Psychoacoustics

Li Su

2019/03/05

Reference

- David M. Howard and Jamie A.S. Angus, “Acoustics and Psychoacoustics,” on ScienceDirect.com, Fourth Edition, 2012
- Meinard Mueller, “Fundamentals of Music Processing,” Springer, 2015

Elements of sounds

- (Roughly speaking) 4 perceptual elements of sounds

Perceptual elements	**Related physical elements**
Pitch (high or low)	Fundamental frequency, fundamental periods, and others
Loudness (strong or weak)	Energy intensity and distribution, and others
Timbre (cold, warm, bright, sweet, …)	Wave shape function, spectral envelope, attack-decay-sustain-release (ADSR) curve, spectral skewness, and many others
Direction	Multiple channels

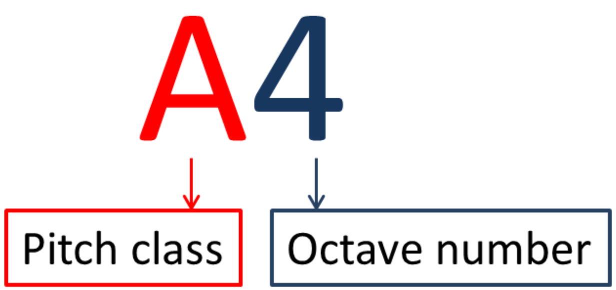
- More critical cases
  - Pitched and non-pitched sounds
  - Consonance and dissonance

Frequency and pitch

- The higher the frequency of a sinusoidal wave, the higher it sounds
- Human's audible frequency: 20 Hz – 20,000 Hz (20 kHz)
- Dog's: $\sim$ 45 kHz; cat's: $\sim$ 64 kHz
- Ultrasound: > 20 kHz; infrasound: < 20 Hz

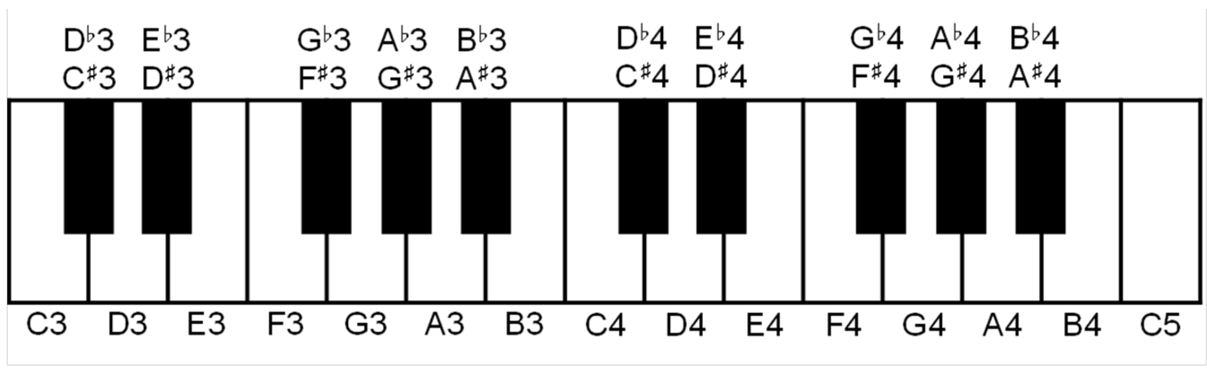
Scientific pitch notation and MIDI number

- Musical Instrument Digital Interface (MIDI): 21 – 108 for piano
- Concert pitch: A4 = 440 Hz
- MIDI number of C0 = 1



F0 = 440 Hz

MIDI = 69



Pitch

- Octave equivalence: two frequencies differing by a power of 2 sounds similar
- Semitone: two frequencies (i.e., f_1 and f_2, $f_1 > f_2$) differ by 1 semitone when their ratio is $f_1/f_2 = 2^{1/12} \approx 1.059463$
- One octave contains 12 semitones
- The center frequency $F_{pitch}(p)$ of each pitch with MIDI number p is
- $F_{pitch}(p) = 440 \times 2^{(p-69)/12}$
- Example: we have $F_{pitch}(p+12) = 2F_{pitch}(p)$
- , $\frac{F_{pitch}(p+1)}{=F_{pitch}(p)} = 2^{1/12} \approx 1.059463$

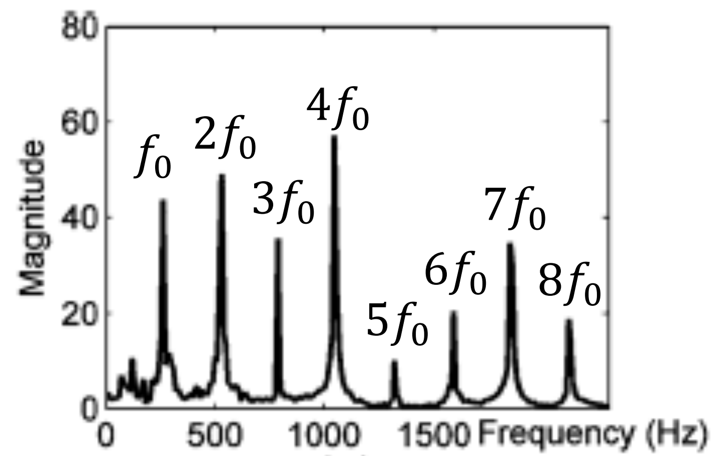
Tuning

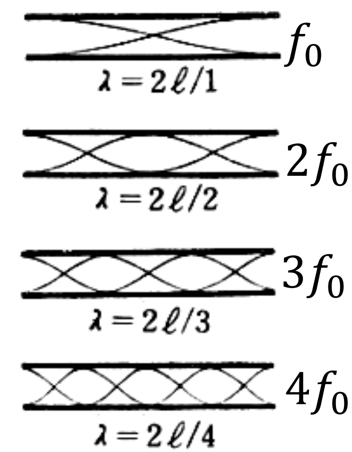
- We (usually) like the sounds with simple frequency ratio (i.e., ratio of small whole number)

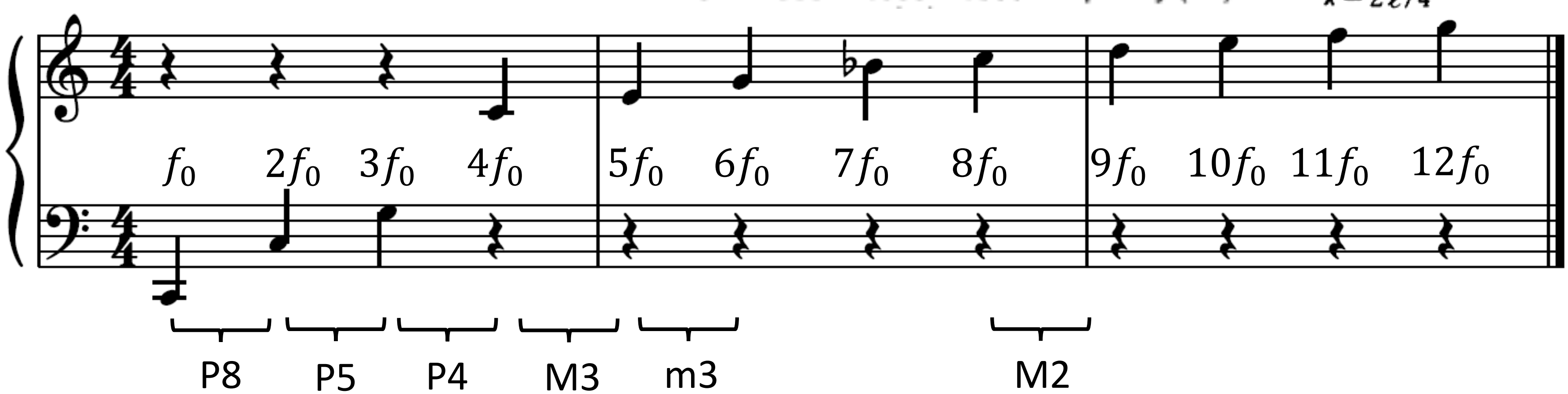
#.	Interval	JI ratio	Pyt ratio	12-TET
0	Unison	1:1	1:1	$1:1$
1	Minor second	15:16	$3^5:2^8$	$1:2^{1/12}$
2	Major second	8:9	$2^3:3^2$	$1:2^{2/12}$
3	Minor third	5:6	$3^3:2^5$	$1:2^{3/12}$
4	Major third	4:5	$2^6:3^4$	$1:2^{4/12}$
5	Perfect fourth	3:4	$3:2^2$	$1:2^{5/12}$
6	Augmented fourth	32:45	$3^6:2^{10}$	$1:2^{6/12}$
7	Perfect fifth	2:3	$2:3$	$1:2^{7/12}$
8	Minor sixth	5:8	$3^4:2^7$	$1:2^{8/12}$
9	Major sixth	3:5	$2^4:3^3$	$1:2^{9/12}$
10	Minor seventh	5:9	$3^2:2^4$	$1:2^{10/12}$
11	Major seventh	8:15	$2^7:3^5$	$1:2^{11/12}$
12	Perfect octave	1:2	1:2	$1:2$

Harmonic series and pitch

- Natural harmonics corresponds to musical pitch

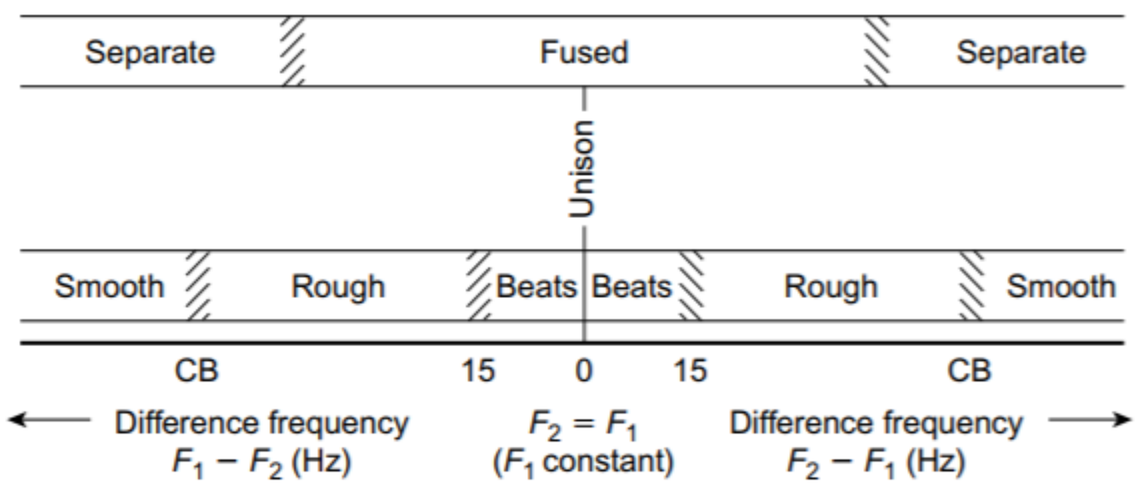






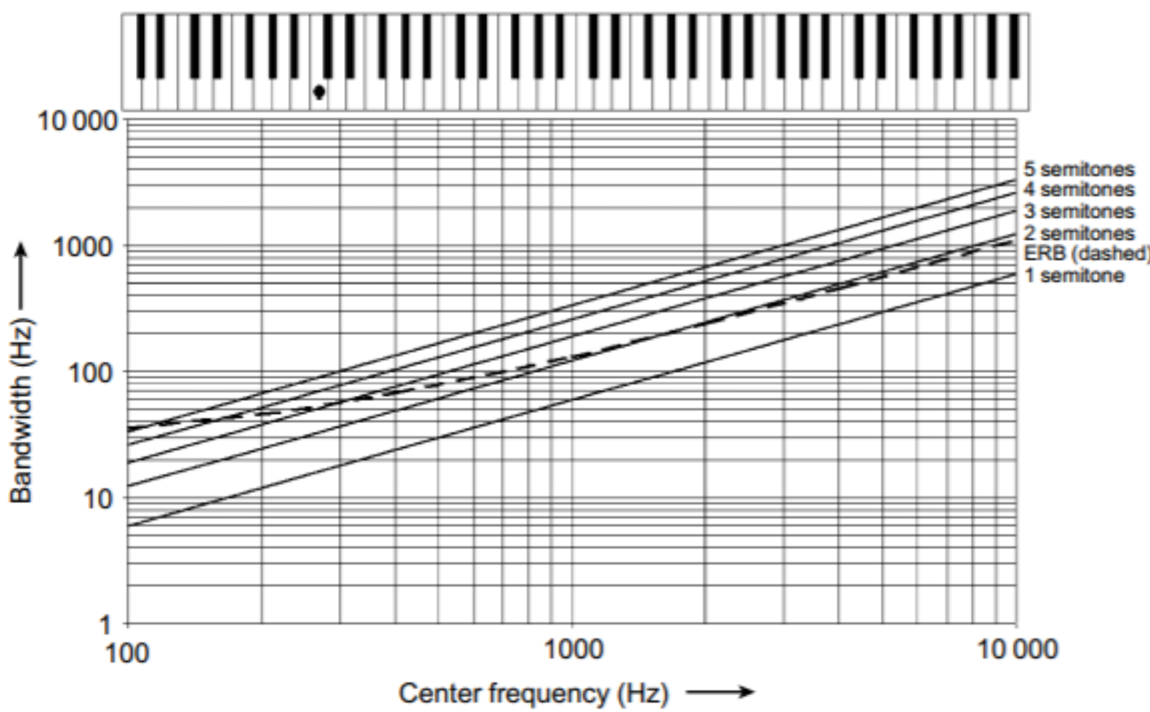
Critical bands

- Experiments on two pure tones (i.e. sinusoidal waves)
- Subject test: are the two tones (1) the same (2) beats (3) rough/smooth separated sources



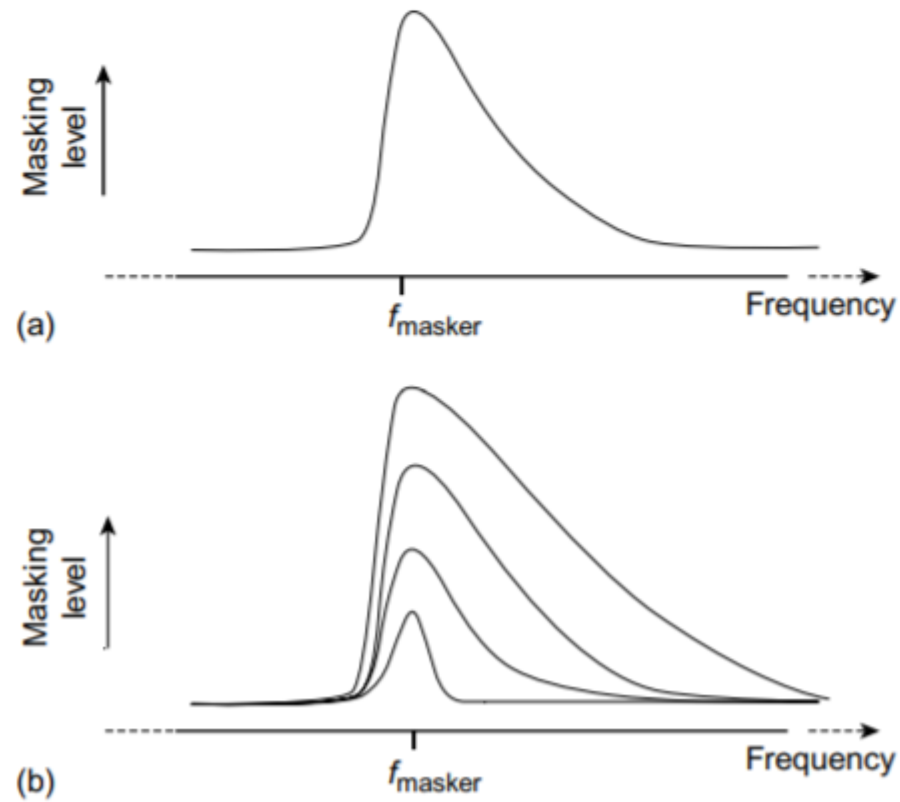
Equivalent rectangular bandwidth (ERB)

- Glasberg and Moore equation
- $ERB = \{24.7 \times [(4.37 \times f_c/1000) + 1]\}$ Hz



Masking effect

- Masking of one sound by another
- (a) Idealized masking level to illustrate the "low masks high," or "upward spread of masking effect" for a masker of frequency f_{masker} Hz. (b) Idealized change in masking level with different levels of masker of frequency f_{masker} Hz.
- Low masks high



Dynamic, loudness, and intensity

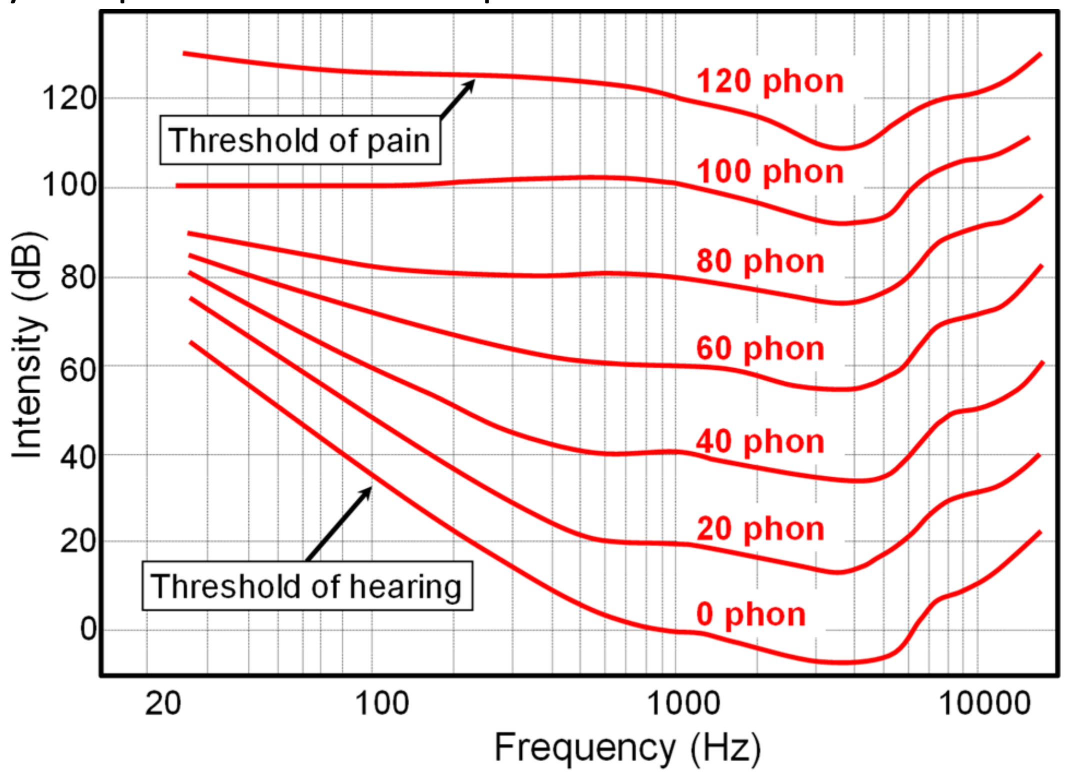
- Dynamic: a term referring to the musical symbols that indicate the volume, like forte (f) or piano (p)
- Loudness: a perceptual, subjective property, depending on sound intensity, duration and frequency, where the sound can be ordered from quite to loud
- Intensity: a physical property, defined as the sound power per unit area (e.g., W/m^2)
- Threshold of hearing (TOH): the minimal sound intensify of a pure tone (i.e., a sinusoid) a human can hear, $I_{TOH} \coloneqq 10^{-12} W/m^2$
- Threshold of pain (TOP): $I_{TOH} \coloneqq 10 W/m^2$
- dB-scaled sound intensity: $dB = 10\text{\textbackslash}\log_{10}(I/I_{TOH})$

Sound intensity

Source	Intensity	Intensity level	× TOH
Threshold of hearing (TOH)	10^{-12}	**0 dB**	1
Whisper	10^{-10}	**20 dB**	10^{2}
Pianissimo	10^{-8}	**40 dB**	10^{4}
Normal conversation	10^{-6}	**60 dB**	10^{6}
Fortissimo	10^{-2}	**100 dB**	10^{10}
Threshold of pain	10	**130 dB**	10^{13}
Jet take-off	10^{2}	**140 dB**	10^{14}
Instant perforation of eardrum	10^{4}	**160 dB**	10^{16}

Equal loudness curve

- Loudness is related with intensity and frequency
- Human ears are most sensitive to sounds around 2--4 kHz
- Frequency-dependent unit: phon

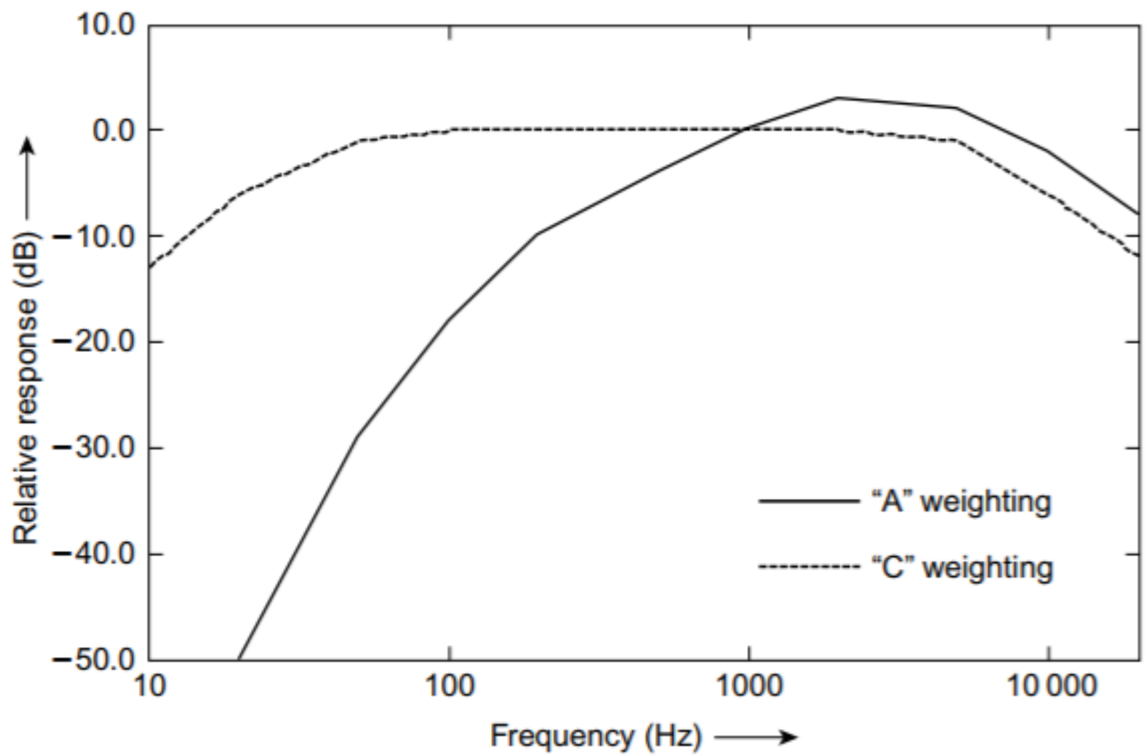


Phon

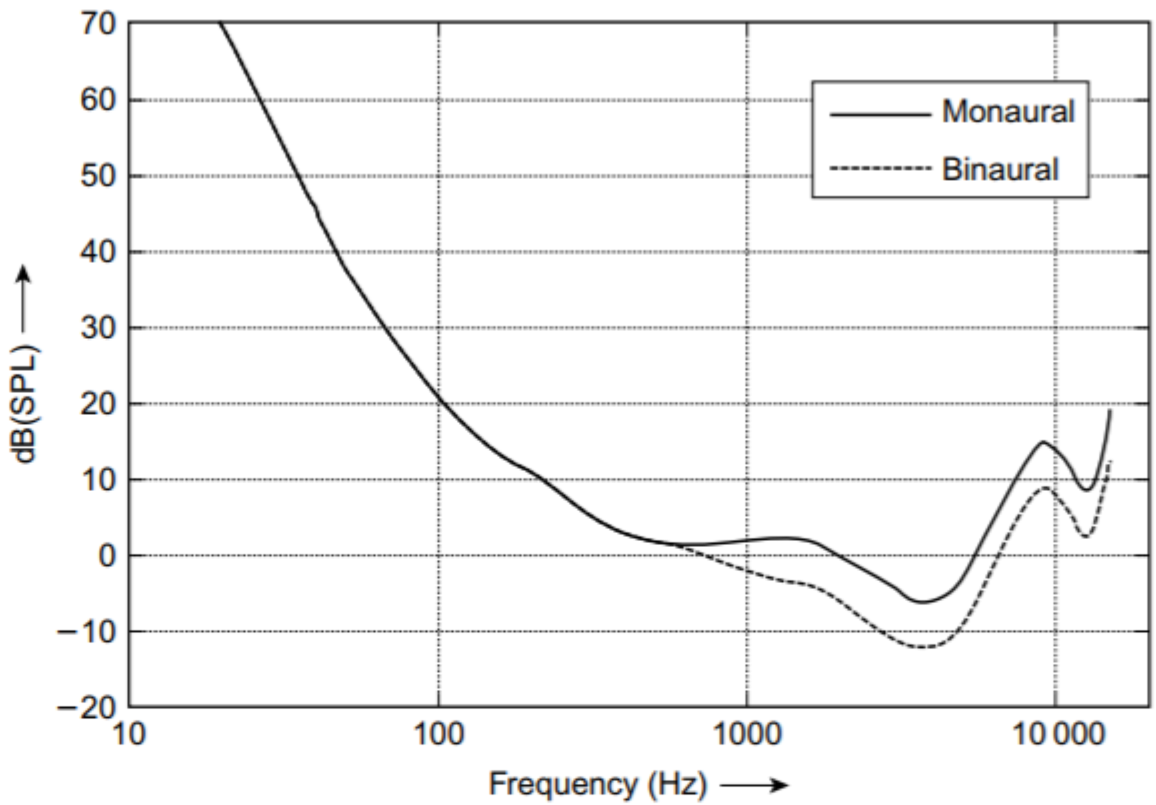
- The loudness of sine wave signals, as a function of frequency and sound pressure levels, is given by the “phon” scale. The phon scale is a subjective scale of loudness based on the judgments of listeners to match the loudness of tones to reference tones at 1kHz.

Measuring loudness

- Measuring loudness: using the sound pressure level but frequency weighting it to compensate for the variation of sensitivity of the ear as a function of frequency



Threshold of hearing

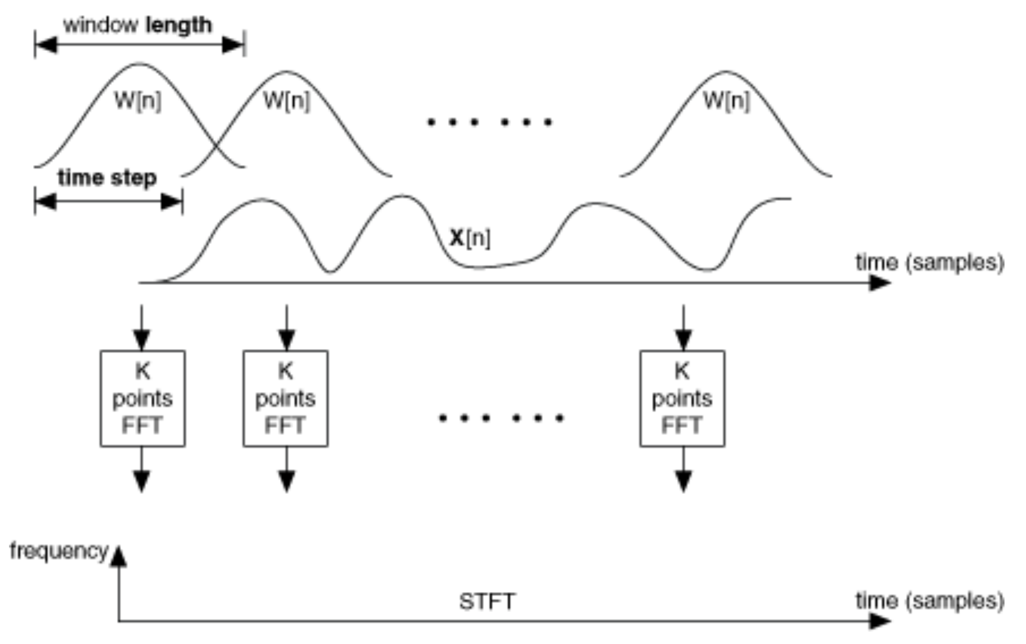


Timbre

- Timbre is the attribute whereby a listener can judge two sounds as dissimilar using any criterion other than pitch and loudness
- Timbre information allows us to tell apart the sounds of a violin, oboe and trumpet, even when the pitch and loudness of them are the same
- Words describing timbre: bright, dark, warm, harsh, cold, ...
- Timbre in signal processing: the *shape* of the signal representation (including time-domain signal and spectrum)

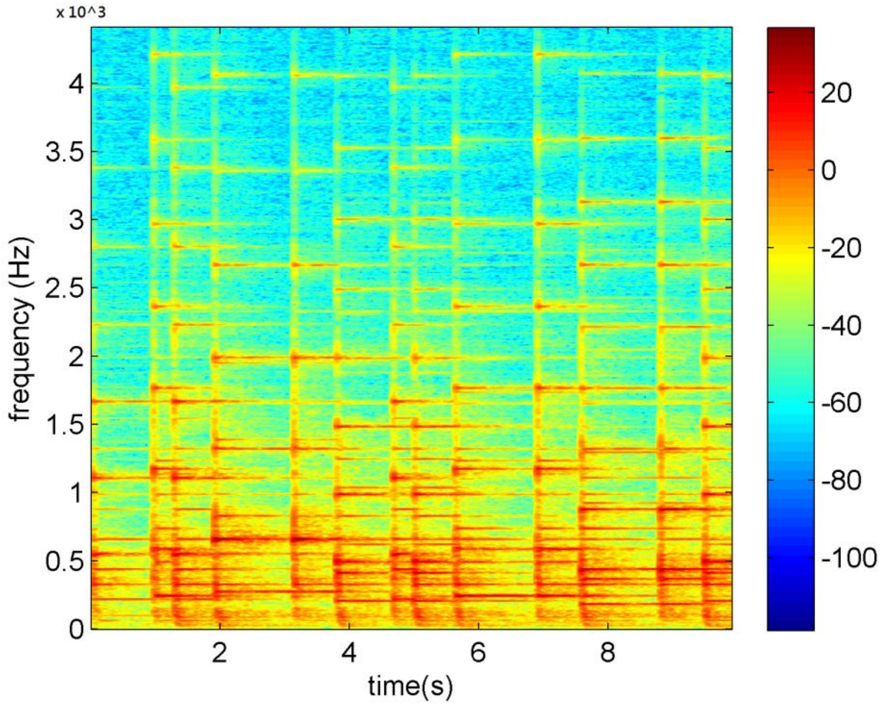
Preliminaries: spectrogram

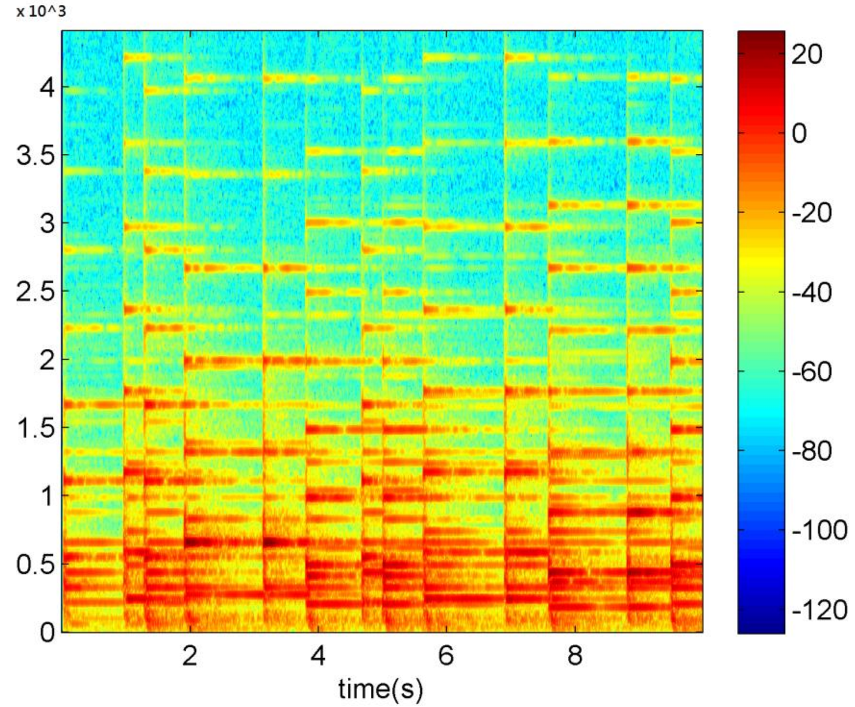
- Slice the signal into frames of segments (usually overlapped)
- Multiply the short segments by a window function
- Do discrete Fourier transform for each segment
- Fast Fourier transform (FFT)



Window size

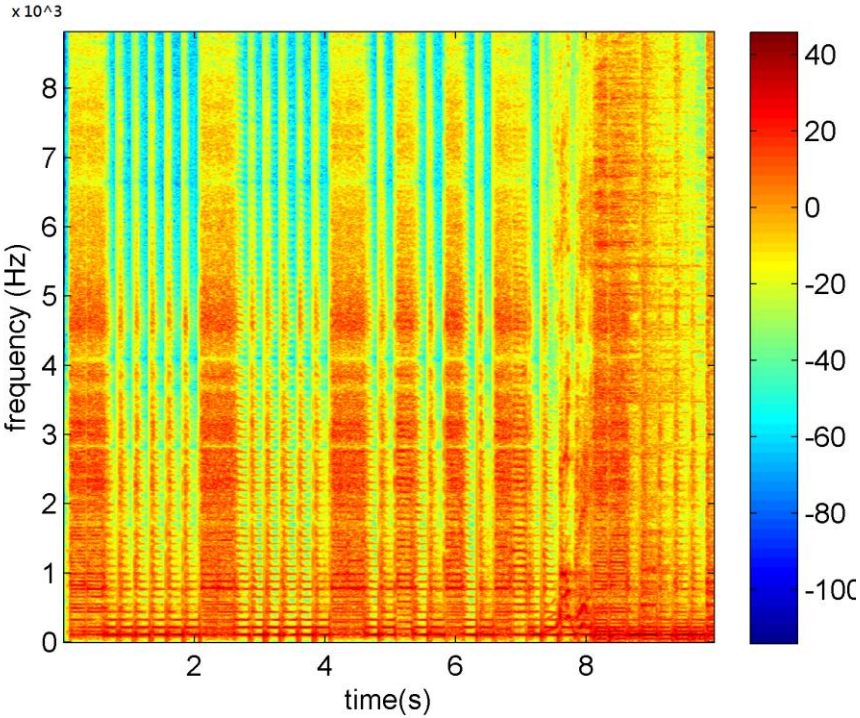
- Large window: better frequency resolution, worse time resolution
- Small window: better time resolution, worse frequency resolution
- Example: piano (window size = 4096 or 1024)

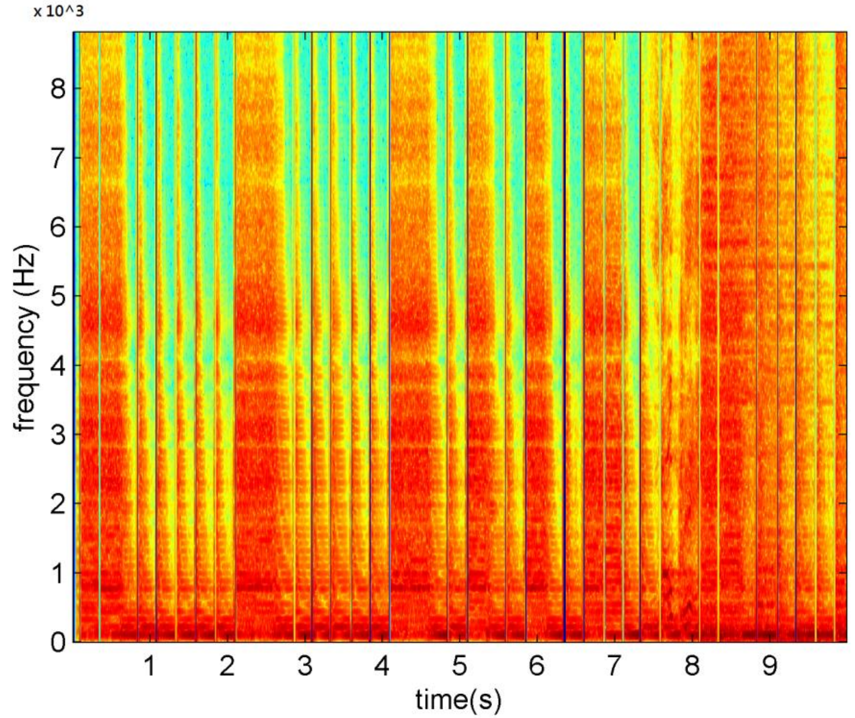




More example: rock

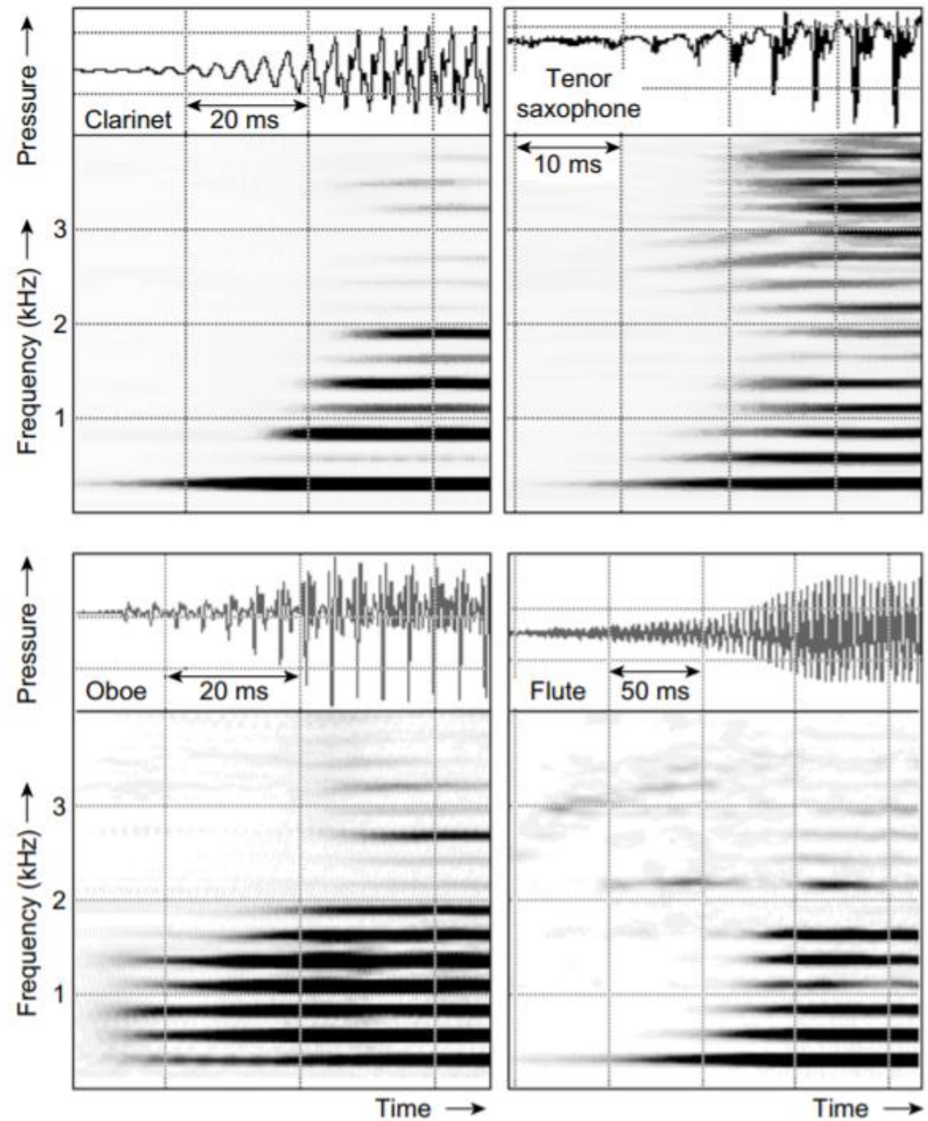
- Window size = 4096 or 1024





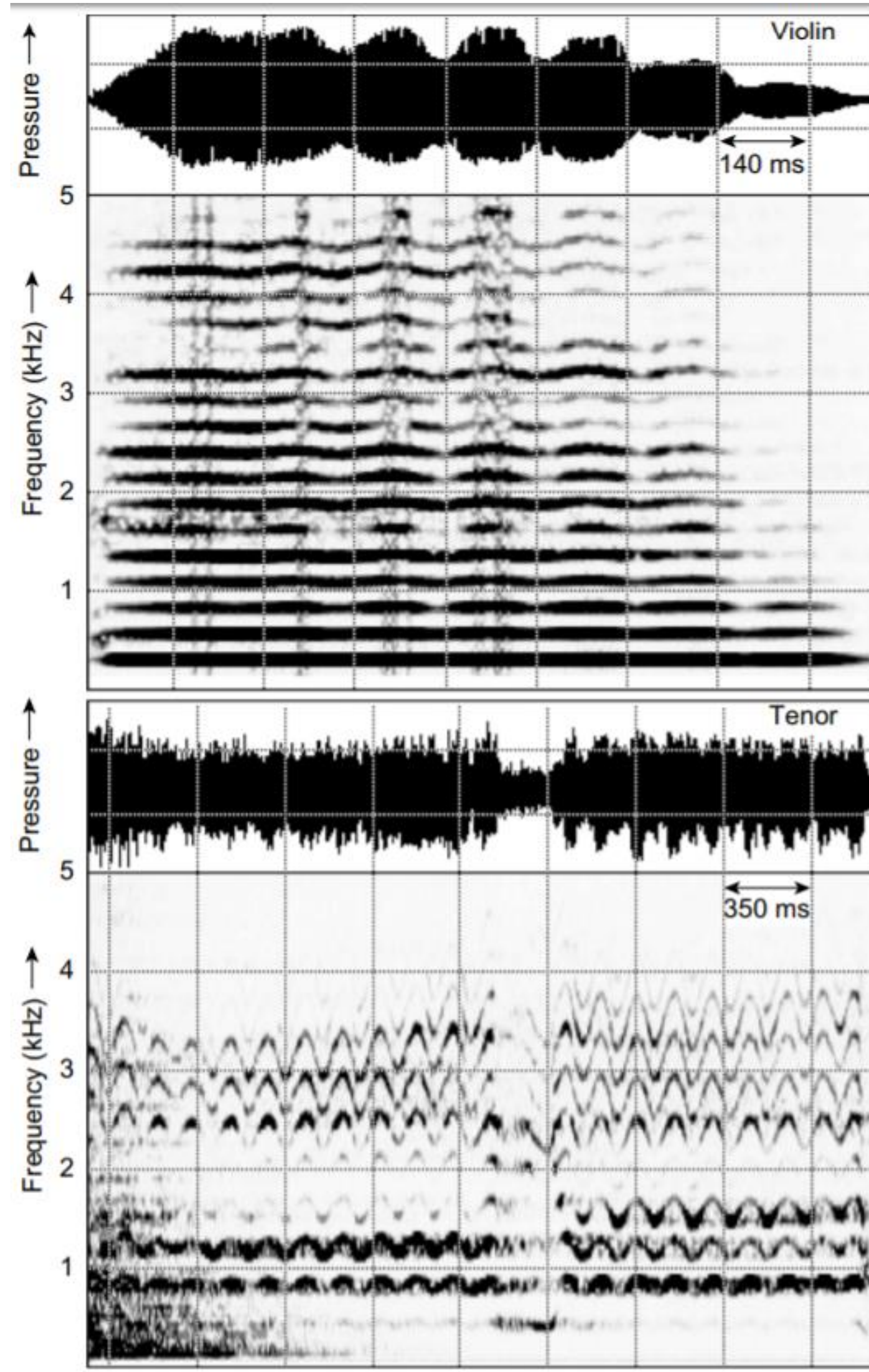
Timbre: note onset

- Waveform (upper) and spectrogram (lower) of the note onset phase for C4 played on a clarinet, flute, oboe and tenor saxophone. LTAS for the clarinet and tenor saxophone are shown in Figure 4.24.



Vibrato/tremolo

- Frequency/amplitude modulation
- UPPER: Waveform and spectrogram of C4 (262Hz) played on a violin. LOWER: Waveform and spectrogram of the last three syllables of the word Vittoria from the second act of Tosca by Puccini sung by a professional tenor ($f_0 = $ Bb4) from a CD recording.

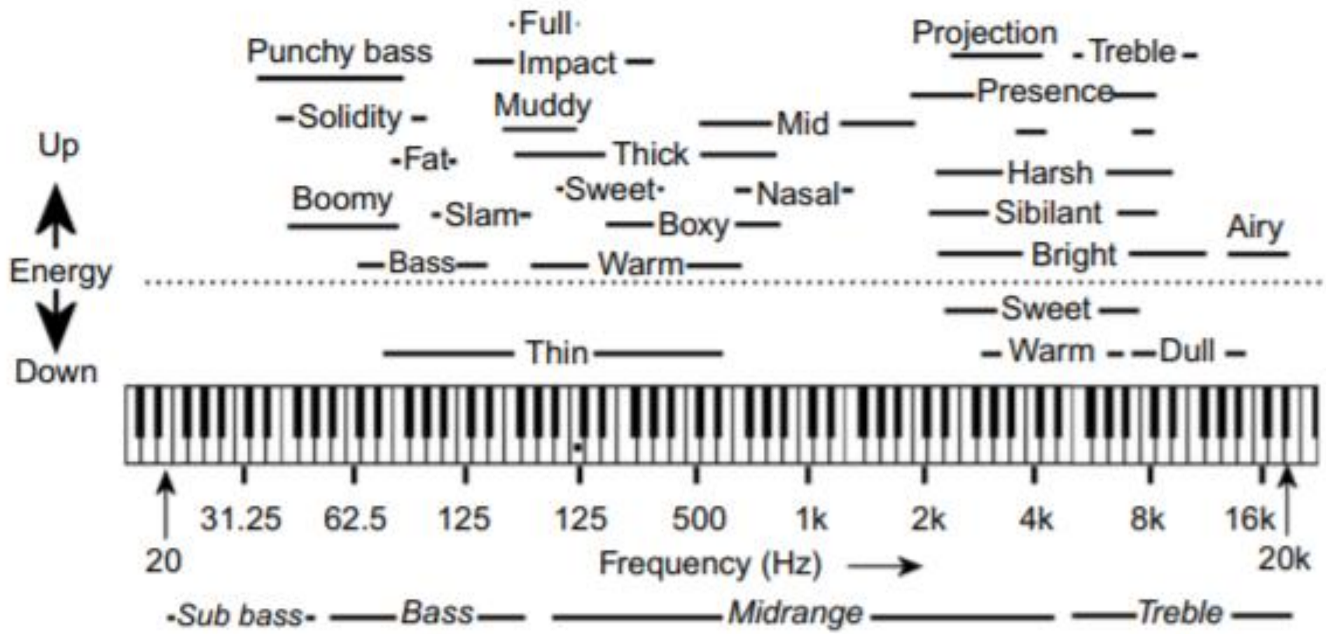


Helmholtz rule (1877)

Helmholtz rule	Human hearing modeling spectrogram	Example timbre descriptors	Example acoustic instruments
1	f_0 dominates	Pure Soft Simple Pleasant Dull at low pitch Free from roughness	Tuning fork Wide stopped organ flues Baroque flute
2	Harmonics dominate	Sweet and soft Rich Splendid Dark Dull Less shrill Bland	French horn, tuba Modern flute Recorder Open organ flues Soft sung sounds
3	Odd harmonics dominate	Hollow Nasal	Clarinet Narrow stopped organ flues
4	Striations dominate	Cutting Rough Bright Brilliant Shrill Brash	Oboe, bassoon Trumpet, trombone Loud sung sounds Bowed instruments Harmonium Organ reeds

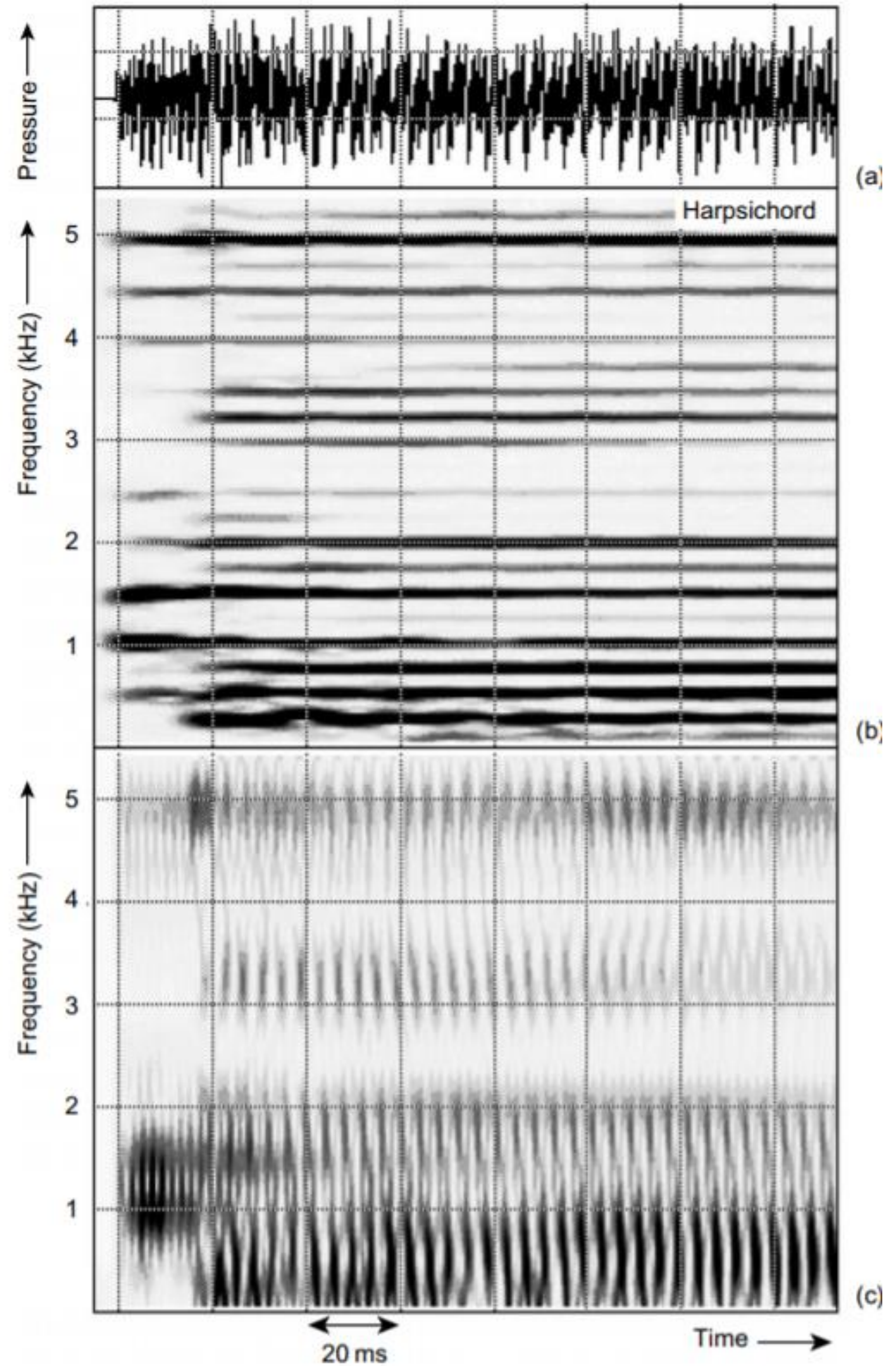
Timbral descriptors and frequency

- Timbral descriptors used to describe the effect of boosting (above) and reducing (below) the spectral energy in various frequency regions set against a keyboard (middle C marked with a spot).



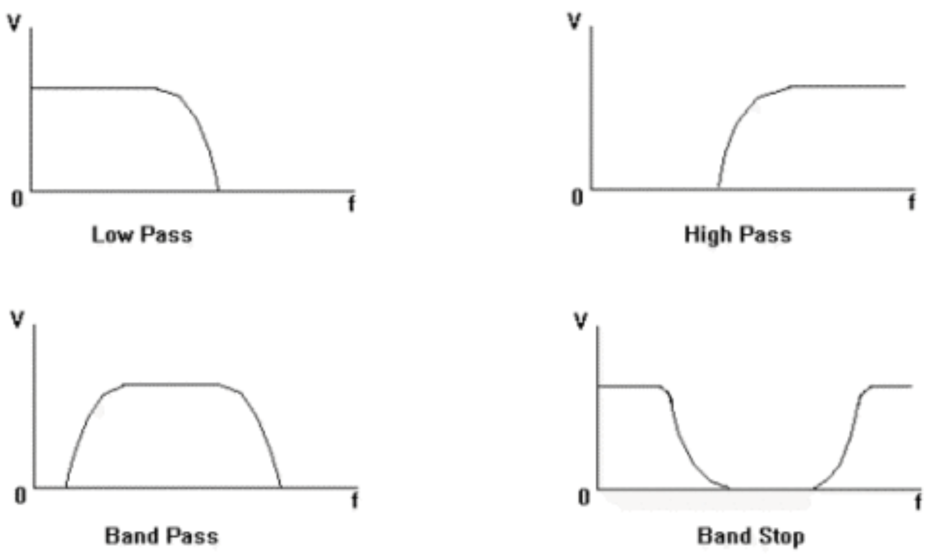
Window size / bandwidth

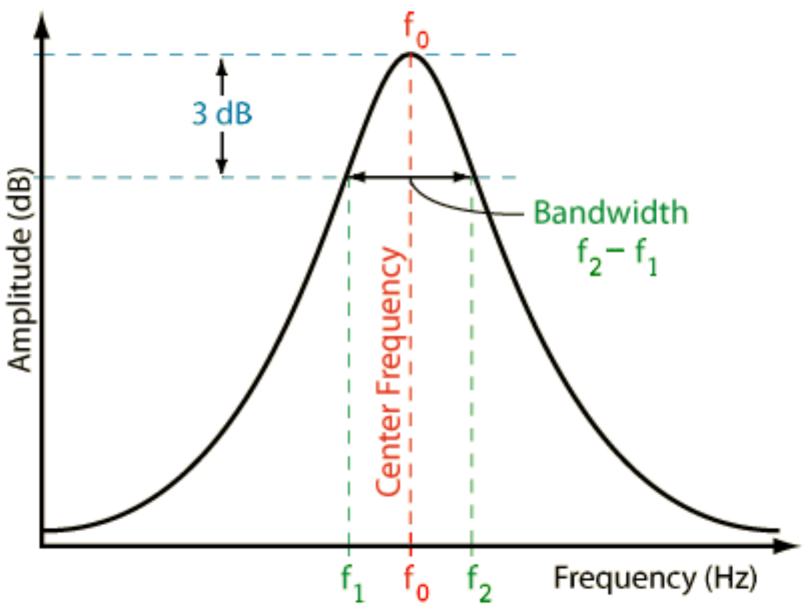
- Acoustic pressure waveform (a) narrowband (40Hz analysis filter) (b) and wide-band (300Hz analysis filter) (c) spectrograms for middle C played on a harpsichord.
- Pitch component
- Striation



Digital audio effects: filter

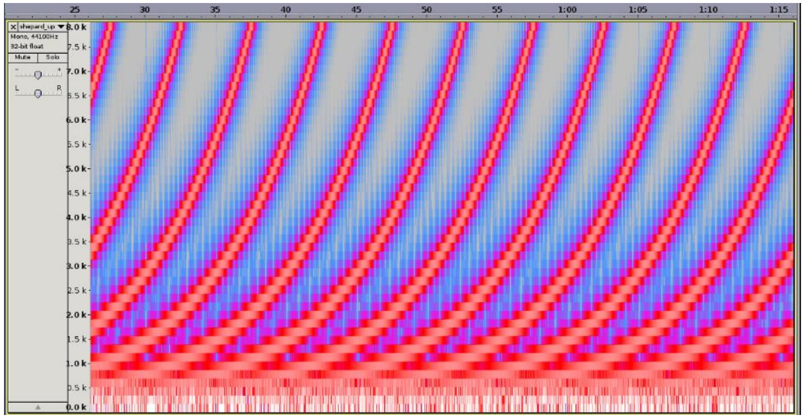
- Suppress or remove specific components in a given frequency band
- Example: what will happen if we use a high-pass filter (e.g., suppress low-frequency components) on a signal?



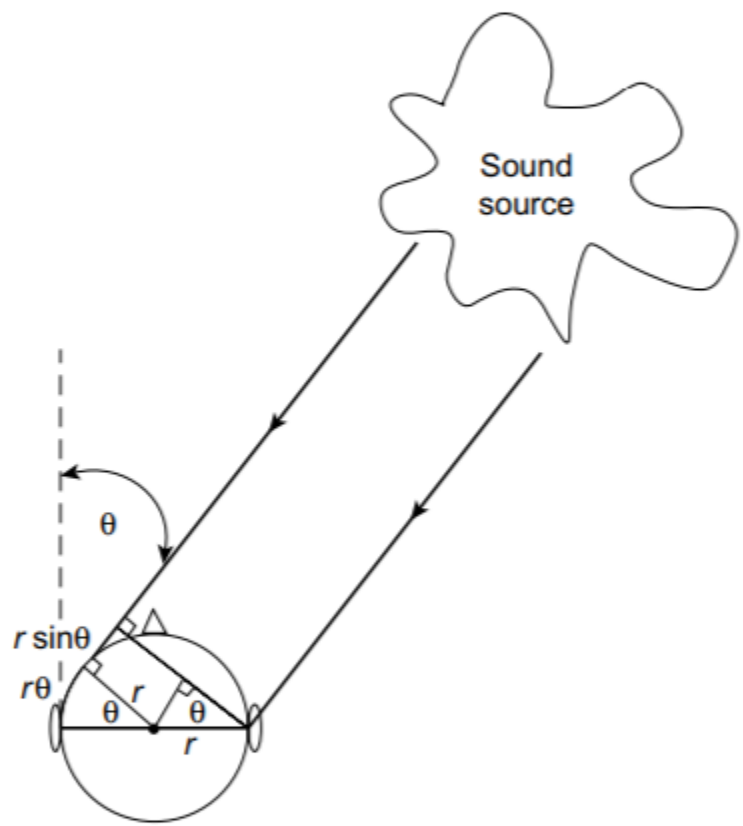


Digital audio effects: flanging

- Flanging: combining two identical signals together, with a small time difference (around 20 ms)
- Behaves like a comb filter
- The history of flanging
- Other audio effects (e.g., phasing, chorus effect, etc.): visit Wikipedia for resources
- "Infinite flanging": the Shepard tone effect (the sonic barber pole)

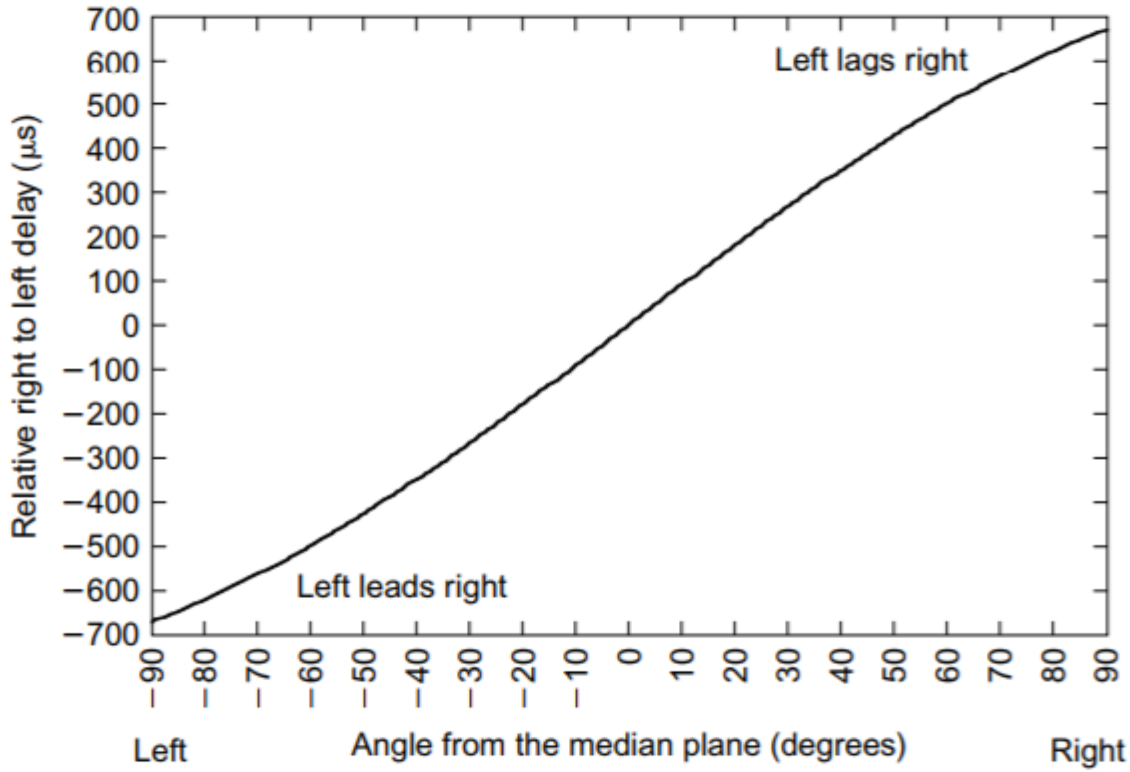


Interaural time difference (ITD)

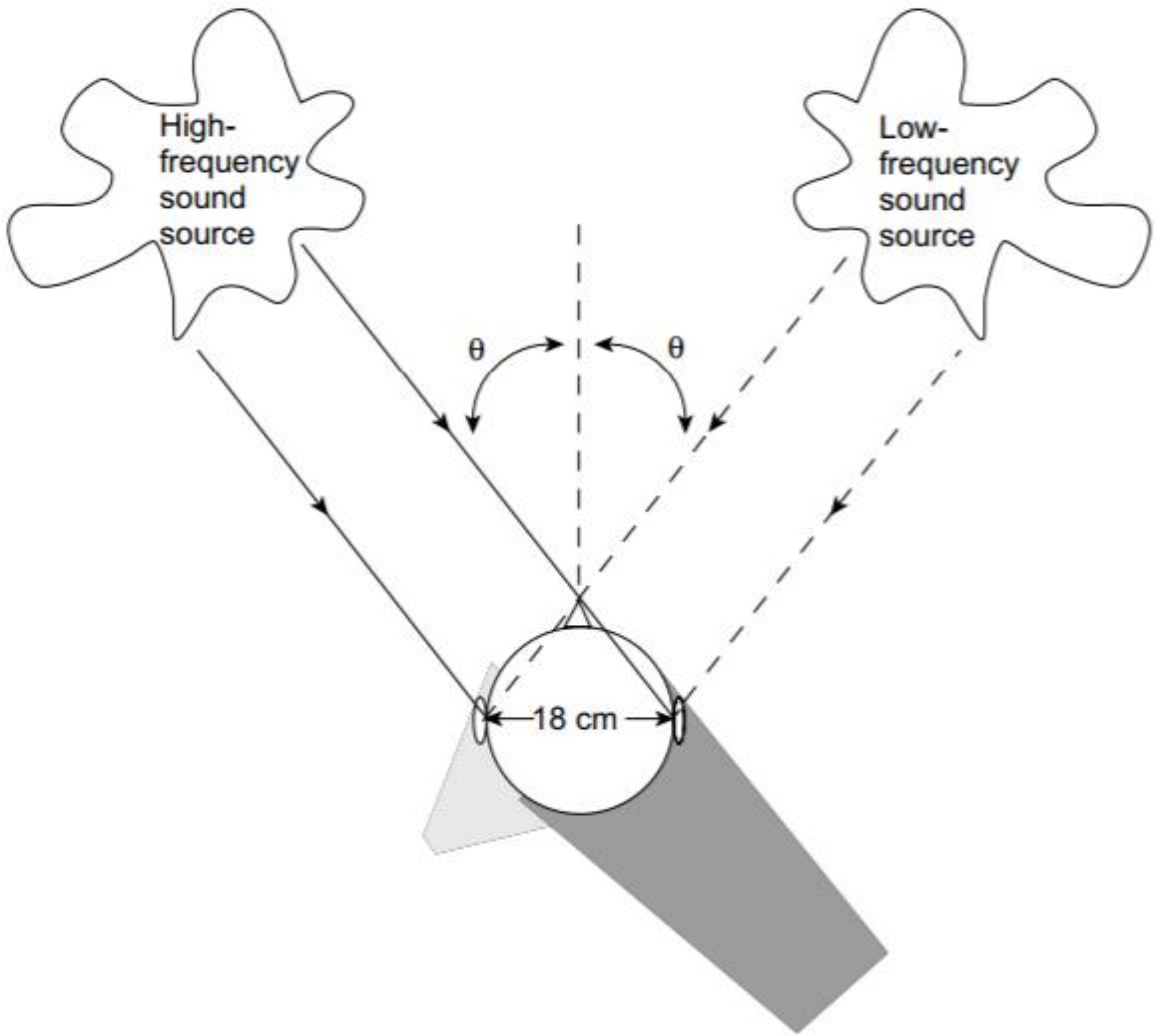


ITD and angle

- The interaural time difference (ITD) as a function of angle

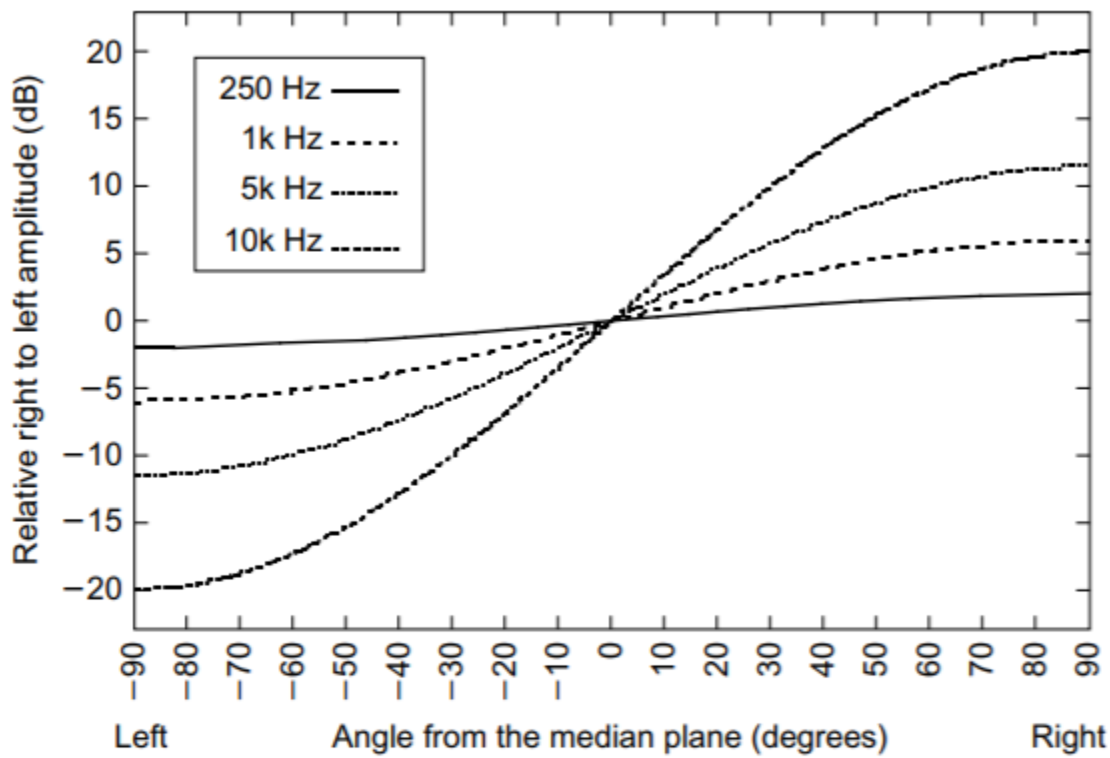


Interaural intensity difference (IID)



IID and angle

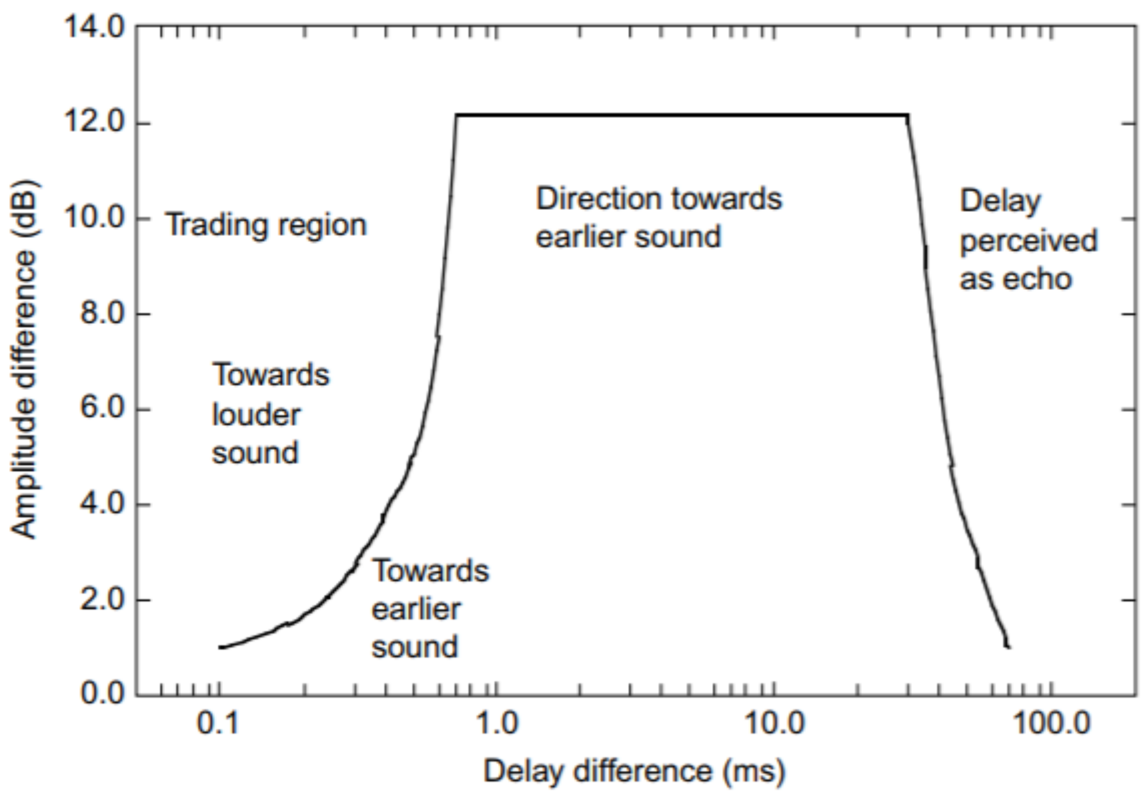
- The interaural intensity difference (IID) as a function of angle and frequency (data from Gulick, 1971)



IID and ITD

- Scattering of sound depends on frequency
- The interaural intensity difference (IID) is a cue for direction at high frequencies
- The interaural time difference (ITD) is a cue for direction at low frequencies

ITD and IID trading



The Haas effect

- Two discrete sounds are interpreted as one sound when their
  - Time delay less than 35 ms
  - Loudness difference less than 10 dB
- Making stereo sound from mono sound
- https://www.youtube.com/watch?v=9Ka7Bq9vzr8