Professor Brendan Morris, SEB 3216, brendan.morris@unlv.edu

EE482: Digital Signal Processing Applications

Audio Signal Processing

Outline

- Audio Coding
- Audio Equalizers
- Audio Effects

Audio Signal Processing

- Digital audio processing used in many consumer electronics
  - Mp3 players, televisions, etc.
- CD audio format:
  - 16-bit PCM @ 44.1Khz → stereo 1411.2 kbps
  - Great for uncompressed CD-quality sound
  - Not well-suited for modern media consumption
    - Uncompressed storage and transmission
    - Multi-channel audio (e.g. surround sound systems)
    - "Professional" audio high sampling rate (96 kHz)
- Techniques are required to enable high quality sound reproduction efficiently

Audio Coding

- Differences with speech
  - Much wider bandwidth (not just 300-1000 Hz)
  - Uses multiple channels
  - Psychoacoustic principles can be utilized for coding
    - Do not code frequency components below hearing threshold
- Lossy compression used based on noise shaping
  - Noise below masking threshold is not audible
- Entropy coding applied
  - Large amount of data from high sampling rate and multi-channels

Audio Codec

- Codec = coder-decoder

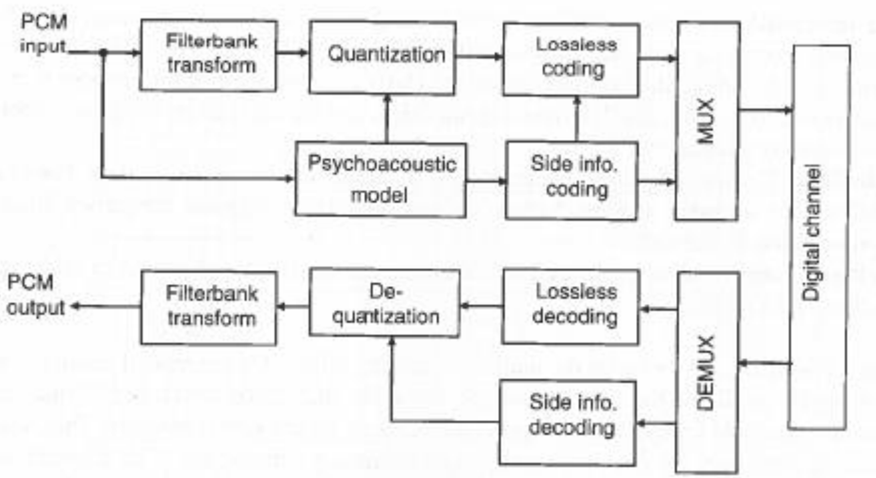


Figure 10.1 Basic structure of audio CODEC

- Filterbank transform
  - Convert between full-band signal (all frequencies) into subbands (modified discrete cosine transform MDCT)
- Psychoacoustic model
  - Calculates thresholds according to human masking effects and used for quantization of MDCT
- Quantization
  - MDCT coefficient quantization of spectral coefficients
- Lossless coding
  - Use entropy coding to reduce redundancy of coded bitstream
- Side information coding
  - Bit allocation information
- Multiplexer
  - Pack all coded bits into bitstream

Encoded Bit Stream

Figure 10.2 Typical encoded audio bit-stream format

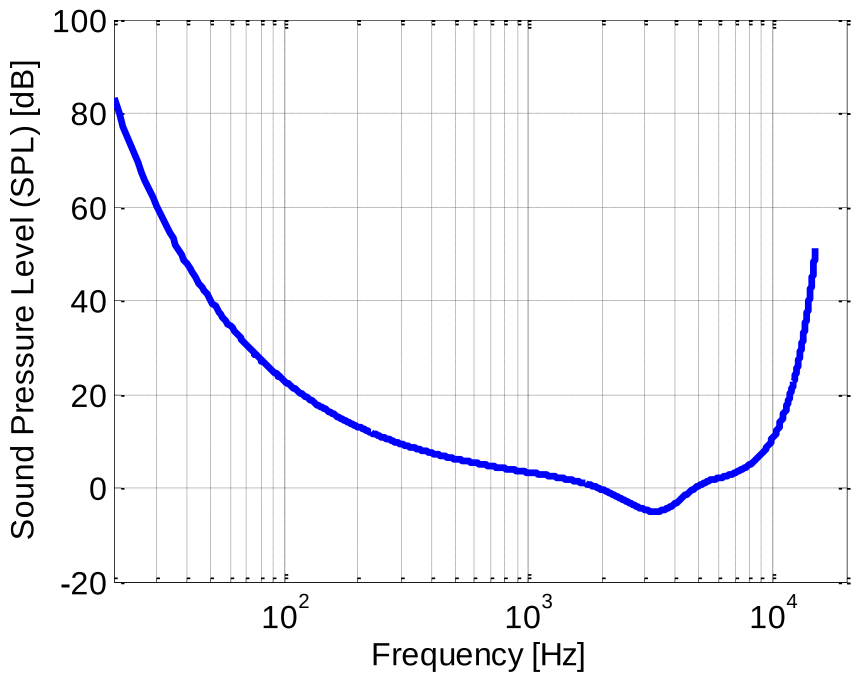
- Header
  - Contains frame format information and synchronization word (e.g. bit rate, sampling frequency, etc.)
- CRC – cyclic redundancy check
  - Error detection code to protect the header
- Side information
  - Decoder information (parameters)
- Main data
  - Coded spectral coefficients and lossless encoded data
- Ancillary information
  - User defined info (e.g. track title, album, etc.)

Auditory Masking Effects

- Psychoacoustic principle that a low-level signal (maskee) becomes inaudible when a louder signal (masker) occurs simultaneously
- Human hearing does not respond equally to all frequency components
- Auditory masking depends on the spectral distribution of masker and maskee
  - These will vary in time
- Will do noise shaping during encoding to exploit human hearing

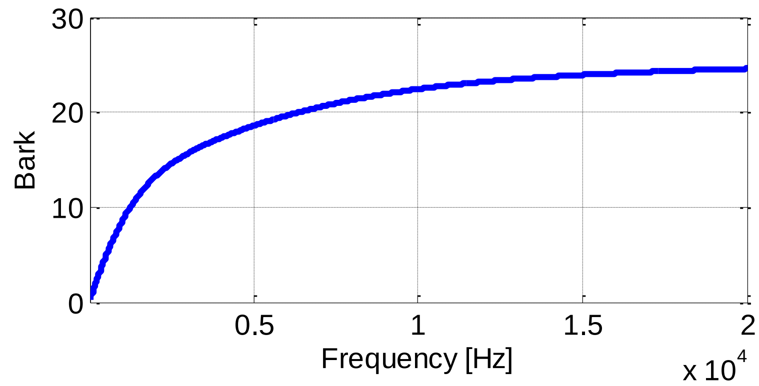
Quiet Threshold

- First step of perceptual coding
  - Shape coding distortion spectrum
- Represent a listener with acute hearing
  - No signal level below threshold will be perceived
- Quiet (absolute) threshold
  - $T_q(f) = 3.64\left(\frac{f}{1000}\right)^{-0.8} - 6.5e^{-0.6\left(\frac{f}{1000}-3.3.\right)^2} + 10^{-3}\left(\frac{f}{1000}\right)^4$ dB
- Most humans cannot sense frequencies outside of 20-20k Hz
  - Range changes in time and narrows with age



Masking Threshold

- Threshold determined by stimuli at a given time
  - Time-varying threshold
- Human hearing non-linear response to frequency components
- Divide auditory system into 26 critical bands (barks)
  - $z(f) = 13\tan^{-1}(0.00076f) + 3.5\tan^{-1}[(f/$



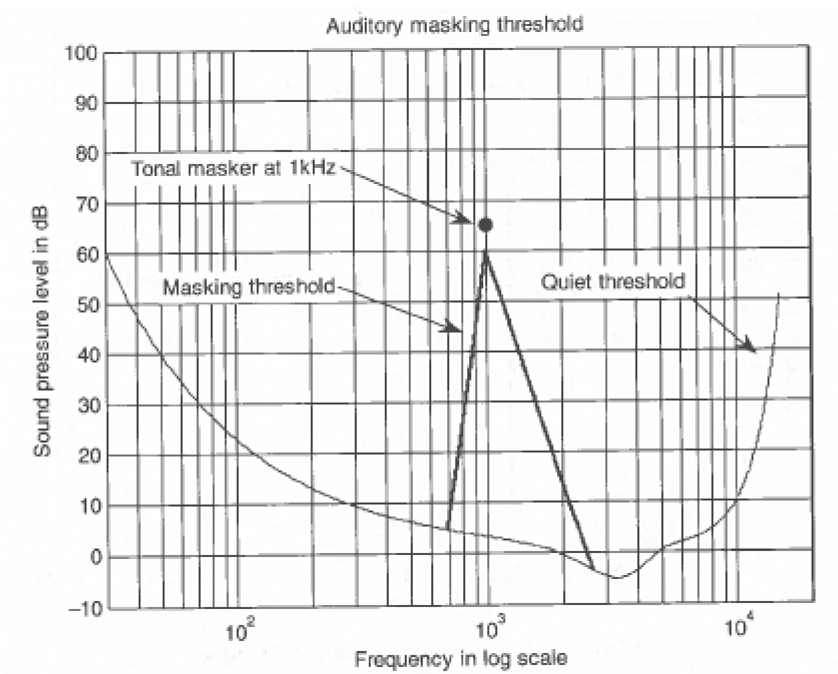


Figure 10.3 Auditory masking thresholds

Example 10.1

- Masking with multiple tones
  - 65 dB tone at 2kHz
  - 40 dB tone at 1.5 and 2.5 kHz
- Use quiet threshold first
  - All pass absolute threshold
- Using masking threshold
  - 65 dB tone is dominant
- Simultaneous masking
  - Examine barks (no overlap)
- Masking spread
  - 2.5 kHz masked
  - 1.5 kHz needs to be coded

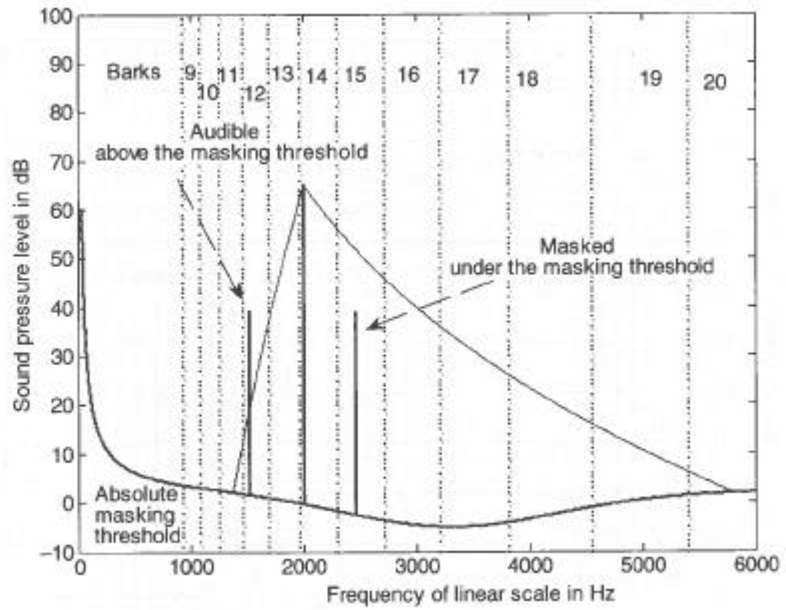


Figure 10.4 Masking effect of a 2 kHz tone

Frequency Domain Coding

- Representation of frequency content of signal
- Modified discrete cosine transform (MDCT) widely used for audio
  - DCT energy compaction (lower # of coefficients)
  - Reduced block effects
- MDCT definition
  - $X(k) = \sum_{n=0}^{N-1} x(n) \cos\left[\left(n + \frac{N+2}{4}\right)\left(k + \frac{1}{2}\right)\frac{2\pi}{N}\right]$
  - $x(n) = \sum_{k=0}^{N/2-1} X(k) \cos\left[\left(n + \frac{N+2}{4}\right)\left(k + \frac{1}{2}\right)\frac{2\pi}{N}\right]$
    - $n = 0,1, \dots, N-1$
    - $k = 0,1, \dots, (N/2) - 1$
  - Notice half coefficients for each window
    - Lapped transform (designed with overlapping windows built in)
- Like with FFT, windows are used but muse satisfy more conditions (Princen-Bradley condition)
  - Window applied both to analysis (MDCT) and synthesis (iMDCT) equations

Audio Coding

- Entropy (lossless) coding removes redundancy in coded data without loss in quality
- Pure entropy coding (lossless-only)
  - Huffman encoding – statistical coding
    - More often occurring symbols have shorter code words
    - Fast method using a lookup table
  - Cannot achieve very high compression
- Extended lossless coding
  - Lossy coder followed by entropy coding
  - 20% compression gain
    - MP3 – perceptual coding followed by entropy coding
- Scalable lossless coding
  - Can have perfect reproduction
  - Input first encoded, residual error is entropy coded
  - Results in two bit streams
    - Can choose lossy lowbit rate and combine for high quality lossless

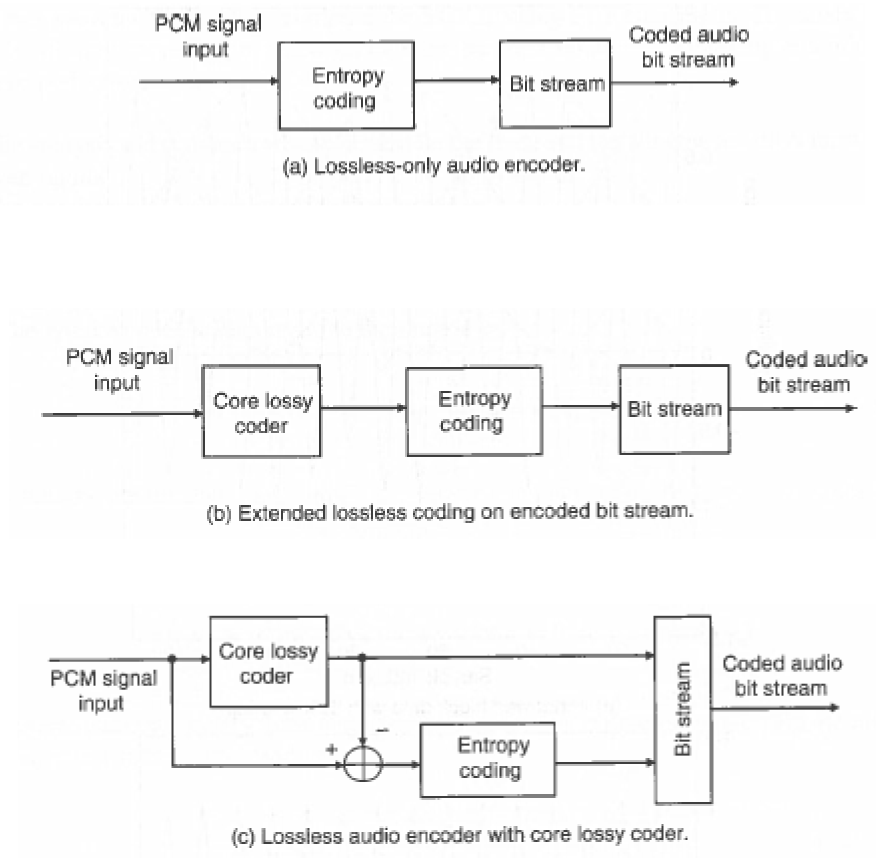


Figure 10.6 Three lossless coder structures

MP3 Algorithm

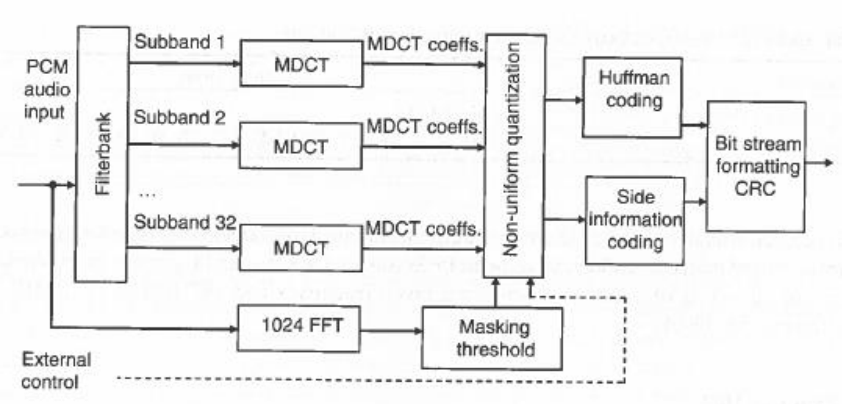


Figure 10.7 Block diagram of MP3 encoder

- Filterbank splits audio into 32 subbands
  - Each decimated to 32-36 MDCT coefficients
  - 32 x 36 = 1152 samples per frame
- Each band processed separately
  - MDCT block length of 18 and 6
    - Requires windows of 36 and 12 for 50% overlap
  - Longer blocks give better frequency resolution (stationary signals)
  - Shorter block length for better time resolution during transients
- 1152 samples per frame (~26 msec)
  - 1152/2 = 576 MDCT coefficients/frame
- Coefficients quantized using psychoacoustic model with masking threshold computed using 1024-pont FFT coefficients
- Control parameters
  - Sampling rate (kHz) – 48, 44.1, 32
  - Bit rate (kbps) – 320, 256, …, 32
- Huffman coding of quantized MDCT coefficients
  - Arrange coefficients in order of increasing frequency
    - More energy in lower frequencies
  - Results in more efficient Huffman coding
- Frequency bins divided into three regions for efficient coding
  - Run-zero – high frequency area with no energy
  - Count-1 – are containing values of [-1, 0, 1]
  - Big-value – coded with high precision
    - Further divided into 3 sub regions and each is Huffman coded

Audio Equalizers

- Spectral equalization uses filtering techniques to reshape magnitude spectrum
  - Useful for recording and reproduction

- Example uses
  - Simple filters to adjust bass and treble
  - Correct response of microphone, instrument pick-ups, loudspeakers, and hall acoustics

- Parametric equalizers provide better frequency compensations but require more operator knowledge than graphic equalizers

Graphic Equalizers

- Use of several frequency bands to display and adjust the power of audio frequency components

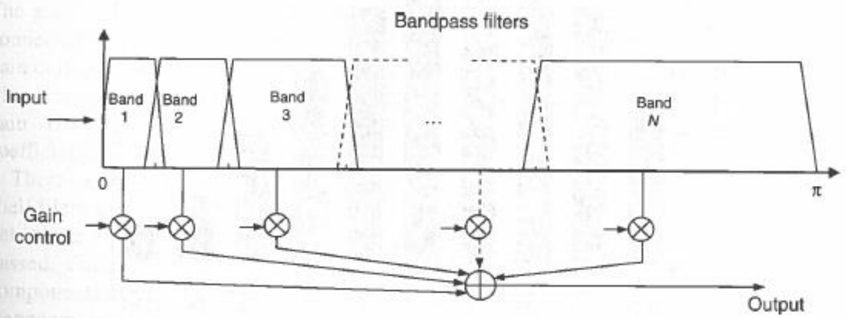


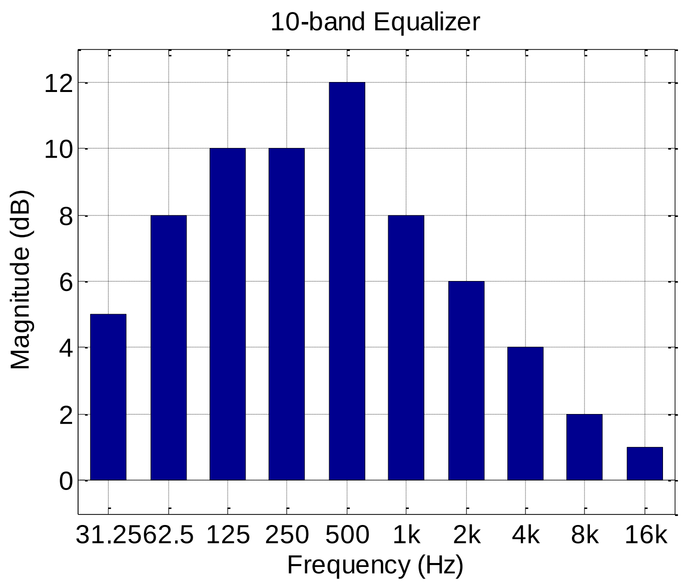
Figure 10.9 An example of an *N*-band graphic equalizer using octave scale

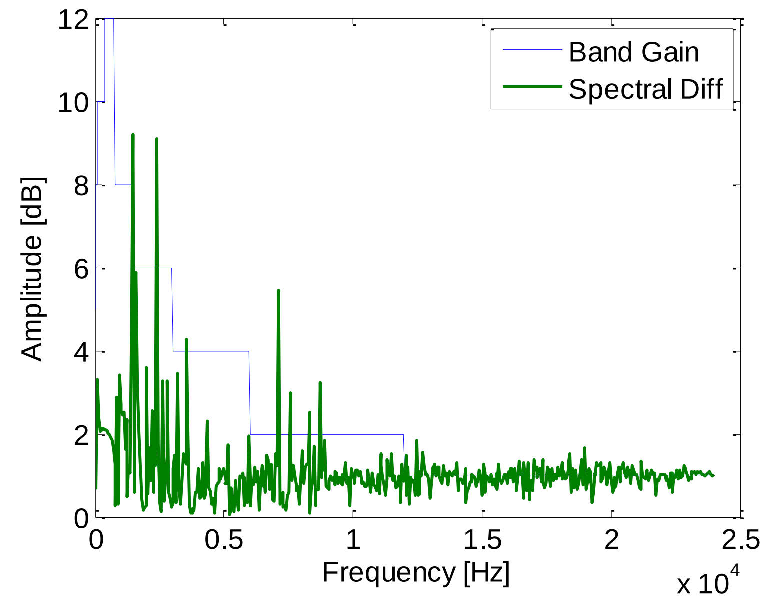
- Input signal decomposed with bank of parallel bandpass filters
- Separate gain control for each band
- Signal power in each band estimated and displayed graphically with a bar
- Divide spectrum using octave scale (doubling scale)
- Bandpass filters can be realized using IIR filter design techniques
- DFT bins of audio signal $X(k)$ need to be combined to form the equalizer frequency bands
  - Use octave scaling to combine

Example 10.4

- Graphic equalizer to adjust signal

- Select bands
  - Use octave scaling
  - `bandFreqs = {'31.25','62.5','125','250','500', '1k','2k','4k','8k','16k'};`





Parametric Equalizers

- Provides a set of filters connected in cascade that are tunable in terms of both spectral shape and filter gain
  - Not fixed bandwidth and center as in graphic
  - Use 2nd-order IIR filters

- Parameters:
  - f_s - sampling rate
  - f_c - cutoff frequency [center (peak) or midpoint (shelf)
  - Q – quality factor [resonance (peak) slope (shelf)]
  - $Gain$ – boost in dB (max $\pm$12 dB)

Shelf Filters

- Low-shelf
  - Boost frequencies below cuttoff and pass higher components
- High-shelf
  - Boost frequencies above cuttoff and pass rest
- See book for equations

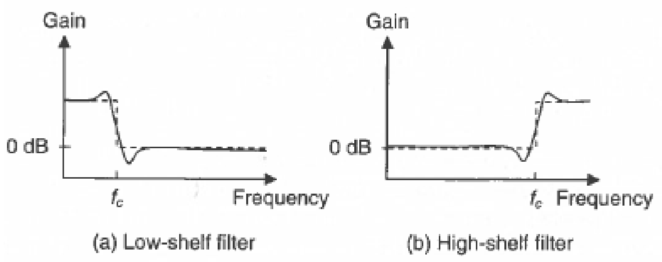
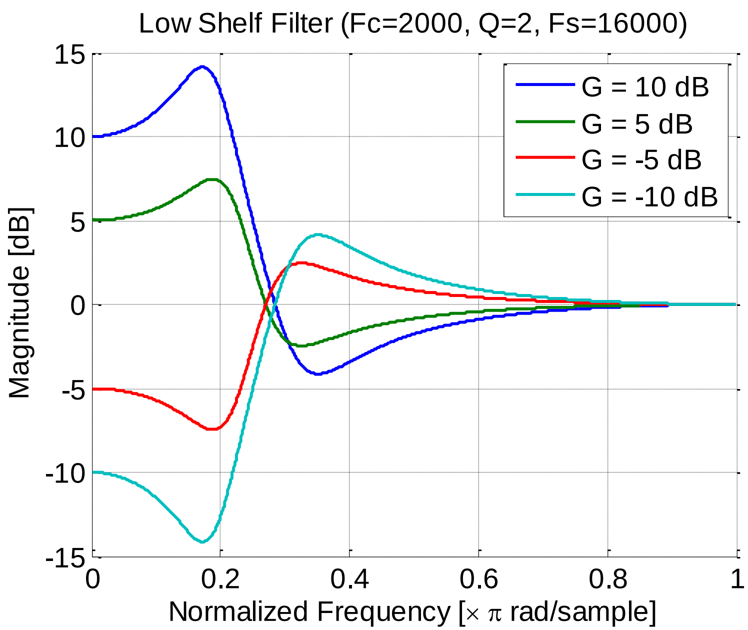
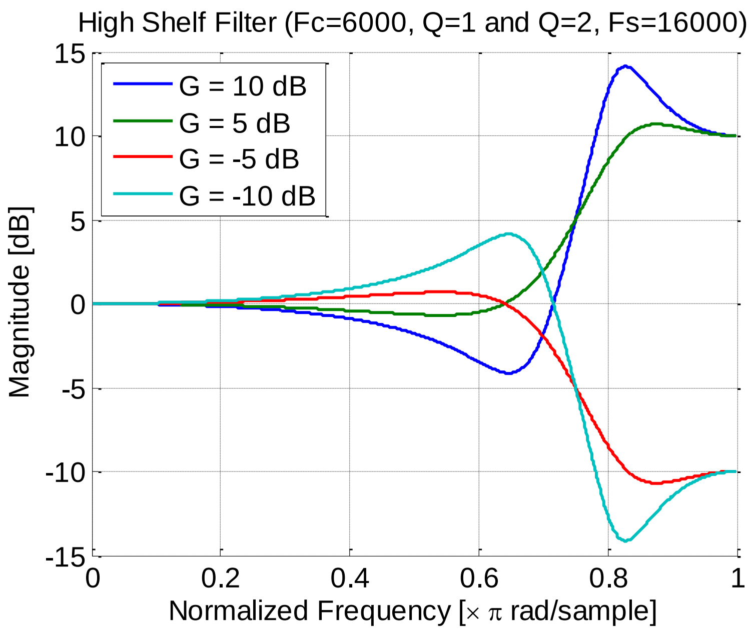


Figure 10.11 Magnitude responses of shelf filters

- Ex 10.6
  - Shape of shelf filter with different gain parameters

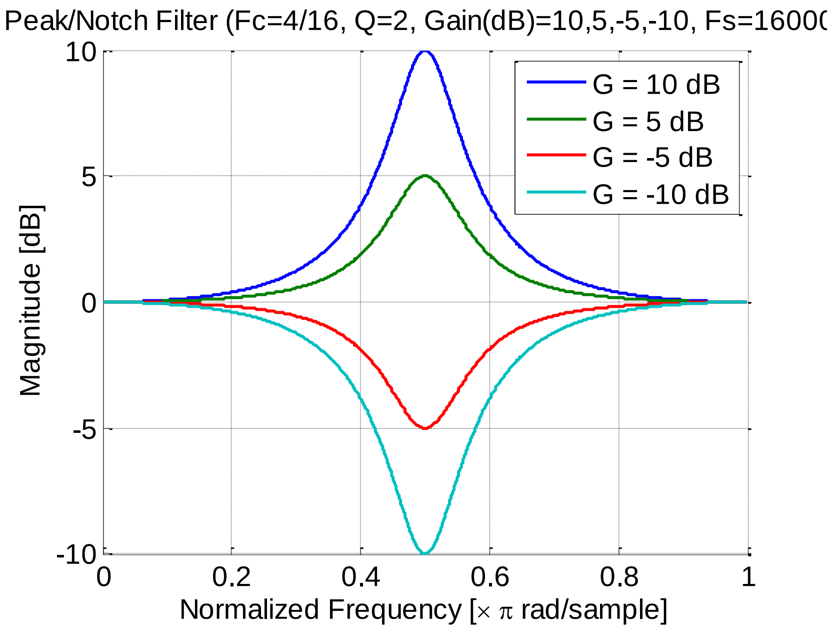




Peak Filter

- Peak filter – amplify certain narrow frequency bands
- Notch filter – attenuate certain narrow frequency bands
- E.g. loudness of certain frequency
- See book for equations

- Ex 10.5
  - Shape of peak filter for different parameters



Example 10.7

- Implement parametric equalizer
  - $f_s = 16{,}000$ Hz
- Cascade 3 filters:
  - Low-shelf filter
    - $f_c = 1000, Gain = -10\ dB, Q = 1.0$
  - High-shelf filter
    - $f_c = 4000, Gain = 10\ dB, Q = 1.0$
  - Peak filter
    - $f_c = 7000, Gain = 10\ dB, Q = 1.0$
- Play example file outside of powerpoint
  - Left channel – original signal
  - Right channel - filtered

Audio (Sound) Effects

- Use of filtering techniques to emphasize audio signal in "artistic" manner

- Will only mention and give examples of some common effects
  - Not an in-depth look

Sound Reverberation

- Reverberation is echo sound from reflected sounds
- The echoes are related to the physical properties of the space
  - Room size, configuration, furniture, etc.
- Use impulse response to measure

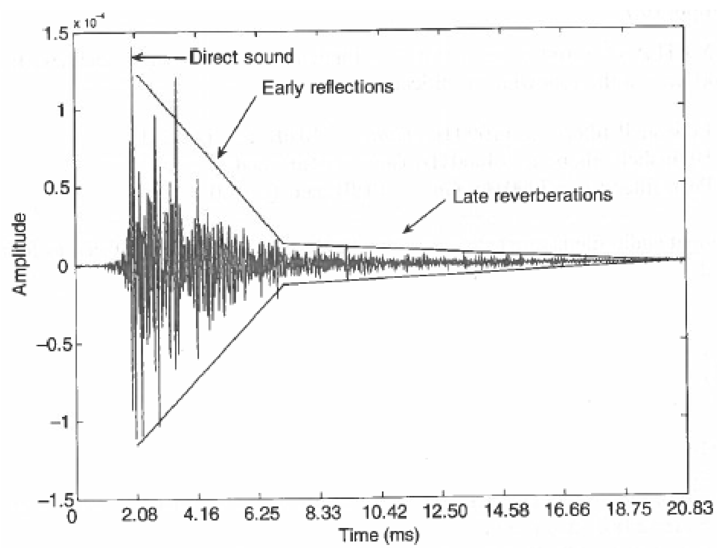


Figure 10.15 An example of a room impulse response

- Direct sound
  - First sound wave to reach ear
- Reflected sound
  - The echo waves that arrive after bouncing off a surface

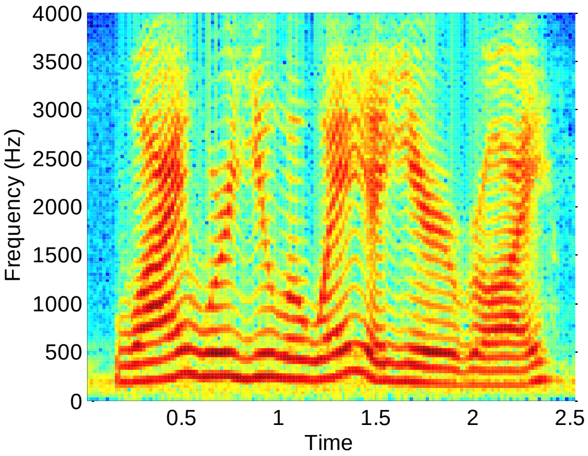
- Example 10.8
- Use hall impulse response to simulated reverberated sound

- Input

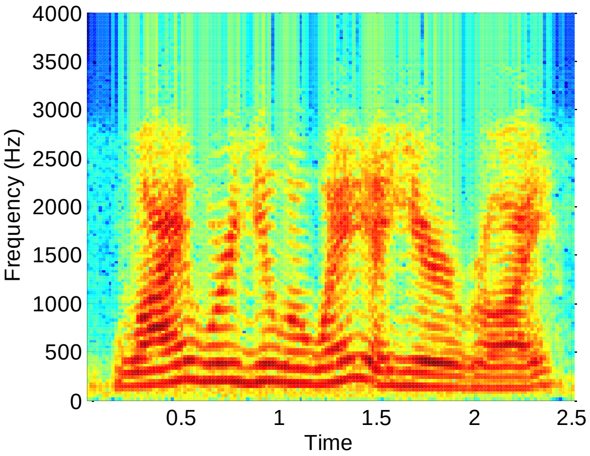
- Output

Pitch Shift

- Change speech pitch (fundamental frequency)
- All frequencies are adjusted over the entire signal
  - Chipmunk voice
- Example 10.9a
  - Adjust pitch
- See audio files



original

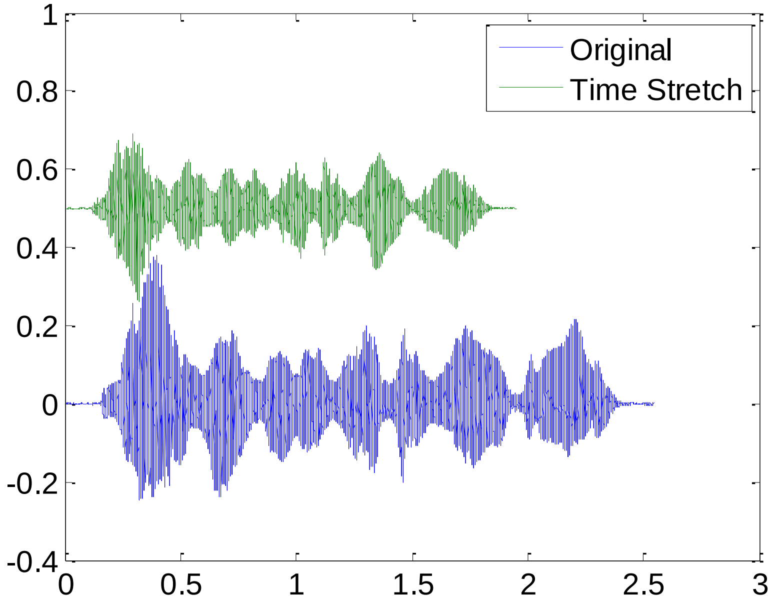


pitch shifted

Time Stretch

- Change speed of audio playback without affecting pitch
- Audio editing: adjust audio to fit a specific timeline

- Example 10.9b
  - Adjust play time
- See audio files



Tremolo

- Amplitude modulation of audio signal
  - $y(n) = [1 + AM(n)]x(n)$
    - A – max modulation amplitude
    - $M(n)$ – slow modulation oscillator

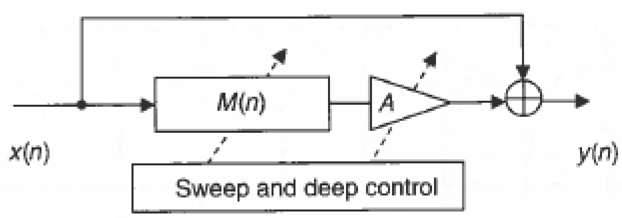
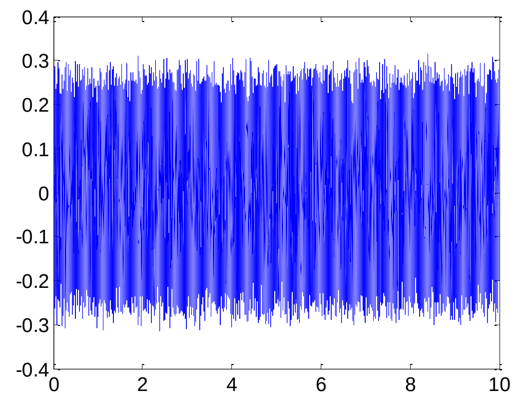
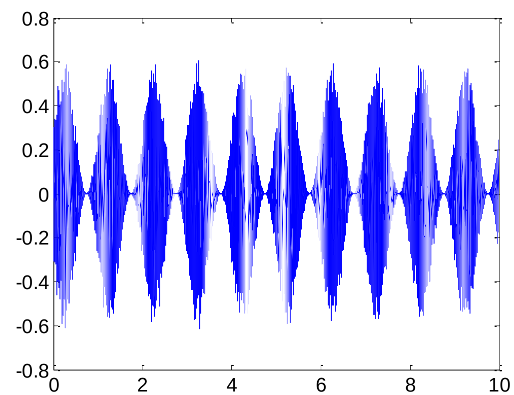


Figure 10.23 A block diagram of tremolo using low-frequency modulation

  - $M(n) = \sin(2\pi f_r nT)$
    - f_r - modulation rate

- Example 10.10
  - $A = 1, f_r = 1\ Hz$
  - White noise input at $f_s = 8000\ Hz$





Spatial Sounds

- Audio source localization determined by the way it is perceived by human ears
  - Time delay and intensity differences

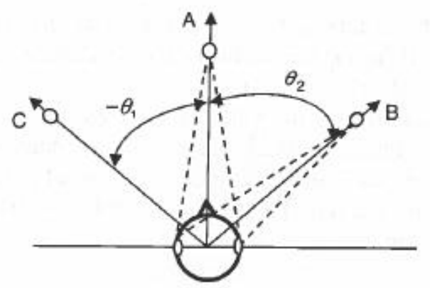


Figure 10.25 The sound source A directly in front of the head and the source B displaced at the θ_2^o azimuth

- Sounds in different positions arrive differently at ears
  - Interaural time difference (ITD) - delay between sounds reaching ear for localization
  - Iteraural intensity difference (IID) - loudness difference for localization

- Binaural audio demos
  - Great home fun
  - http://www.youtube.com/watch?v=IUDTlvagjJA
  - http://www.youtube.com/watch?v=3FwDa7TWHHc
  - http://www.qsound.com/demos/binaural-audio.htm
  - http://www.studio360.org/story/126833-adventures-3d-sound/