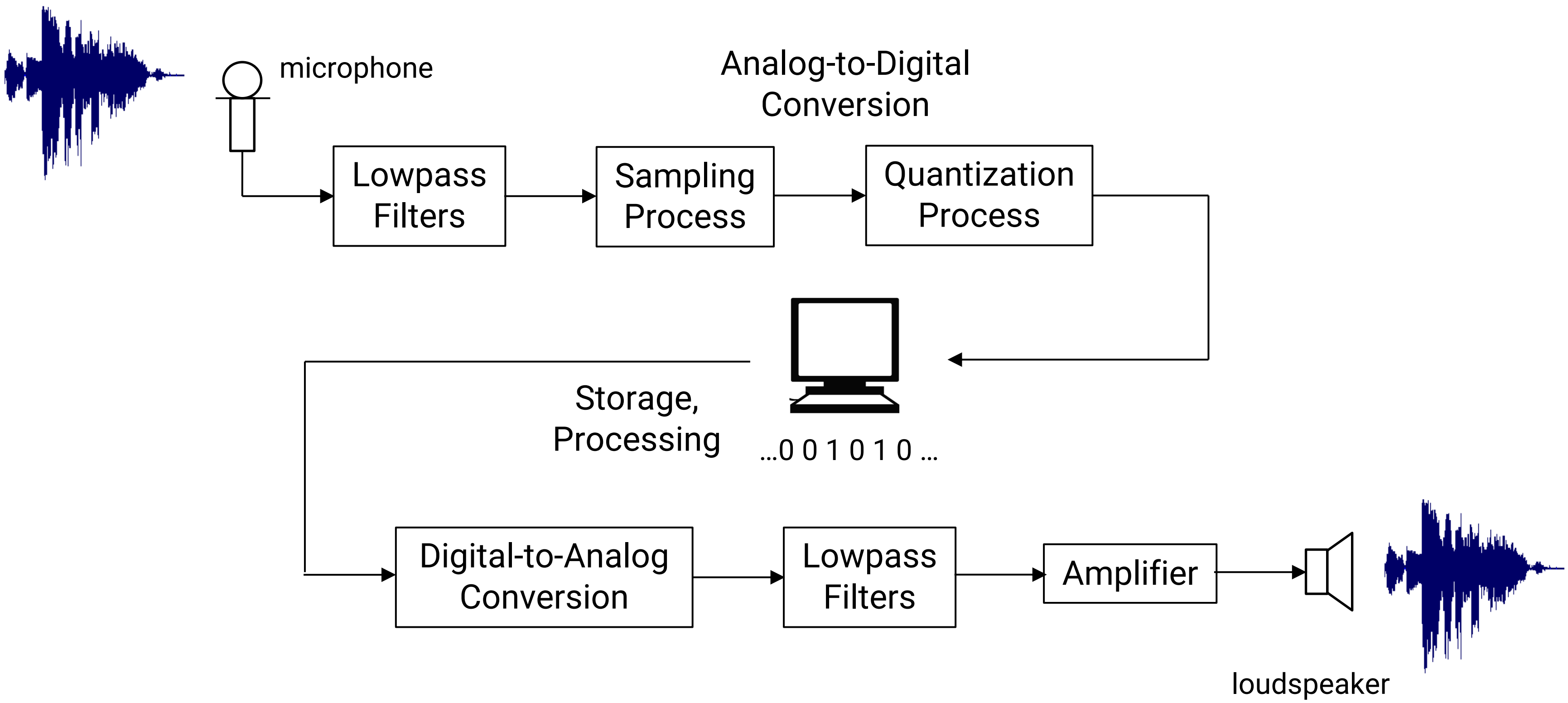
CTP431: Fundamentals of Computer Music

Digital Audio

KAIST Graduate School of Culture Technology

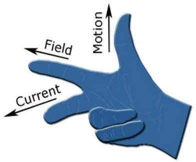
Juhan Nam

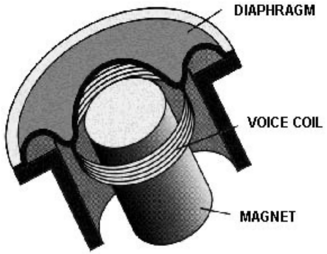
Digital Audio Chain



Microphones

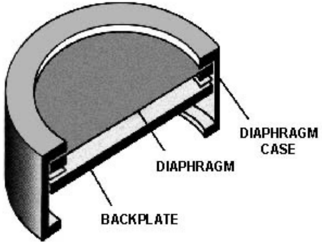
- Convert vibrations in air to electrical signals
- Dynamic mic
  - Fleming's right-hand rule
  - Vocal or drum in a live setting
- Condenser mic
  - Q = CV (capacitor)
  - Need phantom power
  - Studio recording
    - Sensitive and accurate: capture delicate sounds (quite or high-frequency sound)
  - Expensive





Shure SM58

Dynamic mic



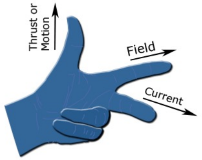
Rode NT1A

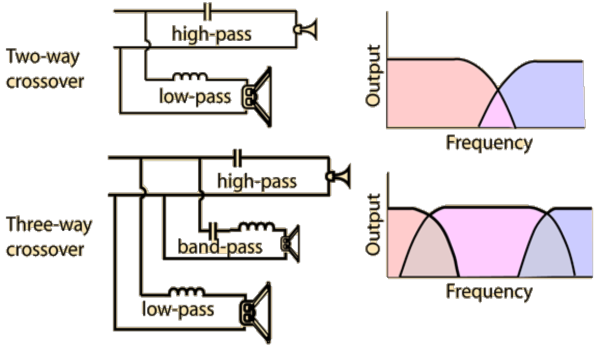
Condenser mic

Source: http://www.shure.com/americas/support/find-an-answer/difference-between-a-dynamic-and-condenser-microphone
Source: http://en.rode.com/microphones/nt1-a
Source: https://www.shure.com/en-US/products/microphones/sm58

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





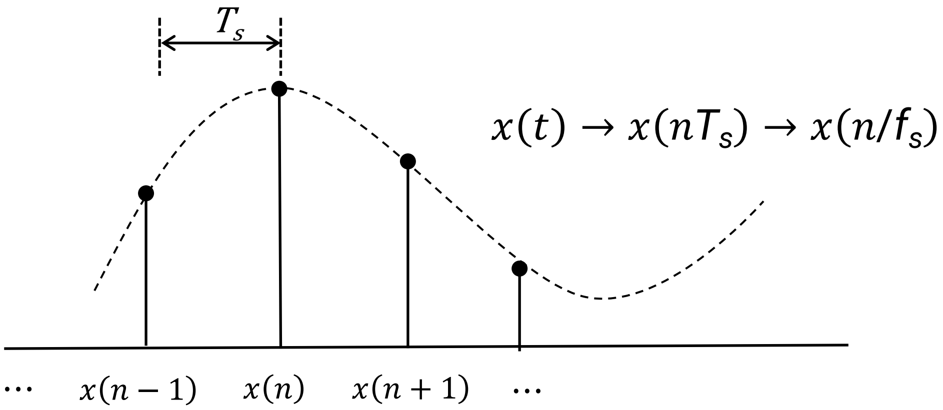
Genelec 8050B

2-way monitor speaker

Source: http://hyperphysics.phy-astr.gsu.edu/hbase/Audio/cross.html
Source: https://www.genelec.com/active-2-way-studio-monitors

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

$$\boxed{f_s > 2 \cdot f_M}$$

f_s: sampling rate

f_M: maximum frequency of the signal

Sampling Theorem

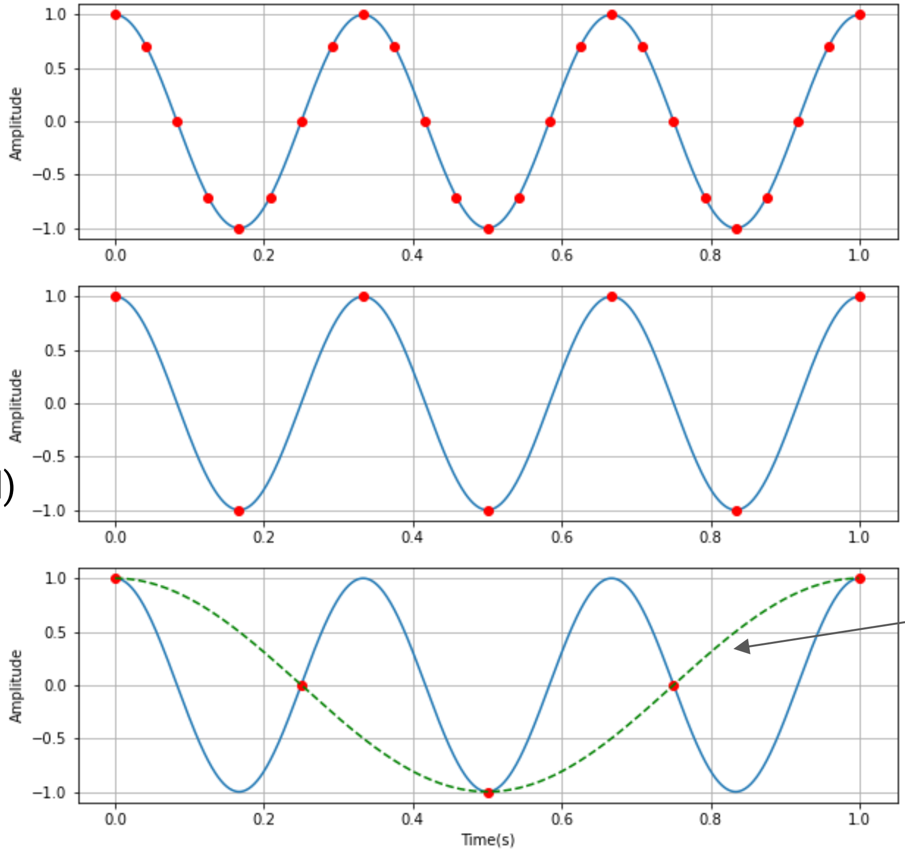
$f_s > 2f_M$
(oversampled)

$f_s = 24\ Hz \quad f_M = 3\ Hz$

$f_s = 2f_M$
(critically sampled)

$f_s = 6\ Hz \quad f_M = 3\ Hz$

$f_s < 2f_M$
(undersamped)



The reconstructed sinusoid with 1Hz frequency

$f_s = 4\ Hz \quad f_M = 3\ Hz$

Aliasing

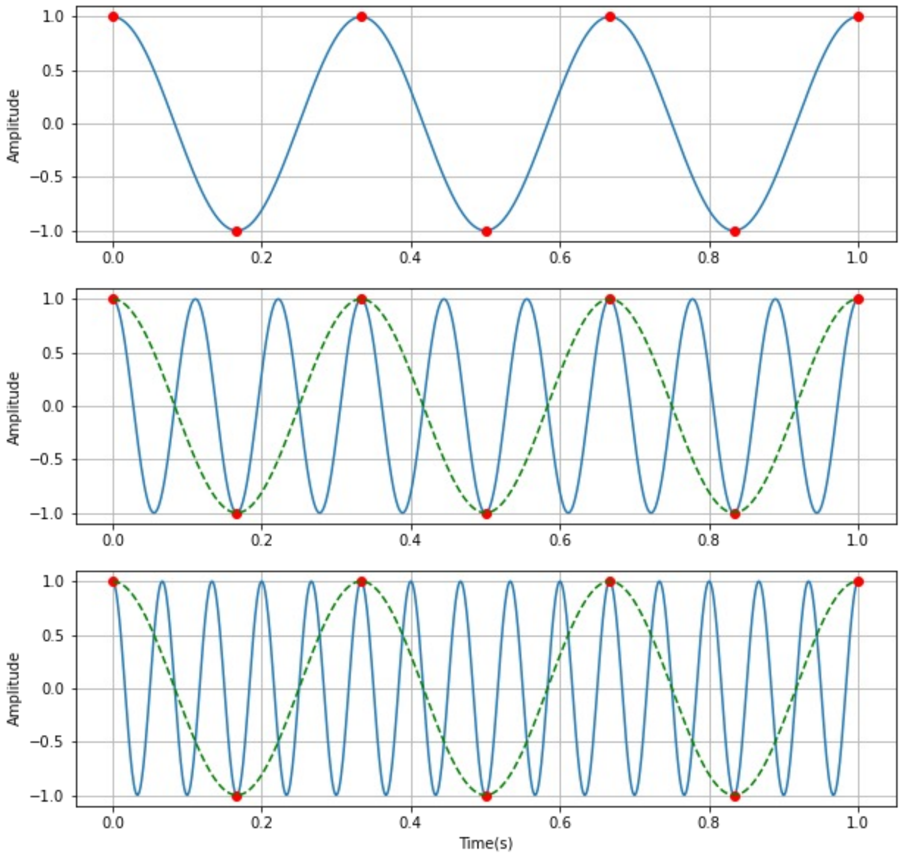
$f = f_0$

$f = f_0 + f_s$

(Alias)

$f = f_0 + 2f_s$

(Alias)



$f_s = 6\ Hz \quad f = 3\ Hz$

$f_s = 6\ Hz \quad f = 9\ Hz$

$f_s = 6\ Hz \quad f = 12\ Hz$

Aliasing

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

$$x(t) \rightarrow x(n/f_s)$$

$$x(t) = A\sin(2\pi f_0 t + \phi) \longrightarrow \boxed{x(n) = A\sin(2\pi f_0 n/f_s + \phi)}$$

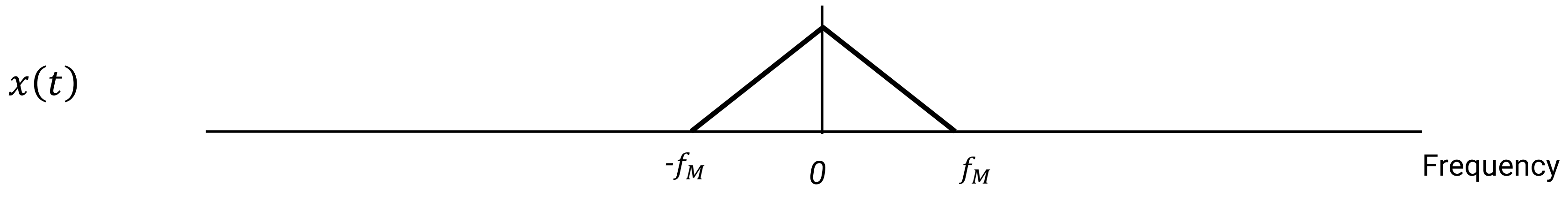
$$x(t) = A\sin(2\pi(f_0 \pm lf_s)t + \phi) \longrightarrow \boxed{x(n)} = A\sin(2\pi(f_0 \pm lf_s)n/f_s + \phi)$$

$$(l = 1, 2, 3, \ldots) \qquad = A\sin(2\pi f_0 n/f_s \pm 2\pi l n + \phi)$$

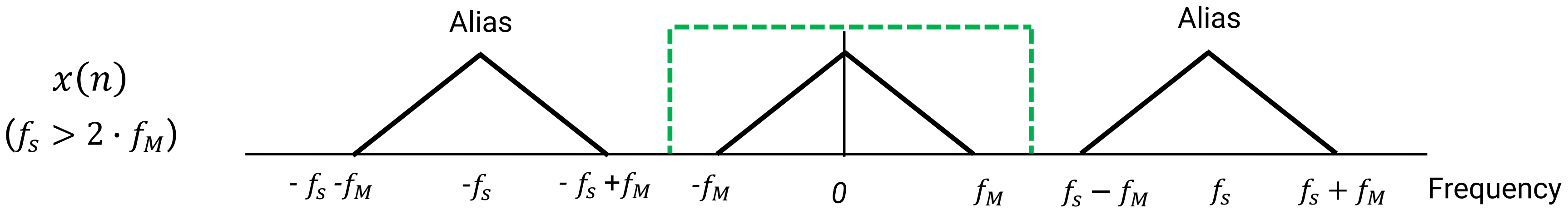
$$\boxed{= A\sin(2\pi f_0 n/f_s + \phi)}$$

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

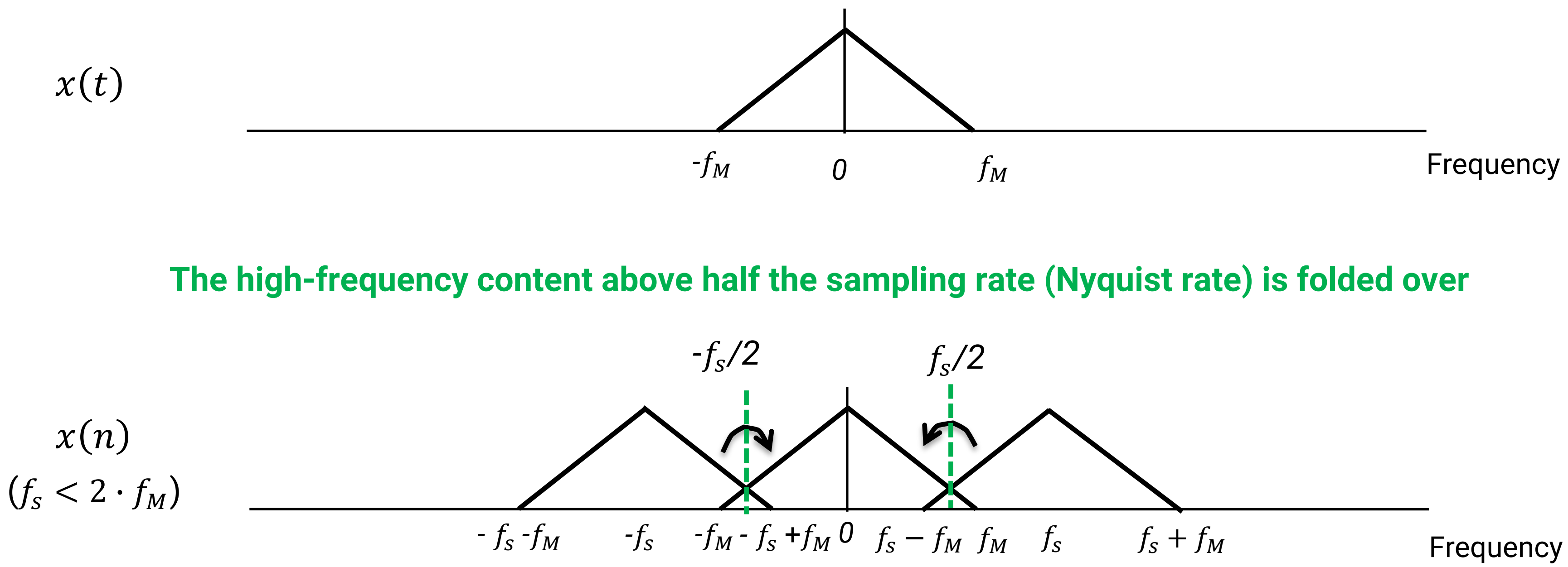
Sampling in the Frequency Domain



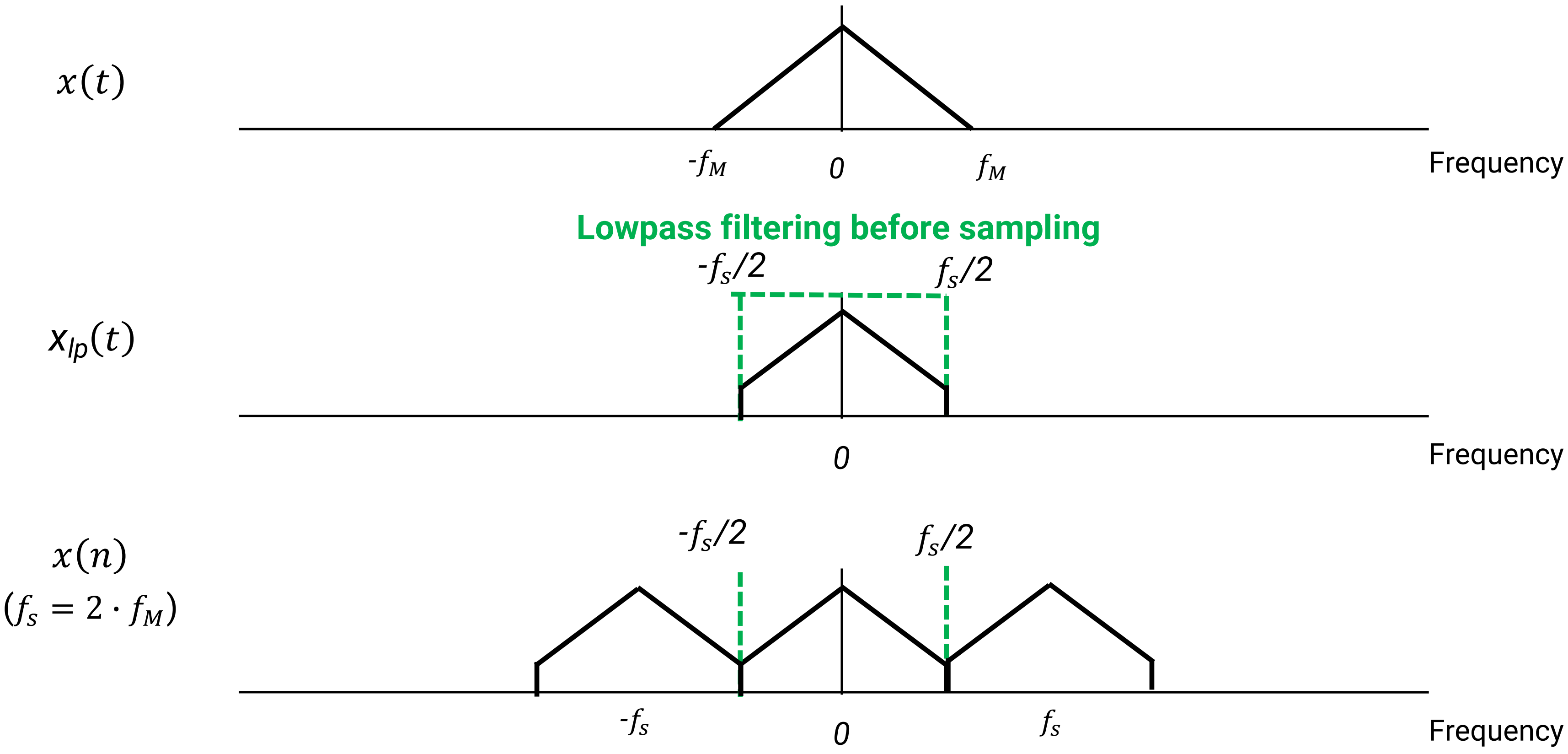
The original signal can be reconstructed using a lowpass filter



Sampling in the Frequency Domain: Aliasing

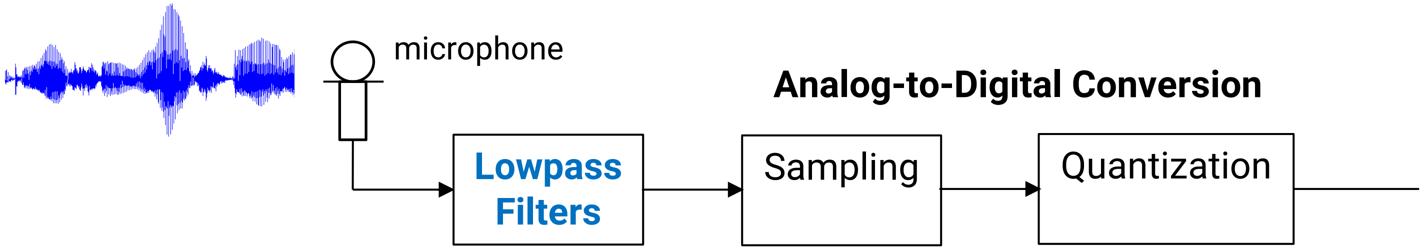


Sampling in the Frequency Domain: Lowpass filtering



Lowpass filter in Sampling

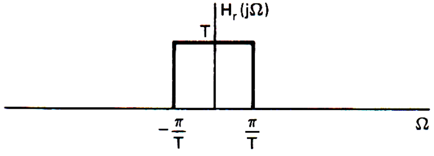
- The lowpass filter is implemented as an analog circuit when it is used before the audio-to-digital conversion in the digital audio system



Reconstruction of Sampled Audio

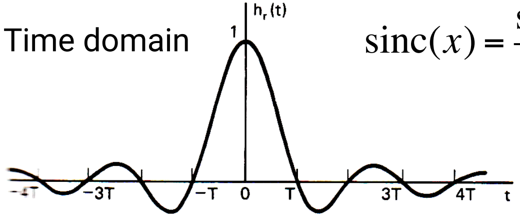
- The ideal low-pass corresponds to convolution with the sinc function in the time domain
  - In practice, DACs are composed of sample-and-hold and low-pass filtering circuits

Frequency domain



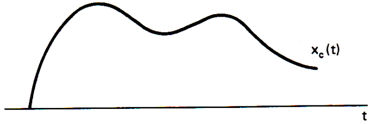
(b)

Time domain



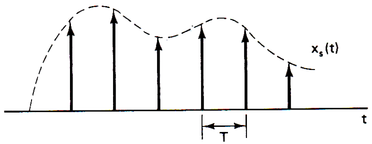
(c)

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



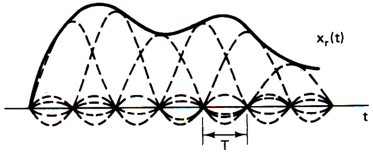
(a)

Before sampling



(b)

After sampling



(c)

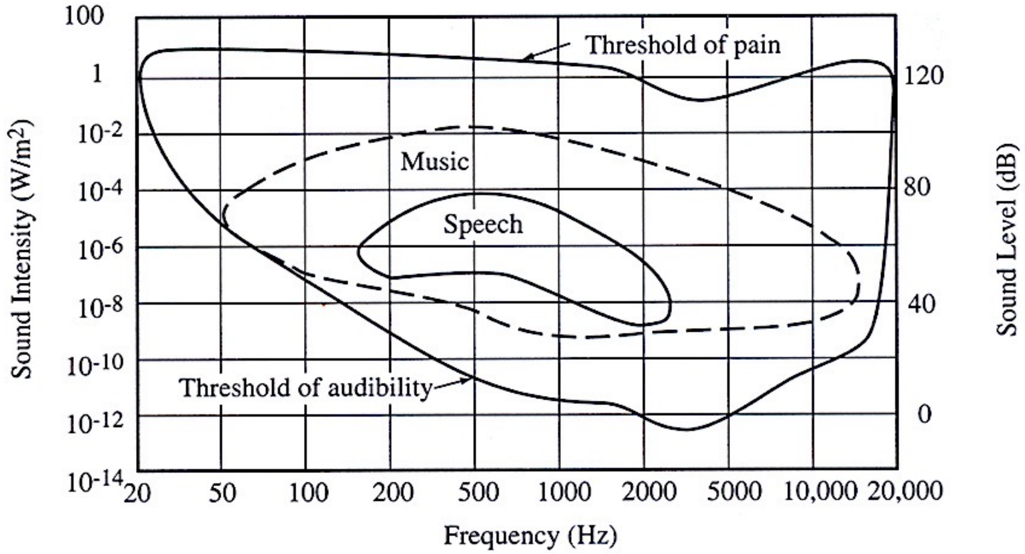
Reconstruction

Sampling Rate in Audio Standard

- Determined by the maximum frequency of signals

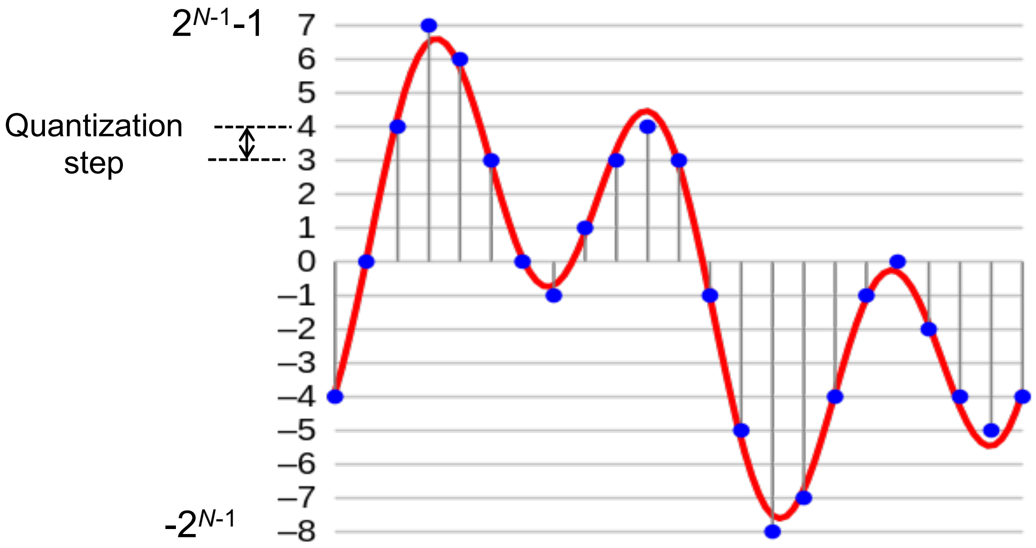
Music ≈ 20 kHz **Covered by 44.1 kHz**

Speech ≈ 3~4 kHz **Covered by 8 kHz**



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)

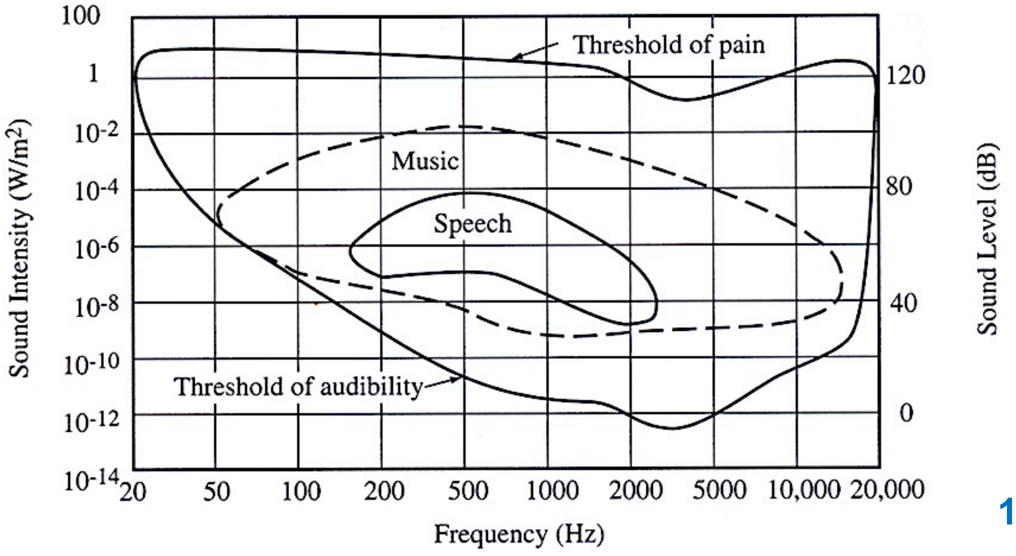


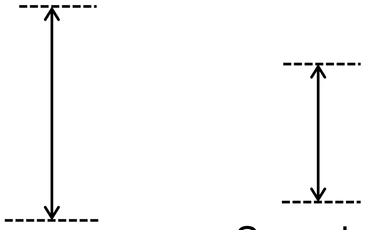
Bit Depth and Dynamic Range

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





Music ≈ 80 dB

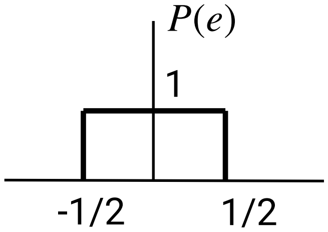
Covered by 16 bits (96dB)

Speech ≈ 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

RMS of full-scale sine wave

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

(With 16bits, SNR = 98.08dB)

  - Based on the max levels

$$20\log_{10}\frac{S_{\max}}{N_{\max}} = 20\log_{10}\frac{2^{B-1}}{1/2} = 6.02B\ \mathrm{dB}$$

(With 16bits, SNR = 96.32 dB)

Dynamic Range

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

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

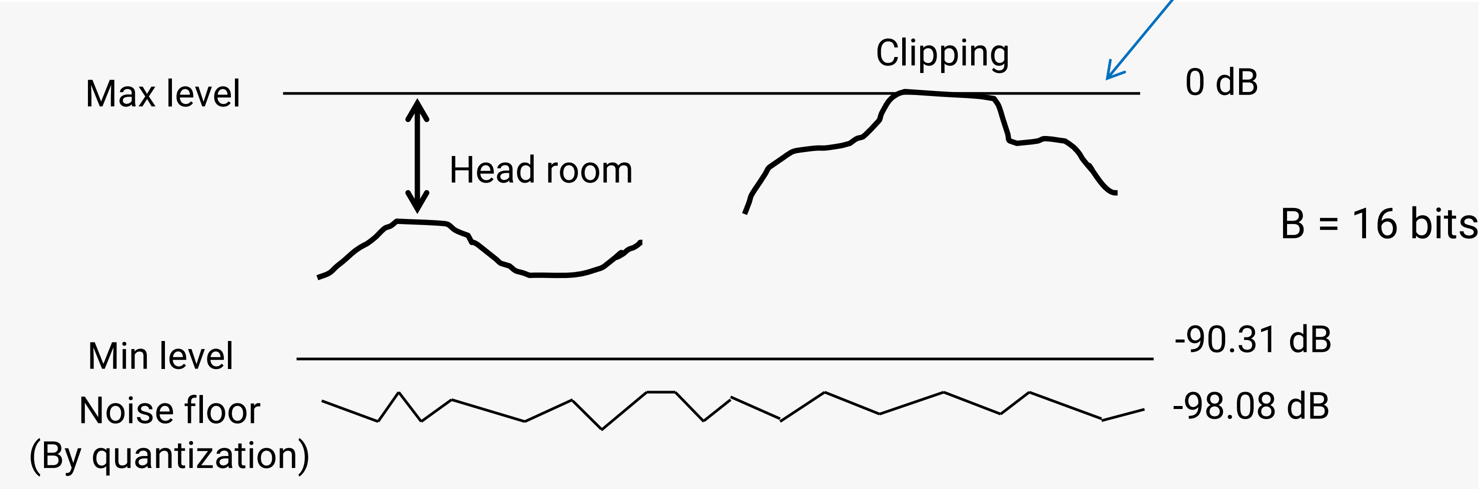
(With 16bits, Dynamic Range = 90.31 dB)

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