Perceptual Dimensions of Sound

Sound Perception

- Physical dimensions:
  - Aspects of a physical stimulus that can be measured with an instrument (e.g., a light meter, a sound level meter, a spectrum analyzer, a frequency meter, etc.)
- Perceptual dimensions:
  - Mental experiences that occur inside the mind of the observer. Perceptual dimensions can be measured, but not with a meter. Measuring perceptual dimensions requires an observer (e.g., a listener).

Sound Perception

Auditory Perception

Physical Properties of Sound	Perceptual Dimensions
Intensity	Loudness
Frequency	Pitch
Complexity (frequency content & time)	Timbre

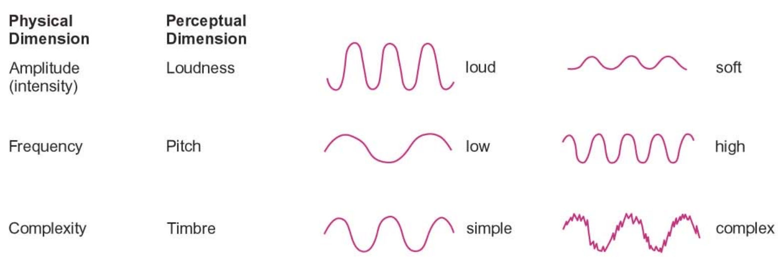
Visual Perception [analogy]

Physical Properties of Light	Perceptual Dimensions
Wavelength	Hue
Luminance	Brightness
Contour/Contrast	Shape

Sound Perception

Auditory Perception

Physical Properties of Sound	Perceptual Dimensions
Intensity	Loudness
Frequency	Pitch
Complexity (frequency content & time)	Timbre



PERCEPTUAL DIMENSION: LOUDNESS

77

77

Recall: Sound Pressure Level (SPL) in Decibels (dB)

- Sound Pressure Level (SPL)

Intensity of target sound

$$\boxed{L = 10log_{10}\frac{I}{I_0}}$$

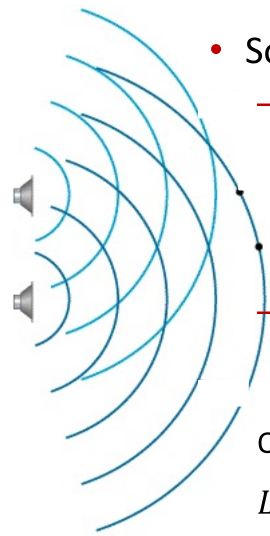
Reference Intensity:
Human absolute hearing threshold
$I_0 = 10 \text{ x } 10^{-12}\ (W/m^2)$
(related to $P_0 = 20\ \mu Pa$)

  - SPL is technically a unitless measure; but uses unit of Decibels or dB or dB-SPL
  - Decibels provide a **relative** measure of sound intensity.

78

78

Adding SPL from multiple sources



- Sounds from multiple sources:
  - Intensity is additive

$$I_{total} = I_1 + I_2 + \cdots + I_N$$

(Recall: intensity is power/area;
so sum powers from all sources)

  - SPL is not additive

$$L_{total} = 10log_{10}\frac{I_{total}}{I_0}$$

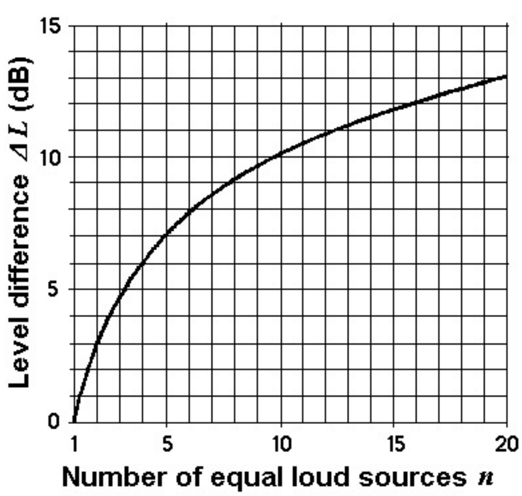
Or:

$$L_{total} = 10log_{10}\left(10^{\frac{L_1}{10}} + 10^{\frac{L_2}{10}} + \cdots + 10^{\frac{L_N}{10}}\right)$$

79

79

Adding SPL from EQUAL sources



Level increase ΔL for n equal loud sound sources	
Number of n equal loud sound sources	Level increase ΔL in dB
1	0
2	3.0
3	4.8
4	6.0
5	7.0
6	7.8
7	8.5
8	9.0
9	9.5
10	10.0
12	10.8
16	12.0
20	13.0

Level difference → $\boxed{\Delta L = 10log_{10}n}$ ← Number of sources (equally loud)

80

Loudness scaling

- Let’s get a feel for the decibel scale
- listen to 3 noise sounds at 3 levels:
  - regular, loud and soft
  - Any guesses about relative dB relationship between them?
    - 0dB, +20dB, -20dB

81

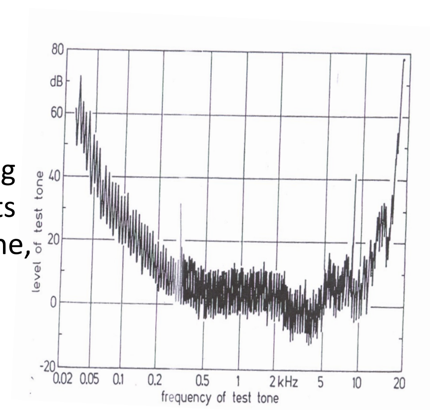
81

Threshold of hearing

- Measuring threshold of hearing:

  The listener pushes a button as long as s/he can hear a test tone and lets go when s/he stops hearing the tone,

  [the level goes up and down around threshold]

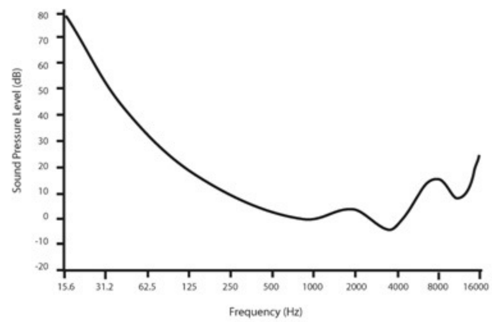


82

82

Threshold of hearing

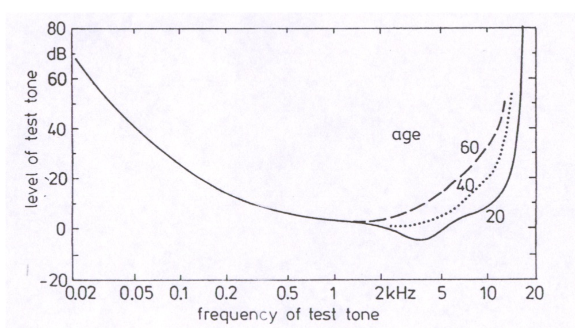
- The threshold of hearing varies with frequency
- The maximum sensitivity at about 3500 to 4000 Hz is related to the resonance of the auditory canal.



83

83

Threshold of hearing



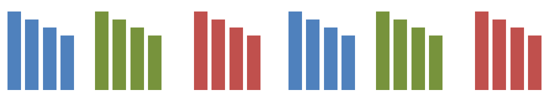
- Hearing sensitivity decreases with age especially at high frequencies
- Presbycusis: hearing loss because of age

84

84

Loudness across frequencies

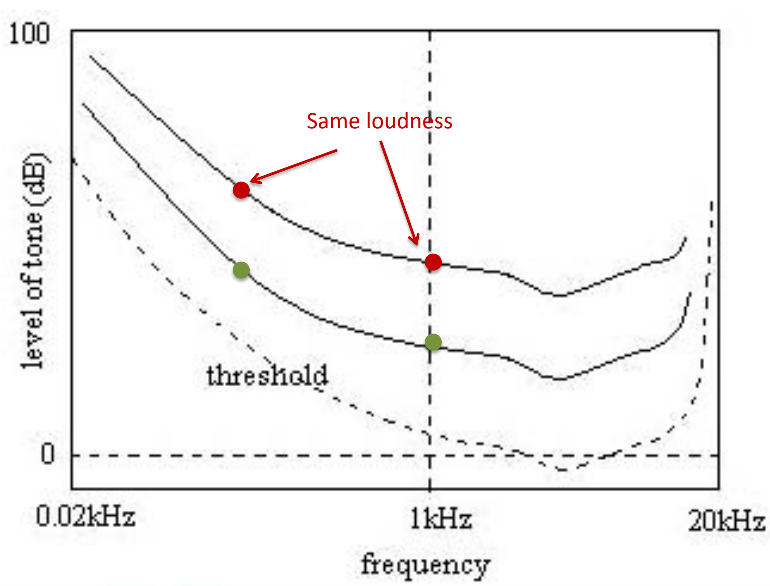
- Let's assess loudness across frequency
  1. Adjust the calibration tone to be just audible
  2. You will hear tones with several frequencies (125, 250, 500, 1000, 2000, 4000, and 8000 Hz) in 10 decreasing steps of **5dB** each
  3. Count the number of steps you hear at each frequency
  4. Frequency staircases are presented twice



85

85

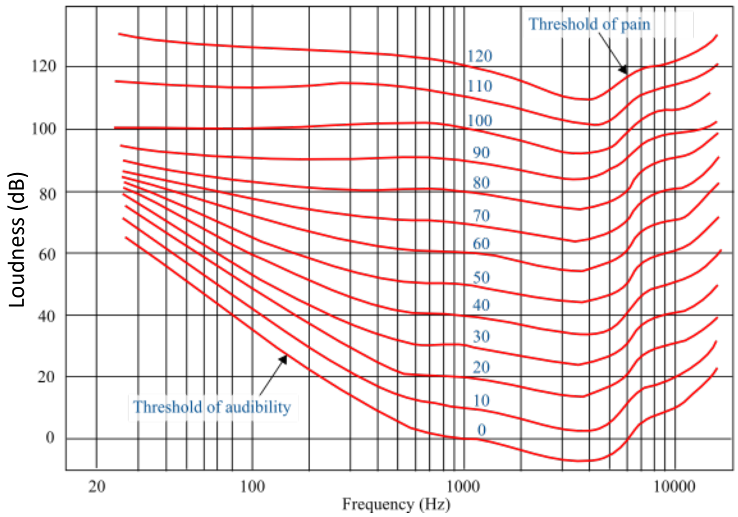
Equal Loudness Curves



86

86

Equal Loudness Curves/Contours



Each contour represents equally-perceived tones

87

87

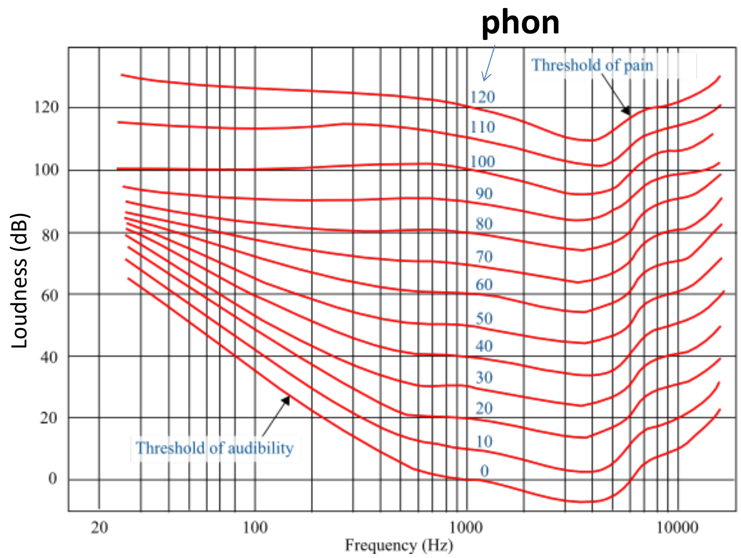
Equal Loudness Curves/Contours

- Mid-frequency signals (in the range from ~1000-4000 Hz) are louder than lower or higher frequency signals.
- 3 tones at equal: 125 Hz, 3000 Hz, 8000 Hz

- The 3000 Hz signal is perceived as louder than the 125 or the 8000 signal

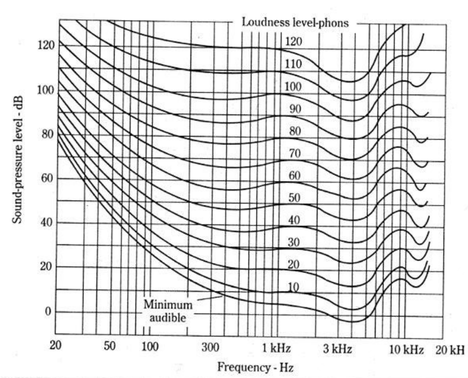
88

Equal Loudness Curves/Contours



Phon scale

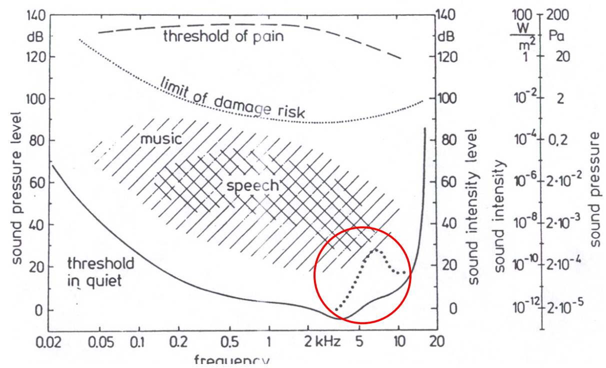
- Sounds judged to have equal loudness are assigned the same "phon" value
  - (e.g., all tones judged as having the same loudness as a 20 dB 1KHz tone have a loudness of 20 phons).



Loudness

- Hearing area is the area between the threshold in quiet and the threshold of pain.

  Note: Shift in threshold of quiet for those who listen to loud music

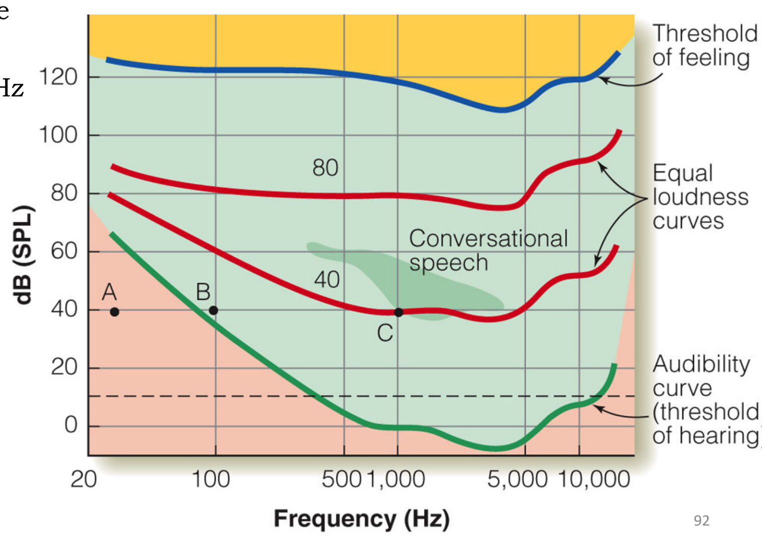


91

91

Range of Human Hearing

Human hearing range – 20 to 20,000 Hz

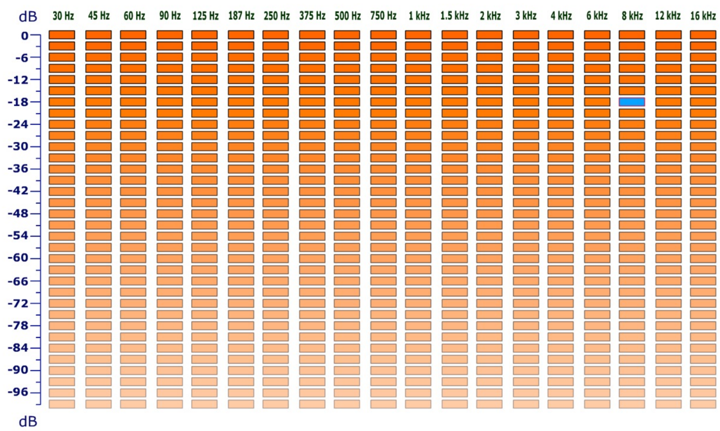


92

92

Plot your own equal loudness curves

www.phys.unsw.edu.au/~jw/hearing.html



93

93

Loudness of complex sounds

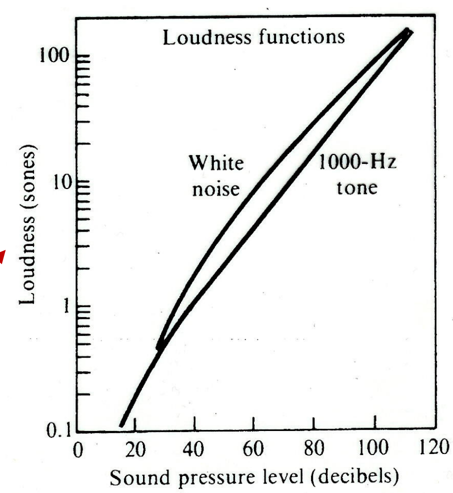
- Loudness depends mainly on sound intensity
- But is is also sensitive to other attributes
  - Frequency → equal loudness curves
  - Spectrum
  - duration

94

94

Loudness and spectrum/bandwidth

- Loudness of a broadband noise is greater than that of a pure tone



Note:
Sones are another unit of loudness on a linear scale

$$S \cong 2^{\frac{L-40}{10}}$$

Doubling 'perceived loudness' doubles sones

95

95

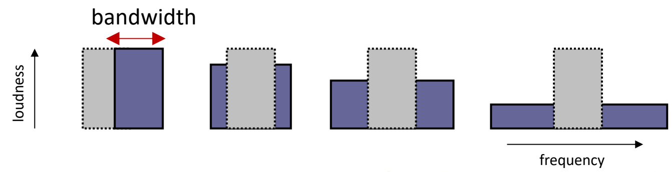
Critical band

- A critical band is a frequency region over which energy is integrated
  - Idea is:
    - If two frequencies are close enough to each (within a critical band), they activate *same* region in the ear; so the sound is not perceived as loud.
    - If they are far apart (larger than a critical band), a new region in the ear is activated making the sound seem louder

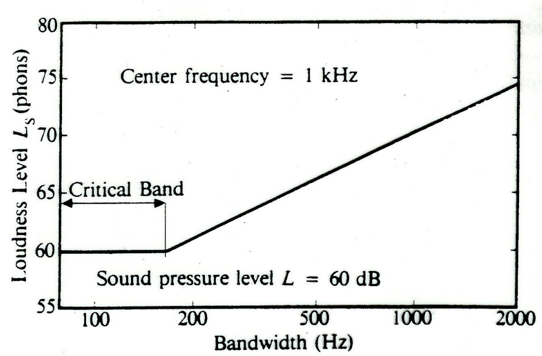
96

96

Critical band

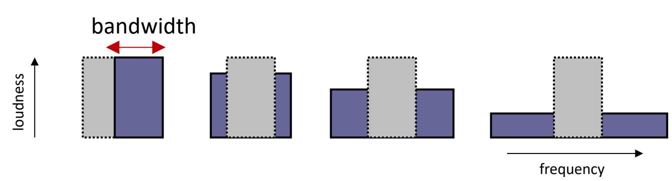


- A reference noise band compared to test noise band with increasing bandwidth (constant power).
- When the bandwidth of the test noise exceeds a certain level (**critical band**), the loudness begins to increase.



97

Critical band



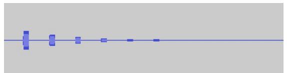
<u>**Demo:**</u>

- A reference noise will be followed by a test noise with the same center frequency
- The bandwidth of the test band is initially the same as the reference
- then increased in 7 steps of 15% each
- the amplitude is decreased to keep the power constant.

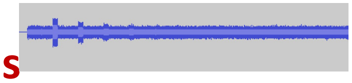
When the bandwidth exceeds the critical bandwidth, the loudness begins to increase.

98

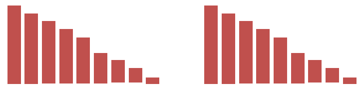
98



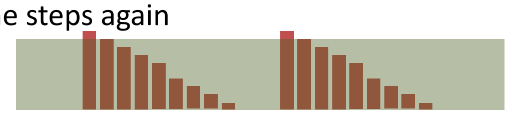
Critical bands



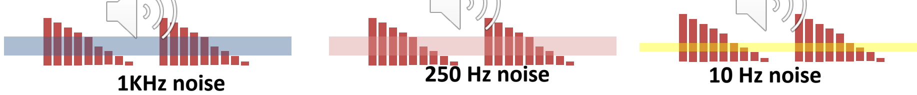
- You will hear a 2k tone in ten decreasing steps of 5 dB. Count how many you can hear (presented twice)



- Now the signal is masked with broadband noise – count the steps again



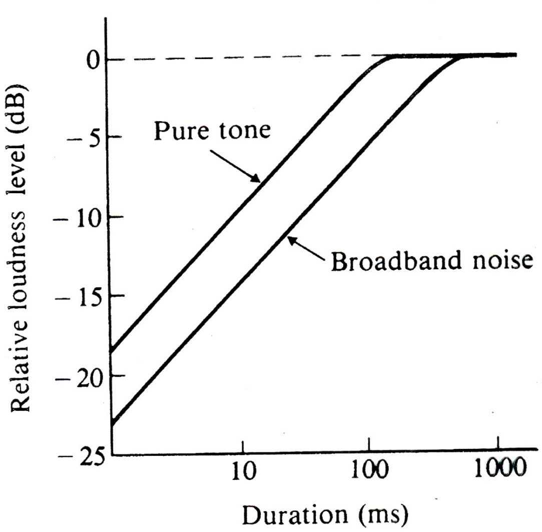
- Now, we will change the bandwidth of the noise



The critical band around 2KHz is ~280Hz so we hear more steps for 250Hz and10Hz noise

99

HOW DOES LOUDNESS DEPEND ON DURATION?



As sound gets longer (up to a given duration ~100’s ms), it sounds louder

100

Loudness vs. time

Demo:

- The level of a broadband noise sample decreases in 8 steps for several signal durations.
  - Durations: 1000, 300, 100, 30, 10,3, and 1 ms
- Count the number of steps you hear in each case.
- Each 8 step sequence is presented twice
- Note: there is a background noise than is running throughout the entire sequence
  - focus on the impulsive noise bust

As the duration gets shorter, you will hear fewer steps

101

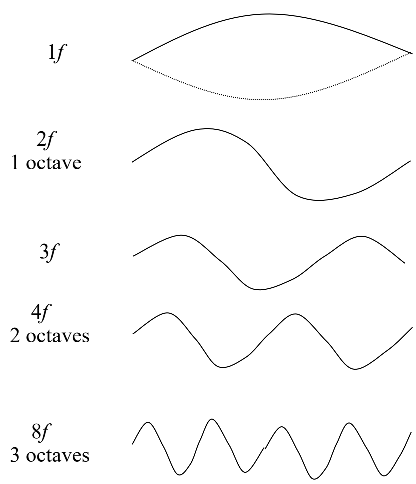
101

PERCEPTUAL DIMENSION: PITCH

102

102

Perceptual Dimensions: Pitch



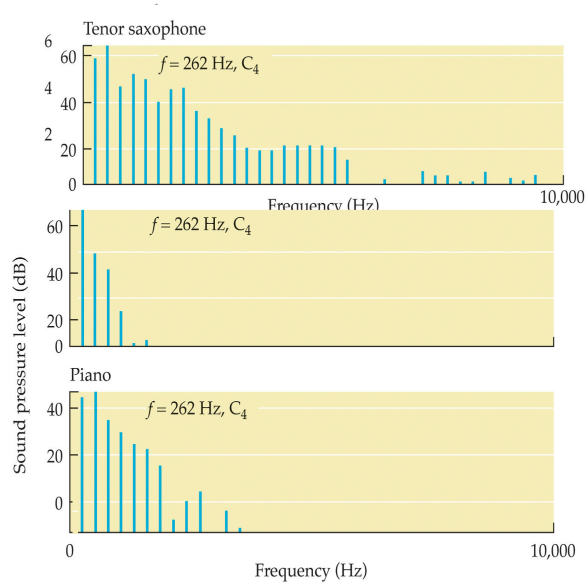
- At first approximation, the pitch of a simple periodic signal is determined by its frequency.
- Most oscillators (guitar string, vocal chords) naturally oscillate at a fundamental frequency (F_0) as well as its integer multiples (called harmonics/partials/overtones).
- The pitch of a *complex* period signal is often determined by its *fundamental* frequency (F_0)

103

103

Fundamental Frequency & Harmonics

- Some harmonics may be damped or enhanced.
- The fundamental frequency (F_0) determines the musical note being played (C4).
- F_0 is the fundamental frequency of the periodic waveform



104

Pitch Perception

The ear is more sensitive to F_0 differences in the low frequencies than the higher frequencies. This means that:

300 vs. 350 $\neq$ 3000 vs. 3050

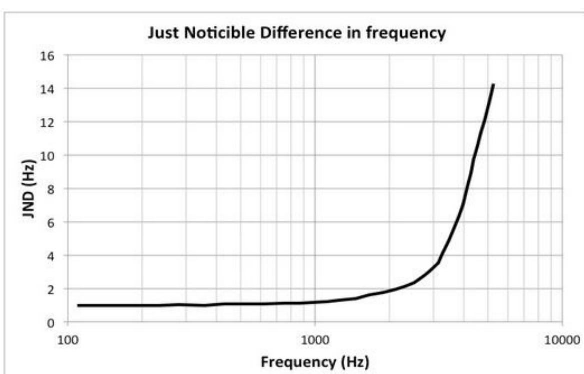
That is, the difference in ***perceived pitch*** (not F_0) between 300 and 350 Hz is **NOT** the same as the difference in pitch between 3000 and 3050 Hz, even though the ***physical*** differences in F_0 are the same.

300-350: **3000-3050:**

105

105

Pitch Perception



- Pitch is non-linearly related to frequency.
- Pitch Just-noticeable difference (JND) is threshold at which a frq change is perceived => it depends on the tone's frequency
- We are more sensitive to pitch changes at low frequency.

106

106

Measuring JND

- Difference limen (DL) or just noticeable difference (JND)
  - Two stimuli cannot be consistently distinguished from one another if they differ by less than a JND
- **Demo**
  - 10 groups of 4 tone pairs are presented
  - Ref. tone is 1000Hz

Group	ΔF	Key			
1	10Hz	$+\Delta F$	$-\Delta F$	$+\Delta F$	$+\Delta F$
2	9Hz	$+\Delta F$	$-\Delta F$	$-\Delta F$	$-\Delta F$
3	8Hz	$-\Delta F$	$+\Delta F$	$+\Delta F$	$-\Delta F$
4	7Hz	$-\Delta F$	$+\Delta F$	$+\Delta F$	$-\Delta F$
5	6Hz	$+\Delta F$	$-\Delta F$	$+\Delta F$	$-\Delta F$
6	5Hz	$+\Delta F$	$-\Delta F$	$+\Delta F$	$+\Delta F$
7	4Hz	$-\Delta F$	$-\Delta F$	$+\Delta F$	$+\Delta F$
8	3Hz	$+\Delta F$	$-\Delta F$	$+\Delta F$	$-\Delta F$
9	2Hz	$-\Delta F$	$-\Delta F$	$-\Delta F$	$+\Delta F$
10	1Hz	$-\Delta F$	$+\Delta F$	$+\Delta F$	$-\Delta F$

According to JND plot, we should hear steps up to about group 10

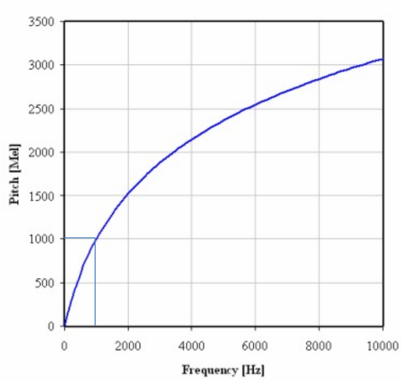
107

107

Pitch Perception – a Pitch scale

- Perceptual scale of pitch: mel scale
  - *How far in frequency do we have to be in order to feel a tone as doubled in pitch?*

It's a relative scale, based on pitch comparisons



$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

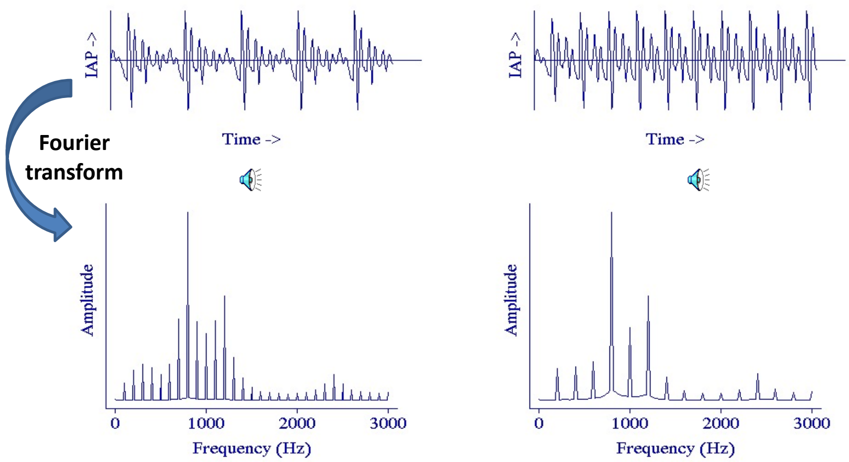
✓ Mel-scaling is used in signal processing to build filters that approximate human pitch perception (MFCC)

108

108

Pitch

All else being equal, the higher the F_0, the higher the perceived pitch.



Lower F_0, lower pitch **Higher F_0, higher pitch**

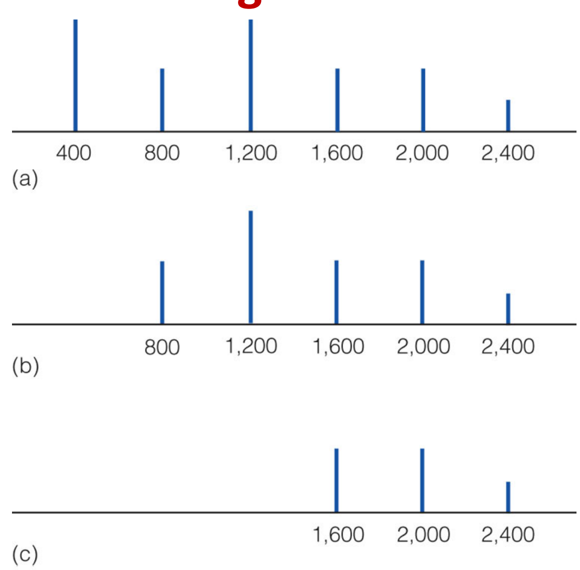
109

The effect of the missing fundamental

Removing the fundamental frequencies does not change the pitch percept (it can weaken it).

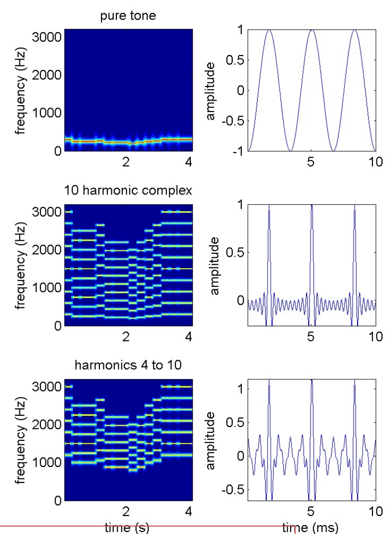
→ Called 'phantom F0'

The perception of the pitch of complex tones cannot be explained by the physical existence of the F0 frequency



Missing fundamental

- Listen to a melody produced with pure tones only
- Now same melody with harmonics 1-10
  - Pitch is the same, though sound quality (timbre) is different
- Now, we remove the fundamental freq, keep only harmonics 4-10



Pitch perception does not require presence of fundamental

111

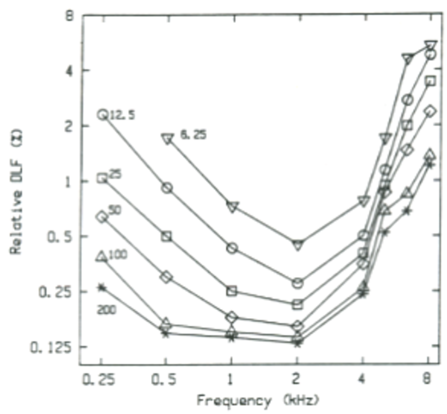
Perfect (absolute) Pitch

- Ability to identify the note name of a sound, or to reproduce a specific note, <u>without</u> reference to an external standard [relative pitch].
- One person in 10,000 claims to have it
- Absolute pitch (AP) is not necessarily "perfect"
  - people can identify 70-100% of midrange tones (chance level is 8.3%)
- Develops during early life; nature vs. nurture source hard to determine
- Absolute pitch can be developed through ear training
- Composers with AP include Mozart, Scriabin, Messiaen, Boulez
- AP can be a negative(constant awareness of pitch labels hurts enjoyment of music)

112

Pitch and Duration

- Subjects can discriminate differences in frequency of 2 tones
- As tone duration is reduced from 200 ms (bottom curve) to 6.25 ms (top), performance falls off, especially for very low or very high frequencies.



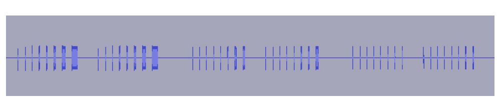
Effects of duration on pure-tone frequency discrimination
DLF (difference limen of frequency)

113

Pitch and time

Demo:

- You hear tones of 300, 1000, and 3000 Hz in bursts of 1, 2 4, 8, 16, 32, 64, and 128 periods. Each sequence of duration is heard twice.
  - How many periods are necessary to establish a sense of pitch?



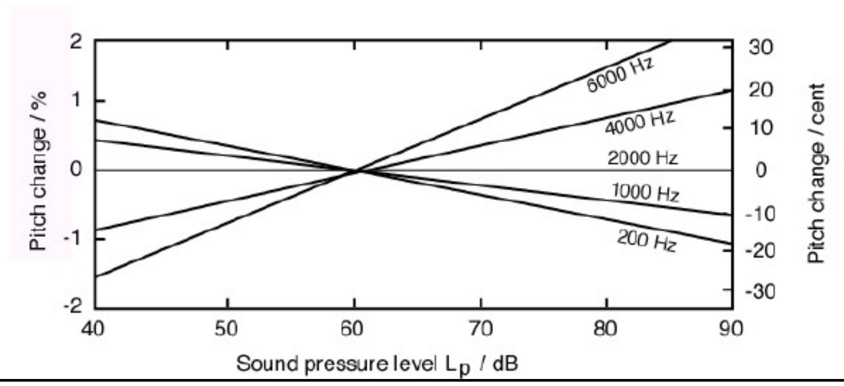
✓ Experiment in 1830s showed that a sense of pitch develops after only two cycles.
✓ Very brief tones are described as "clicks," but as the tones lengthen, the clicks take on a sense of pitch

114

114

Pitch and intensity

- Intensity affects pitch perception as well
  – High pitch tones (>2KHz) tend to be perceived higher in pitch with high intensity, lower with low intensity
  – Low pitch tones tend to be perceived lower in pitch with high intensity, higher with low intensity



115

Pitch and intensity

- First we have a calibration tone at 200Hz.
  – Adjust level so that it is just audible
- Now, 6 tone pairs are presented at various frequencies
- Compare the pitches for each tone pair
- Note: the frequency for each pair is fixed at 200, 500, 1000, 3000, and 4000 Hz

  ... But second tone is 30dB louder
- You will likely hear a slight **change** in pitch

116

116

Pitch in presence of noise

- The pitch of a tone is influenced by the presence of noise/tone near to it in frequency.
  - If interfering tone is lower in frequency, an upward shift in pitch of test tone
  - If interfering tone is higher in frequency, a downward shift is observed, at least at low frequency (< 300 Hz)
- Example
  - A 1000Hz note is presented alone vs. masked by noise
  - Compare the pitch differences between clear and masked

Masked tone appears slightly higher in pitch…
Even though they are identical

117

117

PERCEPTUAL DIMENSION: TIMBRE

118

118

Perceptual dimensions: Timbre

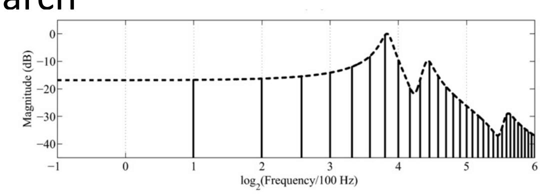
- Timbre, also known as sound quality or tone color, is oddly defined in terms of what it is not.
- ANSI Definition of timbre
  that attribute that allows us to distinguish between sounds having the same perceptual duration, loudness, and pitch, such as two different musical instruments playing exactly the same note
- Example of different sounds with unique sound quality (timbre)
  - ✓ a clarinet, a saxophone, and a piano all play a middle C at the same loudness and same duration.
  - ✓ two vowels (e.g., /a/ and /i/) spoken at the same loudness and same pitch differ from one another in timbre

119

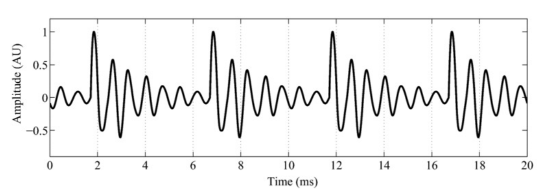
119

Timbre

- Physical attributes that relate to timbre are still a topic of research
  - Spectrum (spectral envelope)



  - Time signal (Amplitude envelope)

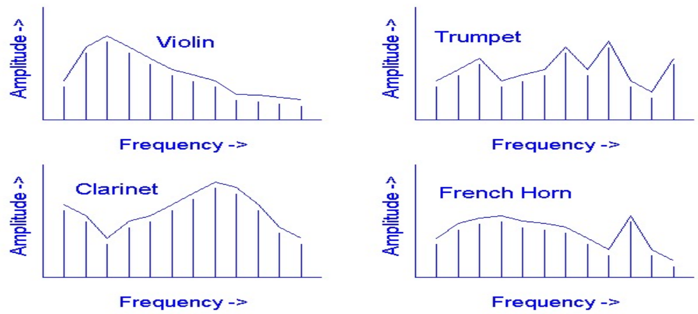


  - Likely BOTH time-freq

120

120

Timbre and spectral envelope



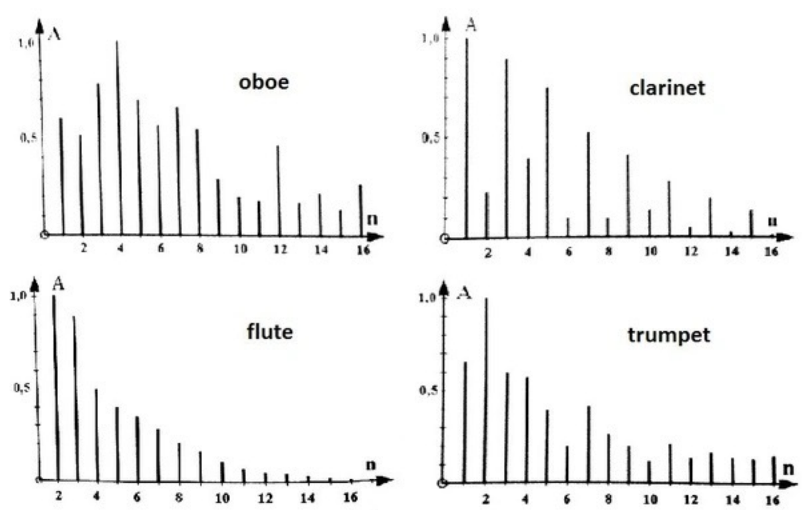
Timbre differences between one musical instrument and another are related to differences in spectral envelope

- i.e. differences in the relative amplitudes of the individual harmonics

121

121

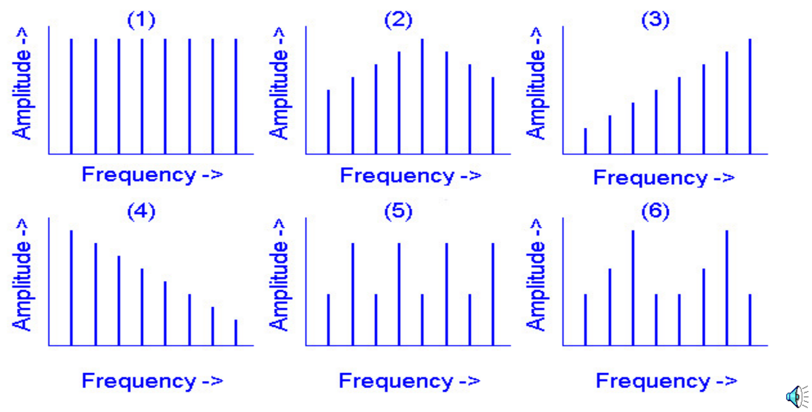
Timbre and spectral envelope



122

122

Timbre and spectral envelope



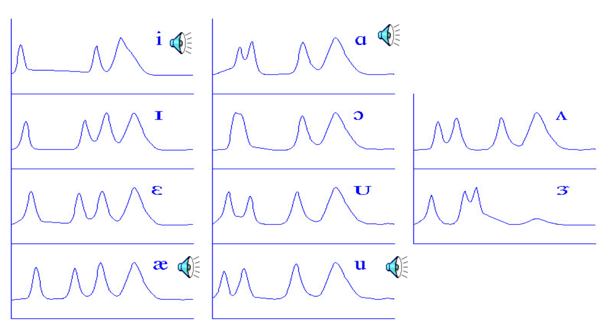
Six Synthesized Sounds Differing in Spectral Envelope

<u>Note:</u> similarities in pitch (due to constant F_0/harmonic spacing) and the differences in timbre or sound quality.

123

123

Timbre and spectral envelope

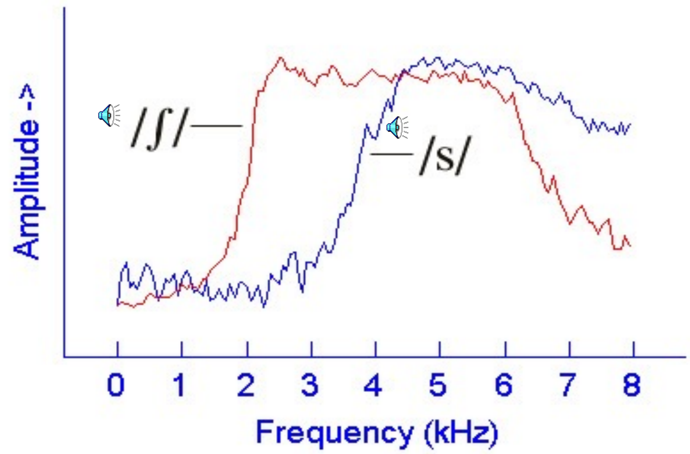


- Timbre differences between vowels. Each vowel has a unique spectral envelope.
- Timbre is a term used often in musical acoustics. In phonetics, timbre differences among vowels are typically referred to as differences in vowel quality or vowel color.

124

124

Timbre and spectral envelope



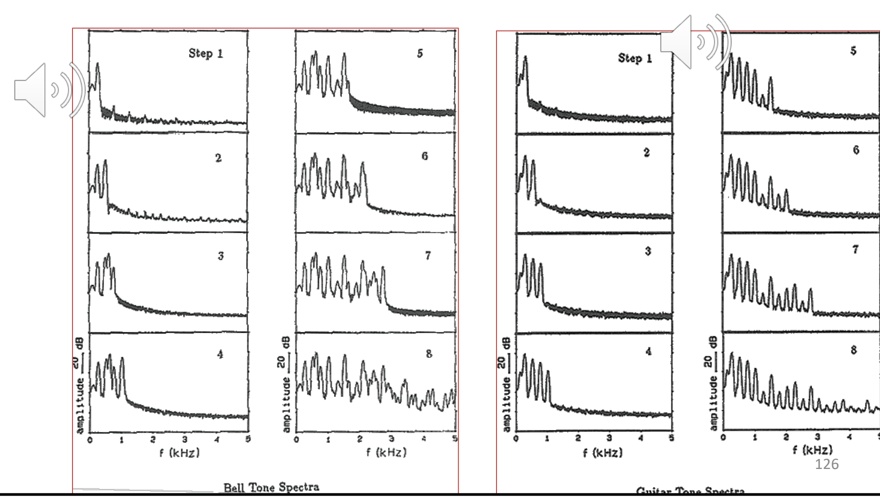
Aperiodic sounds can also differ in spectral envelope, and the perceptual differences are properly described as timbre differences.

125

125

Timbre and spectral envelope

- You will hear 2 timbres (both at a fundamental pitch of 251Hz) synthesized in 8 steps by adding successive partials of F0

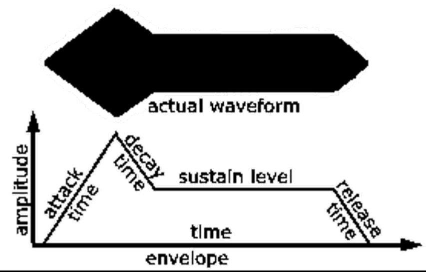


126

126

Sound Envelope

- Timbre also affected by amplitude envelope (envelope in time)
  - Attack – Time it takes for sound to rise from nothing to its greatest intensity. Usually short.
  - Decay – Time it takes for a sound to fall from its attack level to its sustaining level. Decay time is usually short
  - Sustain – The time during which the initial vibrating source continues to supply energy to the sound. Usually perceived as the duration and intensity of the sound
  - Release – Time it takes for the sound to drop from its sustain level to inaudibility after vibrating object stops supplying energy



127

Timbre and amplitude envelope

- First, we listen to a Bach chorale played on piano

- Next the chorale is played backward from end to beginning

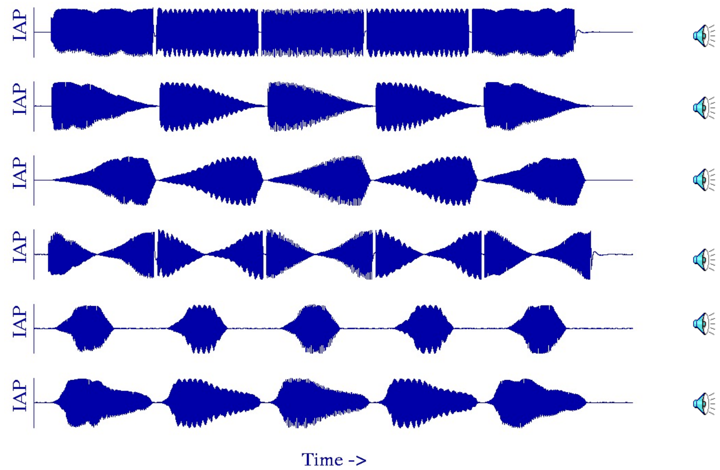
- Now, the last recording (backward chorale) is played in reverse, yielding the original (forward) chorale

  - Note that notes are now played backwards (affecting their amplitude envelope)
  - Instrument resembles a reed organ rather than piano

128

128

Timbre and Amplitude Envelope

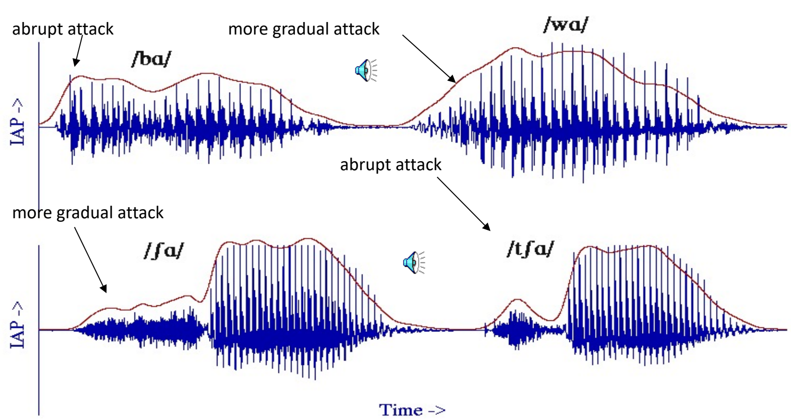


Same melody, same spectrum envelope (if sustained), different amplitude envelopes (i.e., different attack and decay characteristics).

Note differences in *timbre* or *sound quality* as the amplitude envelope varies**.**

129

Timbre and Amplitude Envelope



Timbre differences related to amplitude envelope also play a role in speech. Note the ***differences in the shape of the attack*** for /b/ vs. /w/ and /S/ vs. /tS/.

130

Timbre shift with transposition

- Shifting the frequency scale (transposition) Example:
  - A bassoon playing a 3 octave scale followed by a transposed 3 octave scale

✓Identity of instrument (timbre) sounds different

131

131

Physical vs. Perceptual attributes of sound

Physical attributes	Perceptual attributes			
	Loudness	Pitch	Timbre	Duration
Intensity	+++	+	+	+
Frequency	+	+++	++	+
Spectrum	+	+	+++	+
Duration	+	+	+	+++
A. Envelope	+	+	++	+

132

132

Equalizers

- Changing the amplitudes of specific frequencies across the audible frequency spectrum.
- Affects audio wave shape components:
  - Fundamental - The perceived note
  - Overtones - Define the **timbre** of the sound

133

133

Audio equalizers

- Let you attenuate some frequencies
- Boost others
  - ☞ They can be thought of as frequency-specific volume knobs

20-band equalizer

ps. iTunes has an Equalizer

134

Audio equalizer

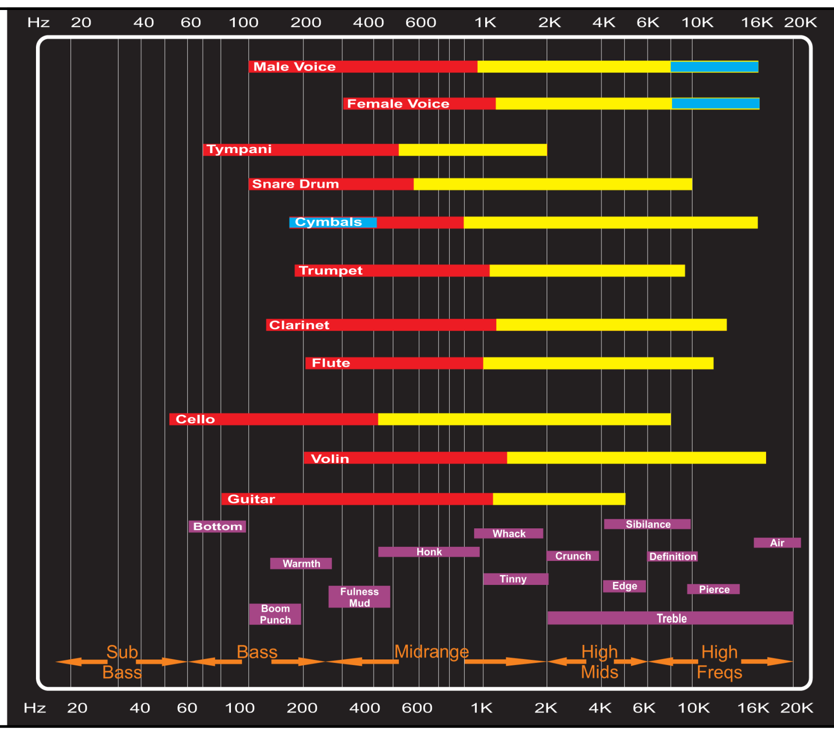
- Basic equalizers focus on treble and bass adjustments

High frequencies (high notes)
~ 2 to 16KHz (C_7–C_{10})
[flute, piccolo, soprano]

Low frequencies (deep tones)
~16 to 256 Hz (C_0 to middle C_4)
Double bass, bassoon, bass guitar

- Advanced equalizer can tailor frequency profile to audio

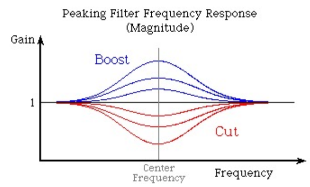
135



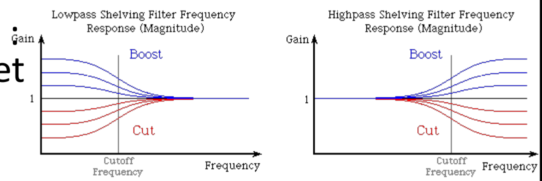
136

Equalizer curves

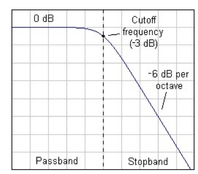
- **Bell curve**: Boost/cut around a center frequency.



- **Low/High shelving** : boost/cut from a set frequency equally onward.

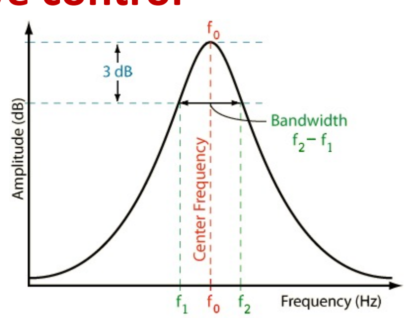


- **High/Low Pass Filter**: cuts lows/highs from a set frequency onward by a slope rate



137

Bell curve control

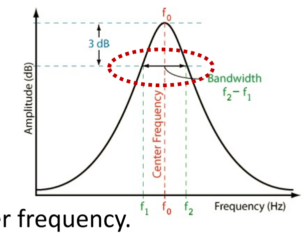


User controls:

1. Gain: boost/cut (dB)
2. Center Frequency (Hz)
   - Incremental: discrete steps
   - Continuous or sweep
3. Bandwidth (Hz) or Q (no units)

138

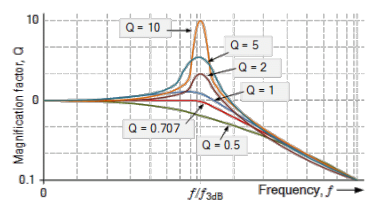
Bandwidth and Q



- Bandwidth:
  - Measurement across points -3 dB from center frequency.
- Q factor:
  - Another way to express bandwidth.

$$Q = \frac{center\ frequency\ (Hz)}{Bandwidth\ (Hz)}$$

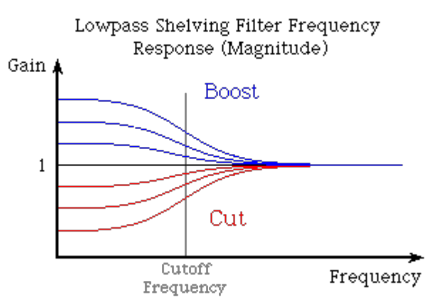
- Inverse relationship. As BW increases, Q decreases.



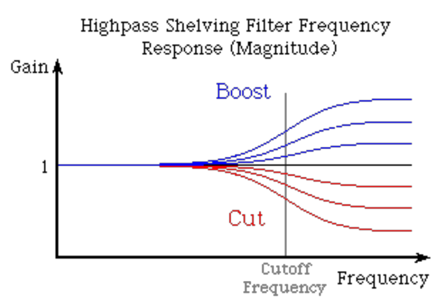
139

Shelving control

- Boost/cuts from a cutoff (turnover) frequency equally onward



Low Shelving Curve



High Shelving Curve

140

Shelving control

User controls:

1. Gain: boost/cut (dB)
2. Turnover frequency (Hz)

Usually no BW or slope control

Note that the turnover frequency is 3 dB from the shelf.

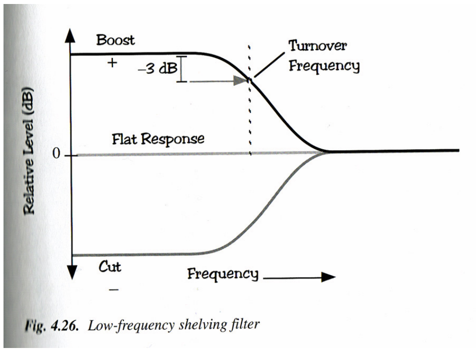


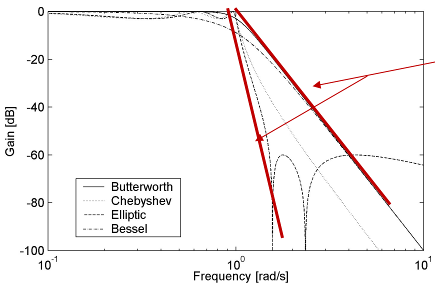
Fig. 4.26. Low-frequency shelving filter

141

Lowpass/Highpass control

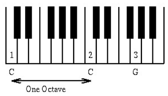
User controls:

1. Filters only cut, no gain control from defined freq.
2. Slope: usually a preset rate.
   - Common Slope Rates: 6, 9, 12, 18 dB/oct



Slope in units of db/oct: Decibels per octave

An octave is a doubling of frequency



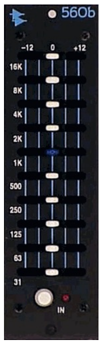
$$oct = \log_2\left(\frac{f_2}{f_1}\right)$$

If filter is a <u>bandpass</u> filter, user controls bandwidth and center frequency (no gain control)

142

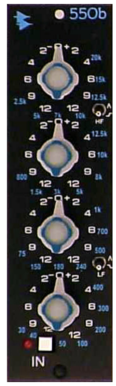
Types of equalizers

Graphic



Multiple bands (broken into octave increments). Fixed frequency, fixed bandwidth. Variable +/- gain.

Semi-parametric



preset or limited (switched) bandwidth adjustment.

Parametric

sweepable (continuous) bandwidth adjustment.

143

MASKING

144

144

Masking

- Hearing phenomenon
  - When the perception of one sound is affected by presence of *another* sound
  - one sound being *masked* by another
- Term masking is used to describe effects of noise and interference in sound perception
- We experience masking everyday

145

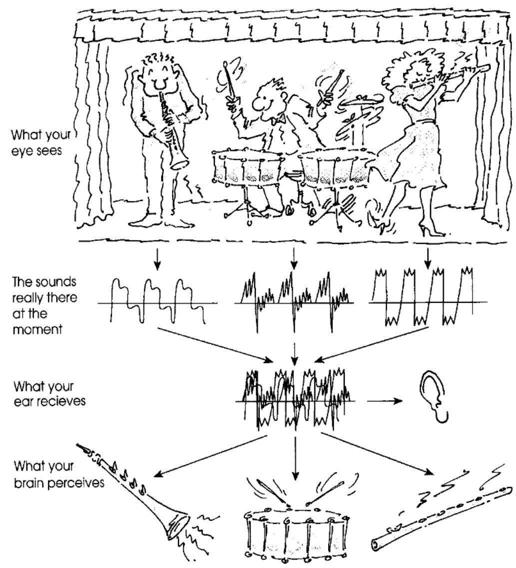
145

Masking

146

146

Masking



147

147

Masking

- Masking generally occurs when a softer sound becomes <u>inaudible</u> when a louder sound masks it.
    - Soft sound: Maskee
    - Loud sound: Masker
  - E.g. you have a having a conversation in the street when a loud truck passes by and your friend can't hear you anymore
    - Conversation: Maskee
    - Truck: Masker

✓Beneficial for applications such as audio compression

148

148

Masking

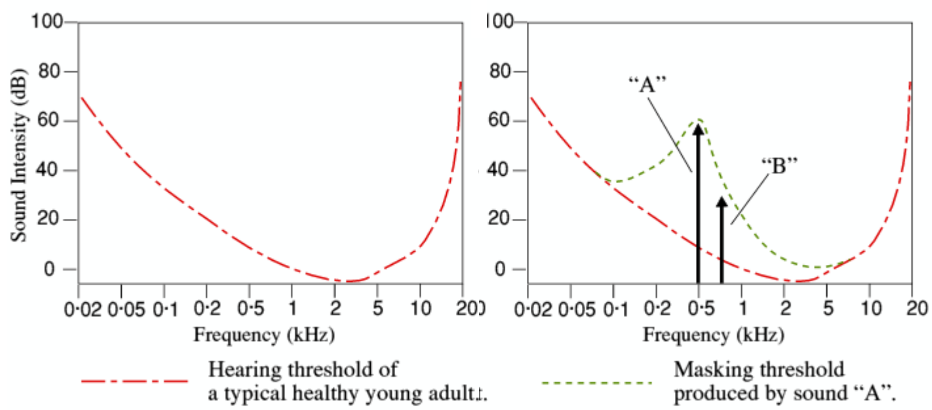
- Masking occurs in a variety of domains
  - Frequency masking
    - Simultaneous masking
  - Temporal masking
    - Non-simultaneous masking

149

149

Frequency Masking

(simultaneous masking)

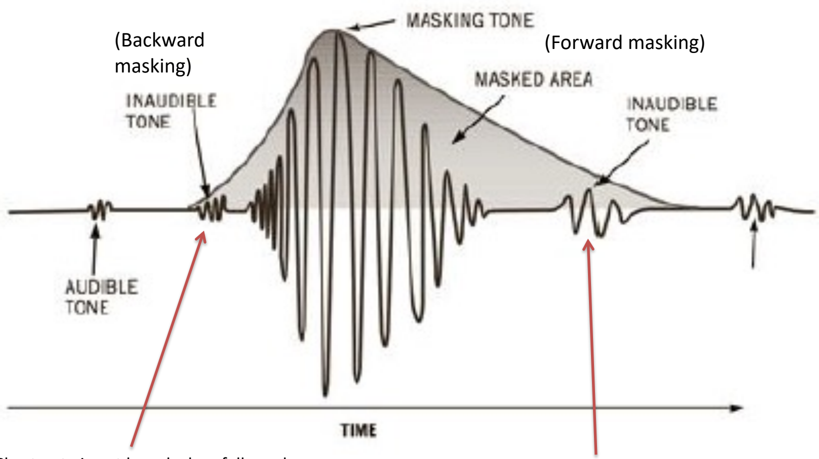


150

150

Temporal Masking

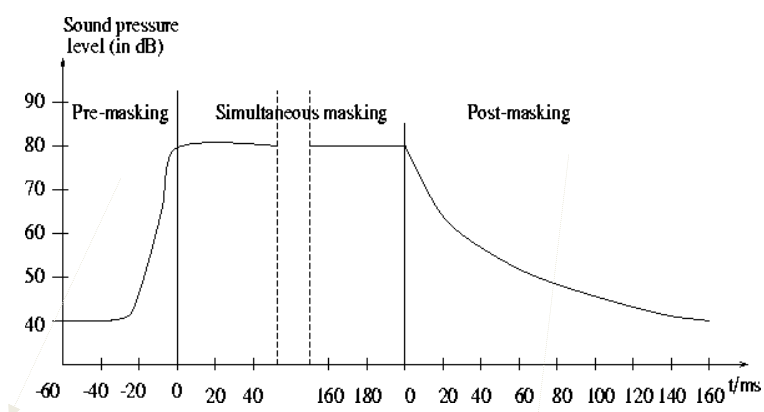
(non-simultaneous masking)



151

151

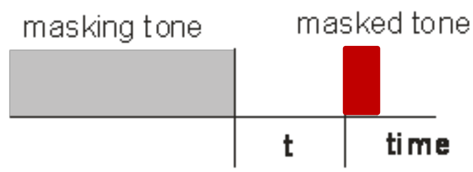
Temporal Masking



152

152

Forward masking



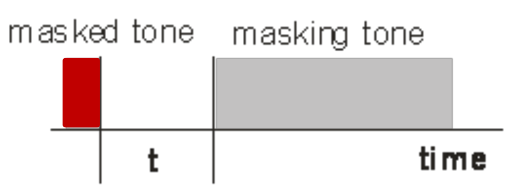
- A maskee (masked tone) is played at decreasing sound levels.
  1. Preceded by masker **100ms** prior
  2. Preceded by masker **10ms** prior

Forward masking

153

153

Backward masking



- A maskee (masked tone) is played at decreasing sound levels.
  1. Followed by masker after **100ms**
  2. Followed by masker less than **10ms**

Backward masking

154

154

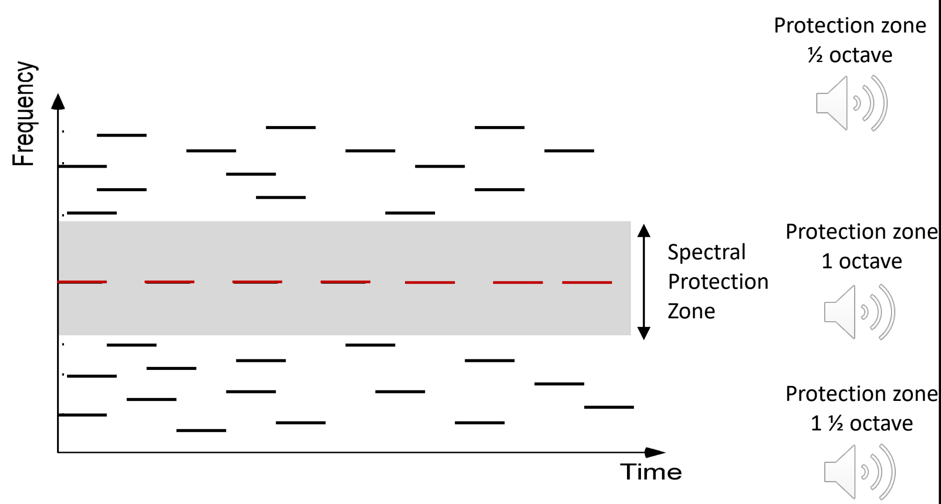
Energetic masking

- When 2 sounds (masker/maskee) share frequency components or are in close temporal proximity of each other
  - **Energetic masking**
- Energetic masking encompasses
  - Frequency masking
  - Temporal masking
- In contrast with another phenomenon called **informational masking**

155

155

Informational Masking



156

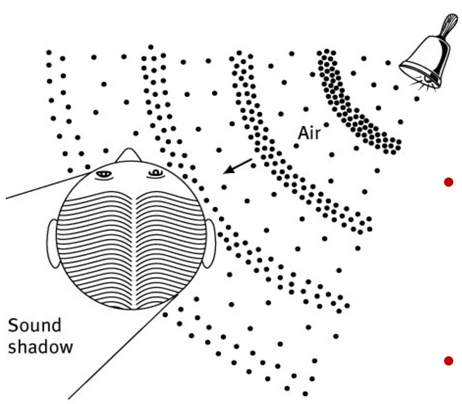
156

SOUND LOCALIZATION

157

157

Sound Localization



- The direct path from the acoustic source to the two ears will generally be different.
- The signal needs to travel further to more distant ear
- More distant ear partially occluded by the head

158

158

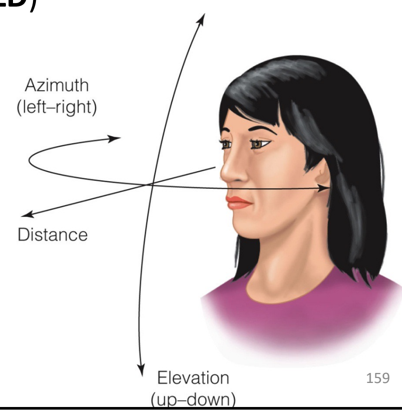
Auditory Localization cues

between 2 ears

1. inter-aural timing differences (**ITD**)
2. inter-aural level differences (**ILD**)
3. monaural cues (pinnae)

…

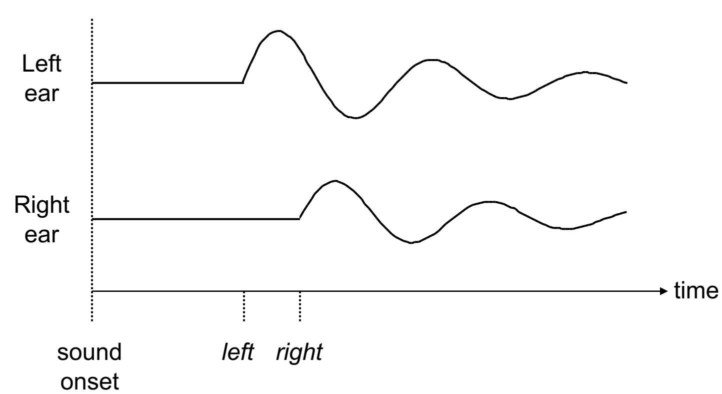
4. head movements



159

159

Interaural cues

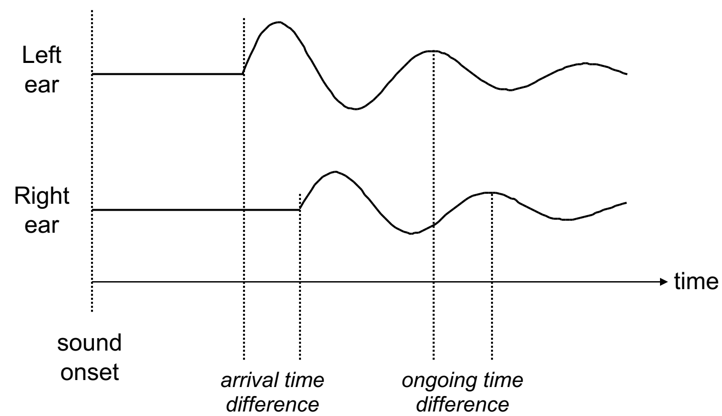


160

160

Interaural cues

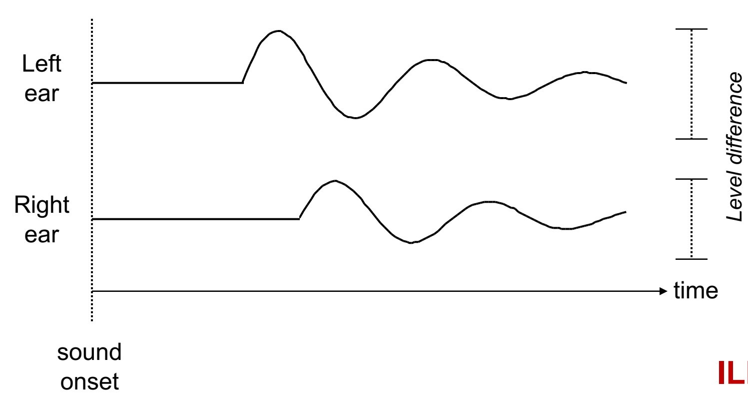
ITD



161

161

Interaural cues



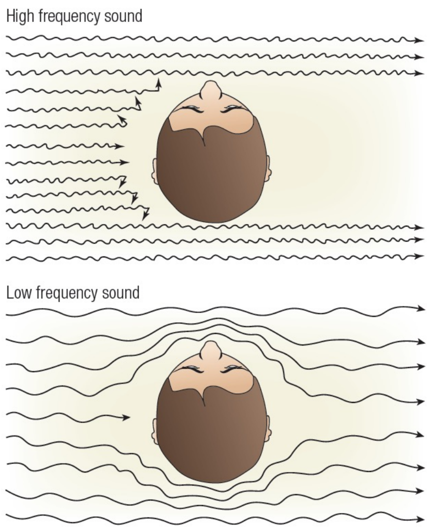
ILD

Also called **IID**:
interaural intensity difference

162

162

Inter-aural cues – Head shadowing



High-frequency sound waves are "blocked" by the human head and cast a "shadow" at the far ear **(Strong ILD cue)**

Low-frequency sound waves wrap easily around the head and cast little or no sound shadow **(Weak ILD Cue)**

163

Interaural cues

Interaural level differences (ILDs)

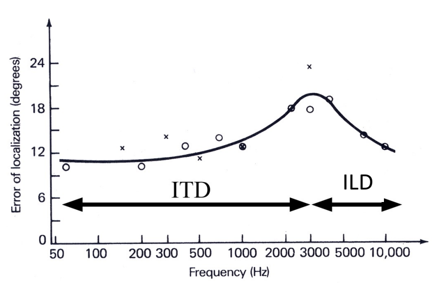
- Threshold ILD ≈ 1 dB

  => Effective for high frequencies

Interaural time differences (ITDs)

- Threshold ITD ≈ 10-20 μs (~ *0.7 cm)*

  ⇒ Effective for low frequencies



164

Interaural cues

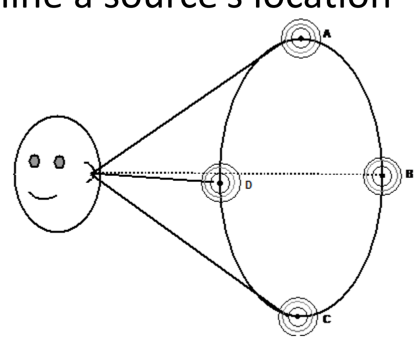
- **ILD**
  - Head-size dependent: larger heads create bigger ILDs for the same frequency
  - Very-frequency dependent – larger effect at higher frequencies
- **ITD**
  - Head-size dependent: larger heads create bigger range of ITDs
  - Less-frequency dependent – works over large freq range
  - Requires extraordinarily exquisite temporal mechanisms (10 – 20 μs sensitivity)

165

165

Interaural cues are not sufficient

- ITD and IID values cannot be used to uniquely determine a source's location



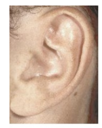
For a spherical head, points lying on the "Cone of Confusion" have the same ITD and IID values

166

166

Monaural cues

- Rely only on 1 ear -> monaural
  - Effective because of filtering properties of outer ear
  - Mostly pinnae



- Pinnae acts as directional filter
  - ✓ It amplifies sounds above and below differently
  - ✓ It acts mostly on high frequencies (above 5KHz)
  - ✓ Shoulder reflection causes changes in signal in 2-3KHz

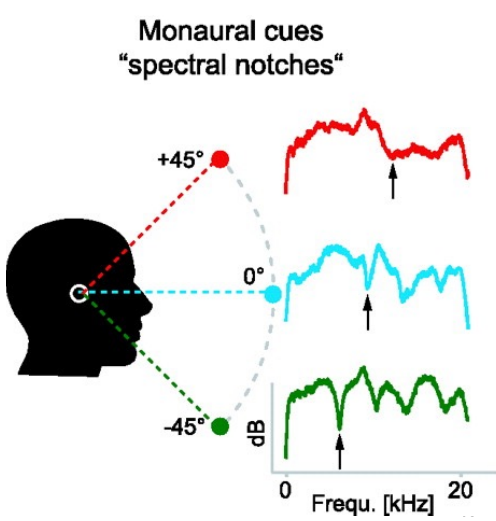
➡ Monaural localization is not as accurate as binaural localization

167

167

Monaural cues

- Monaural localization most effectively helps distinguish sound elevation (vertical dimension)
- A central notch in the spectrum (black arrow) moves to high/low freq when sound moves above/below the horizon
  - Pinnae reflections color the sound spectrum



168

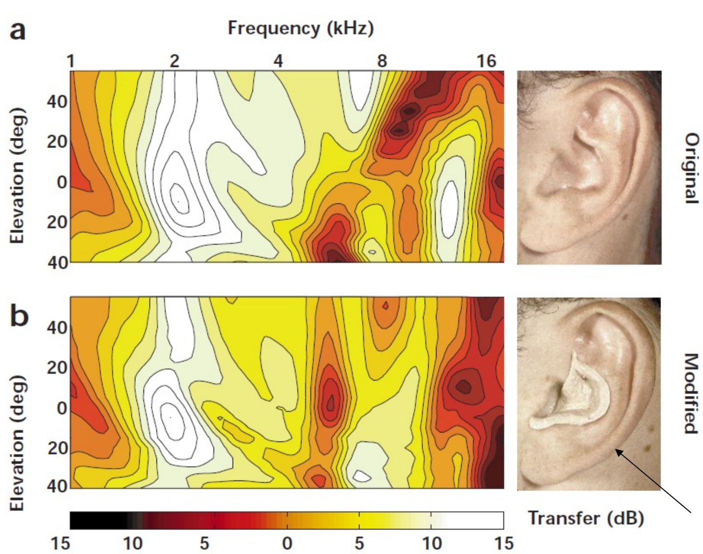
Monaural cues

- Monaural cues are starting to be considered like **earprints**
  - They are highly individual, depending on the shape and size of the outer ear
  - Often create problems in binaural sound effects recorded using a generic head with different-shaped outer-ear surfaces
    - Foreign ears cause localization confusion especially in judging elevation
    - Can cause inside head localization

169

169

Modifying the Pinna Transfer Function



Listeners can learn to hear with a new pinnae

170

Head-related transfer function

- A **Head-related transfer function** (**HRTF**) is a function that characterizes how a particular ear receives sound from a point in space
- It allows an audio system to simulate effects of sound in 3D space by mixing audio tracks with right filtering and delay parameters

171

171

Head-related transfer function

- The Head Related Transfer Function for each ear can be defined as:

$$X_L(\omega) = H_L(\omega, \theta, \phi) X(\omega)$$

$$X_R(\omega) = H_R(\omega, \theta, \phi) X(\omega)$$

- Where:
  - ω represents frequency and θ, ϕ represent elevation and azimuth respectively
  - $H_L(\omega, \theta, \phi)$ and $H_R(\omega, \theta, \phi)$ are the left and right HRTFs
  - $X_L(\omega)$ and $X_R(\omega)$ are the Fourier Transforms of the signals received by the Left and the Right ears
  - $X(\omega)$ is the Fourier Transform of the source signal

172

172

HRTF – Transfer function

$$x(t) \longrightarrow \boxed{\text{HRTF}} \longrightarrow y(t)$$

- Measuring a transfer function of a system involves presenting inputs and measuring outputs
  - Direct method: broadcast test signals from different directions and record received signals at entrance of ears (typically using in-ear microphones)
  - Reciprocal method: exchange location of broadcast and reception (switch inputs and outputs)

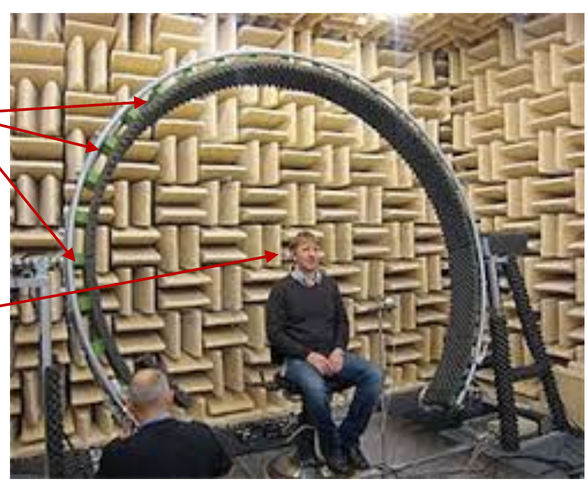
173

173

HRTF measurement

Vary speaker location

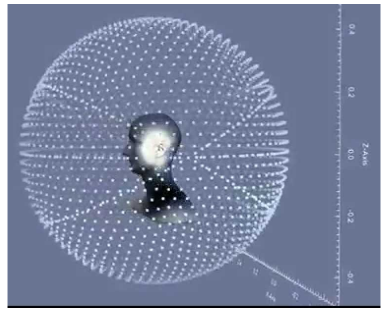
Record signal received



174

174

HRTF



- Special setup to sweep over entire space
- Note foam on room walls to make it anechoic (no reverberation)

175

175

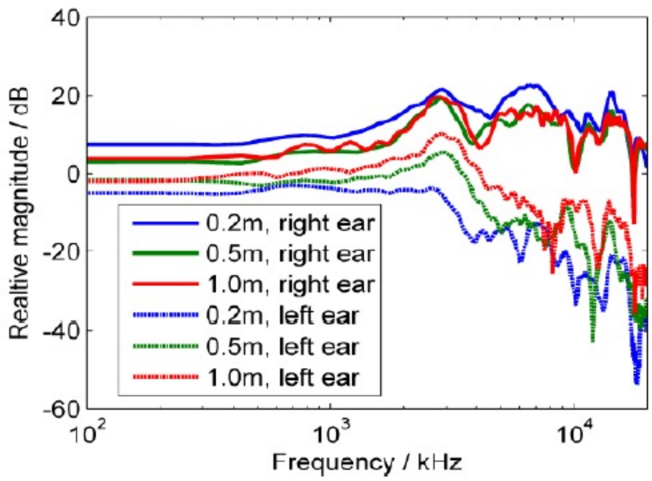
HRTF measurement

- HRTFs are highly individualized transfer functions (depends on head shape, size, pinnae)
- Use of dummy heads

176

176

Example HRTF



HRTF of dummy head at r=0.2m, 0.5m and 1m and 90deg right

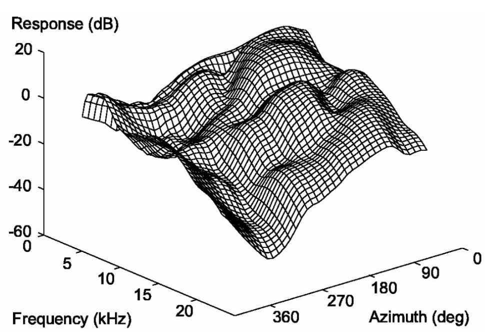
177

177

Example HRTF

There are databases available of many HRTFs measured from a large pool of listeners

See:



178

178

HRTFs

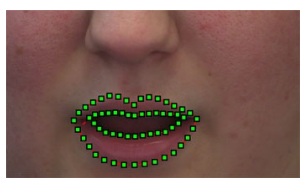
- HRTFs are widely used in gaming and virtual reality simulations to offer an immersive, realistic 3Dexperience without relying on visual cues

179

179

Audio-visual integration

- Perception of sounds is heavily affected by other sensory modalities, importantly vision
  - We rely on it heavily especially to complement audio signals and navigate noisy environments .. Particularly true for people with hearing impairment -> **Lip reading**



  - Many cases where the 2 modalities clash

180

180

AV integration – McGurk effect

- “Compromise” between conflicting sound and visual cues in speech understanding
  - Compare auditory perception with eyes open vs. eyes closed

181

181

AV integration – Motion bounce illusion

- What you hear is what you see
  - check with and without audio

http://www.michaelbach.de/ot/mot-bounce/index.html

182

182

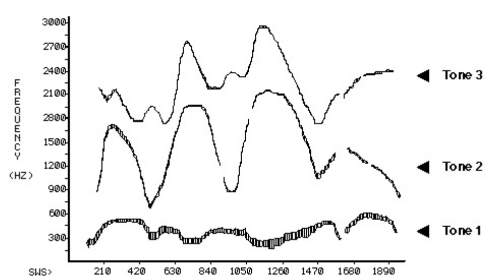
AV integration – Ventriloquism effect

- The ventriloquism effect is what makes us think the ventriloquist's dummy is talking

183

183

Other complicated phenomena … e.g. sinewave speech

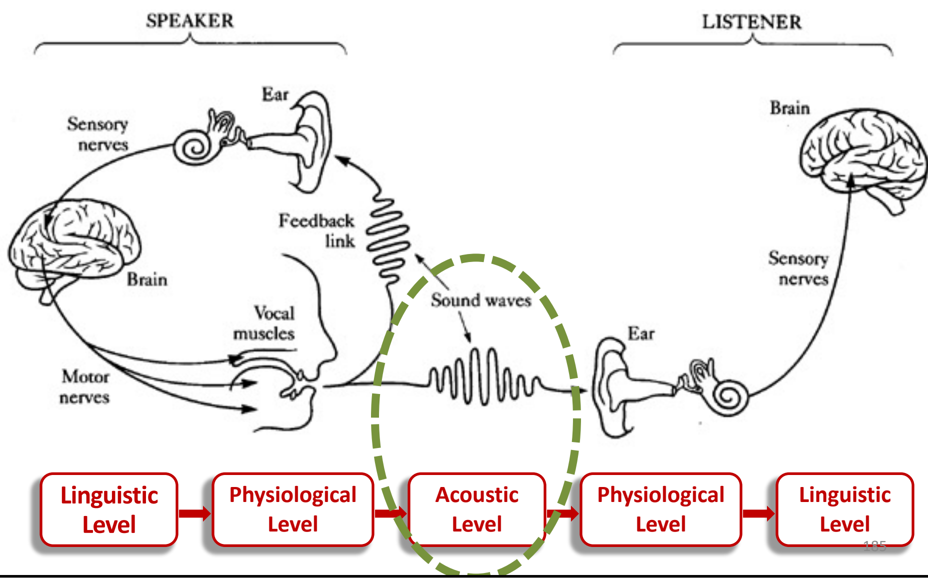


Wave examples

184

184

The speech chain



185