Audio: Rendering

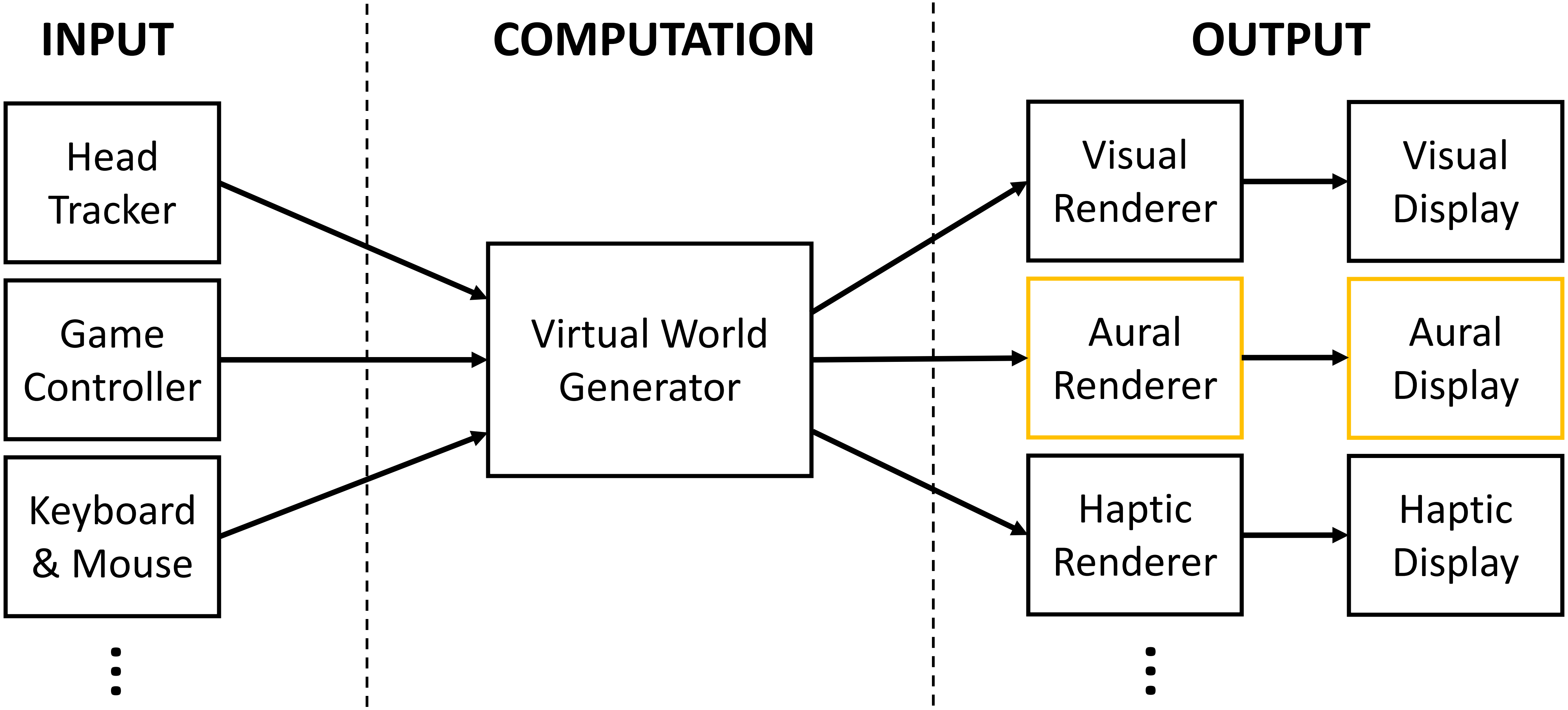
CS 6334 Virtual Reality

Professor Yapeng Tian

The University of Texas at Dallas

A lot of slides of course lectures borrowed from Professor Yu Xiang's VR class

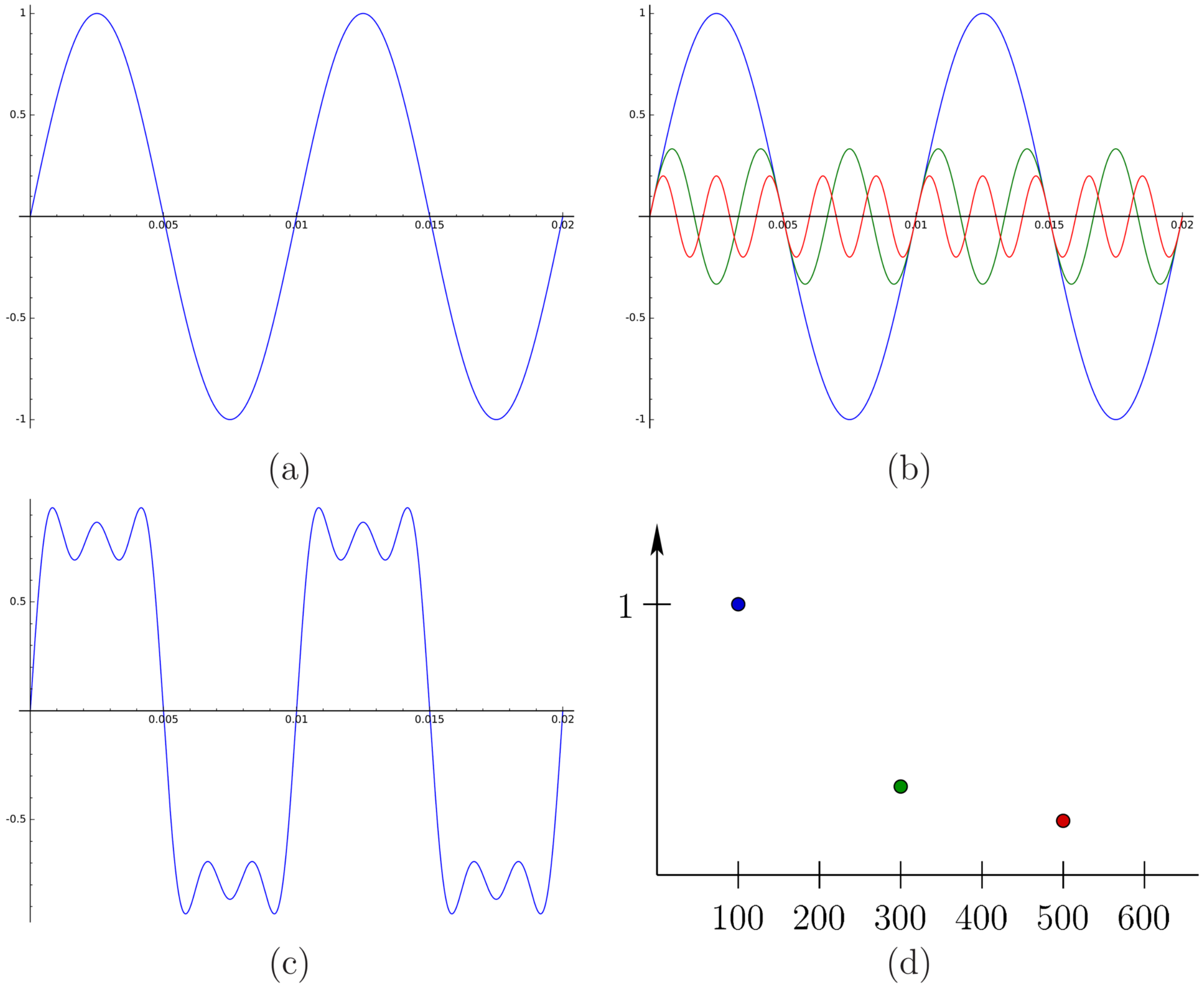
Review of VR Systems



Auditory Rendering

- Producing sounds for the virtual world
- Aural displays: speakers
- The generated sounds should be consistent with visual cues and with past auditory experiences in the real world

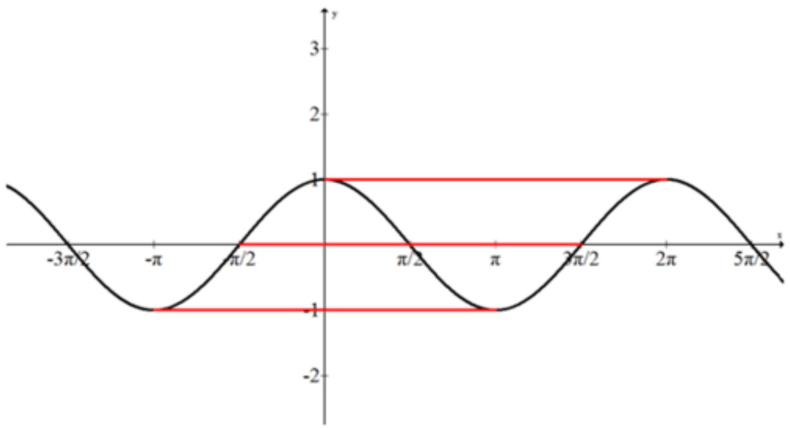
Spectral Decomposition



Fourier analysis: any periodic function can be decomposed into sinusoids

Frequency of Sinusoidal Functions

- Period: length of a complete cycle



- Frequency: number of cycles in 1 unit

$$period = \frac{1}{frequency}$$

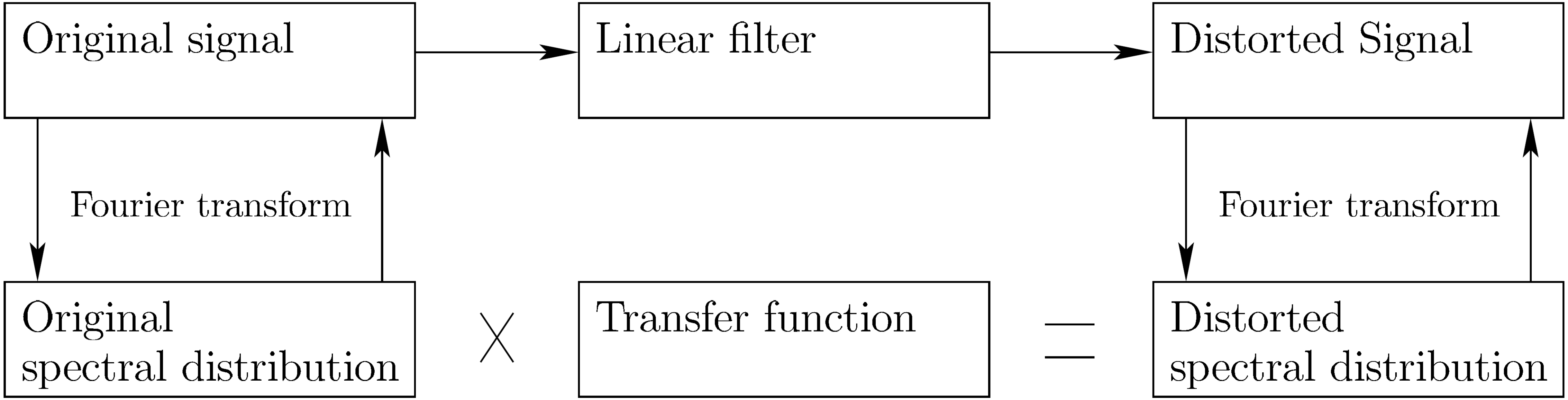
$$f(x) = \sin x$$

$$f(x) = \sin \frac{1}{2} x$$

$$f(x) = \sin bx$$

$$f(x) = \sin 2\pi f x$$

Signal Processing



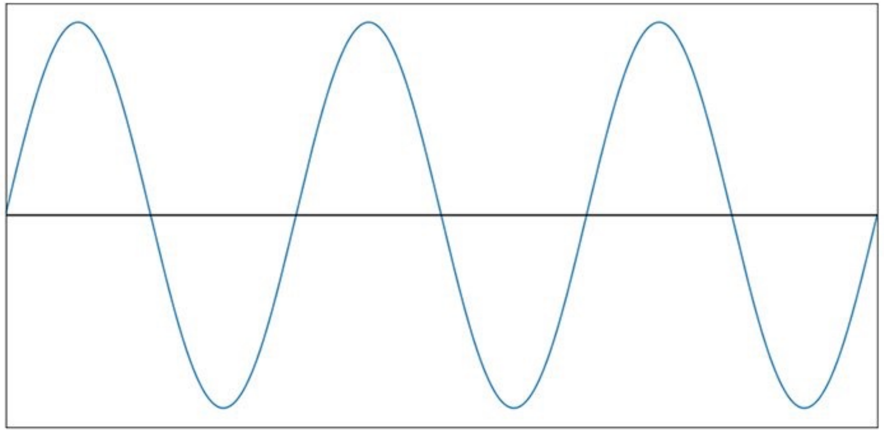
Sampling Rate

- Continuous-time signal $x(t)$
- Discrete-time signal, how computers process signals
- Sampling interval Δt
- Sampling rate (sampling frequency) Hz $1/\Delta t$
  - 1000Hz sampling rate, Δt is 1ms
  - How many samples per second

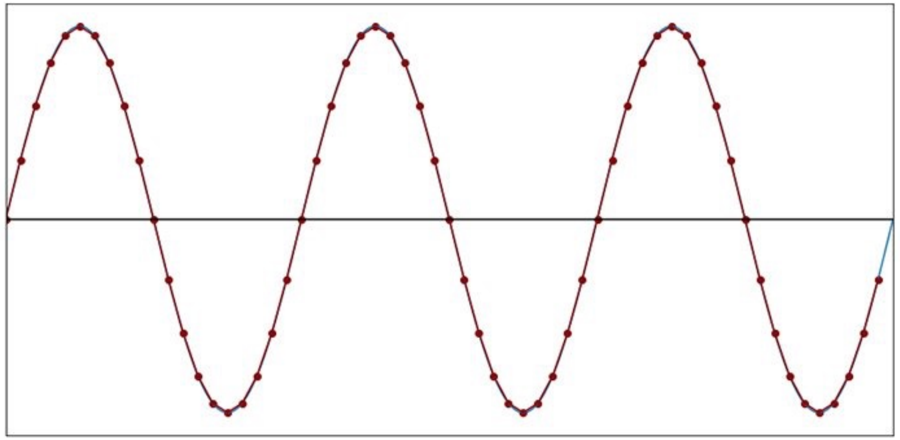
Nyquist–Shannon Sampling Theorem

- The sampling rate should be at least **two times** the highest frequency component in the signal
- The highest frequency for audio is 20,000 Hz, sampling rate at least 40,000 Hz
- Sampling rate of CDs and DVDs: 44,100 Hz and 48,000 Hz
- kth sample $x[k] = x(k\Delta t)$

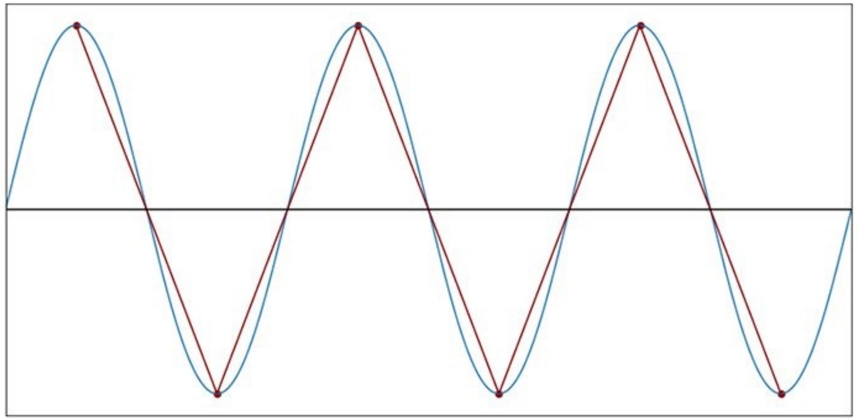
Nyquist–Shannon Sampling Theorem



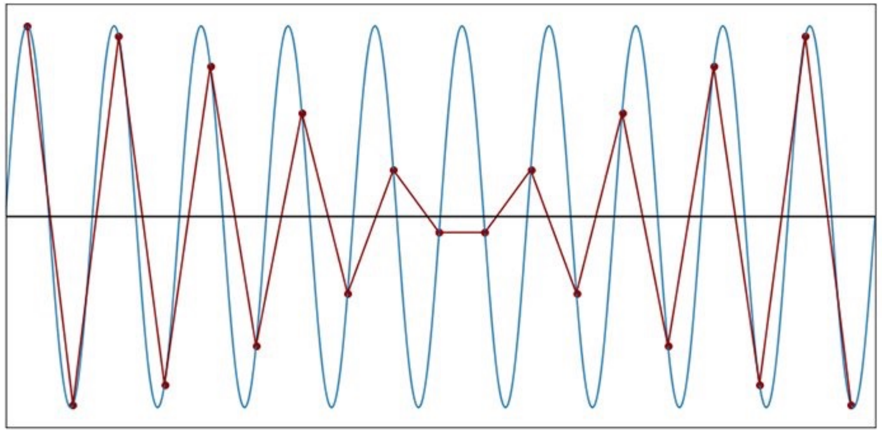
Continuous-time signal



20 samples per cycle



2 samples per cycle

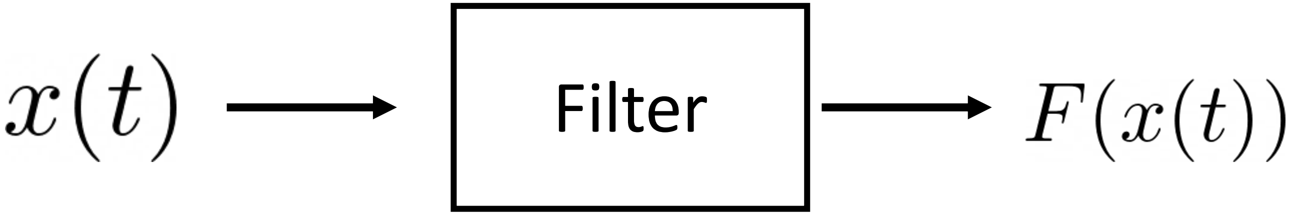


1.9 samples per cycle

https://www.allaboutcircuits.com/technical-articles/nyquist-shannon-theorem-understanding-sampled-systems/

Linear Filters

- A filter is a transformation that maps one signal to another



- Linear filters
  - Additivity $F(x + x') = F(x) + F(x')$
  - Homogeneity $cF(x) = F(cx)$
  - A general form

$$y[k] = c_0 x[k] + c_1 x[k-1] + c_2 x[k-2] + c_3 x[k-3] + \cdots + c_n x[k-n]$$

Examples of Linear Filters

- Moving average

$$y[k] = \frac{1}{3}x[k] + \frac{1}{3}x[k-1] + \frac{1}{3}x[k-2]$$

- Exponential Smoothing (exponentially weighted moving average)

$$y[k] = \frac{1}{2}x[k] + \frac{1}{4}x[k-1] + \frac{1}{8}x[k-2] + \frac{1}{16}x[k-3]$$

Nonlinear Filters

- Any filter that does not follow the following form

$$y[k] = c_0 x[k] + c_1 x[k-1] + c_2 x[k-2] + c_3 x[k-3] + \cdots + c_n x[k-n]$$

- Human auditory system is almost a linear filter, but contains nonlinear behaviors

Fourier Analysis

- Fourier transform for discrete-time systems

$$X(f) = \sum_{k=-\infty}^{\infty} x[k] e^{-i2\pi f k}$$

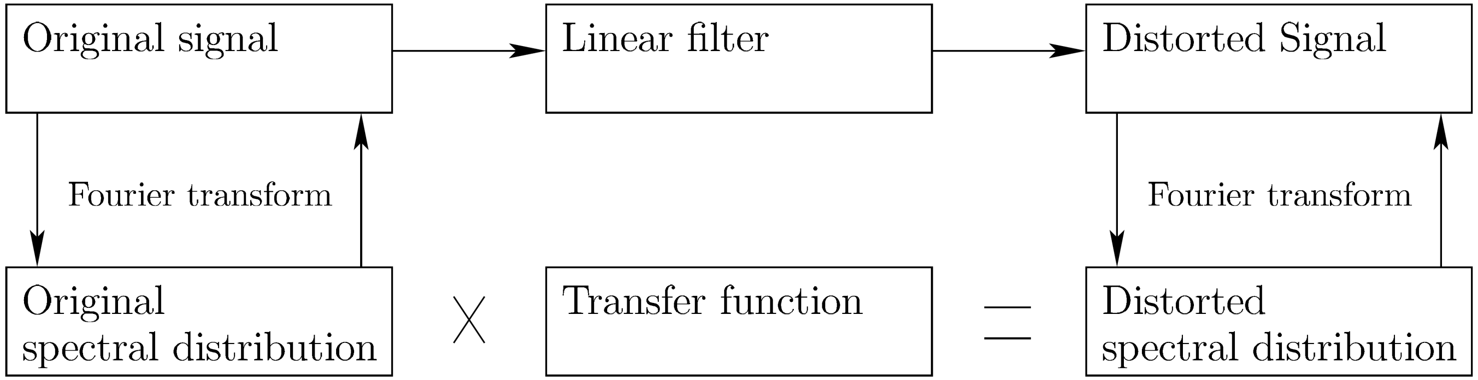
frequency

Spectral distribution: a function of the frequency

Euler's formula

$$e^{-i2\pi f k} = \cos(-2\pi f k) + i \sin(-2\pi f k) \qquad i = \sqrt{-1}$$

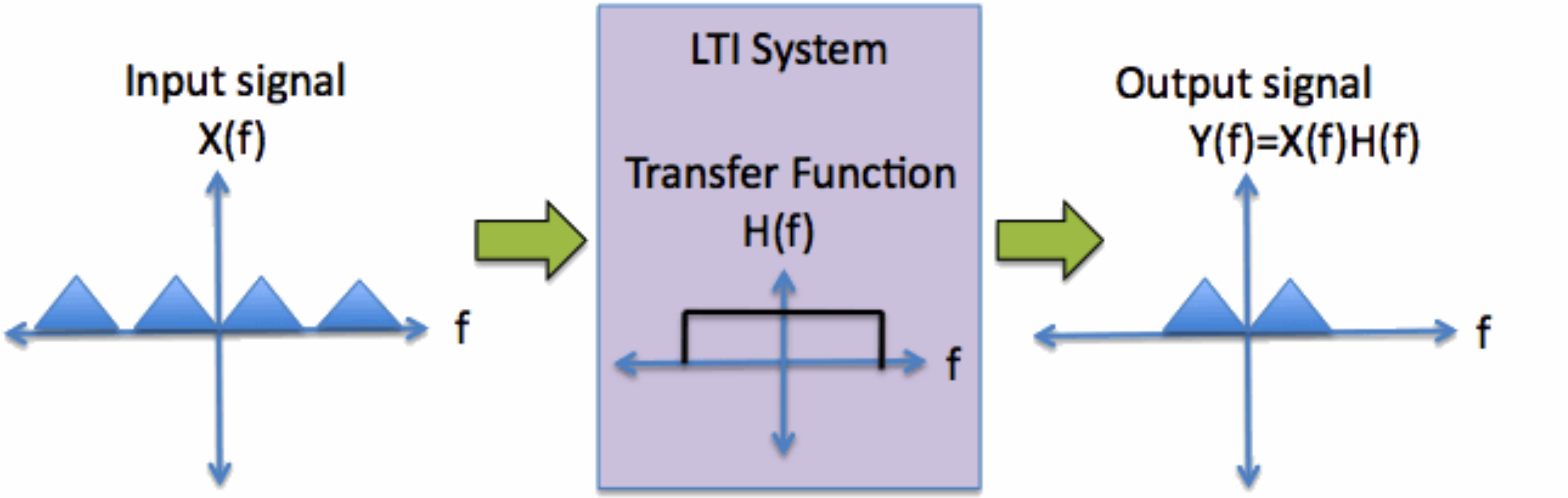
Transfer Function



- A linear filter can be designed to modify the spectral distribution
  - Amplify some frequencies, while suppressing others

- Applying a transfer function
  - Transforming the original signal using the Fourier transform
  - Multiplying the transfer function
  - Applying the inverse Fourier transform

Transfer Function



$$X(f) = \Im\{x(t)\} \qquad H(f) = \Im\{h(t)\}$$

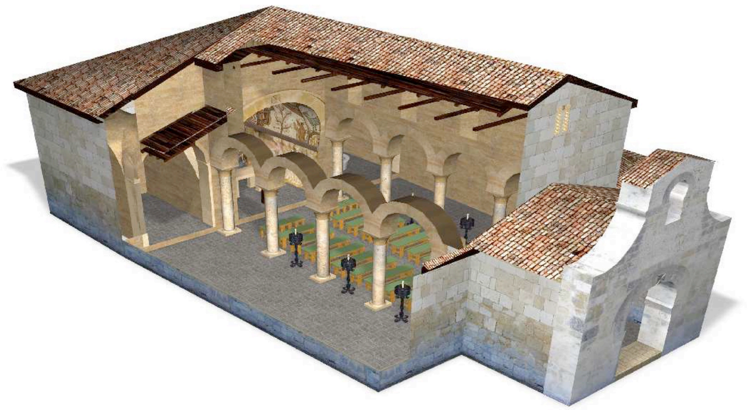
$$Y(f) = X(f)H(f)$$

$$y(t) = \Im^{-1}\{Y(f)\}$$

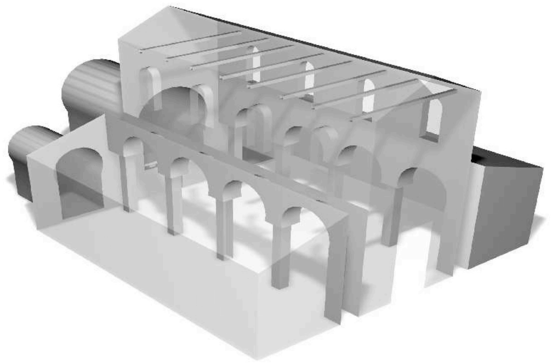
www.thefouriertransform.com

Acoustic Modeling

- The same geometry model can be used for both visual modeling and auditory modeling
  - E.g., walls can reflect lights and sound waves



(a)



(b)

The acoustic model needs to have a spatial resolution of only 0.5m

Figure 11.13: An audio model is much simpler. (From Pelzer, Aspock, Schroder, and Vorlander, 2014, [253])

Acoustic Modeling

- Sound source in the virtual environment
  - White noise, equal weight of all frequencies in the audible spectrum
  - Interesting sounds, high concentration among specific frequencies

- Sound reflection (depends on wavelength)
  - Specular reflection for a large, smooth, flat surface
  - Diffuse reflection for smaller objects, surface with repeated structures (difficult to characterize for sounds)

Propagation of Sounds

- Method 1: simulating the physics as accurately as possible
  - When waves are large relative to objects in the environment
  - Low frequency, detailed environment
- Method 2: Shooting visibility rays and characterize the dominant interactions between sound sources, surfaces, and ears
  - Higher frequency, simpler model

Numerical Wave Propagation

- Helmholtz wave equation
  - Constraints at every point in R^3 in terms of partial derivatives of the pressure function

$$\nabla^2 p + \frac{\omega^2}{s^2} p = 0 \qquad \omega = 2\pi f$$

Laplacian operator sound pressure sound speed

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

Visibility-based Wave Propagation

- Paths of sound rays that emanate from the source and bounce between obstacles

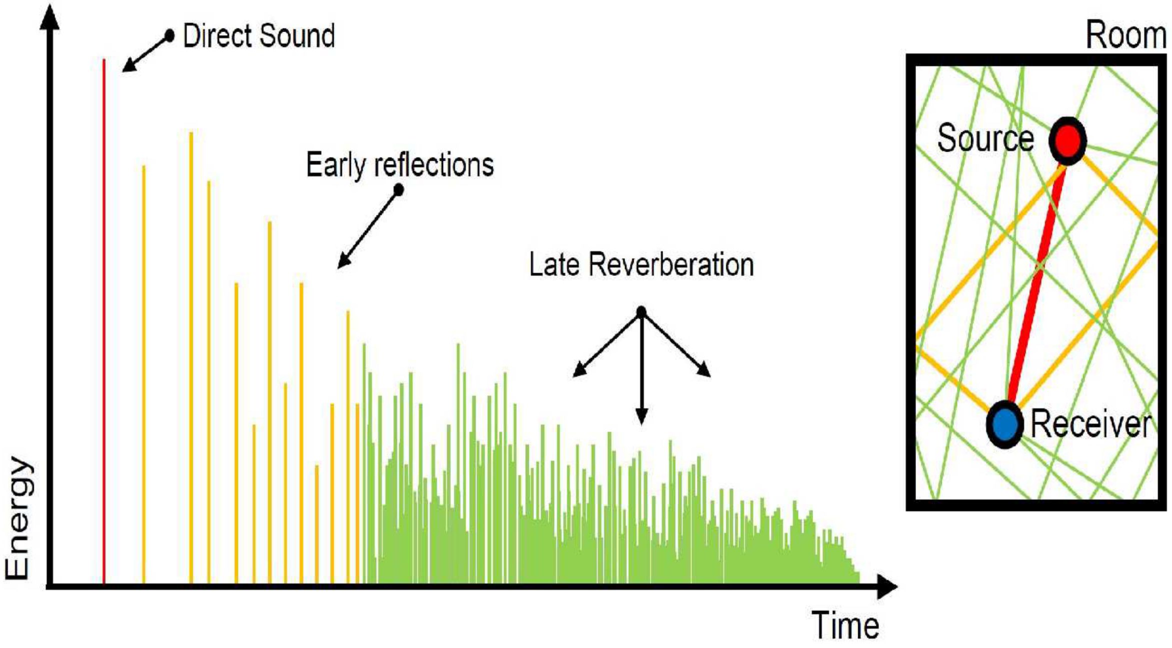


Figure 11.15: Reverberations. (From Pelzer, Aspock, Schroder, and Vorlander, 2014, [253])

Sound Simulation Results

https://www.youtube.com/watch?v=aJOCcaEeLUA

Entering the Ear

- A virtual microphone positioned in the virtual world captures the simulated sound waves

- Convert into audio output through a speaker in front of the ear

- ILD and ITD can be simulated by accounting for both ears
  - Interaural Level Difference (ILD), Interaural Time Difference (ITD)
  - Need to model the physical head in the virtual world
  - Head related transfer function (HRTF)

Tracking the Ears

- If the user turns her head, the sound should be adjusted accordingly

- Perception of stationary for sounds
    - Fixed sound source should be perceived as fixed

- Tracking the ear poses to determine the “viewpoint” for sounds

Further Reading

- Section 11.4, Virtual Reality, Steven LaValle