Auditory Scenes Analysis

An Overview

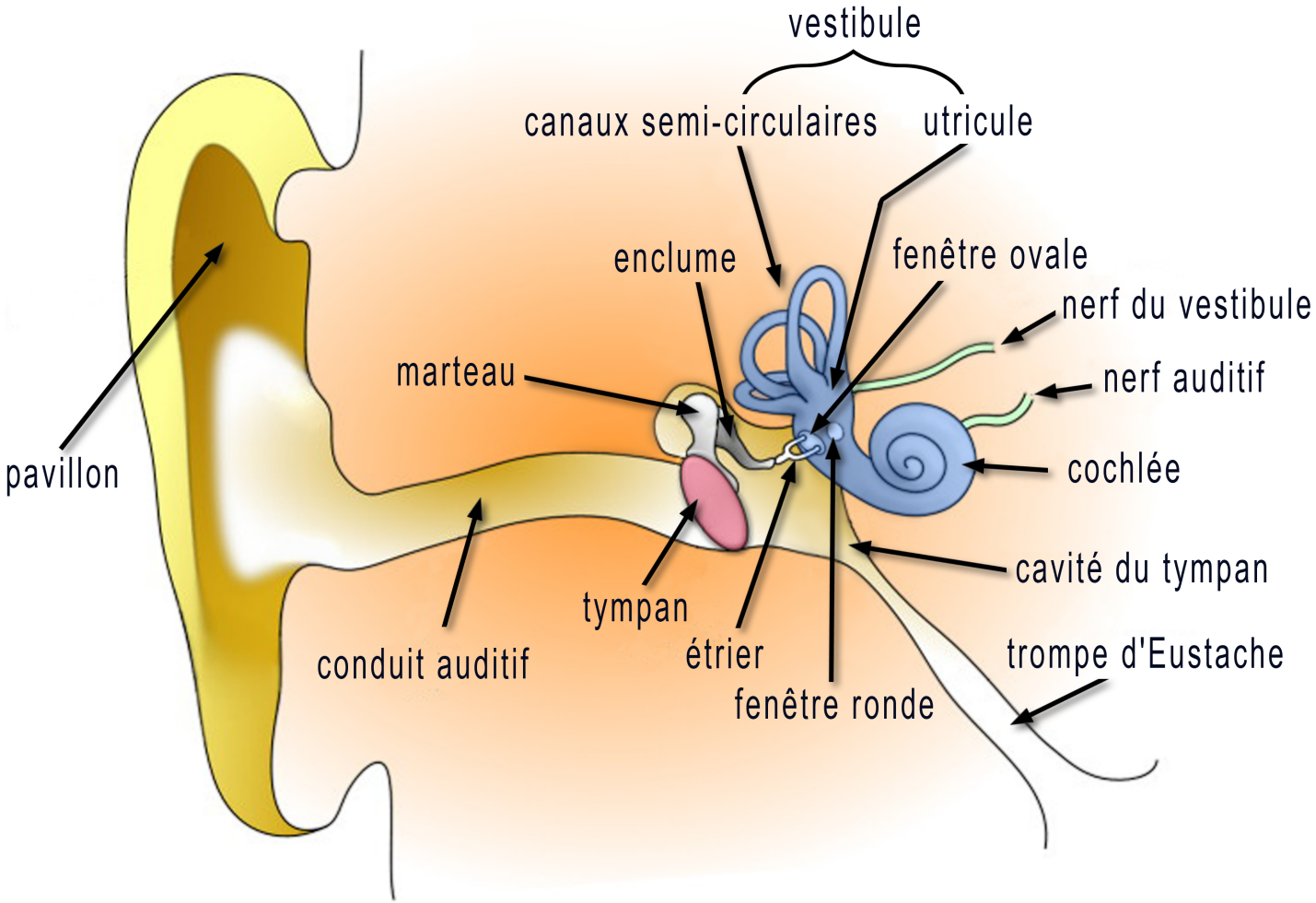
Grégoire Lafay & Mathieu Lagrange

mathieu.lagrange@cnrs.fr

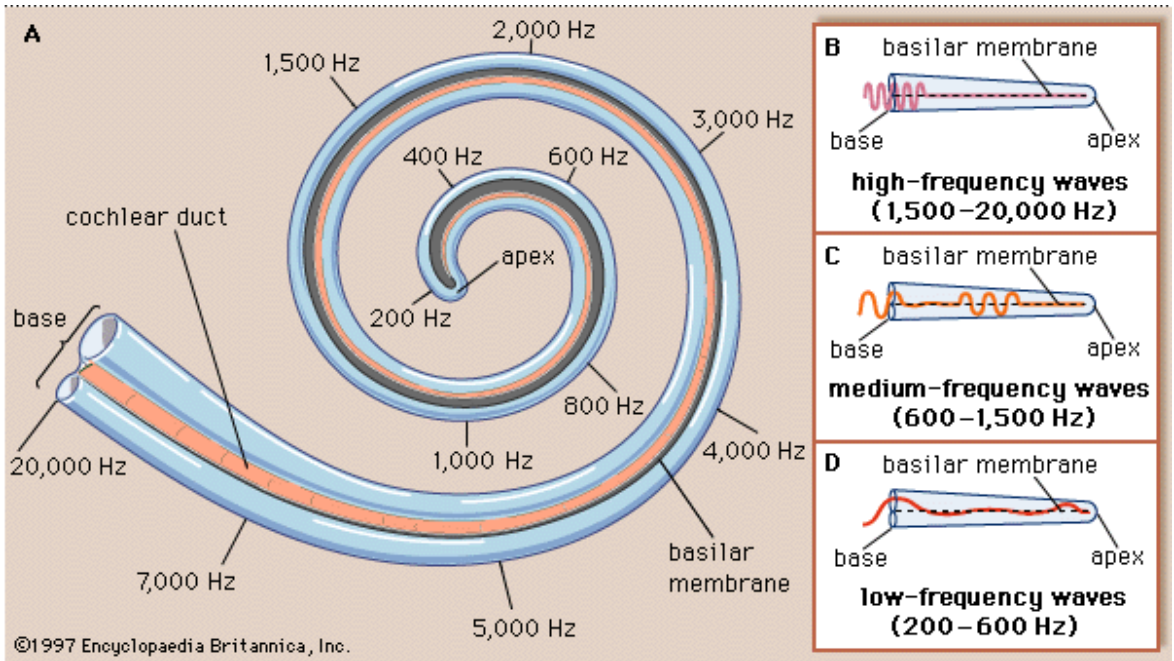
March 6, 2019

Table of Contents

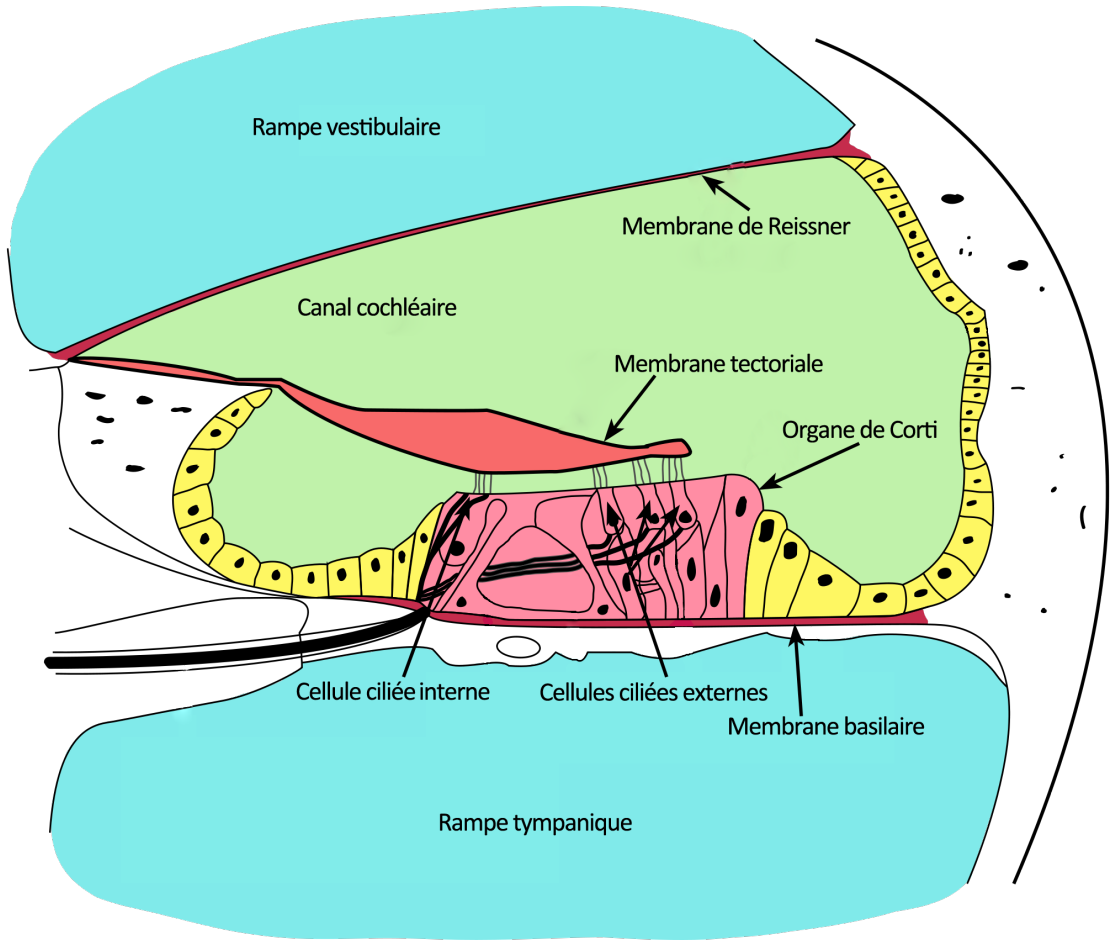
Inner Ear


vestibule
canaux semi-circulaires
utricule
enclume
fenêtre ovale
nerf du vestibule
marteau
nerf auditif
pavillon
cochlée
cavité du tympan
tympan
conduit auditif
étrier
trompe d'Eustache
fenêtre ronde

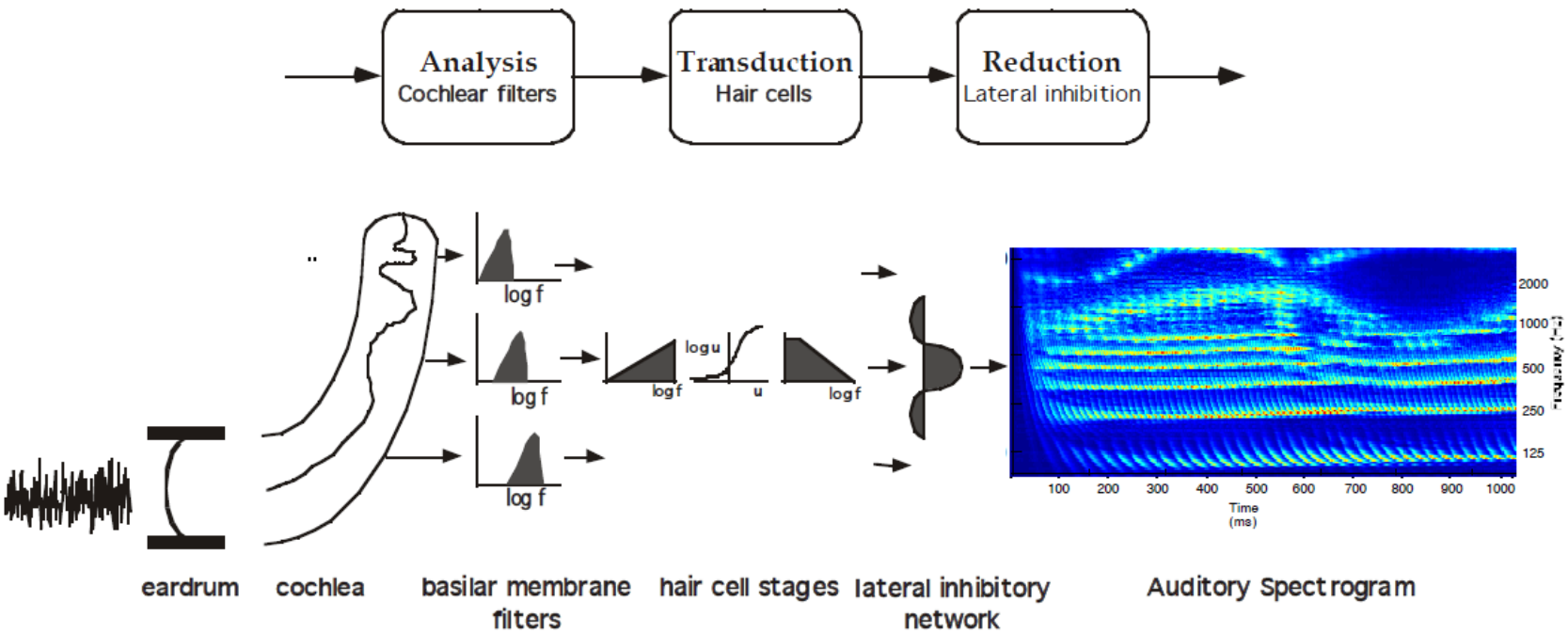
Tonotopic axis



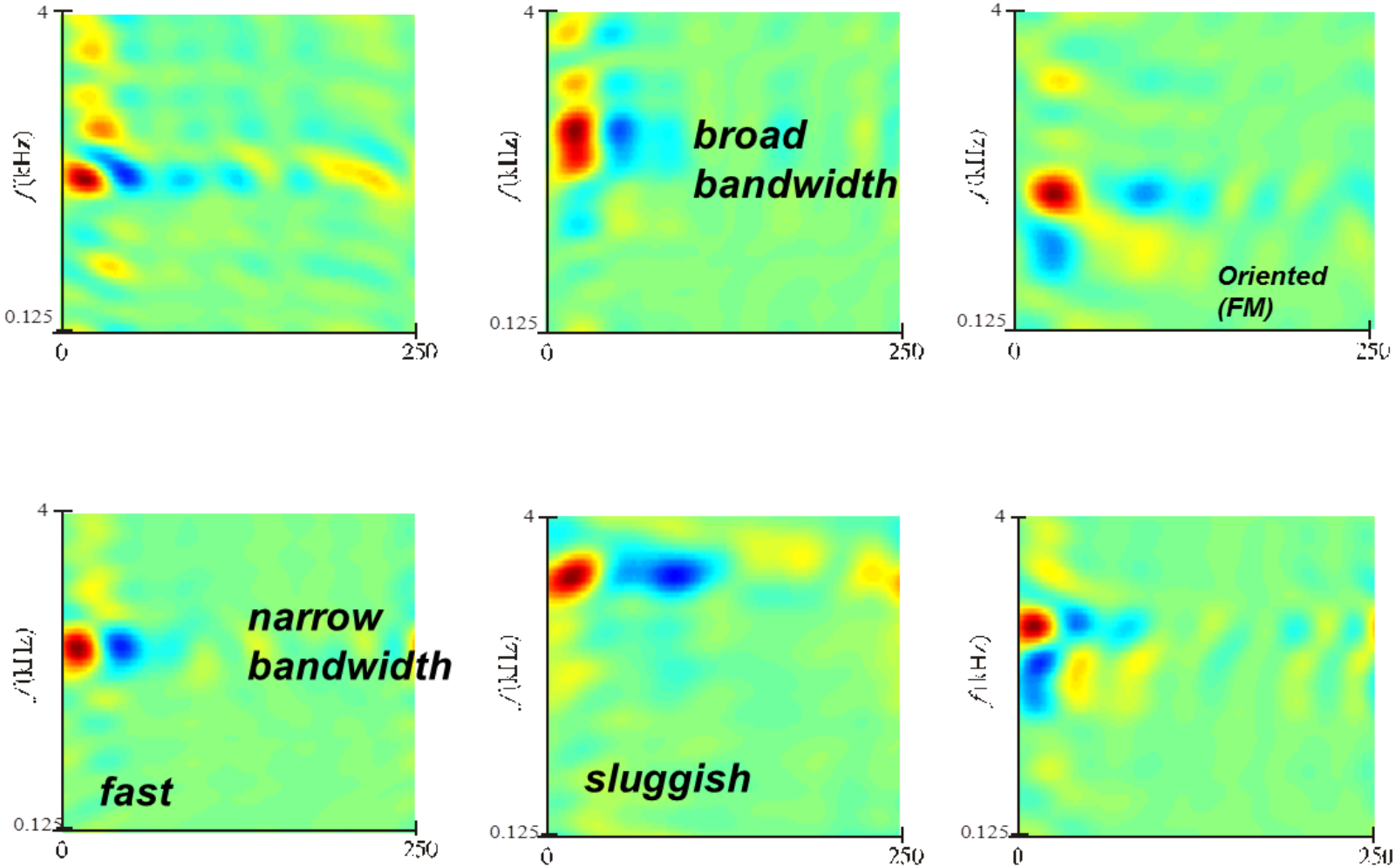
Inner Ear


Rampe vestibulaire
Membrane de Reissner
Canal cochléaire
Membrane tectoriale
Organe de Corti
Cellule ciliée interne
Cellules ciliées externes
Membrane basilaire
Rampe tympanique

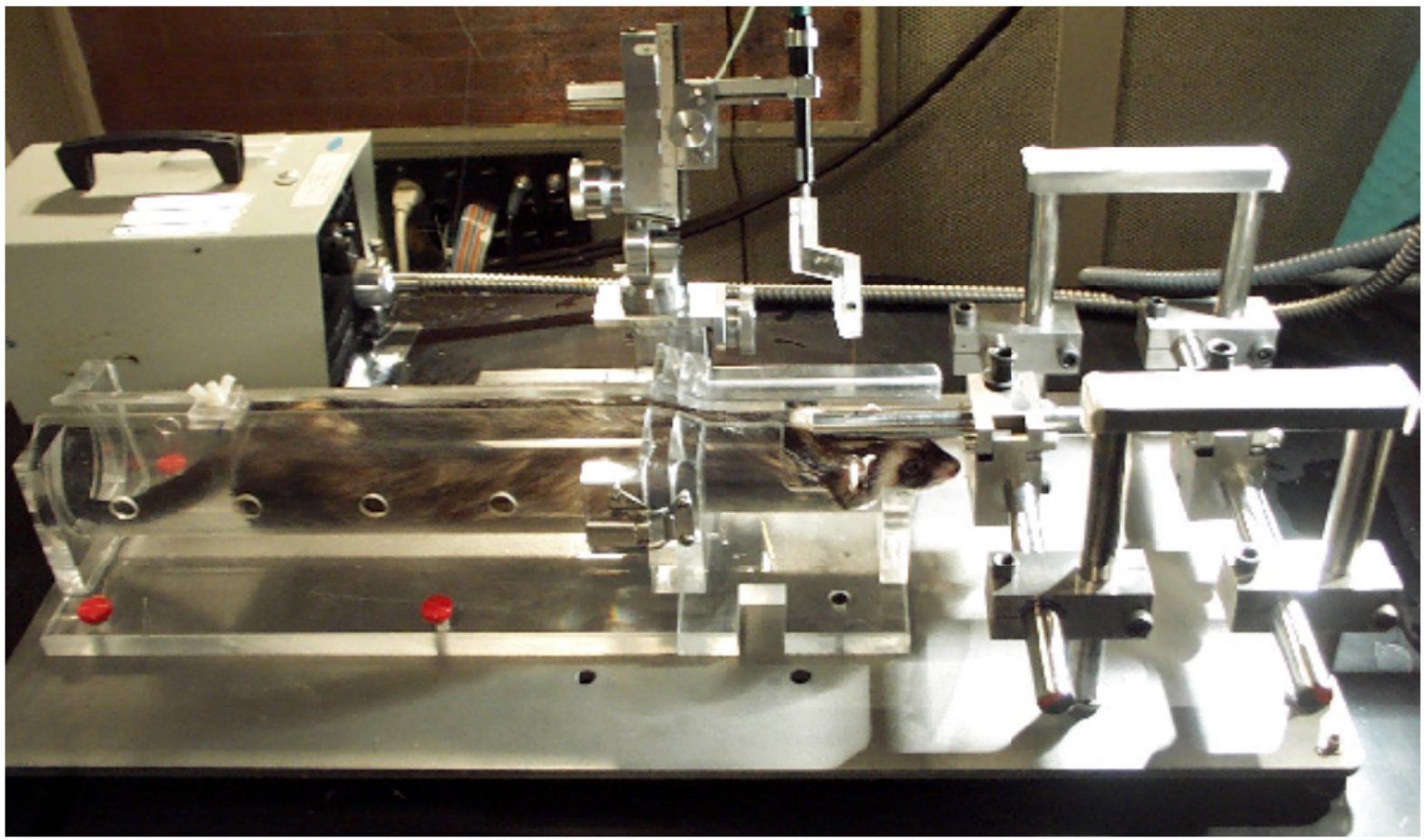
Early Auditory Processing Stages



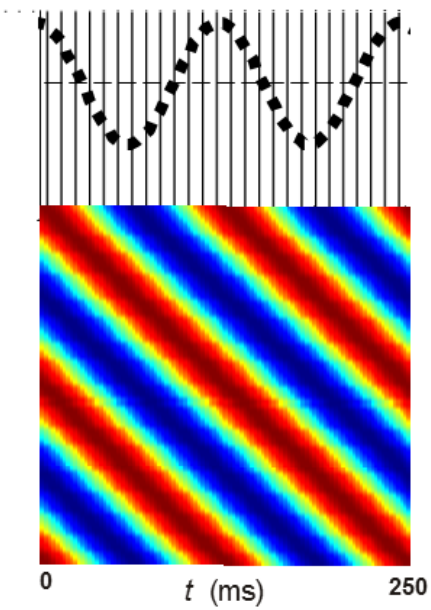
Spectro-Temporal Response fields (STRF) ...

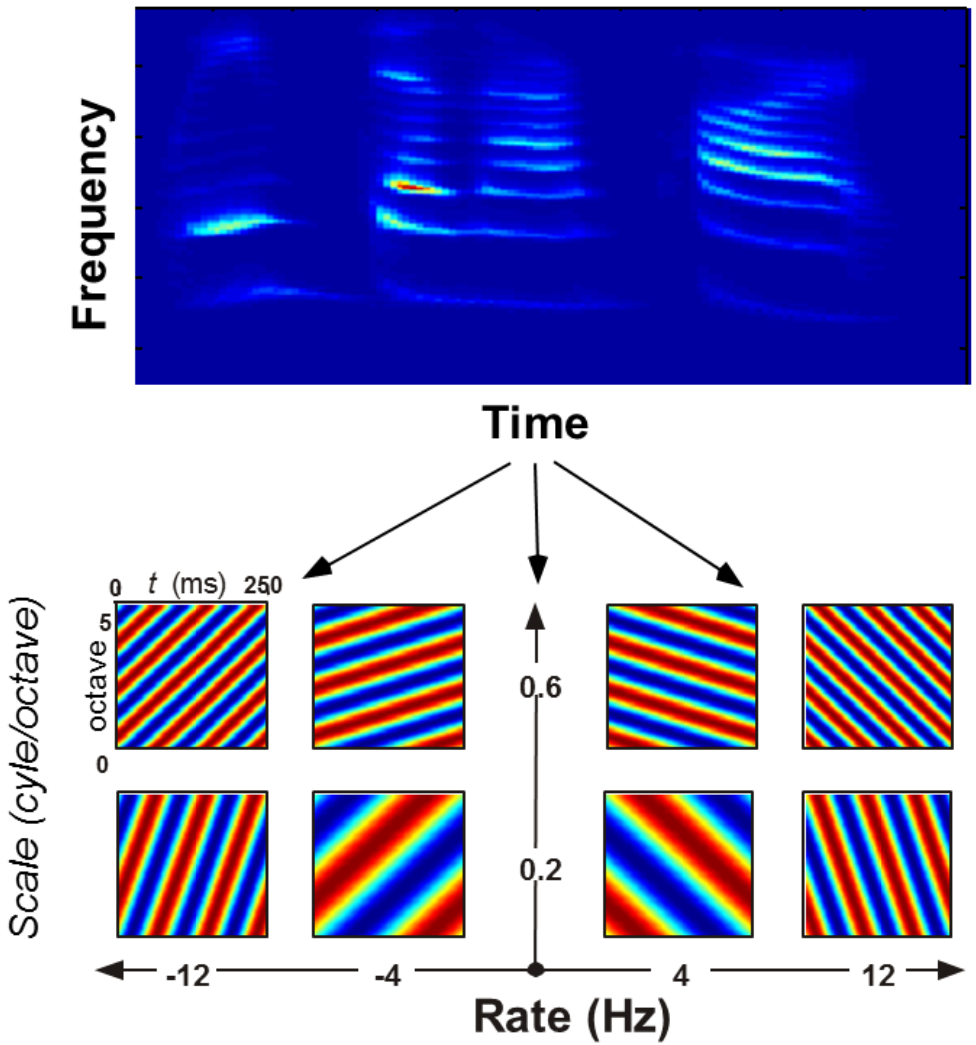


... of ferret

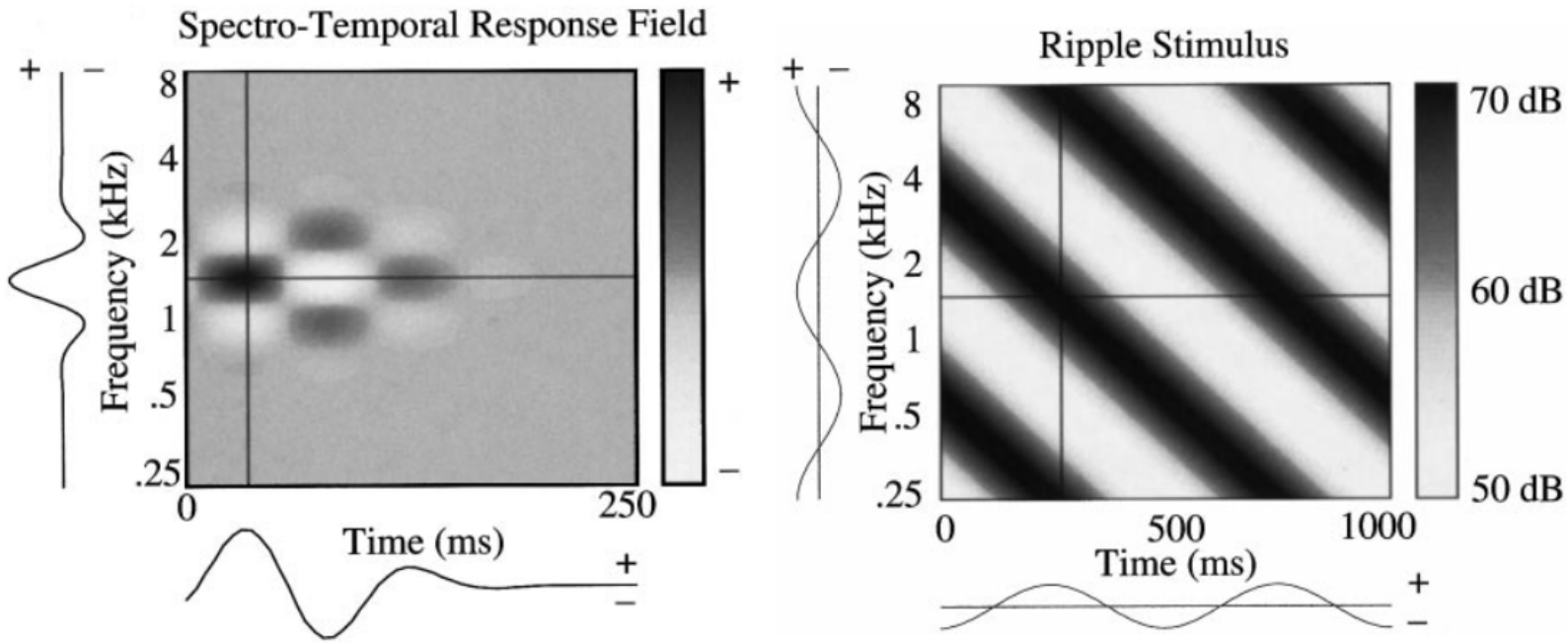


Spectro-Temporal Modulations





Spectro-Temporal Modulations



[Depireux et al., 2001]; see Shihab Shamma !

Cortical views of the spectrogram

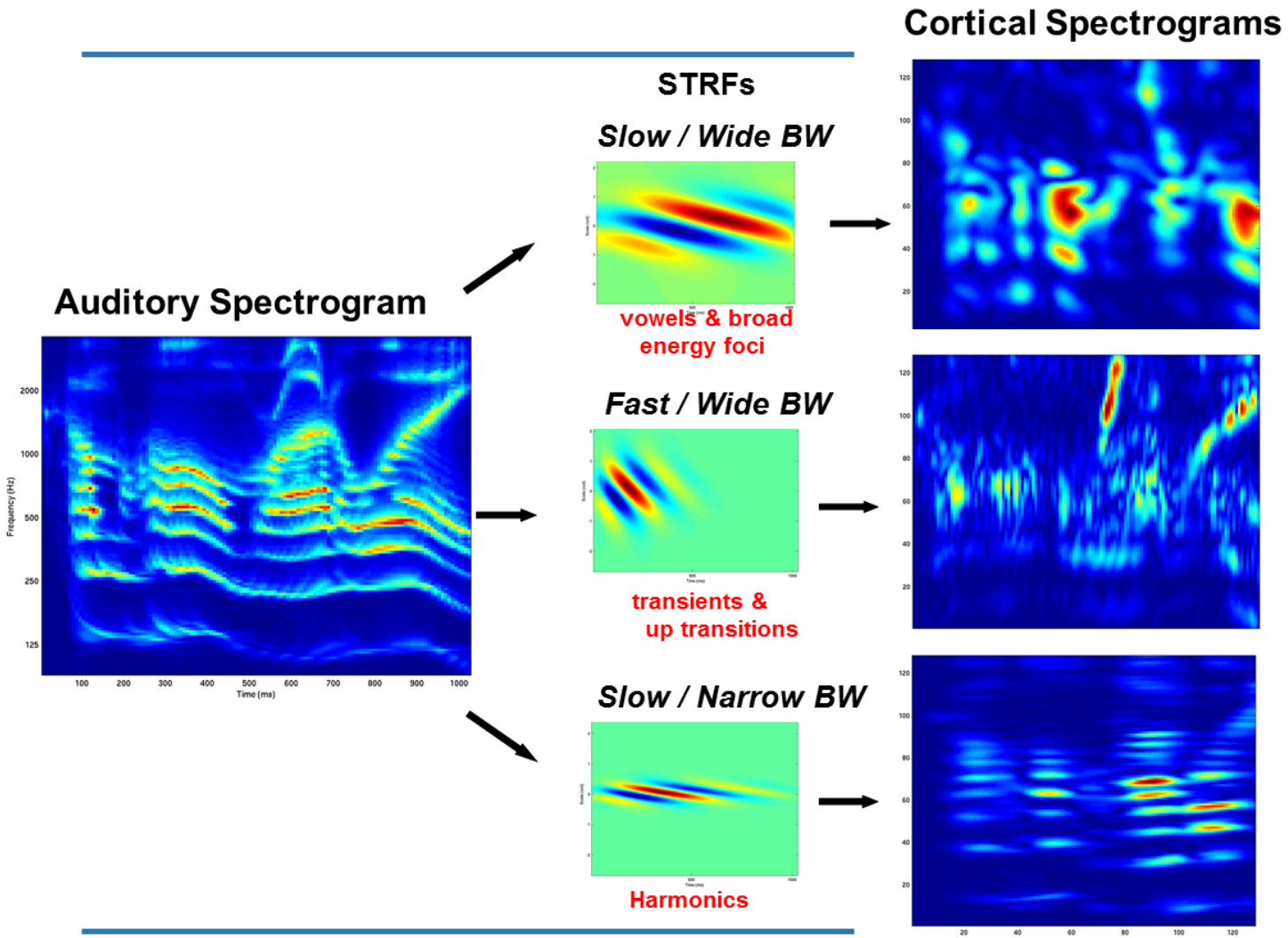
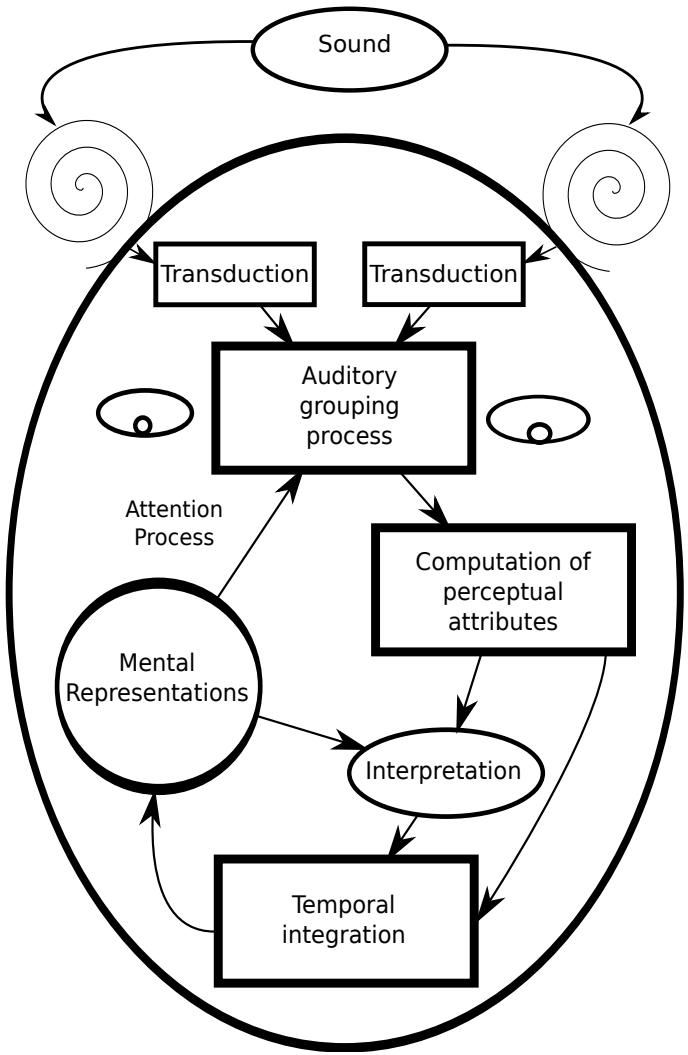


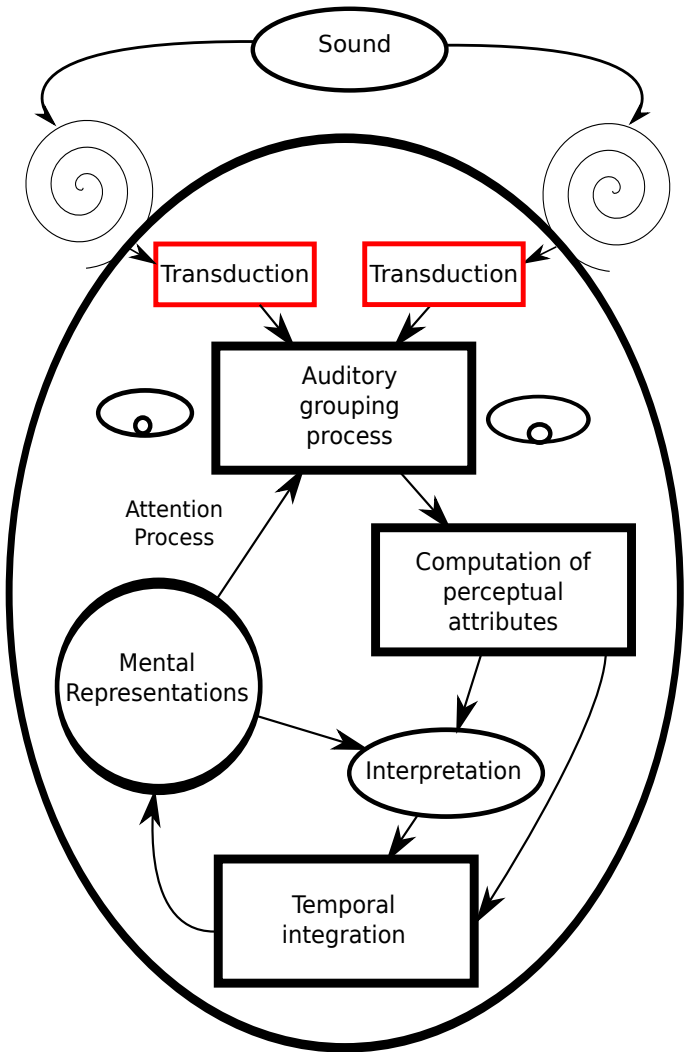
Table of Contents

Mental Processes Involved in Sound Perception



(adapted from [McAdams and Bigand, 1994])

Mental Processes Involved in Sound Perception

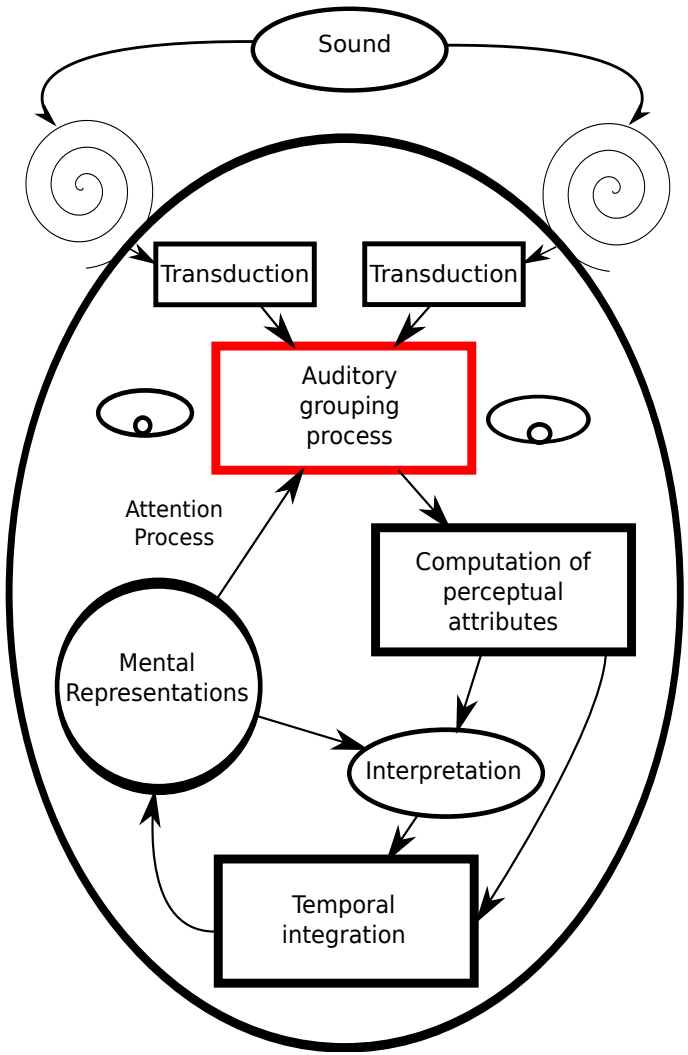


(adapted from [McAdams and Bigand, 1994])

Transduction

Hair cells in the cochlea transduce sounds into electrical signals. [Nelken, 2008]

Mental Processes Involved in Sound Perception

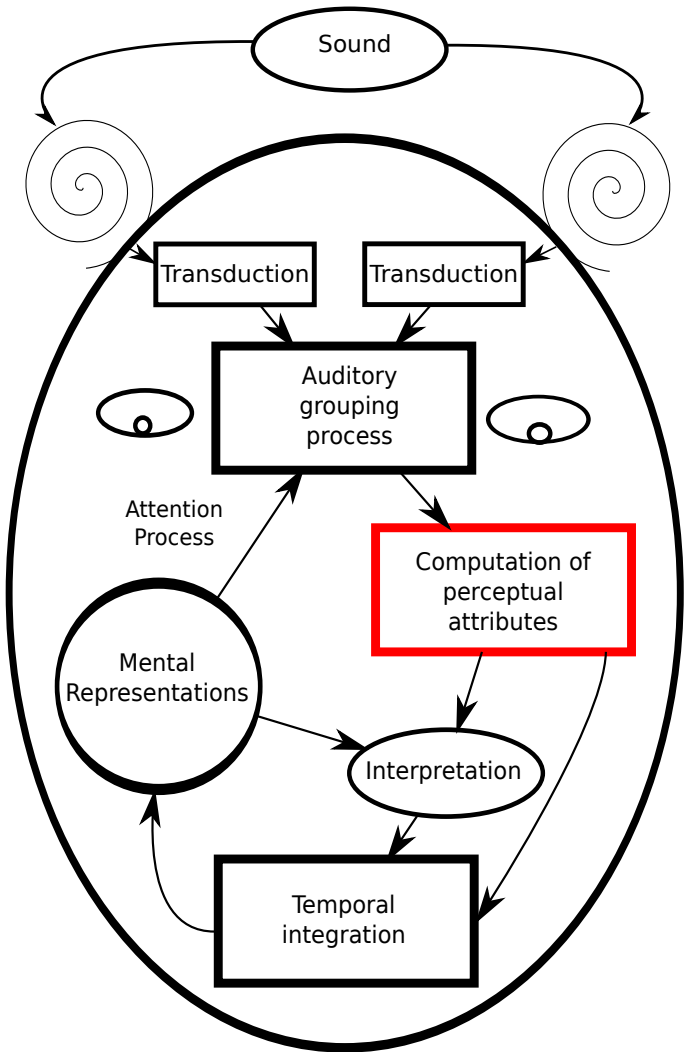


(adapted from [McAdams and Bigand, 1994])

Auditory grouping process

- sequential grouping
- simultaneous grouping

Mental Processes Involved in Sound Perception

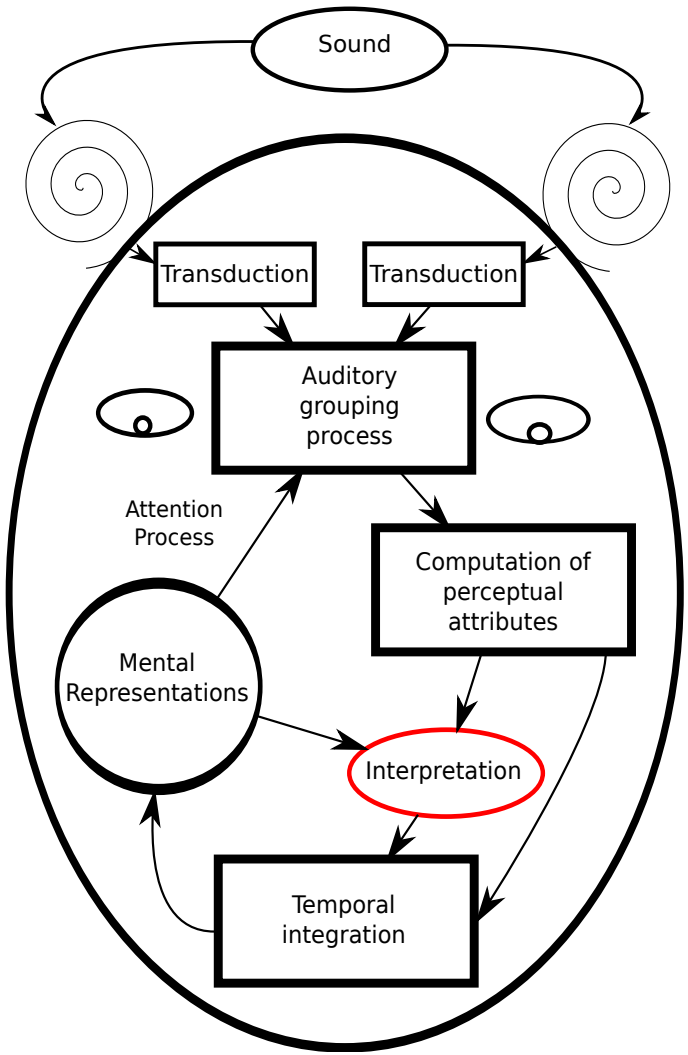


(adapted from [McAdams and Bigand, 1994])

Computation of perceptual attribute

- loudness
- timbre
- pitch
- ...

Mental Processes Involved in Sound Perception

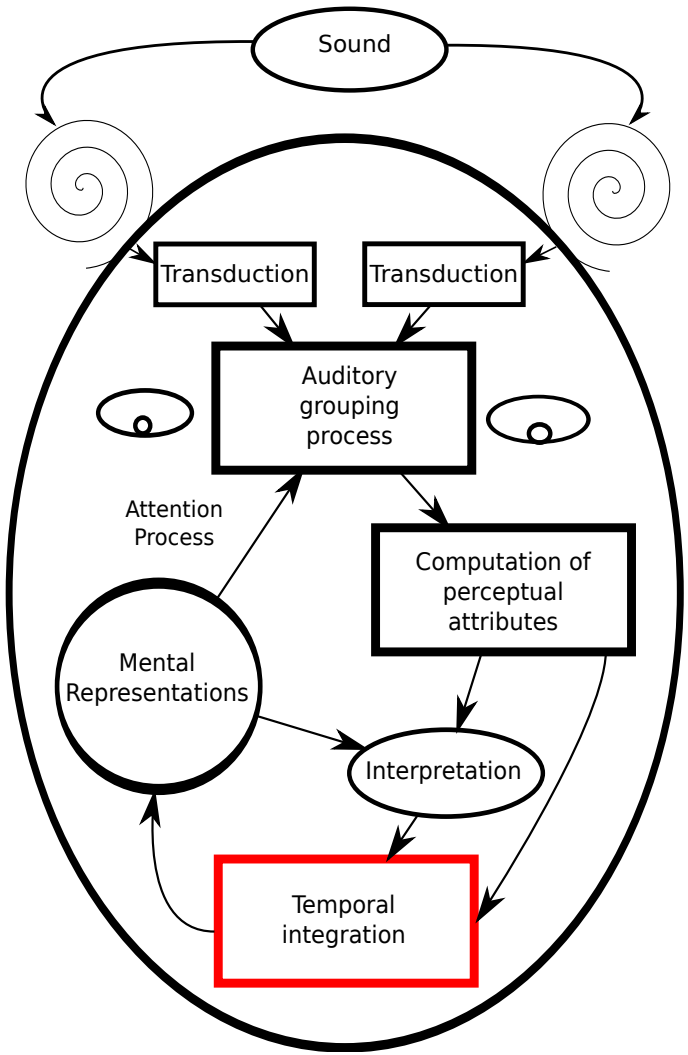


(adapted from [McAdams and Bigand, 1994])

Interpretation

- identification
- qualitative evaluation
- signification

Mental Processes Involved in Sound Perception

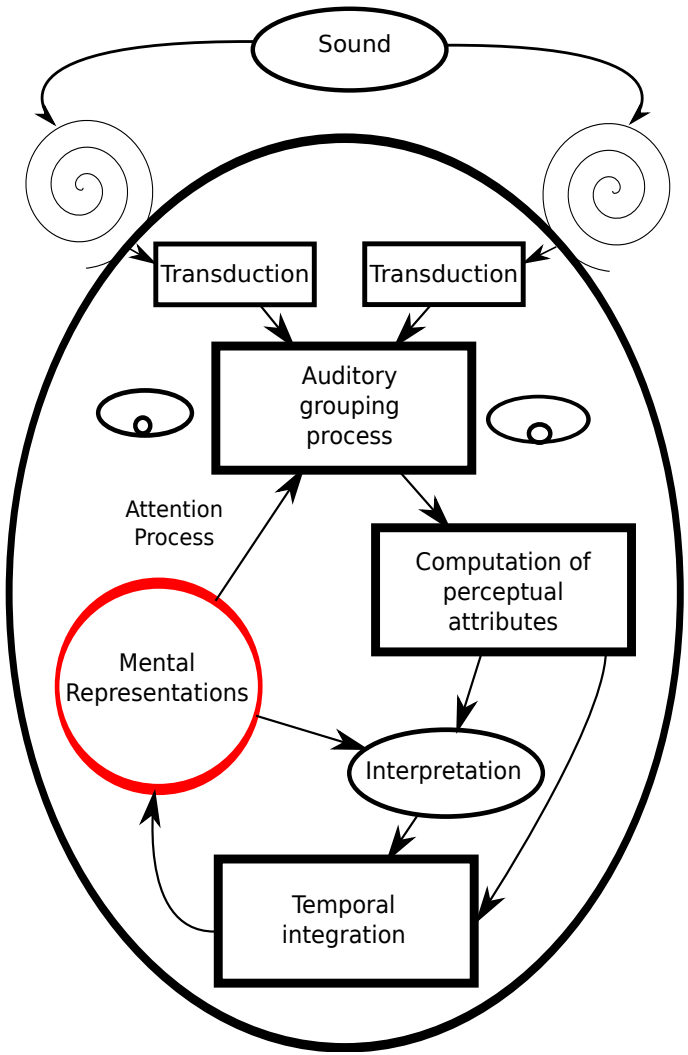


(adapted from [McAdams and Bigand, 1994])

Temporal integration

- Interpretation of large temporal variation (music performance) [McAdams and Bigand, 1994]

Mental Processes Involved in Sound Perception



(adapted from [McAdams and Bigand, 1994])

Mental Representation

- Storing past information to optimize future processes

Nature of the stored information

- semantic [Dubois et al., 2006]
- statistic [McDermott et al., 2013]
- signal [Agus et al., 2010]

Two types of processes

Two types of processes

Bottom-up process

- Sensory input
- Data-driven

Two types of processes

Bottom-up process

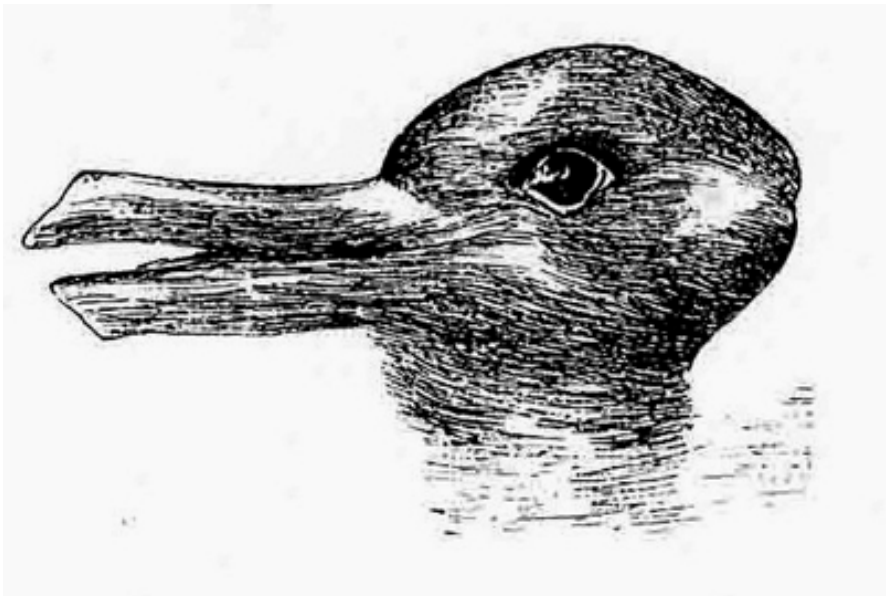
- Sensory input
- Data-driven

Top-down process

- Retroactive process
- Concept- / Knowledge-driven
- Subject Individuality

Two types of processes

Bi-stability phenomenon – exclusive allocation principle



how do we listen to?

how do we listen to?

According to Gaver, there is 2 main modes of listening

- every day listening
- musical listening

how do we listen to?

According to Gaver, there is 2 main modes of listening

- every day listening
- musical listening

They can be reformulated as

- **holistic** listening: fast screening based on pattern matching (low power processes)
- **analytical** listening: intensive search of correlation between various cues (high power processes)

What are we searching for ?

According to Schaeffer [1966], we can interpret the acoustic scene according to three different levels of similarity:

What are we searching for ?

According to Schaeffer [1966], we can interpret the acoustic scene according to three different levels of similarity:

- Acoustic: similarity of acoustical properties

What are we searching for ?

According to Schaeffer [1966], 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

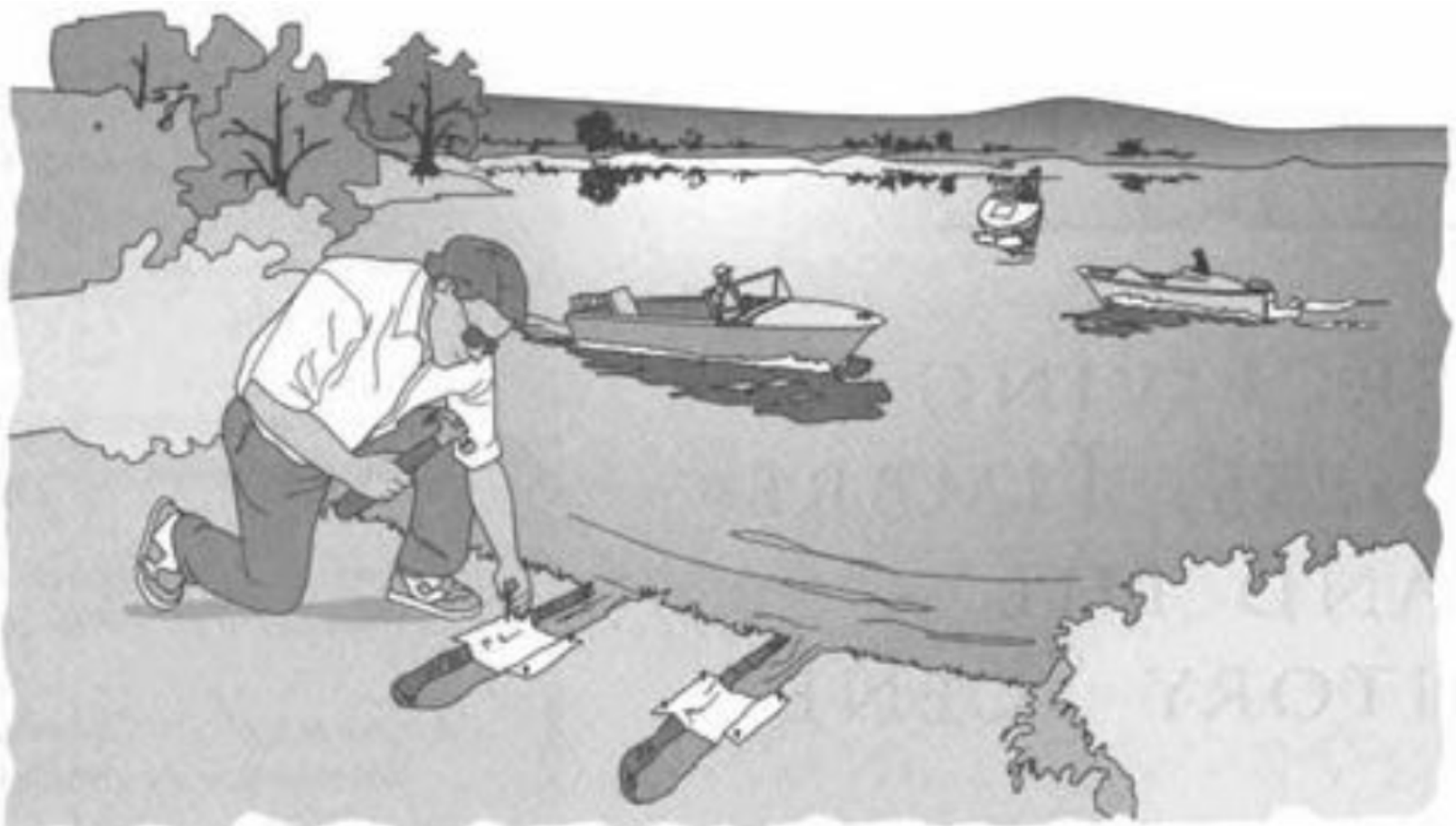
What are we searching for ?

According to Schaeffer [1966], 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 identified objects or event

Table of Contents

Auditory Scene Analysis (ASA)



ASA?

What is not ASA

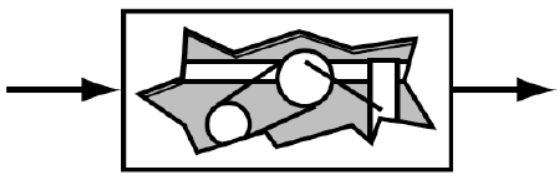
:

ASA?

What is not ASA

:

- Physiology: implementation

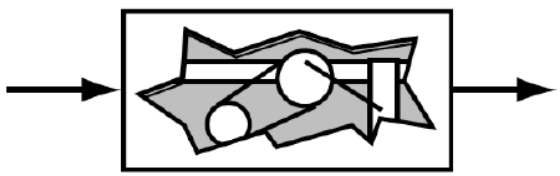


ASA?

What is not ASA

:

- Physiology: implementation



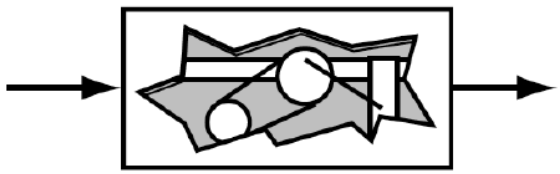
- Psycho-physics: function/behavior

ASA?

What is not ASA

:

- Physiology: implementation



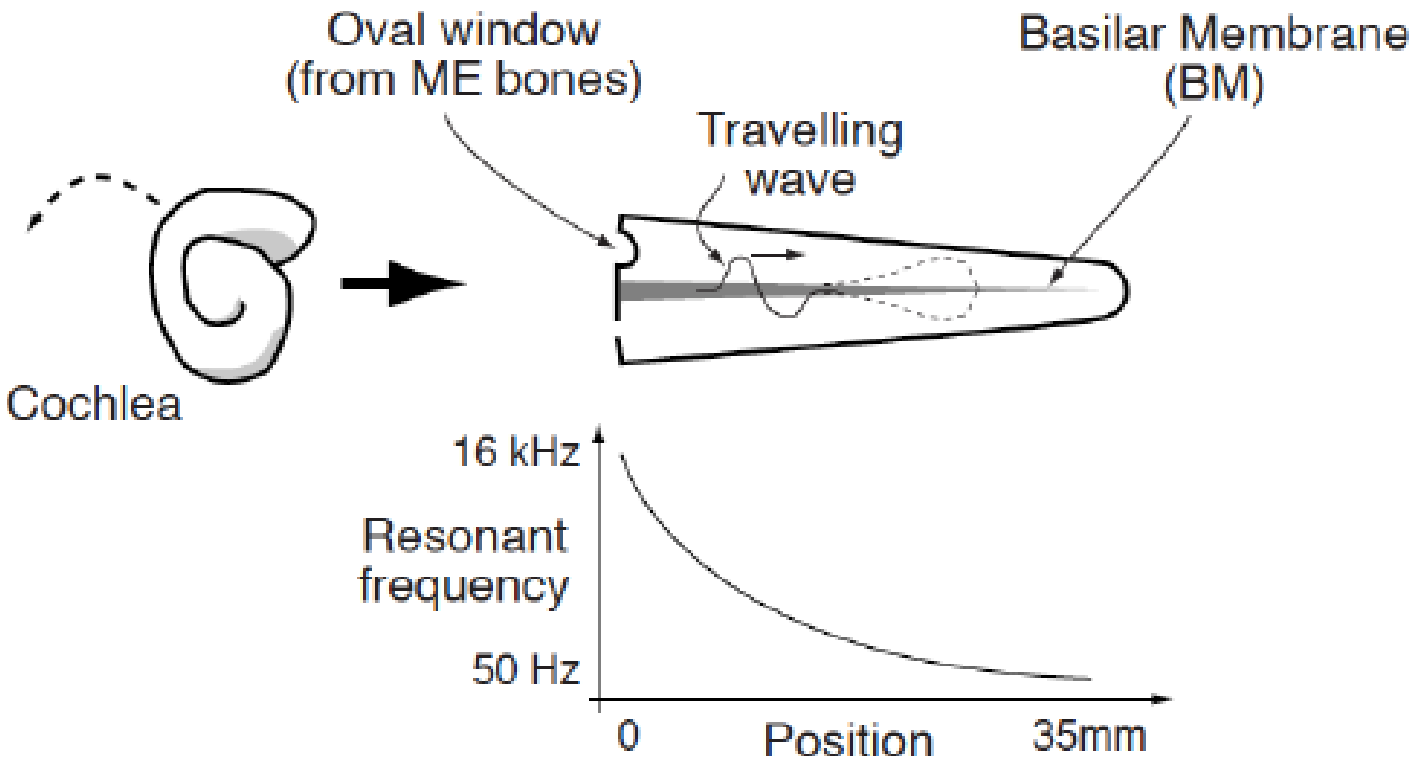
- Psycho-physics: function/behavior

ASA looks at:

- Information processing models

Physiology

Inner Ear

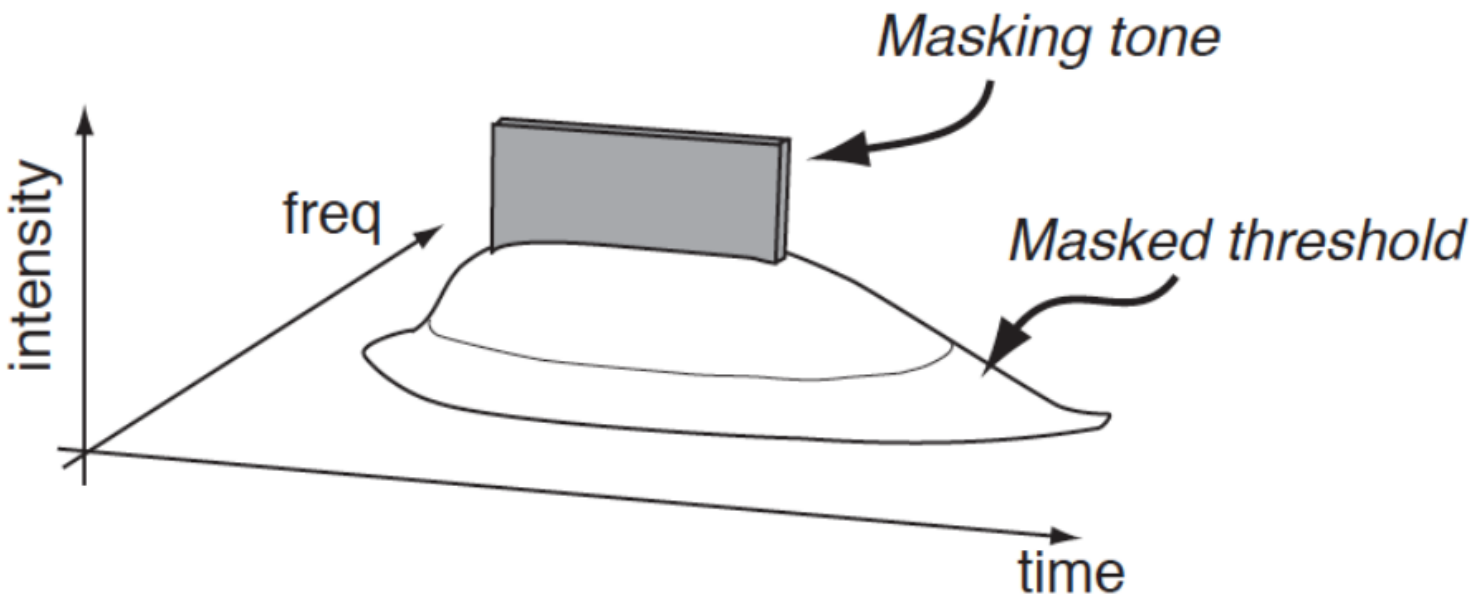


Psycho-physic

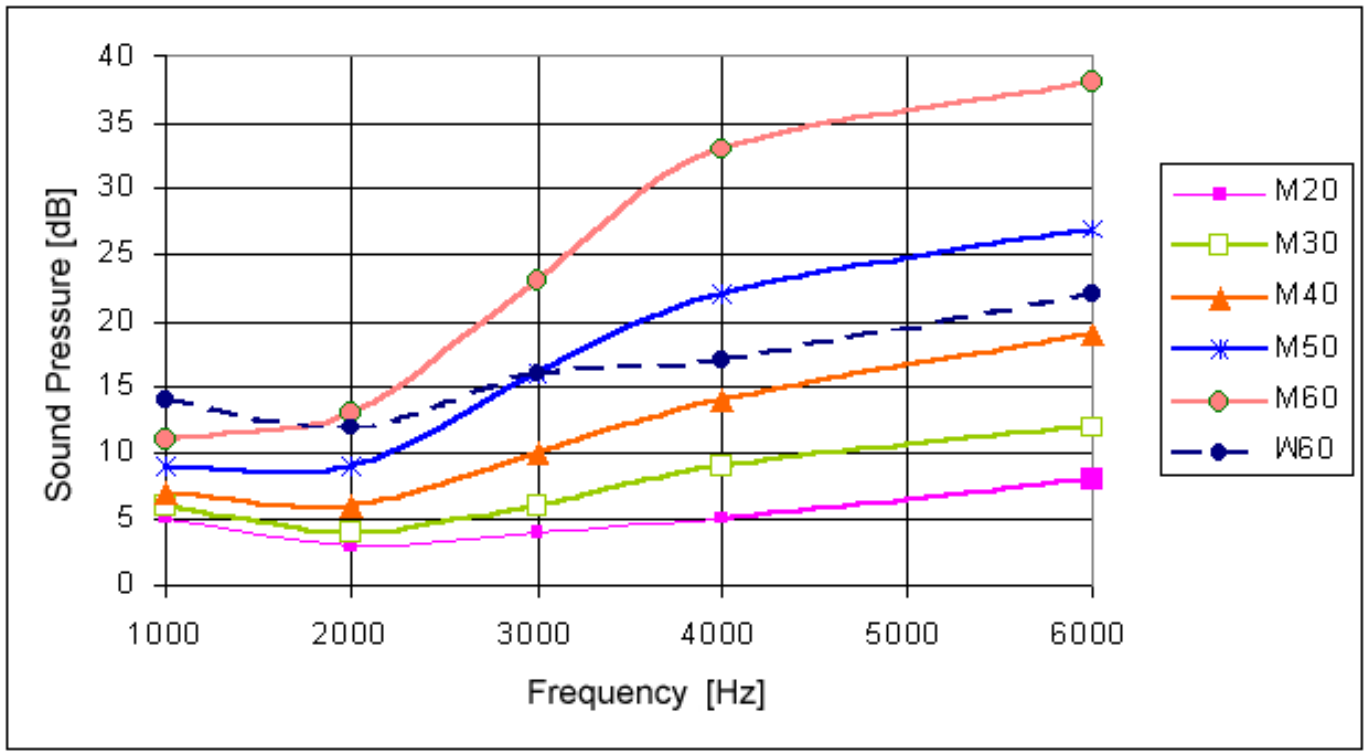
Relate physical and perceptual variables

- intensity → loudness
- frequency → pitch

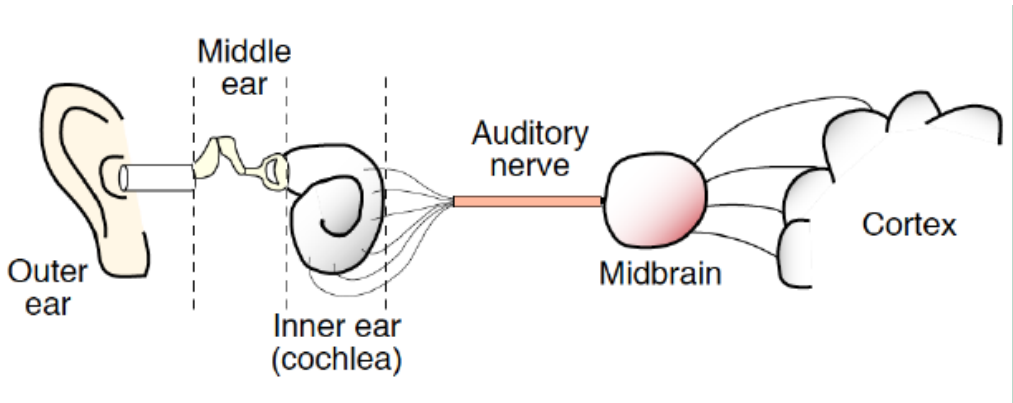
Time/Frequency Masking



Threshold of hearing



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

The Auditory Scene Analysis problem

- Introduced by Bregman [1990]

The Auditory Scene Analysis problem

- Introduced by Bregman [1990]
- **Hypothesis**:

  "If we are to make sense of the auditory world and interact with it effectively, it is necessary for the brain to isolate the information relating to different sound sources" Winkler et al. [2009]

What is a stream?

- A stream is a "perceptual entity"

What is a stream?

- A stream is a "perceptual entity"
- It can incorporate more than one sound (footsteps, a soprano with a piano)

What is a stream?

- A stream is a "perceptual entity"
- It can incorporate more than one sound (footsteps, a soprano with a piano)
- It depends on our perceptual representation of the world (Urban dweller in a jungle)

Auditory Streaming

"A perceptual phenomenon in which a sequence of sounds is perceived as consisting of two or more auditory streams." Winkler et al. [2009]

Auditory Streaming

"A perceptual phenomenon in which a sequence of sounds is perceived as consisting of two or more auditory streams." Winkler et al. [2009]

Two Processes involved in stream segregation:

- Primitive (innate constraints)
- Schema-based (learned constraint)

Primitive Process (*bottom-up*)

- **Innate constraints**: there is constant properties of the environment that are dealt with by every human everywhere

Primitive Process (*bottom-up*)

- **Innate constraints**: there is constant properties of the environment that are dealt with by every human everywhere
- "When a harmonically structured sound changes over time, all the harmonics in it will tend to change together" [Bregman, 1990]

Primitive Process (*bottom-up*)

- **Primitive Process**: Generic rules that arise out of our experience of regularities in the auditory world [Ballas and Howard, 1987]

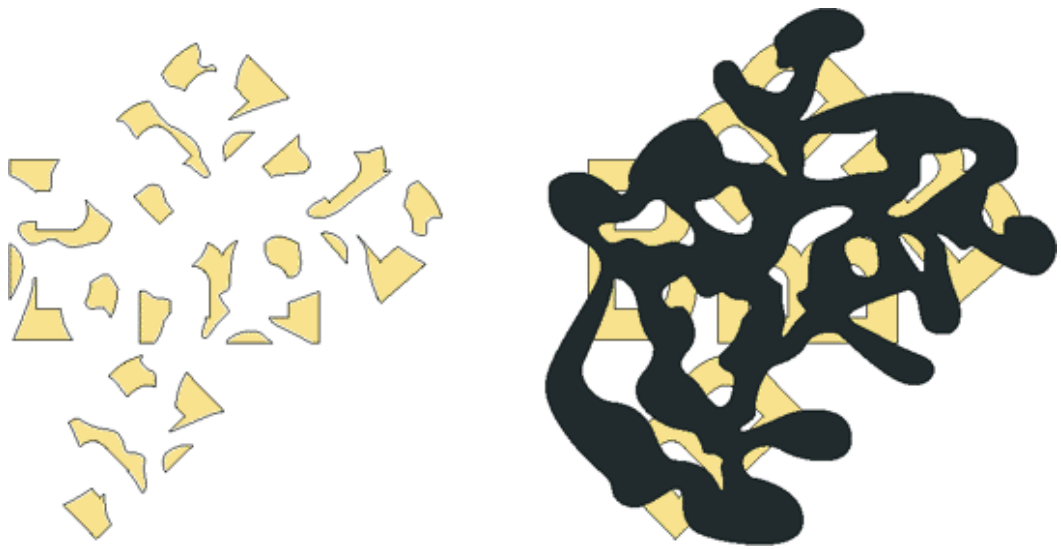
Primitive Process (*bottom-up*)

- **Primitive Process**: Generic rules that arise out of our experience of regularities in the auditory world [Ballas and Howard, 1987]
- Rules are similar to the Gestalt principles of visualization organization [Ballas and Howard, 1987]

Gestalt-like principles: Closure

The mind may experience elements it does not perceive through sensation, in order to complete a regular figure

Gestalt-like principles:Continuity



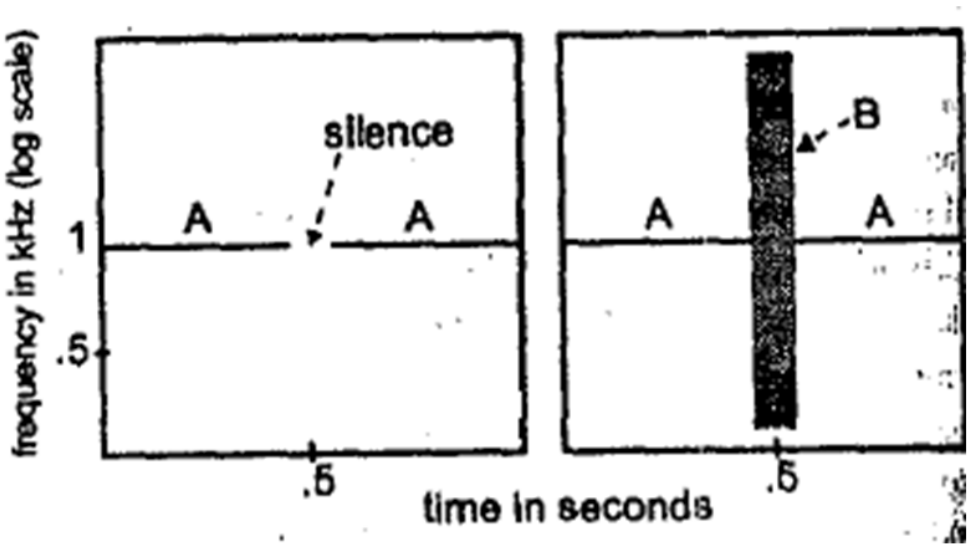
Adding occlusion causes the fragments on the boundaries form to be grouped

Continuity: does it works with sounds?

Perceived continuity

: Sine tone and burst of noise (Warren 1984)

- Apparent continuity

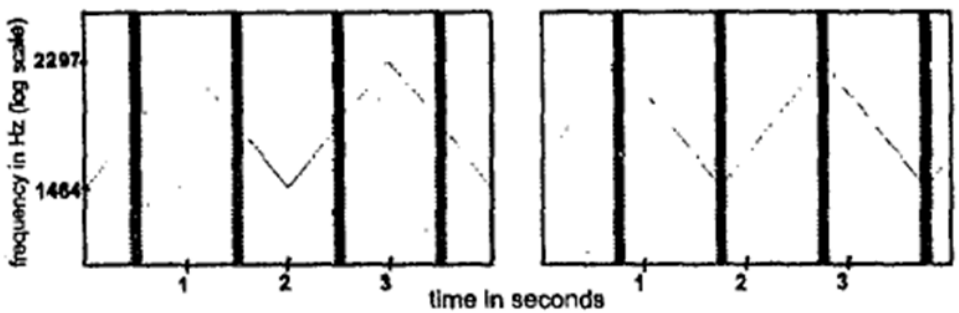


Continuity: does it works with sounds?

Perceived continuity

Sine tone and burst of noise (Warren 1984)

- Perceptual continuation of a gliding tone through a noise burst



Continuity: does it works with sounds?

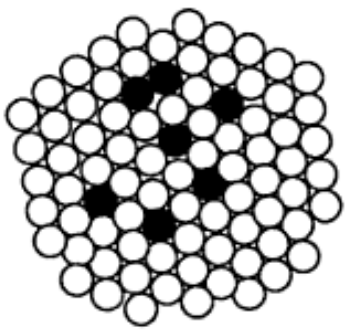
Perceived continuity

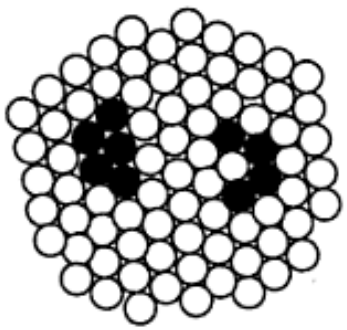
: Sine tone and burst of noise (Warren 1984)

- Picket fence effect

Gestalt-like principles

Similarity and Proximity [Bregman, 1990]

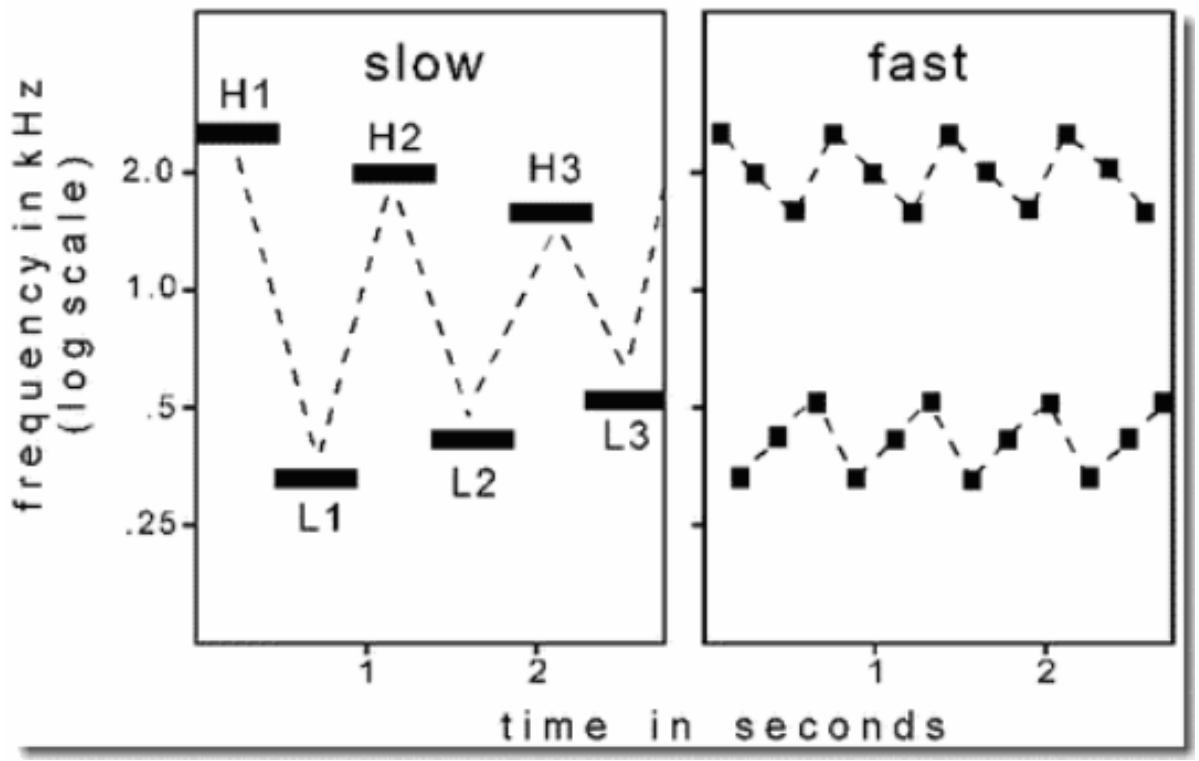




Does it work with sounds ? 

Tones similarity [a]

[a] http://webpages.mcgill.ca/staff/Group2/abregm1/web/



At a rapid rate, the top three notes will be heard as one stream, and the bottom three as another

Grouping strategies

There is two grouping strategies:

- sequential grouping
- simultaneous grouping

Grouping strategies

There is two grouping strategies:

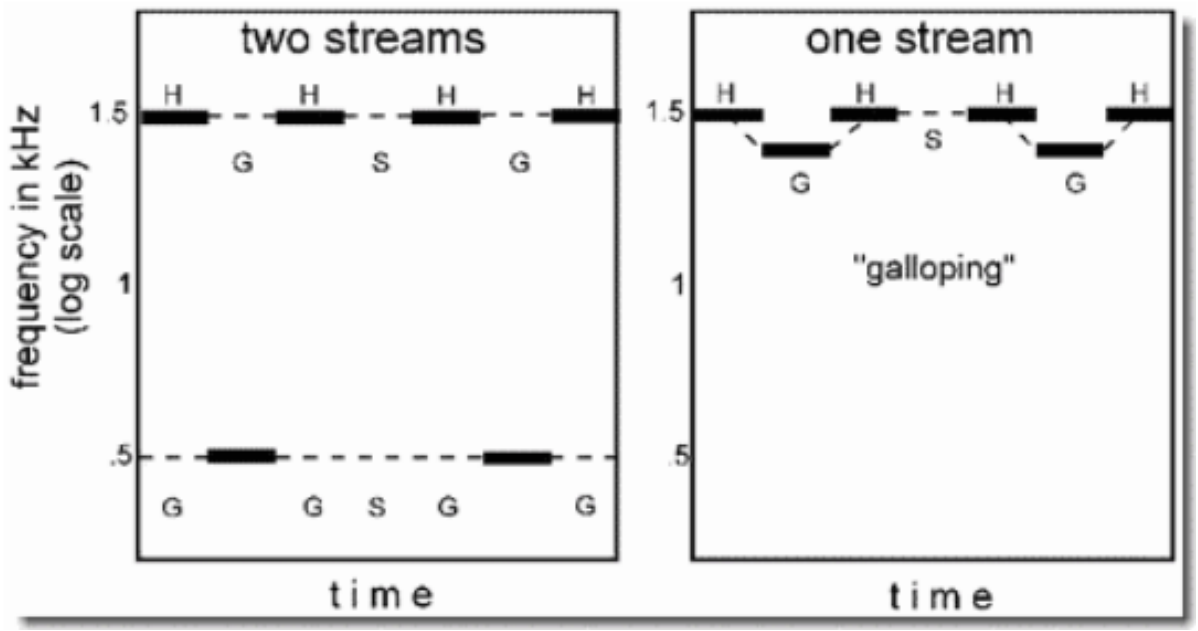
- sequential grouping
- simultaneous grouping

Both are based on:

- acoustic cues: common onset, spectral regularity, spectral harmonicity
- perceptual attributes: timbre, loudness, pitch, duration ...

Sequential Grouping

Linking together sounds, whose onsets are separated in time

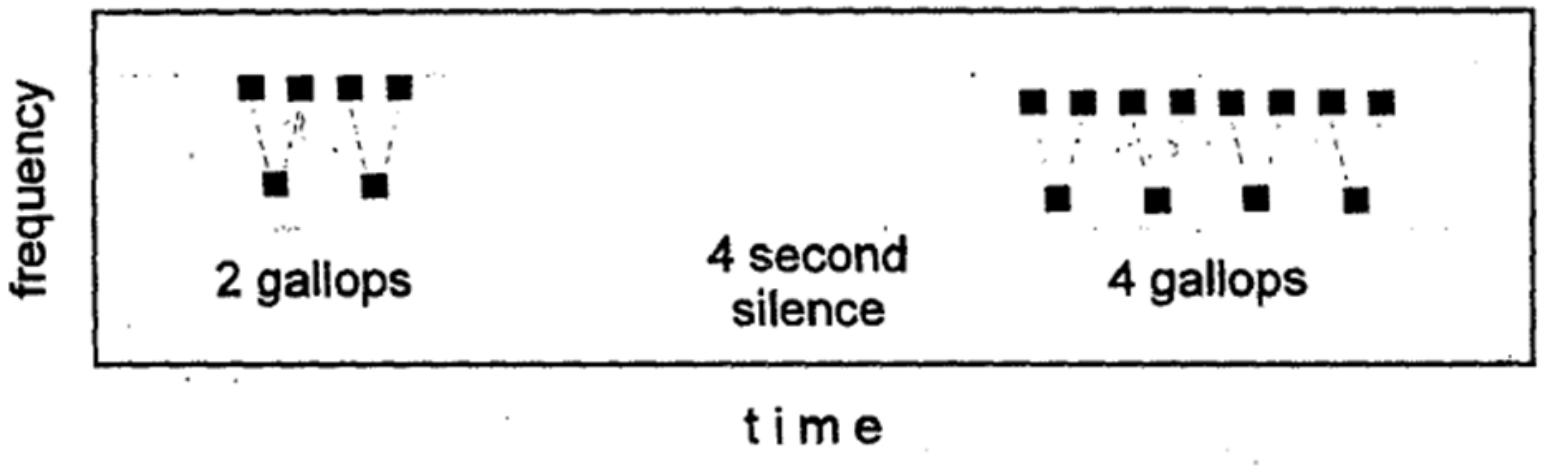


- importance of speed
- importance of frequency

[0] http://webpages.mcgill.ca/staff/Group2/abregm1/web/

Sequential Grouping

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

[0]http://webpages.mcgill.ca/staff/Group2/abregm1/web/

Simultaneous Grouping (spectral grouping)

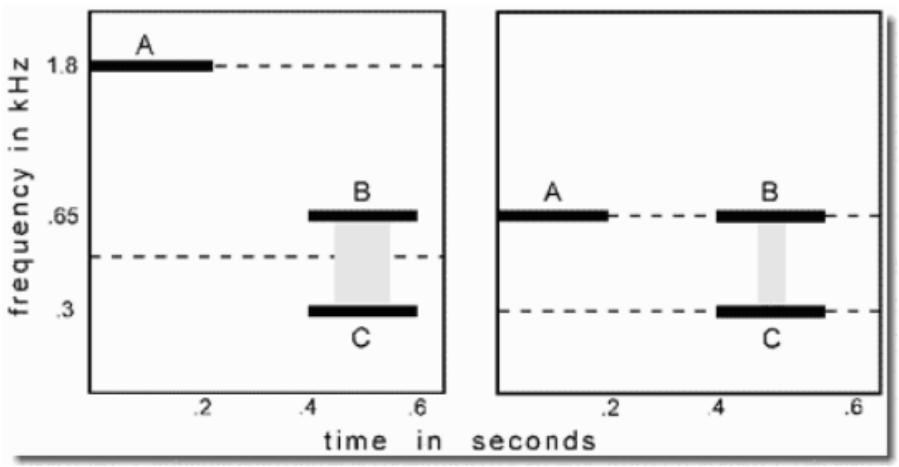
Assign concurrently active features to one or more objects

- grouping by harmonicity 

[0] http://webpages.mcgill.ca/staff/Group2/abregm1/web/

The Old-Plus-New Heuristic

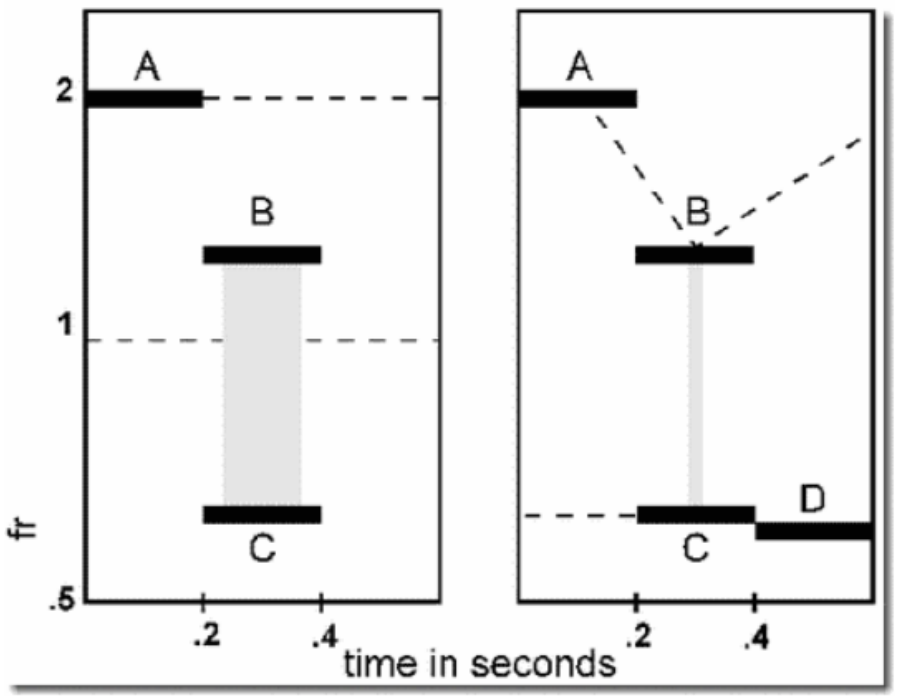
If a spectrum becomes suddenly more complex or more intense, the auditory system tries to interpret this as a continuing old sound joined by a new one.



- spectral fusion
- Old-Plus-New Heuristic

[0] http://webpages.mcgill.ca/staff/Group2/abregm1/web/

Sequential vs. Simultaneous Grouping



- spectral fusion
- sequential grouping

Sequential vs. Simultaneous Grouping

- Sequential grouping override the organization formed by simultaneous grouping

Sequential vs. Simultaneous Grouping

- Sequential grouping override the organization formed by simultaneous grouping
- "Ecologically this makes sense as most informative sounds, especially communication sounds, are intermittent, and it is necessary to form associations between events which may be separated in time by fairly long intervals" [Winkler et al., 2009]

Competition

In case of competition

- The winner is the grouping that considers the cues that the HAS prefers

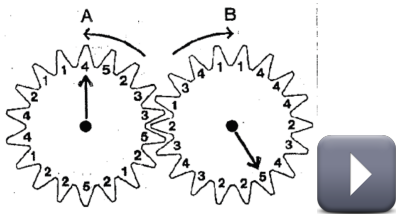
Though, this preference depends on many factors

- Prior, attention, context ...

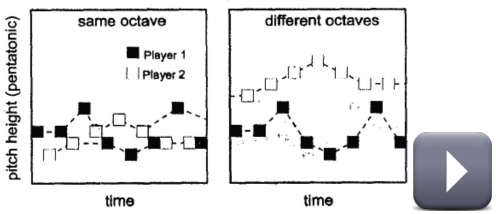
Competition

Illustration with xylophone duet

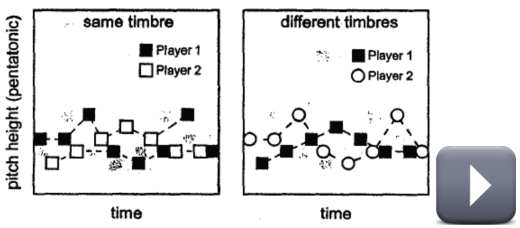
- Normal



- Change of pith range



- Change of timbre



Attention

Consider the High/Low experiments with varying speed and δ_{freq}

- Ask the listeners to integrate the sequence as much as possible
  - Trade-off between between speed and δ_{freq}
- Ask the listeners to segregate the sequences as much as possible
  - Trade-off between between speed and δ_{freq}

Evidence that some primitive mechanisms can be controlled up to a certain level

schema based processing

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

schema based processing

How did we learned such schemas in the real world?

Categorization Theory: semantic

Prototype theory [Rosch and Lloyd, 1978]

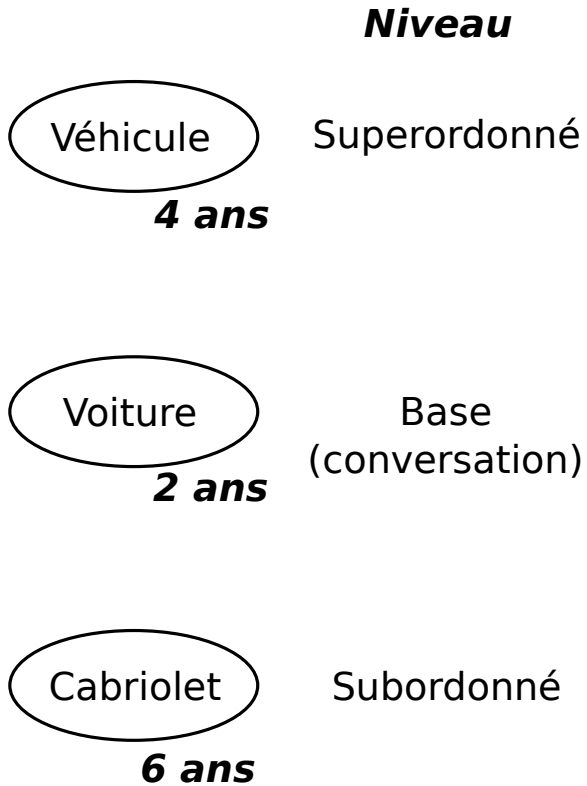
- categories are not defined by a set of precise rules
- items are related to one another through family resemblance
- family resemblance is defined by the similarity to a prototype

Categorization Theory: semantic

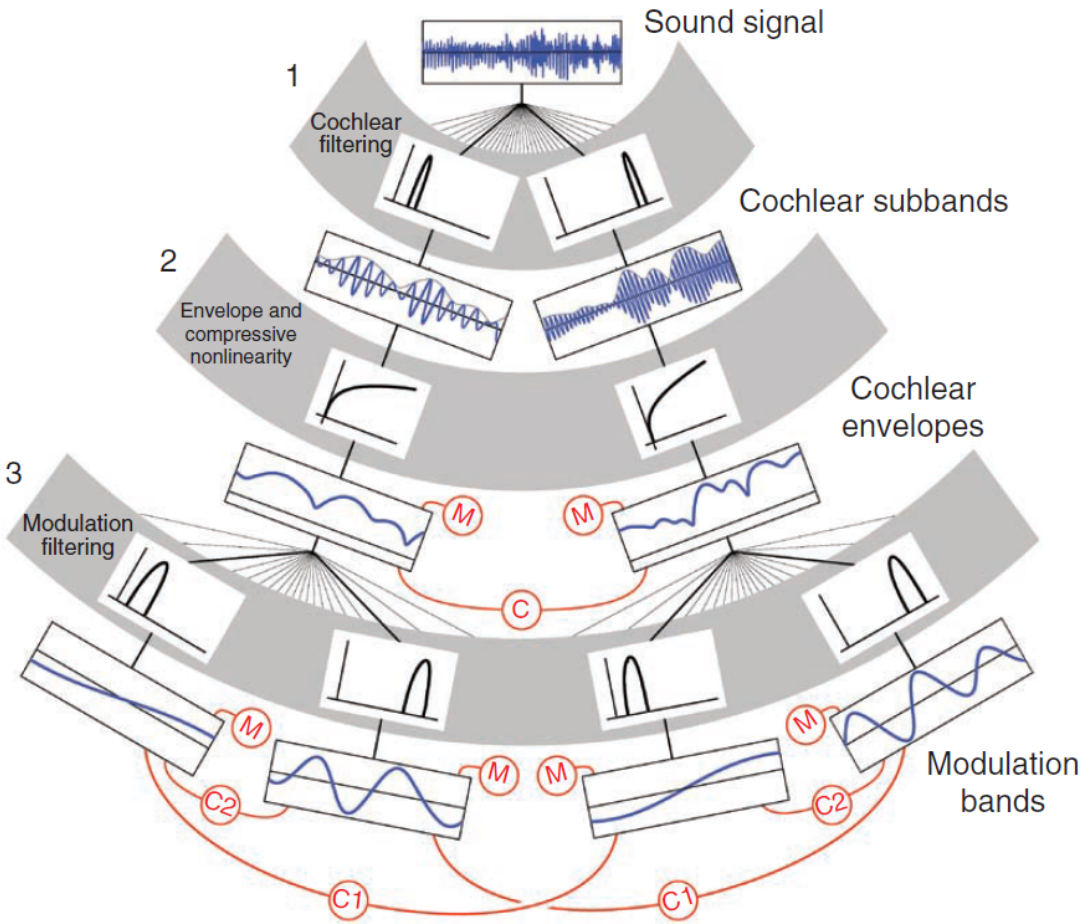
Prototype theory [Rosch and Lloyd, 1978]

- categories are not defined by a set of precise rules
- items are related to one another through family resemblance
- family resemblance is defined by the similarity to a prototype

Three levels
[Dubois, 1991]

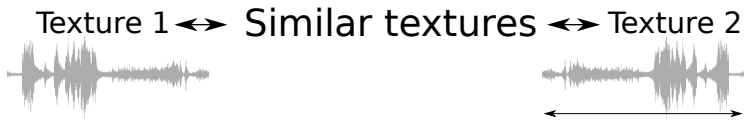


schema based processing: summary statistics

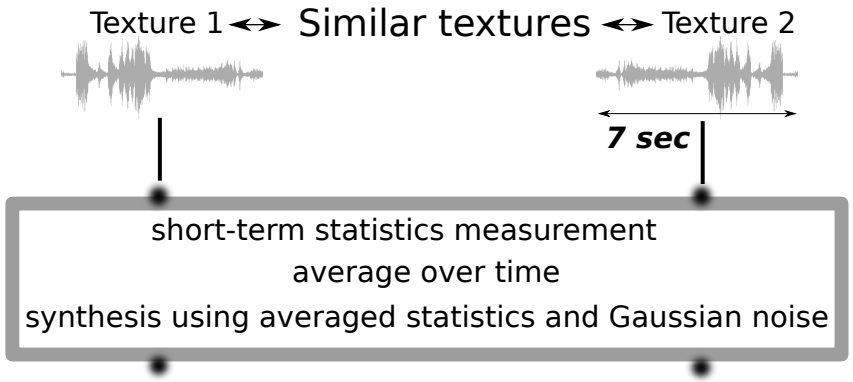


Texture Perception [McDermott and Simoncelli, 2011, McDermott et al., 2013]

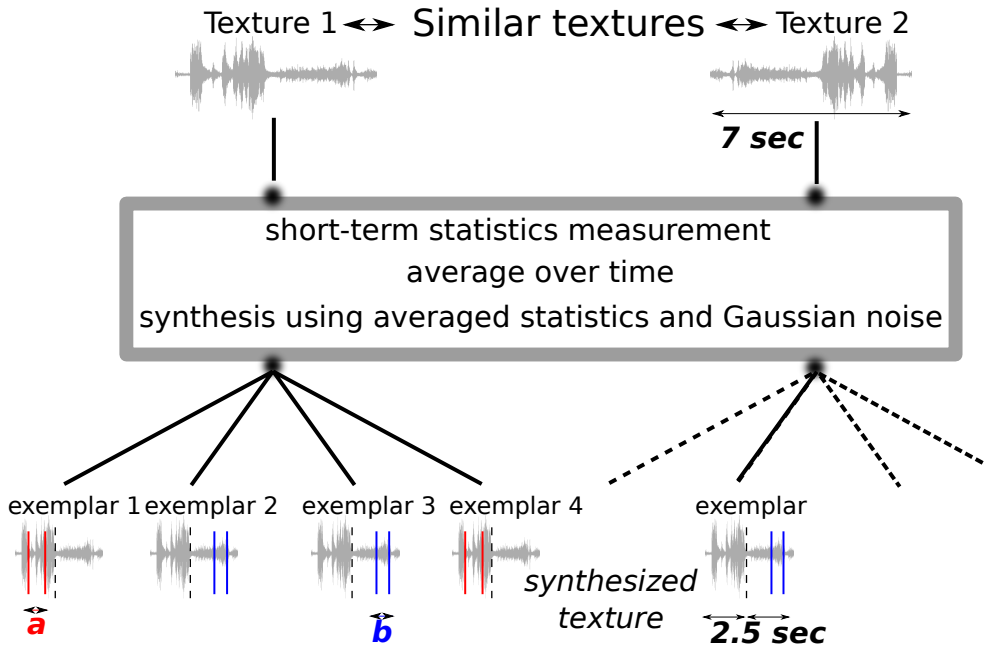
schema based processing: summary statistics



schema based processing: summary statistics

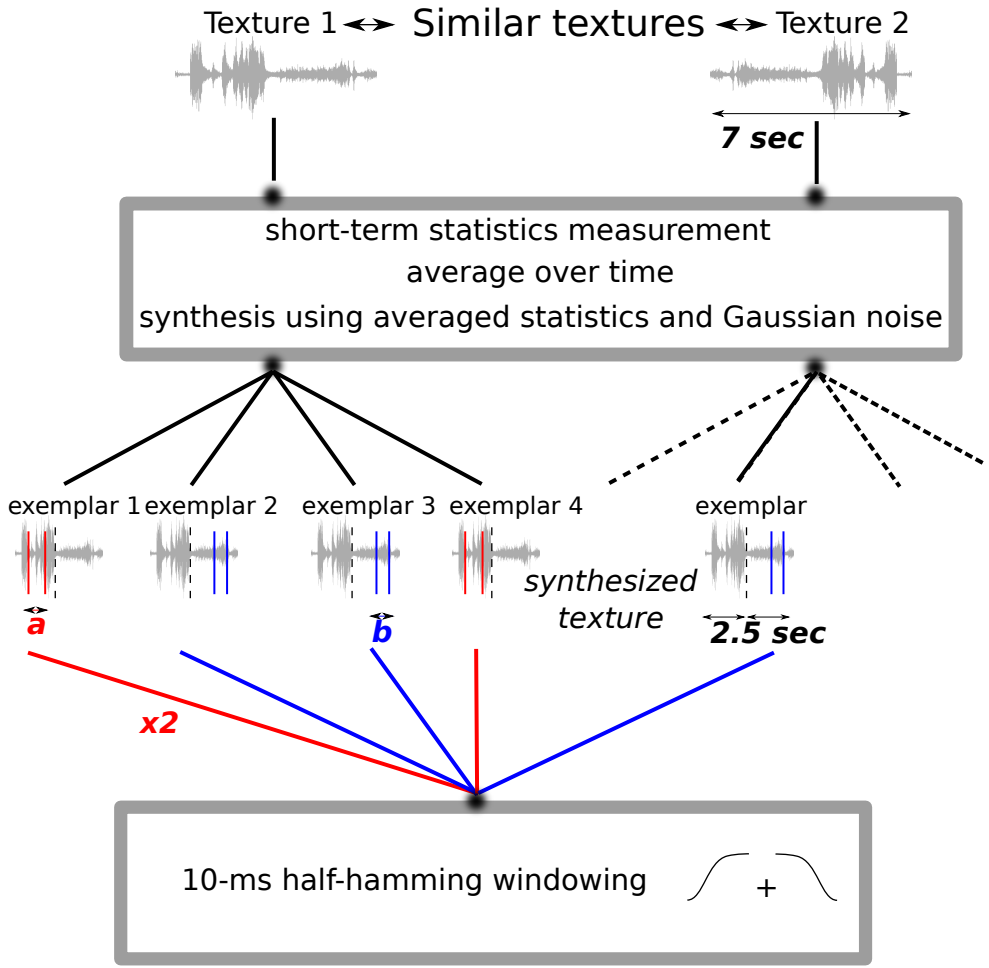


schema based processing: summary statistics



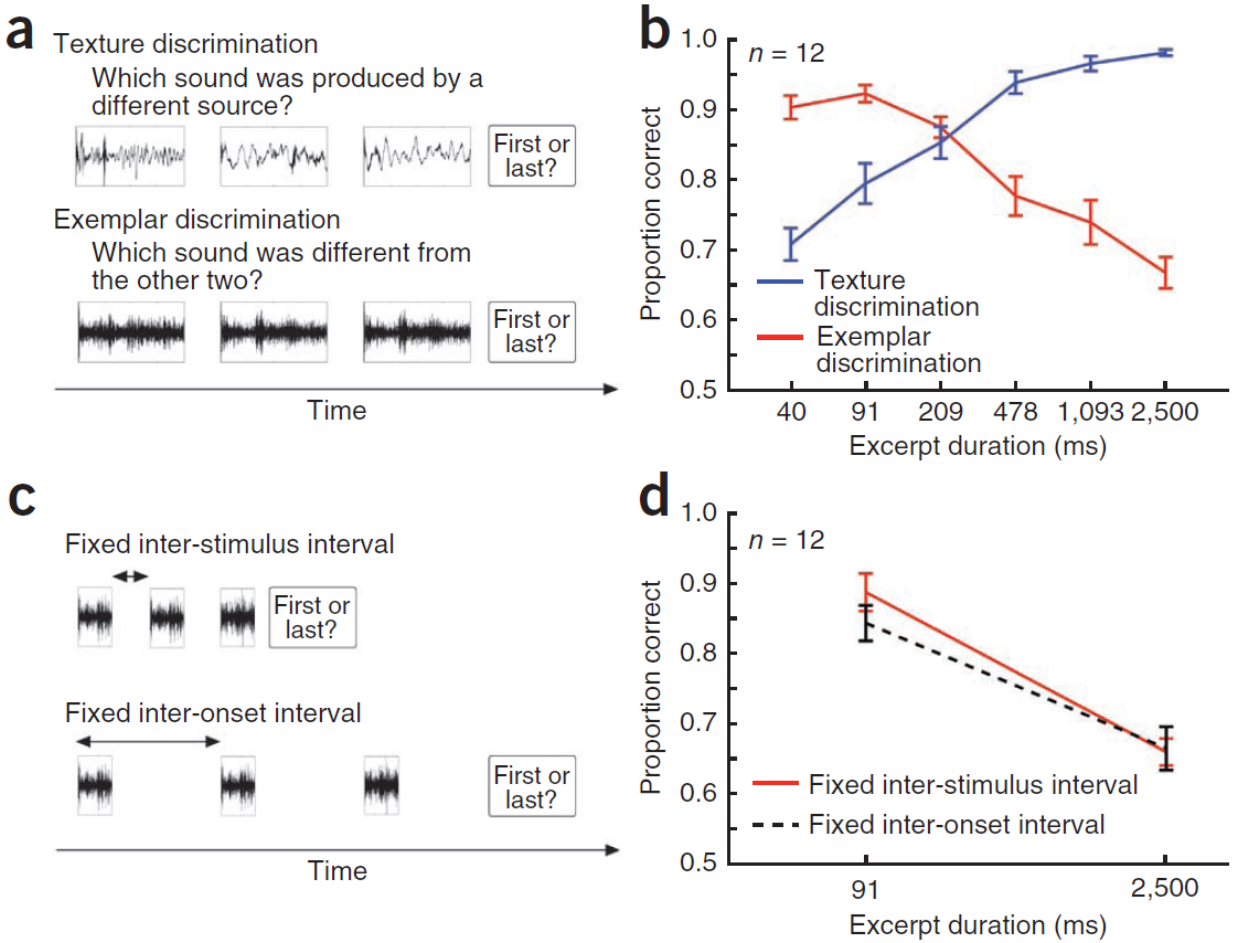
***a* = *b* = [40, 91, 209, 478, 1.093 , 2.500] ms (log scale)**
red: Exemplars discrimination
blue: Textures discrimination

schema based processing: summary statistics



***a* = *b* = [40, 91, 209, 478, 1.093 , 2.500] ms (log scale)**
red: Exemplars discrimination
blue: Textures discrimination

schema based processing: summary statistics

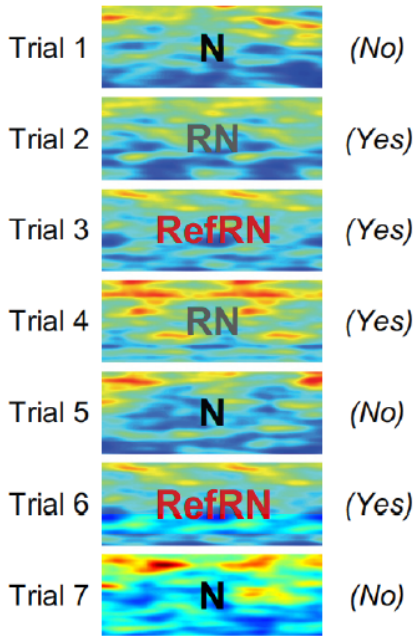


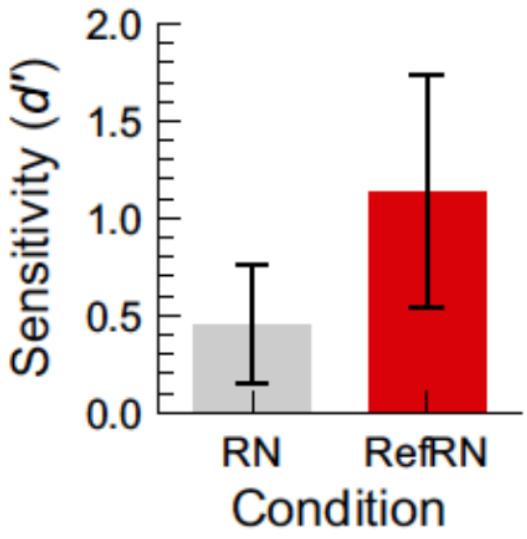
- Textures: $F(5, 55) = 50.96,\ p < 0.001$
- Exemplars: $F(5, 55) = 40.66,\ p < 0.001$

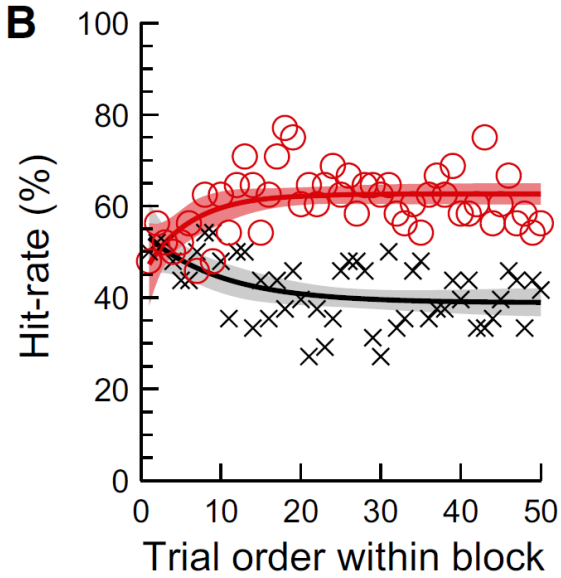
Implicit Learning of Schema

According to Agus et al. [2010] low level (acoustic) schema

- can be learned very rapidly, only few exposition necessary
- are available for several weeks
- does not require ANY meaningful structure (noise stimuli)



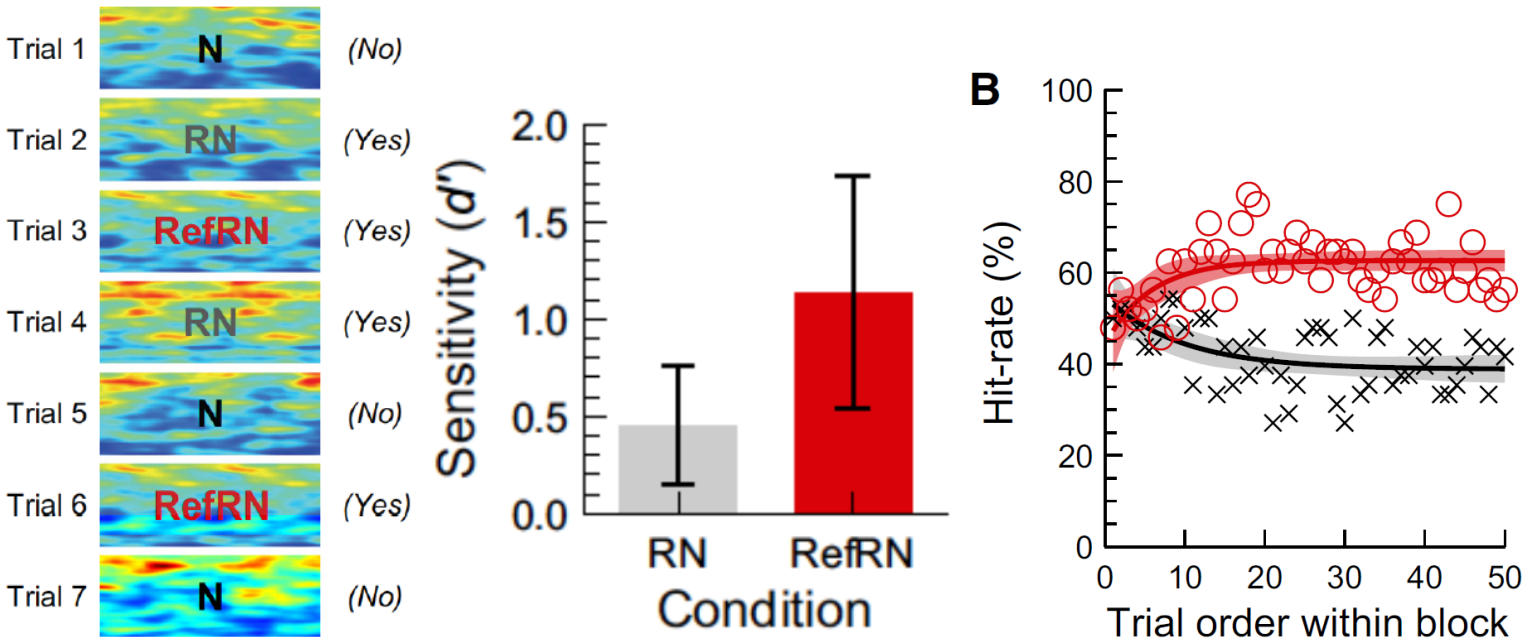




Implicit Learning of Schema

According to Agus et al. [2010] low level (acoustic) schema

- can be learned very rapidly, only few exposition necessary
- are available for several weeks
- does not require ANY meaningful structure (noise stimuli)



Unrepeated: *Repeated*: *Repeated 10 times*:

Influence of Context

The Sheppard Tones illusion

- 
- 

Implicit Learning of Schema

Most occidental people are implicit expert of tonal music

Tonal system

- Restricted set of components
- Statistical regularities (chord, tonality)

One note is dependent of the context

- Linked to the tonal hierarchy

Other systems

- Artificial ones
- System coming from other cultural contexts

Artificial languages

Simple systems (Tillman)

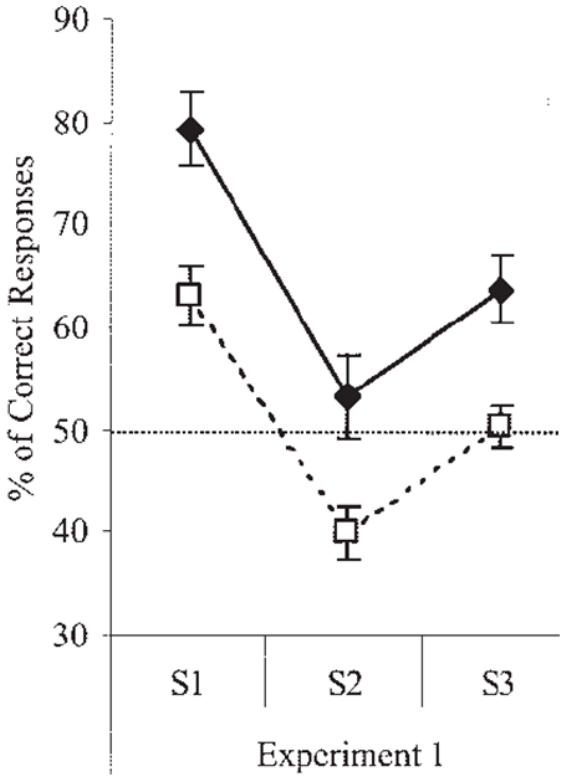
- Triplets of syllables or musical tones
- Exposition: listening passively to some triplets
- Test: choose between two word or melody which one is coming from the exposed set of triplets
- Results: 75 % (well above chance)

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

(Tillman)

- 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

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

Summary

ASA

Summary

ASA

- Primitive and Schema-based process

Summary

ASA

- Primitive and Schema-based process
- Primitive: innate constraint (Gestalt theory)

Summary

ASA

- Primitive and Schema-based process
- Primitive: innate constraint (Gestalt theory)
- sequential and simultaneous grouping

Fun, fun, but scary at the same time

https://www.youtube.com/watch?v=vJG698U2Mvo

Trevor R. Agus, Simon J. Thorpe, and Daniel Pressnitzer. Rapid formation of robust auditory memories: Insights from noise. *Neuron*, 66(4):610–618, May 2010.

James A Ballas and James H Howard. Interpreting the language of environmental sounds. *Environment and behavior*, 19(1):91–114, 1987.

Albert S Bregman. *Auditory scene analysis: The perceptual organization of sound*, pages 1–773. The MIT Press, Cambridge, Massachusetts, 1990.

Didier A Depireux, Jonathan Z Simon, David J Klein, and Shihab A Shamma. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *Journal of neurophysiology*, 85(3):1220–1234, 2001.

Danièle Dubois. *Sémantique et cognition: catégories, prototypes, typicalité*. Diffusion, Presses du CNRS, 1991.

Danièle Dubois, Catherine Guastavino, and Manon Raimbault. A cognitive approach to urban soundscapes: Using verbal data to

access everyday life auditory categories. *Acta Acustica united with Acustica*, 92(6):865–874, 2006.

Stephen McAdams and Emmanuel Bigand. *Penser Les Sons : Psychologie cognitive de l'audition*. Psychologie et sciences de la pensée, Presses Universitaire de France. 1994.

Josh H. McDermott and Eero P. Simoncelli. Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis. *Neuron*, 71(5):926–940, September 2011. ISSN 08966273. doi: 10.1016/j.neuron.2011.06.032.

Josh H McDermott, Michael Schemitsch, and Eero P Simoncelli. Summary statistics in auditory perception. *Nature neuroscience*, 16(4):493–498, 2013.

Israel Nelken. Processing of complex sounds in the auditory system. *Current opinion in neurobiology*, 18(4):413–417, 2008.

Eleanor Rosch and Barbara B Lloyd. *Cognition and categorization*, pages 1–327. Hillsdale, New Jersey, 1978.

Pierre Schaeffer. *Traité des objets musicaux*. 1966.

István Winkler, Susan L Denham, and Israel Nelken. Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in cognitive sciences*, 13(12):532–540, 2009.