Physiological and acoustic characteristics of the male music theatre voice

Tracy Bourne[a)]
Federation University, Arts Academy, Ballarat, Victoria 3353, Australia

Maëva Garnier[b)]
CNRS, GIPSA-lab, 11 rue des Mathématiques, Grenoble Campus BP46, F-38402 Saint Martin d'Hères Cedex, France

Adeline Samson
Laboratoire Jean Kuntzmann, UMR CNRS 5225, University Grenoble-Alpes, Grenoble, France

(Received 19 March 2015; revised 1 June 2016; accepted 8 June 2016; published online 26 July 2016)

Six male music theatre singers were recorded in three different voice qualities: legit and two types of belt ("chesty" and "twangy"), on two vowels ([e] and [ɔ]), at four increasing pitches in the upper limit of each singer's belt range (∼250–440 Hz). The audio signal, the electroglottographic (EGG) signal, and the vocal tract impedance were all measured simultaneously. Voice samples were analyzed and then evaluated perceptually by 16 expert listeners. The three qualities were produced with significant differences at the physiological, acoustical, and perceptual levels: Singers produced belt qualities with a higher EGG contact quotient (CQ_{EGG}) and greater contacting speed quotient (Qcs), greater sound pressure level (SPL), and energy above 1 kHz (alpha ratio), and with higher frequencies of the first two vocal tract resonances (f_{R1}, f_{R2}), especially in the upper pitch range when compared to legit. Singers produced the chesty belt quality with higher CQ_{EGG}, Qcs, and SPL values and lower alpha ratios over the whole belt range, and with higher f_{R1} at the higher pitch range when compared to twangy belt. Consistent tuning of f_{R1} to the second voice harmonic ($2f_0$) was observed in all three qualities and for both vowels. Expert listeners tended to identify all qualities based on the same acoustical and physiological variations as those observed in the singers' intended qualities. *© 2016 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4954751]

[ZZ] Pages: 610–621

I. INTRODUCTION

The terms *belt*, *legit*, and *mix* are commonly used in the professional music theatre industry and in tertiary training institutions to describe vocal qualities that have their origins in both classical and popular musical styles. The distinctive sounds of these qualities reflect the different emotional states, characterizations, and musical influences that are intrinsic to this performance genre. But how distinct are these qualities, and how consensual are these terms? After a first study conducted on female voices (Bourne and Garnier, 2012), this present study aims to examine how these qualities differ for male voices in significant and reproducible ways, at the physiological and acoustic levels, as intended by the singers. Further, we aim to examine how expert listeners agree (between them, and with the singers) on the identification and distinctiveness of these three vocal qualities.

A. Previous knowledge on the female music theatre voice

Previous studies of female singers in contemporary commercial music (CCM) styles have outlined some differences between *belt* and *classical* vocal productions. Typically, female *belt* is characterized by higher subglottal pressure than for classical voice (Sundberg *et al.*, 1993; Bjorkner *et al.*, 2006) with higher formant frequencies (Sundberg *et al.*, 1993; Bestebreurtje and Schutte, 2000) and a more open articulation (Sundberg *et al.*, 1993; Lovetri *et al.*, 1999). Female belt has also been characterized by the tuning of the first formant ($F1$) or vocal tract resonance (f_{R1}) to the second harmonic ($2f_0$) (Schutte and Miller, 1993; Bestebreurtje and Schutte, 2000; Bourne and Garnier, 2012) at pitches where classical sopranos either demonstrate no formant tuning, or tune f_{R1} to f_0 (Joliveau *et al.*, 2004; Garnier *et al.*, 2010).

Fewer studies have specifically compared the music theatre sub-styles. Perceptually and pedagogically, music theatre *belt* appears to share many similarities to CCM *belt*, while *legit* production is closer to the classical voice in a number of parameters (Edwin, 2003; Balog, 2005; AATS, 2008; Bourne and Kenny, 2016). *Belt* articulation typically includes a more open mouth, a higher and more forward tongue, a higher larynx, and a narrower pharynx than *legit* (Sundberg *et al.*, 1993), although there may be some exceptions (Lovetri *et al.*, 1999). This more open and forward articulation for the *belt* quality is accompanied by consistently higher frequencies of the first two resonances (Schutte and Miller, 1993; Sundberg *et al.*, 1993; Bourne and Garnier, 2012). The first resonance (f_{R1}) is generally tuned to $2f_0$ for *belt* sounds (Lebowitz and Baken, 2011), while *legit* demonstrates no consistent tuning of resonances to harmonics (Bourne and Garnier, 2012). *Belt* tends to be produced with a higher sound pressure level (SPL), a lower glottal open quotient (OQ), vocal fold contacting speed quotient (Qcs), and speed quotient than *legit* (Sundberg *et al.*, 1993; Lebowitz and Baken,

[a)]Also at: University of Sydney, Conservatorium of Music, Sydney, NSW 2000, Australia. Electronic mail: tracy.marie.bourne@gmail.com
[b)]Also at: School of Physics, University of New South Wales, Sydney, New South Wales 2052, Australia and University Grenoble-Alpes, F-38040 Grenoble, France.

 0001-4966/2016/140(1)/610/12/$30.00

2011; Bourne and Garnier, 2012). These studies supported the idea that female singers may produce *belt* in laryngeal mechanism M1 and *legit* in laryngeal mechanism M2 (Schutte and Miller, 1993; Bestebreurtje and Schutte, 2000; Bourne and Garnier, 2012) at pitches where classical sopranos systematically sing in M2 (Henrich, 2006). Furthermore, CCM vocal qualities are characterized by greater activation of the thyroarytenoid (TA) muscles and increased adduction of the vocal processes in chest and chest-mix qualities than for head and head-mix qualities (Kochis-Jennings *et al.*, 2012).

B. What about the male music theatre voice?

Things are less clear for the male voice. In a comparison of voice source and formant frequencies of operatic and music theatre male singers, Bjorkner (2008) found that the music theatre singers tended to use a slightly higher subglottal pressure than the classical singers, as well as higher vocal intensity, higher closed quotient (CQ) values, higher formant frequencies, higher SPL values as well as systematic tuning of $F2$, $F4$, and $F5$. Both classical and music theatre singers demonstrated similar levels of normalized amplitude quotient, suggesting that music theatre vocal production is no more "pressed" than opera despite the higher values of CQ and SPL in music theatre subjects (Bjorkner, 2008). There is some evidence to suggest that male CCM singers produce *belt* sounds with a more open and wider mouth shape than classical singers (Titze and Worley, 2009). Sundberg *et al.* (2011) found that classical and non-classical singers used different formant tuning strategies with $F1$ and $F2$ frequencies just under the second voice harmonic ($2f_0$), whereas the CCM singers tended to tune $F1$ above or at $2f_0$ frequencies.

Some pedagogues question the existence of male *belt* as a quality in its own right (Bourne and Kenny, 2016). Since men predominantly sing in laryngeal mechanism M1, it is unclear whether the male voice qualities can vary to a significant degree within the music theatre style. In particular, it is unclear whether a *legit* quality can really be defined for male music theatre singers and if it exists, whether *legit* is differentiated from *belt* by vocal adjustments, or by the use of the laryngeal mechanism M2, as seems to be the case in female music theatre singers.

C. Goals of the study

In short, while we have some understanding of the physiological and acoustic characteristics of the female music theatre voice, there is almost no research on the male voice. We can assume that there are similarities of production between men and women, however the physiological differences between the genders do affect pitch and register and are likely to have an impact on the production and perception of these vocal qualities. This study aims to determine what these differences are in the context of music theatre voice by objectively measuring their acoustic and physiological characteristics, interpreting them in terms of vocal tract adjustments and laryngeal mechanisms, and comparing these results with perceptual evaluations of each quality.

II. MATERIAL AND METHOD

A. Acoustic and physiological database

1. Participants and tasks

Six male Australian music theatre singers agreed to participate in this study. Four of the singers were professional (Singers S1, S2, S4, and S6) and two of them were advanced tertiary students in a Bachelor of Music Theatre course (S3 and S5). All singers had received between 5 to 10 years of vocal training in both classical and CCM vocal styles.

Singers were asked to sustain a single note for 4 s with no change in pitch or tone and without vibrato at four frequencies up to their highest comfortable belt range (see Table I).[1]

Each singer was asked to produce these notes in three qualities: chesty belt, twangy belt, and legit on two vowels ([e], [ɔ]) and to produce five repetitions for each sample. S1 recorded chesty belt on both vowels, but was able to produce legit and twangy belt qualities on the [e] vowel only. No technical instructions were given to the singers in relation to vocal production of these qualities. Each singer was given 10 min to warm up prior to the recording session and was provided with water and encouraged to take vocal breaks.

2. Measured signals

The audio signal was recorded with a $\frac{1}{4}$-in. pressure microphone (Bruël and Kjær 4944-A) attached to the front of a stand. The height was adjusted so that the microphone rested gently upon the singer's lower lip during phonation. The audio signal was amplified (Bruël and Kjær Nexus 2690), and digitized at 16 bits and a rate of 44.1 kHz using a Firewire audio interface (MOTU 828). A small, flexible tube was placed alongside the microphone and was connected to a loudspeaker via an impedance matching horn. This acoustic source was used to excite the vocal tract with a synthesized broadband signal (with a frequency resolution of 11 Hz) during the final remaining three seconds of phonation. The microphone recorded the response of the vocal tract to that excitation, enabling the measurement of the frequencies of the first three vocal tract resonances (see Epps *et al.*, 1997; Joliveau *et al.*, 2004; or Garnier *et al.*, 2010 for more details about this technique).

The EGG signal was simultaneously recorded with a two-channel electroglottograph (Glottal Enterprises EG2) using medical gel to improve electric contact between the skin and the electrodes. Electrodes were placed on both sides of the

TABLE I. Singers' voice type and pitch range as investigated in this study. The pitches selected for the perceptual tests are in bold font.

	Recorded pitch range as intended by the singer						
	D4	E4	F4	$F^{\#}4$	G4	$G^{\#}4$	A4
S1		**X**		X	**X**	X	
S2		**X**		X		**X**	X
S3		**X**		X	**X**		X
S4	X	**X**		X		**X**	
S5	X	**X**	X	**X**			
S6	X	**X**	X		**X**		

thyroid cartilage while the singer was singing in his comfortable middle range. No automatic gain control was used. The high-pass filter was set to a 10 Hz cutoff frequency. The EGG signal was then digitized at 16 bits and a rate of 44.1 kHz using the same Firewire audio interface (MOTU 828).

B. Objective characteristics

Using MATLAB software, nine objective descriptors of the sound and the phonation gesture were extracted from the recorded signals.

Two acoustic descriptors were measured from the first clean second of phonation (no broadband excitation noise):

- The mean SPL was measured accurately, using the internal calibration signal of 1 V-RMS at 1 kHz delivered by the conditioning amplifier (Bruël and Kjær Nexus 2690), and knowing its V/Pa transduction coefficient. It is expressed in dB(Z), meaning that no weighting was applied to account for the frequency sensitivity of the human ear.
- The alpha ratio (or alpha measure, α), defined as the ratio (in dB) of energy above and below 1 kHz (Frøkjaer-Jensen, 1976; Sundberg and Nordenberg, 2006), was calculated from the long term average spectrum (on 4096 points).

Three glottal descriptors were extracted from the EGG signal during the full 4 s of phonation, defined from the closing (positive) and opening (negative) peaks detected in the derivative of the electroglottographic signal (DEGG, see Henrich *et al.*, 2004):[2]

- The mean fundamental frequency (f_0), measured from the time interval between two consecutive closing peaks (Henrich *et al.*, 2004).
- The mean EGG contact quotient (CQ_{EGG}) defined as the ratio between the time interval between a closing peak and the next opening peak, and the fundamental period of the glottal cycle ($1/f_0$). This parameter corresponds to 1-OQ, as defined in our previous companion article on the female music theatre voice (Bourne and Garnier, 2012).
- The mean vocal fold Qcs, defined as the ratio in amplitude of the closing and opening peaks of the DEGG signal. Qcs reflects the degree of asymmetry of the EGG waveform. The high sampling frequency of the EGG signal (44.1 kHz) guarantees a reliable measure of this parameter.

Finally, the frequencies of the first three vocal tract resonances (f_{R1}, f_{R2}, and f_{R3}) were measured during the second to fourth seconds of phonation, by detecting manually the first three maxima of the pressure ratio γ.

C. Perceptual evaluations

1. Listeners

Sixteen expert teachers and vocal coaches from Australia, Canada, United Kingdom, and U.S.A. were invited to undertake a two-part listening test from a webpage.

2. Stimuli

The first part of the listening test consisted of an introductory session during which the expert teachers evaluated 28 sustained pitches extracted from musical phrases of commercially available recordings of music theatre songs from popular Broadway and West End shows. The purpose of this introductory session was to prepare the listeners by presenting more familiar sound examples than those in our database, and to establish whether expert listeners agreed in their perceptual evaluation of commercial samples as a starting point. For this pre-test session, the sustained pitches were 3–5 s in duration, at pitches between C4 and B4, produced on different vowels with varying voice qualities.

The second and most important element of the listening test was an evaluation of a subset of 68 sound examples selected from the recorded database. Samples were chosen as clear and representative examples of each intended quality (chesty belt, twangy belt, and legit), from each of the six singers on two pitches (E4 and G4) and for the two vowels [e] and [ɔ] (see Table I). Singer S1 was only able to produce legit and twangy belt qualities on the [e] vowel, so that we selected 8 samples for that singer instead of the 12 samples chosen for all the other singers bringing the total number of sound examples selected to 68. For S2, S4, and S5, who did not actually produce G4 pitches, we selected samples produced at F#4 and at G#4 (see Table I). Using the PSOLA module in Praat software, we artificially shifted the pitch of these samples up or down to G4 in order to compare the stimuli at a similar pitch for all the singers. This pitch manipulation enabled us to modify pitch without affecting formants and vowel duration. For a slight pitch manipulation of a semitone, it neither affected the spectral envelope nor the perceived voice quality. The stimuli consisted of the "clean" second of phonation, i.e., without any excitation noise, and were normalized in mean intensity. The order of the samples was randomized for the test.

3. Task

For both the introductory and second part of the listening test, listeners were asked to indicate through a forced choice question (Q1) whether they thought the sample was (1) a belt sound, (2) a legit sound, or (3) another quality which they were asked to describe. If they indicated that the sample was belt, they were then asked to answer a second forced choice question (Q2) and specify further whether they thought the sample was (1.1) a "chesty" belt sound, (1.2) a "twangy" belt sound, or (1.3) another type of belt sound, which they were again asked to describe.

D. Statistical analysis

Several statistical analyses were conducted using the R software. The conventional notation was adopted to report statistical results: $^{*}p < 0.05$, $^{**}p < 0.01$, ***, $p < 0.001$, and ns (not significant) $p > 0.05$.

1. Analysis of the production data

First, an analysis of variance (ANOVA) was conducted for each objective parameter (except f_0), in order to examine the effect of the factor QUALITY (as intended by the singers, with three levels: legit, chesty belt, twangy belt) on the value of these voice descriptors, and to determine whether the differences observed between the three qualities could be considered as statistically significant depending on the other factors VOWEL (qualitative factor, with two levels: [e] and [ɔ]) and f_0 (quantitative factor). We conducted the ANOVA test from a mixed model of the data, which aimed to explain the variance of each objective parameter not only by fixed effects (of the factors QUALITY, VOWEL, and f_0) but also by a random effect (of the factor SINGER, on the intercept) using the R package lme.

For each objective parameter, we searched for the simplest model to best explain the variance of this parameter, using a descending approach (function step in R), based on the minimization of the Bayesian Information Criterion. Hypotheses about the model's normality and homoscedasticity have been validated by looking at the residuals graphs.

After examining the effects of the interaction terms remaining in the simplified model, we tested more specifically for the global effect of the factor QUALITY using a likelihood ratio test (LRT).

Specific contrasts were also examined—applying Bonferroni adjustments—between legit and the two belt qualities, and between chesty belt and twangy belt qualities, to determine whether voice parameters were significantly different between these qualities (using the package multcmp in R).

2. Analysis of the perceptual data

In analyzing the results of the perceptual test, we examined the inter-listener agreement separately for the sustained pitches extracted from music phrases and for the sustained sound samples from the recorded database. We then examined the first and second questions of the perceptual test separately (Q1: Legit, Belt, or Other, $Q2_{/Q1=\mathrm{Belt}}$: Chesty Belt, Twangy Belt, or Other kind of belt). For these four cases, we computed the Fleiss's κ as a global indicator of the inter-listener agreement. We arbitrarily chose a threshold of 60% of inter-listener agreement (i.e., more than 10 listeners over 16, for the first question) to determine whether a quality was consensually vs unclearly perceived. Using the 60% threshold, we examined the inter-listener agreement on the second question for the 59 samples that had been rated as belt by more than 6 listeners (i.e., that were not consensually evaluated as legit). Second, we examined the match between the singer's intended quality and the quality actually perceived by the listeners, by drawing confusion matrices and calculating the percentage of "successful" recognition by the listeners. Finally, we conducted logistic regressions on the qualities perceived by the expert listeners, in order to determine whether voice quality (as perceived by the listeners) could be predicted from the combined variation of the seven acoustical and physiological descriptors of the corresponding voice productions. We considered four binary variables: (1) Perceived belt (or not); (2) Perceived legit (or not); (3) Perceived chesty belt (or not), for the samples evaluated as belt in Q1; (4) Perceived twangy belt (or not), for the samples evaluated as belt in Q1. For each of these binary variables, we made a binary logistic regression from the following mixed model:

$$\text{Perceived Quality} \sim f_{R1} + f_{R2} + f_{R3} + \text{SPL} + \alpha + \text{CQ}_{\text{EGG}} + \text{Qcs} + 1|\text{LISTENER}.$$

We reported the area under the Receiver Operating Characteristic curve (AUC) as a quality index of this model.

III. RESULTS

A. Objective comparison of intended qualities

1. Vocal tract resonances

a. First resonance frequency (f_{R1}). Globally, f_{R1} followed the variations of f_0 with a significantly positive slope in each vowel and each quality (Mean f_{R1}:f_0 slope of 1.18 Hz/Hz, $p < 0.001$) (see Fig. 1). At the group level, the variations of f_{R1} followed those of f_0 with a greater slope for the vowel [ɔ] than [e], a greater slope for the belt qualities compared to legit (Δslope = +0.54 Hz/Hz, $p < 0.001$), and a greater slope for chesty belt compared to twangy belt (Δslope = +0.31 Hz/Hz, $p < 0.001$). However, this difference did not reflect significantly different strategies of resonance tuning to voice harmonics in these qualities and vowels. Indeed, at an individual level, the first vocal tract

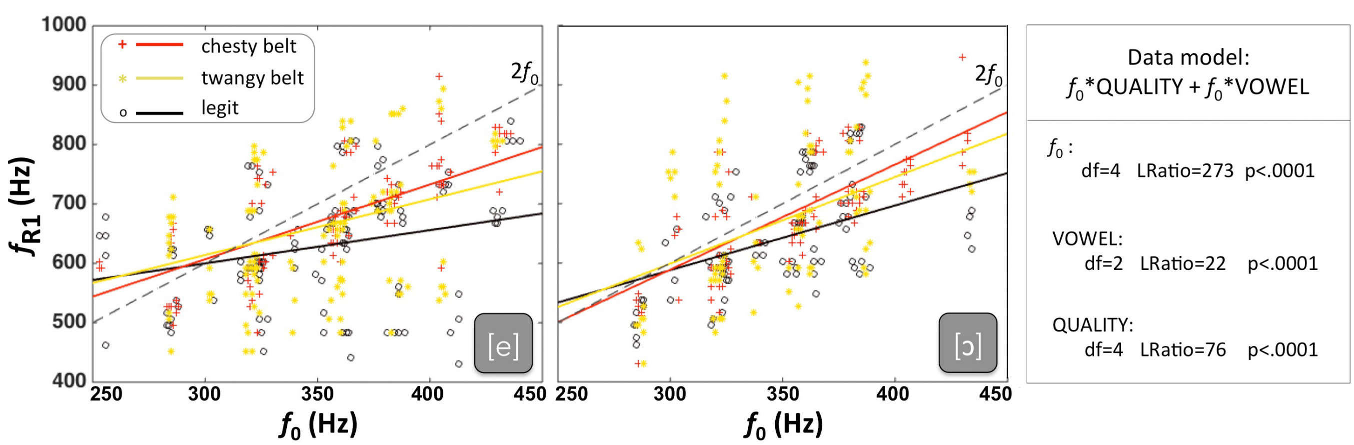

FIG. 1. (Color online) Variations of the first resonance frequency of the vocal tract (f_{R1}) as a function of increasing pitch (f_0), vowel ([e] or [ɔ]), and singing quality (legit, chesty belt, or twangy belt). The plain lines represent, for each quality and each vowel, the slope and intercept coefficients obtained from the regression model (taking into account a random effect of the singer on the intercept). The dashed lines represent the frequency of the nearest voice harmonic ($2f_0$).

resonance was found to be adjusted to the second source harmonic (i.e., at a distance closer to ±30% of $2f_0$) in the three qualities, though with a varying degree of reproducibility over the different repetitions of the task (see Fig. 2). This tendency was observed for chesty belt in both vowels for all singers (reproducibility of 89 ± 14%, depending on the singer and the vowel) and for twangy belt in both vowels for all singers with the exception of the vowel [e] for singer S1 and the vowel [ɔ] for singer S2 (reproducibility of 70 ± 16%). The same was observed for legit in both vowels for all singers with the exception of the vowel [e] for singer S1 and the vowel [ɔ] for singer S3 (reproducibility of 86 ± 15%).

Despite the similarity of tuning strategies for f_{R1} observed in all three qualities across the whole belt range, small but significant differences in f_{R1} values were observed between these vocal qualities at higher pitches. Thus, at the top of the belt range, f_{R1} tended to be greater in belt qualities compared to legit (85 Hz, $p < 0.001$) and greater in chesty belt than in twangy belt (35 Hz, $p < 0.01$), regardless of the vowel.

b. Second resonance frequency (f_{R2}). Globally, f_{R2} followed the variations of f_0 with a significantly positive slope only for the vowel [ɔ] and both belt qualities [+2.75 Hz/Hz ($p < 0.001$) and +1.43 Hz/Hz ($p < 0.001$) in chesty and twangy belt, respectively] (see Fig. 3). It did not vary significantly with pitch in the other cases, reflecting singer-specific strategies of resonance tuning to voice harmonics observed in five of the singers (over six). Indeed, all of the five singers who produced twangy belt on [ɔ] tuned f_{R2} to a voice harmonic ($3f_0$ or $4f_0$). Only three of the singers (S4, S5, S6) demonstrated a similar tuning of f_{R2} in chesty belt and only one singer (S4) in legit. In the other cases no specific tuning of f_{R2} was observed (see Fig. 2).

Apart from resonance tuning considerations, belt qualities tended to be produced with significantly greater f_{R2} values than legit. This tendency was significant for the vowel [e] (+158 Hz on average, $p < 0.001$) and only at the top of the belt range for the vowel [ɔ] (+291 Hz, $p < 0.001$). No general tendency could be found to differentiate between the two kinds of belt.

c. Third resonance frequency (f_{R3}). Globally, f_{R3} tended to follow the variations of f_0 with a positive slope for the vowels [e] and with a negative slope for the vowels [ɔ] (see Fig. 4). These slopes were in any case very small, and significant in only some cases [+1.10 Hz/Hz for the vowels [e] in legit ($p < 0.001$); +0.64 Hz/Hz for the vowels [e] in twangy belt ($p < 0.01$); −0.95 Hz/Hz for the vowels [ɔ] in chesty belt ($p < 0.001$)]. Some significant differences were observed in the f_{R3} values of the three qualities. However, they never exceeded 78 Hz (i.e., about 2%–3% of typical f_{R3} values) and depended on pitch and vowel. As a result, no general tendency could be determined for differences between f_{R3} for the three qualities.

2. Glottal descriptors

a. EGG contact quotient (CQ_{EGG}). CQ_{EGG} increased with f_0 with a small but significantly positive slope that did not

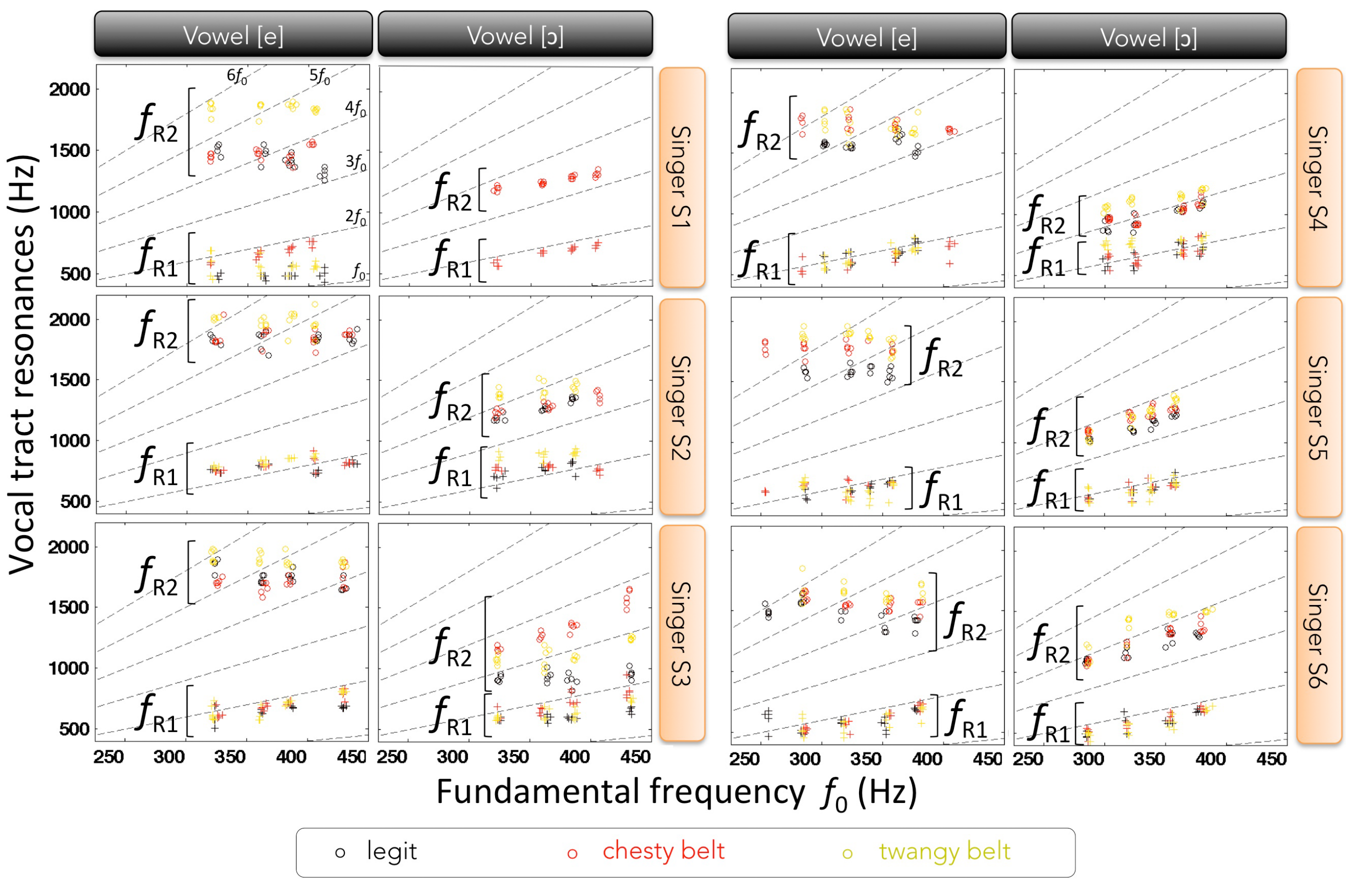

FIG. 2. (Color online) Variations of the first two vocal tract resonance frequencies (f_{R1} and f_{R2}) as a function of increasing pitch (f_0), vowel ([e] or [ɔ]), and singing quality (legit, chesty belt, or twangy belt), for six music theatre singers. The dashed lines represent the frequency of the nearest voice harmonics (f_0 to $6f_0$).

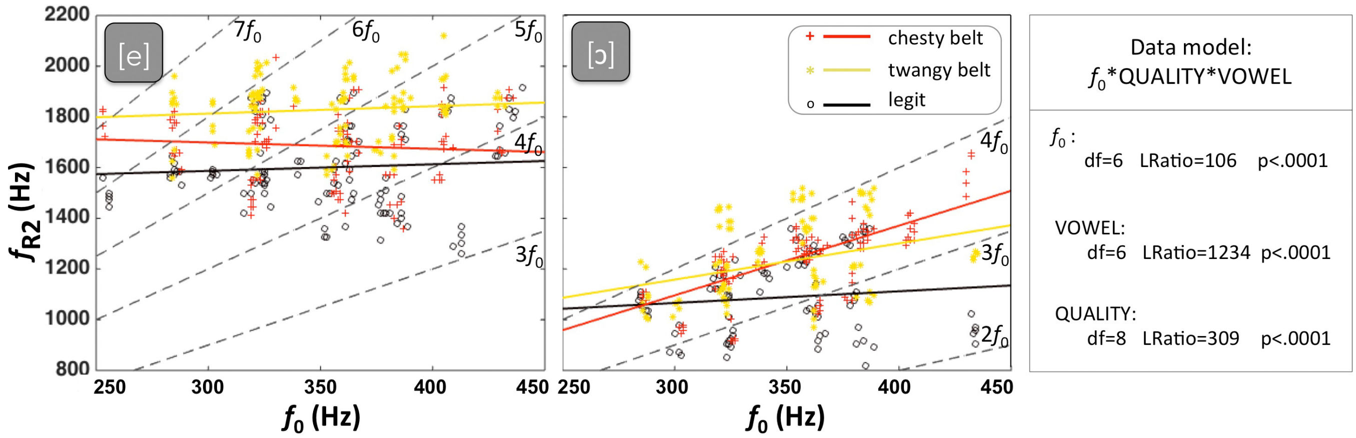

FIG. 3. (Color online) Variations of the second resonance frequency (f_{R2}) as a function of increasing pitch (f_0), vowel ([e] or [ɔ]), and singing quality (legit, chesty belt, and twangy belt). The plain lines represent, for each quality and each vowel, the slope and intercept coefficients obtained from the regression model. The dashed lines represent the frequency of the nearest voice harmonics ($2f_0$ to $7f_0$).

depend on the vowel or quality (Mean CQ_{EGG}:f_0 slope of 2.5×10^{-4}/Hz, $p < 0.001$) (see Fig. 5). CQ_{EGG} was significantly influenced by the quality, with slightly higher values observed in belt qualities, compared to legit (+0.027 on average, $p < 0.001$), and slightly higher values in chesty belt than for twangy belt for the vowel [e] only (+0.017, $p < 0.001$). Despite the statistical significance of these observations, variations in CQ_{EGG} remained small for all the singers but S1, and the measured values were in a similar range (0.45–0.65, with an intra-singer variability lower than 0.1). For the singer S1, however, CQ_{EGG} values were measured in a distinct range for his legit productions on the vowel [e] ($CQ_{EGG} < 0.45$), which was significantly lower than for belt ($0.55 < CQ_{EGG} < 0.6$).

b. Vocal fold Qcs. Globally, Qcs followed the variations of f_0 with a slope that was significantly negative for all conditions except the vowels [e] produced in chesty belt quality (see Fig. 6). Qcs showed significantly greater values in belt than in legit (+3.45 on average, $p < 0.001$). Chesty and twangy belt qualities were also significantly different, with greater Qcs values observed in chesty belt (+0.92 on average, $p < 0.01$). Although four of the six singers showed a significantly distinct range of Qcs values for their productions of chesty belt and legit qualities, it is interesting to mention that only the singer S1 contrasted with the results of other singers by showing particularly low Qcs values (<2) in the legit quality.

3. Descriptors of the radiated spectrum

a. SPL. The SPL increased significantly with f_0 in the belt qualities (Mean SPL: f_0 slope of 0.035 dB/Hz, $p < 0.001$), but not in legit (Mean SPL: f_0 slope of 0.008 dB/Hz, $p > 0.4$) (see Fig. 7). Consequently, at the top of the belt range, belt qualities were produced with significantly higher SPL values compared to legit (6.5 dB, $p < 0.001$). A smaller but significant difference was still observed at the bottom of the belt range, but only for the vowel [e] (2.1 dB, $p < 0.01$). Furthermore, chesty belt sounds were produced with higher SPL values compared to twangy belt. This difference tended to be greater in [e] than [ɔ] vowels (+1.0 dB) and again, it increased with pitch [from 2.7 dB ($p = 0.001$) at the bottom to 5.6 dB ($p < 0.001$) at the top of the belt range].

b. Alpha ratio. Belt qualities were always produced with higher alpha ratios compared to legit (see Fig. 8). This difference tended to be greater in [e] than [ɔ] vowels (+1.1 dB), and increased with pitch [from 1.8 dB ($p = 0.004$) at the

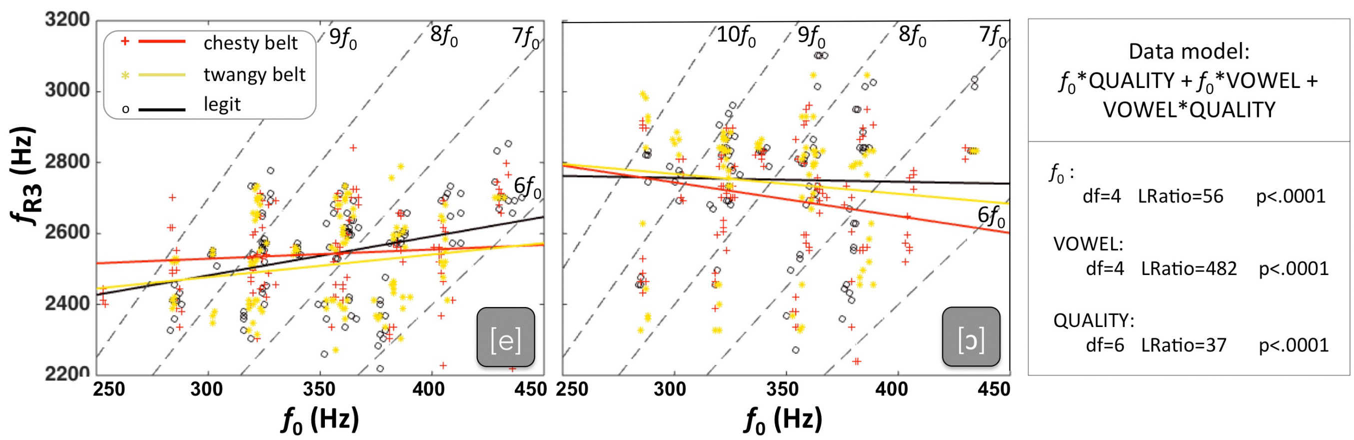

FIG. 4. (Color online) Variations of the third resonance frequency (f_{R3}) as a function of increasing pitch (f_0), vowel ([e] or [ɔ]), and singing quality (legit, chesty belt, and twangy belt). The plain lines represent, for each quality and each vowel, the slope and intercept coefficients obtained from the regression model. The dashed lines represent the frequency of the nearest voice harmonics ($6f_0$ to $10f_0$).

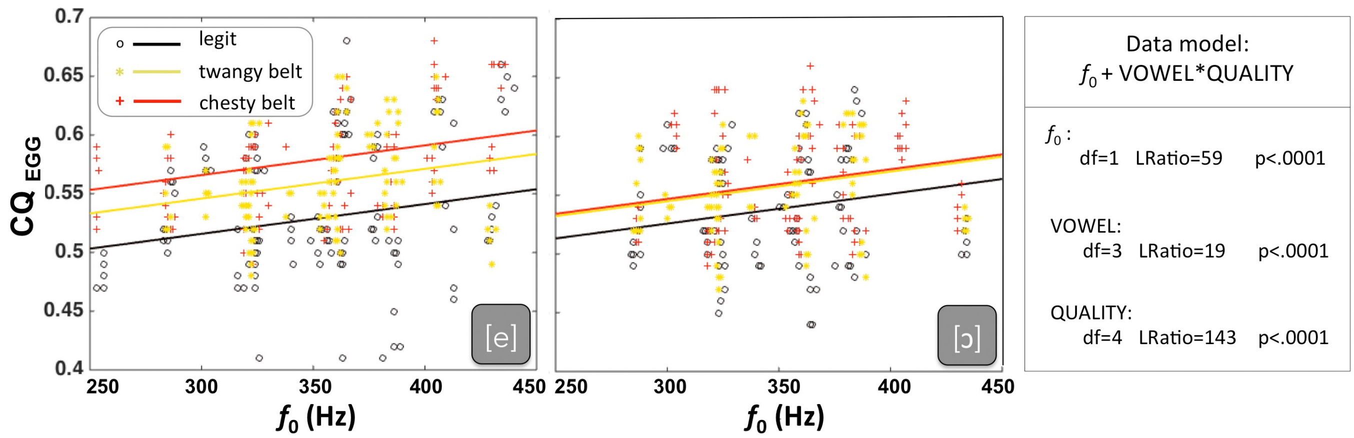

FIG. 5. (Color online) Variations of CQ_{EGG} as a function of increasing pitch, vowel ([e] or [ɔ]), and singing quality (legit, chesty belt, and twangy belt). The plain lines represent, for each quality and each vowel, the slope and intercept coefficients obtained from the regression model.

bottom to 5.5 dB ($p < 0.001$) at the top of the belt range]. Furthermore, twangy belt sounds were produced with higher alpha ratios than was the case for chesty belt productions. This difference tended to be greater in [e] than [ɔ] vowels (+2.0 dB) and remained fairly constant over pitch (1.8 dB on average, $p = 0.02$).

B. Agreement and prediction of perceived qualities

1. Inter-subject agreement by expert listeners

The terms "belt" and "legit" appeared to be relevant for the listeners as they could evaluate the samples using one of these two terms for 96% of the sustained pitches extracted from music phrases and for 94% of the sound samples from the recorded database. Likewise, the terms chesty belt and twangy belt were recognized in the majority of the cases as appropriate sub-categories for sounds that were initially evaluated as belt in the first question (Q1); 81% for musical samples and 86% for database samples.

Expert listeners generally agreed on whether a sustained pitch extracted from music phrases was produced in legit or not (at 79.2%), or whether it was produced in belt or not (at 80.5%). However, for the sound samples from the recorded database, the listeners showed much less agreement on the identification of these qualities (65.0% of agreement for legit, and 59.4% for belt). In other words, all but one of the musical samples were clearly identified as being produced in legit or belt with never less than 70% of inter-listener agreement. For the database samples, however, slightly less than three quarters of the samples had their quality "clearly" identified, (i.e., with an inter-listener agreement greater than 60%). These different results can be summarized by a global indicator of inter-listener agreement on answers to the first question of the perceptual test: the Fleiss's κ is of 0.55 for the musical samples (indicating a moderate inter-subject agreement) and of 0.15 for the database samples (indicating a only slight inter-subject agreement).

Inter-listener agreement on evaluation of the subtype of belt (second question Q2) was examined, considering only the 15 musical samples and the 59 database samples that were not clearly evaluated as legit in Q1 (i.e., that were rated as belt by at least a third of the listeners). Expert listeners agreed only moderately on whether a musical sample was produced in chesty belt or not (at 66.1%), or whether it was produced in twangy belt or not (at 57.2%). This level of agreement was even less for database samples, for which listeners agreed at 56.1% on whether a sample belonged to the chesty belt subcategory or not, and at only 44.7% on whether it belonged to the twangy belt category or not. This means that only a little more than half of the samples were "clearly" identified (i.e., with more than 60% of agreement) as either chesty or twangy belt.

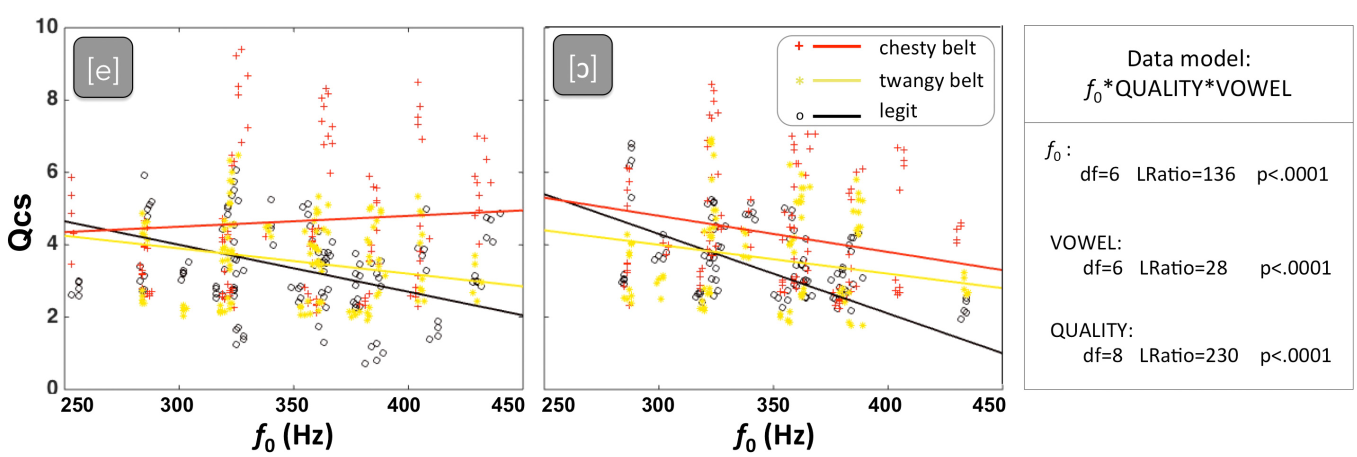

FIG. 6. (Color online) Variations of Qcs as a function of increasing pitch (f_0), vowel ([e] or [ɔ]), and singing quality (legit, chesty belt, and twangy belt). The plain lines represent, for each quality and each vowel, the slope and intercept coefficients obtained from the regression model.

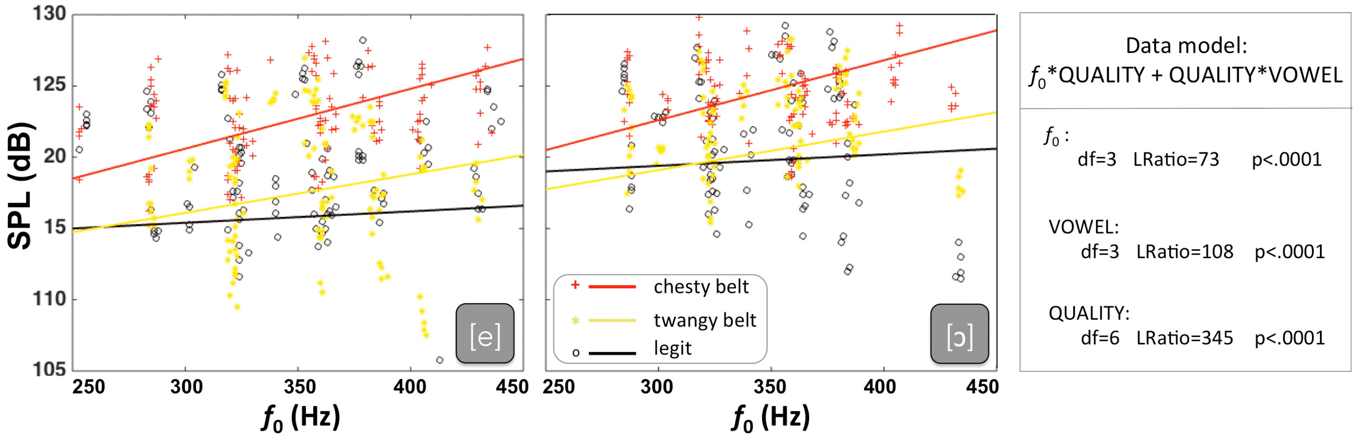

FIG. 7. (Color online) Variations of SPL as a function of increasing pitch (f_0), vowel ([e] or [ɔ]), and singing quality (legit, chesty belt, and twangy belt). The plain lines represent, for each quality and each vowel, the slope and intercept coefficients obtained from the regression model.

All these observations summarize into a Fleiss's κ of 0.15 and 0.10 for the answers to the second question on musical and database samples, respectively, indicating in both cases an only slight level of inter-listener agreement.

Database samples were evaluated with comparable inter-rater agreement for both vowels and both pitches.

2. Agreement between singers and listeners (intended vs perceived quality)

A very good match was observed for belt qualities between the intention of the singer and the quality perceived by the listeners: The samples intended as belt were indeed clearly perceived as belt for the most part (recognition rate of 69.6%) (see Table II). The legit quality showed a more moderate match: samples intended as legit were very often misperceived as belt or unclearly perceived (recognition rate of 27.3%). However, the samples that were clearly perceived as legit were generally intended as such.

The samples intended as twangy belt were generally clearly recognized as twangy belt sounds (recognition rate of 71.4%), contrary to samples intended as chesty belt, which were generally unclearly perceived (recognition rate of 27.3%).

For productions of the legit quality, the match between the intention of the singer and the quality perceived by the listeners did not depend on the vowel or the pitch (see Table III). On the other hand, samples produced with an intended belt quality were less well recognized for the vowel [e] at low pitch. Likewise, the recognition of the intended twangy belt quality was not influenced by the vowel or the pitch, although intended chesty belt samples were slightly better recognized as such for the vowel [e] at low pitch.

3. Prediction of the perceived quality from the acoustical and physiological characteristics of the productions

Table IV summarizes the results of the logistic regression designed to predict expert listeners' perception of voice qualities in relation to variations of acoustical and physiological parameters.

The logistic regression showed that voice samples were more likely to be perceived as belt when f_{R1}, SPL, α, and CQ_{EGG} were greater. On the contrary, the legit quality was more likely to be perceived as such when f_{R1}, α, and CQ_{EGG} decreased. These predictions of the perceived quality are in complete agreement with the variations of the acoustical and physiological parameters observed in production, between belt and legit qualities as intended by the singers (see Sec. III). In production, however, additional differences were also

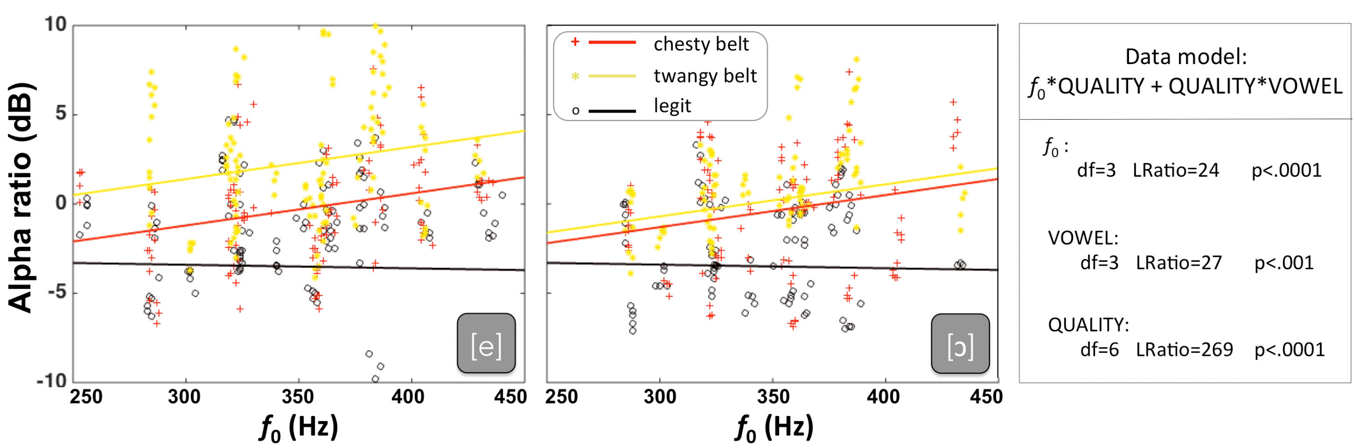

FIG. 8. (Color online) Variations of the alpha ratio as a function of increasing pitch (f_0), vowel ([e] or [ɔ]), and singing quality (legit, chesty belt, and twangy belt). The plain lines represent, for each quality and each vowel, the slope and intercept coefficients obtained from the regression model.

TABLE II. At the top, number of samples consensually perceived as legit or belt, or unclearly perceived, in the first question (Q1), as a function of the quality actually intended by the singers when they produced these samples. At the bottom, the number of samples not clearly perceived as legit in Q1, and then consensually perceived as chesty belt or twangy belt, or unclearly perceived, in the second question (Q2), as a function the quality actually intended by the singers.

	INTENDED		
		Belt	
	Legit	Chesty	Twangy
PERCEIVED			
Q1:68 samples			
Legit	6	2	1
Belt	8	17	15
Unclear	8	5	6
Q2:59 samples			
Chesty belt	7	6	1
Twangy belt	0	3	15
Unclear	9	13	5

observed between intended belt and legit qualities in their f_{R2} frequencies and their Qcs values. Listeners did not appear to rely significantly on these two parameters to evaluate whether a sound is belt or legit.

Furthermore, the logistic regression also showed that voice samples were more likely to be perceived as chesty belt when Qcs increased and when α decreased. The opposite was observed for the twangy belt quality. These predictions of the perceived subtypes of belt were in complete agreement with the variations of the acoustical and physiological parameters observed in production, between chesty belt and twangy belt qualities as intended by the singers (see Sec. III). Some slight acoustical and physiological differences were also observed between intended chesty belt and twangy belt qualities, in SPL, f_{R1} frequencies, and CQ_{EGG} values. Nevertheless, listeners did not appear to rely primarily on these indices to evaluate whether a belt sound was chesty or twangy.

IV. DISCUSSION

A. Can we define and distinguish different voice qualities in the male music theatre (MT) voice?

The results of this study support the idea that the legit quality can be defined in the male music theatre voice, and that it is significantly different from the belt sound physiologically, acoustically, and perceptually. In our study, belt differed from legit by higher alpha ratios, Qcs values, and CQ_{EGG} values over the whole belt range, and by higher f_{R1} at the upper pitch range. Belt also differed from legit by higher SPL and f_{R2} values over the whole belt range for [e] vowels, and only at the upper pitch range for [ɔ] vowels. In addition, the perceptual test showed that the terms legit and belt were meaningful and consensual for expert listeners. An only moderate match was observed between the singers' intended quality and the quality actually perceived by the listeners. However, the logistic regression showed that the listener's evaluation of these qualities was based on the same variations of acoustical and physiological descriptors as those observed between the singers' intended qualities: The belt quality was more likely to be perceived by the listeners when f_{R1}, SPL, α, and CQ_{EGG} increased. Conversely, the legit quality was more likely to be perceived when f_{R1}, α, and CQ_{EGG} decreased.

These results also support the idea that different subtypes of belt can be defined in the male music theatre voice, and are distinguished physiologically, acoustically, and perceptually. We found that chesty belt differed from twangy belt by higher CQ_{EGG}, Qcs, and SPL values and lower alpha ratios, over the whole belt range, and by higher f_{R1} at the higher pitch range. The perceptual test showed that the qualifiers chesty and twangy were meaningful and consensual belt sub categories for expert listeners. The logistic regression confirmed that the listeners' evaluation of these two belt subtypes were based on the same variations of acoustical and physiological descriptors as those observed between the singers' intended qualities: The chesty belt quality was more likely to be perceived when Qcs increased and when α decreased. The opposite was observed for the twangy belt quality.

B. How do the differences between belt and legit, and between chesty and twangy belt relate to vocal gestures?

The male singers in this study were able to produce belt and legit with significantly different glottal and vocal tract adjustments, resulting in significant differences in the radiated sound. The higher Qcs and CQ_{EGG} values observed in belt may simply be related to the greater SPL. However, the higher CQ_{EGG} values may also be caused by increased posterior vocal fold adduction (cartilaginous adduction) or by bulging of the vocal folds via TA muscle contraction (membranous adduction) while singing in chest register (Herbst *et al.*, 2009; Herbst *et al.*, 2011). In any case, for five of the singers, the observed differences in CQ_{EGG} and Qcs values between belt and legit were small and the values measured for these parameters remained in a similar range for both qualities, typical of the laryngeal mechanism M1 (Henrich *et al.*, 2005; Roubeau *et al.*, 2009). On the contrary, the CQ_{EGG} and Qcs values measured for singer S1 were in a significantly distinct and lower range in legit compared to belt, supporting the idea of a change of laryngeal mechanism for this singer. However, only a direct endoscopic examination could enable us to conclude with certainty on the laryngeal mechanism underlying these productions.

The higher f_{R1} and f_{R2} frequencies observed in belt when compared to legit may correspond to both a more open mouth and an anterior tongue position as well as a higher larynx, in agreement with empirical studies by teachers and

TABLE III. Percentage of successful recognition by the listeners of the voice quality intended by the singers, as a function of pitch and vowel.

	PRODUCTION			
	[e]		[ɔ]	
	E4	A4	E4	A4
INTENDED & RECOGNIZED				
Legit	33.3% (2/6)	16.7% (1/6)	40.0% (2/5)	20.0% (1/5)
Belt	33.3% (4/12)	91.7% (11/12)	72.7% (8/11)	81.8% (9/11)
Chesty Belt	60% (3/5)	16.7% (1/6)	20.0% (1/5)	16.7% (1/6)
Twangy Belt	80.0% (4/5)	66.7% (4/6)	80.0% (4/5)	60.0% (3/5)

TABLE IV. Results of the binary logistic regressions that aimed at predicting the probability for a voice sample to be evaluated by the listeners in the first question (Q1) as belt (or not), or as legit (or not), and in the second question (Q2) as chesty belt (or not), or as twangy belt (or not) from the value of acoustical and physiological descriptors of the corresponding voice productions.

	Belt		Legit		Chesty belt		Twangy belt	
	Model AUC = 0.682		Model AUC = 0.713		Model AUC = 0.682		Model AUC = 0.682	
	Odds ratio	p	Odds ratio	p	Odds ratio	p	Odds ratio	p
f_{R1}	1.0036	0.002**	0.9969	0.009**	0.9984	0.30	1.0021	0.17
f_{R2}	1.0007	0.07	0.9994	0.12	0.9995	0.36	1.0005	0.38
f_{R3}	0.9995	0.44	1.0007	0.25	1.0004	0.65	0.9996	0.58
SPL	1.0478	0.023*	0.9760	0.26	0.9682	0.32	1.0287	0.38
α	1.1064	0.0001***	0.8510	2.108×10^{-08}***	0.7079	2.2×10^{-16}***	1.3953	2.2×10^{-16}***
CQ_{EGG}	185.1852	0.009**	0.0011	0.001**	0.0121	0.15	49.7512	0.18
Qcs	0.9761	0.71	1.0537	0.44	1.2594	0.014*	0.7958	0.012*

researchers (Estill, 1988; Miles and Hollien, 1990; Sundberg *et al.*, 1993; Lovetri *et al.*, 1999; Edwin, 2004; Balog, 2005; Burdick, 2005; Titze and Worley, 2009; Titze *et al.*, 2011; Bourne and Kenny, 2016). Nevertheless, for five of the singers, both belt and legit qualities were produced with a close distance between the frequency of the first vocal tract resonance (f_{R1}) and that of the second voice harmonic ($2f_0$), so that the two qualities differed by slight articulatory modifications rather than by two fundamentally different articulatory strategies. Only singer S1, again, demonstrated a different tuning strategy between both qualities (f_{R1}:$2f_0$ in belt but not in legit), conjointly with significant variations in glottal parameters. If one interprets these results as a change in laryngeal mechanism from belt to legit for this singer, then his resonance modifications are consistent with those observed in male operatic singers in laryngeal mechanisms M1 and M2 (Henrich *et al.*, 2014). On the contrary, they do not follow the trend of the operatic tenor in Echternach (2010), who made minimal changes in vocal tract adjustment when transitioning from modal to falsetto voice.

The greater alpha ratios observed in belt are also consistent with the increased SPL, and with the variation of glottal parameters that was observed in the singers' production. It may reflect a louder and brighter sound with flatter spectral slope, with greater perceived effort, as described by expert teachers (Stanley, 1929; Estill, 1980; Edwin, 2004; AATS, 2008; LeBorgne *et al.*, 2010).

The singers were also able to produce two distinctive subcategories of belt that differed significantly in glottal and acoustical descriptors. Again, the slightly higher f_{R1} frequencies observed in chesty belt, compared to twangy belt, may reflect a slightly more open articulation. The slightly higher Qcs and CQ_{EGG} values of chesty belt may be related to the greater SPL values of that voice quality, compared to twangy belt, but may not correspond to a significant change in laryngeal mechanism. Interestingly, variations of the alpha ratio did not follow those of the SPL. Thus, the greater alpha ratios observed in twangy belt may reflect a specific spectral enhancement in high frequencies that relates to articulatory adjustments rather than from the spectral composition of the voice source. Such specific enhancement of voice energy in the 2–4 kHz region has already been associated with perceived "brightness," "ring," or "twang" in the voice of male operatic singers who demonstrate a singing formant (Sundberg, 1974, 2001), in the projected voice of pop singers (Borch and Sundberg, 2002), male stage actors (Nawka *et al.*, 1997; Pinczower and Oates, 2005), and in the voice of country singers (Cleveland *et al.*, 2001).

C. How do these qualities used in male MT singing compare to male classical singing?

Since there are no published studies comparing belt and legit in the male music theatre voice, we can only compare our results with studies of generic music theatre and classical qualities. A comparative study of a male belt singer and an operatic singer noted a more open mouth and higher larynx, with a more forward tongue in the belt sound (Titze and Worley, 2009). Bjorkner (2008) observed higher frequencies of the first

TABLE V. Summary table of the differences observed in the male voice between belt and legit qualities, and between chesty and twangy belt qualities, in comparison to the differences observed in the female voice and reported in a previous article (Bourne and Garnier, 2012). Non-significant differences are reported with the symbol "ns."

	Belt-Legit		Chesty-Twangy	
	Men	Women	Men	Women
f_{R1}	ns (bottom) to 85 Hz (top)	187 Hz	ns (bottom) to 35 Hz (top)	ns
f_{R2}	[e]: 158 Hz [ɔ]: ns (bottom) to 291 Hz (top)	205 Hz	[e]: −90 Hz (bottom) to +192 Hz (top) [ɔ]: −123 Hz	−66 Hz
CQ_{EGG}	0.027	0.21	[e]: 0.017 [ɔ]: ns	ns
Qcs	3.45	0.76	0.92	0.22
SPL	ns (bottom) to 6.5 dB (top)	10.7 dB	2.6 dB (bottom) to 5.5 dB (top)	2 dB
α	1.8 dB (bottom) to 5.5 dB (top)	4.4 dB	−1.8 dB	−1.7 dB

two resonances and higher CQ and SPL values in music theatre singers compared to classical singers. These results compare to some extent with the differences in f_{R1}, f_{R2}, and CQ_{EGG} values that we observed between belt and legit qualities.

Sundberg *et al.* (2011) compared formant tunings in classical and non-classical singers performing nine note scales from E4–G4 on the vowels [ae], [a], [u], and [i] and found that CCM singers tuned $F1$ at or above the second harmonic, unlike the classical singers who tuned the first formant below the second harmonic. These results are consistent with the f_{R1}:$2f_0$ tuning that we observed here in male singers for belt qualities as well as legit, albeit with a lesser proximity of tuning for legit.

While legit can be considered similar to classical voice production in terms of acoustic and physiological parameters, there are likely to be stylistic differences such as onset of sound, duration and amplitude of vibrato, vowel length, and other approaches to vocal phrasing that distinguish these qualities from each other. These questions would be well worth examining in future studies.

D. How do these qualities used in male MT singing compare with female MT singing?

In this study of male voices, voice parameters were found to vary in similar ways between belt and legit qualities, and chesty and twangy belt qualities, to those observed in our previous study of the female voice (Bourne and Garnier, 2012). Thus, for male as well as female voices, belt tended to be produced with greater SPL and alpha ratios than legit, higher f_{R1} and f_{R2} values, and higher CQ_{EGG} and Qcs values. Likewise, for both genders, chesty belt tended to be produced with greater SPL and lower alpha ratios than twangy belt, higher f_{R1} and lower f_{R2} values, and higher CQ_{EGG} and Qcs values.

The differences between belt and legit tended to be greater in women than men (except for the parameter Qcs) and were observed over the whole belt range whereas for men, differences tended to be significant at the top of the belt range only (see Table V). The greatest difference between male and female results tended to be in the comparison of CQ_{EGG} values in belt and legit [$\Delta = 0.21$ on average for women, with very distinct ranges of values in the two qualities: around 0.6 for belt and 0.3 for legit; $\Delta = 0.027$ on average for men, with a comparable range of values (0.45–0.65) for both qualities]. The exception was male singer S1 who produced CQ_{EGG} values lower than 0.45 for legit. Another difference between the results for men and women was in their strategies for tuning the first vocal tract resonance: Female singers consistently tuned f_{R1} to the second voice harmonic ($2f_0$) for belt but not for legit, whereas all men except singer S1 demonstrated the same resonance tuning strategy (f_{R1}:$2f_0$ tuning) for both qualities. Again, the exception was male singer S1 who followed the same trend as female singers. These combined observations support the idea that both qualities may be produced in the same laryngeal mechanism (M1) by men, whereas women may produce belt in M1 and legit in M2.

On the other hand, the differences between chesty belt and twangy belt tended to be comparable for both genders or slightly greater in men than women (see Table V). For both male and female voices, CQ_{EGG} values remained in a similar range in chesty belt and twangy belt qualities, and the same resonance tuning strategy (f_{R1}:$2f_0$ tuning) was observed for both qualities, suggesting that these two subtypes of belt may be produced in the same laryngeal mechanism by both genders, though with subtle laryngeal and vocal tract adjustments.

V. CONCLUSION AND PEDAGOGICAL IMPLICATIONS

Results from our study suggest that belt and legit qualities can be defined in the male voice and distinguished by significantly different physiological and acoustical features that can be measured objectively and observed perceptually. We found that belt differed from legit by higher alpha ratios, Qcs values and CQ_{EGG} values over the whole belt range, and by higher f_{R1} at the upper pitch range. Belt also differed from legit by higher SPL and f_{R2} values over the whole belt range for [e] vowels, but only at the upper pitch range for [ɔ] vowels. Although these differences were significant and similar to those observed for female singers, there was a smaller degree of difference for males, indicating a more subtle distinction between these qualities at the source and vocal tract than was observed in female singers. Male singers produced both belt and legit with a similar resonance tuning strategy (f_{R1} to $2f_0$) for the two vowels [e] and [ɔ] and may use the same laryngeal mechanism (M1), unlike female singers who produced these qualities with fundamentally different glottal and resonance tuning strategies.

Furthermore, our results also suggest that at least two categories of belt; chesty and twangy, can be defined in the male voice, and can be distinguished physiologically, acoustically, and perceptually. Results were similar to those for female music theatre singers: Chesty belt was louder than twangy belt, with higher resonance frequencies and higher CQ_{EGG} values, possibly related to greater adduction of the vocal folds in chest or to heavier registration, but most likely not related to a change in laryngeal mechanism. Both belt qualities were produced with similar resonance strategies (f_{R1} to $2f_0$).

The production of belt and legit may require different pedagogical approaches when teaching both male and female singers, due to the significantly distinct glottal and vocal tract configurations of each style. Subtle technical adjustments rather than fundamentally different pedagogical approaches are likely to be most appropriate for teaching different types of belt.

ACKNOWLEDGMENTS

We thank our volunteer subjects. T.B. was supported by an Early Career Researcher Grant from the University of Ballarat. This work has also been partially supported by Australian Research Council Grant No. DP 0771208 and by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01). We thank the anonymous reviewers for their refined comments and advice, as well as Nathalie Henrich for fruitful discussions on the analysis of the EGG signal, Joe Wolfe, John Smith, and Noel Hanna on the measurement of vocal tract impedance.

[1]This pitch range also corresponds to the typical transition range and possible overlap between the two laryngeal mechanisms M1 and M2 (Roubeau *et al.*, 2009).

[2]It should be noted that maxima of the DEGG signal may not always coincide exactly with the moments of glottal closure and opening, as observed directly from simultaneous videokymographic or high speed video imaging (Herbst *et al.*, 2010; Herbst *et al.*, 2014). Phase differences along the superior-inferior (Baer, 1981; Titze *et al.*, 1993) and anterior–posterior dimensions (Tanabe *et al.*, 1975; Krenmayr *et al.*, 2012; Orlikoff *et al.*, 2012; Yamauchi *et al.*, 2013) suggest that contacting and de-contacting of the vocal folds should be seen as an interval of time during which closing and opening occur (Herbst *et al.*, 2014). Despite these limitations, EGG is still a reliable non-invasive method that provides information about vocal fold vibration and the likely underlying laryngeal mechanism: Indeed, both CQ_{EGG} and Qcs parameters extracted from the EGG signal decrease significantly during a transition from laryngeal mechanism M1 to M2 (on a glide or a decrescendo); CQ_{EGG} tends to be greater in M1 compared to M2, corresponding to a longer contacting time over the glottal cycle; Qcs tends to be weak in laryngeal mechanism M2 (Qcs $\sim$ 1) and significantly greater in laryngeal mechanism M1 (Qcs $\sim$ 4), corresponding to a greater degree of asymmetry of the EGG signal and a greater contacting speed of the vocal folds (Henrich *et al.*, 2003; Roubeau *et al.*, 2009).

AATS (**2008**). "American Academy of Teachers of Singing: In support of contemporary commercial music (nonclassical) voice pedagogy," An American Academy of Teachers of Singing Paper, pp. 1–4.

Baer, T. (**1981**). "Observation of vocal fold vibration: Measurements of excised larynges," in *Vocal Fold Physiology*, edited by K. N. Stevens and M. Hirano(University of Tokyo Press, Tokyo), pp. 119–133.

Balog, J. E. (**2005**). "A guide to evaluating music theater singing for the classical teacher," J. Singing **61**, 401–406.

Bestebreurtje, M., and Schutte, H. K. (**2000**). "Resonance strategies for the belting style: Results of a single female subject study," J. Voice **14**, 194–204.

Bjorkner, E. (**2008**). "Musical theatre and opera singing—Why so different? A study of subglottal pressure, voice source, and formant frequency characteristics," J. Voice **22**, 533–540.

Bjorkner, E., Sundberg, J., Cleveland, T., and Stone, E. (**2006**). "Voice source differences between registers in female musical theatre singers," J. Voice **20**, 187–197.

Borch, D. Z., and Sundberg, J. (**2002**). "Spectral distribution of solo voice and accompaniment in pop music," Log. Phon. Vocol. **27**, 37–41.

Bourne, T., and Garnier, M. (**2012**). "Physiological and acoustic characteristics of the female music theatre voice," J. Acoust. Soc. Am. **131**, 1586–1594.

Bourne, T., and Kenny, D. T. (**2016**). "Vocal qualities in music theater voice: Perceptions of expert pedagogues," J. Voice **30**, 128.e1–128.e12.

Burdick, B. (**2005**). "Vocal techniques for music theater: The high school and undergraduate singer," J. Singing **61**, 261–268.

Cleveland, T., Sundberg, J., and Stone, E. (**2001**). "LTAS characteristics of country singers during speaking and singing," J. Voice **15**, 54–60.

Echternach, M., Sundberg, J., Markl, M., and Richter, B. (**2010**). "Professional opera tenors' vocal tract configurations in registers," Folia Phoniatrica et Logopaedica **62**, 278–287.

Edwin, R. (**2003**). "A broader Broadway," J. Sing. **59**, 431–432.

Edwin, R. (**2004**). "Belt yourself," J. Sing. **60**, 285–288.

Epps, J., Smith, J., and Wolfe, J. (**1997**). "A novel instrument to measure acoustic resonances of the vocal tract during speech," Measure. Sci. Technol. **8**, 1112–1121.

Estill, J. (**1980**). "Observations about the quality called belting," in *The Ninth Symposium, Care of the Professional Voice*, edited by B. Weinberg and V. Lawrence (Voice Foundation, The Juilliard School, New York), pp. 82–88.

Estill, J. (**1988**). "Belting and classic voice quality: Some physiological differences," Med. Prob. Perform. Art. **3**, 37–43.

Frøkjaer-Jensen, B., and Prytz, S. (**1976**). "Registration of voice quality," Brüel & Kjaer Tech. Rev. **3**, 3–17.

Garnier, M., Henrich, N., Smith, J., and Wolfe, J. (**2010**). "Vocal tract adjustments in the high soprano range," J. Acoust. Soc. Am. **127**, 3771–3780.

Henrich, N. (**2006**). "Mirroring the voice from Garcia to the present day: Some insights into singing voice registers," Logopedics Phon. Voc. **31**, 3–14.

Henrich, N., D'Alessandro, C., Doval, B., and Castellengo, M. (**2004**). "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation," J. Acoust. Soc. Am. **115**, 1321–1332.

Henrich, N., D'Alessandro, C., Doval, B., and Castellengo, M. (**2005**). "Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency," J. Acoust. Soc. Am. **117**, 1417–1430.

Henrich, N., Roubeau, B., and Castellengo, M. (**2003**). "On the use of electroglottography for characterisation of the laryngeal mechanisms," in *Stockholm Music Acoustics Conference* (Stockholm, Sweden), pp. 1–4.

Henrich, N., Smith, J., and Wolfe, J. (**2014**). "Vocal tract resonances in singing: Variation with laryngeal mechanism for male operatic singers in chest and falsetto registers," J. Acoust. Soc. Am. **135**, 491–501.

Herbst, C., Fitch, W. T. S., and Svec, J. (**2010**). "Electroglottographic wavegrams: A technique for visualizing vocal fold dynamics noninvasively," J. Acoust. Soc. Am. **128**, 3070–3078.

Herbst, C., Lohscheller, J., Švec, J., Henrich, N., Weissengruber, G., and Fitch, T. (**2014**). "Glottal opening and closing events investigated by electroglottography and super-high-speed video recordings," J. Exp. Biol. **217**, 955–963.

Herbst, C., Qiu, Q., Schutte, H. K., and Svec, J. (**2011**). "Membranous and cartilaginous vocal fold adduction in singing," J. Acoust. Soc. Am. **129**, 2253–2262.

Herbst, C., Ternstrom, S., and Svec, J. (**2009**). "Investigation of four distinct glottal configurations in classical singing—A pilot study," J. Acoust. Soc. Am. **125**, EL104–EL109.

Joliveau, E., Smith, J., and Wolfe, J. (**2004**). "Vocal tract resonances in singing: The soprano voice," J. Acoust. Soc. Am. **116**, 2434–2439.

Kochis-Jennings, K. A., Finnegan, E. M., Hoffman, H. T., and Jaiswal, S. (**2012**). "Laryngeal muscle activity and vocal fold adduction during chest, chestmix, headmix, and head registers in females," J. Voice **26**, 182–193.

Krenmayr, A., Wöllner, T., Supper, N., and Zorowka, P. (**2012**). "Visualizing phase relations of the vocal folds by means of high-speed videoendoscopy," J. Voice **26**, 471–479.

LeBorgne, W. D., Lee, L., Stemple, J. C., and Bush, H. (**2010**). "Perceptual findings on the Broadway belt voice," J. Voice **24**, 678–689.

Lebowitz, A., and Baken, R. J. (**2011**). "Correlates of the belt voice: A broader examination," J. Voice **25**, 159–165.

Lovetri, J., Lesh, S., and Woo, P. (**1999**). "Preliminary study on the ability of trained singers to control the intrinsic and extrinsic laryngeal musculature," J. Voice **13**, 219–226.

Miles, B., and Hollien, H. (**1990**). "Whither belting?," J. Voice **4**, 64–70.

Nawka, T., Anders, L. C., Cebulla, M., and Zurakowski, D. (**1997**). "The speaker's formant in male voices," J. Voice **11**, 422–428.

Orlikoff, R. F., Golla, M. E., and Deliyski, D. D. (**2012**). "Analysis of longitudinal phase differences in vocal-fold vibration using synchronous high-speed video-endoscopy and electroglottography," J. Voice **26**, 816.e13–816.e20.

Pinczower, R., and Oates, J. (**2005**). "Vocal projection in actors: The long-term average spectral features that distinguish comfortable acting voice from voicing with maximal projection in male actors," J. Voice **19**, 440–453.

Roubeau, B., Henrich, N., and Castellengo, M. (**2009**). "Laryngeal vibratory mechanisms: The notion of vocal register revisited," J. Voice **23**, 425–438.

Schutte, H. K., and Miller, D. G. (**1993**). "Belting and pop, nonclassical approaches to the female middle voice: Some preliminary conclusions," J. Voice **7**, 142–150.

Stanley, D. (**1929**). *The Science of Voice* (Carl Fischer, New York).

Sundberg, J. (**1974**). "Articulatory interpretation of the 'singing formant,'" J. Acoust. Soc. Am. **55**, 838–844.

Sundberg, J. (**2001**). "Level and center frequency of the singer's formant," J. Voice **15**, 176–186.

Sundberg, J., Gramming, P., and Lovetri, J. (**1993**). "Comparisons of pharynx, source, formant, and pressure characteristics in operatic and musical theatre singing," J. Voice **7**, 301–310.

Sundberg, J., La, F., and Gill, B. (**2011**). "Professional male singers' formant tuning strategies for the vowel /a/," Logopedics Phon. Vocol. **36**, 156–167.

Sundberg, J., and Nordenberg, M. (**2006**). "Effects of vocal loudness variation on spectrum balance as reflected by the alpha measure of long-term-average spectra of speech," J. Acoust. Soc. Am. **120**, 453–457.

Tanabe, M., Kitajima, K., Gould, W. J., and Lambiase, A. (**1975**). "Analysis of high speed motion pictures of the vocal folds," Folia Phoniatrica **27**, 77–87.

Titze, I., Jiang, J. J., and Hsiao, T. Y. (**1993**). "Measurement of mucosal wave propagation and vertical phase difference in vocal fold vibration," Ann. Otol. Rhinol. Laryngol. **102**, 58–63.

Titze, I., and Worley, A. (**2009**). "Modeling source-filter interaction in belting and high-pitched operatic male singing," J. Acoust. Soc. Am. **126**, 1530–1540.

Titze, I., Worley, A., and Story, B. (**2011**). "Source-vocal tract interaction in female operatic singing and theater belting," J. Sing. **67**, 561–572.

Yamauchi, A., Imagawa, H., Sakakibara, K., Yokonishi, H., Nito, T., Yamasoba, T., and Tayama, N. (**2013**). "Phase difference of vocally healthy subjects in high-speed digital imaging analyzed with laryngotopography," J. Voice **27**, 39–45.