JAIST-SAST2015, 11th Nov. 2015

Speech Signal Processing based on Concept of Amplitude Modulation

Masashi UNOKI

School of Information Science

Japan Advanced Institute of Science and Technology

* Email: unoki@jaist.ac.jp

URL: http://www.jaist.ac.jp/~unoki/

Masashi Unoki, Ph.D

- Master/Ph.D @ JAIST (Prof. Akagi) (1994-1999)
- Visiting Researcher @ ATR, HIP (Prof. Irino) (1999-2000)

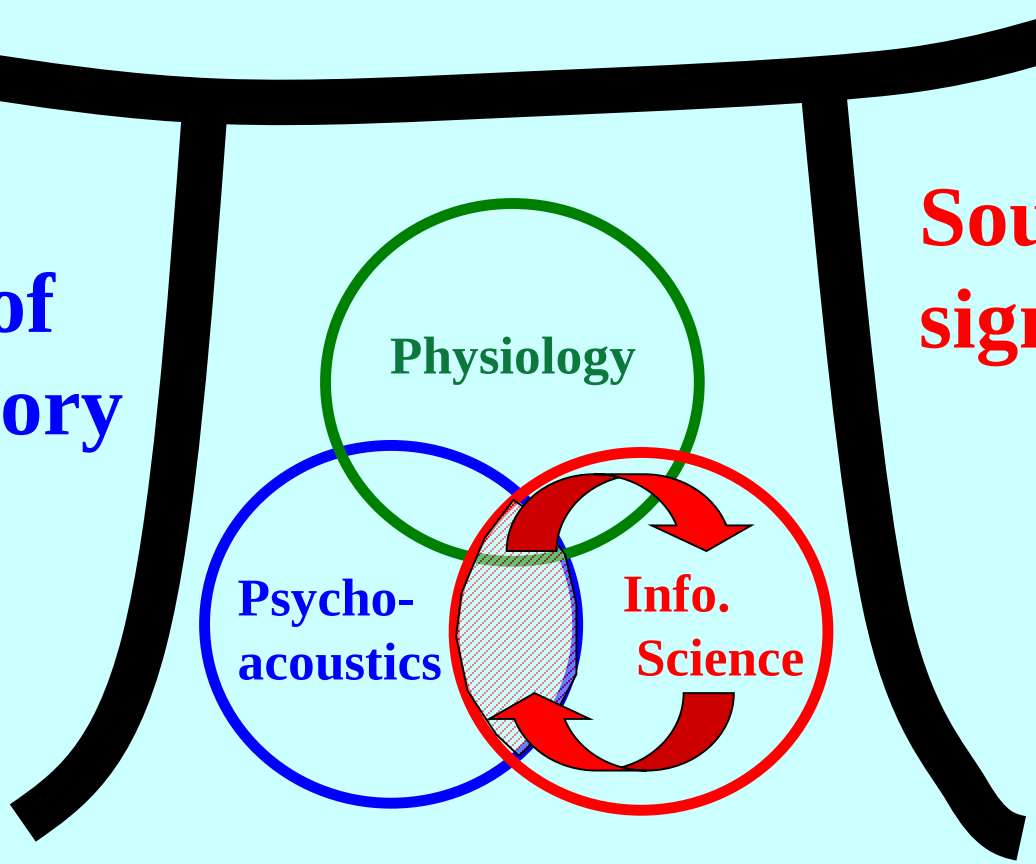
- Visiting Researcher @ Cambridge Univ. (2000-2001)
- Associate@JAIST (2001-2005)
- Associate Professor @JAIST (2005 -)
- Visiting Researcher @ Tech. Univ. Dresden (2010)

- **Associate Prof. Unoki, Assistant Prof. Miyauchi**
- **Seven PhD candidates, Five Master students**

Research stance (π-shaped approach)

Auditory-motivated signal processing

Investigation & modeling of human auditory system



Sound (speech) signal processing

Computational approach

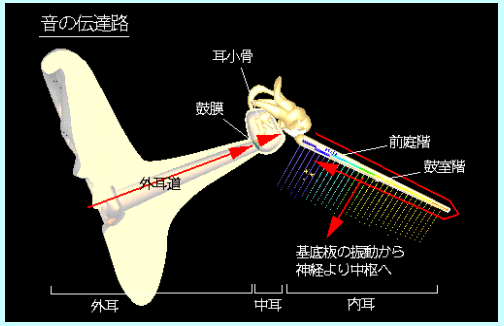
Projects

Auditory-motivated signal processing

Investigation & modeling of human auditory system

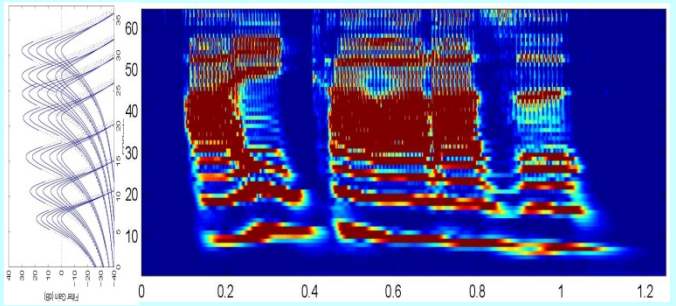
Sound (speech) signal processing

* Cochlear modeling



* Frequency selectivity

* Human auditor filterbank



* Speech enhancement

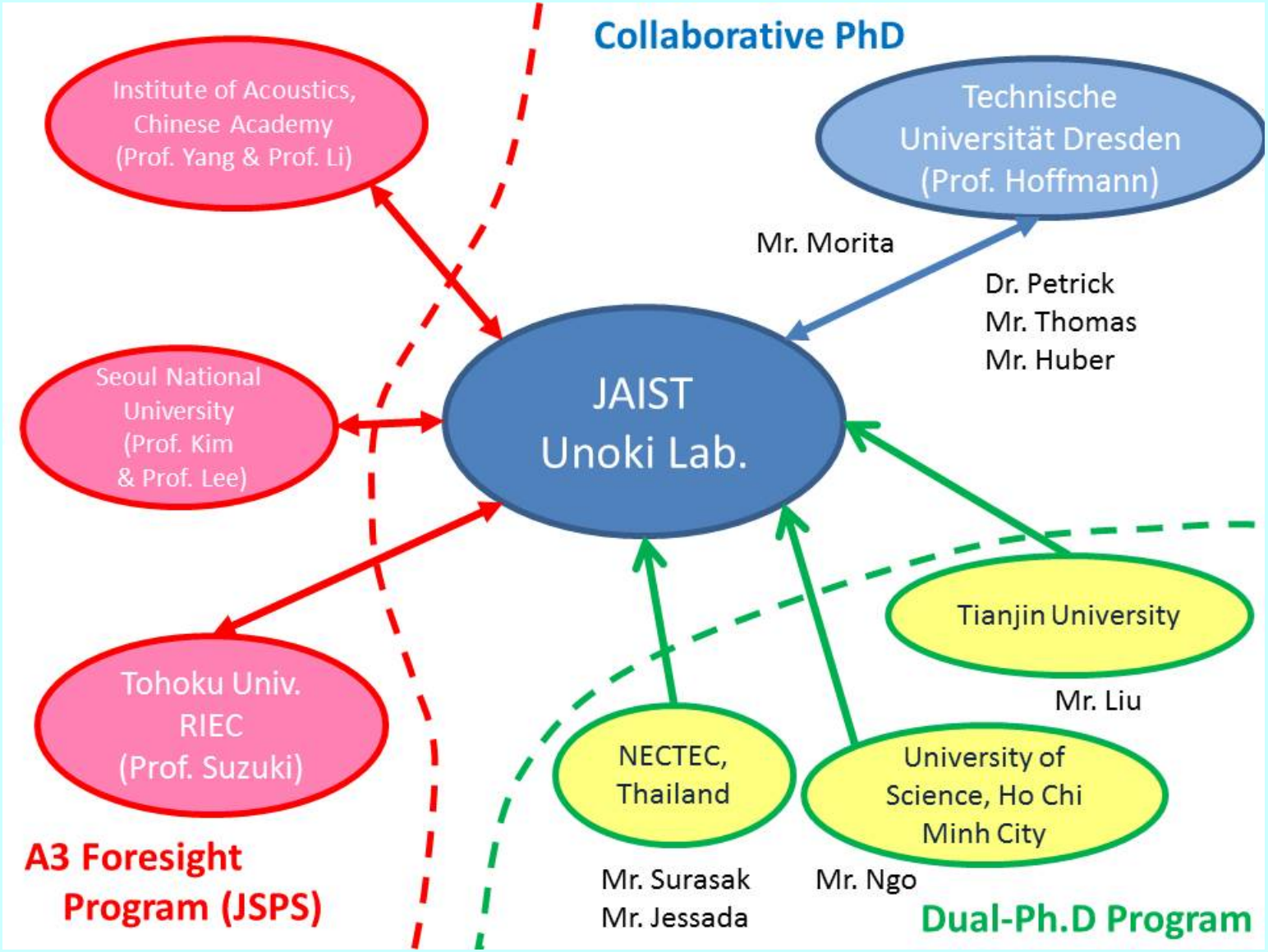
* Audio/Speech hiding

Research keywords

- Human auditory filterbank (cochlear filterbank)
- Auditory nonlinearities and attention
- **Selective listening and attention (ASA)**
- Auditory feedback

- **Acoustical Information Hiding (watermarking)**
- Denosing and dereverberation for speech perception
- Auditory motivated frontend for ASR systems
- Robust F0 estimation and VAD
- Blind estimation of the STI and reverberation time

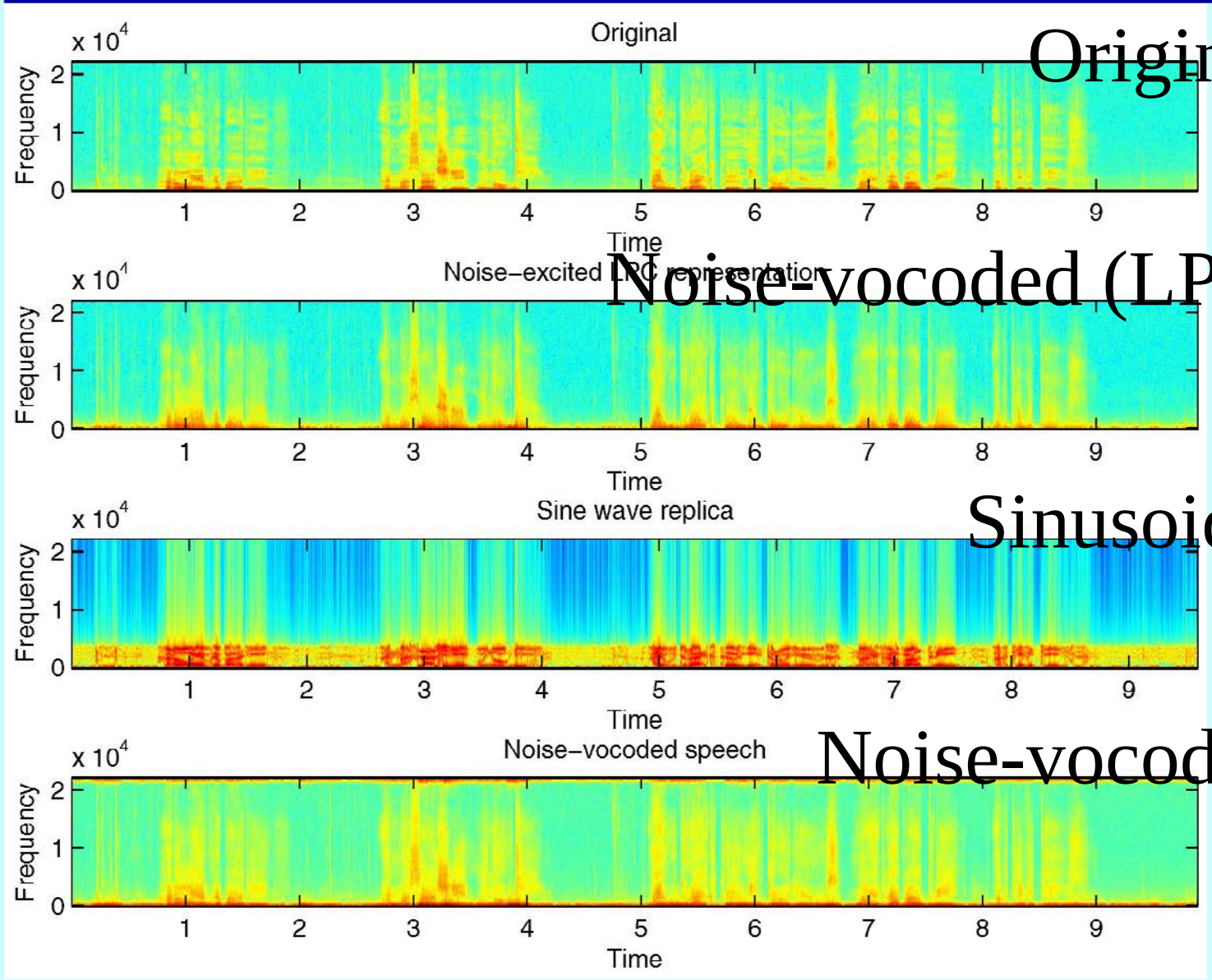
Collaboration networks



- **"Speech" has *redundancy* for representations.**

- **"Hearing" system has *robustness* for perception.**

Robust perception/Redundant info.



Original speech

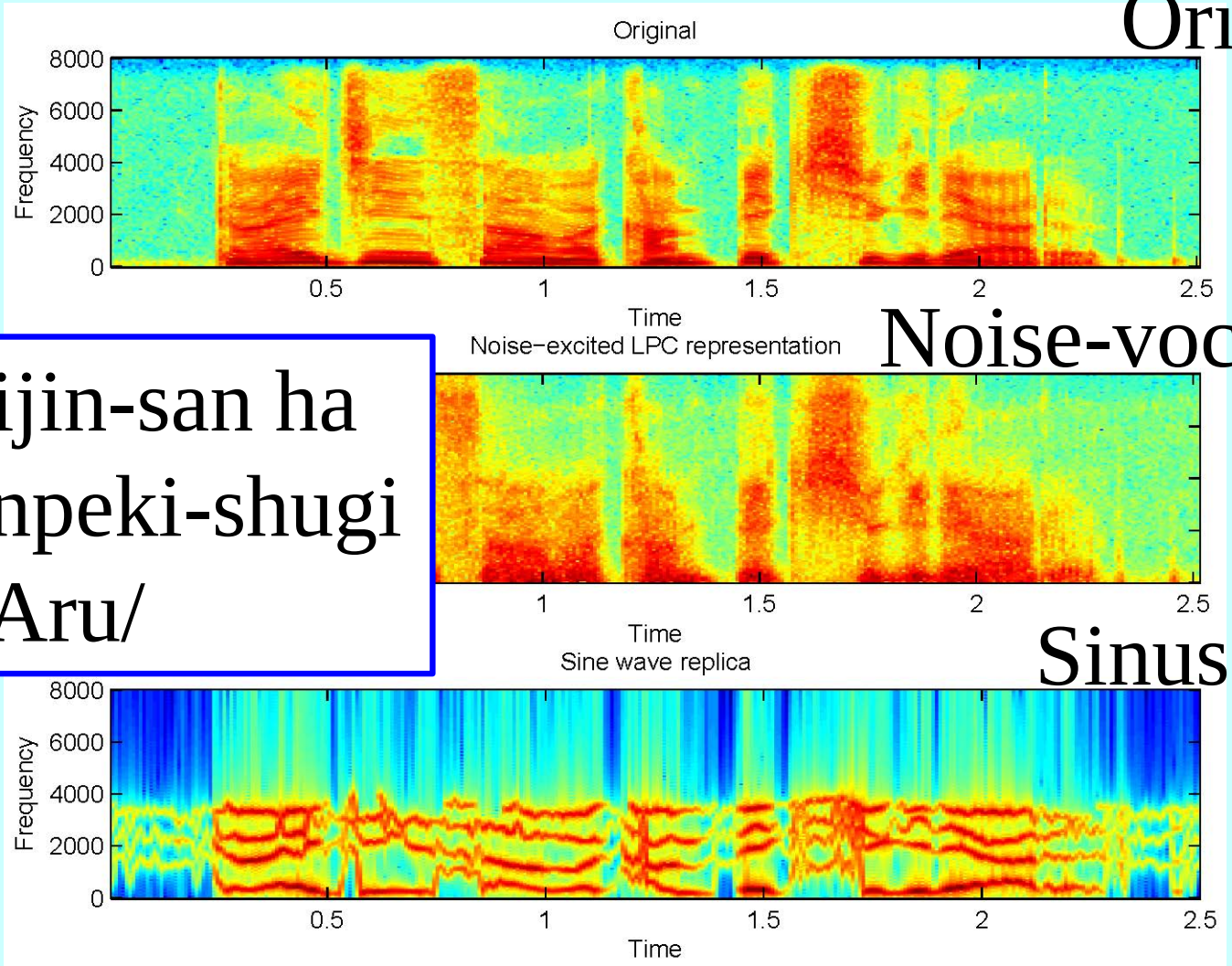
Noise-vocoded (LPC) speech

Sinusoidal speech

Noise-vocoded speech

Japanese

Original speech

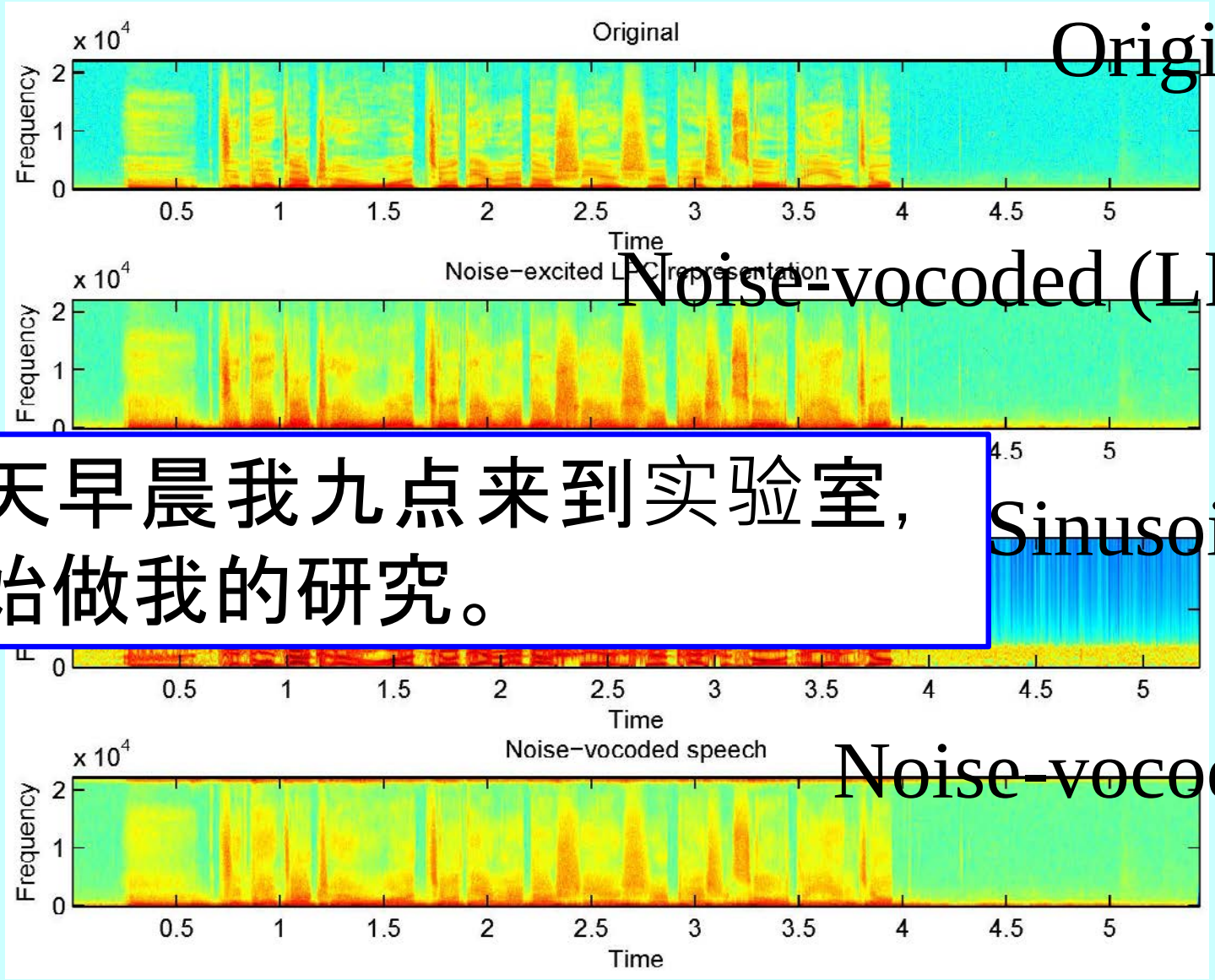


Noise-vocoded speech

Sinusoidal speech

/Gaijin-san ha Kanpeki-shugi de Aru/

Chinese



今天早晨我九点来到实验室，
开始做我的研究。

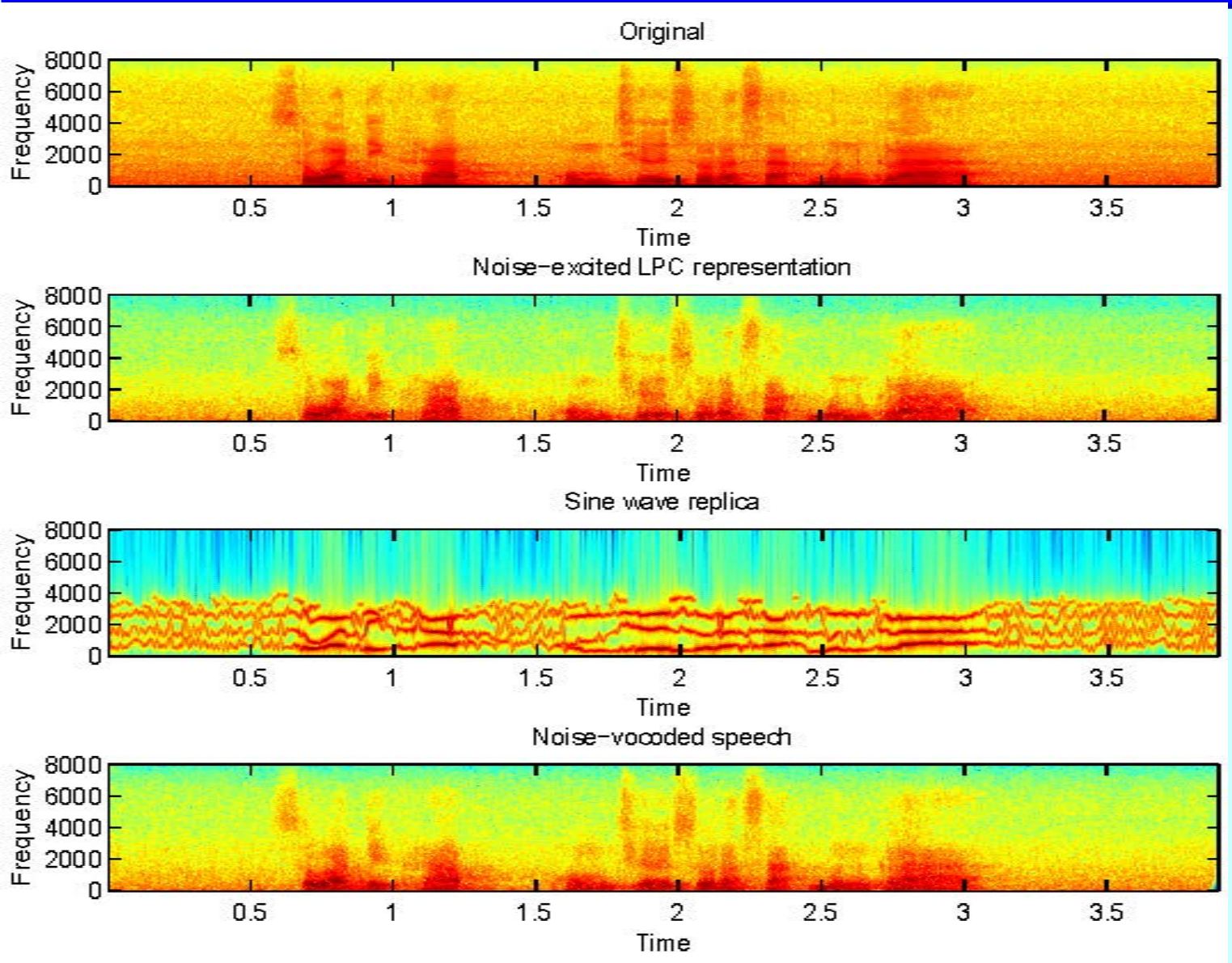
Original speech

Noise-vocoded (LPC) speech

Sinusoidal speech

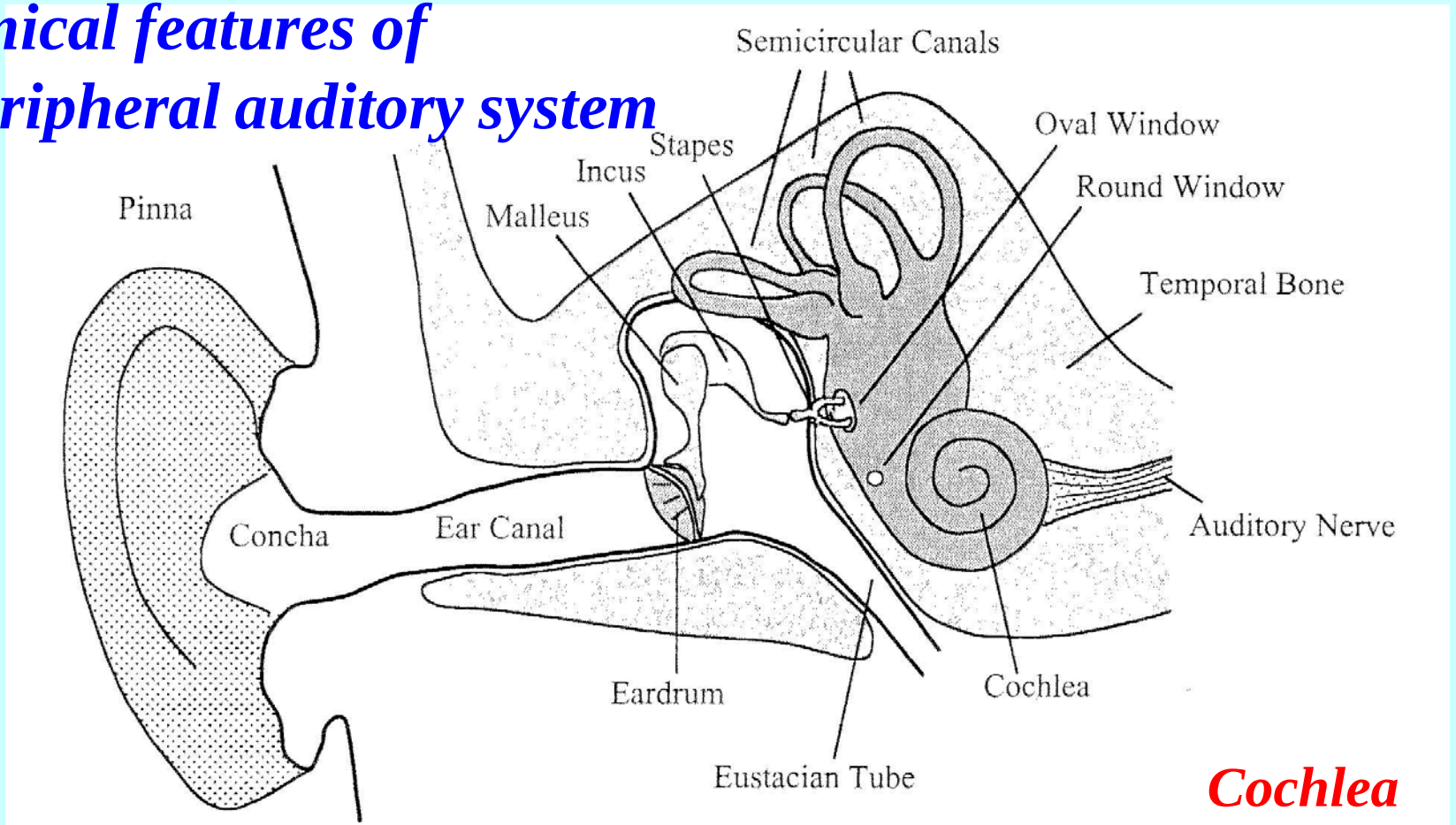
Noise-vocoded speech

Thai


Original
Frequency
8000
6000
4000
2000
0
0.5
1
1.5
2
2.5
3
3.5
Time
Noise-excited LPC representation
Frequency
8000
6000
4000
2000
0
0.5
1
1.5
2
2.5
3
3.5
Time
Sine wave replica
Frequency
8000
6000
4000
2000
0
0.5
1
1.5
2
2.5
3
3.5
Time
Noise-vocoded speech
Frequency
8000
6000
4000
2000
0
0.5
1
1.5
2
2.5
3
3.5
Time

From Air to Ear

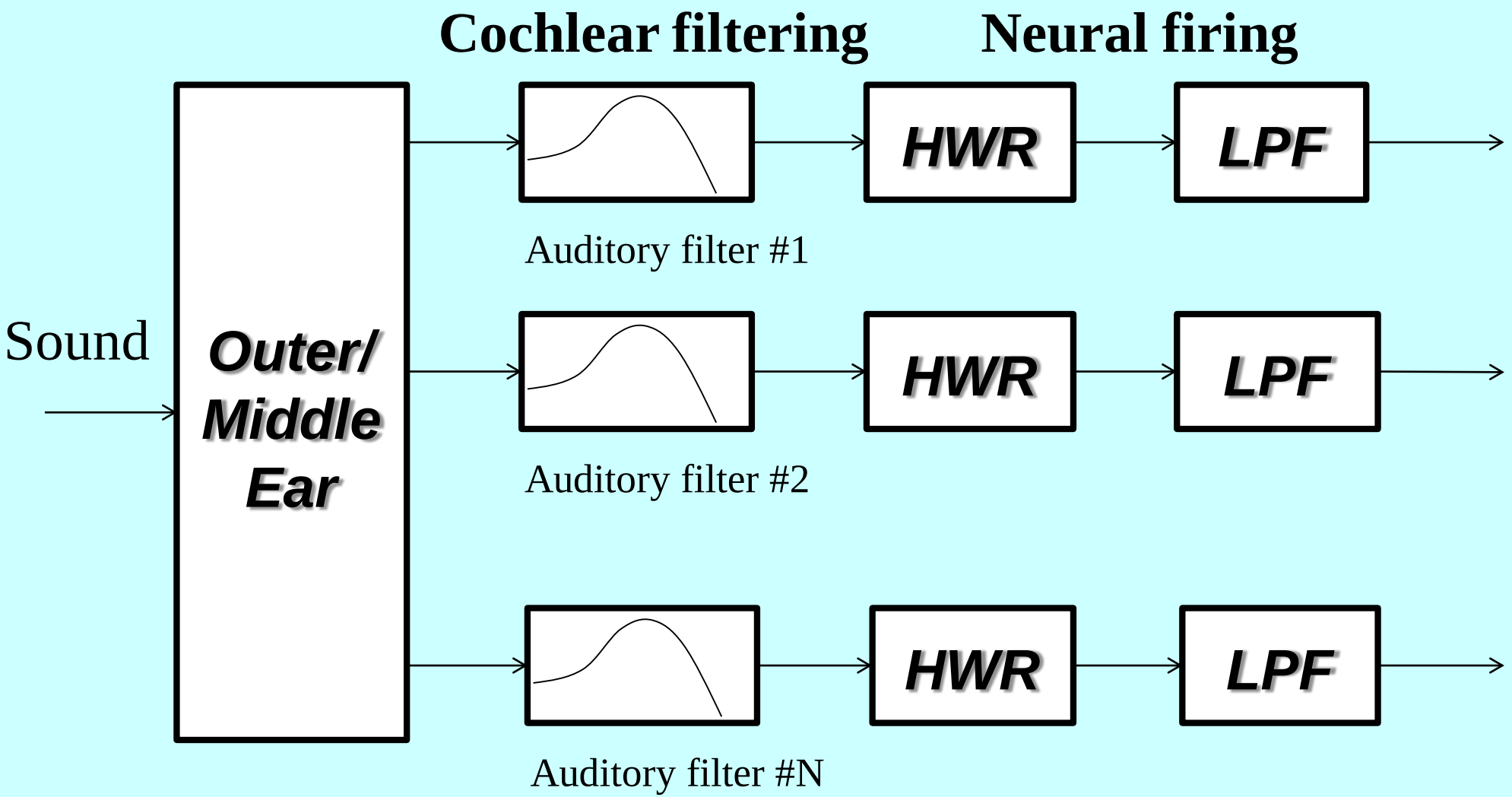
Anatomical features of the peripheral auditory system



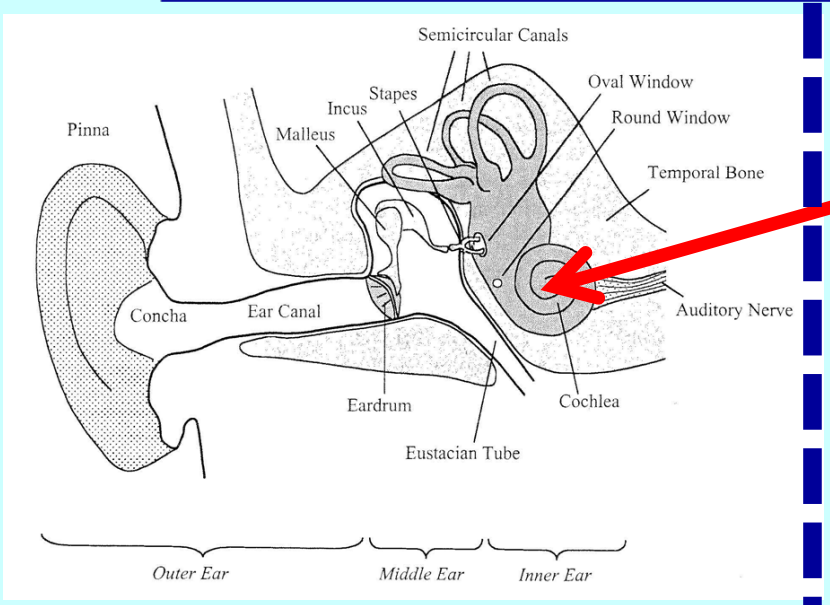
Cochlea

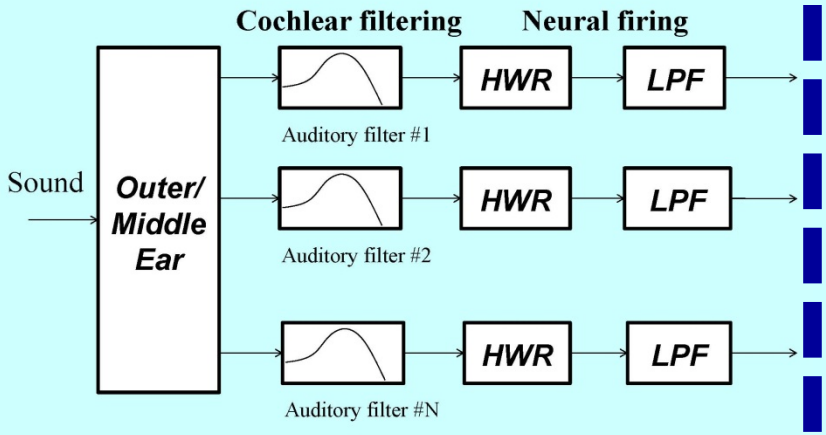
Outer ear → **Middle ear** → **Inner ear**

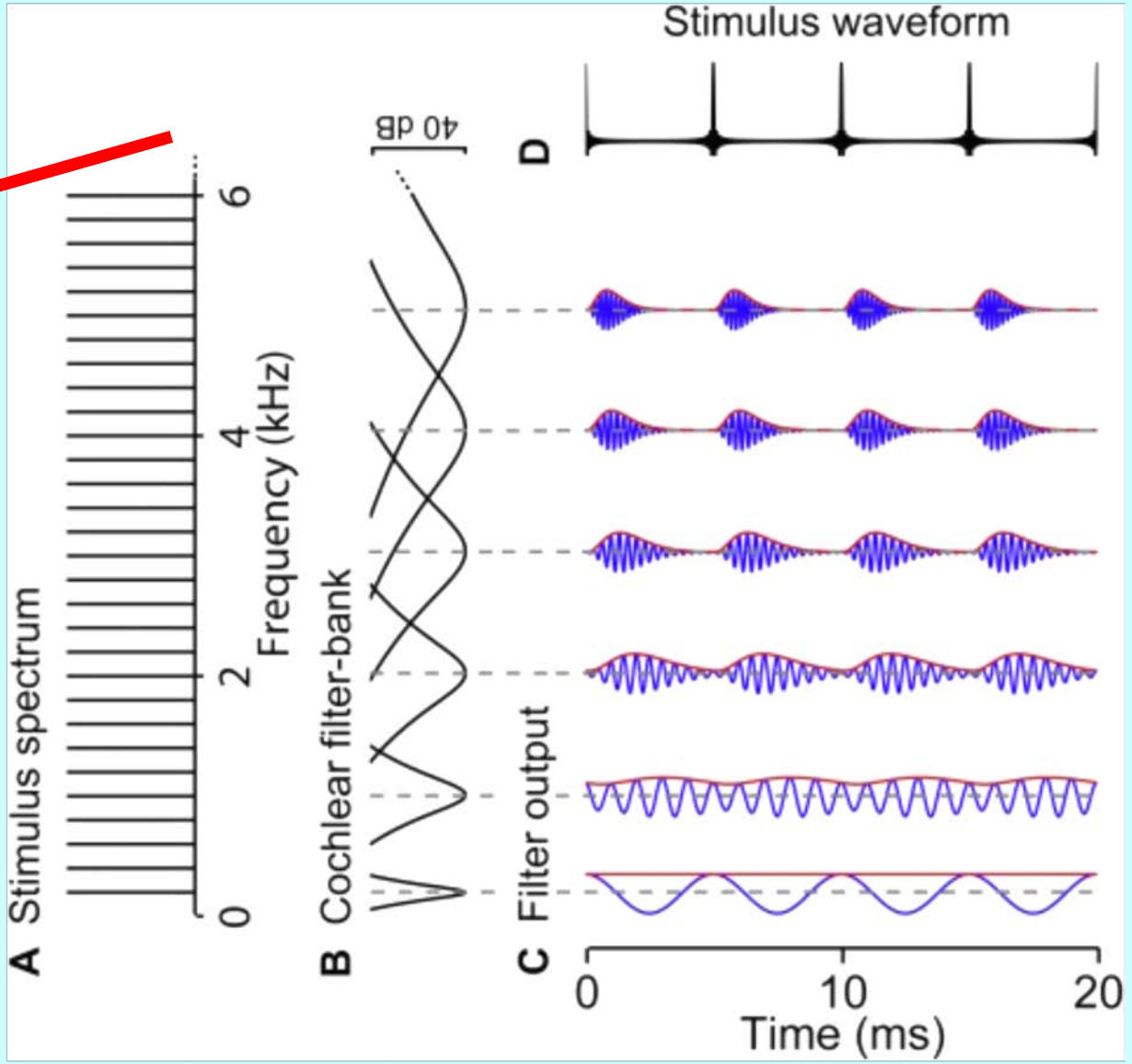
Scheme of Auditory Processing



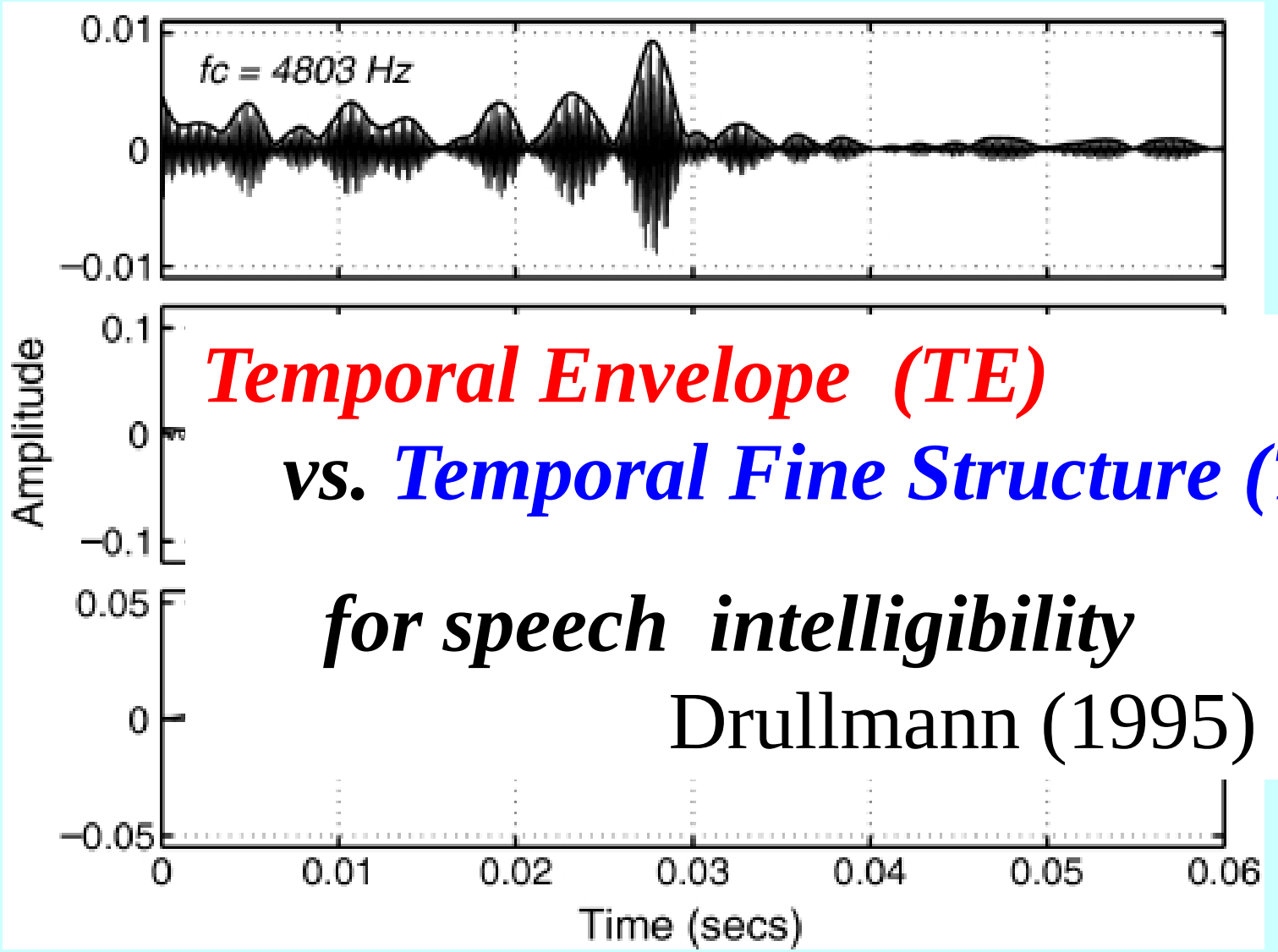
Speech Perception/Filterbank







Speech information (Linguistic info.)



Temporal Envelope (TE)

vs. Temporal Fine Structure (TFS)

for speech intelligibility

Drullmann (1995)

Current projects

- Concept of AM (amplitude modulation) can be used to account for speech perception.

- Current projects in my lab are as follows
  - **Speech enhancement based on AM concept**
  - **Estimation of room acoustics**
  - **Linguistic/non-linguistic processing for Cochlear implant systems**