Normalized time-domain parameters for electroglottographic waveforms

Sten Ternström
Department of Speech, Music and Hearing, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden
stern@kth.se

Abstract: The electroglottographic waveform is of interest for characterizing phonation non-invasively. Existing parameterizations tend to give disparate results because they rely on somewhat arbitrary thresholds and/or contacting events. It is shown that neither are needed for formulating a normalized contact quotient and a normalized peak derivative. A heuristic combination of the two resolves also the ambiguity of a moderate contact quotient, with regard to vocal fold contacting being firm versus weak or absent. As preliminaries, schemes for electroglottography signal preconditioning and time-domain period detection are described that improve somewhat on similar methods. The algorithms are simple and compute quickly.

[AL]
Date Received: May 16, 2019 **Date Accepted:** June 28, 2019

1. Introduction

The principle for electroglottography (EGG) is accessibly described by Baken,[1] and a comprehensive, up-to-date overview of its scientific and clinical applications was recently presented by Herbst.[2]

The EGG contact quotient is usually taken to represent the relative amount of vocal fold (VF) adduction in phonation. It has been defined in the literature using various thresholds or events, leading to rather disparate results.[3] Here, we propose threshold-free formulations that require only a reliable period marker, for the EGG contact quotient, the EGG contacting speed, and an index that combines the two. Normalization circumvents the problem that the amplitude of EGG signals is notoriously variable with electrode placement and other external factors.

2. METHOD

2.1 Preconditioning the EGG signal

EGG signals are often noisy, for four reasons. (1) There is usually a large, irrelevant, fluctuating near-DC low frequency (LF) component below, say, 5 Hz, onto which the interesting but much smaller AC component is superimposed. EGG devices commonly incorporate high-pass filters at 5–20 Hz, but the residual LF energy can remain an inconvenience. (2) The conductive path through the informant's body is only weakly modulated by the VF contacting. A poor signal-to-noise ratio (SNR) can result if the electrode-to-VF distance becomes too large, as, for instance, if the larynx height varies a great deal, or the informant has a thyroid cartilage (Adam's apple) that is not prominent. (3) Spurious frequencies in the EGG signal can result if the high-frequency carrier of the EGG, typically at 2 MHz, interferes with other HF modulation, such as sampling artefacts of the A/D-converter (if the anti-aliasing filtering is inadequate), or HF modulation for inductive bands for plethysmography. (4) Some EGG devices develop a crackling noise over time, if their accumulator power cells are overcharged. Replacing the power cells will restore their proper function.

Against LF noise (1), a steep high-pass filter may be applied, below the lowest expected fundamental frequency (f_o). This filter must be linear-phase, to minimize distortion of the waveform at low f_o. At a sampling rate $f_s = 44.1$ kHz, we use a finite impulse response (FIR) filter of 1024 points, cutoff 80 Hz (−3 dB) and >60 dB attenuation below 20 Hz. A stricter filter would introduce too much delay, in a real-time implementation.

System noise (2) and spurious frequencies (3) may be inaudible but are visible in high-resolution long-time average spectra and/or spectrograms of the EGG signal. We have encountered various spurious frequencies above 12 kHz, so we routinely band-limit the EGG signals to 10 kHz (see also Sec. 4). The EGG spectrum usually has a very uniform slope toward the highest frequencies, so little information is lost in this spectral truncation. Against device noise (4), a nine-point running median filter is helpful. When

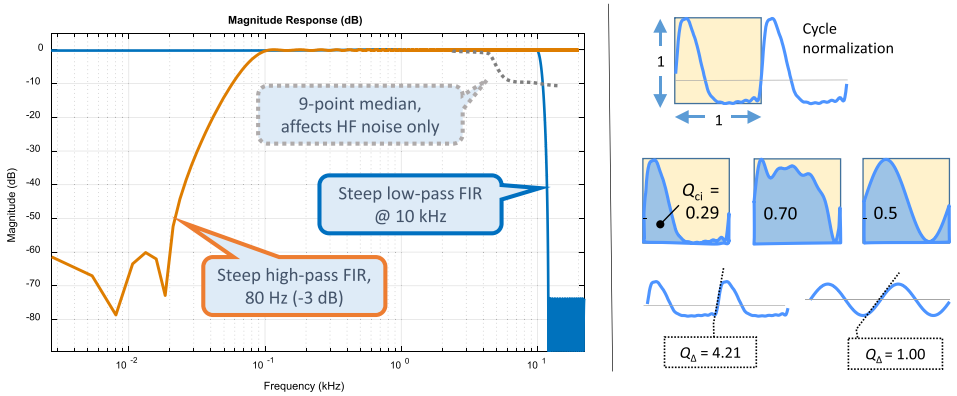

Fig. 1. (Color online) Left: Frequency responses of the filters used to pre-condition the EGG signal. Right, top to bottom: normalization of the EGG pulse in time and amplitude; Q_{ci} is the area below the pulse; Q_Δ is the normalized peak derivative, where a sine wave has $Q_\Delta = 1$.

the vocal folds are vibrating softly without actually colliding, the EGG signal is almost sinusoidal and very weak and noisy. The median filter lowers the noise floor above 5 kHz by about 10 dB, with negligible effect on the periodic signal. Figure 1 (left) shows the magnitude plots of the transfer functions of the above components.

2.2 Period detection

Although the waveform of the EGG signal is much simpler than that of the airborne acoustic voice signal, a robust period segmentation is not quite trivial. The steepest part of the contacting phase, corresponding to the maximum positive derivative of the EGG signal, is a good candidate for marking periods, but not ideal, due to the different pulse shapes that may occur for partial contact of the vocal folds.

Cycle segmentation by means of unwrapping the phase and plotting the phase portrait is a well-known technique; and it has been applied to EGG signals before.[4,5] The variant proposed here performs somewhat better in terms of noise rejection and polarity, and for real-time implementation. We do not claim that these cycle markers locate the precise instants of glottal closing or opening; that is a different issue.[6]

Figure 2 shows the period detection scheme, with signal examples in panels (a)–(c). The pre-conditioned EGG input sample sequence y_n is passed through a leaky integrator $x_n = y_n + \alpha x_{n-1}$ with a leakage coefficient $\alpha = 0.999$ at 44.1 kHz sampling rate. Even in the preconditioned EGG, the cycle average can be non-zero. Short-term accumulation in the integrator is blocked by a high-pass filter (50 Hz, second order Butterworth). The linear-phase requirement does not apply here.

A phase portrait [Fig. 2(b)] may be created with gx_n on the horizontal axis and y_n on the vertical axis, where g is a constant factor to make the magnitude of x_n similar to that of y_n. While the EGG phase portrait has been analyzed before,[7] we are here concerned only with obtaining a stable segmentation of cycles. Usually, phase

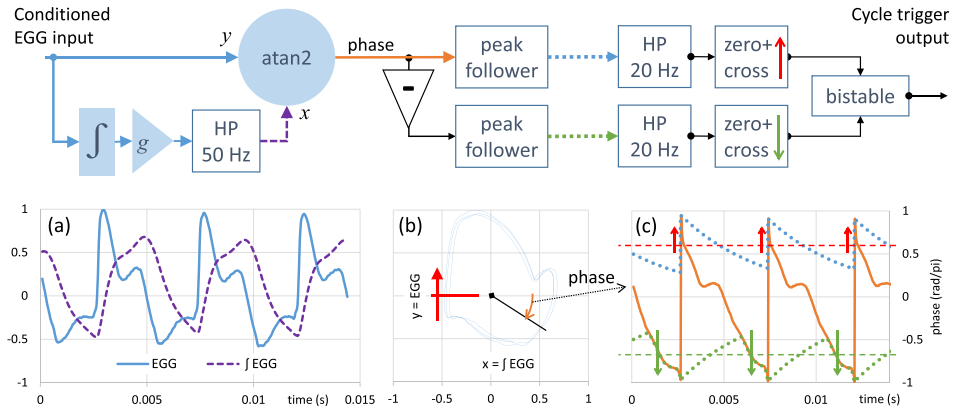

Fig. 2. (Color online) Block diagram of the period detection scheme, as applied to an unusual EGG signal with several peaks. (a) The pre-conditioned signal y (solid line) and its integrated version x (dashed); (b) the phase portrait of y vs x; (c) transients in the phase signal (solid) are selected using peak followers (dotted), which are then high-passed; zero-crossing triggers (arrows through dashed lines) drive the bistable gate that outputs the period marker.

portraits are made with the input sequence on the x axis, and on the y axis a delayed input sequence, or the time derivative, or the Hilbert transform of the input sequence. The latter two-valued signal (x_n, y_n) is often called the "analytic" signal, with the instantaneous phase $\varphi_n = \text{atan2}(x_n, y_n)$. Here we improve noise rejection by instead taking the integrated signal as the abscissa x and the original signal as the ordinate y. The phase φ_n then decreases with time [Fig. 2(c), solid line], wrapping around from $-\pi$ to $+\pi$ at the positive-going zero-crossings of x_n. This wrap-around of the phase is the most abrupt of the transients in the phase signal, although other local extrema are possible, if the phase is not monotonic over a cycle. The main transient is singled out *ad modum* Dolanský:[8] both the positive and the negative peak envelopes are found by hold-and-decay [Fig. 2(c), dotted]. The two envelope signals are high-pass filtered, and their zero-crossings [Fig. 2(c), arrows and dashed zero lines] are then used to toggle the state of a bistable "flip-flop" gate, whose positive flank becomes the period marker pulse. In moderate and strong phonation, this marker aligns closely with the peak derivative of the EGG. In weaker phonation, with only partial or no vocal fold contact, it aligns to the zero-crossing of the underlying sinusoidal wave (Fig. 3).

2.3 EGG pulse normalization

Let the EGG amplitude over one period be represented by a periodic function $a(t)$ with the period time T and hence $a(t) = a(t - T)$. Let $e(\theta)$ be the normalized EGG pulse shape, ranging from 0 to 1 in cycle time $\theta = t/T$, $\theta \in [0, \ldots, 1)$; and from 0 to 1 in amplitude. Let a_{min} and a_{max} be the minimum and maximum amplitudes of $a(t)$ observed on the interval $t \in [0, \ldots, \text{T})$, where $a_{min} < a_{max}$; and let the peak-to-peak amplitude $A_{p-p} = a_{max} - a_{min}$. The normalized pulse shape (Fig. 1, top right) is then obtained as $e(\theta) = [a(\theta T) - a_{min}]/A_{p-p}$.

2.4 Contact quotient by integration: Q_{ci}

For this EGG pulse, normalized to $0, \ldots, 1$ in both time and amplitude, we observe that the *area under the curve* $e(\theta)$ can vary from nearly zero for a very narrow EGG pulse, to nearly 1 for a very broad EGG pulse, regardless of whether the EGG waveform is simple or complicated (Fig. 1, right center). In this sense, this area can be interpreted as a dimensionless contact quotient, which we will call $Q_{ci,}$ for "quotient of contact by integration." Note that its definition does not require any thresholds or events other than the period markers, which can be located at any consistent phase in the period; nor does it matter if there are several local extrema in the EGG pulse (which is unusual, but does happen). Furthermore, if the function a is known to be DC free over the period, that is, if

$$\int_0^T a(t)\,dt = \int_0^1 a(\theta T)d\theta = 0, \tag{1}$$

then we can compute the area Q_{ci} simply by finding a_{min} and a_{max} during the period, since

$$Q_{ci} = \int_0^1 e(\theta)d\theta = \int_0^1 \left(\frac{a(\theta T) - a_{min}}{A_{p-p}}\right)d\theta = \frac{1}{A_{p-p}}\left(\int_0^1 a(\theta T)d\theta - a_{min}\int_0^1 1 \cdot d\theta\right)$$
$$= \frac{0 - a_{min}}{A_{p-p}} = \frac{-a_{min}}{A_{p-p}}, \tag{2}$$

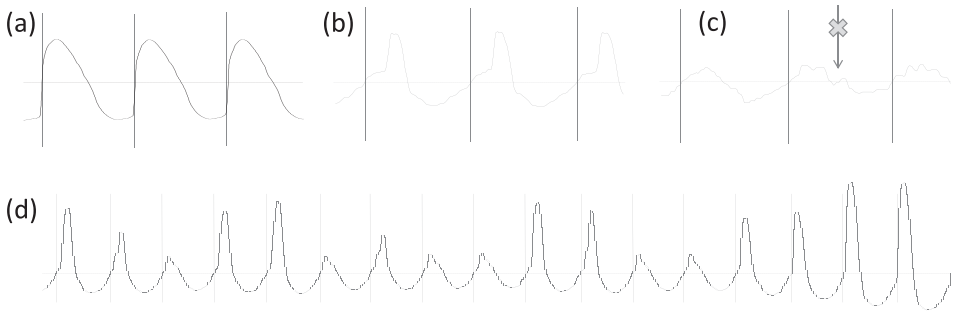

Fig. 3. Cycle segmentations: (a) the "standard" EGG shape: zero-crossings coincide with the maximum positive slope; (b) a common pattern of *partial* vocal fold contact: period markers remain at the zero crossing; (c) multiple zero-crossings due to noise are skipped, if in the decontacting part; (d) the period detector is not easily disrupted by rapid changes in amplitude, which are common at the threshold of vocal fold contacting.

where $A_{p-p} = a_{max} - a_{min}$ as before. Thanks to the preconditioning of the EGG signal, its DC component is very small, so $Q_{ci} \approx -a_{min}/A_{p-p,}$ requiring only tracking of a_{min} and a_{max} from one period marker to the next. This estimate of Q_{ci} may be inexact, due to incomplete low-frequency attenuation and to inevitable discretization artefacts in the digital domain. Such errors remain small if $f_s \gg f_{o,}$ and A_{p-p} is much larger than both the residual LF and the quantization step. In typical situations, both these conditions are comfortably met; and also, Q_{ci} will be averaged over very many EGG cycles.

In weak phonation, when the vocal folds are colliding only a little or not at all, $a(t)$ approaches a sine wave and Q_{ci} thus approaches 0.5. This remains inconvenient, and in itself offers no improvement over the earlier CQ definitions.

Most conventional variants of the contact quotient CQ_{EGG} are defined as the fraction of cycle *time* spent above a fixed amplitude threshold; for instance, 1/4 of A_{p-p}. A given CQ_{EGG} can therefore correspond to different amounts of VF contacting, depending on the shape of the pulse. With Q_{ci} we define the relative amount of contacting to be instead the area under the normalized pulse. A given value of Q_{ci} can of course correspond to different pulse shapes, but we argue that those shapes would represent a similar amount of contacting.

2.5 Amplitude-normalized peak derivative: Q_Δ

In very soft phonation, the EGG signal is very weak and practically sinusoidal. In principle, there are wave shapes, e.g., the triangle, with a smaller peak derivative than that of a sine wave. In practice, however, the maximum derivative for real EGG signals is unlikely to be less than that of the sine; and we have never seen any examples of this. It appears that any non-zero degree of VF contact will give rise to a peak derivative greater than that of the sine wave.

The sine function without any scaling has a period of 2π, and an amplitude of ± 1. Its derivative is the cosine function, which has a maximum amplitude of $+1$, at phase zero. From the period detector, we have obtained the period length T in integer samples, separated in time by the sampling interval $1/f_s$. Let δ_{max} be the largest positive difference observed over the period between two consecutive sample points in the discretized $a(t)$. We may then approximate the amplitude-normalized peak derivative to $Q_\Delta \approx 2\delta_{max}/[A_{p-p} \cdot \sin(2\pi/T)]$, which for a sinusoidal waveform of any period length T samples and any amplitude A_{p-p} will be very close to 1, if the number of sample points per period is large (say >20). For EGG signals obtained in phonation with any non-zero contacting, Q_Δ will be greater than 1 (Fig. 1, bottom right).

This definition of Q_Δ makes it the inverse, but for a scaling factor, of a similar metric suggested by Herbst,[9] as well as of the glottal *flow* parameter normalized amplitude quotient introduced by Alku,[10] and of the R_d parameter of the L-F model of glottal flow.[11] Q_Δ has the practical advantages (a) of increasing with increasing vocal effort; and (b) of approaching a minimum value of unity for the case of a sinusoidal signal, which corresponds to VF oscillation with minimal contacting, mainly at the anterior end of the VFs.

2.6 Index of contacting: I_c

The EGG gives no direct information on the position, amplitude or speed of the vocal folds. Still, it would be valuable to have a single number that in some sense represents the deceleration force with which the VF tissues collide. Ideally, this number should be high for strong phonation, and approach zero when, in very soft phonation, VF contacting ceases. Clearly, a high value of Q_{ci} will be obtained in pressed phonation, but a fairly high $Q_{ci} = 0.5$ can also indicate complete lack of contact. Similarly, a high value of Q_Δ does not necessarily indicate forceful VF collision. When Q_Δ is combined with Q_{ci}, however, these ambiguities can be heuristically resolved. If Q_{ci} and Q_Δ are both large, then the VFs are likely to be firmly adducted, the subglottal pressure is probably high, and the VF masses are probably subject to a large deceleration at collision. If Q_{ci} is small while Q_Δ is large, then we have some contacting, but probably gentle and brief collision. If Q_Δ is small (close to 1) while Q_{ci} is around 0.5, then we may infer that the VFs are vibrating without colliding.

One may now observe that the number $I_c = Q_{ci} \times \log(Q_\Delta)$ fits the wish list. When there is no VF collision, Q_Δ is 1, and then $\log(Q_\Delta)$ and hence I_c are zero. In strong phonation, Q_Δ is on the order of 10, so the base-10 logarithm is appropriate. It remains to relate this heuristic "index of contacting" more directly to the physical loading of the VF tissues. Such studies are being planned.

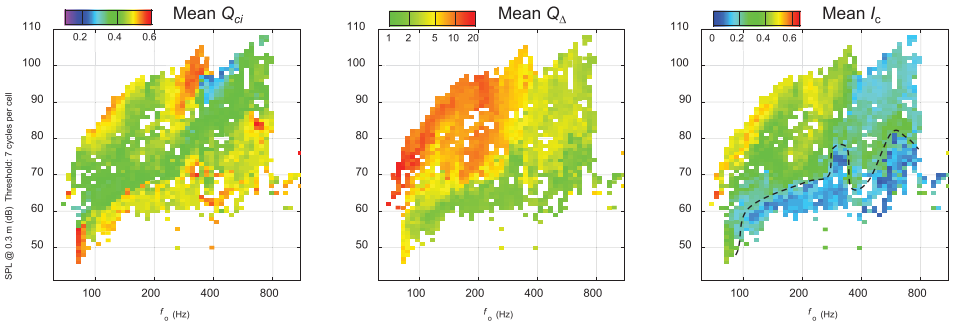

Fig. 4. Voice maps of Q_{ci}, Q_Δ, and I_c. An untrained male speaker vocalized on an [ɑ] vowel over his entire voice range, low to high, soft to loud. Vertical axis is SPL@0.3 m in dB(C), horizontal is f_o in Hz (log scale). Horizontal bars give the color mapping, which is different in each panel. Each cell (1 semitone × 1 dB SPL) represents an average for at least 7 but typically hundreds of EGG cycles. For further comments, see the text.

3. Application

To illustrate how the VF contacting behaviour of a person can be portrayed, Fig. 4 shows Q_{ci}, Q_Δ, and I_c in so-called "voice maps," over the entire voice range of a male subject vocalizing on the vowel /a/. Note how Q_{ci} and Q_Δ report on different aspects of VF contacting. These maps are quite consistent across repetitions within individuals, but can look very different for different individuals.[12] At the lowest sound pressure levels (SPLs), Q_{ci} increases to about 0.5 and Q_Δ descends nearly to 1, because the weak EGG signal from phonation without contacting approaches a sine wave. In the map of I_c, this is rendered as <0.1 (blue). The dashed line indicates approximately the SPL at which VF contacting is established.

4. Discussion and conclusion

The described EGG metrics $Q_{ci,}$ Q_Δ and I_c can quantify heuristically the degree of VF contacting, from zero to observed maximum, on the basis of pulse shape alone, disregarding gain fluctuations in the EGG signal chain.

Q_{ci} is bounded by definition to [0,…,1], while $1 \le Q_\Delta < \infty$ with no upper bound. If some parts of the medial faces of the vocal folds were to be perfectly flat and parallel on collision, then Q_Δ would go to infinity. In practice, Q_Δ is constrained by the bandwidth of the EGG signal. With our pragmatic 10 kHz brickwall limit, Q_Δ rarely exceeds 30. This 10 kHz limit works with all sampling rates ≥20 kHz and is a practical choice for real-time implementations. Since Q_{ci} rarely exceeds 0.7, this gives a practical maximum for I_c in the vicinity of $0.7 \times \log(30) \approx 1$, which is also convenient.

A noisy low-level EGG signal has small random dEGG spikes throughout the cycle. These are likely to prevent Q_Δ from fully reaching the desired lower bound of 1 for very soft phonation, when the SNR is poor. Median-filtering the original EGG signal helps, but only a little. Unfortunately, simply low-passing the EGG to remove such spikes will also introduce a waveform distortion that will change with f_o. If the EGG signal instead is constrained to its N lowest partials, by cycle-synchronous Fourier descriptor analysis and resynthesis,[4,12] then the approximate maximum of Q_Δ becomes N, as for a similarly constrained sawtooth wave. This will give an estimate of $Q_\Delta \le N$ that is far less sensitive to noise, descends truly to 1 for no contacting, and is comparable over changing values of f_o. The Fourier operations are however somewhat harder to achieve in real time.

The methods described here are implemented in an interactive real-time freeware called FONADYN,[13] which also does Fourier-based EGG classification, voice map displays, and can combine the EGG with other signals. FONADYN was released into the public domain in 2018, and an update is planned in 2019.

Acknowledgments

This work was carried out with KTH faculty funding, based on earlier work supported by the Swedish Research Council, Projects Nos. 2010-4565 and 2013-0642. Andreas Selamtzis and two anonymous reviewers offered valuable comments on the manuscript.

References and links

[1]R. J. Baken, "Electroglottography," J. Voice **6**, 98–110 (1992).
[2]C. T. Herbst, "Electroglottography—An update," J. Voice, in press (2019).

[3]C. Herbst and S. Ternström., "A comparison of different methods to measure the EGG contact quotient," Logoped. Phoniatr. Vocol. **31**, 126–138 (2006).
[4]A. Selamtzis and S. Ternström, "Analysis of vibratory states in phonation using spectral features of the electroglottographic signal," J. Acoust. Soc. Am. **136**, 2773–2783 (2014).
[5]S. B. Sunil Kumar, T. Mandal, and K. Sreenivasa Rao, "Robust glottal activity detection using the phase of an electroglottographic signal," Biomed. Sign. Process Control **36**, 27–38 (2017).
[6]C. T. Herbst, H. K. Schutte, D. L. Bowling, and J. G. Švec, "Comparing chalk with cheese—The EGG contact quotient is only a limited surrogate of the closed quotient," J. Voice **31**, 410–409 (2017).
[7]A. Macerata, A. Nacci, M. Manti, M. Cianchetti, J. Matteucci, S. O. Romeo, B. Fattori, S. Berrettinib, C. Laschi, and F. Ursino, "Evaluation of the electroglottographic signal variability by amplitude-speed combined analysis," Biomed. Sign. Process Control **37**, 61–68 (2017).
[8]L. O. Dolanský, "An instantaneous pitch-period indicator," J. Acoust. Soc. Am. **27**, 67–72 (1955).
[9]L. Enflo, C. T. Herbst, J. Sundberg, and A. McAllister, "Comparing vocal fold contact criteria derived from audio and electroglottographic signals," J. Voice **30**, 381–388 (2016).
[10]P. Alku, T. Bäckström, and E. Vilkman, "Normalized amplitude quotient for parametrization of the glottal flow," J. Acoust. Soc. Am. **112**, 701–710 (2002).
[11]G. Fant. "The LF-model revisited. Transformations and frequency domain analysis," Speech Trans. Lab. Quart. Prog. Stat. Rep. **36**, 119–156 (1995), available at http://www.speech.kth.se/prod/publications/files/qpsr/1995/1995_36_2-3_119-156.pdf.
[12]S. Ternström, S. D'Amario, and A. Selamtzis, "Effects of the lung volume on the electroglottographic waveform in trained female singers," J. Voice, in press (2018).
[13]S. Ternström, D. Johansson, and A. Selamtzis, "FonaDyn—A system for real-time analysis of the electroglottogram, over the voice range," SoftwareX **7**, 74–80 (2018).