# Audio Fundamentals

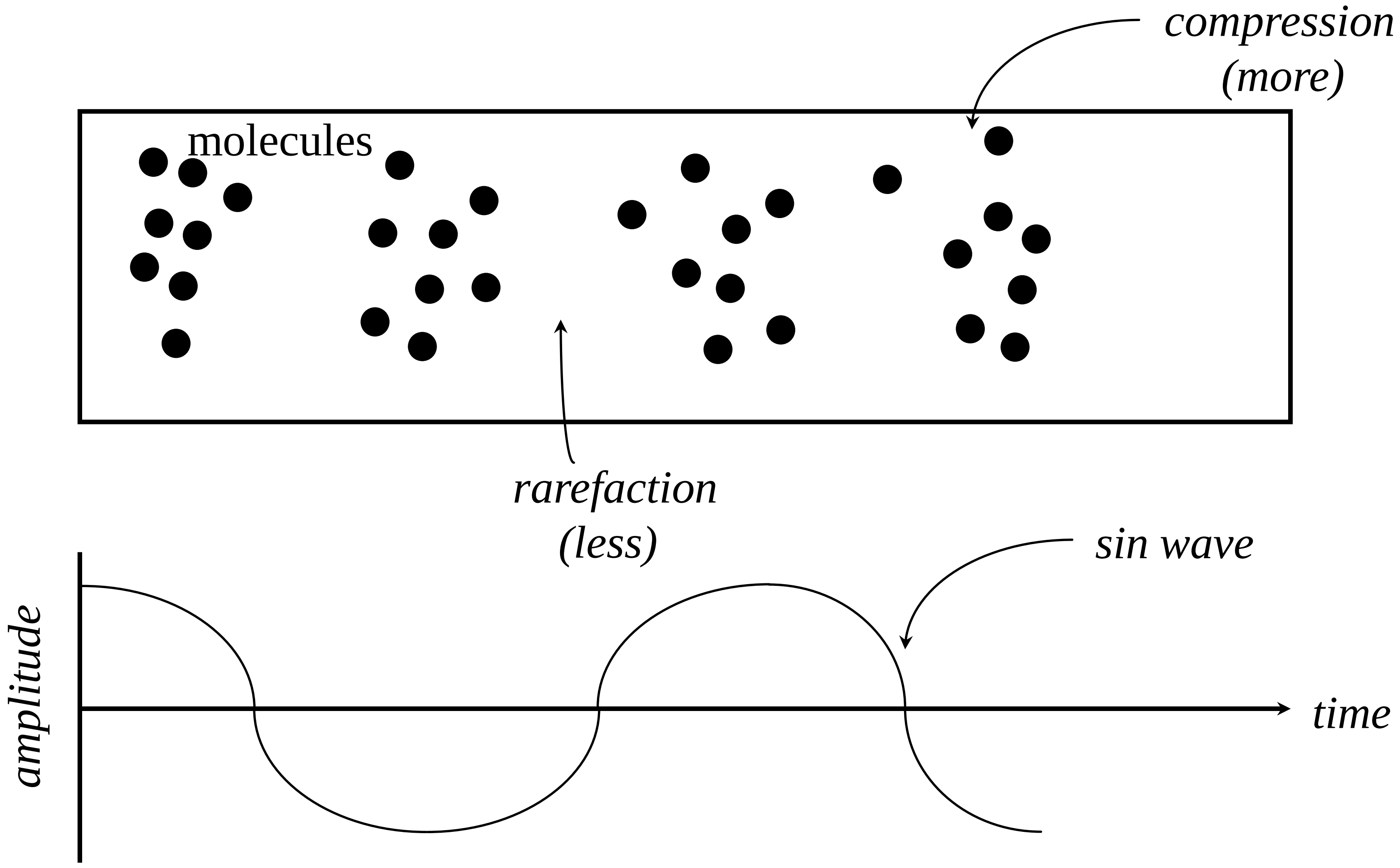
- Sound, Sound Wave and Sound Perception
- Sound Signal
- Analogy/Digital Conversion
- Quantuzation and PCM Coding
- Fourier Transform and Filter
- Nyquest Sampling Theorem
- Sound Sampling Rate and Data Rate
- Speech Processing

# Sound

- Sound, sound wave, acoustics
  - Sound is a continuous wave that travels through a medium
  - Sound wave: energy causes disturbance in a medium, made of pressure differences (measure pressure level at a location)
  - Acoustics is the study of sound: *generation, transmission, and reception* of sound waves

- Example is striking a drum
  - Head of drum vibrates => disturbs air molecules close to head
  - Regions of molecules with pressure above and below equilibrium
  - Sound transmitted by molecules bumping into each other

# Sound Waves


compression
(more)
molecules
rarefaction
(less)
sin wave
amplitude
time

# Sound Transducer

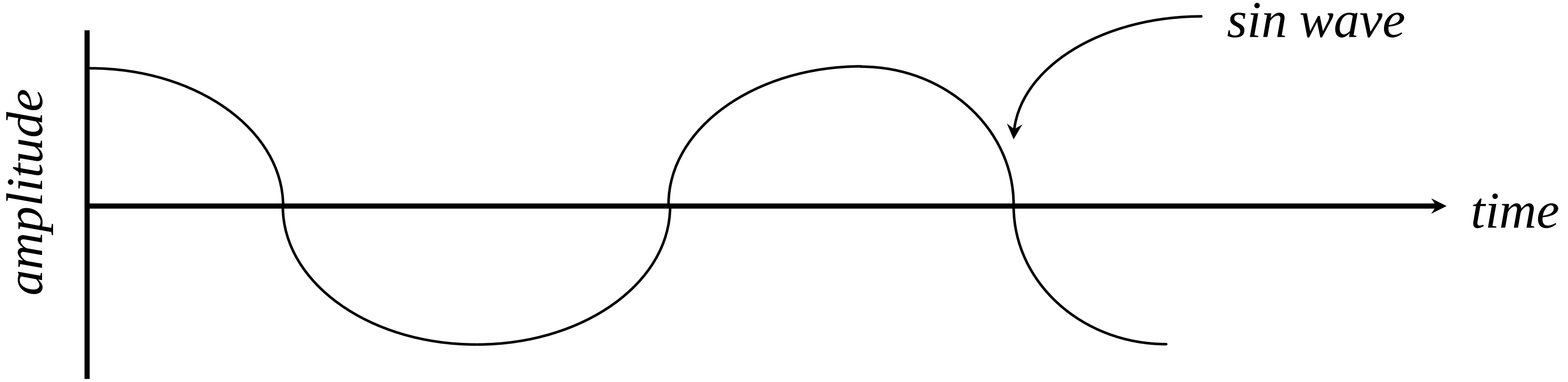
- Transducer

  - A device transforms energy to a different form (e.g., electrical energy)
- Microphone

  - placed in sound field and responds sound wave by producing electronic energy or signal
- Speaker
  - transforms electrical energy to sound waves

# Signal Fundamentals

- Pressure changes can be periodic or aperiodic




- Periodic vibrations
  - –*cycle* - time for compression/rarefaction
  - –*cycles/second* - frequency measured in hertz (Hz)
  - –*period* - time for cycle to occur (1/frequency)

- Human perception frequency ranges of audio [20, 20kHz]

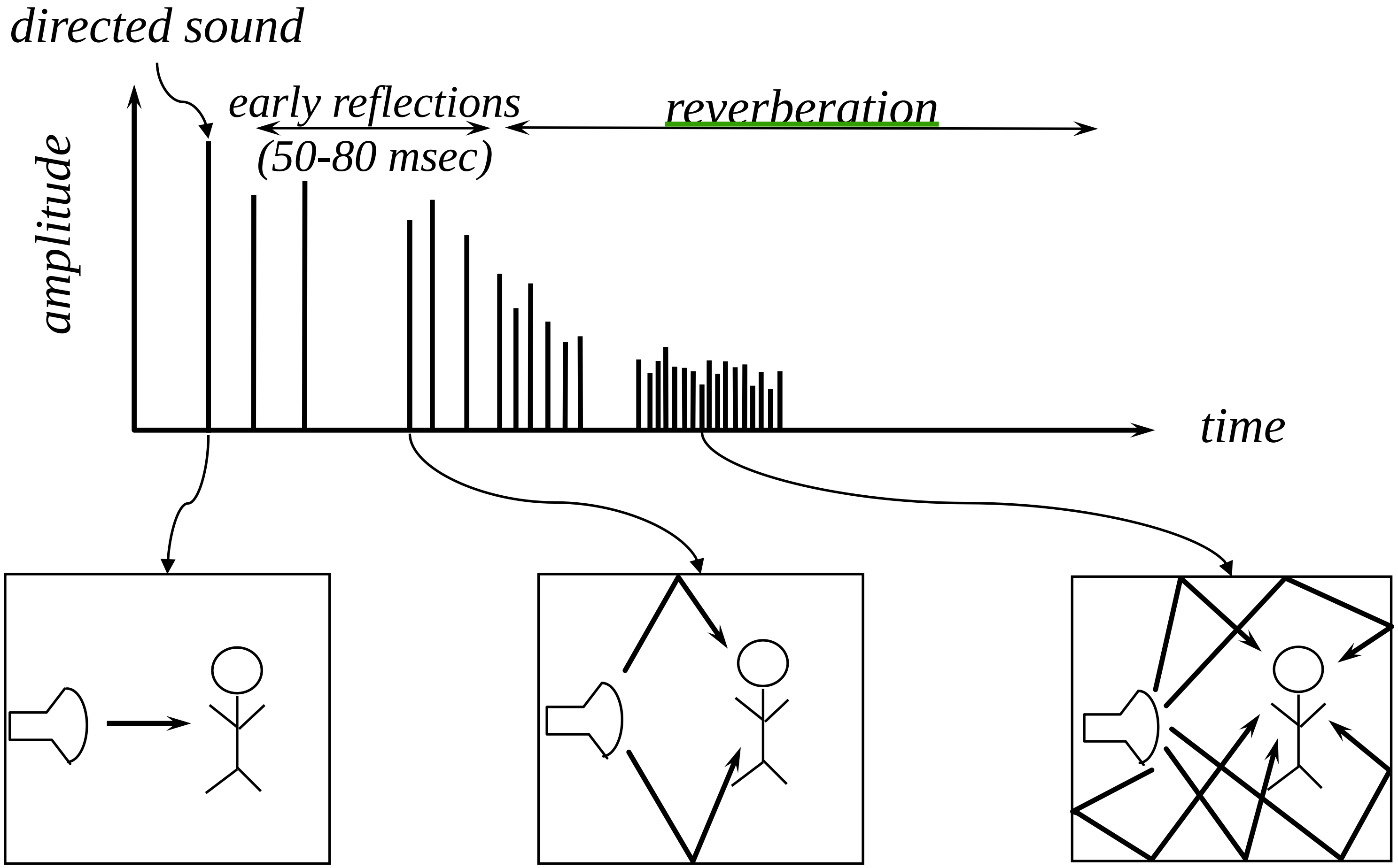
# Measurement of Sound

- A sound source is transferring energy into a medium in the form of sound waves (acoustical energy)
- Sound volume related to pressure amplitude:

  *- sound pressure level (SPL)*
- SPL is measured in *decibels* based on ratios and logarithms because of the extremely wide range of sound pressure that is audible to humans (from one trillionth=$10^{-12}$ of an acoustic watt to one acoustic watt).
  - SPL = 10 log (*pressure*/reference) decibels (dB)
  - where reference is $2*10^{-4}$ dyne/cm$^2$
  - 0 dB SPL - no sound heard (hearing threshold)
  - 35 dB SPL - quiet home
  - 70 dB SPL - noisy street
  - 110 dB SPL - thunder
  - 120 dB SPL - discomfort (threshold of pain)

# Sound Phenomena

- Sound is typically a combination of waves
  - Sine wave is fundamental frequency
  - Other waves added to it to create richer sounds

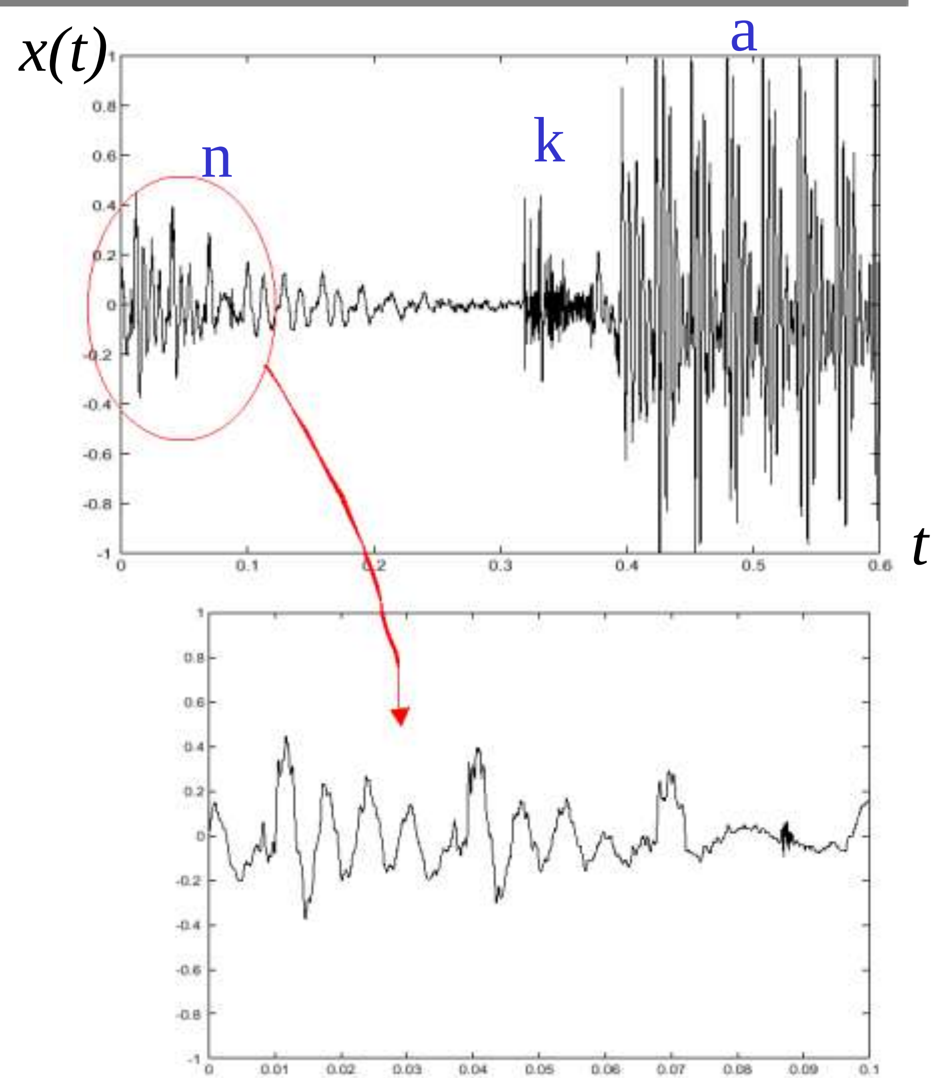


# Human Perception

- Perceptable sound intensity range 0~120dB
  - Most important 10~100dB
- Perceptable frequency range 20Hz~20KHz
- Humans most sensitive to low frequencies
  - Most important region is 2 kHz to 4 kHz
- Hearing dependent on room and environment
- Sounds masked by overlapping sounds
- Speech is a complex waveform
  - Vowels (*a,i,u,e,o*) and bass sounds are low frequencies
  - Consonants (s,sh,k,t,…) are high frequencies

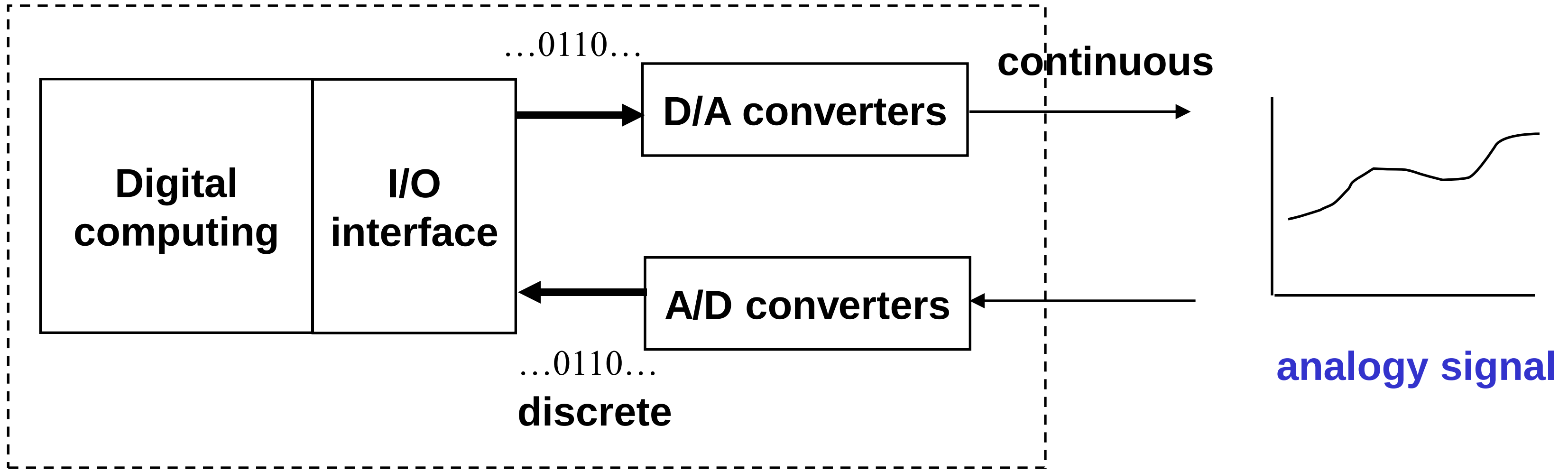
# Sound Wave and Signal

- For example, audio acquired by a microphone
  - Output voltage *x(t)* where *t* is time (continuous) and *x(t)* is a real number
  - One dimensional function
  - Called electronic sound wave or sound signal

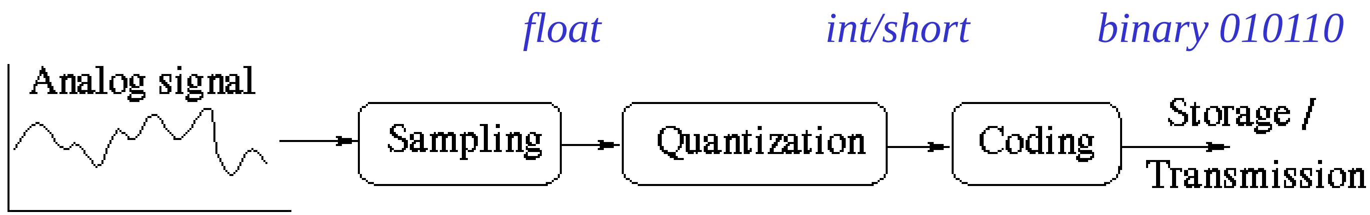


# Analog/Digital Conversion

- Analog signal (continuous change in both temporal and amplitude values) should be acquired in digital forms (digital signal) for the purpose of
  - Processing
  - Transmission
  - Storage & display
- *How to digitize ?*



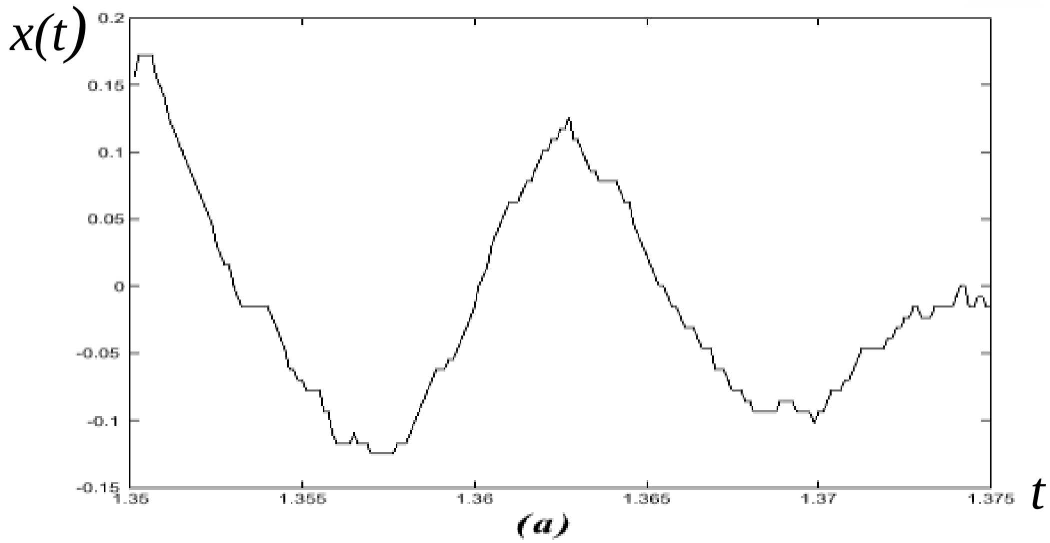
# Process of AD Conversion



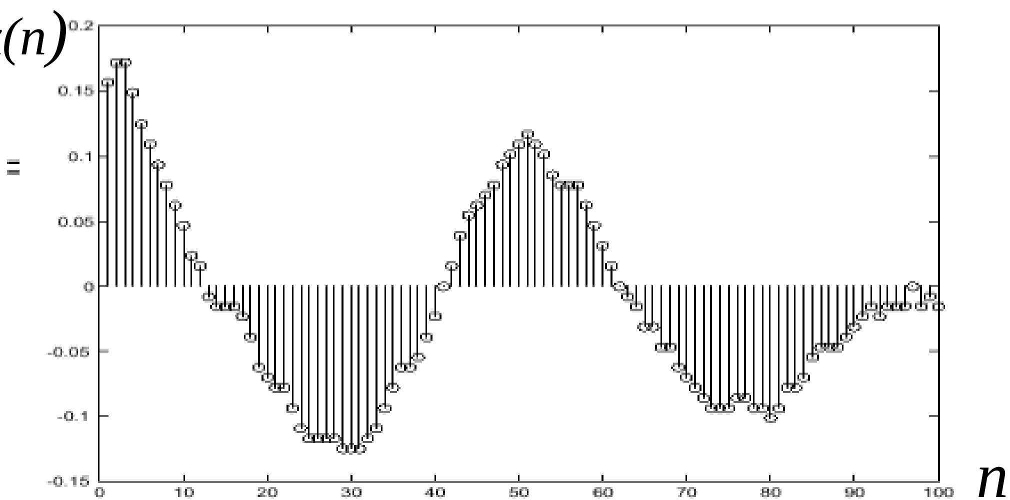

- Sampling (*horizontal*): *x(n)=x(nT),*
  *T* -- sampling period
  Opposite transformation,
  *x(n)* → *x(t),* interpolation.
- Quantization (*vertical*) :

$$\hat{x}(n) = Q(x(n))$$

  *Q()* is a rounding function which maps the value *x(n)* (real number) into value
  in one of *N* levels (integer)
- Coding:
  Convert discrete values to binary digits

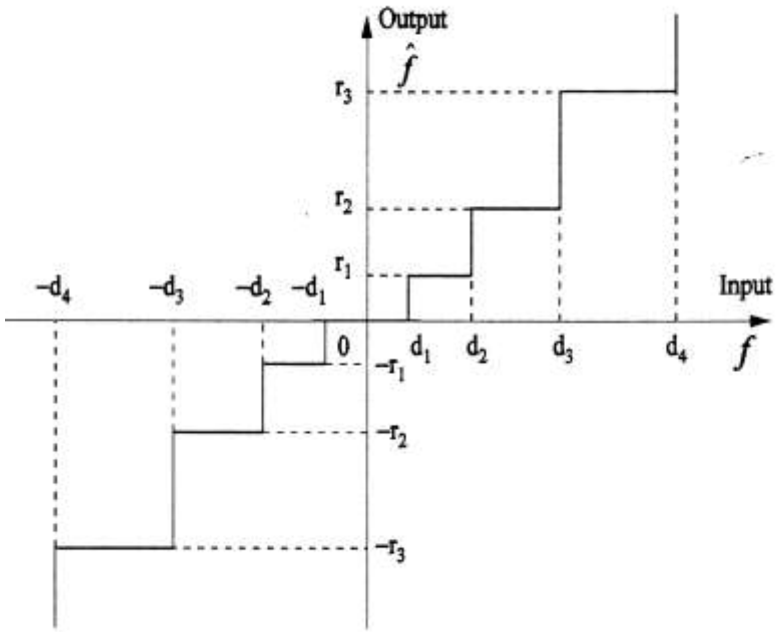


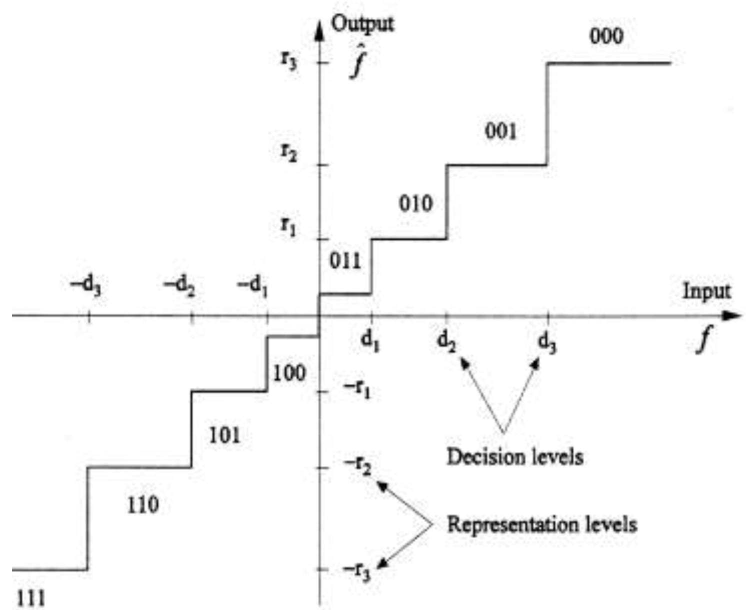

$\hat{x}(n) =$



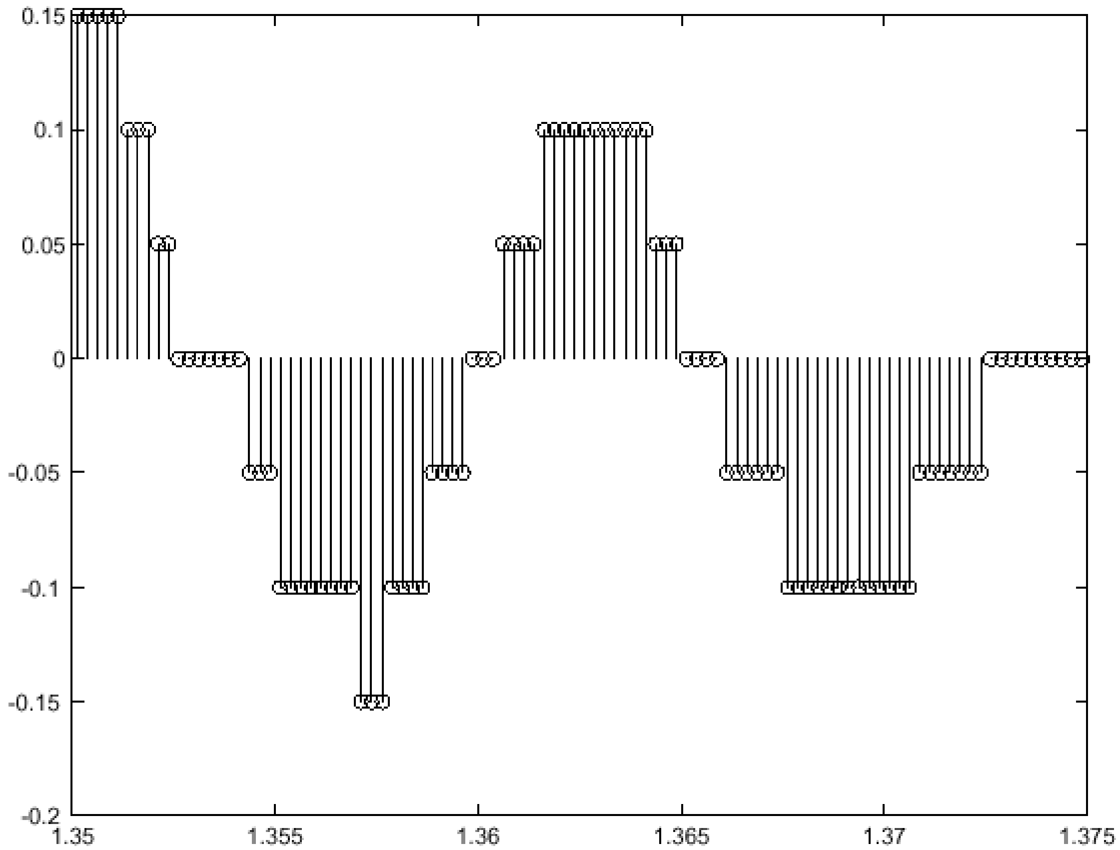
# Quantization and PCM Coding

- Quantization: maps each sample to the nearest value of N levels (*vertical*)
- Quantization error (or quantization noise) is the difference between the actual value of the analog signal at the sampling time and the nearest quantization interval value
- PCM coding (Pulse Code Modulation): Encoding each N-level value to a m-bit binary digit
- The precision of the digital audio sample is determined by the number of bits per sample, typically 8 or 16 bits
-



# Quantized Sound Signal




$(\Delta = 0.05)$

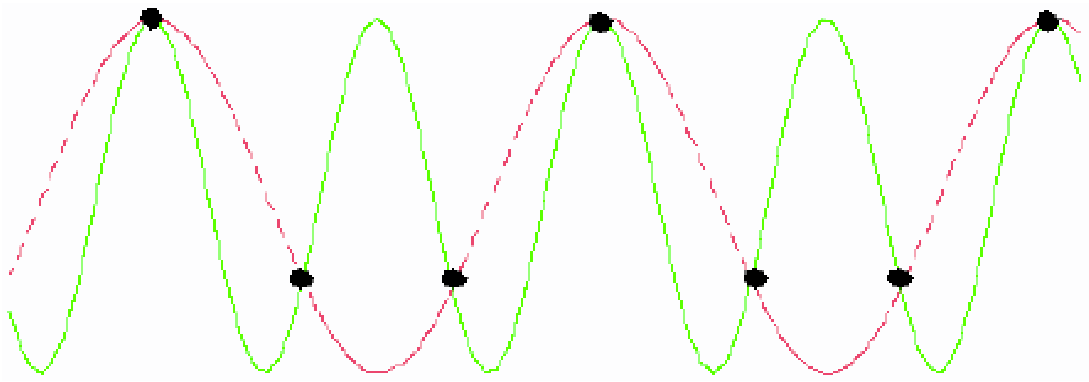
Quantized version of the signal

# Sampling Rate and Bit Rate

- Q. 1: What is the **bit-rate** (bps, bits per second) of the digitized audio using PCM coding? E.g.: CD.
- Sampling frequency is $F$=44.1 KHz
  (Sampling period $T=1/F$=0.0227 ms)
- Quantization with $B$=16 bits ($N=2^{16}$=65,536).
- Bit rate = $BXF$ = 705.6 Kbps = 88.2KBytes/s
  E.g.: 1 minute stereo music: more than 10 MB.
- Q.2: What is the "correct" sampling frequency F? If $F$ is too large, we have too high a bit rate. If $F$ is too small, we have distortion or aliasing . Aliasing means that we loose too much information in the sampling operation, and we are not able to reconstruct ( interpolate ) the original signal $x(t)$ from $x(n)$ anymore.

# Nyquist Sampling Theorem

- Intuitively, the more samples per cycle, the better signal
- A sample per cycle ->constant
- 1.5 samples per cycle -> *aliasing*



- *Sampling Theorem*: a signal must be sampled at least twice as fast as it can change (2 X the cycle of change: *Nyquist rate*) in order to process that signal adequately.

# Fourier Transform

- Fourier transform tells how the energy of signal distributed along the frequencies

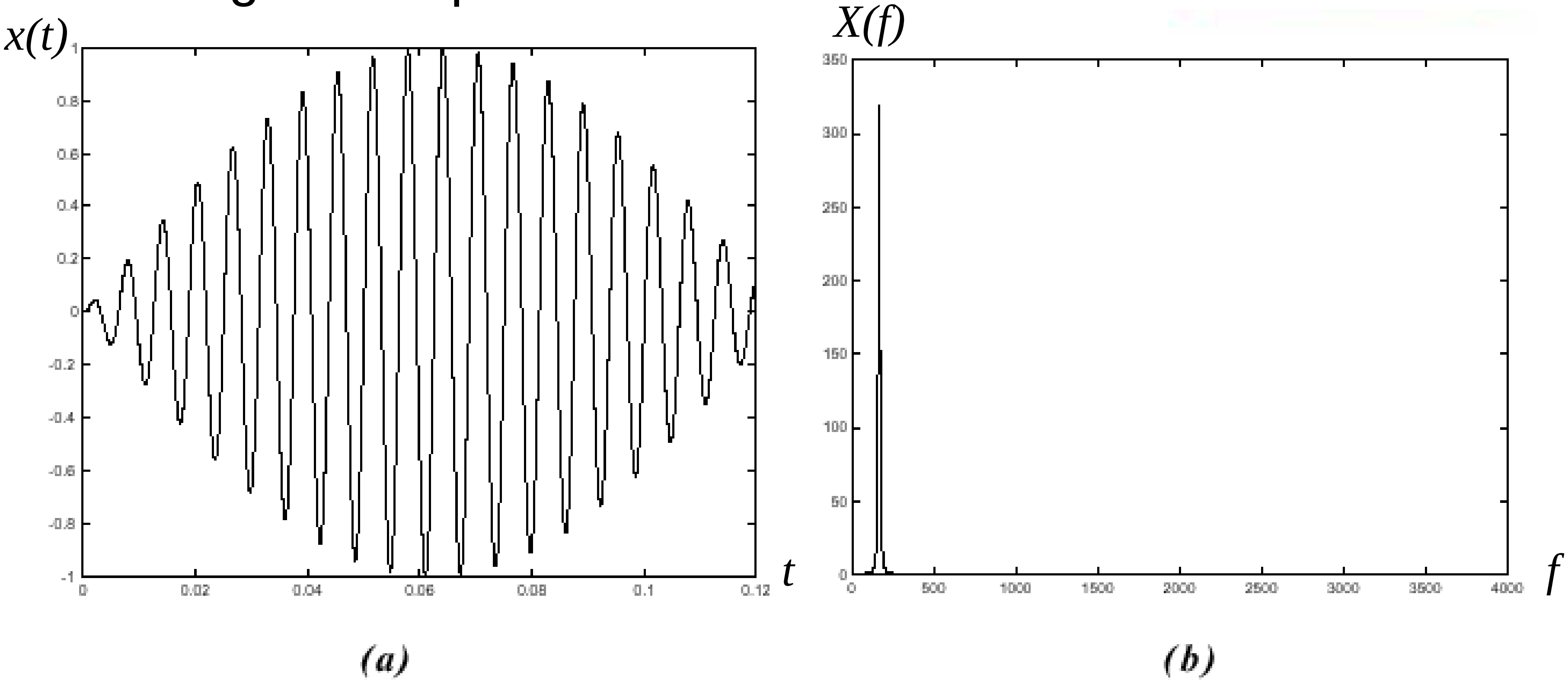
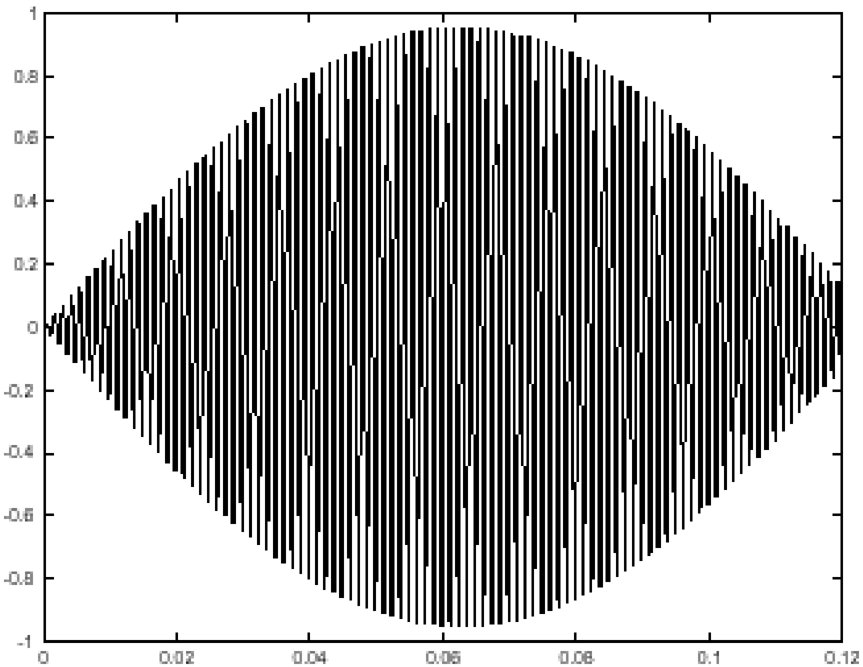
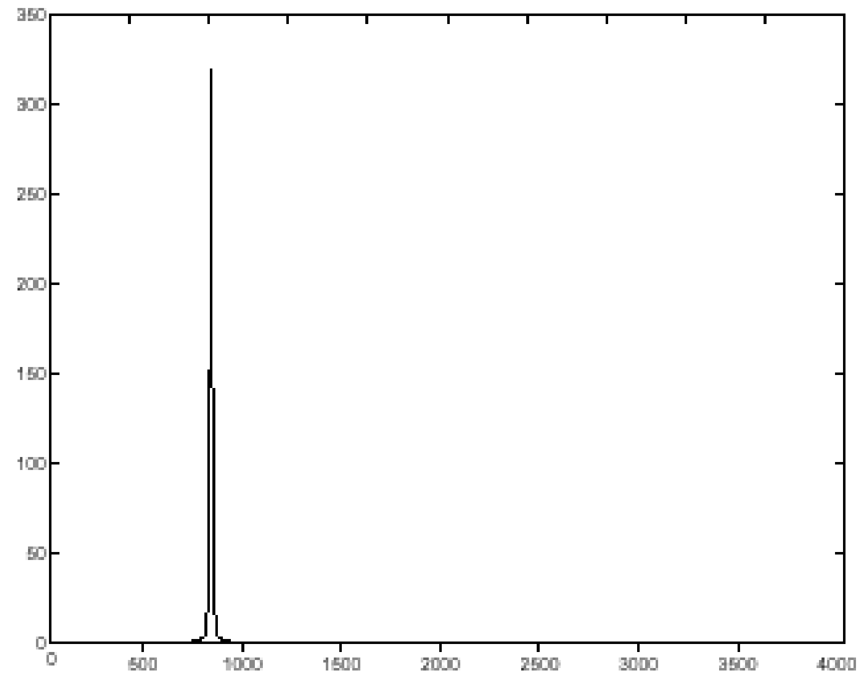



***Figure 5: (a): A tone at 200 Hz. (b): Its Fourier Transform.***

# Fourier Transform (Cont…)




*(a)* *(b)*

*Figure 6: (a): A tone at 800 Hz. (b): Its Fourier Transform.*

# Fourier Transform (Cont…)

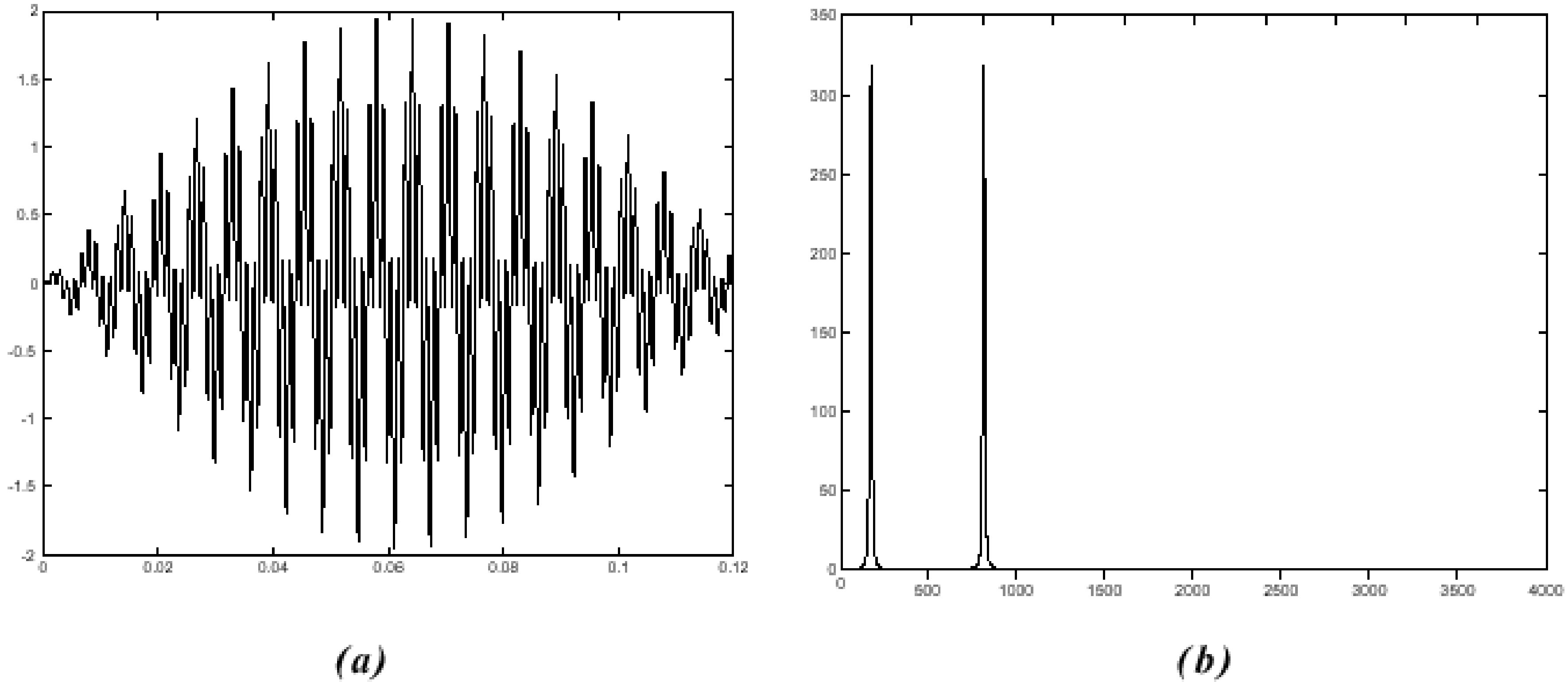



*Figure 7: (a): The sum of a tone at 200 Hz and a tone at 800 Hz. (b): Its Fourier Transform.*

# Fourier Transform (Cont…)

- Using the Fourier's theorem, "*any periodic or aperiodic waveform, no matter how complex, can be analyzed, or decomposed, into a set of simple sinusoid waves with calculated frequencies, amplitudes, phase angles*"
- Change the discussion from time domain to frequency domain
- The mathematical manipulations required for Fourier analyses are quite sophisticated. However, human brain can perform the equivalent analyses almost automatically, both blending and decomposing complex sounds.

# Filters

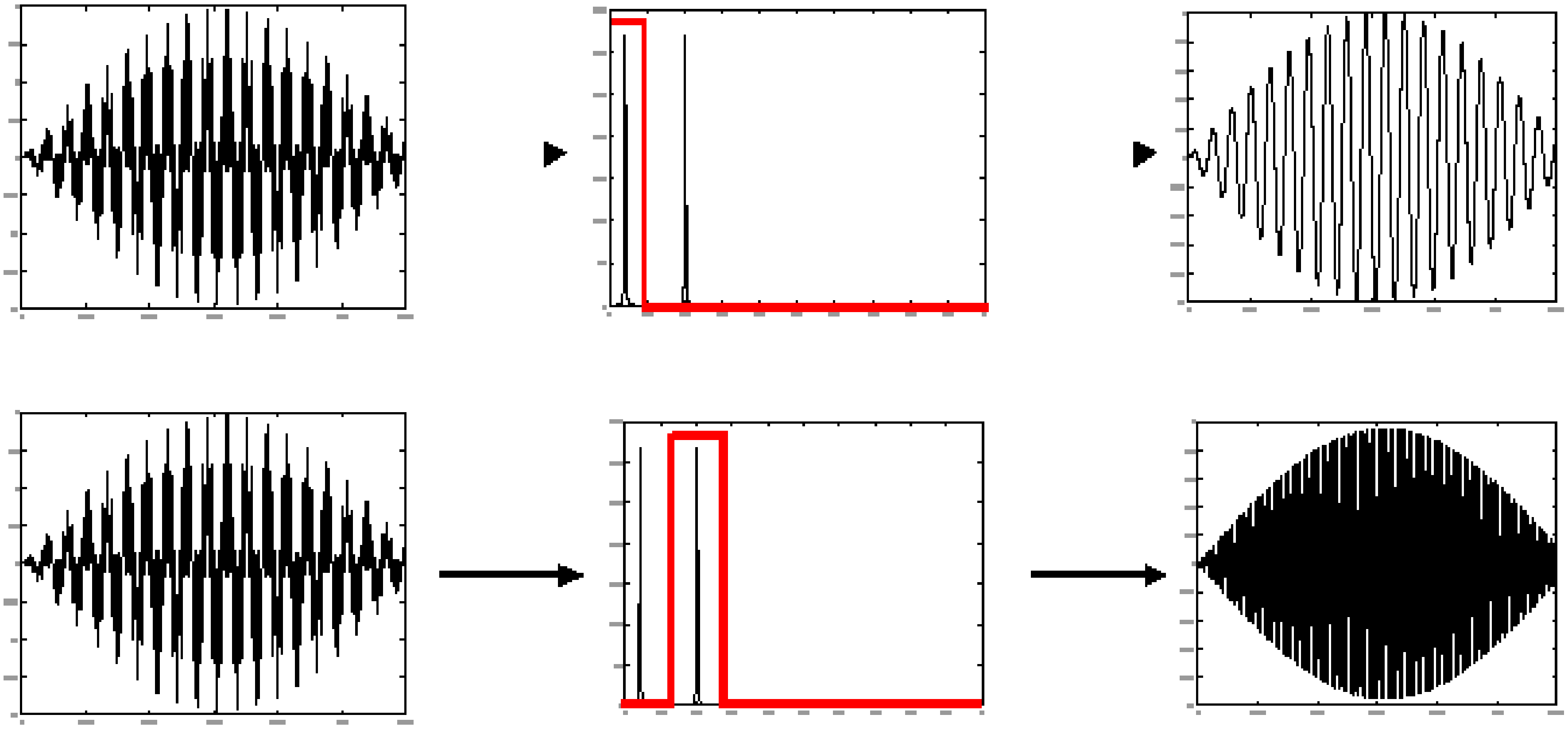
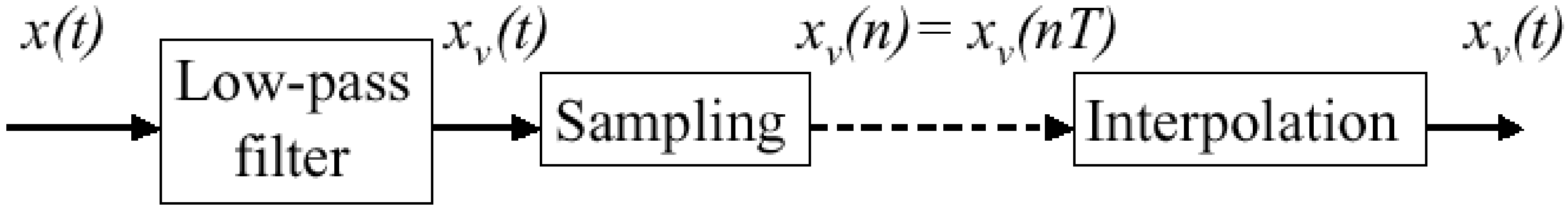


*Figure 8: Filters with bandwidth between 0 and 400 Hz (first row) and between 400 and 1200 Hz (second row) and their action to the signal of Figure 7.*

# Sampling

- Sequence of sampling




- A signal bandwidth-limited to *B* can be fully reconstructed from its samples, if the sampling rate is *at least twice of the highest frequency of the signal*, i.e., the sampling period is less than *1/2B* – **Nyquist sampling rate**
- Subsampling: a technique where the overall amount of data that will represent the digitized signal has been reduced (because this violate the sampling theorem, many types of distortion/aliasing may be noticeable)

# Sampling Rate and PCM Data Rate

| Quality | Sampling Rate (KHz) | Bits per Sample | Data Rate Kbits/s Kbytes/s | Freq. Band |
|---|---|---|---|---|
| Telephone | 8 | 8 (Mono) | 64<br>8 | 200-3,400 Hz |
| AM Radio | 11.025 | 8 (Mono) | 88.2<br>11.0 | 100-5,000 Hz |
| FM Radio | 22.050 | 16 (Stereo) | 705.6<br>88.2 | 50-10,000 Hz |
| CD | 44.1 | 16 (Stereo) | 1411.2<br>176.4 | 20-20,000 Hz |

# Speech Processing

- Speech enhancement
- Speech recognition
- Speech understanding
- Speech synthesis

- Transcription
  - dictation, information retrieval
- Command and control
  - data entry, device control, navigation
- Information access
  - airline schedules, stock quotes