16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

THE SIGMA ALGORITHM FOR ESTIMATION OF REFERENCE-QUALITY GLOTTAL CLOSURE INSTANTS FROM ELECTROGLOTTOGRAPH SIGNALS

Mark R. P. Thomas and Patrick A. Naylor

Department of Electrical and Electronic Engineering
Imperial College London, SW7 2AZ, UK
email: {mrt102, p.naylor}@imperial.ac.uk
web: www.commsp.ee.ic.ac.uk/∼sap

ABSTRACT

Accurate estimation of glottal closure instants (GCIs) in voiced speech is important for speech analysis applications which benefit from glottal-synchronous processing. Electroglottograph (EGG) recordings give a measure of the electrical conductance of the glottis, providing a signal which is proportional to its contact area. EGG signals contain little noise or distortion, providing a good reference from which GCIs can be extracted to evaluate GCI estimation from speech recordings. Many approaches impose a threshold on the differentiated EGG signal which provide accurate results during voiced speech but are prone to errors at the onset and end of voicing; modern algorithms use a similar approach across multiple dyadic scales using the stationary wavelet transform. This paper describes a new method for EGG-based GCI estimation named SIGMA, which is based upon the stationary wavelet transform, peak detection with a group delay function and Gaussian Mixture Modelling for discrimination between true and false GCI candidates.

In most real-world environments, it is necessary to estimate GCIs from a speech signal recorded with a microphone placed at some distance from the talker. The presence of reverberation, noise and filtering by the vocal tract render GCI detection from real speech signals relatively difficult to achieve compared with the EGG, so EGG-based references have often been used to evaluate GCI detection from speech signals. Evaluation against 500 hand-labelled sentences has shown an accuracy of 99.35%, a 4.7% improvement over a popular existing method.

1. INTRODUCTION

Identification of glottal closure instants (GCIs) in voiced speech is important for speech processing algorithms such as prosodic speech modification [1], speech dereverberation [2], glottal-synchronous processing in speech synthesis [3] and some fields of speech therapy [4]. A method for detecting GCIs is through the examination of the Electroglottograph (EGG) (or Laryngograph) signal [4], which is a measurement of the electrical conductance of the glottis. It passes a low-voltage, high-frequency signal through a pair of electrodes on the subject's neck in line with the glottis and measures the conductance. The signal is proportional to the glottal contact area, whose derivative (DEGG) during voiced speech is an impulse train-like signal. An example of a voiced speech segment, the corresponding EGG recording and its derivative is shown in Fig. 1. Many approaches exploit this property [5, 6, 7, 8] in conjunction with dynamic thresholds to obtain an accurate estimate of GCIs during voiced speech. However, they are often prone to errors at the onset and end of voicing as shown in Section 2 using the High-Quality Tx (HQTx) algorithm [8] as an example.

Recent approaches have applied multiscale analysis to detect GCIs as singularities in the EGG signal [9] and from the speech signal [10]. This paper describes the Singularity detection In EGG with Multiscale Analysis (SIGMA) algorithm, which uses the same back-end processing. Thereafter, peak detection is performed on the multiscale product using a group delay approach [11], where the negative-going zero crossings of the average slope of the negative unwrapped phase of the Fourier transform of the EGG derivative are identified, calculated over a sliding window. This method has been applied previously to GCI estimation from an LP residual [12] and it circumvents the need for dynamic thresholds as the method is based entirely upon phase information. A number of false candidates may arise and these are removed by modelling three-dimensional feature vectors as Gaussian distributions and clustering with an unsupervised EM algorithm. The result is a parameterless algorithm which makes no assumptions about the nature of the EGG signal other than the maximum glottal frequency and that a GCI is characterized by a singularity. It may therefore have many further uses, making it suitable for singularity detection in almost any signal.

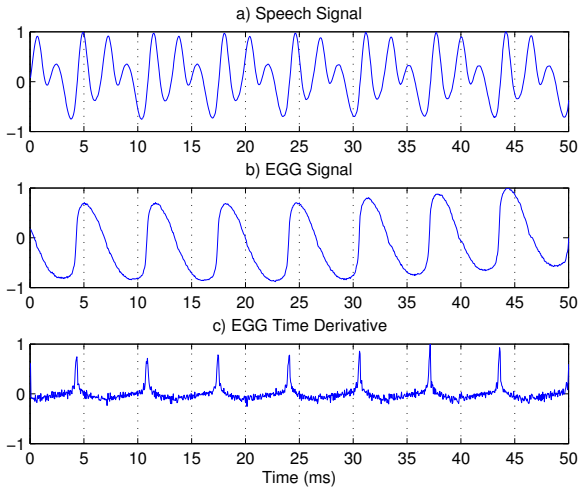

Figure 1: Speech signal, the corresponding EGG signal and the EGG time derivative. In this instance, the negative peaks due to glottal opening are very weak.

In practical applications, GCIs are usually derived from an estimation of the excitation source from real speech recordings which can be distorted in combination of four distinct ways:

(i) Filtering by the vocal tract. This may be modelled as an all-pole filter by LP analysis [13] and then inverse-filtered to give an LP residual. The residual contains strong spikes at times of glottal closure which are a good indication of their true location.

(ii) Filtering by the nose and vocal articulators. Nasal phonemes (such as /m/) and the use of the tongue, teeth and lips can introduce zeros into the vocal transfer function. Methods for blindly determining poles in any transfer function [13] are generally more accurate than those for determining zeros [14]. In addition, full knowledge of the locations of zeros may not always be helpful as they are not necessarily minimum phase, leading to unstable inverse filters and the need for specialized inversion algorithms [15].

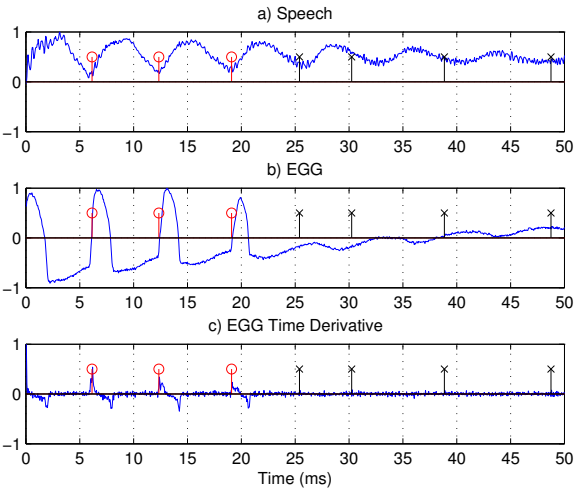

Figure 2: A speech signal (a), EGG signal (b), its time derivative (c) and HQTx GCI estimation markers at the end of a voiced speech segment. The first three GCIs are identified correctly (marked ○) but the last four (marked ×) are erroneous. In this instance, the negative peaks due to glottal opening are significant.

(iii) Background noise. This can be periodic (computer fans), impulsive (doors closing) or quasi-periodic (other talkers). Any type of noise can cause spurious peaks in the LP residual and may fool many algorithms into recognising them as GCIs.

(iv) Reverberation. Sound reflected from walls, desks and other hard objects cause further unwanted peaks in the LP residual, which may be indistinguishable from the true GCIs.

Distortion by (iii) and (iv) may be avoided with recordings in anechoic environments, but this is unrealistic for real-world applications. Estimation and inversion of the all-pole vocal tract filter is a powerful technique but the effect of zeros in the transfer function is usually ignored. Recent developments in speech-based GCI algorithms such as Multichannel DYPSA [16] give good GCI estimation from noisy, reverberant speech recordings and have been shown to work well with glottal-synchronous algorithms [17]. GCI estimation from distorted speech recordings is therefore viable, but with the increased interest in glottal-synchronous processing there has been a corresponding demand for more accurate GCI detection algorithms. The proposed algorithm uses EGG signals to provide 'ground-truth' GCIs, against which speech-based GCI detection algorithms may be evaluated.

This paper is organized as follows: Section 2 reviews the interpretation of an EGG signal. Section 3 describes multiscale analysis and the use of the group delay function for peak detection in the multiscale product. The proposed SIGMA algorithm and the HQTx algorithm are evaluated against 500 hand-labelled sentences in Section 4 and conclusions are drawn in Section 5.

2. INTERPRETING EGG RECORDINGS

A voiced speech signal, the corresponding time-aligned EGG signal and the EGG derivative are shown in Fig. 1. Time alignment is achieved by measuring the source-microphone propagation distance and calculating the delay. The convention used in this paper is for a positive EGG signal to correspond to high contact area of the glottis. During glottal closure there is a large positive transient in the EGG waveform and a corresponding spike in the derivative at each GCI.

The detection of GCIs in the middle of a segment of voiced speech is a relatively straightforward task as the positive peaks in the derivative are distinct. Difficulty arises at end of voiced speech when the air velocity can drop to a point where the glottis no longer snaps shut but is "flapping in the breeze" [18], as seen in Fig. 2. In this region, the EGG signal is a damped sinusoid of decreasing frequency with low corresponding speech energy. In the example of Fig. 2, the hand-labeller would not mark any GCIs from 20 ms onwards as there is no visible instant which defines the periodicity. However, the HQTx algorithm flags a number of erroneous GCIs until the amplitude of the DEGG signal drops below a threshold level. The proposed algorithm is robust to these errors.

Large changes in amplitude of EGG can also cause errors in dynamic threshold-based algorithms, sometimes causing missed peaks if the threshold gain is set too high, or run the risk of flagging erroneous GCIs from noise if it is set too low. GCIs can sometimes be unclear when they are spread out in time [9] which is also often difficult to detect with dynamic thresholds but this is also addressed by the proposed approach.

A glottal closure instant must always be followed by a glottal opening instant (GOI), which manifests itself as a weaker peak of opposite sign in the EGG derivative [5]. GOIs can be useful for closed-phase analysis of speech signals [19] and determining open quotients (OQ) [7]. Some of the aforementioned methods attempt to detect GOIs with the same approach as GCI detection, but it is sometimes impossible as the GOI can be buried in noise and rendered undetectable. Comparing EGG time derivatives in Figs. 1 and 2, the negative peaks caused by GOIs are almost non-existent in the former case and very clear in the latter. Although GOI detection is outside the scope of this paper, our robust GCI detector can provide much useful information for this task.

3. SINGULARITY DETECTION WITH SIGMA

Detection of glottal activity from an EGG signal involves isolating regions of discontinuty (or singularities). Finding the derivative of the EGG signal is a useful approach, where strong peaks and weaker peaks of opposite sign correspond to glottal closure and opening respectively. True and false peaks are often discriminated by assessing the peak amplitude of the EGG derivative (DEGG) and a longer-term measure of the change in EGG amplitude based upon some predetermined window.

3.1 Multiscale Analysis

Let us consider a generalisation of this approach. The dyadic wavelet transform [20] involves iteratively decomposing a signal into decimated subbands; a three-level decomposition is shown in the upper plot of Fig. 3, 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. Singularities can be detected by finding the regions in which the maxima of the multiscale decompositions, $d_j(n)$, converge. A wavelet with n vanishing moments is a multiscale differential operator of order n with some degree of smoothing, so a wavelet with one vanishing moment detects discontinuities in a signal's smoothed derivative, displaying maxima at the discontinuity across multiple scales [21]. It is shown in [22] that one vanishing moment is suitable for singularity detection in EGG signals; as the signal traverses deeper into the tree of filter banks, the derivative is estimated at increasing levels of smoothing. Biorthogonal spline wavelets with one vanishing moment are often chosen for singularity detection as they approximate the first derivative of a Gaussian function [23], giving the smoothing and differention we require.

The dyadic wavelet transform is dyadic in both scale and time; however, in this case we only wish to determine the projection of $x(n)$ on different subspaces, so we do not decimate as shown in the lower plot of Fig. 3. Instead, the filters $g(n)$ and $h(n)$ must be upsampled by 2 at each iteration to implement the 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 [21] 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 result is a signal whose length is unchanged throughout the filterbank tree, allowing sample-by-sample multiplication of the signal at different scales to find converging maxima.

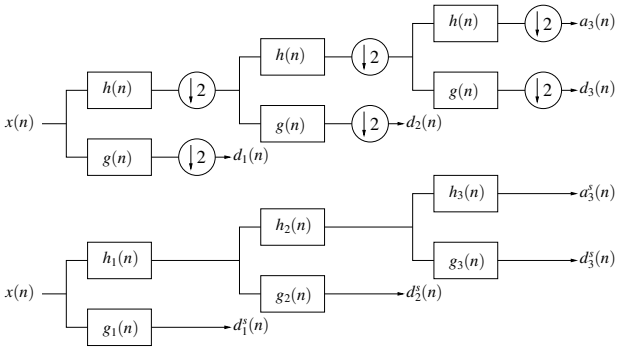

Figure 3: Three-level dyadic signal decomposition on a signal $x(n)$ into detail $d_j(n)$ and approximation $a_j(n)$ parts. The upper figure is the dyadic wavelet transform, where each iteration involves a downsampling by a factor of two. The lower figure is the stationary wavelet transform, where no downsampling is performed on the signal but the filters $g_j^s(n)$ and $h_j^s(n)$ (s indicating stationary) are instead upsampled by 2 upon each iteration.

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, $x(n)$, $1 \le n \le N$ at scale j is

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

where $J = log_2 N$, plus the remaining coarse scale information denoted $S_J(n)$. 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 de-noising effect of the $h(n)$ 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 the centre plot of Fig. 4. 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 $h_i(n)$ and $g_i(n)$ becomes prohibitively large, demanding high processing resources and smoothing adjacent discontinuities. $j_1 = 3$ is a good compromise [24].

3.2 Group Delay Function

A group delay function (GD) was used in [11] for detection of peaks in linear prediction residuals of speech and can be applied to locate spikes in any signal if their minimum separation is known. In the case of GCIs, the maximum frequency of the singularities due to GCIs is in the order of 400 Hz, leading to a window size of 2.5 ms.

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 Fourier transform of $x_n(r)$ at a frequency $\omega = 2k\pi/R$ is

$$X_n(k) = \sum_{r=0}^{R-1} x_n(r)e^{-j\frac{2\pi}{R}rk} \tag{5}$$

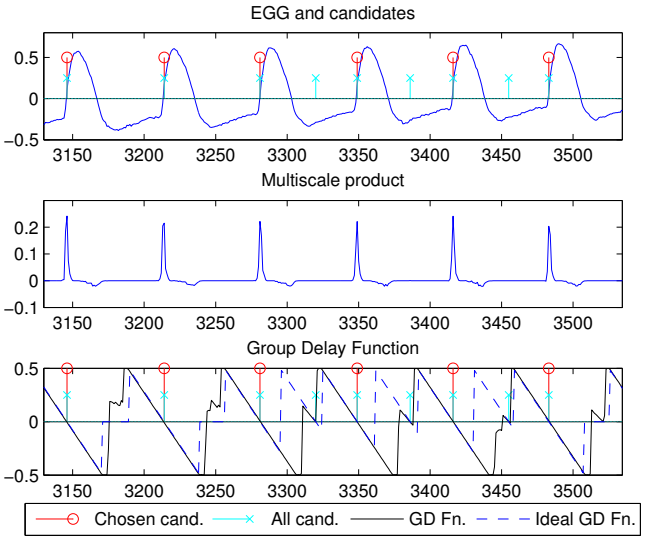

Figure 4: EGG waveform, multiscale product and Group Delay Function. Candidates are marked '×' and chosen candidates are marked 'o'. The ideal slope, marked dashed on the lowest plot, is the simulated slope from perfect impulses at the GCIs.

where k can vary continuously. The group delay of $x_n(r)$ is given by [12]

$$\tau_n(k) = \frac{-\mathrm{d}\arg(X_n)}{\mathrm{d}\omega} = \Re\left(\frac{\tilde{X}_n(k)}{X_n(k)}\right) \tag{6}$$

where $\tilde{X}_n(k)$ is the Fourier transform of $rx_n(r)$. If $x_n(r) = \delta(r-r_0)$, where $\delta(r)$ is a unit impulse function, it follows from (6) that $\tau_n(k) \equiv r_0 \forall k$. In the presence of noise, $\tau_n(k)$ becomes noisy, so an averaging procedure needs to be performed over k; different approaches are reviewed in [11]. The *Energy-Weighted Group Delay* was deemed the most appropriate [16], defined as

$$\gamma(n) = \frac{\sum_{k=0}^{R-1}|X_n(k)|^2\tau_n(k)}{\sum_{k=0}^{R-1}|X_n(k)|^2} - \frac{R-1}{2}. \tag{7}$$

Manipulation yields the simplified expression

$$\gamma(n) = \frac{\sum_{r=0}^{R-1} rx_n^2(r)}{\sum_{r=0}^{R-1} x_n^2(r)} - \frac{R-1}{2} \tag{8}$$

which is an efficient time-domain formulation and can be viewed as the 'centre 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 peak in a function as depicted in the lower plot of Fig. 4. Additionally, if a peak is spread in time then the group delay will tend to find its centre, which is particularly useful in the case of the 'redoubled' GCI discussed in [9].

3.3 Candidate Selection

The negative-going zero crossings of the $\gamma(n)$ will usually occur at the location of the true GCIs, with additional false crossings during unvoiced speech, silence and occasionally between GCIs. Let the number of candidates be M_{cand} occurring at samples n_m^{cand}, $m = \{0,1,\ldots,M_{cand}-1\}$. 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, $\mathbf{F} = [\mathbf{f}_0\ \mathbf{f}_1\ \ldots\ \mathbf{f}_{M_{cand}-1}]$. The measures are:

(i) *Consistency of the group delay gradient.* In the case of an impulse, $\gamma(n)$ is a negative unit slope, with a zero crossing at the location of the impulse and width R samples, as shown in

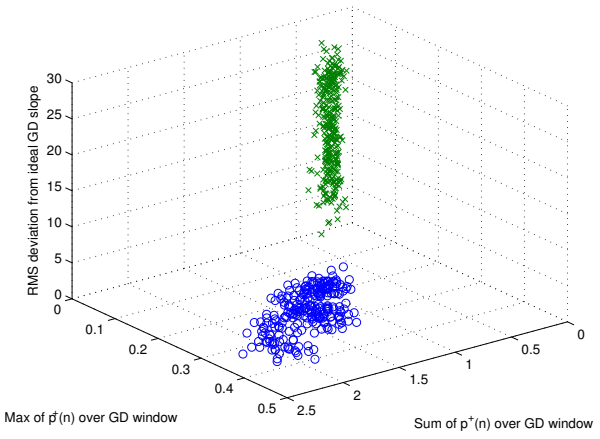

Figure 5: A typical distribution of feature vectors for a segment of voiced/unvoiced/silent speech. The chosen cluster, whose members are marked '○' is the one whose mean f_3 is furthest from the origin. Rejected candidates are marked '×'.

the lower plot of Fig. 4. Spread pulses or noise will cause the slope to deviate from the ideal shape, denoted $I(n)$. The RMS error between ideal and measured is calculated:

$$f_{m,1} = \sqrt{\frac{1}{R}\sum_{n=-(R-1)/2}^{(R-1)/2}(\gamma(n+n_m^{cand}) - I(n+n_m^{cand}))^2}. \quad (9)$$

(ii) *Peak value of multiscale product's j_1^{th} root inside group delay window*. It is shown in [22] that the j_1^{th} root of $p^+(n)$ gives 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.

$$f_{m,2} = \max \sqrt[j_1]{p^+(n)},\ n_m^{cand} - \frac{R-1}{2} \le n \le n_m^{cand} + \frac{R-1}{2} \quad (10)$$

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

$$f_{m,3} = \sum_{n=-(R-1)/2}^{(R-1)/2} \sqrt[j_1]{p^+(n+n_m^{cand})} \quad (11)$$

The feature vectors are modelled as a Gaussian distribution and are divided into two clusters using an unsupervised EM algorithm [25] (k-means), initialized with two random data points. Fig. 5 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 '○'. Rejected candidates are marked '×'.

A system diagram for SIGMA is shown in Fig. 6.

4. RESULTS AND DISCUSSION

The APLAWD database [26] contains speech and contemporaneous EGG recordings of five short sentences, repeated ten times by five male and five female talkers. GCIs were hand-labelled on every sentence independently of the algorithms under test. The same EGG recordings were run through the HQTx and SIGMA algorithms and evaluated by finding the number of estimated GCIs per true period then classified as follows as shown in Fig. 7:

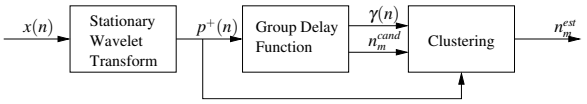

Figure 6: SIGMA system diagram. The EGG signal, $x(n)$, is decomposed into multiple scales from which the half-wave rectified multiscale product, $p^+(n)$, is derived. Peak detection is performed on $p^+(n)$ by the negative-going zero crossings of the group delay function, $\gamma(n)$, at samples n_m^{cand}. 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^{est}.

1. Hit. One GCI per true larynx cycle.
2. Miss. No GCIs per true larynx cycle.
3. False Alarm. More than one GCI per larynx cycle. In this case the closest estimate is a hit and the remaining are false alarms.

Denote n_m^{est}, $m = \{0,1,\ldots,M_{est}-1\}$, the sample locations of the estimated GCIs and n_m^{true}, $m = \{0,1,\ldots,M_{true}-1\}$, the ground-truth. The measures are defined as:

1. *Hit%=n. hits/*$M_{true}\times 100$
2. *Miss%=n. miss/*$M_{true}\times 100$
3. *False Alarm%=n. false alarms/*$M_{est}\times 100$
4. *Overall%=n. hits/(*M_{true} *+ n. false alarms)*$\times 100$

The overall figure of merit combines the measured values. Hit accuracy, δ, and hit bias, ζ, are the the RMS and mean errors between all hits and the corresponding ground-truth GCIs respectively.

The results in Table 1 show that SIGMA performs significantly better than HQTx and this is reflected in the derived measure, *Overall*, where HQTx achieves 94.88% and SIGMA 99.35%. Hit and miss rates are similar between the two algorithms, though it is not necessarily indicative of good performance as indicated by the large number of HQTx's false alarms. This agrees with the qualitative analysis of HQTx's performance in Section 2 which showed that it is prone to false alarms at the onset and end of voiced speech. Hit accuracy is very good for both algorithms, showing negligible bias of under one sample. This agrees with the statement in Section 3.3 that the true GCIs are almost always a subset of the SIGMA candidate GCIs before clustering.

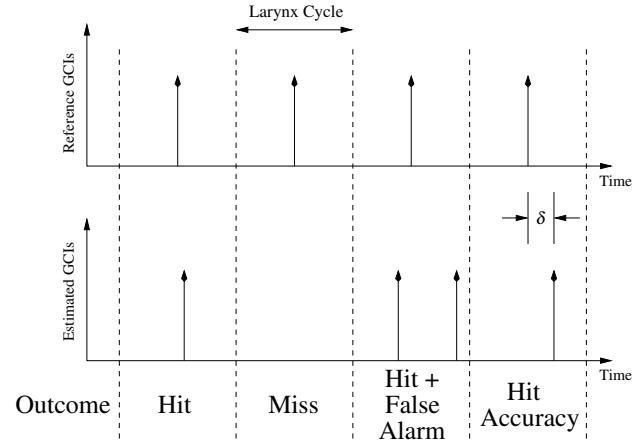

Figure 7: Testing strategy. A hit is when one or more estimated GCIs occur during a reference cycle. A miss is the absence of an estimated GCI per reference cycle. A false alarm is every estimated GCI when more than one estimated GCI occurs per reference cycle. Hit accuracy is the RMS error between a hit and the corresponding reference GCI. Accuracy and bias are the RMS and mean errors between hits and false alarms and the corresponding reference GCIs.

Table 1: Performance comparison between HQTx and SIGMA on the APLAWD database.

	Hit (%)	Miss (%)	FA (%)	Hit Acc., δ (ms)	Hit Bias, ζ (ms)
HQTx	99.68	0.32	4.82	0.1427	0.0231
SIGMA	99.73	0.27	0.37	0.0673	0.0259

5. CONCLUSIONS

It has been seen that robust detection of GCIs from EGG signals is very challenging at the onset and ending of voiced regions of speech. A new method for GCI detection from EGG recordings has been presented which is accurate even in these challenging areas. It first reinforces singularities in the EGG signal by the multiscale product of three dyadic scales, followed by a group delay function which detects peaks in the multiscale product. False candidates are removed by clustering of three-dimensional feature vectors using an unsupervised EM algorithm. A comparison was made between the proposed approach and a popular existing method, HQTx, by evaluating their performance against 500 hand-labelled sentences. The overall figure of merit shows near-perfect GCI detection with the proposed method. Additionally, the algorithm can be used for singularity detection in almost any signal provided the minimum separation of singularities is known, as no further assumptions are made about the input signal.

REFERENCES

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

[2] N. D. Gaubitch, P. A. Naylor, and D. B. Ward, "Multi-microphone speech dereverberation using spatio-temporal averaging," in *Proc European Signal Processing Conf*, Vienna, Austria, Sept. 2004, pp. 809–812.

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

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

[5] D. G. Childers, D. M. Hooks, G. P. Moore, L. Eskenazi, and A. L. Lalwani, "Electroglottography and Vocal Fold Physiology," *Journal of Speech and Hearing Research*, vol. 33, no. 2, pp. 245–254, June 1990.

[6] D. M. Howard, "Variation of Electrolaryngographically Derived Closed Quotient for Trained and Untrained Adult Female Singers," *Journal of Voice*, vol. 9, no. 2, pp. 121–1223, June 1995.

[7] 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, March 2004.

[8] M. Huckvale, "Speech Filing System: Tools for Speech," University College London, Tech. Rep., 2004. [Online]. Available: http://www.phon.ucl.ac.uk/resource/sfs

[9] A. Bouzid and N. Ellouze, "Multiscale Product of Electroglottogram Signal for Glottal Closure and Opening Instant Detection," in *Proc. IMACS Multiconference on Computational Engineering in Systems Applications*, vol. 1, 2006, pp. 106–109.

[10] ——, "Open Quotient Measurements Based on Multiscale Product of Speech Signal Wavelet Transform," *Research Letters in Signal Processing*, vol. 2007, pp. 5 – Actual pages not quoted, 2007.

[11] 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 Processing*, vol. 14, 2006.

[12] R. Smits and B. Yegnanarayana, "Determination of Instants of Significant Exitation in Speech using Group Delay Function," *IEEE Trans. Speech Audio Processing*, vol. 5, no. 3, pp. 325–333, September 1995.

[13] J. Makhoul, "Linear Prediction: A tutorial review," *Proc IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.

[14] N. D. Gaubitch, M. K. Hasan, and P. A. Naylor, "Generalized Optimal Step-Size for Blind Multichannel LMS System Identification," *Signal Processing Letters, IEEE*, vol. 13, no. 10, pp. 624–627, Oct. 2006.

[15] B. D. Radlović and R. A. Kennedy, "Nonminimum-phase equalization and its subjective importance in room acoustics," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 6, pp. 728–737, Nov. 2000.

[16] M. R. P. Thomas, N. D. Gaubitch, and P. A. Naylor, "Multichannel DYPSA for estimation of glottal closure instants in reverberant speech," in *Proc European Signal Processing Conf*, Poznan, Poland, Sept. 2007.

[17] M. R. P. Thomas, N. D. Gaubitch, J. Gudnason, and P. A. Naylor, "A Practical Multichannel Dereverberation Algorithm Using Multichannel DYPSA and Spatiotemporal Averaging," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, Oct. 2007.

[18] J. C. Catford, *Fundamental Problems in Phonetics*. Indiana University Press, 1977.

[19] J. Deller, "Some Notes on Closed Phase Glottal Inverse Filtering," *IEEE Trans Acoustics, Speech and Signal Processing*, vol. 29, pp. 917–919, Aug 1981.

[20] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 7, pp. 710–732, 1992.

[21] S. Mallat and W. Hwang, "Singularity detection and processing with wavelets," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 617–643, March 1992.

[22] A. Bouzid and N. Ellouze, "Local Regularity Analysis at Glottal Opening and Closure Instants in Electroglottogram Signal using Wavelet Transform Modulus Maxima," in *Eurospeech*, 2003, pp. 2837–2840.

[23] B. M. Sadler, T. Pham, and L. C. Sadler, "Optimal and Wavelet-Based Shock Wave Detection and Estimation," *J. Acoust. Soc. Amer.*, vol. 104, no. 2, pp. 955–963, Aug. 1998.

[24] B. M. Sadler and A. Swami, "Analysis of multiscale products for step detection and estimation," *IEEE Trans Information Theory*, vol. 45, no. 3, pp. 1043–1051, 1999.

[25] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[26] G. Lindsey, A. Breen, and S. Nevard, "SPAR'S Archivable Actual-Word Databases," University College London," Technical Report, June 1987.