Return to Main

Objectives

Short-Time Measurements:
Energy
Sums and Filters
Examples

Windows:
Spectrograms
Rectangular Windows
Frequency Response
Popular Windows
Recursive
Control Systems

On-Line Resources:
Signal Modeling
The HTK Book
Windows

LECTURE 11: TEMPORAL ANALYSIS

- Objectives:
  - Understand the relationship between sums and filters
  - Understand the relationship between temporal resolution and frequency resolution
  - Introduce common window functions
  - Explain their use in speech processing
  - Understand how we compute our first recognition feature: energy

One of the best explanations of this material can be found in:

> L.R. Rabiner and B.W. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, Upper Saddle River, New Jersey, USA, ISBN: 0-13-015157-2, 1993.

This textbook is unfortunately out of print. Another excellent reference is:

> J. Deller, et. al., *Discrete-Time Processing of Speech Signals*, MacMillan Publishing Co., ISBN: 0-7803-5386-2, 2000.

ECE 8463: FUNDAMENTALS OF SPEECH RECOGNITION

Professor Joseph Picone
Department of Electrical and Computer Engineering
Mississippi State University

email: picone@isip.msstate.edu
phone/fax: 601-325-3149; office: 413 Simrall
URL: http://www.isip.msstate.edu/resources/courses/ece_8463

Modern speech understanding systems merge interdisciplinary technologies from Signal Processing, Pattern Recognition, Natural Language, and Linguistics into a unified statistical framework. These systems, which have applications in a wide range of signal processing problems, represent a revolution in Digital Signal Processing (DSP). Once a field dominated by vector-oriented processors and linear algebra-based mathematics, the current generation of DSP-based systems rely on sophisticated statistical models implemented using a complex software paradigm. Such systems are now capable of understanding continuous speech input for vocabularies of hundreds of thousands of words in operational environments.

In this course, we will explore the core components of modern statistically-based speech recognition systems. We will view speech recognition problem in terms of three tasks: signal modeling, network searching, and language understanding. We will conclude our discussion with an overview of state-of-the-art systems, and a review of available resources to support further research and technology development.

Tar files containing a compilation of all the notes are available. However, these files are large and will require a substantial amount of time to download. A tar file of the html version of the notes is available here. These were generated using wget:

```
wget -np -k -m http://www.isip.msstate.edu/publications/courses/ece_8463/lectures/current
```

A pdf file containing the entire set of lecture notes is available here. These were generated using Adobe Acrobat.

Questions or comments about the material presented here can be directed to help@isip.msstate.edu.

LECTURE 11: TEMPORAL ANALYSIS

- Objectives:
  - Understand the relationship between sums and filters
  - Understand the relationship between temporal resolution and frequency resolution
  - Introduce common window functions
  - Explain their use in speech processing
  - Understand how we compute our first recognition feature: energy

One of the best explanations of this material can be found in:

> L.R. Rabiner and B.W. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, Upper Saddle River, New Jersey, USA, ISBN: 0-13-015157-2, 1993.

This textbook is unfortunately out of print. Another excellent reference is:

> J. Deller, et. al., *Discrete-Time Processing of Speech Signals*, MacMillan Publishing Co., ISBN: 0-7803-5386-2, 2000.

ENERGY AND POWER

Energy:

$$E \equiv \sum_{n=-\infty}^{\infty} |x(n)|^2$$

Average Power:

$$P = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} |x(n)|^2$$

Finite Energy:

$$E_N = \sum_{n=-N}^{N} |x(n)|^2$$

$$E = \lim_{N \to \infty} E_N$$

$$P = \lim_{N \to \infty} \frac{1}{2N+1} E_N$$

Comments:

(1) If a signal's energy is finite, $P = 0$.

(2) If a signal's energy is infinite, its power may or may not be zero.

(3) RMS value is the square root of the power.

Examples:

(1) The average power of a sinewave is $\frac{A^2}{2}$.

(2) What does the following compute?

$$E(n) = E(n-1) + \alpha x^2(n)$$

FINITE SUMS AND FILTERS

- Consider a simplified equation for energy:

$$E = \sum_{n=0}^{N-1} |x(n)|^2 \qquad n = 0, 1, \ldots$$

- We can write this as a digital filter:

$$\begin{aligned}
x(n) &= s^2(n) \\
E(n) &= x(n-(N-1)) + x(n-(N-2)) + \ldots + x(n) \\
&= \sum_{k=0}^{N-1} |x(n-k)|^2 \\
H(z) &= 1 + z^{-1} + z^{-2} + \ldots + z^{-(N-1)}
\end{aligned}$$

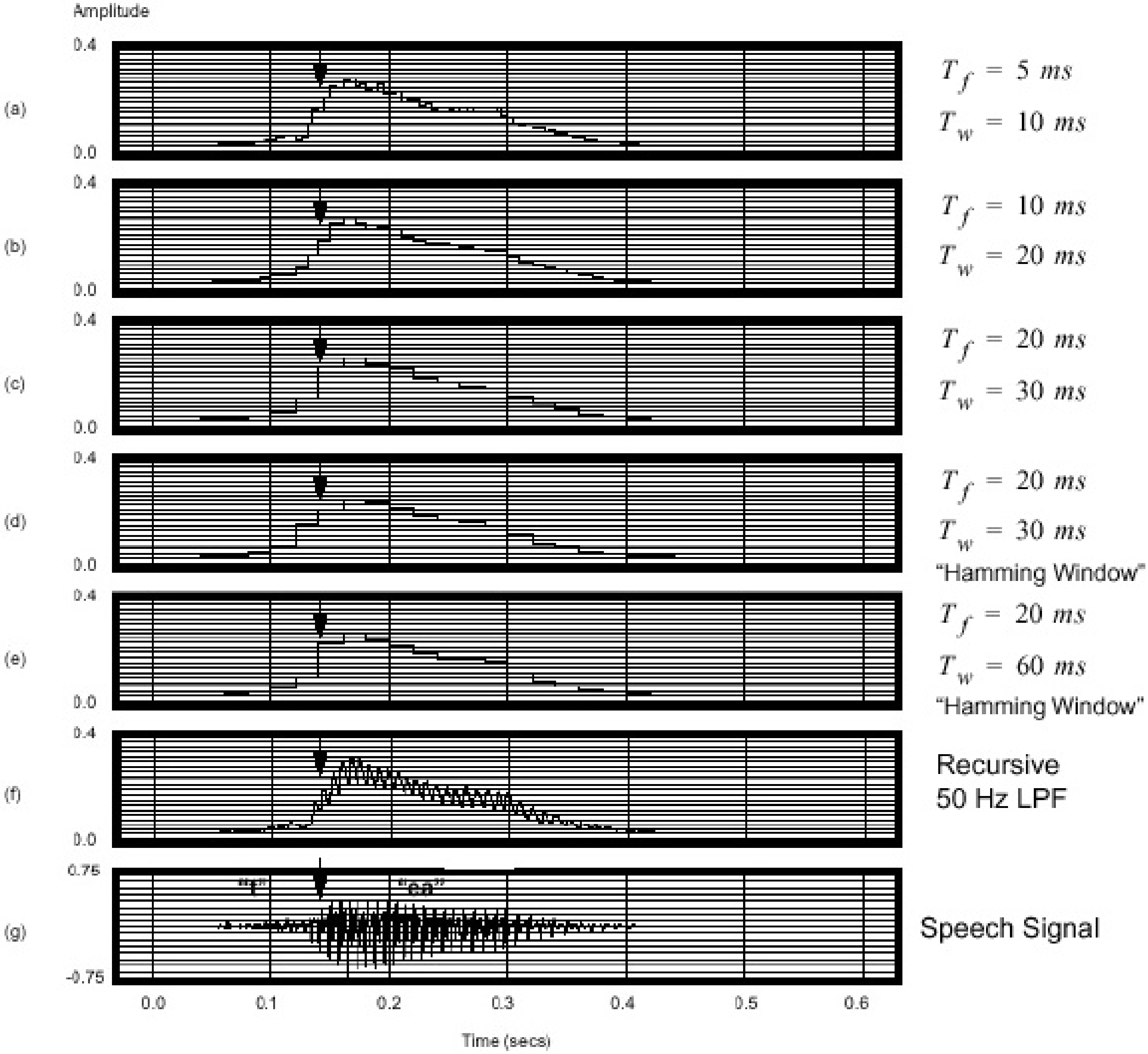
- What is the frequency response of this filter? (Hint: FIR)

- Are there other ways we can implement such a filter? (Hint: IIR)

  Consider these three approaches:

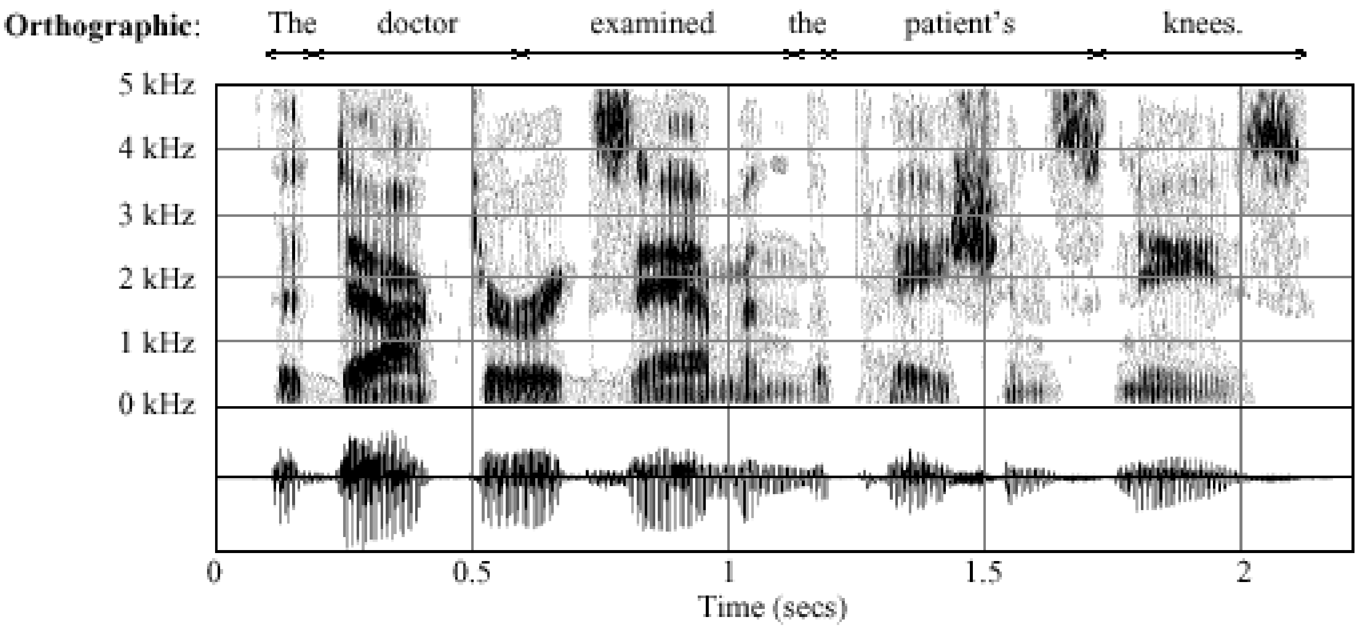
$$\begin{aligned}
E(n) &= E(n-1) - x(n-(N-1)) + x(n) \\
E(n) &= \alpha E(n-1) + x(n) \\
E(n) &= \alpha E(n-1) + \beta E(n-2) + x(n)
\end{aligned}$$

EXAMPLES OF ENERGY COMPUTATIONS

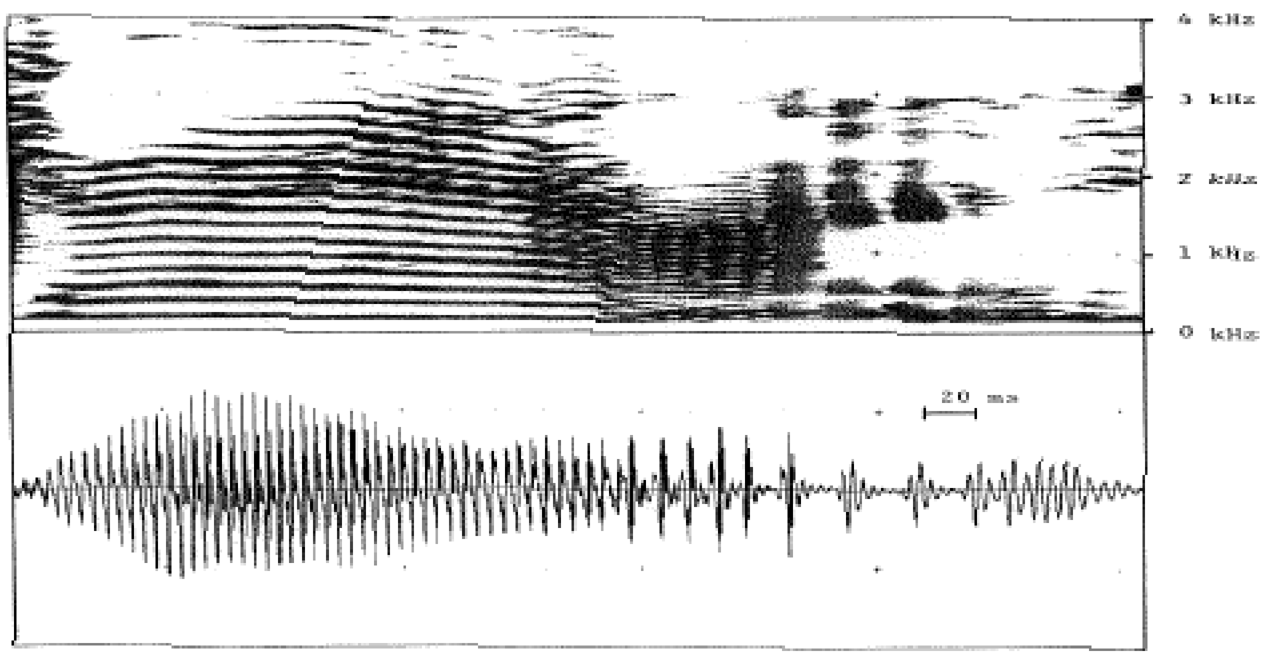


WHAT DOES A SPEECH SIGNAL LOOK LIKE?

Standard wideband spectrogram ($f_s = 10\ kHz$, $T_w = 6\ ms$):

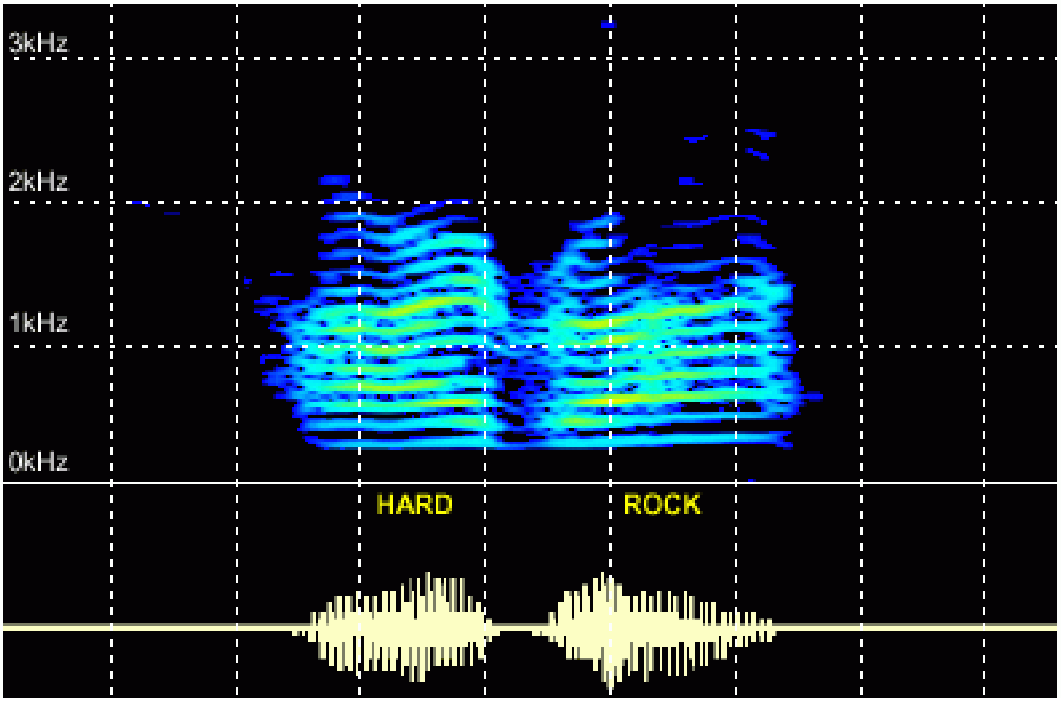


Narrowband Spectrogram ($f_s = 8\ kHz$, $T_w = 30\ ms$):

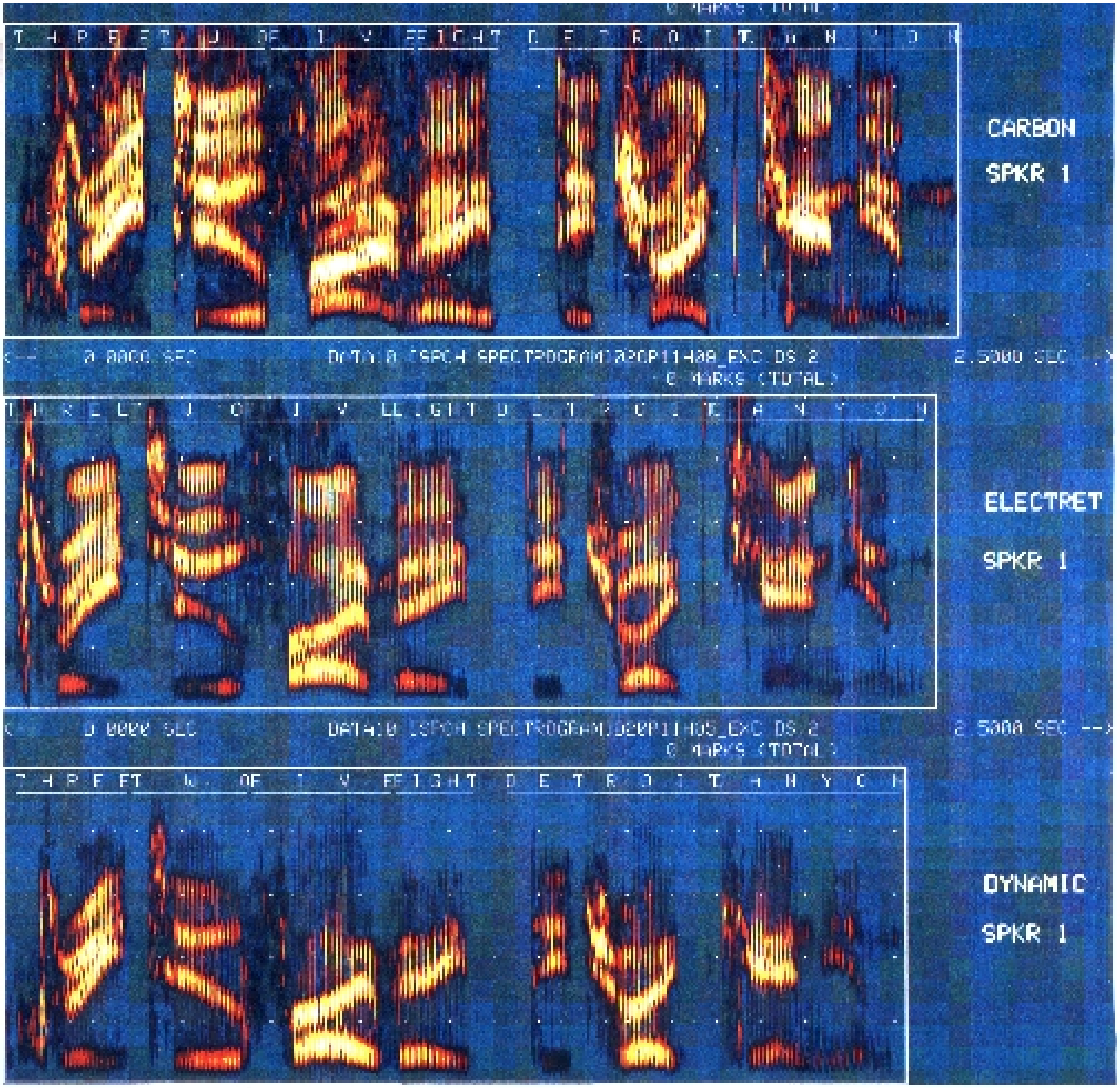


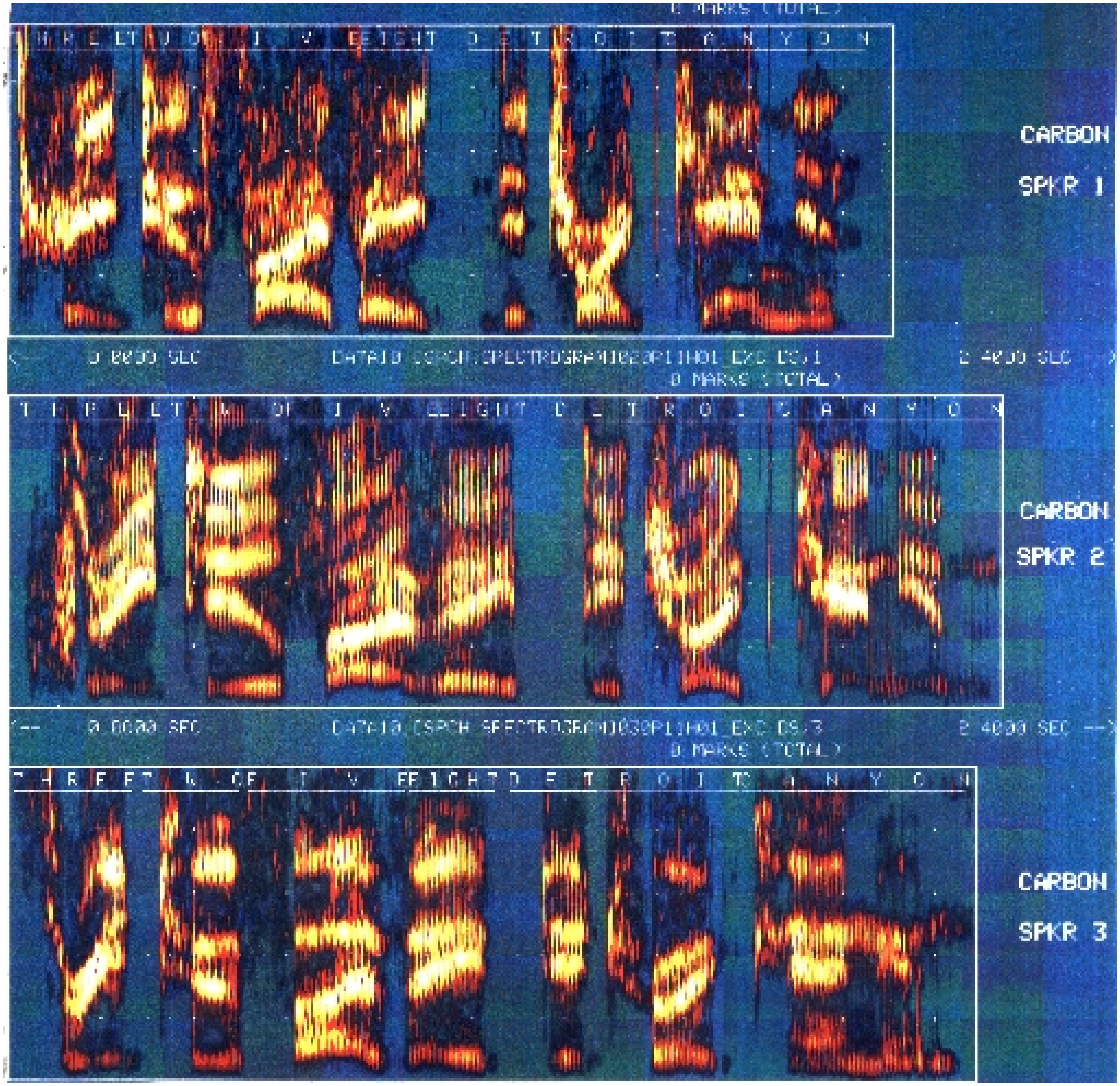
"Drown" (female)

- We often prefer to view a spectrogram using a color visualization in which spectral log magnitude is mapped to "temperature" (the color that emanates from a steel bar when it is heated):



- Here are more examples of color spectrograms using the ever-popular Texas Instruments color map:


CARBON
SPKR 1
ELECTRET
SPKR 1
DYNAMIC
SPKR 1
0.0000 SEC
2.5000 SEC


CARBON
SPKR 1
CARBON
SPKR 2
CARBON
SPKR 3

RECANGULAR WINDOWS

Let $\{x(n)\}$ denote a sequence to be analyzed. Let's limit the duration of $\{x(n)\}$ to L samples:

$$\hat{x}(n) = x(n)w(n)$$

where $w(n)$ is a rectangular window and is defined as

$$w(n) = \begin{cases} 1, & 0 \le n \le L-1 \\ 0, & otherwise \end{cases}.$$

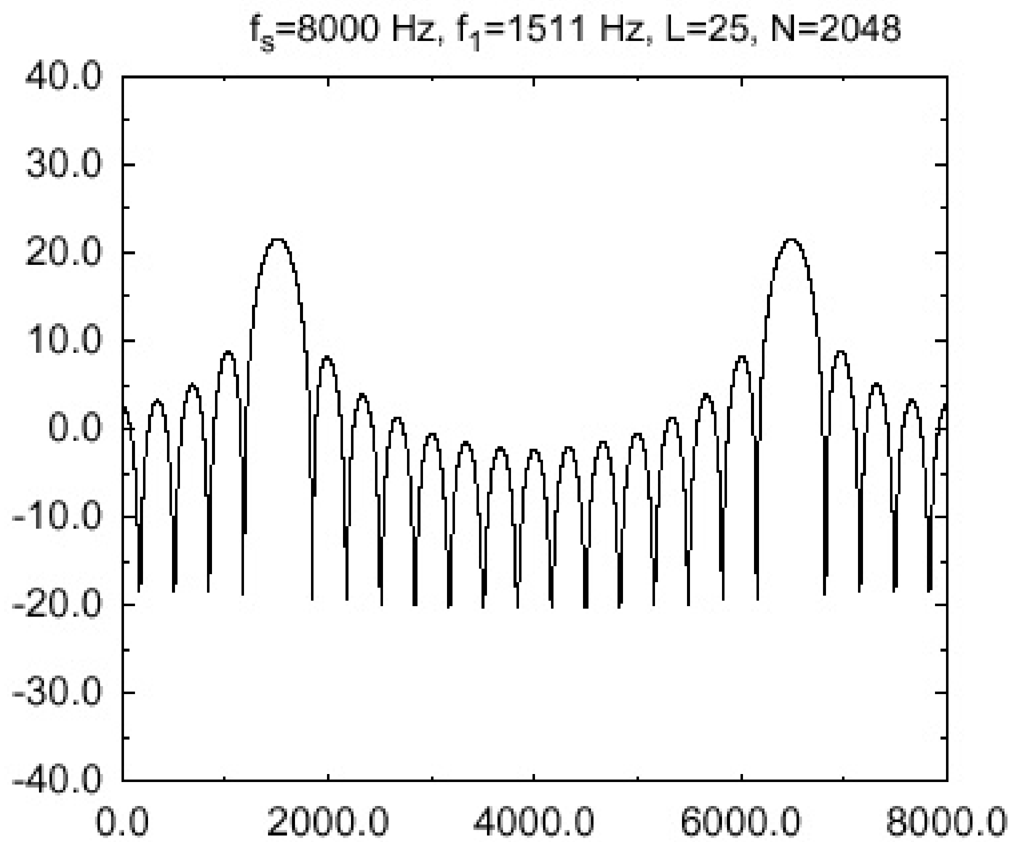
The Fourier transform of $w(n)$ is given by:

$$W(\omega) = \frac{\sin(\omega(L/2))}{\sin(\omega/2)} e^{-j\omega((L-1)/2)}$$

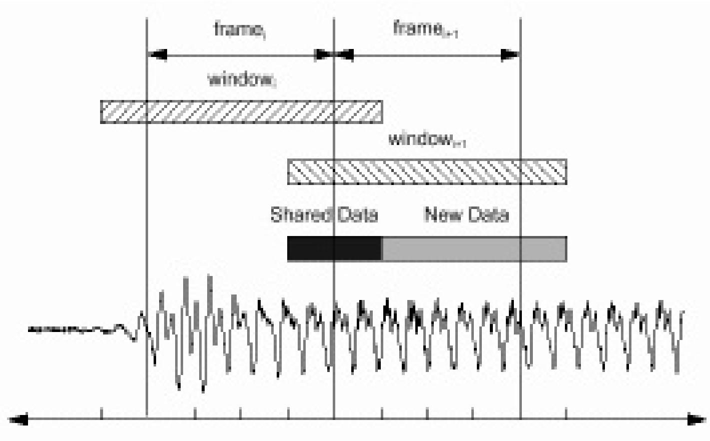
The transform of $\hat{x}(n)$ is given by:

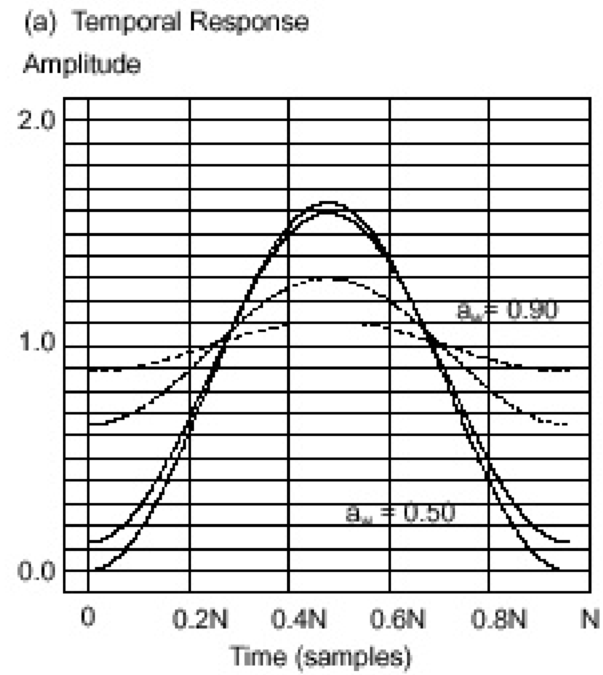
$$\hat{X}(\omega) = \frac{1}{2}[W(\omega - \omega_o) + W(\omega + \omega_o)].$$

This introduces frequency domain aliasing (the so-called picket fence effect):

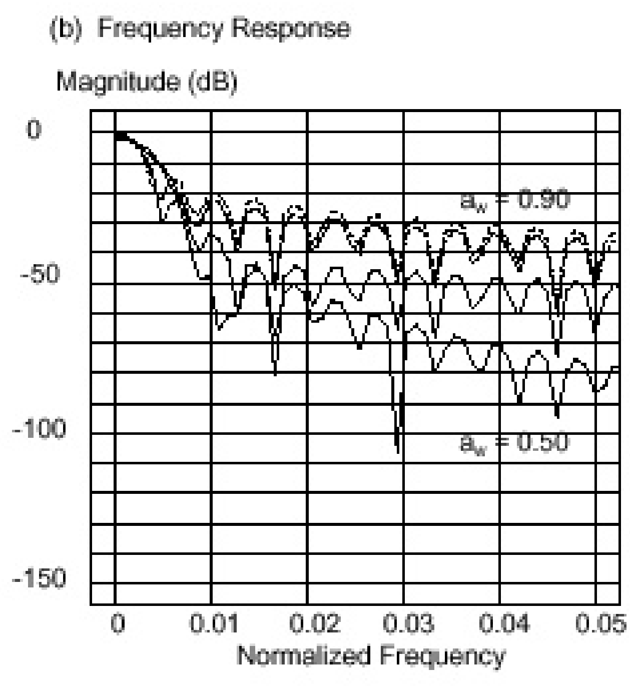


TEMPORAL/FREQUENCY RESPONSE





$$\%Overlap = \frac{(T_w - T_f)}{T_w} \times 100\%$$



POPULAR WINDOW FUNCTIONS

1. Rectangular:

$$w(k) = \begin{cases} 1, & |k| \le N \\ 0, & otherwise \end{cases}$$

2. Generalized Hanning:

$$w_H(k) = w(k)\left[\alpha + (1-\alpha)\cos\left(\frac{2\pi}{N}k\right)\right] \qquad 0 < \alpha < 1$$

$$\alpha = 0.54, \qquad \textit{Hamming window}$$
$$\alpha = 0.50, \qquad \textit{Hanning window}$$

3. Bartlett

$$w_B(k) = w(k)\left[1 - \frac{|k|}{N+1}\right]$$

4. Kaiser

$$w_K(k) = w(k)I_0\left(\alpha\sqrt{1 - \frac{K^2}{N}}\right) / I_0(\alpha)$$

5. Chebyshev:

$$w_N(k) = 2(x_0^2 - 1)w_{N-1}(k) + x_0^2[w_{N-1}(k-1) + w_{N-1}(k+1)] - w_{N-2}(k)$$

6. Gaussian

$$w_G(k) = \begin{cases} \exp\left[-\frac{1}{2}k^2\tan^2\left(\frac{\theta_0}{2}\right)\right] & |k| < N \\ w_G(N-1) / \left[2N\sin^2\left(\frac{\theta_0}{2}\right)\right] & |k| < N \\ 0 & |k| > N \end{cases}$$

There are many others. The most important characteristics are the width of the main lobe and the attenuation in the stop-band (height of highest sidelobe). The Hamming window is used quite extensively.

Define the short-term estimate of the power as:

$$P(n) = \frac{1}{N_s} \sum_{m=0}^{N_s-1} \left(w(m)s(n - \frac{N_s}{2} + m) \right)^2$$

We can view the above operation as a moving-average filter applied to the sequence $s^2(n)$.

This can be computed recursively using a linear constant-coefficient difference equation:

$$P(n) = -\sum_{i=1}^{N_a} a_{pw}(i)P(n-i) + \sum_{j=1}^{N_b} b_{pw}(j)s^2(n-j)$$

Common forms of this general equation are:

$P(n) = \alpha P(n-1) + s^2(n)$ (Leaky Integrator)

$P(n) = \alpha P(n-1) + (1-\alpha)s^2(n)$ (First-order weighted average)

$P(n) = \alpha P(n-1) + \beta P(n-2) + s^2(n) + \gamma s^2(n-1)$ (2^{nd}-order Integrator)

Of course, these are nothing more than various types of low-pass filters, or adaptive controllers. How do we compute the constants for these equations?

In what other applications have we seen such filters?

The first-order systems can be related to physical quantities by observing that the system consists of one real pole:

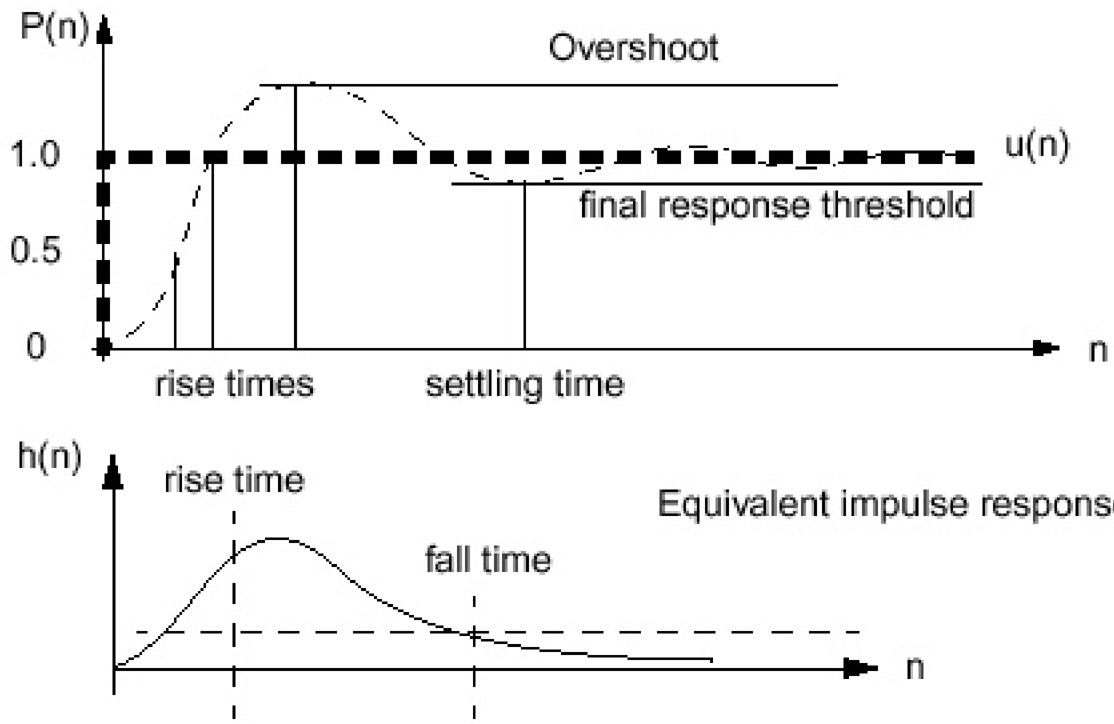
$$H(z) = \frac{1}{1 - \alpha z^{-1}}$$

α can be defined in terms of the bandwidth of the pole.

For second-order systems, we have a number of alternatives. Recall that a second-order system can consist of at most one zero and one pole and their complex conjugates. Classical filter design algorithms can be used to design the filter in terms of a bandwidth and an attenuation.

An alternate approach is to design the system in terms of its unit-step response:

There are many forms of such controllers (often known as



servo-controllers). One very interesting family of such systems are those that correct to the velocity and acceleration of the input. All such systems can be implemented as a digital filter.

Index of /publications/journals/ieee_proceedings/1993/signal_modeling

	Name	Last modified	Size	Description
	Parent Directory	01-Jan-1999 12:07	-	
	paper_v2.pdf	22-Jun-1999 16:14	452k	

Advanced DSP

FFT windows

Index

This is a module of the BORES Signal Processing advanced DSP course - FFT windows.

To follow this course module properly, you should be familiar with the basic ideas of DSP which are introduced in the earlier course:

- Introduction to DSP

and specifically with the section on windowing

- Introduction to FFT window functions

This course module covers the following subjects:

The complete course - Introduction to DSP - is presented regularly as a one day, 'hands on' workshop where delegates use DSP hardware and software to complete exercises intended to help in understanding the concepts which are introduced.

Go to first page

Go to end of module

Go back to Advanced courses index

Go back to Introductory course

Go back to BORES Home Page

BORES Signal processing

Last updated: 29th January 1998

File http://www.bores.com/courses/advanced/windows/index.htm

Fordwater, Pond Road, Woking, Surrey GU22 0JZ
Telephone: 01483 740138 fax: 01483 740136 email: bores@bores.com Web: http://www.bores.com