Journal of the IETE
Vol No 40, Nos 2&3, March-June 1994, pp 105-108

A PC-Based Multi-resolution Spectrograph

T G THOMAS, P C PANDEY AND S D AGASHE
Department of Electrical Engineering, Indian Institute of Technology, Bombay 400 076, India.

A spectrogram is a visual representation of the temporal variation of the component frequencies of dynamic signals like those associated with speech, biomedical phenomena, underwater sounds and Doppler ultrasound signals. Here we describe a PC-based spectrograph which, in addition to the facility of generating spectrograms with specified frequency resolution, also incorporates the "combined" spectrogram that can obtain good time and frequency resolutions simultaneously. Readouts from the spectrogram can be obtained directly from the PC monitor.

Indexing terms : Spectrogram, Speech processing, Multi-resolution spectrograph

THE spectrogram is a three-dimensional visual representation of the analysed signal with time and frequency being displayed along the x-and y-axes and spectral magnitude as an intensity of shading. In addition to wide use in the area of speech processing, it has now begun to find use in music analysis, in analysis of Doppler ultrasound signals, in vibration analysis, etc.

The sound spectrograph [1,2] was the instrument of choice for the analysis of speech from 1940 through the 1970s. In the traditional sound spectrograph machine, the analysis involved a short (roughly 2 s) speech utterance being first recorded onto a magnetic storage medium. This signal was then repeatedly played and analysed by a single bandpass filter. The centre frequency of this filter was increased at the start of each playback cycle and the stylus on the plotting drum was moved across an electrically sensitive recording paper. The output current from the filter was rectified and smoothed and then fed to the stylus. In passing from the stylus to the metallic drum, the current burnt the paper and the intensity of the burning was approximately proportional to the logarithm of the current. Thus a three-dimensional frequency/time/spectral magnitude plot of the recorded signal was created, using a single filter. Depending upon the filter bandwidth used, two different spectrographic plots were obtainable: the narrowband spectrogram and the wideband spectrogram corresponding to filter bandwidths of 45 Hz and 300 Hz respectively. The narrowband spectrogram is characterised by good frequency resolution and poor time resolution. The wideband spectrogram exhibits good time resolution at the cost of frequency resolution.

The sound spectrograph could analyze only a very limited duration segment in one pass and it took nearly 10 minutes for the analysis. The dynamic range of the recording paper used was limited and it was also not possible to accurately measure the spectral magnitudes from the hard copy. The sound spectrograph is not as widely used any longer because digitally generated spectrograms with a much greater dynamic range can be displayed on a graphics terminal [3]. Spectrograms of adjustable time and frequency resolutions can be obtained and readouts can be taken directly from the monitor. The availability of DSP chips have helped in producing cost-effective real time spectrograms [4].

The system presented here employs a PC equipped with a data acquisition card and a VGA card. In addition to trade-off between time and frequency resolutions, the "combined" spectrogram [5] has also been implemented in which good time and frequency resolutions can be obtained simultaneously.

SPECTROGRAPHIC ANALYSIS OF SPEECH

The most common method of generating a digital speech spectrogram is by computing the short-time Fourier transform (STFT) of the sampled speech utterance $x(n)$ as given by

$$X(n, k) = \sum_{m=-(N/2)}^{(N/2)-1} w(m)\ x(n-m) \exp(-j2\pi km/N) \qquad (1)$$

where n is the sample number (discrete time), k is the frequency sample, N is the discrete Fourier transform (DFT) size, and $w(m)$ is an L-point Hamming window given by

$$w(n) = 0.54 - 0.46 \cos[2\pi n/(L-1)],\ 0 \le n < L \qquad (2)$$

For each slice of windowed data, the frequency spectrum is computed using the fast Fourier transform (FFT). The spectrogram is obtained by displaying $|X(n, k)|$ versus n (time) and k (frequency).

The size of the analysis window, L, depends upon the nature of the signal being considered. For a given type of window, the duration of the window is inversely proportional to its spectral bandwidth [6]. The DFT size, N, used is 512 points. In order to have the same number of spectral samples for different values of L, the sequence of length L is padded with zeroes to make it of length N before performing the DFT.

In the case of speech sampled at 10 K samples/s, and with the use of Hamming window, the traditional narrowband spectrogram (filter bandwidth = 45 Hz) can be obtained by using a window length of about 30 ms (L=290 samples). A window length of 4 ms (L=43 samples) yields the traditional wideband spectrogram (300 Hz). The bandwidth and frequency resolution could easily be changed under program control thus

Paper No 190-B ; received 1993 July 9;

lending added flexibility. In the analog spectrograph machine this would have entailed the need for physically realizable filters of variable bandwidth.

THE COMBINED SPECTROGRAM

Traditional spectrograms have either a good time resolution or a good frequency resolution, both of which exhibit certain characteristic features. Because of this tradeoff between time and frequency resolution, it is common to employ both narrowband and wideband spectrograms in the spectral analysis of speech. The narrowband spectrogram is helpful in fundamental frequency measurement and accurate estimation of formant frequencies during vowel segments. The good temporal resolution of the wideband spectrogram is useful for word boundary location and estimating formant trajectories.

Methods for obtaining a spectrogram-like representation with good time and frequency resolution in the display simultaneously have been reported [7]. However, most of these methods are computationally intensive and the resulting displays are also difficult to interpret. A method for preserving the features of both wideband and narrowband spectrograms in a single "combined" spectrogram has been reported by Cheung and Lim [5]. The combined spectrogram X_{cb} is obtained by evaluating the geometric mean of the wideband spectrogram X_{wb} and narrowband spectrogram X_{nb}. For each (n,k) the value $X_{cb}(n, k)$ of the combined spectrogram is given by

$$|X_{cb}(n, k)| = [\,|X_{wb}(n, k)|\;|X_{nb}(n, k)|\,]^{1/2} \qquad (3)$$

The geometric mean operation preserves the lighter levels of each of the two spectrograms and hence both the horizontal and vertical striations remain clearly visible in the combined spectrogram.

DIGITAL SPECTROGRAM IMPLEMENTATION

For display of the digital spectrogram, the IBM Video Graphics Adapter (VGA) with 16 simultaneous colours and 640 × 480 screen resolution has been used. The display monitor is monochrome and a 16-level gray-scale gradation is obtained using an appropriate colour sequence. The spectrogram is 500 pixels wide thus allowing 500 overlapping frames (cross-sections) independent of segment length to be displayed. This leaves sufficient space for the gray-scale plot and vertical axis calibration. The spectrogram is 256 pixels high leaving adequate space for time waveform plot, readouts and user directives.

Figure 1 shows a flowchart describing the digital spectrograph action. The digitized data created using a data acquisition card, are first read from a file and the amplitude/time waveform is displayed on the monitor. The data are windowed and passed through an optional pre-emphasis filter which emphasizes the high frequencies before processing. Pre-emphasis is needed in the case of signals like speech to approximately flatten the spectral envelope and lower the dynamic range requirement. This is achieved by passing the windowed speech samples through a filter with transfer function $(1 - az^{-1})$, where $a = 0.9375$. The STFT of the resulting

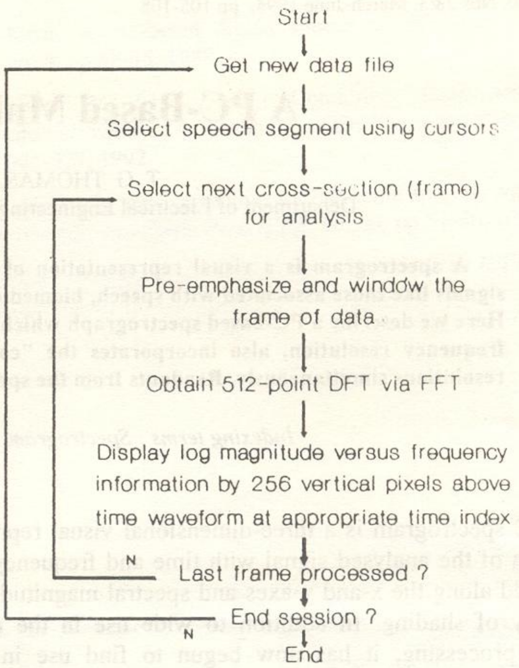


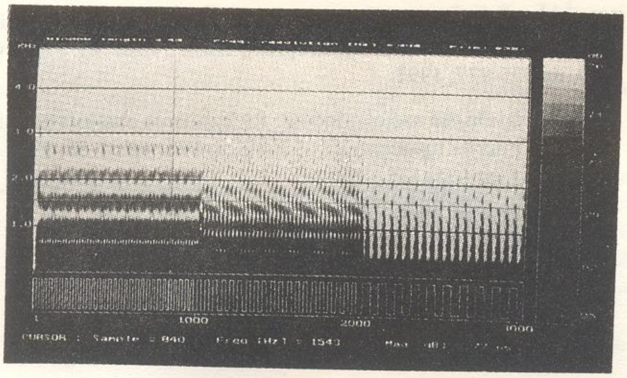
Fig 1 Flowchart for digital spectrograph action

samples is calculated via the FFT. The log magnitude in dB is computed and the spectral information for the frame is displayed by 256 vertical pixels above the last sample of the time window. The above procedure is repeated for a new cross-section by shifting the window appropriately. The amount of this shift is obtained such that the spectral information of 500 cross-sections is displayed. The display shows log magnitude as a function of frequency (y-axis) for the time duration (x-axis) of the waveform. The mapping shows high spectral magnitude as black, intermediate magnitude levels in shades of gray and absence of significant magnitude as white. The user can set values for maximum and minimum intensities in order to adjust the dynamic range of the display. Readouts from the spectrogram can be obtained by using time and frequency cursors.

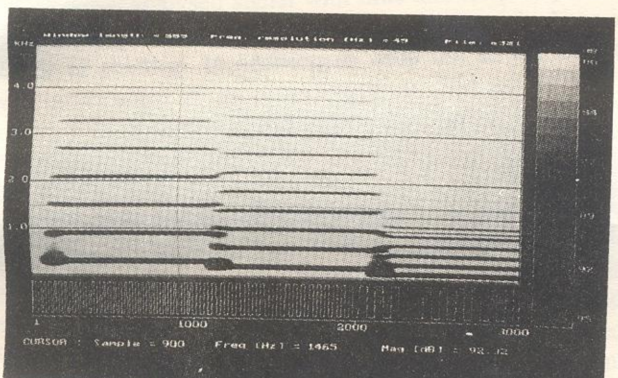
RESULTS

Typical analysis results using the spectrogram program are shown in Figs 2 and 3. The time waveform is shown at the bottom of each spectrogram and the gray scale plot is indicated to the right.

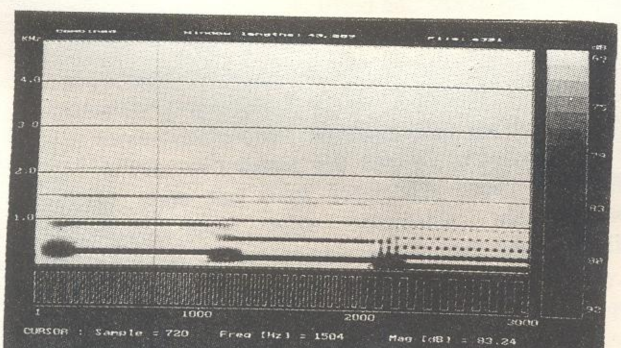
Fig 2*a* shows the wideband spectrogram (300 Hz) for a square wave whose frequency has a step variation from 300 Hz to 200 Hz and then down to 100 Hz. The fundamental and odd harmonics are clearly seen for the 300 Hz and 200 Hz portions but not for the 100 Hz portion. The abrupt variations at the boundaries as well as the vertical striations are characteristic of good time resolution. The intensities of the higher harmonics get fainter because their spectral magnitudes reduce with increasing harmonic number. Figure 2*b* shows the narrowband spectrogram (45 Hz) for the same waveform. The frequency components have come out as thin horizontal bands thus indicating improved frequency resolution. However, time resolution has suffered as is seen from the smears at the points of frequency changes. Figure 2*c* shows the combined



(*a*)

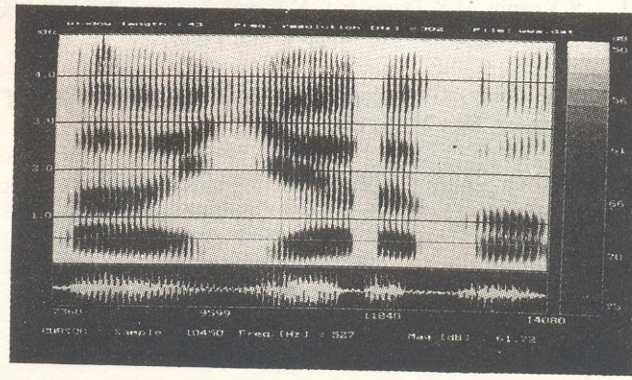


(*b*)

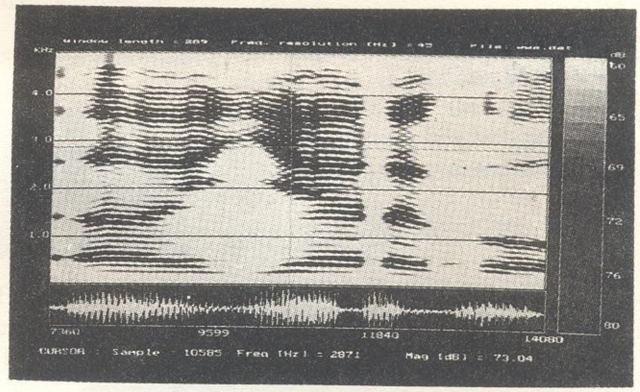


(*c*)

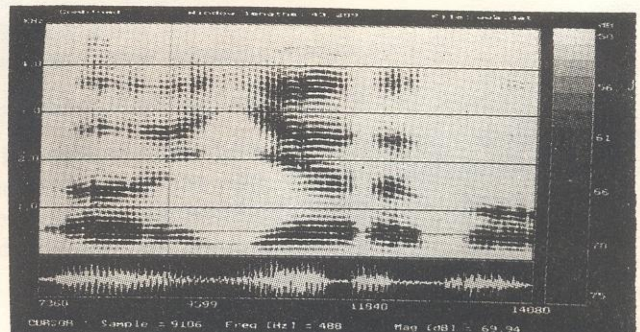
Fig 2 Spectrograms for the square wave with step frequency changes (*a*) wideband (*b*) narrowband, and (*c*) combined cases



(*a*)



(*b*)



(*c*)

Fig 3 Spectrograms for the utterance "… a year ago" (*a*) wideband (*b*) narrowband, and (*c*) combined cases

spectrogram with good resolution along both the time and frequency axes. The fundamental and odd harmonics of the 100 Hz portion, which were not seen in the wideband case, are clearly seen here.

Figure 3*a* shows the wideband spectrogram for the utterance "...a year ago" spoken by the first author. The familiar vertical striations corresponding to glottal pulses and the formant transitions are clearly seen. Figure 3*b* shows the narrowband spectrogram for the same utterance. The horizontal striations signify the improved frequency resolution. Figure 3*c* shows the combined spectrogram. The features of both wideband and narrowband spectrograms are seen here. The simultaneous improvement in resolution along both the time and frequency axes is seen from the "matted" appearance of the combined spectrogram.

REFERENCES

1. R Koenig, H K Dunn & L Y Lacey, The sound spectrograph, *J Acoust Soc Amer*, vol 18, pp 19-49, 1946.

2. R K Potter, G A Kopp & H C Green, *Visible Speech,* Van Nostrand, New York, 1947.

3. A V Oppenheim, Speech spectrograms using the fast Fourier transform, *IEEE Spectrum*, vol 7, pp 57-62, 1970.

4. L R Morris, A PC-based digital speech spectrograph, *IEEE Micro*, vol 8, pp 68-85, 1988.

5. S Cheung & J S Lim, Combined multiresolution (wideband/narrowband) spectrogram, *IEEE Trans Sig Proc*, vol 40, pp 975-977, 1992.

6. F J Harris, On the use of windows for harmonic analysis with the discrete Fourier transform, *Proc IEEE*, vol 66, pp 51-66, 1978.

7. L Cohen ,Time-frequency distributions-a review, *Proc IEEE*, vol 77, pp 941-981, 1989.