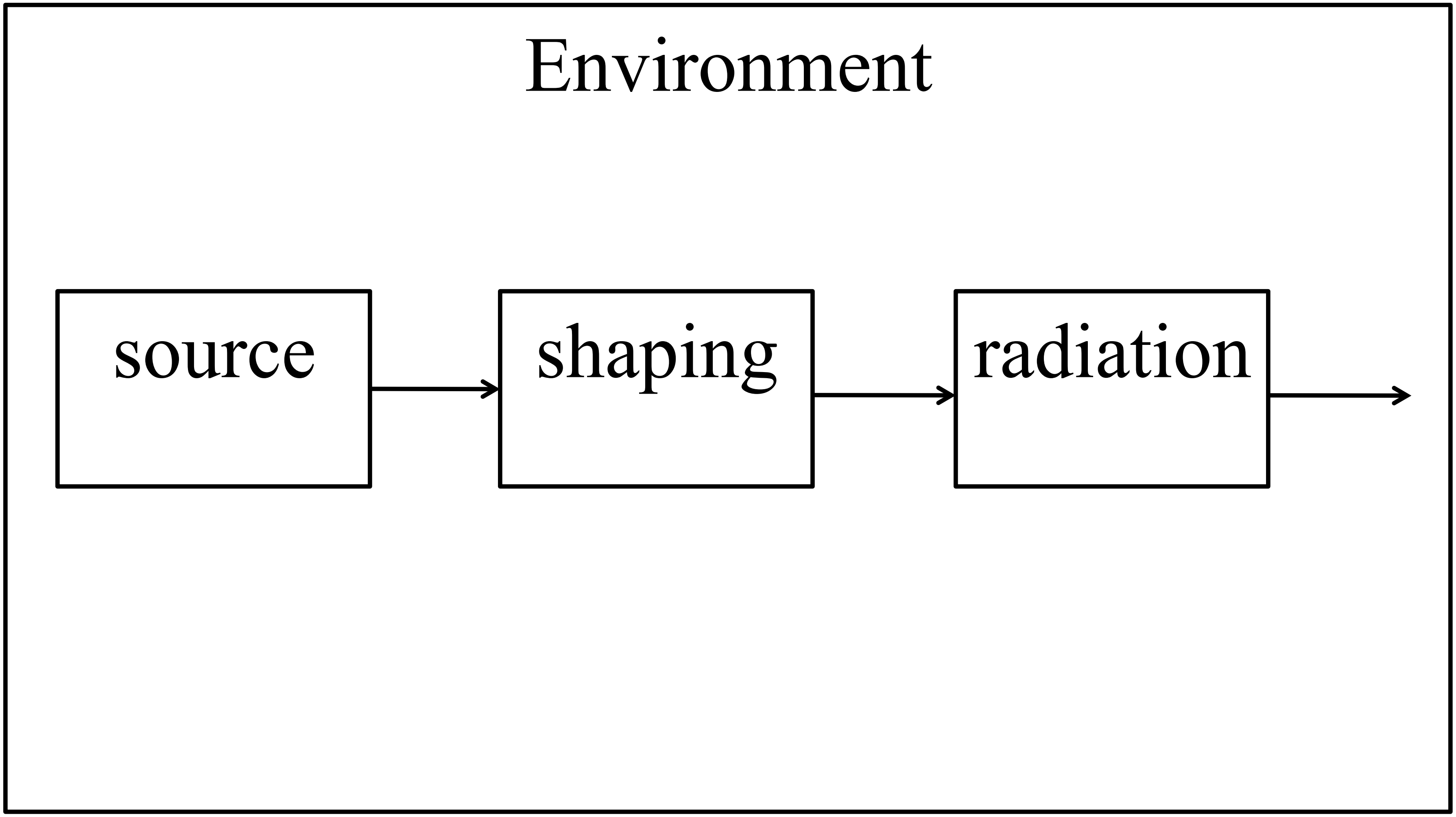
Last time: “small acoustics”

- Voice, many instruments, modeled by tubes
- Traveling waves in both directions yield standing waves
- Standing waves correspond to resonances
- Variations from the idealization give the variety of speech sounds, musical timbre

This time: “large (room) acoustics”

- Resonances still a factor
- Other points about time domain characteristics
- Wave propagation, energy
- Some properties of hearing


Environment
source
shaping
radiation

Spherical wave equation

- In polar coordinates:

$$(\delta^2 p/\delta r^2) + 2/r\,(\delta p/\delta r) = 1/c^2\,(\delta^2 p/\delta t^2)$$

- A solution is:

$$p(r, t) = P_0 \exp[j(\omega t - kr)]/r$$

$$= P_0 \exp[j(2\pi/\lambda)(ct - r)]/r$$

Where $k = 2\pi/\lambda$ and $\omega = kc = 2\pi c/\lambda$

(k is the *wavenumber*, λ is the *wavelength*)

So p is inversely proportional to r

Sound waves

- $c = (331.4 + 0.6T)$ m/s, or 343.4 m/s at 20° C
- 1127 ft/s
- A little less than 1 ms per foot
- Wavelength $\lambda = c/f$
- Actual velocity also depends on humidity

Intensity

- Definition: sound energy flowing across a unit area surface in a second

  $I = p^2 / (\rho_0 c)$

- Where p is the average pressure, ρ_0 is the medium density, c is the speed of sound (as before), and $\rho_0 c$ is the characteristic impedance.

- So, for spherical wave, I is inversely proportional to r^2
  (think of surface area for sphere)

Sinusoidal Intensity

- Let P_0 be the maximum pressure for a sinusoidal wave. Then

  $p^2 = P_0^2 / 2$

And

$I = P_0^2 / 2(\rho_0 c)$

dB sound levels

- Level in dB = $10 \log_{10} I/I_{ref} = 20 \log_{10} p/p_{ref}$

- We choose ref values to correspond to typical threshold of hearing at 1 kHz

  $I_{ref} = 10^{-12}$ Watts/m^2 $p_{ref} = 2 \times 10^{-5}$ Newtons/m^2

- And the corresponding dB levels are Intensity Level (IL) and Sound Pressure Level (SPL)

SPL for typical sound sources

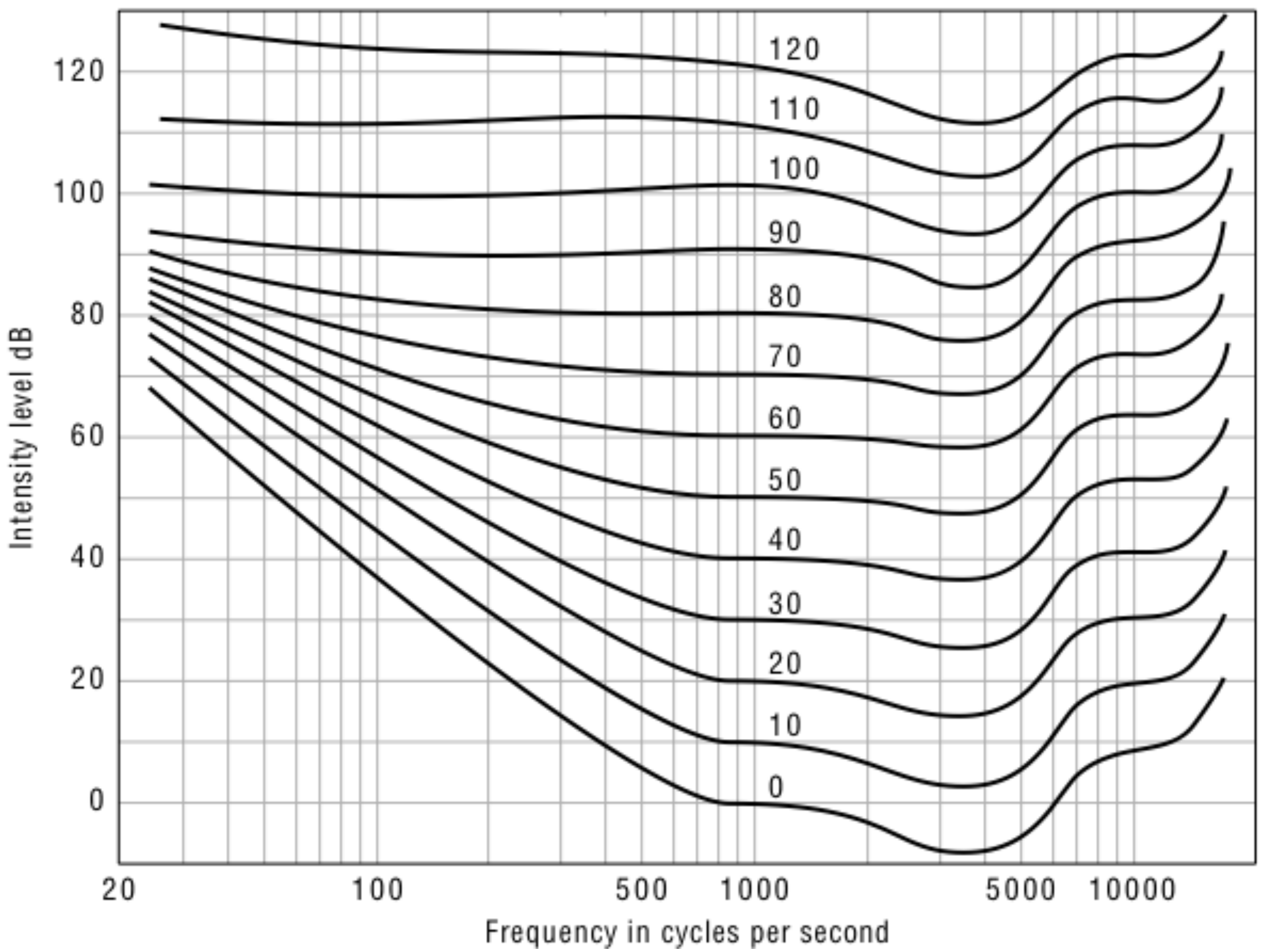
16 inch distant, hemispherical radiation,
about 1m^2 surface area

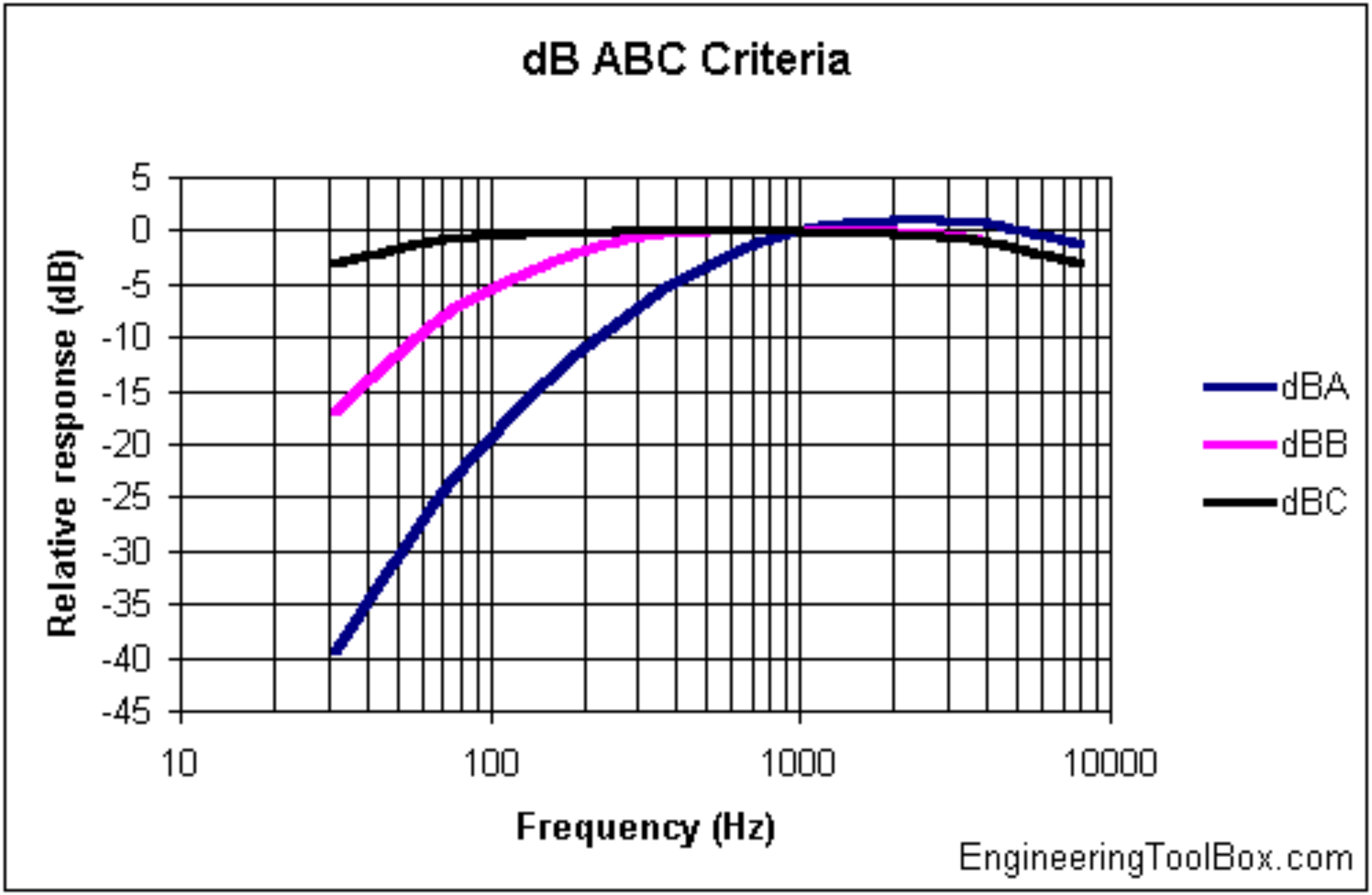
- Whispered speech: 1 nW power, 30 dB SPL
- Average speech: 10 μW power, 70 dB SPL
- Loud speech: 200 μ power, 83 dB SPL
- Shouting: 1 mW power, 90 dB SPL
- Ignoring boundaries, SPL would be 6 dB lower for every doubling of distance

SPL is not loudness

- Psychoacoustics -> cube root approximation
- Frequency dependencies
- Weighting curves for measurement

Equal Loudness curves





dB ABC Criteria
Relative response (dB)
5
0
-5
-10
-15
-20
-25
-30
-35
-40
-45
10
100
1000
10000
Frequency (Hz)
dBA
dBB
dBC
EngineeringToolBox.com

Room modes

- Same math as for strings, tubes
- Standing waves at resonant frequencies
- Mostly a significant issue at low frequencies (given the large dimensions of rooms)
- Essentially a continuum at high frequencies

Acoustic reverberation

- Reflection vs absorption at room surfaces
- Effects tend to be more important than room modes for speech intelligibility
- Also very important for musical clarity, tone

Energy growth

- Assume sound source (e.g., noise) with power W turns on at $t = 0$
- Assume it is uniformly distributed and diffuse
- Let $\overline{\alpha}$ be the average absorption, S be the surface area in the room, and $S\overline{\alpha}$ be the total absorption
- $S\overline{\alpha} = s_1\alpha_1 + s_2\alpha_2 + \ldots$
- Intensity = $(W/S\overline{\alpha})(1 - \exp[-t/\tau])$ where

$\tau = 4V / cS\overline{\alpha}$ (V is the room volume, c is speed of sound)

Steady state intensity for large t

Intensity = Power/S $\overline{\alpha}$

= Power / Total absorption

Reverberation time

- When steady state signal is turned off, overall energy decays roughly exponentially
- Time to decay by 60 dB from steady state after turnoff = RT60
- Sabine formula, RT60 = 0.049V/S$\bar{\alpha}$ in feet,
- RT60 = 0. 163V/S$\bar{\alpha}$ in meters
- Frequency dependent

Air effects

- Additional denominator term
- RT60 = $0.049V/(S\bar{\alpha} + 4\,mV)$
- Air effects dominate at very high frequencies, negligible at low frequencies

Example: 237 Cory Hall

- Dimensions: 20 x 24 x 16 (feet)
- Volume: 7680 ft^3
- Surface area: 2368 ft^2
- Mid-freq RT60: 1.2 sec (empty)
- We infer average $\overline{\alpha}$ of .13
- Air absorption less than 10% for 1 kHz, 50% for 4 kHz

Steady state, 237 Cory

- For a 10 μW source,

Intensity = Power/($S\bar{\alpha}$) = 3.2×10^{-8} W/ft^2
= 3.5×10^{-7} W/m^2

- This is about 55 dB IL, and 120 ms after cutoff, the IL will be about 49 dB. Will this effect interfere with intelligibility?

Losses at boundaries

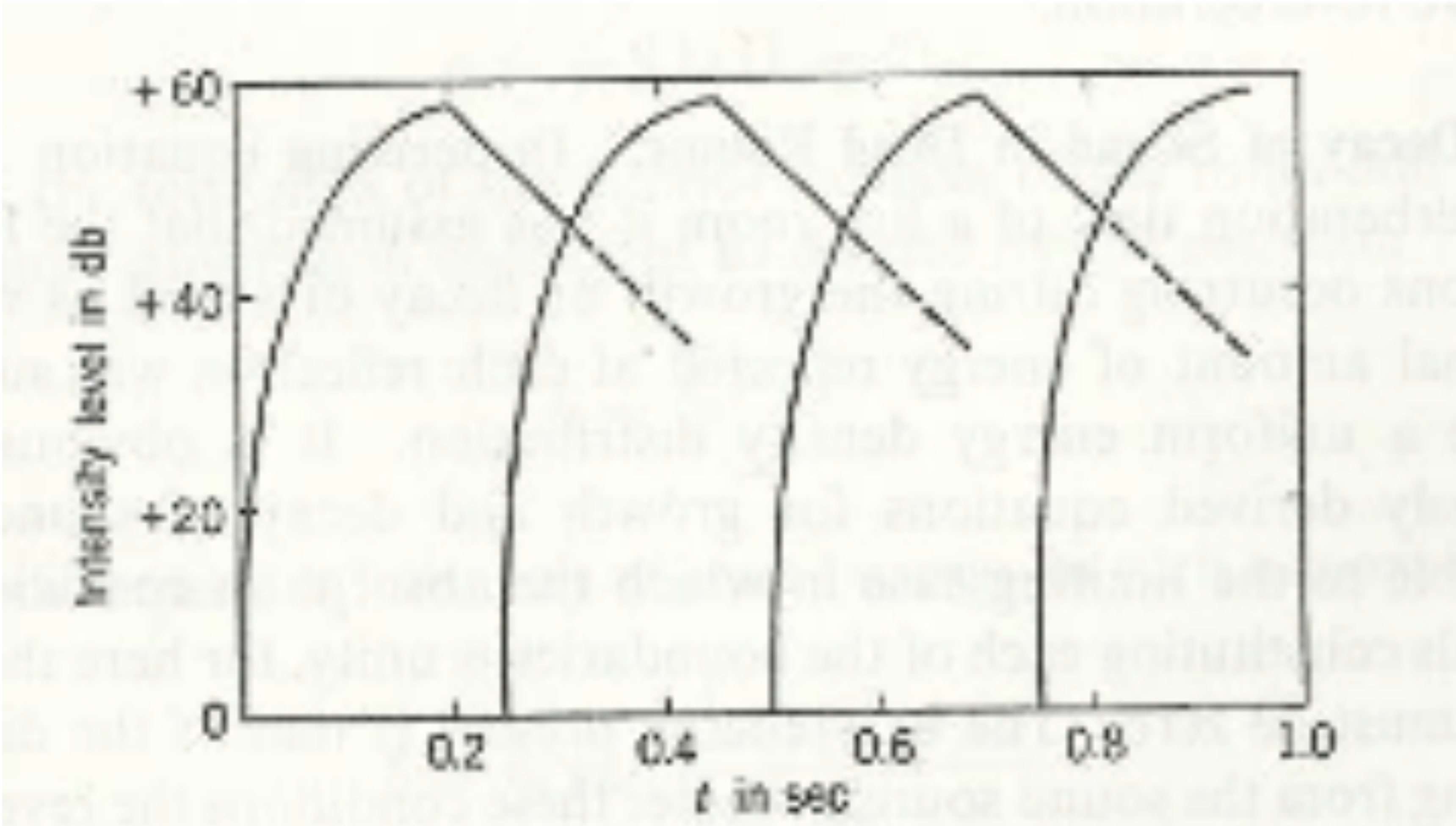
- Sound energy reaching boundary can do 3 things:
  - Reflect
  - Transform into heat
  - Pass through boundary
- The energy fraction for the last two constitutes the absorption coefficient $\overline{\alpha}$

Absorption coefficients, common building materials

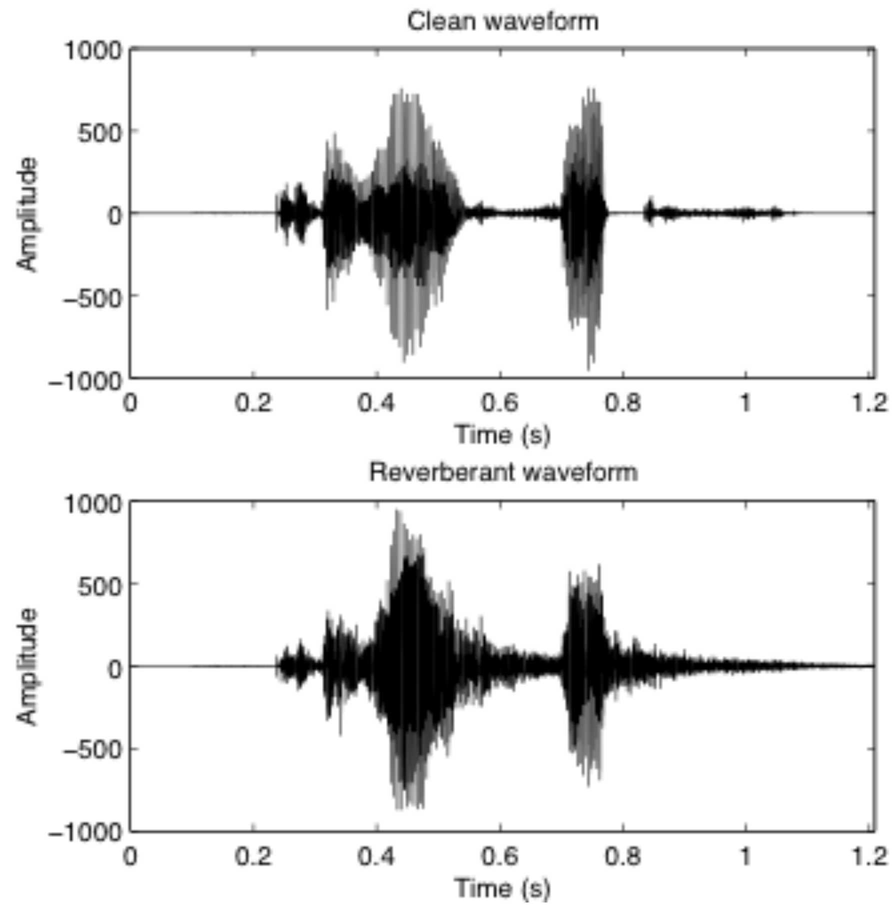
	125 Hz	500 Hz	2000 Hz
Material			
Acoustic paneling	0.16	0.50	0.80
Acoustic plaster	0.30	0.50	0.55
Brick wall, unpainted	0.02	0.03	0.05
Draperies, light	0.04	0.11	0.30
Draperies, heavy	0.10	0.50	0.82
Felt	0.13	0.56	0.65
Floor, concrete	0.01	0.02	0.02
Floor, wood	0.06	0.06	0.06
Floor, carpeted	0.11	0.37	0.27
Glass	0.04	0.05	0.05
Marble or glazed tile	0.01	0.01	0.02
Plaster	0.04	0.05	0.05
Wood paneling, pine	0.10	0.10	0.08

Effect on intelligibility

- Energy is larger than without reverberation
- Decay stretches sounds out in time
- Colorations due to frequency-dependent absorption
  - Spectral measures will be different



Intensity level in a “live” room as produced by successively spoke syllables



The phrase "two oh six" convolved with impulse response from .5 second RT60 room

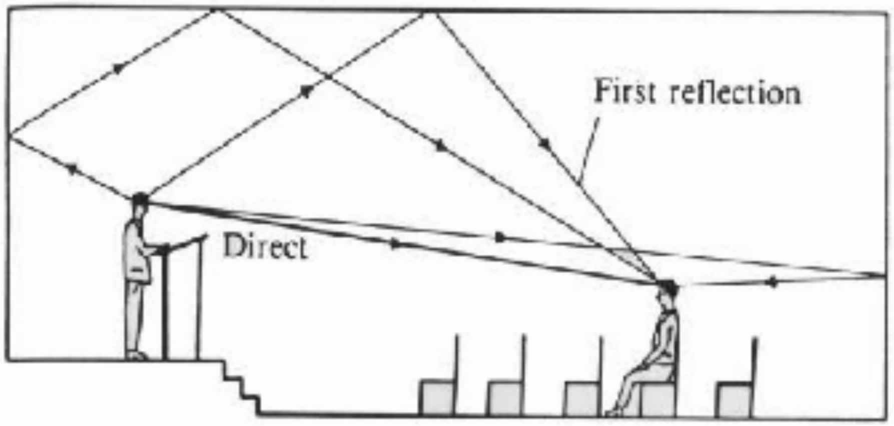
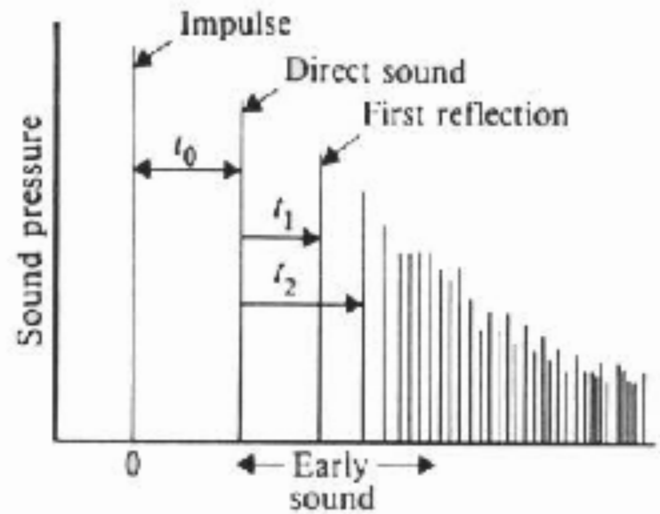


FIG. 23.3
Paths of direct and reflected sound from source to listener with corresponding time delays for a sound impulse. (From *Music, Acoustics, and Architecture.* © 1988 Leo Beranek. Used with author's permission.)



Initial time delay gap = t_1

Measuring room responses

- Impulsive sounds
- Correlation of mic input with random signal source (since $R(x,y) = R(x,x) * h(t)$)
- Chirp input
- Also includes mic, speaker responses
- No single room response (also not really linear)

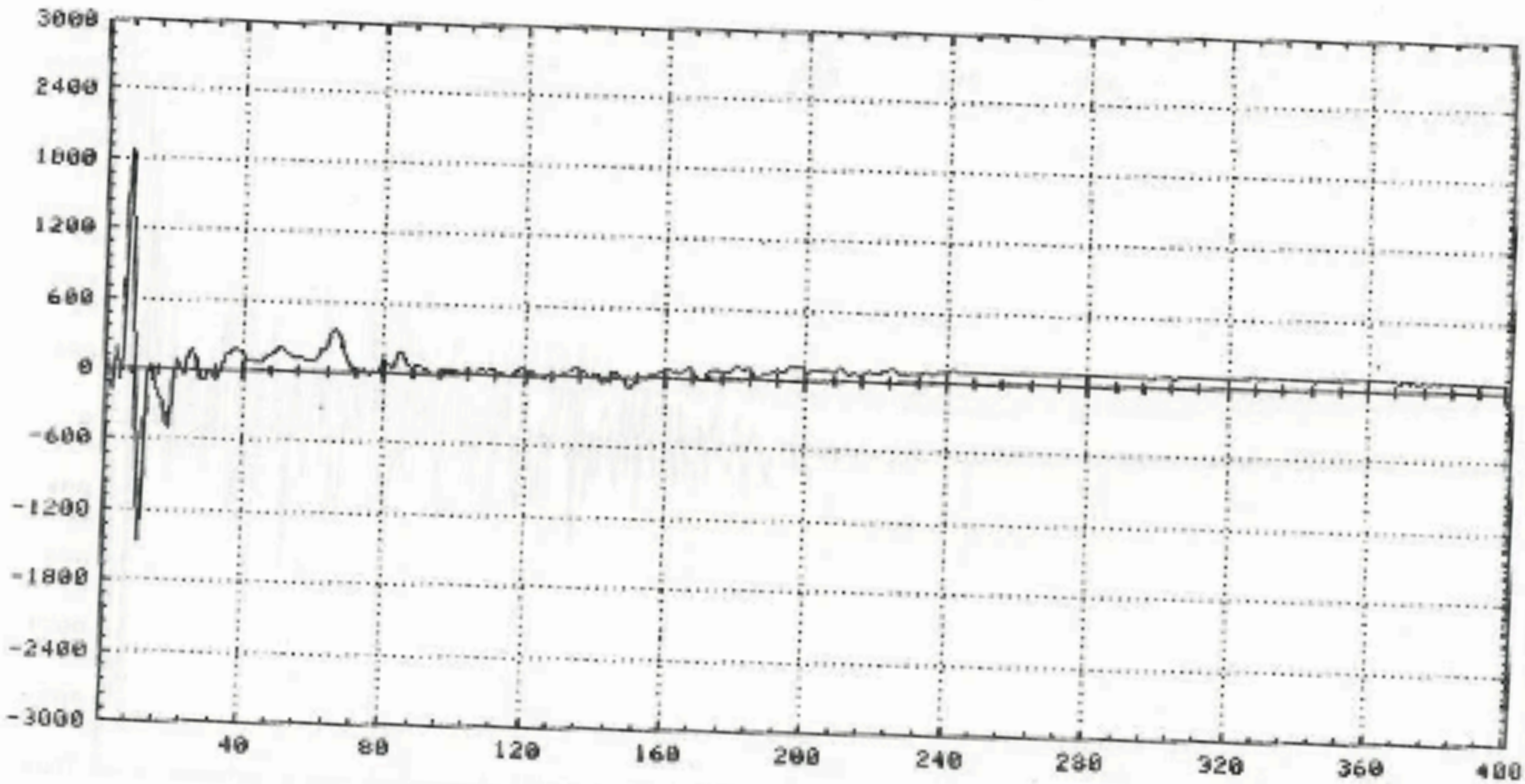


Figure 13: Anechoic direct impulse response, Rogers speaker

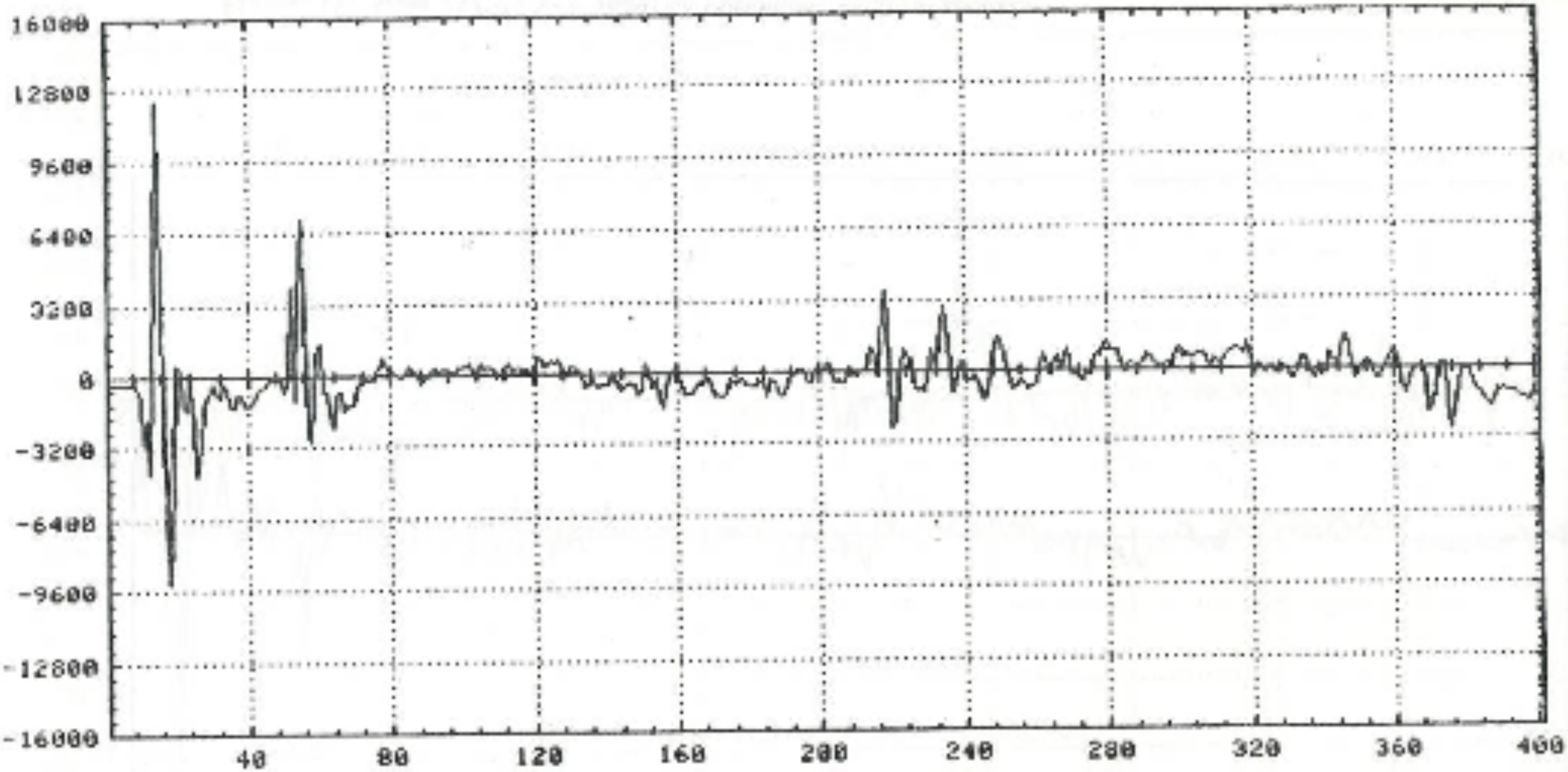


Figure 14: Room 237 direct impulse response, Rogers speaker

Effects of reverberation

- Increases loudness
- “Early” loudness increase helps intelligibility
- “Late” loudness increase hurts intelligibility
- When noise is present, ill effects compounded
- Even worse for machine algorithms

Effects of reverberation (2)

- Reverb acts like a smearing of the spectrogram
- "Spectrum of the short-term spectral components" is the modulation spectrum
- Reverberation acts like a low-pass filter on the modulation spectrum
- Treatment of the modulation spectrum can help ASR

Effects of reverberation: Word error rate for ASR

	Close-mic'ed speech	**RT60 = .55s, distance mic**
Human	0.3%	0.3%
ASR	5.9%	22.2%
ASR w/ treatment	4.7%	13.0%

Numbers '95 corpus;
"treatment" corresponds to modulation spectral features

Dealing with reverberation

- Microphone arrays - beamforming
- Reducing effects by subtraction/filtering
- Stereo mic transfer function
- Using robust features (for ASR especially)
- Statistical adaptation

Artificial reverberation

- Physical devices (springs, plate, etc.)
- Simple electronic delay with feedback
- FIR for early delays (think of "initial time delay gap" in concert halls), IIR for later decay
- Explicit convolution with stored response

Some factors affecting concert hall design (Beranek)

- Reverberation time
- Envelopment (diffusion)
- Clarity

Reverberation time

- 500Hz/1 kHz average,
  RT60, ~2 s for concert halls
- ~1.4 s for opera houses
- want ~80-90% at 2 kHz/4 kHz (“brilliance”)
- want 110-125% at 125/250 Hz (“warmth”)
- Based on subjective assessments of concert masters

Envelopment

- Impression of sound surrounding you
- Requires adequate diffusion (reflections everywhere)

Clarity

- Ratio of energy in early sound to later sound C_{80} corresponds to first 80 ms after direct;

  typically want 0 to -4 dB

- Typically want a short initial time delay gap

  e.g., 15 ms or less

Some “Large Acoustics” Conclusions

- Frequency-domain resonances only a part of the effects; time-domain reflections important
- Short-time reverberation aids in intelligibility, affects spectrum, related to room geometry
- Long-time reverberation acts like exponential decay of signals, can yield overlapping speech
- Factors beyond reverberation time are important for high quality acoustic spaces

What’s next:

- Homework due Wednesday on chap. 9, 10, 14 questions :

  Provide answers to problems (9.2, 9.4, 10.1, 10.5, 14.4, 14.5) and answer the question: "Describe phase locking in the auditory nerve. Over roughly what frequencies does this take place? "

- Wednesday talk is by Prof. Keith Johnson (linguistics) on speech sound categories