
# The Neuroscience of Game Audio

**Zak Belica**
Epic Games

**Seth S. Horowitz**
NeuroPop, Inc.

GAME DEVELOPERS CONFERENCE®
MOSCONE CENTER · SAN FRANCISCO, CA
MARCH 2-6, 2015 · EXPO: MARCH 4-6, 2015

# Why is it so hard to talk about sound?

# Because sound drives EVERYTHING

**Game (and reality) elements**

- Visuals – Huge vocabulary of nouns and adjectives
- Characters – Similarly huge vocabulary of social descriptors (age, physique, gender, race, social status, behavior).
- Physics - Rules about how elements interact – very linear vocabulary (even with non-linear rules) over specific time frames.
- Speech – strict linguistic rules.

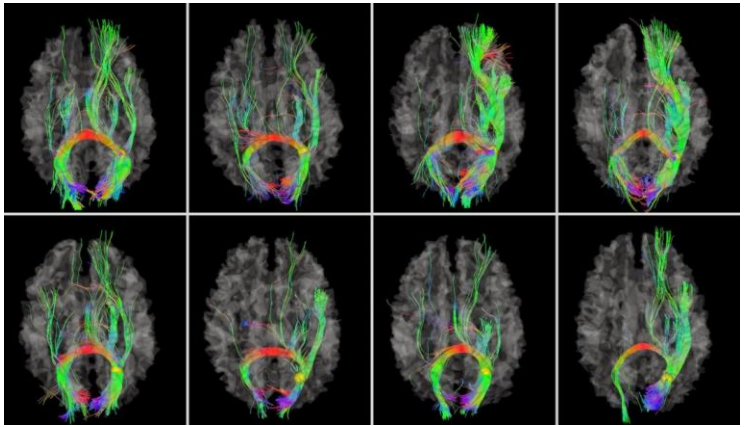
- Sound is about events in time across large time frame (milliseconds to minutes or more).
- Non-scientific vocabulary highly subjective.
- Deeply tied to emotional and unconscious states and reactions.
- Perception of sound works at pre- and subconscious time frames.
- Enormous cultural and demographic differences color perception of situations and events.

# What is a game?

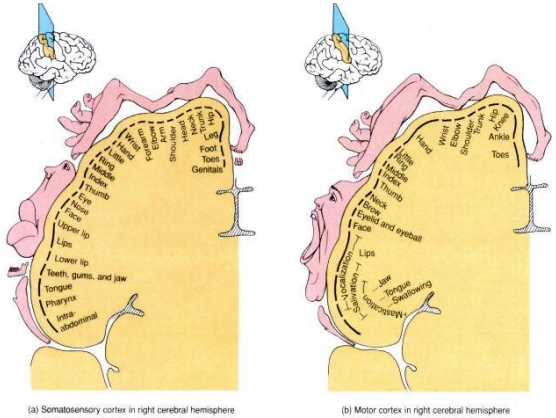
- A reality in a box
- Reality is built from our reactions to input
- Inputs come from our six senses
- But for the most part, game realities have to rely on only two senses: vision and hearing.
- In a well designed game environment, the brain fills in the rest.

# Reality vs. the brain

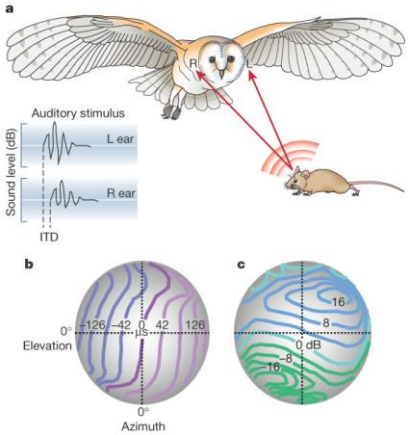

# Maps of the world



Diffusion tensor images of connectivity of memory, vision, language, arousal (Liu et al)



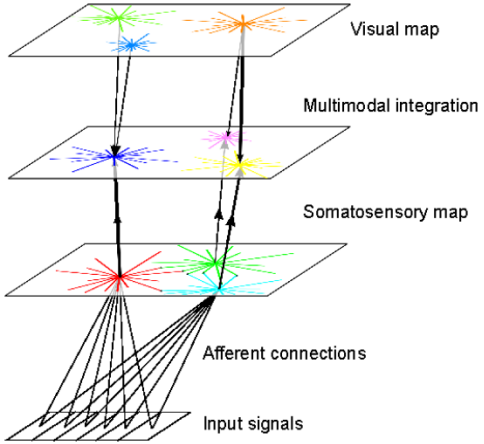

Cortical mapping of somatosensory (left) and motor output (right) in human cortex (Penfield).



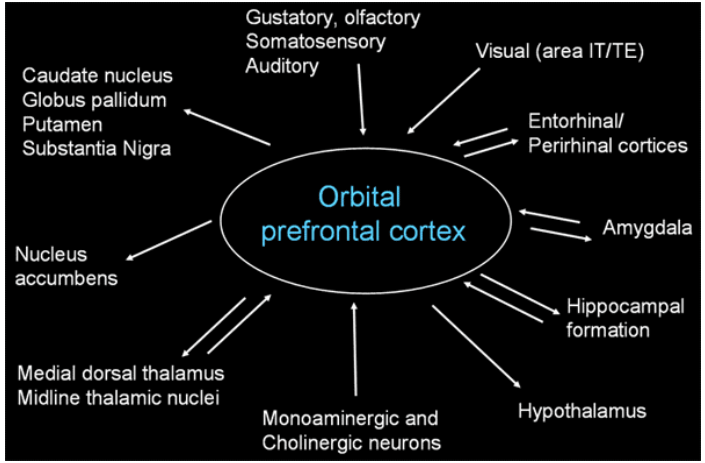

Owls (and humans) make spatial maps with sound that guide their vision (Knudsen)

- Psychophysics – mapping the physics of the outside world onto the psychological internal representation of the world.

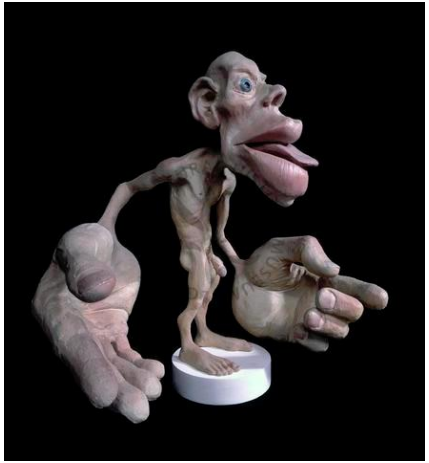
# Lining up the maps




Subcortical regions bring early sensory data into alignment into common coordinates (Pittl et al)



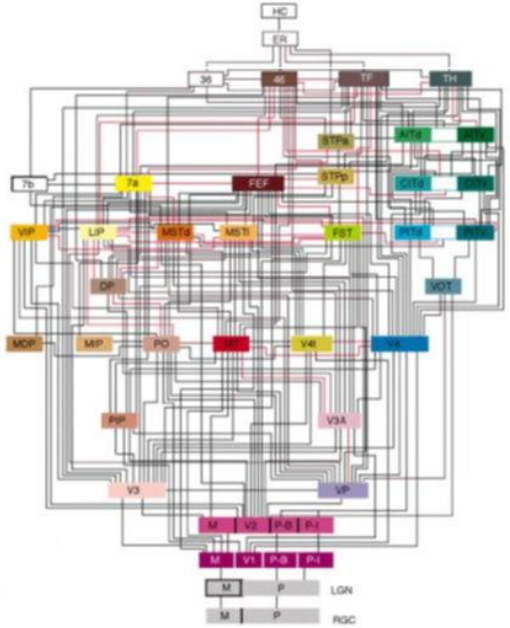

Decision making areas like the pFC integrate senses on a conscious level

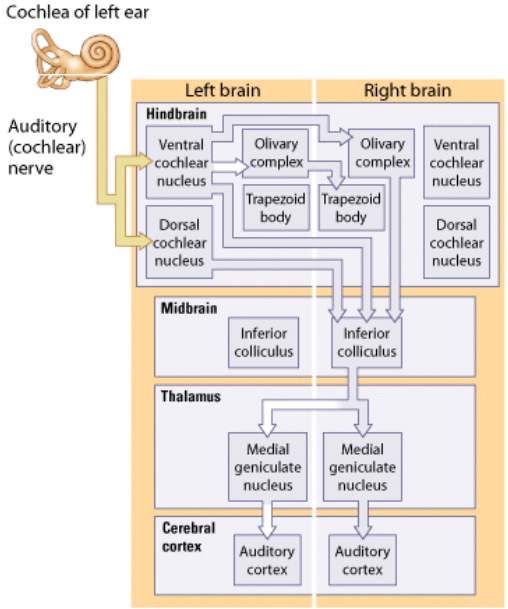


But the map is not the territory (sensory homunculus)

- The brain makes maps by bringing different sensory/motor phenomenon into a common register, using mechanisms like attention.

# Navigating some maps takes more time than others.




- There is no place in the human brain you can't get a visual response.
- Visual recognition takes a minimum of 0.25 seconds, usually more like 0.75 seconds or about the speed of conscious thought

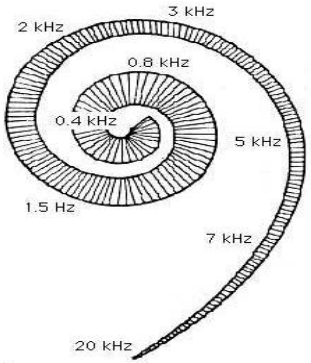
- More discrete projections from the ear throughout the brain (although goes to almost as many places).
- Complex feature processing and responding requires very little brain at all.
- Recognize sounds in 0.05 seconds
- Differences easy to detect down to 0.0003 seconds

# Sensory speeds – Never quite "now"

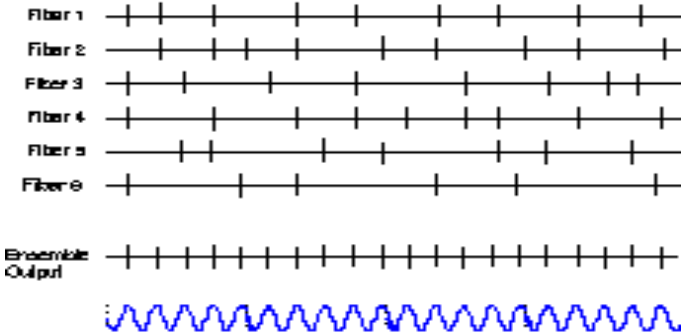
- Vision (150-400 msec)
- Hearing (50-200 msec)
- Touch:
  - Deep pressure (proprioception) skeletal muscle 80-120 m/sec
  - Light pressure (mechanoreceptor) 53-75 m/sec
  - Pain/temperature 5-35 m/sec
- Smell (500-2500 msec + effusion time)
- Taste – extremely variable depending on the components, often not overlapping but close to smell
- Balance – (20 ms to eye correction but up to 480 msec until perceptual onset.)

# Why is hearing different from other senses?

- Hearing is a universal sense.
- Hearing is the fastest sense.
- Even though humans are “visual” animals, our sense of hearing provides us with our primary handle on the environment, out of line of sight and even in the dark.








- Frequency
  - Place of maximal vibration along basilar membrane
  - Which hair cells respond
  - A tonotopic map in the cochlea
- Period
  - Auditory nerve fibers measure the time interval between individual cycles in the sound
  - Neurons “phase-lock” cleanly up to 1500 Hz, “Volley” with other nerve fibers between 1000-5000 Hz.

# The Quick or the Dead

**“The ear doesn’t blink”**
No blind spots - Works in the dark - Works when asleep.

Hearing tells you what is happening in the world:
The environment you’re in shapes the sound.
The shape of your ears and head shapes the sound.
Your age, health and personal history shapes what you can detect.

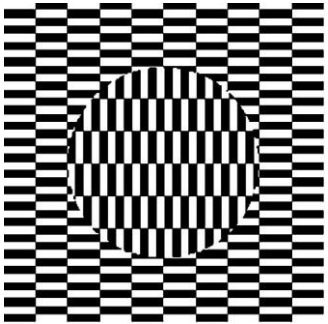
**Within 50** milliseconds:

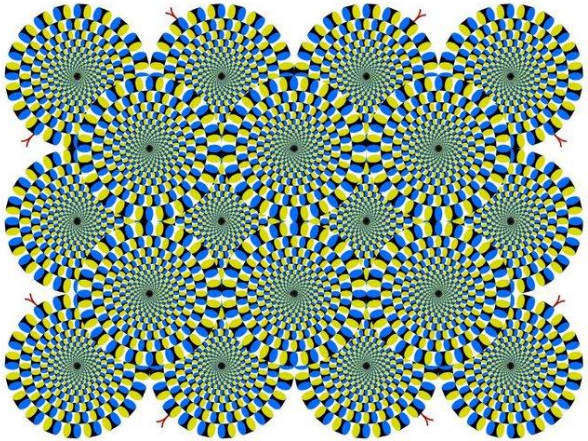
- Where it is.
- What it is.
- Who it is.
- Should you run away from it.

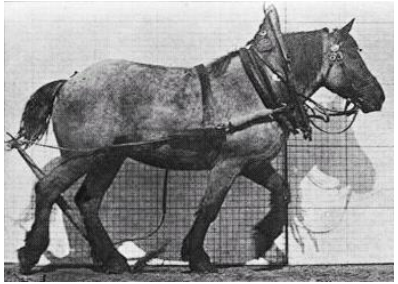
High speed auditory processing underlies our perceptions of complex properties of the environment.

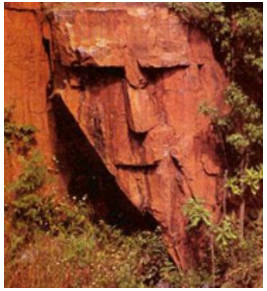
- Material
- Density
- Weight
- Power
- Emotional meaning
- Condition of item
- Time/space/place

# Is Seeing Believing?









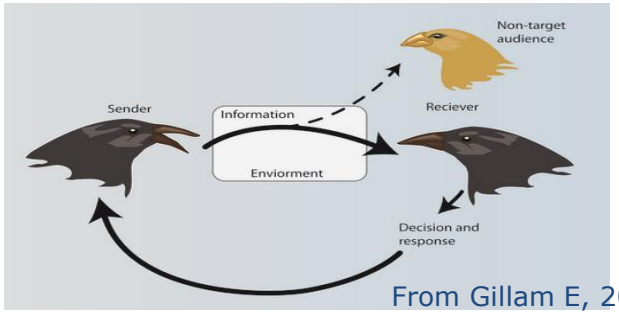
# The “speed” of hearing is why there are so few auditory illusions

- Auditory illusions are rare and usually subtle.
- Require deliberate manipulation usually by technical means.
- Pitch based illusions depend on very complex sounds with multiple harmonics.
- But they can be very revealing about how the brain processes complex sounds.

# Psychophysics & the Non-Linear Mind

- Psychophysics is how we go from the physics of sensation to the psychology of perception.
- Hearing is more than the physics of sound.
- Your ears are not digital receivers.
- Your brain isn’t even CLOSE to linear.
- We *sense* everything within our range, but we pick and choose what we *perceive*.

# The signal and the noise




From Gillam E, 2012)

From Hill, M 2010

- To listen, you need to differentiate between signal and noise.
- Listening implies communication: a sender, a receiver and a signal.
- Signals generated by breaking up noise into temporal or spatial patterns.
- To identify and understand a signal (and tune out the noise), you need to *pay attention*.

# Attention is about synchronized input over time

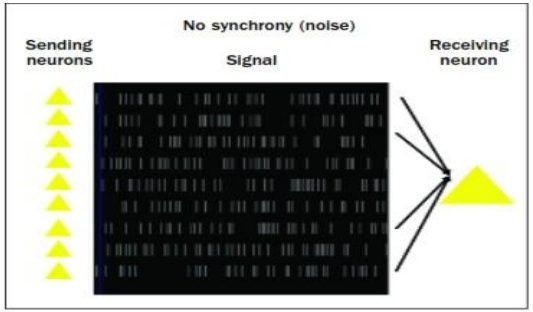
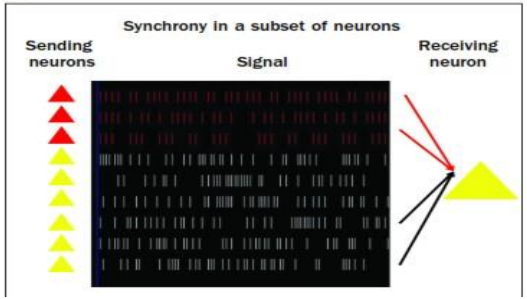
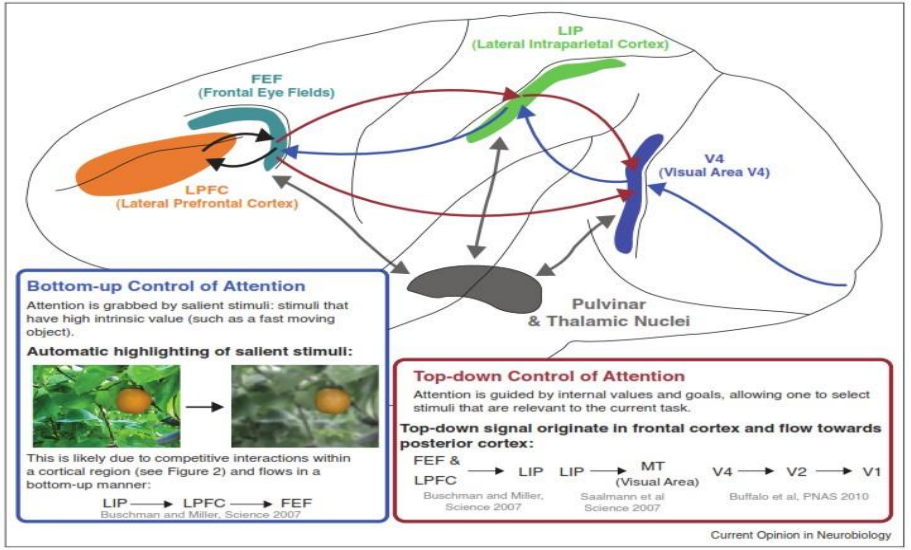



Figure 3.

Desimone 2007

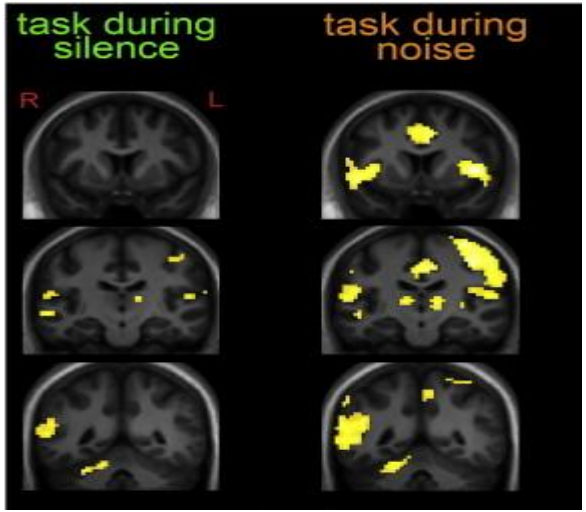



- Signals align in time based on overlap of features
- The greater the synchronization the easier it is to shift attention and the harder it is to ignore the feature.

- Two types of attention: Top down (task driven) and bottom up (stimulus driven).
- Each has separate pathways.
- Final target is the prefrontal cortex.

# Why paying attention is hard.

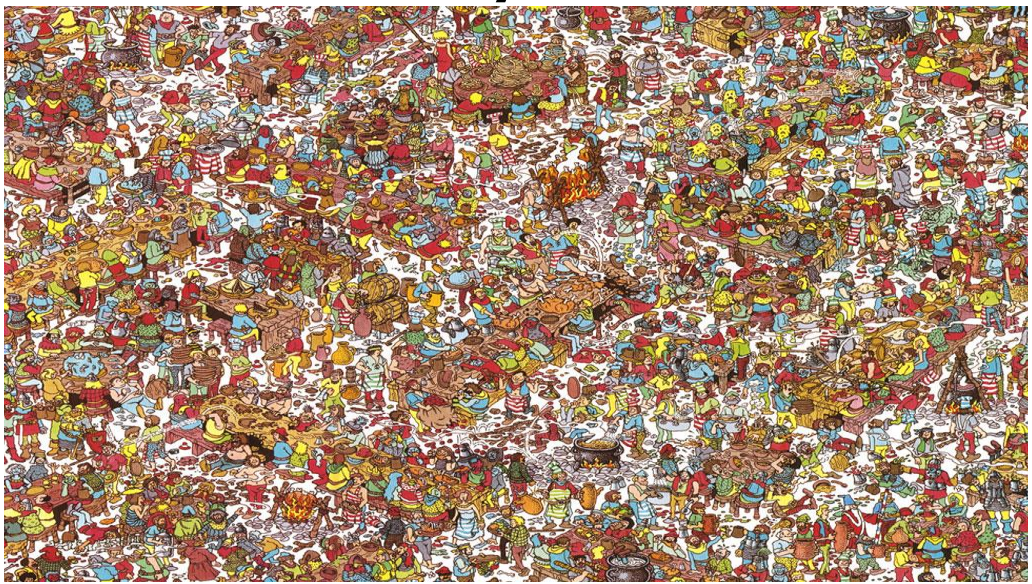



From Tregellas et al, 2012

- Attention span as measure of work brain is doing on tasks.
- Attention NARROWS your input.
- Tremendous natural variation on depth and span of attention.

- Extended listening highly energy intensive.
- Extended listening fighting hundreds of millions of years of evolution.

# Visual vs auditory attention

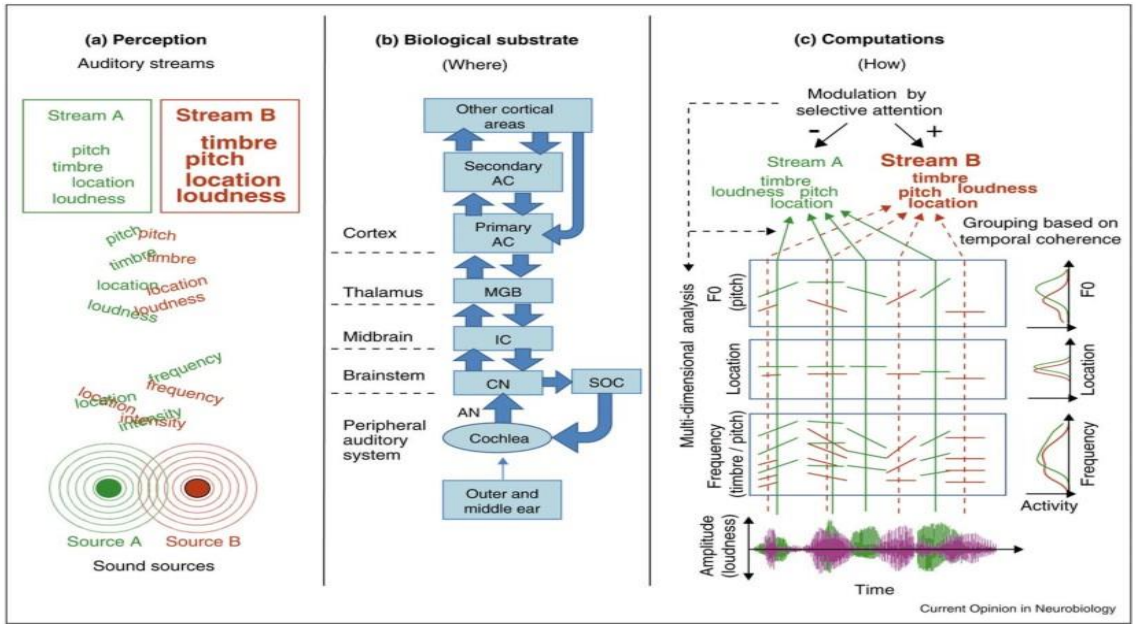


- Finding Waldo can take minutes due to slowness of visual search in a noisy field.
- Cocktail party effect (auditory equivalent of "Where's Waldo?") shows your ears can find relevant sounds in milliseconds.

Cocktail party

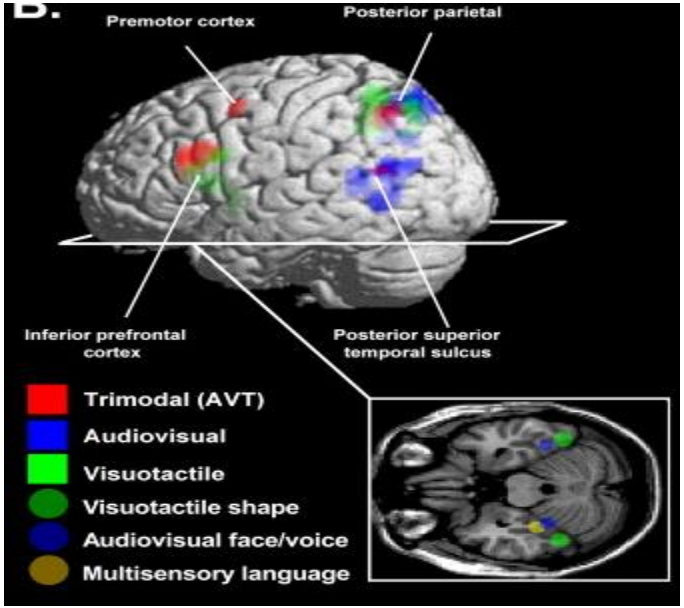
Your Party?

# Auditory Attention

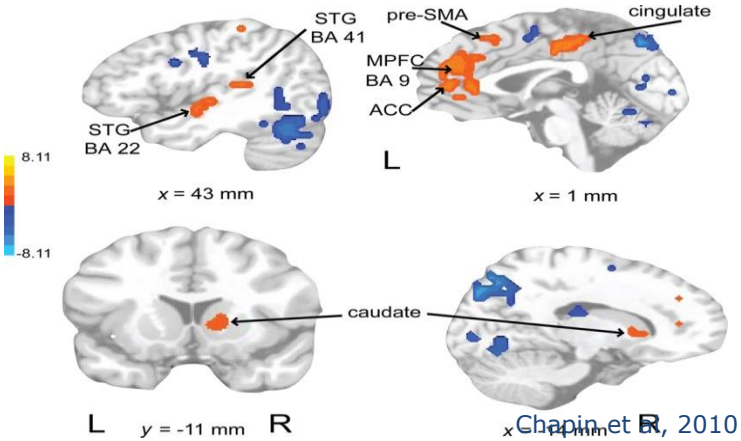



- Auditory attention is different from visual attention even though it feeds into similar pathways.
- Auditory input 5-20X faster than vision.
- Better at stimulus driven, can be harder to sustain for task driven.

# Listening with More Than Ears




Driver & Noesselt, 2008)



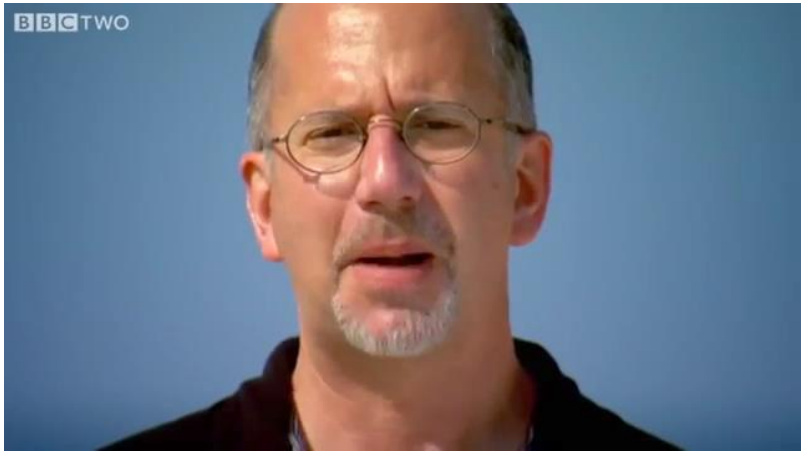

Chapin et al, 2010

- We think about listening as if it's something locked into our ears.
- In reality, there are few places in the brain that respond to only one sense.
- Attention operates across all the senses, and the narrowing of focus can use more than one sense.

# Multisensory attention

- When objects are perceived by multiple sensory systems, they increase measurable attentional loads.
- Multisensory attention, especially when consciously attended to, calls in areas of the prefrontal cortex where it plays a part in decision making.
- This **slows your brain down a lot** and makes it amenable to errors and illusions.

# The McGurk Effect: Vision guides what you hear





# Attention & Cognition

- Attention is a range limiting process (yielding apparent resolution enhancement).
- Attention reinforces familiarity and creates expectations

# The roles of context and expectation

Rain on a roof? Bacon cooking in a pan?

# Violation of expectation = emotional response

# Emotional Listening

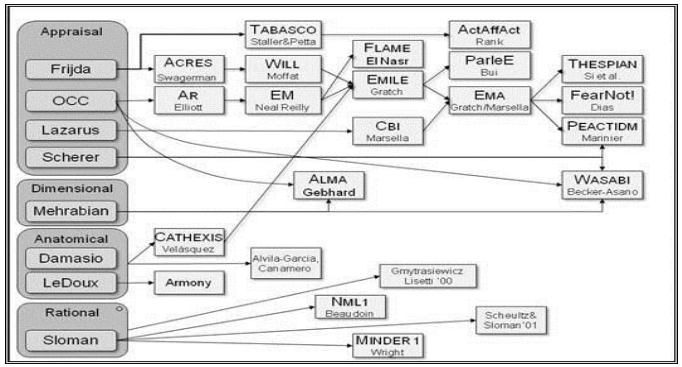
- Sound is the most powerful driver of emotions.
- Music is an emotional language.
- We rapidly and without conscious volition respond to and identify the emotional meaning behind some sounds, especially music…



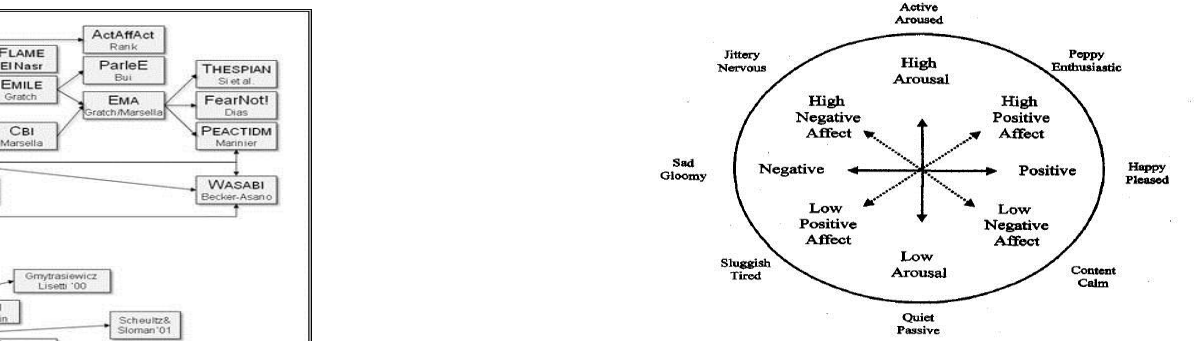
## Some benefits and drawbacks…

"Music activates similar neural systems of reward and emotion as those stimulated by food, sex and drugs.' R. Zatorre. Montreal Institute of Neurology.

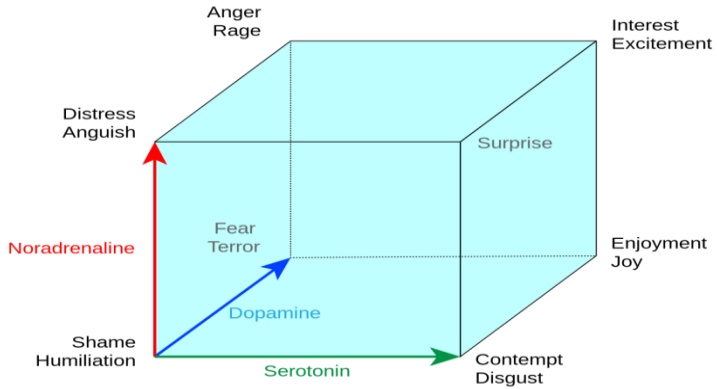
# Quantifying emotion




Many formats for trying to quantify emotions, from sociological to neuropharmacological.



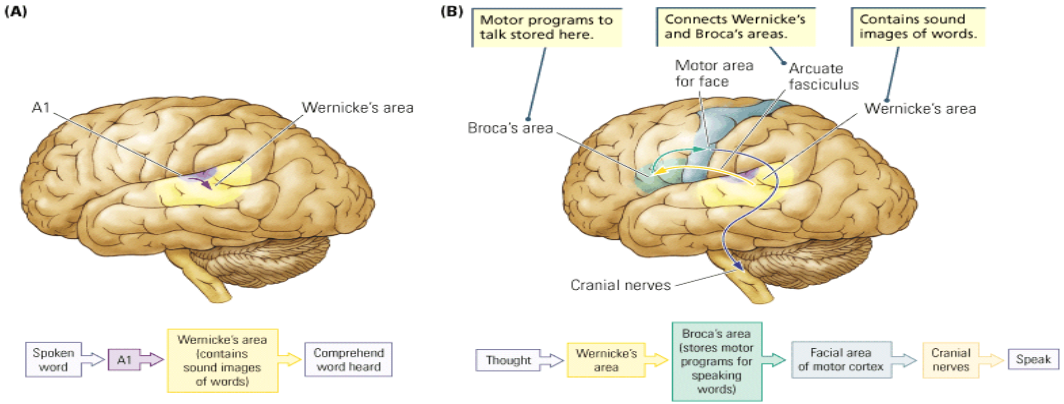

Simple 2 axis circumplex model based on arousal vs valence (Larson & Diener, 1992)

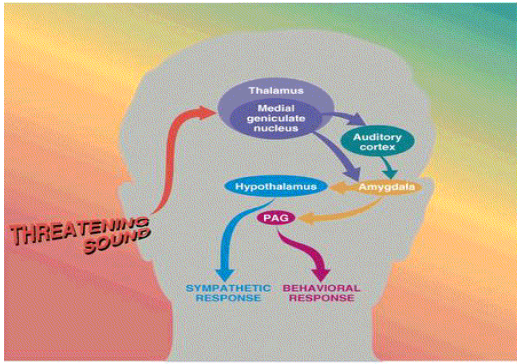


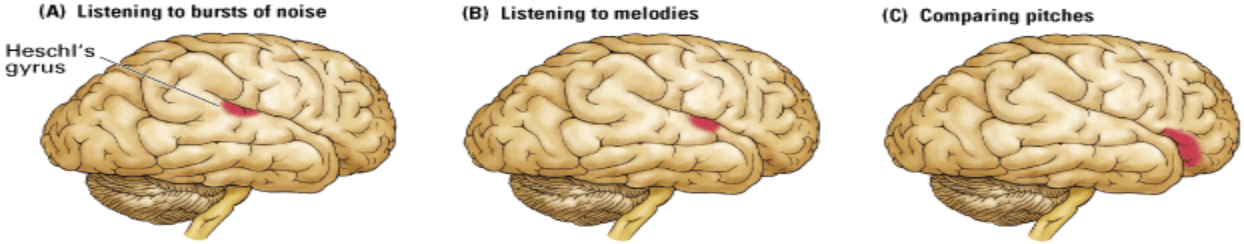

## None of them are entirely satisfactory

"Emotion cube" model based on neurotransmitter release level correlation Lövheim, 2011

# Auditory and attentional systems are deeply wired into emotionally responsive regions of the brain



# Manipulation of sound enables us to trigger specific brain responses

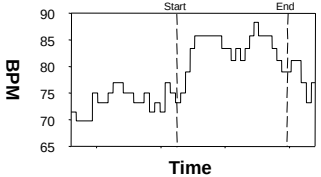
# Tiny changes in fine structure of sound has a huge psychological impact

Pink noise (left) vs pseudorandomly amplitude modulated pink noise (right)

# Rock your body: Sound affects physiology

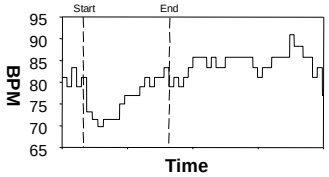
Auditory facilitation: entrainment of respiration and heartrate by sound.

Sympathetic arousal – Pupillary dilation (3Fs)




NSA: "Mr. Furious" Sympathetic stimulator (Wong, 2000)

Parasympathetic arousal (alternating contraction)




NSA: "Afterglow…" Relaxer/Altered state inducer (Wong, 2000)

Arousal?

# So what can you DO to your ~~victim~~ player with proper sound?

Unexpected physical effects: "Eyeball Twitch" modulation at resonance frequency of human eyeball (18.1-21 Hz) causes eye twitching and occasional visual illusions

Emotional manipulation: "Ghost room" – near infrasonic distortion increases listener unease with repeated listening

Make them MAD: "Mister Furious" Sympathetic nervous system activator (Fight/Flight Driver)

Physical relaxation: "Oceanic" Auditory facilitation (breath/heart rate controller & relaxer)

Make them sleep: "RealSleep" – use sound to trigger vestibular systems that control and induce sleep and alertness (Sleep Genius)

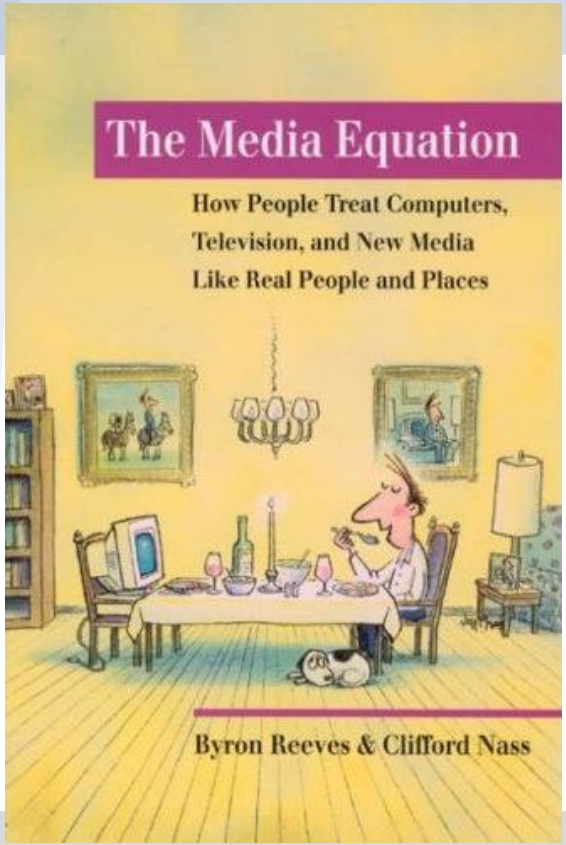
Make them feel like they're moving (or make them throw up). "Vertigo Tour" – trigger vestibular functions with LF sound.

# Expectations of world model

- Reinforce… or break and disappoint
- Giving meaning to player verbs- power, agency, and mastery.
- Audio as **secret game design**

# Does audio make game visuals look better?

Of course! Right? ...
Yes, but not quite like you think...




- The brain accepts AV input as reality, and treats it as such (!).
- Good audio –can- improve visual perception, but mostly…
- Poor audio quality can degrade visual perception
- Poor audio synchronization can degrade total perception

fidelity

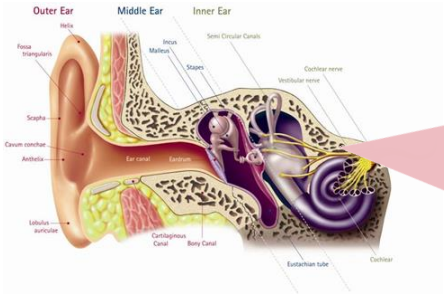


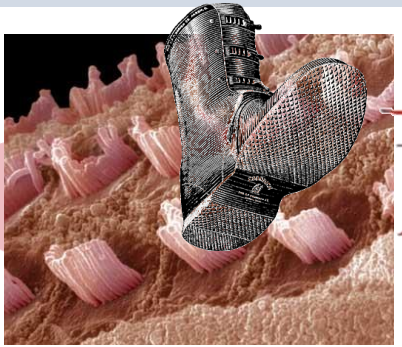
synchronization

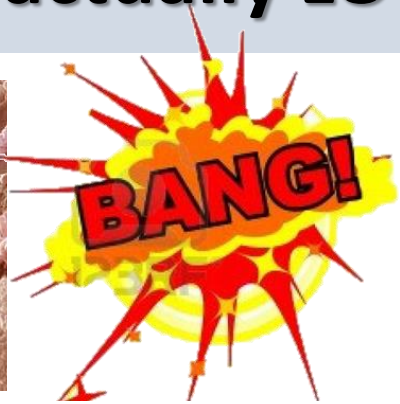
# Actor Performance Consistency

- VO as part of audio score…
- Vocap
- What if you can’t vocap?

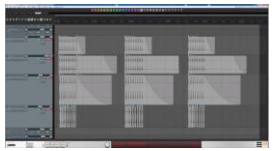
# LOUDER Without making it actually LOUDER

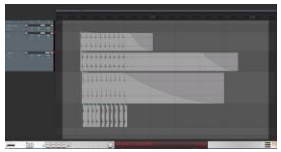


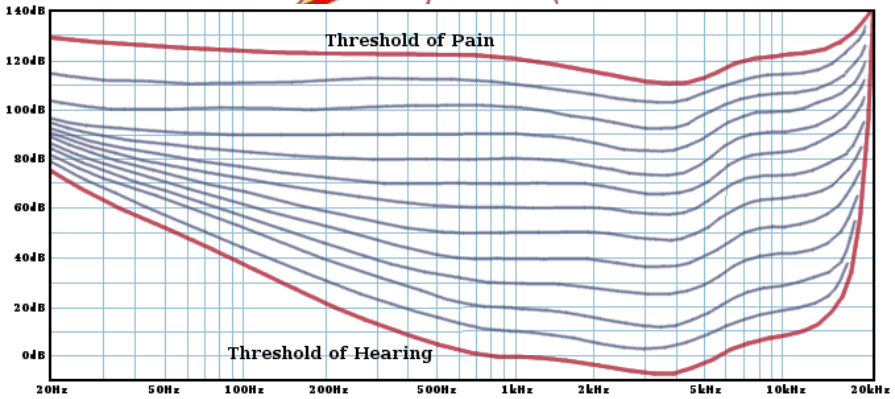






## MIMIC the PAIN







# Talking about sound…

- Fidelity (accuracy) versus inaccuracy
- Naturalness versus awkwardness
- Pleasantness versus annoyance

- Signature audio identification

# Signature Sound

- Tactical
- Branding
- Expected
- Unexpected

# Audio and attention channels

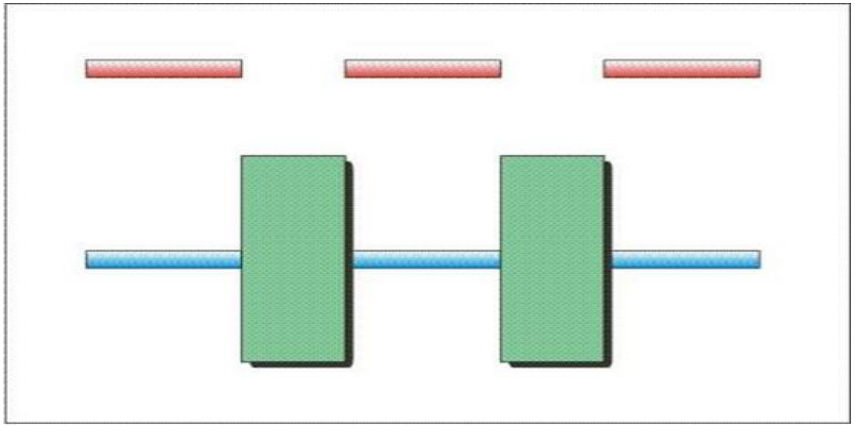
- Impressive sound and still be ineffective.
- Give context, meaning, and a free channel.
- Do not overwhelm.

# ‘wall of sound’ problem

# Bulletstorm case study

- The spatial dogpile…

# The continuity illusion



- One aspect of the wall of sound we try to fight is the Continuity illusion. Here's a visual example, where the red shapes look separate, but the same shapes in blue look continuous due to the green shapes between them.
- You will hear short sine wave beeps, with white noise in the background. As the white noise gets louder, the beeps will start to sound like a continuous tone.
- Too much continuity makes audio confusing and sound as a single audio event.

# The rule of 2½ and ‘tuning’

**Language**

**Encoded**

**Sound Effects**

**Encoded-Embodied**

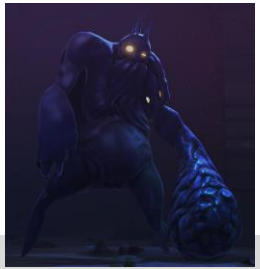
**Music**

**Embodied**

# Strong response sources

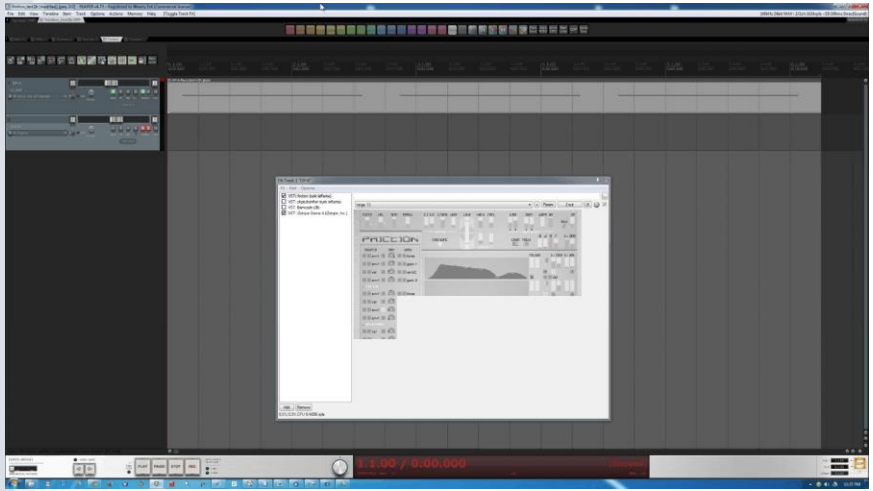
low sounds = scary

Nails on chalkboard



Some examples…

# Friction Monster…




# Speech Perception

Leveraging speech prosody/melody
But not in all cultures…

Important considerations:

- Spatialization
- Masking between 200 and 5000 hz
- Transmitted speech - intelligibility

# Human – Animal - Monster



# Music In Games

Some always turn off music... Why?

- Foreshadowing too obvious
- Distracting emotional beats
- Cognitive dissonance

# Emotional Response to Music

**Built-in:**

Evaluative conditioning

**Learned:**

Episodic memory

Music expectancy

Visual imagery

# Nostalgia: Evaluative + Episodic

Not imprinted, limited response:

Imprinted, response:

Play session style can inform how to approach ‘earworm’ goal

# Audio in a VR Environment

- Movie-Game-VR Contrast
- HRTF issues and Azimuth Effects
- Spatial relation and reflective environment
- Single source observation (TV) vastly different than full VR presentation
- Depth Congruency
- Spatial panning… behind the listener?
- Vestibular system issues

# The Future

- HDR + Meaning + Focus
- From recording to generating.
- Scientific audio ammo!  UX too.
- VR Audio

Zak.belica@epicgames.com
shorowitz@neuropop.com