A Novel Filterbank for Epoch Estimation

Pramod Bachhav
EURECOM, Sophia-Antipolis,
France
pramod.bachhav@eurecom.fr

Hemant A. Patil
Dhirubhai Ambani Institute of Information and
communication technology (DA-IICT),
Gandhinagar-382007, India
hemant_patil@daiict.ac.in

***Abstract*—We present a novel approach for epoch estimation from the simple observation of the speech spectrum. Fundamental frequency *(F_0)* of the speech signal and local variations around F_0 are the characteristics of glottal excitation source. Extraction of this information from the speech spectrum can be used to estimate epochs (since higher harmonics interact with the vocal tract characteristics, they no longer represent the true glottal source). In this paper, we bandpass filter the speech signal through a novel Gaussian filterbank followed by simple peak detection to extract epochs. We do not attempt any post processing to study the effectiveness of F_0 on epoch estimation in the proposed method. The algorithm is validated on various databases and compared with four state-of-the-art methods. The method has shown better or comparable results on the clean speech and found to be highly robust to the additive white noise giving highest IDR at various SNR levels.**

***Keywords*- Glottal closure instant (GCI), epoch, fundamental frequency*(F_o)*, spectrum.**

I. Introduction

According to the source-filter model of speech production, a speech signal can be considered as an output of the time-varying and quasi-stationary vocal tract system excited with a glottal source that generates a sequence of glottal pulses [1]. The glottal airflow is spectrally shaped by the vocal tract to produce a speech signal [1]. Simplicity of the source-filter model is that it allows incorporating radiation at lips, which is modeled by differentiation operator, with the glottal excitation source [1]. Therefore, for voiced speech as the glottis closes suddenly, the derivative of the glottal flow excites the vocal tract with strong impulse-like spikes [2]. The time instants corresponding to these impulses are called as Glottal Closure Instants (GCIs) or *epochs*. In addition, sudden closure of the glottis leads to the burst of energy which is manifested as the sharp changes in amplitude of speech signal around GCIs. Therefore, voiced speech can be modeled as *convolution* of the impulse response of the vocal tract system with the impulse-like excitation signal (due to *sudden* closure of glottis) which has impulses (*i.e.*, singularity function) located at epochs [3].

Accurate and robust estimation of GCIs is important in deriving source-based features in many applications like text-to-speech synthesis (TTS), prosody modification, glottal source estimation, speaker recognition, voice conversion, etc [4]. In addition, epochs are necessary for the accurate analysis of the speech signal to extract dynamic characteristics of the vocal tract [2]. Therefore, many signal processing techniques has been developed for estimation of epochs employing different approaches for preprocessing of the speech signals. Linear prediction (LP) analysis combined with some preprocessing such as epoch filtering of linear prediction residual (LPR) [5], unwrapped phase spectrum of short-time Fourier Transform (STFT) of LPR [6], Hilbert envelope of LPR and group delay function [7], has been found useful for epoch estimation. Methods like YAGA [8], DYPSA [9] have employed dynamic programming to reduce the insertions suffered by group delay-based methods. The approaches like ZFR [4], SEDREAMS [10] and recently proposed novel filtering- based approach (FBA) [11] use smoothing of the speech signal to detect epoch locations, getting rid from parameter settings required in LP analysis. However, the performance of ZFR and SEDREMS depend on accuracy of average pitch period estimation, which might fail in severely degraded noisy conditions. Therefore, estimation of epochs is a challenging task if average pitch period is unknown and hence, study reported in [12] proposes dynamic plosion index (DPI)-based epoch extraction algorithm. Even though it is threshold-independent, its performance might be affected by the erroneous estimation of the integrated linear prediction residual (ILPR) [13]. The method exploiting phase spectral characteristics of the speech signal has been proposed for epoch extraction which models the phase spectrum as an allpass filter [14]. In [15], an approach based on subband analysis of LPR is presented. Recently, [16] considers voiced speech segments as a spectra and then exploits group delay spectra to locate GCI candidates.

In this paper, we propose a novel approach for epoch estimation which uses a Gaussian filterbank to bandpass filter the fundamental frequency (F_o) component of the speech signal and the local frequency variations around F_o. The discrete Fourier transform (DFT) of the impulse train (with period T_0) is a train of impulses located at F_o *($=1/T_0$)* and its harmonics. It implies that the F_o of the speech signal is an inherent property of the excitation source. In addition, higher pitch harmonics of the speech signal no longer represent true source characteristics because of their strong interaction with the vocal tract formants [4]. Therefore, extraction of F_o can be useful for epoch estimation. This idea forms the basis of the proposed approach.

The rest of the paper is organized as follows: Section 2 explains the motivation for this work, section 3 details the proposed approach. In section 4, details of the databases used for evaluation and the experimental setup, are discussed. In section 5, performance of the proposed method is analyzed.

ISBN 978-0-9928626-7-1 © EURASIP 2017

II. Significance of the F_o of Speech Signal for Epoch Estimation

Figure 1 illustrates how the epoch-related information is embedded in F_o and the local frequency variations around F_o. Figure 1(c) shows the synthetic speech signal $s[n]$ obtained by passing the impulse train (shown in Figure 1(a)) through a cascade of four 2^{nd} order resonators with center frequencies (corresponding to first *four* formants) *500Hz, 1075Hz, 2463Hz, 3558Hz* (*r=0.99* for narrow -*3*dB bandwidth) [1]. Figure 1(d) shows that the magnitude spectrum of the synthetic signal consists of peaks at $F_o = 80Hz$ and its harmonics with dominant peak around *500* Hz which is the frequency corresponding to the 1^{st} formant.

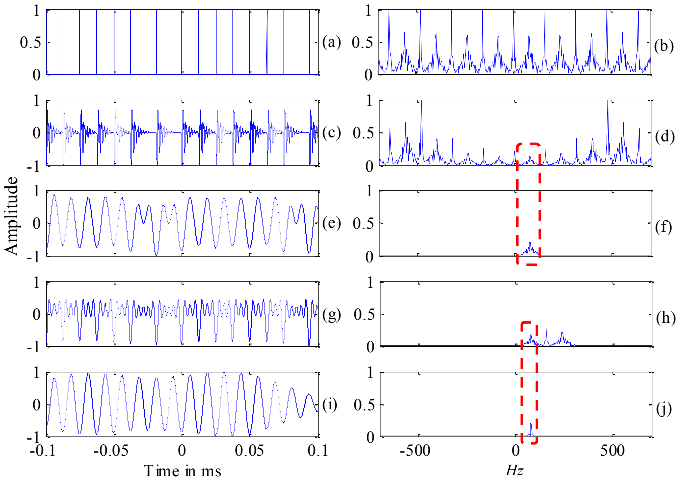

Figure 1. A train of impulses with locally varying period, (b) its magnitude spectrum with F_o=*80* Hz and its harmonics, (c) synthetic speech signal, (d) its magnitude spectrum, (e), (g), (i) are the real parts of the time domain signals with spectra as shown in (f), (h), (j). Red boxes highlight the F_0.

Now, assume that the spectrum consists of an impulse at F_o Hz, i.e., $F(f) = \delta(f - F_0)$.

Taking the inverse DFT we get, $f(t) = e^{j2\pi F_0 t}$.

$$\therefore \quad Real(f(t)) = \cos(2\pi F_0 t).$$

Thus, we extracted the spectrum of the speech signal $s[n]$ between *60 Hz-100 Hz* and took real part of its inverse DFT which is shown the Figure 1(e). It can be observed that the resulting time-domain signal is sinusoidal in nature. In addition, the negative peaks of the resulting signal are able to capture the impulse locations accurately despite of varying pitch period of the impulse train. As shown in Figure 1(g), time-domain signal corresponding to F_o along with the 2^{nd} and 3^{rd} harmonics shows the spurious zero-crossings between two consecutive impulse locations. Because the higher harmonics interact with the frequency response of digital resonator. Therefore, we claim that the fundamental component of the speech signal can be extracted to estimate the epoch locations. The reason behind taking range of the frequencies around F_o is to capture dynamic variations in the pitch period. Extracting a very narrow spectrum around F_0 (Figure 1(j)) gives a time-domain signal (Figure 1(i)) which is unable to capture local variations in the pitch period. Therefore, in this paper, we design a novel Gaussian filterbank to bandpass the F_o and local variations in frequency around F_o, for epoch extraction.

III. Proposed Algorithm

A. Pre-processing

The speech signal $s[n]$ was passed through a filterbank with *L Gaussian* linear time invariant (*LTI*) filters (as they exhibit optimal time-frequency resolution [17], [18]). Figure 2 shows the block diagram for the proposed approach. As a linear-phase FIR filterbank was used, corresponding subband filter outputs were summed to get $y[n]$.

$$\therefore \quad y[n] = \sum_{i=1}^{L} \left(h_i[n] * s[n]\right), \qquad ,i = 1, \ldots, L,$$

with impulse response of the i^{th} subband filter is given by,

$$h_i[n] = w[n] e^{j2\pi f_i t}, \quad i = 1, \ldots, L,$$

$$\text{where } w(n) = e^{-\frac{1}{2}\left(\frac{n}{\sigma}\right)^2}, -\frac{N-1}{2} \le n \le \frac{N-1}{2}, \sigma = \frac{N}{2\alpha}.$$

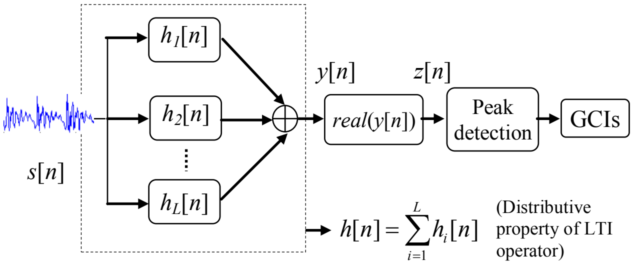

Figure 2. Block diagram of the proposed filterbank for GCI detection.

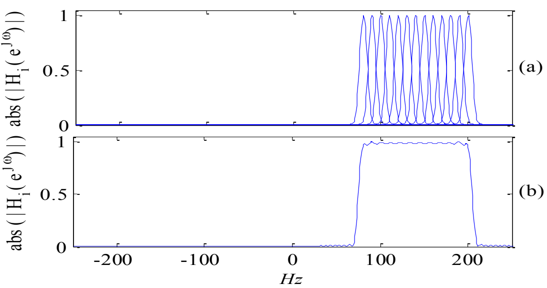

Figure 3. a) Magnitude responses of the L subband filters with f_1=*80 Hz* and f_L=*200 Hz* with f_{i+1}-f_i=*10 Hz*, *i=1,...L-1*,(b) equivalent magnitude response of the filterbank.

DFT of the Gaussian window $w[n]$ is a Gaussian and thus, has a fast decay in the frequency-domain implying smooth time-domain Gaussian window with continuous and bounded derivatives [17]. As multiplication with exponential in the time-domain results in shift in the frequency-domain, magnitude response $|H_i(e^{j2\pi f})|$ of the i^{th} subband filter is Gaussian in nature centered at f_i as shown in Figure 3(a). The center frequencies were spaced in steps of *10 Hz* so that the frequency range f_1-f_L covered the F_0 of the speech signal and variations around F_0 excluding the second and higher harmonics. f_1 and f_L were decided by observing F_0 from the spectrum of the speech signal. The parameters of the window $w[n]$ were selected to be *N=6071* and *α=2.5* for F_s=32 *kHz*. The choice of parameters was to attain nearly constant gain

ISBN 978-0-9928626-7-1 © EURASIP 2017

bandpass characteristics for equivalent magnitude response $|H(e^{j2\pi f})|$ of the filterbank with sharp cut-off at f_l and f_L as shown in Figure 3(b). As $H_i(e^{j2\pi f})$ has causal spectrum, $y[n]$ is analytic in time-domain (by *duality* property of DFT). Therefore, we take its real part to get sinusoidally varying signal $z[n]$,

$$z[n] = Real(y[n]).$$

B. *Peak detection*

Our experimental observations depicted that $z[n]$ exhibits prominent negative peaks around true epochs along with some spurious peaks. To choose a negative peak which corresponds to true epoch, we take maximum negative peak of $z[n]$ between two successive negative zero-crossings which makes the proposed algorithm threshold independent. In order to observe effectiveness of F_0 on GCI detection, we do not employ any post-processing to eliminate spurious negative peaks. Figure 4 illustrates the epoch estimation for a voiced speech segment.

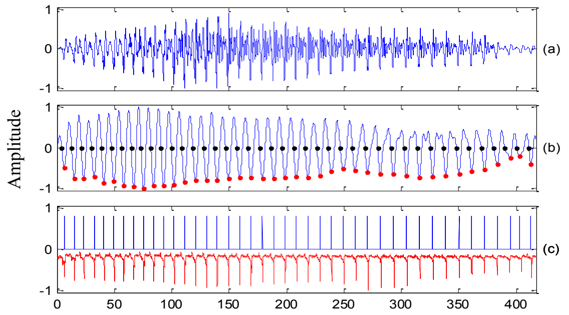

Figure 4. (a) A voiced speech segment of male speaker BDL [19], (b) filtered signal (f_l=*80* and f_L=*170 Hz*) with dotted zero-crossings (black) and negative peaks (red), (c) upper trace-detected epochs, lower trace – delay adjusted differenced electroglottograph (EGG).

IV. EXPERIMENTAL RESULTS

A. *Experimental Setup*

The proposed algorithm was evaluated on CMU-ARCTIC [19] and MOCHA databases [20] which consist of speech signals along with electroglottograph (EGG) signals. The CMU-ARCTIC database (DB) consists of phonetically balanced utterances spoken by speakers SLT, JMK and BDL, TIMIT utterances spoken by KED speaker, a set of nonsense words containing all phone-phone transitions spoken by speaker RAB. MOCHA DB consists of a set of short sentences, uttered by male (M1) and a female (M2) speaker. Details of the databases are given in Table 1. The cut-off frequencies f_l and f_L were selected by observing the spectra of few randomly chosen speech files for each speaker. As the range f_l-f_L to cover F_0 and the local variations do not vary much for a particular speaker, they were kept same for all files of speaker. The maximum negative peaks in the derivative of the EGG (DEGG) were taken as reference epochs (i.e. the ground truth) after adjusting larynx-to-microphone delay of *0.7 ms* for CMU DB [4]. Algorithms were evaluated in the voiced regions with voiced-unvoiced decision made by applying threshold of *1/9* on maximum negative value of DEGG [12]. MOCHA speech and DEGG files were already synchronized and voicing decision was made by applying threshold of *1/6* on maximum positive value of DEGG after passing through *15*-point moving average (MA) filter. The performance measures used for evaluation are: identification rate (IDR), miss rate (MR), false alarm rate (FA), identification accuracy (IDA) and accuracy to 0.25 *ms* (±0.25 Acc.) as defined in [9]. All utterances were resampled at *32 kHz*.

Table 1: Details of the databases used for evaluation

	CMU-ARCTIC DB					MOCHA DB	
Speaker	SLT	BDL	JMK	RAB	KED	M1	M2
No. of utterances	1132	1131	1114	1946	424	460	460
Gender	F	M	M	M	M	M	F
Native	US	US	Canada	UK	US	*	*
$\boldsymbol{f_1(Hz)}$	**100**	**80**	**80**	**70**	**80**	**70**	**100**
$\boldsymbol{f_L(Hz)}$	**250**	**170**	**170**	**120**	**140**	**160**	**300**

*Subjects have variety of accents of English language.

Table 2: Comparison of the results on the clean speech signals

Speaker	Method	IDR (%)	MR (%)	FA (%)	IDA (ms)	±0.25 Acc.(%)
BDL	Proposed	96.65	**0.08**	3.27	0.36	69.76
	DPI	**98.53**	0.28	1.20	**0.32**	**84.04**
	ZFR	97.29	0.11	2.60	0.38	74.71
	SEDREAMS	98.44	0.41	1.15	0.42	81.81
	FBA	98.43	0.98	**0.59**	0.34	63.27
JMK	Proposed	99.50	**0.03**	0.47	0.60	33.48
	DPI	99.02	0.25	0.73	**0.37**	**78.00**
	ZFR	**99.52**	0.09	0.38	0.66	33.58
	SEDREAMS	98.89	0.65	0.46	0.65	67.24
	FBA	97.54	2.18	**0.28**	0.55	47.62
SLT	Proposed	99.20	**0.01**	0.79	**0.21**	**79.37**
	DPI	99.26	0.39	**0.35**	0.27	75.40
	ZFR	**99.47**	0.03	0.50	0.35	78.63
	SEDREAMS	99.46	0.06	0.48	0.32	73.00
	FBA	98.94	0.42	0.64	0.28	69.12
KED	Proposed	99.02	0.15	0.83	0.72	30.90
	DPI	99.30	0.14	0.56	**0.23**	**96.79**
	ZFR	99.38	**0.07**	0.55	0.68	31.92
	SEDREAMS	**99.54**	0.14	**0.32**	0.55	78.96
	FBA	92.06	2.52	5.42	0.91	65.10
RAB	Proposed	**99.03**	0.06	0.90	0.78	29.63
	DPI	96.33	0.14	3.53	**0.43**	**89.34**
	ZFR	96.74	**0.05**	3.20	0.75	33.38
	SEDREAMS	97.43	0.05	2.52	0.69	76.73
	FBA	94.05	0.19	5.75	0.85	37.11
M1	Proposed	**96.43**	**2.72**	0.95	0.84	29.43
	DPI	96.19	3.09	0.72	0.58	**62.41**
	ZFR	96.41	2.79	0.80	0.85	28.00
	SEDREAMS	96.15	3.20	0.65	0.70	46.89
	FBA	94.92	4.59	**0.49**	0.82	35.15
M2	Proposed	97.02	**1.05**	1.93	0.39	54.20
	DPI	96.87	2.88	**0.26**	**0.29**	**77.44**
	ZFR	**97.87**	1.38	0.75	0.40	38.49
	SEDREAMS	97.86	1.53	0.61	0.48	55.77
	FBA	97.02	**1.05**	1.93	0.39	54.20

B. *Performance Analysis*

Table 2 compares the performance of the proposed GCI estimation method on clean speech with three state-of-the art methods, DPI [12], SEDREAMS [10], ZFR [4] and our recently proposed FBA [11]. The evaluation was carried on 7 speakers In addition, performance of the proposed algorithm was studied on the degraded speech signals, for its robustness against noise, at various Signal-to-Noise Ratio (SNR) levels.

ISBN 978-0-9928626-7-1 © EURASIP 2017

Table 3: Comparison of the epoch extraction techniques for additive white noise at various SNR levels averaged over *7* speakers.

	Proposed				DPI				ZFR				SEDREAMS				FBA			
SNR(dB)	-10	-5	0	5	-10	-5	0	5	-10	-5	0	5	-10	-5	0	5	-10	-5	0	5
IDR (%)	**94.4**	**96.6**	**97.6**	**97.9**	83.6	91.2	95.5	97.0	92.7	95.8	97.2	97.8	86.4	93.1	96.7	97.8	78.2	85.7	91.2	94.0
MR (%)	**1.52**	**0.92**	**0.74**	**0.66**	8.52	4.61	2.21	1.33	4.43	2.28	1.32	0.91	12.1	5.84	2.39	1.25	6.72	5.22	3.36	2.41
FA (%)	4.13	2.50	1.68	1.41	7.91	4.15	2.31	1.63	**2.84**	1.96	1.50	1.31	1.56	**1.06**	**0.92**	**0.91**	15.1	9.08	5.46	3.58
IDA (%)	**0.90**	0.73	0.63	0.58	0.96	**0.65**	**0.48**	**0.43**	1.14	0.80	0.65	0.62	1.40	1.10	0.85	0.73	1.05	0.82	0.71	0.64
±*0.2* Acc.	30.1	36.6	41.5	44.2	**38.9**	**54.5**	**63.8**	**67.8**	28.2	37.2	42.9	45.8	19.3	26.9	36.45	45.3	34.3	44.6	50.3	51.5

Table 4: Comparison of the epoch extraction techniques for babble noise at various SNR levels averaged over *7* speakers.

	Proposed				DPI				ZFR				SEDREAMS				FBA			
SNR(dB)	-10	-5	0	5	-10	-5	0	5	-10	-5	0	5	-10	-5	0	5	-10	-5	0	5
IDR (%)	83.5	88.4	92.8	95.8	71.0	77.7	85.6	92.4	**83.8**	88.1	92.3	95.3	82.7	**88.5**	**93.1**	**96.2**	70.6	75.8	82.0	88.2
MR (%)	**3.36**	**2.21**	**1.26**	**0.83**	11.6	9.53	6.58	3.71	6.99	5.42	3.54	2.13	12.5	8.39	4.93	2.43	9.20	8.25	6.97	5.10
FA (%)	13.2	9.42	5.93	3.37	17.4	12.8	7.79	3.94	9.26	6.52	4.17	2.53	**4.80**	**3.10**	**2.02**	**1.36**	20.2	16.0	11.01	6.69
IDA (%)	**1.27**	**1.08**	0.90	0.74	1.62	1.31	**0.89**	**0.61**	1.47	1.20	0.92	0.72	1.56	1.27	0.96	0.76	1.53	1.27	1.00	0.79
±0.25Acc.	17.9	22.6	28.4	35.1	**19.5**	**28.6**	**52.6**	**63.3**	17.1	22.8	29.5	36.2	17.3	26.3	42.4	51.3	19.0	25.5	35.4	45.1

For this purpose, a noise sample from NOISEX-92 DB added to every utterance at respective SNRs [21]. Tables 3 & 4 illustrate the comparison of performances of four methods averaged over seven speakers for various SNR levels. ZFR and SEDREAMS were provided with mean pitch period from the clean speech for noisy evaluations.

The proposed method outperforms all the existing methods in terms of relatively least MR for several speakers and in terms best IDR for RAB, M1 & M2 whereas gives almost comparable IDR for BDL,JMK & SLT. DPI gives lowest IDA for all speakers except SLT. The proposed method and ZFR give low ±0.25 Acc. for few cases. This is because these methods do not refine the estimated epoch locations using LPR or voice source [12]. The combined effect of MR and FA reflects on the IDR, both being the least, give the higher IDR. So when the IDRs are averaged over all the speakers (Table 5) then the proposed method gives highest IDR after SEDREAMS. It can be observed that the proposed method gives relatively higher FA for the speaker BDL. The proposed method is found to be noise robust giving the best IDR and least MR at various SNRs for additive white noise. For babble noise, it gives the least MR with *slightly* less IDR than the best.

Table 5: Performance measures averaged over all the speakers

	Method	IDR	MR	FA	IDA	±0.25 Acc.
All	Proposed	**98.12**	**0.59**	**1.31**	0.56	46.68
	DPI	97.93	1.02	1.05	0.36	80.49
	ZFR	98.10	0.65	1.25	0.58	45.53
	SEDREAMS	**98.25**	**0.86**	**0.88**	0.54	68.63
	FBA	95.49	2.17	2.33	0.63	52.38

Figure 5 illustrates the significance of the frequency variations around F_0 in detail on the GCI estimation. It is evident from Figure(a) and its zoomed versions (as shown in (b), (c)) that the F_0 of the speech signal (for BDL male) falls roughly within 80-160 *Hz*. The IDR is maintained above 99 % for 140<f_L (*Hz*)<170 with f_l =80 *Hz* whereas immediately starts to decrease for f_L>170 *Hz*. In addition, Figure 5(e) shows the IDR variation for utterance of female speaker (SLT) where IDR remains above 99% for *200<*f_L*(Hz)<290*. Hence, the proposed approach estimates GCIs with simple observation of spectrum followed by subband filtering and peak detection which is threshold independent.

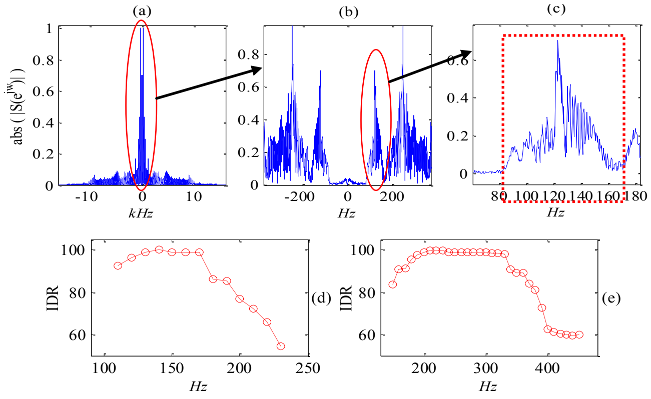

Figure 5. Illustration of the significance of F_0 on IDR, (a) magnitude spectrum of a male (BDL) speech utterance, (b)-(c) zoomed portions of (a), (d) variation in IDR with f_L varying from 120 Hz-240 *Hz* and f_l*=80 Hz*, (e) variation in IDR with f_L varying from 150 Hz-450 *Hz* and f_l*=100 Hz*, for a speech utterance of female (SLT).

I. SUMMARY AND CONCLUSIONS

In this paper, a novel approach based on simple observation of the speech spectrum is presented which extracts F_0 of the speech signal using a Gaussian filterbank. The proposed method is found to be better or comparable than existing state-of-the-art methods and working well on wide range of databases giving best or comparable results in severe additive noisy conditions at various SNR levels. Therefore, extraction of F_0 and variations around F_0 from the speech spectrum is useful for epoch estimation and is illustrated without post-processing and a threshold-independent approach. Our future research will be directed towards validation of the proposed method on extracting GCIs from singing voice and robustness evaluation for additional noisy environments.

II. ACKNOWLEDGEMENTS

The authors thank Department of Electronics and Information Technology (DeitY), New Delhi Govt. of India for the sponsored ASR-phase II consortium project and authorities at DAIICT, Gandhinagar. Authors also thank Dr. A.P. Prathosh for his help.

REFERENCES

[1] T. Quatieri, Discrete-Time Speech Signal Processing: Principles and Practice, 2nd ed., Delhi. Prentice Hall, 2002.

[2] B. Yegnanarayana and R. Veldhuis, "Extraction of vocal-tract system characteristics from speech signals," *IEEE Trans. on Audio, Speech, and Language Processing,* vol. 6, no. 4, pp. 313-327, 1998.

[3] T. Ananthapadmanabha and B. Yegnanarayana, "Epoch extraction of voiced speech," *IEEE Trans. on Acoustics, Speech and Signal Processing,* vol. 23, no. 6, pp. 562-570, 1975.

[4] K. Murty and B. Yegnanarayana, "Epoch extraction from speech signals," *IEEE Trans. on Audio, Speech, and Language Processing,* vol. 16, no. 2, pp. 1602-1613, November 2008.

[5] T. Ananthapadmanabha and B. Yegnanarayana, "Epoch extraction from linear prediction residual for identification of closed glottis interval," *IEEE Trans. on Acoustics, Speech and Signal Processing,* vol. 27, no. 4, pp. 309-319, 1979.

[6] R. Smits and B. Yegnanarayana, "Determination of instants of significant excitation in speech using group delay function," *IEEE Trans. on Speech and Audio Processing,* vol. 3, no. 5, pp. 325-333, 1995.

[7] K. Sreenivasa Rao, S. Prasanna and B. Yegnanarayana, "Determination of instants of significant excitation in speech using Hilbert envelope and group delay function," *IEEE Signal Processing Letters,* vol. 14, no. 10, pp. 762-765, 2007.

[8] M. Thomas, J. Gudnason and P. Naylor, "Estimation of glottal closing and opening instants in voiced speech using the YAGA algorithm," *IEEE Trans. on Audio, Speech, and Language Processing,* vol. 20, no. 1, pp. 82-91, 2012.

[9] P. A. Naylor, A. Kounoudes, J. Gudnason and M. Brookes, "Estimation of glottal closure instants in voiced speech using the DYPSA algorithm," *IEEE Trans. on Audio, Speech, and Language Processing,* vol. 15, pp. 34-43, 2007.

[10] T. Drugman and T. Dutoit, "Glottal closure and opening instant detection from speech signals.," in *INTERSPEECH*, Brighton, UK, 2009.

[11] P. Bachhav, H. Patil and T. Patel, "A novel filtering based approach for epoch extraction," in *International Conference Acoustic Speech and Signal Processing (ICASSP)*, Brisbane, Australia, April, 2015.

[12] A. Prathosh, T. Ananthapadmanabha and A. Ramakrishnan, "Epoch extraction based on integrated linear prediction residual using plosion index," *IEEE Trans. on Audio, Speech, and Language Processing,* vol. 21, no. 12, pp. 2471-2480, 2013.

[13] P. Sujit, A. Prathosh, A. Ramakrishnan and P. Ghosh, "An error correction scheme for GCI detection algorithms using pitch smoothness criterion," in *INTERSPEECH*, Dresden, Germany, 2015.

[14] K. Vijayan and K. Murty, "Epoch extraction from allpass residual of speech signals," in *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, Florence, Italy, 2014.

[15] R. Vikram, K. Girish, S. Harshavardhan, A. Ramakrishnan and T. Ananthapadmanabha, "Subband analysis of linear prediction residual for the estimation of glottal closure instants," in *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, Florence, Italy, 2014.

[16] A. Rachel, P. Vijayalakshmi and T. Nagarajan, "Estimation of glottal closure instants from telephone speech using a group delay-based approach that considers speech signal as a spectrum," in *INTERSPEECH*, Dresden, Germany, 2015.

[17] S. Mallat, A Wavelet Tour of Signal Processing, Academic press, 1999.

[18] D. Gabor, "Theory of Communication. Part 1: The analysis of information.," *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering,* vol. 93, no. 26, pp. 429-441, 1946.

[19] "CMU-ARCTIC Speech Synthesis Databases," Available [Online] : http://festvox.org/cmu_arctic/index.html {Last Accessed: April 20, 2015}.

[20] A. Wrench, "The MOCHA-TIMIT articulatory database," Available [Online] : http://www.cstr.ed.ac.uk/artic/mocha.html {Last Accessed: September 15, 2015}.

[21] "NOISEX-92," Available [Online] : http://www.speech.cs.cmu.edu/comp.speech/Section1/Data/noisex.html.

ISBN 978-0-9928626-7-1 © EURASIP 2017