INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

THE EFFECT OF SINGING MODE AND SEATING ARRANGEMENT ON CHORAL BLEND AND OVERALL CHORAL SOUND

by

Elizabeth Ekholm

Faculty of Music,
McGill University, Montreal

August, 1999

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfilment of the requirements of the degree of
Doctor of Philosophy in Music Education

Abstract

This study examined the effect of choristers' vocal production and seating arrangement on evaluations of choral blend and overall choral sound. Two singing modes were studied: (1) "soloistic" singing, in which choristers attempted to maintain normal solo vocal production; and (2) "blended" singing, in which choristers tried to maximize homogeneity of ensemble sound. Seating arrangements examined were: (1) random sectional seating, and (2) sectional seating according to "acoustic matching" of voices Crossing the 2 factors produced 4 experimental conditions.

An ad hoc choir, composed of 22 university voice majors, was recorded singing 4 choral pieces under each experimental condition. Thirty-seven choral conductors, 33 voice teachers and 32 nonvocal musicians rated the performances according to 7 standard evaluative criteria, and wrote comments. Eight of the choristers were recorded individually during the performances. These choristers were also recorded singing their parts as solo songs. Twelve voice teachers evaluated the vocal production of these 8 choristers. They ranked the 5 performances of each piece (4 choral and solo) by each chorister, and wrote comments. All choristers rated the 4 choral experimental conditions for each piece on a 5-point scale, according to vocal comfort and choral sound, and wrote comments.

Analysis of choral performance ratings revealed a significant seating arrangement effect in favor of acoustic seating over random seating. Singing mode had a significant effect in favor of blended singing over soloistic singing on choral conductors' ratings for

all seven criteria. This effect was observed in the voice teachers only for the criterion "blend/homogeneity." Otherwise, no significant singing mode effect was observed in the voice teachers and nonvocal musicians.

On the basis of individual vocal production, voice teachers ranked solo singing significantly higher than the 2 choral singing modes, and soloistic choral singing higher than blended choral singing. Moreover, they ranked acoustic seating higher than random seating.

Choristers rated acoustic seating higher than random seating for vocal comfort and choral sound. They preferred blended singing to soloistic singing for choral sound. For vocal comfort, sopranos preferred soloistic singing to blended singing, tenors preferred blended singing to soloistic singing. Altos and basses showed no singing mode preference for vocal comfort.

Résumé

Cette étude a porté sur l'effet de la production vocale et de la disposition des choristes sur l'évaluation de l'homogénéité et de la qualité d'ensemble du son choral. Deux modes de chant ont été étudiés: (1) le mode «solistique», dans lequel les choristes ont essayé de conserver leur production vocale normale de soliste, et (2) le mode «fondu», dans lequel les choristes ont tenté de maximiser l'homogénéité du son d'ensemble. Les dispositions des choristes étudiées ont été: (1) la disposition par section (SATB) au hasard, et (2) la disposition par section selon la compatibilité acoustique des voix. La combinaison des deux facteurs a résulté en quatre conditions expérimentales.

Un choeur ad hoc, composé de 22 étudiants universitaires en concentration chant, a enregistré quatre pièces chorales selon chacune des conditions expérimentales. Trente-sept chefs de chorales, 33 professeurs de chant et 32 autres musiciens ont évalué les prestations selon sept critères d'évaluation et ont écrit des commentaires. Huit choristes ont également été enregistrés individuellement durant l'enregistrement de la chorale. Chacun de ces huit choristes a également enregistré sa partie en solo pour chacune des quatre pièces. Douze professeurs de chant ont évalué la production vocale des ces huit choristes. Ils ont rangé les cinq prestations de chacun des choristes pour chaque pièce (quatre chorales et solo) et ont écrit leurs commentaires. Tous les choristes ont évalué les quatre conditions expérimentales pour chacune des pièces, selon une échelle de 1 à 5, pour le confort vocal et le son choral. Ils ont également écrit leurs commentaires.

L'analyse des évaluations des prestations chorales a révélé un effet significatif en faveur de la disposition «acoustique» par rapport à la disposition au hasard. Le mode de chant a eu un effet significatif en faveur du mode fondu par rapport au mode solistique chez les chefs de chorales, pour les sept critères d'évaluation. Cet effet a été observé chez les professeurs de chant uniquement pour le critère d'homogénéité. Aucun autre effet significatif du mode de chant n'a été observé chez les professeurs de chant et les autres musiciens.

Pour la production vocale individuelle, les professeurs de chant ont accorde un rang significativement supérieur au chant solo par rapport aux deux modes de chant choral, et un rang supérieur au mode choral solistique par rapport au mode choral fondu De plus, ils ont rangé la disposition acoustique plus haut que la disposition au hasard

Les choristes ont évalué la disposition acoustique supérieure à la disposition au hasard tant pour le confort vocal que pour le son choral. Ils ont préféré le mode de chant fondu au mode solistique pour le son choral. Pour le confort vocal, les sopranos ont préféré le mode solistique au mode fondu, les tenors ont préféré le mode fondu au mode solistique, tandis que les altos et les basses n'ont montré aucune préférence de mode.

Table of contents

List of tables

List of appendices

Acknowledgments

No undertaking of this scope can be accomplished without the help and support of many people. First of all, I wish to express my gratitude to my thesis supervisor, Prof. Joel Wapnick, whose expertise in music education research has been an invaluable resource. His patience and meticulousness have continuously challenged me to give my best, while inspiring confidence in my abilities to meet his standards. For everything he has taught me, and for his unfailing support, I sincerely thank him.

I would also like to express my appreciation to Prof. John Baboukis. Prof. John Klepko, Mr. Justin Kurtz, Ms. Poppy Crum, Mr. Hugh Cawker and Ms. Pamela Reimer for their expert assistance with the recordings, as well as to Mr. Gary Bowen, Ms. Margaret Parkin and Ms. Margot Rejskind for their kind help with the videotape. My heartfelt thanks also goes out to the singers, whose enthusiasm and professionalism carried us all through the long hours of rehearsing and recording. To all the choral conductors, voice teachers and other musicians who took time out of their busy schedules to participate in this study, I owe a large debt of gratitude.

Financial assistance for this research project came from the Quebec provincial government's Fonds pour la formation de chercheurs et l'aide à la recherche (FCAR) and from McGill University in the form of a Max Stern Fellowship in Music and a Humanities Research Grant.

Finally, I would like to thank my family: my children, Olivia, Caroline and Philippe, for their patience and understanding when "Mom was too busy;" and my husband, André Morin, who was always available when I needed feedback, suggestions, encouragement and inspiration. He helped me to see the forest when I got lost in the trees. I dedicate this work to him with my love.

Introduction

Choral singing has an important place in the field of music education. Choirs provide an opportunity for people of all ages and levels of musical ability to participate in musical performance. For singers who aspire to a career in musical performance, choirs provide excellent musicianship training. Participation in choirs is thought to enhance music reading skills, ensemble skills, understanding of musical styles and performance practices, ear-training skills, diction skills, and to broaden musical experience and knowledge of repertoire. Moreover, many young professional singers rely on choral work to round out their earnings while waiting for their solo career to flourish.

There is, however, much anecdotal evidence to suggest that voice teachers and choral conductors often do not agree on the type of vocal production demanded from singers (Goodwin, 1980a). Voice teachers tend to promote a type of vocal technique developed by solo singers, the aim of which is to achieve maximum resonance with minimum vocal effort. This type of technique makes it possible for a singer's voice to be audible above an orchestral accompaniment. Choral conductors, on the other hand, tend to strive for an ideal ensemble sound in which individual voices are not perceptible. In their quest for homogeneity, or "blend," choral conductors often require singers to alter their vocal production from that taught in the voice studio.

Acoustic studies on the differences in vocal production between solo and choral singing have shown that in trying to blend with an ensemble, singers produce significantly less energy in the "singer's formant" region (around 3 kHz) and a narrower vibrato than

in solo singing (Goodwin, 1980b; Rossing, Sundberg and Ternström, 1985, 1986, 1987). The "singer's formant" is the region of the frequency spectrum associated with the projection or carrying power of the voice (Sundberg, 1987). As energy in the singer's formant region is the principal acoustic factor that distinguishes trained from untrained singers (Teie, 1976; Magill and Jacobsen, 1978), the acoustic studies show that choral singers tend to use a vocal production that differs significantly from that taught by voice teachers. This modified vocal production enables trained singers to achieve greater uniformity of sound with other members of the choir, many of whom may have no voice training (Goodwin, 1980b; Rossing et al., 1985). As a result, voice teachers may find that the vocal habits they teach are being undermined by the type of vocal production their students are encouraged to use during choir rehearsals. Voice teachers have argued that certain kinds of vocal production adopted by singers in order to enhance choral blend may be not only counter-productive to singers' mastery of solo vocal technique but also potentially harmful to their voices. This attitude is reflected in the following statement by the American Academy of Teachers of Singing (AATS):

> Natural, free emission of tone by each individual need never be sacrificed in order to achieve the desired result in choral singing. A student of singing should be encouraged to participate in choral groups, but only if by so doing he can, as an individual, enhance his vocal development, free from strain or tension. In such groups where "imitation," the so-called "straight tone," or other dubious methods are practiced he has nothing to gain and much to lose. (American Academy of Teachers of Singing, 1964).

It is not surprising, therefore, that voice teachers have been known to advise their most promising students against singing in choirs.

Choirs differ a great deal regarding the extent to which blend is required. At one extreme is the vibratoless uniformity of the so-called "Minnesota sound," epitomized by the St. Olaf College Choir, or of the Scandinavian choirs on which the St. Olaf College Choir was modelled, in which the individual singer is required to suppress all individuality of sound. At the other extreme is the opera chorus, in which singers use a solo operatic vocal production (Knutson, 1987; Swan, 1988). Most choirs are somewhere in between these two extremes.

An interesting method for achieving choral blend, associated with Weston Noble and his followers, involves using trial and error to find voices that "match" acoustically and then to seat singers accordingly (Giardiniere, 1991). In a perfect acoustic match voices appear to fuse together so that individual voices are not discernable. An acoustic match can consist of either two voices that are very much alike or two voices that are quite different but somehow complement each other. Singers in the choir are positioned in a seating arrangement so that voices that match are next to one another. Proponents of this method believe that blend is thus enhanced naturally through acoustic phenomena, allowing singers substantial individuality in vocal timbre and freedom in vocal production. Other aspects of ensemble singing such as intonation accuracy, vowel matching, and simultaneity of rhythmic impulse, dynamics and phrasing have been discussed by expert choral conductors (Knutson, 1987) under the broad heading of "choral blend." As these aspects of ensemble singing are largely independent of vocal production (i.e. they can be achieved using various types of vocal production), they are not the primary focus of this study. Although the term "choral blend" is sometimes used as an umbrella term to

encompass all aspects of good ensemble singing, in this study it refers specifically to homogeneity of choral tone.

Purpose of the study

The purpose of the study was to investigate the effect of individual vocal production and seating arrangement on evaluations of choral blend and overall choral sound. The types of vocal production studied in a choral setting were: (1) "soloistic" vocal production, used by choristers who are trying to maintain their normal solo vocal production, and (2) "blended" vocal production used by choristers who are trying to achieve maximum homogeneity of ensemble sound. The seating arrangements examined were: (1) random seating of singers within sections (soprano, alto, tenor and bass), and (2) seating of singers within sections according to trial and error "acoustic matching" of their voices, as derived from the practices of Weston Noble (Giardiniere, 1991).

In the acoustic-matching method of arranging singers, three singers from a given section were asked to stand in a row and sing a familiar short excerpt together in unison, using their most comfortable vocal production. Every possible arrangement of those three singers was heard by three choral conductors, who decided which arrangement produced the best blend. A fourth singer from the same section was then added, and each possible position of that singer was tried before deciding on the most blended arrangement. Additional singers were added one at a time and tested in the same way until all the singers in the section had been placed. It appeared that the phenomenon of vocal fusion was readily perceptible, as agreement among the conductors was easily achieved.

For purposes of this study, the arrangement among sections was kept constant. Singers were place in two rows. The front row consisted of sopranos to the left of the conductor and altos to the right of the conductor. In the back row, basses were placed behind the sopranos, and tenors behind the altos. This formation has been used both for performing and recording by eminent choirs such as the St. Olaf Choir and the Westminster Choir. (Miller, 1992)

Specific issues

Specific questions that were asked were:

(1) Do trained singers modify their normal solo vocal production in a choral setting? If so, how do voice teachers evaluate their modified vocal production?

(2) How is choral blend affected by the type of vocal production used by choristers? Is it possible to achieve an acceptable level of choral blend without modifying solo vocal production?

(3) How is evaluation of overall choral sound affected by the type of vocal production used by choristers?

(4) Does positioning of choristers according to acoustic matching of voices have any effect on choral blend or overall choral sound?

(5) What are the preferences of choristers with regards to singing under the various conditions in this study?

(6) How do aesthetic preferences regarding choral sound compare among choral conductors, voice teachers, and other musicians? What are their levels of interjudge agreement?

Definition of terms

acoustic seating arrangement	- seating of singers within each section according to trial and error "acoustic matching" of their voices, as derived from the practices of eminent American conductor, Weston Noble.
blend (noun)	- a phenomenon resulting when a number of different voices appear to fuse together so that individual voices are not discernable.
blend (verb)	- to mix (components) so that their individuality is obscured in the product (Oxford International Dictionary, 1957)
blended singing mode	- singers were instructed to focus on producing an ensemble sound that was as homogeneous as possible; it was further explained that they should strive to eliminate perceptibility of individual voices.
choral tone quality	- 1. overall aesthetic value of the sound produced by a choir - 2. essential or distinguishing characteristics of the sound produced by a choir.

diction - intelligibility to the listener of the words sung by a singer or group of singers. This generally depends upon two components: enunciation and pronunciation. Enunciation is defined as the clarity of the sounds being articulated by the singer; pronunciation refers to the degree to which the vowel and consonant sounds produced by the singer accord with standards for a given language.

homogeneity - uniformity;

- for purposes of this study, same as "blend;"
- when voices are mixed so that their individuality is obscured in the product, an impression of uniformity or homogeneity is created.

intonation accuracy - degree to which the singers adhere to the pitches notated in the musical score, as perceived by musically sophisticated listeners; extent to which choir is perceived to be "in tune."

pitch accuracy - for purposes of this study, same as "intonation accuracy."

random seating arrangement - for purposes of this study, random positioning of singers within each section.

rhythmic precision - extent to which singers adhere to note durations in the musical score, as perceived by musically sophisticated listeners; can also refer to the degree of simultaneity of rhythmic impulses achieved by a choir.

seating arrangement - order in which singers in a choir are positioned.

singing mode - manner in which a singer uses his/her voice in singing.

soloistic singing mode - singers were instructed to use their voices in the manner taught to them by their voice teachers; they were told to sing as though they were in a quartet of soloists, attending to all other aspects of ensemble singing, such as tempo, dynamics, phrasing, diction and style.

tone quality - 1. essential or distinguishing characteristics of the sound produced by a singer, a group of singers or an instrument;
- 2. degree to which sound produced by a singer, a group of singers or an instrument adheres to standards of beauty.

vocal production - manner in which a singer uses his/her voice in singing.

Hypotheses

(1) The null hypothesis that there would be no significant difference in choral performance ratings between performances in the soloistic singing mode and performances in the blended singing mode was tested.

(2) A second null hypothesis tested in this study was that there would be no significant difference in choral performance ratings between performances in the acoustic seating arrangement and performances in the random seating arrangement.

(3) Furthermore, this study tested the null hypothesis that interjudge agreement among the judges evaluating choral performance would be no greater than that predicted by chance. This result would suggest that evaluation of choral performance is whimsical and not based on any generally accepted standards.

Possible outcomes

Voice teachers could be expected to prefer vocal production in the soloistic singing mode to vocal production in the blended singing mode. Choral conductors, on the other hand, might be expected to prefer the choral blend in the blended singing mode to that in the soloistic singing mode. Herein may lie source of the conflict between voice teachers and choral conductors.

According to the literature on choral conducting, one would expect a preference by both listener and chorister for the acoustic seating arrangement over the random seating arrangement. Proponents of acoustic seating maintain that choristers find it easier to sing freely and in tune when they are acoustically matched with choristers beside them.

There could be, moreover, an interaction effect favoring the combination of soloistic singing mode with acoustic seating arrangement. Under this condition (soloistic/acoustic), one might expect the soloistic singing mode to enhance vocal tone quality, while an acceptable degree of blend is achieved through voice matching. Such a finding would imply that choral directors may be able to achieve an acceptable degree of choral blend through the use of acoustic techniques, without any modification of vocal production.

A comparison of choral tone quality preferences among choral conductors, voice teachers and other musicians might uncover systematic differences in aesthetic preferences among these groups. Interjudge agreement within these groups could be expected to be highest among choral conductors, as they have the greatest experience in judging choral sound. Voice teachers might be expected to exhibit somewhat greater interjudge reliability than nonvocal musicians, due to voice teachers' greater experience in evaluating vocal tone quality. In spite of cultural variations in choral aesthetics, one would, nevertheless, expect to find some general standards operating in the evaluation of Western art music. Interjudge agreement among musically sophisticated judges would thus be expected to be above chance level, though probably not very high. Previous research in musical evaluation, particularly in the evaluation of tone quality, has produced low to moderate interjudge correlations (Abeles, 1973; Burnsed, Hinkle and King, 1985; Campbell, 1971; Ekholm, 1997; Wapnick and Ekholm, 1997).

Limitations of the study

(1) This study focussed on two factors commonly assumed to affect choral sound: individual singing mode and seating arrangement. Of all the possible levels of these factors, only two were selected for testing: soloistic versus blended singing mode; and random versus acoustic seating arrangements, within sections.

(2) This study dealt specifically with university-level voice majors, who have had both solo and choral experience. It was important to ensure that all the singers were capable soloists as well as experienced choristers. The university setting was considered the most appropriate for examining possible conflicts between solo and choral singing. In a university setting, voice majors are often required to participate in both solo and choral singing. Moreover, the students' vocal technique is usually still in its formative stages.

(3) Judges in this study were chosen from professional musicians and/or music teachers at the university level. As in any study in which subjects participate on a voluntary basis, some systematic bias on the part of those who agreed to participate cannot be ruled out.

(4) There was no control over playback equipment, although specific instructions on how to enhance the fidelity were provided.

(5) The repertoire in this study was chosen to represent typical Western choral literature from four periods of music history: Renaissance, Classical, Romantic and Twentieth Century. For purposes of limiting variables, it was thought that all the pieces should be in the same language. The language chosen was Latin, as it is the most widely known language of choral singing in the Western art music tradition. For the sake of

uniformity, the Italian school of pronunciation was chosen. The choice of Latin, however, had the effect of limiting the repertoire to sacred music.

(6) Only immediate effects of the variables were studied. It is possible that long-term effects of seating arrangement or singing mode may differ from the results of this study.

(7) In recording the choral excerpts, the greatest accuracy in sound reproduction was sought. Therefore, artificial enhancement of the choral sound through recording technology or concert hall acoustics was avoided. A small, relatively dry concert hall was used. It was believed that differences in choral sound resulting from the different experimental conditions might be obscured in a more resonant concert hall, or in a recording in which electronic enhancement techniques (such as reverb) were employed.

Assumptions pertinent to the study

It was assumed:

- that the judges were competent and responded honestly in their evaluations.
- that the three groups of participants - choral conductors, voice teachers, and nonvocal musicians - were of comparable musical ability and level of achievement.
- that the singers who were not being individually monitored were, in fact, responding to instructions in the same way as the monitored singers, and that all singers were responding to the best of their ability.

Significance of the study

Research in the field of choral music education has examined many topics, including rehearsal technique, conducting technique, choral curriculum, vocal physiology, vocal pedagogy for the choral rehearsal, choral acoustics, seating arrangements, training of choral conductors, and evaluation of choral performance. A few studies have dealt with choral blend. In two of these studies (Giardiniere, 1991; Knutson, 1987), experienced choral conductors were interviewed on their ideas and practices regarding choral blend. Other studies on choral blend compared individual voices in solo and choral singing from an acoustical perspective (Goodwin,1980b; Rossing et al., 1985, 1986, 1987). One study examined the influence of vibrato rate and extent on judges' perception of blend in pairs of singers (Trevor, 1977). No empirical studies, however, were found which dealt with the effect of individual vocal production on perceived choral blend and overall choral sound, nor with the relationship between seating arrangement and vocal production. Neither were any empirical studies found which compared evaluation of choral sound by choral conductors, voice teachers and other musically sophisticated listeners. The present study may thus be the first empirical study on these important topics.

In this study, I investigated the relationships between singing mode, seating arrangement, and choral sound. Findings from this study might have important implications for choral conductors, voice teachers and singers. They might help to resolve, or at least better understand, the controversy between voice teachers and choral conductors.

Choral blend

The impressive number of articles dealing with choral blend would seem to indicate the relative importance of this aspect of ensemble singing. Interesting as many of these articles are, very few are research-based, being instead mostly anecdotal (Wyatt, 1967a, 1967b, 1968; Bolster, 1983; Bravender, 1984). Nevertheless, there is a growing body of experimental and descriptive research concerned with this important facet of choral performance.

Hunt (1968) performed spectrographic analyses of the choral sound produced when choral groups sang in unison on the vowels /i/, /e/, and /a/. He found that perception of choral blend was directly related to tuning of vowel formant frequencies. Trevor (1977) asked a panel of seven judges to listen to 50 samples of paired voices and judge whether or not the pair was blended. He discovered that the combined effects of vibrato rate and extent accounted for 15% of the variation in judges' perception of blend. As vibrato rates and extents in two singers became more similar, the pair was more likely to be perceived as blended. Goodwin (1980b) recorded sustained tones sung by individual singers in two modes: solo, and while trying to blend with an ensemble heard through headphones. Spectral analysis revealed that the tones produced in the blended mode had stronger fundamentals and first formants, fewer and weaker upper partials, and weaker second and third formants.

Kendall and Carterette (1993) in their study of timbre in orchestral wind instruments found that the perception of blend in a pair of instruments correlated negatively with identification of individual instruments. The oboe, rated as highly "nasal," yielded the lowest blend ratings and highest identifiability among the five instruments studied (oboe, trumpet, flute, clarinet, and alto saxophone). Spectral analyses showed that blend correlated negatively with relative energy in the upper partials. It was observed, moreover, that the presence of vibrato in the flute and trumpet dyads tended to aid in separation and identification of the instruments. In other words, time-variant spectral modulation appeared to be a perceptual cue for instrument identity.

Coleman (1994) asked 20 adult singers from one choral group to produce sustained vowels and to sing a short song excerpt at three different dynamic levels: pianissimo, mezzoforte, and fortissimo. Large differences in dynamic range were found among the singers; sustained vowels tended to be louder than the song excerpt; and mezzoforte levels tended to be closer to fortissimo than to pianissimo levels. Total dynamic ranges among individuals in this study varied from 11 to 33 dB. Coleman opined that choral blend would be enhanced more readily by asking the "powerful voices" to reduce their sound pressure level (SPL) than by attempting to increase the overall SPL of the rest of the choir.

Knutson (1987) compiled the opinions of 13 eminent choral conductors on what constitutes choral blend and how to achieve it. He found that rationales and practices of these conductors with respect to choral blend often differed. Definitions of choral blend were broad and encompassed virtually all aspects of good ensemble singing. He

concluded that the phenomenon of choral blend was largely a matter of personal preference. Furthermore, his study indicated a trend away from the historical practice of subordinating or altering the vocal quality of the individual singer in order to achieve blend. Contemporary choral conductors in his study leaned toward a rehearsal style that strove to achieve an acceptable level of uniformity without sacrificing healthy vocal technique. This trend was based on a belief that restricting the vocal development of the individual singer is unhealthy and detrimental to choral sound.

Giardiniere (1991) reported on Weston Noble's method of voice matching to enhance choral blend. A vocal match was defined as the fusion of two or more voices into a "conglomerate in which individual vocal characters are present but not noticeable." (p.111) Voice-matching trials conducted by the celebrated American choral conductor were recorded and rated by 111 expert choral conductors. Statistical analysis of the level of agreement among conductors regarding the best vocal match in each set of samples led Giardiniere to conclude that there was a perceptible phenomenon of voice matching that transcended individual taste or preference. Giardiniere observed that individual voices in a matched pair or trio could sound either very similar, or very different but complementary. As spectral analysis was not used in this study, acoustic correlates of the vocal match were not determined.

Other aspects of choral sound

Several studies have considered important aspects of choral sound other than blend. Researchers have investigated intonation accuracy in choirs, intensity of sound produced by choristers, and balance of sound intensity among sections of the choir.

Intonation accuracy. Lottermoser and Meyer (1960) examined intonation in commercial recordings of four choirs and found that these choirs tended to sing major thirds sharp (421 cents average, 100 cents = 1 semitone), minor thirds flat (275 cents average), and fourths, fifths and octaves very near to just intonation. They measured the bandwidth (at 70% of peak amplitude) of the distribution of phonation frequencies produced by the choristers singing in unison and observed that the average dispersion in intonation accuracy was +/- 25 cents over all the choirs. The narrowest dispersion was +/- 10 cents, and the widest was +/-50 cents. The sharpened major third and flattened minor third suggest that the choristers' intonation was closer to Pythagorean intonation than to equal temperament. This is consistent with other research on intonation which found that singers and string players tended to use untempered intonation, particularly Pythagorean, when they were not performing with keyboard instruments (Revesz, 1954; Green, 1937; Nickerson, 1949).

Sundberg (1978) examined the effects of vibrato and singer's formant on pitch perception among musicians. When musicians were asked to match a pitch to a synthesized stimulus, it was observed that perceived pitch in a tone with vibrato corresponded closely to the linear average of the time-varying fundamental frequency. Although accuracy of pitch perception was slightly reduced when vibrato was added to low fundamental frequencies (70 and 115 Hz), Sundberg concluded that the addition of vibrato or singer's formant to a vowel sound had no significant effect on accuracy of pitch perception.

When amateur male choristers were asked to sing a given interval above a synthesized stimulus pitch, Ternström and Sundberg (1982) found that the absence of

vibrato, presence of partials in the stimulus tone that were common to the response tone, and presence of high partials facilitated intonation accuracy. The presence of vibrato did not affect the accuracy of pitch perception for a single complex tone. In tuning an interval, however, one clue to pitch accuracy seemed to be the presence of beats or "roughness" that arose when common partials were not exactly tuned. As beats cannot occur in tones that have periodic variation in their fundamental frequencies, this clue to pitch accuracy would be eliminated in tones with vibrato. Ternström and Sundberg further noted that standard deviations of 10-15 cents within a group of singers were typical of musically acceptable performances, and that standard deviations 2 to 3 times greater were observed with certain variations of the stimulus tone. More skilled singers were influenced less by stimulus tone manipulations than were less skilled singers.

Ternström and Sundberg (1986) concluded from research on choral acoustics that intonation accuracy of choristers depended on several factors. One factor was the difference in SPL between the sound the chorister heard from his/her own voice (feedback) and from the other choristers (reference). Accuracy deteriorated when reference was more than 5 dB louder, or more than 15 dB softer, than feedback. Intonation accuracy also depended on spectral characteristics of the sound choristers were hearing. The presence of common partials in intervals provided clues to intonation by producing beats when inexactly tuned. Moreover, the presence of high partials improved intonation accuracy, probably because low partials tended to be masked by the singer's own voice. The presence of vibrato reduced intonation accuracy, probably by eliminating the possibility of using beats between common partials as a clue to intonation accuracy.

Finally, vowels in which the first or second formant was almost identical with one of the common partials in an interval improved intonation by reinforcing common partials.

Ternström, Sundberg and Collden (1988) investigated the phenomenon of "intrinsic pitch" of vowels as it pertained to singing. This phenomenon has been described in speech production as a change of fundamental frequency associated with a change of vowel. Experienced choristers sang sustained tones with a change of vowel in mid-tone, both with normal auditory feedback and with auditory feedback masked by noise through headphones. They found that the vowels /i/ and /y/ tended to raise fundamental frequency, while the vowels /ɛ/ and /a/ lowered it. The presence of auditory feedback resulted in higher intonation accuracy than did the absence of auditory feedback.

Ternström (1993) examined tolerance levels and preferences of experienced listeners for fundamental frequency (F_0) dispersion (also called pitch scatter) and formant frequency (F_3-F_5) dispersion in synthesized unison choir sounds. He found that the "maximum tolerable" pitch scatter was, on the average, 14 cents standard deviation (i.e. 95% of the singers producing an average F_0 within +/- 28 cents of the group average F_0). The preferred pitch scatter was from 0-5 cents standard deviation. The results for formant dispersion (smear) were less conclusive, as this proved more difficult to judge. The "maximum tolerable" standard deviation was 14%, and the preferred standard deviation was 7%.

Ternström concluded that a singer's timbre is largely determined by the higher formants - F_3 , F_4 and F_5 - whose frequencies are directly related to vocal tract length and are largely independent of vowel. This finding has implications for choral blend. In the

strictest sense choral blend involves the combining of individual vocal timbres into a homogeneous choral sound in which no individual timbre can be heard. Attenuation of the higher formants, as observed in choral singing, would tend to enhance choral blend by obscuring differences in vocal timbre among individual singers.

Vocal intensity. Tonkinson (1994) asked singers of various experience levels to sing *The Star Spangled Banner* while listening to a pre-taped choir through headphones On the pretest there was a tendency for singers to increase vocal intensity as intensity of the taped choir increased. This tendency is known as the "Lombard" effect (Tonkinson, 1994, p.24). Singers were then instructed to resist this tendency. Posttest results suggested that choral singers can learn to resist the Lombard effect and to consciously regulate their vocal intensity in the presence of a masking sound.

Ternström (1994) devised a method for measuring the feedback (sound of chorister's own voice) to reference (sound of rest of choir) ratio as experienced by choristers under live performance conditions. The method used binaural microphones worn by choristers and took into account the combination of choir and room acoustics The feedback-to-reference ratio averaged over 12 choristers, chosen at random from all four sections of the choir, was observed to be +3.9 dB, with a range of from +1.5 dB to +7.3 dB. Such a method would make it possible to determine optimal feedback to reference ratios for choral singers. This in turn could be useful in determining ideal acoustic properties for concert halls and amplification for choral performance.

Balance. Killian (1985) investigated preferences of choral conductors and high school students for balance in four-voice chorales. Subjects stated their preferences

and also operated volume controls of a four-channel tape recorder as an operant measure of preferred balance. Although subjects stated a preference for "equally balanced" voices, in the operant measure subjects preferred less bass relative to other voices. Moreover, subjects could discriminate when a single voice part was louder than the other three, but adjusted it significantly louder than the other voices, indicating an initial perception effect. There was no significant difference between students and choral conductors regarding balance preference. Men, however, preferred significantly louder overall levels than women.

These studies have helped to define and measure, either objectively or subjectively, basic elements of choral sound and their effects on choristers and listeners.

Vocal production

Studies of individual vocal production by singers in choirs have examined the effects of voice training on vocal production, differences in vocal production between solo and choral singing, vibrato in choral singing, and vocal pedagogy in choirs.

Effects of voice training. Teie (1976) observed that trained singers had significantly greater energy in the third formant region of the spectrum than untrained singers, and that this was the principal acoustic factor that distinguished the two groups. This finding was supported by Magill and Jacobsen (1978) who also observed that energy in the frequency range between the third and fourth speech formants, known as the singing formant, was a significant aspect of all singing voices, but more so in professional singers. Rossiter and Howard (1994) in their study of voice synthesis also noted that

increasing energy in the singer's formant region resulted in a more "professional" voice quality, albeit with some loss of vowel clarity. They found, furthermore, that an increase in the "closed quotient" (CQ), or ratio of how long vocal folds are in contact relative to how long they are apart, and in energy in the singer's formant region was a function of voice training. The relationship, however, was not strictly a linear function of amount of training. A low CQ produced a soft, breathy sound, and a high CQ resulted in a more professional sound.

Solo versus choral singing. Harper (1967) performed spectrographic analyses on recordings of vowels sung in both solo and choral modes. He found no significant differences between solo and choral enunciation of the vowels studied ("a," "i," and "u"), as indicated by the locations of the first two formants. He did observe, however, that in choral mode there tended to be more energy of partials located between formant regions, than in solo mode. Vennard (1967) explained this phenomenon by pointing out that a good solo singer tended to concentrate the sound energy in relatively few partials, a practice that singers call "focus." The greater intensity of a focussed tone would be an advantage in solo singing, where the soloist needs to "stand out" from the accompaniment. In choral singing, however, the singer is careful not to "stand out," and so the practice of focussing the tone would not be advantageous.

Rossing et al. (1985, 1986, 1987) studied spectral differences between solo and choral singing. For male singers, they found that in solo singing the singer's formant was more prominent and the lower partials (below 500 Hz) less prominent than in choral singing, even when loudness (SPL) was kept constant. In order to amplify the singer's

formant region, singers decreased the distance in frequencies between the third and fifth formants. This resulted in a clustering of the third, fourth and fifth formants, which boosted the amplitude of the sound energy in that spectral region. It was concluded that singers made these changes between solo and choral singing through vocal tract adjustment and glottal adjustment. The basses and baritones adjusted their volume levels to levels of other singers when singing in a choir. In solo mode, however, their volume levels depended much less on the level of the piano accompaniment than they had on the levels of other singers in choral singing. All the sopranos examined, especially two opera singers, produced more energy in the 2-4 kHz range and slightly more vibrato in solo singing than in choral singing. Sundberg (1987) concluded that there were some basic differences in vocal production between solo and choral singing, and that there were differences in the kind of voice timbre that was sought in each type of singing.

Vibrato in choral singing. Trevor's (1977) discovery that similarity of vibrato rates and extents in a pair of singers accounted for 15% of the variation in judges' perceptions of blend has already been cited. Goodwin (1980b) noted that sopranos attempting to blend with an ensemble tended to reduce the extent of their vibrato, especially if they produced tones with strong upper partials. He gave two possible reasons for this: (1) vibrato frequency oscillations were sometimes wide enough among the upper partials to carry them in and out of the third formant frequency range with each vibrato cycle, thereby creating a noticeable on/off switching of strong upper partials; (2) tones with wider vibrato may be perceived by the listener as being louder than tones of the same physical intensity that have less vibrato (Benade, 1976).

Ternström (1991) explained that the greater use of vibrato by solo singers might be an effective way to promote perceptual differentiation of voice from accompaniment (p.139). He stated that it was impossible for any singer to produce perfectly straight tones. In synthesizing the human voice, he found that fluctuations in fundamental frequency were an important cue to perceived naturalness of the voice. Regarding vibrato as a type of "flutter," Ternström stated that "without flutter, a synthesized choir [did] not sound like an ensemble of voices, but more like a mediocre electric organ." (p.141) With increasing flutter, a synthesized choir sounded at first like a few singers, then like a larger choir, then less and less stable, and finally chaotic. Ternström analyzed the voice flutter of eight experienced choristers, and found it to be very personal in nature with respect to amplitude and rate. Amplitude of the flutter varied between 5 and 30 cents. Moreover, it was smaller for the vowel "u" than for the vowel "a."

Weber (1992) examined differences in SPL between straight tones and tones with vibrato produced by soprano choristers. He found that vibrato tones were significantly louder than straight tones at high pitches and at high pitches combined with loud dynamic levels. Otherwise there were no significant differences in SPL between straight and vibrato tones. This finding provides another reason for sopranos' use of less vibrato when attempting to blend with other singers than when singing solo.

Vocal pedagogy in choirs. Decker (1976) reviewed theories of vocal pedagogy presented in the literature published from 1960 to 1970 and interviewed and observed ten choral authorities. He found general agreement in the literature and among the choral directors interviewed regarding vocal pedagogy for the choral rehearsal, although specific

strategies varied. Decker concluded that basic elements of posture, breathing, resonance, relaxation and diction were teachable in a choral rehearsal.

Corbin (1982) found that training in vocal technique during 21 choral rehearsal sessions over a 7-week period resulted in higher scores in diction, precision and tone quality of a high school choir over a similar control group choir. Vocal training focussed on concepts of posture, breathing, diction, resonance, and relaxation.

Positioning of choristers

Many opinions have been expressed on optimum positioning of choristers. Few empirical studies, however, have been conducted to test these opinions. These have dealt with positioning of choristers within a choir and positioning of the choir within a hall according to room acoustics.

Seating arrangement. Lambson (1961) investigated four choral seating arrangements: (1) sectional formation, (2) mixed quartet formation, (3) acoustical formation, and (4) random formation without regard for voice classification or acoustical compatability. He found that for a homophonic piece, 5 of the 10 adjudicators preferred the mixed quartet formation in both the live and recorded performances. For a polyphonic piece, results were less conclusive: sectional formation was slightly preferred in live performance, while mixed quartet formation was slightly preferred in recorded performance. The significance of these differences could not be assessed, as no statistical procedures were applied to the data.

Tocheff (1990) assessed the effects of (a) acoustic versus random positioning of choristers and of (b) sectional versus mixed formation on various elements of choral

sound, as perceived by five judges. He found significantly higher ratings for "overall blend," "intonation," and "rhythmic precision/ensemble/diction" with acoustical placement of choristers over random placement. Although many choristers indicated a preference for mixed formation, sectional formation was found to be superior to mixed formation for "overall blend" and "noticeable individual voices," regardless of the texture of the music (polyphonic versus homophonic).

Daugherty (1996) conducted a pilot study to test the effects of positioning of choristers (acoustic sectional, random sectional, and random mixed) and spacing of choristers (shoulder-to-shoulder, 4" spacing, and 12" spacing) on perceived choral sound in a live performance. Singers were high school students. Auditors were 54 high school students, 16 college level music instructors and 2 parents of one of the singers. Results indicated that variation of the positioning of choristers and the spacing between them produced differences in the choral sound that could be perceived by auditors. Results were inconclusive, however, due to difficulties in the live performance format resulting from singer/auditor fatigue, human error (premature removal of blindfolds by auditors) and distraction of auditors by noise and length of transitions between performances.

The effects of sectional block and mixed choral seating arrangements on individual musical growth and sociological dynamics were studied by Keyne (1992). She found that singing in mixed formation promoted perceived scope of a choral work and group trust. She concluded that mixed formation enhanced individual musical growth and responsibility.

Room acoustics. Husson (1962) studied the effects of varying room acoustics on phonation in speakers and singers. He observed a direct relationship between

reverberation time of the hall and subjective physiological conditions of phonation in solo singing. The greater the reverberation time of the hall, the greater was the singer's impression of ease of phonation. Reverberation time is the time taken for a sound once emitted to diminish to one millionth of its original intensity, i.e. to lose 60 dB. Halls with reverberation times of less than two seconds produced some degree of discomfort at the level of the larynx in solo singers. Optimum reverberation times for music listening were found to be much smaller than for comfortable phonation. Husson suggested that this apparent conflict of interests between the listener and the singer might be resolved by the use of reflecting surfaces in the immediate vicinity of the performer (around the stage). This would have the effect of increasing reverberation around the performer without affecting reverberation around the listener. This suggestion was supported by later research done by Meyer and Marshall.

In a study in which room acoustics were synthesized, Meyer and Marshall (1985) asked choristers to rate the difficulty of singing in various reverberation fields. Previous research with instrumental ensembles had found the effect of early reflections to be of paramount importance (Marshall, 1978). Early reflections are generated by sound-reflecting surfaces in the immediate vicinity of the sound source, and arrive within 100 msec. Choristers preferred strong early reflections in the time range from 15 to 35 ms, which would correspond to distances of 2.5 - 6 meters from the reflecting walls. If early reflections arrived later than 40 msec. (i.e. if the nearest sound-reflecting surface was more than 7 meters away), then loudness of reverberation, rather than early reflections, had the greater effect on singers' impression of ease of phonation. Reverberation times from one to three seconds had no significant effect on apparent ease of phonation.

Ternström and Sundberg (1986) explained that the difference in loudness between feedback (sound of chorister's own voice) and reference (sound from other choristers) depended on two room acoustic factors: reverberation of the room, and spacing between the choristers. As the total energy of the choir's sound would be louder in a more reverberant room, it would be more difficult for individual choristers to hear themselves than if they were singing in a less reverberant room. On the other hand, the further apart the singers were placed, the greater would be the advantage of feedback over reference. Estimates were made of the SPL advantage of feedback over reference for various spacings of 40 choristers, all singing equally loudly, in halls with various reverberation times. As reverberation time increased from 0.6 seconds (an average conference room) through 2 seconds (a large full church) to 4 seconds (a large empty church), the spacing of the choristers became less of a factor. Estimates of SPL advantage of feedback over reference in dB for various spacings of choristers are given in Table 1:

Table 1

Approximate SPL advantage of own voice (feedback) over rest of choir (reference) for various spacings of choristers in halls of different sizes

distance between choristers	conference room (reverb. time = 0.6 s)	large full church (reverb. time = 2 s)	large empty church (reverb. time = 4 s)	large cathedral (reverb. time = 9 s)
150 cm	+7.5 dB	+3.5dB	0 dB	-3.6 dB
100 cm	+5.5 dB	+2.0 dB	-0.5 dB	-3.8 dB
70 cm	+3.0 dB	+1.0 dB	-1.0 dB	-4.0 dB
50 cm	+0.5 dB	-1.0 dB	-2.5 dB	-4.6 dB

Ternström (1991) observed that reverberation enhanced choral blend, "by averaging over directions so that individual singers are not perceived by location" (p.142). It was noted that by standing with their backs against a hard surface, choristers were able to increase their acoustic power by up to 6 dB at low frequencies with no extra effort. It was also found that a choir in a relatively non-reverberant room could emulate to some extent the nature of diffuse reverberation by spreading out. The choir would thus become a diffuse source. Ideally listeners should be equidistant from all singers for maximum effectiveness of this technique.

Musical performance evaluation

One issue of interest in this study is the comparison of three groups of musically sophisticated listeners with regard to their evaluation of choral performance and their level of interjudge agreement.

Reliability. Campbell (1971) applied computer simulation techniques, which had been previously used in the grading of essays (Page and Paulus, 1968), to the evaluation of solo vocal performance. It was found that the simulation produced ratings which correlated with the average ratings of a panel of human judges in the range of .47 to .63. These correlations were similar to correlations of ratings by individual human judges with average panel ratings. It was concluded that objective variables could be identified which would predict subjective responses of a group of musically sophisticated listeners to musical performance.

Vasil (1973) studied the effect of performance length and medium (live versus recorded) on musical performance evaluation. He observed that the rankings of the top 10 to 15 performers in a sample were not significantly affected by the length of performance (30 seconds, one minute, or two minutes), nor by whether the performances were live or recorded. He concluded that rankings were determined within the first 15 seconds of a performance.

Two research studies, one in trumpet adjudication (Fiske, 1975) and the other in piano adjudication (Roberts, 1975) found that instrumental background of the adjudicator had no bearing on judge reliability in the rating of "overall score." Furthermore, "overall score" was significantly correlated with each of the other criteria being rated. It was concluded that judges found it difficult to rate criteria independently, and that perhaps only the rating for "overall score" was meaningful.

Fiske (1977a) found that the average score from a panel of judges was much more consistent than any individual judge, and that even experienced adjudicators rarely demonstrated more than 25% (r = 0.5) consistency in test-retest experiments. He observed that averaging the ratings given by a panel of seven judges, each evaluating independently of the others, could achieve test-retest correlations of approximately 0.8 and higher. Fiske (1977b) also discovered that there was no significant relationship between performing ability and adjudication reliability in a group of 33 recent music education graduates. He found, moreover, a significant inverse relationship between judges' adjudication reliability and their nonperformance (theory/history) grades. He concluded that the ability to evaluate reliably may develop independently of other musical skills.

Results from a study by Miller and Schutte (1983) suggest that the ear of a trained listener and the physical sensations of a singer may be more sensitive than spectral analysis to certain elements of vocal production. Miller and Schutte performed spectral analyses on five different registration timbres sung at the same pitch (F4) by a trained tenor: chest voice, mixed voice, feigned voice, legitimate head voice, and falsetto. Contrary to differences in physical and aural perceptions by the singer and the researchers, only minimal differences were observed in the spectra analyzed. It was concluded that it was possible to maintain similar spectral balance with respect to vowel definition and singer's formant across several vocal registrations.

Wapnick and Ekholm (1997) constructed a rating scale for the evaluation of solo vocal performance, based on 12 criteria derived from interviews with expert voice teachers. The common lore that voice teachers often disagree with each other was supported to a certain extent: interjudge reliability (average $r = 0.49$) was much lower than intrajudge reliability (average $r = 0.70$). Nevertheless, evaluations drawn from averaging ratings of four or more judges were found to achieve considerable interjudge (interpanel) reliability ($r > 0.80$). The most reliable criteria were found to be "intonation accuracy" and "overall score," and the least reliable criterion was "diction." Factor analysis on the criteria yielded three factors: "intrinsic vocal quality," "execution," and "diction." Factor analysis also revealed that experts tended to fall into two groups: one group was primarily influenced by "intrinsic vocal quality," and the other by "execution." Interjudge reliability in this study was similar to that reported by Heller and Campbell (1971, p.7), who found that the average correlation between any two judges on a panel of seven experts rating solo vocal performance was .40.

Choral performance evaluation. Cooksey (1977) used a facet-factorial approach to construct a rating scale for evaluation of high school choral performance. Descriptions of high school choral performances were collected; they were transformed into items to be rated; and then the ratings were submitted to factor analysis. Seven factors of choral performance were obtained: diction, precision, dynamics, tone control, tempo, balance/blend, and interpretation/musical effect. Testing of the rating scale produced high interjudge reliability and high criterion-related validity.

Larkin (1985) constructed 5-point rating scales for each of the following dimensions of choral music performance: tone quality, intonation, rhythmic precision, expression, and balance/blend. These rating scales were then tested with music of Renaissance, Baroque, Classic, and early Twentieth Century styles. Moderate to moderately high reliability coefficients were obtained, with "intonation" being the most reliable criterion and "expression" the least reliable. Intercorrelations among the criteria were moderately high, but were generally lower than interjudge correlations. It was concluded that it was possible to construct reliable rating scales to assist conductors in the measurement of achievement in choral music performance.

Robinson (1988) examined the use of an adjudication form versus a continuous response digital interface device (CRDI) in the evaluation of choral performance. Judges fell into three categories: nonmusicians, undergraduate musicians, and music educators. Data from CRDI ratings indicated that nonmusicians were significantly more positive than the other groups, and that undergraduate musicians and music educators exhibited near agreement across performances. In each three-piece program, operant ratings of the

second piece were higher than those of the first and third pieces. Judges who used the CRDI prior to using the adjudication form were more positive in their operant ratings than judges who used the adjudication form first. Nonmusicians used significantly more words and fewer musical terms in their CRDI written responses than did the other two groups of judges. Written responses on the adjudication forms showed no significant differences among the three groups.

Stutheit (1994) studied preparation and evaluation of contemporary choral music festival performances. It was found that intonation and tone quality were the most frequently selected criteria for choral evaluation by adjudicators and choral directors.

Studies mentioned in the review of research literature dealt with choral blend, other aspects of choral sound, vocal production of choristers, positioning of choristers, choral performance evaluation, and reliability in musical performance evaluation. None of them, however, dealt specifically with the effect of individual vocal production on perceived choral blend and overall choral sound, nor did any of them examine the effect of seating arrangement on vocal production. Moreover, none of the previous studies compared choral conductors, voice teachers and other musically sophisticated listeners in their evaluation of choral sound. It was hoped that findings from the present study might verify, and provide insight into, the controversy between voice teachers and choral conductors, and point the way toward possible solutions.

Design of the study

This experiment had a two-way factorial repeated measures design. The two factors were: singing mode and seating (standing) arrangement. Each factor had two levels. The levels of the singing mode factor were: (1) vocal production used in solo singing, as taught by voice teachers (*soloistic*) and (2) vocal production used by singers trying to maximize choral blend or homogeneity of tone quality (*blended*). The levels of the seating arrangement factor were: (1) random within SATB sections (*random*) and (2) positioned within sections according to best acoustic matching of individual voices (*acoustic*). Crossing the two factors resulted in the four experimental conditions shown in Table 2.

Table 2

Four experimental conditions produced by crossing of two factors - singing mode and seating arrangement

	FACTOR A - SINGING MODE	
	Level 1 - *Soloistic*	**Level 2 - *Blended***
FACTOR B - SEATING ARRANGEMENT Level 1 - *Random*	**Condition 1** - soloistic vocal production with random seating arrangement (within SATB sections)	**Condition 3** - blended vocal production with random seating arrangement (within SATB sections)
Level 2 - *Acoustic*	**Condition 2** - soloistic vocal production with acoustically matched seating arrangement (within SATB sections)	**Condition 4** - blended vocal production with acoustically matched seating arrangement (within SATB sections)

Preparation of recordings

An ad hoc mixed chamber choir consisting of 20 university voice majors was professionally audiotaped. The choir was made up of six sopranos, five altos, four tenors and five basses. It was thought that using an already established choir might introduce some bias into the study, as a group of singers who were accustomed to singing together might have difficulty departing from their accustomed vocal production and level of choral blend. This difficulty might make it impossible for the choristers to follow instructions regarding singing mode. Furthermore, it was important that the singers chosen for the study all be trained soloists as well as experienced choristers.

A potential problem was conductor bias. The choral conductor normally interacts with the choir during performance, each influencing the other. However, for purposes of this study, it was important that the conductor and his/her preferences in choral sound not be a variable. In order to maximize conductor consistency across all four experimental conditions, a videotape was made of a professional choral conductor conducting the four pieces to be used in the study. Care was taken to establish comfortable tempi and to determine the location of each section, so that cues would be accurate. Unfortunately, the choristers were unable to follow the videotape during the recording session. This may have been due to lack of experience singing with a videotaped conductor, lack of experience singing together as a group, and insufficient familiarity with the pieces. Consequently, the choir sang without any conductor. Instead, one of the choristers cued the beginnings and endings of the pieces.

Another problem with using a videotaped conductor was the noise level of the playback equipment. Because of the necessity of having good lighting for the singers to

be able to read their music, a VGA Data Projector was used. This projector produces a high-intensity light output which allows it to playback videotapes under normal lighting conditions. Even after surrounding the video projector with sound-absorbing baffles, the noise level was still high for recording purposes.

Four short choral pieces were performed *a cappella*: Victoria's *O magnum mysterium* (mm. 1 - 39, Schirmer edition, edited by Alice Parker and Robert Shaw), Mozart's *Ave verum corpus* (a cappella version based on L'Ensemble Vocal Philippe Caillard edition), Bruckner's *Locus iste* (Peters edition), and *Un Cygne* by Hindemith (B. Schott's Söhne edition). I chose these pieces for the following reasons: 1) they were representative of four major stylistic periods in standard choral literature - Renaissance, Classical, Romantic and Twentieth Century; 2) homogeneity of tone, or "blend," was an important concern in their performance; 3) their level of difficulty was judged to be well within the capabilities of university music majors.

Treatment

A Latin-square design was used to determine the order of the four experimental conditions for each piece. This was done in order to control for order effects on both singers and judges. The four conditions were: 1) soloistic singing mode, random seating arrangement; 2) blended singing mode, random seating arrangement; 3) soloistic singing mode, acoustic seating arrangement; and 4) blended singing mode, acoustic seating arrangement. Table 3 shows the order of the experimental conditions for the recording.

Table 3

Order of experimental conditions used in recording of choral performances

Mozart	**Bruckner**	**Hindemith**	**Victoria**
1 soloistic/random	2 blended/random	3 soloistic/acoustic	4 blended/acoustic
2 blended/random	1 soloistic/random	4 blended/acoustic	3 soloistic/acoustic
3 soloistic/acoustic	4 blended/acoustic	1 soloistic/random	2 blended/random
4 blended/acoustic	3 soloistic/acoustic	2 blended/random	1 soloistic/random

The order of the pieces was chosen for both vocal and musical considerations. The Mozart was chosen as the best one to warm up the voices of the singers. Mozart is often described by singing teachers as "medicine for the voice." The Bruckner, being the most dramatic, was placed second so that voices would be warmed up but not fatigued. The Hindemith was considered to be the most problematic musically. I thus decided to record this one before too much mental fatigue had set in. I judged the Victoria to be relatively easy vocally and musically, and therefore an appropriate one with which to end.

During the recording session, choristers were asked to rank the four experimental conditions for each piece on the basis of vocal comfort and aesthetic preference for overall choral sound (see Appendix B, p.133). While recording the entire ensemble, eight choristers, chosen at random, were also recorded individually - two sopranos, two altos, two tenors and two basses. For purposes of comparing vocal production in solo and choral settings, these eight choristers were further recorded singing their parts in each piece as though they were solo songs, with piano playing the other parts.

The recordings of the eight individual singers totalled 160 excerpts - 8 singers X 4 pieces X 5 experimental conditions (4 choral plus solo). These were rerecorded onto four audiotapes of 40 excerpts. Each excerpt lasted approximately one minute. Excerpts were presented in blocks consisting of one piece performed by one singer under each of the five experimental conditions. Each of the four tapes thus contained eight blocks. Order of the blocks was random, with the exception that each of the eight singers appeared once on each tape. A panel of expert voice teachers were asked to rank the five performances in each block on the basis of how well the singer was using his/her voice.

The ensemble recordings were evaluated for (1) overall tone quality, (2) blend/homogeneity, (3) intonation accuracy, (4) diction and (5) rhythmic precision by three experienced choral conductors and one voice teacher. The four judges assigned ratings on a scale from 1 (poor) to 7 (excellent) for each criterion.

The recording session took place in a small concert hall with a seating capacity of 100. The choir was recorded using two AKG 414 cardioid condenser microphones, placed in the middle of the hall at a height of about 10 feet above the floor. The eight individual singers within the choir were simultaneously recorded using eight microphones placed on stands, one in front of each singer. These microphones were: four Neumann U-87 condenser mics, two AKG 460 condenser mics and two Sennheiser 441 dynamic mics. The dynamic mics were less sensitive than the condenser mics, and thus were used to record the two sopranos, whose voices projected well. Wind screens were used on all the individual singers' microphones. The mixer was a Mackie 1604 console. The eight individual singers within the choral setting were recorded using a TASCAM DA-38 eight-

track recorder. The choir and solo samples of the eight individual singers were recorded using a Panasonic SV3700 2-track DAT recorder. Duplication of audiotapes for judges was done on a KABA Realtime Professional Duplicating System, using TDK chrome high bias audiotape cassettes.

Room effects were minimized at source by setting the microphones on cardioid pattern, thus increasing the direct-to-reverberant ratio. In other words, sensitivity to sound coming directly from in front of the microphones was heightened, while sensitivity to sound reflected from the walls, ceiling and floor to the sides and behind the microphones was attenuated. A small amount of reverb was applied during subsequent mixing to make the excerpts sound realistic.

Results

Expert ranking of individual vocal production. Judge panel reliability was determined for the panels of expert voice teachers judging each of the four audiotapes of individual singers in the choir. This was calculated using an equation analogous to the Spearman-Brown "prophecy" formula and Rozeboom's "alpha":

$$alpha = \frac{n \times r}{1+(n-1) \times r}$$

where n is the number of judges in the panel and r is the average of the Spearman correlations for each pair of judges. Audiotape number 1 was evaluated by a panel of four voice teachers, whose judge panel reliability coefficient was .87. Two voice teachers evaluated audiotape number 2, and their judge panel reliability coefficient was .83. The four voice teachers who evaluated audiotape number 3 had a judge panel reliability

coefficient of .82. Finally, the three voice teachers who judged audiotape number 4 achieved a judge panel reliability coefficient of .64.

A Friedman rank test was done to compare the experts' rankings of the performances of the eight individual singers in the three singing modes: soloistic singing in the choir, blended singing in the choir, and solo singing with piano accompaniment. Rankings were from 1 to 5, with the lowest number indicating the highest ranking. Significant differences were found among the three singing modes ($z_{\text{rank sum}} = 57.30$, $p<.0001$, $df = 2$). Wilcoxon signed-rank tests were done to compare each pair of singing modes. Results indicated that solo performances (mean ranking = 1.5) were ranked significantly higher than performances in the soloistic choral singing mode (mean ranking = 2.8, $z_{\text{signed-rank}} = -4.82$, $p<.001$) or in the blended choral singing mode (mean ranking = 3.7, $z_{\text{signed-rank}} = -4.94$, $p<.001$). Of the two choral singing modes, performances in the soloistic mode were ranked higher than performances in the blended mode ($z_{\text{signed-rank}} = 4.67$, $p<.001$). In other words, voice teachers judged that singers used their voices significantly better in the solo condition than in either of the choral conditions, and that they used their voices better in the soloistic choral condition than in the blended choral condition.

A Wilcoxon signed-rank test, performed on the rankings of the performances of the eight individual singers, was done to compare the two seating arrangements: random seating within SATB sections, and acoustic seating within SATB sections. No significant differences were found in mean rankings between the two seating arrangements.

Expert ratings of choral performances. Choral performances were rated by a panel of four experts, including three choral conductors and one voice teacher. The judge panel reliability coefficient, calculated in the same way as for the voice teachers, was .63.

A three-way factorial repeated measures analysis of variance was done on the expert ratings of the choral performances. The three factors were musical piece, singing mode, and seating arrangement. A significant three-way interaction effect was discovered for musical piece by singing mode by seating arrangement ($F = 5.59$, $p<.02$, $df = 3, 9$). Fisher's Least Significant Difference post hoc tests revealed that in the Mozart piece choral performances in the blended singing mode received significantly higher ratings (mean = 5.1) than performances in the soloistic singing mode (mean = 4.3, $p<.05$), and that performances in the random seating arrangement (mean = 4.9) were preferred over performances in the acoustic seating arrangement (mean = 4.5, $p<.01$). In the Bruckner and Victoria pieces, there were no significant differences for singing mode or seating arrangement. In the Hindemith piece, performances in the random seating arrangement received significantly higher ratings (mean = 4.4) than performances in the acoustic seating arrangement (mean = 3.9, $p<.01$), while singing mode had no significant effect on ratings.

Further analyses revealed that choral performances in the blended singing mode received significantly higher ratings for blend/homogeneity (mean = 4.4) and rhythmic precision (mean = 4.5) than did performances in the soloistic singing mode (means = 4.0 and 4.3, respectively, $p<.001$ and $p<.05$, respectively). On the other hand, performances in the blended singing mode received significantly lower ratings for intonation accuracy

than performances in the soloistic singing mode (means = 3.9 and 4.2, respectively, $p<.01$). Singing mode had no significant effect on ratings of overall tone quality, or diction. Choral performances in the random seating arrangement received significantly higher ratings for blend/homogeneity (mean = 4.3) and intonation accuracy (mean = 4.2) than did performances in the acoustic seating arrangement condition (means = 4.1 and 3.9, respectively, $p<.02$ and $p<.001$, respectively). Seating arrangement had no significant effect on ratings of rhythmic precision, diction and overall tone quality.

It would be premature to draw any conclusions from these results, due to the small number of judges as well as some problems with the choral recordings. Specifically, one of the tenors, who had recently switched from singing as a baritone, had trouble controlling his upper range. He produced a very loud and pushed quality of tone that was incapable of blending. As a result, this tenor tended to blend more in the soloistic singing mode, as the relatively high overall volume covered him. Also, one of the sopranos seemed incapable of controlling her vibrato, even in the blended singing mode. These anomalies probably made it more difficult for judges to distinguish between the two singing modes. This may explain the lack of a significant difference between the two singing modes in ratings of overall tone quality. In addition, the arrangement of microphones was such that the voices of the uncontrolled tenor and soprano tended to be picked up more in the acoustic seating arrangement than in the random arrangement. This may explain the tendency for judges to give higher ratings to choral performances in the random seating arrangement than in the acoustic seating arrangement.

Choristers' rankings of the four experimental conditions. Statistical analysis was also done on the choristers' rankings of the four experimental conditions for each piece on the basis of vocal comfort and aesthetic preference. A Wilcoxon signed ranks test, conducted on the vocal comfort rankings across all four pieces, found a significant difference between the two singing modes ($z_{signed\text{-}rank}$ = 2.34, $p<.02$) in favor of the soloistic mode over the blended mode. A significant difference between the two seating arrangements was also discovered ($z_{signed\text{-}rank}$ = -2.76, $p<.006$) in favor of the acoustic seating arrangement over the random arrangement. Choristers found that they were more comfortable vocally in a choral setting when singing soloistically (mean ranking = 2.2) than when asked to blend (mean ranking = 2.8). Choristers also indicated that they were more comfortable vocally in an acoustic arrangement (mean ranking = 2.2) than in a random arrangement (mean ranking = 2.8).

A Wilcoxon signed ranks test performed on the choristers' aesthetic preference rankings found a significant difference between the two seating arrangements ($z_{signed\text{-}rank}$ = -2.76, $p<.006$) in favor of the acoustic seating arrangement over the random arrangement. Choristers showed an aesthetic preference for the acoustic positioning of the choristers (mean ranking = 2.2) over the random arrangement (mean ranking = 2.8) No significant difference was found between the two singing modes in choristers' aesthetic preference rankings (soloistic mean = 2.7; blended mean = 2.3).

Conclusions

The results of the pilot study shed some light on the following questions:

(1) Do trained singers modify their normal solo vocal production in a choral setting? If so, how would voice teachers evaluate their modified vocal production?

Expert voice teachers were able to discriminate between the solo, soloistic and blended singing modes to a significant degree when evaluating vocal production of individual choristers. Results suggest that singers modify their vocal production in a choral setting not only when asked to maximize choral blend, but even when asked to use their normal solo vocal production. Voice teachers judged both modes of choral vocal production to be significantly inferior to solo vocal production. Moreover, they rated soloistic choral singing as significantly better than blended choral singing.

(5) What are the preferences of choristers with regards to singing under the various conditions in this study?

The choristers in the pilot study indicated that they were more comfortable vocally when singing in a soloistic mode than in a blended mode. Aesthetically, however, they showed no significant singing mode preference. Choristers, furthermore, expressed a significant preference for the acoustic seating arrangement over the random seating arrangement for both vocal comfort and aesthetic quality.

Changes to main study as a result of pilot study

A research grant from McGill University for the main study made it possible to pay honoraria to the singers, conductor and recording engineer. This allowed the recording to be spread over two sessions. In the pilot study, participation in the recording session was voluntary. Those who were kind enough to participate were, nonetheless, unable to devote as much time to learning the music, rehearsing and recording as might have been possible on a professional basis. Rehearsing and recording for the pilot study had to be completed in one marathon session on a Saturday. Besides the obvious fatigue factor, singers complained that the necessity of alternating between the soloistic and blended singing modes in the pilot study recording session was problematic. They found that blended singing created a greater degree of tension in the vocal apparatus than soloistic singing. After singing in a blended mode for a while, returning to a level of relaxation needed for soloistic singing appeared to be difficult for singers.

For the main study, the recordings were made on two consecutive evenings. It was decided to record all the soloistic singing mode conditions on the first evening and all the blended singing mode conditions on the second evening. The solo recordings for the eight individual choristers were made before the choral recording sessions on both evenings, four on the first evening, and four on the second. As ranking of experimental conditions across two evenings did not seem to be very reliable, choristers were asked to rate each experimental condition for each piece on a 5-point scale on the basis of vocal comfort and choral sound (see Appendix I, p. 143). Furthermore, the term "choral sound" was chosen to replace "aesthetic preference" for greater clarity.

Another problem with the pilot study was the difficulty of isolating the voices of the individual singers from the rest of the choir in the choral conditions and from the piano accompaniment in the solo condition. Not only might this have interfered with the judges' ability to hear the vocal production of the individual singer, but it also could have provided clues to the experimental condition, which may have biased the judges in their evaluations. Ideally, the judges should not have any basis for making discriminations among experimental conditions other than the sound of the individual singer's voice. For the main study it was decided to use headset mics for maximum isolation of individual singers' voices from background sound.

The lack of a conductor in the pilot study took its toll on the ensemble recordings. Entries and cutoffs were ragged, singing was expressionless and mechanical, dynamic contrast was lacking, and the tempo tended to drag because each singer appeared to be waiting for someone else to take the lead. The resulting tape had so many musical problems that it may have been difficult for judges to focus on choral sound quality. For the main study it was decided to use an experienced conductor, who had such a high degree of gestural control that he would be able to conduct the same way over all four experimental conditions of a given piece. The conductor was videotaped in order to have independent judges determine whether there had been any systematic differences in the conductor's gestures, expression or posture which might have influenced outcomes.

Finally, it was found in the pilot study that the French pronunciation in the Hindemith piece was a problem for some singers. It was therefore decided to change the Twentieth Century piece to a Latin piece in the main study. *O sacrum convivium* by

Messiaen (mm. 17 - end, Durand edition) was selected, as it seemed to meet the necessary criteria: it was representative of Twentieth Century choral literature, homogeneity of tone was an important concern in its performance, and its level of difficulty was judged to be appropriate for university music majors.

Design of the study

The design of the pilot study was retained for the main study. There were thus four choral experimental conditions resulting from the crossing of choral singing mode with seating arrangement: 1) soloistic singing mode, random seating arrangement, 2) blended singing mode, random seating arrangement; 3) soloistic singing mode, acoustic seating arrangement; and 4) blended singing mode, acoustic seating arrangement.

Preparation of recordings

Recordings for the main study were made on two consecutive evenings in order to minimize the effects of vocal fatigue on the choristers. In the pilot study, singers had complained of difficulty in alternating between the two choral singing modes. It was therefore decided to record all the soloistic singing mode conditions on the first evening and all the blended singing mode conditions on the second evening. The solo recordings for the eight individual choristers were made before the choral recording sessions on both evenings - four on the first evening, and four on the second.

Singers. An experimental mixed chamber choir, consisting of 22 university voice majors, was professionally audiotaped under the four experimental conditions. The choir was made up of five sopranos, six altos, six tenors and five basses. They were assembled especially for this recording, and otherwise did not sing together. Originally

it was planned to have six singers in each section. However, one bass and one soprano became unavailable due to unforeseen circumstances just before the recording sessions. There was not enough time to replace them.

The reasoning behind using an ad hoc choir rather than an established choir was the same as in the pilot study. A group of singers who were accustomed to singing together might have difficulty departing from their accustomed vocal production and level of choral blend. This difficulty might make it impossible for the choristers to follow instructions regarding singing mode. Furthermore, it was important that the singers chosen for the study all be trained soloists as well as experienced choristers.

Conductor. As in the pilot study, conductor bias posed a potential problem. For purposes of the study, it was important that the conductor and his/her preferences in choral sound not be a variable. However, because of the problems caused by the absence of a conductor in the pilot study, it was decided to use an experienced conductor who would be able to conduct in the same manner over all four experimental conditions of a given piece. His performances were videotaped so that independent judges could determine whether there were any systematic differences in his gestures, expression or posture which could have influenced results.

Evaluation of conductor consistency. The first test of conductor consistency was whether tempi had been kept constant across all four experimental conditions of a given piece. The tempi chosen were as follows: Mozart, half note = 40; Bruckner, half

note = 48; Messiaen, eighth note = 76; Victoria, half note = 50. After checking the tempi of the excerpts on the recording using a metronome, it was found that the tempi were remarkably consistent for any given piece across all experimental conditions. In the Mozart excerpt, the tempo did not fluctuate beyond half note = 38-40. Similarly, in the Messiaen excerpt the tempo remained steady around eighth note = 76 across all experimental conditions. In the remaining two excerpts, however, the tempo fluctuations within the excerpts were greater. What was interesting was that these fluctuations were very similar within a given piece across experimental conditions, i.e. the tempo fluctuations did not seem to be affected by the experimental variables. The greatest tempo fluctuation occurred in the Victoria. Under all four experimental conditions, the Victoria excerpt would begin at half note = 50, and as each voice entered, would slow down until a tempo of half note = 42 was reached. In the Bruckner, the excerpts would all begin at half note = 46-48, slow down at the piano section beginning in measure 21, and settle down into a final tempo of half note = 42 at the reprise in measure 30. Since these tempo fluctuations were consistent across experimental conditions, they did not constitute an extraneous variable in the study.

In order to evaluate other aspects of conductor behavior for consistency across experimental conditions, four independent judges - three choral conductors and one experienced chorister - viewed the videotape of the conductor. They then completed the *Conductor Consistency Observation Form* (Appendix J). This observation form was adapted from Madsen and Yarbrough's (1980) *Music Conductor Observation Form*, with slight adjustments made to reflect the new focus of the evaluation - namely consistency,

rather than competency. The criteria chosen for the *Conductor Consistency Observation Form* were: Movement towards/away from choir (Approaching choir, Departing from choir, or Stationary), Conducting gestures (Strict or Expressive), Eye contact (conductor looking at Group or Other), and Facial expression (Approving, Disapproving, or Neutral), as well as two other criteria, suggested by choral conductors who viewed the videotape, Magnitude of gestures (High or Low) and Body movement other than arms and hands (Much, Some or None).

Interjudge reliability in evaluation of conductor consistency from videotape. A Latin-square design was used to determine the order of experimental conditions presented on the choral audiotape. The videotape of the conductor was edited to follow the same order. The design was such that by selecting the first performance of each piece on the videotape, each of the four experimental conditions could be tested for interjudge reliability. The percentage of agreements out of the total number of ratings was calculated as a measure of interjudge reliability. Average interjudge agreement was found to be 88 % (the equivalent of r = .94), suggesting that this method of measuring conductor consistency was reliable enough for purposes of this study (see Table 4).

Table 4

Interjudge reliability in evaluation of conductor consistency from videotape as measured by percentage of agreements out of total ratings

Pair of judges	Mozart (soloistic /random)	Bruckner (blended /random)	Messiaen (soloistic /acoustic)	Victoria (blended /acoustic)
1 & 2	83	94	86	83
1 & 3	86	86	86	80
1 & 4	94	100	94	97
2 & 3	86	81	83	93
2 & 4	83	94	81	90
3 & 4	86	86	86	83
Mean % of agreements	**86**	**90**	**86**	**88**

Grand mean = 88%

A Friedman rank test was performed on the judges' ratings of the magnitude of gestures criterion. As shown in Table 5, no significant differences were found across the four experimental conditions (X^2_{rank} = 3.98, p<.264, df = 3). A Friedman rank test was also done on judges' ratings of body movement other than arms and hands. Again there were no significant differences across the four experimental conditions (X^2_{rank} = 1.35, p<.717, df = 3) (see Table 6).

Table 5

Differences in magnitude of conductor's gestures among the four experimental conditions

Piece	Friedman X^2_{rank}	Probability	Degrees of freedom
All four pieces	**3.98**	**.264**	**3**
Mozart	1.88	.599	3
Bruckner	1.35	.717	3
Messiaen	5.03	.170	3
Victoria	1.73	.631	3

Table 6

Differences in conductor's body movement other than arms and hands among the four experimental conditions

Piece	Friedman X^2_{rank}	Probability	Degrees of freedom
All four pieces	**1.35**	**.717**	**3**
Mozart	0.98	.807	3
Bruckner	0.60	.896	3
Messiaen	2.03	.567	3
Victoria	5.33	.149	3

Data obtained from the judges' observations under the other four criteria for conductor consistency were all nominal in nature. Pearson chi-square tests revealed no significant differences in conductor behavior among the four experimental conditions (see Tables 7, 8 and 9). The facial expression criterion needed no statistical analysis. All but one of the 368 observations recorded under this criterion were "Neutral."

Table 7

Differences in conductor's movement towards/away from choir among the four experimental conditions

Piece	Pearson X^2	Probability	Degrees of freedom
All four pieces	**5.03**	**.54**	**6**
Mozart	5.12	.53	6
Bruckner	no variation in data		
Messiaen	no variation in data		
Victoria	no variation in data		

Table 8

Differences in conducting gestures among the four experimental conditions

Piece	Pearson X^2	Probability	Degrees of freedom
All four pieces	**1.31**	**.73**	**3**
Mozart	6.40	.09	3
Bruckner	1.67	.64	3
Messiaen	3.30	.35	3
Victoria	6.15	.10	3

Table 9

Differences in conductor's eye contact among the four experimental conditions

Piece	Pearson X^2	Probability	Degrees of freedom
All four pieces	**0.21**	**.98**	**3**
Mozart	0.37	.95	3
Bruckner	3.03	.39	3
Messiaen	5.39	.15	3
Victoria	1.96	.58	3

These analyses suggest that the conductor maintained consistent behavior across all four experimental conditions. Therefore, conductor bias was not a factor in this study.

Musical excerpts. The choice of musical excerpts as well as their order of presentation on the experimental audiotape was the same as in the pilot study, except that the Messiaen piece was substituted for the Hindemith, as already explained. The order was thus: Mozart, Bruckner, Messiaen, Victoria (see Table 10).

Table 10

Order of experimental conditions used in audiotaped presentation of choral performances

Mozart	Bruckner	Messiaen	Victoria
1 soloistic/random	2 blended/random	3 soloistic/acoustic	4 blended/acoustic
2 blended/random	1 soloistic/random	4 blended/acoustic	3 soloistic/acoustic
3 soloistic/acoustic	4 blended/acoustic	1 soloistic/random	2 blended/random
4 blended/acoustic	3 soloistic/acoustic	2 blended/random	1 soloistic/random

Recordings of individual singers. As in the pilot study, eight choristers were recorded individually while singing with the choir. They were two sopranos, two altos, two tenors and two basses, chosen at random. For purposes of comparing vocal production in solo and choral settings, each of these eight choristers was also recorded singing his/her part in each piece as though it were a solo song, with piano playing the

other parts as an accompaniment. Headset microphones were used for maximum isolation of the individual singers' voices from the background sounds.

The same format as in the pilot study was followed in the editing of the recordings of the eight individual singers. The result was thus four audiotapes, each containing eight blocks of five performances of a piece by the same singer (4 choral and solo). All eight choristers sang on each tape.

Recording equipment. The recording session took place in a small concert hall with a seating capacity of 100. The choir was recorded using four Neumann U-87 microphones. One microphone was placed approximately six feet in front of and above each section of the choir. Eight choristers were simultaneously recorded using eight headset mics: four AKG - ATM 71 condenser mics and four Shure SM-10A dynamic mics. The more sensitive AKG condenser mics were used for the altos and basses, since they tended to sing in the lower part of their ranges and therefore did not project as easily as sopranos and tenors. The less sensitive Shure dynamic mics were used for the sopranos and tenors. The same microphones were used for recording the solo samples of the eight individually monitored choristers.

The mixer was a Mackie 1604 console. The eight individually monitored singers were recorded using a TASCAM DA-38 eight-track recorder. The choir was recorded using a Panasonic SV3700 2-track DAT recorder. All recordings were digital sixteen bit, with 44.1 kHz sampling. The two recorders had identical formats aside from the track number. Duplication of audiotapes for judges was done on a KABA Realtime Professional Duplicating System, using TDK chrome high bias audiotape cassettes.

As in the pilot study, room effects were minimized at source by setting the microphones on cardioid pattern, and a small amount of reverb was applied during subsequent mixing to make the excerpts sound realistic.

Evaluations

The recordings of the eight individually monitored singers were evaluated by expert voice teachers on the basis of how well the singers were using their voices. The voice teachers were asked to rank the five performances of a piece by the same singer and to comment on the strengths and weaknesses of the vocal production in each performance. Each of the four audiotapes was evaluated by a different panel of voice teachers.

The choral audiotape consisted of four choral pieces of approximately two minutes' duration, each of which was performed four times under differing experimental conditions as described above. The 16 choral performances were evaluated for blend/homogeneity, diction, dynamic range, phrasing, pitch precision, rhythmic precision, and overall tone quality by experienced choral conductors, voice teachers and other musicians. These criteria were derived from previous research in choral evaluation (Cooksey, 1977; Larkin, 1985; Giardiniere, 1991). Judges were asked to assign a rating from 1 (poor) to 7 (excellent) for each criterion. Moreover, they were encouraged to write comments about each performance in the space provided or, for longer comments, on a separate sheet of paper. Judges were instructed to use a good pair of headphones in order to maximize the quality of sound reproduction, and to evaluate all four performances of any piece in the same session for purposes of more accurate comparison of experimental conditions.

Judges were also encouraged to listen to performances as many times as they found necessary for reliable evaluation.

During the recording sessions, choristers were asked to rate each experimental condition in each piece on a 5-point scale for vocal comfort and for choral sound preference. They also wrote comments.

Judges. Sixty-four choral conductors, recruited from universities in 18 countries in North America, Europe, Africa, Asia and Australia, agreed to participate in this study. Of these, 37, or 58%, actually completed the choral evaluation form within the required time. These 37 conductors came from 12 countries: Austrialia (2), Canada (7), Denmark (3), England (2), Germany (3), Iceland (1), Netherlands (1), Norway (2), Russia (2), South Africa (1), Spain (1) and USA (12).

Fifty voice teachers from 17 countries in North America, Europe, Asia, Africa and Australia, recruited through NATS International (National Association of Teachers of Singing) and universities around the world agreed to participate in this study. Of these, 38, or 76%, completed their tasks: 33 evaluated the choral performances, 12 evaluated the individually monitored choristers (7 voice teachers completed both evaluation tasks). The 33 voice teachers who evaluated the choral performances came from 13 countries: Australia (1), Austria (1), Canada (8), England (1), France (3), Germany (3), Ireland (1), Italy (3), Korea (1), Netherlands (2), Singapore (1), Switzerland (1) and USA (7). The 12 who evaluated the recordings of the individually monitored choristers were from Canada (4), France (1) , South Africa (1), UK (2) and USA (4).

Forty professional musicians who were neither voice teachers nor choral conductors agreed to participate in the study. Of these, 32, or 80%, completed the study. The 32 nonvocal musicians came from nine countries: Canada (7), England (1), Finland (1), Germany (4), Malaysia (2), Russia (1), South Africa (1), Switzerland (1) and USA (14).

Of the 102 judges who evaluated the choral performances, 51 were males and 51 females. In the choral conductors subgroup there were 24 males and 13 females; in the voice teachers subgroup there were 20 females and 13 males; and the nonvocal musicians subgroup were more evenly divided into 18 females and 14 males. Among the 33 voice teachers, 14 had some choral conducting experience; among the choral conductors, two had some voice teaching experience.

Results

Voice teachers' rankings of performances by individual choristers

A Friedman rank test was performed on the rankings, by the 12 voice teachers, of the performances of the eight individually monitored choristers who were recorded while singing with the choir and solo. The test indicated significant differences in performance rankings among the three singing modes ($z_{rank\ sum}$ = 22.17, p<.001, df = 2). Each pair of singing modes was then compared using the Wilcoxon signed-rank test. Results revealed that solo singing was ranked significantly higher than soloistic choral singing ($z_{signed\text{-}rank}$ = 3.06, p<.002) or blended choral singing ($z_{signed\text{-}rank}$ = 3.06, p<.002), and that soloistic choral singing was ranked higher than blended choral singing ($z_{signed\text{-}rank}$ = 2.98, p< 003)

A Wilcoxon signed-rank test discovered a significant difference in favor of the acoustic seating arrangement over the random seating arrangement ($z_{signed\text{-}rank}$ = -1 99, p<.05). In other words, vocal production of individually monitored choristers, recorded while singing in the choir in the acoustic seating arrangement, was ranked higher than vocal production in the random seating arrangement.

Voice teachers' comments on individual choristers

The written comments of the voice teachers were classified by singing mode and further grouped by subject heading. One basis of comparison among the singing modes was frequencies of positive and negative comments. Reliability of judgments regarding whether a comment was positive, negative or neutral was determined by having two

judges separately label one quarter of the comments as positive, negative or neutral. The percentage of agreements between the two judges was 96%. This is equivalent to a correlation coefficient of $r = .98$, indicating high reliability in the judgments.

Blended singing mode. In the blended singing mode, there were 266 negative comments (80%) and 65 positive comments (20%). The most common criticism of this mode of vocal production was its lack of freedom of phonation. Fifty comments dealt with this subject. Typical terms used were: "held back," "restrained," "tight," "pressed," "forced," "tense," "throaty," "pushed," "labored." The next most common criticism dealt with intonation problems. There were 47 comments on this subject, most of which mentioned either intonation problems in general or flatting specifically (17). There were 38 comments criticizing the lack of vibrato or vibrancy or the excessive straightness of the tone; 27 comments on the breathiness of the tone; 26 comments on the lack of breath control; 18 comments on the lack of focus or ring in the tone; and 11 comments describing the tone quality as "weak," "tired," "limp," "dull" or "flabby." The remainder of the criticisms (49) referred to excessive tremolo or wobble, diplophonia or hoarseness, hollowness or colorlessness of the tone, lack of legato line, unnaturalness of the tone, nasality, registration breaks, diction problems and lack of musicality.

Of the positive comments, 12 commended intonation, 7 praised top notes, 7 admired general tone quality, 6 complimented naturalness or clarity of tone, 6 praised legato line, 5 admired freedom of tone, 5 mentioned focus/ring, and 4 each commended diction and placement.

Soloistic singing mode. Of the 310 comments made by voice teachers on the vocal production of individually monitored singers in the soloistic choral mode, 200 were negative (65%) and 110 were positive (35%). The most common criticism dealt with lack of freedom of phonation. There were 42 such comments. Typical terms used to describe the lack of freedom were: "tight," "driven," "pushed," "restrained," "laboured," "pressed" and "throaty." There were almost as many complaints about intonation problems - 39 comments. Twenty-five comments referred to the vibrato as excessive or inconsistent. Of the remaining negative comments: 19 cited breathiness of tone quality; 12 mentioned lack of "resonance," "focus" or "core" in the sound; 11 complained of nasality; 10 described phonation as "weak," "limp," "lifeless" and "crooned"; 10 cited lack of breath support; 9 referred to lack of vibrancy or vibrato; and the remaining 23 complained of poor vocal production in general, instability or lack of control, diplophonia or scratchiness, lack of legato line, diction problems, and "manufactured" sound.

Of the 110 positive comments, 17 commended tone quality in general; 15 praised intonation; 15 complimented "focus," "resonance" or "core;" 11 commended the vibrancy or normal vibrato; 10 admired freedom of vocal production; 9 cited better clarity or less breathiness; 7 mentioned more legato line; 6 described the tone as "richer" or "rounder;" 6 referred to better control; and the remaining 14 praised the naturalness of the tone, breath support, integration of the registers, top notes, and clarity of diction.

Solo singing. Out of 155 comments on the individual singers' vocal production in the solo mode, 112 were positive (72%) and 43 were negative (28%). Twenty-seven of the positive comments commended tone quality in general, or simply stated that this

was the best performance. Of the remainder of the positive comments 15 praised breath control or flow phonation, 14 cited freedom of vocal production, 13 admired resonance, focus or ring, 10 commended intonation, 7 praised richness, warmth or roundness of tone, 6 admired vibrancy or vibrato, and the remaining 20 commended naturalness or clarity of tone, top notes, legato line, integrated registers, diction, and forward placement.

Of the 43 negative comments, 14 criticized intonation, 8 cited excessive vibrato, 7 mentioned lack of freedom in the vocal production, and the remaining 14 criticized breath control, false darkness or fabricated sound, tightness on top notes, straight tones, nasality, breathiness, scratchiness, unsteadiness or lack of control.

Summary of voice teachers' comments. Pearson chi-square analysis was applied to the frequencies of positive and negative comments elicited by the three singing modes Results showed that the solo singing mode evoked a significantly greater number of positive comments and fewer negative comments than the two choral singing modes (X^2 = 126.86, $p<.001$, df = 2). Moreoever, the soloistic singing mode brought out a significantly greater number of positive comments and fewer negative comments than the blended singing mode ($X^2 = 20.25$, $p<.001$, df = 1). The most common criticism of the two choral singing modes referred to lack of freedom of phonation. Fifty comments cited this problem in the blended singing mode, as compared to 41 in the soloistic singing mode, and 7 in the solo mode. The second most common problem mentioned by the voice teachers was lack of intonation accuracy. There were 47 complaints of poor intonation elicited by the blended singing mode, 39 by the soloistic singing mode, and 14 by the solo mode.

Choristers' evaluation of experimental conditions

Vocal comfort. Ratings for vocal comfort, by the 22 choristers who participated in the main study, were submitted to a four-way mixed design analysis of variance, with three repeated measures factors (singing mode, seating arrangement and musical piece) and one independent groups factor (voice classification). The analysis yielded a main effect for seating arrangement ($F = 19.94$, $p<.001$, $df = 1, 18$) in favor of the acoustic arrangement over the random arrangement.

There was also a significant three-way interaction effect for singing mode by seating arrangement by voice classification ($F = 3.68$, $p<.03$, $df = 3, 18$). Fisher's least significance difference (LSD) post hoc tests revealed that sopranos preferred the combination of soloistic singing mode with acoustic seating arrangement (mean = 4.3), followed by the soloistic/random combination (mean = 3.7), then the blended/acoustic combination (mean = 3.3), and least of all the blended/random combination (mean = 3.0). Altos gave the soloistic/random condition (mean = 3.8) significantly lower vocal comfort ratings than the other three conditions - blended/random (mean = 4.3), blended/acoustic (mean = 4.4) and soloistic/acoustic (mean = 4.5) - which were not significantly different. Tenors rated the two blended singing mode conditions - blended/random (mean = 3.7) and blended/acoustic (mean = 3.9) - significantly higher than the two soloistic singing mode conditions - soloistic/random (mean = 3.0) and soloistic/acoustic (mean = 3.0). Basses rated the two acoustic seating arrangement conditions - soloistic/acoustic (mean = 4.2) and blended/acoustic (mean = 4.1) - significantly higher than the two random seating arrangement conditions - soloistic/random (mean = 3.7) and blended/random (mean = 3.6). All differences were significant at the .05 level.

The analysis of variance also yielded significant two-way interaction effects for seating arrangement by musical piece ($F = 4.71$, $p<.005$, $df = 3, 54$) and for musical piece by voice classification ($F = 3.05$, $p<.005$, $df = 9, 54$). Fisher's LSD post hoc tests revealed that the acoustic seating arrangement was significantly preferred over the random seating arrangement for the Mozart, Bruckner and Messiaen pieces. In the Victoria piece, there was no significant difference in vocal comfort ratings between the two seating arrangements. Moreover, the tenors rated the Victoria motet significantly lower than the other three pieces for vocal comfort. The basses on the other hand rated the Messiaen piece significantly lower than the other three pieces.

Choristers' comments on vocal comfort. The choristers produced 211 written comments on vocal comfort during the recording sessions. The soloistic/acoustic experimental condition produced 20 negative comments, 33 positive comments and 3 neutral comments. The soloistic/random condition elicited 34 negative comments, 11 positive and 8 neutral. The blended/acoustic condition brought forth 18 negative comments, 31 positive and 2 neutral. Finally, the blended/random condition yielded 29 negative comments, 17 positive and 5 random. Pearson chi-square analysis revealed that the acoustic seating arrangement elicited a significantly greater number of positive comments and fewer negative comments than the random arrangement ($X^2 = 23.79$, $p<.001$, $df = 2$). No statistically significant difference in the frequencies of positive and negative comments was found between the two singing modes.

Sopranos singing in the blended mode complained of the following: difficulty singing low notes (below G_3) and high notes (above G_4); increased muscular tension in the throat; a tendency to sing flat, caused by trying to control the vibrato; difficulty controlling the voice, especially the pitch; faltering breath support, caused by trying to restrain the sound; difficulty sustaining notes on a straight tone; and a tendency to modify vowels in order to decrease vibrato, resulting in jaw tension. Positive comments on singing in the blended mode mentioned increased ability to tune difficult harmonies, greater attention to dynamics, intonation and vocal tone, which made singing more comfortable, and elimination of the feeling of having to compete with other singers to be heard. One soprano complained, "When singing soloistically we should have the freedom to sing softly, but this was not the case, as certain members insisted on belting out." Some sopranos found the blended singing mode especially comfortable for the Renaissance motet.

Basses complained that in the blended singing mode there was a tendency to make a more breathy unfocussed tone. This made projection more difficult, especially for the low notes. Basses thus felt that they were "pushing" the low notes. One bass complained that it was difficult for him to sing soloistically in a choir, because he had been trained to blend his voice with other basses.

Some altos found it more comfortable to sing in the soloistic mode than in the blended mode because it enhanced breath support. Others mentioned that they could hear the singers beside them better in the soloistic mode, and this gave them a feeling of support. Some also found it easier to sing in the higher range in the soloistic mode.

Negative comments on singing in the soloistic mode mentioned difficulty in hearing oneself, which caused intonation problems and a tendency to oversing. One alto described the singing she heard in the soloistic mode as "blastissimo." Positive comments on vocal comfort in the blended singing mode mentioned ease of hearing and blending with other singers, and less tendency to oversing in order to hear oneself. Some altos criticized singing in the blended mode because of breath support problems which hindered legato line and phrasing, and intonation problems.

Tenors made a few positive comments to the effect that the soloistic singing mode enhanced freedom of vocal production. Most of the tenors' comments, however, were complaints that the dynamic level was too loud in the soloistic mode, causing a tendency to oversing in order to hear oneself. One tenor remarked of the Victoria motet, "It is hard to sing this type of music in a soloistic manner." This comment along with the loud dynamic level suggested that some singers may have equated "soloistic" with "operatic." Another tenor admitted that he was unable "to give an accurate representation of a soloistic sound while singing in a choir." Tenors' comments generally indicated that they found singing in the blended mode more comfortable "mostly because everyone's not yelling now." Some tenors explained that in the blended mode they could use head voice for the higher notes, which they found easier than using their modal register. Others complained that blended singing was tiring and hard on the throat. As one tenor described it, "This type of singing is easy, but less satisfying; breaths don't go as far, and it gets tiring on the throat." Another tenor explained that singing more softly, as in the blended mode, made it easier to "switch" into the high register more smoothly.

Comments on seating arrangement preferences were similar across SATB sections. Positive comments on vocal comfort in the acoustic seating arrangement as compared to the random seating arrangement mentioned: less tendency to oversing because of a feeling of greater vocal projection; greater ability to hear oneself; a feeling of support from standing next to a singer with similar vocal timbre; apparent ease of singing in tune; a feeling of blending more, allowing for greater vocal freedom; and greater ability to hear the other singers in the section and in the rest of the ensemble. Singers indicated that blend and intonation were more conscious concerns in the random seating arrangement than in the acoustic seating arrangement. Several singers commented that in the random seating arrangement they could not hear the other singers around them. Consequently they felt that they had to restrain their voices in order not to be too conspicuous. Other complaints about the random seating arrangement cited problems matching pitch centres with an adjacent singer, and inability to hear oneself without oversinging.

Choral sound preference. Choristers' ratings of the four experimental conditions for each piece according to choral sound preference were analyzed using a four-way mixed design analysis of variance, with three repeated measures factors (singing mode, seating arrangement and musical piece) and one independent groups factor (voice classification). The ANOVA yielded a significant main effect for singing mode in favor of blended singing over soloistic singing ($F = 20.86$, $p<.001$, $df = 1, 18$). A significant main effect for seating arrangement was also found in favor of acoustic seating over random seating ($F = 8.43$, $p<.009$, $df = 1, 18$).

There were two-way interaction effects for singing mode by musical piece ($F = 2.81$, $p<.05$, $df = 3, 54$) and for seating arrangement by musical piece ($F = 6.92$, $p<.001$, $df = 3, 54$). Fisher's LSD post hoc tests revealed that performances of all four pieces in the blended singing mode were given significantly higher choral sound ratings than were performances in the soloistic singing mode. Although the singing mode effect was significant for every musical piece, it was greatest in the Victoria (difference in means = 1.0), followed by the Mozart (difference in means = 0.8), then the Messiaen (difference in means = 0.6), and least in the Bruckner (difference in means = 0.4). Furthermore, choral sound ratings were significantly higher for the Mozart and Messiaen pieces in the acoustic seating arrangement than in the random seating arrangement. Seating arrangement had no significant effect on choral sound ratings of the Bruckner and Victoria motets.

Choristers' comments on choral sound. The choristers wrote 287 comments on aesthetic aspects of the choral performances. The soloistic/acoustic experimental condition evoked 26 positive comments, 38 negative and no neutral comments. The soloistic/random condition aroused 22 positive comments, 50 negative and one neutral comment. The blended/acoustic condition provoked 43 positive comments, 30 negative and 6 neutral comments. Finally, the blended/random condition gave rise to 41 negative comments, 25 positive and 5 neutral comments.

Pearson chi-square analysis revealed a statistically significant difference in the frequencies of positive and negative comments between the two singing modes ($X^2 =$

25.23, $p<.001$, $df = 2$). There was a significantly greater number of positive comments (84) and fewer negative comments (55) in the blended singing mode than in the soloistic singing mode (48 positive and 88 negative). No significant difference was found between the two seating arrangements with respect to frequencies of choristers' positive and negative comments on aesthetic aspects of the choral performances.

Choristers generally commended blended singing for better blend, more appropriate tone quality for the Renaissance motet and the Mozart piece, better intonation, better clarity of line or texture, better shaping of phrases, more rhythmic precision and ensemble togetherness, better matching of vowels due to less vibrato, greater ability to hear harmonies, more inspired interpretation, and greater musicality. Singers criticized blended singing for lack of intensity, lack of line or direction in phrases, tendency to flatten pitches, especially among sopranos and in the upper range, a bland or "dead" sound, narrower dynamic range, apparent tendency for tempo to drag, breathy tone quality, lack of musicality, and lack of vibrancy or vibrato in the tone.

Choristers who expressed their appreciation for choral sound in the soloistic mode cited bigger, warmer, more beautiful and more uniform sound, better intonation, freer sound, especially in the women's top register, greater rhythmic unity, better blend, greater ability to hear individual lines, and fuller, more appropriate sound for the Bruckner piece. Choristers criticized the soloistic mode for intonation problems, especially a tendency towards flatting in the soprano and bass sections, lack of blend, especially in the soprano and bass sections, stylistically inappropriate tone quality (for Victoria, Mozart), poor ensemble, lack of dynamic variation, the sole dynamic level being too loud, lack of

rhythmic precision, lack of precision in vowel uniformity and in cut-offs, lack of refinement in the tone quality, lack of balance due to the disproportionate loudness of the soprano section, a static sound, and excessive vibrato, especially in the soprano section.

In comparing the two seating arrangements with respect to choral sound, singers praised the acoustic seating arrangement for better blend, more dynamic variation, enhanced ability to hear chords or harmony, greater ability to hear the whole ensemble, better intonation, better overall ensemble sound, more uniform vowel sounds, and greater ensemble unity. One bass remarked that changing from one seating arrangement to the other seemed to have less effect on overall sound in the blended singing mode than in the soloistic mode. A soprano commented that she found it easier to hear the basses in the acoustic seating arrangement, which helped her intonation.

Quantitative analyses of choral performance evaluations

A three-way mixed design analysis of variance, was performed on the ratings by all 102 judges of the 16 choral performances. The two repeated measures factors were singing mode and seating arrangement. The independent groups factor was judge subgroup (choral conductors, voice teachers, and nonvocal musicians). A significant interaction effect was found for singing mode by judge subgroup ($F = 6.76$, $p<.002$, $df = 2, 99$). Fisher's least significant difference (LSD) post hoc tests revealed a significant singing mode effect on choral conductors' ratings in favour of blended singing over soloistic singing ($p<.001$, means = 4.56 for blended and 4.08 for soloistic). Singing mode, however, had no significant effect on choral performance ratings given by voice

teachers (means = 4.89 for blended and 4.77 for soloistic) or by nonvocal musicians (means = 4.68 for blended and 4.57 for soloistic.)

A significant main effect for seating arrangement was observed in favor of the acoustic seating arrangement over the random seating arrangement ($F = 8.22$, $p<.005$, $df = 1, 99$, means = 4.64 and 4.55, respectively).

There was a significant main effect for judge subgroup ($F = 3.01$, $p<.05$, $df = 2, 99$). Fisher's LSD post hoc tests revealed that choral conductors gave significantly lower ratings than the other two subgroups ($p<.01$, mean = 4.32). The difference between ratings given by voice teachers (mean = 4.83) and nonvocal musicians (mean = 4.62) was not significant.

Calculations of *eta squared* from the ANOVA summary table showed that for the 102 judges, singing mode accounted for 20% of the total variance in choral performance ratings, and seating arrangement accounted for 7%. For choral conductors, singing mode accounted for 58% of the total variance in choral performance ratings, and seating arrangement accounted for 4%. Singing mode accounted for 7% of voice teachers' choral performance ratings, and seating arrangement accounted for 17%. Finally, in the nonvocal musicians subgroup, singing mode and seating arrangement accounted for 4% and 3% of the total variance, respectively.

Fourteen of the 33 voice teachers indicated that they had some choral conducting experience. They were, therefore, examined as a separate group and compared to the 19 voice teachers who had no choral conducting experience. A two-way repeated measures analysis of variance, with singing mode and seating arrangement as the two factors, was

performed on the choral performance ratings given by each of these groups. Voice teachers with choral conducting experience displayed a significant main effect for singing mode ($F = 5.49$, $p<.04$, $df = 1, 13$) in favour of blended singing over soloistic singing (means = 5.00 and 4.78, respectively). There was no significant seating arrangement effect and no significant interaction effect. Calculations of *eta squared* showed that singing mode accounted for 30% of the variance in choral performance ratings, and seating arrangement accounted for 14%.

The 19 voice teachers with no choral conducting experience showed no significant main effect for singing mode. They did, however, show a significant main effect for seating arrangement ($F = 4.34$, $p<.05$, $df = 1, 18$) in favour of acoustic seating over random seating (means = 4.88 and 4.71, respectively). *Eta squared* calculations revealed that singing mode accounted for 0.6 % of the variance in choral performance ratings, and seating arrangement accounted for 19.4 %. No significant interaction effect was observed.

Ratings of individual criteria. A three-way mixed design analysis of variance, was performed on the ratings for each of the seven evaluative criteria. The two repeated measures factors were singing mode and seating arrangement, and the independent groups factor was judge subgroup. The analyses of variance produced significant main effects for singing mode in favor of blended singing over soloistic singing for each of the seven evaluative criteria (see Table 11b). All criteria, except dynamic range, yielded significant interaction effects for singing mode by judge subgroup (see Table 11a). Significant main effects for seating arrangement in favor of acoustic

observed for the criteria blend/homogeneity ($p<.01$), dynamic range ($p<.001$) and overall tone quality ($p<.003$). A significant interaction for seating arrangement by judge subgroup was found in ratings for phrasing ($F = 3.20$, $p<.05$, $df = 2, 99$). In order to examine the differences in judging among the three subgroups of judges, two-way repeated measures analyses of variance were performed on ratings by each subgroup for each criterion. Table 11b summarizes the results.

Table 11a

Interaction effects of singing mode by judge subgroup on individual evaluative criteria in choral performance evaluations by all 102 judges

Criterion	F-value	Probability	Degrees of freedom
blend/homogeneity	8.03	.001	2, 99
diction	3.40	.04	2, 99
dynamic range	2.46	.09	2, 99
phrasing	4.15	.02	2, 99
pitch precision	4.34	.02	2, 99
rhythmic precision	5.08	.008	2, 99
overall tone quality	7.90	.001	2, 99

Choral conductors gave significantly higher ratings to blended singing than to soloistic singing for each of the seven criteria ($p<.001$). Voice teachers significantly favored blended singing only for the criterion blend/homogeneity. Singing mode had no significant effect on criterion ratings given by nonvocal musicians. The seating arrangement effect in favor of acoustic seating over random seating was found to be significant in the choral conductors subgroup only for blend/homogeneity ($p<.001$). This

effect approached significance in the voice teachers subgroup ($p<.06$) for blend/homogeneity. The seating arrangement effect was significant in the voice teachers subgroup for dynamic range, phrasing and overall tone quality. No significant seating arrangement effect on individual evaluative criteria was found in the nonvocal musicians subgroup.

Table 11b

Effect of singing mode and seating arrangement on individual evaluative criteria

(Significantly preferred singing modes and seating arrangements are given with p-level)

	blend/ homo-geneity	diction	dynamic range	phrasing	pitch precision	rhythmic precision	overall tone quality
all judges (n=102)	blended singing ($p<.001$) acoustic seating ($p<.01$)	blended singing ($p<.001$) seating NS	blended singing ($p<.001$) acoustic seating ($p<.001$)	blended singing ($p<.001$) seating NS	blended singing ($p<.05$) seating NS	blended singing ($p<.001$) seating NS	blended singing ($p<.001$) acoustic seating ($p<.003$)
choral conductors (n=37)	blended singing ($p<.001$) acoustic seating ($p<.02$)	blended singing ($p<.001$) seating NS	blended singing ($p<.001$) seating NS	blended singing ($p<.001$) seating NS	blended singing ($p<.001$) seating NS	blended singing ($p<.001$) seating NS	blended singing ($p<.001$) seating NS
voice teachers (n=33)	blended singing ($p<.01$) seating NS ($p<.06$)	singing NS seating NS	singing NS acoustic seating ($p<.02$)	singing NS acoustic seating ($p<.02$)	singing NS seating NS	singing NS seating NS	singing NS acoustic seating ($p<.004$)
other musicians (n=32)	singing NS seating NS	singing NS seating NS	singing NS seating NS	singing NS seating NS	singing NS seating NS	singing NS seating NS	singing NS seating NS

Criteria intercorrelations. Choral performance ratings from all 102 judges were analyzed using the seven evaluative criteria as variables. A Pearson correlation matrix showed that all seven criteria were significantly intercorrelated. Correlations among the pairs of criteria ranged from .57 (pitch precision with diction) to .79 (blend/homogeneity with overall tone quality), with a grand mean of .67 (see Table 12).

Table 12

Pearson correlation matrix for evaluative criteria

	Blend	Diction	Dynamic range	Phrasing	Pitch precision	Rhythmic precision	Overall tone quality
Blend	1.00						
Diction	0.60	1.00					
Dynamic range	0.67	0.63	1.00				
Phrasing	0.69	0.68	0.75	1.00			
Pitch precision	0.65	0.57	0.61	0.65	1.00		
Rhythmic precision	0.61	0.67	0.63	0.72	0.68	1.00	
Overall tone quality	0.79	0.62	0.69	0.74	0.69	0.65	1.00

Factor analysis on the seven evaluative criteria as variables showed only one factor with an eigenvalue greater than 1 (eigenvalue = 5.0). This factor explained 75 % of the total variance. All seven criteria were highly loaded onto this factor as follows:

Phrasing	.89
Tone quality	.88
Blend	.85
Dynamic range	.84
Rhythmic precision	.84
Pitch precision	.82
Diction	.80
Average of criteria ratings	1.00

Factor analysis of the choral performance ratings for each subgroup, with evaluative criteria as variables, produced in each case only one factor with an eigenvalue greater than 1, onto which all seven evaluative criteria were highly loaded (see Table 13). This factor was the average of the seven criteria ratings.

Table 13

Factor analysis of the ratings of each judge subgroup, with evaluative criteria as variables

37 choral conductors		33 voice teachers		32 other musicians	
Criteria	Component loadings	Criteria	Component loadings	Criteria	Component loadings
Phrasing	.87	Phrasing	.88	Tone quality	.91
Tone quality	.86	Tone quality	.84	Phrasing	.90
Blend	.83	Blend	.84	Rhythmic precision	.88
Dynamic range	.83	Dynamic range	.81	Pitch precision	.87
Rhythmic precision	.83	Rhythmic precision	.76	Dynamic range	.87
Pitch precision	.80	Diction	.75	Diction	.86
Diction	.78	Pitch precision	.74	Blend	.86
Average of criteria ratings	1.00	Average of criteria ratings	1.00	Average of criteria ratings	1.00
Percent of total variance explained	68.7	Percent of total variance explained	64.5	Percent of total variance explained	77.4

Effect of musical piece on choral performance ratings. A preliminary one-way repeated measures analysis of variance revealed a significant order effect on ratings of the choral performances. Ratings were significantly lower for the first performance of a given piece (mean = 4.36) and significantly higher for the third performance (mean = 4.80) than for any other performances ($F = 26.37$, $p<.001$, $df = 3, 303$). Ratings for the second and fourth performances of a piece were not significantly different (means = 4.55 and 4.62, respectively).

In order to study the effect of musical piece on choral performance ratings, it was necessary to eliminate the order effect, as for each musical piece there was only one order of presentation of the four experimental conditions. This was accomplished by averaging the mean ratings for the four presentation positions (first, second, third and fourth) and multiplying ratings in each position by a factor that would bring their mean to equal the mean for the four positions; i.e. if there were no order effect, all four position means would be equal. Thus ratings for the first performance of a piece were multiplied by 1.052; second performance ratings by 1.006; third performance ratings by .954; and fourth performance ratings by .993.

A four-way mixed design analysis of variance was performed on the choral performance ratings, adjusted to eliminate the order effect. The three repeated measures factors were musical piece, singing mode and seating arrangement, and the independent groups factor was judge subgroup. The analysis yielded significant main effects for musical piece, ($F = 6.99$, $p<.001$, $df = 3, 297$), singing mode ($F = 24.17$, $p<.001$, $df = 1, 99$), seating arrangement ($F = 7.63$, $p<.007$, $df = 1, 99$) and judge subgroup ($F = 3.011$,

$p<.05$, $df = 2, 99$). Significant interaction effects were found for seating arrangement by musical piece ($F = 3.53$, $p<.015$, $df = 3, 297$), musical piece by judge subgroup ($F = 2.91$, $p<.009$, $df = 6, 297$), and singing mode by judge subgroup ($F = 6.66$, $p<.002$, $df = 2, 99$).

Fisher's LSD post hoc tests revealed that a significant seating arrangement effect in favour of acoustic seating over random seating was found only for the Mozart ($p<.001$) and Victoria pieces ($p<.03$). According to Fisher's LSD post hoc test results, choral conductors and nonvocal musicians gave significantly lower ratings to the Victoria piece (means = 4.02 and 4.46, respectively) than to the other pieces, which were not significantly different (Mozart means = 4.42 and 4.71, respectively; Bruckner means = 4.39 and 4.65, respectively; Messiaen means = 4.46 and 4.68, respectively). Voice teachers gave significantly lower ratings to the Mozart piece (mean = 4.73) than to any of the other pieces, which did not differ significantly (Bruckner mean = 4.91, Messiaen mean = 4.86, Victoria mean = 4.83).

As observed earlier, the singing mode effect in favour of blended singing over soloistic singing was significant only in the choral conductors subgroup. Moreover, singing mode had no significant interaction with musical piece. Choral conductors preferred blended singing to soloistic singing regardless of the musical selection. Similarly, voice teachers and nonvocal musicians had no significant singing mode preference regardless of the musical selection.

Interjudge agreement within the three subgroups of judges. As a measure of interjudge agreement within each subgroup of judges, the ratings given by each judge were correlated with the mean ratings given by the judge's subgroup. The Pearson correlation coefficients for each judge provided a measure of agreement between that judge and his/her subgroup. All three subgroups of judges had mean interjudge agreement coefficients that were statistically significant: the choral conductors mean interjudge agreement coefficient was $r = .44$ (range = -.11 to .71, $p<.001$); the voice teachers had a mean interjudge agreement coefficient of $r = .32$ (range = -.11 to .58, $p<.001$); and the nonvocal musicians' mean interjudge agreement coefficient was $r = .26$ (range = -.08 to .54, $p<.005$).

A one-way independent groups analysis of variance was performed on judges' Pearson correlation coefficients, with judge subgroup as the independent groups factor. A significant difference in interjudge agreement was found among the three judge subgroups ($F = 10.46$, $p<.001$, $df = 2, 99$). Fisher's LSD post hoc tests revealed that choral conductors had significantly higher interjudge agreement coefficients than voice teachers ($p<.004$) and nonvocal musicians ($p<.001$). Differences between voice teachers and nonvocal musicians were not significant.

Reliability of subgroups of judges. As a measure of evaluation reliability, the ability of each subgroup of judges to predict evaluations by another equivalent group of judges was calculated using an equation analogous to the Spearman-Brown "prophecy" formula and Rozeboom's "alpha":

$$alpha = \frac{n \times r}{1+(n-1) \times r}$$

where n is the number of judges in the panel and r is the average judge-group correlation coefficient. The resulting judge group predictability coefficients were: .97 for choral conductors; .94 for voice teachers; and .92 for nonvocal musicians. These results indicate that the responses of all three subgroups could be used to predict, to a high degree of reliability, responses of other similar groups of musically sophisticated listeners.

Interjudge reliability for each criterion. As a measure of the comparative reliability of the seven evaluative criteria, each judge's ratings for a criterion were correlated with the mean ratings given by the judge's subgroup for that criterion. The Pearson correlation coefficients for each judge provided a measure of agreement between that judge and his/her subgroup for each criterion. A two-way mixed design analysis of variance, with criterion as the repeated measures factor and judge subgroup as the independent groups factor, was performed on the Pearson correlation coefficients as interjudge reliability scores. A significant main effect for criterion was found ($F = 4.16$, $p<.001$, $df = 6, 564$). Table 14 lists the mean reliability scores for the seven criteria for each subgroup of judges.

Fisher's LSD post hoc tests revealed that dynamic range and blend/homogeneity were significantly more reliable criteria than all the others, with the exception of overall tone quality. Overall tone quality was significantly more reliable than pitch precision and rhythmic precision. For choral conductors, blend/homogeneity and dynamic range were significantly more reliable than diction, rhythmic precision or pitch precision; overall tone

quality was more reliable than rhythmic precision or pitch precision. For voice teachers, there were no significant differences among the criteria dynamic range, pitch precision, blend/homogeneity, diction and overall tone quality; rhythmic precision was significantly less reliable than all other criteria, except phrasing; phrasing was less reliable than pitch precision or dynamic range. For nonvocal musicians there were no significant differences in reliability among the seven criteria.

Table 14

Mean interjudge reliability coefficients for the seven evaluative criteria for each subgroup of judges

All judges		**Choral conductors**		**Voice teachers**		**Nonvocal musicians**	
Criterion	**Mean**	**Criterion**	**Mean**	**Criterion**	**Mean**	**Criterion**	**Mean**
Dynamic range	.40	Blend/homogeneity	.49	Dynamic range	.39	Dynamic range	.32
Blend/homogeneity	.38	Dynamic range	.48	Pitch precision	.37	Blend/homogeneity	.28
Overall tone quality	.35	Overall tone quality	.47	Blend/homogeneity	.34	Phrasing	.27
Phrasing	.31	Phrasing	.40	Diction	.32	Rhythmic precision	.24
Diction	.31	Diction	.37	Overall tone quality	.32	Overall tone quality	.23
Pitch precision	.29	Rhythmic precision	.35	Phrasing	.25	Pitch precision	.22
Rhythmic precision	.26	Pitch precision	.30	Rhythmic precision	.19	Diction	.21

Analysis of judges' written comments

Of 102 judges who completed choral evaluation forms, 83 (81%) wrote comments. They were 31 choral conductors (84%), 27 voice teachers (82%), and 25 nonvocal musicians (78%). These comments were subsequently categorized by performance (e.g. Mozart, performance #1, experimental condition - soloistic/random) and then by topic.

The number of comments under each topic was tallied as a measure of the relative importance of the various topics. The topics, listed in order of decreasing importance (as determined by frequency of mention), were: tone quality (22% of comments), balance (12%), overall interpretation (12%), intonation (12%), blend (9%), dynamic range (7%), ensemble (6%), phrasing (6%), vibrato (5%), style (3%), diction (2%), tempo (2%), recording techniques (1%), texture (1%) and breath control (1%). In the same manner, the order of importance of the topics was also calculated for each subgroup of judges. Table 15 lists the topics in order of importance for all 83 judges who wrote comments, as well as for each of the three subgroups of judges.

Tone quality was the topic most frequently cited by judges in each of the three subgroups. Balance was next in order of importance only for the choral conductors (14% of comments) and nonvocal musicians (15%). Among the voice teachers, balance received only about half the percentage of comments (7%) that it did in the other two subgroups. Overall interpretation was of relatively high concern in all three subgroups. Intonation, however, appeared to be of considerably greater importance to voice teachers (14%) and choral conductors (11%) than to nonvocal musicians (8%). On the other hand, phrasing received more than twice the percentage of comments from nonvocal musicians

(11%) that it received from choral conductors (5%) or voice teachers (3%). Blend was of approximately equal concern to all three subgroups. Vibrato and style, however,

Table 15

Topics, according to which judges' comments were classified, listed in order of importance

All judges (n = 83)	**Choral conductors (n = 31)**	**Voice teachers (n = 27)**	**Nonvocal musicians (n = 25)**
tone quality (22%)	tone quality (17%)	tone quality (28%)	tone quality (19%)
balance (12%)	balance (14%)	intonation (14%)	balance (15%)
overall interpretation (12%)	overall interpretation (13%)	overall interpretation (11%)	phrasing (11%)
intonation (12%)	intonation (11%)	blend (10%)	overall interpretation (10%)
blend (9%)	blend (9%)	dynamic range (9%)	ensemble (8%)
dynamic range (7%)	vibrato (7%)	balance (7%)	blend (8%)
ensemble (6%)	dynamic range (6%)	vibrato (4%)	intonation (8%)
phrasing (6%)	ensemble (6%)	ensemble (4%)	dynamic range (6%)
vibrato (5%)	phrasing (5%)	diction (3%)	diction (3%)
style (3%)	style (5%)	phrasing (3%)	recording techniques (3%)
diction (2%)	tempo (2%)	style (2%)	texture (3%)
tempo (2%)	diction (1%)	tempo (1%)	breath control (2%)
recording techniques (1%)	recording techniques (1%)	texture (1%)	style (2%)
texture (1%)	texture (1%)	recording techniques (1%)	vibrato (1%)
breath control (1%)		breath control (< 1%)	tempo (1%)

appeared to be of greater concern to choral conductors (accounting for 7% and 5% of comments, respectively) than they were to voice teachers (4% and 2%, respectively) and nonvocal musicians (1% and 2%, respectively).

In order to examine tendencies associated with each of the four experimental conditions, comments were further classified by experimental condition. It was thus possible to compare the four experimental conditions with regard to comments on any topic. One basis of comparison was frequencies of positive and negative comments. Reliability of judgments regarding whether a comment was positive, negative or neutral was determined by having two judges separately label one quarter of the comments as positive, negative or neutral. The percentage of agreements between the two judges was 96%. This is equivalent to a correlation coefficient of $r = .98$, indicating high reliability in the judgments.

Tone quality. Comments on tone quality appeared to depend more on singing mode than on seating arrangement. In the soloistic singing mode, judges' favorable comments described the tone quality as: "bright," "full," "warmer," "rich," "mature," "free," "well-focussed," "resonant," "clear," "vibrant," "freer on top," "consistent," "even," "more sonorous," "lively," "less forced," "high notes matched low notes better." Negative comments described the tone quality as: "heavy," "forced," "aggressive," "overdone," "driven," "pushed," "too rich," "strident," "harsh," "ugly," "inconsistent," "shrill," "unpolished," "too dark," "too bright."

In the blended singing mode, typical favorable descriptions of tone quality were: "clear," "less strident," "better vowel placement," "warm," "relaxed," "more slender," "ringing," "more forward," "natural," "bright," "younger sound," "slim," "not forced," "light," "pretty," "most even," "consistent," "better high notes," "more focussed high notes," "better *forte*'s," "good in *piano*," "controlled," "assured," "unified." Typical negative comments described tone as: "restrained," "inhibited," "ugly," "shallow," "dull," "sterile," "thin," "closed," "lacking color," "lacking warmth," "lacking depth," "flimsy," "lifeless," "lacking roundness," "tight," "too straight," "muted," "shrill," "strident," "too bright," "straining for high notes," "harsh on top," "white," "naive," "pinched in tenors," "unfocussed in basses."

In order to compare further the comments on tone quality across the four experimental conditions, frequencies of positive and negative comments on this topic were tallied. Table 16 lists the total number of favorable and unfavorable comments on tone quality as well as the frequencies for each subgroup of judges for each experimental condition. There were considerably more unfavorable comments than favorable. The only exception was the soloistic/acoustic condition. Voice teachers wrote 24 favorable comments and 20 unfavorable on performances under this experimental condition. Pearson chi-square analysis revealed that voice teachers wrote a significantly greater number of positive comments and fewer negative comments on choral tone quality in the soloistic singing mode than in the blended mode ($X^2 = 7.84$, $p<.005$, $df = 1$). No other significant differences were found.

Table 16

Comparison of four experimental conditions with regard to judges' comments on tone quality

Experimental Condition	Comments	Choral conductors	Voice teachers	Other musicians	Total
Soloistic/ Random	Favorable	7	14	7	28
	Unfavorable	18	20	13	51
Soloistic/ Acoustic	Favorable	9	24	2	35
	Unfavorable	19	20	14	53
Blended/ Random	Favorable	8	8	3	19
	Unfavorable	18	30	11	59
Blended/ Acoustic	Favorable	5	15	3	23
	Unfavorable	21	31	10	62

Balance. The comments on balance were similar across all four experimental conditions. Judges either commented that balance was good, or they complained that basses were relatively too weak and sopranos too loud. Pearson chi-square analysis of the frequencies of positive and negative comments on balance for each of the four experimental conditions showed that judges wrote significantly fewer negative comments and a significantly greater number of positive comments for the blended singing mode than for the soloistic singing mode ($X^2 = 24.60$, $p<.001$, $df = 1$). Moreover, there were significantly fewer negative comments on balance in the acoustic seating arrangement than in the random seating arrangement ($X^2 = 4.59$, $p<.03$, $df = 1$).

Overall interpretation. Negative comments on overall interpretation were similar across the four experimental conditions: "unmusical," "soulless," "dull," "insensitive," "not pleasing," "unexpressive," "amateur," "no magic," "boring." Positive comments were more varied among the four experimental conditions. In the soloistic/random condition, typical favorable comments were: "professional," "effective opening," "effective ending." In the soloistic/acoustic condition, typical positive comments were: "most lively," "expressive," "lovely," "very musical," "more alive," "imaginative," "exciting," "lyrical," "energetic," "assured," "confident," "sounds easy," "effective opening," "effective ending." In the blended/random condition, typical favorable comments were: "professional," "more listening," "cleaner," "moving," "musical," "expressive," "good instrumental performance." In the blended/acoustic condition, typical positive comments were: "professional," "beautiful," "good feeling," "very musical," "more sensitive," "spiritually involved," "made musical sense," "nice restraint."

Pearson chi-square analysis of the frequencies of positive and negative comments on overall interpretation revealed that choral conductors wrote a significantly greater number of positive comments and fewer negative comments on blended singing than on soloistic singing ($X^2 = 12.72$, $p<.001$, $df = 1$). No other significant differences were found.

Intonation. Of 51 references to intonation problems in specific sections of the choir, 35 mentioned the soprano section, 3 cited the altos, none mentioned the tenors, and 13 referred to the basses. Pearson chi-square analysis found no statistically significant differences among experimental conditions in the frequencies of positive and negative comments on intonation in the choral performances. There was, however, a tendency for judges to make more negative comments about intonation in the blended singing mode (91) than in the soloistic singing mode (67).

Blend. Out of 49 comments on blend problems in particular sections of the choir, 39 mentioned the sopranos, one mentioned the altos, one mentioned the tenors, and eight mentioned the basses. Pearson chi-square analysis of the frequencies of positive and negative comments on blend among the four experimental conditions revealed that there were significantly fewer negative comments and more positive comments for choral performances in the acoustic seating arrangement than for those in the random seating arrangement (X^2 = 7.28, p<.007, df = 1). Moreover, the choral conductors wrote a significantly greater number of positive comments on blend for the blended singing mode than for the soloistic singing mode (X^2 = 4.30, p<.04, df = 1). There were no other significant differences.

Dynamic range. Unfavorable comments typically described the dynamics as "too loud" or the dynamic range as "too narrow." Typical favorable comments were: "greater variation in dynamics" or "more appropriate dynamic level." Pearson chi-square

analysis of the frequencies of positive and negative comments on dynamic range among the four experimental conditions found that there were significantly fewer negative comments and more positive comments for performances in the blended singing mode than for performances in the soloistic singing mode (X^2 = 8.56, p<.003, df = 1).

Ensemble. Judges' comments on this topic were generally concerned with rhythmic precision and aspects of unity or simultaneity within sections or in the choir as a whole. Typical favorable comments were: "good sense of rhythm," "good sense of togetherness," "rhythmically precise." Typical unfavorable comments were: "not rhythmically precise," "sloppy," "ragged," "not together." Pearson chi-square analysis of the frequencies of positive and negative comments on ensemble precision among the four experimental conditions showed that there were significantly fewer negative comments and more positive comments elicited by blended singing than by soloistic singing (X^2 = 9.24, p<.002, df = 1).

Phrasing. Positive comments on phrasing were typically as follows: "good phrasing," "better legato," "more flowing," "good sense of direction," "better sense of line," "good rhythmic vitality," "smoother." Typical negative comments classified under this topic were: "phrasing not good," "broken phrases," "lacks sense of line," "lacks text emphasis," "lacks legato," "lacks sense of direction," "lacks dynamic tension release," "too square," "plodding," "lacks syllabic stress," "dull phrasing," "not well shaped." Pearson chi-square analysis of the frequencies of positive and negative comments on phrasing

among the four experimental conditions revealed that choral conductors wrote significantly fewer negative comments and more positive comments for performances in the blended singing mode than for performances in the soloistic mode ($X^2 = 7.16$, $p<.007$, $df = 1$). No other significant differences were found.

Vibrato. Complaints of too much vibrato, especially in the soprano section, could be found across all four experimental conditions. In the soloistic singing mode, negative comments typically described the vibrato as "excessive," "not appropriate for the style of music" and "too operatic," and referred in particular to excessive vibrato in the soprano and tenor sections. Positive comments described the vibrato as "good," "more vibrant," "the right amount," and "vibrantly present." Other comments indicated that there seemed to be less vibrato in the soloistic/acoustic condition than in the soloistic/random condition. In the blended singing mode, negative comments described the vibrato either as "excessive" and "uncontrolled," particularly in the soprano section, or as "lacking in vibrancy" and "too straight." Positive comments in the blended singing mode qualified the vibrato as "better because less." Pearson chi-square analysis found that frequencies of positive and negative comments on vibrato were not significantly different among the four experimental conditions.

Style. Judges' comments on style depended to some extent on the musical piece. In the Mozart *Ave verum corpus*, performances in the soloistic singing mode were described favorably as "richly Romantic" or unfavorably as "too Romantic," "too

dramatic," "too operatic," "old-fashioned." One judge found the style in the blended/acoustic condition "too dramatic/operatic" for this piece. There were no negative comments on style for the blended/random condition in this piece. An example of a neutral comment on style was one that described the blended singing mode performances as "coolly classic."

In the Bruckner *Locus iste*, the blended/random condition produced two negative comments on style: "bad" and "too much like Mozart." One judge described the style in the soloistic/random condition as "not good" for this piece. The only positive comment on style for this piece was in the blended/acoustic condition: "good, more in Bruckner style." The style in this piece was labelled as "operatic" by at least one judge in each of the four experimental conditions.

In the Messiaen motet, the soloistic/acoustic condition received no comments on style. Each of the other three experimental conditions received one positive comment on style. The style in the blended/acoustic condition was preferred by one judge because it was "less operatic." There were no negative comments on style in this piece.

The greatest number of comments on style (18 out of 50) were elicited by the Renaissance motet. The soloistic/acoustic condition received 10 negative comments such as: "inappropriate," "too Romantic," "too much like Brahms." The remaining comments were fairly evenly distributed among the other three experimental conditions.

Comparison of the four experimental conditions on the basis of frequencies of positive and negative comments on style proved inconclusive, due to the relatively small number of comments.

Texture. Although there were too few comments (20) on this aspect of the choral performances for any conclusive results, the comments suggested that the judges tended to find the texture more clear or transparent in the blended singing mode than in the soloistic singing mode.

Analysis of the judges' comments on diction, tempo, recording technique and breath control proved inconclusive due to their relatively small numbers.

Summary of results

(1) A panel of 12 expert voice teachers ranked the vocal production of two individually monitored choristers in each section (SATB). They gave significantly higher vocal production rankings to solo singing than to the two choral singing modes ($p<.001$). Moreover, they preferred the soloistic choral mode to the blended choral mode ($p<.003$).

(2) The panel of voice teachers gave significantly higher vocal production rankings to the acoustic seating arrangement than to the random seating arrangement ($p<.05$).

(3) Choristers' vocal comfort ratings for the four experimental conditions indicated that they felt more comfortable singing in the acoustic seating arrangement than in the random seating arrangement ($p<.001$). In addition, ratings indicated that sopranos preferred soloistic singing to blended singing, and acoustic seating to random seating. Altos were least comfortable under the soloistic/random condition, with no other significant differences. Tenors preferred blended singing to soloistic singing. Basses preferred acoustic seating to random seating.

(4) Choristers' choral sound ratings for the four experimental conditions revealed that the blended singing mode was preferred to the soloistic singing mode. Furthermore, acoustic seating was preferred to random seating. The seating arrangement effect was significant for performances of the Mozart and Messiaen pieces, but was nonsignificant for the Bruckner and Victoria pieces.

(5) Choral conductors preferred blended singing to soloistic singing ($p<.001$). Voice teachers and nonvocal musicians indicated no significant singing mode effect on choral performance ratings. Singing mode accounted for 20% of the total variance in choral performance ratings overall, 58% among choral conductors, 7% among voice teachers, and 4% among nonvocal musicians.

(6) A significant seating arrangement effect in favor of acoustic seating over random seating was discovered. Seating arrangement accounted for 7% of the total variance in choral performance ratings, 17% among voice teachers, 4% among choral conductors, and 3% among nonvocal musicians.

(7) Singing mode had a significant effect in favor of blended singing over soloistic singing on choral conductors' ratings for all seven evaluative criteria. This effect was found in the voice teachers subgroup only for blend/homogeneity. Otherwise, no significant singing mode effect was observed in the voice teachers or nonvocal musicians subgroups.

(8) Seating arrangement had a significant effect, in favor of acoustic seating over random seating, on choral conductors' ratings for blend/homogeneity. This effect was also found in voice teachers' ratings for dynamic range, phrasing, and overall tone

quality. This effect approached significance in voice teachers' ratings for blend/homogeneity ($p<.06$).

(9) All seven criteria were significantly intercorrelated (mean = .67). Factor analysis revealed that they all were highly loaded onto one factor, the average of the seven criterion ratings. This factor explained 75 % of the total variance.

(10) A significant order effect was found in the choral performance ratings. Ratings were significantly lower for the first performance of a piece and significantly higher for the third performance than for any other performances ($p<.001$). Ratings for the second and fourth performances of a piece were not significantly different.

(11) Analysis of the effect of musical piece on choral performance ratings showed that singing mode had no significant interaction with musical piece. Choral conductors preferred blended singing to soloistic singing for all four musical selections. Similarly, voice teachers and nonvocal musicians had no significant singing mode preference regardless of the musical selection. The seating arrangement effect in favor of acoustic seating over random seating was significant only for the Mozart and Victoria pieces. Choral conductors and nonvocal musicians gave significantly lower ratings to the Victoria motet than to the other three pieces. Voice teachers gave significantly lower ratings to the Mozart than to the other three.

(12) All three judge subgroups had significant levels of interjudge agreement. Choral conductors had significantly higher interjudge agreement (mean $r = .44$) than did voice teachers (mean $r = .32$) and nonvocal musicians (mean $r = .26$). Differences between voice teachers and nonvocal musicians were not significant.

(13) Reliability of each subgroup of judges, as determined by their ability to predict evaluations by another equivalent group of judges was: a = .97 for choral conductors; a = .94 for voice teachers; and a = .92 for nonvocal musicians.

(14) Dynamic range and blend/homogeneity were significantly more reliable criteria than phrasing, diction, pitch precision and rhythmic precision. Overall tone quality was significantly more reliable than pitch precision and rhythmic precision.

(15) Eighty-one percent of the judges wrote comments on their choral evaluation forms. The most frequently addressed topics were tone quality (22% of comments), balance (12%), overall interpretation (12%), intonation (12%) and blend (9%).

(16) Pearson chi-square analyses of the frequencies of positive and negative comments across the four experimental conditions revealed the following:

- voice teachers wrote a significantly greater number of positive comments and fewer negative comments on choral tone quality in the soloistic singing mode than in the blended singing mode.
- judges wrote a significantly greater number of positive comments and fewer negative comments on balance, dynamic range and ensemble precision for performances in the blended singing mode than they did for performances in the soloistic singing mode.
- choral conductors wrote significantly more positive comments and fewer negative comments on overall interpretation, blend and phrasing for performances in the blended singing mode than for performances in the soloistic singing mode.
- there was a tendency for judges to make more negative comments about intonation in the blended singing mode conditions (91) than in the soloistic singing mode (67).

- judges wrote significantly fewer negative comments and more positive comments on blend in the acoustic seating arrangement than in the random seating arrangement.
- frequencies of positive and negative comments on vibrato were not significantly different among the four experimental conditions.

Conclusions

The first null hypothesis, that there would be no significant difference in choral performance ratings between performances in the soloistic singing mode and performances in the blended singing mode, must be rejected on the basis of the results of this study. The results indicated that choral conductors preferred blended singing to soloistic singing. This preference was significant across four different musical styles and seven different evaluative criteria.

The second null hypothesis, that there would be no significant difference in choral performance ratings between performances in an acoustic seating arrangement and performances in a random seating arrangement, also must be rejected. Acoustic matching of voices within sections had a beneficial effect on ratings of choral blend and overall tone quality as compared to random arrangement of choristers within sections.

Furthermore, results from this study lead to rejection of the third null hypothesis, that interjudge agreement would be no greater than that predicted by chance. All three judge subgroups attained interjudge agreement levels that were statistically significant, indicating that some generally accepted standards were informing judge evaluations.

The results shed light on the questions asked at the beginning of the study:

(1) Do trained singers modify their normal solo vocal production in a choral setting? If so, how do voice teachers evaluate their modified vocal production?

Voice teachers' rankings of performances by individually monitored choristers indicated perceptible differences in singers' vocal production across the three singing modes: solo, soloistic mode in a choral setting, and blended mode in a choral setting. This result is consistent with previous research findings on spectral differences between solo and choral singing which revealed basic differences in vocal production between these two modes of singing (Rossing et al., 1985, 1986, 1987; Sundberg, 1987).

The results imply that trained singers modify their vocal production when asked to blend in a choral setting, and that the resulting vocal production is inferior to both their normal solo vocal production and their soloistic vocal production in a choral setting. These findings further suggest that singers alter their vocal production in a choral setting even when instructed to use their normal solo vocal production, and that this soloistic vocal production is inferior to their solo vocal production. It thus appears that soloistic choral vocal production may be a compromise between solo vocal production and blended choral vocal production. This tendency to alter solo vocal production in a choral setting may be due to choristers' preference for the aesthetic quality of blended choral sound, acculturation to a choral aesthetic which does not allow the perceptibility of individual voices, and/or to acoustic properties of choral sound masking individual auditory feedback.

An interesting result is the effect of seating arrangement on vocal production of choristers. Vocal production was judged to be better in the acoustic seating arrangement than in the random seating arrangement. This result implies that positioning choristers so that adjacent voices are acoustically matched enhances vocal production. Choristers' written comments indicated that the acoustic seating arrangement gave them a greater impression of blending naturally. Thus they felt less need to restrain their voices in order to achieve an acceptable degree of homogeneity. Further research on voice matching, using spectral analysis, could shed light on this phenomenon and how it may enhance vocal production.

Voice teachers' comments revealed that their primary concern in evaluating vocal production was freedom of phonation. Terms such as "forced," "tight," "pressed" and "restrained," which denote a lack of vocal freedom, were widely used. According to the frequency of such comments, there appeared to be a tendency for singers to inhibit their freedom of phonation when singing in a choral setting, and that this tendency was greater when trying to blend than when singing soloistically. The voice teachers' attention to freedom of phonation is consistent with the views expressed by the American Academy of Teachers of Singing. In the opinion of AATS, freedom from vocal strain or tension is a major goal of voice training, and engaging in singing activities that sacrifice freedom of phonation could compromise students' vocal development (AATS, 1964). Thus the voice teachers' criticisms of vocal production in the choral setting may be more than mere expressions of preference for one style of singing (solo) over another (choral). They may be diagnoses of undesirable vocal habits that they believe to be detrimental to development of solo vocal technique.

The effect of choral singing on vocal development was not the focus of this study. Nevertheless, there is some evidence from the pilot study that singers had difficulty switching from blended singing mode to soloistic singing mode. Singers' comments indicated that singing in the blended mode created tension in the vocal apparatus, and that regaining a level of relaxation needed for the soloistic mode was subsequently difficult. However, the long-term effects of choral singing on development of solo vocal technique have yet to be studied.

(2a) How is choral blend affected by the type of vocal production used by choristers?

Results imply that individual vocal production has an effect on choral blend, and that trained singers are able to modify their normal solo vocal production in order to enhance choral blend. Written comments by judges and choristers suggest that strategies adopted by singers to maximize blend may involve singing with less vibrato and lower volume levels than in solo singing. Moreover, tenors may switch into falsetto register to maximize blend in their upper range. It is possible that singers develop these strategies for enhancing choral blend through experience with choral conductors who, either explicitly or implicitly, encourage this type of vocal production. The preference of choral conductors in this study for the blended singing mode over the soloistic singing mode supports this hypothesis. Furthermore, singers may be influenced by acculturation to a choral aesthetic that values homogeneity of tone. This is evidenced by the choristers' preference for choral sound in the blended singing mode over the soloistic mode. Singers in a choral setting may thus face a dilemma. They may be divided between a need to

respect the choral aesthetic of homogeneity of tone and a desire to sing in a way that promotes freedom of phonation. But are these necessarily mutually exclusive? This leads to the secondary question:

(2b) Is it possible to achieve an acceptable level of choral blend without modifying solo vocal production?

As the results suggest that singers modify their solo vocal production in a choral setting, even when asked not to do so, the following question might be more pertinent than the original question: Can an acceptable level of choral blend be achieved when singers try to use their normal solo vocal production?

Judges' evaluation of choral performances suggest that the answer generally depends upon the degree of choral conducting experience of the listener. The choral conductors and voice teachers with choral conducting experience significantly preferred performances in the blended singing mode over performances in the soloistic singing mode. Voice teachers without choral conducting experience and other musicians showed no significant preference for one singing mode over the other. The voice teachers in this study demonstrated an awareness of a higher level of blend/homogeneity in the blended singing mode. However, this did not seem to influence their choral performance ratings overall. It appears that musically sophisticated listeners other than choral conductors find choral tone quality equally acceptable in both singing modes.

(3) How is evaluation of overall choral sound affected by the type of vocal production used by choristers?

Findings suggest that the effect of singing mode on evaluation of choral performance depends upon the degree of choral conducting experience of the listener. It appears that choral conductors have a different choral aesthetic from other musically sophisticated listeners. Choral conductors rated all seven criteria according to their preference for blended singing over soloistic singing. Voice teachers rated blended singing higher than soloistic singing only for blend/homogeneity. Otherwise, voice teachers and nonvocal musicians showed no significant effect of singing mode on any of their criterion ratings.

Moreover, the seven evaluative criteria were highly intercorrelated and had high component loadings onto one factor - the average of the criteria ratings. These results indicate that the criteria ratings were all ratings of the same construct, namely, a global evaluation of the performance. This finding is consistent with research in performance evaluation which suggests that judges find it difficult to rate criteria independently, and that one global performance rating may be more reliable and meaningful than several criteria ratings (Fiske, 1975; Roberts, 1975; Ekholm, Papagiannis and Chagnon, 1998).

Tone quality was by far the most frequently addressed topic in the judges' written comments. This is not surprising in light of Stutheit's (1994) finding that tone quality and intonation were the most frequently selected criteria for choral evaluation. Comments concerning tone quality revealed that those who preferred the soloistic singing mode generally appreciated the fullness, freedom and vibrancy of the resultant tone. They

found that blended singing produced a tone quality that was too restrained, thin and lifeless. On the other hand, those who preferred the blended singing mode appreciated the lightness, clarity and control of the choral tone quality. They found that the soloistic singing mode produced a tone that was too heavy, aggressive and overdone.

Of particular interest is the finding that voice teachers wrote a significantly greater number of positive comments and significantly fewer negative comments on tone quality for choral performances in the soloistic singing mode than for performances in the blended singing mode. This was especially the case when soloistic singing was combined with acoustic seating. The acoustic seating arrangement may have added enough blend to the choral tone in the soloistic/acoustic condition to satisfy the demands of the voice teachers' choral aesthetic. From the analysis of vocal production rankings of the eight individually monitored choristers, one would expect voice teachers to prefer individual tone quality in soloistic singing to that in blended singing. The soloistic mode was, after all, an attempt by choristers to transfer the vocal technique taught by their voice teachers to a choral setting. The voice teachers' comments on tone quality, however, did not concur with their ratings of tone quality in the choral performances, which indicated no significant singing mode preference.

An unexpected result was the judges' preference for choral balance in the blended singing mode over the soloistic singing mode. As the number of singers in each section of the choir remained constant throughout the recording process, balance was not expected to vary from one experimental condition to another. Judges' comments, however, indicated that the soprano section was disproportionately loud in the soloistic mode.

This result may be partially explained by Weber's (1992) finding that vibrato tones were significantly louder than straight tones at high pitches, and particularly at high pitches combined with loud dynamic levels. The increased vibrato and louder dynamic levels in soloistic singing would have affected the SPL in the soprano section more than the other sections, as sopranos produced the highest pitches. Moreover, as the fundamental frequency goes above 500 kHz (C_4), the human ear becomes more sensitive than at lower frequencies, and the sound is perceived as being louder. Thus, as there was a tendency for all singers to get louder in the soloistic singing mode, the sopranos' increase in loudness may have been disproportionately amplified.

In addition, sopranos tend to tune their first formant to the fundamental frequency in their higher range, by a lowering of the jaw (Sundberg, 1977). The first formant thus enhances the amplitude of the fundamental and amplifies the sound. This amplification might occur to a greater extent in soloistic singing than in blended singing if there were a lower level of neck and jaw tension in the soloistic mode, which would allow for greater freedom of jaw movement.

Comments on overall interpretation suggest that choral performances in the blended singing mode were preferred to those in the soloistic singing mode. The terms "energetic," "exciting" and "more alive" were reserved for soloistic singing, while the terms "more listening," "cleaner" and "good instrumental performance" were reserved for blended singing. These terms suggest that soloistic singing would appeal to listeners who prefer a more energetic, dynamic interpretation. On the other hand, the terms used to describe blended singing evoke a more controlled, restrained interpretation.

Judges' comments on intonation accuracy revealed no significant differences among the experimental conditions. This is an interesting result in light of the importance of intonation accuracy for choral evaluation (Stutheit, 1994). It is, however, inconsistent with the finding that choral conductors gave significantly higher pitch precision ratings to choral performances in blended singing than they did to performances in soloistic singing. Intonation was the least reliable evaluative criterion for the choral conductors. It is therefore likely that quantitative analysis of judges' ratings for this criterion would be more sensitive to a halo effect in favour of blended singing than would qualitative analysis of judges' comments.

Voice teachers and nonvocal musicians showed no singing mode preference in their intonation ratings. Moreover, judges' written comments showed a tendency, though not statistically significant, to make a greater number of negative comments on intonation in the blended singing mode than in the soloistic singing mode. For the voice teachers who evaluated individually monitored choristers, intonation inaccuracy was the second most common problem, after lack of freedom of phonation, mentioned in their comments. The blended singing mode elicited the greatest number of complaints of poor intonation, followed by the soloistic singing mode, and finally the solo singing mode.

Research has shown that the main differences in vocal production between solo and blended choral singing are more vibrato and more energy in the upper partials (2-4 kHz range) in solo singing (Goodwin, 1980b; Rossing et al., 1985, 1986, 1987; Ternström, 1991). It might therefore be assumed that these were probably also the main acoustic differences between soloistic and blended singing in this study, albeit to a smaller

extent given the difference between solo and soloistic singing. Ternström and Sundberg (1982) found that the addition of vibrato did not affect the accuracy of pitch perception for a single complex tone. They observed, moreover, that the presence of upper partials facilitated intonation accuracy. They did, however, find that the absence of vibrato facilitated intonation accuracy in tuning intervals, by allowing beats to be heard as a clue to inaccurate tuning of partials common to both tones in the interval.

The present study, however, dealt with the perception of pitch precision by human judges without any form of objective measurement. It is possible that the increased presence of vibrato in soloistic singing may have masked slight inaccuracies in tuning not only from the singers but also from the judges, thereby obscuring any effect of singing mode on judges' perception of intonation accuracy. The greater number of negative comments on intonation accuracy in blended singing than in soloistic singing may also be explained by this masking effect.

There were over four times as many negative comments related to vibrato in soloistic singing as in blended singing, though there were approximately the same number of positive comments in both singing modes. The presence of vibrato in choral performances appeared to be of greater concern to choral conductors than to other musically sophisticated listeners. Ternström suggested that vibrato might be an effective way to promote perceptual differentiation of the solo voice from an accompaniment (Ternström, 1991, p.139). Research on perception of blend in orchestral instruments (Kendall and Carterette, 1993) revealed that blend correlated negatively with relative energy in the upper partials and with the presence of vibrato. It was concluded that

vibrato may aid in the separation and identification of individual instruments. Vibrato would thus be disadvantageous to choral blend.

Another unexpected result was that singing mode had no significant interaction with musical piece. Choral conductors preferred blended singing to soloistic singing regardless of musical selection, even though the selections represented four very different musical styles. Similarly, voice teachers and nonvocal musicians had no significant preference for singing mode regardless of the musical selection. One might have expected blended singing to be preferred for the Renaissance motet, and soloistic singing to be preferred for the Romantic piece. It must be remembered, however, that choristers had been instructed to attend to the style of each musical piece as much as possible in both singing modes. Despite evidence that some singers equated the soloistic singing mode with an operatic style of singing, there may nevertheless have been a significant number of singers who were able to adhere to the stylistic conventions of each musical piece to a similar extent in both singing modes.

(4) Does positioning of choristers according to acoustic matching of voices have any effect on choral blend or overall choral sound?

Ratings for the acoustic seating arrangement were significantly higher than they were for the random seating arrangement regardless of judge subgroup. In particular, ratings for blend/homogeneity, dynamic range, and overall tone quality were significantly higher for the acoustic seating arrangement than for the random seating arrangement. Judges also wrote a significantly greater number of positive comments and fewer negative

comments on blend for performances in the acoustic seating arrangement than for performances in the random arrangement. These results suggest that acoustic matching of voices may be beneficial to choral blend and overall tone quality.

The effect of seating arrangement, although significant, was not very strong, however. It accounted for only 7% of the total variance in choral performance ratings. Its effect was overshadowed by the singing mode factor, which accounted for 20% of the total variance. The effect of musical piece preference also accounted for 7% of the total variance. It is possible that the seating arrangement effect would be stronger in a live performance, although attempts to test this have proven inconclusive due to methodological problems (Tocheff, 1990; Daugherty, 1996).

It is not clear why the seating arrangement effect was significant only in the Mozart and Victoria pieces. Perhaps judges preferred a more homogeneous choral sound in these two earlier pieces, and therefore the added blend contributed by the acoustic seating arrangement may have been appreciated. Another reason may be that in trying to observe stylistic conventions in the two earlier pieces, singers made their soloistic singing more similar to their blended singing in these works than in the Bruckner or Messiaen pieces, thereby lessening the effect of singing mode and increasing the effect of seating arrangement.

Judges' written comments revealed that they preferred acoustic seating to random seating for choral balance. A reason for this may be found in the singers' comments about being better able to hear the ensemble in the acoustic seating arrangement, and therefore to adjust volume levels accordingly. Other comments indicated that the vibrato

was less prominent in the soloistic/acoustic condition than in the soloistic/random condition. This may be due to a possible vibrato-matching effect of the voice-matching procedure that determined the acoustic seating arrangement.

(5) What are the preferences of choristers with regards to singing under the various conditions in this study?

One result of choristers' evaluations was that sections of the choir were affected differently by the experimental conditions. Sopranos' preference for soloistic singing over blended singing, on the basis of vocal comfort, is not surprising in view of the written comments of sopranos and voice teachers indicating greater freedom of phonation in the soloistic mode. In addition, sopranos preferred acoustic seating to random seating for vocal comfort. This result is consistent with the finding that voice teachers preferred the vocal production of individually monitored choristers in the acoustic seating arrangement to their vocal production in the random seating arrangement.

Altos and basses, were not significantly affected by singing mode with regard to vocal comfort. It is possible that the reduced vocal freedom associated with blended singing may be less problematic when one is not required to sing at the top of one's range. Furthermore, as altos and basses tend to sing in a lower tessitura where the voice does not project as much, less modification of vocal production may be necessary to achieve choral blend. Like the sopranos, the altos and the basses found singing in the acoustic formation more comfortable vocally than singing in the random formation. According to their comments, in the acoustic arrangement they felt less need to restrain their voices in order to achieve an acceptable degree of choral blend.

A surprising finding is tenors' preference for blended singing over soloistic singing. One might expect the same result in the tenor section as in the soprano section, as both often sing in a high tessitura. Although in their written comments a few of the tenors mentioned the enhanced freedom of vocal production in the soloistic mode, other comments indicated that some of the tenors equated "soloistic" with "operatic," in spite of instructions to attend to the style of the music. The result was a dynamic level in the soloistic singing mode that may have been too loud, causing a tendency to oversing in order to hear oneself. Another explanation for the difference in singing mode preferences of tenors and sopranos is that the softer dynamic level in the blended singing mode made it possible for tenors to switch less perceptibly into their falsetto register for the higher passages, making high notes easier to attain than in their normal modal register. This practice is generally unacceptable in solo singing, and would account in part for the reduced vocal production rankings assigned by voice teachers to the individually monitored tenors in the blended singing mode.

Choristers' preference for choral sound in the blended singing mode is not surprising in light of the tendency in this direction in the pilot study. Part of the explanation may lie in the choristers' inclination to sing more loudly in the soloistic mode than in the blended mode, making it more difficult for individual singers to hear, and be sensitive to, the ensemble as a whole. This tendency at times may have become a vicious cycle, as one by one singers raised their volume level in order to hear their own voices. This, in turn, could cause a general increase in loudness, which would necessitate even louder singing, and so on. It is not clear in the present study whether this effect, known

as the Lombard effect, was generated by a lack of control and/or musicality on the part of a few loud singers, or a natural tendency arising from particular spectral characteristics of soloistic singing.

One of the main spectral differences between solo and choral singing is the greater amount of energy in the 2-4 kHz range in solo singing. The human ear is particularly sensitive to this frequency range. This is one reason that concentration of energy in this range is associated with projection of the voice. Increasing energy in this range, known as the singer's formant range, is an important goal in voice training. If choristers produce more energy in the 2-4 kHz range in the soloistic singing mode than in the blended singing mode, sounds with identical SPL would be perceived as louder in the soloistic mode than in the blended mode. This perception of increased loudness might be enough to generate a Lombard effect in a group of singers who are unaccustomed to singing together. If energy in the singer's formant range is more of a factor in soloistic singing than in blended singing, soloistic singing could also be expected to bring out differences in singers' vocal projection more than blended singing would. In the soloistic mode, singers with less powerful or less well trained voices (i.e. singers with less energy in the singer's formant range) would have a greater tendency to be overpowered by those with more powerful or better trained voices than in the blended mode. As Coleman (1994) suggested, one of the simplest strategies for enhancing choral blend is to ask the more powerful voices to reduce their SPL to match the dynamic level of the less powerful voices. This was probably one of the strategies employed by choristers in the blended mode, as was evidenced by the written comments of the choristers and the voice teachers indicating greater restraint in this mode.

One solution to the problem of the Lombard effect may be in greater spacing of the choristers. Choristers in this study were standing approximately four inches apart. Allowing a greater distance between choristers would increase the ratio of feedback (sound of one's own voice) to reference (sound of the other singers' voice), allowing individual singers to hear themselves more and thereby reducing the Lombard effect (Ternström and Sundberg, 1986, p.14). Moreover, in a study in which shoulder-to-shoulder spacings of zero, four and twelve inches were tested, high school choristers indicated that they thought that greater spacing between singers contributed positively to choral sound (Daugherty, 1996, p.25).

Choristers' written comments on appropriateness of tone quality for the various musical styles indicated that they found blended singing more appropriate for the Victoria and Mozart pieces, while some found soloistic singing more appropriate for the Bruckner. This could explain the differences in magnitude of the singing mode effect on the various musical pieces. The greatest difference in favor of blended singing was in the Victoria piece, while the least difference, albeit still significantly in favor of blended singing, was in the Bruckner.

Many of the written comments of both choristers and voice teachers cited vibrato. These subjects complained of too much vibrato in soloistic singing or not enough vibrato in blended singing. The inhibition of vibrato, or "straightening" of the tone, appeared to be a strategy adopted by the choristers in order to enhance blend. If vibrato aids in differentiation and identification of voices and instruments (Ternström, 1991; Kendall and Carterette, 1993), then it would be expected to have a detrimental effect on blend in choral singing.

Choristers' aesthetic preference for acoustic seating was consistent with the pilot study results. As the acoustic seating arrangement was made on the basis of enhancement of blend, the effect was probably similar to that of the blended singing mode, in that choristers were better able to hear, and be sensitive to, the ensemble as a whole. This was reflected in the choristers' written comments.

The effect of seating arrangement on choristers' choral sound ratings was significant only for the Mozart and Messiaen pieces. In the Victoria, stylistic appropriateness of the singing mode was of such concern that it may have overshadowed the seating arrangement effect. It is not clear why seating arrangement appeared to have no significant effect on choristers' choral sound ratings of the Bruckner performances. Perhaps the greater dynamic range of the piece, with its high and loud passages, made it more difficult for choristers to hear the ensemble as a whole. This would obscure the effect of acoustic seating.

(6) How do aesthetic preferences regarding choral sound compare among choral conductors, voice teachers, and other musicians? What are their levels of interjudge agreement?

The results revealed basic differences among the three subgroups of judges. Choral conductors were primarily influenced by singing mode, in favor of blended singing over soloistic singing. Singing mode accounted for 58% of the total variance in choral performance ratings given by choral conductors. Although seating arrangement had a statistically significant main effect in favor of acoustic seating over random seating,

among the choral conductors this was largely overshadowed by the singing mode factor. Seating arrangement accounted for only 4% of the total variance in choral conductors' ratings. There appears to be a strong relationship between choral conducting experience and preference for blended singing. Written comments indicated that choral conductors preferred the blended choral performances for their clarity of texture, lightness, control, and the impression they gave that choristers were listening more to each other. The work of a choral conductor involves training a group of singers to respond as one to directions. A choral conductor, therefore, must be concerned that choristers learn to be sensitive to his/her directions, both verbal and gestural, and to the rest of the ensemble. Furthermore, choristers' ability to control their vocal production, in order to be able to follow directions, would be important to choral conductors. It is not surprising that choral conducting experience should be associated with a highly developed appreciation for control, unity, and sensitivity to the ensemble and to the director.

The choral conductors' preference for choral performances in the blended singing mode may also be explained partly by one particular problem with the choral performances in the soloistic singing mode. In the soloistic mode, singers in this study tended to oversing due to problems in hearing themselves. This tendency for singers to increase vocal intensity as intensity of the choir increases is known as the "Lombard" effect. This effect was the probable cause of excessively loud dynamic levels, limited dynamic range and a balance in which sopranos were disproportionately loud. In fact, given these problems, it is surprising that the other two subgroups of judges did not also favor the blended singing mode. Research has shown that choristers can learn to resist

the Lombard effect and to consciously regulate their vocal intensity in the presence of a masking sound, when instructed to do so (Tonkinson, 1994). Thus specific instructions to resist the Lombard effect may help counteract the tendency to oversing in the soloistic mode. Another solution to the problem of the Lombard effect may be found in greater spacing of the choristers. Choristers in this study were standing approximately four inches apart. Allowing a greater distance between choristers would increase the ratio of feedback (sound of one's own voice) to reference (sound of the other singers' voice), allowing individual singers to hear themselves more and thereby reducing the Lombard effect (Ternström and Sundberg, 1986, p.14).

The voice teachers differed from the choral conductors in that their choral performance ratings were more affected by seating arrangement (17% of total variance in ratings) than by singing mode (7% of total variance in ratings). As the voice teachers were more tolerant in their singing mode preferences than the choral conductors, they may have been more open to the subtler differences caused by seating arrangement changes within a given singing mode. If the problems caused by the Lombard effect could be resolved, a significant effect in favor of soloistic singing might be expected among voice teachers. There was evidence from their written comments that they preferred the tone quality in soloistic singing to that in blended singing. Furthermore, voice teachers' rankings of vocal production of individually monitored choristers favored soloistic singing over blended singing. After all, soloistic singing is an attempt at transferring voice teachers' instructions from the voice studio to the choral setting.

For nonvocal musicians, the effects of singing mode and seating arrangement were small. The fact that the two main variables in this study only accounted for 7% of the total variance in nonvocal musicians' ratings is consistent with the low average interjudge agreement coefficient of .26 found in this subgroup. It appears thus that musically sophisticated listeners who are not singers, choral conductors or voice teachers may not be particularly sensitive to the blend-enhancing effects of blended singing and acoustic seating. Their evaluations of choral performance may be more influenced by musical considerations other than homogeneity of tone.

One can only speculate regarding the reasons for this difference in choral performance evaluation in nonvocal musicians. If one considers the sound of an orchestra, in which instruments of perceptibly different timbres often sound together on the same pitches, it seems that orchestral blend may be less a question of homogeneity of timbre and more a matter of creating interesting combinations of tone colors, only some of which have a homogeneous quality. Homogeneity of tone may thus not be as important in orchestral performance as in choral performance. It would then be understandable that musicians who have more experience with orchestral sound than with choral sound should be less sensitive to homogeneity of tone than are choral musicians. In any case, results suggest that a large group of musically sophisticated listeners may accept a wider range of choral sound than is generally accepted by choral musicians.

All three subgroups of judges had significant, though rather low, mean interjudge agreement coefficients. It seems that in spite of regional and personal differences in choral tone quality preference, some generally accepted standards informed judges' ratings.

Choral conductors' mean interjudge agreement coefficient was significantly higher than those of voice teachers and nonvocal musicians. Choral conductors' significant preference for blended singing over soloistic singing - an effect not found in the other two subgroups - and the relatively high percentage of total variance explained by this factor would account for the higher interjudge agreement in this subgroup.

The choral conductors' mean interjudge agreement coefficient is consistent with the research literature in vocal performance evaluation. Campbell (1971) found correlations of ratings by individual judges with average judge panel ratings to be in the range of .47 to .63 for the evaluation of solo vocal performance. Heller and Campbell (1971) found that the average correlation between any two judges on a panel of seven experts rating solo vocal performance was .40. Wapnick and Ekholm (1997) found an average judge-group correlation coefficient of .49 for 21 expert voice teachers who evaluated solo vocal performances. Such findings may be due to the inherent difficulty and subjectivity of the task of evaluating vocal performance, whether choral or solo.

Mean interjudge agreement coefficients for voice teachers and nonvocal musicians are lower than in previous studies. In this study, the four experimental conditions were controlled to eliminate extraneous variables as much as possible. Variables kept as constant as possible across all four performances of a musical piece were singers, tempo, conductor, conducting gestures, recording technique, concert hall, attention to musical and stylistic considerations of the piece. The result was four performances that were very similar except for the two independent variables of singing mode and seating arrangement. Singing mode, however, had no significant effect on ratings by these two subgroups of

judges, and the seating arrangement effect was small (approximately 10% of total variance in ratings by these two subgroups). It stands to reason that the more similar performances are, the more subtle discriminations among them become, and therefore the lower interjudge agreement tends to be.

Interjudge reliability was examined also for each of the seven evaluative criteria. Among the three most reliable evaluative criteria were the two of greatest interest to this study - blend/homogeneity and overall tone quality. This is not surprising when one considers that these are the two criteria most likely to be affected by singing mode and seating arrangement, and therefore most likely to differ from one performance to another. The relatively high reliability of dynamic range suggests that there were clearly perceptible differences in this aspect of choral sound among the experimental conditions. It appears that the performances in the soloistic singing mode may have been louder and less dynamically varied than those in the blended singing mode, probably due to a Lombard effect. The acoustic seating arrangement, moreover, had a positive effect on dynamic range ratings. This may have been due to choristers' enhanced ability to hear the ensemble. Judges' written comments on dynamic range revealed that they found more dynamic variation and more appropriate dynamic levels in blended singing than in soloistic singing.

The three least reliable evaluative criteria were diction, pitch precision and rhythmic precision. Diction probably varied the least from one performance to another, as it was not likely to be appreciably affected by the independent variables. Of particular interest is the lack of consensus among judges in the evaluation of pitch precision. Other

studies have found this criterion to be relatively reliable. Larkin (1985) found intonation to be the most reliable criterion and expression the least reliable in the measurement of achievement in choral music performance. Wapnick and Ekholm (1997) also found intonation accuracy to be the most reliable criterion along with overall score in the evaluation of solo vocal performance. The importance of reliable evaluation of this aspect of choral singing is indicated by Stutheit's (1994) finding that intonation and tone quality were the most frequently selected criteria for choral evaluation by adjudicators and choral directors.

The explanation for the relatively low reliability coefficients in ratings of pitch precision and rhythmic precision in this study probably lies in the fact that singing mode and seating arrangement had little effect on these aspects of the performances. This made the discriminations among performances too subtle to be very reliable. Only in the choral conductors subgroup did one of the independent variables, singing mode, have any significant effect on ratings of pitch precision and rhythmic precision. This may have been due to a halo effect in favor of the blended singing mode, however.

The moderate to low average interjudge correlation coefficients found among the criterion ratings are consistent with the research literature. Campbell (1971) reported interjudge correlations ranging from -.17 to .68 for a panel of seven expert judges who evaluated solo vocal performances based on intonation, vibrato, rhythm, dynamics and over-all. In Campbell's study, average interjudge correlations for these criteria were .45, .30, .27, .23 and .44, respectively, which lie in the same range as results from the present study.

Implications of findings

Results from this study are compatible with research showing spectral differences between solo singing and choral singing. Differences in vocal production were readily perceived by expert voice teachers who listened to individual singers in solo and choral settings. Moreover, the difference between solo and choral singing was greater when choristers were asked to blend their voices with those of singers around them than when they tried to maintain their normal solo vocal technique in the choral setting. This study found that the blended vocal production preferred by choral conductors was least preferred by voice teachers who evaluated individually monitored choristers. Herein may lie the basis of the conflict between voice teachers and choral conductors that frequently arises when voice students sing in choirs. The singers in this study were all trained and experienced in both solo and choral singing. Nevertheless, they had trouble maintaining freedom of phonation, intonation accuracy, vibrancy and resonance in the choral setting, especially when they were asked to blend.

This study has several practical implications for choral conductors, voice teachers and singers. It appears that consciously trying to blend one's voice with other voices results in a type of vocal production that is discouraged by voice teachers, as it inhibits "natural, free emission of tone" (AATS, 1964). Asking choristers to blend may thus be promoting vocal habits that are contrary to the teachings of their voice teachers. There seems to be a tendency for singers to inhibit freedom of phonation in a choral setting, even when not asked to blend. Nevertheless, when singers were allowed to sing soloistically in a choir, their vocal production, although modified from solo singing, was judged to be superior to their vocal production when asked to blend.

According to voice teachers, singers' vocal production in the blended mode was unsatisfactory due to a lack of freedom, intonation problems (especially flatting in the higher range), lack of vibrancy due to straightening of the tone, lack of breath control, lack of focus or ring in the tone, and a dull, weak and breathy tone quality. Blending with a choir may be more problematic for higher voices than for lower ones. Sopranos, in particular, indicated vocal discomfort and difficulty reaching notes at the extremes of their range when singing in a blended mode. Tenors may have to switch into the falsetto register to reach their higher range comfortably in blended mode. Altos and basses may have fewer problems with choral blend. However, when forced to sing in the lower extreme of their range, basses complained of difficulty focussing the tone in the blended mode and a resultant need to "push" in order to project the tone.

On the other hand, when asked to sing soloistically singers tended to oversing due to problems in hearing themselves and the resulting Lombard effect. Specific instructions to resist the Lombard effect may help counteract the tendency to oversing in the soloistic mode (Tonkinson, 1994). Greater spacing between choristers has also been recommended to increase the auditory feedback to reference ratio (Goodwin, 1980a, Ternström and Sundberg, 1986; Daugherty, 1996). Results from the present study suggest that arranging choristers so that voices are acoustically matched may enhance choral blend, dynamic range, phrasing, and overall choral tone quality. Moreover, acoustic seating may benefit choristers' vocal production, vocal comfort and aesthetic satisfaction.

Finally, results from this study suggest that choral conductors' standards for choral blend may be relaxed somewhat without affecting the aesthetic satisfaction of musically

sophisticated listeners. Herein may lie some room for compromise. Instead of encouraging blended vocal production, conductors might try to find other means of enhancing blend so that choristers, particularly sopranos, would be allowed to sing with greater freedom of vocal production. There is already evidence in the research literature of a trend among eminent choral conductors in the United States away from the historical practice of subordinating or altering the vocal quality of individual singers in order to achieve blend (Knutson, 1987). Alternative means of enhancing choral blend may include acoustic seating, asking certain sopranos, perhaps with more powerful or more resonant voices, to sing on a lower part rather than to restrain their voices, and perhaps choosing repertoire in which blend is less critical. Voice teachers might develop strategies to help students adapt their solo vocal technique to a choral setting with minimal vocal tension. For example, voice students might be taught to attend to proprioceptive feedback to monitor their vocal technique in situations where auditory feedback is insufficient.

Choral singing appears to involve some compromise in vocal technique for trained singers. There is clearly a difference between the vocal techniques used in solo singing and in choral singing. This is partly due to opposing aesthetic requirements of the two modes: the requirement that the solo voice be differentiated from the background sound versus the need for the choral voice to blend in with the surrounding sound. The acoustic environment of the choir, with its generally lower level of individual auditory feedback, may also account for some changes in vocal technique. If voice students are to benefit from choral singing while persuing the goal of free emission of tone promoted by their voice teachers, and if choral conductors are to benefit from the services of talented voice

students, it would be important for choral conductors and voice teachers to realize that some compromise is involved in choral singing. Greater cooperation between voice teachers and choral conductors could promote the development of choral techniques that would be acceptable to both groups, and which would better serve the needs of singers, particularly young singers in training.

Suggestions for further research

This study dealt only with short-term effects of singing mode and seating arrangement. It is possible that longer-term effects of these variables may differ from the results of this study. Recordings could be used to study long-term effects of specific choral techniques on choral sound and on vocal development, and to examine strategies of adaptation to the choral setting used by voice students. In view of the relatively high frequency of comments dealing with balance, it may be wise to include balance as one of the criteria for choral performance evaluation in future research.

A logical next step would be to test longer-term effects of soloistic choral singing and to determine how the Lombard effect might be successfully avoided. Greater spacing of choristers and instructions to resist the tendency to oversing are two possible interventions. The resulting choral sound could be evaluated by choral conductors, voice teachers and other musically sophisticated listeners. Individual choristers' vocal production could be evaluated by voice teachers to ascertain whether singers who strive to maintain solo vocal technique in the choral setting could succeed in doing so in the longer-term.

In addition to spectral analyses and perceptual studies of the differences between solo and choral singing, there is a need for laryngoscopic studies of physical techniques employed by trained singers to adapt solo vocal production to a choral setting. Researchers might then be better able to assess possible long-term effects of certain practices (e.g. inhibiting the vibrato, or singing in falsetto) on vocal development. The literature on vocal pedagogy for choirs deals mainly with basic vocal technique for untrained singers (Decker, 1976; Corbin, 1982). Specific techniques for adapting solo vocal production in trained singers to the requirements of a choral setting while promoting freedom of phonation and minimizing vocal tension could be collected from interviewing voice teachers and choral conductors. These techniques could then be tested empirically by means of spectral analyses, perceptual studies and laryngoscopic examinations.

Spectral analysis could be used to study the phenomenon of acoustic matching of voices. Findings might help explain its beneficial effects on choral sound as perceived by listeners, on the sensations of choristers with regard to vocal comfort and aesthetic satisfaction, and on the vocal production of choristers as assessed by voice teachers. Spectral analysis could also be used to investigate the matching of vibrato rates and extents, as well as the spectral composition of matched pairs or trios of voices in order to clarify how matched voices may be similar or complementary.

Results from the present study showed no significant interaction between musical style and either of the two independent variables under investigation. It would, however, be premature to generalize from the four musical pieces used in this study to the four style periods of which they were representative. In future research musical repertoire

could be varied in order to determine whether the results of this study would hold for music of different styles and periods. Several pieces from one period or style could be tested to determine the effects of seating arrangement and singing mode on choral performance of a specific style of repertoire.

Finally, results of this study provide evidence of differences in the factors that influence choral conductors, voice teachers, and nonvocal musicians in the evaluation of choral sound. Choral conductors were more influenced by singing mode than any other factor, voice teachers were more influenced by seating arrangement than by singing mode, and nonvocal musicians were barely influenced by either. Further research is needed to explain these differences, and to understand the role of experience in aesthetic perception.

Abeles, H. (1973). Development and validation of a clarinet performance adjudication scale. ***Journal of Research in Music Education, 21*** (4), 246-255.

American Academy of Teachers of Singing (1964). ***Choral Singing and the Responsibility of the Choral Director.***

Benade, A.H. (1976). ***Fundamentals of Musical Acoustics.*** New York: Oxford University Press. (cited in Goodwin, A., 1980b, p. 127)

Bolster, S.C. (1983). The fixed formant theory and its implications for choral blend and choral diction. ***The Choral Journal, 23*** (6), 27-33.

Bravender, P.E. (1984). Use of the Bernoulli effect in achieving choral blend. ***American Music Teacher, 34*** (1), 24-25.

Burnsed, V., Hinkle, D., and King, S. (1985). Performance evaluation reliability at selected concert festivals. ***Journal of Band Research, 21*** (1), 22-29.

Campbell, W.C. (1971). A computer simulation of musical performance adjudication. In E. Gordon (Ed.), ***Experimental Research in the Psychology of Music: VII,*** 1-40. Iowa City, Ia: Iowa University Press.

Cleall, C. (1970). ***Voice Production in Choral Technique.*** London: Novello.

Coleman, R.F. (1994). Dynamic intensity variations of individual choral singers. ***Journal of Voice, 8*** (3), 196-201.

Cooksey, J.M. (1977). A facet-factorial approach to rating high school choral music performance. ***Journal of Research in Music Education, 25*** (2), 100-114.

Corbin, L.A. (1982). Vocal pedagogy in the choral rehearsal: the influence of selected concepts on choral tone quality, student understanding of the singing process, and student attitudes toward choir participation. Ph.D. Dissertation, The Ohio State University.

Daugherty, J.F. (1996). Differences in choral sound as perceived by auditors and choristers relative to physical positioning and spacing of singers in a high school choir: a pilot study. Unpublished paper, presented at MENC Conference, April, 1996, Kansas City, MO.

Decker, W. (1976). A study of vocal pedagogy for the choral rehearsal based on theories presented in published literature from 1960-1970 and on interviews and observations of selected choral conductors. D.M.A. Dissertation, Temple University.

Ekholm, E. (1997). The effect of guided listening on evaluation of solo vocal performance. ***Contributions to Music Education, 24*** (2), 53-72.

Ekholm, E., Papagiannis, G.C., and Chagnon, F.P. (1998). Relating objective measurements to expert evaluation of voice quality in Western classical singing: critical perceptual parameters. ***Journal of Voice, 12*** (2), 182-196.

Fiske, H.E. (1975). Judge-group differences in the rating of high school trumpet performances. ***Journal of Research in Music Education, 23*** (3), 186-96.

Fiske, H.E. (1977a). New Insights into performance evaluation. ***Music Educators Journal,*** Dec., 23-5.

Fiske, H.E. (1977b). Relationship of selected factors in trumpet performance adjudication reliability. ***Journal of Research in Music Education, 24*** (4), 256-63.

Giardiniere, D.C. (1991). Voice matching: a perceptual study of vocal matches, their effect on choral sound, and procedures of inquiry conducted by Weston Noble. Ph.D. Dissertation, New York University.

Gonzo, C. (1973). Research in choral music: a perspective. ***Bulletin of the Council of Research in Music Education, 33***, 21-31.

Goodwin, A. (1980a). Resolving conflicts between choral directors and voice teachers. ***The Choral Journal, 21*** (1), 5-7.

Goodwin, A. (1980b). An acoustical study of individual voices in choral blend. ***Journal of Research in Music Education, 28*** (2), 119-128.

Green, P.C. (1937). Violin intonation. ***Journal of the Acoustical Society of America, 9,*** 43-44.

Harper, A.H., Jr. (1967). Spectrographic comparison of certain vowels to ascertain differences between solo and choral singing, reinforced by aural comparison. Ph.D. Dissertation, Indiana University.

Heller, J.J., and Campbell, W.C. (1971). Music performance analysis. ***Bulletin of the Council for Research in Music Education, 24***, 1-10.

Hunt, W.A. (1968). Spectrographic Analysis of the acoustical properties of selected vowels in choral sound. Ed.D. Dissertation, North Texas State University.

Husson, R. (1962). How the acoustics of a hall affect the singer and the speaker. ***The NATS Bulletin, 18*** (February), 8-13, 17.

Hylton, J. (1983). A survey of choral education research: 1972-1981. ***Bulletin of the Council for Research in Music Education, 76,*** 1-29.

Jordan, J.M. (1984). False blend: a vocal pedagogy problem for the choral conductor. ***The Choral Journal, 24*** (10), 25-26.

Kendall, R.A., and Carterette, E.C. (1993). Identification and blend of timbres as a basis for orchestration. ***Contemporary Music Review, 9*** (1&2), 51-67.

Keyne, L.V. (1992). Choral seating arrangements and their effects on musical and social elements. A.Mus.D. Dissertation, University of Arizona.

Killian, J.N. (1985). Operant preference for vocal balance in four-voice chorales. ***Journal of Research in Music Education, 33*** (1), 55-67.

Knutson, B.J. (1987). Interviews with selected choral conductors concerning rationale and practices regarding choral blend. Ph.D. dissertation, Florida State University.

Lambson, A.R. (1961). An evaluation of various seating plans used in choral singing. ***Journal of Research in Music Education, 9*** (1), 47-54.

Larkin, M.H. (1985). The construction and validation of rating scales for the objective measurement of five dimensions of achievement in choral music performance. D.M.A. dissertation, Temple University.

Lottermoser, W., and Meyer, J. (1960). Frequenzmessungen an gesungenen Akkorden. ***Acustica, 10,*** 181-184. (cited in Ternström and Sundberg, 1986, p. 12)

Madsen, C.K. and Yarbrough, C. (1980). ***Competency-Based Music Education.*** Englewood Cliffs, N.G.: Prentice-Hall, p. 61-3.

Magill, P., and Jacobsen, L. (1978). A comparison of the singing formant in the voices of professional and student singers. ***Journal of Research in Music Education, 26,*** 456-469.

Marshall, A.H. (1978). Acoustical conditions preferred for ensemble. ***Journal of the Acoustical Society of America, 64,*** 1437-1442.

Meyer, J. and Marshall, A.H. (1985). Schallabstrahlung und Gehörseindruck beim Sänger. ***Acustica, 58***, August, 130-40. (cited in Ternström and Sundberg, 1986, p.12)

Miller, A.W. (1992). Choral recordings as history: a study of the recording techniques 6 five choral organizations. Ph.D. dissertation, Florida State University.

Miller, R.D., and Schutte, H. (1983). Spectral analysis of several categories of timbre in a professional male (tenor) voice. ***Journal of Research in Singing,*** *7* (1), 6-10.

Nickerson, J.F. (1949). Intonation of solo and ensemble performance. ***Journal of the Acoustical Society of America, 21,*** 593.

Oxford International Dictionary of the English Language, Unabridged. (1957). Toronto: Leland Publishing Company, Ltd.

Page, E.B. and Paulus, D.H. (1968). The analysis of essays by computer. Final report, HEW Project No. 6-1318. Storrs: University of Connecticut.

Revesz, G. (1954). ***The Psychology of Music,*** trans. G. De Courcy. Norman: University of Oklahoma Press, 30-31.

Roberts, B. (1975). Judge-group differences in the ratings of piano performances. Unpublished master's thesis, University of Western Ontario.

Robinson, C.R. (1988). Differentiated modes of choral performance evaluation using traditional procedures and a continuous response digital interface device. Unpublished doctoral dissertation, The Floride State University.

Rossing, T.D., Sundberg, J., Ternström, S. (1985). Voice timbre in solo and choir singing: is there a difference? ***Journal of Research in Singing and Applied Vocal Pedagogy, 8*** (2), 1-8.

Rossing, T.D., Sundberg, J., Ternström, S. (1986). Acoustic comparison of voice use in solo and choir singing. ***Journal of the Acoustical Society of America, 79*** (6), 1975-1981.

Rossing, T.D., Sundberg, J., Ternström, S. (1987). Acoustic comparison of soprano solo and choir singing. ***Journal of the Acoustical Society of America, 82*** (3), 830-836.

Stutheit, S.A. (1994). Adjudicators', choral directors' and choral students' hierarchies of musical elements used in the preparation and evaluation of high school choral context performance. Unpublished D.M.A. dissertation, University of Missouri-Kansas City.

Swan, Howard. (1988). The development of a choral instrument. In ***Choral Conducting: A Symposium***, 2nd edition. Edited by Harold A. Decker and Julius Herford. Englewood Cliffs, N.J.: Prentice-Hall.

Sundberg, J. (1977). The acoustics of the singing voice. ***Scientific American, 236,*** 82-91.

Sundberg, J. (1978). Effects of the vibrato and the 'singing formant' on pitch. **Musicologica Slovaca, 6,** 51-69.

Sundberg, J. (1987). ***The Science of the Singing Voice.*** Dekalb, Ill.: Northern Illinois University Press.

Teie, E. (1976). A comparative study of the development of the third formant in trained and untrained voices. Doctoral dissertation, University of Minnesota.

Ternström, S. (1991). Physical and acoustic factors that interact with the singer to produce the choral sound. ***Journal of Voice, 5*** (2), 128-143.

Ternström, s. (1993). Perceptual evaluations of voice scatter in unison choir sounds. ***Journal of Voice, 7*** (2), 129-135.

Ternström, S. (1994). Hearing myself with others: sound levels in choral performance measured with separation of one's own voice from the rest of the choir. ***Journal of Voice, 8*** (4), 293-302.

Ternström, S., and Sundberg, J. (1986). Acoustics of choir singing. In S. Ternström (Ed.), ***Acoustics for Choir and Orchestra. Kungliga Musikaliska Akademiens skriftserie, Vol. 52,*** 12-22. Stockholm, Sweden: The Royal Academy of Music.

Ternström, S., and Sundberg, J. (1982). Acoustical factors related to pitch precision in choir singing. ***Royal Institute of Technology, Stockholm, Sweden. Speech Transmission Laboratory: Quarterly progress and status report, 2/3***, 76-90.

Ternström, S., Sundberg, J., and Collden, A. (1988). Articulatory F_0 perturbations and auditory feedback. ***Journal of Speech and Hearing Research, 31*** (2), 187-92.

Tocheff, R.D. (1990). Acoustical placement of voices in choral formations. Ph.D. dissertation, Ohio State University.

Tonkinson, S. (1994). The Lombard effect in choral singing. ***Journal of Voice, 8*** (1), 24-29.

Trevor, B.W. (1977). The influence of vibrato upon vocal blend. D.M.A. dissertation, University of Washington.

Vasil, T. (1973). The effects of systematically varying selected factors on music performing adjudication. Doctoral dissertation, University of Connecticut.

Vennard, W. (1967). ***Singing: The Mechanism and the Technic***. New York: Carl Fischer, Inc.

Wapnick, J. and Ekholm, E. (1997). Expert consensus in solo voice performance evaluation. ***Journal of Voice, 11*** (4), 429-36.

Weber, S.T. (1992). An investigation of intensity differences between vibrato and straight tone singing. Doctoral dissertation. Arizona State University.

Wyatt, L. (1967a). Blend in choral sound: factors related to its achievement - vowels. ***Choral Journal, 8*** (1), 15-18.

Wyatt, L. (1967b). Factors related to choral blend. ***Choral Journal, 8*** (2), 7-9.

Wyatt, L. (1968). Factors related to choral blend. ***Choral Journal, 8*** (4), 22-3.

Appendix A: Instructions to singers before audiotaping of choral performances for both pilot study and main study

Purpose of study: This is a study of various aspects of choral tone quality and choral blend. I am interested in finding out more about how certain variables affect choral tone quality. These variables are:

(1) vocal production (or the way the singers are producing their voices), and

(2) arrangement of the singers (or the way singers are positioned within each section)

If you have any questions, I will be happy to give you more information after the recordings are done.

Explanation of "soloistic vocal production:" Please sing the piece using the same kind of vocal production or vocal technique as if it were a solo piece. Sing as you normally would in your singing teacher's studio, producing your voice as your teacher has taught you. This does not mean that you must sing loudly. Try to respect the dynamics and musical style of the piece as much as possible while using your soloistic vocal production. Remember also that you are still in an ensemble setting, and so you must attend to all other aspects of ensemble singing, such as tempo, dynamics, phrasing, diction, and so on, much as you would if you were a soloist in a quartet, for example.

Explanation of "blended vocal production:" As you sing this piece, please focus on producing a blended or homogeneous ensemble sound. You should strive to eliminate perceptibility of individual voices.

Appendix B: Form used by choristers to evaluate experimental conditions in pilot study

SINGERS' PREFERENCE SHEET

Please rank the 4 experimental conditions under which you performed each piece according to vocal comfort and aesthetic preference. (In cases where there is no preference, please assign the same ranking, e.g. 1, 2, 2, 4 or 1, 1, 1, 1.)

1. Mozart - Ave verum

CONDITION	DESCRIPTION	VOCAL COMFORT 1 = most preferred 4 = least preferred	AESTHETIC PREFERENCE 1 = most preferred 4 = least preferred
1	- soloistic vocal production - random arrangement		
2	- blended vocal production - random arrangement		
3	- soloistic vocal production - acoustic arrangement		
4	- blended vocal production - acoustic arrangement		

Comments:

2. Bruckner - Gradual (Locus iste)

CONDITION	DESCRIPTION	VOCAL COMFORT 1 = most preferred 4 = least preferred	AESTHETIC PREFERENCE 1 = most preferred 4 = least preferred
1	- blended vocal production - random arrangement		
2	- soloistic vocal production - random arrangement		
3	- blended vocal production - acoustic arrangement		
4	- soloistic vocal production - acoustic arrangement		

Comments:

3. Hindemith - Un cygne

CONDITION	DESCRIPTION	VOCAL COMFORT 1 = most preferred 4 = least preferred	AESTHETIC PREFERENCE 1 = most preferred 4 = least preferred
1	- soloistic vocal production - acoustic arrangement		
2	- blended vocal production - acoustic arrangement		
3	- soloistic vocal production - random arrangement		
4	- blended vocal production - random arrangement		

Comments:

4. Victoria - O magnum mysterium

CONDITION	DESCRIPTION	VOCAL COMFORT 1 = most preferred 4 = least preferred	AESTHETIC PREFERENCE 1 = most preferred 4 = least preferred
1	- blended vocal production - acoustic arrangement		
2	- soloistic vocal production - acoustic arrangement		
3	- blended vocal production - random arrangement		
4	- soloistic vocal production - random arrangement		

Comments:

Demographic data

Voice classification: S M A T Bar B Age:

Years of voice training: Years of choral experience:

Appendix C: Instructions to judges evaluating choral performances in pilot study

Dear Colleague,

Thank you for your kind participation in my research project on choral tone quality. As you probably are aware, this project is a study of the effects that individual modes of vocal production and positioning of choristers may have on choral sound. Although many opinions have been expressed on this subject, very little empirical research has been conducted to test those opinions.

Enclosed you will find an experimental audiotape and an evaluation form. The tape contains four choral pieces, performed by a 20-voice mixed ensemble. Each piece is performed four times under differing experimental conditions. There is thus a total of 16 performances, which take about 40 minutes total time. Your task is to rate the performances, on a scale of 1 (poor) to 7 (excellent), according to a set of criteria: "blend/homogeneity," "intonation accuracy," "rhythmic precision," "diction" and "overall choral tone quality." You are encouraged to use the "comments" section of the evaluation form to help you remember the performances for comparison purposes and to enrich my understanding of your ratings.

It is not necessary to complete the entire tape at one sitting. You may do one piece at a time, as long as you complete all four performances of any given piece at the same sitting (for comparison purposes). Please feel free to listen as many times as necessary for a reliable evaluation. Please try to focus solely on the criterion that you are evaluating, and try not to be overly influenced by the unpolished nature of the performances. Bear in mind that this is a group of singers who were thrown together with very little time to rehearse, and that they were singing with no conductor (this was in order to eliminate conductor bias towards any of the experimental conditions). Moreover, the acoustics of the recording hall are purposely unflattering, in order to be able to hear any differences among the experimental conditions.

I truly appreciate your donating time and effort to this project, which will form the basis of my doctoral dissertation. I believe that this is an important and original research topic with considerable practical implications for singers, singing teachers and choral conductors.

Appendix D: Letter of invitation to choral conductors to participate in main study

Dear Colleague:

I would like to invite you to participate in an international research project on choral pedagogy being conducted at McGill University. As you know, much has been written and said about various techniques for enhancing choral sound quality, techniques involving vocal production, diction, positioning of choristers, and so on. Very little empirical research, however, has been conducted to study the effects of these variables on listeners' perception of choral sound. Moreover, there is much anecdotal evidence to suggest that voice teachers and choral conductors often do not agree on the type of vocal production demanded from singers. What are the preferences of expert choral conductors, voice teachers, as well as other musically sophisticated listeners around the world regarding choral sound, and how do they compare?

The purpose of this research project is to investigate some techniques that are commonly used to enhance choral sound and their effects on listeners' perception of choral sound. Your participation in this project would take a total of about 40-60 minutes and could be divided into four shorter segments to be completed at your convenience. Your task would involve listening to an audiotape of short excerpts from the standard choral repertoire, performed by a 22-voice mixed choir, and filling out a short evaluation form. Of course, any ratings, comments or other information that you might give would be strictly confidential.

If you are interested in participating and feel that you would be able to complete the task sometime this summer (before the end of September), please contact me, and I will send you the materials. Please feel free to call me collect or contact me by e-mail (or regular mail, of course). As a voice teacher and choral conductor myself, I believe that this research project will have important practical implications for choral conductors, voice teachers and singers. Your participation in this project would be very much appreciated.

Appendix E: French letter of invitation to choral conductors to participate in main study

Monsieur,

Par la présente, je vous invite à participer à un projet international de recherche (au niveau du doctorat) en pédagogie chorale de l'université McGill à Montréal. Comme vous savez, beaucoup d'opinions ont été exprimées au sujet de différentes techniques pour améliorer le son des chorales, techniques relatives à la production vocale, la diction, la disposition des chanteurs, etc. Cependant, très peu de recherche expérimentale a porté jusqu'à maintenant sur les effets de ces variables sur la perception du son choral par les auditeurs. De plus, beaucoup de faits suggèrent que les professeurs de chant et les chefs de choeurs sont souvent en désaccord au sujet de la production vocale qu'ils demandent des chanteurs.

Le but de ce projet de recherche est d'investiguer quelques techniques souvent utilisées pour améliorer le son choral et leurs effets sur la perception des auditeurs. Ce projet engagera la participation de chefs de choeurs, de professeurs de chant et d'autres connaisseurs de musique de plusieurs pays. Il permettra de comparer les préférences des divers groupes et d'identifier des points de convergence. Votre participation à ce projet demanderait un total de 40 à 60 minutes, qui pourrait être divisé en quatre segments a compléter à votre convenance. Votre tâche consisterait à écouter un enregistrement de courts extraits tirés du répertoire choral, interpretés par une chorale mixte de 22 voix, et à remplir une courte formule d'évaluation. Naturellement tous vos commentaires, évaluations et autres informations seraient strictement confidentiels.

Si vous désirez participer et croyez être en mesure de compléter la tâche cet été, s'il-vous-plaît contactez-moi par téléphone à frais virés (en PCV) ou par courrier électronique (e-mail). Il me fera plaisir de vous envoyer le matériel nécessaire. Comme professeur de chant et chef de choeur moi-même, je crois que ce projet de recherche aura des conséquences pratiques importantes pour les professeurs de chant, les chefs de choeurs et les chanteurs. Votre participation serait très appréciée.

Veuillez agréer, Monsieur, l'expression de mes sentiments distingués.

Appendix F: Letter of invitation to voice teachers to participate in main study

Dear Colleague:

I would like to invite you to participate in an international research project on vocal/choral pedagogy being conducted at McGill University. As a voice teacher you are undoubtedly aware of problems that can arise when voice students sing in choirs. There is much evidence to suggest that voice teachers and choral conductors often do not agree on the type of vocal production demanded from singers. As a result, some voice teachers have been known to advise their most promising students against singing in choirs. However, for many singers who aspire to a career in musical performance, choirs provide excellent musicianship training. Moreover, many young professional singers rely on choral work to round out their earnings while waiting for their solo careers to flourish.

The purpose of this research project is to investigate vocal production in choral ensembles and its effect on listeners' perception of choral sound. This project will involve the participation of expert voice teachers, choral conductors and other musicians around the world. Your participation in this project would take a total of about 40-60 minutes and could be divided into four shorter segments to be completed at your convenience. Your task would consist of listening to an audiotape of short excerpts from the standard choral repertoire, performed by a 22-voice mixed choir, and filling out a short evaluation form. Of course, any ratings, comments or other information that you might give would be strictly confidential.

If you are interested in participating and feel that you would be able to complete the task sometime this summer (by the end of September), please contact me, and I will be happy to send you the necessary materials. Please feel free to call me collect or contact me by e-mail. As a voice teacher myself and fellow NATS member who also conducts choirs, I believe that this research project will have important practical implications for choral conductors, voice teachers and singers. Your participation in this project would be very much appreciated.

Appendix G: French letter of invitation to voice teachers to participate in main study

Monsieur,

Par la présente, je vous invite à participer à un projet international de recherche (au niveau du doctorat) en pédagogie vocale/chorale de l'université McGill à Montréal. Comme professeur de chant, vous êtes sûrement familier avec les problèmes des étudiants de chant qui participent aux activités chorales. D'ailleurs, beaucoup de faits suggèrent que les professeurs de chant et les chefs de choeurs sont souvent en désaccord au sujet de la production vocale qu'ils demandent des chanteurs. En conséquence, il arrive que des professeurs de chant conseillent à leurs étudiants les plus prometteurs de ne pas chanter dans les chorales. Par contre, le chant choral fournit un excellent entraînement musical et permet souvent aux chanteurs qui aspirent à une carrière professionelle de gagner leur vie en attendant que leur carrière de soliste prenne de l'ampleur.

Le but de ce projet de recherche est d'étudier les types de production vocale utilisés dans les chorales et leur effet sur le son de l'ensemble. Ce projet engagera la participation de professeurs de chant, de chefs de choeurs et d'autres connaisseurs de musique de plusieurs pays. Votre participation à ce projet demanderait un total de 40 à 60 minutes, qui pourrait être divisé en quatre segments à compléter à votre convenance. Votre tâche consisterait à écouter un enregistrement de courts extraits tirés du répertoire choral, interpretés par une chorale mixte de 22 voix, et à remplir une courte formule d'évaluation. Naturellement tous vos commentaires, évaluations et autres informations seraient strictement confidentiels.

Si vous désirez participer et croyez être en mesure de compléter la tâche cet été, s'il-vous-plaît contactez-moi par téléphone à frais virés (en PCV) ou par courrier électronique (e-mail). Il me fera plaisir de vous envoyer le matériel nécessaire. Comme professeur de chant et chef de choeur moi-même, je crois que ce projet de recherche aura des conséquences pratiques importantes pour les professeurs de chant, les chefs de choeurs et les chanteurs. Votre participation serait très appréciée.

Veuillez agréer, Monsieur, l'expression de mes sentiments distingués.

Appendix H: Letter of invitation to nonvocal musicians to participate in main study

Dear Colleague:

I would like to invite you to participate in an international research project on choral pedagogy being conducted at McGill University. Much has been written and said about various techniques for enhancing choral sound quality, techniques involving vocal production, diction, positioning of choristers, and so on. Very little empirical research, however, has been conducted to study the effects of these variables on listeners' perception of choral sound. Moreover, there is much anecdotal evidence to suggest that voice teachers and choral conductors often do not agree on the type of vocal production demanded from singers. What are the preferences of choral conductors, voice teachers and other musicians around the world regarding choral sound, and how do these three groups compare?

The purpose of this research project is to investigate some techniques that are commonly used to enhance choral sound and their effects on listeners' perception of choral sound. Your participation in this project would take a total of about 40-60 minutes and could be divided into four shorter segments to be completed at your convenience. Your task would involve listening to an audiotape of short excerpts from the standard choral repertoire, performed by a 22-voice mixed choir, and filling out a short evaluation form. Of course, any ratings, comments or other information that you might give would be strictly confidential.

If you are interested in participating as a musician, who is not a choral conductor nor a voice teacher, and feel that you would be able to complete the task within the next month, please contact me, and I will be happy to send you the audiotape and evaluation form. As a choral conductor and voice teacher, I believe that this research project will have important practical implications for music educators. Your participation in this project would be very much appreciated.

Appendix I: Forms used by choristers to evaluate experimental conditions in main study

SINGERS' EVALUATION OF EXPERIMENTAL CONDITIONS

Please rate from 1 to 5 the experimental conditions, under which you performed each piece, according to vocal comfort and overall choral sound. **Please include comments** (English or French), if possible, with your ratings.

1. Mozart - Ave verum

CONDITION	**VOCAL COMFORT RATING** 1= very uncomfortable 5= very comfortable	**CHORAL SOUND RATING** 1= poor 5= excellent
soloistic vocal production / position **B**		
soloistic vocal production / position **A**		

2. Bruckner - Gradual (Locus iste)

CONDITION	**VOCAL COMFORT RATING** 1= very uncomfortable 5= very comfortable	**CHORAL SOUND RATING** 1= poor 5= excellent
soloistic vocal production / position **A**		
soloistic vocal production / position **B**		

3. Messiaen - O sacrum convivium

CONDITION	VOCAL COMFORT RATING 1= very uncomfortable 5= very comfortable	CHORAL SOUND RATING 1= poor 5= excellent
soloistic vocal production / position **B**		
soloistic vocal production / position **A**		

4. Victoria - O magnum mysterium

CONDITION	VOCAL COMFORT RATING 1= very uncomfortable 5= very comfortable	CHORAL SOUND RATING 1= poor 5= excellent
soloistic vocal production / position **A**		
soloistic vocal production / position **B**		

DEMOGRAPHIC DATA

Name: Years of voice training:

Voice classification: S M A T Bar B Years of choral experience:

Age:

Appendix I continued: (Forms used by choristers to evaluate experimental conditions in main study)

SINGERS' EVALUATION OF EXPERIMENTAL CONDITIONS (2nd recording session)

Please rate from 1 to 5 the experimental conditions, under which you performed each piece, according to vocal comfort and overall choral sound. **Please include comments** (English or French), if possible, with your ratings.

1. Mozart - Ave verum

CONDITION	VOCAL COMFORT RATING 1= very uncomfortable 5= very comfortable	CHORAL SOUND RATING 1= poor 5= excellent
blended vocal production / position **B**		
blended vocal production / position **A**		

2. Bruckner - Gradual (Locus iste)

CONDITION	VOCAL COMFORT RATING 1= very uncomfortable 5= very comfortable	CHORAL SOUND RATING 1= poor 5= excellent
blended vocal production / position **A**		
blended vocal production / position **B**		

3. Messiaen - O sacrum convivium

CONDITION	VOCAL COMFORT RATING 1= very uncomfortable 5= very comfortable	CHORAL SOUND RATING 1= poor 5= excellent
blended vocal production / position **B**		
blended vocal production / position **A**		

4. Victoria - O magnum mysterium

CONDITION	VOCAL COMFORT RATING 1= very uncomfortable 5= very comfortable	CHORAL SOUND RATING 1= poor 5= excellent
blended vocal production / position **A**		
blended vocal production / position **B**		

Please choose the type of vocal production, **soloistic or blended** (S or B), which you preferred on the basis of vocal comfort and overall choral sound for each piece.

PIECE	VOCAL COMFORT	CHORAL SOUND
Mozart - Ave verum		
Bruckner - Gradual (Locus iste)		
Messiaen - O sacrum convivium		
Victoria - O magnum mysterium		

Appendix J: Form used to evaluate conductor consistency from videotape

Conductor Consistency Observation Form

Piece:___________________________________

	Movement to/ from choir			**Conducting gestures**		**Magnitude of gestures**		**Eye contact**		**Facial expression**			**Body movement (except arms/hands)**		
	Approaching / Departing / Stationary			Strict / Expressive		High / Low		Group/ Other		Approving / Disapproving / Neutral			Much	Some	Non
1	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
2	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
3	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
4	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
5	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
6	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
1	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
2	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
3	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
4	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
5	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
6	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
1	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
2	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
3	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
4	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
5	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
6	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
1	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
2	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
3	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
4	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
5	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non
6	A	D	S	S	E	H	L	G	O	A	D	N	Much	Some	Non

Appendix K: Instructions to voice teachers evaluating individually monitored choristers in both pilot study and main study

Dear Colleague:

Thank you again for agreeing to participate in this research project on choral tone quality. The project is a study of the effects that individual modes of vocal production and positioning of choristers may have on choral tone quality. Although many opinions have been expressed on this subject, very little empirical research has been conducted to test those opinions.

Enclosed you will find an experimental tape (# 2) and a **Vocal Production Evaluation Form**. The tape contains excerpts of choral parts sung by individual singers within a choral setting as well as solo. Eight singers were chosen for these individual recordings - two from each choral section (soprano, alto, tenor, bass). Each singer is heard on the tape in a block of five performances of an excerpt under different experimental conditions. **Your task is to compare the vocal production in the five performances within each block and rank them 1 to 5 (best to worst).**

Please try to focus solely on how well the singer is using his/her voice, from the point of view of a singing teacher, rather than on musical or artistic concerns. You will not have to compare one singer to another; you will compare only performances by the same singer. Therefore you need not complete the entire tape at one sitting, but only as many blocks as you wish at one time. There are eight blocks of five performances each on the tape. The excerpts range from 30 seconds to a minute in length. Please feel free to listen as many times as necessary for a reliable evaluation. If you cannot hear any difference between two or more performances, please give them an equal ranking (e.g. 1 2 2 3 4, or 1 2 3 3 3, etc.).

Please use a good pair of headphones for more reliable reproduction. As Dolby B was used in the recording, it is recommended that Dolby B be added in the playback as well, although this is not necessary. You may try with or without Dolby and use whichever you feel gives a more realistic sound, as long as the same playback conditions are used for all performances in any block. You are encouraged to use the "comments" section of the evaluation form to help you remember the performances for purposes of ranking them and to enrich my understanding of your rankings. (You can use an additional sheet for longer comments.) You can simply return the completed evaluation form in the enclosed self-addressed stamped envelope.

I truly appreciate your donating time and effort to this project. As a singer, singing teacher and choral conductor, I believe that this is an important topic for research and one that has considerable practical implications.

Appendix L: Evaluation form used by voice teachers in evaluation of individually monitored choristers

VOCAL PRODUCTION EVALUATION FORM (tape # 2)

Victoria - O magnum mysterium - soprano line

PERFORMANCE	RANKING	COMMENTS
1		
2		
3		
4		
5		

Mozart - Ave verum - soprano line

PERFORMANCE	RANKING	COMMENTS
1		
2		
3		
4		
5		

Messiaen - O sacrum convivium - tenor line (# 1)

PERFORMANCE	RANKING	COMMENTS
1		
2		
3		
4		
5		

Bruckner - Locus iste - alto line

PERFORMANCE	RANKING	COMMENTS
1		
2		
3		
4		
5		

Messiaen - O sacrum convivium - tenor line (# 2)

PERFORMANCE	RANKING	COMMENTS
1		
2		
3		
4		
5		

Victoria - O magnum mysterium - alto line

PERFORMANCE	RANKING	COMMENTS
1		
2		
3		
4		
5		

Mozart - Ave verum - bass line

PERFORMANCE	RANKING	COMMENTS
1		
2		
3		
4		
5		

Victoria - O magnum mysterium - bass line

PERFORMANCE	RANKING	COMMENTS
1		
2		
3		
4		
5		

CHORAL EVALUATION FROM (C)

On the experimental audiotape, you will hear four short choral excerpts (of approximately 2 minutes' duration), each of which is performed four times under differing experimental conditions.

Please assign a rating, from 1 (poor) to 7 (excellent) for each performance, under each of the criteria. The "comments" column is for brief comments. For longer comments, please write them on a separate sheet, being careful to indicate the piece and performance number to which they refer.

Please use a good pair of headphones for more reliable reproduction. As Dolby B was used in the recording, it is recommended that Dolby B be added in the playback as well, although this is not necessary.

Please evaluate all four performances of any piece in the same session (for more reliable comparison). It is not necessary, however, to complete the entire tape in one session.

Please feel free to listen as many times as you find necessary for a reliable evaluation.

Please use the self-addressed envelope to return the evaluation form (and extra sheets, if any). **There is no need to return the audiotape.**

Mozart - Ave verum:

Perf.	blend/ homo-geneity	diction	dynamic range	phras-ing	pitch preci-sion	rhythmic preci-sion	overall tone quality	comments
# 1								
# 2								
# 3								
# 4								

Bruckner - Locus iste:

Perf.	blend/ homo-geneity	diction	dynamic range	phras-ing	pitch preci-sion	rhythmic preci-sion	overall tone quality	comments
# 1								
# 2								
# 3								
# 4								

Messiaen - O sacrum convivium:

Perf.	blend/ homo-geneity	diction	dynamic range	phras-ing	pitch preci-sion	rhythmic preci-sion	overall tone quality	comments
# 1								
# 2								
# 3								
# 4								

Victoria - O magnum mysterium:

Perf.	blend/ homo-geneity	diction	dynamic range	phras-ing	pitch preci-sion	rhythmic preci-sion	overall tone quality	comments
# 1								
# 2								
# 3								
# 4								

Demographic data

Please fill in the following table to give some idea of your predominant experience as a choral director:

type of choir (e.g.church, children's, opera chorus, university mixed choir, men & boys, women's, men's, etc.)	**# years of experience**	**predominant repertoire** (e.g. mixed repertoire early music, Gregorian chant, 20th C, jazz popular, folk, etc)

What is your primary instrument? (If voice, please indicate classification, e.g. SMATBarB)_____

What, if any, other instruments do you play?_____

CHORAL EVALUATION FORM (V)

On the experimental audiotape, you will hear four short choral excerpts (of approximately 2 minutes' duration), each of which is performed four times under differing experimental conditions.

Please assign a rating, from 1 (poor) to 7 (excellent) for each performance, under each of the criteria. The "comments" column is for brief comments. For longer comments, please write them on a separate sheet, being careful to indicate the piece and performance number to which they refer.

Please use a good pair of headphones for more reliable reproduction. As Dolby B was used in the recording, it is recommended that Dolby B be added in the playback as well, although this is not necessary.

Please evaluate all four performances of any piece in the same session (for more reliable comparison). It is not necessary, however, to complete the entire tape in one session.

Please feel free to listen as many times as you find necessary for a reliable evaluation.

Please use the self-addressed envelope to return the evaluation form (and extra sheets, if any). **There is no need to return the audiotape.**

Mozart - Ave verum:

Perf.	blend/ homo-geneity	diction	dynamic range	phras-ing	pitch preci-sion	rhythmic preci-sion	overall tone quality	comments
# 1								
# 2								
# 3								
# 4								

Bruckner - Locus iste:

Perf.	blend/ homo-geneity	diction	dynamic range	phras-ing	pitch preci-sion	rhythmic preci-sion	overall tone quality	comments
# 1								
# 2								
# 3								
# 4								

Messiaen - O sacrum convivium:

Perf.	blend/ homo-geneity	diction	dynamic range	phras-ing	pitch preci-sion	rhythmic preci-sion	overall tone quality	comments
# 1								
# 2								
# 3								
# 4								

Victoria - O magnum mysterium:

Perf.	blend/ homo-geneity	diction	dynamic range	phras-ing	pitch preci-sion	rhythmic preci-sion	overall tone quality	comments
# 1								
# 2								
# 3								
# 4								

Demographic data

How many years have you been teaching voice?_____

What is your voice classification (e.g.soprano, mezzo, baritone, etc.)?_____

Have you also directed choirs?_____ If so, please fill in the following table to give some idea of your predominant experience as a choral director:

type of choir (e.g.church, children's, opera chorus, university mixed choir, men & boys, women's, men's, etc.)	**# years of experience**	**predominant repertoire** (e.g. mixed repertoire early music, Gregorian chant, 20th C. jazz popular, folk, etc)

CHORAL EVALUATION FORM (M)

On the experimental audiotape, you will hear four short choral excerpts (of approximately 2 minutes' duration), each of which is performed four times under differing experimental conditions.

Please assign a rating, from 1 (poor) to 7 (excellent) for each performance, under each of the criteria. The "comments" column is for brief comments. For longer comments, please write them on a separate sheet, being careful to indicate the piece and performance number to which they refer.

Please use a good pair of headphones for more reliable reproduction. As Dolby B was used in the recording, it is recommended that Dolby B be added in the playback as well, although this is not necessary.

Please evaluate all four performances of any piece in the same session (for more reliable comparison). It is not necessary, however, to complete the entire tape in one session.

Please feel free to listen as many times as you find necessary for a reliable evaluation.

Please use the self-addressed envelope to return the evaluation form (and extra sheets, if any). **There is no need to return the audiotape.**

Mozart - Ave verum:

Perf.	blend/ homo-geneity	diction	dynamic range	phras-ing	pitch preci-sion	rhythmic preci-sion	overall tone quality	comments
# 1								
# 2								
# 3								
# 4								

Bruckner - Locus iste:

Perf.	blend/ homo-geneity	diction	dynamic range	phras-ing	pitch preci-sion	rhythmic preci-sion	overall tone quality	comments
# 1								
# 2								
# 3								
# 4								

Messiaen - O sacrum convivium:

Perf.	blend/ homo-geneity	diction	dynamic range	phras-ing	pitch preci-sion	rhythmic preci-sion	overall tone quality	comments
# 1								
# 2								
# 3								
# 4								

Victoria - O magnum mysterium:

Perf.	blend/ homo-geneity	diction	dynamic range	phras-ing	pitch preci-sion	rhythmic preci-sion	overall tone quality	comments
# 1								
# 2								
# 3								
# 4								

Demographic data

What is your primary instrument? (If voice, please indicate classification, e.g. SMATBarB)_____

What, if any, other instruments do you play?_____

How often do you listen to choral music?(circle one) often/sometimes/seldom/hardly ever

Have you ever directed a choir?_____ If so, please fill in the following table to give some idea of your predominant experience as a choral director:

type of choir (e.g.church, children's, opera chorus, university mixed choir, men & boys, women's, men's, etc.)	**# years of experience**	**predominant repertoire** (e.g. mixed repertoire early music, Gregorian chant, 20th C, jazz popular, folk, etc)

Appendix P: Scores of choral excerpts performed in pilot and main studies

Mozart - ***Ave verum corpus*** (a cappella version based on L'Ensemble Vocal Philippe Caillard edition)

Bruckner - ***Locus iste*** (Peters edition)

Hindemith - ***Un cygne*** (B. Schott's Söhne edition, used only in pilot study)

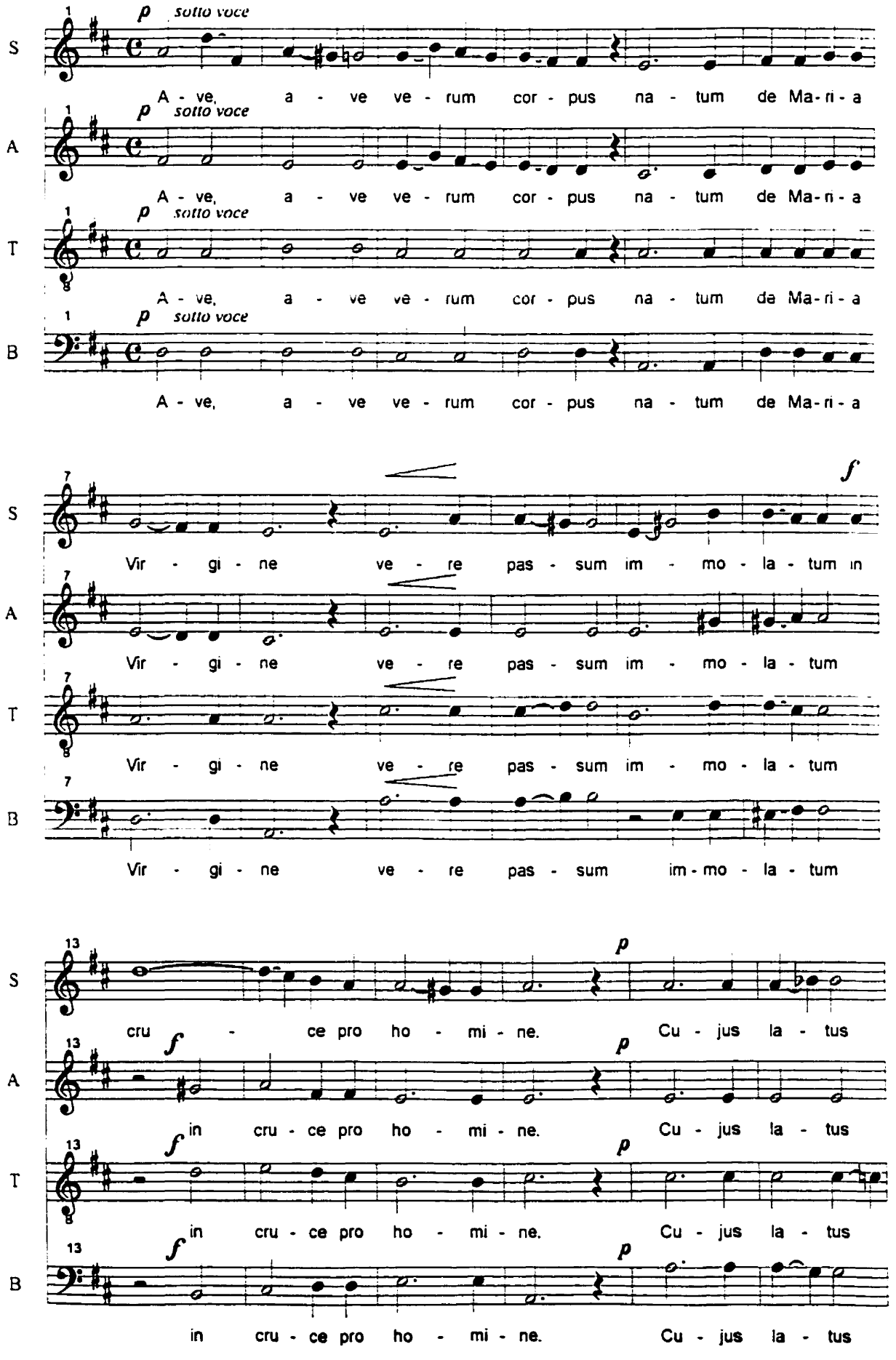
Victoria - ***O magnum mysterium*** (mm. 1 - 39, Schirmer edition, edited by Alice Parker and Robert Shaw)

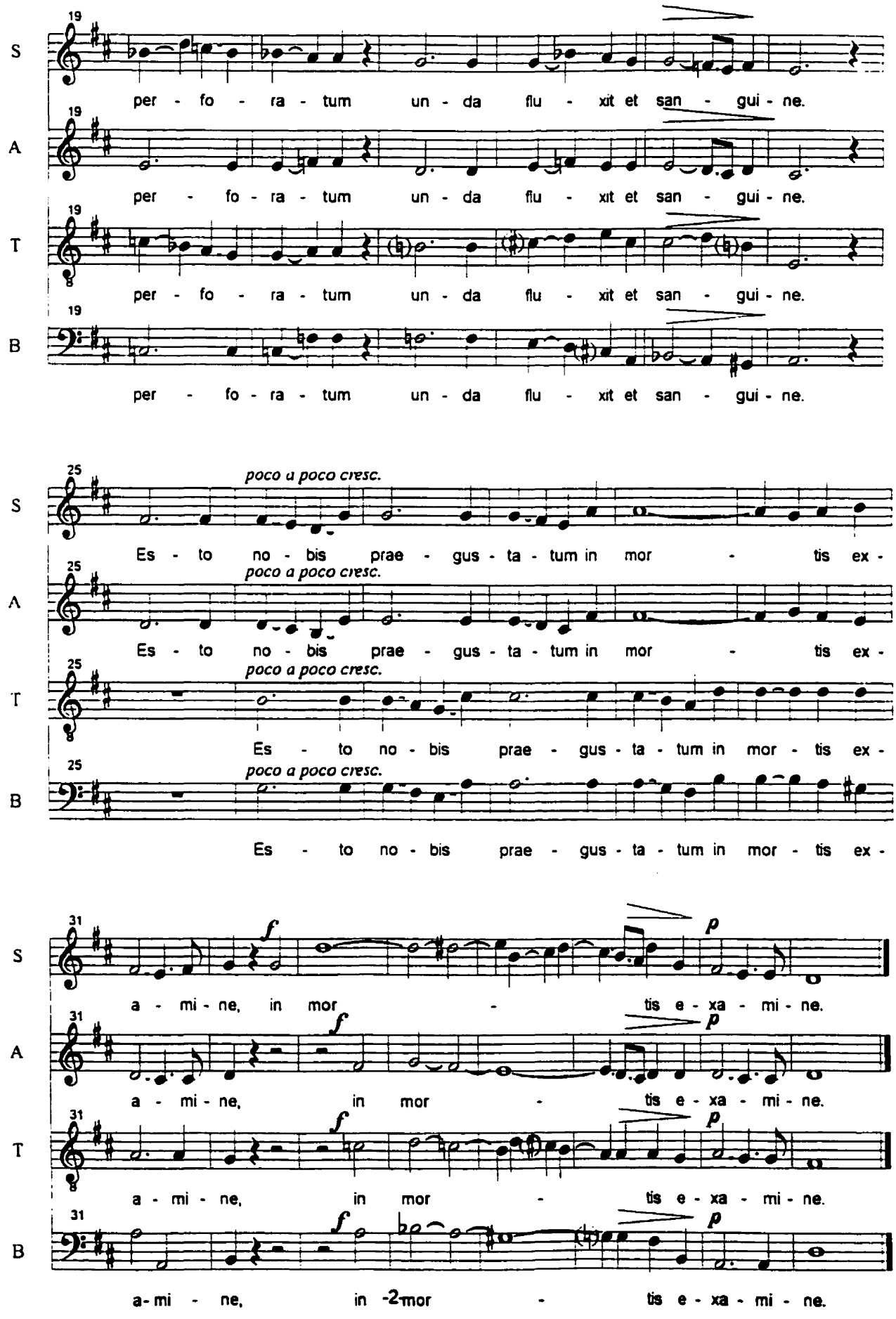
Messiaen - ***O sacrum convivium*** (mm. 17 to end, Durand edition)

AVE VERUM CORPUS

W.A. Mozart

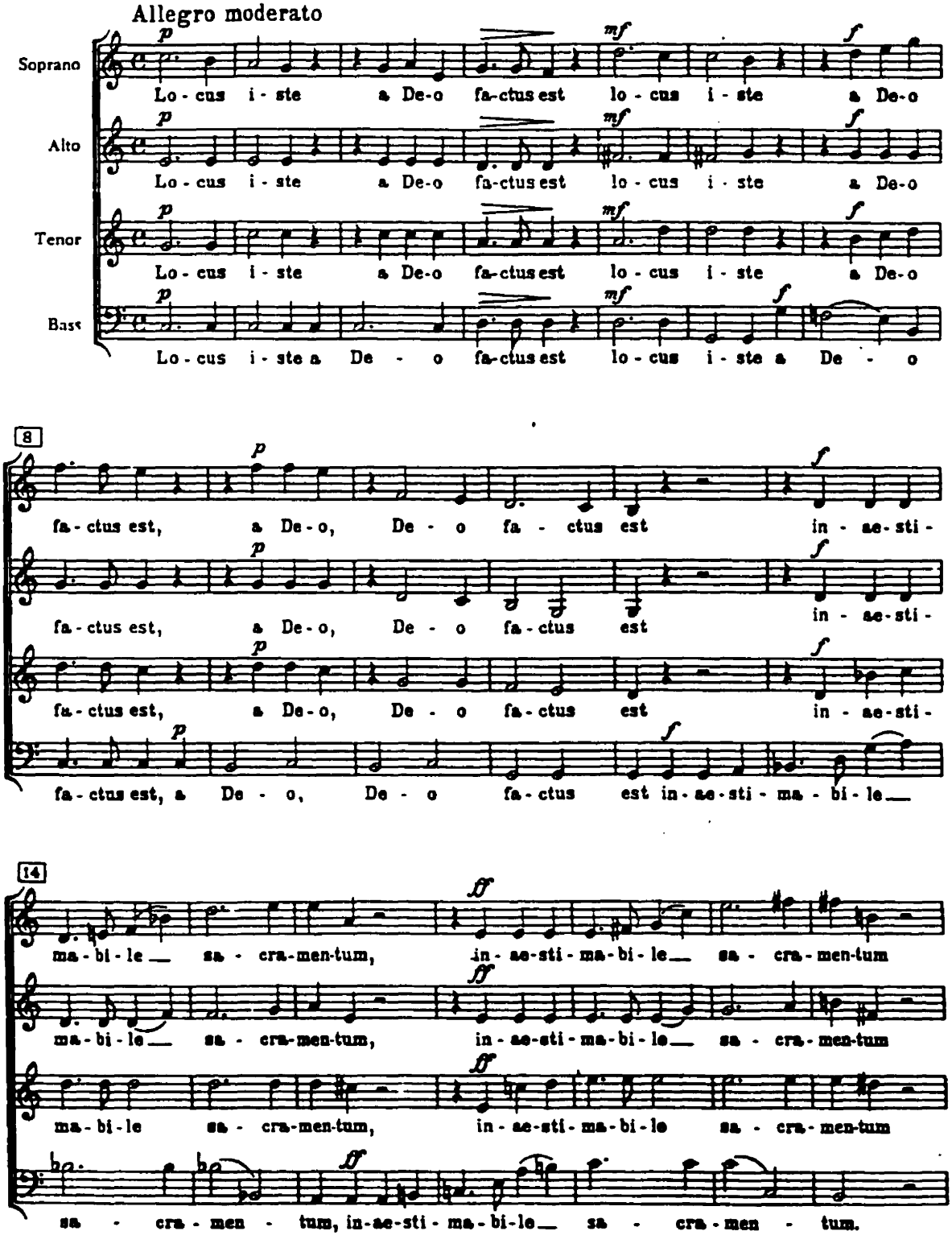
Adagio

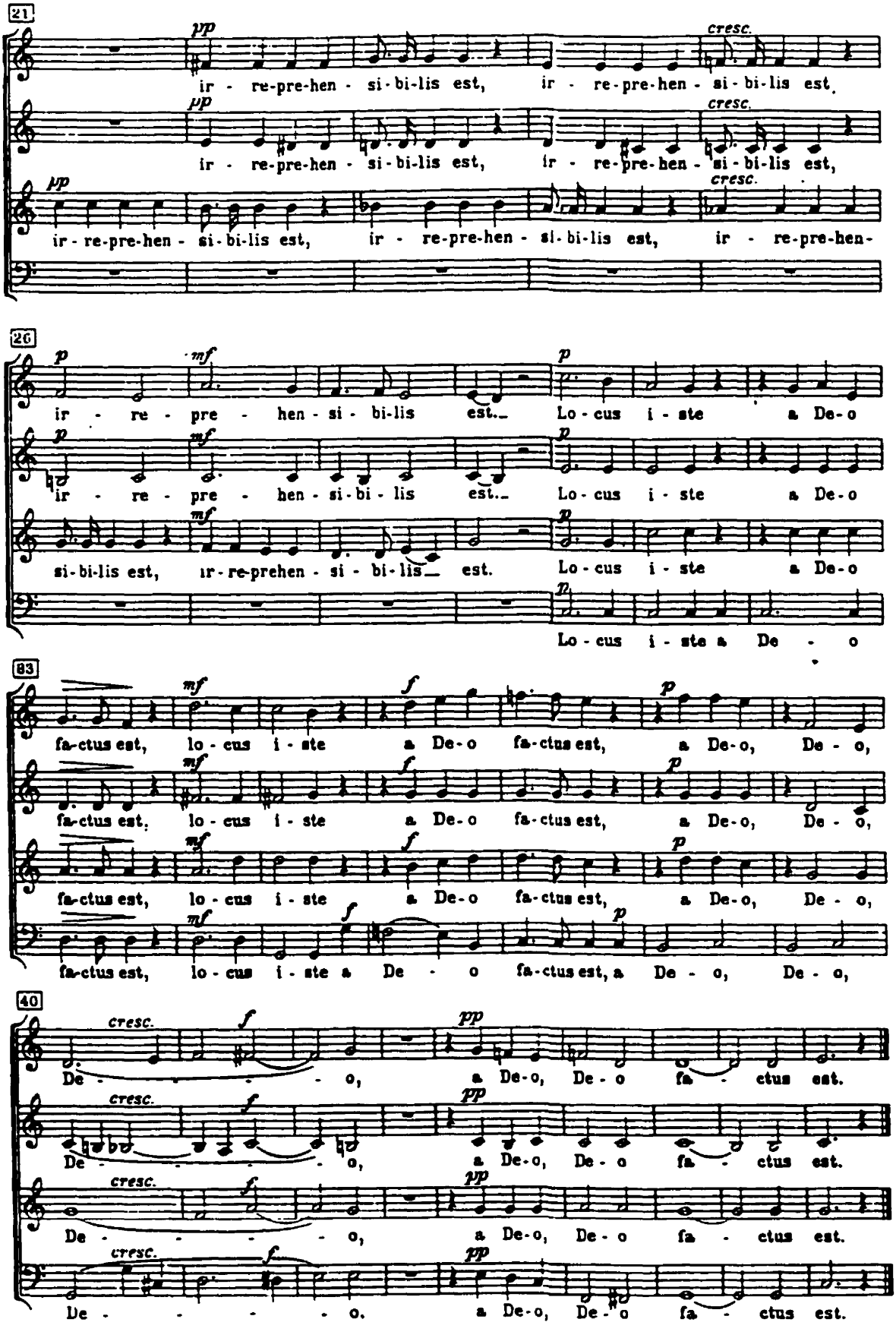



19
S
per - fo - ra - tum un - da flu - xit et san - gui - ne.
A
per - fo - ra - tum un - da flu - xit et san - gui - ne.
T
per - fo - ra - tum un - da flu - xit et san - gui - ne.
B
per - fo - ra - tum un - da flu - xit et san - gui - ne.
25
poco a poco cresc.
Es - to no - bis prae - gus - ta - tum in mor - tis ex -
Es - to no - bis prae - gus - ta - tum in mor - tis ex -
Es - to no - bis prae - gus - ta - tum in mor - tis ex -
Es - to no - bis prae - gus - ta - tum in mor - tis ex -
31
f
p
a - mi - ne, in mor - tis e - xa - mi - ne.
a - mi - ne, in mor - tis e - xa - mi - ne.
a - mi - ne, in mor - tis e - xa - mi - ne.
a - mi - ne, in mor - tis e - xa - mi - ne.

Gradual

Anton Bruckner (1824-1896)




21
pp
cresc.
ir - re-pre-hen - si-bi-lis est, ir - re-pre-hen - si-bi-lis est,
ir - re-pre-hen - si-bi-lis est, ir - re-pre-hen - si-bi-lis est,
ir - re-pre-hen - si-bi-lis est, ir - re-pre-hen - si-bi-lis est, ir - re-pre-hen-
26
p
mf
ir - re - pre - hen - si - bi-lis est. Lo - cus i - ste a De-o
ir - re - pre - hen - si - bi - lis est. Lo - cus i - ste a De-o
si-bi-lis est, ir-re-prehen - si - bi-lis est. Lo - cus i - ste a De-o
Lo - cus i - ste a De - o
83
f
fa-ctus est, lo - cus i - ste a De-o fa-ctus est, a De-o, De - o,
fa-ctus est, lo - cus i - ste a De-o fa-ctus est, a De-o, De - o,
fa-ctus est, lo - cus i - ste a De-o fa-ctus est, a De-o, De - o,
fa-ctus est, lo - cus i - ste a De - o fa-ctus est, a De - o, De - o,
40
cresc.
De - - - - - o, a De-o, De - o fa - ctus est.
De - - - - - o, a De-o, De - o fa - ctus est.
De - - - - - o, a De-o, De - o fa - ctus est.
De - - - - - o. a De-o, De - o fa - ctus est.

Paul Hindemith

(1895–1963)

UN CYGNE

from *SIX CHANSONS*

(1939)

Text: RAINER MARIA RILKE

A foe of atonality and serialism, Hindemith sought to create a musical language that was both contemporary and founded on tonal centers. Wherein does this chanson differ from a work in traditional E minor or E major? By what means is E projected as the tonic? Is the harmony triadic? How is dissonance regulated?

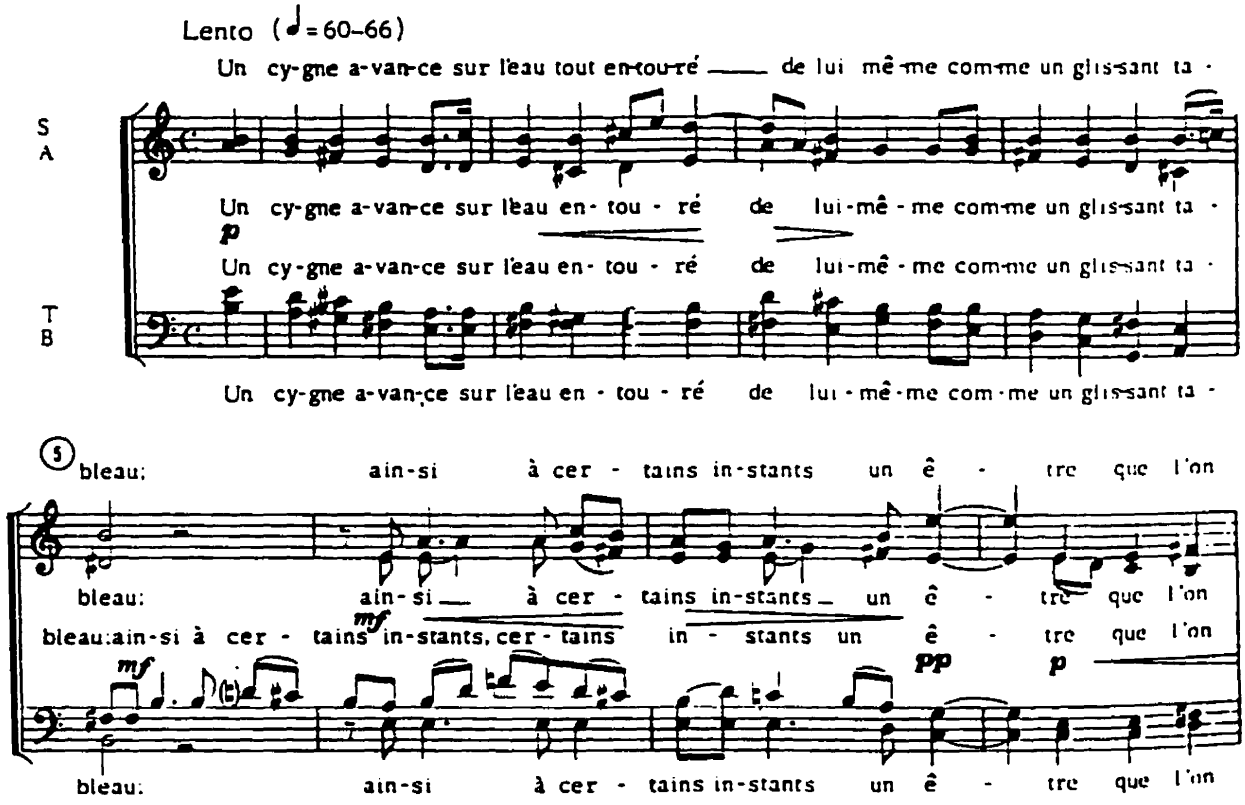
"A Swan" is the second of a set of six chansons for four-part unaccompanied chorus.

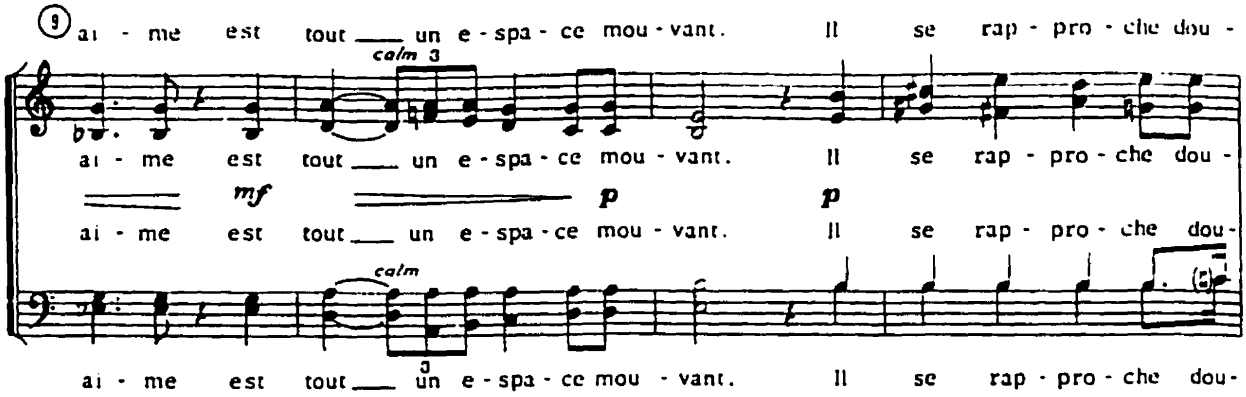
A swan advances on the water,
  Quite surrounded by himself,
Like a gliding picture.

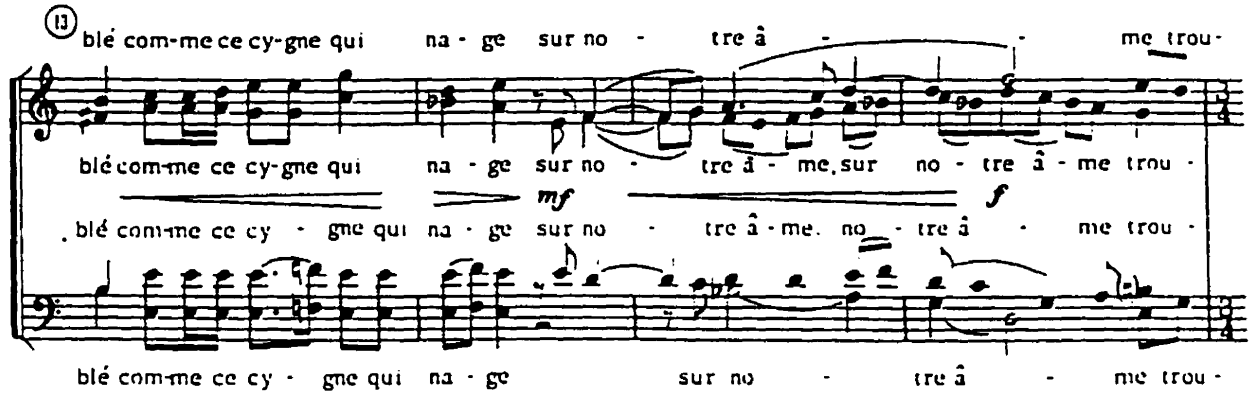
Thus, at certain moments,
  A being that one loves
Is [seen as] a moving space.

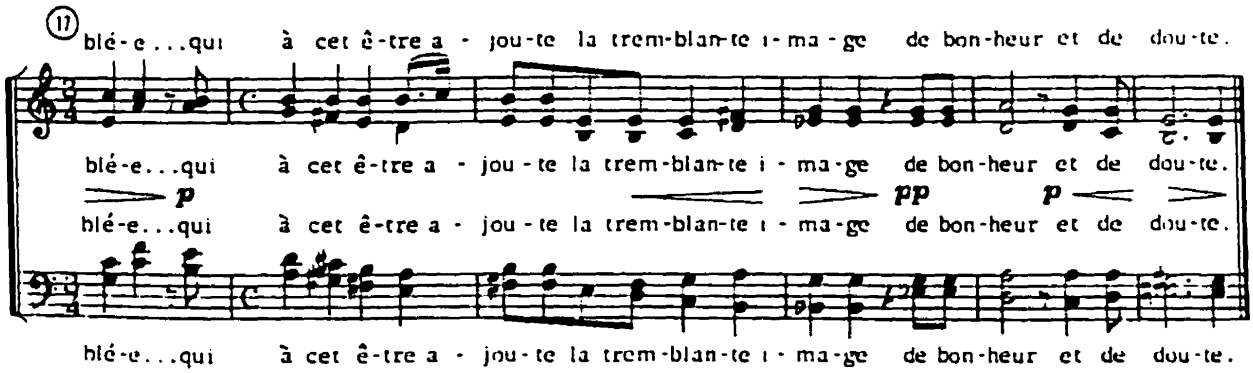
It draws near, doubled,
  Like this swan that swims
Before our troubled soul,

Which adds to that being
  The trembling image
Of happiness and doubt.




9
ai - me est tout un e - spa - ce mou - vant. Il se rap - pro - che dou -
calm 3
ai - me est tout un e - spa - ce mou - vant. Il se rap - pro - che dou -
mf p p
ai - me est tout un e - spa - ce mou - vant. Il se rap - pro - che dou -
calm
ai - me est tout un e - spa - ce mou - vant. Il se rap - pro - che dou -


13
blé com-me ce cy-gne qui na - ge sur no - tre â - me trou -
blé com-me ce cy-gne qui na - ge sur no - tre â - me, sur no - tre â - me trou -
mf f
blé com-me ce cy - gne qui na - ge sur no - tre â - me. no - tre â - me trou -
blé com-me ce cy - gne qui na - ge sur no - tre â - me trou -


17
blé - e . . . qui à cet ê - tre a - jou - te la trem-blan-te i - ma - ge de bon-heur et de dou-te.
blé-e...qui à cet ê-tre a - jou - te la trem-blan-te i - ma-ge de bon-heur et de dou-te.
p pp p
blé-e...qui à cet ê-tre a - jou - te la trem-blan-te i - ma-ge de bon-heur et de dou-te.
blé-e...qui à cet ê-tre a - jou - te la trem-blan-te i - ma-ge de bon-heur et de dou-te.

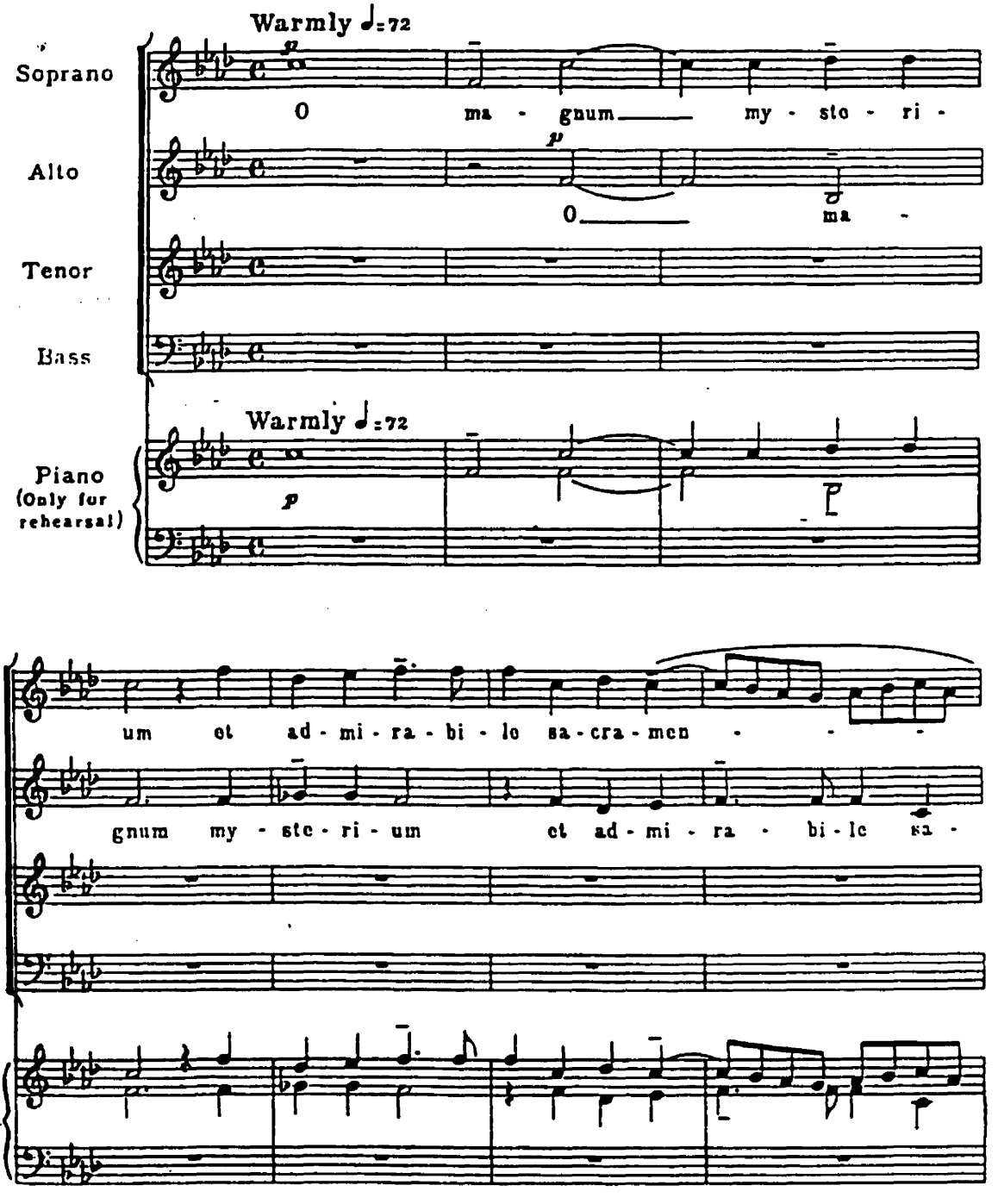
G. Schirmer, Octavo No. 10193

O Magnum Mysterium

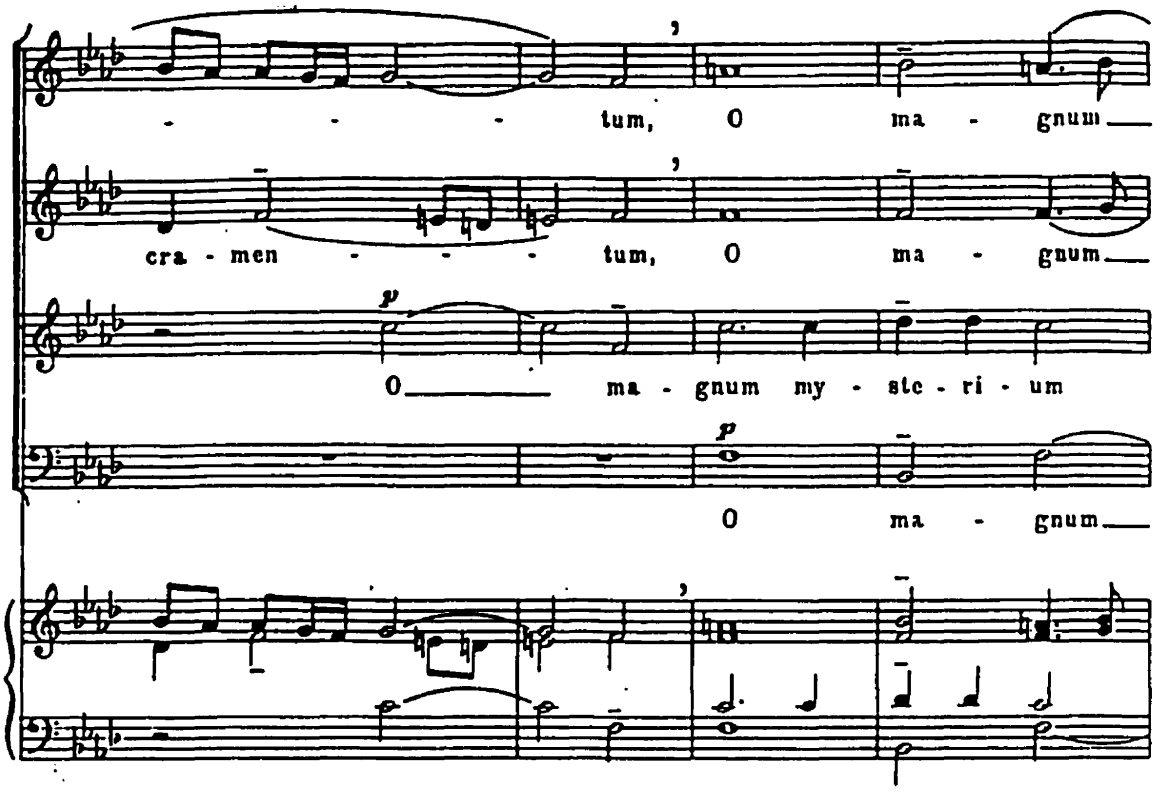
For Four-Part Chorus of Mixed Voices

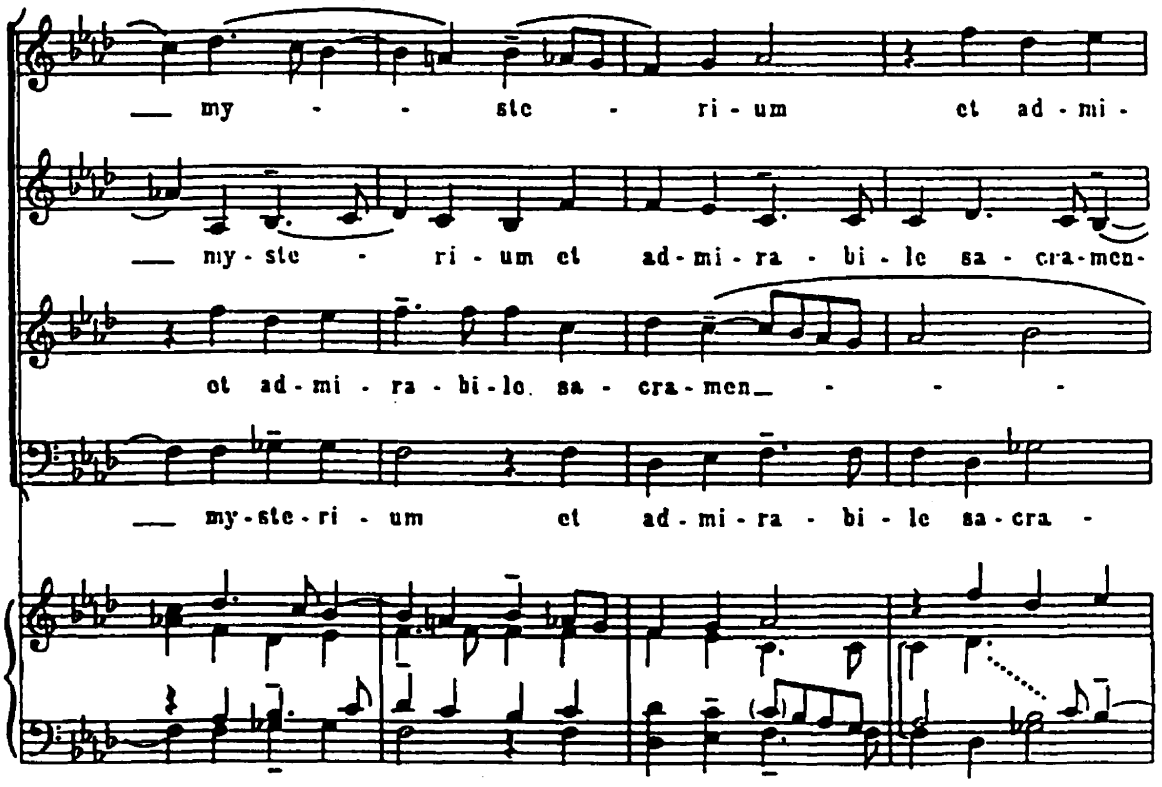
a cappella

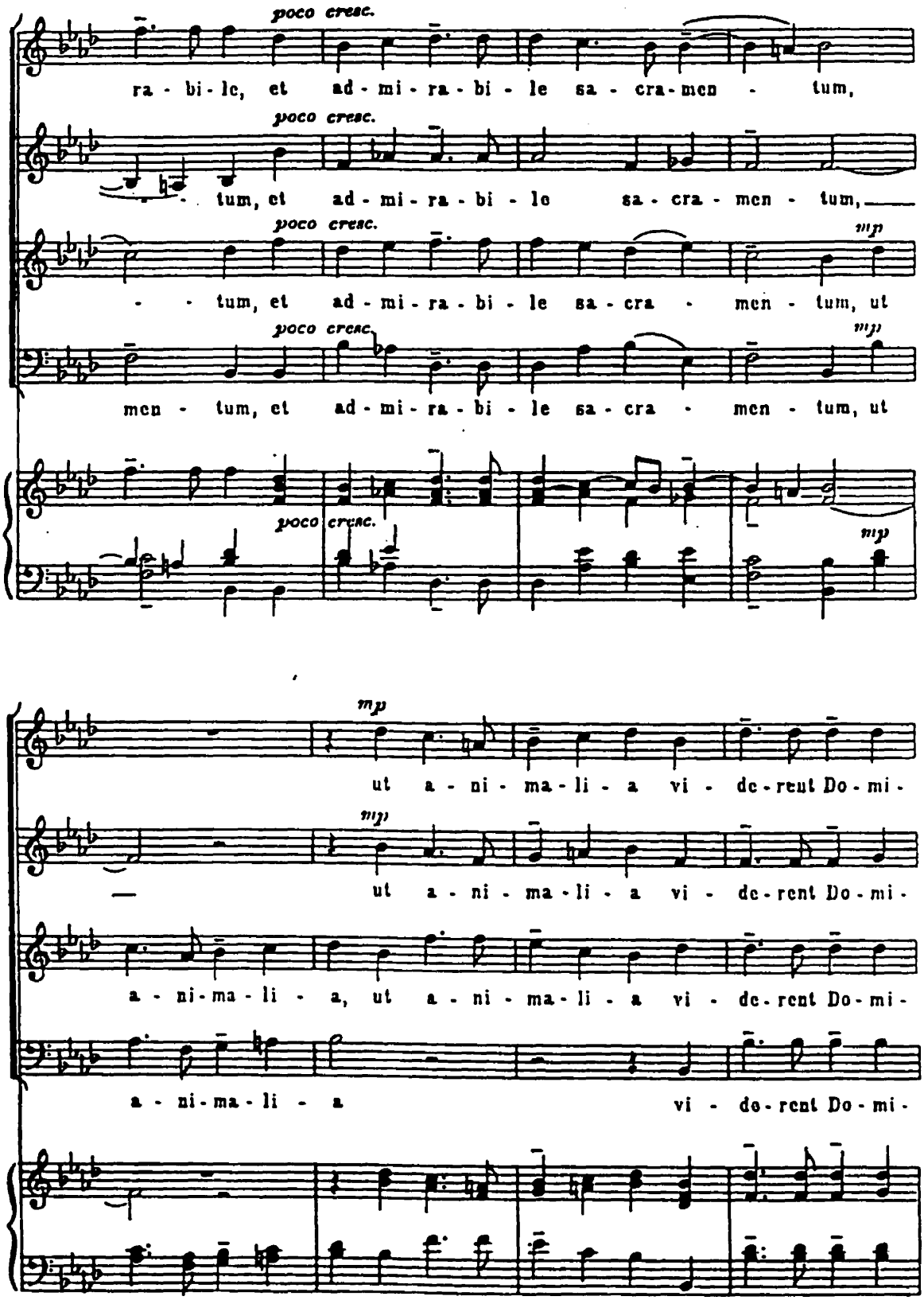
Tomás Luis de Victoria
Edited by Alice Parker
and Robert Shaw

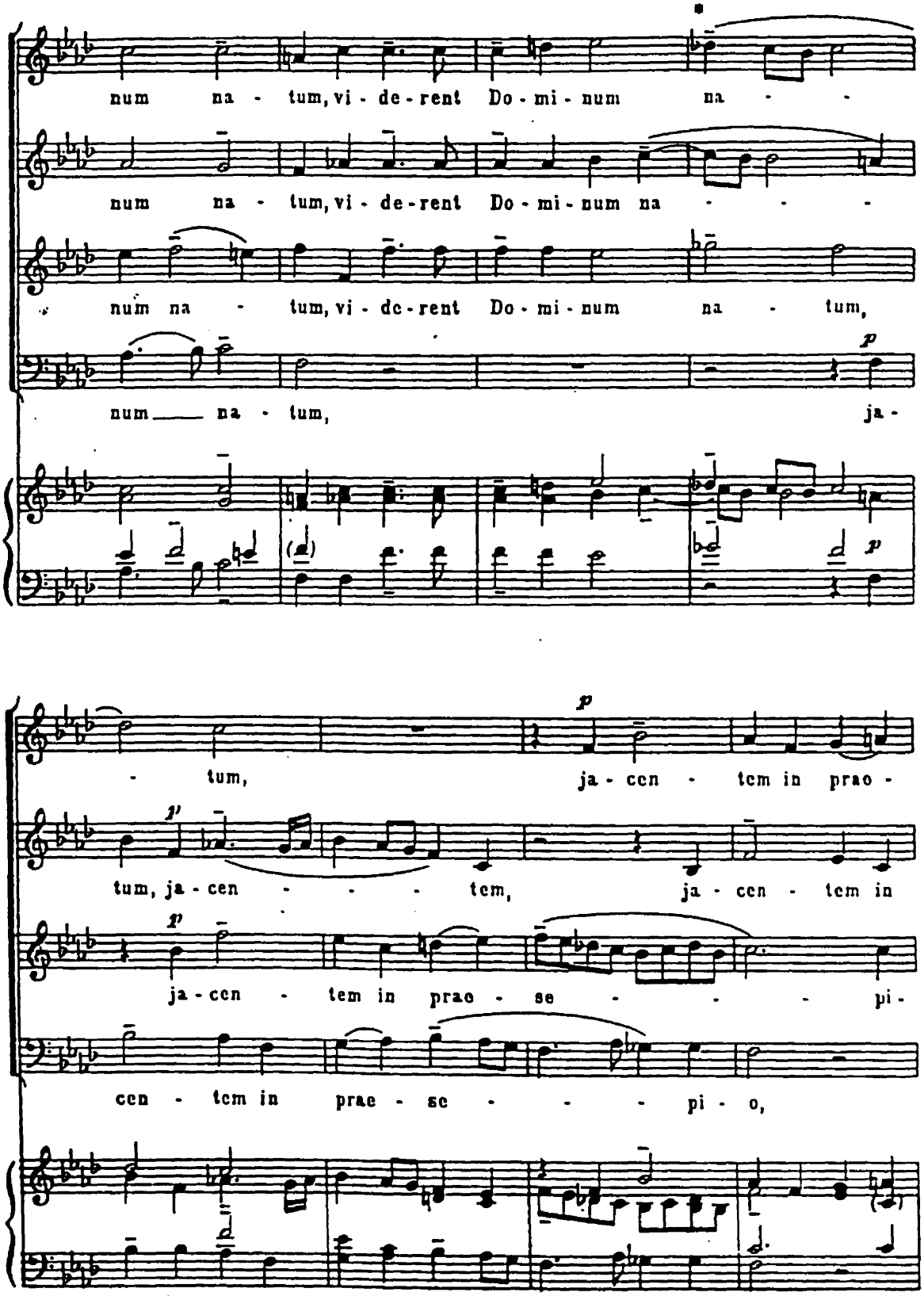


This motet has been recorded by the Robert Shaw Chorale in RCA Victor LM 1711, Volume II of *Christmas Hymns and Carols.*

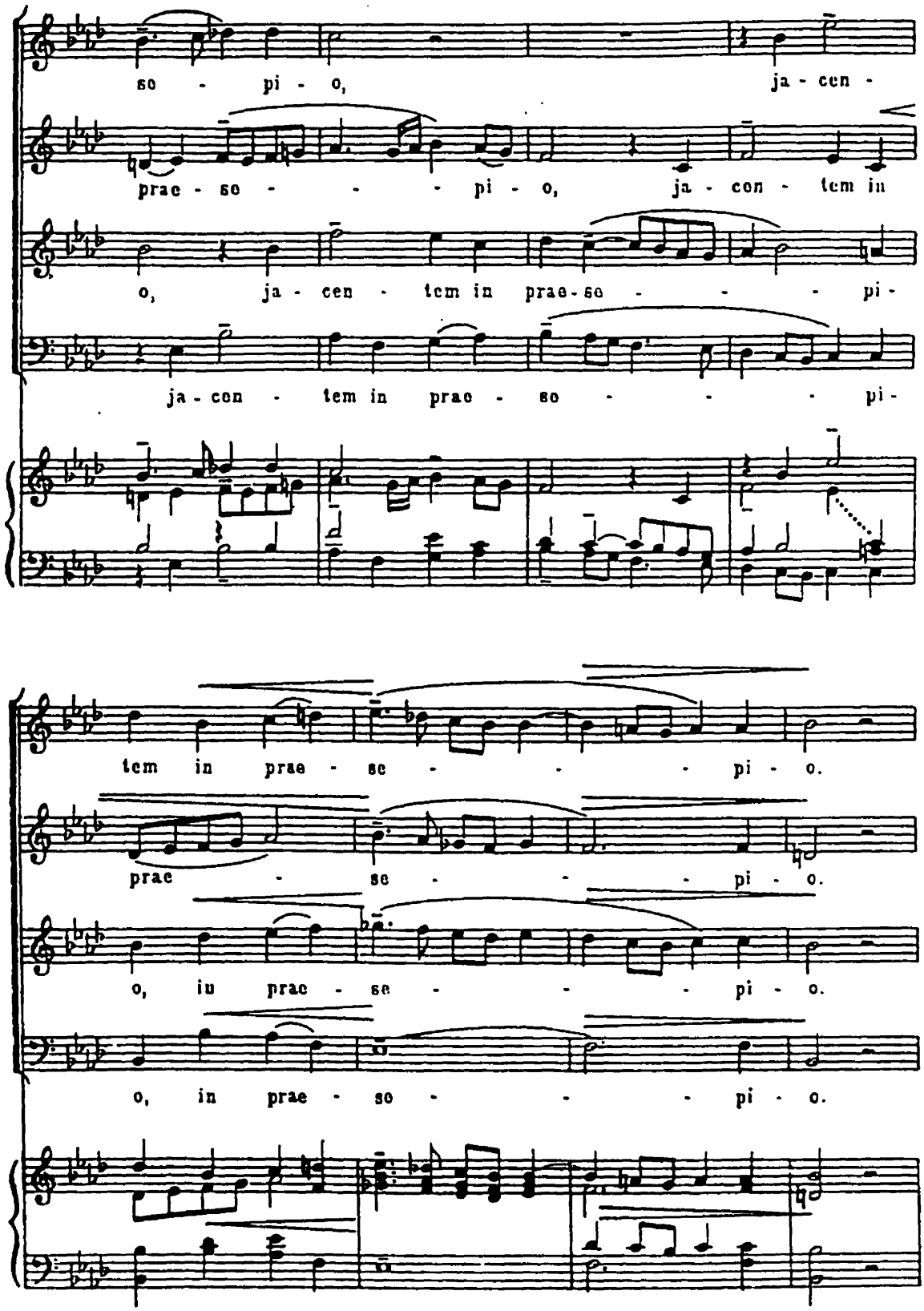

tum, O ma - gnum
cra - men tum, O ma - gnum
O ma - gnum my - ste - ri - um
O ma - gnum


my - ste - ri - um et ad - mi -
my - ste - ri - um et ad - mi - ra - bi - le sa - cra - men -
et ad - mi - ra - bi - le sa - cra - men
my - ste - ri - um et ad - mi - ra - bi - le sa - cra -


poco cresc.
ra - bi - le, et ad - mi - ra - bi - le sa - cra - men - tum,
poco cresc.
- tum, et ad - mi - ra - bi - le sa - cra - men - tum,
poco cresc.
mp
- tum, et ad - mi - ra - bi - le sa - cra - men - tum, ut
poco cresc.
mp
men - tum, et ad - mi - ra - bi - le sa - cra - men - tum, ut
poco cresc.
mp
mp
ut a - ni - ma - li - a vi - de - rent Do - mi -
mp
— ut a - ni - ma - li - a vi - de - rent Do - mi -
a - ni - ma - li - a, ut a - ni - ma - li - a vi - de - rent Do - mi -
a - ni - ma - li - a vi - de - rent Do - mi -



*In some editions the rhythm of the Soprano line is ♩. ♬ ♩

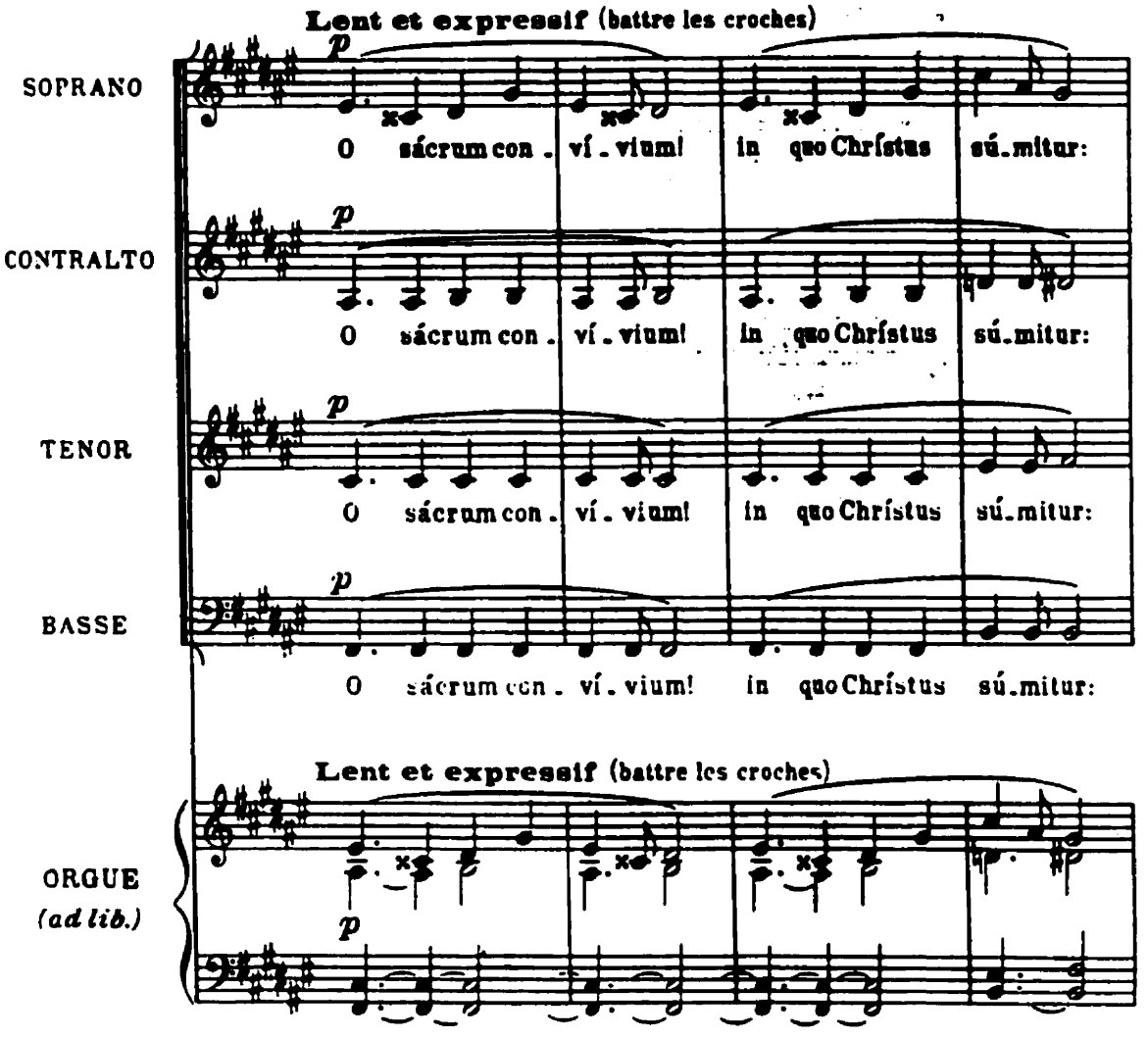

so - pi - o, ja - cen -
prae - so - - - pi - o, ja - con - tem in
o, ja - cen - tem in prae - so - - - pi -
ja - con - tem in prae - so - - - pi -
tem in prae - se - - - pi - o.
prae - se - - - pi - o.
o, in prae - se - - - pi - o.
o, in prae - so - - - pi - o.

O sacrum convivium!

motet au Saint-Sacrement

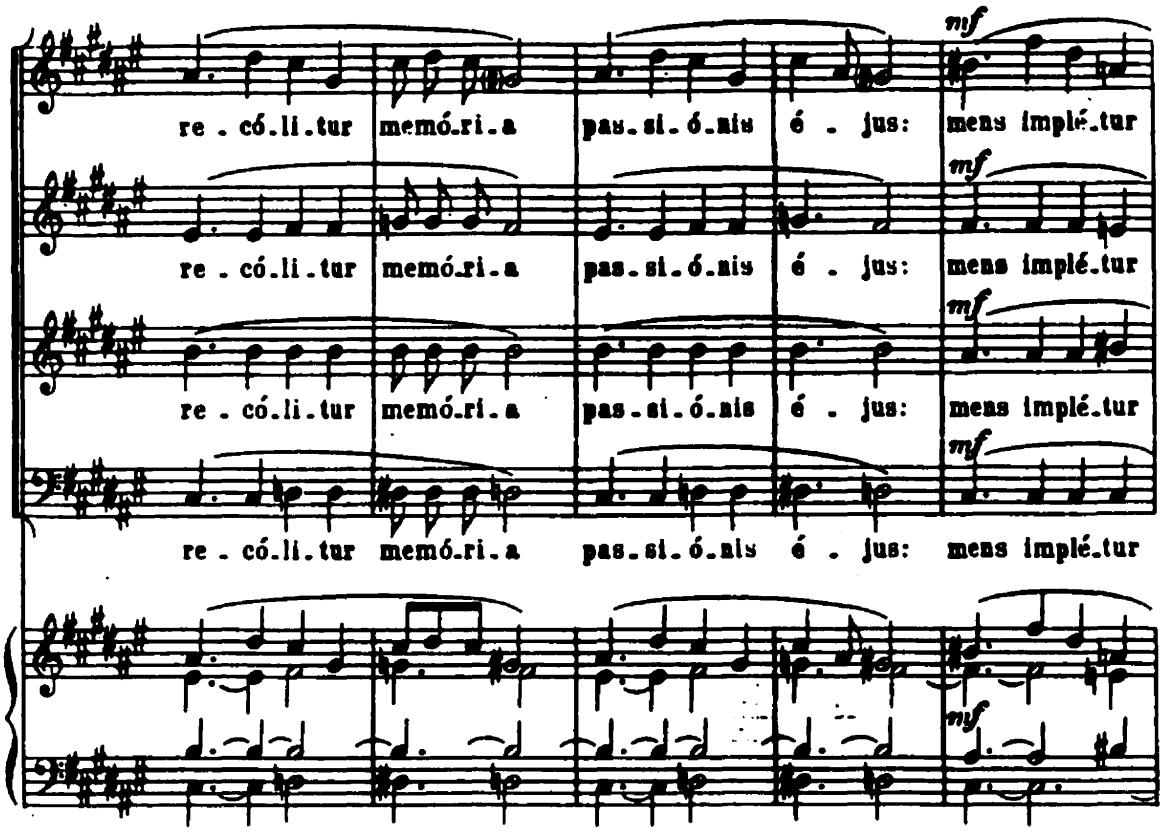
pour chœur à quatre voix mixtes
ou quatre solistes
(avec accompagnement d'orgue *ad libitum*)

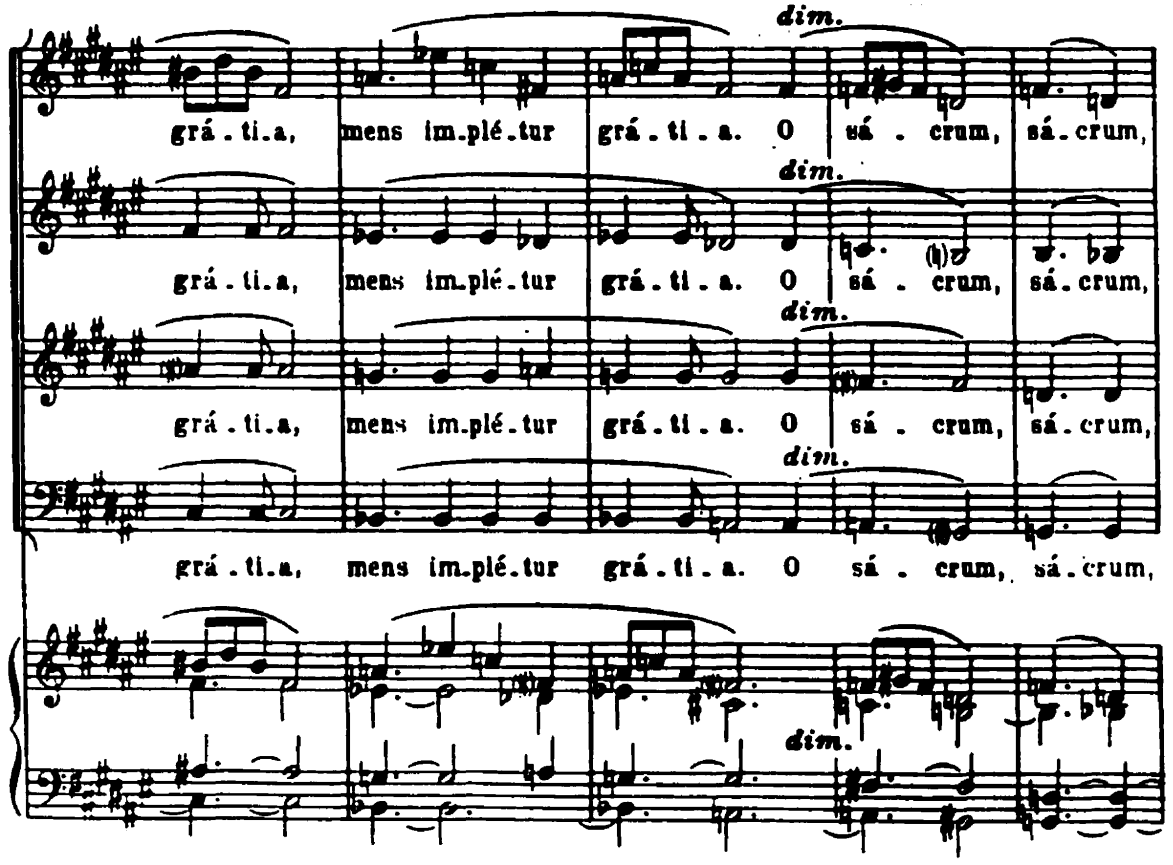
Olivier MESSIAEN

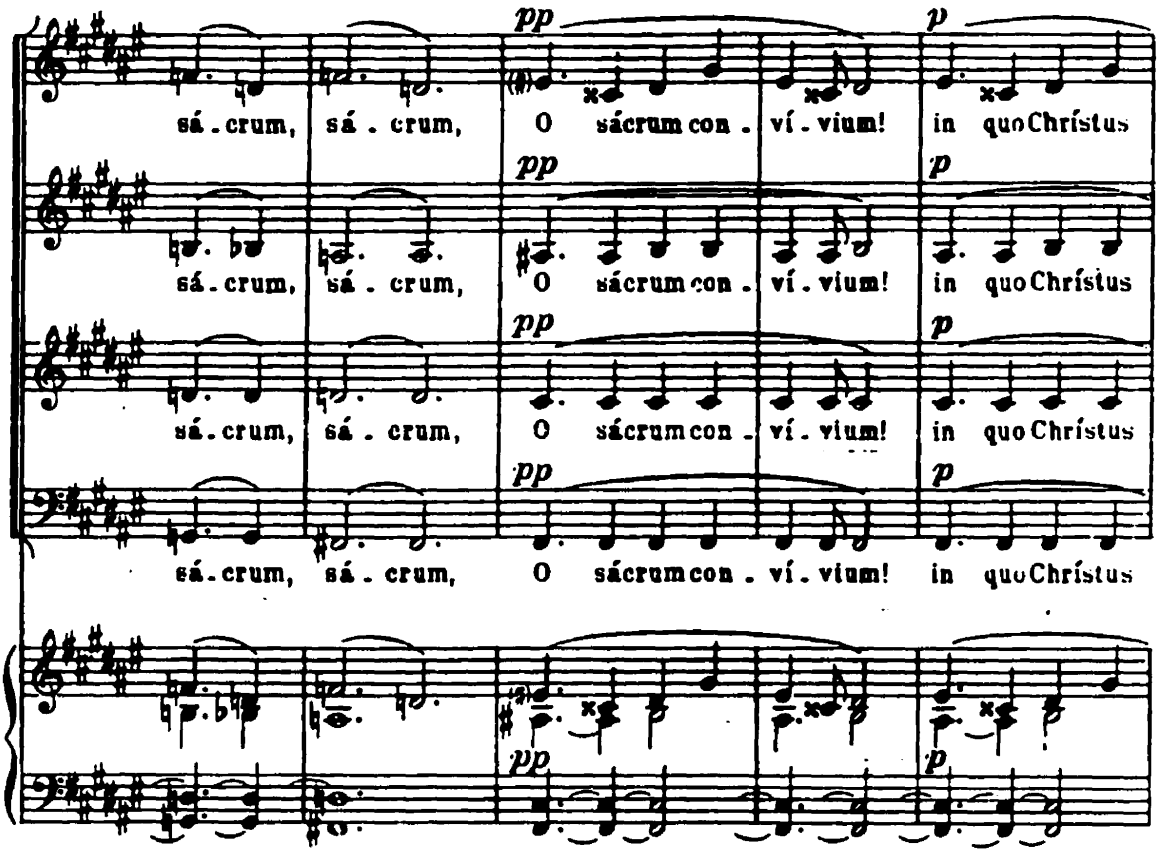


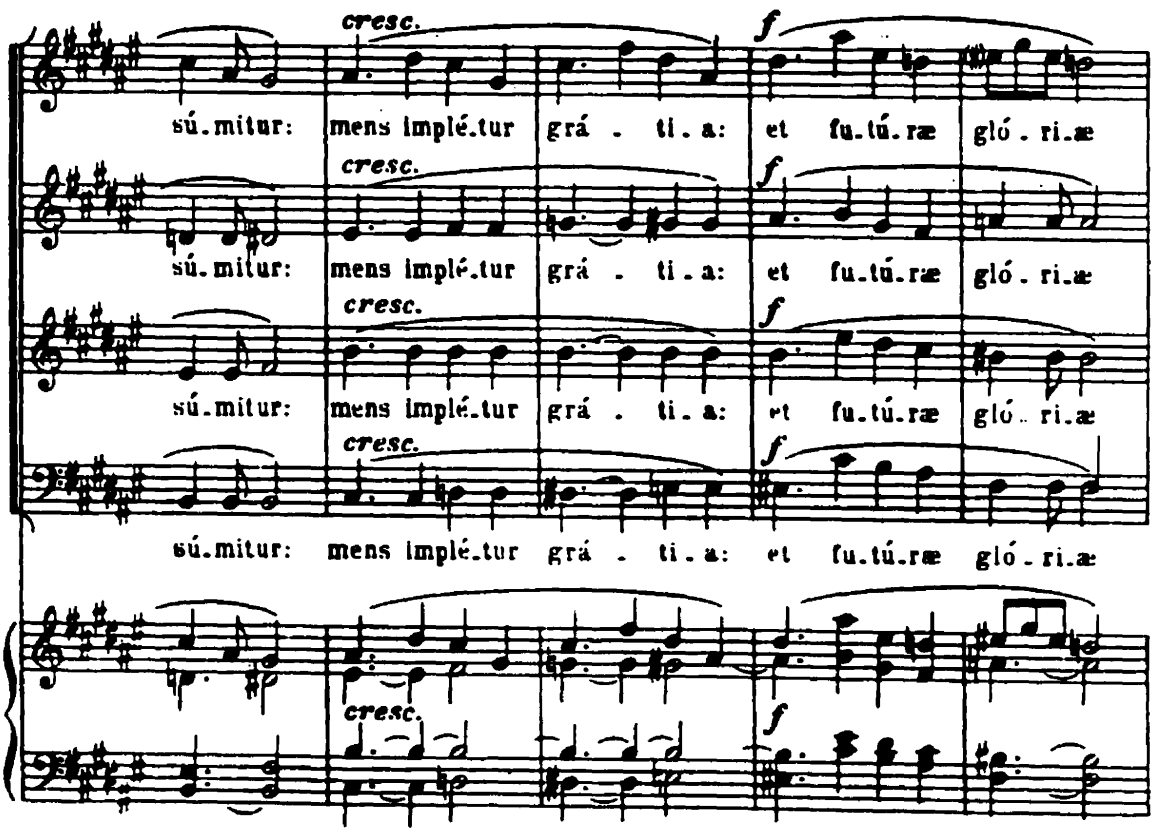
D. & F. 12.742

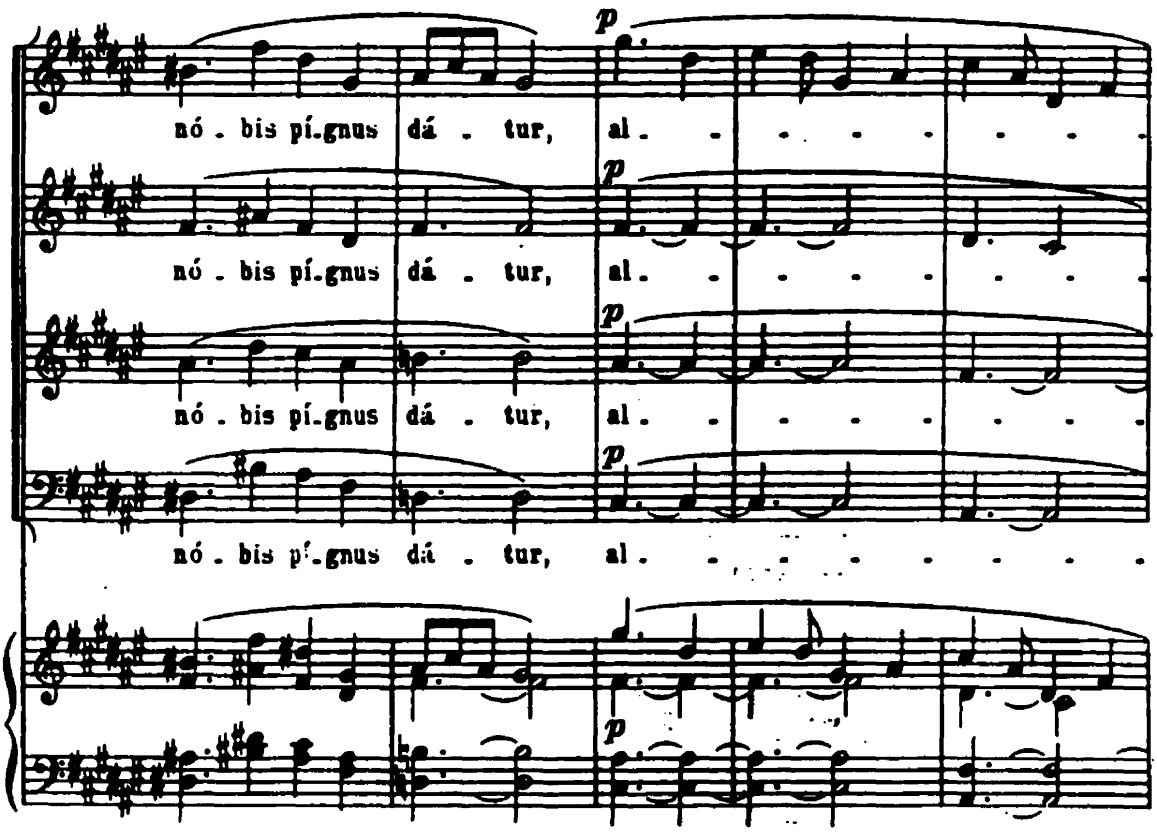
DURAND Editeurs Musicales
215, RUE DU Fbg ST HONORÉ 75008 PARIS

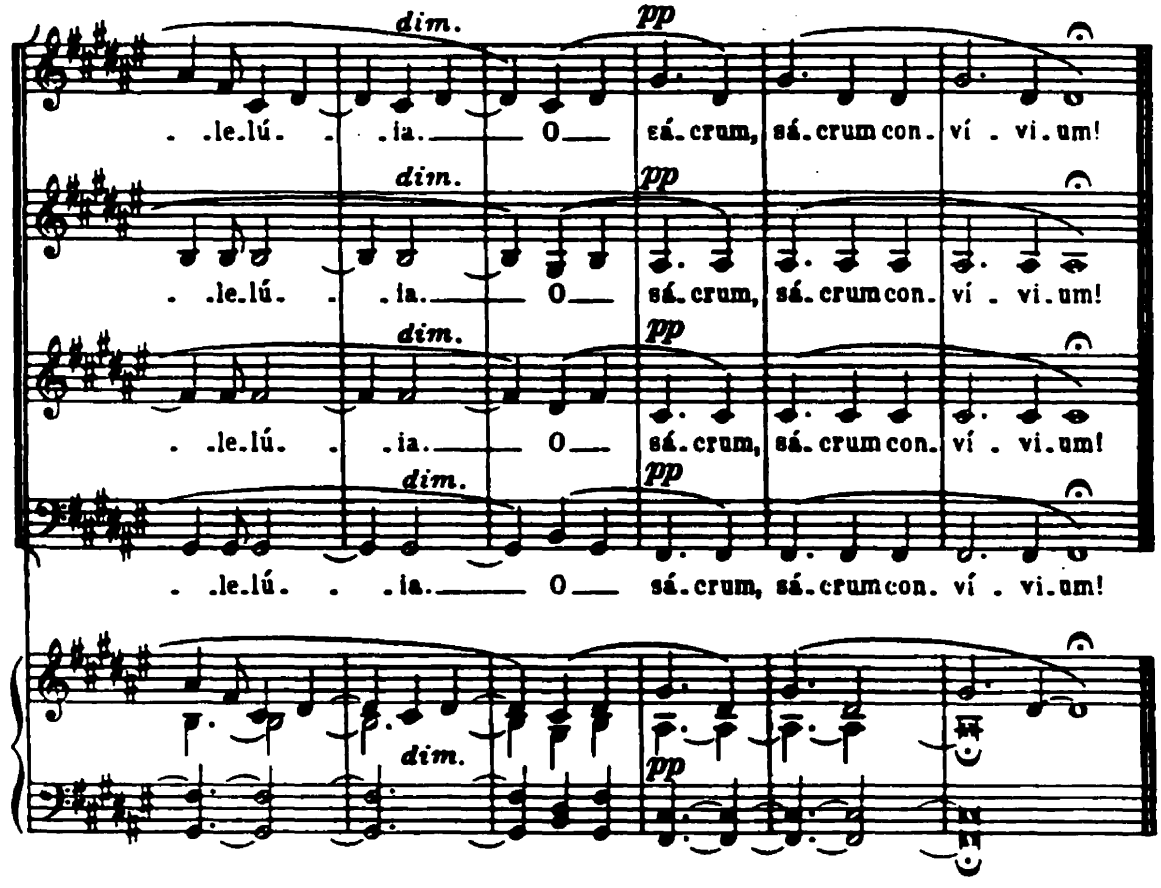

mf
re . có.li.tur memó.ri.a pas.si.ó.nis é . jus: mens implé.tur
mf
re . có.li.tur memó.ri.a pas.si.ó.nis é . jus: mens implé.tur
mf
re . có.li.tur memó.ri.a pas.si.ó.nis é . jus: mens implé.tur
mf
re . có.li.tur memó.ri.a pas.si.ó.nis é . jus: mens implé.tur
mf


dim.
grá . ti.a, mens im.plé.tur grá . ti . a. O sá . crum, sá.crum,
dim.
grá . ti.a, mens im.plé.tur grá . ti . a. O sá . crum, sá.crum,
dim.
grá . ti.a, mens im.plé.tur grá . ti . a. O sá . crum, sá.crum,
dim.
grá . ti.a, mens im.plé.tur grá . ti . a. O sá . crum, sá.crum,
dim.


pp
p
sá . crum, sá . crum, O sácrum con . ví . vium! in quo Chrístus


cresc.
f
sú . mitur: mens implé . tur grá . ti . a: et fu . tú . ræ gló . ri . æ


p
nó . bis pí.gnus dá . tur, al . . . . . .
p
nó . bis pí.gnus dá . tur, al . . . . . .
p
nó . bis pí.gnus dá . tur, al . . . . . .
p
p


dim.
pp
. .le.lú. . .ia. 0 sá.crum, sá.crum con. ví . vi.um!
dim.
pp
. .le.lú. . .ia. 0 sá.crum, sá.crum con. ví . vi.um!
dim.
pp
. .le.lú. . .ia. 0 sá.crum, sá.crum con. ví . vi.um!
dim.
pp
. .le.lú. . .ia. 0 sá.crum, sá.crum con. ví . vi.um!
dim.
pp