HIGH RESOLUTION ANALYSIS OF BIRD SOUNDS

C. Rogers
Department of Physics
East Texas State University

ETSU-Physics, Commerce, Texas 75429 USA

ABSTRACT

Bird songs and calls have been studied using modern signal processing techniques in both time and frequency domains. The complex modulations of these natural sounds are interpreted by a popular bioacoustic model of avian sound production. High resolution signal processing techniques have been applied to the problem of dual-voice detection within bird sounds. Improved time and frequency analyses of these signals have been achieved.

1. Introduction

The vocalizations of many birds are produced as a combination of sound, waves orginating from two modulated channels of air flow within the avian syrinx. In contrast, mammals and most woodwind musical instruments use a single channel of air flow as the primary source of excitation for their sound generation. It is now believed [1-2] that the vibrations of membranes within a bird's symix affect the air column of a bird's trachea in a manner similar to the action of the vocal cords on formation of vowel sounds in human speech. There are a minimum of two independent sets of syringeal membranes found in song birds which affect tracheal waveforms. This suggests that multiple non-harmonically related excitation frequencies could occur as birds sing or call.

There is limited spectrographic evidence that dual-voice signals are present in the sounds of song birds and waterfowl. Previous investigators [3-5] had available the conventional wideband and narrowband sonographic filterbanks (or the equivalent short-time Fourier transform - STFT spectrogram) for the analysis of these sounds. The spectrographic images offer the tradeoff of opposing time and frequency resolutions as required by the uncertainty principle but fails to accomplish either well. It is therefore difficult to isolate closely spaced multiple components for these sounds accurately.

In this effort, several quadratic high resolution signal processing techniques were considered. These algorithms have been developed specifically for studying signals with nonstationary spectra. These include improved methods (a) for time-frequency and time-scale analysis, and (b) for precision waveform demodulation.

2. Analytical Techniques

2.1 Time-Frequency Representation

The Wigner-Ville distribution, WVD and its variants [6-7] provide significant advantages in time and frequency resolution over the standard STFT. Starting with the WVD of a signal, s(t) in integral form is

$$WVD(t,\omega) =$$

$$\int_{-\infty}^{\infty} w(\tau/2)\, w^*(-\tau/2)\, f(t+\tau/2)\, f^*(t-\tau/2)\, e^{-j\omega\tau}\, d\tau$$

Writing this in the frequency domain gives

$$WVD(t,\omega)$$

$$= \frac{1}{\pi} F(t,2\omega) * F^*(t,2\omega)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} F(t,\omega+\nu/2)\, F^*(t,\omega-\nu/2)\, d\nu$$

It is well known that this quadratic analyzer has high resolving power but is plagued with troublesome interferences especially when the signal contains close multiple frequency components.

Stankovic [10] has recently developed an effective form that supresses the spurious crossterms produced by the WVD. The method uses a windowing kernel in the frequency domain. Stankovic has also suggested a method for computing this distribution efficiently by recursion. The resulting smoothed pseudo WVD will be called the SPWVD. The discrete-time version is

$$SPWVD(n,k) = \sum_{i=-L}^{L} W(i)\, F(n,k+i)\, F^*(n,k-i)$$

0-7803-2431-5/95 $4.00 © 1995 IEEE

where L controls the length of the filter. If L is zero, the window shrinks to a Dirac delta at $v = 0$ and the SPWVD operates as the STFT. As the length of the window increases, the SPWVD operation approaches that of the WVD without any crossterms suppression.

A multiple component synthetic test signal was generates by adding (a) a constant AM tone with step discontinuities in both amplitude and frequency at t = 1 second and (b) a higher frequency sinusoidal FM warble. Figure1presents the STFT, PWVD, and SPWVD representations of this signal without added noise. The SPWVD offers the best overall performance.

2.1 Waveform Demodulation

Many years ago Gabor [9] developed a quadratic demodulator for signals of the form $s(t) = a(t)\cos(\phi(t))$. He suggested that the Hilbert transform of the signal be used to construct an analytic signal representation

$$x(t) = s(t) - j\,\mathcal{H}[s(t)]$$

From the analytic signal the instantaneous amplitude and frequency are readily estimated by

$$a(t) = \sqrt{x^*(t)\,x(t)}$$

$$\omega(t) = \frac{d}{dt}\tan^{-1}\left[\frac{H[s(t)]}{s(t)}\right]$$

Kaiser has described a discrete-time form of the Teager quadratic energy operator [11-13]. This operator is defined as

$$\Psi_D[x(n)] = x^2(n) - x(n-1)\,x(n+1)$$

By using the forward time difference $y(n) = x(n) - x(n-1)$ and the backward time difference $z(n) = x(n+1) - x(n)$, we can estimate of the instantaneous frequency and amplitude by the following:

$$a(t) = \sqrt{\frac{\psi[x(n)]}{1 - G^2(n)}}$$

$$\Omega(n) = \cos^{-1}[G(n)]$$

where

$$G(n) = 1 - \frac{\Psi[y(n)] + \Psi[z(n)]}{4\,\Psi[x(n)]}$$

Figure 2 shows application of the Kaiser-Teager of the demodulation technique to the synthetic signal used in figure 1. The input was pre-filtered at 0.4 Nyquist to reject one band of components. The output of the demodulator is seen to acurrately track the test signal. At the discontinuities oscillations are present, but are controlled by post-median filtering.

3. Results

The analysis of sound fragments from two songbirds are reported. These data were digitized at 44.1 kHz with a linear 16 ADC from an audio CD whose original recordings were produced by Allen and Kellogg [14]. Figure 3a shows sounds produced by a Bewick's wren song note and figure 4a sounds from the song of a woodthrush. SPWVD analysis of these sounds is given in figures 3b and 4b for the wren and the woodthrush, respectively. The SPWVD for the wren shows two closly spaced frequencies with energy alternating between them. The SPWVD for the woodthrush is more complicated with two nonharmonically related primary frequencies and other secondary energy transients present. Figures 3c and 4c give the Kaiser-Teager demodulator output for the song fragments. The instantaneous amplitude tracks the signal envelope as would be expected, and the instantaneous frequency rises to a near constant level. Also we observe that AM and FM modes are coupled as would be expected for two oscillators producing beats. In figure 4c, the peaks in the instantaneous frequency may be indicating pitch intervals for these sounds; however, more research is needed to substantiate this. The instantaneous amplitudes follow the envelope.

4. Summary

Several modern signal processing techniques have been applied to the problem of dual-voice detection in bird sounds. Improved time-frequency representations and demodulation of these signals have been achieved. The results obtained are encouraging, and this work continues. Future goals include estimating the pitch values for dual-voice avian sounds.

References

[1] M. Miskimen, "Sound Production in Passerine Birds",The Auk, **68**, pp. 493-504, October 1951.

[2] D. Chamblain, W. Gross, G. Cornwell, and H. Mosby, "Syingeal Anatomy in the Common Crow", The Auk, **85**, pp. 244-252, April 1968.

[3] C. Greenewalt, "How Birds Sing", Scientific American, **221**, pp.126-139, November 1969.

[4] R. Stein, "Modulation in Bird Sounds", The Auk, **85**, pp. 229-243, April 1968.

[5] A. Gaunt, "An Hypothesis Concerning the Realtionship of Syringeal Structure to Vocal Abilities", The Auk, **100**, pp. 853-862, October 1983.

[6] M. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer, and L. Raphael, Wavelets and Their Applications, Jones and Barlett, Inc., 1992.

[7] Cohen, L. "Time-Frequency Distribution - A Review", Proc IEEE, **77**, pp. 941-981, 1989.

[8] Hlawatsch, F. and Boudreaux-Bartels, G.F., "Linear and Quadratic Time-Frequency Signal Representations", IEEE Signal Processing Magazine, **9**, pp. 21-67, 1992.

[9] Gabor, D., "Theory of Communication", J. IEE **93**, pp. 429-457, 1946.

[10] Stankovic, L. "A Method for Time-Frequency Analysis", IEEE Trans. Signal Processing, **42**, pp. 225-229, 1994.

[11] J. Kaiser, "On a Simple Algorithm to Calculate the 'Energy of a Signal', pp. 81-384, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Albuquerque, New Mexico, April 1990.

[12] P. Maragos, J.F. Kaiser, and T.F. Quatieri, "On Amplitude and Frequency Demodulation Using Energy Operators", IEEE Trans. Signal Processing, **41**, pp. 1532-1550, April 1993.

[13] A. Potamianos, and P. Maragos, "A Comparsion of the Energy Operator and the Hilbert Transform Approach to Signal and Speech Demodulation", Signal Processing, **37**, pp. 95-120, 1994.

[14] A Field Guide to Bird Songs, Houghton Mifflin Company, 1990.

Figures

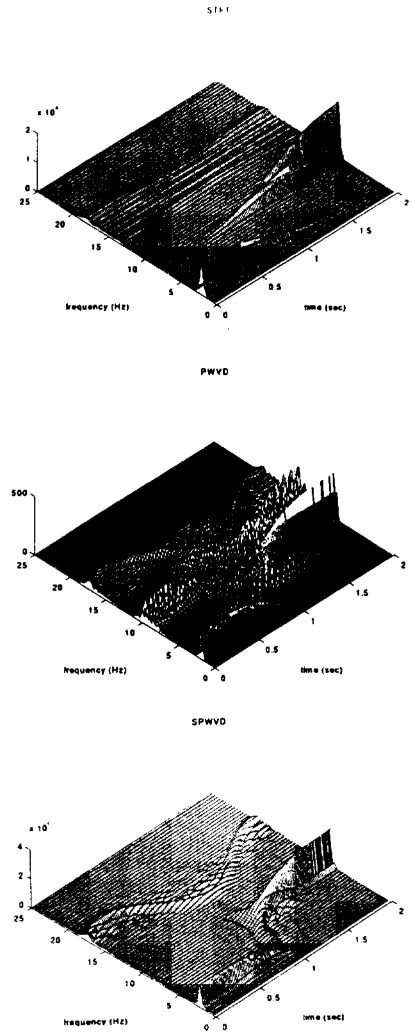

Figure 1

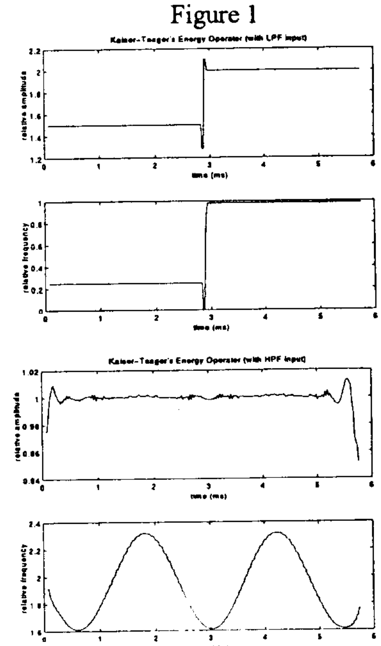

Figure 2

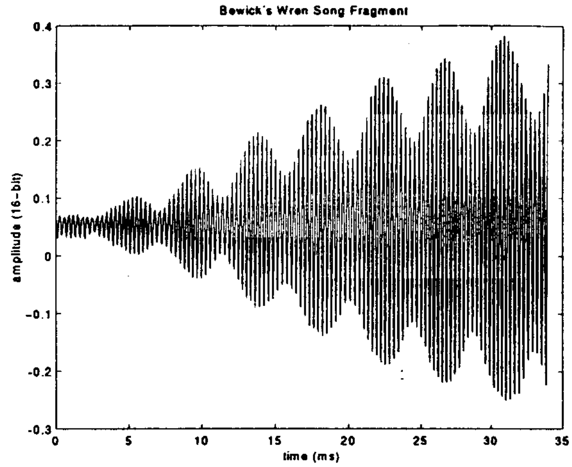

Figure 3a

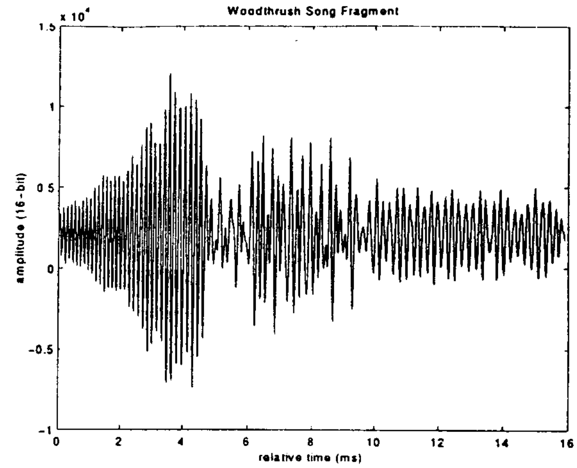

Figure 4a

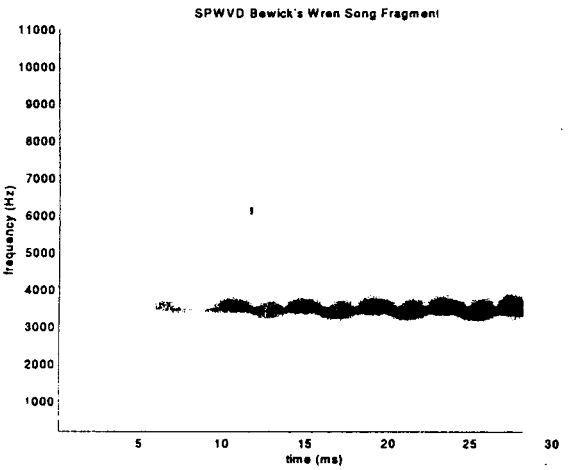

Figure 3b

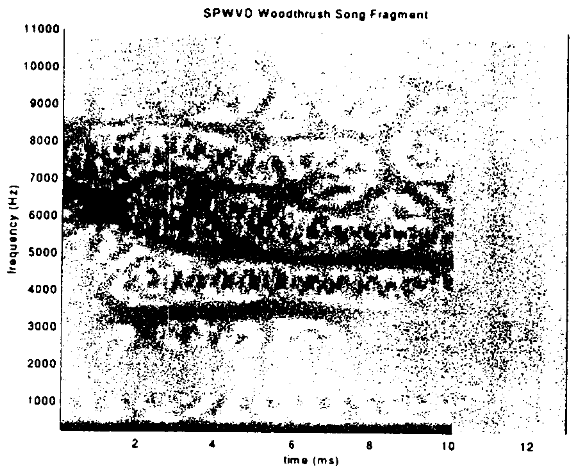

Figure 4b

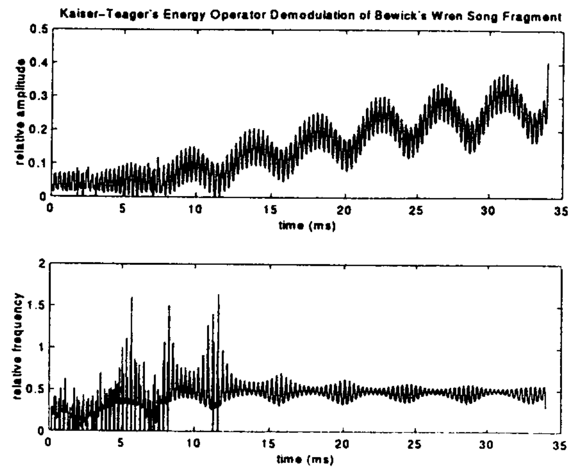

Figure 3c

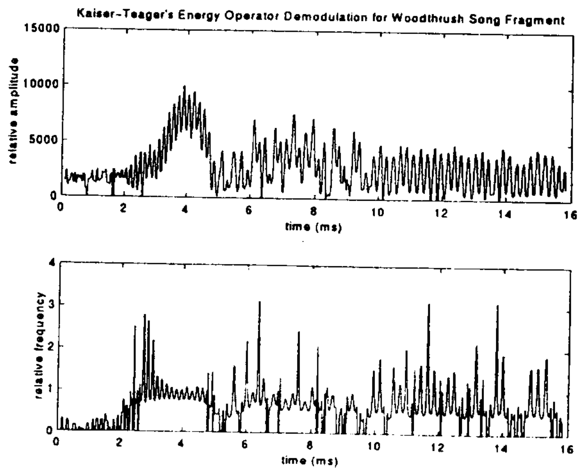

Figure 4c