Audio DSP

Professor Deepa Kundur

University of Toronto

Sound

- Sound: vibration transmitted through a medium (gas, liquid, solid and plasma) composed of frequencies capable of being detected by ears.
  - Note: sound cannot travel through a vacuum.
- Human detectable sound is often characterized by air pressure variations detected by the human ear.
- The amplitude, frequency and relative phase of the air pressure signal components determine (in part) the way the sound is perceived.

Sinusoids and Sound: Amplitude

- A fundamental unit of sound is the sinusoidal signal.

$$x_a(t) = A\cos(2\pi F_0 t + \theta), \qquad t \in \mathbb{R}$$

  - $A \equiv$ volume
  - $F_0 \equiv$ pitch (more on this ...)
  - $\theta \equiv$ phase (more on this ...)

Sound Volume

- Volume = Amplitude of sound waves/audio signals
- quoted in dB, which is a logarithmic measure; $10\log(A^2)$
  - no sound/null is $-\infty$ dB
- Loudness is a subjective measure of sound psychologically correlating to the strength of the sound signal.
  - the volume is an objective measure and does not have a one-to-one correspondence with loudness
  - perceived loudness varies from person-to-person and depends on frequency and duration of the sound

Music Volume Dynamic Range

Tests conducted for the musical note: C6 ($F_0 = 1046.502$ Hz).

Dynamic Level	Decibels
Threshold of hearing	0
ppp (pianissimo)	40
p (piano)	60
f (forte)	80
fff (fortississimo)	100
Threshold of pain	120

Sinusoids and Sound: Frequency

- A fundamental unit of sound is the sinusoidal signal.

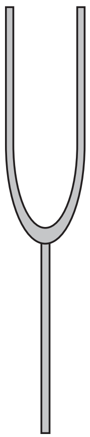
$$x_a(t) = A\cos(2\pi F_0 t + \theta), \qquad t \in \mathbb{R}$$

  - $A \equiv$ volume
  - $F_0 \equiv$ pitch
  - $\theta \equiv$ phase (more on this ...)

Pure Frequency

- **Q:** What type of sound does a pure frequency produce?
  - **A:** A pure tone with a single pitch.
- **Q:** Can any instrument produce a pure tone by playing a single note?
  - **A:** No.

Tuning Forks



- A tuning fork is a two-pronged instrument that is an acoustic resonator. It is usually made out of steel and resonates at a specific constant pitch which is a function of the length of the prongs.
  - Striking the tuning fork will produce the required sounds although initially there may be overtones that die out quickly.
  - A very common tuning fork used by musicians produces the A note ($F_0 = 440$ Hz), which is international concert pitch used to tune orchestras.

Frequency and Pitch

- Sinusoids can be represented either as:

$$x_a(t) = A\cos(2\pi F_0 t + \theta), \qquad t \in \mathbb{R}$$

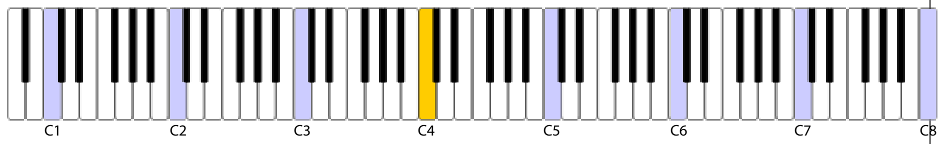
  or for mathematical convenience when interpreting as Fourier signal components as:

$$x_a(t) = Ae^{j(2\pi F_0 t + \theta)}, \qquad t \in \mathbb{R}$$

- Pitch is directly related to the frequency F_0.
- To be able to hear a frequency F_0, it has to be in the human audible range (20 Hz to 20,000 Hz).

Harmonically Related Frequencies and Pitch

Scientific Designation	Frequency (Hz)	k for $F_0 = 8.176$
C1	32.703	4
C2	65.406	8
C3	130.813	16
C4 (middle C)	261.626	32
C5	523.251	64
C6	1046.502	128
C7	2093.005	256
C8	4186.009	512



Harmonically Related Frequencies

- Recall harmonically related sinusoids have the following analytic form for $k \in \mathbb{Z}$:

$$x_{a,k}(t) = A\cos(2\pi k F_0 t + \theta)$$

  or

$$x_{a,k}(t) = Ae^{j(2\pi k F_0 t + \theta)}$$

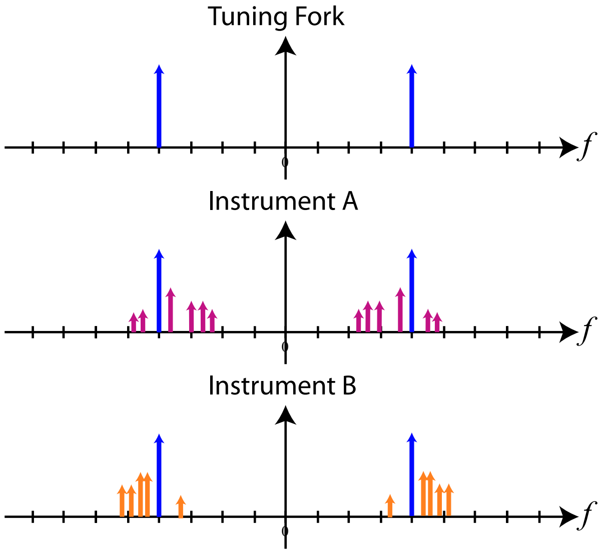
- They are used in the context of the Fourier Series to build periodic signals:

$$x(t) = \sum_{k=-\infty}^{\infty} X(k) e^{j(2\pi k F_0 t)}$$

Signature Sounds

- **Q:** If two different people sing the same note or two different instruments play the same note, why do they sound different?
  - The notes are not pure tones. There are natural overtones and undertones that provide distinguishing signatures that can be viewed in the associated spectra.

Fourier Transforms of the Same Note



Human Audible Range

- Hearing is usually limited to frequencies between 20 Hz and 20 kHz.
- The upper limit decreases with age.
  - The audible frequency range is different for animals

Animal Audible Range

Species	Approx Range (Hz)
human	20 - 20,000
dog	67 - 45,000
rabbit	360 - 42,000
bat	2,000 - 110,000
goldfish	20 - 3,000

Reference: R.R. Fay (1988), Hearing in Vertebrates: A Psychophysics Databook.

Sinusoids and Sound: Phase

- A fundamental unit of sound is the sinusoidal signal.

$$x_a(t) = A\cos(2\pi F_0 t + \theta), \qquad t \in \mathbb{R}$$

  - $A \equiv$ volume
  - $F_0 \equiv$ pitch
  - $\theta \equiv$ phase

Phase and Sound

- An audio signal is represented by a real function $x(t)$.
- The function $x(-t)$ represents playing the audio signal backwards.
- Since $x(t)$ is real:

$$\begin{aligned} X(F) &= X^*(-F) \\ |X(F)| &= |X^*(-F)| = |X(-F)| \quad \text{since } |c| = |c^*| \text{ for } c \in \mathbb{C} \end{aligned}$$

- Therefore,

$$|X(F)| = |X(-F)|$$

  That is, the CTFT magnitude is even for $x(t)$ real.

Phase and Sound

- Recall, $\quad x(t) \overset{\mathcal{F}}{\longleftrightarrow} X(F) \qquad x(-t) \overset{\mathcal{F}}{\longleftrightarrow} X(-F)$
- Therefore,

$$\underbrace{|X(F)|}_{\text{spectrum magnitude of } x(t)} = \overbrace{|X(-F)|}^{\text{spectrum magnitude of } x(-t)}$$

Therefore, the magnitude of the FT of an audio signal played forward and backward is the same!

Phase and Sound

- Therefore, for

$$\begin{aligned} x(t) &\overset{\mathcal{F}}{\longleftrightarrow} X(F) \\ x(-t) &\overset{\mathcal{F}}{\longleftrightarrow} X(-F) \end{aligned}$$

- $|X(F)| = |X(-F)| \Rightarrow$ the CTFT magnitudes for forward and reverse sound signals are exactly the same.
- $\angle X(f) \neq \angle X(-f) \Rightarrow$ the CTFT phases for forward and reverse sound signals are different.
- Therefore, the relative phase of the sinusoidal components of sound contains very salient perceptual information much like for images.

Auditory Masking

- occurs when the perceived quality of one (primary) sound is affected by the presence of another (secondary) sound
  - Simultaneous masking: the secondary sound is heard at the same time as the primary sound
- Can be exploited to mask <u>non-ideal signal processing</u>.

Why Digitize Audio?

- Fidelity of digital audio is much higher than analog audio.
- Manipulation tools for digital audio are much more sophisticated than those available for analog audio.
- Compression of digital audio provides significantly reduced storage requirements.
- Storage of digital audio (e.g., CDs) are much more convenient and compact.
- Duplication of digital audio is exact in contrast to analog audio.

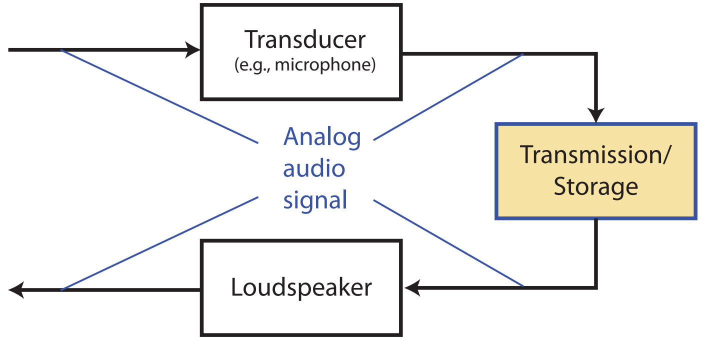
Benefits of Digital Audio

- Convenient recording, enhancement, mass-production and distribution.
  - CDs, online stores such as iTunes, etc.
  - data files are distributed instead of physical media storing the information such as records and tapes.

Concerns about Digital Audio

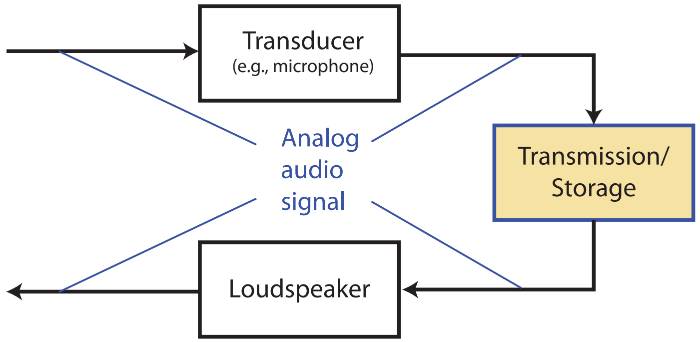
- Convenient recording, enhancement, mass-production and distribution.
  - unlawful manipulation of recorded audio is difficult to detect
  - piracy: unlawful copying and redistribution of copyrighted content

Analog vs. Digital Audio: Analog Audio System



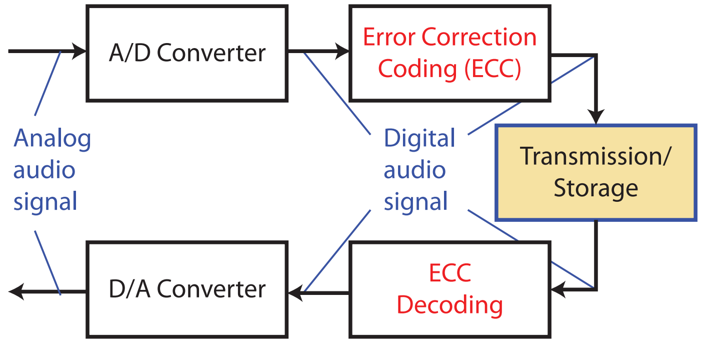
- microphone: converts sound into an electrical signal;
  air pressure → motion of conductor/coil → magnetic field → electrical signal
- loudspeaker: converts electrical signal into acoustic waves;
  electrical signal → magnetic field → motion → air pressure

Analog vs. Digital Audio: Analog Audio System



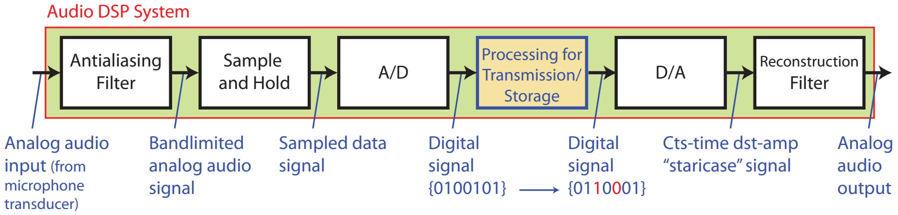
- associated circuits suffer from inherent noise (noise floor)
- capacitance and inductance of the circuits limit bandwidth, and resistance limits amplitude

Analog vs. Digital Audio: Digital Audio Chain

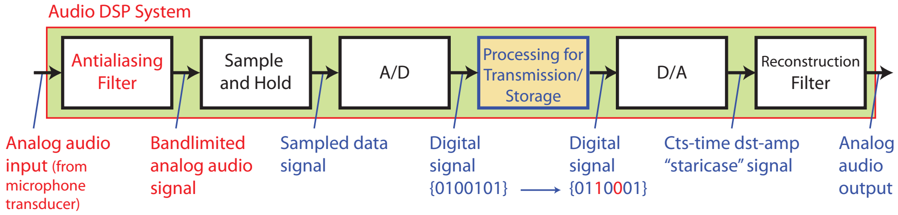


- fidelity limited by quantization noise
- bandwidth limited by sampling rate
- dynamic range limited by bit resolution

Digitizing Audio



Digitizing Audio

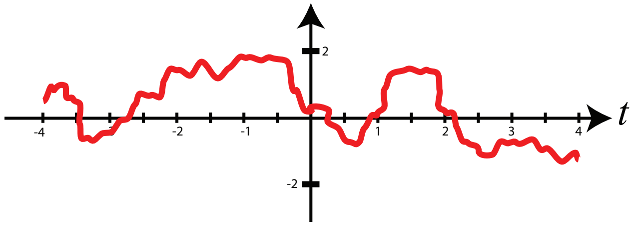


Anti-aliasing Filter:

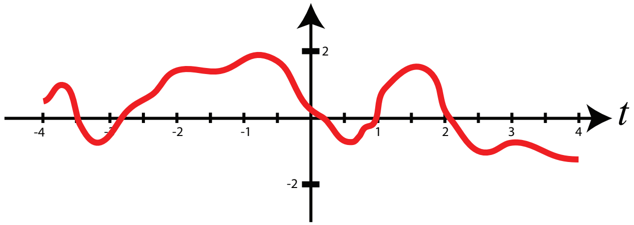
- ensures that analog audio input does not contain frequency components higher than half of the sampling frequency (to avoid aliasing)
- Example: C6713 DSP, $F_s = 8$ kHz, therefore anti-aliasing filter must have a passband of 0 Hz to 4000 Hz.

Digitizing Audio

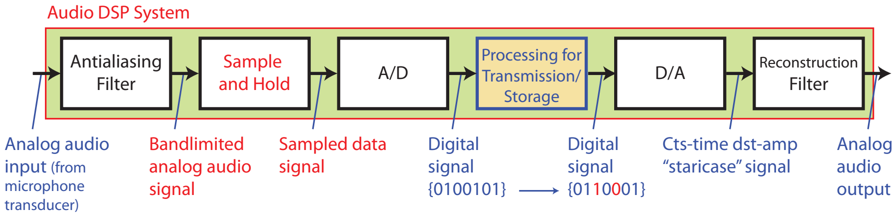
Input Signal



Anti-aliased Signal



Digitizing Audio

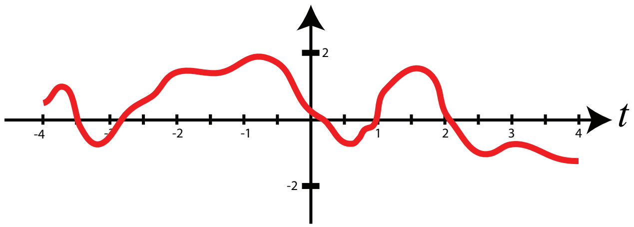


Sample and Hold:

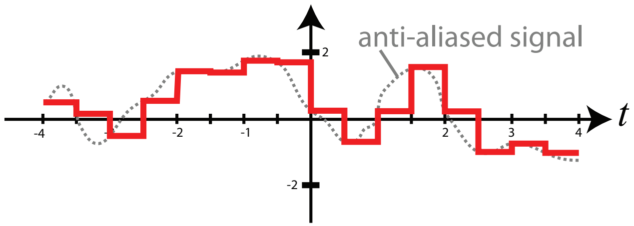
- holds a sampled analog audio value for a short time while the A/D converts and interprets the value as a digital

Digitizing Audio

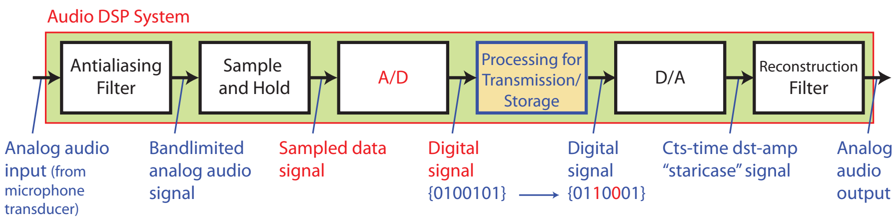
Anti-aliased Signal



Sampled Data Signal



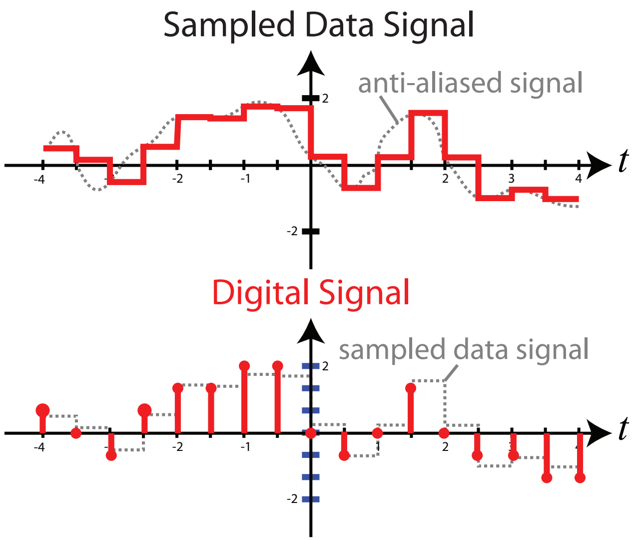
Digitizing Audio



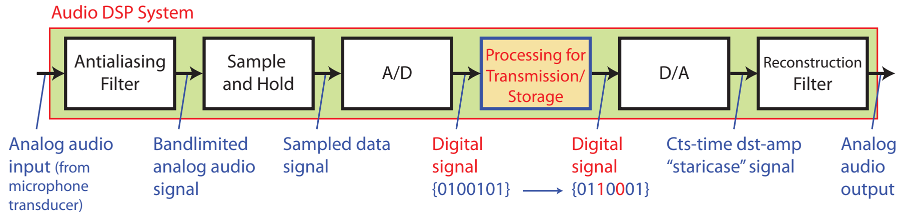
A/D:

- converts a sampled data audio value into a digital number, in part, through quantization of the amplitude

Digitizing Audio



Digitizing Audio



Processing for Transmission/Storage:

- transmission/storage contains inherent non-idealities that cause errors in the received/retrieved data symbols
- error correction coding (ECC) is employed to add redundancy to the digital signal so that errors can be compensated for during decoding

Error Correction Coding

Example: N-repetition code

Input Signal Bit	Coded Sequence
0	$\underbrace{0\ 0\ 0\ \cdots\ 0}_{N \text{ zeros}}$
1	$\underbrace{1\ 1\ 1\ \cdots\ 1}_{N \text{ ones}}$

Therefore, for $N = 3$ the following input signal sequence: 0 0 1 would be coded as follows: 0 0 0 0 0 0 1 1 1.

Error Correction Coding

Q: How would you interpret receiving the following coded sequence (with possible error):

1 1 1 0 1 0 0 0 0?

$$\underbrace{1\ 1\ 1}_{1}\quad \underbrace{0\ 1\ 0}_{0}\quad \underbrace{0\ 0\ 0}_{0}$$

A: Decoding can make use of majority vote logic.

Error Correction Coding

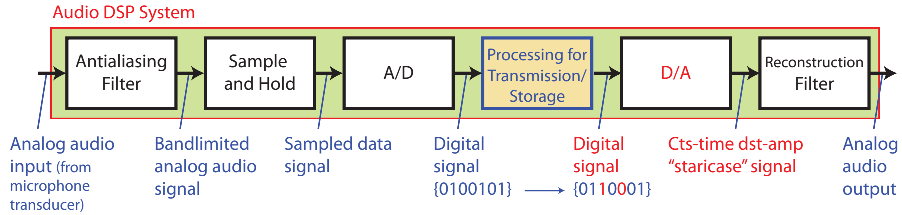
Coder for $N = 3$:

Input Signal Bit	Coded Sequence
0	0 0 0
1	1 1 1

Majority vote logic decoder for $N = 3$:

Received Coded Seq	Decoded Signal Bit
0 0 0	0
0 0 1	0
0 1 0	0
0 1 1	1
1 0 0	0
1 0 1	1
1 1 0	1
1 1 1	1

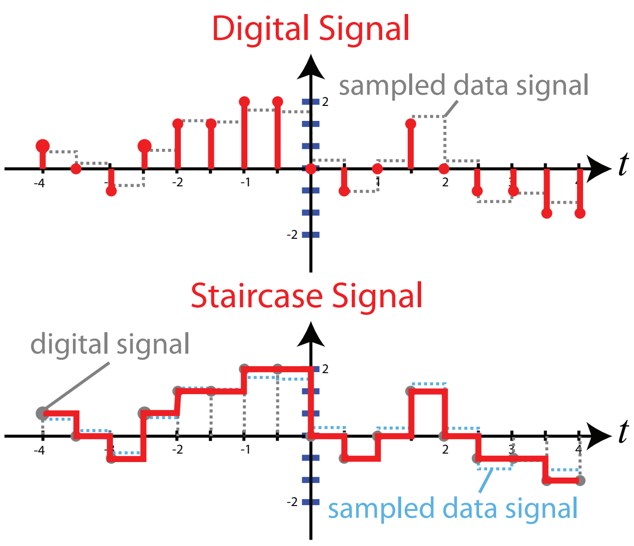
Digitizing Audio



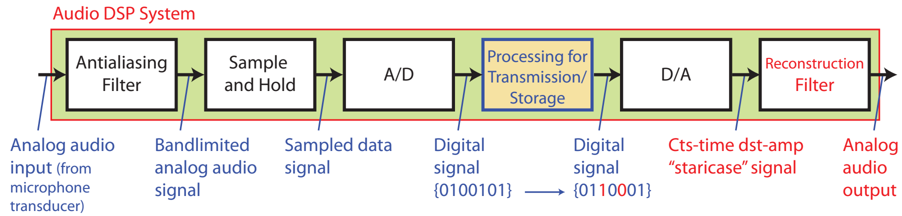
D/A:

- converts a digital audio signal into a "staircase"-like signal for further reconstruction

Digitizing Audio



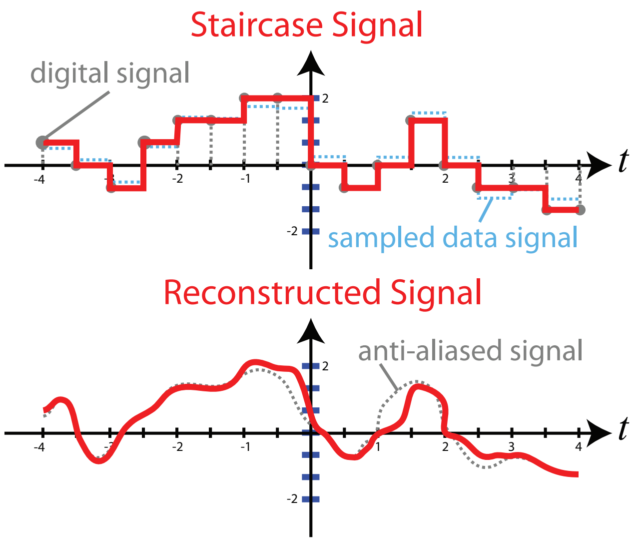
Digitizing Audio



Reconstruction Filter:

- converts a "staircase"-like signal into an analog filter through lowpass filtering
- depending on the application the filter can be similar to the anti-aliasing filter, or may be very cheap (e.g., compact disk receivers), or may using a different sampling rate for special effects

Digitizing Audio



Digitizing Audio

The "quality" of digitizing audio is related to the following parameters:

- sampling rate (Hz)
- bit depth (bits/sample) and dynamic range (related to number of quantization levels)
- mono vs. stereo

Digitizing Audio

Note: For the same cost, digital audio provides higher signal-to-noise ratio or lower mean-square error between the real sound and what is recorded/played.

- It is less expensive to increase sampling rate and quantization depth (i.e., reduce quantization noise) than to use less noisy analog circuitry (i.e., reduce noise floor)
- When signals are represented digitally the natural noise in the circuits can be circumvented via error correction coding. Thus, it is possible to have near perfect storage/transmission.

Audio Quality and Sampling Rate

Audio Quality as a Function of Sampling Rate:

Sampling Rate (Hz)	Quality Similar to
8,000	telephone
11,025	AM radio
22,050	FM radio
44,100	CD
48,000	DAT

Audio Quality, Sampling Rate, and Bit Depth

Audio Quality as a Function of Sampling Rate, Bit Depth and Stereo/Monophony:

Sampling Rate (Hz)	Bit Depth	Stereo/Mono	Quality
8,000	8	mono	telephone
11,025	8	stereo	low
22,050	8	stereo	·
22,050	16	mono	·
22,050	16	stereo	·
44,100	16	mono	good
44,100	16	stereo	CD quality

Audio Quality

Q: Why do some people insist that analog audio is superior to digital audio?

A: What they think sounds good isn't the exact original sound, but a nonlinearly distorted version generated from the analog components.

Note: Some digital audio companies now make digital amplifiers that mimic the distortion from analog audio amplifiers.

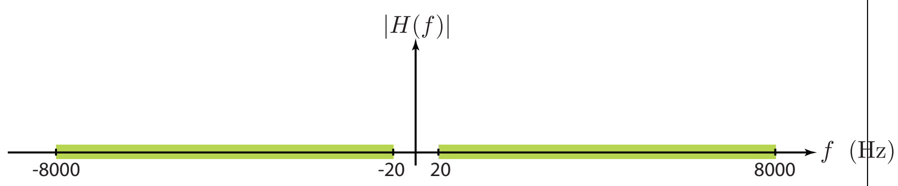
Quality of audio is a qualitative and psychological measure that is user-specific.

Audio Equalization

- Equalization $\equiv$ Equalisation $\equiv$ EQ
  - amplifying or attenuation different frequency components of an audio signal
  - Example: bass/treble control in inexpensive car radios
- Common goals of equalization:
  - provide fine granularity of frequency amplification/attenuation control without affecting adjacent frequencies.
  - correct for unwanted frequency attenuation/amplification during recording processes
  - enhancing the presence of certain sounds
  - reducing the presence of unwanted signals such as noise

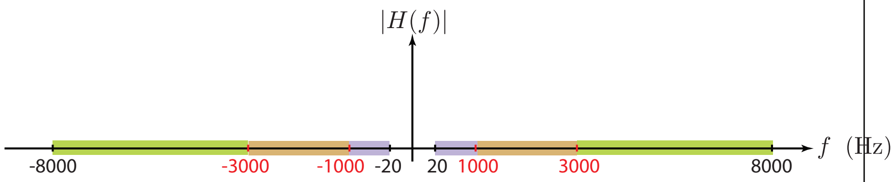
Equalizer Design Basics

1. Determine the processing band of your audio signal.
   - human audible range is: 20 Hz to 20 kHz
   - if sampling rate of a DSP is F_s then, the bandwidth of the audio signal to process is: 20 to $\frac{F_s}{2}$ Hz
   - Example: $F_s = 16,000$ Hz



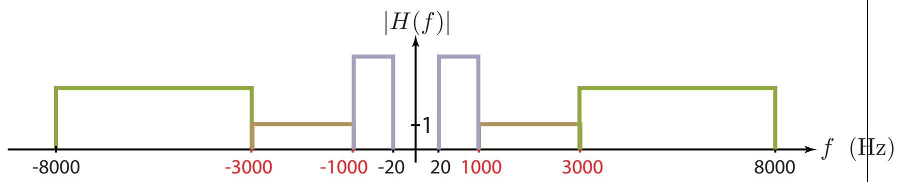
Equalizer Design Basics

2. Determine the granularity of your equalizer (i.e., number of frequency bands to independently control).
   - one approach might be to equally partition the audio signal bandwdith
   - more popular approaches suited to human auditory system models have bands that increase in width by two
   - Example: 3 frequency bands



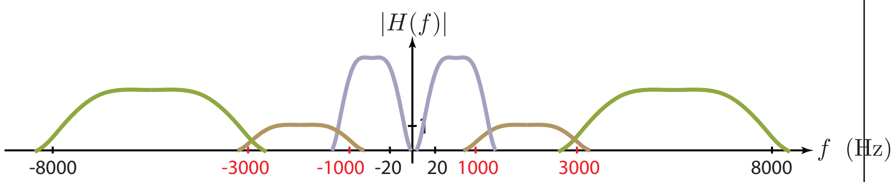
Equalizer Design Basics

3. Design your bandpass filters.
   - each bandpass filter is independently set/controlled from the others
   - ideally, many people would like shelving EQ
   - Example: Ideal bandpass filters



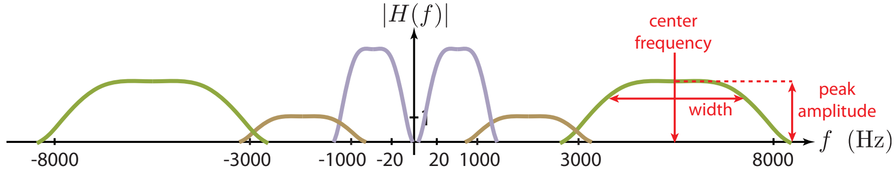
Equalizer Design Basics

3. Design your bandpass filters.
   - each bandpass filter is independently set/controlled from the others
   - ideally, many people would like shelving EQ
   - Example: Bell EQ



Common Types of Equalizers

- All bell filters and many other bandpass filters can be characterized by three parameters:
  - center frequency
  - width of the bell curve
  - gain (i.e. peak) of the bell curve



Common Types of Equalizers

- Parametric Equalizers: the center frequency, passband width and peak amplitude can be independently selected for each filter
  - most powerful EQ, predominantly used for recording and mixing
- Graphic Equalizers: the center frequency and passband width of each filter are pre-set; the gains of each filter can be independently controlled
  - used for live applications such as concerts

Common Types of Equalizers

- Notch Filters: the passband width is small and fixed for each filter; center frequencies and gains are variable.
  - used in multimedia applications/audio mastering

■