Harvard-MIT Division of Health Sciences and Technology

HST.725: Music Perception and Cognition

Prof. Peter Cariani

HST 725

Music Perception & Cognition

Lecture 2

What we hear:

Basic dimensions of auditory experience

(Image removed due to copyright considerations.)

www.cariani.com

What we hear: dimensions of auditory experience

- Hearing: ecological functions (distant warning, communication, prey detection; works in the dark)
- Detection, discrimination, recognition, reliability, scene analysis
- Operating range: thresholds, ceilings, & frequency limits
- Independent dimensions of hearing & general properties
  - Pitch
  - Timbre (sound quality)
  - Loudness
  - Duration
  - Location
  - Distance and Size
- Perception of isolated pure tones
- Interactions of sounds: beatings, maskings, fusions
- Masking (tones vs. tones, tones in noise)
- Fusion of sounds & the auditory "scene":
  - how many objects/sources/voices/streams?
- Representation of periodicity and spectrum
- Power spectrum and auditory filter metaphors
- Analytical (Helmholtz) vs. Gestalt (Stumpf) perspectives

Hearing: ecological functions

- Distant warning of predators approaching
- Identification of predators
- Localization/tracking of prey
- Con-specific communication
  - Mating/competition
  - Cooperation (info. sharing)
  - Territory
- Navigation in the dark

http://www.pbs.org/wgbh/nova/wolves/

http://www.pbs.org/lifeofbirds/songs/index.html

bat-eared fox
http://www.essex.ac.uk/psychology/hearinglab/index.htm

The auditory scene:
basic dimensions

Temporal organization

- Events
- Notes
- Temporal patterns of events

Organization of sounds

- Voices, instruments
- Streams
- Objects
- Sources

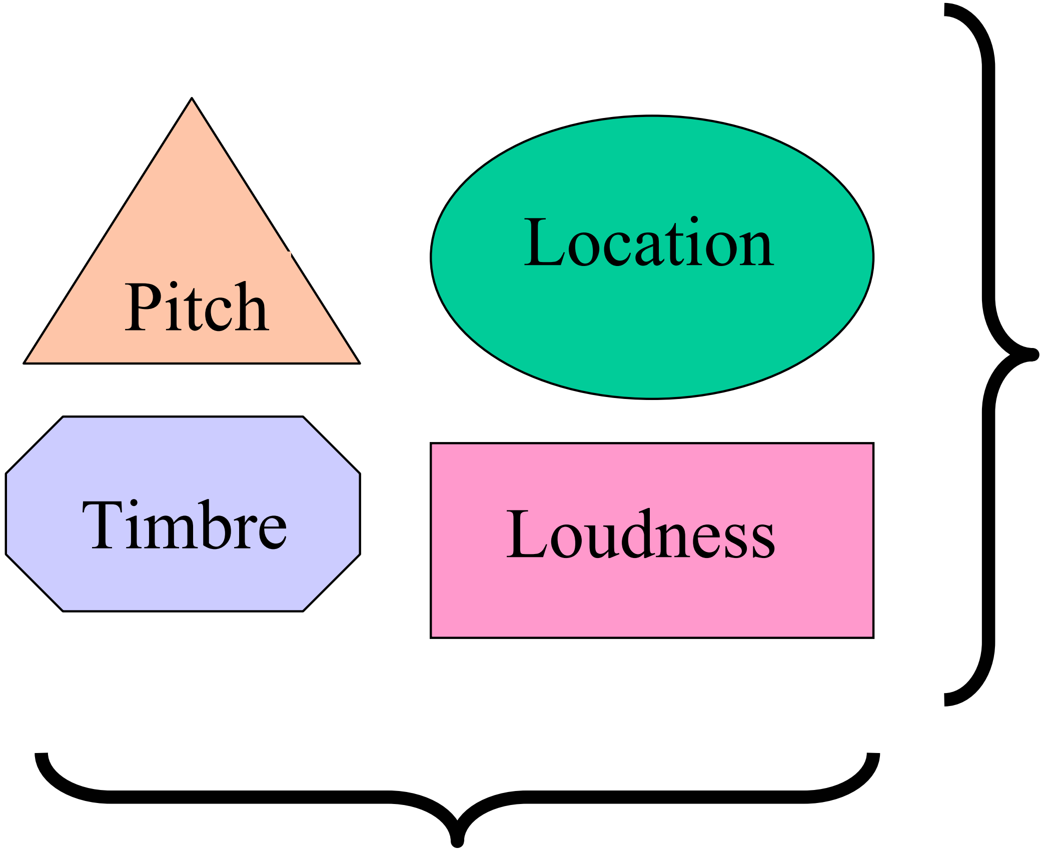
Attributes of sounds

- **Loudness** (intensity)
- **Pitch** (dominant periodicity)
- **Timbre** (spectrum)
- **Duration**
- **Location** (bearing, range)

Auditory qualities in music perception & cognition

• **Pitch**	**Melody, harmony, consonance**
• **Timbre**	**Instrument voices**
• **Loudness**	**Dynamics**
• **Organization**	**Fusions, objects. How many voices?**
• **Rhythm**	**Temporal organization of events**
• **Longer pattern**	**Repetition, sequence**
• **Mnemonics**	**Familiarity, novelty**
• **Hedonics**	**Pleasant/unpleasant**
• **Affect**	**Emotional associations, meanings**
• **Semantics**	**Cognitive associations/expectations**

Basic auditory qualities
Dimensions of auditory perception



TEMPORAL EVENT STRUCTURE

Meter, sequence

FUSION

Grouping into separate objects

Temporal co-occurrence

harmonic structure

John Lurie
Car Cleveland
Music from Stranger than Paradise

Visual scene

Line
Shape
Texture
Lightness
Color
Transparency

Objects
Apparent distance
Apparent size
etc.

LIFE MAGAZINE COVER,
Margaret Bourke-White
Fort Peck Dam, Montana (1^{st} Life Cover)
November 23, 1936.

Sound level basics

- Sound pressure levels are measured relative to an absolute reference
- (re: 20 micro-Pascals, denoted Sound Pressure Level or SPL).
- Since the instantaneous sound pressure fluctuates, the average amplitude of the pressure waveform is measured using root-mean-square RMS. (Moore, pp. 9-12)
- Rms(x) = sqrt(mean(sum(x_t^2)))
  - Where x_t is the amplitude of the waveform at each instant t in the sample
  - Because the dynamic range of audible sound is so great, magnitudes are expressed in a logarithmic scale, decibels (dB).
- A decibel of amplitude expresses the ratio of two amplitudes (rms pressures, P1 and P_reference) and is given by the equation:

  dB = 20 * log10(P1/P_reference)

  20 dB = 10 fold change in rms level

Decibel scale for relative amplitudes (levels) (rules of thumb)

20 dB = fold change amplitude

10 dB = 3+ fold change

6 dB = 2 fold change amplitude

3 dB = 1.4 fold change

2 dB = 1.26 fold change (26 %)

1 dB = 1.12 fold change (12%)

0 dB = 1 fold change (no change)

-6 dB = 1/2

-20 dB = 1/10 fold change

Perceptual functions

Subjective vs. objective measures
Subjective measures
Magnitude estimation
Objective measures
Detection: capability of distinguishing the presence or absence of a stimulus (or some aspect of a stimulus, e.g. AM detection)
Threshold: the value of a stimulus parameter at which a stimulus can be reliably detected
Sensation level (SL): sound level re: threshold

Discrimination: capability of distinguishing between two stimuli
Difference limen: the change in a stimulus parameter required for reliable discrimination, just-noticeable-difference (jnd)
Weber fraction: Difference limen expressed as proportional change (e.g. $\Delta f/f$)

Matching task
Two-alternative forced choice (2AFC)

Recognition: correct identification of a particular stimulus

Dynamic range

0 dB SPL is set at 20 microPascals
60 dB SPL is therefore a 1000 fold change in RMS over 0 dB

A typical background sound level is 50-60 dB SPL.

Dynamic range describes the range of sound pressure levels.

The auditory system registers sounds from 20 dB to >> 120 dB SPL

The auditory system has a dynamic range in excess of 100 dB (!) or a factor of 10^5 = 100,000 in amplitude.

It is quite remarkable that musical sounds remain recognizable over most of this range. This a fundamental aspect of hearing that all auditory theories must address -- how auditory percepts remain largely invariant over this huge range (perceptual constancy).

Typical sound levels in music

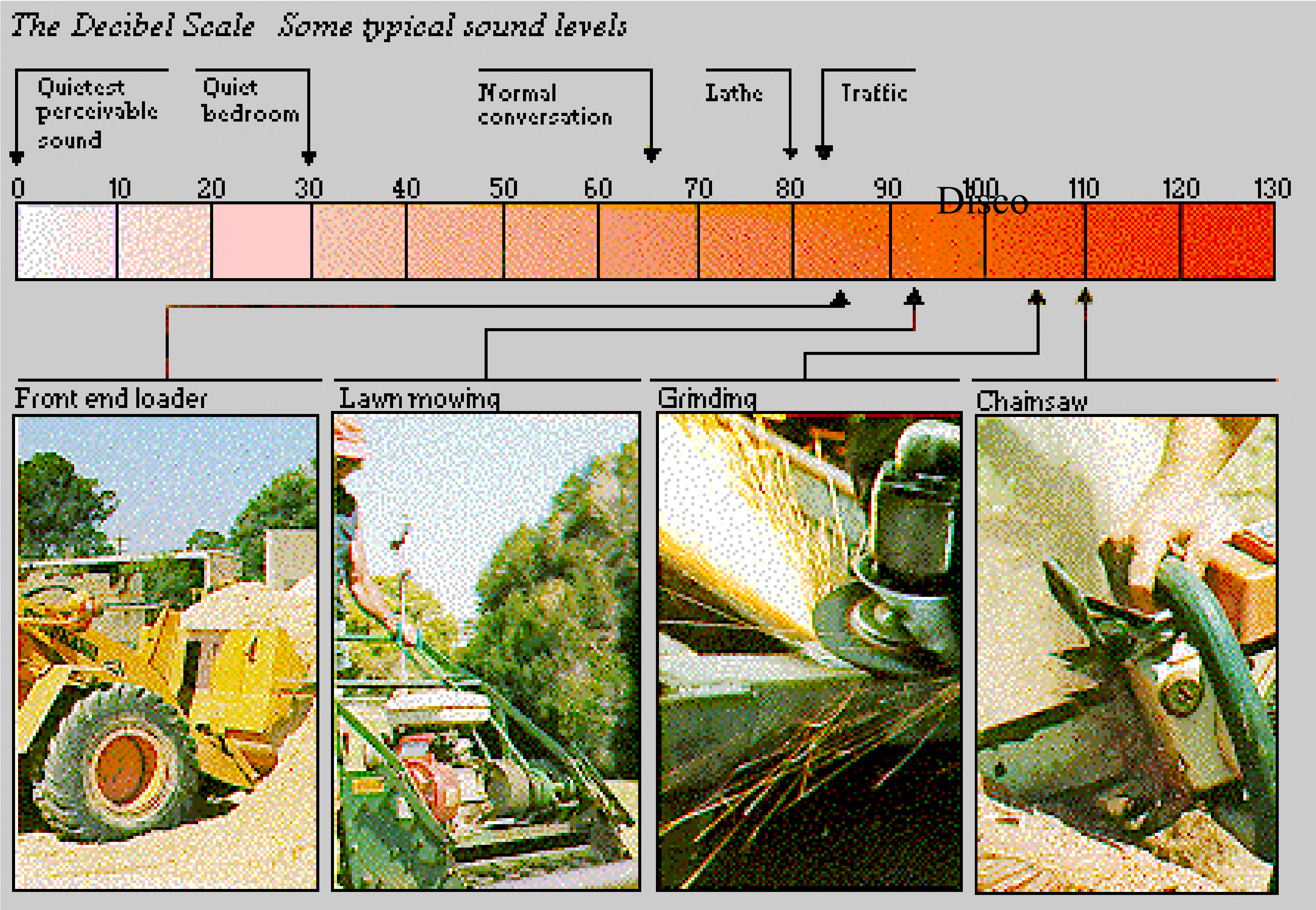
• Pain	> 130 dB SPL
• Loud rock concert	120 dB SPL
• Loud disco	110 dB SPL
• *fff*	100 dB SPL
• *f (forte, strong)*	80 dB SPL
• *p (piano, soft)*	60 dB SPL
• *ppp*	40 dB SPL
• Lower limit	
• Theshold of hearing	0 dB SPL

- Musical notation ranges from Pierce, Science of Musical Sound, p. 325

On origins of music dynamics notation
http://www.wikipedia.org/wiki/Pianissimo

In music , the word **dynamics** refers to the volume of the sound . The renaissance composer Giovanni Gabrieli was one of the first to indicate dynamics in music notation .The two basic dynamic indications in music are *piano* , meaning "softly" or "quietly", usually abbreviated as ***p***; and *forte* , meaning "loudly" or "strong", usually abbreviated as ***f***. More subtle degrees of loudness or softness are indicated by ***mp*** , standing for *mezzo-piano* , and meaning "half-quiet"; and ***mf*** ,*mezzo-forte* , "half loud". Beyond ***f*** and ***p***, there is ***ff*** , standing for "fortissimo", and meaning "very loudly"; and ***pp*** , standing for "pianissimo", and meaning "very quietly". To indicate even more extreme degrees of intensity, more ***p***s or ***f***s are added as required. ***fff*** (*fortississimo*) and ***ppp*** (*pianississimo*) are found in sheet music quite frequently, but more than three ***f***s or ***p***s is quite rare. It is sometimes said that ***pppp*** stands for *pianissississimo* , but such words are very rarely used either in speech or writing, even when present in a score. There is some evidence that this use of an increasing number of letters to indicate greater extremes of volume stems from a convention dating from the 17th century where ***p*** stood for *piano* ,***pp*** stood for *più piano* (literally "more quietly") and, by extension, ***ppp*** indicated *pianissimo* .Antonio Vivaldi seems to have written using this convention, but it was largely replaced by the above, more familiar, system by the middle of the 18th century .

Typical sound pressure levels in everyday life



(Reproduced courtesy of WorkSafe, Department of Consumer and Employment Protection, Western Australia (www.safetyline.wa.gov.au).
The graphic being that at the bottom of: http://www.safetyline.wa.gov.au/institute/level2/course18/lecture54/l54_03.asp)

Demonstrations

- **Demonstrations using waveform generator**
- Relative invariance of pitch & timbre with level
- Loudness matching
- Pure tone frequency limits
- Localization

Loudness

Dimension of perception that changes with sound intensity (level)

Intensity ~ power;

Level~amplitude

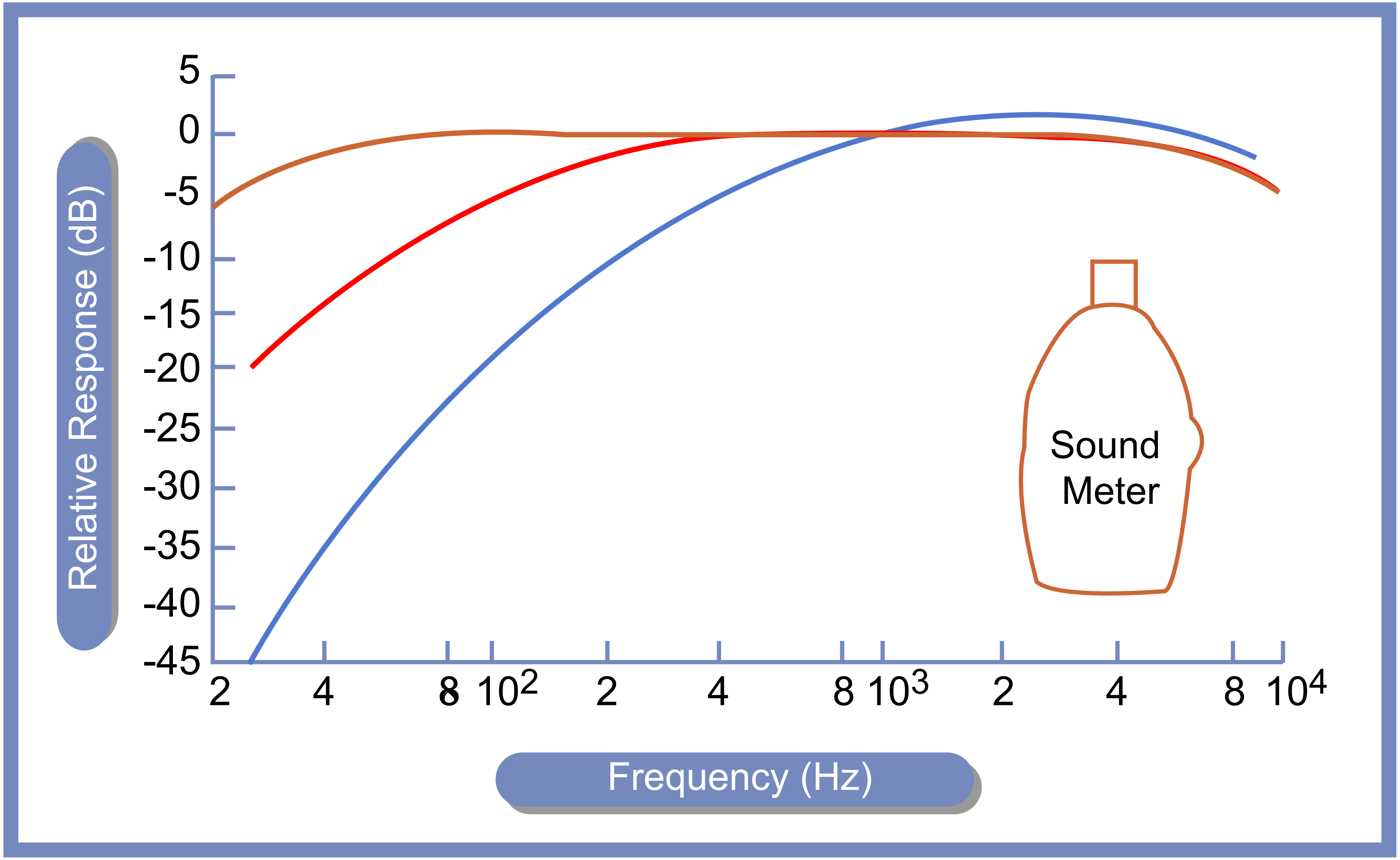
Demonstration using waveform generator

Masking demonstrations

Magnitude estimation

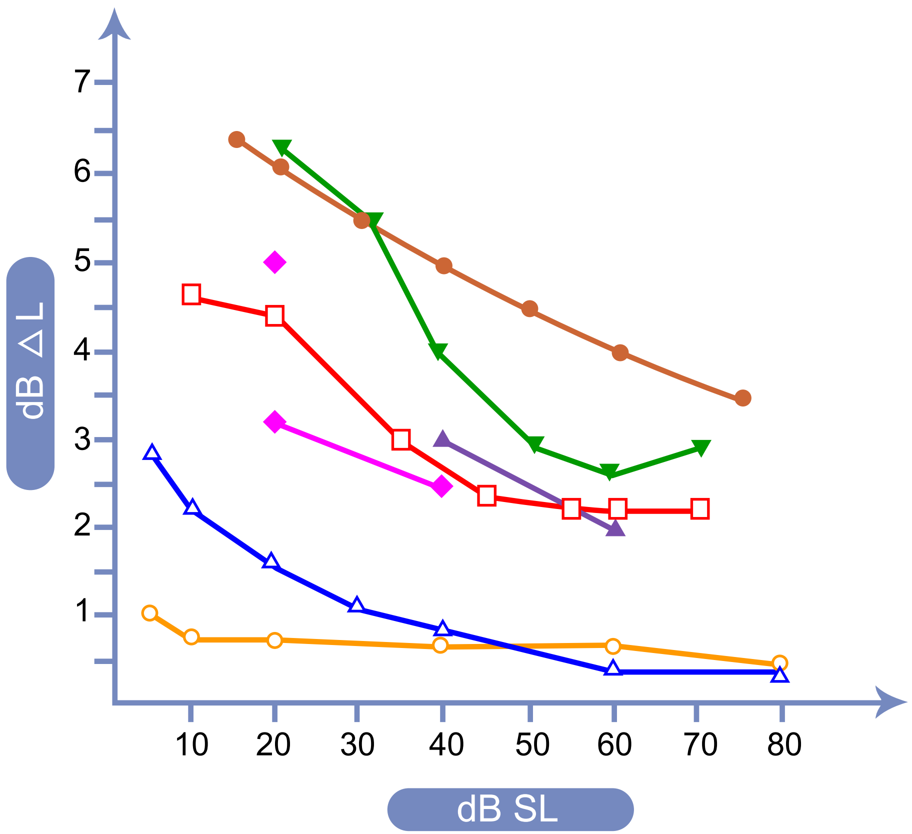
Loudness matching

Sound level meters and frequency weightings



Intensity discrimination improves at higher sound levels

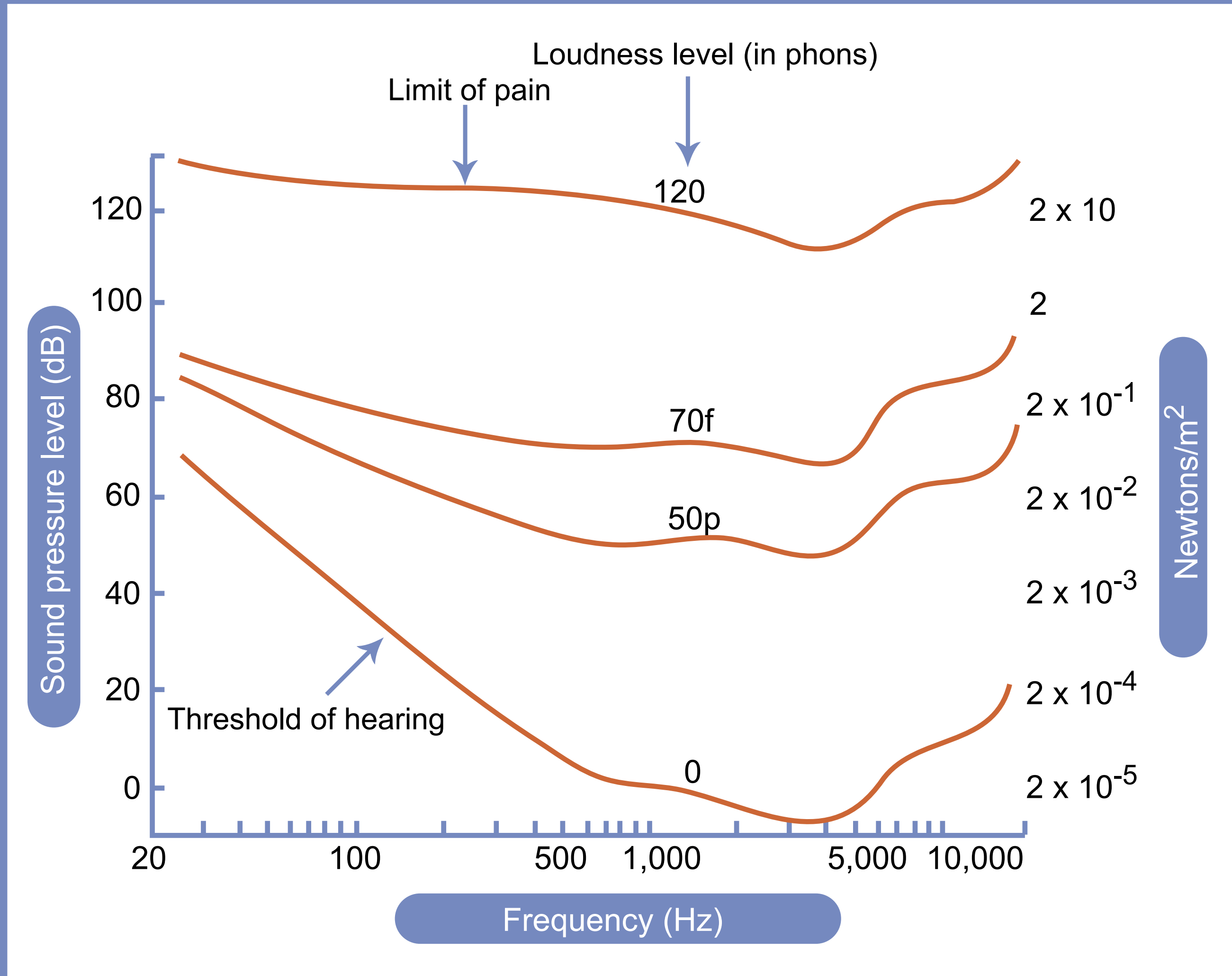
Best Weber fraction $\Delta L/L$ is about 1 dB



A comparison of just noticeable intensity differences (averaged across frequencies) for various species. Man (open symbols): red (Dimmick & Olson, 1941), orange (experiment I), blue (experiment III, Harris, 1963); cat: purple (Raab & Ades,1946; Elliott & McGee, 1965); rat: pink (Henry, 1938; Hack, 1971); mouse: brown (Ehret, 1975b); parakeet: green (Dooling & Saunders, 1975b).

Figure adapted from cited sources above.

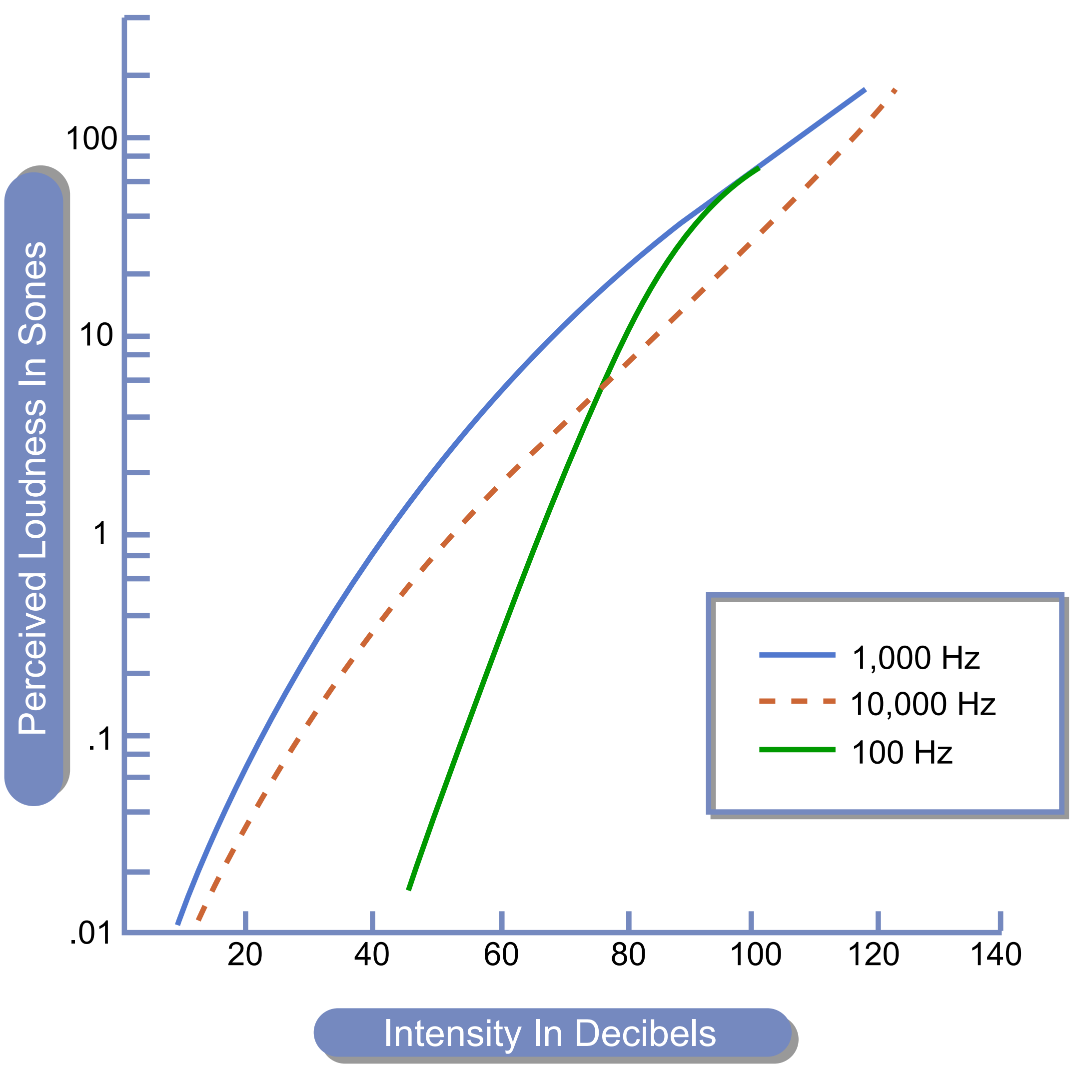
Loudness as a function of pure tone level & frequency



Constant-loudness curves for persons with acute hearing. All sinusoidal sounds whose levels lie on a single curve (an *isophon*) are equally loud. A particular loudness-level curve is designated as a loudness level of some number of *phons.* The number of phons is equal to the number of decibels only at the frequency 1,000 Hz.

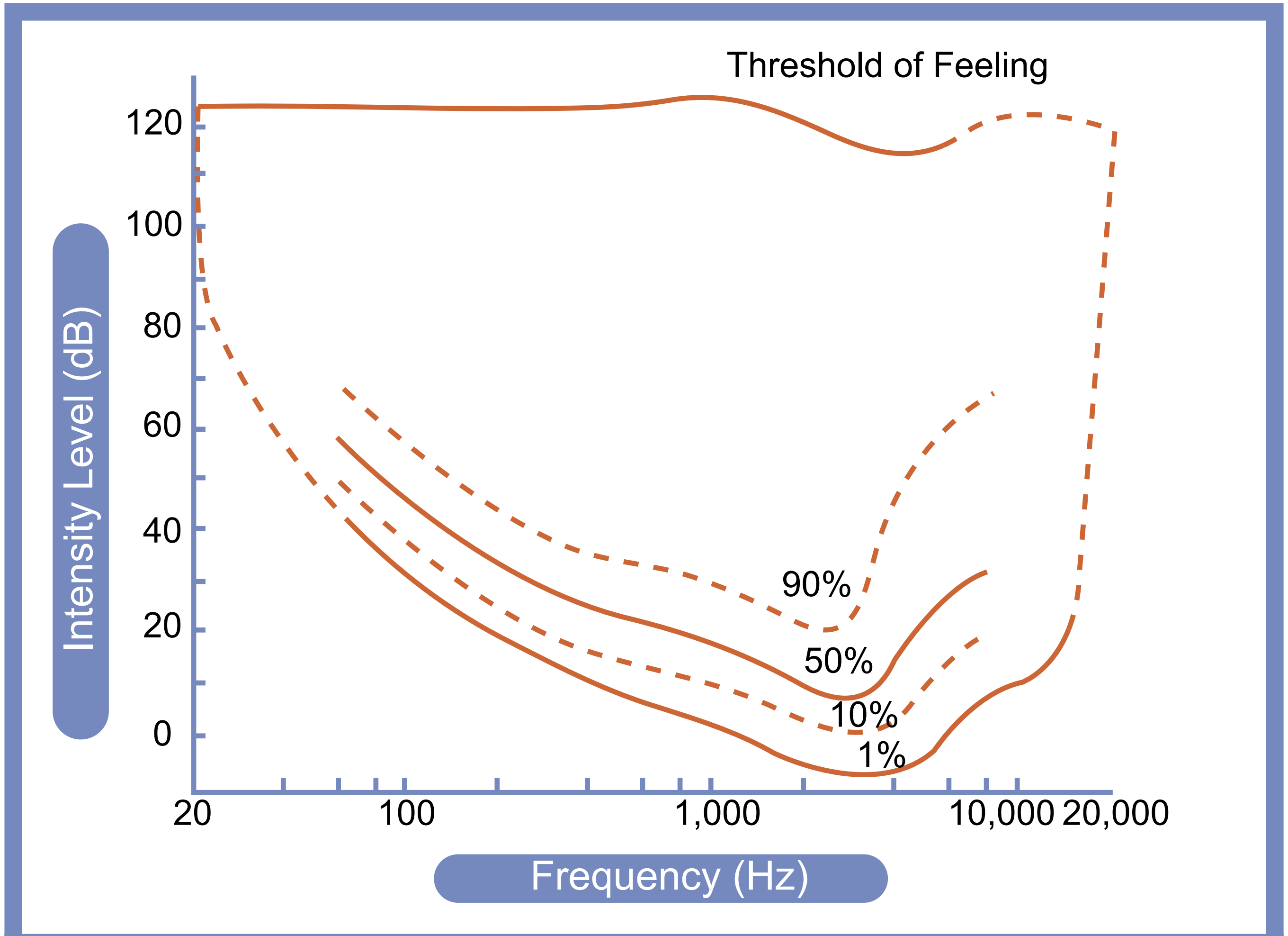
Absolute detection thresholds on the order of
1 part in a million,
Δ pressure ~1/1,000,000 atm
(Troland, 1929)

Loudness perception:
perceived growth of loudness w. level



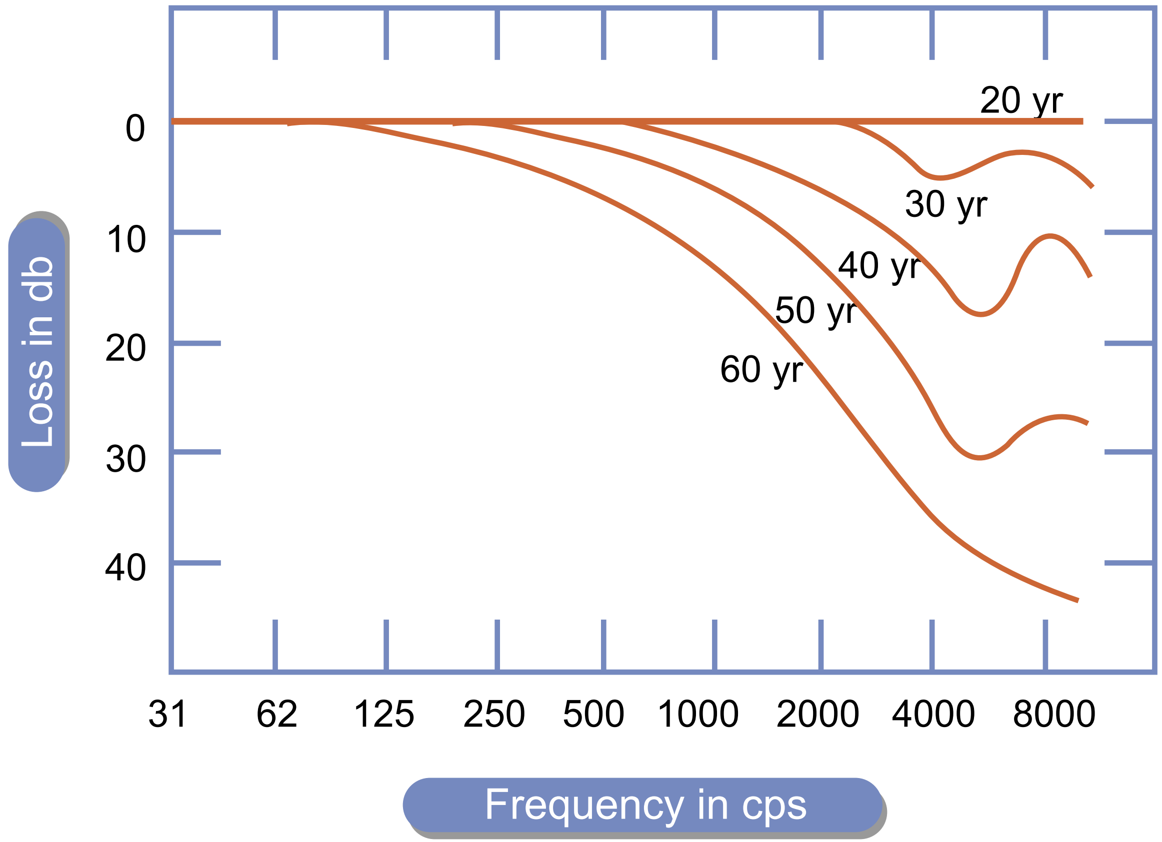
Perceived loudness of tones of various frequencies as a function of physical intensity.

Loudness perception: population percentiles



Curves showing threshold of hearing at various frequencies for a group of Americans: 1 percent of the group can hear any sound with an intensity above the 1 percent curve; 5 percent of the group can hear any sound with an intensity above the 5 percent curve; and so on.

Hearing loss with age



Progressive loss of sensitivity at high frequencies with increasing age.
The audiogram at 20 years of age is taken as a basis of comparison.

(From Morgan, 1943, after Bunch, 1929.)

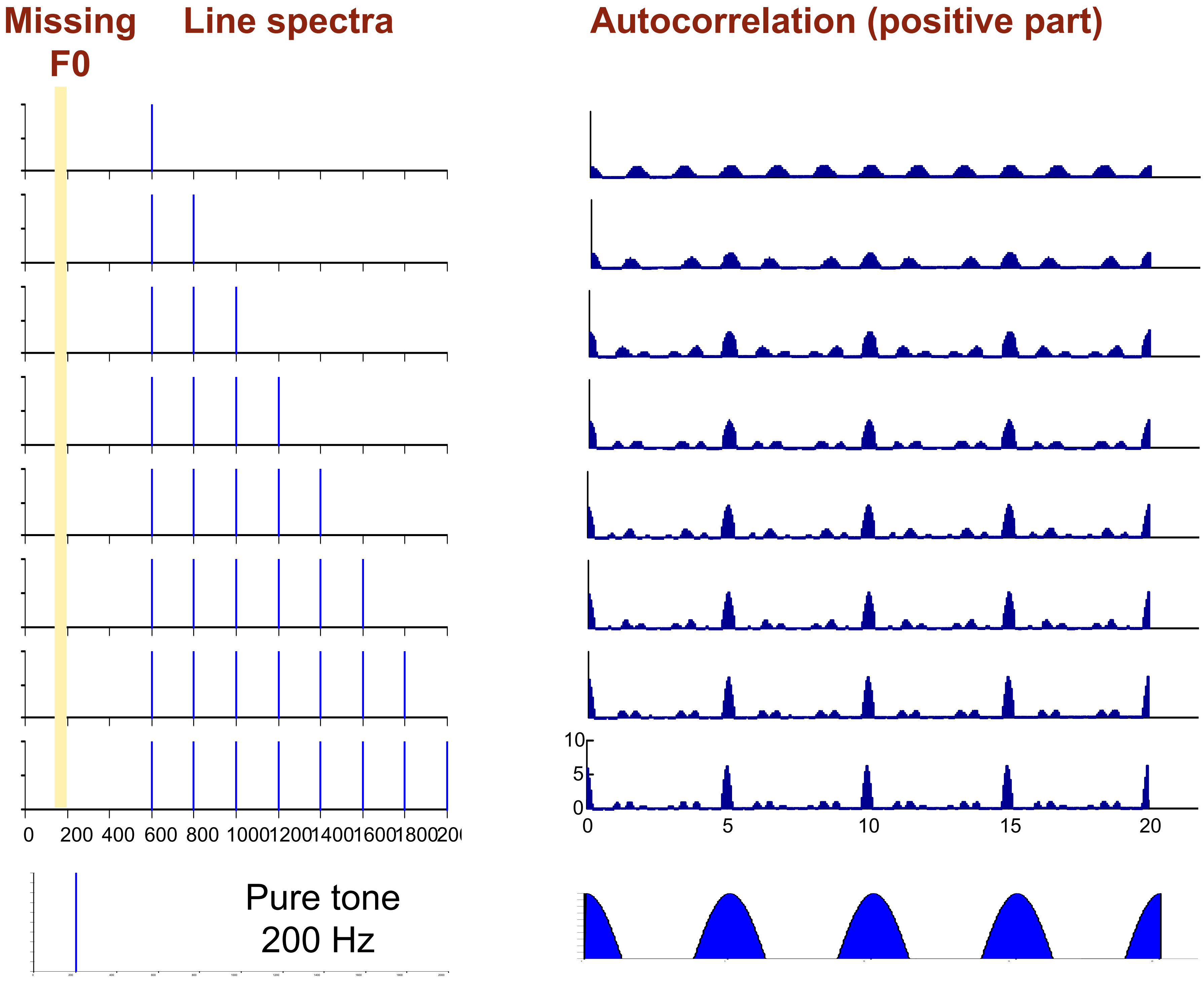
Dynamic range of some musical instruments

Please see Figure 8.5 in The science of musical sound. John R. Pierce. Edition: Rev. ed. Published: New York: Freeman, c1922. ISBN: 0716760053.

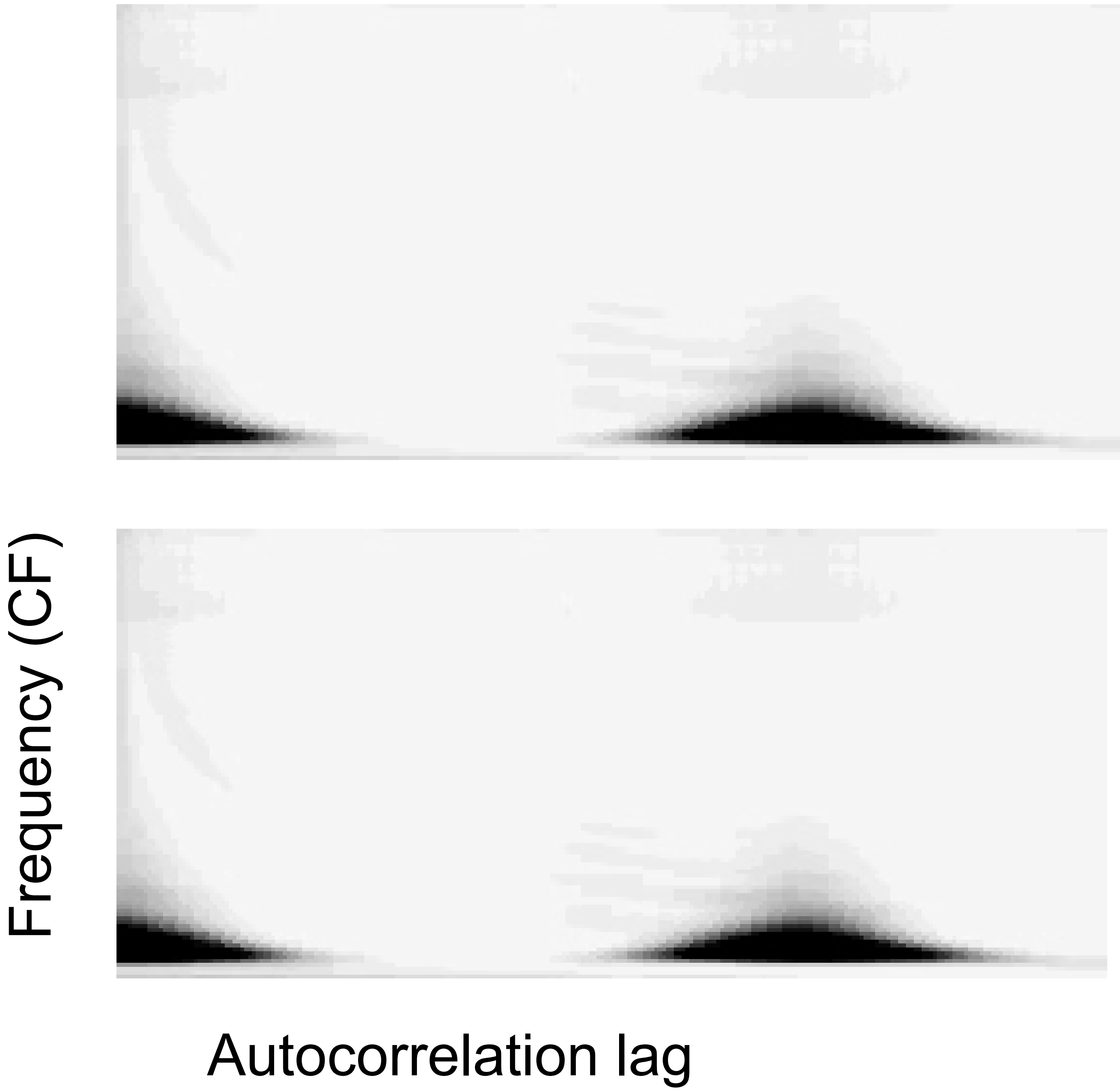
Range of pitches of pure & complex tones

- **Pure tone pitches**
  - Range of hearing (~20-20,000 Hz)
  - Range in tonal music (100-4000 Hz)
- **Most (tonal) musical instruments produce harmonic complexes that evoke pitches at their fundamental frequencies (F0's)**
  - Range of F0's in tonal music (30-4000 Hz)
  - Range of missing fundamental (30-1200 Hz)

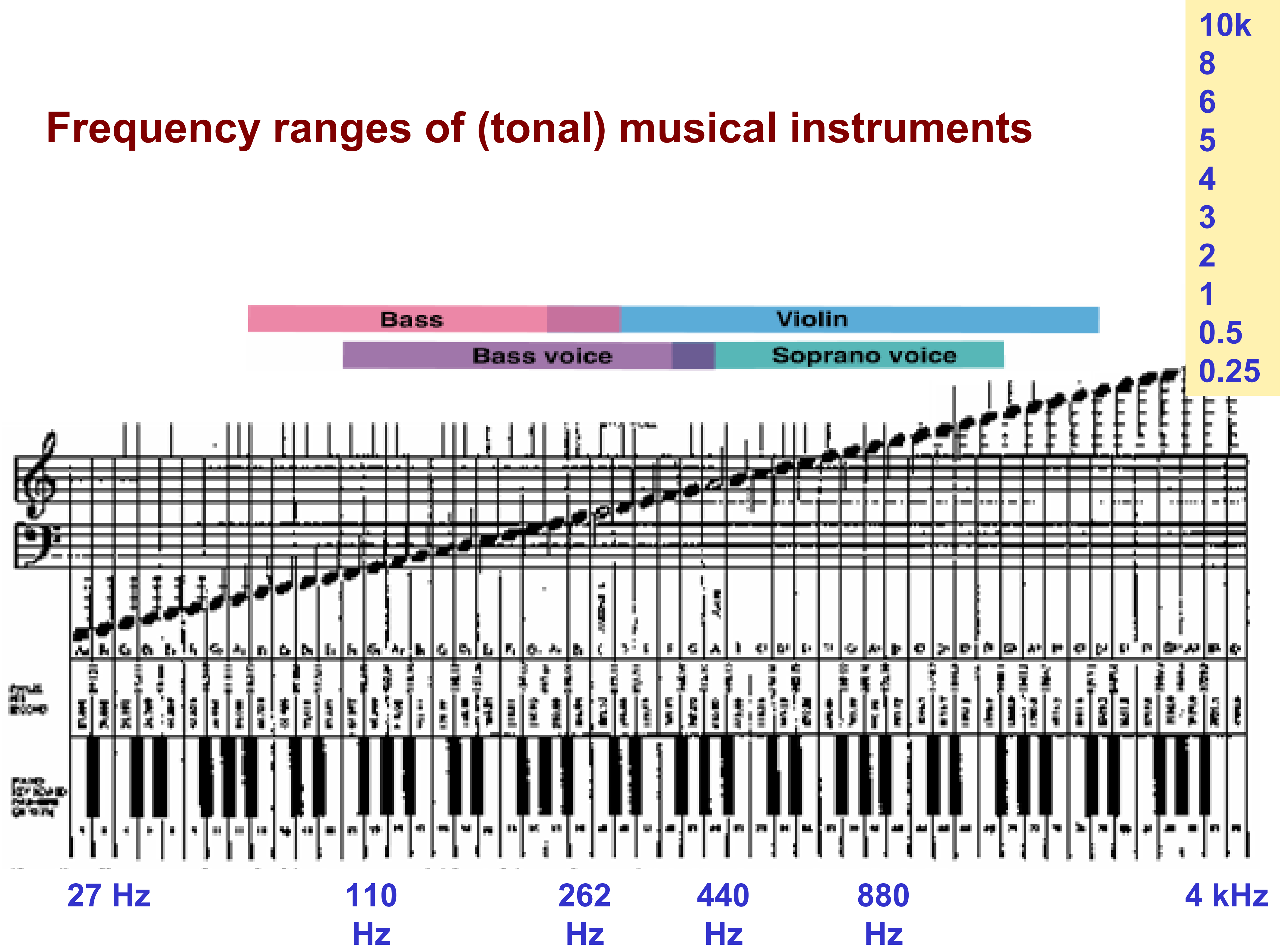
Emergent pitch



Correlograms: interval-place displays (Slaney & Lyon)

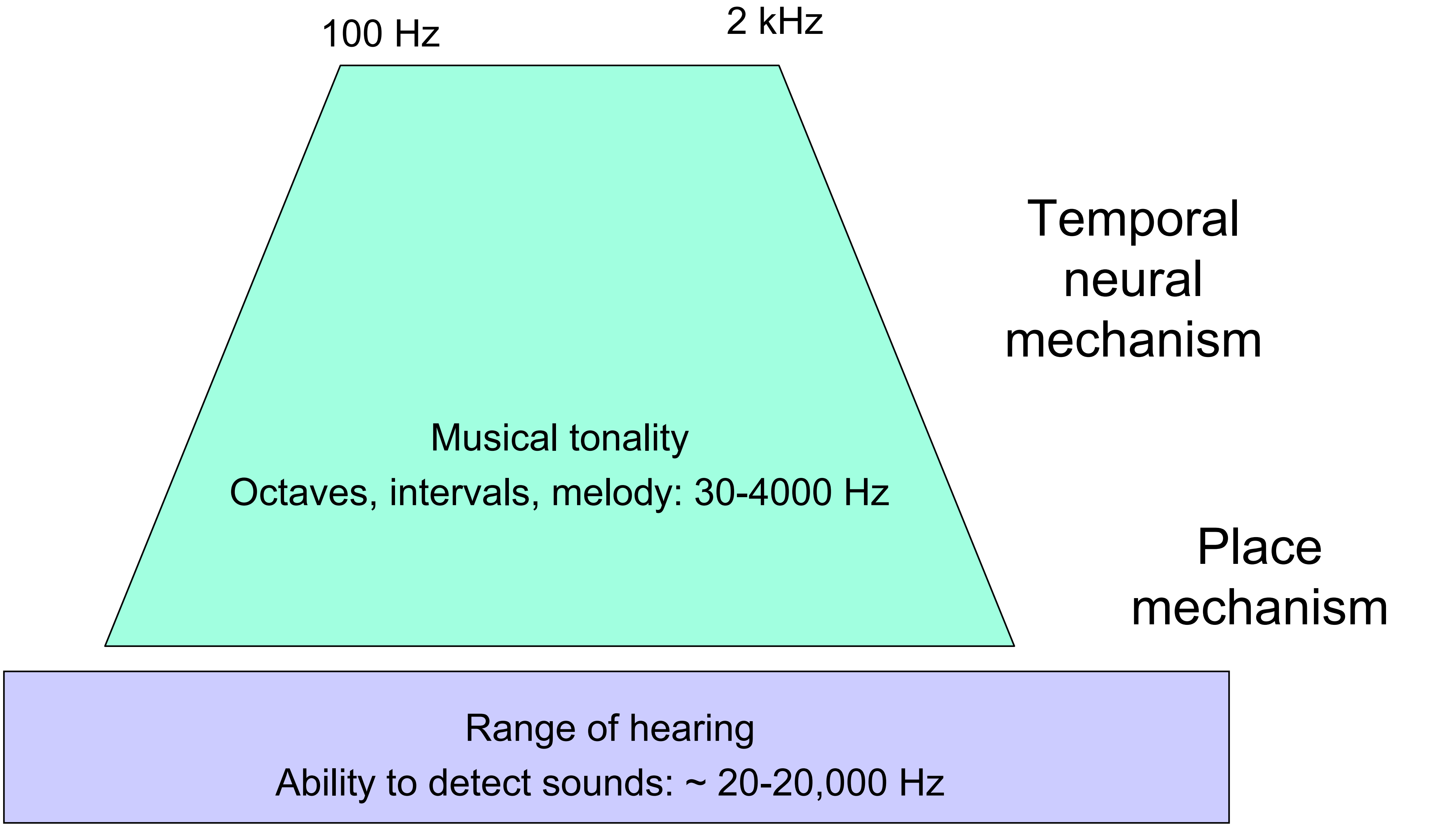


Frequency ranges of (tonal) musical instruments

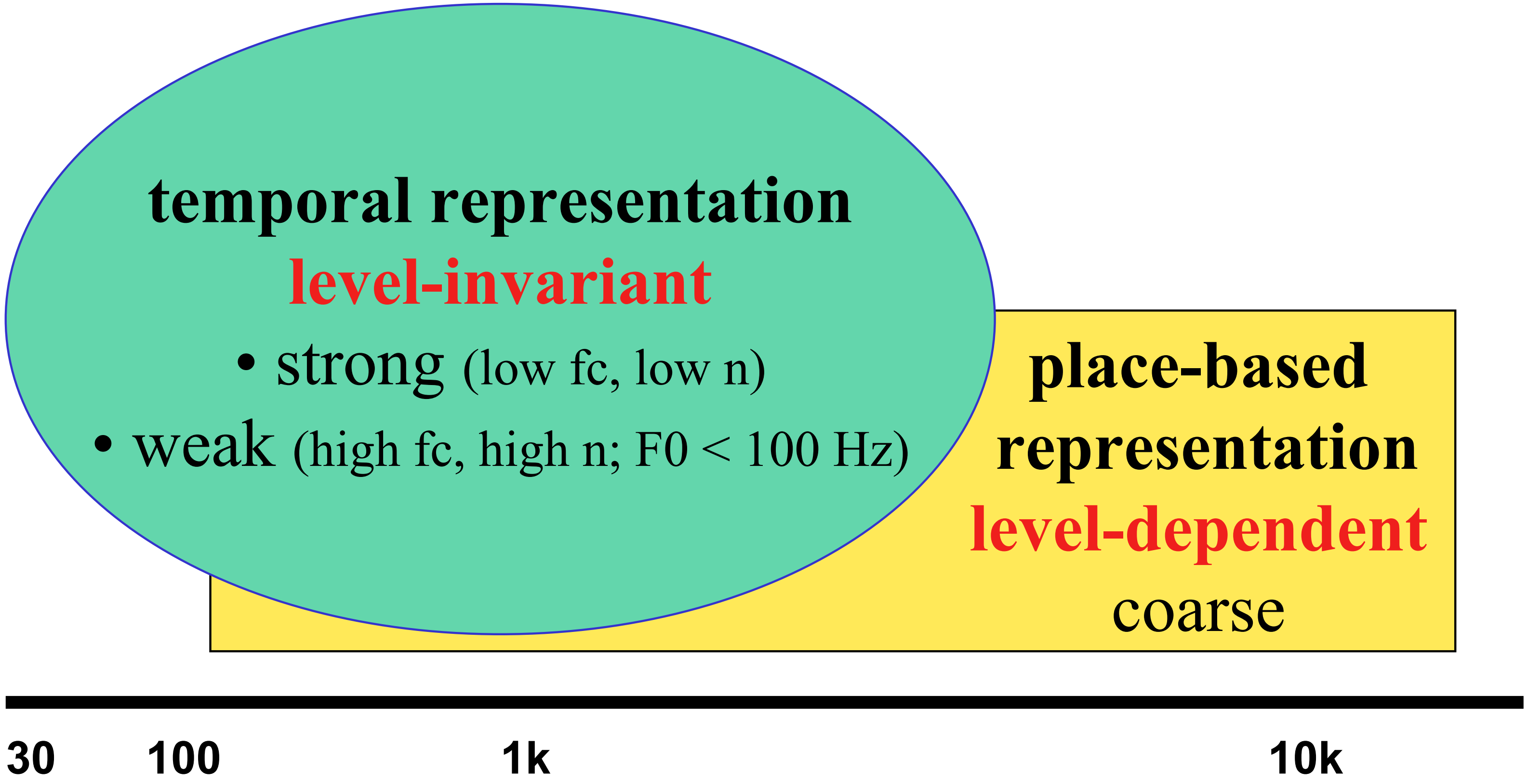


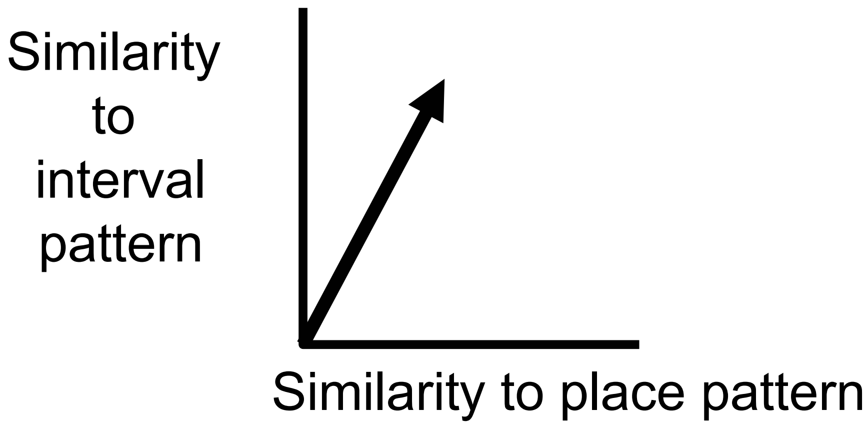
Frequency ranges: hearing vs. musical tonality

(Courtesy of Malcolm Slaney (Research Staff Member of IBM Corporation). Used with permission.)



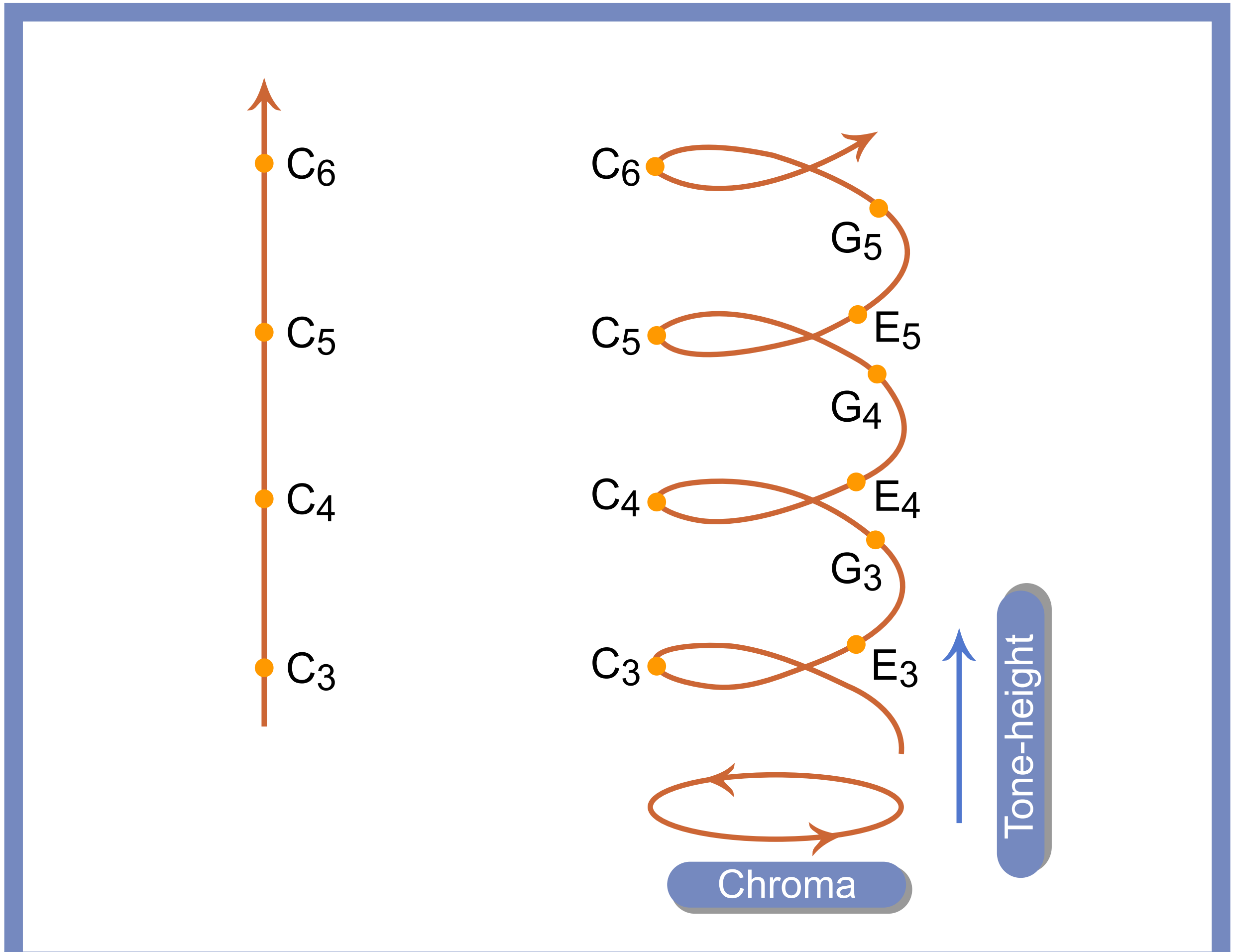
Duplex time-place representations



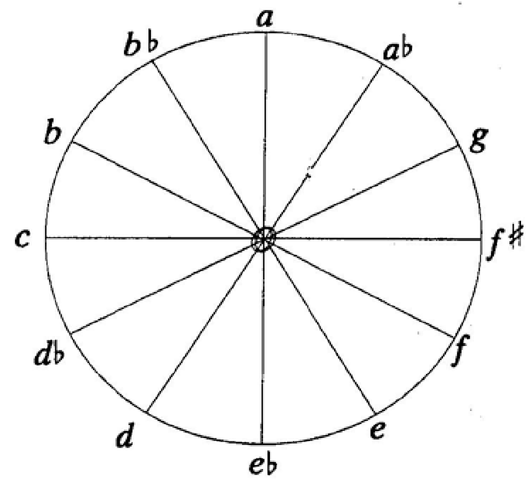


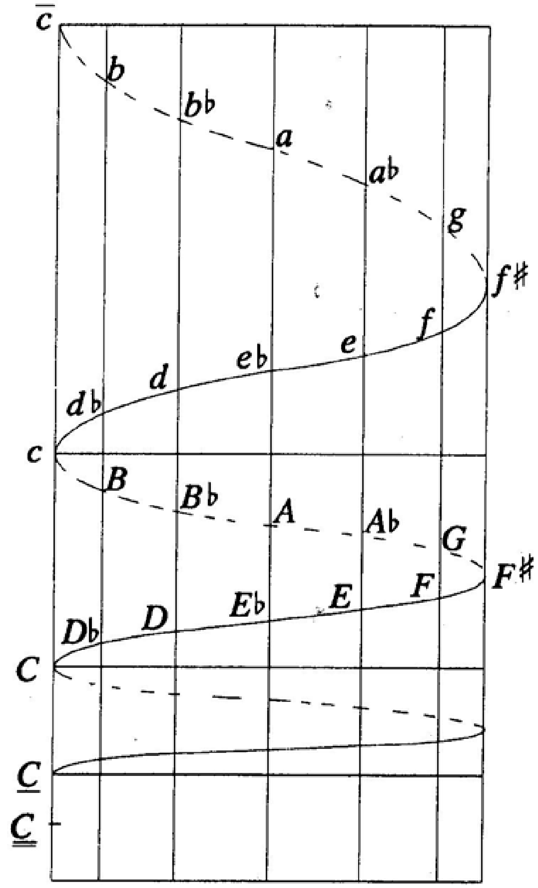
<u>cf. Terhardt's spectral and virtual pitch</u>

Pitch dimensions: height & chroma



Contrast between one-dimensional and two-dimensional models of pitch perception. Notes of a scale played on an ordinary instrument spiral upward around the surface of a cylinder, but computer-generated notes can form a Shepard scale that goes around in circle.



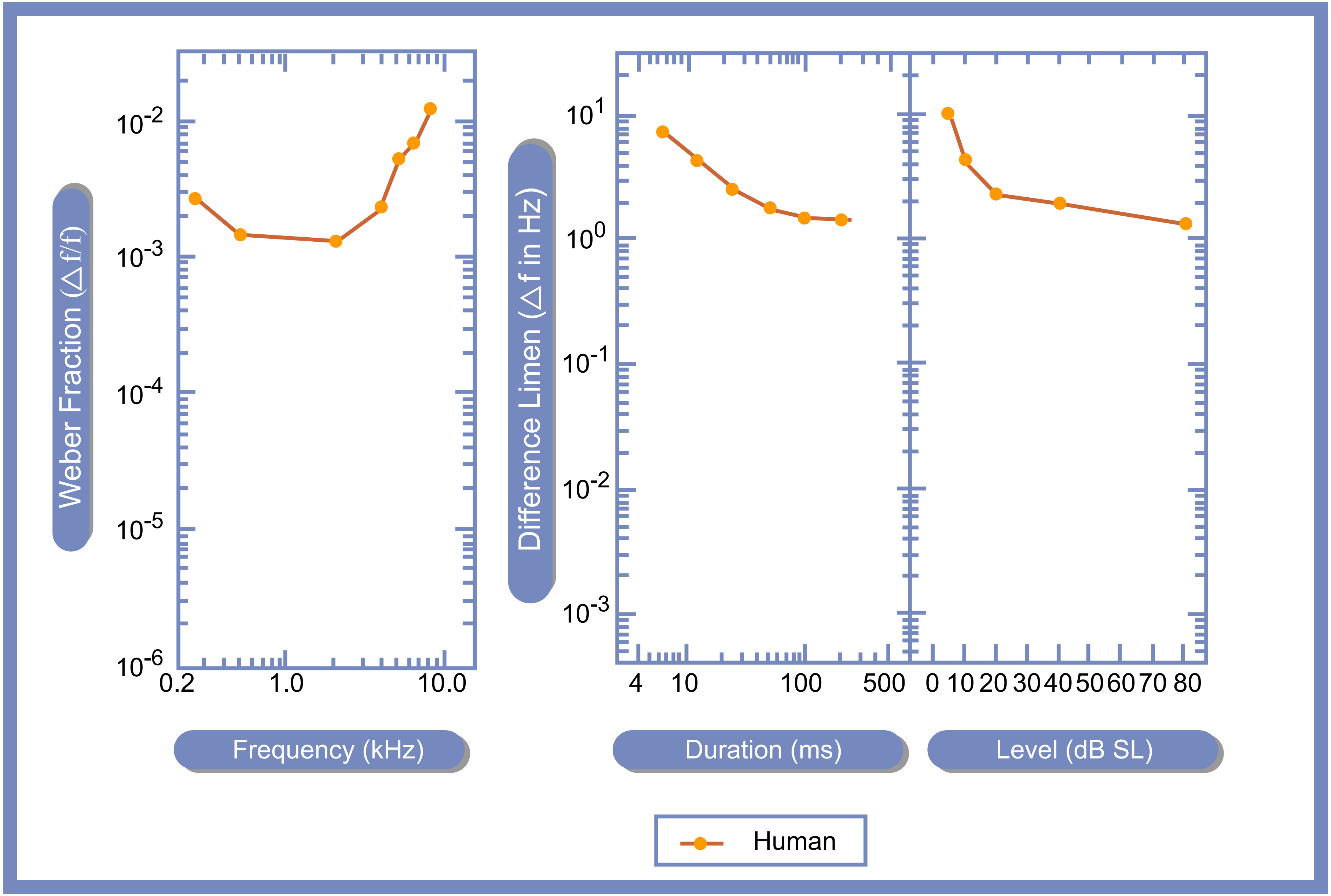


Chroma helix after Drobisch (1852).

Pitch height and pitch chroma

Please see figures 1, 2, and 7 in Roger N. Shepard. Geometrical approximations to the structure of musical pitch. Psychological Review 89 (4): 305-322, 1982.

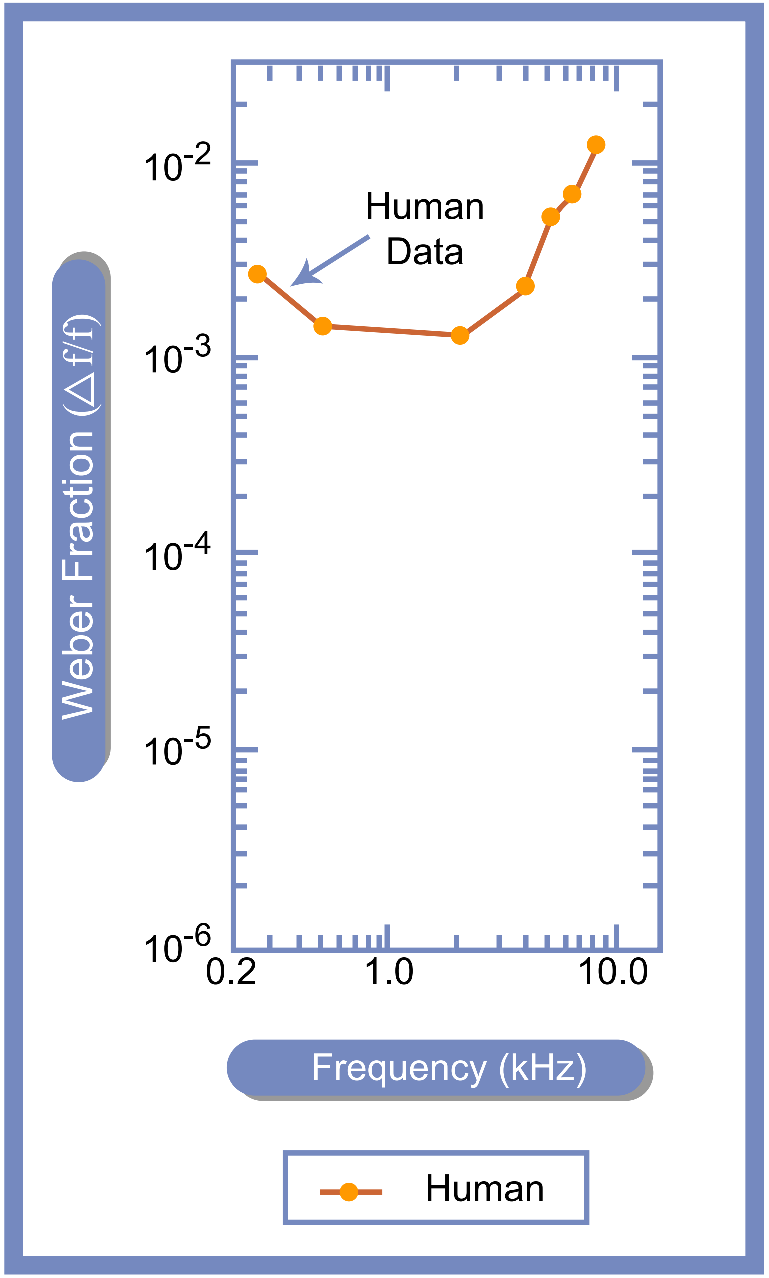
JND's



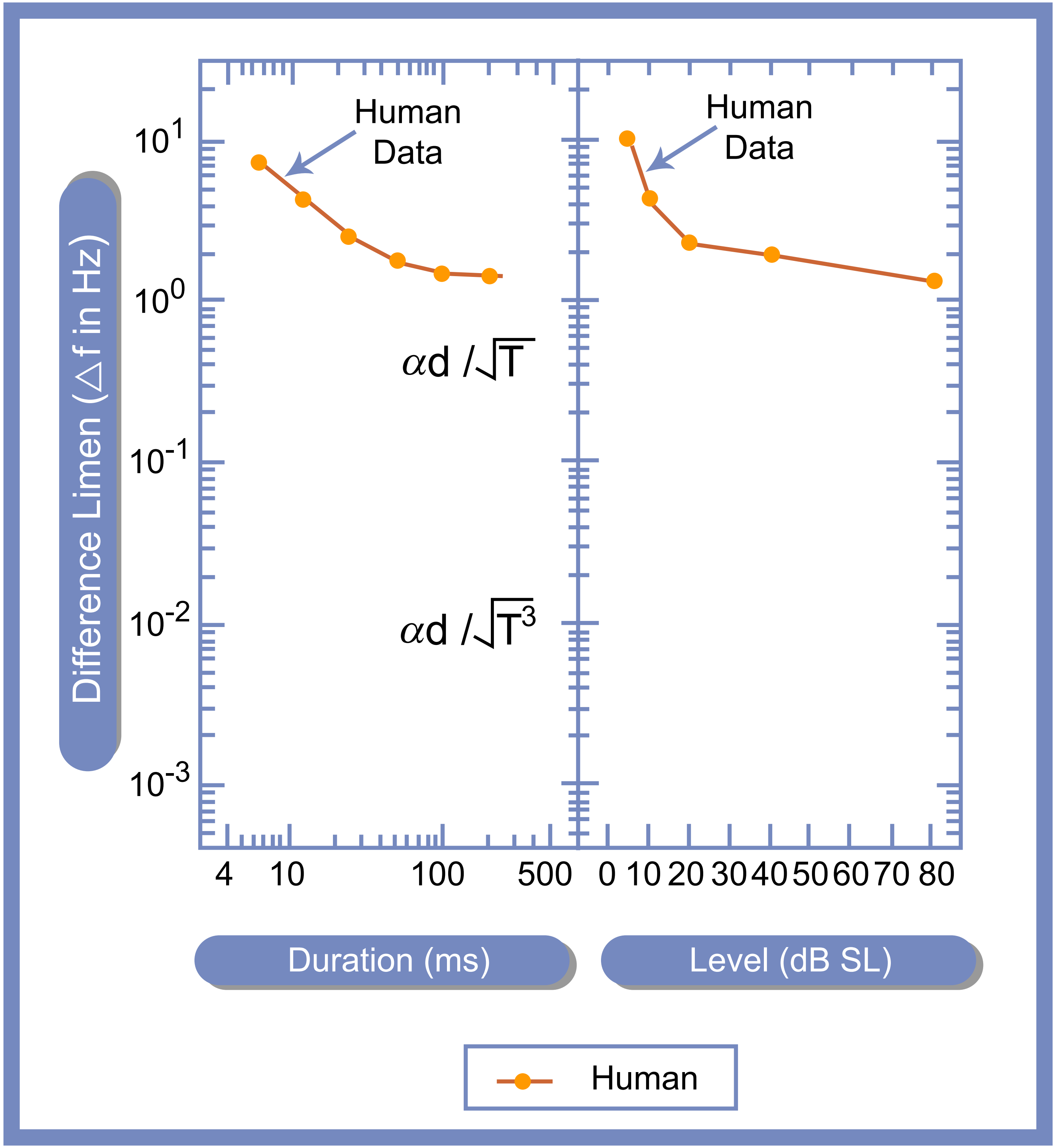
Typical human performance for pure-tone frequency discrimination.

Pure tone pitch discrimination becomes markedly worse above 2 kHz

Weber fractions for frequency (∆f/f) increase 1-2 orders of magnitude between 2 kHz and 10 kHz



Pure tone pitch discrimination improves at longer tone durations and at higher sound pressure levels



Note durations in music

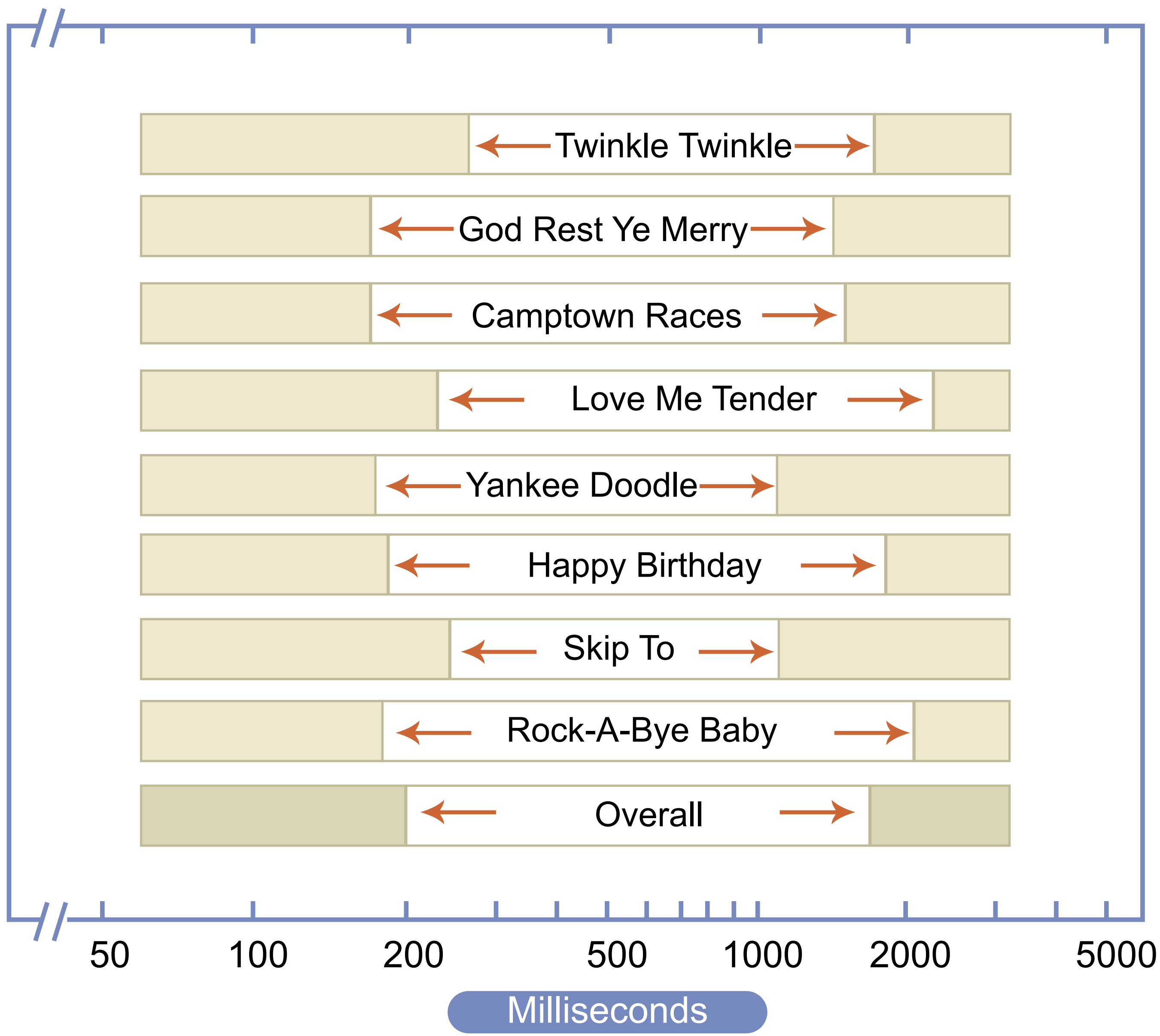


Image adapted from: McAdams, and Bigand. *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford University Press, 1993.

Timbre: a multidimensional tonal quality

tone texture, tone color
distinguishes voices, instruments

(Photo Courtesy of Pam Roth. Used with permission.)

Stationary Aspects

(spectrum)

Dynamic Aspects

Δ spectrum
Δ intensity
Δ pitch
attack
decay

Photo Courtesy of Per-Ake Bystrom. Used with permission.)

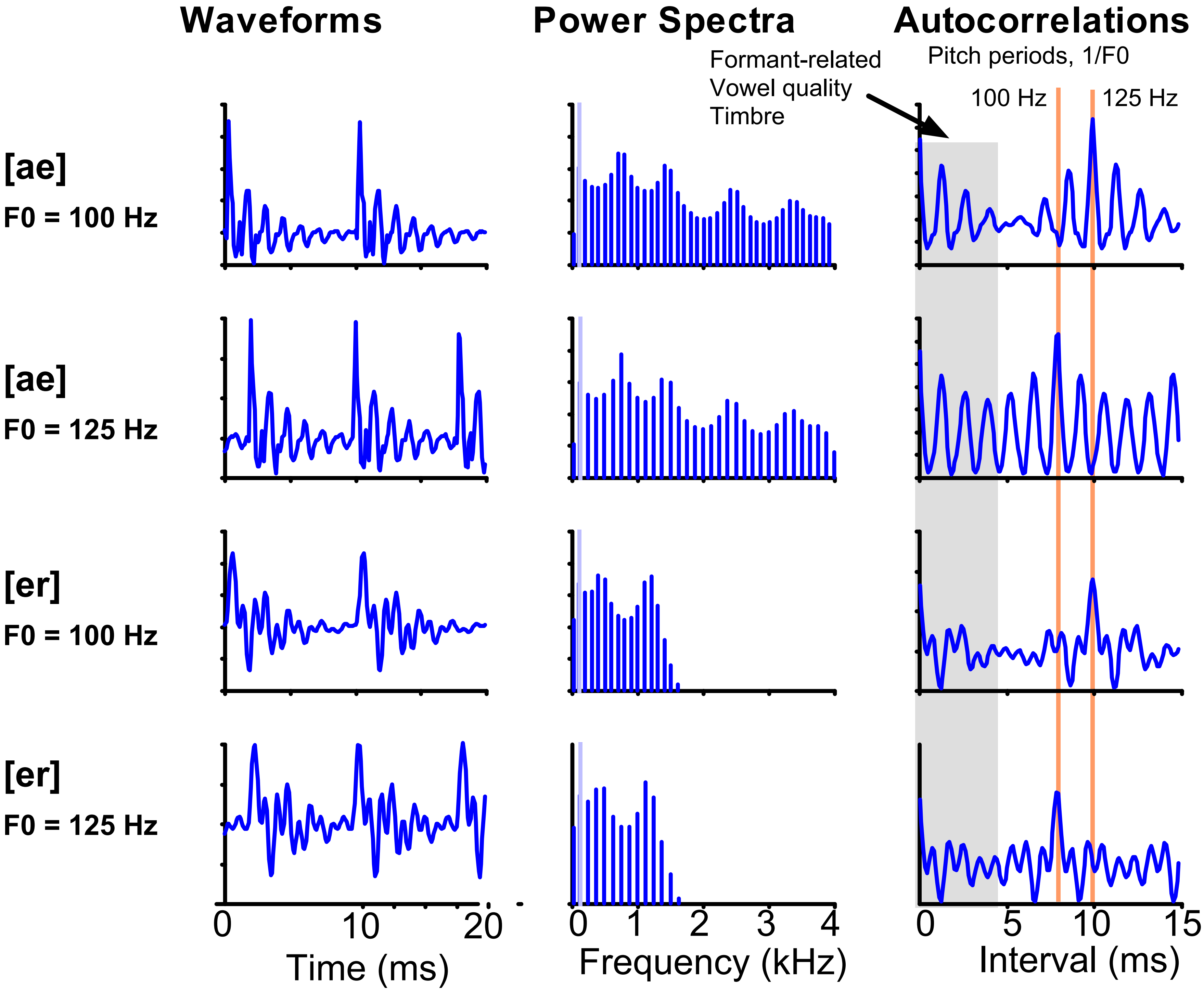
Vowels

Consonants

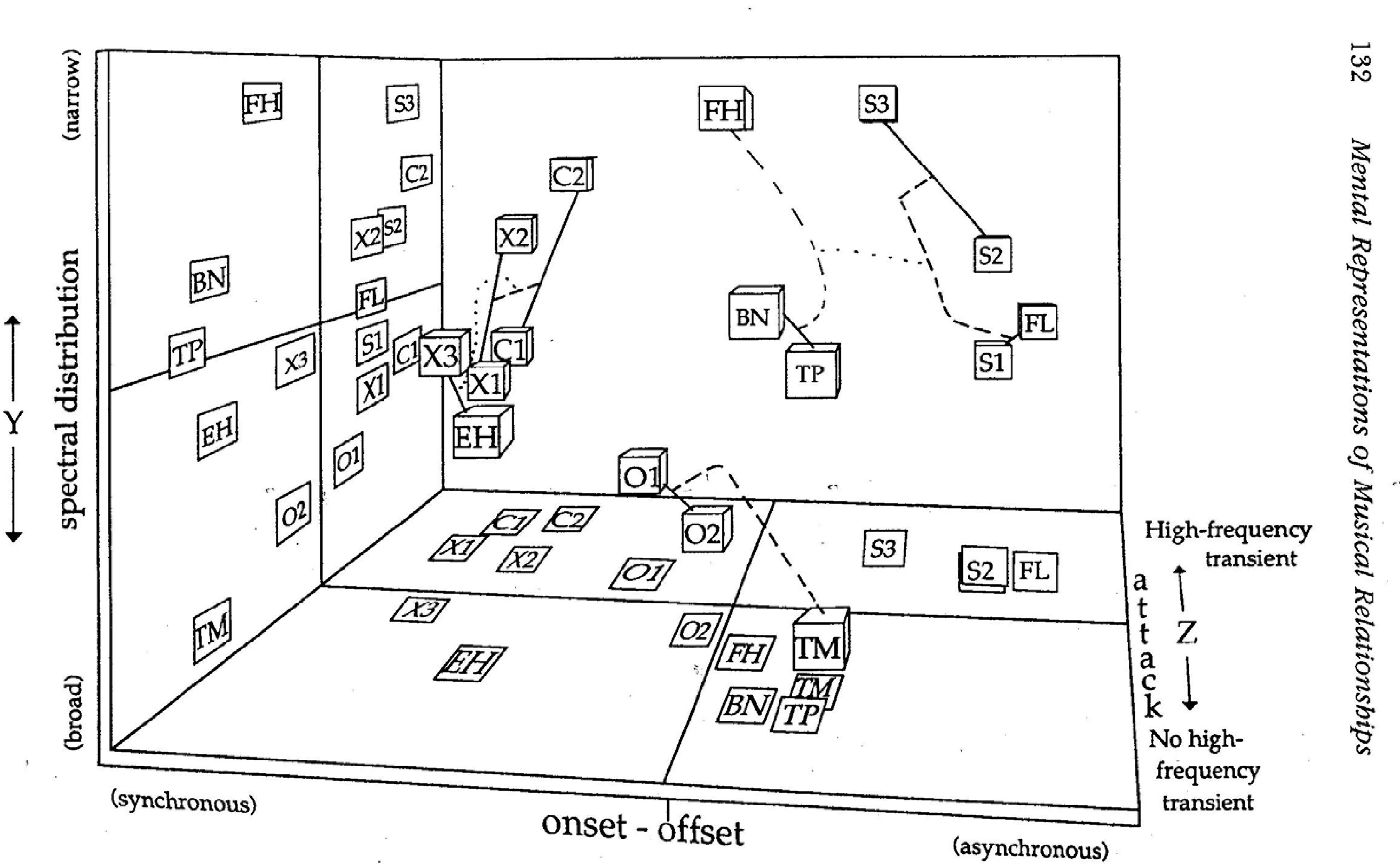
Photo Courtesy of Miriam Lewis. Used with permission.)

http://www.wikipedia.org/

Stationary spectral aspects of timbre



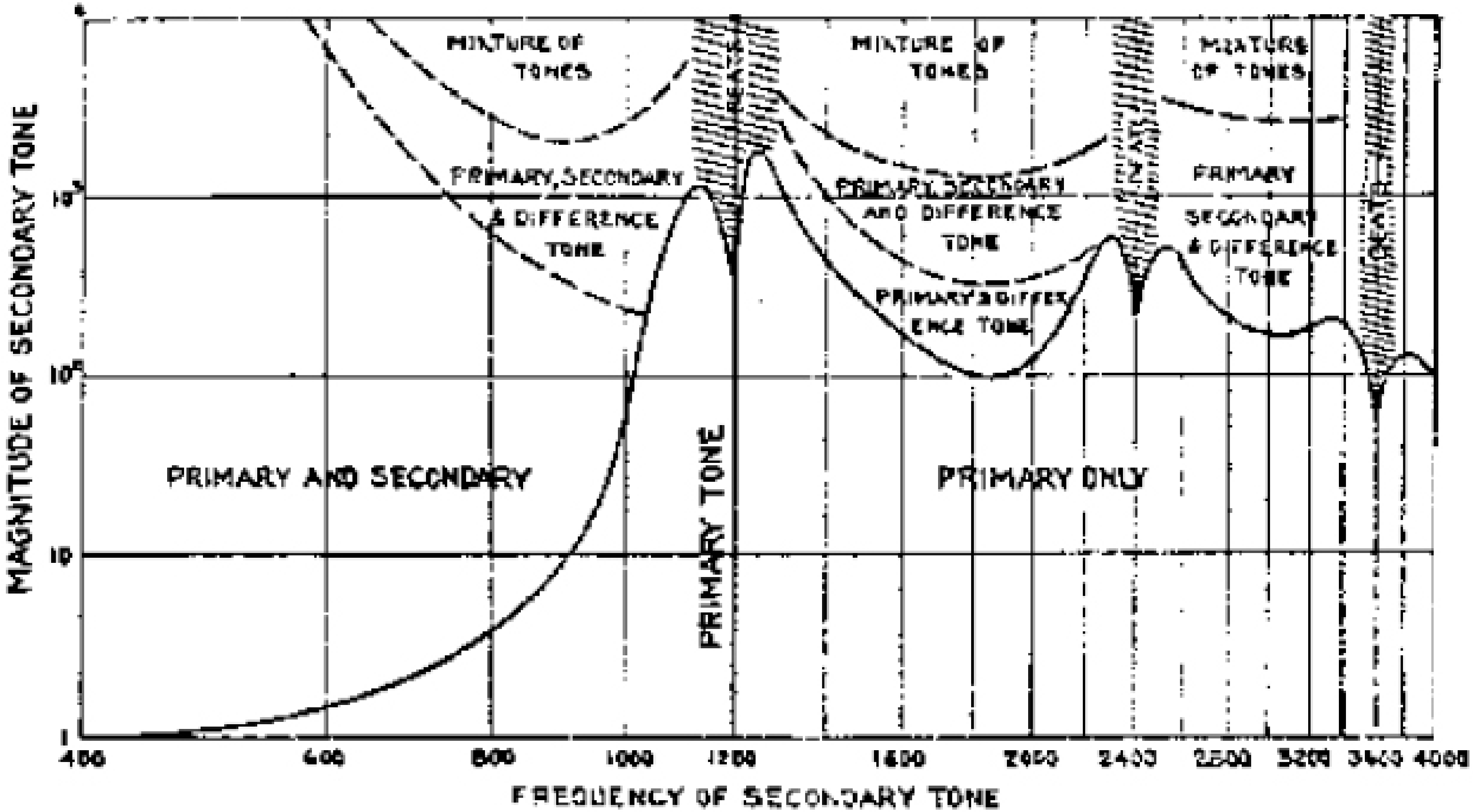
Timbre dimensions: spectrum, attack, decay



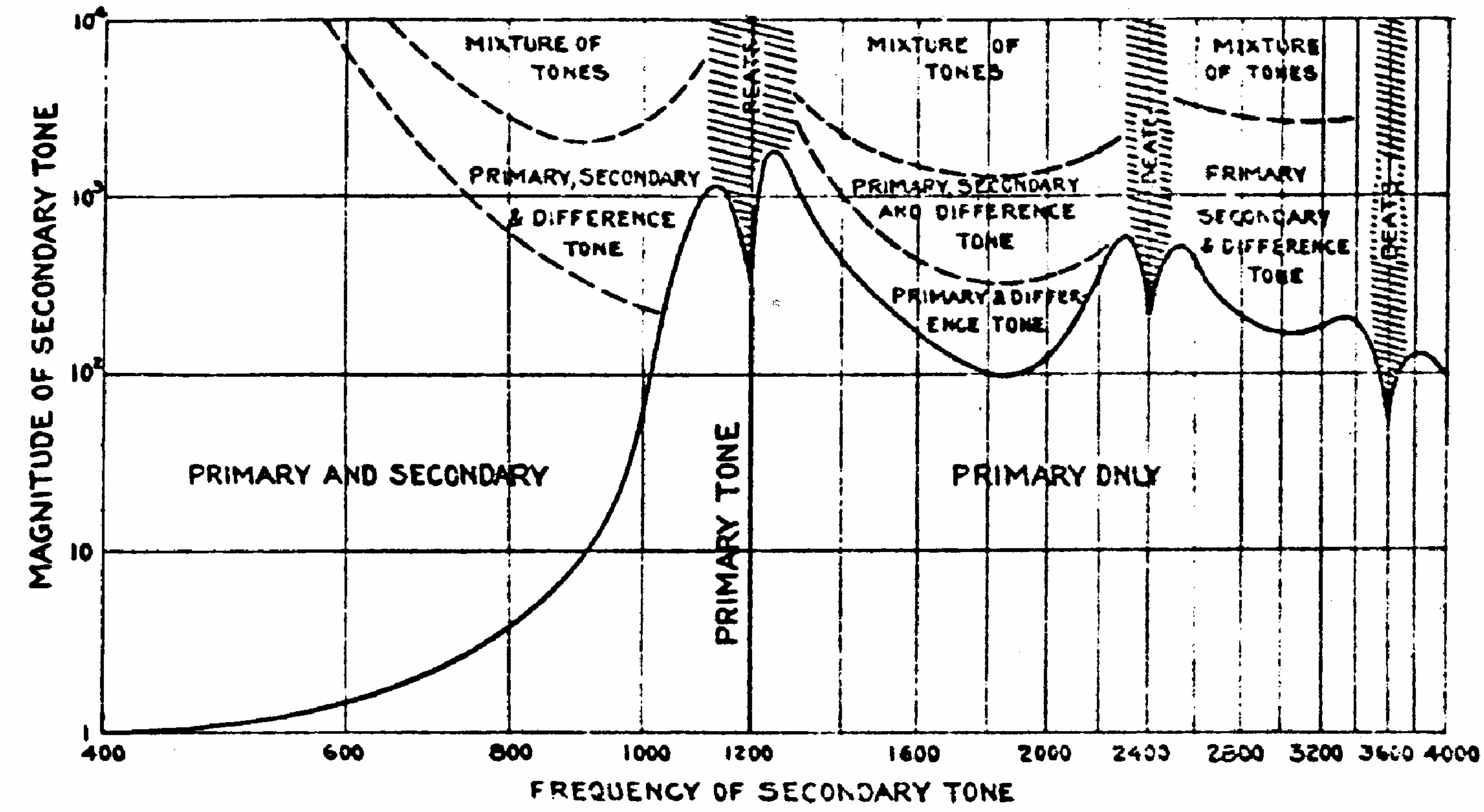
Series of figures from Handel, S. 1989. Listening: an Introduction to the Perception of Auditory Events. MIT Press. Used with permission.

Masking (tone vs. tone)

Demonstration: tones in noise; tones vs. tones



Masking audiograms

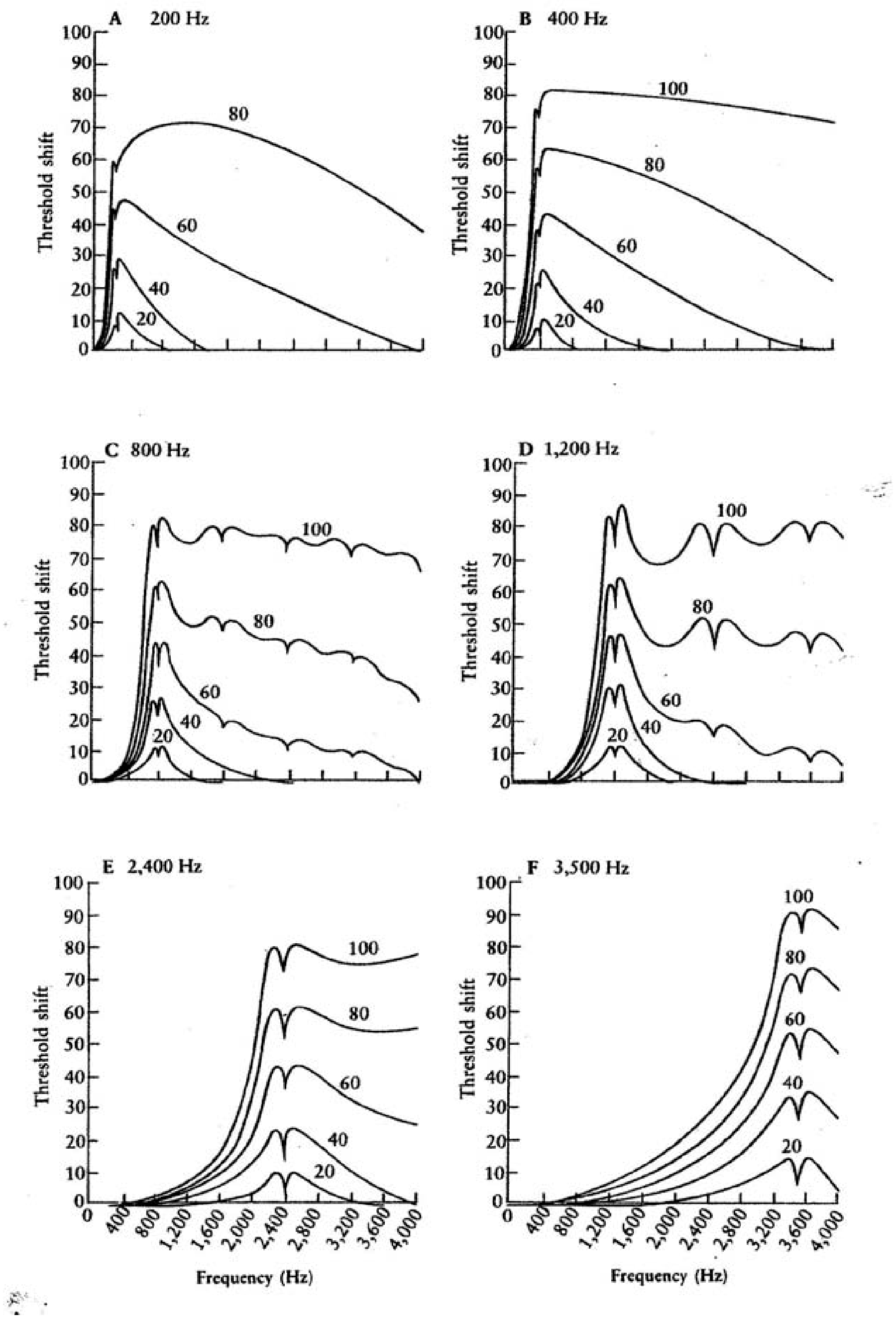


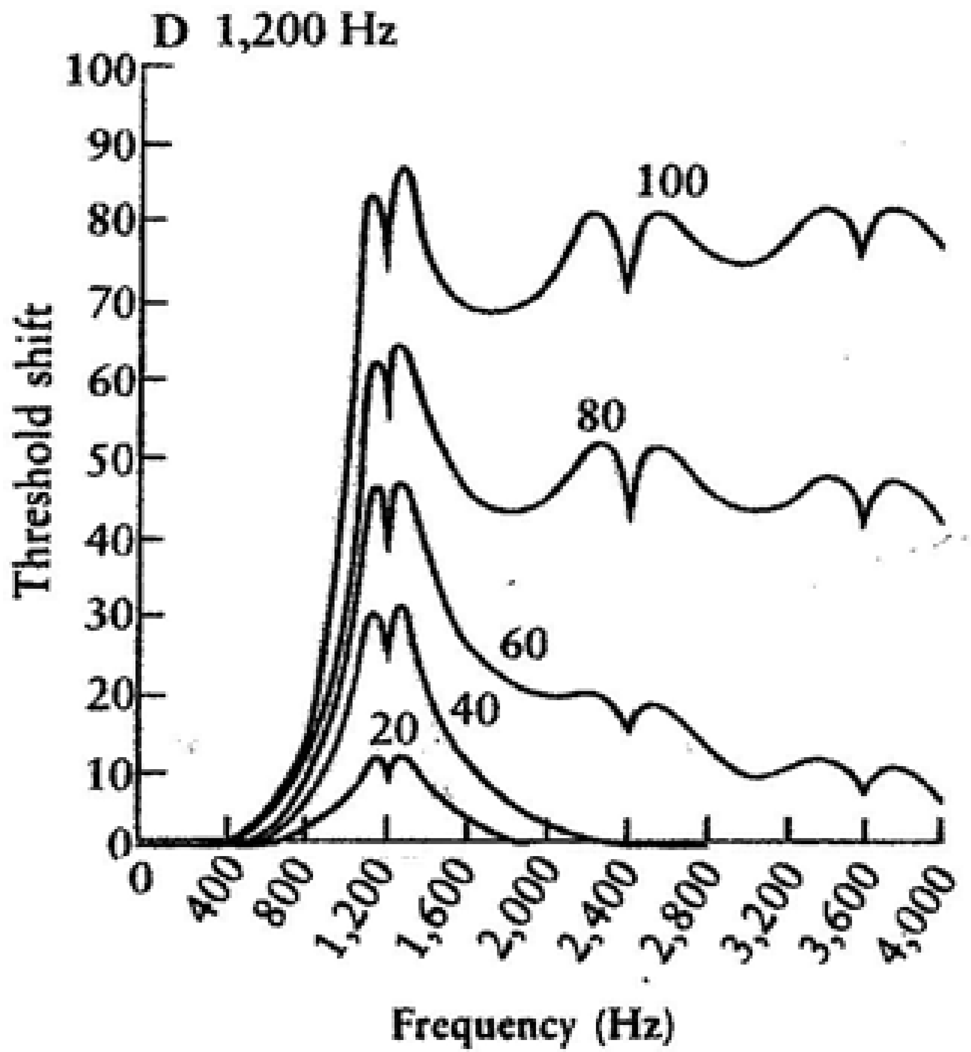
Wegel & Lane, 1924

1000 Hz pure tone masker

Please see http://www. zainea.com/masking2.htm for a discussion of masking.

Tone on tone masking curves (Wegel & Lane, 1924)



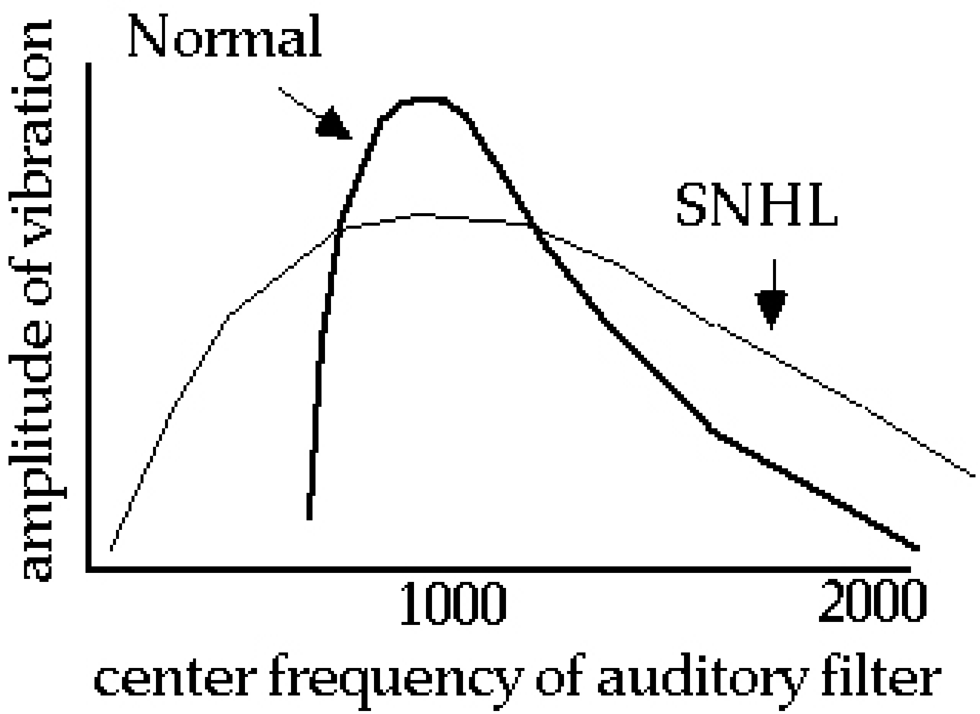


From masking patterns to "auditory filters" as a model of hearing

Power spectrum Filter metaphor

Notion of one central spectrum that subserves perception of pitch, timbre, and loudness

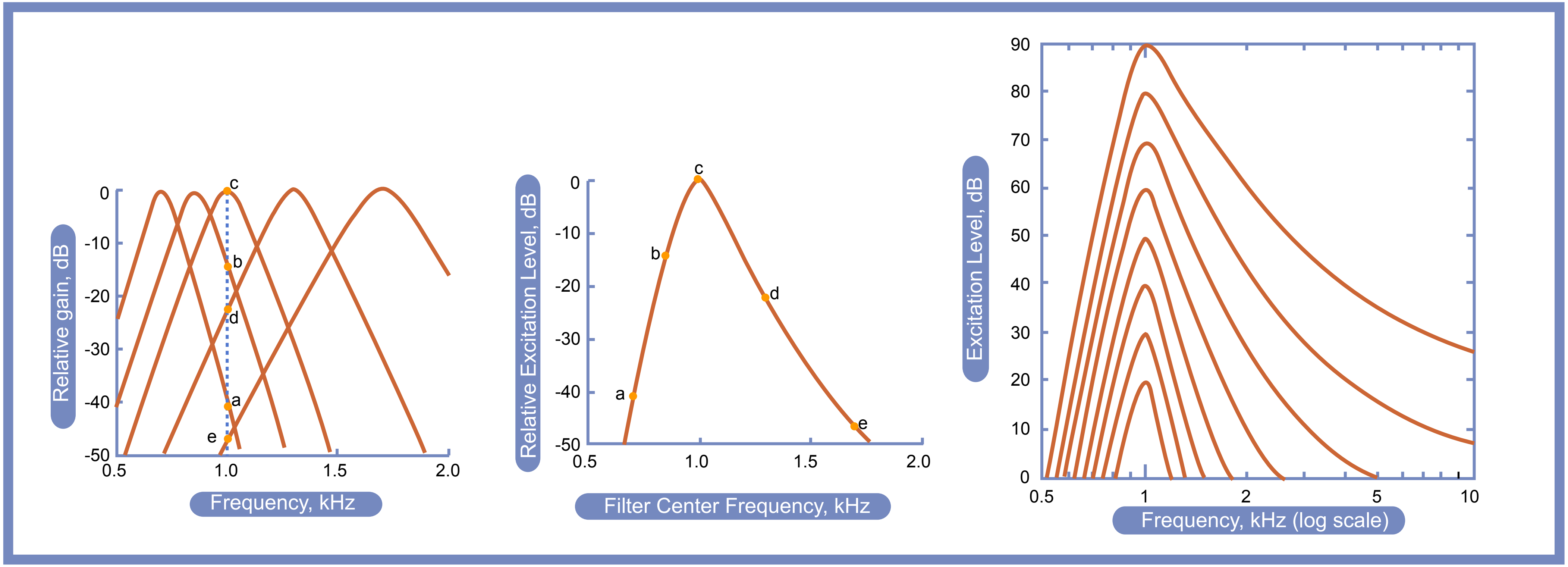
(Courtesy of Prof. Chris Darwin (Dept. of Psychology at the University of Sussex). Used with permission.)



2.2. Excitation pattern Using the filter shapes and bandwidths derived from masking experiments we can produce the excitation pattern produced by a sound. The excitation pattern shows how much energy comes through each filter in a bank of auditory filters. It is analogous to the pattern of vibration on the basilar membrane. For a 1000 Hz pure tone the excitation pattern for a normal and for a SNHL (sensori-neural hearing loss) listener look like this: The excitation pattern to a complex tone is simply the sum of the patterns to the sine waves that make up the complex tone (since the model is a linear one). We can hear out a tone at a particular frequency in a mixture if there is a clear peak in the excitation pattern at that frequency. Since people suffering from SNHL have broader auditory filters their excitation patterns do not have such clear peaks. Sounds mask each other more, and so they have difficulty hearing sounds (such as speech) in noise. --Chris Darwin, U. Sussex,

http://www.biols.susx.ac.uk/home/Chris_Darwin/Perception/Lecture_Notes/Hearing3/hearing3.html

Shapes of perceptually-derived "auditory filters" (Moore)



Binaural localization

Azimuth:

interaural time differences (20-600 usec)

interaural level differences

Elevation:

received spectrum of broadband sounds (pinna effects)

Please see Figure 2.1 in Woodworth, Robert Sessions, 1869-1962. Experimental Psychology. New York: H. Holt and company, c1938.

Interaural time difference and localization of sounds