Can genre Be 'Heard' in Scale as Well as Song Tasks?

An exploratory study of female singing in Western Lyric and Musical Theater styles

Gillyanne Kayes, Independent Researcher, Powys, UK
Graham F. Welch, Department of Culture, Communication and Media, UCL Institute of Education, University College London, UK

Summary: Objectives. Using an empirical design, this study investigated perceptual and acoustic differences between the recorded vocal products of songs and scales of professional female singers of classical Western Lyric (WL) and non-legit Musical Theater (MT) styles.

Methods. A total of 54 audio-recorded samples of songs and scales from professional female singers were rated in a blind randomized testing process by seven expert listeners as being performed by either a WL or MT singer. Songs and scales that were accurately perceived by genre were then analyzed intra- and inter-genre using long-term average spectrum analysis.

Results. A high level of agreement was found between judges in ratings for both songs and scales according to genre. ($p<.0001$). Judges were more successful in locating WL than MT but accuracy was always >50%. For the long-term average spectrum analysis intra-genre, song and scale matched better than chance. The highest spectral peak for the WL singers was at the mean fundamental frequency, whereas this spectral area was weaker for the MT singers, who showed a marked peak at 1kHz. The other main inter-genre difference appeared in the higher frequency region, with a peak in the MT spectrum between 4 and 5kHz—the region of the "speaker's formant".

Conclusions. In comparing female singers of WL and MT styles, scales as well as song tasks appear to be indicative of singer genre behavior. This implied difference in vocal production may be useful to teachers and clinicians dealing with multiple genres. The addition of a scale-in-genre task may be useful in future research seeking to identify genre-distinctive behaviors.

Key words: Female singing - Genre differences – LTAS – Perception – Classical - Musical Theater

INTRODUCTION

The vocal performance styles of Musical Theater (MT) singing and Western Lyric (WL) singing are considered to be somewhat different, evidenced especially in musicals composed post-1970. [1-4] The term Western Lyric has been used to embrace the performance practice of Western classical music across epochs and geographical regions. [5, 6] The differences between these two styles have cultural origins, in that WL music is primarily a written tradition and far older than MT, the latter emerging as distinctive form only in the 1920s. [7, 8] The performance practice and pedagogy of WL singing is thus well documented with records dating at least as far back as the 13th century. [9, 10] MT is a relatively new performance genre with varied origins: European operetta, American Vaudeville, and British Musical Comedy are amongst those noted. [11] Although strongly influenced by neighboring Contemporary Commercial Music (CCM) styles, for example, Blues, Jazz, Rock and Pop, MT is essentially theatrical in nature. [3, 12] Thus, MT is better considered a distinct genre within the umbrella of a meta-genre identified by pedagogues and practitioners as CCM.[12-14] Because of these ongoing CCM influences it may well be that no one specific mode of vocal production is predominant in MT singing. [2, 4, 15, 16] Key stylistic differences between the genres are discussed in more detail below.

In MT singing, intelligibility of text is paramount [3, 12], whereas in WL singing beauty of voice is usually considered to prevail over the needs of text. [17] Typically, in MT singing there is a shorter vowel-to-consonant relationship than in WL, and consonants are likely to be louder. [14, 18] WL singers customarily perform without the aid of microphones and are able to amplify the upper harmonics of their sound output in a region where the orchestral sound is weaker, allowing their voice to be heard above a group of musical instruments. [19] By contrast, since the late 1950s, MT singers are routinely amplified during performance and, in current practice, each artist's voice will be amplified using head microphones and mixed with the orchestral sound via a sound desk. This allows for the use of sound qualities and pitch ranges that are more conversational because there is no need for the voices to project acoustically through an orchestra. [3, 12, 17, 20] In terms of desirable voice quality, a stated goal of WL singing in the traditional *bel canto* style is that of *chiaroscuro*, a tonal quality that is perceived as having both clarity and depth. [18-21] This tonal ideal is thought to be achieved with a relatively low-held

larynx, a widened pharynx, and articulatory strategies to achieve relatively equal resonance across the vowel set, and fullness and "ring" overall in the sound. [22-26]

Taking into account the various voice categories, female singers as a collective might cover the fundamental frequency range of 160-2000Hz [27], implying use of different register mechanisms. Of these, 'chest' and 'head' registers are the most relevant to the present study. The aesthetic of WL music demands an extensive use of head register in female voice. Chest register is confined to the lower portion of the pitch-range and considered detrimental if taken above F_4-G_4. [26, 28-30] In general, the expected tessitura in MT singing for female voice is lower than in WL singing and the pitch range less wide. [3, 4, 31] In MT singing it is considered essential for female performers to be proficient in both chest and head registers. These singers are documented as taking chest register up to D_5 (590Hz) and beyond. [3, 20, 31, 32] In MT performance practice, the term 'legit' is typically used to describe a sound quality considered closest to classical head register. [4, 15, 29, 33] Voice science studies comparing the spectra of chest and head register concur that 'chest' register is characterized by a less steeply sloping spectrum envelope than 'head' register, the latter having its highest spectral energy at the fundamental frequency. [29, 34-38]

Of the substantial body of research in voice science, relatively little has been devoted to so-called 'non-classical' singing styles in female voice, including the Musical Theatre (MT) genre. [39, 40] Many existing studies of MT singing in female voice focus on the phenomenon of 'belting'. Belting is a voice quality considered to be based on chest register, apparently produced with a high laryngeal posture [4, 41]. Early single-subject studies indicated a high closed quotient of >50%. [42-44] However, in a later study of a larger subject group, Lebowitz and Baken [45] reported mean closed contact quotient to be <50% in both 'belting' and in 'legit' styles, and noted a consistently higher speed of closure in "belting". As regards laryngeal height in belting, LoVetri et al. [13] reported that four of seven female subjects lowered their larynx during belting, whereas two retained the same laryngeal height. The study does not discuss whether use of the "dark" or "bright" vowels used in its protocols made any difference to laryngeal height. Comparisons have been made between "belting" and "classical" or "operatic", as well as between "belting" and "legit", and "belting" and "neutral". [41-43, 45, 46] Other

studies have investigated chest, "chest-mix", and "head-mix", or compared "MT" style with "classical style"; similarly, "Broadway" with "Operatic". [35, 47] According to Kochis-Jennings [35], "chest-mix" is a voice quality used in commercial singing (which would include the MT genre), whereas in classical singing "head-mix" is used. In further reports of the same study of seven female singers, spectral tilt [48]was found to increase incrementally as the singers moved from "chest" to "chest-mix" and "head-mix" to "head". [49, 50]

To meet the distinctive performing expectations of these two contrasting genres – WL and MT– different training regimes have evolved to meet their respective needs. [3, 12, 17, 33, 51] It has also been suggested that this need for a different professional preparation may well be more acute in the case of female MT singers because of the expectation that they will need to take their sung chest register higher than their WL counterparts when singing contemporary MT repertoire. [51] Additionally, because of the specific needs of their performing industry, working professional MT singers are noted as being subject to heavy vocal loading, which might imply an increased risk of vocal health problems. [52] Introspection suggests that if female MT singers use their voices differently compared to their WL counterparts, such differences might be detectable in sung products, such as music scales (as used in vocal warm-up exercises), as well as in their respective song repertoires. An investigation of both vocal and perceptual relationships of these two genre communities might assist in clinical pedagogical understanding of issues that are faced by individual singers.

RESEARCH AIMS AND METHODOLOGY

This study was exploratory in nature: its empirical design allowed for investigation of vocal and perceptual relationships in a small group of singers from professional backgrounds with different musical genre biases. Drawing from established case study methodology, each singer was allowed to sing a different song for a song performance test. To further improve ecological validity of the investigation, the song products were from singers' normal and current repertoire. An important and novel aspect of the study was an additional Scale-in-Genre task. This task was designed to investigate whether or not any genre-distinctive use of voice detected in the song performances of individual singers could also be detected in their sung scale tasks. The overall aim was to gain a better understanding of similarities and differences that might exist within the group studied, which might be revealing of genre distinctive behavior. Embedded in the research design were two specific foci:

a. To ascertain from perceptual tests whether or not a singer's Intended Sung Genre could be detected in sung scale tasks as well as in songs; and
b. To examine those spectra of singers' tasks clearly identified as belonging to a particular genre for any timbral characteristics that might be genre-distinctive, and therefore indicative of differences in use of voice that are perceptually accessible.

Participants

In this exploratory study seven female professional singers with a mean age of 29 years (SD 6.46, range 20 – 38) were recorded performing song and scale tasks. All participants completed a questionnaire about their training and professional activities, which included a matrix derived from a Taxonomy of Singers used as Subjects in Scientific Research [53], according to "type" and "category", relabeled as Genre and Level for this study. Of the seven singers, two self-identified as WL and five as MT. However, of the MT singers, two had trained early and extensively in the WL style of singing before retraining for MT. Of these two singers, one was recorded three times on different occasions, performing in both WL and MT genres, the intention being that her song and scale tasks would act as 'genre blind' samples for the perceptual tests. The singer subject group, type and category are summarized in Table 1.

TABLE 1. Participant group summary: Age, Level, and Genre			
Participant	Age	Level	Genre
Singer 1	38	2	WL
Singer 2	25	4	MT
Singer 3	28	4	WL and MT
Singer 4	21	3	MT
Singer 5	20	4	MT
Singer 6	37	3 (6)	WL
Singer 7	34	3 (6)	WL and MT
Note: Where two genres are indicated, the first is the original training for that singer			

For the listening test, seven judges were selected from professional contacts of the lead author. A pre-requisite for judge selection was known expertise in working with singers of both WL and MT genres. All participant singers and judges gave their informed consent to participate in the study according to the ethical guidelines of the British Educational Research Association recommendations current at the time (https://www.bera.ac.uk/researchers-resources/publications/revised-ethical-guidelines-for-educational-research-2004, last accessed July 20, 2015).

Recording procedure and task protocols

Singers were audioirecorded using was the Laryngograph™ Speech Studio, which employs a pressure-sensitive electret omni-directional microphone with a relatively flat frequency response of +/-0.1dB 1Hz-20kHz. The microphone was positioned at a fixed distance of 30 cm from singers' mouths using a simple home-made head set. Recorded data were gathered in two stages across a period of 8 months at the Voice Clinic of Queen Mary's Hospital, Sidcup, UK. The rooms were approximately 4.6m x 6.1m and 8.2m x 7.3m and not acoustically treated. Additionally, singers were digitally imaged via fiberoptic nasopharyngoscopy, the results of which are not reported here. All singers were cleared by an otolaryngologist for vocal health.

The song task was either a song or aria from the WL canon of classical music, or an MT song composed post-1970 (Table 2). The specification of song material from the post-1970 era for the MT singers was because the musical writing and style of singing pre-1964 for this genre is thought to be closer to that of WL singing. [4, 33] The intention therefore was to avoid using 'legit' song material for the MT singers in this study. Singers self-selected their songs from their normal performance repertoire, in collaboration with the lead researcher. Immediately after her song, each singer

performed a scale, henceforward labeled as Scale-in-Genre. For this task, singers were instructed to sing the scale in the manner of the song that they had just performed. In pursuit of this aim, the researcher assisted the singers in devising a scale that matched with the key of the song, also encompassing its pitch range. Singers were allowed to choose the direction and speed of their scale tasks, and encouraged to use the phoneme that, according to their own perception, allowed for the best match to the sound quality used in their song performance. Separate identification of songs and scales according to genre was an important aspect of the study design because the musical and word cues of a song might well help listeners identify a sung product with its musical genre. Such perceptual cues are absent from scales and, if judges were able to correctly identify the respective musical genres separately in songs and scales, it was hypothesized that this would indicate an underlying genre-distinctive vocal behavior on the part of the singer. Altogether 18 distinct stimuli were generated from the 7 singers, of which Intended Genre were WL n=3 songs, 3 scales and MT n=6 songs, 6 scales.

Perceptual tests of singer genre

For the listening test, songs and Scale-in-Genre tasks were validated separately by means of a seven-point Likert-like confidence rating scale [54], using the descriptors "sure it's a classical singer" and "sure it's a musical theater singer" as main anchor points at either end of a continuum. An interim anchor was displayed at the mid-point of the rating scale, using the descriptor 'could be either'. For statistical purposes, these points were logged as 1, 4 and 7 respectively (Figure 1). The rating scheme was in paper format, returned by judges either by mail or as a scanned document by email.

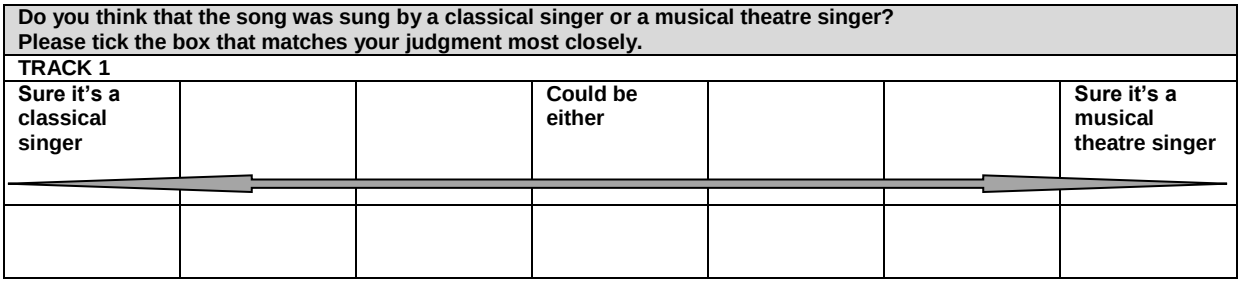

FIGURE 1. *Confidence rating scale* for perception of singer genre**.** Where judges marked the rating scale was logged as a number between **1** (leftmost point) and **7** (rightmost point).

To test for the reliability of the panel members' perceptual judgments of sung genre, each task was presented three times in a randomized order. This procedure was followed for both songs and scales,

which were presented separately in the listening tests: first the songs, then the scales. Together, the songs and scales provided 54 vocal products as stimuli for perceptual ratings according to perceived genre. Because the singers had performed different songs, the individual sound files were of different lengths, ranging from 21 to 41 seconds (songs) and from 10 to 47 seconds (scales). The stimuli gathered from all singers' songs and scales are summarized in Table 2, including the phonemes used by each singer for her scale task.

TABLE 2. Stimulus summary: Song extracts, and Scale Phonemes

Stimulus	Song extract	Self-Selected Scale Phoneme
S1	*Qui d'amor* G.F Handel	/a:/ and /i:/
S2	*Summer in Ohio* Jason Robert Brown	/ja:/and /i:/
S3	*My Husband Makes Movies* Yeston	/a:/ and/ /ri:/
S4	*Wade in the water* Trad. Arranged, Cassidy	/ba/ and /i:/
S5	*There's a Fine, Fine Line* Lopez, Marx	/la:/
S6	*Widmung* R. Schumann	/ri:/ and /a:/
S7	*What a Feelin'* (Flashdance) Moroder, Forsey, Brown	/jeə/
S8	*Caro Nome* G. Puccini	/a:/ and /i:/
S9	*Still Hurting* Jason Robert Brown	/a:/ and /i:/

Judges' ratings by genre for each recorded vocal product were then collated and entered into the statistical analysis software *SPSS Statistics for Windows Version 18* (SPSS Inc., Chicago IL) for tests of judge reliability, and inter-rater agreement. Following these tests, summaries were made of judges' success in locating singers according to Intended Genre, and descriptive tests were used to assess which singers' tasks had been perceived as clear cases of either genre (WL or MT).

Acoustic analyses of sung vocal products

From the perceptual tests of genre, mean scores were calculated for each participant, first for song and scales together, then for songs and scales separately. Singers' products that were rated as clearly agreed cases of genre were selected for acoustic analysis. After removal of silences and pauses from the sound files, long term average spectrum (LTAS) analysis was performed via the *Soundswell Signal Workstation* (Soundswell Music Acoustics HB, Stockholm, Sweden(. [55] LTAS computes the mean sound energy in different frequency bands of relatively long sound samples, averaging phonemes and frequency across the sample [56, 57] and is thus suitable for assessment of timbre and long-term settings such as singer voice quality. To take into account the different frequency distributions and different musical keys used by singers across the data set, a relatively wide bandwidth of 300Hz was used for the spectrographic analysis, with a cut-off point of 8kHz for the upper end. Because the pitch range of the song tasks was between G_3 and G_5, the bandwidth was ample to allow for the behavior of interest, that is, from the first to the seventh partial. Amplitude values of the spectral peaks were normalized to allow for different loudness levels between the singers by assigning a zero decibel value to the highest partial in each case. Using this procedure, it was possible to assess spectral similarities between song and scales intra-singer, and between singers inter-genre. Spectral differences between the genres were then calculated from mean frequency and sound pressure values of LTAS for all clear cases of WL and MT songs and scales respectively.

RESULTS

Perceptual tests of singer Intended Genre

Test reliability

Test reliability was assessed using Cronbach Alpha (α), which provides a correlation coefficient of internal consistency. The Cronbach α also affords a global reliability analysis, because test-retest is treated as if from a different judge. Using this established measure, it was possible to investigate and assess the quality of the responses overall. An excellent level of consistency of ratings across songs and scales was found between the participant judges: Cronbach α = 0.961 ($p<.0001$).

Judge agreement

Judge agreement was tested via Kendall's W analysis. Kendall's W is a test of relationships between multiple cases: it evaluates the degree of inter-rater reliability and agreement between ranked data and subjects or objects. The results of Kendall's W test yielded a correlation coefficient of .821 ($p<.0001$), indicating high level of agreement between the seven judges on their ranking of singers according to either WL or MT genres.

Judges' accuracy

Judges' success rates in locating songs and scales according to singers' Intended Genre were calculated using an un-weighted binary right-wrong assessment of each task from each judge; thus a score of 0 for incorrect identification, and a score of 1 for correct identification of a singer's Intended Genre. The determinative factor for right-wrong judgment was the positioning of each judge's response to either the right or left of the mid-point rating 4 (could be either), as shown in Figure 1.

A summary of judges' success in locating singers according to Intended Genre is given in Table 3a,b. Note that in Table 3a, a grand total of 18 would represent 100% success rate across 9 distinct song tasks and 9 distinct scale tasks for genre WL, whereas – for genre MT – a 100% success rate would be represented by a grand total of 36 (Table 3b).

The results indicated that all judges were more successful in locating WL singers than MT singers, with four judges gaining 100% WL accuracy in this respect. Judges were somewhat less

confident in locating the MT singers by genre, with ratings for some singers' tasks tending towards the middle ('could be either') and a few inaccuracies. However, across the scores of songs and scales for both WL and MT genres, no judge achieved <50% accuracy.

Judges' (n=7) success rates were next calculated by stimulus (n=54), using the same un-weighted binary right-wrong assessment as shown in Table 3c. In this case, 100% accuracy in identifying a singer's Intended Genre in all 3 runs of both tasks (21 songs + 21 scales) would be represented by a grand total of 42. Factored in this manner, judges' success rates were found to be different from those calculated across all tasks (Tables 3a,b.): 0% accuracy was achieved for stimulus 3 scale task and scores for the song task stimulus 9 were less than chance (40.5%), indicated by the two arrows (Table 3c). It should be noted that these two singers had each trained first in the WL genre, then MT.

Means of each singer each task according to perceived genre

Following the results of the perceptual tests, mean scores of each judge were calculated for each participant, first for song and scales together, then for songs and scales separately, including minimum, maximum and SD values. Because the aim of the tests and further analyses at this point was to assess which *singers* had been most clearly identified by genre, it was important to code the results according to the individual participants. Thus, the 18 stimuli (9 songs and 9 scales) were now relabeled singers 1-7 (S1-7), and singer 3's tasks identified as S3A-C. This system of task and participant identification is used henceforward. A summary of the means for each singer and task is shown in the chart in Table 4, with standard deviations shown in parenthesis.

TABLE 3
(a) Judges' Success Rate (Unweighted) in Locating Song and Scale Tasks Accurately With *Factor Judge* (n=7) for *Stimuli* (n=3) Intended Sung Genre WL (Classical); **(b)** Judges' Success Rate (Unweighted) in Locating Song and Scale Tasks Accurately With *Factor Judge* (n=7) for *Stimuli* (n=6) Intended Sung Genre MT (Musical Theater);
(c) Judges' Success Rate (Unweighted) in Locating Song and Scale Tasks Accurately With *Factor Stimulus* (n=9) the Arrows Indicate the Two Tasks in Judges Scored Poorly

(a) Genre WL (Classical)

	Sum of J1	**Sum of J2**	**Sum of J3**	**Sum of J4**	**Sum of J5**	**Sum of J6**	**Sum of J8**
Scale (n=9)	9	9	9	9	9	9	9
song (n=9)	7	9	9	6	9	6	9
Grand Total	16	18	18	15	18	15	18
Success rate (100%/18)	88.89	100.00	100.00	83.33	100.00	83.33	100.00

(b) Genre MT (Musical Theater)

	Sum of J1	**Sum of J2**	**Sum of J3**	**Sum of J4**	**Sum of J5**	**Sum of J6**	**Sum of J8**
Scale (n=18)	7	13	13	8	11	13	13
Song (n=18)	13	15	12	15	12	8	15
Grand Total	20	28	25	23	23	21	28
Success rate (100%/36)	55.56	77.78	69.44	63.89	63.89	58.33	77.78

(c)

Stimulus		**scale**	**song**		**Grand Total**	**Success% (6 runs * 7 judges = max 42)**
1		21	21		42	100.0
2		18	17		35	83.3
3	→	0	12		12	28.6
4		15	21		36	85.7
5		20	21		41	97.6
6		21	13		34	81.0
7		13	14		27	64.3
8		21	21		42	100.0
9		12	5	←	17	40.5

Table 4 shows where the mean ratings for each singer's Song and Scale-in-Genre tasks matched well and vice versa. Singers 1 and 6 had been perceived by judges as clear cases of genre WL, with singers 2, 4 and 5 perceived as clear cases of genre MT. Singer 3's second song and scale-in-genre (3B WL song), had also been rated as genre WL. Note that judges appeared confused by S3A scale and S7 song, both Intended Genre MT, as indicated by the curved brackets in the table. Each of these

singers had received their initial training in the WL genre, before retraining for MT. In terms of closeness of ratings per singer between song and scale, the weakest match between scores was for S3A MT song and scale (rows 7-9).

TABLE 4.
Summary of Mean Genre Ratings *Per Participant* (n=7) and *Task (n=2)* by All Judges (n=7).

Participant	Composite Song Genre Ratings	Composite Scale-in-Genre Ratings
Singer 1	1.14 (0.38)	1.43 (0.79)
	1.14 (0.38)	1.43 (0.79)
	1.14 (0.38)	1.29 (0.49)
Singer 2	5.86 (0.69)	5.43 (1.51)
	5.71 (1.38)	6.14 (1.21)
	5.71 (1.38)	6.14 (1.07)
Singer 3A Intended genre MT	4.57 (1.62)	2.86 (1.46)
	4.71 (1.50)	3.14 (1.21)
	3.71 (1.11)	3.14 (1.21)
Singer 4	6.86 (0.38)	5.00 (1.29)
	6.71 (0.49)	6.00 (1.15)
	6.43 (0.79)	6.00 (1.15)
Singer 5	6.86 (0.38)	6.86 (0.38)
	6.57 (0.79)	6.86 (0.38)
	6.71 (0.49)	6.43 (1.13)
Singer 3B Intended genre WL	2.86 (1.68)	1.43 (0.79)
	3.14 (1.21)	1.57 (0.79)
	3.29 (1.80)	1.57 (0.79)
Singer 3C Intended genre MT	5.00 (1.83)	4.71 (1.25)
	6.00 (1.15)	4.57 (0.98)
	5.14 (1.57)	5.00 (1.00)
Singer 6	1.43 (0.79)	1.14 (0.38)
	1.43 (0.53)	1.14 (0.38)
	1.29 (0.49)	1.29 (0.49)
Singer 7	3.57 (1.62)	4.86 (1.07)
	4.57 (1.81)	4.57 (1.40)
	4.00 (1.29)	5.14 (1.07)

Note: Standard deviations shown in parenthesis; each task had been presented 3 times.

LTAS Analysis of Singers' Genre Tasks

Results of the LTAS analysis of song and scale task for each singer rated as a clear case of either WL or MT genres are shown in Figure 2. The scale used on the x-axis is logarithmic, and amplitude levels between singers have been normalized on the y-axis to show the highest spectral peak at 0dB. Except for singers S1 and S2, spectral peaks between song and scale for individual singers matched reasonably well. The differences between song and scale for S1 and S2 are most likely due to those

singers' choice of phoneme for their scale task. Scale phonemes used by the singers, together with the frequencies of the highest peak in the spectrum are shown in Table 5.

TABLE 5.
Frequency values of LTAS Peaks for the Singers Rated in the Perceptual Tests as "Clear Cases" of Genre.

Singer	LTAS Peak Song	LTAS Peak Scale	Scale Phoneme
S1	625 Hz	1250 Hz	/a:/ and /i:/
S3B	514 Hz	562.5 Hz	/ri:/ and /a:/
S6	562.5 Hz	750 Hz	/a:/ and /i:/
S2	750 Hz	1625 Hz	/ja:/ and /i:/
S4	1062Hz	876 Hz	/ba/ and /i:/
S5	1250 Hz	1250 Hz	/la:/

When inter-genre was plotted by task, results of the LTAS analysis showed that, overall, the song and scale tasks matched rather well (Figure 3A, B). A clear overall difference between the two task types (songs and scales WL and MT) was observed in the sound levels above 1kHz: higher for the MT singers than the WL singers in each case. Additionally, in the scale tasks, the MT singers showed a more noticeable lower sound level in the lower partials (to the left of the chart at approximately 375Hz), indicating a weaker fundamental frequency than the WL singers.

To demonstrate overall differences between the two genres, mean frequency and sound pressure values were calculated for each genre by task (song or scale) and plotted on a chart showing LTAS *differences* between MT and WL songs and scales (Figure 4).

In Figure 4, the most marked difference between MT and WL genres can be seen in the higher frequency region, where there is a peak in the spectrum between 4 and 5kHz that is especially noticeable in the scales. The distinct dip in the spectrum between 500 and 600Hz, represents the difference between MT and WL genres at singers' average fundamental frequencies.

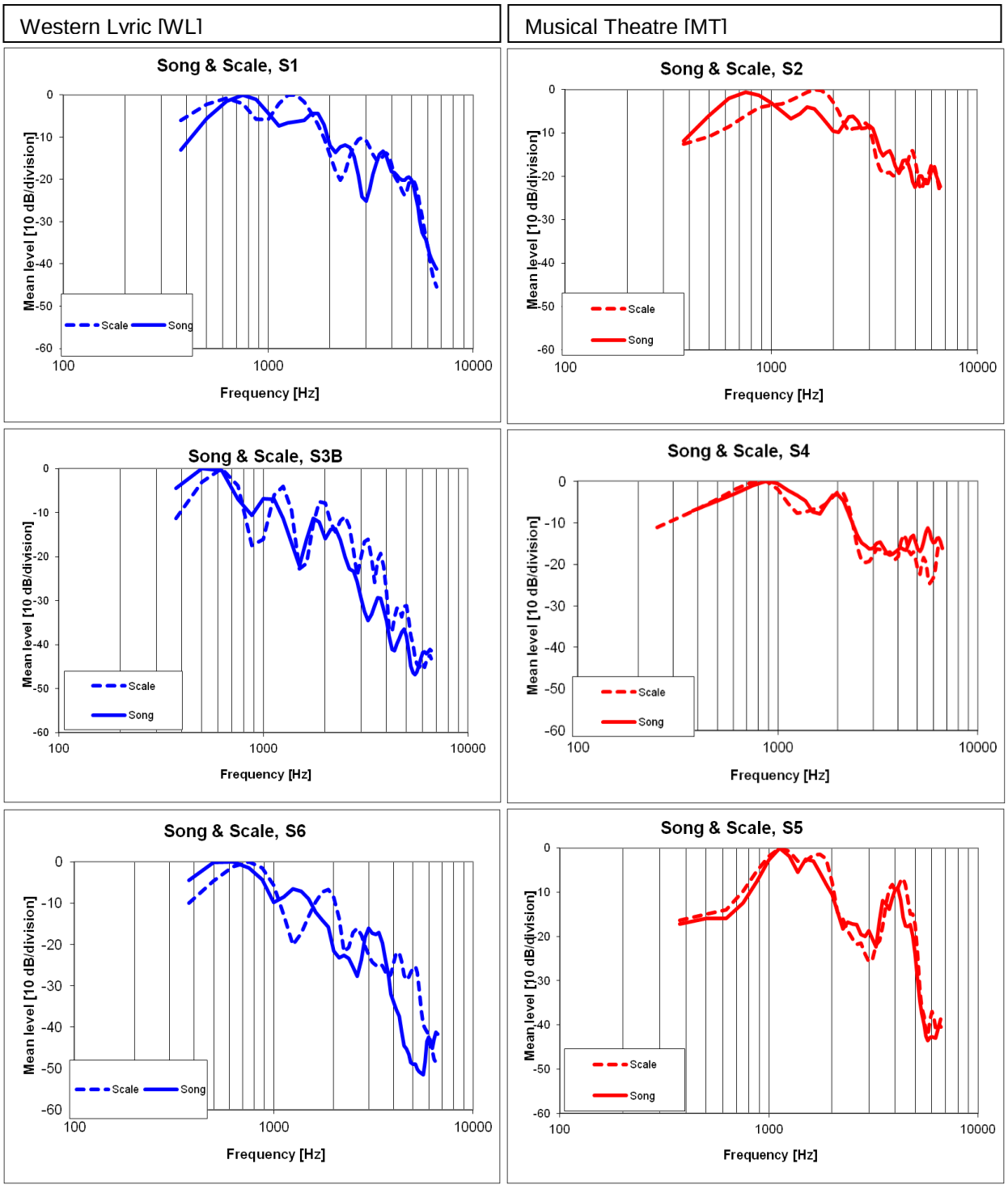

FIGURE 2: LTAS intra-singer of song (unbroken lines) and scale tasks (dashed lines); LTAS settings were bandwidth 300 Hz; upper F0 limit of 8000 Hz; FFT 1024
WL Singers are shown in blue, MT singers in red; SPL in each case on the X axis is between -60dB and 0dB with highest level for each singer normalized to 0dB.

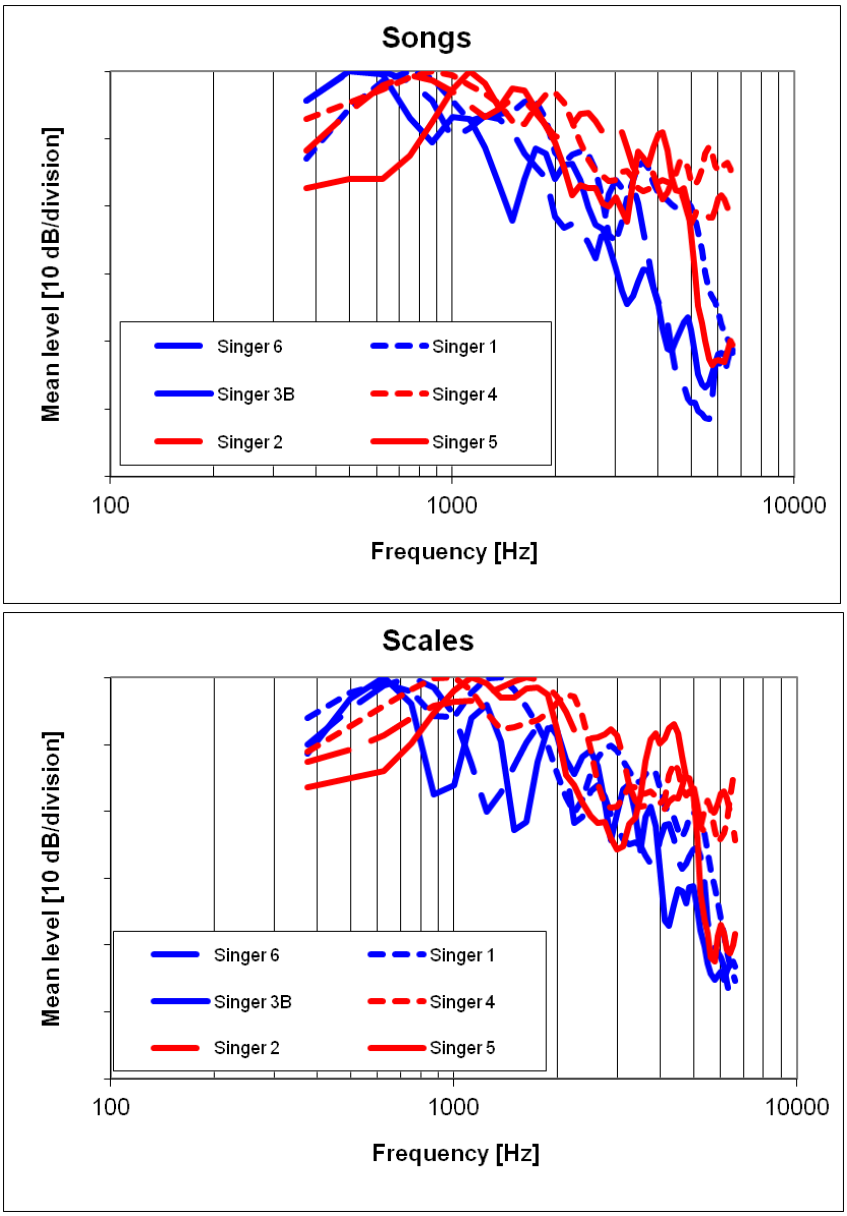

FIGURE 3. (A, B) LTAS of the two tasks inter-genre for singers assessed as clear cases of genre; **(A)** (above) songs; **(B)** (below) scales; WL = blue; MT = red; highest level for each singer is normalized to 0dB.

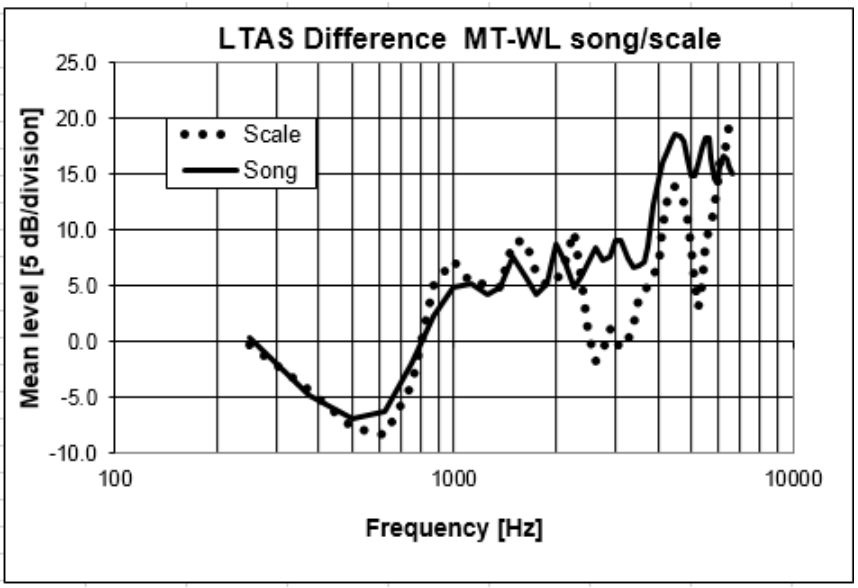

FIGURE 4. Differences LTAS of MT-WL songs (solid line) and scales (dashed line). The data series are calculated from means of the 3WL and 3 MT singers rated as clear cases of their respective genres.

DISCUSSION

The overall aim of the investigation had been to find out if genre could be heard in sung scales as well as song tasks. That being the case, it would be worthwhile to examine the spectra of the stimuli used, to find out if there was anything distinctive about the respective genres that judges were responding to.

Although the study involved a small number of participants, and the number of singers in each sub-grouping was unequal, the test protocols nevertheless produced 18 distinct vocal products and 54 stimuli. Results of intra-judge reliability, judge accuracy and inter-judge agreement in the genre-perception tests proved encouraging. In terms of accuracy, judges had shown more confidence in identifying the WL singers by genre than the MT singers, but none had scored less than 50%. The slightly lower success rate of MT over WL identification may have been due to the larger number of MT singers in the group, or it might have been due to the fact that two MT singers had trained in both genres. A separate factor to consider is the possibility that judges might have recognized or remembered individual singers' voices across the song and scale tasks. However, in the design of the present study, songs and scales were clearly separated into two groups: 27 song excerpts were

followed by 27 scales on a continuous audio track. Additionally, the judges had not been told how many voices they were listening to, so it seems unlikely that even an expert listener would be able to recognize and recall an individual singer's voice across so many excerpts. The one judge who commented in her covering e-mail that she thought Singer 2 and Singer 4 "might" be the same participant in the song tasks was a highly experienced singing teacher. That she was incorrect in this respect tends to further support the supposition that judges would not have recognized individual singer voices between the two task sets (songs and scales).

The LTAS analyses confirmed the results of the genre-perception test, showing that, where song and scale products had been rated 'clear cases' of their respective genre, their spectra were consistent intra-genre, and also distinctively different between the two genres (inter-genre). LTAS gives a reading of the time average of the spectrum and thus reflects overall sound timbre. Broadly speaking, the findings from the LTAS analyses were in good accord with existing research studies, demonstrating a main spectral peak for WL singers in the fundamental frequency region, whereas for MT singers there was a dip in that region (Figure 3A, B approximately 375Hz). Previous research had indicated that female singers of WL and MT genres have been noted as likely to use different resonance strategies, and to sing in different register mechanisms. [4, 32, 41] A strong fundamental and steep spectral slope is considered to be characteristic of head register, particularly in WL singing. [29, 35-37] In WL singing, chest register is only used across a relatively small portion of the singer's range, so this finding was unsurprising from the WL singers in the group. [26, 30, 58] MT singers have been found to have a weaker fundamental, particularly if they are singing with a higher level of activity in the adductor muscles, using thick vocal folds that are typical of chest register. [59] These differences represent an adjustment of voice quality at the sound source. Additionally, the MT singers' main peaks were above the 1-kHz region, with a marked peak in the 4- to 5-kHz region, identified in voice research studies as the "speaker's formant". [60] A number of other studies have also noted the use of speaker's formant as a characteristic of MT and other sub-genres of commercial singing styles. [18, 59, 61, 62] The significance of the peak at 1 kHz for MT singers is somewhat less clear, because there are only a few studies of female MT singing that use LTAS for analysis of higher frequency regions. In a recent study of belting in female voice, Sundberg and Thalén [63] also reported a marked spectrum

peak at 1kHz in all their subjects. The authors attributed this to an effect of the combination of melody and key in which the most frequently used pitch was C_5 (approximately 500Hz) of which the second partial is 1000Hz. However, in the current study the singers performed different songs in different keys. The effect of the 1kHz peak is more likely to have been due to a type of voice production for MT singing that uses a high first formant (F1) production. Stone et al. [47] found a high spectral LTAS level in the 800-1600Hz region in a study of a single female singing in Broadway style, and Sundberg et al. [41] noted use of high F1 in selected vowel spectra of 'chest-mix' singing, also in a single subject study. Likewise, Sundberg et al. [64] noted a high LTAS peak just below 1kHz in their study of a professional MT singer's belting styles. Therefore, it seems probable that the 1kHz spectral peak found in the MT singers was the outcome of a greater use of chest register in their singing.

A key aspect of the method used for the study was that of the Scale-in-Genre task. A scale is a type of 'genre-neutral' task in which the perceptual cues of lyrics, melodic patterns and implied harmonic structure are absent. Despite the individual differences found between song and scale LTAS for singers 1 and 2, when all songs and scales were considered together, the LTAS plots of song and scale tasks matched rather well. Thus results of both the perceptual tests and LTAS analysis support a proposition that singer genre (at least as exampled in the genre biases of this study, WL and MT), can be detected in scale tasks as well as in song repertoire tasks. This in turn tends to imply that singer use of voice in terms of overall vocal production is likely to be differently biased between WL and MT genres, at least in the case of female singers when the MT repertoire chosen is 'non-legit'. Very few comparative studies have been made of singers performing scales in the manner of their preferred performance genre. A study by Sundberg et al., [65] using only scale tasks to investigate sung genre behavior, obtained good results from perceptual tests of differences between classical and non-classical approaches to vocal tract tuning during register transitioning in the male voice. In that study, classically trained singers sang in both styles. Although such a test design undoubtedly makes for better experimental control, it might also account for the reported inter-singer variability in the formant tuning strategies used. In the present study, the intention was to address the issue of ecological validity by selecting participants on the basis of genre speciality with only one of the

singers performing in both genres. Care had also been taken to devise a "genre-contextualized" scale which matched the key of the song and included the song's pitch-range. Inevitably, this method produced a number of variables: different songs in different frequency ranges, and scale tasks performed on different phonemes. Nevertheless, together with the differences in voice quality found in the LTAS, the high levels of judge agreement between song and scale tasks tend to imply that voice quality is an important factor in musical genre identification.

It is worth considering what might have confused judges as to the genre identity of S3 and S7 in the perceptual tests. Table 6 gives a summary of the mean ratings of each of their tasks according to perceived genre.

TABLE 6.
Summary of Mean Ratings by Genre for Songs and Scales for Singers 3 and 7

Participant	Composite Song Genre Ratings	Composite Scale-in-Genre Ratings
Singer 3A MT song	4.57 (1.62)	2.86 (1.46)
	4.71 (1.50)	3.14 (1.21)
	3.71 (1.11)	3.14 (1.21)
Means per task	*4.33*	*3.04*
Singer 3B WL song	2.86 (1.68)	1.43 (0.79)
	3.14 (1.21)	1.57 (0.79)
	3.29 (1.80)	1.57 (0.79)
Means per task	*3.09*	*1.52*
Singer 3C MT song	5.00 (1.83)	4.71 (1.25)
	6.00 (1.15)	4.57 (0.98)
	5.14 (1.57)	5.00 (1.00)
Means per task	*5.83*	*4.76*
Singer 7	3.57 (1.62)	4.86 (1.07)
	4.57 (1.81)	4.57 (1.40)
	4.00 (1.29)	5.14 (1.07)
Means per task	*4.04*	*4.85*

Note: standard deviations are shown in parenthesis

S7's song task was assessed closer to WL than her scale task (song 4.04, scale 4.85) and S3's first scale task (Intended Sung Genre MT) was assessed closer to WL than her equivalent song task (scale 3.04, song 4.33). Even though S7 was a working MT singer, she had trained as a WL singer to conservatoire (higher education) level, so perhaps this result is not so surprising. Moreover, the spectral profile of her song task revealed some interesting aspects that may have influenced judges' perception of her Intended Genre. S7's song, *Still Hurting*, is typical of much contemporary MT

writing, and the section chosen by S7 begins with a dramatically loud section, followed by rather quiet singing that is approximately 12 semitones lower. When calculated to match with these two time-frames in the song, the LTAS analyses revealed that the first 9 seconds (the louder, opening section) were dominated by the sound level of the harmonics at 1 kHz, whereas in the remaining quieter 25 seconds, the sound level near 1 kHz was similar to that of the mean fundamental, as shown in Figure 5. Because a stronger fundamental is associated with WL singing in females, this longer section of the song product may well have influenced judges' ratings in comparison with song products from other MT singers.

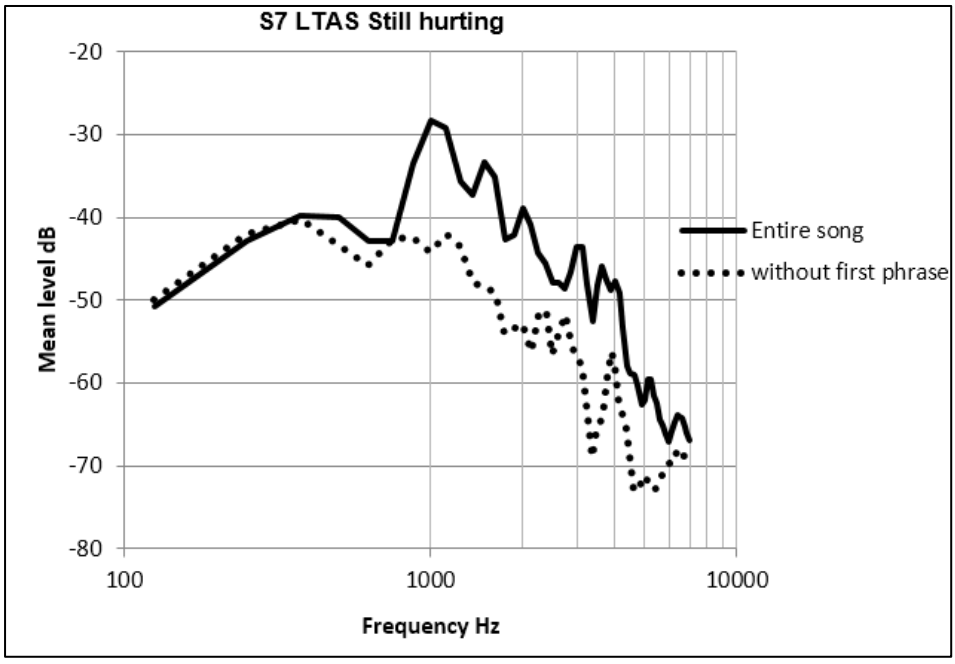

FIGURE 5.
LTAS plot of S7's song, showing the entire song (solid line) and LTAS calculated without the first phrase (dotted line).

No judge had correctly identified S3's first scale task according to Intended Genre. Mean ratings of the three repeated presentations were all < 4 ("could be either"). In this case the LTAS profile of the singer's scale task showed peaks both at the average fundamental frequency and between 1 and 2 kHz. The peak between 1and 2 kHz may have been due to a strong second formant of the front vowels /ae/ and /i:/ that the singer selected for her scale task. It is possible that the listeners responded to this vowel effect as being that of a classically trained (WL) singer. Because S7 and S3 were trained

extensively in WL singing beforeMT, this acoustic finding tends to strengthen results of the tests, rather than the opposite. We might expect that the boundaries of perception would be less distinct in cases where the original training was in WL singing.

It was not the purpose of this study to discover which singers were the most skilled, or most comfortable in terms of their genre performance. However, judges' confusion as to the genre of MT singers trained originally in the WL style does raise a potential issue of how to ensure genre authenticity in voice research studies. MT is a sub-genre of CCM singing and many pedagogues appear to agree that training for CCM needs to be different from that of a WL singer. [3, 4, 33, 51, 66] This need may be more acute in the case of female MT singers, in that a more extensive use of chest register is a requirement where the material sung is "non-legit." [12, 20, 32] In tightly controlled studies that either use the same song material sung in different styles, or the same singers performing in different styles, it is possible that some distinctive aspects of vocal behavior from the comparative genres might not be displayed. An arguable strength of the current study is that the design sought to ensure as far as possible the genre integrity of the vocal products, by allowing the participants to select and perform their own choice of song material and scale phoneme.

CONCLUSIONS

In this study, expert listeners were able to identify singer Intended Genre separately in both song and scale products of WL and MT singers, where the latter were contemporary in style. Overall, the listeners showed greater confidence in identification of Intended Genre WL than MT, but none scored less than chance. Listeners were confused as to Intended Genre of two sung products from MT singers who had earlier trained extensively in WL singing. LTAS comparison between the two genres confirmed the similarities in intra-genre acoustic profile that were reported in the perceptual data, as well as the acoustic differences inter-genre. These results imply a habitual difference in the patterned use of voice between the two genres studied: WL singing was characterized by a stronger fundamental, typical of singing in head register, whereas the fundamental frequency in MT singers tended to be weaker, reported elsewhere to be a characteristic of singing in chest register. MT singers were found to use the speaker's formant region as a resonance strategy, also showing their most prominent spectral peak at approximately 1 kHz, most likely due to a high first formant. It should be borne in mind that, using a case study approach to the genres of interest, the participant group for this study was small. It is also acknowledged that these differences found between the two genres might not have been so marked had the MT singers been performing 'legit' song material. Findings therefore cannot be considered determinative, but offer useful indicators for future research. As far as we know, this was the first empirical study in which song and scale tasks were compared inter and intra genre. The outcome of the perceptual tests and acoustic analyses implies that a Scale-in-Genre task may well be a useful protocol in research studies that seek to characterize different sung genres.

Acknowledgements

The authors gratefully acknowledge the assistance of Professor Johan Sundberg with preparation of the LTAS procedure and LTAS interpretation. Statistical procedures were performed by Dr Evangelos Himonides. Thanks are also due to the singers and judges who gave their time and expertise to this research.

1. LoVetri J. Letters to the Editor. Vol 69. Journal of Singing: National Association of Teachers of Singing; 2012.

2. Meylan M. Pop Music Is In: The Changing Sound of Musical Theatre The Voice Foundation; 2004.
3. Melton J. *Singing in Musical Theatre*. New York: Allworth Press; 2007.
4. Balog JE. A guide to evaluating music theater singing for the classical singing teacher. *Journal of Singing.* 2005;61:401-406.
5. Castellengo M, Chuberre B, Henrich N. Is Voix Mixte, the vocal technique used to smooth the transition across the two main laryngeal mechanisms, an independent mechanism? . *International Symposium on Musical Acoustics*. Nara, Japan2004.
6. Prame E. Vibrato extent and intonation in professional Western lyric singing. *The Journal of the Acoustical Society of America.* 1997;102:616-621.
7. Middleton R. Reading Pop: Approaches to Textual Analysis in Popular Music. Oxford: Oxford University Press; 2000.
8. Potter J. *Vocal Authority*. Cambridge: Cambridge University Press; 1998.
9. Duey P. *Bel Canto in Its Golden Age*. New York: King's Crown Press; 1951.
10. Large J. Acoustic Study of Register Equalization in Singing. *Folia Phoniatrica et Logopaedica.* 1973;25:39-61.
11. Gänzl K, Lamb A. *Gänzl's Book of the Musical Theatre*. London: The Bodley Head; 1988.
12. Fisher J, Kayes G, Popeil L. Pedagogy of different sung genres. In: Nix J, Howard DM, Welch GF, eds. *The Oxford Handbook of Singing*: Oxford University Press; 2014.
13. LoVetri JL, Lesh S, Woo P. Preliminary study on the ability of trained singers to control the intrinsic and extrinsic laryngeal musculature. *Journal of Voice.* 1999;13:219-226.
14. Banfield S. Stage and screen entertainers in the twentieth century. In: Potter J, ed. *The Cambridge Companion to Singing.* Vol 1. 3 ed. Cambridge, UK: Cambridge University Press; 2001:63-82.
15. LoVetri JL, Means Weakly E. Contemporary Commercial Music (CCM) survey: Who's teaching what in non-classical music. *Journal of Voice.* 2003;17:207-215.
16. LoVetri JL, Means Weakly E. Follow-Up Contemporary Commercial Music (CCM) Survey: Who's Teaching What in Nonclassical Music. *Journal of voice : official journal of the Voice Foundation.* 2008.
17. Callaghan J, Emmons S, Popeil L. Solo Voice Pedagogy. In: McPherson G, Welch GF, eds. *The Oxford Handbook of Music Education.* Vol 1. Oxford: Oxford University Press; 2012.
18. Sundberg J, Romedahl C. Text Intelligibility and the Singer's Formant - A Relationship? *Journal of Voice.* 2009;23:539-545.
19. Sundberg J. Research on the singing voice in retrospect. *TMH-QPSR.* 2003;45:011-022.
20. LoVetri JL. Female chest voice. *Journal of Singing.* 2003;60:161-164.
21. Stark J. *Bel Canto A History of Vocal Pedagogy.* 2 ed. Toronto Buffalo London: University of Toronto Press; 2003.
22. Shipp T. Vertical laryngeal position: Research findings and application for singers. *Journal of Voice.* 1987;1:217-219.
23. Sundberg J. Articulatory interpretation of the "singing formant". *Journal of the Acoustical Society of America.* 1974;55:838-844.
24. Hertegård S, Gauffin J, Sundberg J. Open and covered singing as studied by means of fiberoptics, inverse filtering, and spectral analysis. *Journal of Voice.* 1990;4:220-230.
25. Dromey C, Heaton E, Hopkin JA. The Acoustic Effects of Vowel Equalization Training in Singers. *Journal of Voice.* 2011;25:678-682.
26. Miller R. *The Structure of Singing: System and Art in Vocal Technique*. New York: Schirmer; 1986.
27. Smith J, Wolfe J, Henrich N, Garnier M. Diverse resonance tuning strategies for women singers. *Stockholm Music Acoustics Conference SMAC 2013*2013:306-310.
28. Callaghan J. The implications of voice science of the teaching of singing: vocal registers. *Australian Voice.* 1996;2:27-31.
29. Miller DG. *Registers in Singing: Empirical and Systematic Studies in the Theory of the Singing Voice*. Groningen: University of Groningen, Groningen; 2000.
30. Chapman J. *Singing and Teaching singing*. Oxford
Plural Publishing Inc.; 2006.
31. Kayes G, Fisher J. *Successful Singing Auditions*. London: A&C Black; 2002.
32. Björkner E, Sundberg J, Cleveland T, Stone E. Voice Source Differences Between Registers in Female Musical Theater Singers. *Journal of Voice.* 2006;20:187-197.
33. Kayes G. *Singing and the Actor.* 2 ed. London: A&C Black Publishing Limited; 2004.
34. Titze IR. A framework for the study of vocal registers
Journal of Voice. 1988; 2 1-12.
35. Kochis-Jennings KA. *Intrinsic laryngeal muscle activity and vocal fold adduction patterns in female vocal registers: chest,*

chestmix, and headmix. , University of Iowa; 2008.
36. Henrich N. Mirroring the voice from Garcia to the present day: Some insights into singing voice registers. *Logopedics Phoniatrics Vocology.* 2006;31:3-13.
37. Thurman L, Welch GF, Theimer A, Klitzke C. Addressing Vocal Register Discrepancies: An Alternative, Science-based theory of Register Phenomena. *Second International Conference. The Physiology and Acoustics of Singing*. Denver, Colorado, USA2004.
38. Sundberg J. *The Science of the Singing Voice*. Dekalb, Illinois: Northern Illinois University Press; 1987.
39. Wilson P. Sinful Modern Music: Science and the Contemporary Commercial Singer. *Australian Voice* 2003;9:12-16.
40. Bourne T, Garnier M, Kenny D. Music Theatre Voice: Production, Physiology and Pedagogy. In: Harrison SD, ed. *Perspectives on Teaching Singing: Australian Vocal Pedagogues Sing Their Stories*. Brisbane: Australian Academic Press; 2010.
41. Sundberg J, Gramming P, Lovetri J. Comparisons of Pharynx, Source, Formant, and Pressure Characteristics in Operatic and Musical Theatre Singing. *Journal of Voice.* 1993;7:301-310.
42. Bestebreurtje ME, Schutte HK. Resonance strategies for the belting style: Results of a single female subject study. *Journal of Voice.* 2000;14:194-204.
43. Schutte HK, Miller DG. Belting and pop, nonclassical approaches to the female middle voice: some preliminary considerations. . *Journal of Voice.* 1993;7:142-150.
44. Estill J. Belting and classic voice quality: some physiological differences. *Medical Problems of Performing Artists.* 1988;3:37-43.
45. Lebowitz A, Baken RJ. Correlates of the Belt Voice: A Broader Examination. *Journal of voice : official journal of the Voice Foundation.* 2011;25:159-165.
46. Sundberg J, Thalén M. Respiratory and Acoustical Differences Between Belt and Neutral Style of Singing. *Journal of Voice.* 2015;29:418-425.
47. Stone RE, Cleveland TF, Sundberg PJ, Prokop J. Aerodynamic and acoustical measures of speech, operatic, and broadway vocal styles in a professional female singer. *Journal of voice : official journal of the Voice Foundation.* 2003;17:283-297.
48. Fant G, Lin Q. Frequency domain interpretation and derivation of glottal flow parameters. *STL-QPSR.* 1988;29:1-21.
49. Kochis-Jennings KA, Finnegan EM, Hoffman HT, Jaiswal S, Hull D. Cricothyroid Muscle and Thyroarytenoid Muscle Dominance in Vocal Register Control: Preliminary Results. *Journal of Voice.* 2014;28:652.e621-652.e629.
50. Kochis-Jennings KA, Finnegan EM, Hoffman HT, Jaiswal S. Laryngeal Muscle Activity and Vocal Fold Adduction During Chest, Chestmix, Headmix, and Head Registers in Females. *Journal of Voice.* 2012;26:182-193.
51. Bourne T, Garnier M, Kenny D. Music Theater Voice: Production, Physiology and Pedagogy. *Journal of Singing.* 2011;67:437-444.
52. Phyland DJ, Pallant JF, Benninger MS, et al. Development and Preliminary Validation of the EASE: A Tool to Measure Perceived Singing Voice Function. *Journal of Voice.* 2013;27:544-461.
53. Bunch M, Chapman J. Taxonomy of singers used as subjects in scientific research. *Journal of voice : official journal of the Voice Foundation.* 2000;14:363-369.
54. Sergeant DC, Welch GF. Gender Differences in Long-Term Average Spectra of Children's Singing Voices. *Journal of voice : official journal of the Voice Foundation.* 2009;23:319-336.
55. Ternström S. Soundswell Signal Workstation (Version 3.4)[Computer software]. *Stockholm, Sweden: Soundswell Music Acoustics HB.* 1996.
56. Howard DM, Angus J. *Acoustics and Psychoacoustics.* 3 ed. Oxford: Focal Press; 2006.
57. Sergeant DC, Welch GF. Age-Related Changes in Long-Term Average Spectra of Children's Voices. *Journal of voice : official journal of the Voice Foundation.* 2008;22:658-670.
58. Doscher BM. *The Functional Unity of the Singing Voice*. Metuchen: Scarecrow Press Inc.; 1988.
59. Björkner E. Musical Theater and Opera Singing--Why So Different? A Study of Subglottal Pressure, Voice Source, and Formant Frequency Characteristics. *Journal of Voice.* 2008;22:533-540.
60. Leino T. Long-Term Average Spectrum in Screening of Voice Quality in Speech: Untrained Male University Students. *Journal of voice : official journal of the Voice Foundation.* 2009;23:671-676.
61. Zangger Borch D, Sundberg J. Some Phonatory and Resonatory Characteristics of the Rock, Pop, Soul, and Swedish Dance Band Styles of Singing. *Journal of voice : official journal of the Voice Foundation.* 2002.
62. Cleveland TF, Sundberg J, Stone RE. Long-Term-Average Spectrum Characteristics of Country Singers During Speaking and Singing. *Journal of voice : official journal of the Voice Foundation.* 2001;15:54-60.

63. Sundberg J, Thalén M. Respiratory and Acoustical Differences Between Belt and Neutral Style of Singing. *Journal of Voice.* 2015.
64. Sundberg J, Thalén M, Popeil L. Substyles of Belting: Phonatory and Resonatory Characteristics. *Journal of voice : official journal of the Voice Foundation.* 2012.
65. Sundberg J, Lã FMB, Gill BP. Professional male singers' formant tuning strategies for the vowel /a/. *Logopedics Phoniatrics Vocology.* 2011;36:156-167.
66. Bartlett I. One Size Doesn't Fit All: Tailored training for Contemporary Commercial Singers. In: Harrison SD, ed. *Perspectives on Teaching Singing*. Bowen Hills: Australian Academic Press; 2010.