Jones, M. (1993). An assessment of audiation skills of accurate and inaccurate singers in Grades one, two, and three. UPDATE, Spring/Summer 93, 14-17.

Kaye, K. (1982). The Mental and Social Life of Babies: How Parents Create Persons. Chicago: University of Chicago Press.

Kramer, S. J. (1986). The effects of two different programs on third and forth grade children's ability to match pitch vocally. (Doctoral dissertation, Rutgers University, 1985). Dissertation Abstract International, 46, 2609A.

Krumhansl, C. L., & Keil, F. C. (1982). Acquisition of the hierarchy of tonal function in music. Memory and Cognition, 10, 243-251.

Lynch, M. P., & Eilers, R. E. (1992). A study of perceptual development for musical tuning. Perception and Psychophysics, 52(6), 599-608.

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception. Psychological Science, 1, 272-276.

Martin, B. A. (1991). Effects of hand signs, syllables, and letters on first graders' acquisition of tonal skills. Journal of Research in Music Education, 39(2), 161-170.

Michel, P. (1973). The optimum development of musical abilities in the first years of life. Psychology of Music, 1(2), 14-20.

Montgomery. T. D. (1988). A study of the associations between two means of vocal modeling by a male music teacher and third-grade students' vocal accuracy in singing pitch patterns (Doctoral Dissertation, University of North Carolina at Greenboro, 1988). Dissertation Abstracts International, 49, 08A, 2142.

Morrongiello, B. A. (1984). Auditory temporal pattern perception in 6 and 12-month-old infants. Developmental Psychology, 20, 441-448.

Morrongiello, B. A. (1986). Infants' Perception of multi-group auditory patterns. Infant Behavior and Development, 9, 307-319.

Morrongiello, B. A. (1992). Effects of training on children's perception of music: A review. Psychology of Music, 20, 29-41.

Neisser, U. (1976). Cognition and Reality. San Francisco: Freeman.

Nye, R. E., & Nye, V. T. (1992). Music in the elementary school (6th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Petzold, R. G. (1963). The development of auditory perception by children. Journal of Research in Music Education, 11(1), 21-43.

Petzold, R. G. (1969). Auditory perception by children. Journal of Research in Music Education, 17(1), 82-87.

Papousek, H., & Bornstein, M. H. (1992). Didactic interactions: Intuitive parental support of vocal and verbal development in human infants. In Papousek, M., Jurgens, U., & Papousek, H. (Eds.), Nonverbal Vocal Communication (pp. 262-282). Cambridge University Press.

Papousek, M., & Papousek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. [Special issue]. First Language, 9, 137-158.

Preisler, A. (1993). The influence of spectral composition of complex tones and of musical experience on the perceptibility of virtual pitch. Perception and Psychophysics, 1993, 54(5), 589-603.

Ramsey, J. H. (1983). The effects of age, singing ability and instrumental experience on preschool children's melodic perception. Journal of Research in Music Education, 31(2), 133-145.

Rosenthal, M. K. (1982). Vocal dialogues in the neonatal period. Developmental Psychology, 18(1), 17-21.

Sadie, S. (Ed.). The New Grove Dictionary of Music and Musician. London: Macmillan Publishers Ltd., 1980.

Schleuter, S. L. (1991). Research studies relevant to the *Musical Aptitude Profile*: A bibliography. The Quarterly, University of Northern Colorado, 2(1&2), 78-85.

Schoenberg, A. Structural Functions of Harmony. London: Faber and Faber, 1983.

Siegel, J. A., & Siegel, W. (1977a). Absolute pitch identification of notes and intervals by musicians. Perception and Psychophysics, 21, 143-52.

Siegel, J. A., & Siegel, W. (1977b). Categorical perception of tonal intervals: musicians can't tell *sharp* from *flat*. Perception and Psychophysics, 21, 399-407.

Simonton, D. K. (1984). Melodic structure and note transition probabilities: a content analysis on 15,618 classical themes. Psychology of Music, 12, 3-16.

Sims, W. L., Moore, R. S., & Kuhn, T. L. (1982). Effects of female and male vocal stimuli, tonal pattern length, and age on vocal pitch-matching abilities of young children from England and the United States [Special Issue]. Psychology of Music, 104-108.

Small, A. R., & McCachern, F. L. (1983). The effect of male and female vocal modeling on pitch-matching accuracy of first-grade children. Journal of Research in Music Education, 31(3), 227-233.

Smith, R. B. (1963). The effect of group vocal training on the singing abilities of nursery school children. Journal of Research on Music Education, 11(2), 137-141.

Stern, D., Jaffe, J., Beebe, B., & Bennet, S. (1975). Vocalizing in unison and in alteration: Two modes of communication within the mother-infant dyad. Annuals of the New York Academy of Science, 263, 89-100.

Thorpe, L., Trehub, S., Morrongiello, B. A., & Bull, D. (1988). Perceptual grouping by infants and preschool children. Developmental Psychology, 24, 484-491.

Thurman, L., Chase, M., & Langness, A. P. (1987). Reaching the young child through music: Is prenatal and infant music education possible? International Music Education, ISME Year Book, Vol. XIV (PP. 21-28). Austria: International Society for Music Education.

Trehub, S. E. (1990). The perception of musical patterns by human infants: The provision of similar patterns by their parents. In W. C. Stebbins & M. Berkley (Eds.), Comparative Perception: Vol. 1 Discrimination (PP. 429-459). New York: Wiley Interscience.

Trehub, S., Bull, D., & Thorpe, L. (1984). Infants' perception of melodies: The role of melodic contour. Child Development, 55, 821-830.

Trehub, S., Cohen, A., Thorpe, L., & Morrongiello, B. A. (1986). Development of the perception of musical relations: Semitone and diatonic structure. Journal of Experimental Psychology: Human Perception and Performance, 12, 295-301.

Trehub, S., Thorpe, L., & Morrongiello, B. A. (1987). Organizational processes in infants' perception of auditory patterns. Child Development, 58, 741-749.

Wassum, S. (1979). Elementary school children's vocal range. Journal of Research in Music Education, 66-67, 96-101.

Welch, G. F. (1979). Vocal range and poor pitch singing. Psychology of Music, 7, 13-31.

Welch, G. F. (1985). A schema theory of how children learn to sing in tune. Psychology of Music, 13(1), 3-18.

Wendrich, K. A. (1982). The wellsprings of musical meaning: Some new observations of pitch imitation in early childhood. Essays in Music Education in American Education and Society. University Press of America, Lanham, M. B.

Werker, J. F., Gilbert, J. H. V., Humprey, K., & Tee, R. C. (1981). The developmental aspects of cross-language speech perception. Child Development, 52, 349-355.

Werker, J. F., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development, 7, 49-63.

Woodward, S. C. Discoveries in the fetal and neonatal worlds of music. Paper submitted for publication, 1992.

Yarbrough, C., Green, G., Benson, W., & Bowers, J. (1991). Inaccurate singers: an exploratory study of variables affecting pitch-matching. Bulletin of the Council for Research in Music Education, 107(4), 23-34.

Zatorre, R. J., & Halpern, A. R. (1979). Identification, discrimination, and selective adaption of simultaneous musical intervals. Perception and Psychophysics, 26, 384-95.

APPENDIX A

QUESTIONNAIRE ON SINGING ACTIVITIES IN THE CLASSROOM

1. How often do you sing songs with your students?

    a) Daily
    b) Almost every day
    c) Several times per week
    d) Once or twice per week
    e) Very seldom

2. How many songs do you usually sing with your students each time?

    a) 5 or more
    b) 4
    c) 3
    d) 2
    e) 1

3. Please estimate the total amount of time students engaged in singing activities in your classroom **per week** (i.e. excluding the 2 music periods).

    a) more than 3 hours
    b) 2 .5 hours
    c) 2 hours
    d) 1.5 hours
    e) 1 hour
    f) Less than 1 hour
    g) Nil

4. Do you sing to or with your students during these singing activities?

    a) Yes, I sing to/with them all the time.
    b) I sing to/with them most of the time, and use other stimuli (e. g. musical instruments, recorded songs etc.).
    c) I use other equipment (e.g. musical instruments, recorded songs etc.) with my voice together on a more or less equal basis.
    d) I use other equipment (e.g. musical instruments, recorded songs etc.) in stead of my voice most of the time.
    e) I very seldom sing to/with them.

APPENDIX B

COMPATIBILITY OF THE TWO MUSIC SPECIALISTS IN TERMS OF QUALIFICATION AND EXPERIENCE

Factors	Female (School 1)	Male (School 2)
University	Bachelor of Music	Bachelor of Music, Master of Arts
Teaching Qualifications	Teachers Certification Certification in the Orff approach	Teachers Certification Certification in Orff and Kodaly approaches
Teaching Experience	1. Elementary (23 yrs.) and University (10 yrs.) [concurrent]	1. Secondary (6 yrs.), 2. Elementary (19 yrs.) and University (12 yrs.) [concurrent]
Teaching Approaches	Orff	Orff over Kodaly

APPENDIX C

A COMPARISON BETWEEN THE EMPHASIS OF THE TWO TEACHERS

Levels	Female Teacher (School 1)	Male Teacher (School 2)
Kindergarten	pitch matching individually with a graphic illustration of high and low, locomotive skills, dramatic play, use of the singing voice.	A sense of spatial relations, awareness of others, singing versus speaking, being able to sing individually, listening skills.
Grade One	Social cooperation, singing games, rhythm and beat, beat confidence, quarter and half notes (ta ti ti), echo melodic patterns out of songs, pitch concepts (high/low, same/ different, steps within an octave), listening and echo as a group.	Beat awareness (notation of beat and tracking beat notation), widen the vocal range, singing melodic intervals out of songs accurately, pitch concept (high/low)in relation to movement and singing.

APPENDIX D

QUESTIONNAIRE ON THE TWO MUSIC TEACHERS

Teacher: ____________________

Academic Background:

University	Degree	Year

Certification/others	Qualification	Year

Teaching Experience:

School	Type	Grades	Years

Teaching Approaches:

Teaching Emphasis (Major objectives and activities)

Kindergarten

Grade One

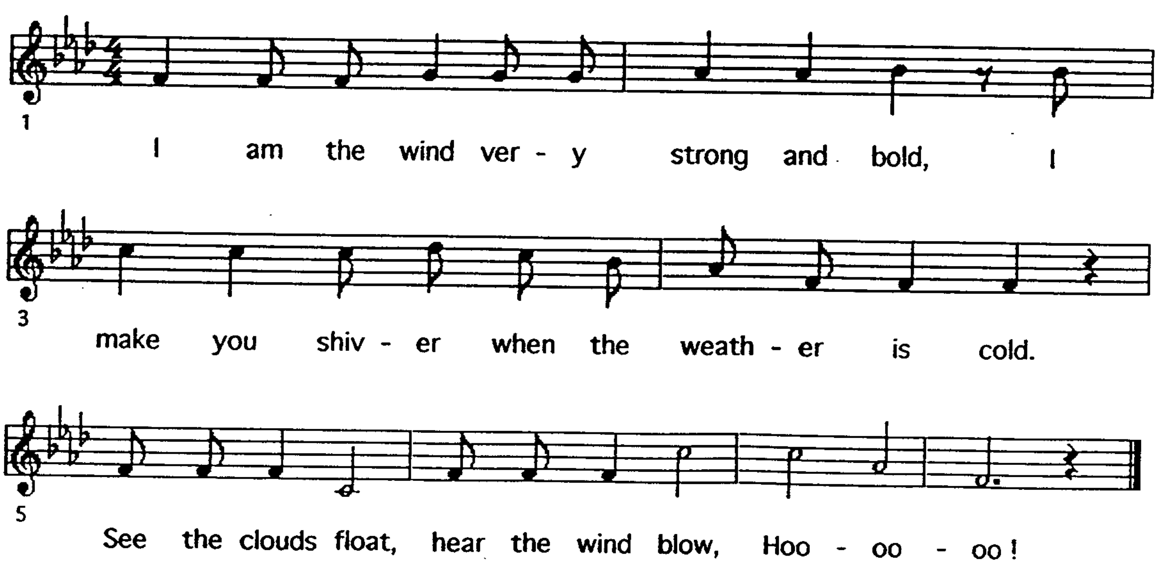
APPENDIX E

Criterion Song 1

The Mailman

Criterion Song 2

I Am the Wind



APPENDIX F

Subject I. D.: ____________ Score: ________

Criterion Song 1

The Mailman

Scores	Description of singing achievement
5	The child maintained a tonal centre without modulation and sang most of the melodic patterns accurately.
4	The child was able to maintain a tonal centre and the general melodic contour but sang some incorrect melodic patterns.
3	The child exhibited a sense of the melodic contour but was unable to maintain a tonal centre or modulated the tonal centre established at the beginning of the song.
2	The child consistently uses a singing voice but exhibited no sense of the general melodic contour.
1	The child occasionally uses a singing voice.

APPENDIX G

Subject I. D.: ____________ Score: _________

Criterion Song 2

I Am the Wind

Scores	Description of singing achievement
5	The child maintained a tonal centre without modulation and sang most of the melodic patterns accurately.
4	The child was able to maintain a tonal centre and the general melodic contour but sang some incorrect melodic patterns.
3	The child exhibited a sense of the melodic contour but was unable to maintain a tonal centre or modulated the tonal centre established at the beginning of the song.
2	The child consistently uses a singing voice but exhibited no sense of the general melodic contour.
1	The child occasionally uses a singing voice.

APPENDIX H

ACOUSTIC DATA RECORD

Subject:: ____________ **CS1 score:** ______ **CS2 score:** ______ **Composite score:** ______

Criterion Song 1

m3 M3 P4 M6 M3 P5 m3 M3 P4 P4

	s	*m*	*d*	*s*	*m*	*d*	*s*	*m*	*d*	*s*	*d*
Hz											
Cent											

P4 P4 M3 M3 M3 m3 m3 M3 P4 P4

	d	*s*	*d*	*m*	*d*	*m*	*s*	*m*	*d*	*s*	*d*
Hz											
Cent											

Criterion Song 2

M2 m2 M2 M2 m2 m2 M2 M2 m3 P4 P4 P5 M3 m3

	l	*t*	*d*	*r*	*m*	*f*	*m*	*r*	*d*	*l*	*l*	*m*	*l*	*m*	*m*	*d*	*l*
Hz																	
Cent																	

APPENDIX I
ACOUSTIC DATA EVALUATION SHEET

Subject I. D. : ________

Criterion Song 1	**Standard Model**
Range	(the actual criterion song for comparison)
Lowest pitch ______ Hz; Highest pitch _____ Hz	
Range: ______ cents	200 cents
Tonic	
Highest ______ Hz; Lowest ______ Hz.	
Deviation: _____ cents	
Mean: ______ Hz / _____ = _____ Hz	
Deviations	
Lowest pitch _____ Hz; deviation from tonic _____ cents	(L) P4; 500 cents
Highest pitch _____ Hz; deviation from tonic _____ cents	(H) P5; 700 cents
Interval Accuracy	
Correct intervals: _____ out of _____ sung	Total: 20 intervals
Intervals sung in correct direction: _____	
Score on the 5 point rating scale __ 1 __ 2 __ 3 __ 4 __ 5 ___	
Criterion Song 2	
Range	
Lowest pitch ______ Hz; Highest pitch______ Hz	
Range: ______ cents	1300 cents
Tonic	
Highest ______ Hz; Lowest ______ Hz.	
Deviation: ______ cents	
Mean: ______ Hz / _____ = _____ Hz	
Deviations	
Lowest pitch _____ Hz; deviation from tonic _____ cents	(L) P4; 500 cents
Highest pitch _____ Hz; deviation from tonic _____ cents	(H) m6; 800 cents
Interval Accuracy	
Correct intervals: _____ out of______ sung	Total: 14 intervals
Intervals sung in correct direction: _______	
Score on the 5 point rating scale __ 1 __ 2 __ 3 __ 4 __ 5 __	

APPENDIX J
SCORING PROCEDURES FOR DATA DERIVED FROM ACOUSTIC ANALYSIS

Scores	Description of singing achievement	Description of Acoustic Data
5	The child maintained a tonal centre without modulation and sang most of the melodic patterns accurately.	The range was comparable to the standard model with a well established tonal centre (the deviation from the highest tonic sung to the lowest tonic sung was less than 350 cents) and at least 85% of the intervals were accurate.
4	The child was able to maintain a tonal centre and the general melodic contour but sang some incorrect melodic patterns.	The range was comparable with the standard model with a well established tonal centre (the deviation from the highest tonic sung to the lowest tonic sung was less than 350 cents). Intervals were sung in correct directions but only between 50 to 85% of them were accurate.
3	The child exhibited a sense of the melodic contour but was unable to maintain a tonal centre or modulated the tonal centre established at the beginning of the song.	The range was comparable with the standard model and although more than 75% of the intervals were sung in correct directions, less than 50% were sung accurately. The tonic was not securely established (the deviation from the highest tonic sung to the lowest tonic sung was more than 350 cents).
2	The child consistently uses a singing voice but exhibited no sense of the general melodic contour.	The total range from the highest to the lowest pitch sung was wider than 550 cents but less than 75% of the intervals were sung in their correct directions.
1	The child occasionally uses a singing voice.	The total range from the highest to the lowest pitch sung was narrower than 550 cents.

APPENDIX K
HISTOGRAMS OF *PMMA TONAL* SCORES