Tassu Takala 2019

Emergent User Interfaces
CS-E4200

Introduction to Multimodal Interaction 1a

-

Sound & Haptics

these slides adapted from 2018 course materials
https://mycourses.aalto.fi/course/view.php?id=16936&section=1 (David McGookin)
https://mycourses.aalto.fi/course/view.php?id=16924&section=1 (Mark Billinghurst)

Project Assignment

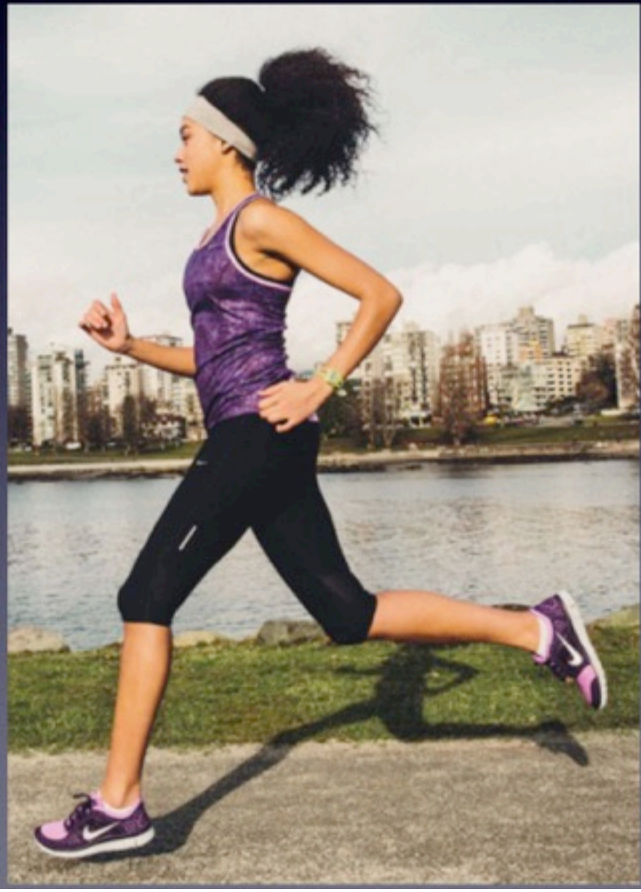
- Project Assignment sheet has been released on MyCourses
- Covers all the details
- First Step Think! : Where could a tangible approach work?
  - wellbeing, navigation, information search, communication, fun …
- Fill out the Google Form with as many ideas as you have!

Multimodality

Figure I.1
How the computer sees us.

Multimodality

- We don't fit fit that model any more than we fit the WIMP paradigm!




Real life in the wild…

[redacted]


THERE'S NOT
AN APP FOR THAT
MOBILE USER EXPERIENCE DESIGN FOR LIFE
MK
SIMON ROBINSON
GARY MARSDEN
MATT JONES

Senses

- How an organism obtains information for perception:
  - Sensation part of **Somatic Division** of **Peripheral Nervous System**
  - Integration and perception requires the **Central Nervous System**
- Five major senses:
  - Sight (Opthalamoception)
  - Hearing (Audioception)
  - Taste (Gustaoception)
  - Smell (Olfacaoception)
  - Touch (Tactioception)

Other Lesser Known Senses..

- Proprioception = sense of body position
  - what is your body doing right now
- Equilibrium = balance
- Acceleration
- Nociception = sense of pain
- Temperature
- Satiety (the quality or state of being fed or gratified to or beyond capacity)
- Thirst
- Micturition
- Amount of CO_2 and Na in blood

Multimodality

- Multiple different definitions
- In this course we mean the use of one or more senses to communicate information

= multisensory UI

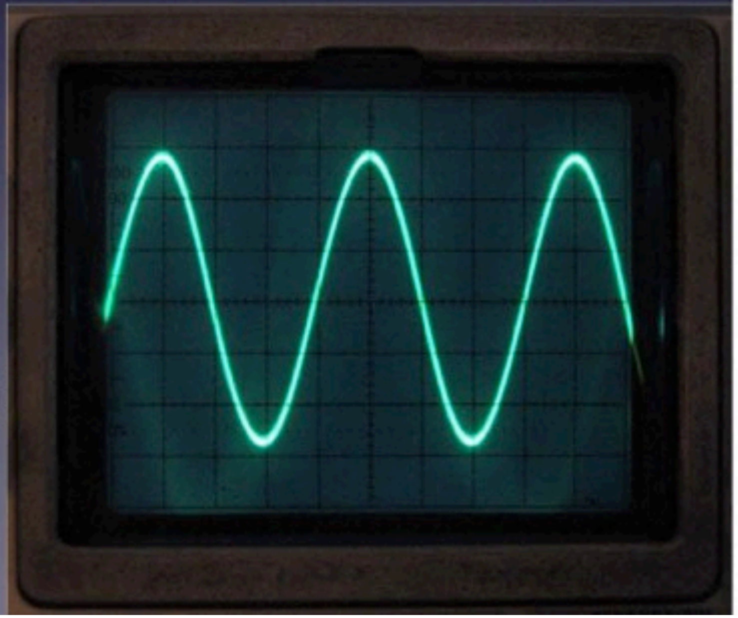
  - Sound
  - Touch
  - Taste
  - Smell
  - *Vision*
- And why this is a useful thing to do!

Sound

- What is Sound?
- How do we hear it?
- How can we create sounds?
- How can we use it in a User Interface?

What is Sound

- A vibration (repeating waveform) that propagates through an elastic media and is detected by a receiver*
- Two key properties
  - Frequency
  - Pressure Level (Amplitude)



*) ears, microphone...

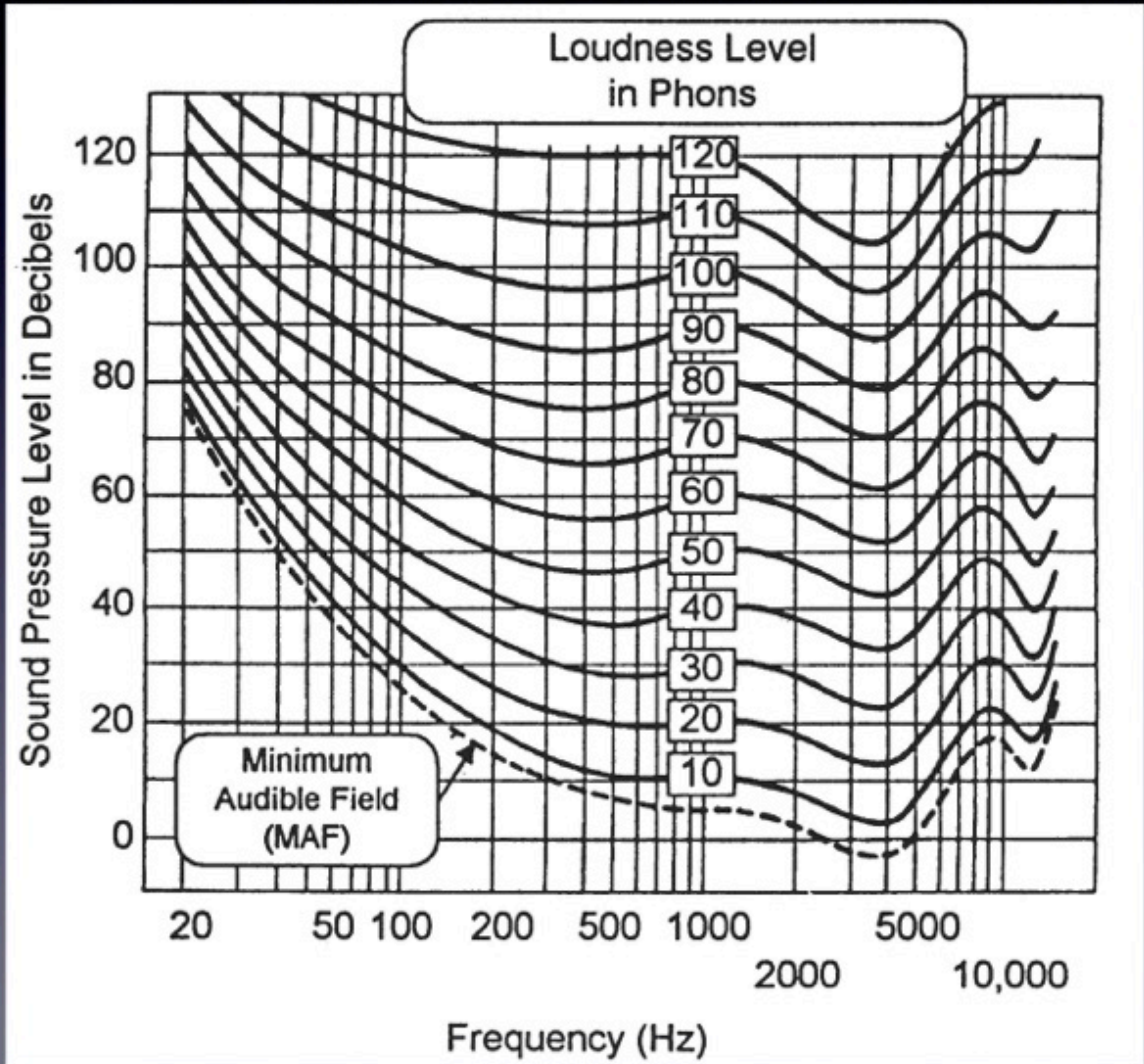
What is Sound?

- Frequency
    - Measured in Hertz (HZ)
        - 1HZ = 1 cycle per second
- Humans can hear approx. 20Hz - 20kHz
    - reduces with age to about 15kHZ
- Pressure Level
    - Variations in sound increase and decrease air pressure.
    - This change is measured in relation to “standard” sound pressure level on a logarithmic scale (decibels dB)
        - normal conversation 65dB Chainsaw 120dB
        - Above 85 is to be avoided
    - Note this is not loudness

Loudness

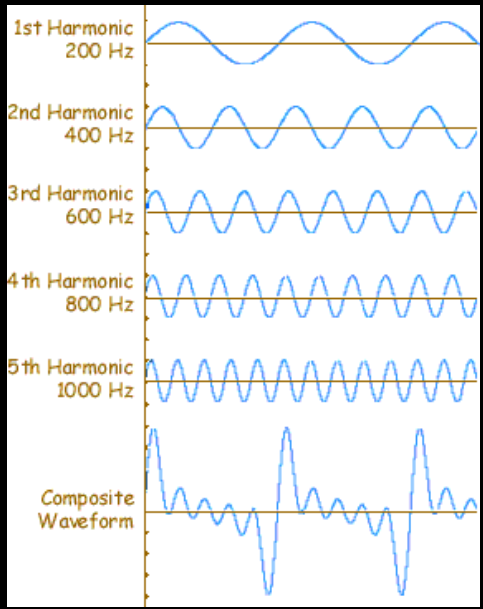
Pressure Level and Frequency both contribute to loudness

Measured in Phons

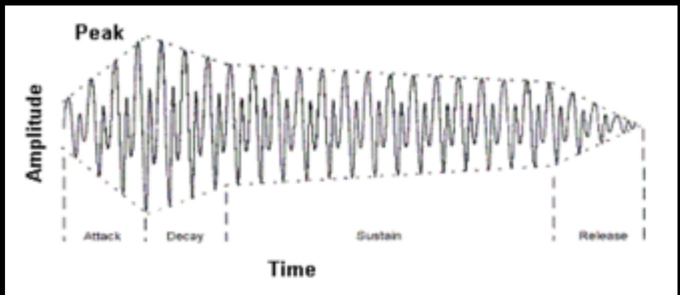


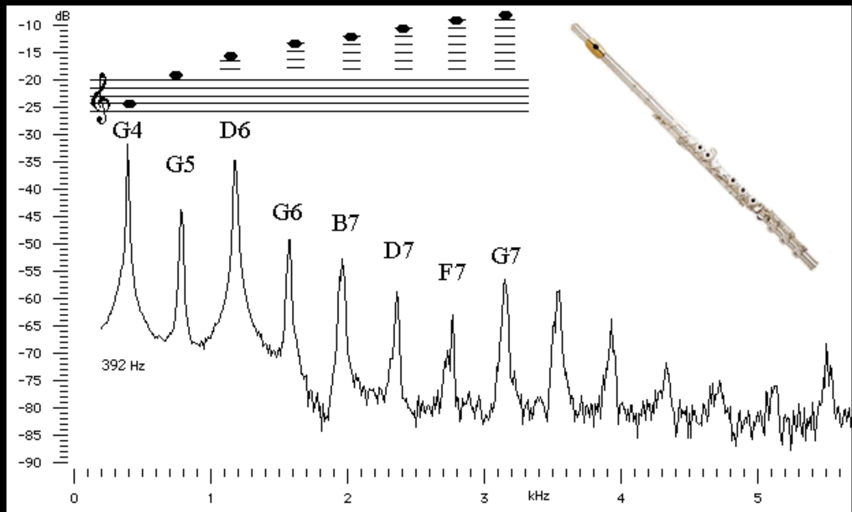
Taken from: S. A. Gelfand. Hearing: An Introduction to Psychological and Physiological Acoustics. Marcel Dekker Inc, New York, 3rd edition, 1998.

Timbre



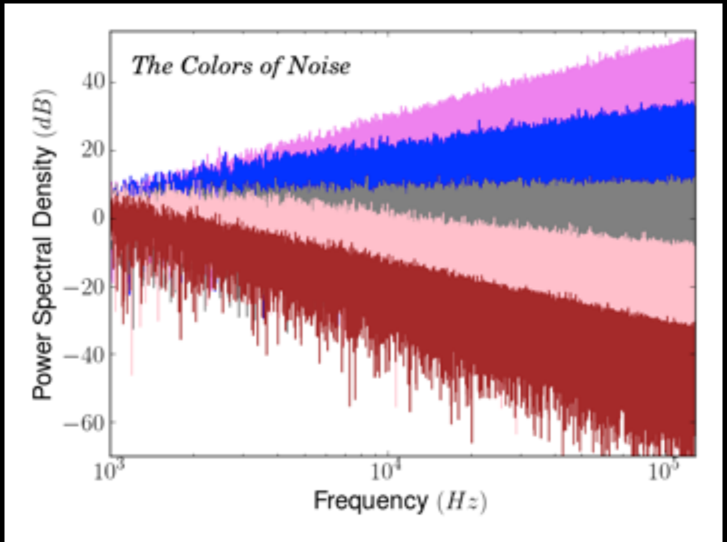
- Not only a single frequency
  - non-sinusoidal waveform
  - tonal sound: base frequency (F_0) + harmonic overtones ($n \times F_0$)
- Not constant loudness
  - sound envelope





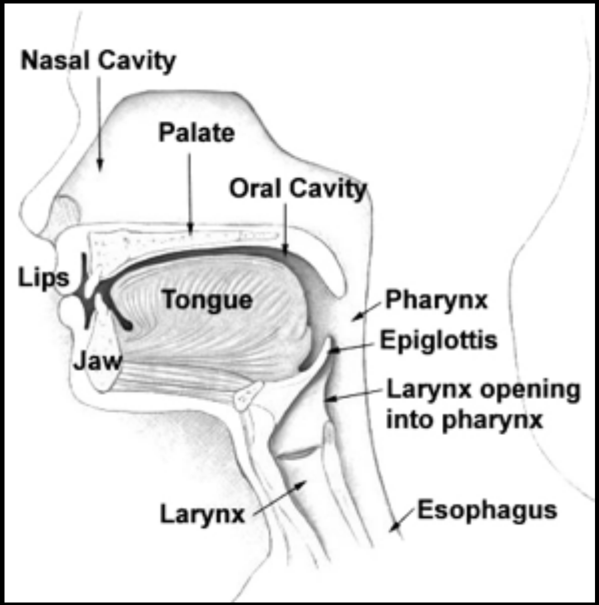
Noise

- Non-tonal sound (non-repeating waveform)
  - continuous, non-harmonic spectrum
- Random (not predictable) signal
- But the distribution of values/frequencies matters
  - white noise, pink/blue noise, popcorn noise, etc.



Voice and Speech

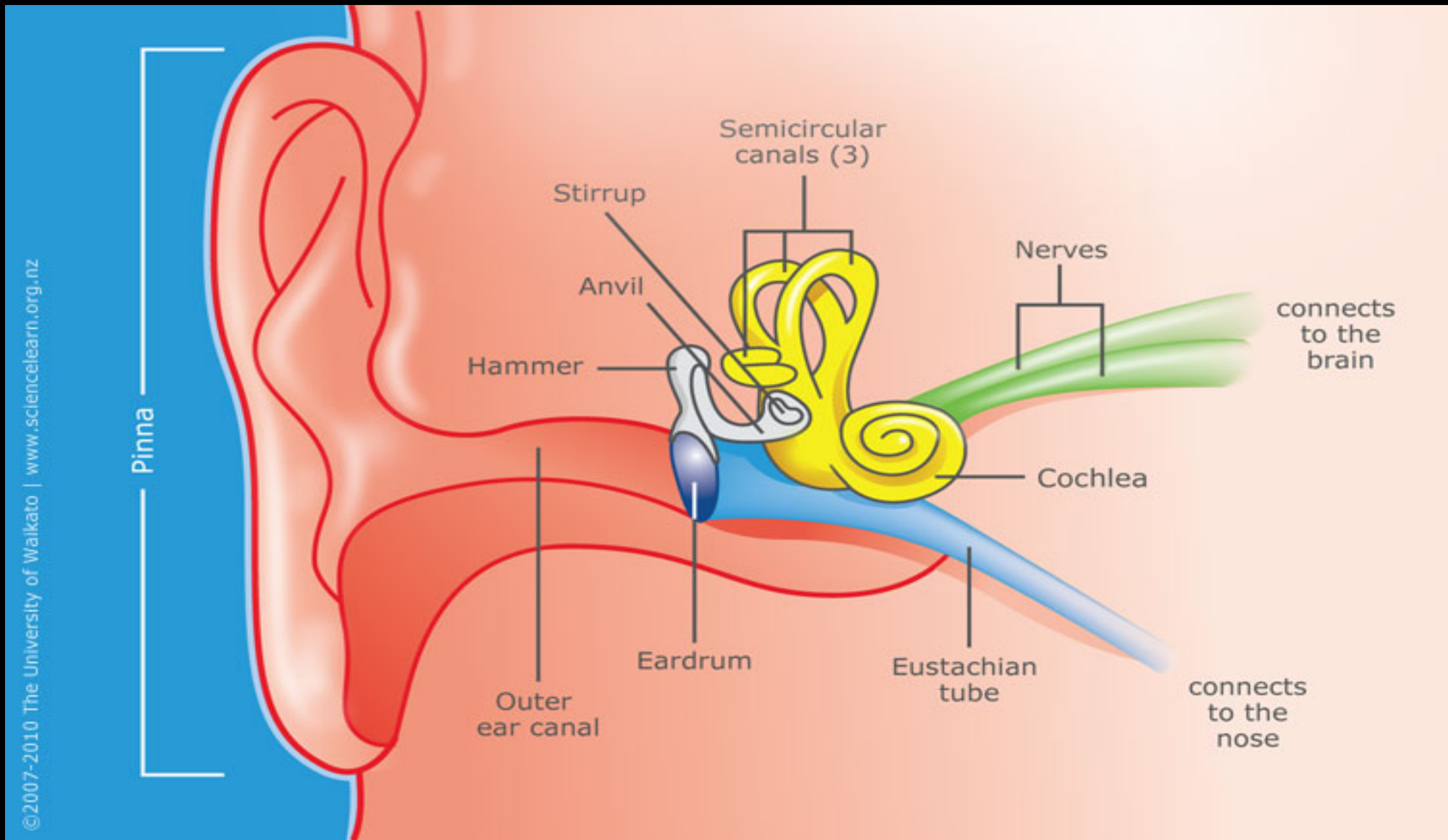
- Formed by the vocal tract
- Vowels (*a, e, i, …*) have a harmonic spectrum characterized by *formants* (peaks at certain kHz frequencies)
- Consonants
  - harmonic: voiced nasals (*b, m, n, j, …*)
  - transients (*k, p, t*)
  - continuous noise (*f, s, …*)



https://en.wikipedia.org/wiki/Articulation_(phonetics))


Hearing

Anatomy of the Ear



How the Ear Works

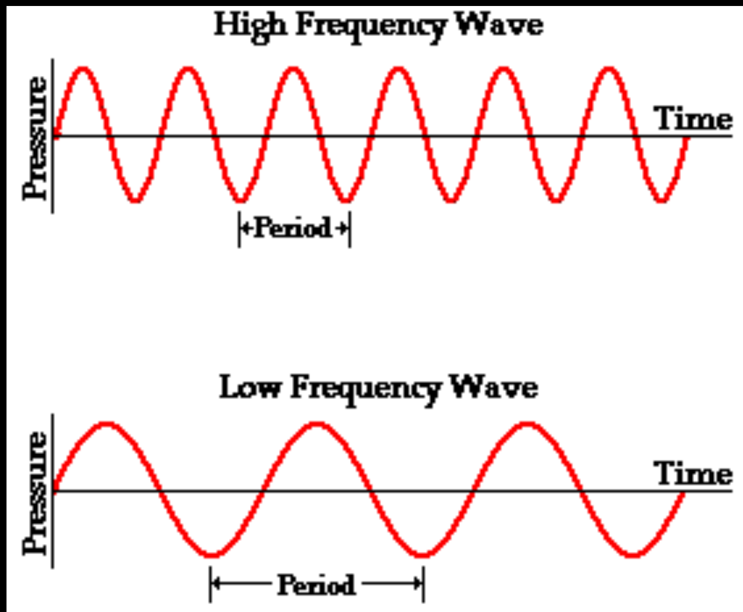
- https://www.youtube.com/watch?v=pCCcFDoyBxM

Sound Frequency and Amplitude

- Frequency determines the *pitch* of the sound
- Amplitude relates to intensity of the sound
  – *Loudness* is a subjective measure of intensity

High frequency = short period

Low frequency = long period

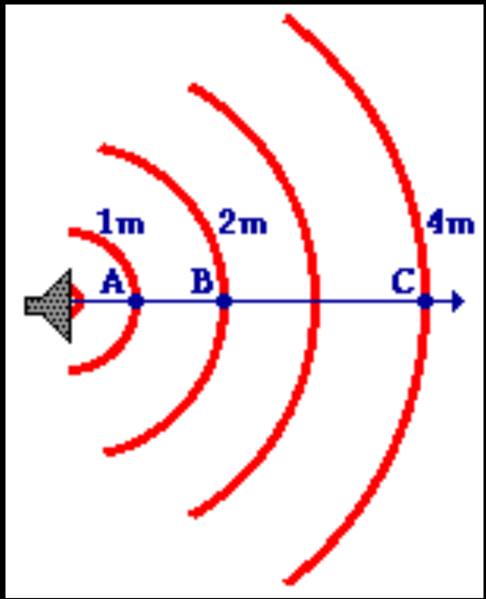


Distance to Listener

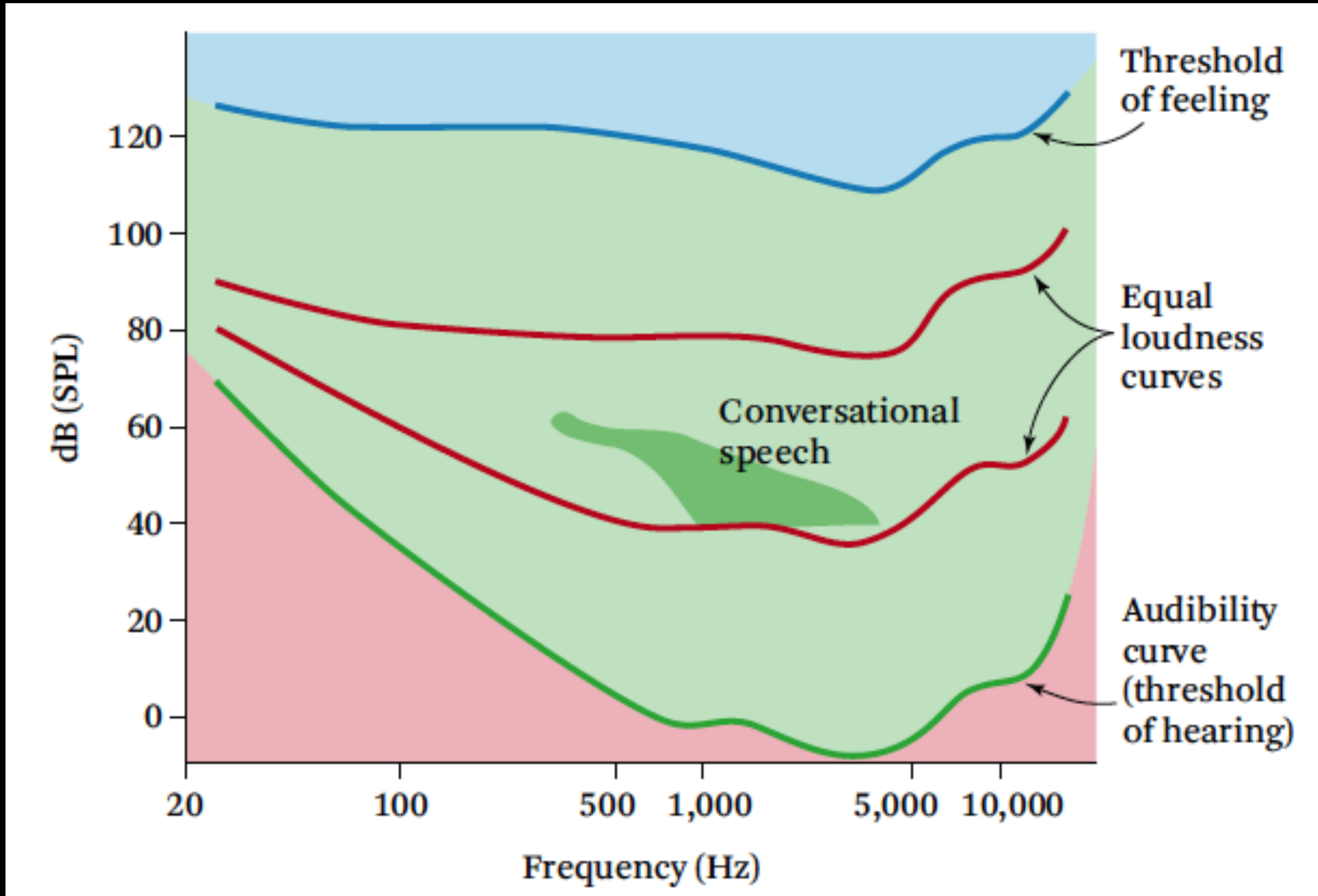
- Relationship between sound intensity and distance to the listener

  Inverse-square law

  - The intensity varies inversely with the square of the distance from the source. So if the distance from the source is doubled (increased by a factor of 2), then the intensity is quartered (decreased by a factor of 4).



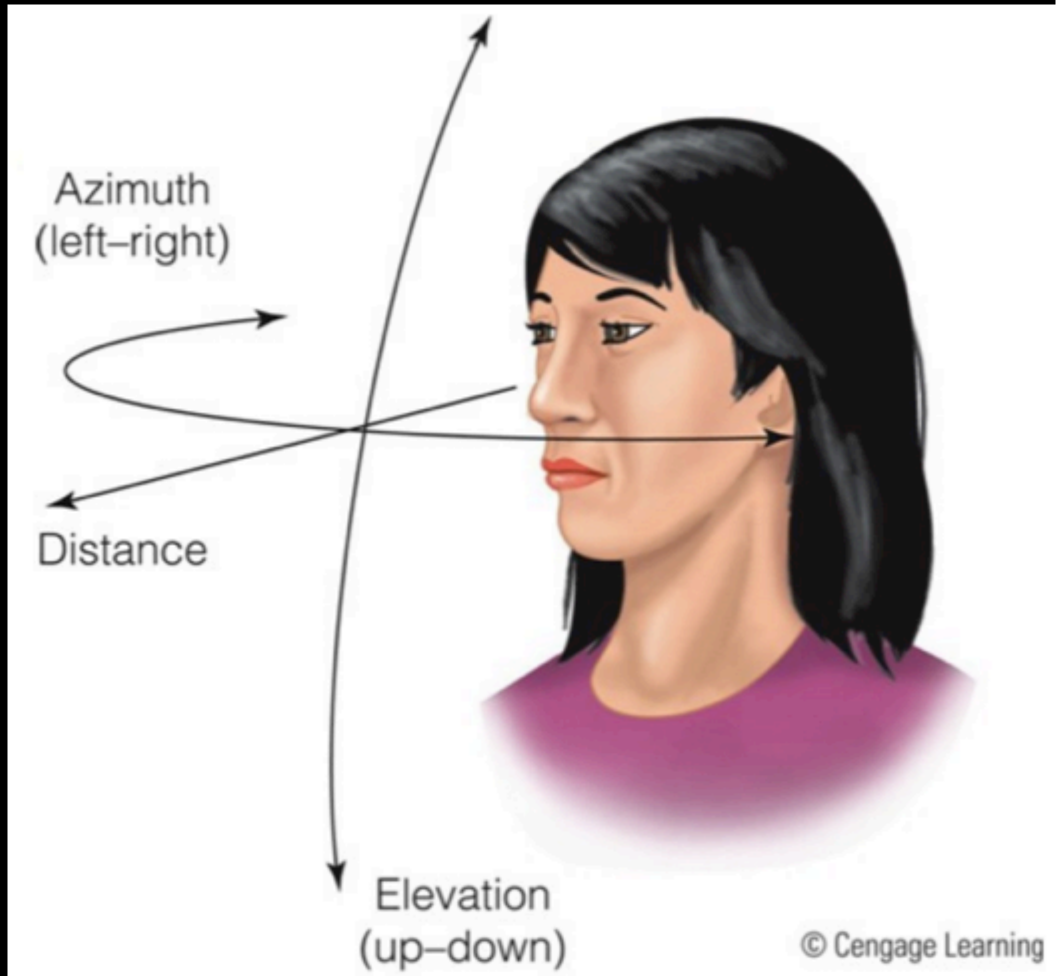
Auditory Thresholds



- Humans hear frequencies from 20 – 22,000 Hz
- Most everyday sounds from 40 – 80 dB

Sound Localization

- Humans have two ears
  - localize sound in space
- Sound can be localized using 3 coordinates
  - Azimuth, elevation, distance



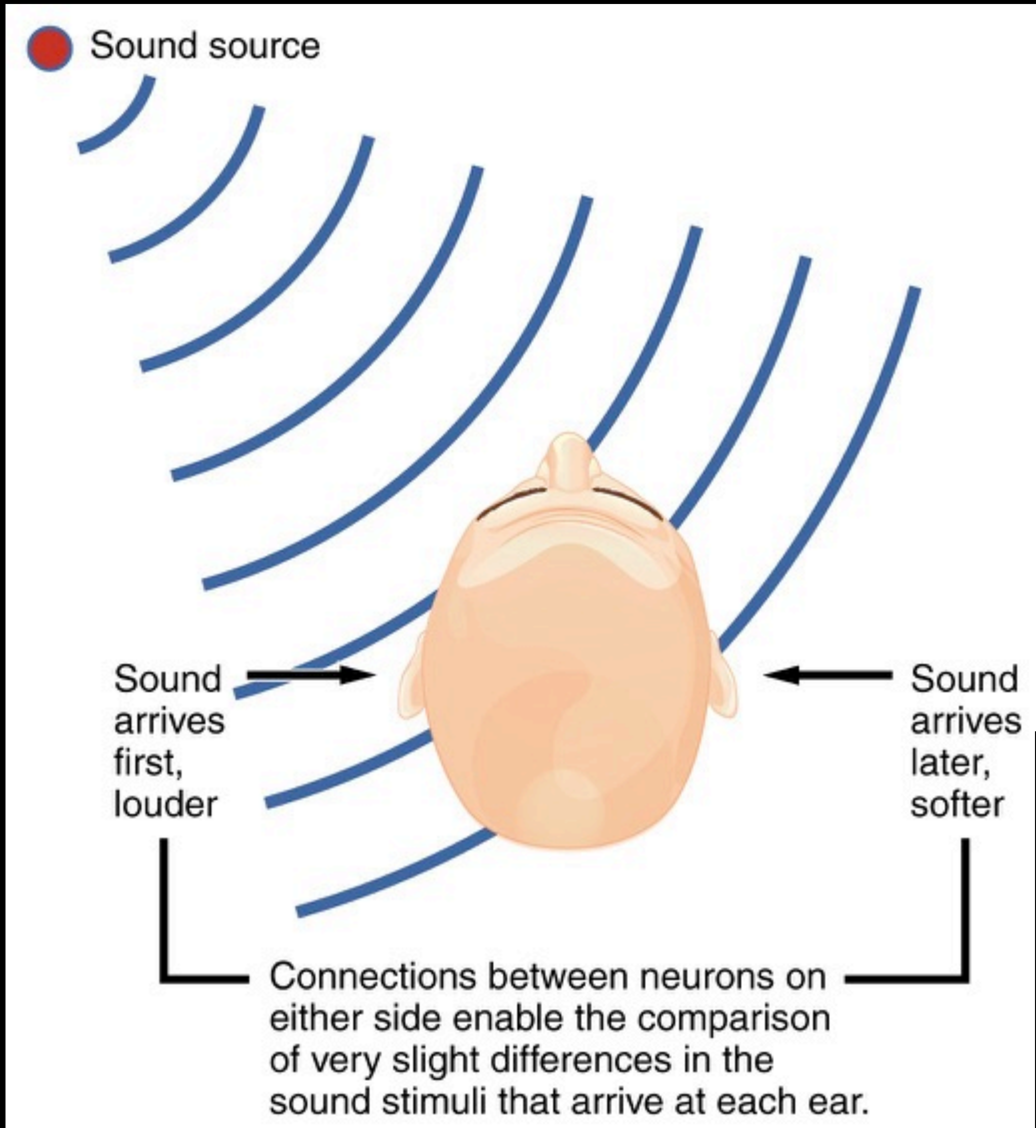
Sound Localization

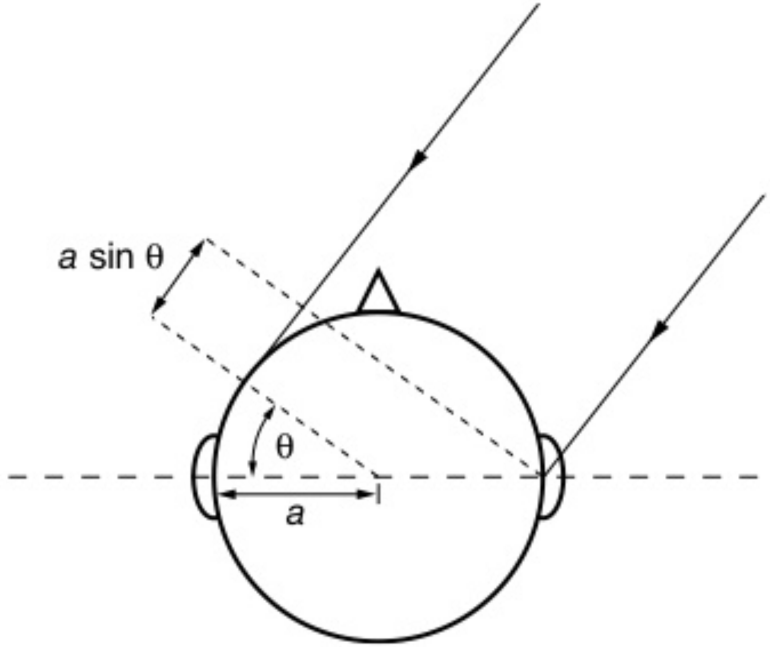
- Azimuth Cues
  - Difference in **time** of sound reaching two ears
    - Interaural time difference (ITD)
  - Difference in sound **intensity** reaching two ears
    - Interaural level difference (ILD)
- Elevation Cues
  - Monaural cues derived from the pinna (ear shape)
    - Head related transfer function (HRTF)
- Range Cues
  - Difference in sound relative to range from observer
  - Head movements (otherwise ITD and ILD are same)

Sound Localization

- https://www.youtube.com/watch?v=FIU1bNSlbxk

Sound Localization (Azimuth Cues)



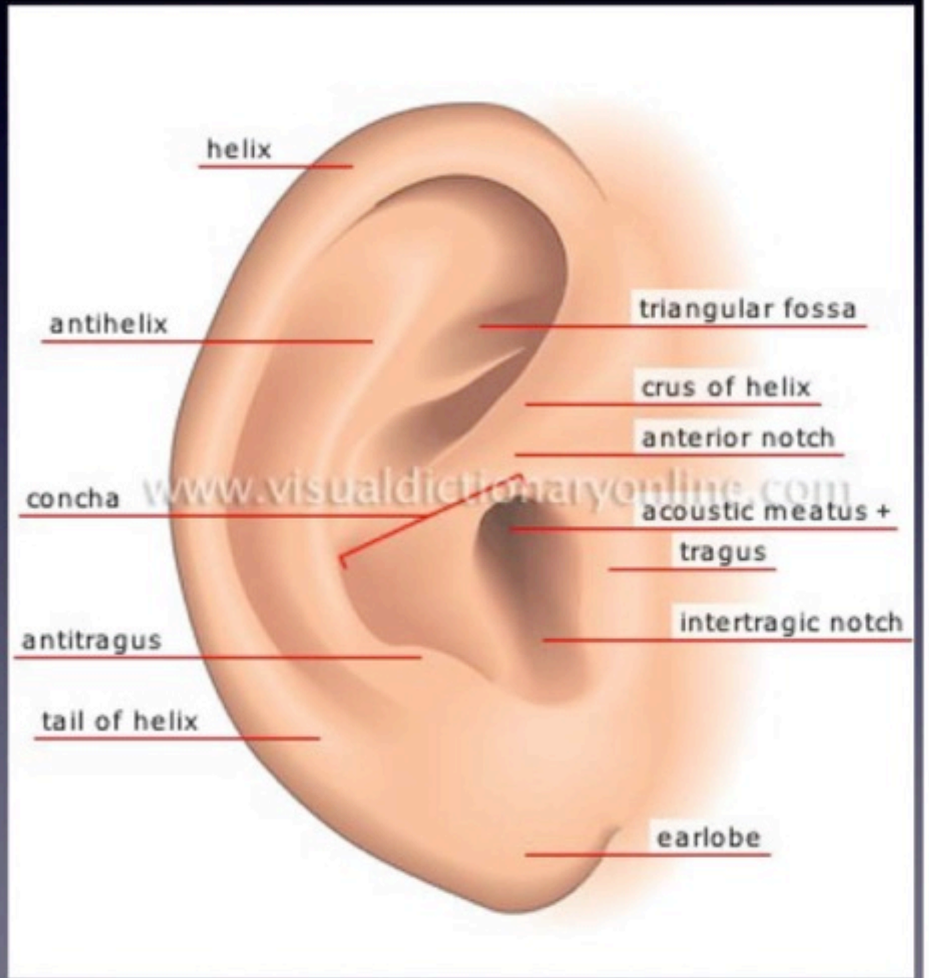


Interaural Time Difference (ITD)

Interaural Level Difference (ILD)

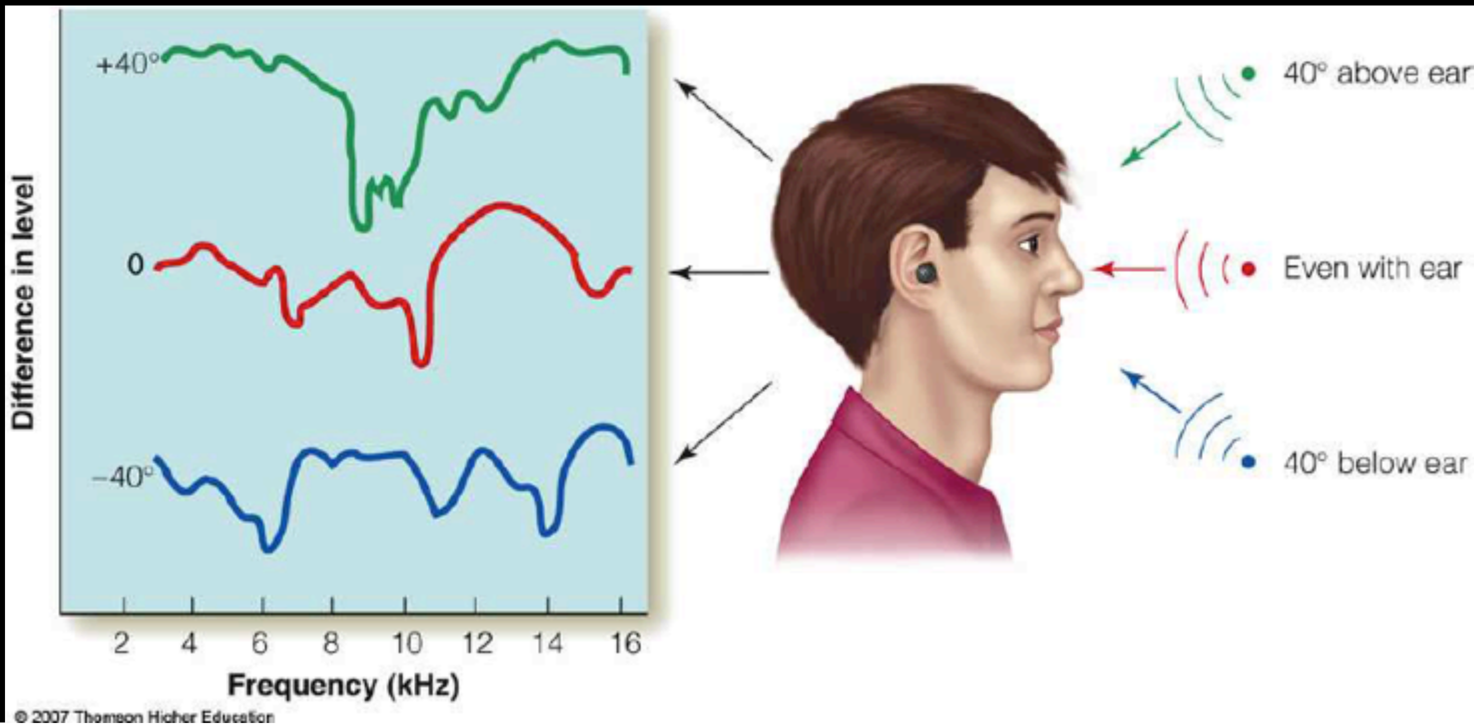
Pinnae Cues

- The outer ears also modify the sound
- Help to deal with front/ back confusion
- Primary Cue for Elevation
- Very Individual, we “learn” our ears during childhood.



HRTF (Elevation Cue)

- Pinna and head shape affect frequency intensities
- Sound intensities measured with microphones in ear and compared to intensities at sound source
  - Difference is HRTF, gives clue as to sound source location



Accuracy of Sound Localization

- People can locate sound
  - Most accurately in front of them
    - 2-3° error in front of head
  - Least accurately to sides and behind head
    - Up to 20° error to side of head
    - Largest errors occur above/below elevations and behind head
- Front/back confusion is an issue
  - Up to 10% of sounds presented in the front are perceived coming from behind and vice versa (more in headphones)

BUTEAN, A., Bălan, O., NEGOI, I., Moldoveanu, F., & Moldoveanu, A. (2015). COMPARATIVE RESEARCH ON SOUND LOCALIZATION ACCURACY IN THE FREE-FIELD AND VIRTUAL AUDITORY DISPLAYS. In*Conference proceedings of» eLearning and Software for Education «(eLSE)*(No. 01, pp. 540-548). Universitatea Nationala de Aparare Carol I.

How can we generate sound

- Record and playback real sounds
- Or samples of them
    - E.g speech, or musical instruments
- Build a virtual model and generate sounds
- What about in 3D?

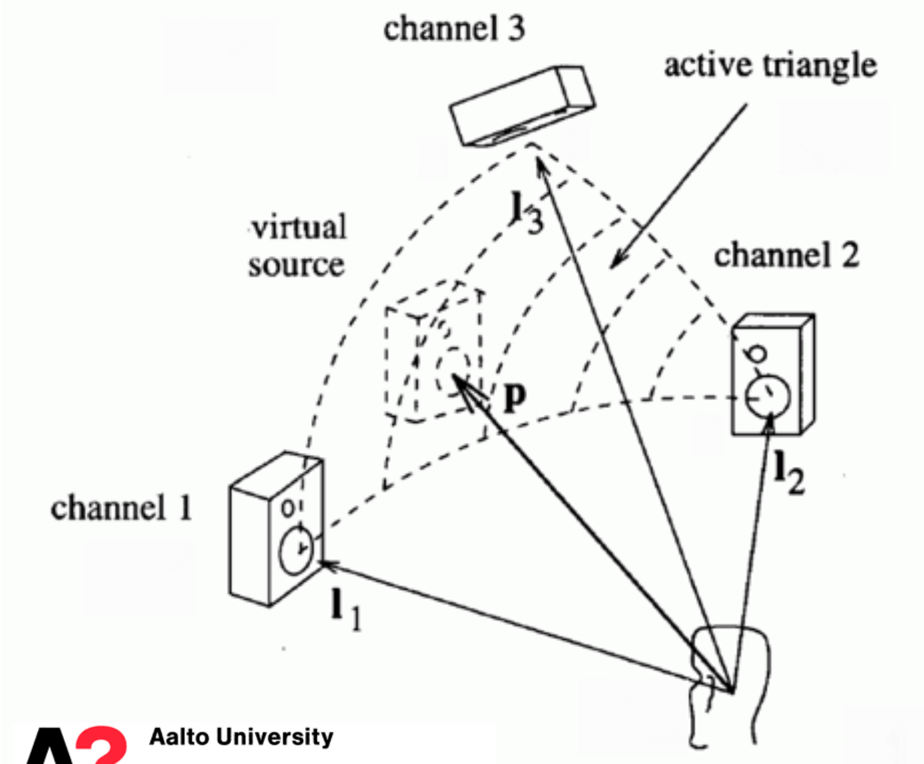
[redacted]

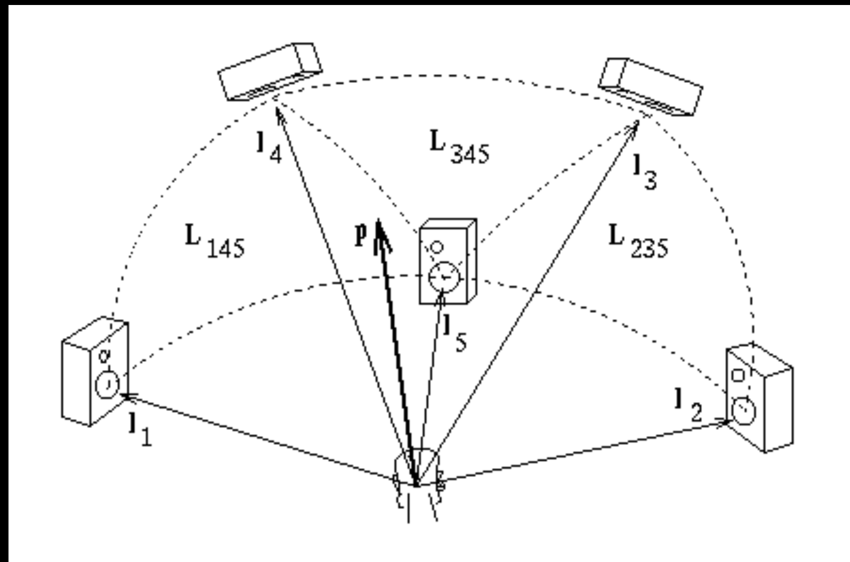
3D Synthetic Sound

- An extension on Stereo Sound
  - Vary sound intensity in each ear
- Can apply to speaker array
  - Requires calibration
- Or virtually over headphones

Vector Base Amplitude Panning

- Divide a virtual sound source's signal among the three loudspeakers nearest to source direction
- Can be realized with any number of loudspeakers





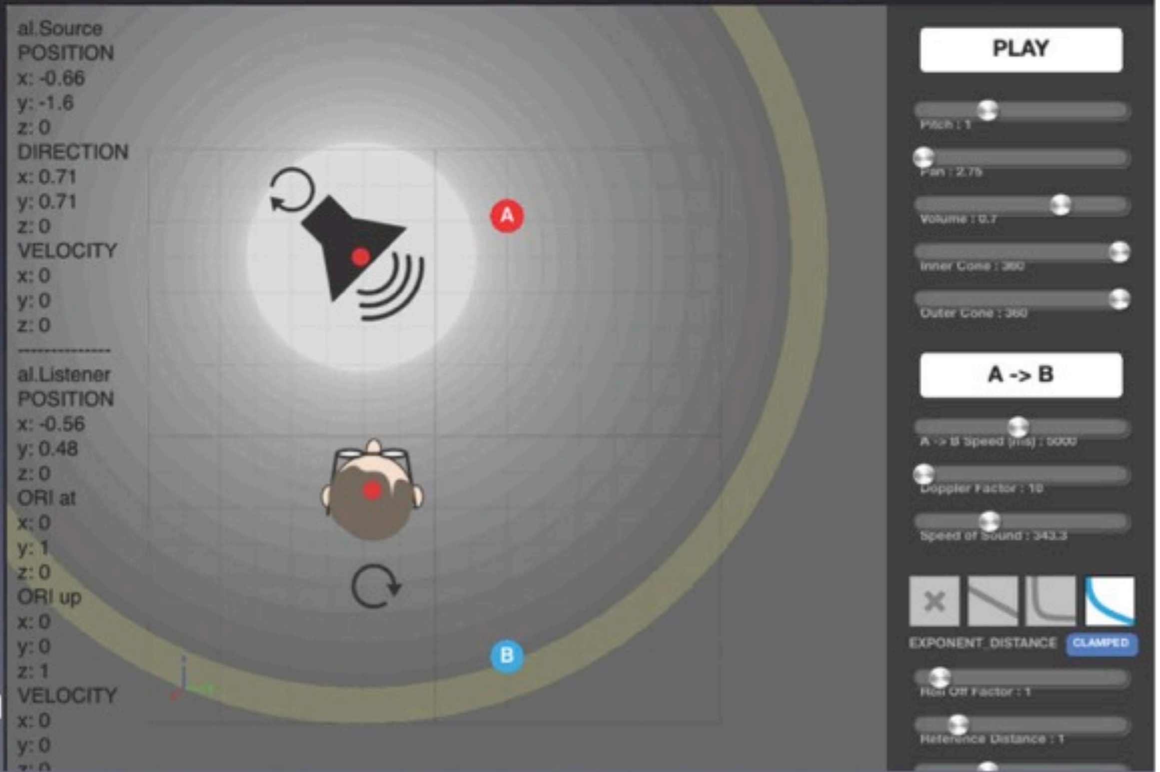
http://legacy.spa.aalto.fi/research/cat/vbap/

Virtual 3D

- Builds an HRTF (Head Related Transfer Function)
- Functions map how sound changes as it interacts with the head.
- Creates filters at different angles
- Can be individualised or Generalised -> GHRTF

Virtual 3D Audio

- GHRTFs are most common
  - Built into smartphones now
- Programmed using an OpenGL model
- I-HRTFs are better
- But direction is comparable, not elevation



- Can combine with sensors to create AAR (Audio Augmented Reality)

Sound in the user interface

- Output: Auditory Display
  - analogous to visual display, but sound is a temporal medium
  - sonification = representing information with sound
    - continuous sound: system state / object property
    - transients: events
- Input: Sound Recognition
  - analysing sound structure
  - interpreting as events / input values

more info & some examples: http://www.icad.org/audio.php

Types of Audio

- Speech
- Music
- Sound Effects
- Encoded Representations
    - Earcons
    - Auditory Icons
    - Musicons
    - Spearcons
- Sonification
    - Direct Parameter Mapping

[redacted]

Auditory Icons

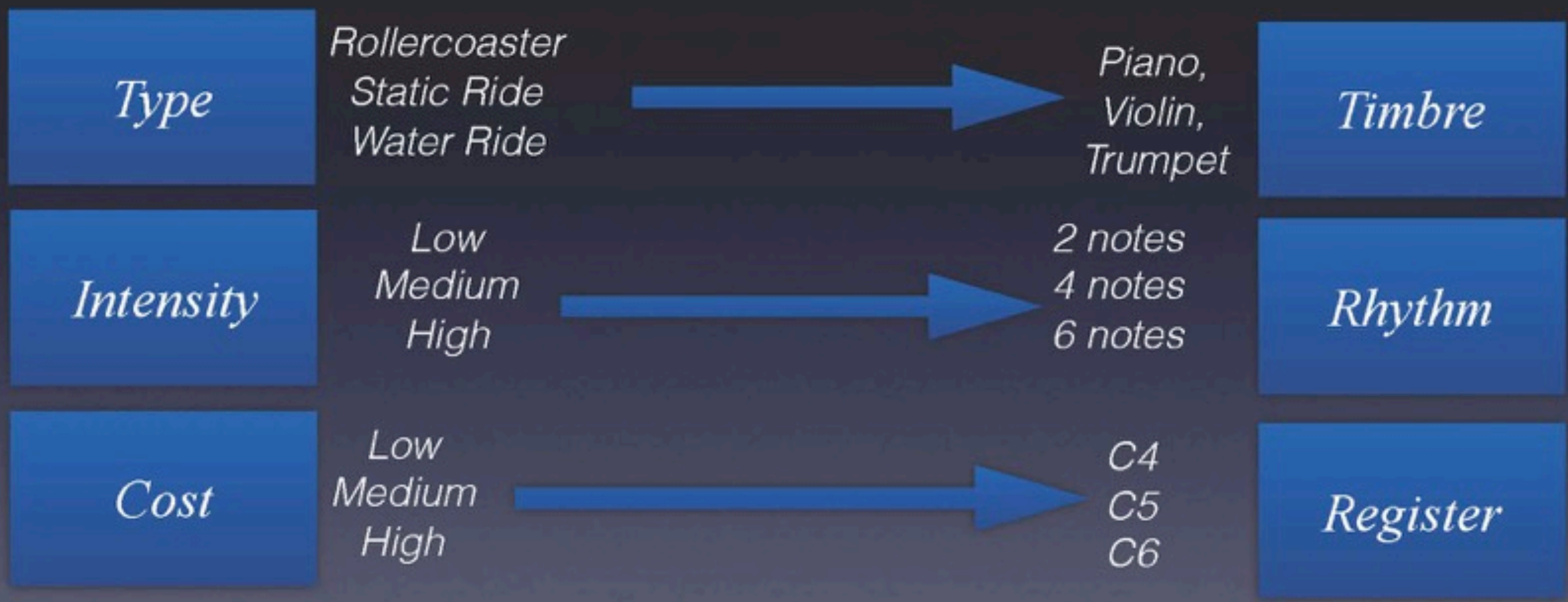
- *"Everyday sounds mapped to computer events by analogy with everyday sound-producing events" - Gaver (1989)*
- Familiar sounds map to computer events
  - E.g. breaking glass for error
  - Lots of material properties for selection of windows
- Application in many scenarios but sometimes difficult to get the right sounds.

Earcons

- *"non-verbal audio messages used in the user-computer interface to provide information to the user about some computer object, operation, or interaction". - Blattner (1989)*
- Essentially, short musical sounds that are systematically altered to encode information
- Mapping is abstract, but encoding allows a lot of information in just a few rules

Example of mapping features to sounds

Theme Park Rides



Just need to learn the rules and can understand any

[redacted]

Spearcons

- Walker et al. (2006)
- Earcons vs Auditory Icons Problem
- Solution -> Speeded up Speech
- Users can easily pickup and retain these sounds
- Useful in e.g. menus?

Musicons

- McGee-Lennon et al. (2011) studied home reminders
- REALLY short snippets of music (< 0.1 sec)
- Found high accuracy in remembering (94%)
- Participants treated music as memes
  - Friends theme -> Putting the trash out
- People will form their own mappings

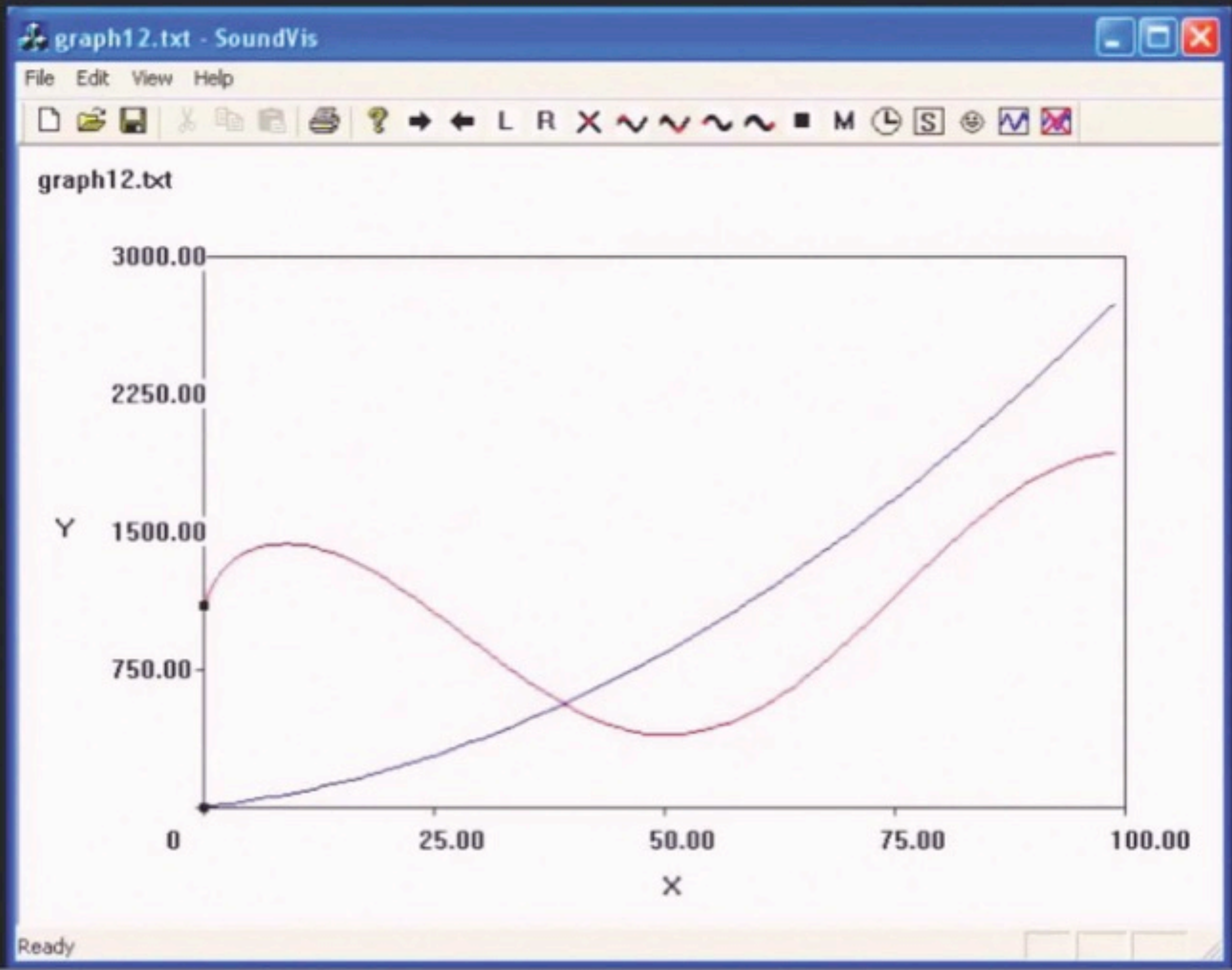
Sonification

- The use of non-speech audio to perceptualise data
    - Generally considered continuous data (e.g. graphs)
- You've seen an example in Tangible Graph Builder
    - Great for VI users
    - Used in "big data"
    - Or non-visual scenarios (navigation, heart rate monitoring?)
- Usually uses direct parameter mapping
- 1 data parameter => 1 auditory attribute

[redacted]

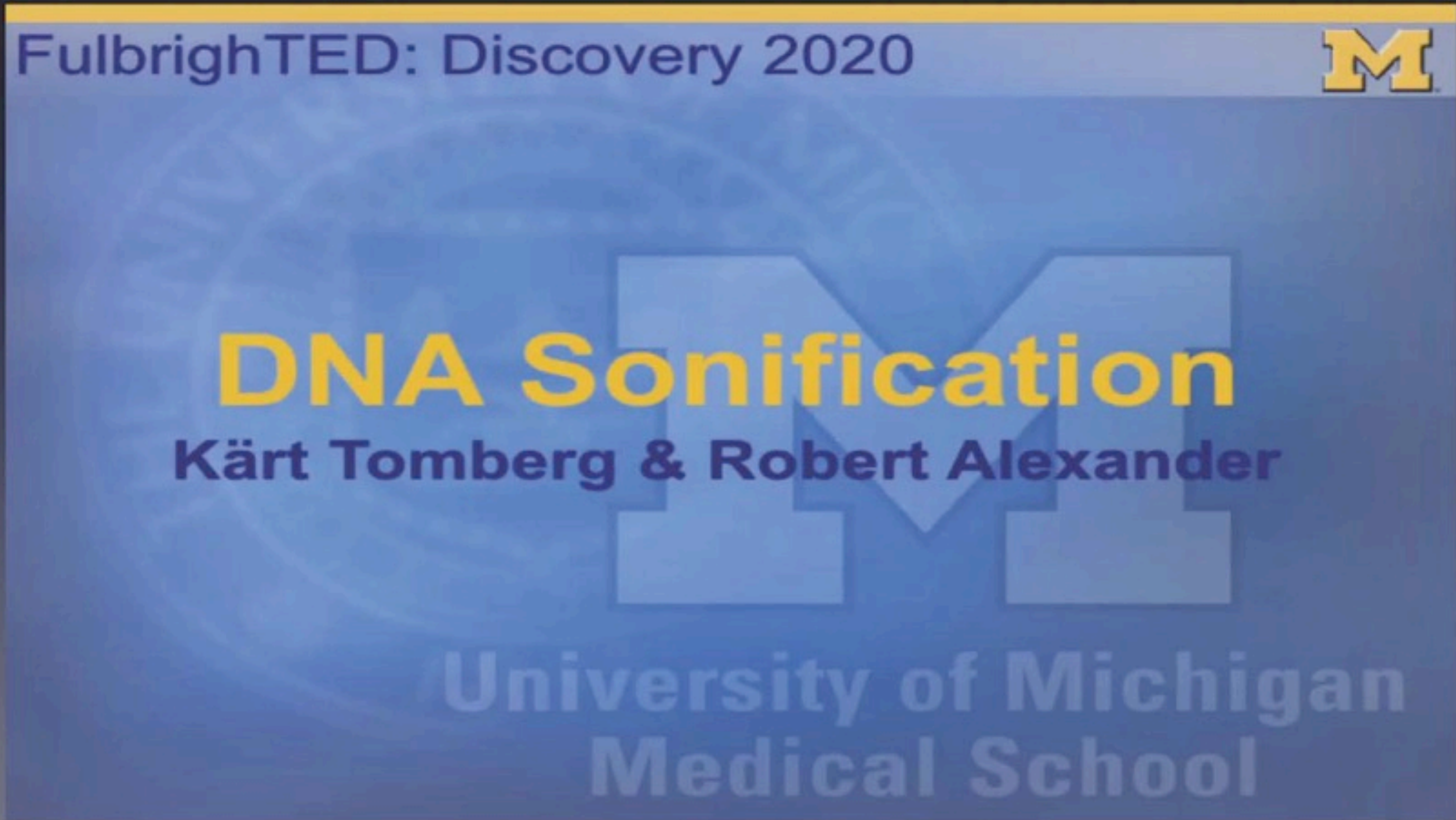
https://www.youtube.com/watch?v=dplpCW-P77o

Auditory Graph



http://dnasonification.org

Big Data Sonification



Where to Use Sound

- **Advantages**
  - Sound in Omni-directional
    - Don’t need to attend to it
  - Can also be discrete
  - Selective attention
- **Disadvantages**
  - Annoying
  - Attention Grabbing
  - Can take away from the environment
  - Encoded sound (sometimes) to be learned

CONTINUES …