



Perception and verbalisation of voice quality in western lyrical singing: Contribution of a multidisciplinary research group

Nathalie Henrich Bernardoni, Pascal Bezard, Robert Expert, Maëva Garnier, Christian Guerin, Claire Pillot-Loiseau, Sophie Quattrocchi, Bernard Roubeau, Boris Terk

► To cite this version:

Nathalie Henrich Bernardoni, Pascal Bezard, Robert Expert, Maëva Garnier, Christian Guerin, et al.. Perception and verbalisation of voice quality in western lyrical singing: Contribution of a multidisciplinary research group. CIMO 2007 - 3rd Conference on Interdisciplinary Musicology, Aug 2007, Tallinn, Estonia. hal-00204136

HAL Id: hal-00204136

<https://hal.science/hal-00204136>

Submitted on 12 Jan 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Perception and verbalisation of voice quality in western lyrical singing: Contribution of a multidisciplinary research group

Nathalie Henrich

Department of Speech and Cognition, GIPSA-lab, CNRS/INPG/UJF/Univ. Stendhal, France
nathalie.henrich@gipsa-lab.inpg.fr - <http://icp.inpg.fr/~henrich>

Pascal Bezard, Robert Expert, Maëva Garnier, Christian Guerin, Claire Pillot, Sophie Quattrocchi, Bernard Roubeau, Boris Terk

Department of Lutherie Acoustique Musique, Institut Jean Le Rond d'Alembert, France

In: K. Maimets-Volt, R. Parncutt, M. Marin & J. Ross (Eds.)

Proceedings of the third Conference on Interdisciplinary Musicology (CIM07)
Tallinn, Estonia, 15-19 August 2007, <http://www.gewi.uni-graz.at/cim07/>

Background in music performance. In the field of lyrical singing, an extensive terminology is dedicated to voice quality description. Among the many terms, some are used with consistent meaning by virtually all voice specialists, whereas others, which are more metaphorical or aesthetic, have multiple meanings despite frequent use. The descriptors used by voice specialists deal not only with the perceived sound, but also with its production. Imitation is often used as a complement to verbal description. In parallel, physicians and speech therapists have developed a vocabulary in their practice to describe voice quality in the case of pathological voices. Their listening is primarily oriented to the search of « defects » for diagnosis. Much effort has been expended in the field to retain the most consensual and appropriate terms for perceptual discrimination of different vocal pathologies.

Background in acoustics. Acousticians do not have a specific vocabulary for voice description. They often make use of terms related to timbre. Many studies conducted on the determination of physical criteria for voice-quality description imply a listening focused on voice spectral content and transient phenomena. Because of source-filter theory, acousticians make a distinction between voice-quality aspects related to vocal-fold vibratory movements and those related to vocal-tract configurations.

Aims. Perception of voice quality is not objective, as it depends on the listener's own experiences and expectations. However, a consensus may be found on its verbal description, in a similar way to the technical vocabulary in wine-tasting. Our aim is to elaborate a common terminology for voice-quality description in voice pedagogy, voice therapy and musical acoustics. This paper presents a three-year study conducted by a research group composed of musical acousticians, speech therapists, singers, singing teachers and choir directors, in an attempt to characterise the notion of voice quality and to describe perceived voice quality in the case of lyrical voices with the help of a listening grid based on consensual terms and illustrative sound examples.

Main contribution. Voice quality is a term with multiple meanings related to acoustical aspects (intensity, pitch, spectral content, etc) and aesthetic or subjective aspects. Its definition also depends strongly on the context. When trying to elaborate a common language to describe the perception of voice quality, several fundamental aspects of perception of a sensory process must be considered. These aspects have a direct consequence on its verbal description. Perception tends first to identify an object, so as to compare it to pre-established mental categories. Analytical perception may occur thereafter and the verbal descriptors then depend on the object's initial categorisation. Therefore, the idea of a listening-oriented grid was suggested. It allows the listener to concentrate on a given aspect of voice quality. Such a grid has been established on three main axes: Perception of vocal gesture or vocal technique, perception of sound, and perception of performance. The first axis is mainly presented here. Perception is also differential: An object is always evaluated or described by comparison to a reference in memory. The descriptive terms are defined, and illustrated with reference sound examples.

Implications. The proposed listening-oriented grid facilitates the perceptual and verbal description of voice quality in singing. It may be used as a tool for vocal pedagogy. It will also provide voice professionals with a consensual terminology for expressing singing voice-quality perception.

In the field of speech processing, voice quality is defined by what differs between two vocal productions with identical lexical content. These differences can be prosodic or acoustic, related to variations of rhythm, pitch, intensity, spectral content, etc. They

can be observed at all levels of speech segmentation: At the phoneme level, by a variation of a sustained-sound spectral content; at the word level, in relation to local variations in timing and spectrum; and at the level of the phrase or sentence, in relation to global variations in timing and spectra. How can we describe the perceived quality of a voice? This question introduces a more general problem of sound description: What are the listening modes and the terms to be used to describe what we hear? Voice quality perception is a complex notion, highly subjective and listener dependent. The listening modes and the descriptive terms vary among the specialties: Physicians and speech therapists, actors, singers, and voice teachers, voice coaches, and voice scientists share neither the same listening mode nor a common vocabulary to describe the perceived quality of a voice. In their daily practice, physicians and speech therapists have developed a common language to describe the quality of pathological voices. Their listening is mainly oriented towards the finding of "defects" for diagnosis. In this field, much effort has been expended to keep the more consensual and adequate terms to discriminate perceptually among different voice pathologies. Perceptual evaluation scales of voice quality are commonly used, such as the GRBASI «Grade, Roughness, Breathiness, Aesthenia, Strain» (Issiki & Takeuchi, 1970; Hirano, 1981, 1989), or the RBS scale «Roughness, Breathiness, Hoarseness» (Wendler et al., 1986). In the case of nonpathological voice, and singing voice in particular, such a consensus needs to be found. In the field of lyrical singing, a varied terminology is found (Guerin, 2006; Vennard, 1967; Miller, 1986). Among the many descriptors used to describe voice quality, some are common to specialists whereas others, more metaphorical or aesthetic, are characterised by either multiple or individually-defined meanings despite their frequent use. The terms used by specialists deal not only with the perceived sound, but also with the production mode (Wapnick and Ekholm, 1997; Garnier et al., 2005). Imitations are also often used as a complement to verbal description.

Voice experts from different backgrounds (voice pedagogy, voice therapy, and musical acoustics) lack a common language to describe voice quality (Ekholm et al., 1998). The search for such a common language has led to the setting up of a multidisciplinary research group, in which musical acoustics researchers, voice therapists, singers and singing teachers participate actively. The desire to find a unified terminology for a sensory object, one that would transcend disciplinary boundaries, is not confined to voice. Other disciplinary fields, such as oenology (Guinard and Noble, 1986) or textile engineering (Philippe et al., 2001), have found a consensus for allowing discussions between the different specialists in these fields, mainly for commercial reasons. In this paper, we present the result of a three-year study conducted by this multidisciplinary research group. Within the framework of adult Western lyrical singing, the aims of the research group were

1. to clarify the notion of voice quality and the related listening modes,
2. to explore the voice-quality terminology and to retain the most consensual terms or criteria among the different specialties,
3. to define precisely the terms or criteria retained, and to illustrate them with a collection of sound examples, so as both to establish a consensual base to verbal exchanges between disciplines, and to train new listeners to analytical listening to voices.

Some fundamental aspects of the perception of a sensory process will first be addressed. Free verbalisation about voice-quality will be discussed, together with the terms and listening modes that emerge. On the basis of these observations, two major axes of a listening-oriented grid will be presented, and the consensus will be assessed by a listening test on the grid. In conclusion, the relevance of this approach and the proposed tool for perceptual evaluation of voice quality in lyrical singing will be discussed.

From perception to verbalisation

Several fundamental aspects of the perception of a sensory process have a direct consequence on its verbal description.

A categorical perception

Many studies have shown that perception tends first to identify an object, in order to place it within the listener's existing mental categories (Castellengo, 1986). Analytical perception may occur thereafter (Schaeffer, 1966), and the verbal descriptors are then dependent on the initial object categorisation (Dubois, 1991). For instance, we do not describe a spoken or singing voice in the same way, nor do we use the same description for a lyrical and non-lyrical singing voice. Within this framework, the research group had to choose in first place the vocal-production category on which to investigate voice quality. The choice was made to work on adult Western lyrical singing (in French: *Chant savant occidental de l'adulte*).

An individual perception

Qualitative listener assessments involve interpretation through the filter of a listener's mental representation. Therefore, the past experiences of each listener, his expectations and listening aims (which depend on his areas of expertise) will influence his/her perception and the cues to which s/he would pay attention. It seems necessary to guide perception, to direct the listening to the aspects to be shared. Rapidly, the research group was led to elaborate a listening-oriented grid, for which one of main goals was to guide perception to selected cues related to voice quality.

A differential perception

Human perception is differential: No evaluation or description is absolute. Rather, it involves comparison with another presented object or with a remembered prototype of the object category. As a consequence, verbal description of a sensory object can take advantage of comparatives, and it often involves the object's defects (or its differences from standards) than its qualities (Faure, 2000). This is even stronger in aesthetic fields such as lyrical singing, where

the hedonistic aspect is part of the object and can not be held apart during evaluation. During the elaboration of a methodology for listening and description of voice quality, the differential aspect of human perception has to be taken into account. The elaboration of shared memory reference can benefit from the training with prototypic sound objects. Therefore, the research group recorded a database of reference sound examples, which perceptually illustrates the selected voice-quality criteria.

What words best express the perceived quality of a lyrical voice?

Several glossaries are provided in the literature (e.g. Vennard, 1967; Miller R, 1986; Titze, 1995). They illustrate the variability and redundancy of the terms. Each specialist has his/her own vocabulary to speak about voice, and this vocabulary is only partly shared with the other specialists.



Figure 1. Categorisation of terms used during free verbalisation of voice quality (from Garnier et al., 2005). The terms used for voice quality description (*qualité vocale*) can be divided into four main categories: Sound (*son*, 39% of the terms), technique and physiology (*technique et physiologie*, 22.5%), hedonism (*hédonisme*, 21%), and performance (*interprétation*, 17.5%).

To gather each expert's vocabulary and the way it is organised, a first study was conducted to determine how and with which words we speak about voice quality. Each expert gave an unconstrained verbal description of a set of several commercial and experimental sound examples. From this

exploratory phase came a rich vocabulary, which was organised into categories. We based the categorisation process on a psycholinguistic study of voice-quality verbalisation that was conducted on singing teachers (Garnier et al., 2005). In that study, singing teachers were asked to speak freely about voice quality while listening to sound examples recorded in the laboratory. The linguistic analysis of their discourse has provided a great part of the lexica related to the notion of voice quality, and the corresponding concepts. In applying a psycholinguistic method developed for categorisation and verbal expression of sensory processes (Dubois, 1997, 2000), the lexica used by singing teachers has been separated into four main categories, which are shown in Figure 1.

Toward an oriented listening

These four main categories have inspired the choice of the major axes of the listening-grid. The hedonistic aspects related to pleasure and value judgment have been omitted. Three axes have been proposed by the research group:

- Perception of vocal gesture or vocal technique
- Perception of sound
- Perception of performance

By giving pre-eminence to perception, we wish to make clear that we make here no claim to describe the vocal gesture or its acoustical characteristics, but only the perception that we have of these. Indeed, perception can sometimes be far away from physiological or physical realities of vocal production. The placement terms "forward" and "backward" commonly used in singing are good examples of terms that refer to placement feelings without any demonstrated link to a physiological reality (Vurma and Ross, 2003).

In a first analysis, the third axis concerning perception of performance was discarded, so that we could concentrate on the first two axes. Figures 2 and 3 present the French version of the listening-oriented grid along these two axes. Each axis is divided into poles segmented to give labels on which the

listener can focus his/her attention. The descriptive terms used for each pole were selected during discussions and listening tests done by the research group as the less ambiguous and most representative terms. Synonymous and imprecise terms have been discarded.

Listening oriented along the first axis: "Perception of vocal gesture or vocal technique"

The first pole of this listening axis deals with the **dynamics of inhalation and exhalation**. Two kind of inhalation are distinguished: *Sonorous* and *silent inhalation*. Sonorous inhalation can be *breathy*, when air breathing involves turbulence, or *voiced*, when breathing has both turbulence sound and a glottal vibration. The dynamics of inhalation and exhalation are also characterised by the *breathing pauses*, which can be *frequent* or *infrequent*. The airflow management during the phrase is also of importance.

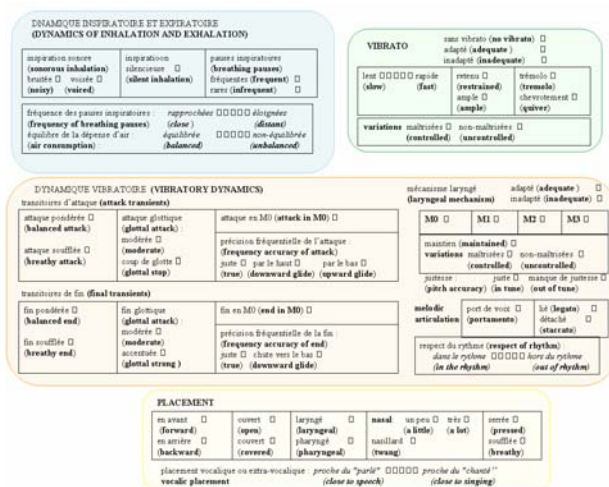


Figure 2. Listening-oriented grid along the first axis "Perception of vocal gesture or vocal technique". The English translation of French terms is given in (bold).

A second pole relates to **vibratory dynamics**. It concerns the *attack and final transients* to which the ear is very sensitive. A sound attack or end can be produced silently, with no audible noise (*balanced*). It can be associated with a breath noise (*breathy*) or with an abrupt vocal-fold contact (*glottal*). When the contact is marked, it could characterise a strong glottal attack (*glottal stop*) or end. An attack at the final quiescent

pitch (*true*) is set apart from an attack with slight *upward glide* (i.e. starting from a lower pitch) or *downward glide*. The attack can be produced in laryngeal mechanismⁱ *MO*, synonymous with vocal fry or pulse registers. The final transients can be described similarly (*true*, *downward glide*, in *MO*), though a final sound with slight upward glide has only rarely been observed. The use of *laryngeal-mechanism* is another aspect of the vibratory dynamics. Sometimes, the same mechanism is used through the whole sentence (*maintained*). When different laryngeal mechanisms are used in the sentence, the listener may perceive a good control of the transition phases between mechanisms (*controlled variations*) or a poor one, for which transitions can be heard (*uncontrolled variations*). *Pitch accuracy*, *melodic articulation*, and *rhythm* are also considered in the vibratory dynamics. The melodic line can be sung *legato*, *staccato*, or with a *portamento*, which is a continuous slide in the melodic variations.

A third pole deals with **vibrato**, its presence or absence and the way it is used in a musical phrase. Vibrato corresponds to a frequency and amplitude modulation of the laryngeal vibration, which induces pitch and loudness modulations in the perceived sound. The modulation frequency can be low (*slow vibrato*) or high (*fast vibrato*). Its amplitude, or frequency extent, can be reduced (*restrained vibrato*) or important (*ample vibrato*). Either a fast laryngeal-frequency modulation (*tremolo*) or a slow and ample one (*quiver*) can be perceived. Both cases may be associated with instabilities. The frequency and amplitude variations of vibrato can be well or poorly *controlled* over the musical phrase.

The listener's assessment of **acoustic source localisation, or 'placement'**, is another important aspect of this perceptual axis. The acoustic source can be perceived as '*forward*' or '*backward*' in the head, in the larynx (*laryngeal*), in the throat (*pharyngeal*), or in the nose (*nasal*). The nasality, for which a contribution of posterior nasal cavities is perceived, is set apart from the *twang* quality, for which anterior nasal cavities seem also to contribute. Voice can be perceived as *breathy*, or giving an impression of laryngeal

tension (*pressed*). Covering is also part of the placement assessment (*open* or *covered* sound). *Vocalic placement* is mentioned, depending on whether the singer's vocal production seems *closer to speech* or *closer to singing*.

Listening oriented along the second axis: "Perception of sound"

The first pole of this listening axis deals with **phonetic aspects** at the segmental and suprasegmental levels. At the *segmental level*, the stress is put on perception of vocalic contrast (*close* or *contrasted vowels*) and vocalic identification (*vowels easily recognisable* or *not*), on perception of consonant control (*short* or *long consonants*) and on consonant pronunciation (*unstressed* or *stressed consonants*). At the *suprasegmental level*, the *respect of phrase and accents* is considered. More generally, sentence *intelligibility* is taken into account.

A second pole concerns the **sound colour**, mainly *timbre* aspects: *high- or low-pitched*, *timbré/détimbré*, *balanced/unbalanced* in respect to energy spectral distribution, *homogeneous/inhomogeneous* on the musical sentence. The *dark* and *light* characters are also considered.

ASPECTS PHONÉTIQUES (PHONETIC ASPECTS)

Segmental:

voyelles (vowels): proches (close) 00000 contrastées (contrasted) 00000
 identifiable (identifiable) 00000 non-identifiable (non-identifiable) 00000
 (recognisable) (unrecognisable)

consonnes (consonants): courtes (short) 00000 longues (long) 00000
 prononciation efficace (effective pronunciation) 00000 prononciation inefficace (ineffective pronunciation) 00000

Suprasegmental:

respect de la phrase (respect of phrase): oui 00000 non 00000 N/A 0
 respect des accents (respect of accents): oui 00000 non 00000 N/A 0
 intelligibilité (intelligibility): oui 00000 non 00000 N/A 0

COLORIS SONORE (SOUND COLOUR)

Timbre: aigu (high-pitched) 00000 grave (low-pitched) 00000
 équilibré (balanced) 00000 déséquilibré (unbalanced) 00000
 homogène (homogeneous) 00000 non-homogène (inhomogeneous) 00000

fonibre (dark) 0 clair (light) 0

Autres qualificatifs (Other terms):

Perception de la fréquence fondamentale (PITCH PERCEPTION)

adéquate 00000 inadéquate (par rapport au tempérament, par rapport à la posture de l'orchestre) 00000
 (adequate) (inadequate)

justesse absolue (perfect pitch) adéquate 00000 inadéquate 00000
 justesse relative (relative pitch) adéquate 00000 inadéquate (justesse de timbre - justesse de hauteur) 00000

Évaluation d'efficacité vocale (LOUDNESS ASSESSMENT)

efficace (efficiency): très efficace (very efficient) 00000 peu efficace (inefficient) 00000
 puissance (power): très puissante (very powerful) 00000 peu puissante (weak) 00000
 fermant du chœur (singing fermant): présent (present) 00000 absent (absent) 00000

Rapport entre la dynamique et la hauteur (possibilité dynamique)

(VOICE RANGE POSSIBILITIES)

grave (low pitches): grande dynamique (great dynamic) 00000 faible dynamique (weak dynamic) 00000 N/A 0
 aigu (high pitches): grande dynamique (great dynamic) 00000 faible dynamique (weak dynamic) 00000 N/A 0

Figure 3. Listening-oriented grid along the second axis "Perception of sound". The English translation of French terms is given in (bold).

A third pole deals with aspects related to **sound intensity and pitch**. The *pitch* can be perceived in *absolute* (perfect pitch) or, more usually, *relative* to the temperament and tuning of the accompaniment. Aspects related to loudness concern efficiency (*very efficient / inefficient*), power (*powerful or weak voice*), and the perceived *presence or absence of a*

singing formant. The *voice range possibilities*, and in particular the relation between vocal intensity and pitch, are considered at both the low and high-pitched part of the singer's tessitura.

Exploration of the consensus

The listening grid presented in the previous part has been established to allow an oriented listening of voice quality, in a view to share a consensual description among specialists of different fields. The relevance of this grid and the description consensual properties have been tested on a group of 18 listeners (mean age 38 (+/- 11) years old), including eight professional musicians, eight amateurs, and two non-musicians. 13 of these listeners were familiar with lyrical technique, either by a regular practice as a singer, or by frequent listening of this music style. All except one were familiar with voice quality verbalisation, ten listeners describing voice quality occasionally and seven very often.

Description of the perceptual test

The perceptual test took place in a meeting room with small groups of 5 to 8 listeners. The test was divided into three parts:

1. listening and free verbalisation of sound examples
2. presentation of the research group work on voice quality, and description of two main axes of the listening grid. The verbal presentation of the first axis (perception of vocal gesture or vocal technique) was complemented by perceptual illustration using prototypic sound examples recorded by the group.
3. replay of the previous sound examples. The subjects were asked to mark the parameters that seemed relevant for them in the grid. When a quality seemed to occur occasionally in the musical sentence, this particularity could be mentioned in the grid.

At the end of the test, the subjects filled a form about their musical skills and knowledge, their feelings about the test and the relevance of using such a grid.

Six sound examples were played, corresponding to two professional male singers performing a reference example and two variations. The first singer (B1) was singing a French sentence composed for the purpose of a previous study (Sotiropoulos, 2004). He was recorded prior to this test, and the variations were chosen according to listening-grid elements (first axis). The second singer (B2) was singing a Latin sentence, the first beats of Gounod's Ave Maria, recorded during a previous study on voice quality (Henrich, 2001).

Two listening modes were tested. First, the reference example was presented alone. Secondly, the variations were presented in comparison with the reference example (listening of sound examples in pairs). The example to be described was then repeated as many times as necessary.

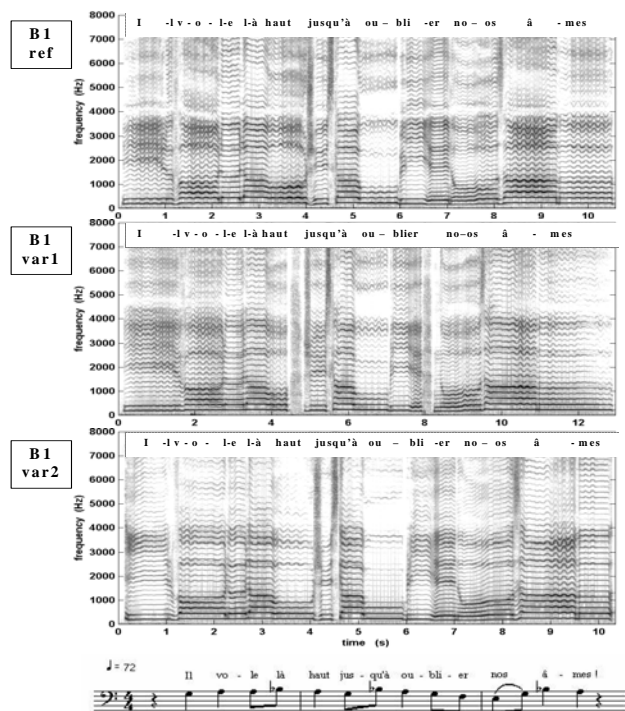


Figure 4. Musical sentence sung by baritone B1 with three different voice qualities. B1ref is the reference example, B1 var1 and var2 are two variations.

Subjects' feelings about the grid

In the form, the following question was asked: *Do you think that, after this test, such a listening-oriented grid could help you in the perception and verbalisation of voice quality ?*

yes, a lot; yes, a little; no, not much; no, not at all.

83% of the subjects considered that this grid could be helpful: 39% chose *yes, a lot*, and 44% *yes, a few*. Two subjects (11%) had no opinion, and one subject (6%) considered that the grid would not be of much help to him.

The next question was about the consensus: *Do you think, after this test, that such a listening-oriented grid could provide a more consensual dialogue on voice quality between the different voice specialists? yes, certainly; yes, possibly; no.*

All the subjects considered that such a grid could lead to a more consensual dialogue: Six subjects (33%) chose *yes, possibly* and 12 subjects (67%) *yes, certainly*.

Perception of salient characteristics

We wished to determine whether listeners would perceive salient voice qualities selected from the grid and performed by singer B1. In complement to his normal singing phonation, which was recorded as the reference example (see Figure 4, B1 ref), the singer performed the following two variations:

- variation 1 (Figure 4, B1 var1): noisy inhalations, frequent breath takes and noticeably unbalanced air supply, breathy attack and final transients, voice production in M1, staccato melodic articulation and out of rhythm, breathy placement.
- Variation 2 (Figure 4, B1 var2): silent inhalations with infrequent breath pauses, without vibrato, with glottal stops and strong glottal final transients, voice production in M1, portamento melodic articulation and in the rhythm, laryngeal and pressed placement.

These characteristics were perceived and verbalised through the grid by a majority of subjects. Results are shown in Figure 5.

After listening to sound example 1, 89% of the listeners considered that breath takes were sonorous and noisy, and that breath pauses were frequent (an opinion shared by 83% of the listeners). The imbalance of breath supply was noticed (89%). The attack transients were more perceived as glottal

(44%) than breathy (33%). Nevertheless, the majority of listeners perceived breathy final transients (55%). The use of laryngeal mechanism M1 was well perceived (78%).

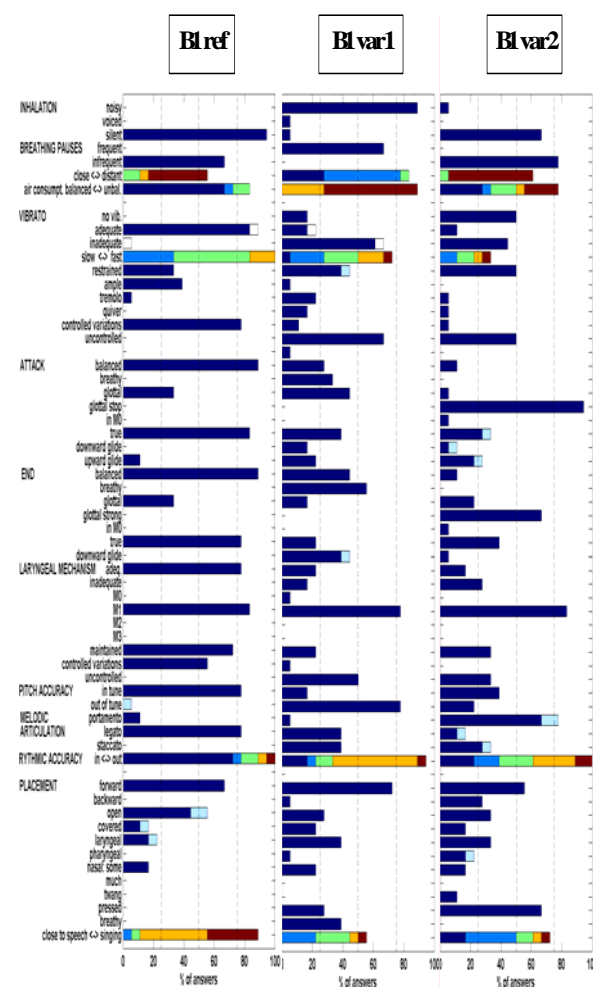


Figure 5. Results of the description given by 18 listeners of singer B1's three examples, along the first axis of perception of vocal gesture or vocal technique. For each listening-grid parameter, the dark blue horizontal bars present the percentage of listeners who have indicated it. The complements in light blue correspond to the cases for which the quality was only occasionally perceived. The white complements correspond to the cases where the parameter is judged to be inapplicable (n/a). Bars with multiple colours present the 5-points bipolar scale answers (from left to right: 1-dark blue, 2-light blue, 3-green, 4-orange, and 5-dark red).

The characteristics of melodic articulation were not consensual: 39% perceived a legato melody and 39% a staccato one. 67% of the listeners perceived that the singer seemed relatively out of rhythm. The breathy placement was not saliently perceived, as it was only mentioned by 39%.

After listening to sound example 2, 67% of the listeners perceived silent inhalations and rare breath pauses (78%). Only half of the listeners (50%) mentioned the lack of vibrato. This could be explained by the fact that the singer did in fact sing with vibrato at the end of his phrase. This vibrato, even expressed briefly, may have been perceived and so taken into account by listeners. The listeners who did not mark the lack of vibrato have all mentioned an inadequate vibrato (44%), and/or a restrained vibrato (50%). The glottal stops were unanimously perceived (94%). The strong glottal ends were also well perceived (67%). The use of laryngeal mechanism M1 was well perceived (83%), together with the portamento (78%). The rhythm was not perceived in a consensual manner. Most of the listeners did not detect a laryngeal placement (33%). They perceived a pressed voice (67%), placed forward (56%).

Discussion on the consensus for description of voice quality

The listeners' answers are illustrated in Figure 5 for singer B1 and in Figure 6 for singer B2. These figures present, for each parameter of the first-axis grid, the number of listeners (expressed as a percentage of the total number of listeners) who marked this parameter. We consider that a consensus on description is noticeable among the listeners when a majority of them (more than 50%) have marked the same box. In the 5-point bipolar scales case, the proportion of each choice is given. In this case, the colour variations and their corresponding proportions indicate the degree of agreement among the listeners.

By analysing the listeners' answers to the six listening, an inter-listener agreement was found on some parameters, whereas others showed a disagreement.

The **dynamics of inhalation and exhalation** is described in a consensual manner by the listeners. Within this pole, the listeners did not always agree on the perception of air supply balance, e.g. for examples B1 var2 and B2 var2.

On the **vibratory dynamics** pole, listeners often agreed on the description of attack and final transients' perception, and this

agreement was stronger when the type of transient was not varied along the sentence (examples B1ref and B2ref). The perception of laryngeal mechanism was consensual for these examples. On the contrary, judgment of adequate/inadequate character was not consensual. Interestingly, the listeners who considered that laryngeal mechanism was maintained throughout the phrase also sometimes noted variations. It seems therefore that the notion of controlled or uncontrolled laryngeal-mechanism variations has not been understood by these listeners. In the selected examples, the laryngeal mechanism was not varied, as the singers were always singing in M1. The listeners generally shared the perception of in or out of tune, except for examples B1 var2 and B2 var2. This is also the case for melodic articulation, which was described in a consensual way, except for example B1 var1. Rhythm is a parameter for which description differs considerably among listeners. Almost all listeners noted it at each listening (between 78% and 100%). Yet, in many cases, the rhythmic adequacy was perceived very differently (examples B1 var2, B2 var1 and var2). The absence of a musical accompaniment may explain this disagreement.

Description of **vibrato** was not consensual among the listeners. Agreement was observed on vibrato adequacy or inadequacy (except for the previously-mentioned example B1 var2, which was either perceived with no vibrato or with an inadequate one). However, the vibrato frequency was always perceived very differently by the listeners. The ample or restrained characters were also barely consensual. The listeners only shared a common perception of the control of vibrato variations.

As with vibrato, the listeners did not agree clearly on '**placement**' perception. A good agreement was found in the case of a salient quality (e.g. forward or pressed in examples B1 ref, B1 var1, var2 and B2 var2). However, much often, the listeners noted different qualities, sometimes opposite ones (such as for open / covered qualities). This difficulty to find a consensus on placement description was already observed during the discussions and previous inner verbalisation tests

conducted by the research group. One reason may be that the notion of placement has no clear meaning. This observation calls for further research on this pole.

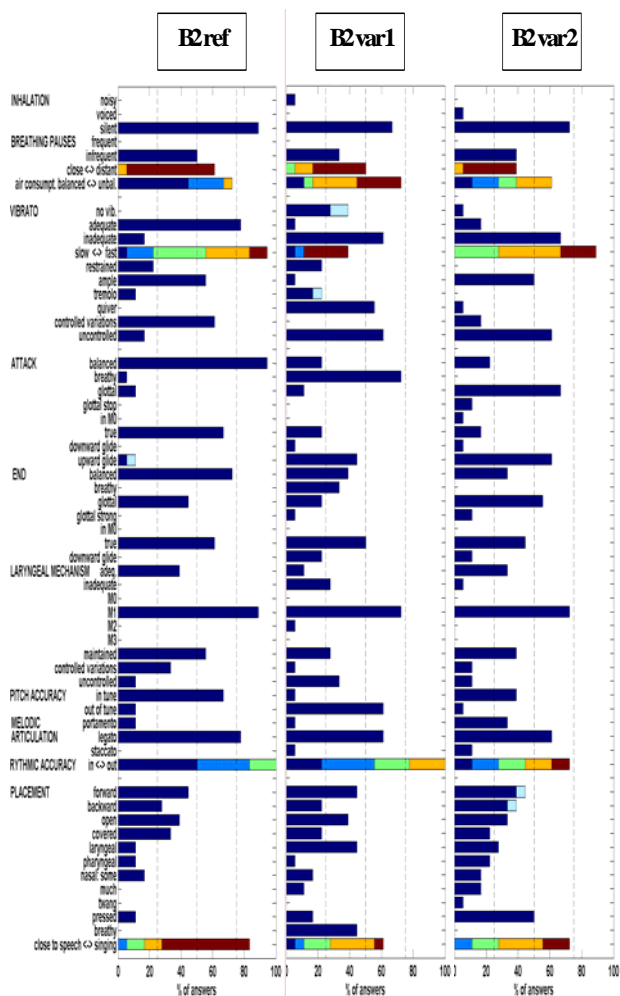


Figure 6. Results of the description given by 18 listeners of singer B2's three examples, along the first axis on perception of vocal gesture or vocal technique. The legend of Figure 5 gives more details.

Conclusion

A three-year study, conducted by a multidisciplinary research group working on perception and verbalisation of voice quality in western lyric singing, has established a listening-oriented grid to describe voice-quality perception. It has two axes: Perception of vocal gesture or vocal technique, and perception of sound. The relevance of this grid has been tested on 18 listeners with different disciplinary backgrounds. The relevance was unanimously

approved by the listeners. The analysis of their answers regarding the first axis shows that a good consensus has been obtained on perception of respiratory and vibratory dynamics. However, no clear inter-listener agreement has been observed concerning vibrato and vocal placement perceptions. Unshared references in memory, different listening modes or the vagueness of the definition of vocal placement could explain this result, which calls for further research.

In addition to its contribution to the search for a consensual terminology, the listening grid constitutes an interesting pedagogical tool. On the one hand, it is a training tool for learning to categorise different voice-quality parameters. On the other, it guides the identification and perceptual evaluation of these parameters during voice listening. Finally, it may constitute a very useful discussion aid for experts from the different voice disciplines.

Acknowledgments. We are grateful to the listeners who volunteered to participate in the listening test. We acknowledge the Laboratoire d'Acoustique Musicale for regularly hosting the research group meetings. We would like to express our thanks and appreciation to Joe Wolfe for his kind help with the English translation.

References

- Castellengo, M. (1986). Les sources acoustiques, in D Mercier (Ed.). Le livre des techniques du son, tome 1, Paris, Editions Fréquences, pp. 45-80.
- Dubois, D. (1991). Prototypes et typicalité. In Sémantique et cognition.
- Dubois, D. (1997). Catégorisation et cognition: De la perception au discours. Collectif, Kimé Ed., Paris.
- Dubois, D. (2000). Categories as acts of meaning: The case in olfaction and audition. Cognitive Science Quarterly, vol.1, 35-68.
- Ekholm E., Papagiannis G.C. and Chagnon F.P. (1998). *Relating objective measurements to expert evaluation of voice quality in western classical singing: Critical perceptual parameters*. Journal of Voice 12(2): 182-196.

- Faure, A. (2000). Des sons aux mots: Comment parle-t-on du timbre musical ? Thèse de doctorat. EHESS.
- Garnier, M., Henrich, N., Dubois, D., Castellengo, M., Poitevineau, J., and Sotiropoulos, D. (2005). Etude de la qualité vocale dans le chant lyrique. *Scolia* 20, pp.151-169.
- Guerin, C. (2006) Glossaire du chant lyrique. Website: <http://chanteur.net/glossair.htm>
- Guinard, J.X., Noble, A.C. (1986). Proposition d'une terminologie pour une description analytique de l'arôme des vins. *Sci. Aliment* 6, pp.657-662.
- Henrich N. (2001). Etude de la source glottique en voix parlée et chantée: Modélisation et estimation, mesures acoustiques et électroglottographiques, perception. PhD thesis, Université Paris 6, 2001.
- Henrich N. (2006). Mirroring the voice from Garcia to the present day: Some insights into singing voice registers, *Logopedics Phoniatrics Vocology*, vol. 31, pp. 3-14.
- Issiki N., Takeuchi (1970). Factor analysis of hoarseness, *Stud. Phonol.*, 5, 37-44.
- Hirano M. (1981). Clinical examination of voice, Springer Verlag, New York.
- Hirano M. (1989). Objective evaluation of human voice: Clinical aspects. *Folia Phoniatr.*, 41, 89-144.
- Miller, R. (1986). The structure of singing: System and art in vocal technique. Ed Schirmer/Thomson Learning.
- Philippe, F., Schacher, L., Adolphe, D., Dacremont, C. (2001) Développement d'une méthodologie d'analyse sensorielle tactile des textiles. *International Journal of Clothing* 15, pp. 268-. 275.
- Roubeau B. (1993) Mécanismes vibratoires laryngés et contrôle neuro-musculaire de la fréquence fondamentale, PhD thesis, Université Paris-Orsay, Orsay.
- Schaeffer, P. (1966). *Traité des objets musicaux*. Ed du seuil, Paris, pp. 712.
- Sotiropoulos, D. (2004). Analyse acoustique et catégorisation d'un ensemble de qualités vocales pertinent pour la description de voix lyriques masculines. Mémoire DEA ATIAM.
- Titze, I. R. (1995). Definitions and nomenclature related to voice quality. in *Vocal Fold Physiology*, edited by Fujimura O. and Hirano M. (Singular, San Diego), pp. 335-342.
- Vennard, W. (1967). *Singing, the mechanism and the technic*. Ed. Carl Fisher.
- Vurma A, Ross J. (2003). *The perception of 'forward' and 'backward placement' of the singing voice*. *Logoped Phoniatr Vocol*. Vol. 28(1): 19-28.
- Wapnick J. and Ekholm E. (1997). *Expert consensus in solo voice performance evaluation*. *Journal of Voice* 11(4): 429-436.
- Wendler J., Rauhut A., Krüger H. (1986). Classification of voice qualities. *J. Phonet.*, 14, 483-488.

ⁱ We refer the reader to Roubeau (1993) or Henrich (2006) for a definition of laryngeal mechanism notion.