

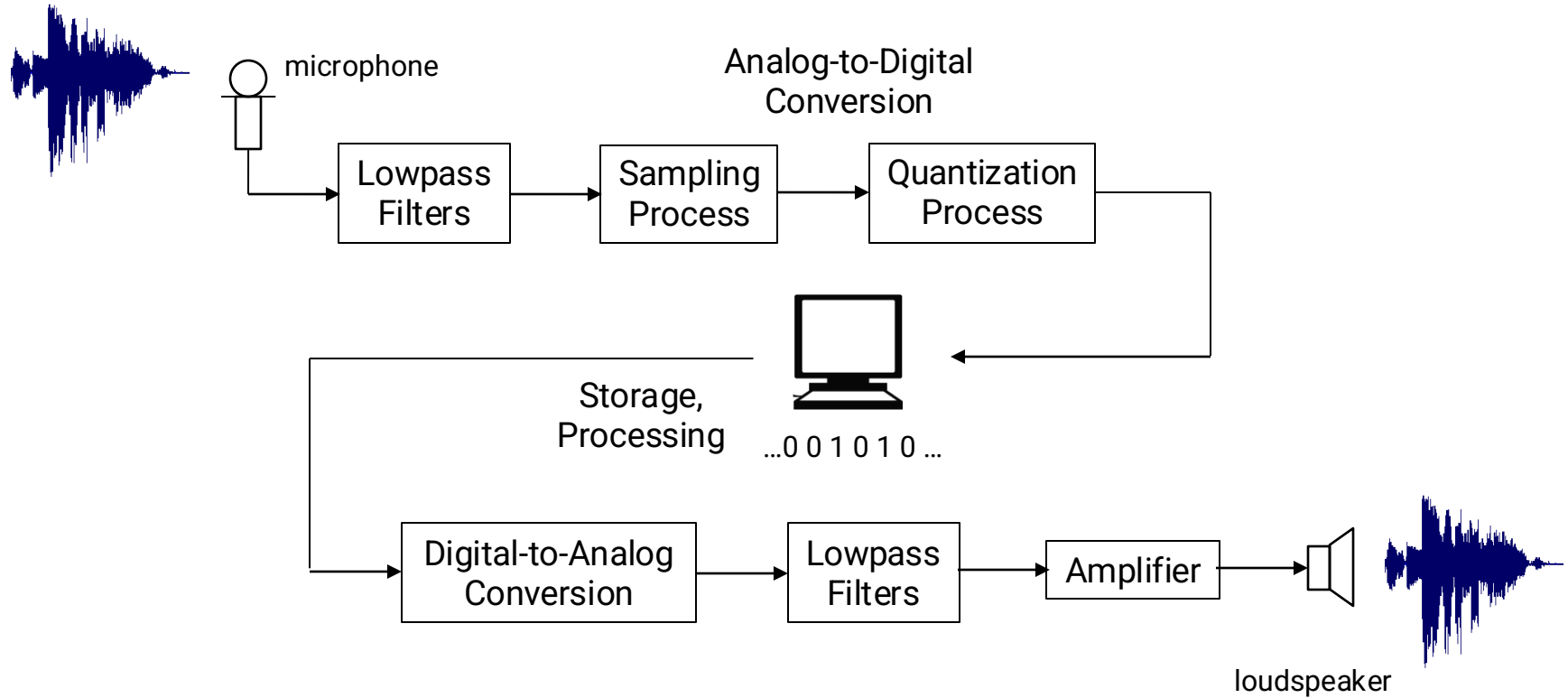
CTP431: Fundamentals of Computer Music

Digital Audio



Juhan Nam

Digital Audio Chain

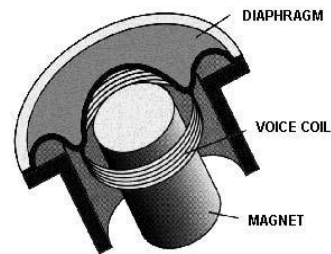
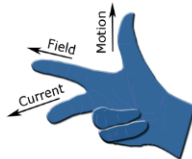


Microphones

- Convert vibrations in air to electrical signals

- Dynamic mic

- Fleming's right-hand rule
- Vocal or drum in a live setting

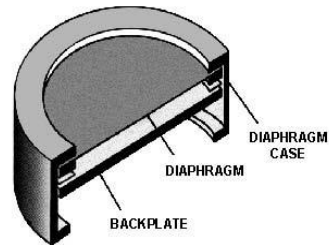


Shure SM58

Dynamic mic

- Condenser mic

- $Q = CV$ (capacitor)
- Need phantom power
- Studio recording
 - Sensitive and accurate: capture delicate sounds (quite or high-frequency sound)
- Expensive

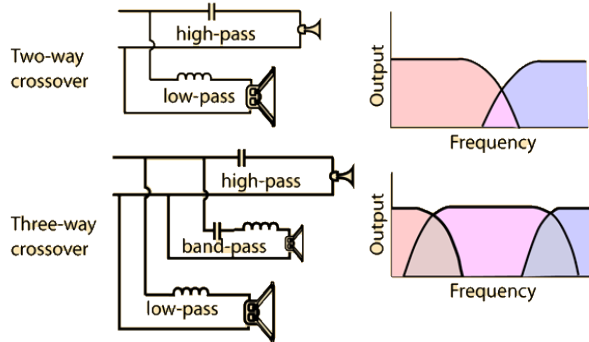
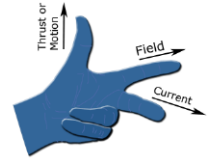


Rode NT1A

Condenser mic

Loudspeakers

- Convert electrical signals back to vibrations in air
- Similar to dynamic mic in principle: Fleming's left-hand rule
- Crossover networks: bass / midrange / tweeter

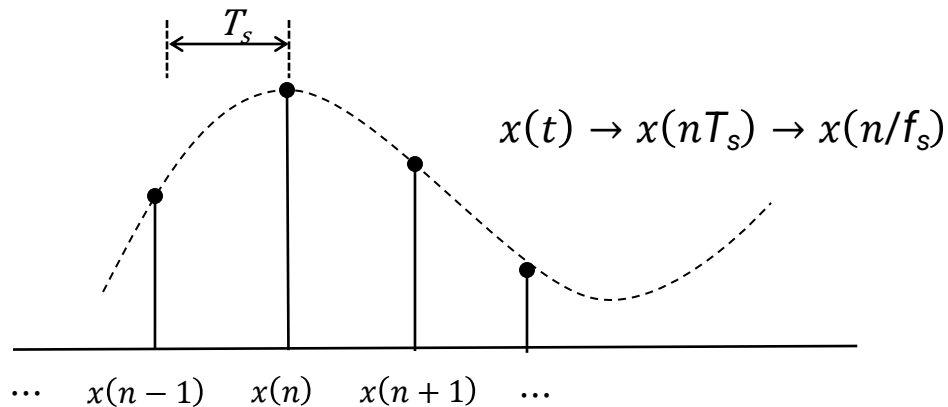


Genelec 8050B

2-way monitor speaker

Sampling

- Convert **continuous-time** signals to **discrete-time** signals by periodically picking up the instantaneous values
 - Represented as a sequence of numbers
 - Sampling period (T_s): the amount of time between samples
 - **Sampling rate** ($f_s = 1/T_s$): the number of samples per second



Sampling Theorem

- How can we determine the sampling rate?
 - Too high: increase the number of samples
 - Too low: cannot reconstruct the original signal
- Sampling Theorem
 - The sampling rate must be greater than twice the maximum frequency in the signal in order to reconstruct the original signal

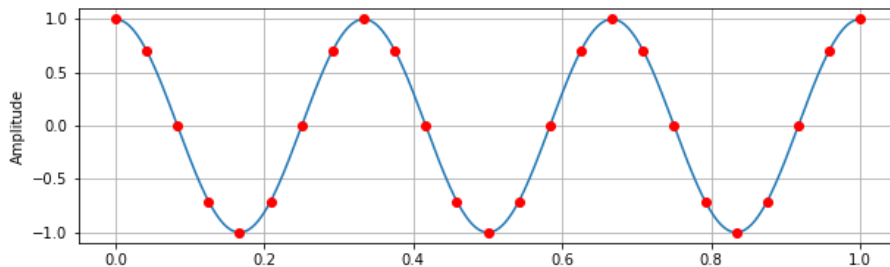
$$f_s > 2 \cdot f_M$$

f_s : sampling rate

f_M : maximum frequency of the signal

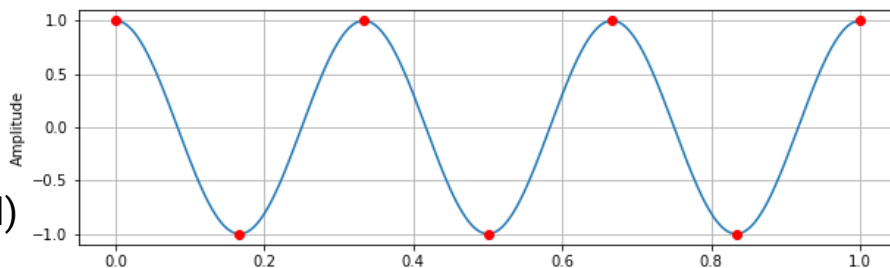
Sampling Theorem

$f_s > 2f_M$
(oversampled)



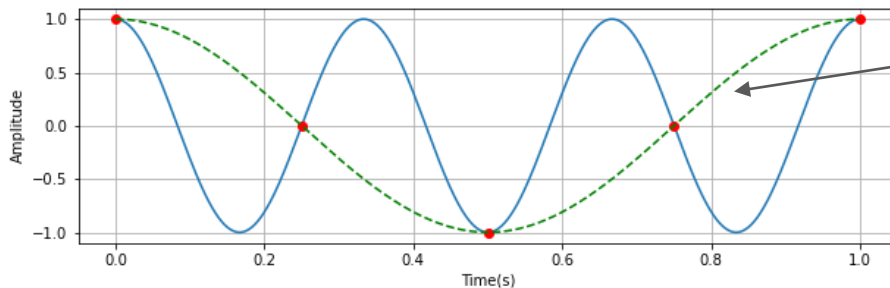
$f_s = 24 \text{ Hz}$ $f_M = 3 \text{ Hz}$

$f_s = 2f_M$
(critically sampled)



$f_s = 6 \text{ Hz}$ $f_M = 3 \text{ Hz}$

$f_s < 2f_M$
(undersampled)

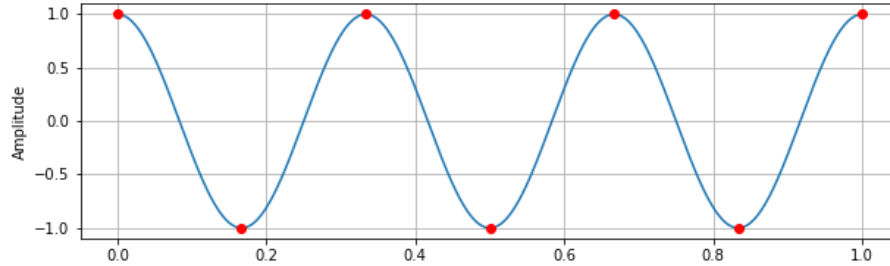


The reconstructed sinusoid
with 1Hz frequency

$f_s = 4 \text{ Hz}$ $f_M = 3 \text{ Hz}$

Aliasing

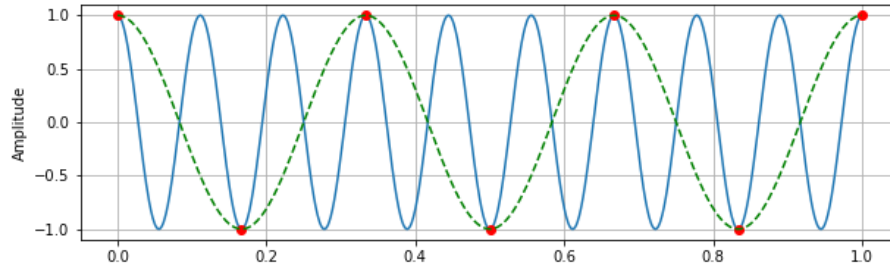
$$f = f_0$$



$$f_s = 6 \text{ Hz} \quad f = 3 \text{ Hz}$$

$$f = f_0 + f_s$$

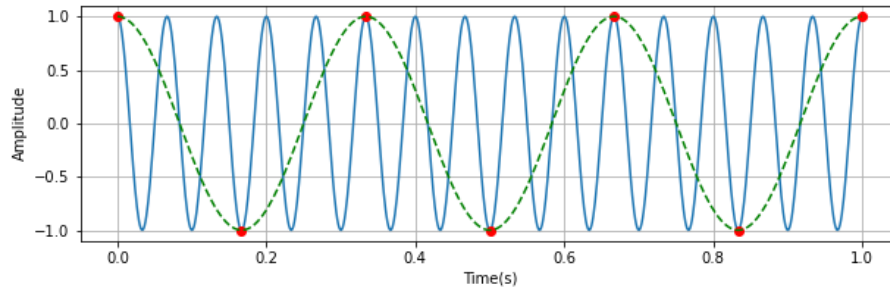
(Alias)



$$f_s = 6 \text{ Hz} \quad f = 9 \text{ Hz}$$

$$f = f_0 + 2f_s$$

(Alias)



$$f_s = 6 \text{ Hz} \quad f = 15 \text{ Hz}$$

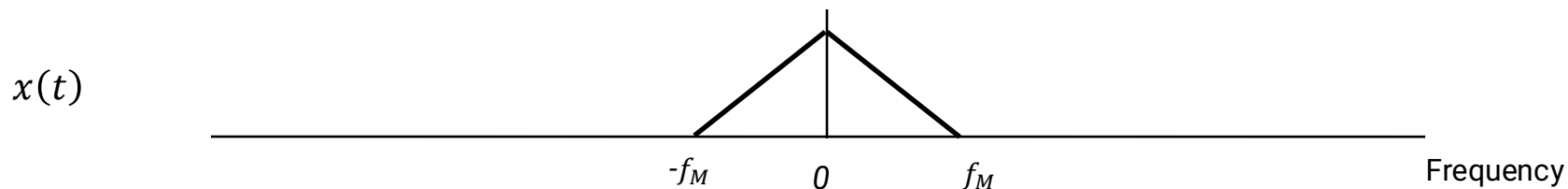
Aliasing

- A continuous-time sinusoid with frequency f_0 is sampled:

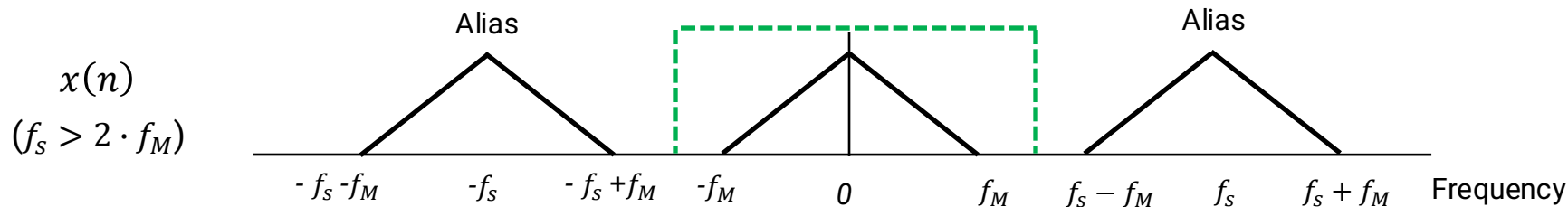
$$\begin{array}{lcl} x(t) = A \sin(2\pi f_0 t + \phi) & \xrightarrow{x(t) \rightarrow x(n/f_s)} & x(n) = A \sin(2\pi f_0 n / f_s + \phi) \\ x(t) = A \sin(2\pi(f_0 \pm l f_s)t + \phi) & \xrightarrow{\hspace{1.5cm}} & x(n) = A \sin(2\pi(f_0 \pm l f_s)n / f_s + \phi) \\ & (l = 1, 2, 3, \dots) & = A \sin(2\pi f_0 n / f_s \pm 2\pi l n + \phi) \\ & & = A \sin(2\pi f_0 n / f_s + \phi) \end{array}$$

- The sampled sinusoid with frequency f_0 has **aliases** at $f_0 \pm l f_s$

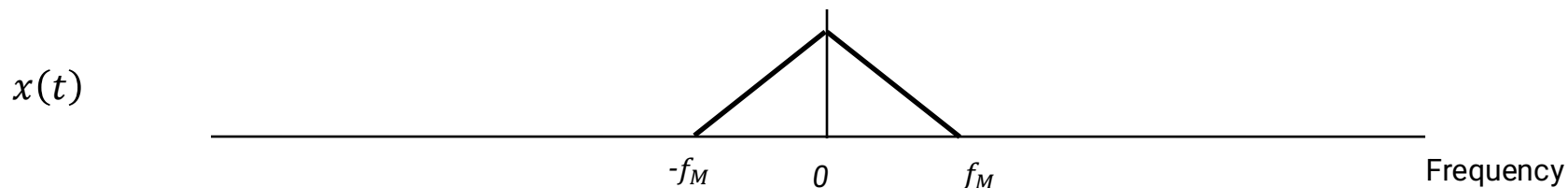
Sampling in the Frequency Domain



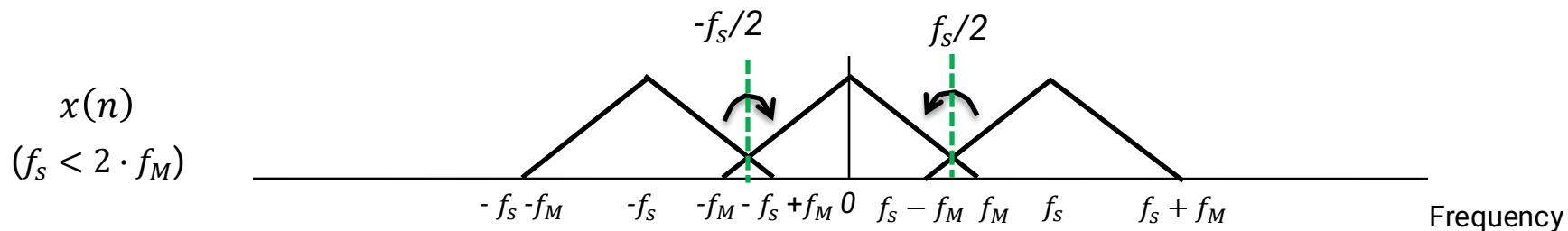
The original signal can be reconstructed using a lowpass filter



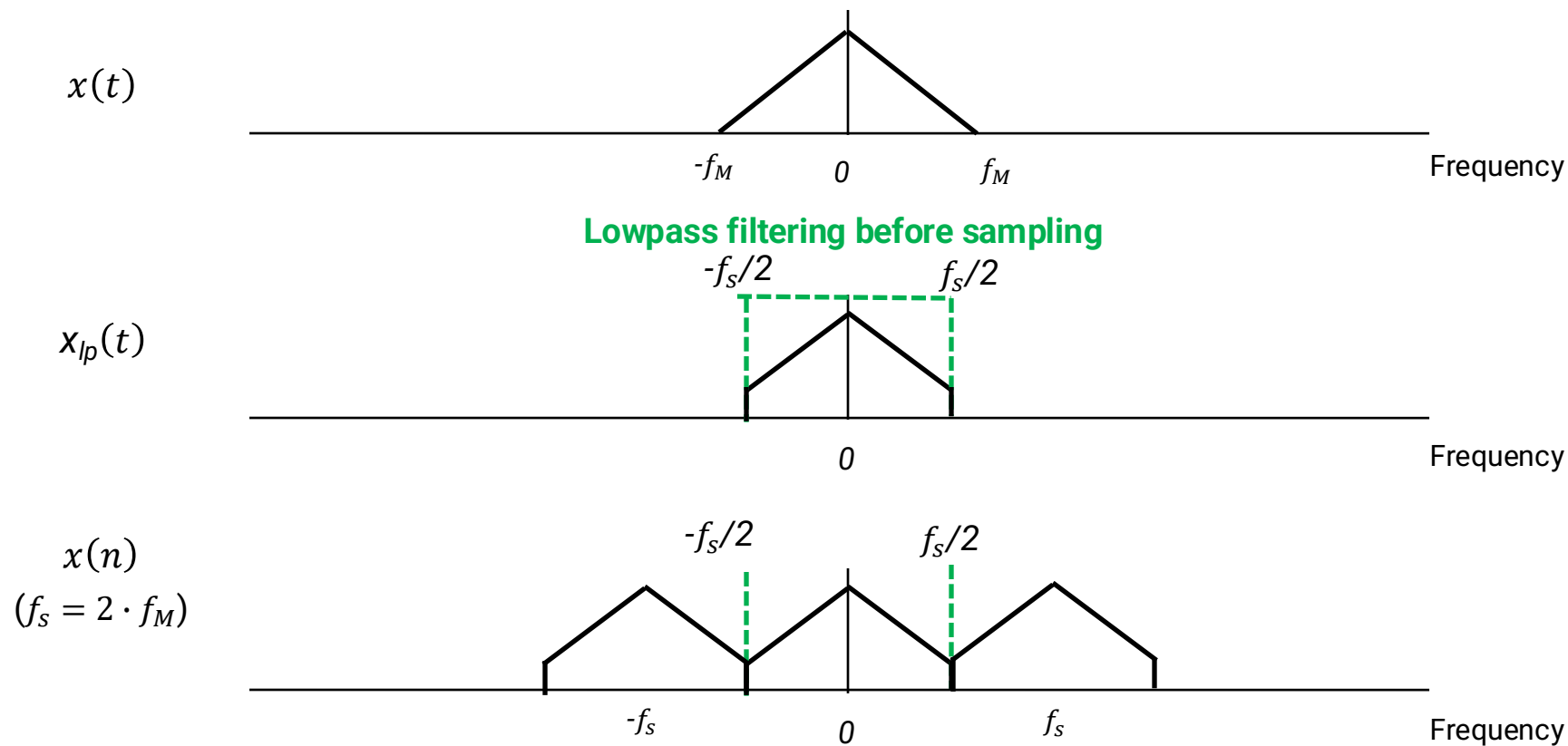
Sampling in the Frequency Domain: Aliasing



The high-frequency content above half the sampling rate (Nyquist rate) is folded over

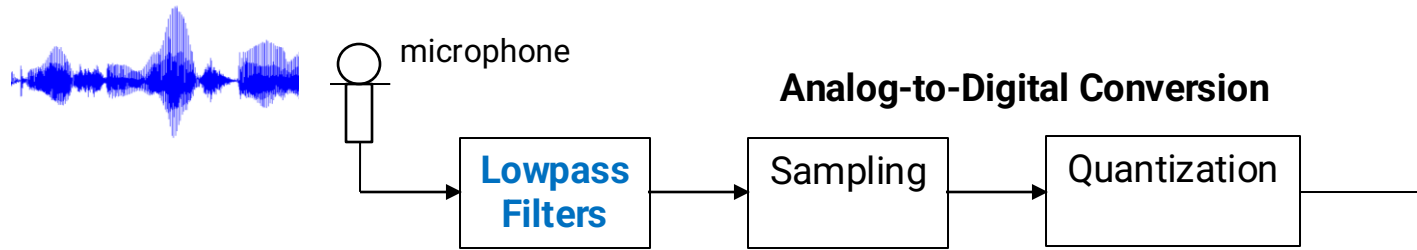


Sampling in the Frequency Domain: Lowpass filtering



Lowpass filter in Sampling

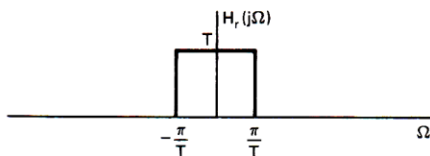
- The lowpass filter is implemented as an analog circuit when it is used before the audio-to-digital conversion in the digital audio system



Reconstruction of Sampled Audio

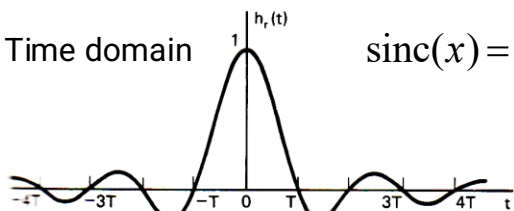
- The ideal low-pass corresponds to convolution with the sinc function in the time domain
 - In practice, DACs are composed of sample-and-hold and low-pass filtering circuits

Frequency domain



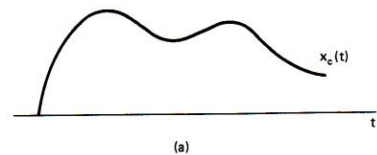
(b)

Time domain



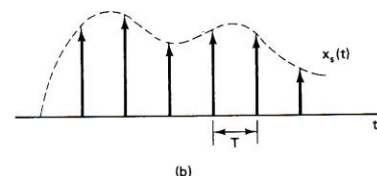
(c)

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$$



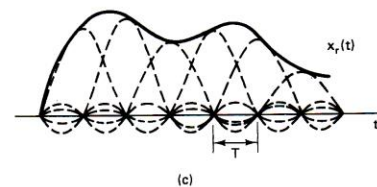
(a)

Before sampling



(b)

After sampling

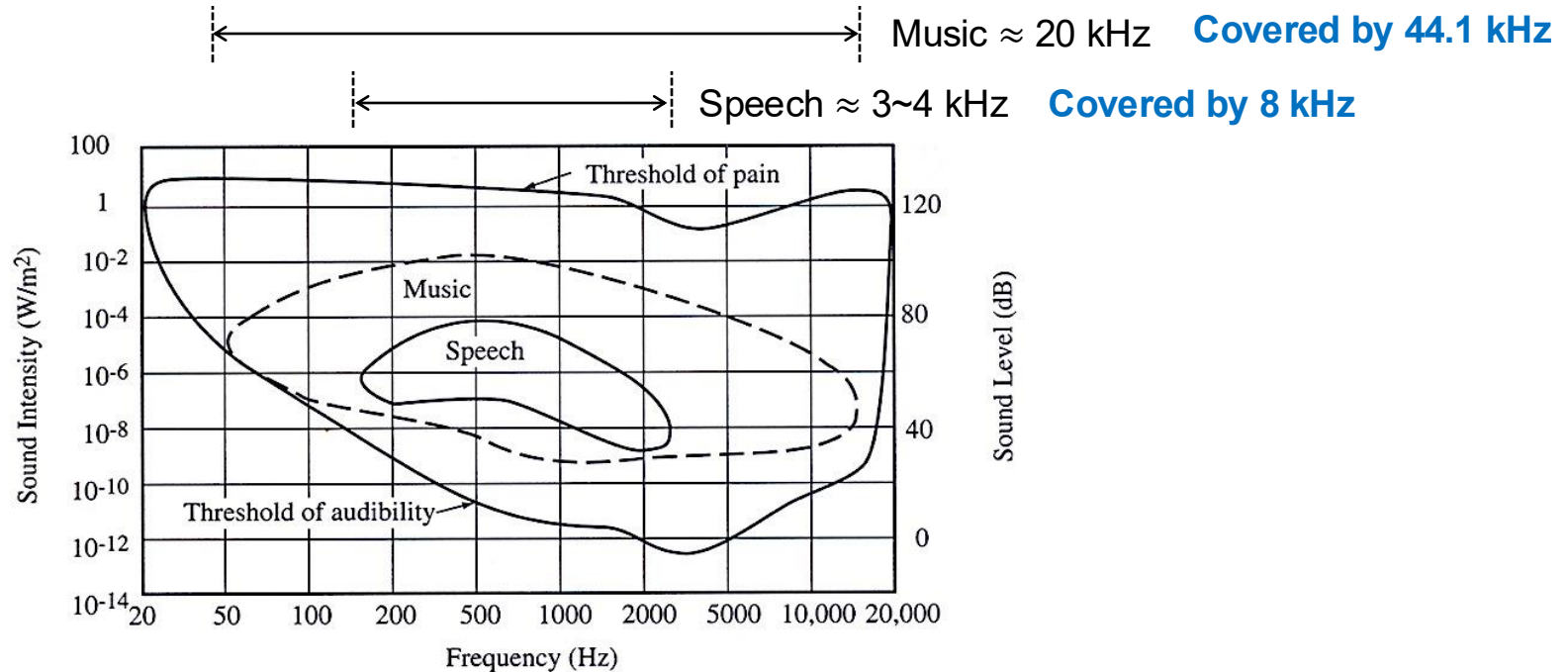


(c)

Reconstruction

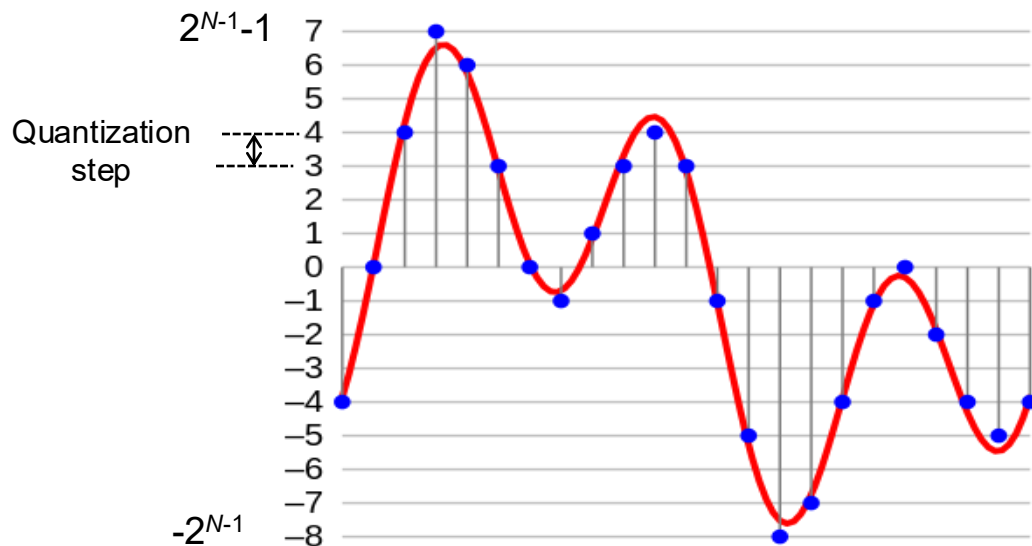
Sampling Rate in Audio Standard

- Determined by the maximum frequency of signals



Quantization

- Round the amplitude to the nearest discrete steps
 - The number of discrete steps are determined by the bit depth
 - N bits range from -2^{N-1} to $2^{N-1}-1$: 8 bit (-128 to 127), 16 bit (-32767 to 32766)

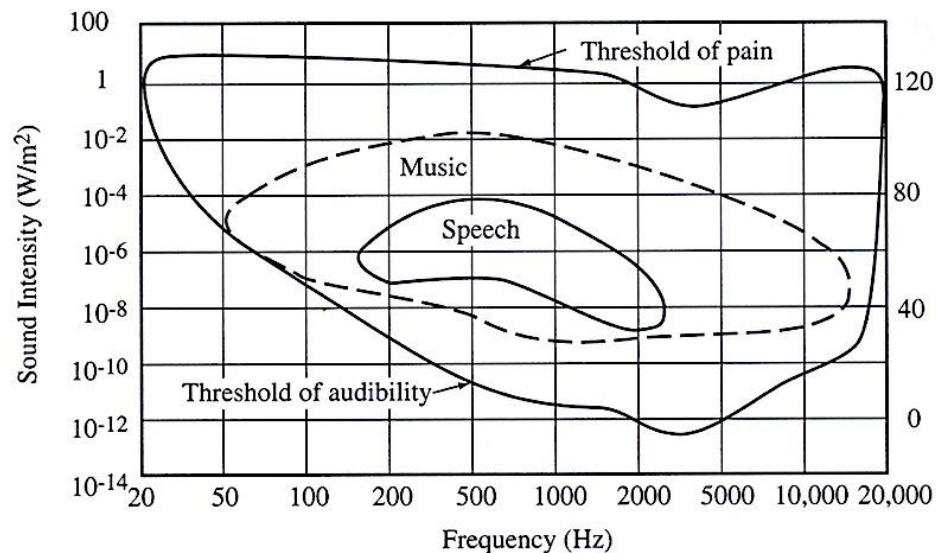


Bit Depth and Dynamic Range

- Bit depth determines dynamic range of digital signals
- Dynamic range = $20\log_{10}(\text{maximum value} / \text{minimum value})$
 - 8 bits = $20\log_{10}(127/1) \approx 48\text{dB}$
 - 16 bit = $20\log_{10}(32766/1) \approx 96\text{dB}$
 - Adding one bit (x2) increases 6dB : N bits $\approx 6N$ dB

Bit Depth in Audio Standard

- Determine bit depth to cover the dynamic range of audio content



Sound Level (dB)

Music \approx 80 dB

**Covered by
16 bits (96dB)**

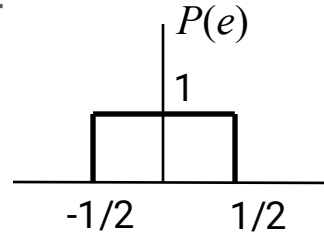
Speech \approx 48 dB

**Covered by
8 bits (48dB)**

Error in Uniform Quantization

- Quantization causes noise

- Average power of quantization noise: obtained from the probability density function (PDF) of the error



Root mean square (RMS) of noise

$$\sqrt{\int_{-1/2}^{1/2} x^2 p(e) dx} = \sqrt{1/12}$$

- Signal to Noise Ratio (SNR)

- Based on average power

$$20 \log_{10} \frac{S_{\text{rms}}}{N_{\text{rms}}} = 20 \log_{10} \frac{2^{B-1} / \sqrt{2}}{\sqrt{1/12}} = 6.02B + 1.76 \text{ dB}$$

(With 16bits, SNR = 98.08dB)

RMS of full-scale sine wave

- Based on the max levels $20 \log_{10} \frac{S_{\text{max}}}{N_{\text{max}}} = 20 \log_{10} \frac{2^{B-1}}{1/2} = 6.02B \text{ dB}$ (With 16bits, SNR = 96.32 dB)


Dynamic Range

- Dynamic range
 - The ratio between the loudest and softest levels

$$20 \log_{10} \frac{S_{\text{rms,max}}}{S_{\text{rms,min}}} = 20 \log_{10} \frac{2^{B-1} / \sqrt{2}}{1 / \sqrt{2}} = 6.02B - 6$$

(With 16bits, Dynamic Range = 90.31 dB)

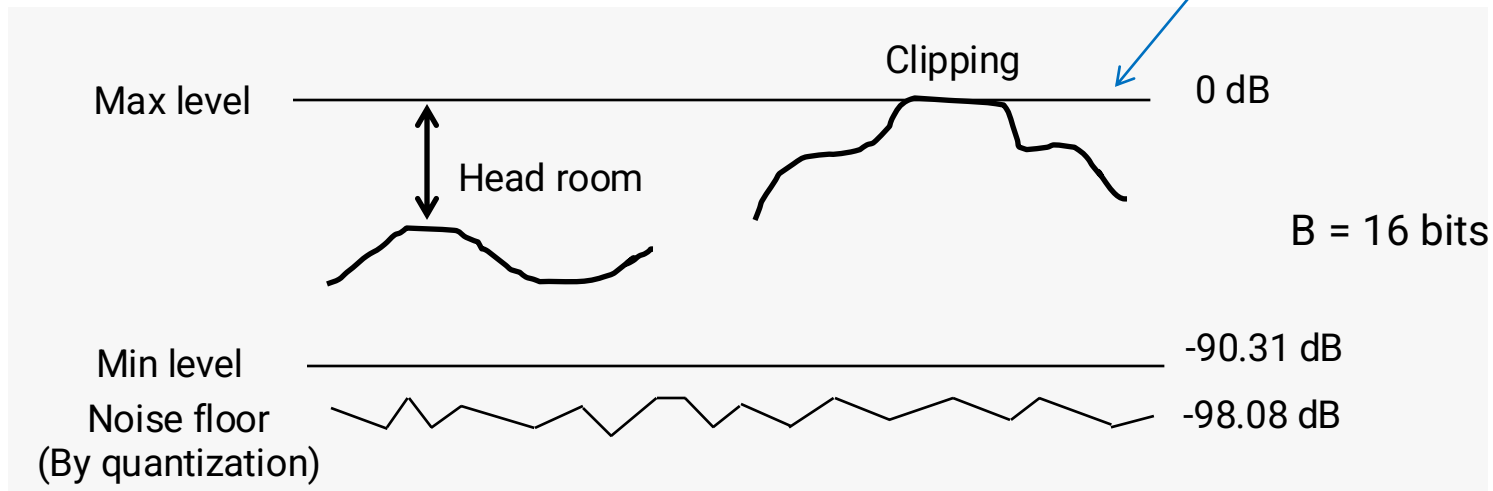
Again, RMS of full-scale sine wave
for both loudest and softest



Clipping and Headroom

- Clipping
 - Non-linear distortion that occurs when a signal is above the max level
- Headroom
 - Margin between the peak level and the max level

In digital audio, 0dB is regarded as the maximum level



Digital Audio Standards

- Compact disc
 - Sampling rate: 44.1 kHz: $> 2 \times 20$ kHz
 - Bit depth: 16 bits: SNR = 98.08dB
- Blu-ray disc / professional audio
 - Sampling rate: 48 / 96 / 192kHz: $> 2 \times 20$ kHz
 - Bit depth: 16 / 20 / 24 bits
- Telephone
 - Sampling rate: 8 / 16 kHz
 - Bit depth: 8 bits