Proc. Int. Conf. Recent Advances in Biomedical Engineering, Osmania University, Hyderabad, 6-8 January, 1994

A PC-BASED SPECTROGRAPH FOR SPEECH AND BIOMEDICAL SIGNALS

T.G. THOMAS, P.C. PANDEY AND S.D. AGASHE

DEPARTMENT OF ELECTRICAL ENGINEERING, INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY, POWAI, BOMBAY-400 076

Abstract

A PC-based spectrograph is described for displaying the spectral content of dynamic signals like those associated with speech, underwater sounds, and biomedical phenomena. In addition to the facility of generating spectrograms with specified frequency resolution, it incorporates the "combined" spectrogram that can obtain good time and frequency resolutions simultaneously.

1 Introduction

A spectrograph is a machine that translates a given signal into a visual representation of its component frequencies as a function of time. With time and frequency on the horizontal and vertical axes respectively, spectral magnitude is indicated by the darkness of the pattern. In the case of speech, the spectrogram contains many features that can be related to parameters known to be important to human speech perception [1 , 2]. It has use in the area of clinical speech pathology where speech characteristics such as vowel distortions, rate of speech changes, and abnormal pitch variations can be viewed and measured. It has also been used in vibration analysis, in analysis of Doppler telemetry signals and measurements on Doppler ultrasound signal [3].

In a sound spectrograph machine [2], the spectrogram of a short (2 s) speech utterance is recorded by an ingenious electro-mechanical system onto heat-sensitive teledeltos paper. The entire process takes about 10 minutes and the dynamic range of the paper used is limited to 12 dB, and therefore these machines are not as widely used any longer. Digitally generated spectrograms can be displayed on a graphics terminal [4 , 5], which can have a much greater dynamic range. Further, spectrograms of adjustable time and frequency resolutions can be obtained and readouts can be taken directly from the screen. Several such spectrograph systems are becoming commercially available.

Here we describe a method of obtaining spectrograms on a PC equipped with a data acquisition card and a VGA card. This system permits trade-off between time and frequency resolutions. In addition, "combined" spectrogram has been implemented in which good time and frequency resolution can be obtained simultaneously.

2 Spectrographic Analysis

Among the various methods of spectrogram generation, one convenient method is to compute the short-time Fourier transform (STFT) [4] from the sampled waveform $x(n)$ as given by:

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

where n represents the discrete time axis, k the discrete frequency axis, and N the discrete Fourier transform (DFT) size. The spectrogram is obtained by displaying the STFT magnitude. The analysis window used, $w(m)$, is an L-point Hamming window [6]. The size of the analysis window, L, depends upon the nature of the signal being considered. 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.

Traditional spectrograms have either a good time resolution or a good frequency resolution, both of which may exhibit certain characteristic features. For a given window shape, the duration of the window is inversely proportional to its spectral bandwidth [6]. The choice of window duration trades off time and frequency resolution. For example, in speech analysis, wideband and narrowband spectrograms are obtained using filter bandwidth of 300 Hz and 45 Hz respectively. With speech digitized at 10 K samples/s, and with the use of Hamming window in the analysis, the corresponding window lengths are about 4 ms ($L = 43$ samples) and about 30 ms ($L = 289$ samples) respectively.

3 Combined Spectrogram

Methods have been reported for obtaining a spectrogram-like representation with good time and frequency resolution available in the display simultaneously [7]. Most of these methods are computationally intensive and the resulting displays are also difficult to interpret. In one of the simpler methods for preserving the features of both wideband and narrowband spectrograms [8], a "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:

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

The geometric mean operation preserves the valleys (lighter levels) of each of the two spectrograms and hence both the horizontal and vertical features remain visible in the combined spectrogram.

4 The Spectrogram Program

The spectrogram is displayed by using a VGA card with 16 simultaneous colours and 640×480 screen resolution and a monochrome monitor. The spectrogram is 500 pixels wide, allowing display of 500 overlapping frames, independent of segment length. It is 256 pixels high. The display also includes gray-scale plot, vertical axis calibration, time-waveform, cursor readouts, and user directives.

The digitized data files are created using a data acquisition card. The data are read from a file. The selected segment is 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. The STFT is calculated via the fast Fourier transform (FFT). The log magnitude in dB is computed and the spectral information for the frame is displayed by 256 vertical pixels above the trailing edge 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 corresponding to 500 cross-sections is ultimately 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 program allows user set values for maximum and minimum magnitudes so that the dynamic range of the display can be adjusted. Time and frequency cursors are provided whereby readouts of the STFT log magnitude versus frequency at any selected time position can be obtained.

5 Results

Two samples of analysis results are shown in Figures 1 and 2. Although photographic reproduction from the PC monitor results in the loss of dynamic range, actual results are comparable to that of traditional spectrograms.

Figure 1(a) shows the wideband spectrogram for a square wave whose frequency has a step variation from 200 Hz down to 100 Hz and then back to 200 Hz. The abrupt variations in frequency are distinctly seen. The fundamental and odd harmonics are clearly seen for the 200 Hz portions but not for the 100 Hz portion. The vertical striations are characteristic of good time resolution. Figure 1(b) shows the narrowband spectrogram for the same waveform. The improvement in frequency resolution is clearly apparent. However, time resolution has suffered as is seen from the smears at the points of frequency changes. Figure 1(c) shows the combined 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 2(a) shows the wideband spectrogram for the speech utterance "ada". The vertical striations corresponding to glottal pulses and the formant transitions are clearly seen. Figure 2(b) shows the narrowband spectrogram for the same utterance. The horizontal striations signify the improved frequency resolution. Figure 2(c) shows the combined spectrogram in which the features of both wideband and narrowband spectrograms are present.

References

[1] R.K. Potter, G.A. Kopp, and H.C. Green, *Visible Speech*, Van Nostrand, New York, 1947.

[2] R. Koenig, H.K. Dunn, and L.Y. Lacey, "The sound spectrograph," *J. Acoust. Soc. Amer.*, Vol. 18, pp 19–49, 1946.

[3] *Kay DSP Sona-graph Workstation Brochure.*

[4] A.V. Oppenheim, "Speech spectrograms using the fast Fourier transform," *IEEE Spectrum*, Vol. 7, pp 57–62, 1970.

[5] L.R. Morris, "A PC-based digital speech spectrograph," *IEEE Micro*, Vol. 8, pp 68–85, 1988.

[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.

[8] S. Cheung and J.S. Lim, "Combined multiresolution (wideband/narrowband) spectrogram," *IEEE Trans. Sig. Proc.*, Vol. 40, pp 975–977, 1992.