This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Alku, Paavo; Murtola, Tiina; Malinen, Jarmo; Geneid, Ahmed; Vilkman, Erkki

Skewing of the glottal flow with respect to the glottal area measured in natural production of vowels

Published in:

Journal of the Acoustical Society of America

DOI:

[10.1121/1.5129121](https://doi.org/10.1121/1.5129121)

Published: 16/10/2019

Document Version

Peer reviewed version

Please cite the original version:

Alku, P., Murtola, T., Malinen, J., Geneid, A., & Vilkman, E. (2019). Skewing of the glottal flow with respect to the glottal area measured in natural production of vowels. *Journal of the Acoustical Society of America*, 146(4), 2501-2509. <https://doi.org/10.1121/1.5129121>

Skewing of the glottal flow with respect to the glottal area measured in natural production of vowels

Paavo Alku,^{1, a)} Tiina Murtola,¹ Jarmo Malinen,² Ahmed Geneid,³ and Erkki Vilkmán³

¹*Department of Signal Processing and Acoustics, Aalto University, Espoo, FI-00076, Finland*

²*Department of Mathematics and Systems Analysis, Aalto University, Espoo, FI-00076, Finland*

³*Department of Otorhinolaryngology and Phoniatrics - Head and Neck Surgery, Helsinki University Hospital and University of Helsinki, Helsinki, FI-00240, Finland*

(Dated: 16 October 2019)

In the production of voiced speech, glottal flow skewing refers to the tilting of the glottal flow pulses to the right, often characterized as a delay of the peak, compared to the glottal area. In the past four decades, several studies have addressed this phenomenon using modeling of voice production with analog circuits and computer simulations. However, previous studies measuring flow skewing in natural production of speech are sparse and they contain little quantitative data about the degree of skewing between flow and area. In the current study, flow skewing was measured from the natural production of 40 vowel utterances produced by 10 speakers. Glottal flow was measured from speech using glottal inverse filtering and glottal area was captured with high-speed videoendoscopy. The estimated glottal flow and area waveforms were parameterized with four robust parameters that measure pulse skewness quantitatively. Statistical tests obtained for all four parameters showed that the flow pulse was significantly more skewed to the right than the area pulse. Hence, this study corroborates the existence of flow skewing using measurements from natural speech production. In addition, the study yields quantitative data about pulse skewness in simultaneous measured glottal flow and area in natural production of speech.

©2019 Acoustical Society of America. [<http://dx.doi.org/DOI number>]

[XYZ]

Pages: 1–11

I. INTRODUCTION

The classical source-filter model of speech production assumes that the source (the glottal flow) and the filter (the vocal tract) form a linear system. In a linear system, there is no two-way interaction between the source and the filter, and the system output, the speech signal, is obtained as a convolution of the input and the impulse response of the filter. However, many studies conducted over the past four decades have shown that the production of voiced speech includes strong source-filter interactions. According to Titze (2008), source-filter interactions can be described on two levels. Level-1 interaction corresponds to the feedback effects of vocal tract pressures on the glottal flow. This interaction manifests itself in two ways in the glottal flow: as the occurrence of a ripple component in the glottal opening phase and as glottal flow skewing, a phenomenon seen as tilting of the glottal flow pulse to the right relative to the glottal area and generally characterized as a delay in the peak of the glottal flow relative to the peak of the glottal area. Level-2 interaction corresponds to the feedback effects of vocal tract pressures on the vocal fold tissue movements. Level-

2 becomes notable when a high-energy harmonic occurs in the vicinity of a formant resulting in disturbed vibration of the vocal folds, seen as sudden frequency jumps and chaotic vibration (e.g., Titze *et al.*, 2008; Zañartu *et al.*, 2011). While both interaction categories are always present in speech production, Level-1 can be considered more prevalent because its effects are easily detectable not only in high-pitched voices but in all speech (Titze, 2008). In addition, from the two manifestations of Level-1 interaction, glottal flow skewing is regarded as the primary effect (Titze and Palaparthi, 2016).

Skewing of the glottal flow relative to the glottal area is caused by the inertia of the air column of the vocal tract, as reported in several studies published in the 1980s using the theory of analog circuits in modeling human voice production (e.g., Ananthapadmanabha and Fant, 1982; Koizumi *et al.*, 1985; Rothenberg, 1981a). Since these early investigations, interactions between the source and tract have been elaborated in many computer simulation studies (e.g., Titze, 2004, 2008; Titze and Story, 1997; Zañartu *et al.*, 2007). These studies have demonstrated, for example, that skewing of the glottal flow due to inertial vocal tract loading is regulated to a large extent by the epilarynx tube diameter (Titze, 2004). In addition, previous simulation studies have investigated source-filter interaction from the point of voice produc-

^{a)}paavo.alku@aalto.fi

FIG. 1. (Color online) Pulse skewness and glottal flow skewing: By the convention used in speech research, a pulse with a positive skewness is said to be tilted or skewed to the left, while a pulse with a negative skewness is said to be tilted or skewed to the right (upper panel). Glottal flow skewing is the degree to which glottal flow (dashed) is more right-skewed than glottal area (solid) (lower panel).

tion efficiency, showing that skewing of the glottal flow is favorable in increasing the maximum flow declination rate of the voice excitation and thereby vocal intensity (Titze, 2008).

The majority of the previous studies on source-filter interaction are based on theoretical modeling of voice production and computer simulations. In addition, other approaches such as excised larynx experiments have been adopted in studying, for example, the effect of source-filter coupling on phonation threshold pressure (e.g., Smith *et al.*, 2013). Source-filter interactions have also been studied in natural production of speech, but only a few of these studies have focused on glottal flow skewing. These studies have used glottal inverse filtering (GIF) to estimate the glottal flow and high-speed filming or digital high-speed videoendoscopy (HSV) to capture the glottal area. When both glottal flow and area are studied simultaneously, it becomes necessary (see Fig. 1) to parameterize *pulse skewness* (i.e., the degree a pulse tilts either left or right). By quantifying pulse skewness, it is then possible to study *flow skewing*, that is, the degree to which a flow pulse is more tilted to the right (i.e., has a more negative skewness) than the area pulse of the same glottal cycle. In older studies (Berouti *et al.*, 1977; Childers *et al.*, 1985; Krishnamurthy and Childers, 1981), the glottal flow and glottal area estimates were not necessarily synchronized, which hindered the analysis of glottal flow skewing. In Rothenberg (1981b), glottal flow was computed using GIF with a flow mask, and the obtained waveform was compared to the synchronized glottal area signal, captured using photoglottography. The main aim of the study by Rothenberg (1981b) was the computation of the normalized inductance for breathy and normal phonations, but the study also demonstrated an example of a flow-area signal pair computed from a natural spoken sentence. In the study by Granqvist *et al.* (2003),

a modern digital HSV system was used with simultaneous GIF-based estimation of the glottal flow. The study demonstrated glottal flow skewing by plotting the glottal flow as a function of the glottal area using Lissajous figures. The Lissajous figures showed signs of non-linear interaction, but no quantitative analyses were computed to measure the degree of flow skewing.

It should be noted that all the previous investigations on flow skewing in the natural production of speech have not only adopted a qualitative, rather than a quantitative, approach in reporting the flow-area relationship, but they have also included only a small number of speakers. In the study by Rothenberg (1981b), only one speaker was included and in the study by Childers *et al.* (1985), four speakers were investigated but the flow signal was shown only for one speaker. Granqvist *et al.* (2003) studied the flow skewing in two speakers. Compared to the large number of computer simulation studies published on non-linear source-filter interaction, the small number of studies published on glottal flow skewing in natural production of speech can be explained by the following two reasons. First, studying glottal flow skewing in the natural production of voiced speech calls for simultaneous measurements of the glottal flow and glottal area that are difficult and time-consuming to conduct and call for special equipment, such as HSV. Second, since direct measurement of the glottal flow from natural production of speech is not feasible, the flow waveform must be estimated indirectly, commonly using GIF from the speech pressure signal. The available GIF methods are almost exclusively based on linear source-filter models. Therefore, the applicability of GIF in studying the non-linear skewing phenomenon might not be immediately obvious, thus explaining the smallish number of previous investigations studying glottal flow skewing in natural speech. However, the assumptions of GIF do not prevent its use in estimating glottal flow even if non-linear skewing is taking place, as long as the variable of interest is the skewed glottal flow, rather than its non-skewed, interaction-free counterpart. This rationale is explained further in Section II B.

In the current study, glottal flow skewing is studied in the natural production of vowels. Given that the flow-area relationship in natural speech has been addressed in previous studies qualitatively using a few voice samples, the present study aims to provide quantitative data from multiple speakers to measure the difference in the skewness of the glottal flow and glottal area. In order to estimate the glottal flow, the speech pressure signal recorded outside lips is inverse-filtered with a recently developed GIF method, quasi closed phase (QCP) analysis (Airaksinen *et al.*, 2014). The glottal area is measured from vocal fold videos captured using HSV. The hypothesis of the study is set as follows. Given that glottal flow skewing, caused by source-filter interaction, has been observed in many computer simulation studies (e.g., Titze, 2008; Titze and Palaparthi, 2016; Titze and Story, 1997), it is expected that this phenomenon would be observable also in natural production of speech as a larger tilt to the

right of the GIF-estimated glottal flow compared to the simultaneously recorded glottal area waveform.

II. MATERIAL AND METHODS

A. Data collection

The data used in this study is taken from a recently collected multichannel dataset (Alku *et al.*, 2019) consisting of simultaneous recordings of speech pressure signal, electroglottography (EGG) and HSV in vowel production. The data was recorded from 10 Finnish speakers (five males, five females) with no history of voice disorders. The ages of the speakers varied between 25 and 61 years, and the mean age was 32 years. Each speaker was instructed to produce six perceptually different variants of the vowel [i] by using two phonation types (normal and breathy) and three pitches (low, medium, and high). The speakers were allowed to choose comfortable degrees of breathiness and pitch levels. The vocalization was monitored by an experienced experimenter, who asked the subject to repeat the vocalization if a sufficiently large perceptual difference was not achieved. The speakers were asked to produce the vowel [i] with their tongues as far forward as possible so that the clearest possible view of the glottis could be obtained. Due to the hindrance caused by the HSV endoscope, there was phonemic variance in the produced utterances so that the recorded vowels ranged perceptually between Finnish phonemes [æ] and [œ]. The measurements took typically 2–3 h per speaker. The recordings resulted in a total of 60 utterances (10 speakers x 2 phonation types x 3 pitches), 200 ms each.

The setup used for recording the multichannel data is shown schematically in Fig. 2. The HSV recordings were made using the KayPentax Color High-Speed Video System (model 9710) with a rigid endoscope. Spatial and temporal resolution of the video images was 512 x 512 pixels and 2000 frames/s respectively. The speech pressure signal was recorded using an omnidirectional condenser headset microphone (DPA, model 4065-BL), which was placed approximately 6.5 cm from the center of the speaker's mouth. EGG was captured using a Glottal Enterprises electroglottograph (EG2-PCX2). Analog-to-digital conversion of speech and EGG was conducted using a MOTU UltraLite-mk3 Hybrid interface with a sampling rate of 44.1 kHz. A custom synchronization signal was played during each measurement and recorded with both the HSV video and the speech–EGG signal pair. This signal contained binary frequency-shift keyed code at the beginning of each second, enabling the alignment of all signals. The recorded EGG signals are not used in the present study, and therefore their processing is not described in the following.

Low-frequency disturbances were removed from the speech signals by filtering them with a linear-phase finite impulse response (FIR) filter with a cut-off frequency of 60 Hz. The speech signals were synchronized with the HSV videos by aligning the synchronization signals and

FIG. 2. Setup for simultaneous measurement of the speech microphone signal, HSV and EGG. In the current study, only the microphone signal and HSV are used.

shifting them to account for propagation delays and internal delays within and between the measurements systems. The delays were estimated to be approximately 1.6 ms for males and 1.5 ms for females. After the completion of this alignment, the maximum error in the synchronization of the speech signal to the video is ± 0.5 ms (one frame).

In the current study, some of the collected vowel samples were left out from the further analyses because the vocal folds were not completely visible, thereby affecting the estimation of the glottal area waveforms from HSV. Samples with $f_0 > 270$ Hz were also excluded as the quality of both the glottal area waveform (at the used frame rate) and glottal flow estimate suffer at high fundamental frequencies. After the exclusion of these samples, 25 samples from males and 15 from females remained as listed in Tables I–II. The remaining data comprises utterances from all the ten recorded speakers but for some of them (e.g., speaker M05 in Table I) one phonation type or some of pitch levels were excluded.

B. Computation of glottal flow

Glottal flow waveforms were estimated from the speech pressure signals using the QCP method (Airaksinen *et al.*, 2014). QCP is an inverse filtering method that is based on the linear source–filter model. QCP estimates the vocal tract using an all-pole modeling method called weighted linear prediction (WLP) (Ma *et al.*, 1993). In computing the WLP analysis, QCP takes advantage of the attenuated main excitation (AME) weighting function, which enables reducing the contribution of the glottal source in the computation of formant frequencies (Alku *et al.*, 2013). QCP was selected as the GIF method in the current study because it was shown in (Airaksinen *et al.*, 2014) to be the most accurate algorithm in comparison to four other widely used GIF methods.

It should be noted that even though QCP models speech production based on the classical linear source–filter model, this GIF method is used in the current inves-

tigation to study glottal flow skewing, i.e., a phenomenon that is produced in natural speech by non-linear interaction between the glottal source and vocal tract. Linear speech production models cannot naturally reveal how much a glottal flow skewing increases when the pulse is combined with an interactive tract compared to when it is combined with a non-interactive tract. However, if flow skewing occurs in the production of natural speech due to interaction by vocal tract inertia, the vocal tract is excited by the airflow pulse that has been skewed by the interaction and this skewed flow pulse *can* be estimated by linear GIF methods such as QCP. By comparing the estimated flow pulse to the simultaneously measured glottal area pulse, one should therefore be able to measure differences in the skewness of the glottal flow compared to the glottal area caused by inertia-induced source-filter interaction.

This justification is clarified in Fig. 3 using a simplified chart of vowel production. The chart describes two cases: a non-interactive vocal tract (Fig. 3(a)) and an interactive vocal tract (Fig. 3(b)). The latter includes non-linear source-filter coupling that causes skewing of the glottal flow, i.e., the phenomenon that is under scrutiny in this study. In the former, the glottal area pulse, denoted by $a(n)$, and the glottal flow pulse, denoted by $g_1(n)$, are of the same pulse as explained in Titze (2004) (see equation 18 in Titze, 2004). In the interactive case, the flow pulse, denoted by $g_2(n)$ in Fig. 3(b), tilts to the right due to vocal tract interaction, even though the area pulse is same as in Fig. 3(a). By assuming (ideally) that the vocal tract filter, denoted by $V(z)$, is all-pole and that QCP is able to compute the exact inverse model of the vocal tract, inverse filtering of the produced speech signal, denoted by $s_1(n)$, yields $g_1(n)$ in the non-interactive case. Due to the linearity assumption in inverse filtering, neither $g_1(n)$ nor $a(n)$ can be obtained with QCP, or with any other linear GIF method, in the interactive case. However, by using simultaneously recorded HSV in the interactive case, $a(n)$ can be measured from vocal fold videos. Therefore, by referring to the notations in Fig. 3(b), the current study first takes advantage of GIF to estimate $g_2(n)$ from $s_2(n)$ and then takes advantage of the simultaneously measured HSV to estimate $a(n)$ to compare the difference in skewness between glottal flow $g_2(n)$ and glottal area $a(n)$.

C. Computation of glottal area

The glottal area waveforms were extracted from the HSV data using the adapted seeded region growing method developed by Lohscheller *et al.* (2007). The extraction of the glottal area was conducted frame by frame from the red channel of the color video. The obtained area waveforms were manually inspected and, where necessary, corrected to counteract the effect of periodical changes in illumination caused by light reflected from the closed glottis.

Recordings of HSV and speech were conducted using different sampling frequencies (2 kHz and 44.1 kHz,

FIG. 3. A simplified chart of vowel production in the case of a non-interactive (a) and interactive (b) vocal tract. Glottal area pulse, denoted by $a(n)$, is same in (a) and (b). Due to source-filter interaction, glottal flow $g_2(n)$ in (b) is more skewed to the right than glottal flow $g_1(n)$ in (a). Skewing of glottal flow with respect to glottal area is measured in the current study as shown in (b) by estimating the skewed glottal pulse from $s_2(n)$ using GIF and by measuring area pulse $a(n)$ from simultaneously recorded HSV.

respectively). Therefore, the discrete-time speech and glottal area signals must be re-sampled to a common sampling frequency. Since 8 kHz sampling was used in Airaksinen *et al.* (2014) and since the inverse filtering parameters of the current study were adjusted according to Airaksinen *et al.* (2014), 8 kHz was selected as a common sampling frequency for speech and glottal area signals in the current study. Sampling rate conversion was conducted using MATLAB's function `resample` with default settings, i.e., an antialiasing low-pass FIR filter and delay compensation. It should be noted that according to the signal processing theory (e.g., Oppenheim and Schaffer, 1989, pp. 101–112), the sampling rate conversion of the glottal area waveforms corresponds to keeping the original frequency contents intact between 0 Hz and 1 kHz (i.e., Nyquist frequency of the original glottal area data) and to generating new, zero-valued frequency components between 1 kHz and 4 kHz (i.e., Nyquist frequency of the up-sampled glottal area data). However, for the speech signals which were originally represented using a much wider bandwidth (of 22.05 kHz), the use of 8 kHz sampling corresponds to narrowing the bandwidth to 0–4 kHz. Therefore, the signal bandwidth of the glottal flows estimated by QCP was also 0–4 kHz. In other words, after the sampling rate conversion, there was an inconsistency in the spectral contents in the 1–4 kHz range between the

glottal area and glottal flow signals: in the former, the spectrum was zero-valued whereas in the latter it was not. In order not to let this affect the comparison of skewness between the glottal flow and glottal area, both signals were finally low-pass filtered using a linear phase FIR with a cut-off frequency of 1 kHz.

D. Parameterization of glottal flow and area

In order to compare glottal flow and glottal area waveforms quantitatively, these time-domain signals were first expressed in a parametric form. Different parameterization methods have been developed, particularly for the glottal flow since the 1950s (see [Alku \(2011\)](#) for a review of the parameterization methods used in GIF), and many of the same methods have also been used in order to parameterize the glottal area ([Chen et al., 2013](#); [Mehta et al., 2011](#)). From the wide range of developed parameters, the current study applies four parameters (the quasi-speed quotient (QSQ), the normalized amplitude quotient (NAQ), the slope quotient (SIQ), and the moment coefficient of skewness (γ_1)) that are described in this section. These four parameters were selected for the current study because of the following rationales. First, in order to quantify skewness, the selected parameter should be capable of measuring the asymmetry of the glottal flow and area pulses. This can be done by considering a pulse as a whole or by utilising specific characteristics of skewness, such as the delay of the pulse peak or the increase in the steepness of the closing phase in right-skewed pulses. Second, the selected parameter should be robust with respect to known challenges in the analysis glottal pulses, such as the ambiguity in extracting the instants of glottal opening and closure. Extraction of the instant of glottal opening is known to be particularly problematic due to smooth, gradual opening of the vocal folds ([Dromey et al., 1992](#)). Third, in order to gather quantitative data about glottal flow skewing that can be easily compared to other voice production studies, it was justified to include two parameters (QSQ and NAQ) that are known in the study area. However, in order to have a more general and reliable comparison between the glottal flow and area waveforms, the parameter set was expanded by including one new parameter (SIQ) and another parameter (γ_1) from outside voice research.

1. Quasi-speed quotient

Speed quotient (SQ) measures the skewness of a pulse as the ratio of the opening and closing phase durations. The classical definition of SQ utilizes opening and closure instants, which often makes it challenging to use with noisy signals, such as glottal flow waveforms estimated using GIF. Quasi-speed quotient (QSQ), which utilizes threshold crossings instead of opening and closure instants, has been proposed as a more robust alternative (see, e.g., [Sapienza et al., 1998](#)). With a suitable threshold, QSQ can be computed identically from both glottal

flow and area pulses as

$$\text{QSQ} = \frac{t_{\max} - t_{qo}}{t_{qc} - t_{\max}}, \quad (1)$$

where $t_{\max}$ is the instant of the pulse peak, t_{qo} is the instant at which the signal crosses the threshold during the opening phase, and t_{qc} is the corresponding threshold crossing instant during the closing phase. A 20% threshold is used, following [Sapienza et al. \(1998\)](#), and it was checked that this threshold removes the segments around the closed phase where the methods of estimating glottal flow and area are known to produce different behavior (e.g., non-flat closed phase produced by GIF due to imperfect formant canceling vs. perfect closure observed in HSV). A symmetric flow or area pulse has $\text{QSQ}=1$, and a larger QSQ value indicates more tilting to the right (negative skewness) in the pulse. Hence, skewing of the flow is expected to manifest as $\text{QSQ}_U > \text{QSQ}_A$, where the subscript U indicates glottal flow and A glottal area.

2. Normalized amplitude quotient

Normalized amplitude quotient (NAQ) was developed by [Alku et al. \(2002\)](#) as a method to quantify the relative steepness of the glottal flow closing phase, and it quantifies skewness using the observation that a delayed pulse peak must be accompanied by a sharper decrease in the flow. NAQ is closely related to closing quotient (CIQ), an older and widely used method to quantify the glottal closing phase ([Monsen and Engebretson, 1977](#)). For a simplified triangular glottal pulse, these two parameters are equal, but in contrast to CIQ, the relative duration of the closing phase is computed in NAQ using amplitude values of the glottal flow and its first time-derivative. This yields the following definition

$$\text{NAQ} = \frac{f_{ac}}{T_o d_{\min}}, \quad (2)$$

where f_{ac} is the AC amplitude of the pulse, T_o is the fundamental period, and $d_{\min}$ is the absolute value of the minimum of the pulse derivative. NAQ has been shown to be more robust for noisy signals than classical time-domain quotients such as CIQ ([Alku et al., 2002](#)). A symmetric flow or area pulse with no closed phase has $\text{NAQ}=0.5$, and a smaller NAQ value indicates more tilting to the right (negative skewness) in the pulse. Hence, flow skewing is expected to manifest as $\text{NAQ}_A > \text{NAQ}_U$, where the subscript A indicates glottal area and U glottal flow.

3. Slope quotient

NAQ utilizes changes in the closing phase of a glottal flow or area pulse to quantify skewness, but these changes are also accompanied by changes in the opposite direction in the steepness of the opening phase. Another amplitude-based measure of skewness can hence be derived to make use of both of these changes simultaneously. NAQ was derived by [Alku et al. \(2002\)](#) by expressing a glottal flow pulse using a simplified triangular

waveform during the glottal open phase and a zero flow during the closed phase. The same approach can be used to define a slope quotient (SIQ)

$$\text{SIQ} = \frac{d_{\min}}{d_{\max}} \quad (3)$$

as described in Fig. 4. In this figure, the length of the glottal opening and closing phase is denoted by T_1 and T_2 , respectively. Given the triangular waveform of the pulse, the corresponding first derivative of the waveform is a rectangular pulse that is positive during the glottal opening phase and negative during the glottal closing phase. Since areas of both of the rectangular pulses must be equal to f_{ac} , the ratio between the length of the glottal opening and closing phase can be computed as follows:

$$T_1 d_{\max} = f_{ac} = T_2 d_{\min} \quad (4)$$

$$\Rightarrow \text{SIQ} = \frac{d_{\min}}{d_{\max}} = \frac{T_1}{T_2} = \text{SQ}. \quad (5)$$

It is worth noting that the latter half of Eq. (5) is true (i.e., SIQ is equal to SQ) only for the simplified triangular pulse. However, similarly to NAQ, SIQ can be computed from amplitude-domain values without extracting instants of glottal opening and closure. Therefore, this parameter is most likely a more robust measure of skewness than the classical SQ. SIQ indicates skewness in a manner similar to QSQ: A symmetric flow or area pulse has $\text{SIQ}=1$, and a larger SIQ value indicates more tilting to the right (negative skewness) in the pulse. Therefore, for flow skewing $\text{SIQ}_U > \text{SIQ}_A$ is expected where the subscript U indicates glottal flow and A glottal area. Finally, even though SIQ was derived above using the approach based on expressing the glottal flow and area with a simplified triangular pulse, a similar measure has been previously used by [Rothenberg \(1981b\)](#) where the quotient was called the slope ratio.

4. Moment coefficient of skewness

The three measures of skewness described above (QSQ, NAQ, SIQ) are all based on a parameterization scheme that is widely used in voice production studies, i.e., characterization of glottal flow and area pulses by measuring differences in their opening and closing phases using time durations or amplitude information. Skewness is, however, also a known measure characterizing probability density functions (PDFs). Several quantitative measures of skewness of PDFs exists, but the most commonly used parameter is the moment coefficient of skewness

$$\gamma_1 = \frac{\mu_3}{\sigma^3}, \quad (6)$$

where μ_3 is the third central moment of the PDF and σ is its standard deviation. The adaptation of γ_1 to quantify the skewness of glottal flow and area pulses requires defining a glottal pulse function $f(t)$ with properties equivalent to a PDF. For the ease of treatment, a continuous signal $y(t)$ containing glottal flow or area pulses is assumed in the following. In practice, $y(t)$ is replaced with the discrete flow or area waveforms.

FIG. 4. A simplified triangular pulse for the open phase of the glottal flow and area (upper pane) and its time-derivative (lower pane). The maximum flow or area amplitude is denoted by f_{ac} . The maximum and minimum of the flow or area derivative is denoted by $d_{\max}$ and $d_{\min}$, respectively. Glottal opening and closing phase durations are denoted by T_1 and T_2 , respectively.

In order to define $f(t)$, a start instant t_1 and an end instant t_2 are identified for each glottal pulse in $y(t)$. For this study, 20% level crossings are used (i.e., $t_1 = t_{qo}$ and $t_2 = t_{qc}$) as skewing is expected to mainly affect the upper part of the waveforms. The glottal pulse function is obtained by normalization

$$f(t) = \begin{cases} \frac{y(t) - \min_t y(t)}{\int_{t_1}^{t_2} (y(t) - \min_t y(t)) dt} & t \in [t_1, t_2], \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

The moment coefficient of skewness can then be computed from Eq. (6) using

$$\mu_3 = \int_{t_1}^{t_2} (t - \mu)^3 f(t) dt \quad \text{and} \quad (8)$$

$$\sigma = \left[\int_{t_1}^{t_2} (t - \mu)^2 f(t) dt \right]^{1/2}, \quad (9)$$

where mean μ is given by

$$\mu = \int_{t_1}^{t_2} t f(t) dt. \quad (10)$$

In contrast to the three other parameters, γ_1 can take positive and negative values. A negative value indicates that the pulse is tilted to the right (i.e., negative skewness) and a positive value that the pulse is tilted to the left (i.e., positive skewness), while $\gamma_1 = 0$ indicates a symmetric pulse. For flow skewing, $\gamma_{1,A} > \gamma_{1,U}$ is expected, where the subscript A indicates glottal area and U glottal flow.

III. RESULTS

Two examples of the extracted waveforms for the glottal flow (denoted by U) and glottal area (denoted

by A) and their derivatives are shown in Fig 5. Normal phonation (e.g., Fig. 5(a)) in the data typically exhibits complete or nearly complete glottal closure (i.e., the glottal area reached zero), a clearly defined d_{min} for both glottal flow and area pulses, and at least some degree of flow skewing is visible. Since the phonation task was to produce perceptually different phonation types, the degree of breathiness in the samples varied. In the most breathy samples (e.g., Fig. 5(b)), a large proportion of the glottis remains open throughout the glottal cycle (i.e., the glottal area is larger than zero at all times), pulse shapes are rounded with typically multiple local minima in the derivative, and while flow and area waveforms may differ, differences in their skewness are not visually obvious.

The values of the four skewness parameters extracted from glottal flow and area waveforms are shown in Fig. 6. Glottal flow skewing can be quantified by the difference between the parameters extracted from glottal flow and area waveform pairs. Hence, for each sample $i = 1, \dots, 40$, parameter differences are computed as

$$\Delta QSQ^i = QSQ_U^i - QSQ_A^i \quad (11)$$

$$\Delta NAQ^i = NAQ_U^i - NAQ_A^i \quad (12)$$

$$\Delta SIQ^i = SIQ_U^i - SIQ_A^i \quad (13)$$

$$\Delta \gamma_1^i = \gamma_{1,U}^i - \gamma_{1,A}^i \quad (14)$$

and these differences are summarized in Tables III–VI. The data showed that flow skewing happened in the vast majority of the compared flow–area sample pairs: for 33 of the 40 samples, regardless of gender or phonation task, *all* four parameters indicated that glottal flow is more right-skewed (has more negative skewness) than glottal area. Only in four samples (M02 breathy phonation at high pitch, M03 normal phonation at high pitch, M04 breathy phonation at high pitch, and M04 normal phonation at medium pitch), three of the parameters indicated the opposite, i.e., that the glottal area was more right-skewed than the glottal flow. In addition, 1–2 of the parameters indicated a similar lack of flow skewing in three further samples. It is worth noting, however, that even though the selected parameters reflect pulse skewness, individual parameters can also be affected by some other differences in the pulse waveforms. For example, M03 exhibited slightly asymmetric vocal fold oscillations at normal phonation with high pitch, resulting in relatively flat-peaked glottal area pulses and inaccuracies in identifying of the instants of pulse peaks for QSQ estimation. Since corresponding flat peaks were not present in the glottal flow signal, QSQ_A was much higher than QSQ_U for this sample.

To test whether the observed flow skewing is statistically significant, paired two-sample Wilcoxon tests were performed on the four skewness parameters. For each set of parameter pairs extracted from the flow and area waveforms, the Wilcoxon test tests the null hypothesis H_0 that the parameter difference (i.e., ΔQSQ , ΔNAQ , ΔSIQ , $\Delta \gamma_1$) is distributed symmetrically around zero against H_1 , which assumes that the distribution is not symmetrical around zero. The results (Table VII) indi-

cate that glottal flow skewing was significant at $p < 0.05$ for all the four parameters.

The degree to which the four parameters provided similar information about flow skewing was inspected by computing Spearman’s rank correlation coefficients for the parameter differences. The results (Table VIII) show that correlation coefficients are moderate but significant at $p < 0.05$. The highest correlation was observed between ΔQSQ and ΔSIQ , indicating correspondence between these two parameters even for non-triangular pulses extracted from natural production of speech. The moment coefficient of skewness shows consistent correlation with all the other three parameters. As γ_1 is based on integrating entire pulses, it is capable of encompassing all the features on which the other parameters are based but it is less responsive to small changes in pulse shapes.

IV. SUMMARY AND CONCLUSIONS

Glottal flow skewing refers to the tilting of the glottal flow pulse to the right compared to the glottal area pulse in the production of voiced speech. This phenomenon, caused by the inertia of the air column in the vocal tract, has been investigated in the past four decades in several studies by using the theory of analog circuits and computer simulations in modeling human voice production. Previous studies analyzing glottal flow skewing in natural production of speech are, however, sparse, and the existing studies have included only a few speech samples. Moreover, there is little, if any, quantitative data available about glottal flow skewing in natural production of speech. In the present study, glottal flow skewing was measured from 40 vowel utterances produced by 10 speakers using high-speed videoendoscopy to capture glottal area and glottal inverse filtering to estimate glottal flow. Skewness of the glottal flow and area pulses was quantified by using four parameters that are robust with respect to ambiguity caused by noise or gradual opening of the vocal folds. Results showed that all the four parameters indicated statistically significant skewing of the glottal flow relative to the area. The measured parameters gave, for the first time, quantitative data about flow skewing in natural production of vowels. For the normalized amplitude quotient and quasi-speed quotient (that are known parameters in voice research) the data showed that the flow pulse was 15% and 45%, respectively, more right-skewed (i.e., more negative skewness) compared to area pulse when the parameters were averaged over genders, phonation types, and pitch levels.

In conclusion, the results corroborate previous modeling and computer simulation studies in showing that skewing of the glottal flow relative to the glottal area is a phenomenon that can also be observed in production of natural speech. Unfortunately, the current study included only 40 of the original 60 samples recorded because vocal folds were not completely visible or the fundamental frequency was too high for reliable amplitude analysis in 20 utterances. Therefore, due to the remaining dataset being unbalanced, it was not possible to con-

FIG. 5. (Color online) Glottal area A (solid) and glottal flow U (dashed) waveforms (upper panes) and their corresponding derivatives (lower panes) for two samples: F03 with normal phonation and low pitch (a) and M02 with breathy phonation and low pitch (b).

FIG. 6. (Color online) Values of the four skewness parameters, quasi-speed quotient (QSQ), normalized amplitude quotient (NAQ), slope quotient (SIQ), and moment coefficient of skewness (γ_1), extracted from both the glottal area and flow waveforms. Subscript A indicates glottal area and U glottal flow. Line $y = x$ is shown for reference. Flow skewing manifests itself in dots that are located above the line in QSQ and SIQ and below the line in NAQ and γ_1 .

duct a more fine-grained statistical analysis to study, for example, whether phonation type, pitch or gender has an effect on the observed difference in skewness of the flow and area. To study the effect of these factors on flow

skewing in natural production of speech, new recordings are needed. The current pilot study, however, indicates that the combination of GIF and HSV can in principle be used to investigate the relationship between glottal flow and glottal area qualitatively in natural production of speech.

ACKNOWLEDGMENTS

This research was supported by the Academy of Finland (project number 312490).

- Airaksinen, M., Raitio, T., Story, B., and Alku, P. (2014). “Quasi closed phase glottal inverse filtering analysis with weighted linear prediction,” *IEEE/ACM Trans. Audio, Speech, and Language Process.* **22**(3), 596–607, doi: [10.1109/TASLP.2013.2294585](https://doi.org/10.1109/TASLP.2013.2294585).
- Alku, P. (2011). “Glottal inverse filtering analysis of human voice production - a review of estimation and parameterization methods of the glottal excitation and their applications,” *Sadhana* **36**(5), 623–650, doi: [10.1007/s12046-011-0041-5](https://doi.org/10.1007/s12046-011-0041-5).
- Alku, P., Bäckström, T., and Vilkmán, E. (2002). “Normalized amplitude quotient for parametrization of the glottal flow,” *J. Acoust. Soc. Am.* **112**(2), 701–710, doi: [10.1121/1.1490365](https://doi.org/10.1121/1.1490365).
- Alku, P., Murtola, T., Malinen, J., Kuortti, J., Story, B., Airaksinen, M., Salmi, M., Vilkmán, E., and Geneid, A. (2019). “OPENGLOT – an open environment for the evaluation of glottal inverse filtering,” *Speech Commun.* **107**, 38–47, doi: [10.1016/j.specom.2019.01.005](https://doi.org/10.1016/j.specom.2019.01.005).
- Alku, P., Pohjalainen, J., Vainio, M., Laukkanen, A.-M., and Story, B. H. (2013). “Formant frequency estimation of high-pitched vowels using weighted linear prediction,” *J. Acoust. Soc. Am.* **134**(2), 1295–1313, doi: [10.1121/1.4812756](https://doi.org/10.1121/1.4812756).
- Ananthapadmanabha, T. V., and Fant, G. (1982). “Calculation of true glottal flow and its components,” *Speech Commun.* **1**, 167–184.
- Berouti, M., Childers, D., and Paige, A. (1977). “Glottal area versus glottal volume-velocity,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 2,

- pp. 33–36, doi: [10.1109/ICASSP.1977.1170184](https://doi.org/10.1109/ICASSP.1977.1170184).
- Chen, G., Kreiman, J., Gerratt, B. R., Neubauer, J., Shue, Y.-L., and Alwan, A. (2013). “Development of a glottal area index that integrates glottal gap size and open quotient,” *J. Acoust. Soc. Am.* **133**(3), 1656–1666, doi: [10.1121/1.4789931](https://doi.org/10.1121/1.4789931).
- Childers, D. G., Naik, J. M., Larar, J. N., Krishnamurthy, A. K., and Moore, G. P. (1985). “Electroglottography, speech, and ultra-high speed cinematography,” in *Vocal Fold Physiology*, edited by I. R. Titze and R. C. Scherer (The Dancer Center For The Performing Arts, Denver, CO), pp. 202–220.
- Dromey, C., Stathopoulos, E. T., and Sapienza, C. M. (1992). “Glottal airflow and electroglottographic measures of vocal function at multiple intensities,” *J. Voice* **6**(1), 44–54, doi: [10.1016/S0892-1997\(05\)80008-6](https://doi.org/10.1016/S0892-1997(05)80008-6).
- Granqvist, S., Hertegård, S., Larsson, H., and Sundberg, J. (2003). “Simultaneous analysis of vocal fold vibration and transglottal airflow: exploring a new experimental setup,” *J. Voice* **17**(3), 319–330, doi: [10.1067/S0892-1997\(03\)00070-5](https://doi.org/10.1067/S0892-1997(03)00070-5).
- Koizumi, T., Taniguchi, S., and Hiromitsu, S. (1985). “Glottal source-vocal tract interaction,” *J. Acoust. Soc. Am.* **78**(5), 1541–1547, doi: [10.1121/1.392789](https://doi.org/10.1121/1.392789).
- Krishnamurthy, A., and Childers, D. (1981). “Vocal fold vibratory patterns: Comparison of film and inverse filtering,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 6, pp. 133–136, doi: [10.1109/ICASSP.1981.1171353](https://doi.org/10.1109/ICASSP.1981.1171353).
- Lohscheller, J., Toy, H., Rosanowski, F., Eysholdt, U., and Döllinger, M. (2007). “Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos,” *Med. Image Anal.* **11**(4), 400–413, doi: [10.1016/j.media.2007.04.005](https://doi.org/10.1016/j.media.2007.04.005).
- Ma, C., Kamp, Y., and Willems, L. (1993). “Robust signal selection for linear prediction analysis of voiced speech,” *Speech Commun.* **12**(1), 69–81, doi: [10.1016/0167-6393\(93\)90019-H](https://doi.org/10.1016/0167-6393(93)90019-H).
- Mehta, D. D., Zañartu, M., Quatieri, T. F., Deliyiski, D. D., and Hillman, R. E. (2011). “Investigating acoustic correlates of human vocal fold vibratory phase asymmetry through modeling and laryngeal high-speed videoendoscopy,” *J. Acoust. Soc. Am.* **130**(6), 3999–4009, doi: [10.1121/1.3658441](https://doi.org/10.1121/1.3658441).
- Monsen, R. B., and Engebretson, A. M. (1977). “Study of variations in the male and female glottal wave,” *J. Acoust. Soc. Am.* **62**(4), 981–993, doi: [10.1121/1.381593](https://doi.org/10.1121/1.381593).
- Oppenheim, A. V., and Schaffer, R. W. (1989). *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, New Jersey).
- Rothenberg, M. (1981a). “Acoustic interaction between the glottal source and the vocal tract,” in *Vocal Fold Physiology*, edited by K. N. Stevens and M. Hirano (University of Tokyo Press, Japan), Chap. 21, pp. 305–328.
- Rothenberg, M. (1981b). “Source-tract acoustic interaction in breathy voice,” in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, edited by I. R. Titze and R. C. Scherer (The Denver Center for the Performing Arts, Denver, CO), Chap. 37, pp. 465–481.
- Sapienza, C. M., Stathopoulos, E. T., and Dromey, C. (1998). “Approximations of open quotient and speed quotient from glottal airflow and EGG waveforms: Effects of measurement criteria and sound pressure level,” *J. Voice* **12**(1), 31–43, doi: [10.1016/S0892-1997\(98\)80073-8](https://doi.org/10.1016/S0892-1997(98)80073-8).
- Smith, B. L., Nemcek, S. P., Swinarski, K. A., and Jiang, J. J. (2013). “Nonlinear source-filter coupling due to the addition of a simplified vocal tract model for excised larynx experiments,” *J. Voice* **27**(3), 261–266, doi: [10.1016/j.jvoice.2012.12.012](https://doi.org/10.1016/j.jvoice.2012.12.012).
- Titze, I. R. (2004). “Theory of glottal airflow and source-filter interaction in speaking and singing,” *Acta Acust. united Ac.* **90**(4), 641–648.
- Titze, I. R. (2008). “Nonlinear source-filter coupling in phonation: Theory,” *J. Acoust. Soc. Am.* **123**(5), 2733–2749, doi: [10.1121/1.2832337](https://doi.org/10.1121/1.2832337).
- Titze, I. R., and Palaparthi, A. (2016). “Sensitivity of source-filter interaction to specific vocal tract shapes,” *IEEE/ACM Trans. Audio, Speech, and Language Process.* **24**(12), 2507–2515.
- Titze, I. R., Riede, T., and Popolo, P. (2008). “Nonlinear source-filter coupling in phonation: Vocal exercises,” *J. Acoust. Soc. Am.* **123**(4), 1902–1915, doi: [10.1121/1.2832339](https://doi.org/10.1121/1.2832339).
- Titze, I. R., and Story, B. H. (1997). “Acoustic interactions of the voice source with the lower vocal tract,” *J. Acoust. Soc. Am.* **101**(4), 2234–2243, doi: [10.1121/1.418246](https://doi.org/10.1121/1.418246).
- Zañartu, M., Mehta, D. D., Ho, J. C., Wodicka, G. R., and Hillman, R. E. (2011). “Observation and analysis of in vivo vocal fold tissue instabilities produced by nonlinear source-filter coupling: A case study,” *J. Acoust. Soc. Am.* **129**(1), 326–339, doi: [10.1121/1.3514536](https://doi.org/10.1121/1.3514536).
- Zañartu, M., Mongeau, L., and Wodicka, G. R. (2007). “Influence of acoustic loading on an effective single mass model of the vocal folds,” *J. Acoust. Soc. Am.* **121**(2), 1119–1129, doi: [10.1121/1.2409491](https://doi.org/10.1121/1.2409491).

TABLE I. Data for male speakers.

Speaker	Age	Phonation type	f_o (Hz)		
			Low	Medium	High
M01	36	normal	120	122	195
		breathy	107	98	122
M02	61	normal	94	118	-
		breathy	99	113	227
M03	28	normal	-	130	173
		breathy	107	147	204
M04	26	normal	114	152	178
		breathy	89	104	168
M05	29	normal	-	-	-
		breathy	94	109	219

TABLE II. Data for female speakers.

Speaker	Age	Phonation type	f_o (Hz)		
			Low	Medium	High
F01	26	normal	168	231	-
		breathy	163	154	-
F02	25	normal	-	256	-
		breathy	-	-	-
F03	28	normal	179	229	-
		breathy	150	188	242
F04	40	normal	186	-	-
		breathy	183	-	-
F05	25	normal	185	-	-
		breathy	213	243	-

TABLE III. Summary of the difference ΔQSQ (mean, standard deviation, and range) and the mean of the relative difference $|\Delta QSQ|/QSQ_A$.

Gender	Task	n	Mean	SD	Range	Mean of the relative difference
Male	normal	10	0.224	0.405	-0.584 – 0.956	43%
	breathy	15	0.360	0.315	-0.183 – 1.094	52%
Female	normal	7	0.667	0.232	0.412 – 1.139	58%
	breathy	8	0.341	0.190	-0.013 – 0.569	35%

TABLE IV. Summary of the difference ΔNAQ (mean, standard deviation, and range) and the mean of the relative difference $|\Delta NAQ|/NAQ_A$.

Gender	Task	n	Mean	SD	Range	Mean of the relative difference
Male	normal	10	-0.011	0.036	-0.070 – 0.057	16%
	breathy	15	-0.060	0.046	-0.141 – 0.022	20%
Female	normal	7	-0.046	0.030	-0.097 – -0.007	20%
	breathy	8	-0.050	0.034	-0.089 – 0.000	18%

TABLE V. Summary of the difference ΔSIQ (mean, standard deviation, and range) and the mean of the relative difference $|\Delta SIQ|/SIQ_A$.

Gender	Task	n	Mean	SD	Range	Mean of the relative difference
Male	normal	10	0.470	0.504	-0.285 – 1.431	65%
	breathy	15	0.458	0.321	-0.162 – 0.992	66%
Female	normal	7	0.696	0.118	0.548 – 0.914	63%
	breathy	8	0.485	0.226	0.150 – 0.862	51%

TABLE VI. Summary of the difference $\Delta \gamma_1$ (mean, standard deviation, and range). Relative differences are not computed since $\gamma_{1,A}$ ranges around zero.

Gender	Task	n	Mean	SD	Range
Male	normal	10	-0.147	0.137	-0.369 – 0.010
	breathy	15	-0.188	0.109	-0.371 – 0.015
Female	normal	7	-0.184	0.081	-0.343 – -0.088
	breathy	8	-0.129	0.087	-0.245 – 0.021

TABLE VII. Results of two-sample Wilcoxon tests (pooled data with $n = 40$ paired samples) on whether the difference between skewing in glottal flow is significantly different from the skewing in the glottal area.

Parameter	Test statistic	p-value	Direction of effect
QSQ	48	$< .001$	$QSQ_U > QSQ_A$
NAQ	762	$< .001$	$NAQ_U < NAQ_A$
SIQ	16	$< .001$	$SIQ_U > SIQ_A$
γ_1	809	$< .001$	$\gamma_{1,U} < \gamma_{1,A}$

TABLE VIII. Spearman's rank correlation coefficients between parameter differences for pooled data ($n = 40$). All correlation coefficients are significant at $p < 0.05$.

	ΔQSQ	ΔNAQ	ΔSIQ
ΔNAQ	-0.404		
ΔSIQ	0.739	-0.480	
$\Delta \gamma_1$	-0.639	0.518	-0.589