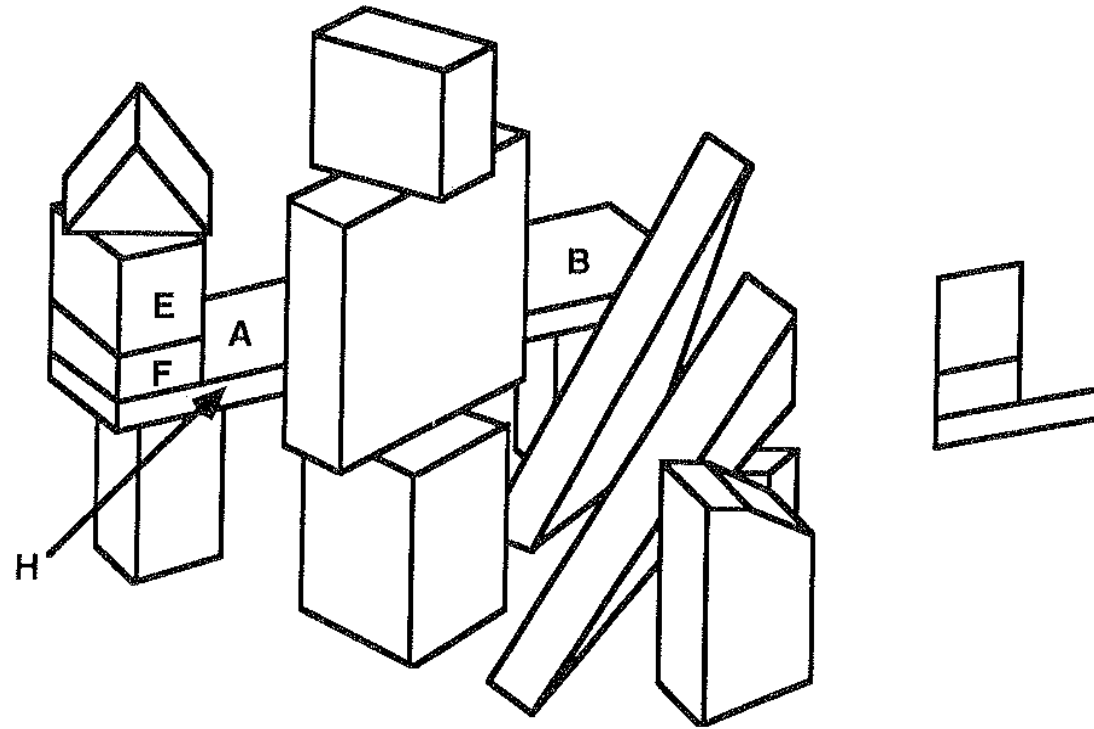


Topic 1

Auditory Scene Analysis

What is Scene Analysis?



(from Bregman's ASA book, Figure 1.2)

Auditory Scene Analysis



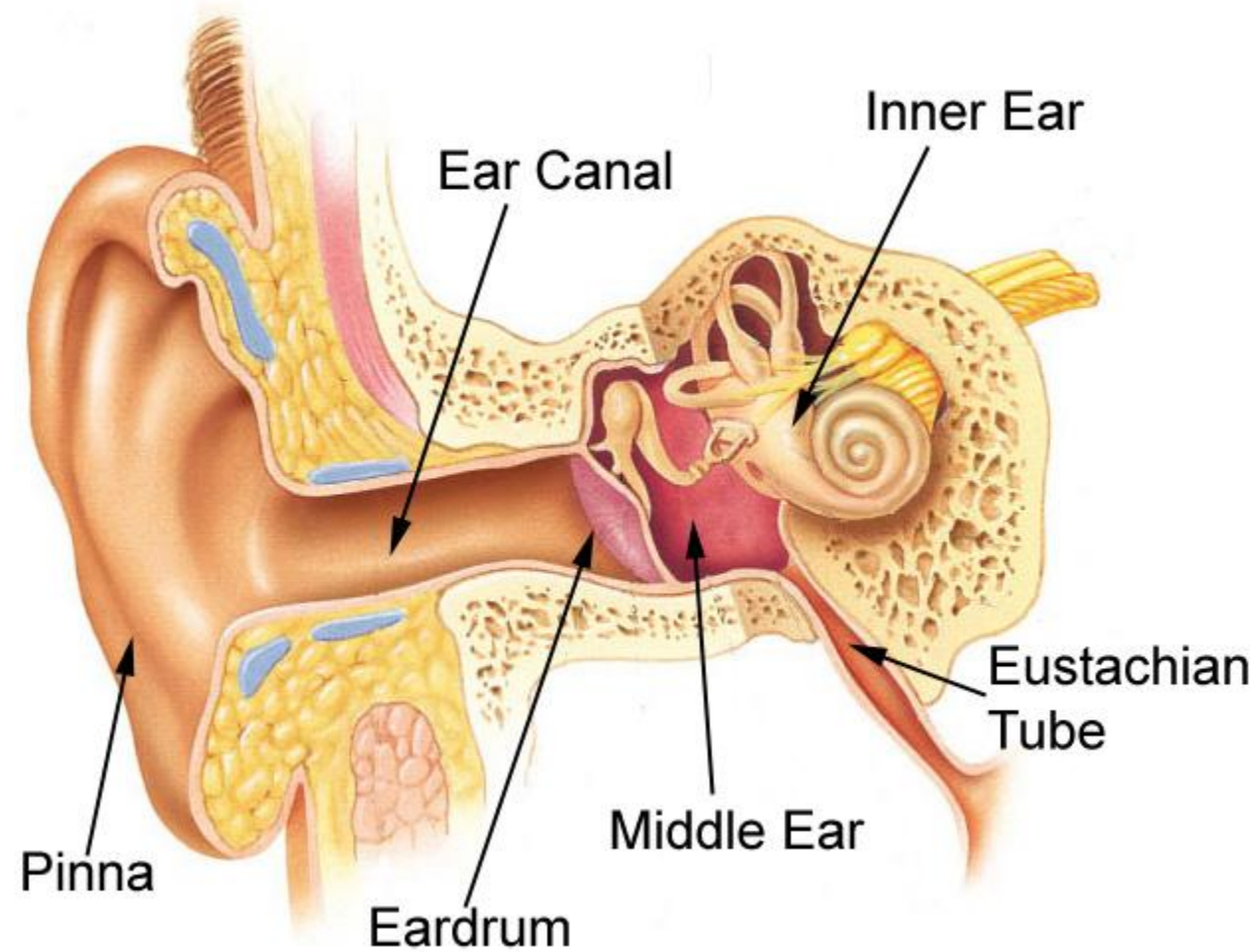
The cocktail party problem

(From <http://www.justellus.com/>)

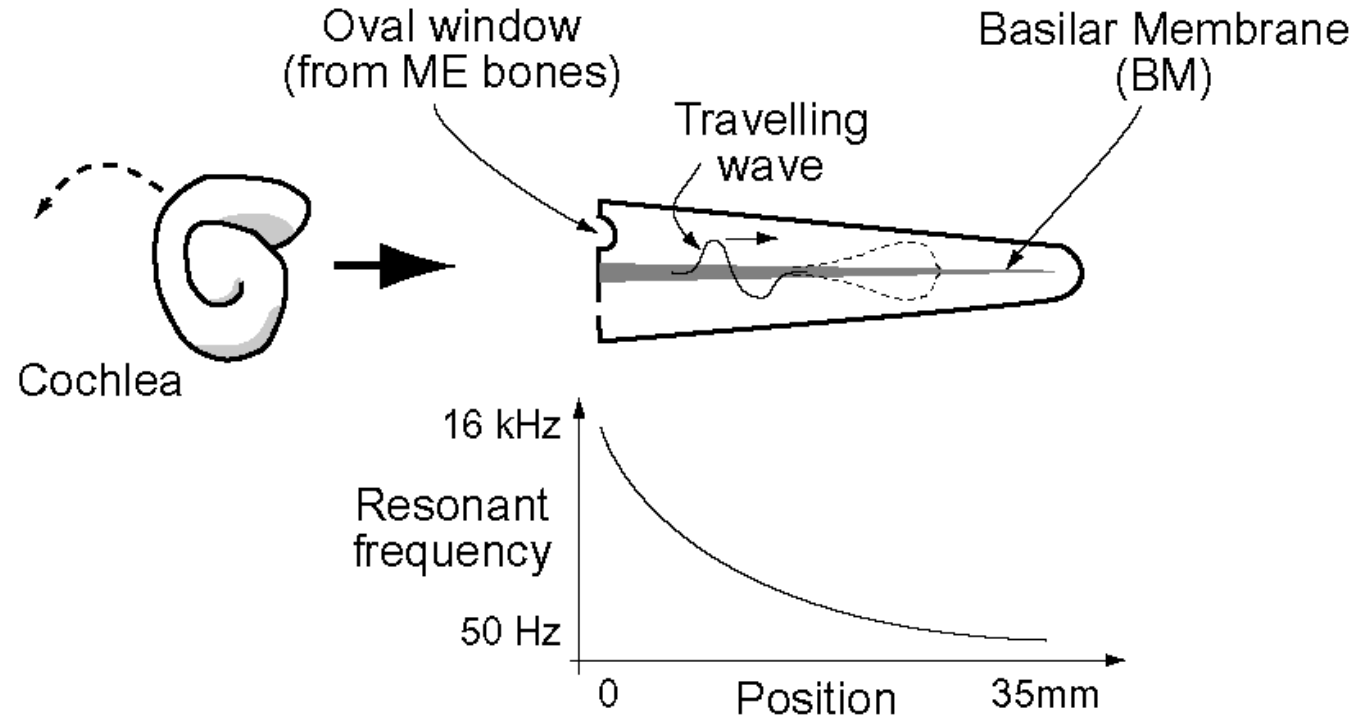
It's very difficult!



The Ear



The Cochlea



- Each point on the basilar membrane resonates to a particular frequency
- At the resonance point, the membrane moves

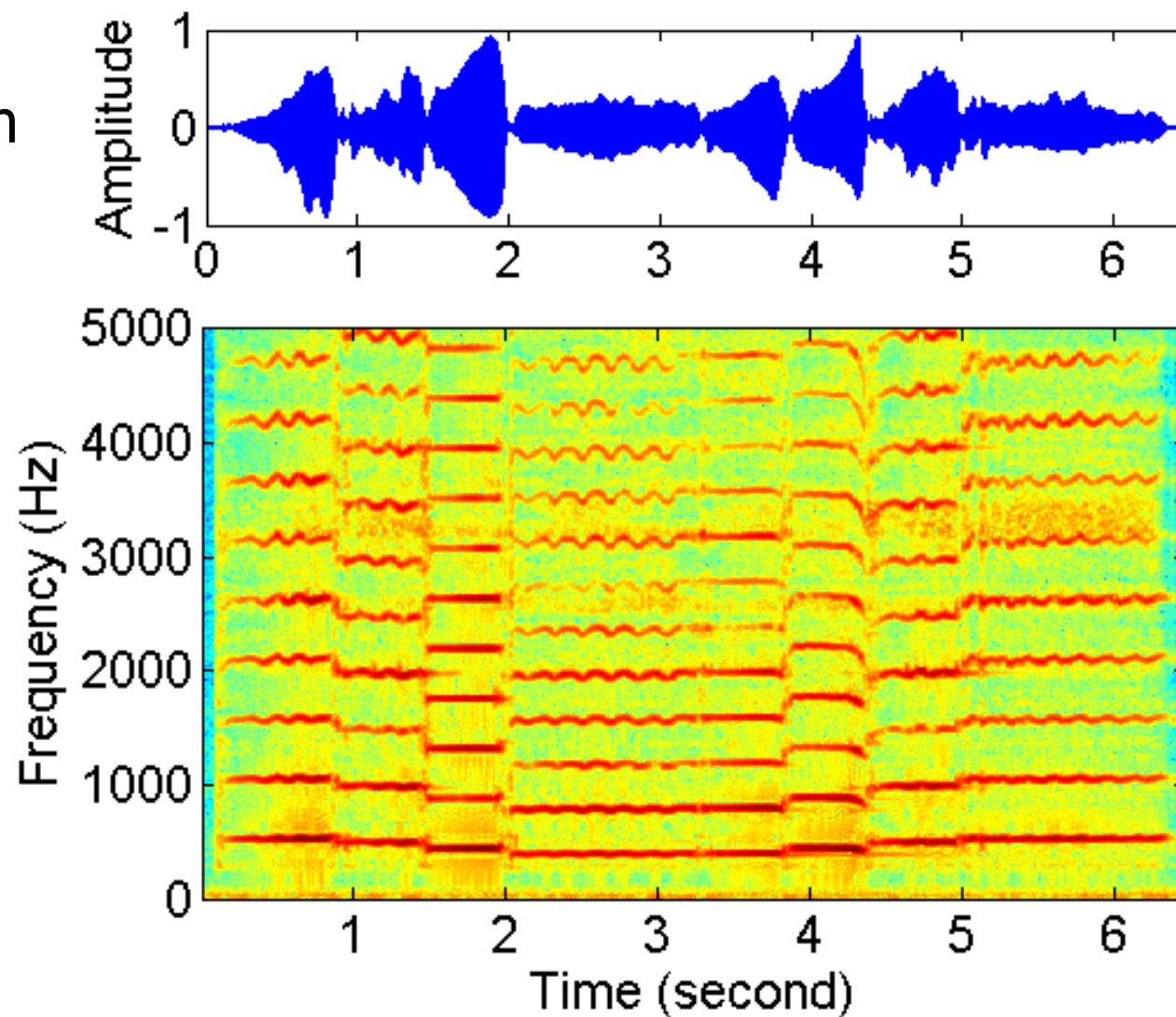
A Movie!



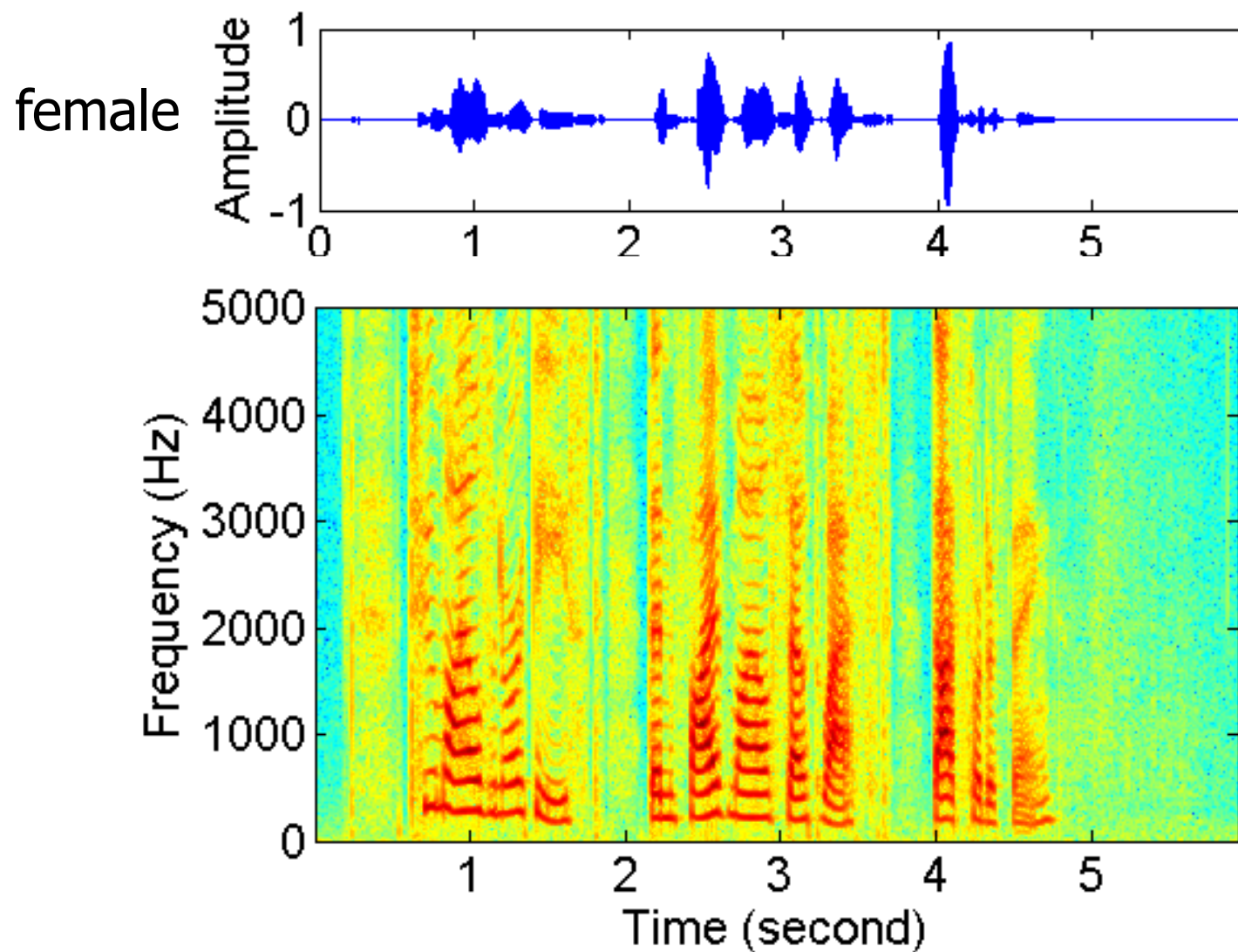
(thanks to Howard Hughes Medical Institute)

Spectrogram

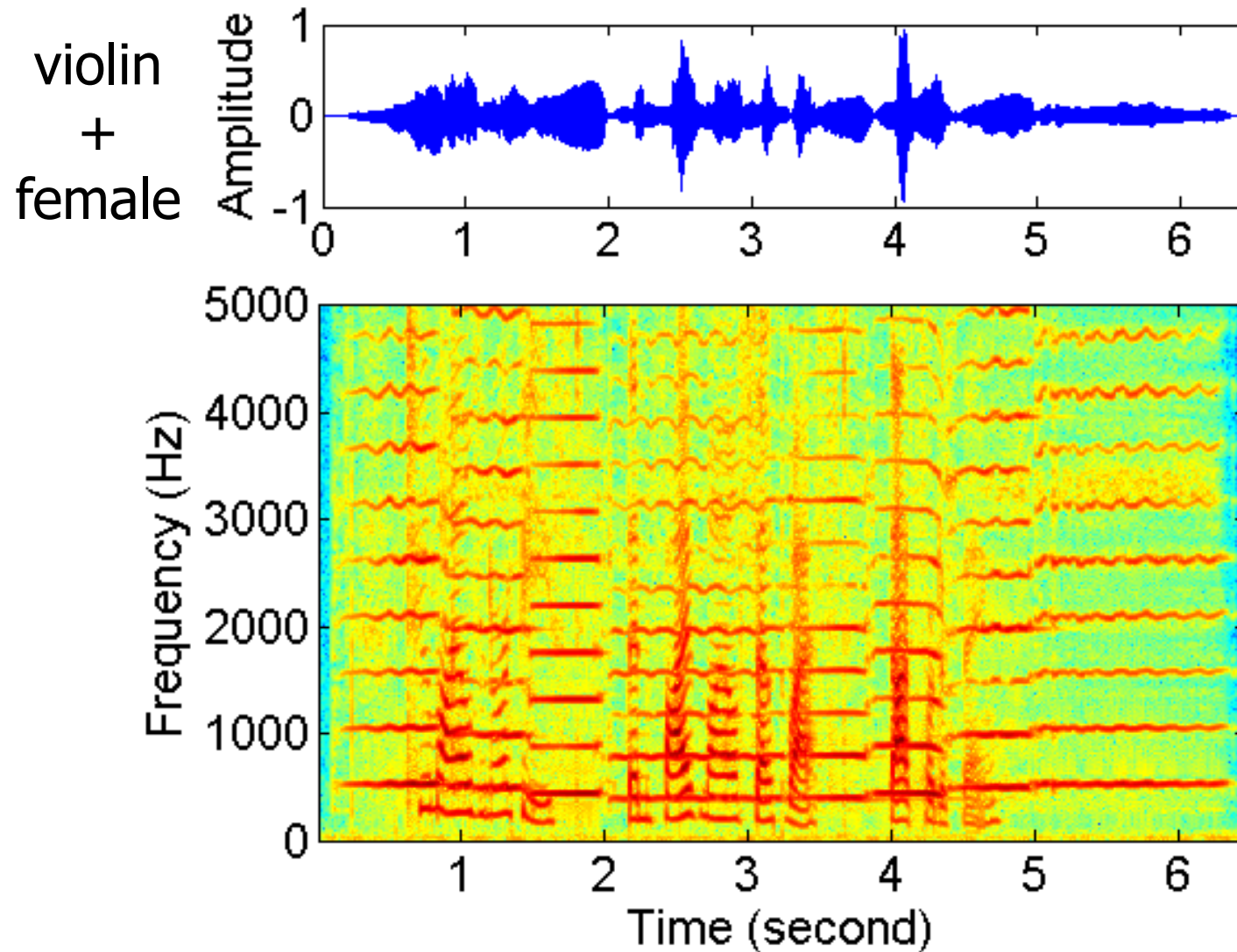
violin



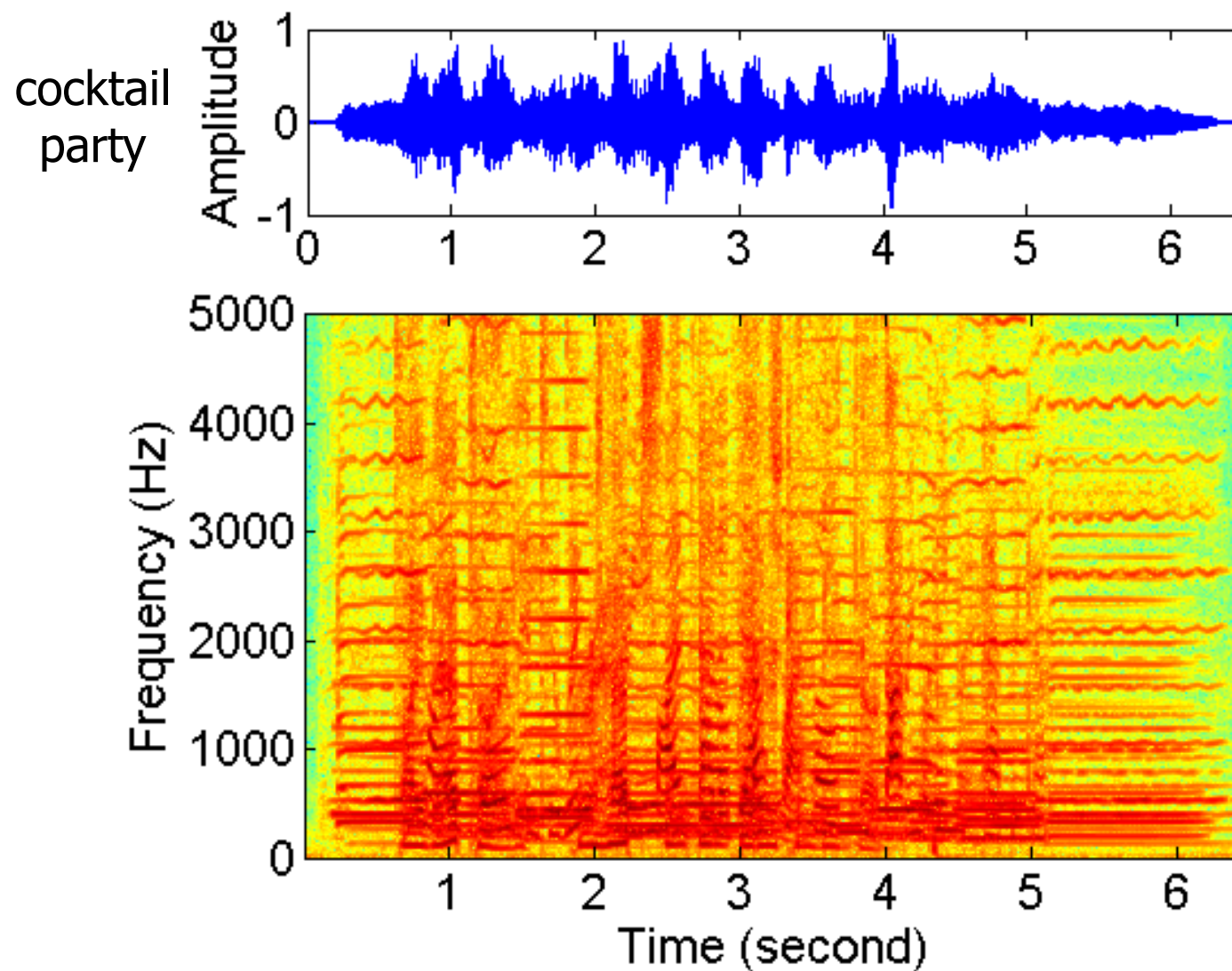
Spectrogram



If they sound together



How about this?



Auditory Scene Analysis

- Studies the mechanism of the human auditory system to answer questions like
 - How many sources at a time?
 - Which frequency components belong to the same source?
 - How does a source evolve?
 - Where are the sources?

Vision vs. Audition

- Visual scenes mainly describe objects that **reflect** light
 - Shape, color, brightness, texture, etc.
- Auditory scenes mainly describe sources that **emit** sound
 - Time, frequency, loudness, location, etc.
- Visual objects occlude; auditory objects overlap

Analyzing auditory scenes is like...

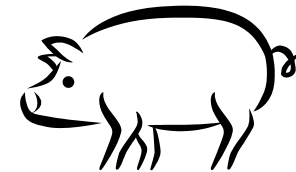
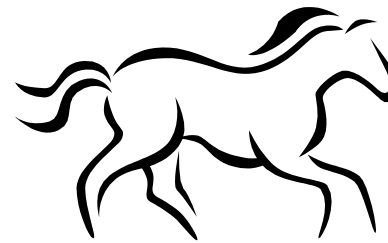
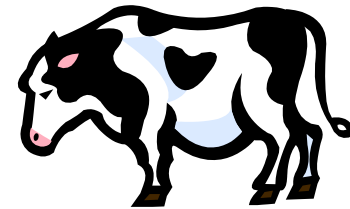
Analyzing visual scenes where

- Objects are half-transparent
- Objects change transparency
- Objects disappear and reappear unexpectedly

Two miles northeast, then five miles southwest -- that sort of thing.

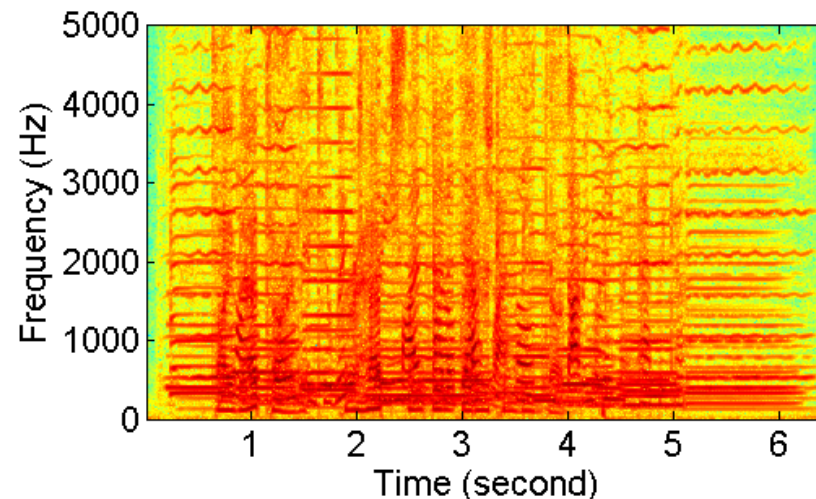
Fold into whipped cream and add a dash of salt and sprinkling of paprika.

By that time, perhaps something better can be done.

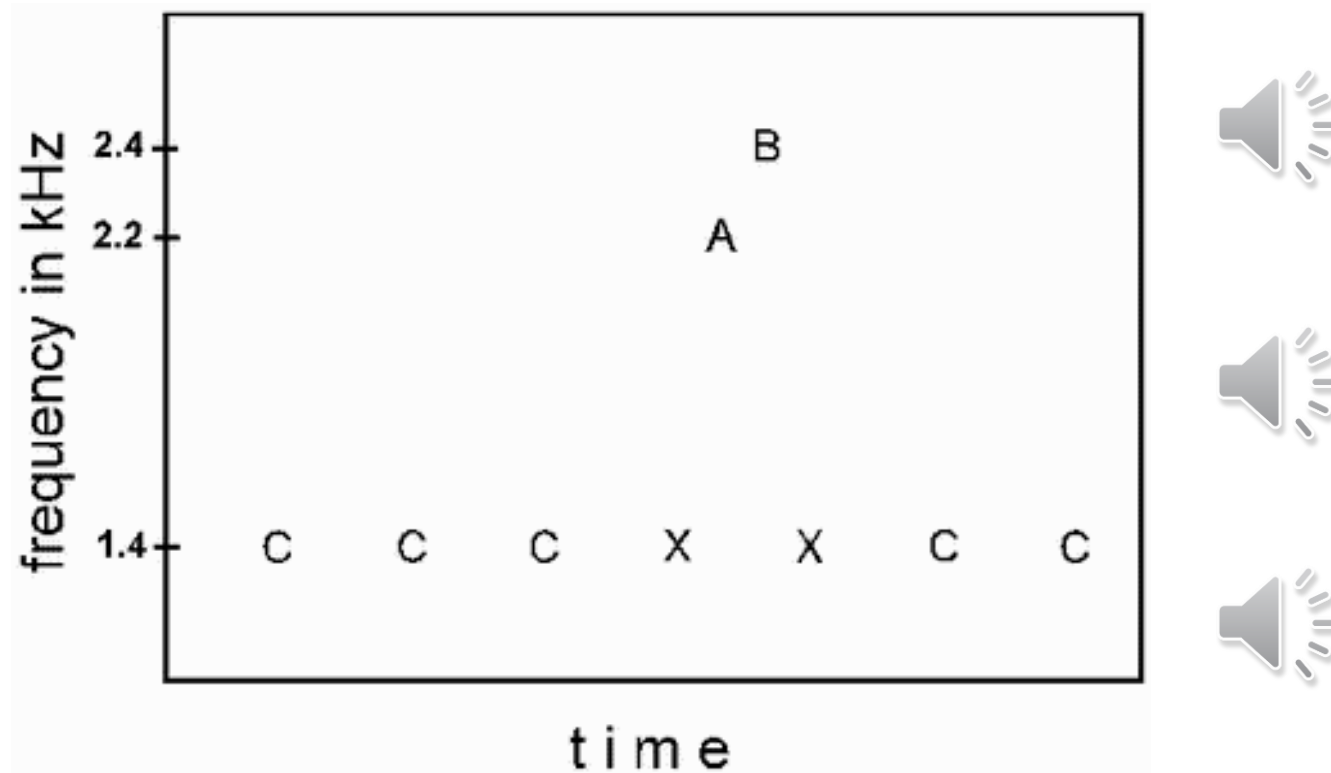


The Analysis-Synthesis Process

- Decompose the acoustic scene into a collection of segments
- Group segments into streams
 - Simultaneous vs. sequential
 - This is the main concern of ASA

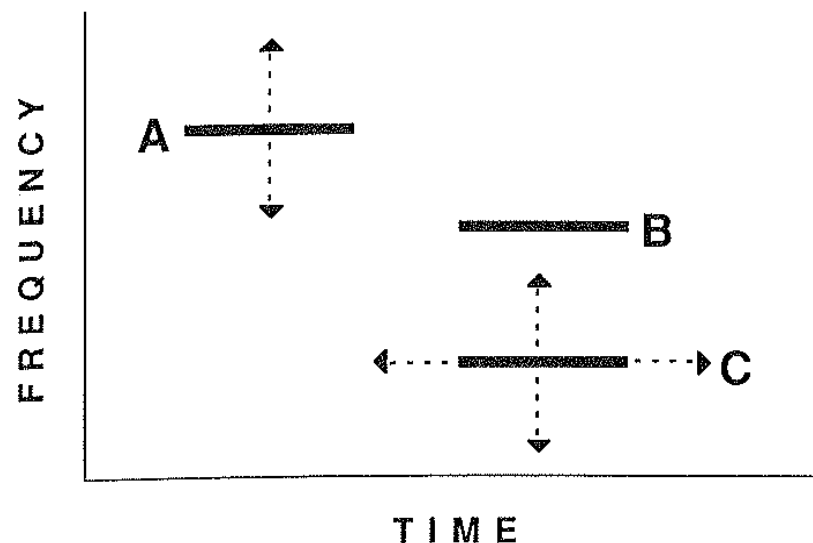


Exclusive Allocation



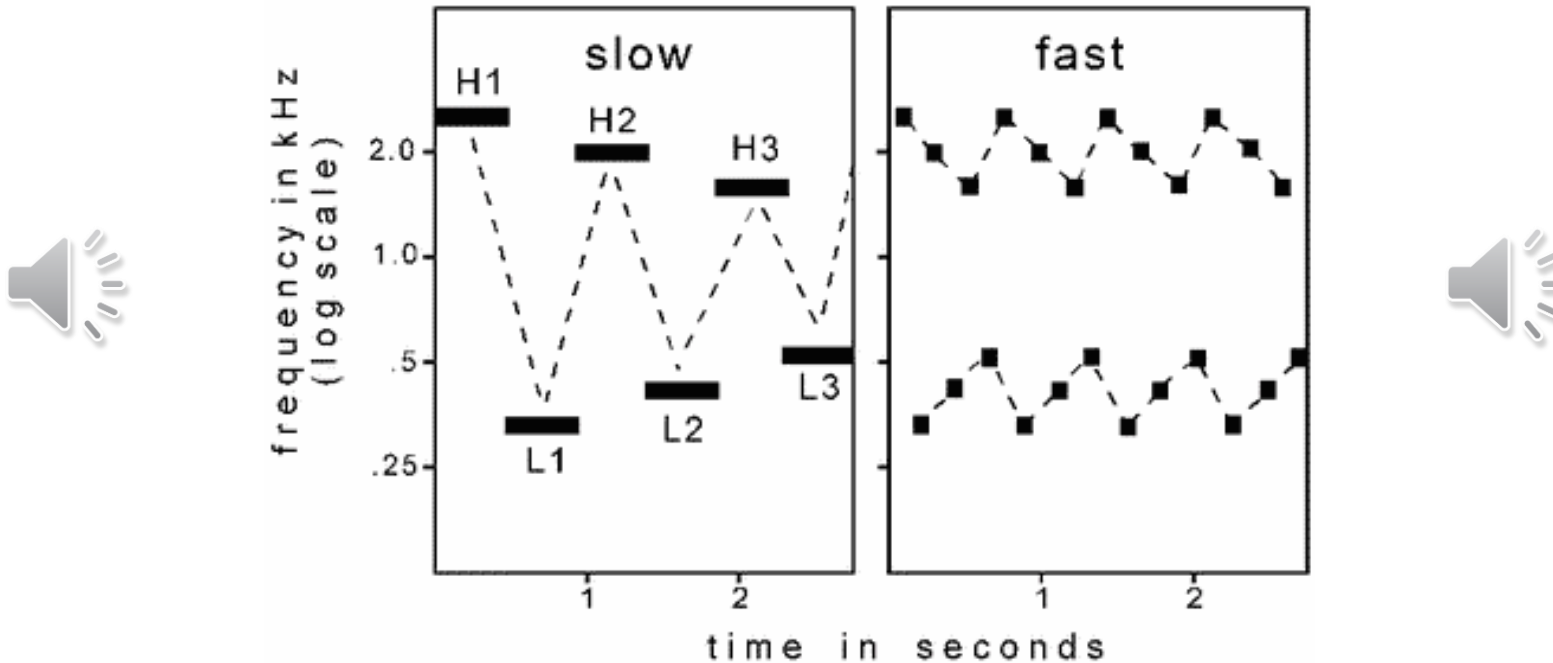
- The allocation of the X tones are different when the C tones are played or not, and it affects our perception of the A and B tones.

Simultaneous vs. Sequential



- Things that affect the grouping of ABC tones
 - Frequency difference between A and B
 - Frequency difference between B and C
 - Synchronization between B and C

Stream Segregation



- High and low tones are segregated when played fast
- Can you tell the order of the tones?

Segregation depends on...

- Time gap between tones within a stream
- Frequency gap between the two streams
- Let's look at a demo
 - <http://auditoryneuroscience.com/scene-analysis/streaming-alternating-tones>

Stream Segregation in Music

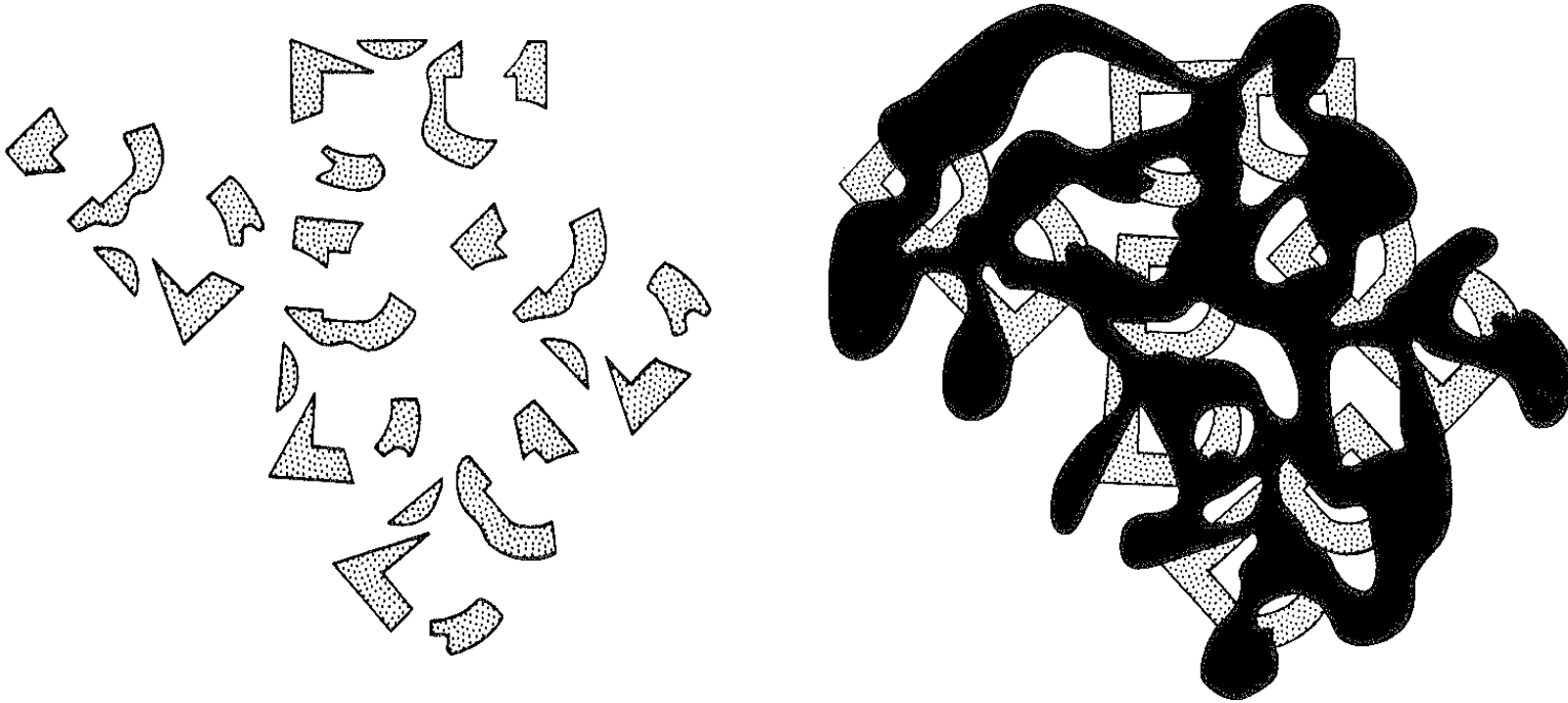
Toccata and
Fugue in D
minor, J.S.
Bach

The image displays a musical score for the Toccata and Fugue in D minor by J.S. Bach, specifically measures 28 through 31. The score is written for a grand piano, with a treble and bass staff. Measures 28 and 29 are in the top system, while measures 30 and 31 are in the bottom system. The key signature is one flat (B-flat). In measures 28 and 29, the right hand plays a series of chords and a melodic line, while the left hand plays a steady eighth-note accompaniment. In measures 30 and 31, the right hand continues the melodic line, and the left hand plays a more complex, rhythmic pattern. Several passages in the score are highlighted with yellow boxes, indicating areas of interest for stream segregation analysis.

https://www.youtube.com/watch?v=R_tu63ypB6I



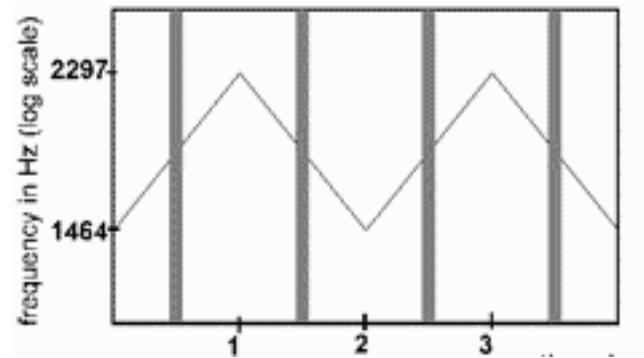
Occlusions in Vision



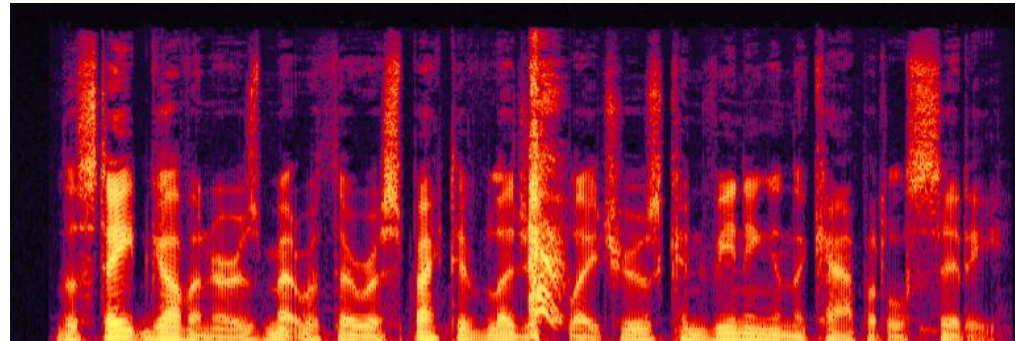
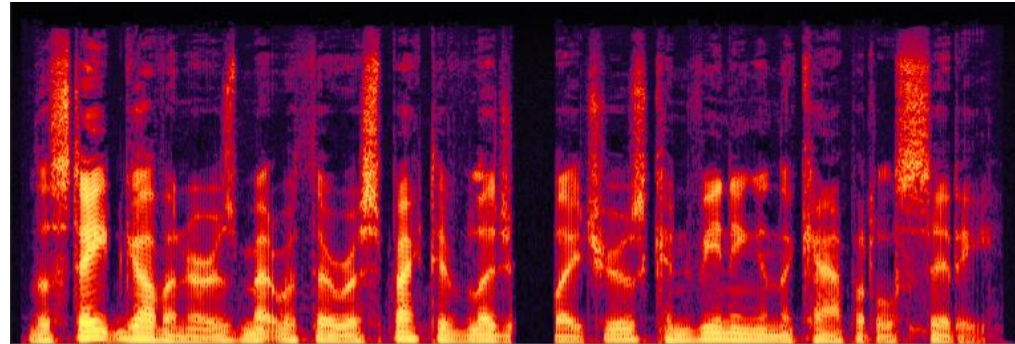
- The occlusion in this example helps with the grouping of the fragments

Masking in Audition

Sinusoids

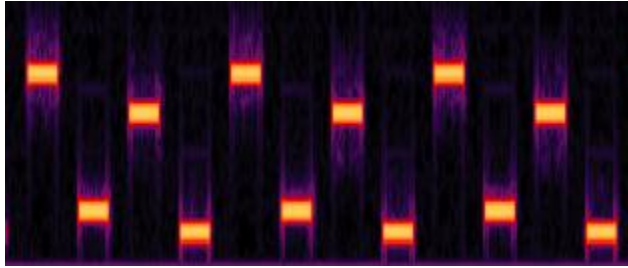


Speech

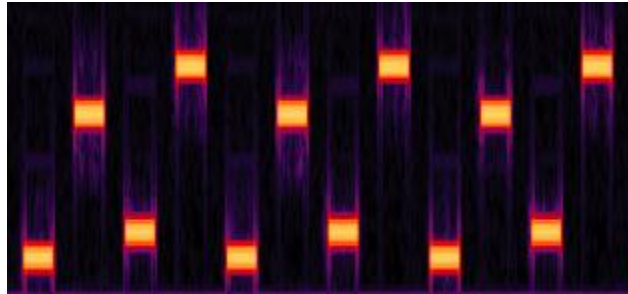


Primitive vs. Learned

H1-L1-H2-L2



L2-H2-L1-H1



- Infants cannot discriminate the two stimuli, which indicates that they perform stream segregation of the high and low tones.

Primitive Grouping Mechanisms

- For simultaneous grouping
 - Periodicity
 - Common onset and offset
 - Common amplitude and frequency modulation
- For sequential grouping
 - Proximity in frequency and time
 - Continuous or smooth transition
 - Related rhythm
- Common spatial location

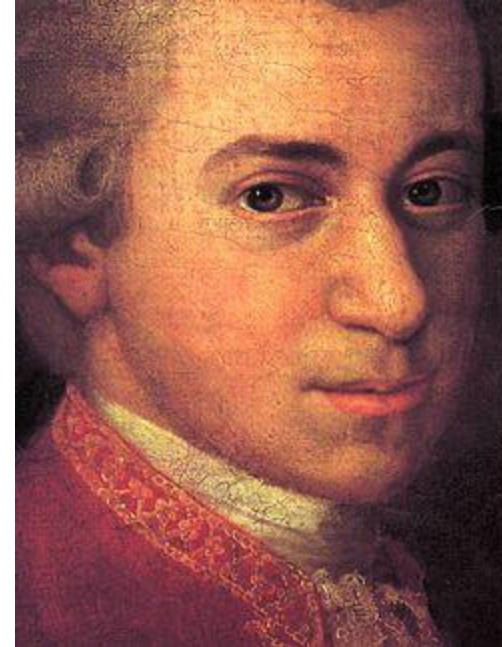
Primitive vs. Learned

- Repeated listening to the stimulus can improve performance in ASA tasks
- Easier to follow a friend's than a stranger's voice in a noisy environment
 - Prior knowledge of timbre helps
- Music training helps analyzing music audio scene
 - Prior knowledge of music theory, composition rules, music style, etc. helps

Extreme Capability in Music ASA

- “In Rome, he (14 years old) heard Gregorio Allegri's *Miserere* **once** in performance in the Sistine Chapel. He wrote it out **entirely from memory**, only returning to correct **minor errors...**”

-- Gutman, Robert (2000).
Mozart: A Cultural Biography



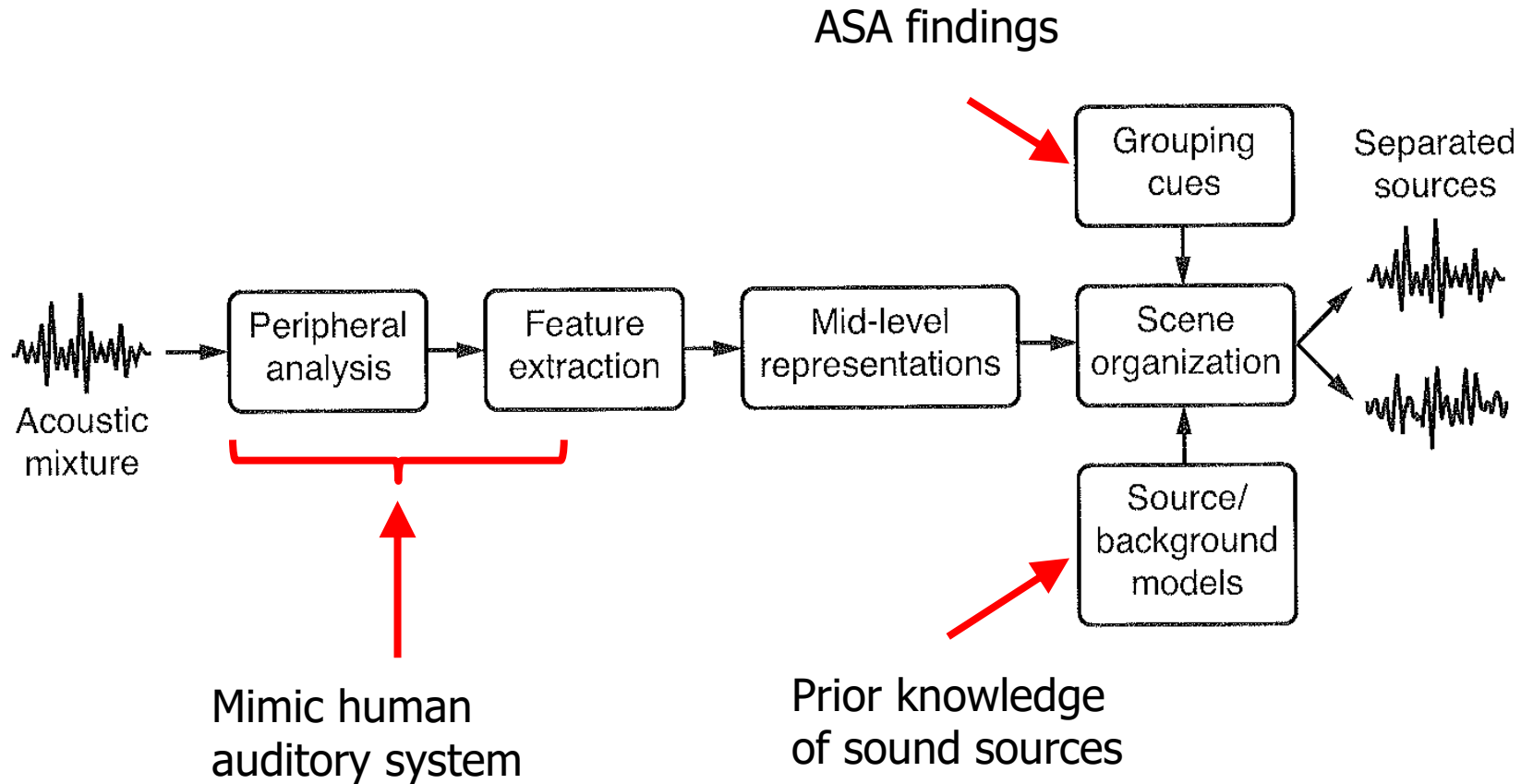
Wolfgang Amadeus Mozart

- Can we make computers compete with Mozart??

Computational ASA

- What is CASA?
 - “the challenge of constructing a machine system that achieves human performance in ASA”
----- E.C. Cherry
 - To computationally extract individual streams from one or two recordings of an acoustic scene
- The definition of CASA makes no reference to the underlying mechanism that a system should adopt, but many systems are based on the principles of processing in the human auditory system.

CASA System Overview



(from the CASA book, Figure 1.5)

CASA vs. Computer Audition

- Both have the same goal
- The term CASA has come to be associated with a perceptually motivated approach
- Computer Audition is open to all kinds of approaches including the purely engineering ones