Tutorial T4

Audio Compression

120th Convention
Audio Engineering Society
May 20, 2006

John Strawn, Ph.D.

S Systems Inc.

jstrawn@s-systems-inc.com

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

What do you hear?

- **Original and 32 kpbs. Lame encoder, Madplay decoder.**

- **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.**

- **Lou Reid's album *Lou Reid and Carolina,* opening (?) of "God loves his children." Rebel Records, 1966. From http://lame.sourceforge.net/gpsycho/.**

What we will cover (I)

- Inside a basic perceptual <u>en</u>coder
- Inside a basic perceptual <u>de</u>coder
- Masking (simultaneous, temporal)
- Windowing
- Transform (Discrete Cosine)
- Codec families

What we will cover (II): selected topics

- Birdies
- Tandem coding
- Classifying codecs
- Evaluating codecs

What we will not cover today

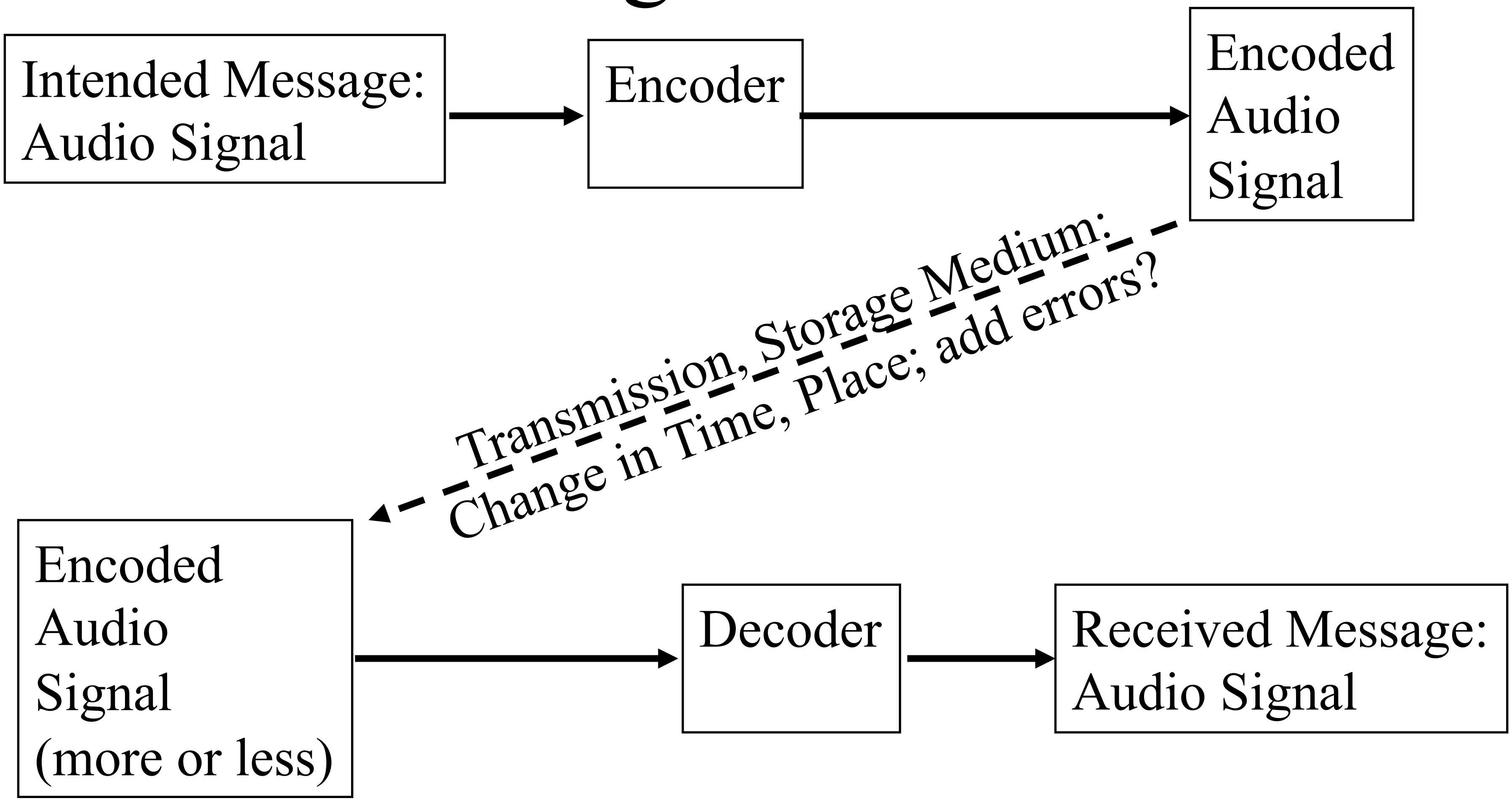
- Stereo, M/S, intensity stereo
- Spatial/Surround (Paper Session P)
- Variable vs fixed bit rate; Scalable
- Critical bands
- Speech compression
- Linear Prediction
- Vector quantization
- QMF filter bank
- Rate distortion function
- Temporal Noise Shaping
- Coupling channel
- ...

Introduction

Model of Communication

Based on Colin Cherry: *On Human Communication*

Coding in General



“Codec”: [en]**COD**er + **DEC**oder

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.

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

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.

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.

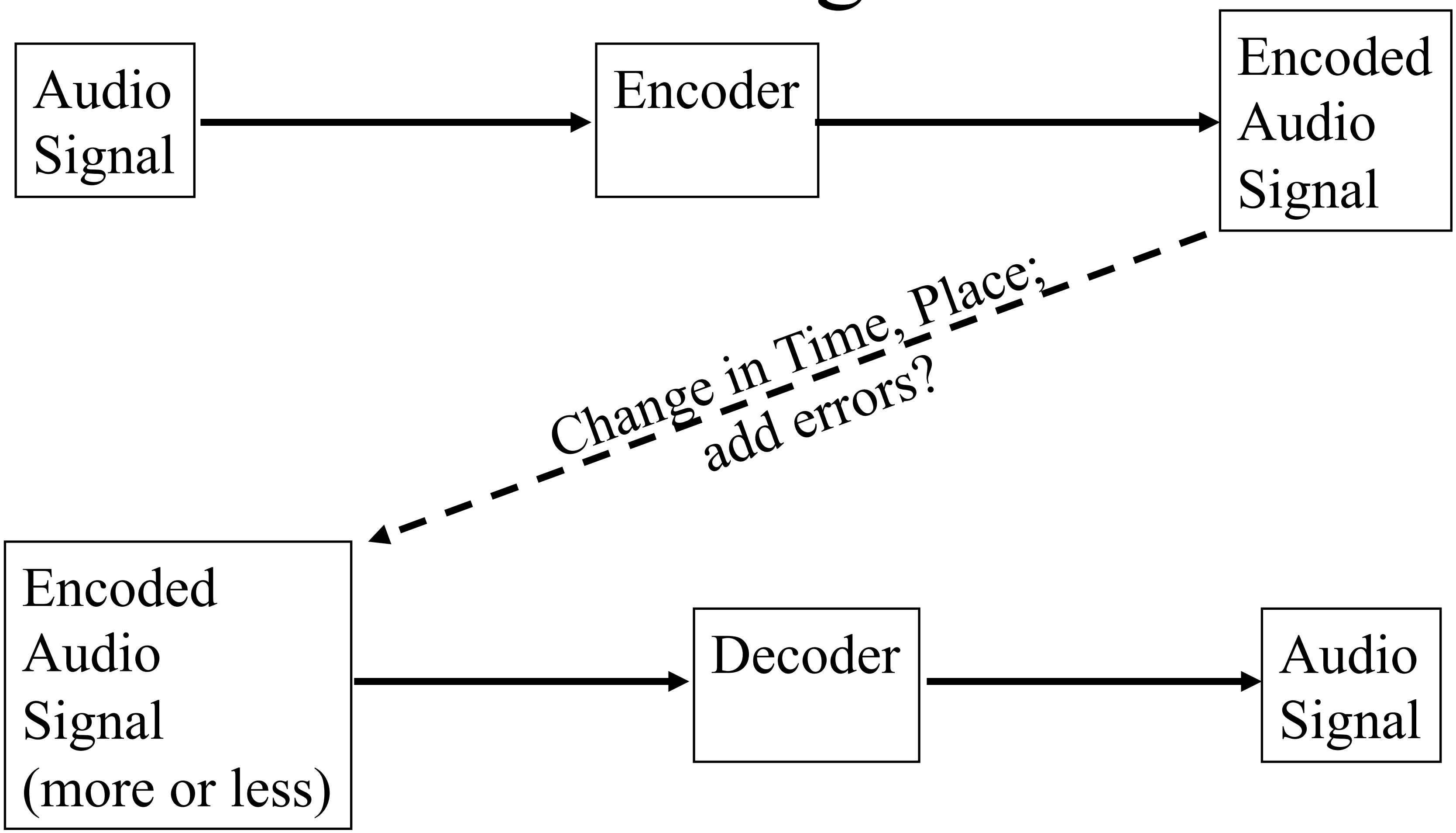
Two broad classes of coders

- Non-perceptual
- Perceptual (based in part on psychoacoustics)

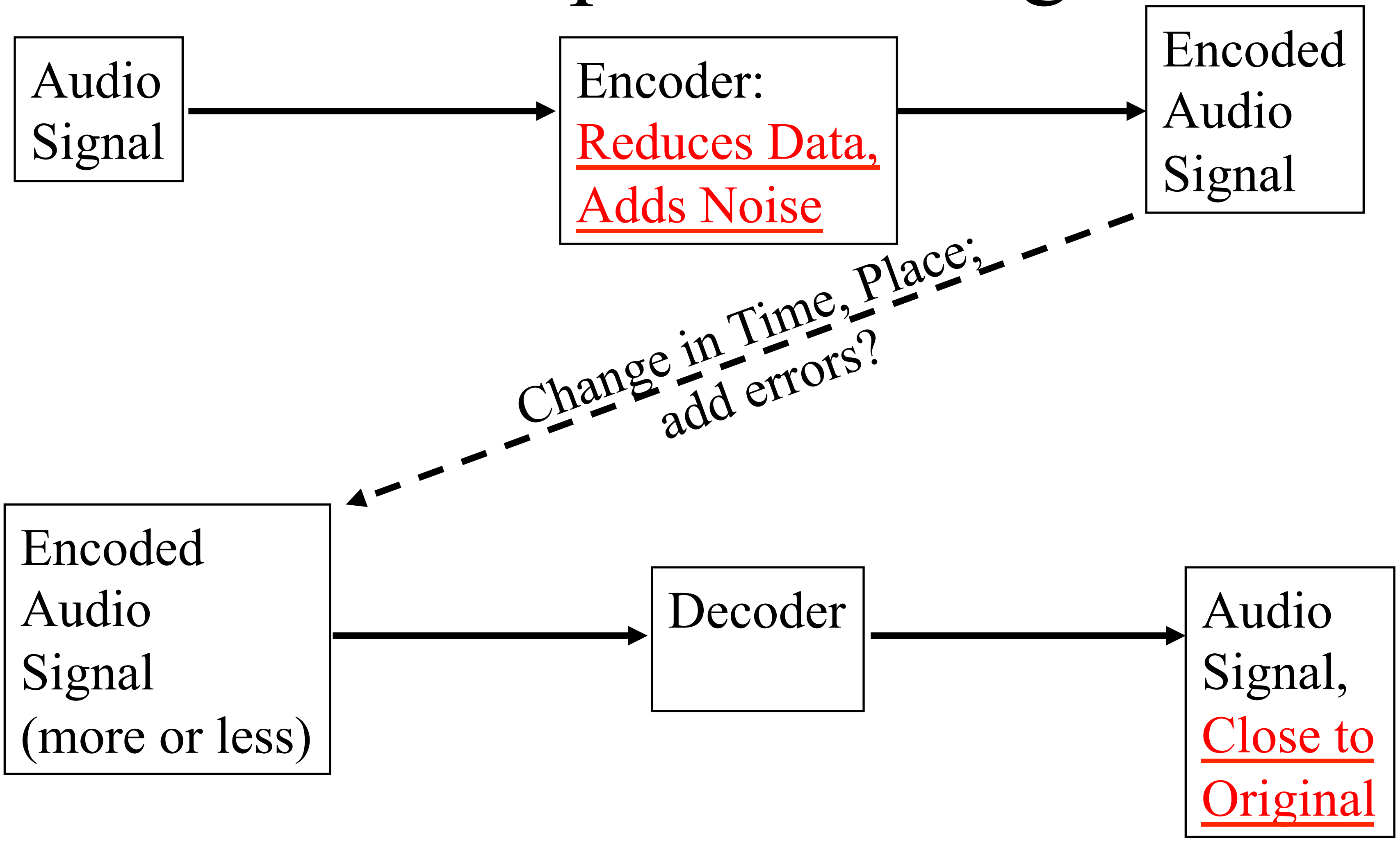
Fundamentals of perceptual coding

- Leave out
  - Irrelevant
  - Redundant
- Allow noise
  - But only in special places

Review: Coding in General



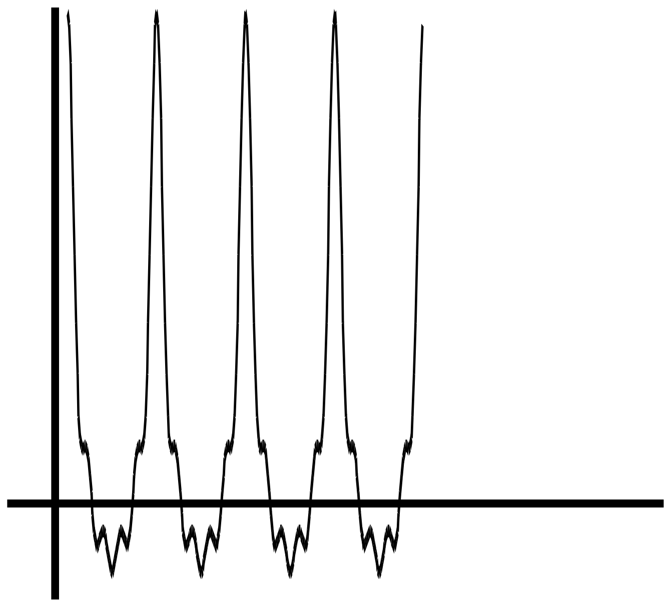
Perceptual Coding



Masking

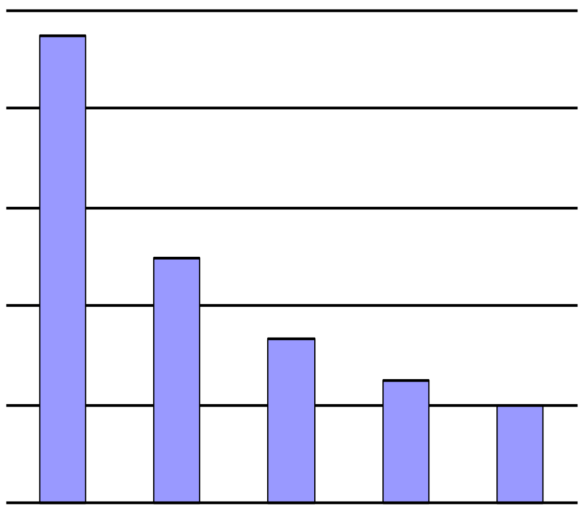
Domains

Waveform



Time Domain

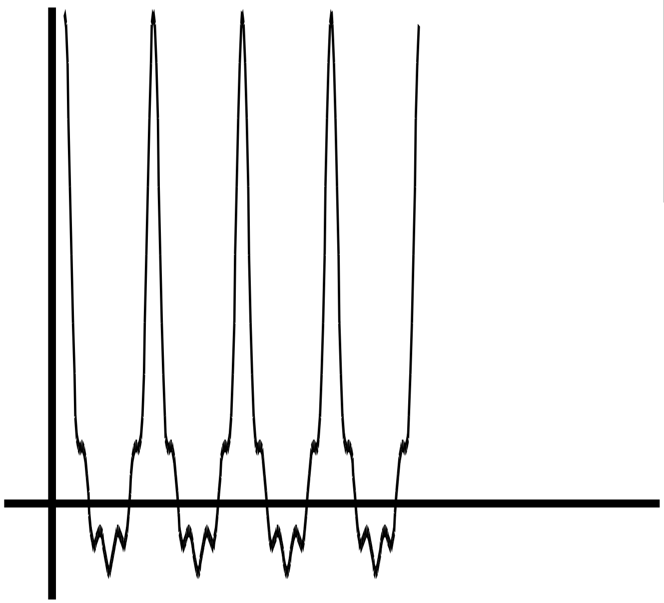
Spectrum



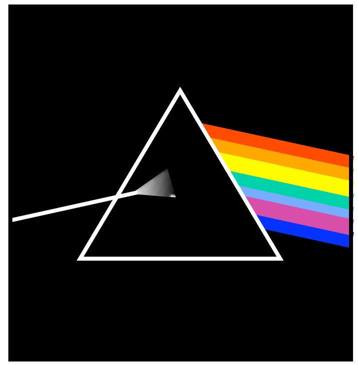
Frequency Domain

Transform

Waveform



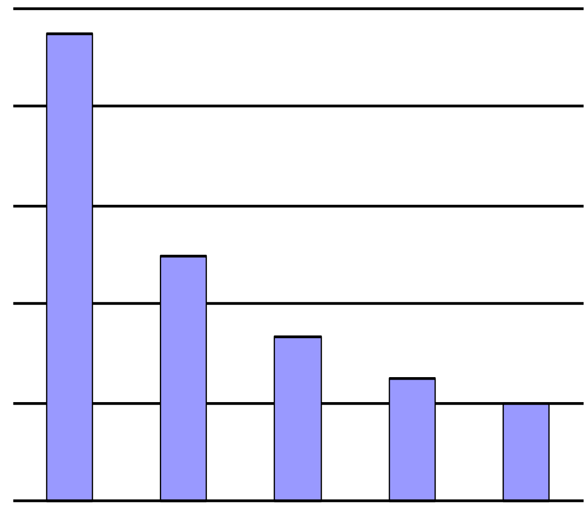
Time Domain



Forward (analyze)

Inverse (Synthesize)

Spectrum



Frequency Domain

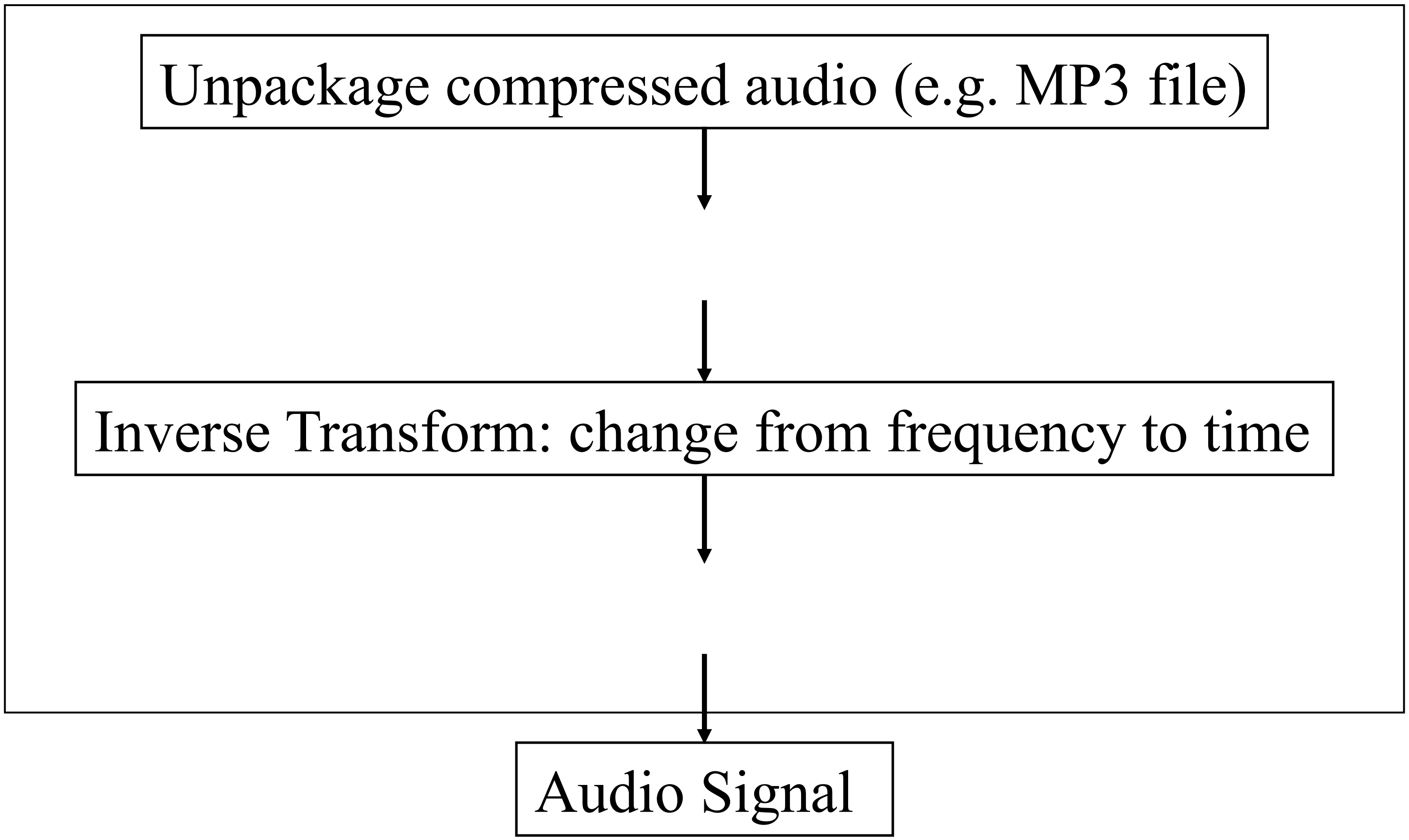
Perceptual Encoder

Audio Signal

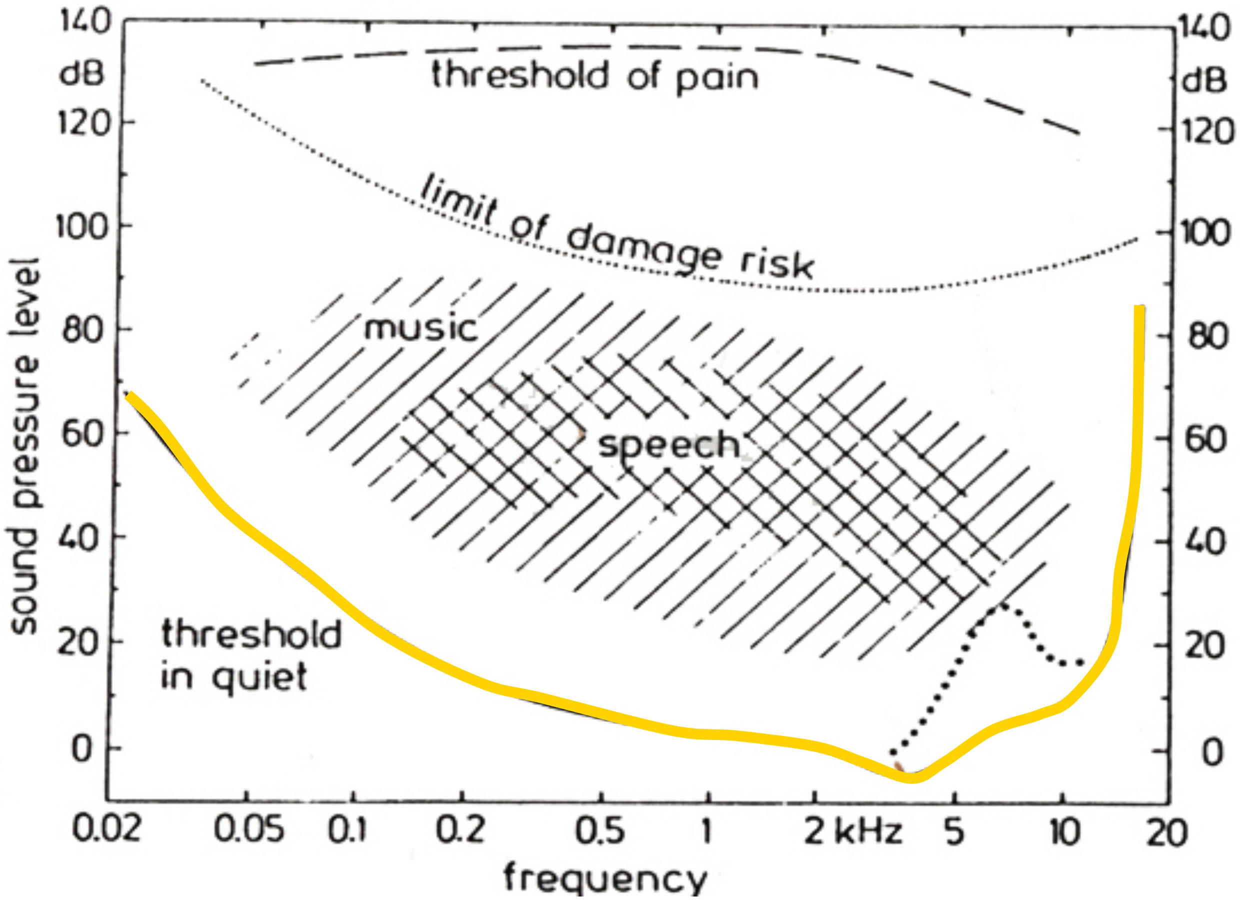
Forward Transform (frequency domain)

Package compressed output (e.g. MP3 file)

Decoder (“mp3 player”)

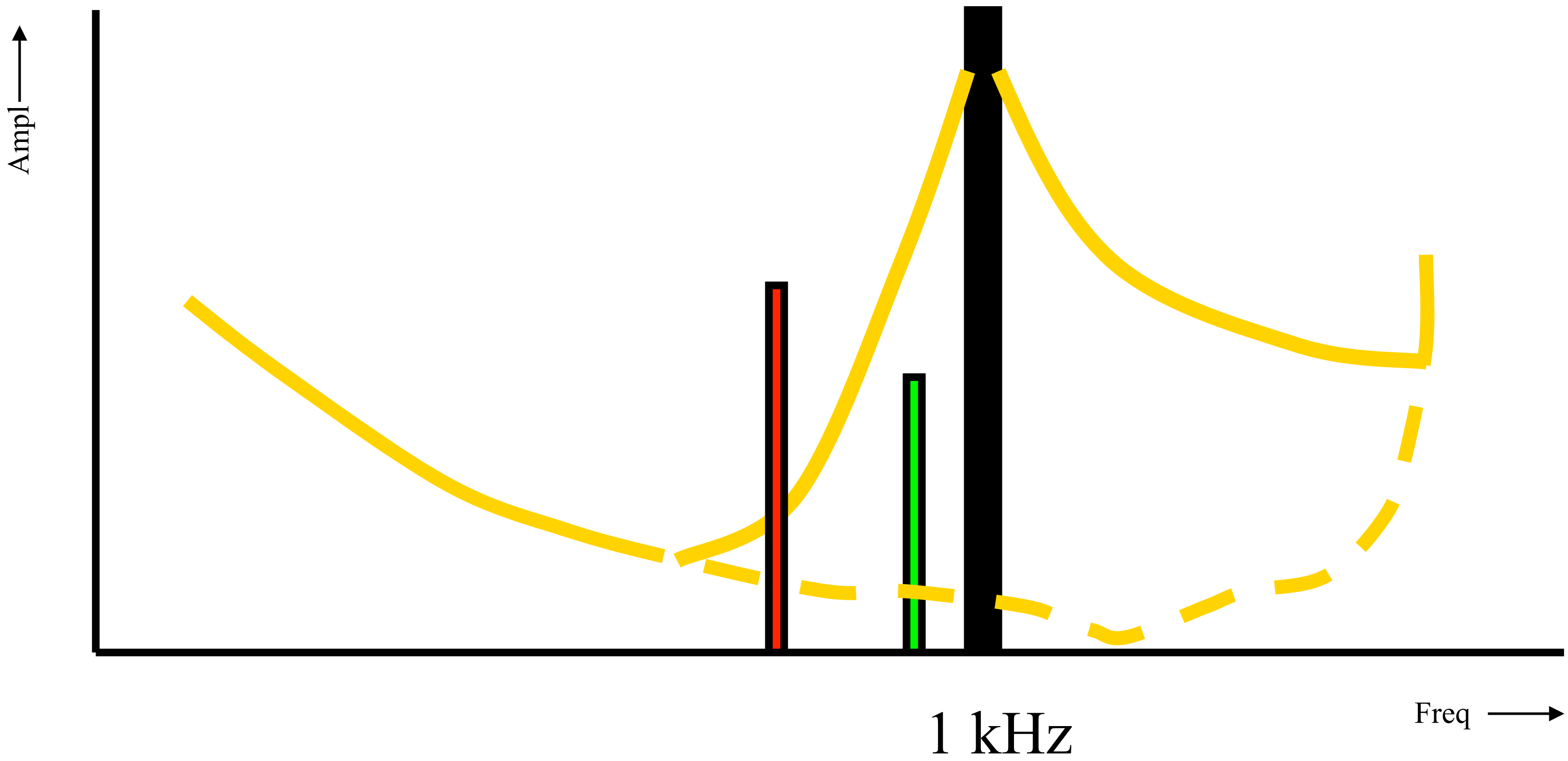


Range of human hearing



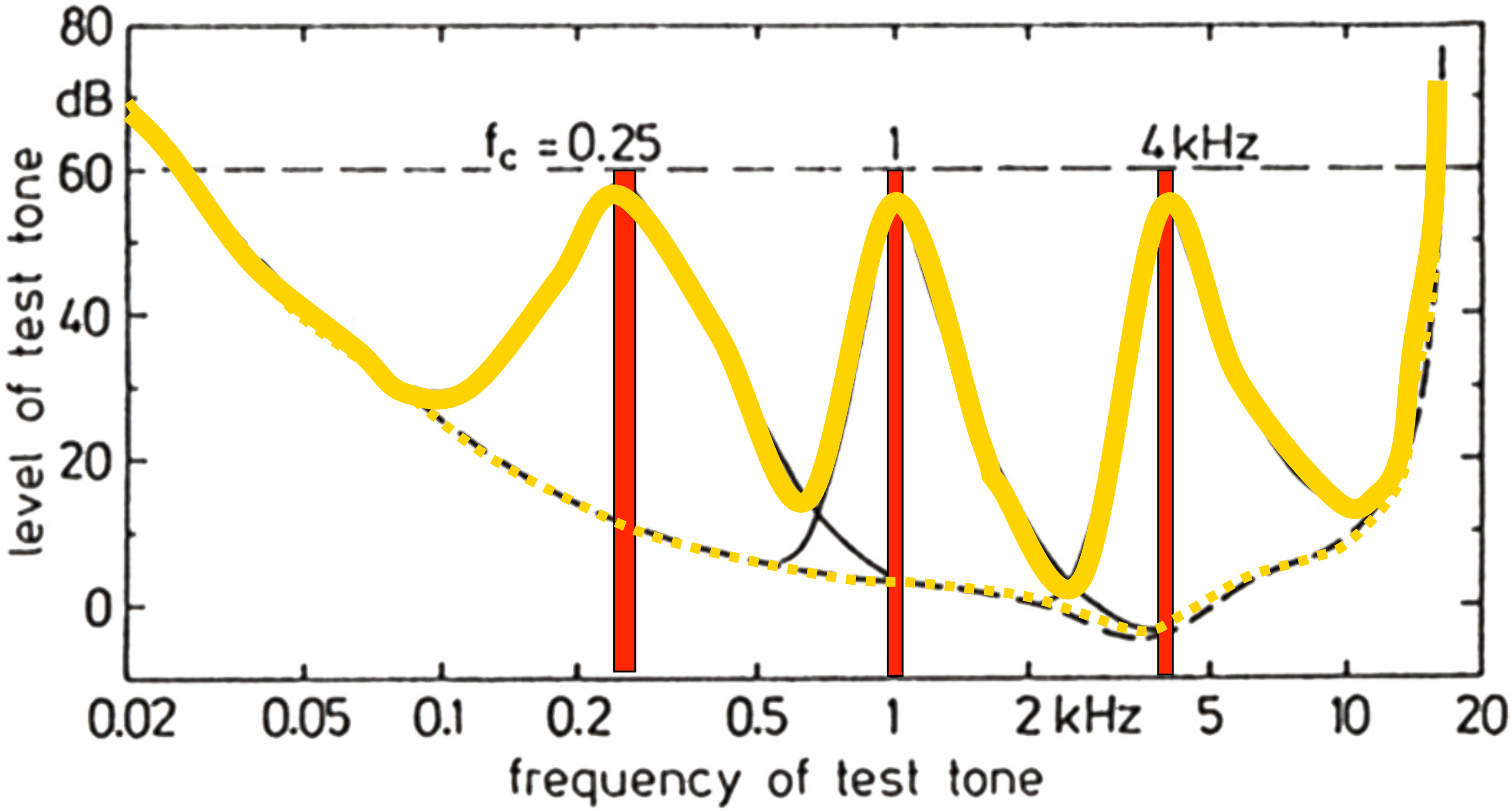
Zwicker/Fastl p. 17

Simultaneous Masking I



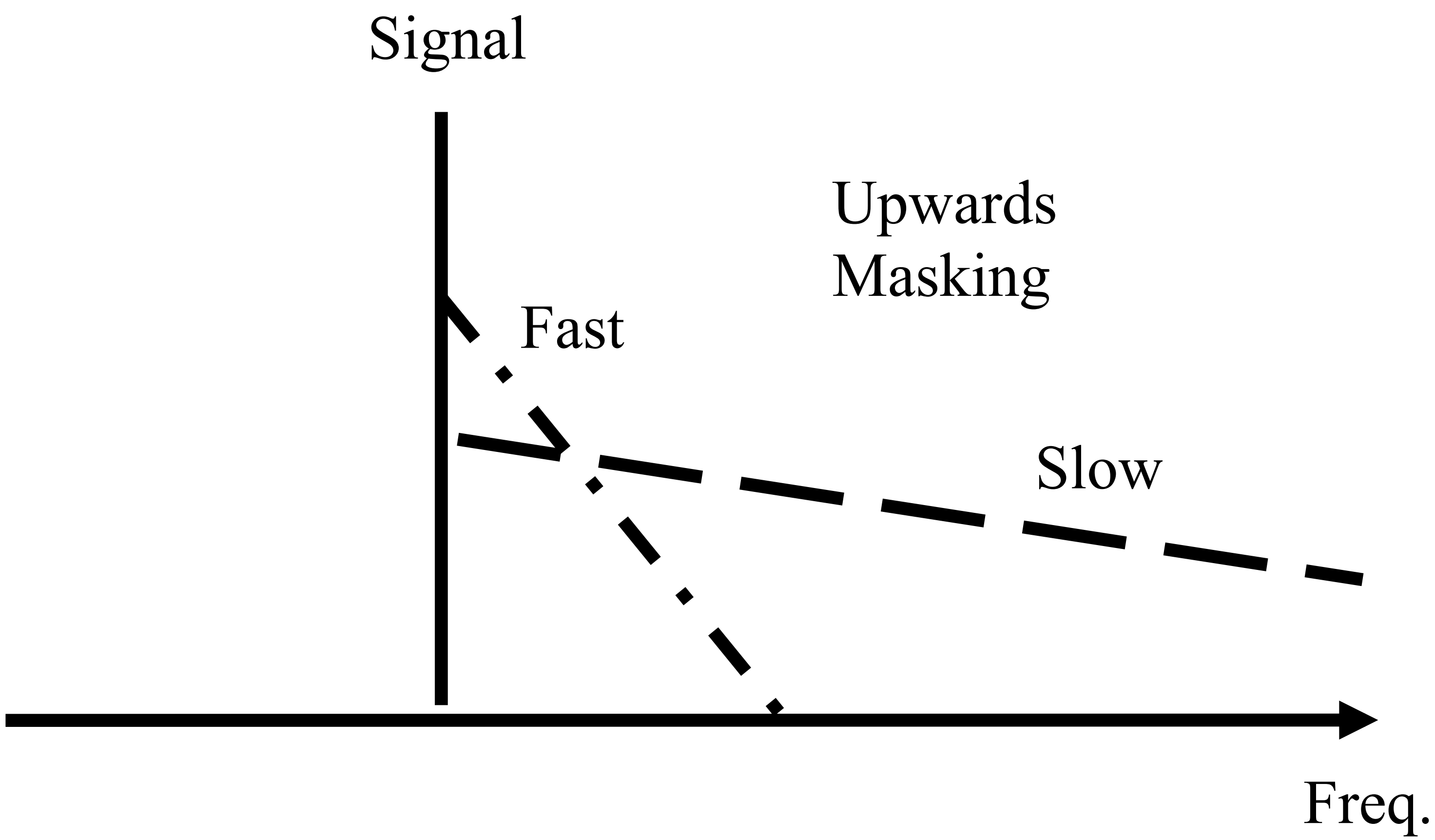
After Zwicker/Feldtkeller p. 60

Simultaneous Masking II: Overall Shape



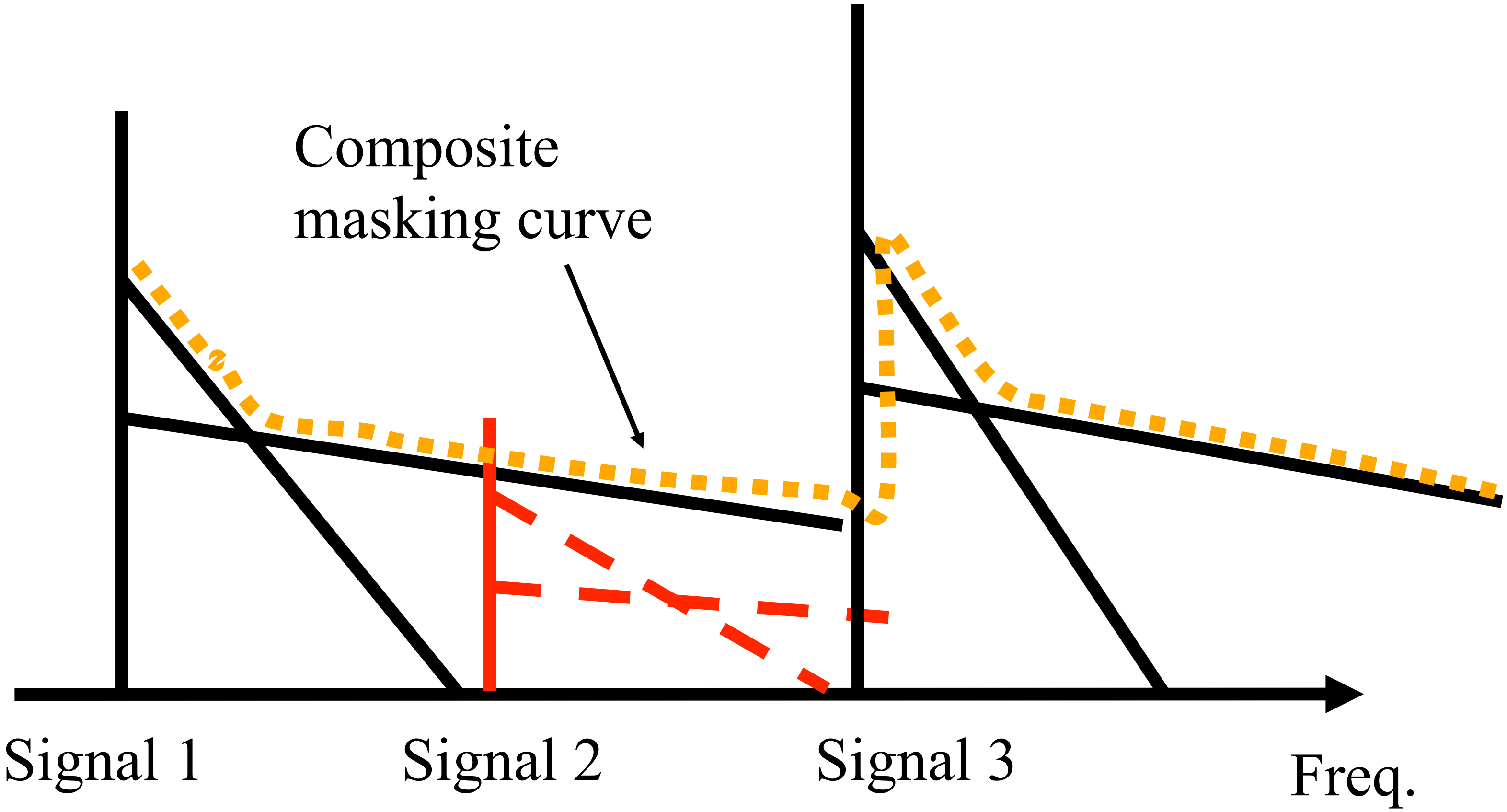
Zwicker/Fastl, p. 64

Composite Masking Curve (1) (AC-3)

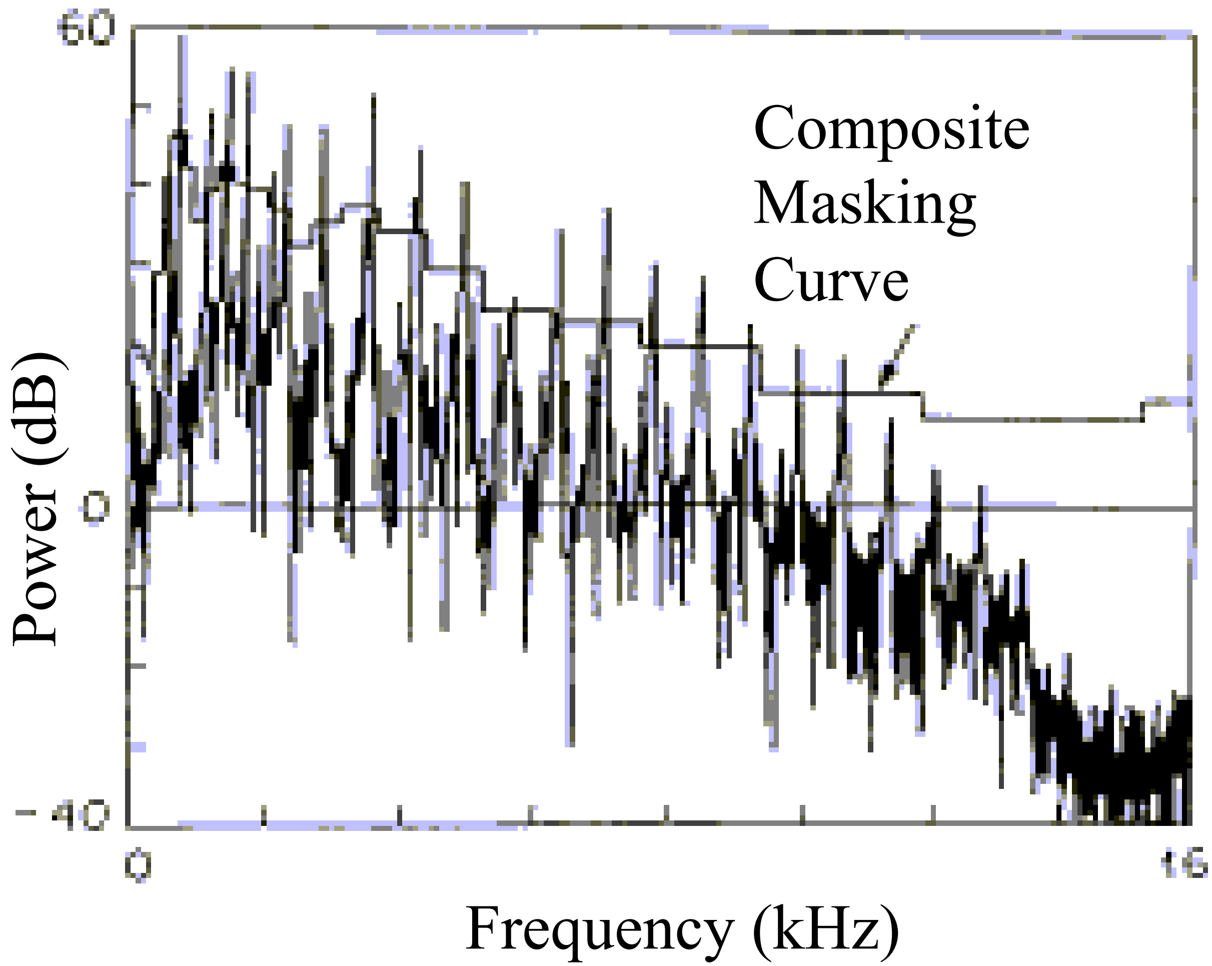


From: Todd et al, AES Preprint 3796, 1994

Composite Masking Curve (2) (AC-3)

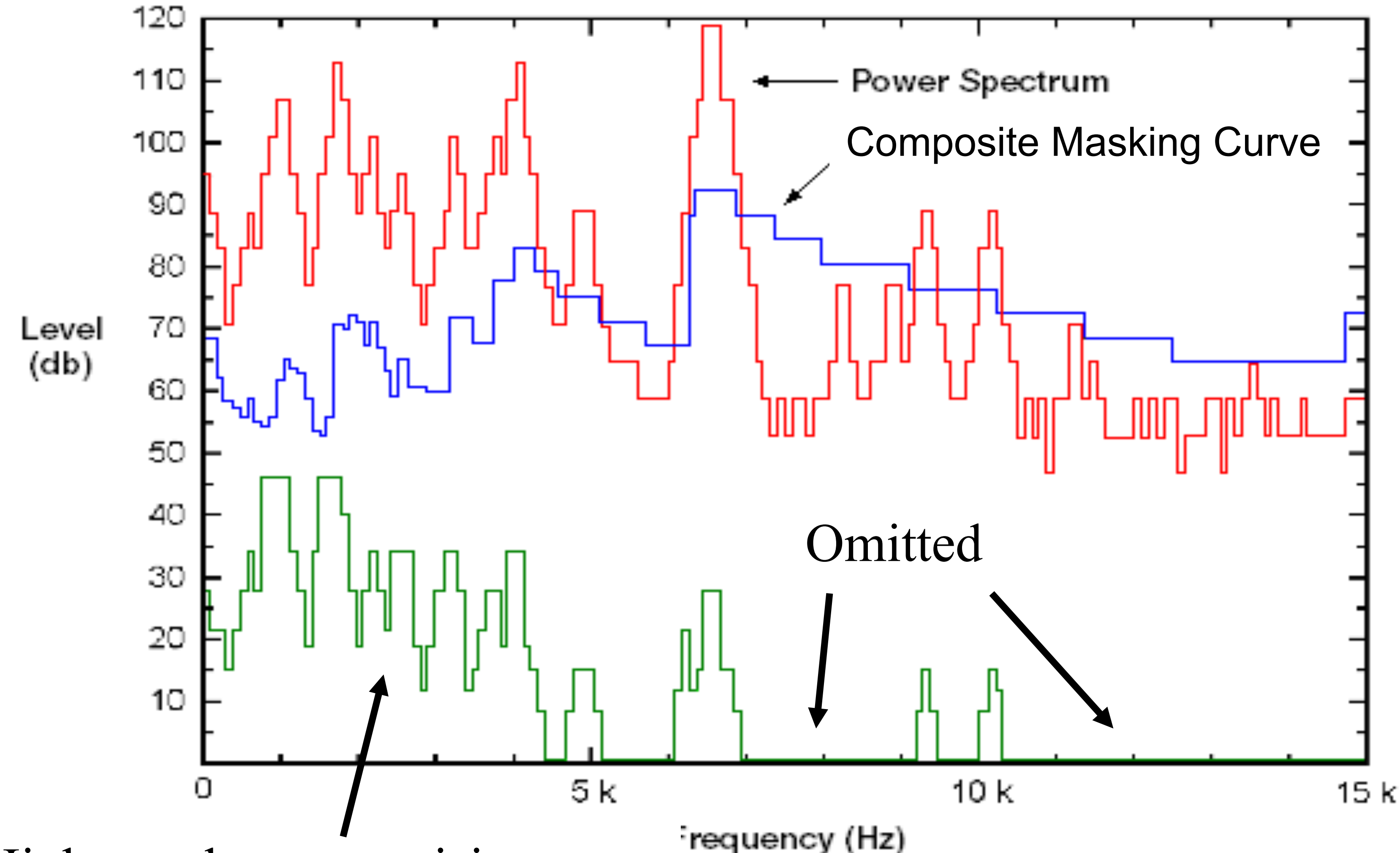


Composite Masking Curve (3) (PAC)



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

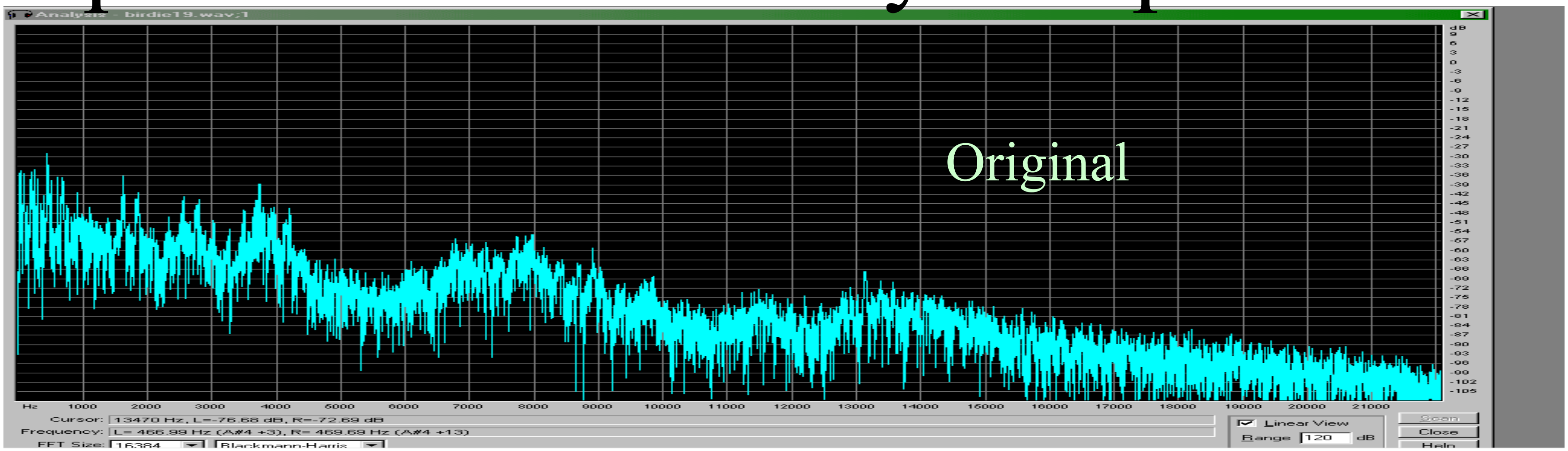
Masking: Leave Something Out

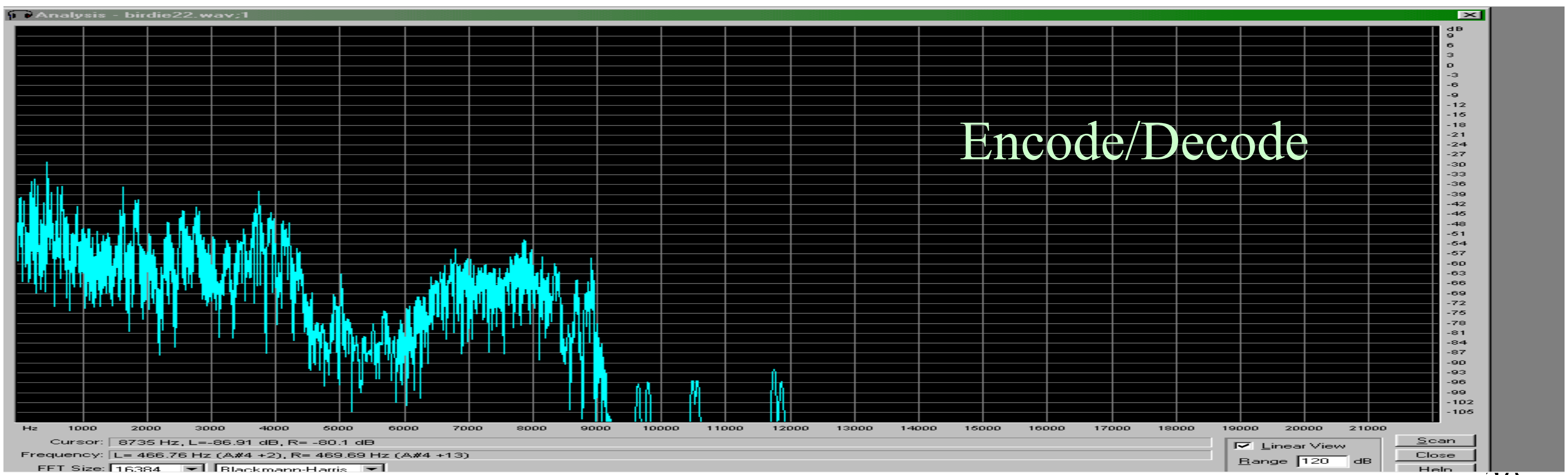


Higher or lower precision

After Davidson et al., 1994

Spectrum modified by compression

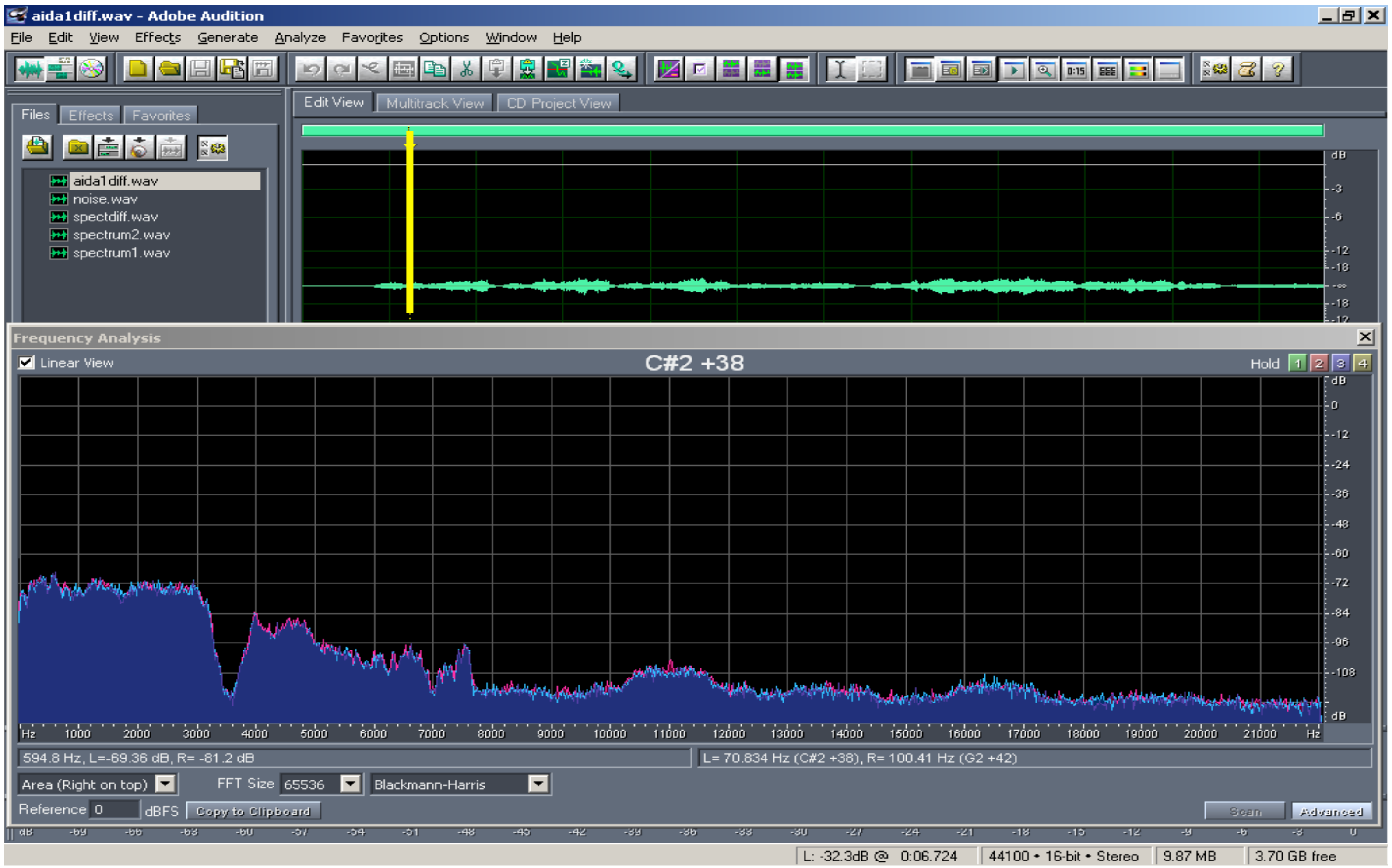




resyn

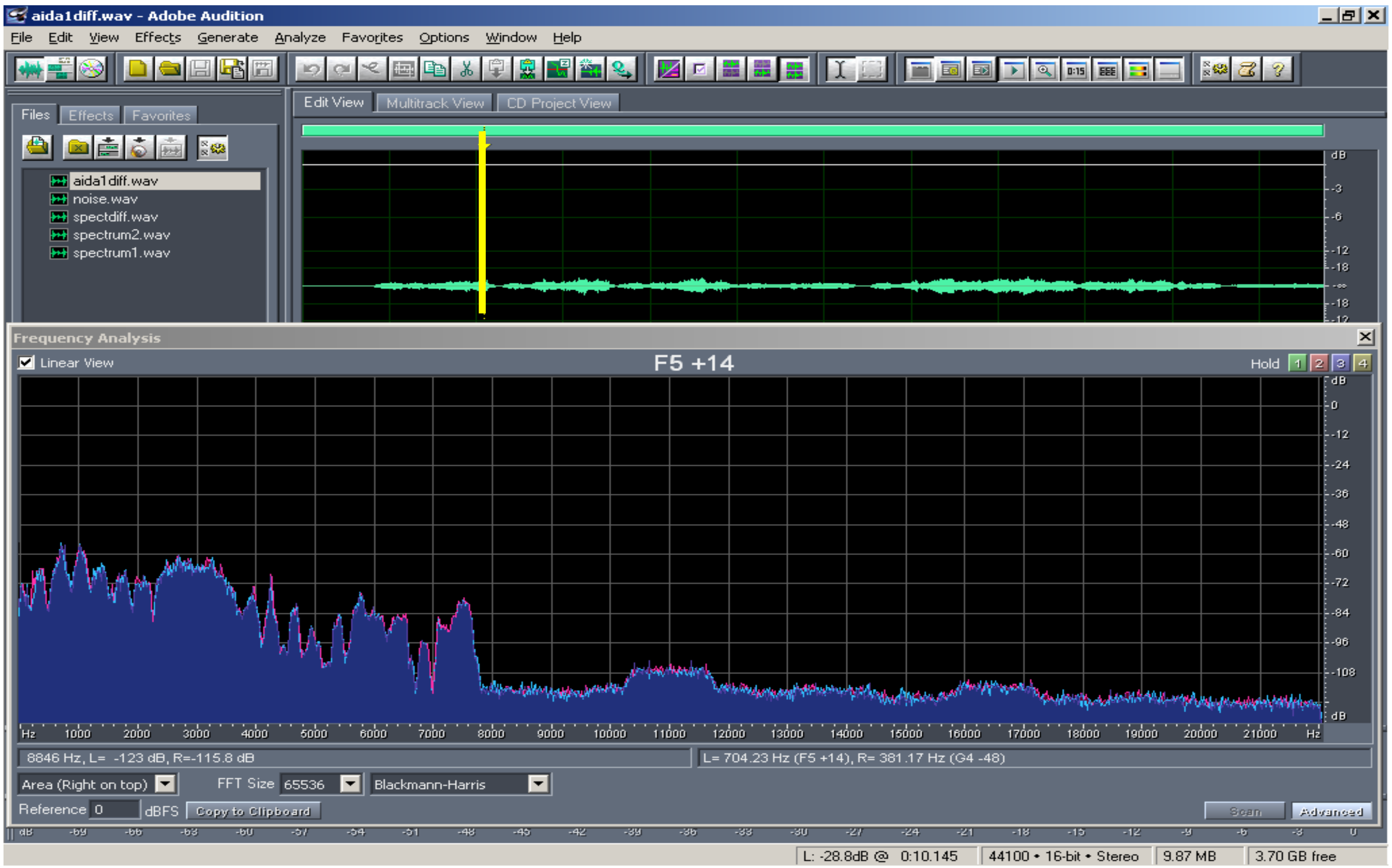
diff

What's left out (I)



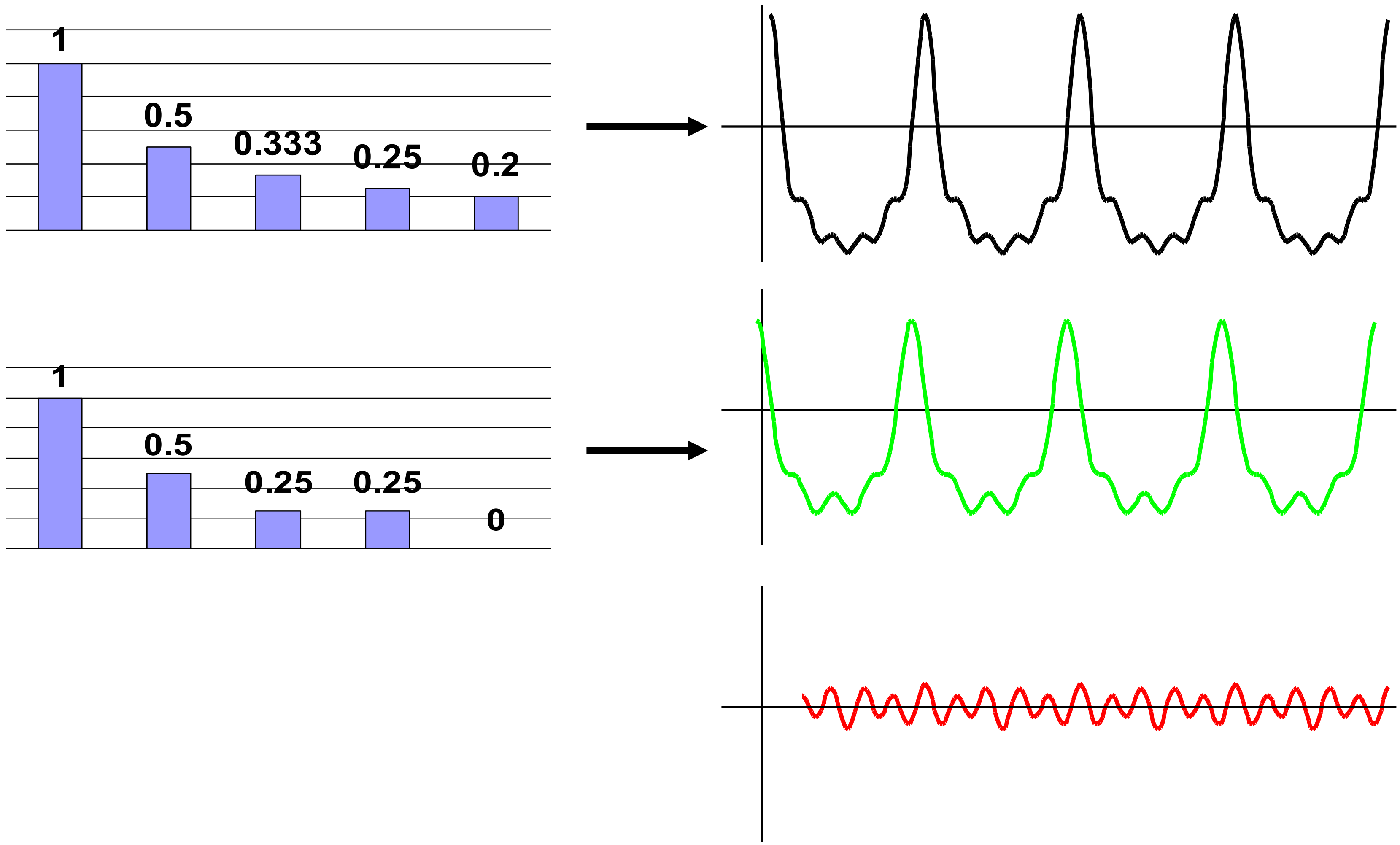
Ben Heppner, *Great Tenor Arias*, Munich Radio Orchestra, conducted by Roberto Abbado, 09026-62504-2, 1995, BMG Music, Approximately one minute from Verdi's *Aida*, "Celeste Aida", track 4. MP3, 56 kpbs.

What's left out (II)

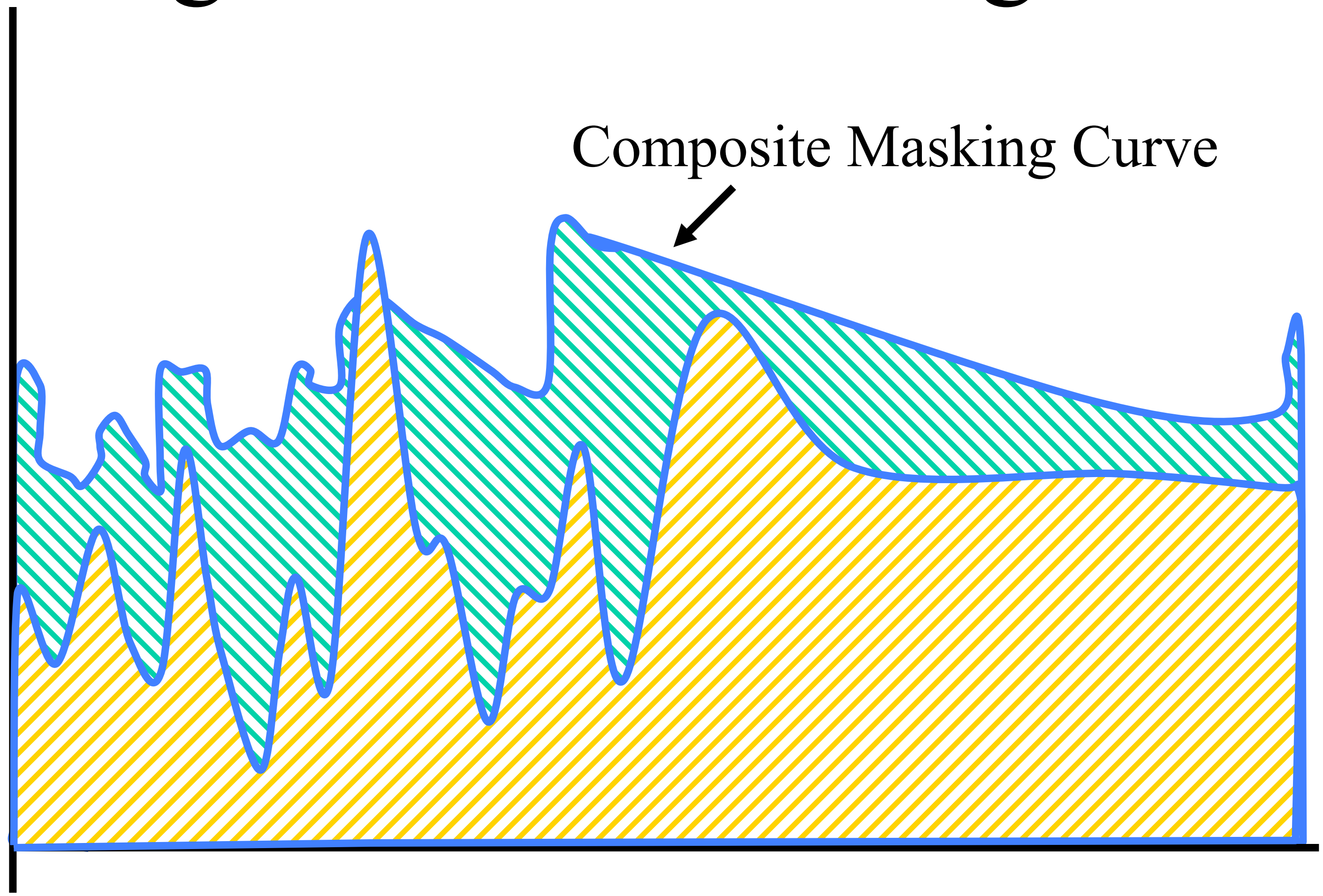


Ben Heppner, *Great Tenor Arias*, Munich Radio Orchestra, conducted by Roberto Abbado, 09026-62504-2, 1995, BMG Music, Approximately one minute from Verdi's *Aida*, "Celeste Aida", track 4. MP3, 56 kpbs.

Quantized Spectrum

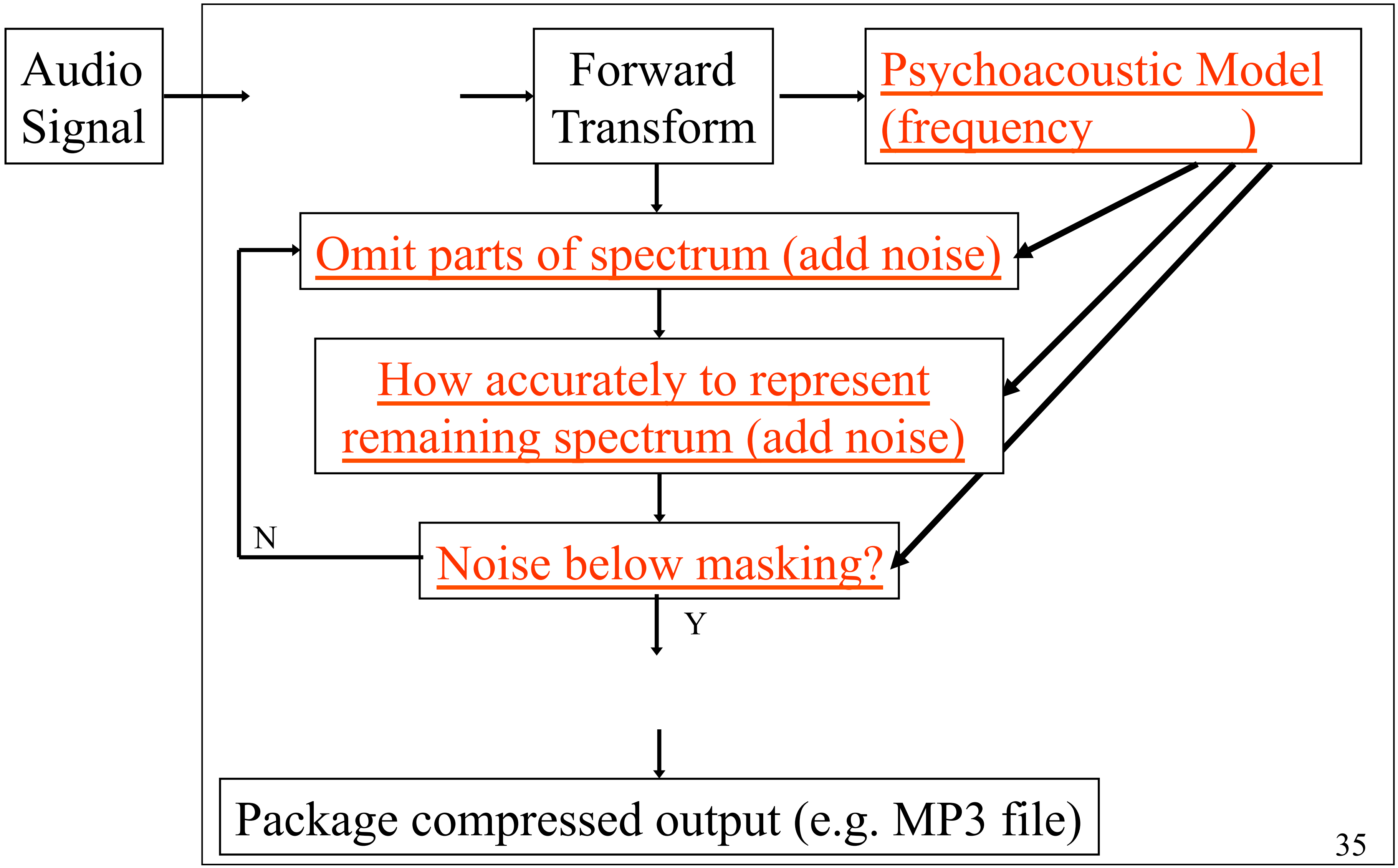


Masking: Where noise might fall

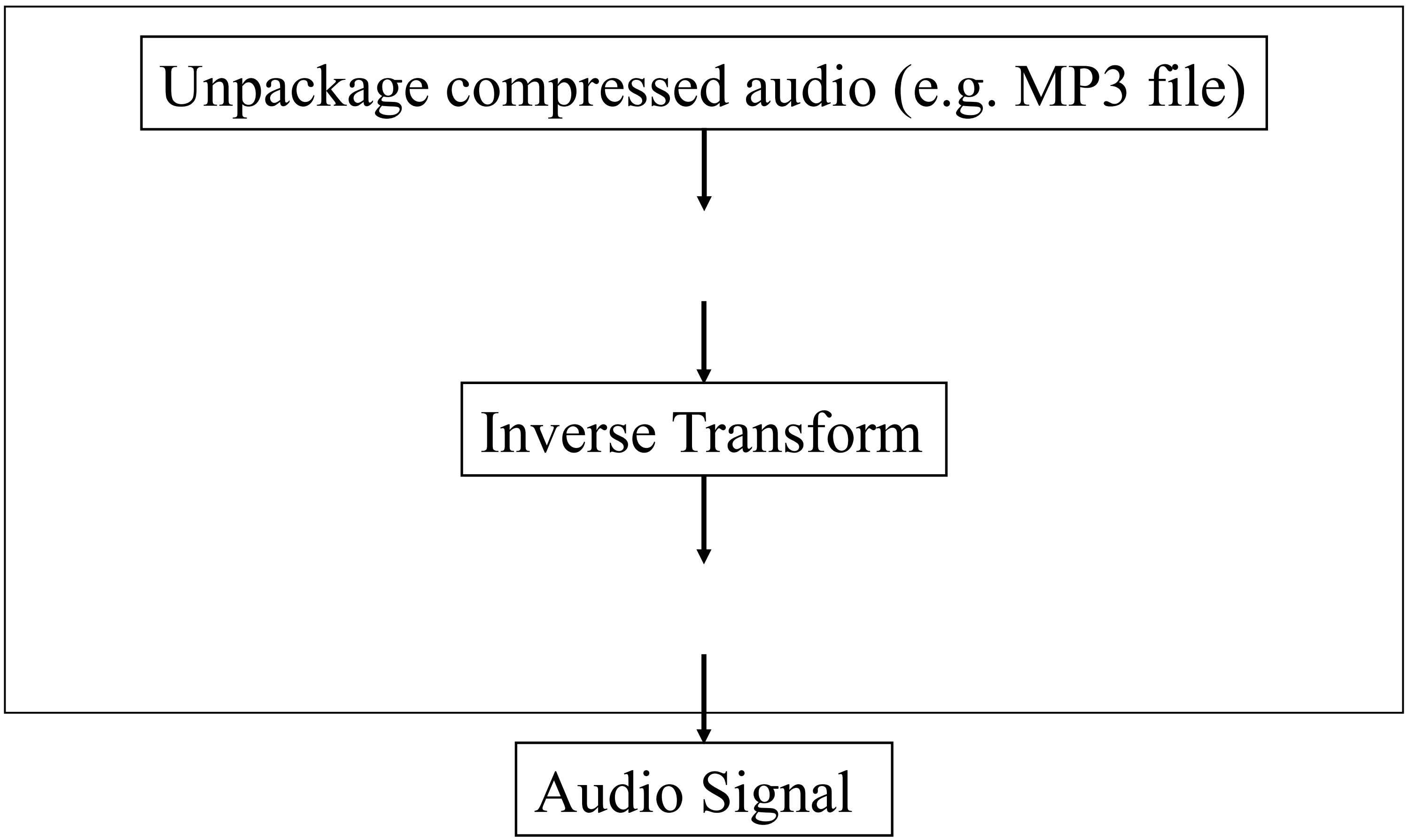


After Davidson et al., 1994

Perceptual Encoder



Decoder ("mp3 player")

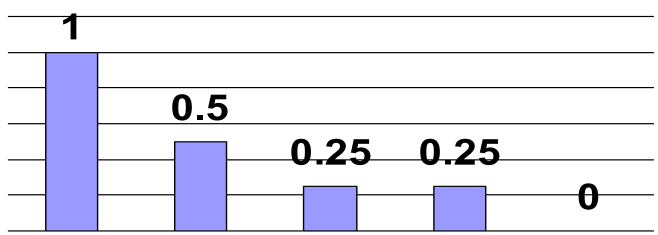


Representing the Spectrum

Fundamentals of perceptual coding

- Leave out
  - Irrelevant
  - Redundant
- Allow noise
  - But only in special places

Quantized Spectrum / bit stream



... 1111111111111111 1000000000000000 0100000000000000 0100000000000000 0000000000000000 ...

“Noiseless” coding (I)

- Huffman coding

5	Mexico City
8	Monterrey
69	Mazatlan
684	Cabo San Lucas

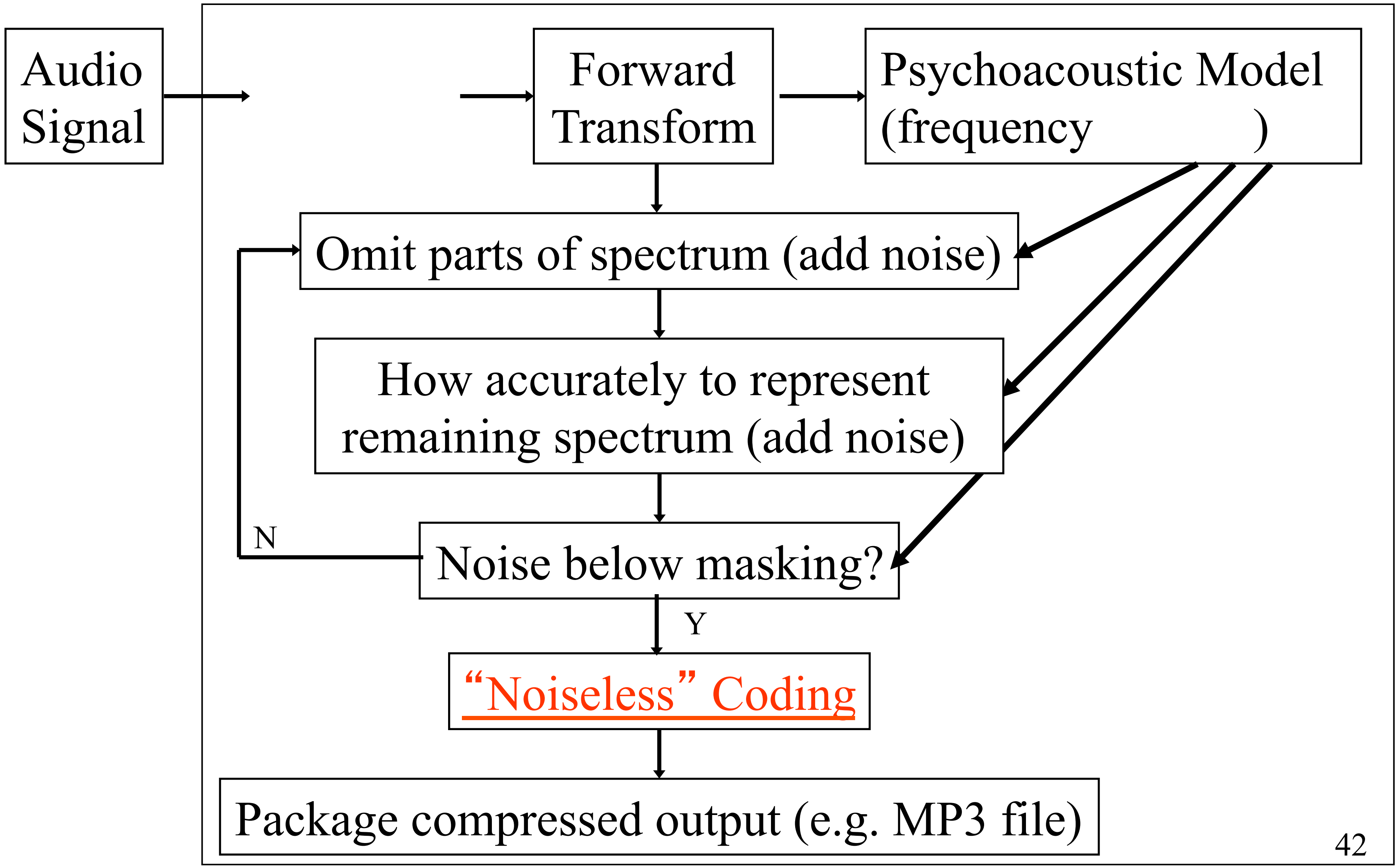
"Noiseless" coding (II)

- Run length:

```
... 1111111111111111 1000000000000000 0100000000000000

01000000000000000 0000000000000000 ...
```

Perceptual Encoder



Decoder ("mp3 player")

Unpackage compressed audio (e.g. MP3 file)

↓

Undo "Noiseless" Coding

↓

Inverse Transform

↓

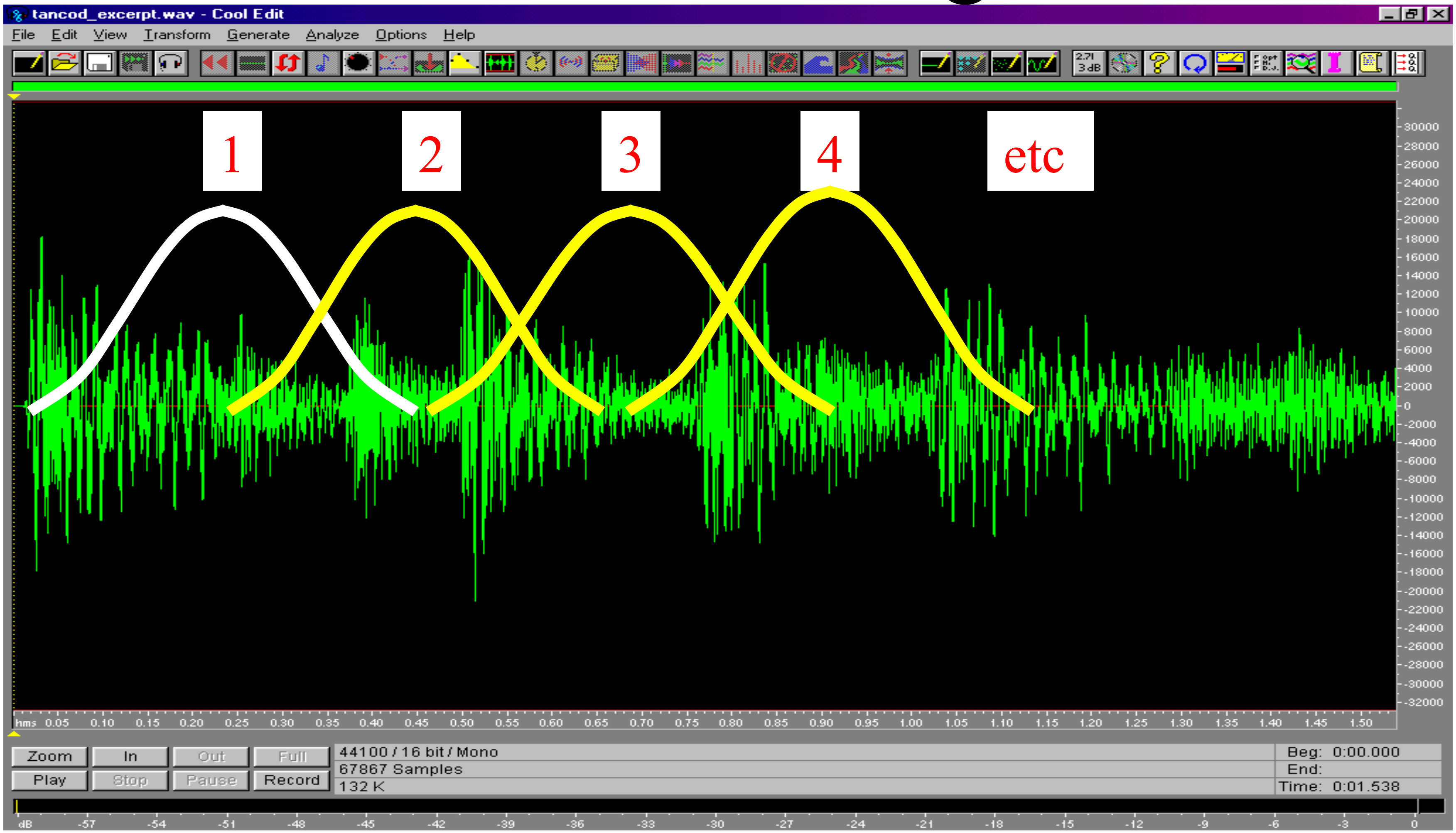
Audio Signal

Review: Fundamentals of perceptual coding

- Leave out
  - Irrelevant
  - Redundant
- Allow noise
  - But only in special places

Window

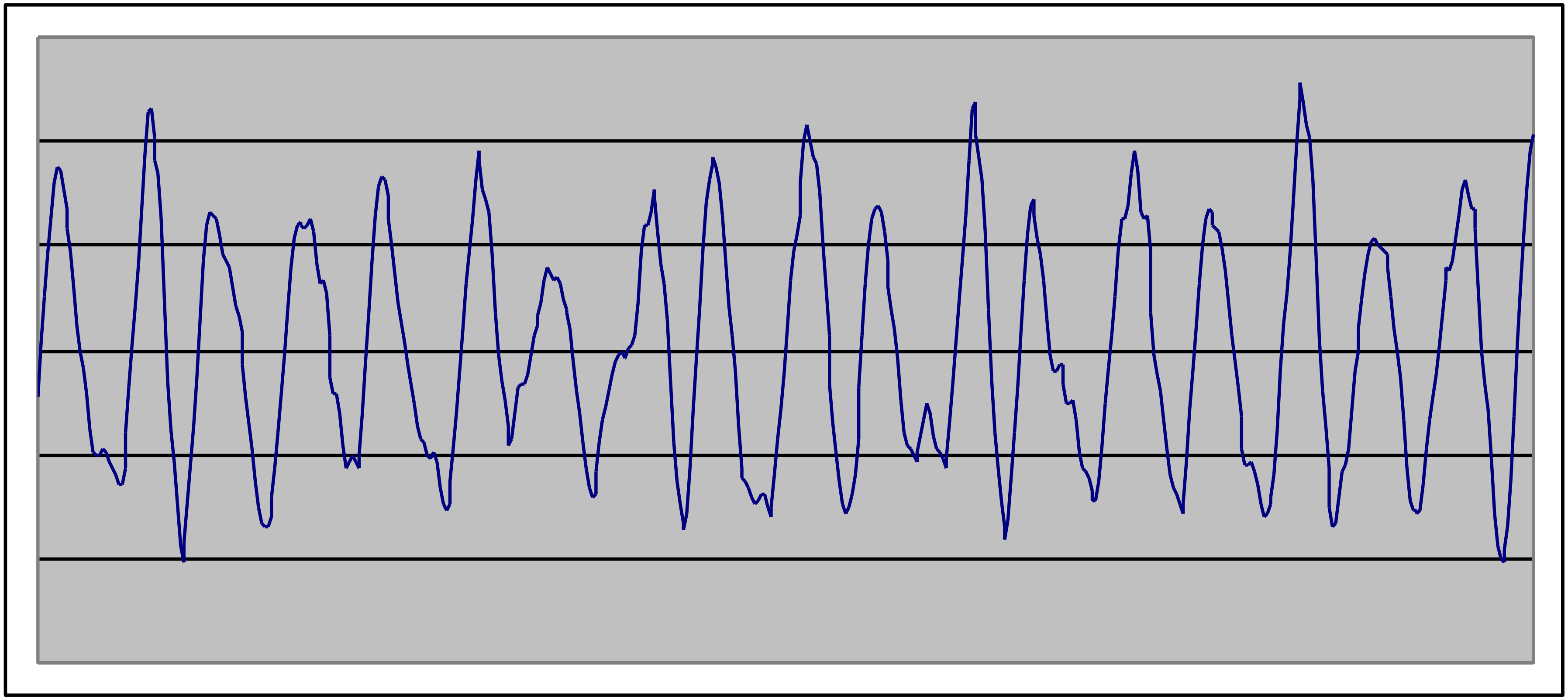
Windowing



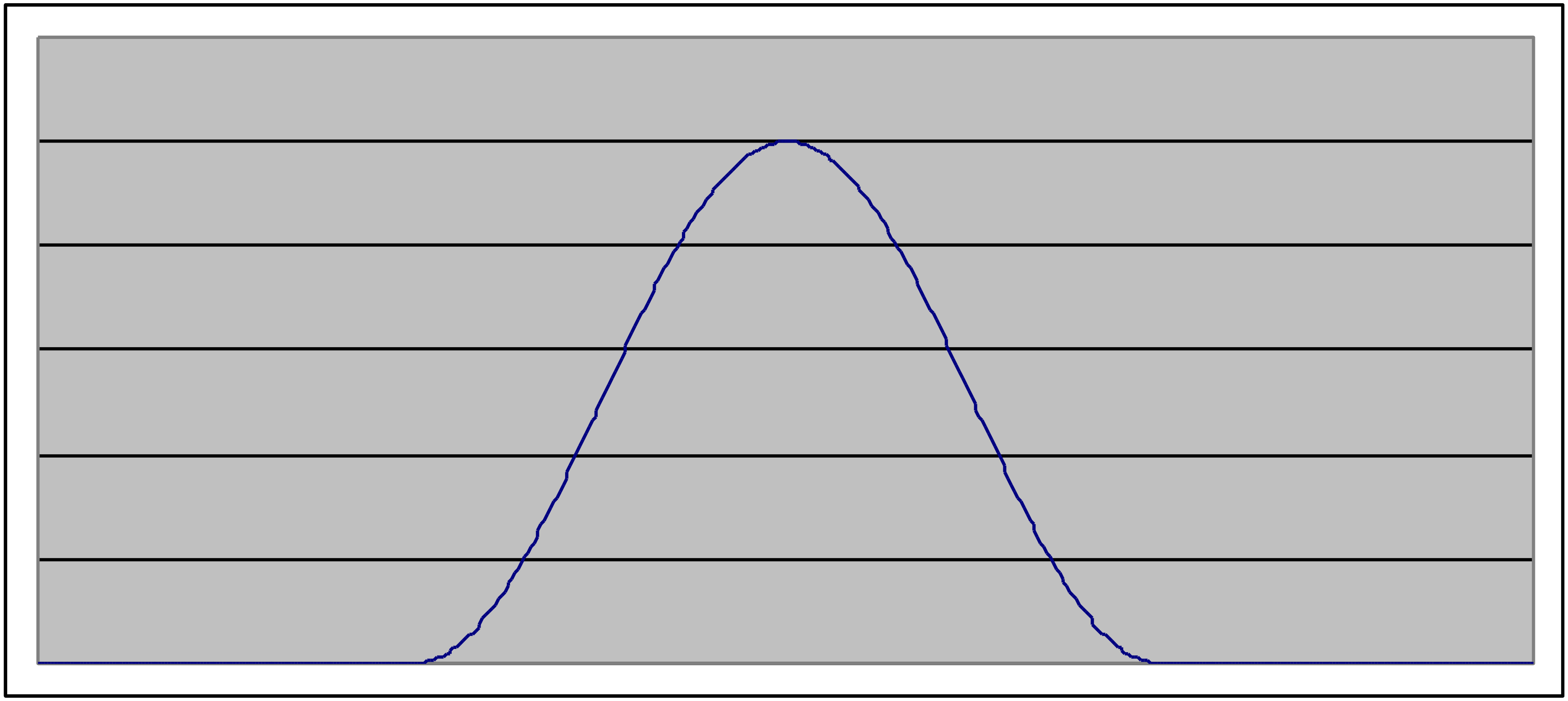
Source: excerpt from tancod55.wav, AES CD-ROM

Note: windows exaggerated in various ways for illustration.

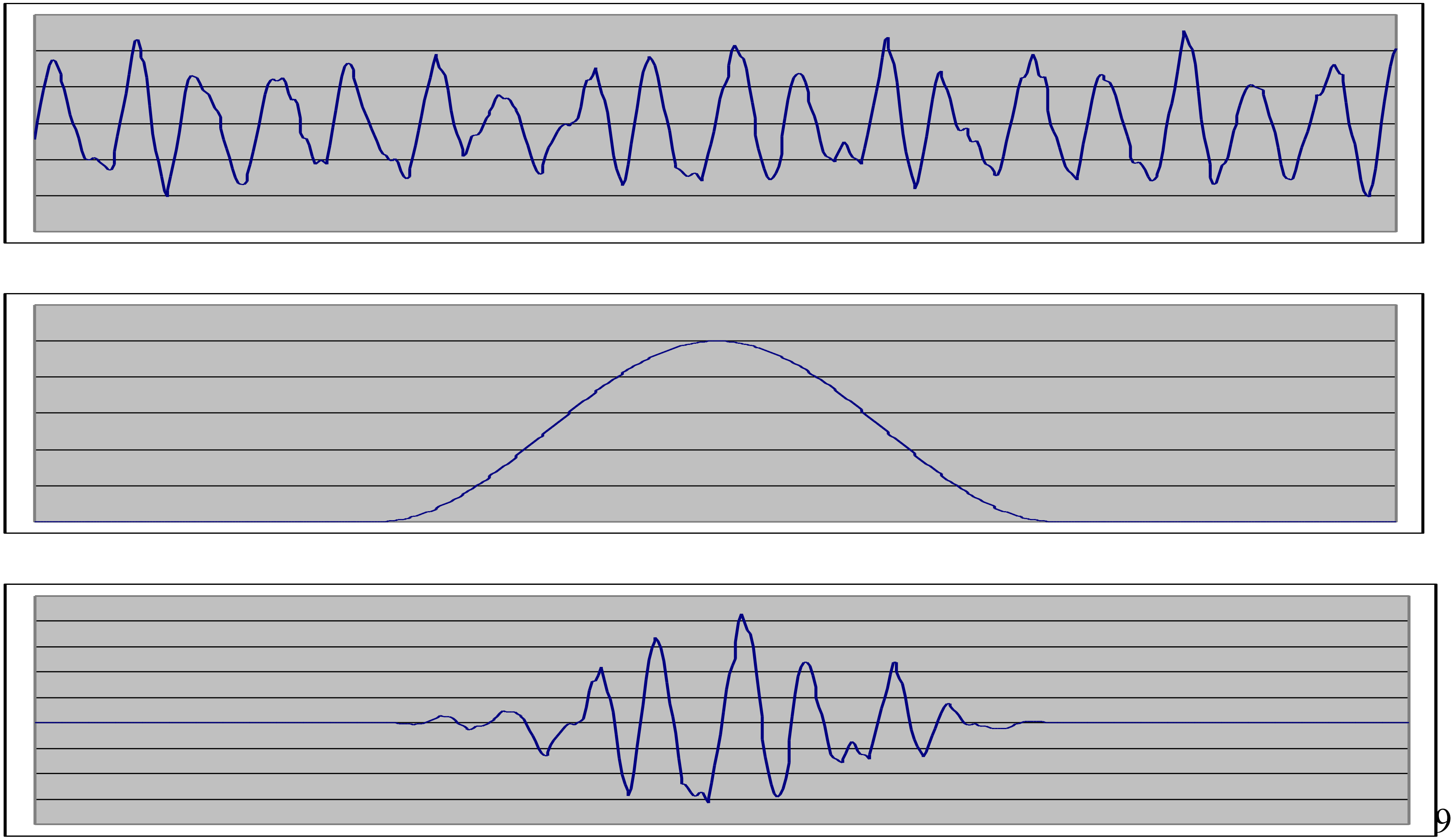
Zoom in on original signal



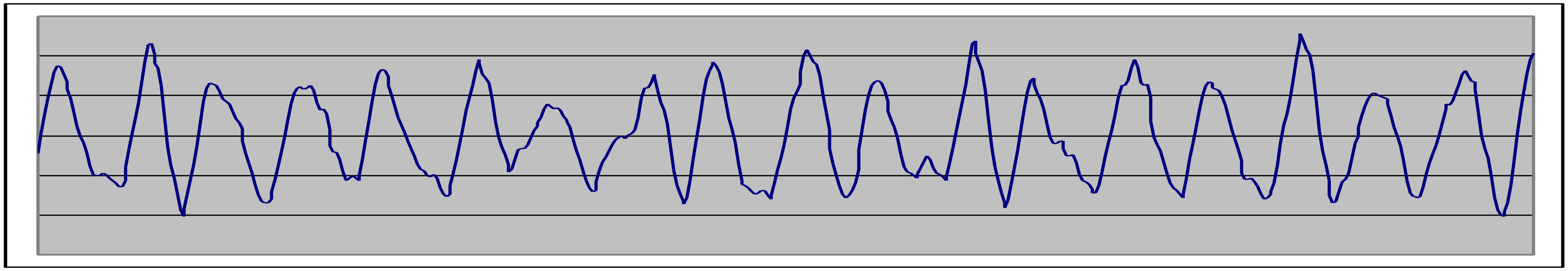
Typical Window

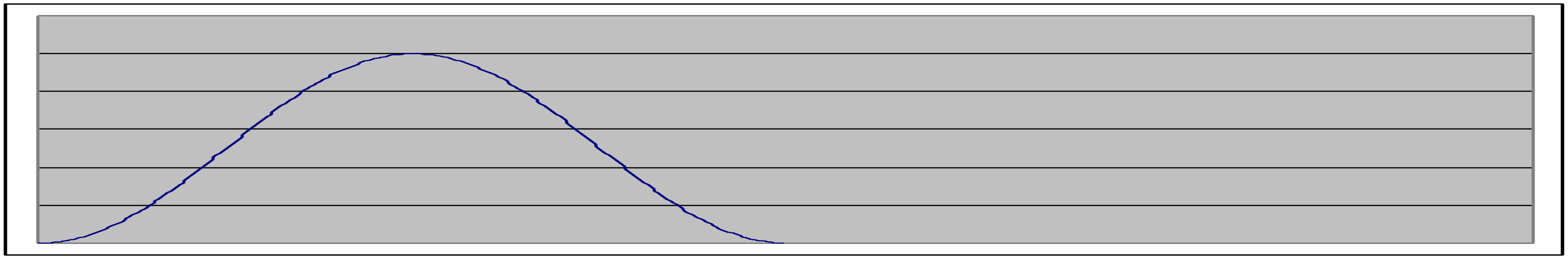


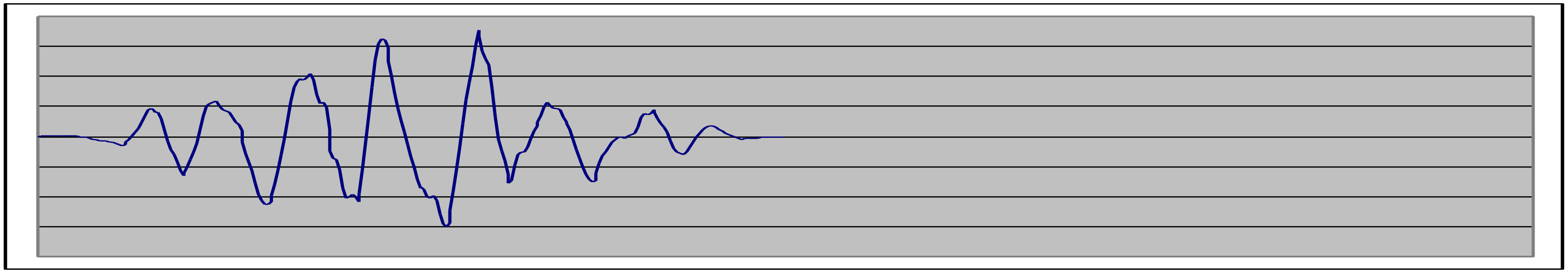
Middle of Original, Windowed



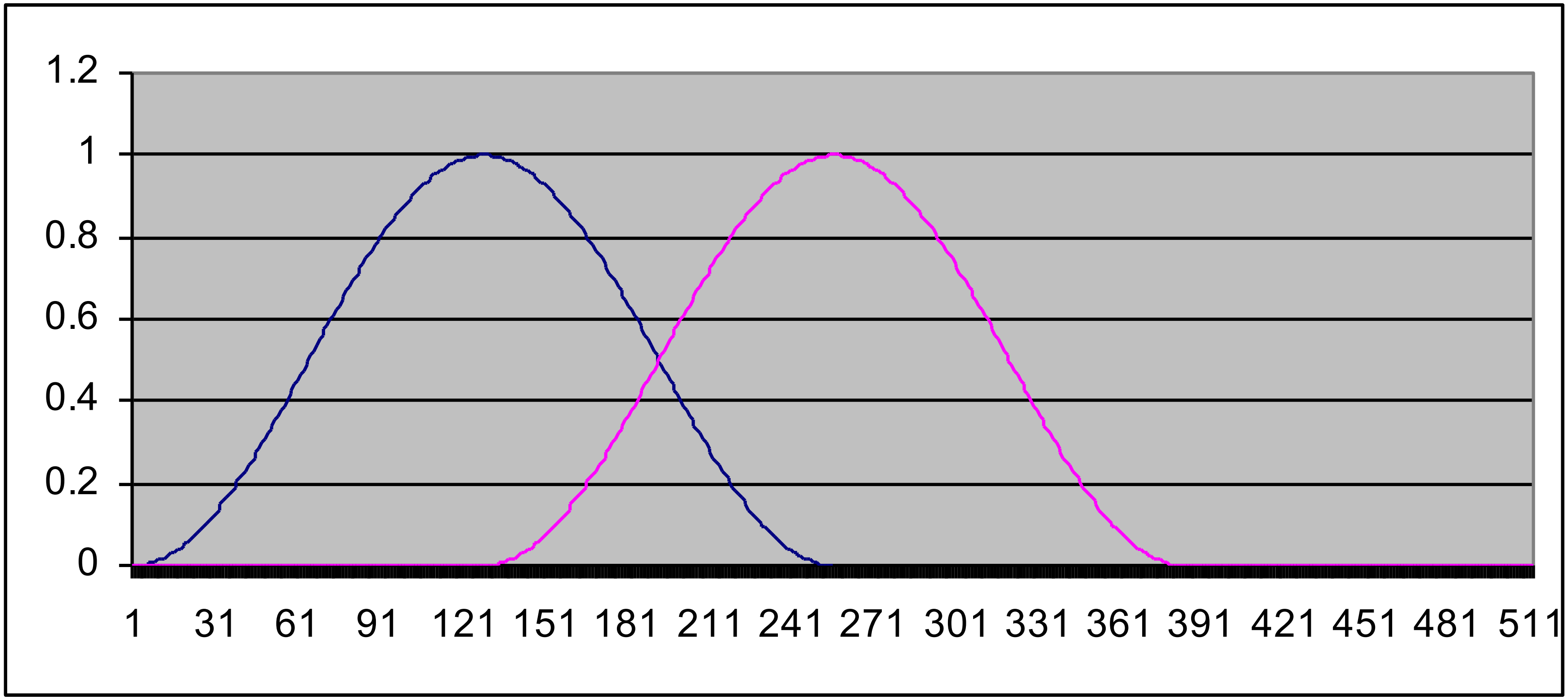
Left side of original, windowed



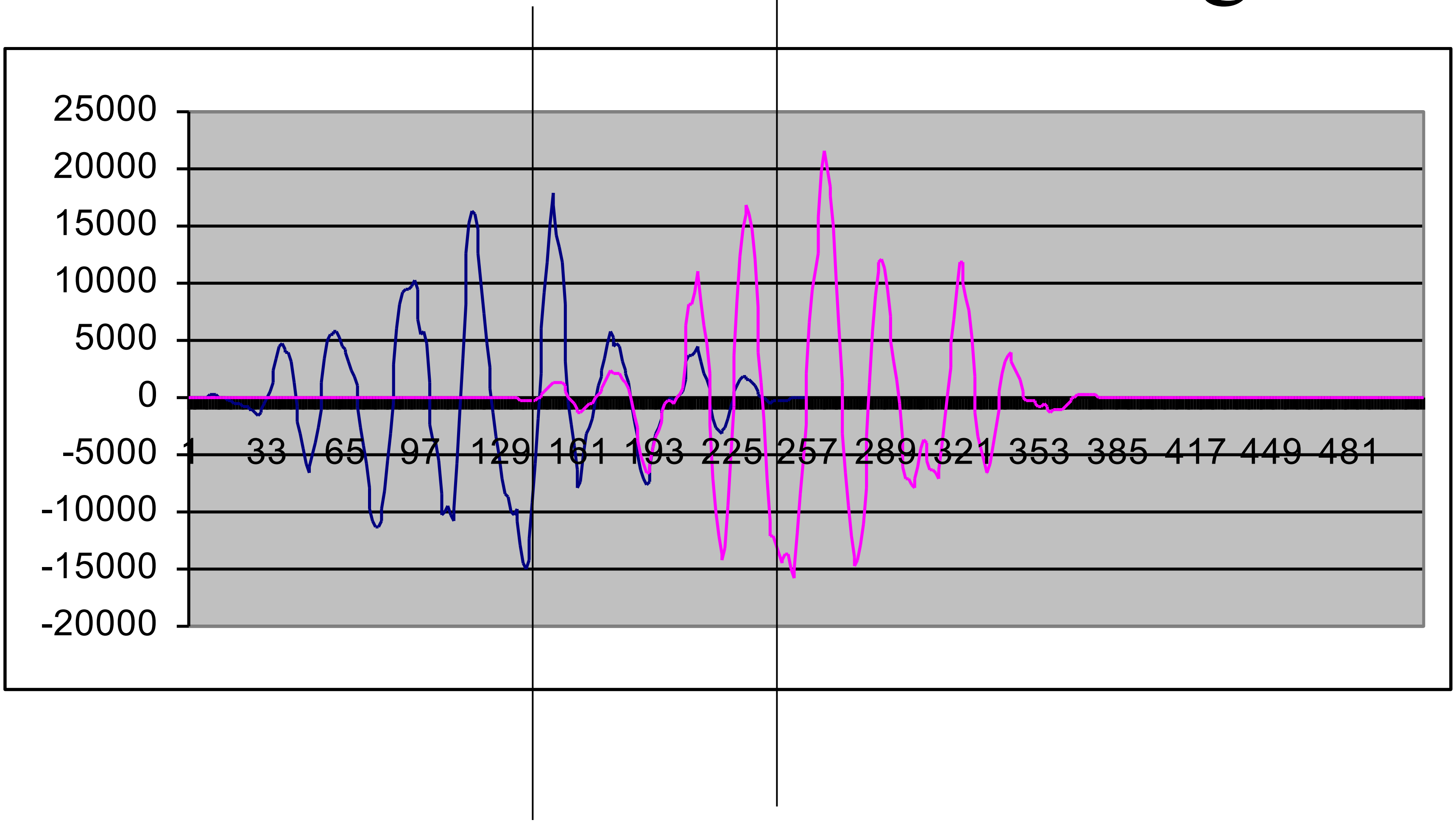




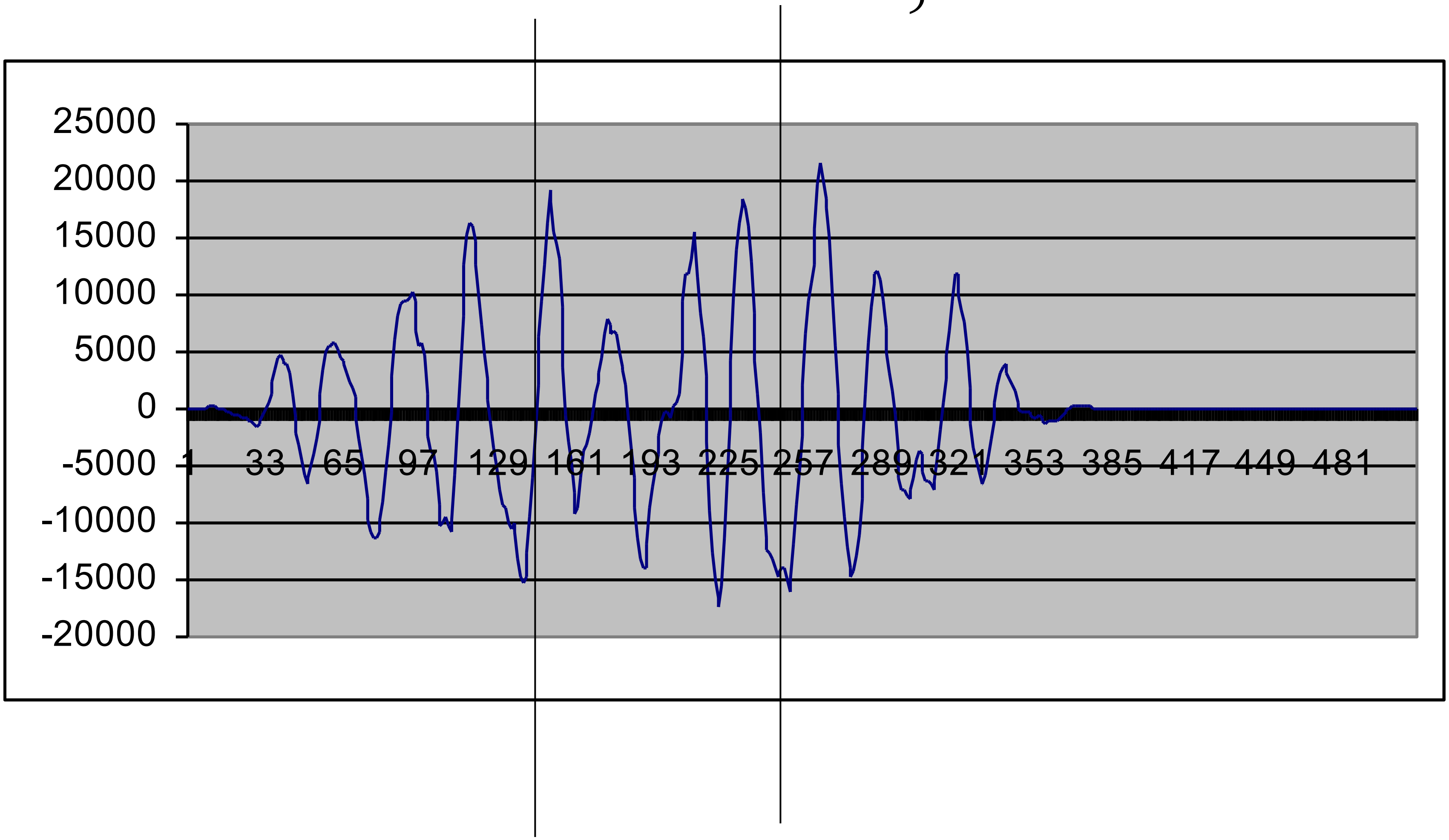
How windows line up



How windowed data align

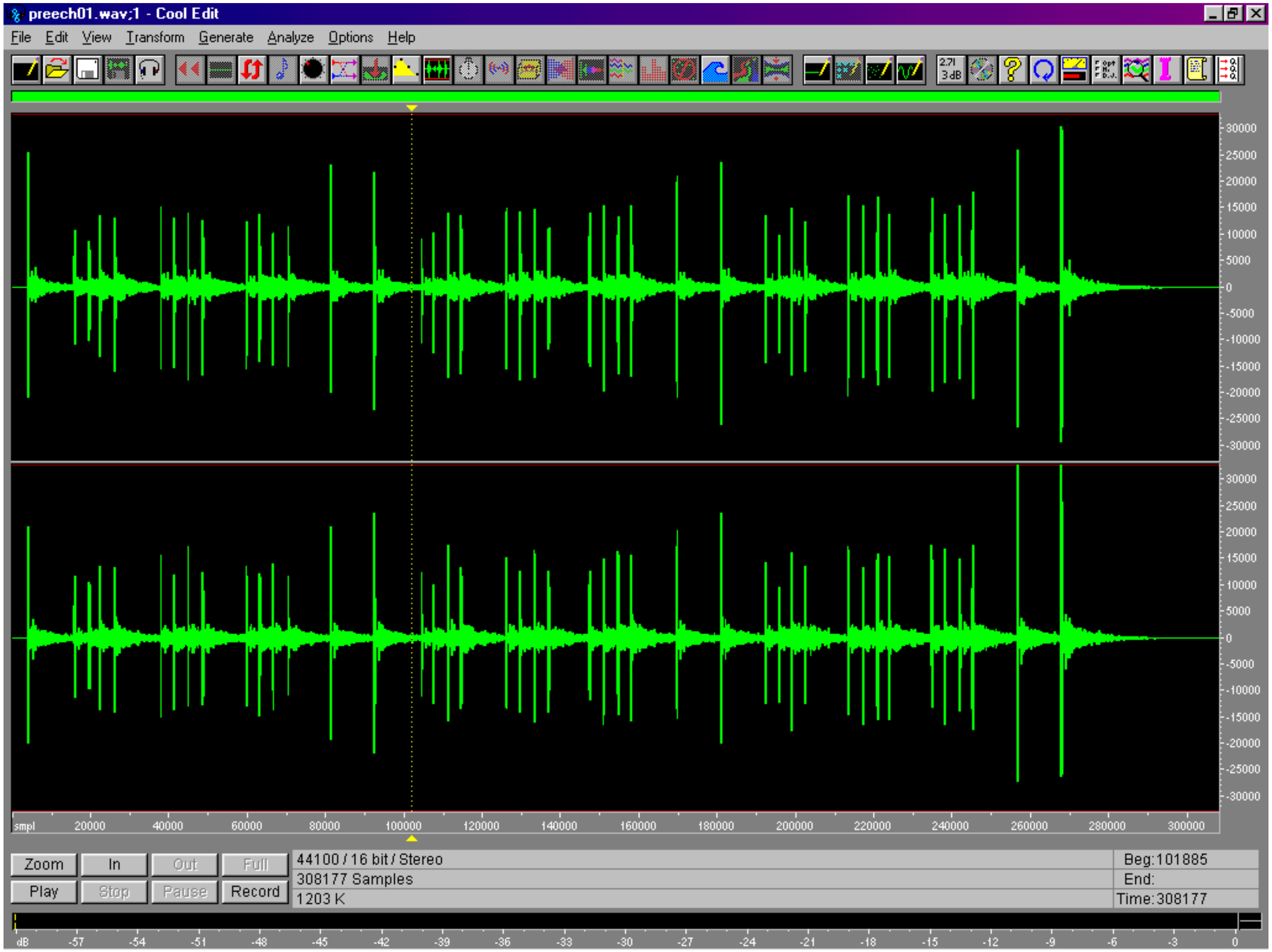


Windowed Data, Summed



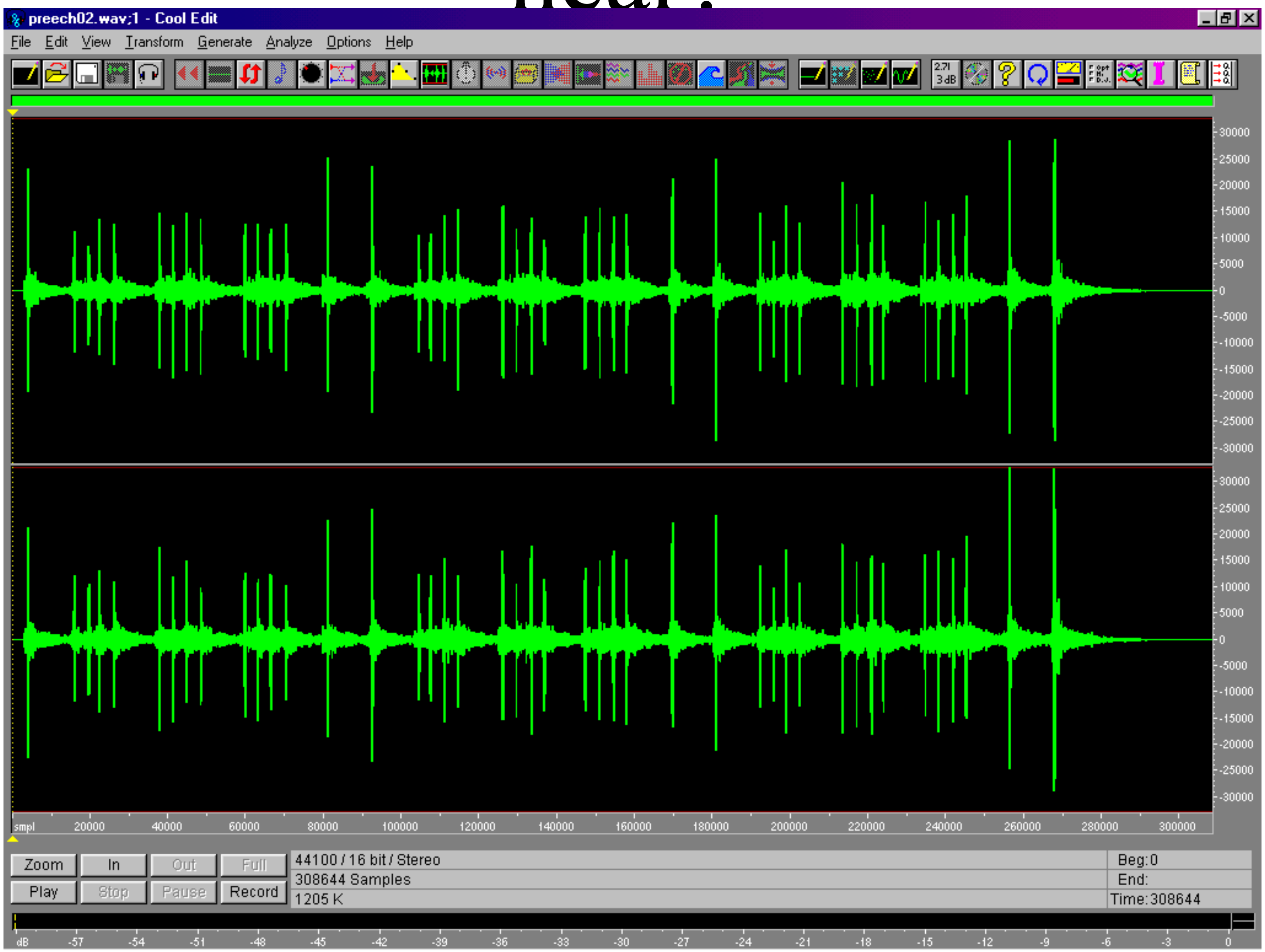
Temporal Masking

Original: Castanets



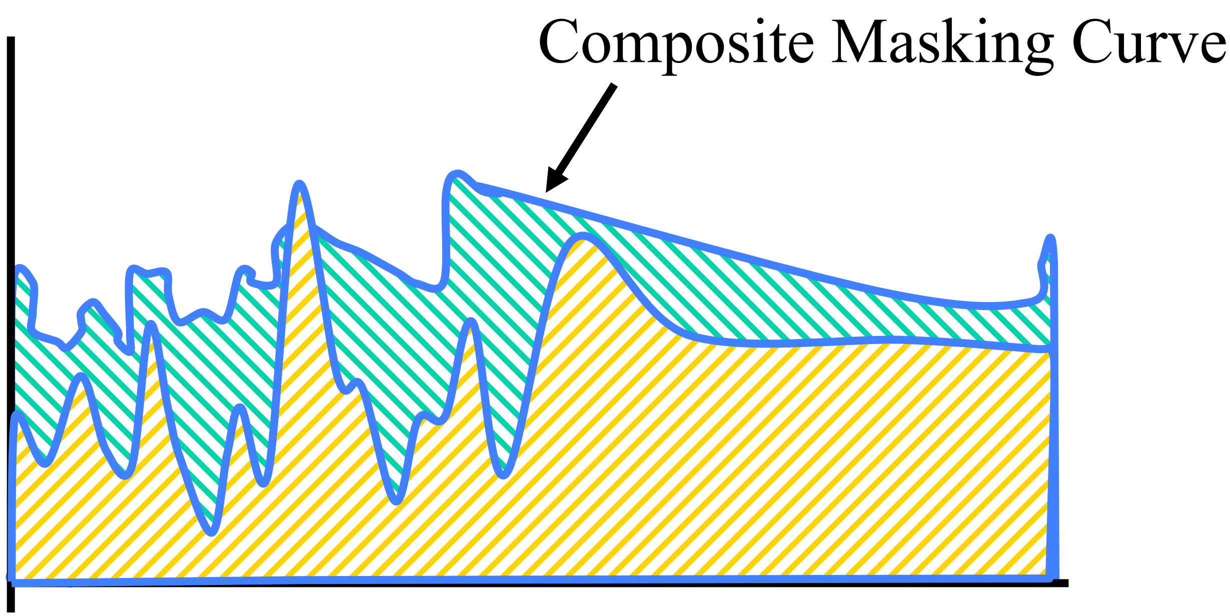
Source: AES CD-ROM

Encoded/Decoded: What do you hear?



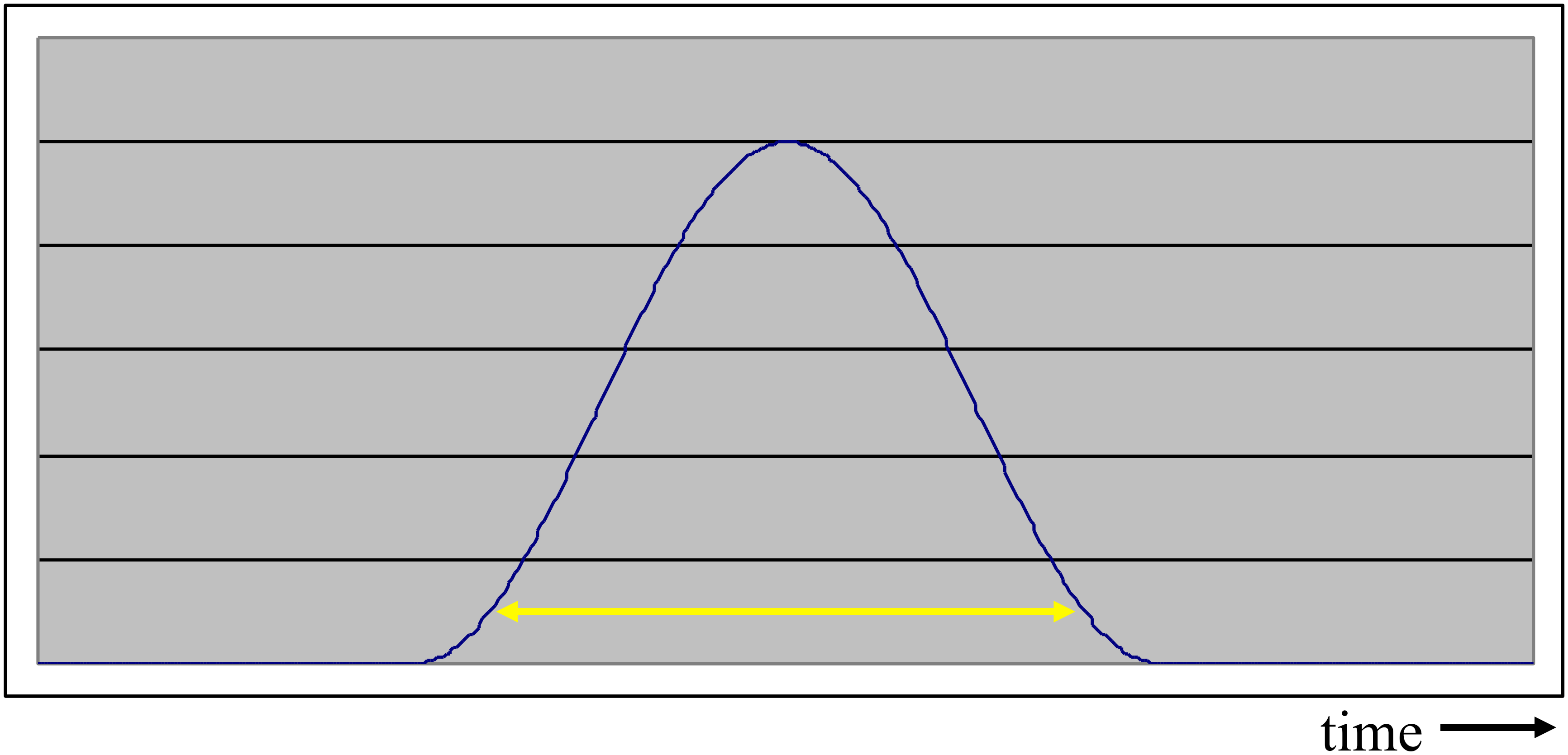
Source: AES CD-ROM

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

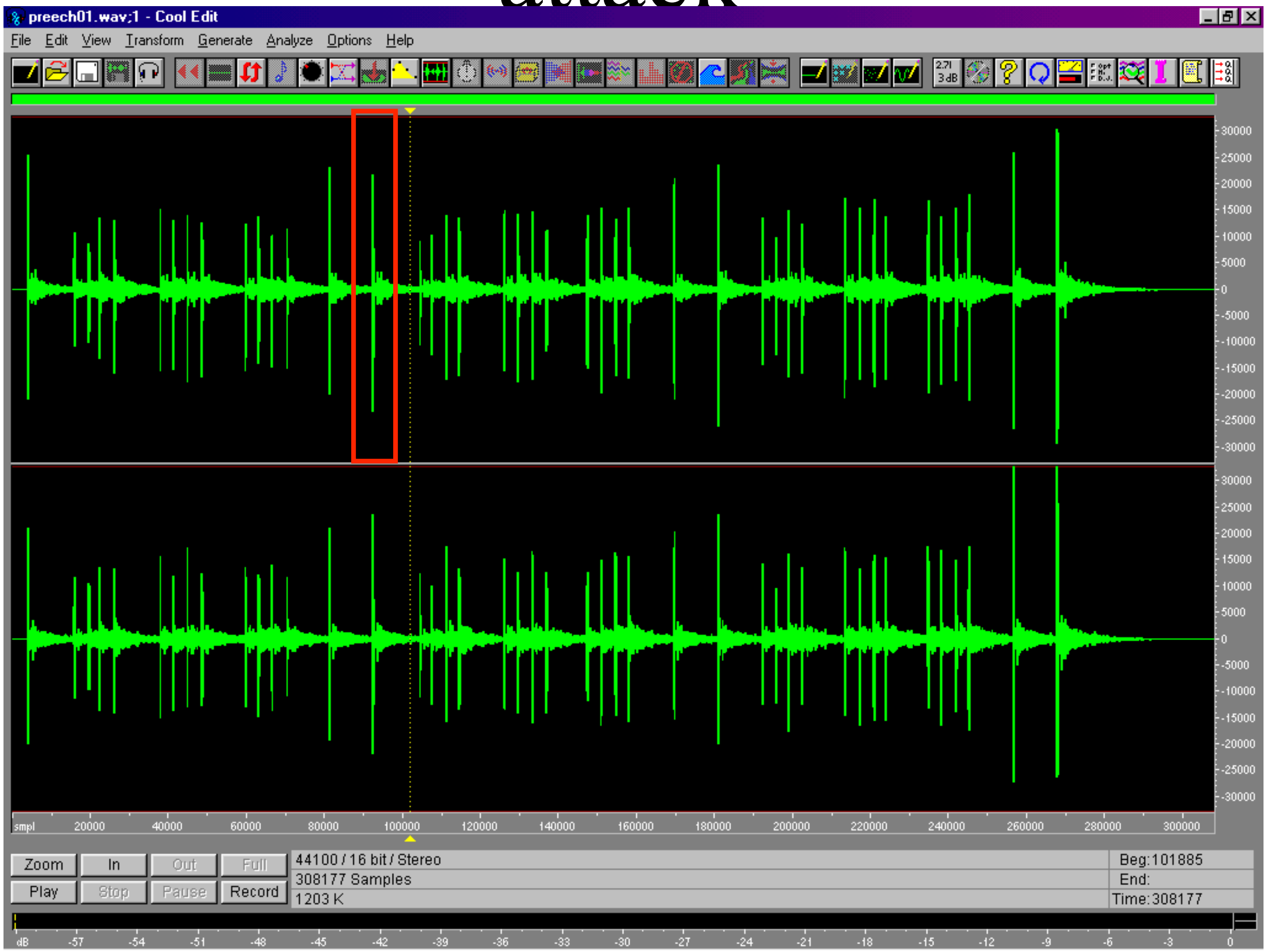


After Davidson et al., 1994

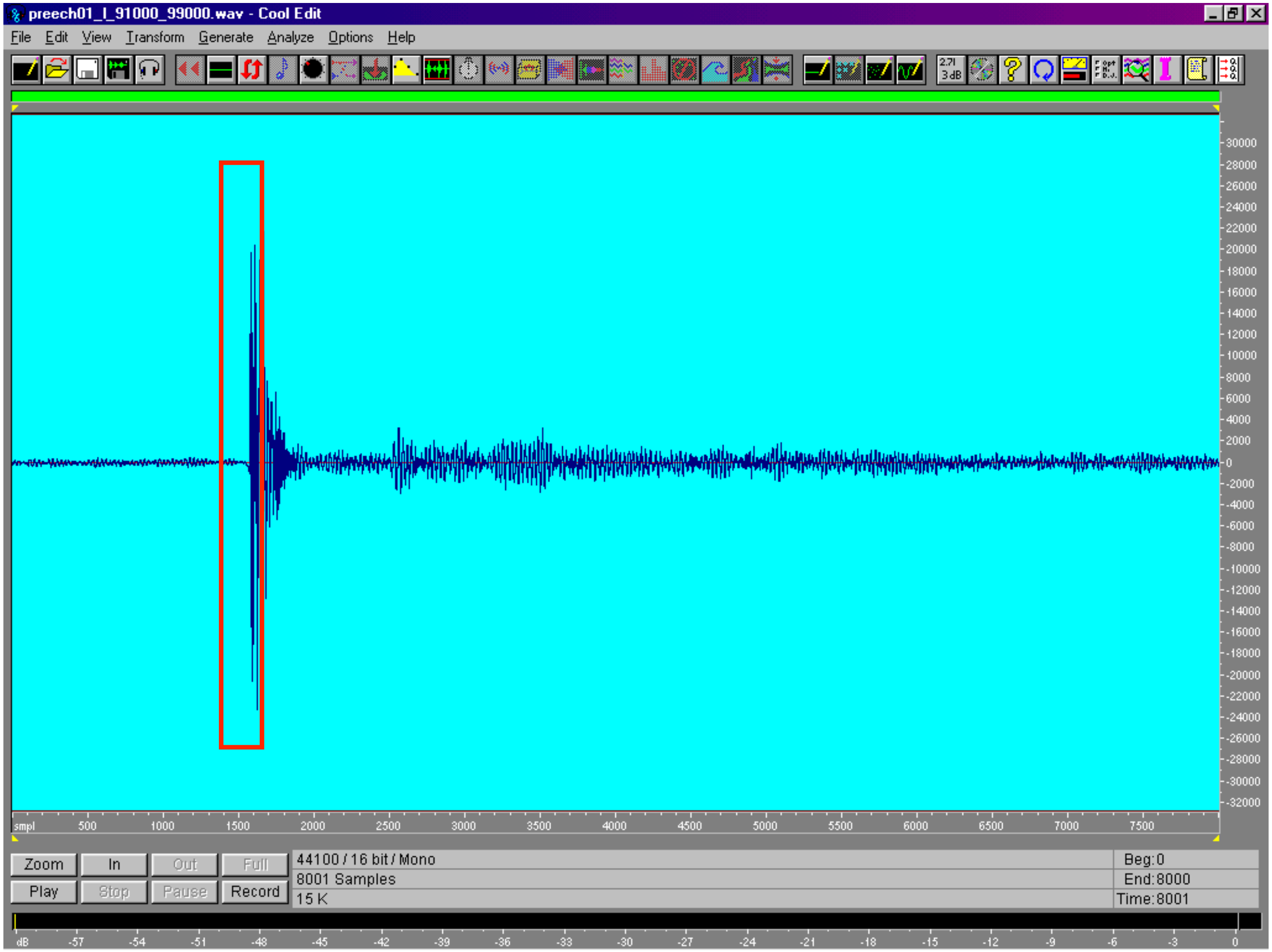
Where does the noise fall in time?



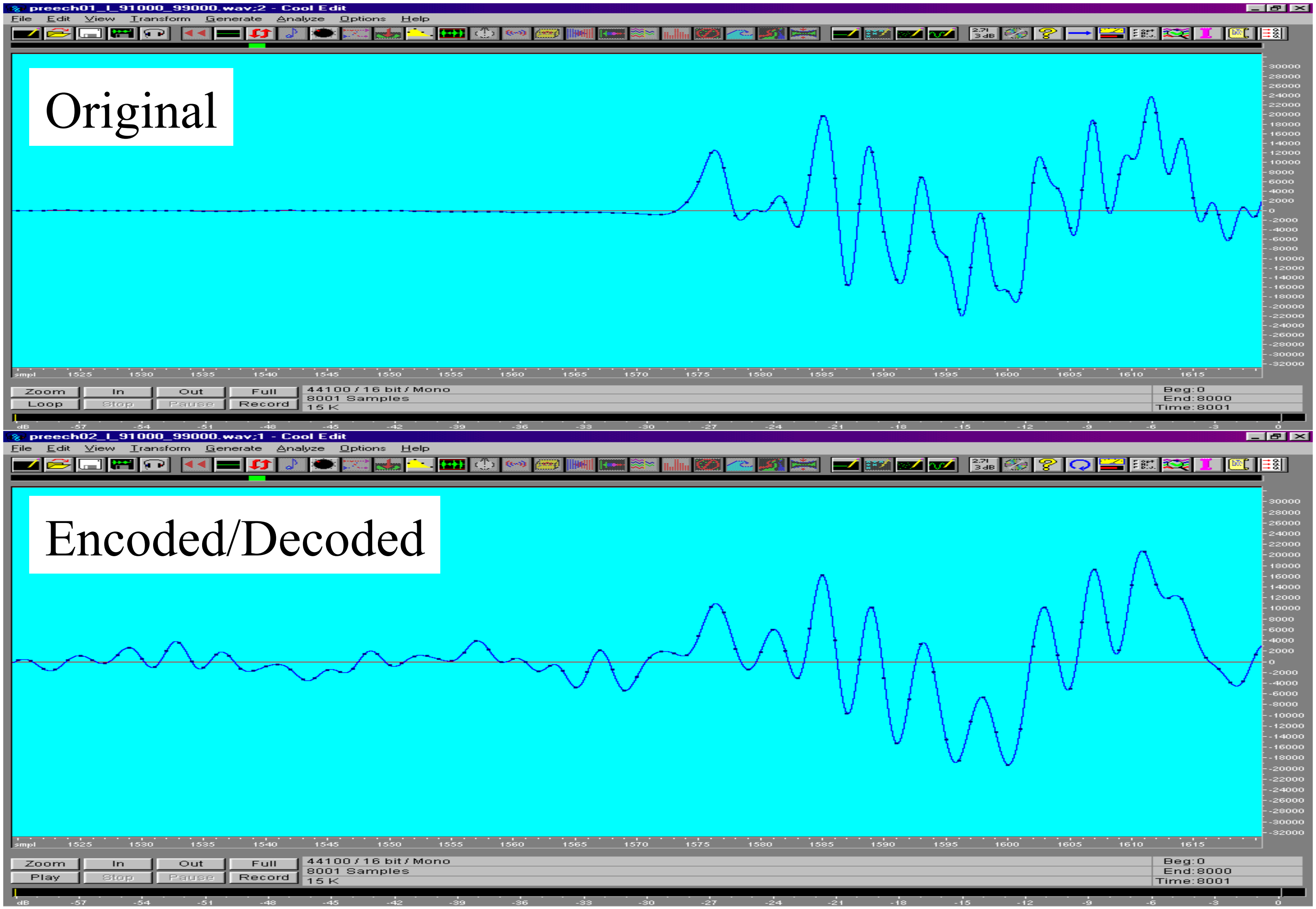
Original: Zoom in on one strike attack



Zoom in on attack



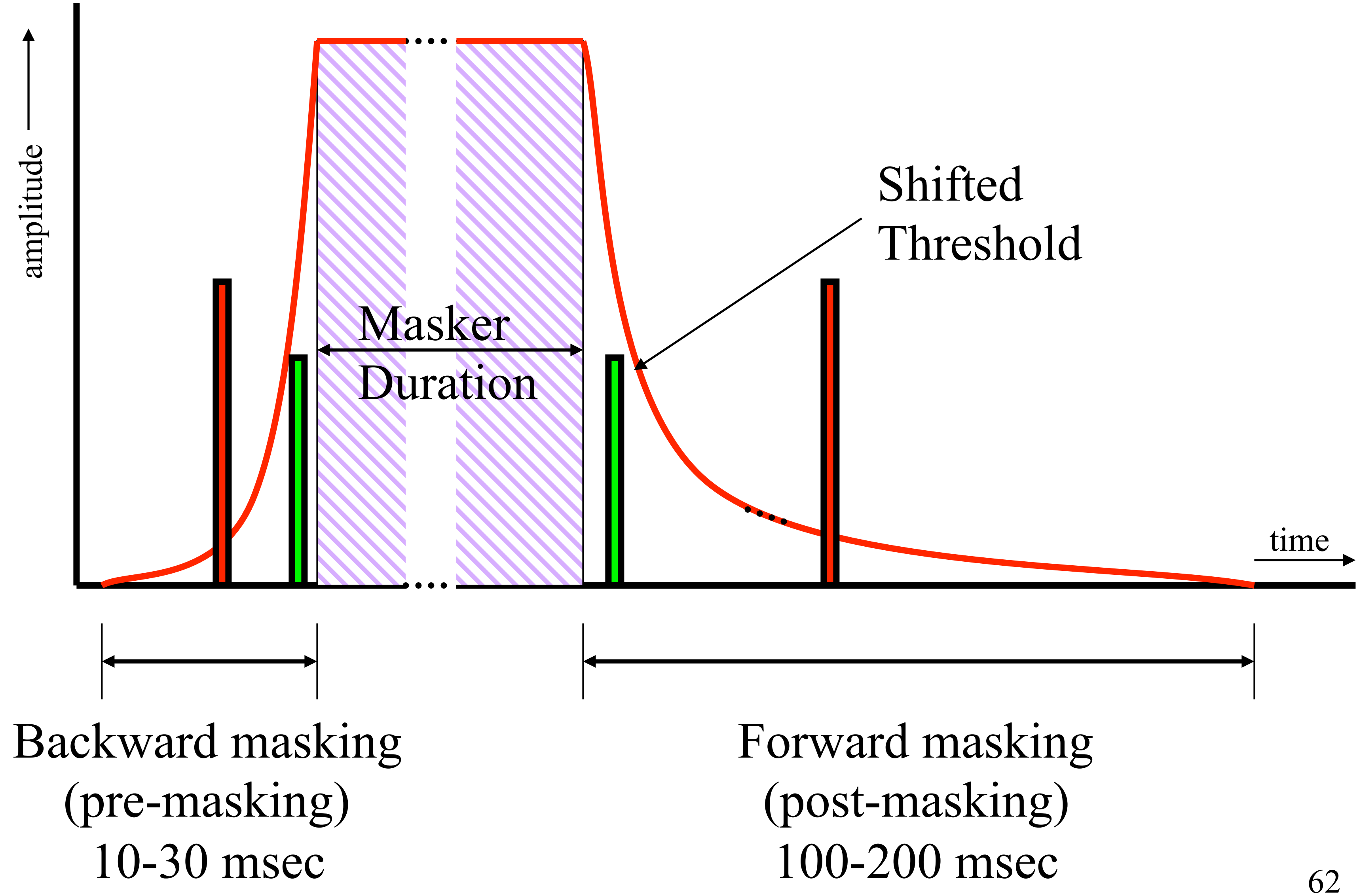
Attack, 2.2 msec (100 samples)



Samples 1520:1620 of preech0X_l_91000_99000.wav

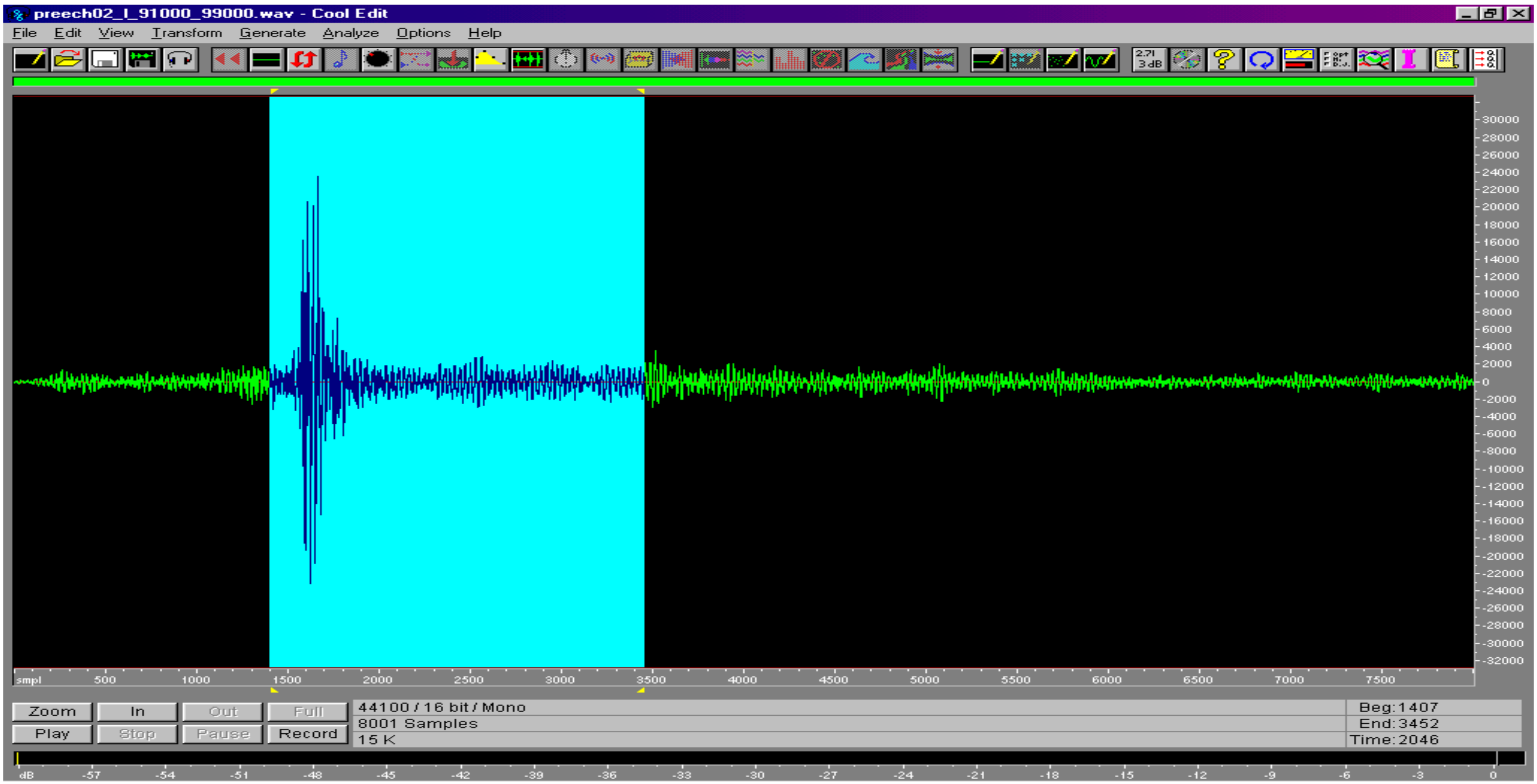
Temporal Masking

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



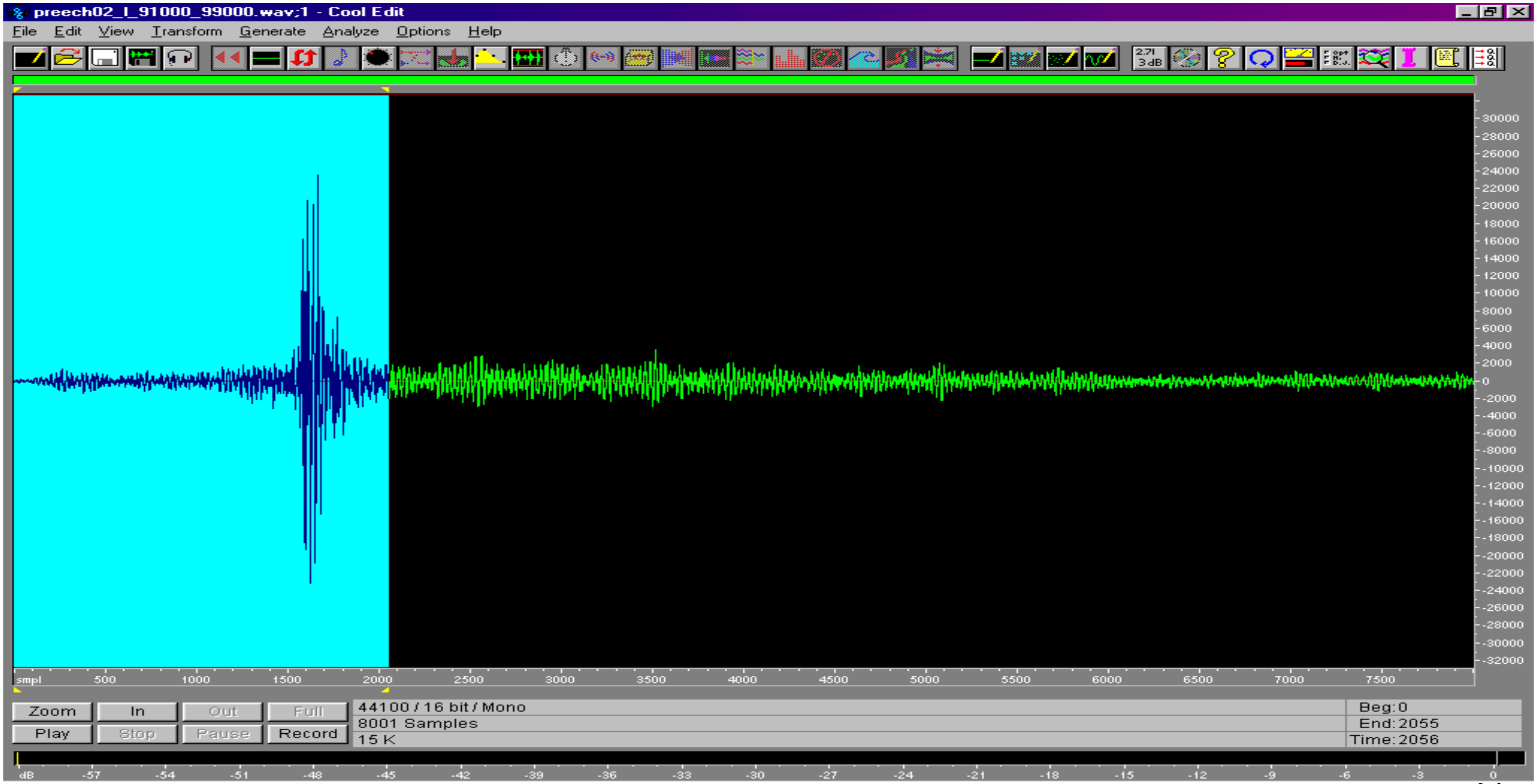
Where the Window Falls (1)

AAC Window width = 2048 samples = 46 msec

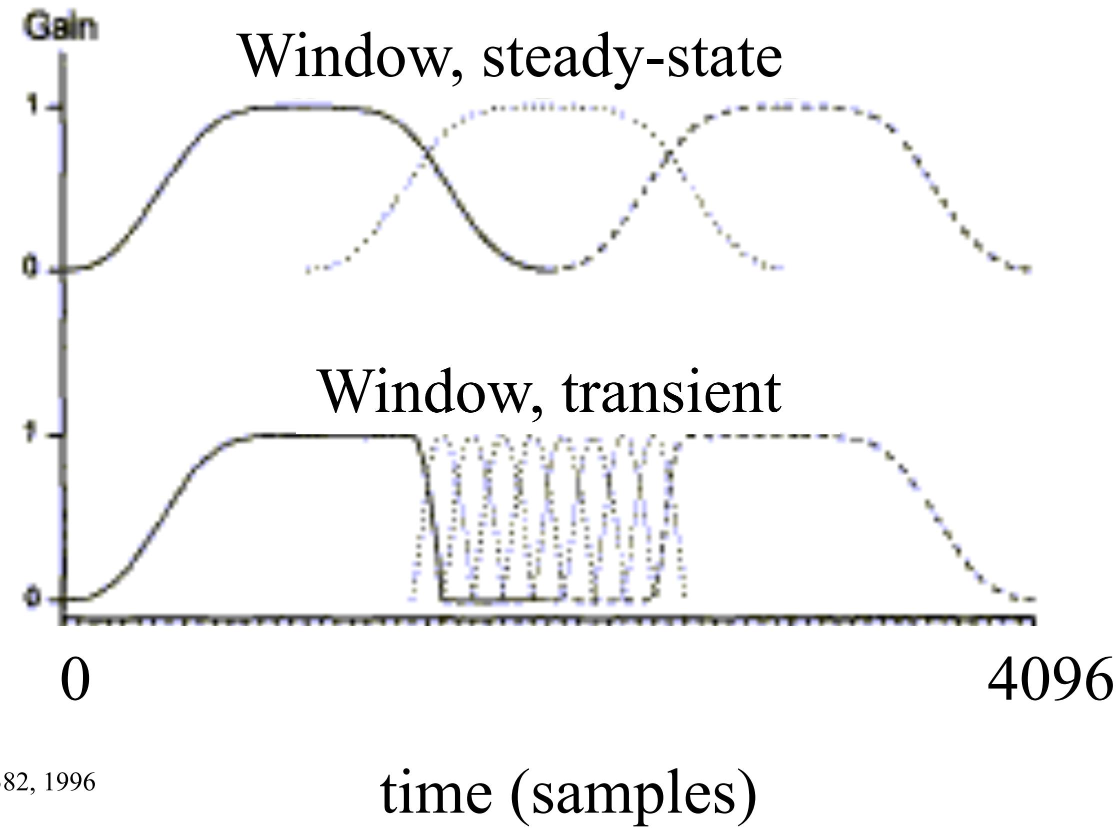


Where the Window Falls (2)

AAC Window width = 2048 samples = 46 msec



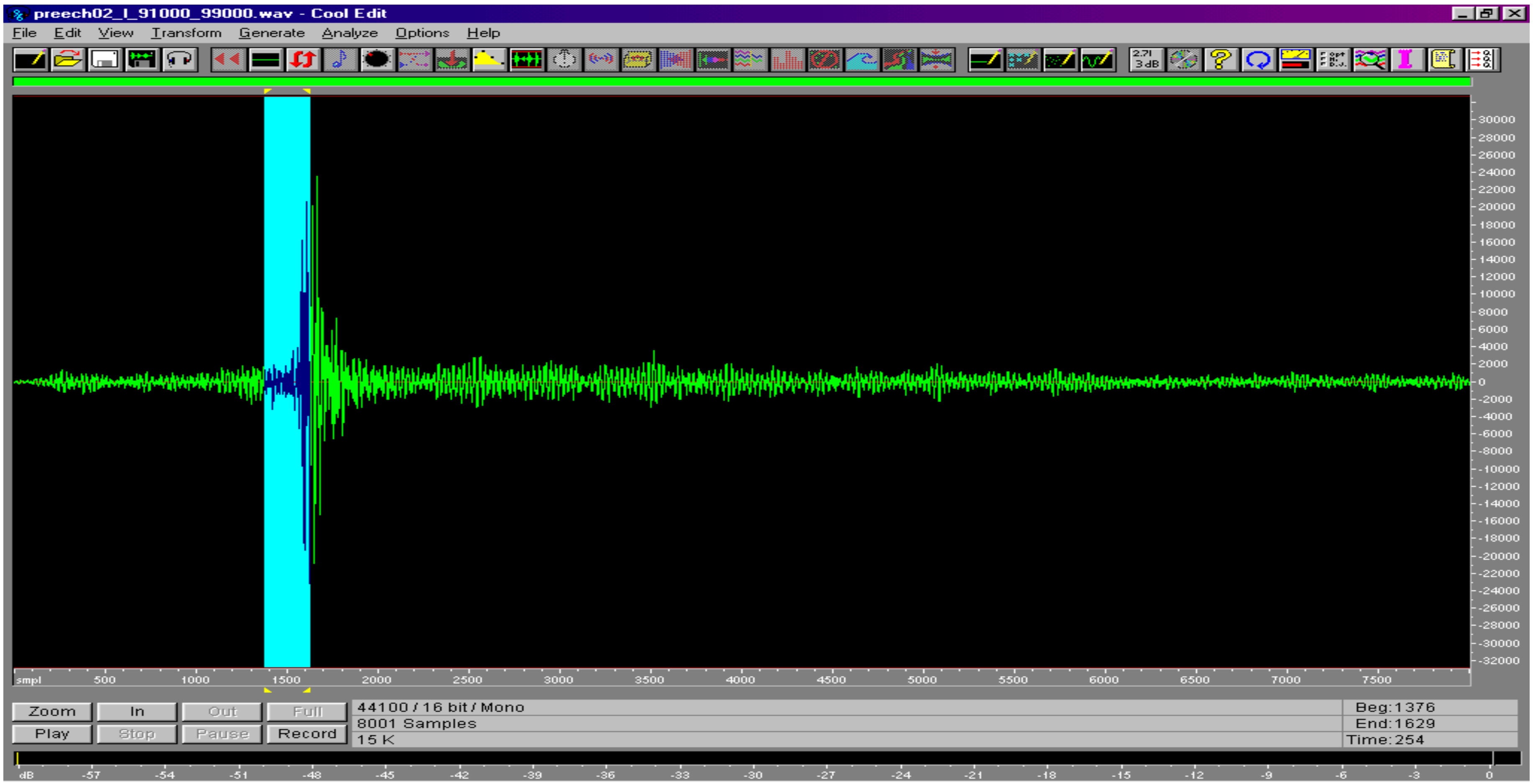
MPEG-2 AAC Windowing



From: Bosi et al, AES preprint 4382, 1996

Where the Window Falls (3)

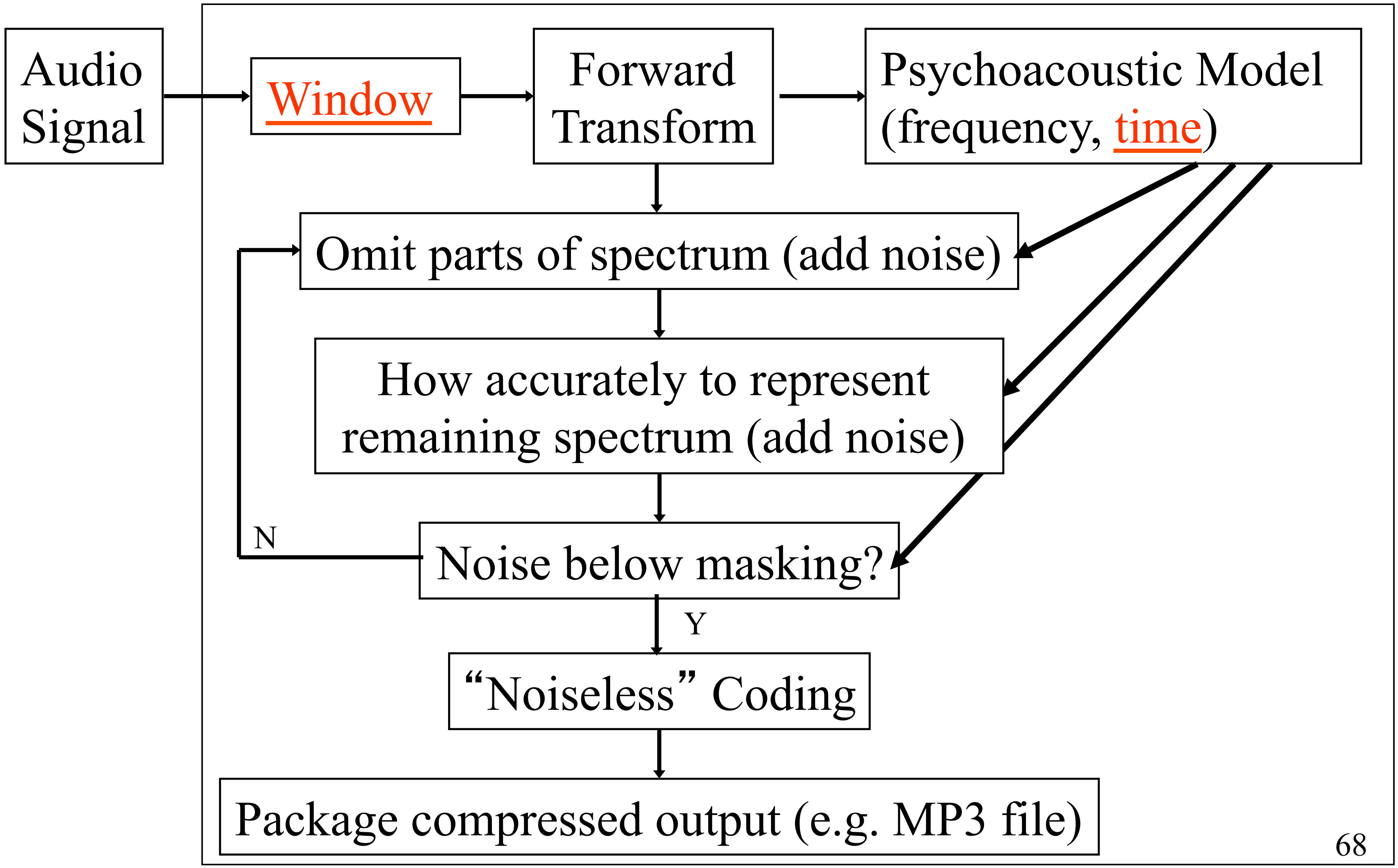
Window width = 256 samples = 5.8 msec



Two kinds of masking

- Simultaneous
- Temporal

Perceptual Encoder



MDCT, TDAC

Textbook DCTs

$$X(m) = \ldots \sum_{n=0}^{N} \ldots x(n) \cos\left[\frac{mn\pi}{N}\right] \qquad \text{I} \qquad x(n) = \ldots \sum_{m=0}^{N} \ldots X(m) \cos\left[\frac{mn\pi}{N}\right]$$

$$X(m) = \ldots \sum_{n=0}^{N-1} x(n) \cos\left[\frac{(2n+1)m\pi}{2N}\right] \qquad \text{II} \qquad x(n) = \ldots \sum_{m=0}^{N-1} X(m) \cos\left[\frac{(2n+1)m\pi}{2N}\right]$$

$$X(m) = \ldots \sum_{n=0}^{N-1} x(n) \cos\left[\frac{(2m+1)n\pi}{2N}\right] \qquad \text{III} \qquad x(n) = \ldots \sum_{m=0}^{N-1} X(m) \cos\left[\frac{(2m+1)n\pi}{2N}\right]$$

$$X(m) = \ldots \sum_{n=0}^{N-1} x(n) \cos\left[\frac{(2m+1)(2n+1)\pi}{4N}\right] \qquad \text{IV} \qquad x(n) = \ldots \sum_{m=0}^{N-1} X(m) \cos\left[\frac{(2m+1)(2n+1)\pi}{4N}\right]$$

Rao and Yip, *Discrete Cosine Transform*, p. 15

"Modified DCT" (MDCT)

$$X(m) = \sum_{n=0}^{N-1} x(n) \cos\left[\frac{\pi}{2N}\left(2n+1+\frac{N}{2}\right)(2m+1)\right]$$

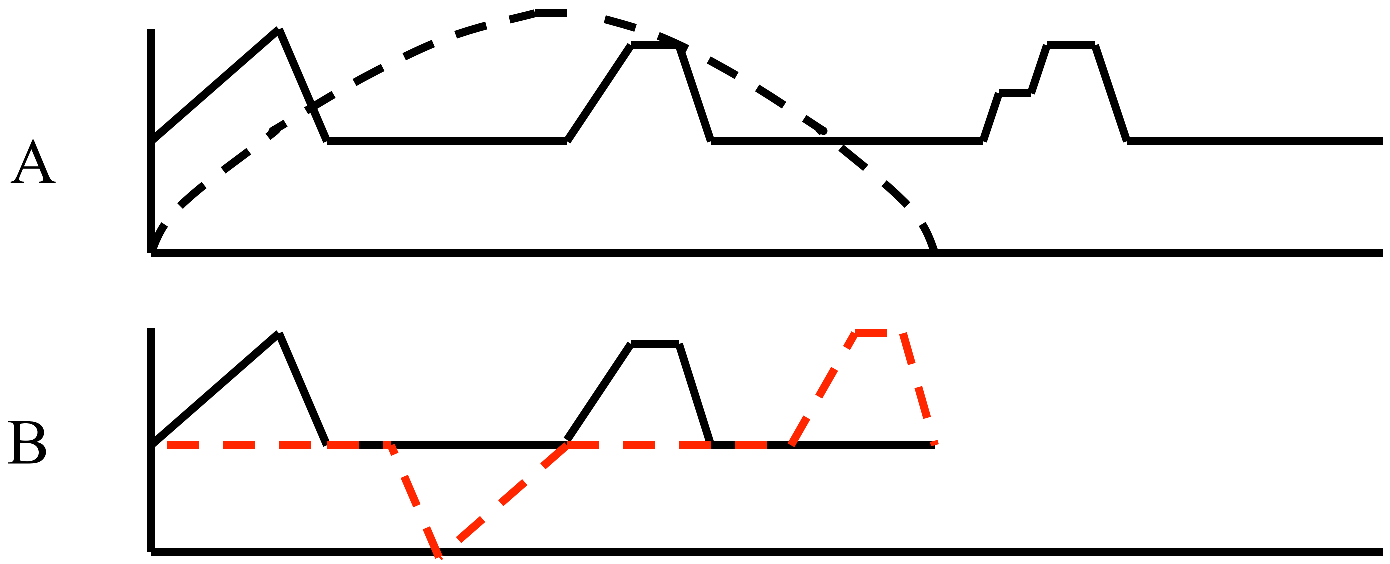
for 0 <= *m* <=*N/2*-1

$$x(n) = \sum_{m=0}^{\frac{N}{2}-1} X(m) \cos\left[\frac{\pi}{2N}\left(2n+1+\frac{N}{2}\right)(2m+1)\right]$$

for 0 <= *n* <=*N*-1

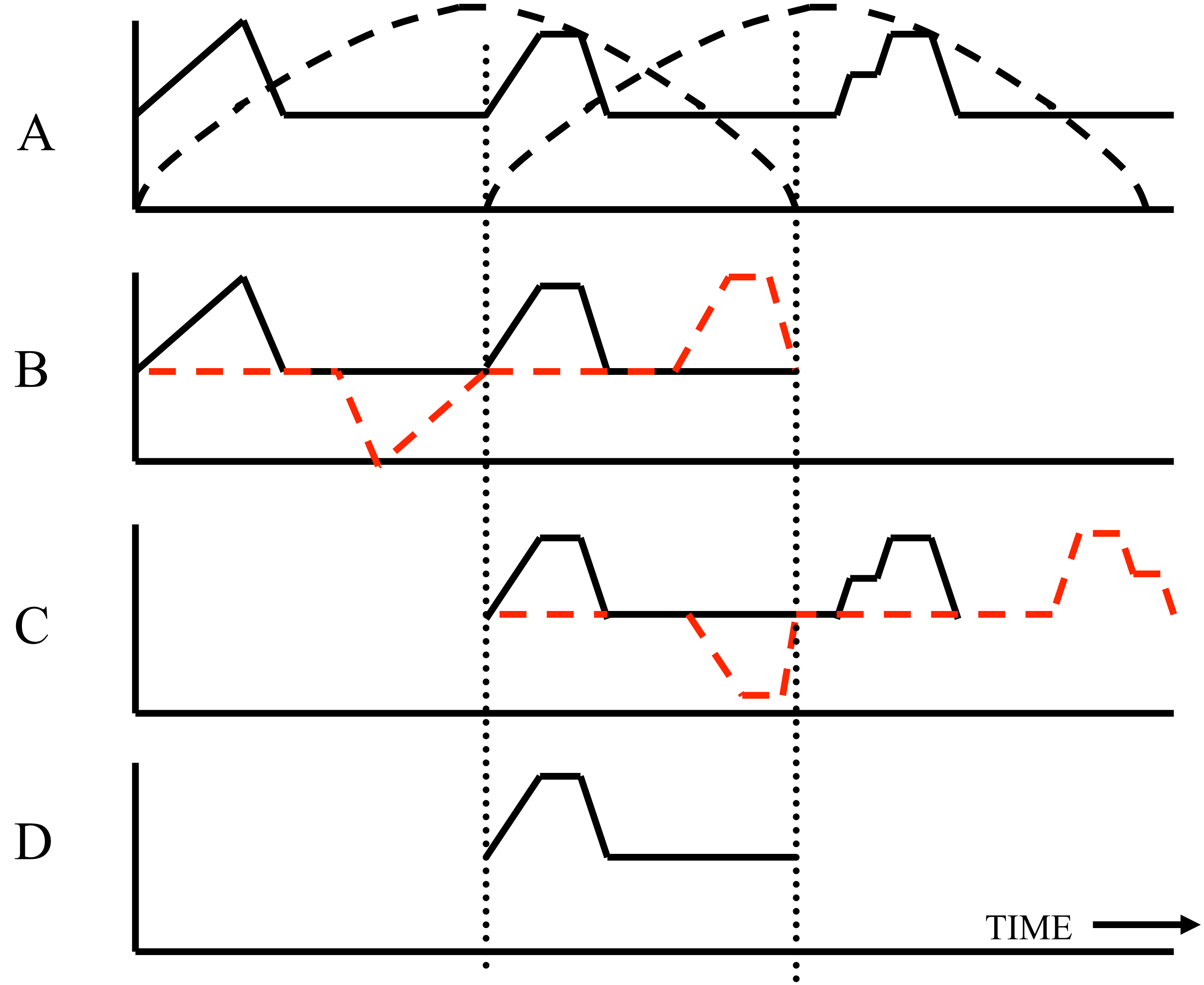
ISO 11172-3, MPEG-1 specification

Time-domain Aliasing



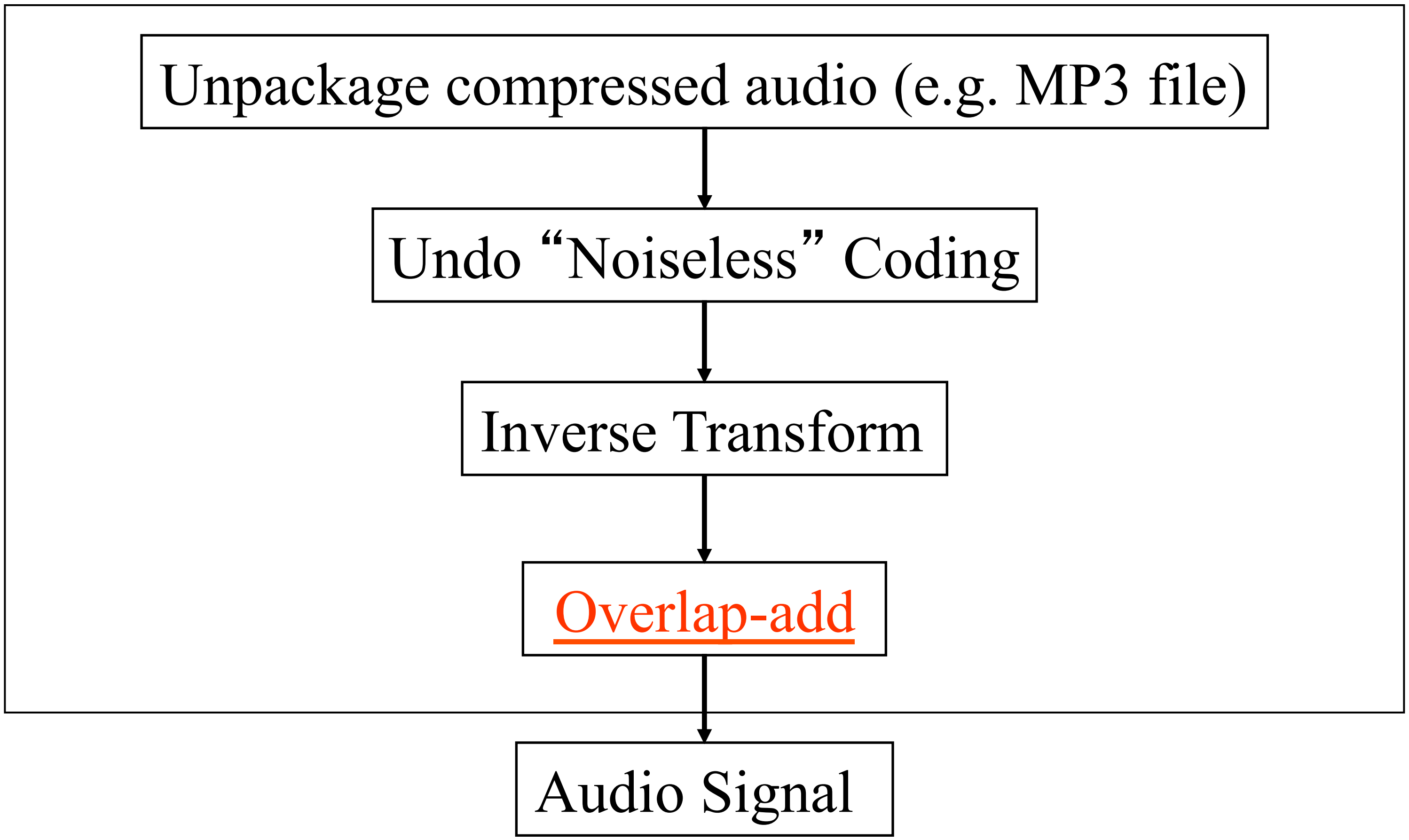
After Wang and Vilermo, JAES 51(1/2):54, 2003

Time-domain Aliasing Cancellation



After Wang and Vilermo, JAES 51(1/2):54, 2003

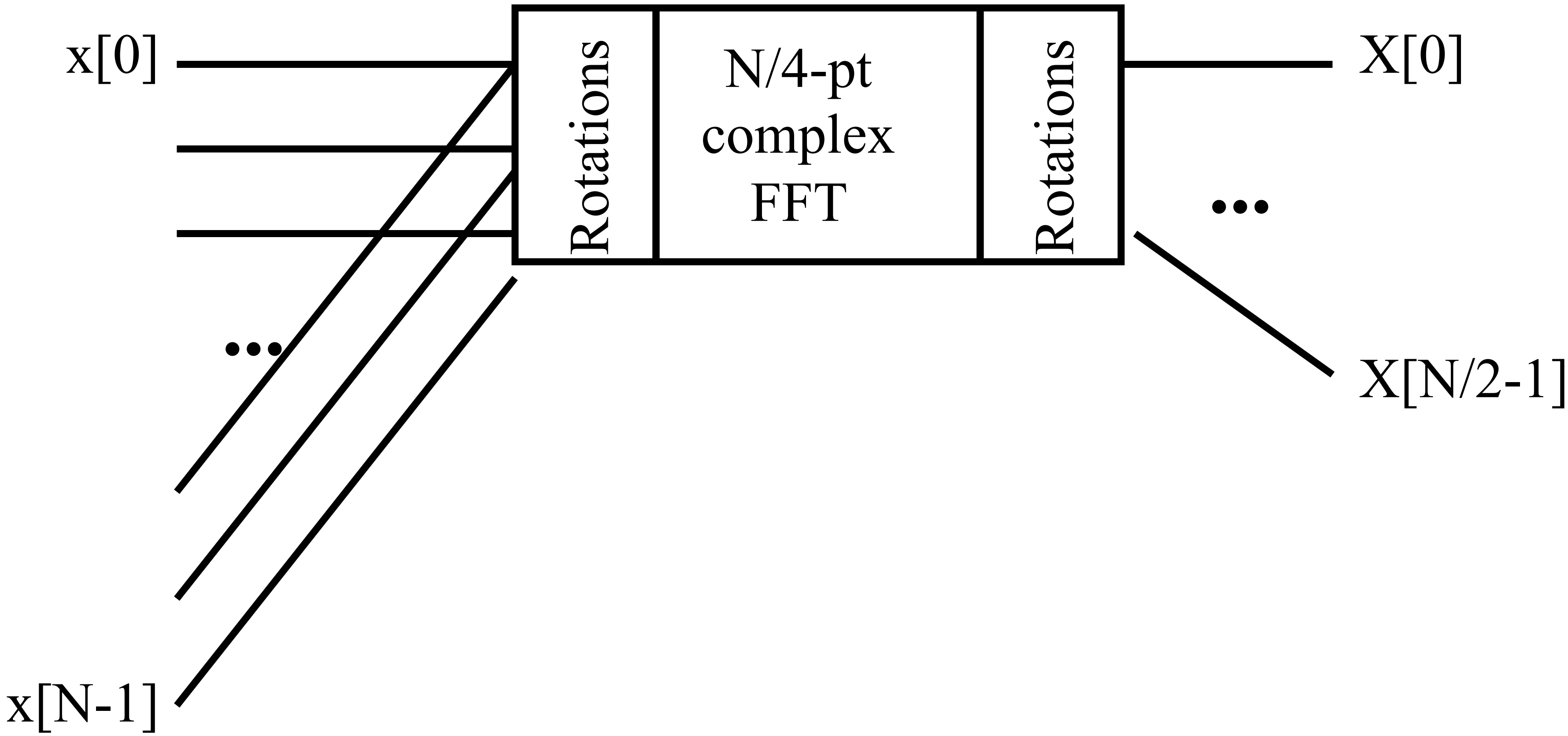
Decoder ("mp3 player")



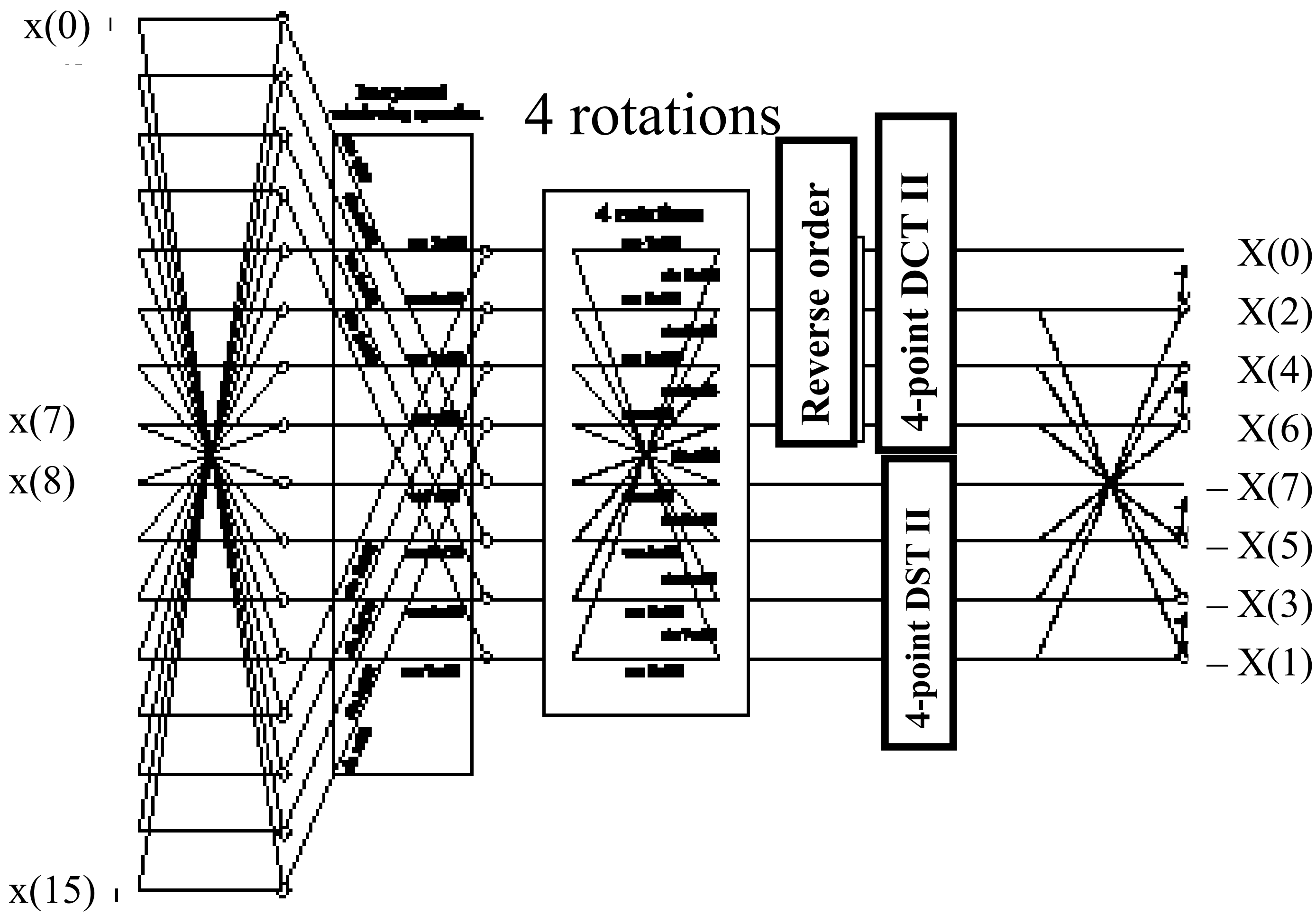
MDCT MIPs

AAC decoder	Initial	After optimize
Unpack bit stream	2.26	0.80
Huffman	4.30	4.30
Prepare spectrum	1.81	1.24
DCT	15.79	4.20
Window, Overlap/Add	2.92	0.26
Misc	2.17	1.50
Total MIPS	29.25	12.29

MDCT implemented with FFT

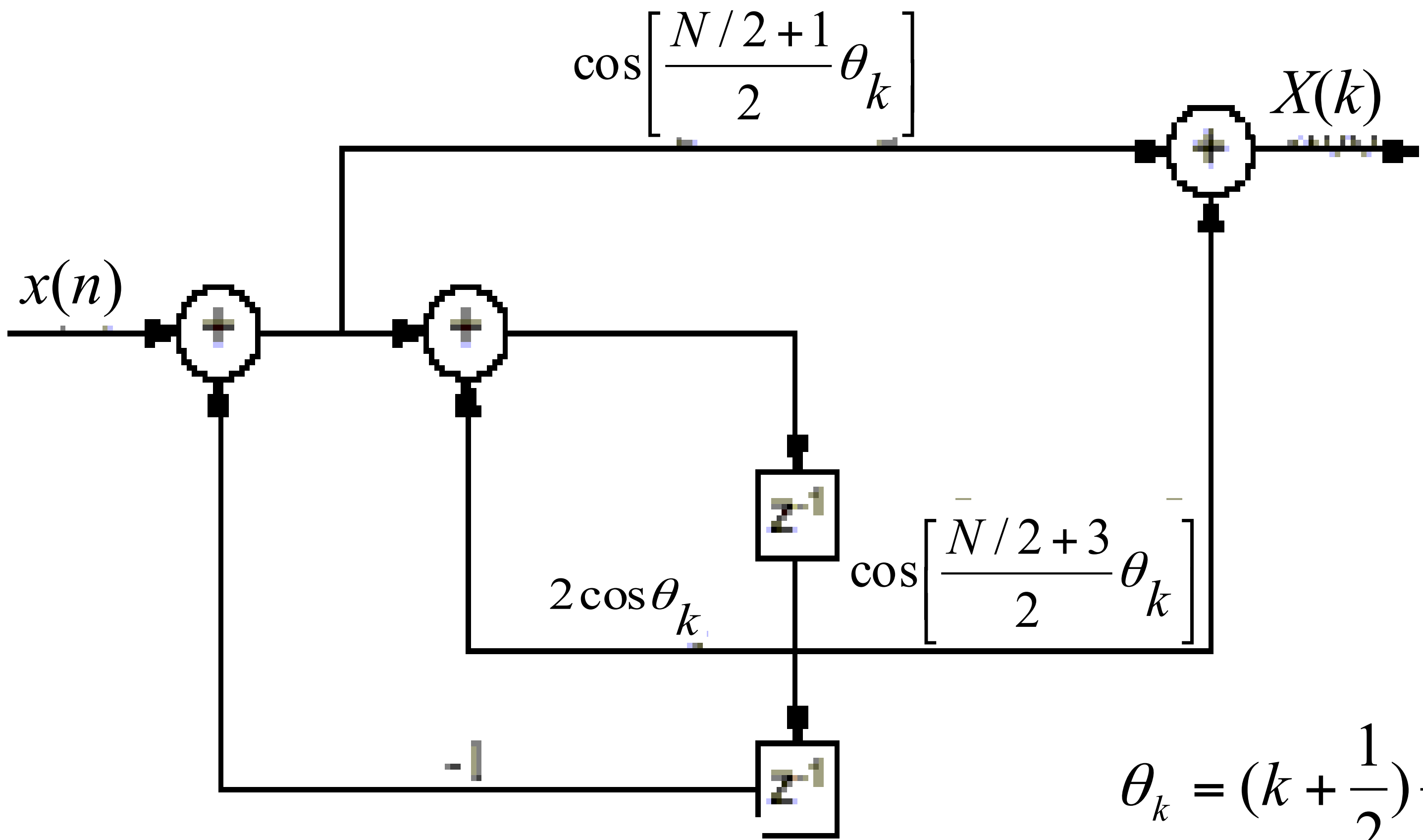


MDCT with shorter DCT/DSTs



Vladimir Britanak and K.R.Rao. "A new fast algorithm for the unified forward and inverse MDCT/ MDST computation." Signal Processing 82:433-459, 2002.

Recursive (regressive) MDCT



$$\theta_k = (k + \frac{1}{2})\frac{\pi}{N/2}$$

Nikolajevic, Vladimir, and Gerhard Fettweis. "New Recursive Algorithms for the Forward and Inverse MDCT." Proc. IEEE Workshop on Signal Processing Systems (SiPS). Antwerp, Belgium, 26.-28. September 2001.

Codec families

MPEG family overview

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

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
- Layer 3: higher frequency and time resolution; ISDN, Internet; most complex decoder
- Bit stream format same
- Layer 3 must decode Layer 2 ...

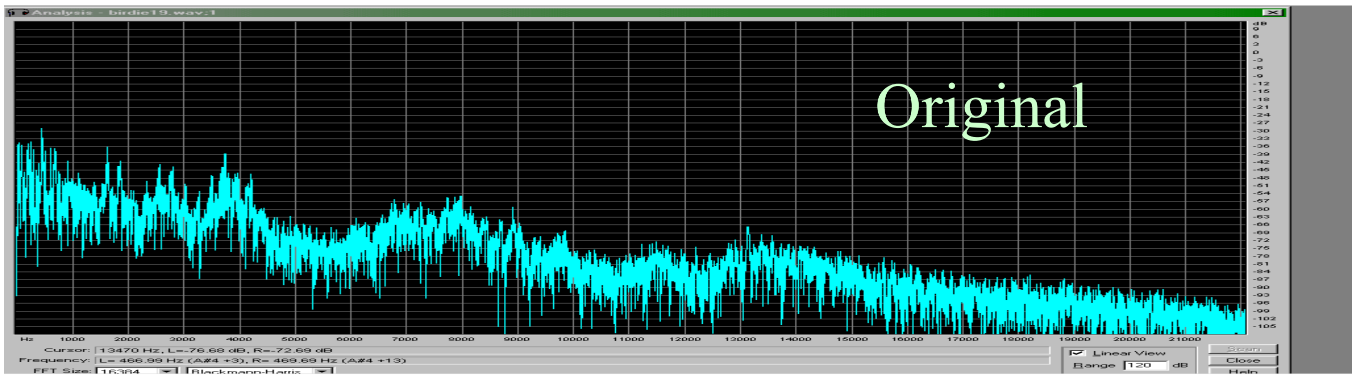
MPEG-2 Audio

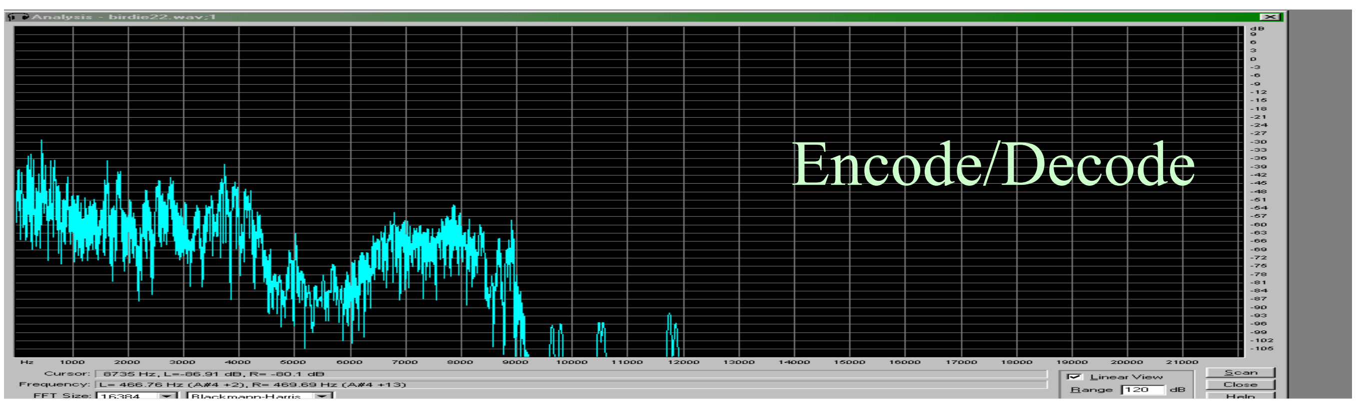
- 1994
- Motivation: video for digital TV
- Backward compatible with MPEG-1
  - Three layers, like MPEG-1
- Broader ranges
  - sample rates: 16, 22.05, 24 kHz
  - data rates: 8 - 1130 kbit/sec
  - channels: 5.1 + up to 7 multilingual/commentary channels
- “MP3” = MPEG-1/2 Layer 3

MPEG-2 Advanced Audio Coding (AAC)

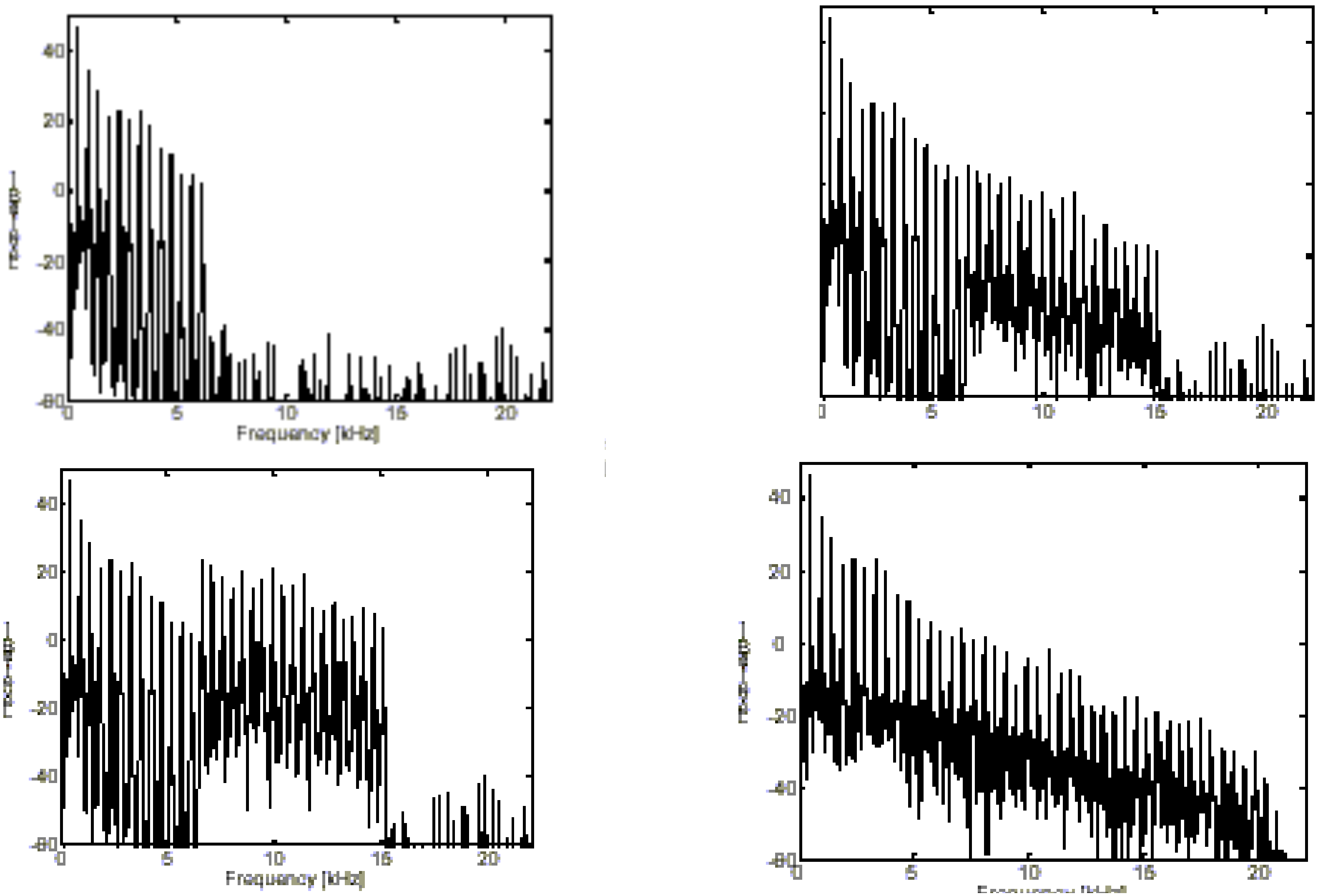
- 1997
- "Indistinguishable" at 384 kbit/sec
- Features:
  - "Non-backward-compatible" ("NBC")
  - Up to 48 channels (stereo, 5.1 …)
  - 8 - 96 kHz sample rate
  - Maximum 48-576 kbit/sec per channel
  - "Tools" combined into "profiles" (LC, SSR, Main)

Review: Spectrum modified by compression





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 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: audio in context

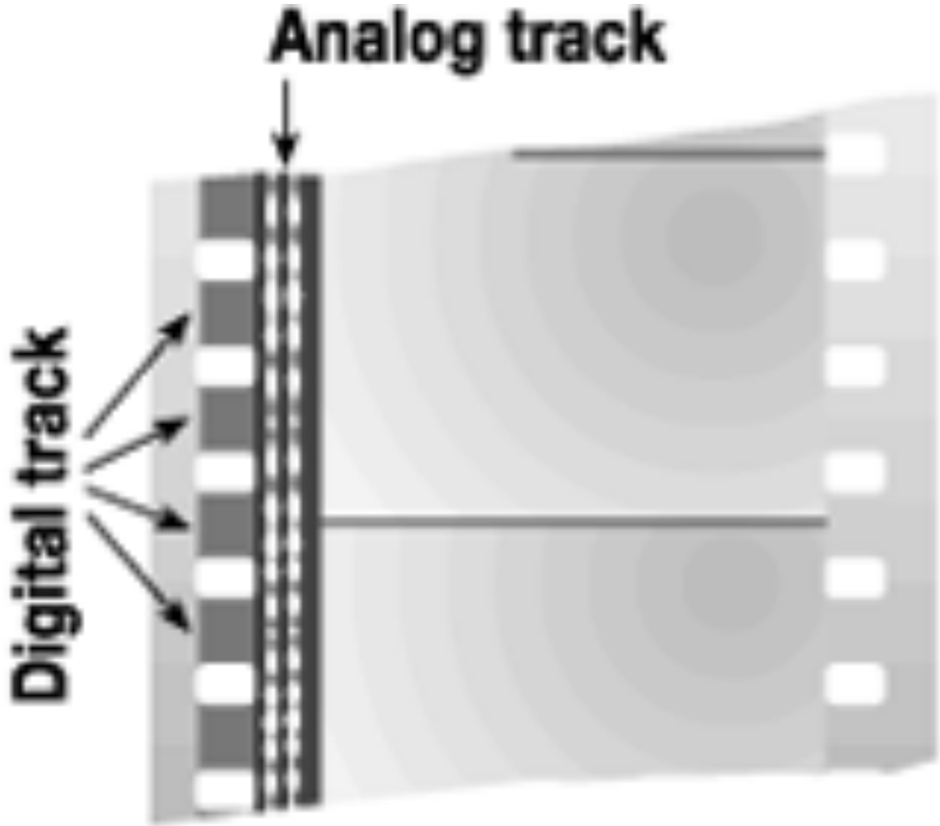
- Media object
- Scene
- Animation
- Interaction
- Fundamental advance over MPEG-1, -2

Algorithms related to Cinema

- DTS
- Sony ATRAC (1992) / SDDS
- Dolby AC-3

Dolby Digital (AC-3)

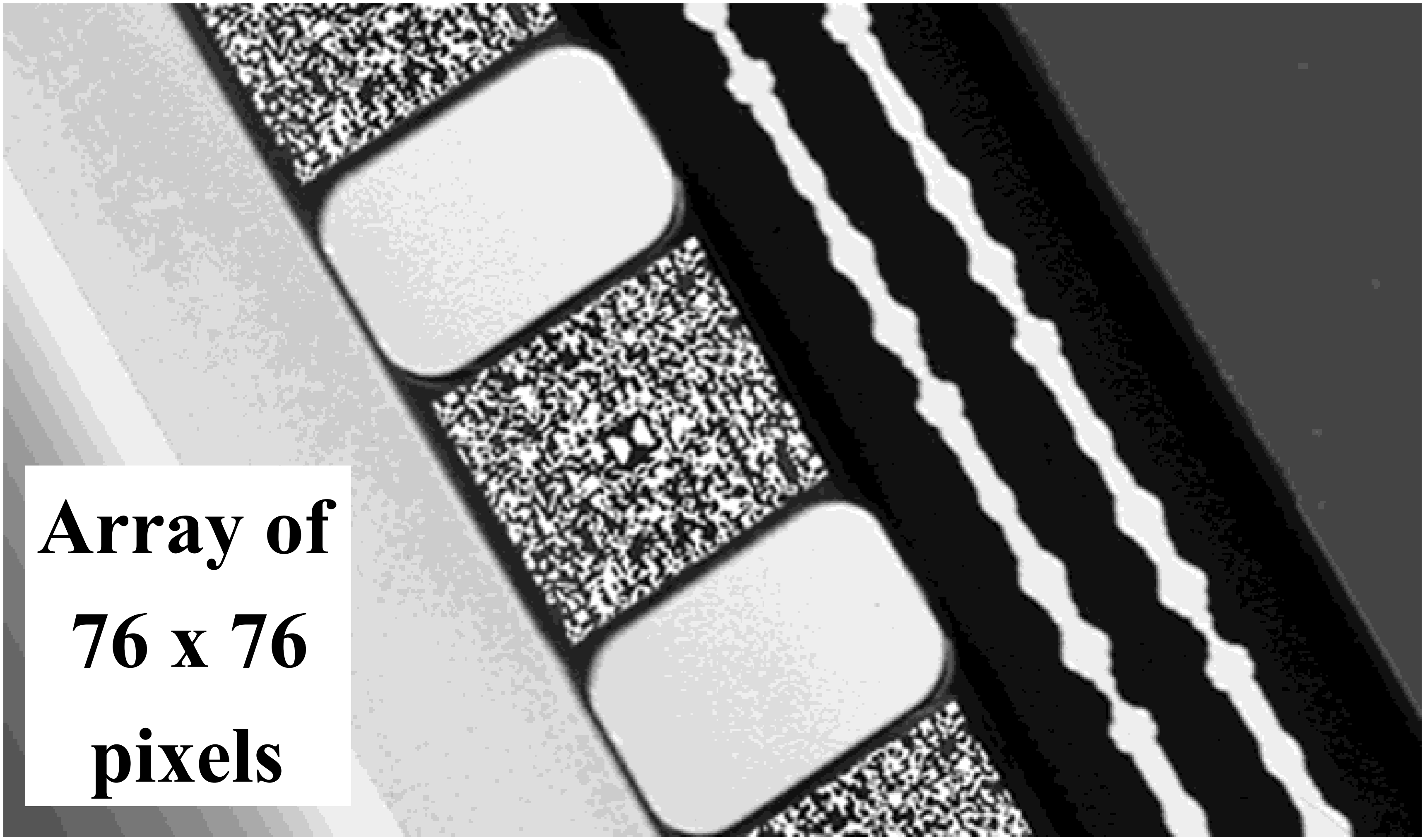
- 1993
- up to 5.1 channels
- MPEG-1 sample rates: 32, 44.1, 48 kHz
- 3 quality levels: 16-, 18-, and 20-bit
- 32 - 640 kpbs
- Applications
    - Cinema
    - DVD in NTSC countries
    - US digital television



Dolby Digital release print

Originally taken from http://www.dolby.com/tech/m.ot.0006.FilmIndFAQ.html

• Dolby Digital Film Format



Courtesy Dolby Laboratories Inc.

A few other algorithms

- WMA
- Qdesign, QuickTime
- Lucent PAC, 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 (www.monkeysaudio.com)
- Free Lossless Audio Codec (flac.sourceforge.net)
- Shorten File (.shn) (www.etree.org/shncom.html)
- DAKX
- …

Selected Topics

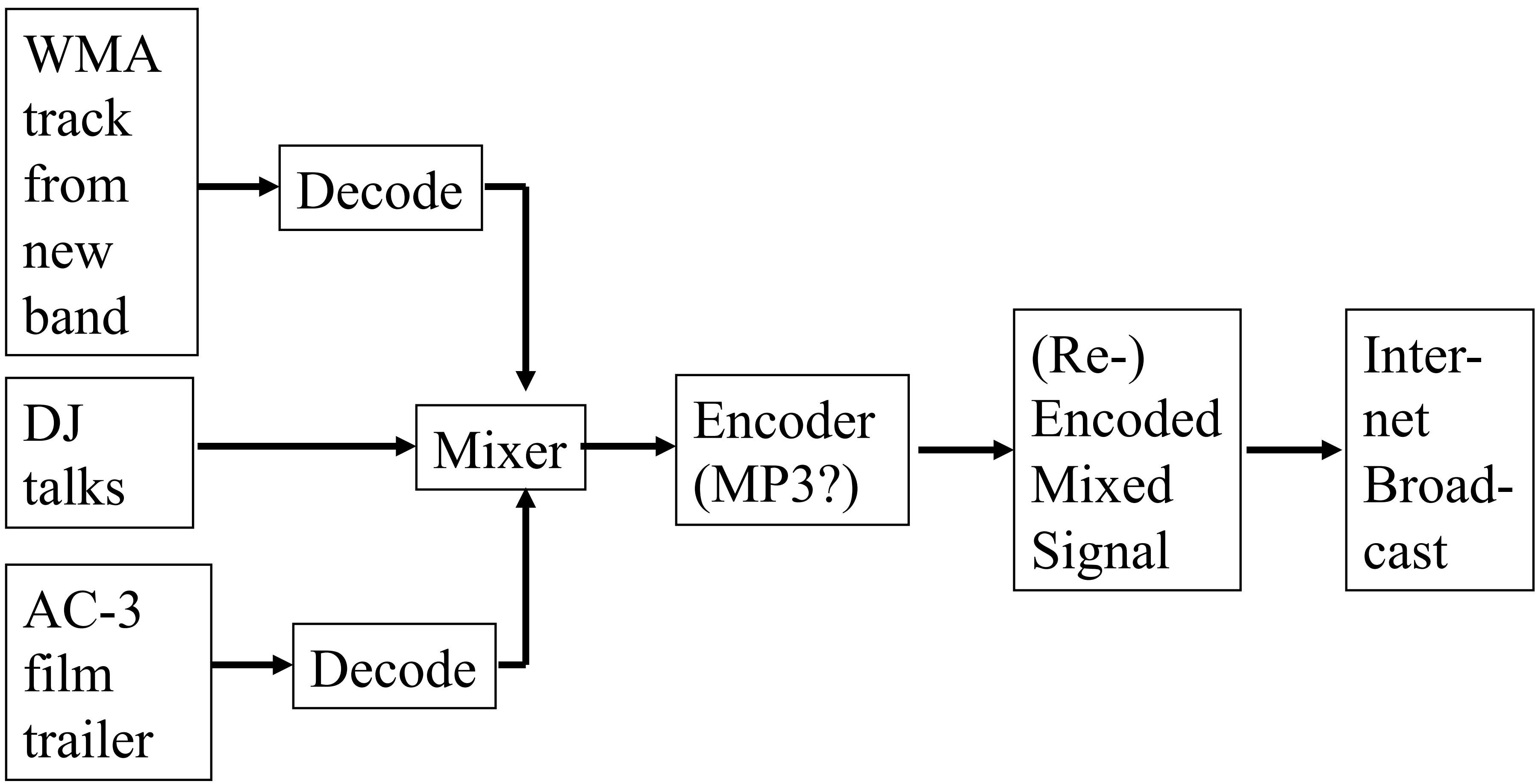
- Tandem Coding
- (name to be revealed)
- Classifying codecs: Data Rates
- Evaluating Codecs

Tandem Coding: What do you hear?

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

Source: AES CD-ROM.

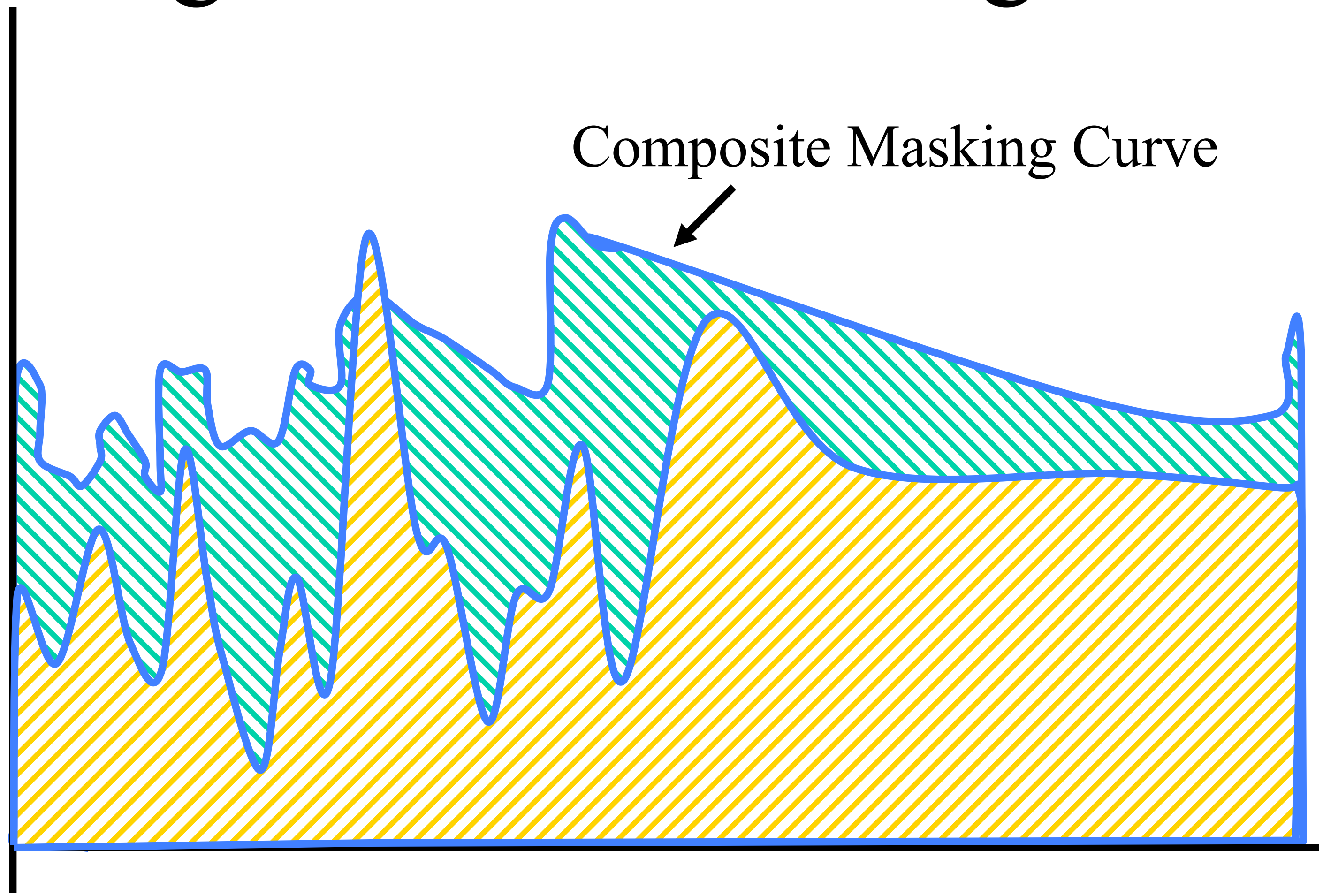
Possible Broadcast Scenario



Tandem Coding

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

Masking: Where noise might fall



After Davidson et al., 1994

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 (example: Meridian Lossless Packing, MLP)

Selected Topics

- Tandem Coding
- (name to be revealed)
- Classifying codecs: Data Rates
- Evaluating Codecs

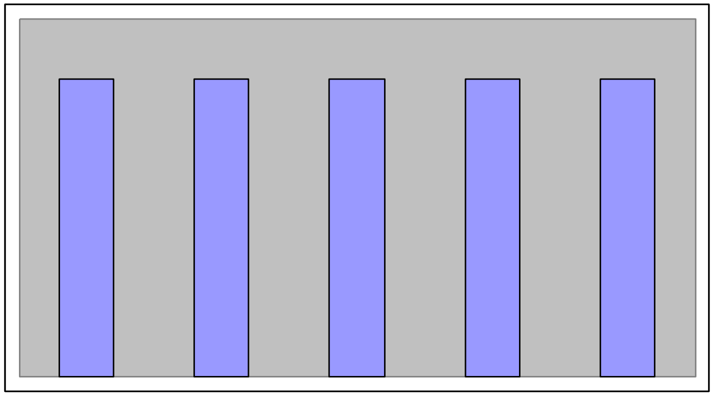
What do you hear?

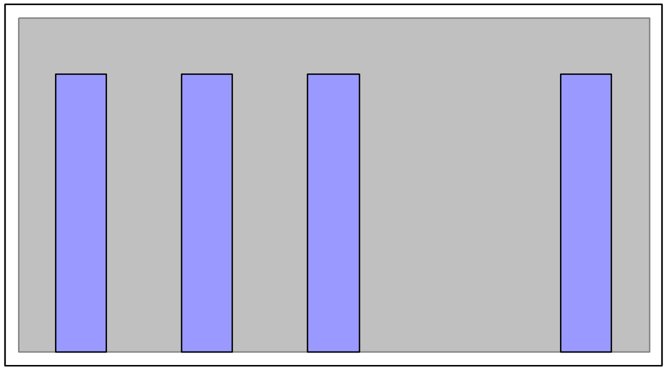
- Gilmour example
  - Original
  - Encoded/decoded
  - Zoom in on original: "Thing I"
  - Zoom in, encoded/decoded

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

Birdies: artificial examples

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

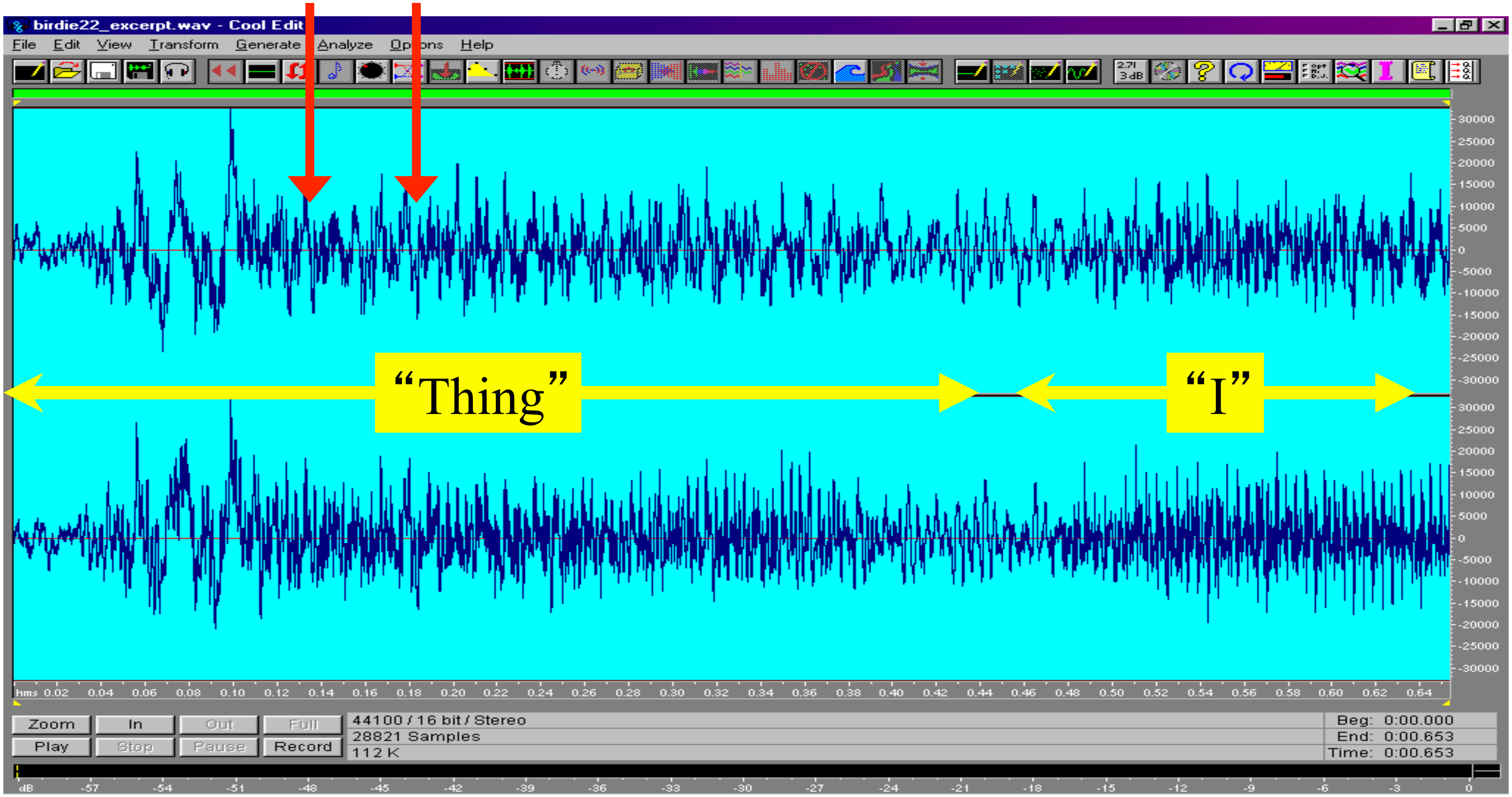




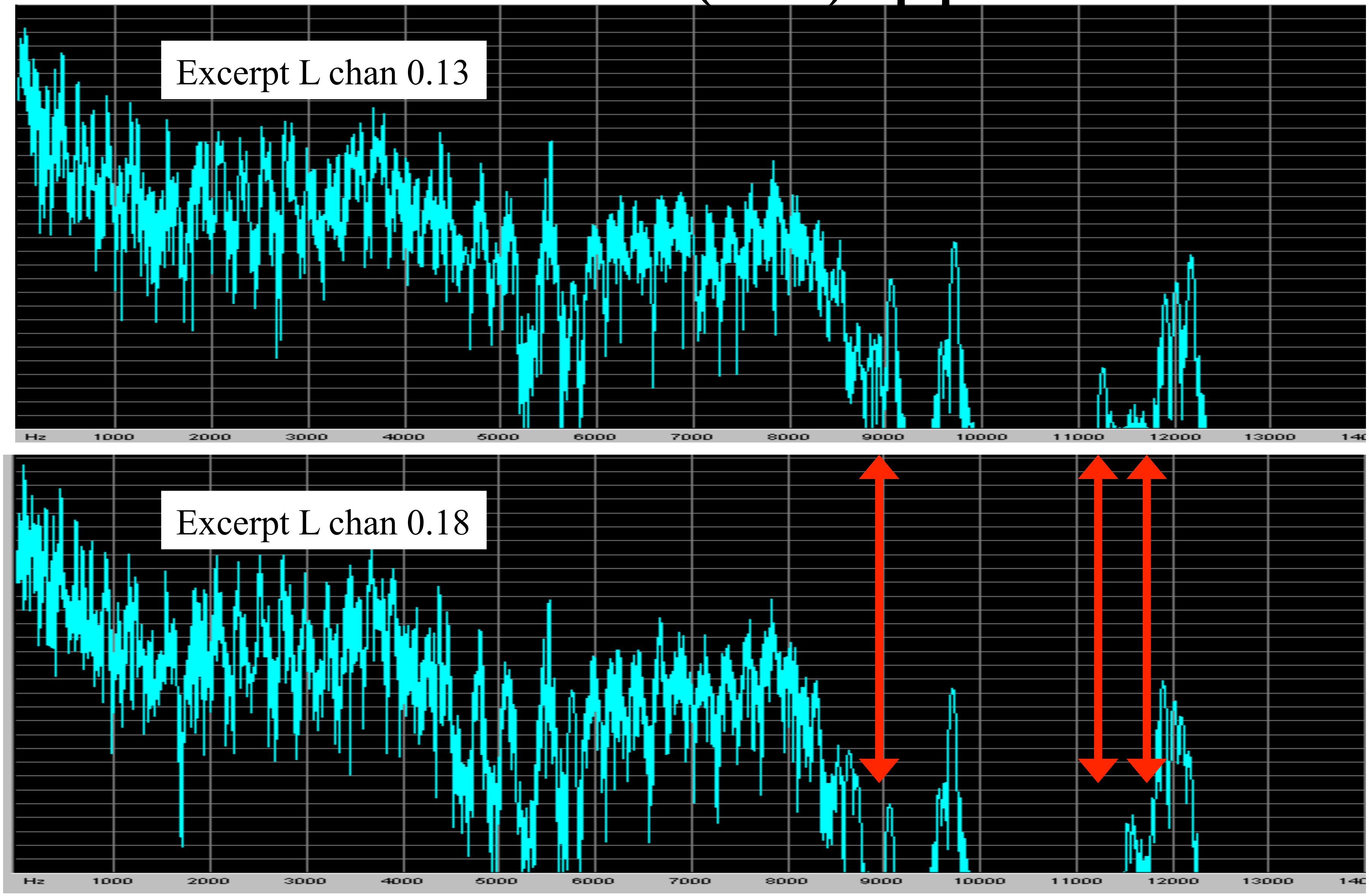
Source: AES CD-ROM.

Birdies: excerpt

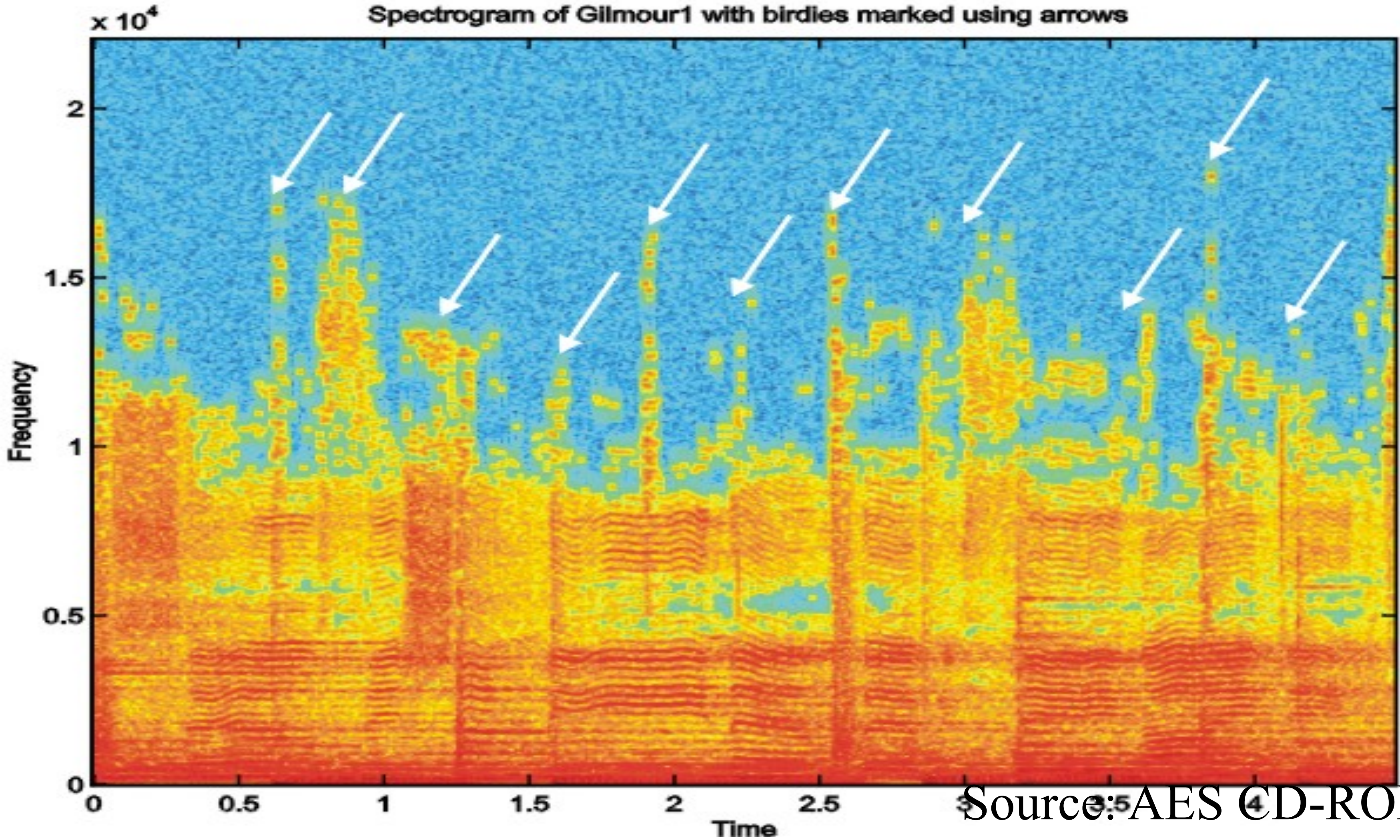
t=0.13 t=0.18



Birdies: Sudden (dis)appearance



Birdies: Even More Sudden (dis)Appearances



Source: AES CD-ROM

Birdies: Solutions

- Increase the bit rate.
- Reduce the complexity of the signal.
- Try different encoders.
- Try different settings on your encoder.

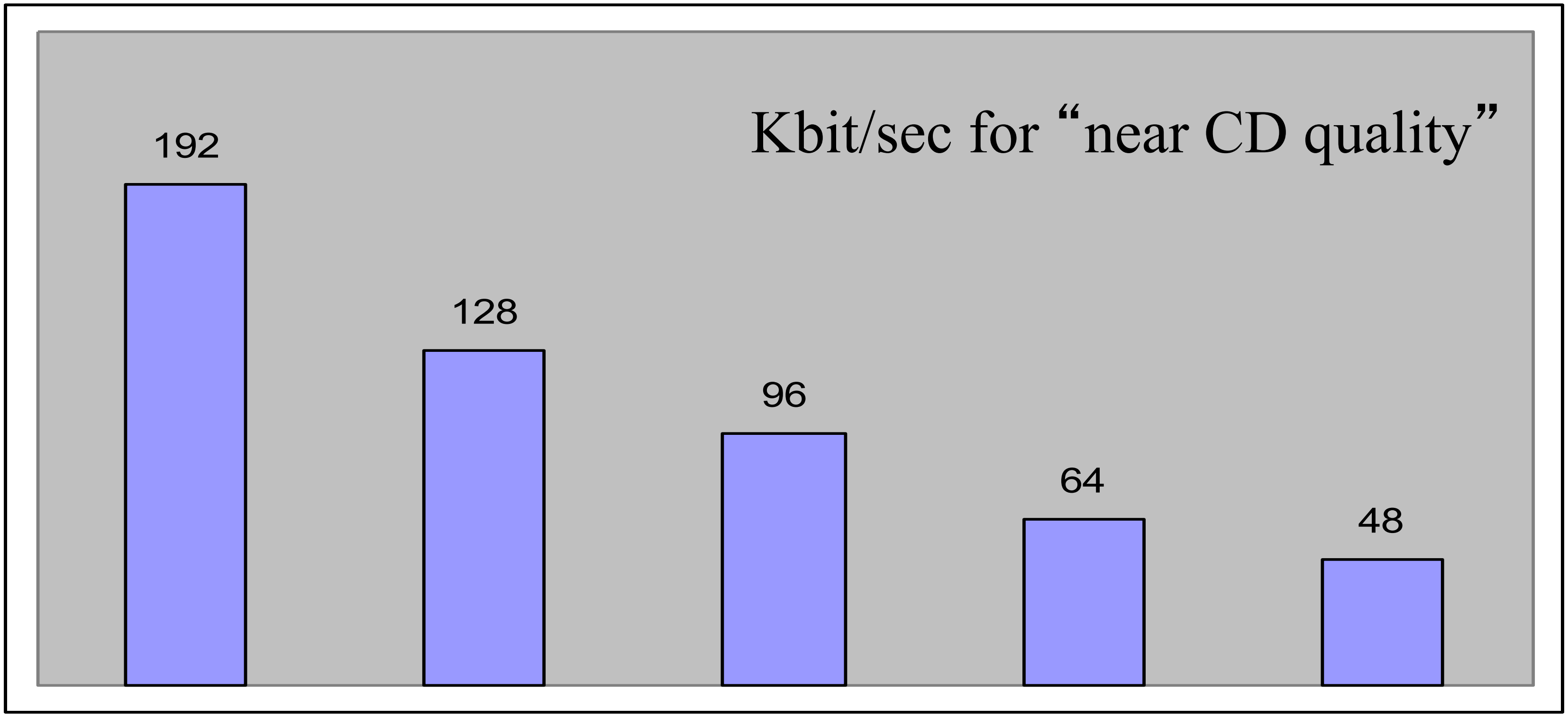
Selected Topics

- Tandem Coding
- Birdies
- Classifying codecs: Data Rates
- Evaluating Codecs

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

Selected Topics

- Tandem Coding
- Birdies
- Classifying codecs: Data Rates
- Evaluating Codecs

Measuring audio

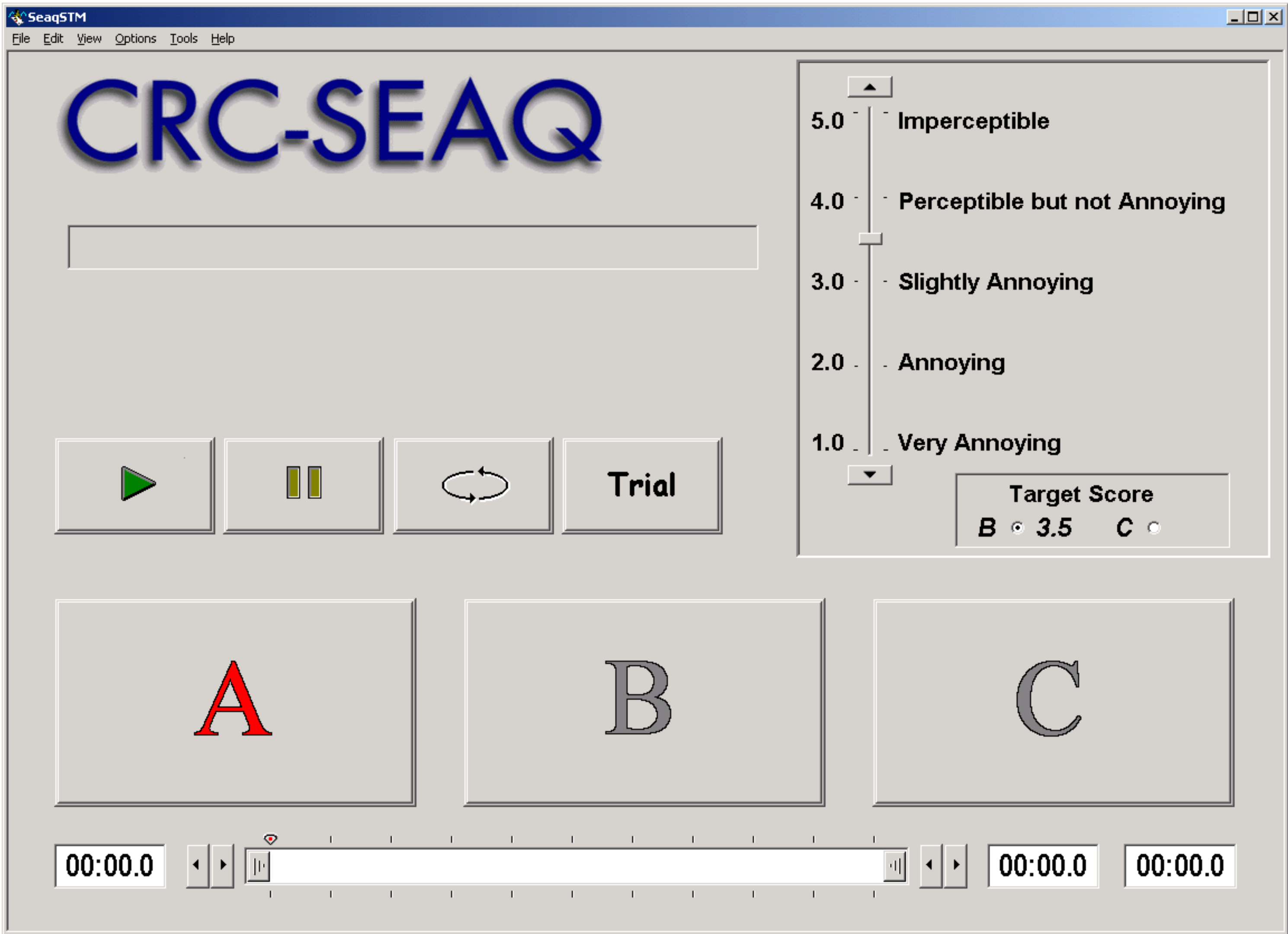
- Uncompressed audio
- Compressed audio

Subjective Compression Evaluation

- ABC listening test: correctly identify A as B or C
- Impairment scale

  0: “difference imperceptible” ... through ...

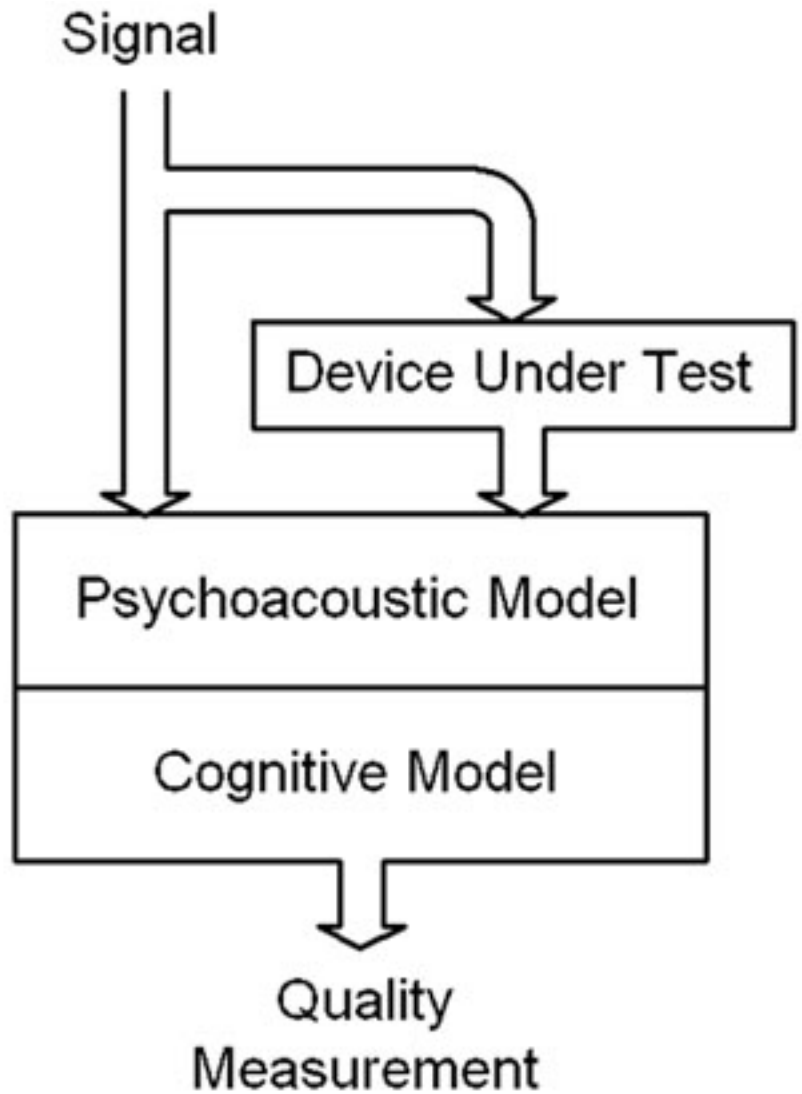
  -4: “difference very annoying”
- ITU-R BS.1116, ITU-R BS.562-3



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 <u>class of five</u> tested in independent trials by Moulton Laboratories. In this test, XXX at <u>96 kbit/sec</u> outperformed the MPEG-2 Advanced Audio Coder (AAC). At <u>96 kbit/sec, XXX</u> also outperformed <u>AAC at 128 kbit/sec</u> based on a repeatable statistical score." (XXX' s press release)

Comparing Codecs

Alg/kbps	ITU score	Characterization	
		Diff perceptible?	Diff "annoying?"
AAC/128 AC-3/192	– 0.5	Yes	Not "annoying"
XXX/160	– 0.8	Yes	"
AAC/96, XXX/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"
XXX/64	– 3.0	Yes	"very annoying"

Taken from Soulodre et al., JAES 1998

What can go wrong?
How to fix it?

What can go wrong?

- Added noise is audible.
- Pre-echo.
- Stereo field collapse.
- Dropouts, fadeouts.
- Example too complicated to encode w/in chosen parameters (--> birdies)

How do we fix it? (1)

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

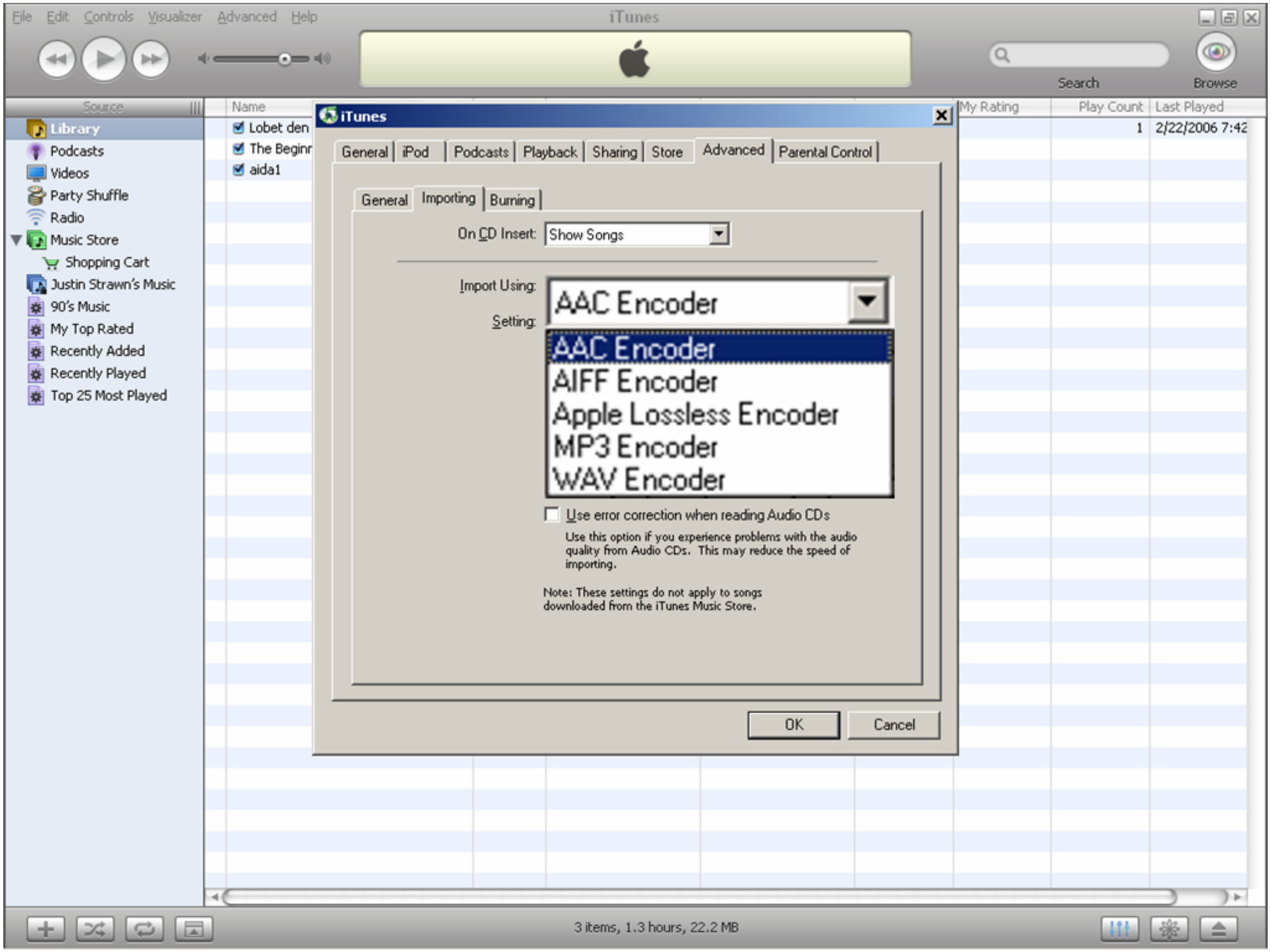
How do we fix it? (2)

- 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.

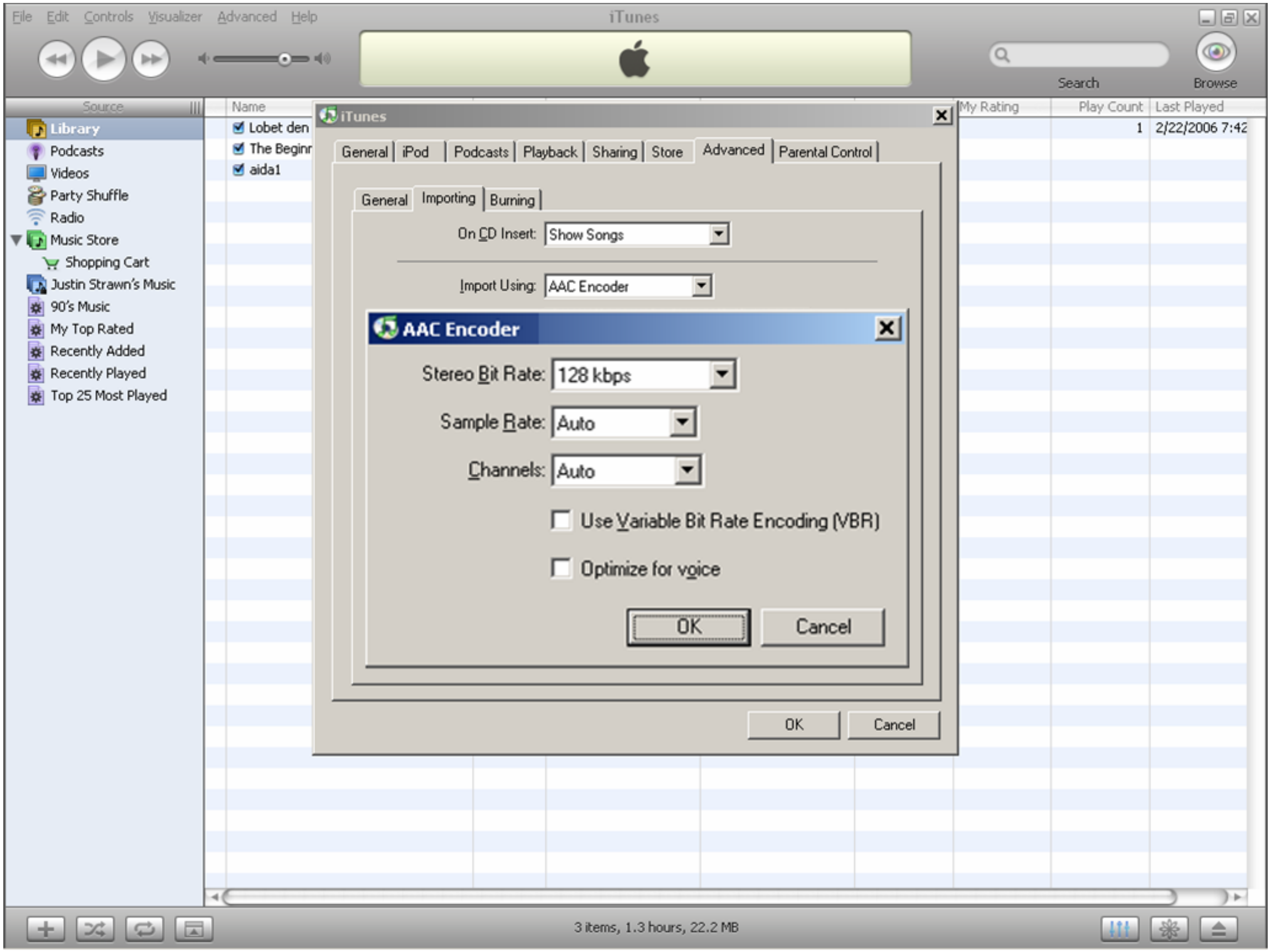
How do we fix it? (3)

- Alter mix to minimize problems.
- Avoid heavy limiting/compression ("wall of sound")
- Decoded output is pure noise? Try swap bytes.

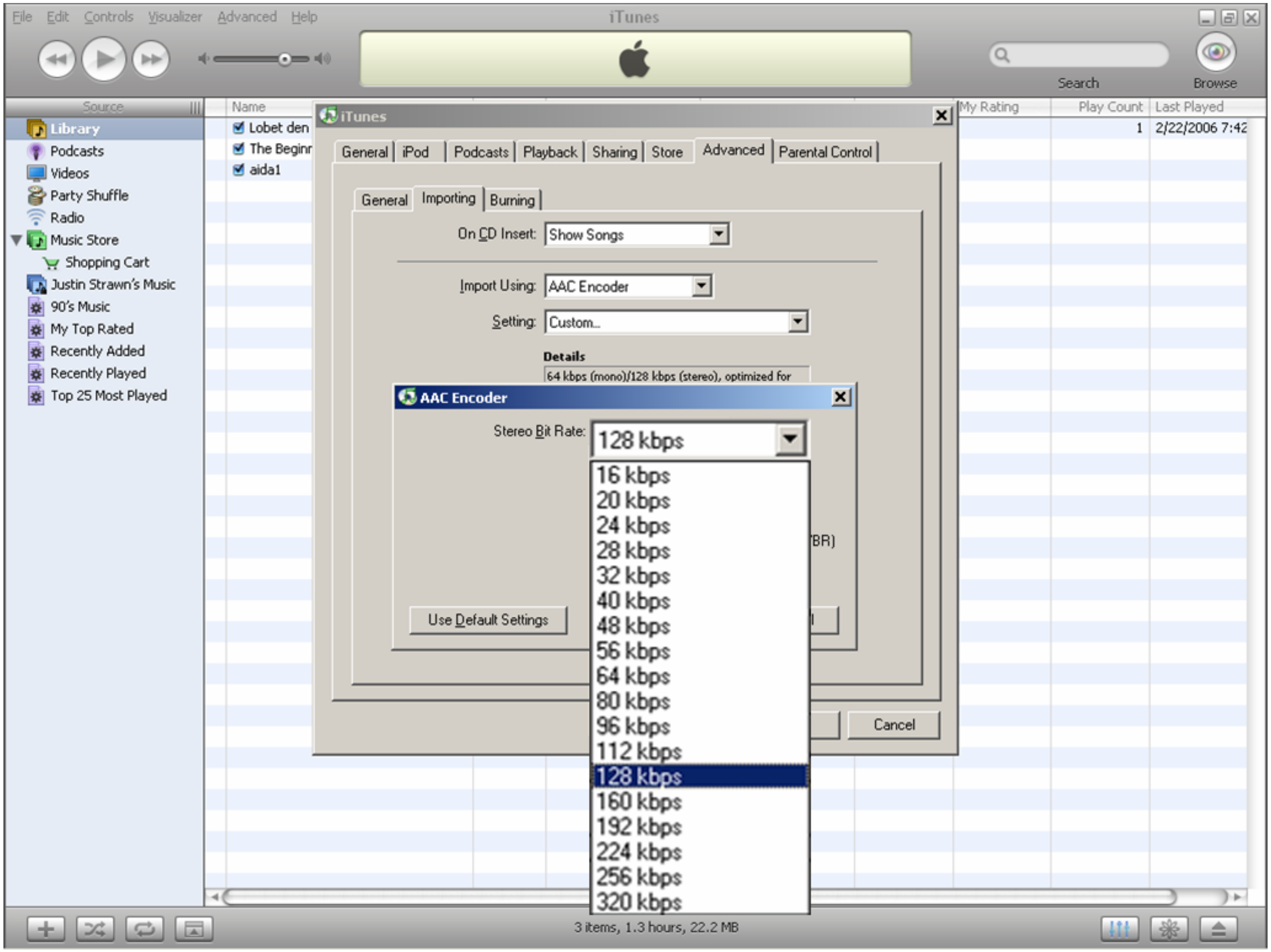
Encoder Controls: iTunes



Encoder Controls: iTunes AAC (1)



Encoder Controls: iTunes AAC (2)



Wrapup

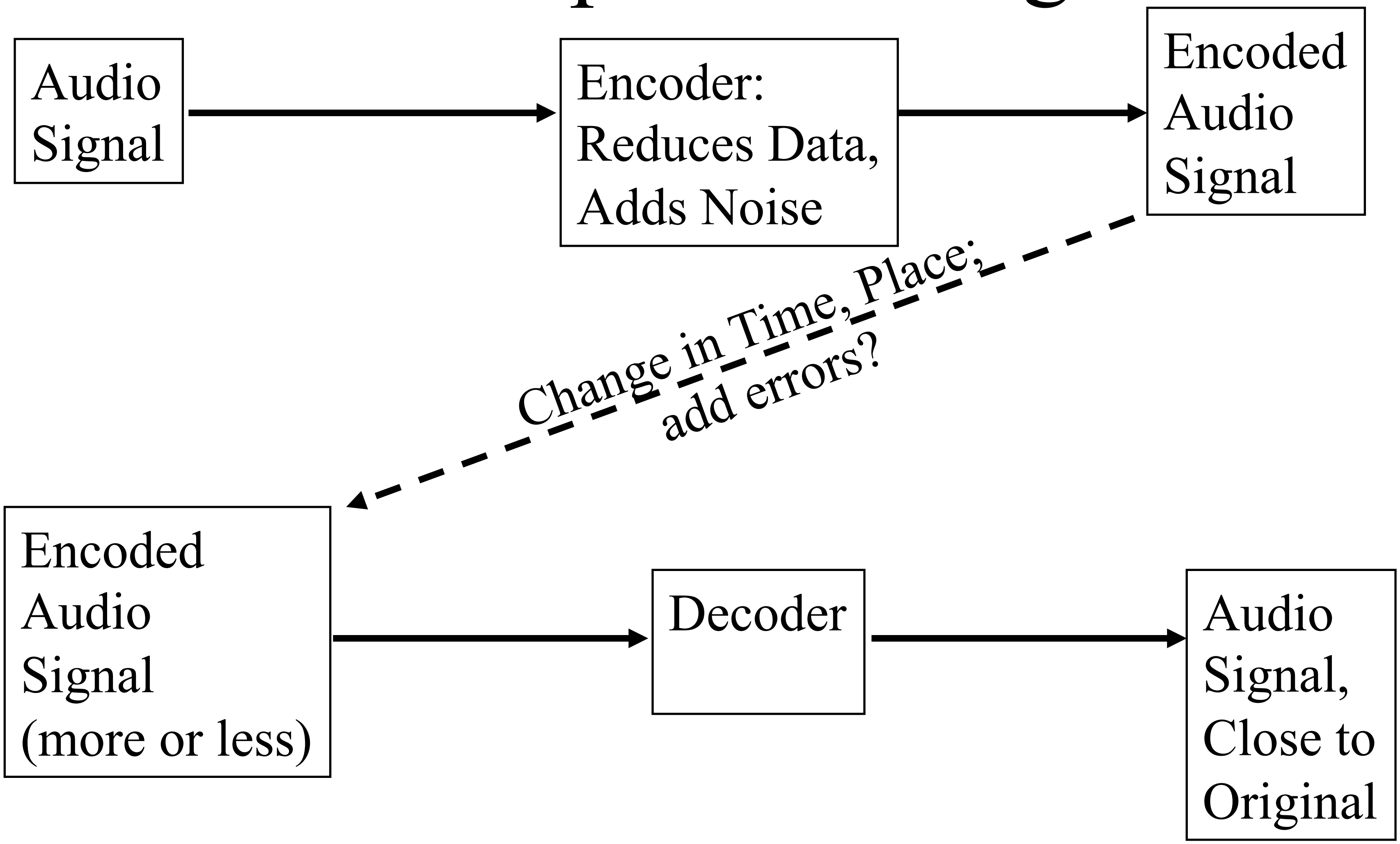
Where I see compression going

- Storage
- Bandwidth
- Domination
- Quality
- Lossless
- DSP
- Scene, not stand-alone
- Synthesis

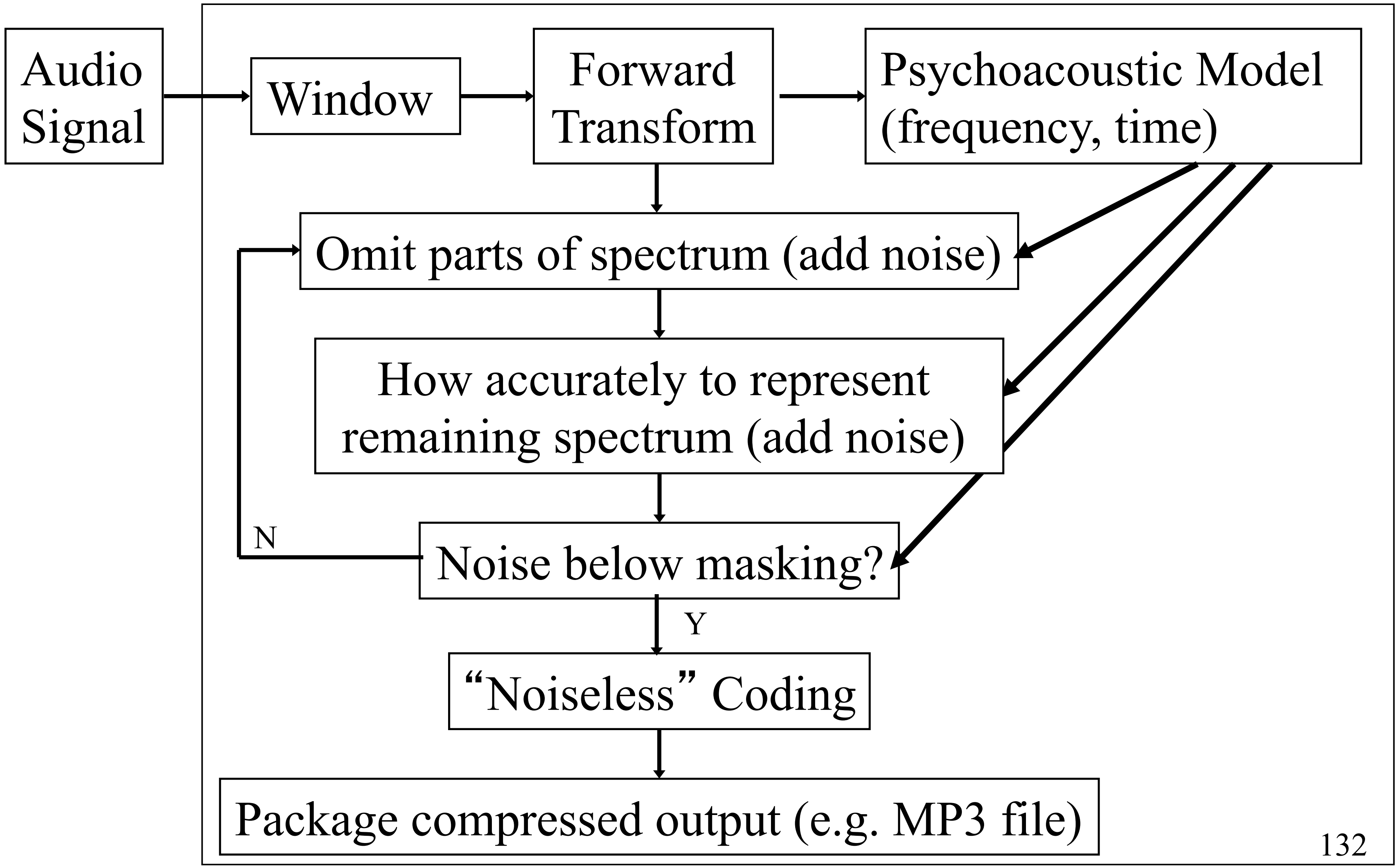
Review: Model of Communication

Based on Colin Cherry: *On Human Communication*

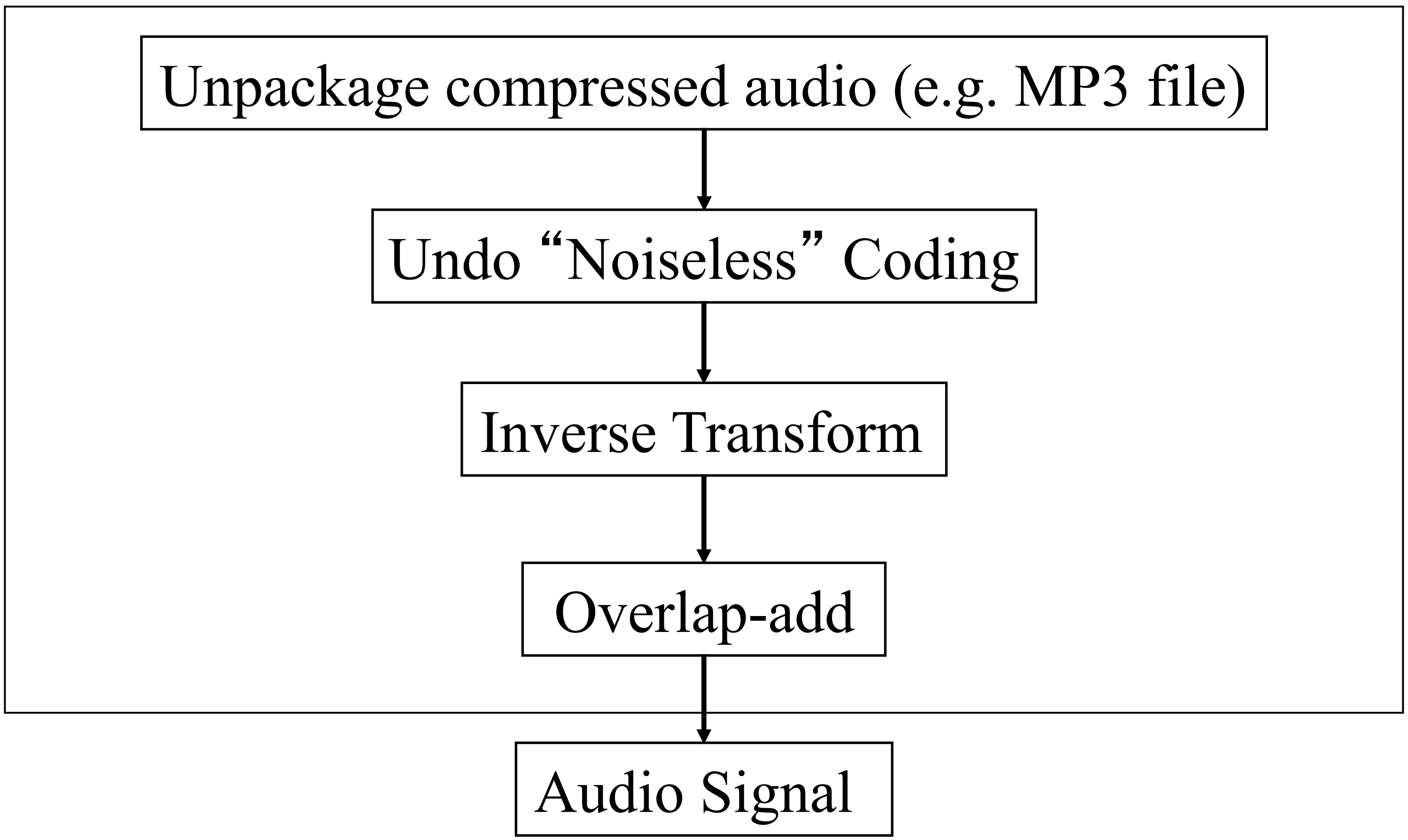
Perceptual Coding



Review: Perceptual Encoder



Review: Decoder (“mp3 player”)



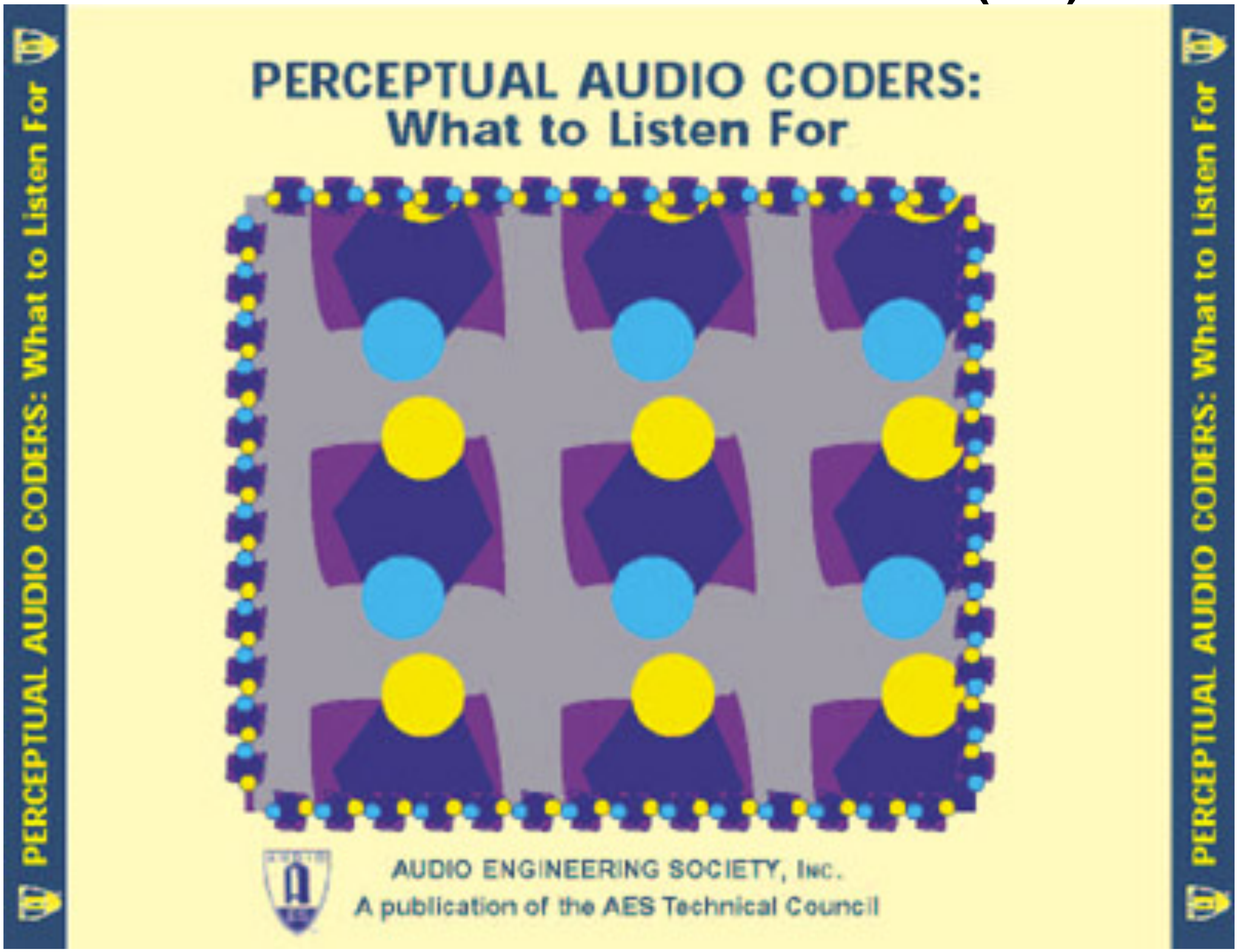
Bonus Question

- Encode/decode full scale
- "Mp3 player" output clips
- Why?
- How to fix?

Compression Resources

- Handouts:
  - List of important URLs
  - Bibliography
  - List of major standards

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/

John Strawn, Ph.D.

S Systems Inc.

jstrawn@s-systems-inc.com

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