

Christian Herbst: Voice timbre in singing

electronic mail: [herbst<at>ccrma<dot>stanford<dot>edu](mailto:herbst@ccrma.stanford.edu)

web: www.christian-herbst.org

Laboratory of Bio-Acoustics, Dept. of Cognitive Biology, University of Vienna,
Althanstraße 14, 1090 Wien, Austria

Department of Biophysics, Faculty of Science, Palacký University Olomouc, 17. listopadu
12, 771 46 Olomouc, Czech Republic

Introduction

From the performer's point of view, the singing voice can be subjectively described with several perceptual attributes, e.g. *pitch*, *loudness*, *brilliance*, *noise components*, *vowel color* and *focus*. Most of these terms are associated with *voice timbre*. This presentation attempts to explain the physical and physiological foundation of the singing voice, relates it to the produced acoustic output (and thus to the aforementioned perceptual qualities), and briefly describes with which structures of the voice apparatus they can be manipulated.

Timbre

The American National Standards Institute defines timbre as "*that attribute of auditory sensation in terms of which a listener can judge two sounds similarly presented and having the same loudness and pitch as dissimilar*". Timbre "*depends primarily on the spectrum of the stimulus, but it also depends upon the waveform, the sound pressure, the frequency location of the spectrum, and the temporal characteristics*"[1]

Pressure	+
Frequency	++
Spectrum	+++
Duration	+
Envelope ^(a)	++

Dependence of the subjective quantity "timbre" on physical parameters. [2]. The number of plus signs is an indicator the strength of this dependence.

^(a)Envelope: attack, release, amplitude variations

A sung sound is (in most cases) a so-called harmonic series. It consists of a series of individual harmonics (also called “partials” or “partial tones”), each having a frequency of an integral multiple of the fundamental frequency (e.g. 240 Hz, 480 Hz, 720 Hz, 960 Hz, 1200 Hz, ...). The perceived pitch of such a harmonic series is usually the frequency of the lowest harmonic, i.e. the greatest common denominator. The timbre of a sung sound is largely determined by the relative strength of each harmonic in the series.

The Source-Filter theory of voice production

“The sound source is most often generated by the vibrating vocal folds that convert the steady (DC) airflow from the lower respiratory system into a periodic train of flow pulses. These pulses, referred to as the glottal flow, are then acoustically filtered by the vocal tract resonances. This process effectively redistributes the amplitudes of the frequency components of the source signal, resulting in a vowel sound.”[3]

According to the source-filter-theory of sound production, the sound quality (i.e. the timbral characteristics of the phonation) can be manipulated independently at the laryngeal level (sound source) and the vocal tract (filter). Even though more recent research showed that this is an over-simplification of matter, the source-filter theory is still useful in explaining the basic concepts of voice production. Thus, the remainder of this discussion focuses first on the (quality of) vocal fold vibration, and then on the vocal tract.

1 Voice source characteristics

1.1 Fundamental Frequency

Fundamental frequency is determined by the frequency of vocal fold vibration. It is mainly controlled by the action of the *cricothyroid muscle* (elevating the cricoid arch and depressing the thyroid lamina, and thus actively lengthening the vocal folds and increasing their tension), and partly by subglottal pressure (i.e. lung pressure) and choice of vocal register (see below). Vibrato is a regular modulation of the fundamental frequency, whereas jitter is an irregular (and potentially pathologic) form thereof. Vocal fold vibration can exhibit certain degrees of irregularity, manifesting itself as biphonation, subharmonics, or chaos.

Fundamental frequency is not to be confused with *pitch*, which is a perceptual quality.

1.2 Amplitude

Loudness, i.e. the subjective perceptual quality that is related to sound pressure level, is mainly related to subglottal pressure as created by the pulmonary apparatus (which influences the amplitude of vocal fold vibration). Glottal efficiency is defined as the ratio of radiated acoustic power to aerodynamic power. The abruptness of vocal fold closure (see also next paragraph) also contributes to sound pressure level.

1.3 Spectral slope

The presence of high-frequency components (partials, harmonics) in the sound source is largely related to the abruptness of vocal folds closure (but can also be greatly influenced by adjustments of vocal tract shape). This abruptness of vocal folds closure is mainly controlled by combinations of the degree of vocal folds adduction (breathy-flow-pressed) and subglottal pressure. Experienced singers can independently control membranous and cartilaginous adduction to a certain degree, thus creating different sound timbres in the voice source. The membranous medialization of the vocal folds is controlled by the *vocalis (thyroarytenoid) muscle*, and it determines the choice of voice register at the laryngeal level.

The audibility of a singing voice in an environment that contains other sound sources (such as an orchestra) is largely determined by the amount of energy between 1000 and 5000 Hertz. This is due to the fact that the human ear is most sensitive in this region.

The presence of strong high-frequency components in a singing voice is sometimes subjectively described by the terms *brilliance* or *focus* (the voice “*carries*”). However, these terms cannot be attributed to the laryngeal level alone, since they might also be used to describe the effect of optimal vocal tract adjustments, i.e. in the singers’ formant or formant tuning (see below).

1.4 Noise components

Noise components can be added to the voice in both (a) the sound source; and (b) the vocal tract, e.g. fricatives (like the consonants [f], [s], [z] or [ʃ]) or plosives (e.g. [p], [k], [g], etc.). In the sound source, noise components are usually generated by turbulent airflow in case of incomplete glottal closure (in the closed phase of vocal fold vibration). Such an incomplete closure can be caused by

- (a) muscular adjustments of the vocal folds, mainly affected by the adductory intrinsic laryngeal muscles (lateral cricoarytenoid, posterior cricoarytenoid, and interarytenoid muscles). These adjustments are either voluntary (as a choice of singing style); or involuntary (which can constitute a pathologic voice condition);
- (b) pathologic changes of the vocal fold structure (such as edema, nodules, cysts, bowing, etc.).

1.5 Registers

“By the word register we mean a series of consecutive and homogeneous tones going from low to high, produced by the same mechanical principle, and whose nature differs essentially from another series of tones equally consecutive and homogeneous produced by another mechanical principle. All the tones belonging to the same register are consequently of the same nature, whatever may be the modifications of timbre or of the force to which one subjects them.”

More than hundred different terms for classifying registers have been used in the literature, thus confusing the issue. The most commonly recognized registers are: (a) vocal fry (or “pulse register”); (b) chest (or modal) register; (c) falsetto register (sometimes also called “head”); (d) and whistle register. Chest and falsetto are considered to be the two main registers for singing. It is still debated whether the “head

voice” or “voix mixte” constitute a register on its own right, and how it is related to glottal and vocal tract configurations in singing.

Registration is mainly determined at the laryngeal level by the configuration of the vocal folds. In chest, the contraction of the vocalis (thyroarytenoid) muscle introduces a vertical phase difference into the movement of the vocal folds, thus effectively shortening the closed phase. This adjustment is likely to enhance the abruptness of vocal folds closure, thus increasing the output of high-frequency partials (as compared to falsetto register). Chest register phonation is generally produced at a lower fundamental frequency as compared to falsetto register, although the ranges of the two registers overlap to a certain degree, varying from singer to singer.

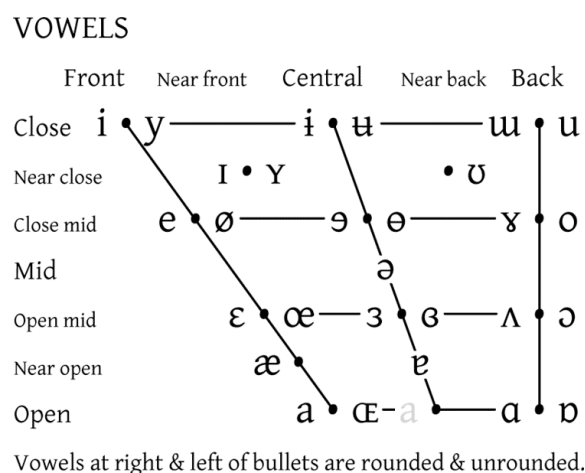
Vocal registration can also be influenced more or less subtly by adjustments of the vocal tract.

2 Vocal tract

2.1 Vowel quality

The vocal tract, i.e. the airways space from the vocal folds to the lips (and the nose), acts as an acoustic filter: based on vocal tract geometry, certain frequency regions allow more acoustic energy to pass (and in effect being radiated from the mouth) than others. These regions are generally called formants. In a uniform tube of 17 cm length (representing a very simple model of a male vocal tract), these formants would be found at ca. 500 Hz, 1500 Hz, 2500 Hz, 3500 Hz, 4500 Hz, etc. The vocal tract shape, and thus the center frequency of the formants, can be influenced by the articulators: mainly the tongue, but also the lips, jaw opening, and the configurations of pharynx and the airway space above the vocal folds in the larynx. The lowest two or three formants are responsible for the perceived vowel quality.

Vowels can be qualified as front vs. back; and close vs. open vowels, largely depending on tongue position and jaw opening, respectively.



IPA (International Phonetic Alphabet) Vowel Chart[4]

2.2 Singers' formant cluster

The singers' formant cluster (also called "singers' formant") is an energy concentration between ca 2500 and 3400 Hz, where the human ear is particularly sensitive. It is caused by clustering of formants 3, 4 and 5, thus enhancing energy output in this frequency region. It is controlled by geometry of the airways space directly above the vocal folds (up to the rim of the epiglottis), and by the vertical larynx position.

The presence of a singers' formant can cause voices to sound "*brilliant*" or "*focused*". Particularly lower (male) voices may depend on the singers' formant in order to be heard over an orchestra or a choir when singing a solo.

2.3 Formant tuning

Acoustic energy is created in the sound source (at the laryngeal level) as a series of harmonics. These harmonics, being integral multiples of the fundamental frequency (of vocal fold vibration) are equally spaced in frequency. Usually, no acoustic energy is present between the harmonics.

In order to be heard, the acoustic energy created in the sound source has to be propagated through the vocal tract. Dependent on its geometrical shape, the vocal tract might let some harmonics pass easily (at frequency bands where the formants are found), whereas other harmonics might be heavily attenuated/damped. Imagine now a constellation where none of the harmonics of the voice source would coincide with any of the vocal tract formants. This would result in a very weak voice, regardless of the singer's effort level (Like screaming into a pillow). On the other hand, if the strongest harmonics "hit" a formant, the energy transfer through the vocal tract would be very effective, resulting in a louder voice whilst potentially reducing the effort level.

Formant tuning is the skill of finding combinations of fundamental frequency and formant frequencies in which energy passing through the vocal tract is optimized. In classical singing, pitch (and thus fundamental frequency) is determined by the composer, whereas vowel quality can be modified to a certain degree without spoiling the lyrics. Thus, in classical singing, formant tuning is achieved by minute adjustments of vowel quality. An experienced singer can produce slightly varying vowel colors for each combination of musical pitch and vowel. These optimized sound production conditions can create standing waves in the vocal tract, which can be perceived by the singer as vibratory sensations in the skull. Terms like "*head resonance*", "*focus*", "*placement*" / "*Stimmsitz*" etc. are used to describe these phenomena on a subjective level.

Resources on the web

Acoustics Group, University of New South Wales: Voice Acoustics: an introduction

<http://www.phys.unsw.edu.au/jw/voice.html>

National Center for Voice and Speech: How Humans Speak, Sing, Squeak and Squeal

<http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/index.html>

W. R. Zemlin Memorial Website (larynx anatomy)

<http://zemlin.shs.uiuc.edu/>

The Bureau of Glottal Affairs

<http://www.surgery.medsch.ucla.edu/glottalaffairs/index.htm>

Praat: doing phonetics by computer

<http://www.fon.hum.uva.nl/praat/>

Madde. An additive, real-time, singing synthesiser

<http://www.speech.kth.se/music/downloads/smptool/madde.exe>

SFS/RTGRAM - Windows Tool for Real-time Speech Spectrogram Display

<http://www.phon.ucl.ac.uk/resource/sfs/rtgram/>

VoceVista ("... developed primarily for singing teachers to analyze the singing voice...")

<http://www.vocevista.com/>

Suggested reading

- Chapman, J. (2006). Singing and teaching singing. A holistic approach to classical voice. San Diego, Oxford, Brisbane: Plural Publishing.
- Doscher, B. M. (1994). The functional unity of the singing voice (2nd Edition). Metuchen, NJ and London: The Scarecrow Press.
- Miller, D. G. (2008). Resonance in Singing. Princeton, NJ: Inside View Press.
- Rossing, T. (1990). The Science of Sound: Addison-Wesley Publishing Company.
- Seidner, W., & Wendler, J. (2004). Die Sangerstimme (4th Edition). Berlin: Henschel Verlag.
- Stark, J. A. (1999). Bel Canto: a history of vocal pedagogy. Toronto/Buffalo/London: University of Toronto Press.
- Sundberg, J. (1987). The science of the singing voice. Dekalb: Northern Ill. University Press.
- Titze, I. (1994). Principles of voice production. Englewood Cliffs, NJ: Prentice-Hall.

References

- [1] ANSI, "USA Standard Acoustical Terminology (Including Mechanical Shock and Vibration)," *Technical Report* vol. S1.1-1960 (R1976), 1960.
- [2] T. Rossing, *The Science of Sound*: Addison-Wesley Publishing Company, 1990.
- [3] B. Story, "An overview of the physiology, physics and modeling of the sound source for vowels," *Acoust. Sci. & Tech.*, vol. 23, 2002.
- [4] Wikipedia. (2010), *IPA (International Phonetic Alphabet) Vowel Chart*. Available: http://en.wikipedia.org/wiki/International_Phonetic_Alphabet