The SIGMA Algorithm: A Glottal Activity Detector for Electroglottographic Signals

Mark R. P. Thomas, *Student Member, IEEE*, and Patrick A. Naylor, *Senior Member, IEEE*

***Abstract*—Accurate estimation of glottal closure instants (GCIs) and opening instants (GOIs) is important for speech processing applications that benefit from glottal-synchronous processing. The majority of existing approaches detect GCIs by comparing the differentiated EGG signal to a threshold and are able to provide accurate results during voiced speech. More recent algorithms use a similar approach across multiple dyadic scales using the stationary wavelet transform. All existing approaches are however prone to errors around the transition regions at the end of voiced segments of speech. This paper describes a new method for EGG-based glottal activity detection which exhibits high accuracy over the entirety of voiced segments, including, in particular, the transition regions, thereby giving significant improvement over existing methods. Following a stationary wavelet transform-based preprocessor, detection of excitation due to glottal closure is performed using a group delay function and then true and false detections are discriminated by Gaussian mixture modeling. GOI detection involves additional processing using the estimated GCIs. The main purpose of our algorithm is to provide a ground-truth for GCIs and GOIs. This is essential in order to evaluate algorithms that estimate GCIs and GOIs from the speech signal only, and is also of high value in the analysis of pathological speech where knowledge of GCIs and GOIs is often needed. We compare our algorithm with two previous algorithms against a hand-labeled database. Evaluation has shown an average GCI hit rate of 99.47% and GOI of 99.35%, compared to 96.08 and 92.54 for the best-performing existing algorithm.**

***Index Terms*—Electroglottograph (EGG), glottal closure instants (GCIs), group delay function, laryngograph.**

I. Introduction

ALL voiced sounds are produced by an excitation signal that is filtered by a passive resonator called the vocal tract. This excitation is produced by the vocal folds, which consist of opposing ligaments that form a constriction at the top of the trachea as it joins the lower vocal tract. When air is expelled from the lungs at sufficient velocity through this orifice—often referred to as the glottis—the Bernoulli Effect results in a partial vacuum that causes the vocal folds to snap shut and disrupt the air flow. This glottal closure instant (GCI) is followed by a period during which the glottis is closed, until muscle tension and air pressure cause the folds to reopen at the glottal opening instant (GOI). The process repeats periodically as a series of pulses that produces "modal" voiced speech. The ratio of open time with respect to glottal period is termed the open quotient (OQ) [1].

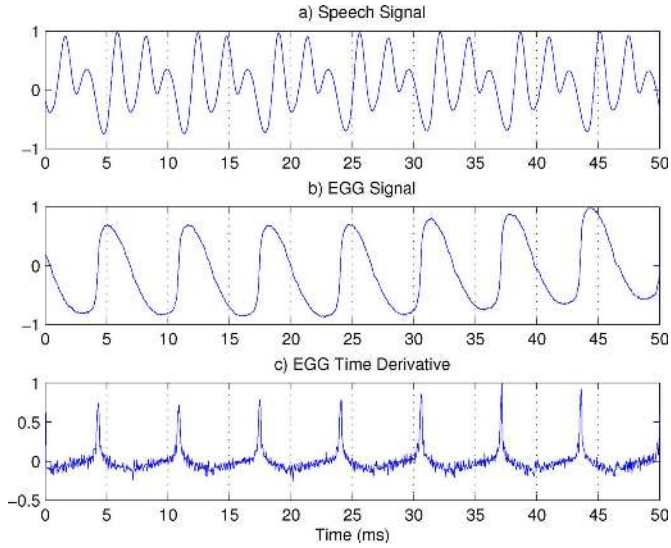

Fig. 1. (a) Speech signal, (b) the corresponding EGG signal, and (c) the EGG time derivative for /ɑ/. Negative peaks due to glottal opening are weak in (c).

Identification of GCIs in voiced speech is important for speech processing algorithms such as pitch tracking [2], prosodic speech modification [3], speech dereverberation [4], glottal-synchronous processing in speech synthesis [5] and voice source modeling [6]. Identification of GOIs is necessary for closed-phase LPC analysis and subsequent inverse filtering to obtain an estimate of glottal volume flow from a speech signal [7] for applications such as feature extraction for speaker identification [8]. Further uses are found in the analysis of pathological speech, including types of dysphonia [9], vocal fold impact stress [10] and essential tremor [11]. We refer to GOIs and GCIs as *glottal activity*.

The Electroglottograph (EGG) (or Laryngograph) signal [12] is a measurement of the electrical conductance of the glottis captured contemporaneously with speech recordings. The measured EGG signal is proportional to the glottal contact area, whose derivative (DEGG) during voiced speech contains short, high-amplitude impulsive temporal features (spikes) due to glottal closure and smaller features of opposite sign due to glottal opening. An example of a voiced speech segment, the corresponding EGG recording and its derivative is shown in Fig. 1. Many approaches analyze the EGG by searching for spikes in DEGG [13]–[16] and compare their amplitudes with thresholds to obtain an estimate of glottal activity during voiced

Manuscript received September 30, 2008; revised March 10, 2009. First published May 08, 2009; current version published August 14, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hiroshi Sawada.

The authors are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: mark.r.thomas02@imperial.ac.uk; p.naylor@imperial.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2009.2022430

speech. Recent approaches have applied multiscale analysis to detect glottal activity as singularities in the EGG signal [17] and speech signal [18]. Existing techniques are, however, often prone to errors around the end of voicing as discussed in Section II.

In cases where only the speech signal is available, new algorithms have recently been proposed which estimate glottal activity from the speech signal alone [19]–[23], and this is an ongoing topic of research with seemingly ever-improving results. These algorithms enable glottal activity information to be determined in real-world applications in which, typically, the EGG signal is not available. However, as such methods improve, their evaluation requires ever more accurate references. This requirement, alongside the application to the study of pathological speech, further motivates the development of better EGG-based detection algorithms.

This paper describes the Singularity in EGG by Multiscale Analysis (SIGMA) algorithm. SIGMA benefits from the use of multiscale processing but it novely extends the approach by performing spike detection on the multiscale product using a group delay method [24] which circumvents the need for thresholding. The robustness of our approach to false detections is further enhanced by Gaussian mixture modeling [25] which is used to remove detections with unlikely features. The proposed method provides GCI estimation with outstandingly high accuracy which also achieves similarly accurate GOI detection. Additionally, the algorithm makes no assumptions about the nature of the EGG signal other than the bounds on the range of glottal frequency and open quotients [26]; SIGMA may therefore have many further uses as it is also suitable for singularity detection in applications outside the field of speech processing.

This paper is organized as follows. Section II reviews the characteristics of the EGG signal and the methodology employed by some existing algorithms. Section III describes multiscale analysis, the use of the group delay function and Gaussian mixture modeling for spike detection in the multiscale product. The proposed SIGMA algorithm is compared with existing techniques and evaluated in Section IV. Conclusions are drawn in Section V.

II. Interpreting EGG Signals

A. HQTx and TXGEN

In Section IV, the performance of SIGMA is compared with two existing algorithms: High Quality Time of excitation (HQTx) and Time of eXcitation GENerator (TXGEN) [16]. The following is a brief description of their operation.

HQTx uses two derived functions: DEGG and an estimation of instantaneous gradient. A threshold function varies dynamically with the EGG signal, whose minimum is set by periods of silence assumed to lie during the first and last 20 ms of the EGG recording. The instants of time when the DEGG and instantaneous gradient exceed this threshold are the estimated GCIs.

TXGEN uses a more straightforward approach but attempts to detect both GCIs and GOIs. After low-pass filtering the EGG signal at 3 kHz, it is differentiated to find DEGG. High and low thresholds are set by the extrema of the DEGG signal from the entire recording multiplied by constant-valued coefficient. If DEGG passes through both thresholds within a set period of time, an estimated GCI is flagged. A GOI is the point in the EGG signal whose amplitude is equal to the amplitude at the preceding GCI.

B. Detection Errors

Both SIGMA and the algorithms described are evaluated against a large hand-labeled database. The remainder of this section describes common features of the EGG signal, those cases where interpretation of the EGG signal requires clarification and the resulting errors made by existing algorithms.

A voiced speech signal, its corresponding time-aligned EGG signal and the EGG derivative are shown in Fig. 1. Time alignment is achieved by ensuring that the lip-microphone propagation distance plus an estimate of the length of the talker's vocal tract is a constant value, then subtracting the corresponding delay. We define a positive EGG signal to be high glottal contact area, giving positive- and negative-going transients for GCIs and GOIs, respectively, with corresponding spikes in the EGG derivative.

Errors in GCI detection can be divided into two categories [19]: *False alarm* errors are made when more than one GCI is detected within a reference cycle; *Miss* errors are made when no GCI is detected within a reference cycle (GOI errors are treated in the same manner). Errors occur when certain types of EGG signal, discussed in the following sections, cause a poor estimate of the signal thresholds described in Section II-A.

C. "False Alarm" Errors

It has been shown that, for normal "modal" voiced speech, the frequency of oscillation of the glottis and the open quotient are dependent on phoneme and voice quality [12], [14]. Studies have further revealed that, for a given talker, the difficulty of detecting glottal closure is largely independent of the sound produced but that interesting effects occur at the boundaries of voiced/unvoiced speech, noting in particular [27]:

1) "Vocal fold vibration does not stop abruptly at the end of voicing, but slowly decays as the vocal folds come to a rest position."
2) "It is possible for vocal fold vibration to continue without the generation of any significant energy," termed "breathy offsets" [28].

This is examined in greater detail in [28], where a third phenomenon is observed at the end of voicing.

1) "A persistence of energy in the speech waveform after the EGG waveform has dropped virtually to zero," termed "breathy voice."

In the case of breathy offsets, GCIs can be detected from the EGG long after the speech amplitude has significantly diminished as the EGG signal remains modal, with increasing open phases that result in a breathier sound [28]. This is demonstrated in Fig. 2, showing 14 cycles of breathy offset terminating in breathy voice when EGG signal finally loses modality.

In the case of breathy voice, observed throughout case (3) and at the very end of case (2), the glottis is "flapping in the breeze" [29] with insufficient contact to register on the EGG waveform. As described in [30], "If the glottis does not shut

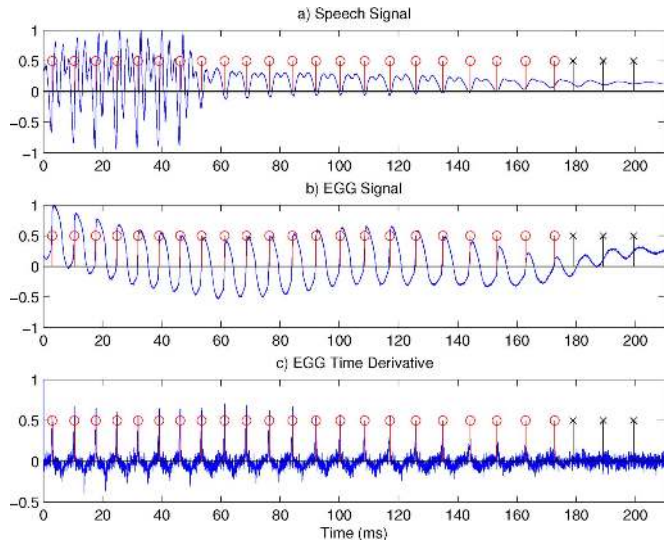

Fig. 2. (a) Speech signal, (b) EGG signal, and (c) its time derivative with overlayed HQTx GCI estimation markers at the end of a voiced speech segment, /u/, exhibiting "breathy offset" (cycles 8–21) and briefly "breathy voice." The first 22 GCIs are identified correctly (marked "o") but the last three (marked "×") are erroneous.

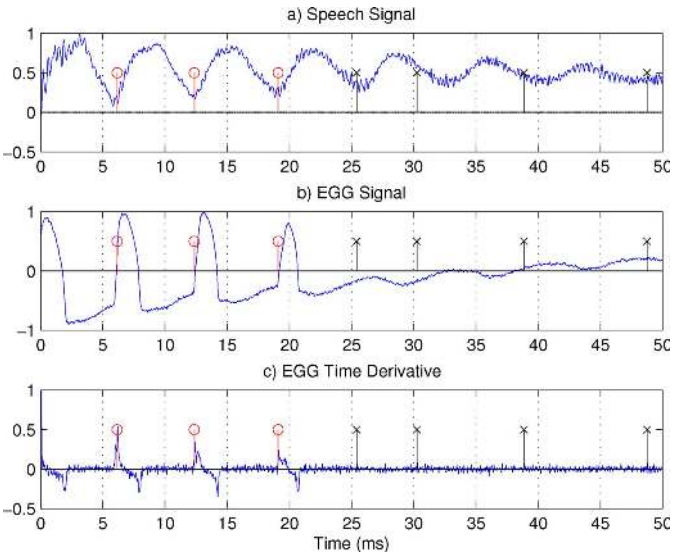

Fig. 3. (a) Speech signal, (b) EGG signal, and (c) its time derivative with overlayed HQTx GCI estimation markers at the end of a voiced speech segment, /I/, exhibiting "breathy voice." The first three GCIs are identified correctly (marked "o") but the last four (marked "×") are erroneous. Negative peaks due to glottal opening are significant in (c).

quickly enough ... no vocal wave is generated in the supraglottic cavity," and is demonstrated in Fig. 3. In both cases, a number of erroneous GCIs are detected by HQTx during segments of breathy voice (×) until its dynamic threshold is no longer exceeded. These errors also often occur at erratic intervals. For the hand-labeled reference, marked "o," the labeler would not mark any GCIs where there is no visible instant defining the periodicity, as would be the case with all instances of breathy voice.

Breathy voice represents a natural transition from modal voiced speech to unvoiced or silence [28]. It is further noted

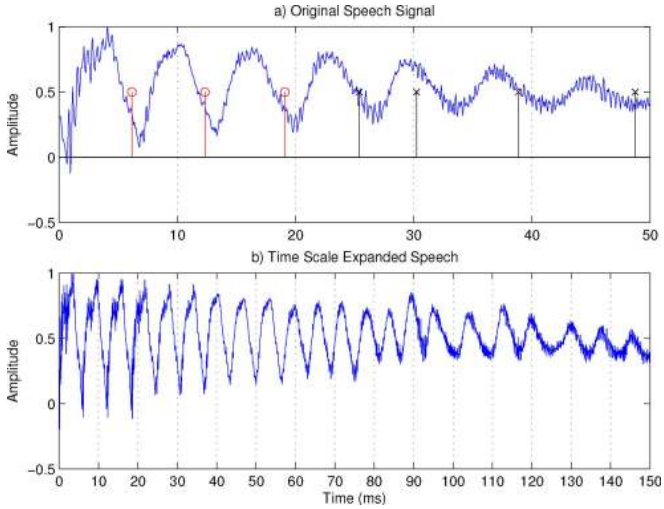

Fig. 4. (a) Original Speech signal with correct GCIs (marked "o") and false alarm errors (marked "×") and (b) time scale expanded by three times with the PSOLA Algorithm. Voiced cycles are copied and concatenated to increase duration; this works well for modal speech but fails when GCIs are detected in the wrong location.

that this usually lasts for just a few cycles of speech but erroneous estimates by a GCI detector during these segments can cause significant problems for glottal-synchronous algorithms. For example, a pitch tracker [2] that calculates pitch on a cycle-by-cycle basis will give highly erratic results. Glottal-synchronous speech processing algorithms such as prosodic speech modification [3], speech dereverberation [4], speech synthesis [5], and voice source modeling [6] all rely upon the manipulation of individual cycles of speech. Any fricatives or plosives following segments of voiced speech will be treated as periodic, giving rise to particularly annoying artefacts [31].

An example is shown in Fig. 4 where HQTx is used to drive the PSOLA algorithm [3] to increase the duration of a speech signal by three times without affecting prosody or formant structure. Applications for increasing the duration of a speech signal include enhancing intelligibility and lip synchronization in motion video. It is achieved by repeating cycles of voiced speech and concatenating them with an estimate of the correct period as shown in the first 70 ms of Fig. 4(b). Unvoiced speech and voiced-unvoiced transitions do not exhibit such periodicity so a common approach is to leave these segments unmodified [31]. This is not the case due to the erroneous detections at the voiced-unvoiced transition from 70–150 ms, leading to strange artefacts that detract from the otherwise natural sound of the processed voiced speech segments.

Sudden changes in EGG amplitude can also cause false alarm errors in dynamic threshold-based algorithms if the threshold is too low. A further problem with dynamic thresholds arises when GCIs have slow rise times [17], causing not a spike but a spread pulse in the DEGG. In this case, we define the GCI as the center of energy of the pulse.

D. *"Miss" Errors*

A common feature at the end of voiced segments is a reduced EGG signal amplitude compared with normal modal voice.

TXGEN's thresholds are proportional to the extrema of the entire signal and it is generally not prone to the false alarm errors exhibited by HQTx. It instead gives miss errors where the EGG amplitude is consistently low, particularly at the very beginning and very end of voiced speech segments. For the majority of glottal-synchronous algorithms this does not pose a significant problem. If, however, the amplitude of the EGG signal momentarily drops below the fixed threshold, TXGEN can miss a small number of isolated cycles which can be problematic for certain applications. Data-Driven Voice Source Modeling [6], for example, derives feature vectors from individual of cycles of voiced speech which are then clustered to determine classes of voice source. This has been demonstrated to have applications in speech compression [6] and artificial bandwidth extension [32]. A missed GCI results in features being derived from multiple cycles of speech, causing misclassification and distorting the processed signal.

HQTx can exhibit miss errors following a sudden decrease in EGG amplitude due to smoothing of the dynamic threshold that is not employed in TXGEN.

E. False Alarm/Miss Tradeoff

In general, HQTx is prone to false alarm errors, particularly at the end of voiced segments. This is verified in Section IV; it is further shown that miss errors are far less common. In contrast, TXGEN is generally prone to miss errors with relatively few false alarms; this is also verified in Section IV.

HQTx fails largely because thresholds are estimated over too short a window and TXGEN because thresholds are based upon single global thresholds for the whole speech utterance. The constant of proportionality used to set the threshold from signal extrema can be varied in TXGEN's function call. The default was empirically chosen to give the best tradeoff between miss and false alarm errors; a marginally lower value can result in increased false alarms and decreased misses. There is therefore a clear tradeoff between false alarms and misses caused by the thresholding approach employed by the majority of existing algorithms. The severity of this type of error is application-specific but, when used as a reference to evaluate speech-based GCI/GOI detectors, neither should be deemed acceptable. SIGMA instead employs a novel method for detecting GCIs and GOIs that does not use thresholding, circumventing the false alarm/miss tradeoff and providing accurate estimates for the entire EGG signal.

F. EGG at Glottal Opening

A glottal closure instant is usually followed by a GOI, which manifests itself as a weaker spike of opposite sign in the EGG derivative [13] whose amplitude is largely speaker-dependent. GOI detection suffers from the same problems as GCI detection but is more challenging because of the low amplitude of the opening pulses. Compare the negative halves of the EGG signals in Figs. 1 and 3. In Fig. 1, glottal opening results in spread pulses, hence the concept of an opening *phase* rather than opening instant is often used. However, as in the case of a spread-pulse GCI, we consistently define the GOI as its center of energy. Fig. 3 represents a speaker for whom the opening spikes in DEGG are easier to locate.

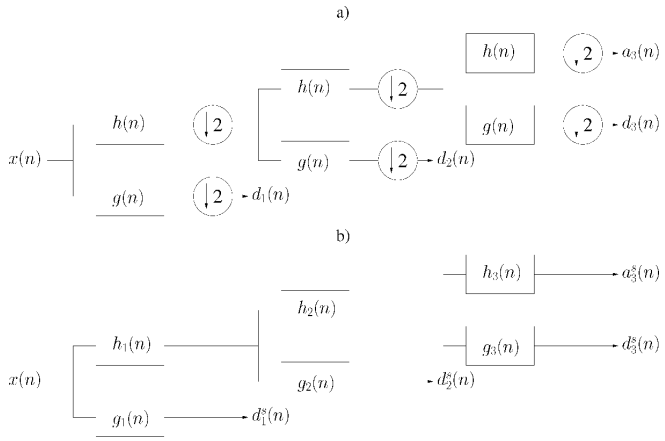

Fig. 5. Three-level dyadic signal decomposition on a signal $x(n)$ into detail, $d_j(n)$, and approximation, $a_j(n)$, signals. (a) is the Dyadic Wavelet Transform (DWT), and (b) the Stationary Wavelet Transform (SWT), an overcomplete version of the DWT useful in the detection of discontinuities.

III. Glottal Activity Detection With the Sigma Algorithm

Detection of glottal activity from an EGG signal involves isolating regions of discontinuity, sometimes referred to as singularities. A common approach is the detection of spikes in the derivative of the EGG signal, whose estimates are refined using the peak amplitude of the DEGG and a longer-term measure of the change in EGG amplitude as a cost function.

A. Multiscale Analysis

Let us consider a generalization of the HQTx approach that employs two estimates of signal gradient. The dyadic wavelet transform [33] involves iteratively decomposing a signal $x(n)$ into decimated subbands; a three-level decomposition is shown in Fig. 5(a), where the downsampling and filtering operations split the signal into octave-wide subbands.

The filters $g(n)$ and $h(n)$ have high- and low-pass characteristics, respectively. It is shown in [34] that, for singularity detection in EGG signals, each filter in the filterbank should be a first-order differentiation operator at increasing levels of smoothing. A wavelet fulfilling this criteria is described as having one vanishing moment and discontinuities in the input signal are seen as converging maxima across scales $d_j(n)$[35].

A derivative-of-Gaussian (dG) approximation with cubic spline wavelet decomposition filters is used in [36] and [34] which provides the differentiation and smoothing we require. However, an arbitrary number of filters exist which fulfil the same criteria. A number of derivations can be found in [37] but give little idea as to their use in the detection of singularities. In order to determine the relative performance, the proposed algorithm was run with five different sets of decomposition filters. Section IV-C presents a performance comparison between the chosen wavelet, whose filters are shown in Fig. 6, and the popular cubic spline dG wavelet.

The dyadic wavelet transform is dyadic in both scale and time. Only scale is of interest in singularity detection, so we do not decimate as shown in Fig. 5(b). Instead, the filters $g(n)$ and $h(n)$ are upsampled by 2 at each iteration to implement the

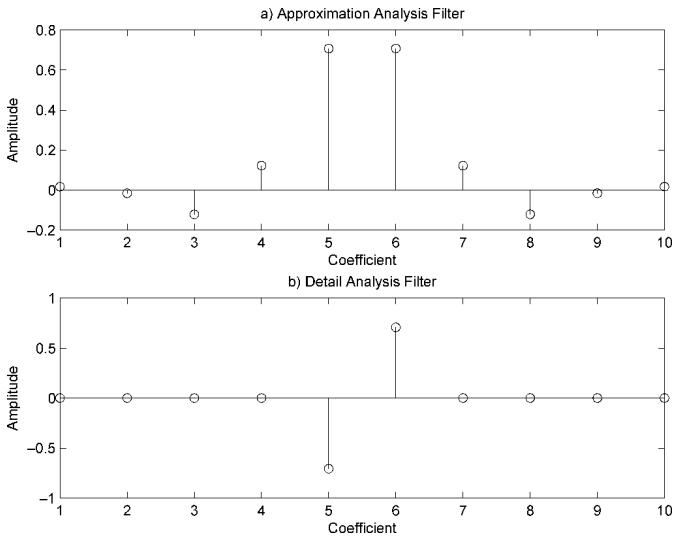

Fig. 6. (a) Approximation and (b) detail analysis filters for multiscale analysis. Iterating these filters through a dyadic filterbank constructs a biorthogonal spline wavelet with one vanishing moment.

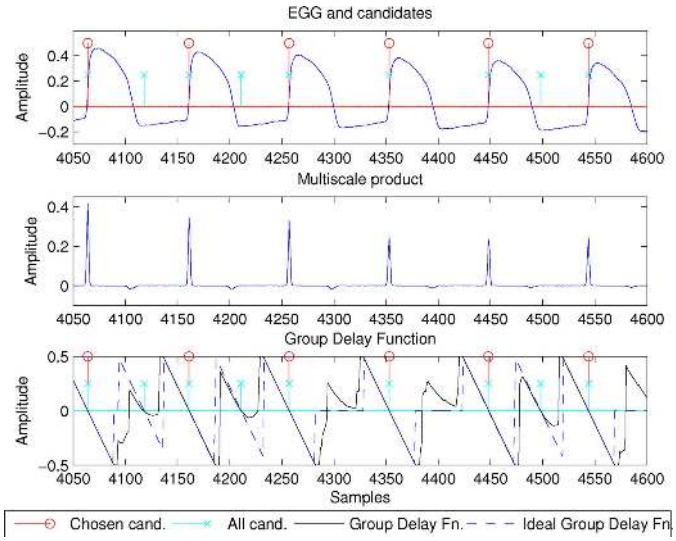

Fig. 7. EGG waveform, multiscale product and group delay function for GCI detection. Candidates are marked "×" and chosen candidates are marked "o." The ideal slope, marked in a dashed line on the lowest plot, is the slope which would exist if the candidates were perfect impulses.

change of scale to form $g_j(n)$ and $h_j(n)$ at scale j. This overcomplete representation of a signal is discussed in detail in [35] and is given many names including: *Stationary Wavelet Transform (SWT)*, *Algorithme à Trous* (Hole Algorithm), *Redundant Wavelet Transform (RWT)* and *Undecimated Wavelet Transform (UWT)*. The signal's length remains unchanged throughout the filterbank tree, allowing simple sample-by-sample multiplication of the signal at different scales to find converging maxima.

Denote the wavelet $\psi_s(t) = (1/s)\psi(t/s)$, where $s = 2^j, j \in \mathbb{Z}$. The SWT of the EGG signal at scale j is

$$d_j^s(n) = W_{2^j}x(n), j = 1, 2, \ldots, J-1 \tag{1}$$

where $J = \log_2 N$, plus the remaining coarse scale information denoted $a_j^s(n)$. This is a simple linear filtering operation

$$d_j^s(n) = W_{2^j}x(n) = \sum_k g_j(k)a_{j-1}^s(n-k) \tag{2}$$

where $d_j^s(n)$ is the SWT of $x(n)$ at scale j and a_{j-1}^s are the approximation coefficients at scale $j-1$. The multiscale product, $p(n)$, is formed by

$$p(n) = -\prod_{j=1}^{j_1} d_j(n) = -\prod_{j=1}^{j_1} W_{2^j}x(n) \tag{3}$$

where it is assumed that the lowest scale to include is always 1. The sign of $p(n)$ is inverted compared with a DEGG using the chosen wavelet, hence a minus sign is included to maintain the convention. The de-noising effect of $h(n)$ at each scale in conjunction with the multiscale product means that $p(n)$ is near-zero except at discontinuities across the first j_1 scales of $x(n)$ as depicted in Fig. 7(b), allowing better identification of discontinuties than the DEGG. The function $p(n)$ can be half-wave rectified to contain peaks pertaining only to GCIs, $p^+(n)$, or GOIs, $p^-(n)$, which aids the group delay function in the following step. The value of j_1 is limited by J, but it is often no greater than $j_1 = 5$ as the region of support (RoS) of $h_j^s(n)$ and $g_j^s(n)$ becomes prohibitively large, demanding high processing resources and smoothing adjacent discontinuties. $j_1 = 3$ is deemed a good compromise [36].

B. Group Delay Function

A group delay function (GD) [24] can be used for detection of peaks in linear prediction residuals of speech and can be applied to locate spikes in any signal if their minimum separation, $T_{\min}$, is known. Consider the multiscale product, $p^+(n)$, and an R-sample windowed segment beginning at sample n

$$x_n^+(r) = w(r)p^+(n+r) \text{ for } r = 0, \ldots, R-1. \tag{4}$$

The group delay of $x_n^+(r)$ is given by [38]

$$\tau_n^+(k) = \frac{-\mathrm{d}\arg(X_n^+)}{\mathrm{d}\omega} = \Re\left(\frac{\tilde{X}_n^+(k)}{X_n^+(k)}\right) \tag{5}$$

where $X_n^+(k)$ is the Fourier transform of $x_n^+(r)$ and $\tilde{X}_n^+(k)$ is the Fourier transform of $rx_n^+(r)$ at frequency $\omega = 2k\pi/R$. If $x_n^+(r) = \delta(r - r_0)$, where $\delta(r)$ is a unit impulse function, it follows from (5) that $\tau_n^+(k) \equiv r_0 \forall k$. In the presence of noise, $\tau_n^+(k)$ remains constant but with a degree of additive noise, so an averaging procedure needs to be performed over k; different approaches are reviewed in [24]. The *Energy-Weighted Group Delay* was deemed the most appropriate [20], defined as

$$\gamma^+(n) = \frac{\sum_{k=0}^{R-1} |X_n^+(k)|^2 \tau_n^+(k)}{\sum_{k=0}^{R-1} |X_n^+(k)|^2} - \frac{R-1}{2}. \tag{6}$$

Manipulation yields the simplified expression

$$\gamma^+(n) = \frac{\sum_{r=0}^{R-1} r x_n^{+2}(r)}{\sum_{r=0}^{R-1} x_n^{+2}(r)} - \frac{R-1}{2} \tag{7}$$

which is an efficient time-domain formulation and can be viewed as the "center of energy" of $x_n^+(r)$, bounded in the range $[-(R-1)/2, (R-1)/2]$. The location of the negative-going zero crossings of $\gamma^+(n)$ give an accurate estimation of the location of a spike in a function as depicted in Fig. 7(c). Additionally, if a spike is spread in time then the group delay method will find its center of energy, which is particularly useful in the case of the "redoubled" GCI discussed in [17]. The same analysis is applied to $p^-(n)$ to provide $\gamma^-(n)$, whose negative-going zero crossings are GOI candidates.

C. Candidate Selection

The true GCIs are usually a subset of the negative-going zero crossings of $\gamma^+(n)$, with additional false crossings during unvoiced speech, silence and occasionally between GCIs. Many existing approaches concentrate only on those areas where false candidates are unlikely to occur. The following candidate selection technique aims to remove all false candidates to provide a set of true GCIs throughout an entire segment of speech. Let the number of candidates be M_{cand}^+ occurring at samples $n_m^{\text{cand}+}$, $m = \{0, 1, \ldots, M_{\text{cand}} - 1\}$. Three measurements construct a feature vector, $\mathbf{f}_m^+ = [f_{m,1}^+ \; f_{m,2}^+ \; f_{m,3}^+]^T$, from which is derived a feature matrix, $\mathbf{F}^+ = [\mathbf{f}_0^+ \; \mathbf{f}_1^+ \; \ldots \; \mathbf{f}_{M_{\text{cand}}-1}^+]$. The features are defined as follows.

1) *Consistency of the group delay gradient*. In the case of a Dirac pulse, $\gamma^+(n)$ is a negative unit slope, with a zero crossing at the location of the impulse and width R samples, as shown in Fig. 7(c). A spread pulse or the presence of noise will cause the slope to deviate from the ideal shape, denoted $I(n)$. The RMS error between ideal and measured is calculated as

$$f_{m,1}^+ = \sqrt{\frac{1}{R}\sum_{n=-(R-1)/2}^{(R-1)/2}(\gamma^+(n+n_m^{\text{cand}+}) - I(n))^2}. \quad (8)$$

2) *Peak value of multiscale product's j_1^{th} root inside group delay window*. It is shown in [34] that the j_1^{th} root of $p^+(n)$ helps to give a "zooming in" on the signal, particularly at weak amplitudes (in this case $j_1 = 3$). Experimentation with this algorithm has shown that the group delay function gives best results on $p^+(n)$ but that its j_1^{th} root has better discriminative properties.

$$f_{m,2}^+ = \max\sqrt[j_1]{p^+(n+n_m^{\text{cand}+})}, -\frac{R-1}{2} \le n \le \frac{R-1}{2} \quad (9)$$

3) *Area beneath multiscale product's j_1^{th} root inside group delay window*. In the case of a spread singularity, the area beneath the multiscale product's j_1^{th} root can provide better discrimination of candidates.

$$f_{m,3}^+ = \sum_{n=-(R-1)/2}^{(R-1)/2}\sqrt[j_1]{p^+(n+n_m^{\text{cand}+})}. \quad (10)$$

The distributions of the feature vectors are modeled as two multivariate Gaussians using the EM algorithm [25], initialized with two random data points. Acceptance or rejection is based upon the likelihood of class ω_i, $i = \{1, 2\}$, given feature vector $\mathbf{f}_m^+$

$$\max_i \; P\{\omega_i | \mathbf{f}_m^+\}. \quad (11)$$

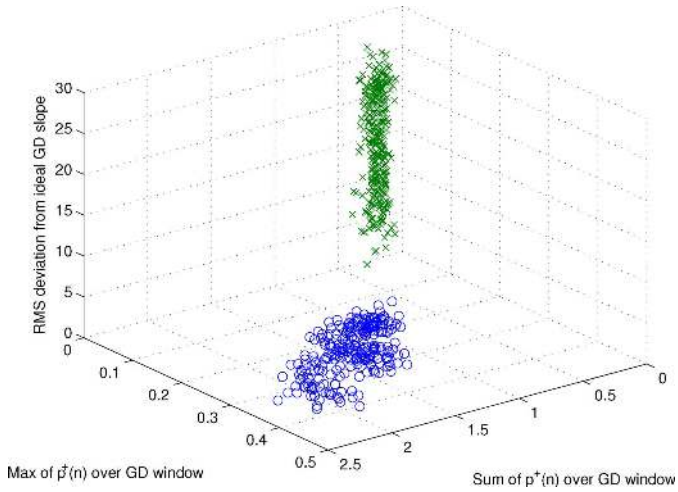

Fig. 8. Typical distribution of GCI feature vectors for a segment of voiced/unvoiced/silent speech. The chosen cluster, whose members are marked "o" is the one whose mean f_3^+ is furthest from the origin. Rejected candidates are marked "×."

Fig. 8 shows a typical distribution of the feature vectors for a segment of mixed voiced/unvoiced/silent speech. It has been found empirically that the cluster whose mean f_3 is furthest from the origin is most likely to contain the chosen candidates, marked "o." Rejected candidates are marked "×." The chosen GCI estimates are defined as n_m^{est}.

GOIs are calculated in the same way but with reversed signs where appropriate.

D. Swallowing

The algorithm proposed thus far performs accurate singularity detection on an input signal without considering any characteristics peculiar to EGG waveforms. It is found that in natural conversional speech, singularities are often caused by swallowing and occasionally by electrical interference in the measurement apparatus and are usually single isolated impulse-like signals. Considering a maximum period $T_{\max}$ all GCIs which are separated from a neighbouring GCI by more than $T_{\max}$ are rejected, else they are kept providing: $(n_m^{\text{est}} - n_{m-1}^{\text{est}}) < T_{\max} f_s > (n_{m+1}^{\text{est}} - n_m^{\text{est}})$.

Experimentation has shown that provided the polarity of the recording is correct, swallowing only causes errors in closure detection so this technique is not applied to opening detection.

E. GOI Postfiltering

GOIs n_m^{est-} are detected from $p^-(n)$ using the same approach as applied to GCI detection (with inverted signs where appropriate). However, the energy imparted by glottal opening is often significantly lower than glottal closure, which results in more erroneous GOI candidates. Assuming that a GOI always accompanies a GCI, postprocessing can be applied to use GCI estimates to improve GOIs accordingly.

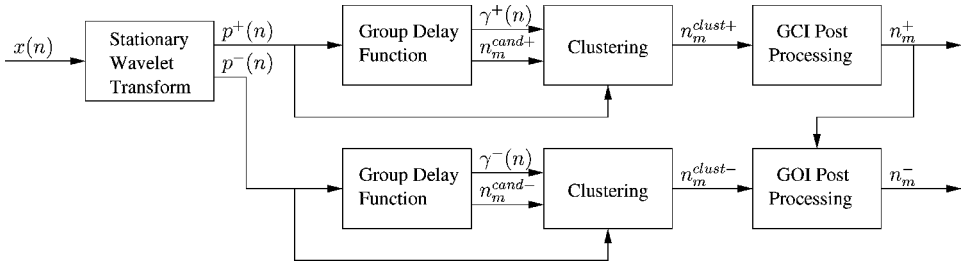

Fig. 9. SIGMA system diagram. The EGG signal $x(n)$ is decomposed into multiple scales from which the half-wave rectified multiscale product $p^+(n)$ is derived. Spike detection is performed on $p^+(n)$ by the negative-going zero crossings of the group delay function $\gamma^+(n)$ at samples $n_m^{\mathrm{cand}+}$. Feature vectors derived from the ideal group delay slope and $p^+(n)$ are clustered by an unsupervised EM algorithm to obtain the GCI estimates n_m^{clust+}. Similarly, GOIs are detected using the negative half-wave of the multiscale product $p^-(n)$. Postprocessing is applied to the GCI estimates to remove isolated clicks from sources other than glottal closure to give n_m^+. GOI postprocessing removes candidates which do not lie within the range of permitted open quotients, using the GCIs as references giving n_m^-.

The main cause of error in GOI post-filtering is small perturbations in $p^-(n)$ immediately preceding a glottal closure which triggers a zero crossing in the group delay function. A region surrounding the closure is therefore isolated, limiting the allowed open quotient, Q_o, to the bounds $Q_o^{\min}$ and $Q_o^{\max}$. The first candidate which lies within these limits is accepted; if no candidate is found, then one is inserted following the current GCI at the previous open quotient.

The SIGMA system diagram is shown in Fig. 9. Symmetry can be seen between closure and opening detection up until the postprocessing stage; prior to this point the algorithm need only know the maximum frequency of the singularities to detect and so is suitable for general singularity detection.

IV. Results and Discussion

The SIGMA algorithm has three parameters and these were set as follows.

- $T_{\min}$: the group delay evaluation window size and therefore the maximum frequency of singularities which can be detected. In the case of voiced speech, the maximum glottal frequency is ~ 400 Hz giving $T_{\min} = 2.5$ ms.
- $T_{\max}$: the maximum glottal period, so that isolated GCI candidates separated from neighboring candidates by more than this value are removed in the GCI postfiltering step. A minimum glottal frequency of 50 Hz leads to a $T_{\max} = 20$ ms.
- $[Q_o^{\min}, Q_o^{\max}]$: the minimum and maximum open quotients for GOI postfiltering. Their purpose is to isolate a region around a GCI inside which a GOI cannot be detected. They are set at 10% and 90%, respectively.

The MATLAB implementation of the chosen biorthogonal spline decomposition filters is called *bior1.5*.

A. Experiment 1: Evaluation With APLAWD and SAM

The APLAWD database [39] contains speech and contemporaneous EGG recordings of five short sentences, repeated ten times by five male and five female talkers. GCIs and GOIs were hand-labeled on the first repetition of every sentence independently of the algorithms under test, denoted $n_m^{\mathrm{true}+}$, $m = \{0, 1, \ldots, M_{\mathrm{true}+} - 1\}$, and $n_m^{\mathrm{true}-}$, $m = \{0, 1, \ldots, M_{\mathrm{true}-} - 1\}$, respectively. A subset of the SAM database [40] contains readings of duration approximately 150 seconds by two male and two female speakers and these were labeled in the same manner. SAM recordings are considered to contain more natural speech with a greater number of swallows and present a more challenging task for a glottal activity detector. The EGG recordings were run through the HQTx (GCI only), TXGEN and SIGMA algorithms and were evaluated by finding the number of estimates per reference cycle then classified as follows, depicted in Fig. 10.

1) Hit. One estimate per true glottal cycle.
2) Miss. No estimates per true glottal cycle.
3) False Alarm (FA). More than one estimate per glottal cycle.
4) False Alarm Total (FAT) Total number of false alarms (the number of estimates which are not hits).

The measures are defined as follows.

1) $Hit\% = n.\ hits/(M_{\mathrm{true}} - 1) \times 100.$
2) $Miss\% = n.\ miss/(M_{\mathrm{true}} - 1) \times 100.$
3) $FA\% = n.\ FA/(M_{\mathrm{true}} - 1) \times 100.$
4) $FAT\% = n.\ FAT/M_{est} \times 100.$
5) $Overall\% = n.\ hits/(M_{\mathrm{true}} - 1 + n.\ FAT) \times 100.$

A glottal cycle is defined as $(n_m^{\mathrm{true}+} - n_{m-1}^{\mathrm{true}+})$ for GCIs and $(n_m^{\mathrm{true}-} - n_{m-1}^{\mathrm{true}-})$ for GOIs. Hit accuracy δ and hit bias ζ are the RMS and mean errors between all hits and the corresponding ground-truth estimates, respectively. The testing strategy is identical to that employed in [19] with the addition of the FAT measure, which counts the *total* number of false alarms as a proportion of total estimates and not the number of reference cycles containing more than one estimate as a proportion of true glottal cycles. The overall figure of merit provides a single-valued measure of performance by expressing the hit rate as a proportion of all reference cycles summed with the number of non-hit estimates (the FAT).

The GCI results in Tables I and III show that SIGMA performs significantly better than HQTx and TXGEN when applied to either database. Notably HQTx is prone to false alarm errors whereas TXGEN is prone to miss errors; this agrees with the qualitative analysis of HQTx's performance in Section II which showed that it is prone to false alarms at the end of segments of voiced speech. HQTx and TXGEN exhibit much greater FAT than FA which suggests that each false alarm is usually followed by successive false alarms within a single reference cycle. SIGMA's miss, FAT and FA measures are broadly similar which tells us that successive false alarms do not usually occur within a

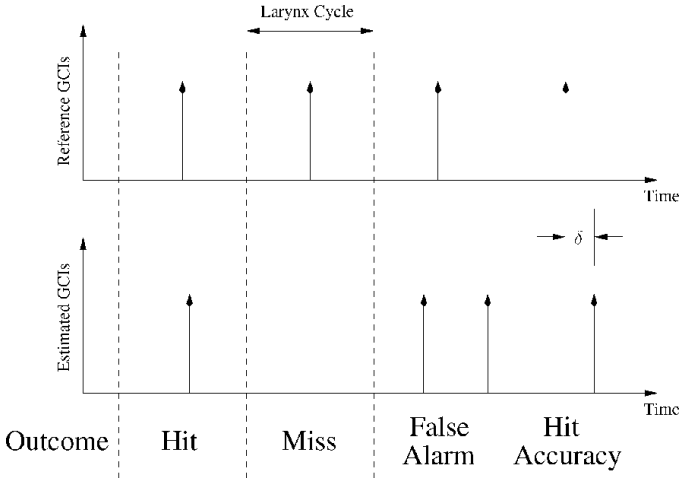

Fig. 10. Testing strategy. A *hit* is one estimate occuring during a reference cycle. A *miss* is the absence of an estimate per reference cycle. If more than one estimate occurs per reference cycle, one *false alarm (FA)* is counted and the total number false alarms in the cycle are added to *false alarm total (FAT)*. Accuracy and bias are the RMS and mean errors between hits and the corresponding reference, respectively.

TABLE I
CLOSURE PERFORMANCE ON THE APLAWD DATABASE BY HQTX, TXGEN, SIGMA (dG AND bior1.5 WAVELET) ALGORITHMS

	Hit (%)	Miss (%)	FA (%)	FAT (%)	Hit Acc., δ (ms)	Hit Bias, ζ (ms)	Overall (%)
HQTx	96.47	1.09	2.44	4.73	0.75	-0.01	91.95
TXGEN	94.78	3.47	1.76	2.63	0.45	-0.13	93.27
SIGMA-dG	99.49	0.25	0.26	0.38	0.04	0.01	99.12
SIGMA	99.59	0.26	0.14	0.18	0.04	0.02	99.41

TABLE II
OPENING PERFORMANCE ON THE APLAWD DATABASE BY TXGEN, SIGMA (dG AND bior1.5 WAVELET) ALGORITHMS

	Hit (%)	Miss (%)	FA (%)	FAT (%)	Hit Acc., δ (ms)	Hit Bias, ζ (ms)	Overall (%)
TXGEN	94.63	3.55	1.82	2.05	0.86	-0.05	93.05
SIGMA-dG	99.38	0.30	0.32	0.43	0.24	0.04	98.96
SIGMA	99.47	0.32	0.21	0.24	0.18	0.04	99.23

TABLE III
CLOSURE PERFORMANCE ON THE SAM DATABASE BY HQTX, TXGEN, SIGMA (dG AND bior1.5 WAVELET) ALGORITHMS

	Hit (%)	Miss (%)	FA (%)	FAT (%)	Hit Acc., δ (ms)	Hit Bias, ζ (ms)	Overall (%)
HQTx	95.68	0.27	4.05	14.56	0.37	-0.01	81.85
TXGEN	90.22	9.75	0.03	0.03	1.08	-0.21	90.19
SIGMA-dG	99.27	0.34	0.39	0.44	0.15	-0.05	98.83
SIGMA	99.35	0.50	0.14	0.17	0.16	-0.04	99.18

given reference cycle and that misses and false alarms have similar likelihood. SIGMA's overall figures of merit are more than an order of magnitude greater than the other algorithms under test.

SIGMA's GCI hit accuracy is in the order of a few samples which agrees with the statement in Section III-C that the true GCIs are usually a subset of the SIGMA candidate GCIs before clustering. SIGMA and HQTx hit bias are universally low but TXGEN's estimates tend to occur slightly early.

TABLE IV
OPENING PERFORMANCE ON THE SAM DATABASE BY TXGEN, SIGMA (dG WAVELET AND bior1.5 WAVELET) ALGORITHMS

	Hit (%)	Miss (%)	FA (%)	FAT (%)	Hit Acc., δ (ms)	Hit Bias, ζ (ms)	Overall (%)
TXGEN	90.44	9.44	0.11	0.14	2.06	-0.11	90.31
SIGMA-dG	98.95	0.34	0.71	0.88	0.29	0.02	98.08
SIGMA	99.23	0.38	0.39	0.50	0.25	0.01	98.74

SIGMA's GOI results in Tables II and IV are also encouraging. The reliance upon the estimated GCIs results in similar hit, miss and false alarm rates, with diminished hit accuracy due to the greater difficulty of precisely locating openings. The gap in the overall figure of merit between SIMGA and TXGEN is again more than an order of magnitude.

B. Experiment 2: Variation in Group Delay Window Size

The group delay evaluation window size was set according to the physical constraints of human speech, whose minimum fundamental period is around 2.5 ms. This experiment assesses the algorithm's sensitivity to variation in the group delay window size on the APLAWD database.

The results presented in Fig. 11 show that 2.5 ms is indeed an optimal choice of window length. The reliance on GCIs to estimate GOIs means that intuitively the overall, hit, miss, and FAT rates should vary in a similar manner which is confirmed by these results.

FAT rates increase with decreasing window sizes due to the fact that more negative zero crossings can occur in the group delay function per unit time. In this case the true candidates remain a subset of all candidates, with a number of additional false ones arising. Providing the clustering algorithm can discriminate against the false candidates, those which are true should always be detected so false alarm rates should therefore increase slowly with decreasing window size.

Miss rates increase with window size as neighboring singularities can occur within a single group delay window and reduce the number of negative zero crossings. It becomes impossible for the GMM to find the correct candidates as they are no longer a subset of the candidate set, hence miss rates climb rapidly with increasing window size.

GCI bias and hit accuracy are relatively immune to variations in window size, suggesting that providing one candidate occurs per true period, is it statistically the correct choice. GOI bias and hit accuracy are more sensitive, showing the most significant increase with reduced window size. Bias increases monotonically with decreasing window length.

This experiment was repeated for male- and female-only speech. The results provide similar curves to the previous experiment that employs both genders, the optimum value being shifted up to approximately 3 ms for male voices and down to approximately 2 ms for female. The experiment with mixed male/female speech shows that variation in group delay size does not have a significant effect upon the results in the range of approximately 1.5 to 3.5 ms, hence performance is weakly dependent on gender.

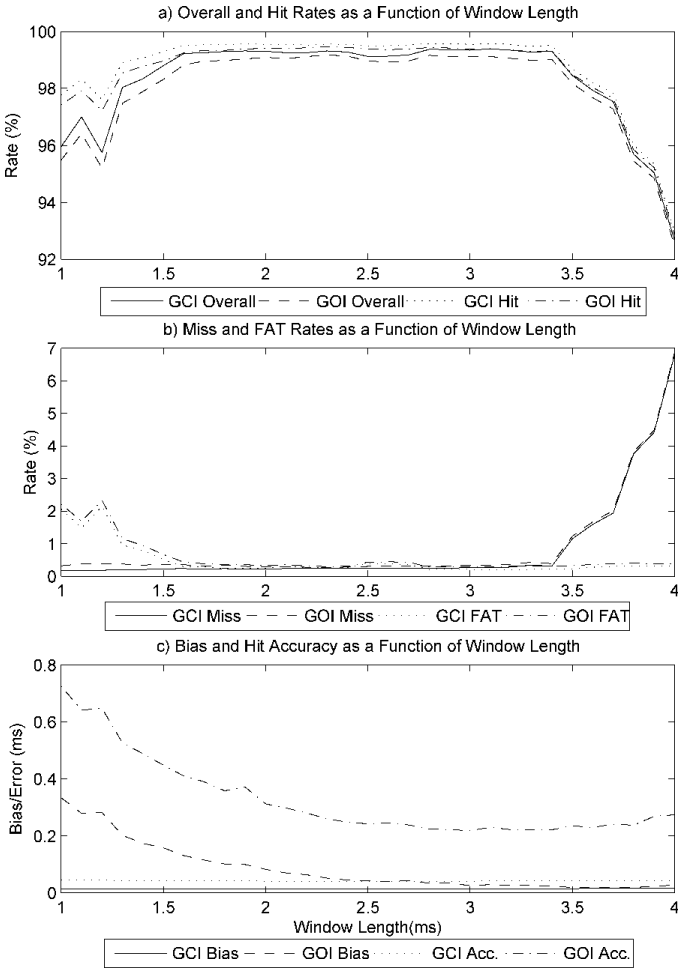

Fig. 11. Effect of varied group delay window length on (a) overall and hit, (b) miss and FAT, and (c) and bias and hit accuracy. The choice of 2.5 ms from physical reasoning is close to the optimal value.

C. Experiment 3: Comparison With Cubic Spline Wavelet

The derivative-of-Gaussian (dG) cubic spline wavelet is the wavelet of choice for multiscale analysis in [17], [18] and [36]. Experiments with other common wavelets have shown that the bior1.5 biorthogonal spline wavelet is more effective for EGG analysis with this algorithm. The results in Tables I–IV show SIGMA using the dG wavelet (labeled SIGMA-dG) as well as the proposed bior1.5 (labeled SIGMA).

The performance of SIGMA is slightly reduced with the dG wavelet, particularly with increased false alarms and increased hit error on the opening tests. Miss rates are slightly reduced but the greater increase in false alarm rate diminishes the overall performance results.

V. Conclusion

We have shown that robust detection of GCIs and GOIs from EGG signals is particularly challenging at the transition regions around the ending of voicing. A new method for glottal activity detection from EGG recordings has been presented which is accurate even in these challenging regions. It first detects singularities in the EGG signal by the multiscale product of three dyadic scales. It then employs a technique based upon the group delay function which detects peaks in the multiscale product. Incorrect estimates are removed by the clustering of three-dimensional feature vectors using the EM algorithm. Postprocessing removes isolated GCIs and uses GCIs to aid GOI detection.

A comparison was made between the proposed approach and two popular existing methods by evaluating their performance against 50 short and four long hand-labeled sentences. An existing testing procedure with some new enhancements was used, showing very accurate GCI and GOI detection with the proposed method, fulfilling our objective of obtaining results that are accurate enough to be used as a reference. Our method enables accurate evaluation of speech-based glottal activity detection algorithms, precise estimation of the closed phase for the estimation of glottal volume flow and could also be applied to the analysis of a number of types of pathological speech. Further, few assumptions are made about the nature of the input signal. This allows the application of the proposed algorithm to singularity detection in almost any signal provided the minimum separation of singularities is known.

References

[1] P. Davies, G. A. Lindsey, H. Fuller, and A. J. Fourcin, "Variation of glottal open and closed phases for speakers of English," *Proc. Inst. Acoust.*, vol. 8, no. 7, pp. 539–546, 1986.

[2] W. Hess and H. Indefrey, "Accurate pitch determination of speech signals by means of a laryngograph," in *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 1984, vol. 9, pp. 73–76.

[3] H. Valbret, E. Moulines, and J. P. Tubach, "Voice transformation using PSOLA technique," *Speech Commun.*, vol. 11, no. 2, pp. 175–187, Jun. 1992.

[4] N. D. Gaubitch, P. A. Naylor, and D. B. Ward, "Multi-microphone speech dereverberation using spatio-temporal averaging," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Vienna, Austria, Sep. 2004, pp. 809–812.

[5] E. Moulines and F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech Commun.*, vol. 9, no. 5–6, pp. 453–467, Dec. 1990.

[6] M. R. P. Thomas, J. Gudnason, and P. A. Naylor, "Data-driven voice source waveform modelling," in *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 3965–3968.

[7] J. Deller, "Some notes on closed phase glottal inverse filtering," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP–29, no. 4, pp. 917–919, Aug. 1981.

[8] J. Gudnason and M. Brookes, "Voice Source cepstrum coefficients for speaker identification," in *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2008, pp. 4821–4824.

[9] R. Colton and J. Casper, *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment*. New York: Williams & Wilkins, 1996.

[10] K. Verdolini, R. Chan, I. R. Titze, M. Hess, and W. Bierhals, "Correspondence of electroglottographic closed quotient to vocal fold impact stress in excised canine larynges," *J. Voice*, vol. 12, no. 4, pp. 415–423, Feb. 1998.

[11] J. Gamboa, F. J. Jiménez-Jiménez, A. Nieto, I. Cobeta, A. Vegas, M. Orti-Pareka, T. Gasalla, J. A. Molina, and E. Garcia-Albea, "Acoustic voice analysis in patients with essential tremor," *J. Voice*, vol. 12, no. 4, pp. 444–452, Feb. 1998.

[12] E. R. M. Abberton, D. M. Howard, and A. J. Fourcin, "Laryngographic assessment of normal voice: A tutorial," *Clinical Linguist. Phon.*, vol. 3, pp. 281–296, 1989.

[13] D. G. Childers, D. M. Hooks, G. P. Moore, L. Eskenazi, and A. L. Lalwani, "Electroglottography and vocal fold physiology," *J. Speech. Hear. Res.*, vol. 33, no. 2, pp. 245–254, Jun. 1990.

[14] D. M. Howard, "Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers," *J. Voice*, vol. 9, no. 2, pp. 121–1223, Jun. 1995.

[15] N. Henrich, C. d'Alessandro, M. Castellengo, and B. Doval, "On the use of the derivative of electroglottographic signals for characterization of nonpathological voice phonation," *J. Acoust. Soc. Amer.*, vol. 115, no. 3, pp. 1321–1332, Mar. 2004.

[16] M. A. Huckvale, "Speech Filing System: Tools for Speech," Tech. Rep. Univ. College London, London, U.K., 2004 [Online]. Available: http://www.phon.ucl.ac.uk/resource/sfs
[17] A. Bouzid and N. Ellouze, "Multiscale product of electroglottogram signal for glottal closure and opening instant detection," in *Proc. IMACS MultiConf. Comput. Eng. Syst. Applicat.*, 2006, vol. 1, pp. 106–109.
[18] A. Bouzid and N. Ellouze, "Glottal opening instant detection from speech signal," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Vienna, Austria, Sep. 2004, pp. 729–732.
[19] P. A. Naylor, A. Kounoudes, J. Gudnason, and M. Brookes, "Estimation of glottal closure instants in voiced speech using the DYPSA algorithm," *IEEE Trans. Speech Audio Process.*, vol. 15, no. 1, pp. 34–43, Jan. 2007.
[20] M. R. P. Thomas, N. D. Gaubitch, and P. A. Naylor, "Multichannel DYPSA for estimation of glottal closure instants in reverberant speech," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Poznan, Poland, Sep. 2007.
[21] K. S. Rao, S. R. M. Prasanna, and B. Yegnanarayana, "Determination of instants of significant excitation in speech using Hilbert envelope and group delay function," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 762–765, Oct. 2007.
[22] K. S. R. Murty and B. Yegnanarayana, "Epoch extraction from speech signals," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 8, pp. 1602–1613, Nov. 2008.
[23] W. Saidi, A. Bouzid, and N. Ellouze, "Evaluation of multi-scale product method and DYPSA algorithm for glottal closure instant detection," in *Proc. 3rd Int. Conf. Inf. Commun. Technol.: From Theory to Applicat. (ICTTA)*, Apr. 2008, pp. 1–5.
[24] M. Brookes, P. A. Naylor, and J. Gudnason, "A quantitative assessment of group delay methods for identifying glottal closures in voiced speech," *IEEE Trans. Speech Audio Process.*, vol. 14, no. 2, pp. 456–466, Mar. 2006.
[25] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc., Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.
[26] M. B. Higgins and J. H. Saxman, "A comparison of selected phonatory behaviours of healthy aged and young adults," *J. Speech Hear. Res.*, vol. 34, pp. 1000–1010, Oct. 1991.
[27] A. K. Krishnamurthy and D. G. Childers, "Two-channel speech analysis," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 4, pp. 730–743, Aug. 1986.
[28] D. M. Howard and G. Lindsey, "Conditioned variability in voicing offsets," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 3, pp. 406–407, Mar. 1988.
[29] J. C. Catford, *Fundamental Problems in Phonetics*. Bloomington, IN: Indiana Univ. Press, 1977.
[30] Y. Lebrun and J. Hasquin-Deleval, "On the so-called 'dissociations' between electroglottogram and phonogram," *Folia Phoniatrica*, vol. 23, pp. 225–227, 1971.
[31] M. R. P. Thomas, J. Gudnason, and P. A. Naylor, "Application of the DYPSA algorithm to segmented time-scale modification of speech," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Lausanne, Switzerland, Aug. 2008.
[32] P. Jax and P. Vary, "On artificial bandwidth extension of telephone speech," *Signal Process.*, vol. 83, pp. 1707–1719, 2003.
[33] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 7, pp. 710–732, Jul. 1992.
[34] A. Bouzid and N. Ellouze, "Local regularity analysis at glottal opening and closure instants in electroglottogram signal using wavelet transform modulus maxima," in *Proc. Eurospeech*, 2003, pp. 2837–2840.
[35] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 617–643, Mar. 1992.
[36] B. M. Sadler and A. Swami, "Analysis of multiscale products for step detection and estimation," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 1043–1051, Apr. 1999.
[37] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
[38] R. Smits and B. Yegnanarayana, "Determination of instants of significant excitation in speech using group delay function," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 3, pp. 325–333, Sep. 1995.
[39] G. Lindsey, A. Breen, and S. Nevard, "SPAR's Archivable Actual-Word Databases," Tech. Rep. Univ. College London, London, U.K., 1987.
[40] D. Chan, A. Fourcin, D. Gibbon, B. Granstrom, M. Huckvale, G. Kokkinakis, K. Kvale, L. Lamel, B. Lindberg, A. Moreno, J. Mouropoulos, F. Senia, I. Trancoso, C. Veld, and J. Zeiliger, "EUROM – A spoken language resource for the EU," in *Proc. Eur. Conf. Speech Commun. Technol.*, Sep. 1995, pp. 867–870.

Mark Thomas (S'06) received the M.Eng. degree in electrical and electronic engineering from Imperial College, London, U.K., in 2006 where he is currently pursuing the Ph.D. degree.

His research interests include glottal-synchronous and multichannel speech processing, involving methods for analysis, prosodic manipulation and reverberation/noise reduction. His previous experience in industry was with the BBC R&D Department, where he worked on audio, video, and RF engineering.

Patrick Naylor (M'89–SM'07) received the B.Eng. degree in electronics and electrical engineering from the University of Sheffield, Sheffield, U.K., in 1986 and the Ph.D. degree from Imperial College, London, U.K., in 1990.

Since 1989, he has been a Member of Academic Staff in the Communications and Signal Processing Research Group, Imperial College London, where he is also Director of Postgraduate Studies. His research interests are in the areas of speech and audio signal processing and he has worked in particular on adaptive signal processing for acoustic echo control, speaker identification, multichannel speech enhancement, and speech production modeling. In addition to his academic research, he enjoys several fruitful links with industry in the U.K., U.S., and in mainland Europe.

Dr. Naylor is an Associate Editor of IEEE SIGNAL PROCESSING LETTERS and a member of the IEEE Signal Processing Society Technical Committee on Audio and Electroacoustics.