Acta Acustica united with Acustica
Vol. 102 (2016) 268 – 283

DOI 10.3813/AAA.918943

The Voice Range Profile: Its Function, Applications, Pitfalls and Potential

Sten Ternström[1)], Peter Pabon[1,2)], Maria Södersten[3,4)]

[1)] Department of Speech, Music and Hearing School of Computer Science and Communication, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden. stern@kth.se

[2)] Institute of Sonology, Royal Conservatoire, The Hague, The Netherlands

[3)] Division of Speech and Language Pathology, Department of Clinical Science, Intervention and Technology, Karolinska Institutet, 171 77 Stockholm, Sweden

[4)] Department of Speech and Language Pathology, Karolinska University Hospital, 171 76 Stockholm, Sweden

Summary

An overview is given of the current status of the computerised voice range profile (VRP) as a voice measurement paradigm. Its operating principles are described, and sources of errors and variability are discussed. The features of the VRP contour and its characterisation are described. Methods for performing statistics on VRP contour and interior data are considered. Examples are given of clinical, pedagogical and research applications. Finally, issues with the models used to interpret VRP data are discussed. It is concluded that, while the VRP offers a convenient frame of reference for a multitude of voice assessment metrics, it also exposes the many degrees of freedom in the voice to an extent that challenges us to improve our models of how the voice functions over a large range and in a dynamic setting.

PACS no. 43.60.Lq, 43.70.-h, 43.70.Aj, 43.70.Bk, 43.70.Gr, 43.70.Jt, 43.75.St

1. Introduction

At first sight, the vocal folds seem quite unremarkable: suspended across the tracheal duct, they are a pair of elastic cords or 'lips' that vibrate when air is forced past them. Yet the act of phonation has so many control aspects and degrees of freedom that an exhaustive description of phonatory phenomena remains elusive. The clinician, the teacher, the training performer and the voice scientist all need ways of presenting multidimensional voice data as succinctly as possible.

The voice range profile (VRP), also known as the phonetogram, is a format for presenting a summary of voice data in a single picture, much like the audiogram quantifies and summarises the basic range of a person's hearing in one graph. While the VRP exists in many variations, it nearly always has a horizontal axis for fundamental frequency (f_o)[1] in semitones, and a vertical axis for sound pressure level (SPL) in decibels (Figure 1). Early VRP's were made manually, using little more than a sound level meter and a keyboard for reference pitch; yet with stringency and care, useful results could be obtained (e.g., [2]). Here, we will be concerned with computerised methods only. A recommendation for the standardised presentation of the VRP was issued in 1983 by the Union of European Phoniatricians (UEP) [3], and it has since been widely respected.

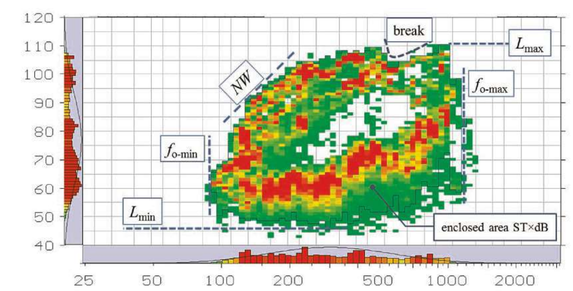

Figure 1. Basic elements and main features of the VRP display, here of a healthy adult female. Conventionally, the horizontal axis shows log f_o in semitones, and the vertical axis shows voice SPL @ 0.3 m in decibels, 40 to 120 dB. The usual bin or cell size is 1 semitone ×1 dB. The cell width:height aspect ratio should be kept at 2:1 to facilitate visual comparisons. Here, the colour scale is assigned to accumulated time, giving a 'density' plot. The thin solid line shows the boundary inside which the accumulated time per cell exceeds 20 ms. The primary metrics of the contour are the minimum f_o (f_{o-min}), maximum f_o (f_{o-max}), f_o range, minimum level (L_{min}), maximum SPL (L_{max}), SPL range and the total enclosed area (ST×dB). The labeled features are discussed in the article text (colour online).

The purpose of making a VRP is usually to generate a representative map of some salient aspect of a given voice.

Received 04 May 2015,
accepted 10 August 2015.

[1] The notation of acoustic voice parameters in this article follows the new recommendations in the recent consensus paper by Titze et al [1].

For the basic form of the VRP, this aspect is simply the ability to phonate at all. The speaker or singer subject is asked to exercise his or her full vocal range from loud to soft to from high to low, usually with real-time visual feedback, and a record is made of the voice range, that is, the attained extremes of f_o and SPL. Thus one obtains the VRP 'contour' or perimeter. This contour is that aspect of the VRP that so far has received the most research attention. We will discuss its potential and its limitations below. In daily life, however, phonation at the extremes is quite rare, and therefore smaller regions inside the maximum VRP contour are also relevant for characterising the practical functioning of a voice. For instance, the speech range profile (SRP) is often appropriate for characterising habitual speech in normal or challenging situations, usually using averages and bounds [4, 5], and some researchers have suggested that the comparison of the VRP to the SRP will give clinically useful insights [6, 7]. Similarly, researchers of singing voice have suggested a 'singing range profile' for characterising those regimes of phonation that are intended for artistic performance [8].

Many conventional voice metrics, based on a brief vowel sustained on a single pitch, will occupy but a single cell in the VRP. Only in special cases could such a 'keyhole view' be taken as representative of a person's voice status. Technically, it is straightforward to map any measured scalar voice entity onto a gray scale or a color gradient. When this is done over all interesting positions of f_o and SPL, and not just at the extremes, the information content of the image can increase dramatically. This is where the VRP really comes into its own, replacing the keyhole with a panorama display, as it were [9]. The most common mapping is the 2-D histogram, in which the colour at any given (f_o, SPL) position represents the accumulated time spent there [10]. This can be useful for studies of habitual ranges, or as a verification of really having tried to phonate in a given part when recording the VRP [4]. But any scalar parameter, such as f_o jitter, spectral balance, or the relative level of the fundamental partial can be visualised, over the entire f_o/SPL plane. This has the potential for revealing trends and regimes of phonation, and for increasing our understanding of phonation in general. With supplementary transducers, mapping can be done also of non-acoustic metrics, for example, from an accelerometer, electroglottograph, or pressure transducer. Some examples of such studies will be given.

A major challenge in obtaining the VRP and making it useful in practice is the large individual variation that it exposes, and the many sources of variation to which it is susceptible. Also, the recording procedure and the precise methods for measuring the independent variables of fo and SPL may have a decisive influence on reproducibility. Several researchers have voiced concern over what has been perceived as the excessive variability of VRP results [11, 12]. In our view, this is more a reflection of the complexity of human voice production; and not that of the VRP. Still, variability and the lack of sufficiently complete underlying models has no doubt been a restraining factor for a more general adoption of the VRP as an analysis framework. In Section 5 below we will point to some important sources of variation, be they inter-subject, intra-subject, procedural, or just technical; and we will point to how their influences might be negotiated. In the Discussion, we will consider why and how current models of voice production need to be refined for the VRP.

The variability of the voice immediately points to the need for performing statistical operations on ensembles of comparable VRPs. Since we are dealing with multidimensional data, this raises several issues as to how averaging might be done, and how to choose meaningful statistics. As always, the right choice of statistical treatment will depend on the research question. Still, some general alternatives will be described.

As with any measurement, it is crucial to have a model against which to interpret observations made using the VRP. The versatility of the VRP actually means that we may need a specific model for every research question. For example, different voice pathologies may affect the level of the lower contour for different reasons. This means that the level of the lower contour, while being very useful as a general metric, is not in itself indicative of any particular pathology. The same applies for any number that we can derive from the VRP. Some examples will be given of how vocal mechanisms manifest themselves in the VRP.

So, how does the VRP relate to evidence-based diagnostics and treatment? As noted, the VRP is not in itself a measurement; rather, it is a mode of presenting compactly a very large number of similar measurements. It thereby produces an image, which, like any other medical image, needs to be interpreted. From this image, some aspects of the voice are immediately clear, while others are more subtle. From a VRP plot, a whole set of metrics can be derived, each of which may be simultaneously dependent on more than one voice production mechanism. Each such metric, such as phonation threshold level or maximum f_o, must be interpreted and validated in the context of a given voice problem or voice characteristic. Hence, the immediate value of the VRP as a mode of presentation lies not in providing direct support for evidence-based inquiry, which on its own it does not. Rather, the strengths of the VRP are that it affords an overview of a voice's range, qualities and changes over training, treatment or time, and that it can provide real-time feedback of hitherto neglected dynamical aspects of voice. This will be further elaborated upon in the Discussion.

The paradigm of the voice range profile/phonetogram/stimmfeld has by now a long history, dating back to the 1930's [13]. The reader interested in the background picture is referred to the extensive literature review in the thesis of Lamarche [14].

2. Principles

2.1. The frame of reference: f_o and SPL

Given that loudness, pitch and timbre are the three basic perceptual aspects of tones, the first two of these dimensions suggest natural independent variables for the generalised mapping of phonatory function: one metric to represent the 'vocal strength', and another metric to represent the f_o of phonation. We here use 'vocal strength' in single quotes, to avoid confusion with the formal psychoacoustics metric of loudness and with the physical acoustics metric of intensity. Conventionally, and mostly for convenience, the metric of sound pressure level (SPL) is taken to represent 'vocal strength'; while the repetition rate of the quasi-periodic voice signal is taken to represent f_o. Other choices of independent variables are possible (such as, in lieu of SPL, the subglottal pressure, or the level from a neck-mounted accelerometer [15]), and in some contexts even preferable. Here, however, we will consider only f_o and SPL. Since they constitute the frame of reference, it is important to understand the implications of adopting them as such. For f_o, the voice signal being analysed must have a periodicity that can be defined, and hence some types of severely dysphonic voices are simply not amenable to VRP analysis. For SPL, the main concern is to account for the combined influences of the frequency dependencies of the power transfer functions of the vocal tract and of the recording chain. The former is time-variant and depends on vowel articulation and especially on the area of the lip opening [16]. The latter transfer function is typically constant, and depends on the microphone(s), the room acoustics, and the frequency weighting function chosen for measuring the sound level (A-weighted or not).

Volumes of research have been dedicated to the estimation of voice signal f_o and SPL over different time scales. Dozens of procedures have been proposed and established in different domains of voice and speech research, and we cannot go into details here. The issue of time scale, however, is central [17]. In particular, one must consider whether the signal analysis should be performed for each glottal cycle ('pitch-synchronously') or over a fixed time window, the latter generally being simpler to implement. This choice will depend on the research question or on the application. If the objective is to establish a VRP contour only, or to map voice signal properties that are not cycle-specific, then a fixed-window analysis may be suitable. Several voice properties, however, such as f_o perturbations, are manifest as deviations from an underlying trend. For instance, in order to map the magnitude of f_o perturbations over f_o and SPL, we might want for the independent variable an f_o metric that is averaged over a relatively long time window (tens of cycles), and for the dependent perturbation variable a metric of cycle-to-cycle deviations from the trend.

2.2. Detecting phonation

In many other types of speech processing systems, the criterion for detecting phonation is an amplitude threshold: if the signal energy in a frequency range that could include f_o exceeds a certain amplitude, phonation is assumed to be present. With the VRP, however, one of the goals is precisely to establish the threshold of phonation. There exists *a priori* no minimum amplitude, and so another criterion must be found. Often a regularity threshold or jitter constraint is used instead [9], for example that the estimated period times of the signal must have a standard deviation *below* a certain threshold, on the order of one or a few percent. This means that very irregular phonation will be rejected, and also that at least a few cycles (3–7) of phonation must occur before they can be recognised as such.

Often there will also be a time threshold, such that the colour of a given cell will be changed from white only when the signal has accumulated a minimum number of milliseconds of voicing at that position. This time threshold may differ between systems, and can be adjustable *post hoc*, such that outlier points can be excluded, for example.

If two VRP systems have different thresholds for detecting phonation, by adjustment or by principle of operation, then the graphs they produce will be slightly different, especially at the lower contour. The estimated phonation time, too, will differ slightly.

A common problem in tracking f_o is that the second harmonic at $2f_o$ occasionally dominates the voice waveform and its spectrum to such a degree ($> 20\,\mathrm{dB}$) that the algorithm will erroneously select $2f_o$ in lieu of f_o. Erratic octave jumps may be displayed as a result. Some VRP systems therefore provide tools to exclude selected regions.

3. The VRP contour

3.1. Lower contour

The sound level of the lower contour represents a lower threshold of phonation. The concept of such a threshold must be carefully defined, for several reasons. Firstly, since onset and offset of phonation may exhibit hysteresis [18], i.e. occur at different levels at the start and end of vocalisations, the exact procedure for obtaining the threshold is important and needs to be relevant to the purpose of making the VRP. Secondly, phonating in a very soft voice is to some extent a motor skill that can be improved with training. Thirdly, the vibratory amplitude of the vocal folds at low glottal flow can be arbitrarily small, especially at the cessation of the driving subglottal pressure, when the abducting vocal folds may come to rest with an amplitude of vibration decaying over several cycles.

In clinical applications, the sound level of the phonation threshold is usually taken as relating to the vocal fold elasticity and pliability, both of which need to be normal to enable soft phonation [19]. In singing training, the lower threshold might instead be defined as the softest tone that would be acceptable for a performance on stage, for example.

3.2. Upper contour

The level of the *upper contour* will be affected by (1) the maximum glottal flow derivative that the subject can produce [16], (2) the area of the mouth opening as in open versus closed vowels [16], (3) in some situations, the acoustic resonances of the vocal tract, (4) a non-linear maximum vibratory amplitude leading to saturation or "clipping" [20], and (5) the voice register (modal/falsetto) or mechanism M1/M2 [21]. Clinically, the level of the upper contour appears to be related to the achievement of tension in the thyroarytenoid and adductor muscles that is necessary for building up subglottal pressure to produce a very loud voice.

3.3. Limits in fundamental frequency f_o

The f_o is governed by both the effective vibrating mass and the stiffness of the vocal folds [22]. The effective vibrating mass and stiffness are both outcomes of a rich set of physical dimensions and muscle activations. *Minimum* f_o is attained for a laryngeal posturing with relaxed, short vocal folds and a modest transglottal pressure. *Maximum* f_o is attained with high activity in the cricothyroid and thyroarytenoid muscles giving lengthened, tense vocal folds, combined with high subglottal pressure, which additionally requires a large adductory force to keep the vocal folds together. In particular, the use of the M2 mechanism (falsetto register) gives access to a much higher maximum f_o than does M1 (modal register). In M2 the vocal fold tension is relatively high and the effective vibrating mass is typically smaller.

The type of voicing that is known as glottal 'fry' (M0) refers to a train of glottal pulses whose repetition frequency is given not by entrainment of the dominant vocal fold mode frequency, but rather by repeated build-up and valved release of subglottal pressure. The period time is then governed more by the average flow and the compliance of the vocal folds. The resulting pulses are often not so periodic, and might thus be classified as non-phonation by most VRP implementations.

3.4. Contour landmarks

In the contour of a full range VRP, there are some common landmarks to observe. The first is the dip in the *upper contour* (Figure 1), at the point where the subject switches between modal and falsetto phonation (M1 to M2). It actually marks the upper intersection of two upward-slanting ellipse-like regions that enclose the productions in these two registers (Figure 5). The second landmark is the *upward sloping portion* of the upper contour from low to medium pitches, sometimes referred to as the "north-west." It has been observed [23] that this is the part of the contour that varies the least between subjects (even of different genders) and also between populations and investigators. This trend is visible in the first and second panels of Figure 5, and also in the intra-subject test-retest comparisons shown in Figure 3. It seems that this upper limit is fairly easy to elicit in a uniform manner, and that it is the least susceptible to SPL variations due to variations in vocal tract resonances. Therefore it could be of interest in relation to vocal pathologies to study especially this portion of the contour. In contrast, the level of the lower contour is much more variable, for the reasons mentioned above.

Voice range profiles produced by vocally healthy subjects exhibit an overall positive slope, i.e., SPL increases with fo, by about 8–10 dB per octave [16], and 10–13 dB per octave in singing [14]. It is not uncommon for the loudest voice at low frequency to be lower in level than the *softest* voice at the high end of the f_o range, particularly when the subject is in a singing mode.

4. The VRP interior

4.1. Mapping a third metric

While the VRP contour, if correctly registered, indicates the operating limits of a voice, a large amount of interesting variation in the voice signal occurs also over the *interior* of the contour. Assuming that f_o and SPL are the independent variables, we may select any interesting scalar metric, map its value on a colour scale, and thus implement a display of how this third metric varies as a dependent variable over fundamental frequency and voice sound level [9], inside (and on) the contour. The subject can then "paint" with her voice so as to fill the contour and produce a "voice map". It does take time to visit all cells, typically at least 10–20 minutes for trained subjects or singers, so in a clinical setting, the choice of a feasible protocol is important. For research purposes, the time taken is less of an issue; on the other hand, there may be effects of vocal warm-up and/or fatigue over the course of the acquisition. Figure 2 shows the VRP renderings of five different voice quality parameters, all computed from the same audio signal.

4.2. Visual rendering of the dependent variable

In the computerised VRP, the value of the dependent variable in a cell at any given coordinate (f_o, SPL) is usually represented by a gray level or colour from a gradient. In the basic form of the VRP, cells that have been visited at all are coloured black on a white background. With a gray scale, the darkness can indicate the accumulated time of phonation in a given cell (Figure 7); this is often referred to as a 'density plot'. Two-coloured gradients can further enhance the visual interpretation (Figure 2).

Since the time dimension is discarded, we must decide what to display in each cell: a time-averaged value, a held peak value, or the most recent value? This too depends on the reason for making the VRP. For a voice status assessment, the average could be right, but if we are looking for isolated events, a peak value might be more useful. In a biofeedback application, the most recent value is usually appropriate (i.e., the value "now", with a minimal technical latency).

It is even possible to have as the dependent variable a multidimensional metric, such as a section of the entire voice spectrum, time-averaged per cell. This becomes cumbersome to render graphically, but scalar features derived from the spectrum, such as its slope or certain relative partial amplitudes, may then be readily displayed (Figure 2, d–f). This is potentially a very powerful mode of analysis, which can unravel functional trends in phonation over the voice range.

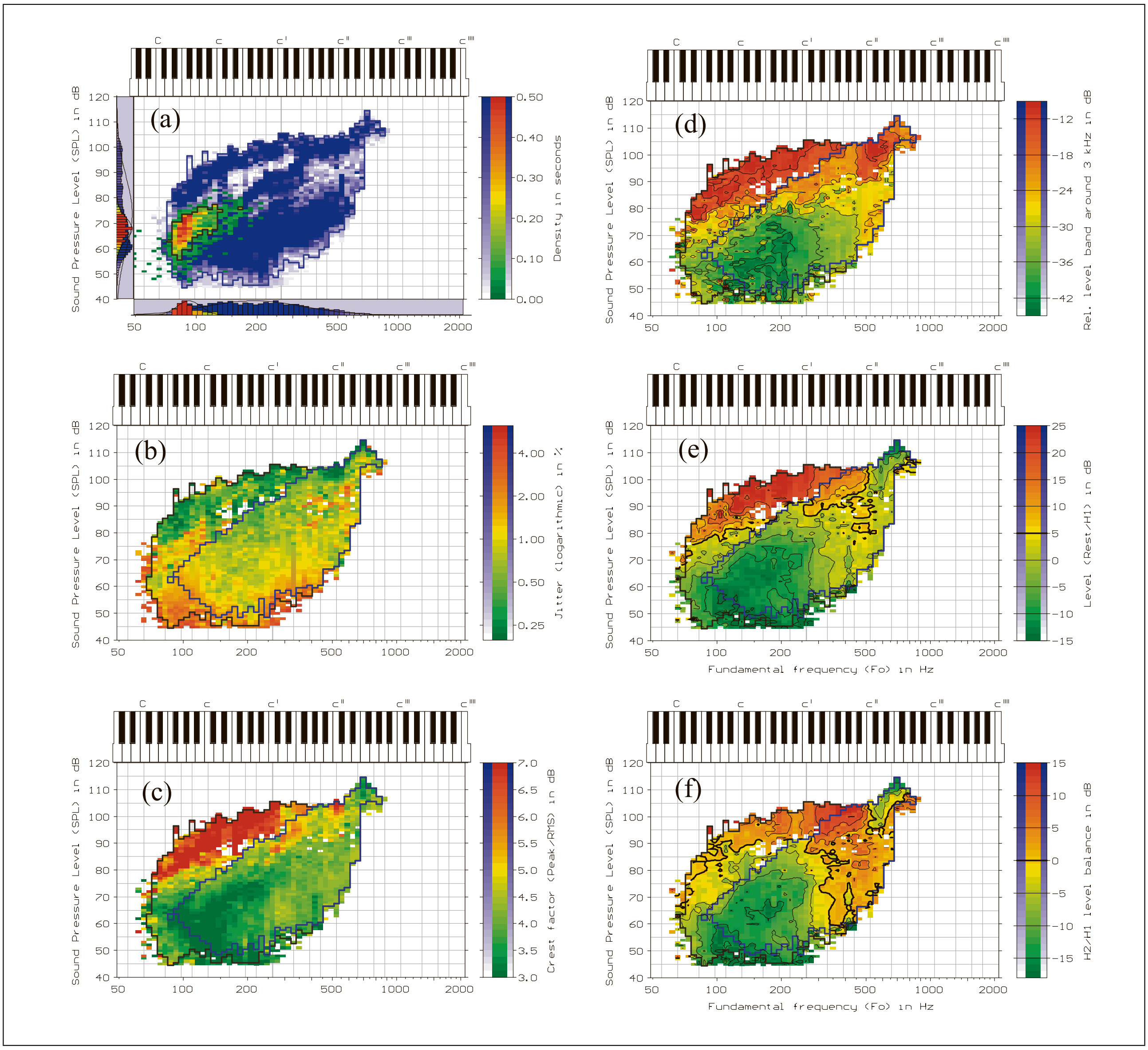

Figure 2. Examples of mapping a third metric into the VRP. The colour scale has a different interpretation with each metric. Male singer. (a) Phonation density (accumulated time); SRP (red-yellow-green) overlaid on VRP area (blue gradient). Registers M1 and M2 were recorded separately, but are merged for this figure. (b) Jitter in % (logarithmic scale). The blue line marks the initial contour seen for register M2. (c) Crest factor ($L_{\text{peak}} - L_{\text{RMS}}$, in dB). (d) Relative level in the 3 kHz band. (e) $L_{H>1} - L_{H1}$. (f) $L_{H2} - L_{H1}$ level difference. The last three metrics are computed from the narrowband spectrum, averaged for each VRP cell. A detailed interpretation of these patterns goes beyond the scope of this paper (colour online).

5. Validity and reliability

5.1. Uniformity of protocol

The VRP outcome is very sensitive to the instructions given to the subject. Often the examiner will need to guide the subject in how to negotiate the phonatory task, and this needs to be done very consistently. It is therefore important to establish a detailed protocol and to verify the examiner's skills and adherence to the protocol. Acquiring a full VRP takes time; approximately 20–30 minutes is suggested for untrained participants and patients [4, 5]. Whether or not a full VRP is actually needed depends on the clinical or research question. With care, the reproducibility can be surprisingly high, as can be seen in Figure 3 when the experimenter followed the same protocol twice [4]. Test-retest reliability was high also in the study by Sanchez *et al.* [5] following the same protocol.

Several different protocols have been suggested [3, 4, 6, 24, 25] that offer a structured and consistent procedure in a given clinical context. Achieving sufficient consistency in the procedure requires explicit training of the persons working with the subjects. It should be remembered also that a VRP contour shows a person's extreme voice capacities, and that a person can achieve this in different ways. Some participants will use the visual feedback and find their voice's potential very much by themselves, while others are much more in need of guidance and prompting from the experimenter. Many participants are helped by

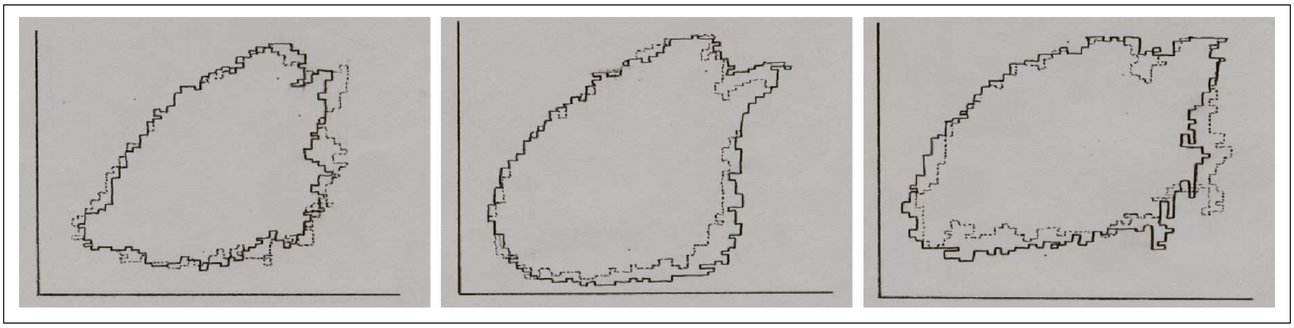

Figure 3. Reproducibility: VRP contours for three vocally healthy male subjects recorded on two occasions (dotted and solid lines), 3–4 months apart. In some parts, the contours from the two recordings are seen to coincide completely. From [4], ©2012, Informa Healthcare. Reproduced with permission of Informa Healthcare.

imitating the experimenter to really understand what is expected.

As a practical example of how to secure validity and reliability, we outline the measures taken with the VRP recording system designed by author P.P. (Voice Profiler, marketed by Alphatron Medical, Rotterdam, The Netherlands). An explicitly goal-oriented recording protocol [24] is supplied as part of the user manual, and a two-day training course is offered. Participants in this course are assessed for their recording skills and receive a certificate on showing that they understand also the rationale that is embedded in the protocol. The Voice Profiler protocol is designed to guarantee the correct recording of a series of VRP-contour targets, together with a series of acoustical voice quality targets that are to be met in approaching of the contour. So, with the contour, also the acoustical qualities that inform on the underlying mechanism are preserved. The protocol is very explicit on how to instruct non-singers to approach a given VRP target. It prescribes an immediate retesting of these targets, as there is an inevitable learning effect that should be anticipated: "the first recording is by its nature also the start of therapy".

5.2. SPL and vowel articulation

Since the voice SPL to some extent depends on the vowel, the recommendation is to use consistently an open /a:/ vowel, when recording the VRP. This is intended to keep most values of f_0 below the frequency of the first formant resonance, so to avoid confounding effects of resonance peaks. It is important to point out that the vocal tract cannot work as an amplifier – that is, the radiated acoustic power cannot exceed the power which is generated at the glottis – but its action as an impedance matcher from the narrow epilarynx to the free field can be more or less effective [16].

There are several studies on the predictions made by source-filter theory on the relationship between vocal tract resonances and the SPL as registered for the VRP. For example, formant-induced ripples in the upper VRP contour were predicted by Titze [26]. However, in our experience, such predictions are rarely borne out in practice.

Lamesch, Doval, and Castellengo [27] studied the role of the vowel in relation to the voice register. They combined and reinterpreted the results from several classic studies of the vowels' influence on maximum intensity for different fundamental frequencies [2, 13, 28, 29], and separated the potential effects of the subjects' sex as well as differences in vocal techniques. Thereafter they shed new light on this matter by supplying averaged VRPs for the two cardinal vowels /a:/ and /i:/ and for the vowel /o:/. All vowels were produced by the same subjects, and with the two voice register mechanisms M1 and M2 recorded separately. The major conclusion from their study is that on the average, the vowel has no influence on the maximum intensities in M2 (falsetto/head), while it does have influence in M1 (modal/chest). In our own VRP recording practice, it is regularly observed that singers claim to be able to produce better results using a different vowel than the prescribed /a:/. This holds in M1, however, when tried while singing in M2, the singers are generally not able to vary significantly the assessed maximum SPL value for the /a:/.

5.3. Influence of unwanted signals

There are several possible sources of unwanted signals that might contaminate the voice signal and thus the measurement outcomes. These include ambient noise, voicing by persons other than the subject, room reflections, and electrical hum and noise.

Measuring the lower contour of the VRP can be quite challenging, especially in office-like environments with a certain amount of ambient noise. The influence of such noise can sometimes be reduced by resorting to A-weighting (next section). A detailed discussion of recording very soft phonation, closely connected to the VRP, is given by Šrámková *et al.* [30], who make recommendations for the choice of equipment, maximum background noise levels and level integration times. In correct conditions, healthy voices on average have a lower contour at around 45 dB or lower, with incidental softer activity. If a VRP recording system never shows any variation in the lowest thresholds, it probably has a problem with background noise.

When acquiring a VRP, it is helpful if the examiner can converse with the subject, but of course only the subject's voice should be registered by the system. Some VRP systems have a 'deaf' switch that momentarily pauses acquisition so that the examiner can talk, while other systems can automatically discriminate between the subject's voice and other sounds, using multiple microphones or other techniques.

The VRP must of course be calibrated for a given microphone distance, and the distance must be constant over the recording. With a head-mounted boom microphone, this is fairly easy to achieve. It should be noted however that any room reverberation will cause a certain amount of imprecision, as reflections from nearby surfaces may interfere and change the sound level somewhat. For running speech this is not a large concern, as the changing f_o in itself has a smoothing effect; but it can be for quasi-stationary signals such as sung vowels. For a subject standing 30 cm from the microphone on a hard floor, the floor reflection, which is usually the strongest one, will add an uncertainty of about ±1 dB to all readings. This is usually of little consequence. At 1 m microphone distance, which might be necessary for loud operatic singing, this uncertainty increases to ±2–3 dB; so absorbents on the floor are advised. Putting the microphone on a stand on a tabletop is not advisable, since reflections from the table can corrupt the spectrum.

A regularity criterion for detecting voicing, as described in Section 2.2, allows the detection even of very weak oscillations. Unfortunately, such a system becomes very sensitive also to weak extraneous signals, such as interference from electrical mains hum at 50/60 Hz, and any periodic components of the ambient noise. In practice this can be a substantial problem. A persistent low-level outlier point at the mains frequency is a sure sign of unwanted electrical hum. With a VRP system that can record the input audio signal on which the VRP is based, one can check for the presence of such unwanted background disturbances. Their effects can often be mitigated by a careful choice of microphone, microphone placement, correct cabling, and two-channel dual-gain acquisition with staggered converters.

Regardless of the criterion for voicing, the electrical noise floor of the system's audio chain needs to be on the order of 10-15 dB lower than the signal level at the phonation threshold [30]. This is readily achieved with high-end computer audio interfaces, but not with inexpensive sound cards.

5.4. A-weighting or not?

In practice, hospital and office settings often suffer from considerable but not so audible low-frequency noise from ventilation and other machinery, motivating the pragmatic use of an A-weighted frequency characteristic, as indeed is recommended by the UEP. For research purposes, when recordings can be made in a sound-proofed environment, we recommend a flat frequency response for the level measurement. The argument for using a flat response is that the VRP should ideally represent voice production and not voice perception [9]. Even if a perceptually more relevant presentation were desired, the A-weighting is not always appropriate, as argued by Hunter *et al.* [31].

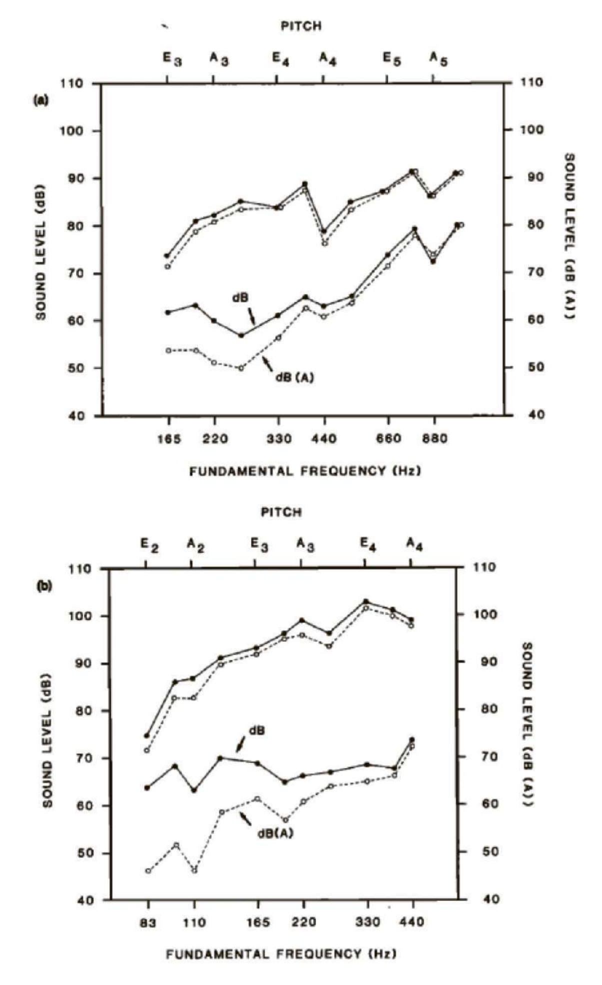

Figure 4. Effect of A-weighting, from Gramming [32]. "Two typical examples of phonetogram registrations for (a) a female and (b) a male subject phonating on the vowel /a/. The solid and dashed curves represent measurements obtained with a flat and an A-weighted frequency curve, respectively." Reproduced with permission from [32], ©1988, Acoustical Society of America.

The typical effect of the A-weighting characteristic on the shape of VRPs of female and male voices was measured by Gramming and Sundberg [32] and is shown in Figure 4. In both cases, the lower but not the upper contour is markedly lowered by the A-weighting. The reason for this discrepancy is that SPL in soft voice is carried by the lowest, fundamental partial, while in loud voice it is carried by partials at higher frequencies, which are less attenuated by the A-weighting characteristic. Strictly speaking, it is not feasible to convert accurately a VRP already computed between one form and the other, unless the spectral information in the voice signal was somehow retained. Hence it is important always to report which characteristic was used.

5.5. Recording loud voices

Making VRPs of singers can present a special technical challenge, in that the voice produced by a trained opera singer can be very loud and even exceed the 120 dB limit of the standardised VRP format. We have seen sopranos produce in excess of 130 dB at 30 cm microphone distance. For a head-mounted boom microphone, as used in theatres, the proximity to the mouth can yield sound levels that easily approach the membrane displacement limit of the microphone, which may distort and underestimate the actual sound level. This issue has been covered in detail by Švec and Granqvist [26]. Some systems use more than one microphone, automatically selecting the better signal.

6. Statistics of VRPs

Given that the VRP is multidimensional and also quite variable, it is essential to have statistical tools for computing norms, for the assessment of intervention effects (therapy or training) and for the comparison of populations (normal, different age groups, different voice disorders).

6.1. Contour statistics

In order to enable a statistical treatment of multiple VRP contours, for relating these to voice characteristics in general and to voice pathologies in particular, several methods have been proposed. The goal is to parameterise the salient features of the VRP contour into a smaller set of numbers. Thus, VRP contours have variously been analysed in terms of multiple ellipses [33], straight line segments and their slopes [25, 34], and Fourier Descriptors [35, 36]. A comparative discussion of these methods, as well as a more general Fourier method, are given in Pabon *et al.* [23]. With the latter method, it is possible to bring together data even from various sources into a common computational framework, for statistical comparison and testing, and it has the advantage that it is so generally applicable. It describes the VRP shape as a whole, independently of the size of the VRP and of the number of points with which the VRP contour was sampled. The method supports scaling and translation, and treats small and large VRPs in the same way. An example of such contour statistics is given in Figure 5.

6.2. Interior statistics

Few researchers have yet worked with data mapped over the interior of the VRP. The most straightforward method for computing averages of ensembles of VRPs is to simply accumulate data, cell by cell, from multiple conditions and/or subjects into one and the same VRP distribution. This is known as the "overlap-count" method. It has been found that the 50% median contour thus obtained usually corresponds well to the average contour obtained using the Fourier Descriptor method [23]. Figure 6 shows some examples of VRP voice quality data obtained by accumulating or averaging the within-cell contents over many VRPs. Note the greater smoothness of the colour gradients as compared to the single-subject VRPs in Figure 2.

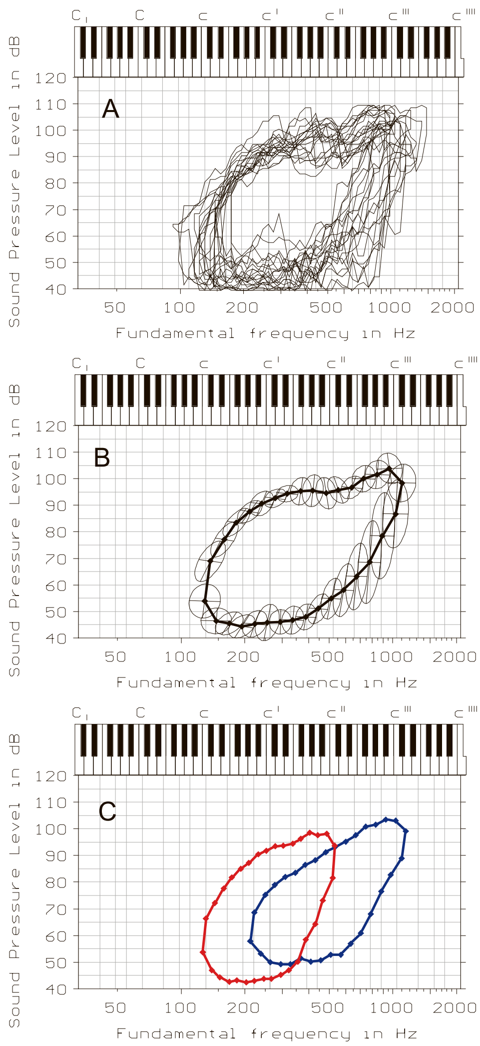

Figure 5. Example of statistical rendering of VRP contours, using the Fourier Descriptor method. A: Individual VRP contours for a group of 20 untrained female voices. B: FD-based average with ellipses showing the SD and direction of variation. C: Averaged contours with register M1 (red, left curve) and M2 (blue, right curve) separated for the same group.

7. Clinical applications

A vocally healthy person can phonate over a substantial frequency range, and can also modulate the voice in sound level to a considerable degree. From a clinical point of view, a limited range or a limited VRP area can indicate vocal dysfunction. The VRP offers a method for documentation of outcomes after interventions such as voice therapy and phonosurgery. Performing a VRP can also help in the diagnostic procedure or confirm a suspected voice disorder. In this section we present some examples of such clinical applications.

7.1. Normative data

Studies have been conducted to collect normative data from vocally healthy women, men as well as from children in different age groups. The results from these studies can be used in voice clinics as reference data to which pathological voice data can be compared. For example, normative data have been presented from 30 vocally healthy untrained Swedish males, aged 21–50 years [4], and 30 healthy untrained male and 33 female Australians aged 21–65 years [5], as well as VRP data from healthy teachers, 43 males and 46 females [35]. VRP data from children have also been collected from 35 vocally healthy boys and 39 girls, 6–11 years, to obtain reference data for those age groups [36]. Differences between male and female VRP metrics were also studied in 224 young adult untrained and trained females and males [37]. Comparisons of some of these data sets are given in [23].

When referencing a VRP data set that purports to be normative, it is important to ensure that the way in which the norm was established is appropriate and relevant for the intended application. An adequate match is needed to both the selected population (e.g, age, gender, pathology, training, genre) and to the protocol. There may also be technical issues of importance in terms of the acquisition, particularly regarding the noise floor and the lower contour of the VRP, as described above.

7.2. Diagnosis and results after treatment

A voice disorder diagnosis cannot be based solely on a VRP. Still, performing a VRP can be helpful in confirming a suspected voice disorder. The VRP results, as well as observing the patient performing the VRP, often provide valuable information. Some examples of VRPs from the clinic are shown in Figure 7, panels (a) through (f). Panel (a) depicts a VRP from a female patient after thyroid cancer surgery. Her major subjective complaint was inability to sing in her upper voice range and when she tried the voice became very strained. The VRP shows a reduced frequency range with a f_{o-max} of 494 Hz, which is very low for a female, as compared to the VRP performed by a vocally healthy woman (panel b). The VRP supported the patient's subjective complaints and shows a typical limit in the *maximum* f_o. This is an example of a diagnosis for which videolaryngostroboscopy gives too little information. Panel (c) depicts a VRP from a female-to-male transgender patient before starting testosterone hormone treatment. The *minimum* f_o dropped from 124 Hz to below 80 Hz after 12 months of treatment (d). Panels (e) and (f) show VRPs before and after phonosurgery in a patient with severe dysphonia that was due to scar tissue on the vocal folds. Before surgery, the voice was extremely unstable and the patient had difficulties for instance to produce stable soft phonation. After surgery, the VRP was smoother and less fragmented.

The examiner needs to understand how different voice problems will affect the VRP, and to take the underlying physiology and diagnosis into account when acquiring the VRP. Clearly, a patient with vocal folds suffering from a vulnerable lesion must not be prompted to vocalise too loudly. Thus, the examiner needs to be well educated in voice disorders and well-trained in the VRP protocol.

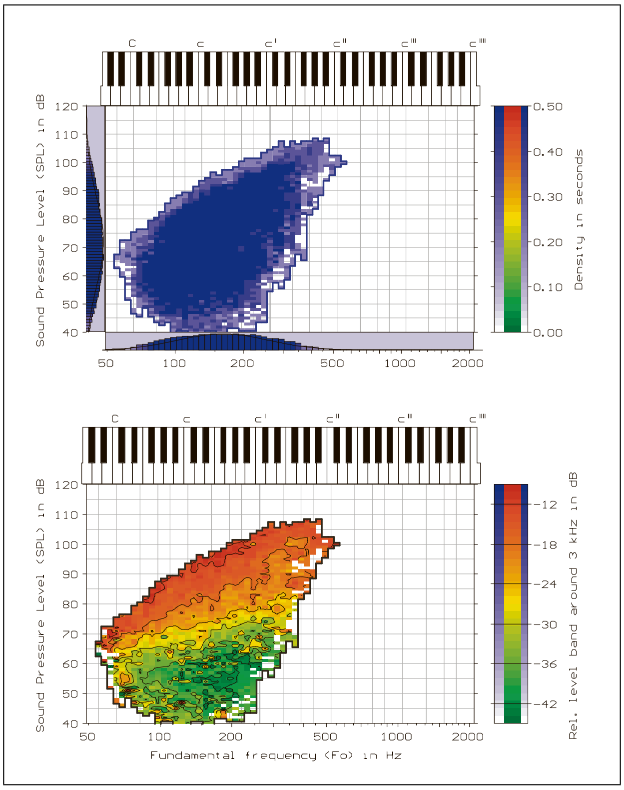

Figure 6. Example of statistical rendering of 3-D VRPs. Average VRP for an undifferentiated group of trained male voices ($N = 7$) using register M1 only. (Top) Overlap count metric, density normalized to 0.1 sec per voice. The total area covered is truncated at the point where less than 3 voices contributed to the average. (Bottom) Average relative level of the 3 kHz band (the "singer's formant" region). Normalized addition (each voice has the same weight) (colour online).

Note also that in the above examples, the speech range profile (SRP) would not necessarily have indicated a problem; rather, it was the failure of the voice to reach the normal extremes which pointed to a particular diagnosis. Indeed, SRPs have been found to be less sensitive as compared to VRPs, in discriminating dysphonic from vocally healthy voice [38]. However, SRPs are found to be useful to document speaking voice changes for transgender individuals [39].

The VRP is valuable also for investigating the efficacy of voice therapy [40, 41, 12]. Again, the interesting feature of the VRP will depend on the objectives of the treatment. With a patient who has had problems with vocal fatigue and inability to speak loudly, for example, the assessment after voice therapy would focus on changes in the upper contour of the VRP [40]. In a patient with vocal nodules, one would expect after therapy a lowering of the lower contour and an extended f_0 range, once the lesion has receded after treatment. For patients with chronic dysphonia, the level of the lower contour and the VRP area seem to be the most sensible metrics of effects of voice therapy [41]. Patients with voice disorders may present different VRP shapes depending on the diagnosis [12]. This has to be taken into account when evaluating effects of the voice treatment.

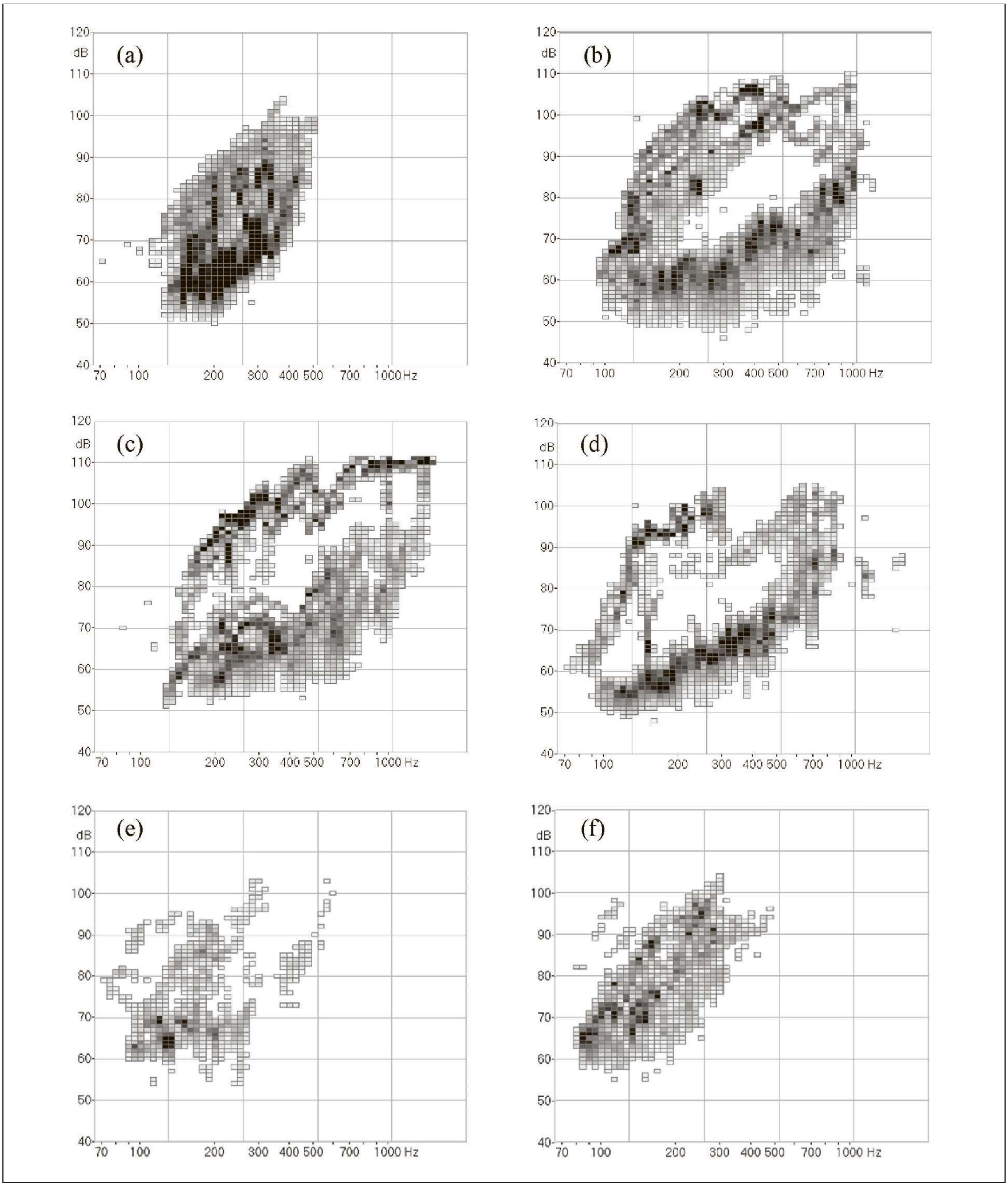

Figure 7. VRPs of clinical interest, with graphs cropped to save space. The gray scale represents the accumulated time spent in each cell, increasing to black for one second. (a) Of a female patient, post thyroid cancer surgery, with a limited maximum f_o; compared to (b) of a vocally healthy woman; (c) minimum f_o of 124 Hz in a female-to-male transgender patient before testosterone treatment and (d) minimum f_o dropped to below 80 Hz after 12 months of treatment; (e) of a male, age 70, before and (f) one year after vocal fold surgery for scar resection.

7.3. Feedback as a part of therapy

A particularly positive aspect of the VRP is the way that the real-time visual feedback captures the patient's interest, and offers an immediate reward for exercising the voice in unfamiliar ways [40]. For example, many people are not aware of the distinction between pitch and loudness, and may become fascinated with these aspects of their own vocal production. The image on the screen can be enticing and facilitates comparisons with previous efforts.

The visual feedback of f_o and SPL also allows for practical exercises of vocal behaviour such as avoiding pitch drops or creakiness at phrase endings, for speaking loudly enough, or with greater variation of prosody. It is not uncommon for patients to ask if they may borrow the system for practicing at home.

An alternative VRP mapping is possible that dispenses with the periodicity criterion for detecting phonation; instead, it reports an f_o derived from single-pulse-durations against SPL. This has proven to be extremely useful for clinical therapy, as it reports the 'islands' where a very disturbed voice still exhibits some form of periodic behaviour. Therapists can use this visualisation to guide patients into regions of stable phonation.

7.4. The VRP/SRP in voice dosimetry

With the large amounts of data that are acquired over a working day or week, the SRP can provide a compact summarising display of average voice use. Speaking in noisy environments, for example, tends to skew the distributions of f_o and SPL upwards, resulting in both a shift and a shape change of the speech range profile. Voice pathologies, too, can in some cases manifest themselves in distribution shapes: for instance, using ambulatory devices and a large data base, Ghassemi *et al.* [42] found that skewing of f_o and SPL distributions over five-minute time windows correlated with the prevalence of vocal fold nodules. While some of the few commercially available voice dosimetry systems do offer a VRP display format, we are not aware of any major research findings so far based on this perspective; but its potential is worth mentioning.

8. Pedagogical applications

Several studies on vocal training (e.g., [43, 44, 45]) have examined whether the VRP might be a useful indicator of voice training outcomes. An overview of these and more studies can be found in [45]. In particular most, but not all of them, find a small extension of the lower contour and/or an increase of the VRP area, as an effect of different types and durations of voice training. Even very short-term effects have been documented, e.g., increased vocal range after one session of voice 'warm-up' exercises [46] for speech-language pathologists in training. The VRP has been used also to research aspects of singing voice of interest primarily to pedagogues, such as how the transition between M1 and M2 is managed by resorting to so-called *voix mixte* [28].

Although a large voice range on the face of it would seem to be a basic indicator of vocal skill, this is not necessarily so. The VRP assesses range, but normally has little to say on the *quality* of the tone produced. For a singer, training involves developing a fine control over the voice within specific domains of the voice type or musical genre, be it bass or soprano, hard rock or chinese opera. Therefore it would be inappropriate to score the outcome of a training programme on the basis of VRP area metrics only.

In a longitudinal study on classical soprano students [45], the VRP area was actually smaller at the end of a three-year programme than at admission, but thanks to a consistent pedagogical methodology, the development of the students' annual VRP outcomes could be attributed to the progression of the study programme. It was also found that, out of the over one hundred students of singing recruited for the study, only ten were tenacious and similar enough, in terms of voice category, genre and training programme, to allow for meaningful VRP statistics to assess the effects of the training.

9. Research applications

Several researchers are using the reference plane of SPL vs. f_o for exploring various voice metrics and how they change over the voice range. A small sample is given below. A rationale for mapping the voice in the context of the VRP, and especially its interior, is given in the Discussion.

9.1. Mapping the voice spectrum over the VRP

The spectrum of the voice changes considerably over the voice range, and a detailed map of such changes could inform us in greater detail on how the voice works. By computing narrow-band spectra at a fixed frame rate of about 40 Hz, and averaging these per cell in the VRP, it is possible to visualise how the voice spectrum changes, on average, over both fo and SPL. From the cell-averaged spectra, scalar metrics can be derived (Figure 2, d–f) and trends for features of the spectrum envelope can be discerned. This mode of analysis opens up a whole new domain of research topics, which we believe will be instrumental in charting the voice production mechanism; unfortunately; it far exceeds the scope of the present paper. However, while exploring this domain, we have seen that the role of many voice metrics tends to change from one part of the voice range to another, as the character of the vocal fold vibration changes, posing additional challenges to their interpretation. Several studies are in progress along these lines.

9.2. Non-acoustic metrics in the VRP

Non-acoustic real-time metrics can also be visualised over the f_o-SPL plane, if they are acquired in parallel with the voice signal. Lamesch [27, 28, 47] acquired the electroglottographic signal, and mapped its amplitude and contact quotient into the VRP, for making M1/M2 register comparisons. He did the same also for the vertical larynx position. Lamarche *et al.* [15] replaced the microphone with a neck-mounted accelerometer, which eliminated the dependency of level on vowel, and also cancelled the upward slope toward high f_o. A further possibility is to create a record of the *subjective sensations* of subjects as they are producing their VRP. In the quest for pinpointing the often subtle voice problems of trained singers, Lamarche *et* *al.* [45] provided them with a hand-held button. They were instructed to press the button to indicate incidents of vocal discomfort or reduced vocal control. Pressing the button would make a mark at the corresponding point in the VRP, for later analysis and discussion.

9.3. Vocal registers as 'vibratory states'

Current VRP implementations, by design, discard the time history of the portrayed phonations. When working with the VRP and its real-time visual feedback, however, it becomes clear that phonation often switches between distinctly different types, sometimes with a degree of hysteresis. In other words, the character of vocal fold vibration at any instant is often determined not only by the instantanenous f_o and/or SPL, but also by the phonation at the preceding instant. The classic example is the voice break or transition between the voice registers modal/chest/M1 and falsetto/head/M2 [8]. Register transitions can occur at different values of f_o and/or SPL depending on where the subject's voice is coming from. Hence the conventional VRP display actually renders a superposition of separate phonatory regimes. We might prefer to display these separately, as done e.g. by Lamesch [47] by instructing subjects to remain within one mechanism at a time. But this is a difficult task and also assumes that the mechanisms are known beforehand.

The different ways in which the vocal folds can vibrate might be called 'vibratory states'. The M1 and M2 mechanisms are the primary instances, but experience using the VRP has led us to believe that there may well exist smaller subregions in the VRP of more subtly different vibratory behaviour. Since the transitions between such states are likely to exhibit hysteresis, it would be particularly interesting to have a VRP system that could automatically track aspects of vocal fold vibration as directly as possible. In an effort to identify such regions, Selamtzis and Ternström [48] used relative metrics derived from spectral components of the electroglottographic signal, to cluster automatically different phonation types across the VRP. For a small number of subjects, the method could discriminate between M1 and M2 phonation, without using as input any of prior knowledge, f_o, SPL or EGG amplitude. The objective of this ongoing work is to create a state-aware VRP that can discriminate between different modes of phonation, such as different voice registers, and assign them to separate overlaid planes.

10. Discussion of VRP modeling

The objective of the VRP is appealing and easy to explain to the person being recorded. To assess voice status by bringing the voice to its extremes seems sensible, and the subject's compliance is easy to obtain. The accepted premise is that, to reach an extreme, all factors and conditions that control the phonation should be in optimal balance, hence detrimental factors will be exposed, while also the physicality of the limiting factors becomes evident. It would thus seem logical that at the extremes, the cause-and-effect relationships would straighten out; and that specific VRP contour characteristics could be straightforwardly connected to specific physiological features of the voice. It is tempting to adopt an immediate physiological interpretation of our measurements, especially at the time of the actual VRP recording, when everything appears to be logically connected. We must be cautious, though – things may not be that simple.

10.1. Inversion problem

In Section 7.2, it was demonstrated how the effects of voice treatment may appear in the VRP; and the physiologic interpretations of the observed VRP characteristics were very plausible. It would not be difficult to extend this paper with more clinical examples or singing voice-training examples [45] that in themselves are similarly logical, precise and convincing. But with each example, how much of the explanation would survive, if there were only the VRP contour for us to interpret? There is an inversion problem, in that the VRP contour is determined by many factors. In general, it is not possible to link definitively a physiological interpretation to a specific VRP contour change. Except for the register-break dip in the upper contour (section 3.4), we are not in a position to assert that any single aspect of voice physiology or particular voice pathology can reliably be linked to a given VRP contour characteristic. Before summoning the reinforcing troops of additional quality metrics, in the quest for supplementary information to secure an interpretation, it is salutary to realise that this would not necessarily solve the inversion problem.

To elucidate this, first an implicit reasoning needs to be made explicit. When interpreting or evaluating a VRP contour we rely on the following paradigm:

> *For a given f_o, all voices comply with a comparable bounding mechanism or limiting factor, such that a change in maximum or minimum SPL can directly be associated with a unique positive or negative influence on the assessed voice production mechanism.*

Although this sounds like a very unrefined model, it is the assumption that is made when, for instance, an averaged VRP contour is seen as normative for a voice.

10.2. Diversity in the limiting mechanism

The validity of the above implicit paradigm has, in a sense, already been tested, when the statistics were done to define the scales of the dysphonia severity index [49]. The DSI is designed as a quantitative correlate of the perceived voice quality. It uses a weighted combination of maximum phonation time, a jitter value and the two VRP metrics f_{o-max} (in Hz) and L_{min}. The data reduction that the DSI offers implies that the information content of the VRP contour is highly limited, as other contour aspects seem not to make any difference, and their contribution is thus not needed. This notion, however, is contradicted by many examples, such as an effect of clinical therapy seeming patently evident in, say, L_{max}. Also, individuals commonly present very distinct VRP contours that are reproducible and unique, so there must be a definite, if as yet unexplained physiological base for the contour. The DSI may point us to a way out of this dilemma. If only $L_{\min}$ and linear $f_{o-\max}$ survive, this could be seen as a sign that only at those extreme settings does the concept of a comparable bounding mechanism or limiting factor hold. In the large and diverse clinical database underlying the DSI there could have been VRP contour changes that were meaningful on an individual basis, but if they lacked generality or a common direction, they would not be resolved. As has been eloquently argued by Baken and Orlikoff [50], dysphonia is a multidimensional phenomenon, and the voice production mechanism may deteriorate in several different directions. The vibratory mechanism can become caught in very different dependencies or states (see 9.3), where dissimilar factors limit the associated VRP areas. In principle, this leads to the situation that in a clinical setting it will be mostly the given voice problem's compliance to the chosen VRP paradigm that is tested. Due to lack of uniformity in the bounding mechanism, the causation is invalidated, and any scaling of VRP deviance becomes arbitrary. The VRP contour variation seen with healthy voices could thus be interpreted as resulting from a different scaling within in a multidimensional space of possible bounding factors. Along these lines, the VRP landmark of the NW corner mentioned above could be interpreted as a region where all voices, despite their mutual physiological differences, tend to defer to the same limiting factor(s).

10.3. Compliance to the same bounding mechanism

Lamesch *et al.* [27] demonstrated that in register M2 the vowel has no influence on $L_{\max}$, while in register M1 it has. This means that for M1 and M2 the limiting factors are known to be different. As long as there is no evidence to what extent M1 or M2 controls the overall VRP contour, we will have no valid criterion against which to rate a local contour change. If voices do not comply with the same limiting mechanism, there is no sense in comparing their VRP contours. This statement could be further generalised: if the control factors that scale the voice production mechanism are not comparable between voices, there is no point in averaging the VRP interior data either, as such an average will have no normative value. The question of how to interpret the VRP is essentially a question of identifying this scaling of the voice production mechanism. A further evaluation of the contour becomes meaningful only when the bounding mechanisms are effectively identified. It is this question of the underlying mechanisms that prompts the further study of data from the VRP interior.

10.4. A step out of the box

It can safely be assumed that with every tone of the voice in normal speech there will be a specific associated vibratory movement of the vocal folds. The prevalent stance is that this vocal fold vibration is directly causing and also limiting the resulting sound. The VRP contour would then be very closely linked to the physiology of the vocal folds,

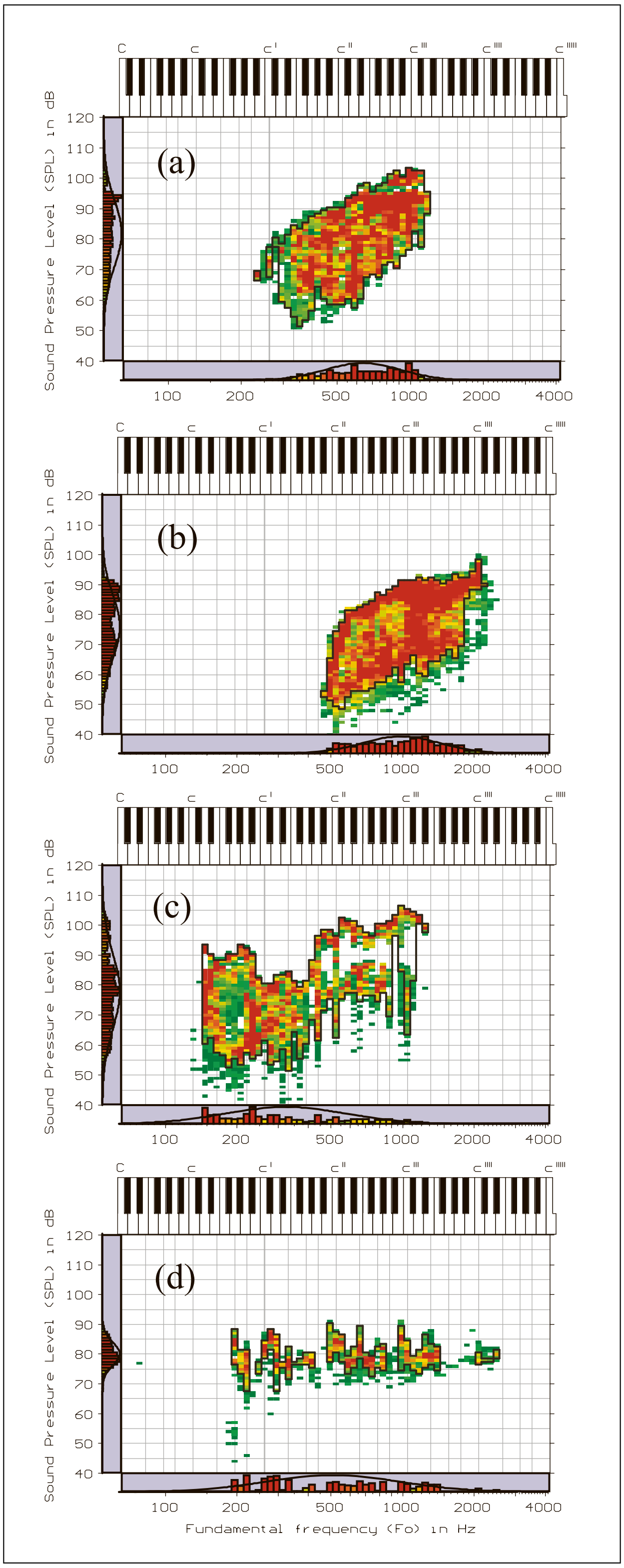

Figure 8. Range profiles from different instruments: (a) The isolated falsetto/head register of an untrained female voice, (b) a mouth whistle, (c) a clarinet in B flat, and (d) a violin (colour online).

where factors like mass, tension and stiffness have their known effects on the vibration. However, there could be more complex interactions going on. To illustrate, Figure 8 shows four example VRP's from different "instruments".

If pitch-dependant changes in SPL offset are ignored, then the three breath-driven instruments (a, b and c in Figure 8) show over their range a very comparable SPL span from $L_{\min}$ to $L_{\max}$. The violin VRP, on the other hand, shows a very limited SPL variation (Figure 8d). The violin does not follow the trend, presumably because it is bow-driven and not breath-driven; the non-linearities in the tone production are of a different nature. Note that the archetypical oval VRP shape appears not only for the voice (register), but can be clearly recognized in the VRP also of the mouth whistle (Figure 8b). With this simple mouth-whistle there are no moving parts, yet there is still a clear nick at the bottom of the pitch range where the upper VRP contour drops abruptly. Also, its $L_{\max}$ marks a saturation point: applying more pressure will not raise the sound level any further. Just as with the voice, in whistling too there are difficulties with producing the soft high notes. Although it is usually assumed that the mechanical characteristics of the vocal folds determine the overall VRP shape, these plots suggest that perhaps the bulk of the VRP shape is largely controlled and bounded by flow phenomena, especially in head/falsetto register. A model that refers only to the mechanics of the vocal folds, or that interprets cause-and-effect relationships based solely on linear source-filter theory to explain the shape of the VRP will probably not be sufficient.

Still, it is essential to model the VRP shape, as only with a model can an interpretation be given to measurements. When reality is too complex for our models, then the realistic choice is to use a statistical model and to rely on statistics for our decision-making. When there is no additional information that accurately informs on the actual mechanism, then an abstract statistical model must be our temporary resort, while we continue trying to untangle the multiplicity of voice production mechanisms.

10.5. Exploiting the VRP interior

A plane holds more information than a line, so, naturally, the way that a metric changes over the interior of the VRP is usually more informative than just the location of the perimeter. As one of countless examples, let us consider the typical behaviour of the jitter metric in approaching the lower contour at high pitches. In this region of the VRP, aspirative noise, typically indicative of insufficient glottal closure, is the dominant cause of jitter. Here, it is not the lower contour that will give an early indication of a voice problem, but rather the jitter. An estimate of the degree of noise due to insufficient glottal closure might be a primary goal when the voice status of a group of female school teachers is to be assessed. However, even when assessment of the contour is secondary, the VRP framework and interactive recording concept are invaluable for organising the voice sampling. Jitter values are no longer seen through the keyhole, as with a single sustained vowel, but become part of a dynamic progression or pattern, from which it can be judged whether the measured values fit a trend or are to be considered deviant. Their location in the VRP provides a link to an underlying mechanism. The “hard” coordinate system of the VRP can bring reproducibility to an acoustic metric, because its variance becomes organised.

The central value of the VRP paradigm is simply that it offers a common frame of reference for a multitude of voice measurements. It allows us to superimpose data from several modalities, and to visually infer relationships that might otherwise be difficult to unravel. In particular, it allows us chart the extent of different modes of phonation or dysphonation. One tends to talk about the VRP as though it were one type of measurement, whereas in fact it can be seen as a context for visualising and relating any number of voice metrics to each other.

In planning this article, we had hoped finally to progress from the ‘legacy’ VRP contour, and to concentrate more on the potential of mapping the VRP interior. However, we found that while the literature on the VRP contour is extensive, the number of publications on its interior is as yet rather small. We can only encourage the voice research community to explore this analysis paradigm that for all its challenges also holds a lot of promise.

Acknowledgement

We gratefully acknowledge the financial support to VRP-related research over many years from several sources, namely: the Swedish Research Council (Vetenskapsrådet), contract 2010-4565; of the Canadian Fondation Baxter & Alma Ricard; and of the former Swedish Council for Working Life and Social Research (FAS), contracts 2001-0341 and 2002-0416. We are grateful also for the support of The Royal Conservatoire, The Hague, Netherlands. Eva B. Holmberg and Andreas Selamtzis offered helpful comments on the manuscript.

References

[1] I. R. Titze, R. Baken, K. Bozeman, S. Granqvist, N. Henrich, C. Herbst, D. Howard, E. Hunter, D. Kaelin, R. Kent, J. Kreiman, M. Kob, A. Löfqvist, S. McCoy, D. Miller, H. Noé, R. Scherer, J. Smith, B. Story, J. Švec, S. Ternström, J. Wolfe: Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. J. Acoust. Soc. Am. **137** (2015) 3005–3007.

[2] P. Gramming: Vocal loudness and frequency capabilities of the voice. J. Voice **5** (1991) 144–157.

[3] H. K. Schutte, W. Seidner: Recommendation by the Union of European Phoniatricians (UEP): Standardizing voice area measurement / phonetography. Folia Phoniatr. **35** (1983) 286–288.

[4] A. E. Hallin, K. Fröst, E. B. Holmberg, M. Södersten: Voice and speech range profiles and voice handicap index for males - methodological issues and data. Logoped. Phoniatr. Vocol. **37** (2012) 47–61.

[5] K. Sanchez, J. Oates, G. Dacakis, E. B. Holmberg: Speech and voice range profiles of adults with untrained normal voices: methodological implications. Logoped. Phoniatr. Vocol. **39** (2014) 62–71.

[6] W. Seidner: Objektive Qualitätsbeurteilung der Stimme mittels Dynamikmessungen. Z. Klin. Med. **40** (1985) Heft 20.

[7] T. Hacki: Comparative speaking, shouting and singing voice range profile measurement: physiological and pathological aspects. Logoped. Phoniatr. Vocol. **21** (1996) 123–129.

[8] Lamarche, S. Ternström, P. Pabon: The singer's voice range profile: Female professional opera soloists. J. Voice **24** (2008) 410–426.

[9] P. Pabon, R. Plomp: Automatic phonetogram recording supplemented with acoustical voice-quality parameters. J. Speech Hear. Res. **31** (1988) 710–722.

[10] J. P. H. Pabon: Objective acoustic voice-quality parameters in the computer phonetogram. J. Voice **5** (1991) 203–216.

[11] R. F. Coleman: Sources of variation in phonetograms. J. Voice **7** (1993) 1–14.

[12] R. Speyer, G. H. Wieneke, I. Warnaar, P. H. Dejonckere: Effects of voice therapy on the voice range profiles of dysphonic patients. J. Voice **17** (2003) 544–556.

[13] S. K. Wolfe, D. Stanley, W. J. Sette: Quantitative studies on the singing voice. J. Acoust. Soc. Am. **6** (1935) 255–266.

[14] A. Lamarche: Putting the singing voice on the map: Towards improving the quantitative evaluation of voice status in professional female singers. Doctoral thesis, KTH Royal Institute of Technology, Stockholm. www.diva-portal.org/smash/get/diva2:173670/FULLTEXT01.pdf, 2009.

[15] A. Lamarche, S. Ternström: An exploration of skin acceleration level as a measure of phonatory function in singing. J. Voice **22** (2008) 10–22.

[16] I. R. Titze, J. Sundberg: Vocal intensity in speakers and singers. J. Acoust. Soc. Am. **91** (1992) 2936–2946.

[17] S. Ternström: Does the acoustic waveform mirror the voice? Logoped. Phoniatr. Vocol. **30** (2005) 100–107.

[18] D. A. Berry, H. Herzel, I. R. Titze, B. H. Story: Bifurcations in excised larynx experiments. J. Voice **10** (1995) 129–138.

[19] I. R. Titze: Phonation threshold pressure: A missing link in glottal aerodynamics. J. Acoust. Soc. Am. **91** (1992) 2926–2935.

[20] S. Ternström, M. Bohman, M. Södersten: Loud speech over noise: Some spectral attributes, with gender differences. J. Acoust. Soc. Am. **119** (2005) 1648–1665.

[21] B. Roubeau, N. Henrich, M. Castellengo: Laryngeal vibratory mechanisms: the notion of vocal register revisited. J. Voice **23** (2009) 425–438.

[22] I. R. Titze: Principles of voice production. Prentice Hall, New Jersey, 1994.

[23] P. Pabon, S. Ternström, A. Lamarche: Fourier descriptor analysis and unification of voice range profile contours: method and applications. J. Speech, Lang. Hear. Res. **54** (2011) 755–776.

[24] kc.koncon.nl/staff/pabon/VRP/VPusersGroupStandardProtocolApril2012PPabon.pdf.

[25] A. M. Sulter, H. P. Wit, H. K. Schutte, D. G. Miller: A structured approach to voice range profile (phonetogram) analysis. J. Speech and Hearing Res. **37** (1994) 1076–1085.

[26] J. G. Švec, S. Granqvist: Guidelines for selecting microphones for human voice production research. Am. J. Speech-Language Pathol. **19** (2010) 356–368.

[27] S. Lamesch, B. Doval, M. Castellengo: Towards a more informative voice range profile: the role of laryngeal vibratory mechanisms on vowels dynamic range. J. Voice **26** (2012) 672.e9–672.e18.

[28] S. Lamesch, R. Expert, M. Castellengo, N. Henrich, B. Chuberre: Investigating voix mixte: A scientific challenge towards a renewed vocal pedagogy. 3rd Conference on Interdiciplinary Musicology, CIM07, Tallinn, Estonia, Aug 2007, http://www–gewi.uni–graz.at/cim07.

[29] W. Seidner, H. Krüger, K.-D. Wernecke: Numerische Auswertung spektraler Stimmfelder [Numerical evaluation of the spectral voice areas]. Sprache - Stimme - Gehör **9** (1985) 10–13.

[30] H. Šrámková, S. Granqvist, C. T. Herbst, J. G. Švec: The softest sound levels of the human voice in normal subjects. J. Acoust. Soc. Am. **137** (2015) 407–418.

[31] E. J. Hunter, J. G. Švec, I. R. Titze: Comparison of the produced and perceived voice range profiles in untrained and trained classical singers. J. Voice **20** (2005) 513–526.

[32] P. Gramming, J. Sundberg: Spectrum factors relevant to phonetogram measurement. J. Acoust. Soc. Am. **83** (1988) 2352–2360.

[33] R. Airainer, F. Klingholz: Quantitative evaluation of phonetograms in the case of functional dysphonia. J. Voice **7** (1993) 136–141.

[34] L. Heylen, F. L. Wuyts, F. Mertens: Evaluation of the vocal performance of children using a voice range profile index. J. Speech, Lang. Hear. Res. **41** (1998) 232–238.

[35] L. Heylen, F. L. Wuyts, F. Mertens, M. de Bodt, P. H. van de Heyning: Normative voice range profiles of male and female professional voice users. J. Voice **16** (2002) 1–7.

[36] F. L. Wuyts, L. Heylen, F. Mertens, M. de Bodt, P. H. van de Heyning: Normative voice range profiles of untrained boys and girls. J. Voice **16** (2002) 460–465.

[37] A. M. Sulter, H. K. Schutte, D. G. Miller: Differences in phonetogram features between male and female subjects with and without vocal training. J. Voice **4** (1995) 363–377.

[38] E. Ma, J. Robertson, C. Radford, S. Vagne, R. El-Halabi, E. Yiu: Reliability of speaking and maximum voice range measures in screening for dysphonia. J. Voice **21** (2007) 397–406.

[39] E. B. Holmberg, J. Oates, G. Dacakis, C. Grant: Phonetograms, aerodynamic measurements, self-evaluations, and auditory perceptual ratings of male-to-female transsexual voice. J. Voice **24** (2010) 511–522.

[40] E. B. Holmberg, E. Ihre, M. Södersten: Phonetograms as a tool in the voice clinic: changes across voice therapy for patients with vocal fatigue. Logoped. Phoniatr. Vocol. **32** (2007) 113–127.

[41] P. H. Dejonckere, I. van Wijck, R. Speyer: Efficacy of voice therapy assessed with the voice range profile (phonetogram). Rev. Laryngol. Otol. Rhinol. (Bord) **124** (2003) 285–289.

[42] M. Ghassemi, J. H. Van Stan, D. D. Mehta, M. Zañartu, H. A. Cheyne, R. E. Hillman, J. V. Guttag: Learning to detect vocal hyperfunction from ambulatory neck-surface acceleration features: Initial results for vocal fold nodules. IEEE Trans. Biomed. Eng. **61** (2014) 1668–1675.

[43] N. Siupsinskiene, H. Lycke: Effects of vocal training on singing and speaking voice characteristics in vocally healthy adults and children based on choral and nonchoral data. J. Voice **25** (2011) 177–189.

[44] W. D. LeBorgne, B. D. Weinrich: Phonetogram changes for trained singers over a nine-month period of vocal training. J. Voice **16** (2002) 37–43.

[45] P. Pabon, R. Stallinga, M. Södersten, S. Ternström: Effects on vocal range and voice quality of singing voice training; the classically trained female voice. J. Voice **28** (2014) 36–51.

[46] K. M. van Lierde, E. D'haeseleer, N. Baudonck, S. Claeys, M. de Bodt, M. Behlau: The impact of vocal warm-up exercises on the objective vocal quality in female students training to be speech language pathologists. J. Voice **25** (2011) e115–e121.

[47] S. Lamesch: Mécanismes laryngés et voyelles en voix chantée. Dynamique vocale, phonétogrammes de paramètres glottiques et spectraux, transitions de mécanismes [Laryngeal mechanisms and vowels in singing, viewed with respect to voice range profiles, glottal and spectral parameters, and laryngeal mechanism transition]. Doctoral thesis, Université Pierre et Marie Curie, Paris, France. www.lam.jussieu.fr/Membres/Lamesch/Publications.html, 2010.

[48] A. Selamtzis, S. Ternström: Analysis of vibratory states in phonation using spectral features of the electroglottographic signal. J. Acoust. Soc. Am. **136** (2014) 2773–2783.

[49] F. L. Wuyts, M. S. de Bodt, G. Molenberghs, M. Remacle, L. Heylen, B. Milet, K. van Lierde, J. Raes, P. H. van de Heyning: The dysphonia severity index: An objective measure of vocal quality based on a multiparameter approach. J. Speech, Lang. Hear. Res. **43** (2000) 796–809.

[50] R. J. Baken, R. F. Orlikoff: Towards a dynamical diagnosis of vocal function. – In: Diagnosis and Treatment of Voice Disorders. J. S. Rubin, R. T. Sataloff, G. S. Korovin (eds.). Plural Publishing, San Diego, California, 2014, 165–176.