# Introduction of Audio Signals

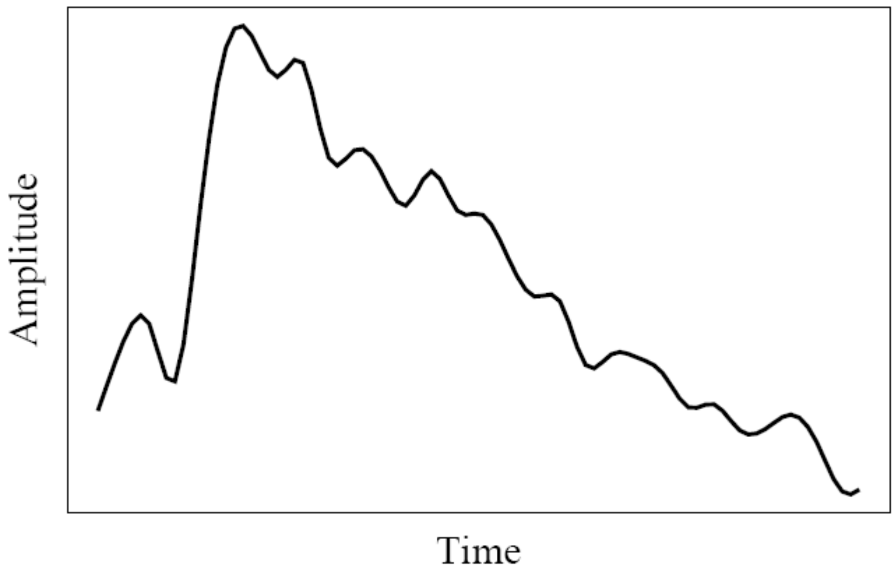
Wei-Ta Chu

Li and Drew, Fundamentals of Multimedia, Prentice Hall, 2004

# Sound



- Sound is a wave phenomenon like light, but is macroscopic and involves molecules of air being compressed and expanded under the action of some physical device.
- Since sound is a pressure wave, it takes on continuous values, as opposed to digitized ones.



# Digitization



- Digitization means conversion to a stream of numbers, and preferably these numbers should be integers for efficiency.
- Sampling and Quantization

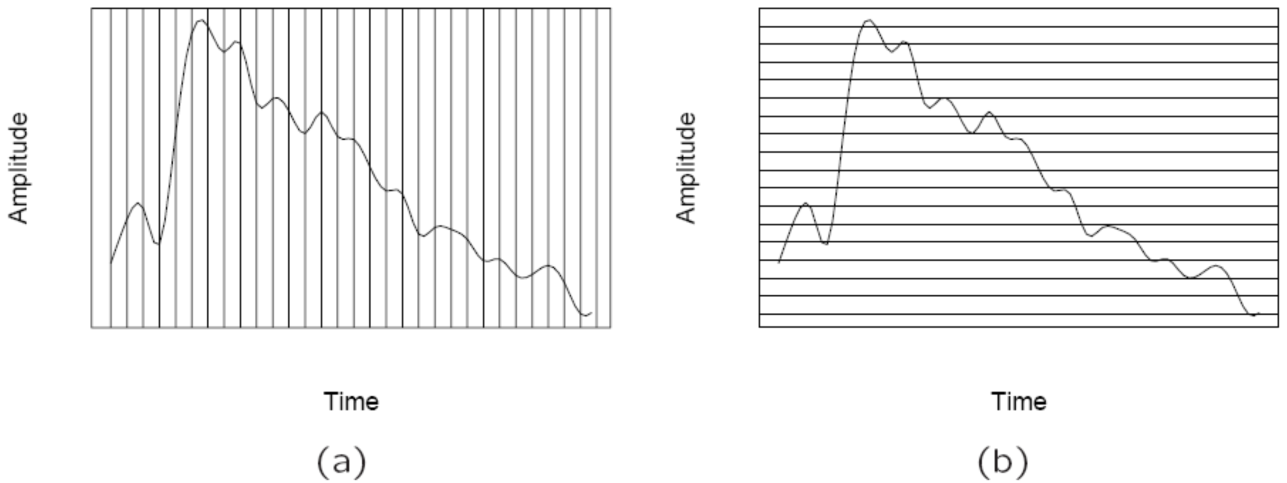



Fig. 6.2: Sampling and Quantization. (a): Sampling the analog signal in the time dimension. (b): Quantization is sampling the analog signal in the amplitude dimension.

# Digitization

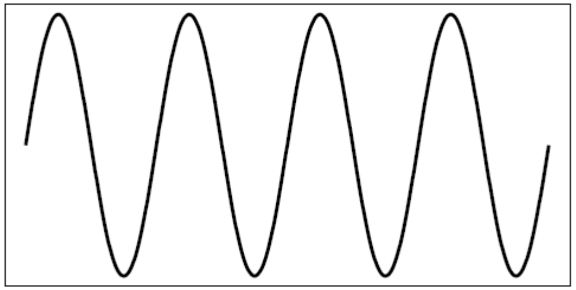


- The first kind of sampling, using measurements only at evenly spaced time intervals, is simply called, ***sampling***. The rate at which it is performed is called the ***sampling frequency***.
- For audio, typical sampling rates are from 8 kHz (8,000 samples per second) to 48 kHz. This range is determined by Nyquist theorem discussed later.
- Typical uniform quantization rates are 8-bit and 16-bit. 8-bit quantization divides the vertical axis into 256 levels, and 16-bit divides it into 65536 levels.

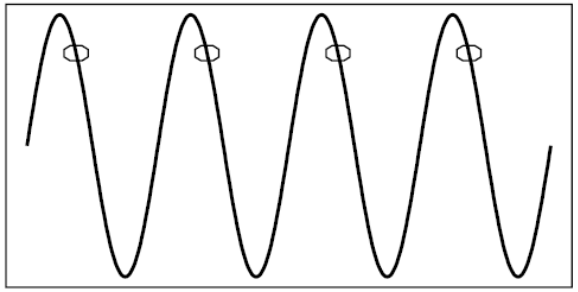
# Nyquist Theorem



- The Nyquist theorem states how frequently we must sample in time to be able to recover the original sound.
- If sampling rate just equals the actual frequency, this figure shows that a false signal is detected: it is simply a constant, with zero frequency.



(a)

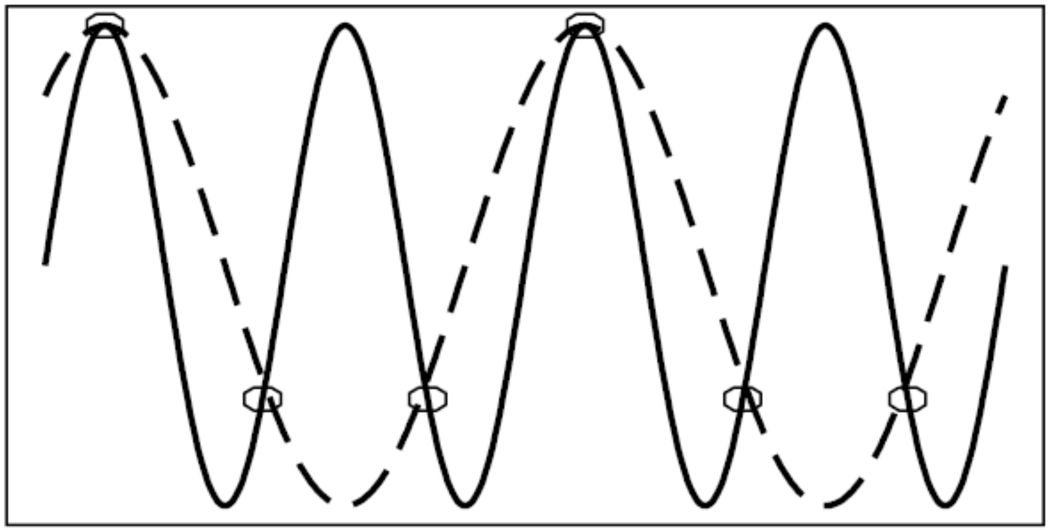


(b)

# Nyquist Theorem



- If sample at 1.5 times the actual frequency, this figure shows that we obtain an incorrect (***alias***) frequency that is lower than the correct one.
- For correct sampling we must use a sampling rate equal to at least twice the maximum frequency content in the signal. This rate is called the ***Nyquist rate***.



# Nyquist Theorem



- **Nyquist Theorem**: The sampling rate has to be at least twice the maximum frequency content in the signal.
- **Nyquist frequency**: half of the Nyquist rate.
  - Since it would be impossible to recover frequencies higher than Nyquist frequency, most systems have an antialiasing filter that restricts the frequency content to a range at or below Nyquist frequency.

# Linear and Nonlinear Quantization



- Non-uniform quantization: set up more finely-spaced levels where humans hear with the most acuity.
  - We are quantizing magnitude, or amplitude – how loud the signal is.
  - Weber's Law stated formally says that equally perceived differences have values proportional to absolute levels:

    $$\Delta \text{Response} \propto \Delta \text{Stimulus} / \text{Stimulus}$$

  - If we can feel an increase in weight from 10 to 11 pounds, then if instead we start at 20 pounds, it would take 22 pounds for us to feel an increase in weight.

# Linear and Nonlinear Quantization



Inserting a constant of proportionality $k$, we have a differential equation that states:

$$dr = k\,(1/s)\,ds \tag{6.6}$$

with response $r$ and stimulus $s$.

– Integrating, we arrive at a solution

$$r = k \ln s + C \tag{6.7}$$

with constant of integration $C$.
Stated differently, the solution is

$$r = k \ln(s/s_0) \tag{6.8}$$

$s_0$ = the lowest level of stimulus that causes a response ($r = 0$ when $s = s_0$).

- Nonlinear quantization works by first transforming an analog signal from the raw $s$ space into the theoretical $r$ space, and then uniformly quantizing the resulting values.

# Linear and Nonlinear Quantization



- For steps near the low end of the signal, quantization steps are effectively more concentrated on the $s$ axis, whereas for large values of $s$, one quantization step in $r$ encompasses a wide range of $s$ values.

# Linear and Nonlinear Quantization



$\mu$-law:

$$r = \frac{\mathrm{sgn}(s)}{\ln(1+\mu)} \ln\left\{1 + \mu\left|\frac{s}{s_p}\right|\right\}, \qquad \left|\frac{s}{s_p}\right| \le 1 \tag{6.9}$$

$A$-law:

$$r = \begin{cases} \frac{A}{1+\ln A}\left(\frac{s}{s_p}\right), & \left|\frac{s}{s_p}\right| \le \frac{1}{A} \\ \frac{\mathrm{sgn}(s)}{1+\ln A}\left[1 + \ln A\left|\frac{s}{s_p}\right|\right], & \frac{1}{A} \le \left|\frac{s}{s_p}\right| \le 1 \end{cases} \tag{6.10}$$

$$\text{where } \mathrm{sgn}(s) = \begin{cases} 1 & \text{if } s > 0, \\ -1 & \text{otherwise} \end{cases}$$

$s_p$ is the peak signal value and $s$ is the current signal value

# Linear and Nonlinear Quantization



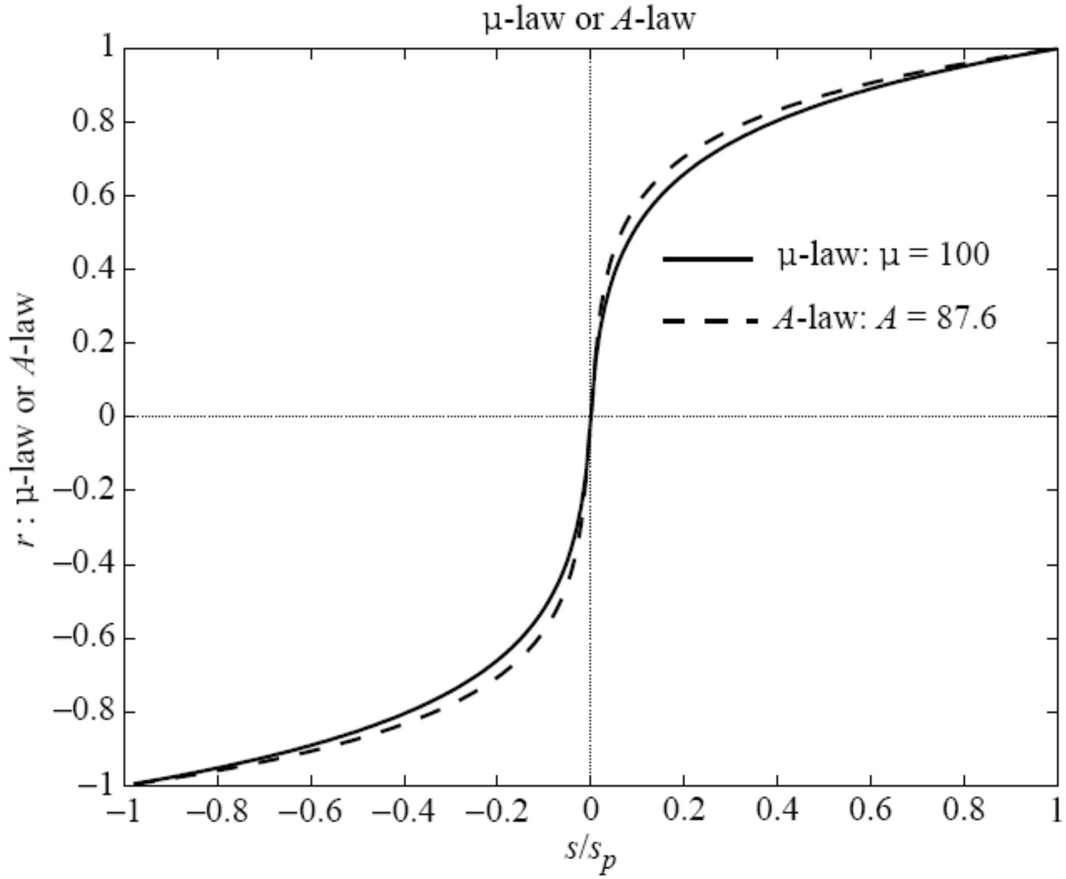



Fig. 6.6: Nonlinear transform for audio signals

- The $\mu$-law in audio is used to develop a nonuniform quantization rule for sound: uniform quantization of $r$ gives finer resolution in $s$ at the quiet end.

# Audio Filtering



- Prior to sampling and analog-to-digital conversion, the audio signal is also usually filtered to remove unwanted frequencies.
- For speech, typically from 50Hz to 10kHz is retained, and other frequencies are blocked by the use of a band-pass filter that screens out lower and higher frequencies.
- An audio music signal will typically contain from about 20Hz up to 20kHz.

# Processing in Frequency Domain



- It's hard to infer much from the time-domain waveform.
- Human hearing is based on frequency analysis.
- Use of frequency analysis often facilitates understanding.

Part of slides are from Prof. Hsu, NTU
http://www.csie.ntu.edu.tw/~winston/courses/mm.ana.idx/index.html

# Problems in Using Fourier Transform



- Fourier transformation contains only frequency information
- No Time information is retained
- Works fine for stationary signals
- Non-stationary or changing signals cause problems
  - Fourier transformation shows frequencies occurring at all times instead of specific times

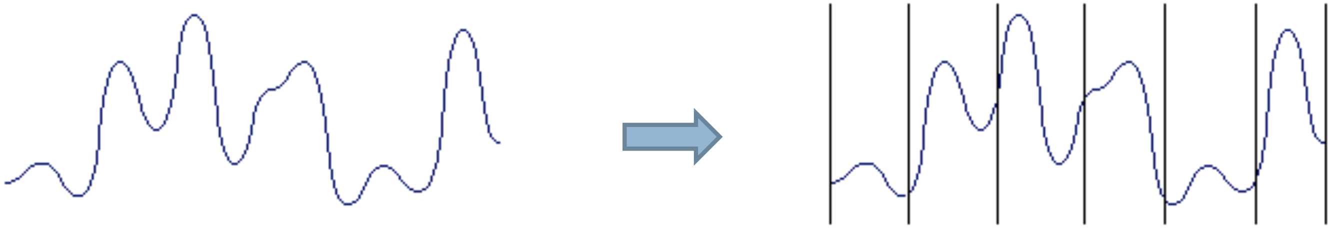
# Short-Time Fourier Transform (STFT)



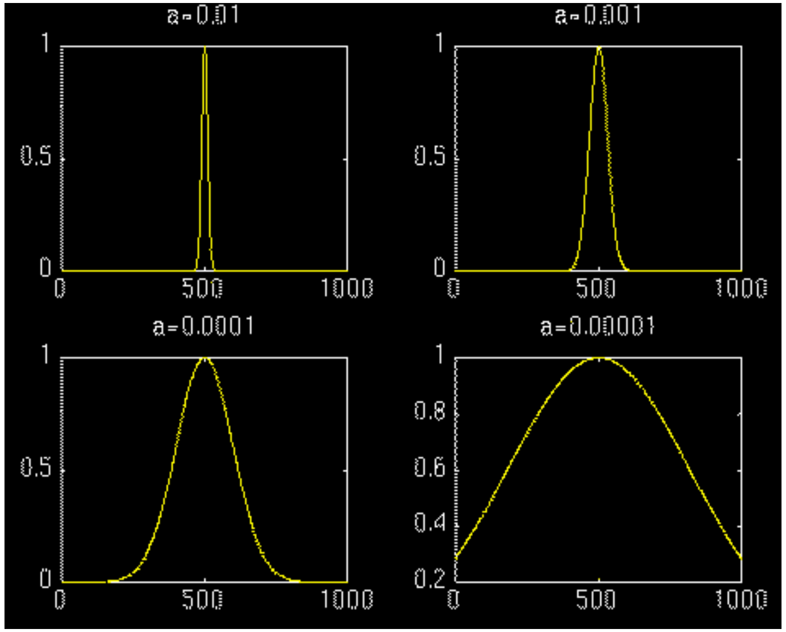
- How can we still use FT, but handle nonstationary signals?
- How can we include time?
- Idea: Break up the signal into discrete windows
- Each signal within a window is a stationary signal
- Take FT over each part

# STFT Example





Window function



# Short-Time Fourier Analysis



- Problem: Conventional Fourier analysis does not capture time-varying nature of audio signals

$$X(\exp^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n] \exp^{-j\omega n}$$

- Solution: Multiply signals by finite-duration window function, then compute DTFT:

$$X[n,\omega] = \sum_{m=0}^{N-1} x[m] w[n-m] \exp^{-j\omega m}$$

# Window Functions



- Rectangular window:

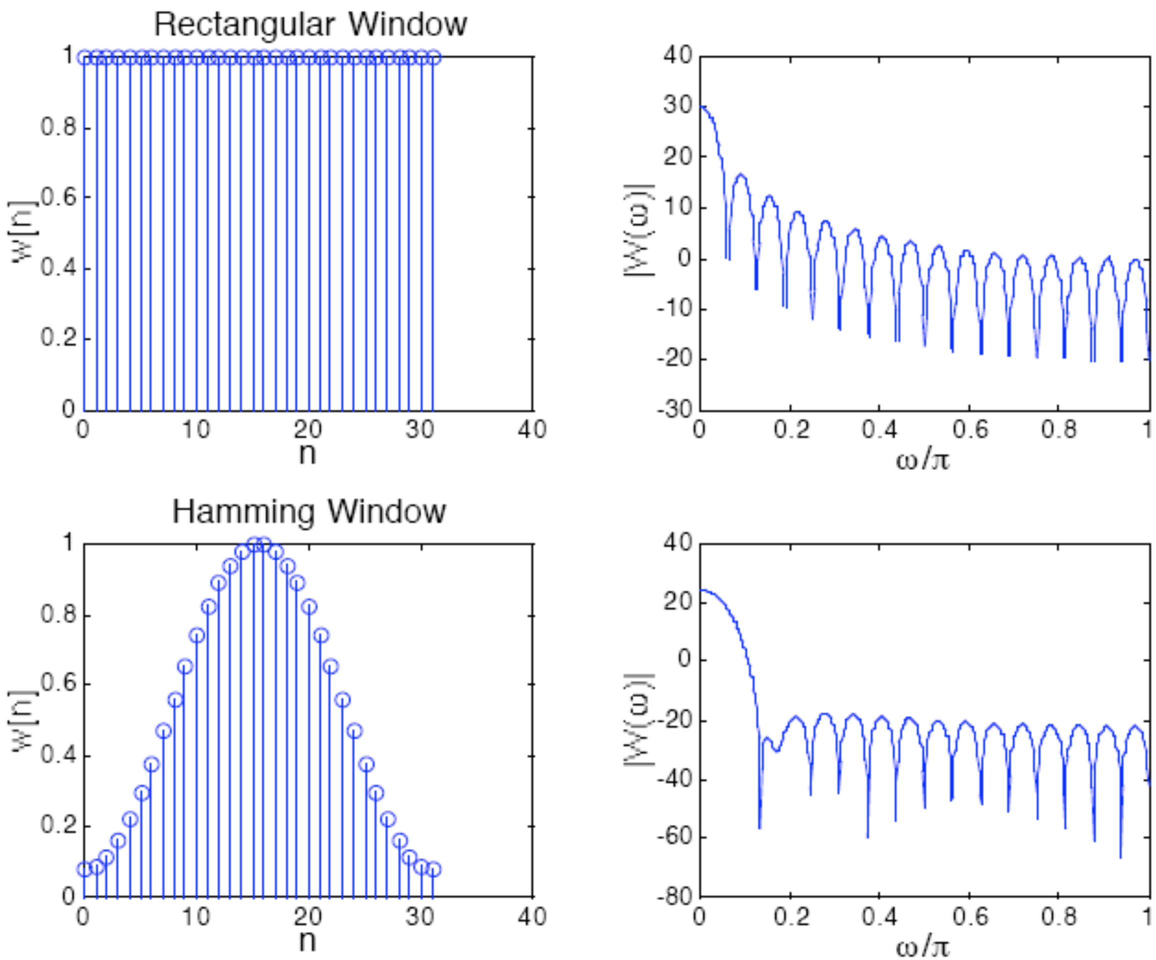
$$w[n] = \begin{cases} 1, & 0 \leq n \leq N \\ 0, & \text{otherwise} \end{cases}$$

- Hamming window:

$$w[n] = \begin{cases} 0.54 - 0.46\cos(2\pi n/N), & 0 \leq n \leq N \\ 0, & \text{otherwise} \end{cases}$$

# Window Functions






Hamming: relatively small side-lobes, fast attenuation which makes it easier for frequency peak detection

# Break Up Original Audio Signals into Frames Using Hamming Window