Published in final edited form as:

J Voice. 2018 September ; 32(5): 572–577. doi:10.1016/j.jvoice.2017.08.007.

Perceptual Differences between Novice and Professional Music Theater Singers

Lynn Maxfield, Ph.D.¹ and Brian Manternach, D.Mus.²

¹Associate Director, National Center for Voice and Speech, University of Utah, 136 South Main Street, Suite 320, Salt Lake City, Utah, 84101, USA

²Assistant Professor (Clinical), Department of Theatre, University of Utah, 240 1500 E, Salt Lake City, Utah, 84112, USA

Abstract

Objective—Research examining contemporary commercial music (CCM) styles of singing has increased significantly over the last ten years. While acoustic analysis has helped determine which characteristics define various vocal genres, discrepancy still exists in how those acoustic characteristics are perceived, described, and evaluated.

Methods—The current study recorded 7 novice and 4 professional musical theatre singers performing belt, legit, and mix vocal samples. Novice singers were defined as 1st and 2nd year students in an undergraduate musical theatre program while professional singers were defined as having played at least one major role in a music theater production in regional Equity theaters. 5 regional Equity casting directors listened to the recordings and rated each sample on the basis of style (belt v. legit) and tone quality (brassy v. flutey, and bright v. dark). Results were compared across experience level (novice/professional) and pitch. Additionally, relationships between style (belt/legit) and quality (brassy/flutey, bright/dark) were examined.

Results—The statistically significant correlation emerged between the raters' perceived singing style and the singers' indicated singing style. Auditors identified the style (belt/legit) more reliably for the professional singers than for novice singers and ratings of other qualities varied significantly between raters.

Conclusions—The singers were successfully able to produce voicing styles that matched the perceptual expectations of the casting directors. Not surprisingly, professional singers were somewhat more successful in this regard than were the novices. There appears to be little consensus among the auditors, however, about which acoustical qualities define a belt, mix, or legit style.

Corresponding Author: Lynn Maxfield, 136 South Main Street, Suite 320, Salt Lake City, Utah, 84101, USA, lynn.maxfield@utah.edu, 801-596-2012.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Singing; Music Theater; Belt; Legit; Voice Perception

1. INTRODUCTION

Research examining CCM styles of singing has increased significantly over the last fifteen years. Despite the plethora of recent information, however, consensus has yet to be reached on many aspects relating to contemporary methods of vocal production. A primary source of confusion is the lack of a uniform, codified language to describe these vocal sounds, which often leads to a disconnect between those in the vocal community (singers, voice teachers, voice scientists) and those in the music theater industry (directors, choreographers, casting agents). The developing lexicon of common terminology has seen frequent use of the words “belt” and “legit” to represent ways of singing that often carry along descriptors such as bright and brassy or dark and fluty, respectively. A third way of singing, “mix,” is often described as on a spectrum between bright and dark or brassy and fluty. Though these terms are regularly used by both voice professionals and music theater industry professionals, their precise meanings prove highly subjective.

In the literature, belting is often associated with the perception of a bright timbre.¹²³⁴⁵ For instance, in one study involving casting directors and music theater majors, all of the participants reported that “belters needed a bright vocal quality to be considered elite.”⁶ In this same study, the authors note that belters do not seem to employ “cover” as they ascend through the *passaggio*, which has been associated with a slight darkening of tone, as reported by Hertegard, Gauffin, & Sundberg, 1990.⁷ Instead, belters avoid the timbral darkening that occurs as $2f_o$ crosses above F1 by modifying vowels to raise F1 as pitch increases and delay this crossing.

Another study examined five “substyles of belting,” one of which was simply referred to as “brassy.”⁸ At least one source for singing voice rehabilitation characterizes belting as both “bright” and “brassy” and, in a chapter called *Speaking Voice Therapy for Singers*, compares “safe yelling” to belting as both produce a “brassy quality.”⁹ Yet another source compared belting to yelling since both are “loud” and “brassy.”¹⁰

The current study, therefore, examines the sounds CCM singers make (specifically, female music theater singers) when asked to sing in belt, mix, and legit styles. It further examines how casting directors perceive these same sounds on a spectrum from belt to legit, bright to dark, brassy to fluty, and “no roughness” to “severe roughness.” In undertaking these tasks, the authors hope to investigate how close the perception of one group (performers) is to the perception of the other group (casting directors) when it comes to the terms belt, mix, and legit. Further, it examines whether pitch and/or experience level influences the ability of a singer to produce sounds that can be believably identified by those terms. Finally, it explores if, from the perspective of casting agents, there is a correlation with bright/dark and/or brassy/fluty to the above modes of singing.

2. MATERIALS AND METHODS

2.1 Subjects and Raters

For the study, the authors recruited four professional female singers, all of whom have performed in regional equity music theater productions, alongside seven additional female singers who were all in their first or second year of vocal study in a BFA program in either theater or music theater.

The expert auditors who analyzed the sounds are all casting directors for equity music theater productions. Though they have all been involved in professional theater in various capacities, none of them have extensive training in singing. The first auditor performed as a dancer on Broadway and in national and international tours and now serves as Executive Artistic Director and director/choreographer for a regional equity theater. The second auditor is a member of the Actor's Equity Association as a stage manager and, as Producing Director of a regional equity theater, assists in casting for musicals. The third auditor is Associate Artistic Director of a regional equity theater and serves as assistant director of their musicals. The fourth auditor serves as director and choreographer for theaters across the country, is Artistic Director of a regional equity theater, and is a voter for the Tony Awards. The fifth auditor performed as a dancer on Broadway, serves as Associate Choreographer and Production Dance Supervisor for a long-running Broadway production, and has assisted in casting for professional productions in the U.S. and in Europe.

2.2 Equipment and Procedures

The singers were recorded using a Countryman Isomax B3 (Menlo Park, CA, USA) head-mounted microphone, worn at a constant mic-to-mouth distance of 6 cm and positioned approximately 45 degrees medial to the mouth to avoid the airstream during phonation. The signal, sampled at 44,100 Hz, was amplified via an FMC (model RNP, Austin, TX, USA) pre-amp and recorded digitally using an ADInstruments (Colorado Springs, CO, USA) Powerlab digital converter with Labchart 7 pro software (proprietary software for the Powerlab).

All 11 singers were instructed to perform an ascending 5-tone major scale, sustaining the highest pitch for at least 5 seconds (indicated by the researcher). This exercise was repeated at four pitches, increasing by one semitone with each repetition, with sustained pitches at C5, Db5, D5, and Eb5. Finally, the complete sequence was repeated in three styles—belt, mix, and legit—for a total of 132 tokens (11 subjects $\times$ 4 pitches $\times$ 3 styles). Singers were instructed to perform an /a/ vowel, but to feel free to adjust as required by the pitch and/or style. They were also allowed to repeat any token if they did not feel that they had successfully performed the intended style or if there were pitch instabilities. The intent was to capture a sample of each singer's ideal belt, mix, and legit sound at each pitch.

Following recording, the researchers excerpted a steady 2.5 s segment of each sustained pitch. Loudness was normalized and a 250ms linear fade-in and fade-out was applied to each sample for listener comfort and to diminish the effect of level on rater perceptions. To the 132 original samples, 10% of the exercises were repeated, adding 13 additional recordings. The resulting 145 recordings were then randomized and uploaded as wav files to a Microsoft

PowerPoint slide presentation. The sound files were played for the auditors on a MacBook Pro laptop (Cupertino, CA, USA) and they listened through Sony stereo headphones (Minato, Tokyo, Japan), model number MDR-XD100.

In an accompanying packet, auditors were asked to rate each excerpt on a spectrum of belt to legit by placing a vertical mark on a 100 mm visual analog scale, anchored on the left by “belt” and “legit” on the right (Appendix A). In the same way, the auditors rated each tone from brassy to fluty, bright to dark, and on a scale of no roughness to severe roughness. Auditors were allowed to listen to each sound excerpt as many times as necessary in order to make a satisfactory evaluation for each category.

3. THEORY/CALCULATION

The raters’ visual analog results from all four characteristics were converted into numerical scores by measuring the distance in mm from the left end of the VAS to the point at which the rater’s tick mark crossed the scale. For example, a tick mark placed at exactly the middle of the scale would result in a score of 50. These scores were then recorded for all five raters and the mean, median, and standard deviation was calculated for each token.

Pearson’s r correlation coefficients were calculated using the raw median rater scores and the performers’ indicated style choices—belt, mix, or legit (coded as 1, 2, and 3, respectively). Additionally, median rater scores were recoded into three distinct categories corresponding to the three style choices. This was accomplished by transforming median rater scores of 0–33.33 to category 1 (belt), 33.34–66.66 to category 2 (mix), and 66.67–100 to category 3 (legit). A second correlation coefficient was then calculated using these values. Two-way interaction plots were created to identify any interaction that the moderating variables of pitch and experience level (professional or beginner) might have had on the relationship between rater scores and the singers’ indicated style choices. Finally, the reliability of the raters was assessed using all cases for inter-rater reliability and the repeated cases for intra-rater reliability. Intraclass correlation coefficients were calculated for all relationships.

4. RESULTS

The results of the analyses described above indicate that a statistically significant relationship exists between the performers’ indicated performance style and the raters’ ability to predict that style. Using the raw scores (0–100), the raters’ scores exhibited a strong positive correlation with the singers’ indicated styles ($r = 0.737$, $p < 0.01$, Table I). However, it was noted that any change in rater score, regardless of magnitude, could correlate with the change in performer style code (1–3). For this purpose, raw rater scores were recoded into three categories and the correlations were calculated again. Using these data, the correlation dropped slightly, but remained strong ($r = 0.700$, $p < 0.01$, Table II), indicating that the correlation truly is a measure of the strength of the relationship between the performers’ style choices and the ability of the rater to identify that style, robust to the magnitude of difference in rater scores necessary to differentiate between three styles across a 100-point scale.

In addition to whether or not raters could classify a sound as belt, mix, or legit, the authors also questioned what effect the performer's experience level had on the rater ability to classify the sound. Separate correlation calculations indicated that the ratings of tokens recorded by the professional population had a stronger correlation ($r = 0.846$, $p < 0.01$) than those from the novice population ($r = 0.612$, $p < 0.01$). Regression analysis also indicated the same disparity between the two populations (Figure 1), with ratings of professionals being more tightly clustered than those of the novices.

Similar analysis was conducted by subdividing the samples by pitch (Figure 2). While differences in agreement between the singers and the raters were observed across the pitch range, no clear trend appeared that would indicate that pitch had a predictable impact on the ability of a rater to identify the singer's intended performance style.

Finally, correlations between style (belt/legit) and two tone qualities (brassy/fluty and bright/dark) were calculated (Table III). This revealed a moderate correlation between ratings of bright/dark and brassy/fluty. Similarly, a moderate correlation presented between ratings of brassy/fluty and style (belt/legit). However, only a very weak correlation existed between ratings of bright/dark and style.

While variance between raters occurred, inter-rater reliability was moderate, with an interclass correlation coefficient (ICC) of 0.657. Intra-rater reliability varied considerably between raters on ratings of repeated tokens. Two of the raters exhibited low reliability while the reliability of the other three raters was moderate (Table IV).

5. DISCUSSION

Results indicate that raters were able to, with reasonable reliability, sort the three samples from each singer into their correct style choices (belt, mix, or legit). Furthermore, that sorting was more reliable among samples from the professional singers than it was for the beginning student singers. Together, these findings provide evidence that the singers were reasonably successful in matching their vocal productions to the expectations of their auditors and, intuitively, that the professional singers were more skilled in doing so than their novice counterparts.

For this experiment, pitches were chosen that purposefully pushed performers to the traditional limit of where a "belt" sound would be possible. As such it was expected that singers would have more difficulty producing tones that could be identified as "belt" as the pitch increased. However, no such effect was observed. While singers may have been uncomfortable belting at higher pitches, they were still able to produce a timbre that the raters identified as acceptable for the style.

Unlike the operatic performance realm, where undergraduate degrees in voice performance are typically followed by graduate studies and apprenticeship experiences before true professional work is expected to begin, undergraduate university programs consistently serve as a primary training ground for aspiring music theater performers before they undertake their professional pursuits. Therefore, the burden is on university faculty to adequately prepare students for the demands and expectations sought by professional casting agents.

While further research with larger sample groups is needed to make more definitive conclusions, this study indicates that singers and casting directors may be closer than sometimes believed regarding a uniform understanding of the terms belt, mix, and legit.

These results may also indicate a reasonable expectation that student singers will continue to strengthen their abilities to create distinguishable belt, mix, and legit sounds through ongoing training and the vocal maturity that comes with age—two advantages enjoyed by the professional singers in this study.

The findings may be especially significant when considering that none of the casting agents who participated in the study are trained as singers. Indeed, those in the vocal community are sometimes criticized for employing a language and standard that does not exactly mirror that of the music theater industry. The fact that the casting agents could consistently identify the intended mode of singing (from both student and professional populations) may suggest that the terms belt, mix, and legit are not only widely used but consistently understood.

6. CONCLUSIONS

Each singer was asked to perform the three styles (belt, mix, and legit) according to their capabilities and were allowed as many “takes” as necessary until they had performed to their satisfaction. There was no other guidance given from the authors, from the students’ voice teachers, or anyone else confirming that the sounds produced were successful examples of belt, mix, and legit singing. Similarly, the auditors were not asked to rate the quality of the singing, merely each singer’s ability to sing in a particular mode of voicing.

The authors acknowledge that the pitches the singers were asked to belt, depending on how each singer defines the word, is perhaps extreme. While lower pitches would likely have made a clearer distinction between styles (belt and legit, for instance), the authors were interested in seeing what would occur in the higher range of notes selected for the study, especially since the expectation of belt-quality singing in this range is occurring more and more frequently in contemporary music theater. While the raters were still able to reliably distinguish between styles at all pitches, it would be useful to examine any relationship between pitch and various acoustic parameters, which may be more sensitive to adjustments in vocal production that the singers are identifying as “uncomfortable.”

Similarly, as many of the subjects pointed out, the /a/ vowel was often not the singers’ preferred vowel for belting. However, the authors felt the /a/ vowel would be a “neutral” choice that could be applied to all three styles. Furthermore, the authors were interested in noting how, if at all, this neutral vowel may be modified as the singers executed the exercises. This analysis has yet to be completed.

It should also be noted that, while other studies have used excerpts from music theater songs to evaluate sung tones,^{6,8,11} this study used sung tones that were approached through the vocal pattern of a 5-note scale without accompaniment or orchestration. Further investigation is needed to see if there is a difference in perception of these tones when placed in the context of a song versus in an a cappella vocal exercise. An additional limitation of the study was the use of a visual analog scale that was anchored by the extremes of each metric.

Such a scale did not allow for appropriateness of each tone. For example, while a rater may have rated two samples as the same style, there was no option for the rater to identify other measures (bright/dark, brassy/fluty) as excessive. Future studies should consider including such a metric.

In light of this study, singers should take heart that it appears that the style choices they are making align with the expectations of casting directors who will be hearing their work and determining their suitability for hire. It may be valuable for novice singers to listen to working singers who appear to have better matched their vocal outputs to the qualities desired by the casting agents. Additionally, while there was moderate agreement between the raters, this study would suggest that casting directors may still have some disagreement regarding what perceptual qualities define various singing styles.

References

1. Miles B, Hollien H. Whither belting? *J Voice*. 1990; 4(1):64–70.
2. Edwin R. Belting 101. *J Sing*. 1998; 55(1):53–55.
3. Edwin R. Belting 101, part two. *J Sing*. 1998; 55(2):61–62.
4. Schutte H, Miller D. Belting and pop, nonclassical approaches to the female middle voice: Some preliminary considerations. *J Voice*. 1993; 7(2):142–150. [PubMed: 8353628]
5. LoVetri JL, Weekly EM. Contemporary commercial music (CCM) survey: Who's teaching what in nonclassical music. *J Voice*. 2003; 17(2):207–215. [PubMed: 12825653]
6. LeBorgne WD, Lee L, Stemple J, Bush H. Perceptual findings on the Broadway belt voice. *J Voice*. 2010; 24(6):678–689. [PubMed: 19900789]
7. Hertegard S, Gauffin J, Sundberg J. Open and covered singing as studied by means of fiber optics, inverse filtering, and spectral analysis. *J Voice*. 1990; 4(3):220–230.
8. Sundberg J, Thalén M, Popeil L. Substyles of belting: phonatory and resonatory characteristics. *J Voice*. 2012; 26(1):44–50. [PubMed: 21439776]
9. Searce L. *Manual of Singing Voice Rehabilitation: A Practice Approach to Vocal Health and Wellness*. San Diego, CA: Plural Publishing, Inc.; 2016. 161245
10. Estill J. Observations about the quality called "belting". In: Lawrence V, editor *Transactions of the 9th Symposium Care of the Professional Voice*. Vol. 2. New York, NY: The Voice Foundation; 1980. 82–88.
11. Lebowitz A, Baken RJ. Correlates of the belt voice: A broader examination. *J Voice*. 2011; 25(2): 159–165. [PubMed: 20236798]

APPENDIX A – Visual Analog Scales and instructions provided to raters

Rating Scales

Instructions

You will be listening to a number of randomized singing samples. Please place a vertical line on the below 100mm visual analogue scales (VAS) to best represents your answer. A line nearest the left side will represent an agreement most closely with the word (anchor) on the left side, and a line all the way to the right side will represent an agreement most closely with the word (anchor) on the right side. The entire length of the VAS can be used

SAMPLE 1 (slide 1)

- What style most closely matches this sample?
belt _____ legit
 - Please rate overall roughness of the sound.
No roughness _____ severe roughness
 - Please rate the tone quality of the sample.
brassey _____ fluty
bright _____ dark
-

Biographies

Lynn Maxfield, PhD is the Associate Director of the National Center for Voice and Speech (NCVS) at the University of Utah. His research areas include motor learning in voice training, occupational voice health, semi-occluded vocal tract exercises, and non-linearity in the source-filter relationship of the vocal tract.

In addition to research, he teaches Instrumentation for Voice Analysis at the Summer Vocology Institute and Doctoral Voice Pedagogy for the University of Utah's School of Music.

Prior to joining the NCVS, he taught voice science, voice pedagogy, and applied voice lessons at Eastern Connecticut State University, Knox College, and Monmouth College.

Brian Manternach, D.M., is an Assistant Professor in the University of Utah's Department of Theatre. He has lectured for the National Center for Voice and Speech, the University of Utah Voice Disorders Center, and TEDxSaltLakeCity, presented research at national conferences for NATS and the Voice Foundation, and received the 2016 Voice Pedagogy Award from the NATS Foundation.

An Associate Editor of the *Journal of Singing*, he also authors “The Singer’s Library” book review column for *Classical Singer* magazine.

Actively performing in opera, music theater, and concert work, his students are pursuing performance careers across the U.S. and abroad.

Figure 1. Interaction plot with quadratic regression lines drawn by experience level (beginner and professional).

Figure 2.
Interaction plot with quadratic regression lines drawn by pitch level.

Table I

Correlations using raw median rater scores and the singers' indicated styles

		Singers' Indicated Style	Raw Median Rater Scores
Singers' Indicated Style	Pearson Correlation	1	.737**
	Sum of Squares and Cross-products	88.000	2451.000
	Covariance	.672	18.710
	N	132	132
Raw Median Rater Scores	Pearson Correlation	.737**	1
	Sum of Squares and Cross-products	2451.000	125618.242
	Covariance	18.710	958.918
	N	132	132

** Correlation is significant at the 0.01 level (2-tailed).

Table II

Correlations using re-coded rater's scores

		Re-coded Median Rater's scores	Singers' Indicated Style
Re-coded Median Rater's scores	Pearson Correlation	1	.700 ^{**}
	Sum of Squares and Cross-products	98.061	65.000
	Covariance	.749	.496
	N	132	132
Singers' Indicated Style	Pearson Correlation	.700 ^{**}	1
	Sum of Squares and Cross-products	65.000	88.000
	Covariance	.496	.672
	N	132	132

^{**}
Correlation is significant at the 0.01 level (2-tailed).

Table III

Correlations between style rating (belt/legit) and two tone qualities (bright/dark and brassy/fluty).

Correlations				
		Median Belt/Legit	Median Brassy/Fluty	Median Bright/Dark
Median Belt/Legit	Pearson Correlation	1	.516 **	.398 **
	Sig. (2-tailed)		.000	.000
	N	132	132	132
Median Brassy/Fluty	Pearson Correlation	.516 **	1	.641 **
	Sig. (2-tailed)	.000		.000
	N	132	132	132
Median Bright/Dark	Pearson Correlation	.398 **	.641 **	1
	Sig. (2-tailed)	.000	.000	
	N	132	132	132

** Correlation is significant at the 0.01 level (2-tailed).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table IV

Intra-rater Interclass Correlation Coefficients for repeated tokens.

Interclass Correlation	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5
Single Measures	.775	.923	.178	.439	.618
Average Measures	.873	.960	.302	.610	.764