Estimation of voice-onset time in continuous speech using temporal measures

A. P. Prathosh

Department of Electrical Engineering,

Indian Institute of Science, Bangalore,

India 560012, email: prathoshap@ee.iisc.ernet.in

A. G. Ramakrishnan

Department of Electrical Engineering,

Indian Institute of Science, Bangalore,

India 560012, email: ramkiag@ee.iisc.ernet.in

T. V. Ananthapadmanabha

Voice and Speech Systems, Malleshwaram,

Bangalore, India 560003, email: tva.blr@gmail.com

(Dated: 6th May 2014)

Abstract

This paper proposes an automatic acoustic-phonetic method for estimating voice-onset time of stops. This method requires neither transcription of the utterance nor training of a classifier. It makes use of the plosion index for the automatic detection of burst onsets of stops. Having detected the burst onset, the onset of the voicing following the burst is detected using the epochal information and a temporal measure named the maximum weighted inner product. For validation experiments are carried out on the entire TIMIT database and two of the CMU Arctic corpora. The performance of the proposed method compares well with three state-of-the-art techniques.

PACS numbers: 43.72.Ar

I. INTRODUCTION

A. Motivation

The production of a stop consonant comprises multiple sub-phonetic events namely the closure interval, the burst onset, the aspiration interval (if any) and the voice onset time (when followed by a voiced phone)[1]. Voice-onset time (VOT) is defined as the interval between the onset of the stop-burst and the onset of the laryngeal vibrations succeeding the burst[2]. It has been extensively studied due to its wide utility. It is an important temporal attribute to discriminate between 'voiced' and 'unvoiced' stops[2], especially when the stops are in word-initial position. It also has applications in psychoacoustic studies[3] and accent identification[4]. It is shown in previous studies[5,6] that inclusion of VOT as an additional feature can improve the phone recognition rate of an automatic speech recognition system. VOT is routinely measured in the context of clinical research studies[7] as related to aphasia, apraxia, etc.

Automatic methods for the measurement of VOT are required in order to reduce the human labor involved in manual measurements and for applications such as automatic speech recognition and accent identification. Several automatic methods have been proposed for the measurement of VOT which broadly fall into two categories: (a) those which explicitly identify the locations of the burst and voicing onsets through a set of customized acoustic-phonetic rules (knowledge-based)[4,6], (b) those which train a learning machine (such as random forest, support vector machine) to estimate the VOT using some acoustic features corresponding to the stop-to-voiced-phone transition event[8,9].

Many of the methods, which report high performance, require phonetic transcription. Some of them use the transcription to identify the segment of the speech signal containing the stop consonant through forced-alignment[4,9]; others use this information to focus the analysis on segments of the signal containing only one stop consonant[8]. Such methods are difficult to employ in a scenario where there is no transcription available. Methods based on statistical classifiers employ training

with high-dimensional feature vectors. Further some methods consider only word-initials stops because the role of VOT in discriminating between voiced and unvoiced stops is more prominent in such occurrences. In this paper we propose an automatic rule-based algorithm for estimating the VOT of both voiced and unvoiced stops occurring at initial and medial positions. This method does not require any *a priori* transcription. This method uses temporal features derived from the examination of the acoustic-phonetic characteristics of the stops and voiced phones. It is validated on the TIMIT database and is compared with three state-of-the-art algorithms.

B. Problem formulation

The problem of automatic estimation of VOT from a given speech signal may be looked upon as a two-stage process: (i) Automatic detection of the instants of the burst onsets corresponding to the stop consonants; (ii) given a burst onset, detection of the onset of the voicing in the following voiced phone (hereafter referred to as the voice onsets). For the former problem of detecting the burst onsets of stops we adopt the solution proposed in our earlier work[10]. In this paper we address the latter problem of detection of the voice onsets.

By the term 'voice onset' we mean the instant at which the laryngeal vibrations begin in the voiced phone (vowels, liquids, semi-vowels, nasals etc.) following the stop under examination. However for some voiced stops (mostly occurring at the word-medial positions) there is a pre-voicing component throughout or for a partial interval of the closure duration. In the literature such stops are said to possess a negative VOT. In this study we consider only the problem of measuring the interval between the burst onset of a stop consonant and the onset of laryngeal vibrations following it, i. e., estimation of positive VOT. The problem of estimating the negative VOT is reserved to our future work.

II. PROPOSED METHOD

In this study the features proposed are based on the temporal cues of the phones under examination. It has been mentioned in an earlier work that the VOT can be more reliably estimated using temporal analysis[11].

A. Maximum Weighted Inner-Product (MWIP)

Inner product is a well known measure used to quantify the similarity between any two vectors. If a segment of the speech signal corresponding to a voiced phone taken between two successive epochs is considered a vector then two such successive vectors possess a high degree of similarity since the response of the vocal tract transfer function corresponding to these segments are highly correlated. Thus the inner product between such segments corresponding to a voiced phone is expected to be higher than that for other phones. Throughout this paper, by the term 'epoch' we mean the instant (point in time) of significant excitation of the vocal tract within a pitch period[12]. An interested reader may refer to more detailed articles[13,14] for a discussion on instant or event based speech processing and epochs.

Further for voiced phones there is a significant amount of energy in the frequency band around the fundamental frequency (F_0) due to the excitation of the supralaryngeal chambers by the voice-source pulse. Equivalently the ratio of the energy within a narrow band of frequencies around F_0 to the total energy is usually higher for a voiced phone than for other phones. The aforementioned characteristics of a voiced phone are quantified using a temporal measure named the weighted inner-product defined between two segments of a speech signal as follows.

Let $s_1[n]$ and $s_2[n]$ be two equal-length segments of a speech signal and $s_1^{'}[n]$ and $s_2^{'}[n]$ be their band-pass-filtered versions. Let $\rho_{s_i^{'}/s_i}$ as the ratio of the l_2 norms of the signals $s_{i^{'}}$ and s_i, respectively. Let $w_i[n] = \rho_{s_i^{'}/s_i}.s_i[n]$ where $i = \{1,2\}$. Now, the weighted inner-product, $w_{s_{1,}s_2}$ between $s_1[n]$ and $s_2[n]$ is defined as $w_{s_{1,}s_2} = \langle w_1[n], w_2[n] \rangle$, where $\langle x,y \rangle$ denotes the Euclidean

inner-product between the vectors x and y. The band-pass filter used here for the computation of WIP is an IIR Butterworth filter of the 4th order with lower and upper 3-dB frequencies at $(0.5).F_{mod}$ and $2.F_{mod}$ respectively, where F_{mod} is the frequency corresponding to the mode of the distribution of all the inter-epoch intervals computed over voiced regions of an entire utterance.

In this work WIP is computed for the speech signals between every pair of successive inter-epoch intervals (IEI), where an IEI is the interval between two successive epochs. This ensures that the beginnings of the segments on which the WIP is being computed coincides with the epochs of the corresponding laryngeal cycles. The computation of WIP needs the segments under consideration to be of equal length. Thus the segments of speech are zero-padded to ensure equal length. The DPI algorithm used here for epoch extraction is shown to place the epochs accurately at instants of significant excitation for voiced phones and at random locations for unvoiced phones[15].

The DPI algorithm has been shown to be temporally very accurate up to 0.25 ms of the true epochs[15]. However since the value of the WIP depends largely on the temporal alignment of the vectors the error made by the epoch extraction algorithm may affect the value of WIP even when the vectors are 'similar'. Hence for a given pair of signals we compute WIP at all lags up to 0.25 ms ($\pm$4 samples at 16 kHz) and use the maximum of those values of WIPs (abbreviated as the MWIP) as one of the temporal measures. The value of MWIP computed between a successive pair of inter-epoch intervals is assigned to the entire former inter-epoch interval. This makes MWIP for an entire utterance a staircase function with a jump discontinuity at every epoch. The MWIP values computed for an entire utterance are normalized by its maximum value for that utterance to ensure that MWIP lies between 0 and 1. MWIP is utilized for voice onset detection by using a threshold as described in Sec. II. C.

B. Zero-crossing difference (ZCD)

It is observed that the MWIP is occasionally high during aspiration intervals of some stops (especially unvoiced velar stops) due to the presence of significant low frequency components and random noise-like structure of the aspiration interval. In order to differentiate such segments from the voice onsets, one more temporal measure termed the zero-crossing difference (ZCD) is proposed.

For a voiced sonorant phone, since the frequency contents of the signal over successive pitch periods do not differ significantly, the difference between the number of zero-crossings in two successive inter-epoch intervals is considerably low. This is not the scenario in the case of aspiration interval because the epochs for unvoiced phones such as stops are placed at random locations and zero-crossing patterns over unequal intervals (between such successive random epochs) are likely to be dissimilar due to the noise-like nature of the aspiration interval. Thus the absolute difference between the number of zero-crossings in two successive inter-epoch intervals called the ZCD, serves as a cue to distinguish between aspiration intervals and voice onsets. As in the case of MWIP, ZCD computed for two inter-epoch intervals is assigned to the entire first inter-epoch interval.

To demonstrate the utility of the MWIP and ZCD we illustrate in Fig. 1 a stop-sonorant segment which comprises a velar stop with a long aspiration interval. It is seen that while MWIP is high over both the aspiration interval and the sonorant, the ZCD (whose value is scaled down by a factor of 20 for ease of visual comparison) is high only over the aspiration interval and low for the sonorant. Thus MWIP and ZCD are jointly used as temporal measures to detect the voice onsets.

C. The voice onset detection algorithm

Burst onsets of the stop consonants are detected using the algorithm proposed in our previous work[10] (The parameters and thresholds of the algorithm used are those which offer the equal error

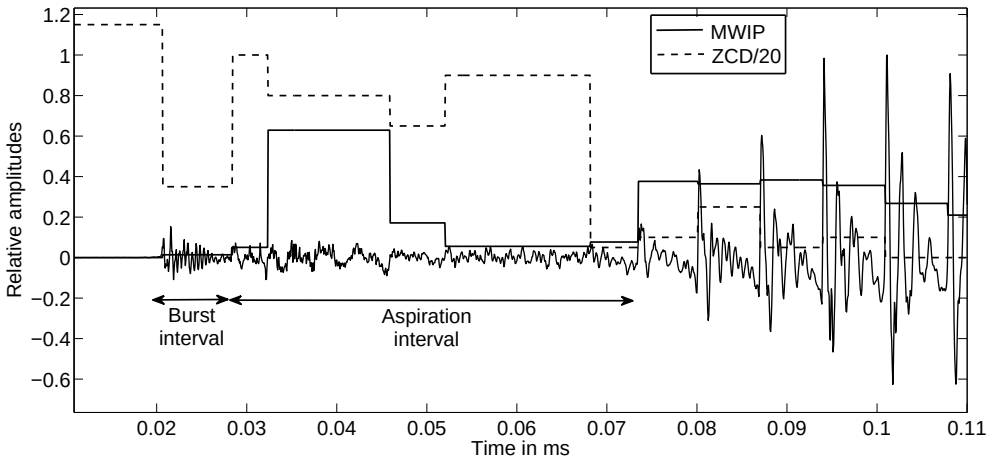

Figure 1: Illustration of the utility of MWIP (solid line) and ZCD (dotted line) as features for voice onset detection from a segment of speech from the TIMIT database (a velar stop followed by a voiced sonorant). While MWIP is high over both the aspiration interval and the sonorant, the ZCD (value scaled down by a factor of 20 for ease of visual comparison) is high over the aspiration interval and low for the sonorant.

rate). For every detected stop-burst, the subsequent voice onset is detected as follows:

1. Let the epoch closest to the detected burst onset be denoted by e_i.
2. Determine whether MWIP over both of the two successive inter-epoch intervals starting from e_i is greater than a threshold T_1 (criterion 1).
3. If criterion 1 is met, determine whether the ZCD over both of the two successive inter-epoch intervals starting from e_i is less than another threshold T_2 (criterion 2).
4. If both the criteria are, met call the e_i^{th} epoch the voice onset and terminate.
5. If either of the criteria is not met, then update e_i to e_{i+1} and repeat steps 1-3 till the voice onset is detected. (The search interval is up to 120 ms, which is assumed to be the longest possible VOT based on the observations of Lisker and Abramson in their study[2] across 18 languages.)

The thresholds T_1 and T_2 are chosen as the modes of the histograms of minimum MWIP and maximum ZCD, respectively, for voiced phones from an arbitrarily chosen small development set (50 samples) taken from the TIMIT training database. The minimum and the maximum required for the histograms are computed over the entire labeled segment of a given phone. The values of T_1 and T_2 thus obtained are 0.06 and 6, respectively.

D. Reference instants for the measurement of VOT

Based on the burst onset detected using the algorithm reported in our earlier work[10] and voice-onset locations detected using the one mentioned in the above section the reference instants are identified as follows for accurate estimate of VOT.

1. **Reference location within the burst interval** - In our earlier work proposed to detect the closure-burst transition (CBT) boundary of a stop[10], the very first instant within a stop burst where the feature plosion index (PI) exceeds a threshold was taken to be the representative CBT for that stop. This may correspond to the beginning of the pre-frication interval. However for measuring VOT, the location at which the value of the plosion index (PI) measure[10] is maximal within an interval between the CBT and the detected voice onset is taken to be the reference instant for burst onset. The rationale for this approach is that the values of the PI within the burst-interval of a stop (interval between the CBT and the voice onset) represent the strengths of the release and the instant with the maximum value serves as a 'better' choice for the burst onset. Often the transcribers also tend to mark this point as the burst onset.

2. **Reference location for the voice-onset instant** - The voice onset should correspond to the very first epoch occurring at the onset of the voiced phone. However the epoch extraction algorithm occasionally misses the very first epoch or the weighting factor ρ may be very low such that MWIP does not exceed the desired threshold. Hence the initial estimates of the voice onsets must be refined. Epochs manifest as prominent negative peaks in the

voice-source signal[15]. Thus the integrated linear prediction residual (ILPR)[16], which is an approximation to the voice-source, is computed for a segment of speech of duration two modal-pitch periods on either side of the initial estimate. The negative extrema of the ILPR are determined and the first extremum which is at least 0.5 times the maximum negative peak in the ILPR is taken to be the final estimate for the voice onset.

Figure 2 depicts a typical case of an unvoiced stop followed by a vowel (taken from the CMU Arctic database) with the initial and refined estimates of the corresponding burst and voice onsets. The corresponding differentiated EGG (dEGG) signal is also shown. It is well known that the negative peaks in the dEGG signal correspond to the epochs[17]. Solid and dot-dash arrows represent the initial and refined estimates of the burst onsets, respectively. Solid and dot-dash downward arrows represent the initial and refined estimates of the voice onset, which coincide in this case. It is seen that the voice onset is detected with a reasonable accuracy as it almost coincides with the very-first negative peak in the dEGG signal following the stop which corresponds to the first glottal closure instant.

III. EXPERIMENTS AND RESULTS

A. Databases and performance measures used

The TIMIT database[18] contains 6,300 utterances hand-labelled at the phone level as spoken by 630 speakers of several dialects of North American English. The algorithm proposed herein for automatically identifying VOT is tested against the hand-placed labels. Further the speech data of two speakers, KED (male) and SLT (female) from the CMU-Arctic database[19] is considered for validating only the detection of voice-onset instants. The CMU Arctic database was created for the purpose of development of TTS systems. This database contain simultaneous EGG recordings along with the acoustic waveform.

The measure used to quantify the performance of the algorithm is the percentage of times the

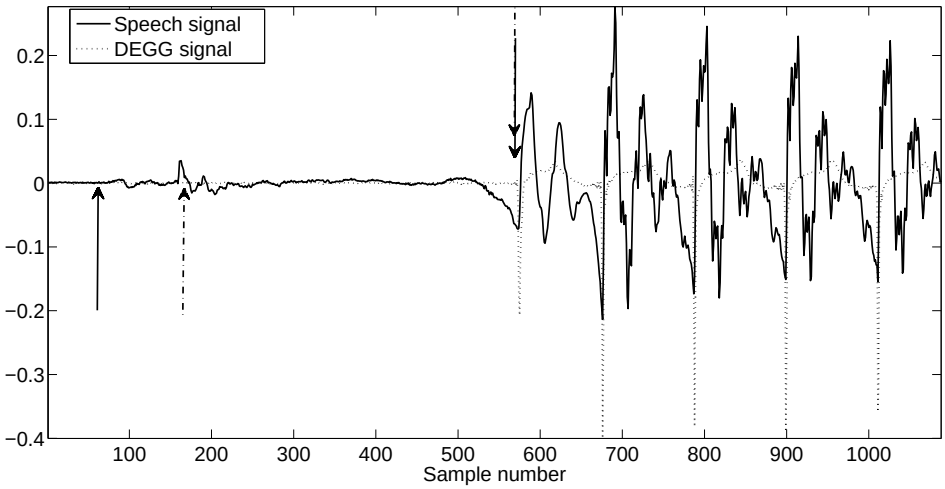

Figure 2: Illustration of the burst and voice onsets detected by the algorithm on a segment of speech from the CMU Arctic database (KED). The acoustic waveform is shown by the solid line and the dEGG signal by the dotted line. Upward and downward arrows denote the estimates of the burst and voice onsets, respectively. In both cases, solid and dot-dash arrows represent the initial and final estimates, respectively. The initial and refined estimates of the voice onset coincide in this case. It is seen that the detected voice onset coincides with the first negative peak in the dEGG.

estimated VOT (or the voice-onset instant in the case of CMU Arctic) is within certain temporal tolerances (5 to 25 ms) of the ground truth. The ground truth is taken to be the hand-labeled boundaries of the burst and voice onsets for the TIMIT database. For the CMU Arctic database, the ground truth for voice onsets is computed automatically using the dEGG signal since phone-level transcriptions are unavailable. It is known that a negative threshold on dEGG signal separates voiced from unvoiced speech[17]. Hence the boundaries between the obstruent and voicing for the following voiced phone are obtained by applying a negative threshold to dEGG, where obstruents can be stops, affricates or fricatives. Within such segments, the voice onset detection algorithm is applied and the temporal deviation of the detected voice onset from the very first peak in the dEGG signal is taken as the performance measure. While validating with the CMU database, the relative delay between the EGG signal and the acoustic signal is compensated for manually for

each speaker. Note that this validates only the detection of the voice onset following any unvoiced phone, of which the problem considered here is a subset. The usage of the CMU Arctic database serves two purposes: (i) objective validation of the algorithm for detection of voice onsets using the EGG signal; and (ii) verification of the scalability of the features and thresholds learned using the TIMIT database.

B. Results and discussion

Table I: Performance comparison of the proposed algorithm (PA) with the state-of-the-art algorithms. The figures listed are the percentage of the number of times the estimated VOT is within the given temporal tolerances of the ground truth. The two values given for PA for the case of all TIMIT stops correspond to the cases of: (i) automatic detection of both the burst and voice onsets; and (ii) detection of only voice onset taking the burst onset from the ground truth, respectively.

Temporal tolerance	< 5 ms	< 10 ms	< 15 ms	< 20 ms	< 25 ms
TIMIT (all stops)	**PA - 61.6, 63.3**	**PA - 85.0 , 88.9**	**PA - 93.9, 95.3**	**PA - 96.9, 97.2**	**PA 98.0, 98.0**
	RS[5] - 50.3	RS -76.1	RS - 88.7	RS - 91.4	RS -93.9
TIMIT (word-initials)	**PA - 62.5**	**PA - 85.9**	**PA - 94.6**	**PA -97.3**	**PA- 98.4**
	RF[9] - 57.2	RF- 83.4	RF - 93.4	RF - 96.5	RF - NA
TIMIT (word-initial UV)	PA - 64.4	**PA - 87.4**	**PA - 95.1**	PA - 97.6	PA **-** 98.3
	SP[8] - 67.2	SP - 85.0	SP - 94.7	**SP - 98.1**	**SP - 99.0**
	Results for the detection of voice-onsets only				
CMU Arctic	PA - 80.1	PA - 91.0	PA - 93.4	PA - 95.1	PA - 96.14

Table I compares the results of the proposed algorithm (abbreviated as PA) with those of three recent algorithms viz., the method based on re-assignment spectra by Stouten and Van hamme

(RS)[5], the random-forest-based method by Lin and Wang (RF)[9] and structured-prediction-based method by Sonderegger and Keshet (SP)[8]. All of these studies report results on validation against the TIMIT database using the same validation criterion as described here. However only the present work and RS consider all the stops in TIMIT, while RF examines the word-initial voiced as well as unvoiced stops and SP validates only on word-initial unvoiced stops. Our results are evaluated separately for each category of stops for a fair comparison. The performance of the proposed method exceeds that of the RS by 4 to12% for different tolerances. For each tolerance, the second entry in the first row indicates the results of PA, when the burst onset for each stop is assumed to be known (taken to be the hand-labeled boundary) and only voice-onset detection is validated. It is seen that, on an average, there is an improvement of 2% for lower tolerances when the burst onsets are assumed to be known. The second row of Table I compares PA and RF on word-initial stops in the TIMIT database. It is observed that PA performs better than the RF for all tolerances. The results of SP are compared with those of PA in the third row of Table I. SP reports accuracies of 67 and 98% at 5 and 20 ms, respectively, while PA offers 64 and 97.6%. However the performance of PA exceeds that of SP for 10 and 15 ms tolerances. If the feature ZCD is omitted from the algorithm, the percentage of times the estimated values are within 5 ms of the ground truth on all TIMIT stops reduces from 61 to 54.

On the CMU Arctic databases, the performance of the PA algorithm appears significant in that about 76% and 80% of the time, the detected voice onset lies within 2 and 5 ms of the ground truth, respectively. This suggests that the features, thresholds and thus proposed algorithm are scalable. Also the lower performance of the PA (and also of other algorithms) on the TIMIT database may be due to the use of human transcription for validation which may not be as accurate as the ground truth generated automatically from the EGG signal.

From the above comparison, the advantages of the proposed method over the state-of-the-art can be listed as follows: (i) The proposed algorithm requires no a-priori transcription unlike the other algorithms; (ii) It employs only two temporal measures derived out of acoustic phonetic

observations with a simple rule based classification, compared to high-dimensional feature vectors (e.g., 56 dimensions in RF, 63 feature maps in SP) and trained classifiers (e.g., random forest in RF, discriminative large margin classifier in SP). In spite of this, the performance of the proposed algorithm compares well with the state-of-the-art; (iii) The thresholds are determined using only 50 voiced-phone tokens here, whereas RF uses all the utterances in the TIMIT training database for training forced alignment HMMs and 40 utterances to train the RF classifier (SP uses 250 examples for training); and (iv) The number of tokens used for validation in this study is the highest (18,885 from TIMIT).

IV. CONCLUSION AND FUTURE RESEARCH

In this paper we presented a simple acoustic-phonetic method for estimating the VOT of stop consonants from speech without any transcription. This method makes use of two temporal measures based on the acoustic-phonetic characteristics of stops and voiced phones along with the epochal information. Experiments on two large corpora demonstrated that the algorithm is accurate and comparable to the state-of-the-art. Our future work will be directed towards detection and estimation of negative VOT for syllable-medial voiced stops and the usefulness of VOT for performing stop consonant classification tasks.

Acknowledgments

The authors thank Dr. Morgan Sonderegger and Dr. Hugo Van hamme for providing data required for experimentation and comparison.

REFERENCES AND LINKS

[1] K. N. Stevens, *Acoustic phonetics* (Cambridge, MA:MIT Press, 1998), chap. 1-8, 1–485.

[2] L. Lisker and A. Abramson, "A cross-language study of voicing in initial stops: acoustical measurements," Word **20**, 384–422 (1964).

[3] J. Jiang, M. Chen, and A. Alwan, "On the perception of voicing in syllable-initial plosives in noise," J. Acoust. Soc. Am. **119**, 1092–1105 (2006).

[4] J. H. Hansen, S. S. Gray, and W. Kim, "Automatic voice onset time detection for unvoiced stops (/p/, /t/, /k/) with application to accent classification," Speech Communication **52**, 777–789 (2010).

[5] V. Stoute and H. V. hamme, "Automatic voice onset time estimation from reassignment spectra," Speech Communication **51**, 1194–1205 (2009).

[6] P. Niyogi and P. Ramesh, "The voicing feature for stop consonants: Recognition experiments with continuously spoken alphabets," Speech Communication **41**, 349–367 (2003).

[7] P. Auzou, C. Ozsancak, R. J. Morris, M. Jan, F. Eustache, and D. Hannequin, "Voice onset time in aphasia, apraxia of speech and dysarthria: a review," Clinical linguistics & phonetics **14**, 131–150 (2000).

[8] M. Sonderegger and J. Keshet, "Automatic measurement of voice onset time using discriminative structured prediction," J. Acoust. Soc. Am. **132**, 3965–3979 (2012).

[9] C.-Y. Lin and H.-C. Wang, "Automatic estimation of voice onset time for word-initial stops by applying random forest to onset detection," J. Acoust. Soc. Am. **130**, 514–525 (2011).

[10] T. V. Ananthapadmanabha, A. P. Prathosh, and A. G. Ramakrishnan, "Detection of the closure-burst transitions of stops and affricates in continuous speech using the plosion index," J. Acoust. Soc. Am. **135**, 460–471 (2014).

[11] A. L. Francis, V. Ciocca, and J. M. C. Yu, "Accuracy and variability of acoustic measures of voicing onset," J. Acoust. Soc. Am. **113**, 1025–1031 (2003).

[12] J. L. Flanagan, *Speech Analysis Synthesis and Perception* (Springer,Newyork, 1972).

[13] B. Yegnanarayana and S. Gangashetty, "Epoch-based analysis of speech signals," Sadhana **36, part 5**, 651–697 (Oct. 2011).

[14] B. Yegnanarayana and R. Veldhuis, "Extraction of vocal-tract system characteristics from speech signals," IEEE Trans. on Speech and Audio Proc., **6**, 313–327 (1998).

[15] A. P. Prathosh, T. V. Ananthapadmanabha, and A. G. Ramakrishnan, "Epoch extraction based on integrated linear prediction residual using plosion index," IEEE Trans. on Audio, Speech, and Lang. Process. **21, no-12**, 2471–2480 (Dec. 2013).

[16] T. V. Ananthapadmanabha, "Acoustic factors determining perceived voice quality," *Vocal fold Physiology - Voice quality control*, edited by O.Fujimura and M. Hirano (Singular publishing group, San Diego,Cal., 1995), chap. 7, 113–126.

[17] D. G. Childers and A. K. Krishnamurthy, "A critical review of electroglottography," CRC Crit. Rev. Bioeng. **12**, 131–164 (1985).

[18] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgrena, *DARPA-TIMIT, Acoustic-phonetic continuous speech corpus.*, US Department of Commerce, Washington, DC (1993). (NISTIR Publication No.4930).

[19] "cmu corpora," URL `festvox.org/cmu_arctic/`, (date last viewed 20/3/14).

Collected figure captions

FIG. 1. Illustration of the utility of MWIP (solid line) and ZCD (dotted line) as features for voice onset detection from a segment of speech from the TIMIT database (a velar stop followed by a voiced sonorant). While MWIP is high over both the aspiration interval and the sonorant, the ZCD (value scaled down by a factor of 20 for ease of visual comparison) is high over the aspiration interval and low for the sonorant.

FIG. 2. Illustration of the burst and voice onsets detected by the algorithm on a segment of speech from the CMU Arctic database (KED). The acoustic waveform is shown by the solid line and the dEGG signal by the dotted line. Upward and downward arrows denote the estimates of the burst and voice onsets, respectively. In both cases, solid and dot-dash arrows represent the initial and final estimates, respectively. The initial and refined estimates of the voice onset coincide in this case. It is seen that the detected voice onset coincides with the first negative peak in the dEGG.