Music & Entertainment Industry Studies

College of Arts & Media

University of Colorado at Denver

© Copyright 2008 John Strawn

MSRA 5500 - 002

MUS 4500 - 002

Topics in Professional Audio:

Audio (Data) Compression---

Making "MP-3" Sound Great

March 14-16 2008

John Strawn, Ph.D.

S Systems Inc.

15 Willow Avenue

Larkspur California 94939

tel. 415 927 8856

email jstrawn@s-systems-inc.com

http://www.s-systems-inc.com

MSRA 5500 - 002
MUS 4500 - 002

March 14, Session 1, Part 1

The Challenge

What do you want to cover?

- Lossless compression: how is it lossless?
- Newer surround formats, Dolby True-HD, DTS Master Audio
- effects of kbps, how does that affect audio. At what point does it still have an effect?
- Is compression here to stay? On its way out due to storage, Blu-ray?

What do you want to cover?

- General practices: how to mix for an MP3 release
- How can fara can compression go? How much compression can you achieve while having great quality?
- What are more widely used formats, what is on its way out, what's on its way in?

What do you want to cover?

- Calculations for final results?
- Encoding / Decoding?
- History of coding; how did this come about?
- Perceptual coding

What do you want to cover?

- Audio streaming, starting from the beginning
- How important is the algorithm and the encoder to get the same results? Does LAME sound better than other encoders?

What I plan to cover so far (1)

- Fundamentals of compression
- Building blocks of perceptual compression
  - Windowing
  - Transform
  - Simultaneous masking
  - Temporal masking
  - Stereo coding
  - Basic structure of encoder, decoder
- Lossless compression

What I plan to cover so far (2)

- Some main families of audio compression.
- Evaluating and comparing techniques.
- Where compression is used and misused.
- Unwanted artifacts.
  - LOTS of listening!
  - How to minimize or avoid artifacts.
- What else you want to cover, see above

What we may not cover

- Discrete cosine transform
- QMF filter banks
- ...

Schedule

Friday	9:00-12:00
Friday	1:00-5:00
(extra:	5:15)
Saturday	9:00-12:00
Saturday	1:00-5:00
Sunday	9:00-12:00

How are you graded?

- No written homework, no test.
- Sign attendance sheets in classroom.
- Sign attendance sheets in listening rooms.
- Participate here.
- Written notes in your handouts.
- CU Denver staff will examine above, determine grade, especially based on your notes in handouts.

If you are willing ...

- Volunteer *your* recording for me to process + us to listen to Sunday a.m.
- Bring your laptop Saturday for hands-on examination of a recording before + after MP3 (Cool Edit Pro? Adobe Audition? ProTools?)
- *40 Year Old Virgin* --- who can lend a copy of DVD?

What we will cover

- Compression: not in a vacuum
- What problems does compression solve? (Why bother?)
- Forerunners to perceptual compression

Where compression used?

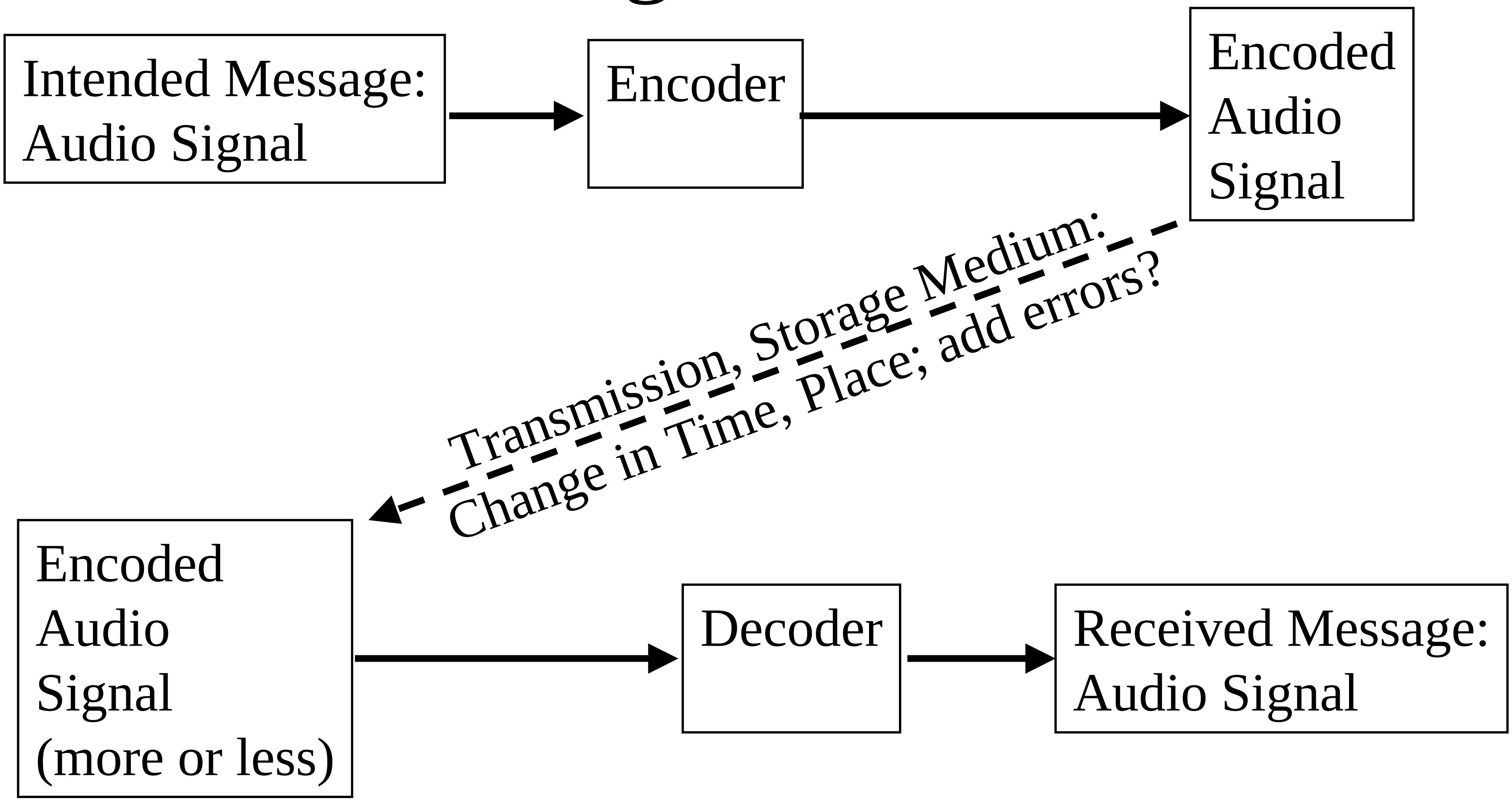
The Politics of Production

- Traditional model
- Model in transition

Model of Communication

Based on Colin Cherry: *On Human Communication*

Coding in General



"Codec": [en]**COD**er + **DEC**oder

What we will cover

- Compression: not in a vacuum
- What problems does compression solve? (Why bother?)
- What was tried that didn't work well?

1989: statement of the problem

In transmitting [digital] musical signals, what level of quality can we achieve at what data rates and using which techniques?

Brandenburg, Karlheinz. *A contribution to the procedures for, and the evaluation of quality of, high-quality musical coding*. Ph.D. Dissertation, University of Erlangen, 1989, p. 2. My translation.

Consumer Digital Audio: CD

1977 Mitsubishi, Hitachi & Sony show digital audio disc prototypes at the Tokyo Audio Fair

1981 Compact Disc Standard agreed

1982 Compact Disc Technology is introduced to Europe and Japan in the fall.

1983 Compact Disc Technology is introduced in the United States in the spring

CD Audio Data Rates

44,100 16-bit samples / sec x

16 bits / sample x

2 channels =

1,411,200 bits per second, or

1,411.2 kbps (kilobits per second, kbit/s), or

1.4 Mbps (megabits per second, Mbit/s)

RealNetworks Data Rates (stereo)

Transmission medium	Max	How far off from real time?
28.8 kbps modem	20 kbps	71
56 kbps modem	32 kbps	44
112 kbps dual ISDN	64 kbps	22
Corporate LAN	132 kbps	11
256 kbps DSL/cable modem	176 kbps	8
512 kbps DSL/cable modem	352 kbps	4

Derived from: RealNetworks, http://service.real.com/help/library/guides/production8/htmfiles/audio.htm, retrieved 4 Feb 2008

CD Audio Storage Requirements

- 74 1/2 minutes x
  60 sec /minute x
  1.4 Megabits per second /
  8 bits/byte
  = about 3/4 Gigabyte per CD
- Today’s computers ship with: 320 Gbyte drive: over 200 audio CDs?
- Ipod ships with 80 Gbyte (“stores up to 40,000 songs”)

Storage Capacity: 4-min songs

Bit rate	32 MB RAM	64 MB RAM	CD (650 MB)
64 kbps	17	34	355
80 kbps	14	28	284
96 kbps	11	22	237
128 kbps	9	18	177
160 kbps	7	14	142
192 kbps	6	12	118
256 kbps	4	8	89

From: http://www.teamcombooks.com/mp3handbook/16.htm

1989: The challenge

For a series of possible applications, the same distortion-free sound quality [as on the CD] should be achieved without using the transmission bandwidth or the storage capacity of the CD. Examples of this would include: future digital terrestrial broadcast (DAB, Digital Audio Broadcasting), sound for digital video recorders, tape recording devices with stationary tape heads (S-DAT), and many others.

Brandenburg, Karlheinz. *A contribution to the procedures for, and the evaluation of quality of, high-quality musical coding*. Ph.D. Dissertation, University of Erlangen, 1989, p. 2. My translation.

Listening

- Pick a listening room: 285D, 285F, 285H, 285J. All rooms have same tracks.
- Listen to 3 “Tallis” tracks in Tallis project, compare, contrast. [not Tallis_PVOC]
- Take notes and discuss among yourselves:
  - What do you hear?
  - How are they different?
  - What do you like?
  - What don’t you like?
- Back here at 10:15 to discuss. (15 minutes)

Notes on Tallis Sound Examples

Thomas Tallis. “God grant we Grace,” from
Spem in Alium. Carlton 5-030366-9527.

Tallis: Discussion of what you heard

Tallis_1

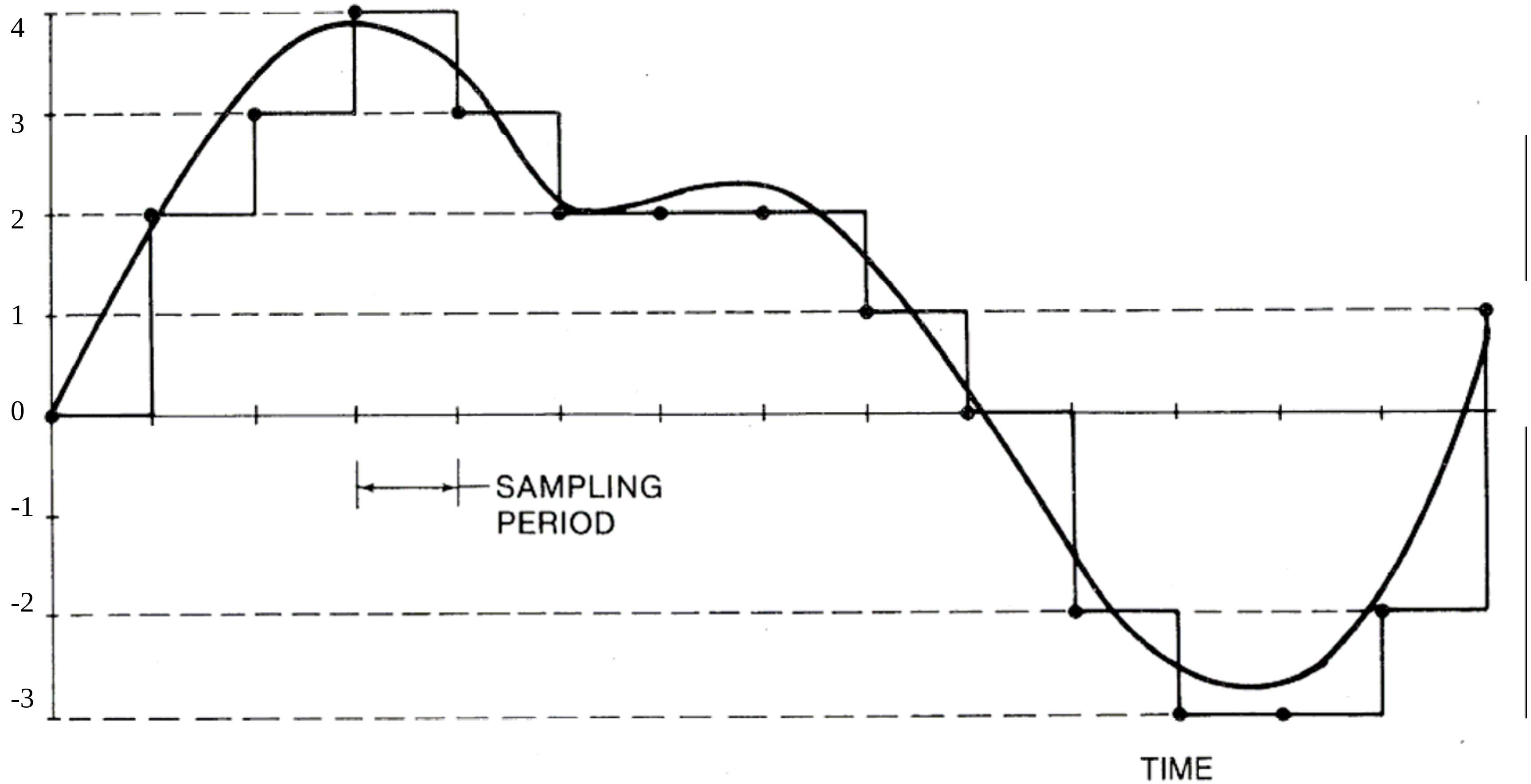
Tallis_2

Tallis_3

What we will cover

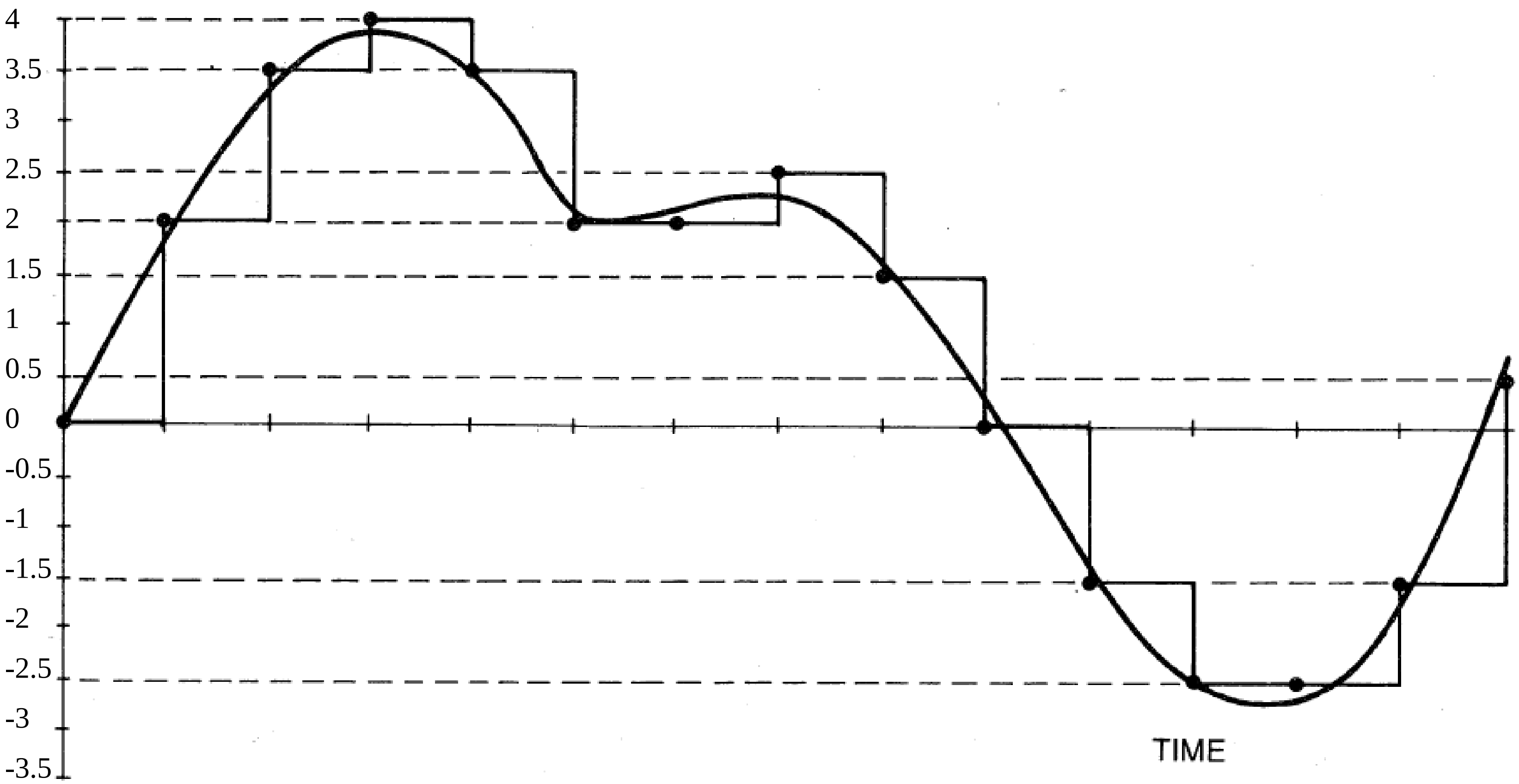
- Compression: not in a vacuum
- What problems does compression solve? (Why bother?)
- Forerunners to perceptual compression

Coarser Quantization



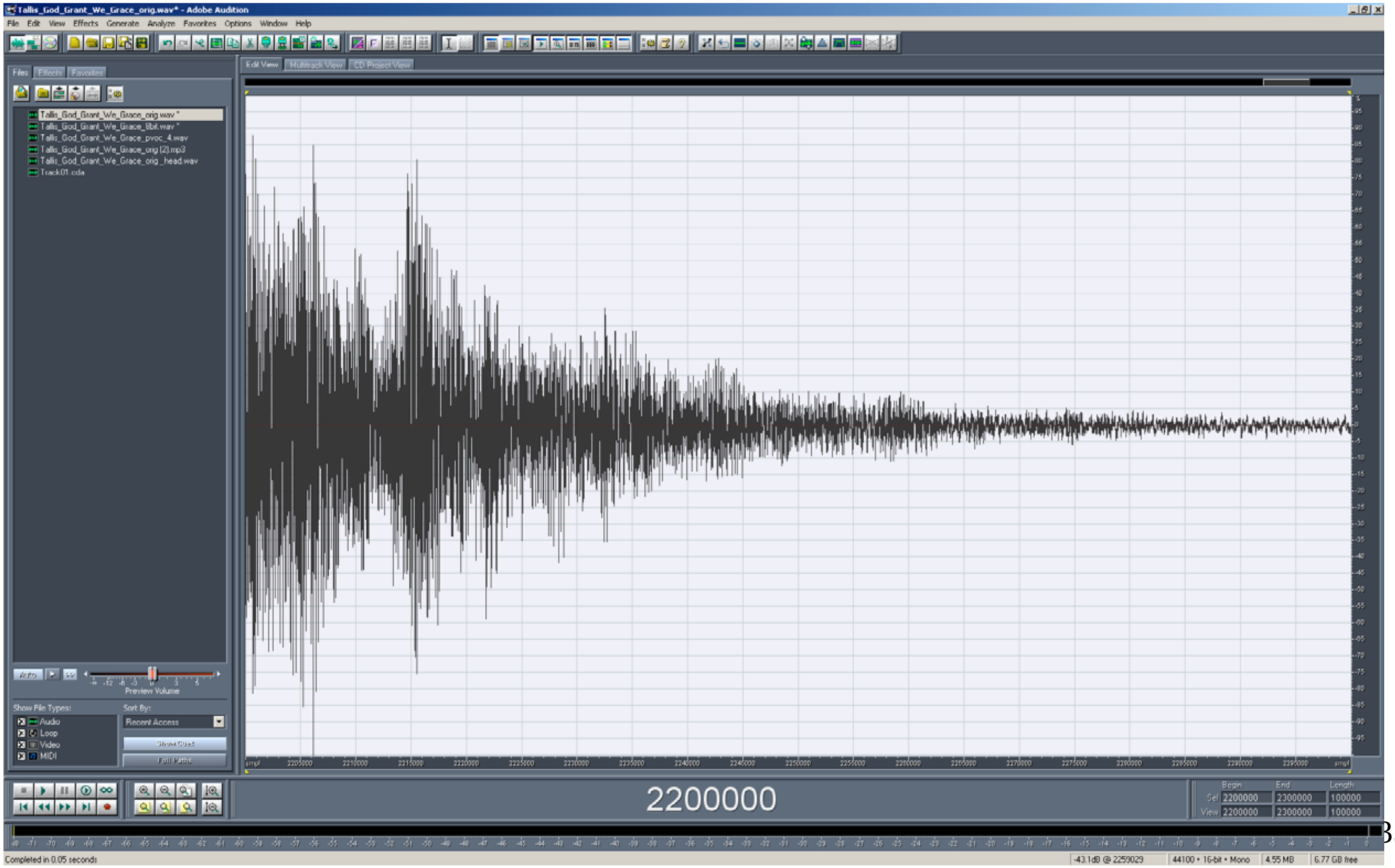
From: Ken Pohlmann, *Principles of Digital Audio*, Sams

Finer Quantization

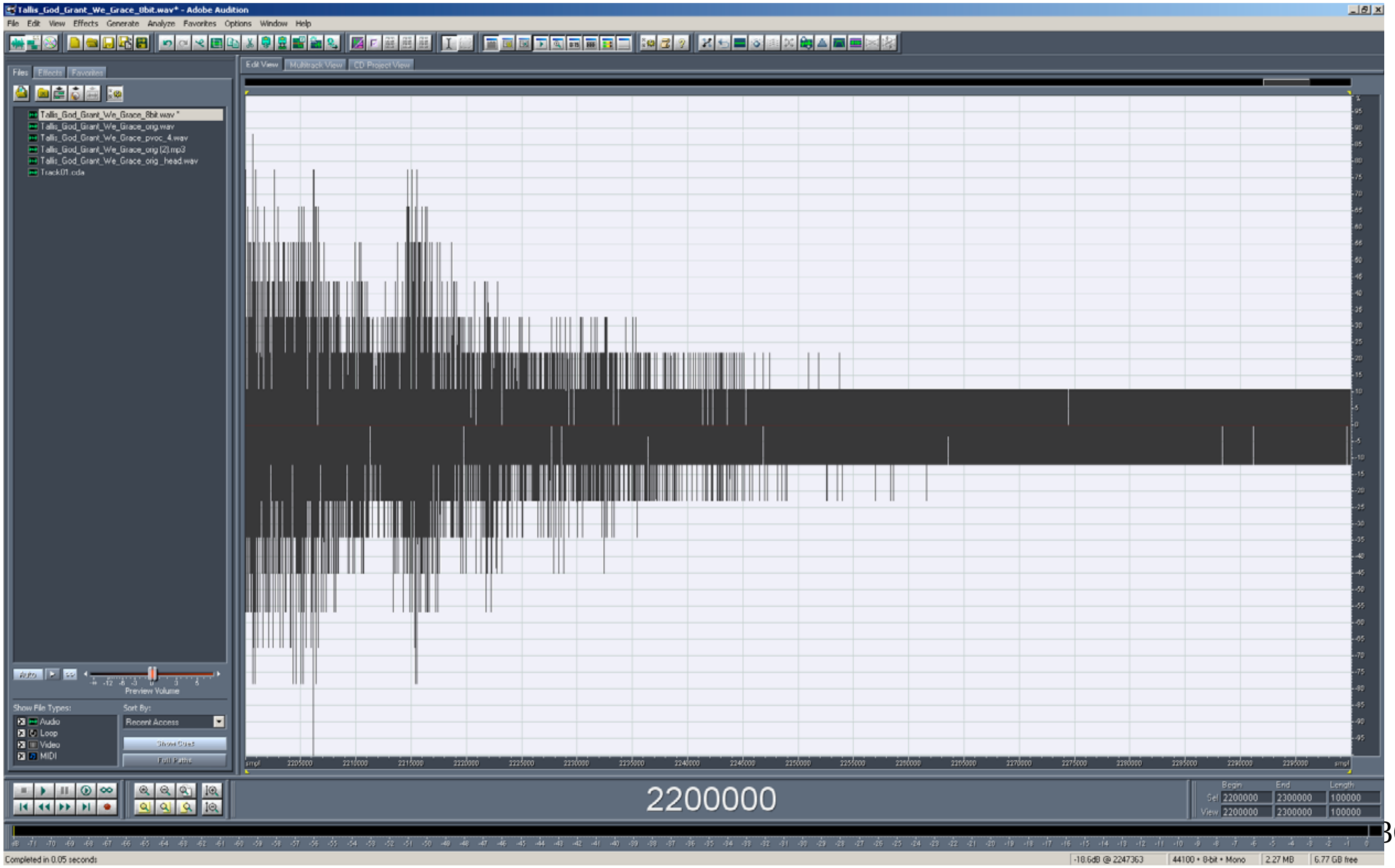


From: Ken Pohlmann, *Principles of Digital Audio*, Sams

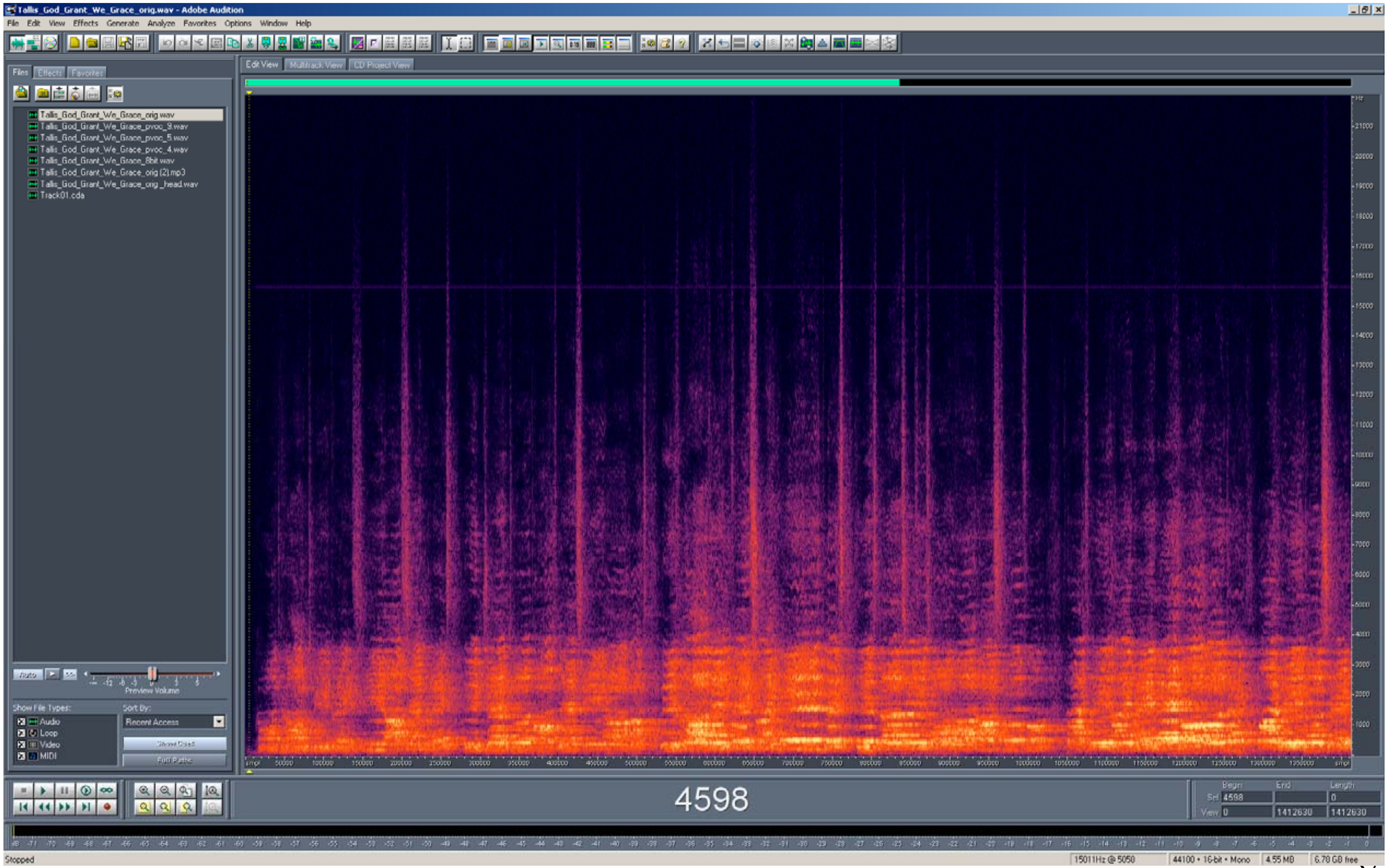
Tallis 1: Tail



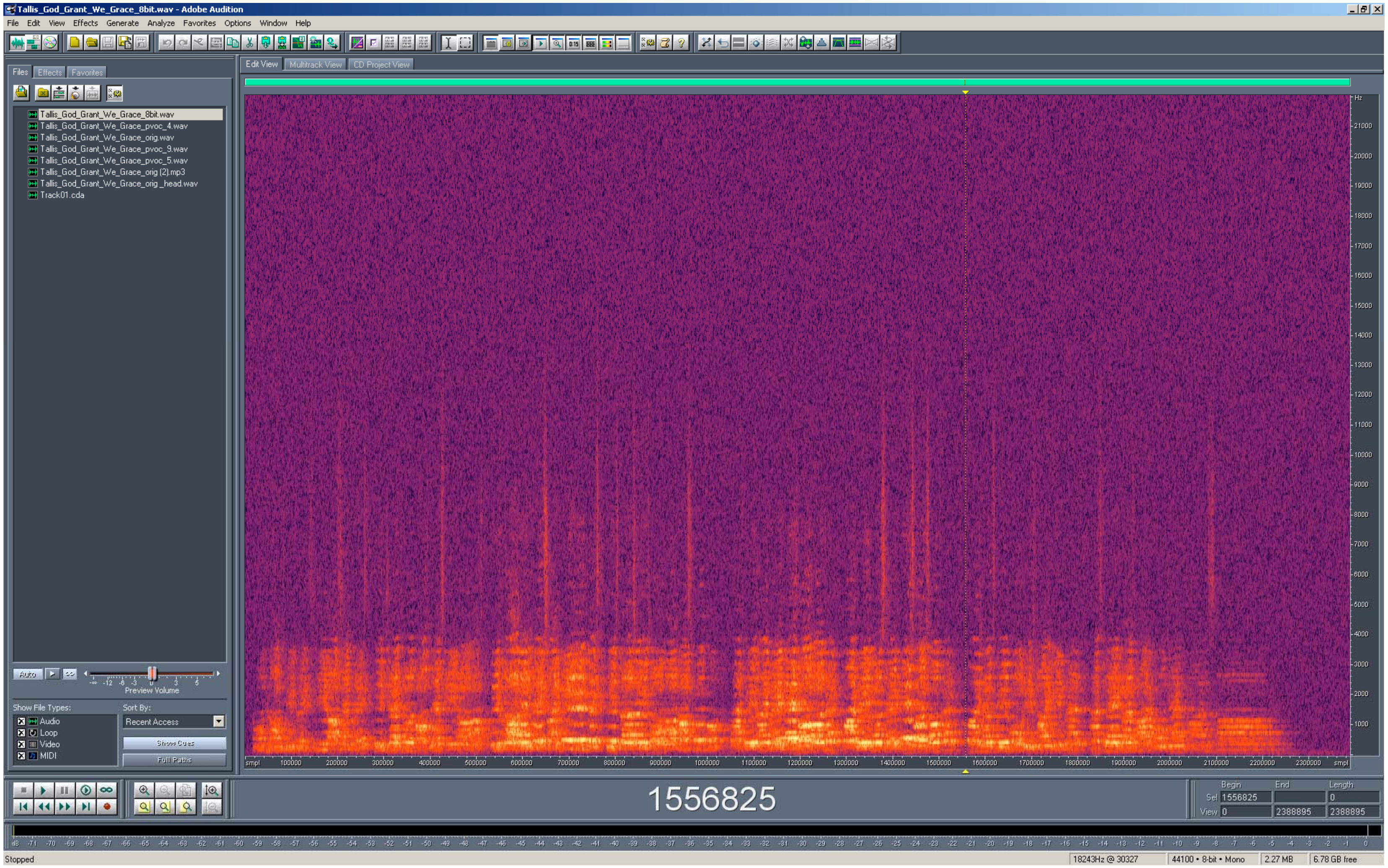
Tallis 2: Tail



Tallis 1: spectrum



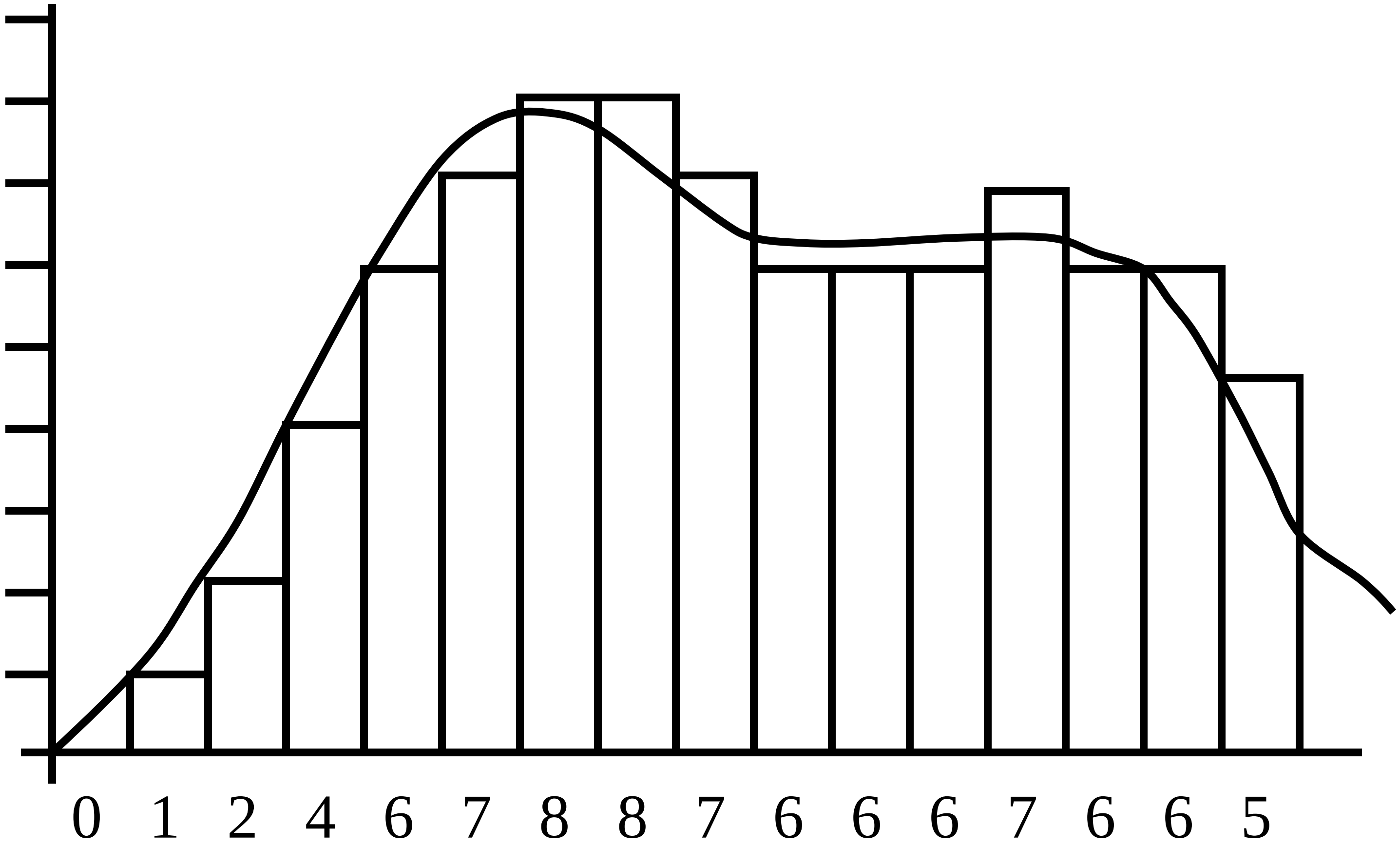
Tallis 2: Spectrum



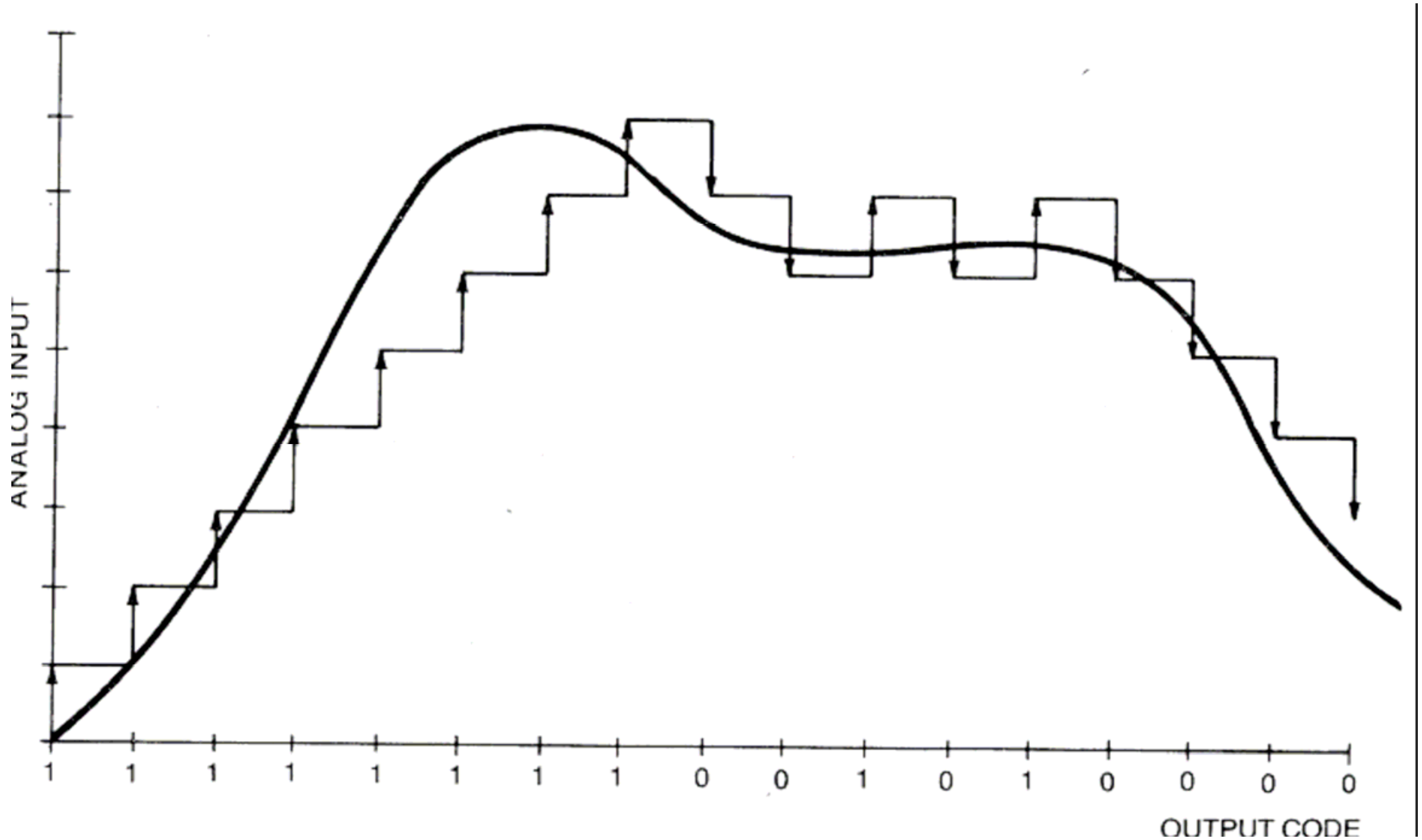
Meeting the challenge

- Coarser Quantization

PCM

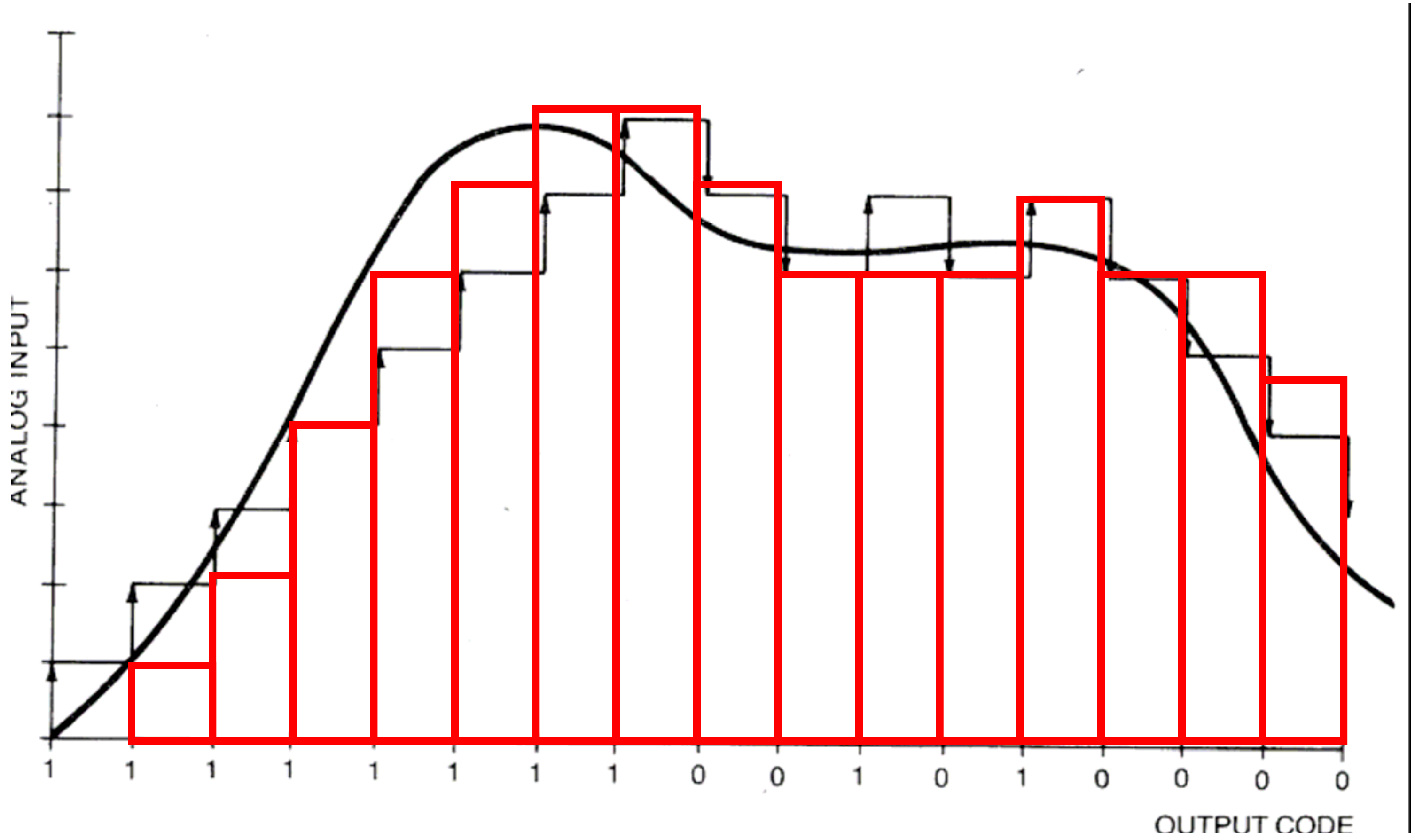


DPCM



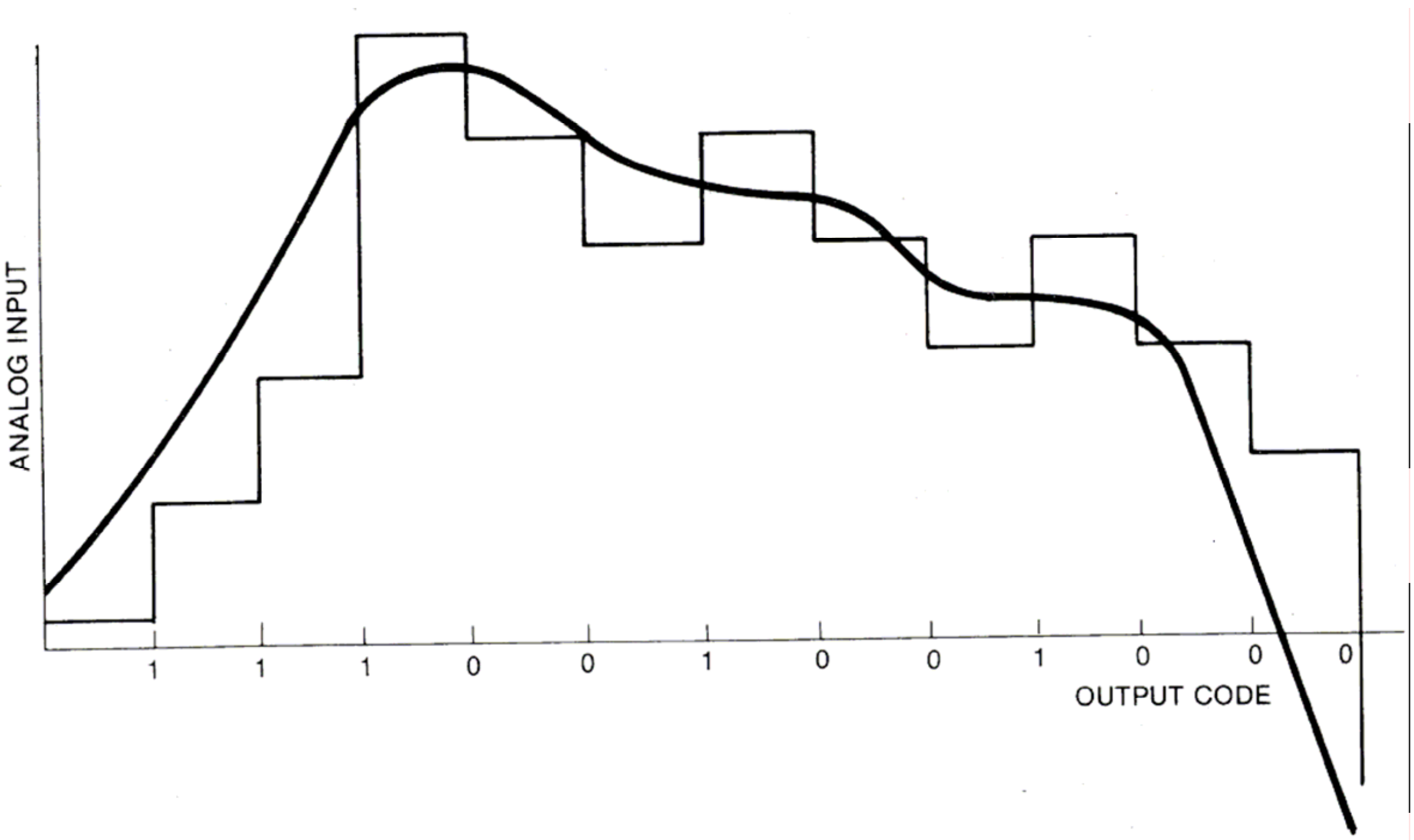
From: Ken Pohlmann, *Principles of Digital Audio*, Sams

DPCM vs. PCM



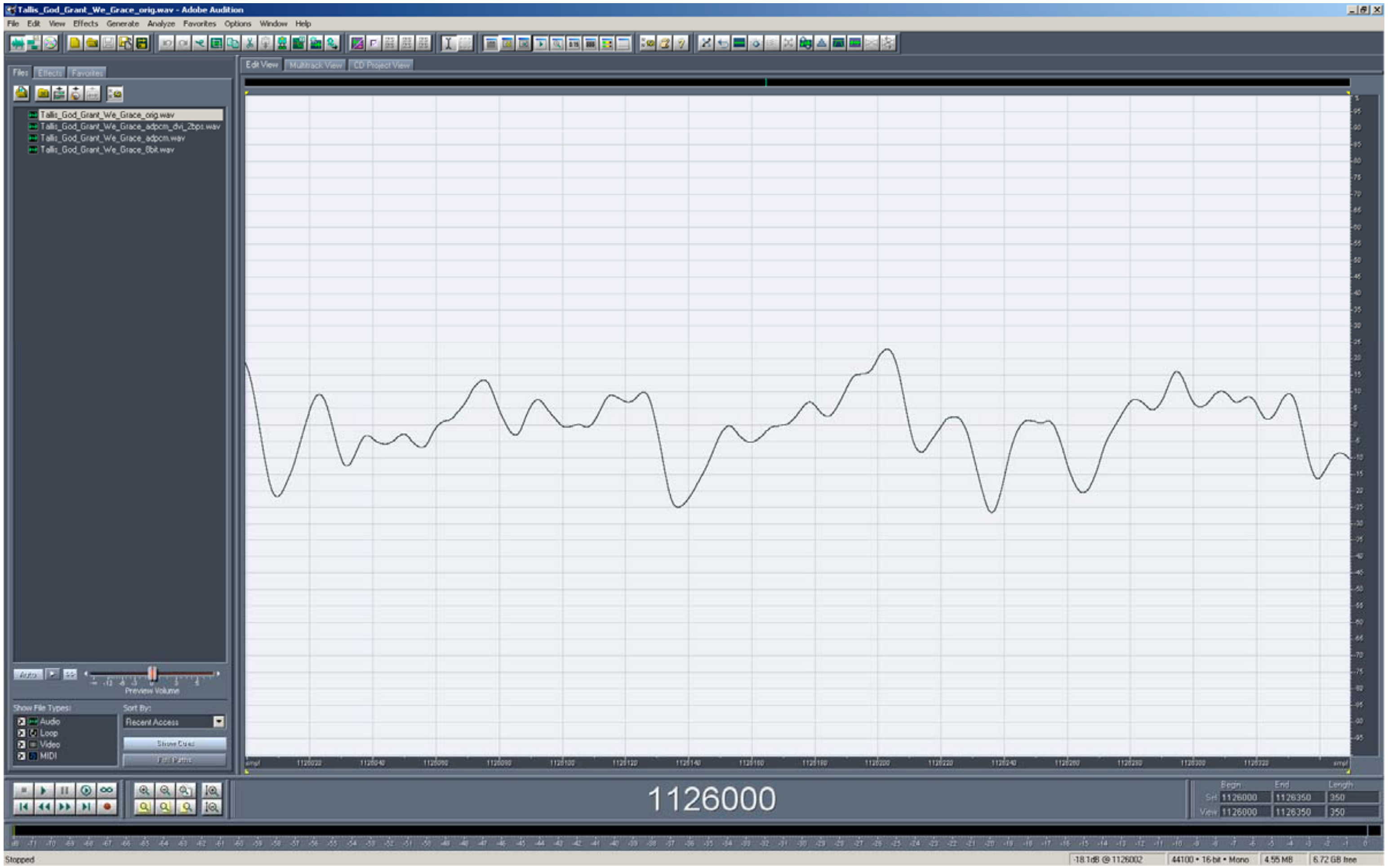
From: Ken Pohlmann, *Principles of Digital Audio*, Sams

ADPCM

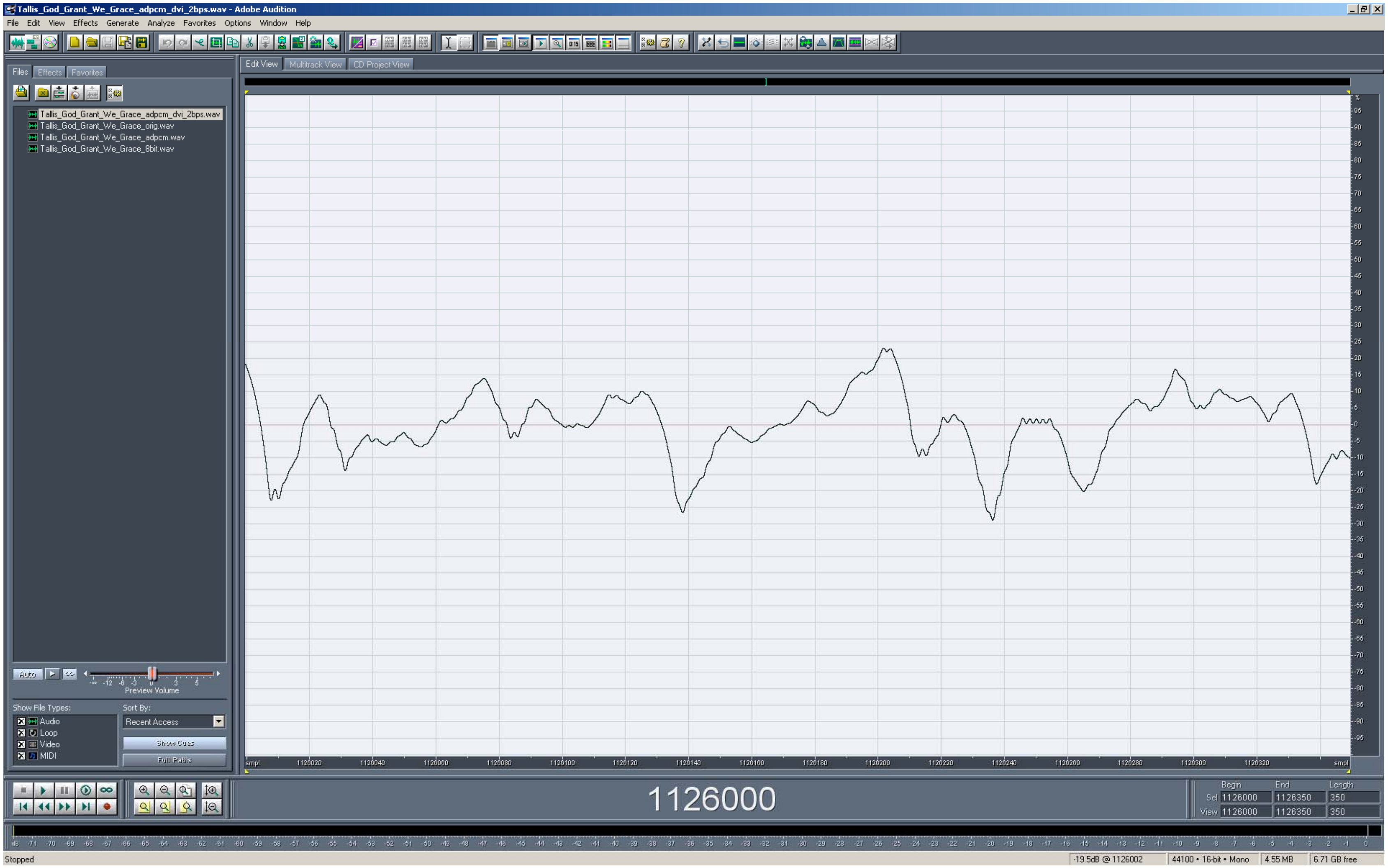


From: Ken Pohlmann, *Principles of Digital Audio*, Sams

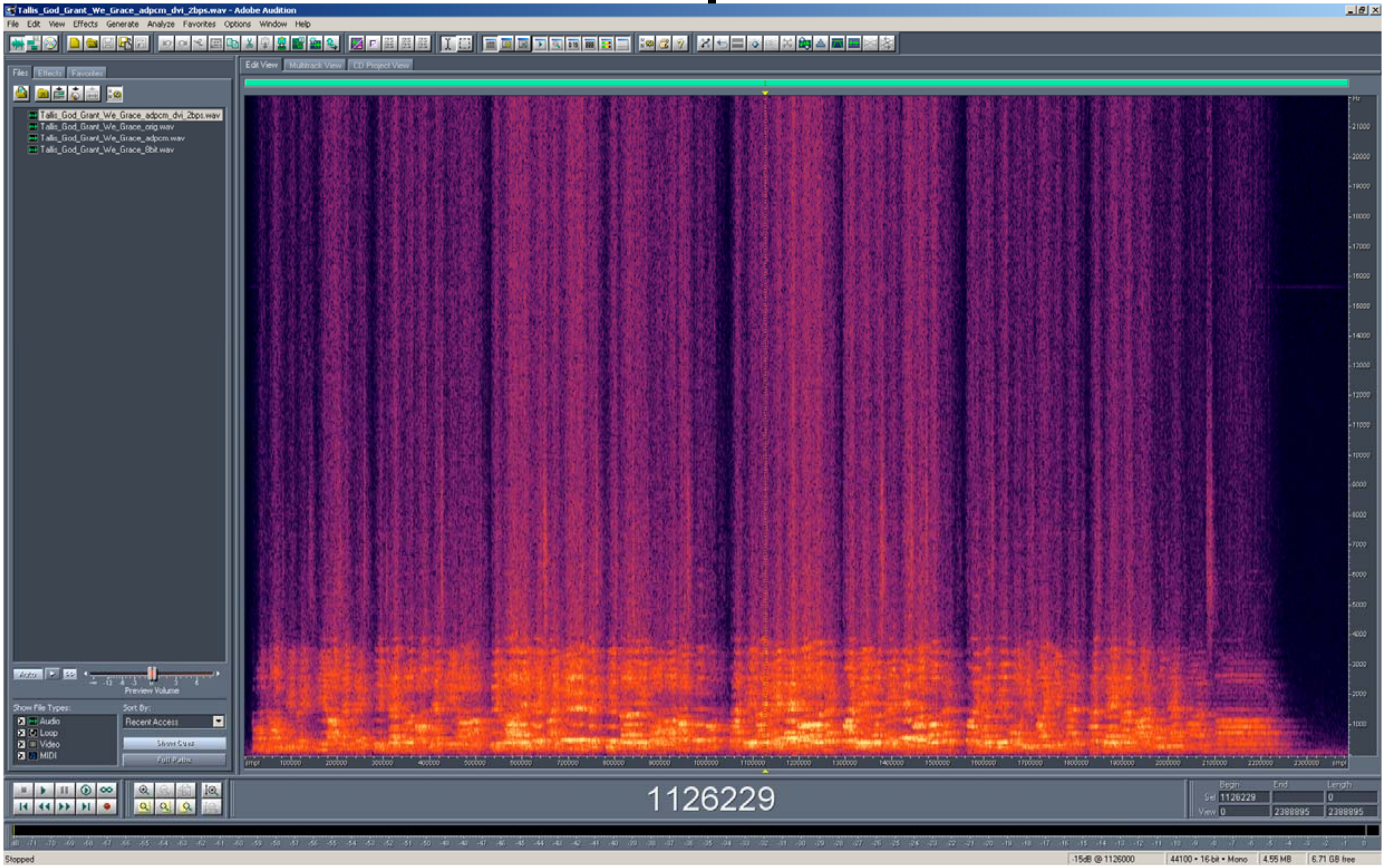
Tallis 1


Tallis_God_Grant_We_Grace_orig.wav - Adobe Audition
Tallis_God_Grant_We_Grace_orig.wav
Tallis_God_Grant_We_Grace_adpcm_dvi_2bps.wav
Tallis_God_Grant_We_Grace_adpcm.wav
Tallis_God_Grant_We_Grace_8bit.wav
1126000

Tallis 3


Tallis_God_Grant_We_Grace_adpcm_dvi_2bps.wav - Adobe Audition
File Edit View Effects Generate Analyze Favorites Options Window Help
Files Effects Favorites
Tallis_God_Grant_We_Grace_adpcm_dvi_2bps.wav
Tallis_God_Grant_We_Grace_orig.wav
Tallis_God_Grant_We_Grace_adpcm.wav
Tallis_God_Grant_We_Grace_8bit.wav
Edit View Multitrack View CD Project View
1126000
Stopped

Tallis 3: Spectrum



Meeting the challenge

- Coarser Quantization
- DPCM, ADPCM

Next Listening Session

- Meet by _____ in 285D, 285F, 285H, 285J. All rooms have same tracks.
- Listen to tracks Specod76/Specod77; and Specod80/Specod81 in Specod project.
- Take notes and discuss:
  - What do you hear?
  - How are they different?
  - What do you like?
  - What don't you like?
- Back here at 11:00 to discuss. (10 min)

Notes on Specod Sound Examples

Discuss Specod listening examples

76 77 80 81

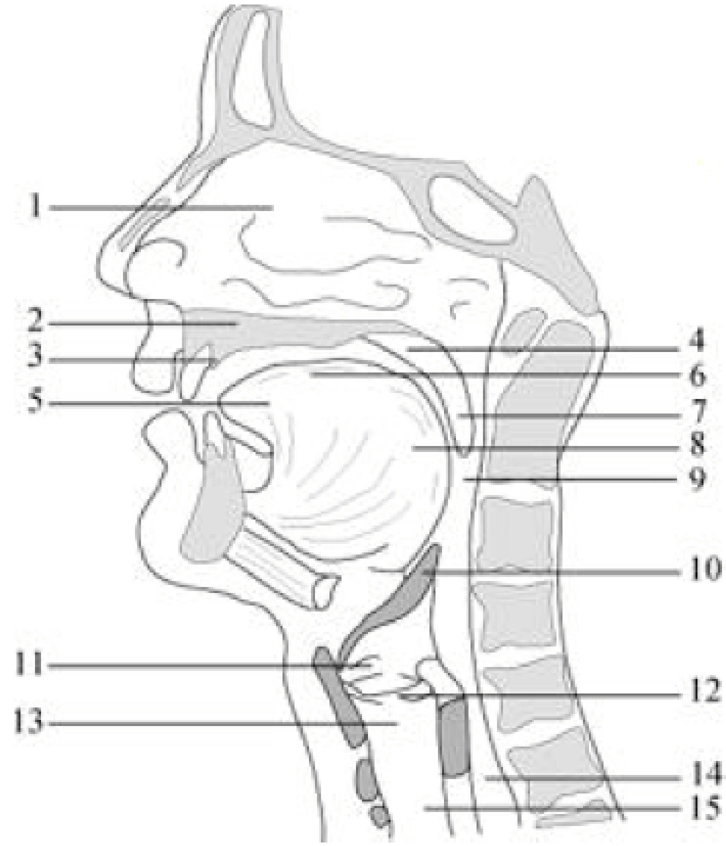
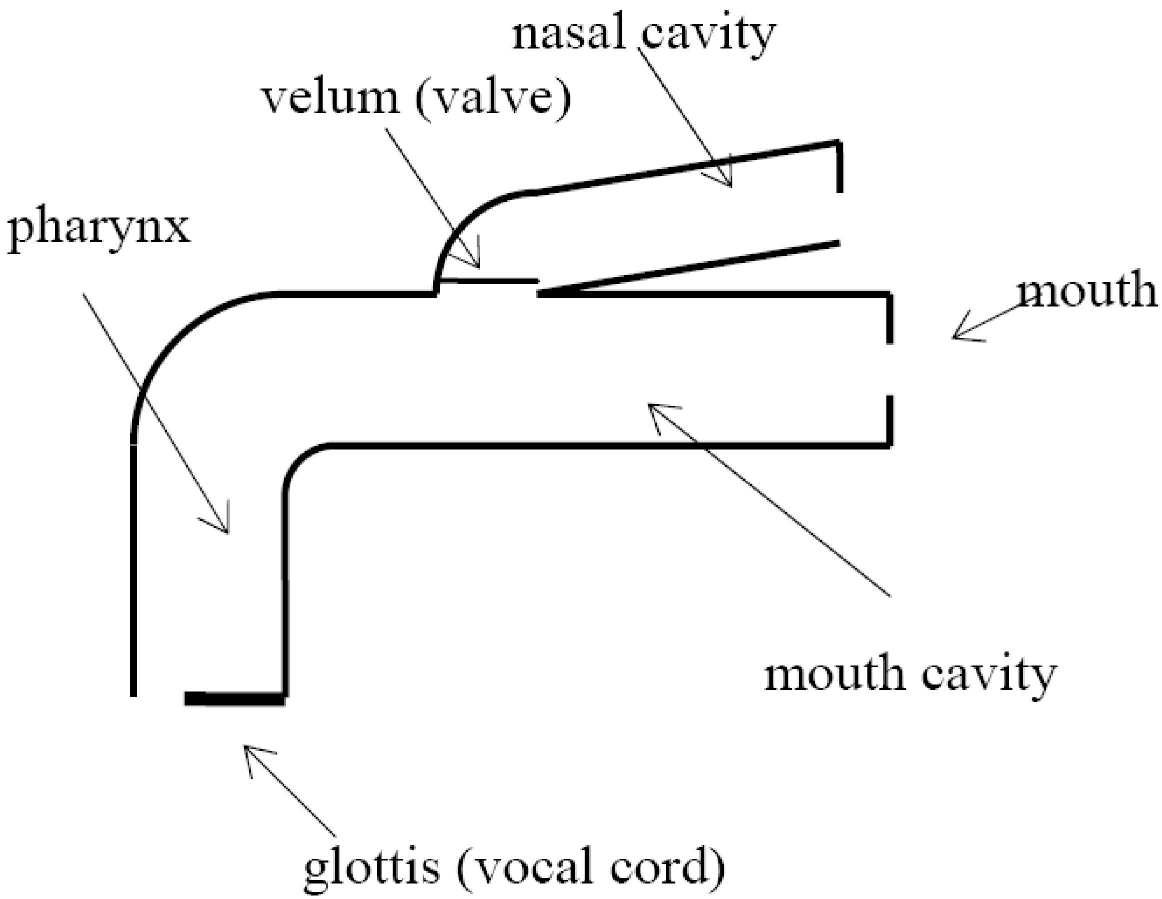


Fig. 3.5. The human vocal organs. (1) Nasal cavity, (2) Hard palate, (3) Alveoral ridge, (4) Soft palate (Velum), (5) Tip of the tongue (Apex), (6) Dorsum, (7) Uvula, (8) Radix, (9) Pharynx, (10) Epiglottis, (11) False vocal cords, (12) Vocal cords, (13) Larynx, (14) Esophagus, and (15) Trachea.

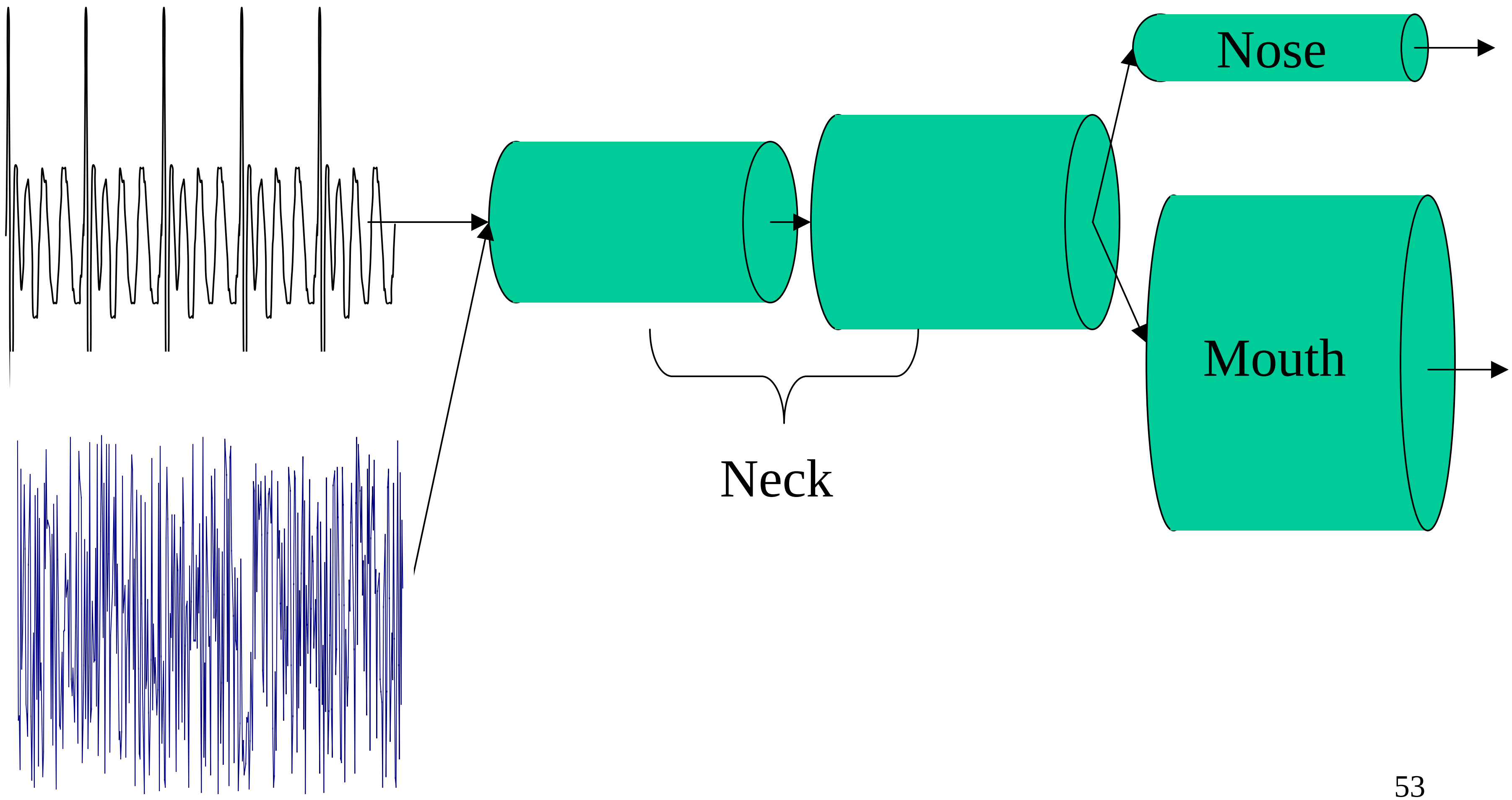
Lemmetty, Sami. "Review of speech synthesis technology." Masters thesis, Department of Electrical and Telecommunications Engineering, Univ. Helsiki, 1999. http://www.acoustics.hut.fi/publications/files/theses/lemmetty_mst/thesis.pdf. Retrieved 080301

Tube model of vocal tract



http://www.engineer.tamuk.edu/

Speech production model



LPC

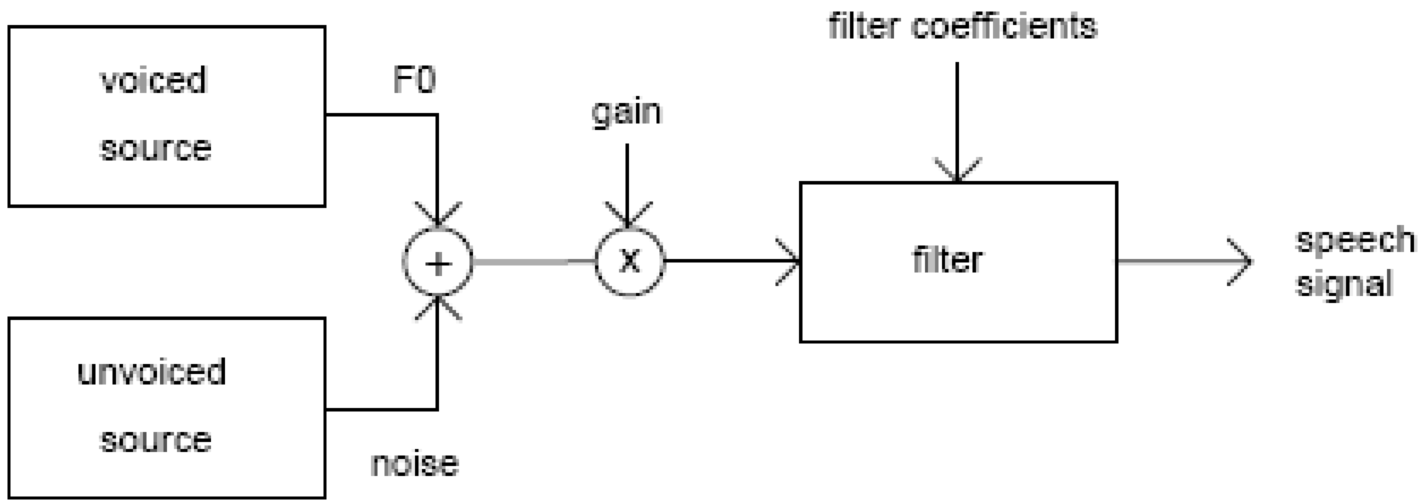


Fig. 1.2. Source-filter model of speech.

Lemmetty, Sami. "Review of speech synthesis technology." Masters thesis, Department of Electrical and Telecommunications Engineering, Univ. Helsiki, 1999. http://www.acoustics.hut.fi/publications/files/theses/lemmetty_mst/thesis.pdf. Retrieved 080301

Speech coding: What you heard

- Male speech, original
- Male speech, coded/encoded with speech coder
- Vocal quartet, original
- Vocal quartet, coded/encoded with speech coder

Source: AES CD-ROM, from SQAM/EBU.

Meeting the challenge

- Coarser Quantization
- DPCM, ADPCM
- Linear Prediction

What we have covered

- Compression: not in a vacuum
- What problems does compression solve? (Why bother?)
- Forerunners to perceptual compression

MSRA 5500 - 002
MUS 4500 - 002

March 14, Session 1, Part 2

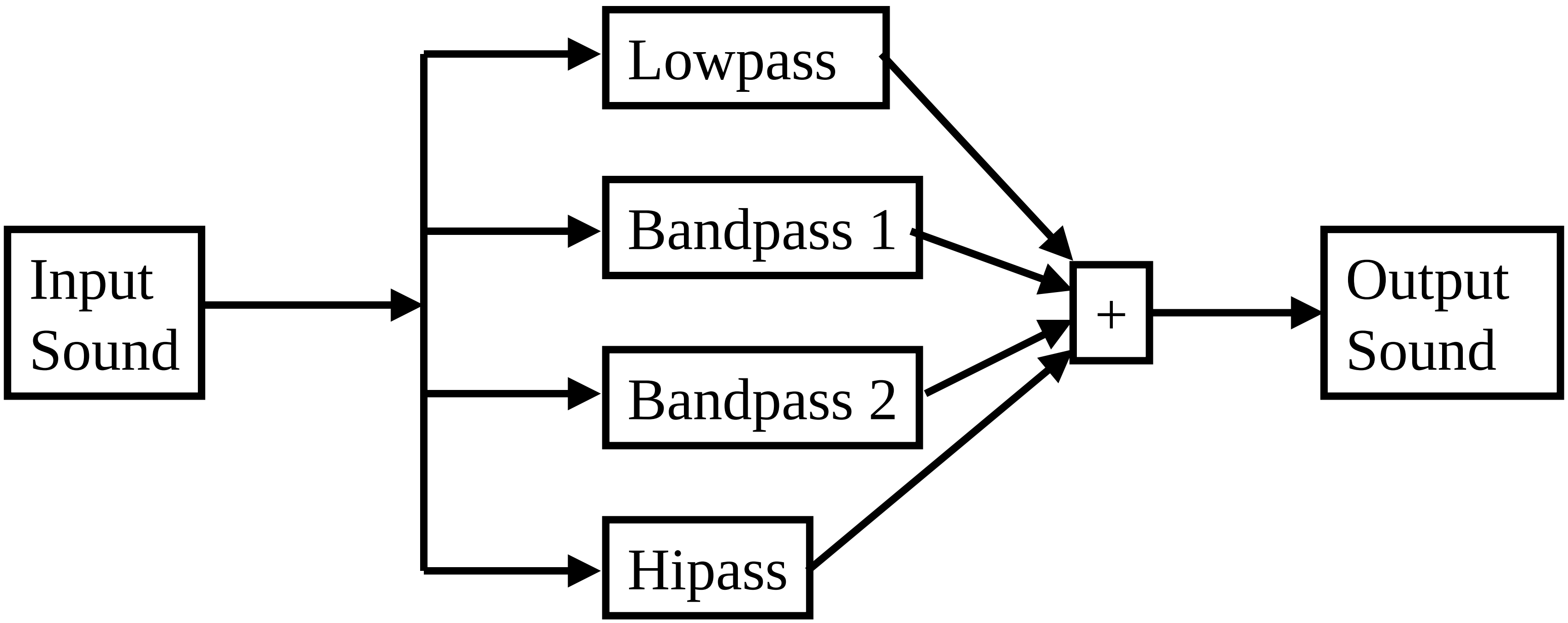
Subband and Transform Coding

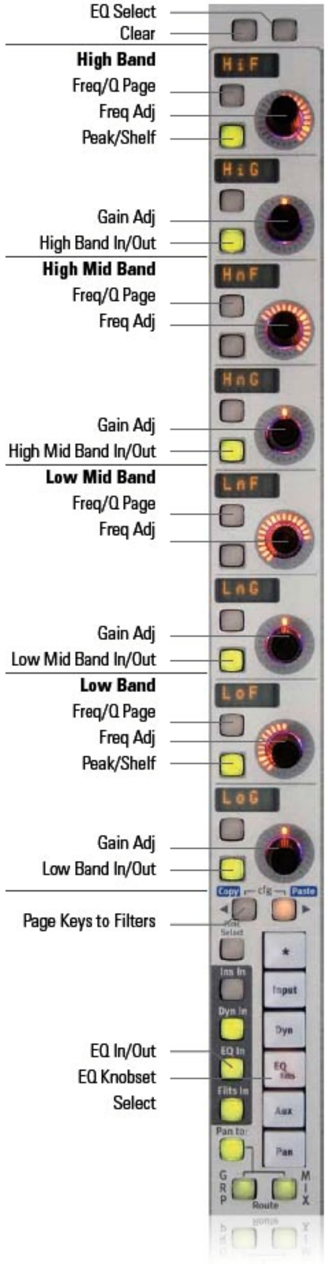
© Copyright 2008 John Strawn

What we will cover

- Subband Coding
- Transform Coding
- Structure of a simple encoder/decoder
- Quantizing transform coefficients

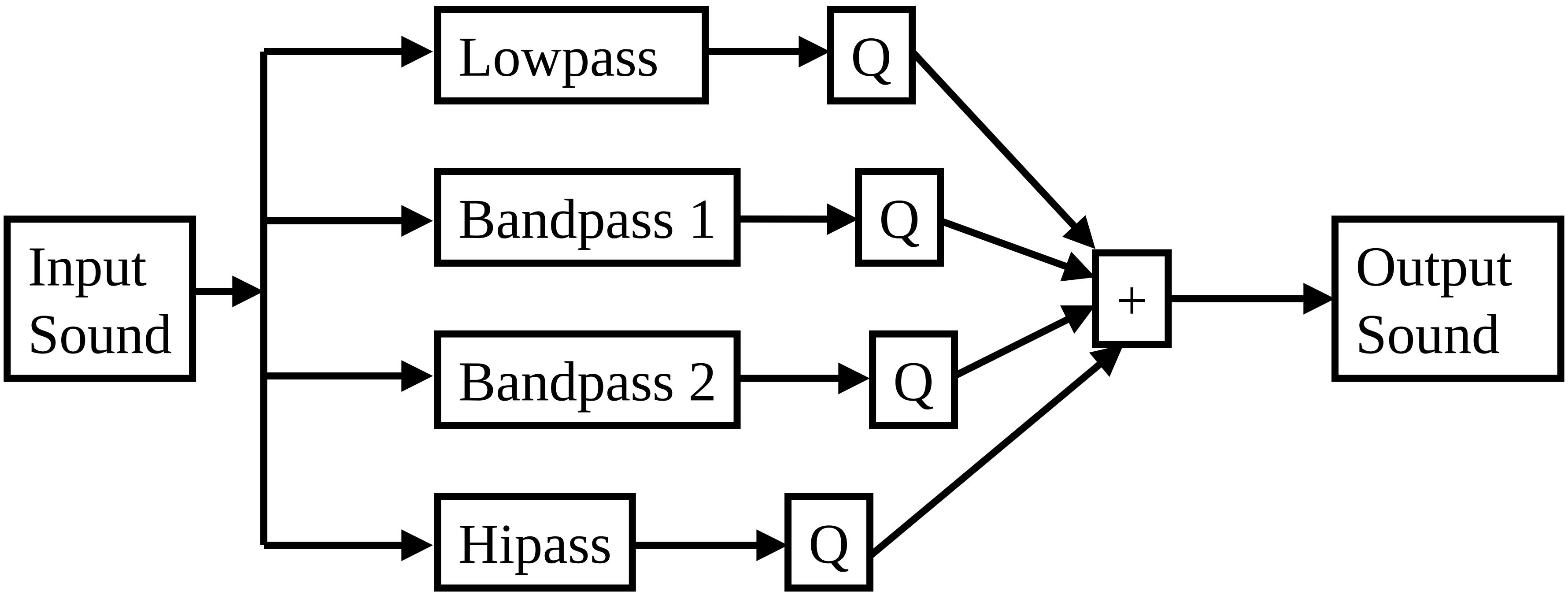
Subband


Input
Sound
Lowpass
Bandpass 1
Bandpass 2
Hipass
+
Output
Sound

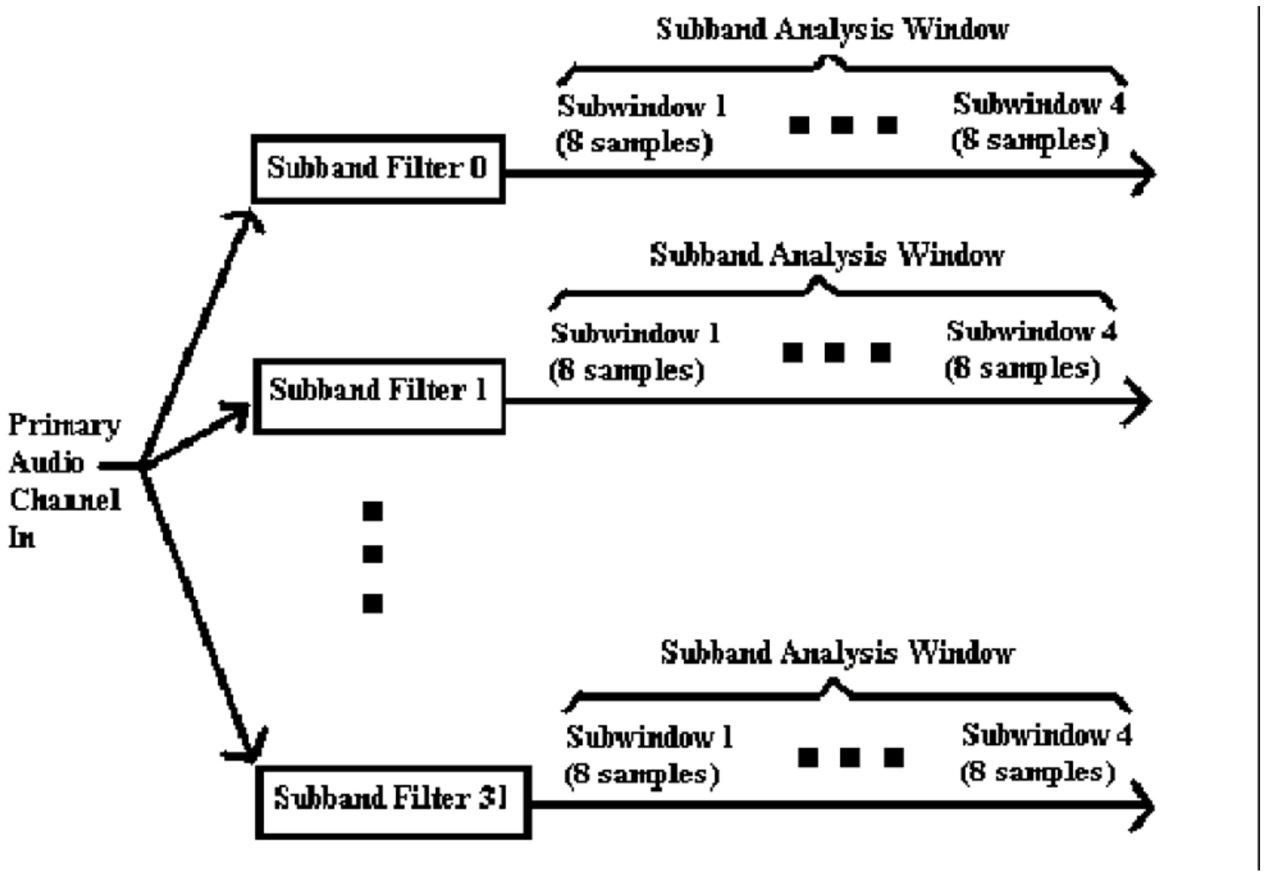


www.euphonix.com

Subband with Quantization

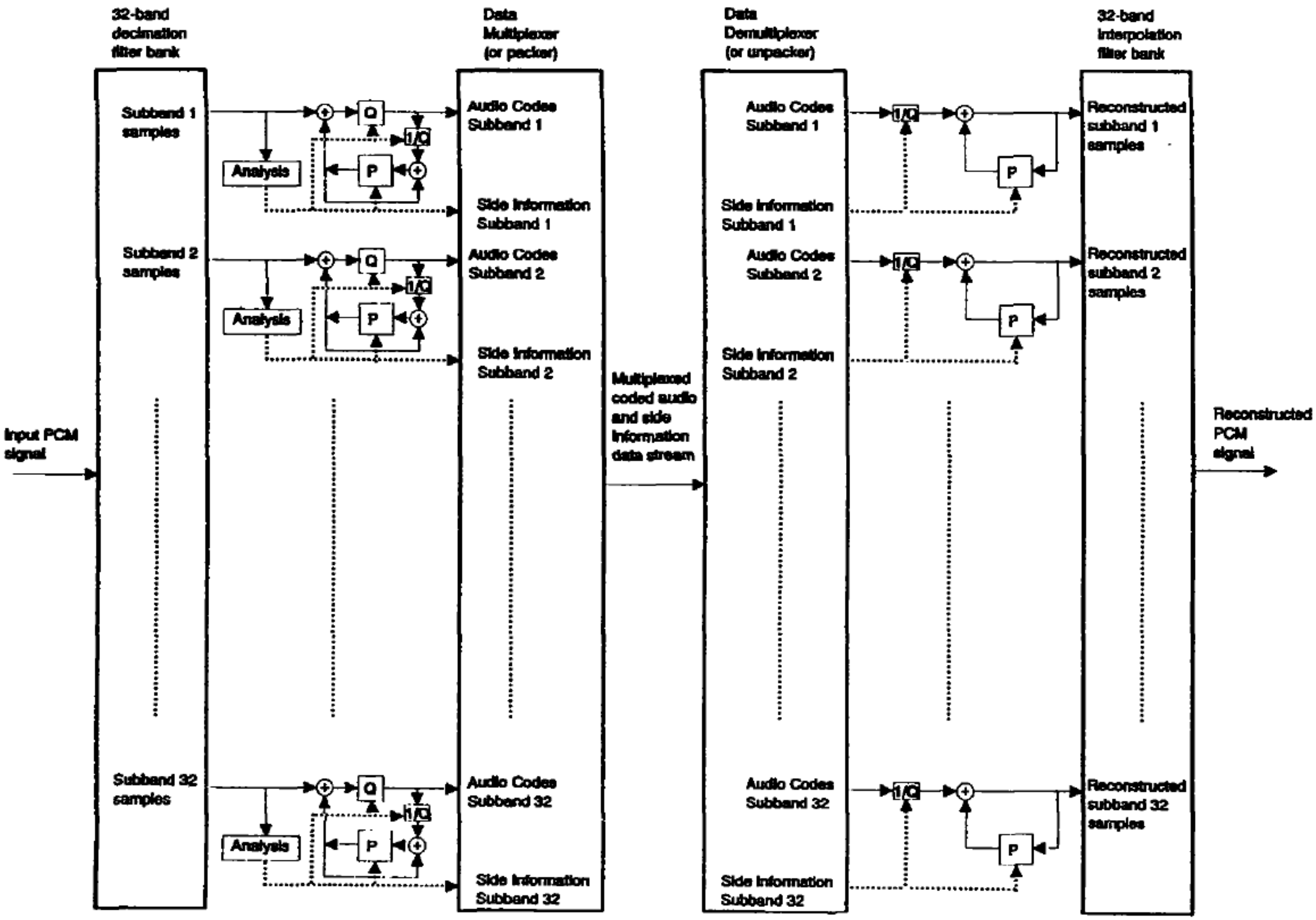


DTS



DTS Coherent Acoustics White Paper: Requirements Specification for Core Audio, Version Draft 2, n.d.

DTS Coherent Acoustics



Smyth, S. M. F.; Smith, W. P.; Smyth, M. H. C.; Yan, M.; Jung, T. "DTS Coherent Acoustics Delivering High-Quality Multichannel Sound to the Consumer." Preprint 4293, AES Convention, May 1996

Meeting the challenge

- Coarser Quantization
- DPCM, ADPCM
- Linear Prediction
- Subband coding

Listening

- Meet by _____ in 285D, 285F, 285H, 285J. All rooms have same tracks.
- Listen to tracks in Tallis PVOC project, compare, contrast.
- Take notes and discuss among yourselves:
  - What do you hear?
  - How are they different?
  - What do you like?
  - What don't you like?
- Was one the original?
- Back here at ______ to discuss. (15 minutes)

Notes on Tallis PVOC Sound Examples

Tallis PVOC: Discuss Listening examples

Tallis 4

Tallis 8

Tallis 16

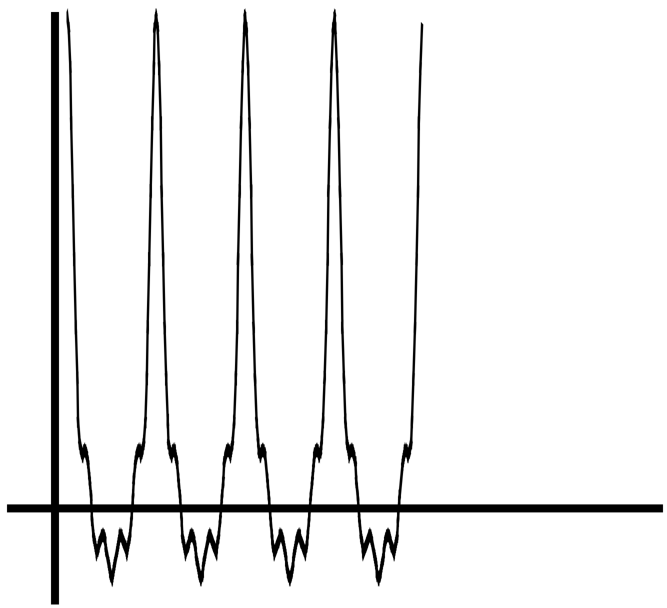
Tallis 128

What we will cover

- Subband Coding
- <u>Transform Coding</u>
- Structure of a simple encoder/decoder
- Quantizing transform coefficients

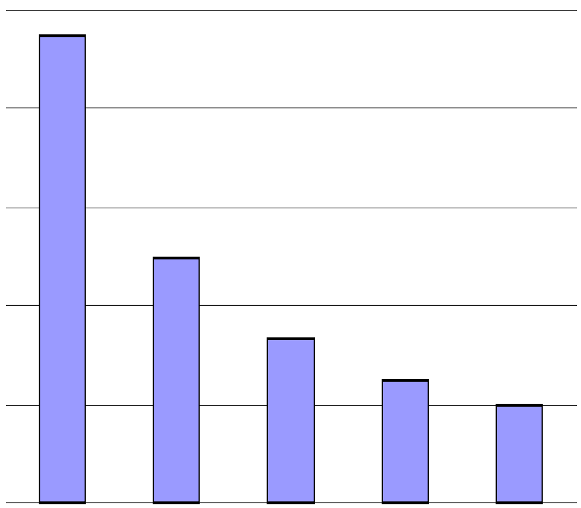
Domains

Waveform



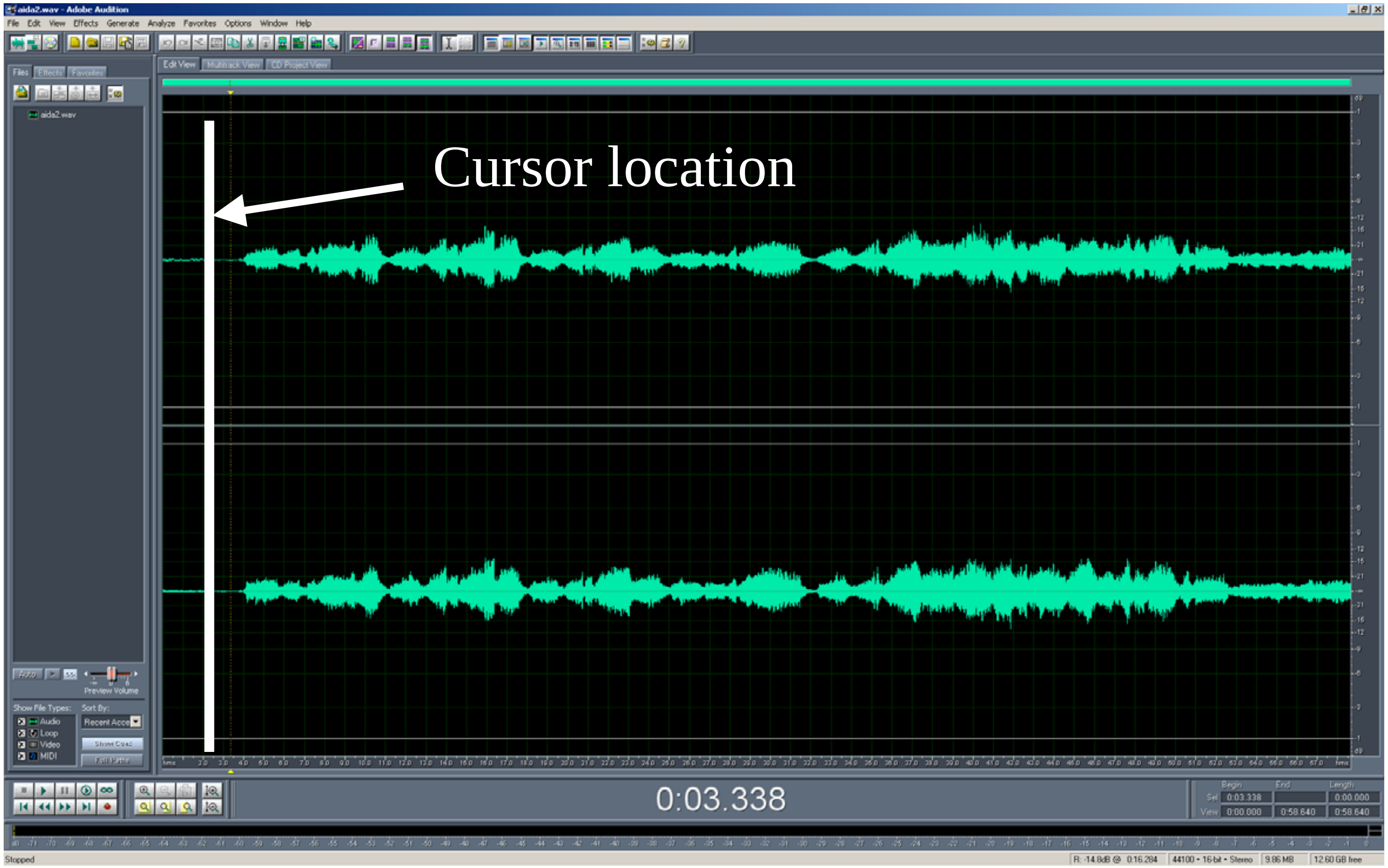
Time Domain

Spectrum

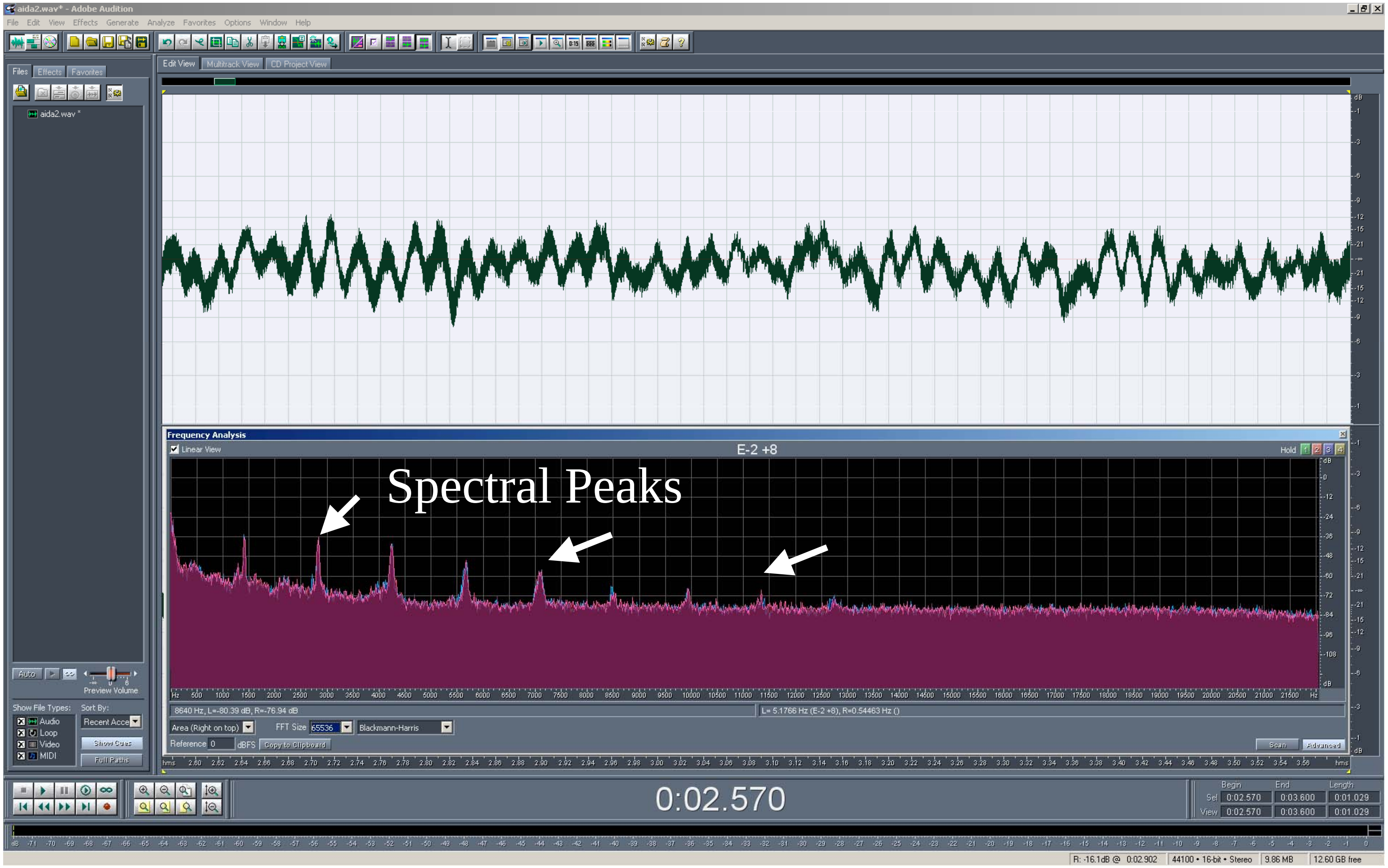


Frequency Domain

Waveform vs Spectrum

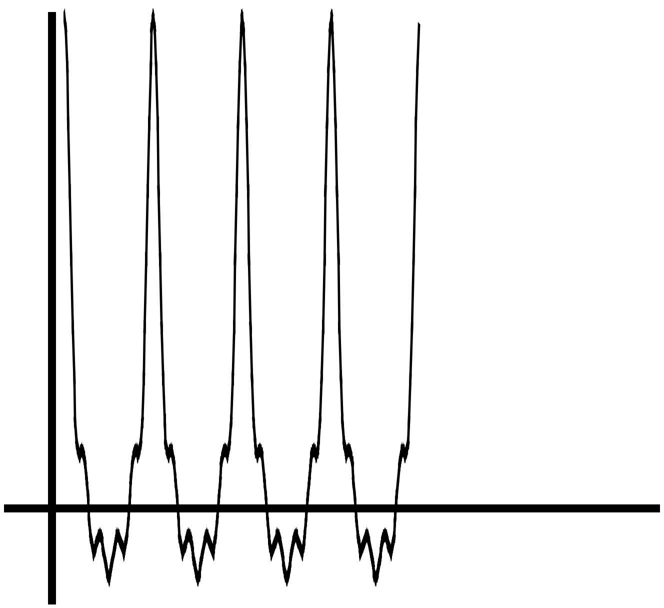


Waveform vs Spectrum

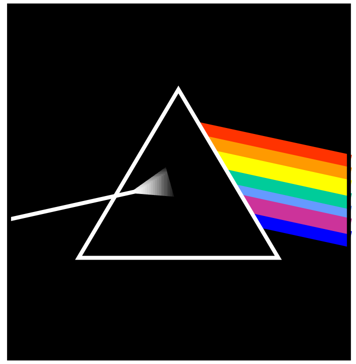


Transform

Waveform



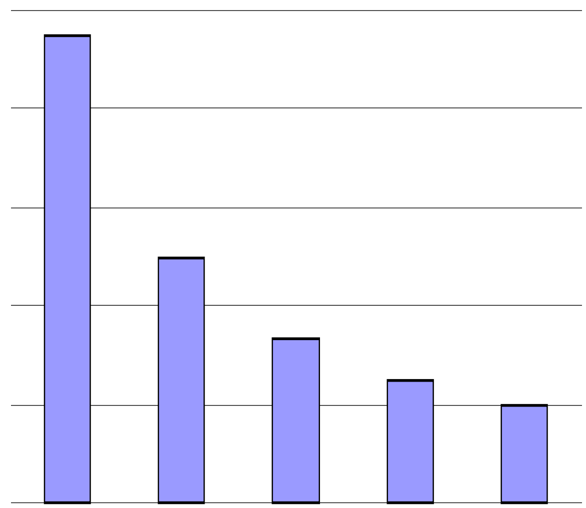
Time Domain



Forward
(analyze)

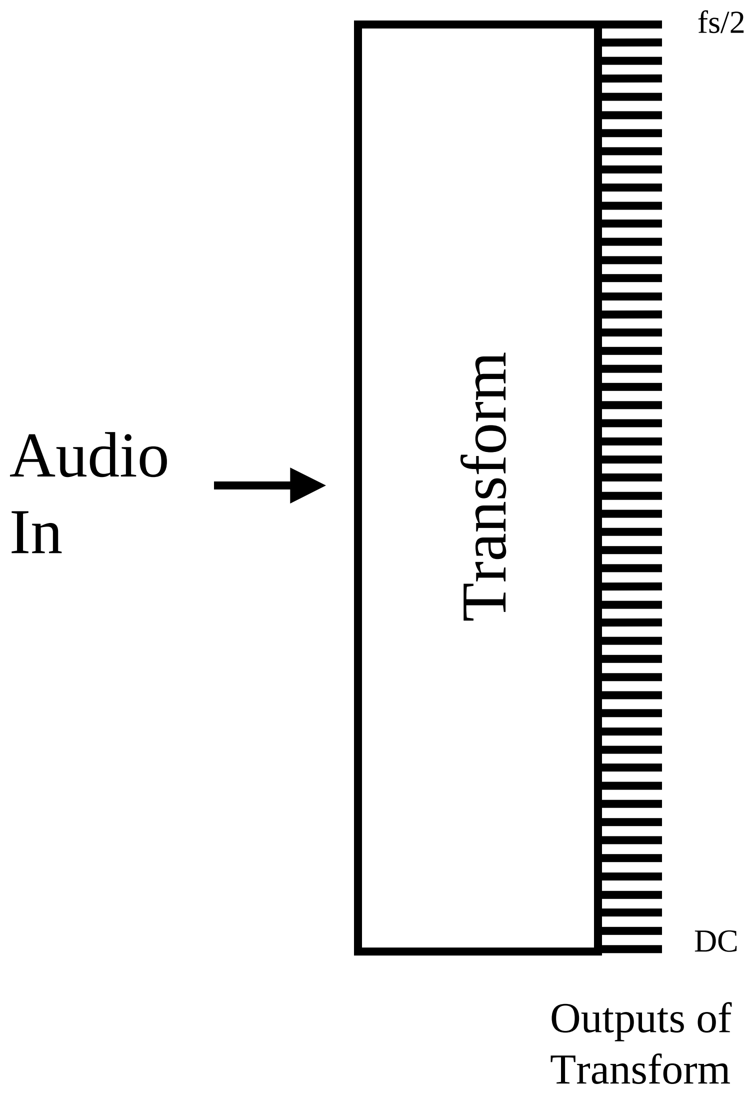
Inverse
(Synthesize)

Spectrum



Frequency Domain

Transform



Tallis PVOC: What you heard (and didn't hear)

- Tallis Orig
- Tallis 128
- Difference

What we will cover

- Subband Coding
- Transform Coding
- Structure of a simple encoder/decoder
- Quantizing transform coefficients

Transform Encoder

Time-Domain Audio Signal

Forward transform (enter frequency domain), get spectrum

Package compressed output (e.g. MP3 file)

Decoder (“mp3 player”)

Unpackage compressed audio (e.g. MP3 file)

↓

Inverse Transform: change from frequency to time

↓

Time-Domain Audio Signal

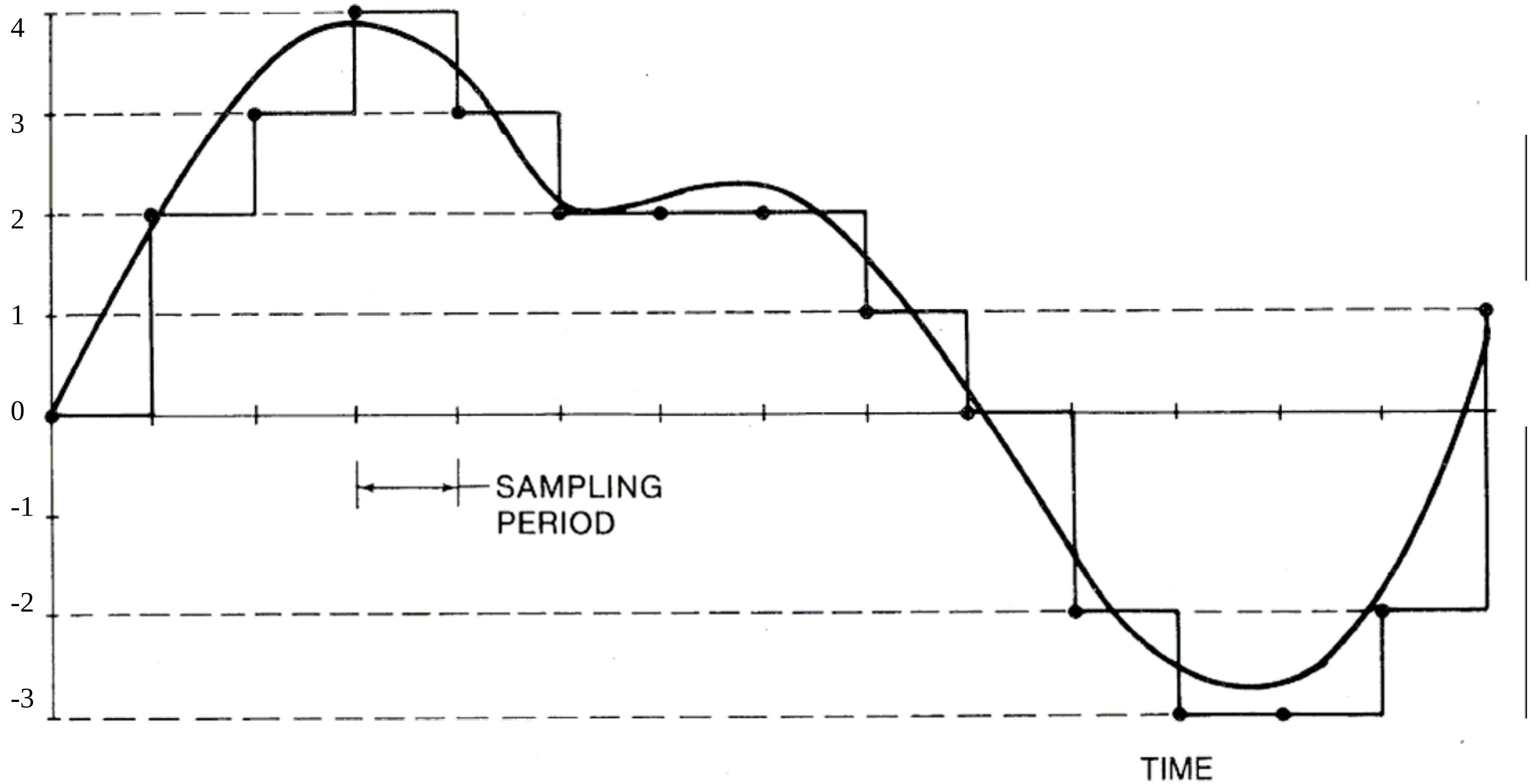
"Is MP3 coding related to the FFT?"

- Fourier transform
- Discrete Fourier transform
- Fast Fourier transform (FFT)
- Discrete Cosine Transform (DCT)
- Modified Discrete Cosine Transform (MDCT) (More today in optional session)
- [subband transforms; filter banks; quadrature mirror filters (QMF)]

What we will cover

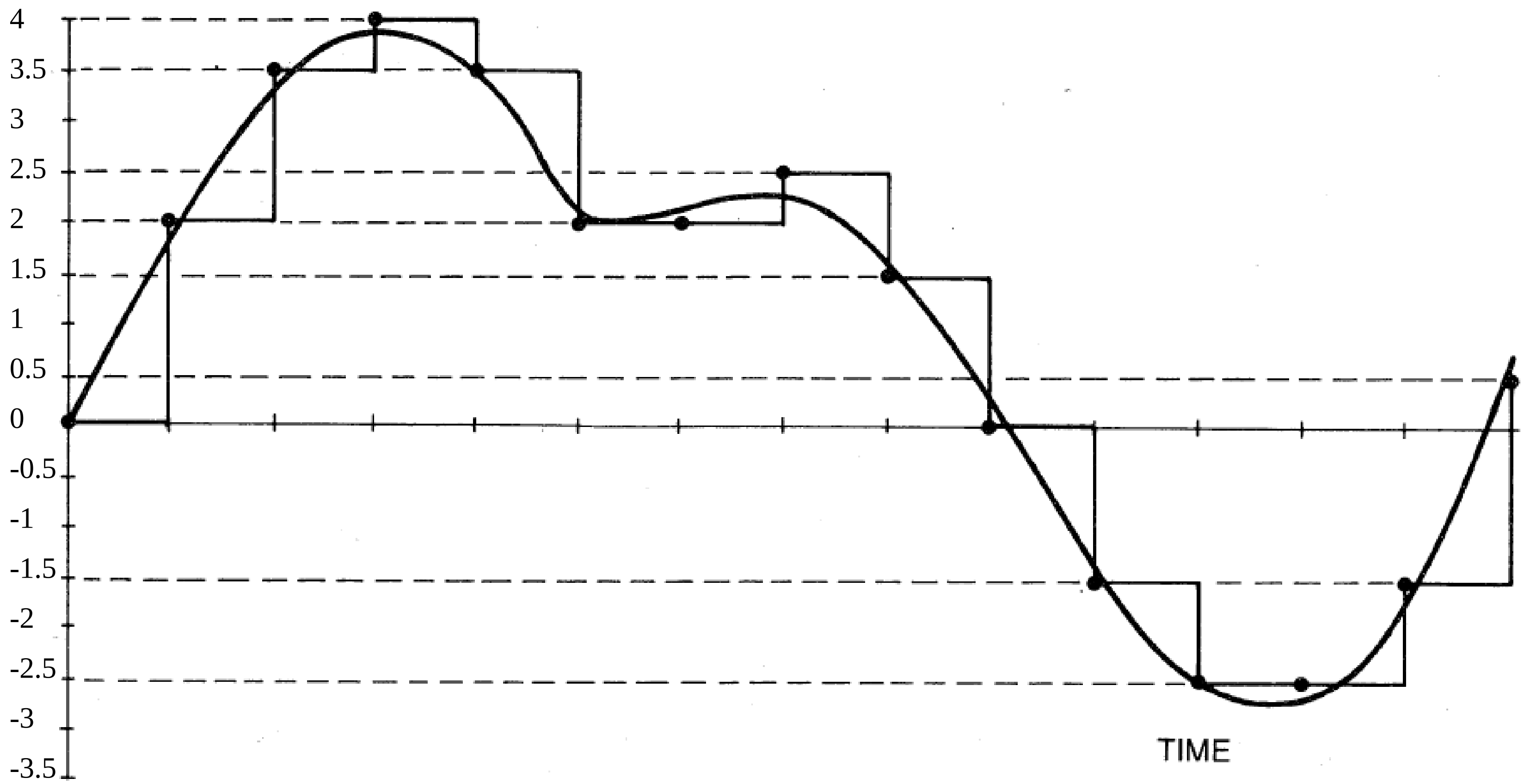
- Subband Coding
- Transform Coding
- Structure of a simple encoder/decoder
- <u>Quantizing transform coefficients</u>

Review: Coarser Quantization



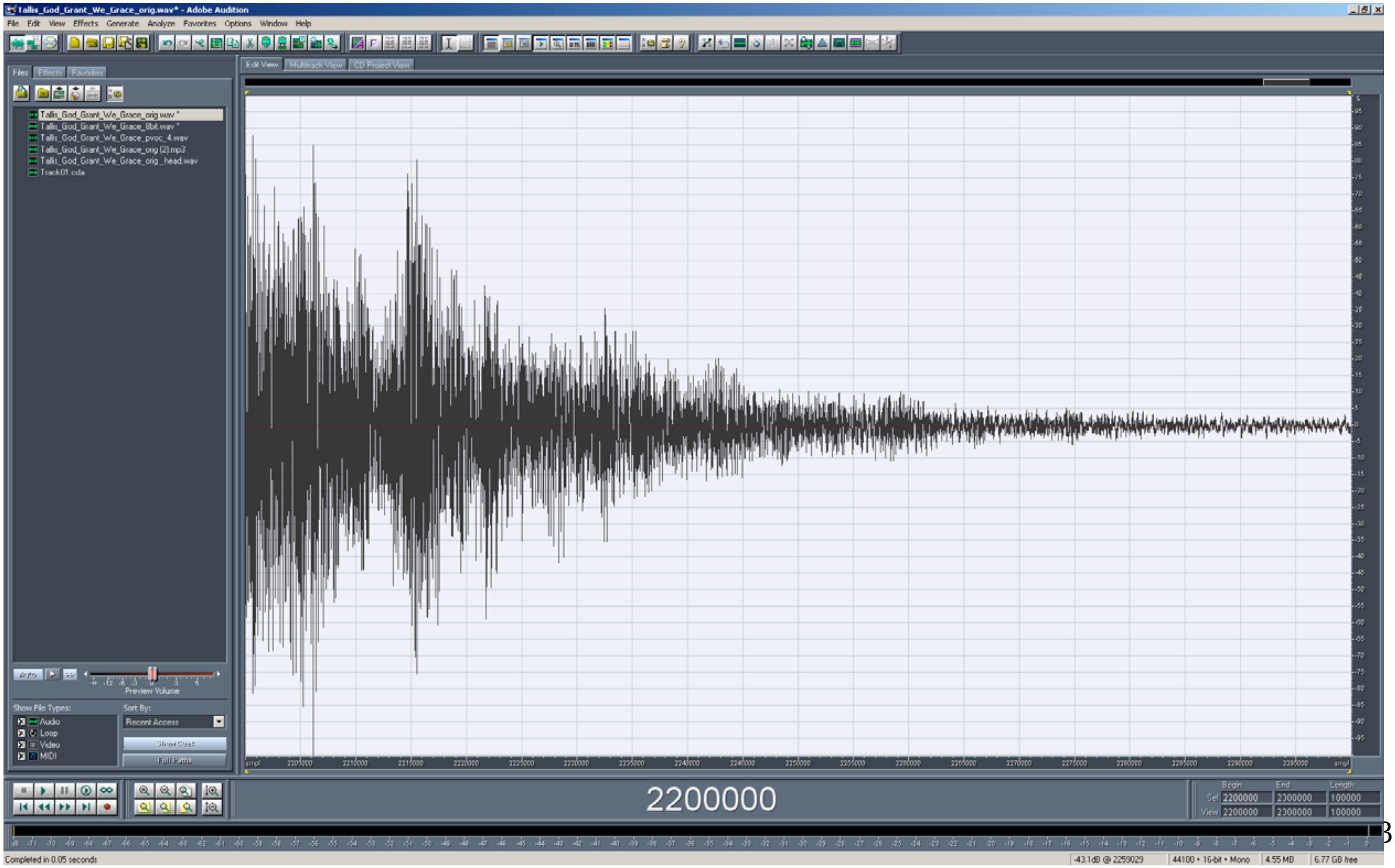
From: Ken Pohlmann, *Principles of Digital Audio*, Sams

Review: Finer Quantization

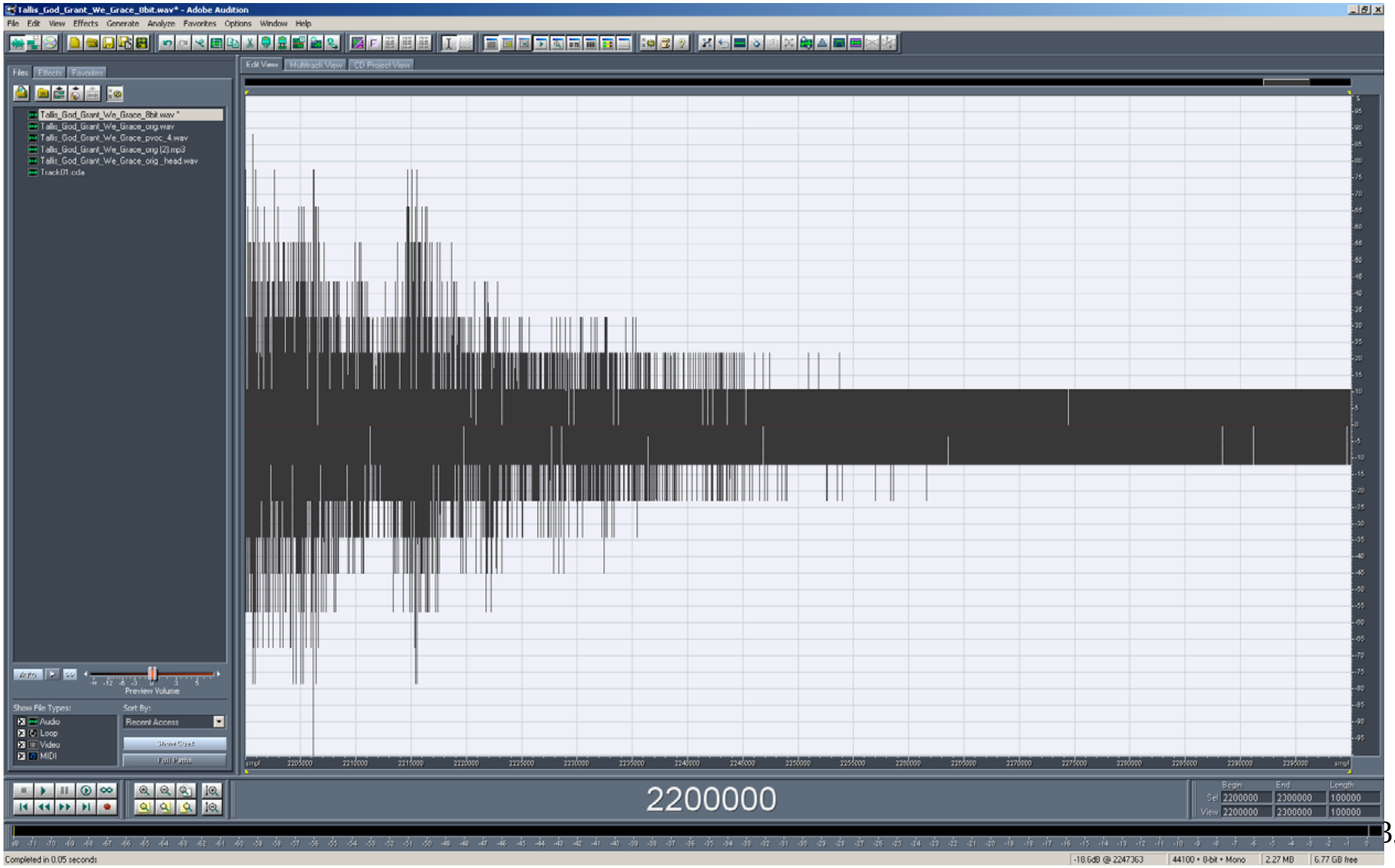


From: Ken Pohlmann, *Principles of Digital Audio*, Sams

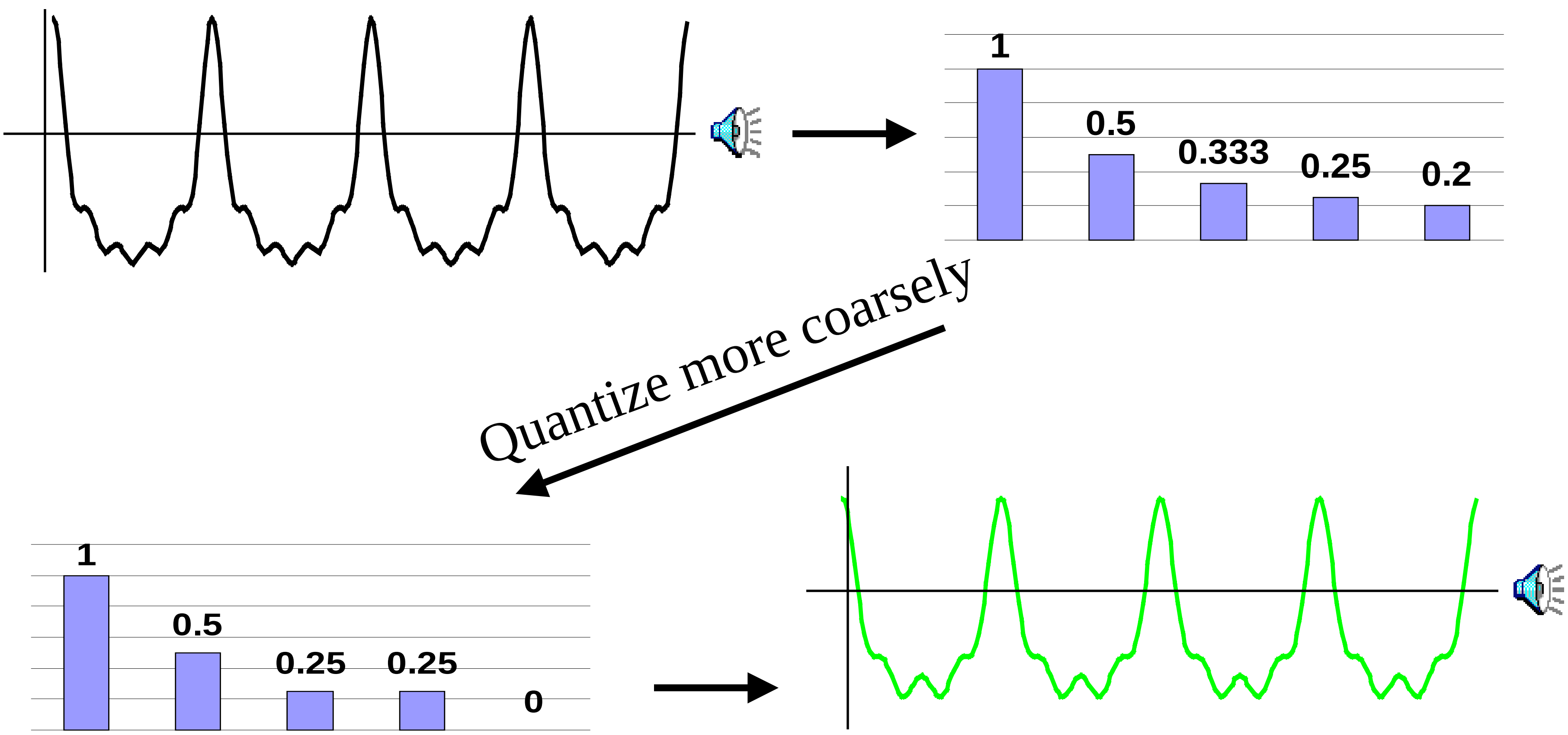
Review: Tallis 1: Tail



Review: Tallis 2: Tail

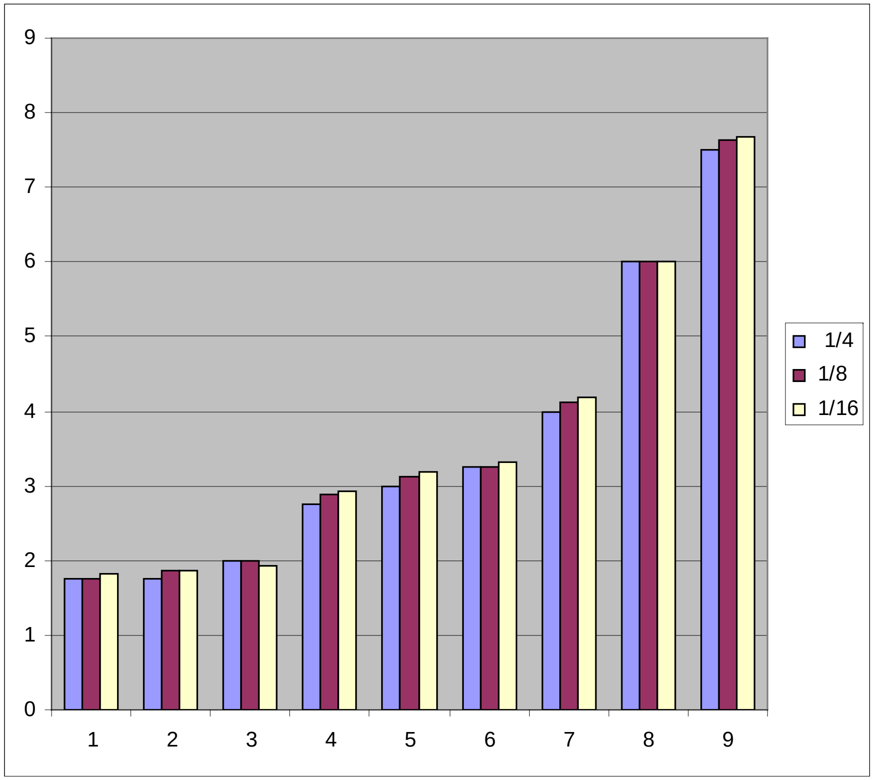


Leave something out -> Noise

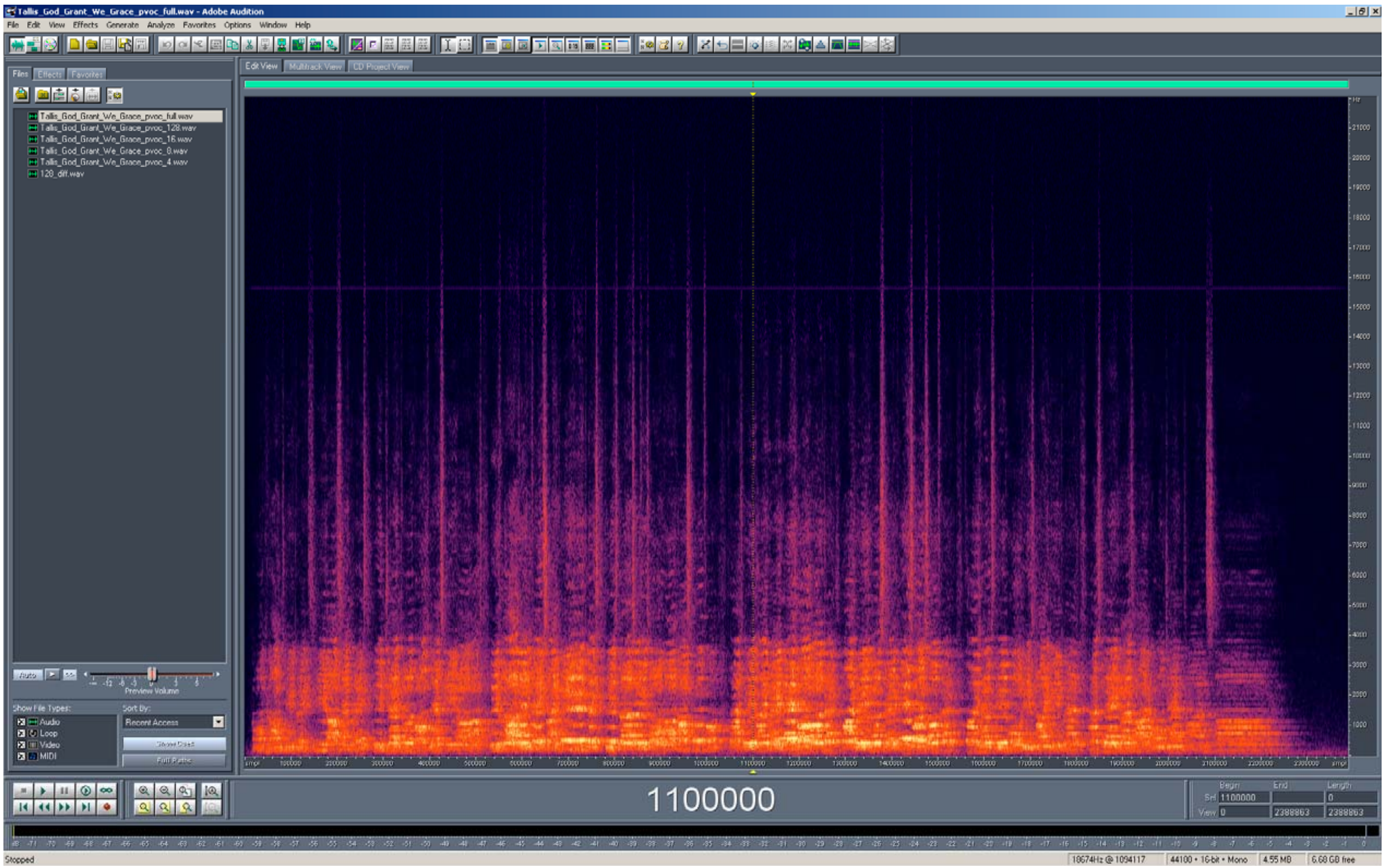


Tallis PVOC Values

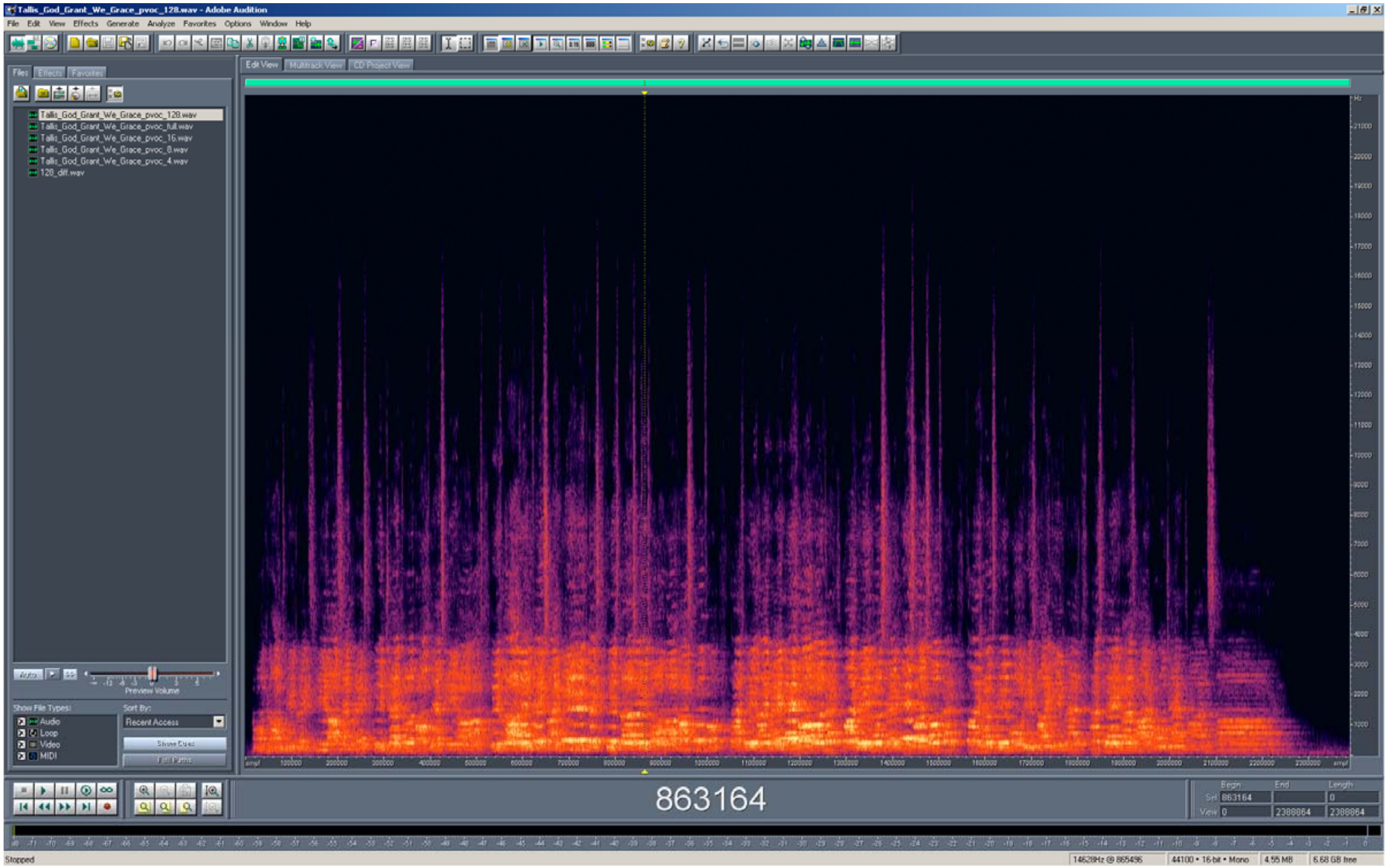
1/4	1 /8	1/16	full
1.75	1.75	1.8125	1.8438
1.75	1.875	1.875	1.8906
2	2	1.9375	1.9375
2.75	2.875	2.9375	2.9609
3	3.125	3.1875	3.1875
3.25	3.25	3.3125	3.3438
4	4.125	4.1875	4.2344
6	6	6	6.0469
7.5	7.625	7.6875	7.7031



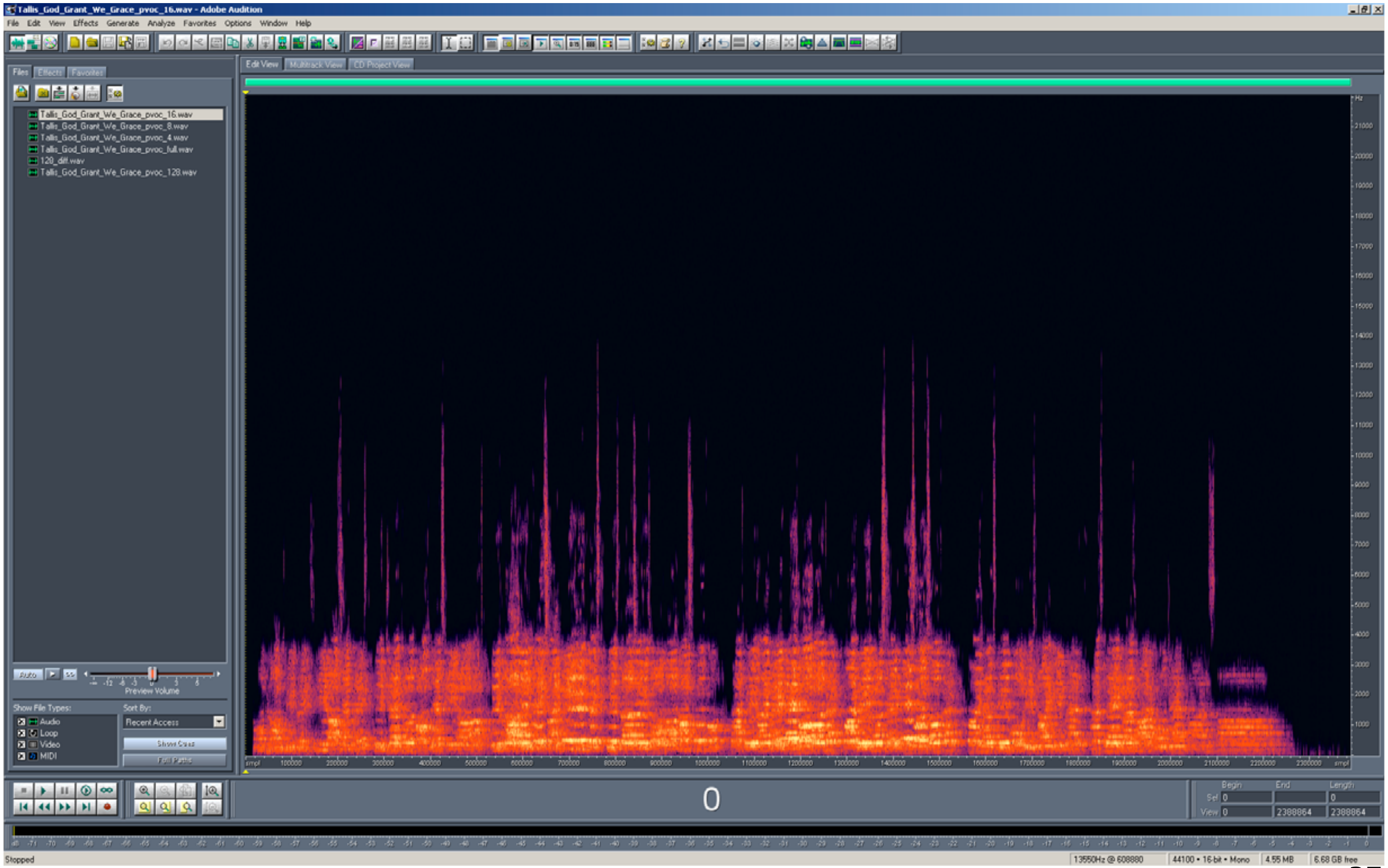
Spectrum, full resolution



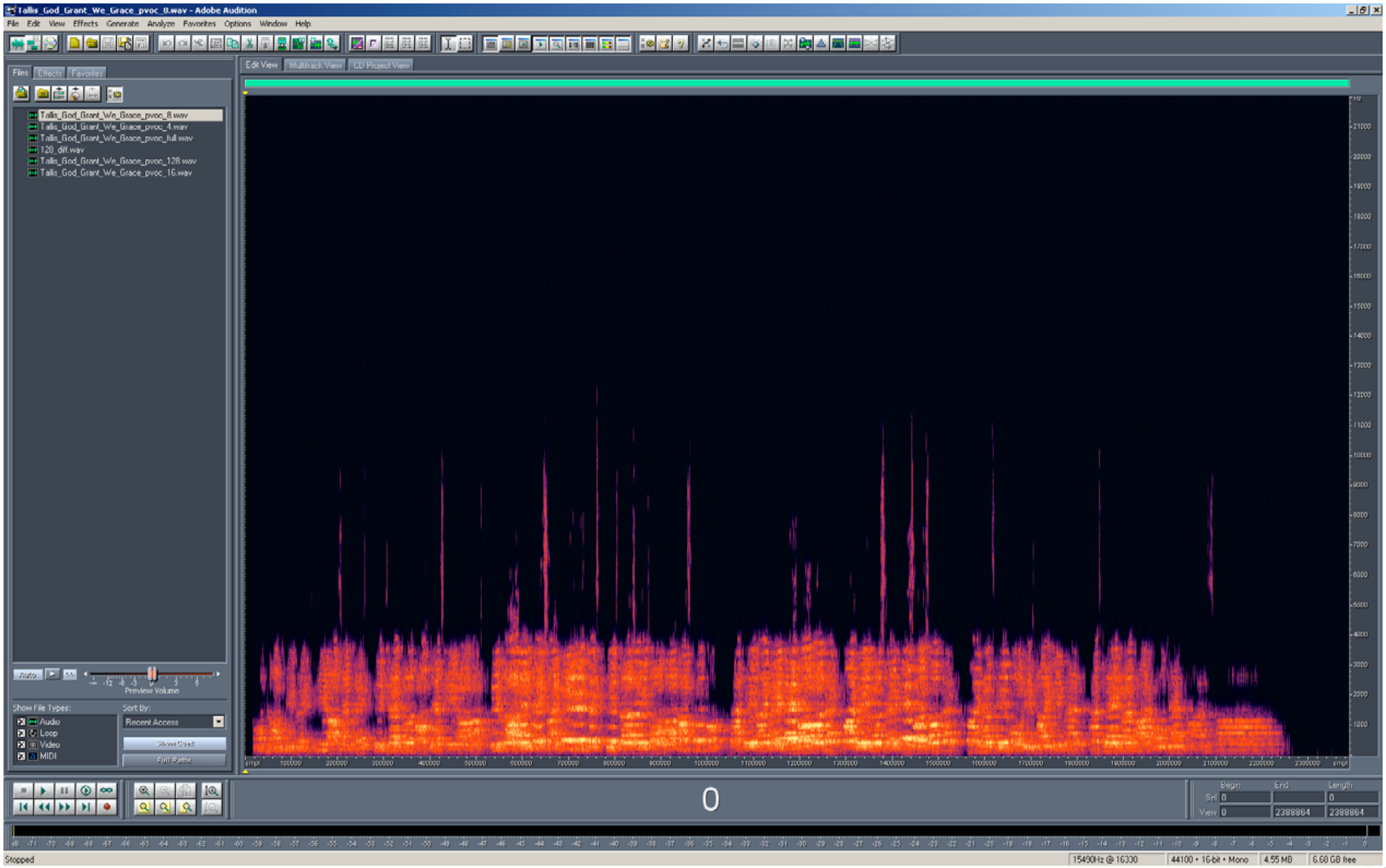
Spectrum, quantized to 1/128



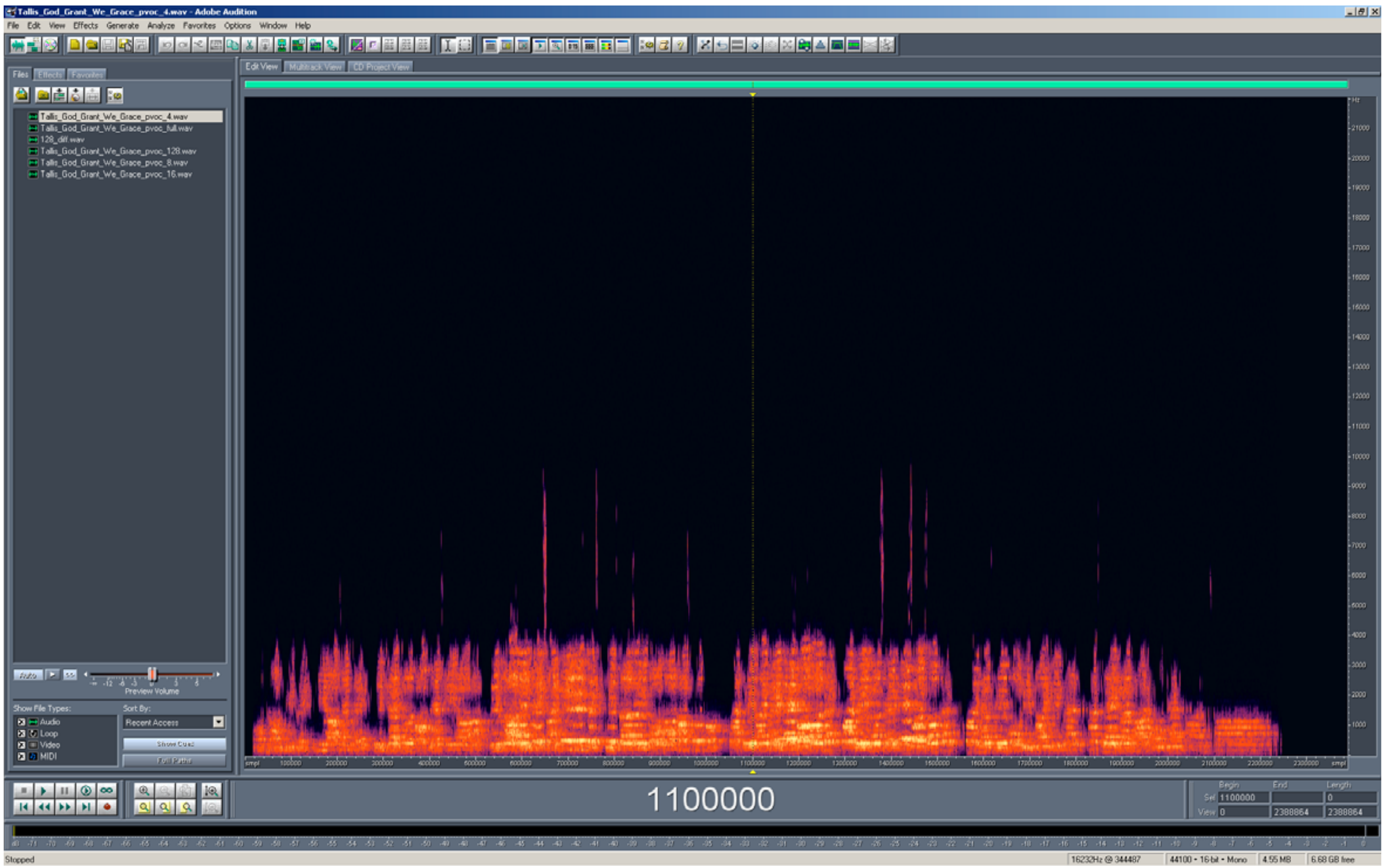
Spectrum quantized to 1/16



Spectra quantized to 1/8



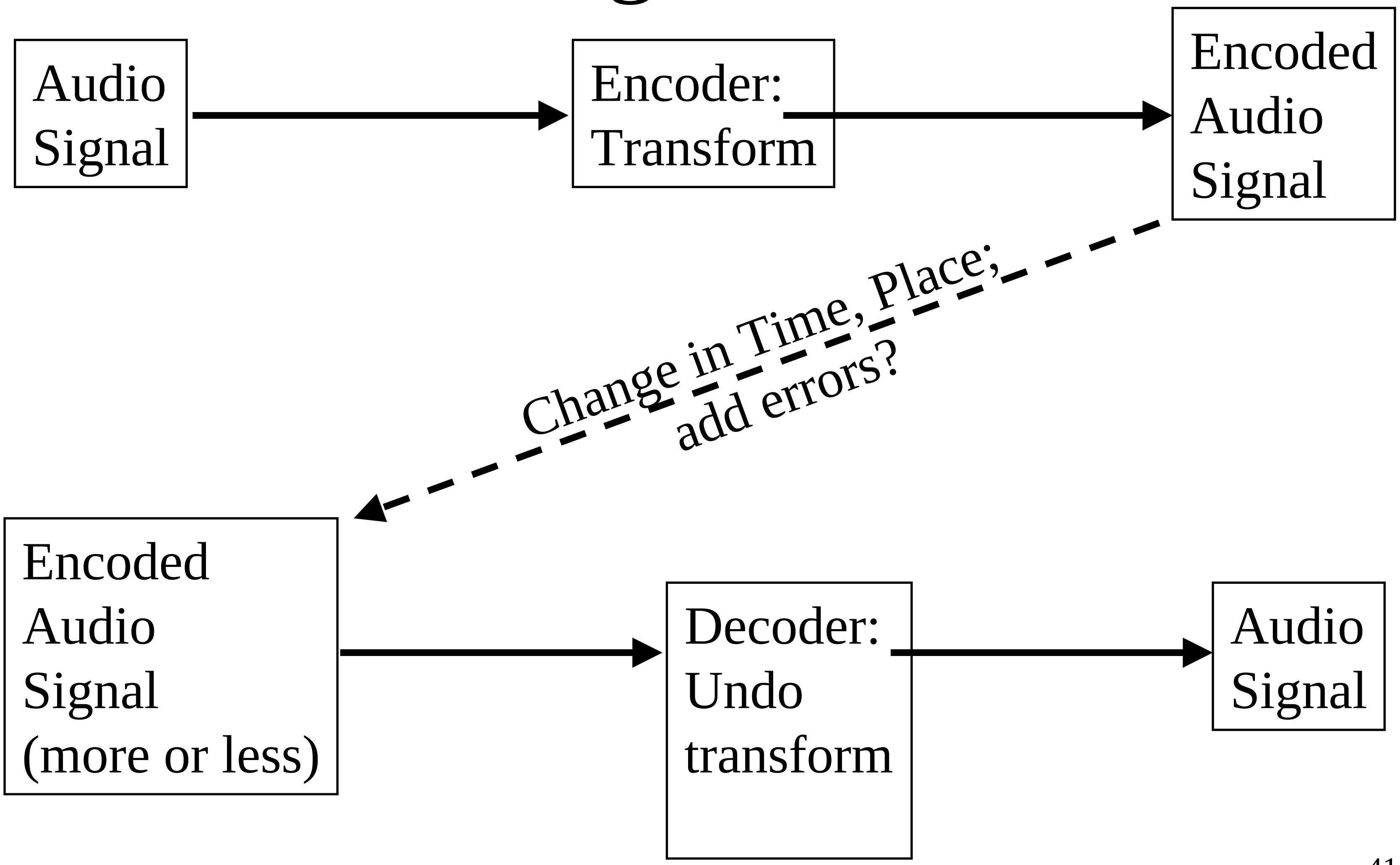
Spectra quantized to 1/4



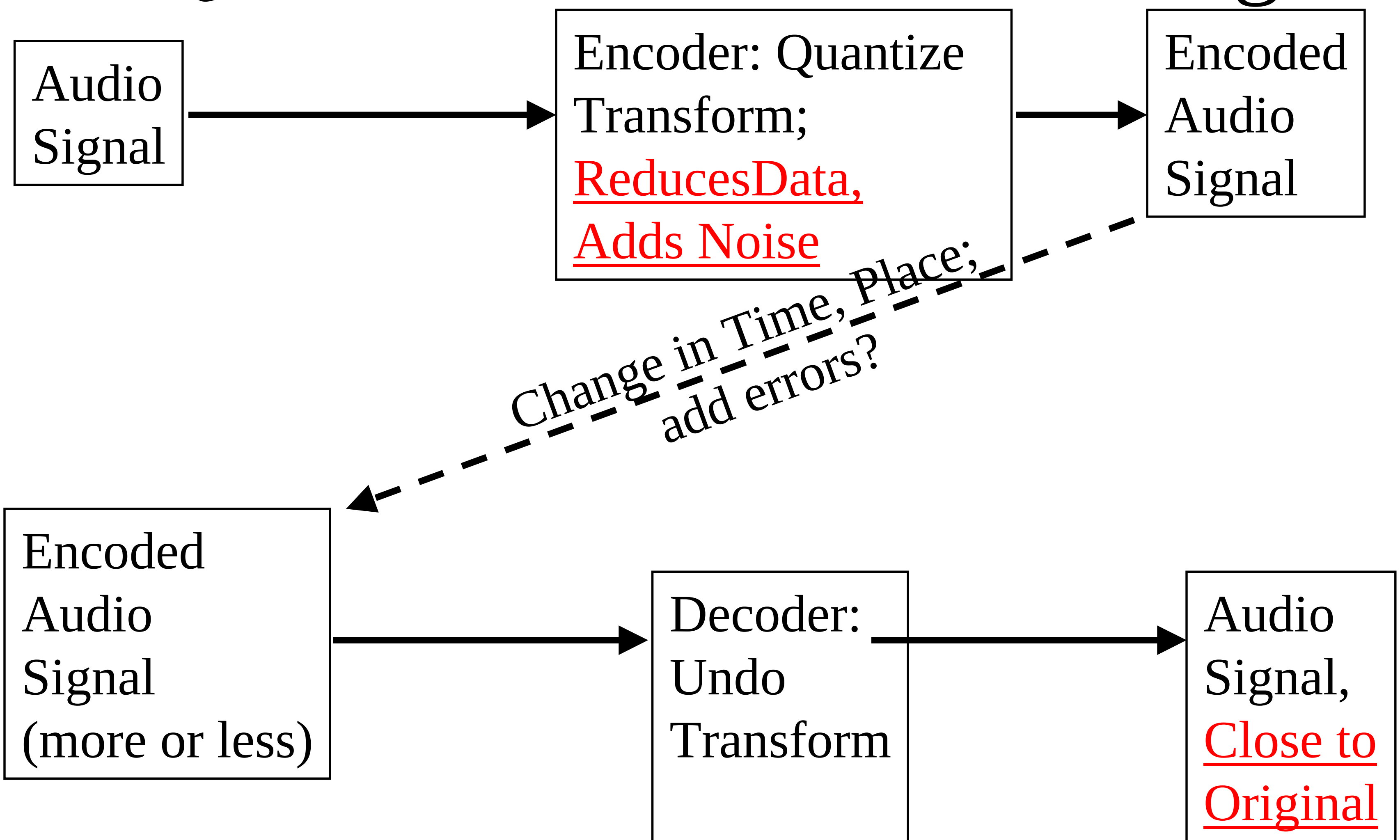
Tallis PVOC: What you Heard

- Tallis 4
- Difference, Tallis 4, original
- Tallis 8
- Tallis 16
- Tallis 128

Coding in General



Quantized Transform Coding



Quantized Transform Encoder

Time-Domain Audio Signal

Forward transform (enter frequency domain), get spectrum

Omit parts of signal (adds noise!)

Package compressed output (e.g. MP3 file)

Decoder (“mp3 player”)

Unpackage compressed audio (e.g. MP3 file)

Inverse Transform: change from frequency to time

Time-Domain
Audio Signal

So what's really in an MP3 file?

- And why do your MP3's sound better than those sound examples?

Meeting the challenge

- Coarser Quantization (time domain)
- DPCM, ADPCM
- Linear Prediction
- Subband coding
- Transform to frequency domain
- Coarser quantization (frequency domain)

What we have covered

- Subband Coding
- Transform Coding
- Structure of a simple encoder/decoder
- Quantizing transform coefficients

Lunch Break

- Optional math lecture tonight, DCT
- Bring your laptop tomorrow?
- Lend me your project for Sunday?
- Who can lend a DVD of *40-year-old Virgin?*

Deep Listening

- Meet by 13:00 in 285D, 285F, 285H, 285J.
- Each room has different tracks.
- Listen to 1st set of 4 tracks (next slide)
- Take notes and discuss among yourselves:
  - What do you hear?
  - How are they different?
  - What do you like?
  - What don't you like?
- Do likewise for 2nd set of 4 tracks (next slide)
- If you have time, go to another room.
- Back here at 13:30 to discuss. (half hour)

Listening Examples in Rooms

- Aida [NOT: aida_stereo]
- Dinah

- Vbrtest (Lou Reid)
- Akarui

- Vega
- Money

Notes on Listening Examples: First Recording

- 0
- 1
- 2
- 3

- Was one of them the original?

Notes on Listening Examples: Second Recording

- 0
- 1
- 2
- 3

- Was one of them the original?

Notes on Listening Examples: 2nd Room, First Recording

- 0
- 1
- 2
- 3

- Was one of them the original?

Notes on Listening Examples: 2nd Room, Second Recording

- 0
- 1
- 2
- 3

- Was one of them the original?

MSRA 5500 - 002
MUS 4500 - 002

March 14, Session 2, Part 1

Frequency-domain Aspects

© Copyright 2008 John Strawn

Discussion of Listening examples

	0	1	2	3
Aida				
Dinah				
Lou Reid				
Akarui Tsuki				
Money				
Vega				

Listening Examples

- **Ben Heppner, *Great Tenor Arias,* Munich Radio Orchestra, conducted by Roberto Abbado, 09026-62504-2, 1995, BMG Music, Approximately one minute from Verdi's *Aida*, "Celeste Aida", track 4.**
- **"Dinah won't you blow...", Justin Strawn, age 12, harmonica solo, recorded 2002.**

- **Lou Reid's album *Lou Reid and Carolina,* opening (?) of "God loves his children." Rebel Records, 1966. From http://lame.sourceforge.net/gpsycho/.**
- **John Strawn, *akarui tsuki*, (1975), final 1.4 min. Score in Electronic Music Systems, Techniques and Controls by Allen Strange; Computer Music Journal CD #21**

- **Pink Floyd, *Dark Side of the Moon*, Track 5: "Money", First 54 seconds.**
- **Vega: "Tom's Diner," *Solitude Standing*, 3 verses starting with 2nd verse. See Tom's Diner in Wikipedia; known as "Mother of MP3."**

RealNetworks Data Rates (stereo)

Transmission medium	Max	How far off from real time?
28.8 kbps modem	20 kbps	71
56 kbps modem	32 kbps	44
112 kbps dual ISDN	64 kbps	22
Corporate LAN	132 kbps	11
256 kbps DSL/cable modem	176 kbps	8
512 kbps DSL/cable modem	352 kbps	4

(0)
(1): 56
(3): 80

Derived from: RealNetworks, http://service.real.com/help/library/guides/production8/htmfiles/audio.htm, retrieved 4 Feb 2008

MP-3 Quality vs. bit rate: one view

Bit-rate	Mode	Quality	
8 kbps	Mono	Telephone	
16 kbps	Mono	Short-wave Radio	
32 kbps	Mono	AM Radio	(0): 32
64 kbps	Stereo	FM Radio	(1): 56
128 kbps	Stereo	Near CD	(3): 80
256 kbps	Stereo	Equal to CD	

From: http://www.teamcombooks.com/mp3handbook/16.htm, retrieved February 2008;
Click on “16” at http://www.mp3handbook.com/.

Who will contribute a recording?

- Your project
- Used for Sunday morning listening
- I’ll modify (differently)

What we will cover

- Perceptual coding
- Psychoacoustics
- Range of hearing
- Masking (simultaneous)
- More on structure of simple encoder/decoder
- Critical Bands
- Variable, Constant Bit Rate
- “Noiseless” Coding

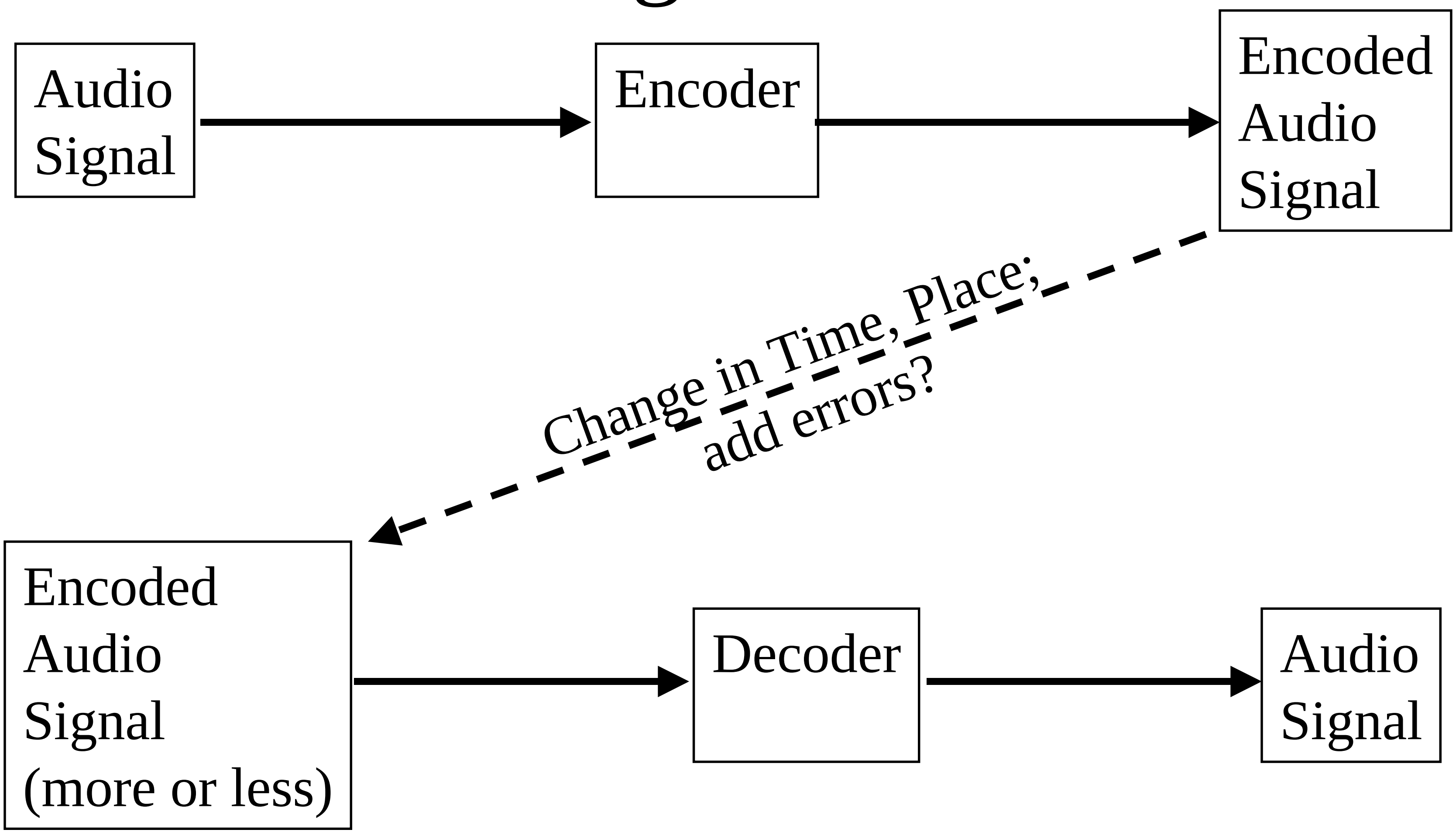
Two broad classes of coders

- Non-perceptual
- Perceptual

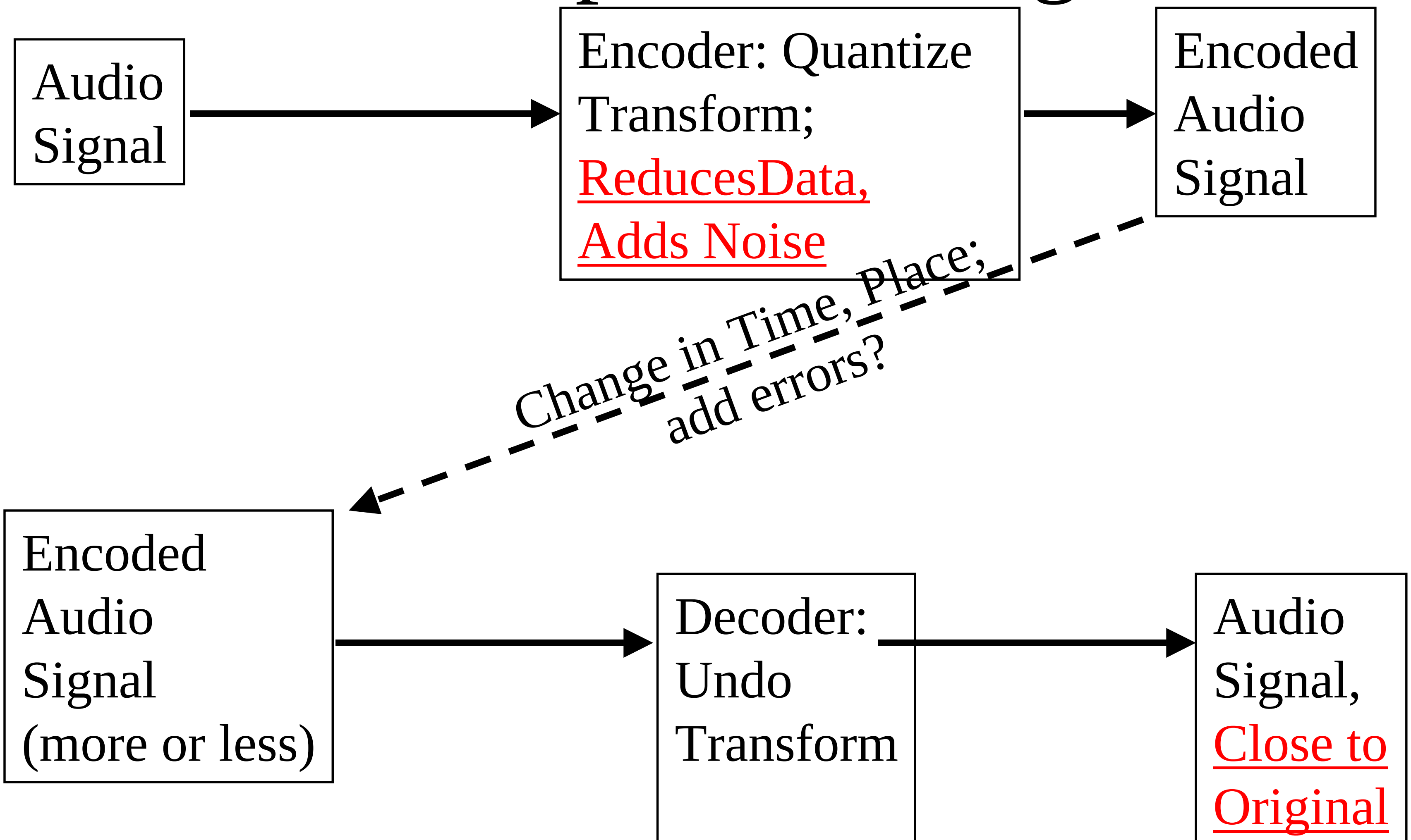
Fundamentals of perceptual coding

- Leave out
  - Irrelevant (easier in frequency domain)
  - Redundant
- Allow noise
  - But only in special places

Coding in General



Perceptual Coding



Perceptual Encoder

Time-Domain Audio Signal →

Forward transform (enter frequency domain), get spectrum

↓

Omit parts of signal (may add noise)

↓

↓

↓

Package compressed output (e.g. MP3 file)

Decoder (“mp3 player”)

Unpackage compressed audio (e.g. MP3 file)

Inverse Transform: change from frequency to time

Time-Domain Audio Signal

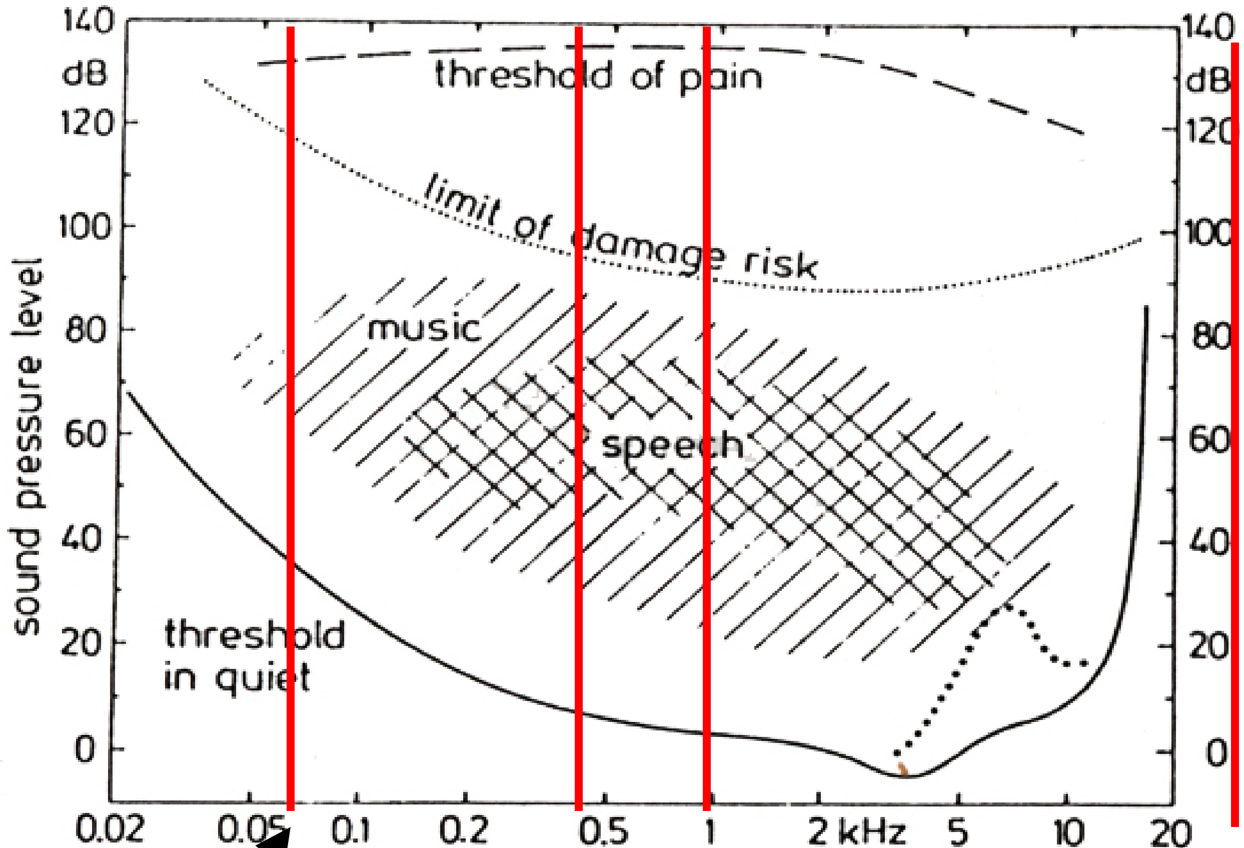
What we will cover

- Perceptual coding
- Psychoacoustics
- Range of hearing
- Masking (simultaneous)
- More on structure of simple encoder/decoder
- Critical Bands
- Variable, Constant Bit Rate
- “Noiseless” Coding

What is a Psychoacoustics?

Range of human hearing

Dental Plaque Removal



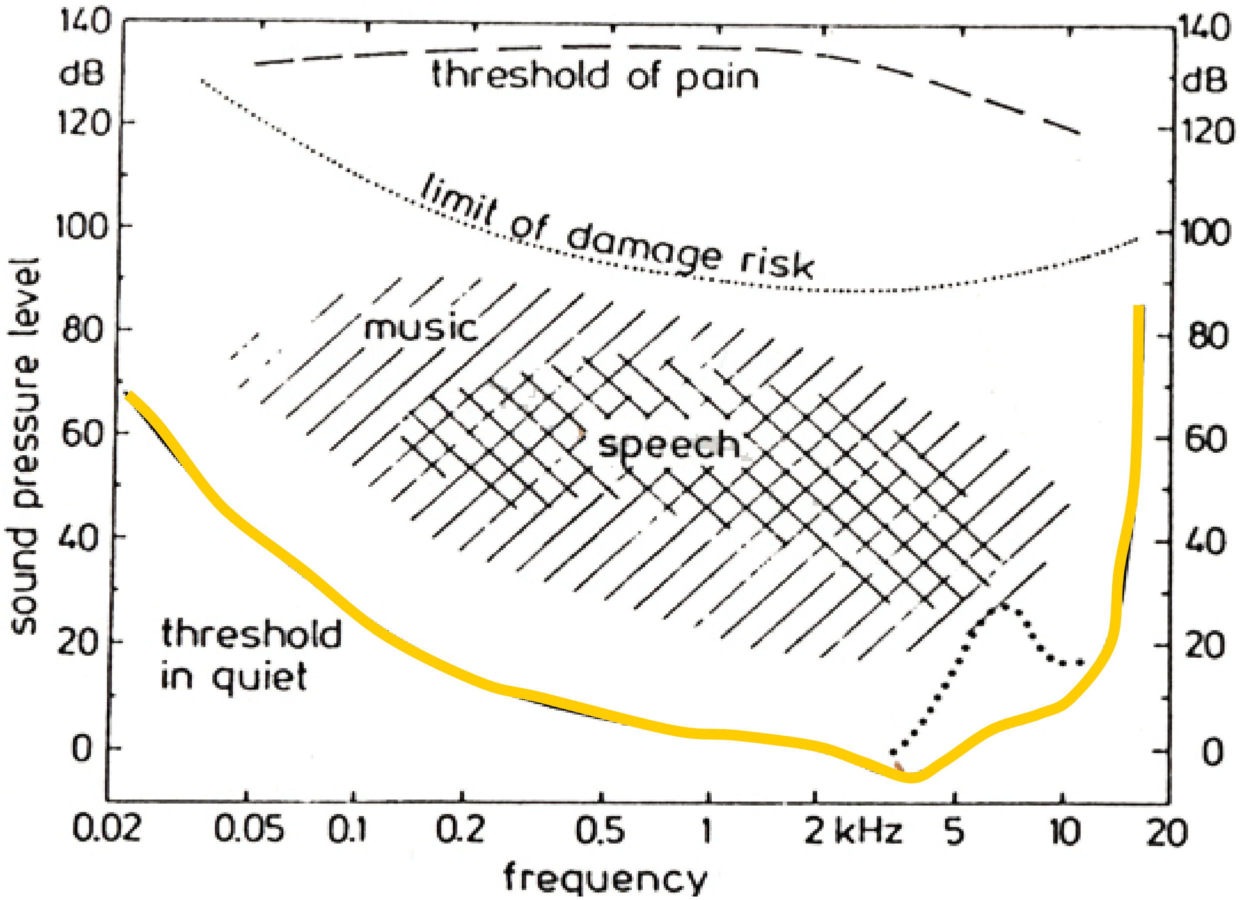
A440

Cello Low C

Soprano High C

Zwicker/Fastl p. 17

Range of human hearing



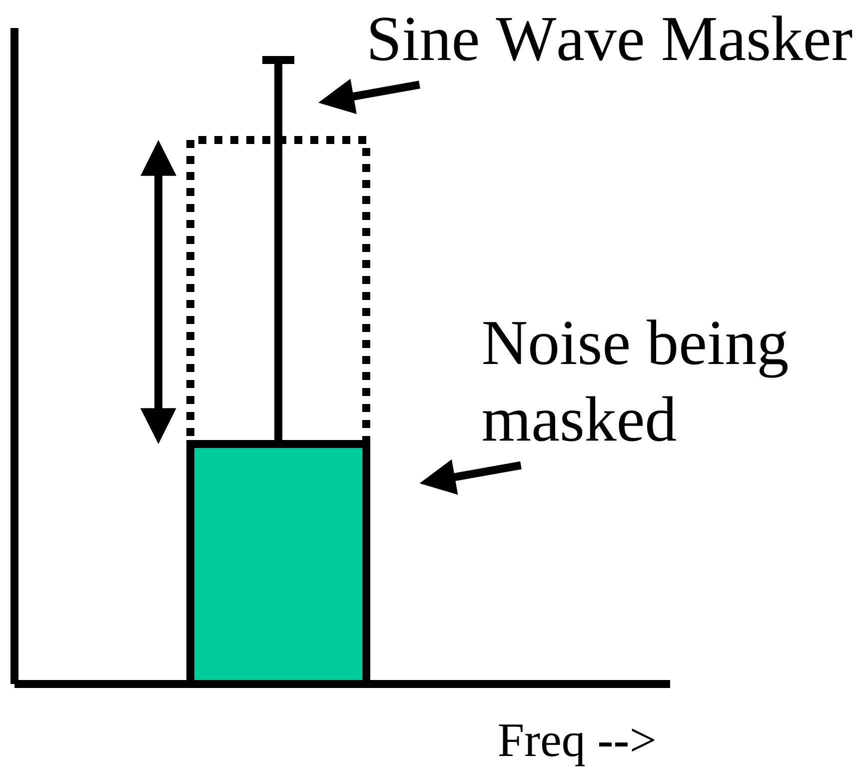
Zwicker/Fastl p. 17

What we will cover

- Perceptual coding
- Psychoacoustics
- Range of hearing
- Masking (simultaneous)
- More on structure of simple encoder/decoder
- Critical Bands
- Variable, Constant Bit Rate
- “Noiseless” Coding

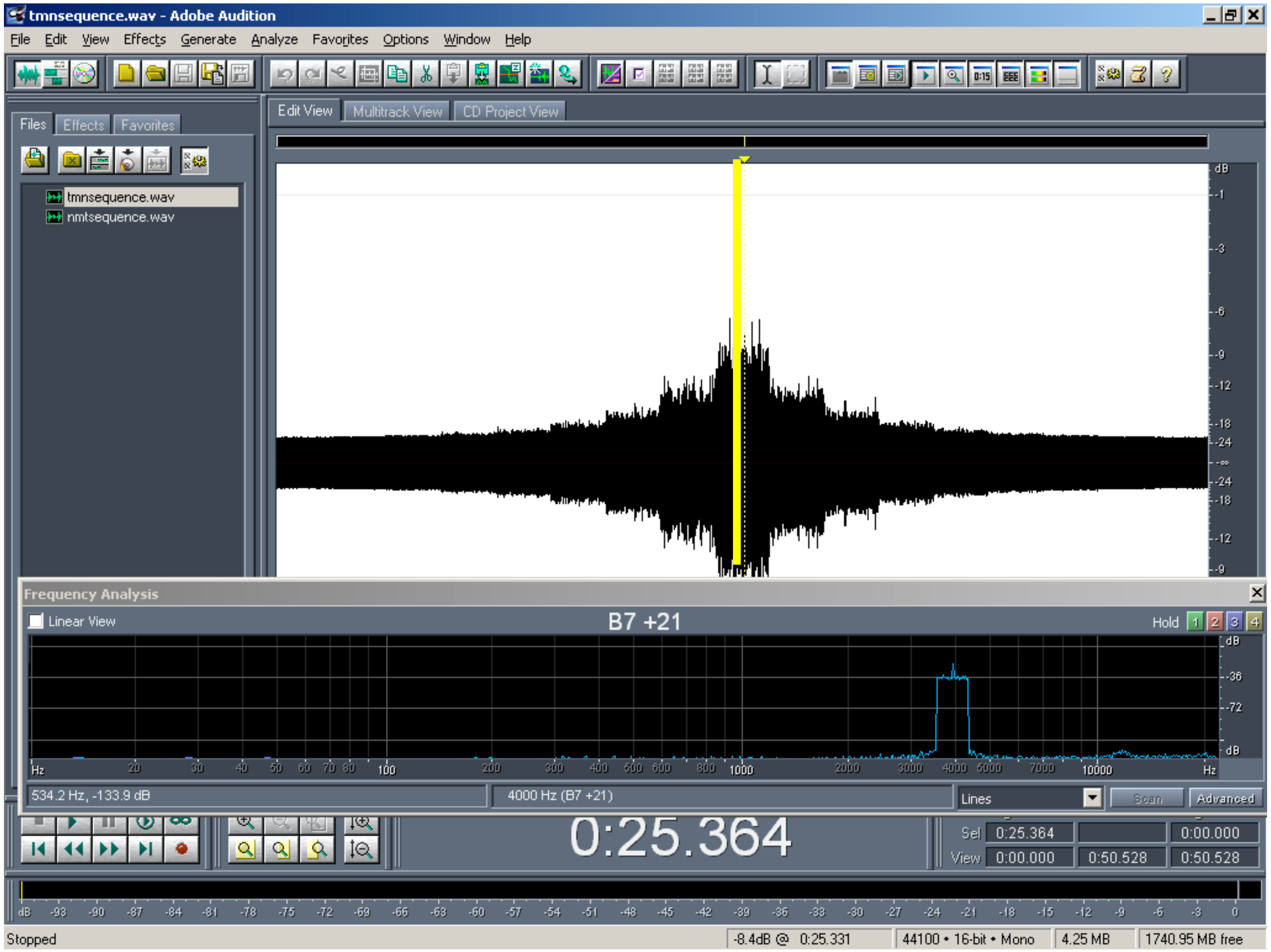
Sound example: Masking

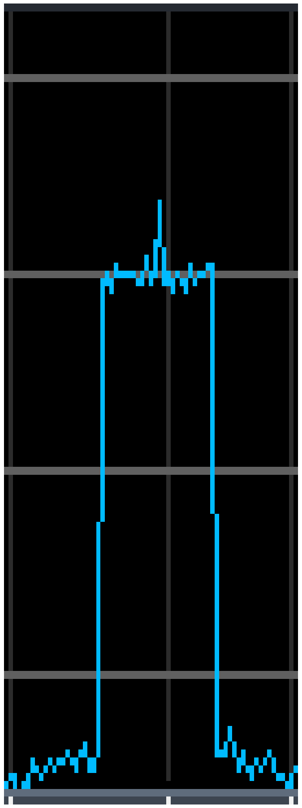
- AES Technical Committee on Signal Processing (forthcoming CD)
- Sine 4 kHz
- Noise band, level varies



After Painter and Spanias, p. 459

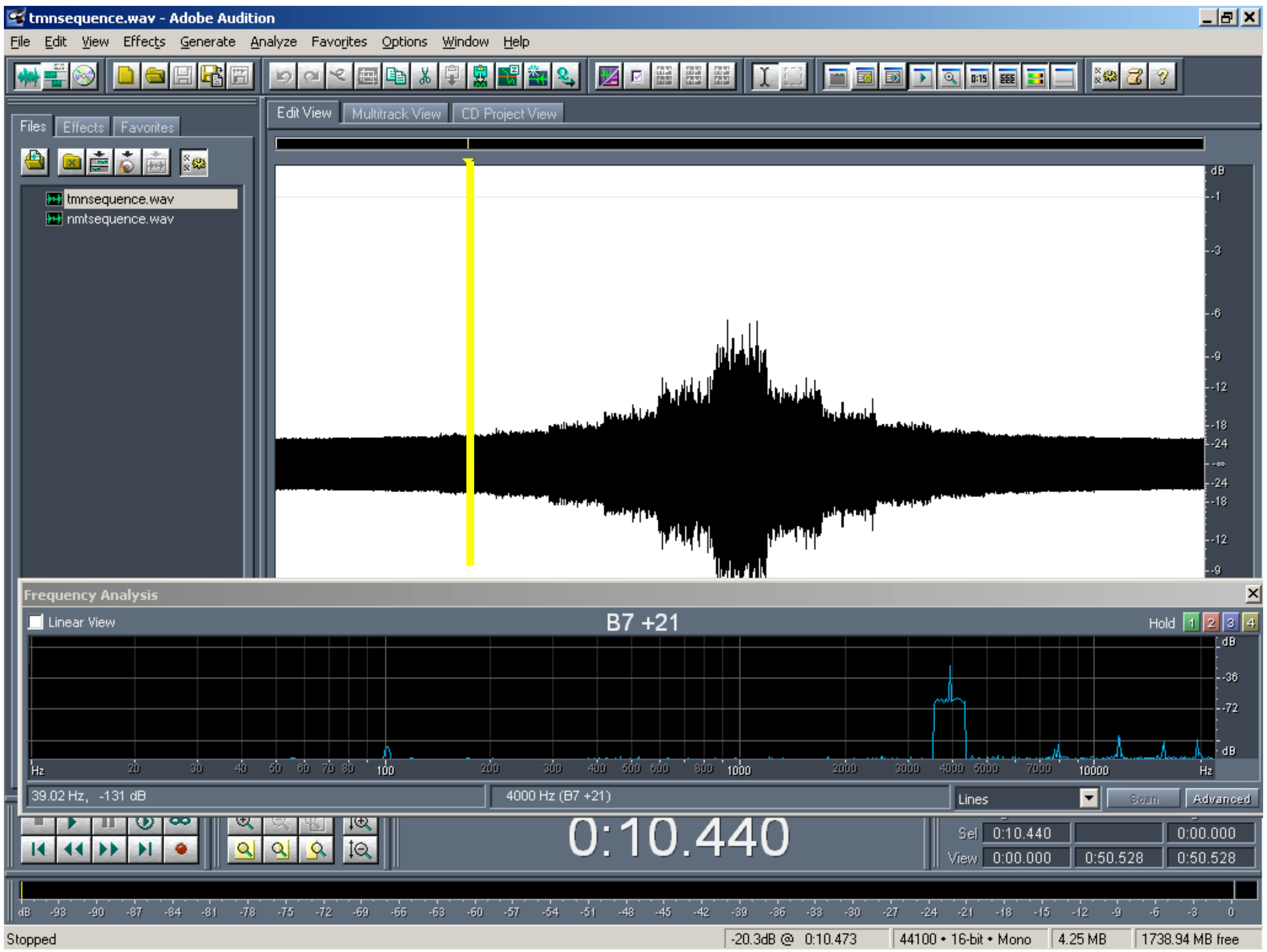
Sine Not Mask Noise

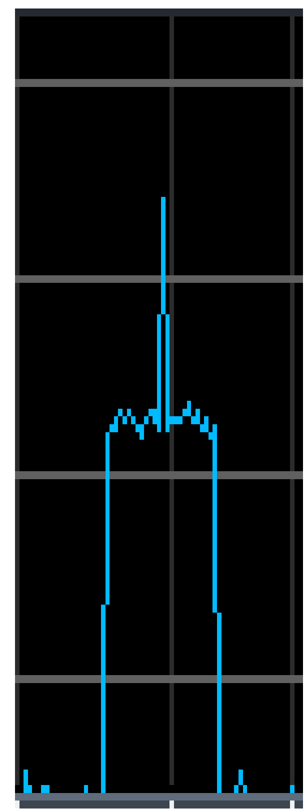




From forthcoming AES CD-ROM

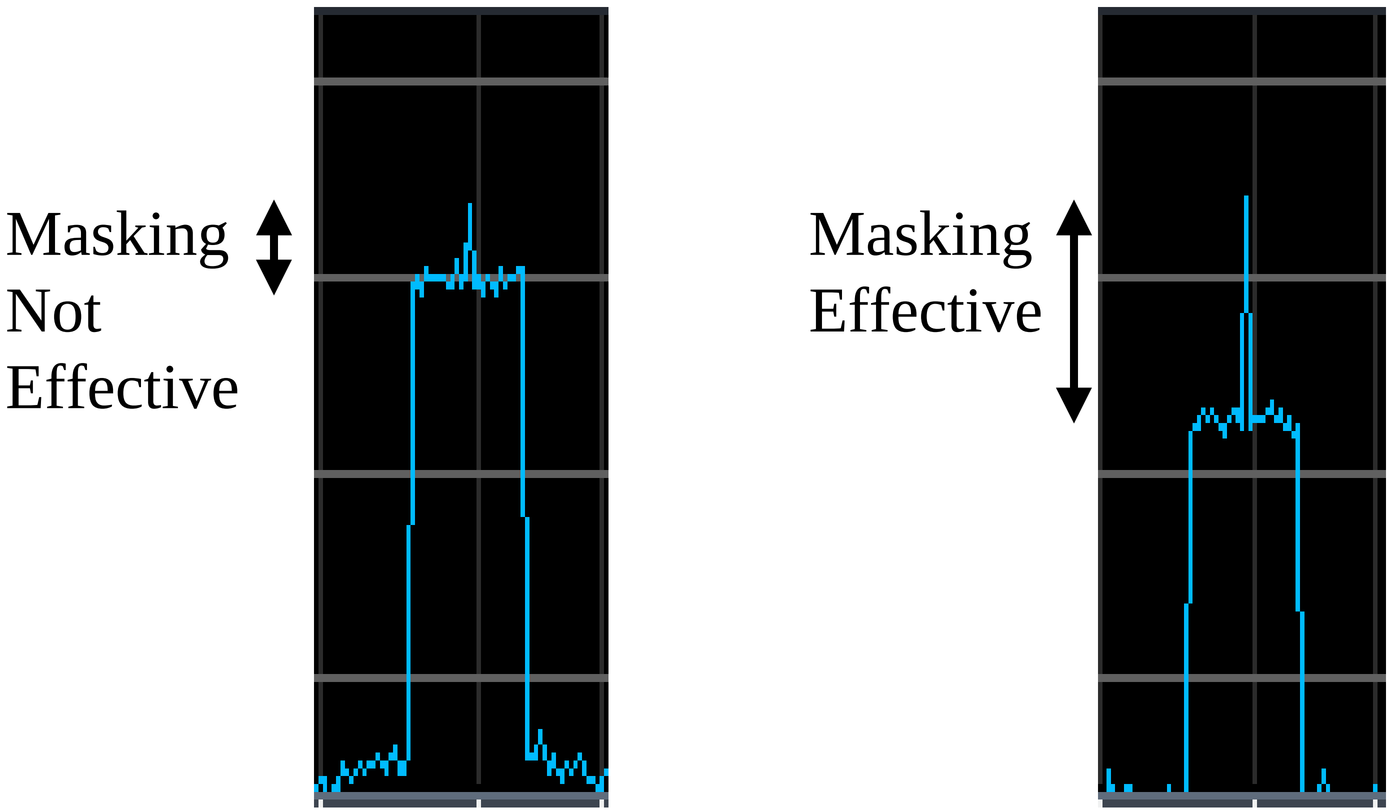
Sine Masks Noise



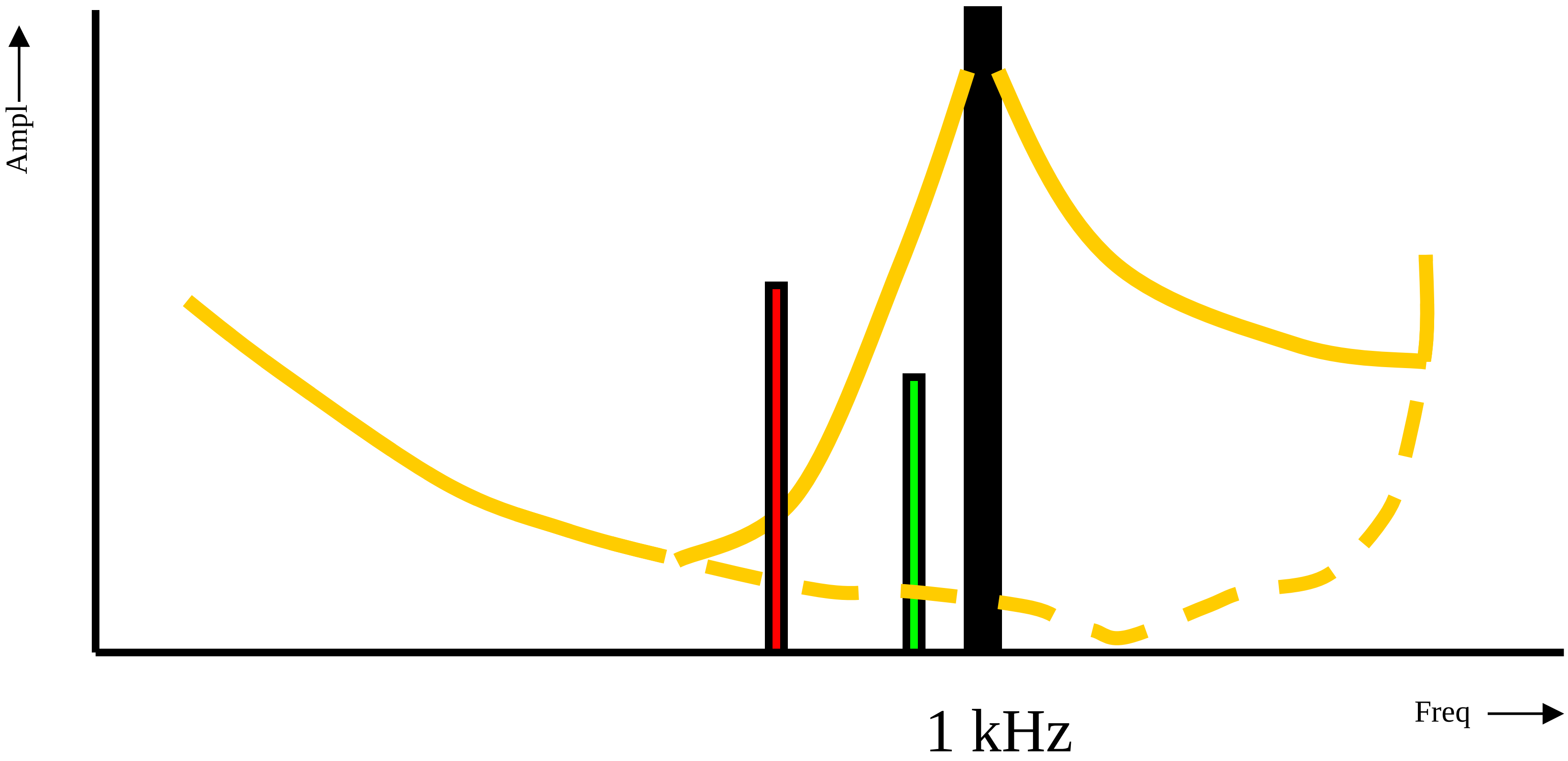


From forthcoming AES CD-ROM

When Sine Mask Noise?

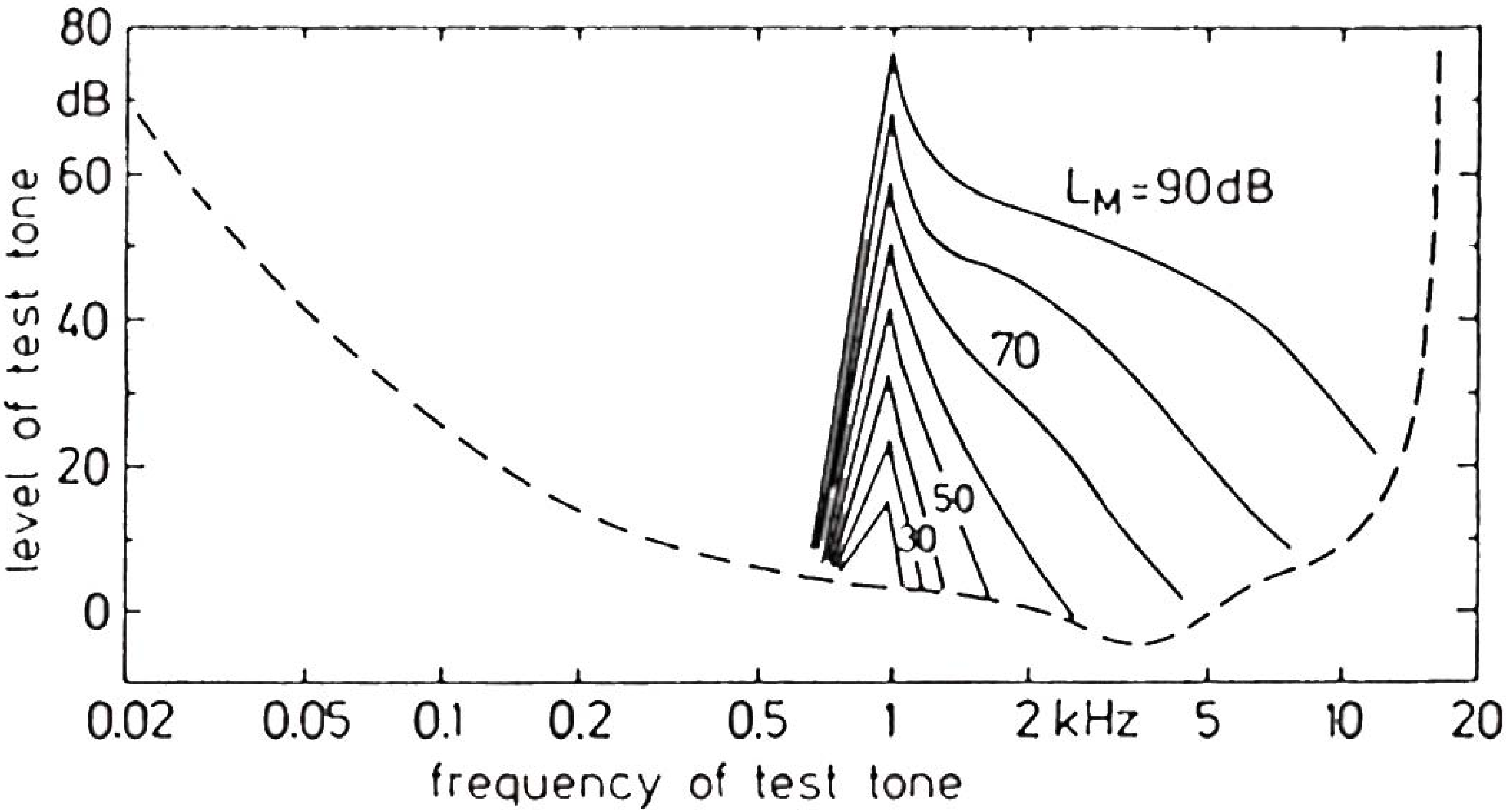


Simultaneous Masking I



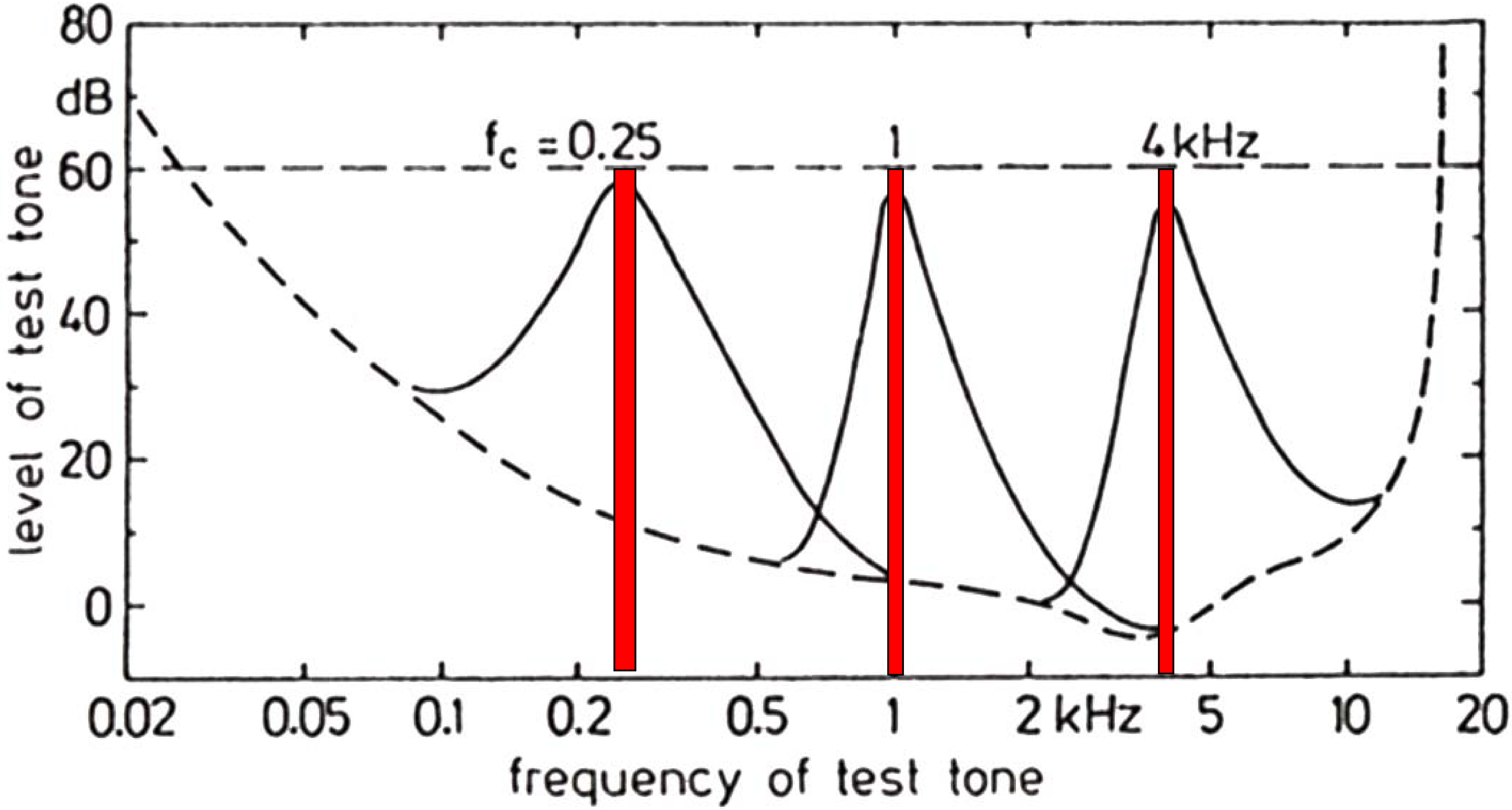
After Zwicker/Feldtkeller p. 60

Simultaneous Masking II: Amplitude affects shape



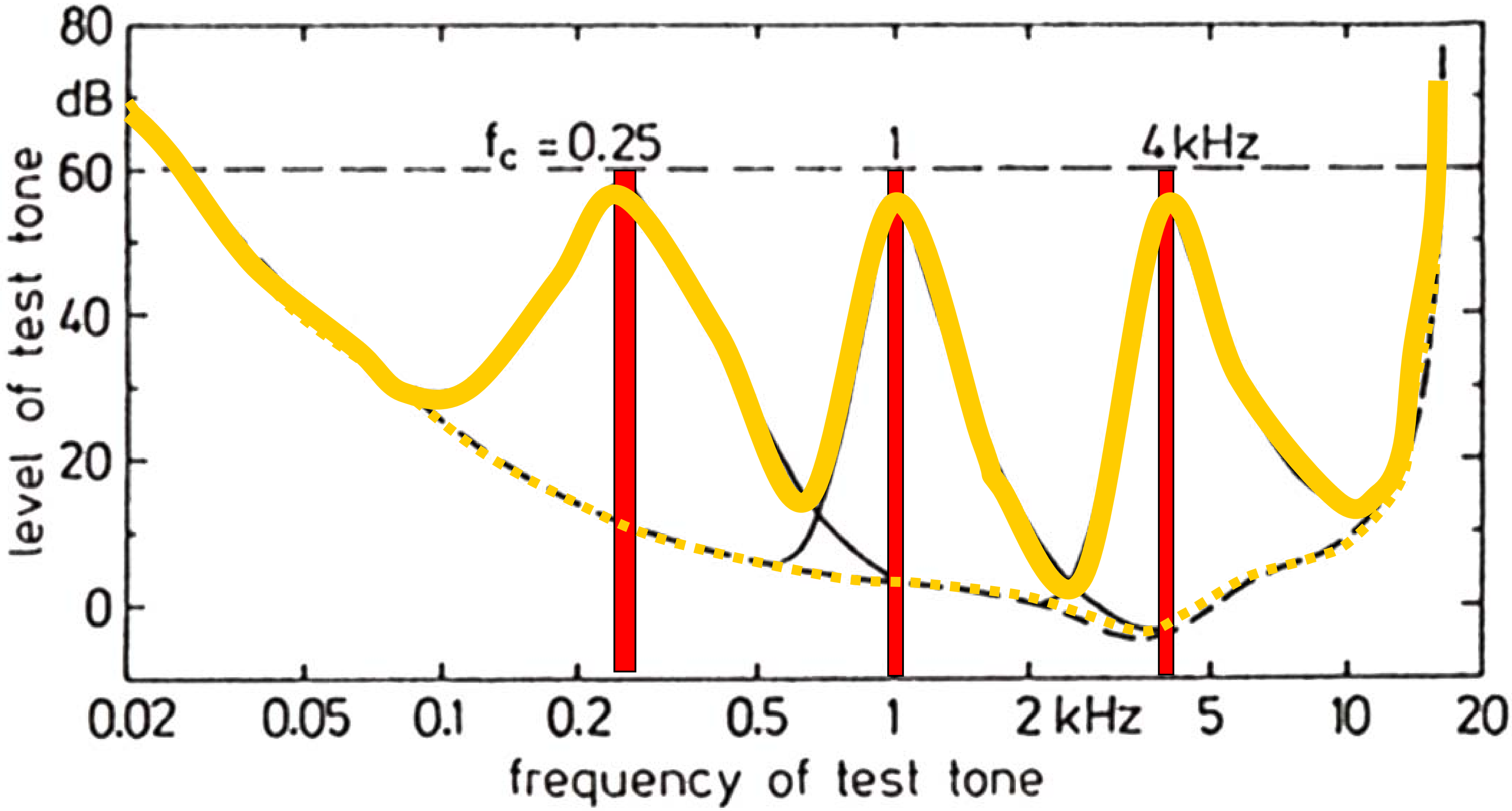
Zwicker/Fastl, p. 68

Simultaneous Masking III: Frequency affects shape



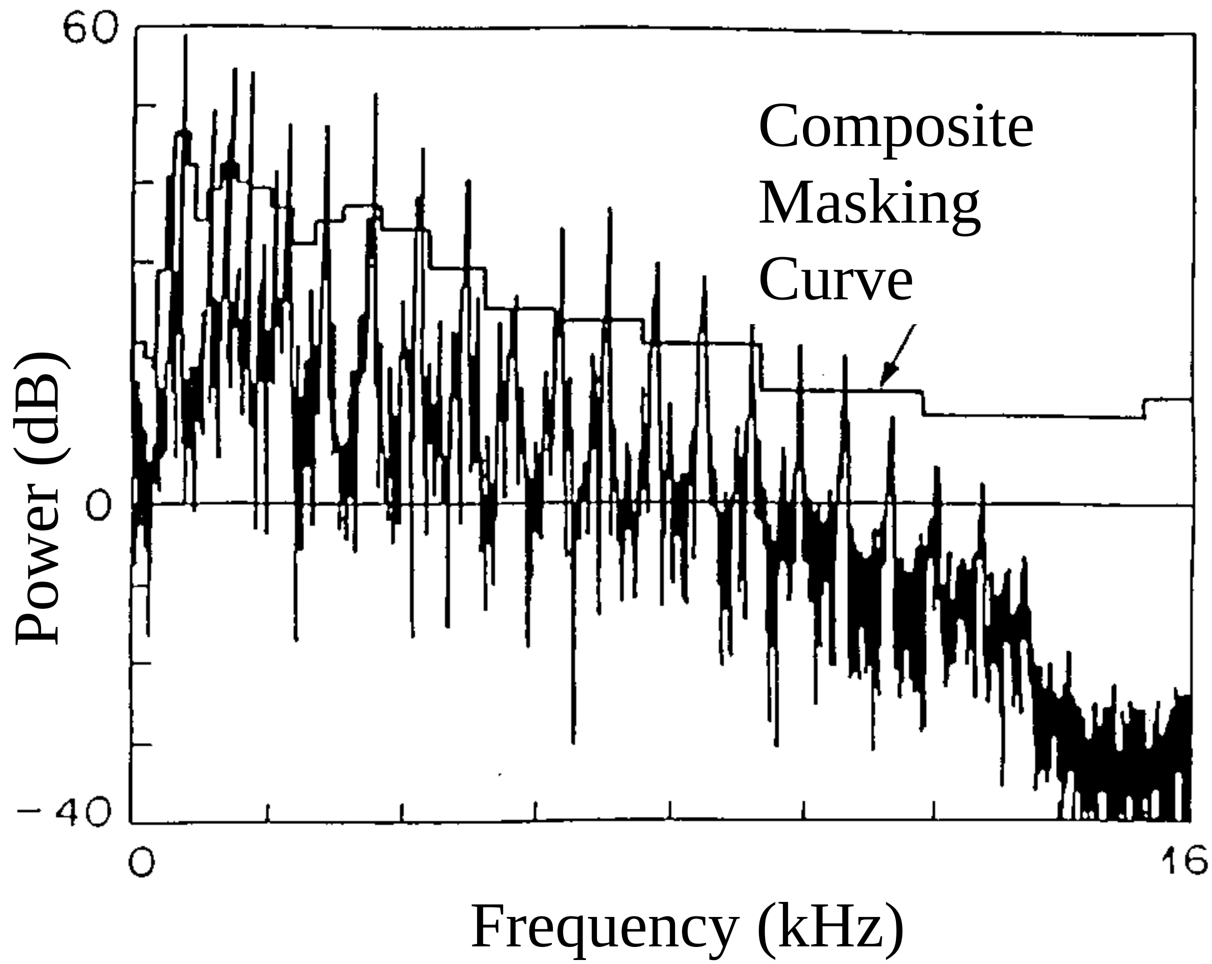
Zwicker/Fastl, p. 64

Simultaneous Masking IV: Composite Masking Curve



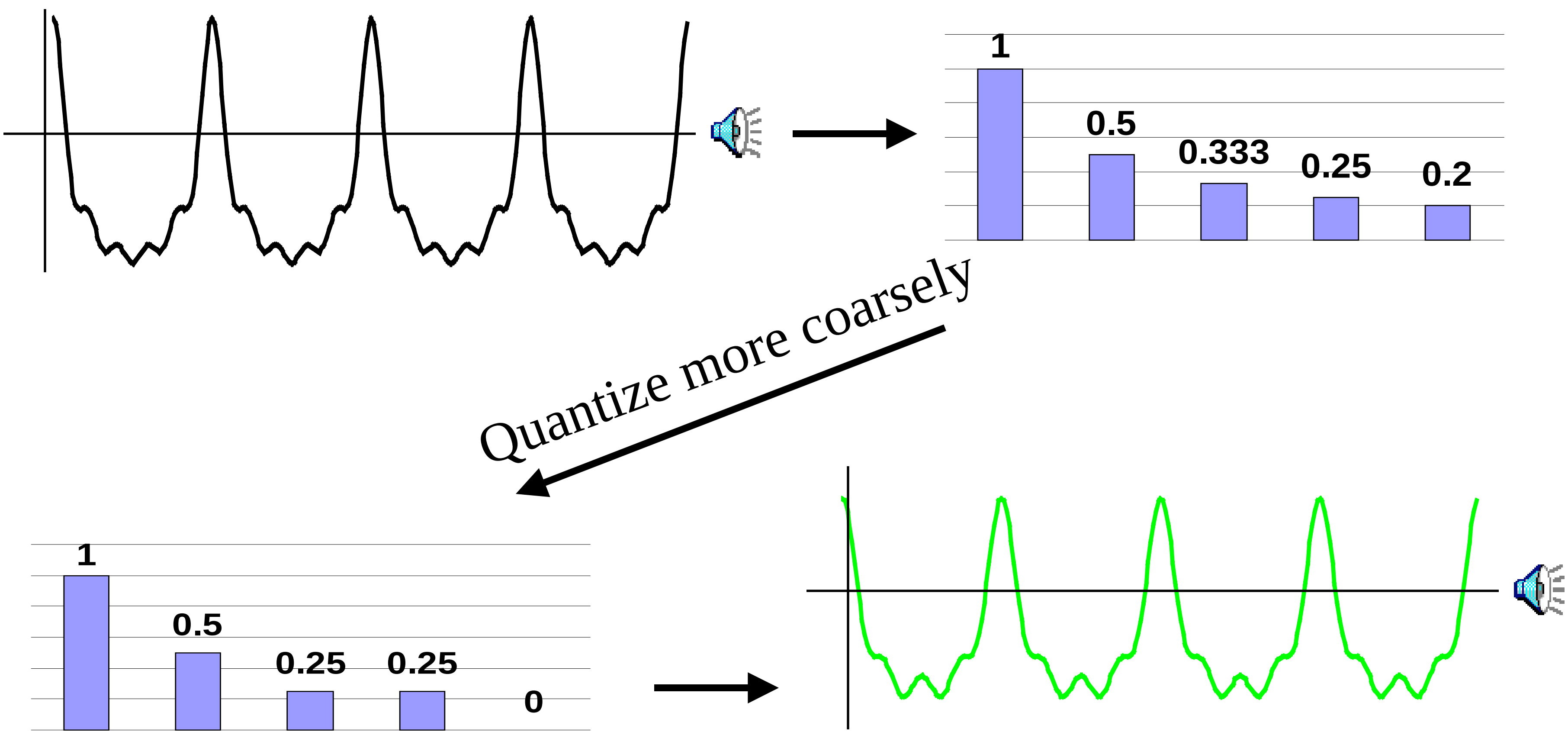
Zwicker/Fastl, p. 64

Composite Masking Curve (PAC)

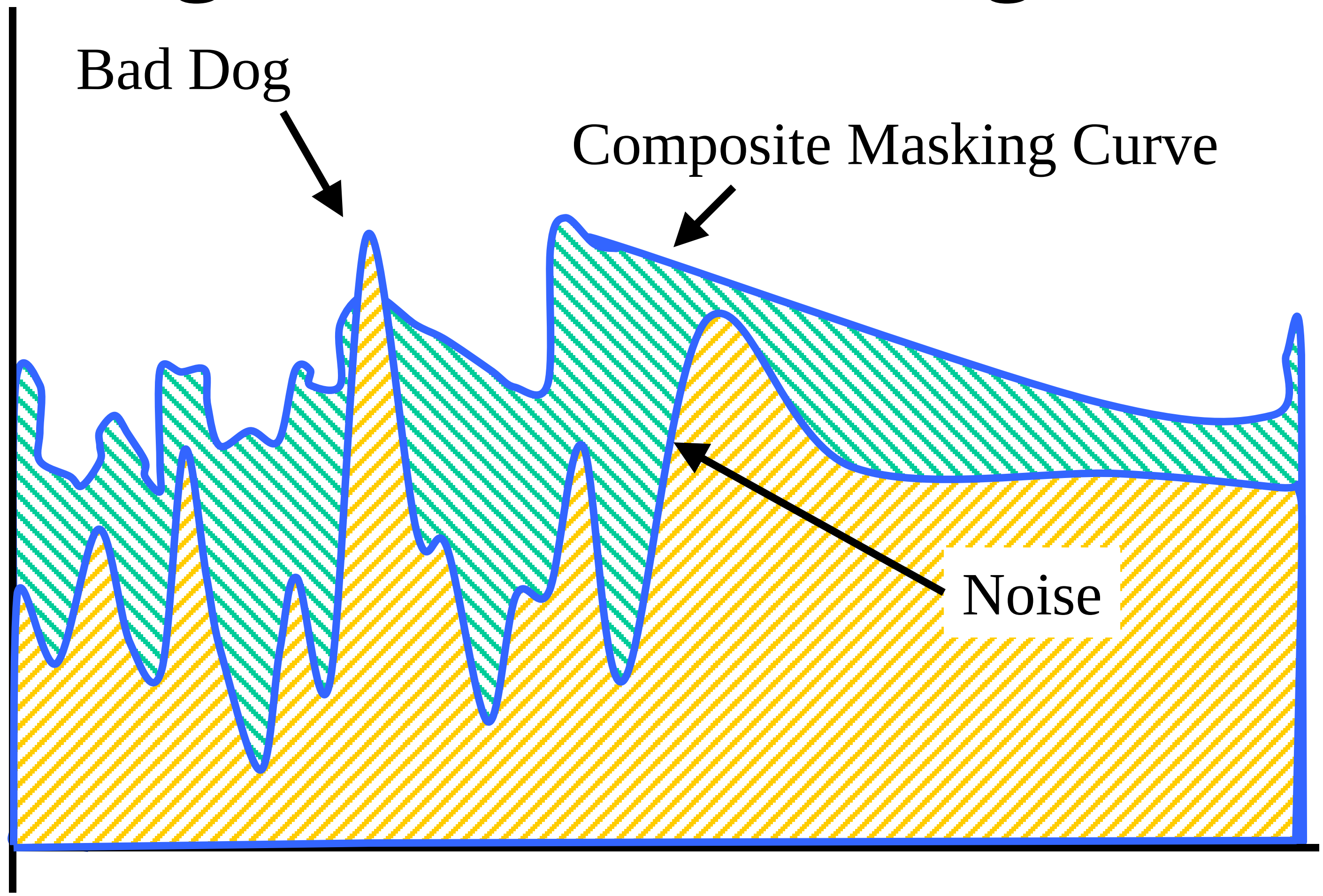


From Johnston et al., AES Collected Papers on Digital Audio Bit-rate Reduction, p. 76

Review: Leave something out -> Noise

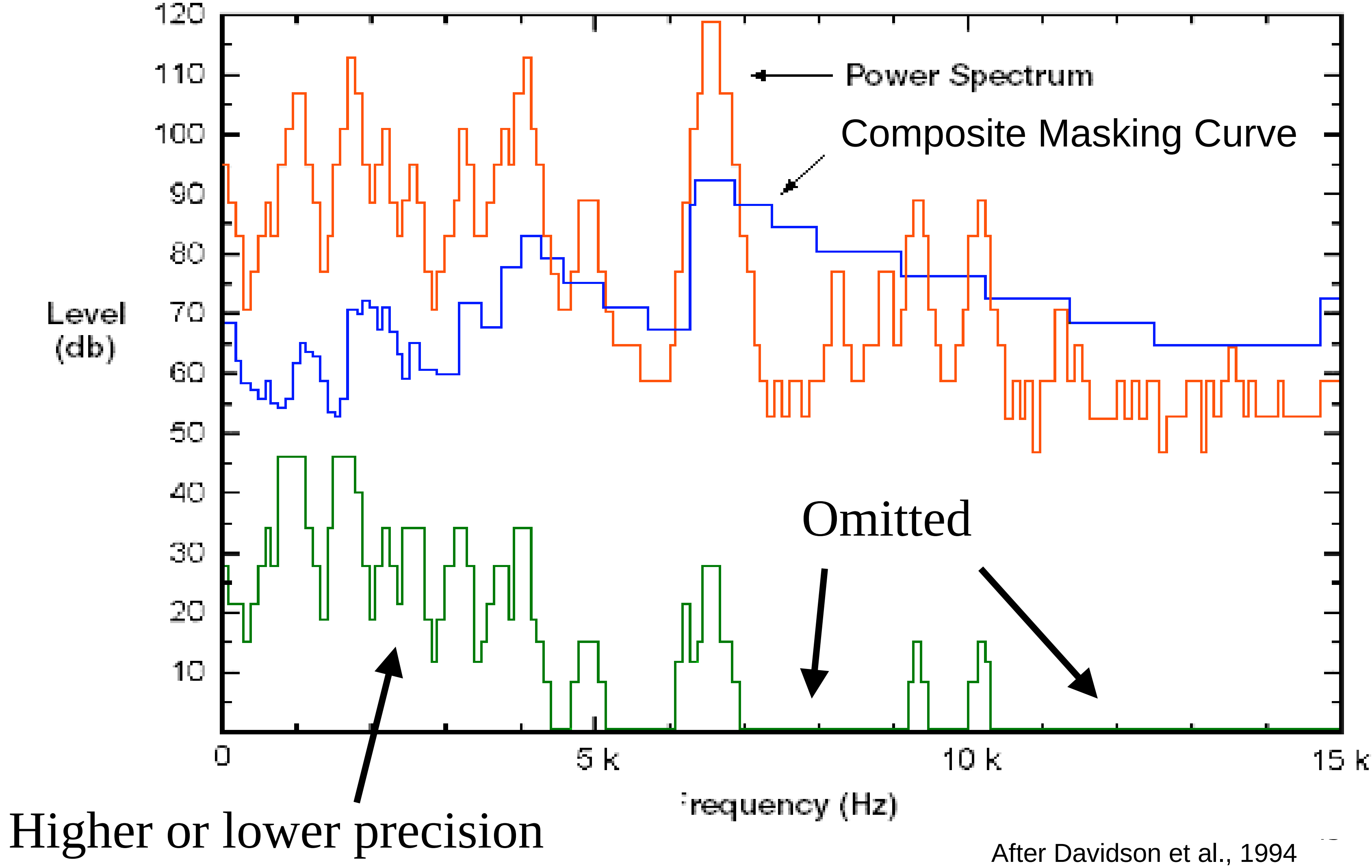


Masking: Where noise might fall



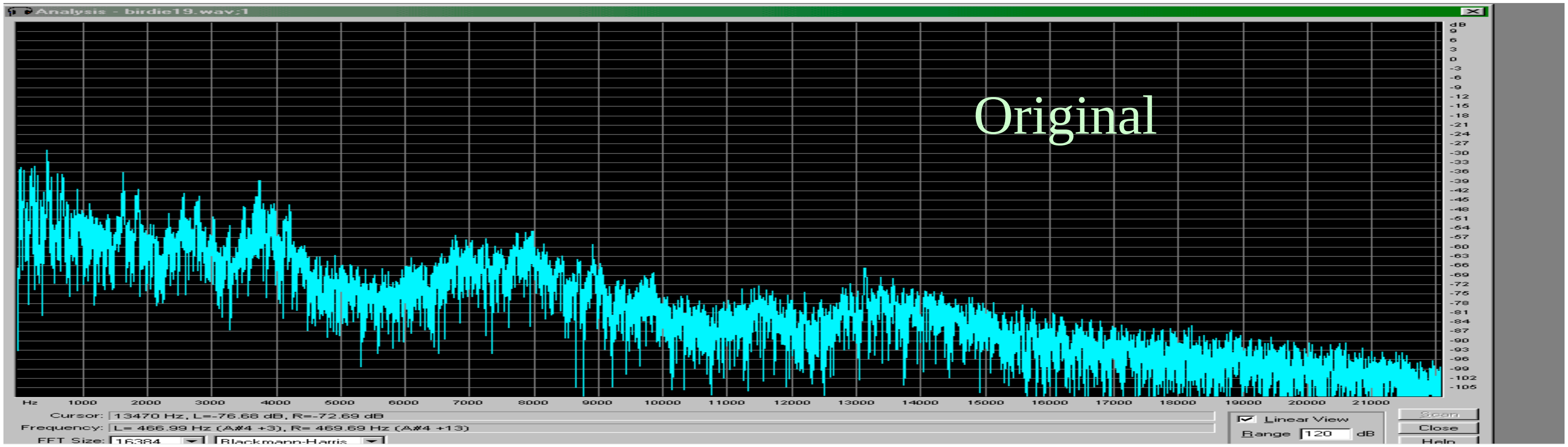
After Davidson et al., 1994

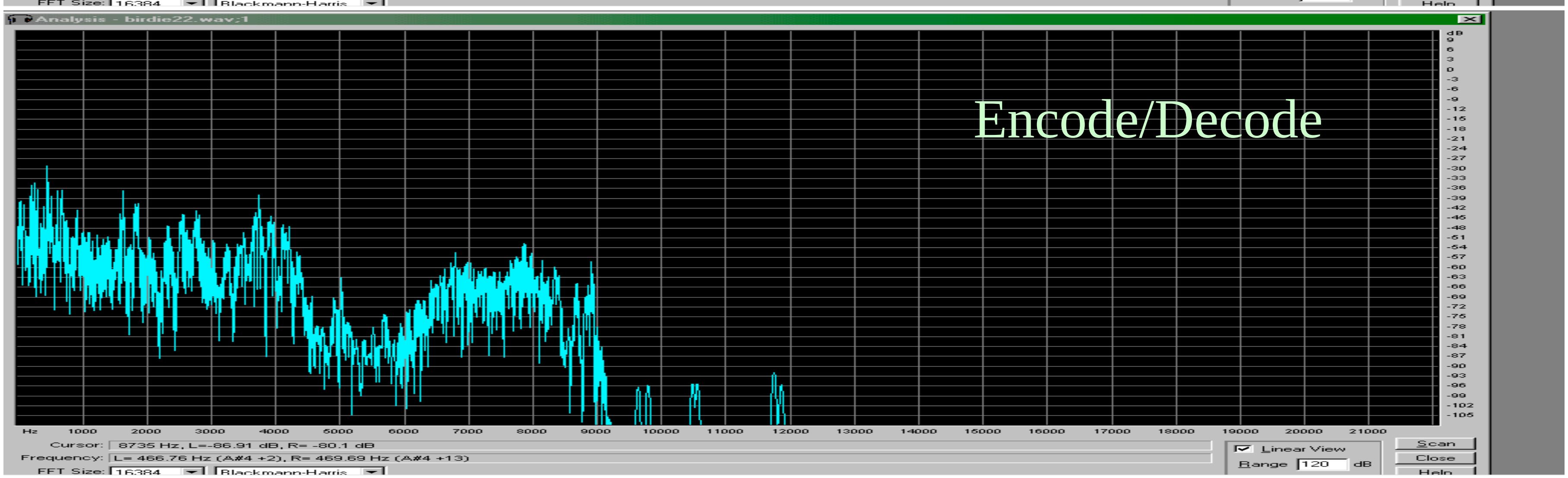
Masking: Leave Something Out



After Davidson et al., 1994

Spectrum modified by compression (Exaggerated)





Demonstration: making it all fit

Perceptual Encoder

Audio Signal

Forward transform, get spectrum

Omit parts of spectrum (add noise)

How accurately to represent remaining spectrum (add noise)

Noise below masking?

N

Y

Package compressed output (e.g. MP3 file)

Decoder (“mp3 player”)

Unpackage compressed audio (e.g. MP3 file)

Inverse Transform: change from frequency to time

Time-Domain
Audio Signal

So what's really in an MP3 file?

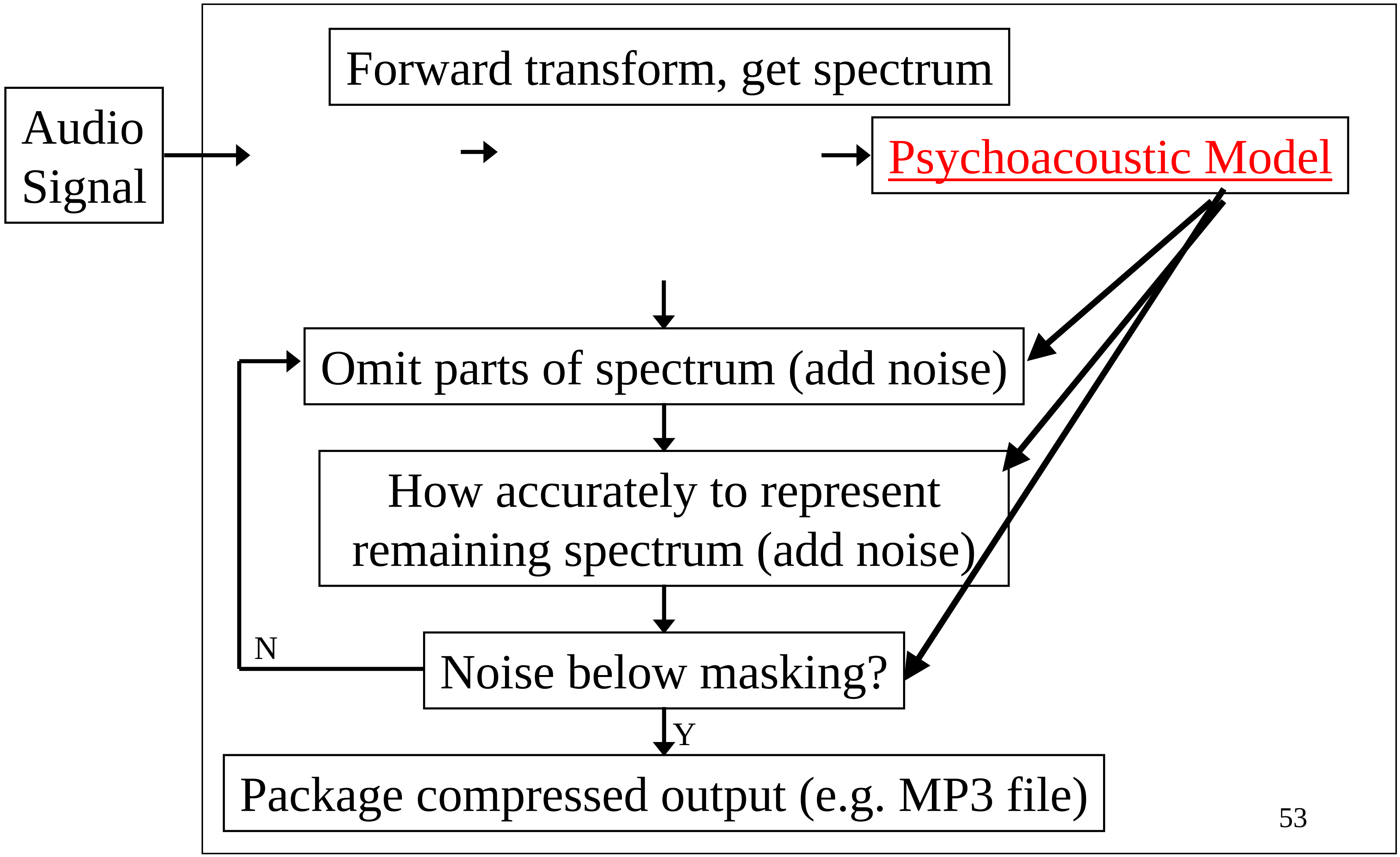
- (again)
- Why were there still artifacts in the sound examples? (Aida, Dinah, Money ...)

What we will cover

- Perceptual coding
- Psychoacoustics
- Range of hearing
- Masking (simultaneous)
- More on structure of simple encoder/decoder
- Critical Bands
- Variable, Constant Bit Rate
- "Noiseless" Coding

The Psychoacoustic Model

Perceptual Encoder



Decoder ("mp3 player")

Unpackage compressed audio (e.g. MP3 file)

↓

Inverse Transform: change from frequency to time

↓

↓

Time-Domain Audio Signal

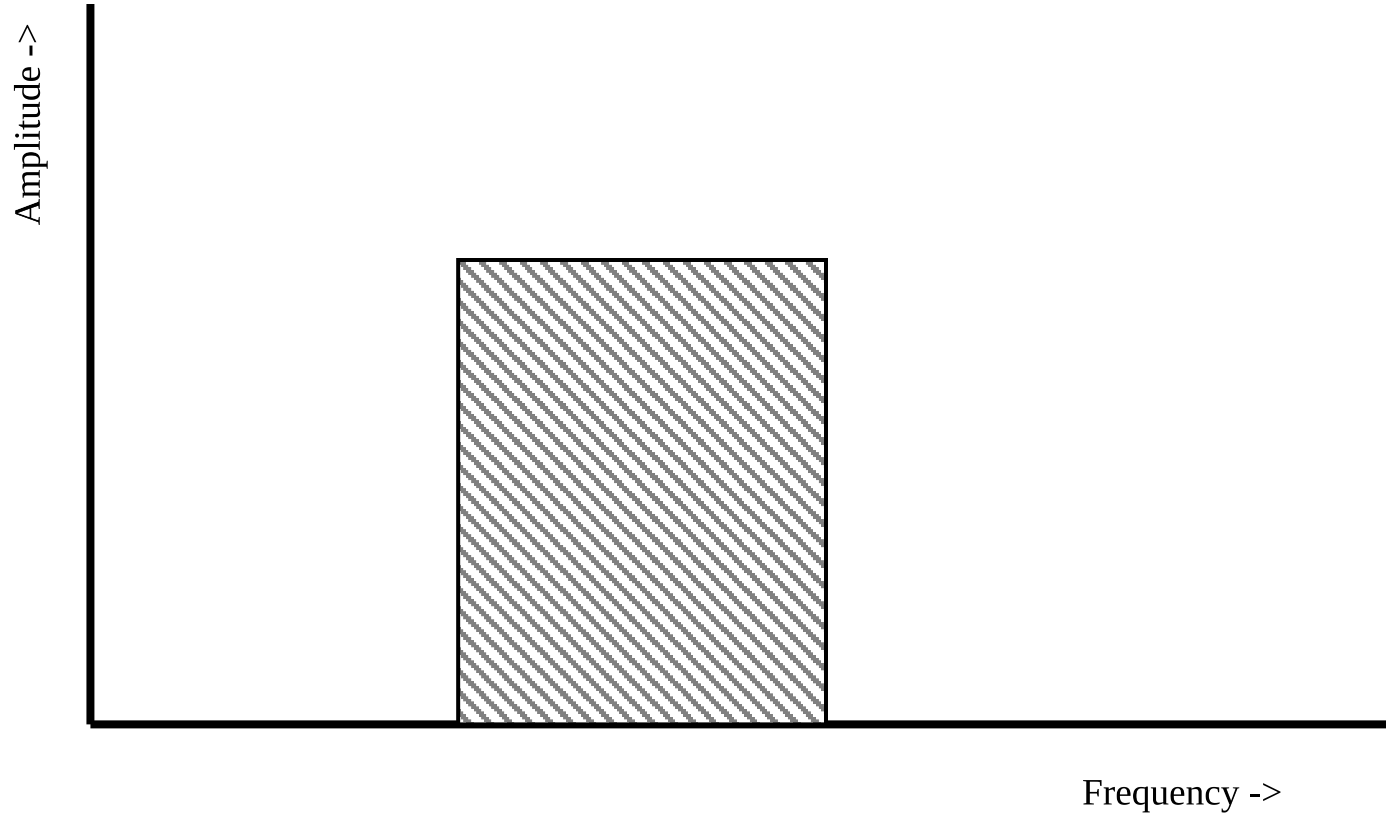
Sizes of sound examples

File	Rate	Size (kbyte)	Ratio
aida0.mp3	32 kbs	235	44:1
aida1.mp3	56 kbs	411	25:1
aida2.wav		10,344	1:1
aida3.mp3	80 kbs	587	18:1
aida2.zip		8,456	1.25:1

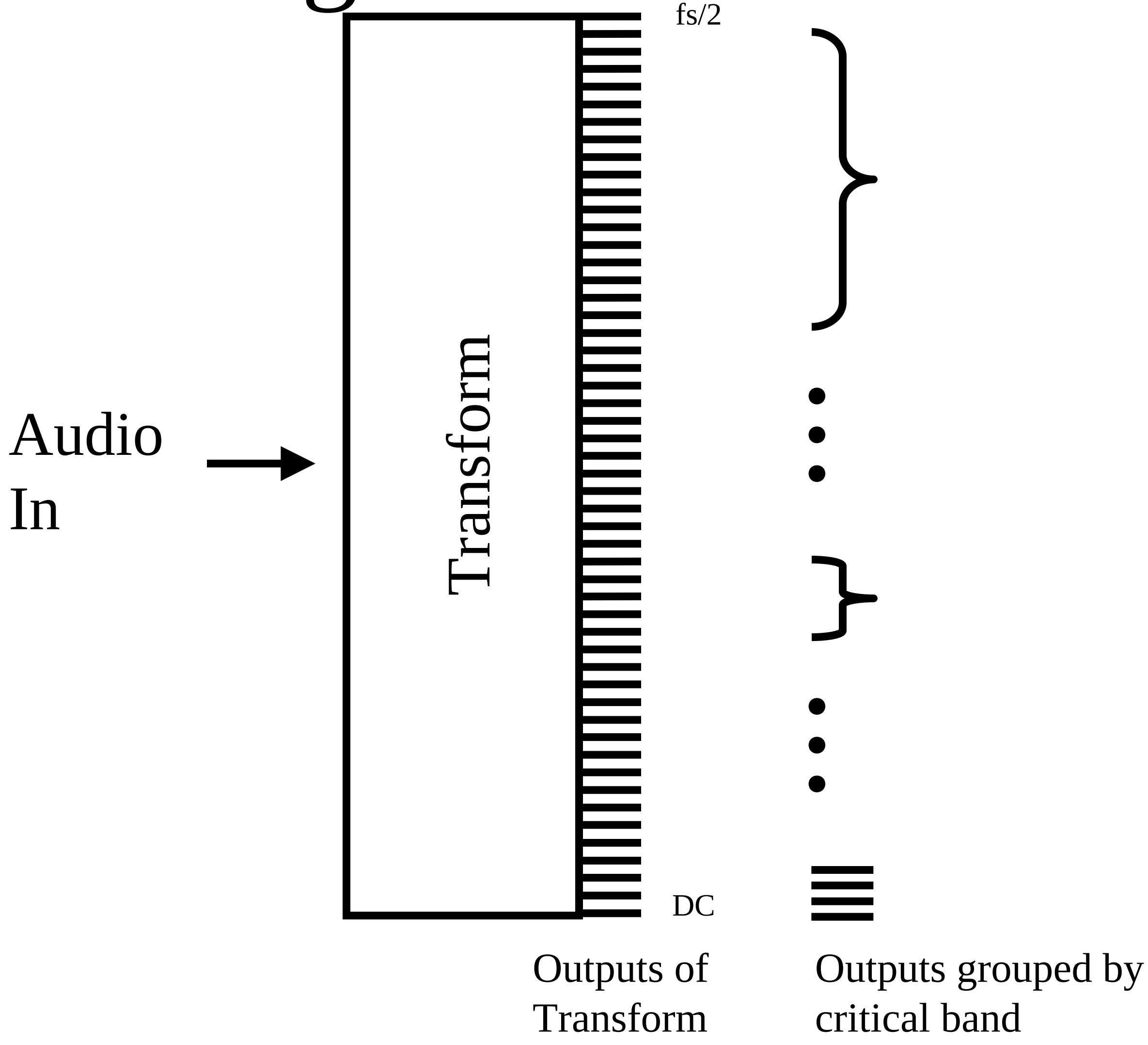
What we will cover

- Perceptual coding
- Psychoacoustics
- Range of hearing
- Masking (simultaneous)
- More on structure of simple encoder/decoder
- Critical Bands
- Variable, Constant Bit Rate
- “Noiseless” Coding

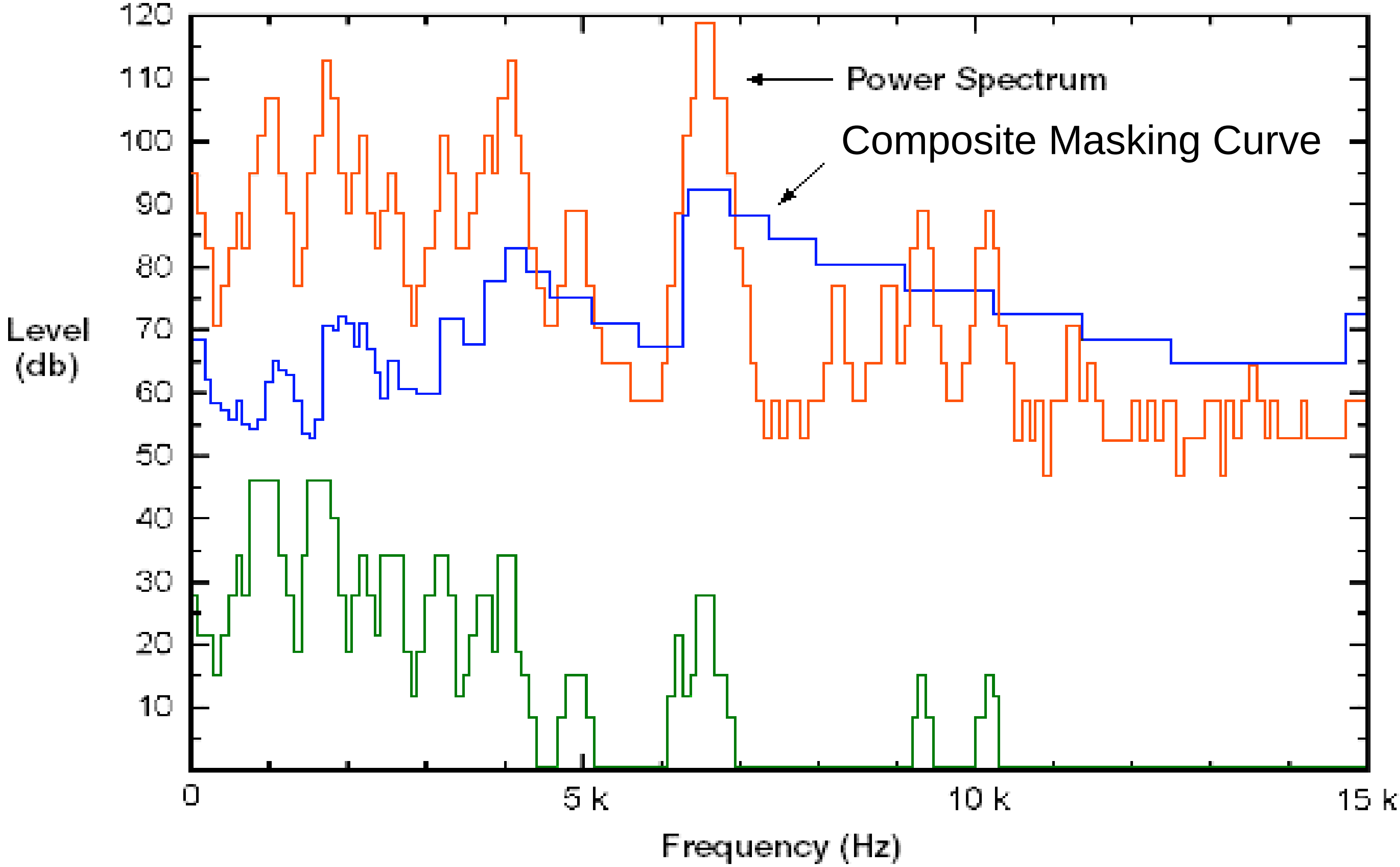
"Critical Bands"



Using Critical Bands



Using Critical Bands



After Davidson et al., 1994

Meeting the challenge

- Coarser Quantization (time domain)
- DPCM, ADPCM
- Linear Prediction
- Subband coding
- Transform to frequency domain
- Coarser quantization (frequency domain)

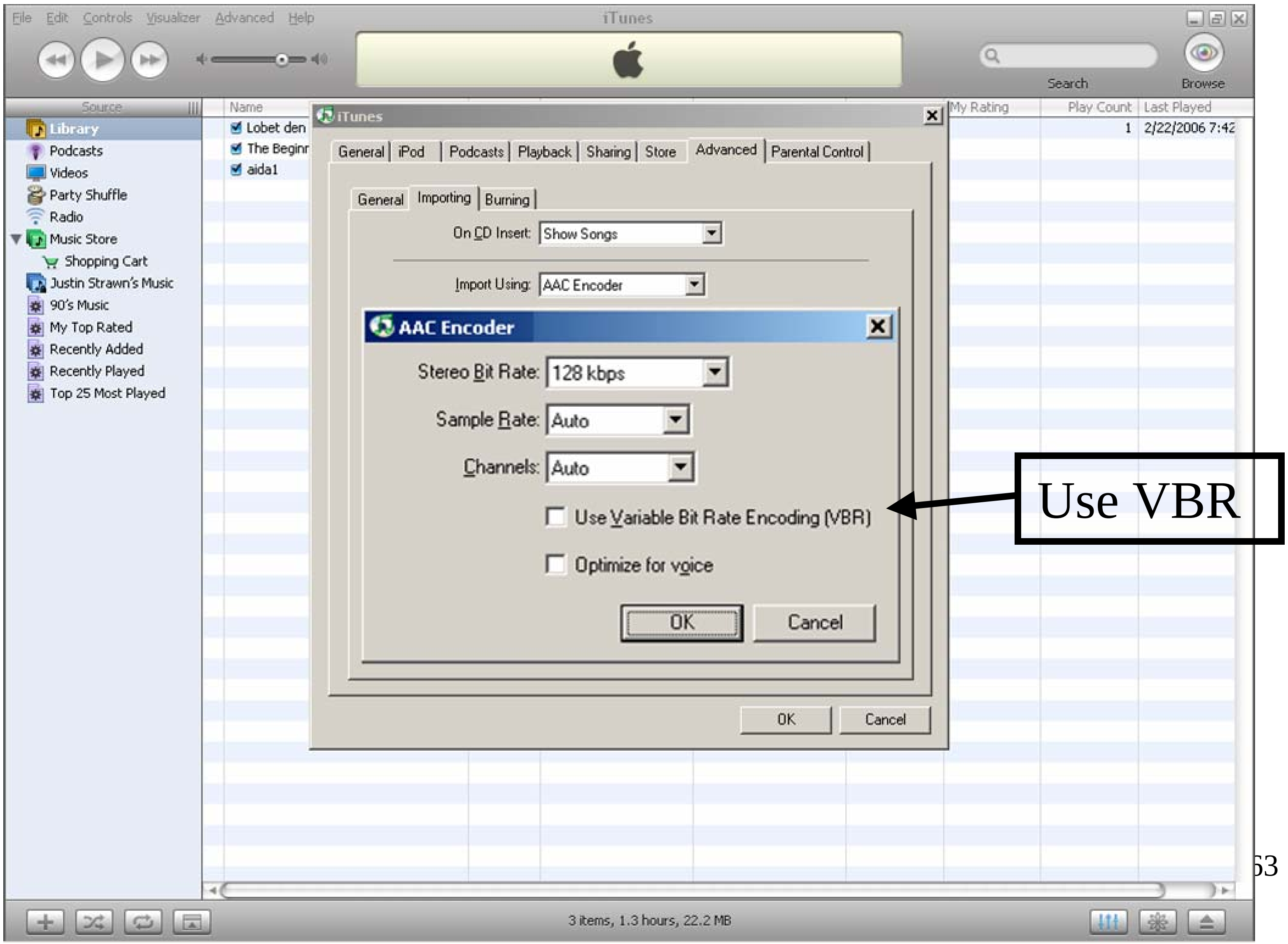
- Psychoacoustics: mask the noise

What we will cover

- Perceptual coding
- Psychoacoustics
- Range of hearing
- Masking (simultaneous)
- More on structure of simple encoder/decoder
- Critical Bands
- Variable, Constant Bit Rate
- "Noiseless" Coding

VBR vs. CBR

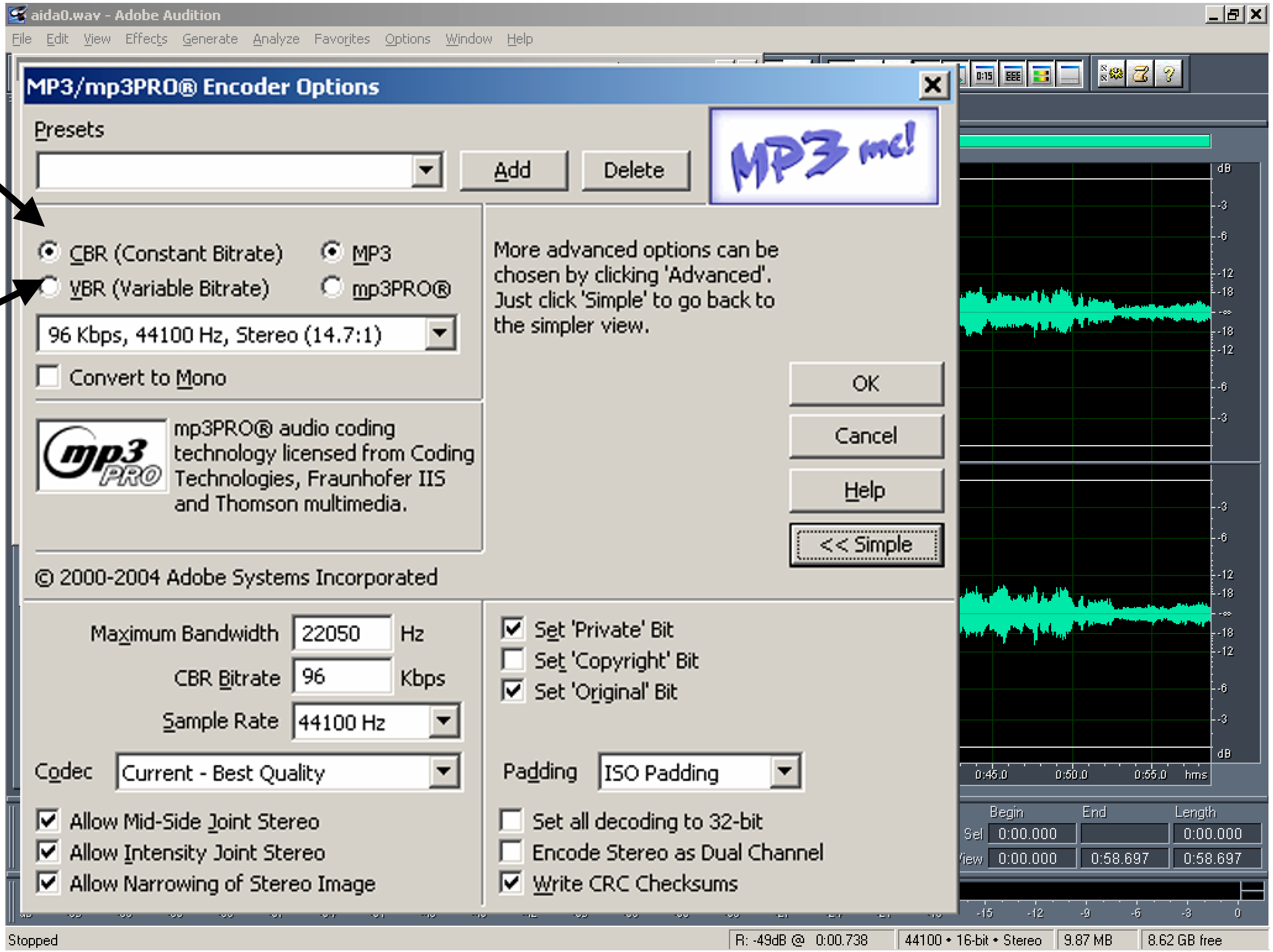
VBR vs. CBR: iTunes



VBR vs. CBR: Audition

Use CBR

Use VBR



What we will cover

- Perceptual coding
- Psychoacoustics
- Range of hearing
- Masking (simultaneous)
- More on structure of simple encoder/decoder
- Critical Bands
- Variable, Constant Bit Rate
- “Noiseless” Coding

Review: Fundamentals of perceptual coding

- Leave out
  - Irrelevant (easier in frequency domain)
  - Redundant
- Allow noise
  - But only in special places

"Noiseless" coding (I)

- Huffman coding

5	Mexico City
8	Monterrey
69	Mazatlan
684	Cabo San Lucas

Bit stream 1

...

0011 0010 1100 1111 0111 1001 0010 0011

0101 1000 0100 0010 1111 0000 0010 1001

...

(from aida0.mp3)

MP3 Huffman Coding

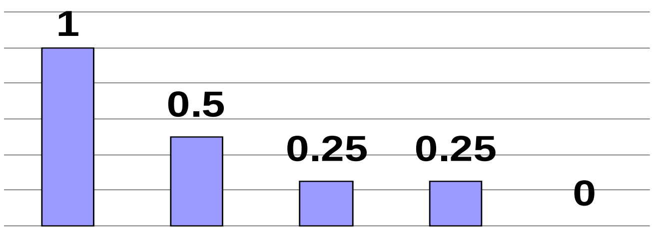
ISO/IEC 11172-3: 1993 (E)

Table B.7 -- Huffman codes for Layer III

Huffman code table for quadruples (A)

v w x y	hlen	hcod
0 0 0 0	1	1
0 0 0 1	4	0101
0 0 1 0	4	0100
0 0 1 1	5	00101
0 1 0 0	4	0110
0 1 0 1	6	000101
0 1 1 0	5	00100
0 1 1 1	6	000100
1 0 0 0	4	0111
1 0 0 1	5	00011
1 0 1 0	5	00110
1 0 1 1	6	000000
1 1 0 0	5	00111
1 1 0 1	6	000010
1 1 1 0	6	000011
1 1 1 1	6	000001

Quantized Spectrum / bit stream 2



... 1111111111111111 1000000000000000 0100000000000000 0100000000000000 0000000000000000 ...

“Noiseless” coding (II)

- Run length:

... 1111111111111111 1000000000000000 0100000000000000

0100000000000000 0000000000000000 ...

Perceptual Encoder

Audio Signal

Forward transform, get spectrum

Omit parts of spectrum (add noise)

How accurately to represent remaining spectrum (add noise)

N

Noise below masking?

Y

Package compressed output (Noiseless coding)

So what's really in an MP3 file

Meeting the challenge

- Coarser Quantization (time domain)
- DPCM, ADPCM
- Linear Prediction
- Subband coding
- Transform to frequency domain
- Coarser quantization (frequency domain)
- Psychoacoustics: mask the noise
- Variable bit rate
- Noiseless coding

What we have covered

- Perceptual coding
- Psychoacoustics
- Range of hearing
- Masking (simultaneous)
- More on structure of simple encoder/decoder
- Critical Bands
- Variable, Constant Bit Rate
- “Noiseless” Coding

Easy Listening Session

- Meet by ____ in listening areas.
- Listen to tracks PreEcho1 and PreEcho2 in Session: MP3 Seminar 3.
- Take notes and discuss:
  - What do you hear?
  - How are they different?
  - What do you like?
  - What don't you like?
- Back here at 4:00 to discuss. (10 min)

Notes on Sound Examples

MSRA 5500 - 002
MUS 4500 - 002

March 9, Session 2, Part 2

Time-Domain Aspects

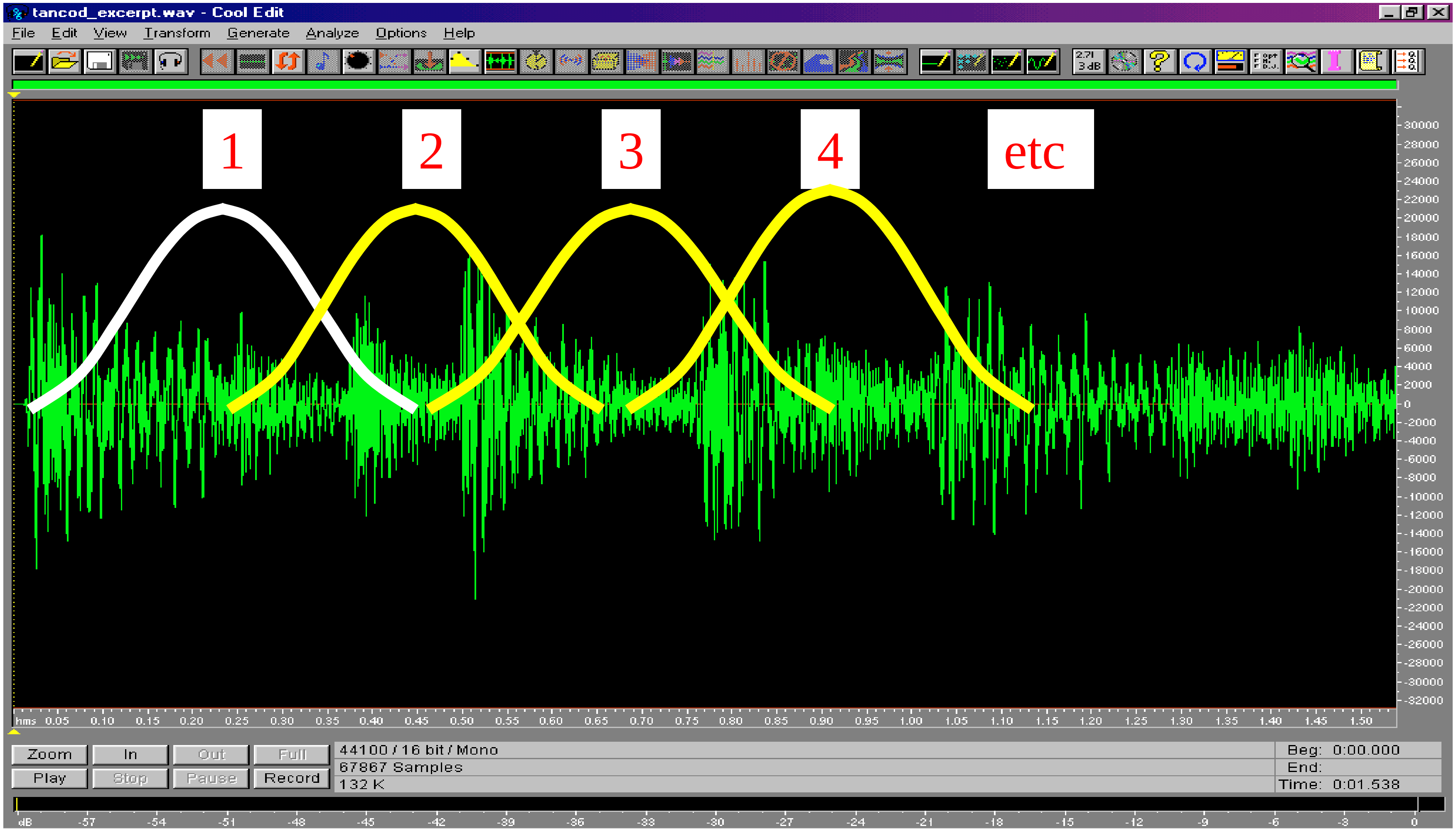
© Copyright 2008 John Strawn

Discussion of Sound Examples

What we will cover

- Windowing
- Pre-echo
  - What is it
  - Why it happens
- Temporal (not simultaneous) masking
- Error Recovery

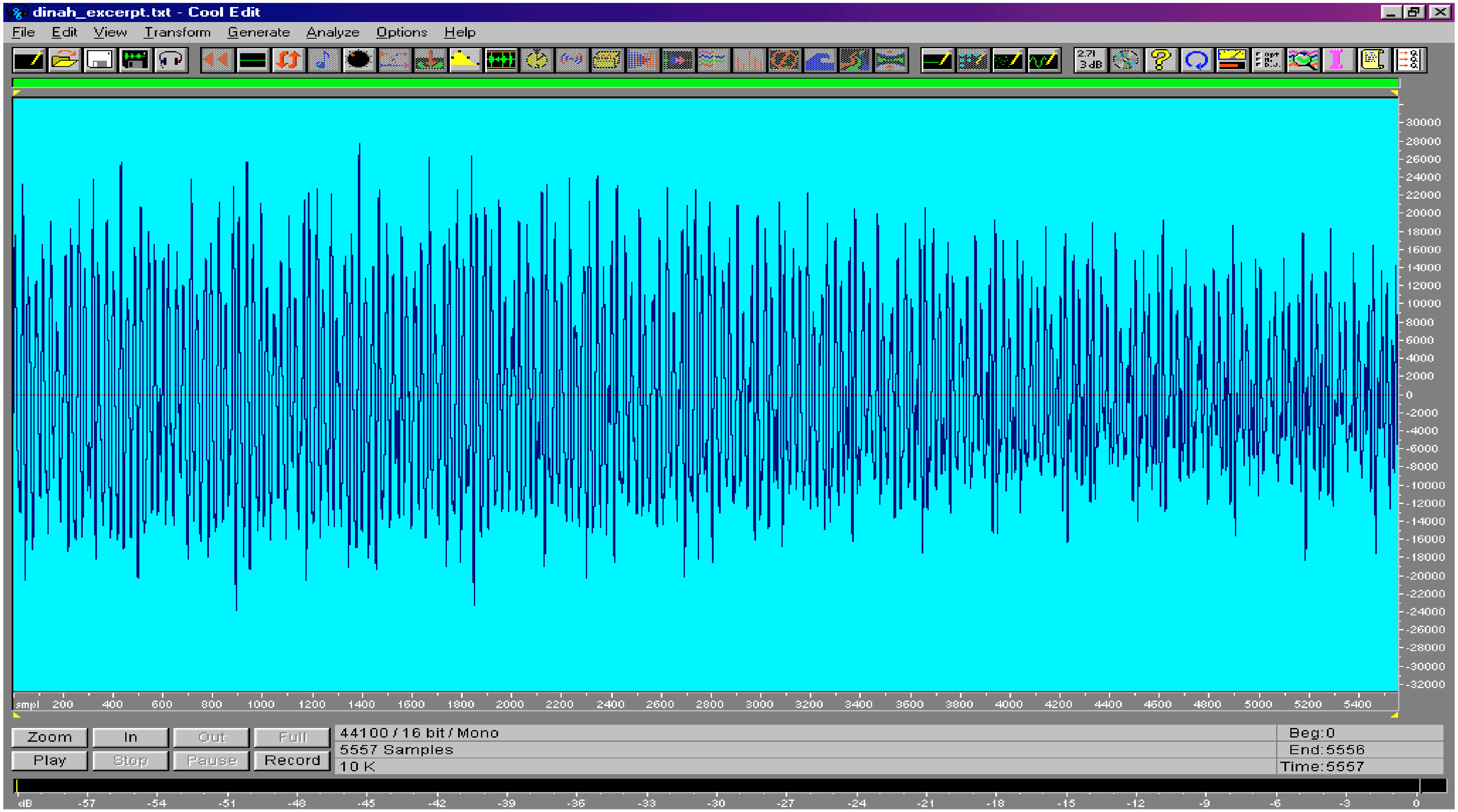
Windowing in Encoder



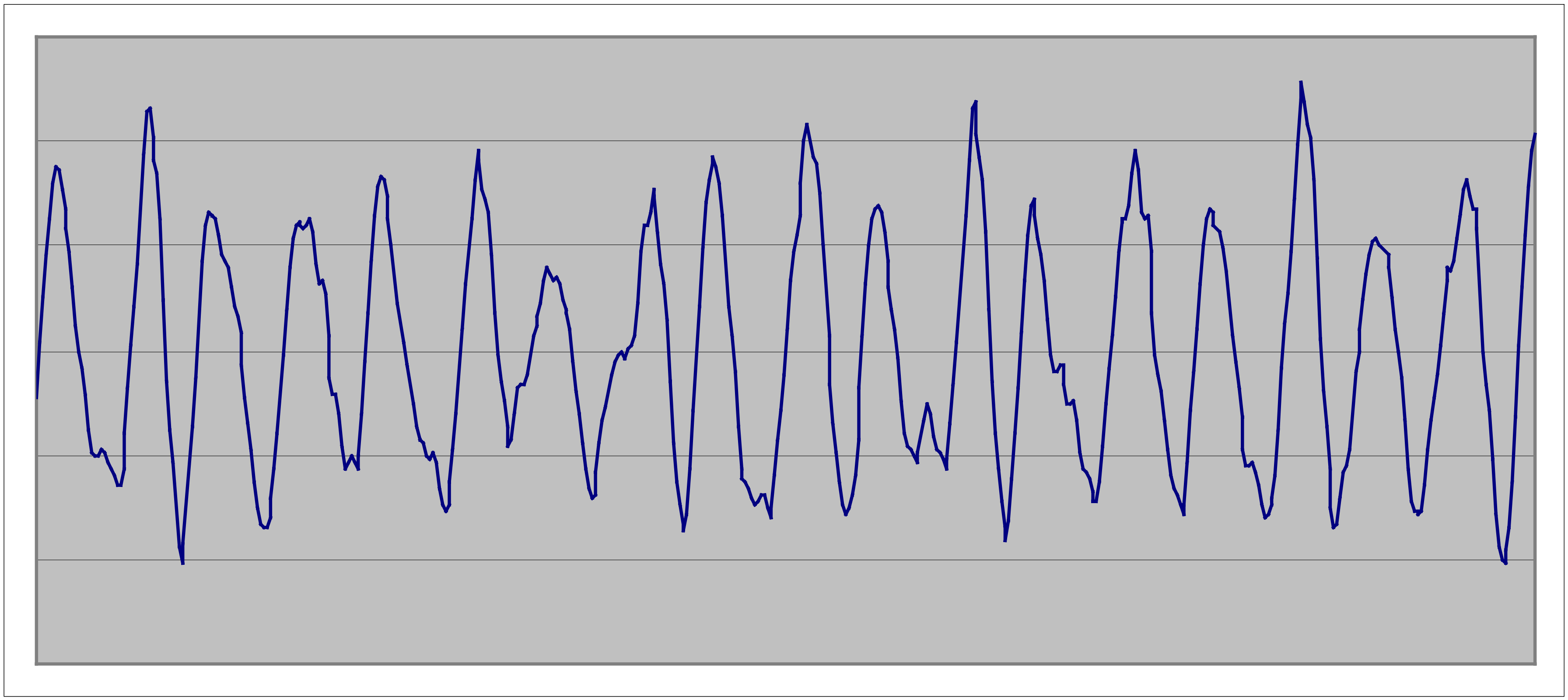
Source: excerpt from tancod55.wav, AES CD-ROM

Note: windows exaggerated in various ways for illustration.

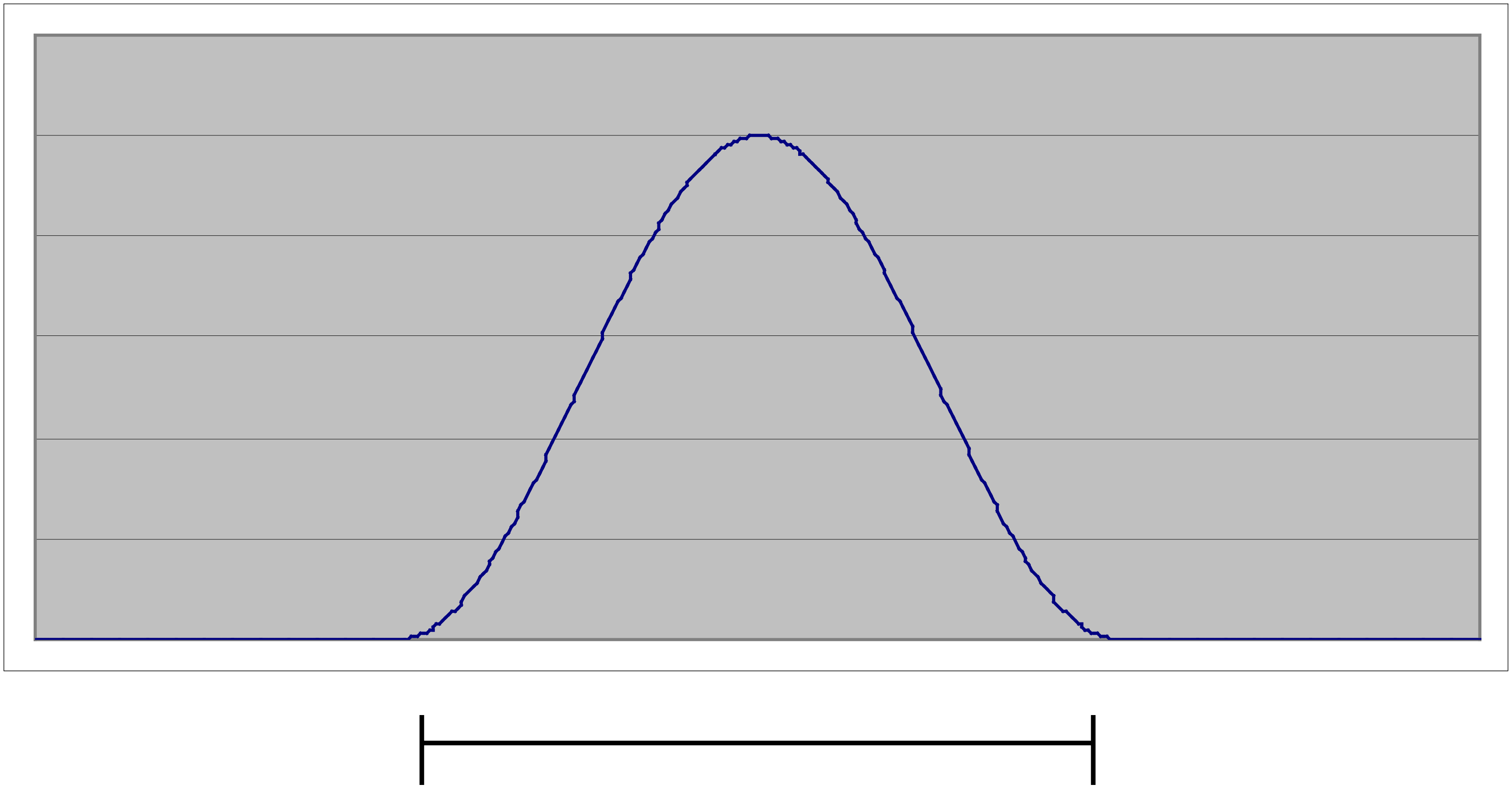
Original Signal



Zoom in on original signal

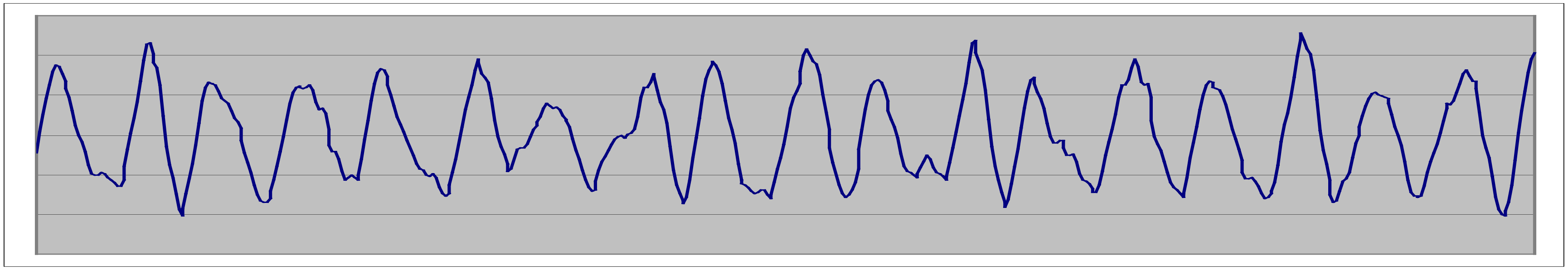


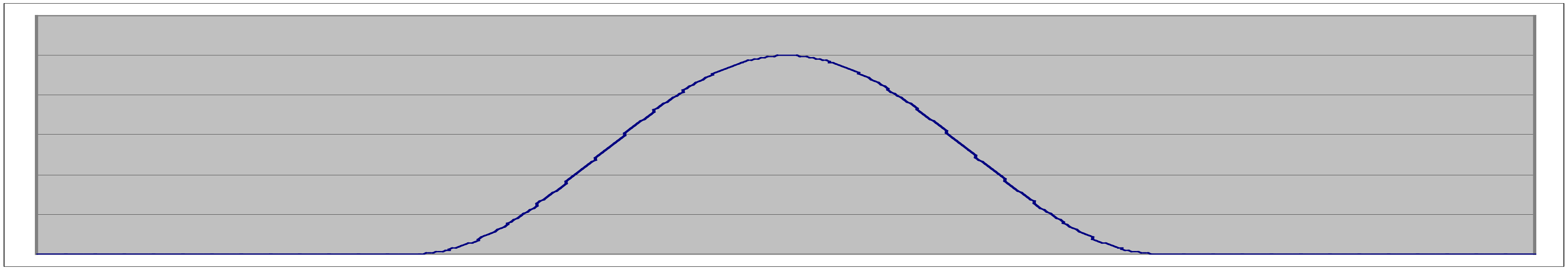
Typical Window

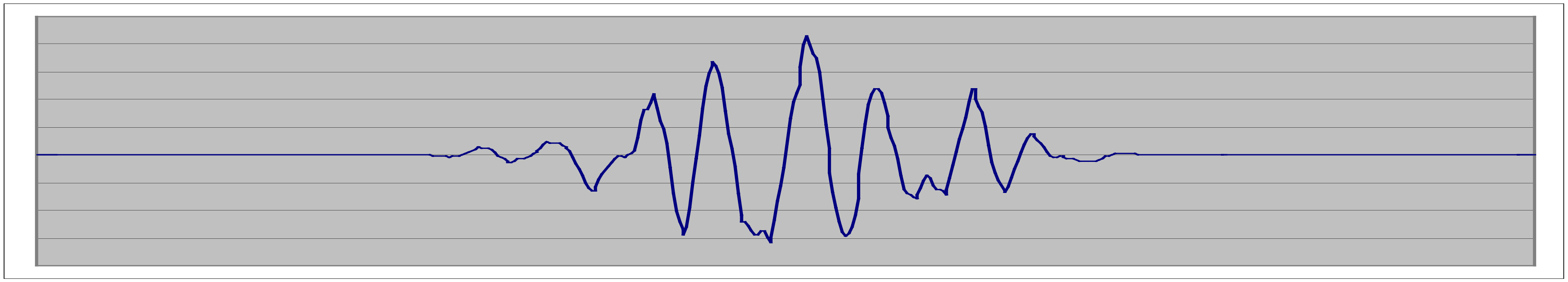


< 1 msec - 40 msec

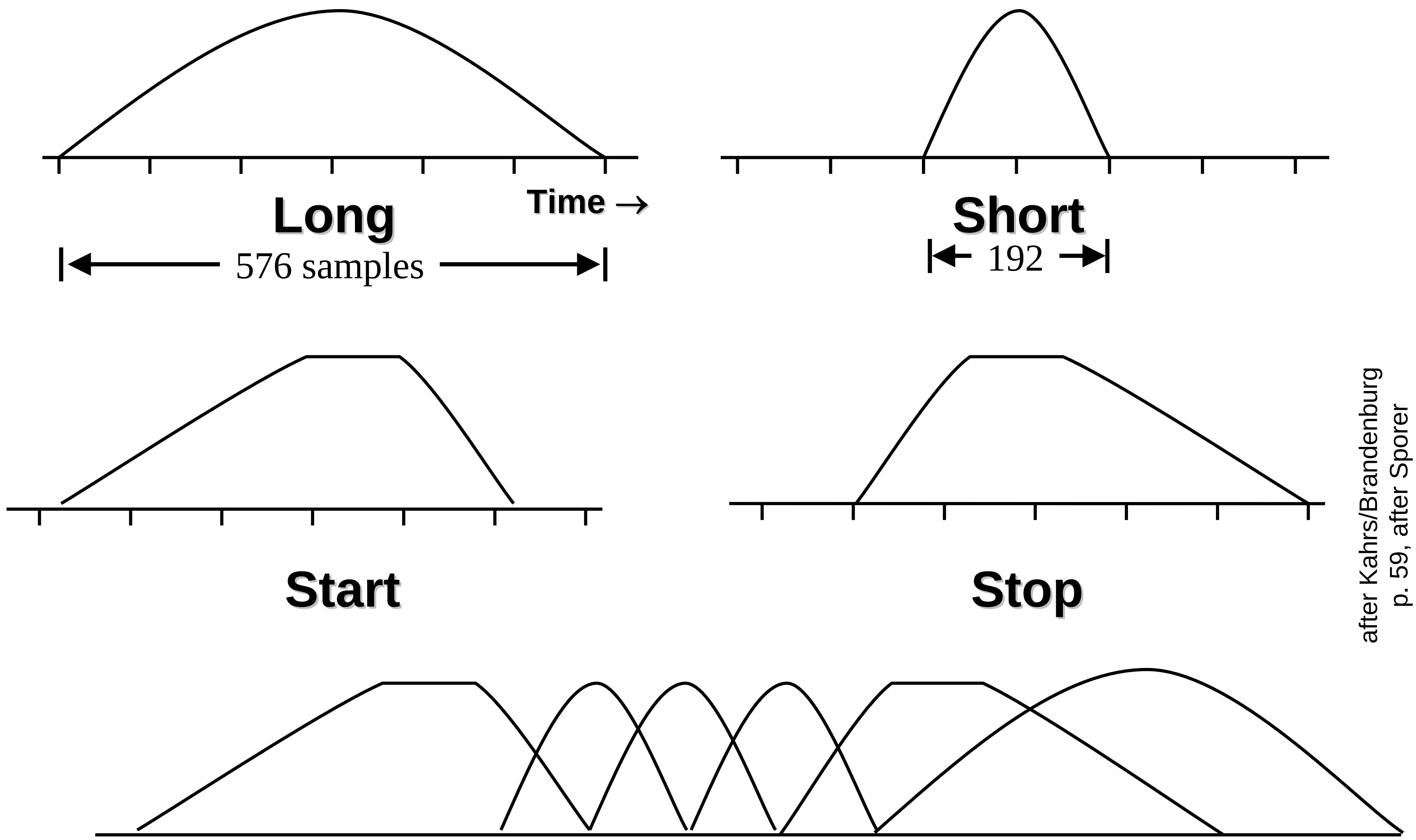
Middle of Original, Windowed





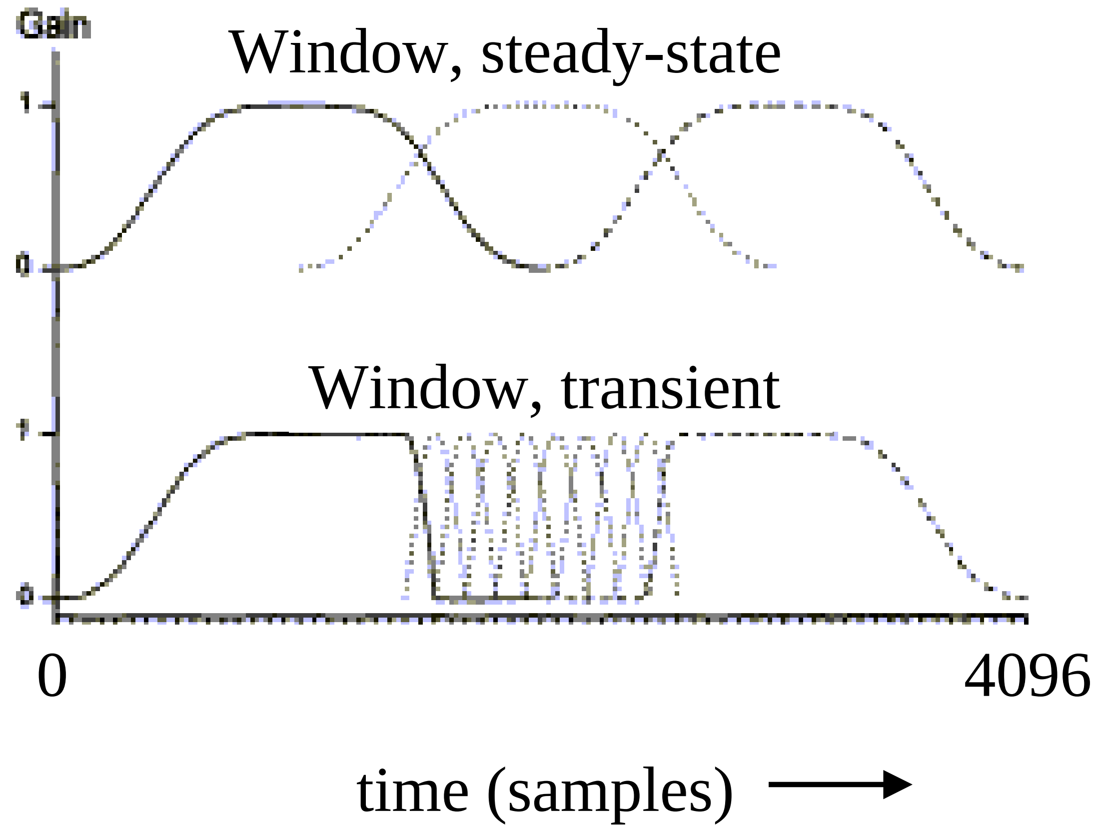


Windows for MP-3



after Kahrs/Brandenburg
p. 59, after Sporer

MPEG-2 AAC Windowing

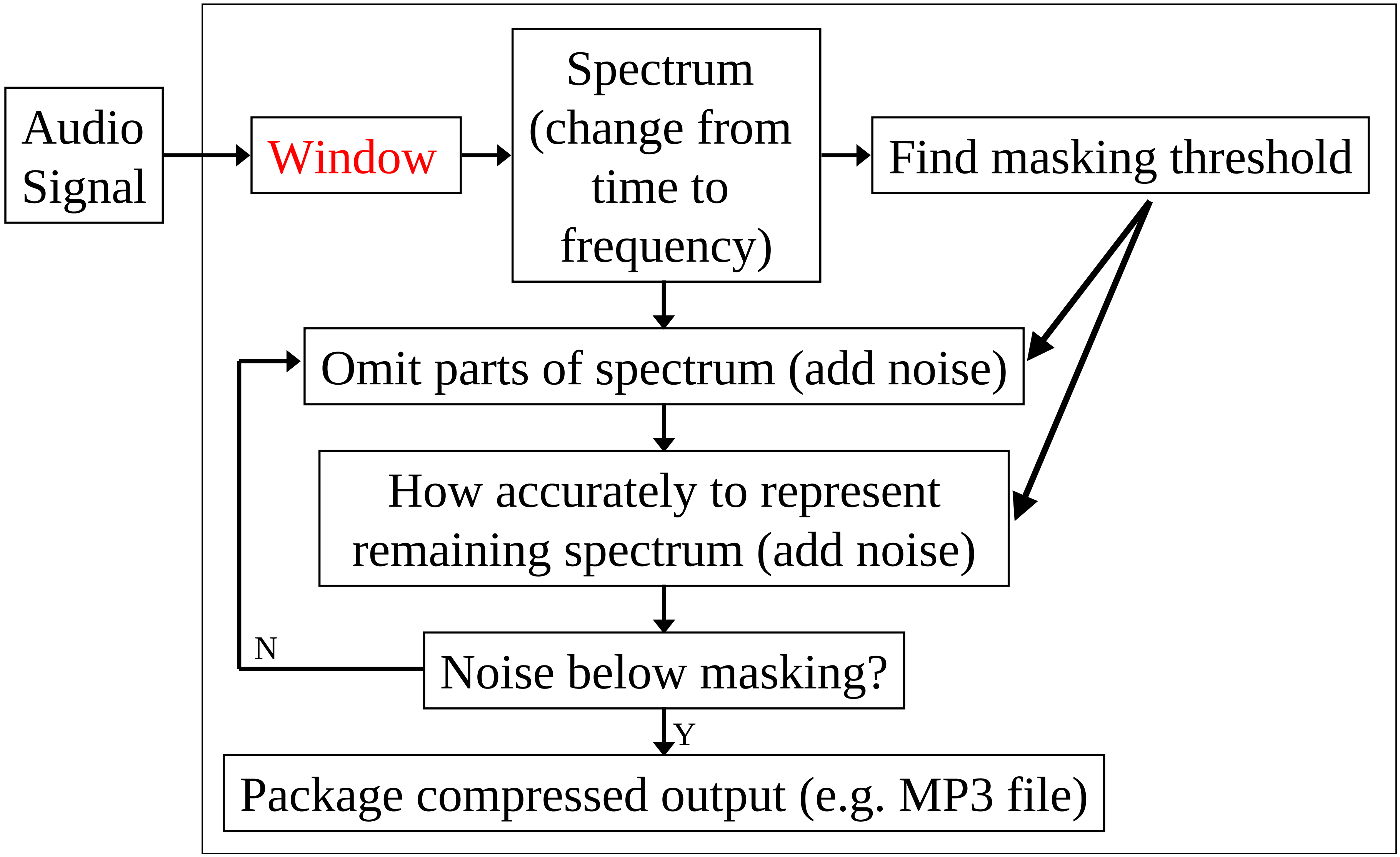


From: Bosi et al, AES preprint 4382, 1996

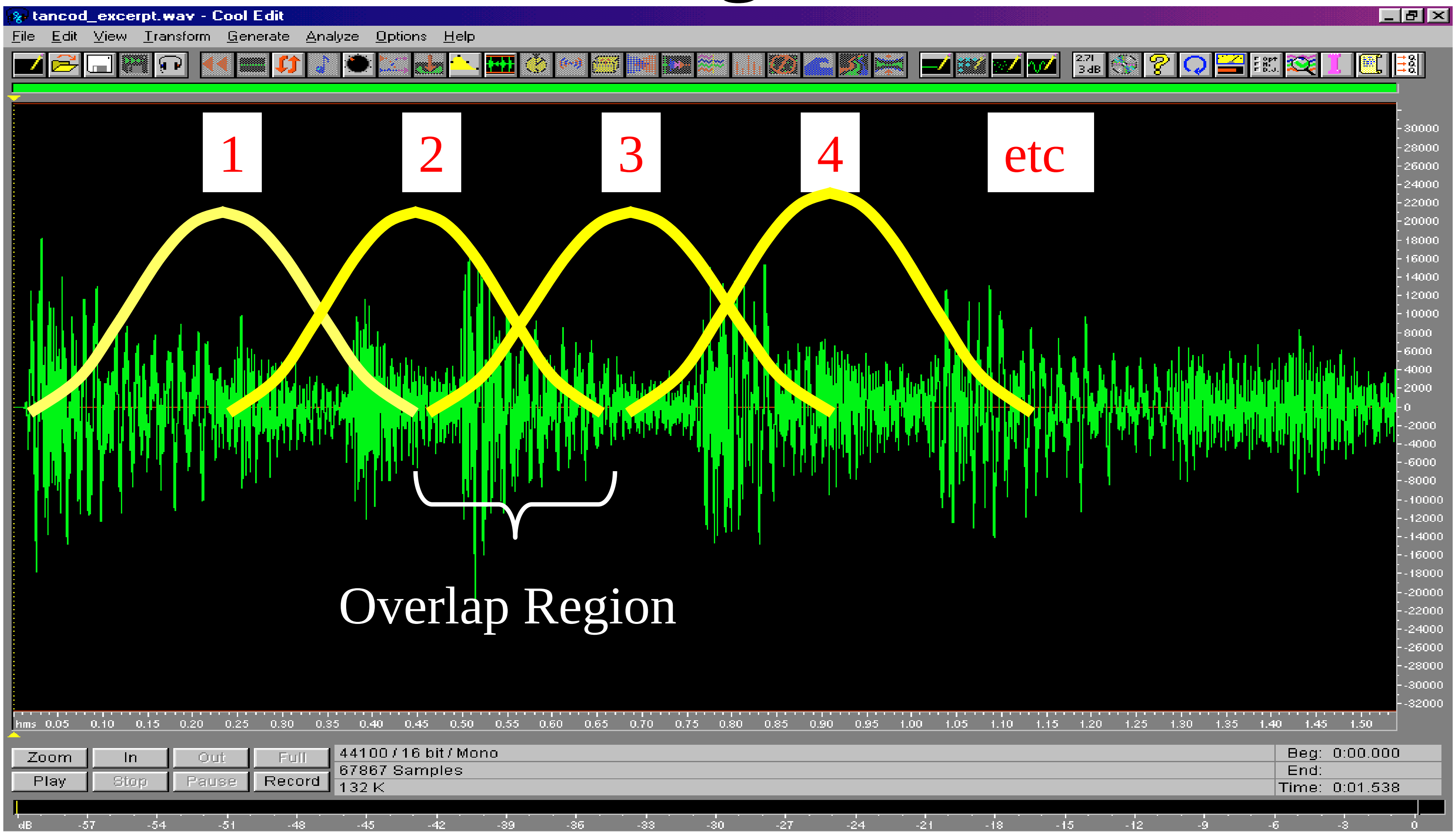
Window lengths (no. samples)

Technique	No. Samples	Sec (48000)	inches @ 7.5 ips
MPEG-1 Layer 3	192, 576		
MPEG-2 AAC	256, 2048		
AC-3	256, 512		
MLP	40-160		
DTS	1024, 2048		
PAC	256, 2048		

Perceptual Encoder



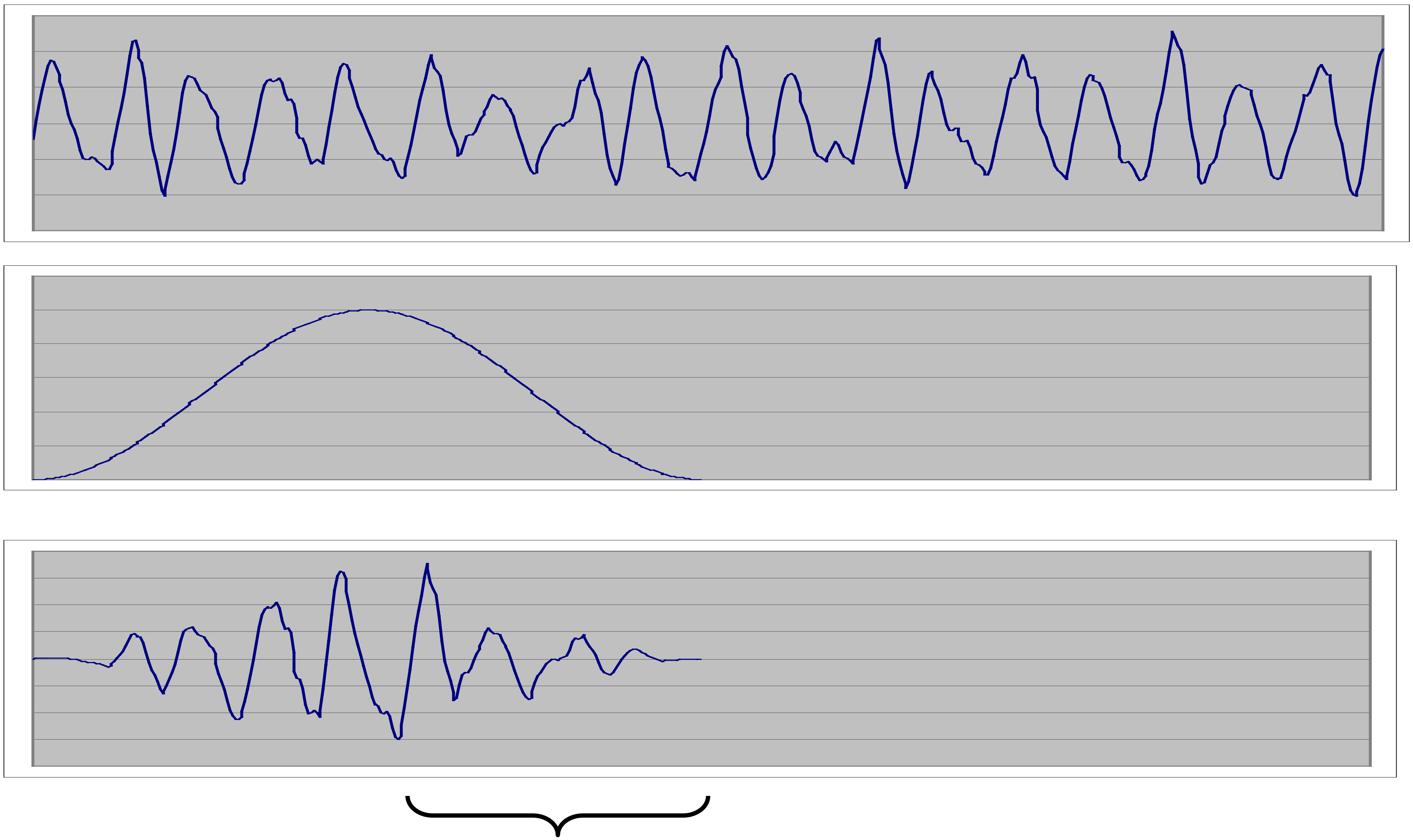
Windowing in Decoder



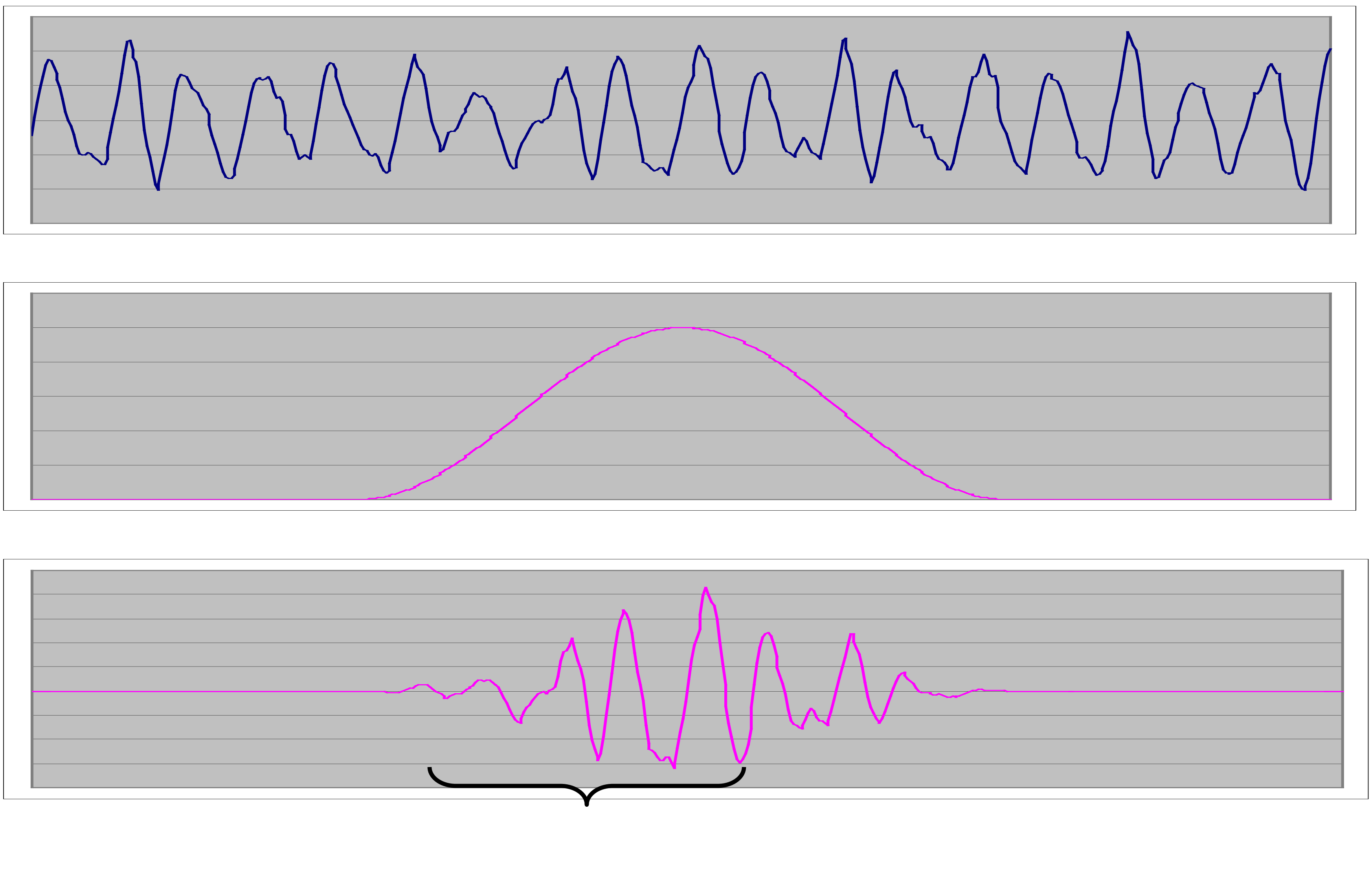
Source: excerpt from tancod55.wav, AES CD-ROM

Note: windows exaggerated in various ways for illustration.

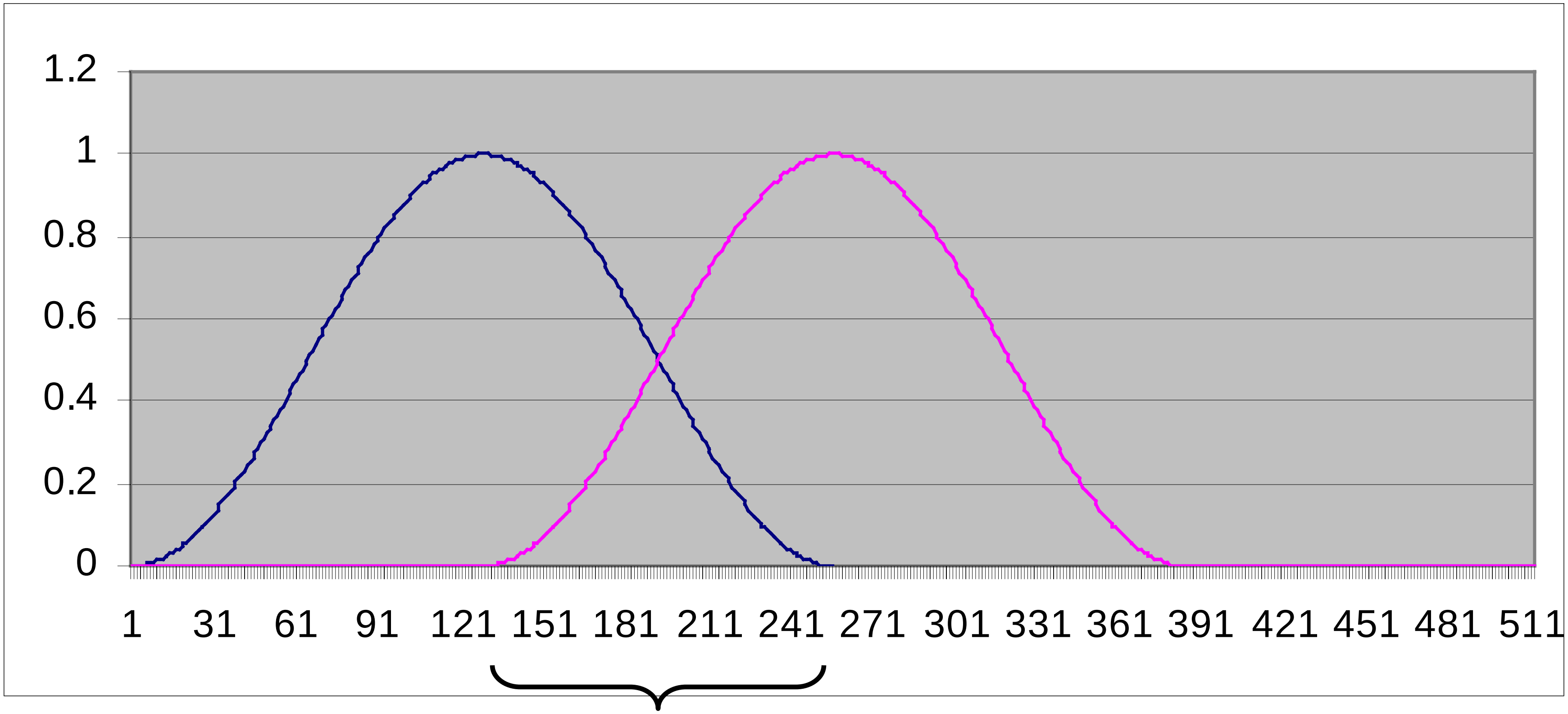
1st Frame, Windowed by Encoder



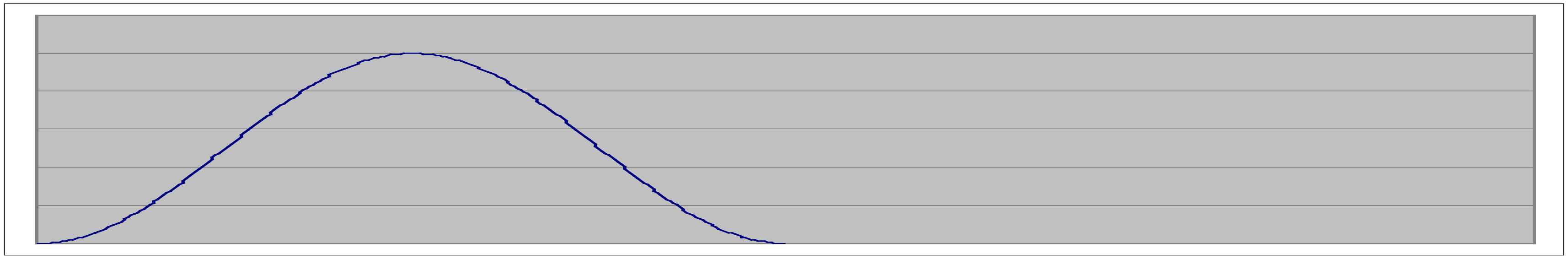
Next Frame, Windowed by Encoder



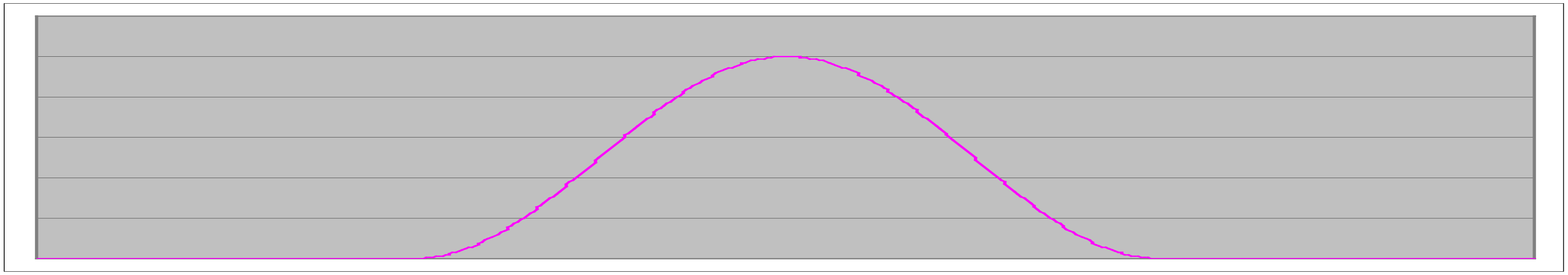
Decoder: How windows align



Decoder: Sum Two Windows

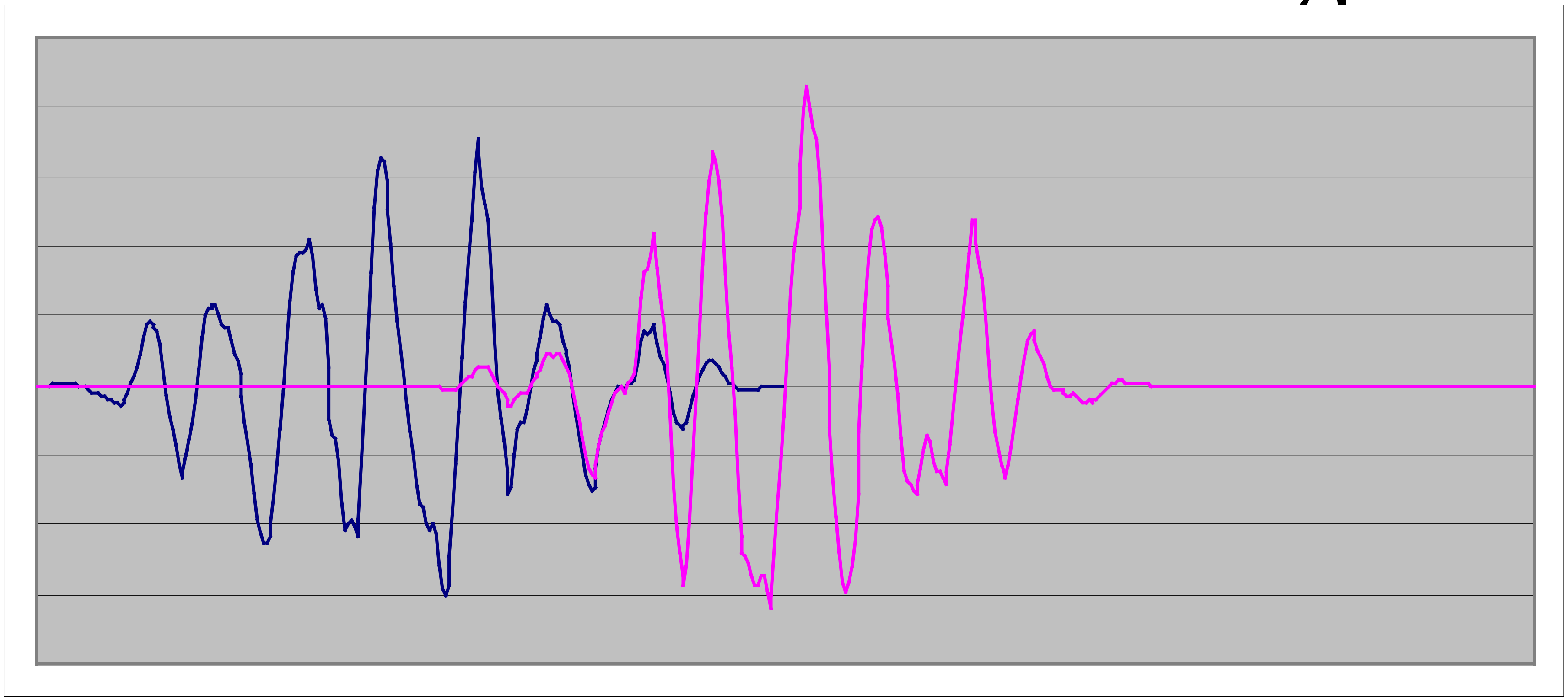


+



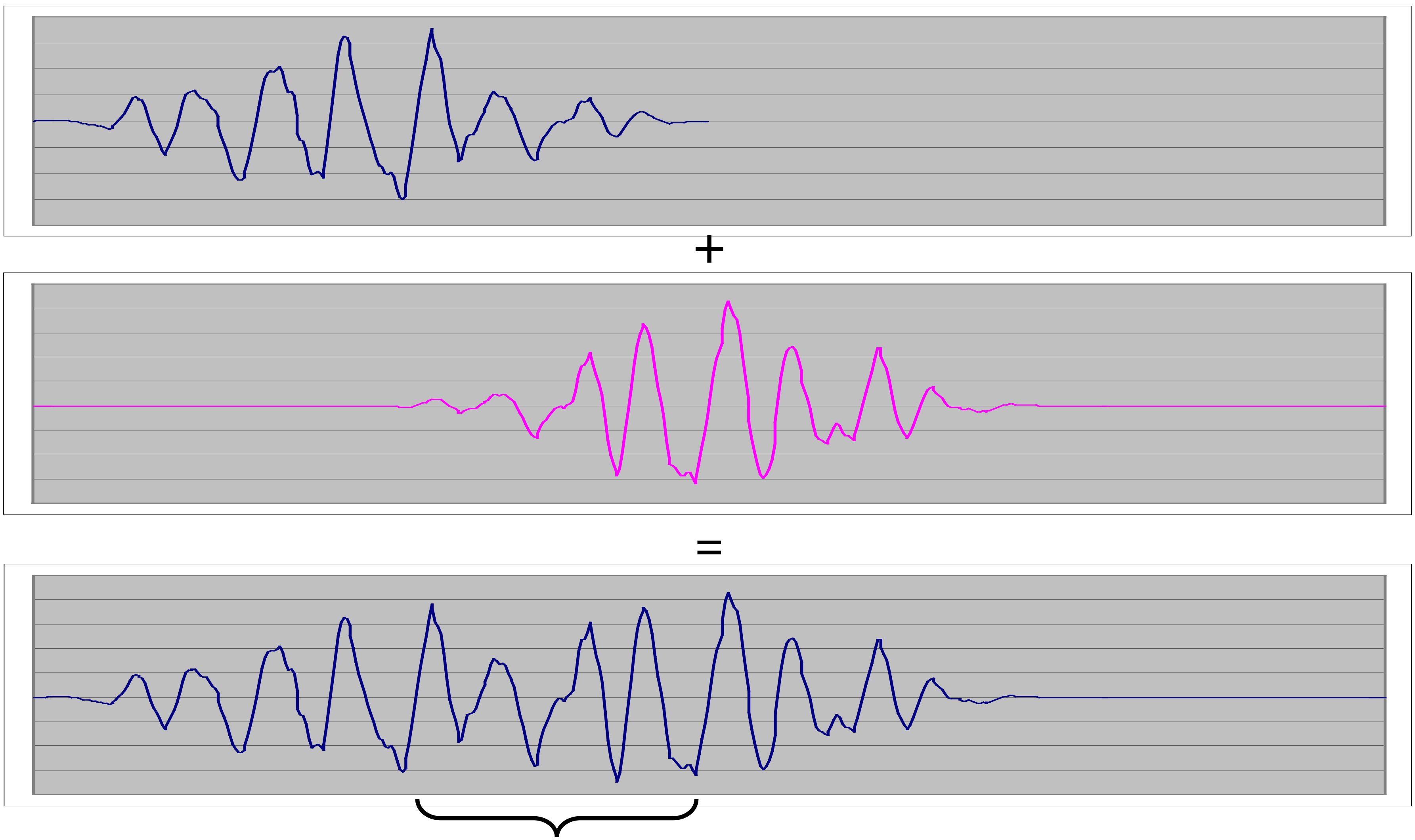
= 1.0 in overlap region

Decoder:
How windowed data align

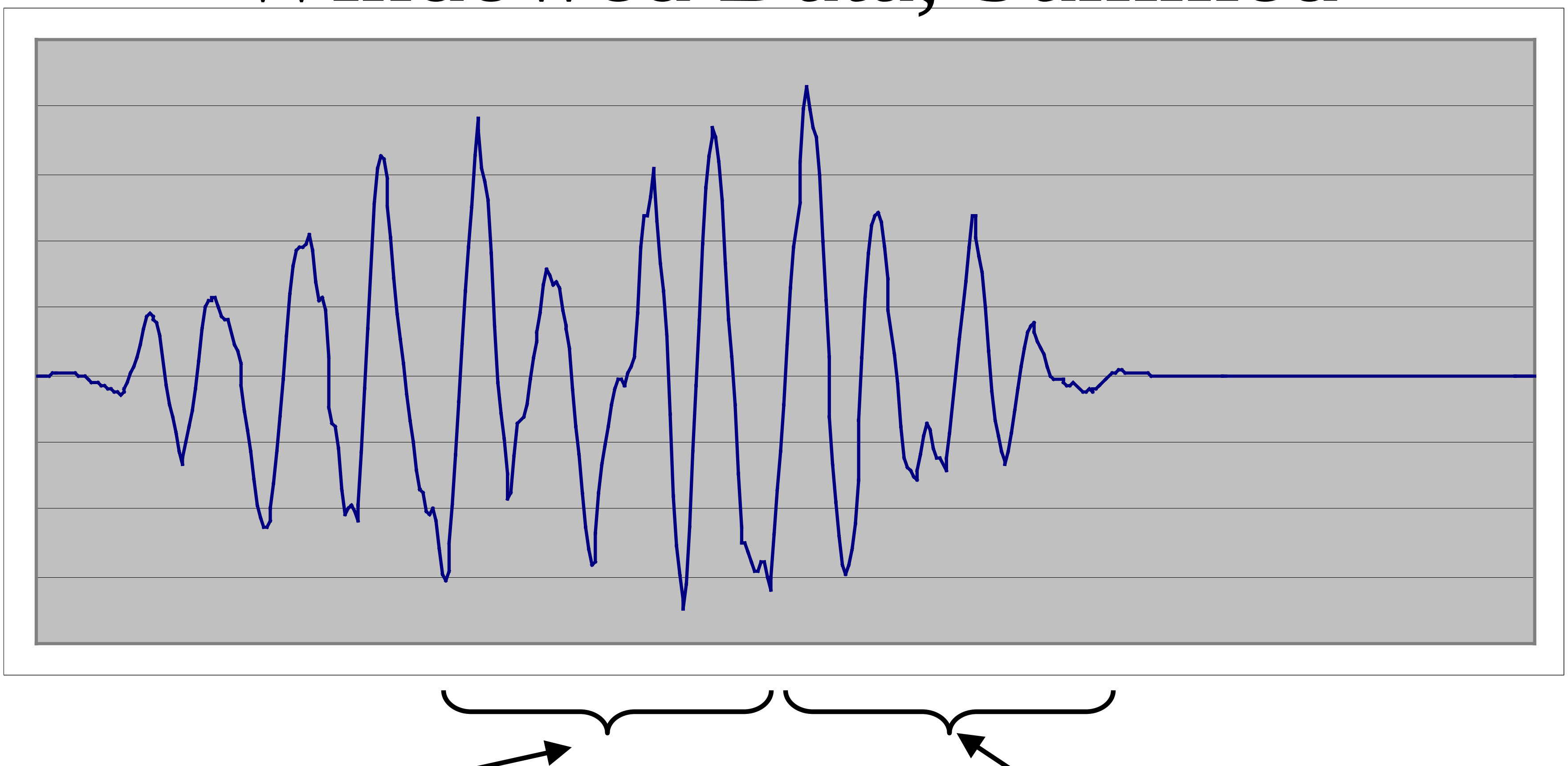


Overlap Region

Decoder:
Summing Two Windowed Frames



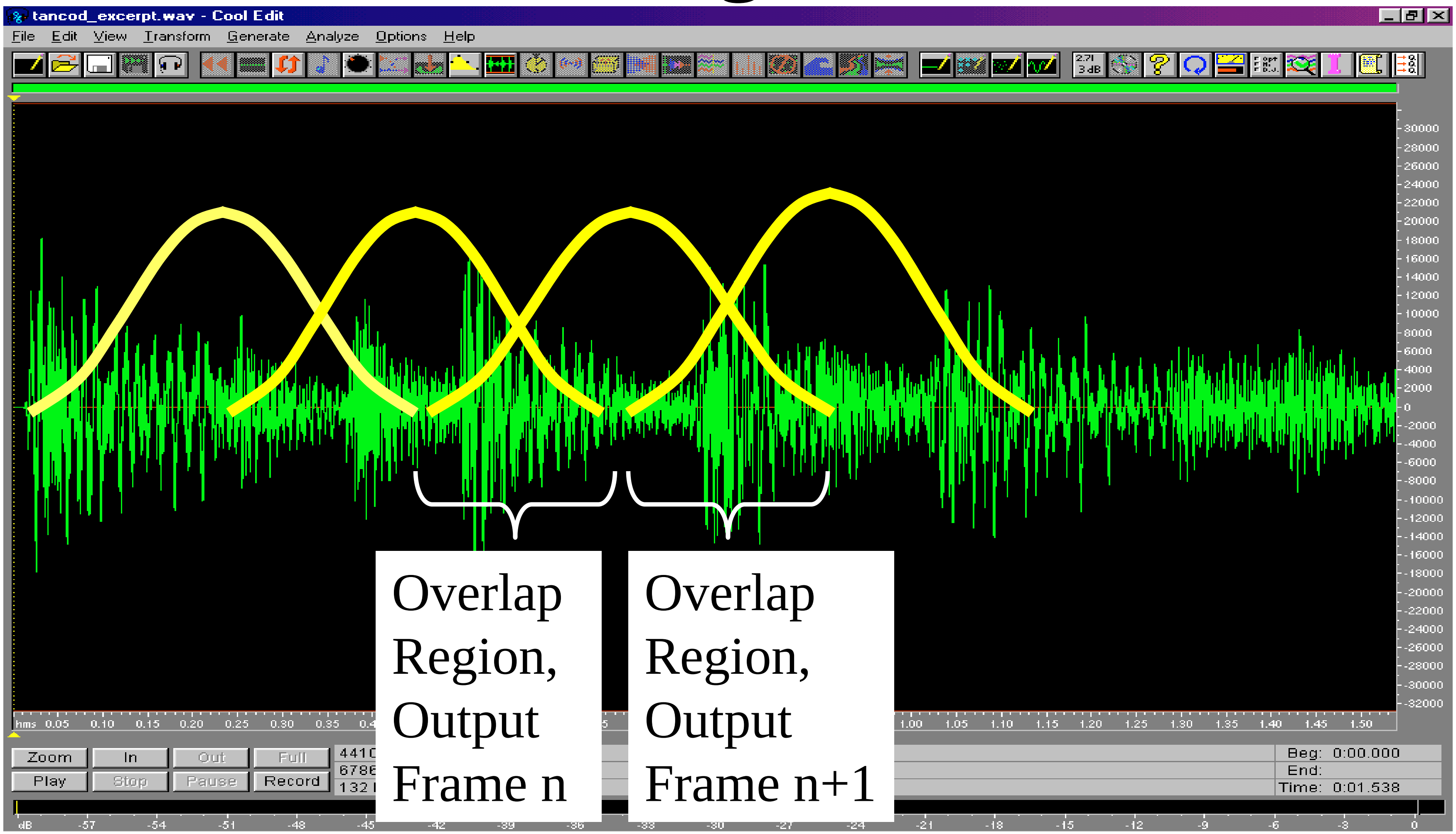
Decoder: Windowed Data, Summed



Overlap Region,
Ready for Sound Output

Saved for
Next Frame

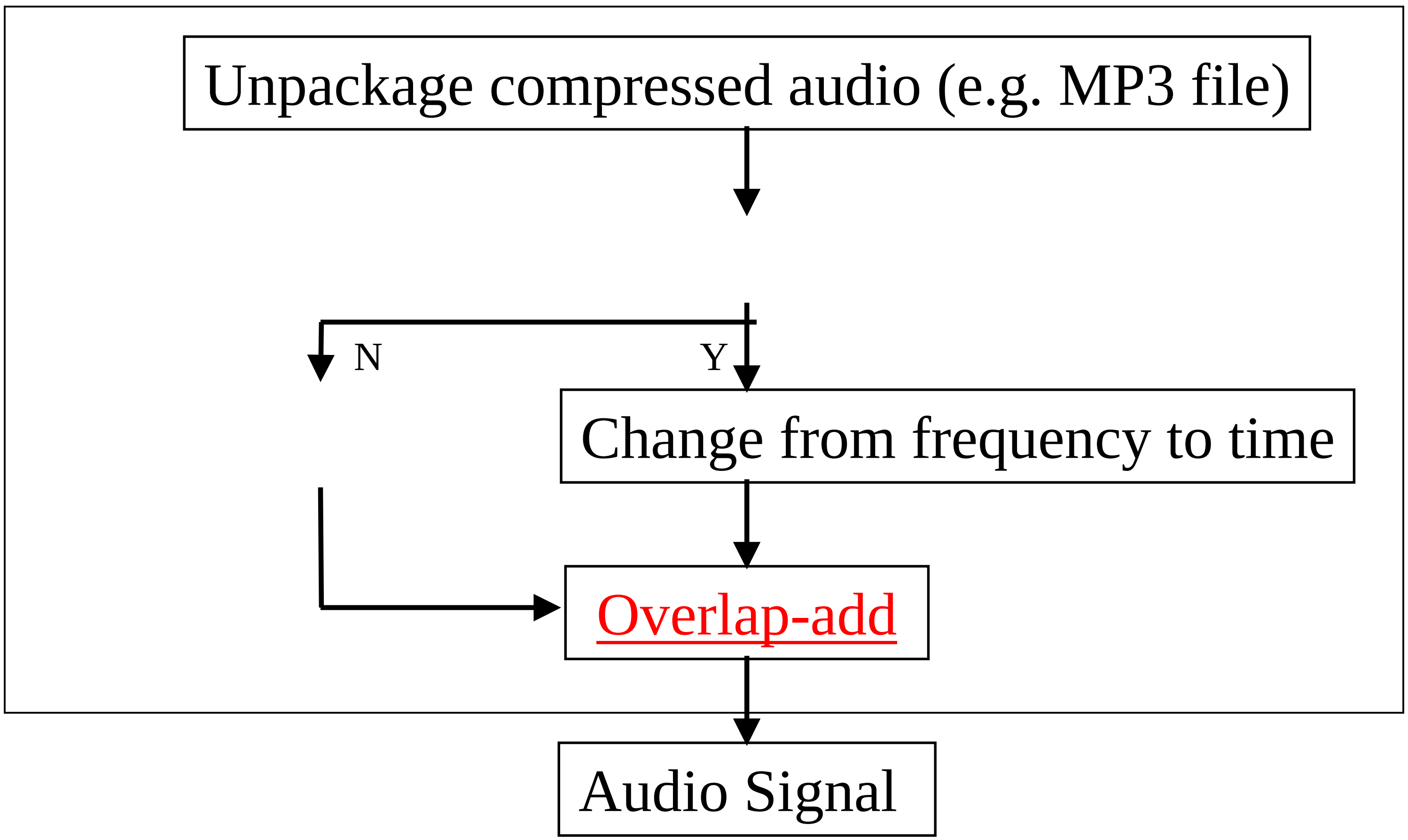
Windowing in Decoder



Source: excerpt from tancod55.wav, AES CD-ROM

Note: windows exaggerated in various ways for illustration.

Decoder ("mp3 player")

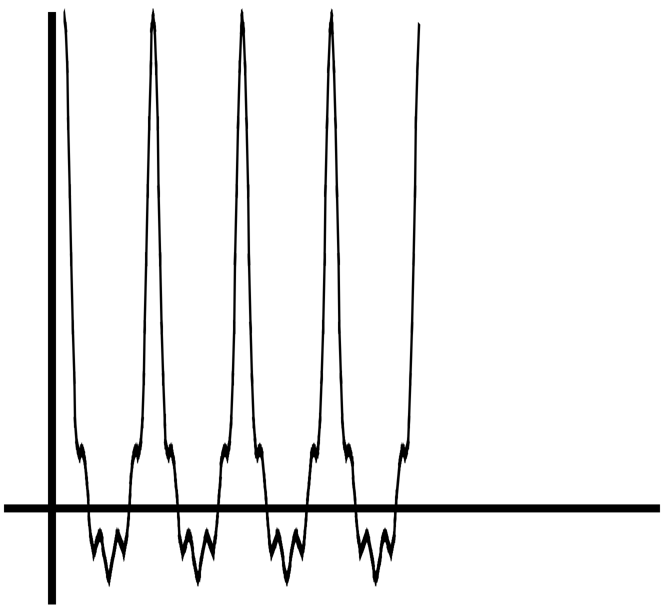


What we will cover

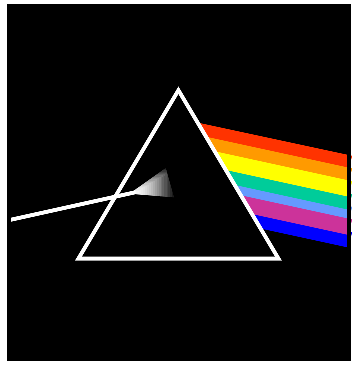
- Windowing
- Pre-echo
  - What is it
  - Why it happens
- Temporal (not simultaneous) masking
- Error Recovery

Review: Transform

Waveform



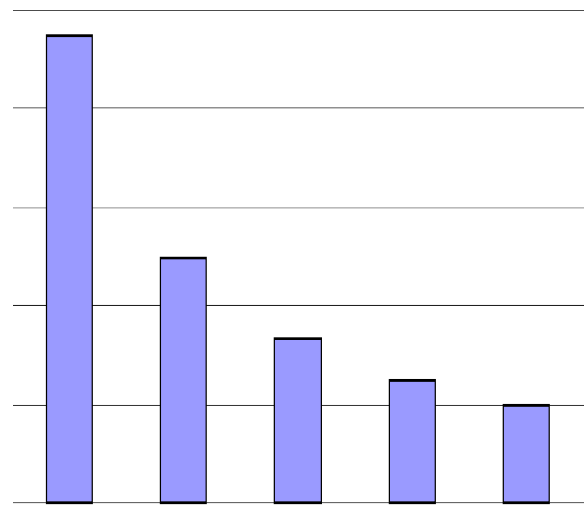
Time Domain



Forward
(analyze)

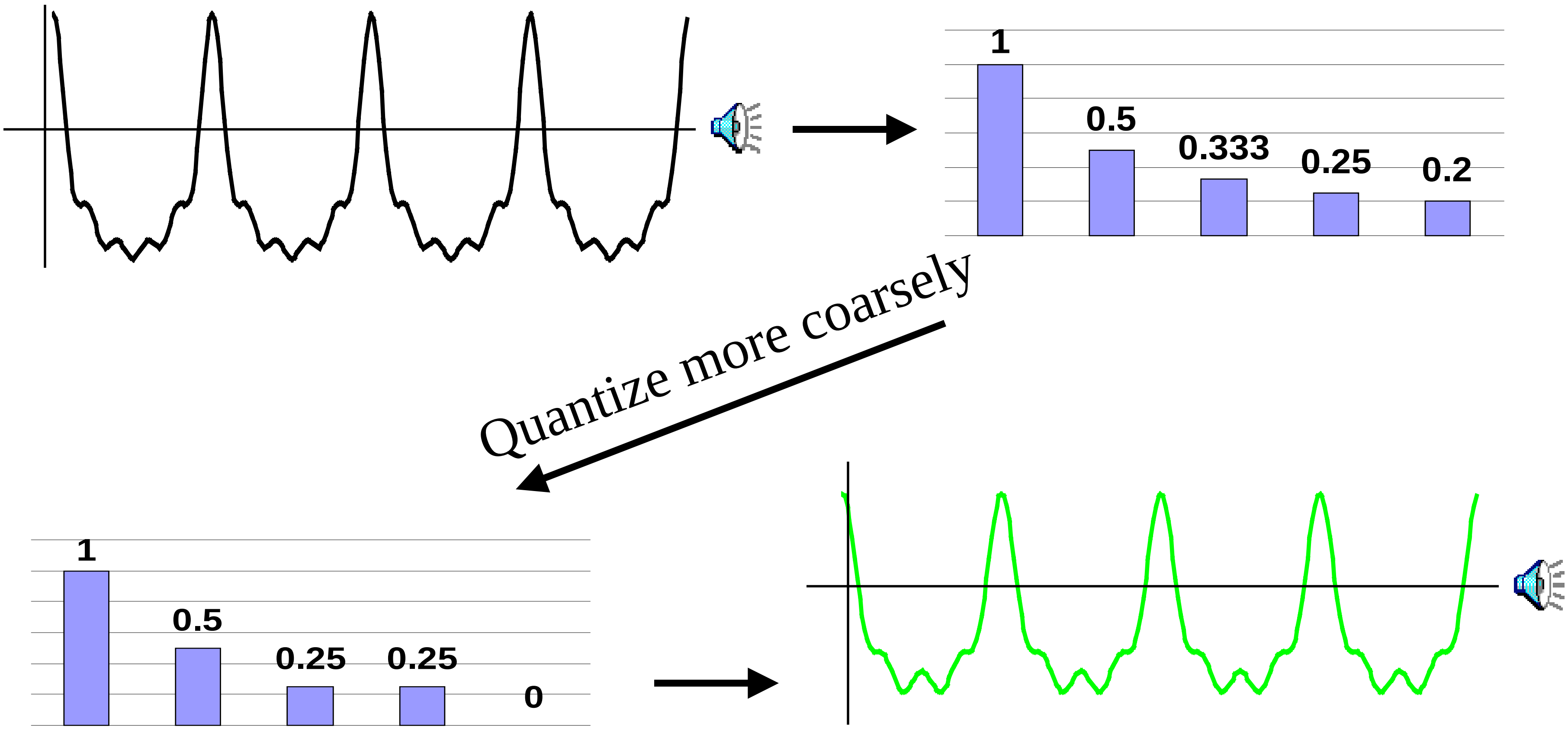
Inverse
(Synthesize)

Spectrum

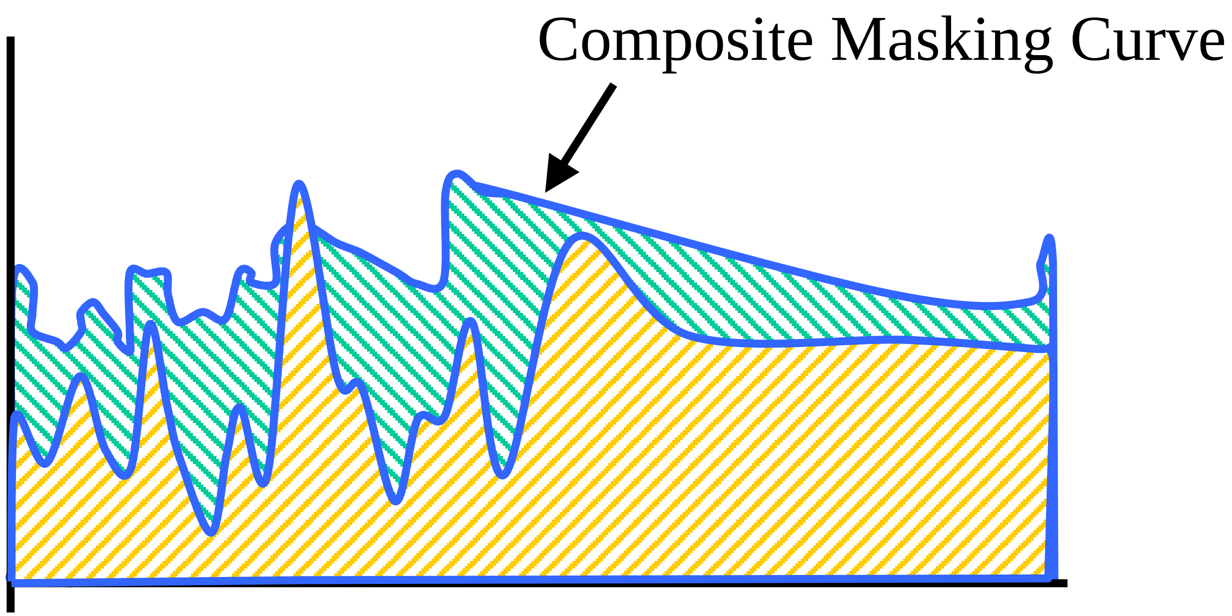


Frequency Domain

Review:
Leave something out -> Noise

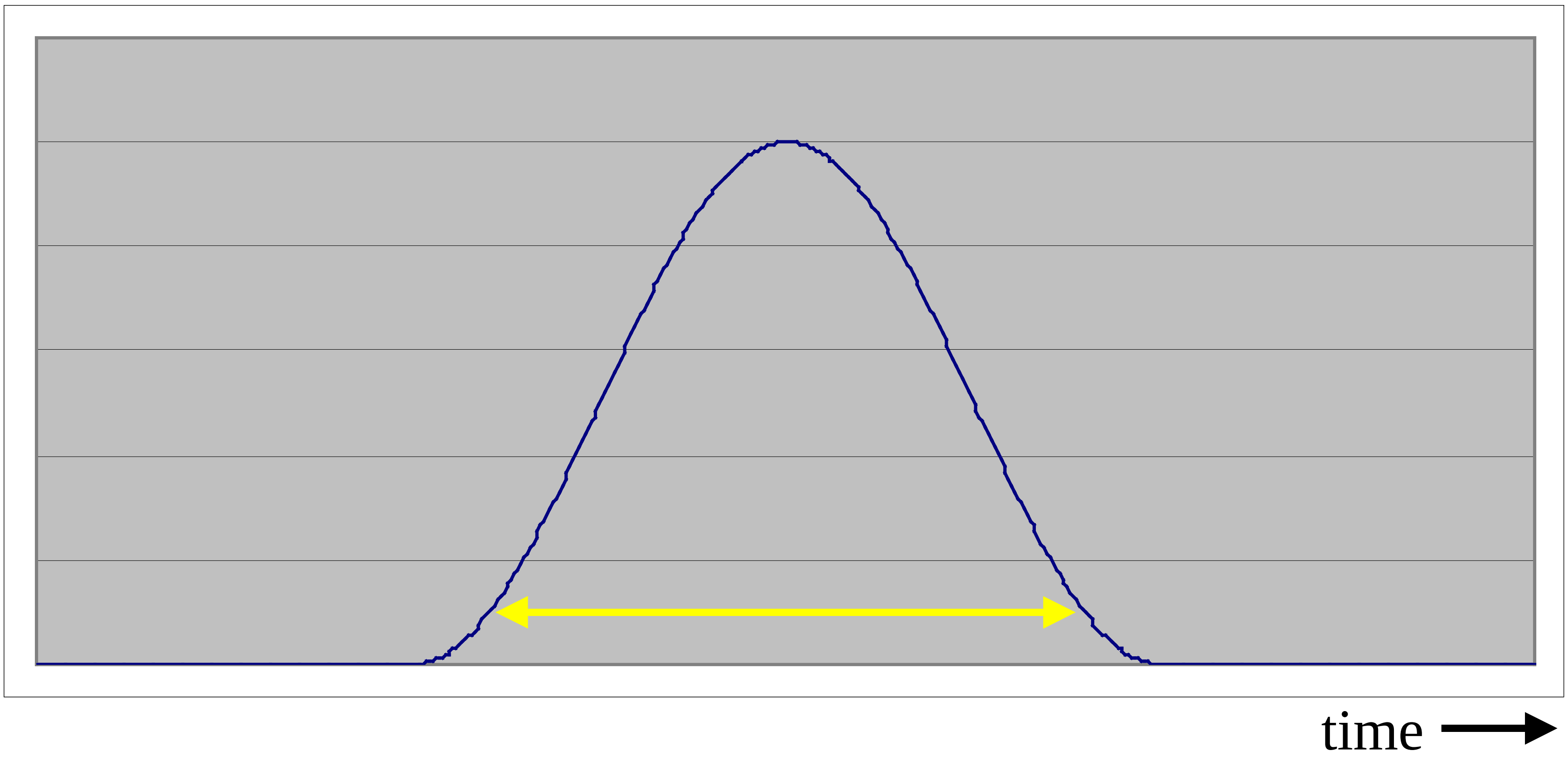


Review:
Masking: Where noise might fall --- in frequency domain



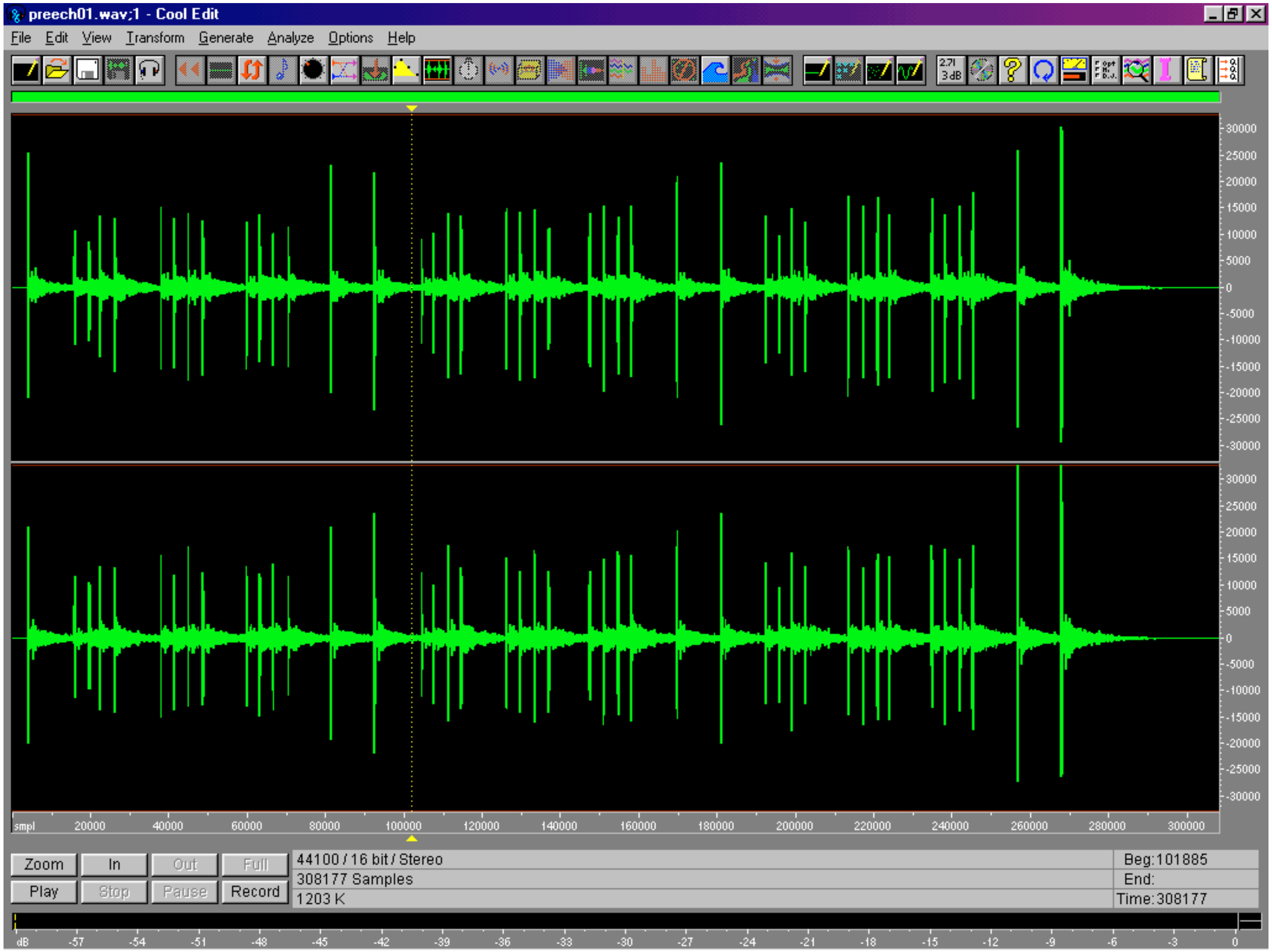
After Davidson et al., 1994

Where does the noise fall in time domain after decoder?



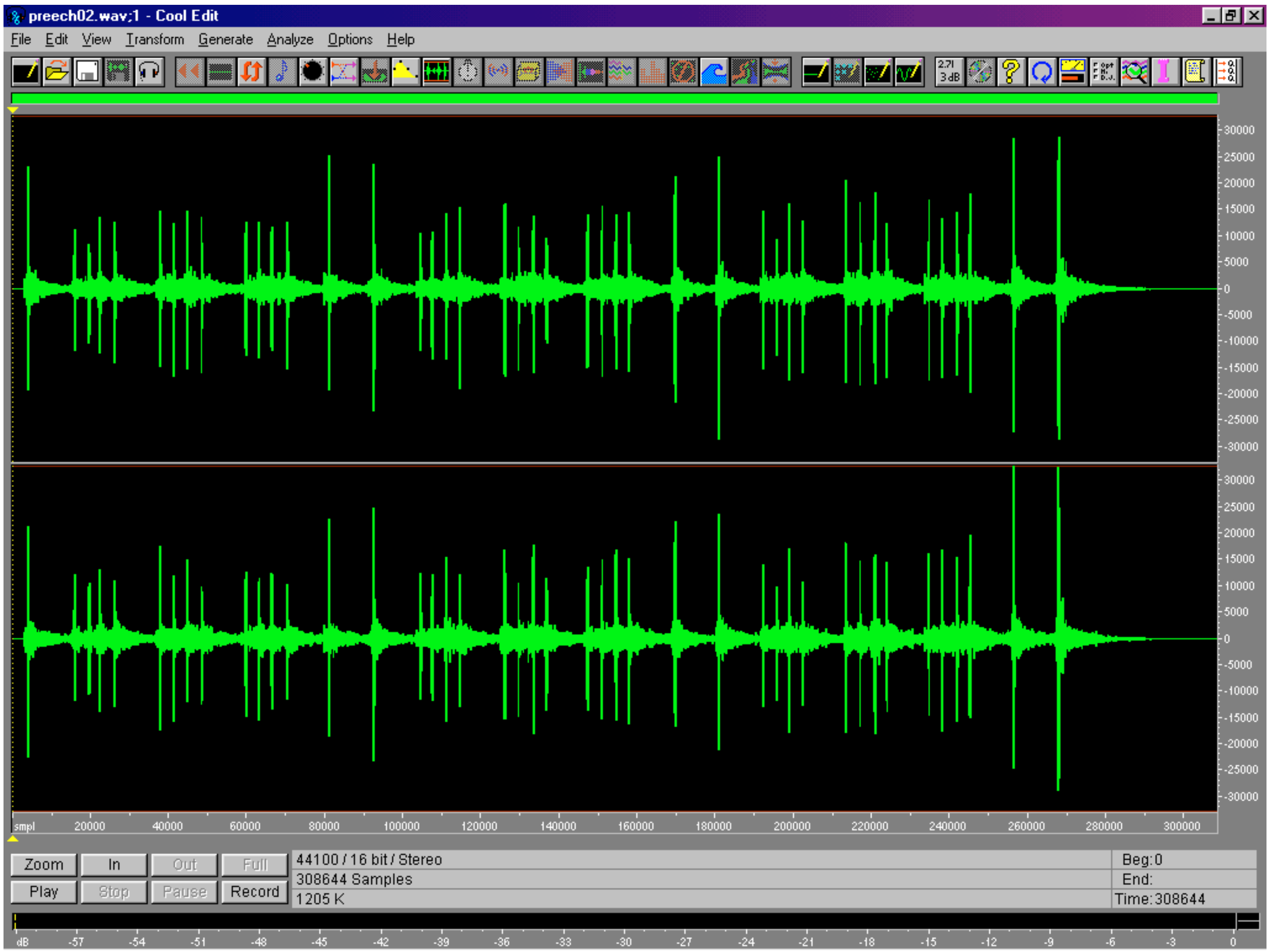
Review Sound Examples

Original: Castanets



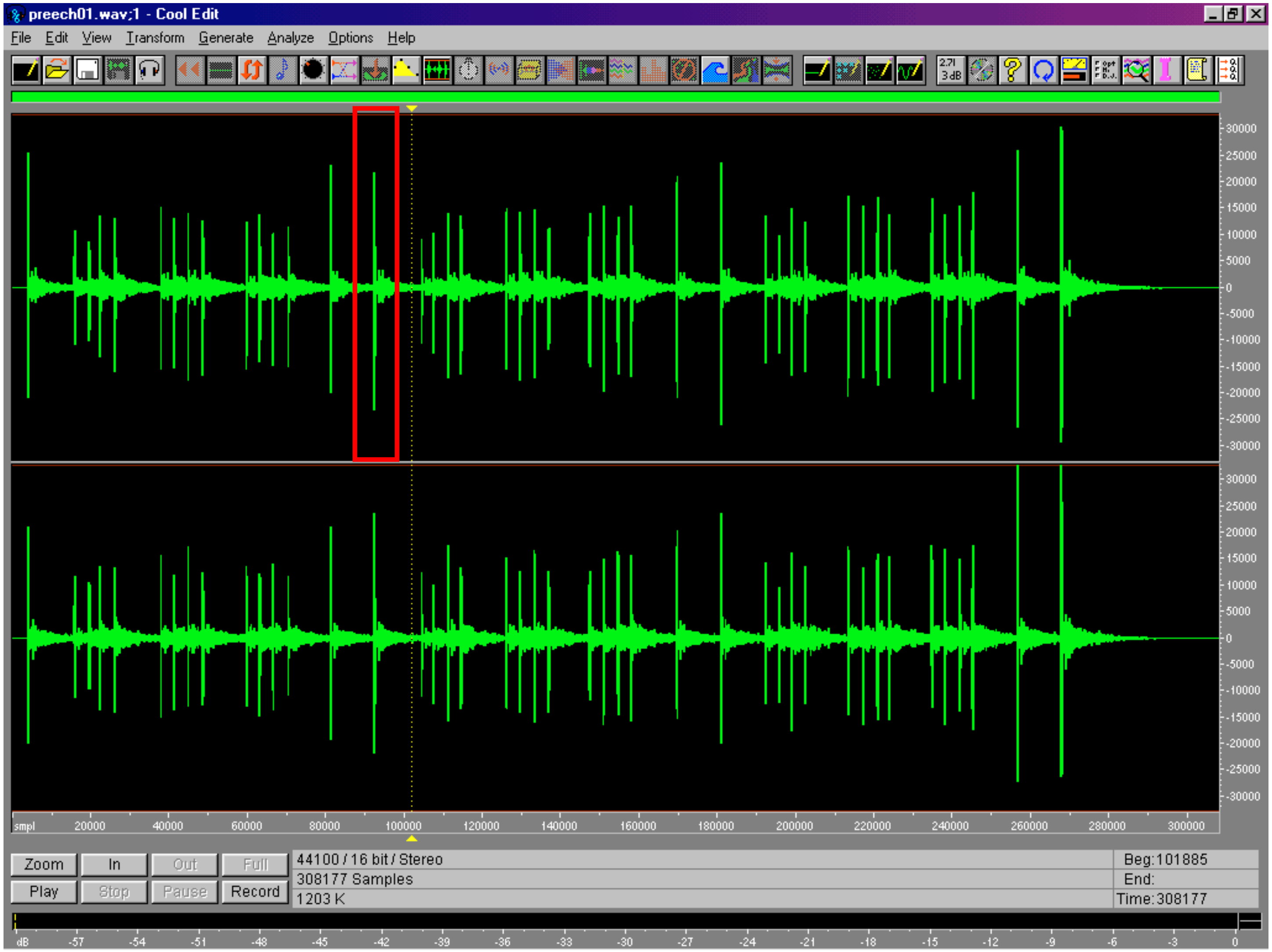
Source: AES CD-ROM

Encoded/Decoded: Pre-echo

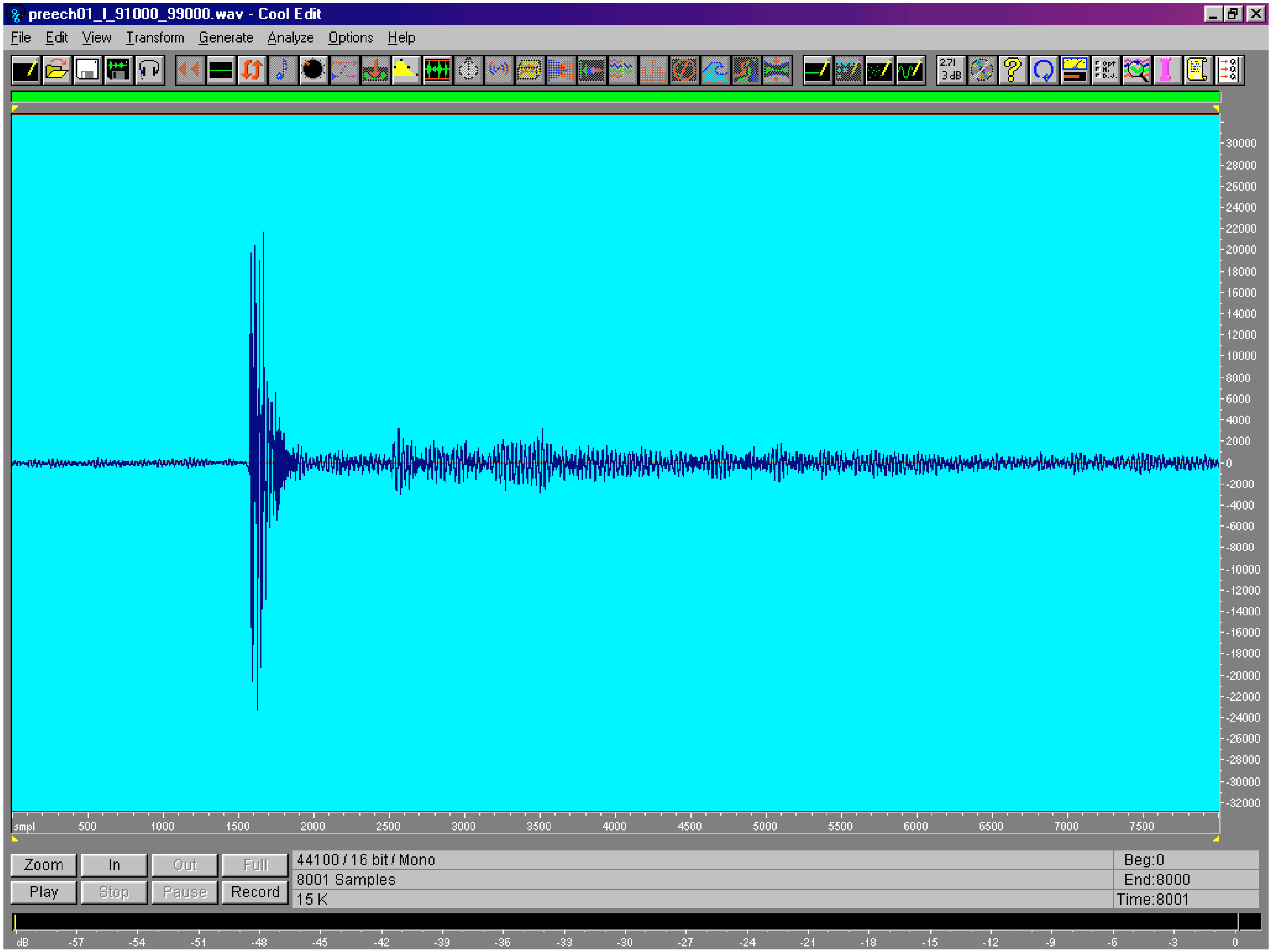


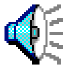
Source: AES CD-ROM

Original: Zoom in on one strike



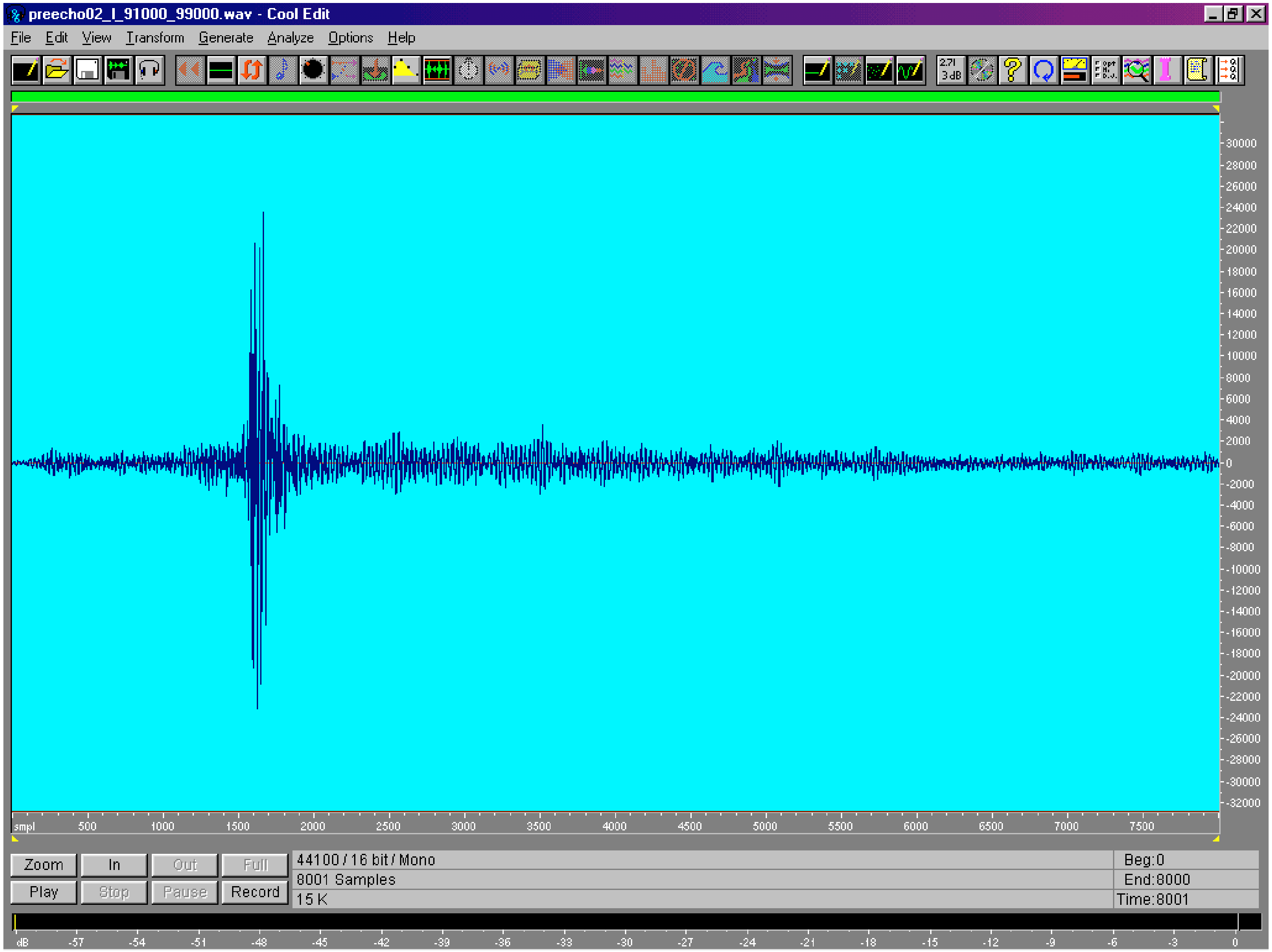
Original, 16.66 msec=8000 samples





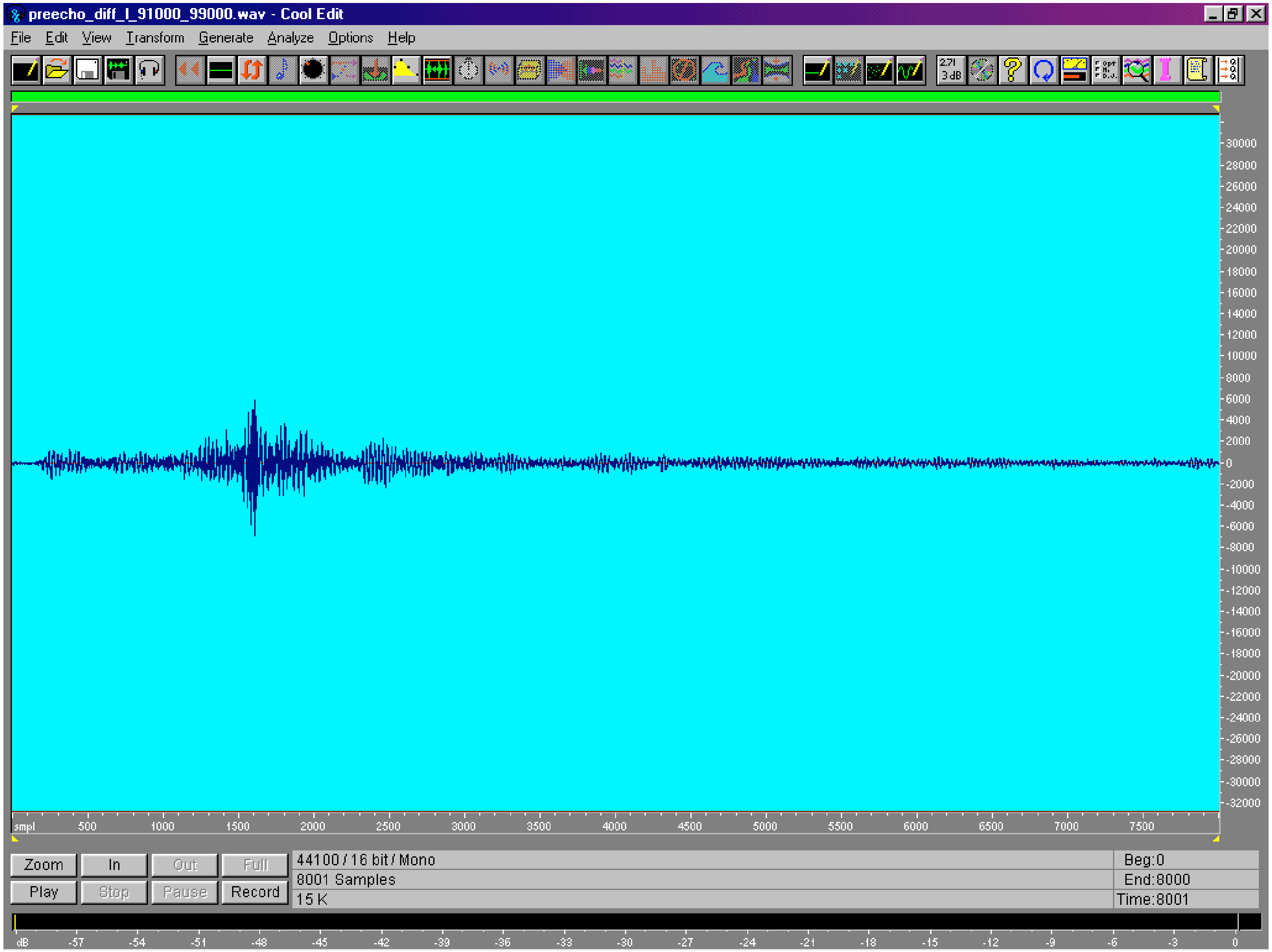
8x

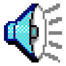
With pre-echo



8x

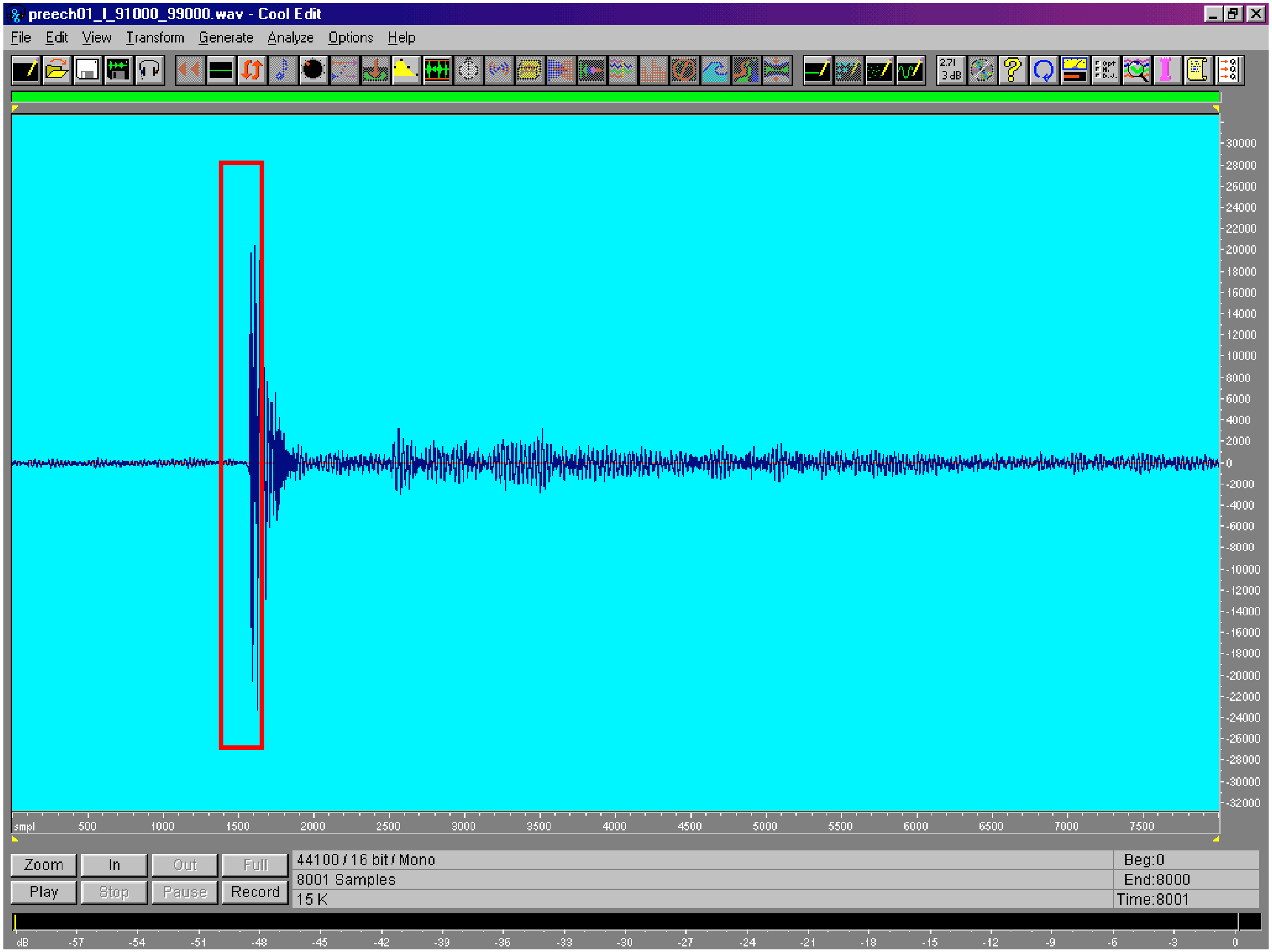
Difference



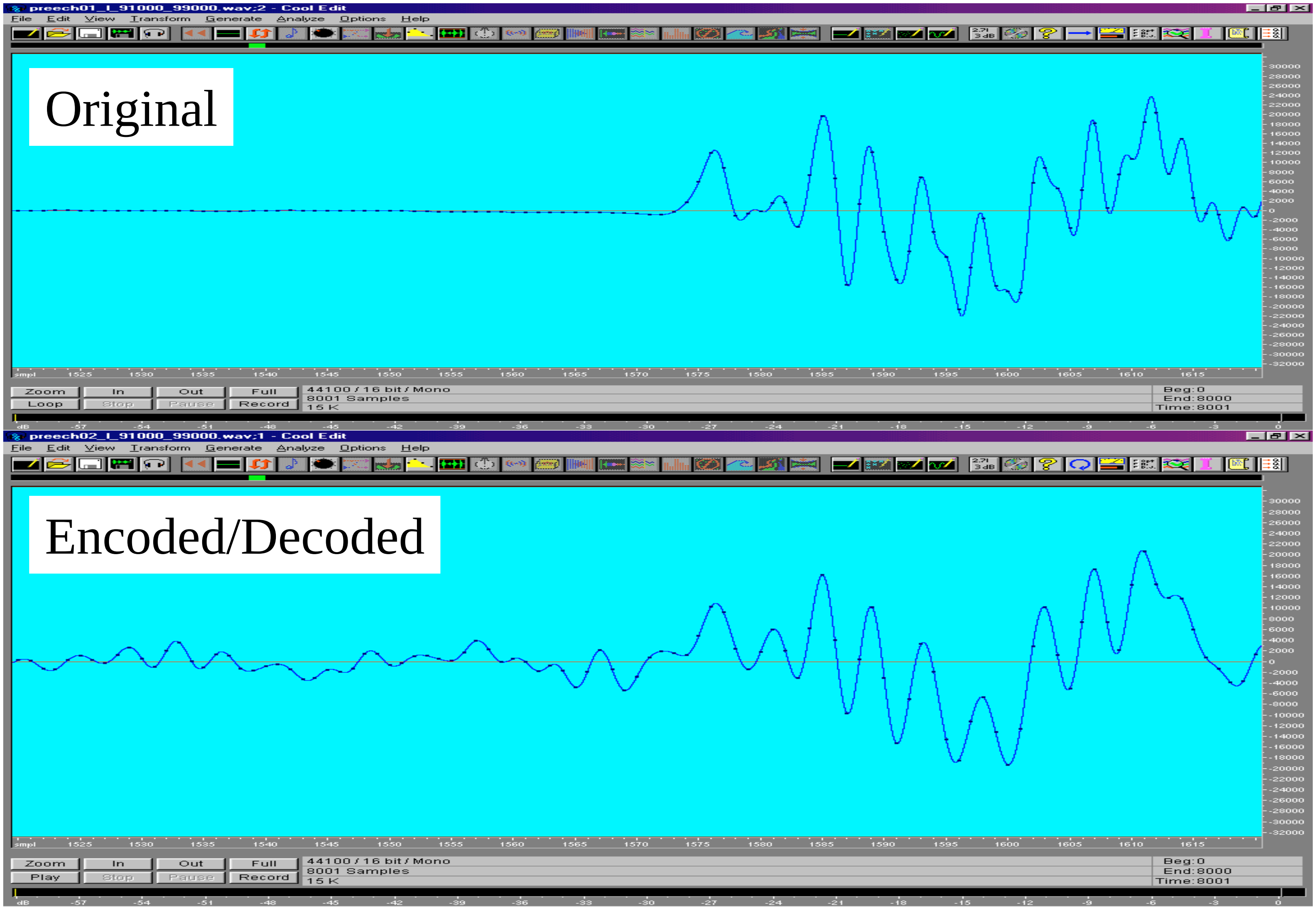


8x

Zoom in on attack



Attack, 2.2 msec (100 samples)



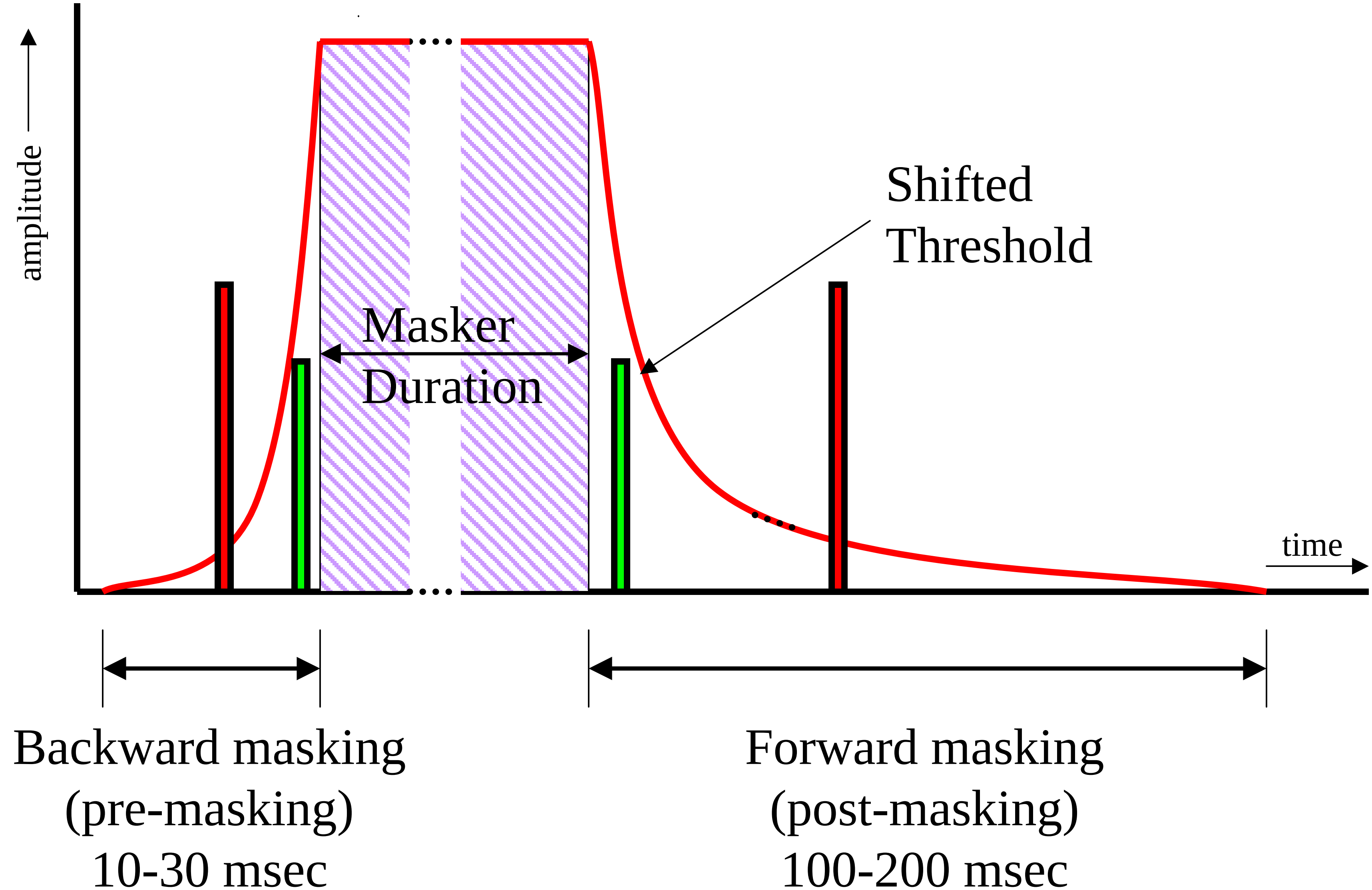
Samples 1520:1620 of preech0X_l_91000_99000.wav

What we will cover

- Windowing
- Pre-echo
  - What is it
  - Why it happens
- Temporal (not simultaneous) masking
- Error Recovery

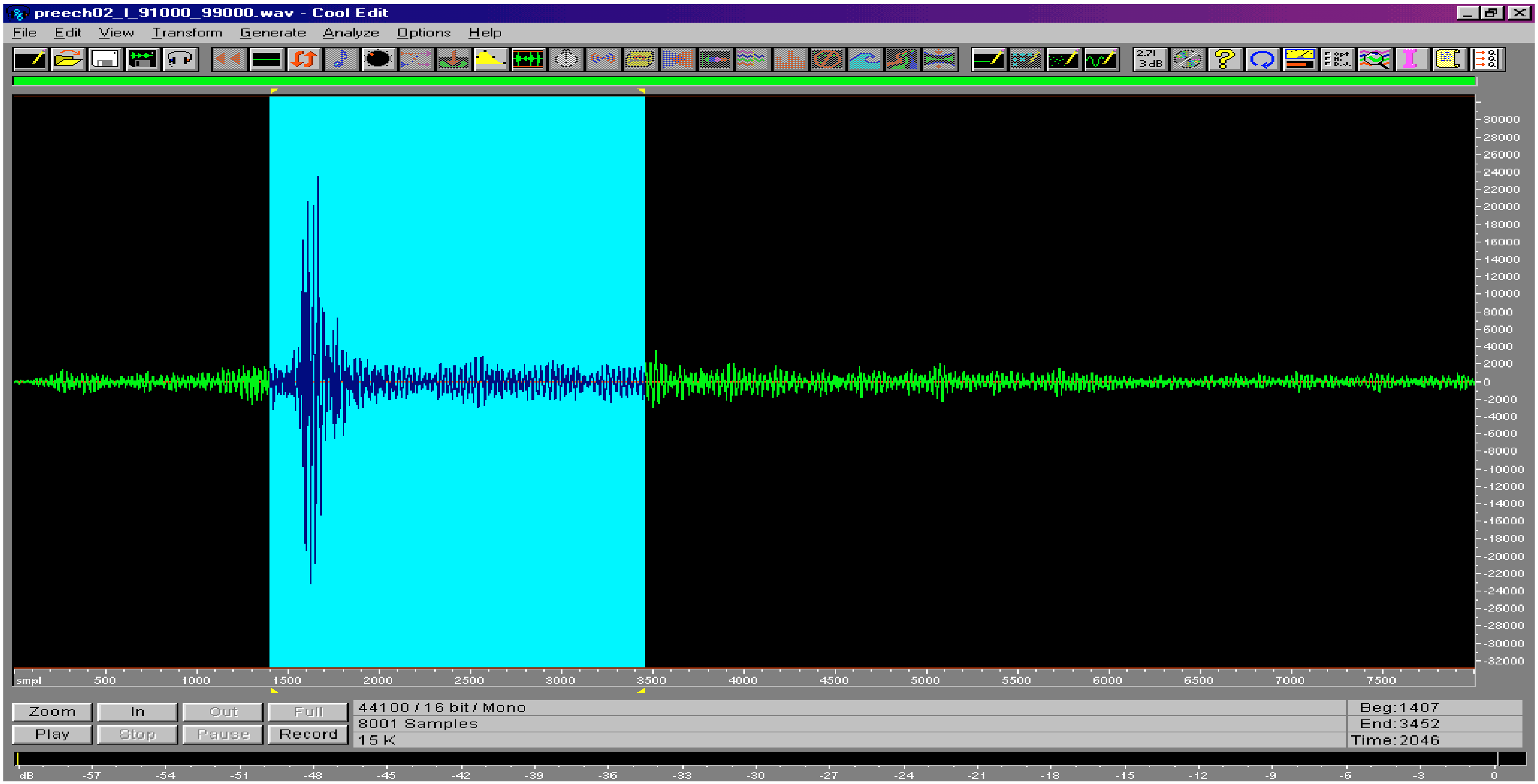
Temporal Masking

Derived from
Zwicker/Fastl p. 78,
Buser/Imbert p. 47



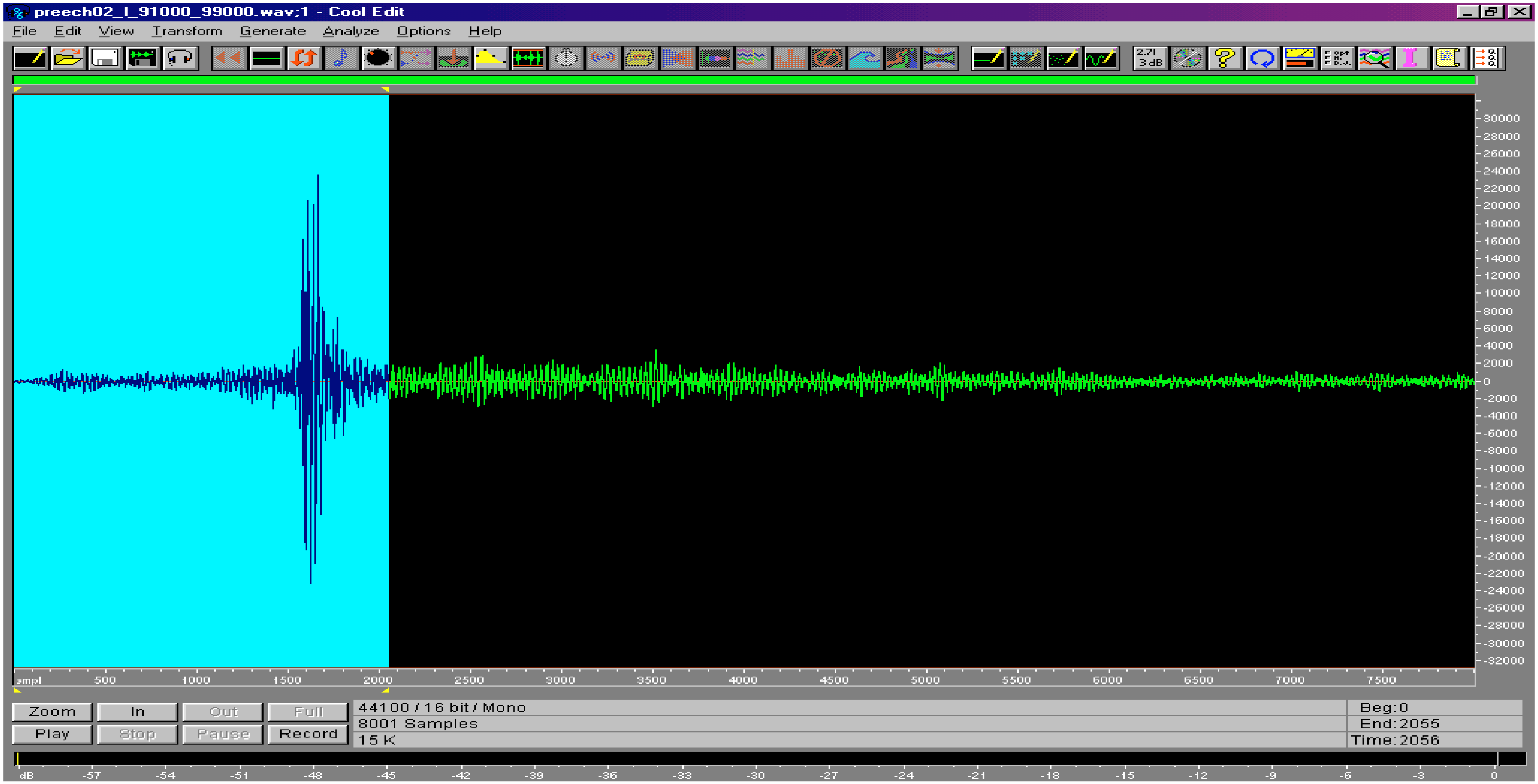
Where is Window? (Best Case)

AAC Window width = 2048 samples = 46 msec



Where is Window? (Worst Case)

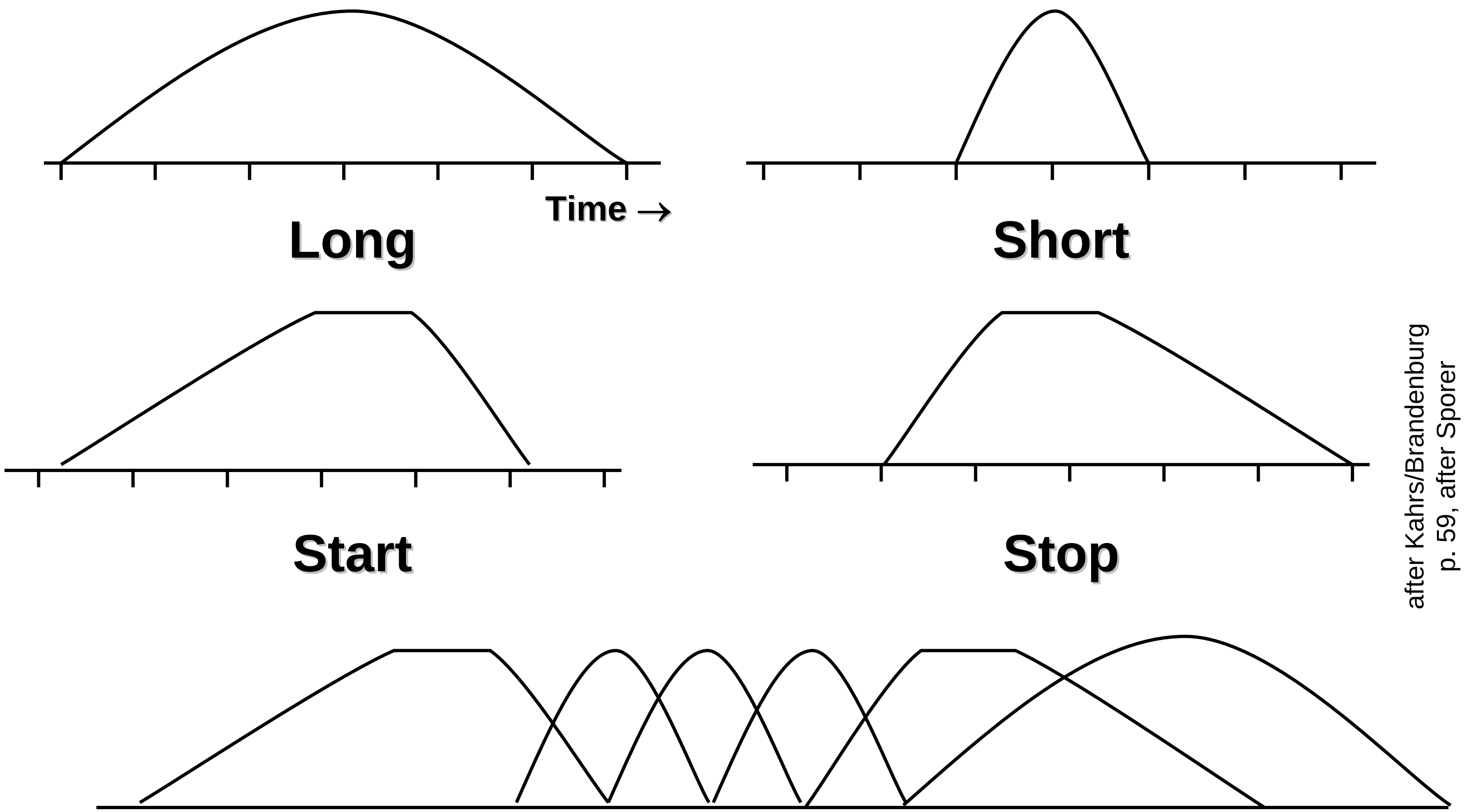
AAC Window width = 2048 samples = 46 msec



Review:
Window lengths (no. samples)

MPEG-1 Layer 3	192, 576
MPEG-2 AAC	256, 2048
AC-3	256, 512
MLP	40-160
DTS	1024, 2048
PAC	256, 2048

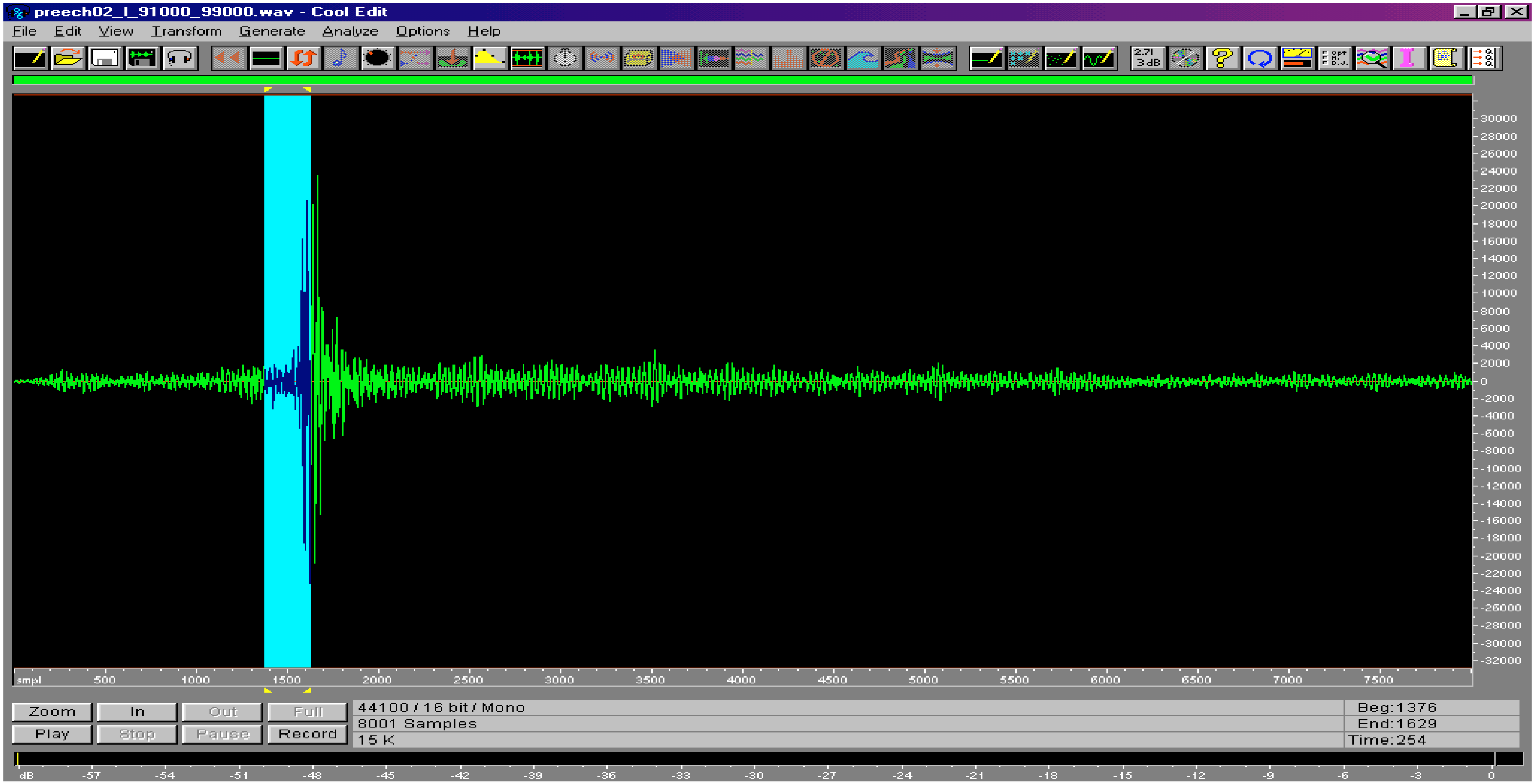
Review: MP-3 windows



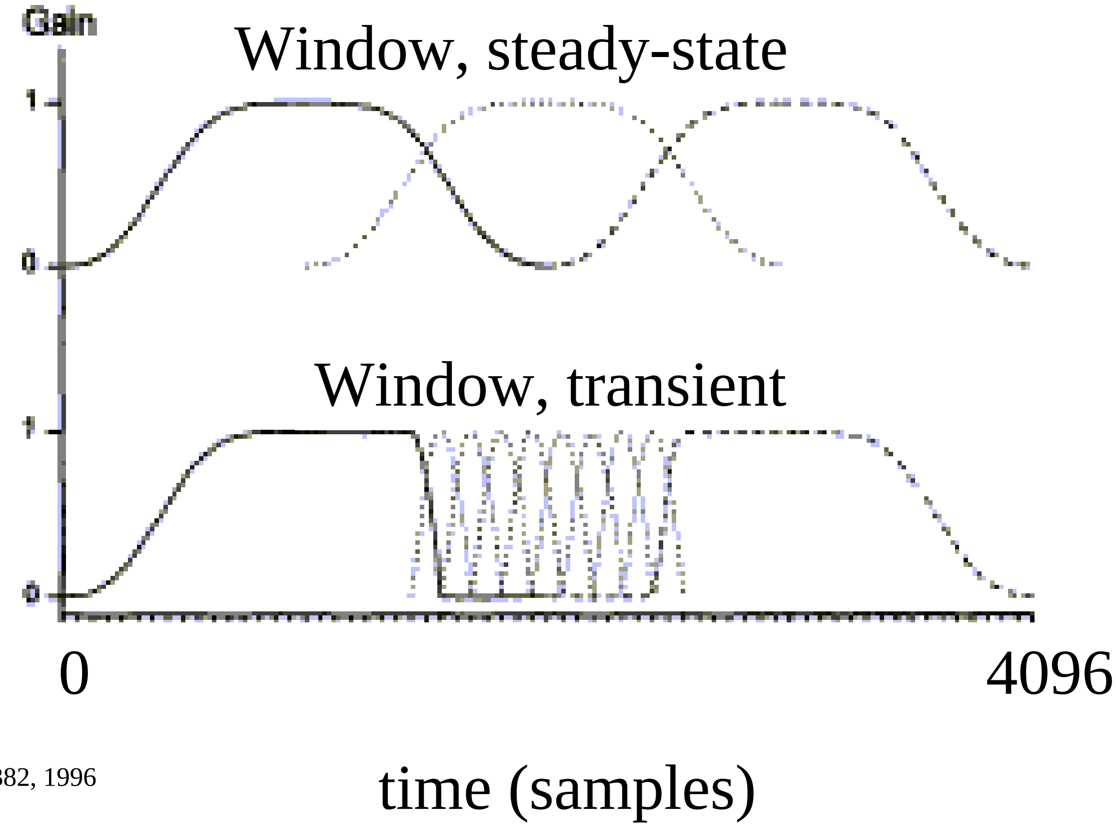
after Kahrs/Brandenburg
p. 59, after Sporer

Where is Window? (Solution)

Short window width = 256 samples = 5.8 msec



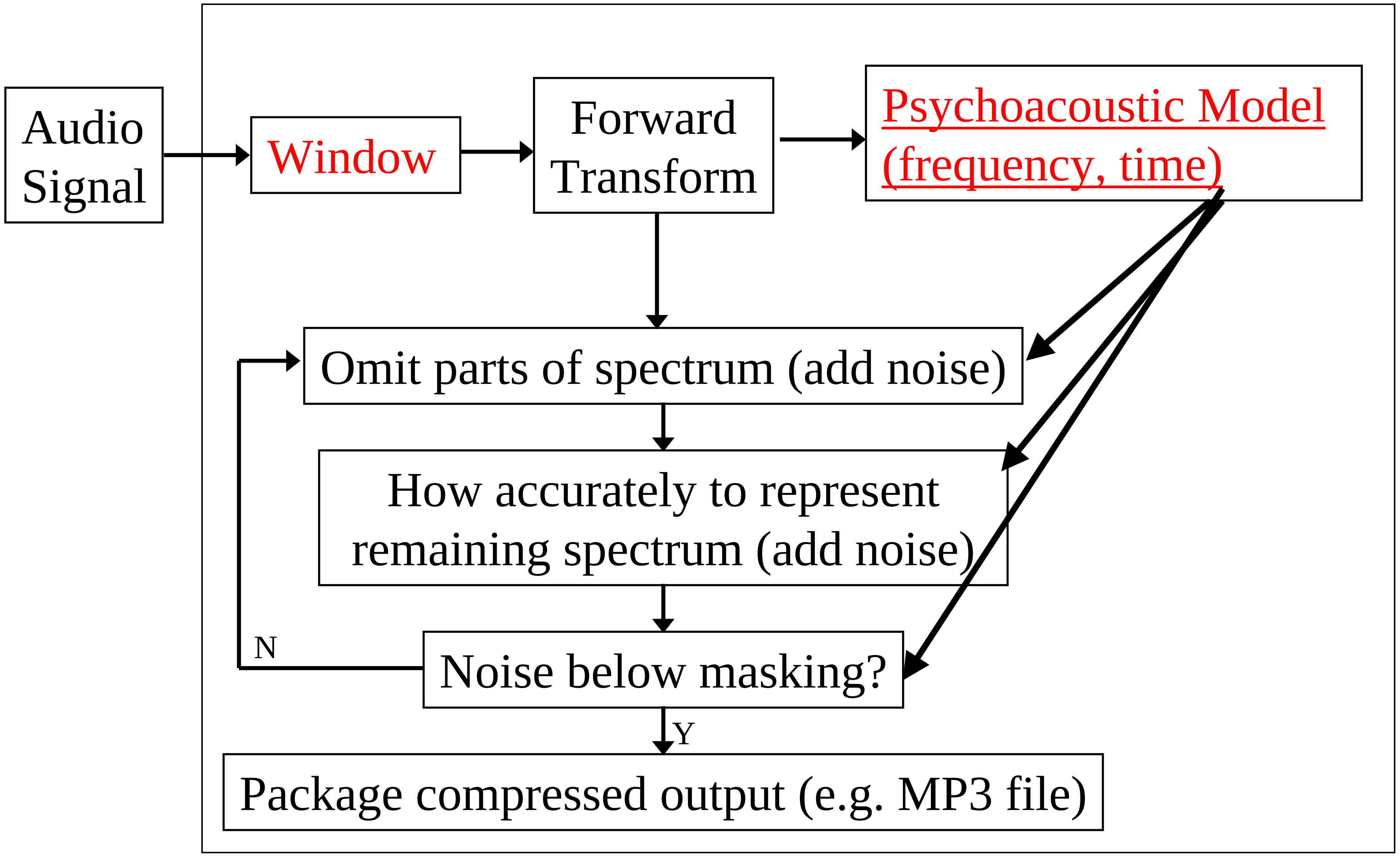
Review: MPEG-2 AAC Windowing



From: Bosi et al, AES preprint 4382, 1996

The Psychoacoustic Model (2)

Perceptual Encoder



Meeting the challenge

- Coarser Quantization (time domain)
- DPCM, ADPCM
- Linear Prediction
- Subband coding
- Transform to frequency domain
- Coarser quantization (frequency domain)
- Psychoacoustics: mask the noise
- Variable bit rate
- Noiseless coding
- Window
- Temporal masking

Even more of what's really in an "MP3" file?

Decoder (“mp3 player”)

Unpackage compressed audio (e.g. MP3 file)

↓

↓

Change from frequency to time

↓

Overlap-add

↓

Audio Signal

Next listening session

- Meet by ______ in listening areas.
- Listen to tracks Aida Stereo 0, 1, 2, 3 in Session: Aida Stereo. Is 3 different from 0?
- Take notes and discuss:
  - What do you hear?
  - How are they different?
  - What do you like?
  - What don’t you like?
- Back here at 9:20 to discuss. (20 min)

Notes on Sound Examples

Discussion of Sound Examples

0

1

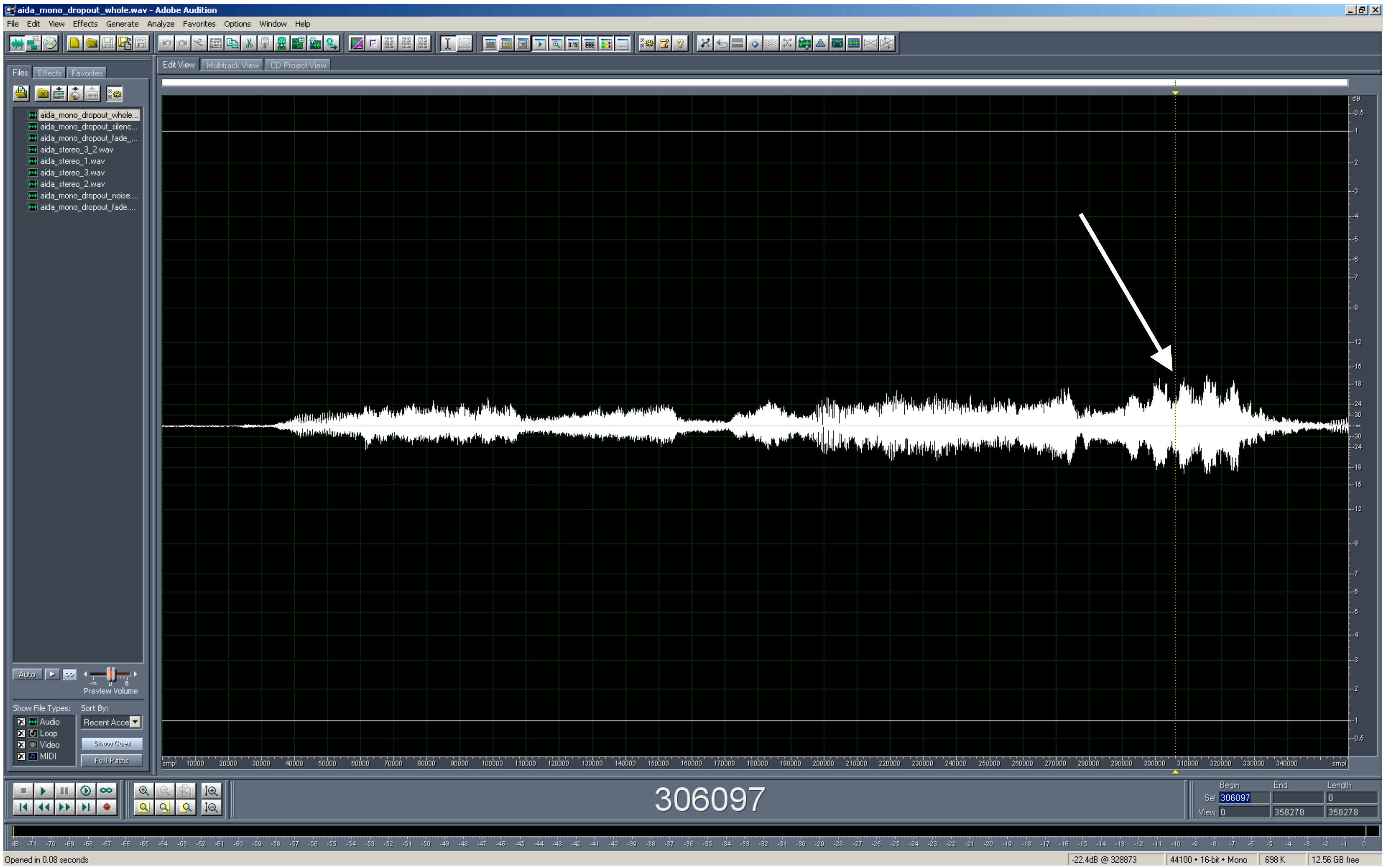
2

3

What we will cover

- Some history of (perceptual) coding
- Main codec families, their names, key features
- Lossless coding
- Error Recovery

Decoder Error Recovery: Original



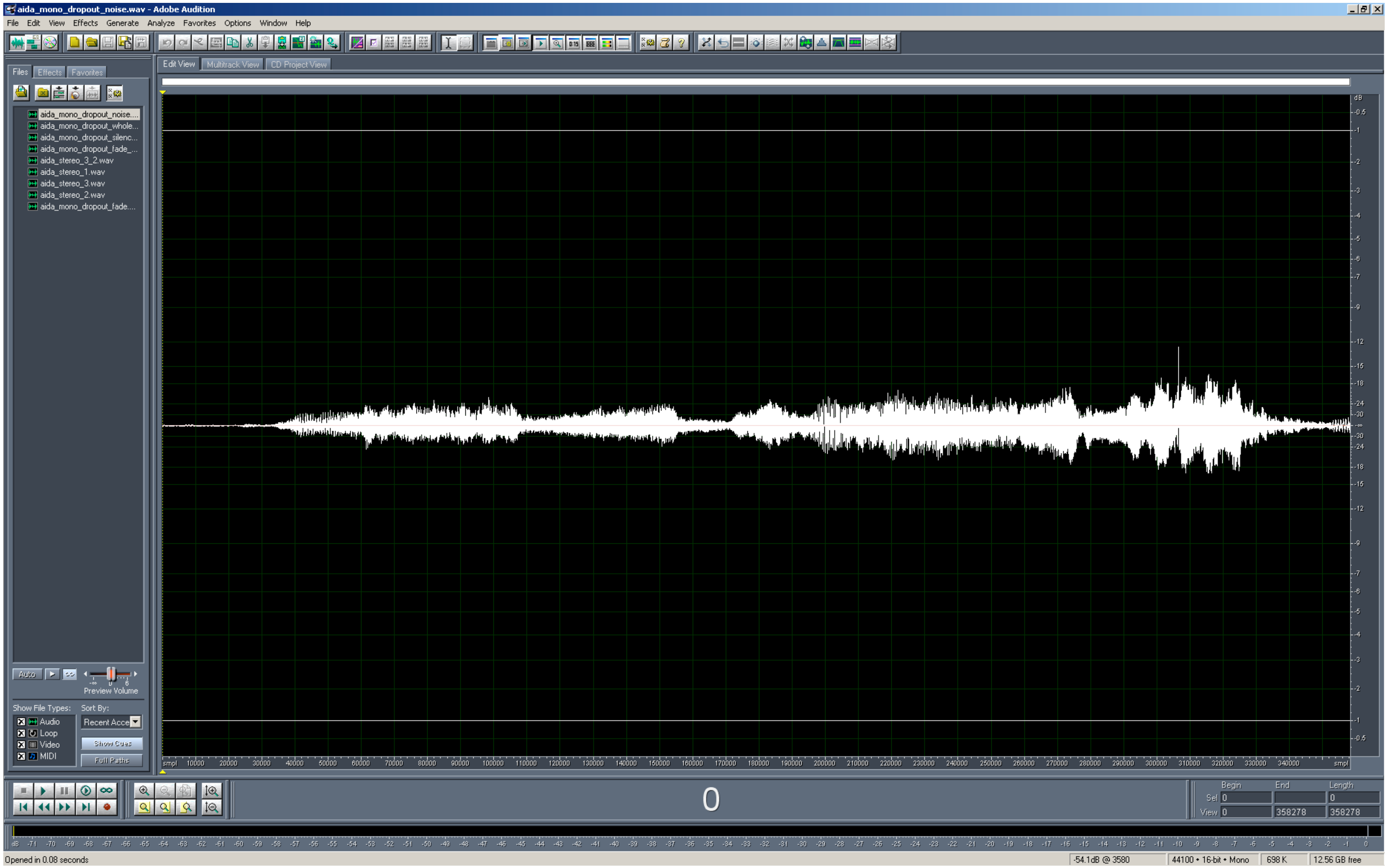
Decoder Error Recovery

- One frame = 1024 samples, 44.1 kHz

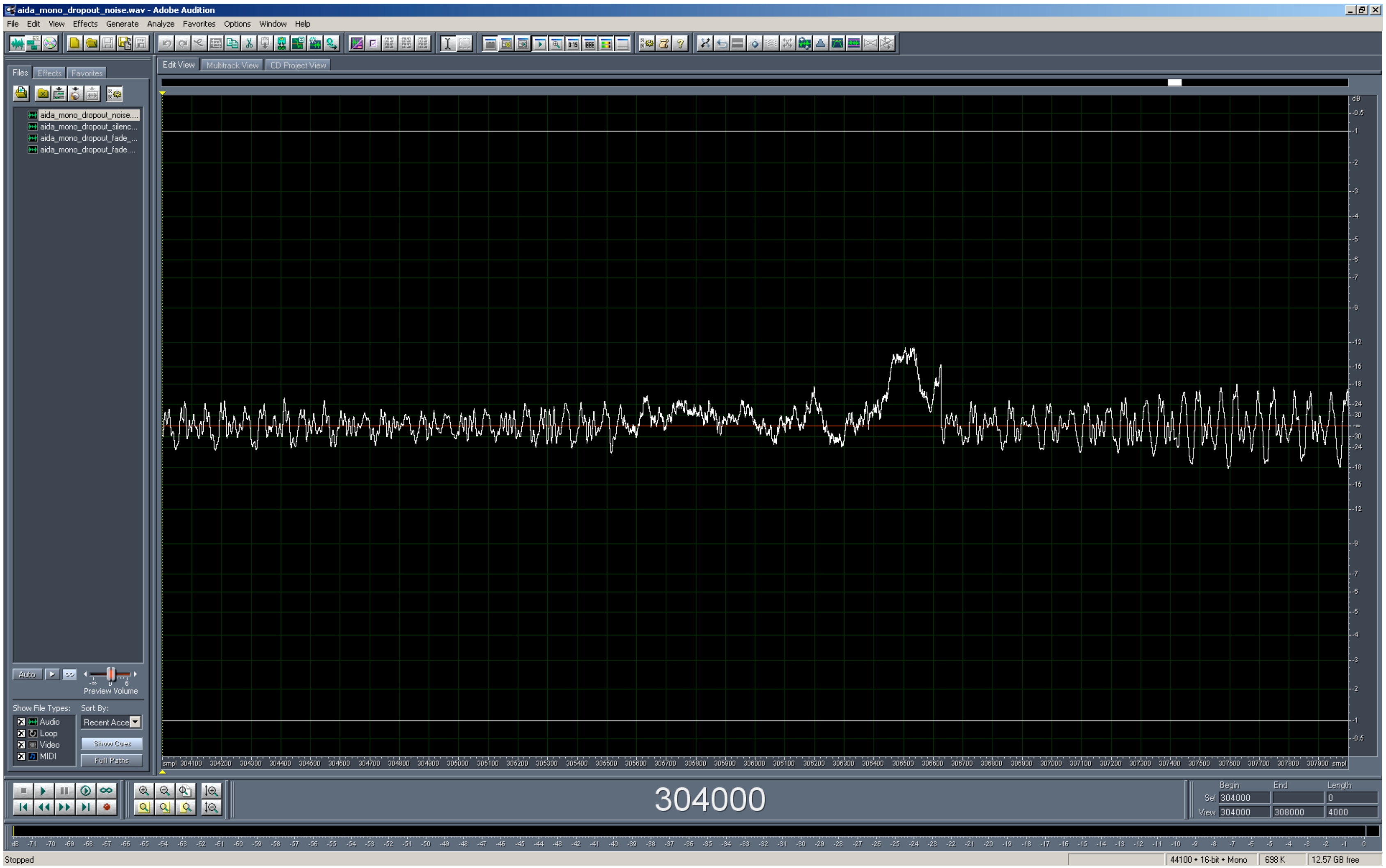
  = ? msec

  = ? Inches @ 7.5 ips?

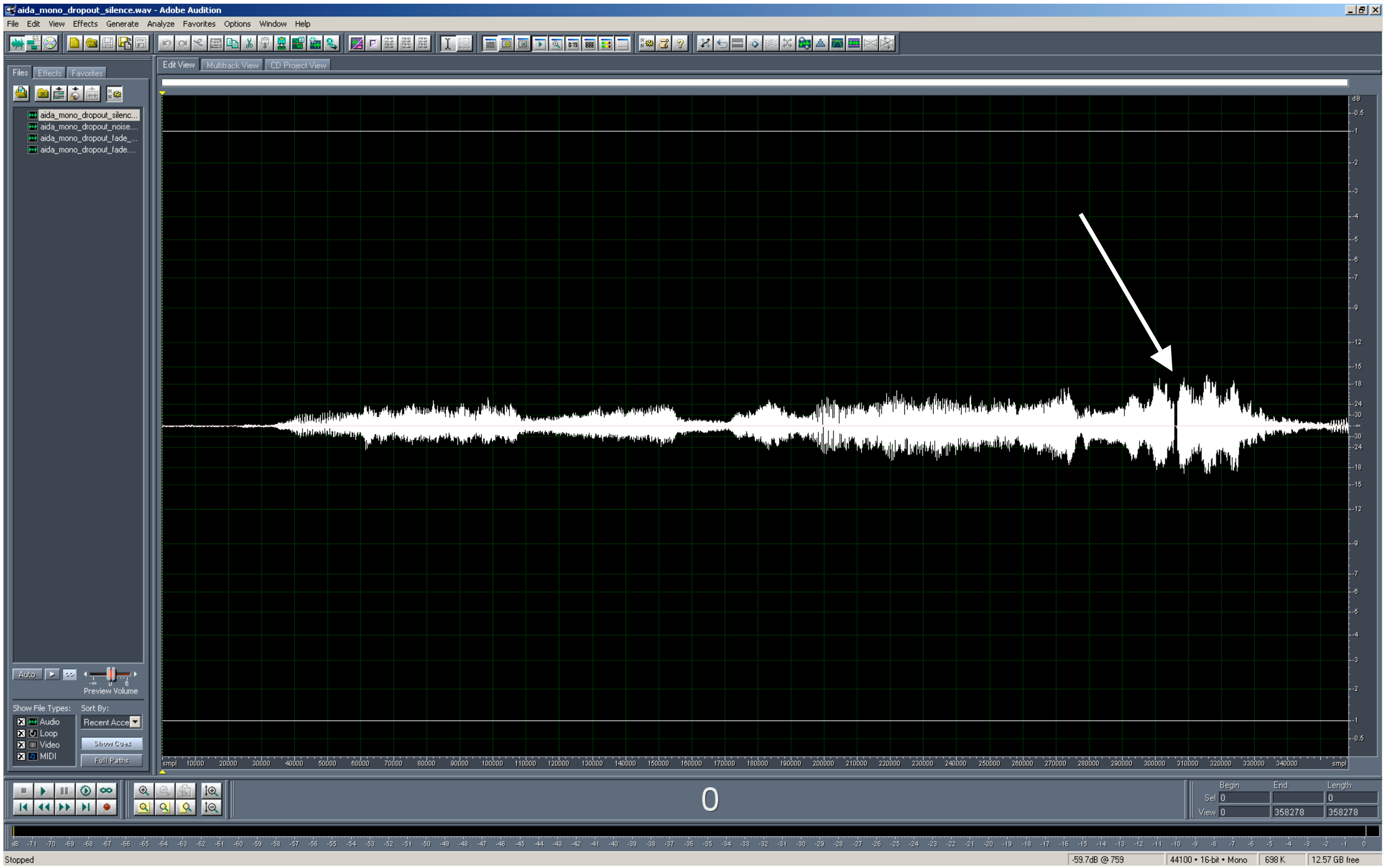
Decoder Error Recovery: Noise?



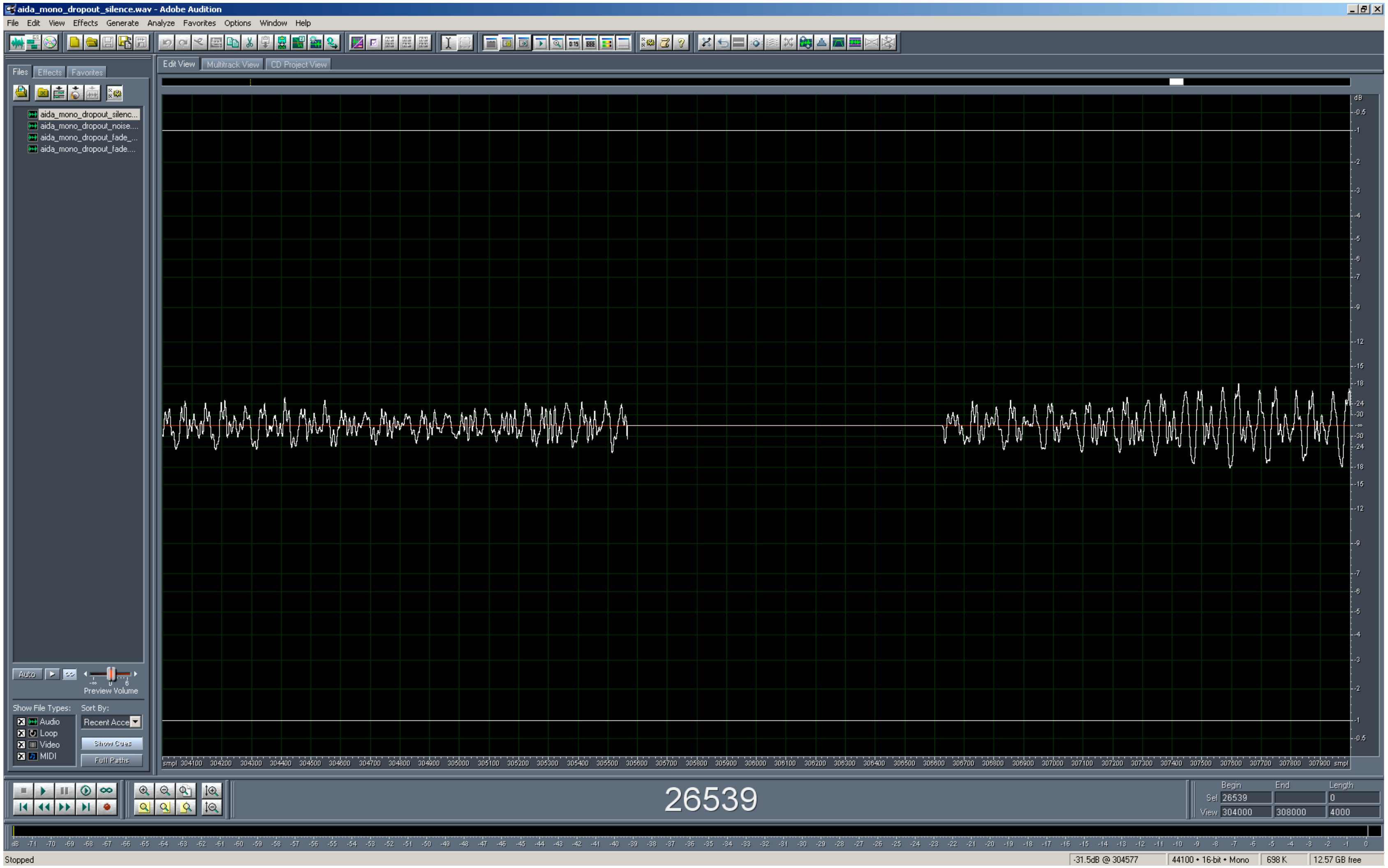
Decoder Error Recovery: Noise?



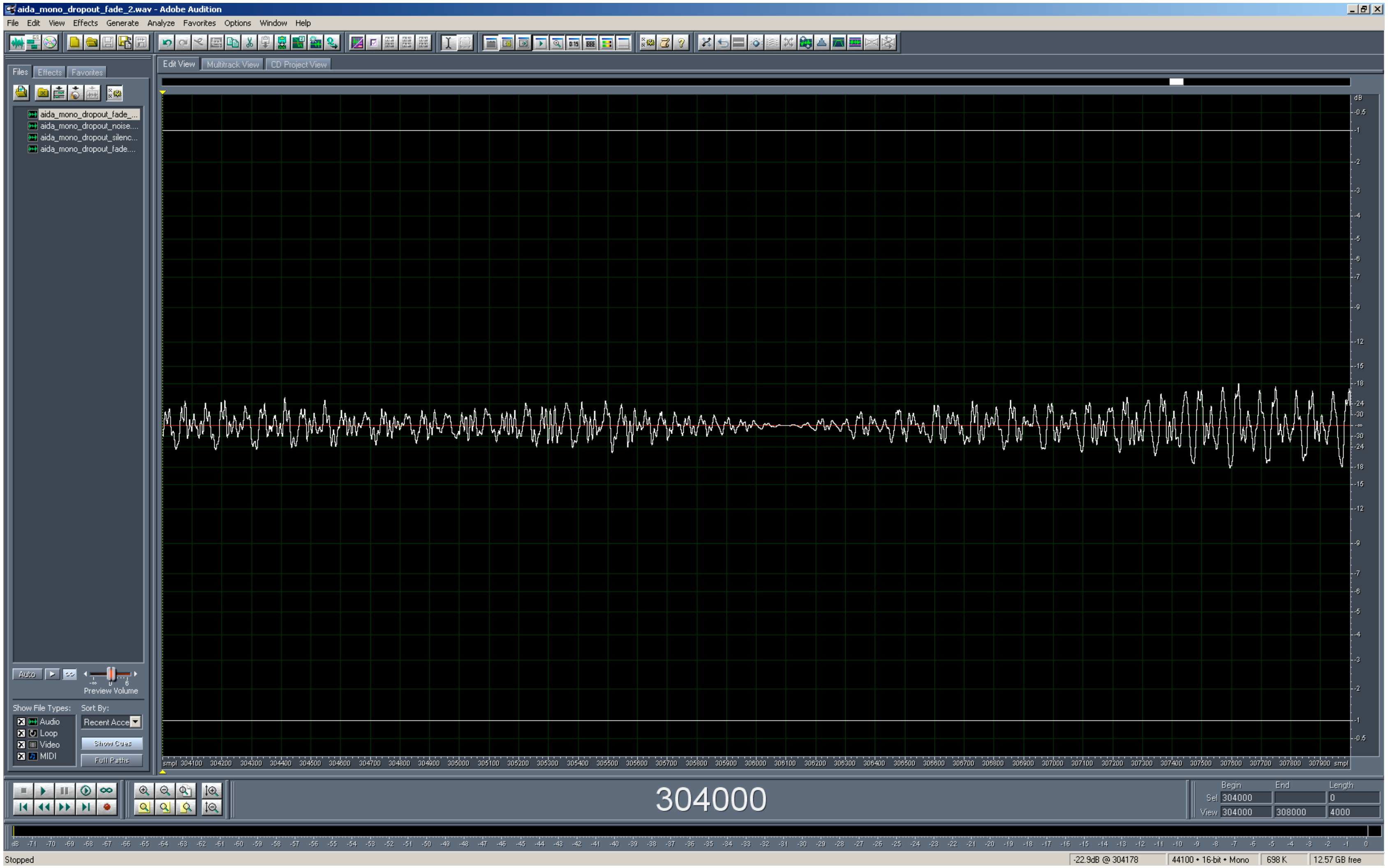
Decoder Error Recovery: Silence?



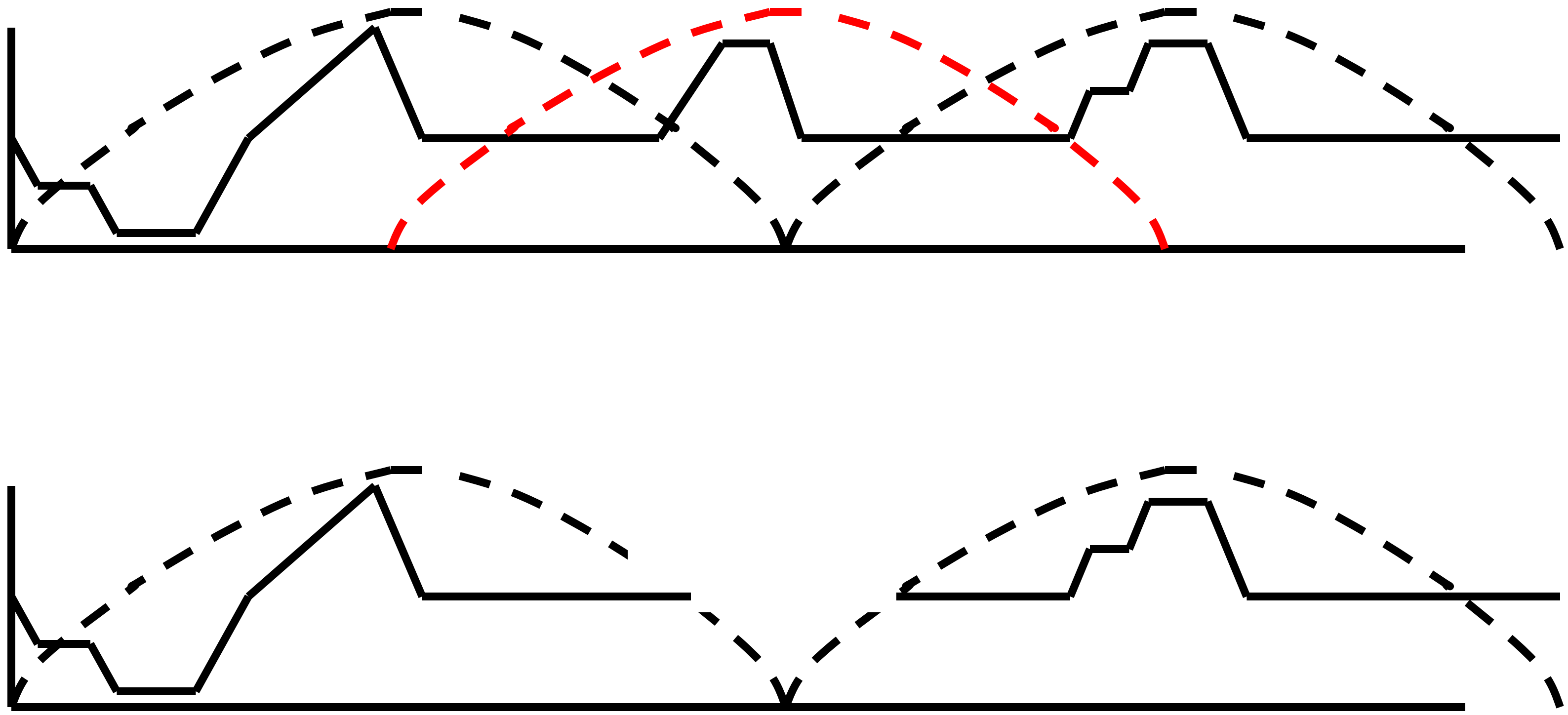
Decoder Error Recovery: Silence?



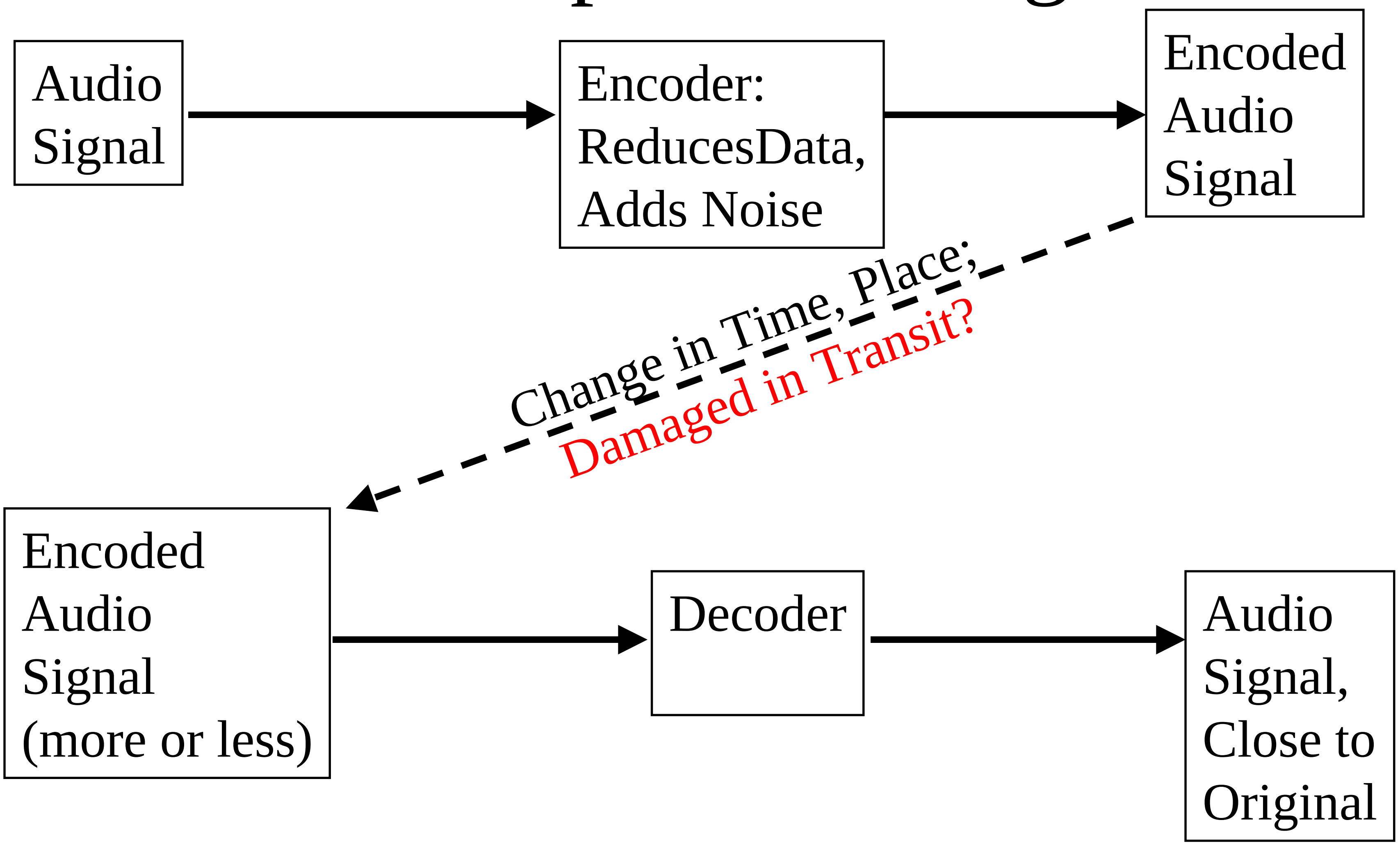
Decoder Error Recovery: Fade?



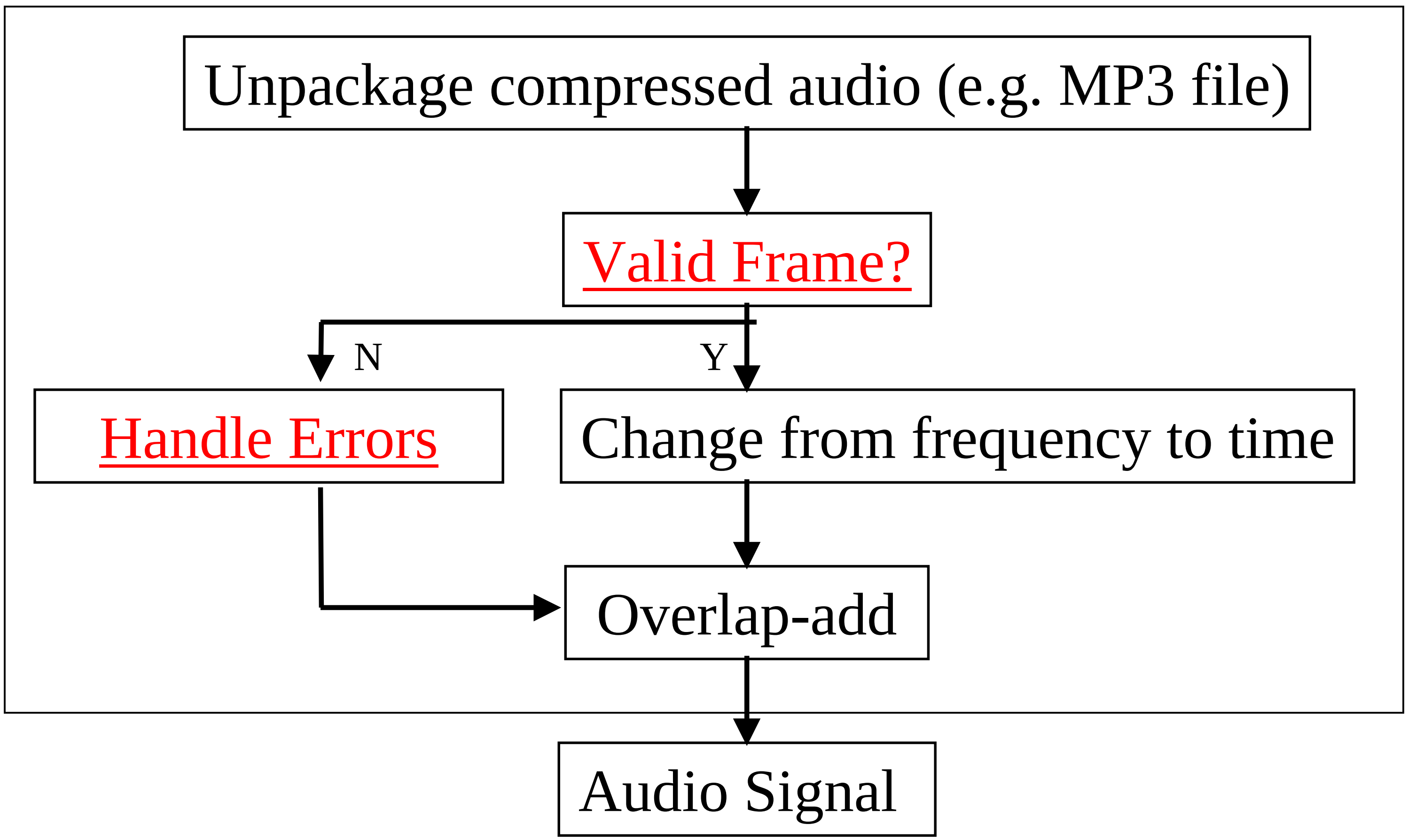
Fadeout: For Free?



Perceptual Coding



Decoder ("mp3 player")



What's really in an "MP3" file ---additional stuff?

Meeting the challenge

- Coarser Quantization (time domain)
- DPCM, ADPCM
- Linear Prediction
- Subband coding
- Transform to frequency domain
- Coarser quantization (frequency domain)
- Psychoacoustics: mask the noise
- Variable bit rate
- Noiseless coding
- Window
- Temporal masking
- Error recovery

What we have covered

- Windowing
- Pre-echo
  – What is it
  – Why it happens
- Temporal (not simultaneous) masking
- Changing window size to prevent pre-echo
- Error Recovery

Look ahead: Sunday a.m.

- Discussion 9:00-10:30
  - Modification of one or two student projects
  - Volunteers?
- RoundTable, 10:30-12:00

Subtle Listening Session

- Meet by 9:40 a.m. in listening areas.
- Listen to tracks Stimag50 and Stimag54 in Session: Stimag. --- short!
- Take notes and discuss:
  - What do you hear?
  - How are they different?
  - What do you like?
  - What don't you like?
- Back here at 9:50 to discuss.

Notes on Sound Examples

MSRA 5500 - 002
MUS 4500 - 002

March 15, Session 3, Part 1

More than one channel

© Copyright 2008 John Strawn

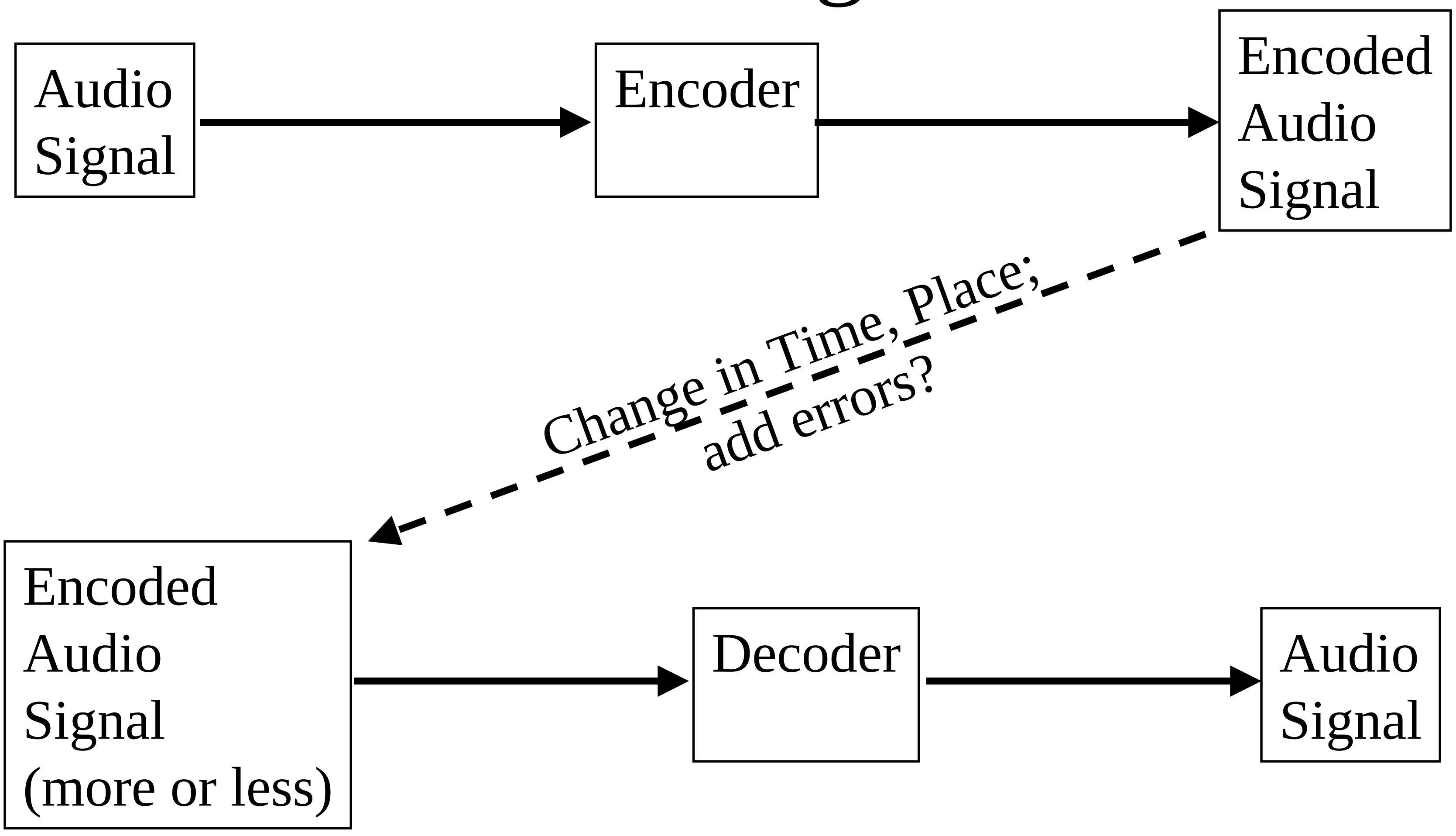
What we will cover

- Stereo perception
- Intensity Stereo coding
- Mid/Side coding

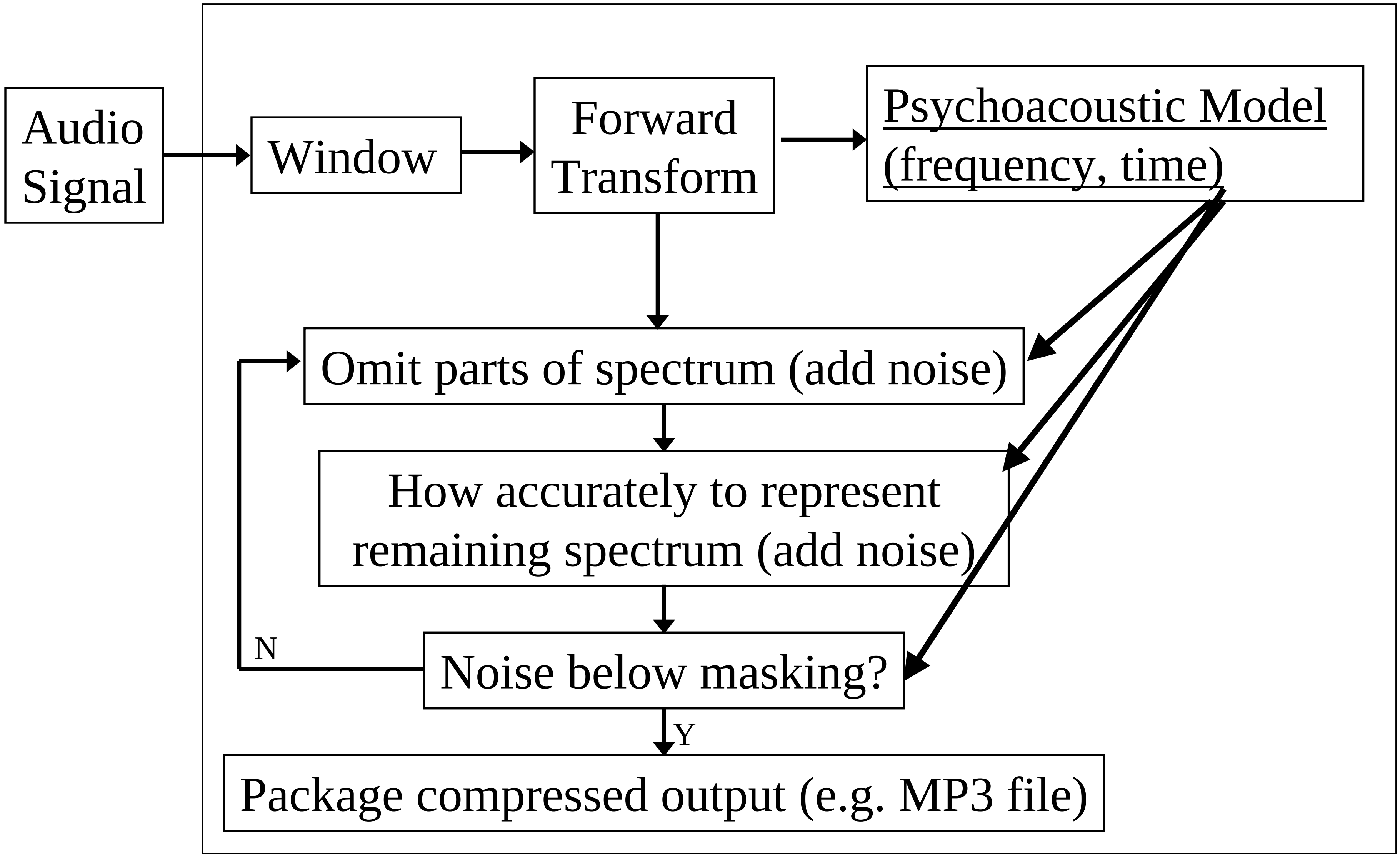
Discuss listening examples

Source: AES CD-ROM. Original recording by Jim Johnston.

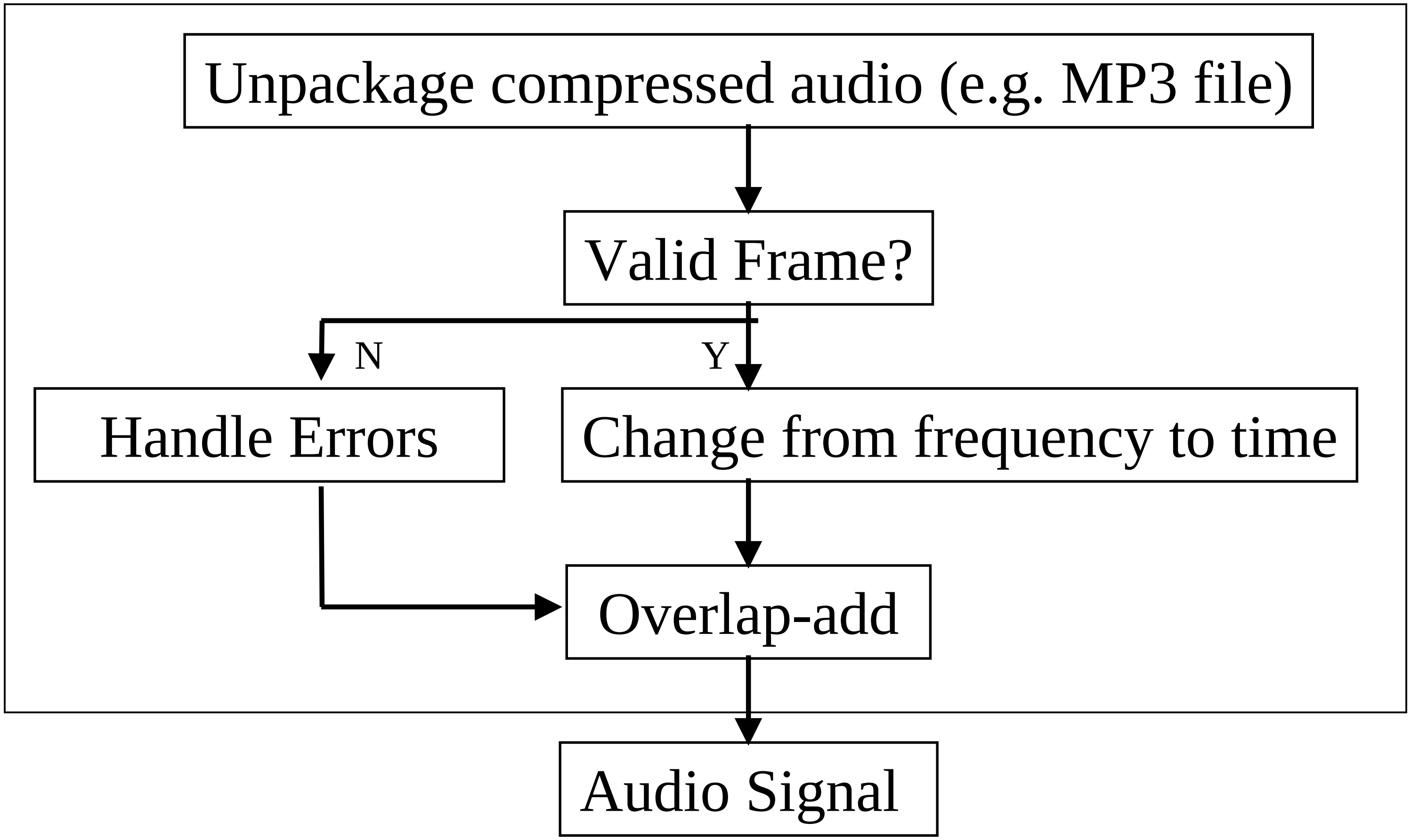
Review: Coding in General



Review: Perceptual Encoder



Review: Decoder (“mp3 player”)

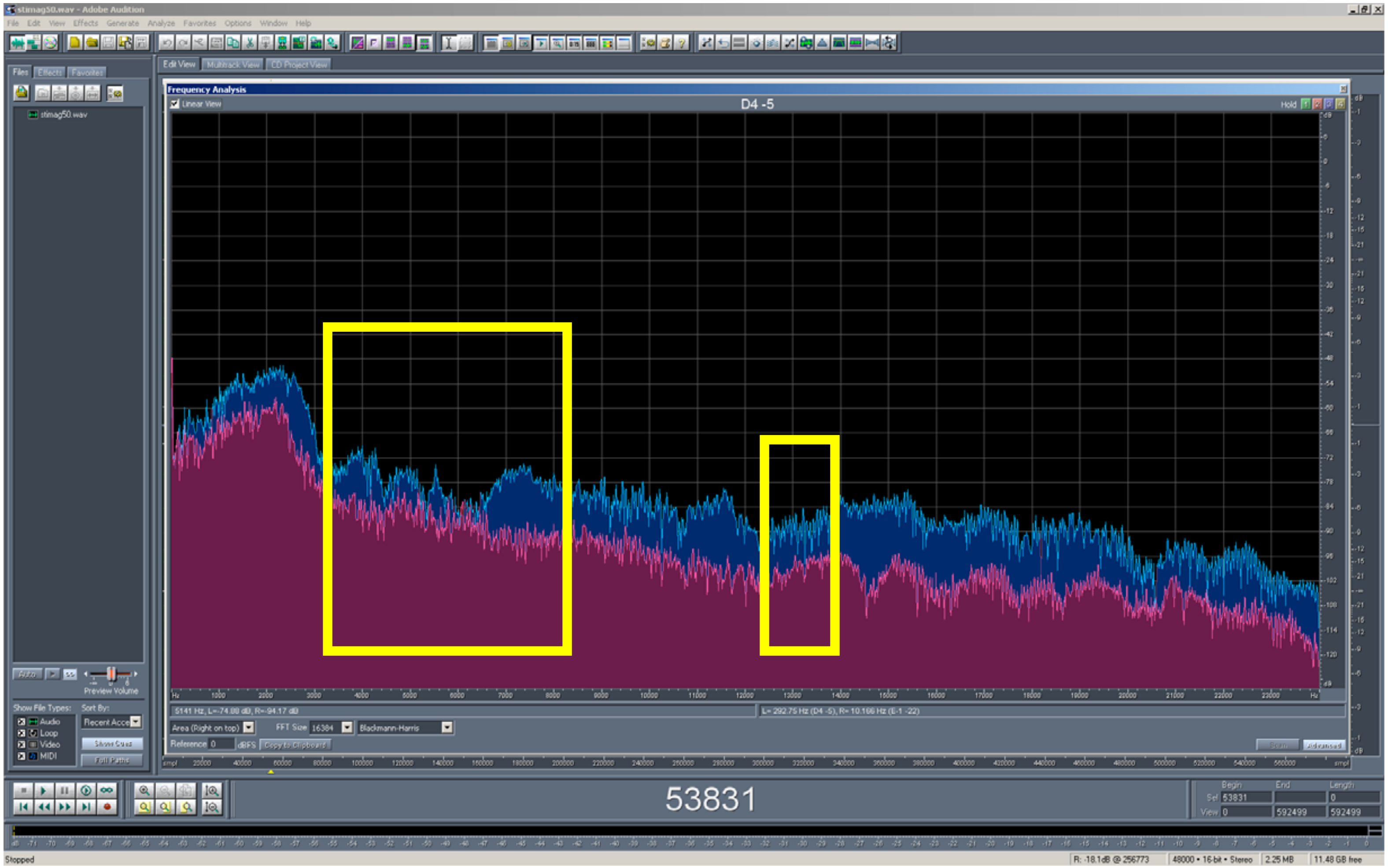


How do we hear stereo?

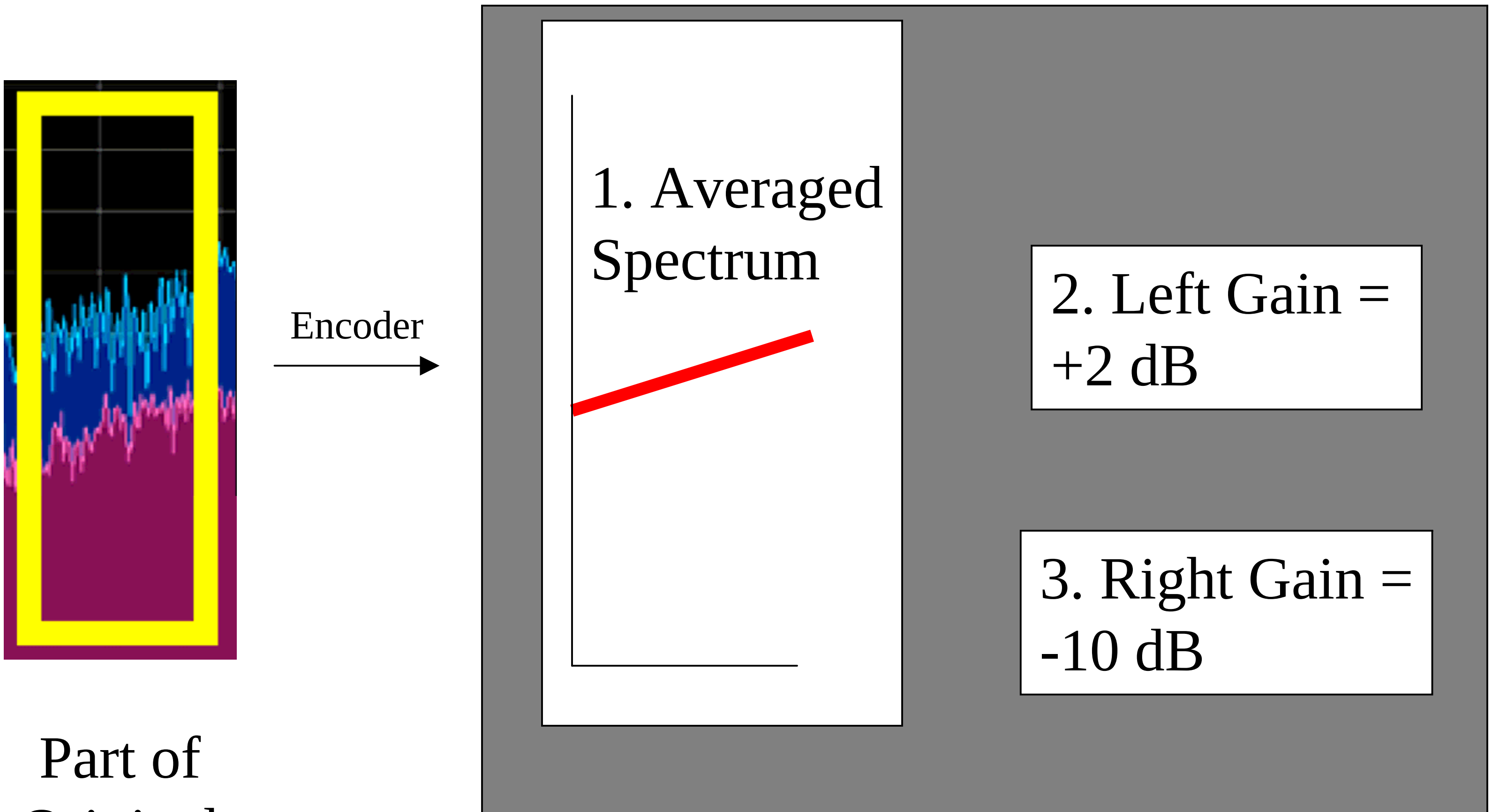
-

How many channels to be handled?

Intensity Stereo: where applied?



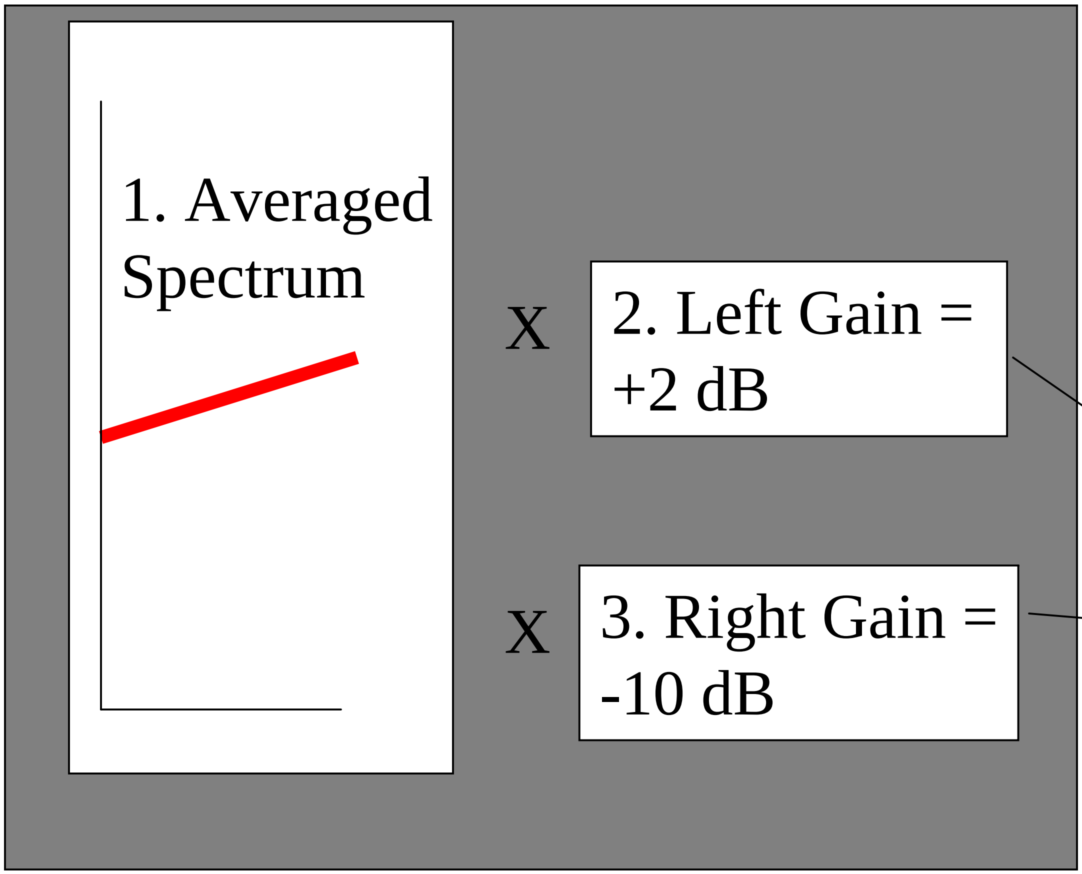
Intensity Stereo: how applied?



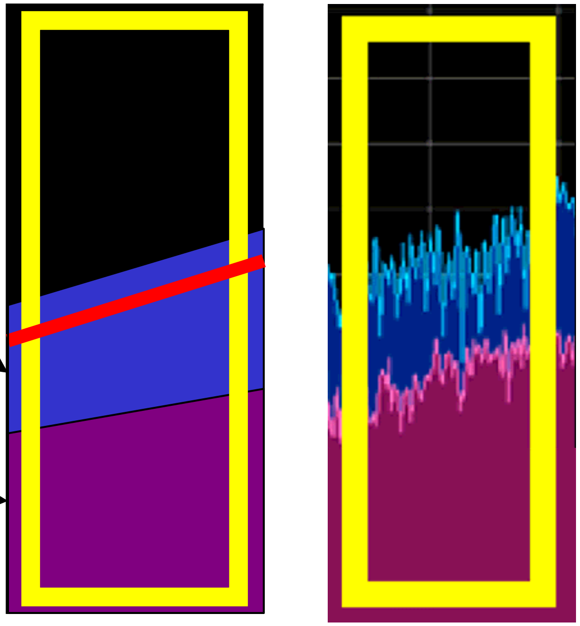
Part of Original Stereo Spectrum

Intensity Stereo Representation in “mp3” file

Intensity Stereo: how recovered?



Intensity Stereo Representation in “mp3” file



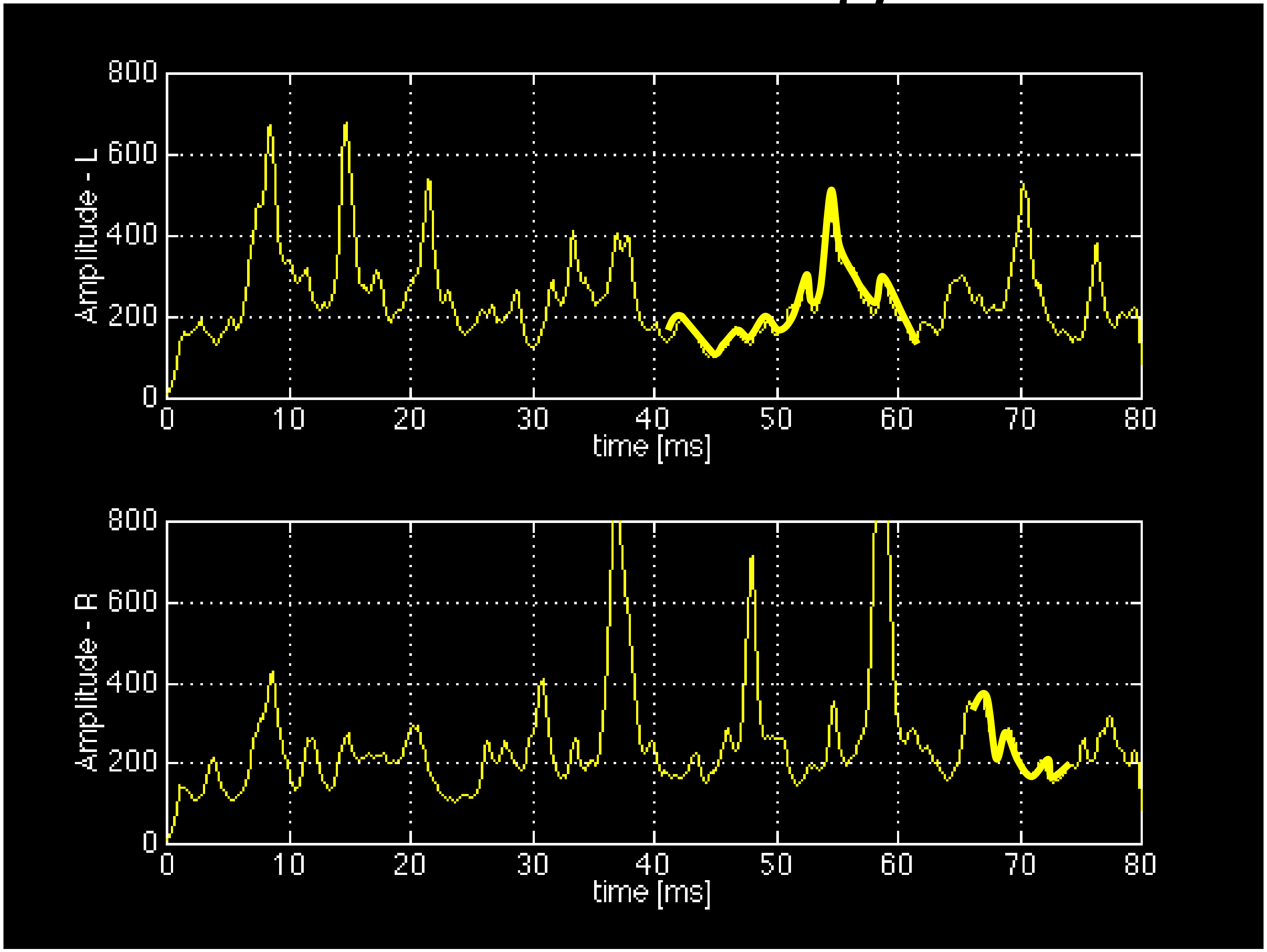
Recon-structed spectrum

(Orig. spectrum)

Intensity Stereo

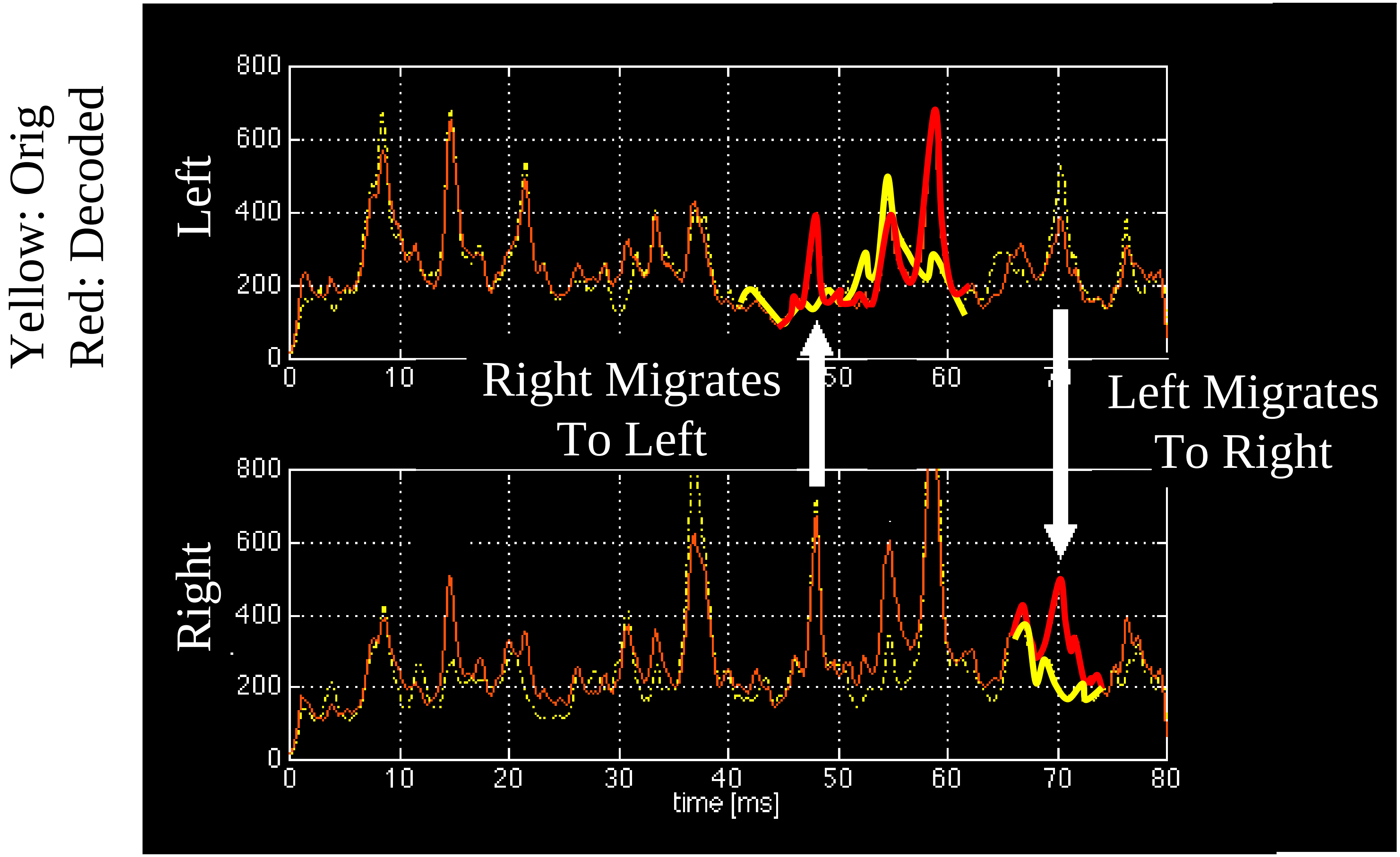
- Spectrum of L and R more or less same?
- In which frequency region?
- If a "same" region found:
  - Make one "template" [AC-3, AAC: "coupling channel"]
  - Transmit template spectrum in one channel.
  - Don't transmit (that part of) spectrum for other channel(s).
  - Transmit scaling factor (gain) for each channel.

I/S Effects: Original



After AES CD-ROM

I/S Decoded Waveforms Migrate



After AES CD-ROM

What can go wrong?
(Intensity) Stereo: Sound Examples

- Original
- Intensity stereo, not done well
- Difference signal

Source: AES CD-ROM. Original recording by Jim Johnston.

Mid/Side: Encoding

- Mid/Side (Sum/Difference)
- Equations for Encoder:

  $M = L + R$

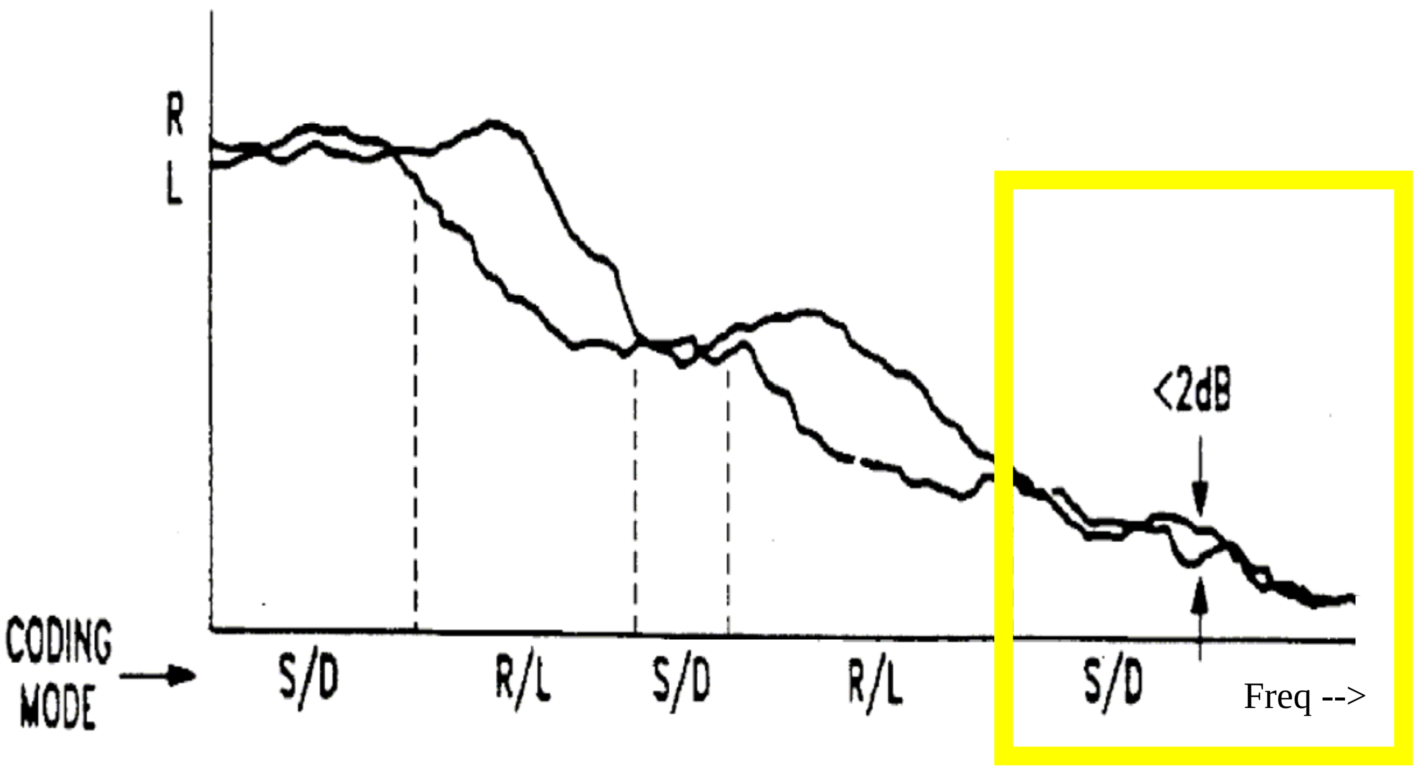
  $S = L - R$

Equations for Decoder:

  Sum M and S: $[(L + R) \mathbf{+} (L - R)]/2 = L$

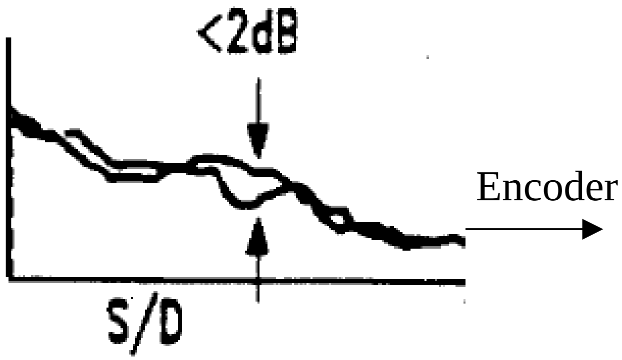
  Diff M and S: $[(L + R) \mathbf{-} (L - R)]/2 = R$

M/S: where to apply

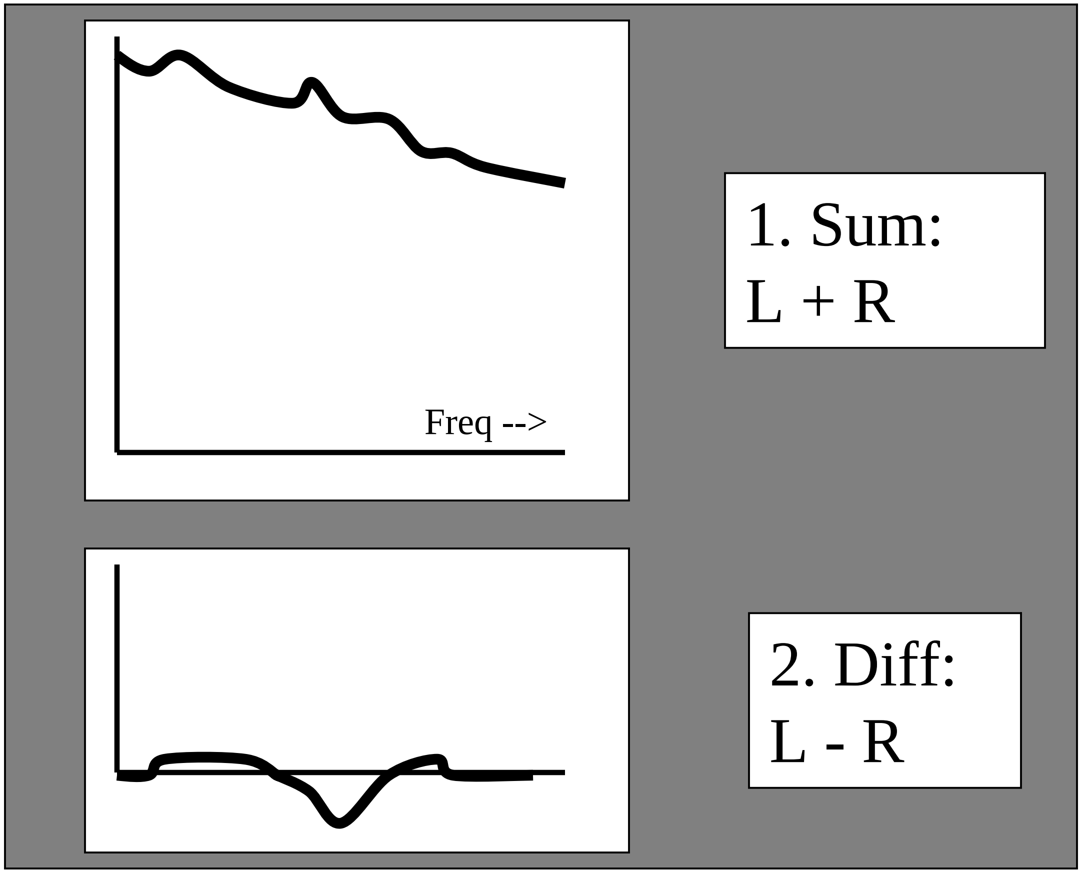


U.S. Patent 5,627,938, James Johnston

M/S Stereo: how applied?

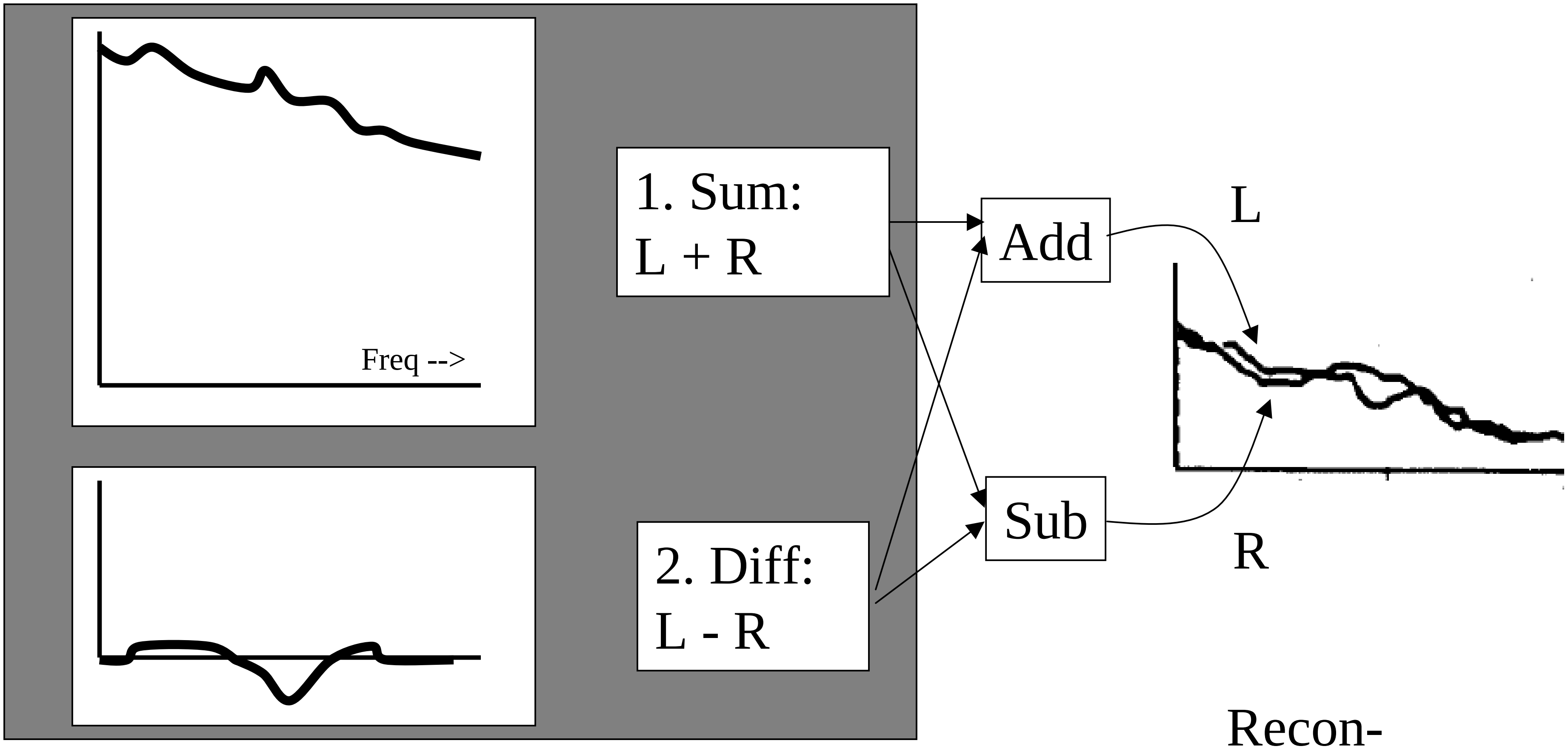


Part of Original Stereo Spectrum



M/S Representation in encoded file

M/S Stereo: how recovered?



M/S Representation in encoded file

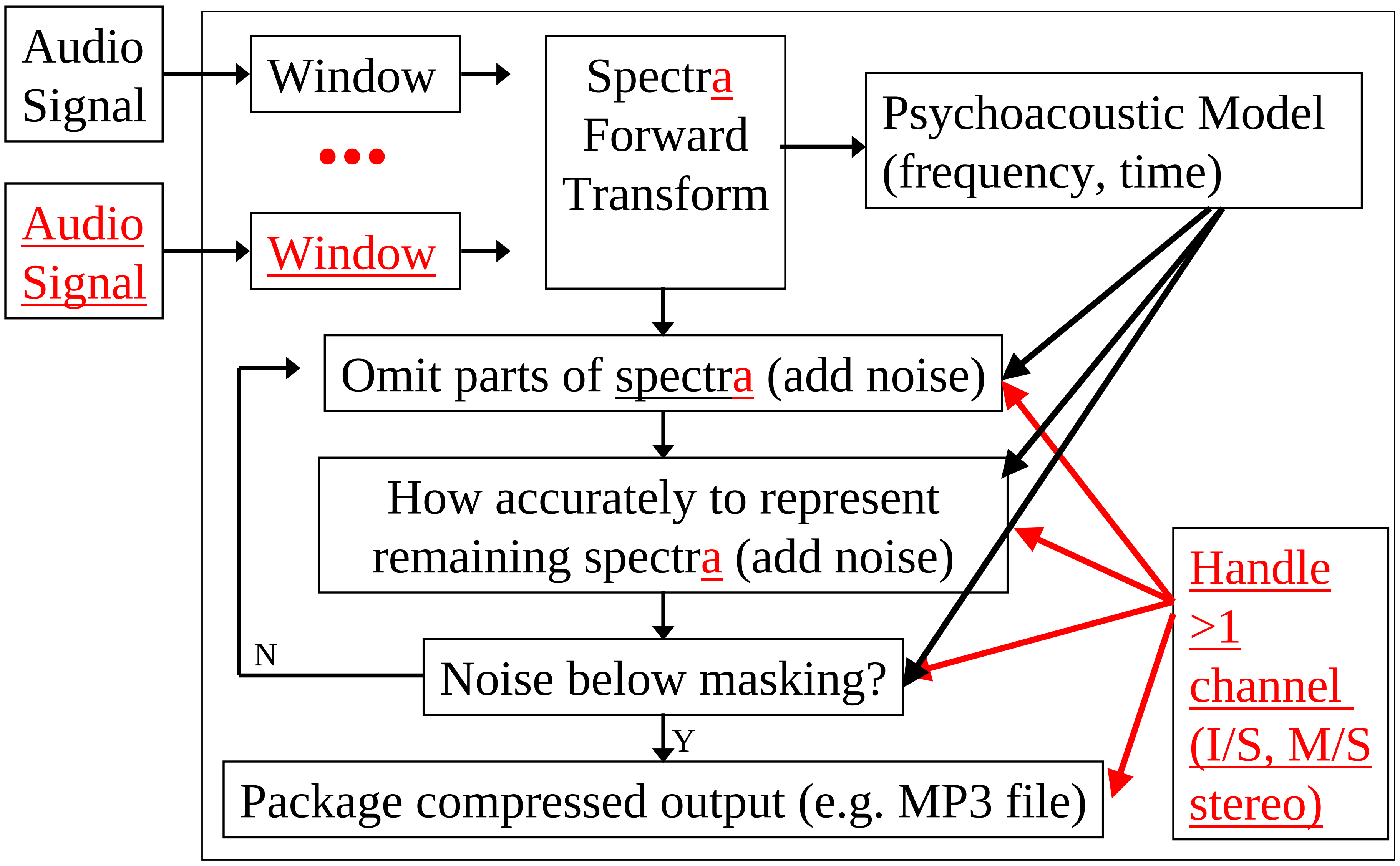
Recon-structed spectrum

Juergen Herre, Suzanne Vega, M/S

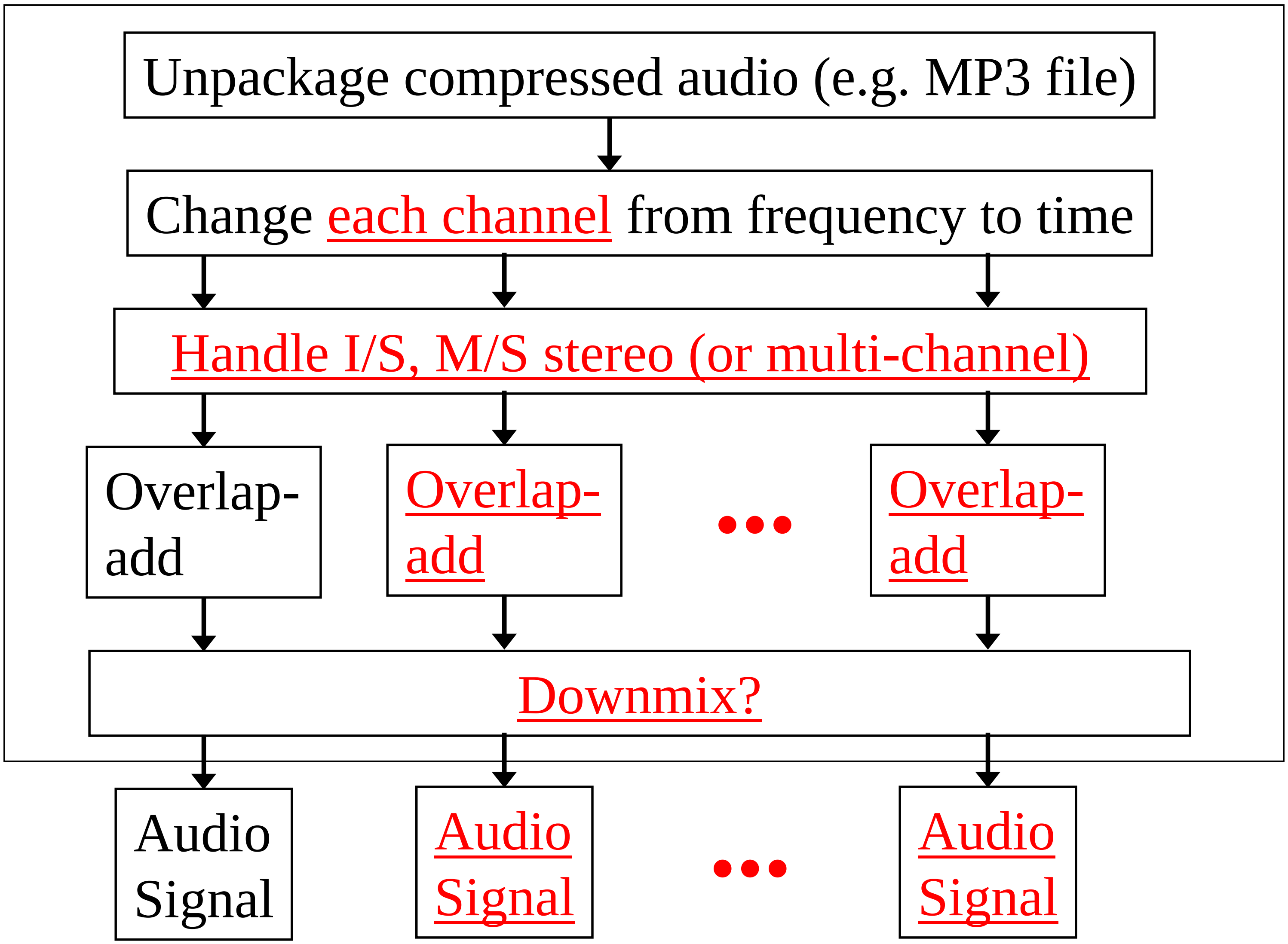
How Many Channels Encoded?

MPEG-2 Layer 3	Up to 5.1 + 2nd stereo program + 7 other multilingual channels
MPEG-2 AAC	1 thru 48
AC-3	1 thru 5.1
MLP	1 thru 63
DTS	1 thru 8 + LFE
PAC	1-16 front channel pairs, 7 surround, 7 auxiliary, 3 effects = 66 channels

Perceptual Encoder



Decoder (“mp3 player”)



Meeting the challenge

- Coarser Quantization (time domain)
- DPCM, ADPCM
- Linear Prediction
- Subband coding
- Transform to frequency domain
- Coarser quantization (frequency domain)
- Psychoacoustics: mask the noise
- Variable bit rate
- Noiseless coding
- Window
- Temporal masking
- Error recovery
- Multichannel redundancy

What's really in an "MP3" file --- additional stuff?

What we have covered

- Stereo perception
- Intensity Stereo coding
- Mid/Side coding

Next Listening Session

- Meet by _____ in listening areas.
- Listen to tracks Tancod 55, 57, 59, 65 in Session: Tancod.
- Take notes and discuss:
  - What do you hear?
  - How are they different?
  - What do you like?
  - What don’t you like?
- Back here at 10:40 to discuss. (15 min)

Notes on Sound Examples

55

57

59

65

MSRA 5500 - 002
MUS 4500 - 002

March 15, Session 3, Part 2

Codec Abuse

© Copyright 2008 John Strawn

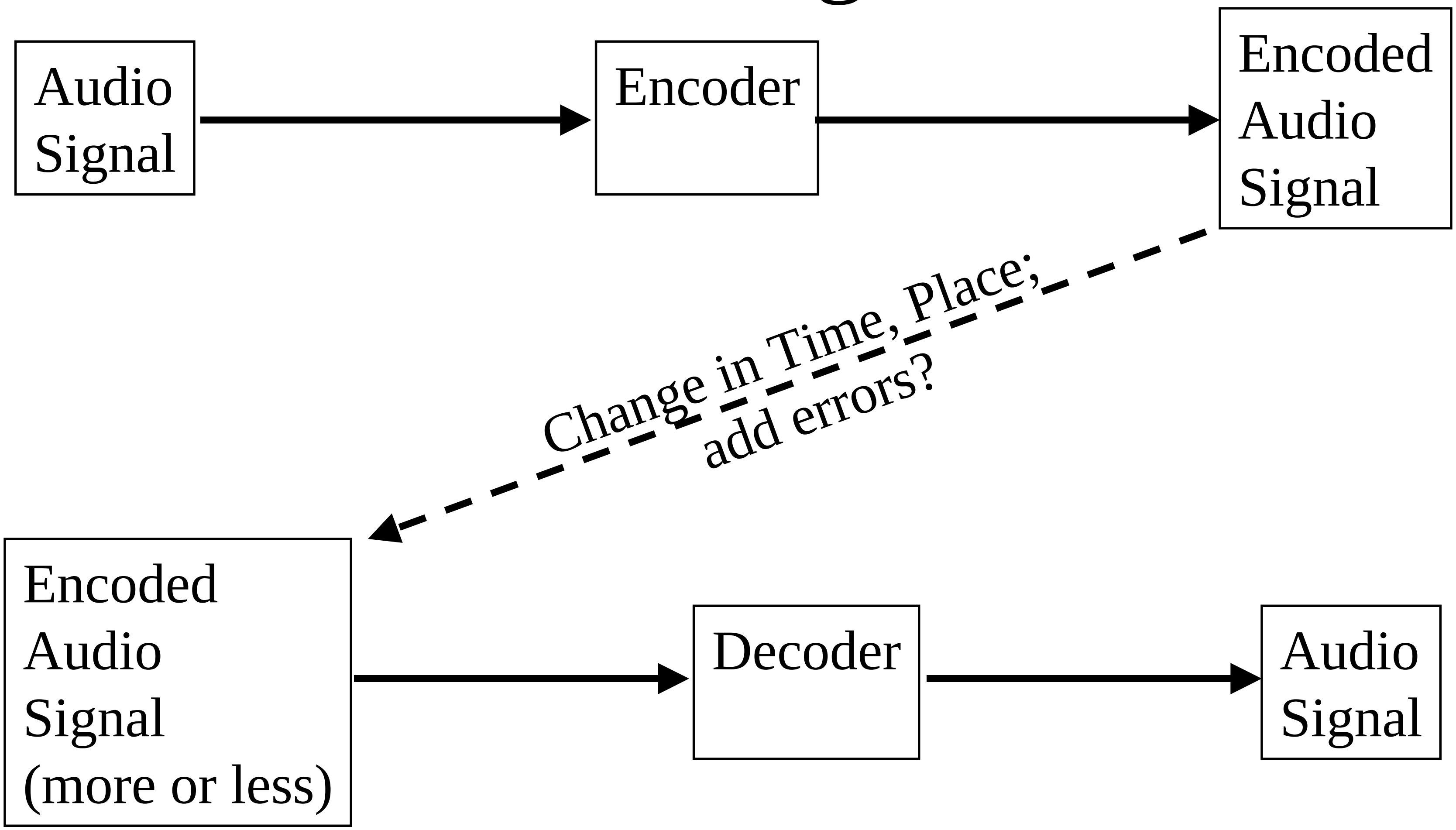
Discuss listening examples

55 57 59 65

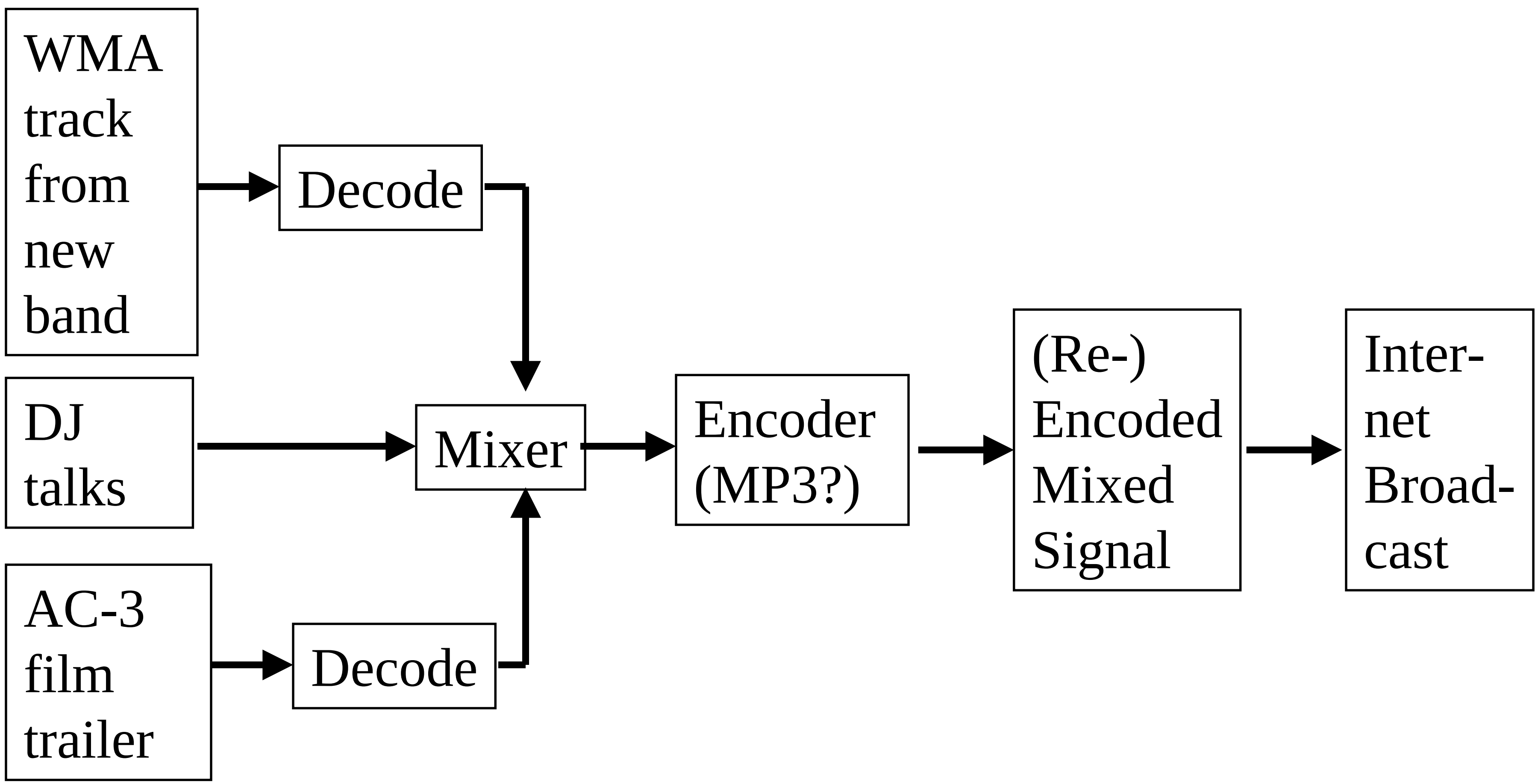
What we will cover

- How in the audio food chain can codecs be (accidentally) misused?
- Problems in real-time transmission.
- Security, Watermark.

Review: Coding in General



Possible Broadcast Scenario



Tandem Coding

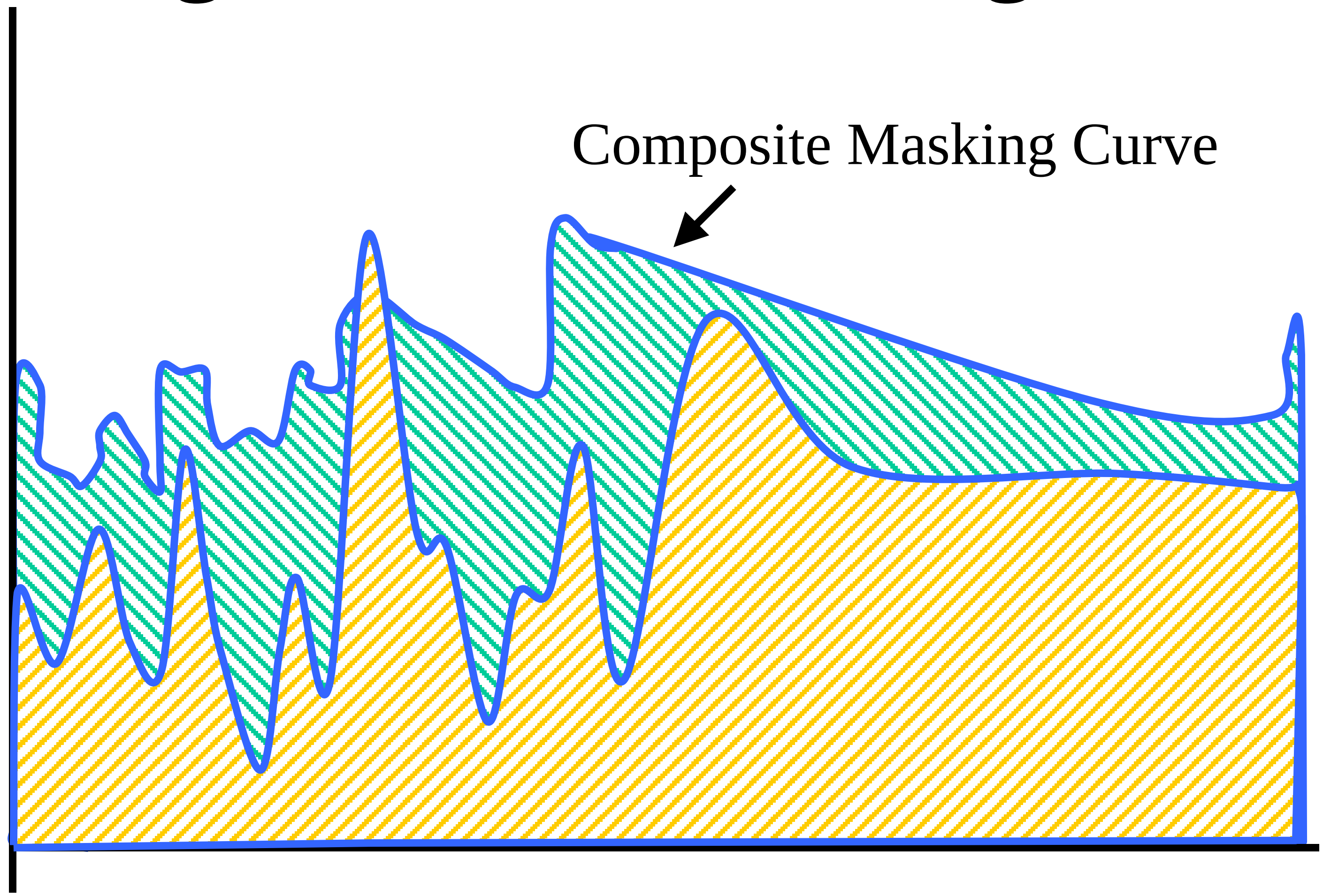
- Encode, decode more than one time in succession
- aka TransCoding

What you heard: Tandem Coding

- Original
- After 2 generations
- After 4 generations
- After 10 generations

Source: AES CD-ROM.

Masking: Where noise might fall



After Davidson et al., 1994

Tandem Coding: Solutions

•

Tandem Coding: Solutions

- Get unencoded original whenever possible.
- Start with high bit rate.
- Special coders/encoders designed to allow
- Embed information in encoded data stream
- Lossless codecs
  - Meridian Lossless Packing, MLP
  - ... (later today)

How Much Headroom?

- Assumptions: 1) All codecs are identical & operate at the same bit-rate 2) Encoder frame alignment is random from codec to codec
- SNR degradation due to N codecs in tandem: 10 log(N) dB

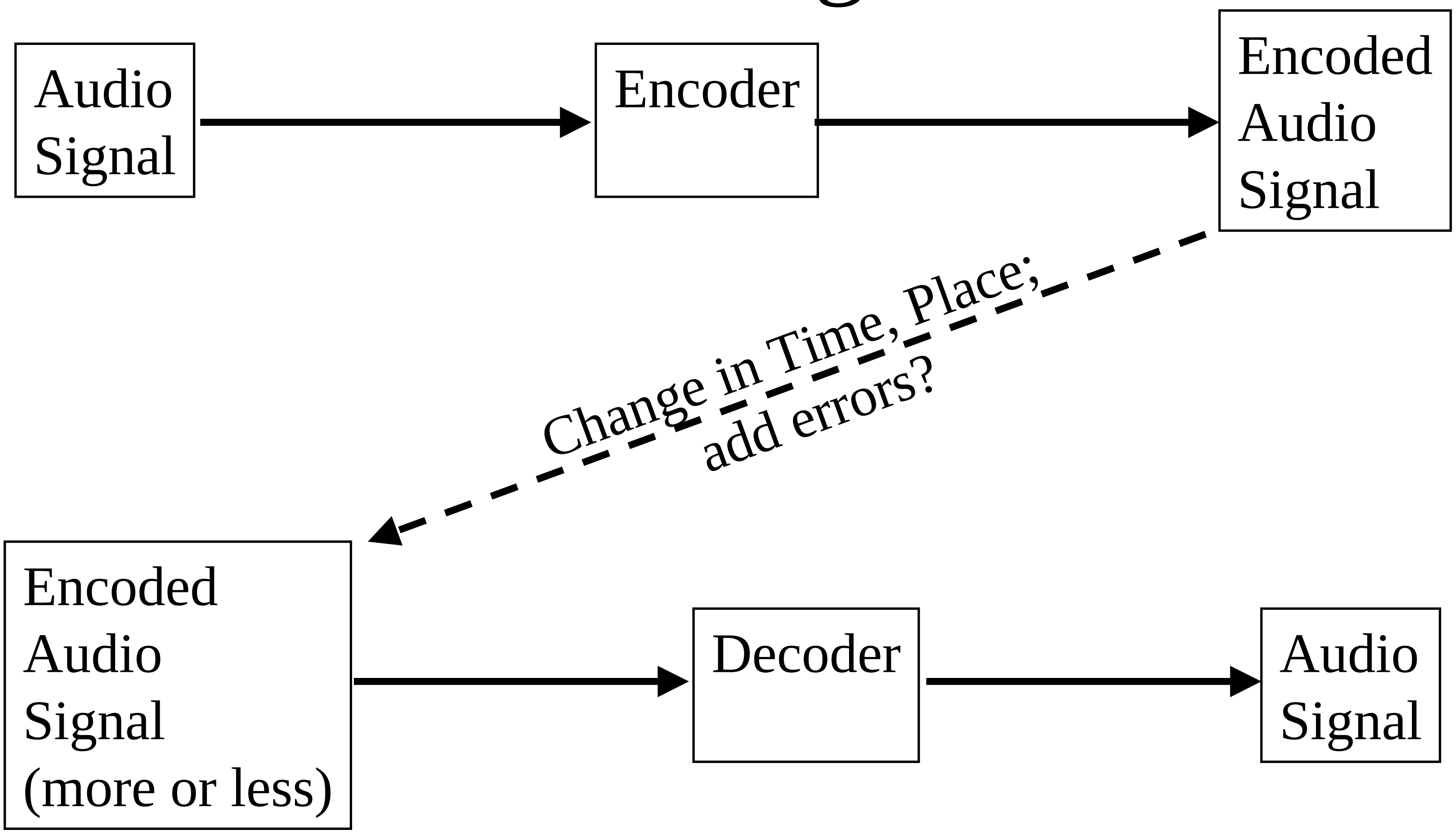
Headroom Required for N Stereo Codecs in Tandem		
N	Headroom (kb/s)	Bitrate for audio quality equivalent to one-pass 128 kb/s
2	40	168
4	80	208
10	133	261

Source: Grant Davidson, Dolby

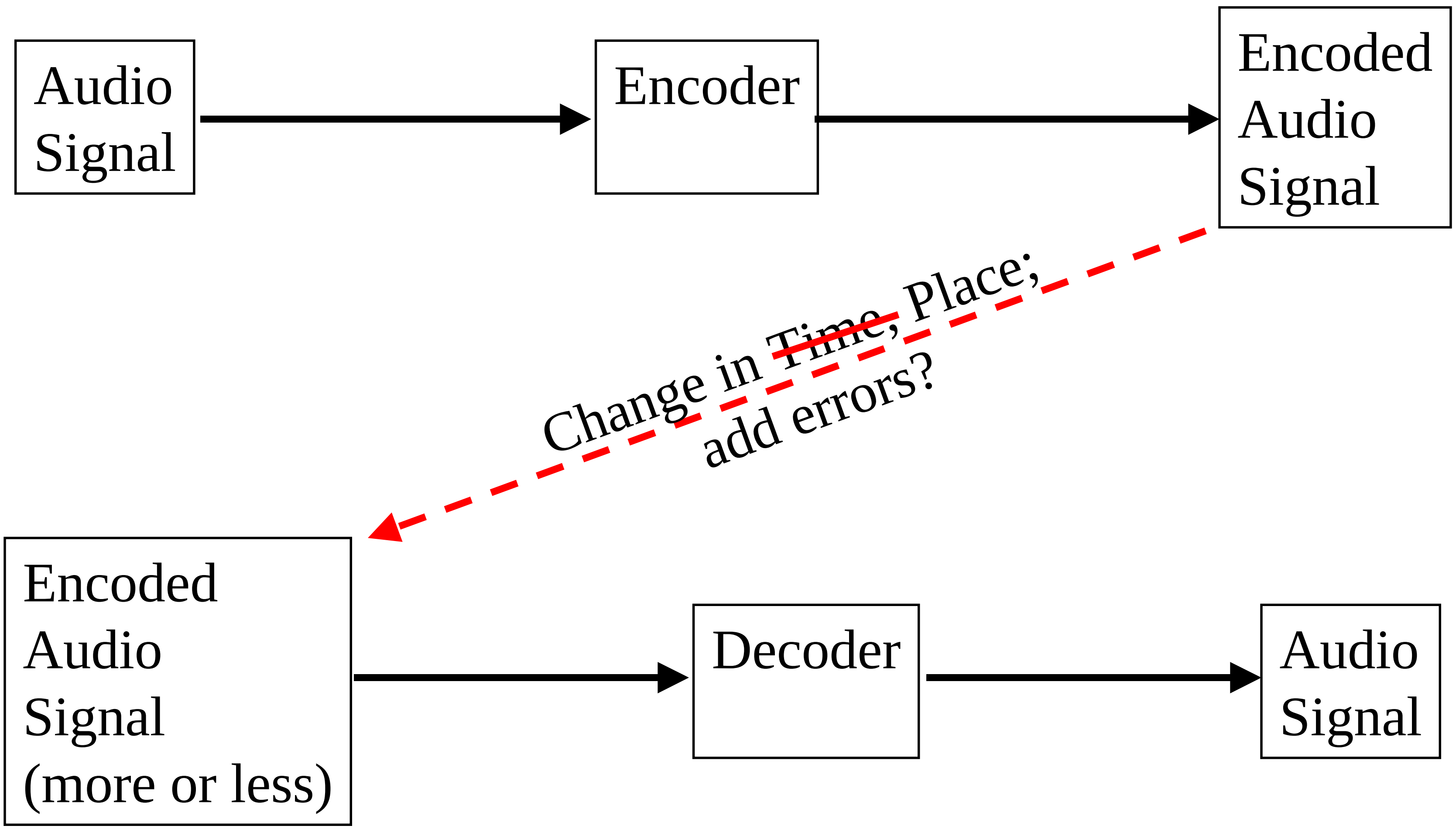
What we will cover

- How in the audio food chain can codecs be (accidentally) misused?
- Problems in real-time transmission.
- Security, Watermark.

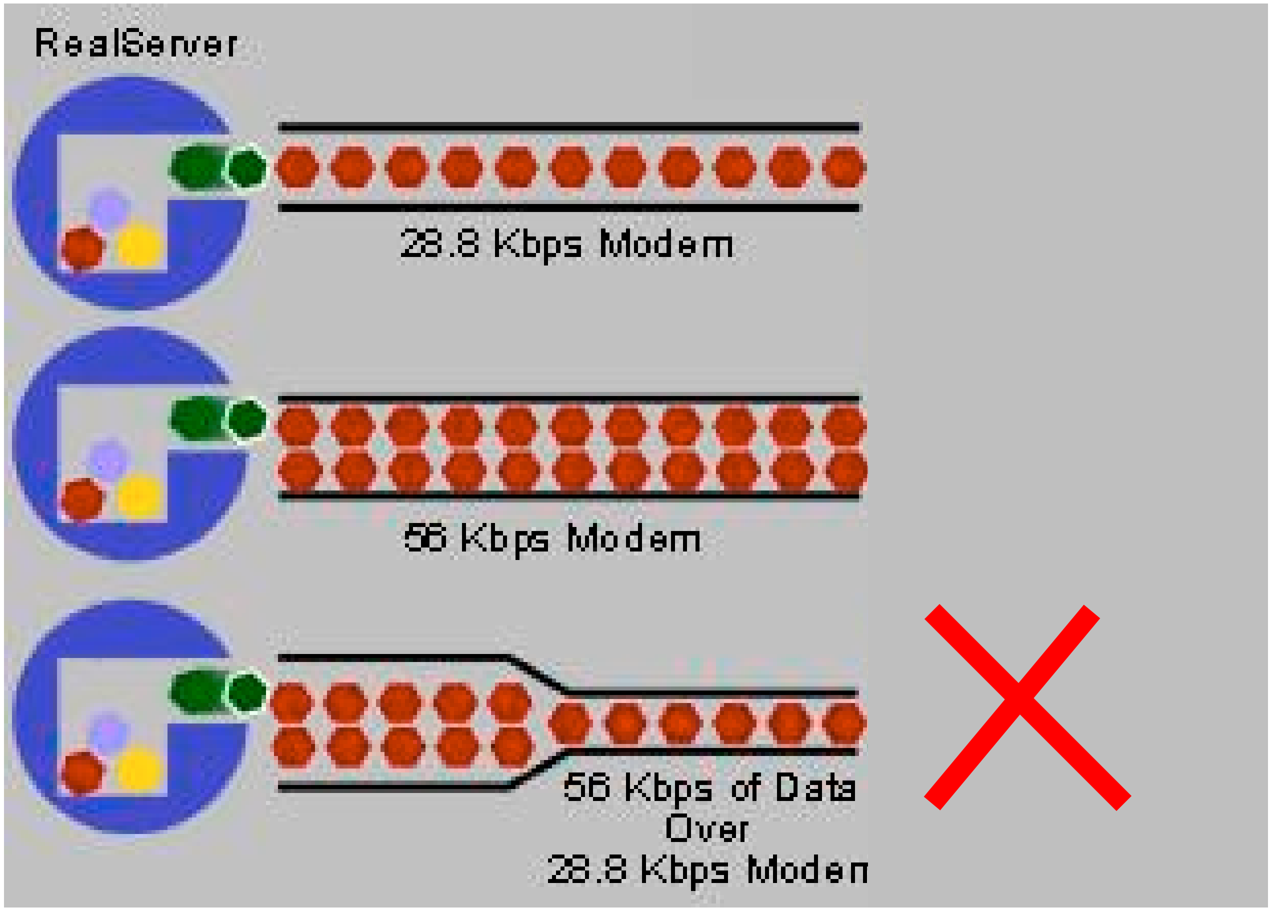
Review: Coding in General



Real-time Transmission

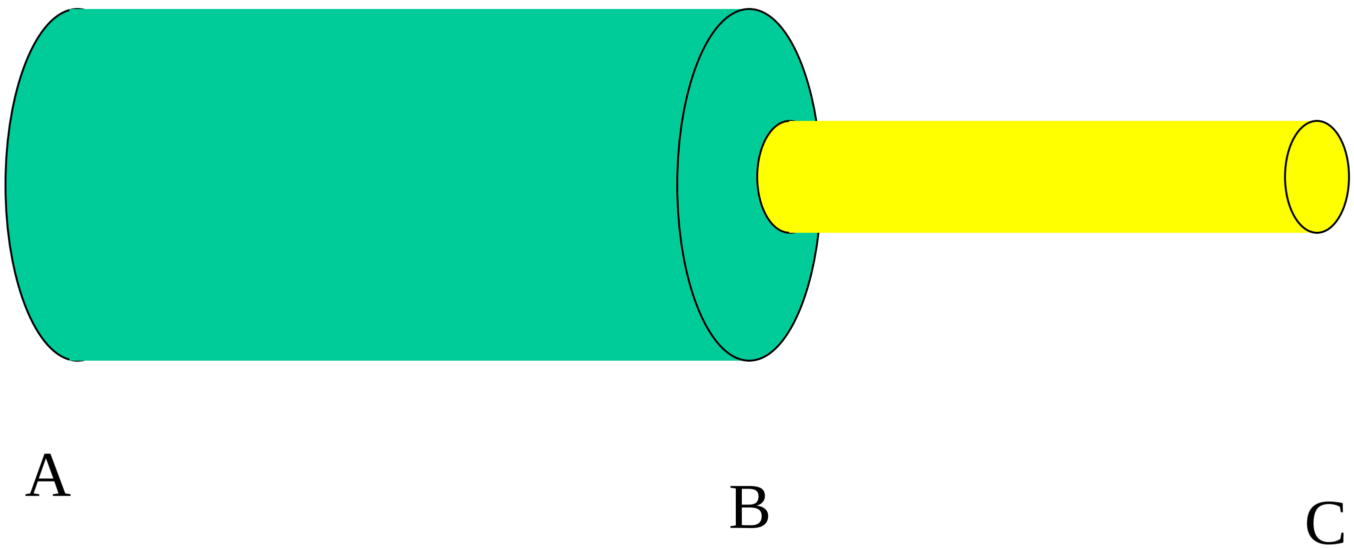


Match Xmit Rate with Bandwidth

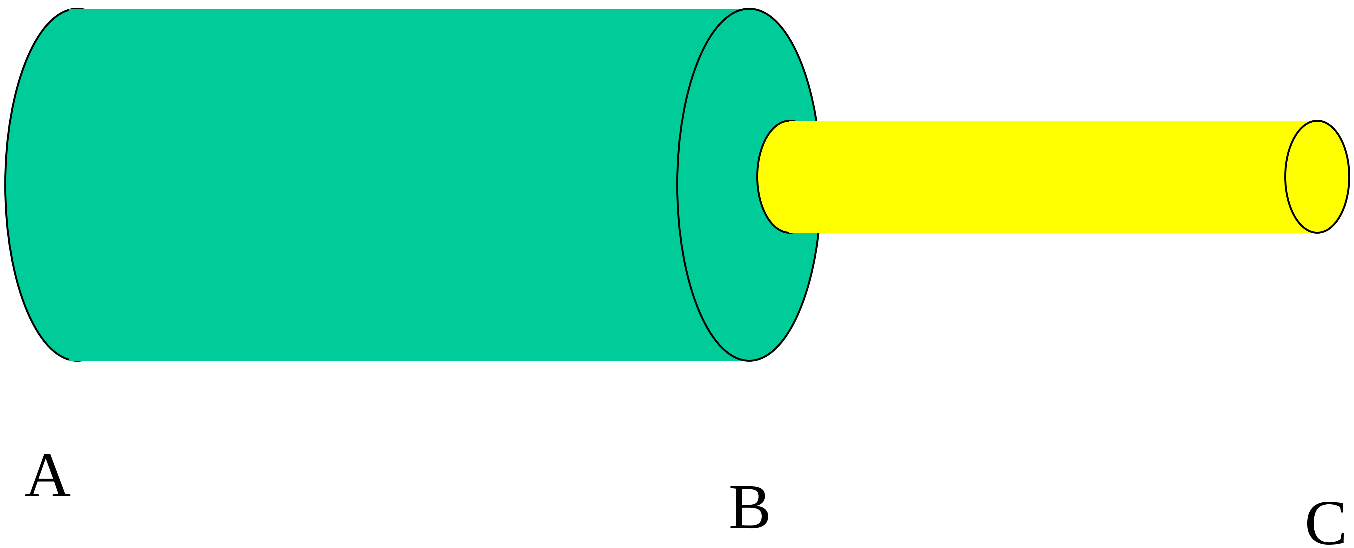


After http://service.real.com/help/library/guides/production8/htmfiles/realsys.htm, retrieved 4 Feb 2008

Real-time Transmission

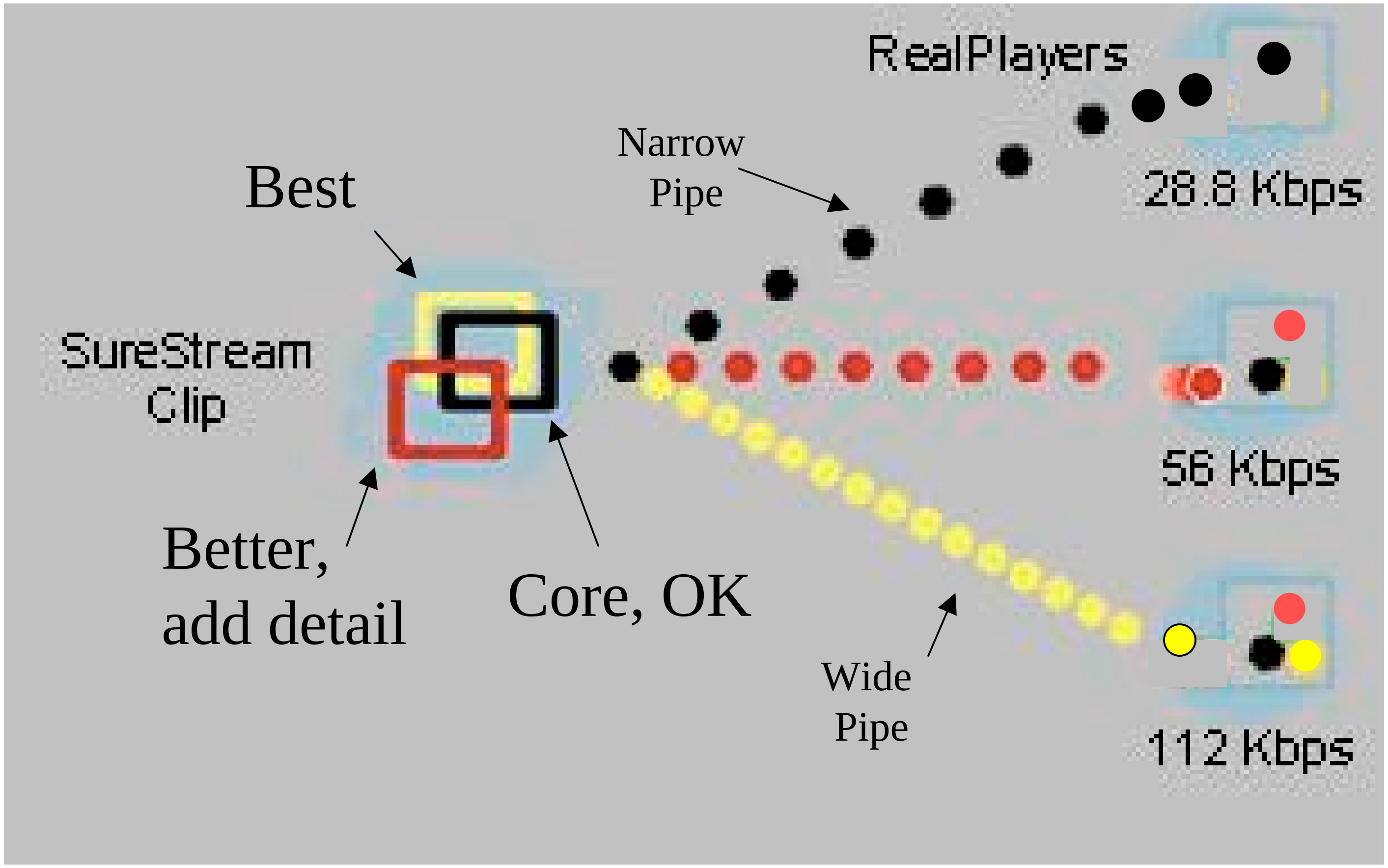


Real-time Transmission



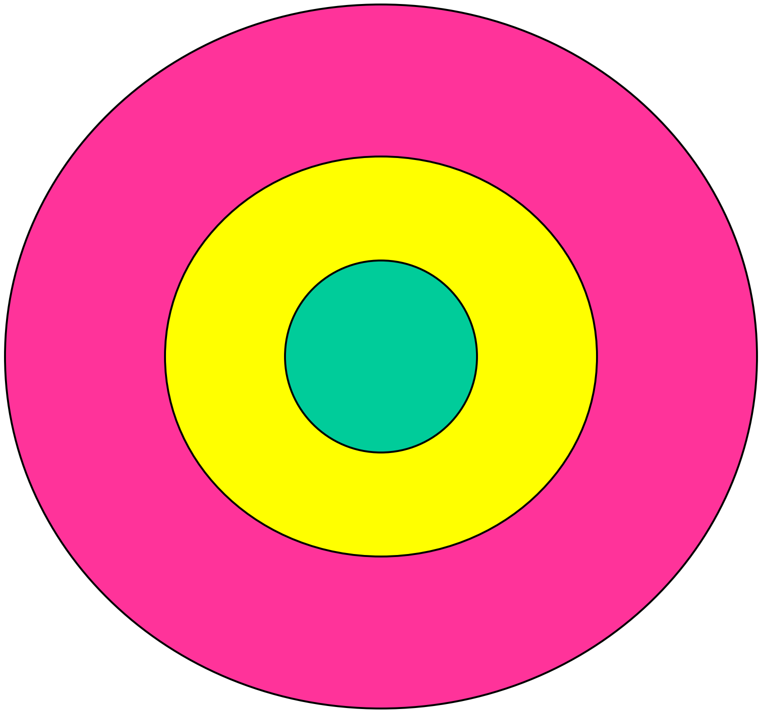
SCALABLE

Example: Real SureStream

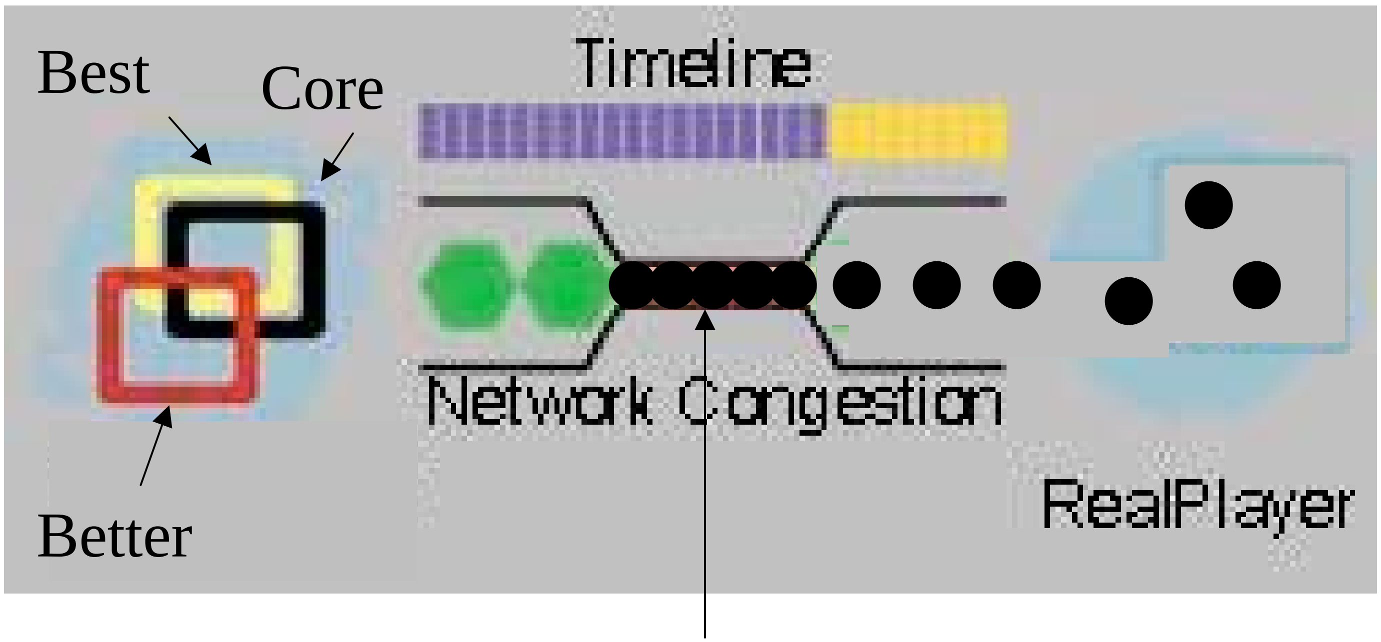


After http://service.real.com/help/library/guides/production8/htmfiles/realsys.htm

Scalable



Example: Real SureStream



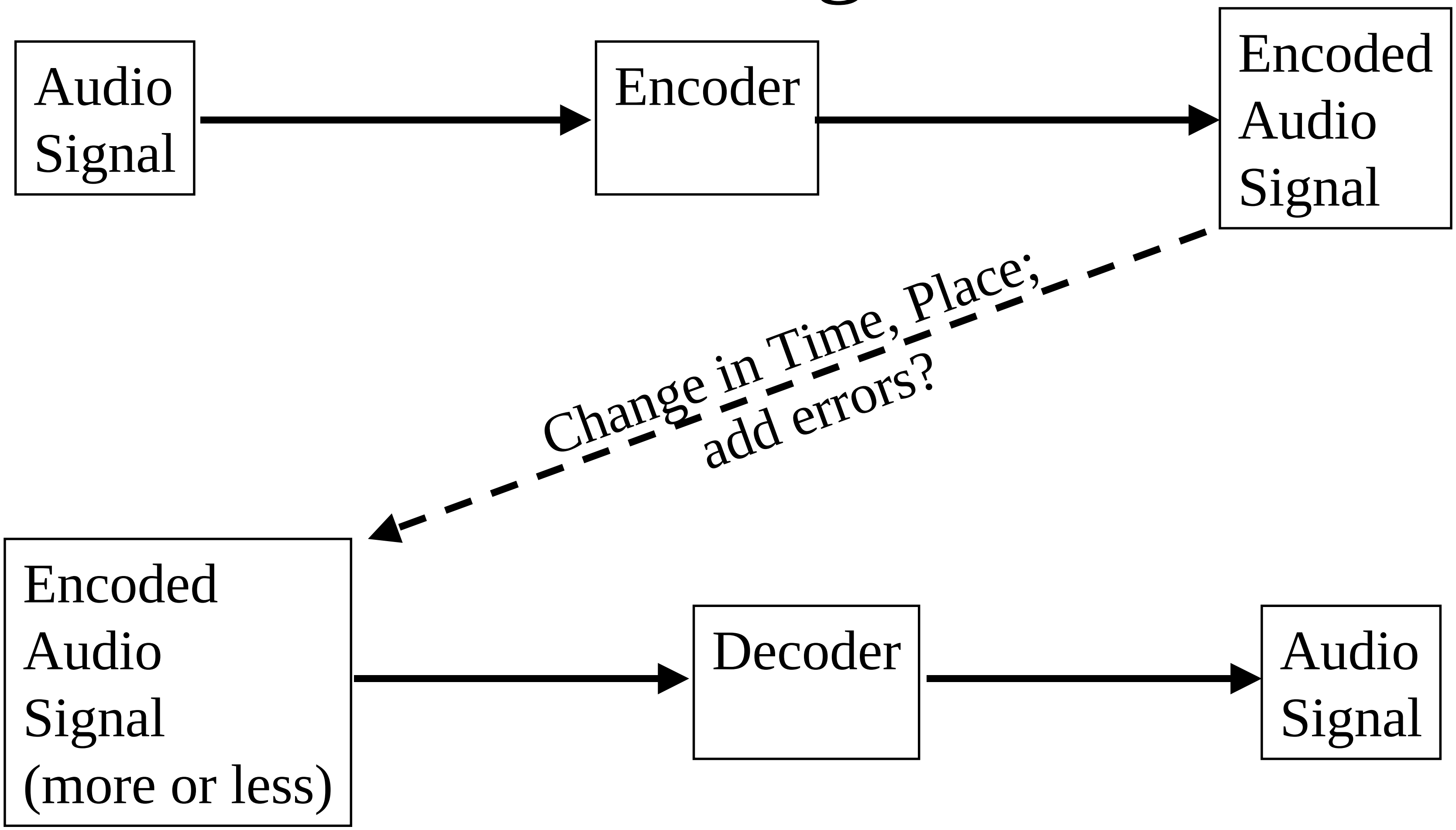
At least the core gets through

After http://service.real.com/help/library/guides/production8/htmfiles/realsys.htm

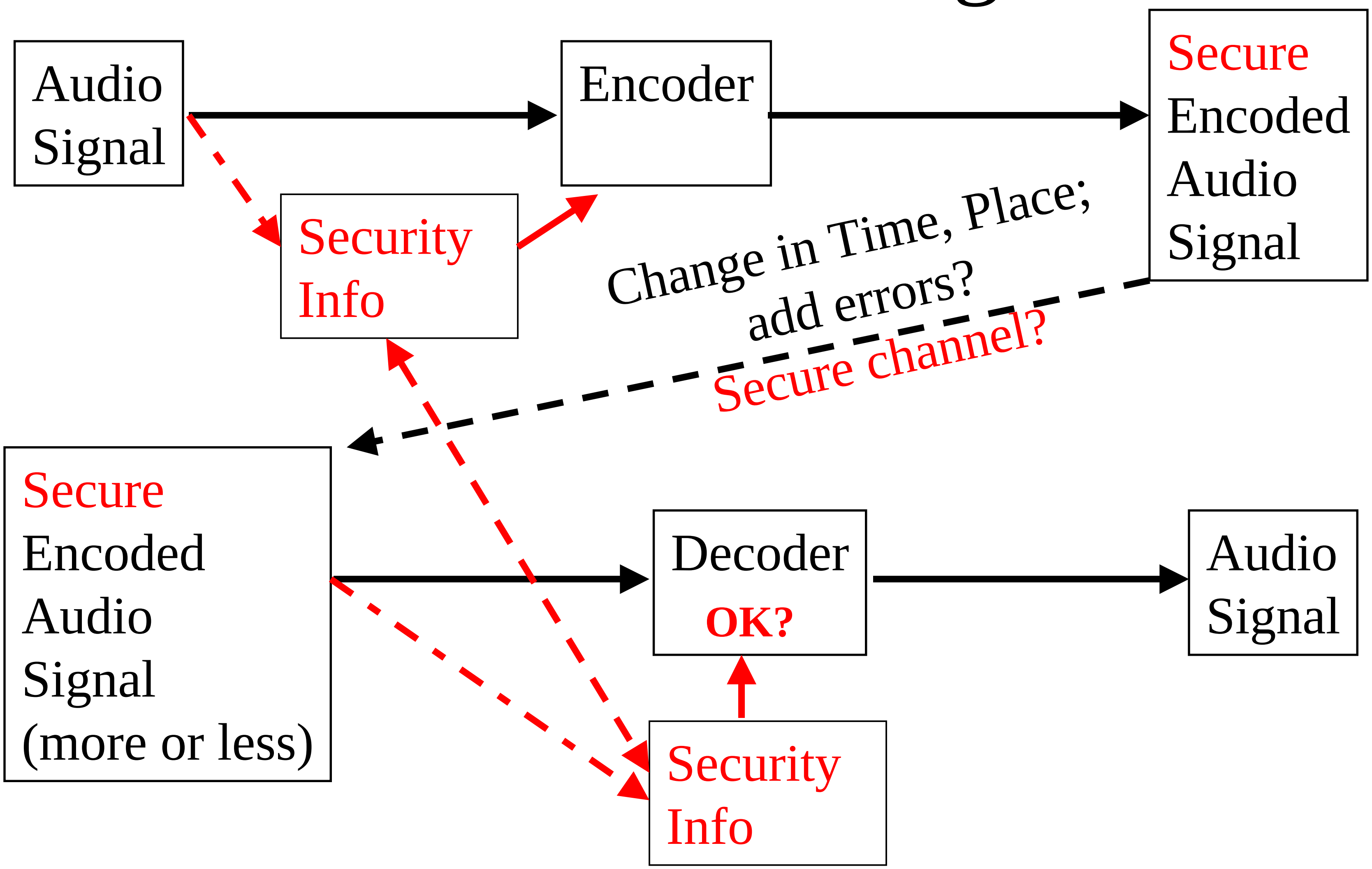
What we will cover

- How in the audio food chain can codecs be (accidentally) misused?
- Problems in real-time transmission.
- <u>Security, Watermark.</u>

Review: Coding in General



Secure Coding



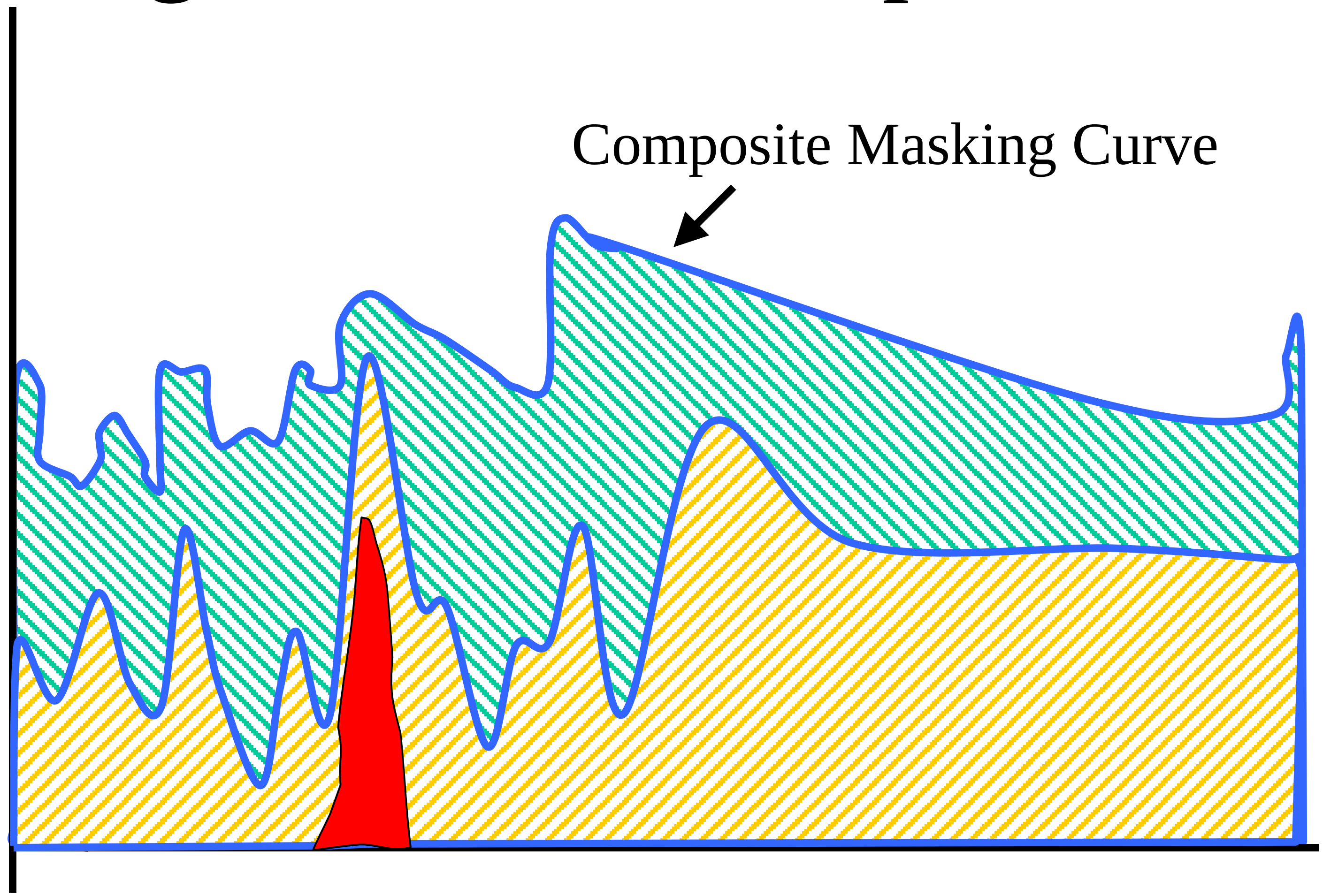
Desireable Watermark

- Inaudible; High data rate
- Difficult to detect/remove without authorization.
- Original corrupted if watermark forcibly removed.
- Minimal false positives/negatives.
- Robust to: filter; D/A-A/D; audio compression; radio transmission ...
- No/minimum increase in data size.

How to make a watermark

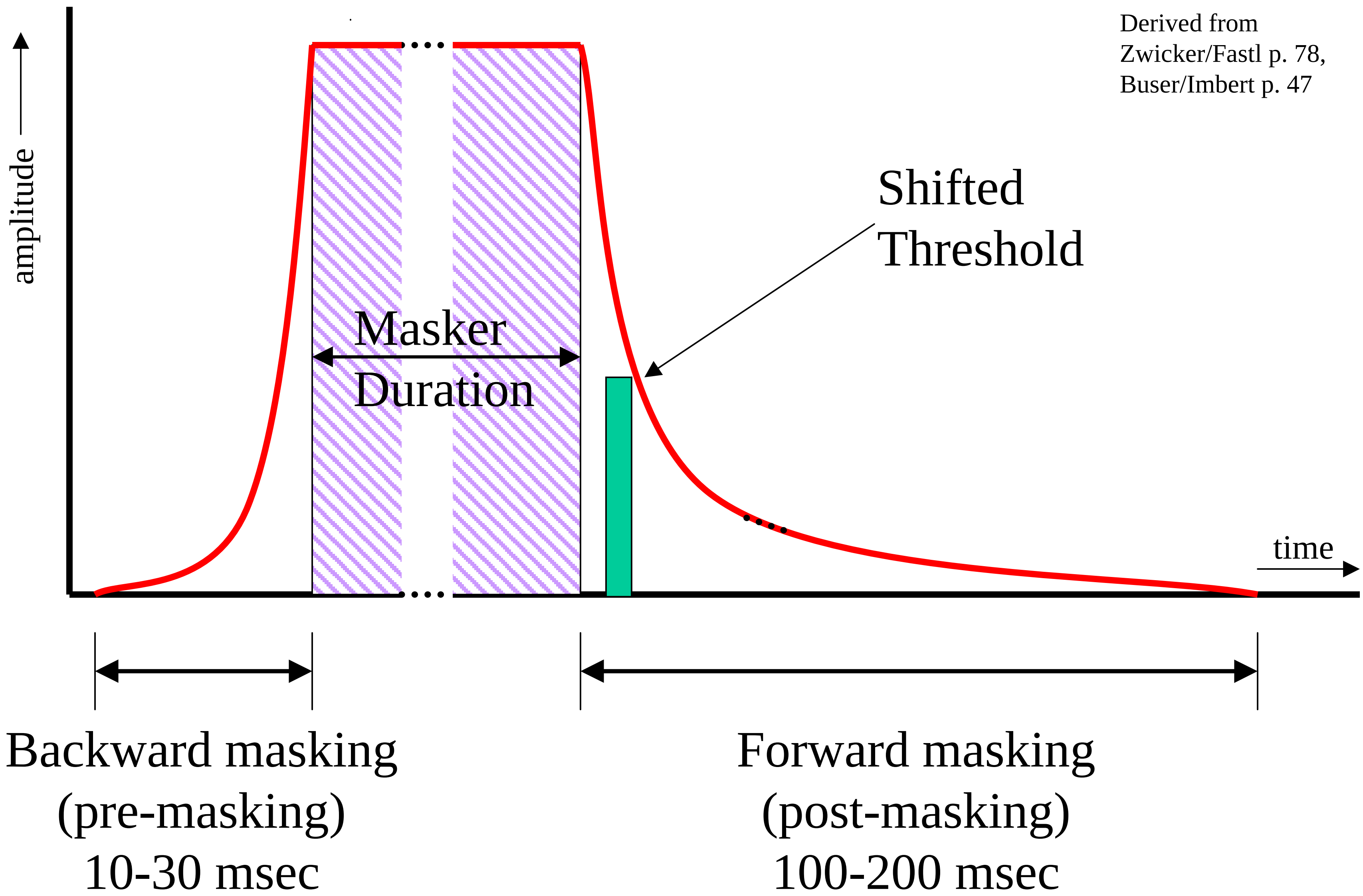
- Twiddle bits
- Add echoes
- Spread spectrum
- ...

Hiding watermark in spectrum



After Davidson et al., 1994

Hiding watermark in time



Some Protection Schemes

- ~~MidBar CDS (“Cactus Data Shield”)~~
- MPEG-4 IPMP
- InterTrust DRM
- Apple FairPlay DRM, iTunes (Jobs’ Letter)
- Windows Media

Some More Players

- SACD (mandatory watermark, track position)
- ~~Sony BMG, XCP from UK-based First4Internet~~
- Verance VCMS

Situation Today

- DRM not unified
- DRM benefits consumer?
- “Rights” unclear
- www.drmwatch.com
- Lawrence Lessig at Stanford
- ...

Watermarking Demonstration: 40-year-old Virgin DVD

What we will cover

- How in the audio food chain can codecs be (accidentally) misused?
- Problems in real-time transmission.
- Security, Watermark.
- [short break?]

MSRA 5500 - 002
MUS 4500 - 002

March 15, Session 4, Part 1

Which Codec to Choose

© Copyright 2008 John Strawn

What we will cover

- Some history of (perceptual) coding
- Main codec families, their names, key features
- Lossless coding
- Analysis of decoded waveform

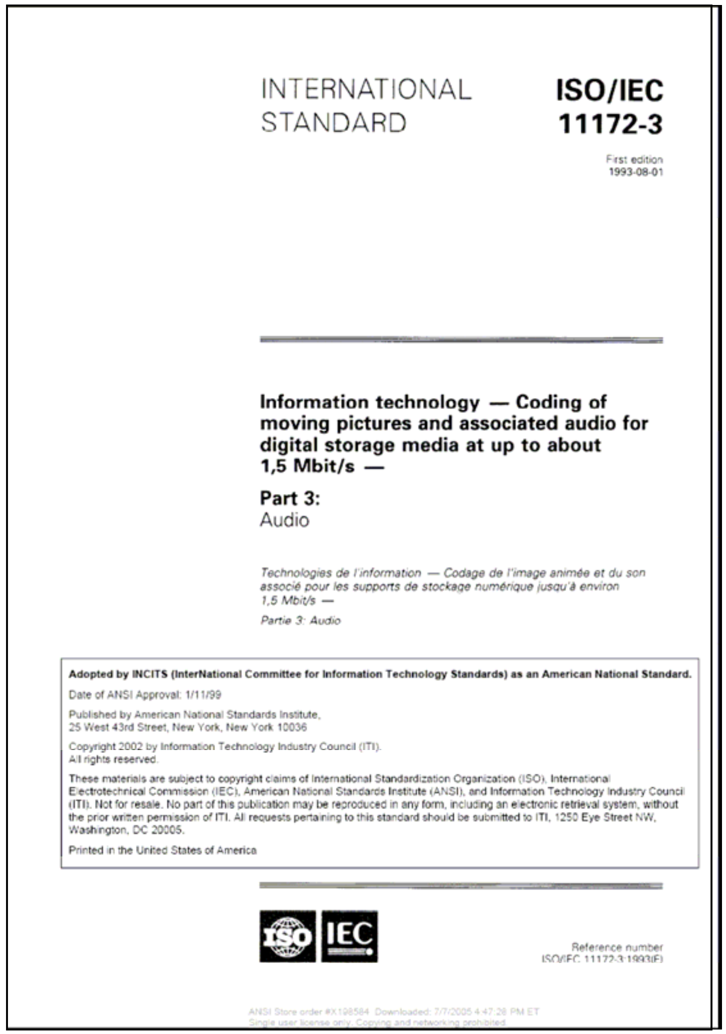
MPEG family overview

- MPEG-1
- MPEG-2
- MPEG-2 AAC
- (no MPEG-3)
- MPEG-4
- (MPEG-7)
- (MPEG-21)

MPEG-1 Standards Family

- Information Technology - Coding Of Moving Pictures And Associated Audio For Digital Storage Media At Up To About 1,5 Mbit/S
  - ISO/IEC 11172-1 Part 1: Systems
  - ISO/IEC 11172-2 Part 2: Video
  - ISO/IEC 11172-3 Part 3: Audio
  - ISO/IEC 11172-4 Part 4: Compliance Testing
  - ISO/IEC 11172-5 Part 5: Software Simulation

The MPEG-1 audio standard



INTERNATIONAL STANDARD

ISO/IEC 11172-3

First edition
1993-08-01

Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s —

Part 3:
Audio

Technologies de l'information — Codage de l'image animée et du son associé pour les supports de stockage numérique jusqu'à environ 1,5 Mbit/s —

Partie 3: Audio

Adopted by INCITS (InterNational Committee for Information Technology Standards) as an American National Standard.

Date of ANSI Approval: 1/11/99

Published by American National Standards Institute,
25 West 43rd Street, New York, New York 10036

Copyright 2002 by Information Technology Industry Council (ITI).
All rights reserved.

These materials are subject to copyright claims of International Standardization Organization (ISO), International Electrotechnical Commission (IEC), American National Standards Institute (ANSI), and Information Technology Industry Council (ITI). Not for resale. No part of this publication may be reproduced in any form, including an electronic retrieval system, without the prior written permission of ITI. All requests pertaining to this standard should be submitted to ITI, 1250 Eye Street NW, Washington, DC 20005.

Printed in the United States of America

Reference number
ISO/IEC 11172-3:1993(E)

ANSI Store order #X198584 Downloaded: 7/7/2005 4:47:28 PM ET
Single user license only. Copying and networking prohibited.

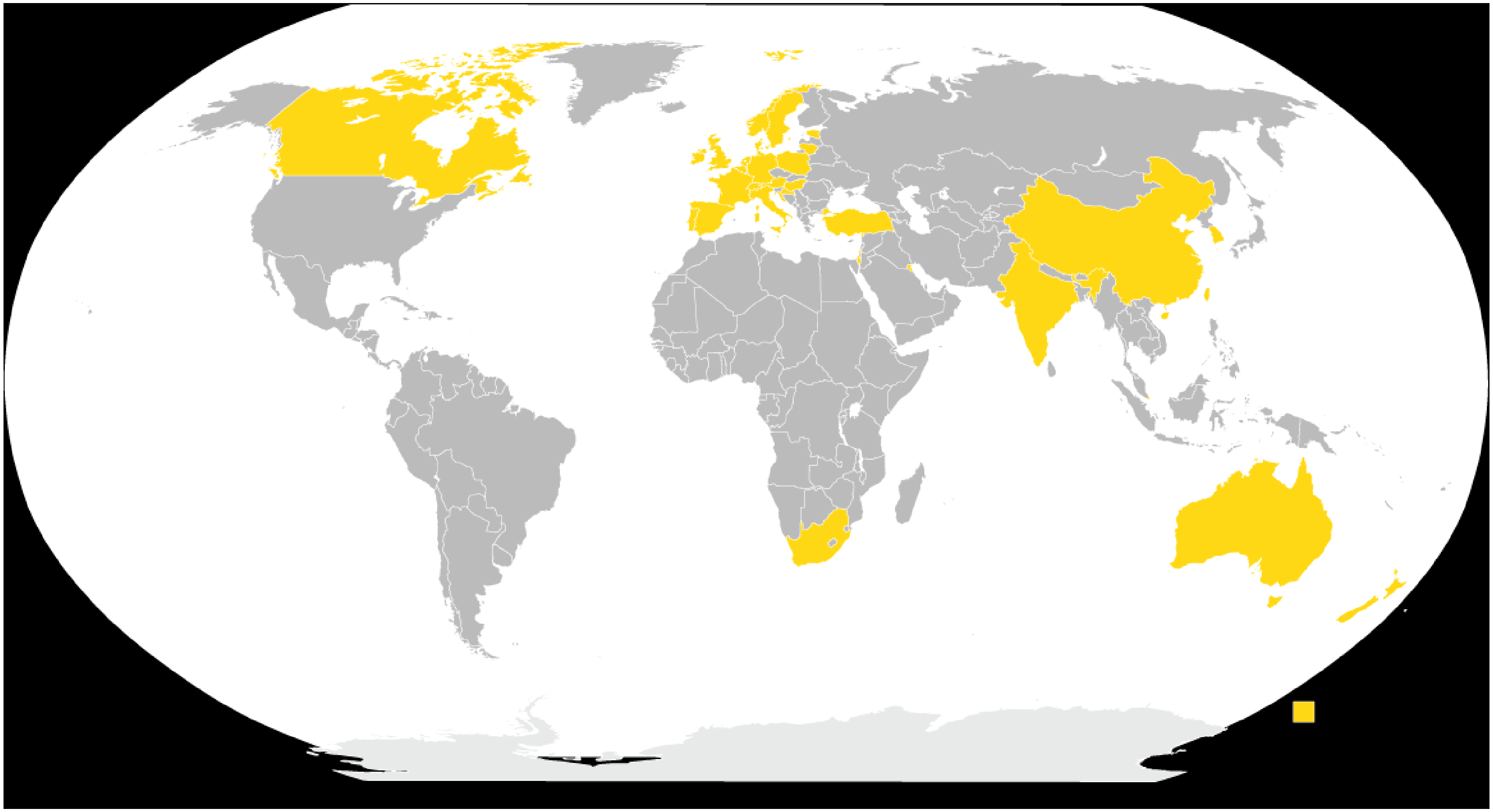
MPEG-1 Audio

- 1992: What gear available then?
- 32 kHz, 44.1 kHz, 48 kHz
- Only up to two channels:
  - Single channel
  - Two independent channels (why?)
  - Stereo
  - Stereo with joint coding

MPEG-1 Audio: “Layers”

- 32- 448 kbit/sec
- Specifies decoder, but not encoder (!)
- Layer 1: simplest; Philips DCC
- Layer 2: more efficient coding; DAB, CD-I
  - KUVO was: 256 kbps, stereo, 2 independent chans
- Layer 3: higher frequency and time resolution; ISDN, Internet; most complex decoder
- Bit stream format same
- Layer 3 must decode Layer 2 ...

DAB



http://en.wikipedia.org/wiki/Digital_Audio_Broadcasting

MPEG-1 Layer 1 and 2

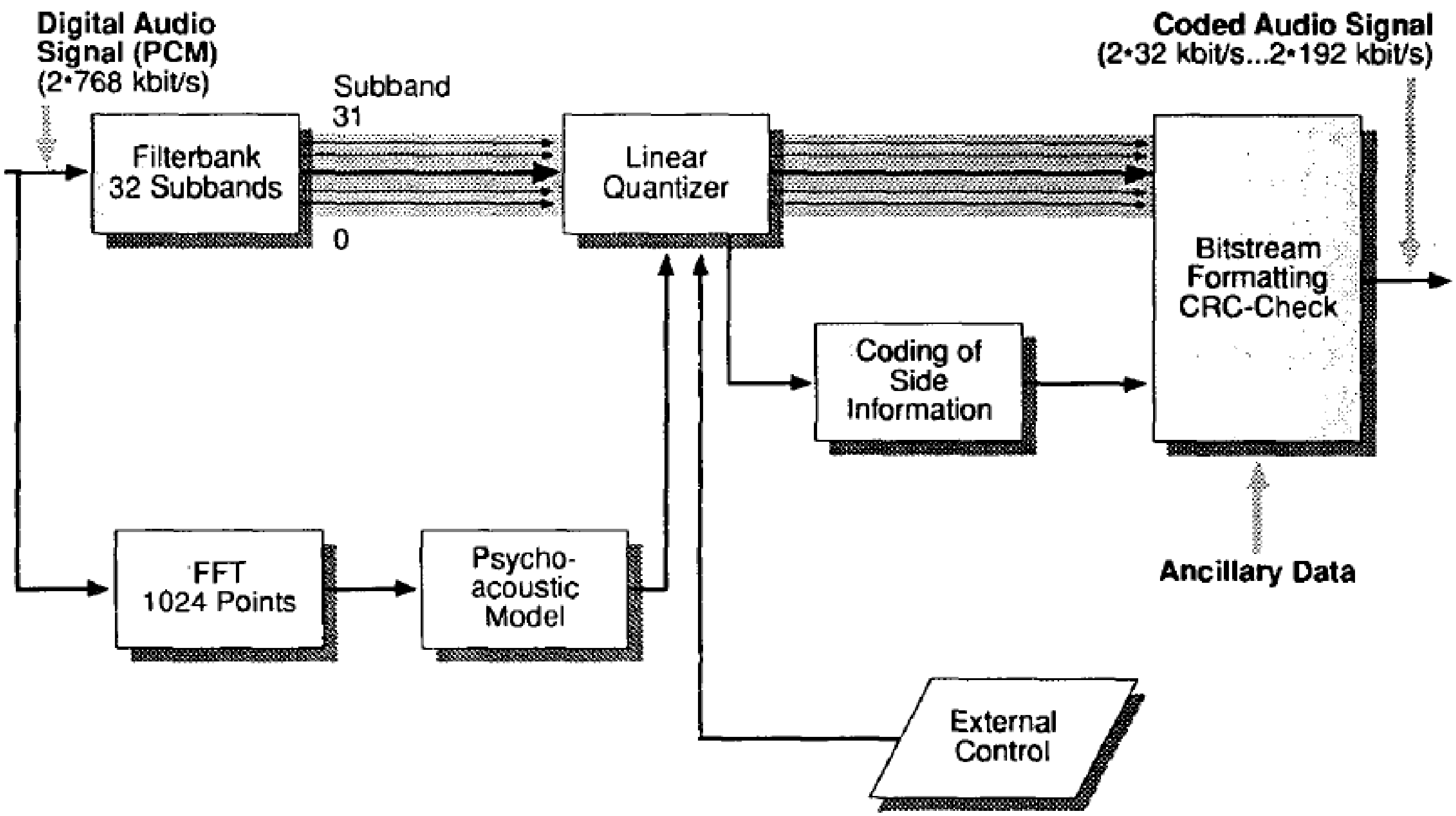


Fig. 2. Block diagram of ISO/MPEG/Audio encoder, Layer I and II (single-channel mode).

MPEG-1 Layer 3

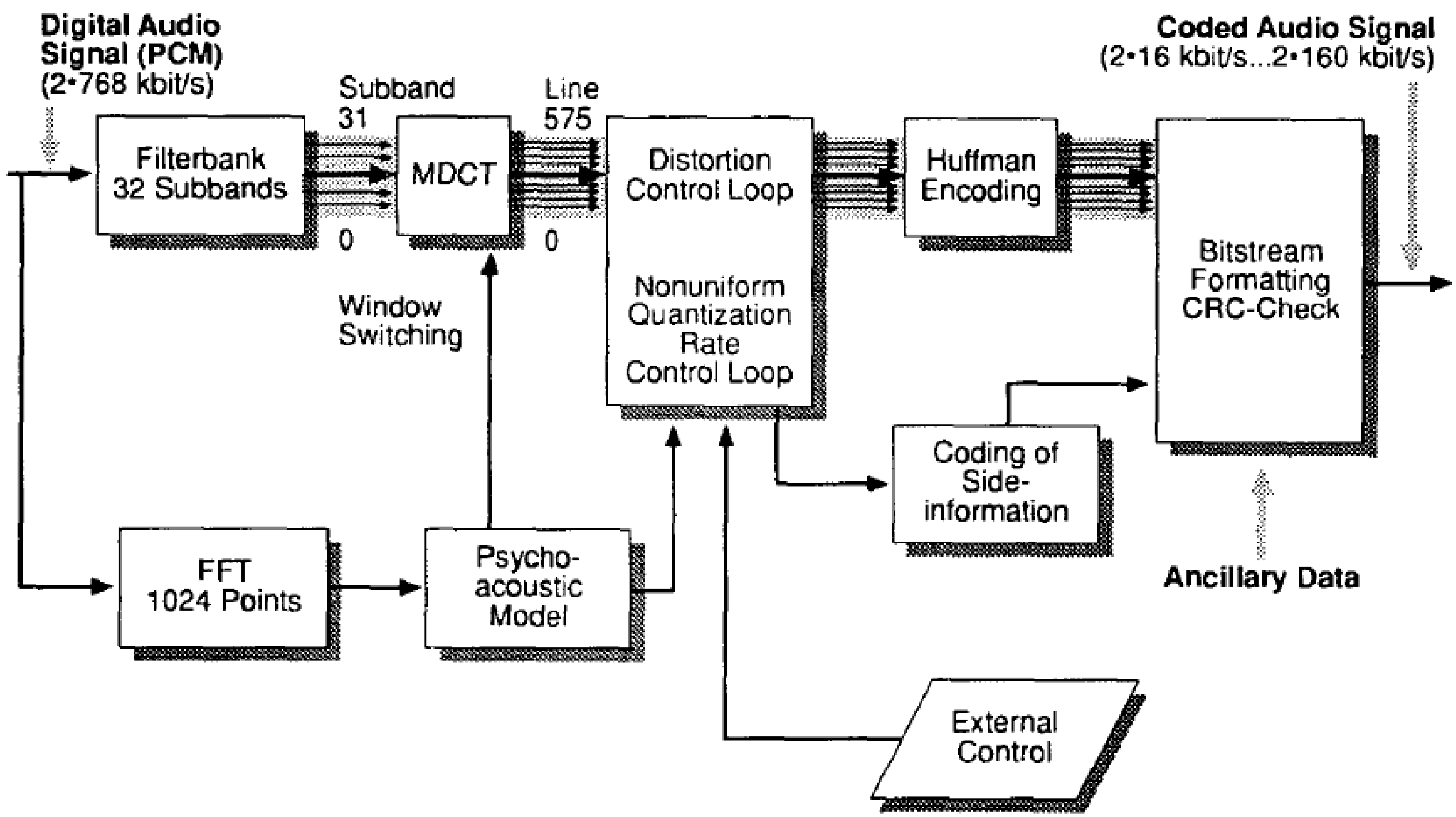


Fig. 5. Block diagram of ISO/MPEG/Audio encoder, Layer III (single-channel mode).

Bit stream: MPEG-1 Frame

Header	CRC	Audio Data	Ancillary Data

Layer 1: 384 PCM samples / chan
Layer 2: 1152 PCM samples / chan
Layer 3: 1152 samples / chan, divided into
2 “granules” of 576 samples each

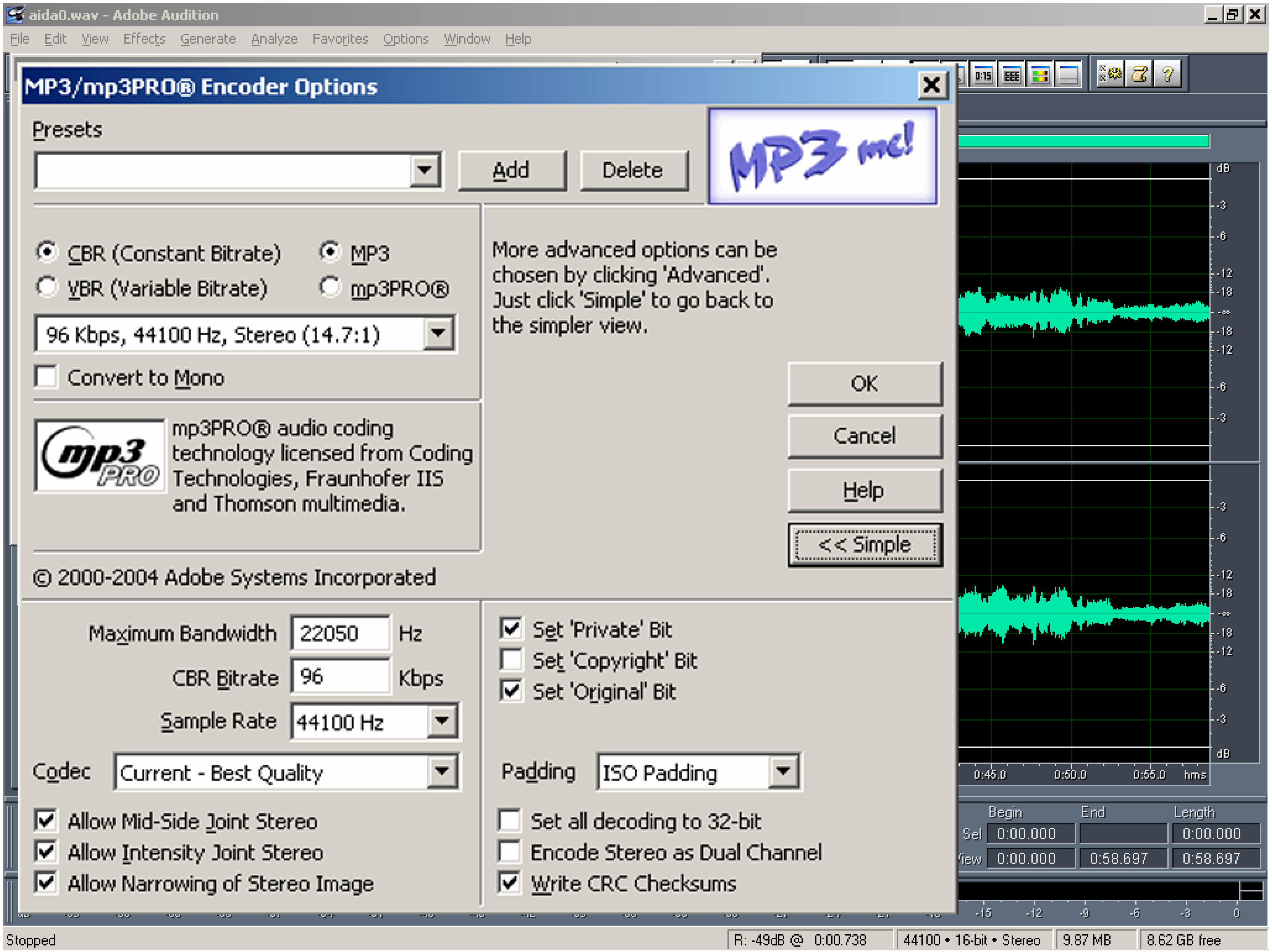
MPEG-1 Frame header

- Sync word
- ID (“1” for MPEG audio, “0” otherwise)
- Layer (2 bits)
- Protection bit
- Bit rate index (4 bits)
- Sampling frequency (2 bits)
- Stereo mode (2 bits)

MPEG-1 Frame header (con’t.)

- Private
- Copyright
- Original/copy
- Emphasis applied?
- Details: ISO/IEC 11172-3

Creating MPEG-1 header



Unravelling an MP-3 header

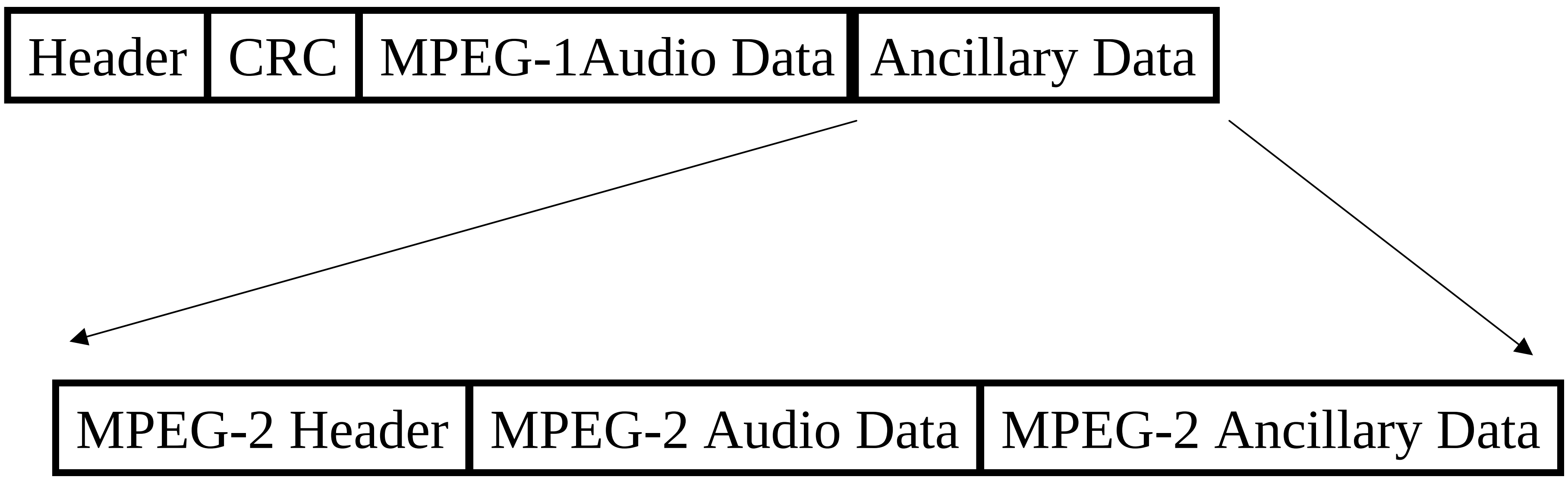
```
ID        1  layer     1  prot.     1  bitr.     1  freq.     0  padd.     1
priv.     0  mode      1  mext.     2  copr.     0  orig.     1  emph.     0
crc       ------
bits_in_frame       840

decoded:     layer III, 44.1 kHz, 32 kbit/s, joint_stereo
```

Thanks to (as always!) Fraunhofer, this time for their mdec program

MPEG-2 Audio

- 1994
- Motivation: video for digital TV
- Backward compatible with MPEG-1
  - Three layers, like MPEG-1
- Broader ranges
  - sample rates: 16, 22.05, 24 kHz
  - data rates: 8 - 1130 kbit/sec
  - channels: 5.1 + up to 7 multilingual/commentary channels
- "MP3" = MPEG-1/2 Layer 3
- MPEG-2.5: 8, 12, 11.025

Bit Stream: MPEG-2 Extensions



Next analysis (?) session

- Meet by ______ in listening areas, break off in groups of 2s and 3s.
- Examine the signals in (use your laptop!)
  - Martha White Orig.wav
  - Martha White mp3.wav
  - (OK to listen too of course!)
- Take notes and discuss:
  - How are they physically different?
- Back here at 12:05 to discuss. (20 min)

Notes on Differences

Rhonda Vincent and the Rage. “The Martha White Theme,” from *The Storm Still Rages*. Rounder Records 11661-0474-2.

Discussion of Differences

Original

Encoded/
Decoded

See also http://lame.sourceforge.net/tech-FAQ.txt and Wikipedia, “gapless playback”

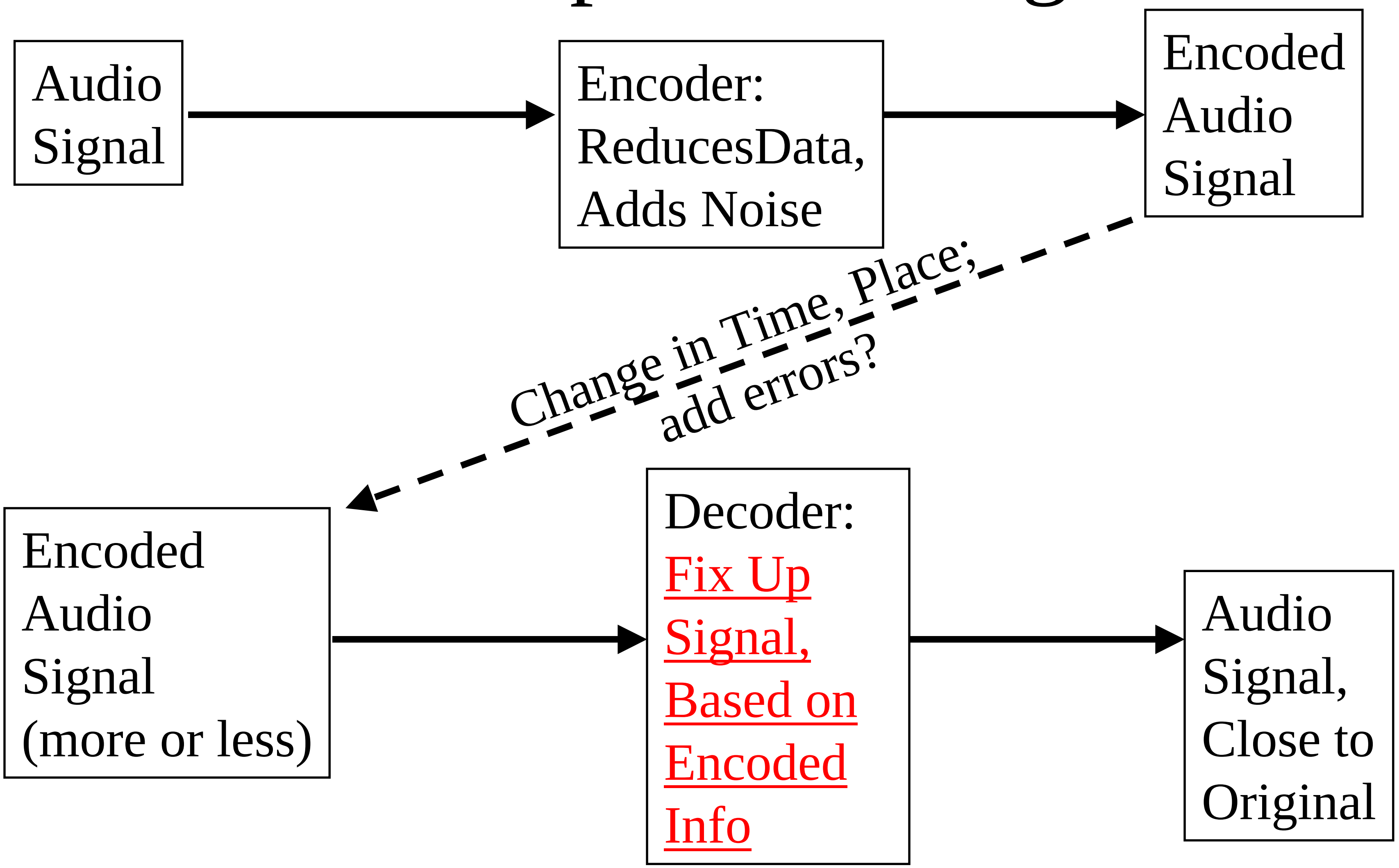
MPEG-2 Advanced Audio Coding (AAC)

- 1997
- “Indistinguishable” at 384 kbit/sec
- Features:
  - “Non-backward-compatible” (“NBC”)
  - Up to 48 channels (stereo, 5.1 …)
  - 8 - 96 kHz sample rate
  - Maximum 48-576 kbit/sec per channel
  - “Tools” combined into “profiles” (LC, SSR, Main)

AAC

- Itunes:
  - Stereo Bit Rate 128 kpbs default
  - 16 kbps thru 320 kbps available
- Flash Player
- MPEG-4

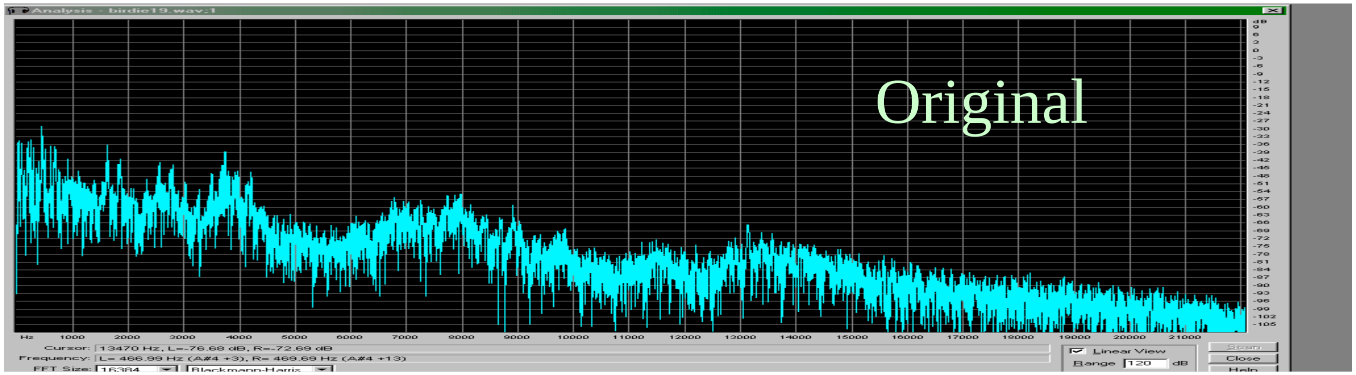
Perceptual Coding

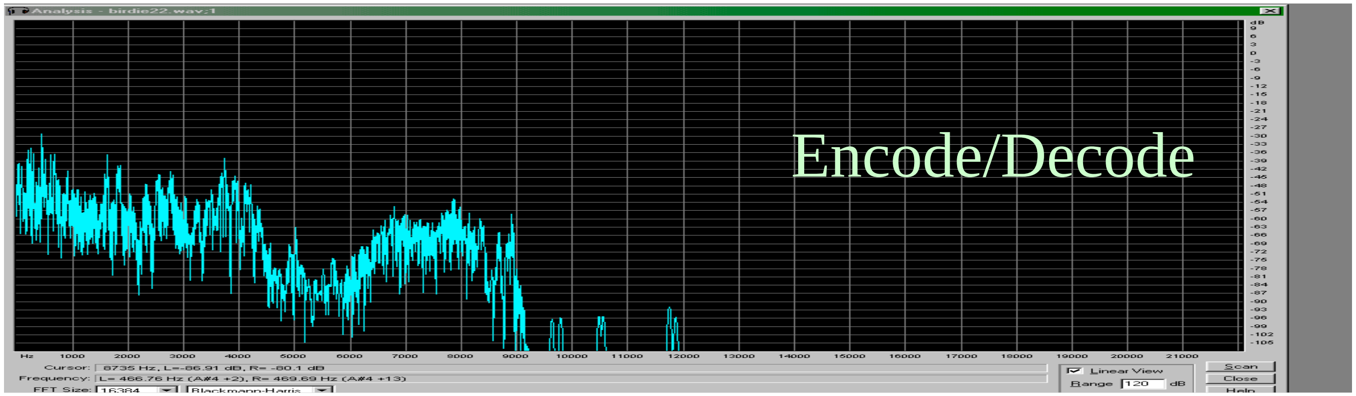


Decoder Postprocessing

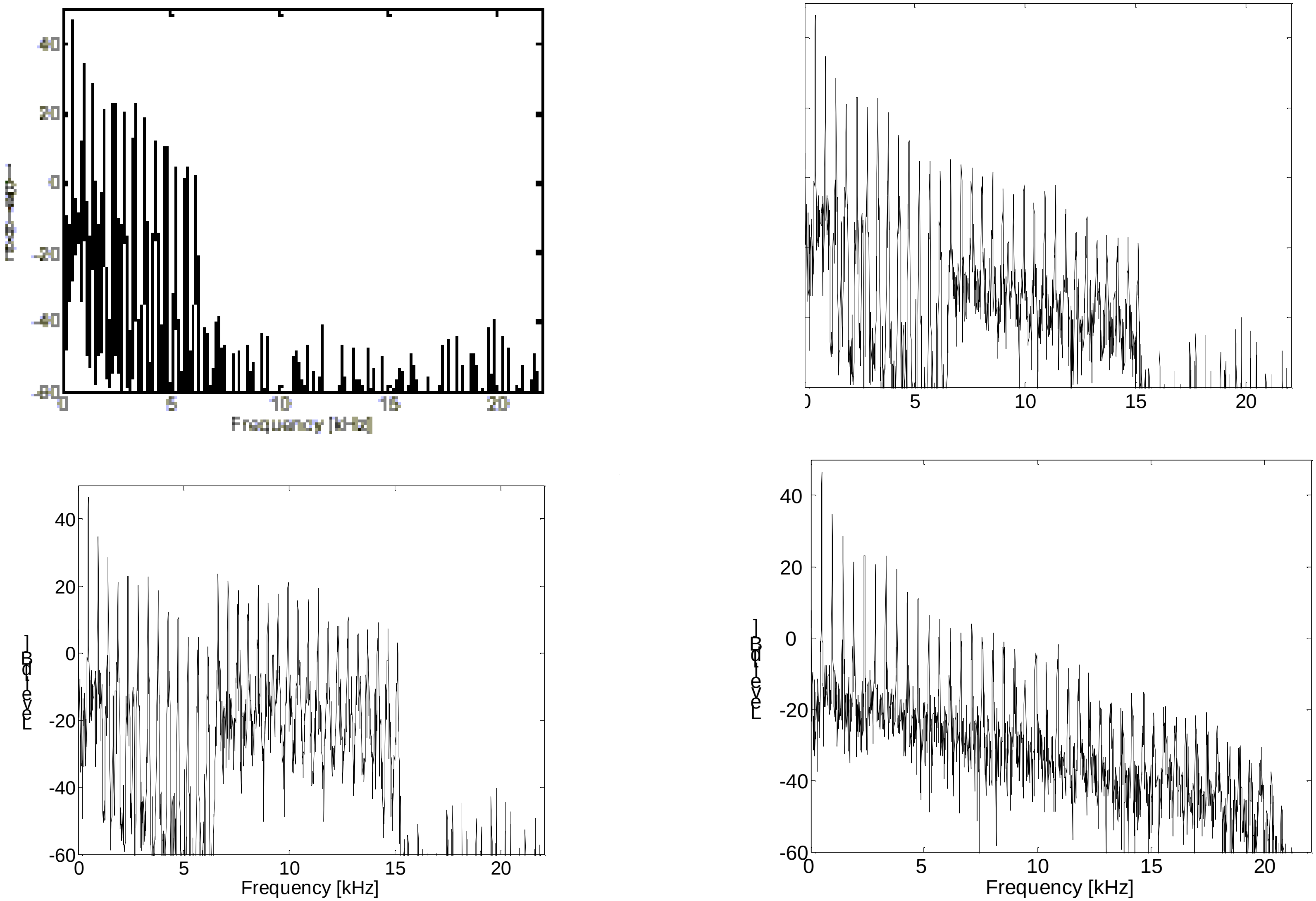
- Coding Technologies
- MP3 Pro
- aacPlus --- XM Radio

Review: Spectrum modified by compression (exaggerated)





Spectral Band Replication



Dietz et al. "Spectral Band Replication, a novel approach in audio coding" AES preprint 5553, 2002.

MPEG-4 Audio

- 1999/2000…
- Different compression techniques for different kinds of audio
- Intellectual property management
- Scalable bit rates

MPEG-4 Standards

- ISO/IEC 14496-3-1999, Information technology - Coding of audio-visual objects
- Part 1: Systems
- Part 2: Visual
- Part 3: Audio (ISO/IEC 14496-3)
- Part 4: Conformance testing
- Part 5: Reference software

MPEG-4 Standards

- Part 6: Delivery Multimedia Integration Framework (DMIF)
- Part 7: Optimized reference software for coding of audio-visual objects
- Part 8: Carriage of ISO/IEC 14496 contents over IP networks
- Part 9: Reference hardware description
- Part 10: Advanced video coding

MPEG-4 Standards

- Part 10: Advanced video coding
- Part 11: Scene description and application engine
- Part 12: ISO base media file format
- Part 13: Intellectual Property Management and Protection (IPMP) extensions
- Part 14: MP4 file format
- Part 15: Advanced Video Coding (AVC) file format

MPEG-4 Standards

- Part 16: Animation Framework eXtension (AFX)
- Part 17: Streaming text format
- Part 18: Font compression and streaming
- Part 19: Synthesized texture stream
- Part 20: Lightweight Application Scene Representation (LASeR) and Simple Aggregation Format (SAF)

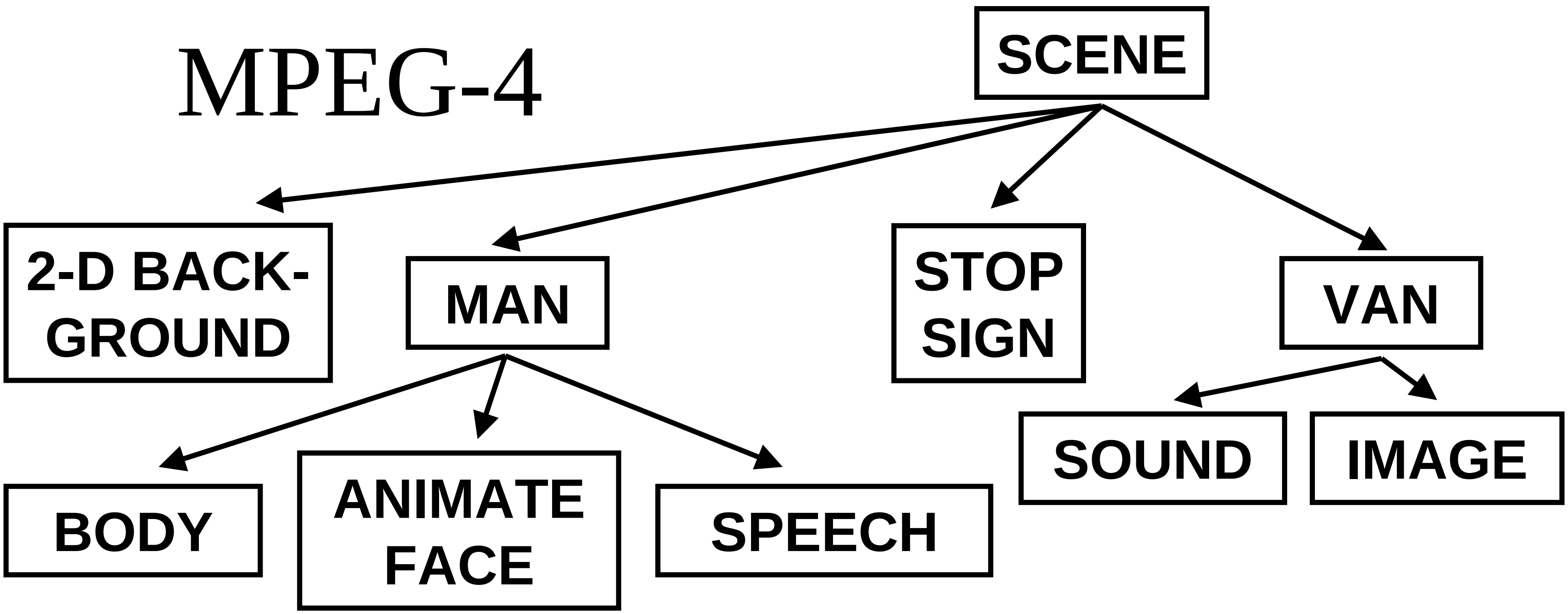
MPEG-4 Standards

- Part 21: MPEG-J Graphics Framework eXtensions (GFX)
- Part 22: Open Font Format
- Part 23: Symbolic Music Representation
- Part 24: Audio and systems interaction

MPEG-4: audio in context

- Media object
- Scene
- Animation
- Interaction
- Fundamental advance over MPEG-1, -2

MPEG-4

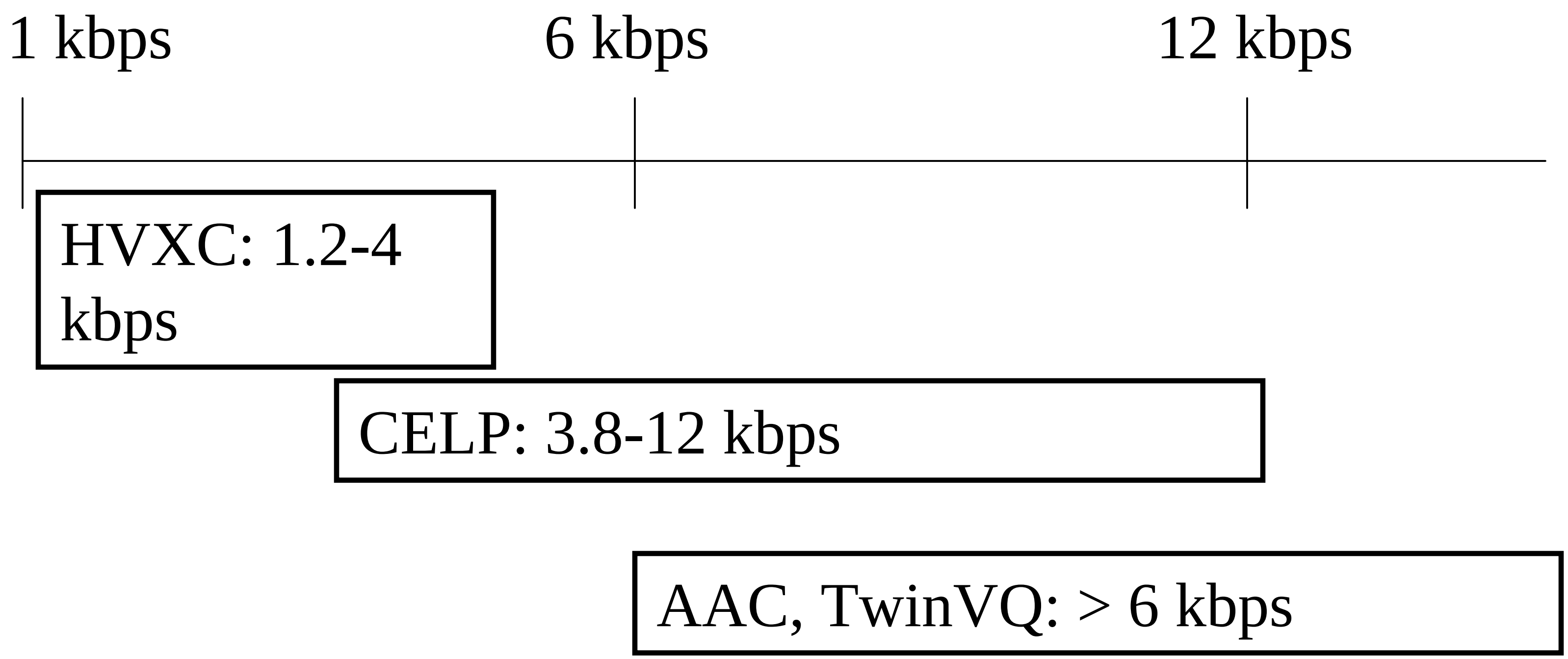

SCENE
2-D BACK-GROUND
MAN
STOP SIGN
VAN
BODY
ANIMATE FACE
SPEECH
SOUND
IMAGE


ALTO

MPEG-4 Audio Profiles

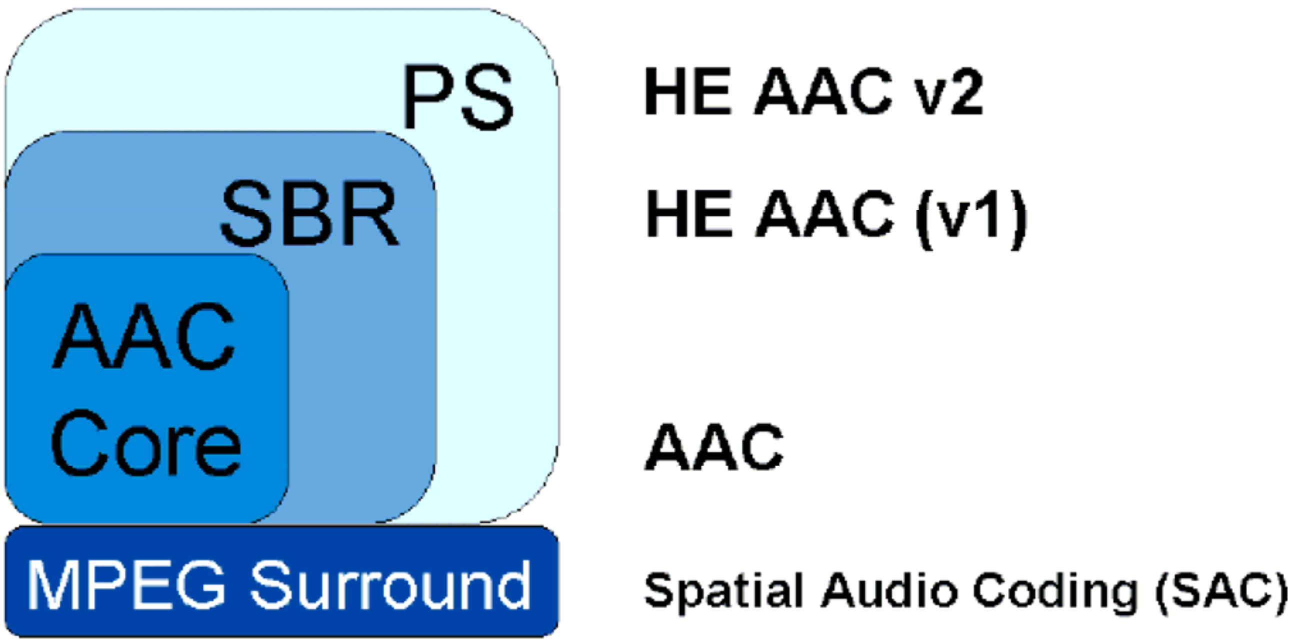
- Speech Profile
  - Two kinds of speech coding, 1.2 - 12 kbit/sec
  - Text-to-speech interface (TTS)
- Scalable Profile
  - AAC (one “tool” added), 4.6-64 kbit/sec/chan
  - TwinVQ
- Structured audio tools, 2-3 kbit/sec
  - SAOL: Structured Audio Orchestra Language
  - SASL: Structured Audio Score Language
- Main

MPEG-4 Scalable



After http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm#12.

AAC Expanded in MPEG-4

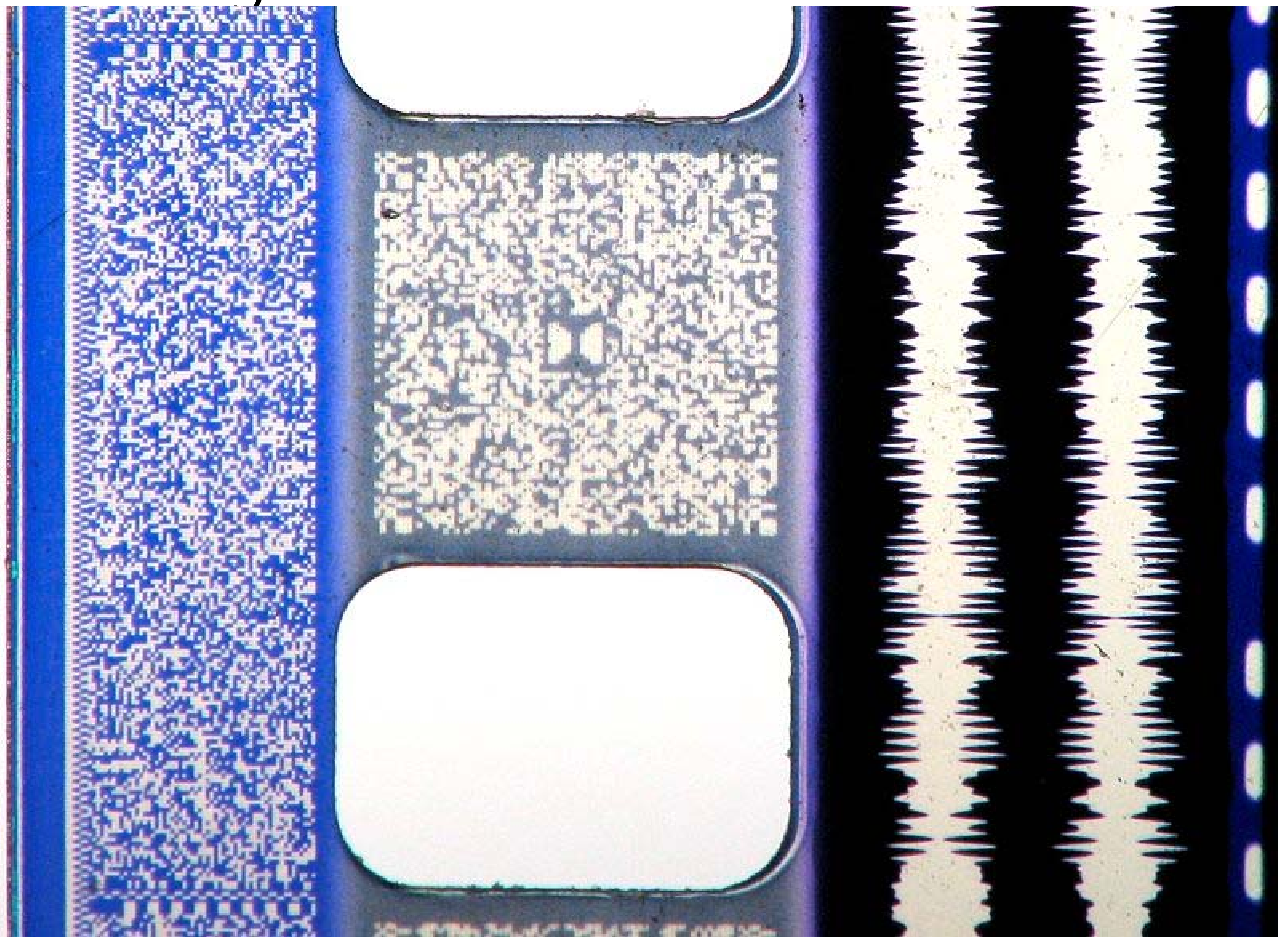


http://www.ebu.ch/en/technical/trev/trev_311-dab_dmb.pdf.

Algorithms related to Cinema

- DTS
- Dolby AC-3
- Sony ATRAC (1992) / SDDS (not in home formats)

They All Fit On the Film



left to right: Sony SDDS, Dolby Digital, analog Optical, and finally DTS time code.

http://en.wikipedia.org/wiki/Image:35mm_film_audio_macro.jpg

AC-3

advancedtelevisionsystemscommitteeinc

Digital Audio Compression Standard (AC-3, E-AC-3)
Revision B

Document A/52B, 14 June 2005

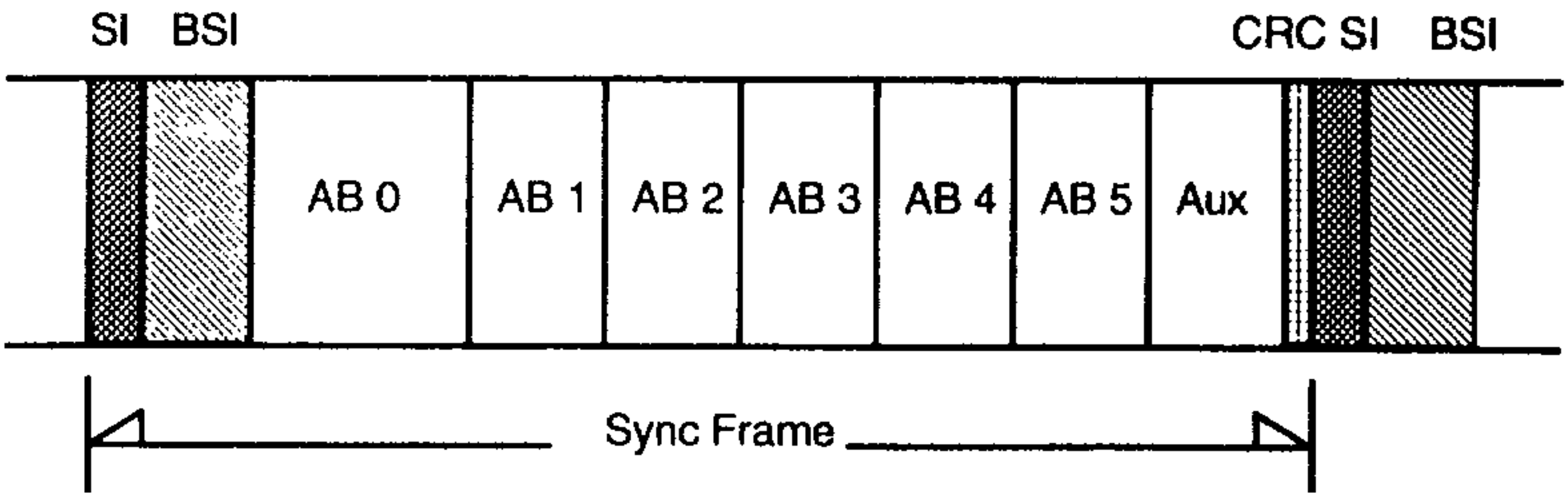
Advanced Television Systems Committee, Inc.
1750 K Street, N.W., Suite 1200
Washington, D.C. 20006

http://www.atsc.org/standards/a_52b.pdf

Dolby Digital (AC-3)

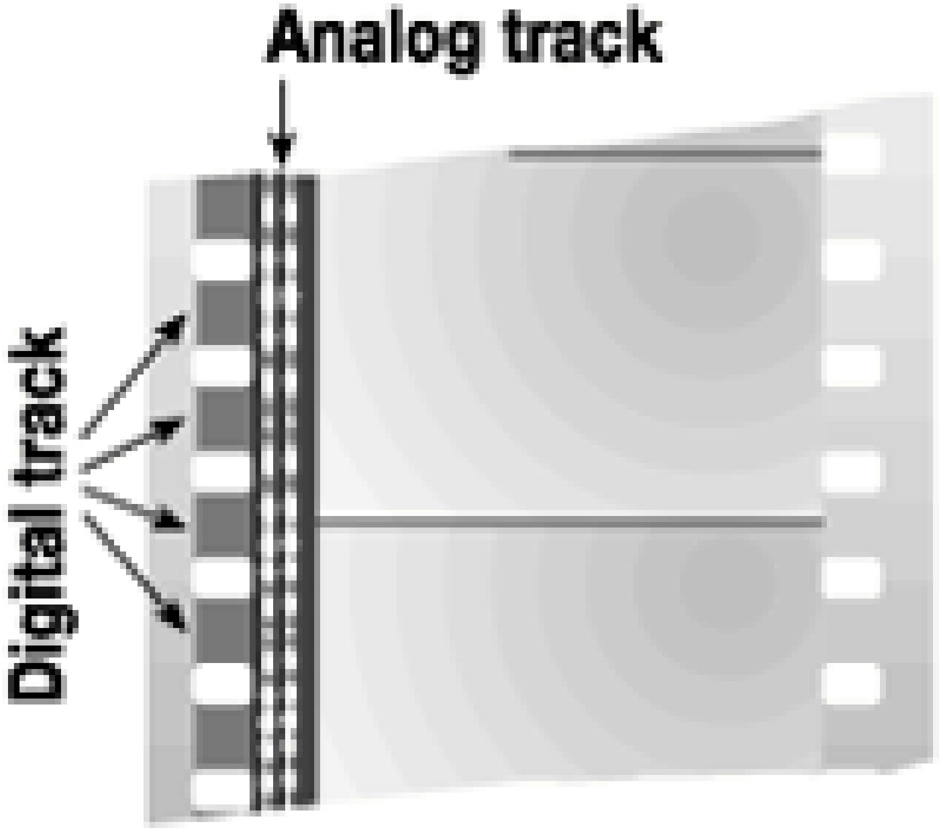
- 1993
- up to 5.1 channels
- MPEG-1 sample rates: 32, 44.1, 48 kHz
- 3 quality levels: 16-, 18-, and 20-bit
- 32 - 640 kpbs
- Applications
    - Cinema
    - DVD in NTSC countries
    - US digital television

Bit Stream: AC-3



From: Fielder et al., AES Preprint 4022, 1995

Dolby Digital (AC-3)



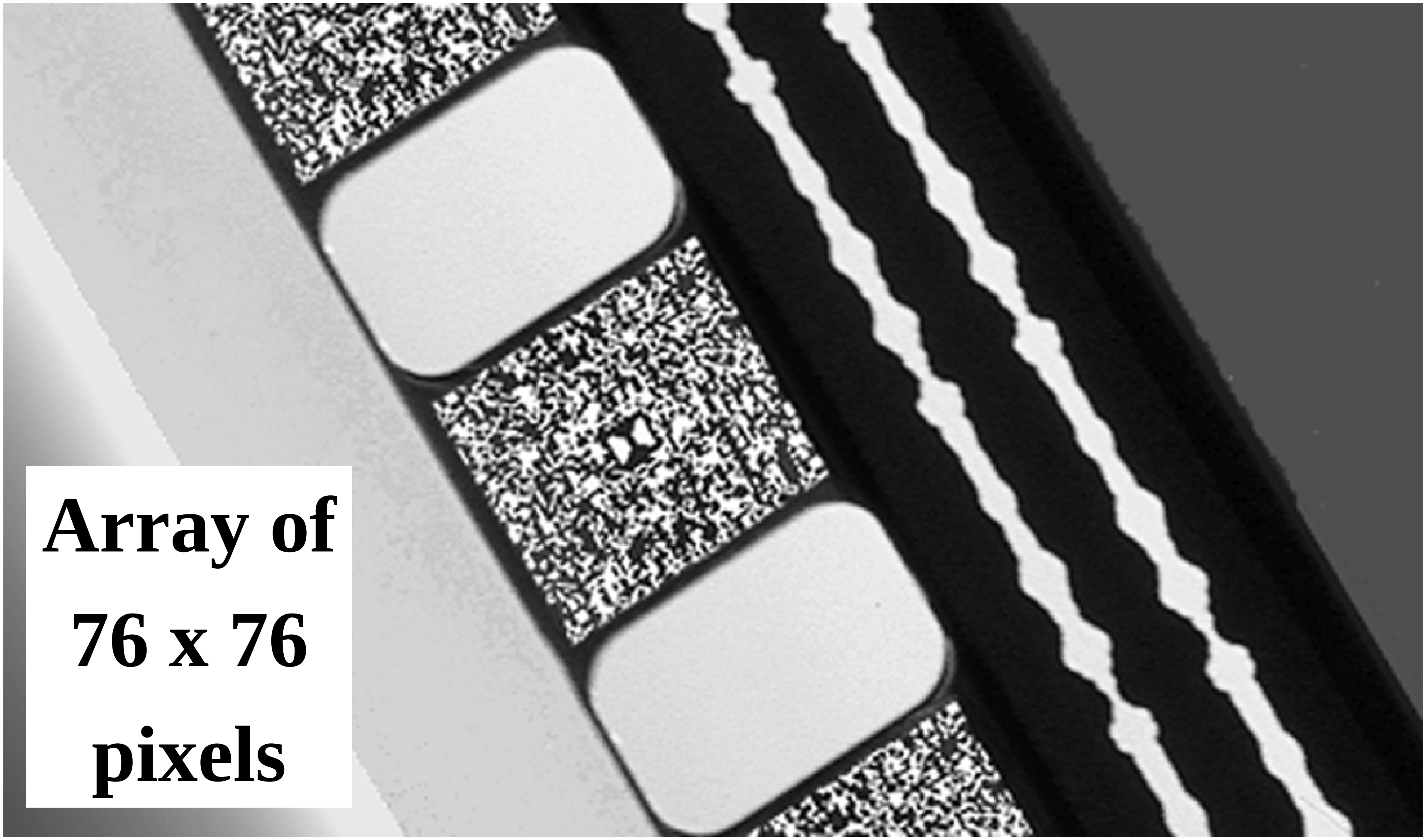
Dolby Digital release print

Taken from http://www.dolby.com/professional/motion_picture/technologies2.html

Two-D Barcode

http://www.barcodeman.com/faq/2d.php

- Dolby Digital Film Format



Courtesy Dolby Laboratories Inc.

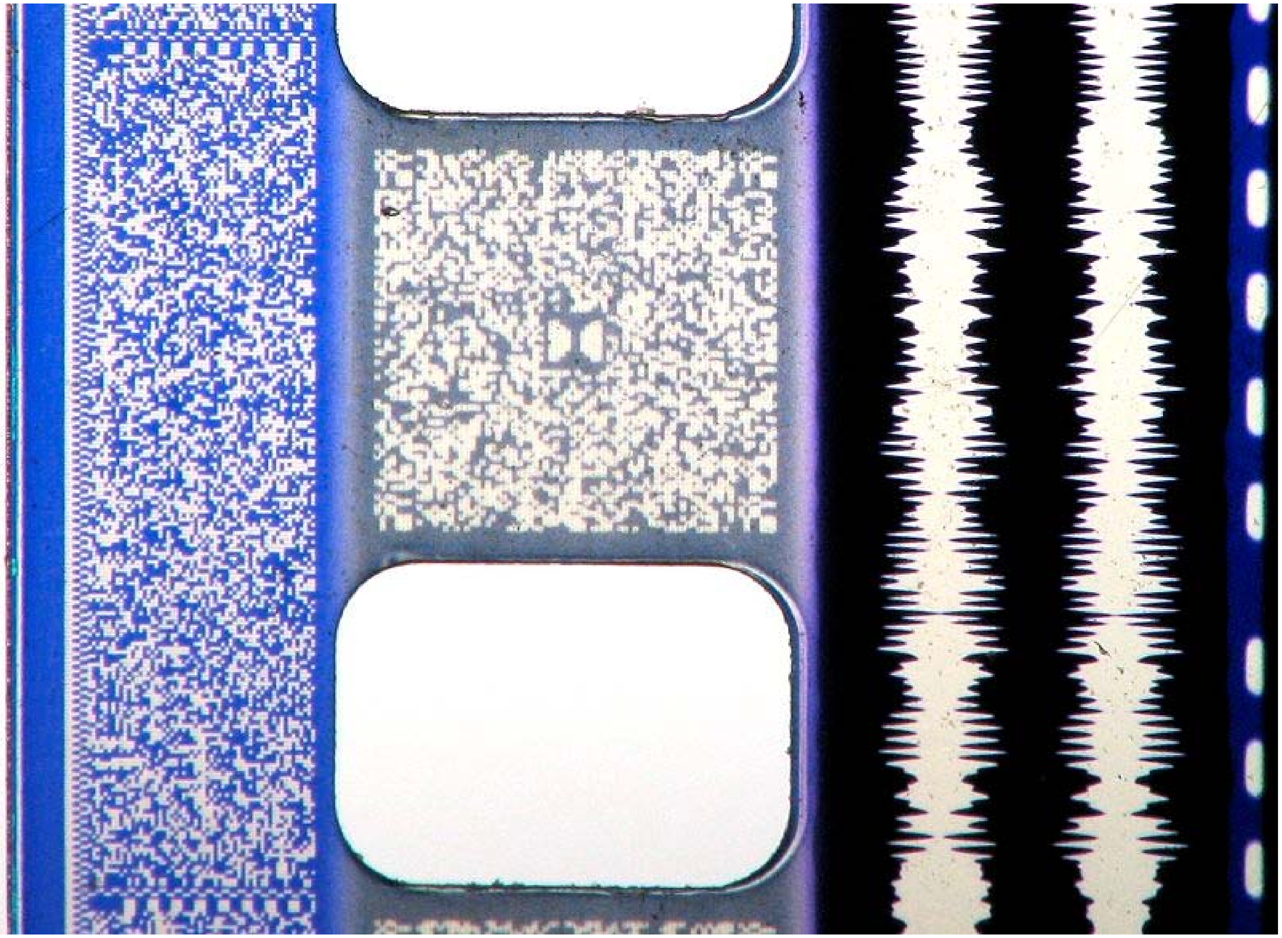
Dolby E

- For inside production, not for consumer
- Up to 8 channels, e.g. 5.1 + (Lt/Rt)
- Frame rate matches video rates
- 20-bit audio (later: 16-, 24-bit), 48 kHz
- “Up to ten encode/decode cycles without degradation.”

Dolby E

- Applications:
  - sending a program to a local station for commercial insertion
  - routing program within the same studio for voice-over editing
  - sending program via satellite to another broadcast facility.

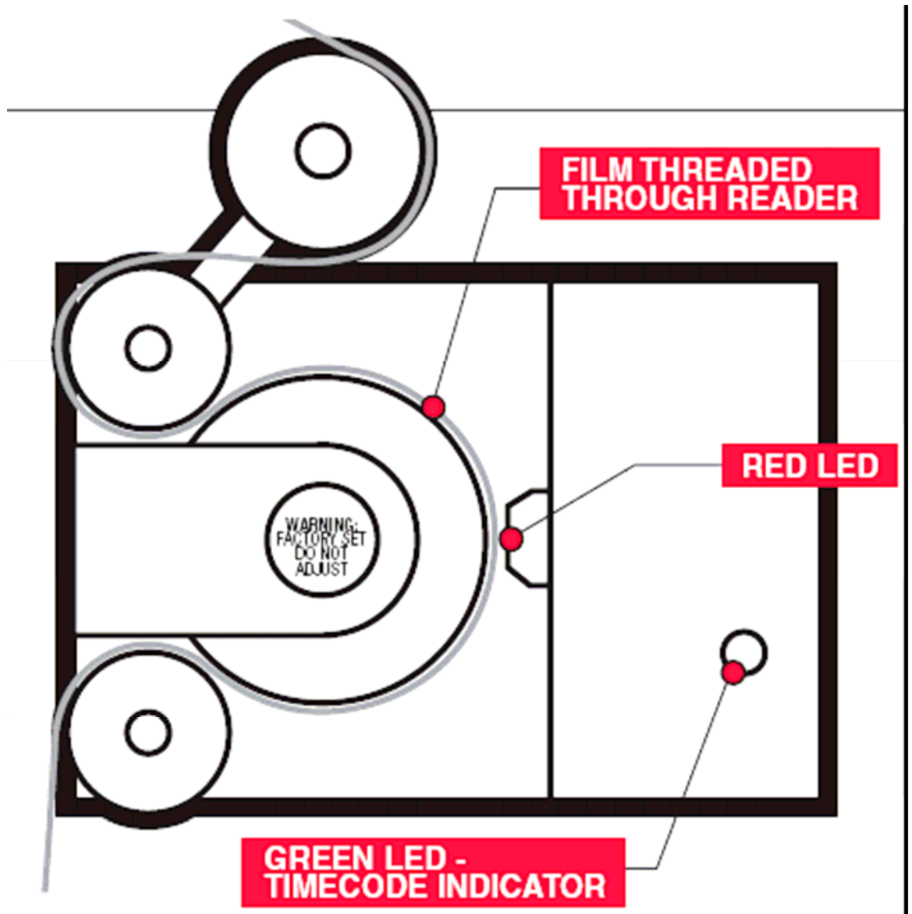
DTS Time Code



left to right: Sony SDDS, Dolby Digital, analog Optical, and finally DTS time code.

http://en.wikipedia.org/wiki/Image:35mm_film_audio_macro.jpg

DTS at the projector

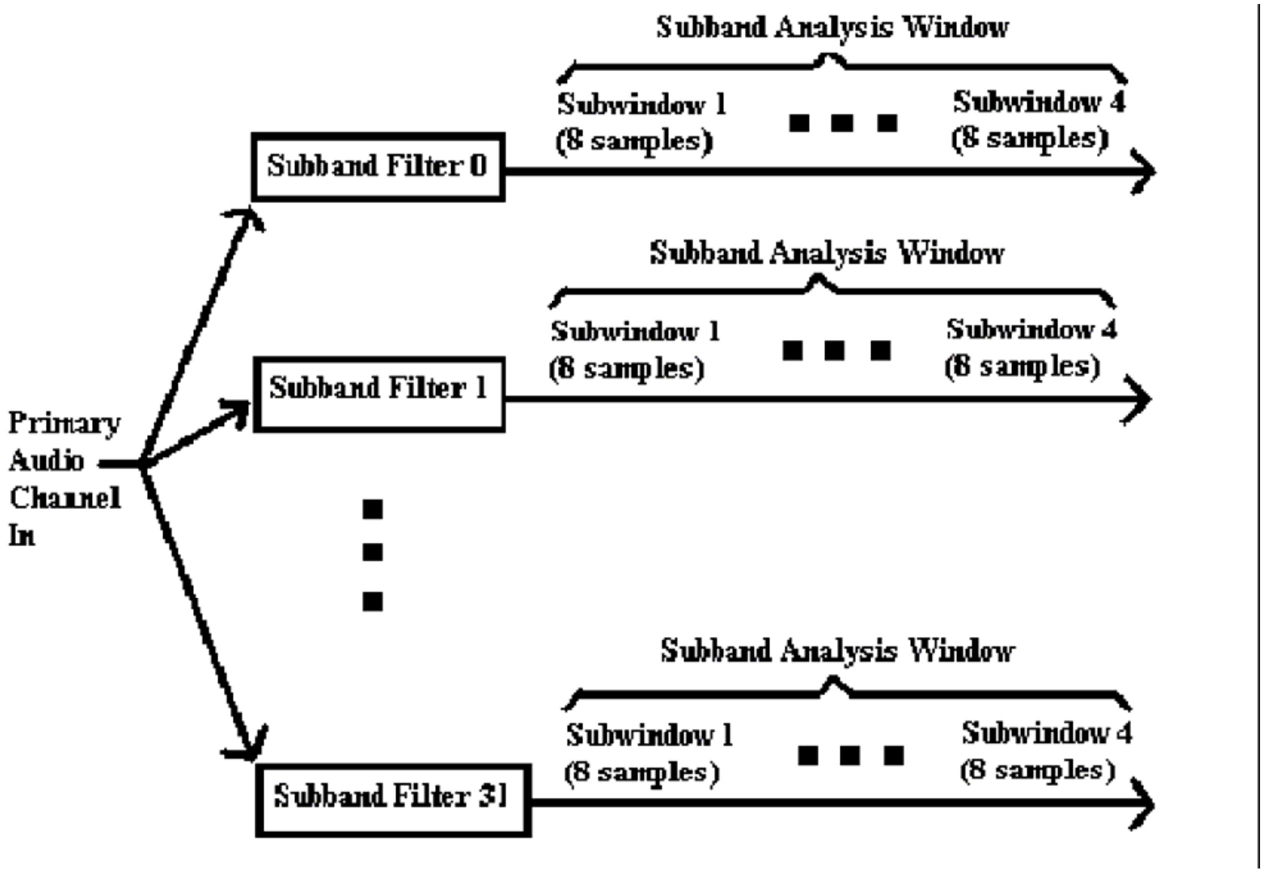


http://www.dts.com/media/2007/support/digitalcinema/products/XD10/TN-E714_(XD10_Wall_Chart).pdf

DTS Theatre CD player

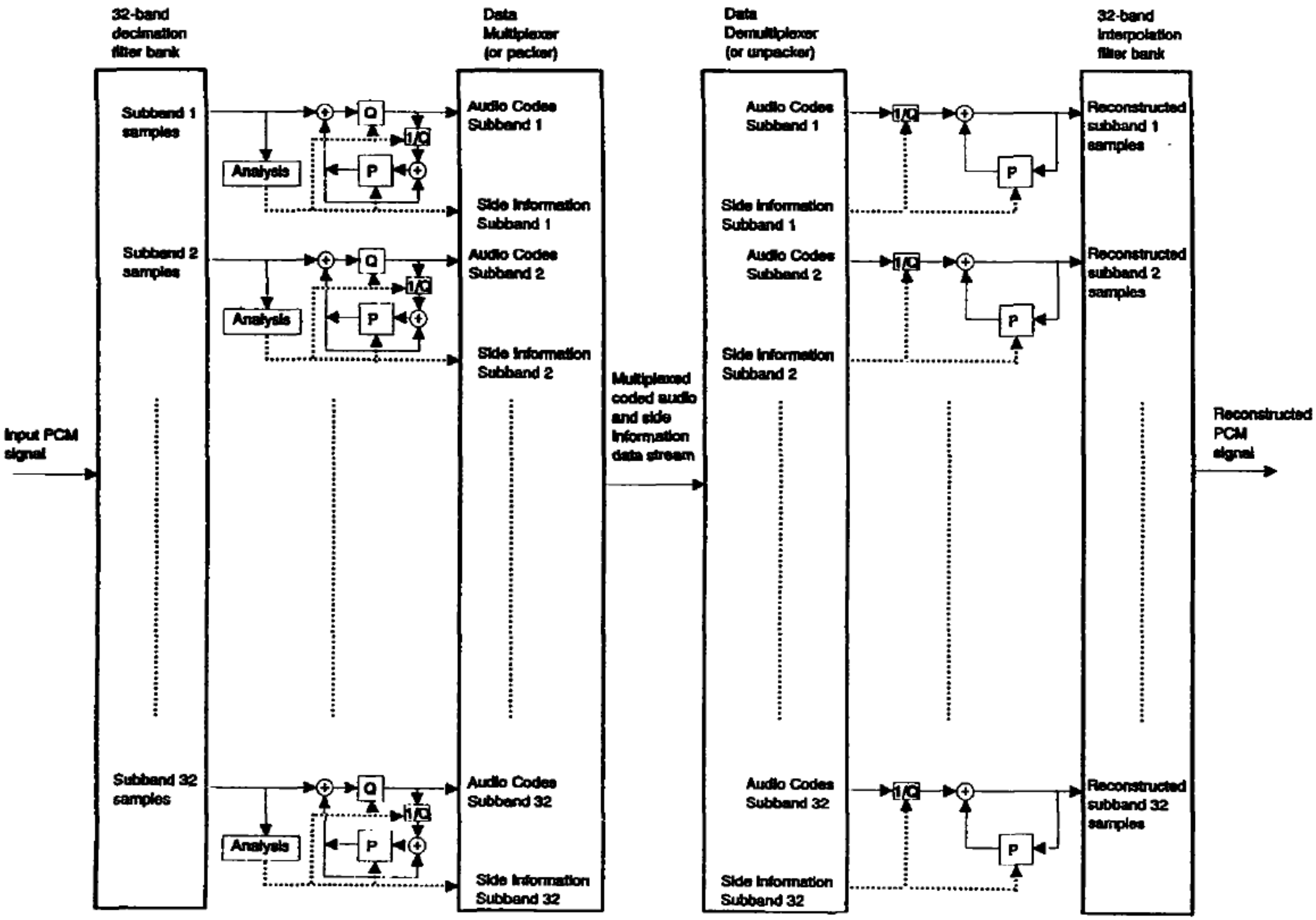
http://www.dts.com/media/2007/support/digitalcinema/products/XD10/XD10%20product%20sheet.pdf

DTS Coherent Acoustics



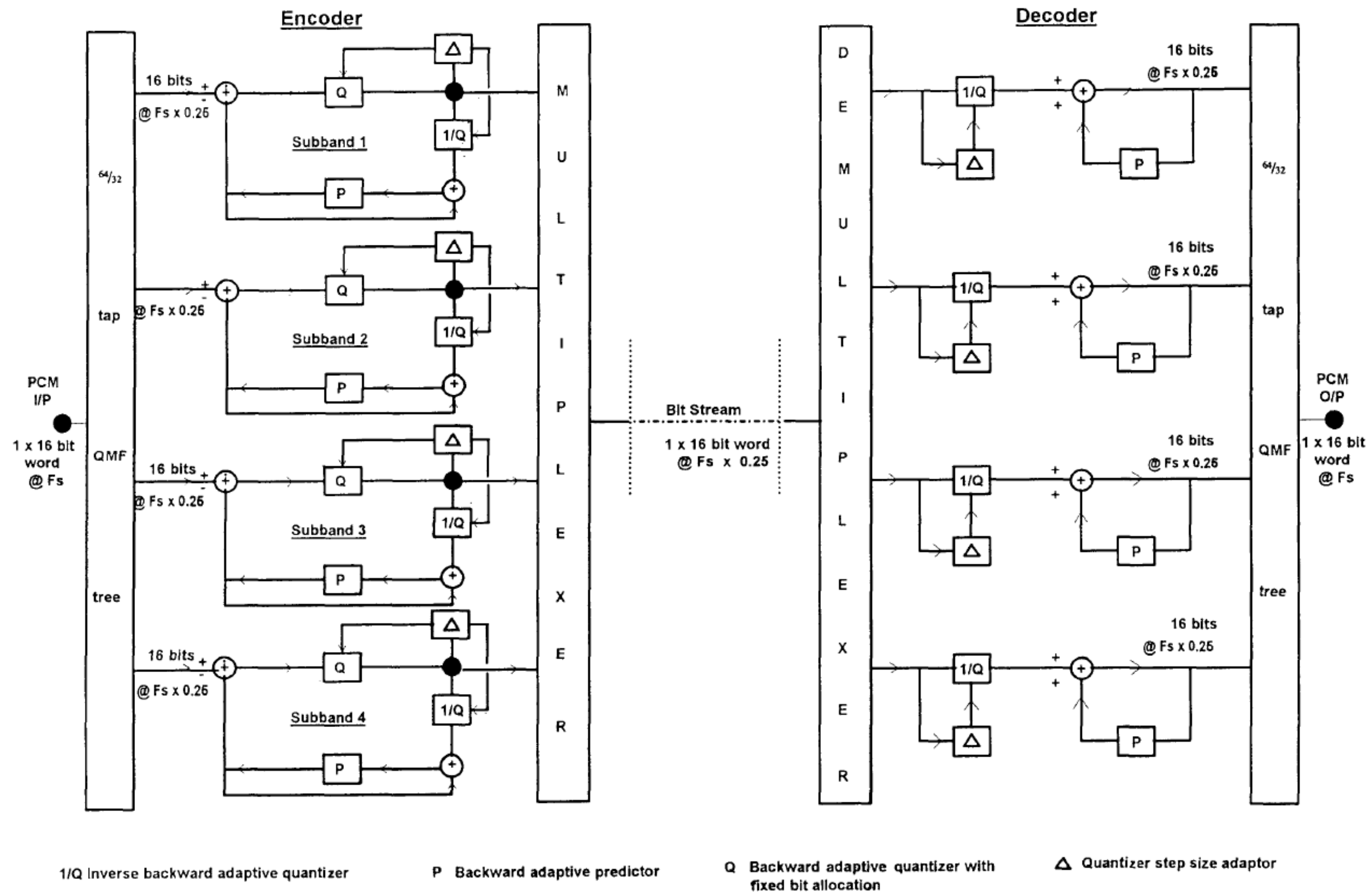
DTS Coherent Acoustics White Paper: Requirements Specification for Core Audio, Version Draft 2, n.d.

DTS Coherent Acoustics



Smyth, S. M. F.; Smith, W. P.; Smyth, M. H. C.; Yan, M.; Jung, T. "DTS Coherent Acoustics Delivering High-Quality Multichannel Sound to the Consumer." Preprint 4293, AES Convention, May 1996

apt-X100


Encoder
Decoder
PCM
I/P
1 x 16 bit
word
@ Fs
QMF
tap
tree
64/32
16 bits
@ Fs x 0.25
Q
1/Q
P
Subband 1
Subband 2
Subband 3
Subband 4
M
U
L
T
I
P
L
E
X
E
R
D
E
M
U
L
T
I
P
L
E
X
E
R
Bit Stream
1 x 16 bit word
@ Fs x 0.25
PCM
O/P
1/Q Inverse backward adaptive quantizer
P Backward adaptive predictor
Q Backward adaptive quantizer with fixed bit allocation
Quantizer step size adaptor

Other algorithms

- WMA
- Qdesign, QuickTime
- Lucent PAC, Sirius, iBiquity, IBOC
- ...

Meridian Lossless Packing (MLP)

- 1997
- Lossless
  - Therefore cascadable
- Up to 64 channels, up to 24 bits
- DVD-Audio Version 1 “Packed Audio”
- Variable bit rate

Some Other Lossless Algorithms

- Monkey’s Audio
- Free Lossless Audio Codec (FLAC)
- Shorten File (.shn)
- DAKX
- Apple Lossless
- WMA Lossless
- MPEG-4 Lossless
- WavPack

Licensing Algorithms

Audio for Blu-Ray

- Linear PCM (LPCM) - up to 8 channels of uncompressed audio. (mandatory)
- Dolby Digital
- DTS (and variants)

http://www.blu-ray.com/faq/#bluray_audio_codecs
http://www.emedialive.com/articles/readarticle.aspx?articleid=11397#iii

What we have covered

- Some history of (perceptual) coding
- Main codec families, their names, key features
- Lossless coding
- Analysis of decoded waveform

Next Listening Session

- Meet by _____ in listening areas.
- Listen to project Birdie: Birdie 19 (stereo pair), Birdie 22 (stereo pair)
- Take notes and discuss:
  - What do you hear?
  - How are they different?
  - What do you like?
  - What don’t you like?
- Back here at 15:00 to discuss. (10 min)

Notes on sound examples

MSRA 5500 - 002
MUS 4500 - 002

March 15, Session 4, Part 2

Bye Bye Birdies

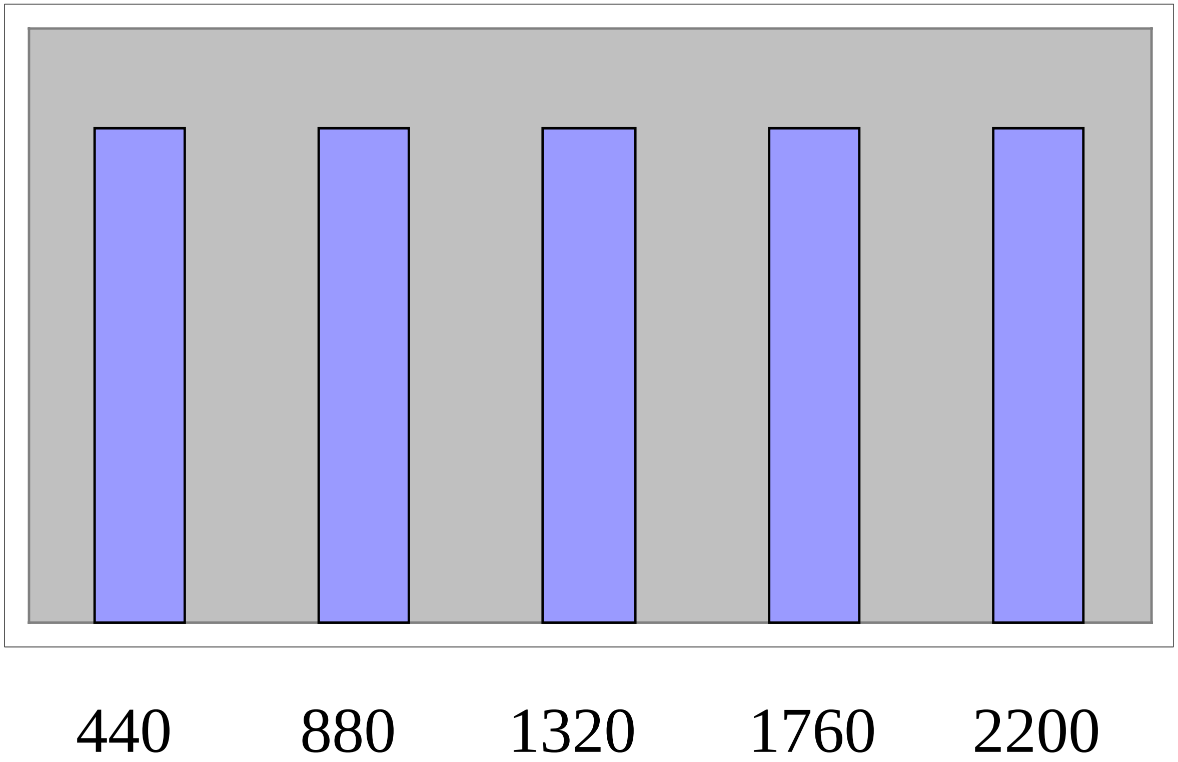
© Copyright 2008 John Strawn

What we will cover

- The origin and extermination of birdies.
- Classifying compression techniques.
- Evaluating compression techniques.
- Comparing compression techniques.
- Some typical encoder controls.
- Market forces on compression
- What to try when something goes wrong.

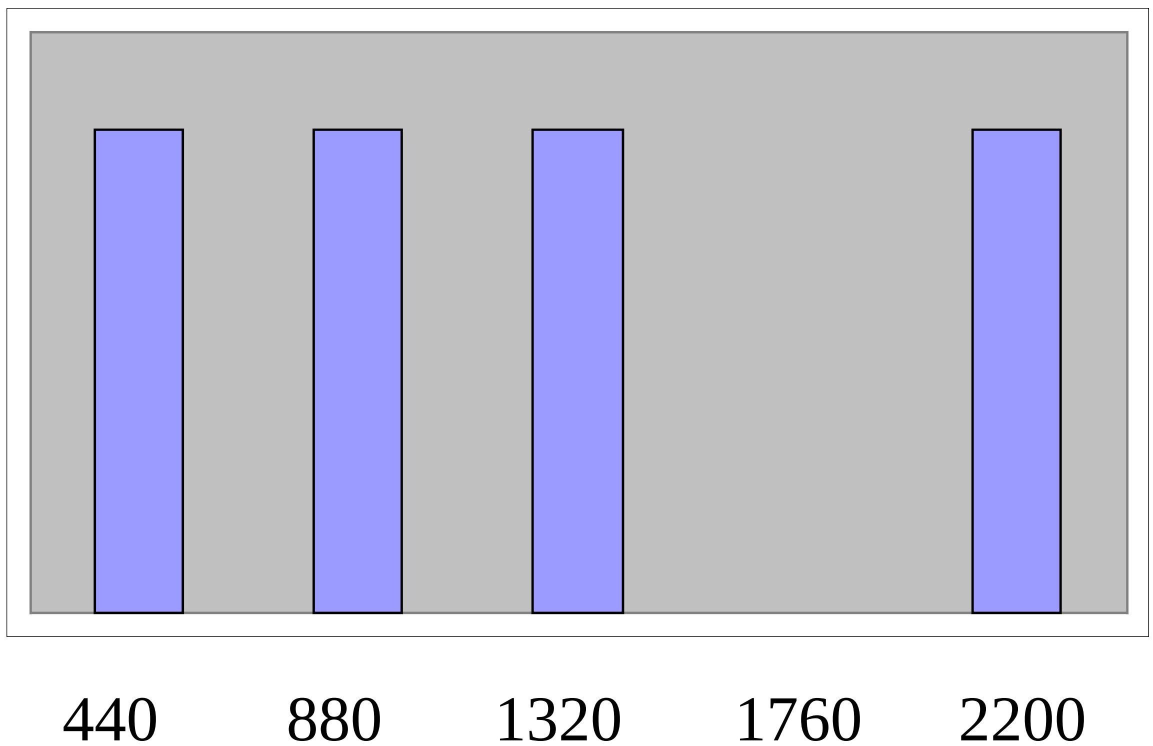
Discuss listening examples

Birdies: Artificial examples 5 harmonics



Source: AES CD-ROM.

Birdies: Artificial Examples
1760 Gone



Source: AES CD-ROM.

Birdies: artificial examples

- A440 + 4 harmonics: 880, 1320, 1760, 2200
- As above, but no 1760
- As above, 1760 jumps in and out

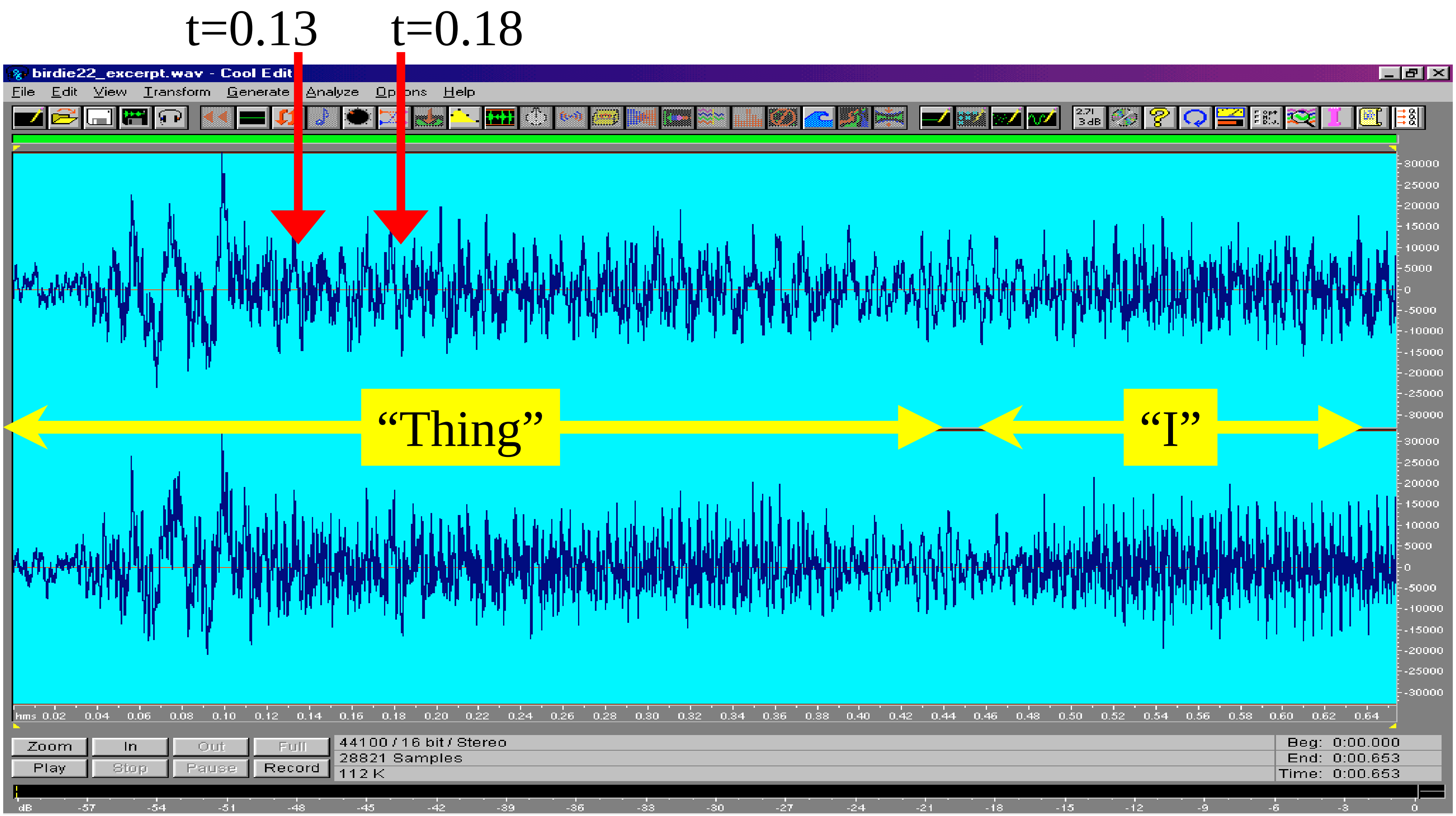
Source: AES CD-ROM.

Birdies: Sound Examples

- Speech example
- *Aida* example
- "Money" example
- Gilmour example
  - Original
  - With birdies
  - Zoom in on original: "Thing I"
  - Zoom in, with birdies

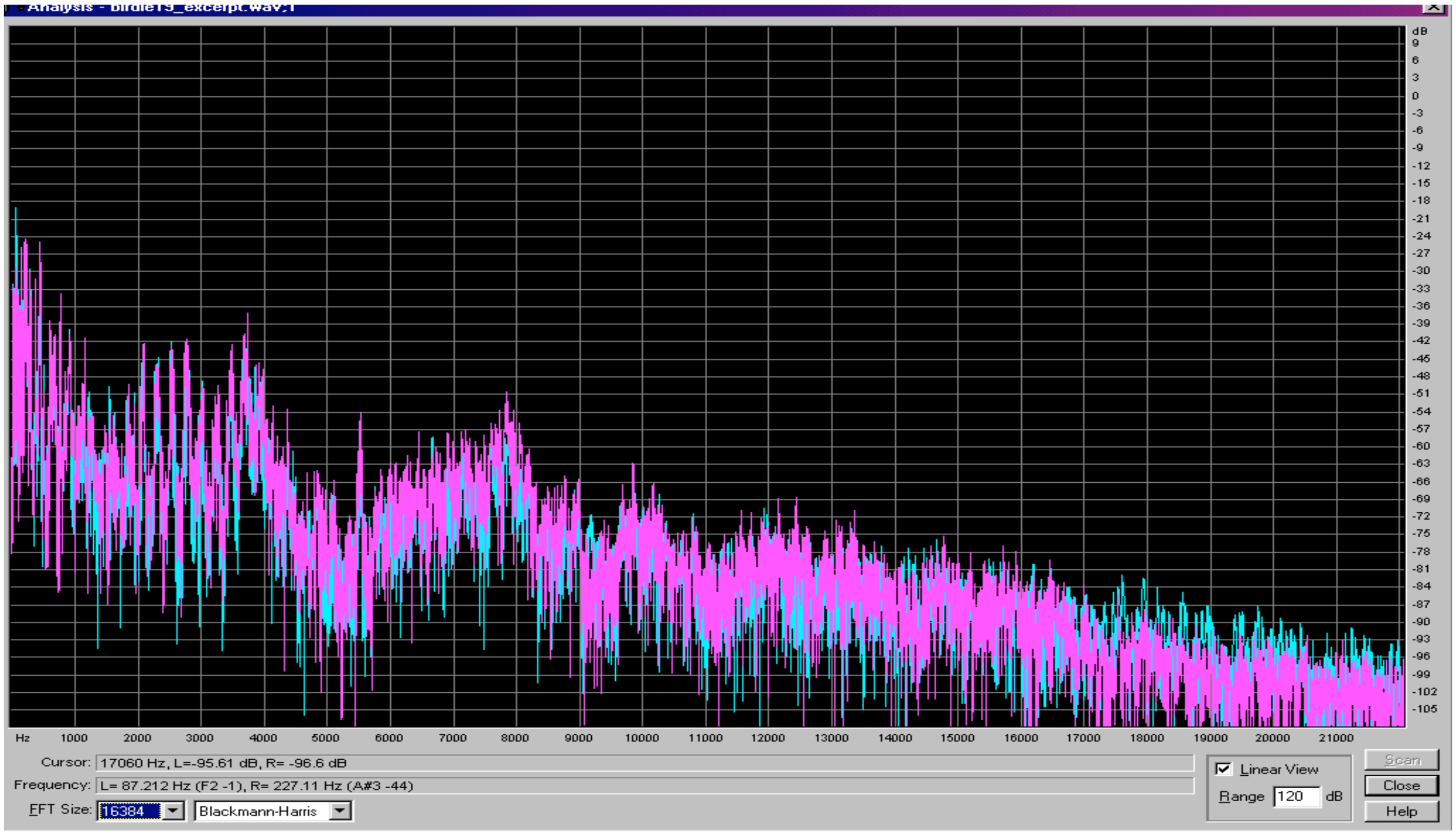
AES CD-ROM. Brian Gilmour, *Too Many Lies*, 1995.

Birdies: excerpt

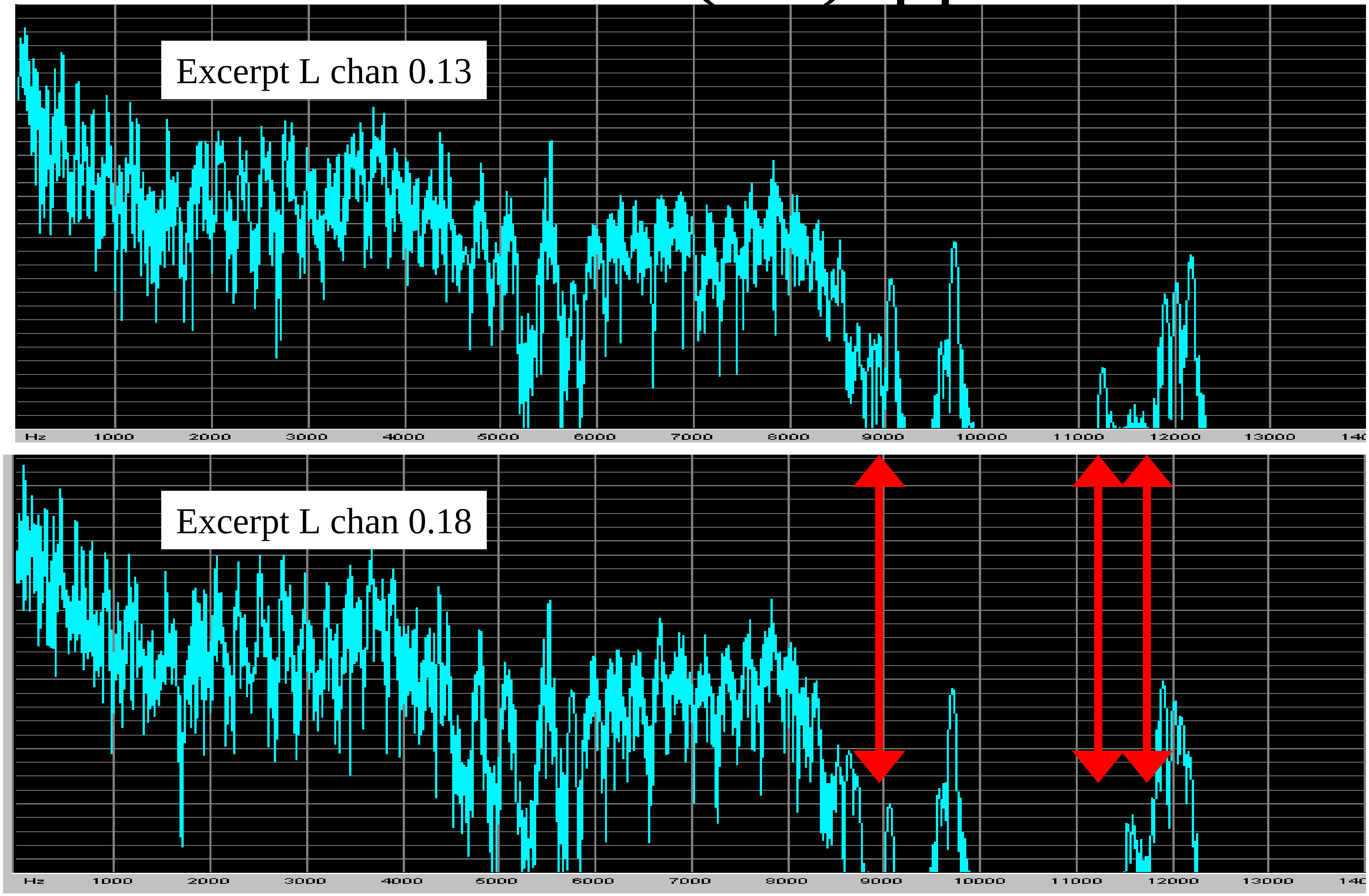


Birdies

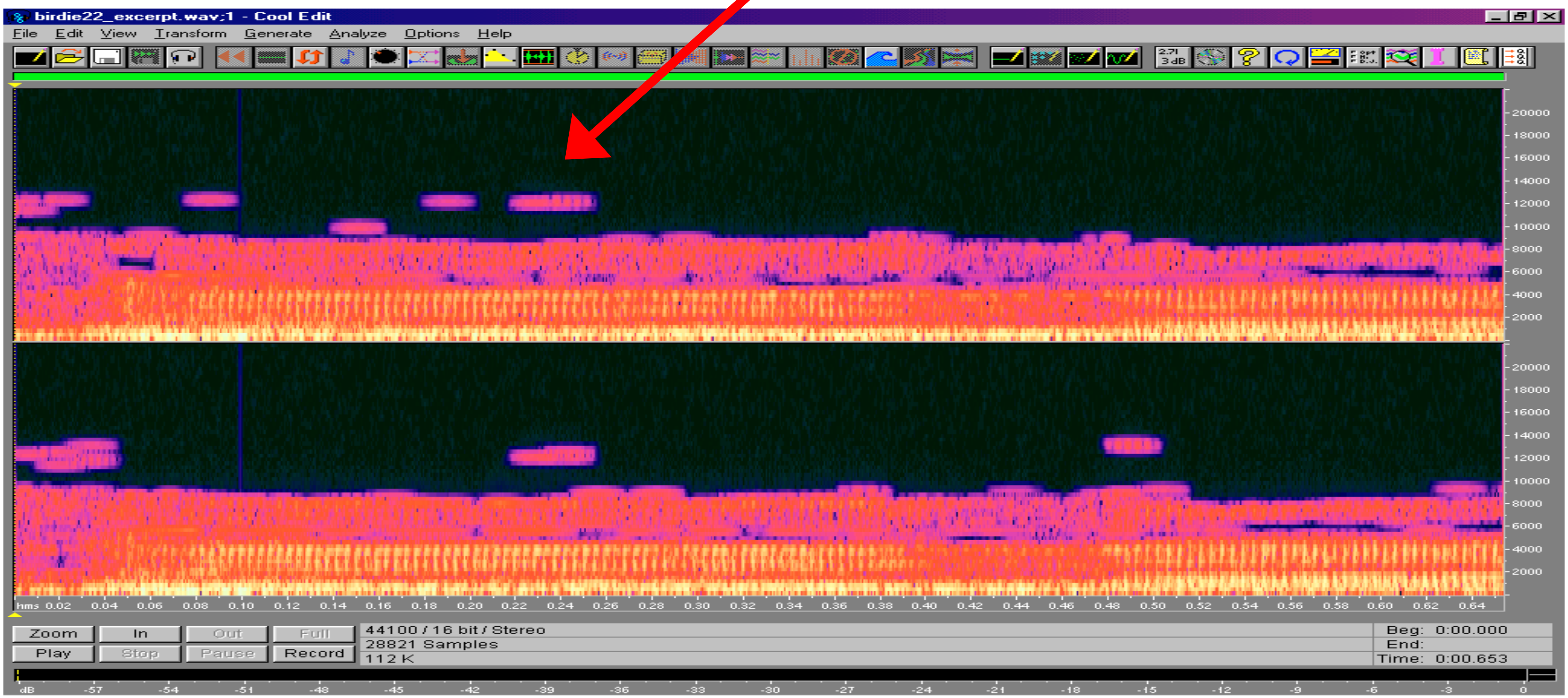
Typical Spectrum of “Thing I” Excerpt, original, 2 channels



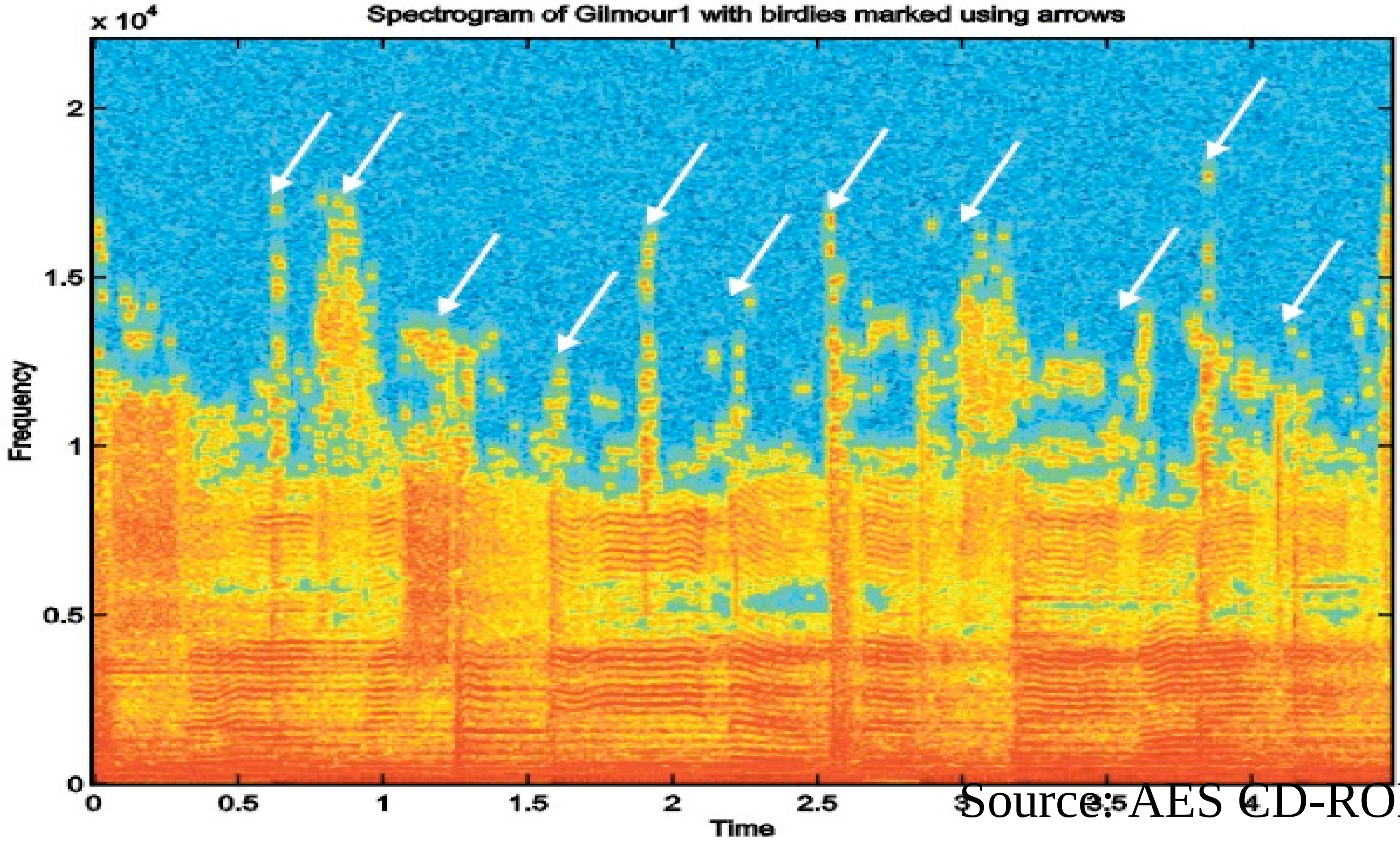
Birdies: Sudden (dis)appearance



Birdies: More Sudden (dis)Appearances

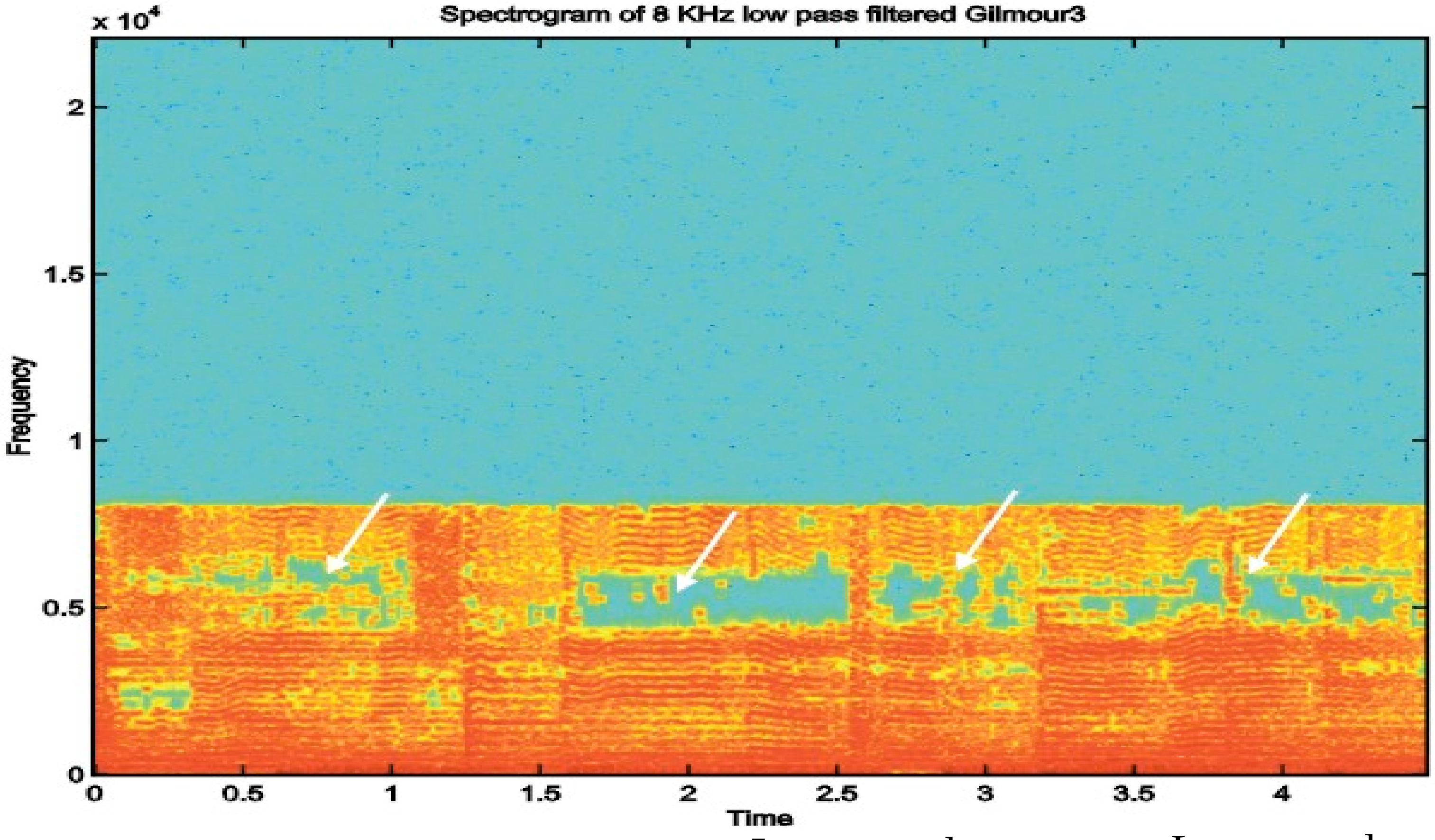


Birdies: Even More Sudden (dis)Appearances



Source: AES CD-ROM

Birdies: Low-pass may not help



Orig Birdies Lowpassed Orig Lowpassed, Birdies

Source: AES CD-ROM

What we will cover

- The care and starving of birdies.
- Classifying compression techniques.
- Evaluating compression techniques.
- Comparing compression techniques.
- Some typical encoder controls.
- Market forces on compression
- What to try when something goes wrong.

Classifying & evaluating compression techniques

RealNetworks Data Rates (stereo)

Transmission medium	Max	How far off from real time?
28.8 kbps modem	20 kbps	71
56 kbps modem	32 kbps	44
112 kbps dual ISDN	64 kbps	22
Corporate LAN	132 kbps	11
256 kbps DSL/cable modem	176 kbps	8
512 kbps DSL/cable modem	352 kbps	4

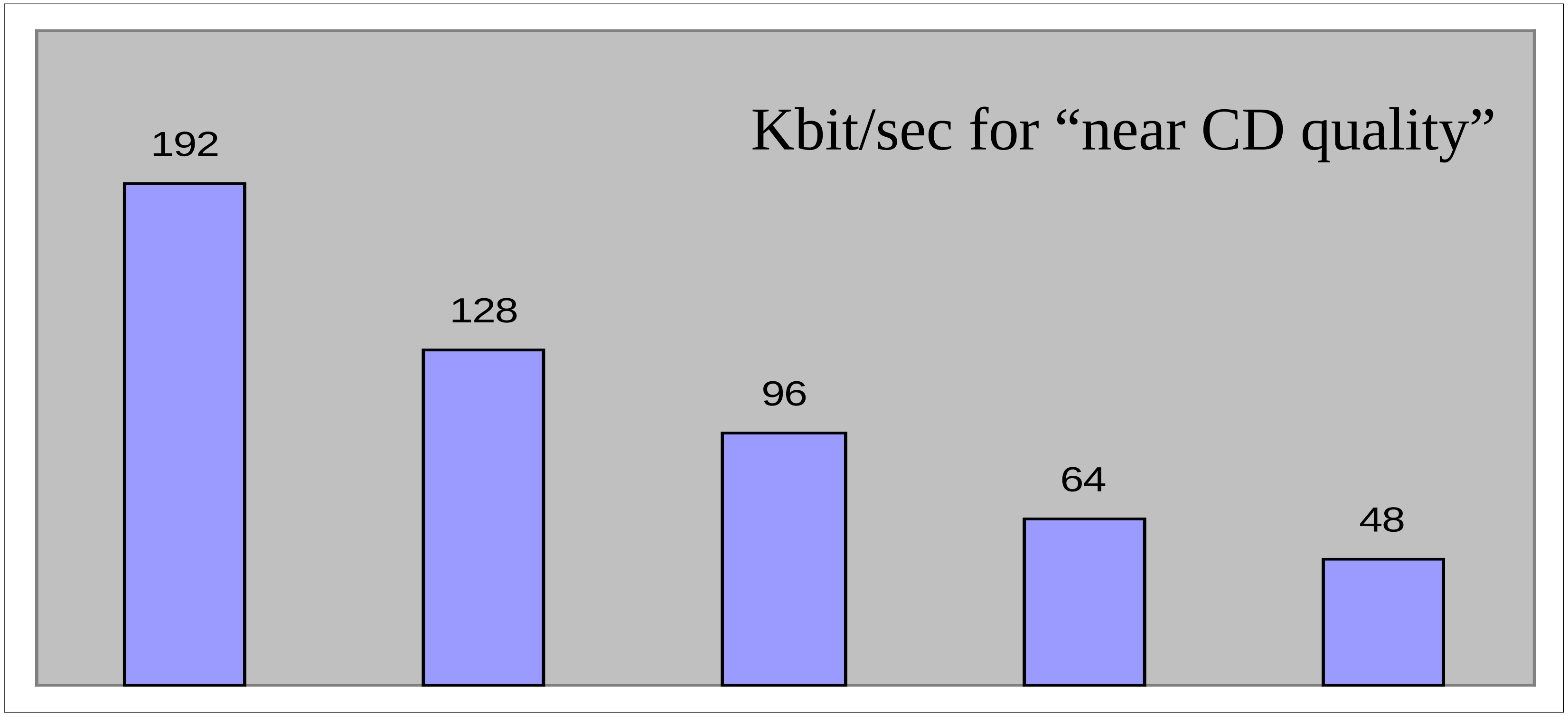
(0)
(1): 56
(3): 80

Derived from: RealNetworks, http://service.real.com/help/library/guides/realone/ProductionGuide/HTML/realpgd.htm?page=htmfiles/audio.htm%23optimize

Bit rate ranges (kbps)

	MIN	MAX
MPEG-1 Layer 3	32	1024
MPEG-2 AAC	<=8	576/chan
AC-3	32	640
MLP	?	9830
DTS	32	3072
PAC	32	1024

Quality improves over time



1992	1994	1997	2000	2001
MPEG-1	MPEG-2	MPEG-2	MPEG-4	WMA
Layer 3	Layer 3	AAC	AAC	Ver. 8

Source: Brandenburg, AES Burlingame; Microsoft press release

Measuring audio

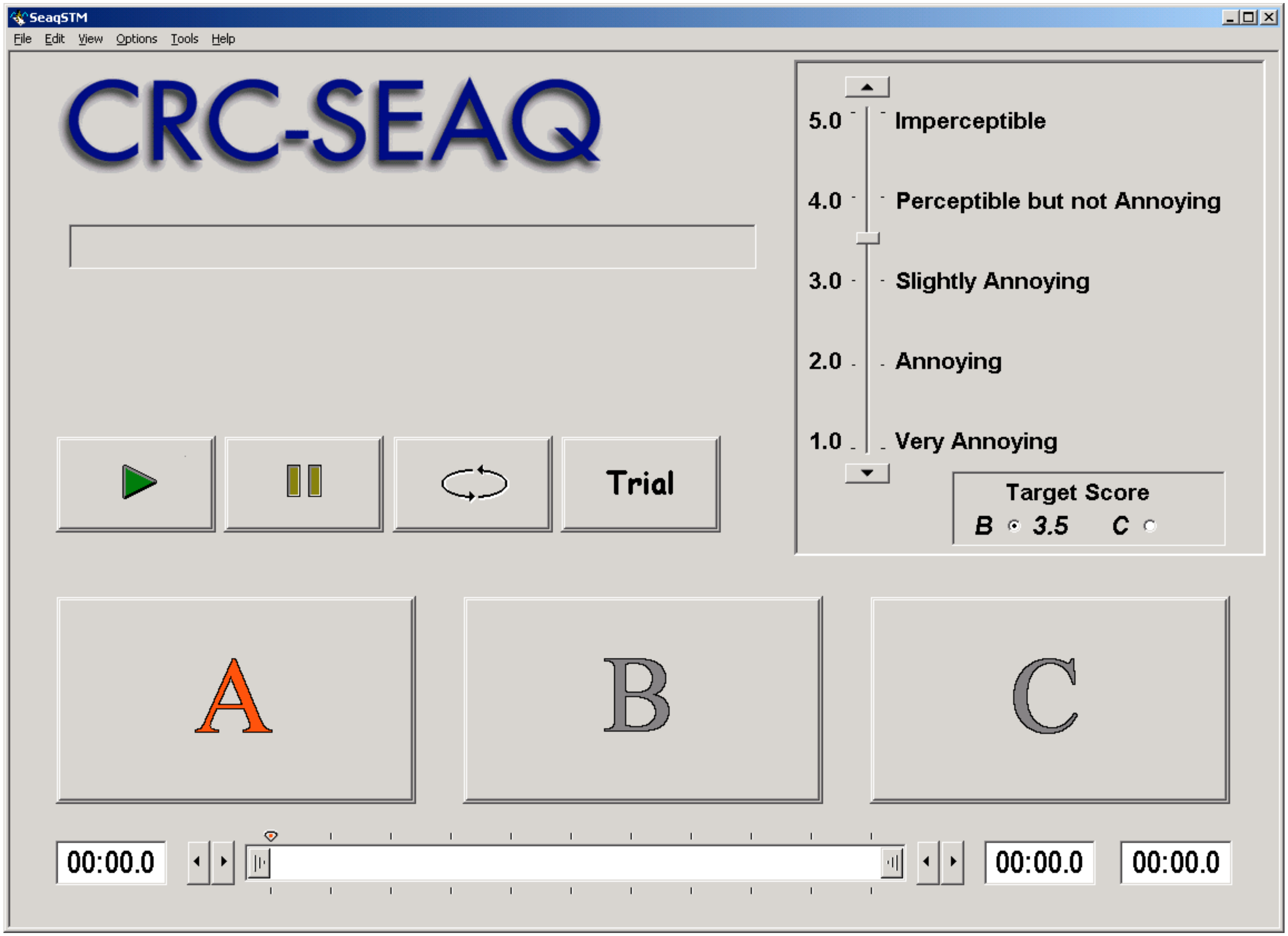
- Uncompressed audio
- Compressed audio

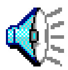
Subjective Compression Evaluation

- ABC listening test: correctly identify A as B or C
  - AAB, ABA
- Rating scale

  0: "difference imperceptible" ... through ...

  -4: "difference very annoying"
- ITU-R SB.1116

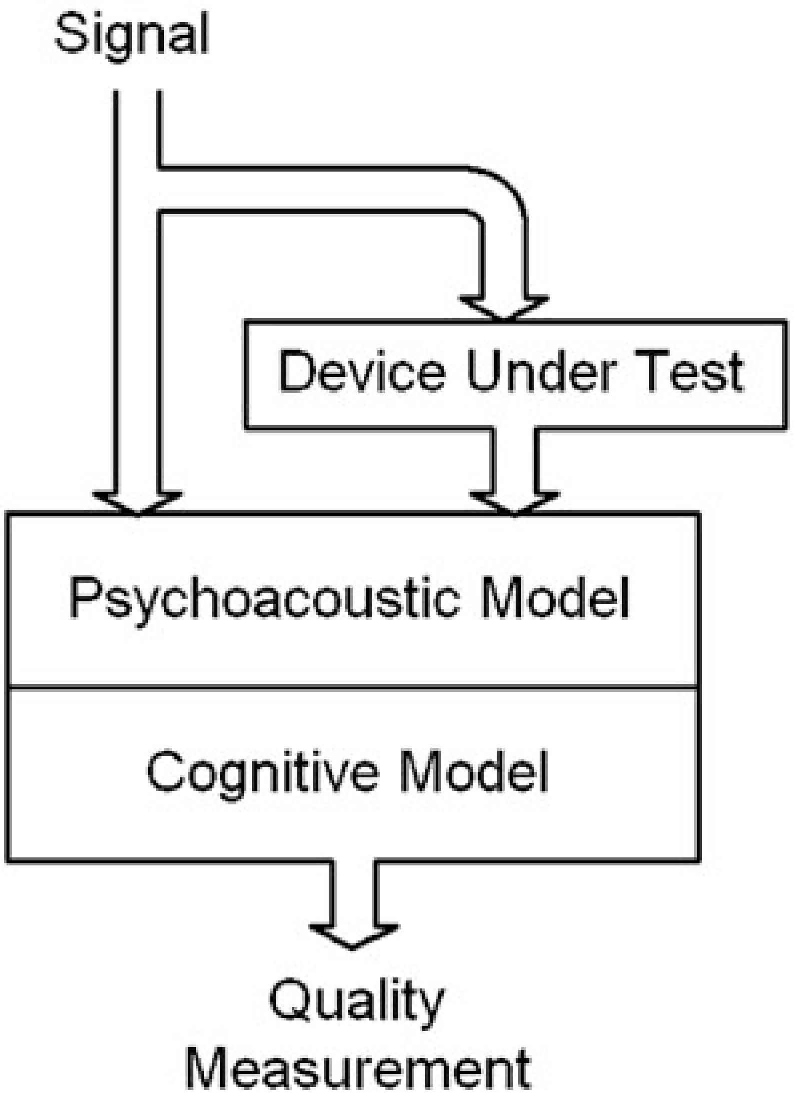




Source: Grant Davidson, Dolby

PEAQ

- Perceptual Evaluation of Audio Quality
- OPTICOM
  - OPERA Voice/Audio Quality Analyzer
- Not:
  - Predictive Equations for Alfalfa Quality



http://www.crc.ca/en/html/aas/home/peaq/peaq

Comparing Codecs

- "XXX was recently rated the best performing audio coding technology in a class of five tested in independent trials by Moulton Laboratories. In this test, XXX at 96 kbit/sec outperformed the MPEG-2 Advanced Audio Coder (AAC). At 96 kbit/sec, XXX also outperformed AAC at 128 kbit/sec based on a repeatable statistical score." (from a press release)

Comparing Codecs

Alg/kbps	ITU score	Characterization	
		Diff perceptible?	Diff “annoying?”
AAC/128 AC-3/192	– 0.5	Yes	Not “annoying”
PAC/160	– 0.8	Yes	“
AAC/96, PAC/128, AC-3/160, MP2/192	– 1.1	Yes	“slightly annoying”
XXX/96, MP3/128, MP2/160	– 1.8	Yes	“
AC3/128, MP2/128	– 2.1	Yes	“annoying”
PAC/64	– 3.0	Yes	“very annoying”

Taken from Soulodre et al., JAES 1998

What we will cover

- The care and starving of birdies.
- Classifying compression techniques.
- Evaluating compression techniques.
- Comparing compression techniques.
- Some typical encoder controls.
- Market forces on compression
- What to try when something goes wrong.

Encoder controls (AC-3) (1)

- Sample rate code (default 0 = 48 kHz)
  - 0 = 48 kHz
  - 1 = 44.1 kHz
  - 2 = 32 kHz
- Audio bandwidth code
  - 2 = 6.80 kHz 10 = 15.80 kHz
  - 3 = 7.92 kHz 11 = 16.92 kHz
  - ...
  - 8 = 13.55 kHz 16 = 22.55 kHz
  - 9 = 14.67 kHz 17 = 23.67 kHz

Encoder controls (AC-3) (2)

- Audio coding mode
  - 0 = 1+1 (L, R)
  - 1 = 1/0 (C)
  - 2 = 2/0 (L, R)
  - 3 = 3/0 (L, C, R)
  - 4 = 2/1 (L, R, l)
  - 5 = 3/1 (L, C, R, l)
  - 6 = 2/2 (L, R, l, r)
  - 7 = 3/2 (L, C, R, l, r)
- Low frequency effects channel on/off

Encoder controls (AC-3) (3)

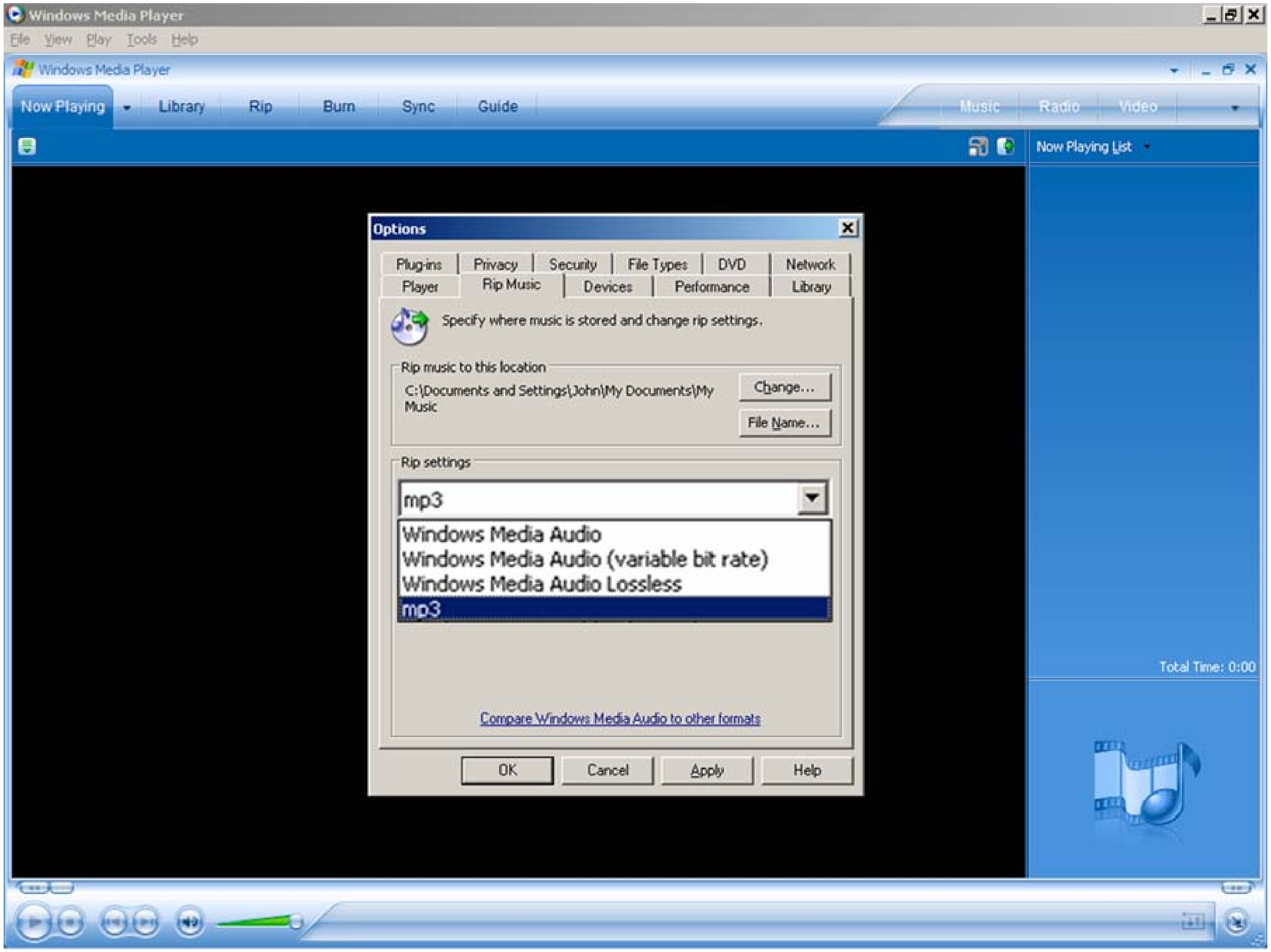
- Coupling channel on/off
- Coupling begin frequency code
  - 0 = 3.42 kHz 8 = 12.42 kHz
  - 1 = 4.55 kHz 9 = 13.55 kHz
  - 2 = 5.67 kHz 10 = 14.67 kHz
  - 3 = 6.80 kHz 11 = 15.80 kHz
  - 4 = 7.92 kHz 12 = 16.92 kHz
  - 5 = 9.05 kHz 13 = 18.05 kHz
  - 6 = 10.17 kHz 14 = 19.17 kHz
  - 7 = 11.30 kHz 15 = 20.30 kHz

Encoder controls (AC-3) (4)

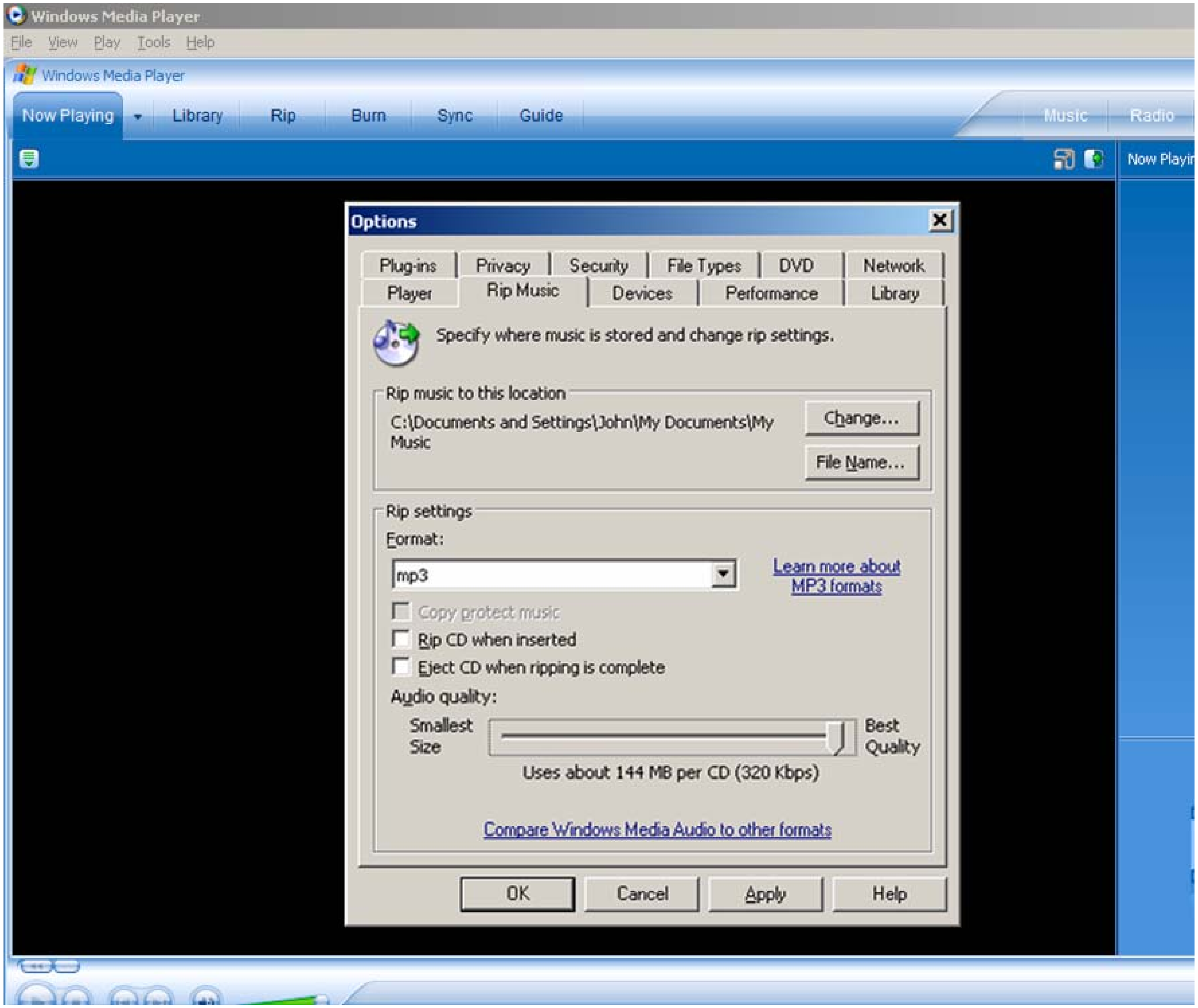
- Frame size code

– 0 = 32 kbps	7 = 112 kbps	14 = 384 kbps
– 1 = 40 kbps	8 = 128 kbps	15 = 448 kbps
– 2 = 48 kbps	9 = 160 kbps	16 = 512 kbps
– 3 = 56 kbps	10 = 192 kbps	17 = 576 kbps
– 4 = 64 kbps	11 = 224 kbps	18 = 640 kbps
– 5 = 80 kbps	12 = 256 kbps	
– 6 = 96 kbps	13 = 320 kbps	

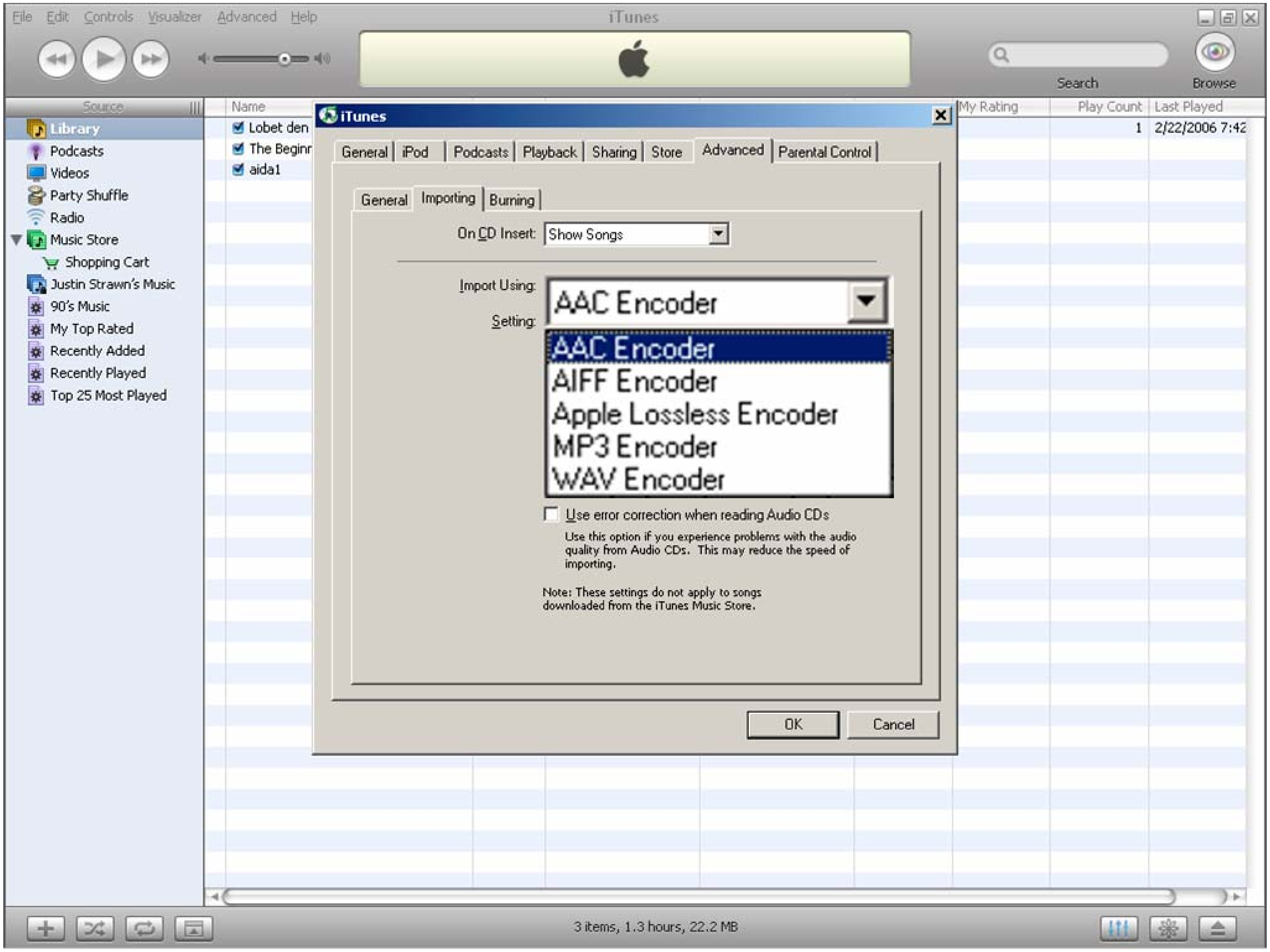
Encoder Controls: WMA (1)



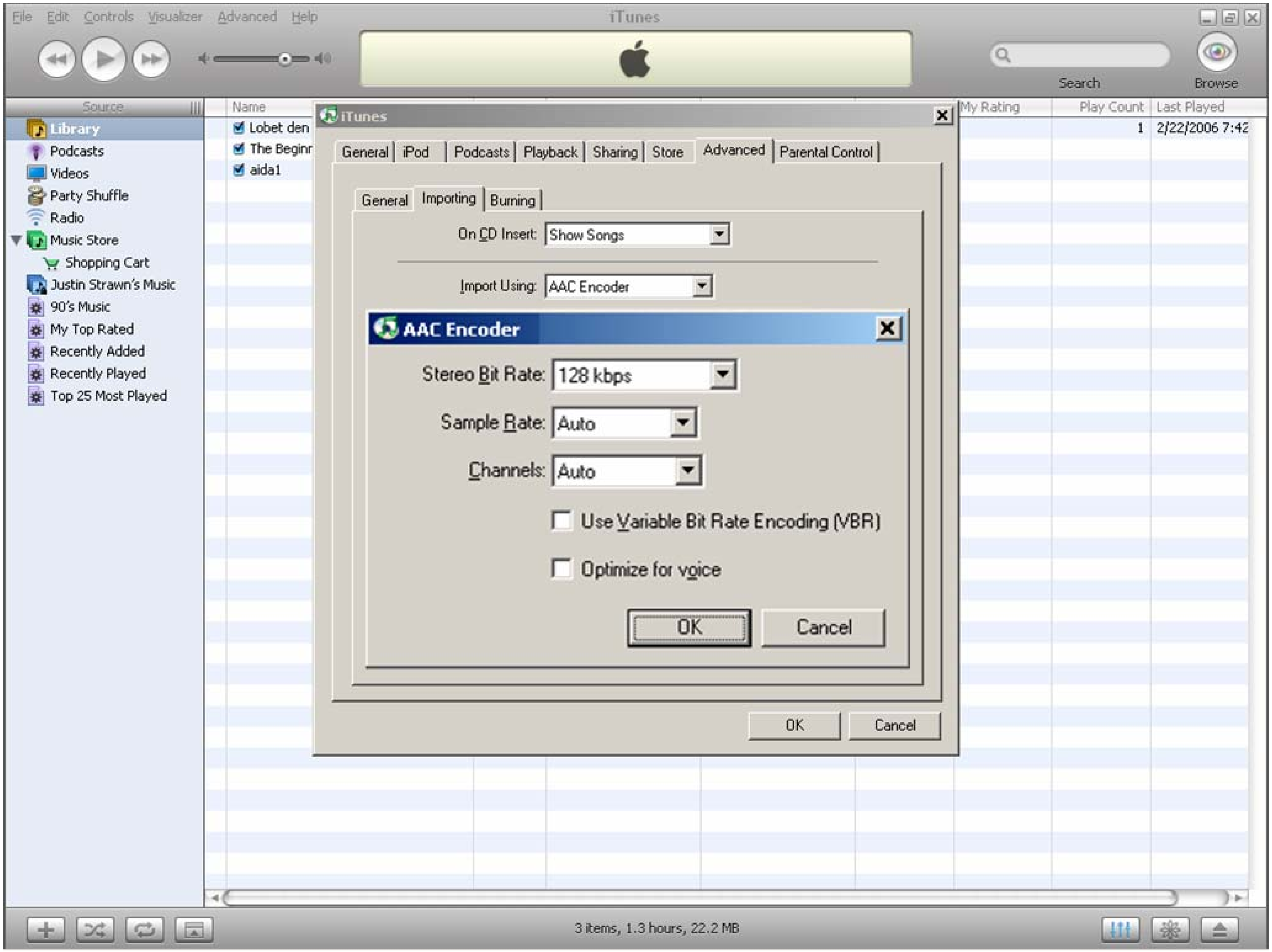
Encoder Controls: WMA (2)



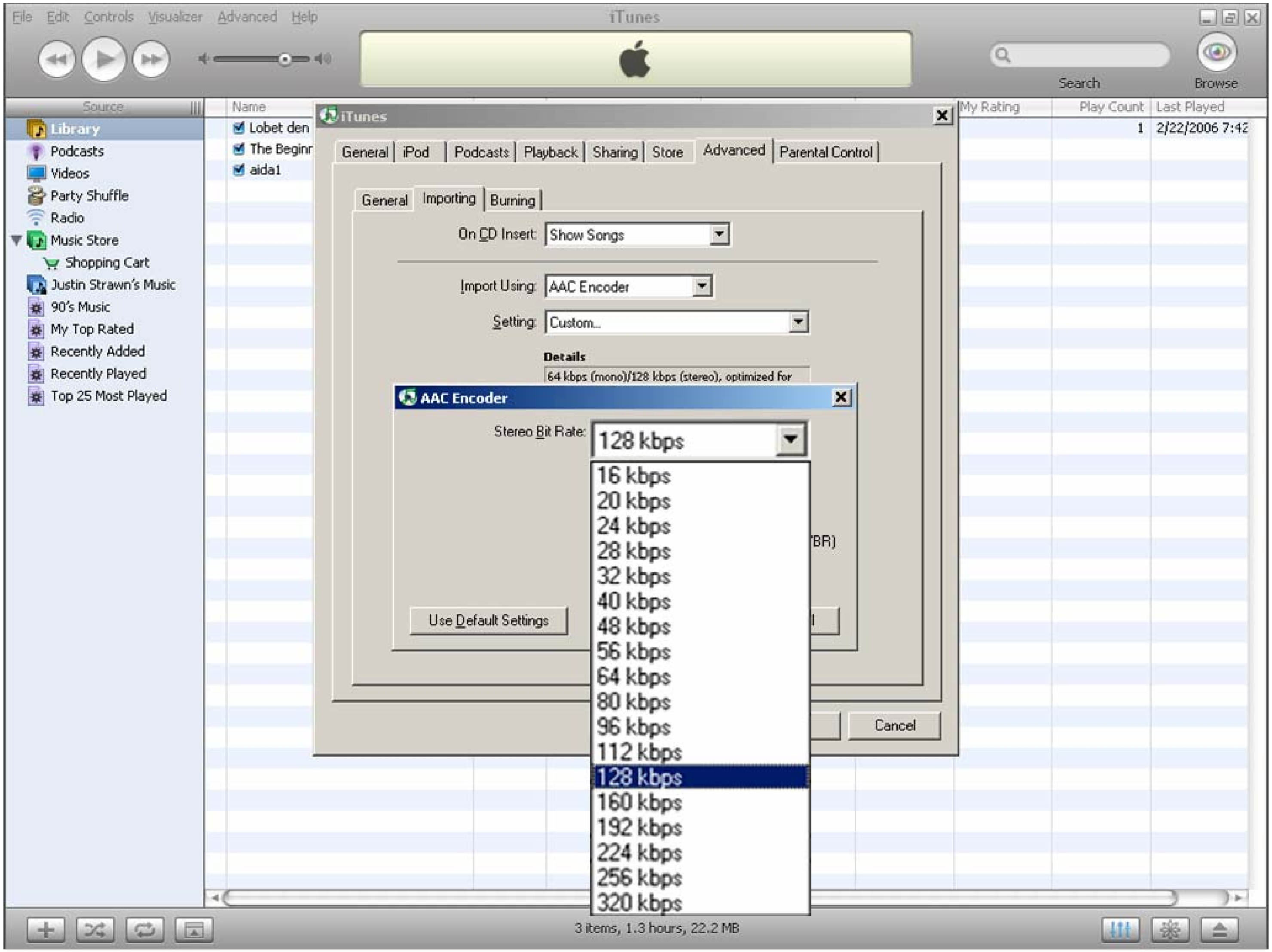
Encoder Controls: iTunes



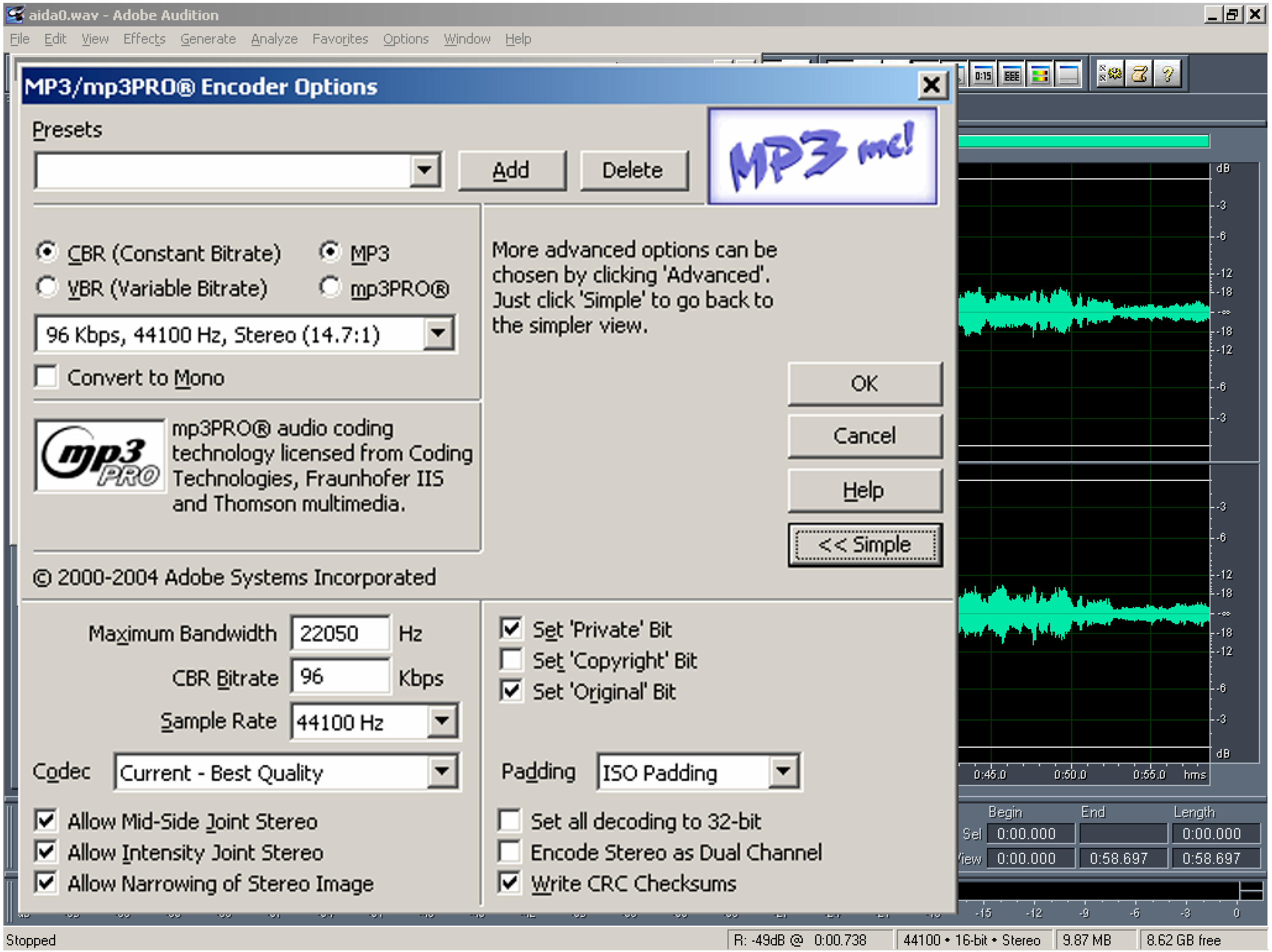
Encoder Controls: iTunes AAC (1)



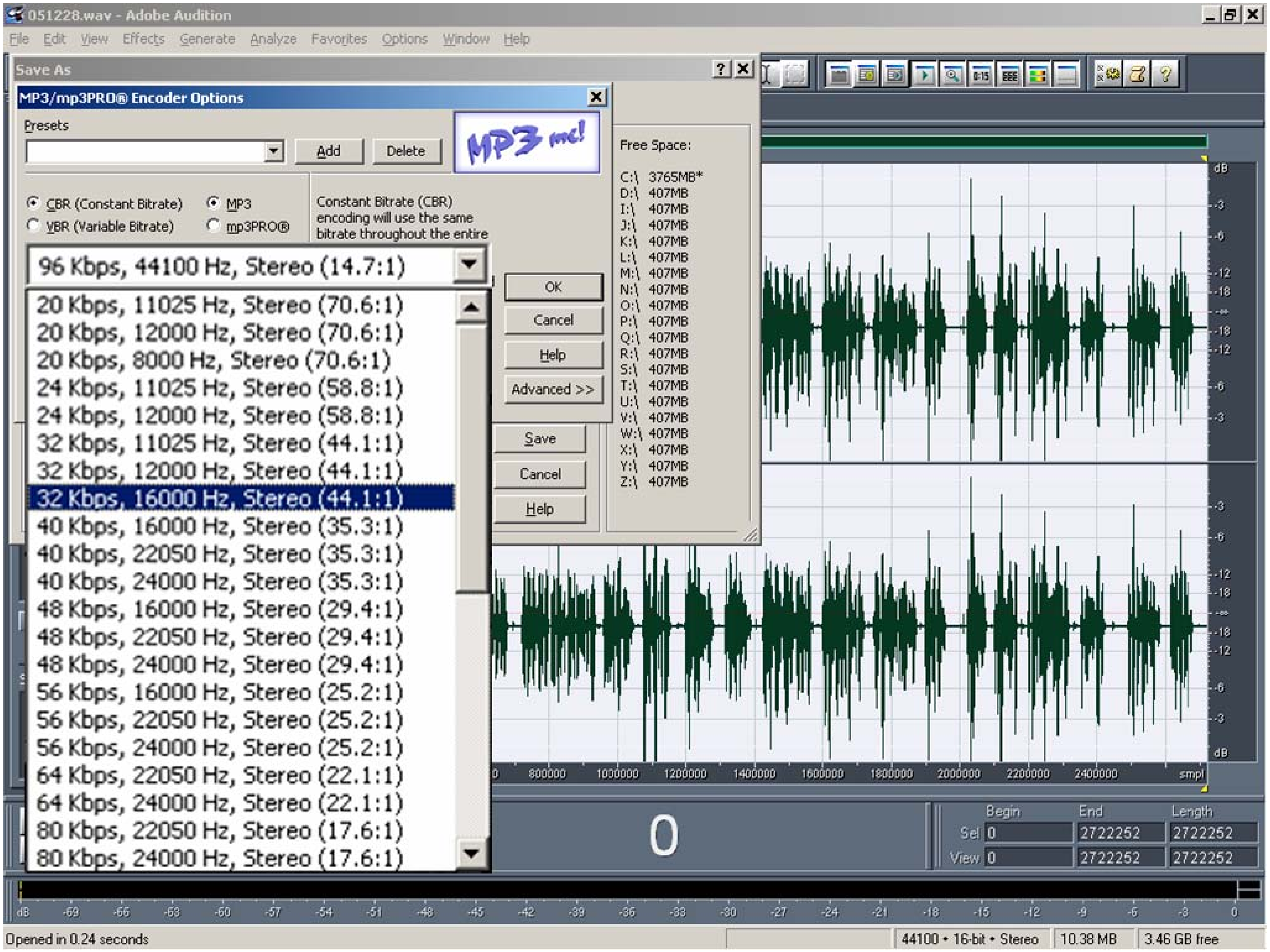
Encoder Controls: iTunes AAC (2)



Encoder Controls: Audition (1)



Encoder Controls: Audition (2)



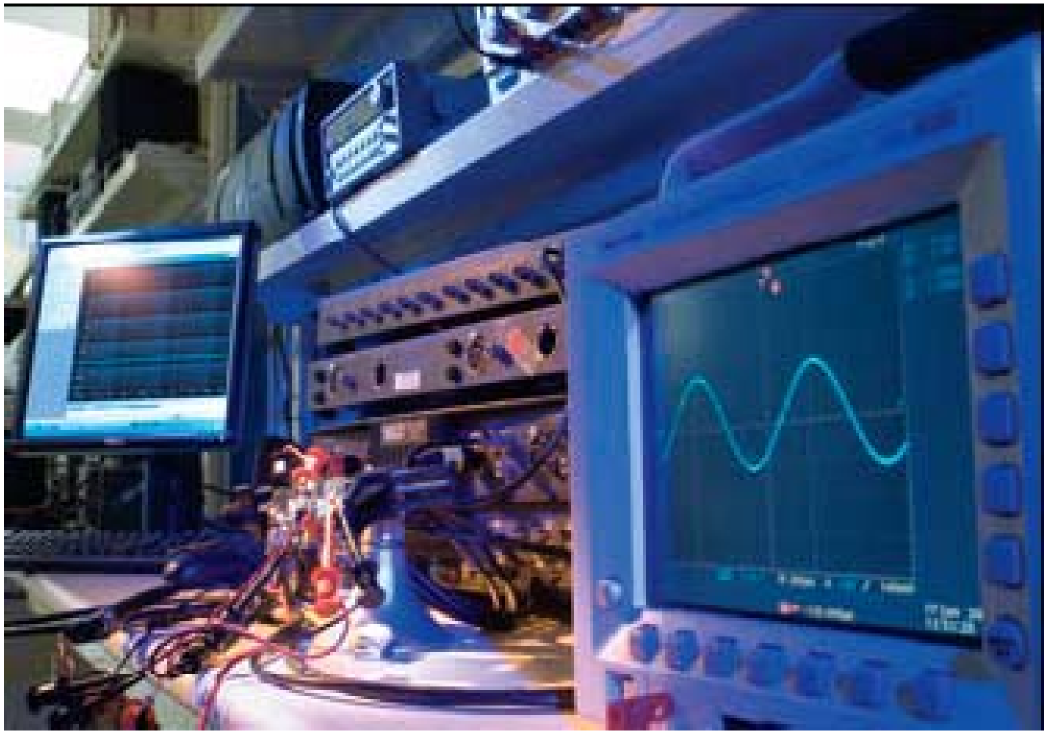
What we will cover

- The care and starving of birdies.
- Classifying compression techniques.
- Evaluating compression techniques.
- Comparing compression techniques.
- Some typical encoder controls.
- Market forces on compression
- What to try when something goes awry.

Market forces on compression

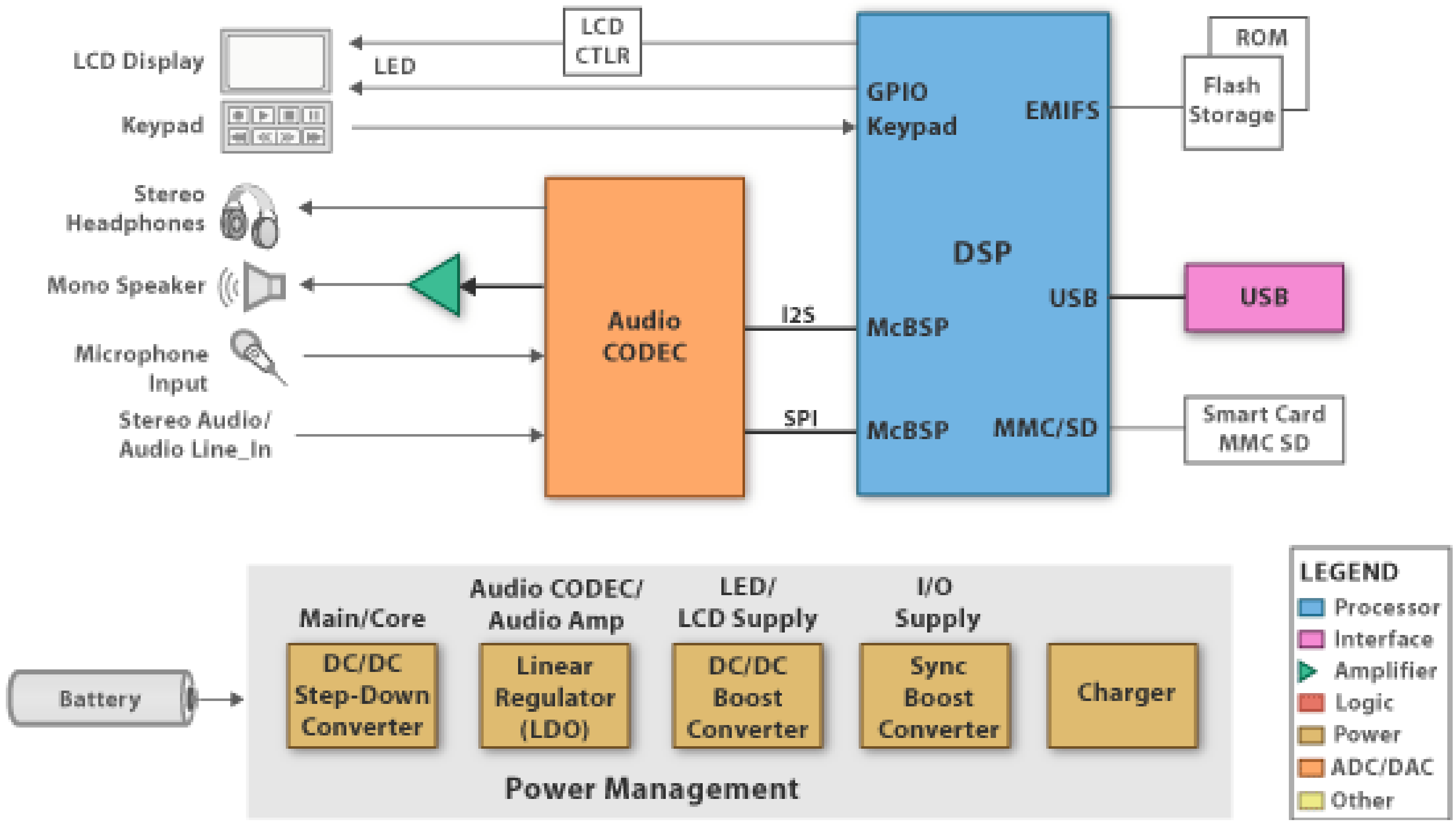
- As good as CD?
- HD
- Commoditization: music is the accessory?
- Manufacturability
- Licensing
- Patent lawsuits

Manufacturability



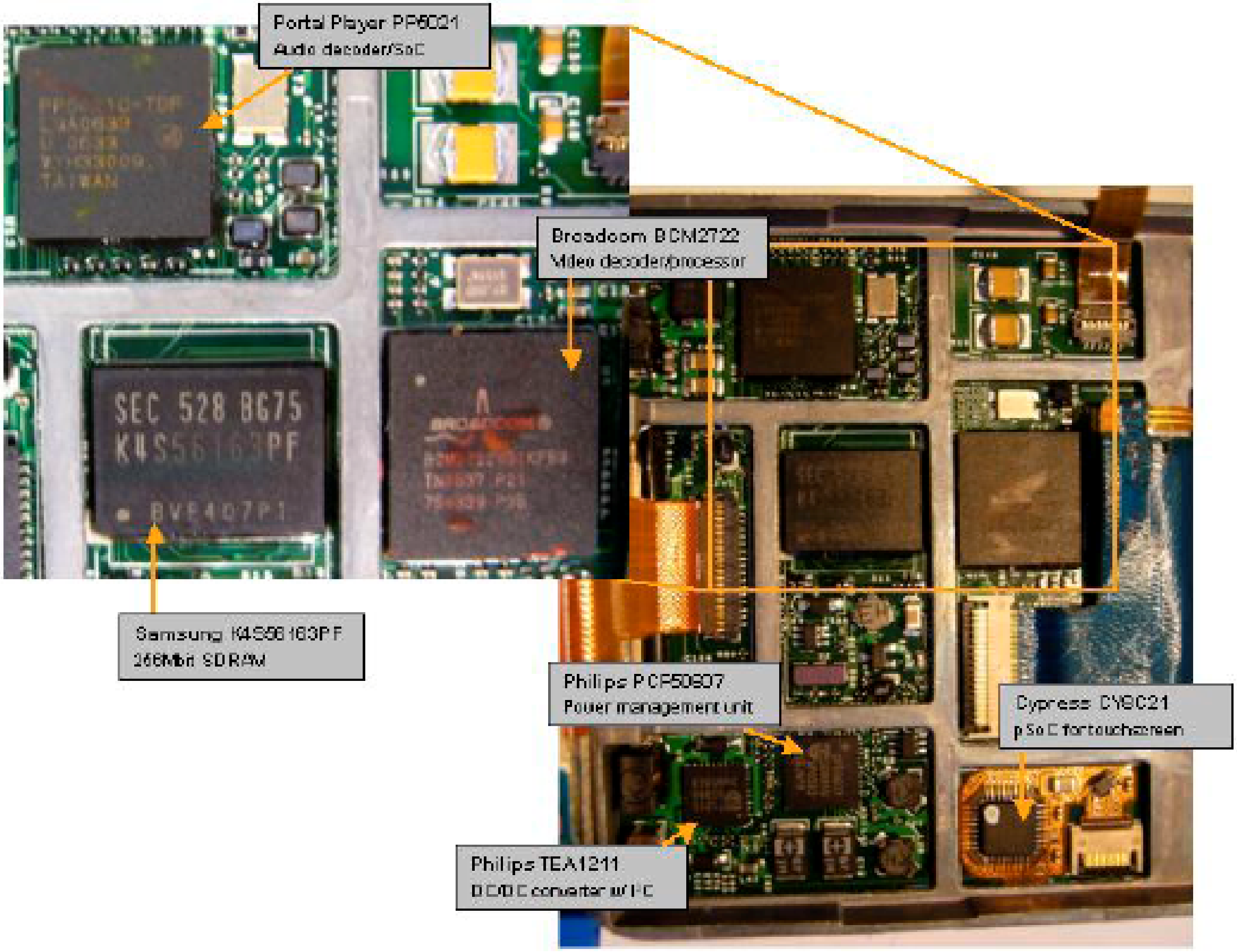
"Power this: testing audio ICs"
EDN, March 1, 2007
http://www.edn.com/index.asp?layout=article&articleid=CA6418209&text=%2C+Class+D?text=power+this

Inside an MP3 player



http://focus.ti.com/docs/solution/folders/print/12.html; see also http://focus.ti.com/docs/solution/folders/print/267.html

iPod Video 5G Autopsy



https://jefferies.bluematrix.com/docs/pdf/31086.pdf; http://www.ipodhacks.com/article.php?sid=1583

Minty

http://www.ladyada.net/make/minty/

Compression: the Future

•

What we will cover

- The care and starving of birdies.
- Classifying compression techniques.
- Evaluating compression techniques.
- Comparing compression techniques.
- Some typical encoder controls.
- Market forces on compression
- What to try when something goes awry.

When (not) to use compression

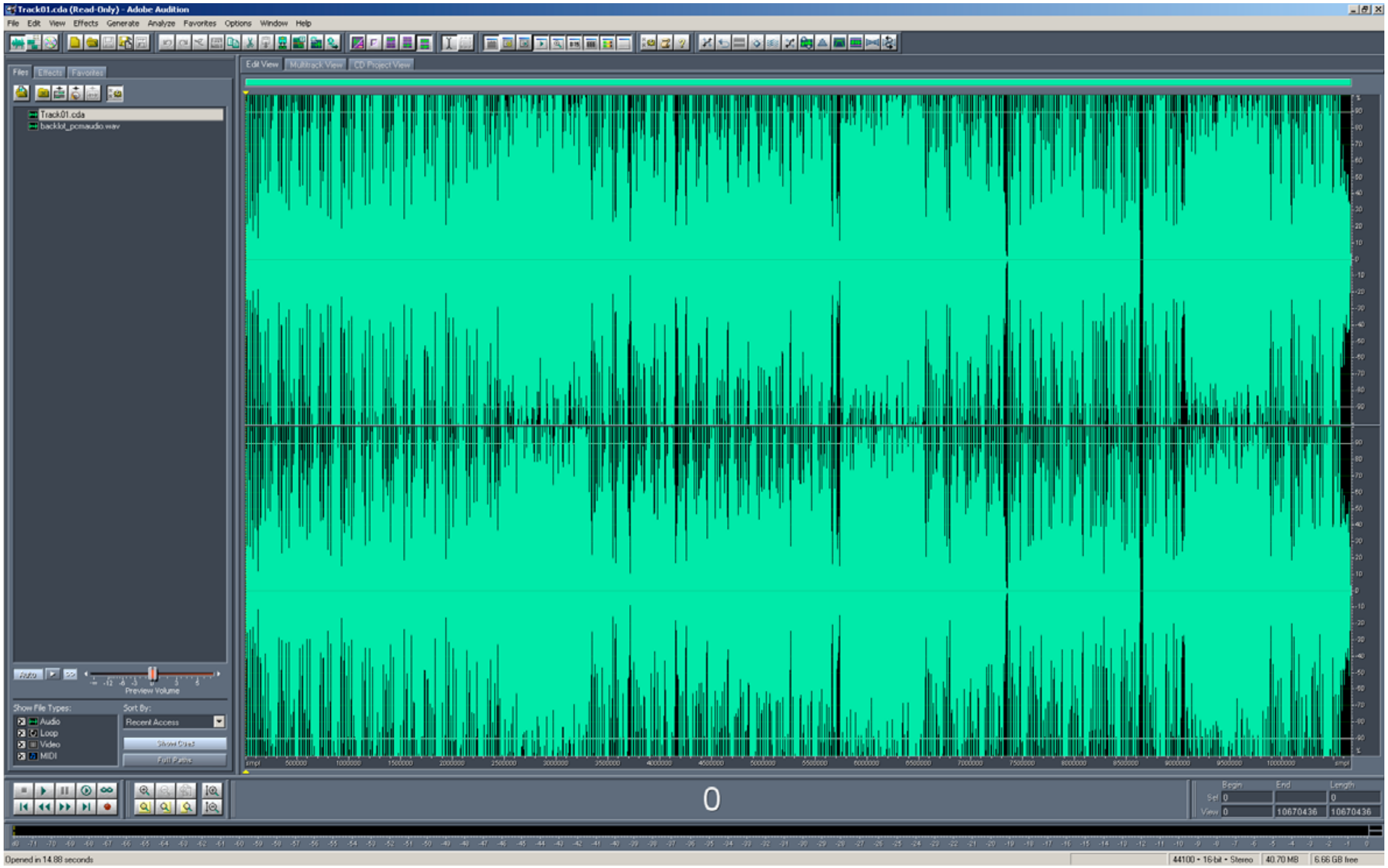
What can go wrong with compression?

•

How do we fix it? (1)

- Highest quality in original (before encode).
- Encoder/Decoder implemented well?
- Level: close to full scale as possible.
- Avoid Wall of Sound
- Low-pass filter?
- No DC?

(Wall of Sound)



Justin Timberlake, *FutureSex/LoveSounds*, Track 1

How do we fix it? (2)

- Study encoder settings / switches.
- Raise bit rate.
- Reduce the stereo separation.
- Experiment with M/S vs I/S
- Experiment with constant versus variable bit rate.

How do we fix it? (3)

- Alter mix to minimize problems.
- Avoid heavy limiting/compression
- Decoded output is pure noise? Try swap bytes.

Bonus Question

- Encode/decode full scale input
- “Mp3 player” output clips
- Why?
- How to fix?

What we have covered

- The origin and extermination of birdies.
- Classifying compression techniques.
- Evaluating compression techniques.
- Comparing compression techniques.
- Some typical encoder controls.
- Market forces on compression
- What to try when something goes wrong.

What did you want to cover?

What do you want to cover?

- Lossless compression: how is it lossless?
- Newer surround formats, Dolby True-HD, DTS Master Audio
- effects of kbps, how does that affect audio. At what point does it still have an effect?
- Is compression here to stay? On its way out due to storage, Blu-ray?

What do you want to cover?

- General practices: how to mix for an MP3 release
- How far can compression go? How much compression can you achieve while having great quality?
- What are more widely used formats, what is on its way out, what's on its way in?

What do you want to cover?

- Encoding / Decoding?
- History of coding; how did this come about?
- Perceptual coding

What do you want to cover?

- Audio streaming, starting from the beginning
- How important is the algorithm and the encoder to get the same results? Does LAME sound better than other encoders?

Meeting the challenge

- Coarser Quantization (time domain)
- DPCM, ADPCM
- Linear Prediction
- Subband coding
- Transform to frequency domain
- Coarser quantization (frequency domain)
- Psychoacoustics: mask the noise
- Variable bit rate
- Noiseless coding
- Window
- Temporal masking
- Error recovery
- Multichannel redundancy
- Survive in Marketplace

Look ahead: Sunday a.m.

- Discussion 9:00-10:30
  - Modification of student projects
- RoundTable, 10:30-12:00

Roundtable Participants

- Justin Davis (Rocky Mountain Recorders)
- Lorne Bregitzer (Colorado Sound; UCD)
- Rich Sanders (UCD)
- Leslie Gaston (UCD)

For Sunday Morning

- Meet by 9:00 in D, F, & J: Examples TBD
- Take notes and discuss among yourselves:
  - What do you hear?
  - How are they different?
  - What do you like?
  - What don’t you like?
- If you have time, go to another room.
- Back here at 9:30 to discuss.

Notes, 1st Recording

Notes, 2nd Recording

Notes, 3rd Recording

MSRA 5500 - 002
MUS 4500 - 002

March 16, Session 5

Discussion of Recordings Roundtable

© Copyright 2008 John Strawn

What we will cover

- Discussion 9:00-10:30
  - Questions from last 2 days
  - Discuss modifications of three student projects
- RoundTable, 10:30-12:00

Discuss Recordings

- 1
- 2
- 3

Tracks You Heard

- 1 Jen Nives, Beside Myself
- 2 David Vorhees, Viking Seminar
- 3 Bard Helzer, Atlas Rough

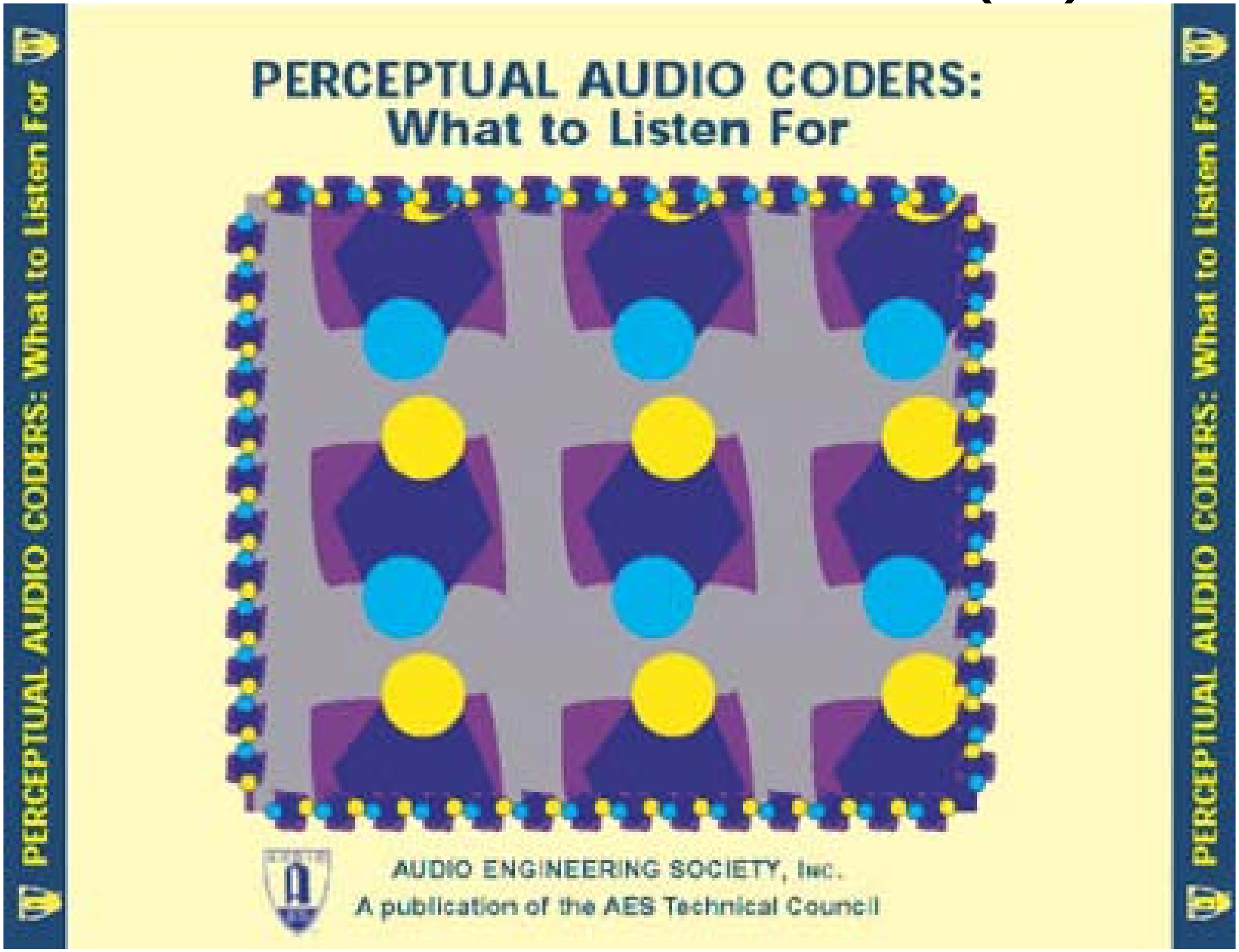
How tracks were modified

- 1 AAC (not Plus) (32 kHz picked by encoder, 48 kpbs)
- 2 AACPlus (32 kHz, 48 kpbs)
- 3 MP3 (Lame), variable bit rate (32->56 kbps)
- 4 original
- 5 MP3 (Lame), constant bit rate (48 kpbs)

More information: handouts

- Bibliography
- List of URLs
- List of standards
- PDFs of slides available today --- memory stick

AES TC CD-ROM (1)



AES TC SP CD-ROM (2)

- AES Technical Committee on Signal Processing digital audio education CD
- To be released in the near future
- To include demos of
  - masking
- Watch http://www.aes.org/technical/

Roundtable Participants

- Justin Davis (Rocky Mountain Recorders)
- Lorne Bregitzer (Colorado Sound; UCD)
- Rich Sanders (UCD)
- Leslie Gaston (UCD)