CMU-CS-79-108

On Improving The Reliabilty Of Cepstral Pitch Estimation

B. Yegnanarayana and T. V. Ananthapadmanabha
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, PA 15213
February 1979

This research was sponsored by the Defence Advanced Research Projects Agency (DOD), ARPA Order No. 3597, and monitored by Air Force Laboratory under Contract F33615-78-C1151.

ABSTRACT

Identification of relatively high SNR regions in the short-time spectrum of a speech segment is very useful in speech processing applications. Such regions usually occur around the peaks in the spectral envelope. In this paper we propose a method for determining such regions automatically for a given speech segment. The method is based on a recently developed technique for pole-zero decomposition of speech spectra. It is shown that by selectively processing the high SNR regions of the spectrum, an unambiguous pitch peak in the high quefrency portion of the cepstrum can be obtained. The processing involves computation of Hilbert envelope of the selectively filtered cepstrum. Several examples of speech segments are considered to illustrate the improvement provided by the proposed method.

I. INTRODUCTION

Speech is the output of a time varying vocal tract system excited by a time varying excitation. Due to nonstationary nature of the speech signal, speech analysis is usually performed on short segments (10-40 msec) of speech. Signal to Noise ratio (SNR) of speech signal is different for different segments of the data. Further, for a given segment, the SNR is a function of frequency in the short-time spectrum. For additive white noise, it is reasonable to assume that SNR is relatively higher over the regions corresponding to peaks in the envelope of the short-time spectrum. Identification of such high SNR regions in the signal spectrum would be very useful in accurate analysis of speech, especially in obtaining a reliable estimate of voice pitch. So far there has been no convenient method available for automatically identifying such regions in the short-time spectrum. Recently it was shown that the derivative of phase spectrum (DPS) of the minimum phase correspondent of a given signal provides directly the information corresponding to peaks and valleys of the spectral envelope[1]. The objective of this paper is to show the use of such an information in improving the reliability of pitch estimation by cepstrum.

Pitch extracton is a classical problem in speech analysis which has been eluding speech researchers of a complete solution. Although there is only one intrinsic mechanism causing the quasi-periodic speech waveform, a large number of techniques exists for pitch extraction. In spite of the availability of several diverse techniques, there is still a need for a robust technique for extraction of pitch from noisy and filtered speech data. The problems in pitch detection and the comparative performance of several pitch algorithms are discussed by Rabiner et al [2]. One general requirement of any pitch extraction method is to remove or at least reduce the influence of the formant frequencies from the speech signal. The first formant, in particular, usually has a considerable influence on the accuracy of estimation of pitch. During the process of inverse filtering[3] (or some other similar operation) to remove the effect of formants, the signal to noise ratio of the filtered signal is reduced. This is because the spectral flattening operation gives equal emphasis to peaks and valleys in the spectral envelope in which the valleys correspond to relatively low SNR regions. Moreover

there is no satisfactory method for determining an exact inverse filter for short segments of speech. The effect of inaccurate inverse filtering is to introduce ambiguities in the estimation of pitch epochs from the residual signal[4]. In order to obtain a reliable pitch estimation, low frequency region of the spectrum is generally used[2,3]. But the low frequency region may not necessarily correspond to a high SNR region. Also for a bandpass filtered signal as in telephone quality speech, significant harmonics of the fundamental in the low frequency region are attenuated. It would defnitely be desirable to make use of the high SNR regions in the spectrum provided the location of such regions is known. In the absence of such a knowledge, the low frequency region is the best choice. A lowpass filter with a variable cut-off frequency need to be used depending on the SNR and the spectrum of the speech segment being analysed. However, it is difficult to estimate the cut-off frequency automatically.

Accurate and reliable pitch measurement directly from speech waveform alone is often exceedingly difficult. Because of the imperfect periodicity of glottal pulses, and the interaction of the vocal tract system and the excitation, there is an inherent difficulty in defining the exact beginning and end of each pitch period, and also in distinguishing an unvoiced segment from a low level voiced segment. These difficulties are accentuated by the presence of noise and distortion in communication channels. Time domain methods based on maximum likelihood formulation have been proposed to get a reliable pitch estimation from speech waveform in the presence of noise [5,6]. A highly reliable and sophisticated method is the semiautomatic pitch detector (SAPD) proposed by Mcgonegal et al [7].

Among the frequency domain methods, the cepstrum pitch extractor[8] is a highly reliable method. Cepstrum is the inverse Fourier transform of the log power spectrum of the speech segment. The harmonic structure in the spectrum due to periodic impulse excitation gives rise to a distinct peak in the cepstrum, while the spectral envelope is represented as fictuations mostly near the origin. Thus the formant information is separated from excitation information in the cepstrum. The best flattening of spectral envelope can be obtained by eliminating the cepstral components near the origin (low quefrency region). Because of its

precision, cepstrum method has attained a status of being a standard by which other methods are evaluated. But in case of noisy speech, noise distorts the harmonic structure in the log spectrum and thus reduces the discrimination of the cepstral peak [6].

In order to increase the reliabilty while retaining the advantages of the cepstrum pitch method, it is necessary to improve the SNR of the cepstrum in the high quefrency region so as to enhance the pitch peak against the background noise. In this paper we present a method for automatically selecting the regions around the peaks in the spectral envelope. The cepstrum for each such region is separately computed. Thus there is an increased robustness because of the selective processing and an increased reliability due to the availability of several independent cepstra.

II. SELECTION OF HIGH SNR REGIONS

Selection of regions corresponding to peaks in the envelope of short-time spectrum is based on a technique for pole-zero decomposition of speech spectra[1]. Such a decomposition is achieved based on the properties of the derivative of phase spectrum (DPS) of a minimum phase signal[9].

A. Properties of DPS

A minimum phase signal has, by definition, all its poles and zeros within the unit circle in the z-plane[10]. The pole part of the spectrum of the signal can be represented as a cascade of several first order sections with real poles and second order sections with complex conjugate poles. The DPS of a typical first order pole filter usually has significant values confined to frequencies close to origin. The magnitude of the DPS of a second order pole filter (resonator) around the resonance frequency is approximately proportional to the squared magnitude response of the filter. The DPS of the overall filter is a superposition of the component DPSs of the sections. These properties were shown to be useful in an unambiguous identification of formants from the DPS of the all pole filter obtained by linear prediction analysis[9]. The DPS corresponding to zeros is similar to that of poles except for a change in the sign of the DPS function. Specifically, the DPS of the pole part is negative and

that of the zero part is positive. Any given portion of the short-time spectrum has a considerable influence of all the poles and zeros present in the transfer function, whereas there is a negligible mutual influence amongst the component DPS functions of the poles and zeros. These simple but powerful properties of DPS have been used to accomplish the pole-zero decomposition[1] described in the following section.

B. Pole-zero decomposition

In order to represent the spectrum of a given signal by a pole-zero model, it is sufficient to consider the minimum phase correspondent of the signal. This is because the spectra of the minimum phase correspondent and the original signal are identical by definition. Let $V(\omega)$ be the Fourier transform of the minimum phase correspondent of the signal being modelled. Since all the poles and zeros of $V(\omega)$ lie within the unit circle in the z-plane, $\ln V(\omega)$ can be expressed in Fourier series expansion as follows[10]:

$$\ln V(\omega) = c(0) + \sum_{n=1}^{\infty} c(n)\, e^{-j\omega n} \tag{1}$$

where $\{c(n)\}$ are called *cepstral coefficients*. Writing

$$V(\omega) = |V(\omega)|\, e^{j\theta_V(\omega)}, \tag{2}$$

we get the real and imaginary parts of $\ln V(\omega)$ as

$$\ln |V(\omega)| = \sum_{n=0}^{\infty} c(n) \cos n\omega \qquad \text{(real part)} \quad (3)$$

and

$$\theta_V(\omega) + 2\lambda\pi = - \sum_{n=1}^{\infty} c(n) \sin n\omega \qquad \text{(imaginary part)} \quad (4)$$

where λ is an integer. Note that $\theta_V(\omega)$ represents the phase spectrum of a minimum phase signal. Taking the derivative of $\theta_V(\omega)$, we get

$$\theta'_V(\omega) = - \sum_{n=1}^{\infty} n\, c(n) \cos n\omega \,. \qquad (5)$$

From the properties of DPS we know that the significant portion of DPS due to poles is confined to the negative part of the DPS function and the significant portion of DPS due to zeros is confined to the positive part. Hence the contribution of poles and zeros can be separated by considering the negative and positive portions of $\theta'_V(\omega)$ respectively. Let

$$\theta'_V(\omega) = [\theta'_V(\omega)]^- + [\theta'_V(\omega)]^+ \qquad (6)$$

where

$$[\theta'_V(\omega)]^- = \theta'_V(\omega) \quad \text{for} \quad \theta'_V(\omega) < 0$$
$$= 0 \quad \text{for} \quad \theta'_V(\omega) \geq 0 \qquad (7)$$

and

$$[\theta'_V(\omega)]^+ = \theta'_V(\omega) \quad \text{for} \quad \theta'_V(\omega) \geq 0$$
$$= 0 \quad \text{for} \quad \theta'_V(\omega) < 0 \,. \qquad (8)$$

C. High SNR regions

The spectral envelope of a speech segment is represented by the low quefrency portion of the cepstrum, say the first M_1 cepstral coefficents $\{c(n)\}$ where M_1 is in the range of 10-30

for speech sampled at 10 kHz. The DPS $\theta_1'(\omega)$ of the spectral envelope can therefore be obtained by computing the summation in (5) for the first M_1 cepstral coefficients only. i.e.,

$$\theta_1'(\omega) = - \sum_{n=1}^{M_1} n\, c(n) \cos n\omega \tag{9}$$

The pole part of $\theta_1'(\omega)$ determines the regions around peaks in the spectral envelope and hence can be used to represent the high SNR regions of the spectrum. Fig.1 shows the DPS for a voiced speech segment for three different values of M_1. The negative of the DPS function (NDPS) is shown in the figure because it is convenient to compare the positive part of the NDPS with peaks in the spectral envelope. Speech data was sampled at 10 kHz, passed through a preemphasis filter ($1-0.92z^{-1}$), and then multiplied with a Hamming window before computing the cepstrum and the DPS. The short-time spectrum of the segment is also shown by dotted curve in the figure for comparison. It is interesting to note that regions around the peaks in the spectral envelope are represented by the positive part of the NDPS and the valleys by the negative part of the NDPS. The resolution of the individual poles in the DPS is improved as the value of M_1 is increased. The frequency bands corresponding to high SNR regions are given by the nonzero values of the function $X(\omega)$ which is defined as follows:

$$X(\omega) = \begin{cases} 1 & \text{for } \theta_1'(\omega) < 0, \\ 0 & \text{for } \theta_1'(\omega) \geq 0 \ . \end{cases} \tag{10}$$

Thus we have shown a method for automatically identifying the high SNR regions in the signal spectrum.

III. PITCH EXTRACTION

Although pitch is a low frequency percept, unlike a low frequency sinusoid pitch information is present over the entire freqency range of a speech spectrum. It may be recalled that the quasiperiodicity in a speech waveform arises due to quasiperiodic glottal pulse excitation of the vocal tract system. The quasiperiodic glottal pulses give rise to nearly periodic spectral peaks (harmonic structure) in the signal spectrum. In an earlier study it was shown that pitch epochs can be obtained by processing the high frequency portion of a voiced speech signal[11]. In fact any band of frequencies is good enough to provide the pitch information, and this band need not necessarily enclose the *fundamental*. Thus the different high SNR regions selected based on DPS can be used to obtain the pitch information.

A. DPS of Excitation

Although reliable pitch estimate can be obtained by selective filtering, accuracy is assured only if the influence of the formants is removed. There are several methods for removing the effect of formants. One of the effective methods is by the use of cepstrum. The high quefrency portion of the cepstrum has neglegible influence due to the impulse response of the vocal tract system. The log spectrum corresponding to high quefrency portion of the cepstrum would possess a truly flat spectral envelope with the superposed harmonic structure due to pitch. The harmonic structure is better represented in the DPS of the excitation portion which is given by

$$\theta_2'(\omega) = - \sum_{n=M_2}^{N/2} n\, c(n) \cos n\omega \qquad (11)$$

The upper limit in the summation is determined by the order of FFT used for computing the DFTs to obtain cepstrum. All the DFTs used in the present work were computed for a value of N=512. Fig.2 shows the NDPS for the excitation portion of the cepstrum for the same voiced segment considered for obtaining Fig.1. A value of M_2=21 was used in obtaining the NDPS. The positive and negative parts of NDPS are separately shown as the upper and lower

solid curves respectively. The short-time segment of the segment is also shown by a dotted curve in the figure. The flatness of the spectral envelope that can be obtained in the DPS using the high quefrency portion of cepstrum can be seen in Fig.2. The NDPS curve also shows the improved resolution of the fluctuations observed in the short-time spectrum. It is possible to distinguish the noisy pulses from harmonic pulses in the NDPS curve. For example, the pulses in the region 3.8 to 4.8 kHz is mostly due to noise. Such regions usually occur around the valleys of the spectral envelope.

B. Processed Cepstra

The high SNR portions of the DPS of the excitation can be obtained by multiplying $\theta_2'(\omega)$ with the window function $X(\omega)$. For each band of $\theta_2'(\omega)$ selected by $X(\omega)$, the cepstrum can be obtained using the relation (11). As a result of these operations not only an improvement in SNR is obtained in the resulting cepstrum but also several independent cepstra, one for each nonzero band in $X(\omega)$, are available. Thus there is an increased robustness coupled with reliability if the information in the independent cepstra is properly interpreted. But the narrow frequency bands in $X(\omega)$ reduce the resolution of the pitch peak in the cepstrum for these bands. This is a familiar problem of trade-off between SNR and resolution, which occur in inverse filtering problems as well[12]. The implication of this trade-off will be discussed for specific cases in the next section. The cepstrum for a bandpass DPS will have the modulation effect of bandpass filtering. This makes it difficult to distinguish the pitch peak from other peaks in the cepstrum. Moreover, the abrupt discontinuities at the edges of each band in $\theta_2'(\omega)$ would produce undesirable ripples in the cepstrum. This effect can be reduced by using a suitable window function in place of a rectangular window function for each of the nonzero bands in $X(\omega)$. The fluctuations in the cepstrum caused by the bandpass operation can be overcome by computing the Hilbert envelope[13] of the resulting cepstrum. This method of processing to reduce the ambiguities is similar to the epoch filtering method proposed for the identification of the significant instant of excitation of the vocal tract system from a bandpass signal or its linear prediction residual[4,11,14]. The processing steps involved for generating the cepstra from the DPS are as follows:

(a) Consider each nonzero band of $\theta_2'(\omega)X(\omega)$ separately.

(b) Multiply the band with a suitable extended cosine-bell window[15] in frequency domain. The width of the window depends on the width of the band being processed. For the selected lowpass section the window function is symmetric about the origin.

(c) Compute the cepstrum $\hat{c}(n)$ and its Hilbert transform $\hat{c}_H(n)$ for each of the windowed bands of $\theta_2'(\omega)X(\omega)$.

(d) Compute the Hilbert envelope $c_0(n)$ defined as follows:

$$c_0(n) = [\, \hat{c}^2(n) + \hat{c}_H^2(n) \,]^{1/2} \qquad (12)$$

The signal $c_0(n)$ will be referred to as *processed cepstrum*. We get one processed cepstrum for each nonzero band in $X(\omega)$. The details of the computational scheme are shown in Fig.3. For the purpose of computing the Hilbert transform, the property that the imaginary part of the inverse DFT of a one-sided complex sequence is the Hilbert tranform of the real part, is used. The one-sided complex sequence is obtained by setting all DFT samples beyond N/2 to zero. Speech segments of 25.6 msec duration sampled at 10 kHz is multiplied with an extended cosine-bell data window before computing the DFT. A cosine-bell data window is flat over the middle portion and has raised cosine characteristics at either end. We have found that the cosine-bell window does not significantly distort the signal waveform in different pitch periods of the analysis segment and hence is a suitable choice for cepstral pitch estimation. All the DFT computations are performed using a 512 point FFT. The number of processed cepstra depends on the number of nonzero bands in $X(\omega)$, which in turn depends both on the analysis segment and the number of coefficients M_1 used in obtaining $\theta_2'(\omega)$ in (11).

IV. RESULTS AND DISCUSSION

A segment of vowel sound of telephone quality speech is analysed to illustrate the above method. The waveform, the log spectrum and the cepstrum for this segment are shown in Fig.4. The cepstrum has a clearly distinguishable pitch peak. The DPS for low quefrencies of the cepstrum for a value of M_1=5 is shown in Fig.4f. The corresponding window function $X(\omega)$ has two high SNR regions. It may be noted that these regions correspond to peak regions in the envelope of the log spectrum. The DPS for the high quefrency part of the cepstrum is computed using a value of M_2 equals to 32. The selected portions of DPS of the excitation is shown in Fig.4h. Over the selected regions the DPS of the excitation has clearly marked harmonic structure. The processed cepstrum for both the regions show unambiguous pitch peak. The processed cepstrum corresponding to bandpass region has a sharper pitch peak compared to the pitch peak for the lowpass region because the frequency band for the bandpass region is broader. Although for this case the pitch peaks in the cepstra are unambiguous, the improvemet in SNR is evident when we observe the pitch peaks at twice the pitch period (shown by arrows).

The improvement in SNR is obtained at the cost of resolution. Since we are interested in locating only a single pitch peak, we can afford to sacrifice resolution. The number of independent segments available for processing depends on the value of M_1 and the particular segment being analysed. A large value of M_1 results in more number of nonzero bands with narrower widths in $X(\omega)$. The SNR can be expected to be higher for these bands, but the resolution of pitch peak will be very poor. On the other hand, a low value for M_1 will result in broader bands in $X(\omega)$ thus relatively reducing the SNR. These remarks are illustrated by analysing a voiced stop segment /b/. The cepstrum and the processed cepstra for three diffrent values of M_1 are shown in Fig.5 . The processed cepstra for the low frequency region only are shown. The DPS of excitation for each case is also given in the figure to show the bandwidth of the selected segments. It may be noted from the DPS that the number of segments is more for a larger value of M_1. In the processed cepstrum Fig.5c there is a sharp peak at the pitch period but there are also many spurious peaks present.

These spurious peaks are absent in processed cepstra Figs.5d and 5e but the pitch peaks are much broader. A trade-off exists between the resolution and SNR depending on the value of M_1 chosen. As M_1 is increased, though the number of available independent measurements increases, not all segments may prove to be useful. The optimum value for M_1 to be used depends on the segment, but we have found a value of 5 to be generally satisfactory.

The result of analysis for a high pitched vowel sound spoken by a female speaker is shown in Fig.6. This figure illustrates the automatic choice of the cut-off frequency for lowpass region by the NDPS curve Fig.6e. It is also to be noted that the harmonic structure in the spectrum is poor beyond the cut-off frequency. The improvement in SNR in the processed cepstrum is evident in the appearence of pitch peaks at twice and three times the pitch period.

The result of analysis for a transition segment of noisy speech is shown in Fig.7 . Noisy speech was generated by adding zero mean white Gaussian noise to the speech samples. The average SNR over the utterance was 10 dB. There are noticable dissimilarities in the successive periods of the waveform arising due to noise. The selection of frequency regions containing the harmonic structure is clearly seen. The processed cepstrum for the low frequency region has a distinct peak. But for the bandpass region there are two distinct peaks in the processed cepstrum. Since the clean data for this frame was available, the analysis of the clean data showed that the occurrence of two peaks is not due to noise. From the waveform the measured pitch periods show a large difference for the two periods seen in the waveform. This is a case of a rapid transition. To define a period there must be at least two repetions. Thus for transitions one can not define an average pitch for the segment. In the frequency domain the harmonic spacing would be different in different frequency bands. This also shows that for synthesis of transitions, the instants of significant excitation of the vocal tract cavity have to be deduced.

Processed cepstra for five successive frames in the utterance of the word "*zero*" are shown in Fig.8. The frames are separated by 20 msec. The first frame shown is a segment of voiced fricative. Only one processed cepstrum for each frame is shown. There is no clear evidence of pitch peak in the cepstrum for the voiced fricative segment (frame (a)). But the

ambiguity can be reduced by following the pitch peaks in the adjecent frames. The reliability of a pitch contour derived from the processed cepstra can be further improved by considering the contour derived from the peaks appearing at twice the pitch period. It is also clear that it is much simpler to derive a pitch contour from the processed cepstra than from the unprocessed cepstrum.

The sequence of the processed cepstra for a transition from unvoiced to voiced sound in the word "*six*" is shown in Fig.9. The successive frames in the figure are separated by 10 msec in order to study the results for transition. The processed cepstrum for the unvoiced frame (a) is distinctly different from that for voiced frames. There are many significant peaks even for high quefrencies. This suggests a method of distinguishing voiced frames from unvoiced frames. The processed cepstrum for the transition frame (b) contains several significant high quefrency components besides a distinct pitch peak. It appears that the presence of these significat high quefrency components is an indication of the presence of unvoiced data in the analysis frame (also see processed cepstrum for frame (a) in Fig.8). The pitch peaks in the processed cepstra for voiced frames (c), (d) and (e) are quite distinct. It is easy to visualize the pitch contour from the processed cepstra of successive frames, whereas the unprocessed cepstra do not give a clear indication of pitch contour. The SNR improvement in the processed cepstrum is also evident from the distinct peak at twice the pitch period for frames (d) and (e).

V. CONCLUSIONS

A method for automatically selecting the relatively high SNR regions of the short-time spectrum of a speech signal has been presented. Significat improvement in the discrimination of pitch peak against background noise is possible by processing the selected frequency bands. Ambiguities in the identification of the pitch peak in the cepstrum is reduced by computing the Hilbert envelope of the bandpass filtered cepstrum. Reliability is achieved as a result of availability of several independent cepstra, each representing a high SNR portion of the spectrum. In most cases the reliability of pitch estimation is further enhanced by the presence of an unambiguous pitch peak at twice the pitch period. An algorithm for reliable

pitch contour can be obtained based on the processed cepstra for several successive segments and also on the knowledge of several factors such as spectral level in each band, the width of each band, the trade-off between SNR and resolution.

ACKNOWLEDGMENTS

The authours wish to thank Prof. D. Raj Reddy for many useful discussions and encouragement.

REFERENCES

1 B. Yegnanarayana, " Pole-zero decomposition of speech spectra," Computer Science Department, Carnegie -Mellon University, Pittsburgh, PA 15213, Report CMU-CS-79-101, Jan. 1979.

2. L.R. Rabiner, M.J. Cheng, A.E. Rosenberg and C. A. Mcgonegal " A comparative performance study of several pitch detection algorithms, " IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-24, No.5, Oct. 1976, pp.399-418.

3 J.D. Markel, " The SIFT algorithm for fundamental frequency estimation, " IEEE Trans. Audio Electroacoust., vol. AU-20, No.6, Dec. 1972, pp.367-377.

4 T.V. Ananthapadmanabha and B. Yegnanarayana, " Epoch extraction from linear prediction residual for identification of closed glottis interval, " to appear in IEEE Trans. Acoust., Speech, Signal Processing.

5 J.D. Wise, J.R. Caprio, and T.W. Parks, " Maximum likelihood pitch estimation, " IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP-24, No.5, Oct. 1976, pp.418-423.

6 D.H. Friedman, " Pseudo-maximum likelihood speech pitch extraction, " IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-25, No.3, June 1977, pp.213-221.

7 C. A. Mcgonegal, L.R. Rabiner, and A.E. Rosenberg, " A semiautomatic pitch detector (SAPD), " IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-23, No.6, Dec. 1975, pp.570-574.

8 A.M. Noll, " Cepstrum pitch determination, " J. Acoust. Soc. Am., vol.41, No.2, Feb.1967, pp.293-309.

9 B. Yegnanarayana, " Formant extraction from linear prediction phase spectra, " J. Acoust. Soc. Am., vol.63, May. 1978, No.5, pp.1638-1640.

10 A.V. Oppenheim and R.W. Schafer, *Digital Signal Processing*, Prentice-Hall, New Jersy, 1975, ch. 7, pp.345-352.

11 T.V. Ananthapadmanabha and B. Yegnanarayana, " Epoch extraction of voiced speech, "

IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-24, No.6, Dec.1975, pp.562-570.

12 S. Treitel and E.A. Robinson, " The design of high-resolution digital filters, " IEEE Trans. Geosc., Electronics, vol. GE-4, No.1, June 1966, pp.25-38.

13 J. L. Flanagan, *Speech analysis synthesis and perception*, New York, Springer Verlag, Sec. Edn., 1972, ch.5, p.146.

14 T.V. Ananthapadmanabha and B. Yegnanarayana, " Zero-phase invese filtering technique for extraction of source characteristics, " in IEEE Conf. Rec., Acoust., Speech, Signal Processing, May 1977, pp.336-339.

15 G. D. Bergland, " A guided tour of the fast Fourier transform, " SPEC., vol.6, No.7, July 1969, pp.41-52.

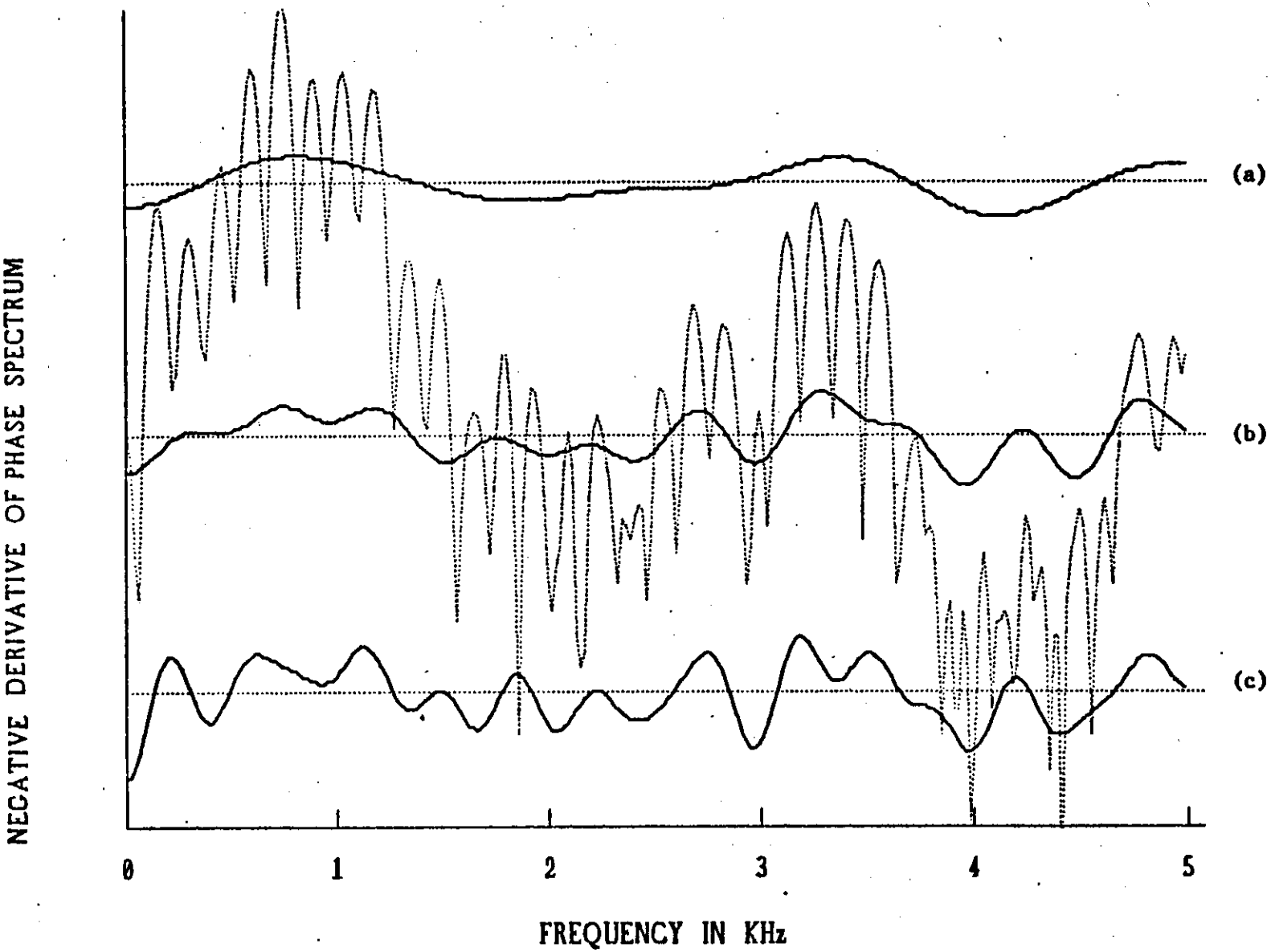

Fig.1 NDPS function for low quefrency cepstal coefficients of a vowel segment for different values of M_1: (a) M_1=10 (b) M_2=20 , (c) M_3=30. The short-time spectrum of the segment is also shown in the figure by a dotted curve.

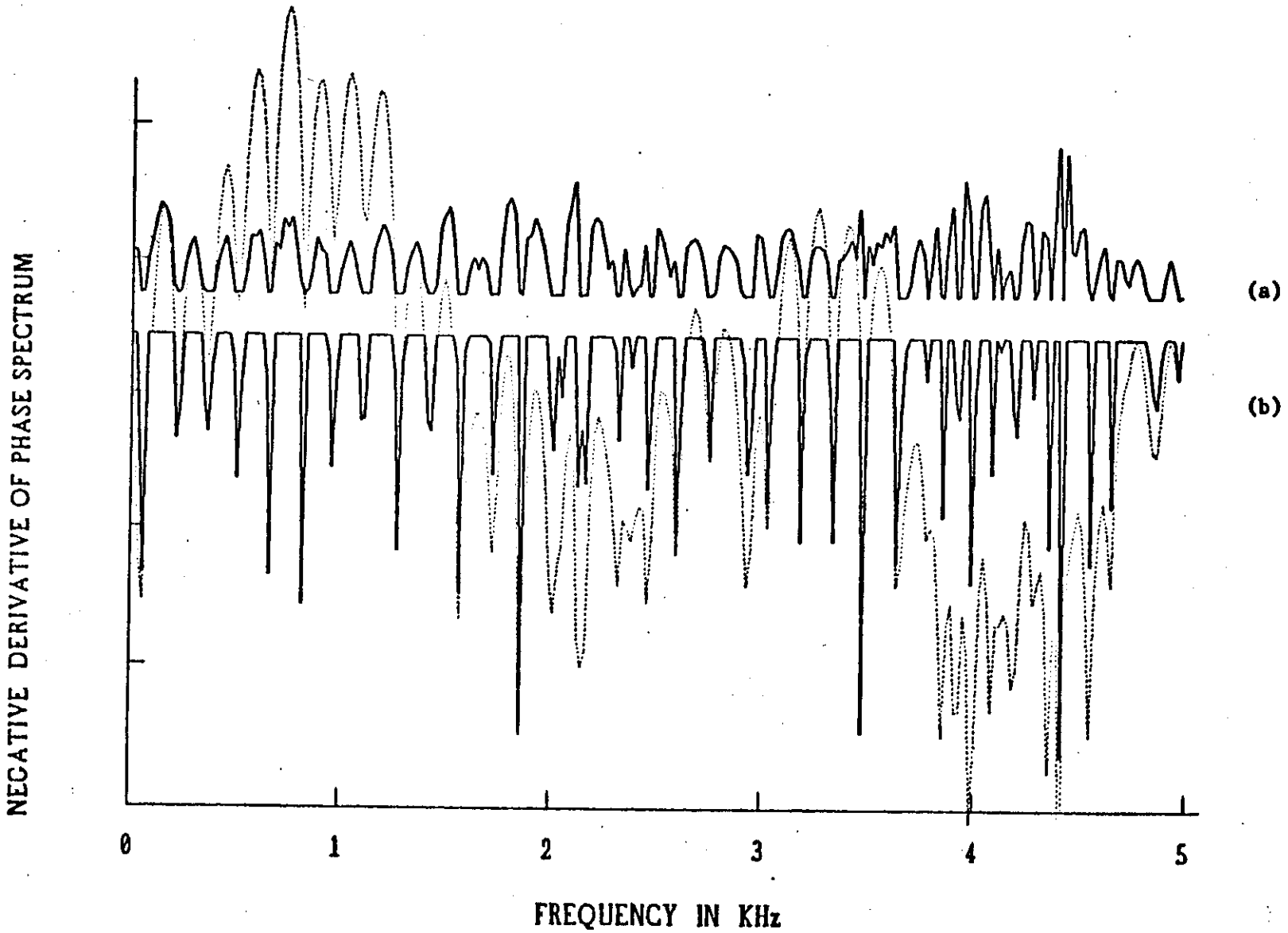

Fig.2 NDPS function for high quefrency cepstral coefficients of the same vowel sound as in Fig.1 (a) pole part of NDPS (upper plot), (b) zero part of NDPS (lower plot). The short-time spectrum of the segment is also shown in the figure by a dotted curve.

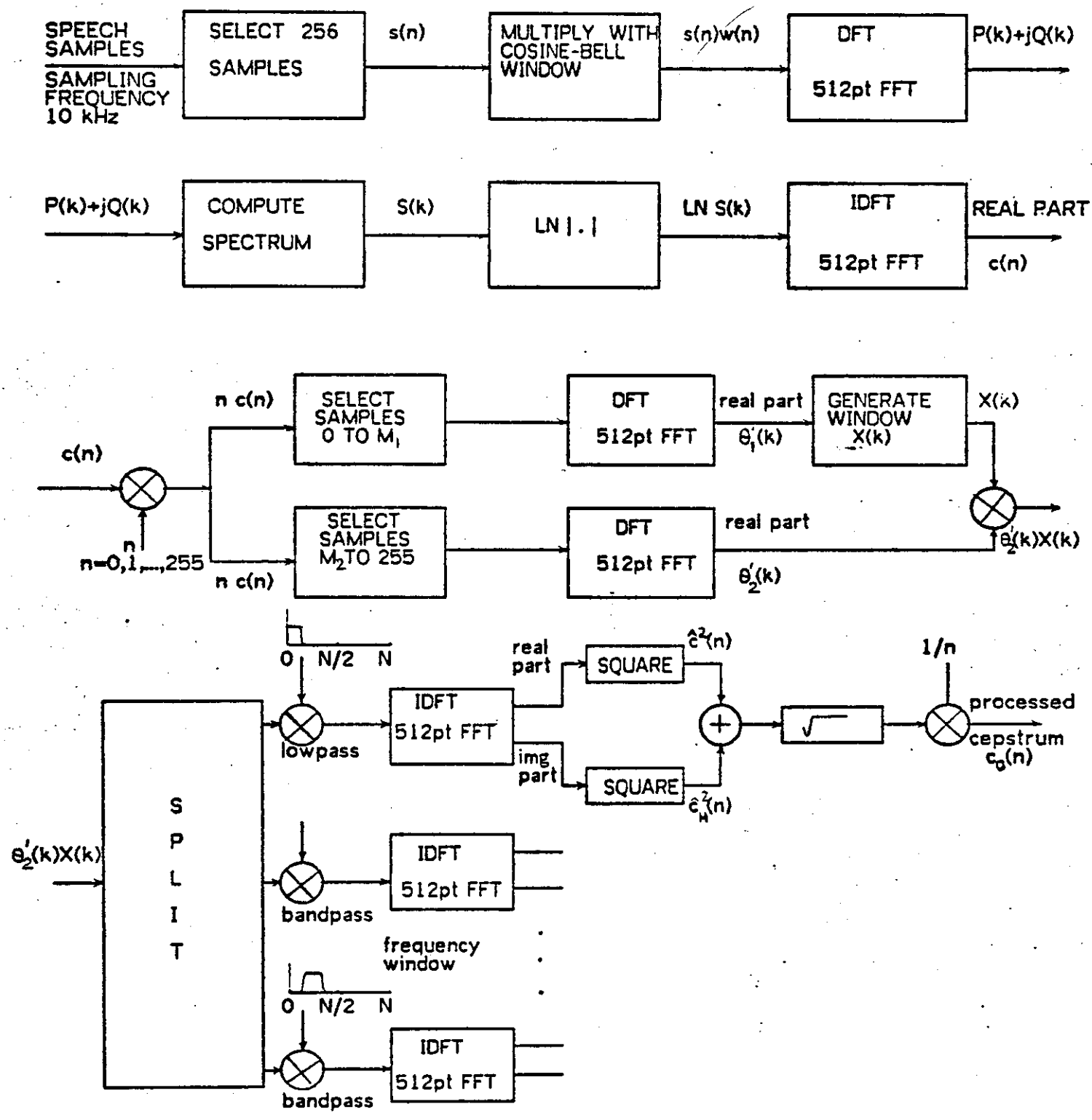

Fig.3 Blockdiagram for computation of processed cepstrum

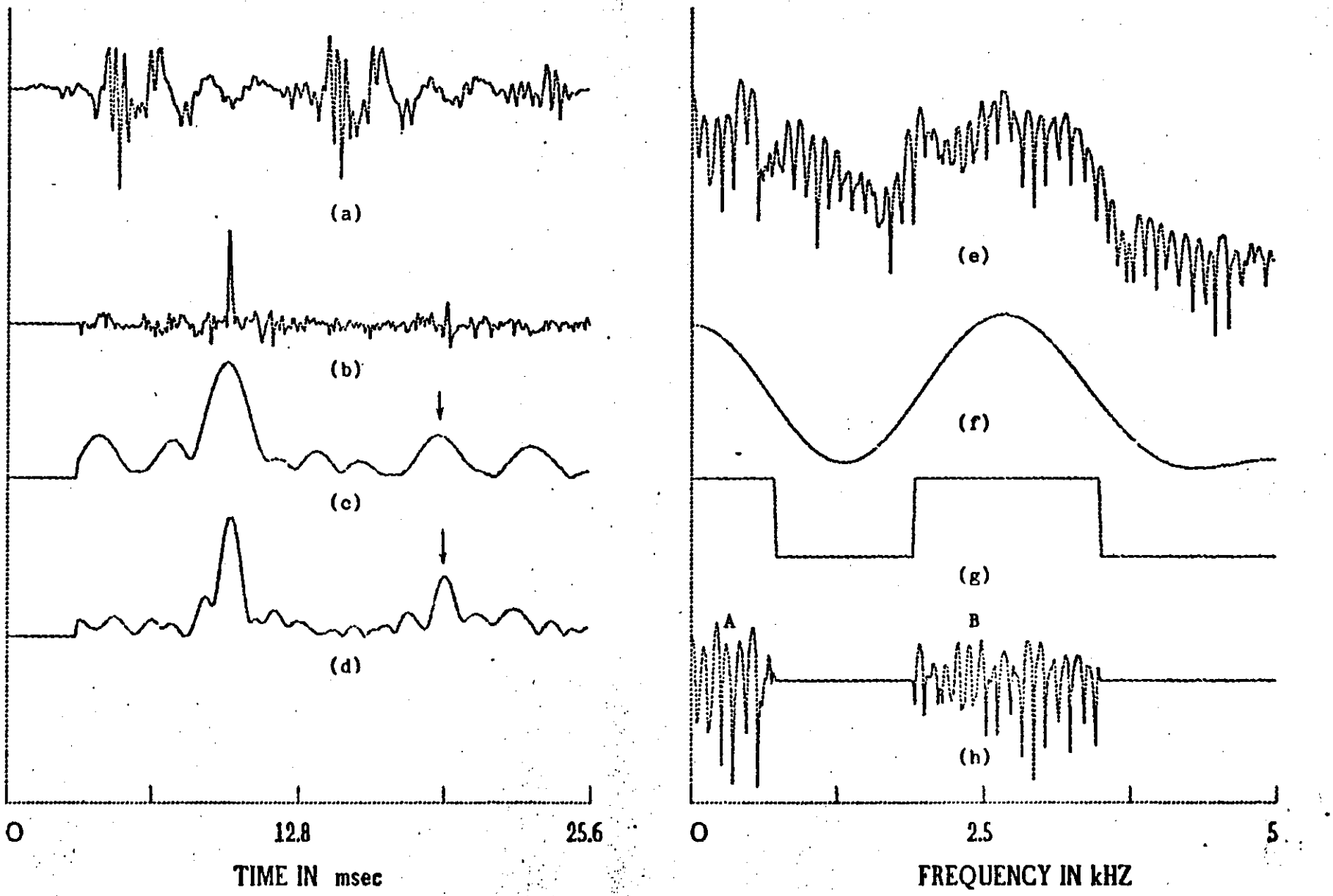

Fig.4 Analysis of a vowel segment of telephone quality: (a) Speech waveform, (b) cepstrum , (c) Processed cepstrum for region A, (d) Processed cepstrum for region B, (e) Log spectrum, (f) DPS of low quefrencies (M_1=5), (g) Window function X(ω) for high SNR regions, (h) NDPS for high quefrencies (M_2=32).

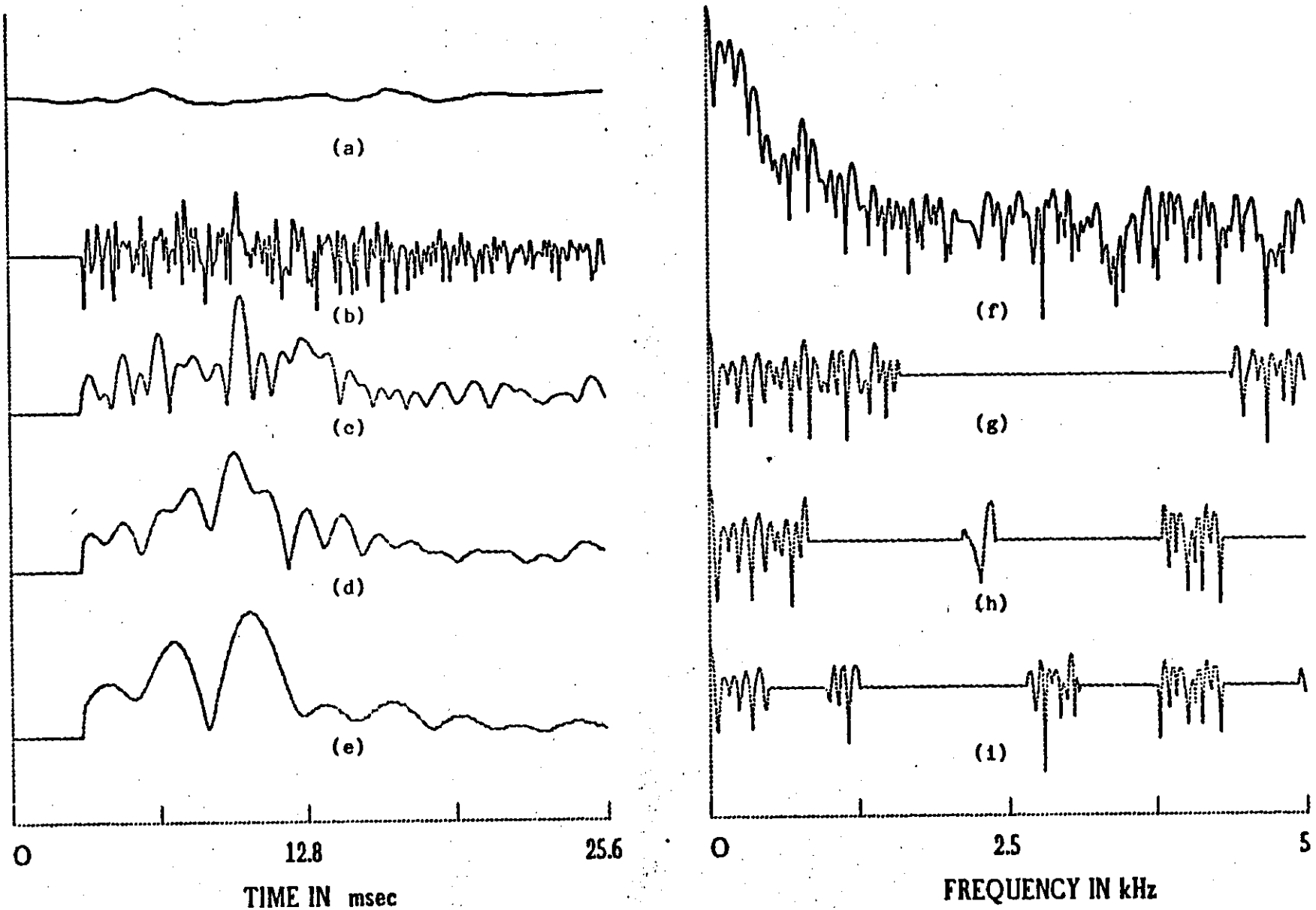

Fig.5 Trade-off between resolution and SNR: (a) voiced stop segment, (b) cepstrum , (c) processed cepstrum for M_1=3, (d) processed cepstrum for M_1=7, (e) processed cepstrum for M_1=13, (f) log spectrum, (g) NDPS for M_1=3, (h) NDPS for M_1=7, (I) NDPS for M_1=13

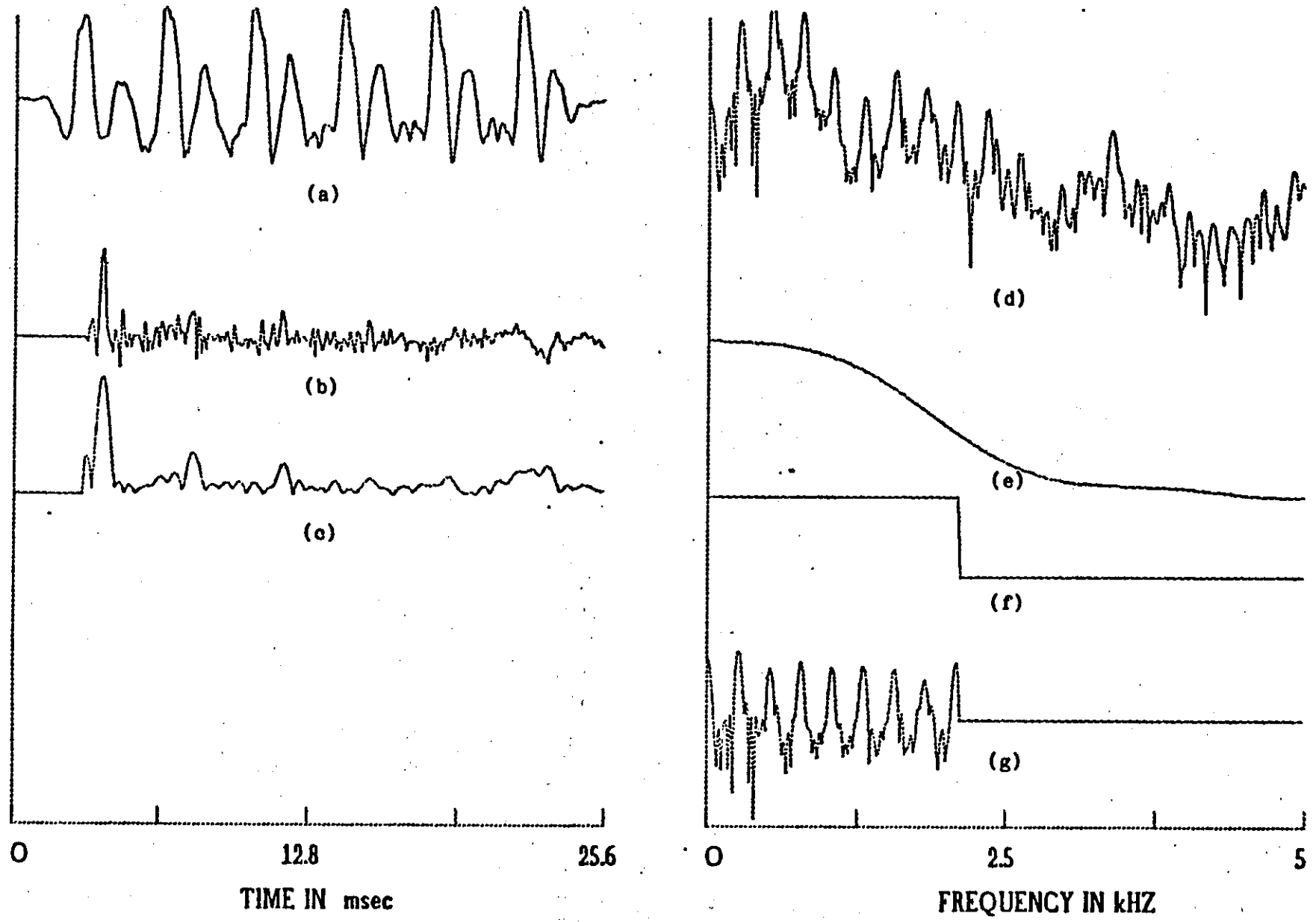

Fig.6 Analysis of a high pitched vowel segment of female voice: (a) vowel segment, (b) cepstrum, (c) processed cepstrum, (d) log spectrum, (e) NDPS for low quefrencies (M_1=5), (f) Window function $X(\omega)$, (g) NDPS for high quefrencies (M_2=32)

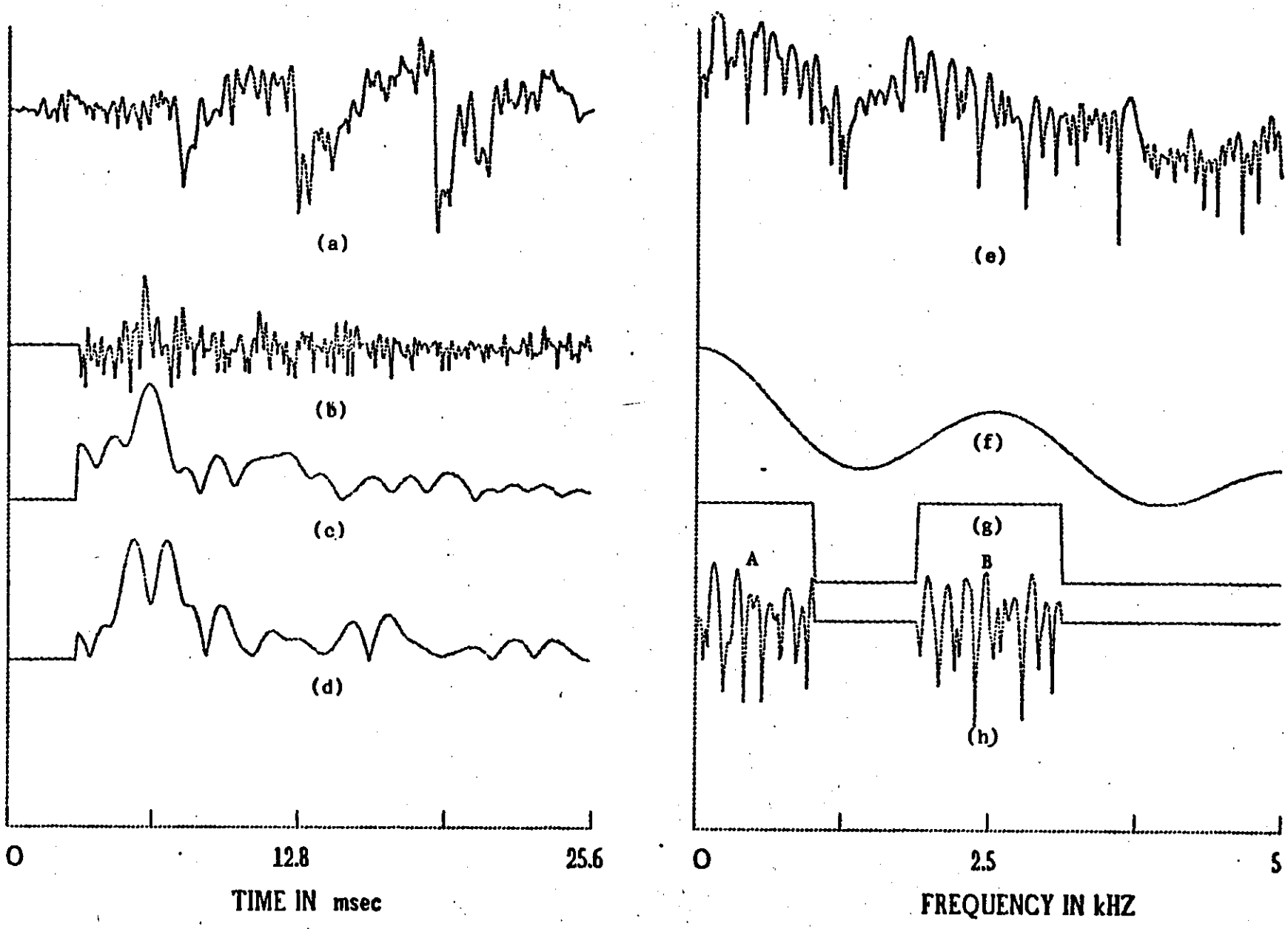

Fig.7 Analysis of a transition segment: (a) transition segment, (b) cepstrum, (c) processed cepstrum for region A, (d) processed cepstrum for region B, (e) log spectrum, (f) NDPS for low quefrencies (M_1=5), (g) Window function X(ω), (h) NDPS for high quefrencies (M_2=32)

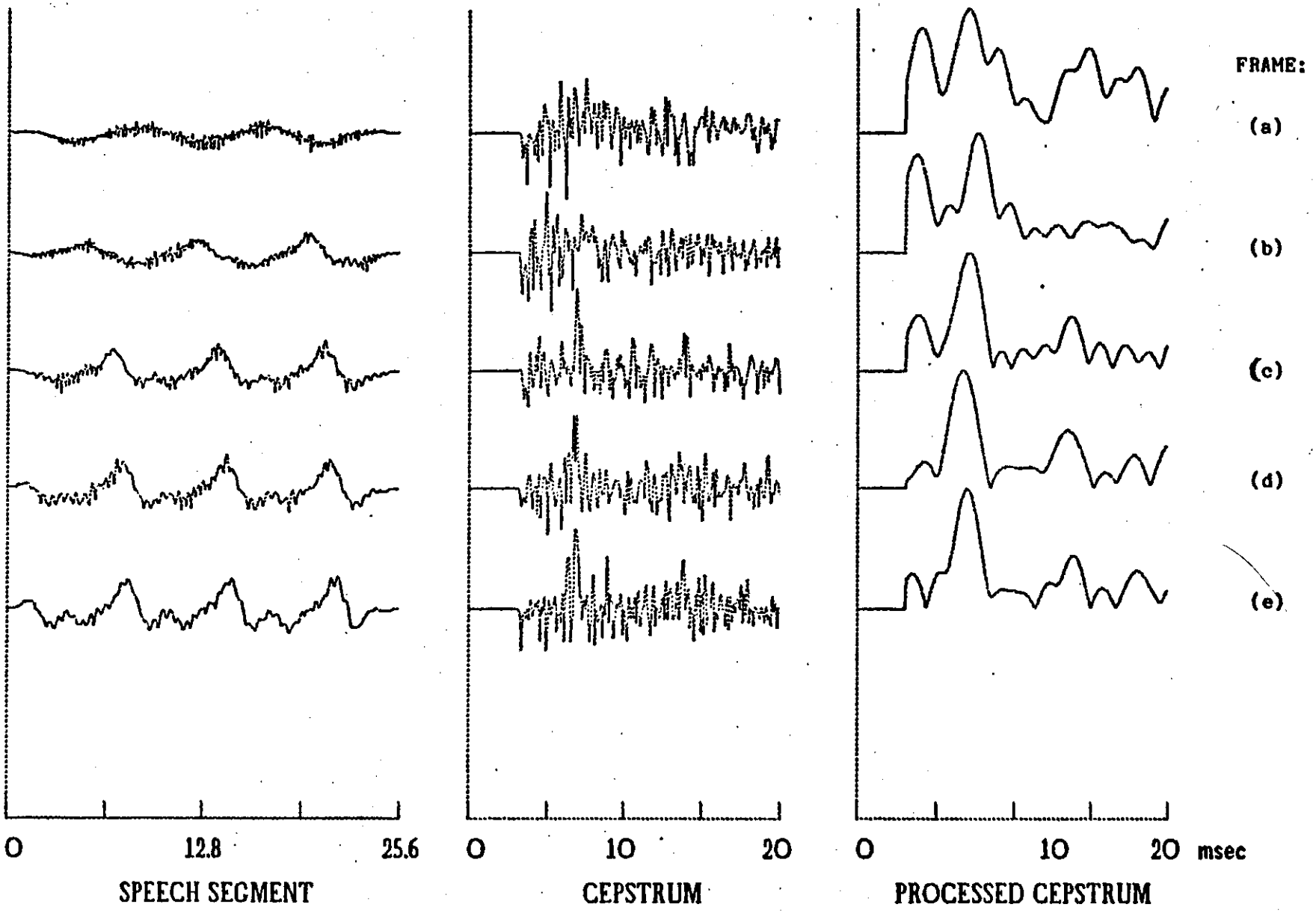

Fig.8 Processed cepstra for several successive frames separated by 20 msec for a transition from voiced fricative to vowel in the word *zero*

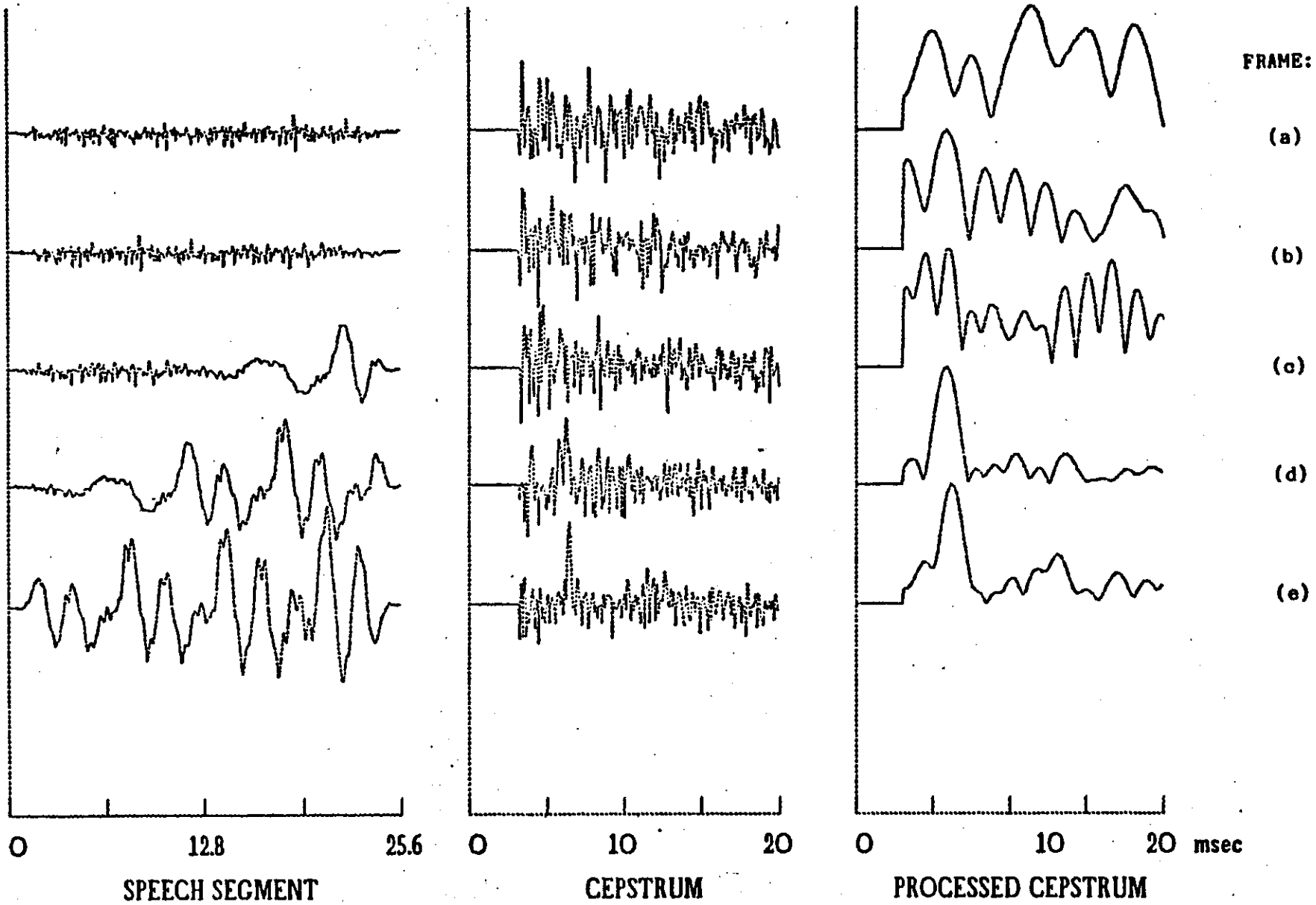

Fig.9 Processed cepstra for several successive frames separated by 10 msec for a transition from unvoiced fricative to vowel in the word *six*