21M.380 · Music and Technology
Recording Techniques & Audio Production

Perception of sound

Session 4 · Monday, September 19, 2016

1 Student presentation (PA1)

- [redacted]

2 Announcements

2.1 I want *you* for schlepping!

- Volunteers needed for Wed, 9/21 class meeting
- 2–3 volunteers at room [redacted], 10 minutes before start of class
- 2–3 volunteers after class (please approach me after class)

2.2 Preview QZ1

3 Physics vs. perception of sound

3.1 Psychoacoustic relationships

Physical property	Perceptual effect
Amplitude	Loudness
Fundamental frequency	Pitch
Spectral composition	Timbre
Sound source position	Perceived direction

Table 1. Some psychoacoustic relationships

- Physics vs. perception of sound can differ radically!
- Field of *psychoacoustics* investigates relationships between them
- For example, we can hear sounds that are not physically present!
- Important to use correct terminology in each situation! For example:
  - Amplitude & frequency refer to *physical* properties
  - Loudness & pitch refer to *perceptual* phenomena
- Yes, amplitude affects loudness perception.
- Yes, frequency affects pitch perception.
- But relationships are interdependent and highly non-linear!

3.2 Just noticeable difference (jnd)

- Smallest change of a physical quantity that results in a perceptual effect
- Examples:
  - Jnd for amplitude ($\approx$ 1 dB)
  - Jnd for frequency (depends on range)
  - Jnd for source position ($\approx 1^\circ$ for front direction)
- Pigeon (2007–2014) provides various blind jnd listening tests online

4 Auditory scene analysis

- Bregman (1990): Imagine you can tell number, sizes, and positions of boats on a lake – from looking at resulting water ripples on the shore.
- Human ear continuously performs aural equivalent of this task![1]
  - But *how* does human ear make sense of 'spectrogram mess'?
  - Field of *auditory scene analysis* investigates underlying principles
- Application example №1: Music composition
  - J. S. Bach – *E major partita for solo violin* (bwv 1006), prelude, mm.13–28
  - Franz Liszt – *Etude III: La Campanella* `http://auditoryneuroscience.com/topics/la-campanella`
- Application example №2: Music mixing
  - First law of music mixing: Balance = coherence + transparency
  - Coherence requires fusion, transparency requires segregation

[1] *That's* what *I* call music technology!

4.1 Sequential grouping (stream segregation)

- How can ear tell which sequential components belong to same stream?
- Determined by *acoustic distance* (Bregman and Woszczyk 2004, p. 40)
- Refers to separation of sounds in terms of:
  - Frequency: `http://auditoryneuroscience.com/topics/streaming-alternating-tones`
  - Time (cf., Bregman and Ahad 1996, ex. 1)
  - Fundamental frequency (cf., Bregman and Ahad 1996, ex. 6)
  - Spectral shape
  - Spatial direction (cf., Bregman and Ahad 1996, ex. 38)
  - Center frequency (e.g., band-passed noise bursts)
- Weaker factors:

  - Differences in intensity
  - Differences in rise times
  - Differences in noisiness
- Abruptness of changes also affects grouping
- Cumulative effect (stronger grouping as evidence grows over time)

4.2 Simultaneous grouping (spectral integration or fusion)

- How can ear tell which concurrent components belong to same source?
- Determined by (non-linear) combination of multiple factors:
  - Principle of harmonicity, e.g., fusion by common frequency change (cf., Bregman and Ahad 1996, ex. 19)
  - Onset and offset asynchrony (cf., Bregman and Ahad 1996, ex. 21, `http://auditoryneuroscience.com/topics/onsets-and-vowel-identity`)
  - Envelope independence
  - Spatial separation
  - Spectral separation
- More robust cues tend to be more dominant

4.3 Competition sequential vs. simultaneous grouping

- Phenomenon of *apparent continuity*
  - Interpreted in terms of a principle referred to as *old-plus-new heuristic*
  - Cf., Bregman and Ahad (1996, exs. 28, 32, 36, 27)
- Combining information from many cues

5 Anatomy of the human ear

5.1 Outer ear

- Pinna (what we call 'the ear')
- Auditory canal
- *Tympanic membrane* (ear drum) at transition to middle ear

5.2 Middle ear

- Filled with air (unless you have a cold)
- Eustachian tube for pressure equalization to outside world
- 3 ossicles (malleus, incus, stapes)
  - Act as impedance transformer for fluid inside cochlea
  - Analogy: Shouting at your diving friend from shore of lake
- Oval window at transition to inner ear

5.3 Inner ear

- Embedded in hardest bone of human body
- Filled with fluid
- *Cochlea* ('curly thing') contains *basilar membrane*
- Different frequencies excite different sections of basilar membrane
- Haircells (inner vs. outer) on membrane connected to auditory nerve

6 Limits of human hearing

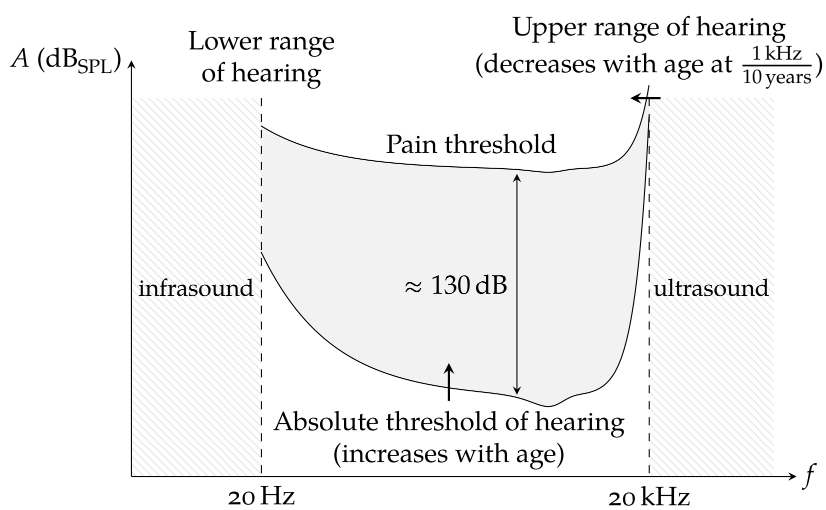

Figure 1. Amplitude and frequency limits of human hearing

6.1 Frequency

- Rule of 👍: 20 Hz to 20 kHz (ca. 10 octaves)
- Upper range of hearing decreases by ca. 1 kHz per life decade
- Lower range of hearing: rhythm-pitch continuum

6.2 Amplitude

- Lower limit: *Absolute threshold of human hearing*
  - In absolute terms: $I_0 \approx 1 \times 10^{-12}\,\mathrm{W\,m^{-2}}$ at 1 kHz
  - Depends heavily on frequency (more later)
  - Increases with age (but not as dramatic as for frequency)
- Upper limit: *Pain threshold* ($I_{pain} \approx 10\,\mathrm{W\,m^{-2}}$ at 1 kHz)
- Exercise: What is the dynamic range of the human ear?

7 Loudness perception

- Perceived loudness is a subjective quality
- Depends on many factors other than physical sound pressure level

7.1 Factors that loudness perception depends on[2]

[2] cf., Farnell 2010, pp. 81 ff.

- Perceived loudness of short sounds increases with their duration
  Sound design application: Stretching gunshots (< 200 ms)
- Perceived loudness for steady sounds decreases with exposure time
  - Depends on absolute sound pressure level
  - Short interruptions 'reset' loudness perception
- Loudness distinction depends on (Farnell 2010, p. 83):
  - Frequency (minimum JND for 1 kHz to 4 kHz)
  - Sound pressure level (minimum JND for 60 $\mathrm{dB_{SPL}}$ to 70 $\mathrm{dB_{SPL}}$)
- Most significantly, perceived loudness depends on *frequency*

7.2 Equal-loudness contours

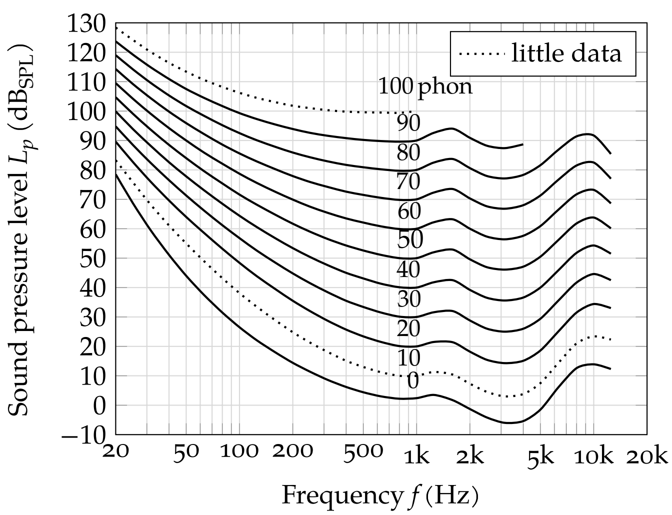

Figure 2. Equal-loudness contours (Iso 2003)

- Describe frequency dependence of loudness perception
- First measured by Fletcher and Munson (1933)
- Multiple revisions since; most recently by Iso (2003)
- Spl at 1 kHz defines *phon* as a measure of equal loudness
  - Zero-phon curve … absolute threshold of human hearing

- 60 phon means [no more than] "as loud as a 1 kHz sine tone at 60 dB"
- But really only expresses *equal* loudness & only applies to sine tones

- Exercise: By how many dB does one need to increase sound pressure level of a 20 Hz tone by comparison to an 1 kHz, 80 phon tone, such that both will be perceived as equally loud?

7.3 Decibel weightings

- Idea:
  - Apply inverted equal-loudness contour to dB_{SPL}
  - Hope is to yield perceptually relevant loudness measure
- *A-weighted decibel* (dB_A):
  - RMS-averaging measure
  - Uses inversion of original 40 phon curve (Fletcher and Munson 1933)
  - Widely used for much louder sounds (e.g., industrial noise) ☹
- *C-weighted decibel* (dB_C):
  - Models flatter equal loudness contours for larger SPLs
  - So more suitable for measurements of sounds > 100 dB
- Old dB_B and dB_D weightings no longer widely used
- Example: Galaxy CM-140 SPL meter in MOSS features A & C weightings
- dB_{ITU} works better for music and speech, since it accounts for transients
- However, consider more recent *loudness unit* LU for music production

7.4 Masking

- Sounds can *mask* (i.e., render inaudible) other sounds
- Rules of 👍: Ability of a sound to mask other sounds increases with
  - Sound pressure level (louder sounds are more likely to mask)
  - Bandwidth (white noise is a better masker than a sine tone)
- Pratical application: Psychoacoustic data compression
  - If we don't hear it, why bother storing or transmitting it?
  - Backbone of lossy audio file formats such as MP3
- Interpretation: Presence of masker causes *temporary threshold shift* for
  1. simultaneous sounds in frequency neighborhood (spectral masking)
  2. sounds that occur just before or after masker (temporal masking)

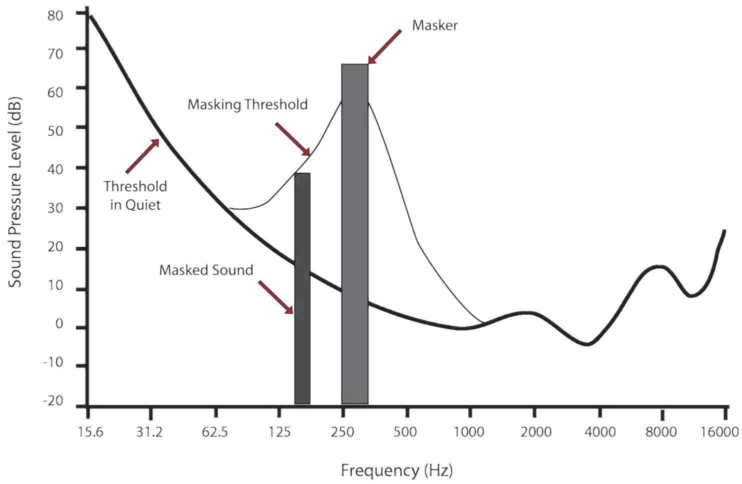

Figure 3. Spectral masking (Public domain image. Source: `https://en.wikipedia.org/wiki/File:Audio_Mask_Graph.png`)

7.4.1 Spectral masking (simultaneous)

- Demo:
  1. Play white noise & sine tone together
  2. Decrease sine tone's SPL until nobody in class can hear it
  3. Stop white noise
- Interpretation:
  - Masked sound above absolute threshold of hearing in quiet
  - Is therefore audible if masker is absent
  - But masker temporarily raises threshold for nearby frequencies
  - Masked sound falls below raised threshold and becomes inaudible

7.4.2 Temporal masking (post- & pre-masking)

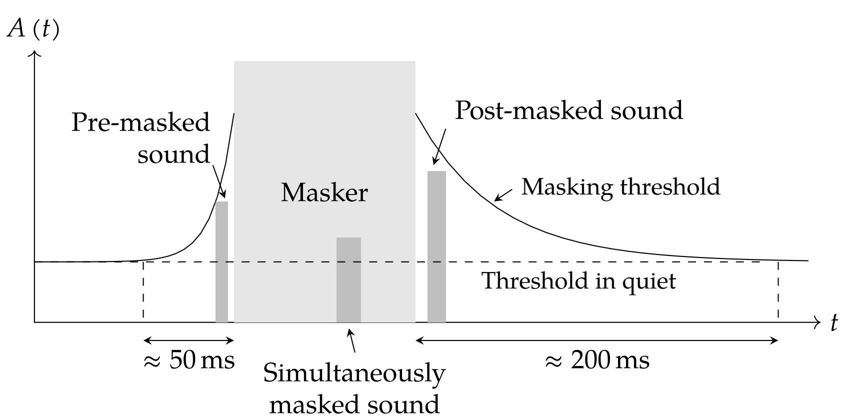

Figure 4. Temporal masking

Again temporary threshold shift, but this time in time domain:

- *Post-masking:* raised threshold for ca. 200 ms after masker ceases
  - Somewhat intuitive
  - Same effect as on day after a loud concert (at shorter timescale)
- *Pre-masking:* raised threshold for ca. 50 ms *before* masker appears (!)[3]
  - How to explain? Auditory perception is *not* instantaneous!
  - Instead, ear 'integrates' perceptual stimuli over short time windows
  - Example: Ear cannot follow amplitude oscillations of a 440 Hz tone

[3] Note that post-masking is occasionally referred to as *forward masking*, while pre-masking is also called *backward masking*. It's all a matter of perspective.

8 Pitch perception

8.1 Combination tones[4]

- Occur when 2 frequencies f_1 and f_2 played simultaneously
- We sometimes hear a third tone that is *not physically present*!
  - In particular *difference tones* (most prominent & reliable)
  - Much less reliable (and somewhat debated): *sum tones*
- Explanation: Non-linear distortions in inner ear
- Sound example:
  - Actually playing: Stationary f_1 & downward glissando f_2
  - Also audible: Rising difference tone $2 \cdot f_1 - f_2$

[4] Combination tones are sometimes also referred to as *Tartini tones*, named after violinist Giuseppe Tartini, one of the people credited with their discovery (albeit not the first).

Table 2. Combination tones

Difference	vs.	sum tones
$f_1 - f_2$		$f_1 + f_2$
$2 \cdot f_1 - f_2$		$2 \cdot f_1 + f_2$
$3 \cdot f_1 - f_2$		$3 \cdot f_1 + f_2$
…		…

8.2 Missing fundamental

- Remember: Harmonic spectrum's pitch determined by fundamental f_1
- Remarkable: Applies even if f_1 itself is absent from spectrum!
- Intuitive, since a given harmonic spectrum can only match a single f_1
- Sound example: Same melody played with
  1. Harmonics f_4 to f_{10}
  2. Fundamental f_1 only
  3. Harmonics f_1 to f_{10}
- Real-world examples & applications:
  - Some woodwind instruments (e.g., oboe)
  - *MaxxBass*™ plugin (© Waves Inc.)
  - Extending the perceived range of subwoofers or organ pipes

9 Sound localization

- Refers to auditory system's evaluation of where a sound 'comes from'
- Again potentially significant differences between:
  - Physics (sound source location)
  - Perception (perceived location of sound)
- *Localization blur* describes (in)accuracy of sound localization

9.1 Localization blur depends on source signal

Rules of 👍 (cf., Farnell 2010, p. 79):

- Broadband sounds are easier to localize than narrow-band sounds.
- High frequencies are easier to localize than low frequencies.
- Sounds with sharp attacks are easier to localize than stationary sounds.
- Free-field conditions (no reflections from walls) facilitate localization.
- Ability of listener to move head increases localization accuracy

9.2 Localization blur depends on source direction

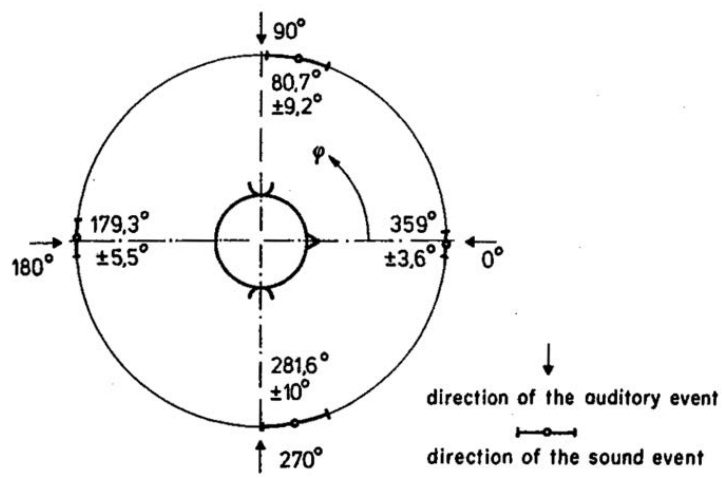

Figure 5. Localization blur in the horizontal plane. Experimental setup: 100 ms white noise pulses, head immobilized (Blauert 1996, p. 41.

- Best localization in horizontal plane
  - Minimum JND ($\approx 1°$) for sounds in front of listener
  - Not quite as good for sounds from behind
  - Localization blur increases further towards sides
- Less reliable localization in median plane (which divides left-right)

- Localization of elevated sounds depends on frequency (!)
- Bad localization also for sounds from below (rarely occurs in nature)

- Significant localization blur also with regards to distance

9.3 Interaural time differences (ITD)

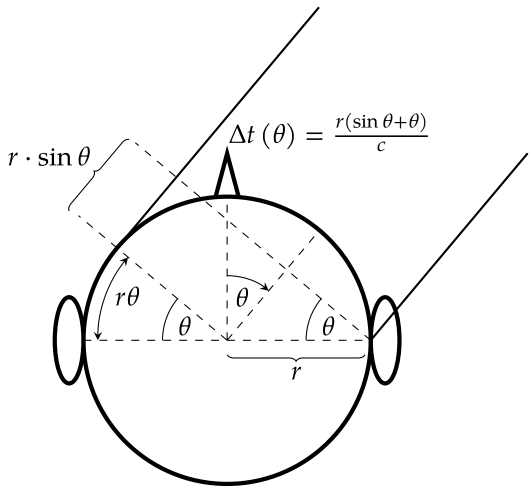

Figure 6. Simple model of interaural time differences

- Sound arrives earlier at the ear closer to the source (since $c < \infty$)
- Results in *interaural*[5] *time difference* (ITD)
- ITD can be modeled geometrically (cf., figure 6)
- Distance between ears: $d \approx 17 \ldots 21\,\text{cm}$
- Exercise: Largest possible ITD?
- But note that we can detect ITD as small as 30 µs!

[5] The term *inter-aural* is rooted in Latin and means "between the ears".

9.4 Interaural level differences (ILD)

- Higher SPL at ear closer to the source
- Two different causes for such *interaural level differences* (ILD):

1. Inverse distance law: $\Delta L = 20 \cdot \log_{10}\left(\frac{p_L}{p_R}\right) = 20 \cdot \log_{10}\left(\frac{r_R}{r_L}\right)$
2. Acoustic shadow of head (affects HF more than diffracted LF)

9.5 ITD & ILD over the frequency range

- ITD & ILD complement each other over frequency range:
  - ITD dominant cue below 700 Hz
  - ILD dominant cue above 1500 Hz

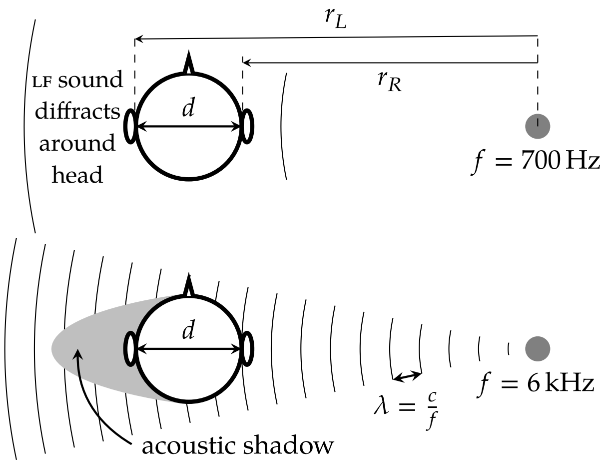

Figure 7. Interaural level differences

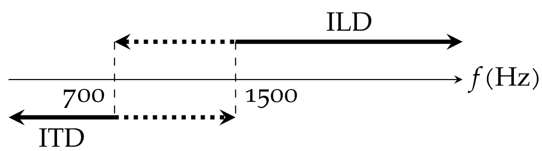

Figure 8. Interaural time and level differences complement each other over the audible frequency range.

- How can we explain this?
  - Ild: ______________________________
  - Itd: ______________________________

9.6 Cone of confusion

- Example: Two sounds that yield identical ITD & ILD:
  - Sound from 45° (front right)
  - Sound from 135° (rear right)
- Generalized to 3D: Sounds on surface of *cone of confusion* around ear yield identical ITD & ILD
- Consequence: ITD & ILD insufficient to explain all aspects of localization!

9.7 Head rotations resolve front-back ambiguities

- Frequent (unconscious) *head rotations* resolve front-back ambiguities
- For example, clockwise head rotation will
  - Decrease interaural differences for sound from 45° (front right)
  - Increase interaural differences for sound from 135° (rear right)
- Localization deteriorates for listening test subjects with fixed head
- But question remains: How does ear determine elevation & distance?

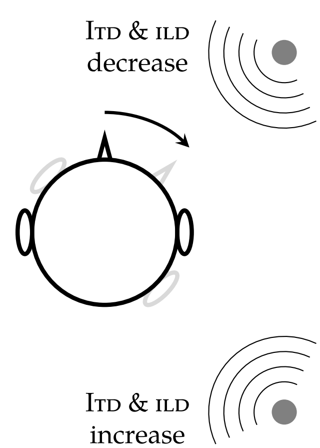

Figure 9. Head rotations resolve front-back ambiguities in sound localization (cf., Blauert 1996, p. 180)

9.8 Elevation cues

- Observation: Pinna asymmetric along front-back & top-bottom axes
  - Reflections from pinna result in incoming sound being *filtered*[6]
  - Reflection pattern & thus filter characteristics depend on direction!
  - Ear decodes this information for front-back discrimination & determining elevation
- Similar cue due to reflections from shoulders (top-bottom asymmetry)
- But localization in median plane not as good as in horizontal plane!
- Depends on frequency (!) more than actual source direction[7]

[6] *Filtering* occurs whenever some frequencies are emphasized, while others are attenuated. We will see in future lectures that reflections at short time intervals Δt inevitably result in such filtering effects. This insight provides the basis for a lot of room acoustics theory, but also for the creation of sound effects such as flangers, phasors, etc., which the guitar players among you might be familiar with.

[7] Cf., Blauert 1996, p. 45.

9.9 Distance cues

- Distance is even harder to judge than elevation
- Especially in absolute terms (arguably true also for visual perception)
- Some cues that indicate increasing source distance to ear:
  - Sound level drop due to inverse distance law $p \propto \frac{1}{r}$
  - High-frequency attenuation due to atmospheric absorption
  - Increasing ratio of reverberant to direct sound (in rooms)

9.10 Precedence effect

- Phenomenon:
  - Same signal arrives from different directions at different times
  - Time delays Δt between signals on the order of 1 ms to 50 ms
  - Sound tends to be localized from direction of first arrival
- Referred to as *precedence effect* or *law of the first wavefront*
- *Haas effect* … special case of precedence effect. Haas showed that:
  - Effect works even if delayed sound has higher SPL than first wave
  - However, the louder the delayed sound, the smaller Δt can be before it becomes distinct echo (Blauert 1996, p. 226)
  - So tradeoff between ΔL and Δt
- Practical application: *Delay lines* in sound reinforcement systems
  - Delayed loudspeakers on the sides of large auditoria with a front PA
  - Idea: Bring up overall volume without compromising localization
  - Used in theaters to increase speech intelligibility

References & further reading

Blauert, Jens (1996). *Spatial Hearing. The Psychophysics of Human Sound Localization*. Revised edition. Cambridge, ma and London: Mit Press. 508 pp. isbn: 978-0-262-02413-6. mit library: 000808775.

Bregman, Albert S. (1990). *Auditory Scene Analysis. The Perceptual Organization of Sound*. Cambridge, ma and London: Mit Press. 790 pp. isbn: 978-0-262-02297-2. mit library: 000435159.

Bregman, Albert S. and Pierre Ahad (1996). *Demonstrations of Auditory Scene Analysis. The Perceptual Organization of Sound. Audio compact disk*. Distributed by Mit Press. Montréal, Canada: Auditory Perception Laboratory, Psychology Department, McGill University. mit library: 002304502. Also available from http://webpages.mcgill.ca/staff/Group2/abregm1/web/downloadsdl.htm.

Bregman, Albert S. and Wieslaw Woszczyk (2004). "Controlling the perceptual organization of sound." In: *Audio Anecdotes: Tools, Tips, and Techniques for Digital Audio*. Ed. by Ken Greenebaum and Ronen Barzel. Vol. I. Natick, ma: A K Peters, pp. 35–63. mit library: 001253727. url: http://webpages.mcgill.ca/staff/Group2/abregm1/web/pdf/2004_Bregman_Woszczyk.pdf.

Farnell, Andy (2010). *Designing Sound*. Cambridge, ma and London: Mit Press. 688 pp. isbn: 978-0-262-01441-0. mit library: 001782567. Hardcopy on course reserve at the Lewis Music Library. Also available through mit libraries as an electronic resource.

Fastl, Hugo and Eberhard Zwicker (2007). *Psychoacoustics. Facts and Models*. 3rd ed. Springer Series in Information Science. Springer. 462 pp. isbn: 978-3-642-51765-5. (Visited on 11/17/2107).

Fletcher, Harvey and Wilden A. Munson (1933). "Loudness, its definition, measurement and calculation." In: *Journal of the Acoustic Society of America* 5, pp. 82–108. doi: 10.1121/1.1915637.

International Organization for Standardization (Aug. 2003). *Acoustics – Normal equal-loudness-level contours*. Iso 226:2003. mit library: 001410672. url: https://www.iso.org/standard/34222.html (visited on 03/22/2017).

Loy, Gareth (2007). "Psychophysical basis of sound." In: *Musimathics. The Mathematical Foundations of Music*. Vol. 1. Cambridge, ma and London: Mit Press. Chap. 6, pp. 149–98. mit library: 001379675. Available at: Mit Learning Modules ‣ Materials.

Moore, Brian C. J. (2014). "Psychoacoustics." In: *Handbook of Acoustics*. Ed. by Thomas D. Rossing. 2nd ed. Modern Acoustics and Signal Processing. New York, ny: Springer. Chap. 13, pp. 475–517. isbn: 978-1493907540.

Pigeon, Stéphane (2007–2014). *Blind Listening Tests*. url: http://www.audiocheck.net/blindtests_index.php (visited on 10/03/2014).

Zwicker, Eberhard (1961). "Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen)." In: *Journal of the Acoustical Society of America* 33.2. doi: https://asa.scitation.org/doi/abs/10.1121/1.1908630.