Determination of Instants of Significant Excitation in Speech Using Group Delay Function

Roel Smits and B. Yegnanarayana, *Senior Member, IEEE*

Abstract—**A new method for determining the instants of significant excitation in speech signals is proposed. Here, significant excitation refers primarily to the instant of glottal closure within a pitch period in voiced speech. The method is based on the global phase characteristics of minimum phase signals. The average slope of the unwrapped phase of the short-time Fourier transform of linear prediction residual is calculated as a function of time. Instants where the phase slope function makes a positive zero-crossing are identified as significant excitations. The method is discussed in a source-filter context of speech production. The method is not sensitive to the characteristics of the filter. The influence of the type, length, and position of the analysis window is discussed. The method works well for all types of voiced speech in male as well as female speech but, in all cases, under noise-free conditions only.**

I. Introduction

VOICED speech is produced by excitation of the vocal tract system with the quasiperiodic vibrations of the vocal folds at the glottis. The vibrations are reflected as the opening and closing of the glottis within each pitch period. The major excitation of the vocal tract system within a pitch period takes place at the instant of glottal closure. We call these instants significant instants. In this paper, we propose a method of determining these instants of significant excitation automatically from a speech signal using the negative derivative of the unwrapped phase (group delay) function of the short-time Fourier transform of the signal [1]. Throughout the paper, we refer to the unwrapped phase function as the phase spectrum.

Many speech analysis situations depend on the accurate estimation of the instant of glottal closure within a pitch period. For example, if such instances are known, the closed glottis region can be identified, and the vocal tract parameters such as formants may be derived accurately by confining the analysis to only those regions [2]. It is also possible to determine the characteristics of the voice source by a careful analysis of the signal, starting with this information [3]. In applications such as text-to-speech (TTS) conversion, especially using methods like PSOLA [4], currently, a lot of manual effort is involved in marking the pitch excitation points since the methods critically depend on the accuracy of locations of the pitch markers. Therefore, determination of these instants reduces this effort considerably.

It is difficult to isolate the major excitation within a pitch period. Usually, there may be several excitations within a period, and many of them may be significant [3], [5]. In fact, at every instant, there is some excitation, although in normal steady vowels, the instant of glottal closure corresponds to the instant of major excitation. In weak voicing, it is difficult even to define the instant of excitation, let alone determine it. Still, it is useful in many cases to assume that the major excitation is at the glottal closure. Note that there will also be a major excitation at the instant of release of a stop burst. All such major excitation instants are included in the category of significant instants in this paper.

As a first approximation, starting from the significant instant, the excitation signal (second derivative of the glottal pulse or glottal volume velocity) within a pitch period can be assumed to be a minimum phase signal [6]. Therefore, one can use the properties of minimum phase signals to derive the instants of significant excitation, provided the excitation signal is available. However, what is available is the speech signal, which is the result of excitation of the vocal tract system. Although the impulse response of the vocal tract system, including the nasal tract, is a minimum phase signal, the overlapping quasiperiodic impulse responses makes the speech signal a nonminimum phase signal in general. The response exactly within a period is still a minimum phase signal, but it is difficult to isolate a period starting from the significant instant of excitation. The difficulty in determining these instants is compounded by the fact that only a finite data window has to be used for analysis of the signal.

Several methods have been proposed for determining the instant of glottal closure [2], [3], [7], [8]. Almost all of them use some kind of block processing to determine the energy of the residual excitation signal in a small interval. The point where the computed energy is maximum is marked as the instant of significant excitation. While these methods work well in most cases, the block processing leaves some uncertainty as to the precise location of the instant of excitation [3], [7], [8].

In this paper, we present a method for determining the instants of significant excitation using the properties of minimum phase signals and group delay functions [1]. In Section II,

Manuscript received September 9, 1992; revised January 18, 1995. The associate editor coordinating the review of this paper and approving it for publication was Dr. Amro El-Jaroudi.

R. Smits was with the Institute for Perception Research (IPO), Eindhoven, The Netherlands. He is now with the Department of Phonetics and Linguistics, University College London, London, England.

B. Yegnanarayana is with the Department of Computer Science and Engineering, Indian Institute of Technology Madras (IITM), Madras, India.

IEEE Log Number 9413733.

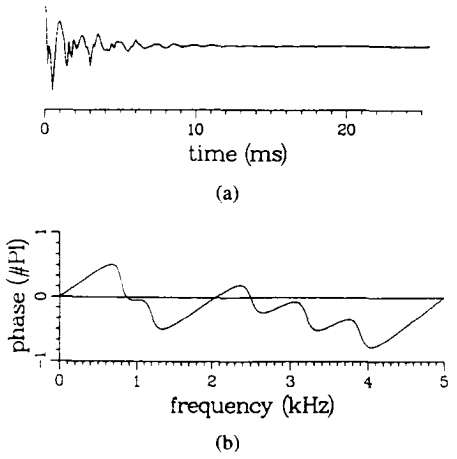


Fig. 1. (a) Impulse response of a minimum phase (five resonator) system and (b) its phase spectrum.

we discuss the basis for the proposed method. Development of the algorithm for determining the significant instants of excitation is described in Section III. The method is illustrated for several cases of synthetic signals in Section IV. In Section V, we consider several examples of natural speech data to demonstrate the applicability of the method. Finally, in Section VI, we briefly discuss potential applications of these results in speech analysis.

II. Basis for the Proposed Method

In this section, we present some properties of discrete-time minimum-phase signals and discuss some issues in processing these signals.

A. Properties of Minimum Phase Signals

A minimum phase signal is a signal for which the time of building up of energy is least among all signals possessing the same spectral amplitude characteristics and the same total energy [9], [10]. The system that generates the minimum phase signal as its impulse response has all the roots (zeros and poles) of its transfer function lying within the unit circle in the z-plane. Let us consider some issues in processing these signals.

Consider the impulse response of a stable all-pole filter (Fig. 1(a)). This is a minimum phase signal. One way to check the minimum phase property is to compute the average slope of the unwrapped phase of the Fourier transform (i.e., phase spectrum) of the signal (see Fig. 1(b)). For a minimum phase signal, the average slope of the phase spectrum is zero [6]. The shifted version of a minimum phase signal will have a phase spectrum similar to the original but with an average slope proportional to the shift. This means that the average slope of the phase spectrum is dictated by the location of the excitation impulse. Note that the system characteristics are manifested as fluctuations in the phase spectrum, whereas the average slope of the phase spectrum is dictated by the instant of excitation relative to the time origin.

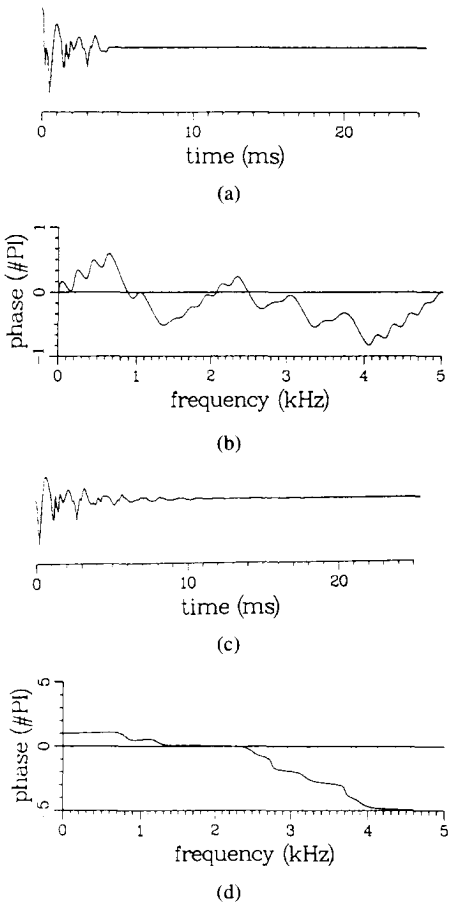


Fig. 2. Effects of truncation on the phase characteristics of the minimum phase signal of Fig. 1: (a) Right truncated signal of Fig. 1(a); (b) phase spectrum of the signal in (a); (c) left truncated signal of Fig. 1(a); (d) phase spectrum of the signal in (c).

B. Effects of Location of Analysis Window

The effects of left and right truncation due to the analysis window on the phase characteristics of a minimum phase signal are illustrated in Fig. 2. Since a minimum phase signal of an infinite impulse response system eventually decays with time, the right truncation (see Fig. 2(a) and (b)) does not affect the average slope of the phase function, except when the truncation is severe, leaving only a few samples of the signal. Of course, the truncation introduces fluctuations (Fig. 2(b)) in the phase spectrum depending on the instant at which the signal is truncated. Another way of interpreting the behavior of the phase spectrum is that if the excitation instants are included in the analysis window, the signal preserves the average phase characteristics of the excitation impulses, as long as the truncation does not introduce large discontinuities.

On the other hand, if the minimum phase signal is truncated on the left (see Fig. 2(c) and (d)), the overall phase characteristics as well as the average slope of the phase spectrum are severely affected. Note that in this case, the truncation is likely to remove the instants of excitation. The average slope of the phase spectrum is not zero any more, and the signal is not a minimum phase signal. The average slope characteristics of the phase spectrum are dictated by the truncating window rather than by the excitation.

C. Effects of Size of the Analysis Window

In this section, we describe the effects of applying rectangular windows of various sizes to periodic signals. Consider an excitation signal consisting of a sequence of periodic impulses, with two impulses in each period, as shown in Fig. 3(a). As long as the analysis window length is less than a pitch period, and if the large peak of a period is enclosed in the analysis interval, the signal behaves like a minimum-phase signal delayed by the distance of the large peak to the $t = 0$ point. Therefore, the average slope of the phase spectrum is proportional to this delay. Moreover, if the $t = 0$ instant coincides with the largest peak, the average slope of the phase spectrum is zero. If, on the other hand, the analysis interval is moved beyond the main peak (see the window in Fig. 3(a)) and if the next main peak is not within the interval, then the average slope is dictated by the minor peak. This is shown in Fig. 3(b) and (c) for a windowed portion of the signal in Fig. 3(a) and the corresponding phase spectrum. For computation of the phase spectrum, the center of the analysis window is taken as the $t = 0$ instant.

The response of an all-pole system to the periodic sequence is shown in Fig. 3(d). The windowed signal and its phase spectrum are shown in Fig. 3(e) and (f), respectively. The average slope in Fig. 3(f) clearly differs from the average slope of the phase spectrum in Fig. 3(c). This shows that the characteristics of the phase spectrum are influenced by the size as well as the position of the analysis window.

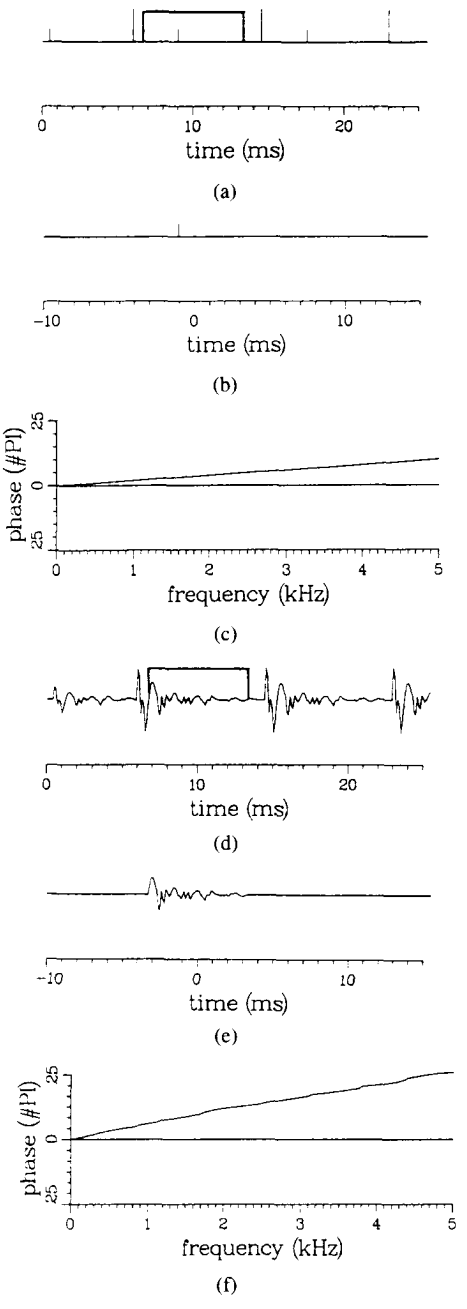


Fig. 3. Rectangular window of 6.5 ms, centered at $t = 10$ ms, is used: (a) Periodic impulses sequence; (b) windowed segment of (a) with the center of window at $t = 0$; (c) phase spectrum of the signal in (b); (d) response of an all-pole filter for the periodic sequence in (a); (e) windowed segment of (d) with the center of the window at $t = 0$; (f) phase spectrum of signal in (e).

III. Method for Determining the Instants of Significant Excitation

In this section, we describe the development of an algorithm for determining the instants of significant excitation from the average slope of the phase spectrum of the signal. First, we discuss the computation of the average slope of the phase spectrum, and then, we discuss a method to extract the desired instants from the phase slope values as a function of time.

A. Computation of Average Slope of the Phase Spectrum

The main step in the identification of the instant of significant excitation is the computation of average slope of the unwrapped phase spectrum. Direct computation of the phase through the real and imaginary parts of the DFT of a signal results in phase values that are confined to the range $-\pi$ to $+\pi$. In other words, the phase values are said to be wrapped around the limits. To compute the average slope, it is necessary to determine the unwrapped phase. Accuracy of computation of the unwrapped phase depends on the (windowed) signal and the resulting phase values. If there are too many zeros of the z-transform of the signal around the unit circle in the z-plane, there will be too many phase jumps in the resulting phase function. The degree (number of 2π's) of wrapping in between two adjacent discrete frequency points around a zero depends on the closeness of the zero to the unit circle. Any sharp discontinuity at the ends of the analysis interval could result in periodic zeros on the unit circle in the z-plane. To overcome this problem, a window function that does not significantly alter the phase characteristics of the signal but tapers the signal to zero at the ends is desirable. A Hanning window is used in the remaining part of this study, although it may be possible to have a window function more suitable for this purpose. The change in the phase characteristics of an analysis segment due to Hanning window is illustrated in Fig. 4. Fig. 4(a) is the same signal as in Fig. 3(d), except that the signal is multiplied with a Hanning window function. The average slope of the phase spectrum (Fig. 4(b)) in this case is equal to the average slope of the phase spectrum for the corresponding windowed excitation signal as shown in Fig. 3(c). Therefore, in contrast with the rectangular window case shown in Fig. 3(f), the

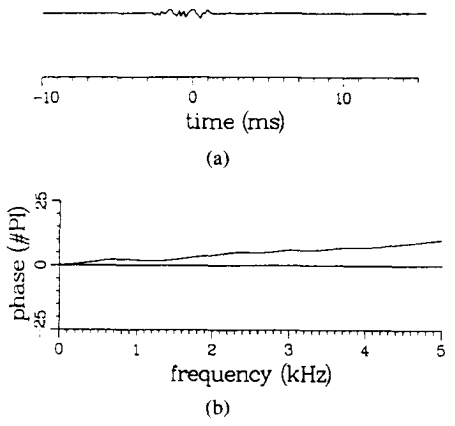


Fig. 4. (a) Signal of Fig. 3(e) multiplied by a Hanning window, and (b) phase spectrum of the signal in (a). The average slope is equal to the average slope in Fig. 3(c). Therefore, it is now dictated by the excitation rather than the window, which was the case in Fig. 3(f).

Hanning window does not influence the average slope of the phase spectrum significantly (see Fig. 3(c), 3(f), and 4(b)).

For phase unwrapping, it is difficult *a priori* to determine how closely the interval between the points in the frequency domain need to be taken in the computation of the DFT of the signal. Since we are interested only in the average slope of the phase spectrum, it is preferable to compute the slope by using a DFT size significantly larger (at least twice) than the size of the analysis window. There are several approaches to obtain the average slope of the phase spectrum, including computation directly from the expression for the group delay function [10]. The approaches differ in the manner the effects of phase wrapping are handled. Phase wrapping produces discontinuities in the group delay spectrum computed by simply taking the differences between the successive wrapped phase values. We have used a direct approach to compute the unwrapped phase spectrum. The unwrapping is done by adjusting (in integral multiples of 2π) the group delay values at each frequency point to make sure that the magnitude of the difference between the adjusted value at this point and the group delay value at the preceding point is minimum. Once the optimal value for adjustment, i.e., $2k\pi$, is determined for a given frequency point, the phase values for all the frequency points from the current point until the end of the phase spectrum are modified by this value. Then, the next frequency point is considered for unwrapping the phase value, and so on. A straight line fit for the resulting phase spectrum is obtained using linear regression. The slope of the line is taken as the average slope of the phase spectrum, i.e., the average group delay function.

The average slope of the phase spectrum is computed at each sample point as a function of time. This is called phase slope function. The analysis window is placed symmetric with respect to $n = 0$ point. Therefore, the phase slope function will be zero when the center of the window coincides with the start of a minimum phase signal.

B. Determination of Zero-Crossing Instants from the Phase Slope Function

Due to discrete nature of computations in obtaining the phase slope function at each sampling instant as well as to effects of finite window size and shape, the phase slope function for real speech data will show many fluctuations that may sometimes make it difficult to identify the positive zero-crossing instants uniquely. Fig. 5(a) and (b) show a segment of voiced speech and its linear prediction (10th order) residual. Fig. 5(c) shows the phase slope function computed for the residual signal shown in Fig. 5(b). The reasons for the choice of the residual signal, instead of the speech signal directly, is explained in Section V.

The first step is to collect all the positive zero-crossing instants from the phase slope function (Fig. 5(d)). To select the ones corresponding to the significant instants, the phase slope function is smoothed using a 13-point Hanning window. Note that this smoothing window size is not very critical, as long as it removes some fine fluctuations. The positive zero-crossing instant corresponding to the major excitation instant in the analysis window is not affected by this smoothing operation. The positive zero-crossing instants from the smoothed phase slope function (Fig. 5(e)) are shown in Fig. 5(f).

For an ideal situation, with an impulse at the excitation instant, the phase slope function will be a straight line with a slope corresponding to the unit sampling delay and the positive zero-crossing instant corresponding to the delay of the excitation instant from the time origin in the analysis window. At each selected instant, we can verify how well the local phase slope function matches the ideal one, i.e., the straight line with ideal slope passing through the instant. This is done by determining the number of points of the phase slope function falling within a range from the ideal slope function line positioned at each positive zero-crossing instant. The range is the region covered by two lines, called threshold lines, with slopes that are the same as the ideal one but passing through two instants that are equidistant on either side of the current zero-crossing instant. The distance on either side was chosen as eight sampling intervals in our method. Each point of the phase slope function falling in this region is given a weightage that decreases in proportion to its horizontal distance from the ideal slope function line, with maximum weight of 1 for points on the ideal slope function line and a weight of 0 for points on the threshold lines. The weighted sum of all the points in the selected range is computed for each zero-crossing point, and it is plotted as confidence levels in Fig. 5(g). Note that the choice of the range (in sampling intervals) for the threshold and the weightage are not critical for our method, as this plot is used only to assign some level of confidence for the identified positive zero-crossing instants in Fig. 5(f). This is useful, especially to distinguish instants due to significant excitation from those due to random noise in the input. Typically, if there is a point due to noise, the corresponding confidence level will be lower.

We have also computed the gain at each of the zero-crossing instants by computing the square root of the average energy per sample in the LP residual for an interval between two successive zero-crossing instants. The interval is centered around the zero-crossing point under consideration. The resulting gain values are shown in Fig. 5(h). These values may be viewed as the strengths of the impulses at the selected instants.

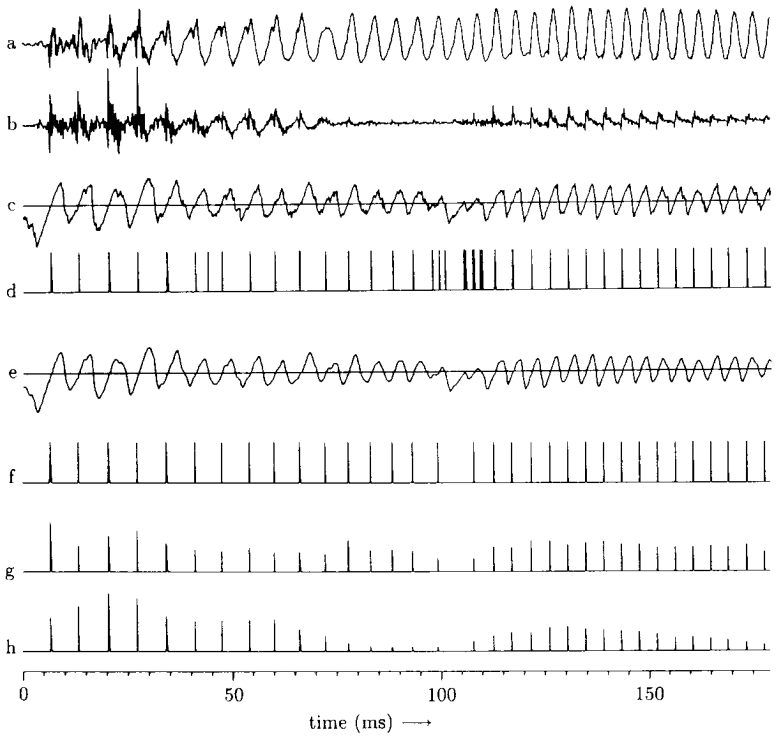


Fig. 5. (a) Segment of voiced speech signal; (b) linear prediction residual of (a); (c) phase slope function for the LP residual signal in (b); (d) unit impulse sequence with impulses located at positive zero-crossing instants of the plot in (c); (e) smoothed phase slope function; (f) unit impulse sequence with impulses located at positive zero-crossing instants of the plot in (e); (g) confidence level plot; (h) gain plot indicating the strengths of the impulses at the positive zero-crossing instants of (e).

The complete procedure for determining the instants of significant instants of excitation is summarized in Fig. 6. In the following sections, the performance of the method is evaluated for several synthetic and natural speech signals.

IV. Illustration of the Method for Synthetic Signals

The phase slope function is computed for several cases of synthetic signals as shown in Figs. 7–12. Examples are given for two different window sizes in each case. Each figure contains three plots:

a) the original time signal,
b) the phase slope function for an analysis window of size 6.4 ms,
c) the phase slope function for an analysis window of size 12.8 ms.

Fig. 7 is for a periodic impulse sequence, Fig. 8 is for a periodic impulses sequence, and Fig. 9 is for the output of an all-pole model excited by the periodic impulses sequence. In Figs. 7–9, we can see that the positive zero-crossing instants correspond to the instants of major excitation within the chosen analysis window. In Fig. 7, a large portion of the phase slope function is linear around the instant of major excitation. This linear part is the result of dominance of one excitation in its surrounding region.

If a signal containing minor excitations is analyzed using a large size window (Figs. 8 and 9), the phase characteristics due to major excitation dominate the phase slope function, and the minor excitations do not produce any positive zero crossings. The presence of these minor excitations may sometimes make it difficult to extract the instants of major excitations. This is one of the reasons why we have proposed in Section III-B some post-processing of the raw phase slope function in which the goodness of the local fit of the function to the expected ideal slope function is determined.

Figs. 7–9 also show that the phase slope function is mostly dictated by the excitation signal. It is interesting to note that neither the minimum phase system nor the location and size of the analysis window has influenced the decision on the excitation instants obtained from the phase slope function. Even for a window size greater than a pitch period, the use of a Hanning window reduces the effects of the surrounding peaks on the resulting excitation instants. Thus, within a window, only one excitation impulse is likely to dominate.

For random noise (Fig. 10), the features in the phase slope function are different for different window sizes. It is interesting to note that whenever there is any major excitation in the noise signal, it will again clearly show up, irrespective of the size of the analysis window. For any noise signal, these excitation instants will be distributed randomly in time. That is why the method will not work well for noisy speech, as the excitations due to noise will show up randomly in between the instants of glottal closure. Fig. 11 shows the behavior of phase slope function for two different window sizes for a signal generated by exciting an all-pole filter with random noise.

For sinusoids (Fig. 12), there are no isolated major points of excitation, and the phase function does not show the characteristic linear part around the zero crossing. Again, the

Algorithm:

Determination of major instants of excitation in speech

1. Calculate the LP residual using a frame of size 25ms, Hanning window, 10th order LP analysis by autocorrelation method, and a frame rate of 100 frames per second.
2. Calculate the short-time Fourier transform of the residual at each sampling instant, with the frames having a size of approximately one to two times the average pitch period. The signal is multiplied with a Hanning window and the center of the window is placed at $n = 0$ point in the DFT computation. Pad the windowed signal with zeros until it has a length of at least twice the frame size.
3. Calculate the phase and the difference phase (group delay) values. Unwrap the FT phase, by adjusting each group delay value at each frequency point in multiples of 2π, so that the adjacent group delay values are continuous, i.e., differ by a minimal amount. The optimum adjustment parameter ($2k\pi$) at each point is used to modify the phase values of the current and all the succeeding points in the spectrum, before taking up the next frequency point for unwrapping the phase.
4. Obtain a best fit straight line through the unwrapped phase spectrum by linear regression.
5. Using the slope of the straight line, obtain the phase slope as a function of time.
6. Smooth the phase slope function using a Hanning window of size 13 samples.
7. Find the instants of positive zero-crossings in the smoothed phase slope function.
8. Calculate the "level of confidence" for these zero-crossing instants, by comparing with an ideal slope function passing through these points. Take the region enclosed by ideal slope lines placed at 8 sampling points to the left and 8 on the right of the zero-crossing point. Each point on the phase slope function in this region is given a weightage. The weightage is 1 if the point is on the ideal slope line passing through the zero-crossing point, and 0 if it is on either of the threshold lines enclosing the region. All other points in the region are given a weightage which decreases linearly with its horizontal distance from the ideal slope line passing through the zero-crossing point. The sum of the weights of all the points in the region is used as the confidence level for the zero-crossing instant to represent an excitation instant. The confidence levels for all the zero-crossing instants give the confidence level plot.
9. Calculate the "gain plot", which indicates the gain, or strength, of an impulse at the excitation instant. The gain at each excitation instant is defined as the squareroot of the average energy per sample in the LP residual for an interval between two successive zero-crossing instants. The interval is centered around the zero-crossing point under consideration.

Fig. 6. Algorithm for determination of major instants of excitation.

features of the phase slope function are different for different window sizes, and the effects of windowing show up clearly.

V. Illustration of the Method for Natural Speech Data

A. Determination of Instants of Significant Excitation

In this section, we discuss the performance of the proposed method on natural speech data. In all the illustrations to follow, the speech signals were sampled at 10 kHz. In order to minimize the effects of the position of the analysis window with respect to the impulse response, it is better to obtain at least an approximation to the excitation signal before computing the average slope of the phase spectrum. Linear prediction residual [11] is a good approximation to the excitation signal, as the correlation between adjacent samples is significantly reduced from what it is in the original signal. Note that since inverse filtering in linear prediction analysis [11] is, in effect, passing the speech signal through a minimum phase system, the phase slope characteristics of the excitation will not be altered in the residual. For the computation of the residual, a 10th-order LPC was used in this study, although the order is not very critical for this analysis.

Since, for speech data, there will be several points of excitation even within a pitch period, the phase slope function will have many fluctuations. In order to determine the instants of significant excitation, the points of positive zero crossing of the phase slope function are obtained by post-processing the function as discussed in Section III-B. The accuracy of estimation of the significant excitation instants by the proposed algorithm is verified by comparing these instants with the instants of glottal closure derived from the electroglottograph (EGG) signal. Fig. 13(a) and (b) show a segment of voiced speech signal and the corresponding LP residual signal, respectively. The derivative of the EGG signal is given in Fig. 13(c). The valleys in Fig. 13(c) correspond to the instants of glottal closure in each pitch period. The significant excitation instants are identified by applying the proposed algorithm on the LP residual (Fig. 13(b)) using a window of 6.4 ms, and the result is shown in Fig. 13(d). Comparing Fig. 13(d) and (c), it can be seen that the significant instants of excitation obtained using the proposed algorithm correspond to the instants of glottal closure. This shows that the algorithm indeed picks up the significant instants of excitation within each pitch period.

Once the major excitations are identified, it is possible to explore for the presence of other excitation instants by

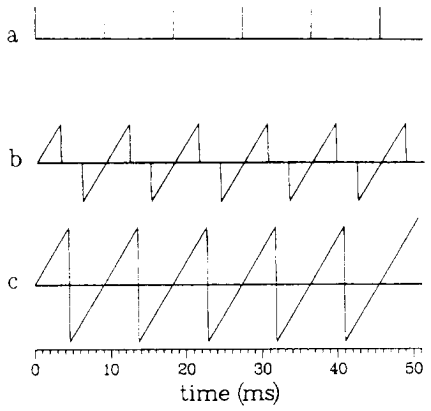


Fig. 7. Phase slope function for a periodic impulse sequence; (a) Periodic impulse sequence; (b) phase slope function for a window size of 6.4 ms; (c) phase slope function for a window size of 12.8 ms.

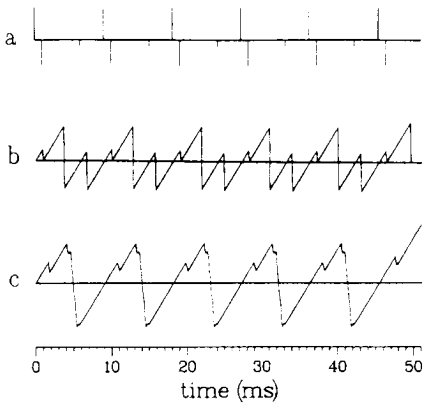


Fig. 8. Phase slope function for a periodic impulses sequence: (a) Periodic impulses sequence; (b) phase slope function for a window size of 6.4 ms; (c) phase slope function for a window size of 12.8 ms.

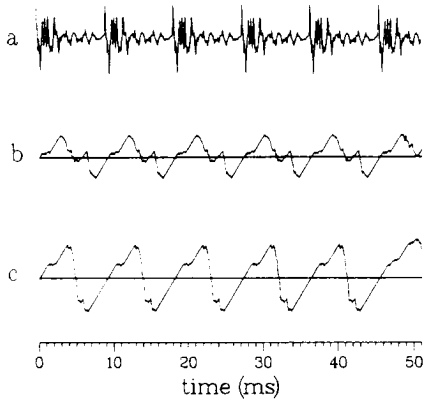


Fig. 9. Phase slope function for the output of an all-pole model excited by the periodic impulses sequence of Fig. 8(a): (a) Output of an all-pole model excited by a periodic impulses sequence; (b) phase slope function for a window size of 6.4 ms; (c) phase slope function for a window size of 12.8 ms.

computing the phase slope function on the residual using a smaller window for analysis, which is typically half the size of the original window. Thus, the method, in principle, enables us to determine other significant instants in the excitation. For identifying a single instant in each pitch period, it is preferable to have an analysis window size in the range of one to two pitch periods.

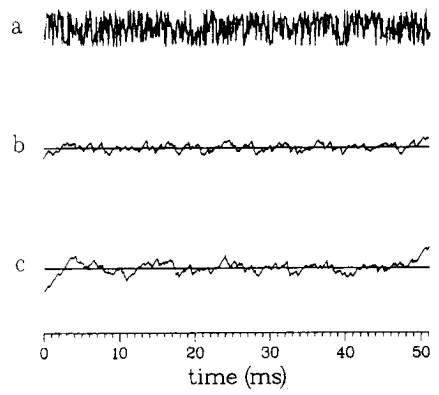


Fig. 10. Phase slope function for a random noise sequence: (a) Random noise sequence; (b) phase slope function for a window size of 6.4 ms; (c) phase slope function for a window size of 12.8 ms.

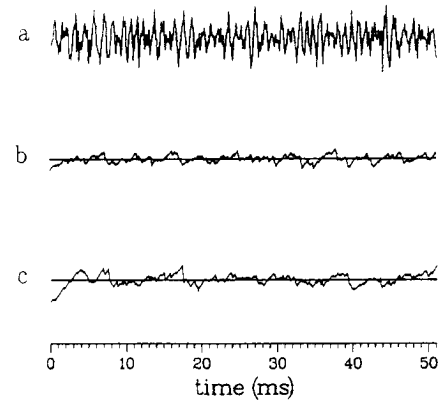


Fig. 11. Phase slope function for an all-pole model excited by a random noise sequence: (a) Output of an all-pole model excited by a random noise sequence; (b) phase slope function for a window size of 6.4 ms; (c) phase slope function for a window size of 12.8 ms.

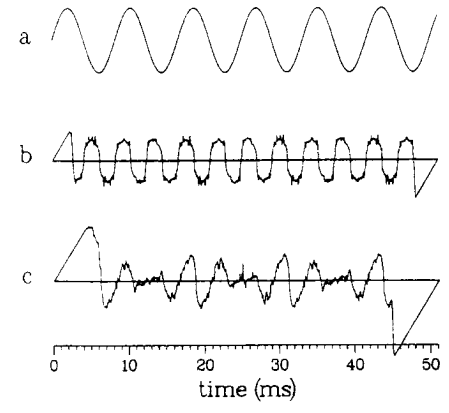


Fig. 12. Phase slope function for a sinusoidal sequence: (a) Sinusoidal sequence; (b) phase slope function for a window size of 6.4 ms; (c) phase slope function for a window size of 12.8 ms.

B. Analysis of Continuous Speech

Fig. 14 illustrates the result of our method the initial part of the utterance "ANY DICTIONARy will give at least . . .," spoken by a male voice. This utterance consists of a variety of segments like voiced, unvoiced, nasal, transition, stop, fricative, etc. The length of the analysis window was 10 ms, and the average pitch period was 8 ms. Fig. 14(a) and

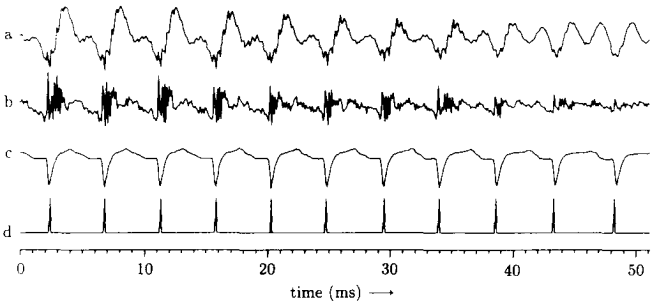


Fig. 13. Comparison of the results of extraction of significant excitation instants with the information from the derivative of EGG signals: (a) Segment of voiced speech signal; (b) linear prediction residual for (a); (c) derivative of the EGG signal; (d) instants of significant excitation from the proposed algorithm.

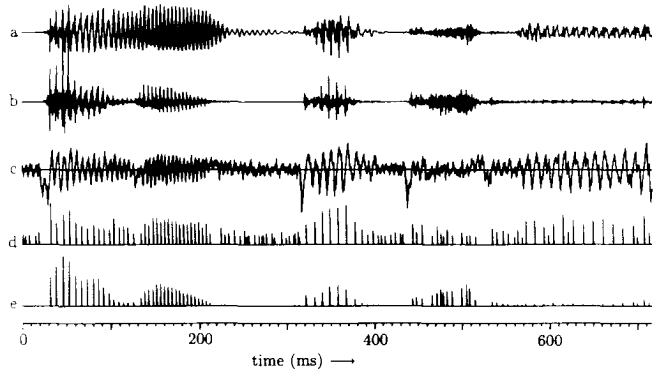


Fig. 14. Illustration of the method for a portion of the utterance for the sentence "ANY DICTIONAry will give ..." by a male speaker: (a) Speech signal; (b) linear prediction residual; (c) phase slope function for the linear prediction residual signal; (d) confidence plot for the positive zero-crossing instants of the smoothed phase slope plot; (e) gain plot showing the strengths of the impulses at the significant instants.

(b) show the speech waveform and the LP residual signal, respectively. Fig. 14(c)–(e) give the phase slope function, the confidence level, and the gain plots, respectively, for the residual signal of Fig. 14(b). The voiced segments have clear quasiperiodic instants of excitation. The unvoiced and silent parts show random instants of excitation. The instants (Fig. 14(d)) identified by the phase slope function indeed correspond to the instants of significant excitation, as can be seen in comparison with the residual signal (Fig. 14(b)).

Let us look at the individual signal categories in some detail. In silence and unvoiced fricative regions, almost no significant excitation was identified. However, whenever there is a transition from one category to another, like at a burst release, the transition point is identified as a significant instant. In many weakly voiced regions, there will not be any significant excitation, as is evidenced by the residual signal. The same is reflected in the phase slope function, although the speech waveform shows low-frequency periodicity. Absence of significant excitation instants in this case can also be verified by observing the amplitude spectra for these regions. Typically, the spectra do not show any formant structure in these cases but only show some energy at the pitch frequency.

Since the technique uses the phase characteristics of the excitation signal, the vocal tract system has very little influence on the proposed method of determining the instants of excitation. That is why it can be seen that the technique works well not only for steady vowels but for diphthongs, transitions, liquids, and nasals as well. Note again that the method shows significant quasiperiodic excitations even in cases where they are not clearly evident in the linear prediction residual.

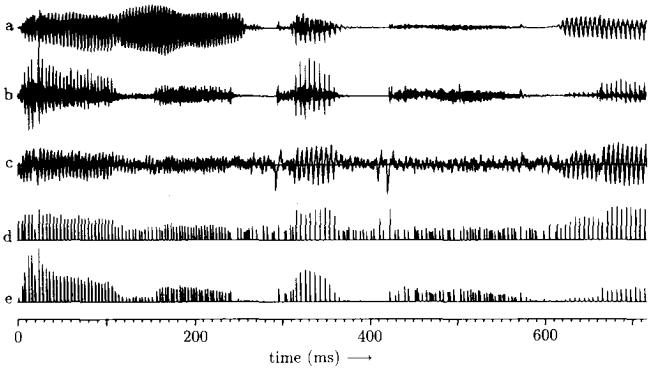


Fig. 15. Illustration of the method for a portion of the utterance for the sentence "ANY DICTIONAry will give ..." by a female speaker: (a) Speech signal; (b) linear prediction residual; (c) phase slope function for the linear prediction residual signal; (d) confidence plot for the positive zero-crossing instants of the smoothed phase slope plot; (e) gain plot showing the strengths of the impulses at the significant instants.

Fig. 15 shows the identification of instants of significant excitation for an utterance by a female voice for the same sentence. The different plots in the figure have the same description as for Fig. 14 for the male voice. The technique works well for the female voice for all categories of segments as well. Note that because of the smaller average pitch period, a smaller analysis window size of 6.4 ms was used.

VI. Discussion and Conclusions

In this paper, we have proposed a method to determine the instants of significant excitation using the average group delay characteristics of minimum phase signals. Here, the term significant for voiced speech refers to the instants of glottal closure in each pitch period. Since the method is based on the phase characteristics of the excitation signal, it is possible to derive the instants of significant excitation for all categories of speech segments. The method works well even for female voices because there is no influence of the vocal tract system on the phase slope characteristics of the signal. The method is also not very sensitive to the choice of analysis parameters, like the size of the window and the placement of the window relative to the significant excitations. It is interesting to note that by block processing, we have been able to mark an instant that does not depend critically on the size of the block and its placement.

This observation leads to several interesting applications of the results of this study. In particular, it is now possible to reliably estimate the pitch markers, which in turn may help in the estimation of the pitch period, V/UV decision, voice onset time, and probably of an accurate analysis of voice source. It is also possible to analyze speech in specific regions, like closed glottis intervals, to study the characteristics of the vocal

tract system more accurately than by the currently available methods.

ACKNOWLEDGMENT

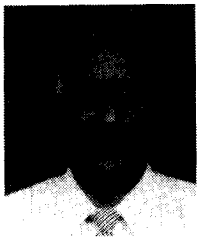
The authors thank G.-J. Plattèl, L. Willems, B. Eggen, and R. Collier for their help and suggestions in this work.

REFERENCES

[1] B. Yegnanarayana, "Significance of group delay functions in signal reconstruction from spectral magnitude or phase," *IEEE Trans Acoust., Speech, Signal Processing,* vol. ASSP-32, no. 4, pp. 610–623, 1984.
[2] A. K. Krishnamurthy, "Glottal source estimation using a sum-of-exponentials model," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 40, no. 3, pp. 682–686, 1992.
[3] T. V. Ananthapadmanabha and B. Yegnanarayana, "Epoch extraction from linear prediction residual for identification of closed glottis interval," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-27, no. 8, pp. 309–319, 1979.
[4] C. Hamon, E. Moulines, and F. Charpentier, "A diphone synthesis system based on time domain prosodic modifications of speech," in *Proc. Int. Conf. Acoust., Speech, Signal Processing,* Glasgow, 1989, pp. 238–241.
[5] C. Ma and L. F. Willems, "A SVD approach to glottal closure determination from the speech signal," Inst. for Perception Res., Rep. no. 761, 1990.
[6] E. A. Robinson and T. S. Durrani, *Geophysical Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1986, ch. 5.
[7] D. J. Wong, J. D. Markel, and A. H. Gray, "Least squares glottal inverse filtering from the acoustic speech wave," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-27, no. 8, pp. 350–355, 1979.
[8] Y. M. Cheng and D. O'Shaughnessy, "Automatic and reliable estimation of glottal source instant and period," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 37, no. 12, pp. 1805–1815, 1989.
[9] A. J. Berkhout, "Related properties of minimum-phase and zero phase time functions," *Geophys. Prospecting,* vol. 22, pp. 683–709, 1974.
[10] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1975, ch. 10.
[11] J. D. Markel and A. H. Gray, *Linear Prediction of Speech.* New York: Springer-Verlag, 1976.

Roel Smits received the Engineer degree in physics from the Eindhoven University of Technology, Eindhoven,, The Netherlands.

From 1990 to 1995, he worked on his doctoral study at the Institute for Perception Research on the analysis and perception of stop consonants. He received the doctoral degree in June 1995. At present, he is with the Department of Phonetics and Linguistics, University College London, UK. His research interests involve speech perception, speech processing, and modeling of human classification behavior.



B. Yegnanarayana (M'78–SM'84) was born in India on January 9, 1944. He received the B.E., M.E., and Ph.D. degrees in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1964, 1966, and 1974, respectively.

He was a Lecturer from 1966 to 1974 and an Assistant Professor from 1974 to 1978 in the Department of Electrical Communication Engineering, Indian Institute of Science. From 1977 to 1980, he was a Visiting Associate Professor of Computer Science at Carnegie-Mellon University, Pittsburgh, PA. He was a Visiting Scientist at ISRO Satellite Centre, Bangalore, from July to December 1980. Since 1980, he has been a Professor in the Department of Computer Science and Engineering, Indian Institute of Technology, Madras. He was a Visiting Professor at the Institute for Perception Research, Eindhoven Technical University, Eindhoven, The Netherlands, from July 1994 to January 1995. During the period of 1966 to 1971, he was engaged in the development of environmental test facilities for the Acoustics Laboratory at the Indian Institute of Science. Since 1972, he has been working on problems in the area of speech signal processing. He is presently engaged in research activities in digital signal processing, speech recognition, and neural networks.

Dr. Yegnanarayana is a member of the Computer Society of India and a fellow of the Institution of Electronics and Telecommunication Engineers of India.