Audio Signal Processing I

Shyh-Kang Jeng
Department of Electrical Engineering/
Graduate Institute of Communication Engineering

*Reference: Marina Bosi, Perceptual Audio Coding-Lecture Note of Music 422/EE367C, CCRMA, Stanford University, Spring 1999

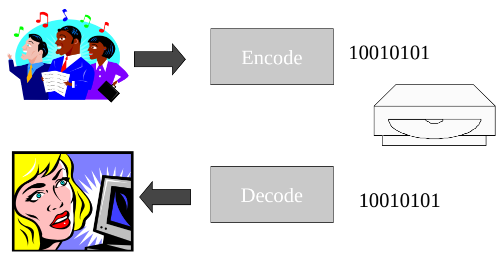
Reference

- M. Bosi and R. E. Goldberg, Introduction to Digital Audio Coding and Standards, Kluwer Academic Publishers, 2003.

Contents

- Introduction
- Quantization
- Time to Frequency Mapping
- Psychoacoustics
- Bit Allocation
- Perceptual Audio Coders
- MPEG-1 Audio

Audio Coder

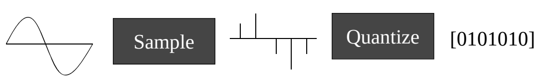


Coding Goals

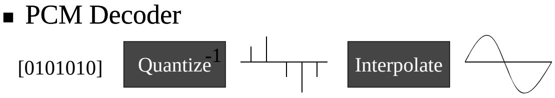
- Maximize the perceived quality of the sound
- Minimize the data rates and complexity
- Related parameters
  - Delay
  - Error robustness
  - Scalability
  - etc.

Pulse-Code Modulation

- PCM Encoder



- PCM Decoder



PCM Example: CD Format

- Sampling frequency: Fs = 44.1 KHz (i.e. one sample every ~0.023 ms)
- Number of bits per sample: R = 16 (i.e. up to 2^{16} = 65536 levels)
- Bit rate: I = Fs*R = 706.5 kb/s per channel
- Total bit rate: I_{stereo} = 1.413 Mb/s
- Signal to noise ratio: SNR ~ 90 dB

7

Fourier Transform

- Fourier transform

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt$$

- Inverse Fourier transform

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft}df$$

8

Energy and Power

- Power

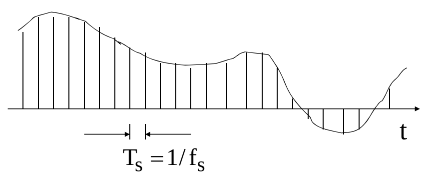
$$p(t) = x(t)^2$$

- Energy

$$E = \int_0^T p(t)dt = T\langle p(t)\rangle$$

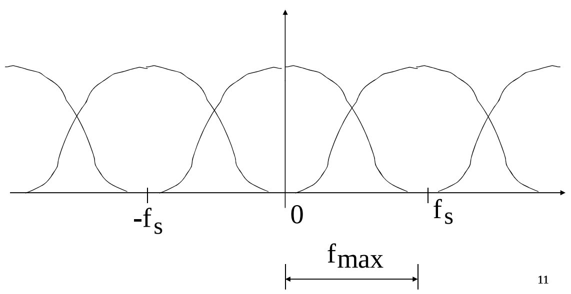
9

Signal Sampling



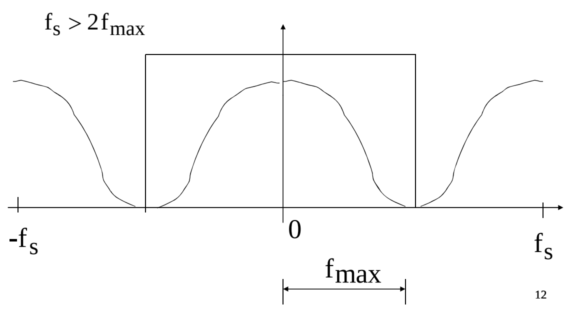
10

Aliasing



11

Sampling Theorem



12

Eliminating Aliasing

- If your application requires a sample rate below the highest frequencies in the signal, you will need to low pass filter the signal before sampling
- Example: The telephone sample rate is 8 KHz and a 4 KHz low pass filter is employed. (speech: ~100 Hz to ~7 KHz, you really do sound different on the phone)

13

Coder Implications

- We can only hear up to ~20 KHz so we should filter out higher frequencies and sample at ~40 KHz to get high quality reproductions of broadband sound
- For example, CDs sample at 44.1 KHz and provide much greater sound quality than telephone system

14

Binary Numbers

- Decimal notation
  - Symbols: 0, 1, 2, 3, 4, …, 9
  - e.g., $1999 = 1*10^3 + 9*10^2 + 9*10^1 + 9*10^0$
- Binary notation
  - Symbols: 0, 1
  - e.g.,

$$[01100100] = 0*2^7 + 1*2^6 + 1*2^5 + 0*2^4 + 0*2^3 + 1*2^2 + 0*2^1 + 0*2^0 = 100$$

15

Negative Numbers

- Folded binary
  - Use the highest order bit as an indicator of sign
- Two's complement
  - Follows the highest positive number with the lowest negative
  - e.g., 3 bits, $3 \equiv [011], -4 \equiv [100] = 2^4 - 4$
- We use folded binary notation when we need to represent negative numbers

16

Two Quantization Methods

- Uniform quantization
  - Constant limit on absolute round-off error $\Delta / 2$
  - Poor performance on SNR at low input power
- Floating point quantization
  - Some bits for an exponent
  - the rest for an mantissa
  - SNR is determined by the number of mantissa bits and remain roughly constant
  - Gives up accuracy for high signals but gains much greater accuracy for low signals

17

Quantization Error

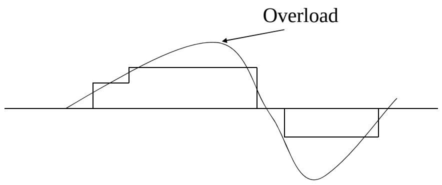
- Main source of coder error
- Characterized by $\langle q^2 \rangle$
- A better measure

  $\mathrm{SNR} = 10\log_{10}(\langle x^2 \rangle / \langle q^2 \rangle)$
- Does not reflect auditory perception
- Can not describe how perceivable the errors are
- Satisfactory objective error measure that reflects auditory perception does not exist

18

Quantization Error (cont.)

- Round-off error
- Overload error



19

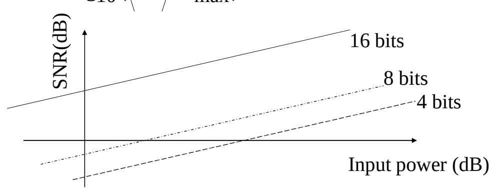
Round-Off Error

- Comes from mapping ranges of input amplitudes onto single codes
- Worse when the range of input amplitude onto a code is wider
- Assume that the error follows a uniform distribution
- Average error power $\langle q^2 \rangle = \int_{-\Delta/2}^{\Delta/2} q^2 \frac{1}{\Delta} dq = \Delta^2/12$
- For a uniform quantizer $\langle q^2 \rangle = x_{\max}^2 / (3 * 2^{2R})$

20

Round-Off Error (cont.)

$$\mathrm{SNR} = 10\log_{10}(\langle x^2 \rangle / \langle q^2 \rangle)$$
$$= 10\log_{10}(\langle x^2 \rangle / x_{\max}^2) + 20\,\mathrm{R}\log_{10} 2 + 10\log_{10} 3$$
$$= 10\log_{10}(\langle x^2 \rangle / x_{\max}^2) + 6.021\mathrm{R} + 4.771$$



21

Overload Error

- Comes from signals where $x(t) > x_{\max}$
- Depends on the probability distribution of signal values
- Reduced for high $x_{\max}$
- High $x_{\max}$ implies wide levels and therefore high round-off error
- Requires a balance between the need to reduce both errors

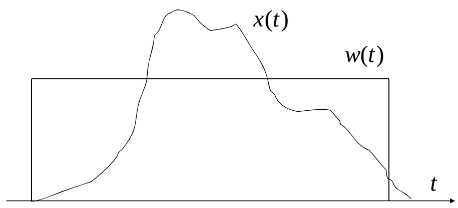
22

Frequency Domain Coding

- Subdivide the input signal into a number of frequency components and quantize these components separately
- Subdivision into frequency components removes redundancy in the input signal
- Number of bits to encode each frequency component can be variable, so that encoding accuracy can be placed in frequencies where is most needed

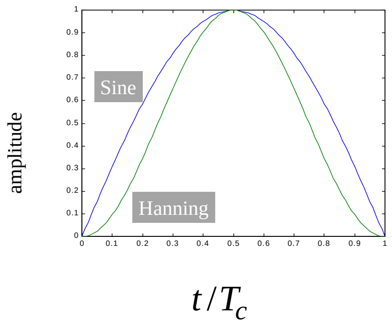
23

Window Function

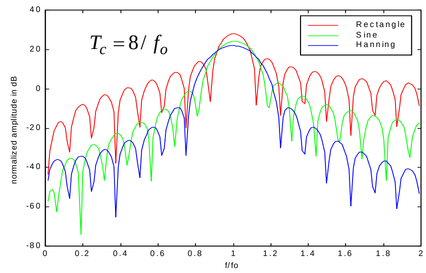


24

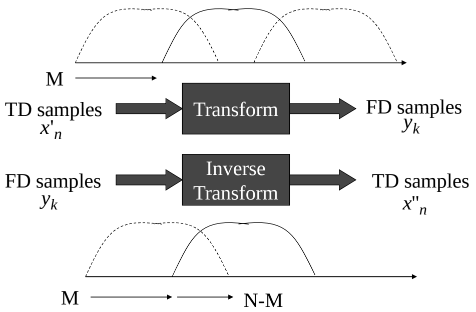
Windows



Fourier Transform of a Sine Wave with Various Windows



Overlap-Add Scheme



Reconstruction

- Window input signal with analysis window

$$x_n \rightarrow x'_n = x_n \cdot w_n^a$$

- Apply transform to the windowed signal

$$x'_n \rightarrow y_k = F(x'_n)$$

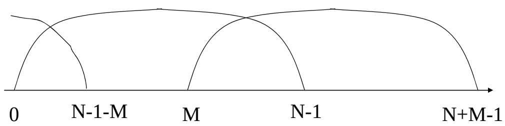
- Apply the inverse transform

$$y_k \rightarrow x''_n = F^{-1}(y_k) = x'_n$$

- Window with the synthesis window

$$x'_n \rightarrow z_n = x'_n \cdot w_n^s = x_n \cdot w_n^a \cdot w_n^s$$

Window Constraints



$$w_n^a \cdot w_n^s + w_{n+M}^a \cdot w_{n+M}^s = 1$$

$$w_n^a \cdot w_n^s = 1 \qquad\qquad w_n^a \cdot w_n^s = 1$$

Perfect Reconstruction

- Assume that the analysis window is the same as the synthesis window
- Assume that the window is symmetrical
- Assume no quantization

$$w(n)^2 + w(n+M)^2 = 1 \quad n = 0, \cdots, N-M-1$$

- A possible window

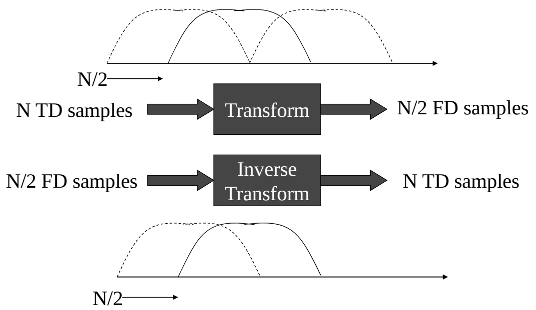
$$w(n) = \begin{cases} \sin(\pi \cdot ((n+1/2)/(N-M))/2) & n = 0, \cdots, N-M-1 \\ 1 & n = N-M, \cdots, M-1 \\ \sin(\pi \cdot ((N-n-1/2)/(N-M))/2) & n = M, \cdots, N-1 \end{cases}$$

Overlapping and Required System Rate

- Overlap N-M samples
  - Slide the window by M samples
  - Perform an N-point transform to obtain N frequency samples
  - Transmit N frequency samples every M time samples
  - If there is no overlap, we need only to transmit N frequency samples every N time samples
  - Thus the required system rate is higher than that of the no-overlapping case, because M<N

31

Perfect Reconstruction TDAC Transform



32

Oddly Stacked TDAC (OTDAC)

- Modified discrete cosine transform (MDCT)

$$X(k) = \sum_{n=0}^{N-1} x'(n)\cos\left(\frac{2\pi}{N}(n+n_o)(k+\frac{1}{2})\right) \quad k = 0,\cdots,\frac{N}{2}-1$$

$$n_o = (\frac{N}{2}+1)/2 \qquad X(k) = -X(N-k-1)$$

- Inverse modified discrete cosine transform (IMDCT)

$$x''(n) = \frac{2}{N}\sum_{k=0}^{N-1} X(k)\cos\left(\frac{2\pi}{N}(n+n_o)(k+\frac{1}{2})\right) \quad n = 0,\cdots,N-1$$

33

Perfect Reconstruction TDAC Transform

- Symmetric analysis and synthesis windows
- Identical analysis and synthesis windows

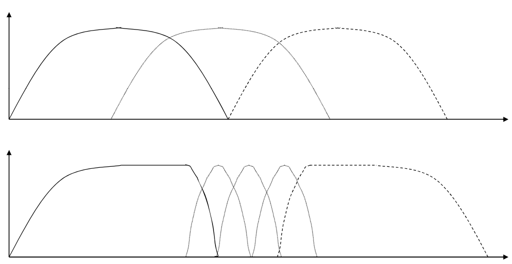
$$[w(n)]^2 + [w(N/2+n)]^2 = 1 \quad n = 0,\cdots,N/2-1$$

- Sine window

$$w(n) = \sin(\pi * ((n+1/2)/(N/2))/2) \quad n = 0,\cdots,N/2-1$$

34

Steady-State vs. Transient Block Selection



35