

A Voice Quality Chart inspired by the Laryngeal Articulator Model

Parham Mokhtari
Toyama Prefectural University, Japan

Keywords — voice quality, chart, laryngeal articulator model

I. INTRODUCTION

Voice quality (VQ) is defined as a long-term, quasi-permanent characteristic of a person's voice, determined by the individual's anatomical constraints and habitual idiosyncracies [1]. It can also refer to the quality used, whether intentionally or not, throughout any given portion of an utterance, regardless of its duration. In any case, VQ is generally regarded as a suprasegmental characteristic of speech, potentially extending over multiple phonetic segments.

The physiological *settings* involved in producing the various VQs are traditionally classified as either *phonatory* (glottal configurations, modes of oscillation) or *articulatory* (supraglottal vocal tract). Meanwhile, much research has pointed to the inadequacy of the source-filter model of speech production in its simplest form where the source is confined to only the vocal folds. From various viewpoints, it appears that voice production models had better consider the entire larynx, including not only the vocal folds but also the various hypopharyngeal structures and cavities such as the ventricles, piriform fossae, ventricular (false) folds, vestibule, and aryepiglottic folds with their sphinctering mechanism. Accordingly, Laver's [1] VQ descriptive framework based on the glottal-vs-supraglottal distinction, was revised and extended in Esling *et al.*'s [2] Laryngeal Articulator Model (LAM), where laryngeal VQ is determined by the configuration of the *entire larynx* and involves the aryepiglottal sphincter as an active articulator. In line with this model, the importance of the hypopharynx (including the supraglottic laryngeal cavity and piriform fossae) as an influence on VQ and speaker individuality, separate from the rest of the vocal tract proper, has also been shown by Honda *et al.* [3].

In parallel with progress in phonetics and speech science, recent technical advances have enabled speech synthesis with the characteristics of a target person's voice, given a sufficiently large recorded dataset of that voice. However, this is achieved without necessarily knowing or learning anything new about the underlying acoustics or physiological mechanisms. Moreover, intentionally and flexibly adjusting a synthesizer's VQ without recourse to recordings, and based on auditory or articulatory descriptors, remains a challenge. Furthermore, regarding VQ research in general, perhaps to circumvent the multiplicity of categories and ambiguity in the meaning of impressionistic labels, researchers often restrict the scope of their investigation to only one or a small number of relatively unambiguous VQs. One example is the preponderance of papers on breathy vs pressed/tense voices; while such research has yielded significant advances, it also runs the risk of portraying VQ as merely one-dimensional, which it certainly is not.

The aim of this contribution is to reassert the multi-dimensional nature of non-pathological laryngeal VQ, by organizing the basic categories in the form of a Voice Quality Chart (Fig.1) inspired by the LAM. While it is inherently limited to two dimensions, multidimensionality is implied by emphasizing relations between categories, motivated by underlying physiological mechanisms.

II. KEY LARYNGEAL MECHANISMS & THE CHART'S AXES

The structure and function of the larynx are well described in numerous works [4, 5]. Here we provide a simplified summary of only the key mechanisms involved in producing the range of VQs covered in the chart.

The chart as a whole is divided into two groups of VQ depending on **voicing**, which refers to whether (voiced) or not (voiceless) any part of the true vocal folds is oscillating. The oscillation mechanism itself has been the subject of decades of research; an updated contemporary account is given by Švec *et al.* [6]. The most basic (neutral) VQ in each group is *modal voice* and *breath*, respectively.

The chart's horizontal axis represents **laryngealization**, which refers to any degree of narrowing/constriction of the epilaryngeal tube, mainly by the sphinctering mechanism of its upper boundary at the aryepiglottic folds. VQs in the left column (aligned with *breath*) have minimal/no sphinctering, with either a neutral or wide-open epilaryngeal tube. VQs in the second column (aligned with *whisper*) are produced with a certain amount of laryngealization. An even greater degree of laryngealization can lead to ventricular adduction, epilaryngeal vertical compression, and aryepiglottal folds oscillation (trilling), generally resulting in perceived harshness (third column). Extreme laryngealization of the entire larynx structure is involved in producing *ventricular voice* at the far right.

The voiced group of VQs is further subdivided vertically into basic categories according to the following three principal mechanisms. **Fundamental frequency (f_0) raising/lowering** refers to control of the oscillation rate of the vocal folds, corresponding to the auditory percept of voice pitch. A basic distinction is between *modal voice* and *false alto*, two VQs that are mutually exclusive owing to their different modes of oscillation, and whose f_0 ranges typically overlap. At the opposite extreme, very low f_0 and at least some laryngealization is required to produce *creaky voice* (or its sporadic/irregular f_0 equivalent *creak*). Another distinction appears in the third column, where at least three types of harsh voice are distinguished according to f_0 range. **Posterior glottal opening (pGO)** refers to an open gap at the posterior end of the vocal folds, between the two arytenoid cartilages. This permits air from the

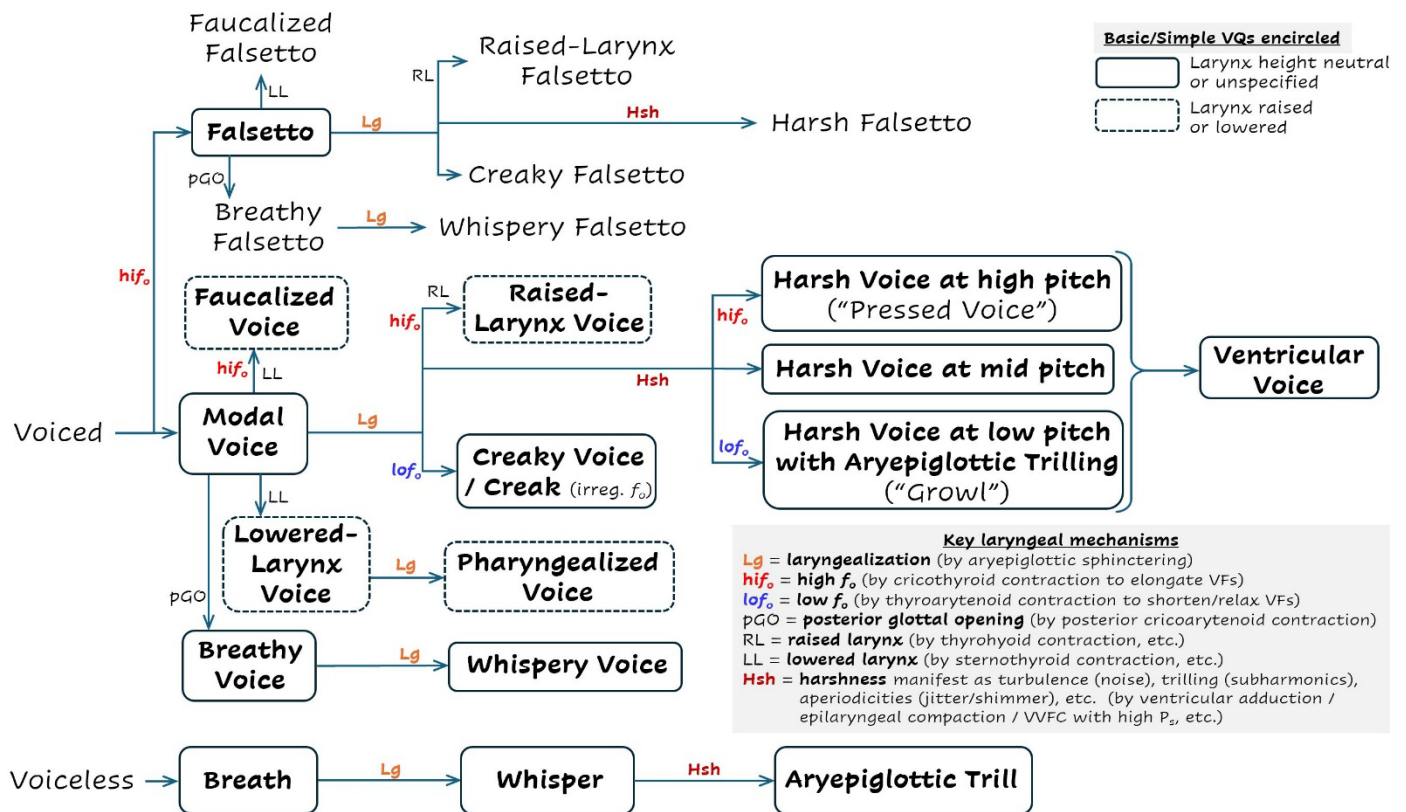


Figure 1. Laryngeal Voice Quality Chart, with simple/basic (as opposed to compound/mixed) categories encircled.

lungs to continually pass through the chink even while the anterior (ligamental) vocal folds oscillate, resulting in *breathy voice* and, when simultaneously laryngealized, *whispery voice*. **Larynx raising/lowering** refers to control of the vertical displacement (height) of the entire larynx above/below its neutral, modal position. While many of the VQs can be produced with a range of possible larynx heights, explicitly *raised-larynx voice* requires concomitant laryngealization and a relatively high f_0 . An explicitly lowered larynx can result in either *lowered-larynx voice* (with mid-to-low f_0), or *faucalized voice* (with high f_0), or *pharyngealized voice* (with laryngealization). In the chart, these four VQs that involve explicit larynx raising/lowering are encircled with dashed lines.

It is worth noting that the chart does not explicitly specify the *vocal effort* (or subglottal pressure) required to produce each VQ. While some VQs naturally tend to be produced either more softly (e.g., breath or breathy voice) or more loudly (e.g., harsh voices), many have a wide range of possible intensities. Similarly, the chart does not explicitly make a distinction between *tense* vs *lax*. These terms have sometimes appeared in the literature, often with ambiguity in regard to the location of the muscular tension or relaxation. It seems preferable to specify tension more precisely in terms of the LAM's key parameters: degree of pharyngealization, degree of larynx raising/lowering, and so on. Also, provided physiological compatibility, two or more *basic/simple* VQs can be combined to produce numerous *compound/mixed* VQs not listed in the chart, such as *harsh whispery creaky voice*.

III. CONCLUDING REMARKS

The proposed chart organizes the basic laryngeal voice qualities in a hierarchy that emphasizes their interrelationships, in terms of the key mechanisms described above. We hope that the chart, together with accompanying physiological descriptions and audio samples to be presented, will facilitate more inclusive and comprehensive research on laryngeal voice quality.

ACKNOWLEDGEMENTS

This work was partially supported by JSPS KAKENHI 24K15013. I am grateful for Prof. John Esling's generous advice.

REFERENCES

- [1] J.Laver, *The phonetic description of voice quality*. Cambridge: Cambridge University Press, 1980.
- [2] J.H.Esling, S.R.Moisik, A.Benner, and L.Crevier-Buchman, *Voice quality: the laryngeal articulator model*. Cambridge: Cambridge University Press, 2019.
- [3] K.Honda, T.Kitamura, H.Takemoto, S.Adachi, P.Mokhtari, S.Takano, Y.Nota, H.Hirata, I.Fujimoto, Y.Shimada, S.Masaki, S.Fujita, and J.Dang, "Visualisation of hypopharyngeal cavities and vocal-tract acoustic modelling," *Computer Methods in Biomech. and Biomed. Eng.*, vol. 13, no. 4, pp. 443–453, Aug. 2010.
- [4] W.R.Zemlin, "Phonation," in *Speech and hearing science: anatomy and physiology*, ch. 3, pp. 100–196, Boston: Allyn and Bacon, 1998 (4th edition).
- [5] N.Gibbins, "Larynx," in S.Standring & R.S.Tubbs (eds.) *Gray's anatomy: the anatomical basis...*, ch. 44, pp. 773–792, Elsevier, 2026 (43rd edition).
- [6] J.G.Švec, H.K.Schutte, C.J.Chen, and I.R.Titze, "Integrative Insights into the Myoelastic-Aerodynamic Theory and Acoustics of Phonation. Scientific Tribute to Donald G. Miller," *Journal of Voice*, vol. 37, no. 3, pp. 305–313, 2023.