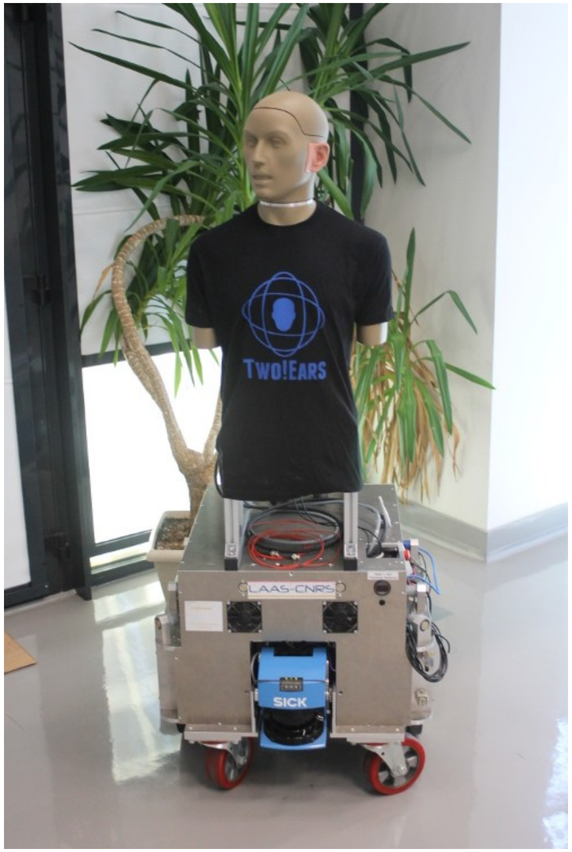
Computational models of the auditory brain

C. D. Salvador[1], R. Teraoka[1], Y.-W. Liu[2], M. Sato[3],
A. Kral[3], S. Sakamoto[1]

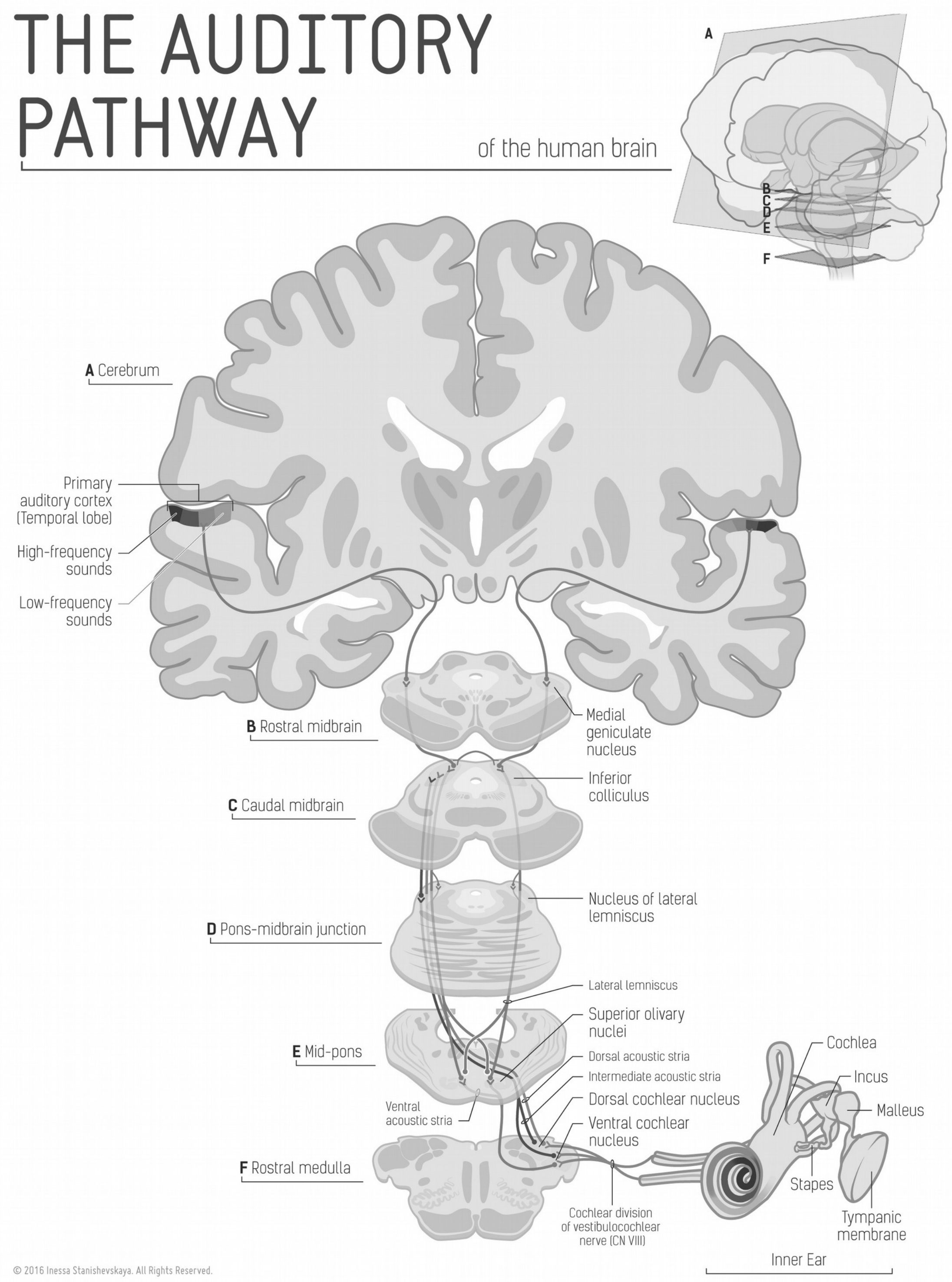
[1]*Research Institute of Electrical Communication, Tohoku University*
[2]*Department of Electrical Engineering, National Tsing Hua University*
[3]*Institute of AudioNeuroTechnology and Department of Experimental Otology, ENT Clinics, Hannover Medical School*

Introduction

- Auditory brain modeling is important to realize human-oriented audio processing tools
- These tools are required in:
  - **Artificial hearing systems**
  - 3D sound for virtual and augmented reality
  - Next-generation hearing aids
- We present a review of auditory models



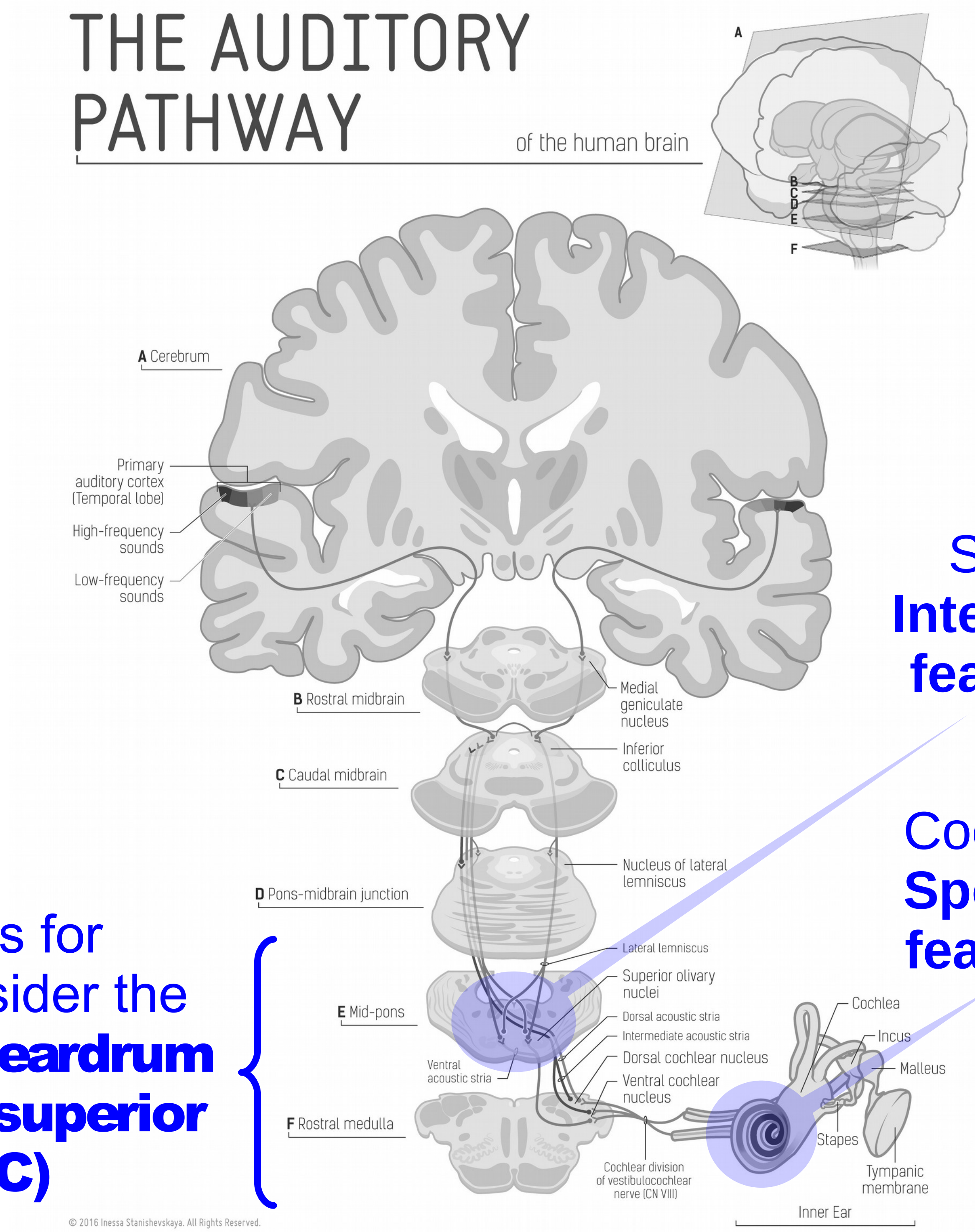
http://twoears.eu/



Cognition and higher functions

Integration with other modes of perception

Extraction of primary auditory features



© 2016 Inessa Stanishevskaya. All Rights Reserved.

SOC:
Interaural features

Cochlea:
Spectral features

State-of-the-art models for audio processing consider the transduction from the **eardrum** to the **cochlea** to the **superior olivary complex (SOC)**

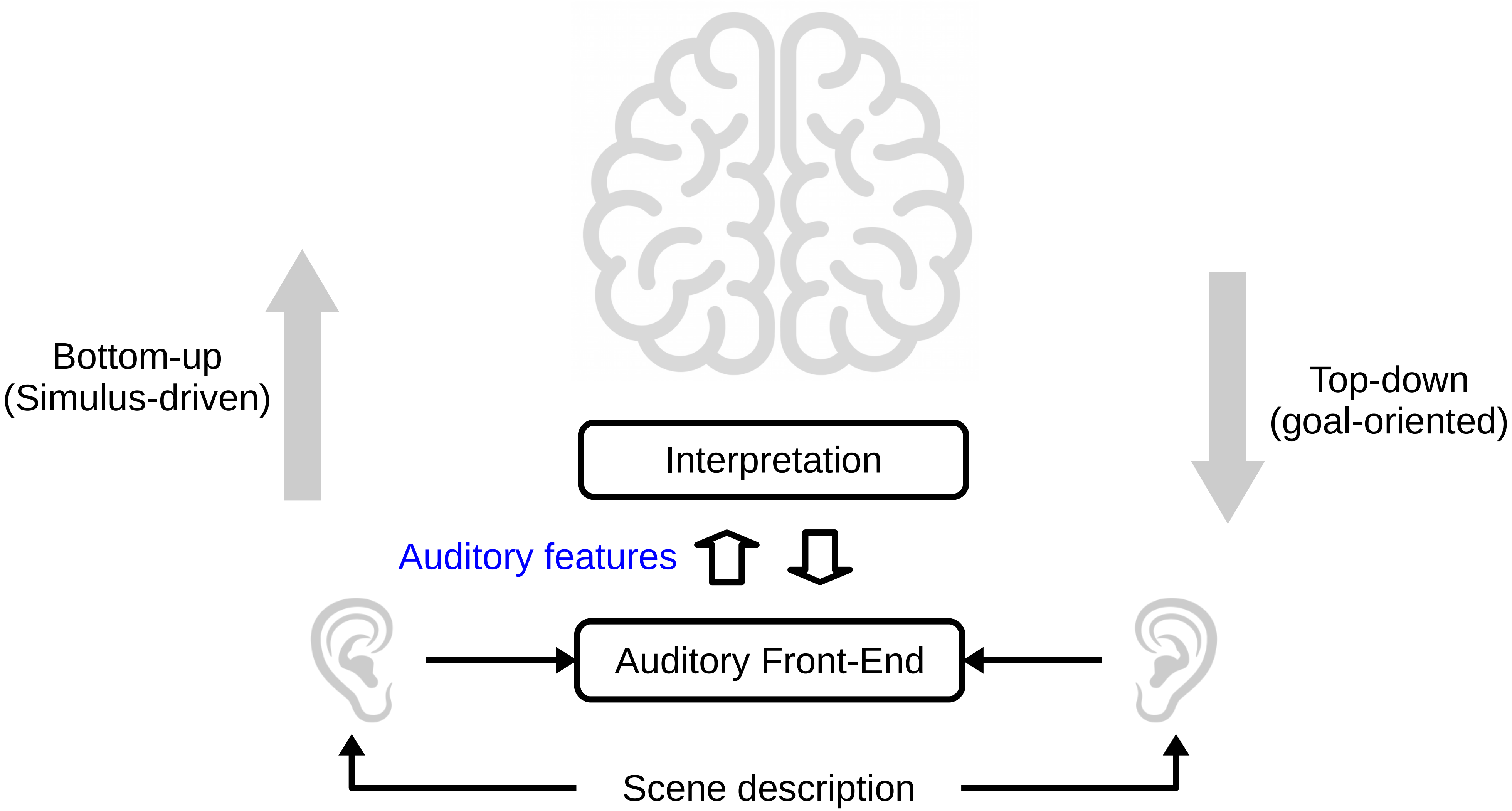
Line of thought

- Detailed aspects of biological structuralism are sacrificed for the sake of large-scale functionality
- Emphasis is placed on algorithmic efficiency, which is crucial in the development of advanced brain-morphic LSI systems

Toolboxes for auditory modeling

- Past projects
  - Development system for auditory modeling (L. P. O’Mard *et al.*, 1986-2007)
  - Auditory toolbox (M. Slaney *et al.*, 1993-1998)
- Recent projects
  - Auditory modeling toolbox (P. L. Søndergaard *et al.*, 2013)
  - Two!Ears Auditory Model (A. Raake *et al.*, 2013-2016)
    - Object-oriented
    - Modular and scalable
    - Batch processing and chunk-based processing

Auditory modeling paradigm



Bottom-up
(Simulus-driven)

Top-down
(goal-oriented)

Interpretation

Auditory features

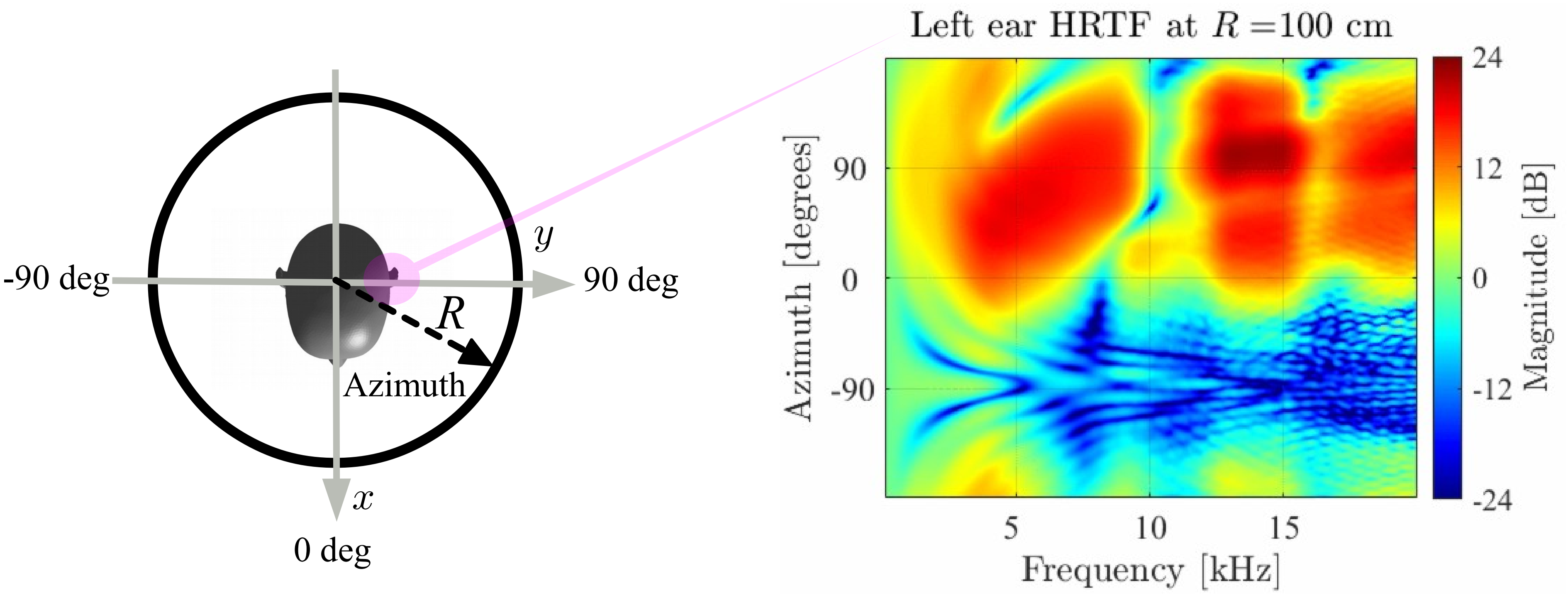
Auditory Front-End

Scene description

http://twoears.eu/

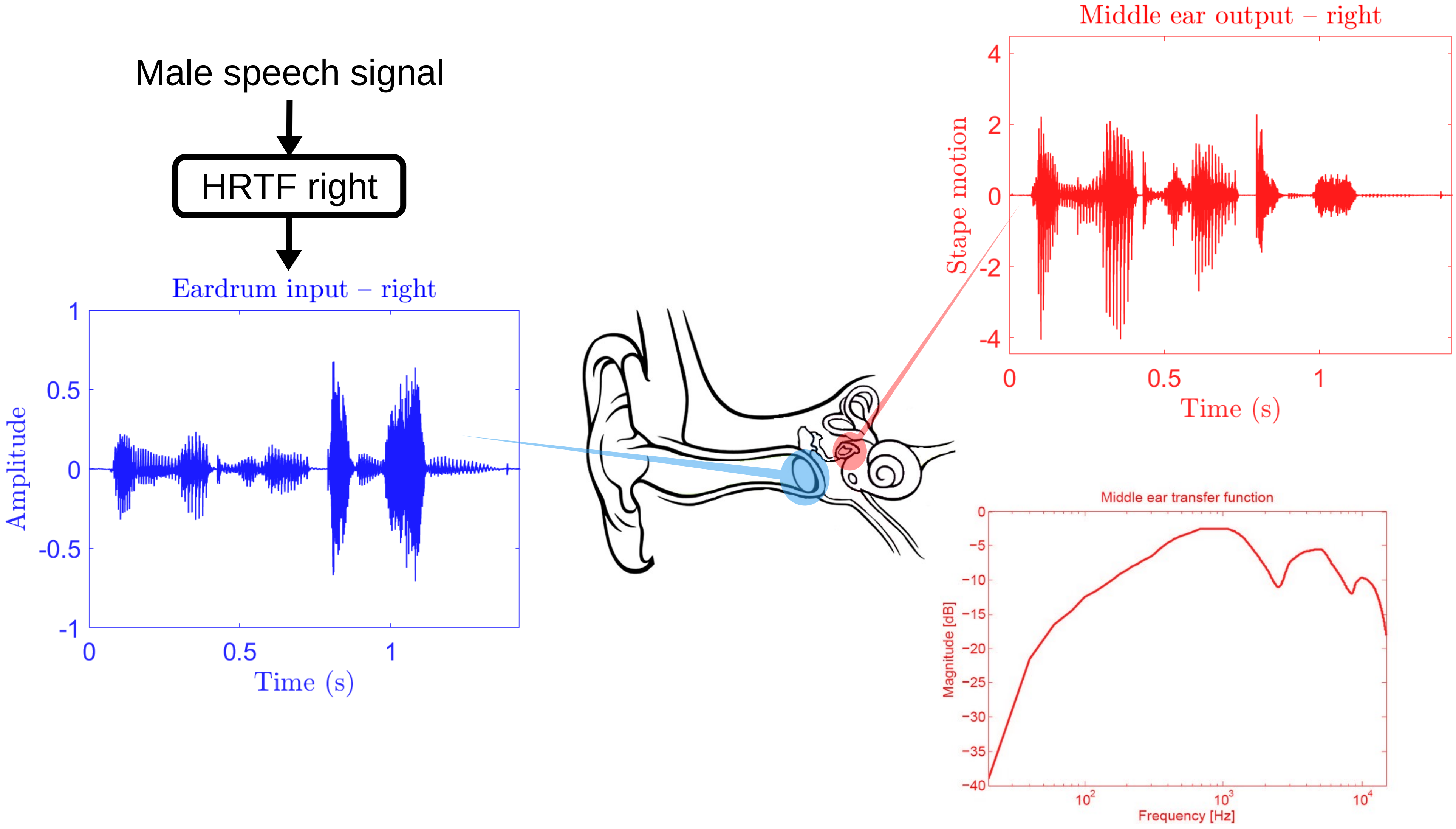
Acoustic filtering of the head and outer ear

Head-related transfer functions (HRTFs)



C. D. Salvador *et al.*, “Dataset of near-distance head-related transfer functions calculated using the boundary element method,” in *Proc. Audio Eng. Soc. Int. Conf. Spatial Reproduction —Aesthetics and Science—*, Tokyo, Japan, 2018.

Middle ear transfer function

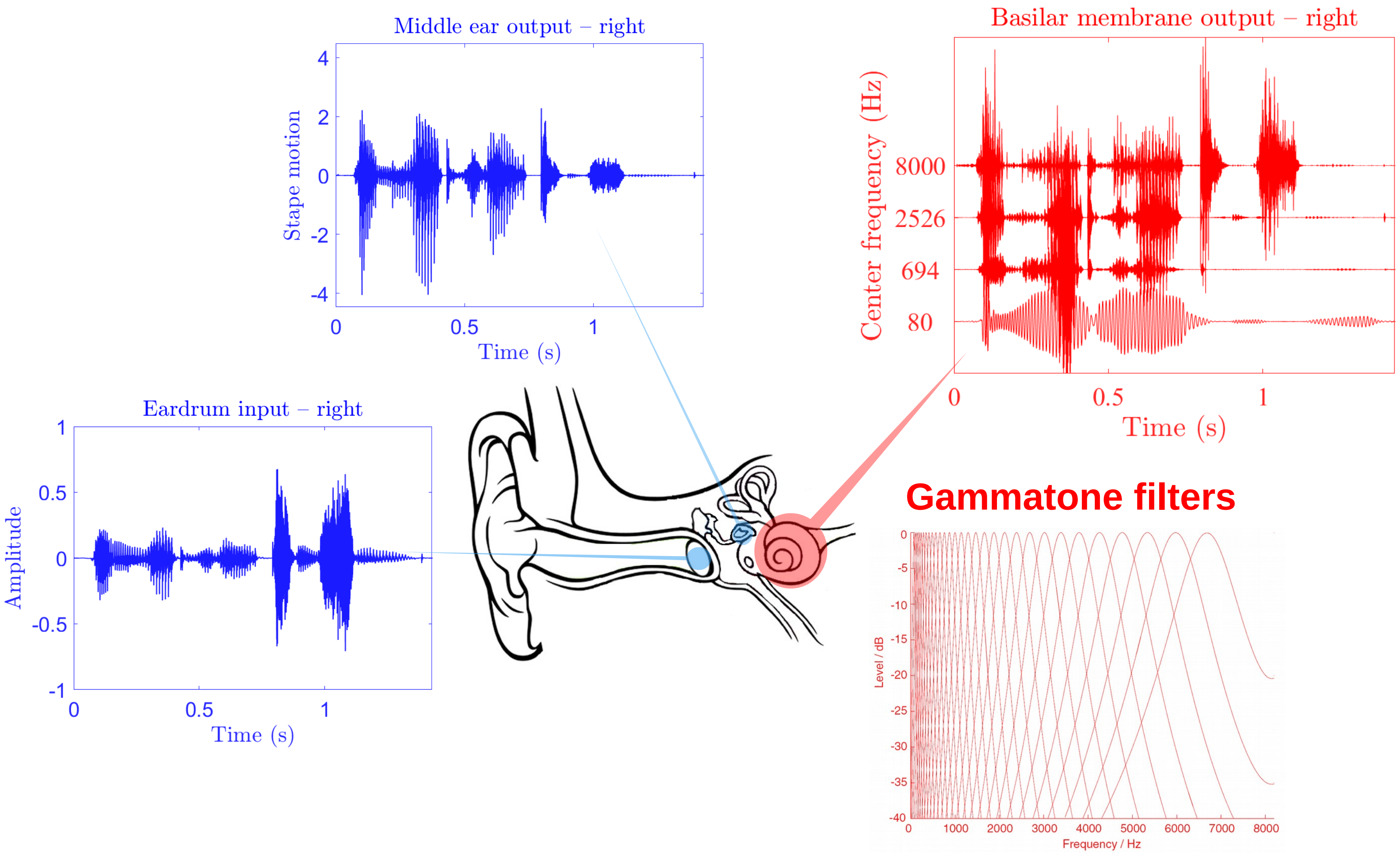


Male speech signal

HRTF right

M. L. Jepsen *et al.*, “A computational model of human auditory signal processing and perception,” *J. Acoust. Soc. Am.*, vol. 124, no. 1, pp. 422–438, 2008.

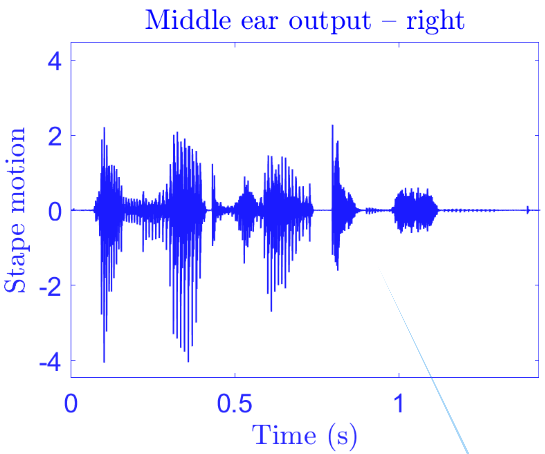
Cochlea: linear filtering

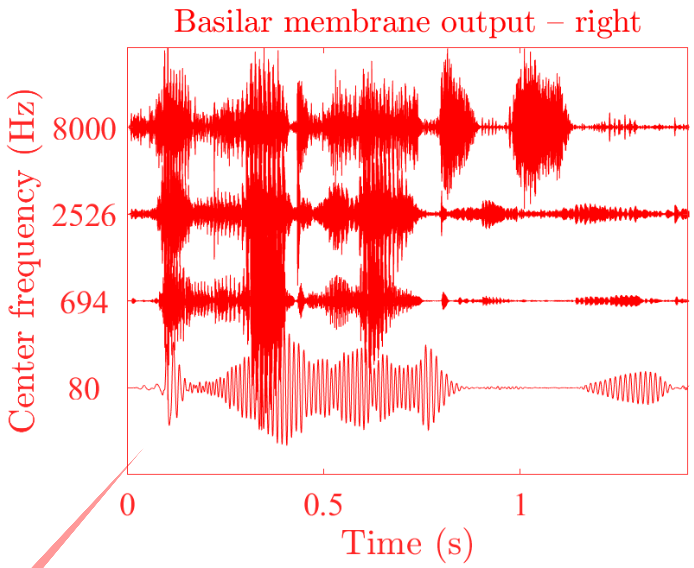


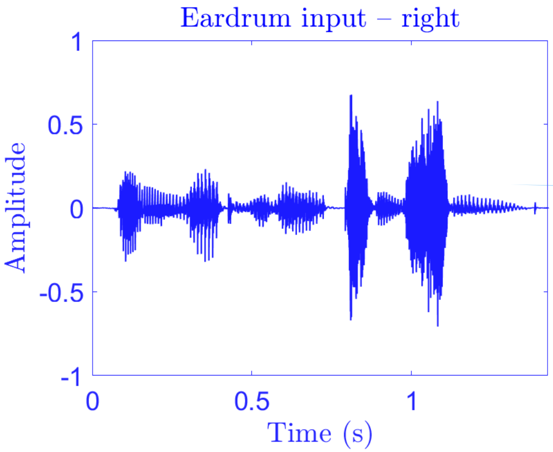
Gammatone filters

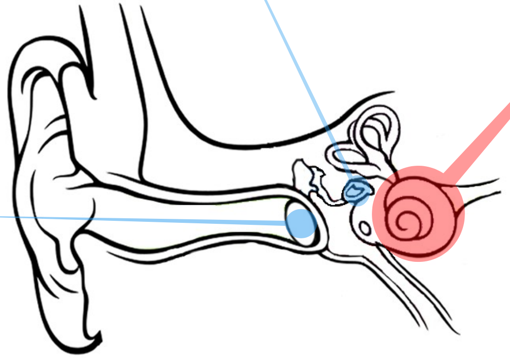
V. Hohmann, "Frequency analysis and synthesis using a gammatone filterbank," *Acta Acust. United Ac.*, vol. 88, no. 3, pp. 433–442, May 2002.

Cochlea: non-linear filtering

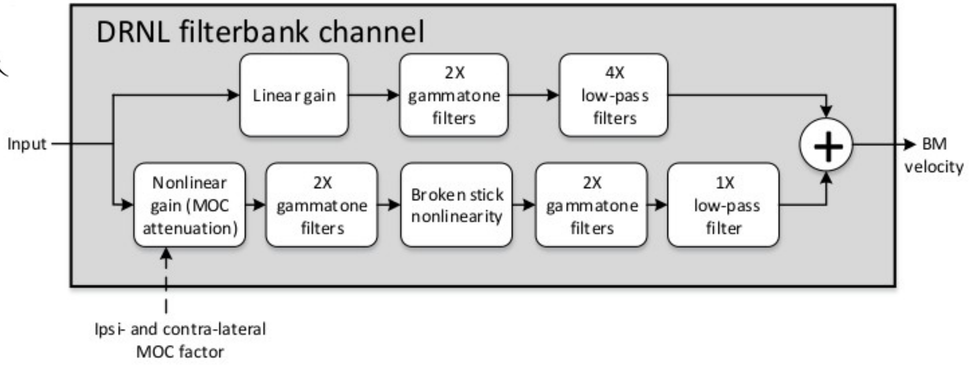






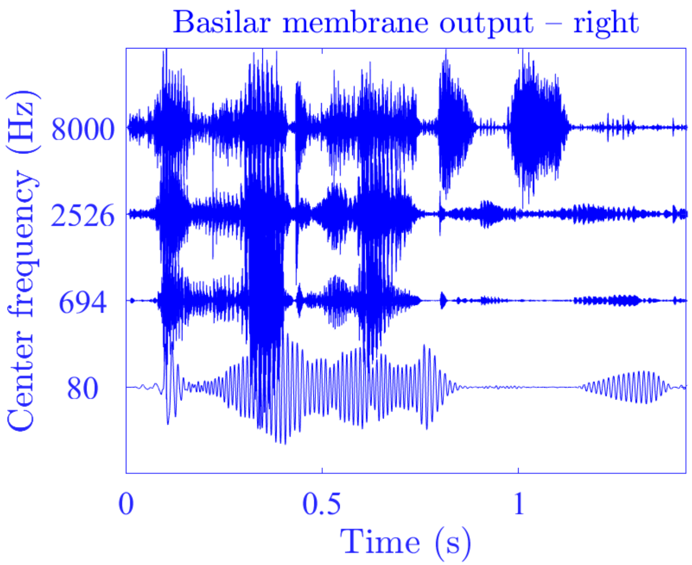


Dual-resonance non-linear filters

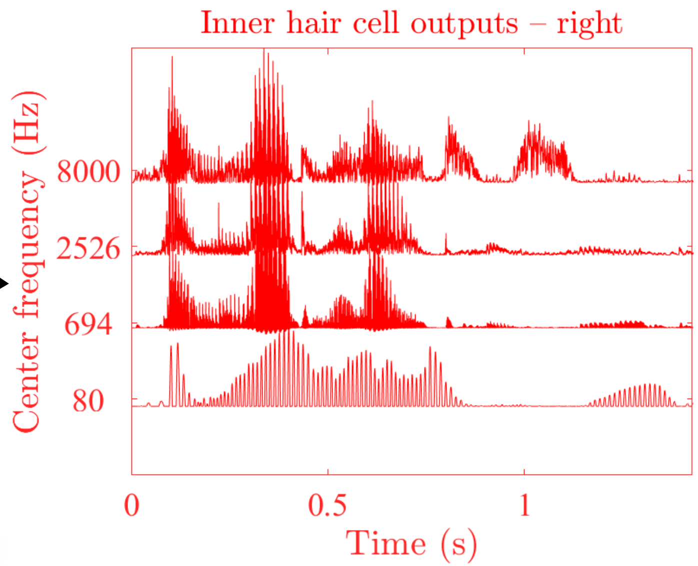


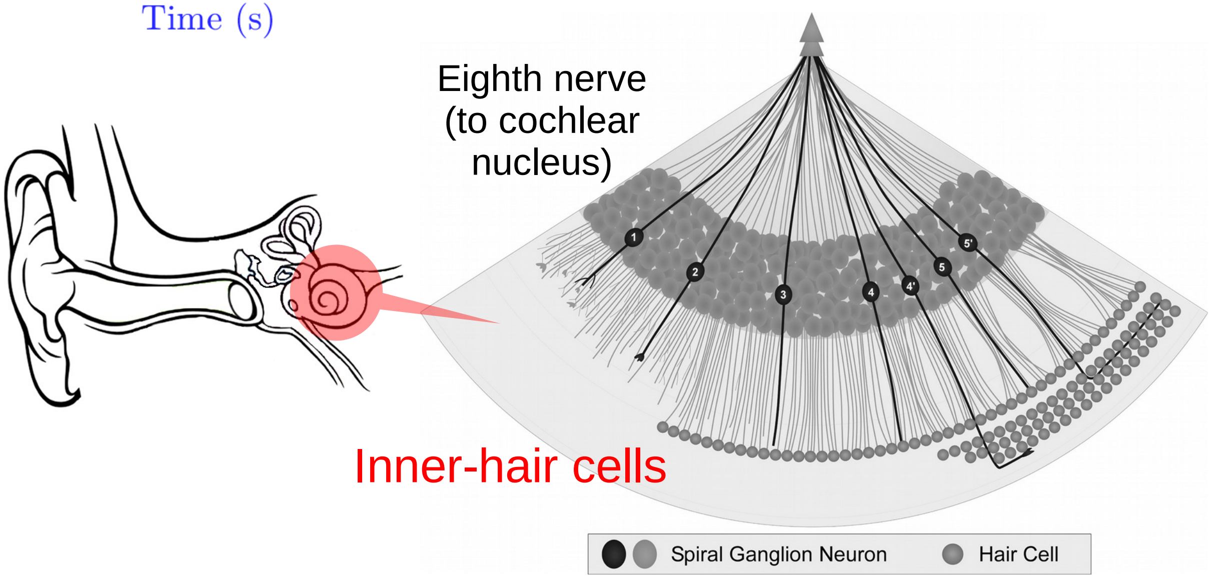
E. A. Lopez-Poveda and R. Meddis, “A human nonlinear cochlear filterbank,” *J. Acoust. Soc. Am.*, vol. 110, no. 6, pp. 3107–3118, 2001.

Hair-cell envelope detector



Envelope detector



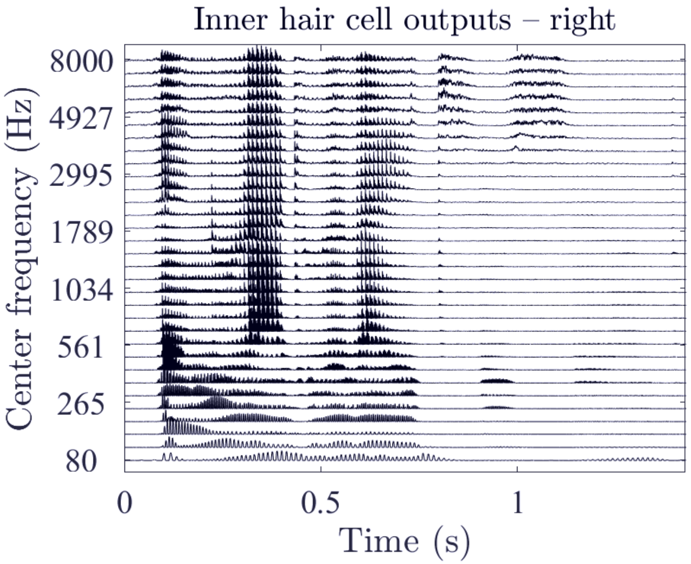


Eighth nerve (to cochlear nucleus)

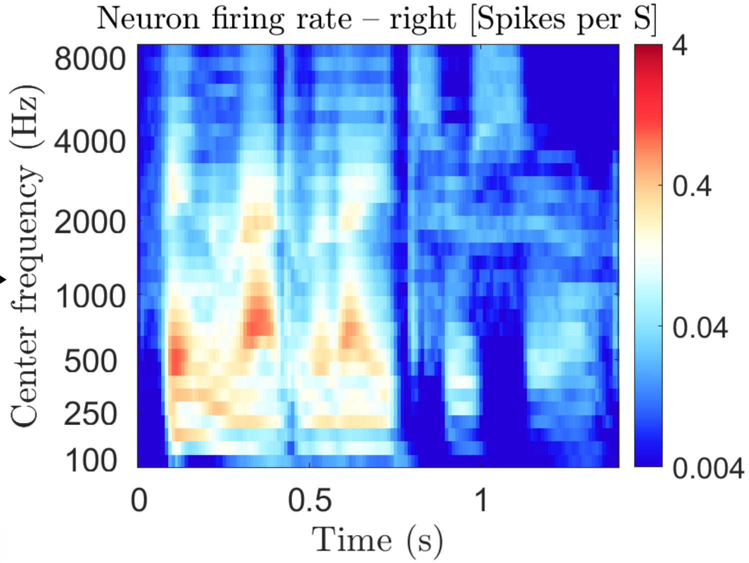
Inner-hair cells

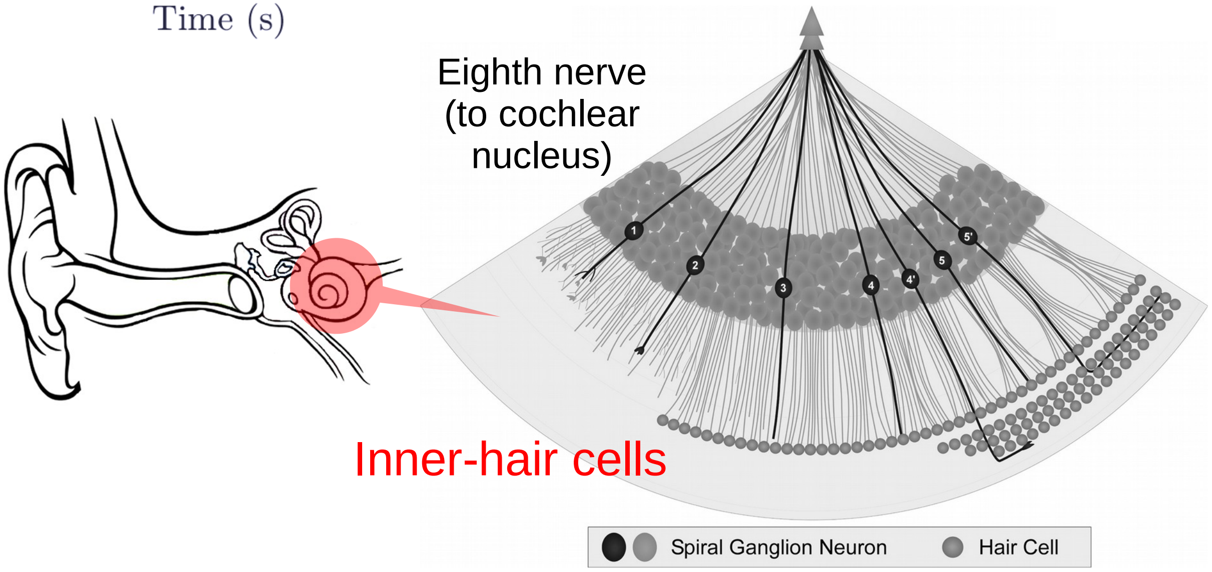
J. M. Appler and L. V. Goodrich, “Connecting the ear to the brain: Molecular mechanisms of auditory circuit assembly,” *Prog. Neurobiol.*, vol. 93, no. 4, pp. 488 – 508, 2011.

Neuron firing rates



Leaky integrator

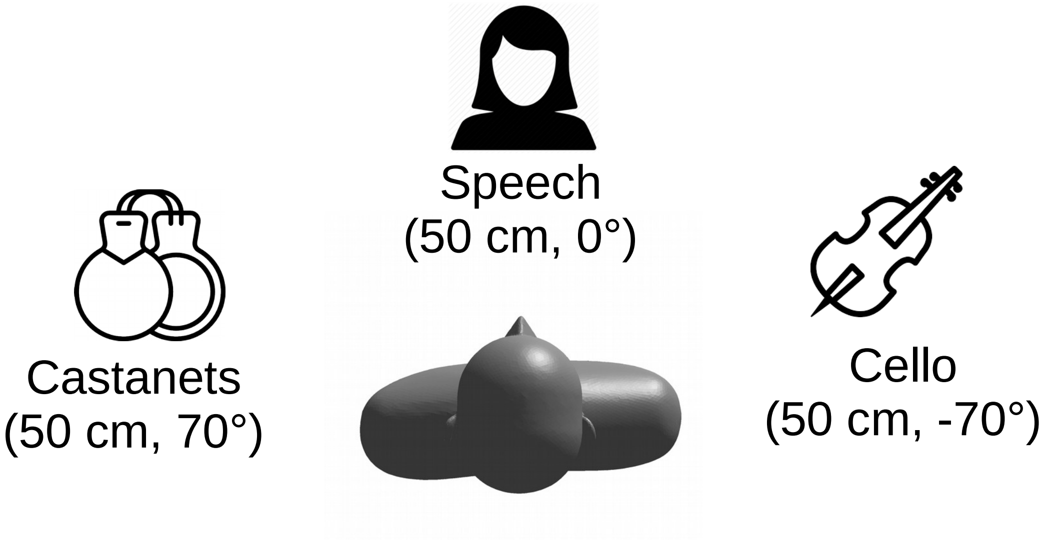




Rate maps provide primary features for localization and identification of sounds

J. M. Appler and L. V. Goodrich, “Connecting the ear to the brain: Molecular mechanisms of auditory circuit assembly,” *Prog. Neurobiol.*, vol. 93, no. 4, pp. 488 – 508, 2011.

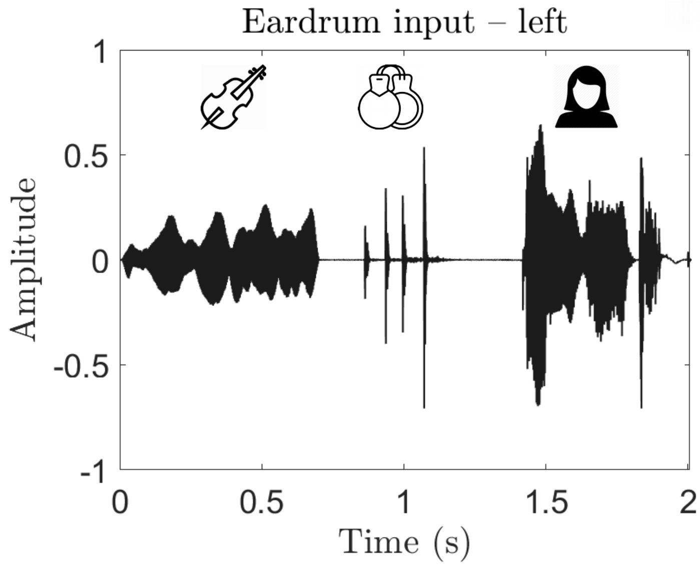
Example: Localization and identification of sounds

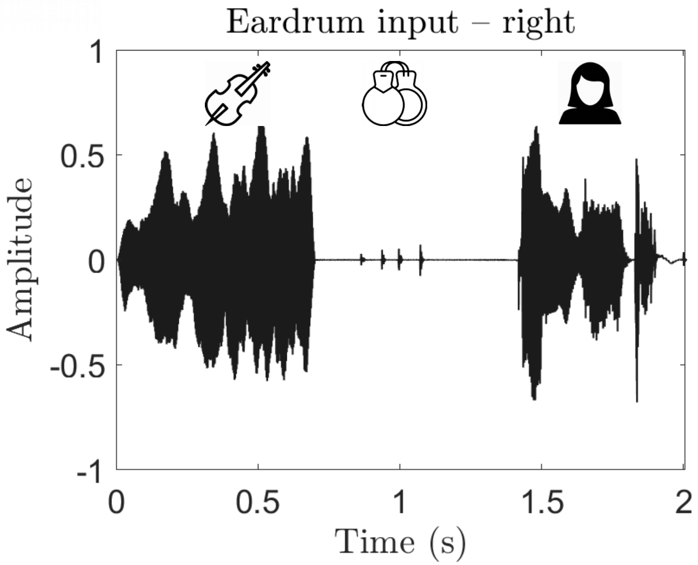


Speech
(50 cm, 0°)

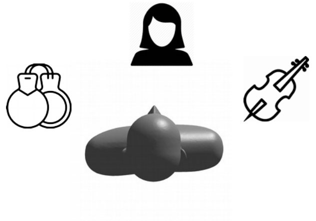
Castanets
(50 cm, 70°)

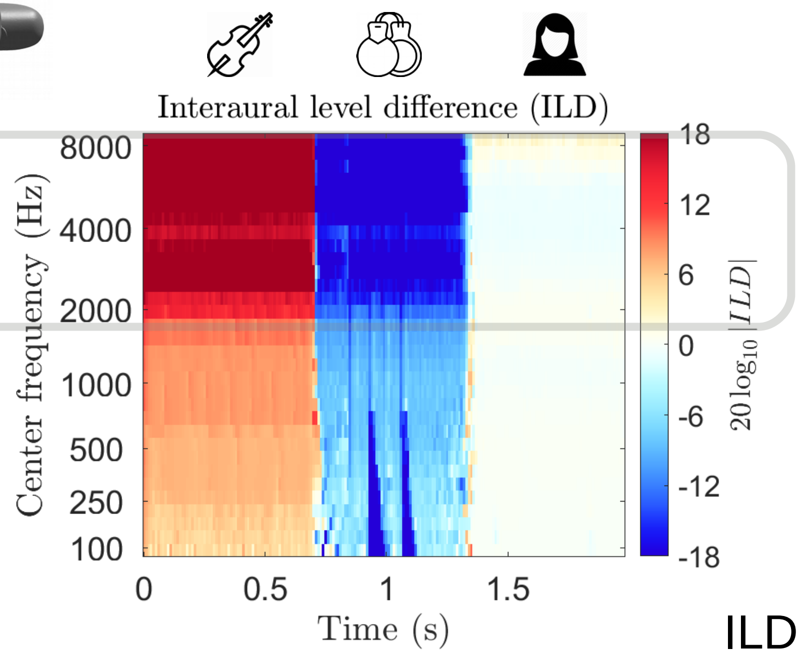
Cello
(50 cm, -70°)

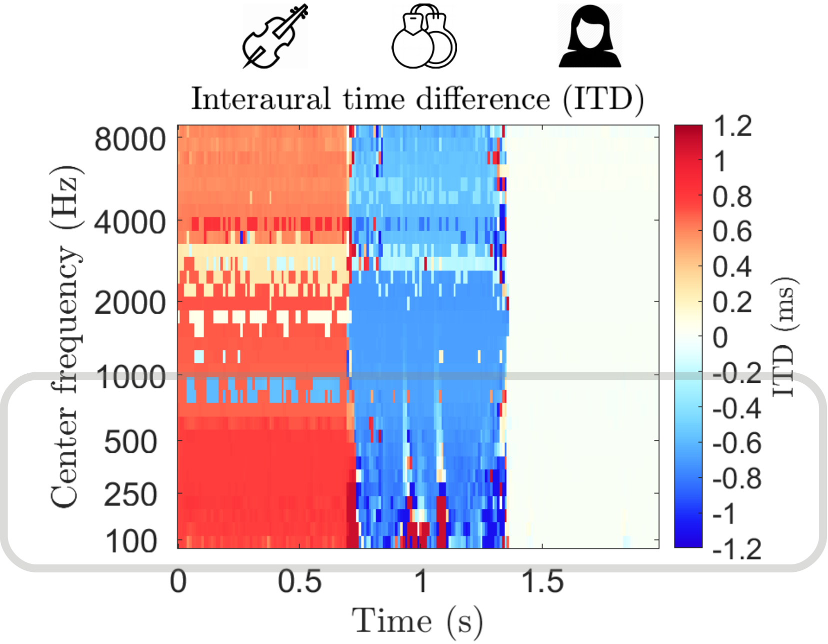


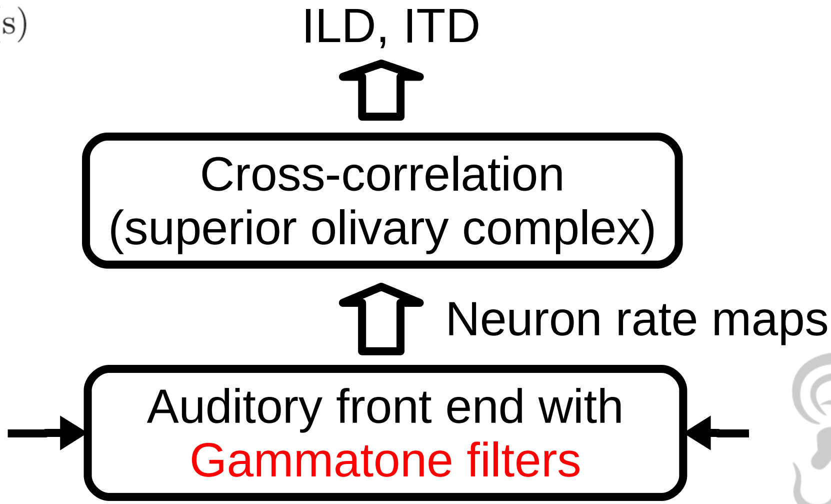


Localization cues (linear cochlea)







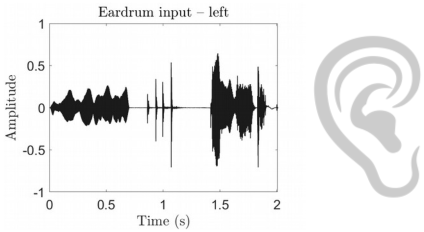


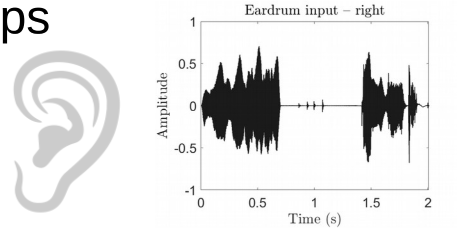
ILD, ITD

Cross-correlation
(superior olivary complex)

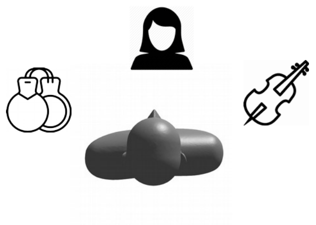
Neuron rate maps

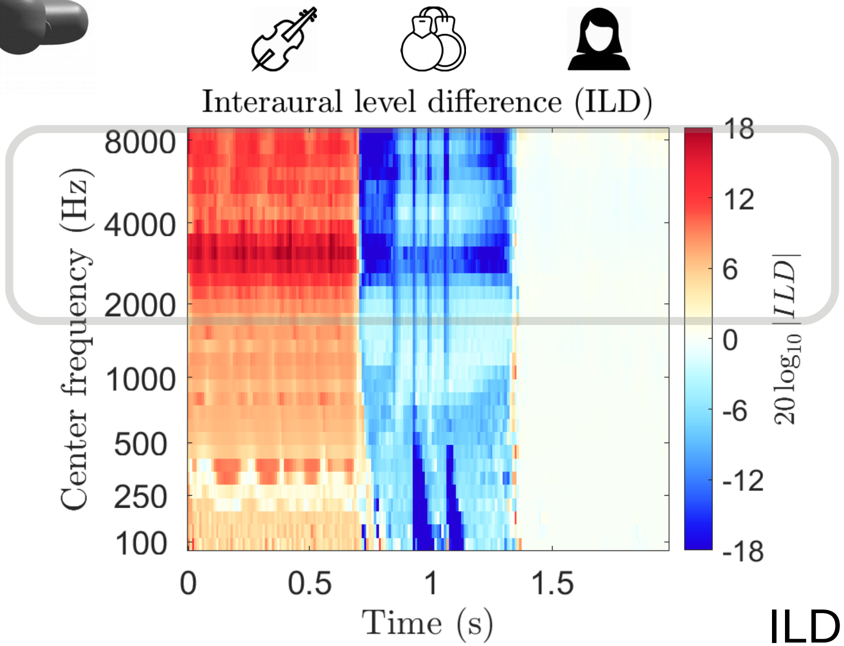
Auditory front end with
Gammatone filters

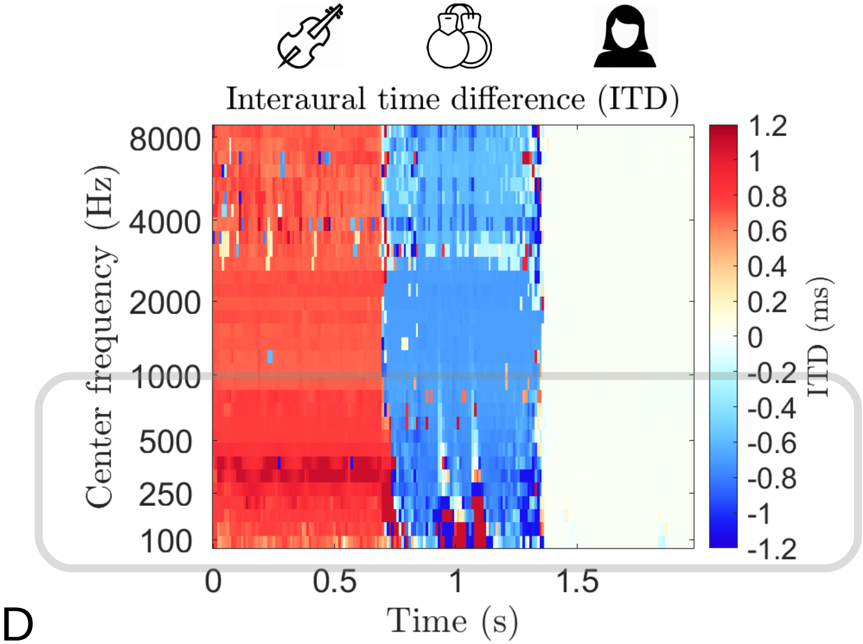




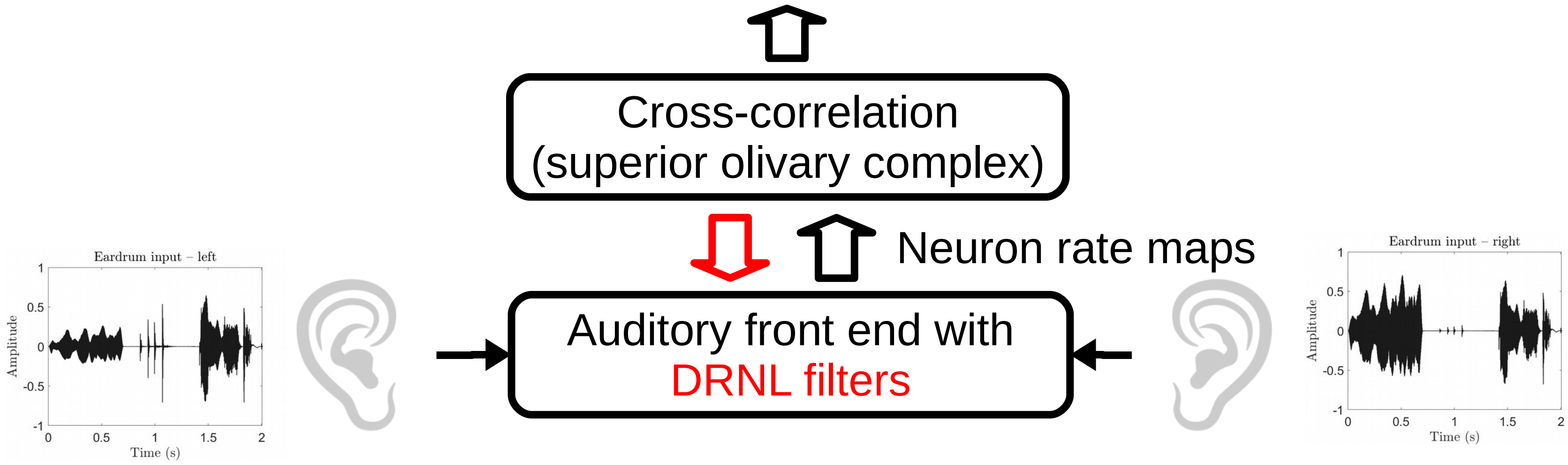
Localization cues (non-linear cochlea)







ILD, ITD

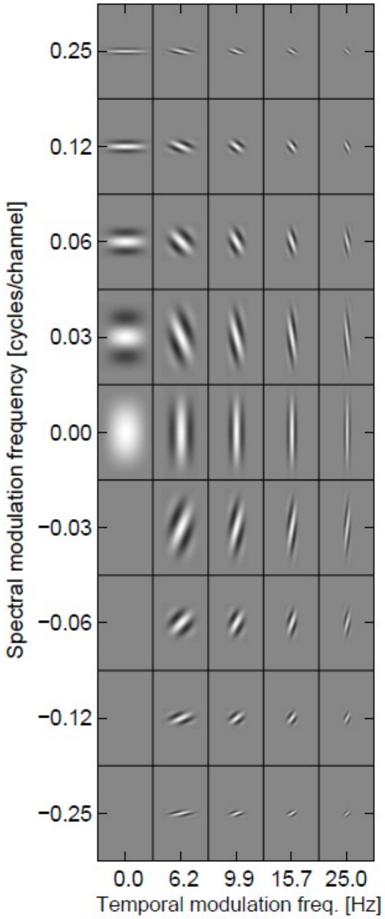


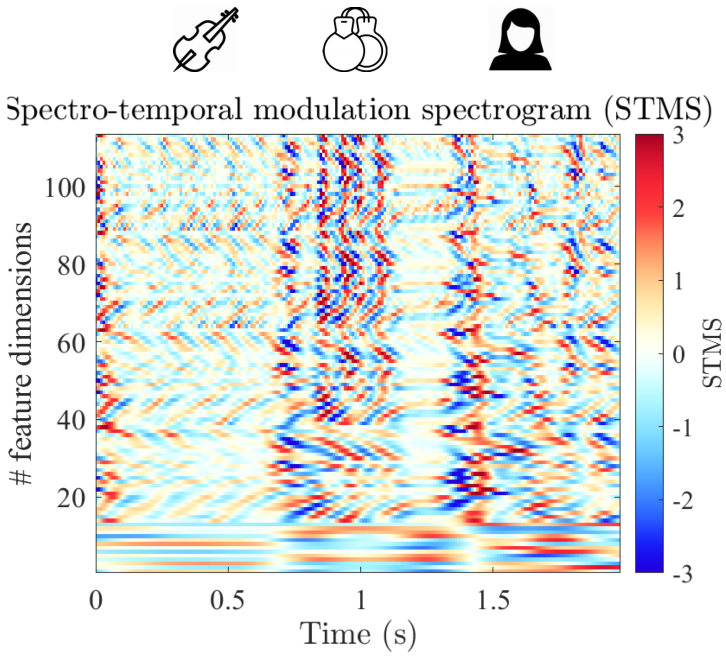
Cross-correlation
(superior olivary complex)

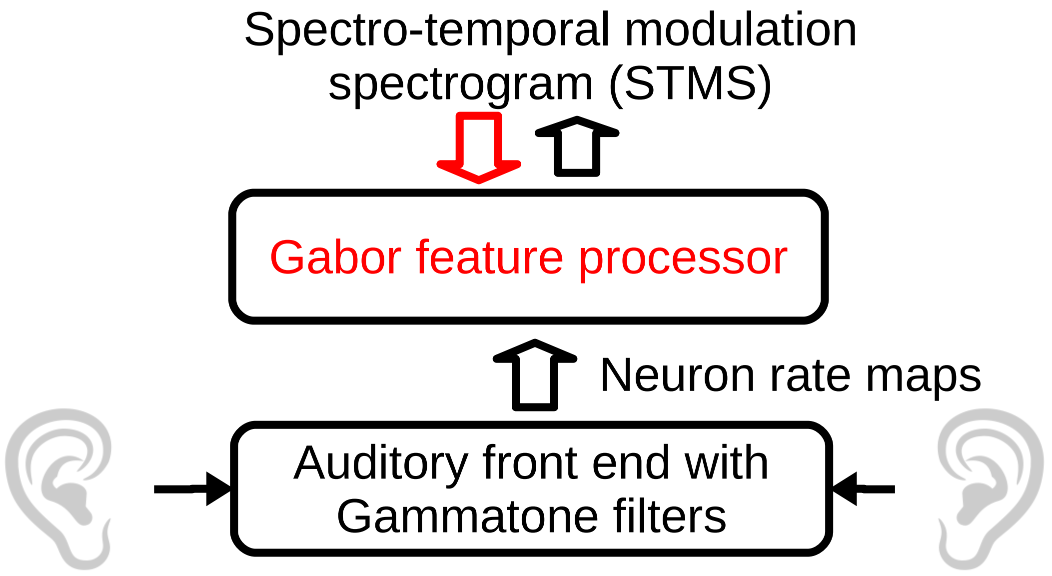
Neuron rate maps

Auditory front end with
DRNL filters

Identification cues







Spectro-temporal modulation spectrogram (STMS)

Gabor feature processor

Neuron rate maps

Auditory front end with Gammatone filters

M. R. Schädler *et al.*, “Spectro-temporal modulation subspace-spanning filter bank features for robust automatic speech recognition,” *J. Acoust. Soc. Am.*, vol. 131, no. 5, pp. 4134–4151, 2012.

Top-down auditory attention model

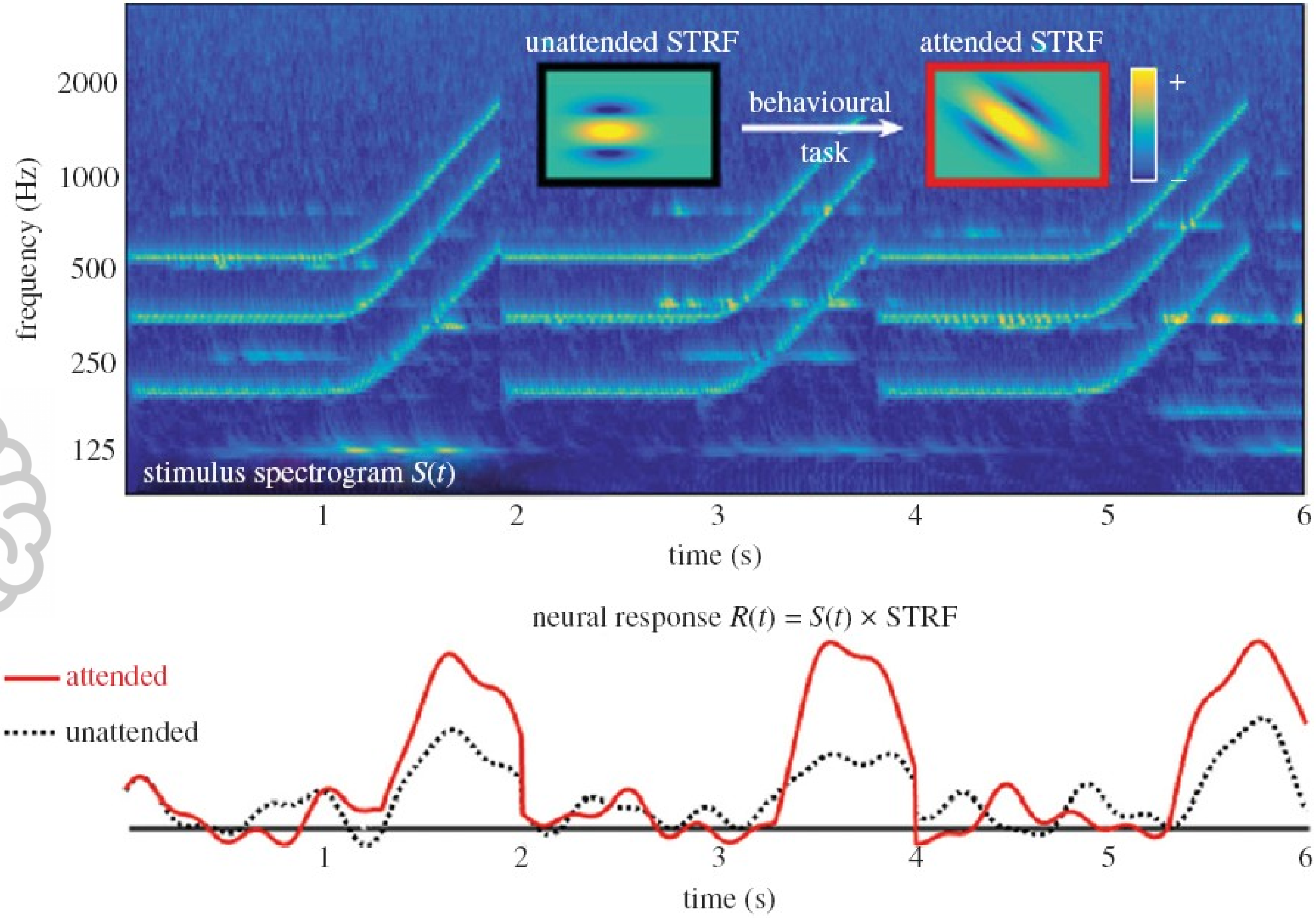


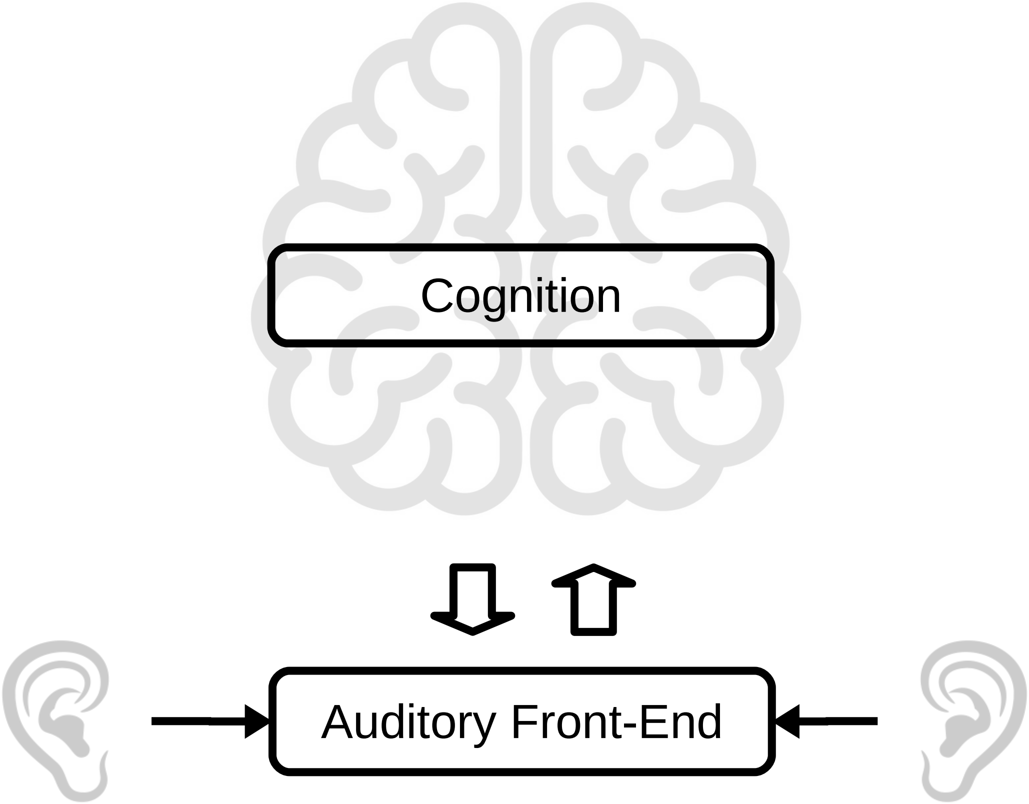
Figure 3. Attending to a particular sound characteristic tunes the neural spectro-temporal receptive fields (STRFs) and boosts the neural signal at times of attended event. Violin notes are overlaid with frequency modulations (FMs), illustrated with the spectrogram $S(t)$. When instructed to attend to the FM segments, the STRF adapts to the orientation of the modulations, resulting in an enhancement in the neural response $R(t)$.

E. M. Kaya and M. Elhilali, “Modelling auditory attention,” Phil. Trans. R. Soc. B Biol. Sci., vol. 372, no. 1714, Jan. 2017.

Summary of auditory front-end modeling

- Localization of sound
  - Non-linear models of the cochlea
    - Include interaural olivocochlear feedback
    - Extract better spectral features
- Identification of sound
  - Gabor feature processors
    - Mimic the spectro-temporal receptive fields of neurons
    - Improves Identification of low-frequency sounds
    - Possible modeling of top-down attention

An approach to include cognition



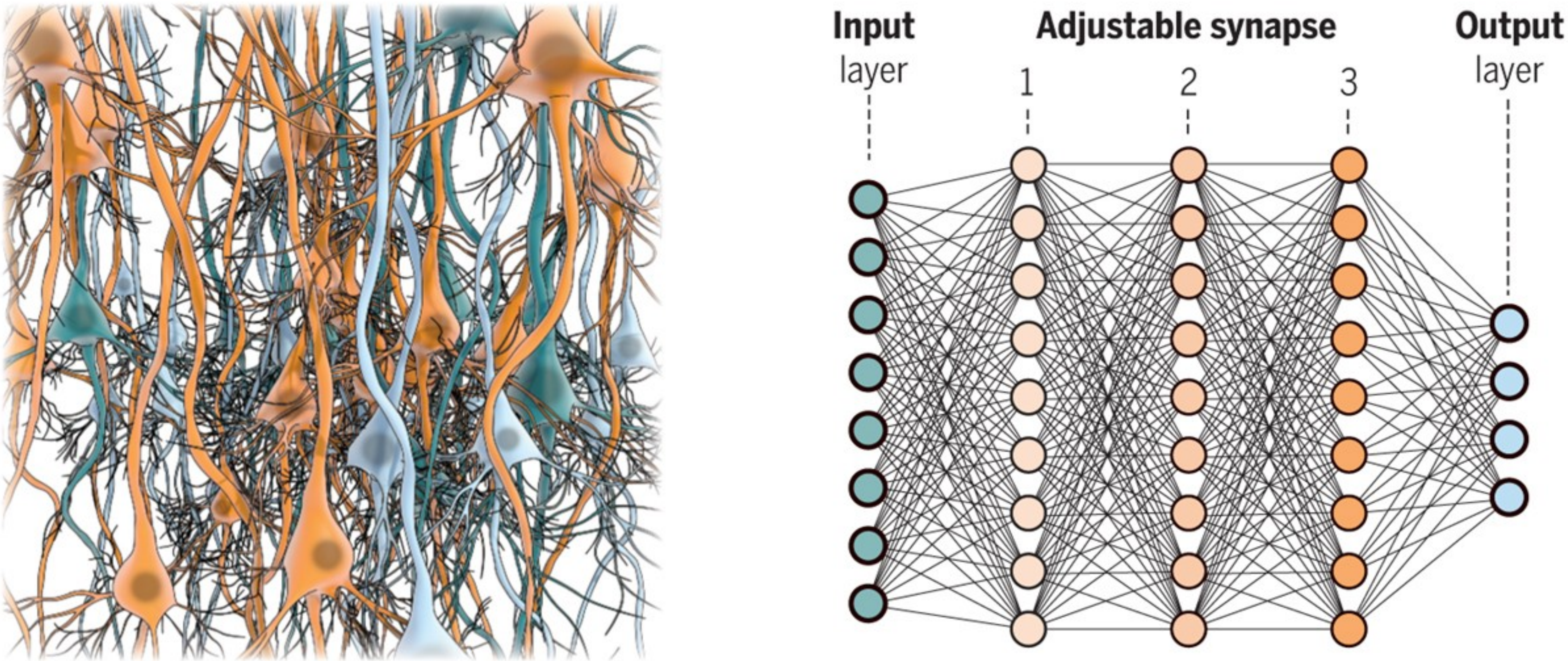
Cognition

Auditory Front-End

Learning innate structures

Brain circuitry and learning

A major open question is whether the highly simplified structures of current network models compared with cortical circuits are sufficient to capture the full range of human-like learning and cognition.



Complex neural network **Informed AI network**

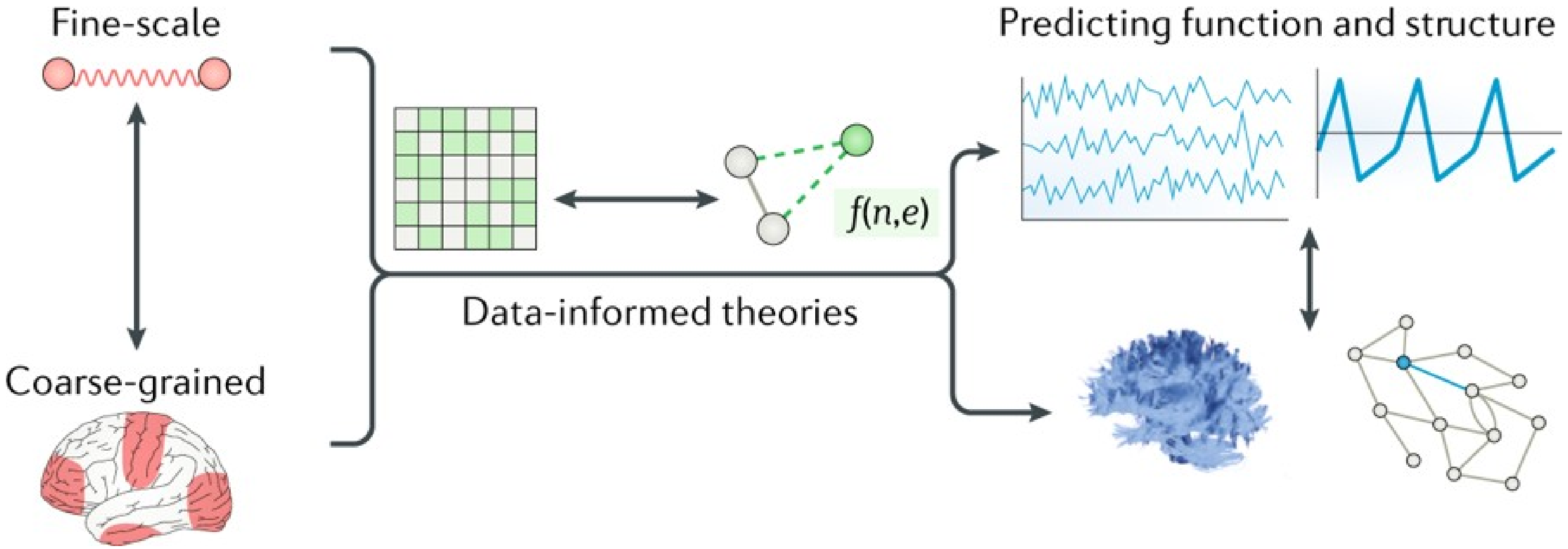
S. Ullman, “Using neuroscience to develop artificial intelligence,” Science, vol. 363, no. 6428, pp. 692–693, 2019.

Network neuroscience

Connectome

Graph theory and complex networks

- Structural connectivity (neural links)
- Functional connectivity (statistical association)



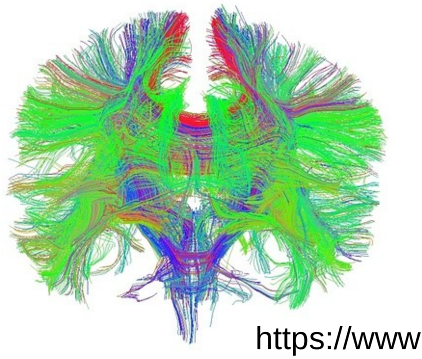
Clinical diagnosis

A connectome is a comprehensive map of neural connections. Structural maps are obtained from noninvasive neuroimaging.

D. S. Bassett and O. Sporns, “Network neuroscience,” Nat. Neurosci., vol. 20, p. 353, Feb. 2017.

Brain-inspired machine hearing

Topological data analysis

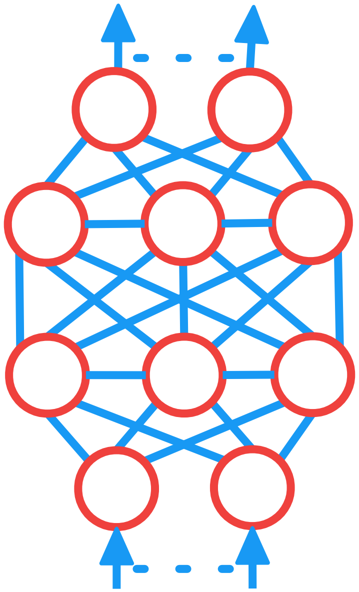


https://www.brain-map.org/

Auditory brain connectome

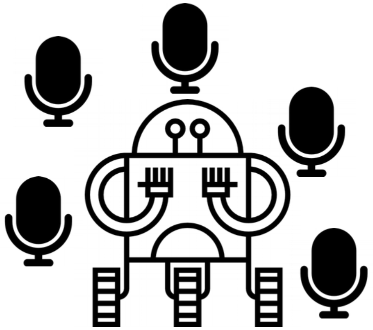
Structure
Function

Acoustic environment recognition

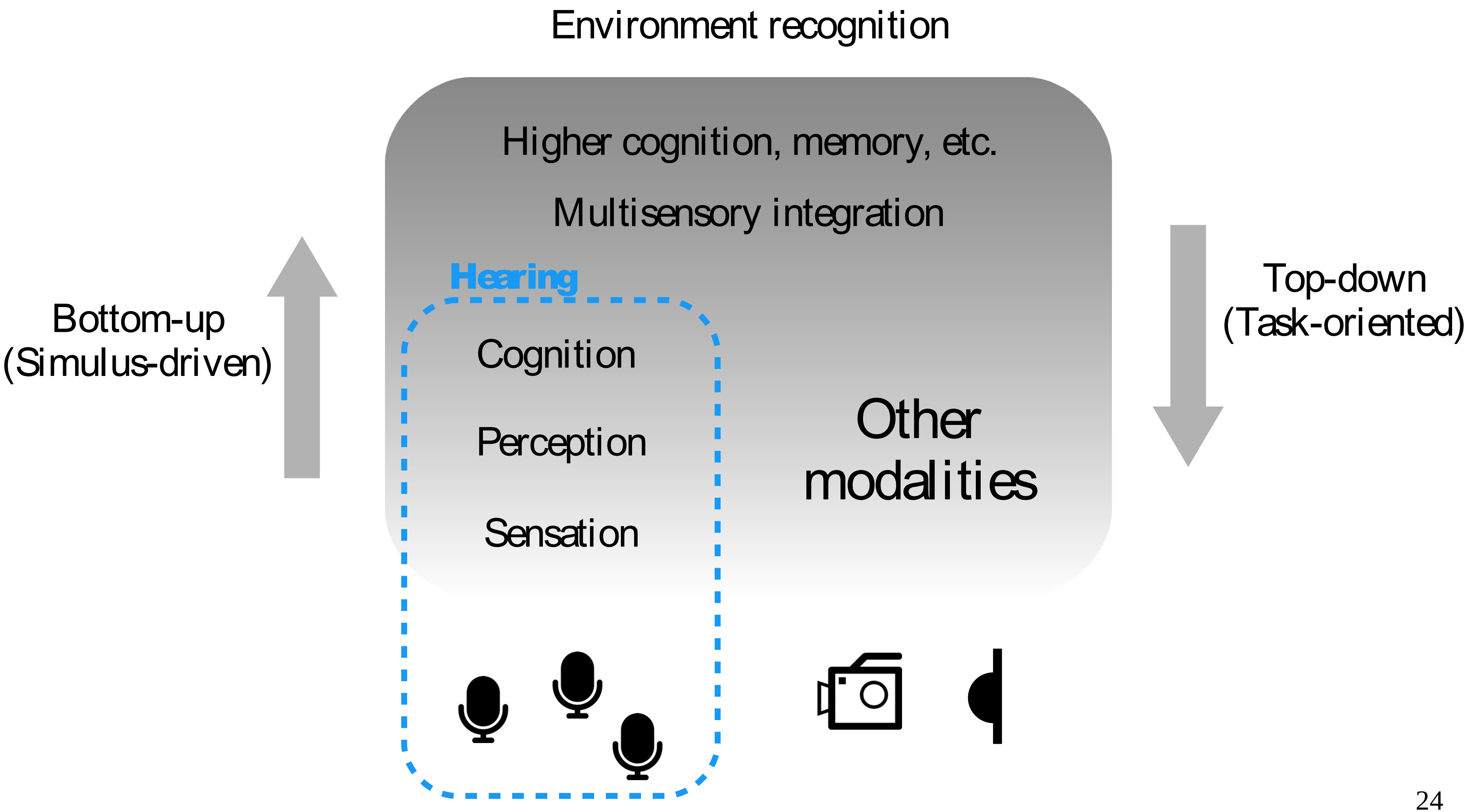


Network with biologically innate connectivy

Multichannnel auditory front-end



Towards a brain-inspired multisensory machine



Environment recognition

Higher cognition, memory, etc.

Multisensory integration

Hearing

Cognition

Perception

Sensation

Other modalities

Bottom-up
(Simulus-driven)

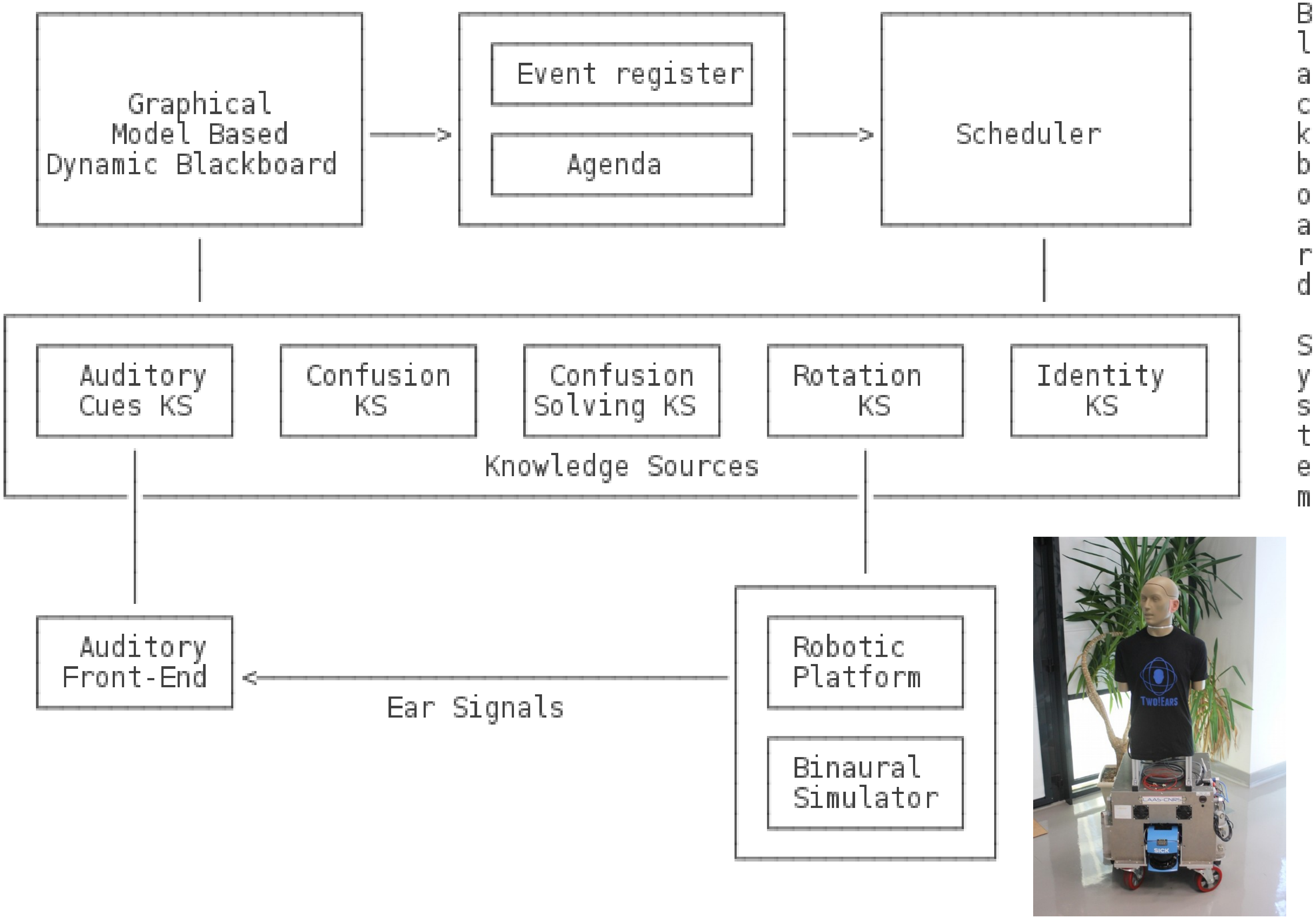
Top-down
(Task-oriented)

Thanks for your attention.

Appendix A:

The Two!Ears Auditory Model

Two!Ears Auditory Model



Graphical Model Based Dynamic Blackboard

Event register

Agenda

Scheduler

Auditory Cues KS

Confusion KS

Confusion Solving KS

Rotation KS

Identity KS

Knowledge Sources

Auditory Front-End

Ear Signals

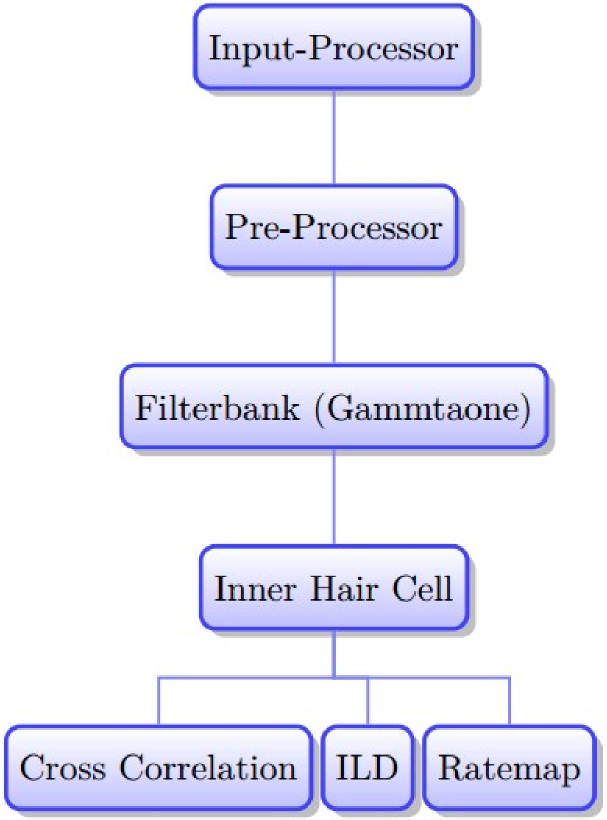
Robotic Platform

Binaural Simulator

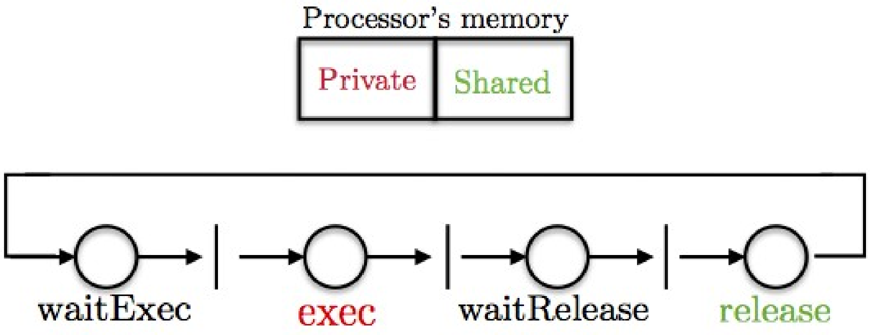
Blackboard System

http://twoears.eu/

Auditory Front-End



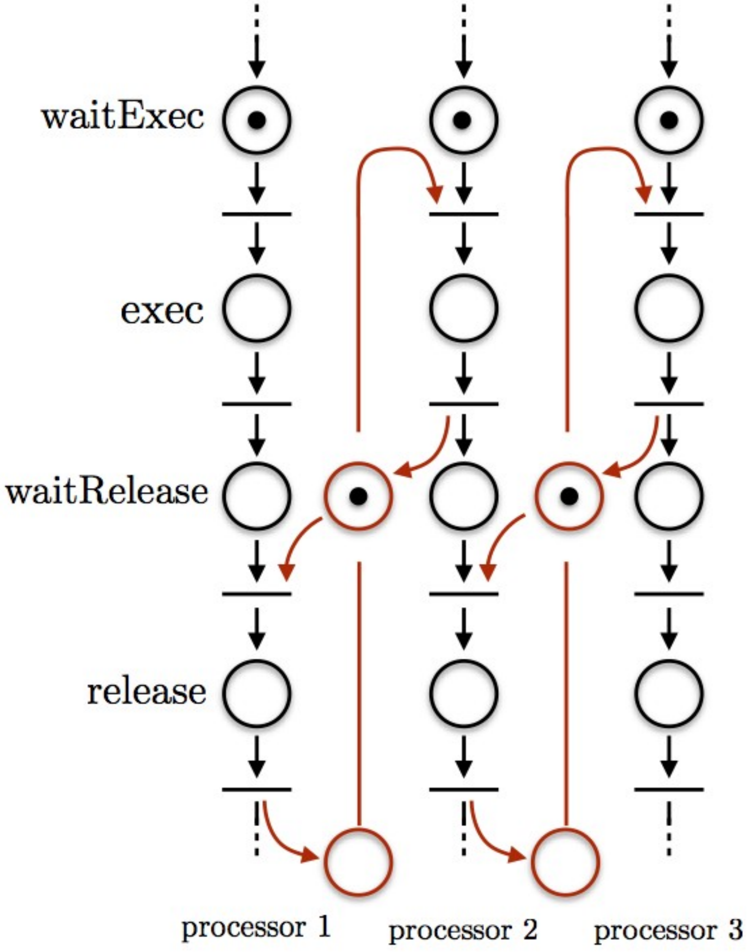
Tree of processors. Each processor is represented as a box, which can be connected to one other. In this tree, Inner Hair Cell is the child of Filterbank, and Filterbank is the parent of Inner Hair Cell.



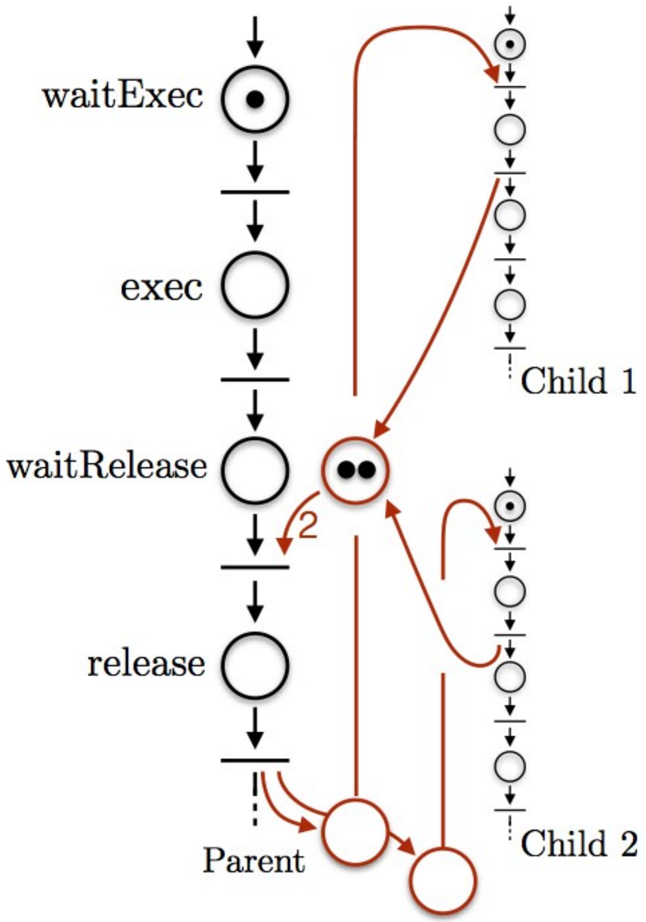
State machine and memory management of a processor.

http://twoears.eu/

Serial and parallel



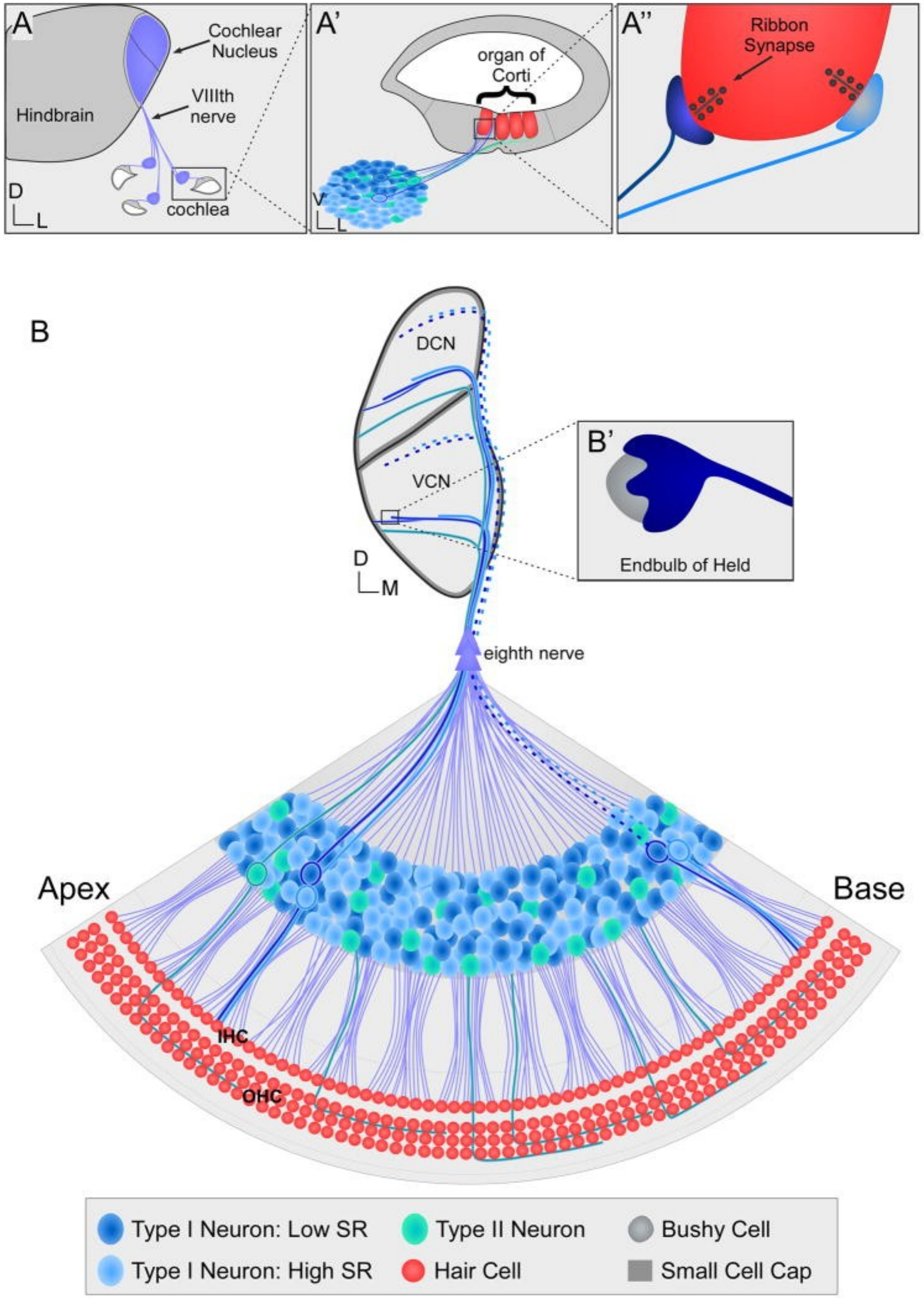
Petri net for a serial chain of processors, highlighting vertical concurrency between 3 processors.



Petri net for a parallel chain of processors, highlighting horizontal concurrency between 2 processors.

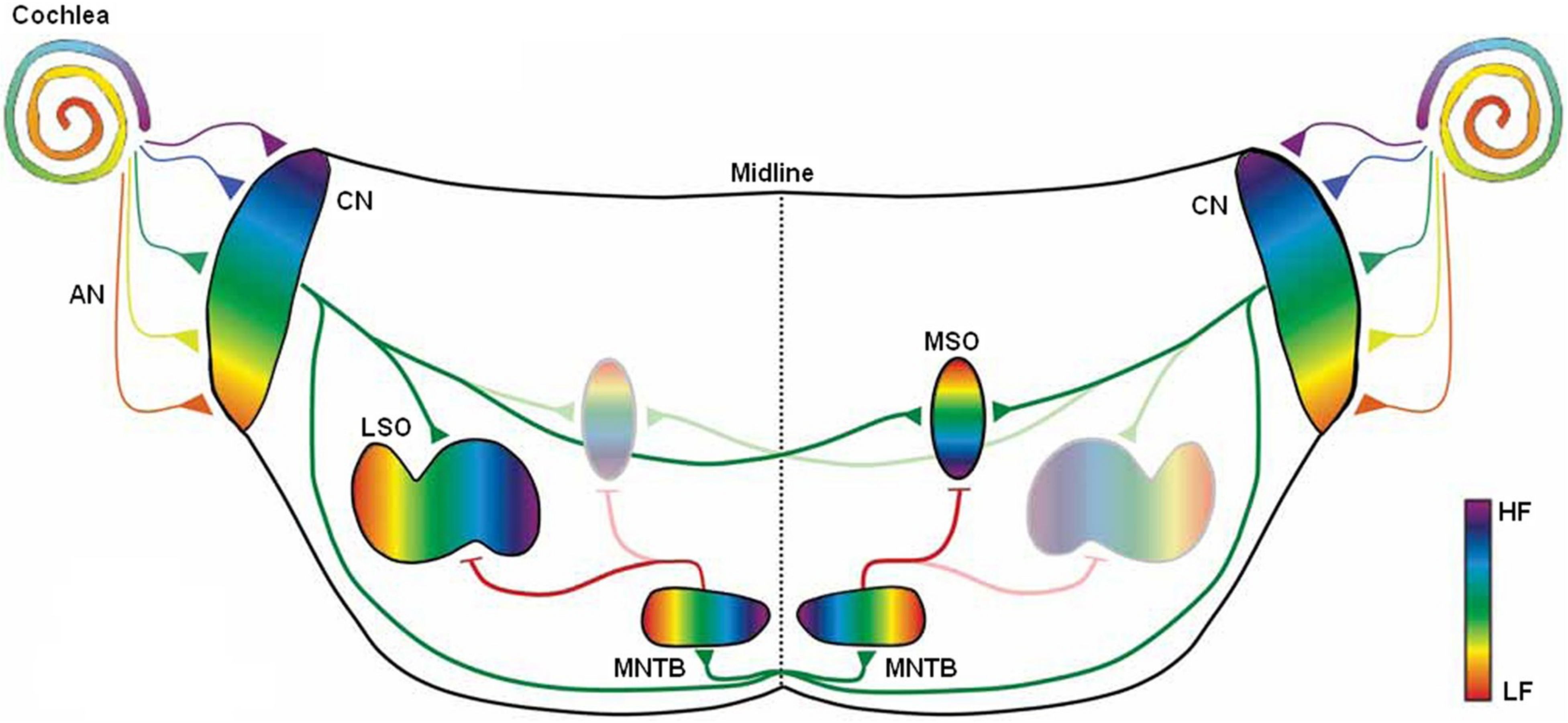
Appendix B:

Bottom-up and top-down auditory pathways



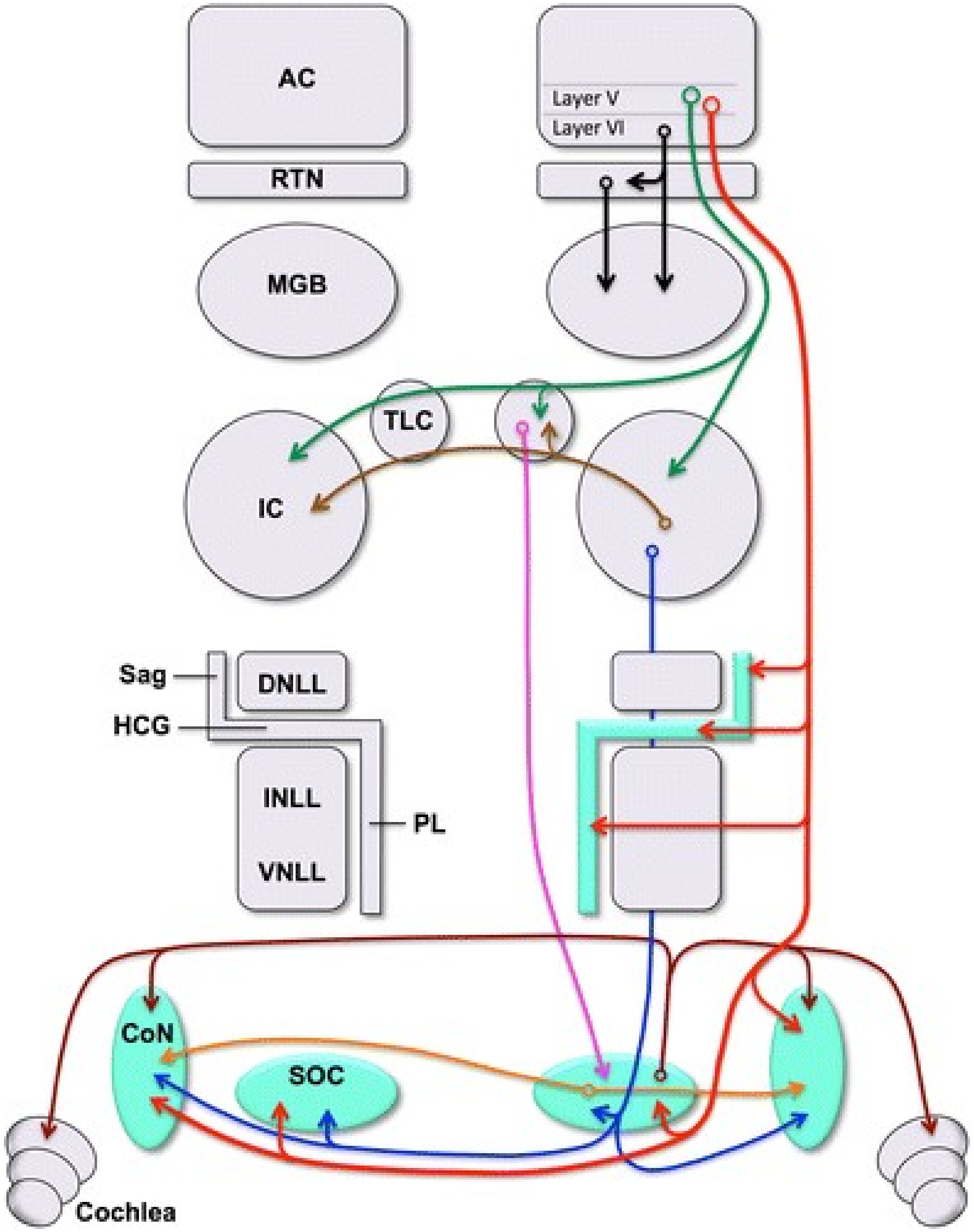
The organization of connections from the ear to the brain
(A) Diagrammatic view of a cross-section through the hindbrain at the level of the cochlea. Dorsal is up; lateral is right. Spiral ganglion neurons (blue) convey auditory information from the cochlea via the eighth (VIIIth) nerve, which arborizes within the cochlear nucleus in the hindbrain. A single turn of the cochlea is boxed and shown at higher power in A′. (A′) By convention, the orientation is flipped such that ventral is now up. Projections from spiral ganglion neurons (light blue, dark blue and green) penetrate the cochlear duct to reach the organ of Corti, which houses the hair cells (red). In this cross-sectional view, spiral ganglion neurons with a low spontaneous rate (SR) (dark blue neurons) are located more ventrally than those with high SR (light blue neurons). The boxed region is shown at higher power in A″. (A″) Spiral ganglion neurons receive information from hair cells via ribbon synapses. Connections made by low SR neurons (dark blue synapse) are located on the neural (also called the modiolar) side of the hair cell, while those made by high SR neurons (light blue synapse) are located on the on the abneural side. In this diagram, neural is to the left and abneural is to the right. (B) Schematic view of a wedge from a flatmounted cochlea (bottom) and its connections with the cochlear nucleus complex (top). In the cochlea, peripheral projections are corralled in radial bundles that pass through the spiral lamina to the hair cells (red). Low SR (dark blue) and high SR (light blue) Type I neurons contact inner hair cells (IHC). Type II neurons (green) are positioned in the ganglia nearest to the hair cells, and extend a projection past the inner hair cells and turn towards the base, with each projection contacting multiple outer hair cells (OHC) along its length. Information is conveyed to the cochlear nucleus by central axons, which bundle together to form the eighth nerve (double arrowhead). Upon entering the brainstem, individual axons bifurcate. The ascending projections terminate with bouton endings in the dorsal cochlear nucleus (DCN), while the descending projections target the ventral cochlear nucleus (VCN), where they form boutons with a variety of post-synaptic target neurons as well as unusual endbulb of Held synapses with bushy cells (B′). Within each division of the cochlear nucleus, auditory axons are tonotopically organized, such that high frequency information from the base of the cochlea is processed dorsally (dotted lines) and low frequency information from the apex is processed more ventrally (solid lines). In addition, the central projections from neurons with low spontaneous firing rates project more laterally than those with high spontaneous firing rates. Type II neurons project to the small cell cap that surrounds the cochlear nucleus complex (dark gray), as do some arbors from low SR fibers.

J. M. Appler and L. V. Goodrich, “Connecting the ear to the brain: Molecular mechanisms of auditory circuit assembly,” Prog. Neurobiol., vol. 93, no. 4, pp. 488 – 508, 2011.



For clarity, only the LSO or MSO are shown on each side. Except for the auditory nerve, excitatory connections are shown in green and inhibitory connections are shown in red. AN, auditory nerve; CN, cochlear nucleus; HF, high frequency; LF, low frequency.

K. Kandler, A. Clause, and J. Noh, “Tonotopic reorganization of developing auditory brainstem circuits,” Nat. Neurosci., vol. 12, p. 711, May 2009.



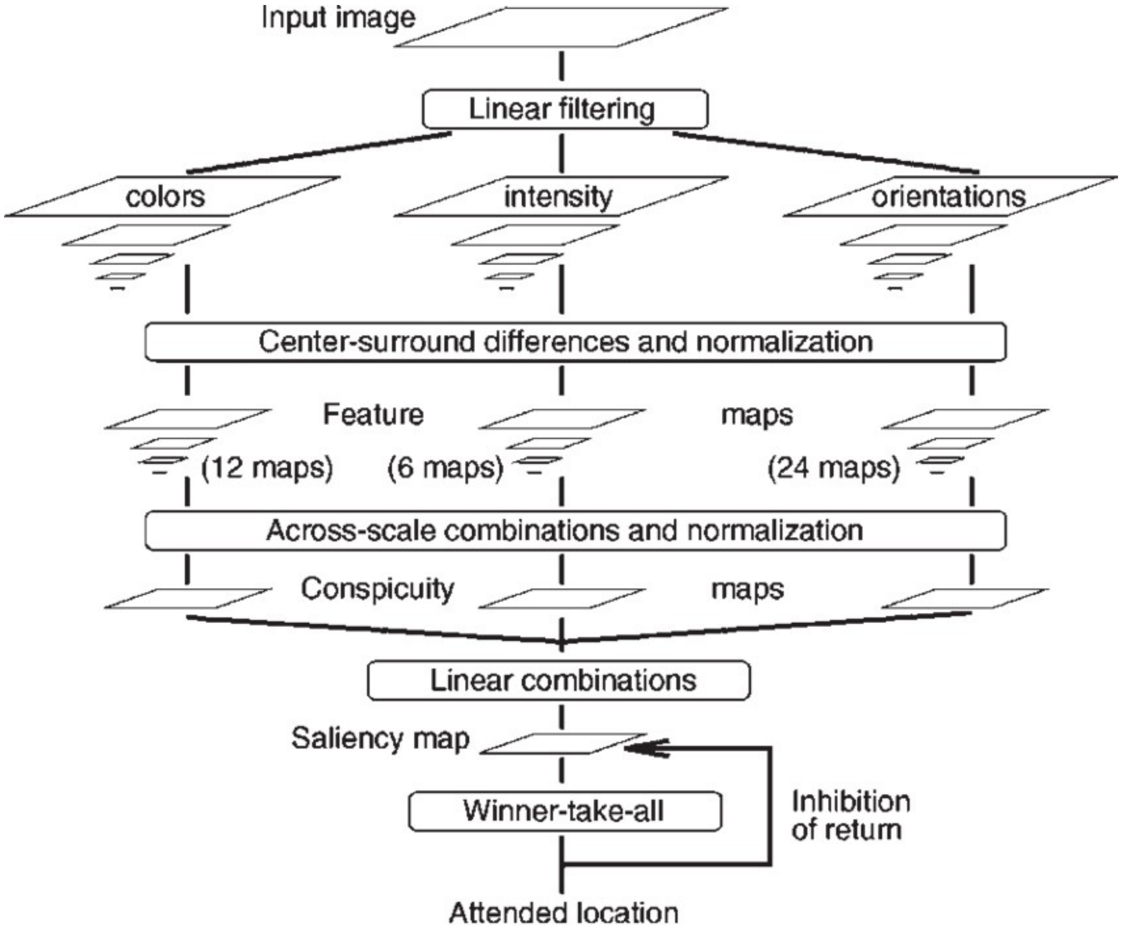
Schematic representation of the main descending auditory pathways, with emphasis on direct corticosubcollicular projections. The scheme illustrates the principal projections through which the auditory cerebral cortex (AC) of the right side exerts its influence over lower auditory centers. The origin of each projection has been represented with an open circle and the targets with arrows. The direct corticosubcollicular projections have been depicted in red, and the nuclei innervated by them are highlighted in green. Other illustrated descending pathways include corticothalamic projections (black), corticotectal projections to the inferior colliculus (IC) and the tectal longitudinal column (TLC) (dark green), commissural projections of the IC (brown), descending projections of the IC (blue), projection from the TLC to the superior olivary complex (SOC) (pink), projections from the SOC to the cochlear nuclei (CoN) (orange), and olivocochlear projections (dark crimson). Note that most of the subcollicular centers innervated by the auditory cortex are also the target of descending projections from other auditory nuclei that are also innervated by the cortex. For the sake of simplicity, no individual nuclei or subdivisions or have been depicted in the medial geniculate body (MGB), IC, SOC, and CoN. DNLL dorsal nucleus of the lateral lemniscus, HCG horizontal cell group, INLL intermediate nucleus of the lateral lemniscus, PL paralemniscal regions (including the medial paralemniscal nucleus), RTN reticular thalamic nucleus, Sag nucleus sagulum, VNLL ventral nucleus of the lateral lemniscus

E. Saldaña, “All the way from the cortex: a review of auditory corticosubcollicular pathways,” The Cerebellum, vol. 14, no. 5, pp. 584–596, 2015.

Appendix C:

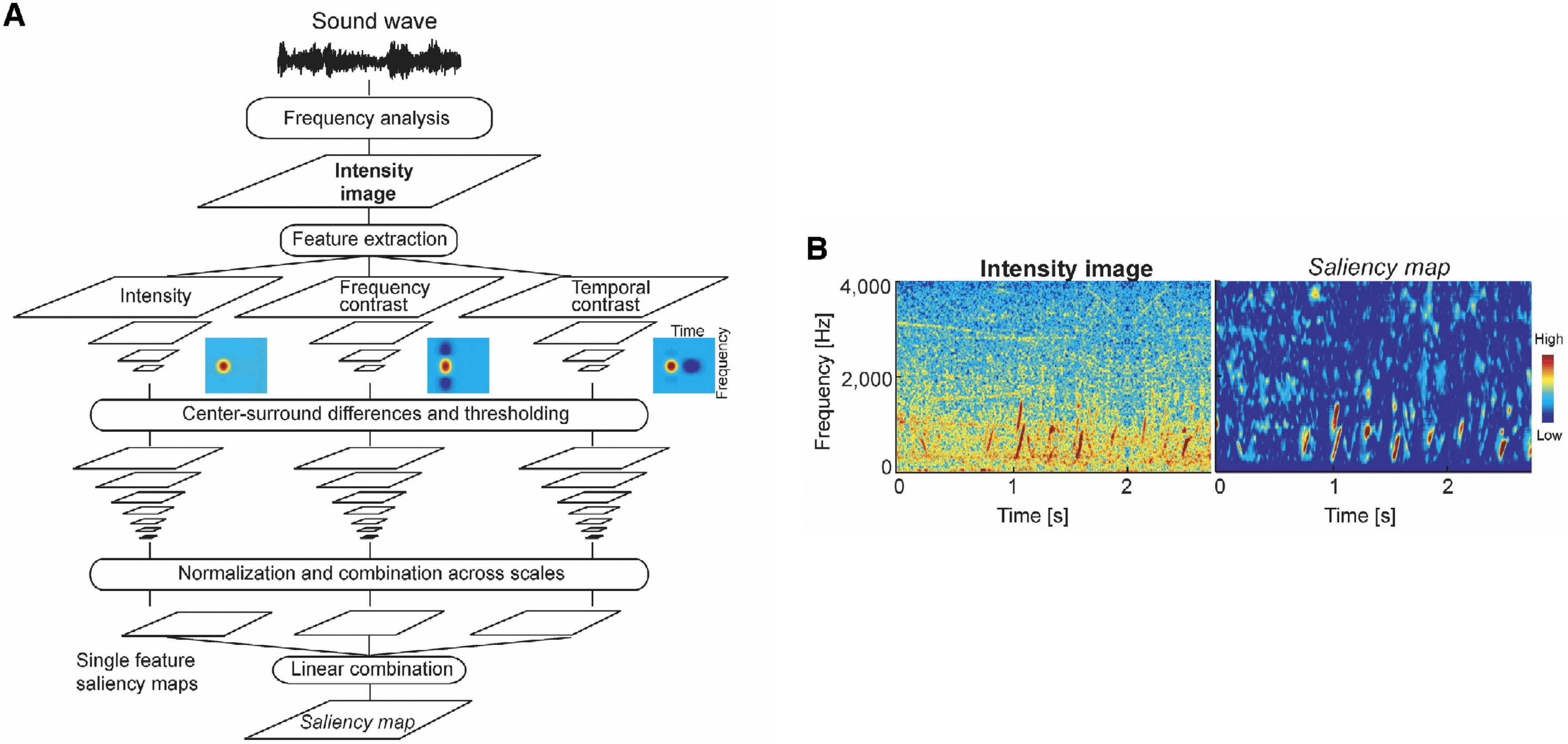
Parallelisms between visual and auditory attention

Bottom-up visual attention model: Saliency map



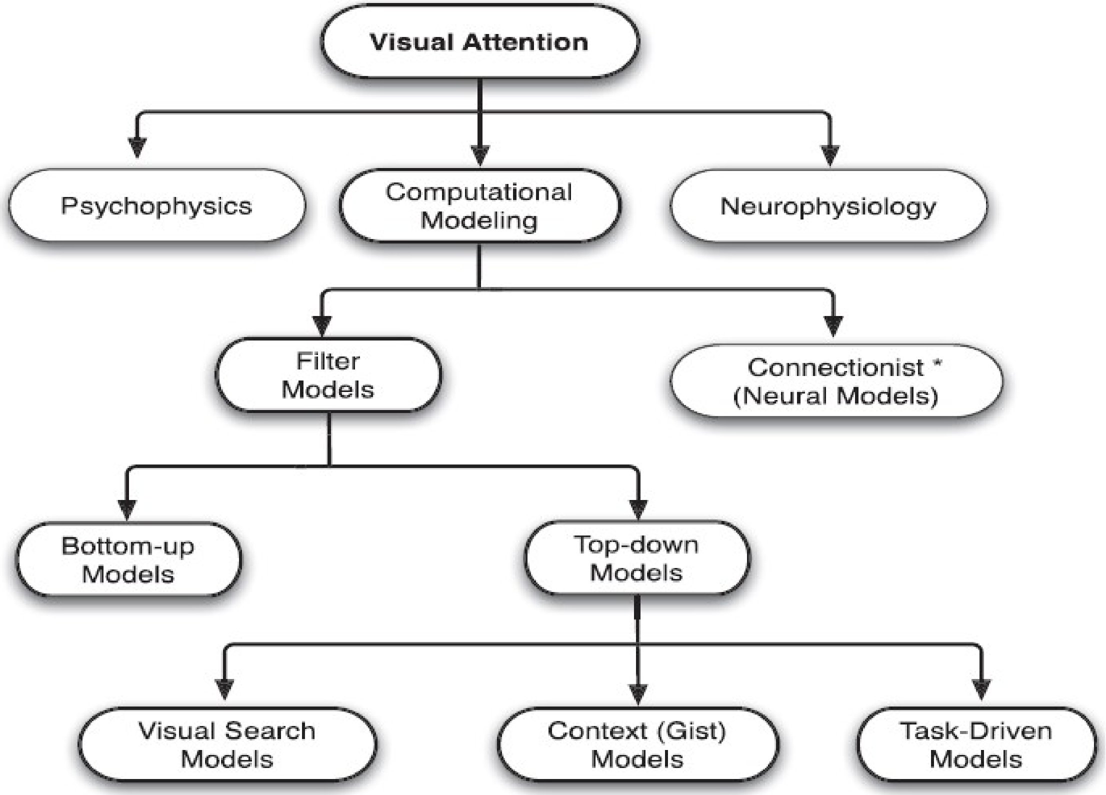
C. Koch and S. Ullman, “Shifts in selective visual attention: towards the underlying neural circuitry.,” Hum. Neurobiol., vol. 4, no. 4, pp. 219–227, 1985.

Bottom-up auditory attention saliency map



C. Kayser, C. I. Petkov, M. Lippert, and N. K. Logothetis, “Mechanisms for allocating auditory attention: An auditory saliency map,” Curr. Biol., vol. 15, no. 21, pp. 1943–1947, 2005.

Top-down visual attention models

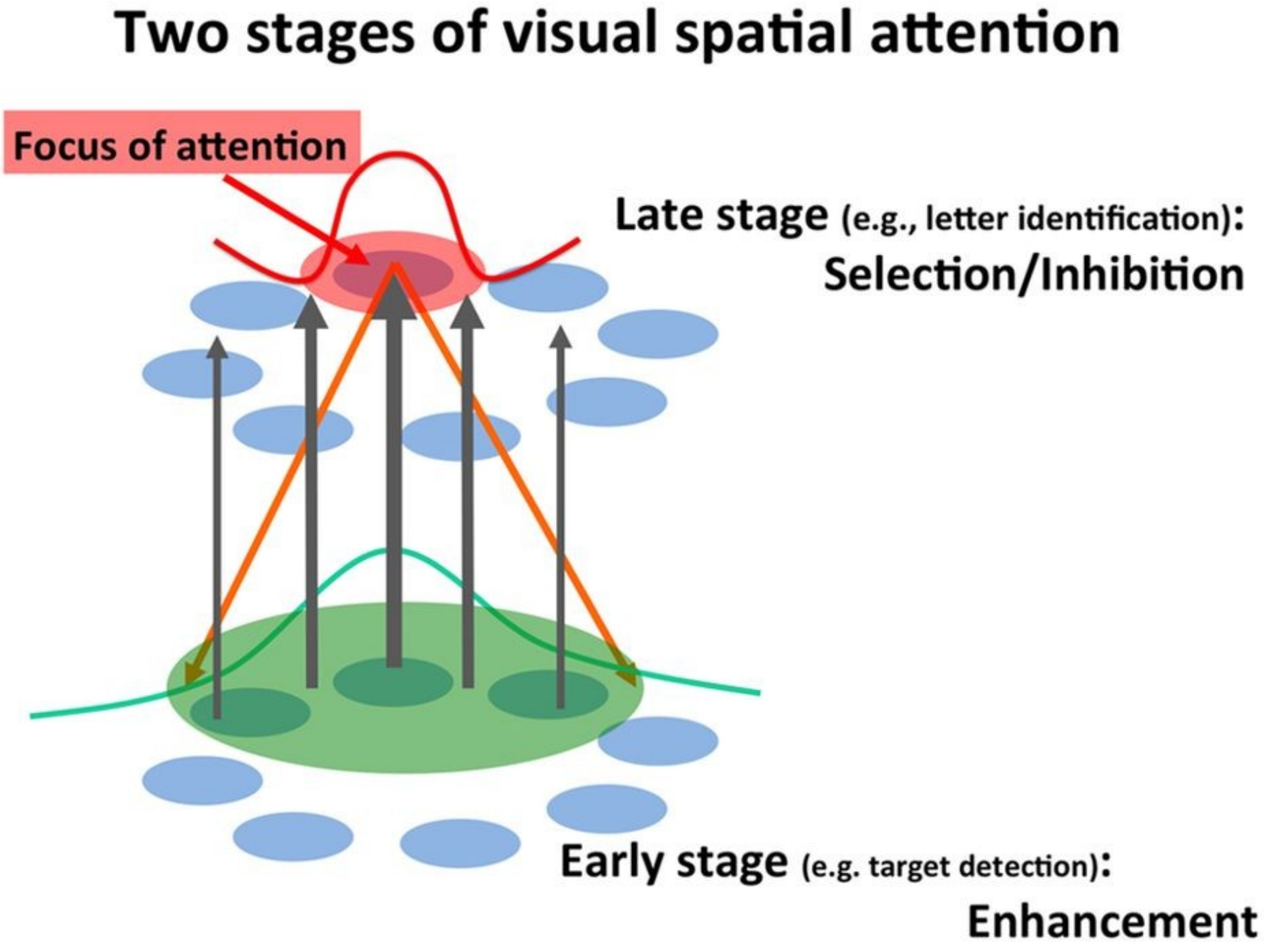


* Connectionist approaches use realistic neuron models while filter models use functions believed to be performed by single neurons or neural networks.

Fig. 1. Taxonomy of visual attention studies. Ellipses with solid borders illustrate our scope in this paper.

A. Borji and L. Itti, “State-of-the-art in visual attention modeling,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 1, pp. 185–207, Jan. 2013.

Top-down visual attention model



S. Shioiri, H. Honjyo, Y. Kashiwase, K. Matsumiya, and I. Kuriki, “Visual attention spreads broadly but selects information locally,” Sci. Rep., vol. 6, p. 35513, Oct. 2016.

Top-down auditory attention model

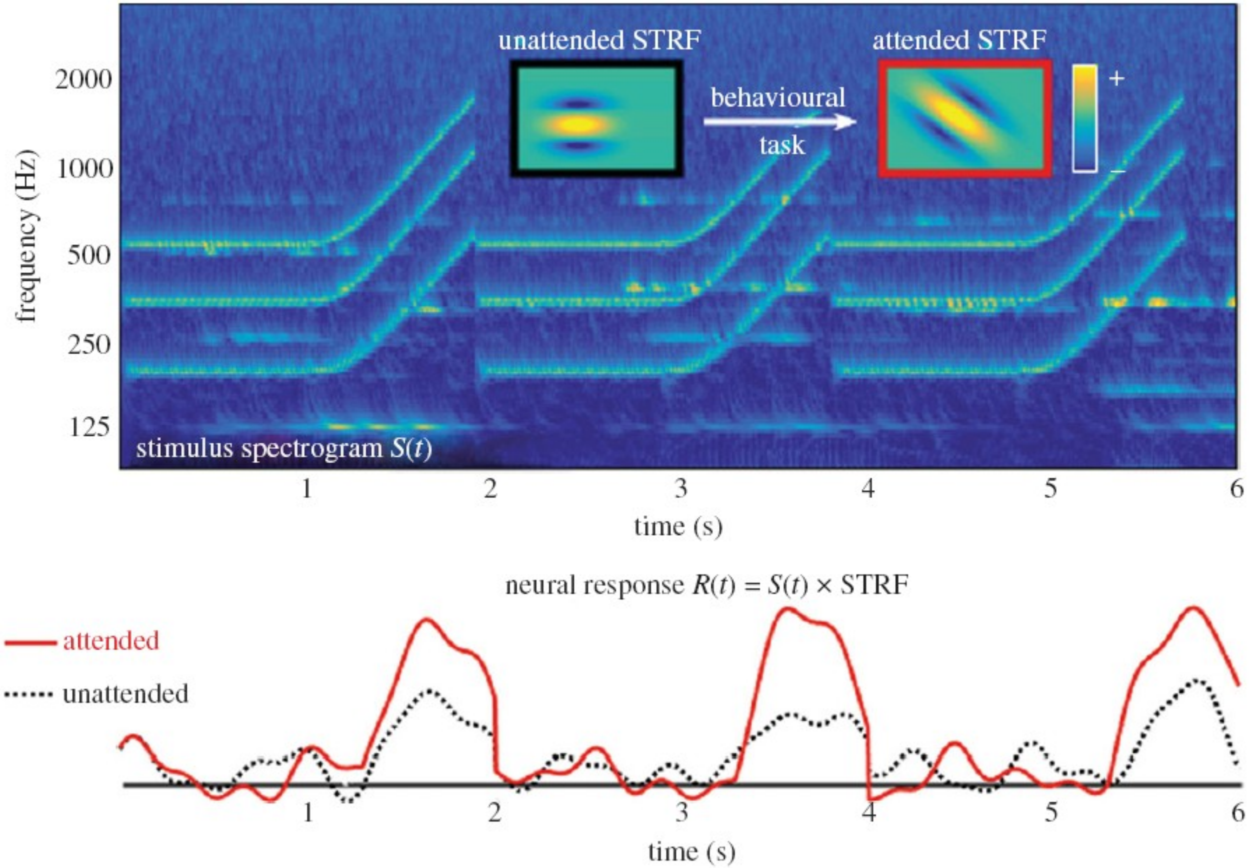
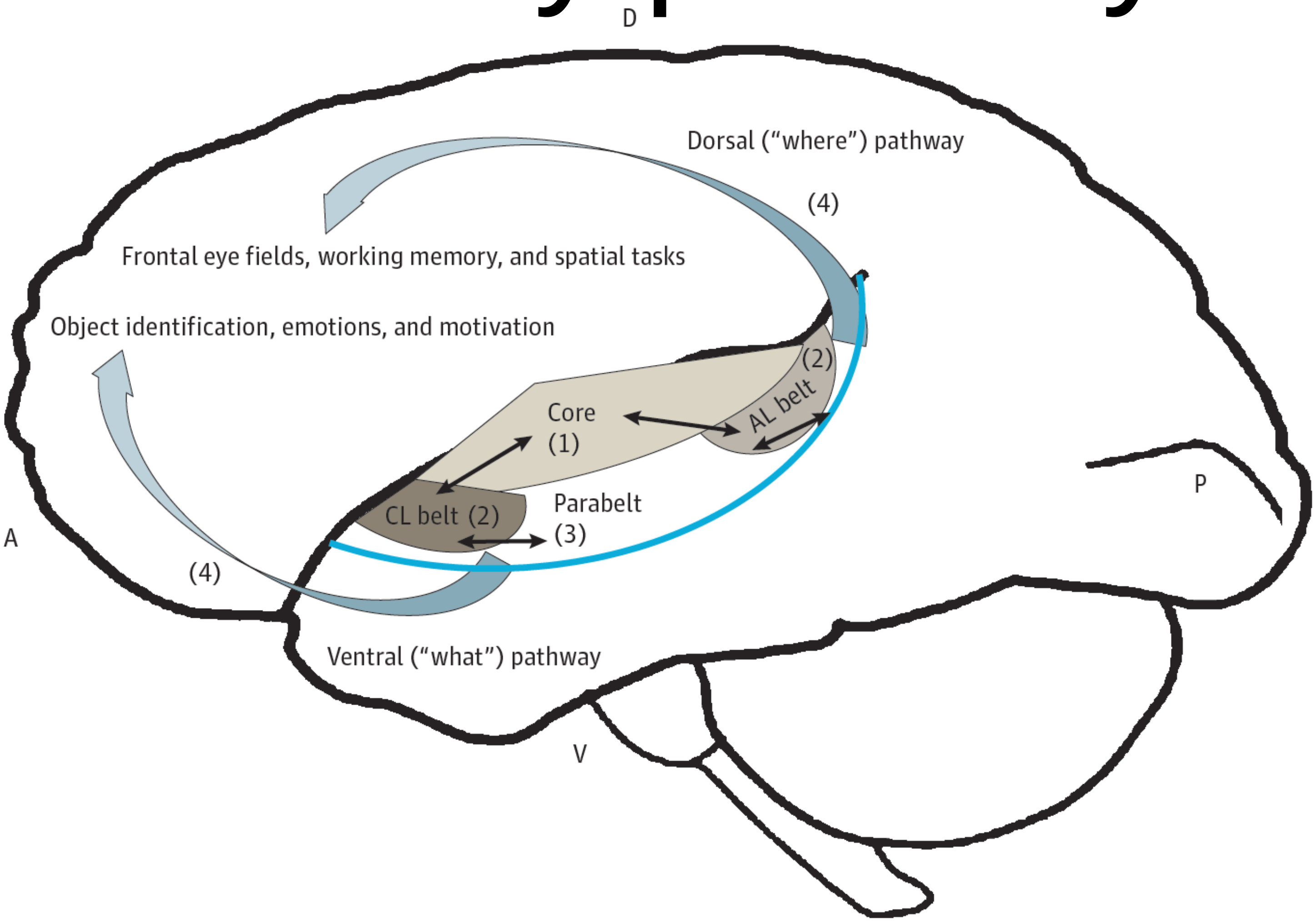


Figure 3. Attending to a particular sound characteristic tunes the neural spectro-temporal receptive fields (STRFs) and boosts the neural signal at times of attended event. Violin notes are overlaid with frequency modulations (FMs), illustrated with the spectrogram $S(t)$. When instructed to attend to the FM segments, the STRF adapts to the orientation of the modulations, resulting in an enhancement in the neural response $R(t)$.

E. M. Kaya and M. Elhilali, “Modelling auditory attention,” Phil. Trans. R. Soc. B Biol. Sci., vol. 372, no. 1714, Jan. 2017.

The “where” and “what” auditory pathways



W.-J. Wang, X.-H. Wu, and L. Li, “The dual-pathway model of auditory signal processing,” Neurosci. Bull., vol. 24, no. 3, pp. 173–182, Jun. 2008.
S. Gokhale, S. Lahoti, and L. Caplan, “The neglected neglect: Auditory neglect,” JAMA Neurology, vol. 70, no. 8, pp. 1065–1069, 2013.

Multisensory integration

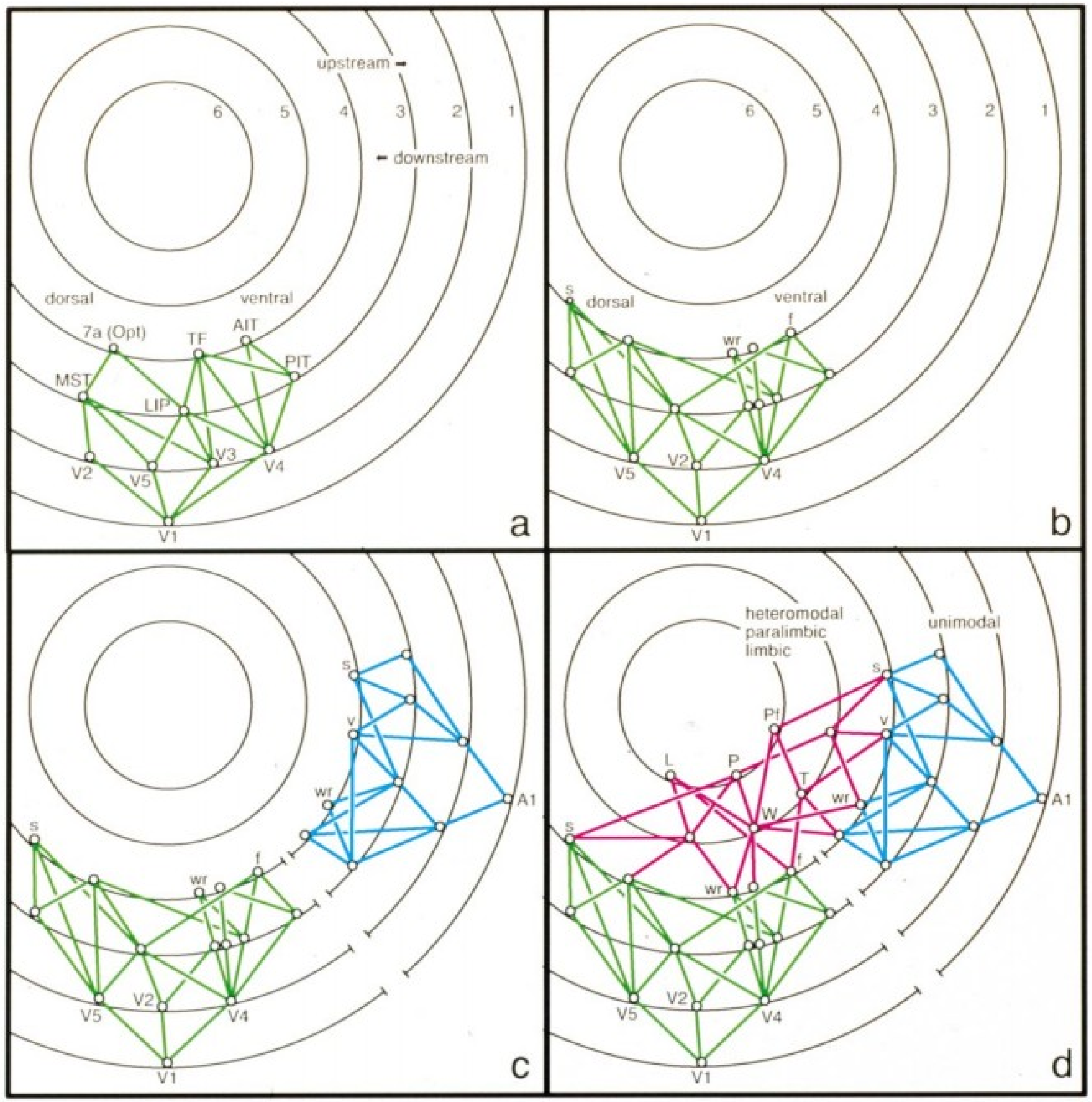


Fig. 2 Each concentric ring represents a different synaptic level. Any two consecutive levels are separated by at least one unit of synaptic distance. Level 1 is occupied by the primary sensory cortex. Small empty circles represent macroscopic cortical areas or 'nodes', one to several centimetres in diameter. Nodes at the same synaptic level are reciprocally interconnected by the black arcs of the concentric rings. Coloured lines represent reciprocal monosynaptic connections from one synaptic level to another. **(a)** Visual pathways as demonstrated by experimental neuroanatomical methods in the macaque brain. **(b)** The inferred organization of the homologous visual pathways in the human brain. **(c)** Visual (green) and auditory (blue) pathways in the human brain. **(d)** Visual (green), auditory (blue) and transmodal (red) pathways in the human brain. In **b**, **c** and **d**, the anatomical details of individual pathways are inferred from experimental work in the monkey. The anatomical identity of many of the nodes is not specified because their exact anatomical location is not critical. This review is guided by the hypothesis that these types of anatomical interconnections and functionally specialized nodes exist in the human brain even though their exact location has not yet been determined. The terms 'dorsal' and 'ventral' in **a** and **b** refer to the separation of visuo-fugal pathways, especially at the fourth synaptic level, into dorsal and ventral streams of processing. The gaps in the circles at the first four levels indicate the absence of monosynaptic connections between modality-specific components of auditory and visual pathways. Abbreviations: A1= primary auditory cortex; AIT = anterior inferotemporal cortex; f = area specialized for face encoding; L = hippocampal–entorhinal or amygdaloid components of the limbic system; LIP = lateral intraparietal cortex; MST = medial superior temporal cortex; P = heteromodal posterior parietal cortex; Pf = lateral prefrontal cortex; s = area specialized for encoding spatial location; PIT = posterior inferotemporal cortex; T = heteromodal lateral temporal cortex; TF = part of medial inferotemporal cortex; v = area specialized for identifying individual voice patterns; V1 = primary visual cortex; V2, V3, V4, V5 = additional visual areas; W = Wernicke's area; wr = area specialized for encoding word-forms; 7a(Opt) = part of dorsal parieto-occipital cortex.

M.-M. Mesulam, "From sensation to cognition," Brain, vol. 121, no. 6, pp. 1013–1052, 1998.