PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Accounting for variability over the voice range

Sten TERNSTRÖM[1]; Peter PABON[2]

[1] KTH Royal Institute of Technology, Sweden

[2] Royal Conservatoire, The Netherlands

ABSTRACT

Researchers from the natural sciences interested in the performing arts often seek quantitative findings with explanatory power and practical relevance to performers and educators. However, the complexity of singing voice production continues to challenge us. On their own, entities that are readily measurable in the domain of physics are rarely of direct relevance to excellence in the domain of performance; because information on one level of representation (e.g., acoustic) is artistically meaningful mostly when interpreted in a context at a higher level of representation (e.g., emotional or semantic). Also, practically any acoustic or physiologic metric derived from the sound of a voice, or from other signals or images, will exhibit considerable variation both across individuals and across the voice range, from soft to loud or from low to high pitch. Here, we review some recent research based on the sampling paradigm of the voice field, also known as the voice range profile. Despite large inter-subject variation, the localizing by f_0 and SPL in the voice field will make the recorded values very reproducible within subjects. We demonstrate some technical possibilities, and argue the importance of making physical measurements that provide a more encompassing and individual-centric view of singing voice production.

Keywords: variability, voice range profile, voice analysis

1. VARIABILITY IN VOICES

Acoustic voice analysis is concerned mostly with identifying and quantifying those qualities or parameters that are relevant for assessing the health or training status of a voice, or that characterize the individual quality, e.g. for identification or for artistic/creative purposes. However, the human voice production mechanism has very many degrees of freedom that affect its acoustic output. While this makes it a very expressive instrument, it also complicates any quantitative analysis. Many studies in voice analysis have taken the approach of sampling metrics (e.g., on spectral features, f_0 perturbations, noise content, etc) at only one or a few points in the voice range. But where the voice varies a lot, this amounts to undersampling from an unknown distribution, whereas in fact it may be the shape of the distribution that carries the relevant information. One consequence of this is that acoustic and physiological voice metrics have so far been difficult to generalize for establishing norms, as needed for providing clinical evidence (1). We submit that this is a large problem that begs for the application of modern 'big data' and visualization techniques. The present paper is an overview of some key points made in Pabon's recent thesis (2), where an in-depth discussion of the measurement paradigm is given; here supplemented with data from other sources. These sources all seek to characterize voices by compiling intra-subject *distributions* of voice characteristics, rather than just a few samples from such distributions.

2. THE VOICE FIELD

The acoustic signals of voiced sounds have two primary attributes: fundamental frequency (f_0) and sound level (SPL). Voiced sounds have also numerous secondary attributes, or 'voice qualities', that can be derived from the acoustic signal, especially from its periodicity and its spectrum. All such

[1] stern@kth.se
[2] pabon@koncon.nl

voice qualities co-vary essentially and individually with the fundamental frequency and the sound level. Therefore, methods for assessing the complete voice need to account for such co-variation and individuality.

From the legacy German term *Stimmfeld*, the 'voice field' is the coordinate system with f_o in semitones on the horizontal axis and sound pressure in dB on the vertical axis; note that this is a log/log format. These axes are the established frame of reference for the voice range profile, a.k.a. the phonetogram (3), but apply equally well for phonation anywhere between the extremes. We may map onto this plane any scalar metric representing some physical aspect of voice production, using a colour scale and/or isolines, as on a geographical map. This can be done for any vocal task, such as running speech, sustained phonation, singing a song, etc.

3. INTRA-SUBJECT REPRODUCIBILITY

Figure 1 shows for two female singer subjects (top and bottom row) how the electroglottographic (EGG) contact quotient Q_{ci} varied over the voice field. The left and middle panels show maps resulting from two disjunct sets of replicated productions, of singing a short song 18+18 times, at different dynamic levels ***p-mf-f*** so as to cover a wide range in SPL; data from (4). The 'islands' are of similar shape, because the song was the same. The right panels show the colour-coded differences between the two. It is clear that the two singers replicate their own productions quite faithfully, and also that Q_{ci} varies systematically, with connected regions of colours across the voice range. But it is also clear that the Q_{ci} patterns of variation are rather different in the two singers, to a thought-provoking extent, for such a basic aspect of phonation. It is a general observation, for all metrics of the voice: subjects are typically quite consistent within themselves, when productions are matched for f_o and SPL. The main exception to this rule is for metrics that display some form of hysteresis. For instance, in singers the contact quotient Q_{ci} can be different on ascending and descending pitch, in local regions of the voice map; but systematically so: the averages of repeated tasks remains stable.

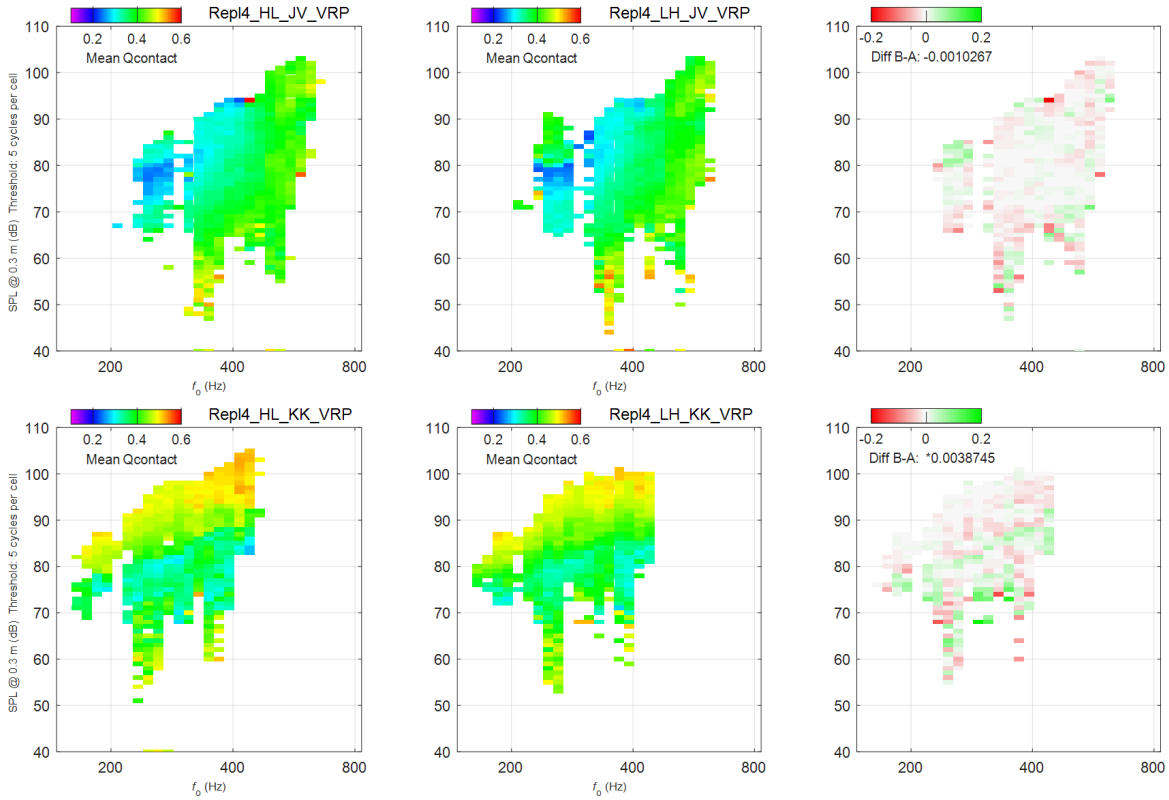

Figure 1: Reproducibility: aggregated voice field maps of the EGG contact quotient of two trained female singers singing the same short song many times at various dynamic levels ***p-mf-f*** . Top, a classical soprano; bottom, a jazz singer, singing in a lower key. The left and center panels show outcomes from disjunct replicated sets of the same task. The right panels show the small differences between the two sets. Each cell (1 semitone × 1 dB) contains a contact quotient average over typically hundreds of phonatory cycles.

Note that the Q_{ci}, just like any other voice metric, follows gradients over SPL and f_o. While these gradients are consistent within subjects, they tend to be particularly steep in the lower left-hand region of the voice range, that is, in the range of habitual speech. Hence, if a clinician asks a patient to phonate "at a comfortable loudness and pitch", it is likely that the gradient for the observed metric is steep, and will give rise to a large variance in repeated measurements. This makes perfect sense in terms of efficiency of communication (a large change in output for a small change in effort), but it also means that the conventional methods of assessing acoustic voice metrics tend to suffer from underestimated sources of variation.

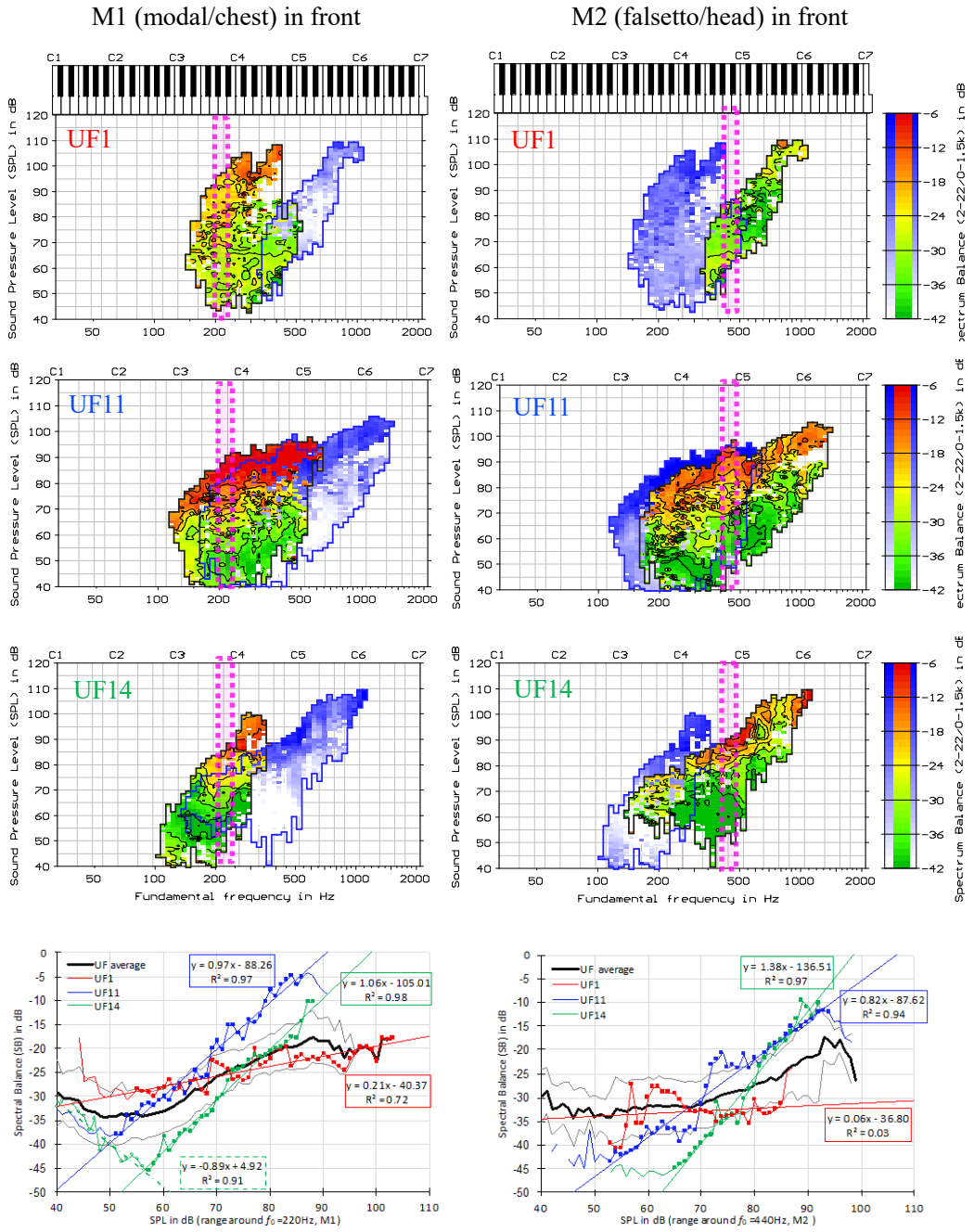

Figure 2. Inter-subject variability: top three rows: voice maps of SB, three untrained female subjects in M1 (modal/chest voice, left) and M2 (head/falsetto voice, right). The spectrum balance SB is mapped to colour (vertical bars). Bottom row: SB as function of SPL in M1 at 220 Hz (left) and in M2 at 440 Hz (right), at the sections indicated by dotted rectangles in the voice maps. Regression lines are based on the parts of the curves with marker dots.

4. INTER-SUBJECT VARIABILITY

Many interesting features can be derived from the per-cell spectra of sustained [ɑ:] vowels, for which a very large data set was recently provided by Pabon & Ternström (5). For example, the spectrum balance (SB) is often used as a metric of the relative power of high frequencies in the voice spectrum. SB is defined as the ratio in decibels between the spectral powers in a high and a low frequency band (here, >2 kHz and <1.5 kHz). Figure 2 shows examples of voice maps of SB, for which three untrained but vocally healthy females explored their full range from soft to loud and from low to high, with guidance from a trained operator following a stringent protocol. In that study, care was taken during the recording to record the chest and head registers separately (left and right panels), since these are two very different phonatory mechanisms. Notice how very different these three voices are, not only in the ranges of their two mechanisms, but also in how the SB varies over the voice range.

5. AVERAGING ACROSS INDIVIDUALS

The large variation across individuals makes it difficult to derive any meaningful population-based normative criteria e.g. for pathologies or skills, and thereby presents a challenge to conventional objective voice assessments. In general, it is possible to quantify the vocal status of students or patients only as compared to themselves across a change, or to a narrowly defined group of individuals that are not only of the same sex and approximate age but also within the same training programme, musical genre, or treatment regime. Making group averages will be meaningful only if the participating voices are sufficiently similar to one another to be relevant to the research question at hand. Such selection is however possible, and there are several ways of averaging multiple voice maps into one (6). By averaging, certain general trends can still be revealed.

For instance, Figure 3 shows SB maps derived from spectra that were power-averaged over 12 female classical singers. The singers were training in the same programme for classical voice, and their individual voice *ranges* were somewhat less diverse than those of the untrained voices in Figure 2, yet their SB patterns over those ranges were still quite different from one another. The sum, or *union*, of their voice maps is however much smoother. These SB distributions suggest that, for this group, SB seems to be largely independent of f_0 in chest voice (left), but not in head voice (right).

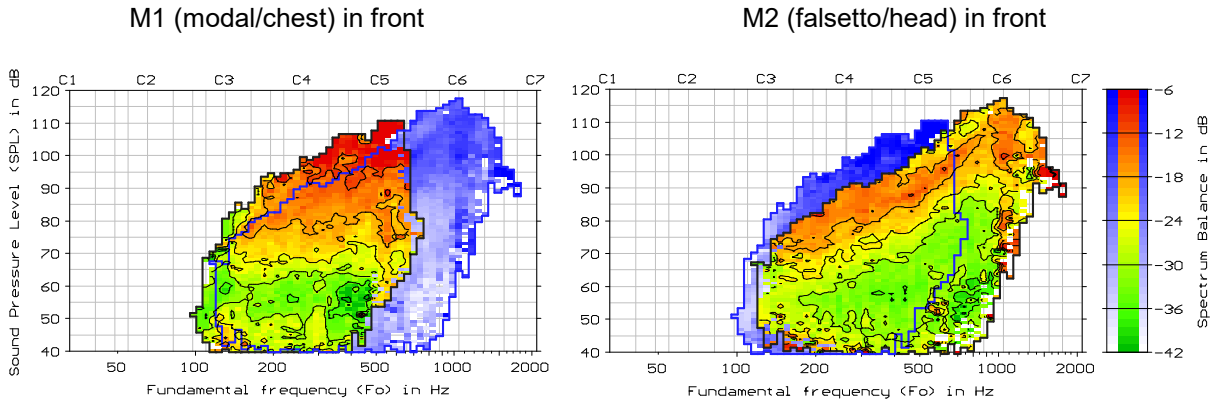

Figure 3. Maps of the spectrum balance SB, derived by averaging the power spectra over the union set of the VRP areas, for 12 trained female voices.

6. LOGARITHMIC SCALING

Fundamental frequencies in performing voices can range from about 2^5 Hz to 2^{11} Hz, spanning some six octaves. In sound level, the range of a single trained voice can be even larger, 40-120 dB SPL, or eight *decades* of acoustic power, 1:10^8. This is a huge range that challenges the SNR of microphones and audio reproduction chains, and exceeds the dynamic range of all other musical instruments that are based on a single sustained-tone oscillator. Clearly the physical reality of soft phonation is very different indeed from that in loud phonation.

Our senses, especially of hearing, have evolved to perceive most real-world entities on logarithmic scales, which compared to linear scales afford a more useful overview and more equal emphasis to small and large things alike. Given that voices operate over such large ranges, our man-made measuring instruments (which, due to the physics, operate on linear scales), must be designed to resolve very fine detail over very large ranges. For times and frequencies, this is readily achieved with digital technology; but for amplitudes it requires great attention to the dynamic range of the measurement system.

7. EXPOSING CONTROL REGIMES ON LOG SCALES

Although it is out of scope for this short paper, it can be shown that systematic relationships of a multiplicative nature will manifest themselves as linear relationships in a log/log/log space.

Voice properties do not necessarily change uniformly or smoothly with SPL or f_o. For instance, the vibratory modes of the vocal folds are in some respects state-bound, and phonation can transition abruptly from one mode to another. Again, this varies greatly from one person to another. Even within one mechanism, such as M2/head voice/falsetto, things can change a great deal. In Figure 2, the bottom row shows two subsets of the data above, created by vertically slicing the M1 voice maps at 220 Hz and the M2 voice maps at 440 Hz. The SB at each SPL is the intra-subject average from the five adjacent cells at f_o ±2 semitones. At f_o=440 Hz, we see how the SB changes abruptly at around 70 SPL for subject UF11 (blue). For UF1 (red) on the other hand, there is little change in SB from 60-85 dB SPL, while UF14 (green) exhibits a practically linear increase in SB on the interval 65-90 dB SPL. Note that a fair approximation can be made of these graphs using piecewise linear segments in three ranges: in M1, we have 40-55 dB, 55-85 dB and >85 dB; in M2, these ranges are shifted rightwards by about 10 dB. This piecewise-linear behaviour suggests that the ranges represent different internal dependencies that take turns in dominating the voice production, as SPL changes. For example, the black curves (the SB mean of 16 subjects) tend to flatten or even turn down at high SPLs, suggesting a spectral saturation effect, as observed earlier by Ternström *et al*. (7).

For the purposes of this overview, we have exemplified variations in the voice using mostly spectral properties. The same general observations hold also for how vocally generated noise and f_o perturbations vary substantially across the voice range (8). Cycle-to-cycle perturbations are of particular interest in charting the voice, since, originating at the glottal source, they are largely independent of the filter, and hence of the spectral aspects of the resulting sound. Since external noise sources tend to impose a minimum level of perturbations, such metrics are also useful for exposing shortcomings of the recording setup.

The central interest in the voice field recording paradigm is to map the *proportional* dependencies that exist between voice parameters. These proportional dependencies are characterized by shared exponential factors, which are exposed by mapping under a pervasive logarithmic scaling, as exemplified, we submit, by the bottom curves in Figure 2. The recording paradigm conforms to a maximum entropy criterion, which has its own specific statistics. This criterion presents an optimization principle for individual voice parameters, and maximizes the probability of revealing a possible connection between voice parameters, given the total information content observed. The recurring outcome is that voice metric distribution shapes differ considerably between voices, but are consistent within a voice. This in-voice consistency makes it possible to navigate on the map of an individual voice, while the differences between voices make it difficult to derive a general picture by averaging. Absolute scales and absolute positions, with which we are so familiar in experimental research, are more a hindrance than a help when studying a phenomenon so variable as human voice production. There is no guarantee that all voices will work from the same point of reference and will scale in the same manner. It is precisely in this dynamic scaling that the individuality of the voice is apparent.

8. CONCLUSION

Accounting for the variability of the voice requires a mapping paradigm that recognizes the variation generated by numerous voice production mechanisms. The relative influence of these mechanisms on the emitted sound will vary across individuals and over the range of the voice. The voice field is one mapping alternative, with the convenient property of requiring only the acoustic signal. By consistently representing data on logarithmic scales, it becomes more likely that underlying multiplicative dependencies can be identified. However, logarithmic scales force us also to reconsider our notions of variances and averages, and of how the significance of differences might be tested for.

ACKNOWLEDGEMENTS

The work referred to herein was carried out over many years and many projects, with several funding sources, hereby gratefully acknowledged, including the Royal Conservatoire, The Hague, NL; KTH Royal Institute of Technology, Stockholm, SE; and the Swedish Research Council contract #2010-4565. We are grateful also for the participation of the many subjects. Svante Granqvist astutely pointed out the communicative benefit of steep gradients in the speech range.

REFERENCES

1. Roy N, Barkmeier-Kraemer J, Eadie T, Sivasankar MP, Mehta D, Paul D, et al. Evidence-Based Clinical Voice Assessment: A Systematic Review. Am J Speech-Language Pathol. 2013; 22(2):212–26.
2. Pabon P. Mapping Individual Voice Quality over the Voice Range: The Measurement Paradigm of the Voice Range Profile. PhD thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2018. ISBN 978-91-7729-958-5.
3. Schutte HK, Seidner W. Recommendation by the Union of European Phoniatricians (UEP): Standardizing voice area measurement/phonetography. Folia Phoniatr Logop. 1983;35(6):286–8.
4. Ternström S, D'Amario S, Selamtzis A. Effects of the Lung Volume on the Electroglottographic Waveform in Trained Female Singers. J Voice 2018 e-pub ahead of print; DOI: 10.1016/j.jvoice.2018.09.006
5. Pabon P, Ternström S. Feature Maps of the Acoustic Spectrum of the Voice. J Voice 2018 e-pub; DOI: 10.1016/j.jvoice.2018.08.014.
6. Pabon P, Ternström S, Lamarche A. Fourier Descriptor Analysis and Unification of Voice Range Profile Contours: Method and Applications. J Speech, Lang Hear Res. 2010; 54(3):755–76.
7. Ternström S, Bohman M, Södersten M. Loud speech over noise: Some spectral attributes, with gender differences. J Acoust Soc Am. 2006; 119(3):1648–65.
8. Brockmann-Bauser M, Bohlender JE, Mehta DD. Acoustic Perturbation Measures Improve with Increasing Vocal Intensity in Individuals With and Without Voice Disorders. J Voice 2018; 32(2):162–8.