Dependencies and Ill-designed Parameters Within High-speed Videoendoscopy and Acoustic Signal Analysis

*Patrick Schlegel, †Michael Stingl, ‡Melda Kunduk, *Stefan Kniesburges, §Christopher Bohr, and *Michael Döllinger, *†*Erlangen, and* §*Regensburg, Germany, and* ‡*Baton Rouge, Louisiana*

Summary: Objective. The phonatory process is often judged during sustained phonation by analyzing the acoustic voice signal and the vocal fold vibrations. Many formulas and parameters have been suggested for qualifying the characteristics of the acoustic signal and the vocal fold vibrations during sustained phonation. These parameters are directly computed from the acoustic signal and the endoscopic glottal area waveform (GAW). The GAW is calculated from laryngeal high-speed videoendoscopy (HSV) recordings and describes the increase and decrease of the glottal area during the phonation process, that is, the opening and closing of the two oscillating vocal folds over time. However, some of the parameters have strong mathematical dependencies with one another and some are ill-defined. The purpose of this study is to identify mathematical dependencies between parameters with the aim of reducing their numbers and suggesting which parameters may best describe the properties of the GAW and the acoustical signal.

Methods. In this preliminary investigation, 20 frequently used parameters are examined: 10 GAW only and 10 both GAW and acoustic parameters.

Results. In total 13 parameters can be neglected because of mathematical dependencies. In addition, nine of these parameters show problematic features that range from unexpected behavior to ill definition.

Conclusions. Reducing the number of parameters appears to be necessary to standardize vocal fold function analysis. This may lead to better comparability of research results from different studies.

Key Words: High speed video endoscopy–Glottal area waveform–Parameters–Mathematical dependencies–Ill design.

INTRODUCTION

Phonation is a complex process, during which the vocal tract and the vocal fold vibratory patterns play a major role. According to the source-filter theory, the tongue and lips modulate the sound while the vocal folds are the vibratory source. During the periodic oscillation of the vocal folds, the area between the vocal folds, called the "glottal area," changes, resulting in periodic interruption of the airflow traveling through the trachea. This leads to a periodic series of flow pulses that produces audible sound during phonation.[1,2]

One powerful tool for the examination of vocal fold vibrations is high-speed videoendoscopy (HSV). As shown in Figure 1, using HSV, the movements of the vocal folds are directly recorded by a high-speed camera at speeds several times faster than the fundamental frequency of the vibrating vocal folds. Further, Figure 1 illustrates the oscillation cycle of the vocal folds. The movies recorded with HSV allowed determination of within-glottic cycle vibratory patterns at the source. Obtained HSV data can be quantified as the function of glottal area over time, that is, the glottal area waveform (GAW).[4–6]

Different definitions of the GAW exist.[6–9] In this work, the GAW is defined as the simple function of the glottal area in pixels over time. Other definitions of the GAW and their possible effects on the analysis will be outlined in the discussion section.

In addition to HSV, other techniques for visualizing the vocal fold vibrations exist, such as videostroboscopy[10,11] and videokymography (VKG).[12,13] Their limitations compared with HSV are already reported in previous publications.[4,14]

HSV currently reaches recording rates of 4 kHz in clinical applications and up to 20 kHz in research applications,[15] which is distinctly higher than the frequency range of the vocal folds, reaching up to 1568 Hz in soprano singing.[16] In general, fundamental frequency ranges are between 70 and 500 Hz in men and between 130 and 1000 Hz in women.[12]

Overall HSV is a superset of both, stroboscopy and VKG, which also means that from HSV data artificial stroboscopy and VKG data can be produced, called "simulated stroboscopy" and "digital kyomography." This shows the superiority of HSV in assessment of vocal fold vibratory patterns.[4,14]

However, even with HSV as an effective tool for evaluation of vocal fold function, there are still challenges to be addressed, predominantly involving the quantitative analysis of the HSV data. One of these challenges is making sense of the clinical value of a great number of HSV parameters and their redundancy in the assessment of vocal function. Whereas in some cases these parameters are exclusively connected to the GAW, such as the Open Quotient, other parameters are also used for acoustic signals, such as mean Shimmer.[17,18] In other cases, some of the parameters show high intra-subject variability[19] or can be problematic even just from their definition, such as Shimmer(%), as will be shown later in this work.

Accepted for publication April 19, 2018.
Conflict of interest: The authors declare that there is no conflict of interest regarding the publication of this paper.
From the *FAU Erlangen-Nürnberg, Medical School, Division of Phoniatrics and Pediatric Audiology at the Department of Otorhinolaryngology Head & Neck Surgery, University Hospital Erlangen, Erlangen, Germany; †Department Mathematics, Applied Mathematics II, FAU Erlangen-Nürnberg, Erlangen, Germany; ‡LSU Speech Language Hearing Clinic, Department of Communication Sciences & Disorders, Louisiana State University, Baton Rouge, Louisiana; and the §FAU Regensburg, Clinic and Polyclinic for Otolaryngology, University Hospital Regensburg, Regensburg, Germany.
Address correspondence and reprint requests to: E-mail: Patrick.Schlegel@uk-erlangen.de
Journal of Voice, Vol. 33, No. 5, pp. 811.e1–811.e12
0892-1997

https://doi.org/10.1016/j.jvoice.2018.04.011

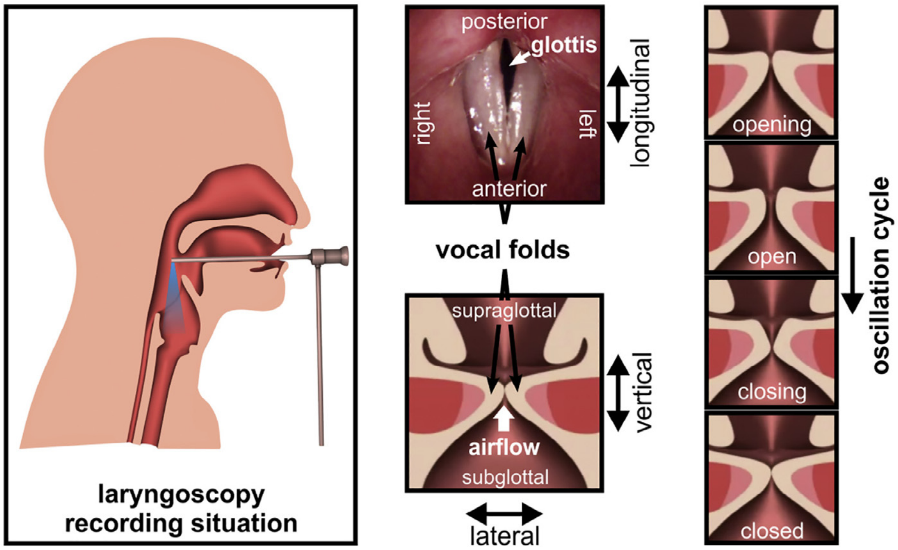

FIGURE 1. Schematic illustration of vocal fold recording by employing an endoscopic high-speed camera (left), an image of the vocal folds (center), and a display of one oscillation cycle (right).[3]

In this work some of the stronger mathematical connections between 20 different parameters or derivatives will be shown. A "strong mathematical dependency" is defined by either of the following criteria:

- within a group of parameters, all of them can be calculated if just one or some are known;
- the parameters are derivatives of another; for example, the normalized version.

In other words, in both cases, there is no gain in information by calculating a strongly mathematically dependent parameter, if the others are known. Hence, this work aims to:

1. determine mathematical dependencies between parameters;
2. find ill-defined parameters, for example, parameters being not defined within a meaningful interval;
3. suggest which of the parameters can be kept and which should be discarded.

METHODS

In this preliminary investigation, 20 frequently used parameters are examined: 10 GAW only and 10 both GAW and acoustic parameters. Because 10 of the parameter formulas are also used for other types of signals (eg, acoustic or time-resolved pressure signals), they are not exclusively devoted to the assignment of GAW parameters.[17,18] Different weaker and stronger correlations between parameters exist. However, in this work we focus only on strong mathematical dependencies. The parameters will be subdivided into thematically connected subsections. For each subsection, the purpose of the parameters will be explained briefly, and subsequently, how to calculate one parameter out of the others will be shown, and characteristics of different formulas will be revealed. Finally, a rationale will be given for the choice of which related parameters should be kept for analyses of vocal fold function.

GAW phase characteristics

This group of parameters is used to measure ratios between different phases of increasing and decreasing glottal area. Figure 2 shows a schematic illustration for a single GAW cycle i, beginning with the opening of the vocal folds and ending with a short interval during which the vocal folds stay closed. Any such cycle can be subdivided into different phases:

- the opening phase ($[C \rightarrow O]_i$), during which the glottal area is increasing and the vocal folds are opening;
- the closing phase ($[O \rightarrow C]_i$), during which the glottal area is decreasing and the vocal folds are closing;
- the open phase duration (t^i_{open}), which is the sum of the two previous phases;
- the closed phase duration (t^i_{closed}), during which the glottis is closed and the glottal area therefore is zero;
- the sum of t^i_{open} and t^i_{closed} is the duration of the entire cycle (T_i).

In Figure 2, below the GAW diagram, four images of male vocal folds in different stages of the glottal cycle from opening to closed phase are shown. For measuring different ratios between these distinct phases of the GAW, six different parameters are discussed:

- The **Open Quotient**,[21] defined as

$$Oq_i = \frac{t^i_{open}}{T_i}$$

- The **Closing Quotient**,[22] defined as

$$Cq_i = \frac{[O \rightarrow C]_i}{T_i}.$$

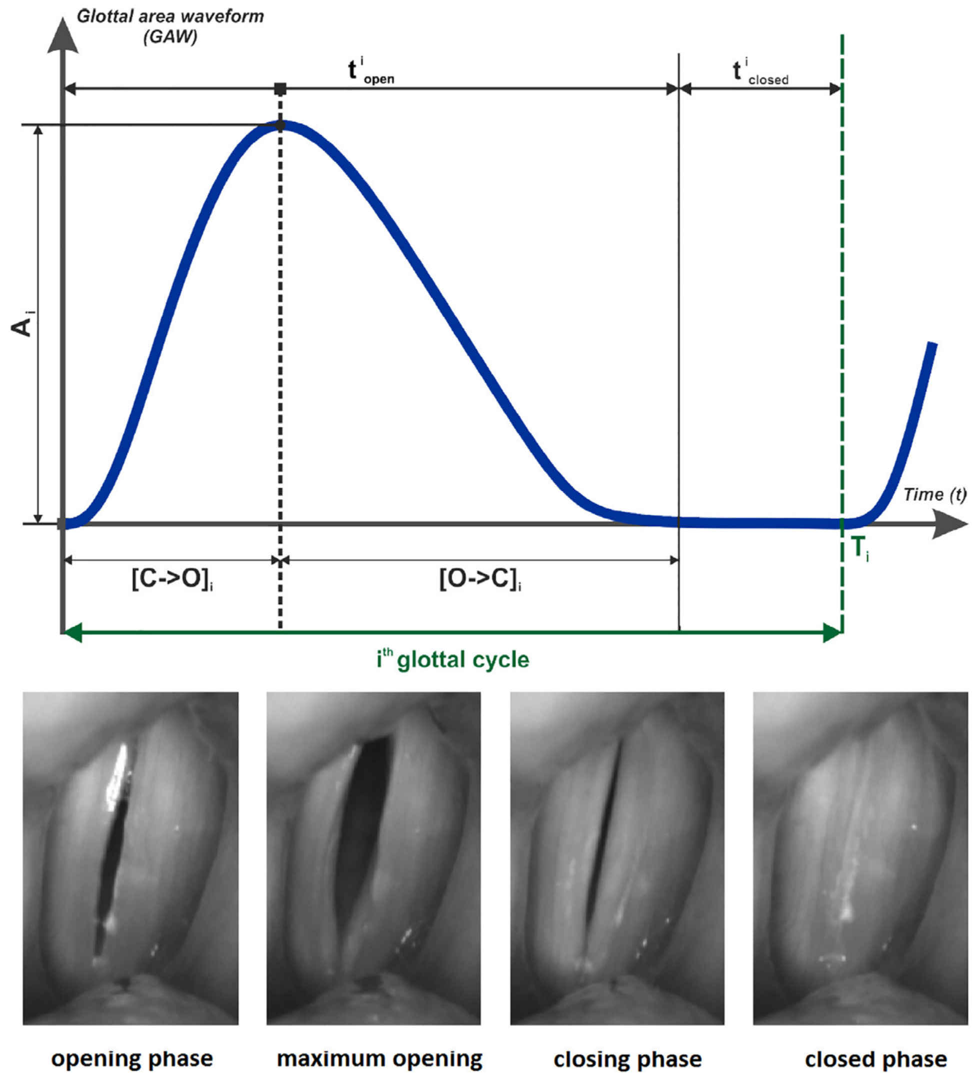

FIGURE 2. Subdivision of one GAW cycle (T_i) into different time intervals, the opening phase ($[C \rightarrow O]_i$), the closing phase ($[O \rightarrow C]_i$), the open phase duration (t^i_{open}), and the closed phase duration (t^i_{closed}). Also shown is the glottal area dynamic range A_i, which describes the maximum changing area during one cycle between glottis opening and closing.[20] Four images of male vocal folds in different stages of the glottal cycle from opening to closed phase are illustrated.

- The Speed Quotient,[21] defined as

$$Sq_i = \frac{[C \rightarrow O]_i}{[O \rightarrow C]_i} = \frac{Oq_i - Cq_i}{Cq_i}.$$

- The Speed Index,[21] defined as

$$Si_i = \frac{[C \rightarrow O]_i - [O \rightarrow C]_i}{t^i_{open}} = \frac{Sq_i - 1}{Sq_i + 1}.$$

- The Rate Quotient,[21] defined as

$$Rq_i = \frac{t^i_{closed} + [C \rightarrow O]_i}{[O \rightarrow C]_i} = \frac{1}{Cq_i} - 1.$$

- The Asymmetry Coefficient,[23] defined as

$$Aq_i = \frac{Sq_i}{1 + Sq_i}.$$

For the above parameters, the last four can be calculated using only the first two highlighted in bold type, which shows the high degree of redundancy within the group of GAW phase characteristic parameters. Hence it would be useful to reduce the number of parameters and limit them to just two parameters that provide independent information. In this case, Oq_i and Cq_i may be considered the most appropriate choice, because they describe the ratios of the different phases in a straightforward manner and are widely used in voice research.[24–27] According to previous research, breathy voice implies an increased open quotient; the closing quotient decreases when intensity and pressedness in voice increase.[22,24] However, depending on how exactly t^i_{closed} is defined and which type of signal is used to calculate Oq_i and Cq_i, these coherencies may vary. Moreover, it should be mentioned that these two parameters contain only relative information about the lengths of the different phases and absolute values are not included. Therefore, to complete the set of quantitative data, one absolute value of phase length is needed. In this case, the duration of the actual cycle T_i is thought to be an appropriate choice.

GAW derivative measures

GAW derivative measures aim to describe the change of the glottal area over time. Two different parameters that measure the maximal declination of the GAW in one cycle, that is, the maximum speed with which the glottis closes, are reviewed here. The first is the Maximum Area Declination Rate (*MADR*),[28] which is the absolute value of the minimum of the first derivative of the GAW in cycle *i* and the second is the Amplitude Quotient (*Amq*), which is a normalized form of *MADR*. Both parameters allow indirect insight in the viscoelastic properties of vocal folds[28] and are, in case of *Amq*, directly related to the closing phase.[29] Respectively the relation between *MADR* and closing phase is inverse. They can be written as follows:

- Maximum Area Declination Rate[28]:

$$MADR : \mathbb{R}^{T_i} \rightarrow \mathbb{R}$$

$$\{f_i(t)\} \rightarrow \left| \min_{t \in \{1,\dots,T_i\}} \left(\frac{d}{dt} f_i(t) \right) \right|$$

where i = cycle number, t = sample number, $f_i(t)$ = ith cycle of the signal, and T_i = period length of the ith cycle in samples.

- Amplitude Quotient[28,29]:

$$Amq_i : \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$\{A_i, MADR_i\} \rightarrow \frac{A_i}{MADR_i} | i = cycle\ number$$

where A_i is the glottal area dynamic range in cycle i (Figure 2), and therefore the amplitude of the GAW during one cycle, that is, distance from most open to most closed status of the glottis. To simplify the comparison of the following formulas for GAW and acoustic recordings, we will refer to this dynamic range as the "amplitude" of the GAW in this work. Furthermore, all amplitudes of GAW and audio signals are defined as greater than or equal to zero. *MADR* and *Amq* are inversely proportional to each other, and both contain information on the maximal declination, with the difference that *Amq* is additionally normalized to one through the dynamic range A_i. By performing the normalization to calculate *Amq*, the maximal declination of the glottal area is independent of the glottal area itself because it is canceled out. Because of this, one advantage is that values of *Amq* that are calculated from different GAW recordings can be compared directly. In contrast, comparing *MADR* values between different recordings is not meaningful because of varying distances between glottis and endoscope tip resulting in different numbers of pixels for the glottis itself.[9] Hence, it is not advisable to keep both parameters because they both contain the interesting information about the GAW declination. For the previously mentioned reason, we suggest that *Amq* should be kept as the relevant parameter.

Another approach for measuring opening and closing velocity and thereby the acceleration of the vocal fold by using GAW data is the calculation of Peak Closing Velocity and Peak Acceleration. These two parameters are based on the similarity assumption between a GAW signal and a sine wave. To avoid the calculation of the actual signal derivatives, the maximal velocity and acceleration of a sine wave are calculated instead. This sine wave is chosen to obtain the same period length T_i and the half amplitude A_i of the GAW. This approach is derived from formulas for the calculation of vocal fold movement, as described by Titze,[30] and leads to the following definitions:

- Peak Closing Velocity:

$$PeakClosingVelocity = 2\pi \cdot \frac{A_i}{2 \cdot T_i}.$$

- Peak Acceleration

$$PeakAcceleration = 4\pi^2 \cdot \frac{A_i}{2 \cdot T_i^2}.$$

The motivation behind this approach is to avoid the direct calculation of the signal derivatives, because this can lead to problems for noisy data. Instead, approximations of the maximal signal speed and acceleration are calculated, using the sine wave as an approximation for the GAW. However, if the information on the amplitude sizes and period lengths is known, these formulas relate the two variables to each other in a simple way. As indicated, measurements for calculating the closing velocity of the vocal folds, such as *MADR* and *Amq*, already exist. It may be true that the approach to approximate the GAW with a sine wave avoids some problems that can occur with noisy signals, but it also can be an oversimplification for non−sine-like GAW shapes of irregular vocal fold vibrations. Figure 3 shows the GAW of a healthy male glottis with a distinctive closed phase. Although no voice disorders were diagnosed for this subject, the variation of the glottal area is obviously not similar to the sinusoidal motion that is required by Peak Closing Velocity and Peak Acceleration parameters. The long closed phase extends the period length, resulting in an underestimation of the maximal GAW slope for a sine function with the same period length. Furthermore, similarly to *MADR*, Peak Closing Velocity and Peak Acceleration are also dependent on the unit of the glottal area. As mentioned before, this is highly disadvantageous when comparing data from different GAW recordings. For these reasons and because of the low explanatory power of the parameters, we suggest that both Peak Closing Velocity and Peak Acceleration should be discarded.

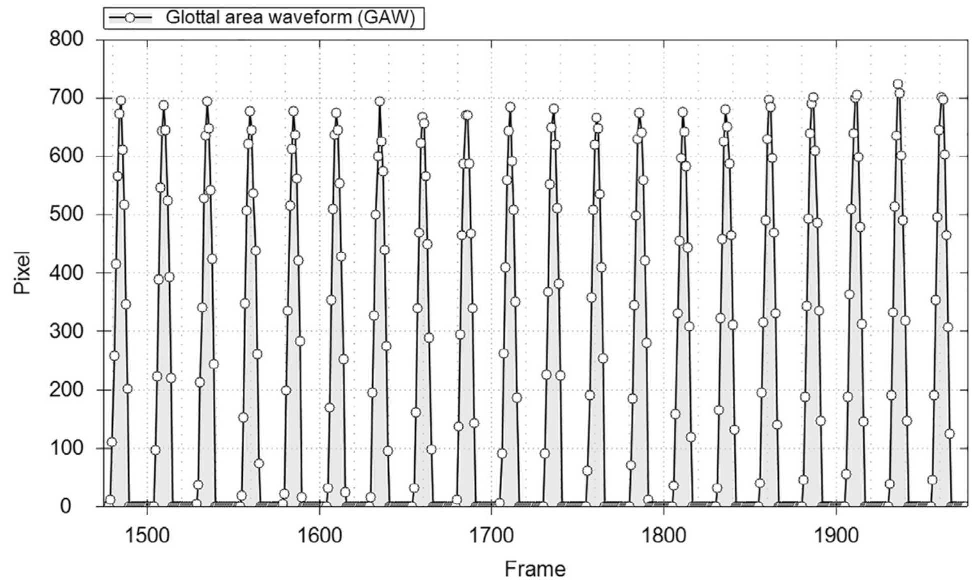

FIGURE 3. The GAW of a male glottis during sustained phonation exhibiting a distinctive closed phase; that is, $GAW = 0$.

Amplitude perturbation measures

Amplitude perturbation measures can be determined both for GAW and acoustic signals. Usually, a high degree of amplitude perturbation, and hence a high value of amplitude perturbation measures, is associated with change in stability of the voice. Furthermore, Shimmer measures are associated with the perception of hoarseness, even though there is no unanimity on this point.[31] For measuring the perturbation of the amplitudes of the GAW and other periodic signals, many different formulas exist.[17,18] However, this investigation focuses on the relationship between mean Shimmer and Shimmer(%). Shimmer(%) represents an attempt to normalize mean Shimmer to the percent scale and is a good example for parameters that are problematic. The mean Shimmer and Shimmer(%) are calculated as follows, whereas the absolute value in the denominator of Shimmer(%) was supplemented to avoid the possibility of obtaining negative values:

- **mean Shimmer**[32,33]:

$$meanShimmer : \mathbb{R}^N \rightarrow \mathbb{R}$$

$$\vec{A} = (A_0, \ldots, A_{N-1}) \rightarrow \frac{20}{N-1} \sum_{i=0}^{N-2} \left| \log_{10} \left[\frac{A_i}{A_{i+1}} \right] \right|.$$

- Shimmer(%)[33]:

$$Shimmer(\%) : \mathbb{R}^N \rightarrow \mathbb{R}$$

$$\vec{A} = (A_0, \ldots, A_{N-1}) \rightarrow \frac{meanShimmer}{\frac{20}{N} \sum_{i=0}^{N-1} |\log_{10} A_i|} \cdot 100,$$

where A_i is again the dynamic range of the GAW during cycle number i and N is the total number of cycles of the GAW. The first thing that catches one's eye is that whereas mean Shimmer is in decibels (dB), Shimmer(%) has obviously the wrong unit, by definition being percent. For accurate evaluation of its formula displayed earlier, the denominator turns out to be in the unit of the dynamic range A_i. Hence, in the case when the GAW was measured in pixels, the unit of Shimmer(%) would be $\frac{dB}{pixel}$. Another problem with Shimmer(%) is stated by *theorem A.1, Appendix A*. The maximum value that Shimmer(%) can reach is 300, which is unexpected for a parameter that should be a normalized measure of amplitude perturbation. *Theorem B.1* predicates a further issue: Shimmer(%) is nonlinearly dependent on the absolute size of the amplitudes if the ratio between the single amplitudes in the formula for mean Shimmer stays the same. Although in general a normalized version of a parameter should be preferable to a non-normalized parameter, in this case, the normalization did not work well. Furthermore, through this normalization, Shimmer(%) also became nonlinearly dependent on the absolute values of the amplitudes. Results suggest use of mean Shimmer instead of Shimmer(%).

Period perturbation measures

As amplitude perturbation, perturbation in period lengths between different oscillation cycles is measured by several parameters.[18] Similar to amplitude perturbation, period perturbation is also associated with the stability of the phonatory system. As the phonatory system is not a machine, no voice is perfectly free of period perturbation; nevertheless, an abnormal larynx should produce a more erratic voice than a healthy one.[34] At least four different variations of Jitter formulas exist, where one is the actual mean Jitter and the others are normalized variations suggested by different research groups. Mathematically, the different forms of Jitter are as follows:

- **mean Jitter**[33,35]:

$$meanJitter : \mathbb{R}^N \to \mathbb{R}$$
$$\vec{T} = (T_0, \ldots, T_{N-1}) \to \frac{\sum_{i=1}^{N-1} |T_i - T_{i-1}|}{N-1}.$$

- Jitter(%)[33,35]:

$$Jitter(\%) : \mathbb{R}^N \to \mathbb{R}$$
$$\vec{T} = (T_0, \ldots, T_{N-1}) \to \frac{meanJitter}{\frac{1}{N}\sum_{i=0}^{N-1} T_i} \cdot 100.$$

- Jitter Ratio[35]:

$$JitterRatio : \mathbb{R}^N \to \mathbb{R}$$
$$\vec{T} = (T_0, \ldots, T_{N-1}) \to \frac{meanJitter}{\frac{1}{N}\sum_{i=0}^{N-1} T_i} \cdot 1000.$$

- Jitter Factor[35,36]:

$$JitterFactor : \mathbb{R}^N \to \mathbb{R}$$
$$\vec{f} = (f_0, \ldots, f_{N-1}) \to \frac{\frac{1}{N-1}\sum_{i=1}^{N-1} |f_i - f_{i-1}|}{\frac{1}{N}\sum_{i=0}^{N-1} f_i} \cdot 100.$$

In this case, T_i is, analogously to amplitude perturbation measures, the duration of the cycle i of the GAW and N is again the total number of cycles. Further, f_i is the fundamental frequency of the ith cycle and therefore is equal to $\frac{1}{T_i}$. The most obvious dependency in this case exists between Jitter(%) and Jitter Ratio, because the latter is 10 times the former. Jitter(%) and Jitter Factor are also strongly connected as the only differences are their function arguments, which are mathematically dependent. Even though a connection exists between these two parameters, they behave differently to the presence of an outlier cycle with increased or decreased duration. The outlier cycle with increased duration will increase Jitter(%) up to a limit of 200, whereas Jitter Factor remains almost unaffected. On the other hand, one outlier cycle with decreased duration will increase the Jitter Factor but Jitter(%) remains almost unaffected. An example is given in Appendix C. Also, as stated in *theorem D.1*, similarly to Shimmer(%), Jitter(%) (and analogous Jitter Factor) can reach values of up to 300, whereas Jitter Ratio reaches values as high as 3000. Despite all these inconsistencies of the Jitter parameters, all of them are independent, and in case of mean Jitter, linearly dependent, of the absolute period lengths, as shown for Jitter(%) in *theorem E.1*. Furthermore, they are, with the exception of mean Jitter, dimensionless and therewith appropriate for comparison between different recordings and subjects.

However, all four Jitter parameters describe the perturbation of the cycle length and thereby of the fundamental frequency without adding further information. Thus, to be consistent with mean Shimmer as chosen parameter for the amplitude perturbation, we suggest that mean Jitter should be used.

In addition to different Jitter formulations, there are also other parameters for measuring period perturbation.[18] One of them is the "Relative Average Perturbation" (RAP), which can be calculated using two different formulas, that is, version 1 (v1)[33] and version 2 (v2)[37,38]:

- Relative Average Perturbation—version 1:

$$RAP_{v1} : \mathbb{R}^N \to \mathbb{R}$$
$$\vec{T} = (T_0, \ldots, T_{N-1}) \to \frac{\sum_{i=1}^{N-2} \left| \frac{T_{i-1}+T_i+T_{i+1}}{3} - T_i \right|}{\sum_{i=0}^{N-1} T_i}.$$

- Relative Average Perturbation—version 2:

$$RAP_{v2} : \mathbb{R}^N \to \mathbb{R}$$
$$\vec{T} = (T_0, \ldots, T_{N-1}) \to \frac{\frac{1}{N-2}\sum_{i=1}^{N-2} \left| \frac{T_{i-1}+T_i+T_{i+1}}{3} - T_i \right|}{\frac{1}{N}\sum_{i=0}^{N-1} T_i}.$$

The second formula is an attempt to normalize the first version and therefore only one of them is needed. As *theorem F.1* states, RAP_{v1} reaches a maximum value of $\frac{4}{3}$. RAP_{v2} is additionally dependent on N and goes up to $\frac{20}{9}$ (see corollary of *theorem F.1* in Appendix F). However, $\frac{20}{9}$ is only reached for the case $N=5,\ T_0=T_1=T_3=T_4=0,\ T_2>0$. Aside from that, compared with Jitter, there is also a less obvious difference, as the following example shows:

$$if \quad T_{i+1} = T_i + c \,\forall\, i, c \in \mathbb{R} > 0$$
$$\begin{aligned} RAP_{v1} &= \frac{\sum_{i=1}^{N-2} \left| \frac{T_{i-1} + T_i + T_{i+1}}{3} - T_i \right|}{\sum_{i=0}^{N-1} T_i} \\ &= \frac{\sum_{i=1}^{N-2} \left| \frac{T_i - c + T_i + T_i + c}{3} - T_i \right|}{\sum_{i=0}^{N-1} T_i} \\ &= \frac{\sum_{i=1}^{N-2} \left| \frac{3T_i}{3} - T_i \right|}{\sum_{i=0}^{N-1} T_i} \\ &= \frac{\sum_{i=1}^{N-2} |T_i - T_i|}{\sum_{i=0}^{N-1} T_i} \\ &= 0. \end{aligned}$$

As one can see, RAP_{v1} (and analogously RAP_{v2}) is insensitive to overall linear deviations, so if the cycle duration increases or decreases by a constant value c over time, this perturbation will not be detected by both RAP measures. Depending on the type of noise that is added to the signal,

TABLE 1.
Summary of the Different Parameters Discussed Together With Our Suggestions Where Further Use Seems Appropriate

Parameter	Source	Reasons to ✘ or ✓	Retain
Phase characteristic parameters			
Open Quotient	GAW	Widely used, straightforward	✓
Closing Quotient	GAW	Widely used, straightforward	✓
Speed Quotient	GAW	Redundant	✘
Speed Index	GAW	Redundant	✘
Rate Quotient	GAW	Redundant	✘
Asymmetry Coefficient	GAW	Redundant	✘
Derivative measures			
Maximum Area Declination Rate	GAW	Not normalized, redundant information	✘
Amplitude Quotient	GAW	Normalized	✓
Peak Closing Velocity	GAW	Oversimplification, not normalized	✘
Peak Acceleration	GAW	Oversimplification, not normalized	✘
Amplitude perturbation measures			
mean Shimmer	GAW, acoustic	widely used, straightforward	✓
Shimmer(%)	GAW, acoustic	ill-design bad normalization	✘
Period perturbation measures			
mean Jitter	GAW, acoustic	Widely used, straightforward, consistent with mean Shimmer	✓
Jitter (%)	GAW, acoustic	Design problems	✘
Jitter Ratio	GAW, acoustic	Design problems, redundant	✘
Jitter Factor	GAW, acoustic	Design problems, redundant	✘
Relative Average Perturbation v1	GAW, acoustic	More consistent maximum value	✓
Relative Average Perturbation v2	GAW, acoustic	Design problems	✘
Signal periodicity measures			
mean Cycle Duration	GAW, acoustic	Redundant, only indirectly determinable	✘
Fundamental Frequency	GAW, acoustic	Directly determinable	✓

Note: The groups of directly connected parameters are separated from each other by horizontal dashed lines if they belong to the same subsection.

this can lead to unexpected low values of RAP. Still, this does not necessarily have to be a disadvantage, but anyone who uses this parameter should know about this characteristic. For standardization of the parameter set in GAW evaluation, just one of the RAP parameters should be kept. In this case, we suggest discarding RAP_{v2} because its maximum range is dependent on the number of cycles N.

Signal periodicity characteristics

For predicting the periodicity characteristics of the GAW, at least two different measurements are in use.[18] One of them is the mean cycle duration (T) in milliseconds (ms), which, as the name indicates, is equal to the mean duration of one oscillation cycle of the GAW. The other is the fundamental frequency (F_0) in Hz, which is calculated as follows:

$$F_0 = 1000 \cdot \frac{1}{T}.$$

The multiplication by 1000 is necessary to obtain F_0 in Hz, as T is commonly measured in ms owing to the relevant frequency range in human phonation. F_0 and T are related to the perceived pitch of the voice. In general, pitch increases with F_0, albeit the relation is not linear because the auditory system is varyingly sensitive to different frequency changes.[39] It is difficult to say which of either measurements is superior in a mathematical sense, and depending on the actual situation, it may be more convenient to choose one of them according to the focus of the study. For the sake of standardization, it should at least be acknowledged that both measurements technically rely on the same information. Thus, for automated and computerized calculation of relevant parameters, only one of them is necessary. In such cases, we suggest that F_0 be used, because it also can be directly determined using Fourier transformation. In contrast, for the calculation of T, the duration of all cycles T_i has to be known beforehand.

RESULTS AND DISCUSSION

As shown, parameters that are used to describe different features of the GAW and acoustic signals are very similar or redundant. Some of the parameters such as $MADR$, $PeakClosingVelocity$, and $PeakAcceleration$, owing to their design and the methods to calculate them, do not allow direct comparison between different recordings. Further, parameters may have unexpected properties, such as RAP_{v1} and RAP_{v2}, which are insensitive to linear increasing or decreasing cycle durations. Even severe design problems were revealed, as in the case of *Shimmer*(%), which may lead to the misinterpretation of data in the worst case. A summary of all parameters that were discussed here, together with our suggestions for keeping or discarding them, is presented in Table 1. In total, out of 20 analyzed parameters, 13 were found to be redundant.

Given the number of different parameters and the potential problems that arise with the use of some of them, the necessity to discard especially the ill-designed ones seems to be indisputable. However, it is not always uncontroversial which of the parameters should be discarded and which should be kept. To support our suggestions, various subtle properties of different parameters were shown, whereby some of these traits indicate severe design problems, as mentioned. Still, in some cases, different versions of parameters may be used to describe different kinds of data, as is the case with T and F_0. The clinical researchers have to be aware of the shortcomings of the GAW and acoustic parameters discussed in this work to ensure that the parameters can be used to determine the treatment effects and are comparable from one recording to the next.

Even the definition of the GAW varies. In some cases it is defined just as a plain function of the glottal area in pixels, as in this work, whereas in other cases additional means of normalization, such as division through the maximal opening of the glottis, are applied.[6–9] Hence, some of the abovementioned problems, such as the comparability of data from different recordings, may or may not arise for some of the parameters, depending on the used GAW definition. For instance, if one uses a normalized form of the GAW, *MADR* values from different recordings may be comparable.

Measurement of vocal fold vibratory characteristics and their relation to perceived vocal quality in vocal health and disease is essential to improve the clinical assessment and treatment of voice disorders. Establishing sound measures of laryngeal biomechanics that give insight into the changes into viscoelastic properties of vocal folds before and after voice intervention is still needed. This study results highlight the need for the clinicians to educate themselves on the limitations of routinely used parameters for the assessment of vocal fold vibratory function as the outcomes of behavioral or medical voice interventions. This study identified mathematical dependencies between some of these parameters with the aim of reducing the numbers of parameters used in clinical settings and suggested which parameters may best describe the properties of the GAW and the acoustical signal.

LIMITATIONS

The selection of parameters discussed in this paper represents only a small fraction of frequently used parameters and does not cover all of the less explicit connections between at least 63 different formulas and variations of formulas for GAW parameter calculation.[18] Also, it was not investigated if the found issues also apply for different GAW calculations.[6–9] Furthermore, there may exist other definitions or derivations of the examined parameters, which were also not investigated here.

CONCLUSION

In this work, one step toward parameter reduction has been suggested. Even though it will require future effort and many further steps of discussion and debate, the final aim of a greatly reduced set of significant parameters should be worth this effort. A small set of standardized parameters will be a major step to enhanced information exchange between different studies and help to increase relevance of evaluated data in clinical settings. Future studies will investigate the presence of the redundancy in other frequently used parameters in voice research and clinic.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under grants BO4399/2-1 and DO1247/8-1 (no. 323308998).

APPENDICES

Appendix A. We prove by contradiction that the maximum value that *Shimmer*(%) can reach is 300

Theorem A.1: (Let N be a non-negative integer >2 and let $\vec{A}$ be a vector of amplitudes in real numbers, which is component-wise greater than or equal to zero and which contains at least one $A_i \neq 1$, then).

a) $Shimmer(\%)(\vec{A}) \leq 300$.
b) Moreover there exists $\vec{A}*$ in $\mathbb{R}_0^{+N}$ for which $Shimmer(\%)(\vec{A}*) = 300$.

Proof:

$$a)\ \frac{\frac{20}{N-1}\sum_{i=0}^{N-2}\left|\log_{10}\left[\frac{A_i}{A_{i+1}}\right]\right|}{\frac{20}{N}\sum_{i=0}^{N-1}|\log_{10}A_i|}\cdot 100 > 300$$

$$\Leftrightarrow N\cdot\sum_{i=0}^{N-2}\left|\log_{10}\left[\frac{A_i}{A_{i+1}}\right]\right| > 3\cdot(N-1)\cdot\sum_{i=0}^{N-1}|\log_{10}A_i|$$

$$\Leftrightarrow N\cdot\sum_{i=0}^{N-2}|\log_{10}A_i - \log_{10}A_{i+1}| > 3N\cdot\sum_{i=0}^{N-1}|\log_{10}A_i| - 3\cdot\sum_{i=0}^{N-1}|\log_{10}A_i|$$

$$\Rightarrow N\cdot\sum_{i=0}^{N-2}(|\log_{10}A_i| + |\log_{10}A_{i+1}|) > 3N\cdot\sum_{i=0}^{N-1}|\log_{10}A_i| - 3\cdot\sum_{i=0}^{N-1}|\log_{10}A_i|$$

$$\Leftrightarrow N\cdot\sum_{i=0}^{N-1}|\log_{10}A_i|-N|\log_{10}A_{N-1}| +N\cdot\sum_{i=0}^{N-1}|\log_{10}A_i|-N|\log_{10}A_0|$$

$$> 3N\cdot\sum_{i=0}^{N-1}|\log_{10}A_i|-3\cdot\sum_{i=0}^{N-1}|\log_{10}A_i|$$

$$\Leftrightarrow \frac{3}{N}\cdot\sum_{i=0}^{N-1}|\log_{10}A_i| > \sum_{i=0}^{N-1}|\log_{10}A_i| + |\log_{10}A_{N-1}| + |\log_{10}A_0|$$

$\Rightarrow$ Contradiction if $N \geq 3$if $N = 2$:

$$\frac{3}{2}\cdot\sum_{i=0}^{1}|\log_{10}A_i| > \sum_{i=0}^{1}|\log_{10}A_i| + |\log_{10}A_1| + |\log_{10}A_0| = 2\cdot\sum_{i=0}^{1}|\log_{10}A_i|$$

$\Rightarrow$ Contradiction if $N = 2$

$\Rightarrow$ Shimmer(%) cannot reach any value higher than 300 for any N

$$b)\ \text{Case } N = 3,\ A_0 = A_2 = 1,\ A_1 > 1$$

$$\frac{\frac{20}{2}\sum_{i=0}^{1}\left|\log_{10}\left[\frac{A_i}{A_{i+1}}\right]\right|}{\frac{20}{3}\sum_{i=0}^{2}|\log_{10}A_i|}\cdot 100 = \frac{\frac{1}{2}(|\log_{10}A_1| + |\log_{10}A_1|)}{\frac{1}{3}|\log_{10}A_1|}\cdot 100$$

$$= \frac{|\log_{10}A_1|}{\frac{1}{3}|\log_{10}A_1|}\cdot 100 = 3\cdot 100 = 300\ q.e.d.$$

therefore, as shown above, the maximum value that *Shimmer*(%) can reach is 300.

Appendix B. We prove that *Shimmer*(%) is dependent on the absolute values of its amplitudes

Theorem B.1: (Let N be a non-negative integer >2 and let $\vec{A}$ be a vector of amplitudes in real numbers, which is component-wise greater than or equal to zero and which contains at least one $A_i \neq 1$, then). The function $Shimmer(\%)(\vec{A})$ is in general not linear.

Proof:

Let $c = 1$, $N = 2$, $A_1 = 100$ and $A_2 = 1000$ be given, then:

$$\frac{\frac{20}{N-1}\sum_{i=0}^{N-2}\left|\log_{10}\left[\frac{c\cdot A_i}{c\cdot A_{i+1}}\right]\right|}{\frac{20}{N}\sum_{i=0}^{N-1}|\log_{10}(c\cdot A_i)|}\cdot 100 = \frac{20\cdot\left|\log_{10}\left[\frac{100}{1000}\right]\right|}{10\cdot(|\log_{10}100| + |\log_{10}1000|)}\cdot 100 = 40$$

if $c = 0.1$

$$\frac{20\cdot\left|\log_{10}\left[\frac{100}{1000}\right]\right|}{10\cdot(|\log_{10}100 + \log_{10}0.1| + |\log_{10}1000 + \log_{10}0.1|)}\cdot 100 = \frac{200}{3}$$

if $c = 10$

$$\frac{20\cdot\left|\log_{10}\left[\frac{100}{1000}\right]\right|}{10\cdot(|\log_{10}100 + \log_{10}10| + |\log_{10}1000 + \log_{10}10|)}\cdot 100 = \frac{200}{7}$$

These different pairs of c and *Shimmer*(%)(c) do not lie on a straight line:

$$\text{I} : 40 = m\cdot 1 + t \Leftrightarrow t = 40 - m = \frac{280}{27}$$

$$\text{II} : \frac{200}{3} = m\cdot 0.1 + t = m\cdot 0.1 + 40 - m \Leftrightarrow m = \frac{800}{27}$$

$$\text{III} : \frac{200}{7} \neq \frac{800}{27}\cdot 10 + \frac{280}{27}\ q.e.d.$$

Therefore no straight line exists that can contain all three value pairs of c_i and $Shimmer(\%)(c_i\vec{A})$. Hence *Shimmer*(%) depends nonlinearly on c.

Appendix C. In the following we give an example of the different reactions to outlier cycles of *Jitter*(%) and *Jitter Factor*

Influence of prolonged outlier cycles on *Jitter*(%):

$$\frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i-1}|}{\frac{1}{N}\sum_{i=0}^{N-1}T_i}\cdot 100$$

$$= \frac{\frac{1}{N-1}\cdot\left(\begin{array}{c}|T_1 - T_0| + \ldots + |T_k - T_{k-1}| + |T_{k+1} - T_k)| \\ + \ldots + |T_N - T_{N-1}|\end{array}\right)}{\frac{1}{N}(T_0 + \ldots + T_k + \ldots + T_N)}\cdot 100$$

$$\approx \frac{\frac{1}{N-1}\cdot(T_k + T_k)}{\frac{1}{N}(T_k)}\cdot 100$$

$$\rightarrow 200\ \textit{for large } N$$

Influence of prolonged outlier cycles on *Jitter Factor*:

$$\frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|\frac{1}{T_i} - \frac{1}{T_{i-1}})\right|}{\frac{1}{N}\sum_{i=0}^{N-1}\frac{1}{T_i}}\cdot 100$$

$$= \frac{\frac{1}{N-1}\cdot\left(\begin{array}{c}\left|\frac{1}{T_1}-\frac{1}{T_0}\right|+\ldots+\left|\frac{1}{T_k}-\frac{1}{T_{k-1}}\right|+\left|\frac{1}{T_{k+1}}-\frac{1}{T_k}\right|\\ +\ldots+\left|\frac{1}{T_N}-\frac{1}{T_{N-1}}\right|\end{array}\right)}{\frac{1}{N}\left(\frac{1}{T_0}+\ldots+\frac{1}{T_k}+\ldots+\frac{1}{T_N}\right)}\cdot 100$$

$$\approx \frac{\frac{1}{N-1}\cdot\left(\begin{array}{c}\left|\frac{1}{T_1}-\frac{1}{T_0}\right|+\ldots+\left|0-\frac{1}{T_{k-1}}\right|+\left|\frac{1}{T_{k+1}}-0\right|\\ +\ldots+\left|\frac{1}{T_N}-\frac{1}{T_{N-1}}\right|\end{array}\right)}{\frac{1}{N}\left(\frac{1}{T_0}+\ldots+0+\ldots+\frac{1}{T_N}\right)}\cdot 100$$

Therefore, the value of *Jitter*(%) converges to 200, whereas the value of *Jitter Factor* is mainly dependent on values of *T*(*i*). Hence a perturbation with single prolonged cycles has almost no influence on the value of *Jitter Factor*. This applies analogously also for perturbation with single shortened cycles, if the role of *Jitter*(%) and *Jitter Factor* are switched.

Appendix D. We prove that analogously to *Shimmer* (%), the maximum value that *Jitter*(%) can reach is also 300

Theorem D.1: (Let N be a non-negative integer >2 and let $\vec{T}$ be a vector of cycle durations in real numbers, which is component-wise greater than or equal to zero and which contains at least one $T_i \neq 0$, then).

a) $Jitter(\%)(\vec{T}) \leq 300$.

b) Moreover there exists $\vec{T}*$ in $\mathbb{R}_0^{+N}$ for which $Jitter(\%)(\vec{T}*) = 300$.

Proof:

$$a)\ \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i-1}|}{\frac{1}{N}\sum_{i=0}^{N-1}T_i}\cdot 100 > 300$$

$$\Leftrightarrow N\cdot\sum_{i=1}^{N-1}|T_i - T_{i-1}| > 3N\cdot\sum_{i=0}^{N-1}T_i - 3\cdot\sum_{i=0}^{N-1}T_i$$

$$\Rightarrow N\cdot\sum_{i=1}^{N-1}T_i + N\cdot\sum_{i=1}^{N-1}T_{i-1} > 3N\cdot\sum_{i=0}^{N-1}T_i - 3\cdot\sum_{i=0}^{N-1}T_i$$

$$\Leftrightarrow N\cdot\sum_{i=0}^{N-1}T_i - N\cdot T_0 + N\cdot\sum_{i=0}^{N-1}T_i - N\cdot T_{N-1} > 3N\cdot\sum_{i=0}^{N-1}T_i - 3\cdot\sum_{i=0}^{N-1}T_i$$

$$\Leftrightarrow \frac{3}{N}\cdot\sum_{i=0}^{N-1}T_i > \sum_{i=0}^{N-1}T_i + T_0 + T_{N-1}$$

Contradiction if $N \geq 3$if $N = 2$:

$$\frac{3}{2}\cdot\sum_{i=0}^{1}T_i > \sum_{i=0}^{1}T_i + T_0 + T_1 = 2\cdot\sum_{i=0}^{1}T_i$$

Contradiction if $N = 2$

$\Rightarrow$ Jitter(%) cannot reach any value higher than 300 for any N

$b)$ Case $N = 3,\ A(0) = A(2) = 1,\ A(1) > 1$

$$\frac{\frac{1}{2}\sum_{i=1}^{2}|T_i - T_{i-1}|}{\frac{1}{3}\sum_{i=0}^{2}T_i}\cdot 100 = \frac{\frac{1}{2}\cdot(T_1 + T_1)}{\frac{1}{3}T_1}\cdot 100 = \frac{T_1}{\frac{1}{3}T_1}\cdot 100$$
$$= 3\cdot 100 = 300\ q.e.d.$$

This shows that the maximum value that *Jitter*(%) can reach is 300.

Appendix E. We prove that *Jitter*(%) is independent of the absolute values of its amplitudes

Theorem E.1: (Let N be a non-negative integer >2 and let $\vec{T}$ be a vector of cycle durations in real numbers, which is component-wise greater than or equal to zero and which contains at least one $T_i \neq 0$, then). For any c in $\mathbb{R}^+$ it is true that $Jitter(\%)(c\cdot\vec{T}) = Jitter(\%)(\vec{T})$; that is, *Jitter*(%) is invariant with respect to uniform scaling of the function argument.
Proof:

$$\frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|c\cdot T_i - c\cdot T_{i-1}|}{\frac{1}{N}\sum_{i=0}^{N-1}c\cdot T_i}\times 100$$
$$= \frac{\frac{1}{N-1}\cdot c\cdot\sum_{i=1}^{N-1}|T_i - T_{i-1}|}{\frac{1}{N}\cdot c\cdot\sum_{i=0}^{N-1}T_i}\times 100$$
$$= \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i-1}|}{\frac{1}{N}\sum_{i=0}^{N-1}T_i}\times 100\ q.e.d.$$

Therefore, *Jitter*(%) is independent of any chosen positive real number *c*.

Appendix F. We prove that the maximum value that RAP_{v1} can reach is $\frac{4}{3}$

Theorem F.1: (Let N be a non-negative integer >2 and let $\vec{T}$ be a vector of cycle durations in real numbers, which is component-wise greater than or equal to zero and which contains at least one $T_i \neq 0$, then).

a) $RAP_{v1}(\vec{T}) \leq \frac{4}{3}$.

b) Moreover, there exists $\vec{T}*$ in $\mathbb{R}_0^{+N}$ for which $RAP_{v1}(\vec{T}*) = \frac{4}{3}$.
c) Furthermore, $RAP_{v1}(\vec{T}) \leq \frac{2}{3}$ if $N = 3$.
d) And $RAP_{v1}(\vec{T}) \leq 1$ if $N = 4$.

Proof:

$$a)\ \frac{\sum_{i=1}^{N-2}\left|\frac{T_{i-1}+T_i+T_{i+1}}{3} - T_i\right|}{\sum_{i=0}^{N-1} T_i} > \frac{4}{3}$$

$$\Leftrightarrow \sum_{i=1}^{N-2}\left|\frac{T_{i-1}+T_{i+1}-2T_i}{3}\right| > \frac{4}{3}\cdot\sum_{i=0}^{N-1} T_i$$

$$\Leftrightarrow \sum_{i=1}^{N-2}|T_{i-1}+T_{i+1}-2T_i| > 4\cdot\sum_{i=1}^{N-2} T_i + 4\cdot T_0 + 4\cdot T_{N-1}$$

$$\Rightarrow \sum_{i=1}^{N-2} T_{i-1} + \sum_{i=1}^{N-2} T_{i+1} + 2\cdot\sum_{i=1}^{N-2} T_i > 4\cdot\sum_{i=1}^{N-2} T_i + 4\cdot T_0 + 4\cdot T_{N-1}$$

$$\Leftrightarrow \sum_{i=1}^{N-2} T_i + T_0 - T_{N-2} + \sum_{i=1}^{N-2} T_i - T_1 + T_{N-1} > 2\cdot\sum_{i=1}^{N-2} T_i + 4\cdot T_0 + 4\cdot T_{N-1}$$

$$\Leftrightarrow 0 > 3\cdot T_0 + 3\cdot T_{N-1} + T_{N-2} + T_1$$

$$\Rightarrow \text{Contradiction since } T_i \geq 0\ \forall\, i \text{ and } \exists\, T_i > 0$$

$$b)\ \text{Case } N = 5, T_0 = T_1 = T_3 = T_4 = 0, T_2 > 0$$

$$\frac{\sum_{i=1}^{3}\left|\frac{T_{i-1}+T_i+T_{i+1}}{3} - T_i\right|}{\sum_{i=0}^{4} T_i} = \frac{\frac{1}{3}T_2 + \frac{2}{3}T_2 + \frac{1}{3}T_2}{T_2} = \frac{\frac{4}{3}T_2}{T_2} = \frac{4}{3}\ q.e.d.$$

This shows that the maximum value that RAP_{v1} can reach is $\frac{4}{3}$.

$$c)\ \frac{\sum_{i=1}^{1}\left|\frac{T_{i-1}+T_i+T_{i+1}}{3} - T_i\right|}{\sum_{i=0}^{2} T_i} > \frac{2}{3}$$

$$\Leftrightarrow \frac{\left|\frac{T_0+T_2-2T_1}{3}\right|}{T_0+T_1+T_2} > \frac{2}{3}$$

$$\Rightarrow \frac{T_0+T_2+2T_1}{3} > \frac{2T_0}{3} + \frac{2T_1}{3} + \frac{2T_2}{3}$$

$$\Leftrightarrow 0 > \frac{1T_0}{3} + \frac{1T_2}{3}$$

This leads to a contradiction because there has to exist at least one T_i that is greater than zero and furthermore T_i cannot be less than zero.

Hence RAP_{v1} cannot reach any value higher than $\frac{2}{3}$ if $N = 3$.

$$d)\ \frac{\sum_{i=1}^{2}\left|\frac{T_{i-1}+T_i+T_{i+1}}{3} - T_i\right|}{\sum_{i=0}^{3} T_i} > 1$$

$$\Leftrightarrow \frac{\left|\frac{T_0+T_2-2T_1}{3}\right| + \left|\frac{T_1+T_3-2T_2}{3}\right|}{T_0+T_1+T_2+T_3} > 1$$

$$\Rightarrow \frac{T_0+T_2+2T_1}{3} + \frac{T_1+T_3+2T_2}{3} > T_0+T_1+T_2+T_3$$

$$\Leftrightarrow 0 > \frac{2T_0}{3} + \frac{2T_3}{3}$$

Analogously to *c*), this also leads to a contradiction.

Hence RAP_{v1} cannot reach any value higher than 1 if $N = 4$.

Corollary F.1: (If theorem F.1 (*a*–*d*) are fulfilled). Also $RAP_{v2}(\vec{T}) \leq \frac{20}{9}$ if $\vec{T}$ is a vector of cycle durations in real numbers, which is component-wise greater than or equal to zero and which contains at least one $T_i \neq 0$.
Moreover, there exists $\vec{T}*$ in $\mathbb{R}_0^{+N}$ for which $RAP_{v2}(\vec{T}*) = \frac{20}{9}$.
Proof:if $N = 3$

$$\max(RAP_{v2}) = \max\left(\frac{3}{3-2}RAP_{v1}\right) = 3\cdot\max(RAP_{v1}) = 3\cdot\frac{2}{3} = 2$$

if $N = 4$

$$\max(RAP_{v2}) = \max\left(\frac{4}{4-2}RAP_{v1}\right) = 2\cdot\max(RAP_{v1}) = 2\cdot 1 = 2$$

if $N \geq 5$

$$\max(RAP_{v2}) = \max\left(\frac{N}{N-2}RAP_{v1}\right) = \frac{5}{3}\cdot\max(RAP_{v1}) = \frac{5}{3}\cdot\frac{4}{3} = \frac{20}{9}$$

Hence the maximum value that RAP_{v2} can reach is $\frac{20}{9}$.

REFERENCES

1. Titze IR. *Principles of Voice Production*. 2nd ed Salt Lake City, Utah: National Center for Voice and Speech; 2000.
2. Stevens KN. *Acoustic Phonetics*. Cambridge, Massachusetts: MIT Press; 1999.
3. Semmler M, Kniesburges S, Parchent J, et al. Endoscopic laser-based 3d imaging for functional voice diagnostics. *Appl Sci*. 2017;7:600.
4. Deliyski D. *Laryngeal Evaluation*. New York city, New York: Georg Thieme; 2010.
5. Döllinger M. The next step in voice assessment: high-speed digital endoscopy and objective evaluation. *Curr Bioinform*. 2009;4: 101–111.
6. Chen X, Bless D, Yan Y. *A segmentation scheme based on Rayleigh distribution model for extracting glottal waveform from high-speed laryngeal images*. In 27th Annual International Conference of the Engineering in Medicine and Biology Society (IEEE-EMBS) 2005, pages 6269−6272. IEEE. 2006.
7. Mendez A, Gracia B, Ruiz I, et al. *Glottal area segmentation without initialization using Gabor filters*. In IEEE International Symposium on Signal Processing and Information Technology (ISSPIT) 2008, pages 18−22. IEEE. 2008.
8. Kunduk M, Yan Y, McWhorther AJ, et al. Investigation of voice initiation and voice offset characteristics with high-speed digital imaging. *Logoped Phoniatr Vocol*. 2006;31:139–144.
9. Noordzij PJ, Woo P. Glottal area waveform analysis of benign vocal fold lesions before and after surgery. *Ann Otol Rhinol Laryngol*. 2000;109:441–446.
10. Patel R, Dailey S, Bless D. Comparison of high-speed digital imaging with stroboscopy for laryngeal imaging of glottal disorders. *Ann Otol Rhinol Laryngol*. 2008;117:413–424.
11. Olthoff A, Woywod C, Kruse E. Stroboscopy versus high-speed glottography: a comparative study. *Laryngoscope*. 2007;117:1123–1126.
12. Švec JG, Schutte HK. Videokymography: high-speed line scanning of vocal fold vibration. *J Voice*. 1996;10:201–205.
13. Schutte HK, Švec JG, Šram F. First results of clinical application of videokymography. *Laryngoscope*. 1998;108:1206–1210.
14. Deliyski D, Hillman R. State of the art laryngeal imaging: research and clinical implications. *Curr Opin Otolaryngol Head Neck Surg*. 2010;18: 147–152.
15. Semmler M, Kniesburges S, Birk V, et al. 3D reconstruction of human laryngeal dynamics based on endoscopic high-speed recordings. *IEEE Trans Med Imaging*. 2016;35:1615–1624.
16. Echternach M, Döllinger M, Sundberg J, et al. Vocal fold vibrations at high soprano fundamental frequencies. *J Acoust Soc Am*. 2013;133: 82–87.
17. Yeğin Y, Çelik M, Şimşek BM, et al. Assessment of effects of septoplasty on acoustic parameters of voice: a prospective clinical study. *Turk Arch Otorhinolaryngol*. 2016;54:146–149.
18. Pedersen M, Jønsson A, Mahmood S, et al. Which mathematical and physiological formulas are describing voice pathology: an overview. *J Gen Pract*. 2016;4:253.
19. Bough D, Heuer R, Sataloff R, et al. Intrasubject variability of objective voice measures. *J Voice*. 1996;10:166–174.
20. Division of Phoniatrics and Pediatric Audiology at the Department of Otorhinolaryngology, Waldstraße 1 91054 Erlangen, Germany. *Glottis Analysis Tools version 16.1—User guide*. 2017.
21. Baken RJ, Orlikoff RF. *Clinical Measurement of Speech & Voice (Speech Science)*. 2 ed New York city, New York: Cengage Learning; 1999:409.
22. Holmberg EB, Hillman RE, Perkell JS. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *J Acoust Soc Am*. 1988;84:511–529.
23. Henrich N, Sundin G, Ambroise D, et al. Just noticeable differences of open quotient and asymmetry coefficient in singing voice. *J Voice*. 2003;17:481–494.
24. Järvinen K, Laukkanen A-M, Geneid A. Voice quality in native and foreign languages investigated by inverse filtering and perceptual analyses. *J Voice*. 2017;31 261.e25−261.e31.
25. Ras Y, Imam M, El-Banna M, et al. Voice outcome following electrical stimulation-supported voice therapy in cases of unilateral vocal fold paralysis. *Egypt J Otolaryngol*. 2016;32:322–334.
26. Benninger M, Sataloff R. 1st ed *Sataloff's Comprehensive Textbook of Otolaryngology Head & Neck Surgery Laryngology*Vol. 4New Delhi, India: Jaypee Brothers Medical Publishers; 2016.
27. Kettlewell B. *The influence of intraglottal vortices upon the dynamics of the vocal folds*. Mathesis, University of Waterloo. 2015.
28. Patel R, Dubrovskiy D, Döllinger M. Measurement of glottal cycle characteristics between children and adults: physiological variations. *J Voice*. 2014;28:476–486.
29. Alku P, Arias M, Björkner E, et al. An amplitude quotient based method to analyze changes in the shape of the glottal pulse in the regulation of vocal intensity. *J Acoust Soc Am*. 2006;120:1052–1062.
30. Titze I. Mechanical stress in phonation. *J Voice*. 1994;8:99–105.
31. Baken RJ, Orlikoff RF. *Clinical Measurement of Speech & Voice (Speech Science)*. 2 ed New York city, New York: Cengage Learning; 1999:130.
32. Baken RJ, Orlikoff RF. *Clinical Measurement of Speech & Voice (Speech Science)*. 2 ed New York city, New York: Cengage Learning; 1999:133.
33. Bielamowicz S, Kreiman J, Gerratt B, et al. Comparison of voice analysis systems for perturbation measurement. *J Speech Hear Res*. 1996; 39:126–134.
34. Baken RJ, Orlikoff RF. *Clinical Measurement of Speech & Voice (Speech Science)*. 2 ed New York city, New York: Cengage Learning; 1999:190–191.
35. Baken RJ, Orlikoff RF. *Clinical Measurement of Speech & Voice (Speech Science)*. 2 ed New York city, New York: Cengage Learning; 1999:201–203.
36. Hollien H, Michel J, Doherty ET. A method for analyzing vocal jitter in sustained phonation. *J Phon*. 1973;1:85–91.
37. Koike Y. Application of some acoustic measures for the evaluation of laryngeal dysfunction. *Studia Phonologica*. 1973;7:17–23.
38. Baken RJ, Orlikoff RF. *Clinical Measurement of Speech & Voice (Speech Science)*. 2 ed New York city, New York: Cengage Learning; 1999:205.
39. Baken RJ, Orlikoff RF. *Clinical Measurement of Speech & Voice (Speech Science)*. 2 ed New York city, New York: Cengage Learning; 1999:148.