journal of interdisciplinary music studies
spring/fall 2008, volume 2, issue 1&2, art. #0821206, pp. 95-110

The Tertiary Singing Audition: Perceptual and Acoustic Differences between Successful and Unsuccessful Candidates

Helen F. Mitchell and Dianna T. Kenny

Australian Centre for Applied Research in Music Performance (ACARMP)
Conservatorium of Music, University of Sydney

Background in music practice. The audition process is used extensively in the arts to identify talent and to determine entry into music conservatoria. Judges must make rapid decisions regarding the quality and training potential of singing voices, however, there is limited understanding of these decisional processes and the criteria on which they are based.
Background in music perception and acoustics. Few scientific studies link perceptual judgments with acoustic parameters. The audition process provided an opportunity to investigate these in an authentic singing assessment setting.
Aims. This is the first study to evaluate the singing audition process by assessing judges' reliability in adjudication and the relationship between audition outcomes (voice major, minor and not accepted), descriptions of voices by adjudicators and acoustic indicators of successful performances of singers who auditioned for entry into a Bachelor of Music degree.
Main contribution. We observed and recorded a singing audition. The panel of judges was required to achieve consensus in their allocation of students into three groups - major or minor streams of the vocal program or non-acceptance into the degree. However, their comments on singers' vocal quality did not identify specific features that were used to make these allocations. There was no effect of time or day on audition outcomes. There were significant differences between mean SPL in performances of females allocated to different groups but there were no differences in SPL between males' performances.
Implications. Assessment of the audition process in an authentic audition setting raised a number of significant methodological issues about data capture and analysis that need to be addressed in future studies. The challenge is to identify and analyse meaningful acoustic data from a performance setting that does not interfere with or alter the performance or assessment process.

Keywords: Singing, audition, assessment, ranking, terminology, vocal quality, vocal performance

•*Correspondence:* Helen Mitchell, Australian Centre for Applied Research in Music Performance (ACARMP), Conservatorium of Music C41, University of Sydney NSW 2006 Australia; e-mail: h.mitchell@usyd.edu.au

Introduction

Auditions have a long history as a means to gain entry into a music performance degree (Legge, 2001; Subotnik, 2003). They often mark the beginning or end of the musical careers of young singers. Adjudicators evaluate technical skill (Subotnik, 2003, 2004), assess singing potential or 'talent' (Davidson & Da Costa Coimbra, 2001; Hollien, 1993; Watts, Barnes-Burroughs, Andrianopoulos, & Carr, 2003) and singing voice quality (Geringer & Madsen, 1998) as part of the adjudication process. Understanding the basis of adjudicator judgments (Davidson & Da Costa Coimbra, 2001; Stanley, Brooker, & Gilbert, 2002) and their reliability (Goldin & Rouse, 2000) within an authentic audition constitutes a major challenge for the field of music research.

Objective measures of quality of sung performance would be invaluable for voice pedagogy and competitive selection processes such as auditions. However, the gap between phenomena that are relevant to excellence in performance and phenomena that are observable through basic voice science is still wide (Ternström, 2005). This paper represents one attempt to relate objective observations of measurable voice quality phenomena from the acoustic waveform to the subtle and general aspects of singing that are assessed at adjudication and to identify perceptual and acoustic features of the singing voice that reliably separated successful from unsuccessful applicants.

Perceptual studies indicate that listeners show some degree of reliability and consistency in their judgments of good and poor vocal and instrumental performances (Ekholm, Papagiannis, & Chagnon, 1998; Geringer & Madsen, 1998; Saunders & Holahan, 1997; Smith, 2004) and in their assessment of excellence in overall voice quality (Kenny & Mitchell, 2006; Stanley et al., 2002; Wapnick & Ekholm, 1997). However, factors such as day, time of day, performer order and listener fatigue can all affect judge consistency (Bergee & Platt, 2003; Elliott, Schneider, & Zembower, 2000; Flores & Ginsburgh, 1996). Judges' reliability increases when teams of adjudicators are used, in which the highest and lowest scores are eliminated and the remaining scores averaged to produce the final score (Dugger, 1997). This and similar processes may remove the effects of judge fatigue and other biases (e.g. gender bias) in the adjudication process (Elliott, 1996; Goldin & Rouse, 2000).

There is a vast literature on the acoustic properties and visual representation of the singing voice that is now incorporated into singing and pedagogy texts [e.g. (McCoy, 2004; Miller, 1996; Nair, 1999; Sundberg, 1987)]. Particular features of the singing voice, although helpful in assessment, do not describe the whole sound and may lead to an incomplete or inaccurate representation of the individual singing voice (Ekholm et al., 1998). These qualities include loudness (Mendes, Rothman, Sapienza, & Brown, 2003; Mürbe, Sundberg, Iwarsson, Friedmann, & Hofmann, 1999), vibrato (Mendes et al., 2003; Mürbe, Zahnert, Kuhlisch, & Sundberg, 2007) and vocal energy (Mürbe et al., 1999; Vurma & Ross, 2000; Watts, Barnes-Burroughs, Estis, & Blanton, 2006). As these features of vocal quality are considered important outcomes of vocal training, we aimed to identify the presence of these features in the performances of singers who were either accepted or not accepted into a conservatorium of music.

Sound pressure level (SPL) is closely related to vocal loudness and there is evidence to suggest that singers undergoing classical training (Akerlund & Gramming, 1994) or with an advanced technique (Mitchell & Kenny, 2004b) produce a louder sound compared with untrained singers and can increase their average SPL over time (Mürbe et al., 1999). Classical vocal training stabilizes variations in SPL throughout the pitch range in young singers (Mürbe et al., 1999). Long term average spectra (LTAS) have been used to analyse spectral characteristics of voice quality. For professional classical and operatic voices, LTAS demonstrate a spectral reinforcement between 2-4 kHz (Barnes, Davis, Oates, & Chapman, 2004; Sundberg, 1974; Thorpe, Cala, Chapman, & Davis, 2001), which achieves ‘carrying power’ in classical singing voices. Measures applied to LTAS reduce the information contained in an LTAS (spectral tilt or slope) to a single meaningful number and enable inter-singer comparisons (Kenny & Mitchell, 2006; Löfqvist, 1986; Löfqvist & Mandersson, 1987; Mitchell & Kenny, 2004a; Omori, Kacker, Carroll, Riley, & Blaugrund, 1996; Thorpe et al., 2001). LTAS of professional and student singers show marked differences (Barnes *et al*., 2004; Kenny & Mitchell, 2006; Thorpe et al., 2001). Acoustic and perceptual assessment of singing voice has identified trained singers from non singers (Brown, Rothman, & Sapienza, 2000; Mendes et al., 2003). However, LTAS and mean or maximum SPL are static measures that cannot capture the singer’s ability to modulate the voice in response to the expressive demands of the music. These measures therefore remain crude proxies of what an expert listener hears, and some research has shown that neither carrying power (Vurma & Ross, 2000) nor measures on LTAS (Kenny & Mitchell, 2006; Mitchell & Kenny, 2004a) are reliably associated with expert ratings of vocal quality.

Identifying talent or potential, as in a singing audition, may be more problematic for younger singers, who will have technical facility appropriate to age, training and experience, but are unlikely at that stage of their musical development to be able to demonstrate a fully mature or developed vocal instrument. However, a recent study by Watts et al. (2006) successfully differentiated between the vocal qualities of untrained talented and non-talented singers and illustrated these differences through spectral measures and LTAS curves. Comparing perceptual and acoustic responses at audition may provide a foundation for assessing voices of elite singing students who gain entry to music conservatoria compared with advanced singing students who are unsuccessful at audition.

This study assessed the relationship between perceptual features, terminology to describe vocal quality and general acoustic characteristics of voices awarded a place in a vocal unit at a prestigious conservatorium of music with unsuccessful applicants.

Method

Ethical approval for the study was obtained from the Human Research Ethics Committee, The University of Sydney.

Participants

Singers (n=60) who auditioned for the Bachelor of Music degree at the Sydney Conservatorium of Music in 2006 were the participants in this study. Auditions took place in the Music Workshop at the Conservatorium of Music, Sydney, Australia. Singers were expected to prepare four songs of their own choice for their audition and typically, most singers would be asked to sing two of their prepared songs (from memory) from the categories:

- One vocal study (e.g. works of Concone, Marchesi etc)
- A 17th or 18th century Italian song
- A Lied (e.g. Schubert or Schumann)
- A song of the candidate's own choice

The audition panel comprised members of the vocal unit at the Sydney Conservatorium of Music, all highly renowned for their singing studios and singing pedagogy. This panel engaged in a group decision making process and arrived at selections and audition scores using discussion and consensus (Subotnik, 2003), that is panel members did not submit individual scores (Davidson & Da Costa Coimbra, 2001; Dugger, 1997). A single score and comments sheet was prepared for each candidate immediately following the group discussion.

Procedure

Applicants were sent information about the recording process prior to their audition time and given the opportunity to opt out of the recording before the audition or on the day of the audition. They were offered a CD copy of their audition performance if they chose to participate in the study.

Recordings

Auditions took place in a performance hall at the Sydney Conservatorium of Music[1]. The audition panel sat on the stage area approximately 9m from the singer and piano. All singers were accompanied by the same expert pianist from the Conservatorium. The expert accompanist and the acoustic environment remained constant during the audition process.

A Head and Torso Simulator (HATS; Brüel & Kjær 4100-D) was placed directly behind the audition panel (height to pinnae 155cm) approximately 10m from the piano and singer. The HATS mannequin has built-in ear simulators that provide a realistic reproduction of the acoustic properties of an average adult human head and torso. HATS approximates the frequency response of the human ear.

Microphone inputs in the Music Workshop were patched to the Conservatorium's main recording studio for audio capture (microphone preamplifier JLM MP8). The acoustic signals were digitised (Digidesign ProTools HD192 A-D convertor) and captured as 24 bit, 48 kHz AIFF lossless files using Pro Tools. Recordings were calibrated (Brüel & Kjær DP 4231) so the singer's absolute sound pressure was known at these microphones. We were able to achieve a high quality recording environment that produced audio files that emulated the experience of the listener (Fletcher & Munson, 1933) in an auditorium environment. Audio files were edited in Adobe Audition and lossless WAV files were labelled by subject number.

Order effects

The singing auditions occurred over 2.5 days in five sessions (morning and afternoon). Judge rankings of singers were assessed against the order in which the singers auditioned in order to identify any effects of day (day 1, 2, 3) or time (morning/afternoon). Singers' audition ranks were compared with their audition order and examined for effects of ordering by day or time.

Audition reports

Each singer's audition report was examined for the candidate's repertoire choices, final mark and assignment to group (major, minor, not accepted). Comments on the performances were examined for descriptions of vocal quality and ability. Major and minor levels reflect different levels of entry and attainment and majors are considered to have more performance potential. Voice majors receive extensive training in public performance while voice minors train without an emphasis on public performance.

Acoustic Analysis

Each singer's first song performance was analysed using PRAAT[2] (Boersma & Weenink, 2006). Measures of mean and maximum sound pressure level (SPL) were calculated in PRAAT. Long-term average spectra analyses (performance LTAS) were performed on each song audio file to describe the acoustic nature of the singers' entire performance and elucidate acoustic cues which may be responsible for the panel's perceptual judgements. Piano accompaniment would certainly have affected the perceptual and acoustic cues available to judges and these are considered in this data. Data were copied into Excel for further calculation. LTAS were calibrated to be representative of known dB levels (Mitchell & Kenny, 2004a).

Perceptual and acoustic relationships

Relationships between acoustic parameters and perceptual ratings of voices awarded a place at major and minor level and voices not accepted into the course were examined to determine the degree of concordance between the acoustic parameters and their group assignment.

Statistical analysis
Separate one-way between groups (males, females) Analysis of Variance (ANOVA) with a full set of four planned comparisons (based on audition outcome) was undertaken on mean and maximum sound pressure level (SPL in dB). Before interpreting the results, an examination of the Levene's test for homogeneity of variance was undertaken to determine whether any of the assumptions for ANOVA had been violated.

Results

Twenty four singers were offered a place, 12 as voice majors and 12 and voice minors. The performances of these 24 top ranked singers were compared with performances of the next 23 ranked singers who were unsuccessful at audition and not offered a place to study voice. Of the 60 original candidates, there were 47 female and 13 male applicants. Table 1 presents study group by gender, age and voice type.

Table 1. Study group demographics: highest 43 ranked audition candidates by gender, audition outcome, voice type with mean age and standard deviation (SD) in years for each group.

	Females			Males		
	Major	Minor	Not Accepted	Major	Minor	Not Accepted
Total	5	9	20	7	3	3
Soprano	4	8	20	-	-	-
Mezzo	1	1	-	-	-	-
Tenor	-	-	-	2	1	1
Baritone	-	-	-	5	2	2
Average Age (years)	18.2	19.7	18.3	20.1	17.9	18.5
(SD)	(0.1)	(2.9)	(0.7)	(2.0)	(0.8)	(0.9)

Effects of day and time of day
There was no statistically significant effect of day or time of day on the audition results (Fisher's Exact Test = 3.472, $p=0.54$).

Repertoire selection
Candidates selected their first audition song, presented in Figure 1. The majority of majors sang their own song choice and minors/NA were most likely to sing an Italian song or German lied. Own choice songs comprised art songs from other traditions (French, English, American) (n=12, 2 male majors, 2 male minors, 1 female minor, 6

NA females, 1 NA male), oratorio (n=3, 1 female major, 1 male major, 1 male NA) and operatic arias or Neapolitan songs (n=6, 1 female major, 3 male majors, 3 minor females).

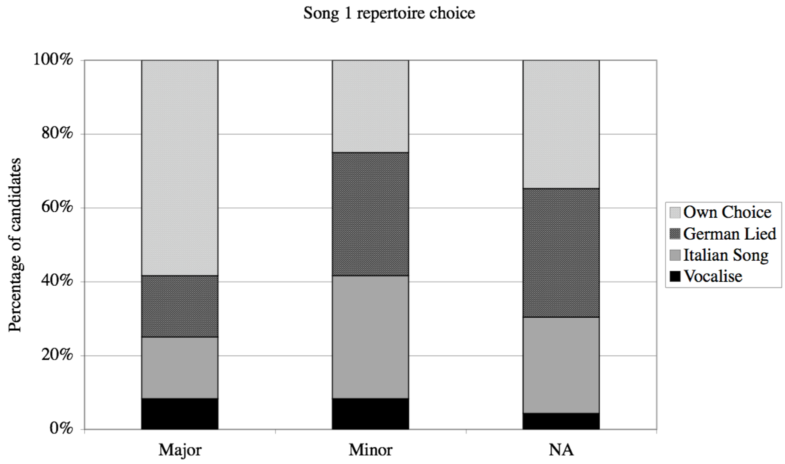


Figure 1. Percentages of candidates' first song selections from the designated audition categories by candidates in each audition group (major, minor, not accepted (NA)).

Summary of adjudicator comments

The panel, through their comments, showed that they were primarily focused on candidates' vocal quality although there was also interest in overall presentation, musicality, style and personality of the singers. Voice majors were deemed to have a superior 'instrument' or a voice of 'high' or 'lovely' quality. In addition, the highest ranked majors were credited with 'potential', 'promise' or 'talent' for a future in the singing profession. Technical concerns, such as manufacturing a 'darker' sound in female majors attracted comment but singers were not penalised if other superior vocal qualities were noted. Minor voices were described as having a 'pleasing' vocal quality but were more likely to attract technical comments or criticisms even if the core voice was deemed to have potential. These voices were described as lacking in 'support' or breath coordination and two female minor voices were described as having a 'pushed' sound leading to problematic intonation, two had a 'breathy' tone and two produced a noticeably 'nasal' sound. However, the panel agreed that these candidates had an aptitude for singing that would benefit from further training and would improve over time. Unsuccessful voices from this audition were variously described as 'sweet' or 'pleasing' in core vocal quality but the panel focused on a lack of 'vocal maturity'. Candidates not accepted into the course were considered not to be vocally or technically equipped for the rigours of the degree and the reports tended not to specify individual technical flaws.

Vocal intensity

Measures of mean and maximum SPL for song 1 by audition outcome group is presented in Table 2.

Table 2. Mean and max sound pressure level (SPL) and standard deviations (SD) of song 1 by gender and audition outcome.

Sex	Audition result	SPL	Number	Mean	SD
Female					
	Major	Max SPL	5	91.77	3.42
		Mean SPL	5	78.06	3.01
	Minor	Max SPL	9	94.20	5.27
		Mean SPL	9	79.10	4.06
	Not accepted	Max SPL	20	91.43	4.20
		Mean SPL	20	74.86	2.87
Male					
	Major	Max SPL	7	89.26	3.23
		Mean SPL	7	75.75	2.88
	Minor	Max SPL	3	88.91	3.39
		Mean SPL	3	76.47	2.96
	Not accepted	Max SPL	3	88.83	1.99
		Mean SPL	3	73.64	1.59

Table 3 shows the contrasts between major, minor and not accepted female and male singers on mean and maximum SPL.

Table 3. Contrasts for maximum and mean sound pressure level (SPL) by gender and between audition outcomes.

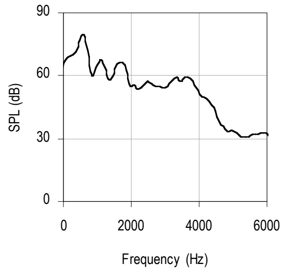
Females	Contrast	Value of Contrast	Std. Error	t (df=31)	P Value (2-tailed)
Max SPL	Accepted vs Not accepted	-3.12	3.16	-0.99	0.33
	Major vs Not accepted	0.35	2.21	0.16	0.88
	Minor vs Not accepted	-2.78	1.77	-1.57	0.13
	Major vs Minor	-2.43	2.46	-0.99	0.33
Mean SPL	Accepted vs Not accepted	-7.44	2.31	-3.22	0.00
	Major vs Not accepted	3.20	1.62	1.98	0.06
	Minor vs Not accepted	-4.24	1.30	-3.27	0.00
	Major vs Minor	-1.04	1.80	-0.58	0.57
Males				t (df=10)	
Max SPL	Accepted vs Not accepted	-0.50	4.11	-0.12	0.91
	Major vs Not accepted	0.42	2.11	0.20	0.85
	Minor vs Not accepted	-0.07	2.49	-0.03	0.98
	Major vs Minor	0.35	2.11	0.17	0.87
Mean SPL	Accepted vs Not accepted	-4.94	3.62	-1.36	0.20
	Major vs Not accepted	2.11	1.86	1.14	0.28
	Minor vs Not accepted	-2.83	2.20	-1.29	0.23
	Major vs Minor	-0.71	1.86	-0.38	0.71

Mean SPL of performances was significantly different between major/minor girls and not accepted and between the minors and not accepted groups but not between major and minor girls. Accepted females produced a greater mean intensity than not accepted females. There were no statistically significant differences between the three male groups.

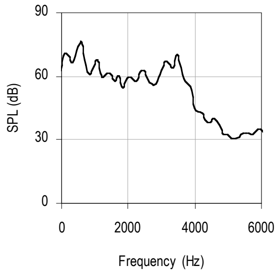
Long term average spectra

Figure 2 presents LTAS performed on selected of song 1 for females and males. Each figure shows LTAS from the performances of four candidates accepted into the course (2 majors, 2 minors) and two candidates not accepted into the course for both males and females. Below each LTAS exemplar is the corresponding assessment from the audition panel including the singer's overall perceptual rating (%), group designation (to major, minor or not accepted) and voice type. The panel's comments illustrate their response to vocal and performance features that shaped their judgments (e.g. vocal quality, technique and overall musical performance). From the adjudicators' comments, acoustic cues appeared to be a critical influence on audition outcome and it was hypothesized that LTAS may provide further clarification of the acoustic signal.

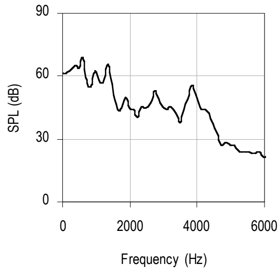
Specifically, we looked for common spectral peaks in these LTAS of singers' first song performances that were common to gender and to each group designation (Major, Minor, NA). The acoustic cues will have been influenced by the song stimuli and the presence of piano in the recording. LTAS therefore may provide indicators of broad timbral similarities or differences between performances. High ranked singer's performance LTAS showed strong spectral reinforcement above 2 kHz. Male singers in particular showed a prominent clustering of energy above 2 kHz although this energy was also evident in minor and NA males (Figure 2). Accepted females showed reinforcement in spectral peaks above 2 kHz (Figure 2), although lower ranked females' LTAS plots showed clearer individual peaks of energy which may be representative of harmonics that occur during the song line and accompaniment.



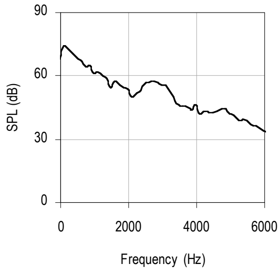
Highest ranked major soprano - 96%
A good instrument and there is obvious performing flair here. Tendency to darken sound. Tremolo detracts. Overblows instrument which diminishes natural beauty in this voice. Lots of promise here.



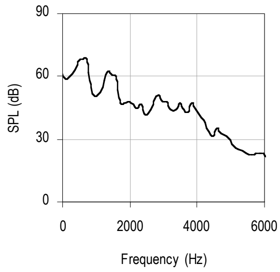
Highest ranked major tenor - 95%
Enormous potential in this tenor voice. Sings 'note to note' at this stage… when he learns to sing legato this will be very special.



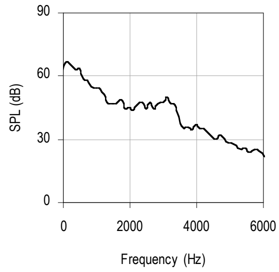
Highest ranked minor soprano – 80%
A lovely clean free flowing soprano. Good presentation.



Highest ranked minor baritone – 80%
A very musical approach and an obvious understanding of the complexities of the text. More technical background would enable the imaginative interpretative ideas to be more fulfilled.



Highest ranked NA soprano – 40%
The voice has sweet tone although under developed at this stage. Sincere and appealing. Technical work needed. Very pleasing presentation.



NA baritone – 40%
A pleasing quality instrument. Voice can have more consolidated connection with the breath. Perhaps at this stage, does not have the vocal maturity to accommodate the course demands.

Figure 2. Highest ranked females and males in each major, minor and not accepted group. LTAS of song 1, audition score and judges' comments.

Discussion

Mean SPL levels differentiated performances of accepted and non-accepted female candidates but not those of accepted and non-accepted male candidates. Visual inspection of LTAS indicated different timbral or acoustic cues between the three group designations of singers to major, minor and not accepted (Figure 2). Female singers accepted into the degree program (major/minor) produced a statistically greater mean SPL in their performance than those not accepted and vocal minors were significantly louder in performance than those not accepted. However, there was no difference between maximum SPL between females across groups and there were no differences in either SPL measure in males' performances. The top ranked male (tenor) and female (soprano) produced the greatest mean SPLs. However, three female voice minors achieved the same high mean SPLs. The loudest female performances were all associated with technical flaws, such as 'overblown' or 'pushed' singing. The not accepted female group may have had insufficient control of vocal 'loudness' or inability to sustain a sufficient mean SPL over the duration of the song (Mürbe et al., 1999).

F0 changes also account for variation in SPL of singers and these may be larger than those due to musical dynamics (Lamarche & Ternström, 2008) and differences in vocal ability between these singers. The top of a singer's F0 range will have the highest SPLs irrespective of musical nuance, and these spectra will tend to dominate the LTAS. A future study should attempt to eliminate the dependency on F0 of SPL by including F0 as a covariate of SPL in the statistical model, which may increase accurate discrimination of different voices. By eliminating the effect of F0 mathematically, the remaining differences in SPL between singers may become clearer.

There may be factors other than intrinsic vocal ability that influenced the loudness or spectrum of the singers, such as variations in song selection and consequently the accompaniment which was not fixed for acoustic comparison. Because this was an opportunistic study, the usual controls over setting, songs and accompaniment could not be achieved. We were able to ascertain, nonetheless, that singers' song choices were not associated with differences in mean SPL. High mean SPLs were achieved equally by performances of lieder, vocalize, Neapolitan song and operatic arias. In fact, singers who performed operatic arias generally achieved a lower mean SPL in their performances (around the mean SPL for their major or minor group). Further, the recordings were made with the microphones in the diffuse field in order that the LTAS were similar to that of the sound to which the judges were exposed. This placement of microphones meant that the effect of the room acoustics on the LTAS could not be controlled, although it was the same for all singers in this study, and that these LTAS cannot be directly compared to LTAS of other studies.

Visual inspection of the LTAS curves showed a similarity in shape, with prominent peaks above 2 kHz, particularly for males and accepted females. All males' performances demonstrated a degree of high energy boost between 2-4 kHz. The different voice types (tenor and baritone) demonstrated expected ranges for this high range energy (< 3kHz for baritones and ≥3kHz for tenors) (Dmitriev & Kiselev, 1979; Sundberg, 2001). LTAS of accepted females' performances also demonstrated a distinctive energy above 2 kHz, which is consistent with previous research that showed that female singers with excellent vocal technique produced more unified peaks above the remainder of the spectrum than those with poor technique (Barnes et al., 2004; Kenny & Mitchell, 2006; Mitchell & Kenny, 2004a). It should be noted that LTAS reflects characteristics of the entire performance only if there are no prominent long and/or loud/high notes in the song. The differences between singers that are manifest in the LTAS are likely to represent the differences between the top notes of the singers. Therefore, it is important to design the musical selection such that individual high notes will not dominate the result. Because singers sang their own selections, we had no control over the number of such notes in the song selection. However, songs for audition are generally chosen to reflect the technical skill and full vocal and dynamic range of the voice.

Professional classical and operatic singers and singers in training demonstrate a key characteristic spectral energy above 2 kHz (Kenny & Mitchell, 2006; Omori et al., 1996; Thorpe et al., 2001; Vurma & Ross, 2000). There has been inconclusive longitudinal evidence to support the development of high range energy of voices in training (Mendes et al., 2003). Indeed, while high frequency energy may increase over time for singing students, it is not necessarily linked to an improvement in overall vocal quality (Vurma & Ross, 2000). We suggest that successful performances at audition demonstrate a degree of characteristic energy above 2 kHz. It may be possible to quantify this common spectra as a general classical or 'operatic timbre' (Davidson & Da Costa Coimbra, 2001). Given that singers may exhibit a degree of high range energy before training commences, high range energy identified in conservatoria auditions in both accepted and not accepted candidates provides an interesting foundation for future longitudinal research that follows singers through tertiary level singing training, using more rigorous techniques (Cabrera, Davis, Barnes, Jacobs, & Bell, 2002) and recording environments than those available in naturalistic settings such as auditions or performances.

There were no statistically significant differences in audition outcome based on day of audition or time of day of audition. The audition panel in this study arrived at their decision collectively after group discussion following each audition. Collective decision making may counter fatigue and extreme assessments over the course of extended auditions (Bergee & Platt, 2003; Elliott, 1996; Flores & Ginsburgh, 1996; Smith, 2004). However, it introduces the possibility that particular judges may exert undue influence on the group. Examination of the comments on the audition reports indicated a certain repetitiveness and stereotypy in their descriptors of vocal quality and reasoning about selection. Even for an expert audition panel, the words used to describe the best and worst voices were essentially uninformative. Previous studies have also shown that assessment of vocal quality is difficult to articulate or itemise (Davidson & Da Costa Coimbra, 2001; Robison, Bounous, & Bailey, 1994; Stanley et al., 2002; Watts et al., 2003). Studies that have attempted to extract different elements of the vocal performance for assessment have found that all elements converged on overall quality (Ekholm et al., 1998).

It may be that verbal encoding is inappropriate for describing a holistic vocal quality and verbalizing descriptions of voices may reduce the ability to discriminate between voices or at least to recall the vocal characteristics of a particular voice. Davidson and Coimbra (2001) found that singing assessors used features of candidates' visual appearance or dress in their notes which effectively served to identify or remind them of singers, although these comments would not appear in the final assessment. Those adjudicators did not address the core vocal quality of the voice and suggested that judges may have accepted the vocal quality as a 'stable element' and without it, candidates would not have been admitted to the course (Davidson & Da Costa Coimbra, 2001). This audition panel used an analytical approach to identify the technical flaws in voices within each category (Major, Minor, NA) after their judgments of intrinsic vocal qualities determined the candidate's audition result. They described specific vocal flaws, areas for vocal improvement, and the techniques needed to address these (Mitchell & Kenny, 2006; Mitchell, Kenny, Ryan, & Davis, 2003).

Authentic performance and assessment settings present a challenge to music researchers. The performance setting ensures that candidates perform at their optimal level and present songs that promote their vocal and technical ability (Legge, 2001). In such settings, it is not possible or appropriate to introduce unfamiliar protocols such as blind auditioning (Goldin & Rouse, 2000), or use recording methodology better suited to precise vocal measurement. It may, however, be possible to require candidates to sing the same pre-advised vocal exercises without piano accompaniment for measurement and comparison of basic vocal parameters. This would ensure the measurements are not affected by musical task or high notes within the stimuli.

For young classical singers, the audition process is an unavoidable reality in determining their future musical career (Legge, 2001). Observing the singing audition informed our understanding of the way experts listen to and describe singing voices. The next challenge is to replicate these findings in a more controlled acoustic environment and refine measurement and assessment of vocal quality in accordance with pedagogical assessments of singers.

Acknowledgements

This project was funded by an ARC Discovery Grant [DP0558186] to Dianna Kenny, Helen Mitchell, Densil Cabrera and Michael Halliwell. Sincere thanks to the singing audition panel from the Vocal Unit, Sydney Conservatorium of Music, University of Sydney, to Peter Thomas and John Bassett for acoustic and recording advice, to Dr Sally Collyer for her suggestions and advice, and to the anonymous reviewers for their detailed and helpful comments and recommendations.

References

Akerlund, L., & Gramming, P. (1994). Average loudness level, mean fundamental frequency, and subglottal pressure: comparison between female singers and nonsingers. *Journal of Voice, 8*(3): 263-270.

Barnes, J. J., Davis, P. J., Oates, J., & Chapman, J. (2004). The relationship between professional operatic soprano voice and high range spectral energy. *Journal of the Acoustical Society of America, 116*(1): 530-538.

Bergee, M. J., & Platt, M. C. (2003). Influence of selected variables on solo and small-ensemble festival ratings. *Journal of Research in Music Education, 51*(4): 342-353.

Boersma, P., & Weenink, D. (2006). Praat: doing phonetics by computer (Version 4.5.04).

Brown, W. S. Jr., Rothman, H. B. and Sapienza, C. M. (2000). Perceptual and acoustic study of professionally trained versus untrained voices. *Journal of Voice, 14*(3): 301-309.

Cabrera, D., Davis, P., Barnes, J., Jacobs, M., & Bell, D. (2002). Recording the operatic voice for acoustic analysis. *Acoustics Australia, 30*(3): 103-108.

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.

Dmitriev, L., & Kiselev, A. (1979). Relationship between the formant structure of different types of singing voices and the dimensions of supraglottic cavities. *Folia Phoniatrica, 31*(4): 238-241.

Dugger, R. (1997). Inter-judge reliability for the 1994 Oklahoma all-state band auditions based on an olympic style judging system. *Journal of Band Research, 32*(2): 66-75.

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.

Elliott, C. A. (1996). Race and gender as factors in judgments of musical performance. *Bulletin of the Council for Research in Music Education*, *127*(Winter): 50-56.

Elliott, C. A., Schneider, M. C., & Zembower, C. M. (2000). Influence of the audition hour on selection to an all-state band. *Journal of Band Research, 35*(2): 20-31.

Fletcher, H., & Munson, W. (1933). Loudness: its definition, measurement, and calculation. *Journal of the Acoustical Society of America, 5*(2): 82-108.

Flores, R. G., Jr., & Ginsburgh, V. A. (1996). The Queen Elisabeth musical competition: how fair is the final ranking? *The Statistician, 45*(1): 97-104.

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.

Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: the impact of "blind" auditions on female musicians. *American Economic Review, 90*(4): 715-741.

Hollien, H. (1993). That golden voice-talent or training? *Journal of Voice, 7*(3): 195-205.

Kenny, D. T., & Mitchell, H. F. (2006). Acoustic and perceptual appraisal of vocal gestures in the female classical voice. *Journal of Voice, 20*(1): 55-70.

Lamarche, A., & Ternström, S. (2008). An exploration of skin acceleration level as a measure of phonatory function in singing. *Journal of Voice, 22*(1): 10-22.

Legge, A. (2001). *The Art of auditioning: a handbook for singers, accompanists and coaches*. London: Peters.

Löfqvist, A. (1986). The long-time-average spectrum as a tool in voice research. *Journal of Phonetics, 14*: 471-475.

Löfqvist, A., & Mandersson, B. (1987). Long-time average spectrum of speech and voice analysis. *Folia Phoniatrica et Logopedica, 39*: 221-229.

McCoy, S. J. (2004). *Your voice, an inside view: Multimedia voice science and pedagogy*. Princeton, N.J.: Inside View Press.

Mendes, A. P., Rothman, H. B., Sapienza, C., & 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.
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.
_________. (2004b). The impact of "open throat" technique on vibrato rate, extent and onset in classical singing. *Logopedics Phoniatrics Vocology, 29*(4): 171-182.
_________. (2006). Can experts identify "open throat" technique as a perceptual phenomenon? *Musicae Scientiae, X*(1): 33-58.
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.
Mürbe, D., Sundberg, J., Iwarsson, J., Friedmann, P., & Hofmann, G. (1999). Longitudinal study of solo singer education effects on maximum SPL and level in the singers' formant range. *Logopedics Phoniatrics Vocology, 24*(4): 178-186.
Mürbe, D., Zahnert, T., Kuhlisch, E., & Sundberg, J. (2007). Effects of professional singing education on vocal vibrato-a longitudinal study. *Journal of Voice, 21*(6): 683-688.
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.
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.
Smith, B. P. (2004). Five judges' evaluation of audiotaped string performance in international competition. *Bulletin of the Council for Research in Music Education* (160): 61-69.
Stanley, M., Brooker, R., & Gilbert, R. (2002). Examiner perceptions of using criteria in music performance assessment. *Research Studies in Music Education, 18*(1): 43-52.
Subotnik, R. F. (2003). Adolescent pathways to eminence in science: lessons from the music conservatory. In P. Csermely & L. M. Lederman (Eds.), *Science Education. Talent Recruitment and Public Understanding* (pp. 295-301). Amsterdam: IOS Press.
_________. (2004). Transforming elite level musicans into professional artists: a view of the talent development process at the Juilliard School. In L. V. Shavinina & M. Ferrari (Eds.), *Beyond knowledge: extracognitive aspects of developing high ability* (pp. 137-167). Mahwah, MJ, London: Erlbaum Associates.
Sundberg, J. (1974). Articulatory interpretation of the "singing formant". *Journal of the Acoustical Society of America, 55*(4): 838-844.
_________. (1987). *The science of the singing voice*. DesKalb, Illinois: Northern Illinois University Press.
_________. (2001). Level and center frequency of the singer's formant. *Journal of Voice, 15*(2): 176-186.
Ternström, S. (2005). Does the acoustic waveform mirror the voice? *Logopedics Phoniatrics Vocology, 30*(3-4): 100-107.
Thorpe, C. W., Cala, S. J., Chapman, J., & Davis, P. J. (2001). Patterns of breath support in projection of the singing voice. *Journal of Voice*, 15(1): 86-104.

Vurma, A., & Ross, J. (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.
Watts, C., Barnes-Burroughs, K., Andrianopoulos, M., & Carr, M. (2003). Potential factors related to untrained singing talent: a survey of singing pedagogues. *Journal of Voice, 17*(3): 298-307.
Watts, C., Barnes-Burroughs, K., Estis, J., & Blanton, D. (2006). The singing power ratio as an objective measure of singing voice quality in untrained talented and nontalented singers. *Journal of Voice, 20*(1): 82-88.

[1] This Music Workshop at the Sydney Conservatorium of Music is used for concerts and opera school productions (http://www.music.usyd.edu.au/talent_facilities/musicws.shtml).
[2] www.praat.org