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, videokymography, 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

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 cappella" 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 /aɪm/ from /sʌmətaɪm/.

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:

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

where $t^{i}_{f_{0max}}$ 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 al[23] 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 ± 2