PHYSIOLOGY OF AUDITION

The Problem of Perception

- "The problem of perception is initially a problem of taxonomy"
  - We must classify things in the world
  - We discard a lot of the total information in order to cope with the continuous stream of information
- "Internal taxonomy of perception is adaptive …"
  - We perceive in a way that changes with the world around us, but our brains cannot get a complete detailed faithful image of everything we see in our heads
  - Perception is adaptive but it doesn't mean our perception is an exact copy of the world we perceive

Physical Properties / Sensations (Perceptual Properties) / Musical Properties

- All information on slide

Stages in "Perception"

- **Transduction**: any change in energy form (Cochlea)
  - Input is physical energy to cholera, the output is electrical/chemical energy
  - From air pressures changes → trains of spikes (electrical)
- **Low-Level Detection**
  - The coding of information happens in parallel
    - Sensations of loudness/pitch/timbre happen at the same time
- **Scene Analysis**
  - Making sense of the sound happening around you
    - Decides foreground/background, left/right side
- **Memory, Anticipation, Emotion, Goals**
  - We take the scene we have analyzed and integrate past experiences with similar stimuli

Perception versus Cognition

- Perception
  - Low-level processes
  - Transduction and basic sensations (pitch, timbre, etc)
  - Happens bottom-up
  - Automatically processed (you can't choose not to get a pitch sensation)
  - Encapsulated (this involves specific circuits that are for auditory information, separate from visual information circuits for example)
- Cognition
  - Involves our consciousness
  - Top-Down processing (from the brain downwards)

        - Electro-chemical patterns become meaningful to us and adaptive
            - You can reason about the parts of music, composer, etc.
        - Consciousness-mediated (involves an active will to do something)

- Cocktail Party Effect
    - Once you hear somebody say your name at a party, you can follow that voice around the party as it moves
    - This is top-down perception… you tell the brain to filter out the rest of the noise to continue tracking this voice

Sensory Systems

- Exteroceptors (our ears): track what happens outside the body
- Interoceptors: keep the brain informed about the status of internal organs
- Proprioceptors: information about the body position, limb movements, kinesthetic system
    - Touch and hearing are the only two physical sensations
    - Embryos don't distinguish the tissue for the two sensations early on

Anatomy of the Ear

- The length of the ear canal acts as an amplifier (filters specific frequencies)
- The middle ear acts as an amplifier
    - Shaped like a lever, which receives input from sound waves, the length of the wave affects the action of the lever
- The cochlea is the strongest bone in the body
    - Cochlea = spectral analyzer?

The Outer Ear

The Middle Ear

The Inner Ear

- Chemical energy before this becomes electrochemical

The Cochlea

- Basilar membrane and tectorial membrane
- There are specialized neurons in the cochlea (hair cells)

Traveling Waves in the Cochlea

- The velocity of the frequency of wave hitting the oval window will have a maximum at one point in the cochlear duct
- Where the maximum amplitude of the wave happens determines the frequency of the sound
    - The higher the frequency, the further in the basilar membrane the sound reaches
- Where the maximum wave stops moves the specific hair cells bend
    - This is the origin of the electrical impulse sent into the brain
- The elasticity of the basilar membrane can be changed
    - This can be responsible for the filtering out of sounds
        - This is the mechanism behind the cocktail party effect

Cochism

Neurons (Hair Cells)

- Sends information upwards to the brain
- Sound of neurons talking = clicks
    - Closer in time, the clicks happen = higher pitch sound
- Neurons have a refractory period while neurotransmitters are replenished

Coding of Acoustic Information

- Separations of spikes while help the system discover the periodicity
- Phase-locking
    - Neurons in the cochlea are phase-locked
- Can use this information to infer the frequency of a sound
    - Look at the number of spikes as well as the separation of spikes

Auditory Pathways

- After cochlea:
    - The first function of the auditory system is to determine the source of a sound
        - Helps to stop unexpected and dangerous sounds from getting ignored
    - The second function is to detect and classify what the sound is
    - The third function is to learn from the sound and its meaning
- Along the pathway are structures to determine specific qualities of the sound that get processed in parallel, BUT before reaching the cortex
- The whole system computes the essential functions of sound identification, NOT just the upper parts of the brain
    - Loudness, pitch, etc. are already available at these places

The Brain

Auditory Cortex

- Located in the temporal lobe
- Primary cortex takes in all previous processed information (loudness/pitch etc)
    - Sends the signals to other parts of the brain (secondary cortex)
        - To make sense of sound/what it may be needed for
- Tertiary cortex
    - Connects sound to other sensory information
    - The place responsible for synesthesia

The Musical Brain

(Just memorize information from the slides)

Timings in The Musical Brain

(Memorize what aspects are part of psychoacoustics vs cognition)

Tonotopicity

- Information is carried in an organized way
- Fibers that leave the cochlea leave together
    - Low or high frequencies travel together
    - Reach auditory cortex in an organized way
    - Different information arrives at different places

<u>RESEARCH METHODS</u>

How to “Measure” (Music) Perception and Cognition?

- Surveys
    - Information gathering
- Experiments
    - Variables that can be manipulated
        - dependent/independent
        - Causal relationship
- Behavioral measures
- EEG
    - -this includes ERPs
- Magneto-Electro-Encelphalography (fMRI)

Physiological Measures

- Heart rate
- Skin conductance
    - These changes in relation to emotions
        - IF YOU WANT TO STUDY EMOTIONS THIS IS A GOOD PROXY
        - CHEAP AND EASY
- Blood pressure
- Temperature
- Muscle Tension

Behavioral Measures

- Errors in solving a task
- Reaction time
- Relatedness, similarity
- Choice
- Eye fixation
- Motion capture

Behavioral Tasks

- Detection: did you hear a sound?
- Discrimination: did you hear a difference?
- Forced choice: which do you prefer?
- Direct Scaling: how loud is the sound on a scale from zero to ten?
- Adjustment: move the slider until it matches the reference
- Chronometric tasks: press button A as quick as possible when you hear a voice
- Verbal Description: what do you perceive

Electro-Encephalography

Electro-Encephalography: Spontaneous EEG

Electro-Encephalography: Event-Related Potentials (ERP)

- Put electrodes in specific areas/places
- Look at the differences between currents in the electrodes
- This occurs after experimental manipulation is done
    - Example: playing a group of chords, with one that is wrong in the context of the song
    - We observe if putting this chord in makes some change in the electrical currents

- Example: Language Processing ERP
    - When you present a sentence, and then afterward a word, the ERP changes if you present a semantically related word vs. an unrelated word
            - N400 appears when you hear a word that relates to the context of the previous sentence
        - Repeated with music:
            - Played some music, and looked at a description that matched the meaning of the piece just heart
            - N400 still elicits semantic association

Frequency-Following Response (FFR, AKA cABR)

Timings in the Musical Brain (According to ERP Studies)

fMRI: Functional Magnetic Resonance Imaging

PET: Positron Emission Tomography

Diffusion Spectrum Imaging

Computer Simulation

PSYCHOPHYSICS OF THE BASIC SOUND DIMENSIONS: FREQUENCY RESOLUTION

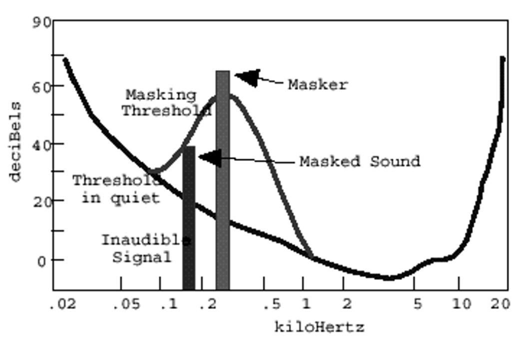
- What is frequency? Pitch, essentially
- What is resolution? Being able to separate simultaneous pitches
    - We generate multiple different ones, and can we really perceive them as different pitches or do we hear them as part of the same sound?
    - How good is our capacity to differentiate sounds?
        - 1k vs 1001hz, we probably cannot separate the difference
            - Some but not many may be able to distinguish these two frequencies

Frequency Resolution of our Hearing System

- If two sinusoids are happening very close in frequency can we perceive them as separate phenomena?
    - RESOLUTION = separation
- Most of our knowledge comes from frequency masking
    - 2 centuries of research about this topic
    - knowing/charactarizing masking effect has implications for loudness, pitch, timbre

Masking

- This is the basis for coding and communication technologies such as streaming services
- Example: two peaks that start the same and separate out in frequency
    - We perceive these two played at the same time as ‘beating’
- WHAT HAPPENS:
    - Two tones
    - One higher in frequency than the other
    - One lower in amplitude than the other
    - We have a threshold for each band of the frequency spectrum necessary to hear when no sound is playing



    - When we have one sound playing it changes the threshold for other frequencies around the sound. So the threshold for sounds lower and higher will be at a higher amplitude than

Basilar Membrane Behavior

- When we have a complex sinusoid with three frequencies, three areas on the basilar membrane will be excited creating peaks
- Because of the mechanical properties of our basilar membrane, when we detect the spectral components of a sum, it makes it more difficult for frequencies to be detected at places close by
- We experience masking IN PART due to the mechanical properties of the basilar membrane

Casual Mechanism of Masking: Basilar Membrane

- A 400 hz tone will mask 800 hz more than 200 hz
  - This is because of asymmetry
  - Masking is easier for higher frequencies than the signal rather than lower frequencies
  - There is asymmetrical bending of the basilar membrane
  - This is called **SPREADING**
  - The louder the tone the MORE RANGE the mask has

Masking Patterns

(all information in slides)

Beating (Fluctuation Strength)

- Two tones close to each other will make the **beating** or **fluctuation of strength**
- Beating is very helpful for tuning instruments (you listen for the slowest beats)
- If we can count the number of beats per second we can calculate the difference in the frequencies

Roughness

- When we have larger numbers of distances between frequencies we get **roughness**
- The beating is happening so fast that our basilar membrane cant perceive two separate sounds
  - The tone becomes dissonant

Full Resolution

- We have now separated the tones enough to be able to perceive two separate pitches

Temporal Masking

- One signal can mask signals before or after
- How can a signal mask one that comes after it?

Causal Mechanisms of Masking: Neural Functioning
- Simultaneous masking can happen due to the basilar membrane
- Temporal masking has to do with neuron pathways as well

Auditory Filters
- Every millimeter of the basilar membrane can be treated as one filter
- Frequency selectivity can be modeled with a bank of bandpass filters with overlapping bands

Critical Bands
- Critical bands = auditory filters
- Each frequency creates its own critical band
- This has to do with the distance in the basilar membrane whereby two tones of different frequencies do not interfere themselves anymore
- When you have two spectral components close together that effect the same common filters, the information that reaches the brain will not be clear or independent
    - Information from two tones perceived in one filter is what causes beating/roughness
- CRITICAL BANDS:
    - The distance at which two components happening simultaneously do not have any filter interaction
- This can be mathematically modeled using gamma tones
- The critical bandwidth:
    - We need some tones to be further separated than the critical bandwidth of a particular frequency in order to have it be perceived as two separate tones

Fletchers Experiment
- At a certain masking noise bandwidth, a common signal threshold is obtained
- At each tone he played another tone at different hz distances away and recorded the threshold at each hz
- When playing close to the

Critical Bands: Bark Estimation or **ERB**
- Formulas to compute for a given frequency what is the critical bandwidth
- Below 1k hz there is not much change in the bandwidth
    - As you get higher in frequency, the bandwidth increases,
- The Bark and ERB graphs relatively match ⅓ octave line
    - This is because above ⅓ octave steps you won’t get enough information
    - Below ⅓ octave sounds will be overlapping in their critical bands
    - ⅓ octaves makes it possible to adjust the change of amplitude in bands without being influenced by neighboring bands

        - If used less, the judgements would be more tricky

Critical Bandwidth Comparison

(all information in slides)

Critical Bandwidth Applications

- Audio coding: how can we encode the perceived spectral shape of sounds in a "compact" way?
    - Analysis of audio using Bark critical bandwidths
    - Comput the energy for each band
    - Can convert the energy into 1's and 0's to create a "Timbral Code-Word"
        - With the idea that we can analyze many types of audio signals/sounds and encode the sounds into binary dictionaries and do operations on them.
        - Can calculate the energy using the 35 +/- coefeccients
        - This is enough to draw correlations between sounds

Perceptual Coding Models

PSYCHOPHYSICS OF THE BASIC SOUND DIMENSIONS: LOUDNESS

Historical Figures

1. Helmholtz: father of modern psychophysics
2. Ernst Heinrich Weber: philosopher/doctor of psychophysics
3. Gustov Fechner: philosopher/doctor of psychophysics

Physical and Perceptual Features of Sounds

- Loudness is the result of the filtering of amplitude information
    - Amplitude is the power of the waveform
        - This is an objective property
    - The way our brain perceives amplitude changes affects our perception of loudness
        - Larger amplitude = louder (generally)
- Pitch
    - Periodicity of the waveform
    - Periods have specific notes with specific names
- Timbre
    - More complex
    - The shape of the waveform (the way that pressure changes over time) is the main originator of timbre sensations
        - We can look inside one period of the waveform to count how many wrinkles there are in the waveform

Psychophysical Laws

- The three researchers started a quest to find the psychophysical laws
    - Convinced of a mathematical way to connect the physical world and world of our subjective sensations/experiences
    - Started long-term exploration for human judgments as a function of stimulation we put into their sensors
    - Labs with subjects who were trained in a way to make sure the subjects were able to explain a phenomenon or provide data
        - Collected “subjective judgments” after being presented with certain sound intensities/ pitches along a single dimension
        - Tried to find the best fit between the physical magnitudes and the perceptual estimations
- Weber’s Law (1831)
    - The Just detectable change in stimulus intensity is proportional to the intensity
        - JND is constantly changing with changing stimulus
- Fechner’s Law (1860)
    - The relationship between physical sensations and our own sensation is more logarithmic

        - R(sensation) = k(constant)log(I)(physical property)
- Steven's Power Law
    - Our subjective sensation is a function of the intensity, but the relationship is mediated by a constant dependent on the modality we are measuring

Absolute Thresholds

- Where are our absolute thresholds for any sensation? What do we need to start perceiving a sound? How much stimulation is needed?

Differential Thresholds

- What is the minimum difference that can be perceived?
    - This includes JND
    - Differential Limen (same as JND?)
    - Quantified as the point at which 75% of population can perceive a difference

Loudness

(all info in slide)

Compression and Rarefaction

- We have a sound source that is vibrating (string, skin, tube, etc.)
- We play the instrument causing air particles to vibrate in an ORAGNIZED WAY
    - Is transferred from particle to particle (and some converted into heat)
    - Particles do not travel long distances
    - We create an energy displacement, waveform is propagated
    - At certain moments/points of space we have an accumulation or lack of particles
        - We can measure the pressure of the air at these places

Absolute Thresholds

- Many different measurement conditions exist
    - These yields slightly different curves for the perceptible threshold
    - dbSPL
        - A measurement of sound pressure (SPL = sound pressure level)
    - 0 db or below means that we need less than a certain amount of pressure in order to hear
    - Why do we hear better around 3K?
        - Our communications use phonemes that put a lot of energy and discrimination around 3K
        - Resonant frequency of our auditory canal is around 3K hertz

The dBspl

- Unit preferred to measure the Sound Pressure Level
    - dBspl = 20log(P/P0)
        - P0 = minimum audible pressure for 1Khz as reference value (equivilent to .00002 pascals)
    - This usually ranges from 0 to 130
    - 0 dBspl:
        - Does not mean there is no pressure
        - There still is sound, at the threshold
    - dBs are not additive… 20db + 20 db != 40db
- The amount of pressure that is needed to be able to hear something
- We should avoid levels below 90 dBspl
- 0 dBspl is the hearing threshold

Differential Thresholds

- What is the minimum amount of change in pressure is necessary to perceive a change in sound
- At lower intensity db sounds (30-40dB) its about 1.0dB of difference
- At higher intensitiies like 50-70 you only need 0.5dB to perceive a change
    - Different frequencies give way to different thresholds

Sones

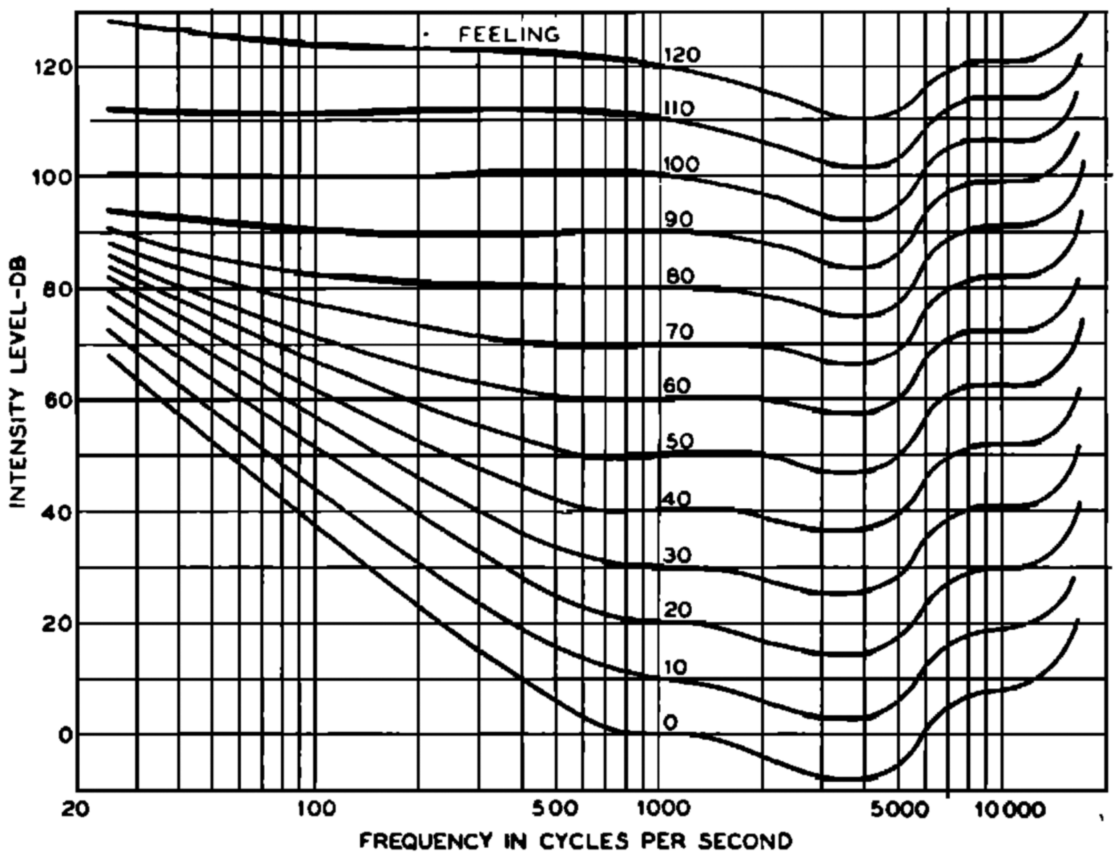
- Subjective scale for loudness
- How much louder a sound is than a nother one?
- 1 sone = loudness sensation for 1KHz @ 40dBspl
- A 10 db increase in level gives a doubling in loudness
- Even though physically doubling an intensity
    - Doubling a sound by 10db is what is necessary to hear twice as loud

Phones

- How much do we need to amplify/attenuate a tone to get equal loudness between both?
- Tones with equal loudness have the same phon level
- “Loudness” function in hi-fi amplifiers and music player presets
    - Example: we have a 1k sound at 40dm
    - Then we play a sound at 100hz and ask the person listening to the sound to change the loudness until 1k and 100hz are the same perceived loudness
    - Then we can measure how much dbSPL is added or subtracted

Phones

- How much do we need to amplify each sound to make a tone equally loud as another one



- Sones: how we PERCEIVE LOUDNESS
- Doubling SONES is equal to multiplying PHONES by a factor of 2
- Increasing 10 decibels is equal to the subjective sensation with doubling
    - (+10dBSPL)
- With two sounds we can decide if they are **isophonic**
    - If we play two sounds they are isophonic when they have the **same** loudness sensation
- This knowledge allows us to modify a sound so that all sounds are perceived as equally loud
- 40 phones = 1khz at 40 dbSPL
    - We are by default comparing to 1khz

Loudness Weighting Scales (application of the loudness curves)

- All dB should have a surname (dBSPL, dBV, dBA)
    - A fridge produces (40dbA) means that based on our hearing this is what loudness you will hear from the fridge
- Loudness weighting is more important for practical applications

        - Motorbike produces 100dB but its low frequencies
            - This would not feel the same as high frequencies at 100dB
- Weighting Curves:
    - Commonly used ones: A, B, C
    - https://noisetestingequipment.com/noise-weighting-scales/
    - Each curve has different assumptions about the way you will hear or listen to a sound and it is used to approximate loudness sensations.

Loudness and Duration

- Loudness is not dependent of other sound sensations
- Shorter sounds will give us the perception of a softer sound
    - This is because of the way firings happen in a basilar membrane
    - These firing need to be integrated
    - There is a time window that must be met for each loudness level in order to give the perception of certain loudnesses
        - A sound at a certain level of loudness must meet the window in order for the signals to be integrated
- 100 ms is the general amount of time necessary to process a sound true to the sound level it actually is

Neural Coding of Intensity

- Two ways:
    - Number of neurons are firing
        - Louder sound = more neurons are firing
        - Louder sounds will increase the amplitudes of more harmonics past the threshold to be recognized by our basilar membrane
    - Firing rate
        - At a single place when acoustic input has a small amount of dB SPL and basilar membrane moves only a little, the rate is slower than when the amplitude is higher
        - This counts the amount of spikes per second

- The firing rate hypothesis works better low-intensity sounds
- The number of neurons hypothesis is better for **mid** and **loud** tones

- The response of the basilar membrane is not linear
    - Around mid-level intensities there is squashing:
        - There are fewer responses per increase in dB
- Intensity is MOSTLY encoded in the early transmission through auditory pathways

Loudness of Complex Sounds

- Two complex sounds of the same energy level are played, one with 500, 520, 540 hz
- Another with 500, 720, 940 hz
- Will the two sounds be isophonic? (percieved at the same loudness)
- This depends on if they share the SAME critical bandwidth or not
    - If they share the same CB they pass through the same filters
    - Whole energy is not processed completely
    - **IF** they do not share the same CB, they will be processed in separate filters and they will each contribute to a louder perception
- We should consider the frequency resolution too

- Peripipheral processing (filtering according to the ouder and middle ear specificites)
- Computation of the excitation pattern considering masking effects
- Conversion of the excitation pattern into a band specific

Auditory Filters and Loudness

Approximating Loudness: LUFS

- By using the typical measurements you can measure the peak amplitude of an audio file
- RMS values: simple way to demonstrate the energy of a waveform
    - This is better than looking at the dBFS(peaks)
    - However NEITHER is better than LUFS
- Loudness unit expression that has been filtered using the K-weighting filter curve
    - Takes into account how we hear different frequencies
        - taking weight away from low frequencies
    - Measures three things: momentary, short term and integrated loudness
        - Momentary: k weighted RMS in 400 ms
        - Short term: integrated loudness in 3 second windows
        - Integrated: average loudness for a whole track
    - TV's check show, commercial, movies LUFS and adjusts the loudness so that the viewer has a seamless experience
    - The same is true for music providers
- This is the goto today
    - Low computational cost
    - Relatively simple
- Each streaming service will ask you to have an integrated LUFS level
    - Ex: spotify = -14LUFS

<u>PSYCHOPHYSICS OF THE BASIC SOUND DIMENSIONS: PITCH</u>

Pitch

- Pitch = sensations
    - Pitch /= fundamental frequency (f0)
    - “That attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high”
- F0 → pitch , periodicity → pitch
    - These are not direct linear correlations

Pitch from a noisy source?

(all info in slide)

Pitch: Different Waves, Different Spectra..

- …but the same pitch sensation?
- If the wave form changes our perception of the timbre changes
    - But the length of the period is the same
    - If periodicity does not change
- With sinusoids:
    - One place is excited, and from this place spikes are sent telling you that a 260hz sound has been played
- With multiple places in the basiliar membrane providing spikes, the information is combined and periodicity is used to determine pitch…
    - As such, the fundamental frequency is not necessary to determine pitch

Frequency Perception: Absolute Thresholds

- 20hz-20k (in young, intact hearing systems. 16khz in adults)
- HF threshold permanent decrease due to:
    - Ototoxic substances: aspirin, quinine
        - Pain killers can harm your hearing!
    - Exposition to high levels of sound pressures for long periods of time

Head Size (Cochlear volume) and threshold for frequency

(all info in slide)

Dimensions of the pitch sensation: chroma and height

- Pitch sensation is more complex than just using f0
  - Chroma:
    - If we play two C’s, we can perceive some similarity between them
    - These two notes share **chroma**:
      - The property of being multiples of themselves

Dimensions of the pitch sensation: Strength

- Determined by:
  - **Microstructure of the waveform**… the simpler the waveform, the stronger the pitch
    - Ripples in the waveform make the pitch less clear
  - **Amplitude**:
    - When given sounds with amplitude modulation:
      - The pitch blurs.. Somehow
    - Band noise:
      - If we have a narrow band we will start to perceive pitch:
        - This is as opposed to white noise

Pitch Sensations Thresholds

- Region for pitch: 30hZ-5000hz
  - This is based on the physiology of the neural transmission
  - Hard to reliably and quickly transmit frequencies above this threshold
  - Because of the unreliability of pitch sensations above 5k, we don’t have many instruments that go above this

Pitch Differential Threshold (JND)

- High frequency tones: JND = 10 cents
- LF tones: around 40 cents
- MF tones: around 4 cents

Percievable “Quanta”

- There are regions/combinations of pitches and sound pressure levels that will give way to equal sensations of pitch
  - We can break this map of freq/amplitude into regions of sounds that are perceived the same way

Duration effects on pitch

- Minimum cycles or minimum duration?
- When we judge very short sounds our pitch sensation becomes very problematic
- Low and very high freq tones need a longer duration to perceive pitch
- Could have to do with the length of the periods involved
- But also with limitation in speed of neural spikes that can be generated

Loudness + Pitch

- Depending on freq and sound pressure as slightly flat or sharp
- At high-pressure levels, high freq is perceived as sharp, low as flat
- Low pressure levels is the opposite effect

Mel: Subjective Scale for Pitch

- Mel filters (?)
- Used in signal processing to build filters that approximate our frequency resolution (MFCC)!!
- Arbitrary reference hertz: 1k at 40 dbSPL = 1000 mels
    - We need to present a little more than 2k to get a sensation of twice as high in pitch

Absolute Pitch

- Skill has been learned
- Hearing one musical note and being able to identify the note
- Two side:
    - How fundamental frequencies are exposed to you
        - Brain functions as a pattern recognition device
    - You need to pair labels with sensations
- Pitch chroma extraction:
    - Pitch height + pitch chroma extraction
- Verbal learning

Factors of AP

- Some genetic factors:
    - Chromosome 8
- Age:
    - Starting formal training in music affects how well you can identify pitch
- Mother tongue
    - The language you were born and raised in helps to develop AP
    - Tonal languages

        - Languages where meaning is derived from changes in pitch
        - Ex: Mandarin has 4 pitches used to convey different meaning

Mechanisms for the pitch sensation of pure tones

- Two explanations proposed:
    1. Place Theory:
        a. Pitch sensation comes from place in basilar membrane where basilar membrane is bending the most. At that point neurons start train of spikes that are most well organized and tell the brain we are neurons detecting 100 hz for example
    2. Timing Theories
        a. Combination of trains of spikes that are generated by different neurons
            i. Wave generates phase locked spikes at certain point in time
            ii. Overall distribution of spikes follows a certain periodicity
            iii. When we combine spikes into an integrator we observe overall when we sample activation coming from neurons coming from auditory nerve, we get equal separation that tells about the length of the period of a wave which can be used to recover pitch information

The problem in Place Theory: Missing Fundamental

- The pitch of a tone having partials at 300, 400, 500, 600 will be 100 hz even though there is no energy present there

Timming Theory: Phase Locking

- When we combine activity of neurons we can get an estimation of the periodicity and use this to extract the pitch sensation

Timing Theory: Unresolved Harmonics in Critical Bands

- Beating in HF auditory filters follows the periodicity of the fundamental
- Looking at the beating in auditory filters allows the periodicity to be computed
- Timing Theories:
    - This contains most of what we need to explain most of the pitch sensation

Two Operating Mechanisms

Two Complementary Mechanisms: Place + Periodicity

PSYCHOPHYSICS OF THE BASIC SOUND DIMENSIONS: TIMBRE

TIMBRE
- Multidimensional
- Waveform Shape has a big effect on the timbre

Defining Attributes of Timbre
(all material in slides)

Timbre = Spectrum?
- Change amplitudes of spectral components changes the timbre sensation
- These do not yield realistic sounds
    - The attacks and changes in amplitude are very important to the sound
    - You can create an example of an acoustic instrument from one idea of this but not perfect

Envelopes: Temporal Factor
- ADSR
- Some sounds don't have sustain
- Partials of a trumpet:
    - Different partials have different ADSR envelopes

Attack
- Contains most source information
- Most noisey,
- Different attacks = different timbre sensation
- The same

Formants
- Regions of spectrum which contain harmonics with a higher amplitude
- Can affect the timbre perception
- Every language chooses the position of vowels in a different way
- Can be plotted using measurement of tongue position and mouth opening
    - Makes learning a language hard if it does not contain neutral vowels
- Singer formant: mostly achieved by males singers
    - Having to sing against an orchestra: needing to be heard… how?
    - Put spectral energy in a region where orchestra does not have as much energy
    - This is what creates the "opera singer voice"

Combination of Tones: Tartini Tones

- At loud and very loud levels two tones close in frequency may create a ghost tone corresponding to the sum and difference of the frequencies. Other extra tones may also appear, all of them obeying to a non linear distortion process happening in the middle-ear or/and in the early auditory processing

Combination of tones: Beating-Roughness

(information already presented above in frequency/pitch notes)

Combination of Tones: Sensory Consonance-dissonance

- Sounds will be most dissonance around 0.25 of the critical bandwidth
- Intervals in music:
    - Subjective judgements of dissonance, easiness, liking/disliking of notes
    - We obtain representations close to what music theory suggests
    - conscience → dissonance: 1:1 (octave) 1:2 (fifth) 2:3 major sixth… 3:4, 4:5, 5:6 ect.
    - This is called **sensory consonance**
        - Can be attributed mostly to how our hearing circuits work, not our culture
        - Sensory consonance can be combined with harmony (tonal context of the music)
            - When the tones and harmony are established by the song and we hear a different set of tones IN the context of a certain harmony
                - This is how musical dissonance arises

Roughness

- This is a subjective sensation with a physical correlate
- Subjective scale: **asper**

Fluctuation Strength

- Measured in **vacils**
- 1 vacil = 1kHz tone fully modulated at 4 hz at a level of 60 phon
    - Fast beating (>20hz) → roughness

Sharpness

- Sharpness = brightness sensation related to position of the spectral centroid, the gravity centre of the spectrum
- Measured in acums

Sound Description

- How can listeners describe the sound? What semantic representation is created by the user?
- Mostly done with likert scales
- OR: paired differentials: list of descriptor words with opposites on a spectrum
    - You choose where a sound sits on the spectrum

Timbre Spaces

- Three dimensional space
    - 1. Attack time (slow / fast)
    - 2. Spectral centroid (lower or higher)
    - 3. Spectral Flux (vairability of the sound as it is played) (stable: greater or lesser)

PERCEPTUAL ORGANIZATION

Sorry, I missed all the notes from this lesson before Figure-Ground :/

Figure Ground Phenomenon

- Listeners attn usually drawn to aspects of sound that are changing
    - Changing sound = figure
    - Stable sound = ground
- Scrambled melodies?
    - Easier to hear them when:
        - Different instruments with a different timbre
        - Different volumes
        - More separation between the pitches
            - **Pitch, loudness, timbre**

Grouping Heuristics

- How can we link together outputs of analysis of timbre, pitch, loudness, location?
- This can be combined via an automatic process
- We have heuristics to help find a solution to a problem but they won't always work
    - **Simultaneous grouping:**
        - We consider audio input as continuous in time
        - Our audio unit is continuously parsing the output
        - Different elements of output are grouped together and one object emerges
            - Example: traffic jam, student shouting in the hall
        - This has to do with analyzing simultaneous spectral information and deciding what goes with what from the output
    - **Temporal Grouping Heuristics**
        - Make it possible to connect one temporal point with the next
        - Something that happened one moment before is still happening
        - Or, changes in the information are changes
            - We are listening to a constant drone, suddenly more energy in freq, or amplitude, we consider this to be one new stream
            - Our brain decides where in spectrum it is located, how it evolves
                - This is based on temporal proximity
                - Spectral shape
                - Brightness
                - Timbre
                - Intensity, including onset and offset
    - Heuristics are based on research done based on the gestalt school
        - Mostly studied visual processing but did a little bit of studying auditory processing as well

Proximity Law

- Stream segregation depends on proximities between them (temporal, pitch, etc)
    - Example: alternating high and low tone melody:
        - We perceive the high notes as one stream and the low notes as part of another
        - This depends on the speed though. A fast enough stream in necessary to get the perception
- When this happens, we can't pay attention to both streams at once
- Pitch proximity: when the pitches are closer together we hear the galloping effect and further away we separate into two streams
- **Bi stability:** an ambiguous pattern that can be coded by two approximations, your brain first hears one way and over time discovers the other
    - Example: hearing two sounds alternating quickly, first you hear the gallop sound, but your brain eventually changes what note you perceive as coming first and the gallop effect is lost

Wessel's Illusion

Common Fate

- Components in sound act together, they start and finish together, or change pitch / intensity together
- Example: vibrato (frequency modulation)
    - This means that your sound / overtones will modulate against a static frequency
        - This creates a new stream which has a common fate that can be followed separately

Closure

- Source may be obscure or absent but we still perceive it
    - Example: drums happen in the middle of a long note don't stop our perception of the whole long note playing
    - **Phoneme restoration:** hearing a sentence spoken while somebody coughs. The coughing breaks up the sound of the sentence but we are still able to perceive the sentence as a whole

MUSICAL AND MEMORY

Recognition of Familiar Melodie

- Can we identify a musical piece with one note?

Simultaneous activations of different "memories" when listening to music

- Some areas are activated by familiarity
    - I have heard this music before
- Some areas respond to autobiographical memories
    - This reminds me of
- Some areas respond to tunes that person enjoys
    - This song makes feel good chemicals

Multi-Storage View of Memory

(all information on slides)

Functional View of Memory Systems

- Echoic Memory: this is where feature extraction happens
    - Also where perceptual binding happens: where we decide what are the streams we need to track or discover
    - This has very fast decay (forgotten very quickly, less than 4 seconds)
    - Either you move this to STM or traces are rewritten as new incoming information
- STM:
    - Segmentation & chunking
        - Break into meaningful parts
        - Can have interference with other traces of memory
- LTM:
    - Explicit learning (permanent)
    - We you explicit rehearsal to encode the information
    - Implicit knowledge (muscle memory, early memory associations, things we cannot explain or express using language) w

Iconic Memory

Echoic or Sensory Memory

Short Term Memory

- **Segmenting**: continuous musical texture is broken into smaller segments
- **Phrasing**: the expression of a meaning of a unit
  - The way notes are jointed into a meaningful unit
    - Example in speach: how you add pauses and what speed you speak can change what you are saying / how you understand it
- **Grouping**
  - If we can make some similarities or differences prevalent we can help encode more efficiently

- **Chunking**
  - Memorize this seris of letters: F-B-I-C-I-A-U-S-A-C-N-N-I-B-M

Serial Position Effect

Pitch Specific Short Term Memory

- Experiment:
  - Play a sequence of notes, task for subjects is to pay attention to the first one. After sequence, another note is presented, then ask if the last presented note was the same as the first one
  - Idea was to fill up short term memory
  - Then compared with distractors as other acoustic information (speech)
  - Found that the musical acoustic information gave way to more errors than the speech

Memory Consolidation

- Waking: acquisition, early and ongoing consolidation
- Slow wave sleep: hippocampal signaling (NREM Sleep)
  - Transfers informationt to neocortical cell groups
  - This means that the hippocampus reherases what we learned during the day

Declarative Memory (Explicit Memory)

- Name of a song?
- Episodic Memory: memory of a specific moment in time, like where and when we learned about a piece of music
- Semantic Memory: memory about things of the world: knowing that Rite of Spring is a song written by a composer

BOTH Declarative and Non-Declaritive are associate and distributed

-we learn based on association and learning does no happen in one place in the brain

Autobiographical Memories

- Music is linked to autobio memes
- It can be helpful for studying patients with Alzheimer's
    - Not true in very advanced cases
- Important to consider **autobiographical bumps:**
    - Humans show certain preferences for stimuli experienced at particular ages
    - Example: music between 15-20 years old leaves a stronger/deeper impact long term

Non-Declarative (Procedural Memory)

- How does a melody go? I can play it without saying it because I automatically know it
- Procedural knowledge: learning by doing
- Grammar is a good example. You learn and internalize this sequential knowledge.
- This type of learning is difficult to retrieve verbally, often causing interference
    - These ideas are difficult to express with language
- After some aount of compilation of knowledge, big chunks are created that are not accessible to introspection
    - Doing a golf swing and all the steps involved, you cant think really about a specific small part

Implicit Musical Knowledge

- Even without musical training we pick up on musical patterns
- We extract them from exposure alone
- Knowledgeable listeners even from other cultures can pick up on cues prvoded a certain amount of exposure

Types of Retrieval

- Recognition: acknowledgment that a pattern in STM is stored in LTM
    - Ive heard this before
- Recall: activation of LTM encoded pattern by a diffuse effort of will
    - I recognize this, who wrote it?
- Reminding: which memories are brought up by this song?

Phrasing

Themes, Variations and Motives

- These work to help us increase the memorability of musical material
- This is beneficial sometimes for the approval of the audience

Neural Darwinism

- We have a view of the world that is very subjective based on our understanding of the world
- Some events and categories are not toally exclusive
- Our personal taxonomy does not line up with the laws of physics
- Things can fit into multiple categories and we may not agree about what should be catagorized where

THE WHEN/WHAT HIERARCHIES

sorry I missed everything before this :/

When What Frameworks

- They frame our listening experience: when we can expect something, what to be expected
  - When: pulse/tempo/beat… how they help to anticipate future musical events
  - What: tonality (esp. In western music)... how this provides meaning to the music
    - Ex: scales, notes, consonance
- Synchopation example: we hear a clicks that quickly increase in tempo
  - As a certain point we perceive this as a note. There is a threshold that is important for this
    - We can see that the difference between rhythm and tonality is only a matter of time
    - We can create different intervals depending on the syncopation
      - 5:4 becomes a perfect 4th
      - 6:5 becomes a major chord

Chills

- Not everybody can experience chills
- People into classical vs not into classical:
  - Experience with a type of music allows you to feel more chills
  - Experience also predicts more intense chills
  - People with experience in classical music could also **anticipate** the climax seconds before reaching it
- Chills happen in concordance with:
  - Skin conductance
  - Respiration: heart rate
- Involves both the caudate nucleus and the nucleus accumbens

Music is a matter of Time… Time is a Matter of Music

- Different musical events need different amounts of time in order to be processed
  - In order:
    - Pitch
    - Phonemes
    - words/syllables
    - Beat perception, locomotion, phrases
    - Musical phrase sentences

Temporal Scales of Musical Events

IOI: Inter Onset Interval

- The relationship between one onset and the next one
- Our brain detects changes in different dimensions (amplidute, pitch, spectral energy, etc.)
- All these changes can be considered “onsets”
- A big change in amplitude is considered an onset but is not the only change in property that can be perceived
- An onset is an **instant** not a **duration**
- Between onsets we have events:
    - Events have certain loudness, pitch, timbre sensations
- By analyzing IOI’s from musical excerpts we see many have simple relationships
    - Usually the relationships consist of very simple ratios

Musical Speed Limits

- The separation between onsets makes it possible to perceive melodies
- The same melody if played very fast or slow:
    - Fast: cannot detect all the pitches/onsets sufficiently
    - Slow: short term memory is filled with 2-3 notes… not enough to store the sound of a full melody

Beat/Tempo

- From the detection of onsets we can deduce the beat
- The beat is in our brain
    - From incoming information our brain decides where/when/how long a beat is
- We are detecting a basic recurrent pulse that accompanies the music
- After hearing the beat for a few seconds we are able to anticipate future events
- We have a preference for tempos around 100 BPM

Entraining

- SOME but not all mammals are able to internalize rhythm
- Beat induction can be observed in some non human species, such as birds
- This is not observed in closely related animals to humans
    - Likely because they have no sonic system

Beat Induction

Rhythm Hierarchies/ Beat Inudction

- General agreement about most appropriate times to tap or clap
- Beat allows us to anticipate future events

        - When we are given a certain beat that helps to anticipate where events happen, we can anticipate events better if the events continue the beat sequence that was previously established

Times(s) and Tempo

- When we combine BMP's we enter the world of tempos
- Humans prefer BPM between 60-120
    - They entrain more easily
- We can accept tempos from 30-240 BPMs

Tempo JND

- Around 10% change for normal listeners

Accent

- With a series of stimuli we can give a special mark to certain events
- Accent does not necessarily mean adding more energy
    - It is just a way to mark a special event as special
    - This breaks the equality of the events

Accent and Meter

Influences of Native Language on Segmentation/ Meter Strategies

Metrical Hierarchy

- The marking of strong / weak is the typical way to consider different moments
    - The first beat is the strongest: often starts a phrase
        - Third beat is second strongest
        - 2nd and 4th next
        - Then the offbeats
    - The music dictates where the strength is
    - The weaker the beat, the more tension we wexperience when an accent is placed on it

Expressive Timing and Expectation

- Timing and meter can be changed at the micro level
    - Good performers can play with this… allowing for expressive timing
- Listeners learn expressive timing is different for different genres

Exposure Dependent Sensitiveity to Timing Alterations

Rythmic Pattern

Rythym without Beat and Meter?

3 Dimensions of Rhythmic Semantic Experience

THE WHAT

- Pitch information
- We experience constancy in properties
  - Even though we change the angle at wich we see perfe, he is still the same person
  - This categorization can be observed when we deal with music
- Can be observed in brain
  - Looking at what brain areas are active based on what type of images/objects/sounds are presentsed
  - The brain does refined caetgorizations
    - Example: living/non living things (one of first categories learned by babies)
    - Human vs non human
    - Music vs speech (this requires more time as opposed to the previous two learned in infancy)

Knowledge Structures: Categories

Explicit Musical Categories

- Notes: discretizing frequencies
- Etc.
- Some categories can be overwritten… music can be created without these categorizations

Melodic Categories: Tuning Systems, Scales Pitches

- Different cultures create different systems to allow or disallow certain frequencies
- Usually simple ratios are preferred when selecting frequencies to make musical notes
- Once the tuning system available we can select a few freqs to become a scale
  - Usually involve few notes/freqs (5-7)

I missed everything after this :/

<u>MUSIC AND EMOTIONS</u>

MUSIC AND EMOTIONS

- In order to create expressions of different emotions we manipulated two main muscles
    - Auricularfrontal (?): forehead
    - Zygomatic: mouth/lips
- Most authors agree: all humans share some universal emotions
    - However much of the research on this is biased
- Darwin: first to study human emotions
- **Cultural differences:** each culture has a different type of emotional expressions that are not able to be translated into our language

What is an Emotion?

- Emotions are linked to survival behaviors
- Involved in expressing, detecting, and announcing danger
- Engaged in competition, loss, and cooperation

Communicating vs. Generating Emotions

- If interested in doing research on music/emotions you must immediately identify WHICH of two perspectives on emotion you are researching
    - Music is a way to communicate an emotional state
        - We learn by exposure / making mistakes the way to decode emotional interpretation from speech
        - We also apply this when we listen to music. We can listen to any type of music and make this decision
    - Music is a brain changer: a way to modify our mood / emotions
        - We play/decide/select certain music in order to assist with things we do throughout the day.
            - For the morning, the night, to forget something sad
        - Music is a stimulation that can change our emotional state
        - This does not have to be two ways: you can hear sad music and not be sad after
        - This field is a little more subjective: will a piece of music change two different people in the same way?
            - Studies show a lot of individual difference in results. There is no universality.
        - Ex: music recommender based on emotional content? Won't work on everybody the same.
            - How can we personalize this

Modes and Emotions

- Historically we think that certain modes/intervals are associated with certain emotions
- This comes from the combined work of centuries from musicologists, philosophers, etc over centuries and still find this is a matter of opinion

Affective Phenomena

- Emotions are a part of larger constellation of subjective phenomena that are relevant to music/music tech
    - Preferences: more general than emotions
        - frameworks/results from a generic aspect or specific items
        - I can show the world what I like and dont like
    - Usually modeled/tuned/developed across life
    - They have strong connections with personality traits, attitudes/ways ot seeing the world, and interpersonal stances
    - Preferences can influence our emotions responses
- Moods can change the state of music understanding
    - These can be short or long duration: a few minutes or all day?
    - These are expressed by emotions that are shown
    - **Mood: LONGER EMOTIONAL STATE**
    - **Emotion: MORE SPECIFIC, WE CAN'T FEEL THEM FOR A WHOLE DAY**

What is an Emotion?

- Complex phenomenon
- Involves interaction between different laywers/physiological structures in brain
    - Involves:
        - Affective experiences: we feel okay, more activated or less activated
            - This is an evaluative component: can be expressed by a positive or negative measure
        - Cognitive processes: when experiencing emotions, my thoughts can go in a certain direction. I can attend more deeply to a certain stimuli in a situation
        - Physiology: hard to control consciously, occur due to the chemicals that emotions release, and can be measured very easily
        - Behavior: approaching, running away from, increasing exposure… depending on what we feel our behavior will be a consequence of this

The Limbic System (Amygdala, Hippocampus, Hypothalamus)

- Amygdala: survival, emotion behavior, fear, excitement
- Hypothalamus: hormonal release, physiological feelings of emotions. Release and regualtionof hormones
- Hippocampus: explicit memory, links us to past experiences

- NOT in the limbic system but important:
    - Auditory coretex: temporal lobe is very close to the hippocampus so it is easy for the coretext to activate memories based on sound/music
    - Nucleus accumbens: this is important for the reward system
        - In any situation where we need to reward the repetitions of a behavior we can use this
    - Caudate Neucleus: this anticipates the reward, works with the neucleus accumbens
    - Prefrontal coretex: modulates our action plans/behavior
- There are two different systems in our emotional behavior
    - Fast system:
        - Includes most basic structures that we share with animals
            - Limbic System: these structures give us the ability to make quic reactions
    - Slow system: prefrontal cortex… can regulate and slow down the most "animal" reactions we would show in certain situations

Why does music convey (express) emotions?
- split up music as a way to express and to generate emotions
- We can often hear emotion based on its resemblance to speech
    - This comes just from the tone and rythm of speech
- We use **connotations** to express certain types fo emotions
    - For example: instrumentation, type of tonality etc.

Musical Features Used to Express Emotions
- The weight of a certain descriptor word is very important for sound classification

Which of 4 Emotions was expressed
1. Neutral
2. Happiness
3. Tenderness
4. Anger

Dimensional Representation with Determining Features
- 4 way map of emotions (two dimensional)
    - Activity scale and valance
        - Activity: related to the events that happen in music. How many things happen per second? How is the tempo? How many voices?
        - Valance: subjective appreciation of the communicative content

Secondary Emotional Dimension Map

- Russell's Circumplex Model
  - Organizes emotions based on their similarity to each other, within regions
    - There are clusters of related emotions
    - Generally, the two dimensions are:
      - Activity level
      - Positive vs negative emotions
- **Hevner's Categorical Model**
  - Circular model that groups together specific models

Emotion in Time

- Music related emotions require time to be developed, perceived, felt
- 15-20 seconds seems to be enough to represent this
- Study: when listeners used a valence tracking software

Why does Music INDUCE emotions?

- **Musical Expectation -** Associations we make in our frontal lobe, associated with chemical activity
- **Arousal** - because we listen to something we consider interesting, this also affects our inner chemicals
- **Mood Contagion** - our moods are effected by the moods of people surrounding us
- **Associations** - linking between special situations in our lives and music… because of this our brain automatically computes concurrent covariations, without paying attention we hear a highly associated music and we will react in a similar way
- **Imagery -** its not just the music, the music can activate our occipital area, can help us to recover, part of autobiographical memories helps us to imagine places

Violations of Musical Regularities Elicit Emotiional Responses

- Beethoven music, measured by listeners wearing skin conductance responses
  - Unexpected music causes big change in SC

Surprise vs Uncertainty

- The most interesting music with the strongest reactions is that which has uncertainty and surprise
  - Surprise: event that is not happening according to expectations
  - Uncertainty: there is no solid context to even base expectations in