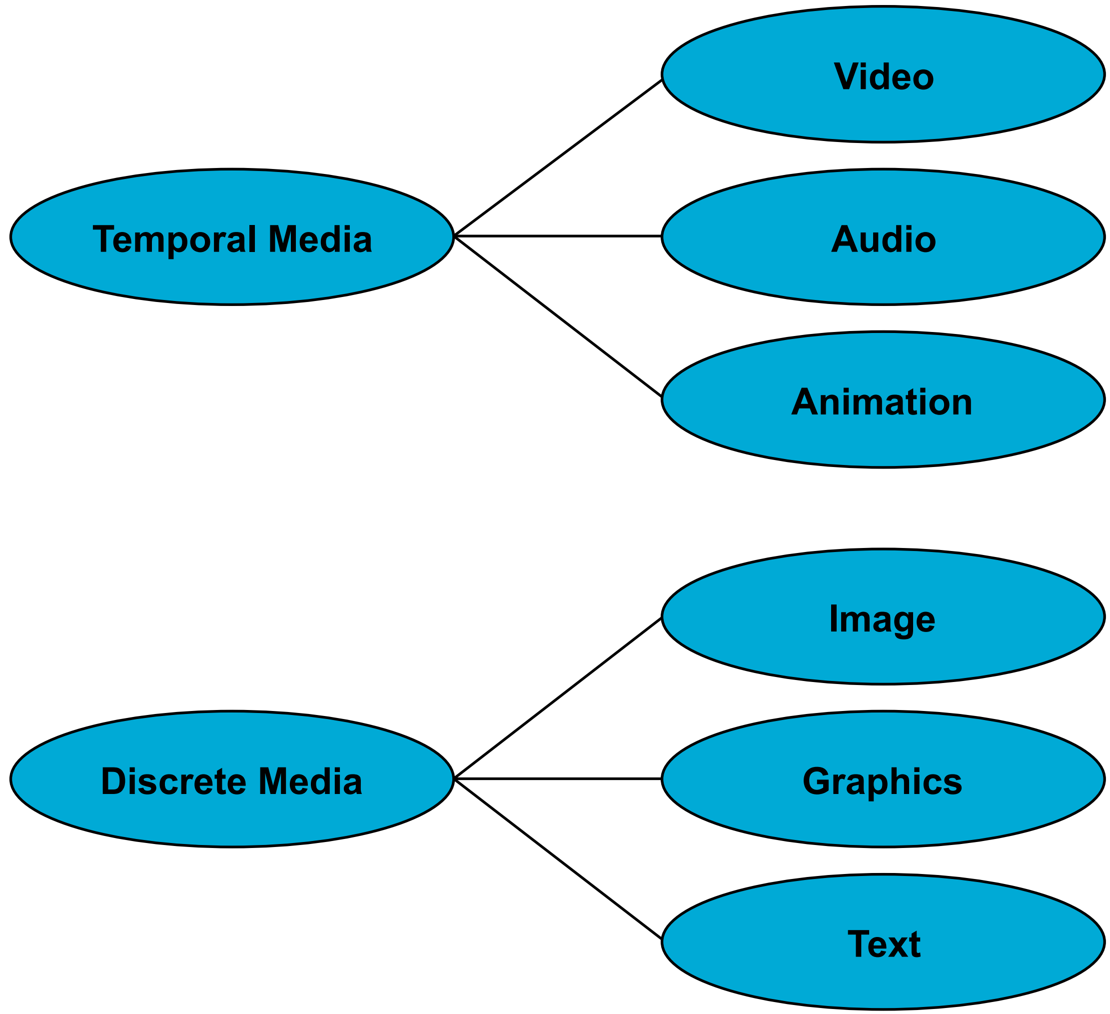
Chapter 2

Media Representations:

Audio Media

Media Hierarchy



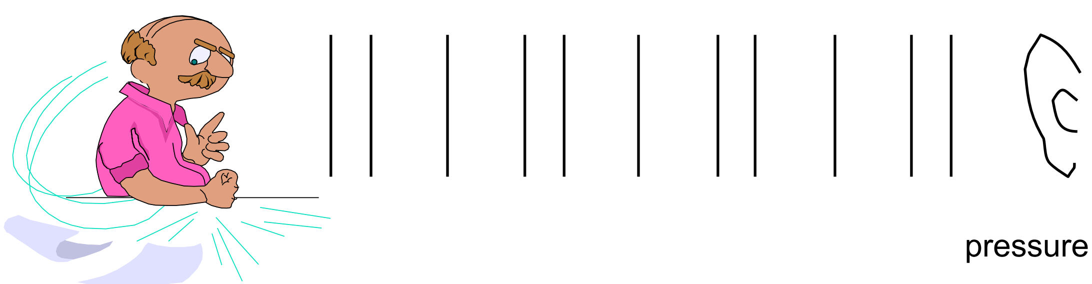
Temporal Media

- Contents and meanings depend on presentation time.

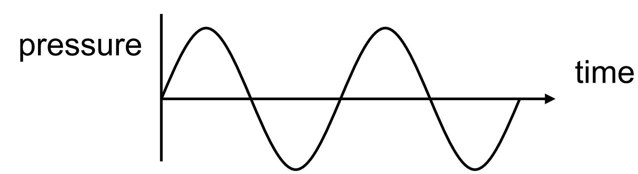
- End-to-end fixed time relationship, from data capture to playback - **isochronous** media. There are lower and upper delay time bound (requirement) for data delivering.

Measuring Audio

- Audio sensation is caused by vibration in air pressure that reaches the human ear-drum.
- Audible frequency of vibration ranges from 20Hz to 20KHz.
- How come audio CD is 44.1 KHz sampled?
- Pressure fluctuation causes sound heard as soft or hard, measured by **amplitude**.



frequency
amplitude



Measuring Audio (2)

- Dynamic range of human hearing is very large.
  - Lower limit is threshold of audibility.
  - Upper limit is threshold of pain.
  - The two can differ by 1,000,000 times (1 Million).
- To work with large range, audio amplitude is often measured in dB (*decibels*).

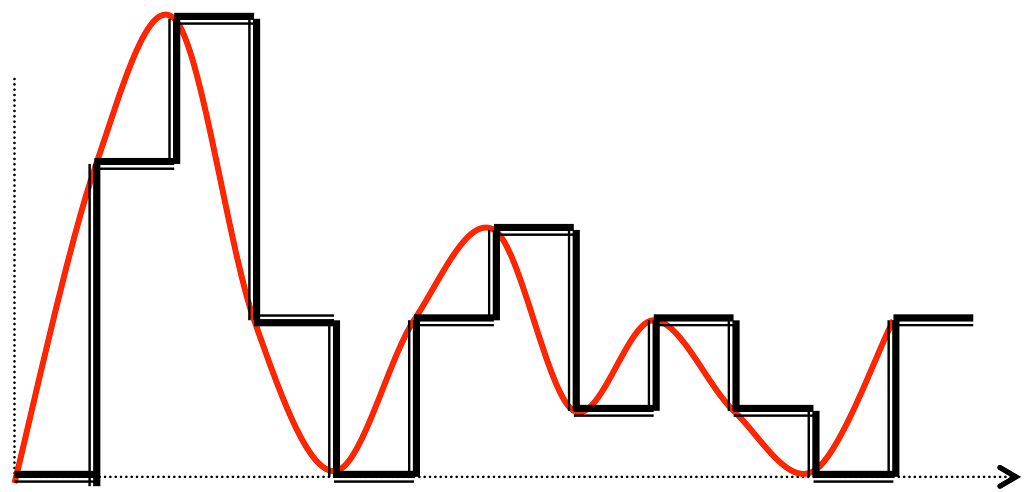
$$dB = 20\log_{10}\left(\frac{x}{y}\right)$$

Also as signal-to-noise ratio (SNR)

If *y* is the amplitude of audibility threshold and *x* is the upper limit, then
the threshold of pain = 20 * log1000000 = 120 dB.

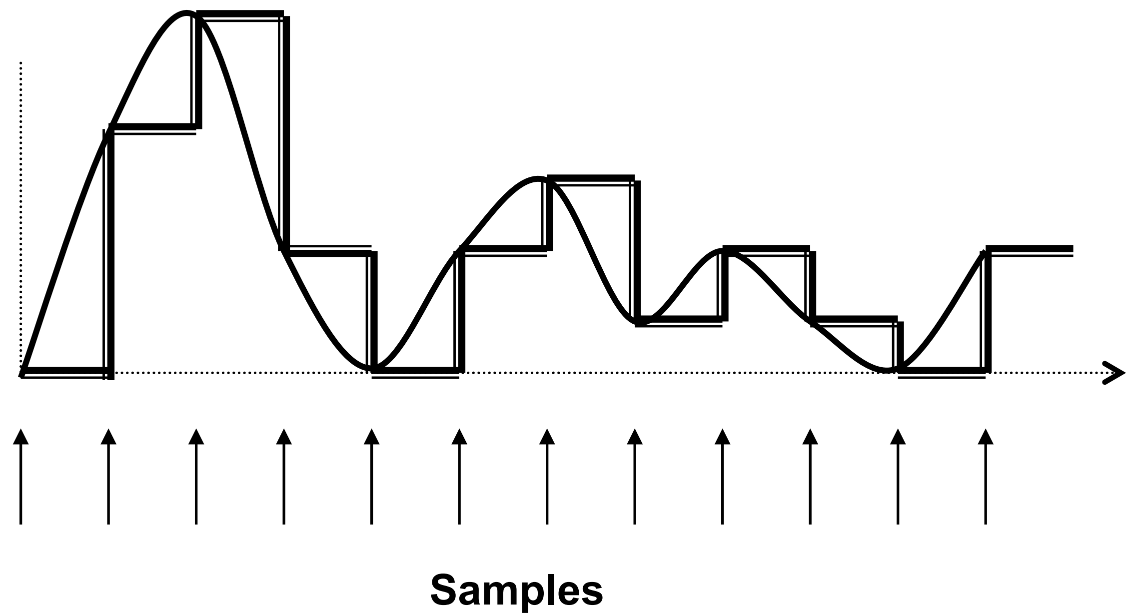
Representation of Audio Data

- Continuous audio waveform → electrical voltage in a microphone (*analog signal*) (red curve)
- Analog → digital for computer processing (*ADC conversion*).
- Digital → analog in soundcard/speaker during playback (*DAC*).
- 3 stages:
  - Sampling
  - Quantization
  - Coding



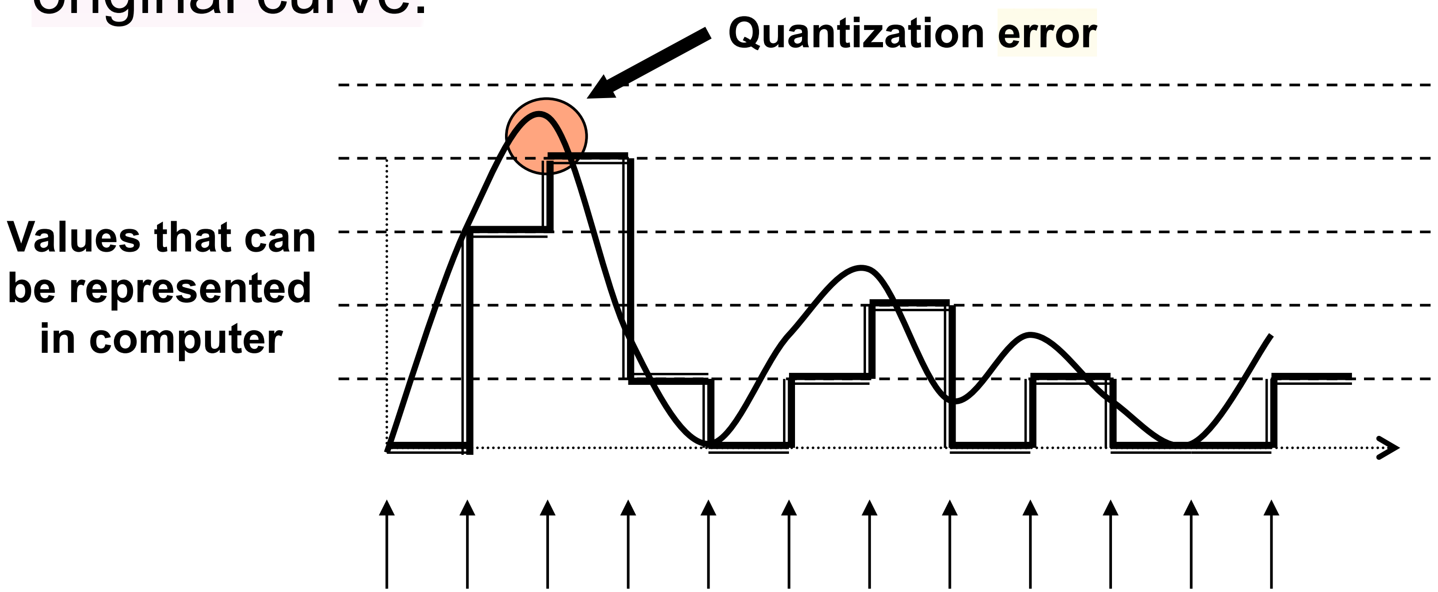
Sampling (in time domain)

- Continuous time → fixed intervals at which the analog signal is read.
- Frequency is called *sampling rate*.
- Higher sampling rate ⇒ better approximation to original curve.



量化
Quantization (in value domain)

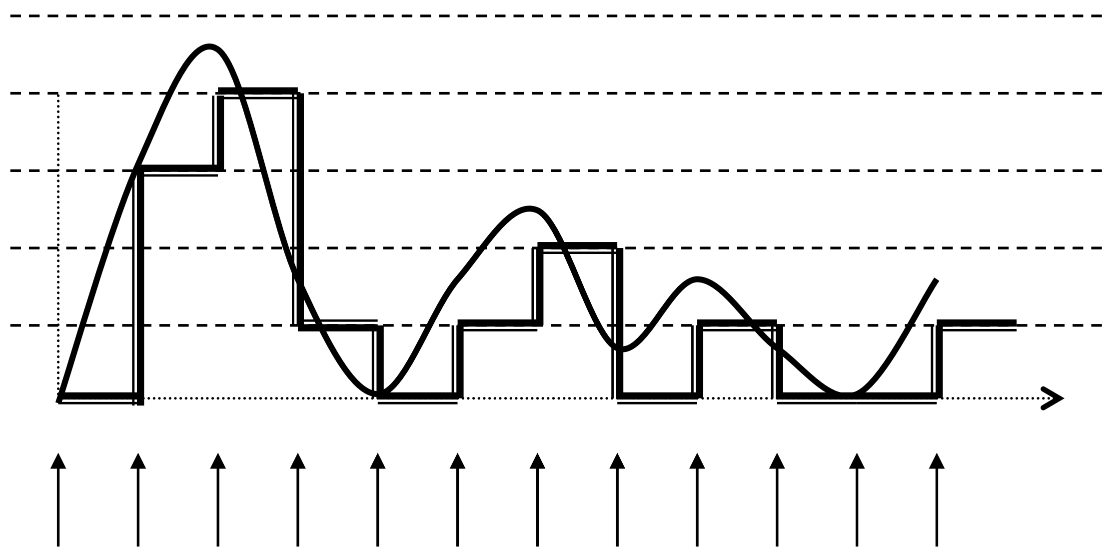
- Continuous signal levels → fixed intervals 整數
  → discrete values
- Size of interval is called *quantization step*.
- Smaller quantization step ⇒ better approximation to original curve.



時間與數值的量化 —>有quantization error

Coding

- Representing quantized values digitally is called *coding*.
- 6 levels, hence 3 bits are enough:
  - 000 011 100 001 000 001 010 000 001 000 000 001



Sampling Rate

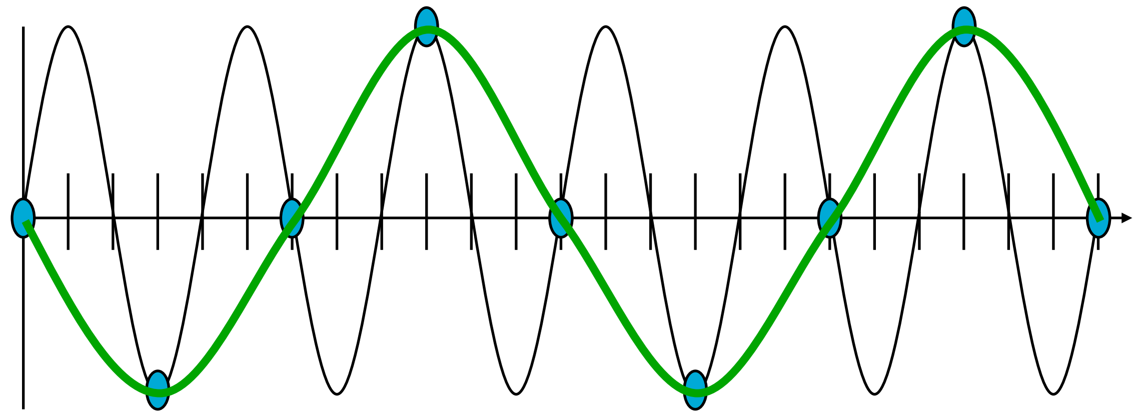
- *Nyquist theorem*: Signal contains frequency components up to *f* Hz, then effective sampling requires

  sampling rate ≥ 2*f* Hz (*critical sampling*)

- Audible frequency range is 20KHz ⇒ CD sampling is 44.1KHz
- Human voice range < 3.1KHz ⇒ digital telephone is 8kHz.
- *Aliasing* : problem of sampling at < critical sampling rate.

Sampling Rate (2)

- For example, actual frequency = *f*
- Let sample at a sampling rate of 1.33 *f*



- We reconstruct a wrong waveform with 1/3 *f*
- This phenomena is called aliasing in sampling theory

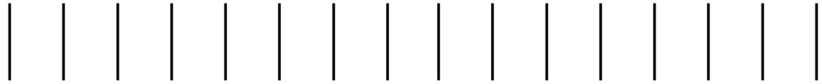
只有藍點不能夠
recover到原本的
wave —> aliasing

Quantization Levels

- Discrepancy between sampled values and original analog signal values gives rise to

  Quantization error (noise)

- Quantization levels affects choice of number of coding bits.
- Quantization levels is manifested in **Signal-to-Noise Ratio** (*SNR).*
- Usually quantization levels are linear



- Logarithmic scale is more uniform in perceptual domain

How good is the signal?

- In analog system, there is voltage (signal) you want to measure, and there is always some random fluctuations (noise).
- Ratio of the power of signal and noise is called the **Signal-to-Noise Ratio** (SNR).
- The large the ratio is, the better is the signal quality.
- Measuring unit: decibels (dB).

How good is the Quantization?

- "How are bits related to *SNR*?"

$$SNR = 20\log_{10}(S/N)$$

N is quant'n noise,
S is signal level

- 1 **more** bit used in coding, max signal increase by 2 and hence increases *SNR* by 6 dB.

$$SNR = 20\log_{10}(2S/N)$$
$$= 20\log_{10}(2) + 20\log_{10}(S/N)$$
$$\approx 6\text{dB} + 20\log_{10}(S/N)$$

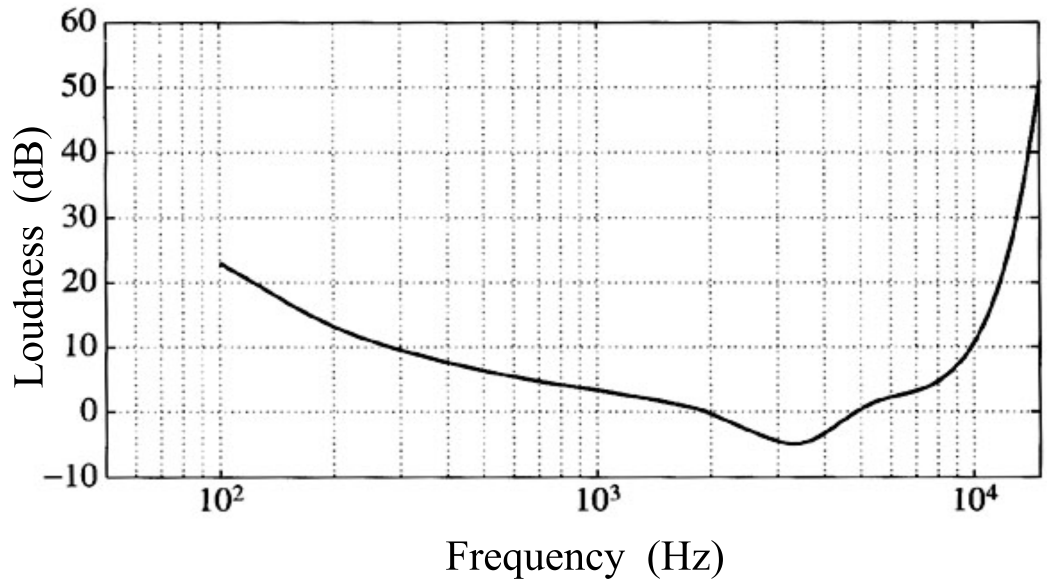
- In practice, 8-bit audio gives 48 dB.
- 16-bit CD-audio gives 98 dB. why not 96?
  - Good because it is close to the audibility limit of 100~120 dB.
  - When *SNR* is close to 120 dB, quantization noise is subdued to audibility threshold.

Limitations of Human Hearing

- Before we go on, let's study the limitation of human hearing perception
- Just like visual media (discussed in next chapter), human hearing has much limitation
- Have you experience that some tones are not audible when they are not loud enough?
- We cannot hear every tone that physically exists
- It is pointless to store the "sound" that we cannot hear
- These limitations are the foundation of modern compression of digital audio
- In this chapter, we only study the limitations, the compression method/standard (such as MPEG Layer 3 or MP3) that utilizes these limitations will be discussed later. (Because we need more knowledge to understand the MP3)

Threshold of Hearing

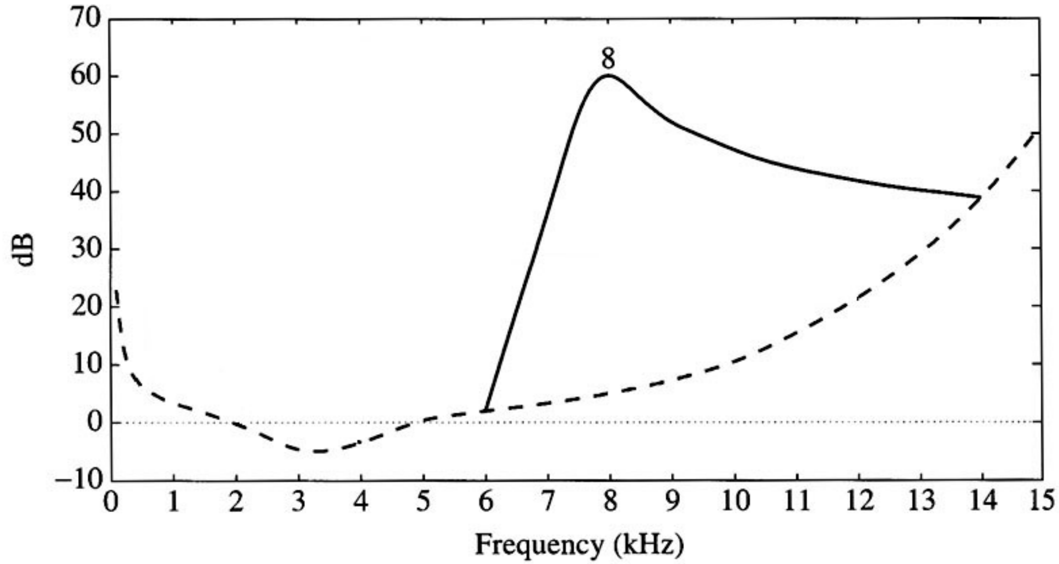
- A frequency is not audible if its loudness (measured in dB) is below the **threshold of hearing**
- The threshold is frequency dependent



- A psychological experiment is done to obtain the plot
- Generate that frequency and turn up its volume (loudness) until it is barely audible. That loudness is the threshold.

Frequency Masking

- When a particular frequency (masking tone) is played at a loud volume, it may mask the nearby frequency
i.e. we cannot hear that nearby frequency
- This is known as **frequency masking**
- As the masking tone changes, frequency masking curve changes
- The following plots show three "add-on" frequency masking curves for 8kHz played with 60dB loudness



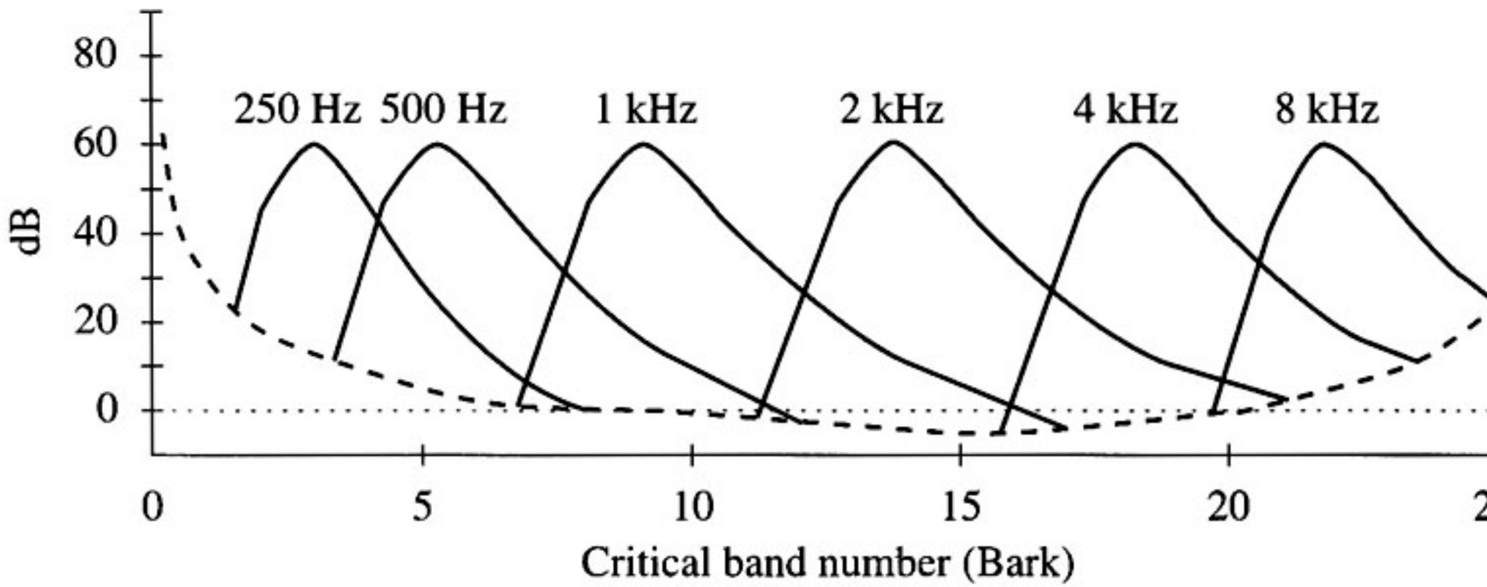
Frequency masking curve when 8kHz is played with 60dB

Critical Bands (1)

- Human hearing range naturally divides into **critical bands**
- Our ears operate like a set of band-pass filters (a limited range of frequencies is passed while others are blocked
- Within a critical band, we human are not very well in resolving frequencies in the same band
- The width of critical band (critical bandwidth) is not constant
- Lower critical bands have smaller bandwidths while high critical bands have larger bandwidths
- For bands above 500Hz, their bandwidths increases roughly linearly
- There are 24 critical bands

Critical Bands (2)

- We can "normalize" the band to form a new unit called "Bark"
- The following diagram shows the frequency masking curves in this Bark domain

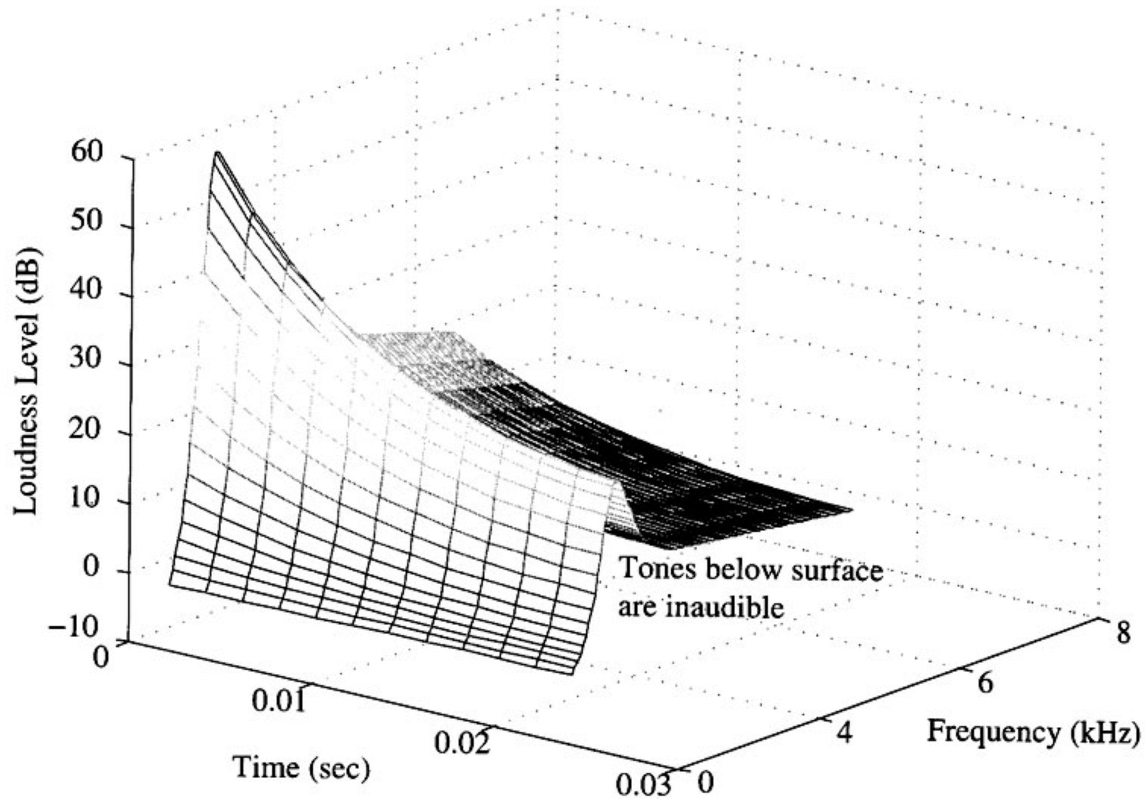


Temporal Masking (1)

- The frequency masking mentioned before assumes the masking takes effect when all frequencies are played simultaneously
- Nearby frequencies may be *temporarily* masked even the masking tone is turned off. This is known as **temporal masking**
- Under loud tone, hearing receptors in our inner ears become saturated and require time to recover

Temporal Masking (2)

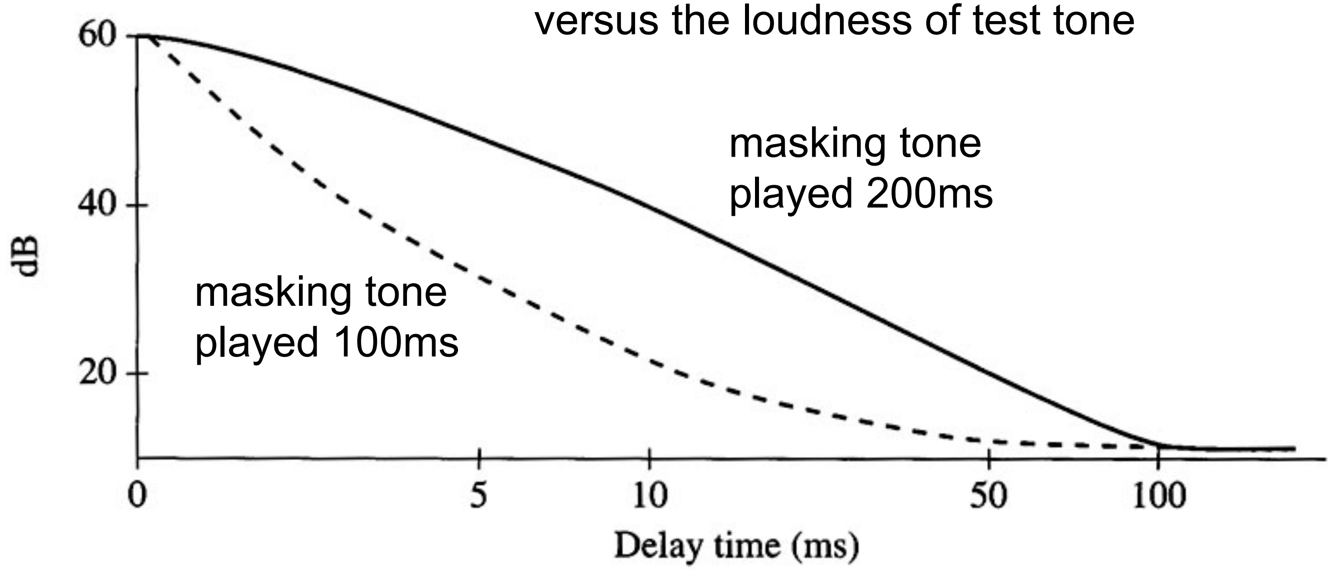
- Therefore, we actually have masking surfaces rather than masking curves
- Again, the masking surface changes as the masking tone changes



Temporal Masking (3)

- Besides **post-masking**, there exists **pre-masking**
- Pre-masking means that frequencies may also be masked out just before the stronger masking tone is played
- If the masking tone is played longer, it takes longer time to recover

The curves show the time needed to recover versus the loudness of test tone



Audio Compression

- Modern audio compression method such as MP3, Dolby AC-3, and Sony ATRAC (MDLP) utilize these limitations to achieve high compression ratio

- However, we shall defer the discussion of the compression method/standard to later chapter, because we need more compression tools to ease our discussion

Audio Encoding Standards

- Telephone-quality audio uses 8-bit logarithmic quantization at 8000Hz sampling rate.
- CD-quality music/audio is sampled at 44.1 KHz with 16-bit PCM quantization. (PCM to be discussed later).
- Popular audio file formats: .au (SUN), .aiff (MAC, SGI), .wav (PC), .mp3 (internet), .ra (RealAudio)

	Samp. rate (KHz)	bps	# of Channe	Date Rate	Freq Band
Telephone	8.00	8	Mono	8.0 KB/s	200-3400Hz
AM Radio	11.03	8	Mono	11.0 KB/s	
FM Radio	22.05	16	Stereo	88.2 KB/s	
CD	44.10	16	Stereo	176.4 KB/s	20-20000 Hz
DAT	48.00	16	Stereo	192.0 KB/s	20-20000 Hz

MIDI

- Is there any better (more compact) way to represent the audio?
- For music, why not store the notes in the song, instead of sampling the sound wave.
- During playback, synthesize the music in real time.
- MIDI (Musical Instrument Digital Interface)
- Industrial standard defines the interface between computer and electronic musical instruments.
- Allows record, playback, synchronization and communication of sound-producing devices.

MIDI (2)

- MIDI defines the
  - Hardware
  - Data format (format of messages between computer and musical instrument). No audio sample is included. It encodes the notes (128 notes or 10 octaves).
- When a musician presses a piano key, a MIDI message. The message indicates the beginning of note and encodes the stroke intensity.
- Device communicates with other devices through channels. There are altogether 16 musical channels, one for each instrument.
- It identifies 128 instruments, e.g. flute, violin, …

MIDI (3)

Two main MIDI devices

- Synthesizer
  - sound generator, e.g. electronic guitar
  - includes microprocessor, keyboard, control panel, memory, etc
  - Some sound cards store the wave table of different instruments. In other words, the MIDI synthesizers are built into the sound card.
- Sequencer
  - Storage server for MIDI data.
  - Stand-alone unit or a software program
  - Nowadays, it is usually a software on computer
- For further information, you can find a MIDI FAQ course homepage.