

Automatic Assessment of Acoustic Parameters of the Singing Voice: Application to Professional Western Operatic and Jazz Singers

*Claudia Manfredi, *Davide Barbagallo, †Giovanna Baracca, *Silvia Orlandi, *Andrea Bandini, and ‡,§Philippe H. Dejonckere, *†Florence and Milano, Italy, and ‡Leuven and §Brussels, Belgium

Summary: Introduction. The obvious perceptual differences between various singing styles like Western operatic and jazz rely on specific dissimilarities in vocal technique. The present study focuses on differences in vibrato acoustics and in singer's formant as analyzed by a novel software tool, named *BioVoice*, based on robust high-resolution and adaptive techniques that have proven its validity on synthetic voice signals.

Material and Methods. A total of 48 professional singers were investigated (29 females; 19 males; 29 Western operatic; and 19 jazz). They were asked to sing "a cappella," but with artistic expression, a well-known musical phrase from Gershwin's *Porgy and Bess*, in their own style: either operatic or jazz. A specific sustained note was extracted for detailed vibrato analysis. Beside rate (s^{-1}) and extent (cents), duration (seconds) and regularity were computed. Two new concepts are introduced: vibrato jitter and vibrato shimmer, by analogy with the traditional jitter and shimmer of voice signals. For the singer's formant, on the same sustained tone, the ratio of the acoustic energy in formants 1–2 to the energy in formants 3, 4, and 5 was automatically computed, providing a quality ratio (QR).

Results. Vibrato rates did not differ among groups. Extent was significantly larger in operatic singers, particularly females. Vibrato jitter and vibrato shimmer were significantly smaller in operatic singers. Duration of vibrato was also significantly longer in operatic singers. QR was significantly lower in male operatic singers.

Conclusions. Some vibrato characteristics (extent, regularity, and duration) very clearly differentiate the Western operatic singing style from the jazz singing style. The singer's formant is typical of male operatic singers. The new software tool is well suited to provide useful feedback in a pedagogical context.

Key Words: Singing voice–Vibrato–Singing formant–Western operatic–Jazz–Singing style–Acoustic analysis.

INTRODUCTION

Traditionally, singing pedagogy relies for an important part on the auditory perception of the teacher and the pupil. Teaching is essentially practice focused, supported by linguistic imagery and, to some extent, by vocal and postural modeling. Singing teachers mainly draw on their personal experience within an essentially oral culture.¹ This applies for, eg, technical aspects (forward placement, appoggio, ring, and so forth) as well as for singing styles (legit, chesty belt, barbershop, and so forth). Actually, a valuable artistic imitation requires neither an analytical process nor an instrumental approach. However, for centuries, singers have been eager for relevant extra-auditory information to either better learn (or teach) how experienced artists are achieving some skills or to receive useful extra-auditory feedback in real time or at least in short term. A well-known example is the flame of a candle kept in front of the mouth during a "messa di voce" exercise for visually controlling the airflow.² In the more recent years, new techniques have emerged that are providing the singing student as well as the singing teacher with clearer representations of aspects of voice. Examples are given by the improvements in the noninvasiveness of endoscopes, vid-

eokymography, and high-speed video recording, magnetic resonance imaging, and so forth,³ as well as by programs providing spectrographic displays, like Voce Vista[®] (www.vocevista.com/dgmiller.html). The information provided may be of pedagogical relevance because many differences in voice qualities are a reflection of changes in muscular and aerodynamic conditions, which are to some extent under direct voluntary control.⁴ Nevertheless, these techniques are primarily used for research purposes in singing science and are not normally available in the singing studio or in the classroom.

As regards singing styles, their obvious perceptual differences rely on clearly distinct vocal techniques and behaviors.⁵ During the past 15 years, quite a lot of research has been devoted to "nonclassical," or "nonlegit" singing although most studies deal only with a limited number of singers or even one single singer.^{4–9} However, a large debate remains among singing teachers on exactly defining the differential properties of styles like Broadway, pop, soul, country, folk, metal, and so forth. Singing techniques like belting also elicit controversies, demonstrating a lack of understanding within the pedagogic voice community.¹⁰ Descriptions of these different techniques have typically evolved from singer's subjective opinions concerning their own performances. Even a majority opinion may not accurately reflect physiological or physical reality.⁵ In the last years, however, specific high-level education programs in "nonclassical" professional singing (and particularly "jazz") have been created in numerous Conservatories or Universities of Music. This suppresses an important research bias because differences in voicing technique can no longer be related to differences in level of (musical and vocal) education or training,³

Accepted for publication September 10, 2014.

From the *Department of Information Engineering, Università degli Studi di Firenze, Florence, Italy; †Division of Head and Neck Surgery, IEO - Istituto Europeo di Oncologia, Milano, Italy; ‡Department of Neurosciences, Exp. ORL, Catholic University of Leuven, Leuven, Belgium; and the §Federal Institute of Occupational Diseases, Brussels, Belgium.

Address correspondence and reprint requests to Philippe H. Dejonckere, Federal Institute of Occupational Diseases, 1, Avenue de l'Astronomie, B-1210 Brussels, Belgium. E-mail: philippe.dejonckere@med.kuleuven.be

Journal of Voice, Vol. 29, No. 4, pp. 517.e1-517.e9
0892-1997/\$36.00

© 2015 The Voice Foundation

<http://dx.doi.org/10.1016/j.jvoice.2014.09.014>

but solely to stylistic purposes, certainly if the singer specialized in “modern” styles also disposes of the whole technical background as the one performing in the classic Western operatic style. All singers in the present study may be considered having comparable technical vocal skills; thus, if, eg, the jazz singer uses less vibrato or singer’s formant, it is his or her stylistic choice and not a limitation of his vocal skills. All these considerations support the need of better analysis of singing styles, particularly with a tool that is noninvasive, user friendly, and that can—when used for short-term biofeedback—be suited for teaching purposes within the singing studio.

Another relevant methodological issue when comparing singing styles is the selection of the vocal material. For the present study, we chose a well-known musical phrase from Gershwin’s opera *Porgy and Bess*: “Summertime, and the livin’ is easy.” “Summertime” (1935) has become one of the most covered songs in the world and is well known by every singer. Its peculiarity is that it is suited for interpretation in Western operatic style as well as in jazz style. Furthermore, “Summertime” also resonates on a deep emotional level and stimulates the expressivity of the singer. A specific sustained note was extracted for detailed vibrato and timbre analyses.

Singers in general and operatic (especially male) singers, in particular, strive to enrich and enhance their singing voice by developing the singer’s formant, which provides the voice with a special “ring,” as well as causing an increase in signal intensity, resulting from the clustering of the third, fourth, and fifth formants near 3 kHz.^{11–13} This is attained by lowering the larynx and widening the hypopharynx. Thus, it is an articulatory phenomenon within the vocal tract that enhances the resonance. As an acoustic correlate, this clustering of formants promotes extra energy in the higher frequency range, allowing the singer to be heard without amplification over an accompanying orchestra.

Vocal vibrato is recognized as a prominent characteristic of classical Western operatic singing, but it is also used by modern and jazz singers. The mechanism of vibrato production is not yet fully understood, but it is plausibly the result of a long-latency neuromuscular reflex loop involving antagonist muscles that can change vocal fold length and tension. Singers appear to increase the gain in the reflex loop to cultivate the vibrato, which grows out of a spectrum of 0–15 Hz physiologic tremors in raw form.¹⁴ It corresponds to a periodic low frequency modulation of fundamental frequency (f_0).

Summarizing, the aim of the present work is as follows:

- (1) to apply a recently developed software tool, *BioVoice*, a tool that has proven its validity on synthetic normal and deviant voice signals^{15–19} for acoustic analysis on a large sample of highly educated professional singers from two clearly distinct styles: Western operatic and jazz. The software tool is based on specific robust high-resolution and adaptive techniques that are required for this application.^{20,21} Two typical stylistic aspects were investigated: the vibrato and the singer’s formant.
- (2) to quantify the expected acoustic differences pertaining to the vibrato and the singer’s formant between the two

singing styles and to check if these differences as demonstrated and visualized by the software tool are of such nature and extent that they could be used in education and training by providing the singing student with short-term feedback.

MATERIAL AND METHODS

Subjects

Forty-eight professional singers were investigated (29 females, sopranos or mezzo-sopranos; 19 males, tenors or baritones; 29 Western operatic; 19 jazz. Males operatic: 14; males jazz: 5; females operatic: 15; females jazz: 14; Age ranges 21–32 years for females and 21–54 years for males). All of them received higher vocal and musical education (at least 3 years of experience) and were free of voice problems or complaints. They were recruited via the Luigi Cherubini Conservatory in Firenze and via private singing schools in Milano and Ravenna (Italy).

All gave their informed consent to participate in this study.

Vocal material and recording conditions

Every singer was asked, after vocal warm-up, to sing “a capella” but with artistic expression and in his/her own musical style (operatic or jazz), a well-known musical phrase from Gershwin’s opera *Porgy and Bess*: “Summertime, and the livin’ is easy.” Each performance was produced twice so that the better might be selected for analysis. In this experiment, no tonal key was imposed to the singers, who were allowed to sing in the tone they felt as the most comfortable one. They were also asked to adapt their loudness to the size of the room (a quiet small auditorium). A specific sustained note was extracted for detailed vibrato analysis: the sustained /aIm/ from /sAmətaIm/.

Recordings were made with a sound board (Tascam US-144-MK2; Tokyo, Japan) and a cardioid unidirectional microphone (Shure SM58, frequency response: 50–15,000 Hz) connected to a computer. The distance between the microphone and the mouth was kept fixed at 20 cm. Sample frequency was 44.1 kHz.

Analyzed parameters

Vibrato characteristics. Vibrato primarily consists of a periodic f_0 modulation.²² Traditionally, the main parameters used to characterize the vibrato are the rate and the extent.

Vibrato rate (Vrate) (s^{-1}) represents the number of f_0 oscillations per second. It is evaluated as the reciprocal of the mean time difference between two subsequent f_0 maxima:

$$\text{Vrate} = \frac{1}{N} \sum_{i=1}^{N-1} \left| \frac{1}{t_{f_{0\max}}^{i+1} - t_{f_{0\max}}^i} \right| \quad (1)$$

where $t_{f_{0\max}}^i$ is the time instant corresponding to the i th maximum of f_0 (ie, the i th cycle) and N is the number of f_0 maxima. It is important to define limits of pulsation rate that may be considered as vibrato. Relying on the observations of Ekholm et al²³ and Ferrante,²⁴ the extreme range may be defined as 4.2–8.1 Hz in females and 4.8–6.6 Hz in males (mean \pm 2

standard deviation [SD]). For security, and because there is no study specifically pertaining to jazz singers, the range 4.2–8.1 Hz was taken for both genders. Henceforth, in the computation, only differences $t_{f_{0\max}}^{i+1} - t_{f_{0\max}}^i$ that are within the interval 4.2–8.1 Hz are taken into account.

Vibrato extent (Vext) is the difference in frequency between a maximum and a minimum of f_0 within a cycle (ie, the amplitude of the cycle). Here, it has been computed as the mean of the differences between f_0 maximum and f_0 minimum in each cycle:

$$\text{Vext} = \frac{1}{N} \sum_{i=1}^N (f_{0\max}^i - f_{0\min}^i) \quad (2)$$

where $f_{0\max}^i$ is the i th maximum of f_0 and $f_{0\min}^i$ the corresponding minimum.

To make the values for Vext comparable among singers, cents of semitone are computed on each cycle ($i = 1, \dots, N$) according to the relation:

$$\text{Cent}(i) = 1200 * 3.322038403 * \log_{10} \left(\frac{f_{0\max}^i}{f_{0\min}^i} \right) \quad (3)$$

In the same way as for rate, limits need to be defined for extent values that may be considered compatible with vibrato. According to reports by Ferrante²⁴ and Anand et al,²⁵ the extremes should be 17–145 cents for females and 38–130 cents for males. Actually, for extent, the relevant limit is the lower one. Once again, because no reports are available for jazz singers, the threshold of 17 cents was taken for all singers. Henceforth, Vext is computed excluding those cycles that exhibit an extent less than 17 cents.

Regularity of vibrato

Moreover, two novel parameters are implemented as a measure of vibrato regularity, by analogy with the usual “perturbation” parameters in voice signals: jitter and shimmer. They are called here Jvib and Svib, respectively:

$$\text{Jvib} = 100 * \frac{\frac{1}{N_c - 1} \sum_{i=1}^{N_c - 1} |t_{f_{0\max}}^{i+1} - t_{f_{0\max}}^i|}{\frac{1}{N_c} \sum_{i=1}^{N_c} t_{f_{0\max}}^i} \quad (4)$$

$$\text{Svib} = 100 * \frac{\frac{1}{N_c - 1} \sum_{i=1}^{N_c - 1} |\text{Vamp}_{i+1} - \text{Vamp}_i|}{\frac{1}{N_c} \sum_{i=1}^{N_c} \text{Vamp}_i} \quad (5)$$

where $t_{f_{0\max}}^i$ and Vamp_i are respectively the i th vibrato cycle and its amplitude and N_c is the number of cycles (possibly $N_c < N$ due to the boundaries imposed both on Vrate and Vext).

Duration of vibrato

Once the boundaries were defined as described previously for what could be considered as vibrato within the /aIm/ from /sAmətaIm/, the duration of this vibrato part was also measured for comparisons.

The software tool further computes the average tone pitch during the vibrato as the mean of maximum and minimum frequency (hertz) in the i th cycle.²⁶

Resonance characteristics (singer’s formant). The singer’s formant consists of an increase in signal intensity due to clustering of the third, fourth, and fifth formants, approximately in the range 2500–4500 Hz.¹¹ A first attempt to objectively measure the singer’s formant was made by Omori et al.²⁷ They computed the ratio of the highest harmonic peak between 2 and 4 kHz to the highest harmonic peak between 0 and 2 kHz and termed it “singing power ratio”. It was calculated semimanually from the Fast Fourier Transform (FFT) power spectrum. Actually, in the ideal case, a quality ratio (QR) should be defined as the ratio of the area under the curve of the Power Spectrum Density (PSD) plot corresponding to the cluster of the first two formants ($\text{Area}_{1,2}$) and the area corresponding to the third, fourth, and fifth formants ($\text{Area}_{3,4,5}$):

$$\text{QR} = \frac{\text{Area}_{1,2}}{\text{Area}_{3,4,5}} \quad (6)$$

Thus, the stronger the singer’s formant, the smaller the QR. Considering that the unit of the vertical axis of a PSD plot is the decibel, a logarithmic transformation is applied. For obtaining an automatic computation of the QR, a major difficulty is finding the reference “threshold frequencies,” F_{ref} , that delimit the PSD integral into $\text{Area}_{1,2}$ and $\text{Area}_{3,4,5}$. Actually, clusters depend on the PSD shape that may vary significantly among singers, sung materials, and singing styles. The estimation method of the QR must therefore be as general as possible, without requiring changes in the program or any manual intervention.

In previous work, both “static thresholds,” $F_{\text{ref_inf}}$, $F_{\text{ref_sup}}$, respectively, set at 2500 Hz and 5500 Hz (ie, midpoint between 2000 and 3000 Hz, and between 5000 and 6000 Hz, approximately corresponding to the second and third formant and the fifth and sixth formant, respectively) and several “dynamic thresholds” have been defined and tested.²⁸ However some mismatch was found because of signal variability. In this work, the following “mixed criterion” was implemented that overcomes problems encountered with previous methods, such as the presence of spurious maxima in the range of interest that may lead to errors. First, the number P of PSD maxima in the range (0–6 kHz) is computed. If $P > 5$, then the PSD is smoothed with a median filter of order 3. If there are still more than five peaks in the smoothed PSD, a median filter of order five is applied to the original PSD. If there are still more than five peaks in the PSD, no further filtering is applied, to avoid excessive smoothing that would result in a distortion of the PSD. Then, $F_{\text{ref_inf}}$ ($F_{\text{ref_sup}}$) is set at the frequency value corresponding to the first minimum after the second (fifth) PSD maximum, if such a minimum is found between 1.5 and 2.5 kHz (4.5 and 6 kHz). In fact, it is reasonable to suppose that the second (fifth) PSD maximum corresponds to the second (fifth) formant. Otherwise, if the second (fifth) PSD maximum is found below 1.5 kHz (4.5 kHz), $F_{\text{ref_inf}}$ is set equal to 2.5 kHz ($F_{\text{ref_sup}} = 4.5$ kHz). Hence, the two thresholds are neither set below 1.5 and 4.5 kHz, respectively, nor above 2.5 and 6 kHz, to include the whole spectral contribution of the formants’ cluster.

Technical issues

BioVoice. The proposed tool for singing voice analysis is a part of the more general *BioVoice* tool for voice analysis developed under MATLAB 2013b® (The MathWorks Inc., Natick, MA) and constantly updated. *BioVoice* performs the analysis on subsequent and partially overlapping short time frames (even few milliseconds) whose length varies adaptively, tailored to varying pitch: the higher the f_0 , the shorter the time window. Moreover, *BioVoice* implements a parametric approach for f_0 and formant estimation and the model order q varies according to the signal characteristics. These features were shown to improve results as pitch variations are tracked more precisely and proved effective in many applications, with enhanced results with respect to classic approaches as far as resolution is concerned.^{20,21,29,30} Specifically, a favorable comparison with other methods (*Praat*, *Ampex*, and multidimensional voice program (MDVP)) performed on synthetic signals is provided in^{15–19} as far as voiced parts selection, jitter and noise estimation are concerned. Coordinates of the PSD maxima (three or five) are provided along with their mean and SD value. A color-coded spectrogram is also provided with formant trajectories superimposed. Specifically for singers, the following parameters are computed: f_0 , F1–F5, vibrato rate (also in semitones), vibrato extent (each with the corresponding SD), Jvib, Svib and QR, according to [Equations 1–6](#). Several plots are provided in the color map and with legends that show all the computed parameters. Specifically for singing voice, they are the signal waveform (with possible unvoiced parts highlighted), f_0 (in hertz and semitones), the spectrogram with F1–F5 superimposed, and the PSD with the estimated QR. To the authors' knowledge, no other software tool provides such data.

Furthermore, most common software tools require the manual setting of the model order q and other parameters (length of the analysis window, frequency range of interest, spectrum estimation method, and so forth), thus implying some expertise required from the user. Default values are in fact commonly set to deal with the f_0 range relative to adult's speech; thus, they are unsuitable for the present application and could distort results. With *BioVoice*, the user does not have to worry about manual settings as these are performed automatically according to the selected type of analysis to be made through a user-friendly interface that allows uploading the audio file(s), selecting the category (adult male or female, infant, and so forth), and the type of analysis (dysphonia, singing, and so forth). Moreover, it allows uploading multiple signals to avoid loss of time to rescue the individual results. At the end of the processing, *BioVoice* automatically provides a folder for each signal containing all the figures (JPEG) and parameters (Excel table) related to the type of analysis selected. [Figure 1](#) shows the interface with the specific options for the singing voice.

A freely available release of *BioVoice* is in preparation.

Vibrato. Estimating V_{rate} and V_{ext} requires properly finding absolute maxima and minima of f_0 , that is the maximum and minimum values in each vibrato cycle, also in case of irregular

vibrato profile that possibly carries out more than one maximum and minimum (with varying amplitude) in each cycle. Therefore, a robust criterion to discriminate between absolute and relative maxima or minima is required. The problem was afforded in²⁶ where a procedure for absolute maxima identification was proposed based on a fixed threshold value. Here, an improved procedure is implemented that takes into account the boundaries imposed on V_{rate} and V_{ext} after smoothing f_0 with a median filter.

Formants. Formant frequencies are principal analytical features of the speech spectrum. However, a reliable estimation of the formants, and in particular formants F1–F5, is a challenging numerical task as often F1, F2, and F3–F5 are quite close to each other because of cluster. Moreover, often F5 exhibits high SD and low energy.

With *BioVoice*, a robust and high-resolution formant estimation method is implemented based on parametric autoregressive power spectral density (AR PSD) evaluation.^{31,32} One of the main advantages of parametric spectral analysis over classical FFT-based approach relies on its high-resolution capability. This method gives a significant improvement over the accuracy that would be expected from an attempt to extract peaks from the unprocessed speech spectrum. However, AR PSD estimators are sensitive to the underlying model order q : overestimation may result in formant splitting and underestimation smoothes the spectrum causing misallocation of spectral peaks.^{31,32}

To overcome this problem, *BioVoice* automatically selects q from short signal frames linked to the estimated (and varying) f_0 values according to the subject and signal characteristics. Specifically, the relation $q = 2LFs/c$ is used, where F_s = sampling frequency, L = vocal tract length (linked to the patient's age and gender), and c = sound speed: 34 300 cm/s.³³ Roughly, for adult female $L = 13$ cm and for adult males $L = 17$ cm, hence, with $F_s = 44.1$ kHz, $q = 33$ or $q = 44$, respectively. This choice was proven effective in many applications with enhanced results as far as spectral resolution is concerned.^{20,21} The full procedure for the automatic selection of voiced parts, f_0 and formant estimation is described in detail in the study by Fort et al²⁰ and Manfredi et al.²¹

RESULTS

The values for all parameters in the compared groups could be considered as normally distributed (Kolmogorov-Smirnov test), which allows the use of parametric statistics.

First, a breakdown and one-way analysis of variance (ANOVA) was achieved for comparing the two styles in all singers (males and females): all relevant parameters (V_{ext} in cents, Jvib, Svib and duration, as well as QR) except V_{rate} significantly differ between Western operatic and jazz singers. At this stage, tone pitch is not taken in account as there are both males and females in each group.

Second, a Tukey's honest significant difference (HSD) test compared the mean values per parameter:

Rate: NS
 Vext: 0.000044

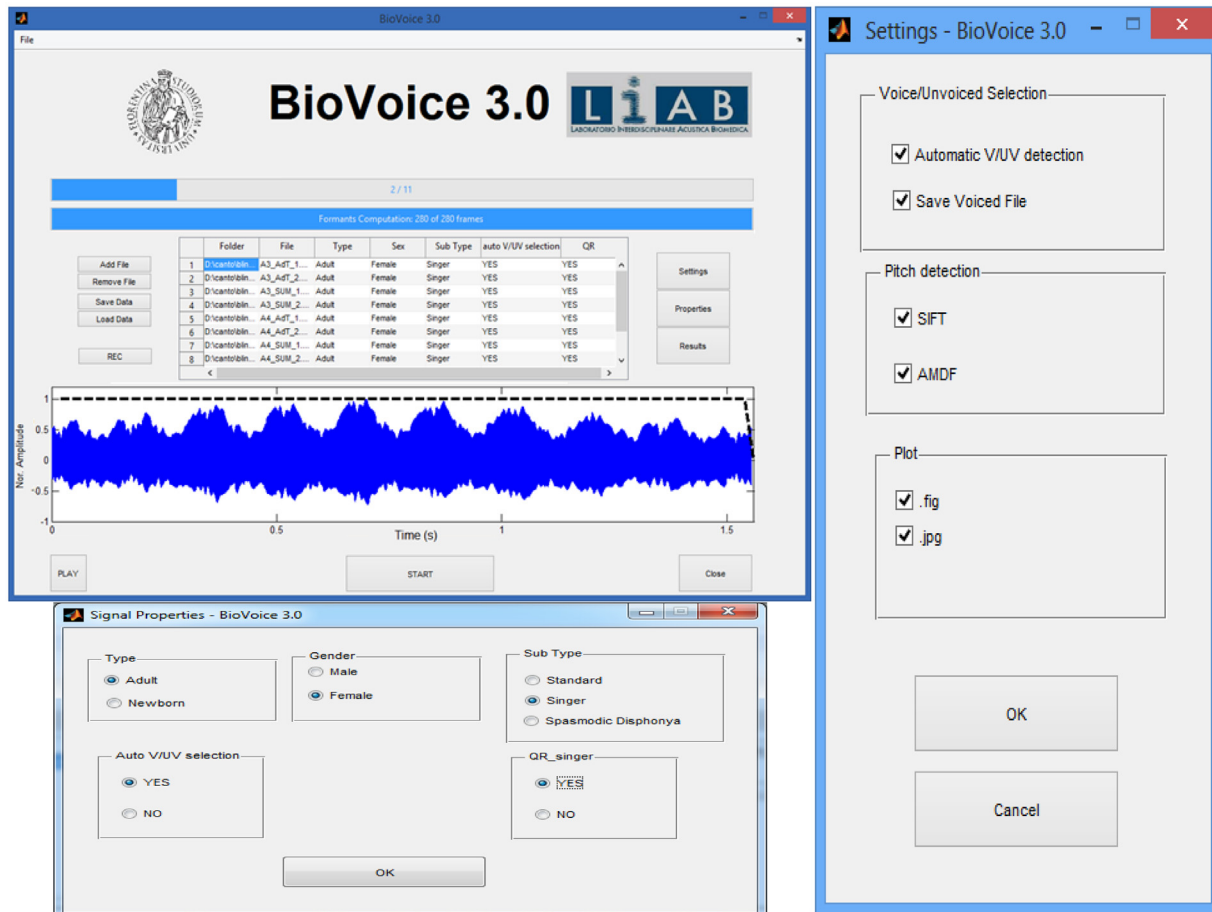


FIGURE 1. BioVoice interface. Upper plot: several audio files can be uploaded and analyzed sequentially. Results for each file are saved in a separate folder. Lower left plot: choice of age, gender, kind of signal, and other options. Lower right plot: More options (default: all): to speed up computations, some can be unchecked.

Jvib: $P = 0.00012$
 Svib: $P = 0.00011$
 duration: $P = 0.00019$
 QR: $P = 0.012$

Third, ANOVA was used for comparing the four groups (male classical, male jazz, female classical, and female jazz). Globally, the same parameters, except again V_{rate} , differ significantly across the four groups (ie, the sample of male classical singers MC, that of male jazz singers MJ, that of female classical singers FC, and that of female jazz singers FJ do not come from the same population). All means ± 1 SD are shown in Table 1.

Fourth, Tukey’s HSD test was again used for comparison of the groups per parameter:

Rate: no single comparison is significant.
 Average tone pitch in hertz:

- MC < FC ($P = 0.00017$) and <FJ ($P = 0.0048$)
- MJ < FC ($P = 0.00017$) and <FJ ($P = 0.00087$)
- FC > FJ ($P = 0.00017$)

Vext in cents:

- MC < FC ($P = 0.00030$)
- MJ < FJ ($P = 0.00086$)
- FC > FJ ($P = 0.000169$)

TABLE 1.
Means ± 1 SD for All Parameters in the Four Groups of Singers

Type of Singer	Average Tone Pitch (Hz)	Vibrato Rate(/s)	Vibrato Extent (cents)	Vibrato Jitter (%)	Vibrato Shimmer (%)	Vibrato Duration (s)	Quality Ratio
Male classical	283.53 \pm 54.06	6.03 \pm 0.95	81.95 \pm 34.58	7.02 \pm 3.31	23.89 \pm 8.08	2.34 \pm 0.77	6.03 \pm 0.95
Male jazz	211.91 \pm 27.91	6.09 \pm 0.62	66.78 \pm 20.04	16.05 \pm 7.34	56.31 \pm 24.25	1.09 \pm 0.32	1.60 \pm 0.67
Female classical	602.14 \pm 118.73	5.82 \pm 0.76	145.43 \pm 51.57	7.49 \pm 2.84	21.12 \pm 10.36	2.62 \pm 0.71	1.51 \pm 0.50
Female jazz	399.50 \pm 81.33	6.16 \pm 0.71	52.89 \pm 16.71	12.63 \pm 3.69	41.49 \pm 10.09	1.23 \pm 0.68	1.60 \pm 0.80

Jvib in percentage:

- MC < MJ ($P = 0.000412$) and <FJ ($P = 0.00298$); no difference MC – FC
- MJ > FC ($P = 0.000628$)
- FC < FJ ($P = 0.0044$); no difference MJ – FJ

Svib in percentage:

- MC < MJ ($P = 0.000183$) and <FJ; no difference MC – FC
- MJ > FC ($P = 0.000171$) no difference MJ – FJ
- FC < FJ ($P = 0.000311$)

QR:

- MC < MJ ($P = 0.013$), <FC ($P = 0.0011$), and <FJ ($P = 0.000494$)

No difference FC – FJ

Duration in seconds:

- MC > MJ ($P = 0.00069$) and >FJ ($P = 0.000852$)
- MJ < FC ($P = 0.000728$)
- FC > FJ (0.00018)

The mean values, with SD and standard error for each parameter in the four groups, are shown in Figures 2–8.

Finally, the correlation between Vext and Vrate was computed: a significant negative correlation appears both in classic ($r = -0.56$; $P < 0.01$) and in jazz singers ($r = -0.49$; $P < 0.05$). Globally, Vext is also positively correlated ($r = 0.51$; $P < 0.01$) with vibrato duration.

DISCUSSION

Operatic and jazz styles

A first interesting finding pertains to the Vrate: there is no significant difference (even no trend) between the operatic and jazz singers. There is also no noticeable gender effect. This could suggest that, referring to Titze's hypothesis¹⁴ and in

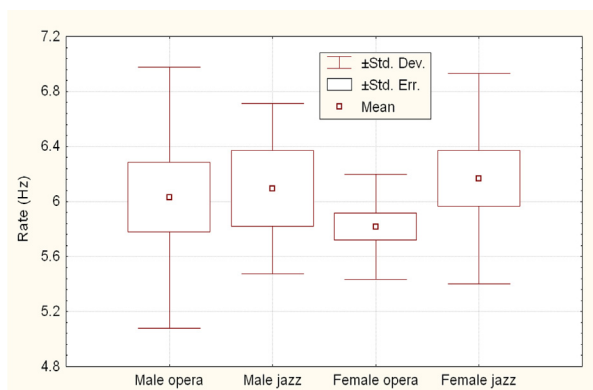


FIGURE 2. Vibrato rate (mean, standard error, and standard deviation) in the four groups of singers.



FIGURE 3. Average tone pitch of the /aɪm/ from /sɪmɔtɑɪm/ (mean, standard error, and standard deviation) in the four groups of singers.

contrast to the other vibrato parameters, the vibrato rate is determined physiologically rather than stylistically. Nevertheless, de Almeida Bezerra et al³⁴ found slightly but significantly lower values (5.1 vs 5.82 Hz) of Vrate in lyric singers compared with sertanejo (Brazilian country western-like) singers.

The significant negative correlation between Vrate and Vext is in line with the results of Ferrante²⁴ in classical female singers. This author also found a small positive correlation between duration and extent: in the present study, there is also a moderate correlation ($r = 0.51$; $P < 0.01$). Indeed, a frequently noticed vibrato pattern is a slight progressive increase of extent when the note is held for a sufficiently long time. Furthermore, in our data, Jvib and Svib appear to be also significantly correlated (negatively) with both duration and extent. However, this effect disappears when the correlation coefficients are adjusted (eg, the negative correlation between Jvib and duration disappears when adjusted for extent, and the negative correlation between Jvib and extent disappears when adjusted for duration).

Classical singers of both genders obviously demonstrate a more important vibrato extent and duration than jazz singers on the same song material. This was expected because vibrato is a traditional feature of Western operatic singing. De Almeida Bezerra et al³⁴ observe mean semitone values of 1.10 in lyric singers versus 0.72 in “sertanejo” singers. However, the new information provided by our results is the significant quantitative

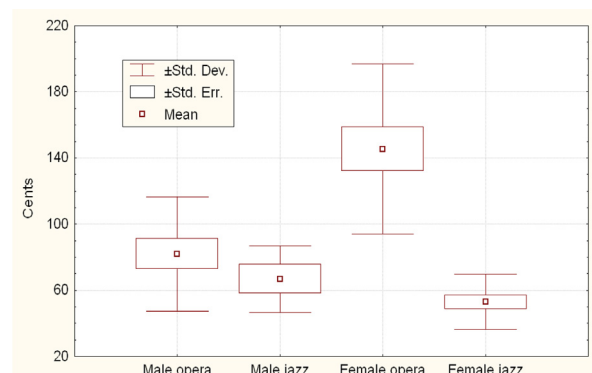


FIGURE 4. Vibrato extent (mean, standard error, and standard deviation) in the four groups of singers.

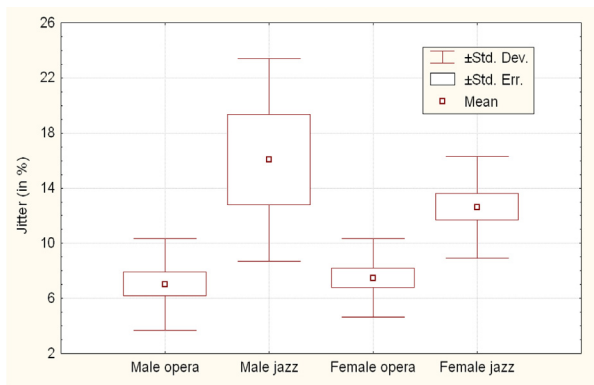


FIGURE 5. Vibrato jitter (mean, standard error, and standard deviation) in the four groups of singers.

difference in vibrato regularity between the two styles. According to Sundberg,²² in skilled (classic) singers, f_0 oscillations are very regular. To quantify this regularity, two original indices were created (Jvib and Svib), by analogy with the perturbation parameters usually computed on voice signals. Vibrato of classic singers of both genders is much more regular than that of jazz singers. In the study of De Almeida Bezerra et al,³⁴ relying on spectrograms, it is reported that “there is no regularity” in the “sertanejo” style, while “there is regularity in terms of frequency oscillation” in the lyric genre.

As for the singer’s formant, the specific presence in male classic singers is again in accordance with expectation based on comparisons of singing styles in the literature: eg, Doskov et al³⁵ already observed a stronger singer’s formant in opera singers than in folk and pop singers. Traditional explanations pertain to the repertoire, the orchestral accompaniment, the working demands, and the environment (amplification),⁴ but the difference in vocal (and musical) education level and in professionalism is an obvious possible bias, that could only be avoided in this study by a selective recruitment. Björkner⁴ specifically investigated professional musical theater singers: musical theater singers showed higher formant frequencies than opera singers presumably because of a somewhat higher larynx position. Unlike opera singers, musical theater singers

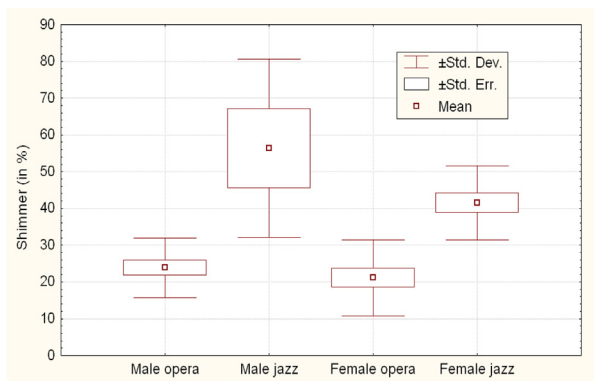


FIGURE 6. Vibrato shimmer (mean, standard error, and standard deviation) in the four groups of singers.

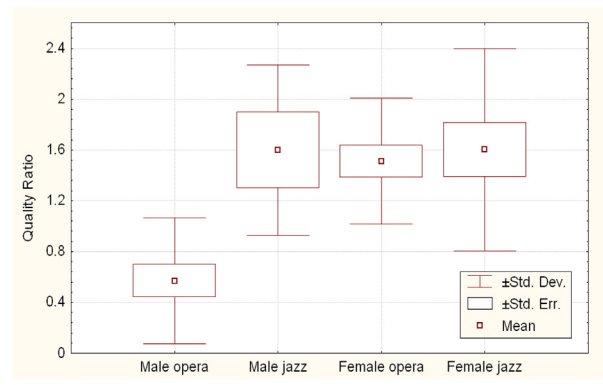


FIGURE 7. Quality ratio (ratio of the acoustic energy in the cluster of the first two formants [Area_{1,2}] and that in the cluster of the third, fourth, and fifth formants [Area_{3,4,5}]). Mean, standard error, and standard deviation) in the four groups of singers.

tended to raise their formant frequencies slightly with rising f_0 . Furthermore, they did not, as the opera singers, tune F3, F4, and F5 closely together into a singer’s formant cluster. As for gender differences, Hollien³⁶ already reported that the singer’s formant has a lower amplitude in female than in male voices. This was confirmed by Omori et al.²⁷

When neither instrumental reference nor accompaniment is provided to the singer, classic singers intone the melody on a significantly higher pitch than jazz singers. This is probably related to the habitual repertoire of the two types of music.

Biofeedback and pedagogic issues with BioVoice

Vibrato develops quasi automatically as voice training successfully proceeds.²² Thus, its objective evaluation may give useful information both as far as quality of the performance and singing style. As for singer’s formant, Omori et al²⁷ also pointed out that singer’s formant quantification may help singing pedagogy refine vocal tract resonance.

The *BioVoice* tool could be very well suited for educational and training purposes. It has a user-friendly interface particularly because it does not require any manual setting to be made by the user. The program provides the singer with easily

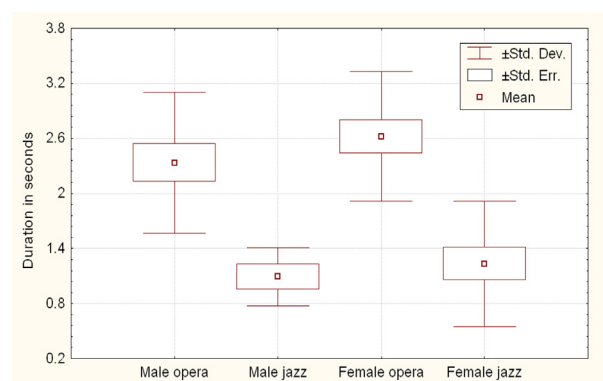


FIGURE 8. Vibrato duration on the /aIm/ from /sAmətaiM/ (mean, standard error, and standard deviation) in the four groups of singers.

readable results pertaining to useful information concerning the acoustic characteristics of his singing voice (as well short-term as long-time averaging), and is optimally convenient for short-term feedback and comparison of different singing trials. Practical training with singing teachers and students is ongoing.

CONCLUSIONS

In this work, a novel tool for the automatic analysis of the singing voice is presented that provides objective and relevant indices pertaining to vibrato and singer's formant. Two groups of skilled singers, with comparable levels of education and professionalism but different singing styles (Western operatic and jazz) were compared on identical vocal material suited for both styles (Gershwin's Summertime).

Some vibrato characteristics (extent, regularity, and duration) very clearly differentiate the Western operatic singing style from the jazz singing style. Extent is significantly larger in classic singers, particularly classic females. Regularity of vibrato, accounted for by vibrato jitter and vibrato shimmer, is significantly better in classic singers. Duration of vibrato is also significantly larger in classic singers. However, vibrato rates do not differ among groups. The cluster of resonances around 3–3.5 KHz is significantly stronger in classic singers, and the presence of the singer's formant is typical for male operatic singers.

All these parameters were computed by means of the *BioVoice* software tool that, unique among similar tools for voice analysis, implements high-resolution analysis techniques particularly suited for high-pitched and time-varying signals as singing voices are. *BioVoice* is provided with a user-friendly interface making the tool fully automatic and easy to use also by nonexpert users, particularly singing teachers and singers, making it very well suited for educational and training purposes.

Acknowledgments

The authors are very grateful to teachers and students of the Luigi Cherubini Academy of Music, Firenze, Italy, for their valuable collaboration.

REFERENCES

1. Welch GF, Howard DM, Himonides E, Brereton J. Real-time feedback in the singing studio: an innovator action-research project using new voice technology. *Music Educ Res*. 2005;7:225–249.
2. Dejonckere PH, Lebacqz J, Bocchi L, Orlandi S, Manfredi C. Automated tracking of quantitative parameters from single line scanning of vocal folds: a case study of the 'messa di voce' exercise. *Logoped Phoniatr Vocol*. 2014;. In press. <http://dx.doi.org/10.3109/14015439.2013.861014>; 2014.
3. Howard D. Acoustics of the trained versus untrained singing voice. *Curr Opin Otolaryngol Head Neck Surg*. 2009;17:155–159.
4. Björkner E. *Why so different? Aspects of Voice Characteristics in operatic and Musical Theatre Singing. Doctoral Thesis*. Stockholm, Sweden: University of Stockholm; 2006:52.
5. Stone RE, Cleveland TF, Sundberg J, Prokop J. Aerodynamic and acoustical measures of speech, operatic and Broadway vocal styles in a professional female singer. *J Voice*. 2003;17:283–297.
6. Bestebreurtje ME, Schutte HK. Resonance strategies for the belting style: results of a single female subject study. *J Voice*. 2000;14:194–204.
7. Thalén M, Sundberg J. Describing different styles of singing: a comparison of a female singer's voice source in "classical", "pop", "jazz" and "blues". *Logoped Phoniatr Vocol*. 2001;26:82–93.
8. Boersma P, Kovacic G. Spectral characteristics of three styles of Croatian folk singing. *J Acoust Soc Am*. 2006;119:1805–1816.
9. Zangger Borch D, Sundberg J. Some phonatory and resonatory characteristics of the rock, pop, soul, and Swedish dance band styles of singing. *J Voice*. 2011;25:532–537.
10. Spivey N. Music theatre singing...Let's talk. Examining the debate on belting. *J Singing*. 2008;64:607–614.
11. Sundberg J. *The Science of Singing Voice*. DeKalb, IL: North Illinois University Press; 1987.
12. Sundberg J. What's so special about singers? *J Voice*. 1990;4:107–119.
13. Sundberg J. Level and center frequency of the singer's formant. *J Voice*. 2001;15:176–186.
14. Titze IR, Story B, Smith M, Long R. A reflex resonance model of vocal vibrato. *J Acoust Soc Am*. 2002;111:2272–2282.
15. Dejonckere PH, Schoentgen J, Giordano A, Fraj S, Bocchi L, Manfredi C. Validity of jitter measures in non-quasi-periodic voices. Part I: perceptual and computer performances in cycle pattern recognition. *Logoped Phoniatr Vocol*. 2011;36:70–77.
16. Manfredi C, Giordano A, Schoentgen J, Fraj S, Bocchi L, Dejonckere PH. Validity of jitter measures in non-quasi-periodic voices. Part II: the effect of noise. *Logoped Phoniatr Vocol*. 2011;36:78–89.
17. Dejonckere PH, Giordano A, Schoentgen J, Fraj S, Bocchi L, Manfredi C. To what degree of voice perturbation are jitter measurements valid? A novel approach with synthesized vowels and visuo-perceptual pattern recognition. *Biomed Signal Process Control*. 2012;7:37–42.
18. Manfredi C, Giordano A, Schoentgen J, Fraj S, Bocchi L, Dejonckere PH. Perturbation measurements in highly irregular voice signals: performances/validity of analysis software tools. *Biomed Signal Process Control*. 2012;7:409–416.
19. Rruqja N, Dejonckere PH, Cantarella G, Schoentgen J, Orlandi S, Barbagallo SD, Manfredi C. Testing software tools with synthesized deviant voices for medicolegal assessment of occupational dysphonia. *Biomed Signal Process Control*. 2014;13:71–78.
20. Fort A, Ismaelli A, Manfredi C, Brusciaglioni P. Parametric and non parametric estimation of speech formants: application to infant cry. *Med Eng Phys*. 1996;18:677–691.
21. Manfredi C, D'Aniello M, Brusciaglioni P, Ismaelli A. A comparative analysis of fundamental frequency estimation methods with application to pathological voices. *Med Eng Phys*. 2000;22:135–147.
22. Dejonckere PH, Hirano M, Sundberg J. *Vibrato*. San Diego, CA: Singular Publishing Group, Inc; 1995.
23. Ekholm E, Papagiannis GC, Chagnon F. Relating objective measurements to expert evaluation of voice quality in western classical singing: critical perceptual parameters. *J Voice*. 1998;12:182–196.
24. Ferrante I. Vibrato rate and extent in soprano voice: a survey on one century of singing. *J Acoust Soc Am*. 2011;130:1683–1688.
25. Anand S, Wingate JM, Smith B, Shrivastav R. Acoustic parameters critical for an appropriate vibrato. *J Voice*. 2012;26:820–826.
26. Sangiorgi T, Manfredi C, Brusciaglioni P. Objective analysis of the singing voice as a training aid. *Logoped Phoniatr Vocol*. 2005;30:136–146.
27. Omori K, Kacker A, Carroll LM, Riley WD, Blaugrund SM. Singing power ratio: quantitative evaluation of singing voice quality. *J Voice*. 1996;10:228–235.
28. Manfredi C. The quality ratio in the singing voice: an electroacoustic study. Proceedings VIII th. International Course of Phoniatrics and Logopedics - The Artistic Voice, Alighieri Theatre, 20–23 October 2011. F. Fussi, Ed., [Ravenna, Italy (in Italian)].
29. Manfredi C, Peretti G. A new insight into post-surgical objective voice quality evaluation. Application to thyroplastic medialisation. *IEEE Trans Biomed Eng*. 2006;53:442–451.

30. Manfredi C, Bocchi L, Orlandi S, Spaccaterra L, Donzelli GP. High-resolution cry analysis in preterm newborn infants. *Med Eng Phys.* 2009;31:528–532.
31. Marple SL. *Digital Spectral Analysis With Applications.* Englewood Cliffs, NJ: Prentice Hall; 1987.
32. Deller JR, Proakis JG, Hansen JHL. *Discrete-Time Processing of Speech Signals.* New York, NY: Macmillan Pub. Co.; 1993.
33. Markel JD, Gray AH. *Linear prediction of speech.* Berlin, DE: Springer-Verlag; 1982.
34. De Almeida Bezerra A, Cukier-Blaj S, Duprat A, Camargo Z, Granato L. The characterization of the vibrato in lyric and sertanejo singing styles: acoustic and perceptual auditory aspects. *J Voice.* 2009;23:666–670.
35. Dostkov D, Ivanov T, Boyanov B. Comparative analysis of singer's high formant in different types of singing voices. *Folia Phoniatr Logop.* 1995;47:291–295.
36. Hollien H. The puzzle of the singer's formant. In: Bless DM, Abbs JH, eds. *Vocal Fold Physiology: contemporary research and clinical issues.* San Diego: College Hill; 1983:368–378.