

# Can experts identify “open throat” technique as a perceptual phenomenon?

Helen F. Mitchell and Dianna T. Kenny

From the Australian Centre for Applied Research in Music Performance (ACARMP), Sydney Conservatorium of Music, The University of Sydney, New South Wales 2006, Australia

Received 25 July 2004. Accepted 7 February 2005.

---

*Musicae Scientiae*; 2006

Singing pedagogues have identified ‘open throat’ as a widely used pedagogical tool in the singing studio and a critical technique to achieve good classical vocal quality. This study is the first to assess the perceptual validity of open throat. Fifteen expert singing pedagogues assessed 48 messa di voce and 24 song samples with six repeats of six advanced singing students under two conditions: ‘optimal’ (O), representing use of maximal open throat technique and ‘suboptimal’ (SO), representing reduced open throat. Correctly identified responses were counted by condition (O/SO), by judge and by singer. Data were analyzed using Cohen’s Kappa. Hypotheses that correct identification would be greater than chance were confirmed for both messa di voce and the song samples, with thirteen of the fifteen judges correctly identifying 82.7% of song samples as O or SO. Singers’ self-ratings attributed their best singing to maximal use of open throat technique. These results indicate that listeners are consistent in making a dichotomous choice in identifying use of a singing technique through auditory-perceptual evaluation. The findings from this study suggest that there is a sound quality associated with the use of open throat technique, which is a perceptual reality to singing pedagogues and singers and that the specific vocal quality in classical singing that it produces can be reliably identified by expert listeners. If singers and expert listeners’ of singing accord, focussing future research on specific singing techniques could enhance singing pedagogy.

*Helen Mitchell, Australian Centre for Applied Research in Music Performance (ACARMP), The Conservatorium of Music, The University of Sydney, New South Wales, Australia 2006 Phone: 61 2 9351 9644; Fax: 61 2 9351 9540 Email: d.kenny@fhs.usyd.edu.au*

---

## INTRODUCTION

In this paper, we report the first study to assess the perceptual validity of a specific singing technique – open throat (OT) – as taught in today’s singing studio. Previous perceptual studies (Ekholm, Pappagiannis, & Chagnon, 1998; Mendes, Rothman, Sapienza, & Brown, 2003; Robison, Bounous, & Bailey, 1994; Vurma & Ross, 2000; Wapnick & Ekholm, 1997) have demonstrated that singing improves over time and through training but have not identified specific components of the training that produce improvements in vocal quality and thus add little to the singing pedagogical literature. In order to evaluate the contribution of one singing technique (OT), we have assessed the acoustic aspects of voice quality produced with and without this technique and the perceptual impressions of listeners, firstly their capacity to identify voices that used the technique and then to describe the vocal quality produced.

### *Open Throat Technique (OT)*

The term “open throat” occurs frequently in the

vocal pedagogy literature (Burgin, 1973; Fields, 1947; Monahan, 1978). It is defined as a complex process that is both a pedagogical instruction and a perceived sensation or action that results in a specific sound quality. Current support for the use of OT in singing technique is widespread (Miller, 1996; Reid, 1975, 1983). When OT is used, pedagogues perceive the sound as resonant (Miller, 1996; Vennard, 1968), round (Joiner, 1998), free (Ware, 1998), pure (Marafioti, 1981) rich and warm (McKinney, 1982). The sound quality is attributed to freedom from ‘constrictor tensions’ (Reid, 1983, p. 83). In the pedagogical literature, the sound quality attributed to open throat is linked to the preparation to sing, on inhalation (Hemsley, 1998; Manèn, 1987; Miller, 1997a, 1997b) through the surprise breath or smelling the rose imagery (Hemsley, 1998; Miller, 1996; Puritz, 1956) and visualizing space within the throat, through an ‘air-ball’ or ‘soap bubble’ (Herbert-Caesari, 1951; Manèn, 1987) to achieve the posture of OT.

The technique of OT is a pedagogical concept transmitted through the oral tradition of singing. Mitchell, Kenny, Ryan & Davis (2003) assessed

current pedagogical practices and perceptions of OT. They noted that expert singing pedagogues identified OT as a critical technique in vocal production of classical singing. Despite the use of different terminology to describe the technique (such as freedom, collar and retraction), OT was defined by these pedagogues as a technique to maximize pharyngeal space and/or abduct the ventricular folds. There was consistency in the vocal instructions to achieve it. Pedagogues teach conscious control of OT using laugh, sob, correct inhalation or maintaining the posture of inhalation. They consistently reported a perceptually distinctive sound quality as a result of mastery of this technique and described their perception of the final sound quality as balanced, coordinated, free, even, consistent, warm and open.

Hypotheses generated from pedagogical beliefs and practices derived in the first study on OT were subsequently tested acoustically by Mitchell and Kenny (2004b) who compared the use of open throat technique (optimal, O) to a reduction of the technique (sub-optimal, SO) using female classical singers. Reduced OT resulted in significant reductions in vibrato extent and increases in vibrato onset time but no change to vibrato rate between the two conditions. These singers relied on OT to produce acceptable vibrato parameters. However, according to the literature, inappropriate vibrato is indicative of poor singing in general; therefore further acoustic tests were required to test the differences in timbre. Visual inspection of long term average spectra (LTAS) confirmed differences between O and SO for each singer, and the O condition produced a rounder peak between 0-2 kHz indicating a warmer sound quality compared to SO. Despite these findings, LTAS did not demonstrate significant differences between SO and O in conventional measures performed on long term average spectra energy peak height [singing power ratio (Omori, Kacker, Carroll, Riley, & Blaugrund, 1996)] and energy peak area [energy ratio (Thorpe, Cala, Chapman, & Davis, 2001)].

Technology to measure voice acoustically has become more sophisticated and is used increasingly in experimental voice research, the findings of which are now being incorporated into texts on singing (Miller, 1996; Nair, 1999; Sundberg, 1977; Thurman & Welch, 2000). To date, research into the singing voice has described acoustic properties of voice and its visual representation with few or no links to pedagogical descriptors or perceptual judgments. The implications of acoustic studies of the singing voice within singing lit-

erature and pedagogy are difficult to ascertain, since most studies do not assess acoustically or perceptually the vocal strategies used in the singing studio (Foulds-Elliott, Thorpe, Cala, & Davis, 2000; Thorpe et al., 2001). Indeed, few studies link acoustic findings with perceptual judgments (Ekholm et al., 1998) or discuss their pedagogical implications (Callaghan, 2000; Miller, 1998).

This study seeks to provide stronger links between empirical research and singing pedagogy, and recognises the importance of singing pedagogues in this enterprise.

### *Perceptual Studies*

Perceptual evaluation of musical performance is important for a number of reasons. Opera singers, their teachers and audiences are primarily interested in how the voice sounds and use the sensory information derived from listening as the primary basis for determining the quality of the sound (Kitch & Oates, 1994). Despite recent advances in the physiological and acoustic measurement of voice, "...voice quality is fundamentally perceptual in nature" (Kreiman, Gerratt, Kempster, Erman, & Berke, 1993, p. 21). The auditory-perceptual aspects of a voice represent a psychological reality for both the singer and the listener and have the potential to provide a common terminology for ease of communication between singers, their teachers, and clinicians. Development of a reliable and valid method of communicating one's perception has significant implications. For example, audience appreciation ensures continuing engagement of performers; expert listeners adjudicate examinations, auditions and other competitions, acts which may seal the fate of performing artists. Because such evaluations can have significant impacts on musical careers, it is important to establish whether listeners are reliable in their judgments, both in comparison to other listeners and across time.

Research indicates that listeners show some degree of reliability and consistency in their perceptual judgments of timbre, between vocal genres (from opera to music theatre) (Sundberg, Gramming, & Lovetri, 1993), and between good and poor vocal and instrumental performance (Ekholm et al., 1998; Geringer & Madsen, 1998; Saunders & Holahan, 1997; Wapnick, Flowers, Alegant, & Jasinkas, 1993), in the assessment of excellence in overall voice quality (Wapnick & Ekholm, 1997) and in rankings and ratings of performers in competitive situations (Davidson & Da Costa Coimbra, 2001).

Recent evidence indicates that judges initially assess the performance as a whole before 'justifying' their ranking by assessing specific criteria (Stanley, Brooker, & Gilbert, 2002). Judges do acknowledge, however, that providing them with specific criteria focuses their assessment on certain elements of performance and that this provides the basis for useful, specific feedback to students.

Perceptual studies of vocal quality generally require listeners to focus on specific vocal dimensions, rather than, or as well as, making an overall judgement. Focusing listeners by using a number of rating criteria improves the consistency of judges' responses (Wapnick et al., 1993). However, analysis of available rating scales that request listeners to make several judgments of vocal quality reveal that these specific dimensions may all be tapping into a single underlying construct, that of vocal quality, thereby rendering individual assessments on each dimension at least partially redundant. For example, a number of researchers (Ekholm et al., 1998; Robison et al., 1994; Wapnick & Ekholm, 1997) have found very high correlations between all dimensions of voice quality studied, (for example, "appropriate vibrato", "resonance/ring", "color/warmth", and "clarity/focus"). All of these dimensions were found to converge with the overall judgement of vocal quality.

There are, of course, other limits to the accuracy of perceptual judgements. Perceptual studies produce an unnatural assessment setting for teachers and assessors of singing. Listeners have commented on their difficulty in assessing singing when the duration of the sample is insufficient (Ekholm et al., 1998). In forced choice situations in which listeners are asked to make dichotomous choices, listeners can reliably make gross judgments such as identifying girl from boy singers (Howard, Szymanski, & Welch, 2002). However, listeners could not identify singers from speakers during speaking rather than singing tasks (Brown, Rothman, Morris, & Sapienza, 2001; Brown, Rothman, & Sapienza, 2000), or the same singers singing at different pitches throughout their singing range (Erickson, 2003; Erickson & Perry, 2003). This lack of identification may have been due to the variability in the stimuli rather than a lack of sensitivity in the listeners, and these issues are addressed in the design of this study.

#### *Perceptual identification of open throat*

To date, there has been no research that assesses the sound qualities associated with particular vocal techniques. In Western classical singing, the

focus remains on the overall good sound quality (Davidson & Da Costa Coimbra, 2001) whereas pedagogical literature focuses on mastery of a wide variety of techniques to achieve this final sound (Miller, 1998; Stark, 1999).

In this study, we investigated the perceptual significance of 'open throat' technique by comparing the same female classical singers when they consciously used maximum open throat technique and reduced open throat technique in their singing of messa di voce (a crescendo-diminuendo on a single note of long duration), a classical song and a romantic lied. The aim of the study was to assess expert pedagogues' perception of OT technique in singing, to explore its acoustic correlates, and to assess the degree of agreement between perceptual and acoustic factors.

## METHOD

### *Participants*

#### Listeners

Listeners were 15 experienced singing pedagogues, 12 females and 3 males aged between 37 and 76 with a mean of 54 years. Six participants had a postgraduate qualification in singing, five had a diploma of music or singing, and three a bachelor degree in music. One cited extensive international performing experience as their qualification in singing. All had taught singing for 4 to 40 years, with an average for 20 years. Thirteen of fifteen participants taught singers in a Conservatorium of Music. Overall, participants' singing studios comprised an average of 39.5% of operatic students and 36.7% classical students. For 11 pedagogues, the majority of their studio comprised these two genres. Six pedagogues also taught  $\geq 20\%$  of musical theatre students in their studio. Eleven taught a proportion of international and national level singers, nine at big city or regional/touring and eight at local community level. Participants were either known to the researchers via affiliations with key music centres in Australia or volunteered in response to an advertisement in a national singing organization newsletter. Participants were sent information about the project and were invited to take part in a perceptual study of singing technique. They were required to participate in a single listening session at a time and location convenient to them.

Prior to commencement, participants completed a questionnaire regarding their musical and teach-

ing experience and current singing studio. Each participant was asked for demographic information and to classify their singing studio according to their students' performing levels and singing styles (Bunch & Chapman, 2000).

### Singers

Six female singers, three sopranos and three mezzo-sopranos volunteered to participate in this project. They were advanced students with excellent technique studying with an experienced singing pedagogue, a Lecturer in Vocal Studies and Opera at a state Conservatorium of Music in Australia. This institution is considered the premier institution for musical education in the country, having produced several singers of international repute. Criteria for participant selection included, through this pedagogue's assessment, singers who: (1) had a good classical singing technique for their level of training and experience; and (2) understood and demonstrated skilful control of 'open throat' or 'retraction' techniques in their singing. Singers were sent information about the project and were required to attend a single recording session. They were told that the object of the study was to investigate acoustical and perceptual features of the use of open throat in singing.

Prior to the voice recording, participants completed a questionnaire. Singer participants were aged between 23 and 30, with a mean of 26 years. All had studied singing for at least 7 years (average 9.8 years) and had spent an average of 5 years studying with their present singing teacher. Each singer held a qualification in singing or music (four had Bachelor of Music degrees and two had diplomas, in music and/or singing) and 5 of 6 were currently undertaking a second degree in singing (3 postgraduate Diploma of Opera and 2 Bachelor of Music degrees). All defined the majority of their singing as operatic (>50%), with the second most common style classical (>20%) (Bunch & Chapman, 2000). All reported that they were in good health and able to perform the tasks.

### Singer Protocol

#### The Musical Tasks

Three musical tasks were chosen in order to test different demands of good singing, but were not musically difficult. They were designed to test the use of open throat, and contained musical features derived from a previous qualitative study on the use of the technique (Mitchell et al., 2003), where use or lack of the technique was deemed to be particularly valuable or noticeable. These features



Figure 1a: First 7 bars used as perceptual stimuli of Mozart song *Ridente la Calma*, K 152.

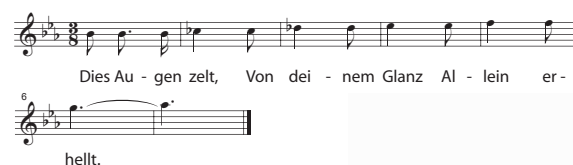


Figure 1b: Bars 54 to 60 used as perceptual stimuli of the Schubert lied, *Du bist die Ruh* D. 776 (Op. 59, No. 3).

were: high tessitura, sustained or legato singing, dynamic range control, and vocal agility.

The messa di voce, a crescendo-diminuendo on a single note is a tool for accomplishing and maintaining uniform timbre throughout the singing range (Miller, 1996). The singers sang three messa di voce across their passaggio beginning with pitches determined by their voice type, soprano or mezzo soprano (A4, G4, G-flat4).

The Mozart song *Ridente la Calma*, K 152, bars 1-27 (Figure 1a) was selected as it is a nominally simple song in the Italian language (Wapnick & Ekholm, 1997) with a mixture of common musical statements involving repeated legato lines as well as the initial stylised leaps of a major 4, and short scale figures. All 6 singers sang this aria in the same key (F major).

The third verse of the Schubert lied, *Du bist die Ruh* D. 776 (Op. 59, No. 3) (Figure 1b), bars 54 to 80 was chosen for its demanding vocal control, sustained musical line and high climactic tessitura. The three sopranos and 3 mezzo-sopranos sang this in an appropriate key depending on voice type (E-flat, D-flat and C major).

#### Experimental conditions

Singers sang each of the two song excerpts under three conditions: optimal, sub-optimal and loud sub-optimal. 'Optimal' (O) provided the best possible sound quality the singer could produce using their best open throat technique. 'Sub-optimal' (SO) condition involved the use of a reduced (open throat) technique but still with an acceptable singing technique and without consciously altering any other aspect of their technique. The loud

sub-optimal condition provided additional acoustic information discussed in earlier papers (Kenny & Mitchell, in press; Mitchell & Kenny, 2004a, 2004b).

Each task was performed twice in the O and SO conditions. On completion of each O and SO condition in each task, participants were asked to self-rate their singing on a 10 cm visual analogue scale (VAS) for 'openness' and 'my best singing'. Each singer completed six of these descriptive self-rating scales.

#### *Reliability check*

A pedagogue was present during the recording sessions to provide accompaniment for warm-ups and practice of the tasks where necessary. Prior to the commencement of the study, the students and the pedagogue had several practice sessions in which the singer was instructed to use either O or SO at random. The pedagogue indicated which version she thought the singer had used. Practice with each singer continued until both the singer and the pedagogue reached 100% agreement on the occasions that the technique was applied or reduced.

#### *Recording*

The voice was recorded using a high-quality microphone (AKG C-477) positioned on a head boom a constant 7 cm distance from the singer's lips. This ensured that direct energy of the performers' voices was recorded rather than room reflections, enabling us to use a studio environment with low ambient noise rather than an anechoic studio (Cabrera, Davis, Barnes, Jacobs, & Bell, 2002). The signal was then amplified using a Behringer Ultragain preamplifier and digitally recorded to a CD recorder (Marantz CDR 630).

Calibration was carried out in each recording by playing pink noise samples immediately following each recording session at the same recording gain used for recording the singer's voice. For calibration of absolute sound pressure levels (SPL), a sound level meter (Rion NL-06 SPL) was placed adjacent to the AKG microphone 7 cm from a speaker (Bose Lifestyle) from which the pink noise was played. The SPL shown on the sound level meter was noted for the pink noise signal and used later for calibration.

Pink noise enables calibration of an audio playback system across the frequency range for tasks such as perceptual testing and comparative analysis on a computer. Pink noise is at least as good as

any other steady-state known signal. A pure sine wave tone is more susceptible to interference in the environment, and hence would give less repeatable results.

This data set was used as a basis for perceptual and acoustic evaluations.

#### *Perceptual Test*

##### *Stimuli*

The audio recordings of the singers were digitally extracted from CD using Audiograbber software ([www.audiograbber.com](http://www.audiograbber.com)) on a standard PC computer to wav audio format in stereo at 41000 Hz and 16 bit sample rate. This methodology ensures as close as possible that the original recorded sound was played back to the judges (no filtering or normalization was applied to minimize the effects of digital artifacts).

The recorded pink noise for each singer was used to equalize the peak levels of each sample to ensure that relative SPL for each singer was the same. The amplification tool in Cool Edit calculated the SPL necessary to equalize each sample's peak SPL, thus making each recording relative to the level (in dB) of the other samples produced on the day.

The files were then edited in Cool Edit Pro 1.2 ([www.adobe.com](http://www.adobe.com)) and final samples saved to CD.

##### *Procedure*

The perceptual test was conducted in a quiet environment and samples were played on CD from a Sony CD Walkman (DEJ885W) via circum-aural closed-back stereo monitoring headphones (Sennheiser HD 270). This enabled the study to be conducted in participants' singing studios. This methodology was favoured as optimal sound quality, rather than using a computer sound card (Erickson & Perry, 2003) or sending tapes to participants (Ekholm et al., 1998; Wapnick & Ekholm, 1997). Prior to presentation of stimuli, participants were given information on the two singing conditions, O and SO, based on qualitative results in a previous study (Mitchell et al., 2003). Pedagogues were instructed to assess each singing sample for use of open throat technique. In the first study, pedagogues were presented with single notes in *messa di voce* and in the second study, extended singing samples in both O and SO conditions.

Single notes as one of the test stimuli were used to assess the timbral qualities of open throat technique without potential distractors such as

intonation (Geringer & Madsen, 1998) or diction (Wapnick & Ekholm, 1997). The term 'timbre' has attracted different definitions over time. Traditionally, 'timbre' was defined as the perceptual difference between two sounds which are equal in pitch and loudness (ANSI, 1973). Following this definition, musical 'timbre' has been investigated as the difference between pairs of sounds rather than the 'timbre' of a single sound. More recently, 'timbre' has been defined as overall sound quality (Handel & Erickson, 2004). Since the aim of this study was to assess the perceived overall sound quality when a pedagogical technique is applied, we assessed the *messa di voce* responses individually.

#### *Study 1: Messa di voce*

For the *messa di voce* pairs, participants were asked to assess timbre in each sung note. Each pair contained a combination of O and/or SO by the same singer. There were four possible paired combinations: O/O, O/SO, SO/O and SO/SO. Each listener had a trial session with four sample pairs, O/O, O/SO, SO/O, SO/SO before commencing the study. There were 720 samples of *messa di voce* in 360 pairs.

#### *Study 2: Song samples*

In the second study, participants were presented with individual samples of Mozart or Schubert. They were presented with the musical score of each musical task and asked to assess each sample for use of open throat technique. Each listener had a trial session with four samples, 2 O, 2 SO for each of the musical tasks (Mozart and Schubert) before starting the second perceptual test. Each of the 15 judges responded to 24 pairs of singing samples plus six repeats.

#### *Study design*

Assessment of inter-rater agreement between judges was calculated using Cohen's Kappa (Siegel & Castellan, 1988). Kappa varies between -1 and 1, where 1 is perfect agreement and 0 is agreement not better than expected by chance. Although its use in medical settings has been criticised because it does not adjust for bias (the difference in rated prevalence of a disorder between judges), this is not a problem in a controlled experimental design where the prevalence of the two conditions being rated is identical.

As a test statistic, Kappa can verify that agreement exceeds chance levels. It is an omnibus index of agreement and does not make distinctions among various types and sources of disagreement.

Further, Kappa is influenced by distributions and base-rates. As a result, Kappas are seldom comparable across studies, procedures, or populations. Thus, attempts to categorize ranges of kappa as "good", "fair", or "poor" may be inappropriate. Krippendorff (1980) advocates agreement of at least .70.

Altman (1991) argued that any value of Kappa below .50 would indicate "poor" agreement. Fleiss (1981) recommends the following Kappa benchmarks: < .40 = poor agreement; 0.40-0.75 = intermediate to good agreement; > .75 = excellent agreement. Gardner (1995) argues that Kappa should exceed .70 before one proceeds with additional data analyses. McGinn, Wyer, Newman, Leipzig & Guyatt, (2004) state that a safe rule of thumb is that .5 indicates moderate agreement, and anything lower is considered fair to poor agreement.

## RESULTS

### MESSA DI VOCE

In the first analysis, judges' individual ratings of samples as either O or SO were assessed. Table 1 shows the results of these ratings.

A number of tests were conducted to determine the reliability of the judges in their ratings of the samples (Cicchetti & Feinstein, 1990). All confidence intervals were set at 95%. All indices demonstrated high levels of accuracy in the ratings of these judges. The measure of efficiency (ie correct overall classification rate) was 0.8597 (95% CI: 0.8322 - 0.8843); sensitivity (ie the proportion of true positives rated as positive) = 0.8111 (95% CI: 0.7668 - 0.8502) and specificity (ie the number of true negatives rated as negative) = 0.9083 (95% CI: 0.8737 - 0.9361). The chance levels of efficiency were 50%; of sensitivity, 45.14% and of specificity, 54.86%. The overall misclassification rate was 0.1403 (95% CI: 0.1157 - 0.2094); the false positive rate = 0.0917 (95% CI: 0.0639 - 0.1263); and the false negative rate = 0.1889 (95% CI: 0.1498 - 0.2332).

Cohen's Kappa was calculated to test the hypothesis that Kappa = 0. The hypothesis was not confirmed ( $z=19.40$ ,  $p=0.0000$ . Cohen's Kappa = 0.7194 (95% CI: 0.6690 - 0.7699), indicating agreement substantially above chance among this sample of judges on this perceptual test.

The reliability of the best and worst judges were

Table 1: *Messa di voce ratings. Percentages of all correctly and incorrectly identified samples by Overall percentages of correctly and incorrectly judged samples. Top row represents row percentages and bottom row in parentheses represents column ratings.*

All judgments			
Sample	O	SO	
O	81.11%	18.89%	
	(89.85%)	(17.22%)	50.00%
SO	9.17%	90.83%	
	(10.15%)	(82.78%)	50.00%
	(45.14%)	(54.86%)	N = 720

tested individually. For the best judge, the measure of efficiency = 0.9792 (95% CI: 0.8893 - 0.9995); sensitivity = 0.9583 (95% CI: 0.7888 - 0.9989) and specificity = 1.00 (85%); overall misclassification rate = 0.0208 (95% CI: 0.0005 - 0.1398); and false negative rate = 0.0417 (95% CI: 0.0011 - 0.2112). Cohen's Kappa = 0.9583 (95% CI: 0.8776 - 1.0391) indicated almost perfect agreement in the best judge's responses. ( $z=6.65$ ,  $p=0.0000$ )

For the worst judge, the measure of efficiency = 0.6458 (95% CI: 0.4946 - 0.7784); sensitivity = 0.5833 (95% CI: 0.3664 - 0.7789) and specificity = 0.7083 (95% CI: 0.4891 - 0.8738). The overall misclassification rate = 0.3542 (95% CI: 0.2216 - 0.5389); the false positive rate = 0.2917 (95% CI: 0.1262 - 0.5109); and the false negative rate = 0.4167 (95% CI: 0.2211 - 0.6336). Cohen's Kappa = 0.2917 (95% CI: 0.0232 - 0.5601) indicated poor agreement in the worst judge's responses ( $z=2.04$ ,  $p=0.0417$ ). These data are presented in Table 2.

Table 2: *Messa di voce responses of the best and worst judges. The top row represents the total number of judgments from 24 responses to O and 24 to SO and the bottom row represents the corresponding percentages of correctly and incorrectly identified samples..*

	Best Judge (3)		Worst Judge (1)	
	O	SO	O	SO
O	23	1	14	10
	(95.8%)	(4.2%)	(58.3%)	(41.7%)
SO	0	24	7	17
	(0.0%)	(100.0%)	(29.2%)	(70.8%)

### Song samples

In the individual song samples, listeners identified the use of open throat in 84% of O samples, and 69% of SO (Table 3). They were more likely to make a correct identification in the Mozart task (85% correct) than in the Schubert task (68%). Listeners identified > 83% of Mozart O and SO, and Schubert O. They were least reliable in judging Schubert SO (53%).

In 180 song samples, for O, judges rated 80 of 90 correctly in Schubert and 70 of 90 correctly in Mozart, that is, judges were less consistent in rating the Mozart task ( $k=.3$ ) than the Schubert task ( $k=.47$ ) in both O and SO.

### Reliability of judges

The 15 judges identified the experimental condition in 372 of 450 samples (82.7%). There is a 95% chance that this sample comes from a population where the true proportion comes from either side of 0.827. Kappa analysis was performed to assess listener consistency. Twelve of fifteen judges were moderately consistent in their judgments ( $k \geq 0.600$ ); judges 1, 5 and 7 were inconsistent in their judgements. These data are presented in Table 4.

### Individual singers

Individual singers' O samples, in both Mozart and Schubert were more readily identified by listeners than SO. Figure 2a-d presents these data graphically, showing the correct identifications and misclassifications of O and SO for each singer in each musical task. Four of six singers (1-4) were consistently identified correctly as singing with open throat in O and less open throat in SO. Singers 5 and 6 were outliers. Judges incorrectly judged Singer 6's intentions in Mozart O and SO

Table 3: *Song responses. Percentages of correct (hit) and incorrect (miss) responses to the song task, by condition (optimal and sub-optimal) and by condition and task (Mozart and Schubert).*

Task	Hit %	Miss %
All O	84%	16%
All SO	69%	31%
Mozart O	86%	14%
Mozart SO	84%	16%
Schubert O	83%	17%
Schubert SO	53%	47%

Table 4: Reliability of judges' responses. Cohen's Kappa, Asymptomatic standard error (not assuming the null hypothesis), asymptotic standard error (assuming the null hypothesis) and approximate significance for listeners 1-15.

Listener	Kappa	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
1	0.400	0.167	2.196	0.028
2	0.867	0.091	4.747	0.000
3	0.733	0.123	4.053	0.000
4	0.800	0.109	4.392	0.000
5	0.467	0.161	2.556	0.011
6	0.867	0.090	4.790	0.000
7	0.200	0.179	1.095	0.273
8	0.600	0.145	3.316	0.001
9	0.733	0.124	4.017	0.000
10	0.667	0.133	3.727	0.000
11	0.733	0.120	4.168	0.000
12	0.667	0.133	3.727	0.000
13	0.667	0.136	3.660	0.000
14	0.600	0.134	3.586	0.000
15	0.800	0.107	4.472	0.000

and identified her O as SO and SO as O in over half of the 30 judgements. Singer 5 was incorrectly identified in Schubert SO only, that is, judges viewed her singing in SO as O.

#### *Singers' self-perceptions of their sound*

Overall means of singers' VAS results indicated that singers associated their performance of the O condition with a high degree of openness and strongly representative of their best singing (Figure 3). Subject 6 was dissatisfied with all of her singing on the test day and cast "informal votes" when asked to rate her singing. Her responses were omitted from the final result.

These responses are in accordance with the singers' intentions in their performance in each condition. Pearson  $r$  varied between the responses to best singing and openness according to condition and task. The O condition responses were moderately correlated in the Mozart task ( $r=0.45$ ), but highly correlated in the Schubert ( $r=0.84$ ) but neither was statistically significant ( $p>0.05$ ). In the

SO condition, responses were both very highly correlated ( $r=0.98$ ,  $r=0.94$ ), and significantly different from zero ( $p<0.01$ ,  $p<0.05$ ). At least 90% of the singers' perception of good singing can be attributed to their perception of openness.

## DISCUSSION

This study confirmed that the use of open throat technique is a perceptual reality to singers and that it produces a specific vocal quality in classical singing that can be reliably identified by expert listeners. While previous acoustic studies proved inconsistent in differentiating between experimental conditions O and SO, the human ear produced the most reliable assessment of vocal quality.

#### *Messa di voce*

Collectively, judges were accurate in their identification of O and SO and in each messa di voce sample, the majority of listeners correctly recognised the use of open throat when it occurred. Reliability of listeners' judgements ranged from fair to highly consistent. The high proportion of correct judgements of O and SO suggests that these singers produced a specific sound quality or timbre that expert pedagogues could reliably differentiate, thus strengthening the claim that open throat technique is associated with a particular desirable vocal quality.

In a previous study, expert pedagogues confirmed their collective understanding of the technique of open throat and its use as an essential pedagogical instruction in the singing studio (Mitchell et al., 2003). This study confirmed their assertions in the previous study that open throat technique produces a specific audible quality, readily identifiable by vocal pedagogues.

Perceptually, we considered it an important methodological advance to assess open throat technique in single notes (Erickson, 2003; Erickson & Perry, 2003) before progressing to extended singing samples (Ekholm et al., 1998; Howard et al., 2002; Robison et al., 1994; Wapnick & Ekholm, 1997). While extended singing samples are more representative of listeners' and pedagogues' experiences in the singing studio and in singing assessment, testing timbre of single notes removes any influence of intonation, diction and musical stimuli on judgements and focuses listeners directly to core sound quality. Proficiency in the messa di voce is a key indicator of vocal mastery (Miller, 1996) and these samples provided listeners with

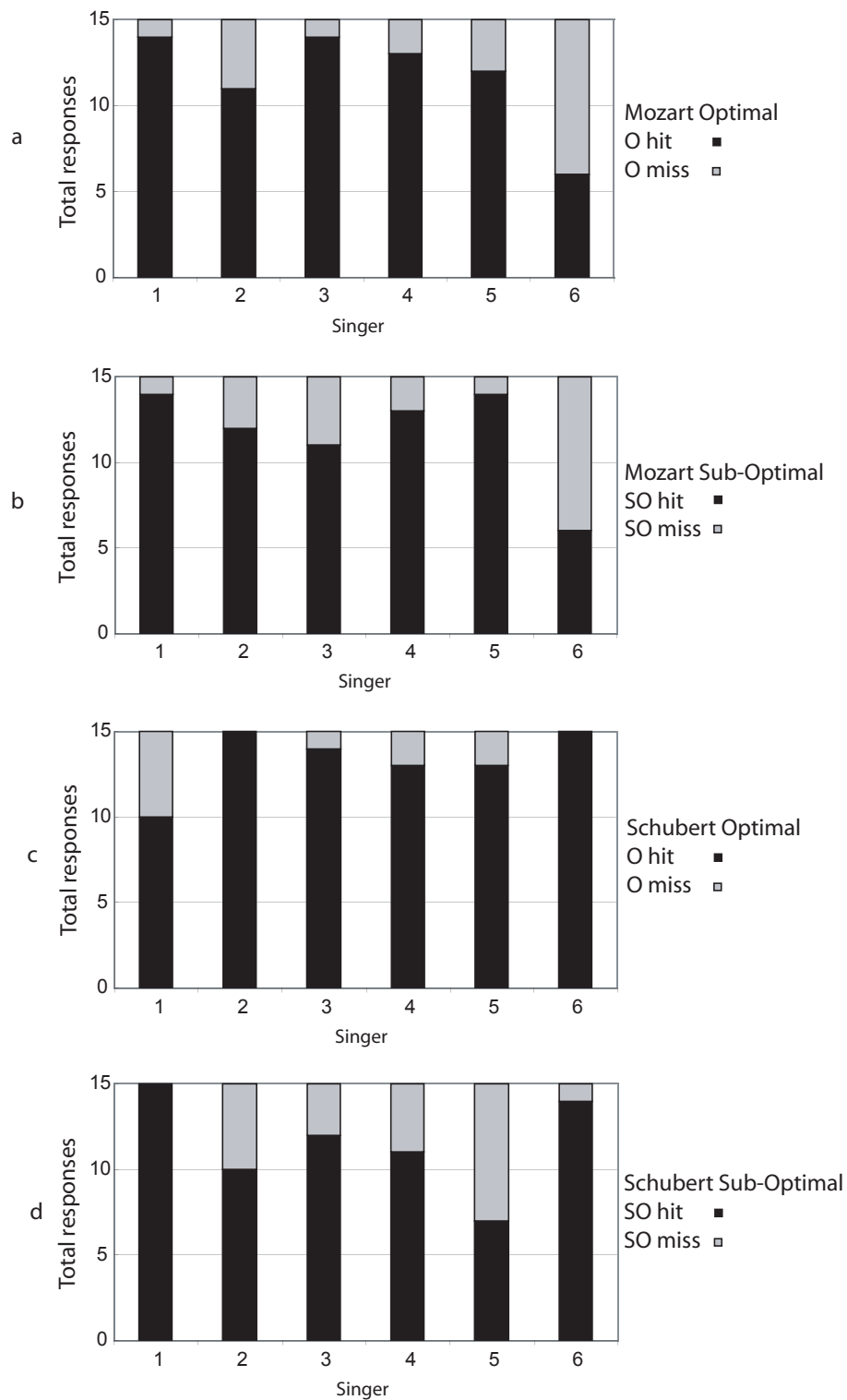


Figure 2 Correct identifications and misclassifications for each singer (1-6) in (a) Mozart optimal (O); (b) Mozart sub-optimal (SO); (c) Schubert optimal and (d) Schubert sub-optimal.

sufficient vocal information to respond to timbre.

As the definition of ‘timbre’ is generally accepted as overall or intrinsic sound quality (Handel & Erickson, 2004), it was less important to the aims of this part of our study to report the difference between one sound and another (ANSI, 1973).

### Song samples

Twelve of fifteen expert listeners consistently identified the use of open throat technique in O and a reduction of the technique in SO in the song samples. Listeners demonstrated reliability in their judgements through the duration of the song task, repeating their judgement of O or SO in the

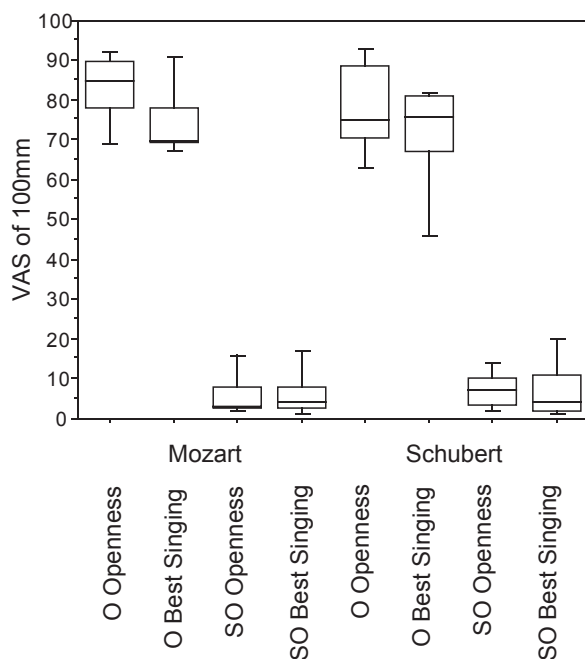


Figure 3: Overall means of singers' VAS scores from 1-100, for singers 1-5 (singer 6 omitted) to report the degree of openness and representation of her best singing for each musical task (Mozart and Schubert) in each condition (optimal and sub-optimal).

six repeated samples in the perceptual test. These results support previous research that indicates high intra-judge agreement (Davidson & Da Costa Coimbra, 2001; Ekholm et al., 1998) in assessing overall vocal quality.

However, when Mitchell and Kenny assessed the use of open throat in these singers through acoustic characteristics of vibrato (Mitchell & Kenny, 2004b) and energy distribution in LTAS (Mitchell & Kenny, 2004b), their findings did not accord with experts' perceptual evaluation (Kenny, Davis, & Oates, in press). While the vibrato analysis revealed significant reductions in vibrato extent and increases in onset time in SO compared to O, conventional measures of singing power ratio (SPR) (the difference between the peak height) and energy ratio (ER) (the difference between peak area) on LTAS of O and SO were not statistically significantly different (Mitchell & Kenny, 2004b). Changes in vibrato rate and extent, inappropriate to genre are considered indicative of poor singing so it was hypothesized that testing the energy distribution in these singers' voices in each condition would identify the timbral changes associated with open throat. LTAS results were not consistent with the vibrato findings for O/SO, and SPR and ER were not sensitive to the timbral differences between O/SO identified perceptually by expert

listeners in this study.

We examined our hypotheses using both extended singing samples as well as single notes (as in the messa di voce) because each musical task tests a different demand of good singing and the former more closely resembles singing as heard and assessed in the studio and in formal assessment situations (Davidson & Da Costa Coimbra, 2001; Ekholm et al., 1998; Howard et al., 2002; Robison et al., 1994; Wapnick & Ekholm, 1997).

An interesting and somewhat unexpected finding in our study was that listeners were more likely to identify the correct experimental condition in the Mozart song sample rather than Schubert. Previous perceptual studies have recommended using the same musical stimuli to reduce inter-subject variables that may influence responses (Howard et al., 2002; Saunders & Holahan, 1997); yet, these findings indicated an effect of musical task. Indeed, as open throat is of critical importance in high sustained singing (Mitchell et al., 2003), the scale motif and high tessitura of the Schubert task may have proved too difficult for singers to truly reduce open throat in SO. Conversely, the Mozart task was less vocally demanding, opening up the possibility that singers require a less robust technique to make the desired sound.

When Mitchell and Kenny (2004a) assessed these singers' vocal qualities in O and SO through the energy distribution in LTAS, the effect of musical task approached significance, that is, both SPR and ER were slightly greater in Schubert than in Mozart. According to current singing literature, lower ER and SPR should indicate improved singing quality, by boosted energy above 2 kHz (Omori et al., 1996; Thorpe et al., 2001). A future study by these authors will examine the significance of SPR and ER rankings with perceptual judgment of overall quality (Kenny & Mitchell, in press).

Examination of correct judgements by singer indicated singers 1-4 were correctly identified more often than singers 5 and 6. In testing the sound quality of open throat by singer, singers 1-4 were consistently identified correctly in O and SO. Singer 5 and 6 were outliers in certain conditions and tasks. Singer 6, in the Mozart task, produced the most inconsistent responses amongst the judges. Her O and SO were incorrectly identified in over 50% of occurrences, suggesting there was little difference between her O and SO, and that judges collectively were unclear whether she was attempting to use maximum open throat. On the other hand, listeners had little trouble identifying

singer 5's Schubert O, but were misled by her SO, judging it as O in 8 of 15 cases. The overall high quality of singer 5's voice may have misled judges to rate her SO as O.

For an accomplished singer, it may be possible to sing using reduced open throat technique but still produce an aesthetically pleasing sound, or equally, use an adequate open throat technique but produce an aesthetically displeasing sound as in the case of singer 6. Clearly, there are many factors that converge in producing an aesthetically pleasing vocal quality, of which open throat is but one. Listeners have clearly attended to multiple dimensions of the vocal production in arriving at their perceptual judgments of the singers. In fact, singers 5 and 6 were the highest and lowest scoring singers respectively in a study of overall vocal quality (Kenny & Mitchell, in press) but these rankings showed no relationship with corresponding rankings of SPR and ER, which are intended as objective measures of voice or voice quality.

Singers in this study were acutely aware of the impact of open throat technique on their overall singing sound. Singers 1-5 equated good singing with use of maximal open throat technique and felt their overall singing sound decreased when they reduced the technique. Matching a performer's intention to a pedagogical concept is valuable to singing pedagogy. Singer 6 was dissatisfied with her singing on the day of recording, which accorded with listeners' judgements of her vocal quality and the similarities between her O and SO. It may be, as one listener pointed out, that it is 'hard to tell if it is a good singer singing badly, or a bad singer singing well'.

Sophisticated acoustic measurement technologies are now being routinely applied to voice science and vocal pedagogy the goal of which is to provide objective measurement of voice. To date, studies that attempt to define a good voice perceptually and acoustically (Ekholm et al., 1998; Wapnick & Ekholm, 1997) or assess vocal mastery throughout musical education (Mendes et al., 2003; Vurma & Ross, 2000) by linking perceptual ratings with acoustic measurements have not assessed the relationship between vocal mastery and the development of specific technical proficiencies. If singers' and expert listeners' reports of singing accord, focussing research on specific singing techniques could enhance singing pedagogy. Matching optimal singing quality with performer technique or intention enhances communication between singers and teachers. These findings will

be strengthened by replication in male voices and professional singers.

Current work in singing has not sufficiently incorporated perceptual ratings and descriptions of sound quality. Vocal quality results from a complex combination of acoustical parameters that interact. To date, no single objective evaluation captures or characterizes vocal quality in a systematic way (Omori et al., 1996; Thorpe et al., 2001). The future of experimental studies in the teaching of singing depends on their incorporation into a theory of the singing voice and on their relation to current practice.

Most singers and teachers of singing understand vocal functioning and control through audiation, sensory feedback and mental imagery. Teachers have a mental image of the particular sound individual students are capable of producing and use verbal explanation, vocal modeling and feedback to teach this. Despite the fact that teachers and assessors of singing process sound using sensory and auditory methods as their primary feedback (Kitch & Oates, 1994), few perceptual studies use these professional skills in assessing vocal quality. If expert listeners can reliably identify a technique, such as open throat, in the overall singing sound, future perceptual and acoustic studies of voice may provide more useful information for the dissemination of current vocal practices, their assessment and documentation in singing pedagogy. It is now feasible to track the methods by which good singers are trained. In linking the use of specific pedagogical techniques with perceptual judgment and acoustic measurement, future studies in voice can provide a basis for a more systematic pedagogy of the singing voice.

## ACKNOWLEDGEMENTS

We thank Ms Maree Ryan for her pedagogical advice and support, Mr Peter Thomas and Dr Denis Cabrera for their advice on acoustical matters, the six singers for their willing participation in the project, and the reviewers for helpful comments on an earlier draft of the manuscript.

## REFERENCES

- Altman DG (1991). *Practical Statistics for Medical Research*. Chapman and Hall.
- ANSI. (1973). *Psychoacoustical terminology: S3.20*. New York: American National Standards Institute.
- Brown, W. S., Jr, Rothman, H. B., Morris, R. J., & Sa-

- pienza, C. M. (2001). Acoustic analyses of trained singers perceptually identified from speaking samples. *Journal of Voice*, 15(1), p. 25-35.
- Brown, W. S., Jr., Rothman, H. B., & Sapienza, C. M. (2000). Perceptual and acoustic study of professionally trained versus untrained voices. *Journal of Voice*, 14(3), 301-309.
- Bunch, M., & Chapman, J. (2000). Taxonomy of singers used as subjects in scientific research. *Journal of Voice*, 14(3), 363-369.
- Burgin, J. C. (1973). *Teaching singing*. Metuchen, N.J.: Scarecrow Press.
- Cabrera, D., Davis, P., Barnes, J., Jacobs, M., & Bell, D. (2002). Recording the operatic voice for acoustic analysis. *Acoustics Australia*, 30, 103-108.
- Callaghan, J. (2000). *Singing and voice science*. San Diego, California: Singular Pub. Group.
- Cicchetti, D. V., & Feinstein, A. R. (1990). High agreement but low kappa: II. Resolving the paradoxes. *J. Clin. Epidemiol*, 43, 551-558.
- Davidson, J. W., & Da Costa Coimbra, D. (2001). Investigating performance evaluation by assessors of singers in a music college setting. *Musicae Scientiae*, 5(1), 33-53.
- Ekholm, E., Papagiannis, G. C., & 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.
- Erickson, M. L. (2003). Dissimilarity and the Classification of Female Singing Voices: A Preliminary Study. *Journal of Voice*, 17(2), 195-206.
- Erickson, M. L., & Perry, S. R. (2003). Can listeners hear who is singing? A comparison of three-note and six-note discrimination tasks. *Journal of Voice*, 17(3), 353-369.
- Fields, V. A. (1947). *Training the singing voice: an analysis of the working concepts contained in recent contributions to vocal pedagogy*. New York: King's Crown Press.
- Fleiss, J.L. (1995) *Statistical Methods for Rates and Proportions* (2nd edition). New York: Wiley 1981.
- Foulds-Elliott, S., Thorpe, C., Cala, S., & Davis, P. (2000). Respiratory function in operatic singing: effects of emotional connection. *Logopedics Phoniatrics Vocology*, 25(4), 151-168.
- Gardner, W. (1995). On the reliability of sequential data: measurement, meaning, and correction. In John M. Gottman (Ed.), *The analysis of change*. Mahwah, N.J.: Erlbaum.
- Geringer, J. M., & Madsen, C. K. (1998). Musicians' ratings of good versus bad vocal and string performances. *Journal of Research in Music Education*, 46(4), 522-534.
- Handel, S., & Erickson, M. L. (2004). Sound Source Identification: The Possible Role of Timbre Transformations. *Music Perception*, 21(4), 587-610.
- Hemsley, T. (1998). *Singing and imagination: a human approach to a great musical tradition*. Oxford; New York: Oxford University Press.
- Herbert-Caesari, E. F. (1951). *The voice of the mind*. London: Robert Hale.
- Howard, D. M., Szymanski, J., & Welch, G. F. (2002). Listeners' perception of English girl and boy choristers. *Music Perception*, 20(1), 35-49.
- Joiner, J. R. (1998). *Charles Amable Battaille: pioneer in vocal science and the teaching of singing*. Lanham, Md.: Scarecrow Press.
- Kenny, D.T., Davis, P. & Oates, J. (2004). Music performance anxiety and occupational stress amongst opera chorus artists and their relationship with state and trait anxiety and perfectionism. *Journal of Anxiety Disorders*, 18, 6, 757-777.
- Kenny, D. T., & Mitchell, H. F. (in press). Acoustic and perceptual appraisal of vocal gestures in the female classical voice. *Journal of Voice*.
- Kitch, J. A., & Oates, J. (1994). The perceptual features of vocal fatigue as self-reported by a group of actors and singers. *Journal of Voice*, 8(3), 207-214.
- Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., & Berke, G. S. (1993). Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *Journal of Speech & Hearing Research*, 36(1), 21-40.
- Krippendorff, K. (1980). *Content Analysis: An Introduction to its Methodology*. Sage Publications, Thousand Oaks, CA.
- Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- Manèn, L. (1987). *Bel canto: the teaching of the classical Italian song-schools its decline and restoration*. Oxford: Oxford University Press.
- Marafioti, P. M. (1981). *Caruso's Method of Voice Production*. New York: Dover Publications.
- McKinney, J. C. (1982). *The Diagnosis and Correction of Vocal Faults*. Nashville, Tennessee: Broadman Press.
- McGinn, P. T., Wyer, C., Newman, T. B., Leipzig, R., & Guyatt, G. (2004). Tips for learners of evidence-based medicine: Measures of observer variability (kappa statistic). *Canadian Medical Association Journal*, 171(11), 1369-1373.
- Mendes, A. P., Rothman, H. B., Sapienza, C. M., & Brown, W. S., Jr. (2003). Effects of vocal training on the acoustic parameters of the singing voice. *Journal of Voice*, 17(4), 529-543.
- Miller, R. (1996). *The structure of singing: system and art in vocal technique*. New York: London: Schirmer Books.
- Miller, R. (1997a). *The Open Throat (La gola aperta)*.

- In Miller, R. (Ed.), *On the Art of Singing* (pp. 17-19). New York: Oxford University Press.
- Miller, R. (1998). The singing teacher in the age of voice science. In Sataloff, R. T. (Ed.), *Vocal Health and Pedagogy* (pp. 297-300). San Diego, CA: Singular.
- Miller, R., 1926-. (1997b). *National schools of singing : English, French, German, and Italian techniques of singing revisited / Richard Miller. ([Rev. ed.]. ed.)*. Lanham, Md.: Scarecrow Press.
- Mitchell, H. F., & Kenny, D. T. (2004a). The effects of open throat technique on long term average spectra (LTAS) of female classical voices. *Logopedics Phoniatrics Vocology*, 29(3), 99-118.
- Mitchell, H. F., & Kenny, D. T. (2004b). The impact of "open throat" technique on vibrato rate, extent and onset in classical singing. *Logopedics Phoniatrics Vocology*, in press.
- Mitchell, H. F., Kenny, D. T., Ryan, M., & Davis, P. J. (2003). Defining open throat through content analysis of experts' pedagogical practices. *Logopedics Phoniatrics Vocology*, 28(4), 167-180.
- Monahan, B. J. (1978). *The art of singing: a compendium of thoughts on singing published between 1777 and 1927*. Metuchen, N.J.: Scarecrow Press.
- Nair, G. (1999). *Voice tradition and technology: a state-of-the-art studio*. San Diego: Singular Publishing Group.
- Omori, K., Kacker, A., Carroll, L. M., Riley, W. D., & Blaugrund, S. M. (1996). Singing power ratio: quantitative evaluation of singing voice quality. *Journal of Voice*, 10(3), 228-235.
- Puritz, E. (1956). *The teaching of Elisabeth Schumann*. London: Methuen.
- Reid, C. L. (1975). *Voice: psyche and soma*. New York: Joseph Patelson Music House.
- Reid, C. L. (1983). *A dictionary of vocal terminology*. New York: Joseph Patelson Music House.
- Robison, C. W., Bounous, B., & Bailey, R. (1994). *Vocal Beauty: A study proposing its acoustical definition and relevant causes in classical baritones and female belt singers*. *Journal of Singing*, 51, 19-30.
- Saunders, T. C., & Holahan, J. M. (1997). Criteria-specific rating scales in the evaluation of High School instrumental performance. *Journal of Research in Music Education*, 45(2), 259-272.
- Siegel, S., & Castellan, J. N. J. (1988). *Nonparametric Statistics for the Behavioural Sciences* (Second ed.): McGraw Hill International.
- Stanley, M., Brooker, R., & Gilbert, R. (2002). Examiner perceptions of using criteria in music performance assessment. *Research Studies in Music Education*, 18, 43-52.
- Stark, J. A. (1999). *Bel canto: a history of vocal pedagogy*. Toronto, Buffalo: University of Toronto Press.
- Sundberg, J. (1977). The acoustics of the singing voice. *Scientific American*(March), 82-91.
- Sundberg, J., Gramming, P., & Lovetri, J. (1993). Comparisons of pharynx, source, formant, and pressure characteristics in operatic and musical theatre singing. *Journal of Voice*, 7(4), 301-310.
- Thorpe, C., Cala, S., Chapman, J., & Davis, P. (2001). Patterns of breath support in projection of the singing voice. *Journal of Voice*, 15(1), 86-104.
- Thurman, L., & Welch, G. F. (Eds.). (2000). *Bodymind & voice: foundations of voice education* (Rev. ed. ed.). Collegeville, Minn: VoiceCare Network.
- Vennard, W. (1968). *Singing: the mechanism and the technic* (5th ed.). New York: Fischer.
- Vurma, A., & Ross, Y. (2000). Priorities in voice training: carrying power or tone quality. *Musicae Scientiae*, 4(1), 75-93.
- Wapnick, J., & Ekholm, E. (1997). Expert consensus in solo voice performance evaluation. *Journal of Voice*, 11(4), 429-436.
- Wapnick, J., Flowers, P., Alegant, M., & Jasinskas, L. (1993). Consistency in piano performance evaluation. *Journal of Research in Music Education*, 41(4), 282-292.
- Ware, C. (1998). *Basics of vocal pedagogy: the foundations and process of singing*. New York: McGraw-Hill.