NYU

Virtual and Augmented Reality

CS-GY 9223/CUSP-GX 6004

https://nyu-icl.github.io/courses/2022fall-vr-ar

Prof. Qi Sun
qisun@nyu.edu | www.qisun.me

Final Presentation Next Week

Submit EO Sun: We need to prepare

You shall have voice-over

No need for ultra-high res: 720P-1080P mp4 is fine enough

Report/code to open soon (don’t worry about them now)

1-1s Today

Spatial Sound

- what is sound? how do we synthesize it?
- the human auditory system
- stereophonic sound
- spatial audio of point sound sources
- surround sound
- ambisonics

What is Sound?

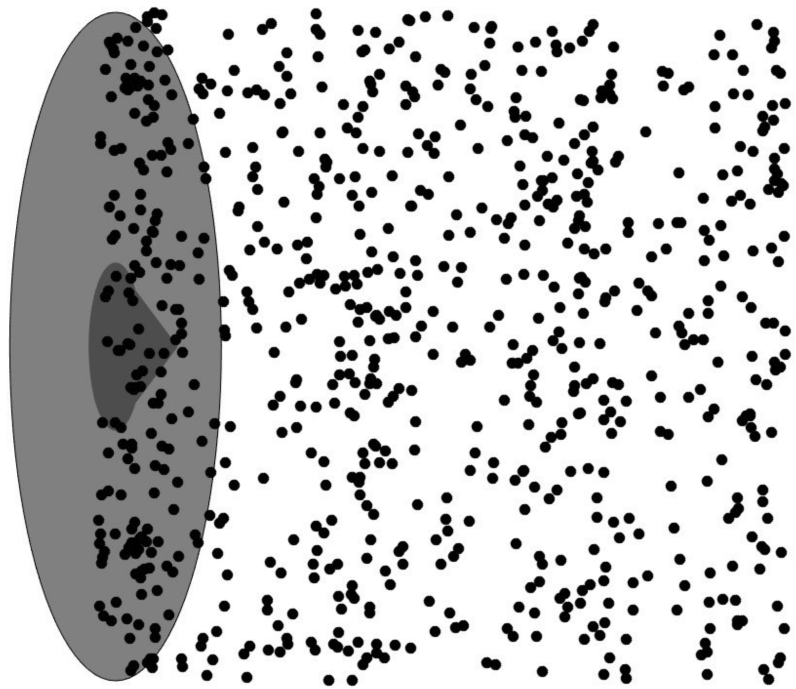
- “sound” is a pressure wave propagating in a medium
- speed of sound is $c = \sqrt{K/\rho}$ where c is velocity, ρ is density of medium and K is elastic bulk modulus
- in air, speed of sound is 340 m/s
- in water, speed of sound is 1,483 m/s

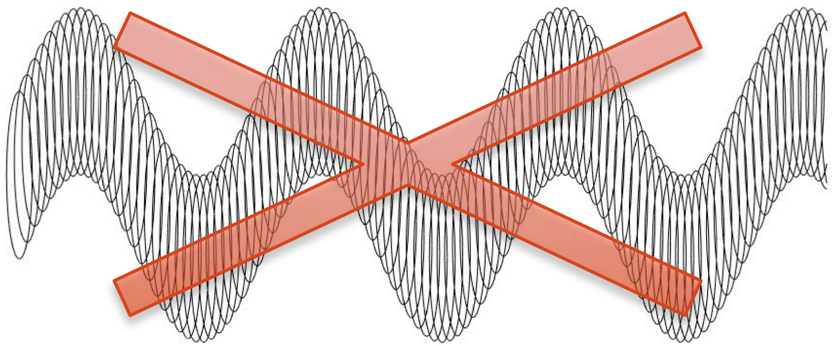
How do we Synthesize Sound?

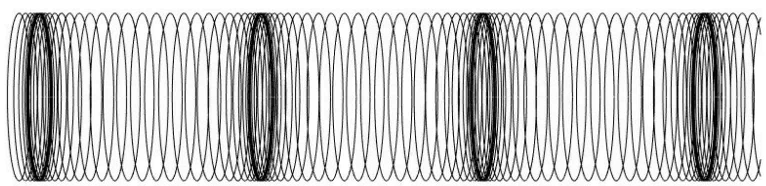
https://www.youtube.com/watch?v=aDrs6EieFCM

Producing Sound

- Sound is longitudinal vibration of air particles
- Speakers create wavefronts by physically compressing the air, much like one could a slinky

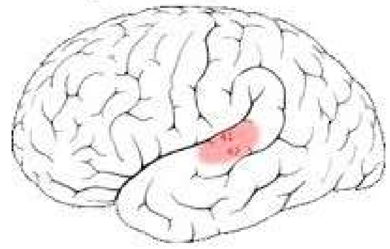




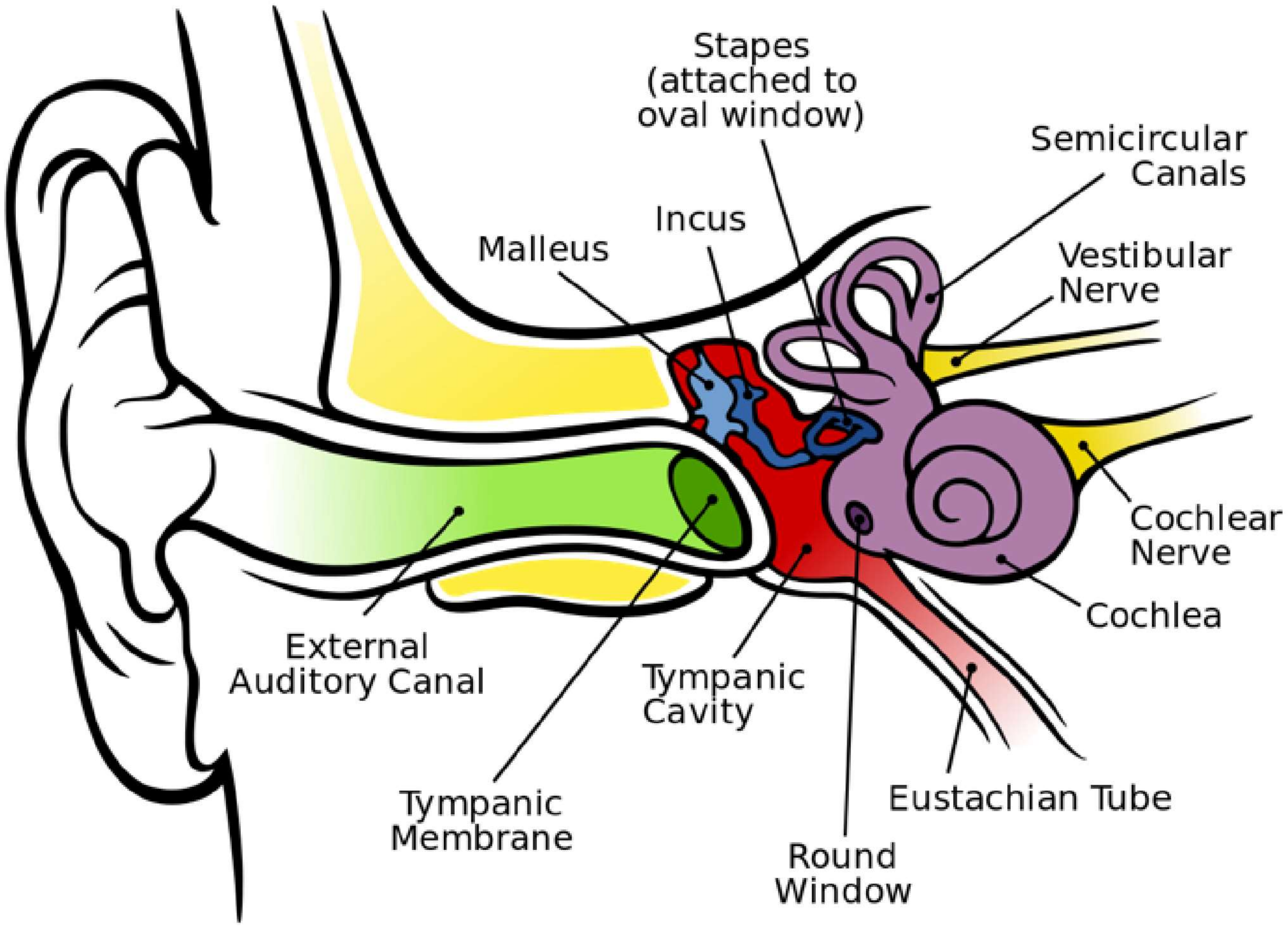


The Human Auditory System

Primary auditory cortex



pinna

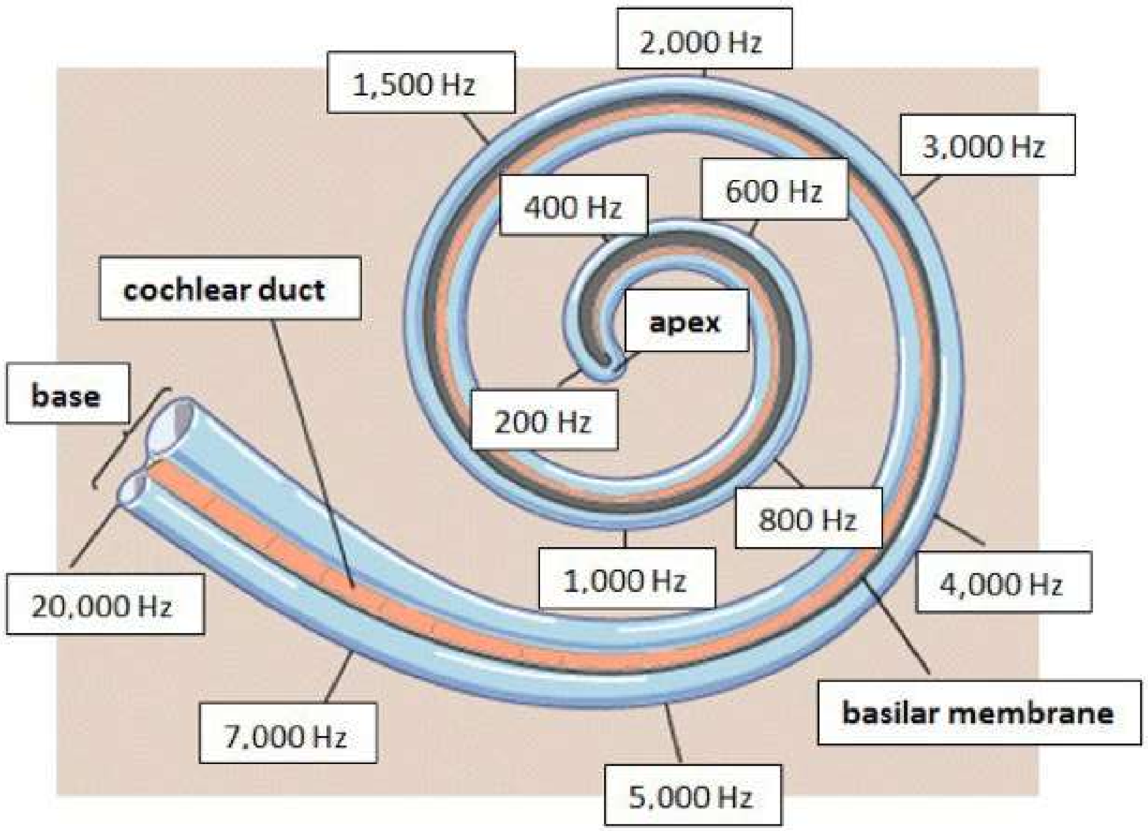


wikipedia

The Human Auditory System

- hair receptor cells pick up vibrations

cochlea

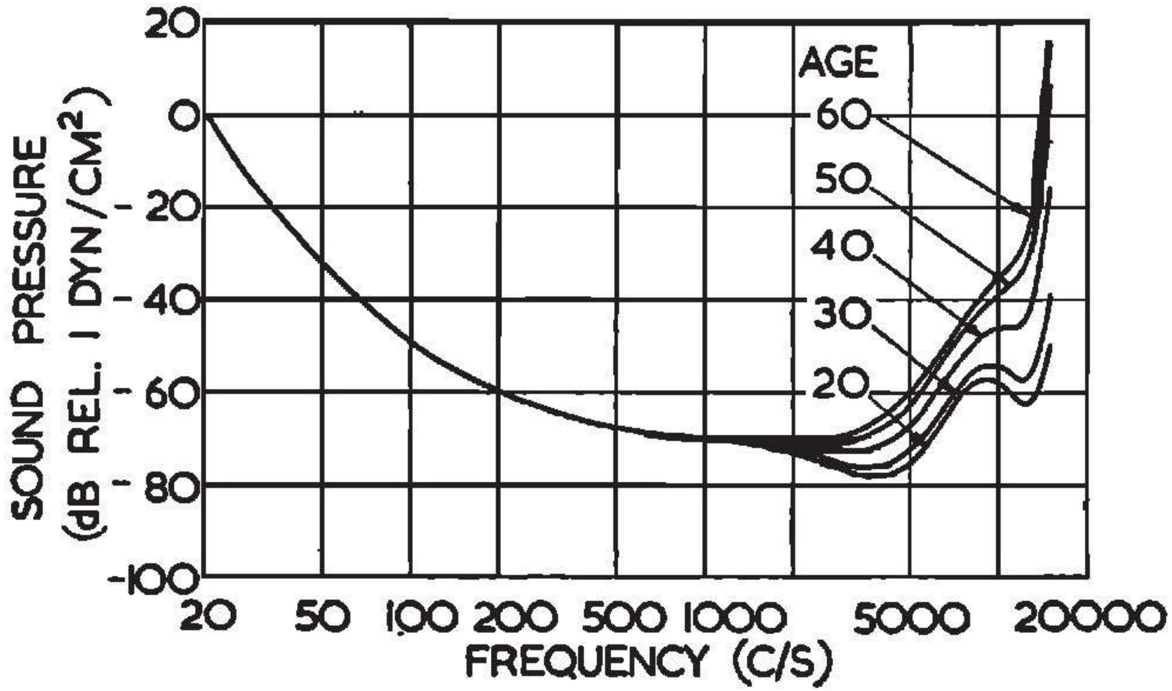


wikipedia

The Human Auditory System

- Human hearing range: ~20–20,000 Hz
- Variation between individuals
- Degrades with age

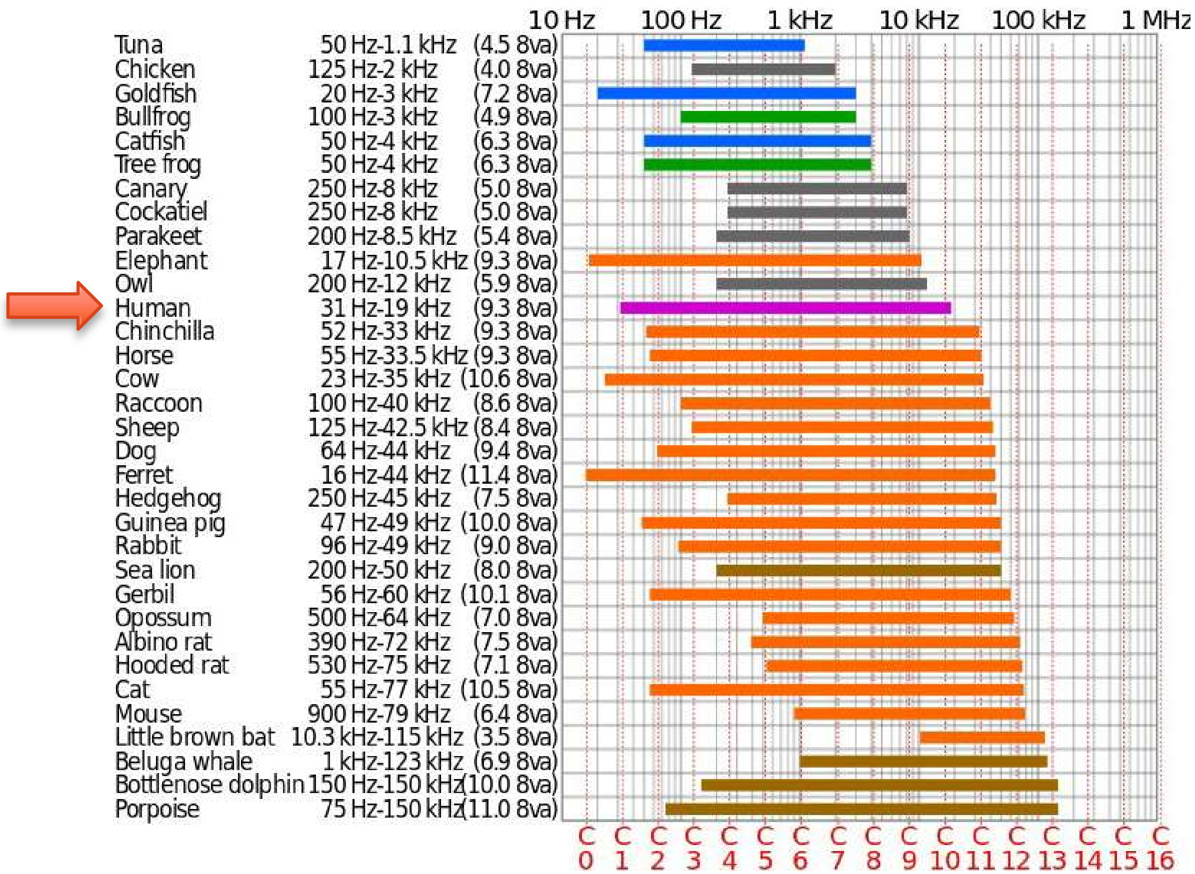
Hearing Threshold in Quiet



D. W. Robinson and R. S. Dadson, 1957

The Human Auditory System

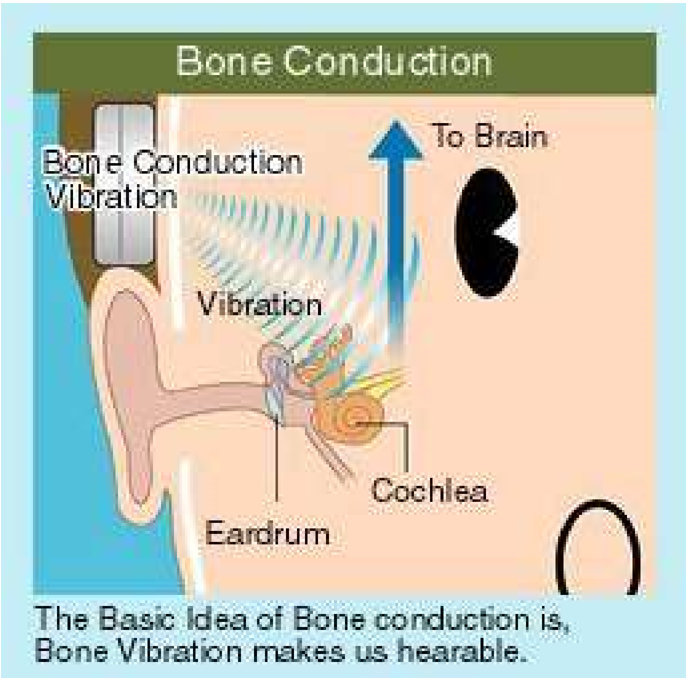
- human hearing range: ~20 – 20,000 Hz
- variation between individuals and changes with age



wikipedia

Bone Conduction

- can stimulate eardrum mechanically to create the illusion of audio, e.g. with bone conduction

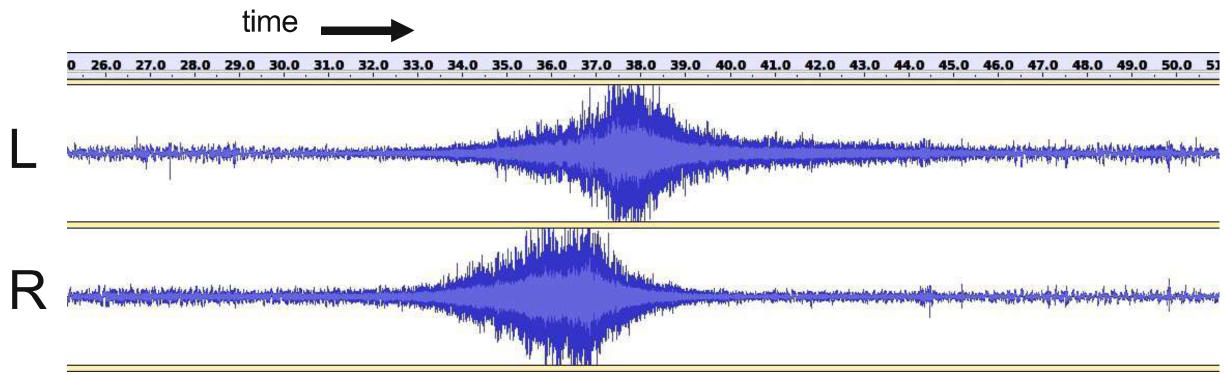


http://www.goldendance.co.jp/English/boneconduct/01.html

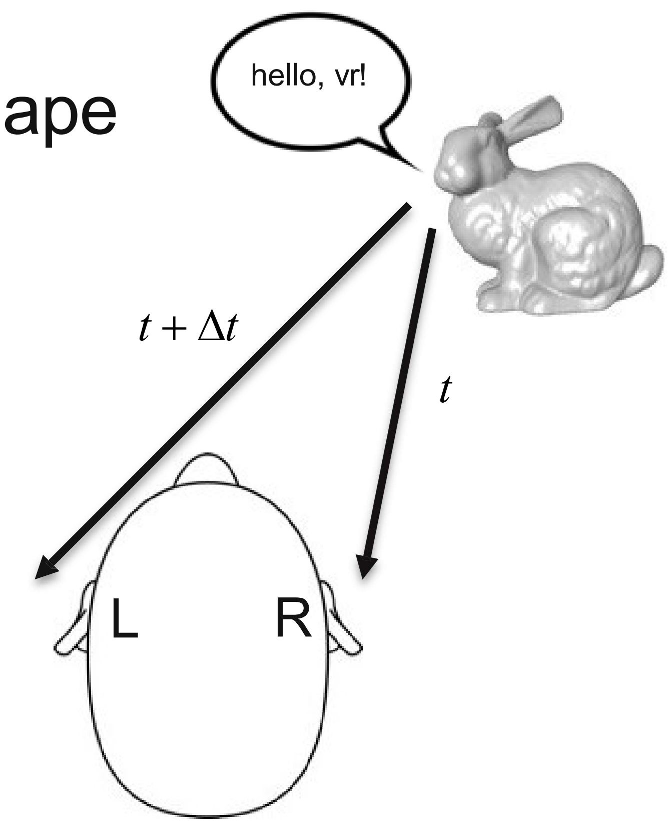
the verge

Stereophonic Sound

- mainly captures differences between the ears:
  - interaural time difference
  - amplitude differences from body shape (nose, head, neck, shoulders, …)



wikipedia

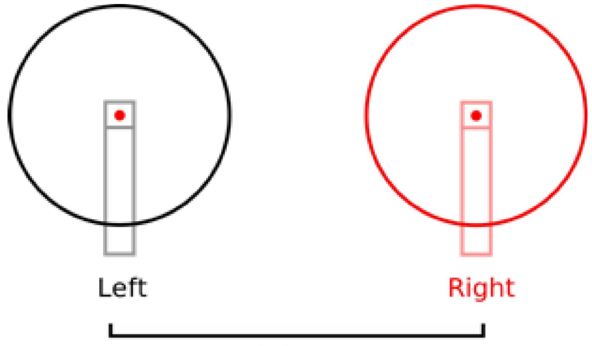


Stereophonic Sound Recording

- use two microphones

- A-B techniques captures differences in time-of-arrival

Olympus



wikipedi

- other configurations work too, capture differences in amplitude

Rode

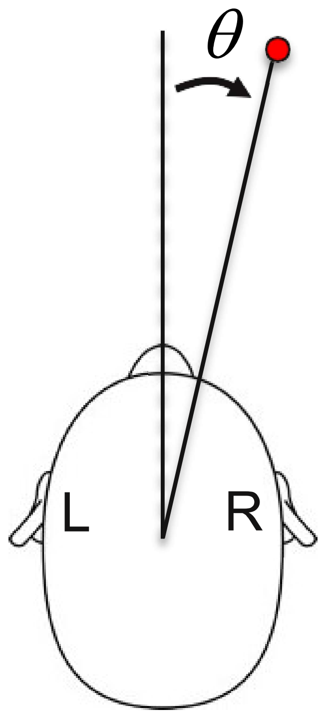
X-Y technique

Head-related Impulse Response (HRIR)

- models phase and amplitude differences for all possible sound directions parameterized by azimuth θ and elevation φ

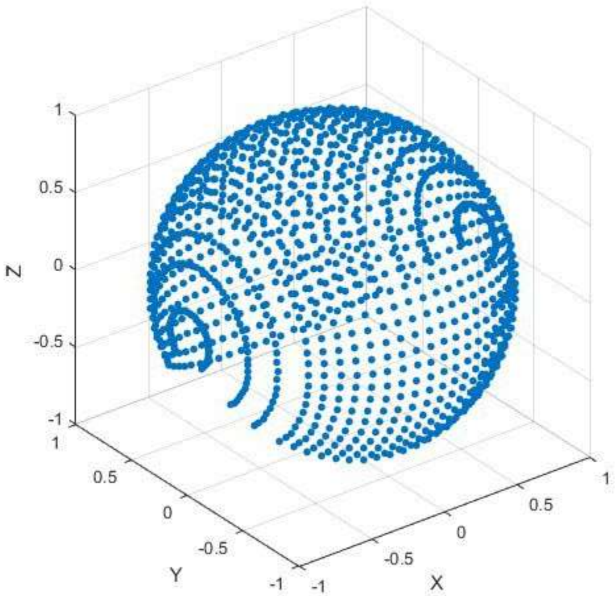
- can be measured with two microphones in ears of mannequin & speakers all around

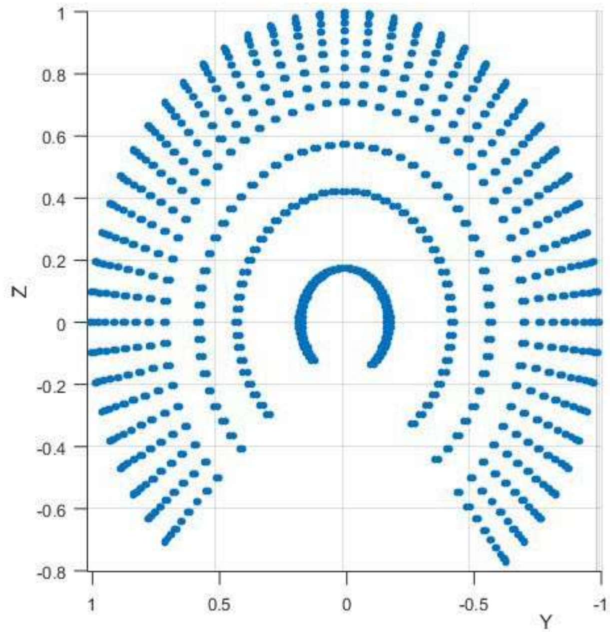
Zhong and Xie, “Head-Related Transfer Functions and Virtual Auditory Display”



Head-related Impulse Response (HRIR)

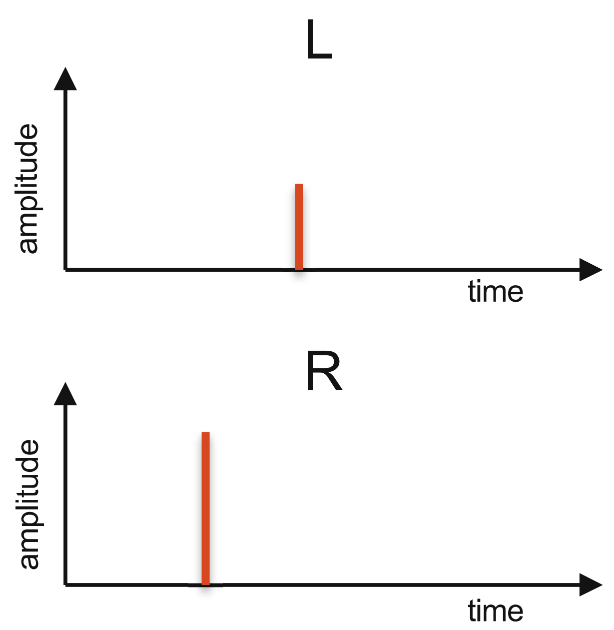
- CIPIC HRTF database: http://interface.cipic.ucdavis.edu/sound/hrtf.html
- elevation: -45° to 230.625°, azimuth: -80° to 80°
- need to interpolate between discretely sampled directions

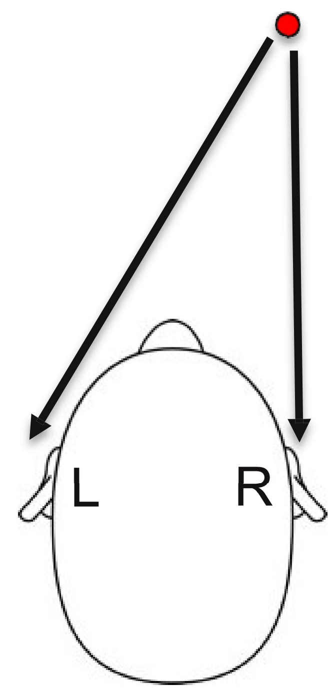




Head-related Impulse Response (HRIR)

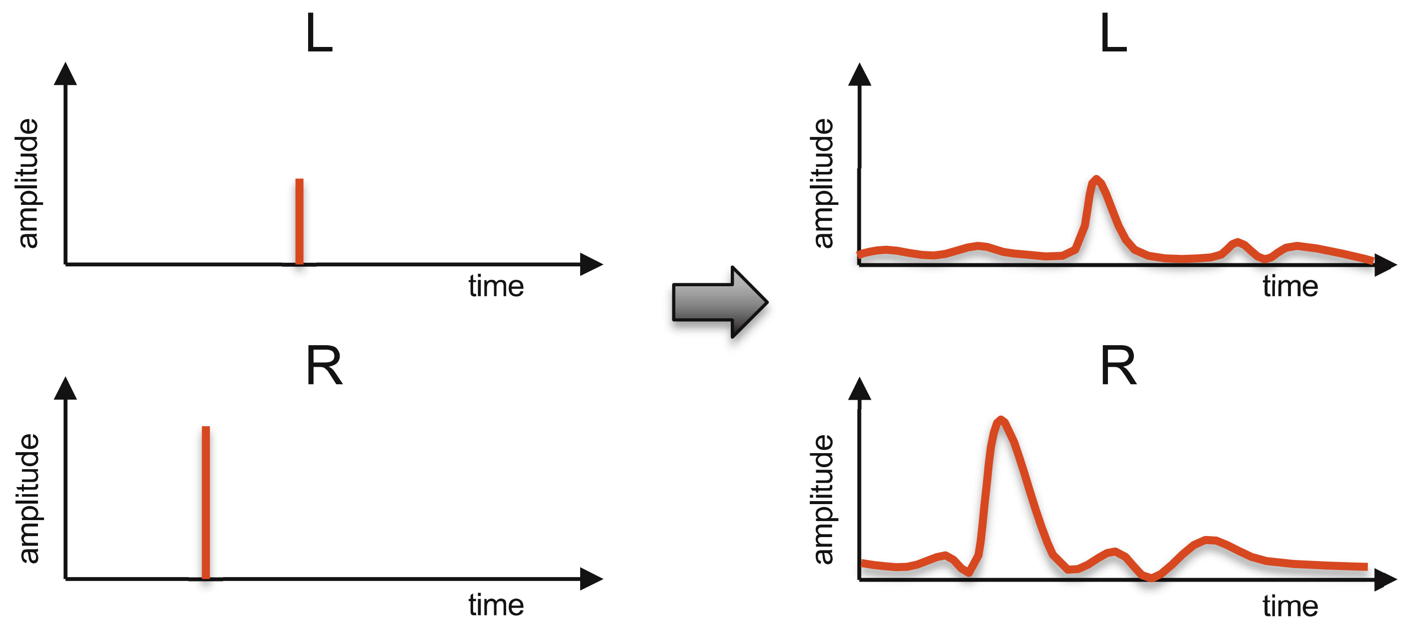
- measuring the HRIR
    - ideal case: scaled & shifted Dirac peaks

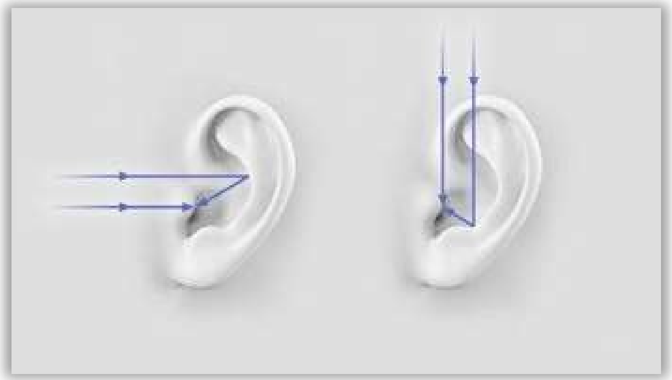




Head-related Impulse Response (HRIR)

- measuring the HRIR
  - ideal case: scaled & shifted Dirac peaks
  - in practice: more complicated, includes scattering in the ear, sholders etc.





Head-related Impulse Response (HRIR)

- measuring the HRIR
  - need one temporally-varying function for each angle
  - total of $2 \cdot N_\theta \cdot N_\varphi \cdot N_t$ samples, where $N_{\theta,\varphi,t}$ is the number of samples for azimuth, elevation, and time, respectively

$$hrir\_l(\theta,\varphi,t)$$

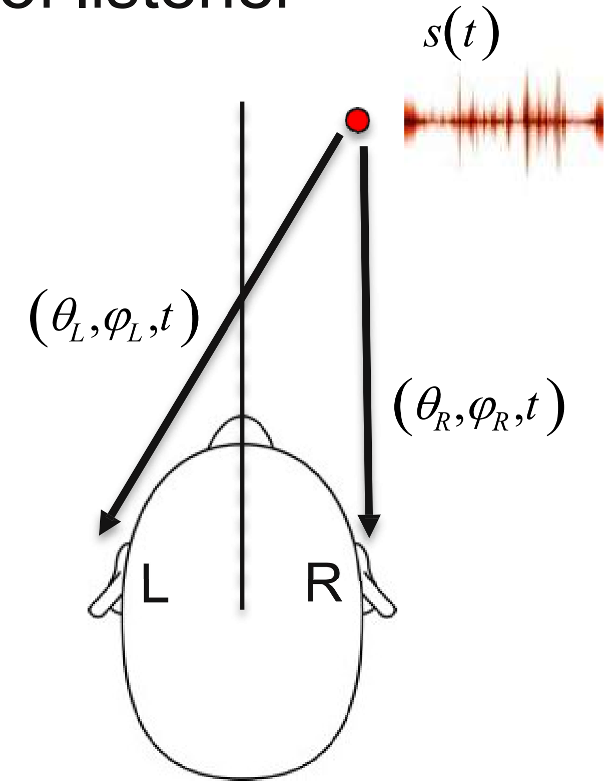
$$hrir\_r(\theta,\varphi,t)$$

Head-related Impulse Response (HRIR)

applying the HRIR:

- given a mono sound source $s(t)$ and it's 3D position

1. compute (θ_L, φ_L) and (θ_R, φ_R) relative to center of listener

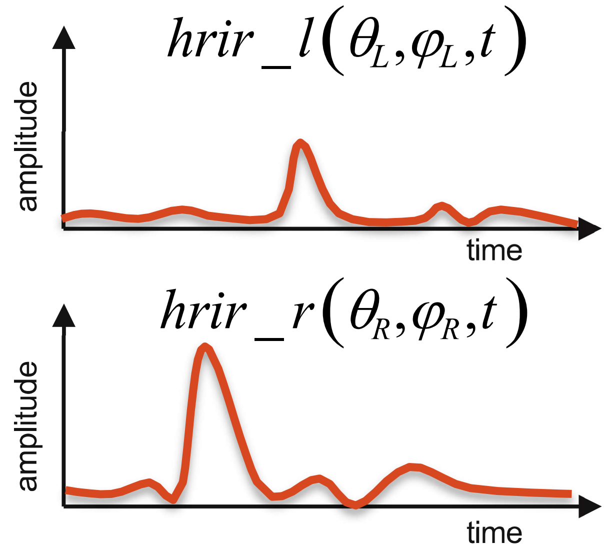


Head-related Impulse Response (HRIR)

applying the HRIR:

- given a mono sound source $s(t)$ and it's 3D position

1. compute (θ_L, φ_L) and (θ_R, φ_R) relative to center of listener
2. look up measured HRIR for left and right ear at these angles



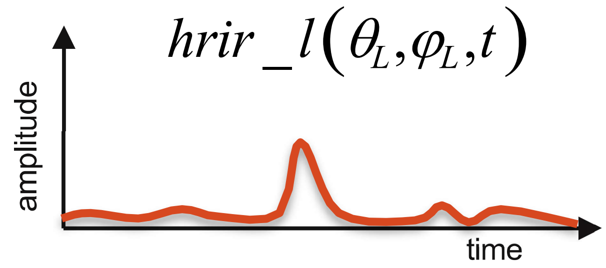
Head-related Impulse Response (HRIR)

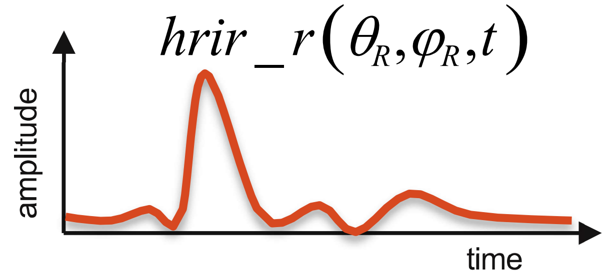
applying the HRIR:

- given a mono sound source $s(t)$ and it's 3D position

1. compute (θ_L, φ_L) and (θ_R, φ_R) relative to center of listener
2. look up measured HRIR for left and right ear at these angles
3. convolve signal with HRIRs to get response for each ear as

$$s_L(t) = hrir\_l(\theta_L, \varphi_L, t) * s(t)$$
$$s_R(t) = hrir\_r(\theta_R, \varphi_R, t) * s(t)$$



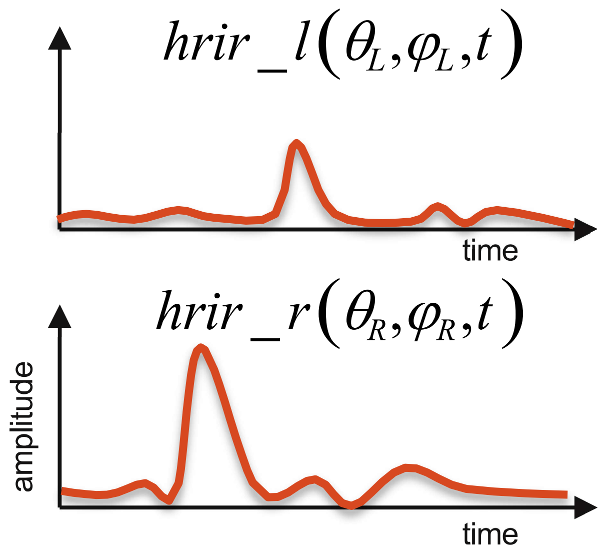


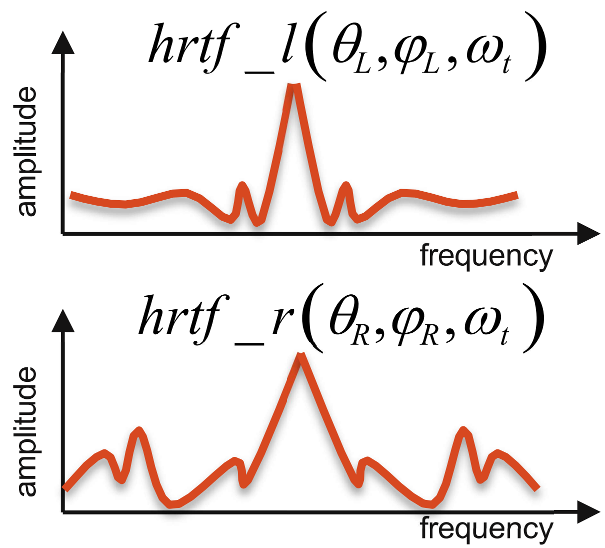
Head-related Transfer Function (HRTF)

- HRTF is Fourier transform of HRIR! (you'll find the term HRTF more often that HRIR)

$$s_L(t) = hrir\_l(\theta_L, \varphi_L, t) * s(t)$$
$$s_R(t) = hrir\_r(\theta_R, \varphi_R, t) * s(t)$$

$$s_L(t) = F^{-1}\left\{hrtf\_l(\theta_L, \varphi_L, \omega_t) \cdot F\{s(t)\}\right\}$$
$$s_R(t) = F^{-1}\left\{hrtf\_r(\theta_R, \varphi_R, \omega_t) \cdot F\{s(t)\}\right\}$$





Head-related Transfer Function (HRTF)

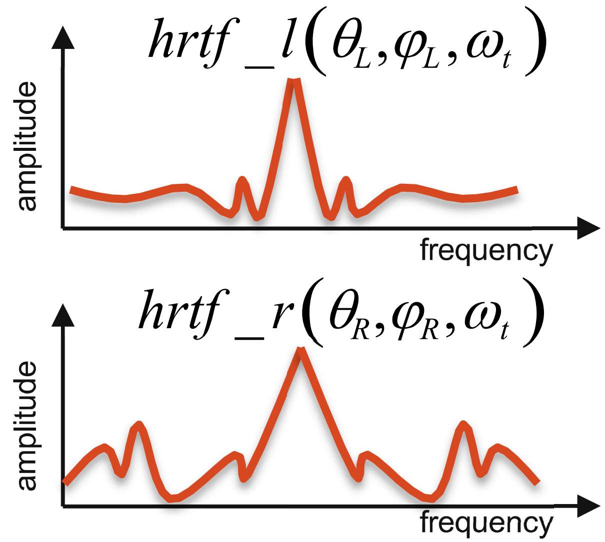
- HRTF is Fourier transform of HRIR! (you'll find the term HRTF more often that HRIR)

$$s_L(t) = hrir\_l(\theta_L, \varphi_L, t) * s(t)$$
$$s_R(t) = hrir\_r(\theta_R, \varphi_R, t) * s(t)$$

$$\Rightarrow$$

convolution theorem

$$s_L(t) = F^{-1}\left\{hrtf\_l(\theta_L, \varphi_L, \omega_t) \cdot F\{s(t)\}\right\}$$
$$s_R(t) = F^{-1}\left\{hrtf\_r(\theta_R, \varphi_R, \omega_t) \cdot F\{s(t)\}\right\}$$



Head-related Transfer Function (HRTF)

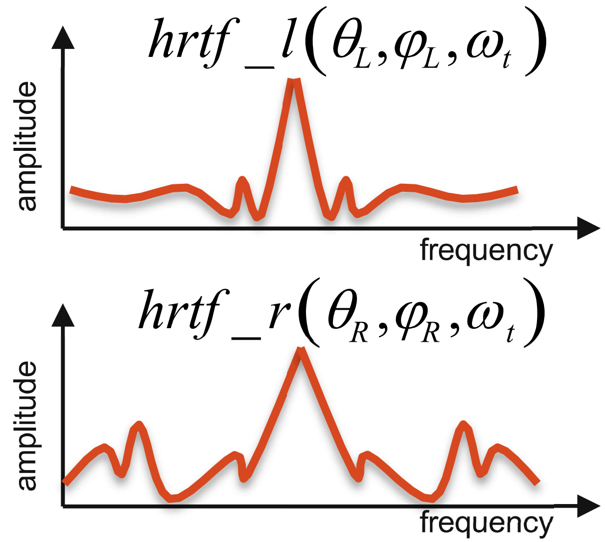
- HRTF is Fourier transform of HRIR! (you'll find the term HRTF more often that HRIR)

$$s_L(t) = hrir\_l(\theta_L, \varphi_L, t) * s(t)$$
$$s_R(t) = hrir\_r(\theta_R, \varphi_R, t) * s(t)$$

$$\Rightarrow$$

$$s_L(t) = F^{-1}\left\{hrtf\_l(\theta_L, \varphi_L, \omega_t) \cdot F\{s(t)\}\right\}$$
$$s_R(t) = F^{-1}\left\{hrtf\_r(\theta_R, \varphi_R, \omega_t) \cdot F\{s(t)\}\right\}$$

- properties of HRTF:
    - complex-valued
    - symmetric (because HRIR is real-valued)



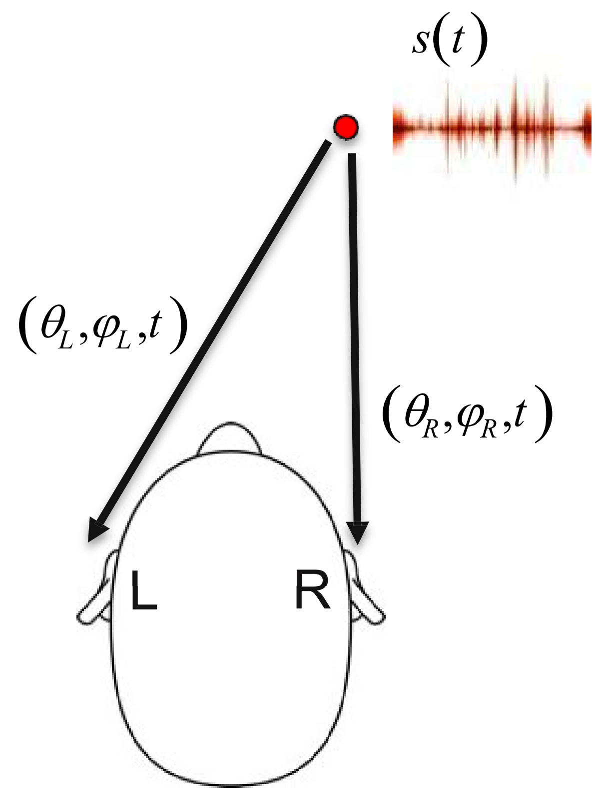
Head-related Transfer Function (HRTF)

$$s_L(t) = F^{-1}\left\{hrtf\_l(\theta_L, \varphi_L, \omega_t) \cdot F\{s(t)\}\right\}$$

$$s_R(t) = F^{-1}\left\{hrtf\_r(\theta_R, \varphi_R, \omega_t) \cdot F\{s(t)\}\right\}$$

Spatial Sound of 1 Point Sound Source

- given $s(t)$ and 3D position, follow instructions from last slides by convolving Fourier transform of s with HRTFs for each each

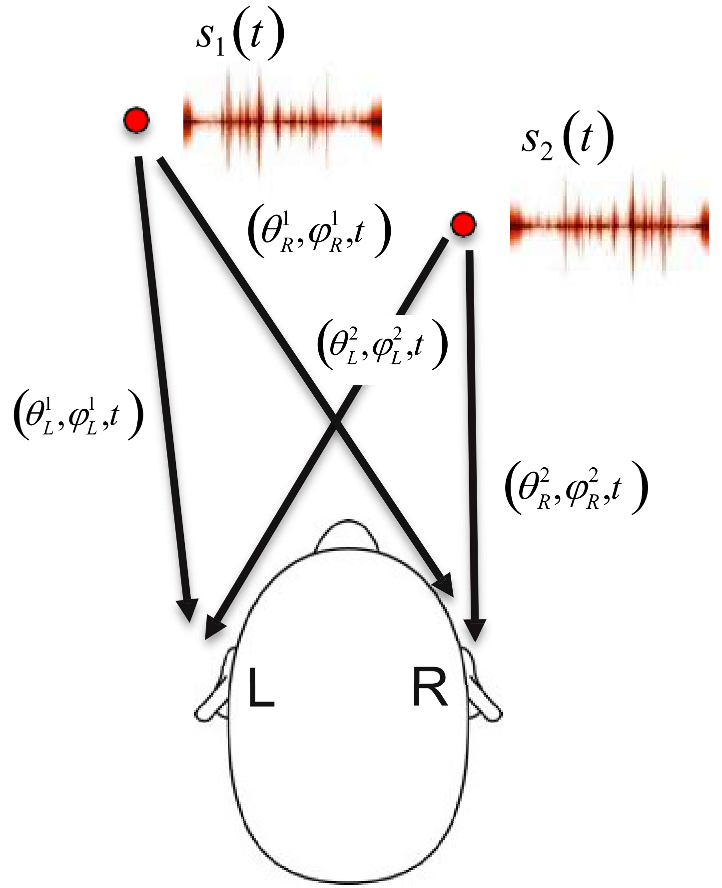


Spatial Sound of N Point Sound Sources

- superposition principle holds, so just sum the contributions of each

$$s_L(t) = \sum_{i=1}^{N} F^{-1}\left\{hrtf\_l\left(\theta_L^i, \varphi_L^i, \omega_t\right) \cdot F\left\{s_i(t)\right\}\right\}$$

$$s_R(t) = \sum_{i=1}^{N} F^{-1}\left\{hrtf\_r\left(\theta_R^i, \varphi_R^i, \omega_t\right) \cdot F\left\{s_i(t)\right\}\right\}$$



Surround Sound

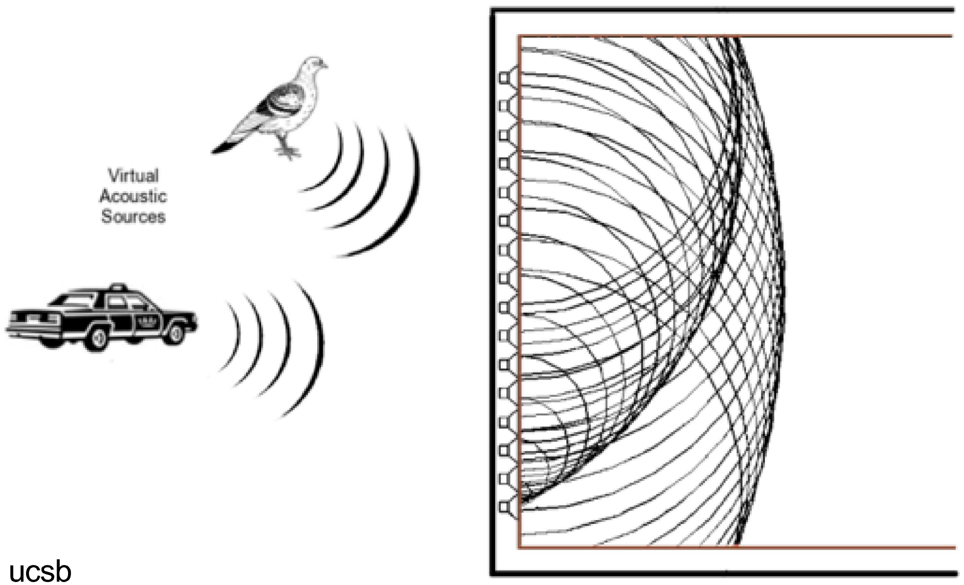
- approximate continuous wave field with discrete set of speakers

- most common:
  5.1 surround sound =
  5 (channels) . 1 (bass)

➔ 6 channels total

Surround Sound

- approximate continuous wave field with discrete set of speakers
- can also use more speakers for “wave field synthesis” (i.e. audio hologram)



http://spatialaudio.net/

Surround Sound

- approximate continuous wave field with discrete set of speakers
- can also use more speakers for “wave field synthesis” (i.e. audio hologram)
- for wave field synthesis, phase of speakers needs to be synchronized, i.e. a phased array!

Surround Sound & HRTF

- for all speaker-based (surround) sound, we don’t need an HRTF because the ears of the listener will apply them!
- speaker setup usually needs to be calibrated

Spatial Audio for VR

- VR/AR requires us to re-think audio, especially spatial audio!
- could use 5.1 surround sound and set up “virtual speakers” in the virtual environment – can use existing content, but not super easy to capture new content; also doesn’t capture directionality from above/below

Spatial Audio for VR

Two primary approaches:

1. Real-time sound engine
    - render 3D sound sources via HRTF in real-time, just as discussed in the previous slides
    - used for games and synthetic virtual environments
    - a lot of libraries available: FMOD, OpenAL, …

Spatial Audio for VR

Two primary approaches:

2. Spatial sound recorded from real environments
    - most widely used format now: ambisonics
    - simple microphones exist
    - relatively easy mathematical model
    - only need 4 channels for starters
    - used in YouTube VR and many other platforms

Ambisonics

- idea: represent sound incident at a point (i.e. the listener) with some directional information

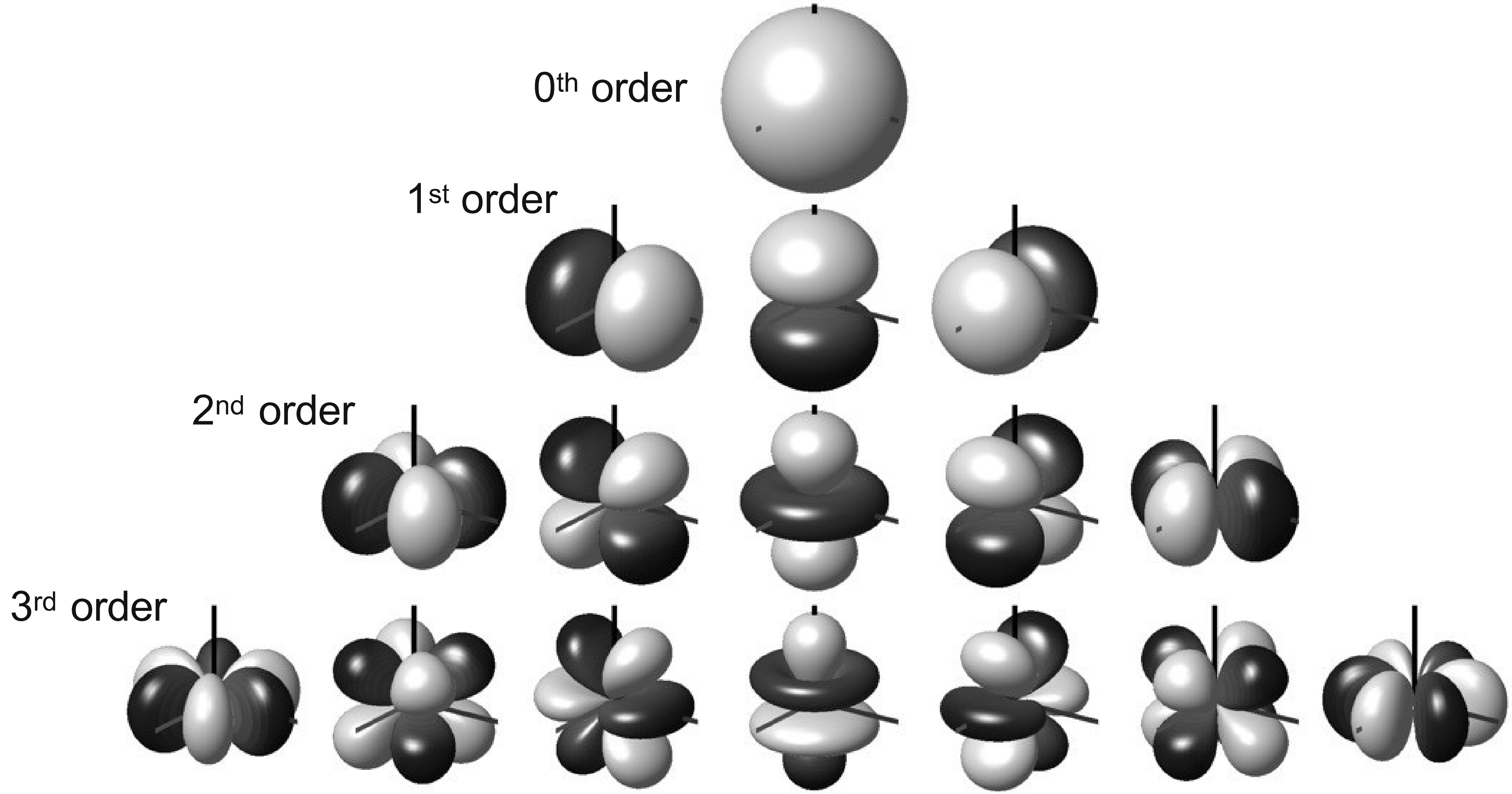
- using all angles θ,φ is impractical – need too many sound channels (one for each direction)

- some lower-frequency (in direction) components may be sufficient ➔ directional basis representation to the rescue!

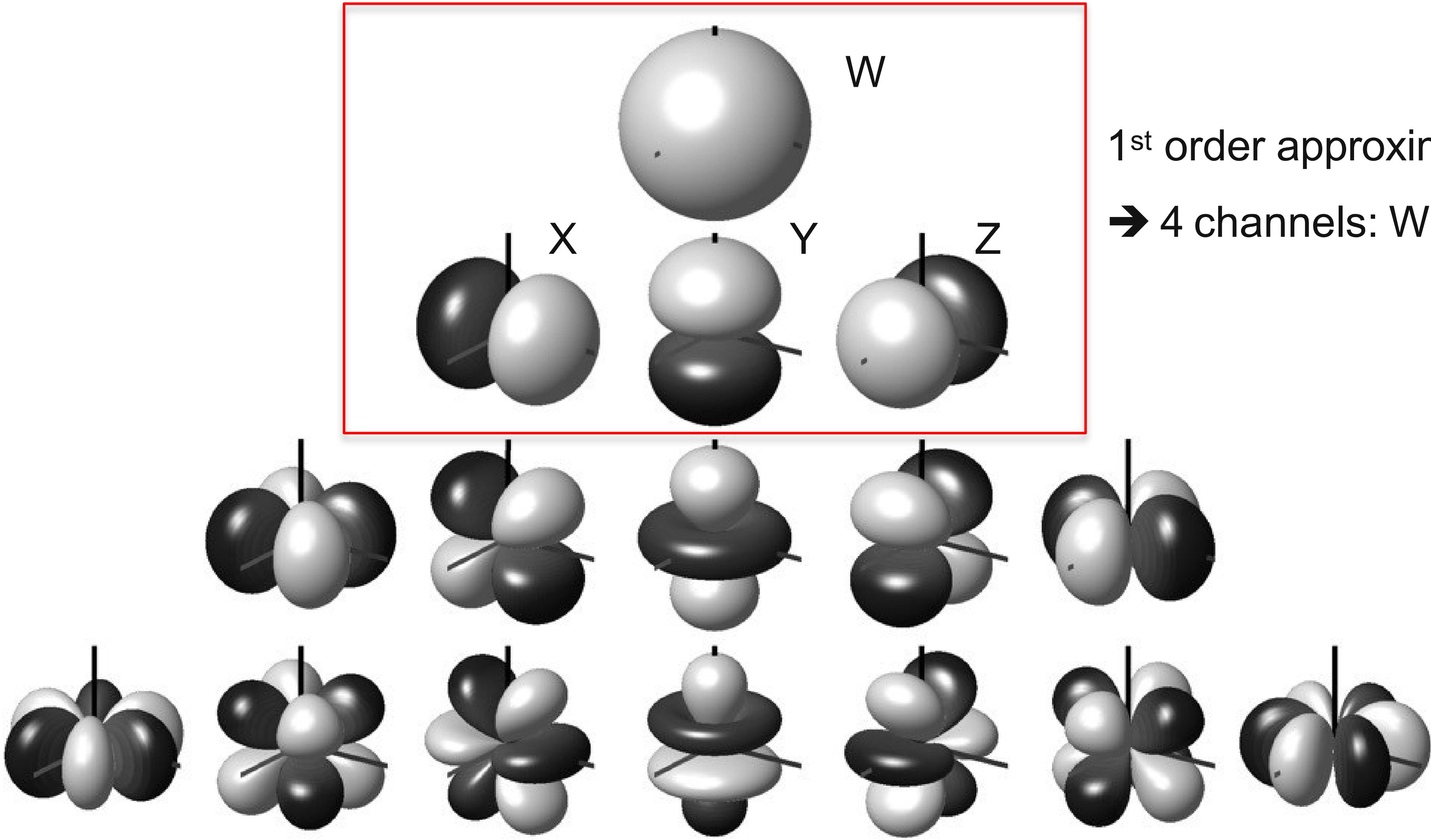
Ambisonics – Spherical Harmonics

- use spherical harmonics!
- orthogonal basis functions on a sphere, i.e. full-sphere surround sound
- think Fourier transform acting on the directions of a sphere

Ambisonics – Spherical Harmonics



Ambisonics – Spherical Harmonics



1st order approximation

➔ 4 channels: W, X, Y, Z

Ambisonics – Spherical Harmonics

- can easily convert a point sound source to the 4-channel ambisonics representation
- given azimuth and elevation θ, φ , compute W,X,Y,Z as

$$W = S \cdot \frac{1}{\sqrt{2}} \quad \longleftarrow \quad \text{omnidirectional component (angle-independent)}$$

$$X = S \cdot \cos\theta \cos\varphi \quad \longleftarrow \quad \text{“stereo in x”}$$

$$Y = S \cdot \sin\theta \cos\varphi \quad \longleftarrow \quad \text{“stereo in y”}$$

$$Z = S \cdot \sin\varphi \quad \longleftarrow \quad \text{“stereo in z”}$$

Ambisonics – Spherical Harmonics

- can also record 4-channel ambisonics via special microphone
- same format supported by YouTube VR and other platforms

http://www.oktava-shop.com/

Ambisonics – Spherical Harmonics

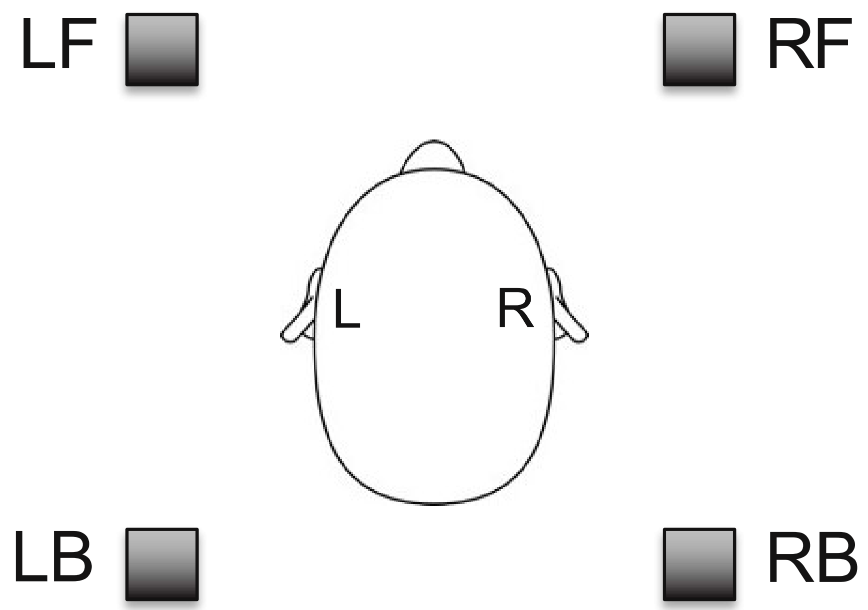
- easiest way to render ambisonics: convert W,X,Y,Z channels into 4 virtual speaker positions
- for a regularly-spaced square setup, this results in

$$LF = (2W + X + Y)\sqrt{8}$$

$$LB = (2W - X + Y)\sqrt{8}$$

$$RF = (2W + X - Y)\sqrt{8}$$

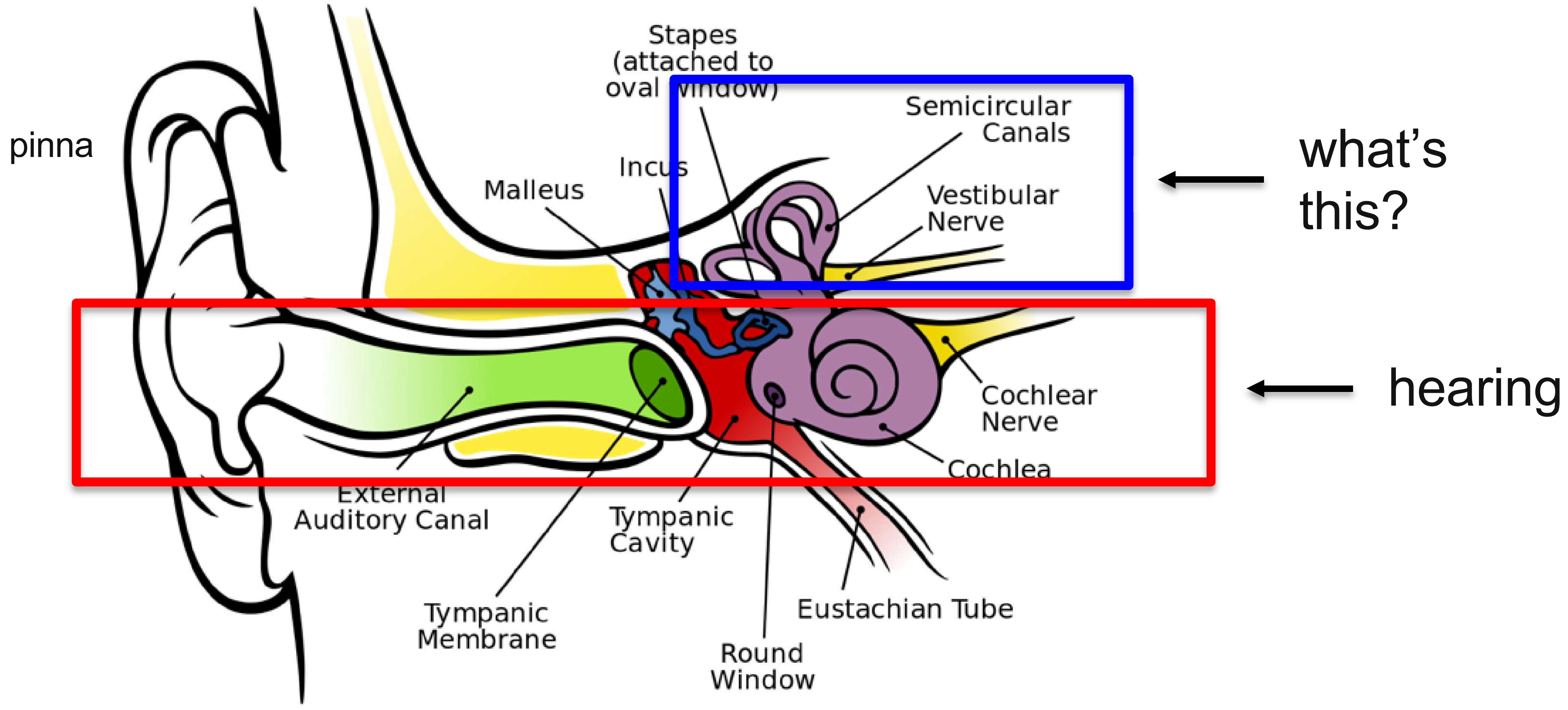
$$RB = (2W - X - Y)\sqrt{8}$$



The Vestibular System

or “What else is happening in the inner ear?”

The Inner Ear

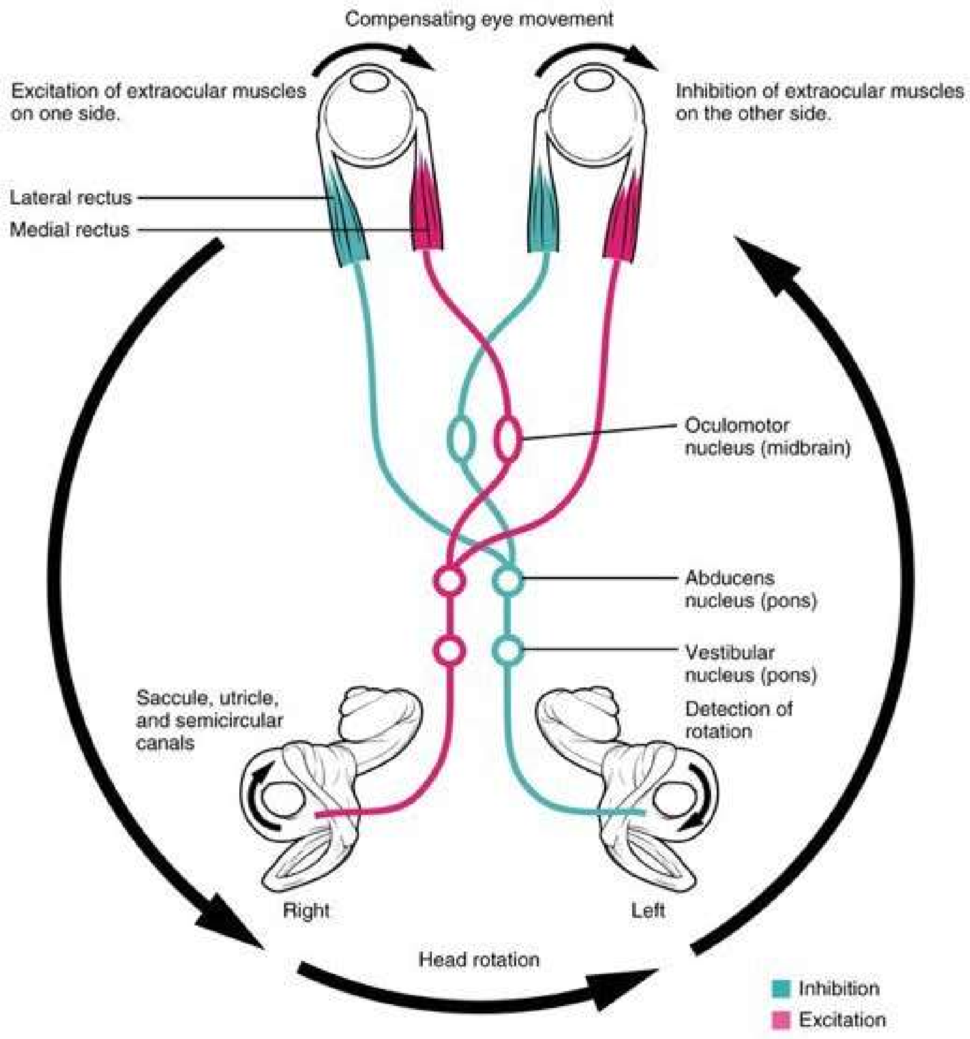


wikipedia

Brief Overview of the Vestibular System

- provides sense of balance & gravity
- like IMUs – one in each ear!
- in each ear, sense linear (3 dof from otolithic organs) and angular (3 dof from 3 semicircular canals) acceleration via hair cells

Vestibulo-Ocular Reflex (VOR)



- vestibular system and ocular system are directly coupled in a feedback system
- enables low-latency “optical image stabilization” of the visual system with head motion

Motion Sickness

3 types of motion sickness (all related to visual-vestibular conflict theory):

1. Motion sickness caused by motion that is felt but not seen
2. Motion sickness caused by motion that is seen but not felt
3. Motion sickness caused when both systems detect motion but they do not correspond.

Motion Sickness

3 types of motion sickness (all related to visual-vestibular conflict theory):

1. Motion sickness caused by motion that is felt but not seen
2. Motion sickness caused by motion that is seen but not felt
3. Motion sickness caused when both systems detect motion but they do not correspond.

Example: car and sea sickness

Motion Sickness

3 types of motion sickness (all related to visual-vestibular conflict theory):

1. Motion sickness caused by motion that is felt but not seen
2. Motion sickness caused by motion that is seen but not felt
3. Motion sickness caused when both systems detect motion but they do not correspond.

Example: VR sickness or visually-induced motion sickness (VIMS)

Motion Sickness

3 types of motion sickness (all related to visual-vestibular conflict theory):

1. Motion sickness caused by motion that is felt but not seen
2. Motion sickness caused by motion that is seen but not felt
3. Motion sickness caused when both systems detect motion but they do not correspond.

Example: motion in low gravity

References and Further Reading

Panoramic Imaging and VR

- M. Brown, D. Lowe “Automatic Panoramic Image Stitching using Invariant Features”, IJCV 2007
- autostitch: http://matthewalunbrown.com/autostitch/autostitch.html
- S. Peleg, M. Ben-Ezra, Y. Pritch “Omnistereo: Panoramic Stereo Imaging” IEEE PAMI 2001
- Konrad et al. “SpinVR: Towards Live Streaming VR Video”, ACM SIGGRAPH Asia 2017

References and Further Reading - Spatial Sound

- Google's take on spatial audio: https://developers.google.com/vr/concepts/spatial-audio

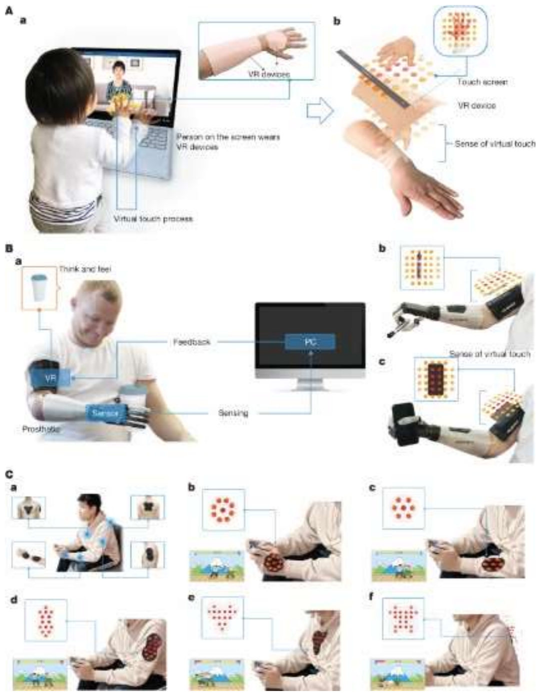
HRTF:

- Algazi, Duda, Thompson, Avendado "The CIPIC HRTF Database", Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics
- download CIPIC HRTF database here: http://interface.cipic.ucdavis.edu/sound/hrtf.html

Resources by Google:

- https://github.com/GoogleChrome/omnitone
- https://developers.google.com/vr/concepts/spatial-audio
- https://opensource.googleblog.com/2016/07/omnitone-spatial-audio-on-web.html
- http://googlechrome.github.io/omnitone/#home
- https://github.com/google/spatial-media/

existing applications of haptics

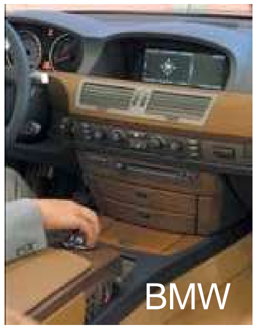


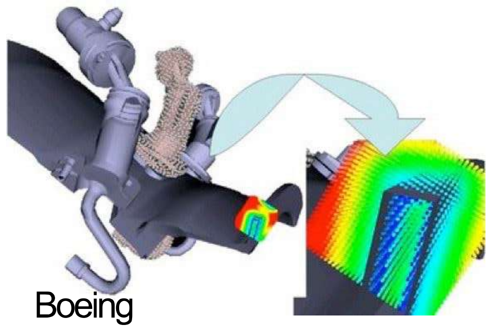
entertainment

education

health

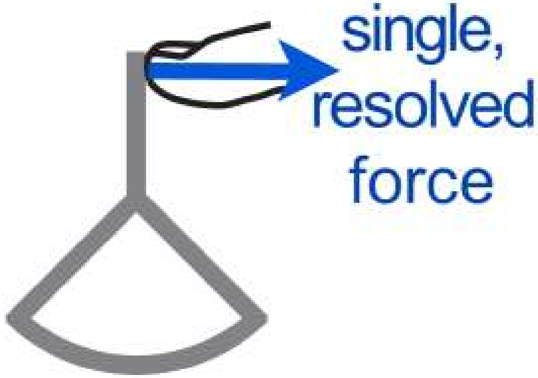
human-computer interfaces





kinesthetic vs. tactile haptic devices

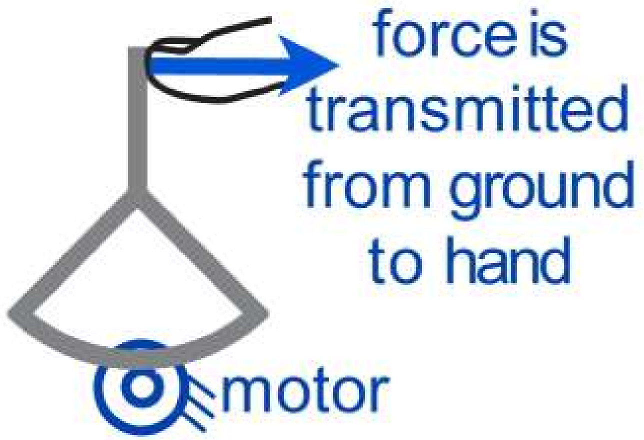
Kinesthetichapticdevices display forces or motions through a tool



Tactile haptic devices stimulate the skin

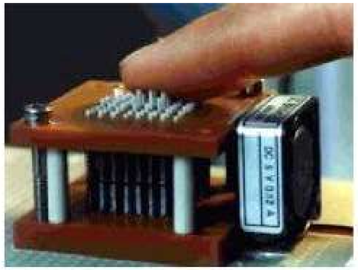
kinesthetic vs. tactile haptic devices

Kinesthetic haptic devices are usually **grounded**



Tactile haptic devices can more easily be **wearable**

Tactile Devices



Stimulate skin to create contact sensations

Hybrid Devices

Attempt to combine tactile and kinesthetic feedback

Kinesthetic Devices

Apply forces to guide or inhibit body movement

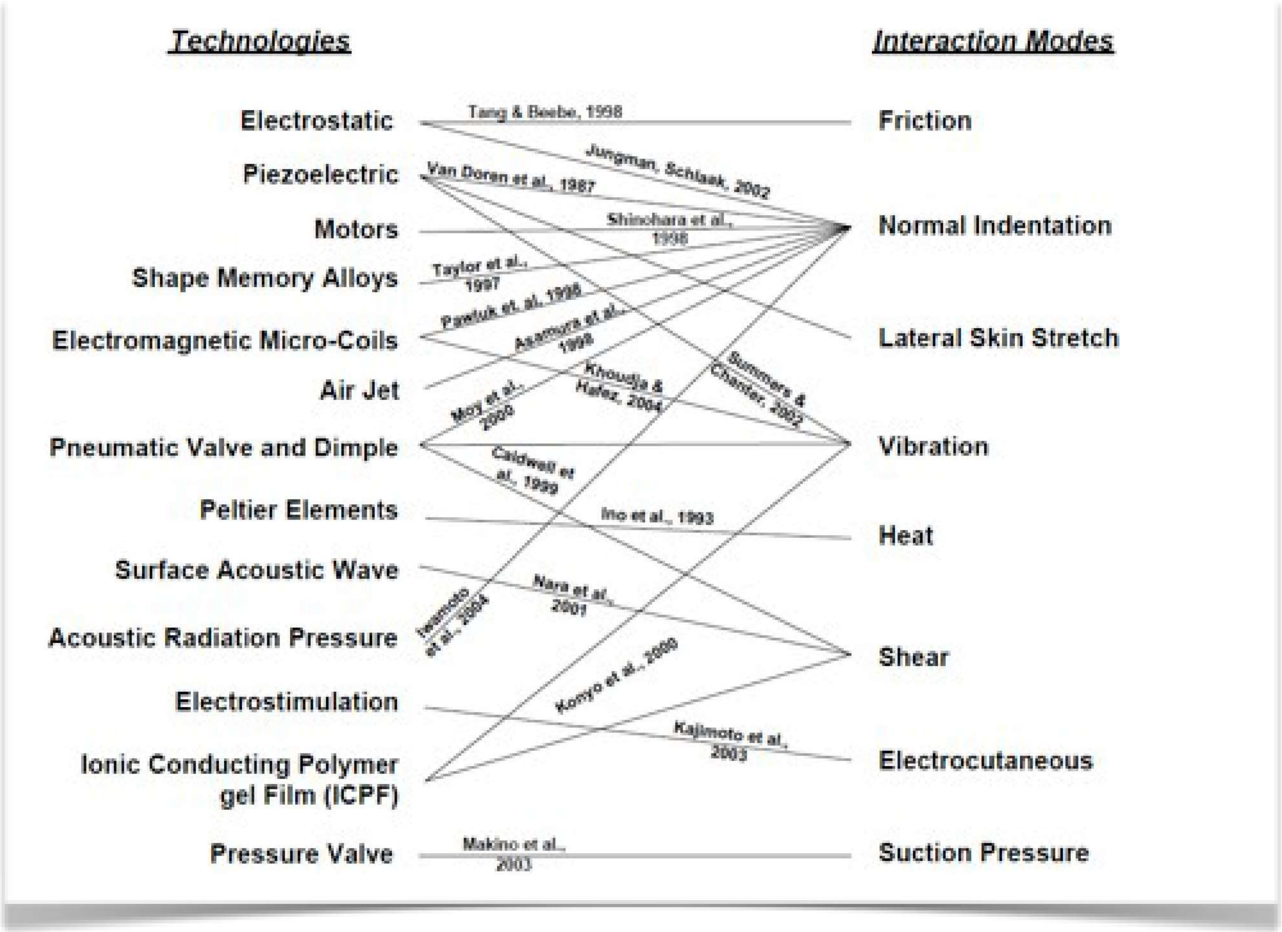
K. Kuchenbecker

tactile
(cutaneous)
device basics

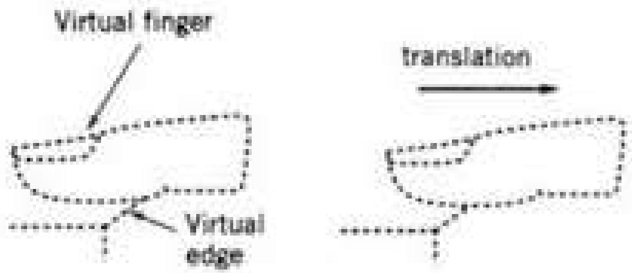
tactile feedback

- goal is to stimulate the **skin** in a programmable manner to create a desired set of sensations
- *sometimes* **distributed** tactile feedback is provided
- tactile feedback is generated by a **tactile device**, sometimes called a **tactile display**
- can aim to recreate real sensations, create novel ones, or communicate information

technologies and interaction modes

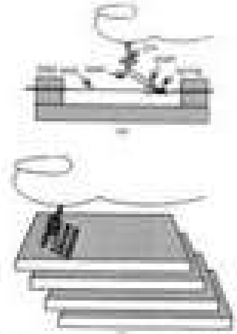


Jerome Pasquero, Survey on Communication through Touch, Technical Report: TR-CIM 06.04, 2006



Burdea & Coiffet (1994)

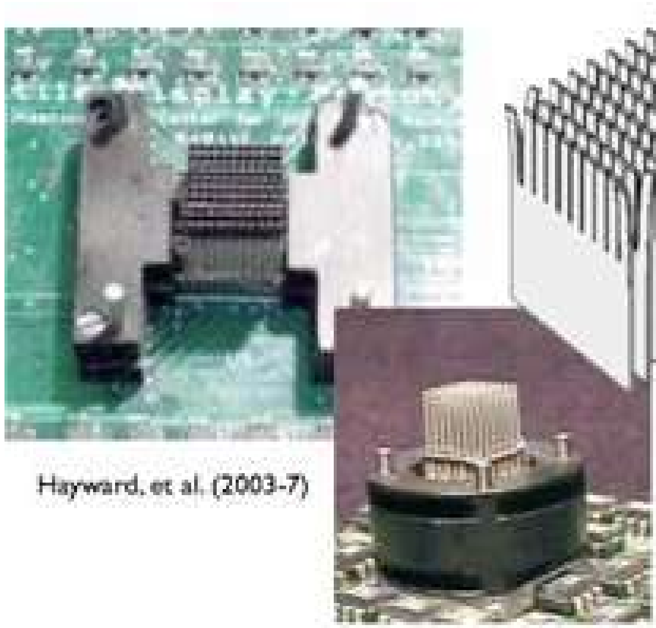
Wagner & Howe (2002)



Kontarinis, et al. (1995)

Kaczmarek, et al. (1995)

Russell


Hayward, et al. (2003-7)

Glassmire thesis (2006)

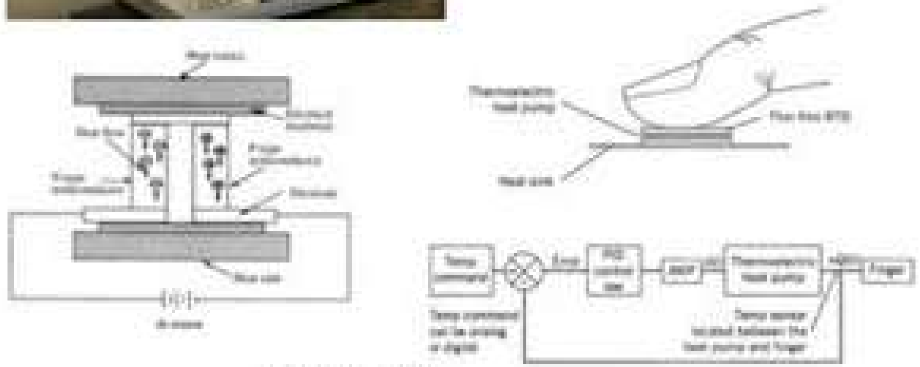


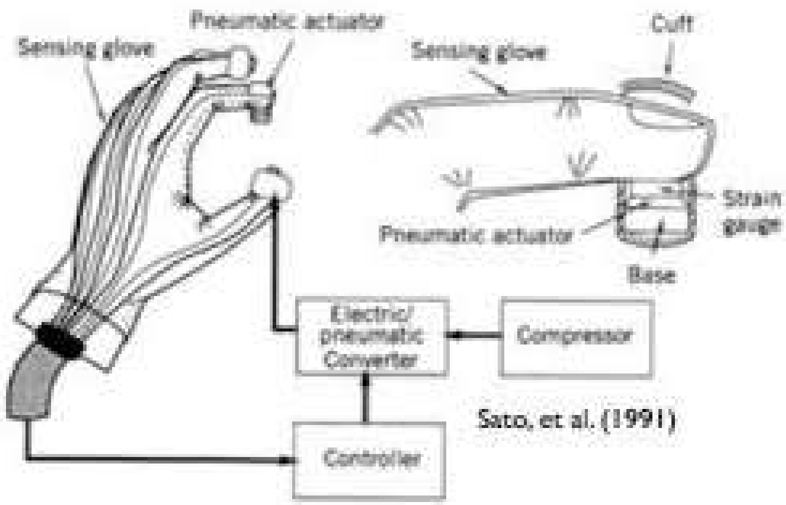
Samsung, et al.

Winfield & Colgate (2007)



Salada, et al. (2002-5)

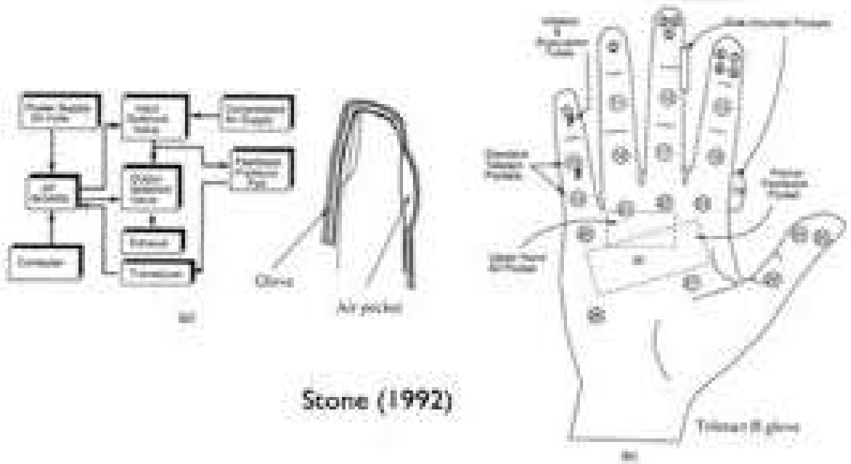

Zerkus (1993)



Sato, et al. (1991)

Immersion CyberTouch Glove

Burdea (1996)



Stone (1992)

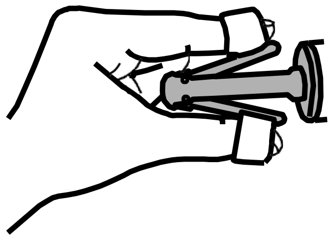
kinesthetic
(force-feedback)
device basics

typical kinesthetic device configurations

manipulandum

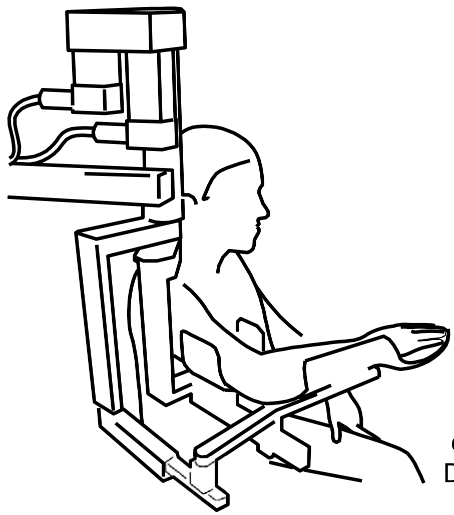
drawing by Jorge Cham

grasp



drawing by Tricia Gibo

exoskeleton



drawing by David Grow

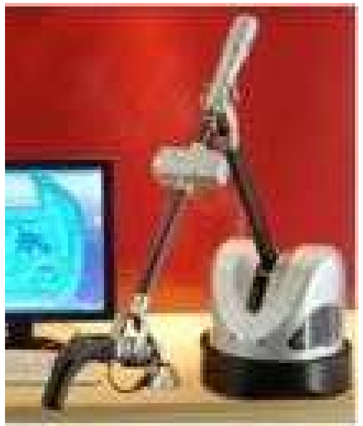
manipulandums (expensive)

Omega
from Force Dimension

delta configuration
3 degrees of freedom

Phantom Premium 1.5
from SensAble/Geomagic

5-bar + rotation
3 degrees of freedom



Virtuose
from Haption

additional “wrist”
6 degrees of freedom

all images from Wikimedia Commons

manipulandums (cheaper)

Falcon
from Novint

delta configuration
3 degrees of freedom

image from Wikimedia Commons

Phantom Omni/Touch
from SensAble/Geomagic

5-bar + rotation
3 degrees of freedom

photographed by Akiko Nabeshima

Sidewinder
from Microsoft

spherical mechanism
2 degrees of freedom

image from Wikimedia Commons

Grip/grasp

Custom haptic gripper for Phantom Premium

© 2007 IEEE. Reprinted, with permission, from L. N. Verner and A. M. Okamura.. Effects of Translational and Gripping Force Feedback are Decoupled in a 4-Degree-of-Freedom Telemanipulator, World Haptics Conference,, pp. 286-291, 2007

Single-finger Cybergrasp from Cyberglove Systems

photograph courtesy
Stanford Center for Design Research

da Vinci Surgical System from Intuitive Surgical, Inc. (no programmable force feedback on gripper)

photographed by Akiko Nabeshima

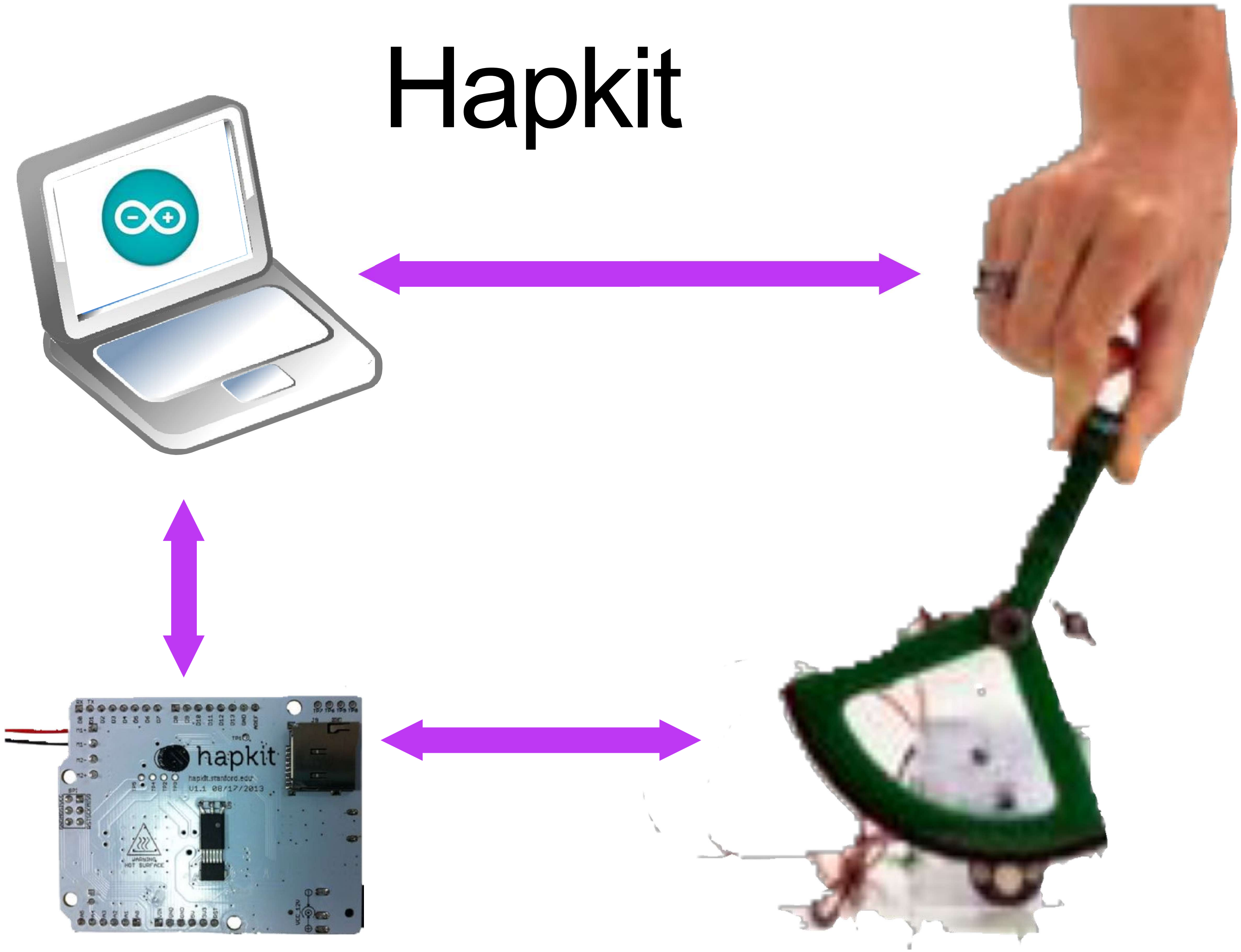
Exoskeletons

KINARM Exoskeleton
from BKIN Technologies

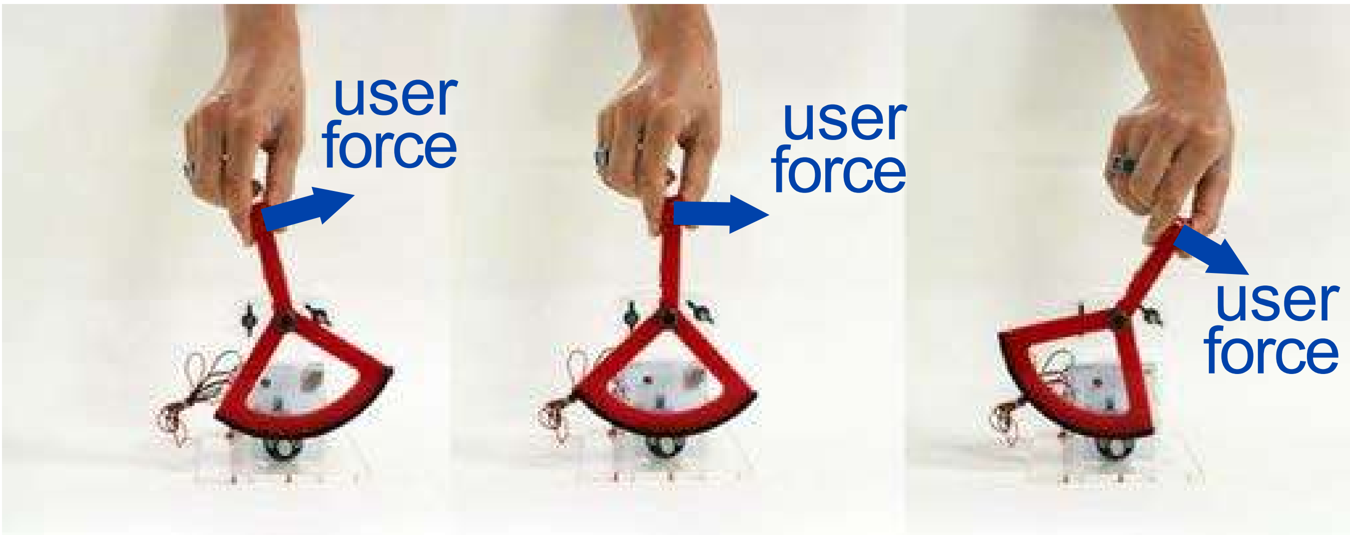
Harvard DARPA

images from Wikimedia Commons

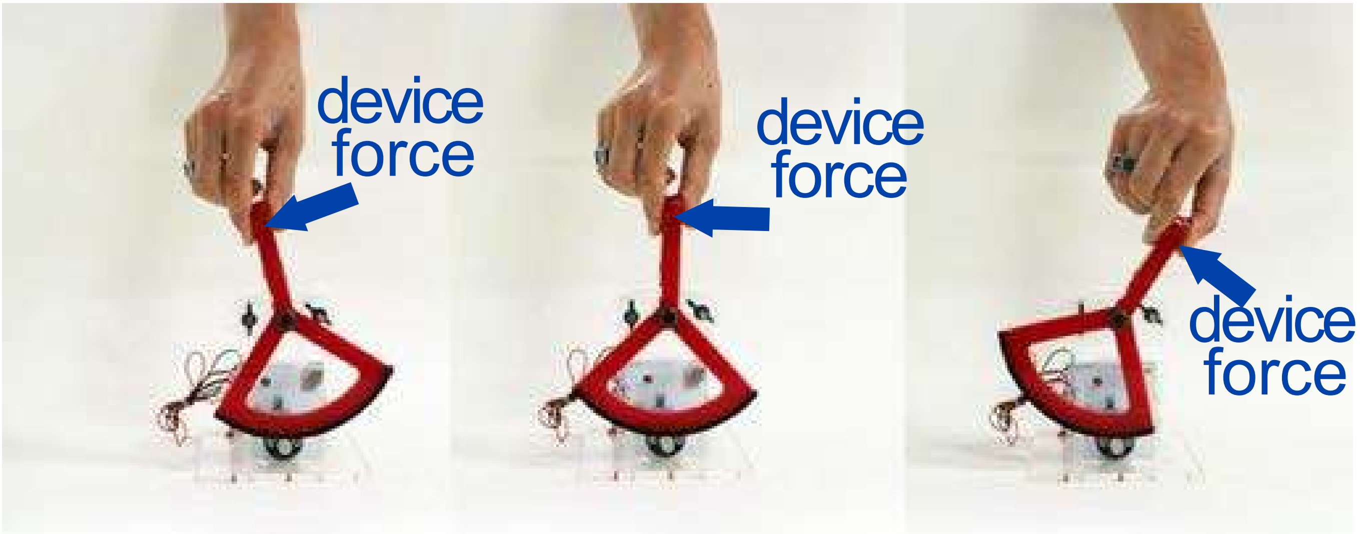
Hapkit



Hapkit



Hapkit



Trends driving haptics

- Networking - constant connectivity
- Ubiquity of computing devices - beyond sparse visual real estate
- Multitasking - doing more things at once may benefit from multiple channels of communcation
- Virtualization – fostering presence

Rendering friction

(in one degree of freedom)

surface properties

- typical haptic display general shape very well, but don’t feel “realistic”
  - add surface properties to increase realism
- how you do it depends on
  - the surface model
  - complexity of the surface property
  - frequency response of your haptic device

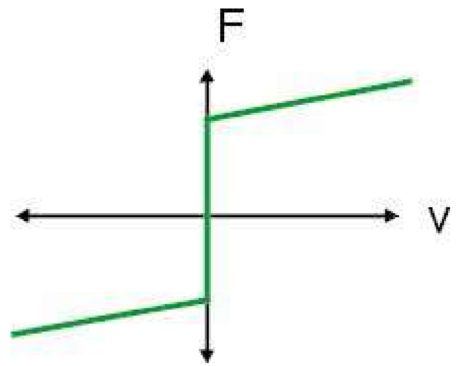
damping for virtual walls

- a pure springforce for a wall may seem to “active”
- add a dissipative term, where b is the damping coefficient
- only damps when going into the wall
- this can also create vibrations upon wall impact

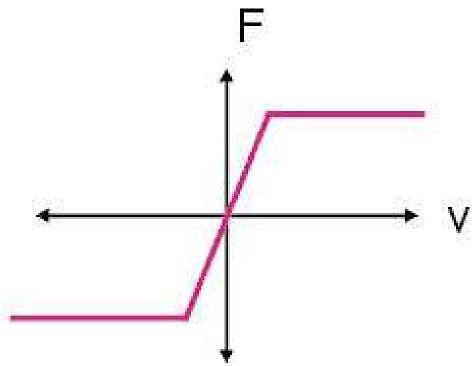
$$F = \begin{cases} k\Delta x + b\dot{x} & \text{for } \dot{x} > 0 \\ k\Delta x & \text{for } \dot{x} < 0 \end{cases}$$

“frictional” damping

- surfaces can feel unnaturallyslippery
- friction would help, but it is difficult to implement
- you can add damping to motion parallel to surface



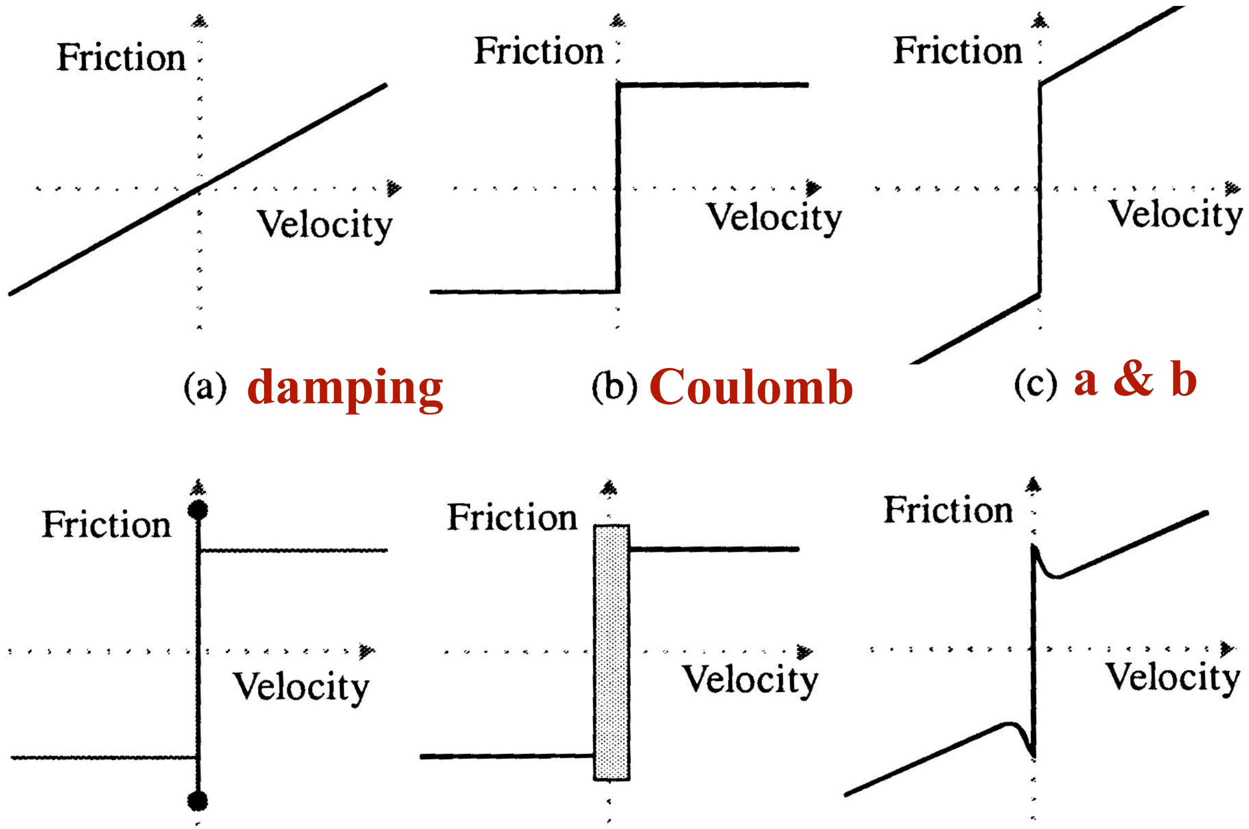
friction



damping

friction display

difficult to render because it is non-linear



(d) stiction (e) static (f) Stribeck effect

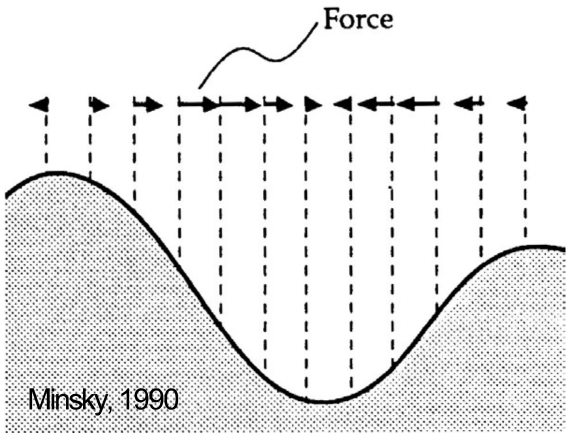
Richard & Cutkosky

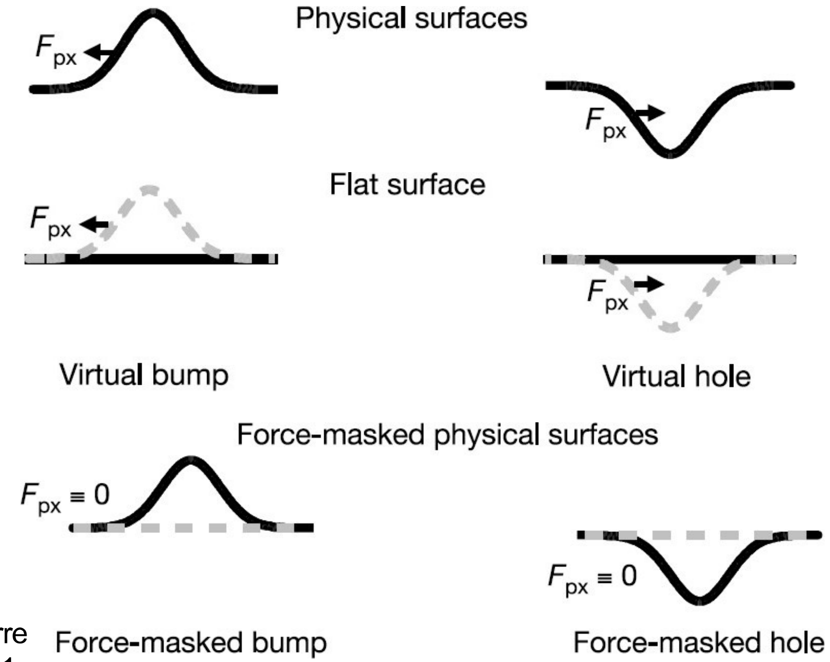
Rendering bumps and textures

(in one degree of freedom)

bumps and valleys

- as a user moves "up" a bump, motion is opposed.
- done in 2D, spring force proportional to height of bump
- force information can overcome geometry!





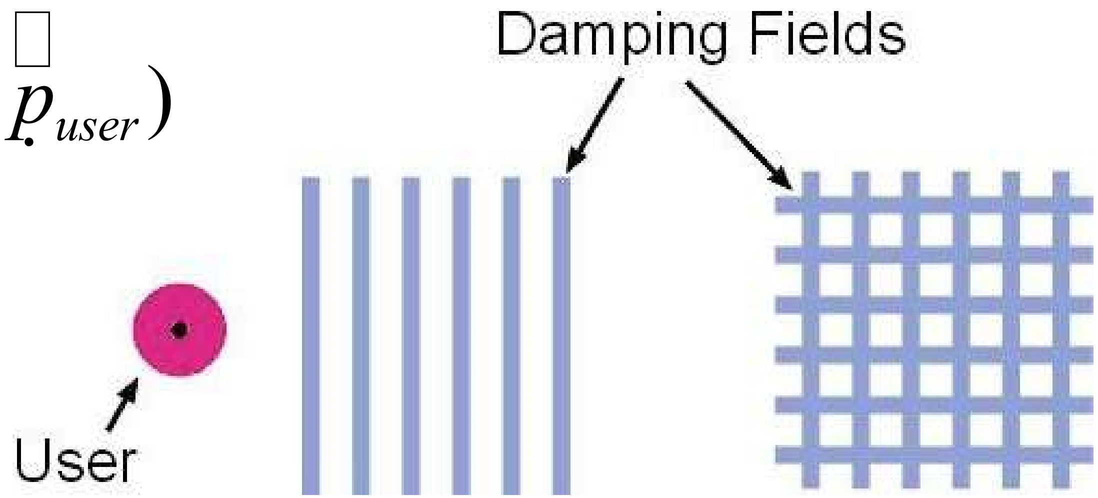
Robles-De-La-Torre & Hayward, 2001

damping textures

if p_{user} is inside a damping area

$$F = b\vec{v}_{user}$$

$$(\vec{v}_{user} = \dot{\vec{p}}_{user})$$



note that vibrations occur due to discontinuity in force simulating and rendering

dynamic objects

(in one degree of freedom)

dynamic simulation of “rigid” bodies

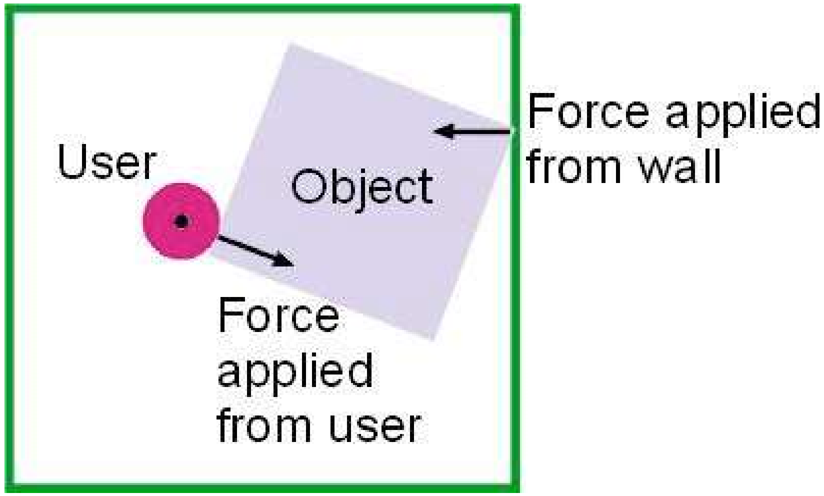
- assumptions
  - you have an impedance-type force-feedback display
  - you are using a linear stiffness model of the surface
- basic approach:
  1. save the state of the moving object
  2. sum the forces on the object
  3. calculate the new state

object state

- the “state” of an object is used to describe its current condition
- made up of variables that will change with time, such as
  - position
  - velocity
  - acceleration
  - rarely: other parameters such as mass, shape, etc. that might be changed by dynamic interaction

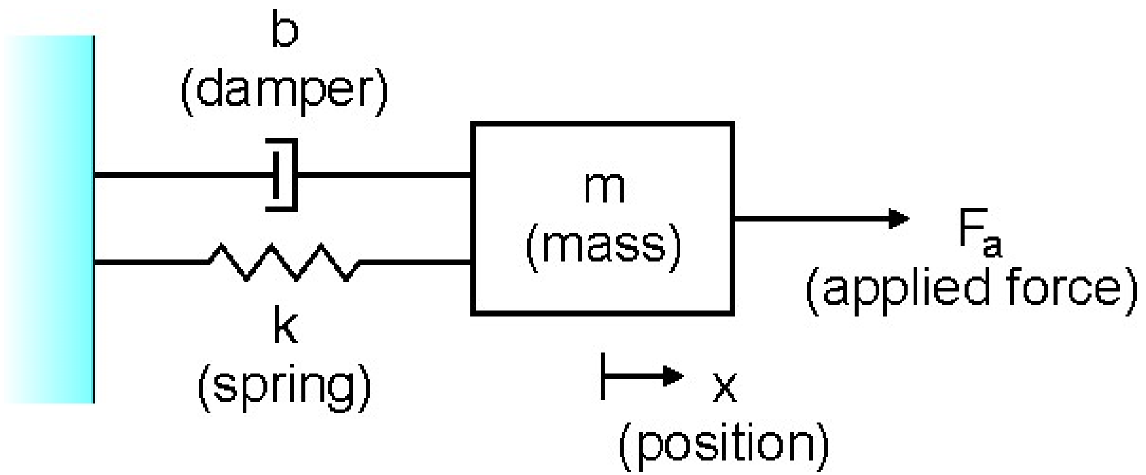
calculating forces on an object

- for forces from the user's hand pushing, this is equal and opposite to the force fed back to the user
- for forces from other objects in the VE, use the same idea: force is proportional to penetration



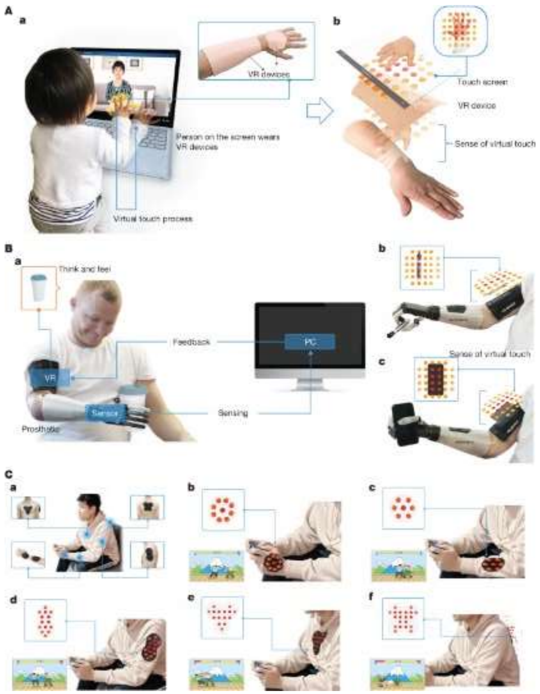
second order dynamic systems

mass-spring-damper



x = 0 at equilibrium

Haptics

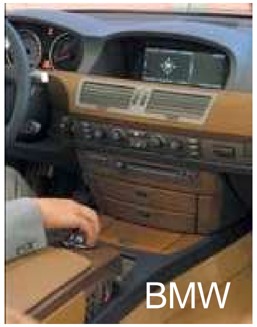


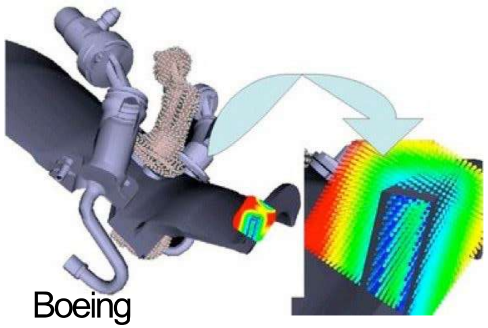
entertainment

education

health

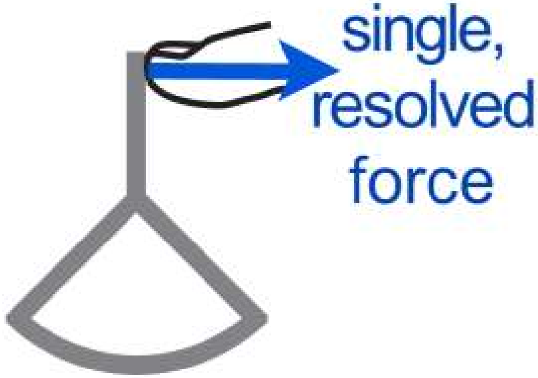
human-computer interfaces





kinesthetic vs. tactile haptic devices

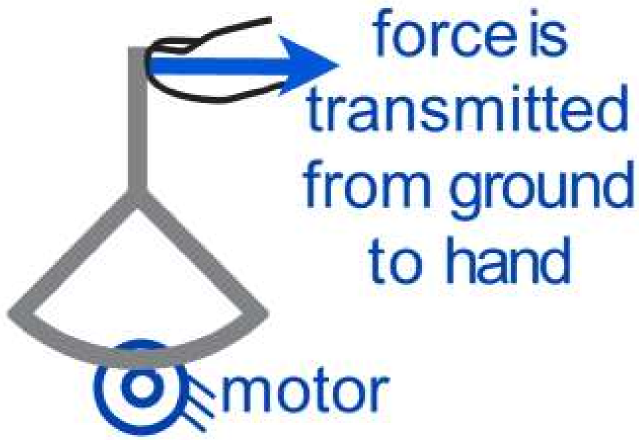
Kinesthetichapticdevices display forces or motions through a tool



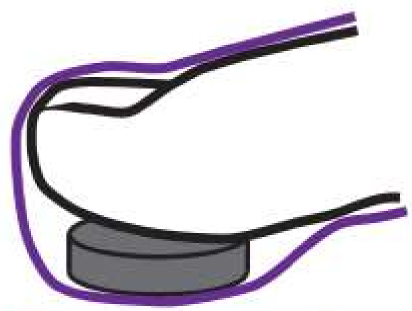
Tactile haptic devices stimulate the skin

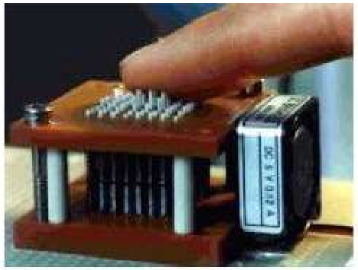
kinesthetic vs. tactile haptic devices

Kinesthetic haptic devices are usually **grounded**



Tactile haptic devices can more easily be **wearable**





Tactile Devices

Stimulate skin to create contact sensations

Hybrid Devices

Attempt to combine tactile and kinesthetic feedback

Kinesthetic Devices

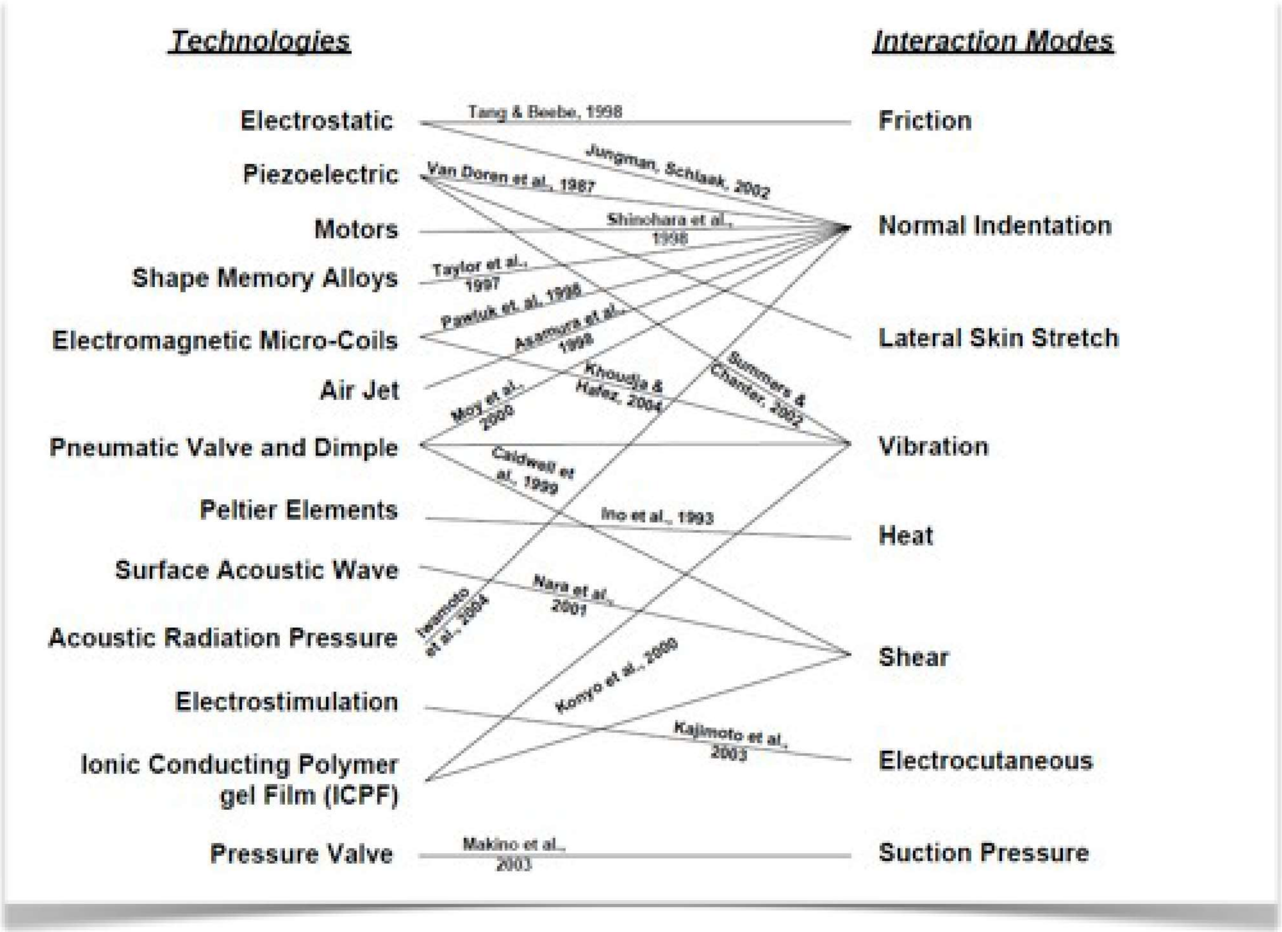
Apply forces to guide or inhibit body movement

tactile (cutaneous) device basics

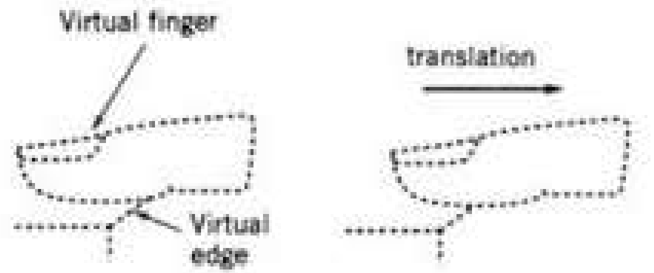
tactile feedback

- goal is to stimulate the **skin** in a programmable manner to create a desired set of sensations
- *sometimes* **distributed** tactile feedback is provided
- tactile feedback is generated by a **tactile device**, sometimes called a **tactile display**
- can aim to recreate real sensations, create novel ones, or communicate information

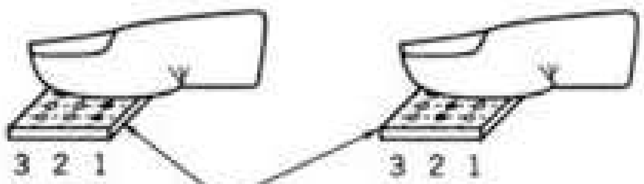
technologies and interaction modes



Jerome Pasquero, Survey on Communication through Touch, Technical Report: TR-CIM 06.04, 2006

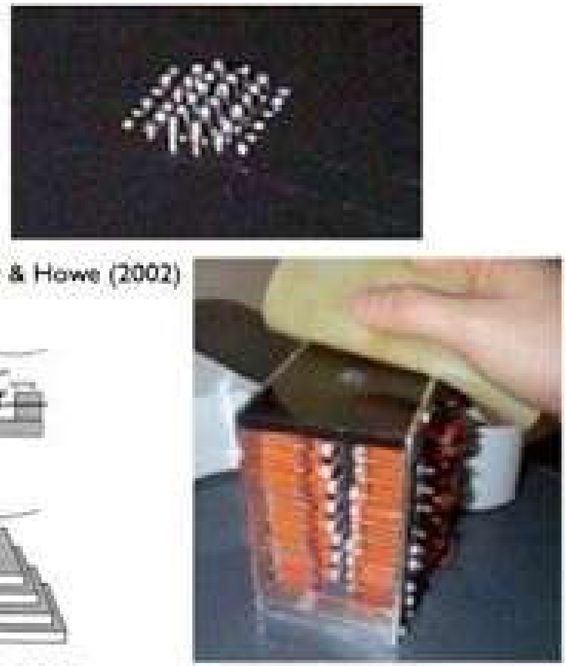


Burdea & Coiffet (1994)

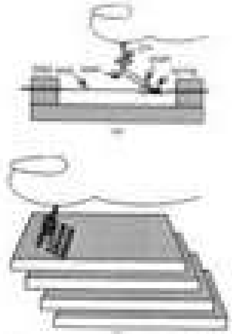


Tactile feedback array

Kaczmarek, et al. (1995)

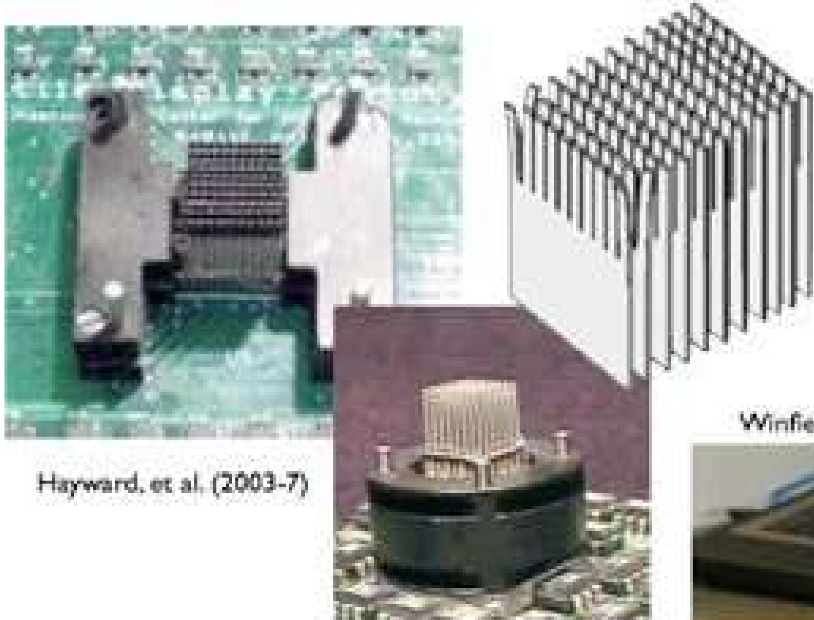


Wagner & Howe (2002)



Kontarinis, et al. (1995)

Russell

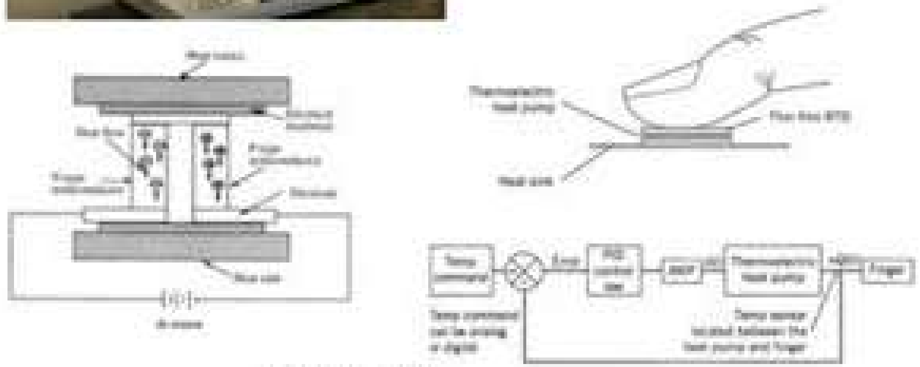

Hayward, et al. (2003-7)

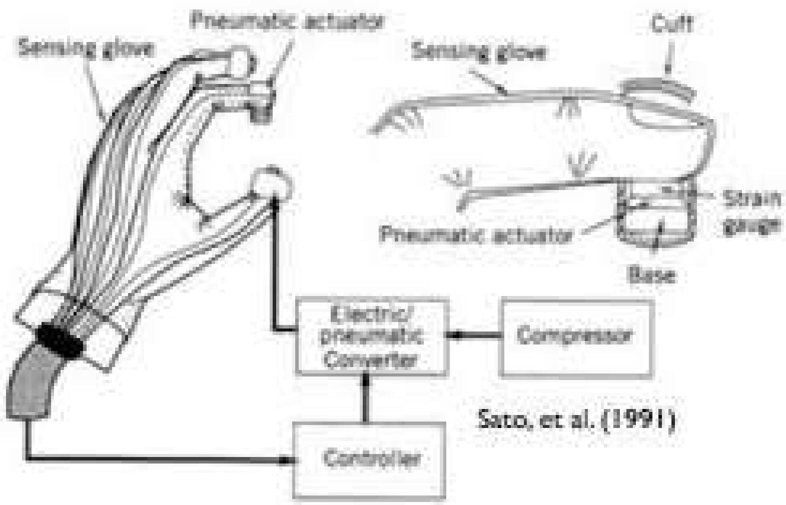

Glassmire thesis (2006)


Winfield & Colgate (2007)


Samsung, et al.


Salada, et al. (2002-5)

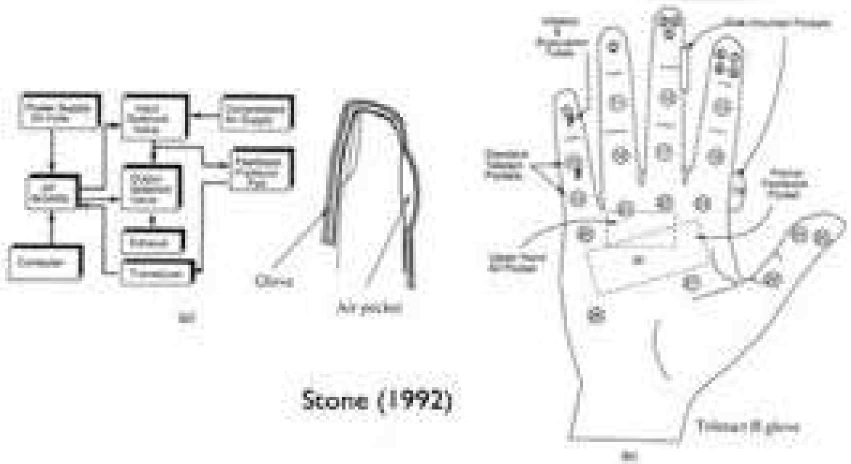

Zerkus (1993)



Sato, et al. (1991)

Immersion CyberTouch Glove

Burdea (1996)



Stone (1992)

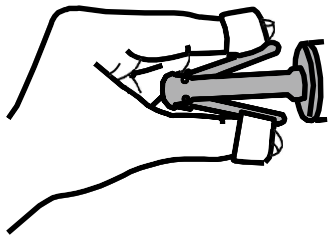
kinesthetic (force-feedback) device basics

typical kinesthetic device configurations

manipulandum

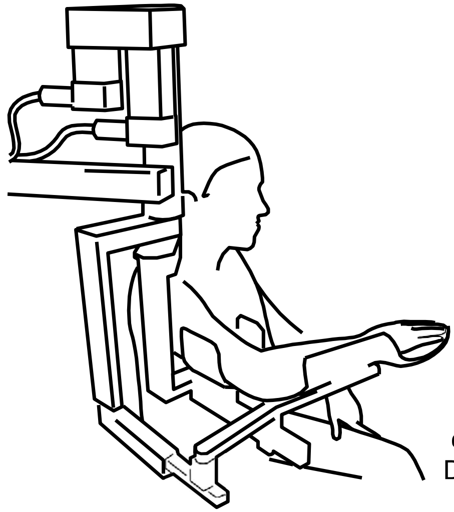
drawing by Jorge Cham

grasp



drawing by Tricia Gibo

exoskeleton



drawing by
David Grow

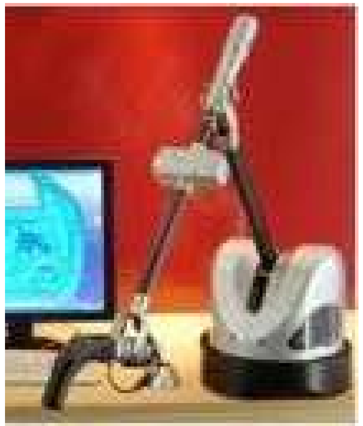
manipulandums (expensive)

Omega
from Force Dimension

delta configuration
3 degrees of freedom

Phantom Premium 1.5
from SensAble/Geomagic

5-bar + rotation
3 degrees of freedom



Virtuose
from Haption

additional “wrist”
6 degrees of freedom

all images from Wikimedia Commons

manipulandums (cheaper)

Falcon
from Novint

delta configuration
3 degrees of freedom

image from Wikimedia Commons

Phantom Omni/Touch
from SensAble/Geomagic

5-bar + rotation
3 degrees of freedom

photographed by Akiko Nabeshima

Sidewinder
from Microsoft

spherical mechanism
2 degrees of freedom

image from Wikimedia Commons

Grip/grasp

Custom haptic gripper for Phantom Premium

© 2007 IEEE. Reprinted, with permission, from L. N. Verner and A. M. Okamura.. Effects of Translational and Gripping Force Feedback are Decoupled in a 4-Degree-of-Freedom Telemanipulator, World Haptics Conference,, pp. 286-291, 2007

Single-finger Cybergrasp from Cyberglove Systems

photograph courtesy Stanford Center for Design Research

da Vinci Surgical System from Intuitive Surgical, Inc. (no programmable force feedback on gripper)

photographed by Akiko Nabeshima

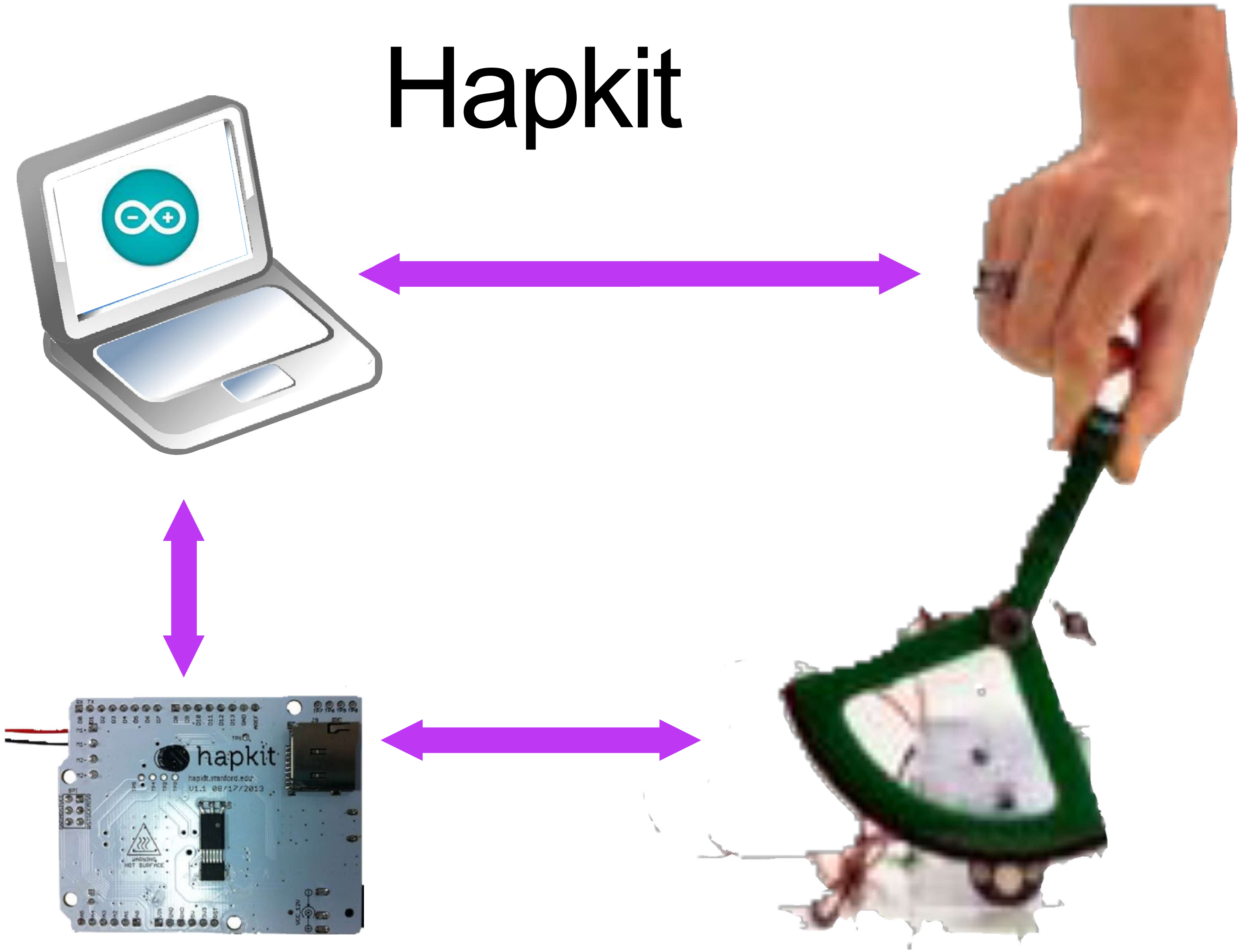
Exoskeletons

KINARM Exoskeleton
from BKIN Technologies

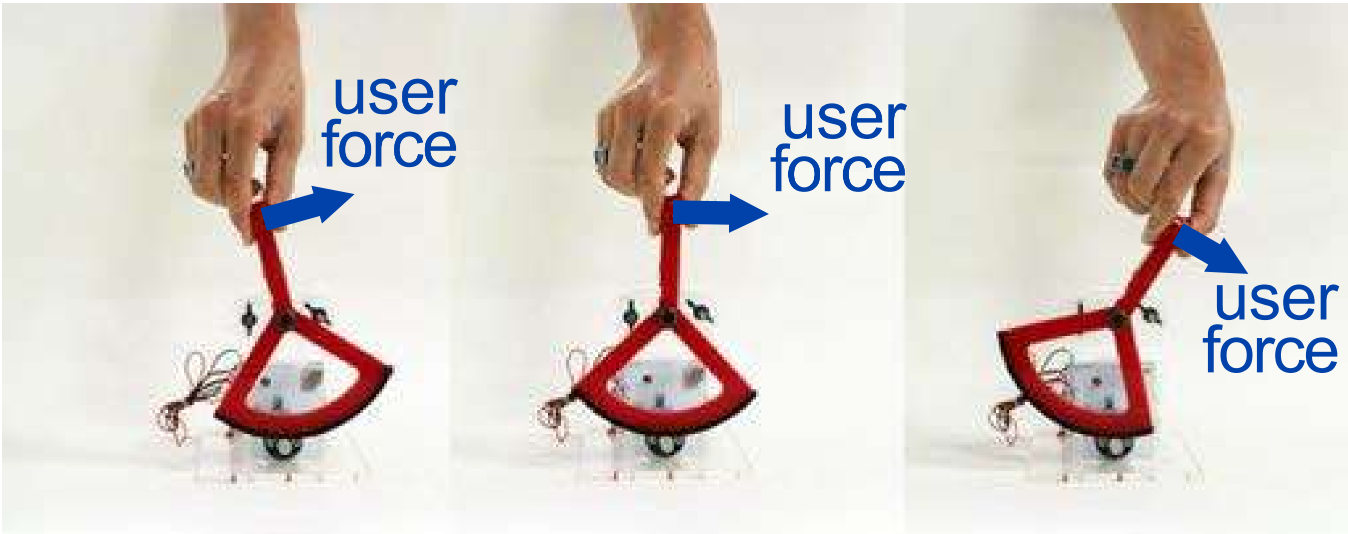
Harvard DARPA

images from Wikimedia Commons

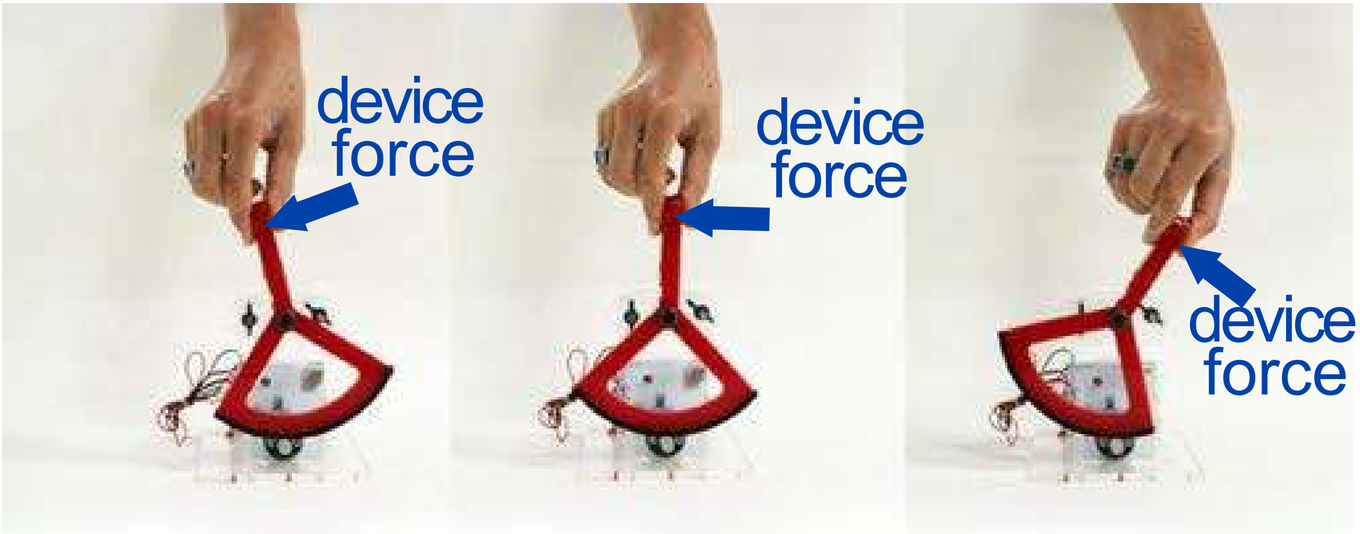
Hapkit



Hapkit



Hapkit



Trends driving haptics

- Networking - constant connectivity
- Ubiquity of computing devices - beyond sparse visual real estate
- Multitasking - doing more things at once may benefit from multiple channels of communcation
- Virtualization – fostering presence

rendering friction

(in one degree of freedom)

surface properties

- typical haptic display general shape very well, but don’t feel “realistic”
  - add surface properties to increase realism
- how you do it depends on
  - the surface model
  - complexity of the surface property
  - frequency response of your haptic device

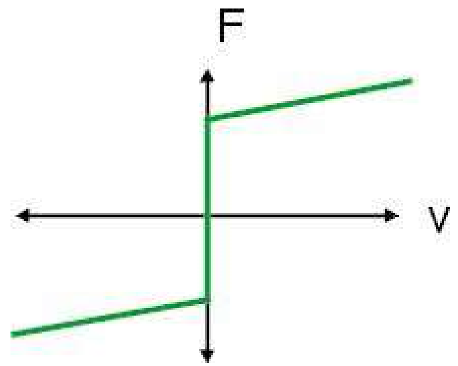
damping for virtual walls

- a pure springforce for a wall may seem to “active”
- add a dissipative term, where b is the damping coefficient
- only damps when going into the wall
- this can also create vibrations upon wall impact

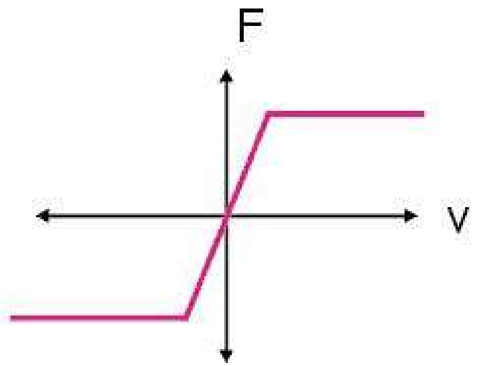
$$F = \begin{cases} k\Delta x + b\dot{x} & \text{for } \dot{x} > 0 \\ k\Delta x & \text{for } \dot{x} < 0 \end{cases}$$

“frictional” damping

- surfaces can feel unnaturallyslippery
- friction would help, but it is difficult to implement
- you can add damping to motion parallel to surface



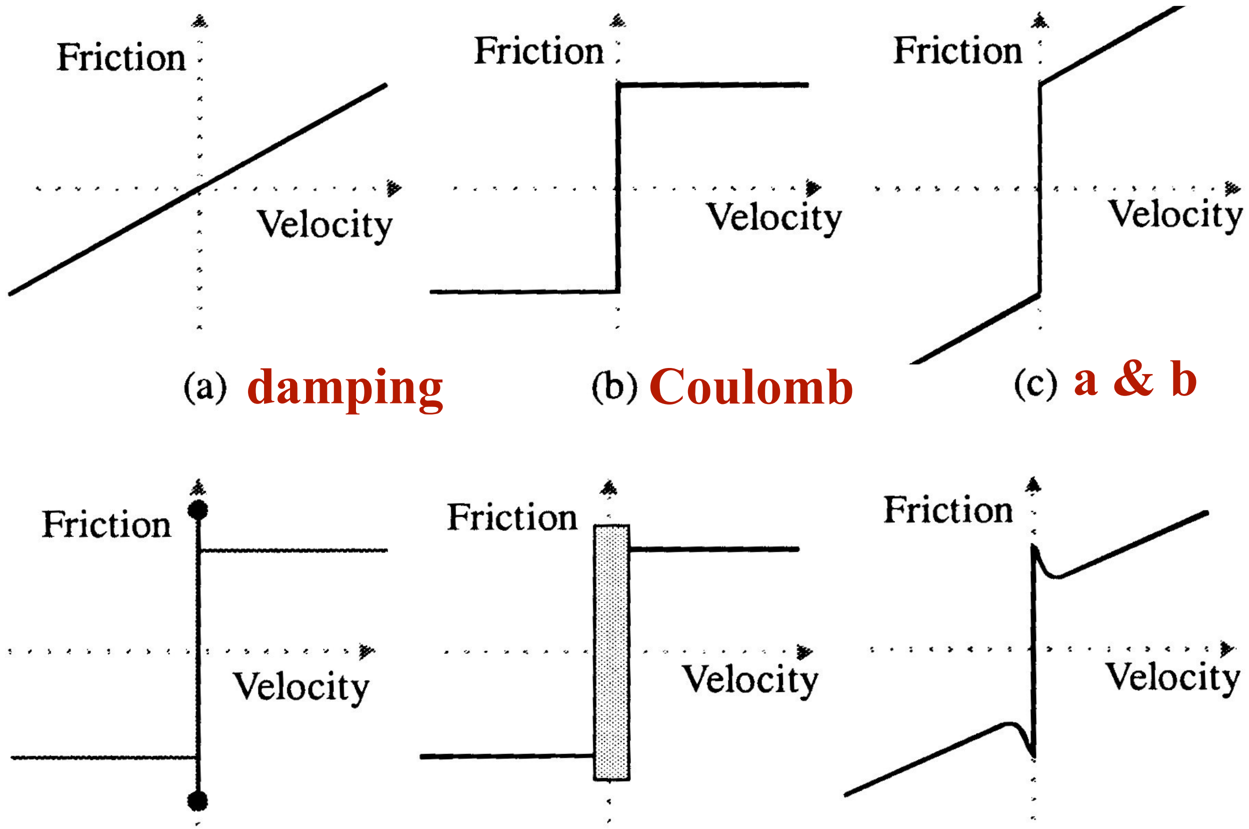
friction



damping

friction display

difficult to render because it is non-linear



(a) **damping** (b) **Coulomb** (c) **a & b**

(d) **stiction** (e) **static** (f) **Stribeck effect**

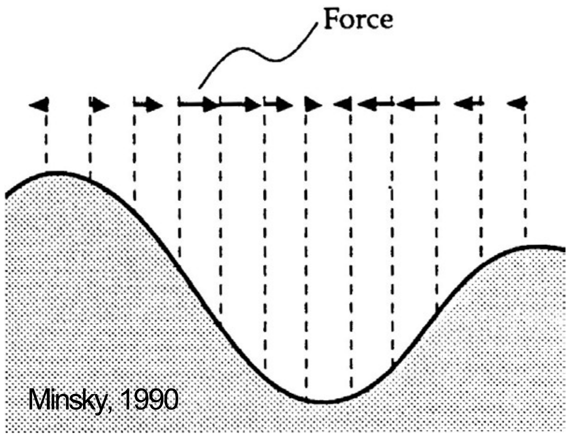
Richard & Cutkosky

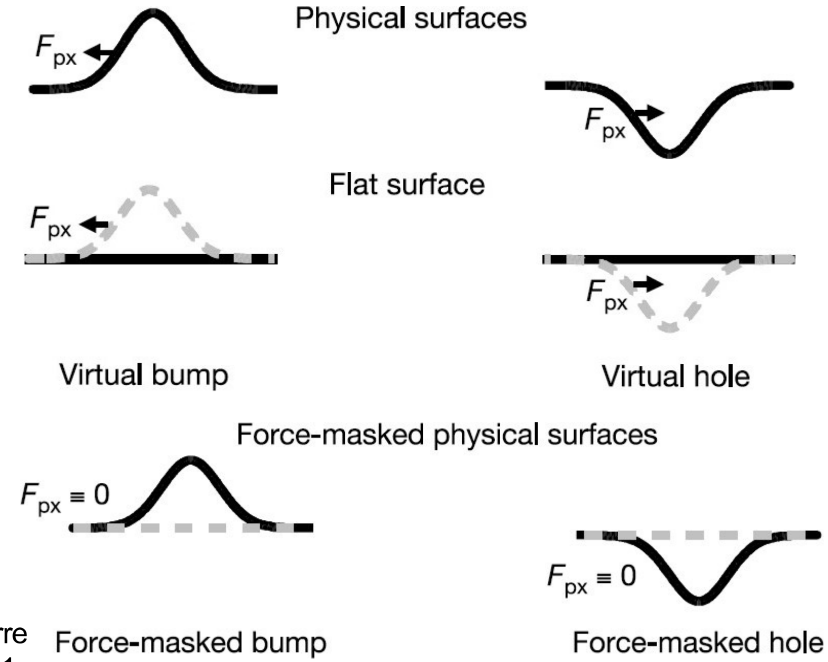
rendering bumps and textures

(in one degree of freedom)

bumps and valleys

- as a user moves "up" a bump, motion is opposed.
- done in 2D, springforce proportional to height of bump
- force information can overcomegeometry!





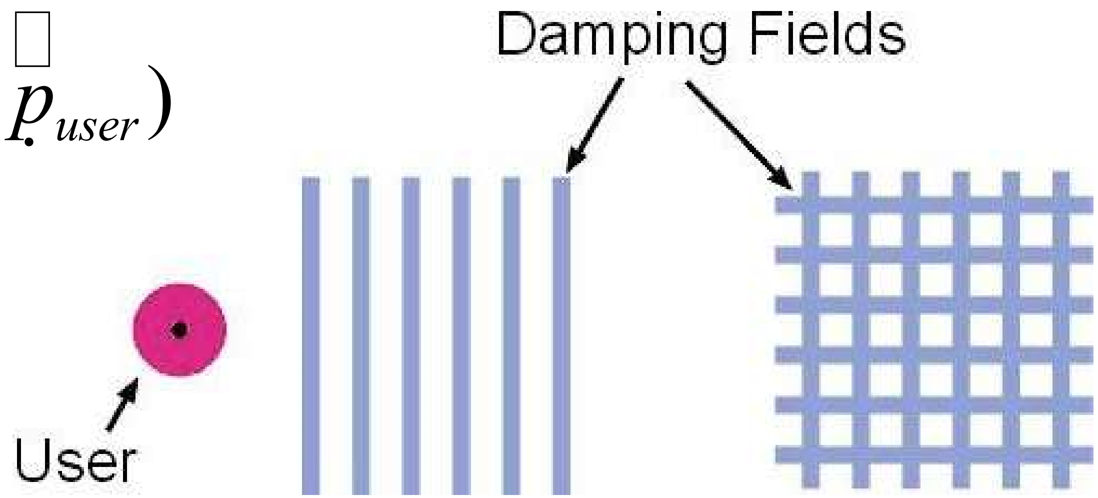
Robles-De-La-Torre & Hayward, 2001

damping textures

if $\vec{p}_{user}$ is inside a damping area

$$F = b\vec{v}_{user}$$

$$(\vec{v}_{user} = \dot{\vec{p}}_{user})$$



note that vibrations occur due to discontinuity in force

simulating and rendering dynamic objects

(in one degree of freedom)

dynamic simulation of “rigid” bodies

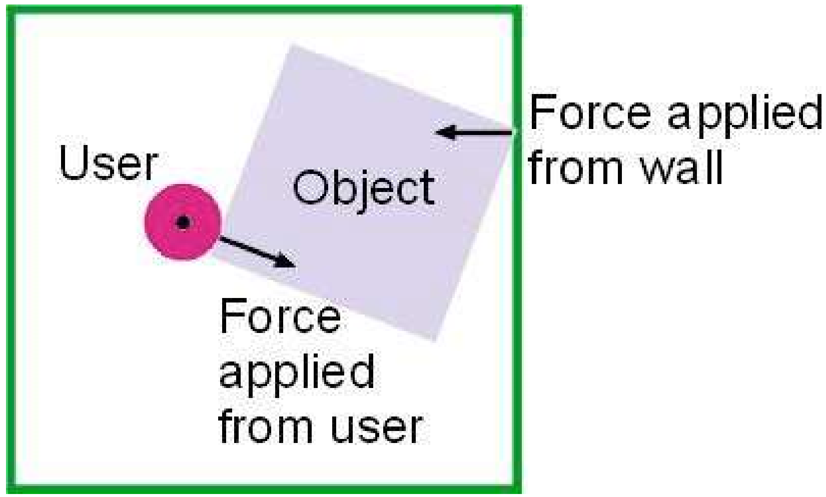
- assumptions
  - you have an impedance-type force-feedback display
  - you are using a linear stiffness model of the surface
- basic approach:
  1. save the state of the moving object
  2. sum the forces on the object
  3. calculate the new state

object state

- the “state” of an object is used to describe its current condition
- made up of variables that will change with time, such as
  - position
  - velocity
  - acceleration
  - rarely: other parameters such as mass, shape, etc. that might be changed by dynamic interaction

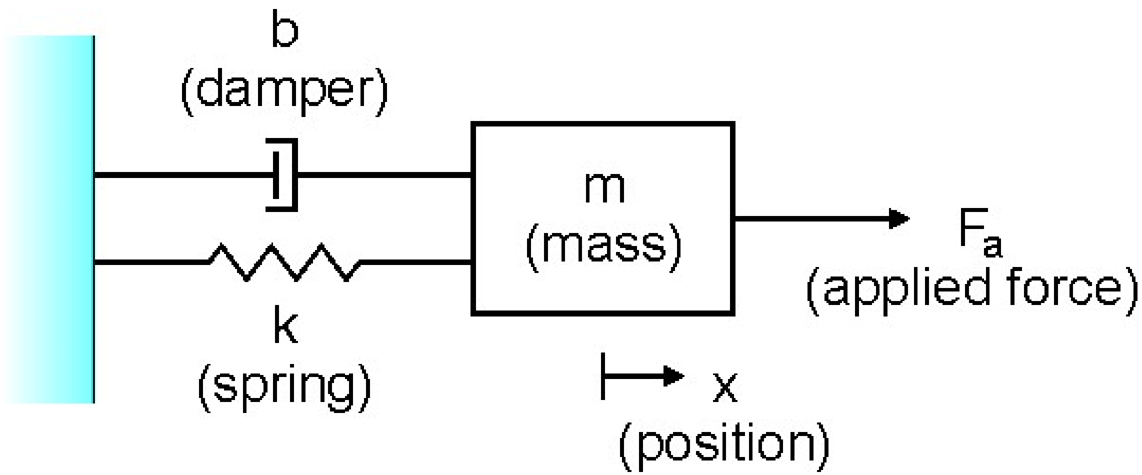
calculating forces on an object

- for forces from the user's hand pushing, this is equal and opposite to the force fed back to the user
- for forces from other objects in the VE, use the same idea: force is proportional to penetration



second order dynamic systems

mass-spring-damper



x = 0 at equilibrium