Harvard-MIT Division of Health Sciences and Technology
HST.723: Neural Coding and Perception of Sound
Instructor: Bertrand Delgutte

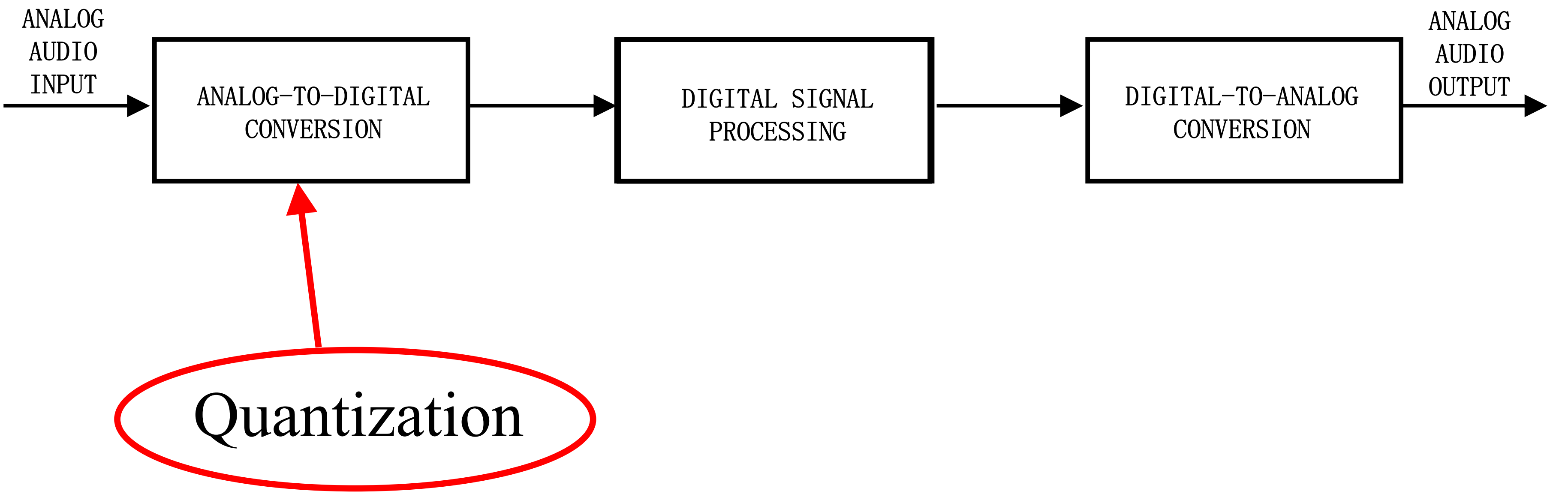
Fundamentals of Perceptual Audio Encoding

Craig Lewiston
HST.723 Lab II
2/25/05

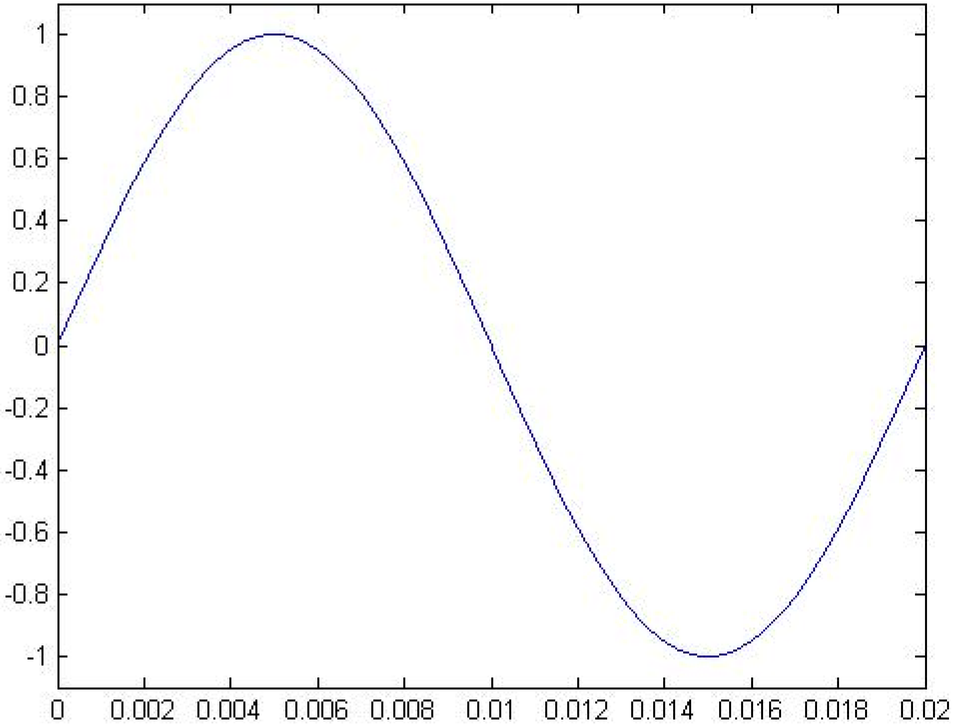
Goals of Lab

- Introduction to fundamental principles of digital audio & perceptual audio encoding
- Learn the basics of psychoacoustic models used in perceptual audio encoding.
- Run 2 experiments exploring some fundamental principles behind the psychoacoustic models of perceptual audio encoding.

Digital Audio

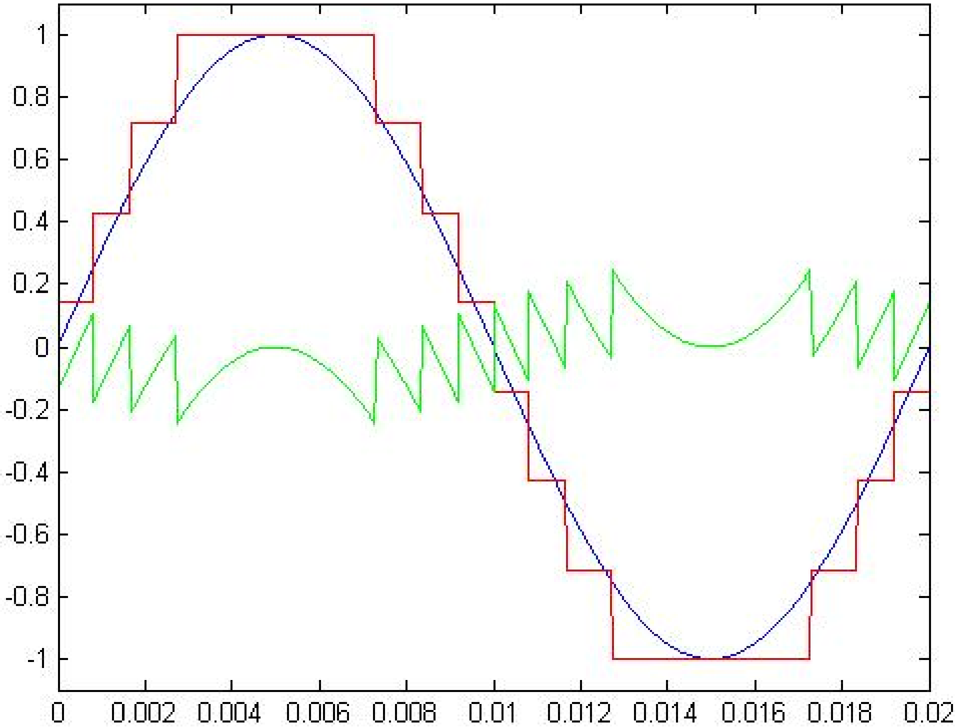


Quantization



N Bits => 2^N levels

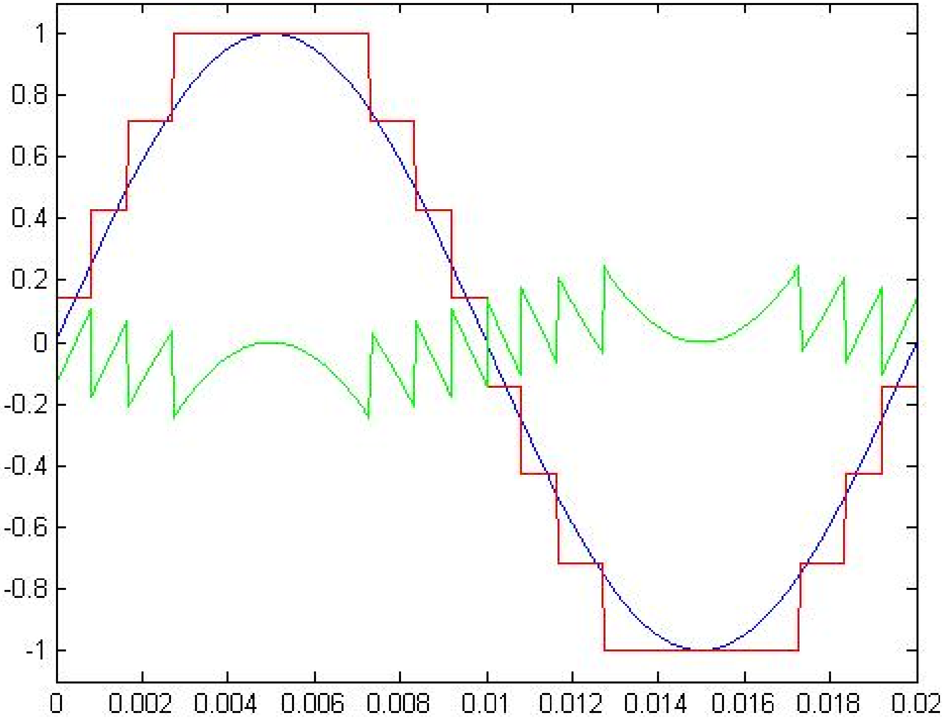
Quantization



Quantization Noise is the difference between the analog signal and the digital representation, and arises as a result of the error in the quantization of the analog signal.

N Bits => 2^N levels

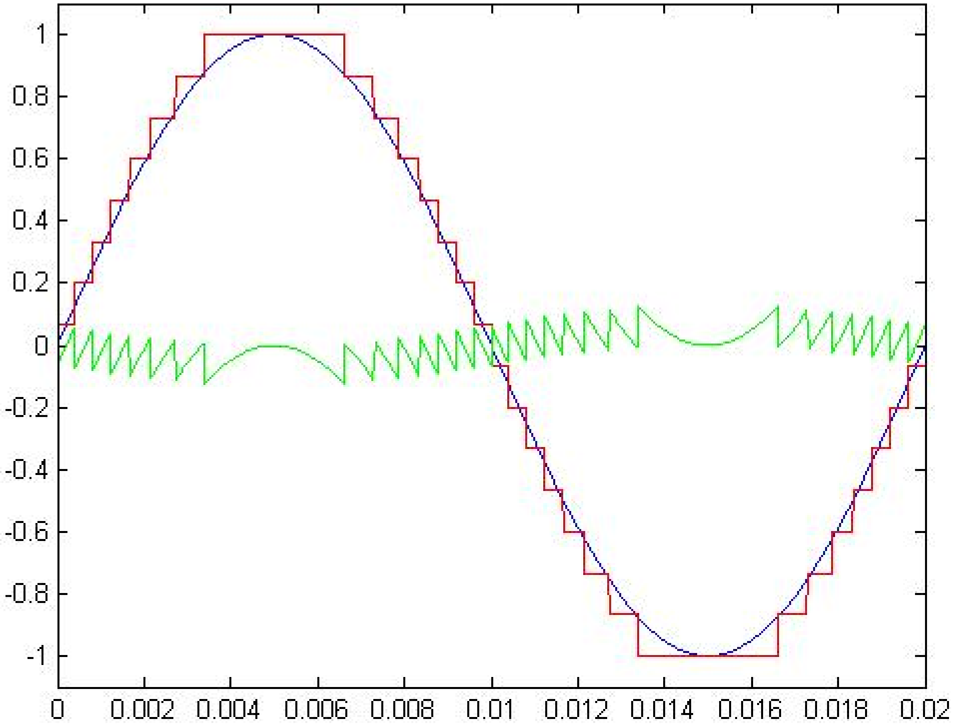
Quantization



Bits	Levels
3	8
4	16
5	32
8	256
16	65536

With each increase in the bit level, the digital representation of the analog signal increases in fidelity, and the quantization noise becomes smaller.

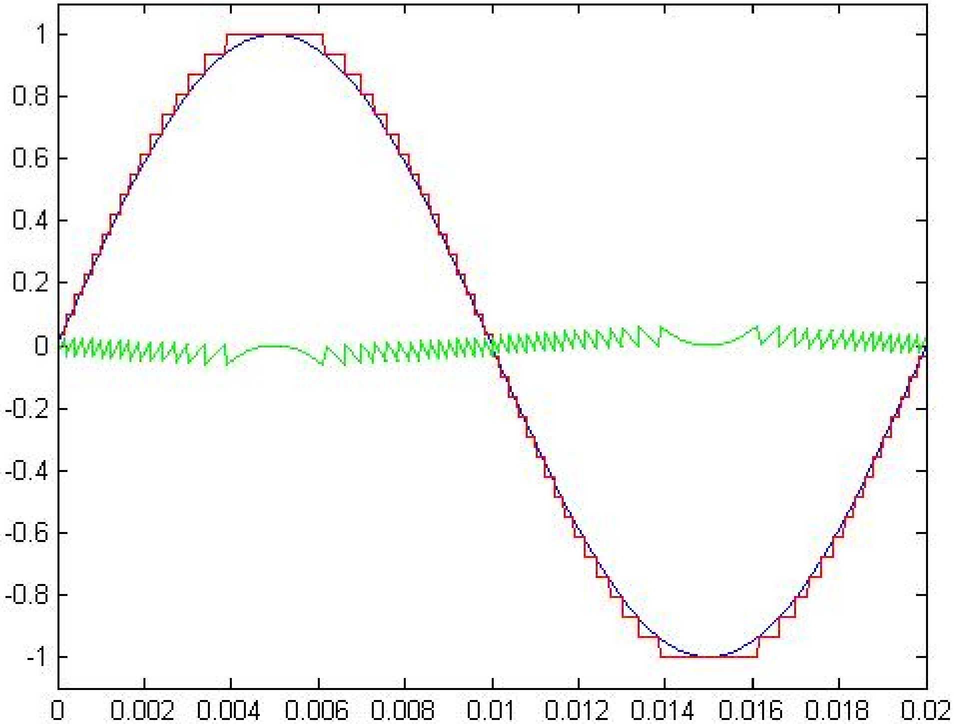
Quantization



Bits	Levels
3	8
4	16
5	32
8	256
16	65536

With each increase in the bit level, the digital representation of the analog signal increases in fidelity, and the quantization noise becomes smaller.

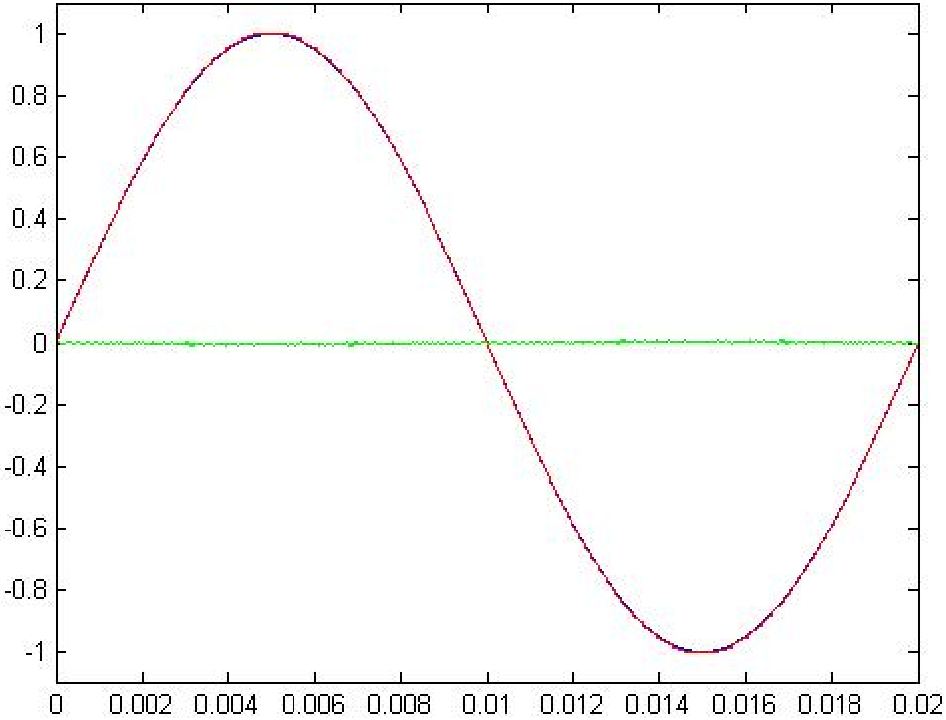
Quantization



Bits	Levels
3	8
4	16
5	32
8	256
16	65536

With each increase in the bit level, the digital representation of the analog signal increases in fidelity, and the quantization noise becomes smaller.

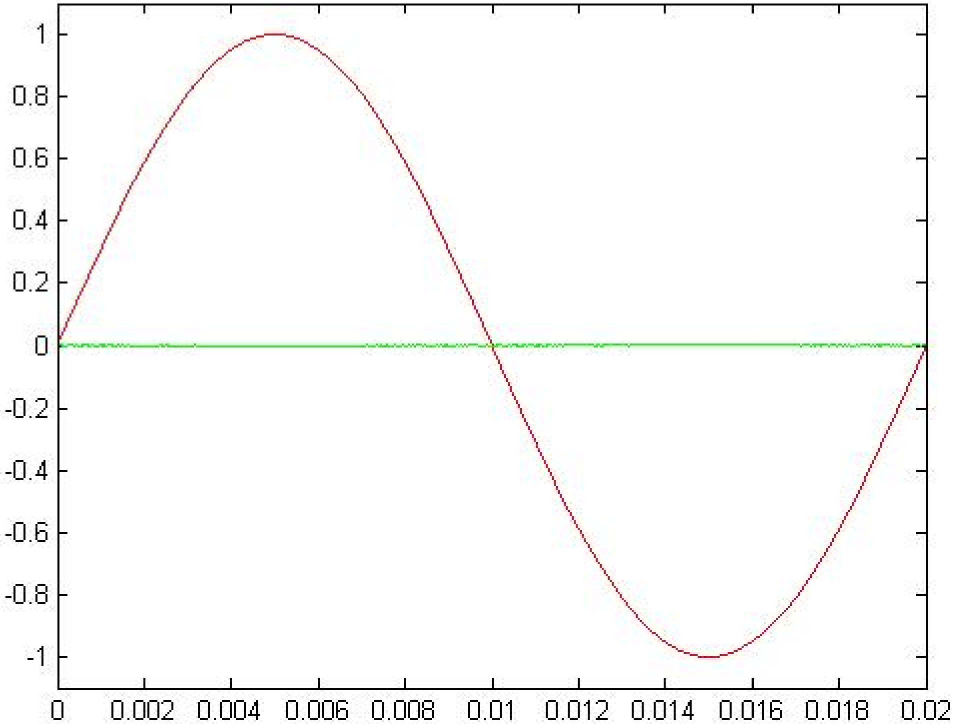
Quantization



Bits	Levels
3	8
4	16
5	32
8	256
16	65536

With each increase in the bit level, the digital representation of the analog signal increases in fidelity, and the quantization noise becomes smaller.

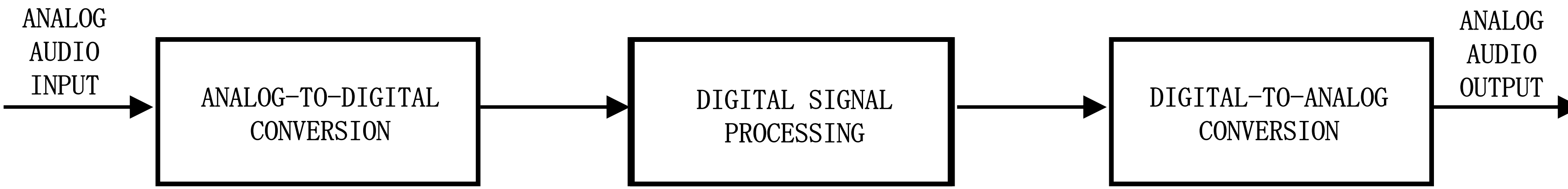
Quantization



Bits	Levels
3	8
4	16
5	32
8	256
16	65536

With each increase in the bit level, the digital representation of the analog signal increases in fidelity, and the quantization noise becomes smaller.

Digital Audio



CD Audio:

- 2 Channels (Stereo)
- 44.1 kHz sampling rate
- 16 bit encoding

2 * 44.1 kHz * 16 bits = 1.41 Mb/s

+

Overhead (synchronization, error correction, etc.)

CD Audio = 4.32 Mb/s

Compression

- High data rates, such as CD audio (4.32 Mb/s), are incompatible with internet & wireless applications.
- Audio data must somehow be compressed to a smaller size (less bits), while not affecting signal quality and reducing quantization noise.
- **Perceptual Audio Encoding** is the encoding of audio signals, incorporating psychoacoustic knowledge of the auditory system, in order to reduce the amount of bits necessary to faithfully reproduce the signal.
    - MPEG-1 Layer III (aka mp3)
    - MPEG-2 Advanced Audio Coding (AAC)

MPEG

MPEG = Motion Picture Experts Group

MPEG is a family of encoding standards for digital multimedia information

- MPEG-1: a standard for storage and retrieval of moving pictures and audio on storage media (e.g., CD-ROM).
    - Layer I
    - Layer II
    - Layer III (aka MP3)

- MPEG-2: standard for digital television, including high-definition television (HDTV), and for addressing multimedia applications.
    - Advanced Audio Coding (AAC)

- MPEG-4: a standard for multimedia applications, with very low bit-rate audio-visual compression for those channels with very limited bandwidths (e.g., wireless channels).

- MPEG-7: a content representation standard for information search

Overview of Perceptual Encoding

Figure removed due to copyright reasons.
Please see:
Painter, T., and A. Spanias. "Perceptual coding of digital audio." *Proceedings of the IEEE* 88 (2000): 451-513.

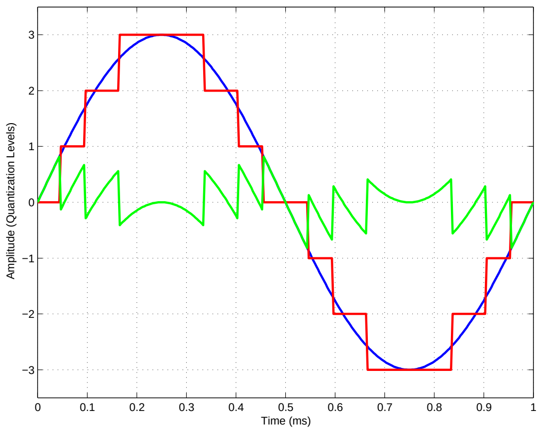
General Perceptual Audio Encoder (Painter & Spanias, 2000):

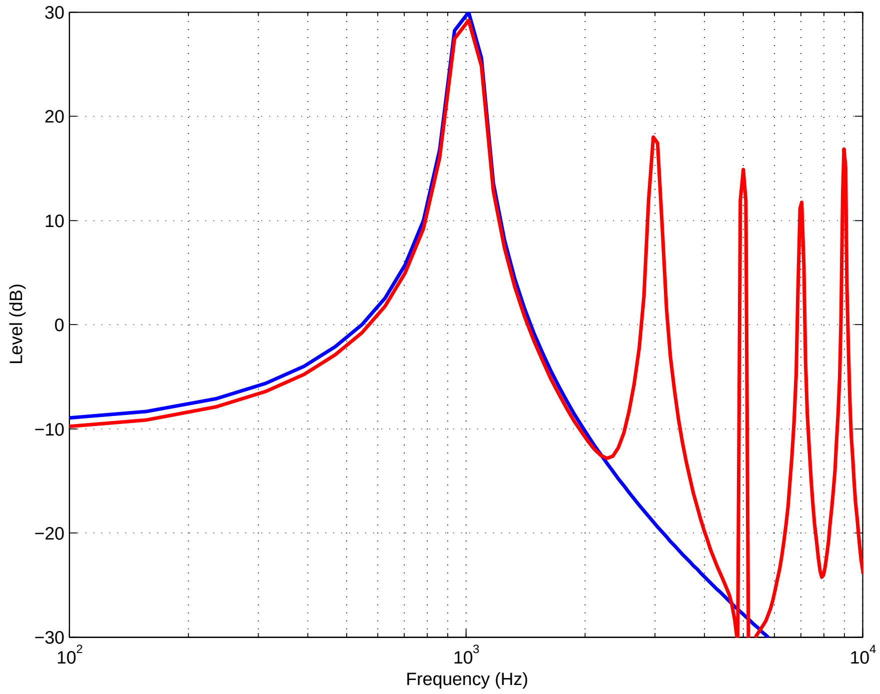
- Psychoacoustic analysis => masking thresholds
- Masking thresholds ≡ Excitation Pattern
- Basic principle of Perceptual Audio Encoder: use **masking pattern** of stimulus to determine the **least number of bits** necessary for each **frequency sub-band**, so as to prevent the **quantization noise** from becoming audible.

Masking

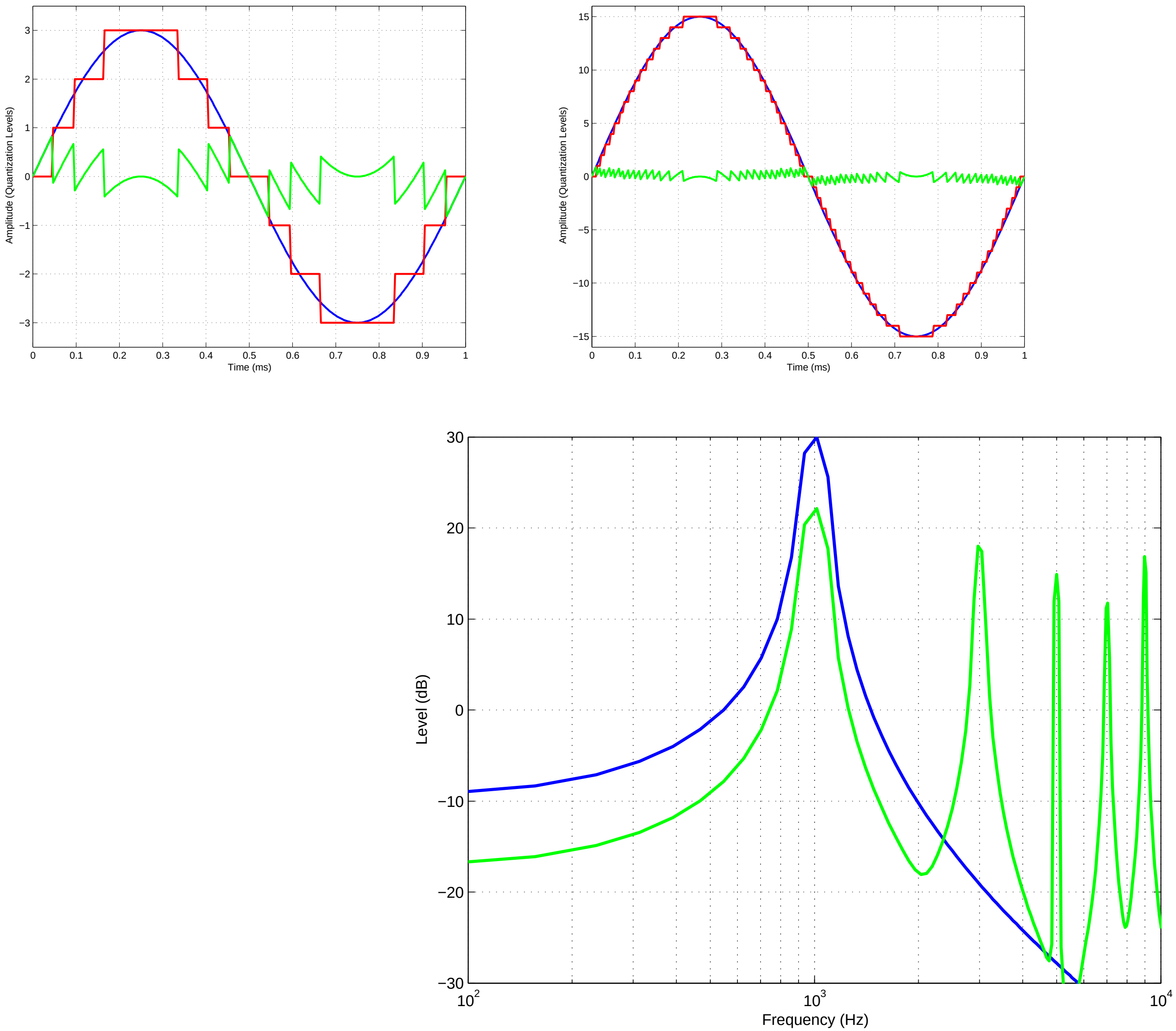
Figure removed due to copyright reasons.
Please see:
Painter, T., and A. Spanias. "Perceptual coding of digital audio." *Proceedings of the IEEE* 88 (2000): 451-513.

Quantization Noise

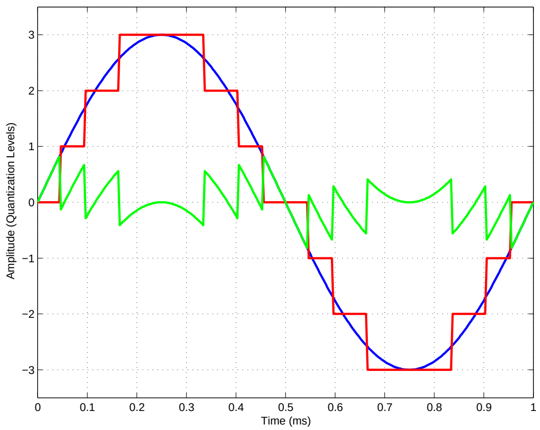


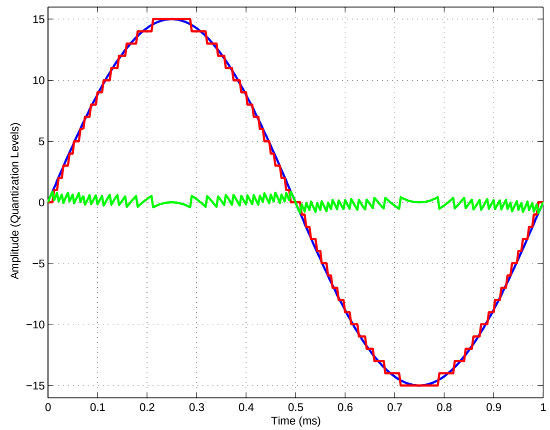


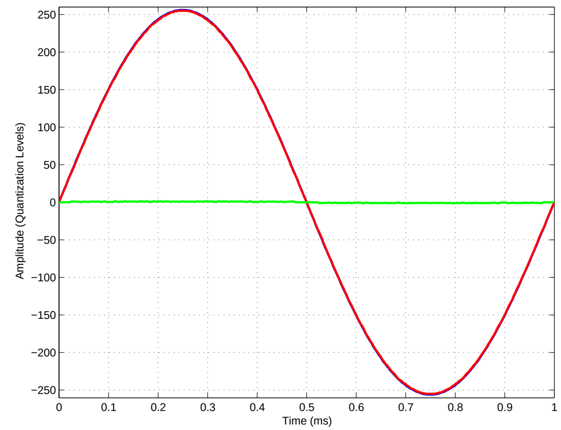
Quantization Noise

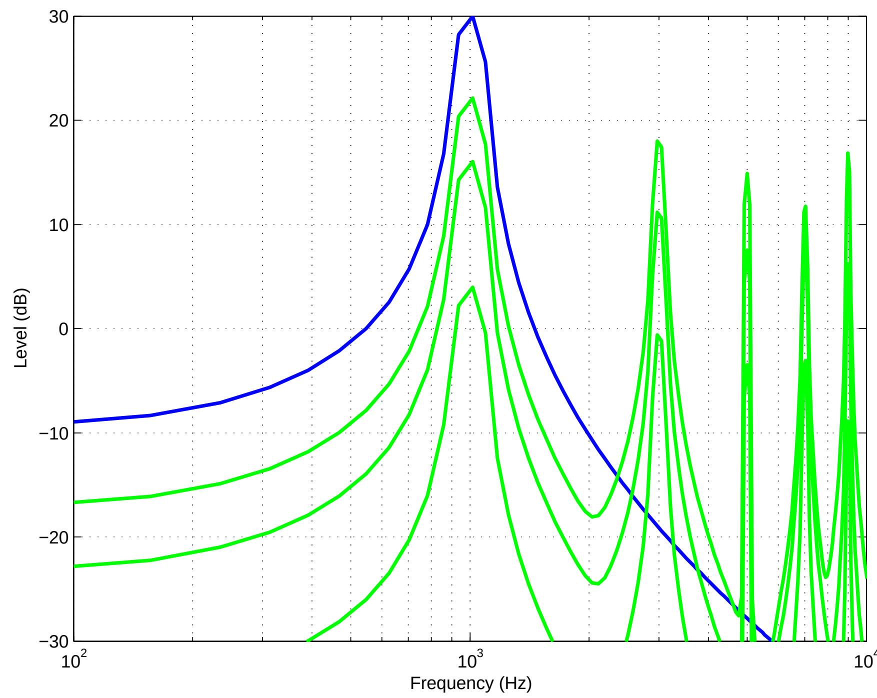


Quantization Noise

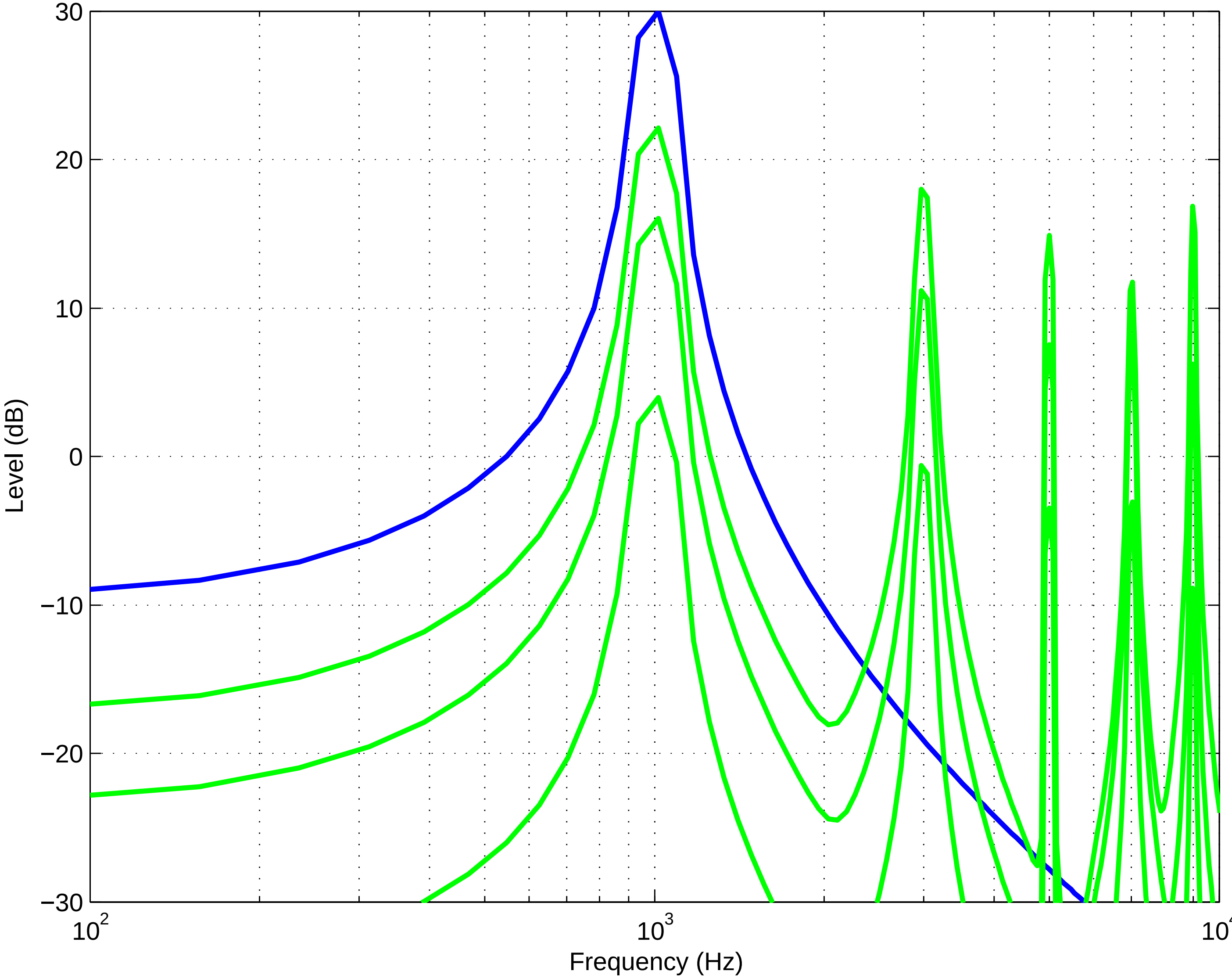




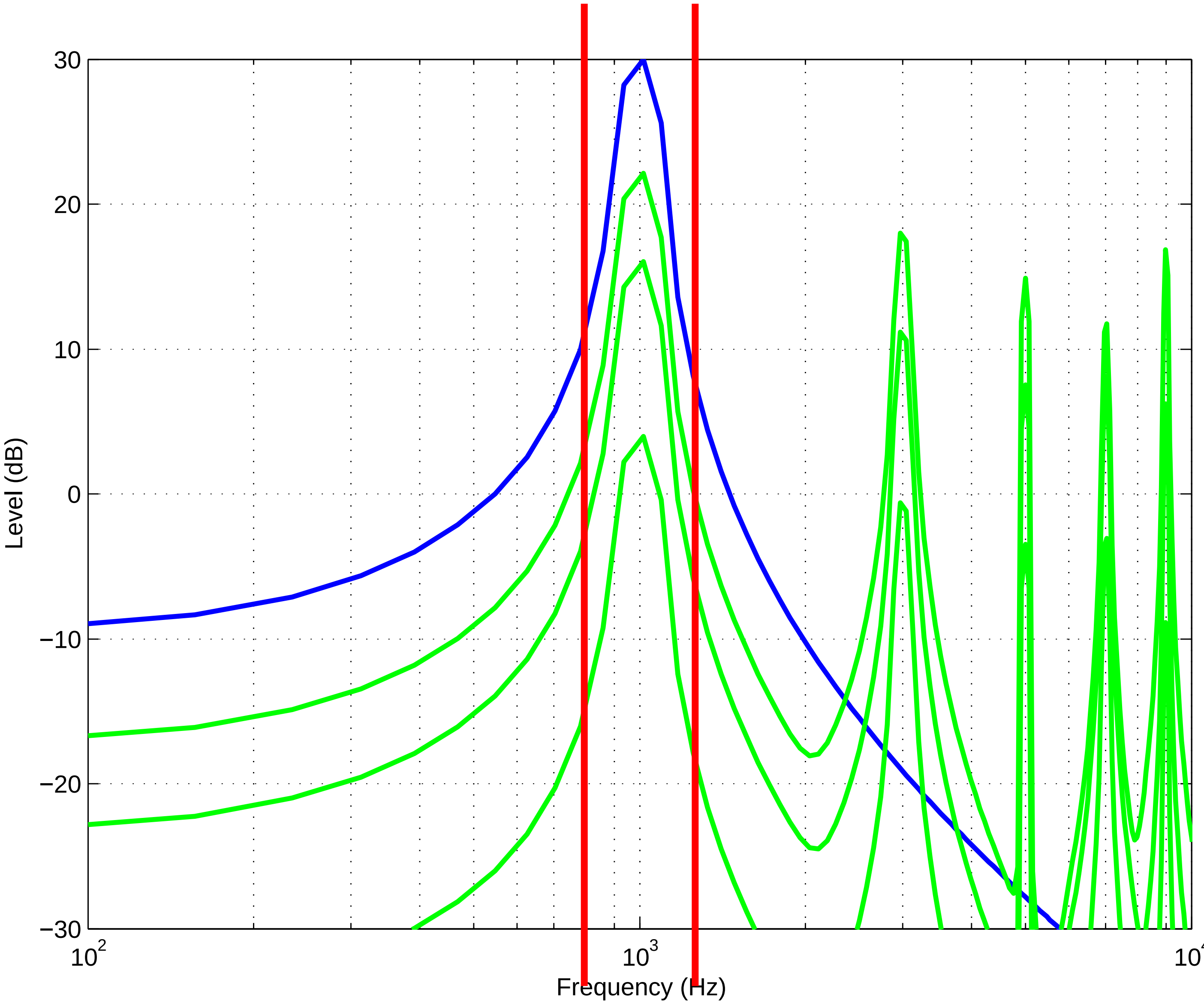




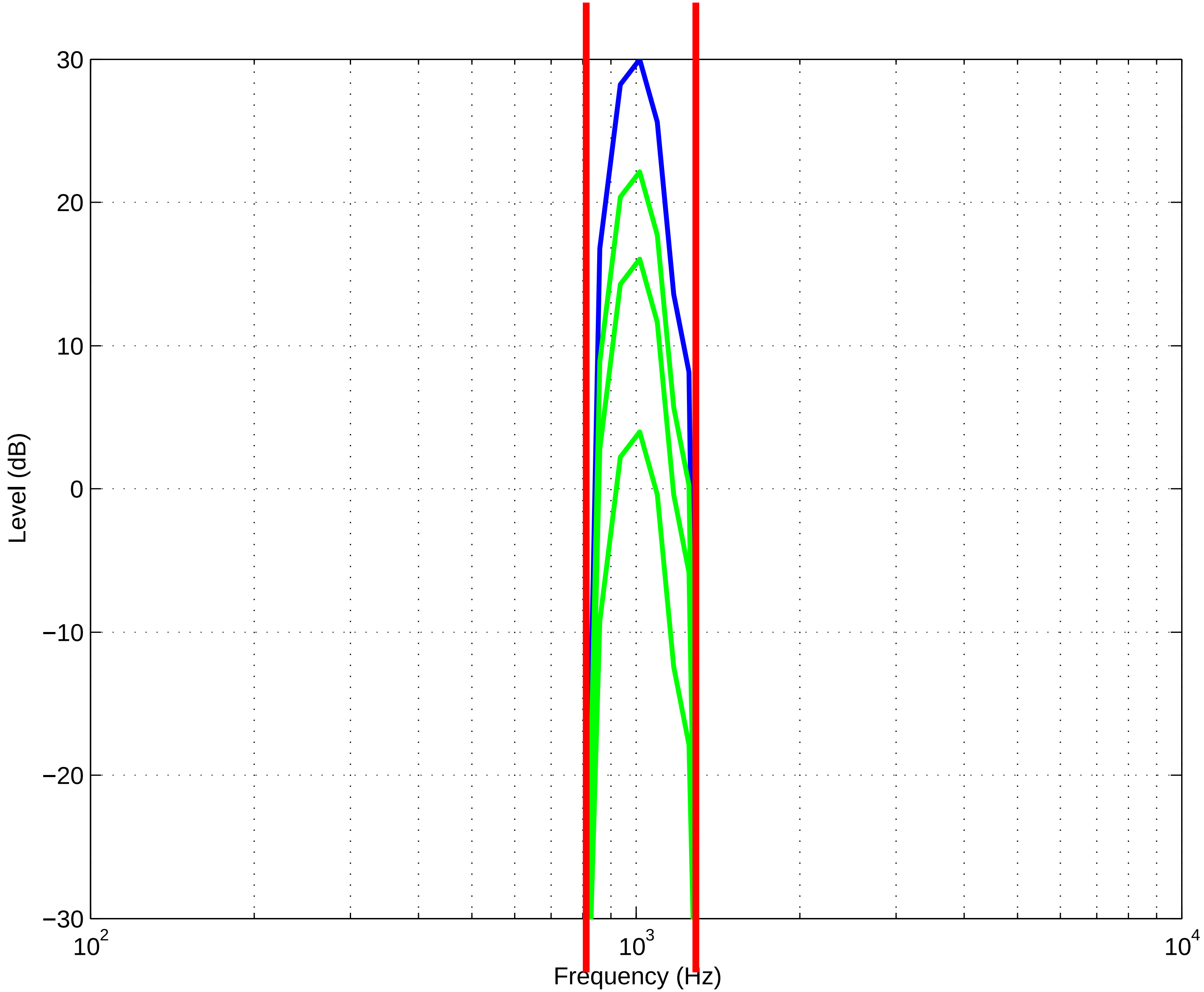
Sub-band Coding



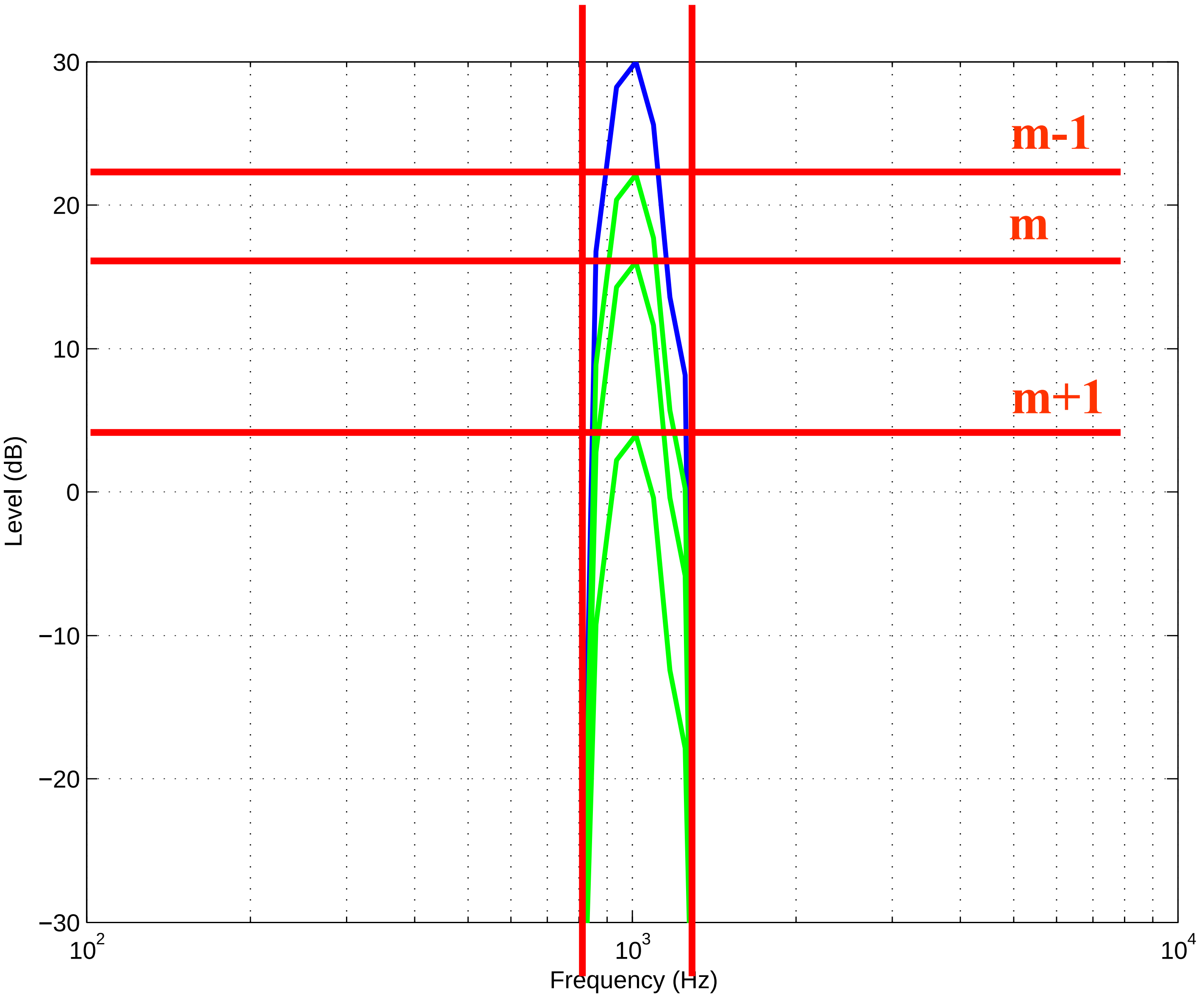
Sub-band Coding



Sub-band Coding



Sub-band Coding



Masking/Bit Allocation

Figure removed due to copyright reasons.
Please see:
Painter, T., and A. Spanias. "Perceptual coding of digital audio." *Proceedings of the IEEE* 88 (2000): 451-513.

The number of bits used to encode each frequency sub-band is equal to the least number of bits with a quantization noise that is below the minimum masking threshold for that sub-band.

Example: MPEG-1 Psychoacoustic Model I

Figure removed due to copyright reasons.
Please see:
Painter, T., and A. Spanias. "Perceptual coding of digital audio." *Proceedings of the IEEE* 88 (2000): 451-513.

1. Spectral Analysis and SPL Normalization

Example: MPEG-1 Psychoacoustic Model I

Figure removed due to copyright reasons.
Please see:
Painter, T., and A. Spanias. "Perceptual coding of digital audio." *Proceedings of the IEEE* 88 (2000): 451-513.

2. Identification of Tonal Maskers & calculation of individual masking thresholds

Example: MPEG-1 Psychoacoustic Model I

Figure removed due to copyright reasons.
Please see:
Painter, T., and A. Spanias. "Perceptual coding of digital audio." *Proceedings of the IEEE* 88 (2000): 451-513.

2. Identification of Noise Maskers &
calculation of individual masking thresholds

Example: MPEG-1 Psychoacoustic Model I

4. Calculation of Global Masking Thresholds

Example: MPEG-1 Psychoacoustic Model I

Figure removed due to copyright reasons.
Please see:
Painter, T., and A. Spanias. "Perceptual coding of digital audio." *Proceedings of the IEEE* 88 (2000): 451-513.

A - Some portions of the input spectrum require SNR's > 20 dB

B - Other portions require less than 3 dB SNR

C - Some high frequency portions are masked by the signal itself

D - Very high frequency portions fall below the absolute threshold of hearing.

Example: MPEG-1 Psychoacoustic Model I

Figure removed due to copyright reasons.
Please see:
Painter, T., and A. Spanias. "Perceptual coding of digital audio." *Proceedings of the IEEE* 88 (2000): 451-513.

5. Sub-band Bit Allocation

Lab Experiments

Exp 1: Masking Pattern

- Measure absolute hearing thresholds in quiet
- Measure absolute hearing thresholds in presence of narrowband noise masker

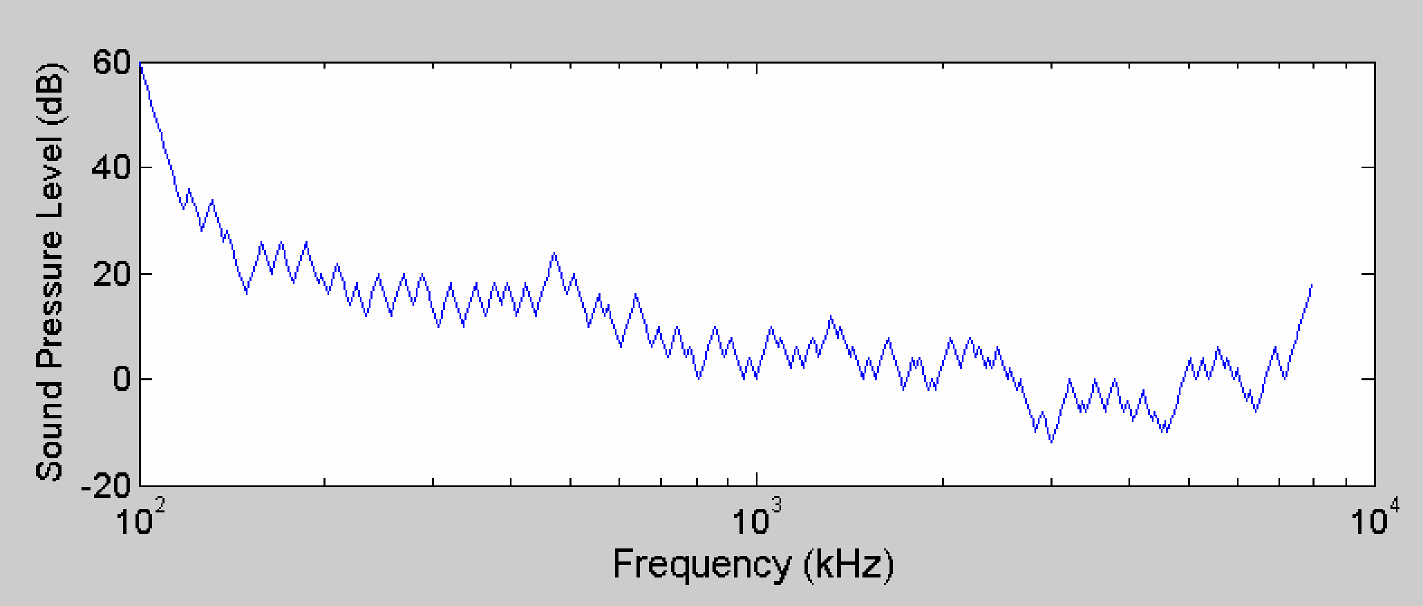
Exp 2: Masking Threshold

- Measure masking threshold of a 1 kHz tone in the presence of four different maskers:
  - Tone
  - Gaussian Noise
  - Multiplied Noise
  - Low-noise Noise

Exp 1: Masking Pattern

Figure removed due to copyright considerations. Please see: Noll, Peter. "MPEG Digital Audio Coding Standards." *The Digital Signal Processing Handbook.* Edited by V. K. Madisetti and D. B. Williams. IEEE Press/CRC Press, 1998, pp. 40-1 - 40-28.

Method of Adjustment



Method of Adjustment (aka Békésy tracking method)

Target tone is swept through frequency range, and subject must adjust intensity of target tone so that it is **just barely detectable**

Exp 2: Masking Thresholds

Figure removed due to copyright reasons.
Please see:
Painter, T., and A. Spanias. "Perceptual coding of digital audio." *Proceedings of the IEEE* 88 (2000): 451-513.

Calculation of tonal & noise masking thresholds:
Tonal & noise maskers have different masking effects…

Asymmetry of Simultaneous Masking

Figure removed due to copyright reasons.
Please see:
Painter, T., and A. Spanias. "Perceptual coding of digital audio." *Proceedings of the IEEE* 88 (2000): 451-513.

Tone masker
SNR ~ 24 dB

Noise masker
SNR ~ 4 dB

Asymmetry of Simultaneous Masking

Why do tones and noises have different masking effects?

$$\text{Signal} = A(t)\, e^{j\omega(t) + \varphi(t)}$$

For narrowband Gaussian noise, $e^{j\omega(t)}$ is approximately the same as a tone centered at the same frequency.

Asymmetry effect is either due to the amplitude term $A(t)$ or to the phase term $\varphi(t)$.

Asymmetry of Simultaneous Masking

Measure masking effects of "modified" noises:

Multiplied Noise: generated by multiplying a sinusoid at 1 kHz with a low-pass Gaussian noise.

Amplitude => Gaussian Noise

Phase => Pure Tone

Low-Noise Noise: Gaussian noise with a temporal envelope that has been smoothed.

Amplitude => Pure Tone

Phase => Gaussian Noise

Exp 2: Masking Thresholds

Figure removed due to copyright reasons.
Please see:
Painter, T., and A. Spanias. "Perceptual coding of digital audio." *Proceedings of the IEEE* 88 (2000): 451-513.

Target (Quantization noise)	**Masker** (Desire signal)
Gaussian noise	Tone
Gaussian noise	Gaussian noise
Gaussian noise	Multiplied noise
Gaussian noise	Low-noise noise

1. Measure masking threshold for four different types of masker
2. Comparing the modified noise thresholds with the tone & Gaussian noise thresholds should indicate which component of the Gaussian noise (Amplitude or Phase) contributes to the asymmetry effect.

Method: Adaptive Procedure

Trial Number

Intensity	1	2	3	4	5	6	7	8	9	10	11	12
75												
74												
73												
72												
71	Y											
70		Y				Y						Y
69			Y		N		Y				N	
68				N				Y		N		
67									N			
66												
65												

Threshold = average of reversal points (usually 6 or 7)

Lab Write-up

1) Describe the fundamental concepts behind digital audio & perceptual audio encoders (e.g. quantization & quantization noise, sub-band coding & bit allocation, tone & noise masking thresholds, etc.)

2) Describe the methods of Experiment 1 and the results you obtained. Explain how the threshold results obtained relate to the masking thresholds used in perceptual audio encoding.

3) Describe the methods of Experiment 2 and the results you obtained, highlighting the amplitude and phase characteristics of the two "modified" noises used. Based on your data, indicate which component (amplitude or phase) contributes to the asymmetry of simultaneous masking observed.

References

- Dau, T., Verhey, J., and Kohlrausch, A. (**1999**). "Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers," J. Acoust. Soc. Am. **106**, 2752-2760.
- Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., and Püschel, D. (**1997**). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," Acta Acustica **83**, 659-669.
- Peter Noll, MPEG Digital Audio Coding Standards, Chapter in: IEEE Press/CRC Press "The Digital Signal Processing Handbook" (ed.: V.K. Madisetti and D. B. Williams), pp. 40-1 - 40-28, 1998
- Painter, T., and Spanias, A. (**2000**). "Perceptual coding of digital audio," Proceedings of the IEEE **88**, 451-513.
- Pan, D. (**1995**). "A tutorial on MPEG/audio compression," IEEE Multimedia Journal