Auditory Brain Models for the Localization and Identification of Sound

César D. Salvador[1], Ryo Teraoka[2] and Shuichi Sakamoto[3]

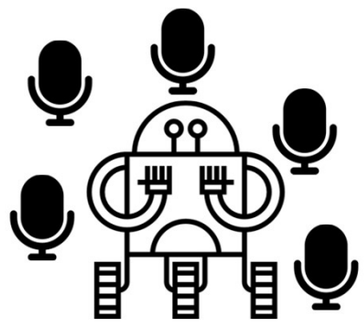
[1]Perception Research

[2]Graduate School of Humanities and Social Sciences, Kumamoto University, Japan

[3]Research Institute of Electrical Communication, Tohoku University, Japan

Importance of auditory brain modeling

- Smart capture, analysis, and rendering of auditory scenes
  - Object-based representation
- Applications
  - Hearing aids
  - 3D audio for VR/AR/XR
  - Machine hearing



Brain modeling concepts

- Neural connectivity
  - Structural: Anatomical links
  - Functional: Statistical association
- Neural pathway
  - Bottom-up: Stimulus-driven (e.g., localization of sound)
  - Top-down: Goal-oriented (e.g., identification of sound)

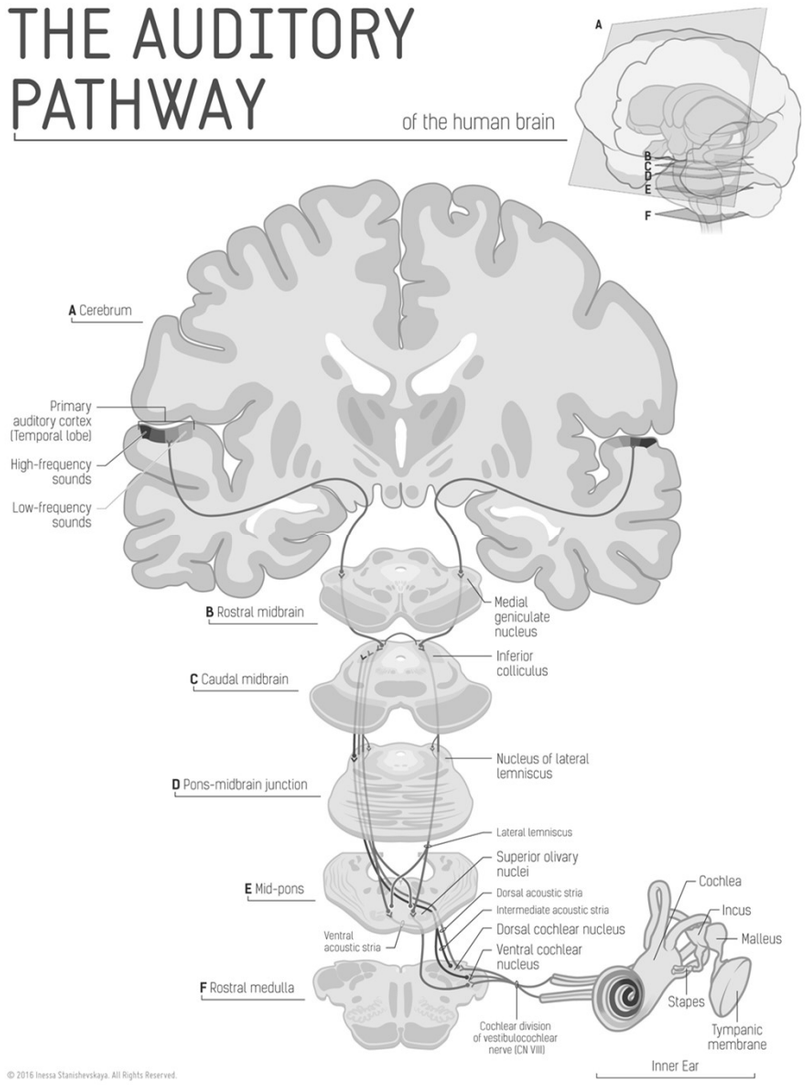
Considerations

↑ High-scale functionality

↑ Algorithmic efficiency

↓ Biological detail

Localization and identification of sound



Cognition: Interpretation of features (e.g. for **identification**)

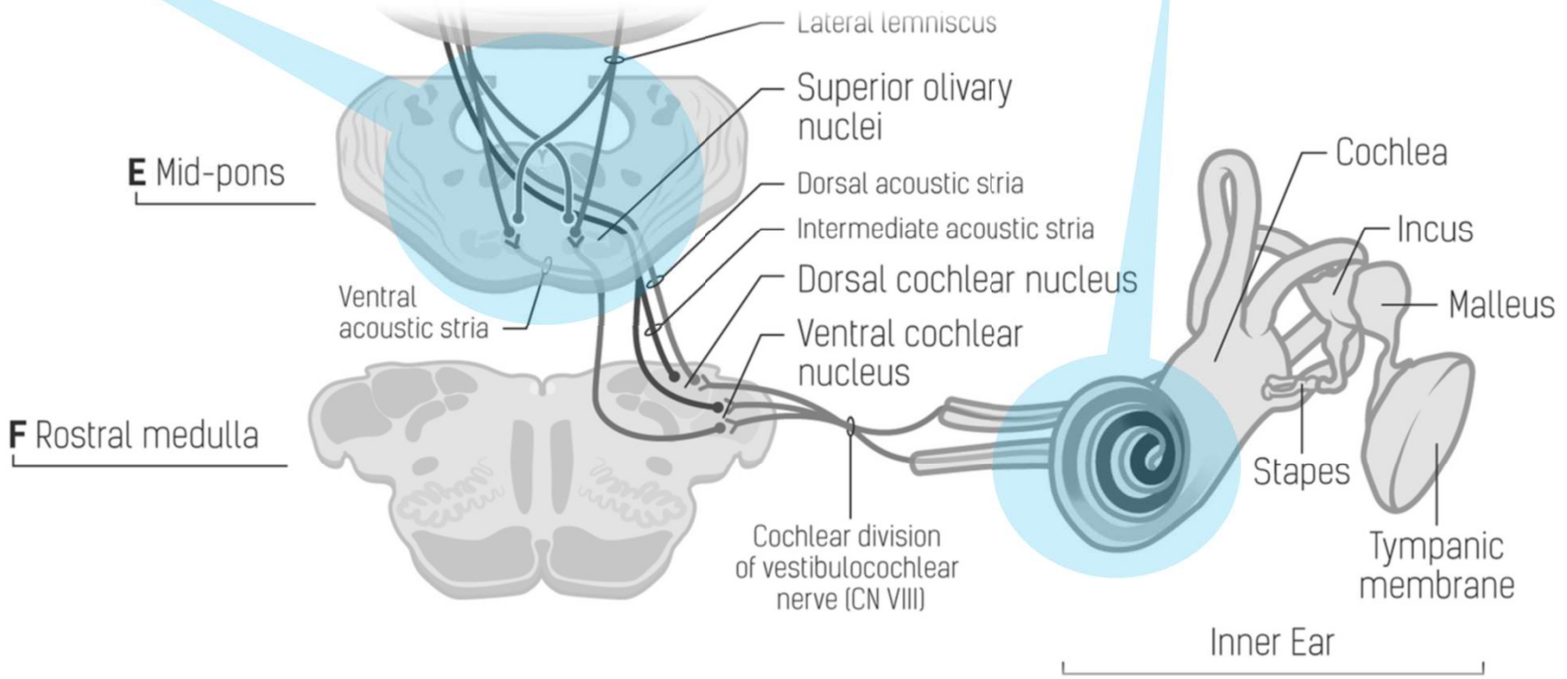
Multimodal integration of hearing, vision, and touch

Sensation: Primary feature extraction (e.g., for **localization**)

Primary features

The superior olivary complex (SOC) extracts **interaural features**

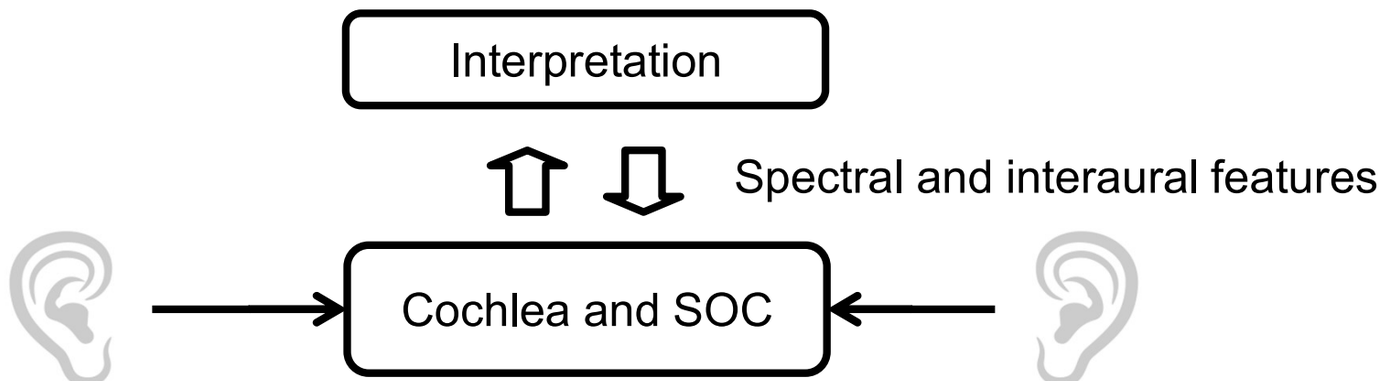
The cochlea extracts **spectral features**



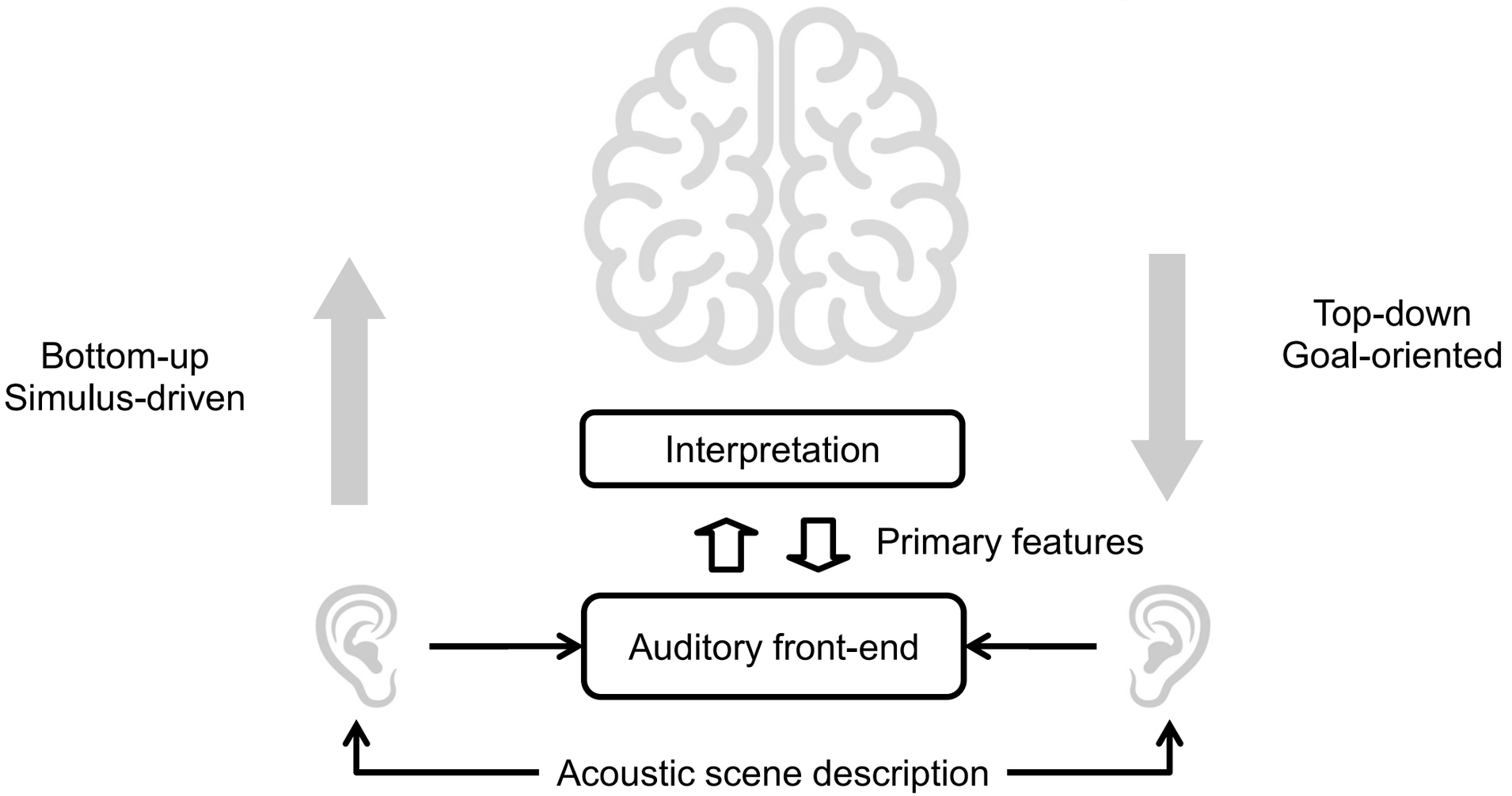
Interpretation of features

"Where" dorsal pathway

"What" ventral pathway



Framework for auditory modeling

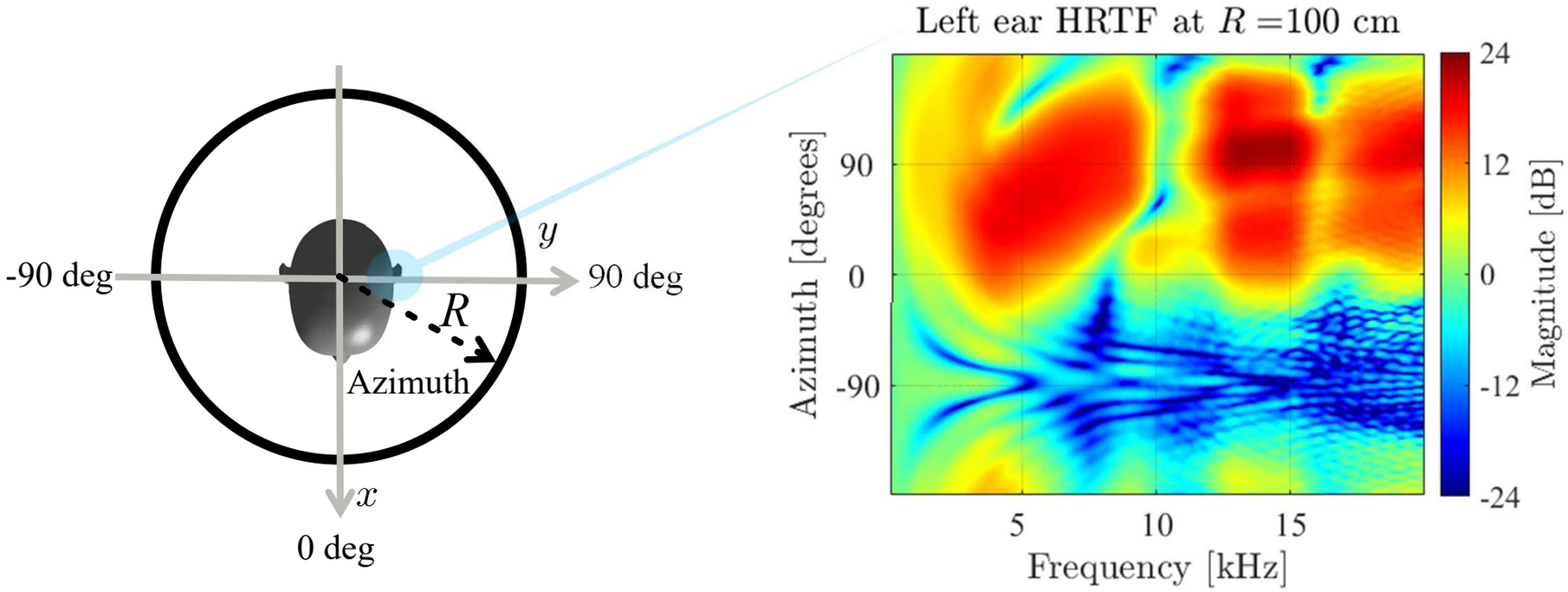


Tools for auditory modeling

- Development system for auditory modeling (L. P. O'Mard et al., 1986-2007)
- Auditory toolbox (M. Slaney et al., 1993-1998)
- Auditory modeling toolbox (P. L. Søndergaard et al., 2013)
- Two!Ears Auditory Model (A. Raake et al., 2013-2016)
  - Modular
  - Scalable
  - Object-oriented
  - Real time

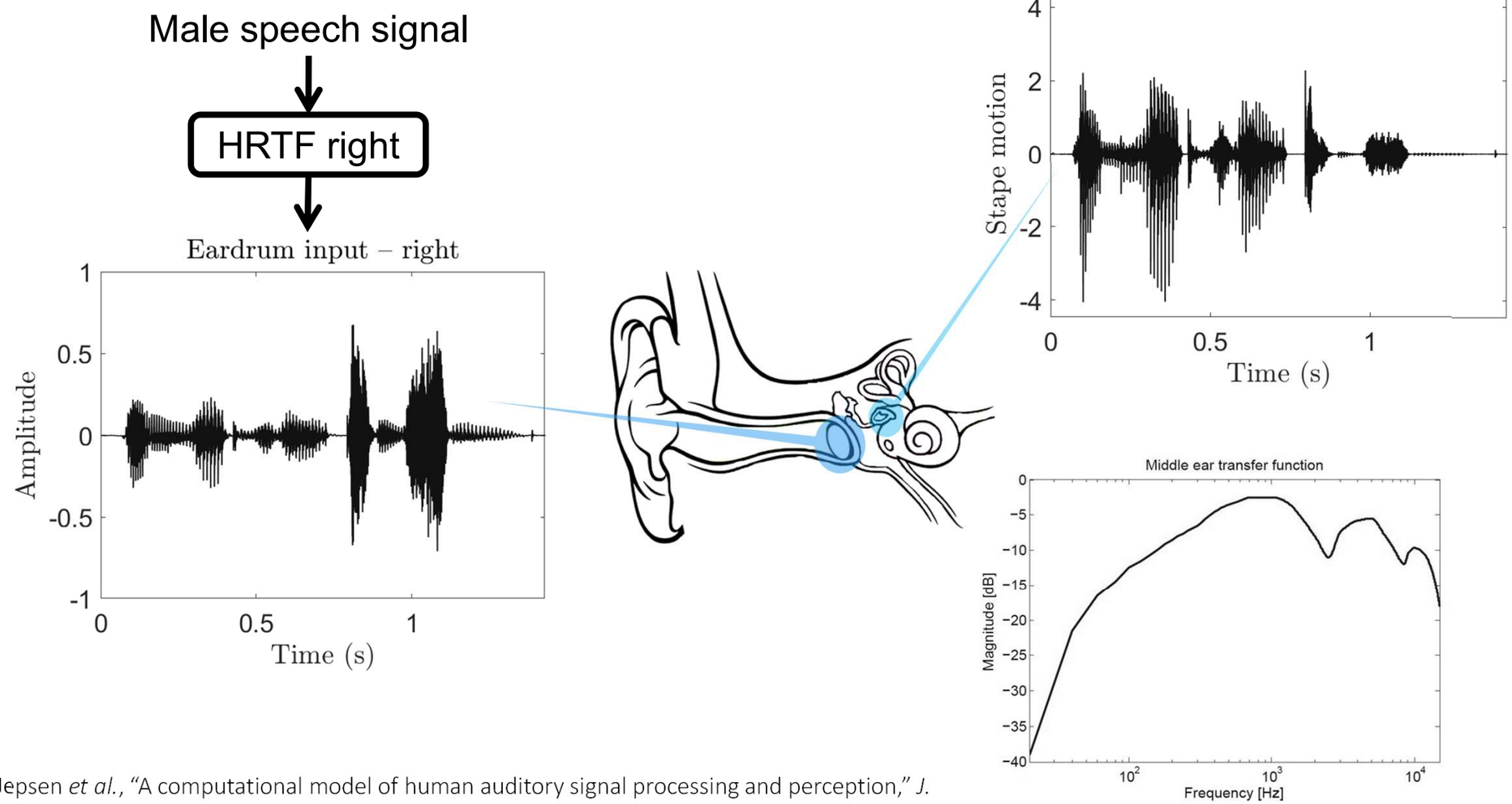
Acoustic filtering of the head and ears

Head-related transfer functions (HRTFs)



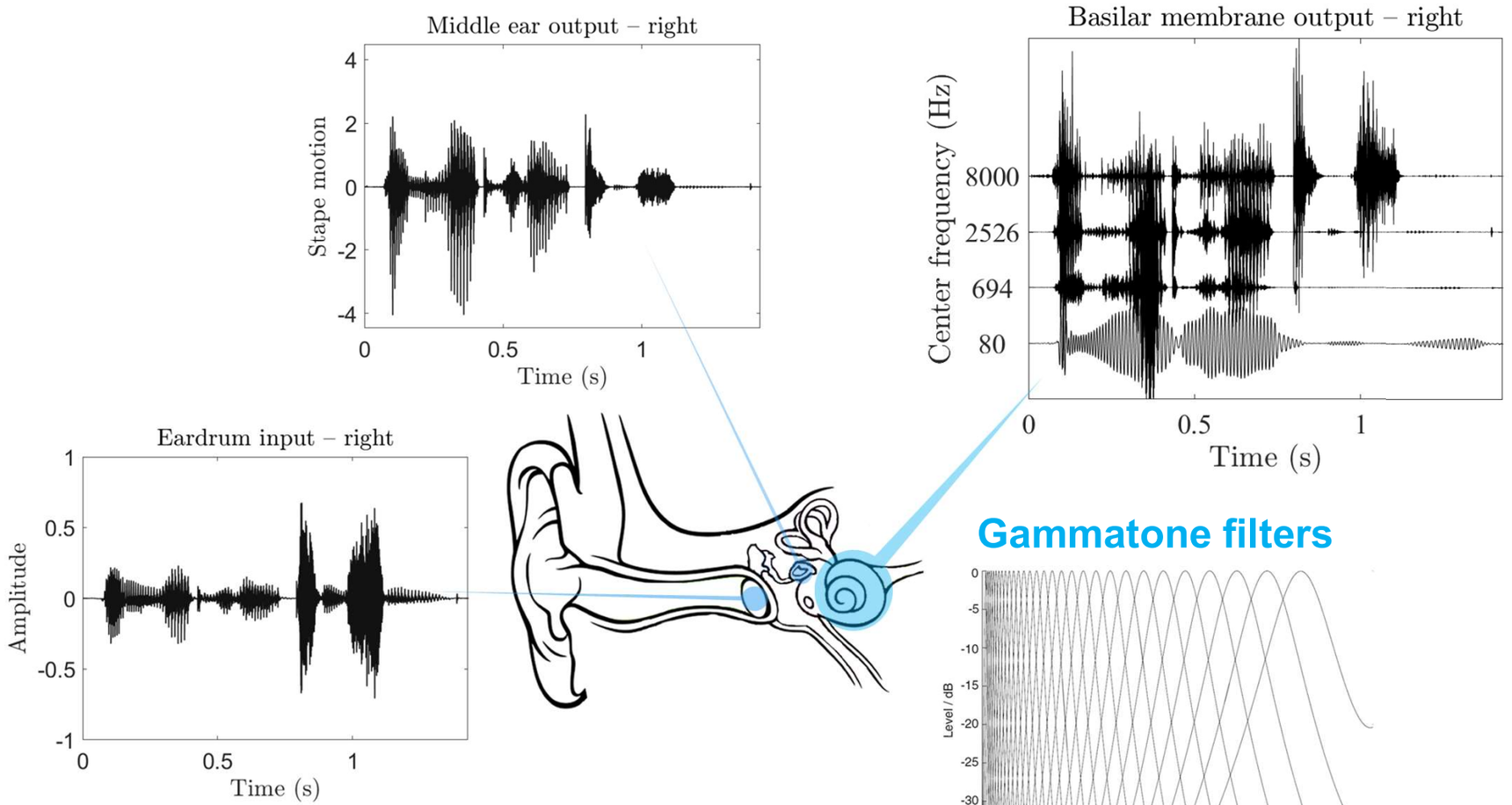
C. D. Salvador *et al.*, "Dataset of near-distance head-related transfer functions calculated using the boundary element method," *Proc. Audio Eng. Soc. Int. Conf. Spatial Reproduction —Aesthetics and Science—*, Tokyo, Japan, 2018.

Middle ear



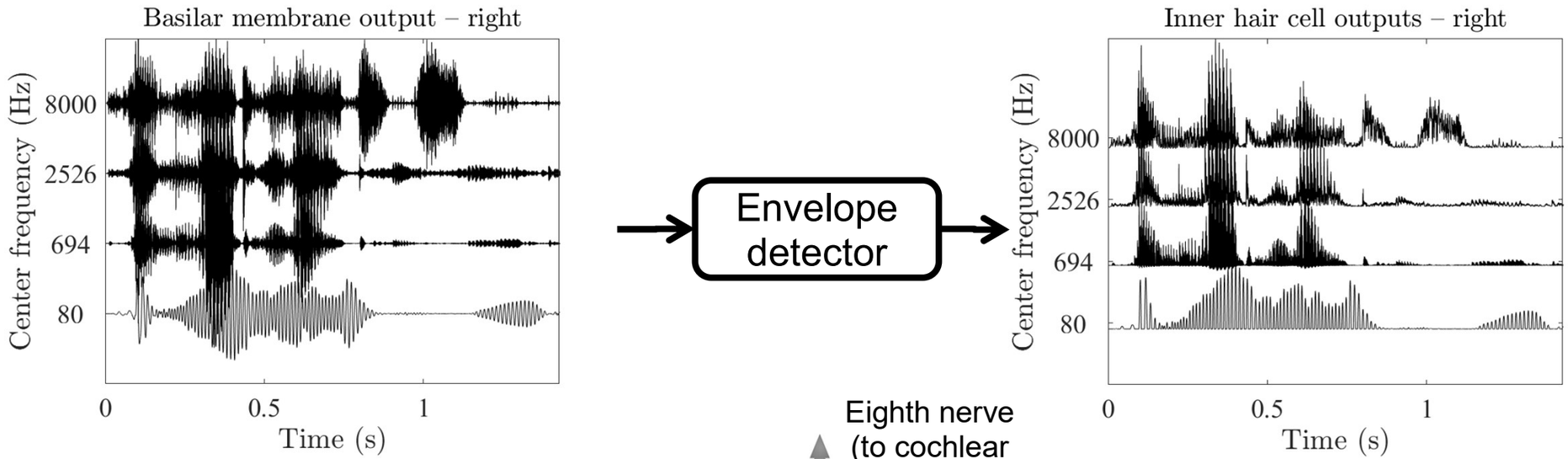
M. L. Jepsen *et al.*, "A computational model of human auditory signal processing and perception," *J. Acoust. Soc. Am.*, vol. 124, no. 1, pp. 422–438, 2008.

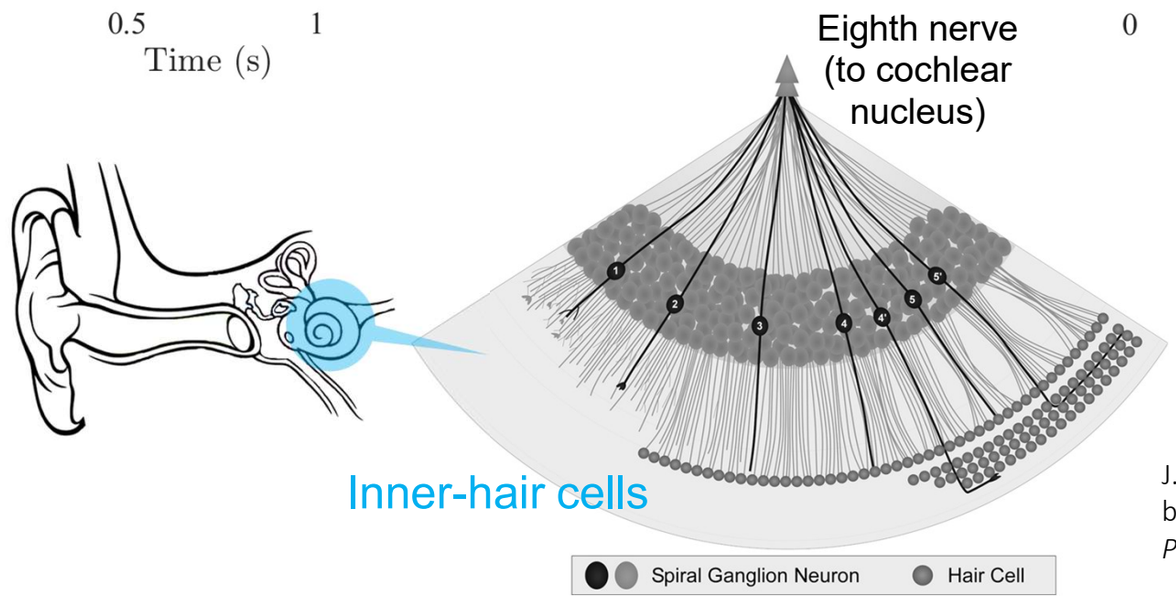
Cochlea



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

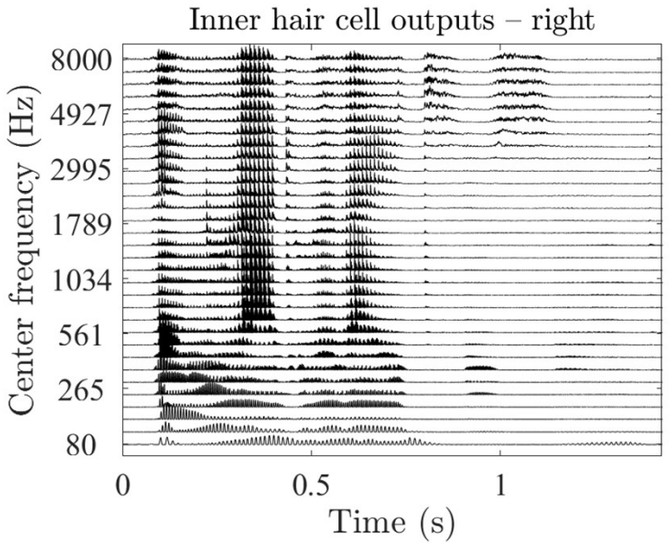
Hair cells



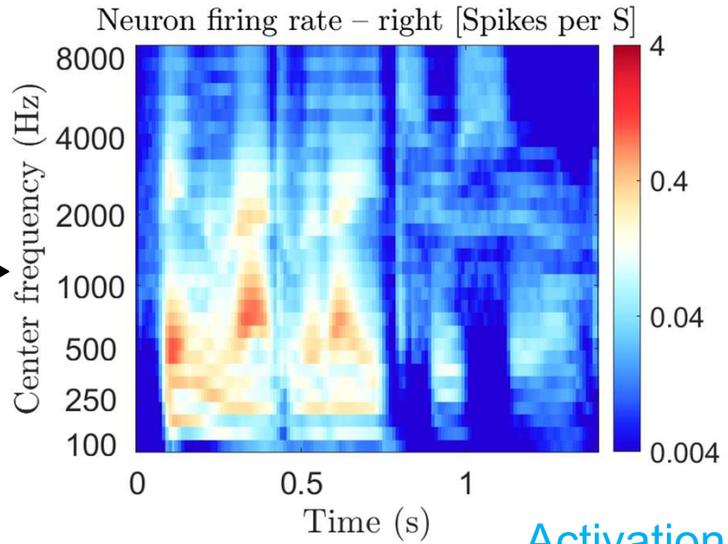


J. M. Appler and L. V. Goodrich, "Connecting the ear to the brain: Molecular mechanisms of auditory circuit assembly," *Prog. Neurobiol.*, vol. 93, no. 4, pp. 488 – 508, 2011.

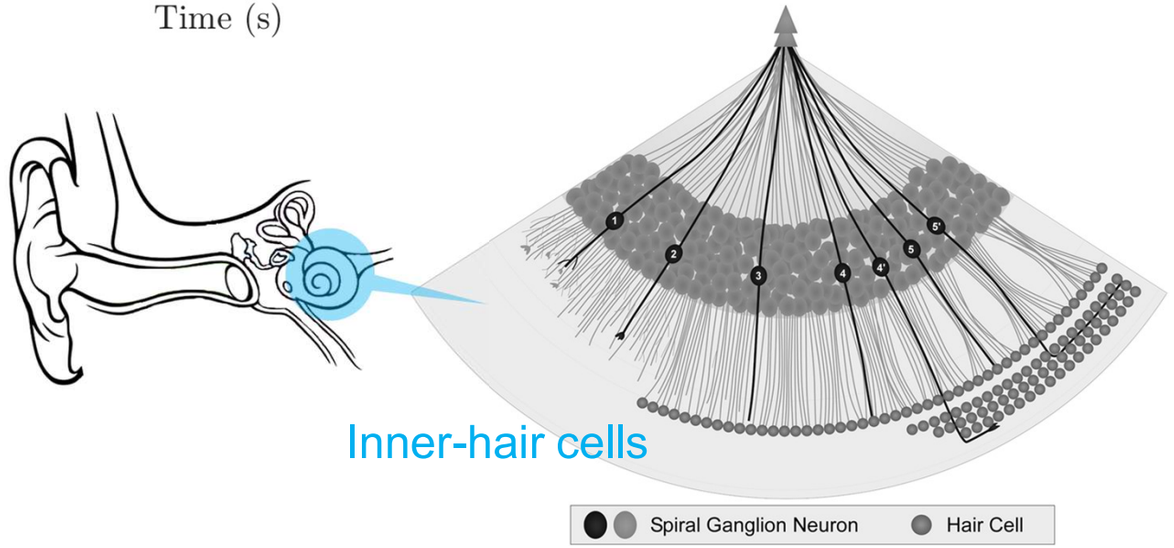
Activation rate map



Leaky integrator



Activation rate maps are the primary spectral features

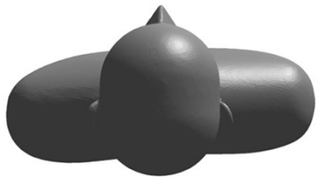


J. M. Appler and L. V. Goodrich, "Connecting the ear to the brain: Molecular mechanisms of auditory circuit assembly," *Prog. Neurobiol.*, vol. 93, no. 4, pp. 488 – 508, 2011.

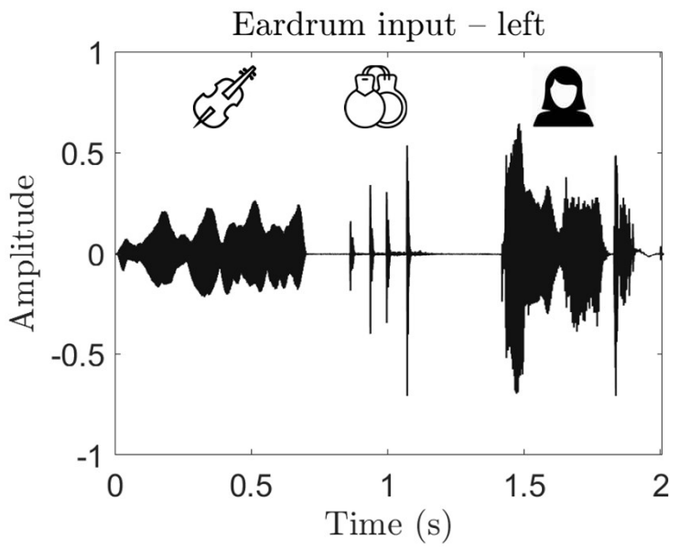
Example: Localization and identification

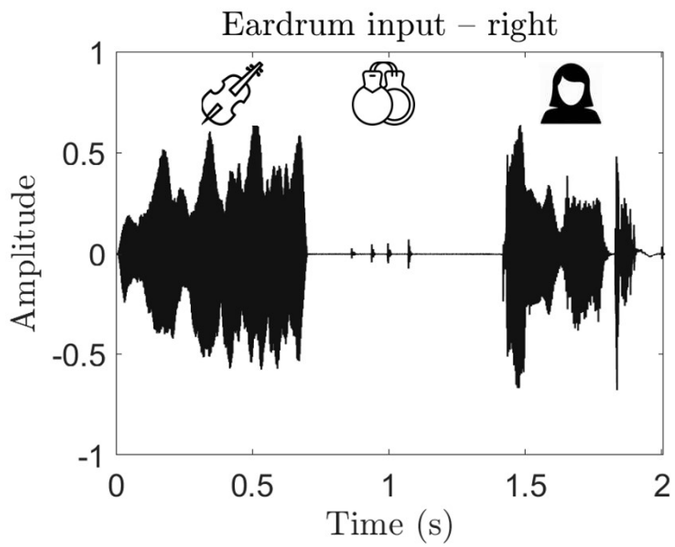
Speech
(50 cm, 0°)

Castanets
(50 cm, 70°)

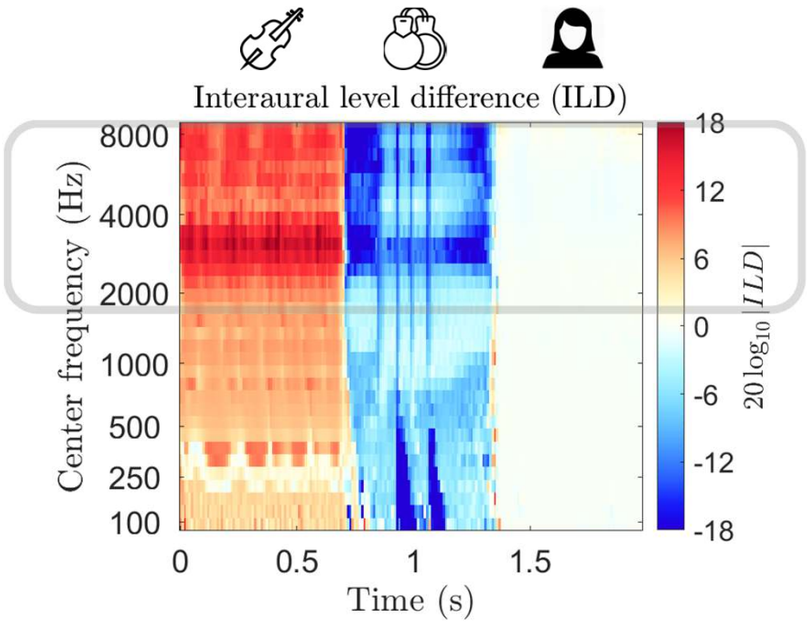


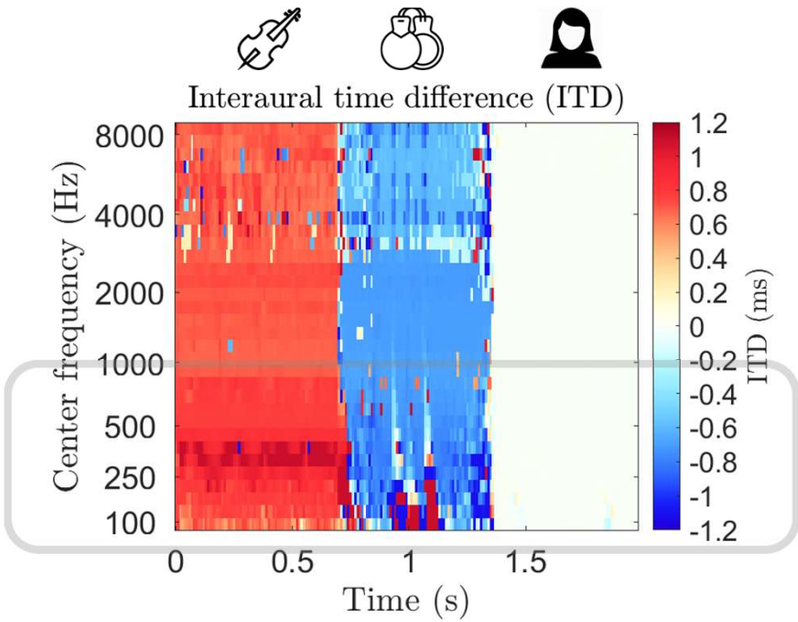
Cello
(50 cm, -70°)





Localization



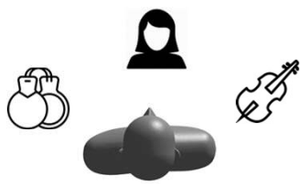


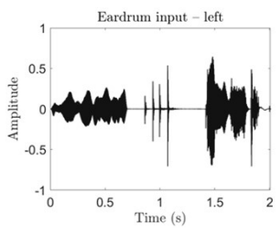
Interaural cues (ILD and ITD)

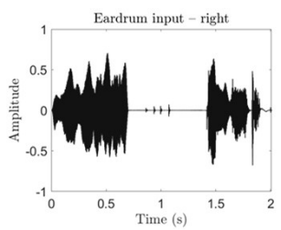
Superior olivary complex
(Cross-correlation)

Spectral cues (rate maps)

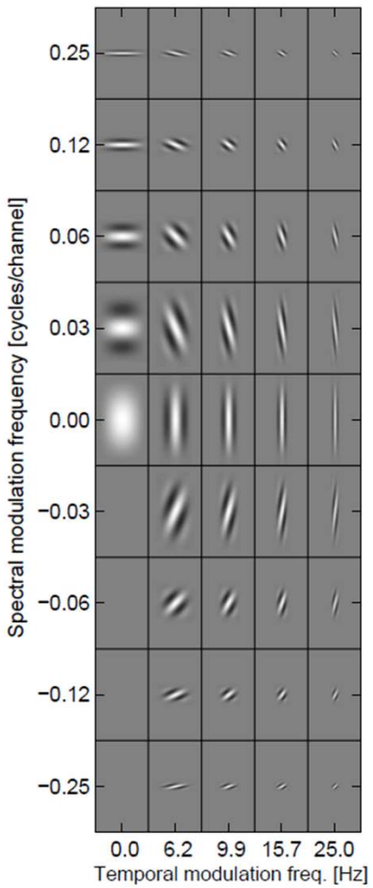
Cochlea
(gammatones, rect., integr.)



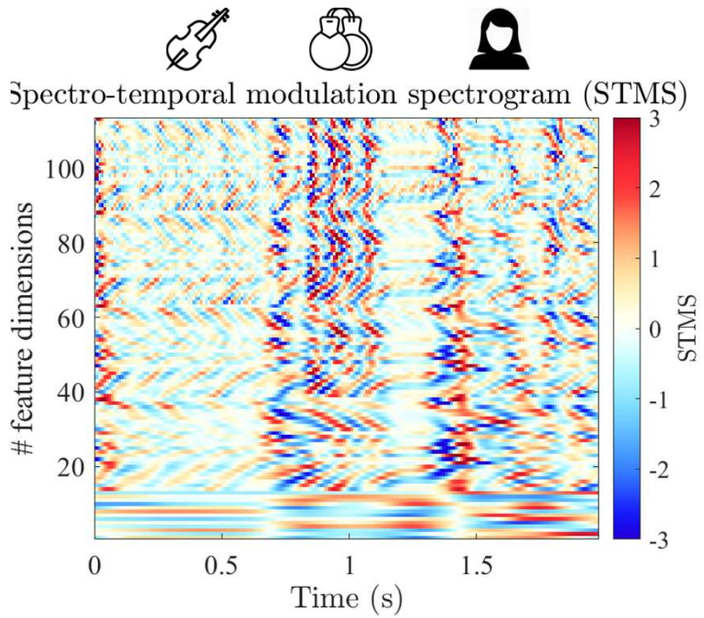


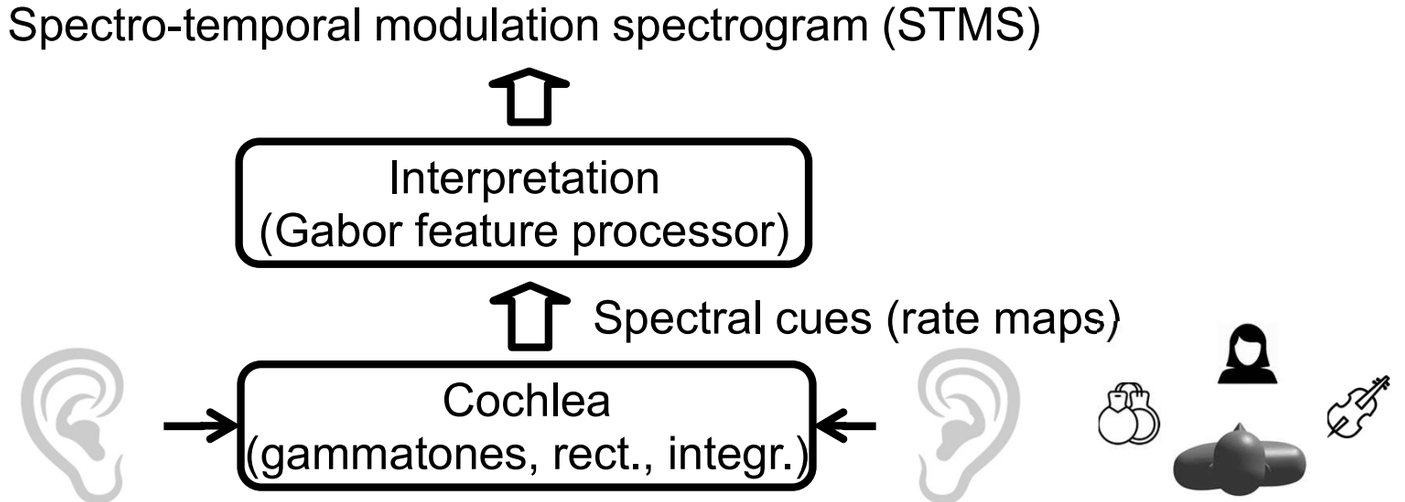


Identification

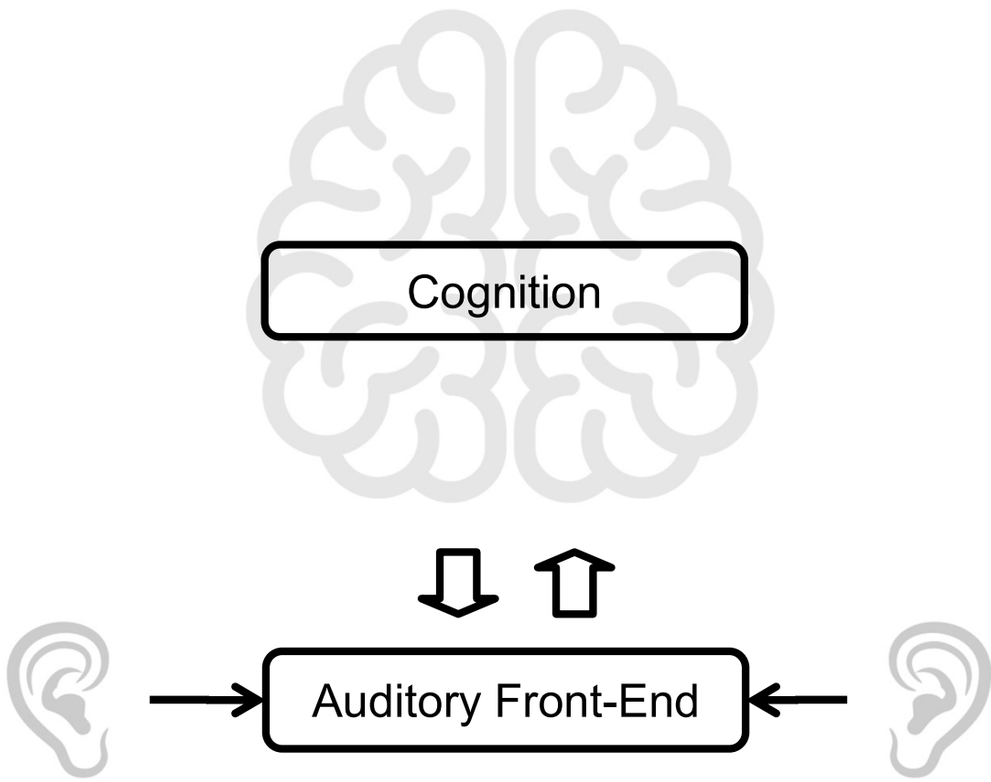


M. R. Schädler *et al.*, "Spectro-temporal modulation subspace-spanning filter bank features for robust automatic speech recognition," *J. Acoust. Soc. Am.*, vol. 131, no. 5, pp. 4134–4151, 2012.

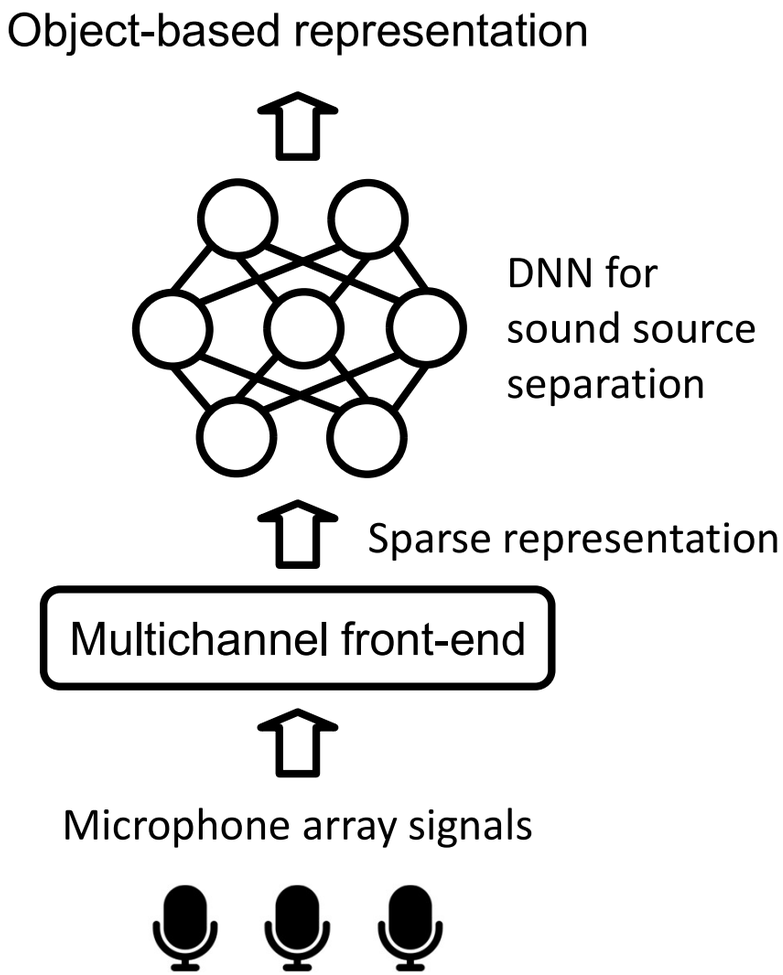




Neuroscience-inspired machine hearing



Neuroscience-inspired machine hearing



- DNN architectures
  - Fully-connected
  - Wide hidden layers
- DNN optimization metrics
  - Denoising: |x-y|_p, p = 1, 2
  - Classification: log(1+exp(-xy))

Thanks for your attention

César D. Salvador

https://cesardsalvador.github.io/

Perception Research

https://www.perception3d.com/