Audio Engineering Society

Convention Paper

Presented at the 119th Convention
2005 October 7–10 New York, New York USA

A Spectrogram Display for Loudspeaker Transient Response

David W. Gunness[1], William R Hoy[2]

[1] David W. Gunness, Eastern Acoustic Works, Inc., 1 Main St., Whitinsville, MA 01588
david.gunness@eaw.com

[2] William R Hoy, Eastern Acoustic Works, Inc., 1 Main St., Whitinsville, MA 01588
bhoy@eaw.com

ABSTRACT

A spectrogram is a two-dimensional depiction of a waveform or transfer function in which frequency is depicted on one axis and time is depicted on the other. The level is plotted against frequency and time by using a color or gray scale. If the time resolution is constant, the display is usually referred to as a Fourier transform spectrogram. If the time resolution is scaled to the frequency, it is usually referred to as a wavelet transform spectrogram. In this paper, we present a novel and efficient method for calculating a wavelet transform spectrogram, which is optimized for the analysis of loudspeaker transient response. The new method employs complex convolution of the frequency response, rather than explicit time domain windowing or the wavelet transform.

0 INTRODUCTION

0.1 Acoustical Response Displays

The most commonly used displays of acoustical transfer functions of loudspeaker systems are frequency response magnitude and phase, and impulse response. A frequency response display gives detailed frequency content but obscures the response against time. The impulse response displays the response against time, but obscures the frequency content. An intermediate form of display, the spectrogram, represents the frequency content as a function of time. These three dimensional displays depict time on one axis, frequency on the other axis, and magnitude as a color or gray scale value. Because the spectrogram displays frequency as well as time, it offers valuable insights into the behavior of loudspeakers. Figures 0.1.1, 0.1.2, and 0.1.3 show three different displays of the transfer function of a high frequency horn. While all three depictions represent the same information, the spectrogram gives the best visual indication of the frequency dependent transient response.

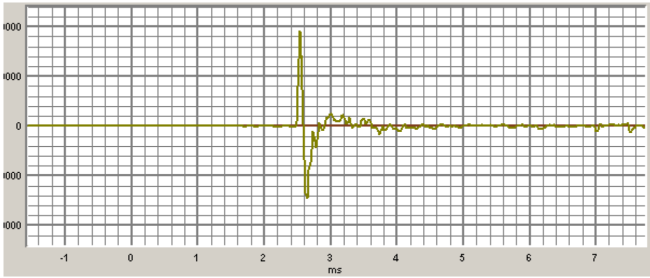

Figure 0.1.1 Impulse Response of an HF Horn

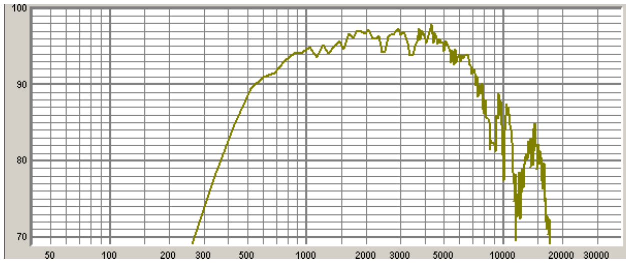

Figure 0.1.2: Frequency Response of an HF Horn

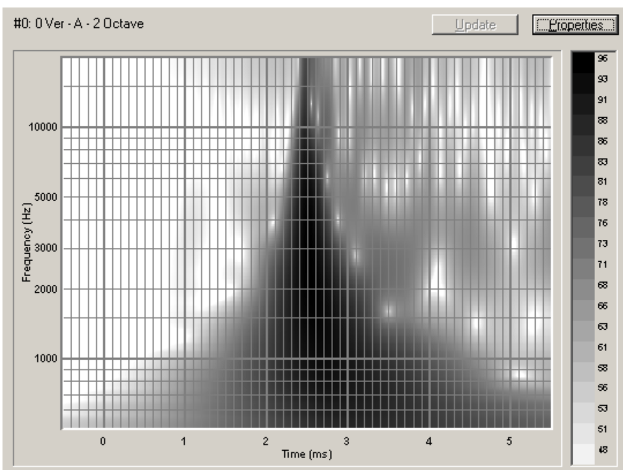

Figure 0.1.3: Wavelet Spectrogram of an HF Horn

0.2 Spectrogram Displays

Several methods exist for calculating spectrograms, including the windowed FFT method (WFT), and the wavelet transform.

The windowed FFT method produces a spectrogram by applying a data window to the impulse response and computing the FFT of the windowed data. The window is progressively shifted in time to obtain a frequency response for each discrete time. Figure 0.2.1 shows the effective basis function for the WFT.

The width of the window corresponds to a small multiple of the period at low frequencies, and to a large multiple of the period at high frequencies. Consequently, low frequencies tend to be displayed with inadequate frequency resolution, and high frequencies tend to be displayed with inadequate time resolution.

The wavelet transform method is based on a mother wavelet, and scaled and time-shifted versions of the mother wavelet. The impulse response is essentially re-stated as the sum of a series of wavelets of various lengths.

Since the width of the wavelets is inversely scaled to frequency, the time resolution is a constant number of periods, which in the frequency domain equates to a constant percentage bandwidth or fractional octave resolution. Short basis functions represent high frequencies with fine time resolution. Long basis functions represent low frequencies with coarse time resolution. The varying length of the basis functions is illustrated in Figure 0.2.1. Unfortunately, the wavelet transform is computationally expensive, especially for optimal wavelet shapes. Many references on wavelet theory are available including [1], [2], [3], and [4].

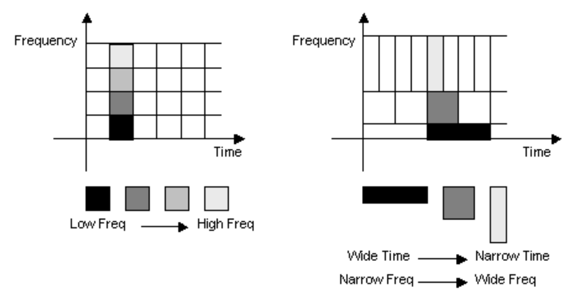

Figure 0.2.1: Windowed FFT Time-Frequency Plane vs. Wavelet Transform Time-Frequency plane

0.3 Overview

In the April, 2000 issue of the Journal of the Audio Engineering Society [5], Hatziantoniou and Mourjopoulos presented methods for calculating fractional-octave smoothing, and identified the windowing property of complex convolution. The method presented here expands on this work, in order to implement an efficient method for calculating spectrograms.

In this new method, a complex frequency response is "smoothed" by applying a low pass filter to the complex frequency data. For each successive time increment, the phase response of the frequency data is shifted by an amount determined by the specified time increment. The efficiency of the method is due in large part to a novel implementation of the smoothing filter. The zero-phase-shift low pass filter is implemented with forward and backward passes of an IIR filter.

In this paper, we will describe the implementation of the new method, and give several examples of its use. We will give examples of the spectrogram's ability to expose subtle loudspeaker transient response aberrations. And, we will discuss the resolution limitations of spectrograms.

Spectrograms are subject to the generalized Heisenberg uncertainty principle, which states that it is not possible to be precise simultaneously in the time domain and the frequency domain. Fine frequency resolution can only be obtained if the time resolution is coarse and vice versa. The best resolution trade-off is achieved when the time window is Gaussian; which can be approached arbitrarily closely by increasing the order and/or complexity of an IIR smoothing filter.

1 THEORETICAL BACKGROUND

1.1 Time and Frequency Resolution

A central concept governing spectrogram displays is the inverse relationship between time and frequency. Precise time information can only be obtained by sacrificing frequency precision.

A brief explanation of the WFT method should illuminate this concept. In the WFT method, the impulse response is multiplied by a window centered at the specified time. Then, the FFT of the windowed portion of the signal is calculated, and displayed as one column of data - the frequency response at that point in time. This process is then repeated for each column of the spectrogram. The process is illustrated in Figure 1.1.1.

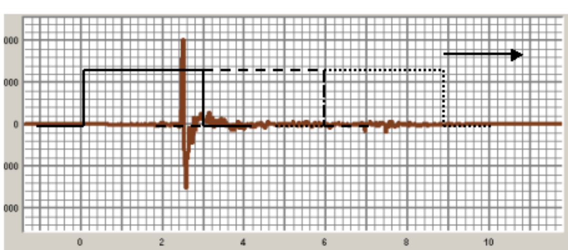

Figure 1.1.1: Rectangular time window, successively repositioned

The width of the time window determines the size of the FFT, and hence the frequency spacing of the resulting frequency response. A wide time window produces closely spaced frequency points; while a narrow time window produces widely spaced frequency points.

This property is also expressed in the Fourier transform of the window shape. Figure 1.1.2 shows a rectangular window and its Fourier transform - which is a sinc function (sin(x)/x). As the width of the time window is increased, the sinc function becomes narrower. In the same way that the width of the time window defines the time resolution, the Fourier transform of the time window defines the frequency resolution.

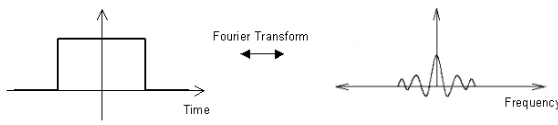

Figure 1.1.2: Rectangular time window and its FFT.

1.2 Window Shape

The width of the Fourier transform of the time window determines the frequency resolution, but its shape is also significant. The ringing that is evident in the sinc function is indicative of the ringing that occurs in the Fourier transform of rectangular-windowed data. If a rectangular window was employed in a spectrogram, a single-frequency sine wave would be displayed as a series of frequencies, rather than a single, broadened frequency.

Figure 1.2.1 shows a windowed 1 kHz sine wave, and Figure 1.2.2 shows its Fourier transform.

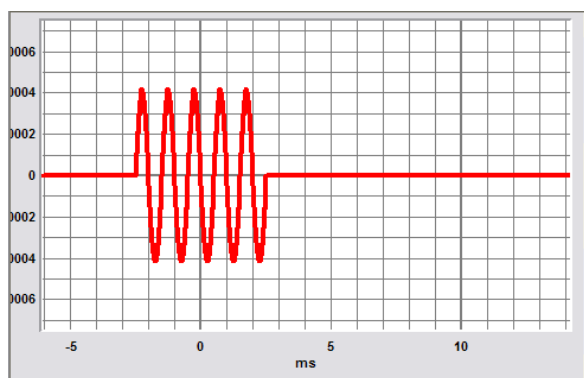

Figure 1.2.1: 1 kHz Sine Wave, Windowed by 5 ms Rectangular Window

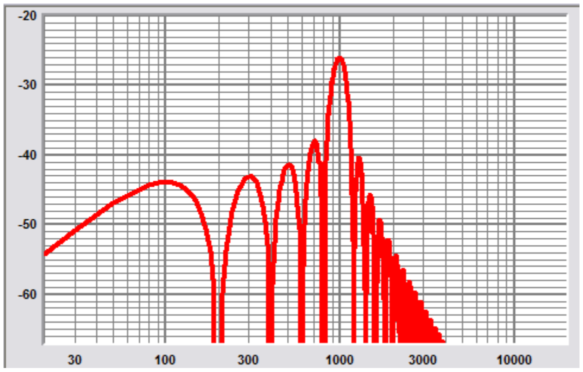

Figure 1.2.2: Fourier Transform of Windowed Sine Wave Shown in Figure 1.2.1

If the frequency response shown in Figure 1.2.2 were displayed as a column of a spectrogram, the display would mislead, by implying the presence of several sine waves of various frequencies. It is desirable, then, to select a window shape that has a Fourier transform which is free of ringing. By way of example, Figure 1.2.3 and 1.2.4 represent the same sine wave windowed by a Hann window.

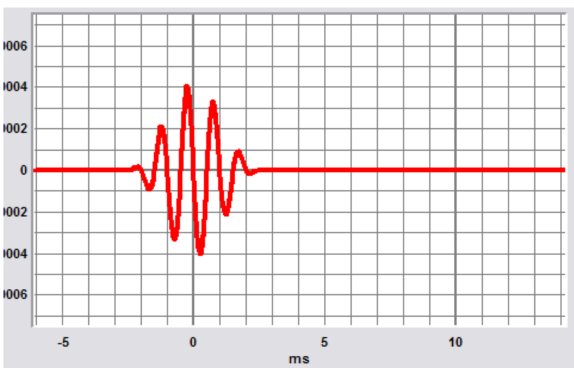

Figure 1.2.3: Hann Windowed Sine Wave

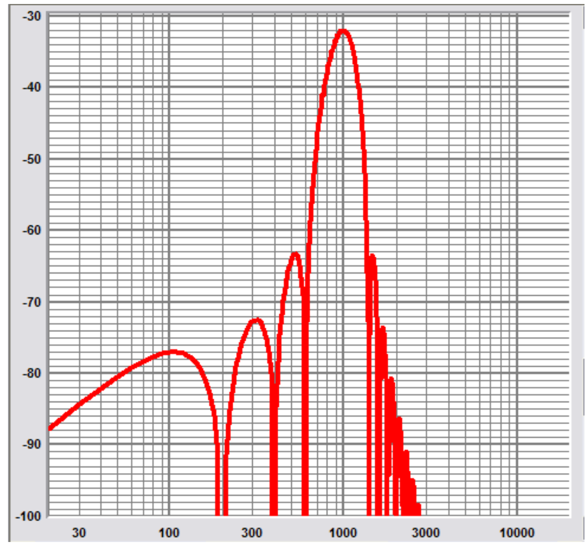

Figure 1.2.4: Fourier Transform of Hann Windowed Sine Wave

The ringing in the Fourier transform of the Hann window is much lower in level than that of the rectangular window, but is still very evident. While each of the well-known window types has unique properties, all suffer from ringing, because they are all truncated. A non-truncated window is better suited. Figures 1.2.5 and 1.2.6 show a Gaussian windowed sine wave and its Fourier transform.

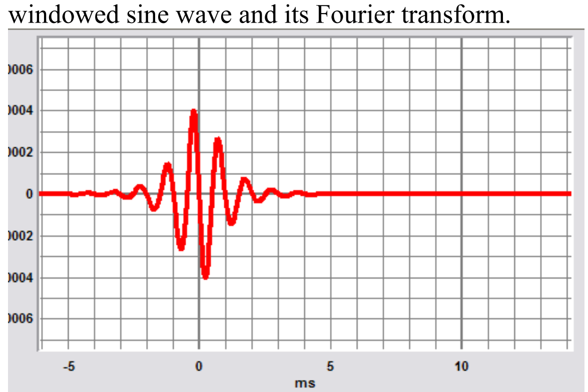

Figure 1.2.5: Gaussian Windowed Sine Wave

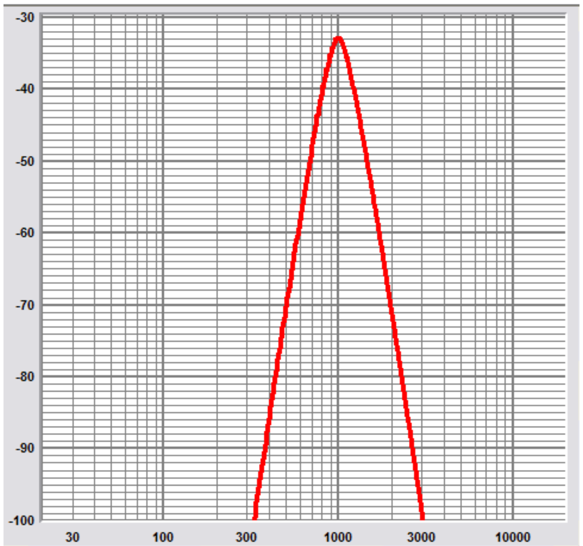

Figure 1.2.6: Fourier Transform of Gaussian Windowed Sine Wave

The complete lack of ringing in the Fourier transform of the Gaussian window makes it an excellent candidate for use in generating spectrogram displays. The next section lends further mathematical support to this observation.

1.3 Heisenberg Uncertainty Principle

Spectrograms are subject to the generalized Heisenberg uncertainty principle, which states that the exact position and the exact momentum of an object cannot both be known simultaneously [6]. In its general form, the uncertainty principle can be extended to apply to any pair of conjugate variables. In this case time and frequency are analogous to position and momentum. Specifically, the product of time resolution and frequency resolution cannot be less than 1/2.

To assess the time-bandwidth product of various windows and their Fourier transforms, the time and frequency resolutions must be quantified. This has been defined in the literature as the variance of the time window function, and the variance of its Fourier transform. Let g(t) represent a window function, and let G(ω) be the Fourier Transform of g(t). Treating the window function as a waveform, Parseval's Theorem states that the energy of the function is:

$$E = \int_{-\infty}^{\infty} \left(\left|g(t)\right|\right)^2 dt$$

in the time domain, and:

$$E = \int_{-\infty}^{\infty} \left(\left|G(\omega)\right|\right)^2 d\omega$$

in the frequency domain. The temporal center, or expected value of g(t), is:

$$t_c = \frac{1}{E}\int_{-\infty}^{\infty} t\left(\left|g(t)\right|\right)^2 dt\,,$$

and the spectral center is:

$$\omega_c = \frac{1}{2\pi E}\int_{-\infty}^{\infty} \omega\left(\left|G(\omega)\right|\right)^2 d\omega\,.$$

The time resolution is defined as the temporal width, or variance of g(t):

$$\Delta_t = \sqrt{\frac{1}{E}\int_{-\infty}^{\infty}\left(t-t_c\right)\left(\left|g(t)\right|\right)^2 dt}\,.$$

The frequency resolution is defined as the spectral width, or variance of G(ω):

$$\Delta_\omega = \sqrt{\frac{1}{2\pi E}\int_{-\infty}^{\infty}\left(\omega-\omega_c\right)\left(\left|G(\omega)\right|\right)^2 d\omega}$$

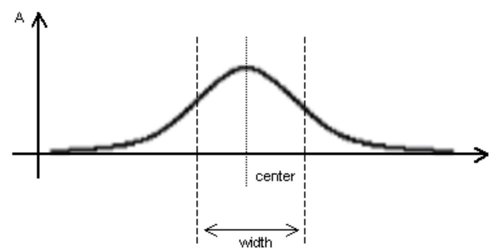

Figure 1.2.1: Center and width

Now we can state the uncertainty principle in terms of the time and frequency resolution:

$$\Delta_t\Delta_\omega \geq \frac{1}{2};$$

or, after substitution:

$$\sqrt{\frac{1}{E}\int_{-\infty}^{\infty}\left(t-t_c\right)\left(\left|g(t)\right|\right)^2 dt}\sqrt{\frac{1}{2\pi E}\int_{-\infty}^{\infty}\left(\omega-\omega_c\right)\left(\left|G(\omega)\right|\right)^2 d\omega} \geq \frac{1}{2}\,.$$

This calculation is presented in detail in [7], [8], [9], and [10]. Gabor proved that the expression above only achieves the theoretical limit of 1/2 if the window function is Gaussian [10]. Combine this result with our earlier observation that the Fourier transform of the Gaussian window is completely free of ringing, and it emerges as the clear choice for an optimal spectrogram display.

A Gaussian window is described by the expression:

$$g(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}},$$

where σ^2 is the variance of the window, which we have now defined as the nominal time resolution. It has the unusual characteristic that its Fourier transform is also a Gaussian function. In intuitive terms, it is the most selective window whose Fourier transform is completely free of ringing.

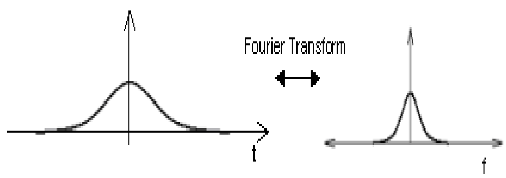

Figure 1.2.2: Gaussian Window and Its Fourier Transform

In section 2.2.4, we introduce an efficient method for achieving an arbitrarily close approximation of the Gaussian function.

2 SPECTROGRAM IMPLEMENTATION

2.1 Algorithm

The new method is similar to the WFT method, the primary distinction being that the time windowing step is accomplished by complex convolution of the complex frequency response data [5]. Implementing windowing in this fashion allows frequency-dependent window widths to be implemented, and drastically reduces the number of FFTs that must be run. Further improvements in calculation speed, and close approximation of a Gaussian window are accomplished by implementing complex smoothing as an IIR filter.

The step-by-step algorithm is:

1.) Set t_0.

The windowing that occurs in the time domain when complex smoothing is performed in the frequency domain is always centered about t=0. Therefore, the time window cannot be shifted to include different parts of the impulse response. Instead, the transfer function is shifted in time. This is accomplished by progressively deducting latency from the transfer function in the form of frequency response phase shift. The phase shift, Φ, as a function of delay, t_d, is:

$$\phi(\omega) = -\omega \cdot t_d$$

2.) Window.

The windowing operation is accomplished by convolving the complex frequency response with a non-truncated Gaussian function. This is accomplished efficiently by using an IIR approximation of the Gaussian function. This step is detailed in section 2.2.2.

3.) Display the Frequency Response Magnitude.

The frequency response represents a single column of pixels of the Spectrogram. The magnitude of the frequency response is mapped to a color or gray scale value.

4.) Increment t_0 and repeat.

2.2 Complex Convolution

The phrase, "frequency smoothing", is used to indicate that a low-pass filter has been applied to the sequence of numbers that represents the frequency response. If the low-pass filter is applied first to the sequence of real parts, then to the sequence of imaginary parts, the result is *complex smoothing*. Complex smoothing of the frequency response is equivalent to, and indistinguishable from, windowing in the time domain [5].

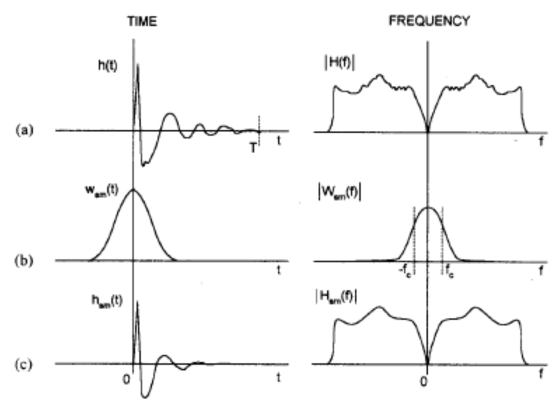

Figure 2.2.1: Windowing and convolution duality (from [5])

The typical means of implementing smoothing is to calculate a moving average, which is essentially an application of a finite impulse response (FIR) low-pass filter. A less obvious means of implementing smoothing is to employ an infinite impulse response (IIR) filter. The IIR approach has three main benefits: calculation speed, the absence of window truncation, and the ability to closely approximate a Gaussian window.

2.2.1 IIR Smoothing

An important attribute for a frequency-domain smoothing function, g(t), is that it must be even; which is to say that g(t) = g(-t). If it were odd, then it would exhibit the unfortunate behavior of changing the frequency of details in the response. A narrow peak in a frequency response should be made broader by a smoothing filter - but it should not be moved.

Strictly speaking, then, all smoothing filters must be acausal, but this is actually of no concern. These filters will be applied to complete data sets as a post-processing operation - never as part of a real-time process.

IIR filters are generally efficient for implementing minimum phase low-pass filters, which are both causal, and odd. So, it may seem unusual to consider them for use as frequency smoothing filters. However, the fact that they are only applied in post-processing means they can be run in reverse order on the frequency response data. The phase lag incurred by an increasing-frequency execution of a filter can be exactly cancelled by a decreasing-frequency execution of the same filter. The convolution of an IIR filter with its reversed-time version is always even.

Figure 2.2.1.1 shows the impulse response of a first-order lowpass filter, and figure 2.2.1.2 shows the same filter convolved with its mirror image. This illustrates how cascading forward and backward passes results in an even function. Figures 2.2.1.3 and 2.2.1.4 show the effect of cascading multiple passes of the even smoothing filter. Each pass brings the window function closer to an ideal Gaussian window.

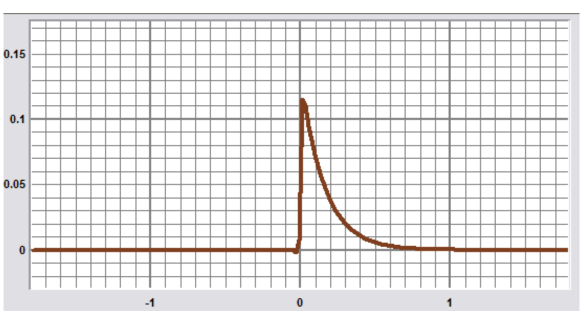

Figure 2.2.1.1: First Order Smoothing Window

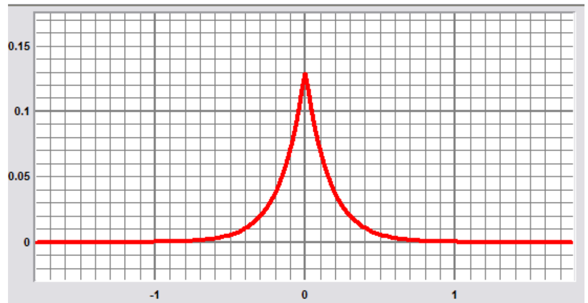

Figure 2.2.1.2: Second Order Smoothing Window (First Order Smoothing Window Convolved with Its Mirror Image)

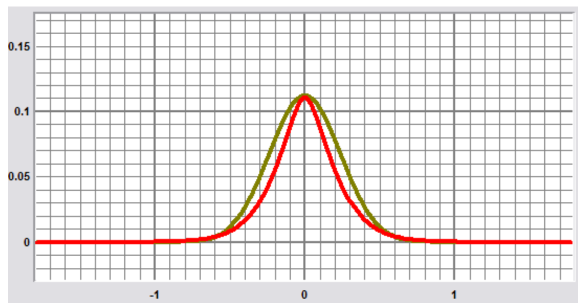

Figure 2.2.1.3: Two Cascaded Second Order Smoothing Windows, Compared to Ideal Gaussian Window

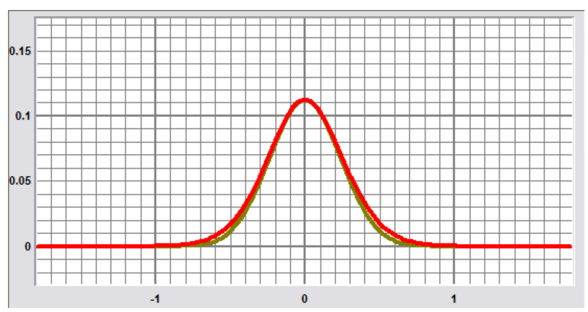

Figure 2.2.1.4: Eight Cascaded Second Order Smoothing Windows, Compared to Ideal Gaussian Window

A Gaussian window can be approximated to an arbitrary degree of precision by executing multiple cascades of this filter. While a rigorous proof of this statement would be possible, the graphical evidence should be sufficient to establish its probable truth. The approximation with 64 cascades is nearly a perfect match (see figure 2.2.1.5).

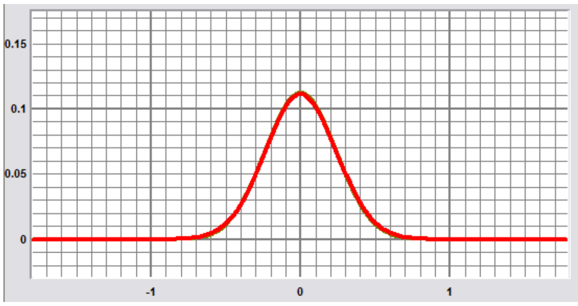

Figure 2.2.1.5: Sixty-Four Cascaded Second Order Smoothing Windows, Compared to Ideal Gaussian Window

While 64 cascades provides a near-perfect match, most of the time-frequency performance benefit can be realized with only four cascades. At this level of precision, the cost of the IIR smoothing window is equivalent to an eight-tap FIR window - but it is effective regardless of the required width. A degree of smoothing that would require a 64-tap FIR filter can be accomplished with the same eight tap IIR filter. Furthermore, the window is non-truncated; which allows its Fourier transform to be completely free of ringing.

2.2.2 Fractional Octave Smoothing

The IIR approach produces fractional-octave results if the frequency response data is logarithmically spaced. If the frequency response data is linearly spaced, fractional-octave smoothing can be accomplished by scaling the filter parameters to frequency. When applying the filter in the increasing-frequency direction, the corner frequency of the smoothing filter is increased for each successive frequency. When applying the filter in the decreasing-frequency direction, the corner frequency of the filter is decreased for each successive frequency.

3 EXAMPLE DISPLAYS

3.1 Analytic Examples

The figures in this section are examples of spectrograms of various calculated waveforms. The waveforms were selected for their interest and to demonstrate the resolution of this new display type. The effects of varying the parameters of the display are also demonstrated.

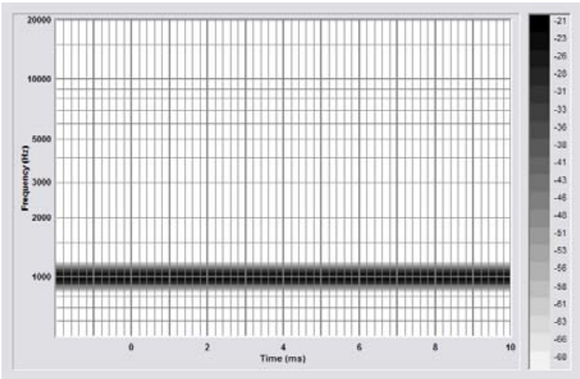

Figure 3.1.1: 1 kHz Sine Wave, Displayed with .25-Octave Resolution, 50 dB Scale

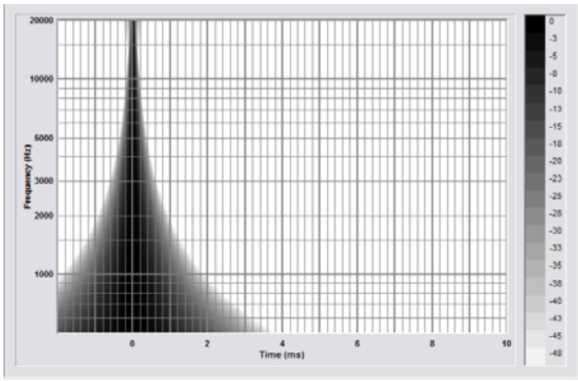

Figure 3.1.2: Pure Impulse, Displayed with 2-Octave Resolution, 50 dB Scale

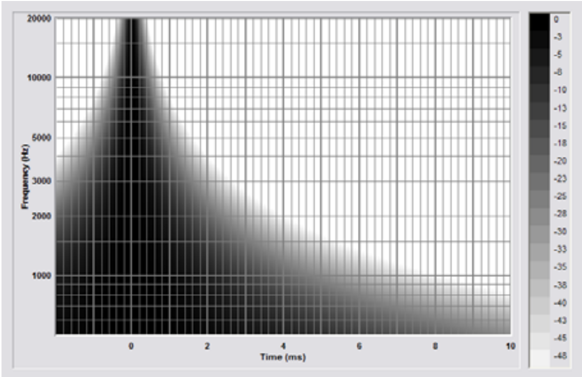

Figure 3.1.3: Pure Impulse, Displayed with 0.5-Octave Resolution, 50 dB Scale

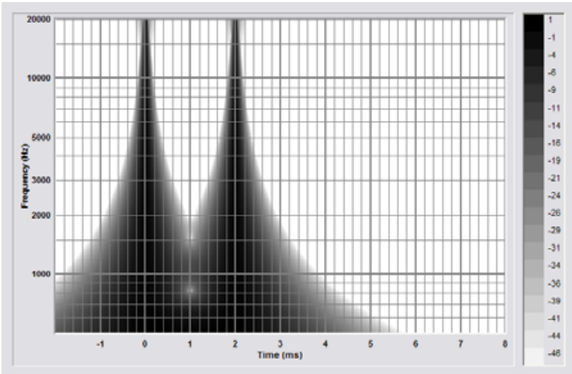

Figure 3.1.4: Two Impulses Separated by 2 ms, 2 Octave Resolution, 50 dB Scale

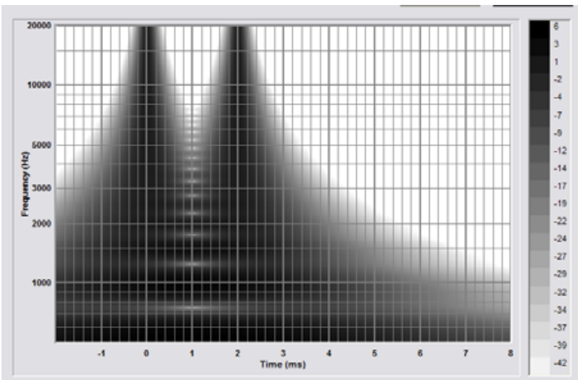

Figure 3.1.5: Two Impulses Separated by 2 ms, 0.5-Octave Resolution, 50 dB Scale

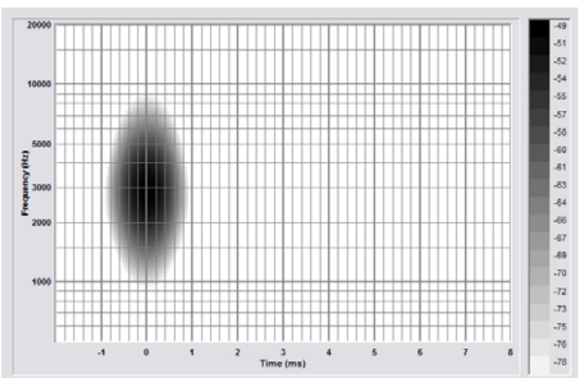

Figure 3.1.6: 3000 Hz Sine Wave, Gated by 2 ms Blackman Window, 2-Octave Resolution, 30 dB Scale

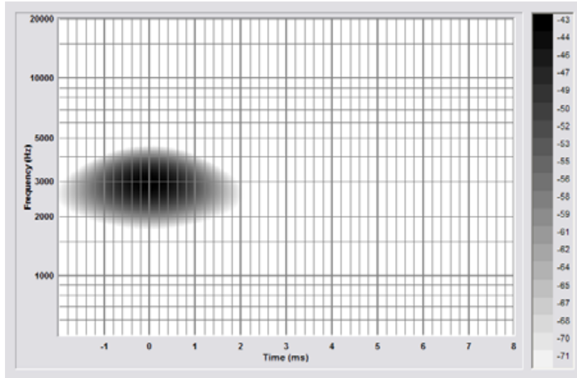

Figure 3.1.7: 3000 Hz Sine Wave, Gated by 2 ms Blackman Window, 0.5-Octave Resolution, 30 dB Scale

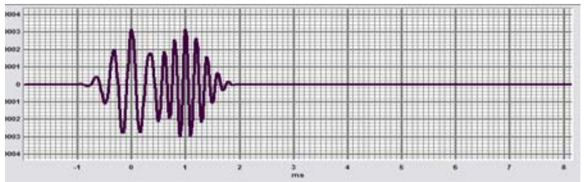

Figure 3.1.8: 3000Hz Burst summed with 5000 Hz, 1 ms delayed, Burst

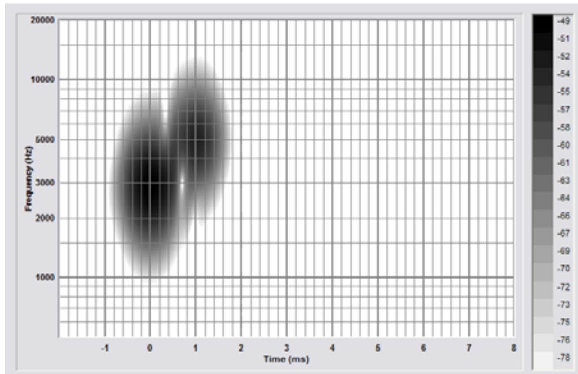

Figure 3.1.9: 3000Hz Burst summed with 5000 Hz, 1 ms delayed, Burst, 2-Octave Resolution, 30 dB Scale

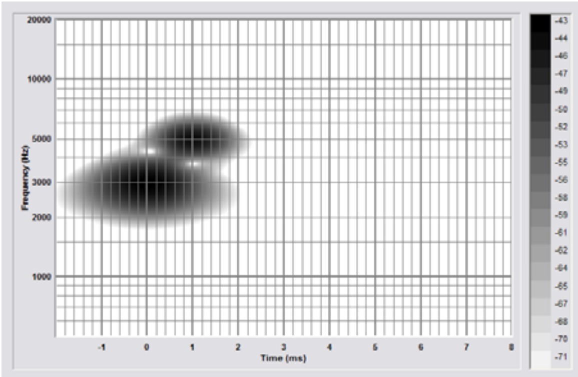

Figure 3.1.10: 3000Hz Burst summed with 5000 Hz, 1 ms delayed, Burst, 0.5-Octave Resolution, 30 dB Scale

3.2 Loudspeaker Response Examples

The example below shows the effect of changing the scale of the display. By selecting a larger (50 dB) scale, the fine details of the loudspeaker's transient decay are emphasized, while a smaller (30 dB) scale might be preferred in some cases, because it emphasizes the strong aberrations and suppresses the weak aberrations.

The transient response displayed in this example shows the combined effects of compression driver phase plug time smear, edge diffraction, horn resonance, and cone resonance. Of these mechanisms, all but edge diffraction can be improved through the use of preconditioning filters. Figure 3.3.3 shows the improvement that is possible.

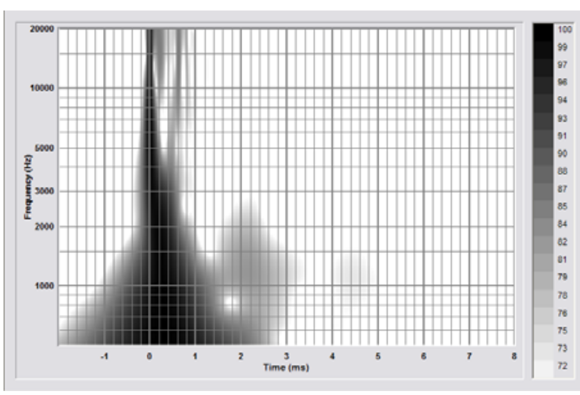

Figure 3.2.1: 2-Way Horn Loaded Loudspeaker, 2-Octave Resolution, 30 dB Scale

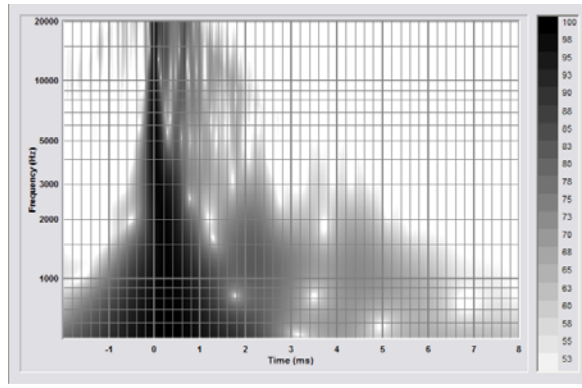

Figure 3.2.2: 2-Way Horn Loaded Loudspeaker, 2-Octave Resolution, 50 dB Scale

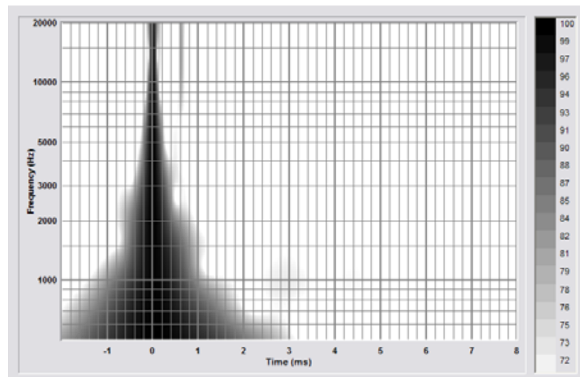

Figure 3.3.3: 2-Way Horn Loudspeaker with Preconditioning Filters, 2-Octave Resolution, 30 dB Scale

4 REFERENCES

[1] Amara Graps, "An Introduction to Wavelets," *IEEE Computational Science and Engineering*, vol. 2, num. 2 (Summer 1995).

[2] G. Evangelista and C.W. Barnes, "Discrete-Time Wavelet Transforms and Their Generalizations, " *IEEE International Symposium on Circuits & Systems*, vol. 3 (1990).

[3] P. Muller, B. Vidakovic, "Wavelets for Kids: A Tutorial Introduction," www.isye.gatech.edu/~brani/wp/**kids**A.ps (1991).

[4] Tim Edwards, "Discrete Wavelet Transforms: Theory & Implementation," qss.stanford.edu/~godfrey/wavelets/ (June 1992)

[5] P. D. Hatziantoniou and J. N. Mourjopoulos, "Generalized Fractional-Octave Smoothing of Audio

and Acosutic Responses, " *J. Audio Eng. Soc*, vol. 48, no. 4 (April 2000).

[6] Jon Claerbout, "Time-Frequency Resolution," http://sep.stanford.edu/sep/prof/fgdp/c4/paper_html.node2.html

[7] Charles K. Chui, *An Introduction to Wavelets*, Academic Press, Boston, 1992, pp 54-59.

[8] Christopher Lang and Kyle Forinash, "Time-Frequency Analysis with the Continuous Wavelet Transform, " *Am. J. Phys*, vol. 66, no. 9 (Sept 1998).

[9] Phil Schniter, "Time-Frequency Uncertainty Principle," cnx.rice.edi/content/m10416/latest (Jan 2003)

[10] Shie Qian and Dapang Chen, "Gabor Expansion, Inverse Sampled STFT" http://zone.ni.com/devzone/conceptd.nsf/webmain/826B40795B46327A862568F80064CD86 (1996)

[11] Christopher Lang and Kyle Forinash, "Time-Frequency Analysis with the Continuous Wavelet Transform, " *Am. J. Phys*, vol. 66, no. 9 (Sept 1998).

[12] Phil Schniter, "Time-Frequency Uncertainty Principle," cnx.rice.edi/content/m10416/latest (Jan 2003)

[13] Shie Qian and Dapang Chen, "Gabor Expansion, Inverse Sampled STFT" *http://zone.ni.com/devzone/conceptd.nsf/webmain/826B40795B46327A862568F80064CD86* (1996)

[14] Panagiotis D. Hatziantoniou and John N Mourjopoulos, "Generalized Fractional0-Octave Smoothing of Audio and Acosutic Responses, " *J. Audio Eng. Soc*, vol. 48, no. 4 (April 2000).

[15] R. C. Heyser, "Determination of Loudspeaker Signal Arrival Times, Part I," *J. Audio Eng. Soc.*, vol. 19, pp. 734-743 (Oct 1971).

[16] Richard Quinnell, "Designing Digital Filters," www.techonline.com/community/ed_resource/feauture_article/26649.

[17] C. Radcliffe, "Recursive, Infinite Impulse (IIR) Digital Filters," egr.msu.edu/classes/me855/radcliff/Notes/**Recursive**_filter_notes.pdf (April 2003).

[18] Emmanuel C. Ifeachor and Barrie W. Jervis, Digital Signal Processing: A Practical Approach, 2nd ed., pp. 468-471 (Pearson Education Limited, Essex, England, 2002).