Understanding the human auditory system

Mathieu Lagrange
Analyse / Synthèse Team, IRCAM
Mathieu.lagrange@ircam.fr

ircam
Centre Pompidou

ATIAM

How do we listen ?

- According to Gaver, there is 2 main modes of listening
  - **everyday** listening
  - **musica**l listening
- They can be reformulated as
  - **holistic** listening: fast screening based on pattern matching (low power processes)
  - **analytica**l listening: intensive search of correlation between various cues (high power processes)

What are we searching for ?

- According to Pierre Schaeffer, we can interpret the acoustic scene according to three different levels of similarity:
  - **Acoustic**: similarity of **acoustical** properties
  - **Causal**: similarity of the identified physical event **causing** the sound
  - **Semantic**: similarity of some kind of **knowledge**, or **meaning**, associated by the listeners to the identifed objects or event

The Auditory System Hardware

- A bit of physiology

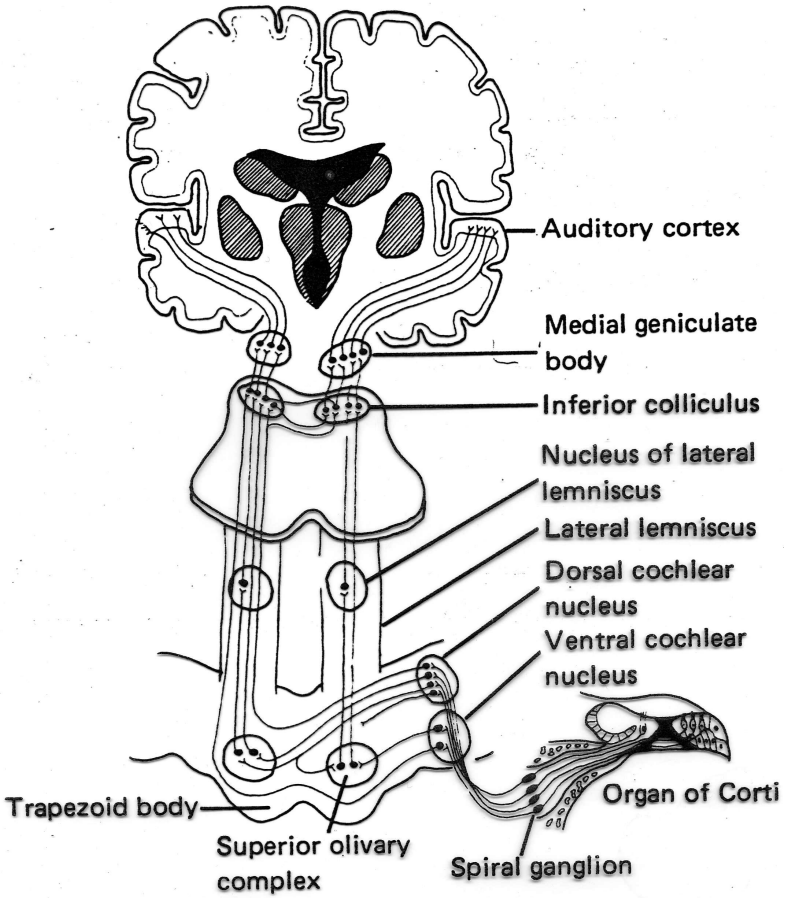


Figure 6-130 The major components of the ascending auditory pathway.

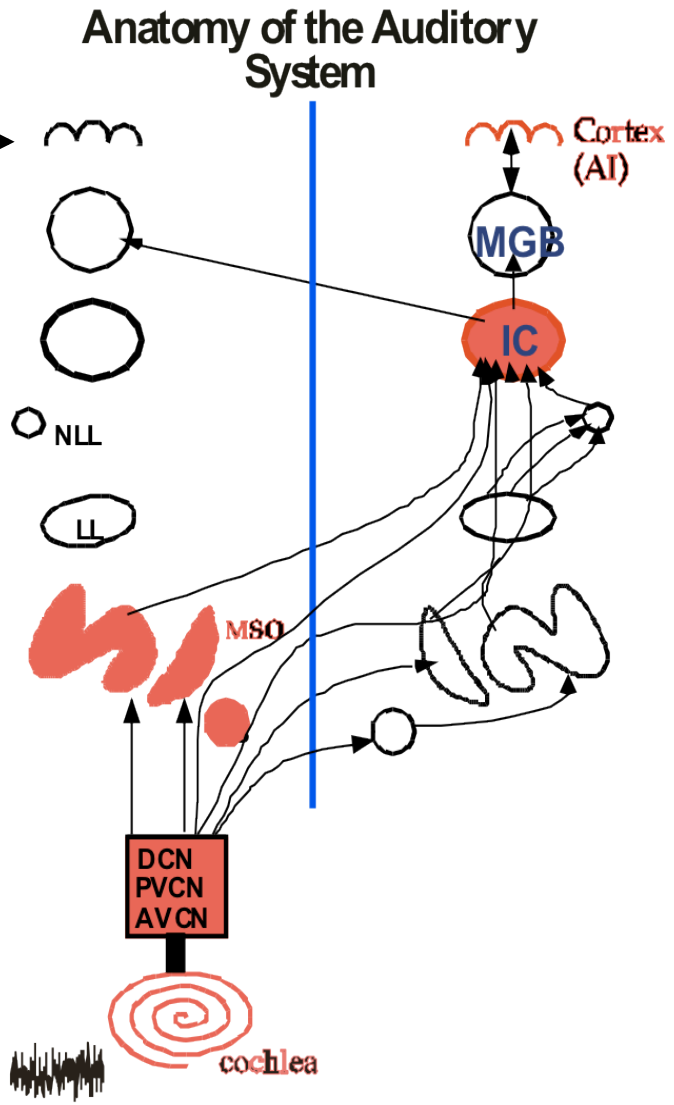
Let us be more abstract

Spectrotemporal modulations

Binaural processing

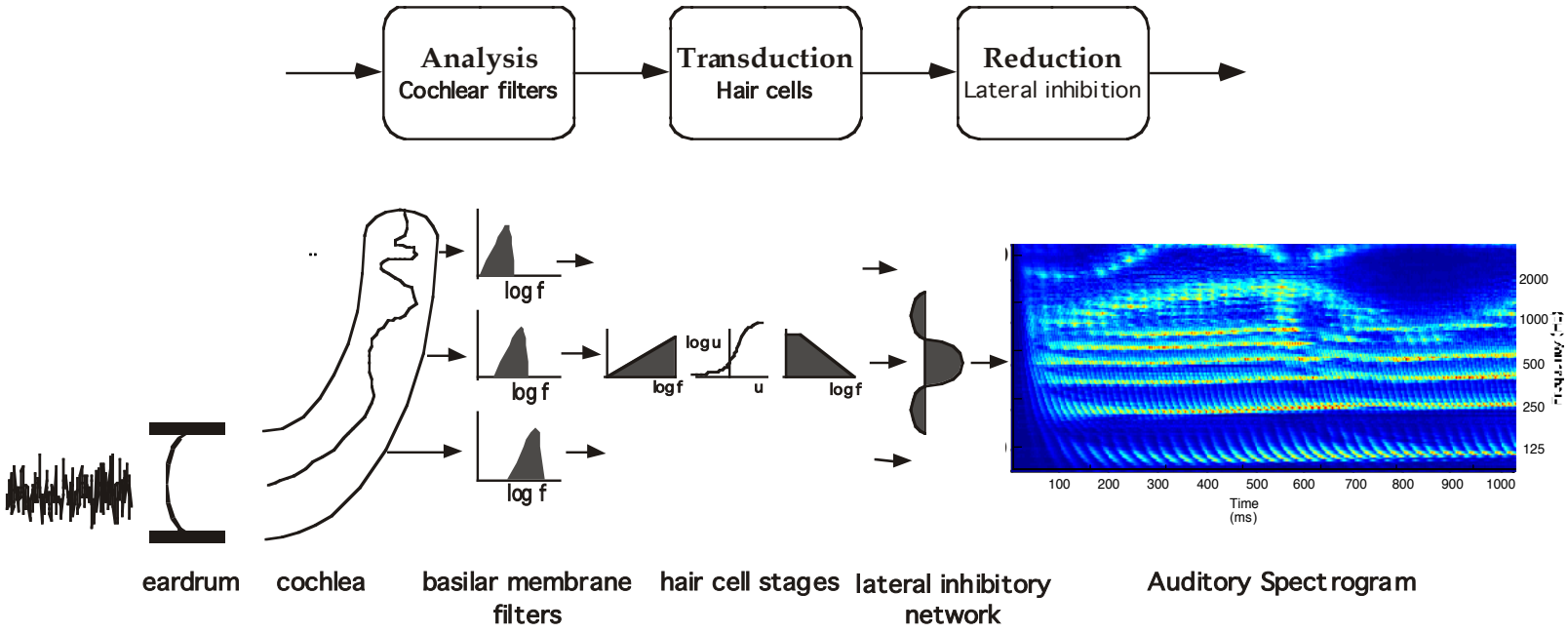
Spectral estimation

Spectral Analysis

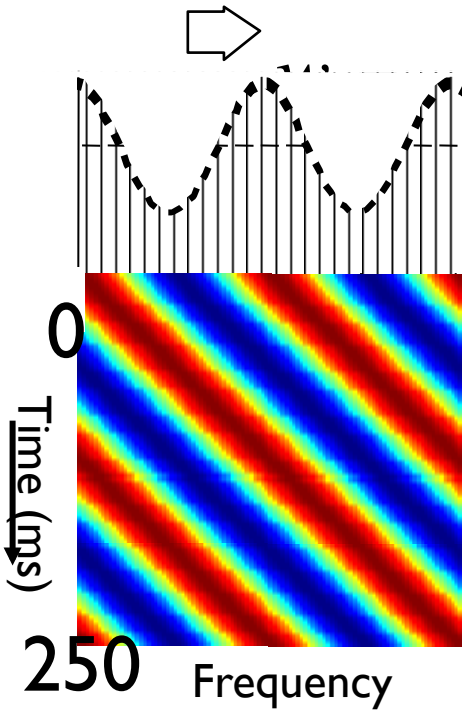


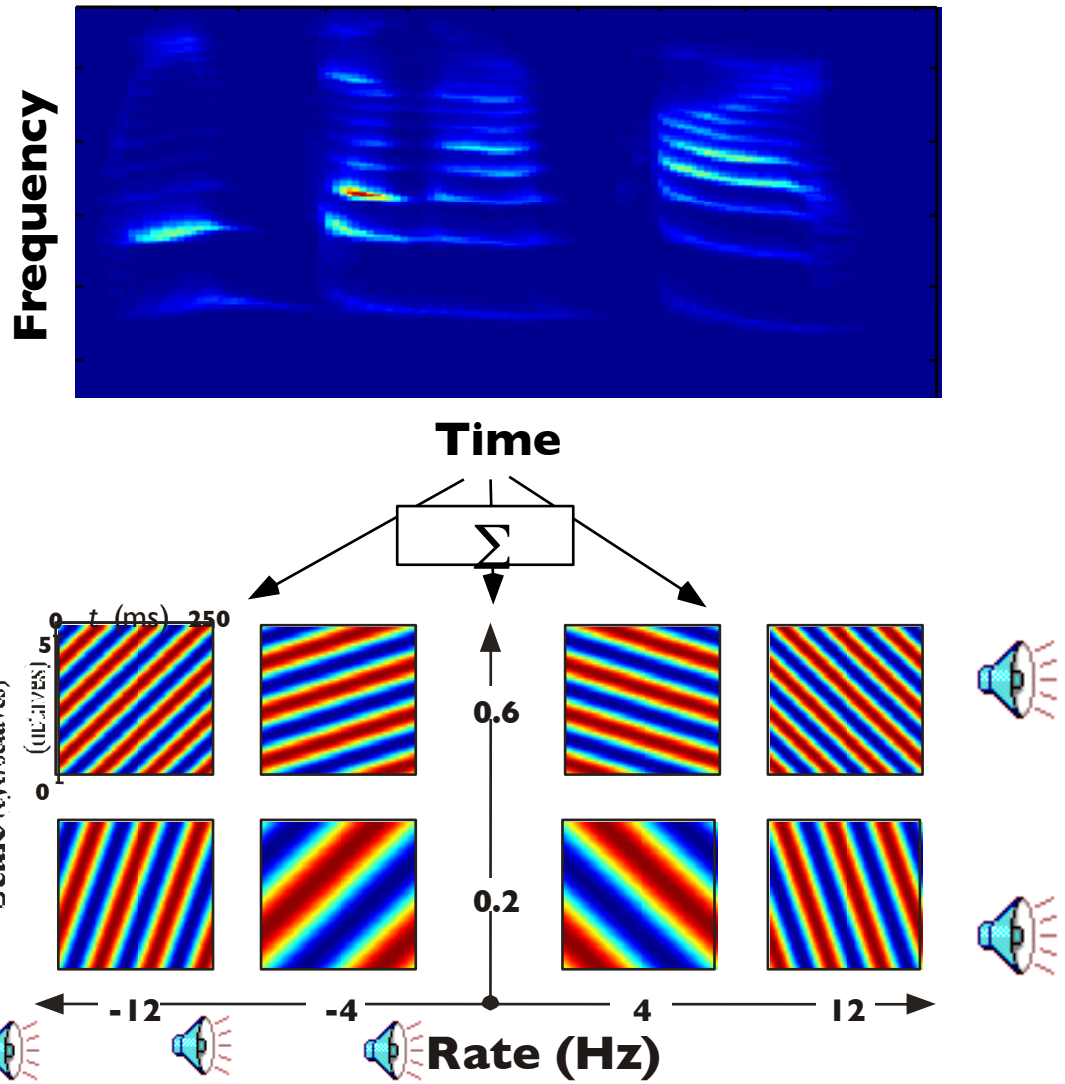
Spectral Analysis

Early Auditory Processing Stages

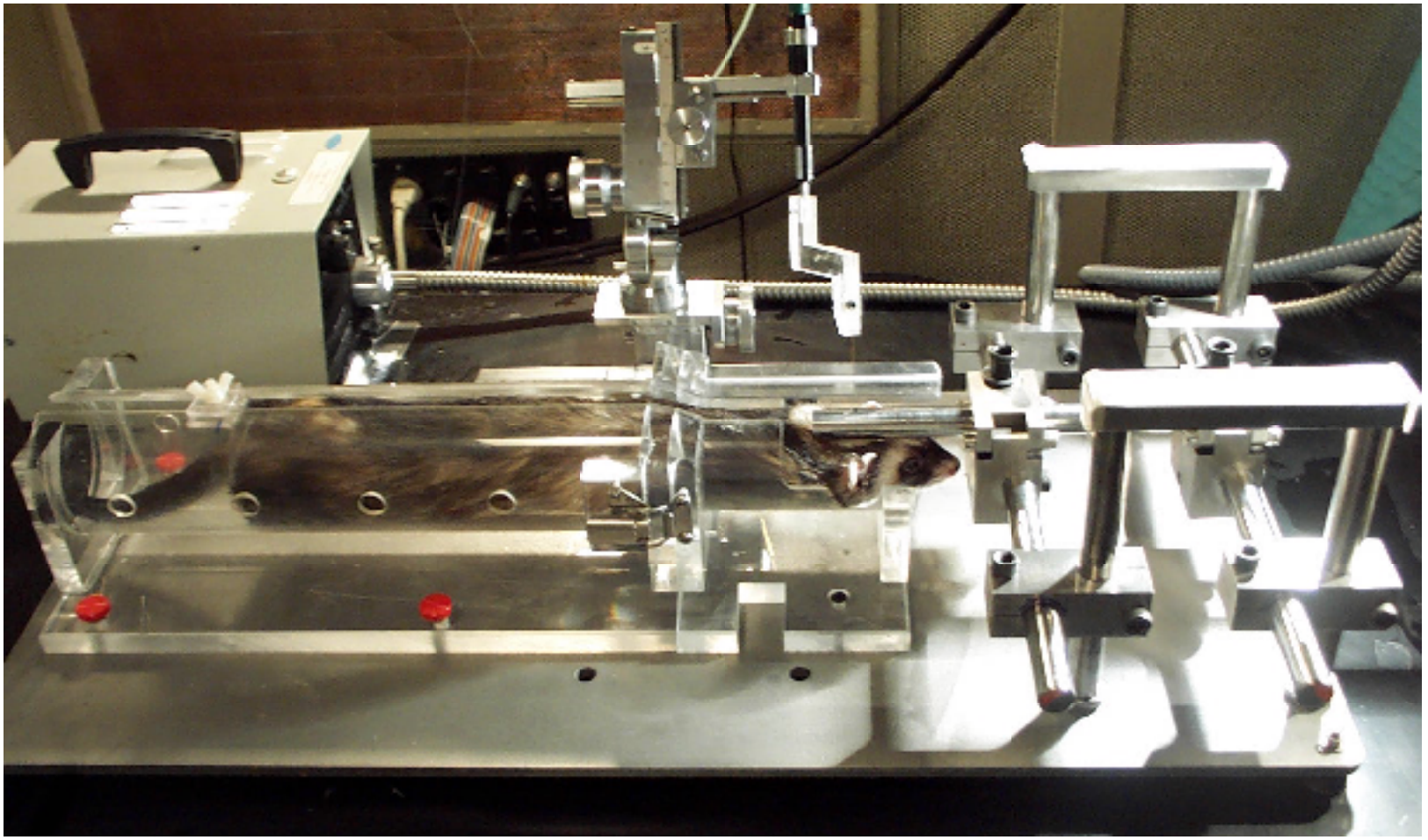


Spectro Temporal Modulations





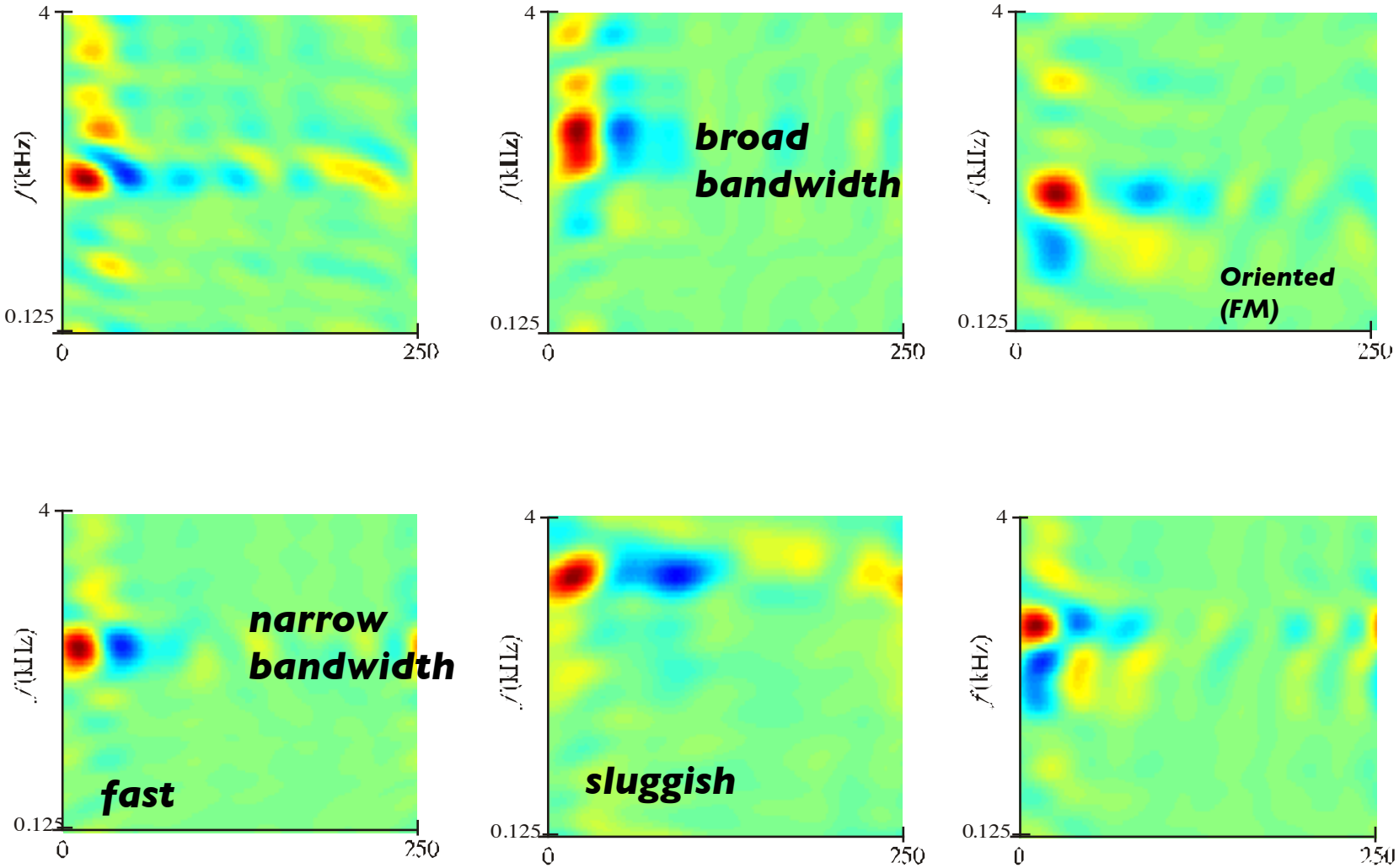
The listener



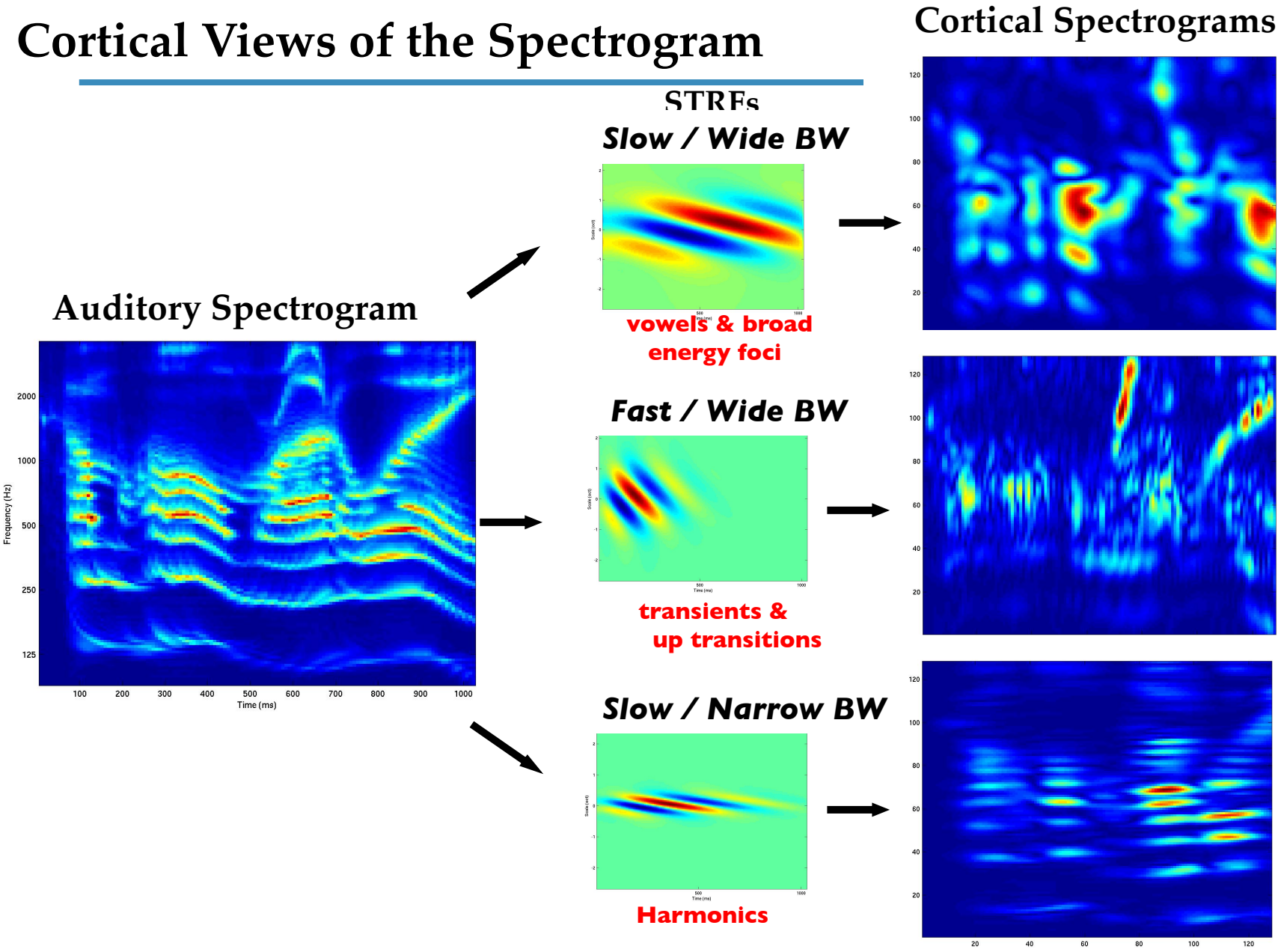
Awake ferret with head restraint in cylindrical holder

The Spectro-Temporal Response Fields

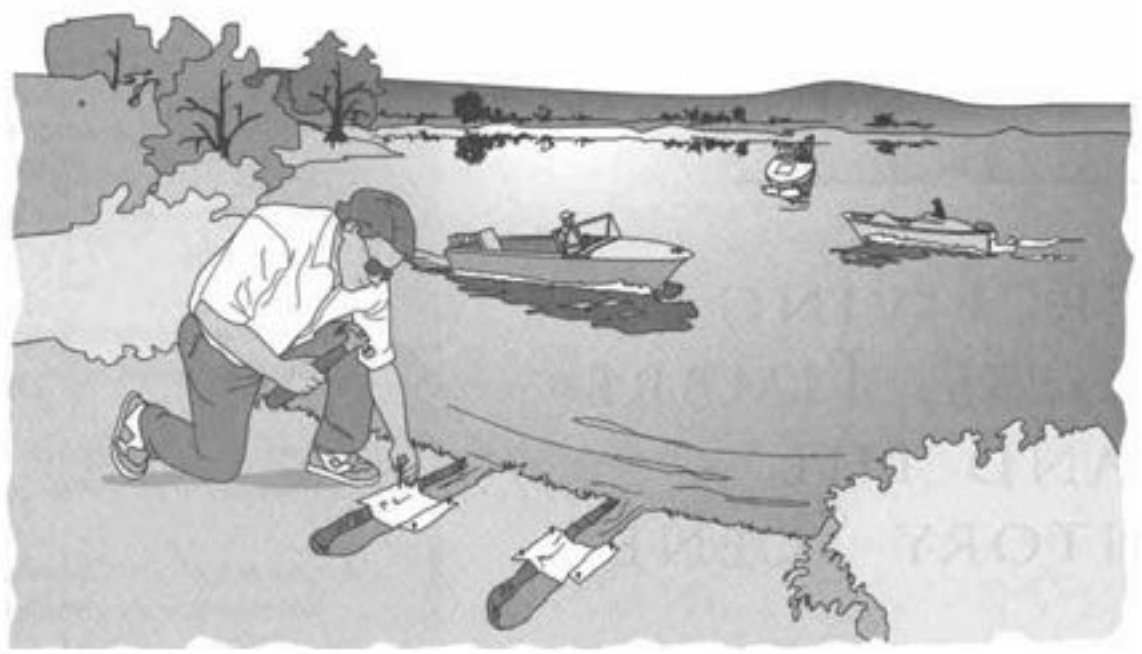
Examples of Different STRF Shapes



Cortical Views of the Spectrogram

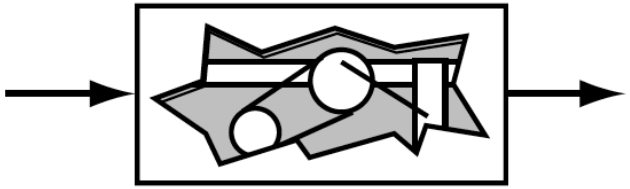


What next : Auditory Scene Analysis ?



ASA ?

- What is not ASA:
  - Physiology: implementation

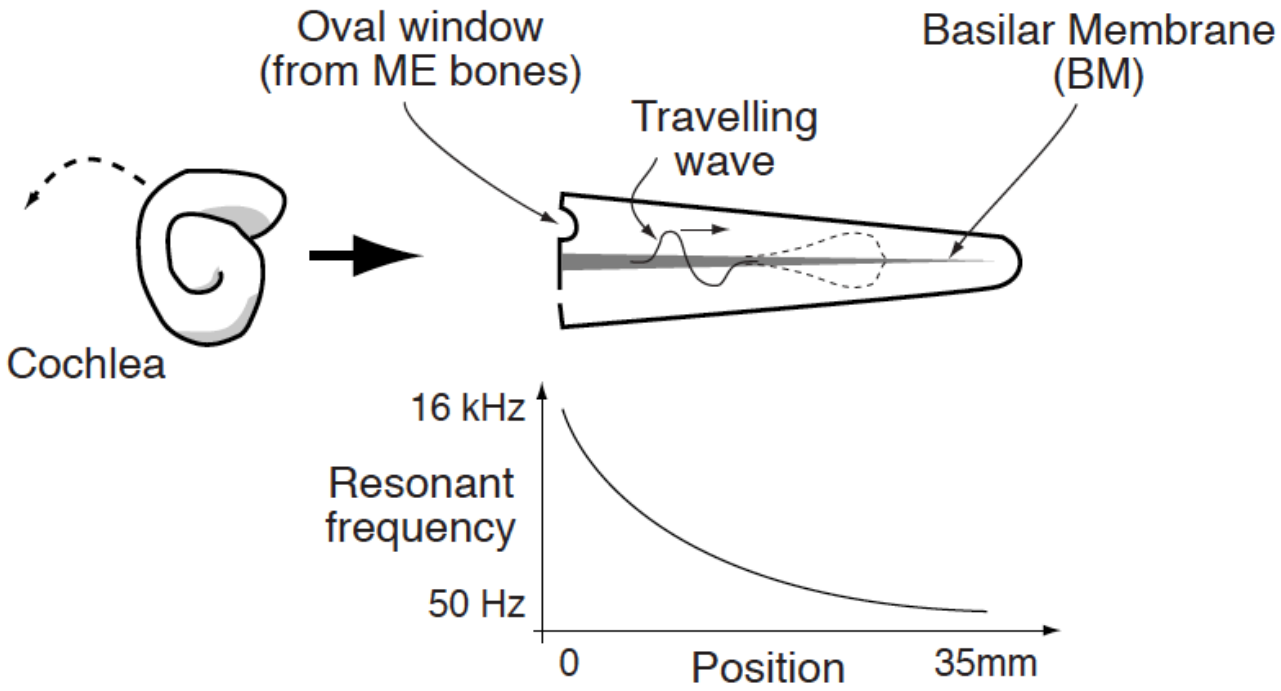


  - Psychophysics: function/behavior

- ASA looks at:
  - Information processing models

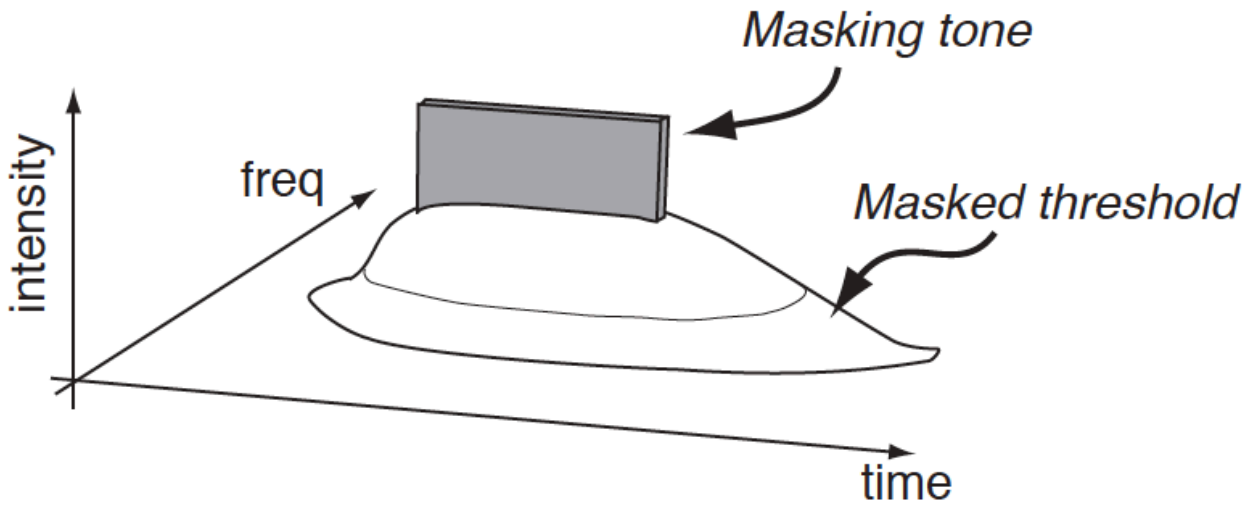
Physiology

- Inner ear

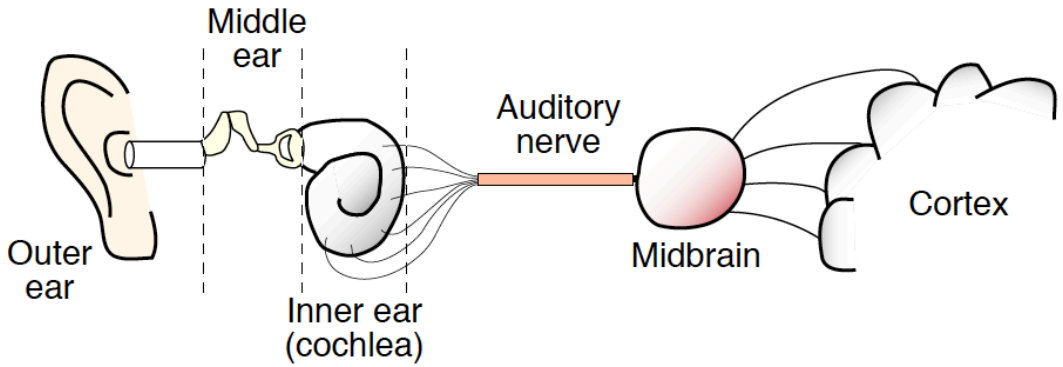


Psychophysics

- Relate physical and perceptual variables
  - Intensity -> loudness
  - Frequency -> pitch
- Time/Frequency Masking



Next ?



- No matter how precise (or imprecise) our measurement system will be
- Signals arriving are non linear mixtures of many components sounds
- Some of those components have to be individually described
  - This is the purpose of ASA

ASA: digging into the unkown

- For most people, performing ASA means
  - Paying attention to one of the sound at a time
  - Very difficult to do better (not ecologically useful ?)
- How do we do presumably ?
  - Activation of learned schemas in a purely automatic way
    - Have you ever mistakenly heard your name in a crowd ?
  - Activation of learned schemas in a voluntary way (attention)
- What are schemas:
  - Mental representation of a particular set of characteristics
  - Implicitly or explicitly formed by prior listening

The methodology of ASA

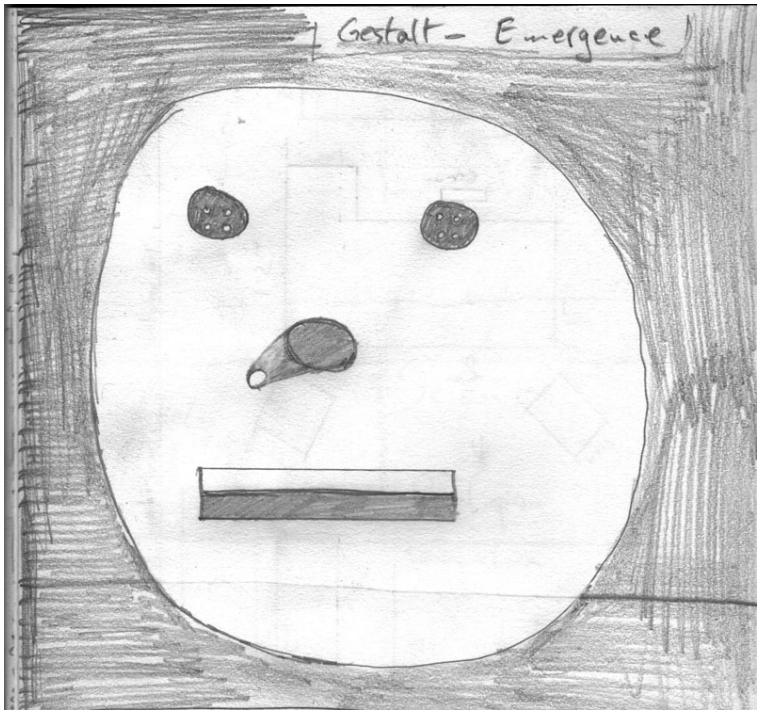
- How did we learned such schemas in the real world?
- Needs for general methods for partitionning an incoming mixture
- Those methods are guided by (ecologically selected ?) cues:
  - Psychophysical complementarity (Shepard 1981)
  - Determining the laws of auditory organization reduces to
    - Discover relations among the components
    - Perform experiments to determine how the Human Auditory System (HAS) uses them

ASA is a Gestaltist theory

- Principle of Totality - The conscious experience must be considered globally.
- Principle of psychophysical isomorphism: A correlation exists between conscious experience and cerebral activity
  - Some scientists states that there is some kind of « tuning »
- Key principles of Gestalt systems are
  - Emergence,
  - Reification,
  - Multistability
  - Invariance

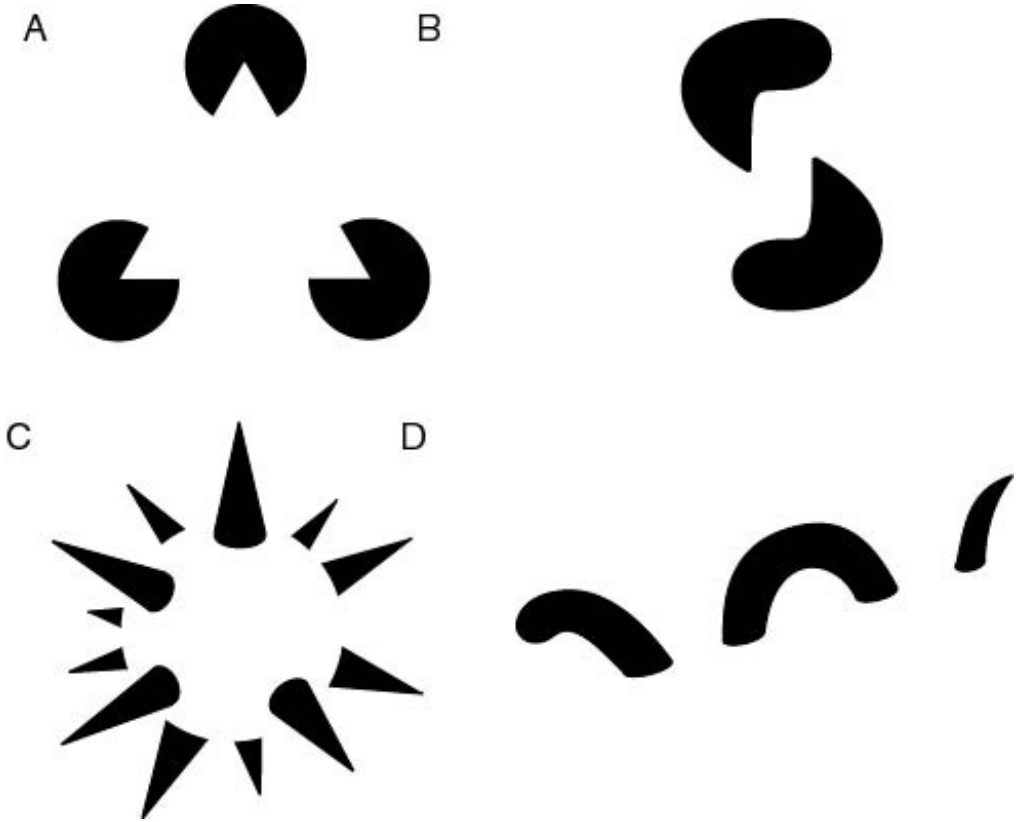
Emergence

- Emergence is the process of complex pattern formation from simpler rules



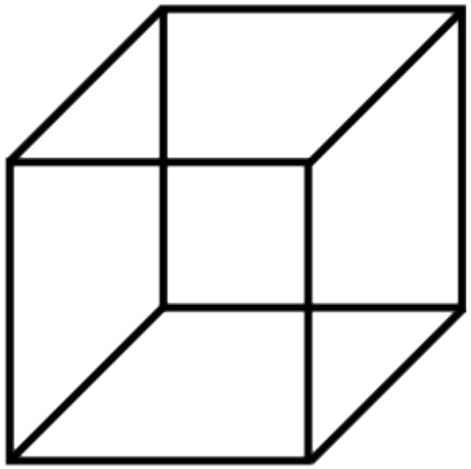
Reification

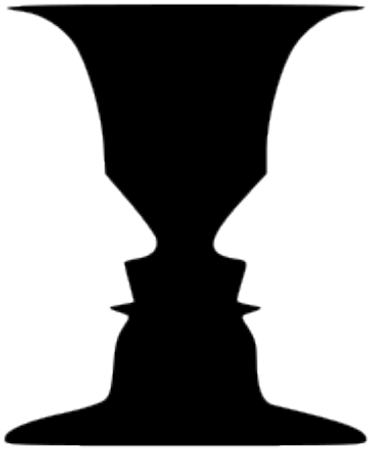
- Reification is the constructive or generative aspect of perception



Multistability

- Multistability (or multistable perception) is the tendency of ambiguous perceptual experiences to pop back and forth unstably between two or more alternative interpretations



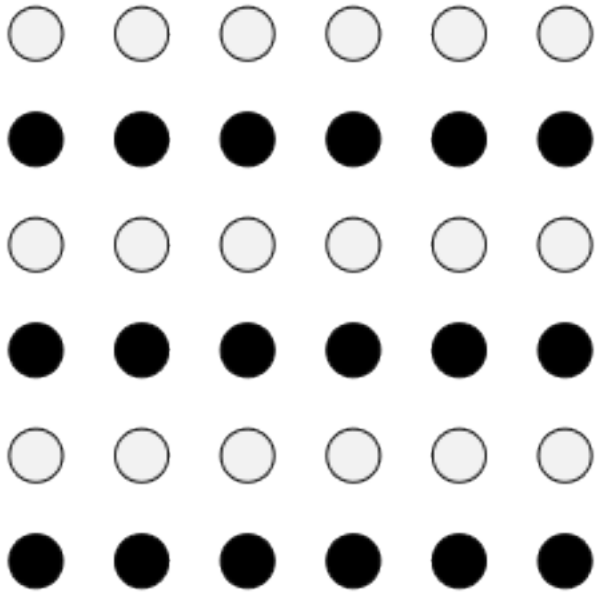


“Prägnanz” rules

- the law of prägnanz (German for pithiness) says that we tend to order our experience in a manner that is
  - Regular
  - Orderly
  - Symmetric,
  - Simple

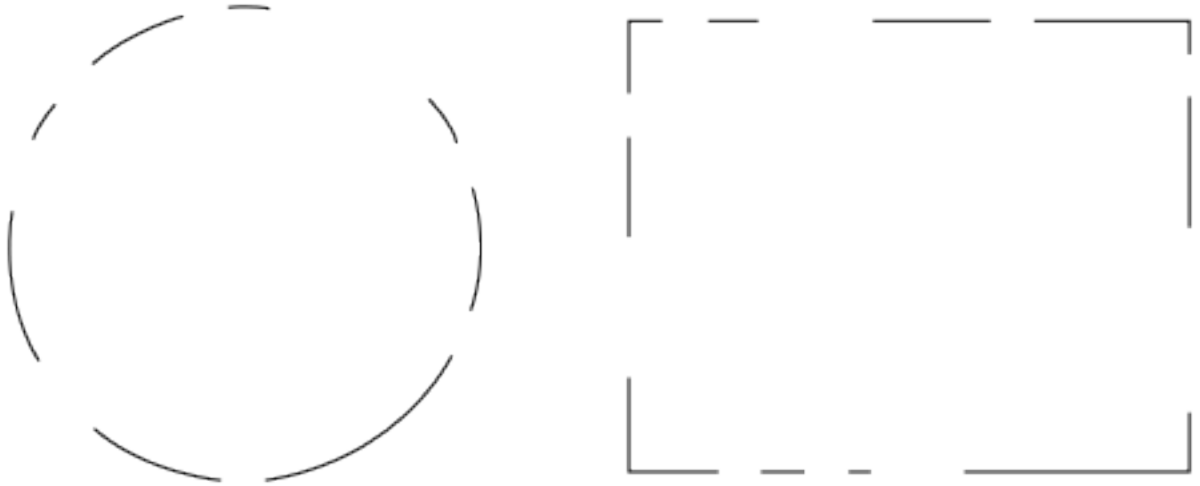
Similarity

- Law of Similarity: the mind groups similar elements into collective entities or totalities. This similarity might depend on relationships of form, color, size, or brightness.



Closure and symmetry

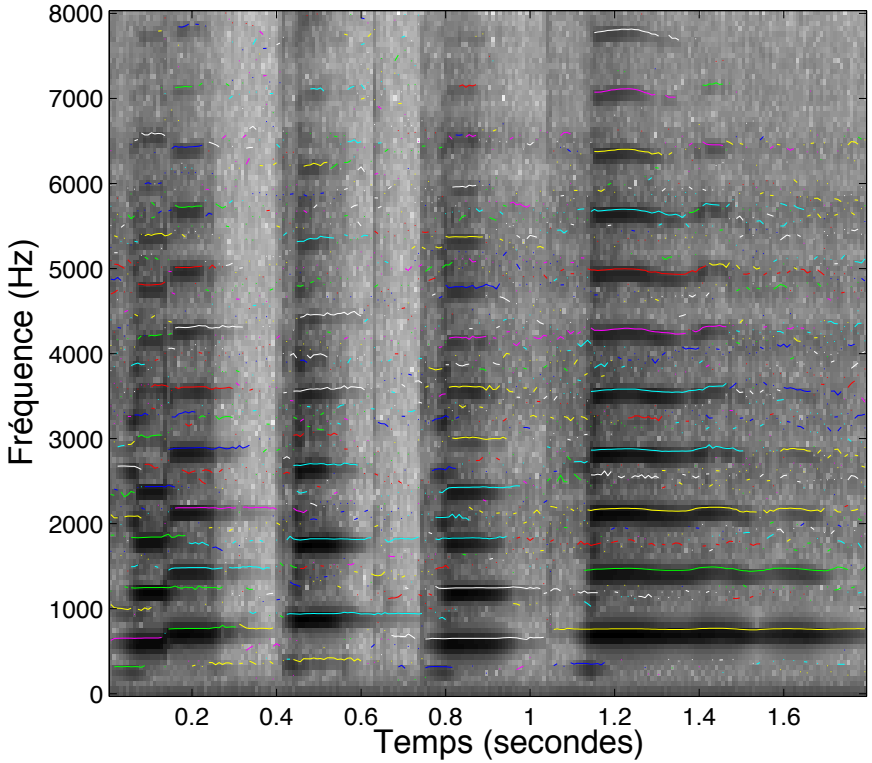
- Law of Closure: the mind may experience elements it does not perceive through sensation, in order to complete a regular figure



- Law of Symmetry: Symmetrical images are perceived collectively, even in spite of distance

Continuity and common fate

- The mind continues visual, auditory, and kinetic patterns.



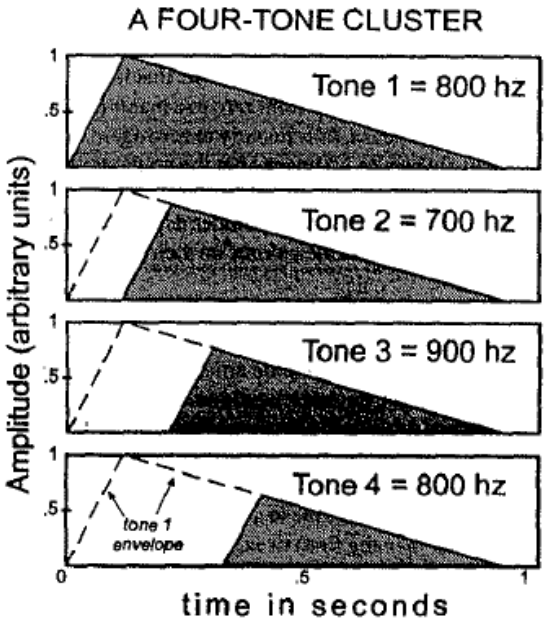
- Law of Common Fate: Elements with the same moving direction are perceived as a collective or unit.

Auditory Demonstrations

- Albert S. Bregman / Pierre A. Ahad "Demonstration of Auditory Scene Analysis, The perceptual Organisation of Sound"
  - http://webpages.mcgill.ca/staff/Group2/abregm1/web/downloadstoc.htm#
- For a comprehensive view of Auditory Scene Analysis:
  - Bregman, A. S. (1990) *Auditory scene analysis: the perceptual organisation of sound.* Cambridge, Mass.: The MIT Press (in library)
  - Other books on auditory perception also give descriptions of ASA

ASA regularity

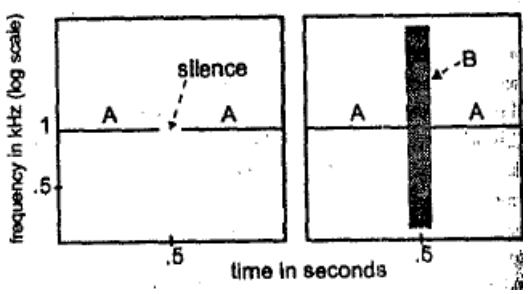
- Gradualness of change
  - A single sound tends to change its properties smoothly and slowly
  - A sequence of sounds from the same source tends to change its properties slowly
- Unrelated sounds seldom start or stop at exactly the same time
  - From abrupt to smooth onsets (Kim 94)



(21)

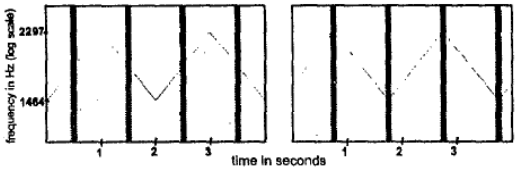
ASA regularity

- Perceived continuity:
  - Sine tone and burst of noise (Warren 1984)
    - Apparent continuity



(28)

    - Perceptual continuation of a gliding tone through a noise burst



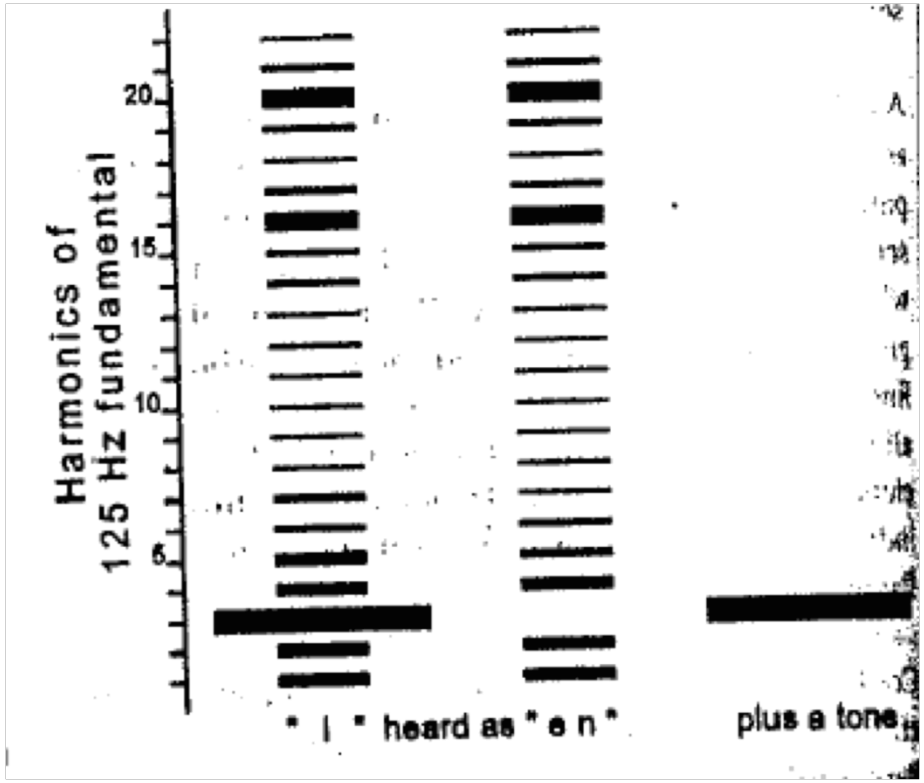
(29)

    - Picket fence effect

(31)

Competition

- Sine tone and vowel (Darwin 1984)
  - Changing a vowel's quality by capturing a harmonic
    - 4 'e' then 4 'en', then 4 'e' with capturing tone



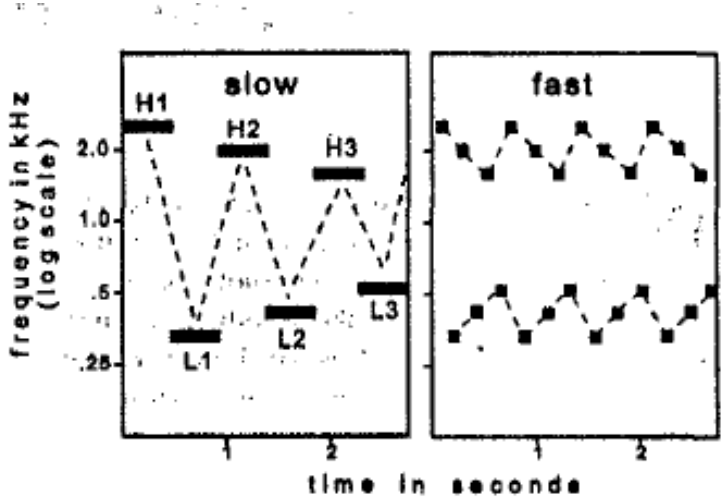
(37)

Old+New heuristic

- Decomposing and interpreting mixtures of sounds as
  - A continuation of previously recieved and interpreted events
  - New events
- This heuristic has strong explanation capabilities
  - Foreground / Background
  - Context / Attention
- Nice bootstraping causal framework:
    - start with/without prior knowldege
    - Start to understand the current snapshot of the scene
    - Consider the next snapshot
      - Remove what can been understood as a continuation
      - Focus on the remaining

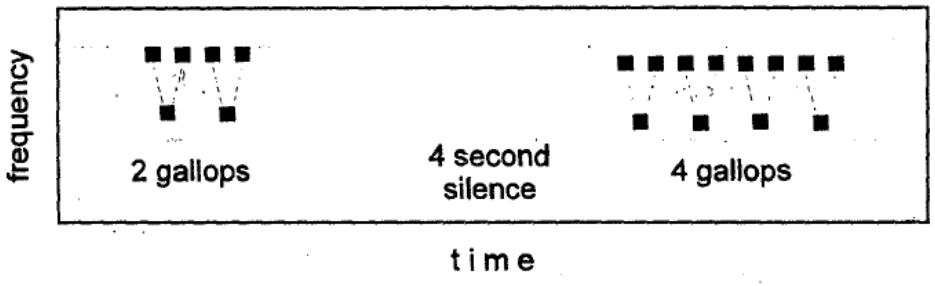
Sequential Streaming

- Given Low frequency tones and High frequency ones



(1)

- There is trade off between speed and frequency difference
  - Segregation sensitivity can be viewed as a rate sensitivity
  - Segregation takes time to build up and remains for at least 4 seconds



(4)

ASA Regularity 2

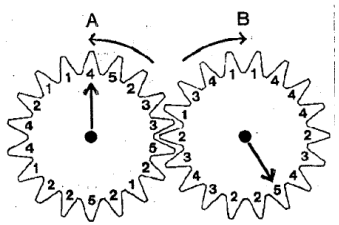
- The ‘grouping by similarity’ rule
  - Take sounds that have similar properties
  - Link them together perceptually into groups
  - Segregate them from one another
    - Actually, segregate the source of interest and discard the rest

Segregation cues

- Frequency
- Spatial position
  - Not mandatory
- Timbre
  - Usually defined as the spectral envelope (stationarity assumption)
  - Though non stationarity are extremely important
- Harmonicity

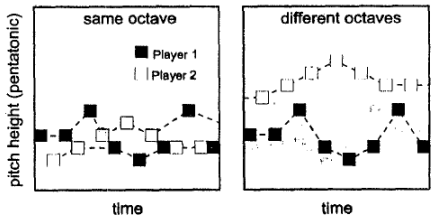
Competition

- In case of competition
  - The winner is the grouping that considers the cues that the HAS prefers
- Though, this preference depends on many factors
  - Prior, attention, context…
- Illustration with xylophone duet
  - Normal



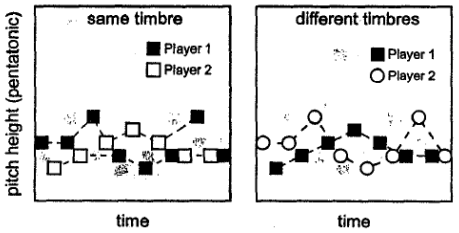
(7)

  - Change of pitch range



(8)

  - Change of timbre



(9)

Attention

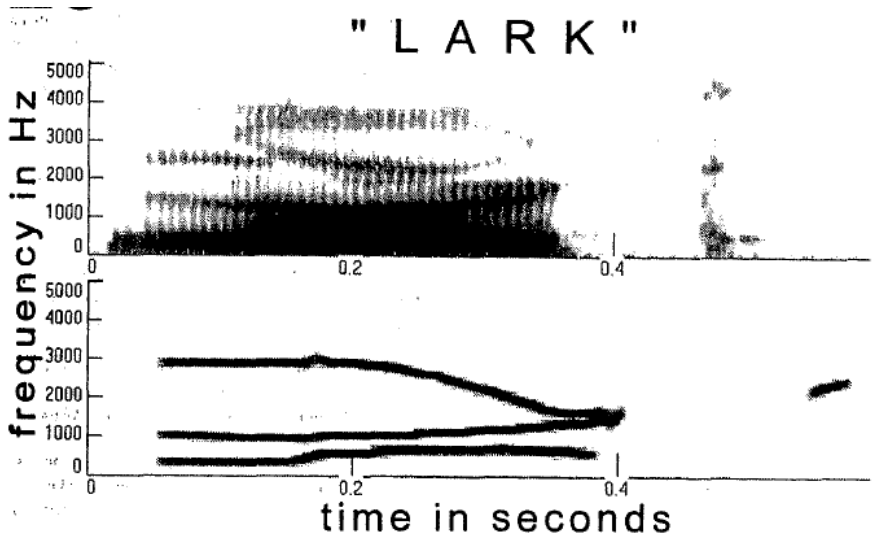
- Consider the High/Low experiments with varying speed and delta
  - Ask the listeners to integrate the sequence as much as possible
    - Trade-off between between speed and delta
  - Ask the listeners to segregate the sequences as much as possible
    - As long as the delta is sufficient, the segregation is done at any rate
- Evidence that some primitive mechanisms can be controlled up to a certain level

Primitive vs. schema based processing

- Vowel recognition
  - Mix 2 vowels with the same pitch (Scheffers 1983)
    - Performance of the listeners well above chance
  - Slightly change the pitch
    - Significant rises of recognition rate

Schema based processing

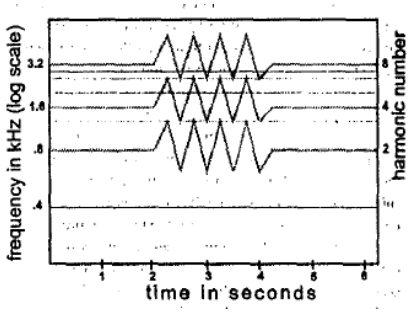
- Sine wave speech
  - One sine wave per formant
  - Monophonic (Bailey 77)
    - Perfect recognition rate
  - Polyphonic (Barker 99)
    - Extremely difficult
  - Solution:
    - "Please say what this word is »
    - "sill, shook, rust, weed, pass, lark, jaw, coop, beak",



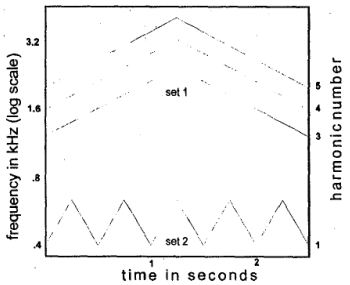
(23)

ASA Regularity

- Many changes that take place in an acoustic event will affect all the components of the resulting sound in the same way and at the same time
  - Synchronized frequency change
    - Intentional modulations

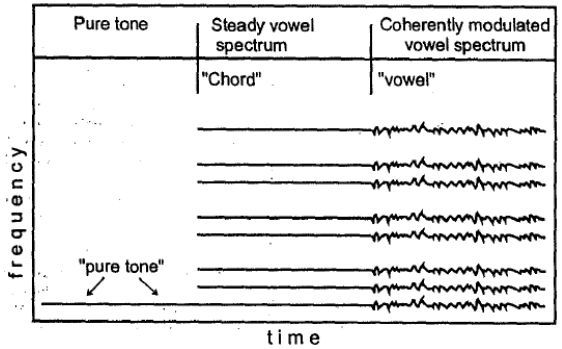


(19)



(20)

    - Micro-modulations



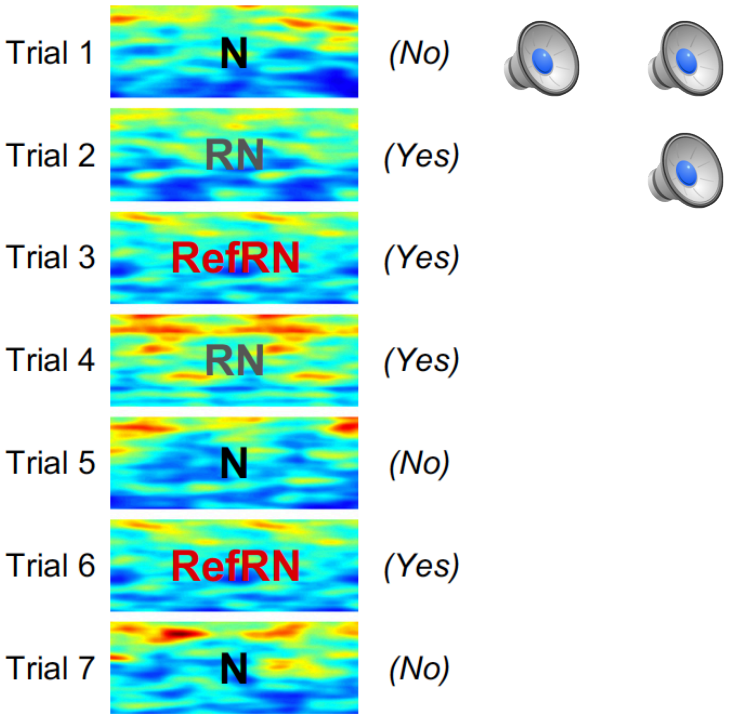
(24)

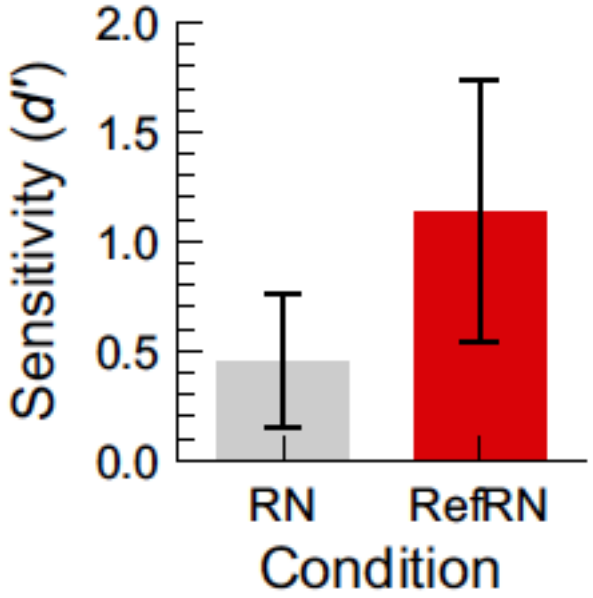
Summary

- Two types of processes
  - Bottom-up: primitive cues (hard-wired ?)
  - Top-down: schemas (learnt priors with relative and adaptive confidence)
  - Non linear influence between those processes
- Two types of integration
  - Simultaneous (from spectral components to notes)
  - Sequential (from notes to melody)
  - Again, non linear influence between those two

Implicit Learning of Schemas

- According to (Agus 10) low level (acoustic) schemas
  - can be learned very rapidly, only few exposition necessary
  - Are available for several weeks
  - Does not require ANY meaningful structure (noise stimuli)





Implicit Learning of Schemas

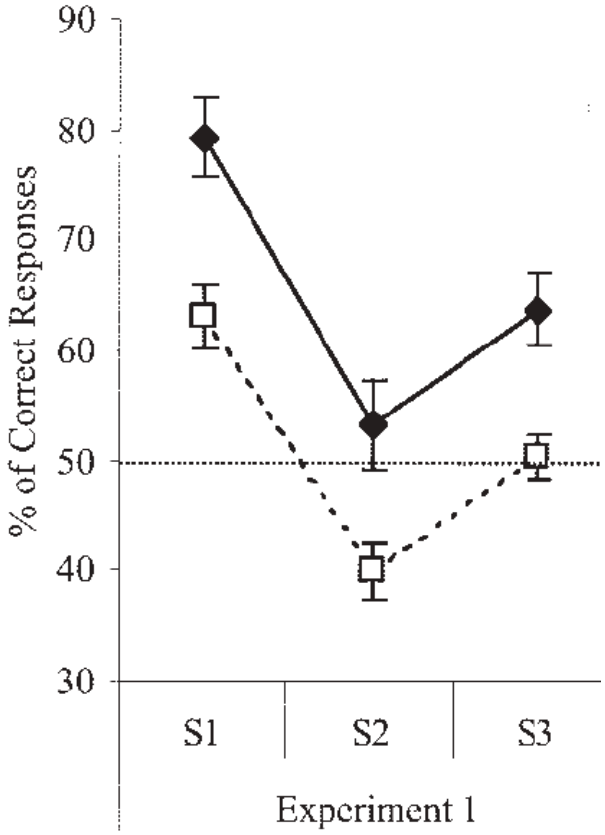
- Most occidental people are implicit expert of tonal music
- Tonal system
  - Restricted set of components
  - Statistical regularities (chord, tonality)
- One note is dependant of the context
  - Linked to the tonal hierarchy
- Other systems
  - Artificial ones
  - System coming from other cultural contexts

Artificial languages

- Simple systems
  - Triplets of syllables or musical tones
    - Exposition: listening passively to some triplets
    - Test: choose between two word or melody which one is coming from the exposed set of triplets
    - Results: 75 % (well above chance)
- More complex grammars gives the same results

Artificial languages

- Acoustical similarities only bias the performance of the implicit learning (Tillman 04)
  - Use of instruments that lies in a given timbre space
  - S1 positive influence of timbre,
    - within triplets, instruments are close
  - S2 negative influence of timbre,
    - within triplets, instruments are far apart
  - S3: neutral
    - no correlation between instrument change and triplets transitions



Atonal music

- One series and some transformations
  - Exposition based on several excerpts from the same series with active listening
  - Test: distinguish between previously heard excerpts and others from a different series
  - Results: around 60 % for musicians and non musicians

Summary

- From low to high level of mental representation, the HAS has a high level of plasticity that allows us to adapt to generate new expectations from an every day changing world
  - According to some studies on vowel perception this does not degrade with time
- Even at very low level, no implicit structure within the stimuli is necessary to allow the HAS to generate reliable expectations