

# Brain Intelligence and Artificial Intelligence

## 人脑智能与机器智能

### Lecture 3 – Auditory System

**Quanying Liu (刘泉影)**

SUSTech, BME department

Email: [liuqy@sustech.edu.cn](mailto.liuqy@sustech.edu.cn)

# Recap Lecture 2 – Visual System

0. Marr's 3 levels of explanation
1. Evolution of the eye
2. Function of the visual system
3. Structure of the eye
4. Photoreceptors
5. Information integration by ganglion cell
6. **Visual pathways:** photoreceptors, interneurons, ganglion cells, LGN, V1, ventral/dorsal pathways

# The Marr's three level of explanation

## Level 1: Computation theory

- What is the problem to be solved?
- What are the inputs and outputs to the computation?
- What is the **goal**, and what is the logic by which it is carried out?

## Level 2: Algorithmic

- How is the information **represented** and **processed** to achieve the computational goal?

## Level 3: Implementation

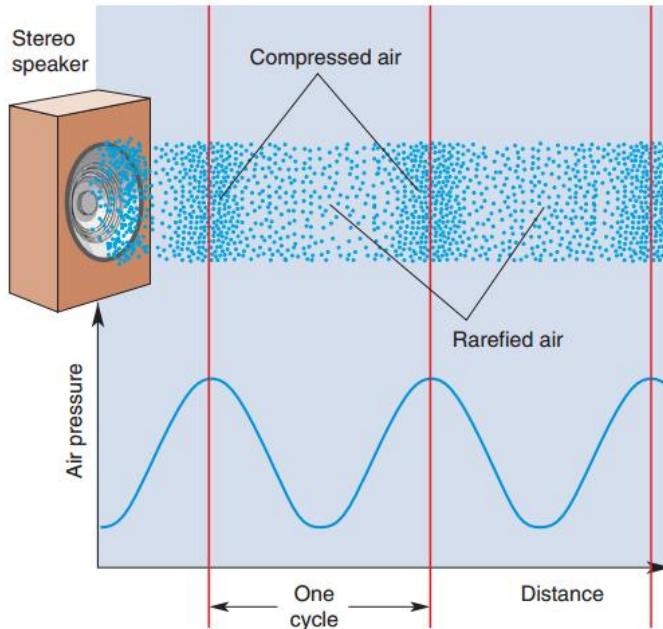
- How is the computation realized in physical or biological **hardware**?

# **Lecture 3 – Auditory System**

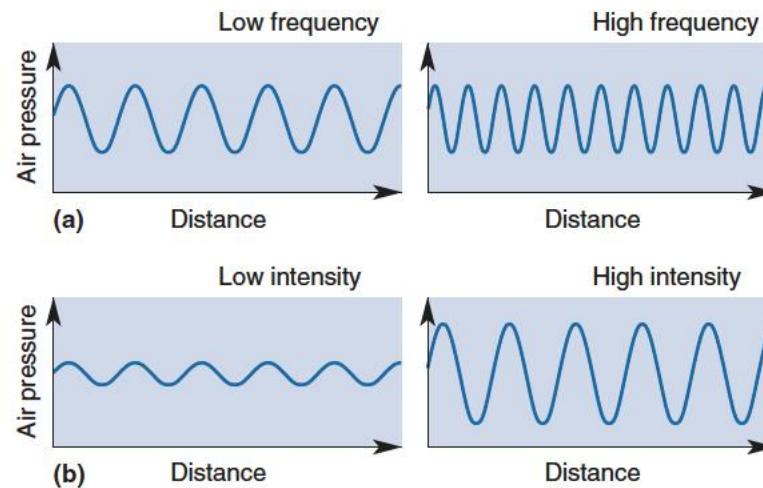
- **Sound:** frequency (Hz) & intensity (dB)
- **Ear**
  - Structure of human ear: outer ear, middle ear, inner ear
  - Cochlea; Organ of Corti; Hair cells
- **Central auditory process**
  - Auditory pathways: SG, (VCN, SO, IC, MGN), A1
  - Encoding sound intensity and frequency
  - Mechanisms of sound localization
  - Auditory cortex
- **Neural representation & plasticity in auditory cortex**
- **Deep learning for auditory processing:** encoding, decoding, interpreting

# Property of sounds

Sounds are audible variations in air pressure.



The **frequency** of sounds is the number of compressed patches of air that pass by our ears each seconds.



Pitch / tone (Hz)

Loudness / intensity (dB)

Real-world sounds **rarely** consist of simple rhythmic sound waves at **one** frequency and intensity.

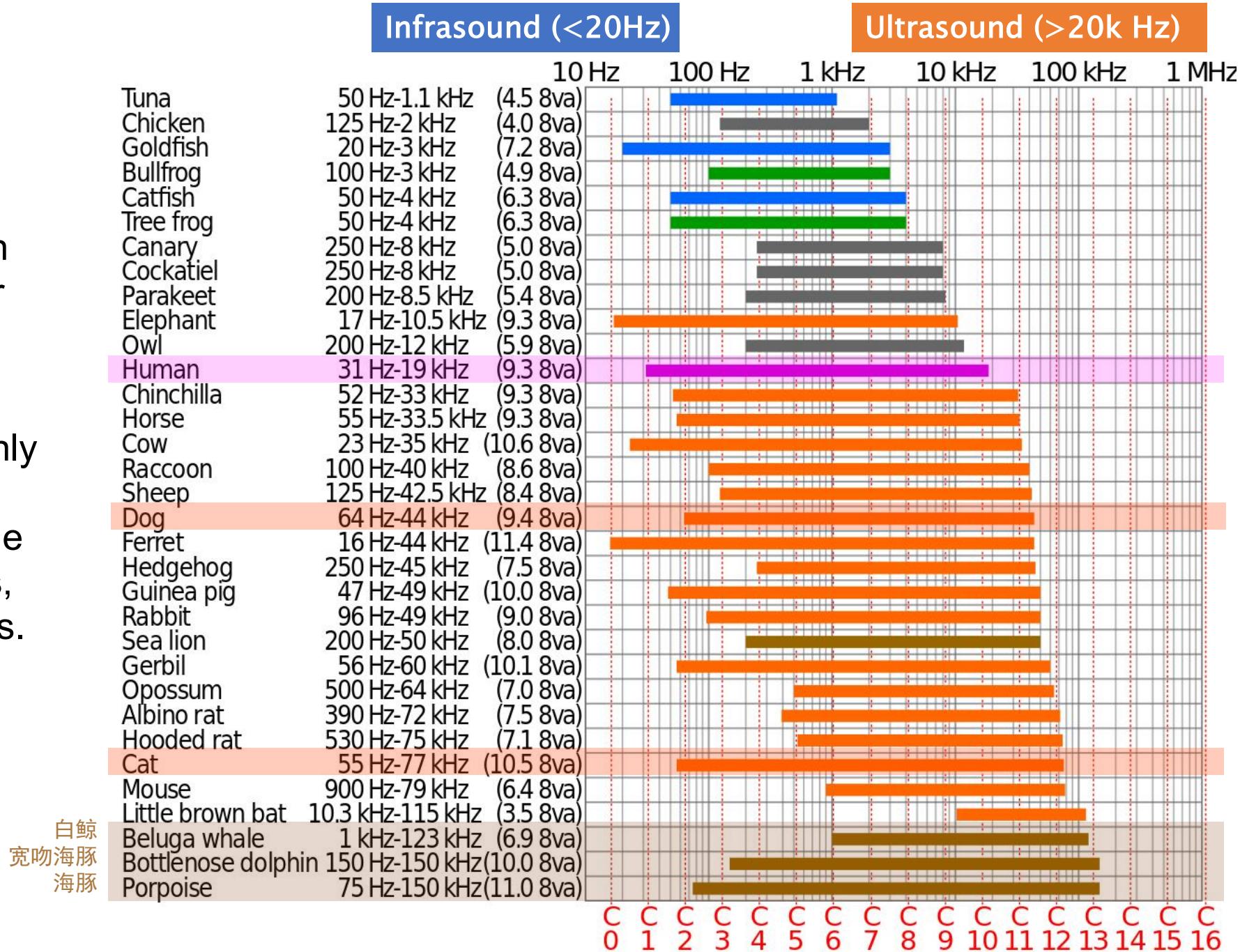
It is the simultaneous **combination** of different frequency waves at different intensities.

# Hearing range

Hearing range describes the range of frequencies that can be heard by humans or other animals, though it can also refer to the range of levels.

The human range is commonly given as **20 to 20K Hz**, although there is considerable variation between individuals, especially at high frequencies. We gradually **lose** the sensitivity to higher frequencies with age.

- **Infrasound (<20Hz)**
- **Ultrasound (>20k Hz)**



## Infrasound (<20 Hz)

- Even though we usually **cannot** hear low frequencies with our ears, they can have **unpleasant subconscious effects** (dizziness, nausea, and headache).
- For example, cars produce low-frequency sound at highway speeds makes sensitive people *carsick*.

## Ultrasound (>20K Hz)

- Ultrasound imaging
- Ultrasound scalpel
- Ultrasound neural stimulation

**A TEST – to hear the low-frequency sound from your body**

*Please carefully put your index finger in your ears.*

What is the mechanism of infrasound weapon? Have it already developed?

次声波武器的原理主要基于生物共振现象，具体原理如下：

## 1. 次声波的特性

次声波是指频率低于20Hz的声波，这种声波低于人耳可感知的范围，因此具有隐蔽性。同时，次声波具有传播距离远、穿透力强、稳定性好等特点。这些特性使得次声波能够在多种环境中有效传播，并穿透障碍物，对人体产生作用。

## 2. 生物共振现象

当人体接收到次声波时，由于次声波的频率与人体器官的固有频率相近，会产生共振现象。这种共振会导致人体器官变形、移位甚至破裂，从而达到杀伤目的。具体来说，次声波武器可以分为“神经型”和“器官型”两类，它们分别通过影响神经系统和器官系统来对人体造成伤害。

## 3. 神经型次声波武器

神经型次声波武器的频率接近人脑的阿尔法节律（8~12Hz），能够刺激大脑，引起共振。当人体大脑接收到此类声波时，神经会被麻痹。在功率较小时，会使人感到眩晕；而当功率较大时，会引起人类整个神经系统的紊乱，使人癫狂，从而使战斗人员丧失作战能力。

## 4. 器官型次声波武器

器官型次声波武器的振荡频率接近于人体某些特定器官的振荡频率。当人体接触到这类次声武器时，其对应器官会产生共振，并吸收一定能量。如果功率较大，就会损伤器官，甚至造成撕裂等严重后果。

## 5. 综合作用

次声波武器的作用不仅仅局限于单一器官或系统，而是能够全面地、直接地造成生物体各个部分组织的生物共振。一方面，次声波能够直接被生物体吸收声波的能量；另一方面，通过次声波对机体的直接刺激，器官会将这种生物共振的感觉传递到神经系统的相应部位，引起神经系统一系列的变化甚至毁坏。

综上所述，次声波武器的原理是利用次声波的特性和生物共振现象来对人体造成伤害。这种武器具有隐蔽性强、杀伤效果好等特点，但同时也面临着伦理和法律等方面的风险。因此，在研发和使用次声波武器时，必须严格遵守相关规范和准则，确保其仅用于正当的国防目的。



The mechanism of an **infrasound weapon** relies on generating and directing **low-frequency sound waves** (below 20 Hz) towards a target. These waves are below the human hearing threshold, but they have the potential to cause significant physiological and psychological effects. Here's an overview of how it works:

### Mechanism:

1. **Low-Frequency Sound Waves:** Infrasound refers to sound waves with frequencies below 20 Hz, which are typically inaudible to humans but can still cause vibrations in objects and living tissues.
2. **Penetration and Resonance:** Infrasound waves can travel long distances, penetrate buildings, and cause resonance in certain materials or organs in the human body. This resonance can lead to discomfort and various physiological effects, including nausea, dizziness, and disorientation. In some cases, exposure to intense infrasound can induce **fear**, **anxiety**, or **extreme stress**, likely because it disrupts normal brain function.
3. **Biological Impact:** Prolonged exposure to infrasound can cause vibrations in body tissues, particularly the chest and abdominal regions, which can lead to symptoms such as:
  - Nausea and vomiting
  - Difficulty breathing or chest tightness
  - Disorientation or loss of balance
  - Severe headaches
  - Feelings of panic or extreme fear (linked to the "fight or flight" response)
4. **Psychological Effects:** Infrasound has also been linked to psychological effects such as a sense of dread, anxiety, and even hallucinations. Researchers believe these effects are tied to the way low-frequency sound waves influence the vestibular system and internal organs.

What is the mechanism of infrasound weapon? Have it already developed?

次声波武器的原理主要基于生物共振现象，具体原理如下：

## 1. 次声波的特性

次声波是指频率低于20Hz的声波，这种声波低于人耳可感知的范围，因此具有隐蔽性。同时，次声波具有传播距离远、穿透力强、稳定性好等特点。这些特性使得次声波能够在多种环境中有效传播，并穿透障碍物，对人体产生作用。

## 2. 生物共振现象

当人体接收到次声波时，由于次声波的频率与人体器官的固有频率相近，会产生共振现象。这种共振会导致人体器官变形、移位甚至破裂，从而达到杀伤目的。具体来说，次声波武器可以分为“神经型”和“器官型”两类，它们分别通过影响神经系统和器官系统来对人体造成伤害。

## 3. 神经型次声波武器

神经型次声波武器的频率接近人脑的阿尔法节律（8~12Hz），能够刺激大脑，引起共振。当人体大脑接收到此类声波时，神经会被麻痹。在功率较小时，会使人感到眩晕；而当功率较大时，会引起人类整个神经系统的紊乱，使人癫狂，从而使战斗人员丧失作战能力。

## 4. 器官型次声波武器

器官型次声波武器的振荡频率接近于人体某些特定器官的振荡频率。当人体接触到这类次声武器时，其对应器官会产生共振，并吸收一定能量。如果功率较大，就会损伤器官，甚至造成撕裂等严重后果。

## 5. 综合作用

次声波武器的作用不仅仅局限于单一器官或系统，而是能够全面地、直接地造成生物体各个部分组织的生物共振。一方面，次声波能够直接被生物体吸收声波的能量；另一方面，通过次声波对机体的直接刺激，器官会将这种生物共振的感觉传递到神经系统的相应部位，引起神经系统一系列的变化甚至毁坏。

综上所述，次声波武器的原理是利用次声波的特性和生物共振现象来对人体造成伤害。这种武器具有隐蔽性强、杀伤效果好等特点，但同时也面临着伦理和法律等方面的风险。因此，在研发和使用次声波武器时，必须严格遵守相关规范和准则，确保其仅用于正当的国防目的。



The mechanism of an **infrasound weapon** relies on generating and directing **low-frequency sound waves** (below 20 Hz) towards a target. These waves are below the human hearing threshold, but they have the potential to cause significant physiological and psychological effects. Here's an overview of how it works:

### Has it been developed?

Infrasound weapons have been **theorized** and **experimented with** over the years, but their development and deployment are limited. Some reports suggest that certain countries have experimented with infrasound as a potential **non-lethal weapon** for crowd control or **psychological warfare**, but there are significant practical and ethical challenges.

1. **Experimental Uses:** There have been **research** and **small-scale tests** on using infrasound as a non-lethal means of controlling crowds, incapacitating opponents, or creating a deterrent effect in certain military or security scenarios. However, the weaponization of infrasound is complex due to:
  - Difficulty in controlling and focusing infrasound waves.
  - Unpredictability of the biological effects, as not everyone reacts the same way to infrasound.
  - Ethical and legal concerns regarding the use of such weapons on humans.
2. **Challenges:** Developing an effective infrasound weapon faces several hurdles:
  - **Control:** Infrasound is difficult to direct with precision, and it can affect unintended targets, including friendly forces or civilians.
  - **Power:** Generating infrasound at intensities high enough to cause significant effects over long distances requires a lot of power.
  - **Legal and Ethical Issues:** The use of infrasound weapons is controversial, as they can inflict harm on individuals without leaving visible injuries. Their potential psychological effects, such as inducing panic or fear, also raise ethical questions about their use in warfare or law enforcement.

在刘慈欣的科幻小说《三体》中，次声波氢弹是一种独特的太空核武器，其原理和设计充满了科幻色彩。以下是次声波氢弹的原理及可行性的分析：

## 原理

次声波氢弹的原理主要基于核爆炸与电磁辐射的相互作用，以及次声波在特定条件下的产生和传播。具体来说：

- 核爆炸与电磁辐射：**次声波氢弹在引爆时，能够以空气中次声波的频率连续发生多次核爆炸。每次核爆都会产生强烈的电磁辐射。
- 电磁辐射与金属外壳的相互作用：**这些电磁辐射与目标飞船的金属外壳相互作用，将电磁能量转化为飞船内部空气的声能。
- 次声波的产生：**在飞船内部，这种电磁能量的转化导致了超强次声波的产生。这些次声波具有足够的能量，能够杀死飞船内部的一切宏观生命体，但对飞船的设施基本没有损坏。

## 可行性分析

从科幻小说的角度来看，次声波氢弹的原理是富有想象力和创意的。然而，从现实科学的角度评估其可行性，则需要考虑多个因素：

- 核爆炸与电磁辐射的转换效率：**虽然核爆炸确实能产生强烈的电磁辐射，但将其高效转化为特定频率的次声波并在飞船内部产生足够的声压级，这在现实中可能是一个巨大的挑战。
- 金属外壳的电磁屏蔽效果：**飞船的金属外壳通常具有良好的电磁屏蔽性能，这可能会削弱电磁辐射对飞船内部的影响。
- 次声波在真空中的传播：**值得注意的是，太空中是真空的，声波（包括次声波）在真空中无法直接传播。但在《三体》的设定中，电磁辐射与金属外壳的相互作用产生了类似次声波的效果，这是一种科幻化的处理。
- 对人体的杀伤效果：**虽然次声波对人体有一定的伤害作用（如引起共振等），但要达到小说中描述的杀伤效果（杀死飞船内部的一切宏观生命体），可能需要极高的能量和精确的频率控制，这在现实中可能难以实现。

综上所述，次声波氢弹在《三体》中作为一种科幻设定是极具想象力和创意的。然而，从现实科学的角度来看，其可行性存在诸多不确定性和挑战。尽管如此，这种设定仍然激发了人们对未来科技发展的无限遐想和探索欲望。



在《三体》小说中，刘慈欣提到了“次声波氢弹”这一概念。这种武器设想基于次声波的独特性质，结合氢弹（热核武器）的威力，用于产生一种更具毁灭性和影响力的武器。不过，这更多的是科幻设想，基于目前已知的物理原理和技术，其可行性是存疑的。下面是对其原理的推测和分析。

## 次声波氢弹的假设原理

- 次声波与氢弹结合：**刘慈欣设想的次声波氢弹，可能是基于氢弹的巨大能量释放，通过某种机制将爆炸产生的能量以低频（次声波）的形式传播。氢弹爆炸产生的冲击波本来就包含不同频率的能量传递，其中也可能包括次声波成分。次声波因其波长较长、衰减较慢，理论上可以传播得更远，影响更大区域的目标。
- 次声波的破坏性：**次声波因为其低频率的特性，可以引起建筑物、地形结构的共振，甚至引发人体内部器官的共振。这种共振可能会造成生理损伤，如恶心、内脏器官损伤等。将次声波与氢弹的巨大冲击力结合，理论上可能会形成一种新型的破坏力，既有高能爆炸的直接物理破坏，也有次声波带来的远距离震荡效果。
- 次声波的传播范围与效果：**氢弹爆炸产生的次声波在传播过程中，其穿透性和传播范围远比普通爆炸波更大。次声波能够在空气或固体介质中传播很远，甚至在地球上产生全球性的影响。这可能是刘慈欣在《三体》中构想该武器的关键逻辑之一。

## 可行性分析

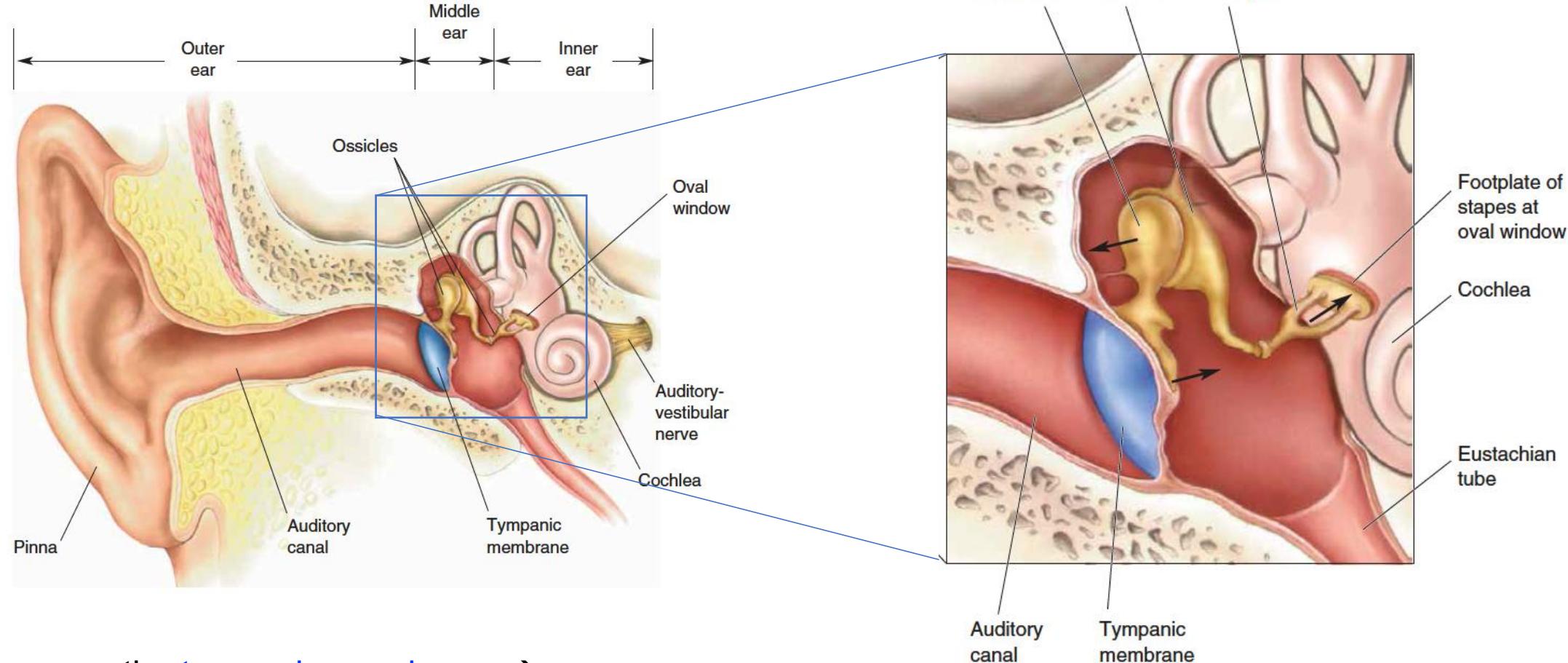
尽管在《三体》中次声波氢弹被描述得非常强大，但从现实科学角度来看，这种武器的可行性存在巨大问题。

- 次声波的威力与控制问题：**目前，虽然次声波确实能够产生某些生理或心理影响，但其能量不足以和氢弹爆炸的直接破坏力相比拟。氢弹爆炸所释放的能量主要以高温、高压和辐射的形式迅速释放，如何将这些能量有效转换为次声波且能够精确控制这种转换，仍是一个巨大的技术难题。
- 能量转换效率低：**氢弹爆炸是瞬间发生的剧烈反应，而次声波则需要较长的波长和较低的频率。要将氢弹爆炸的巨大能量集中并转化为次声波传播，理论上需要非常高的能量转换效率，这在目前的物理和工程技术中尚未实现。
- 次声波的物理限制：**虽然次声波能传播较远，但其实际破坏能力在远距离会迅速下降。即便氢弹爆炸能够产生次声波，如何确保次声波在传播过程中仍然保持足够的破坏力并影响到远距离目标是一个很大的挑战。
- 次声波与氢弹的结合：**从目前的技术来看，次声波的生成和传播与氢弹的高温高压爆炸机制之间存在很大差异，如何将两者有效结合并利用，是一个理论上尚未解决的问题。

## 结论

次声波氢弹在《三体》中是一个引人入胜的科幻设想，具有极大的想象空间，但从现有的物理学和工程学角度来看，其现实可行性很低。虽然次声波确实能对生理、心理产生一定影响，但要将其与氢弹的爆炸威力结合，并产生大范围的次声破坏，目前仍然没有实现的技术路径。

# Structure of human ear



Sound wave moves the **tympanic membrane** →

Tympanic membrane moves the **ossicles** →

Ossicles move the **membrane at the oval window** →

Motion at the oval window moves **fluid in the cochlea** →

Movement of fluid in the cochlea causes a **response in sensory neurons**

# The attenuation reflex

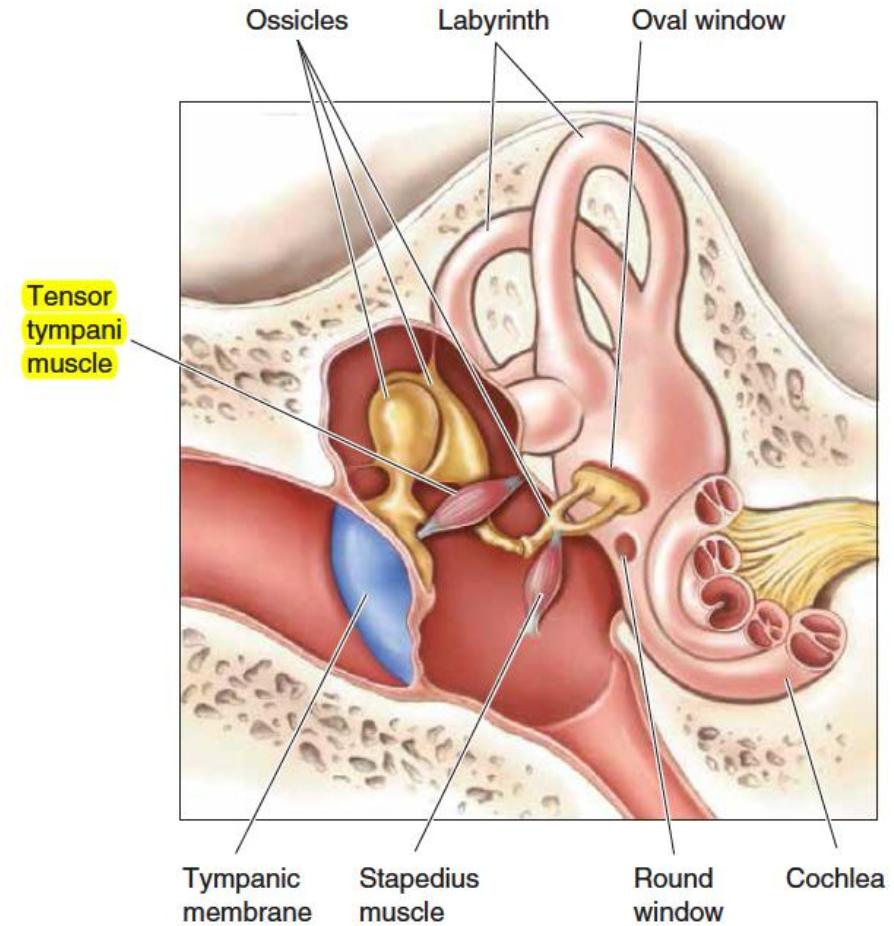
Question: Our own voice is transmitted by the bones or tissues to the ear, which generates much **stronger** vibrations in tympanic membrane.

Why doesn't it impair our hearing?

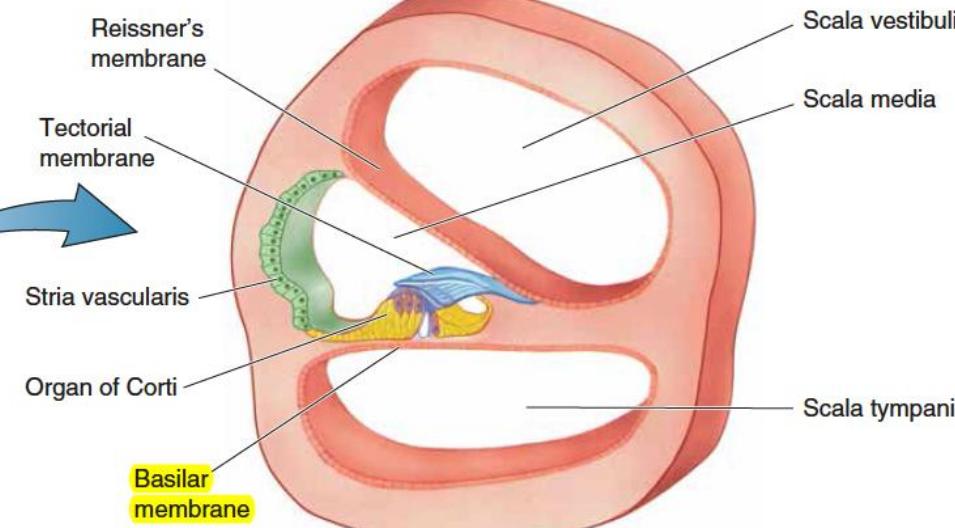
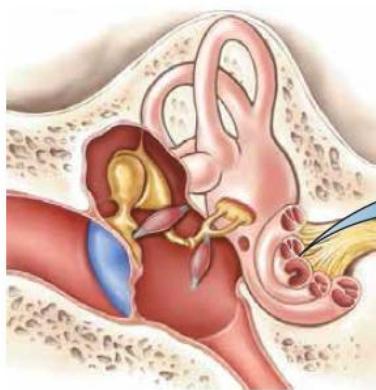
**Two muscles:** *tensor tympani muscle* and *stapedius muscle*

The onset of a loud sound triggers a neural response that causes these muscle to **contract**, a response called the **attenuation reflex**.

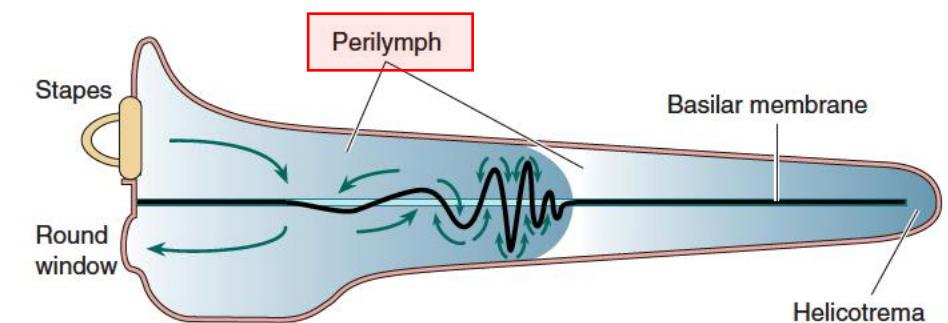
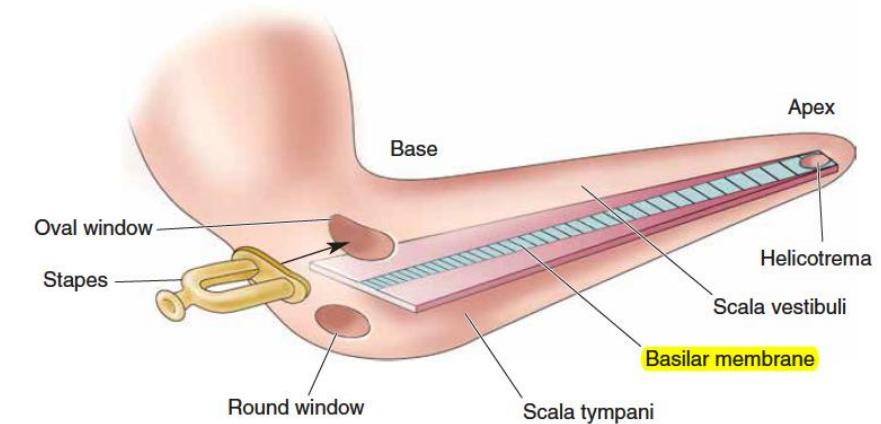
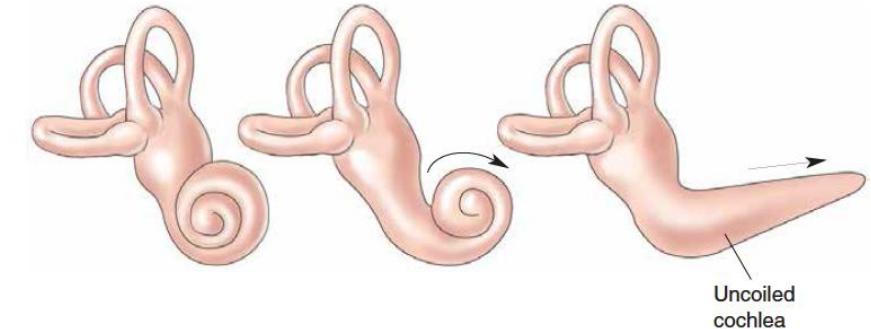
Muscle contraction → Ossicles gets more rigid



# The cochlea



## The basilar membrane in an uncoiled cochlea



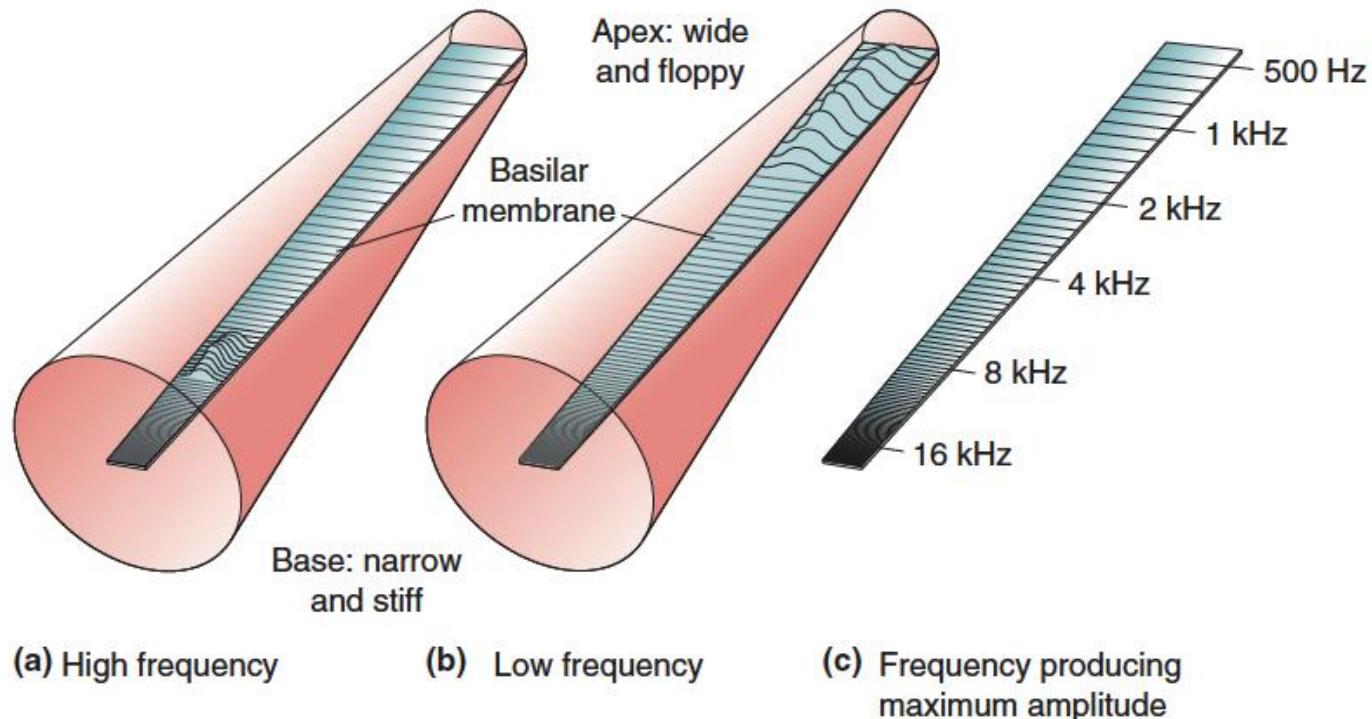
**Perilymph:** the fluid in the scala vestibuli and scala tympani

- relatively low K<sup>+</sup> (7 mM) and high Na<sup>+</sup> (140 mM), similar to CSF

**Endolymph:** the fluid in the scala media

- high K (150 mM) and low Na (1 mM), similar to intracellular fluid

# The response of the basilar membrane to sound → Place code



## Tonotopy

Systematic organization of sound frequency within an auditory structure is called **tonotopy**, analogous to *retinotopy* in the visual system.

*Tonotopic maps* exist on the basilar membrane and within each of the auditory relay nuclei, the MGN, and auditory cortex.

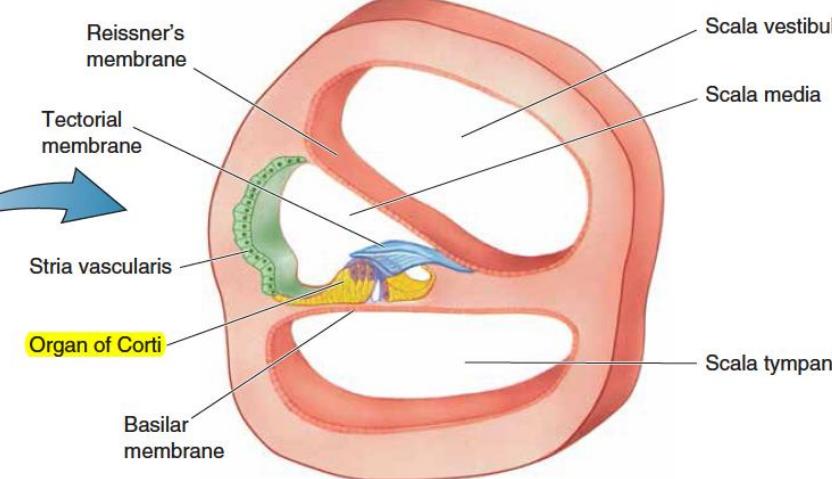
Everything we have discussed to this point involves the **mechanical** transformations of sound energy that occur in the middle and inner ear.

Now, we come to the point in the system where **electrical** signals are first involved.

The *auditory receptor cells*, which convert mechanical energy into a change in membrane polarization, are located in the *organ of Corti*.

# Organ of Corti

皮质器官



## The basilar membrane

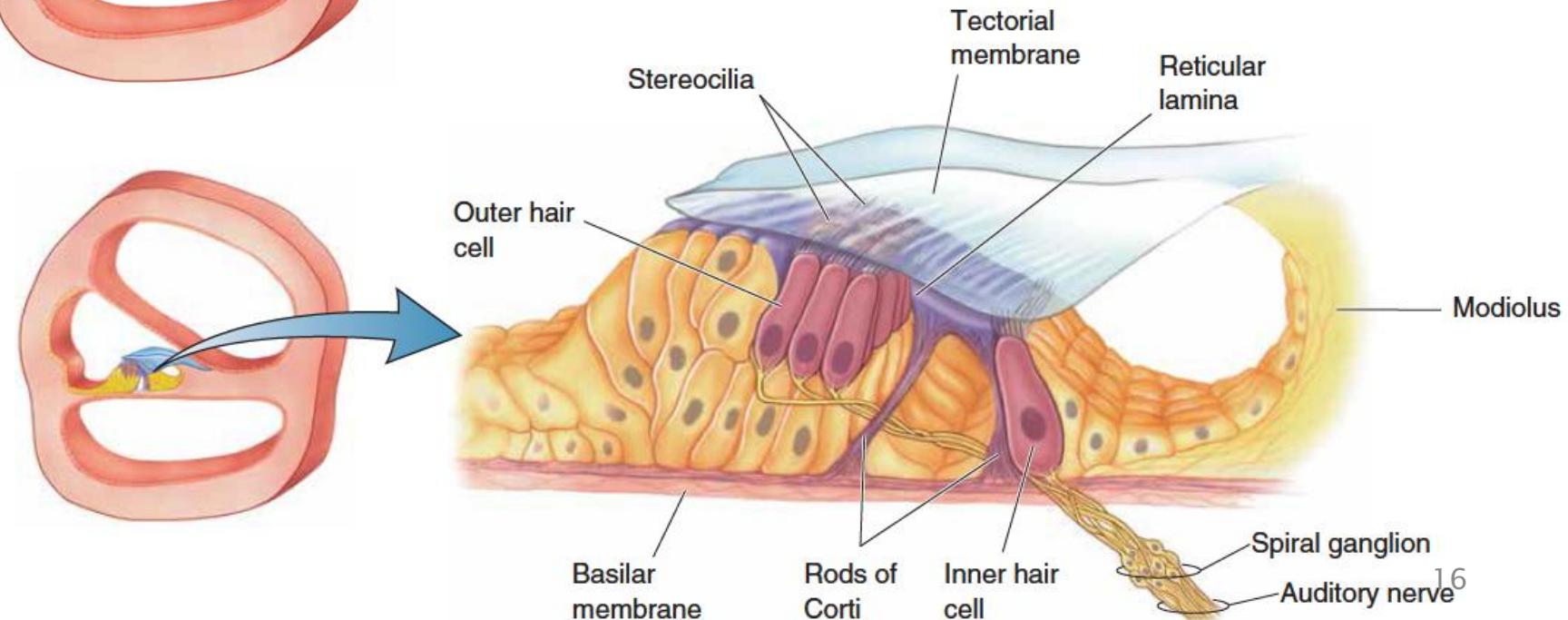
supports tissue that includes the inner and outer hair cells and the stiff rods of Corti.

The tectorial membrane extends from the bony modiolus to cover the stereocilia that protrude from the tops of the hair cells.

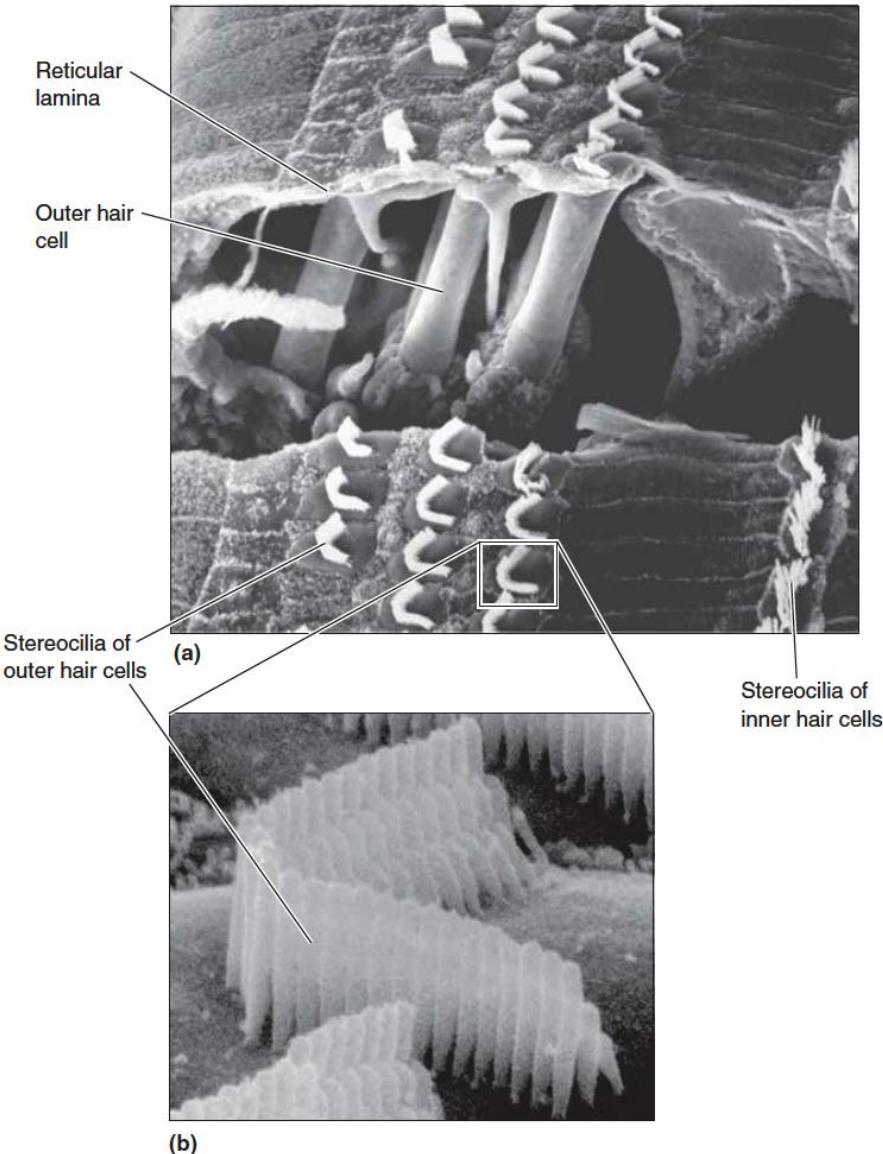
The *auditory receptor cells*, which convert mechanical energy into a change in membrane polarization, are located in the *organ of Corti*.

The *organ of Corti* consists of

- **hair cells** (auditory receptor cell)
- rods of Corti
- various supporting cells



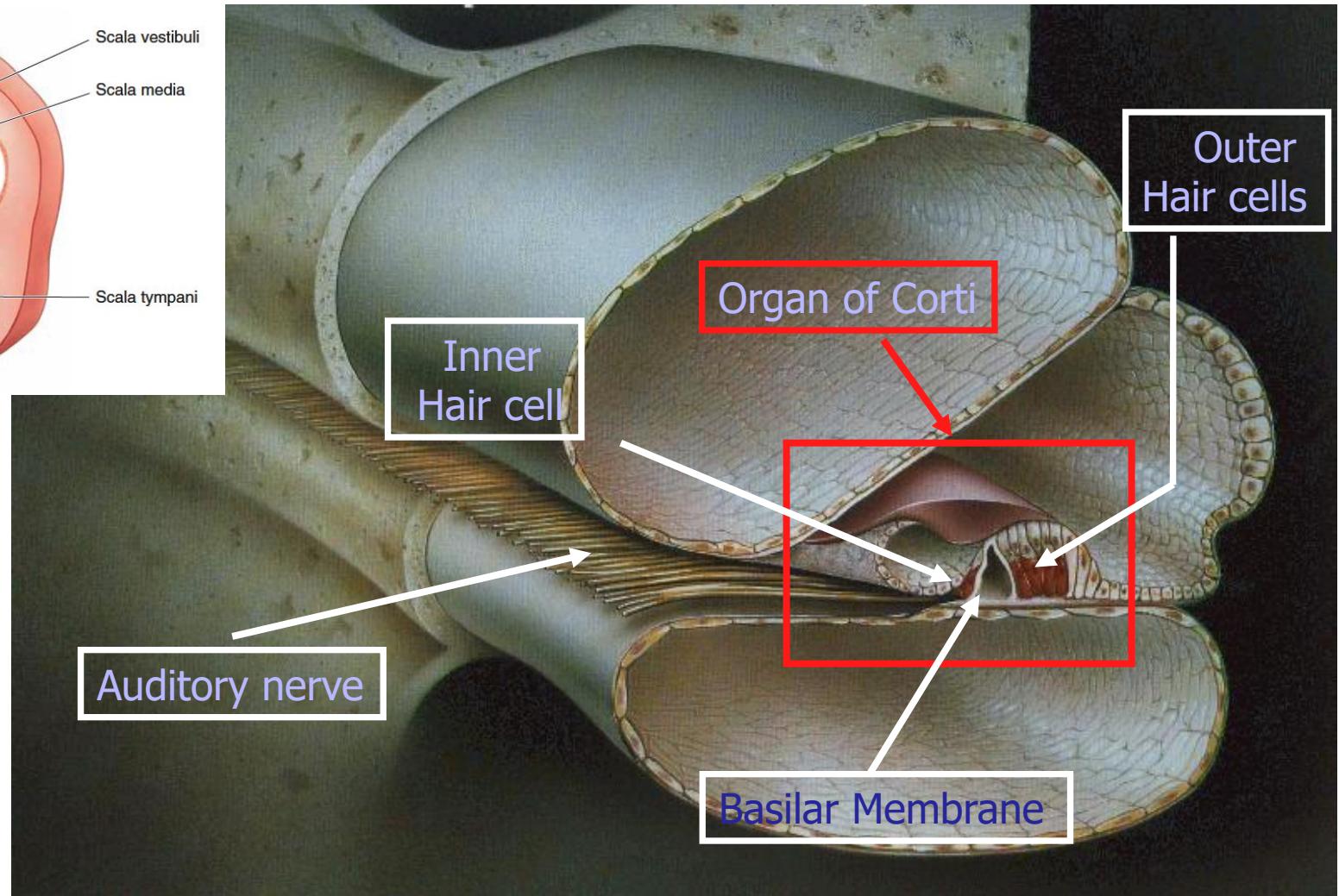
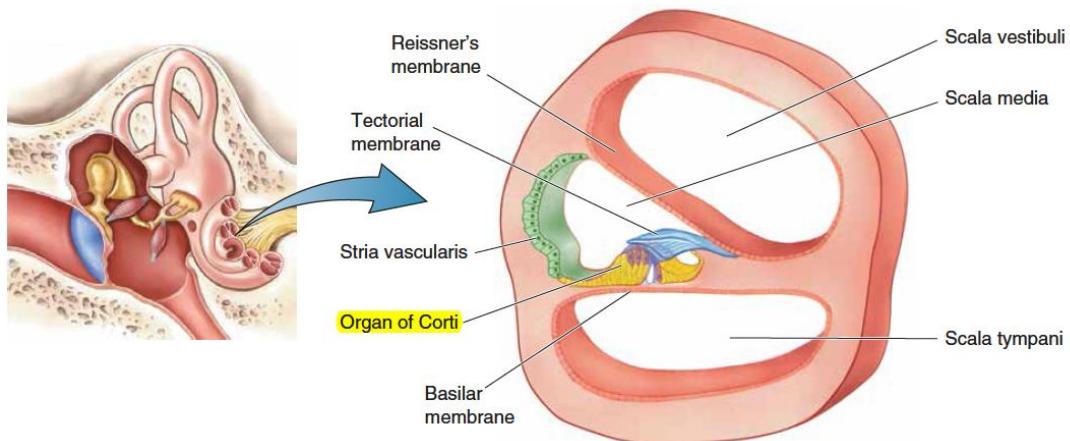
# Hair cells and their stereocilia



**Hair cells viewed through the scanning electron microscope.** (a) Hair cells and their stereocilia. (b) A higher resolution view of the stereocilia on an outer hair cell. The stereocilia are approximately  $5\text{ }\mu\text{m}$  in length. (Source: Courtesy of I. Hunter-Duvar and R. Harrison, The Hospital for Sick Children, Toronto, Ontario, Canada.)

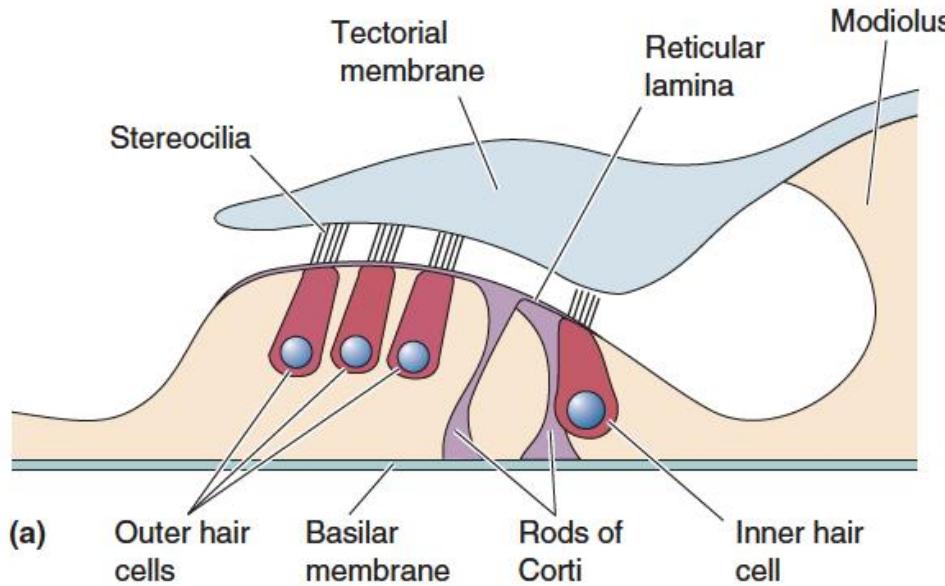
- Each hair cell has 10–300 hairy-looking *stereocilia*.
- Hair cells are **not** neurons. They lack axons, and in mammals, they do **not** generate *action potentials*.
- Outer hair cells outnumber inner hair cells.

# Hair cells and their stereocilia

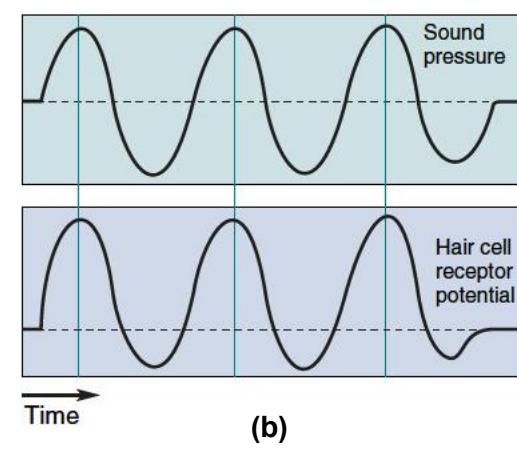
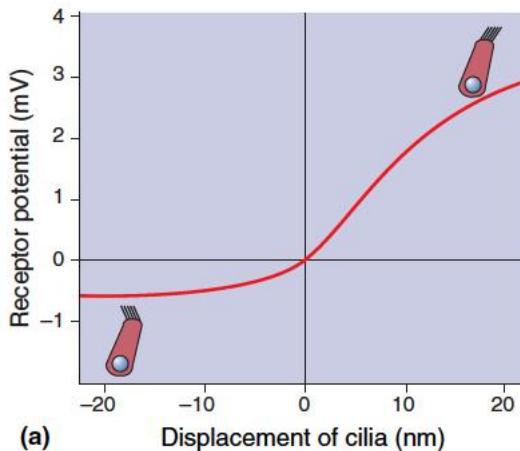
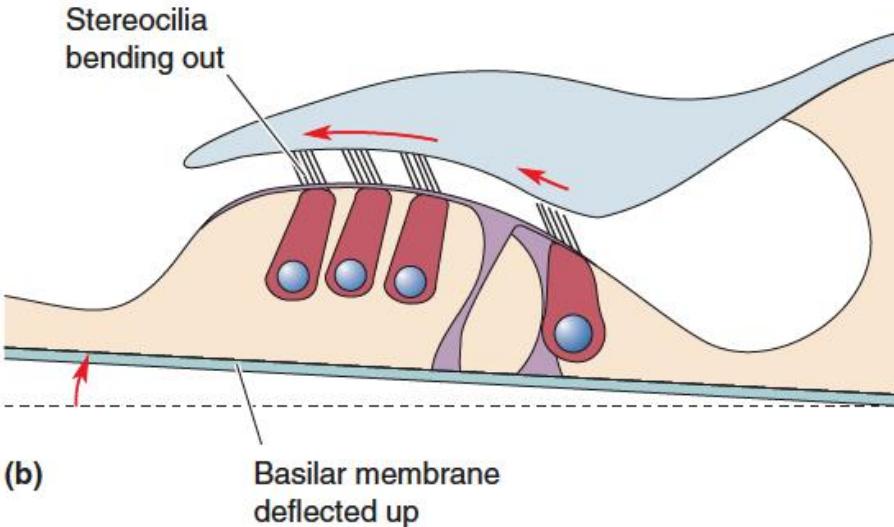


# The bending of stereocilia by the basilar membrane

At rest



When sound causes the basilar membrane to deflect upward



Hair cell receptor potentials.

- (a) The hair cell depolarizes or hyperpolarizes, depending on the **direction** in which the stereocilia bend.
- (b) The hair cell receptor potential closely follows the air pressure changes during a low-frequency sound.

# Cochlear implants

Hair cell damage or death is the most common cause of deafness.

In most of the cases, the *auditory nerve* remains intact

→ Cochlear implants to bypass hair cells.

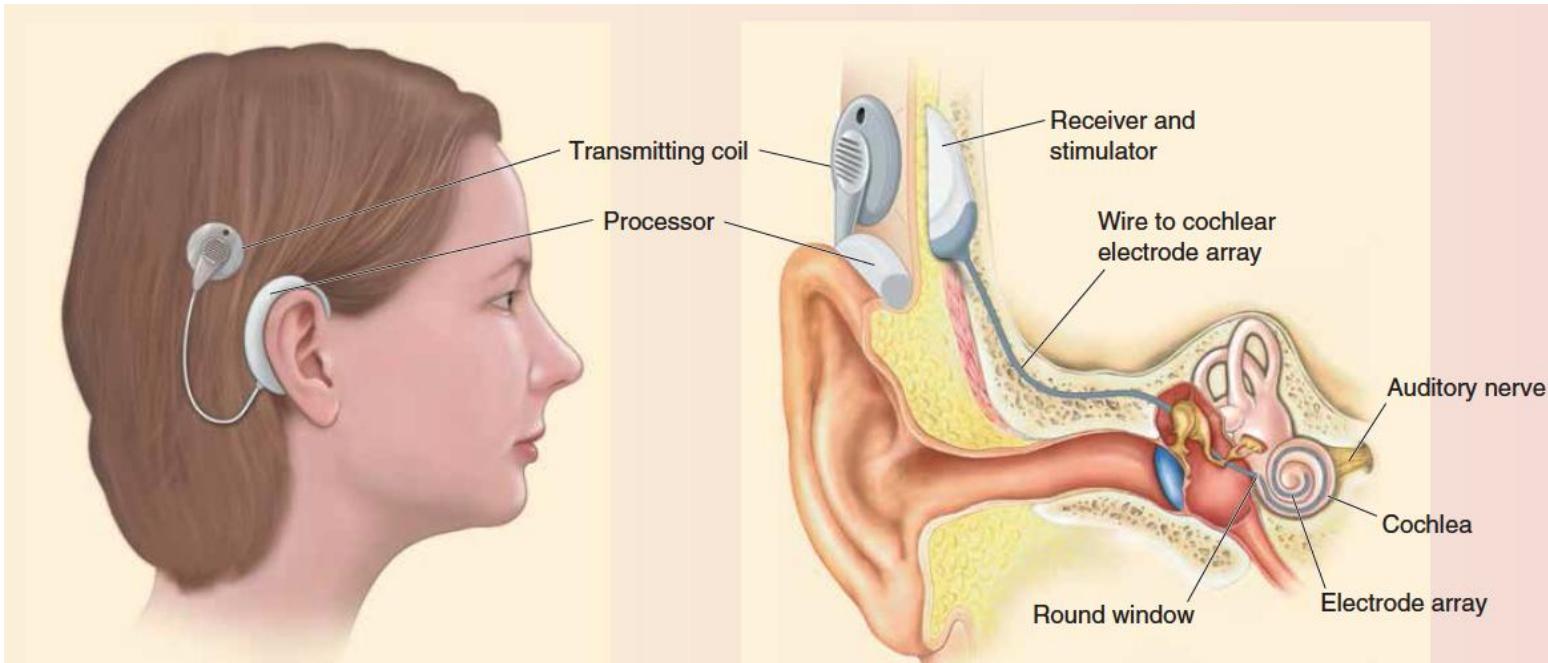


Figure A  
The headpiece of a cochlear implant, positioned behind the ear.

Figure B  
A cochlear implant in place inside the ear.

## How do cochlear implants work?

The cochlear implant takes advantage of the **tonotopic** arrangement of auditory nerve fibers.

**Stimulation** near the **base** of the cochlea evokes a perception of **high-frequency** sounds, while stimulation toward the **apex** evokes **low-frequency** sounds.

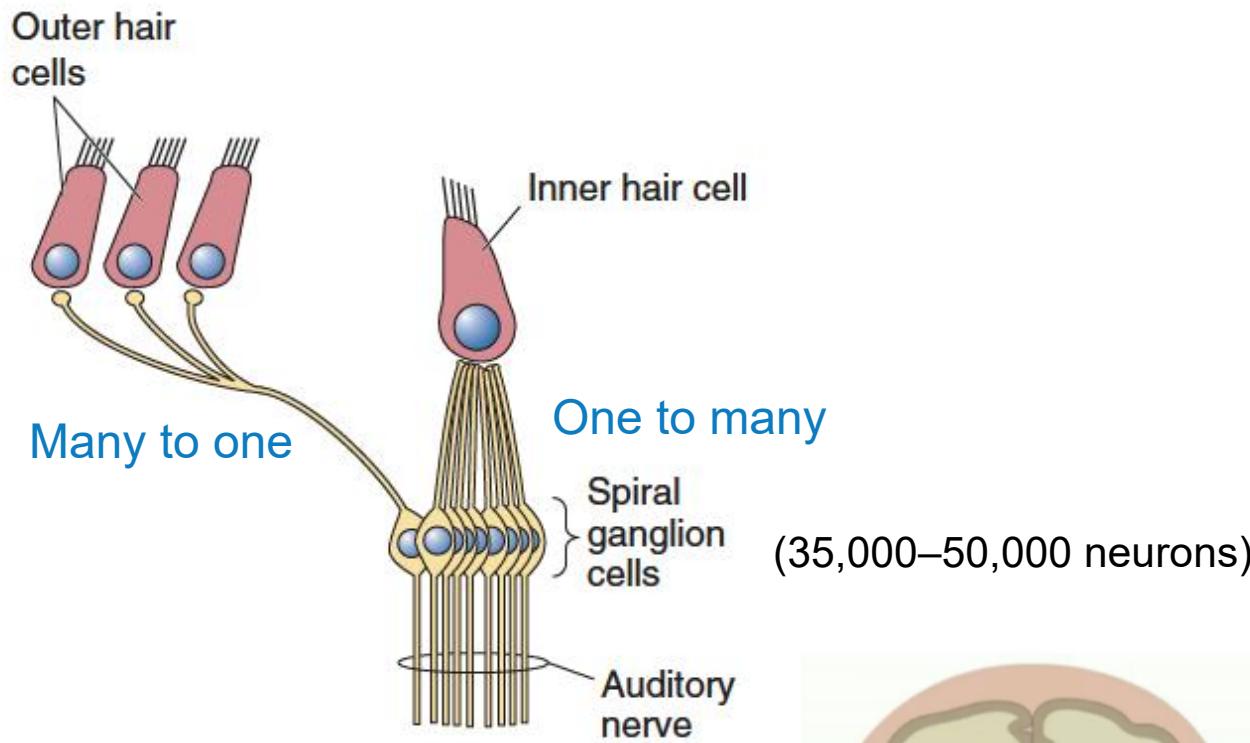
## A limit of cochlear implants

The best candidates for cochlear implants are **young children** (optimally as young as one year).

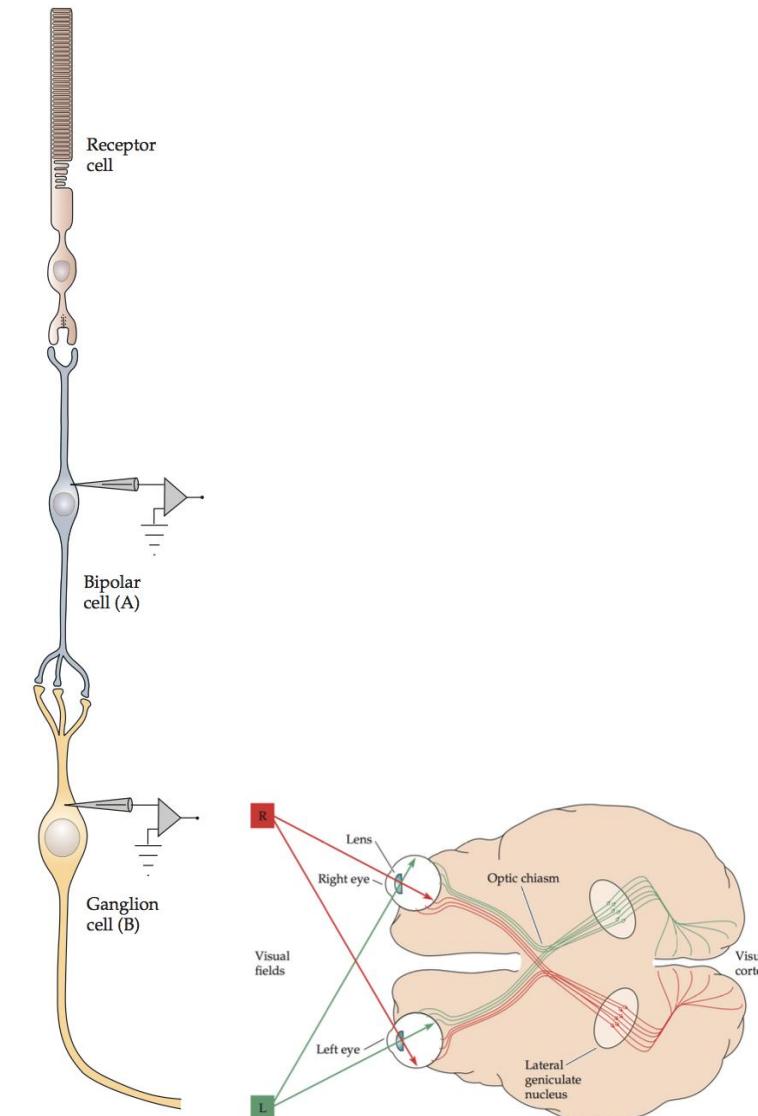
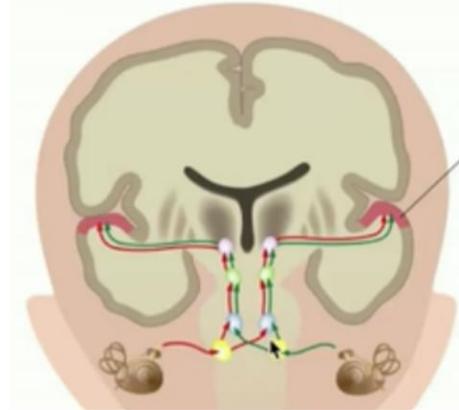
For **adults** whose deafness preceded any experience with speech, on the other hand, cochlear implants seem to provide only a crude perception of sounds.

# Audition: mechanic → electrical

# Vision: electromagnetic → electrical



The **spiral ganglion neurons**, which are the first in the auditory pathway to fire **action potentials**, provide all the auditory information sent to the brain.

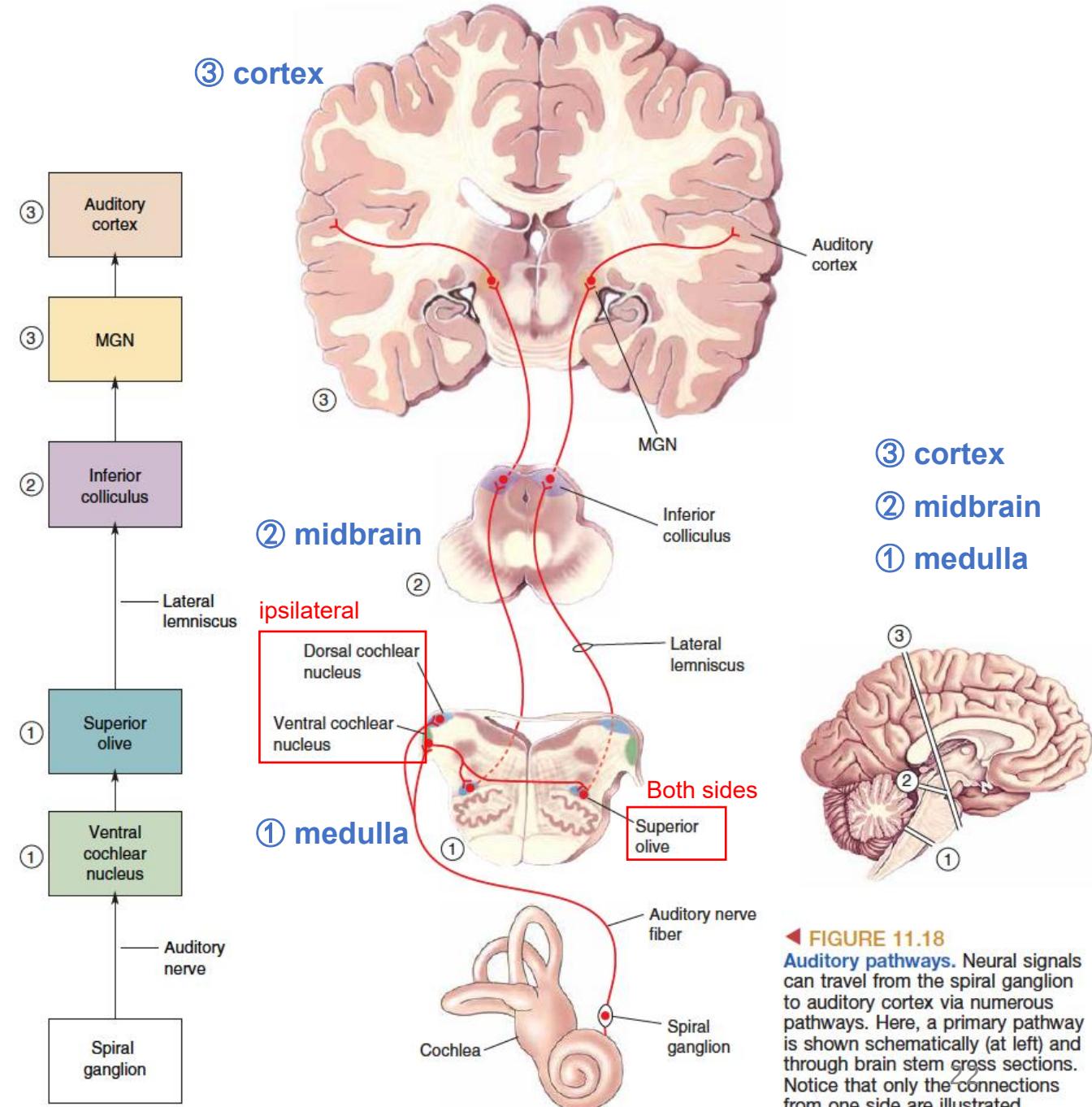


# Central auditory process

Hair cells are excited by deformations of the *basilar membrane*, and each portion of the membrane is **maximally sensitive** to a particular range of frequencies.

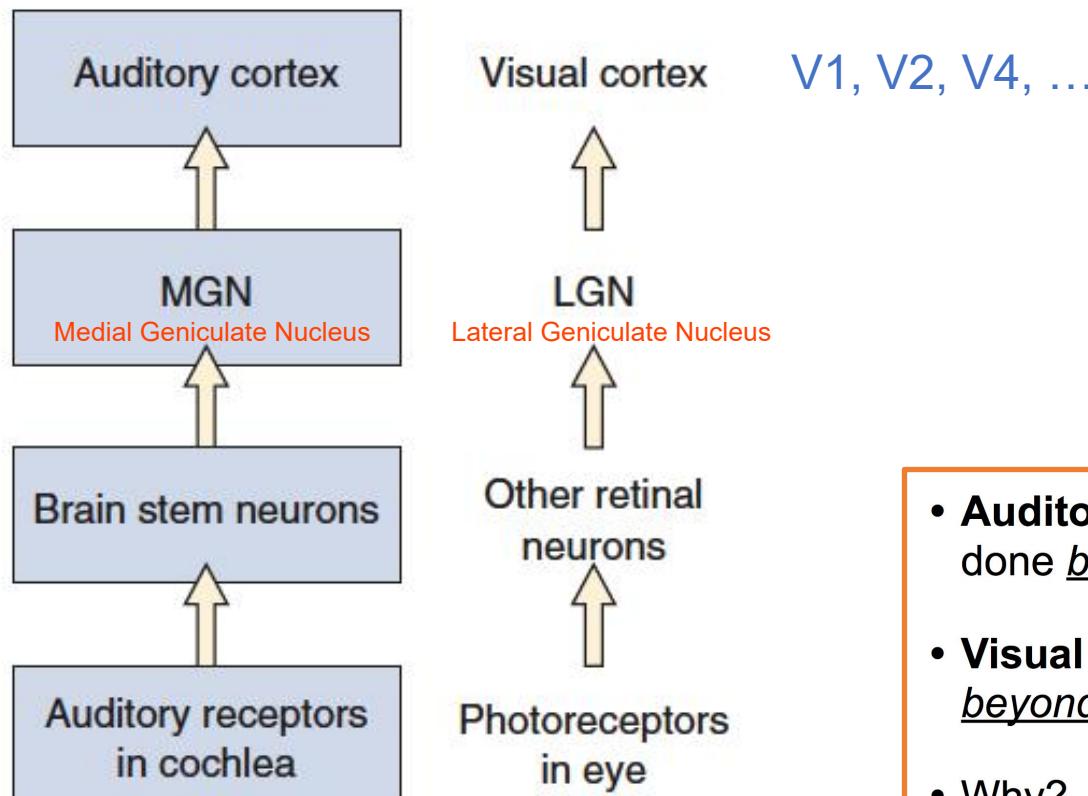
Most *spiral ganglion cells* receive input from a single inner hair cell at a particular location on the basilar membrane, so they fire *action potentials* only in response to sound within **a limited frequency range**.

Spiral ganglion  
→ **dorsal** cochlear nucleus and **ventral** cochlear nucleus **ipsilateral** to cochlea  
→ more complex connections (less well understood)



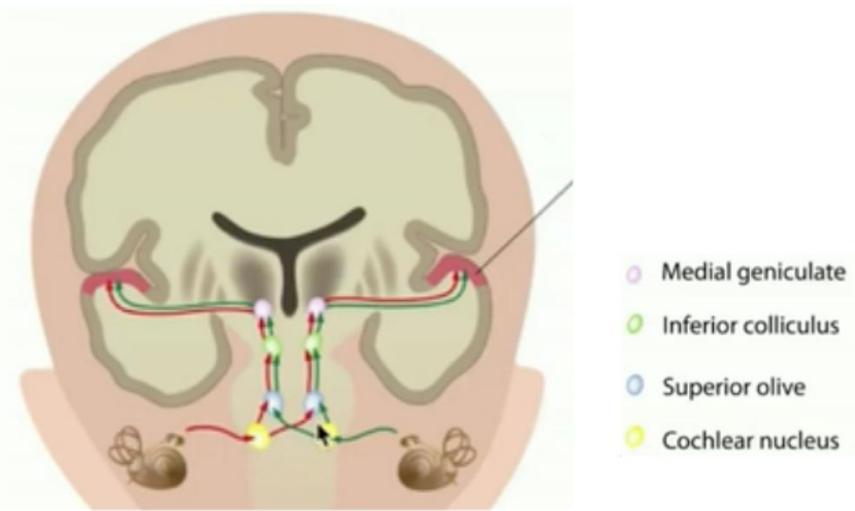
# Auditory vs Visual pathways

Following the sensory receptors, both systems have **early integration** stages, a **thalamic relay**, and a projection to **sensory cortex**.



- **Auditory system:** Large proportion of processing is done before A1
- **Visual system:** Large proportion of processing occurs beyond V1
- Why? May be due to evolutionary reasons...

# General principles of central auditory pathway



- **Parallel** pathways, each analyzing a particular feature
- **Streams separate in cochlear nucleus**: different cell types of project to specific nuclei, similar to ventral “what” pathway and dorsal “where” pathway in visual system
- **Increasing complexity** of responses along the pathway (like vision & touch)

## FUNCTION:

Identify and process complex sounds

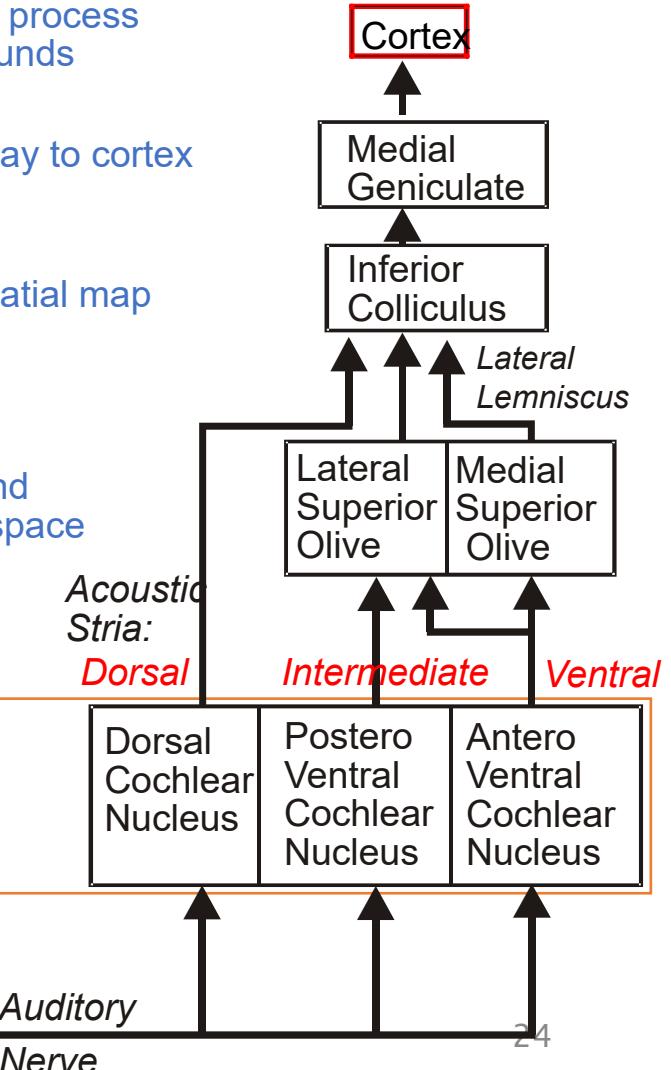
Principal relay to cortex

Form full spatial map

Locate sound sources in space

Start sound feature processing

Cochlea      Auditory Nerve



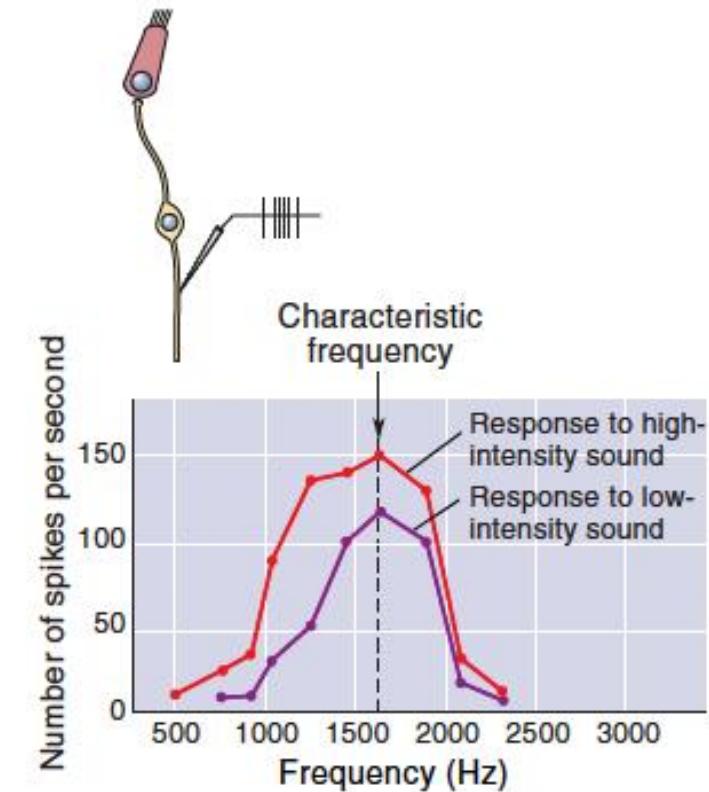
# Encoding sound frequency

The figure in the right side shows the results of an experiment in which action potentials were recorded from a single auditory nerve fiber (i.e., the axon of a ***spiral ganglion cell***).

The neuron is most responsive to sound at one frequency, called the neuron's ***characteristic frequency***, and it is less responsive at neighboring frequencies. This type of frequency tuning is seen in many neurons in each of the relays from cochlea to cortex.

As one ascends, the auditory pathway in the brain stem, the response properties of the cells become more diverse and complex, just as in the visual pathway.

- For instance, some cells in the ***cochlear nuclei*** are especially sensitive to sounds varying in frequency over time (think of the sound of a trombone as it slides from a low note to a high note).
- In the ***MGN***, there are cells that respond to fairly complex sounds such as vocalizations, as well as other cells that show simple frequency selectivity, as in the auditory nerve.



▲ **FIGURE 11.19**  
The response of an auditory nerve fiber to different sound frequencies. This neuron is frequency-tuned and has its greatest response at the characteristic frequency. (Source: Adapted from Rose, Hind, Anderson, and Brugge, 1971, Fig. 2.)

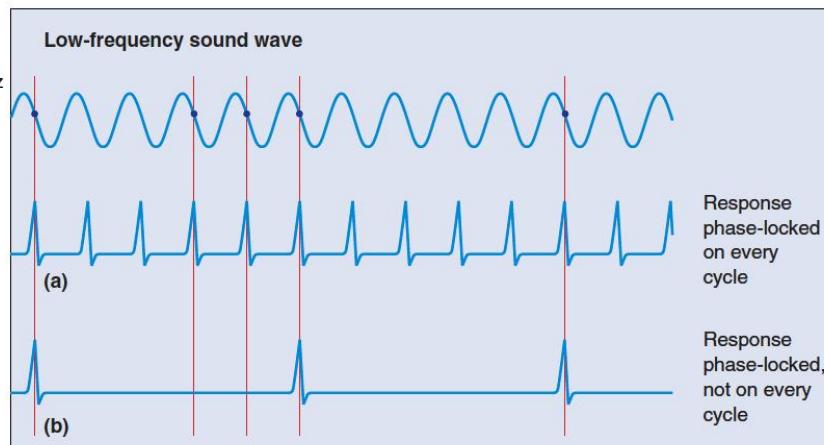
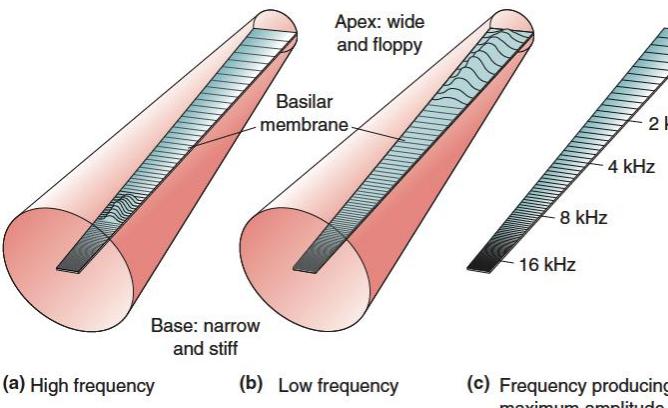
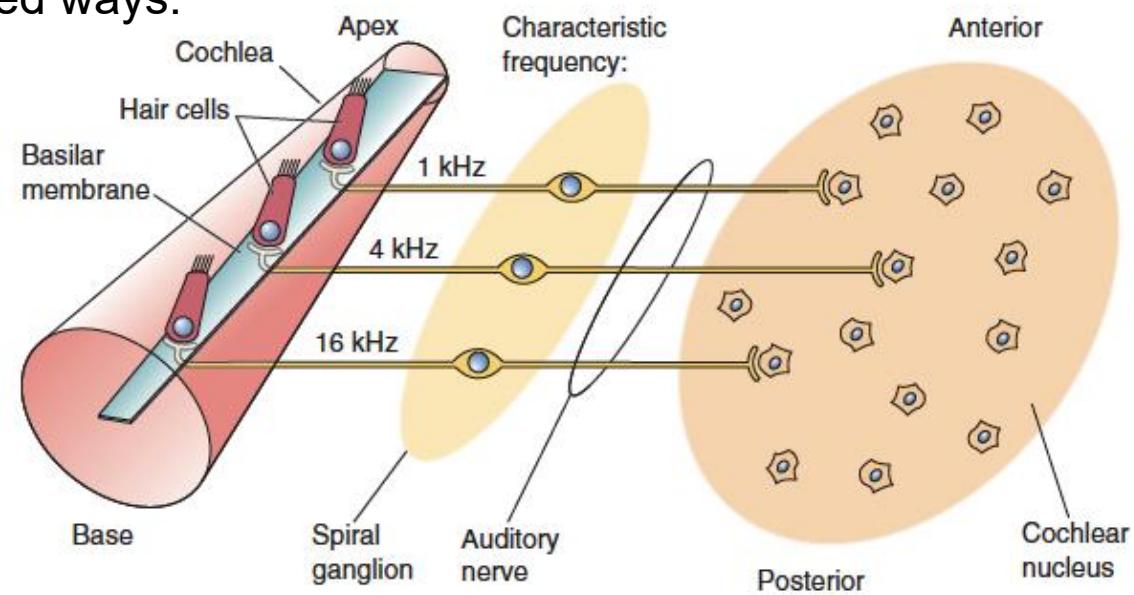
# Encoding sound intensity and frequency

Information about **sound intensity** is coded in two interrelated ways:

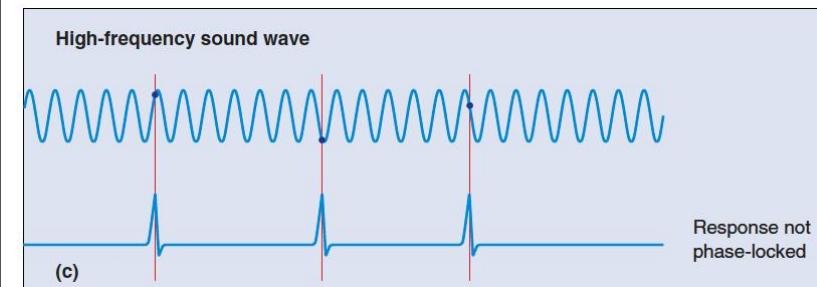
1. the **firing rates** of neurons
2. the **number of active neurons**

Information about **sound frequency**

1. **Tonotopy** in basilar membrane, spiral ganglion, and cochlear nucleus, MGN, and auditory cortex
2. **Phase locking**: The timing of neural firing is phase locking, which means the consistent firing of a cell at the same phase of a sound wave.



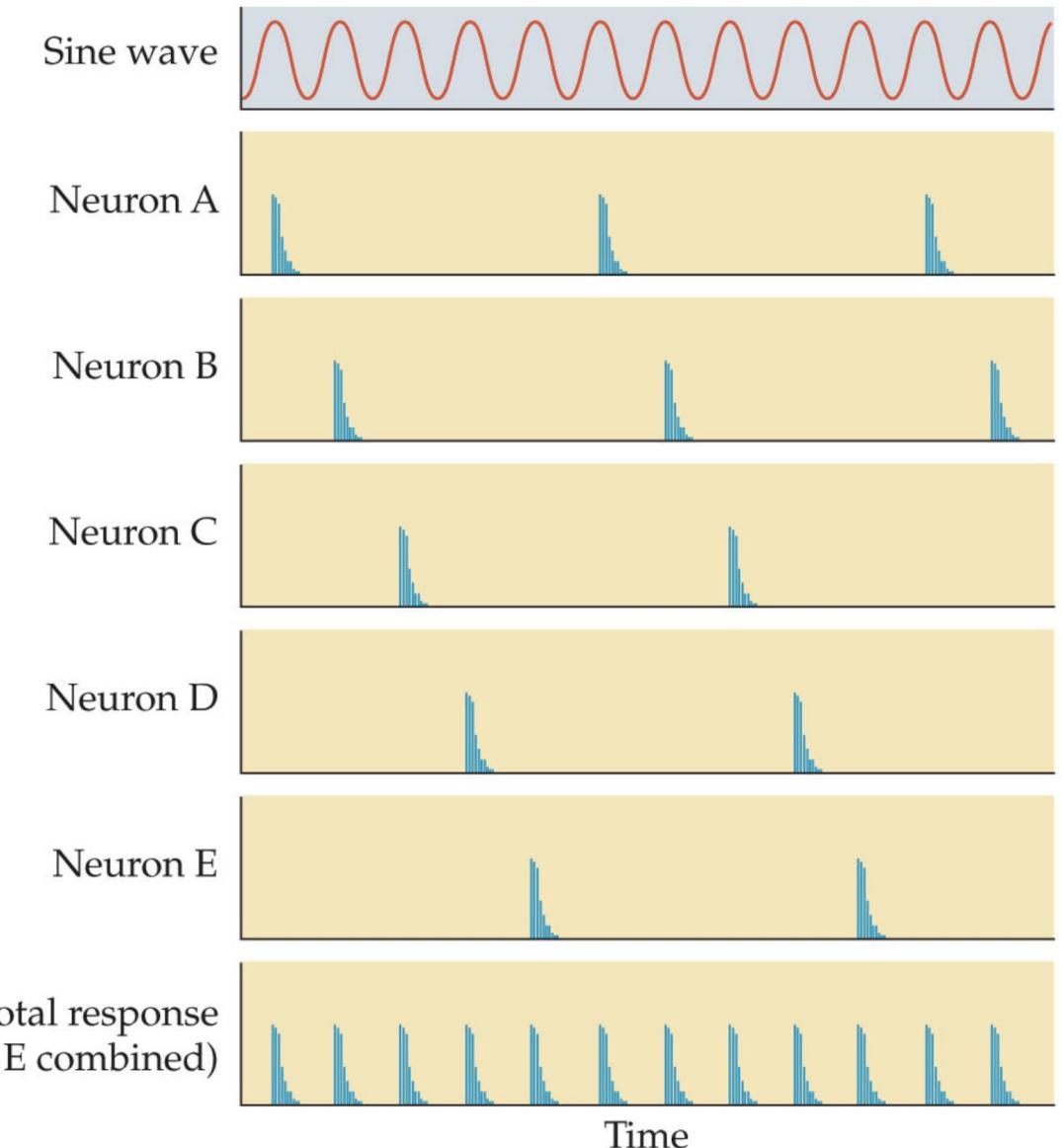
**Phase locking** occurs with sound waves up to about **5 kHz**.



# Encoding sound intensity and frequency

## The volley principle:

- multiple neurons can provide a *temporal code* for frequency by working together.



# Why DECIBELS for sound intensity?

- LOUDNESS perception: also LOGARITHM of the physical variable (intensity).
  - Fechner (1860) noticed: “equal” steps of perceived loudness actually multiples of each other in intensity.  
→ Logarithmic
  - Defined: log scale (Bel)
  - **10  $\log_{10}$  ( $I / I_{th}$ ) Decibels:** 
$$dB_p = 10 * \log\left(\frac{I_{sig}}{I_{ref}}\right)$$
  - Threshold: 0 dB: ( $1/10^{10}$  atmospheric pressure)
  - Max: 5,000,000 larger in amplitude,  $10^{13}$  in power
  - Huge range.
- Encodes loudness with firing rate & numbers
- Adapts to this huge range (like light intensity)

# Psychoacoustics:

- The study of the psychological correlates of the physical dimensions of acoustics
- The study of subjective human perception of sounds

Each line corresponds to tones rated by observers as having the same loudness.

## Physical Property

Frequency

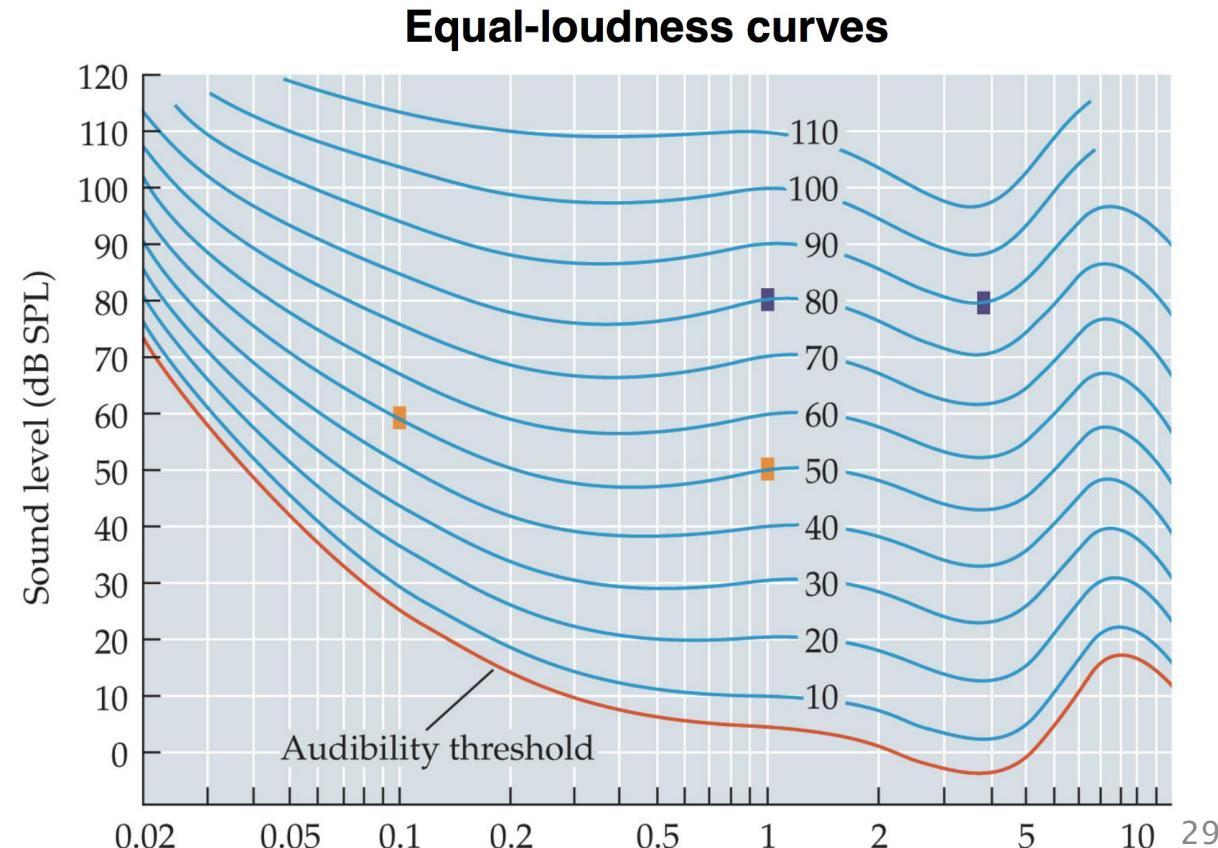


Pitch

Amplitude / Intensity



Loudness

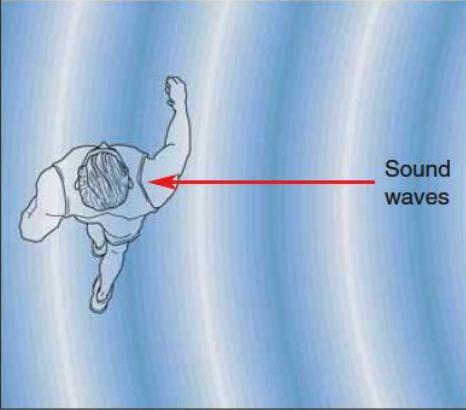


# Mechanisms of sound localization

**Sound localization** can be of critical importance for survival.

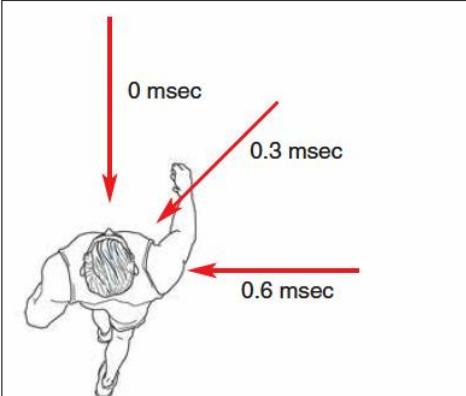
We use different techniques for locating sources in the **horizontal** (left–right) and **vertical** plane (up–down).

## Horizontal plane



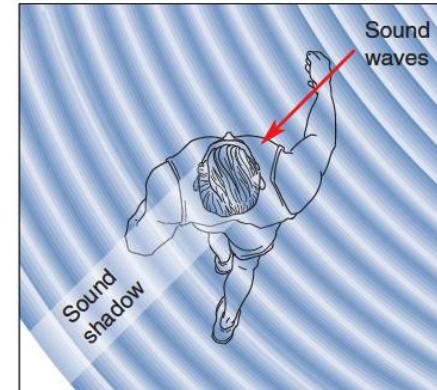
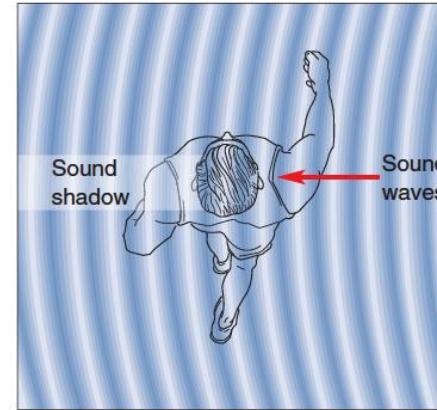
### interaural time delay (ITD)

If the distance between your ears is 20cm, sound coming from the right, perpendicular to your head, will reach your left ear 0.6 msec after reaching your right ear. There is a simple *relationship* between *location* and *interaural time delay*.



People can discriminate directions with a precision of about **2°**, which requires to discriminate the **11  $\mu$  sec** between the time it takes a sound to reach their two ears.

### interaural intensity difference (IID)



An **interaural intensity difference** exists between the two ears because your head effectively casts a sound shadow. Neurons sensitive to differences in intensity can use this information to locate the sound.

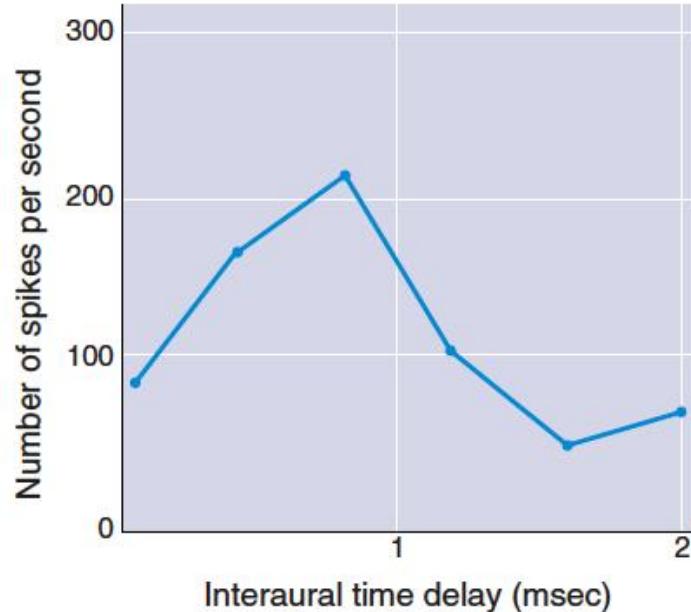
**Low-frequency**: With sounds in the range of 20–2k Hz, the process involves *interaural time delay (ITD)*.

**High-frequency**: From 2k–20k Hz, *interaural intensity difference (IID)* is used.

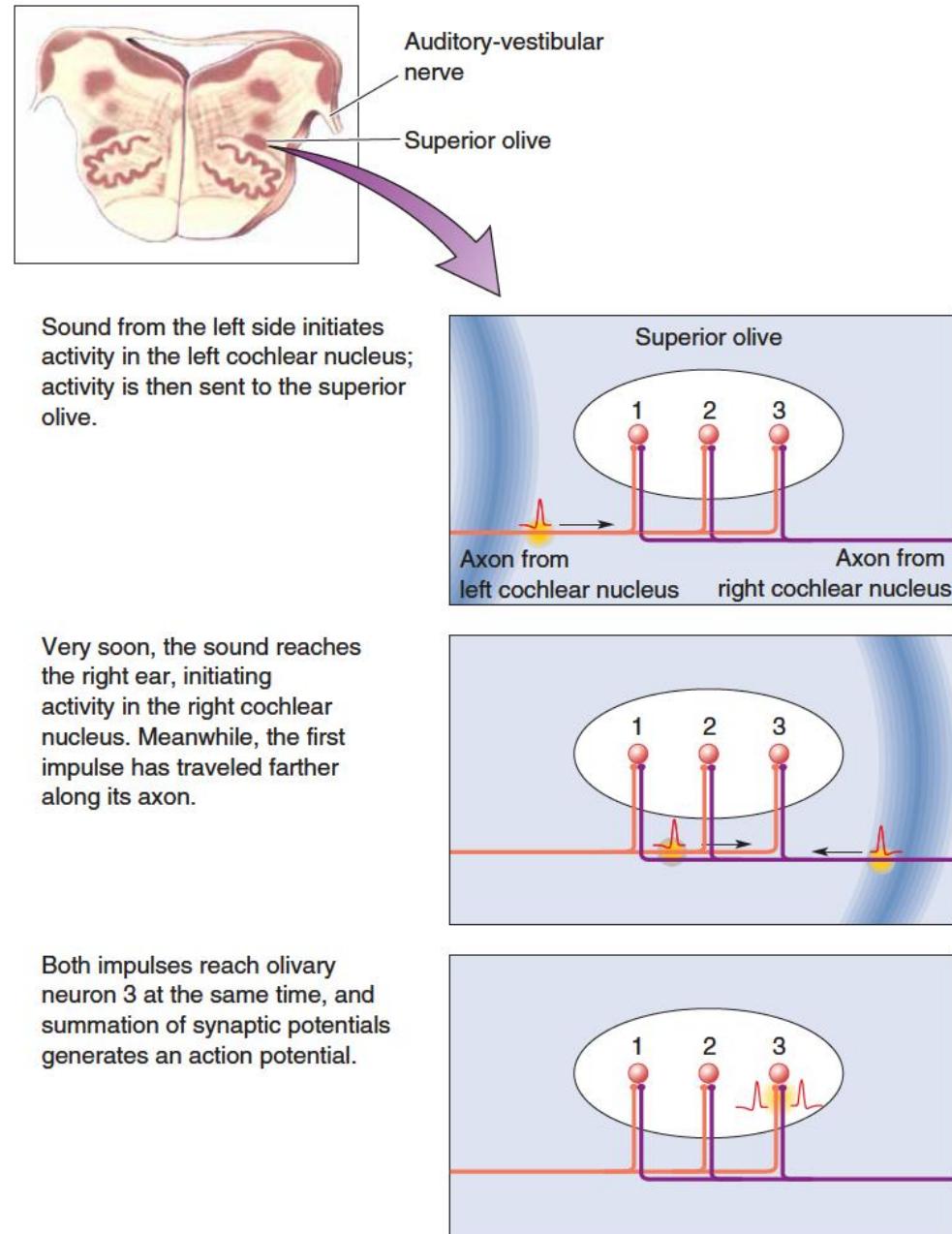
# Mechanisms of sound localization

The first structure where *binaural neurons* are present is the *superior olive*.

*Phase locking* is essential for a precise comparison of the timing of the inputs.



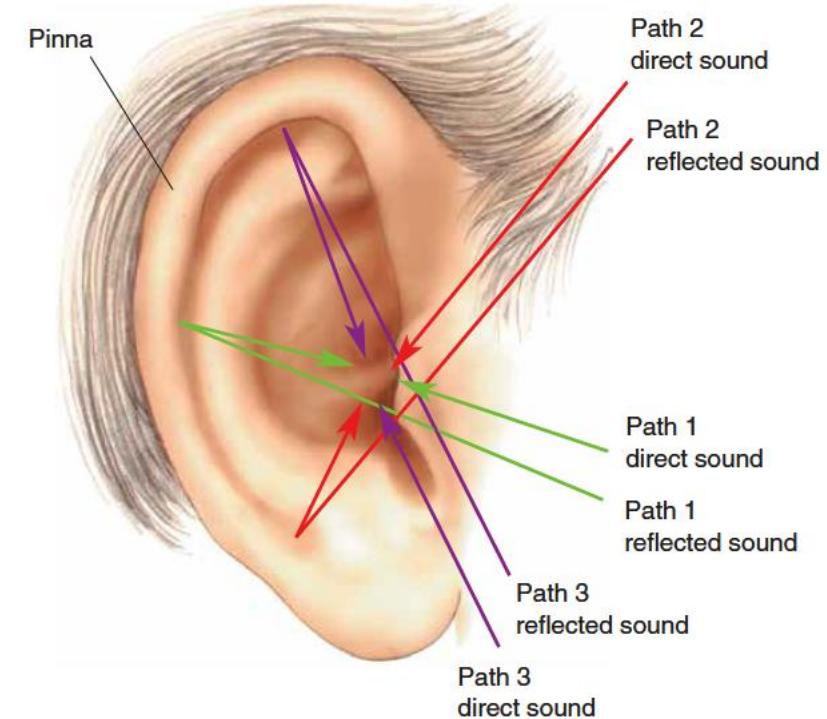
▲ FIGURE 11.24  
Responses of a neuron in the superior olive sensitive to interaural time delay.  
This neuron has an optimal delay of about 1 msec.



# Mechanisms of sound localization

## Vertical plane: reflections from the pinna

- The sweeping curves of the outer ear are essential for assessing the elevation of a source of sound.
- The bumps and ridges produce reflections of the entering sound.
- The **delays** between the direct path and the reflected path change as a sound source moves **vertically**.
- The combined sound, direct and reflected, is subtly **different** when it comes from above or below.



## Animal's skills

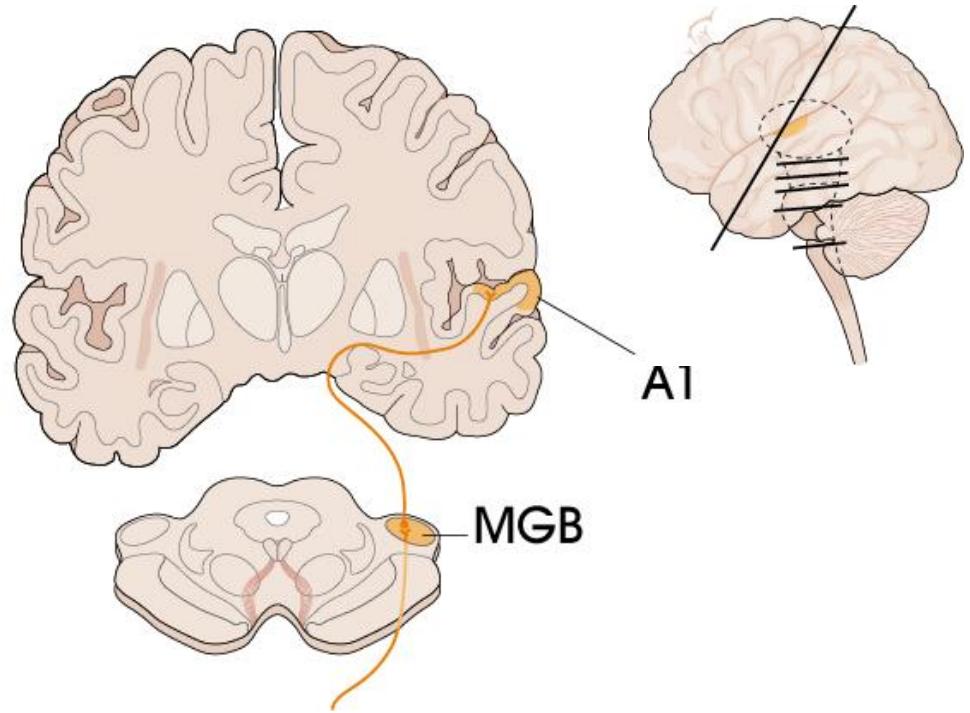
For example, a barn owl can swoop down on a squeaking mouse in the dark, locating accurately by sound, not sight. Although owls do not have a pinna, they can use the same techniques we use for horizontal localization (interaural differences) because their two ears are **at different heights** on their head.

Certain bats **emit sounds** that are reflected off objects, and these echoes are used to locate objects without sight.



# Auditory Cortex

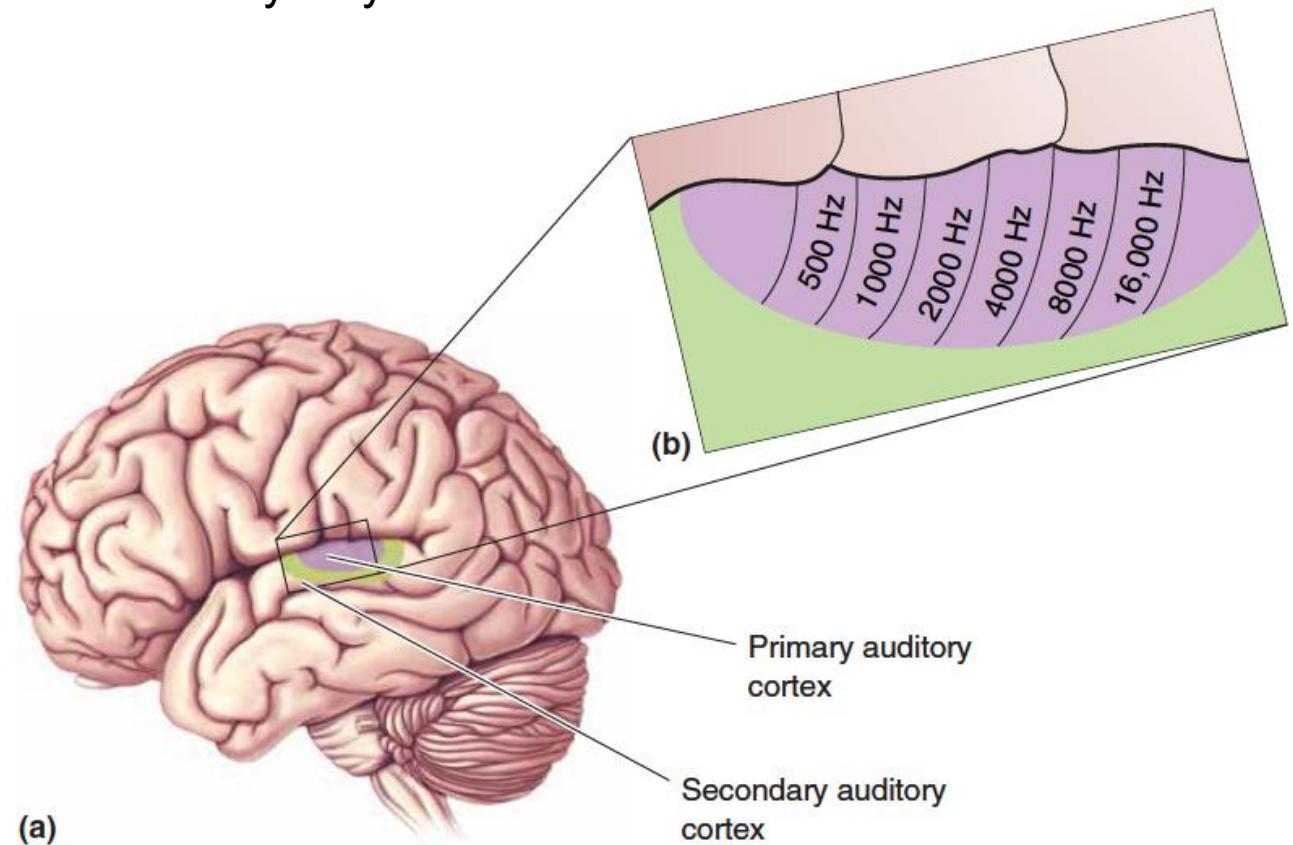
- Progressively more complex
- 15 distinct tonotopic areas.
- A1: Primary Auditory Cortex:  
Superior temporal gyrus
- Like other sensory cortex:
  - 6 layers
  - Input layer: IV,
  - V: back project to MGB.
  - VI: back project to IC
  - Cortical columns (map),
- Logarithmic map of frequency.
- Perpendicular to frequency axis:
  - binaural interactions: EE, EI,
  - rising or falling pitch



# Auditory Cortex

The structure of **A1** and **the secondary auditory areas** is in many ways similar to corresponding areas of the visual cortex.

- Layer I contains few cell bodies, and layers II and III contain mostly small pyramidal cells.
- Layers V and VI contain mostly pyramidal cells that tend to be larger than those in the superficial layers.
- Layer IV, where the medial geniculate axons terminate, is composed of densely packed granule cells.



▲ FIGURE 11.27

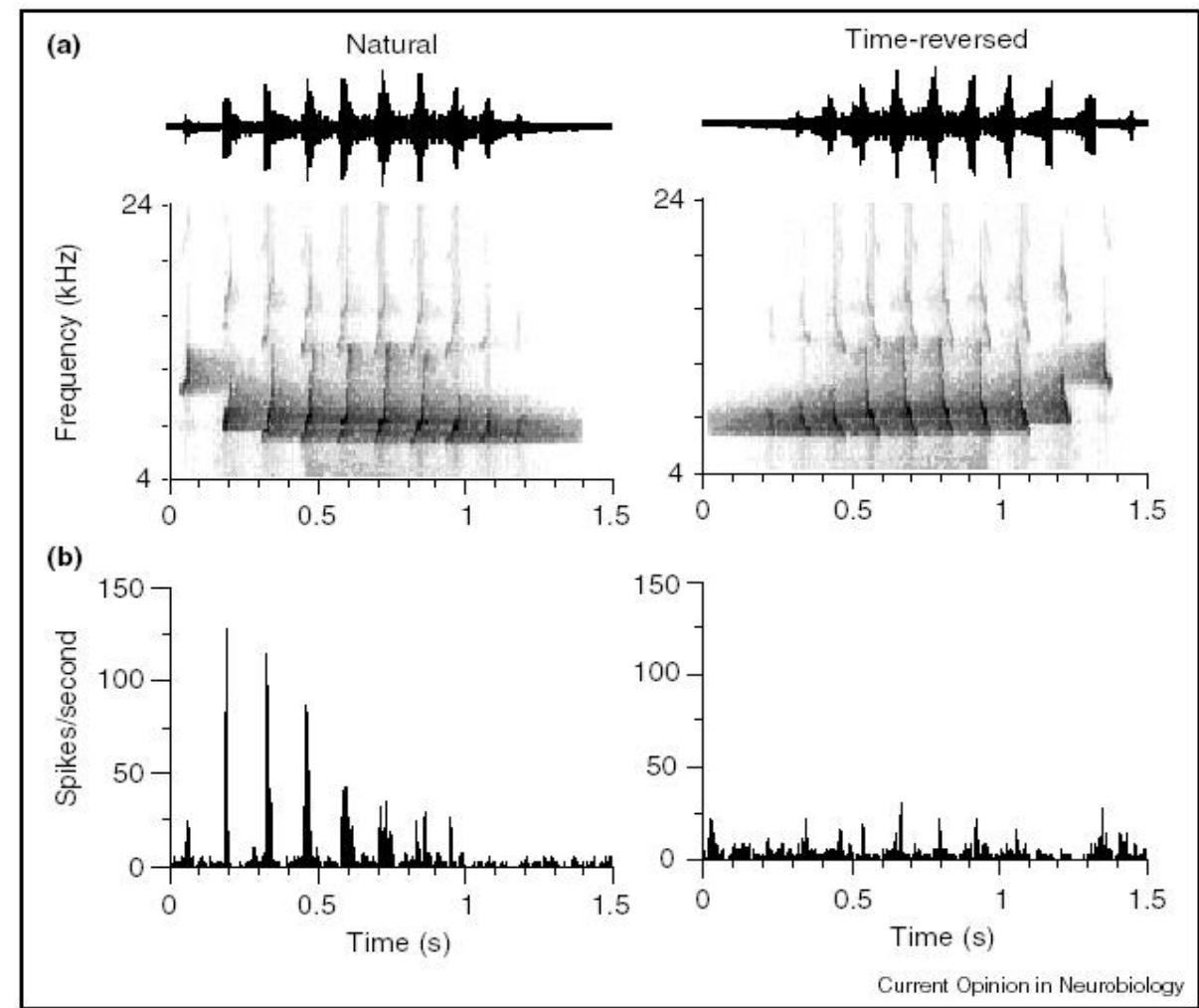
**Primary auditory cortex.** (a) Primary auditory cortex (purple) and secondary auditory areas (green) on the superior temporal lobe. (b) Tonotopic organization within primary auditory cortex. The numbers are characteristic frequencies.

# Auditory Cortex: Complex patterns

- Cortical cells: tuned to precise sequence of complex sounds
- Particularly, ethologically important sounds
- Marmoset A1 response to its own twitter call

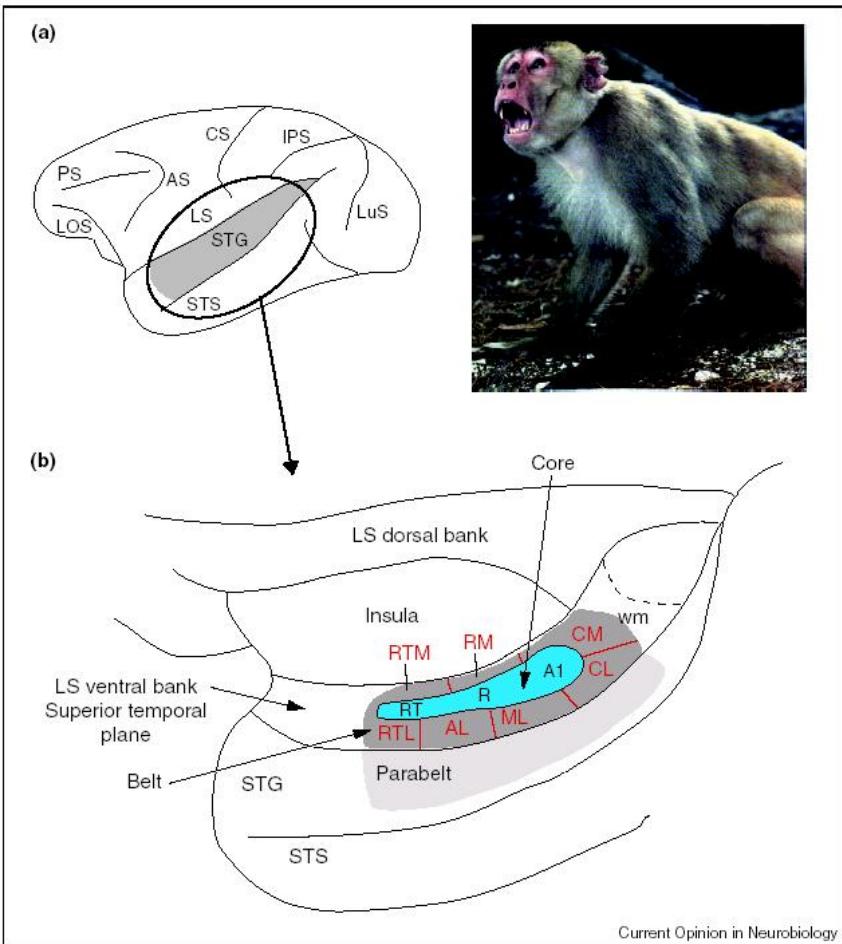


AA Ghazanfar & M D Hauser

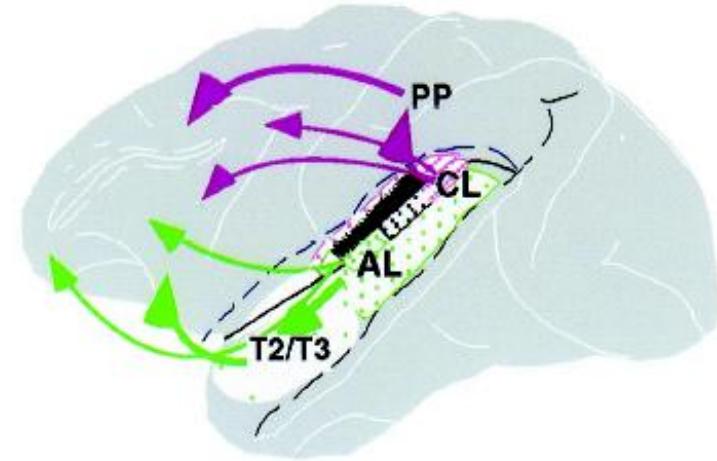


# Auditory Cortex: What & Where pathways

A1 = primary or core auditory area;  
R = rostral core auditory area;  
RM = rostromedial;  
RT = rostral temporal core auditory area;  
RTM = RT medial;  
RTL = RT lateral;  
AL = anterolateral belt auditory cortex;  
ML = middle-lateral belt auditory cortex;  
CM = caudal-medial belt auditory cortex;  
CL = caudal-lateral belt auditory cortex.



"Where"



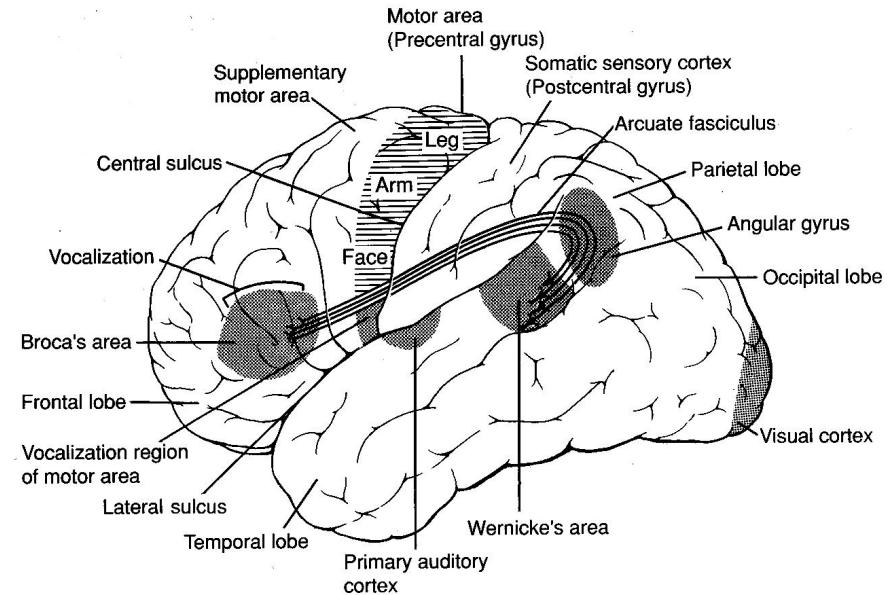
"What"

(e.g. food, or alert, or emotional content)

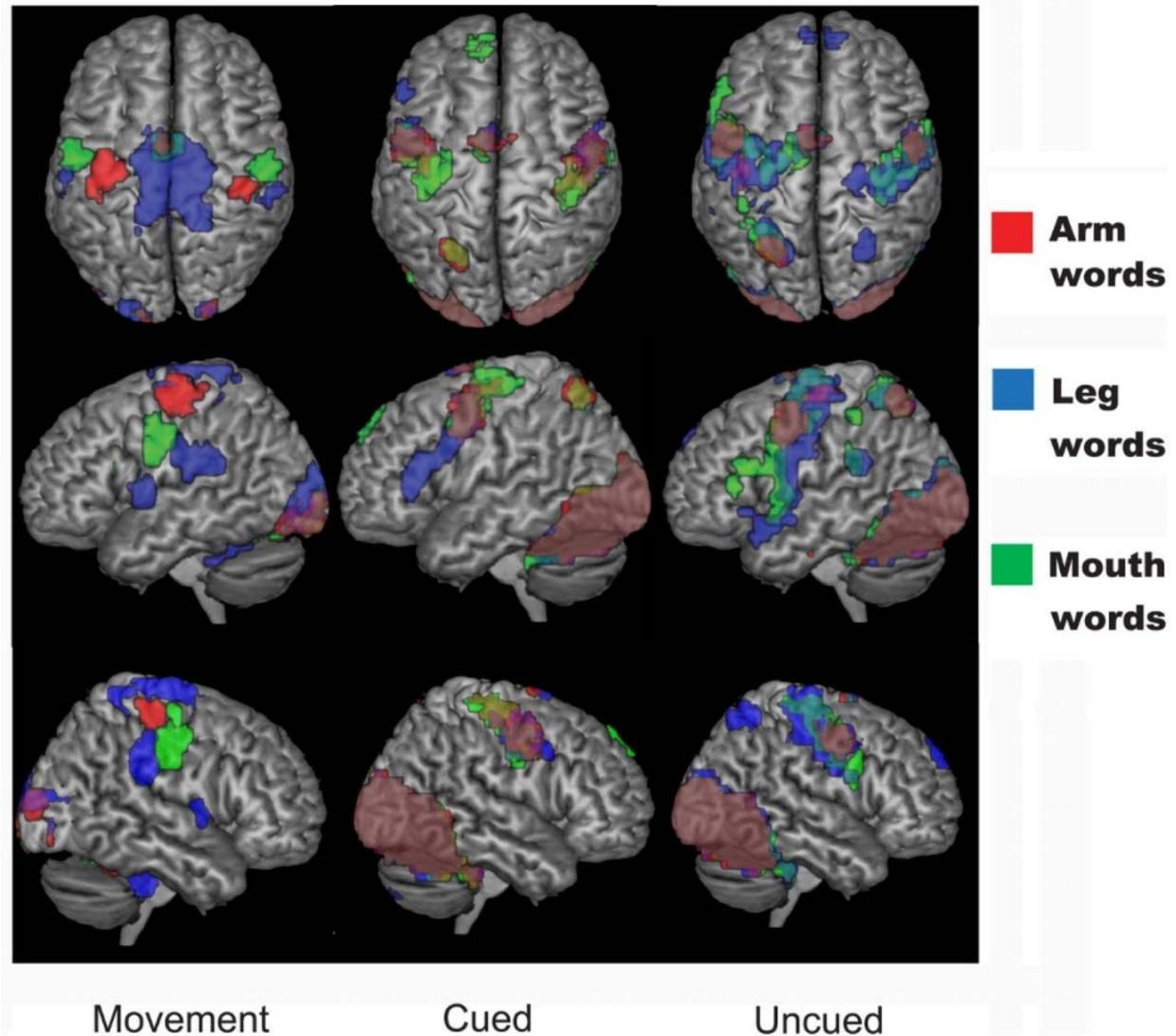
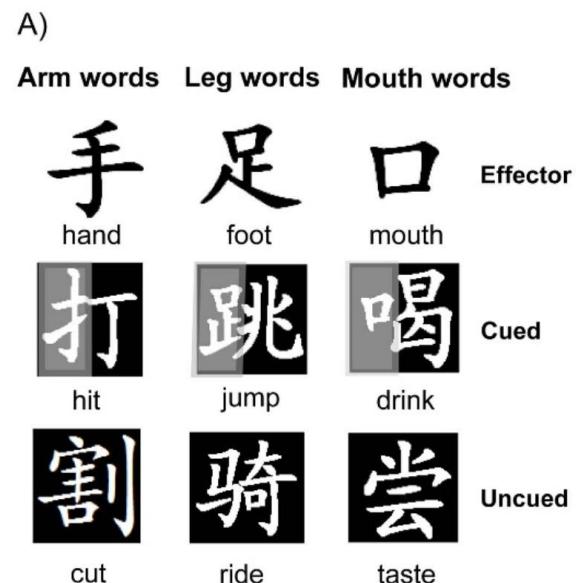
- Rhesus monkey: “belt” or secondary auditory cortex

# Auditory Cortex: Speech areas

- Classical division on basis of aphasia following lesions:
  - **Broca's area**: understand language but unable to speak or write
  - **Wernicke's area**: speaks but cannot understand
- Current understanding: not uniform areas. Rather, category-specific with strongest activation proximal to the sensory or **motor area** associated with that category:
  - Words for manipulable objects (tools) activate reaching / grasping motor areas
  - Words for movement activate next to visual motion areas
  - Words for complex objects (faces) activate visual recognition areas

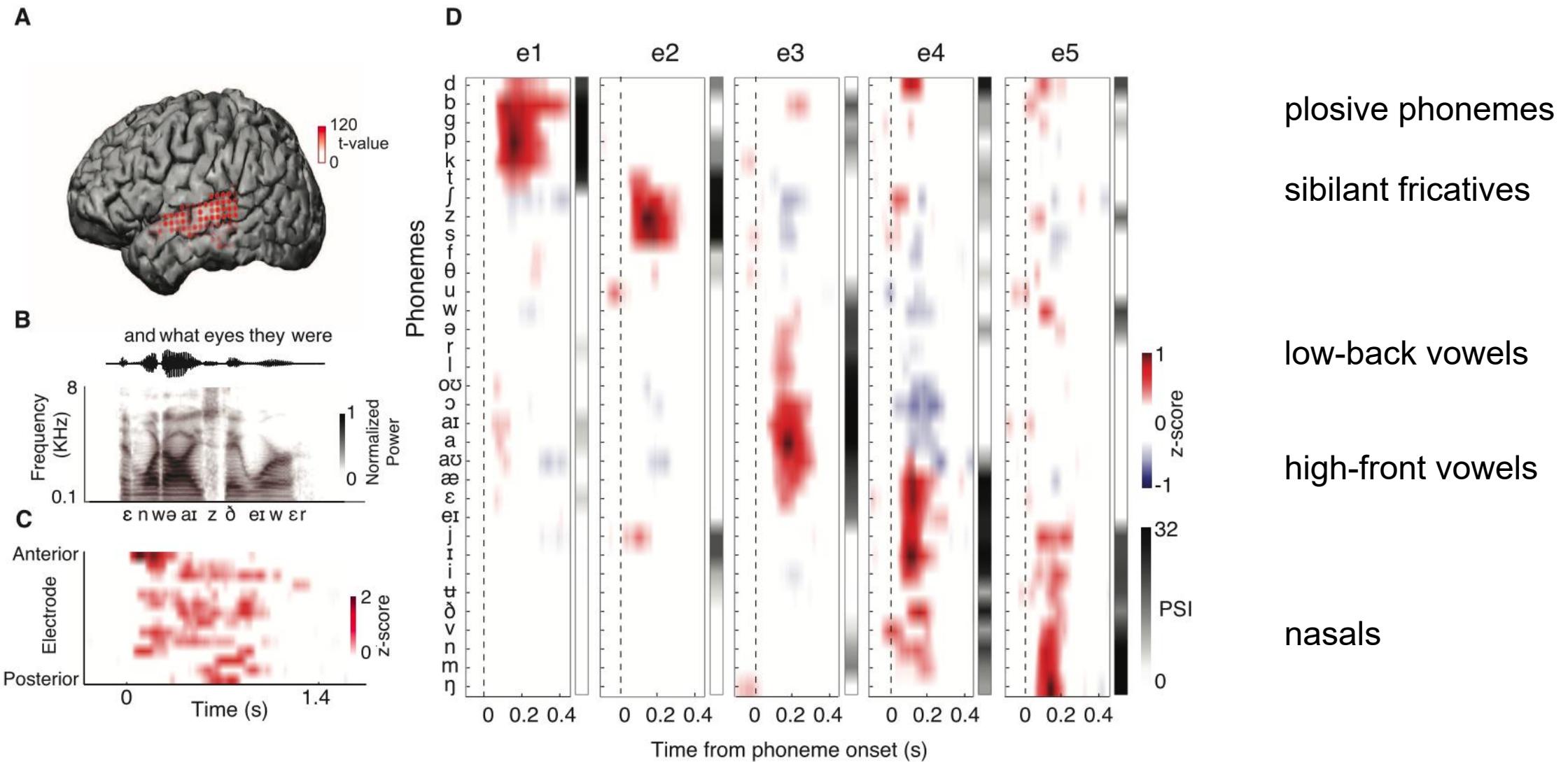


# Dissociable Somatotopic Representations of Chinese Action Verbs in the Motor and Premotor Cortex

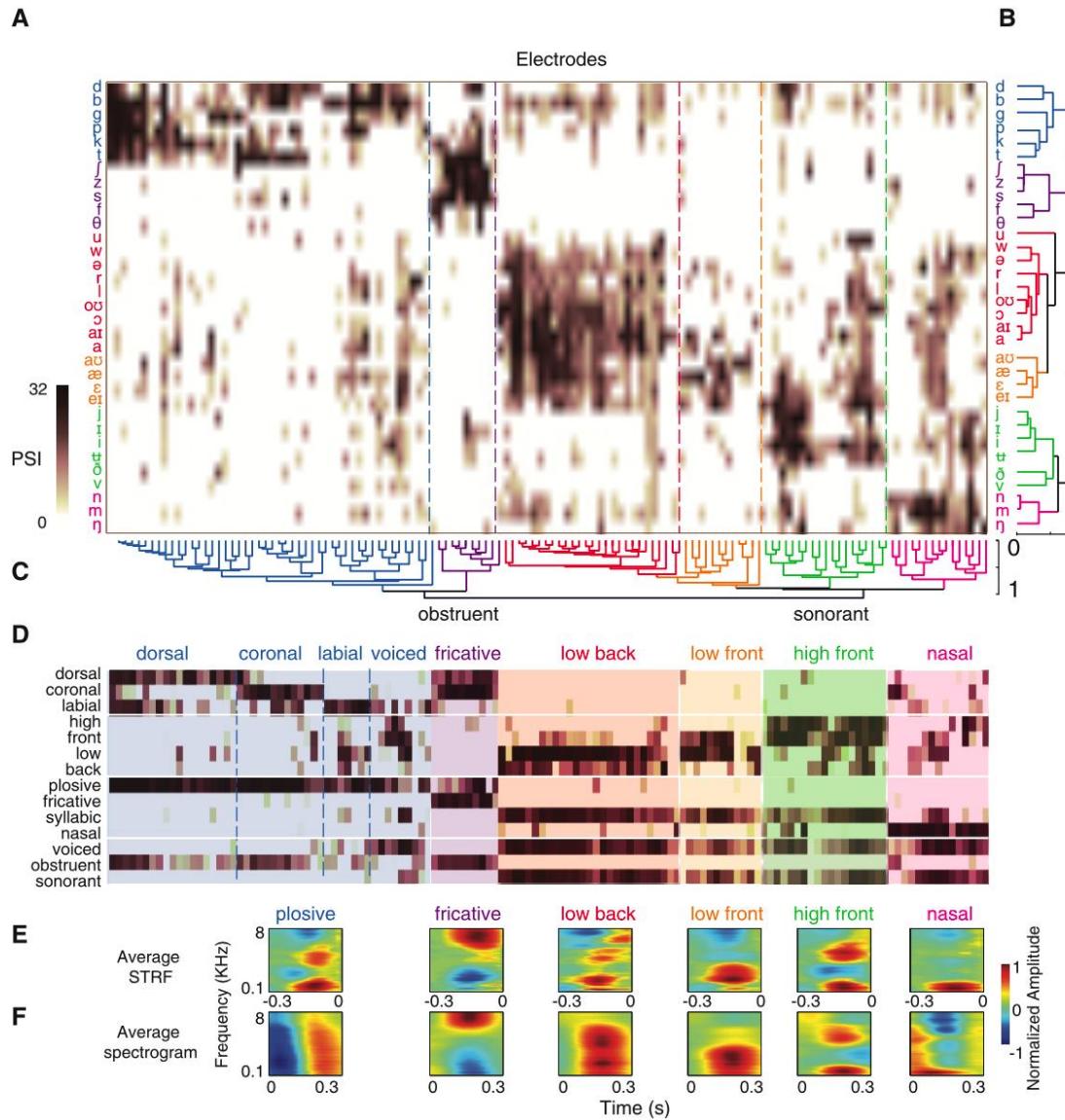


The results support the universality of somatotopic representation of action verbs in the motor system. 38

# Phonetic Feature Encoding in Human Superior Temporal Gyrus



# Phonetic Feature Encoding in Human Superior Temporal Gyrus



Hierarchical clustering of single-electrode and population responses.

(A) phoneme selectivity index (PSI) vectors (33 English phonemes) of selective electrodes across all participants. Rows correspond to phonemes, and columns correspond to electrodes.

(B) Clustering across population PSIs (rows).

(C) Clustering across single electrodes (columns).

(D) Alternative PSI vectors using rows now corresponding to **phonetic features**, not phonemes.

(E) Weighted average STRFs of main electrode clusters.

(F) Average acoustic spectrograms for phonemes in each population cluster.

# Cortical Plasticity in Auditory System

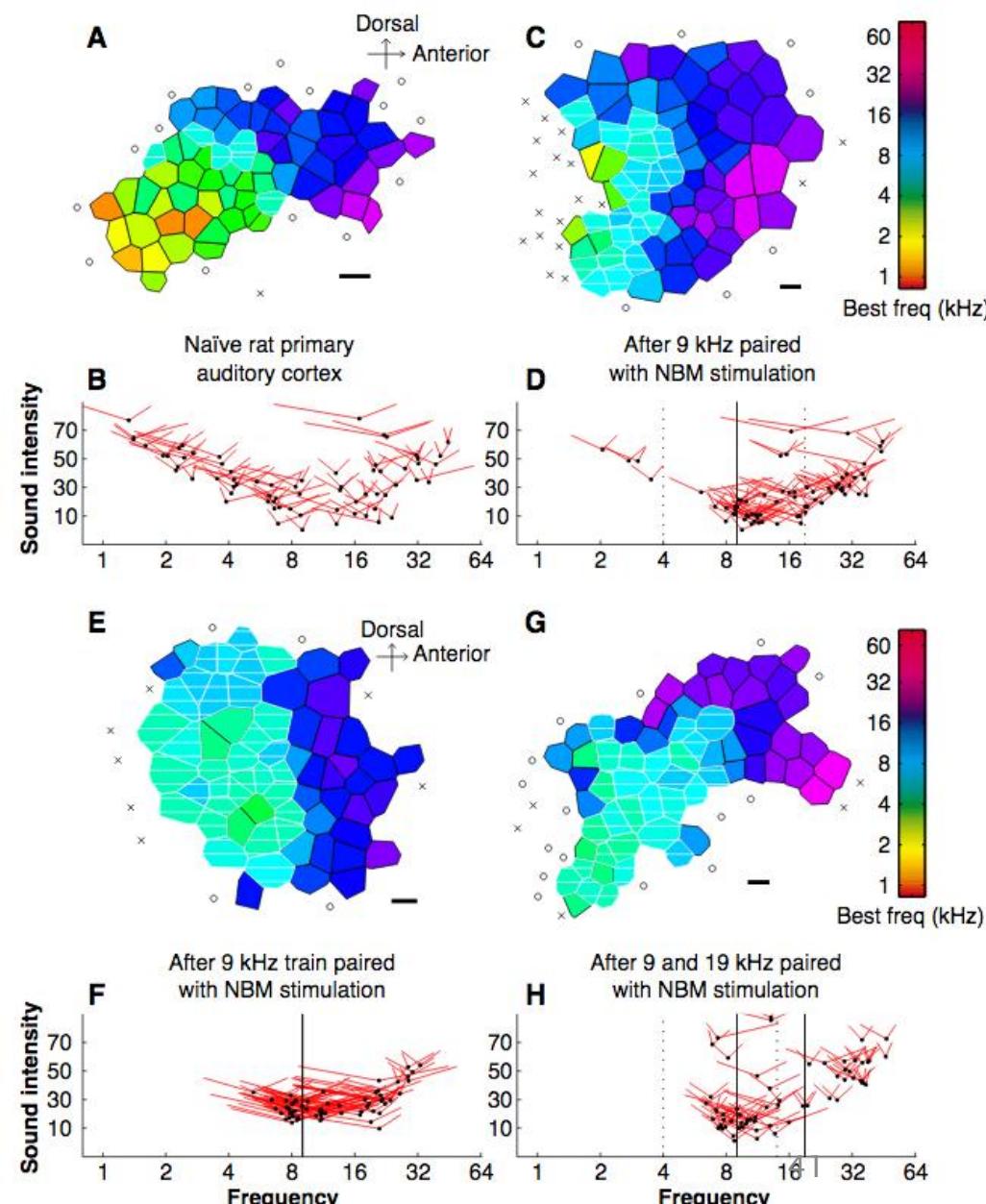
This study demonstrates that episodic electrical stimulation of the **nucleus basalis**, paired with an auditory stimulus, results in a massive progressive reorganization of the primary auditory cortex in the adult rat.

Receptive field sizes can be narrowed, broadened, or left unaltered depending on specific parameters of the acoustic stimulus paired with nucleus basalis activation.

This result suggests that input characteristics may be able to drive appropriate alterations of receptive fields *independently* of explicit knowledge of the task.

These findings also suggest that the basal forebrain plays an active instructional role in representational plasticity.

Kilgard & Merzenich (1998) Science



In humans, the basic elements of speech, called **phonemes**, are common to all languages and are the sounds first babbled by babies, before particular sounds are selected to be combined into words.

One evident challenge for *auditory perception* is to identify a particular temporal pattern of frequency changes as an acoustic object, even as the absolute frequencies vary with different speakers (male vs female voices, for example).

By analogy with the *visual system*, which contains cells that respond to slits, corners, edges, and other geometrical forms, we might expect to find higher-order cells in the human auditory cortex that respond to particular *formants* or, perhaps, *phonemes*.

However, in contrast to the other senses, auditory information is extensively processed **at subcortical levels**, as suggested by its more complex subcortical circuitry.

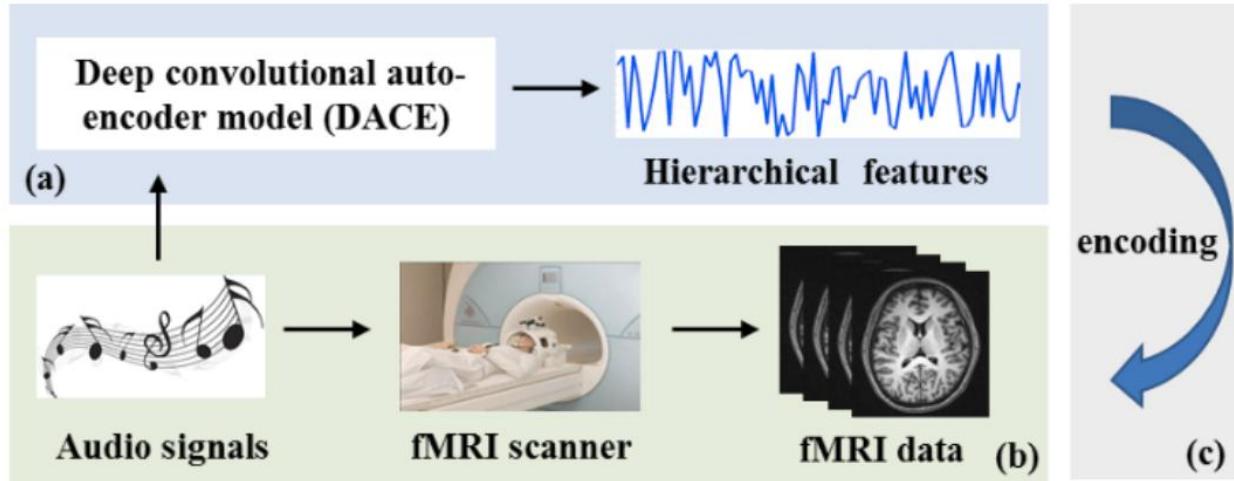
**How to use deep learning to understand the auditory processing in brain?**

**Q1:**

**Can we predict the neural signals  
based on the input auditory signals?**

neural encoding:  
sound → neural activity

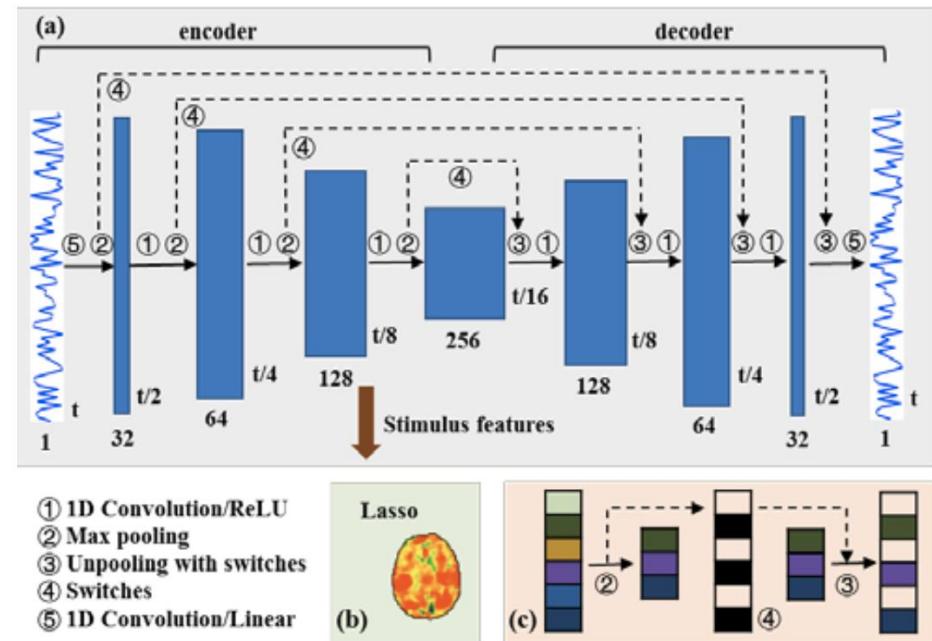
# Deep learning for predicting neural signals



**Fig. 1.** The framework of this study. (a) Hierarchical feature extraction of audio excerpts via DCAE. (b) FMRI acquisition. (c) FMRI encoding models based on LASSO.

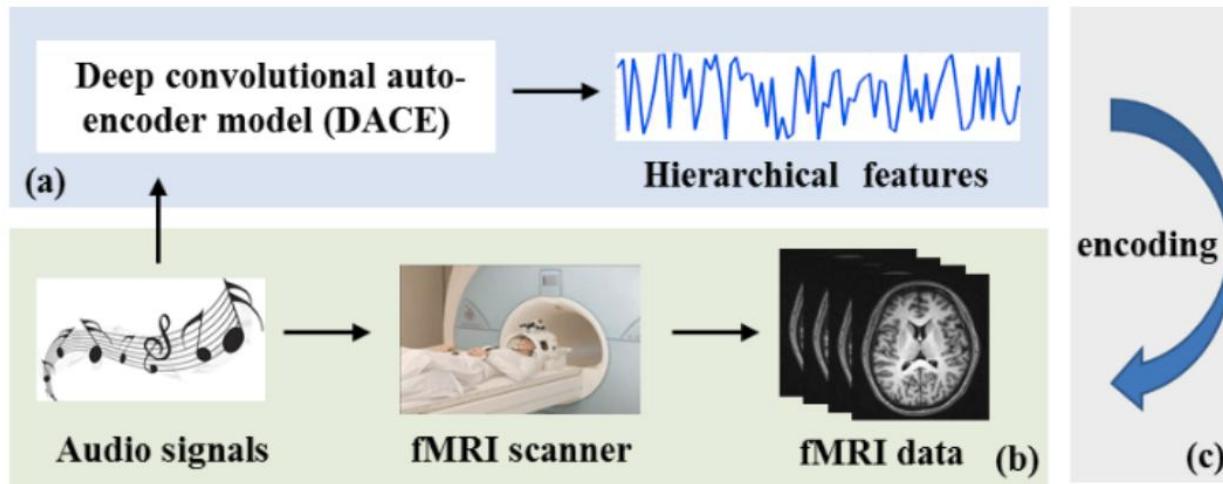
$$\min_{\alpha_i \in R^m} \frac{1}{2} \|x_i - D\alpha_i\|_F^2 + \lambda \|\alpha_i\|_1 \quad (4)$$

where  $x_i$  is the fMRI signal of each voxel,  $D$  is the hierarchical features represented in each layer and  $\alpha_i$  is the encoding coefficient matrix.  $\ell_1$  is a term to enforce the sparsity of  $\alpha_i$  and  $\lambda$  is a balance controlling parameter. In

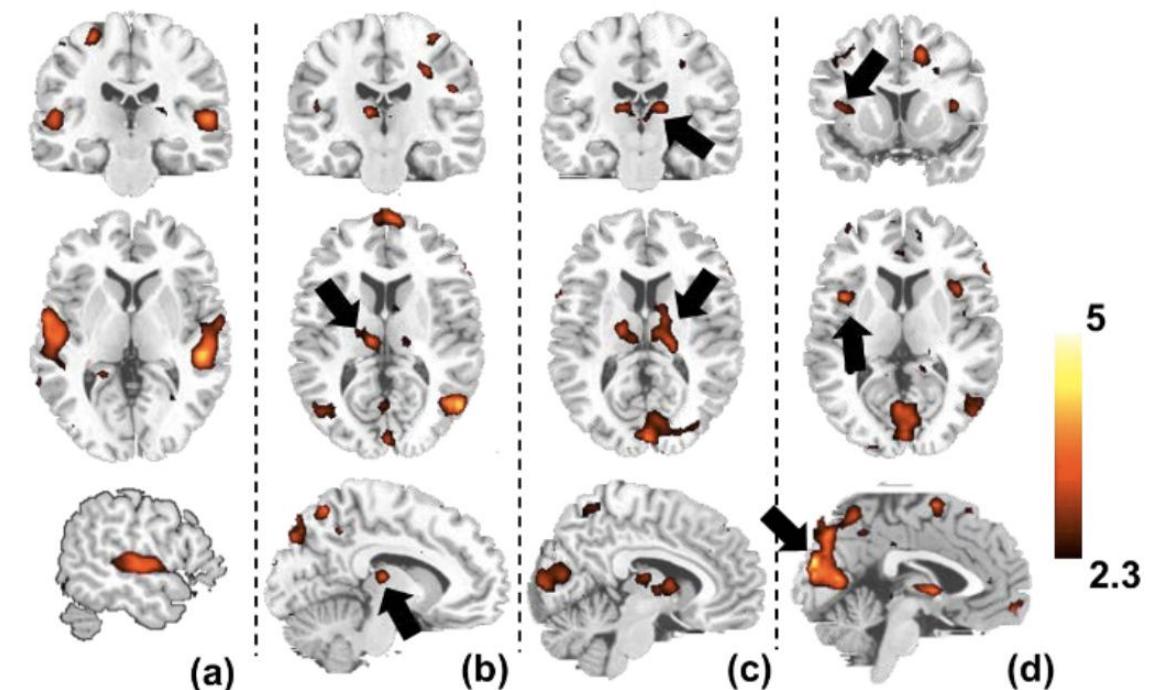


**Fig. 2.** A schematic visualization of the architecture of the DACE neural network. (a) DCAE structure, the hidden states, and the feature maps. (b) The prediction process of fMRI signals in encoder. (c) The detailed processes of the max pooling and unpooling.

# Deep learning for predicting neural signals



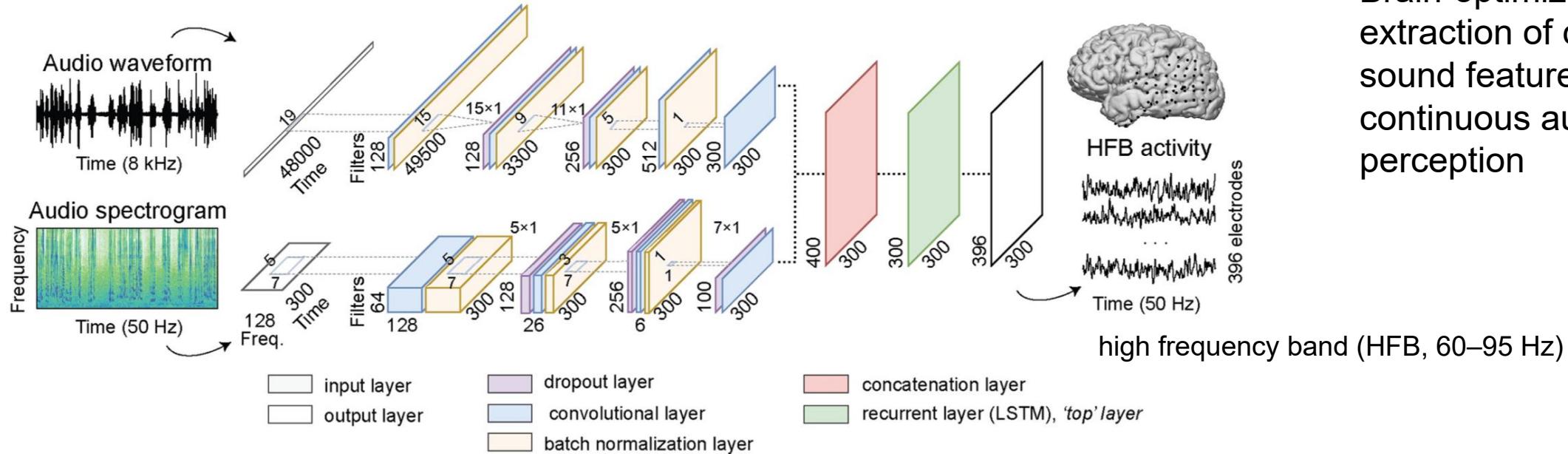
**Fig. 1.** The framework of this study. (a) Hierarchical feature extraction of audio excerpts via DCAE. (b) FMRI acquisition. (c) FMRI encoding models based on LASSO.



**Fig. 4.** The Z-maps in group analysis for each encoder layer.

In the first layer (Fig. 4a), brain activities in the **primary auditory cortex** can be encoded with significantly ( $Z \geq 2.3$ ) high encoding accuracy, indicating that the features learned in the first layer may represent basic acoustic attributes.  
In the fourth layer, **bilateral insula** and **ventral visual cortex** are with significantly high encoding accuracy (Fig. 4d).

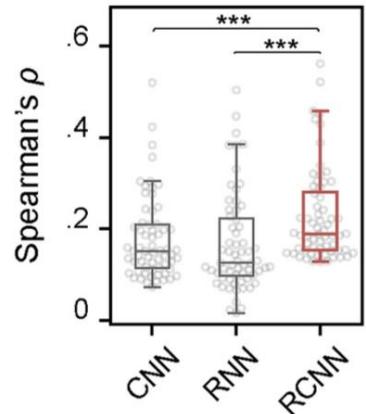
## Architecture of brain-optimized (artificial) neural network (BO-NN)



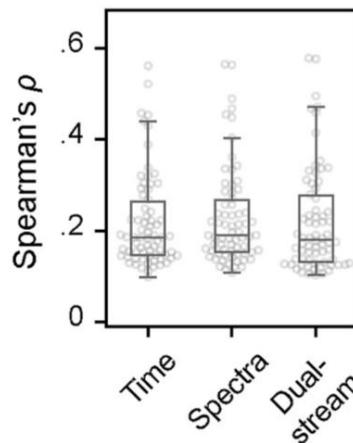
Brain-optimized extraction of complex sound features that drive continuous auditory perception

### Testing various architectural choices

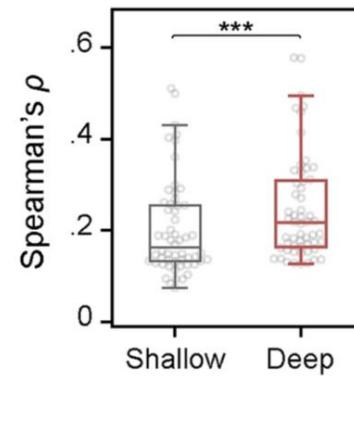
Computations within main nodes



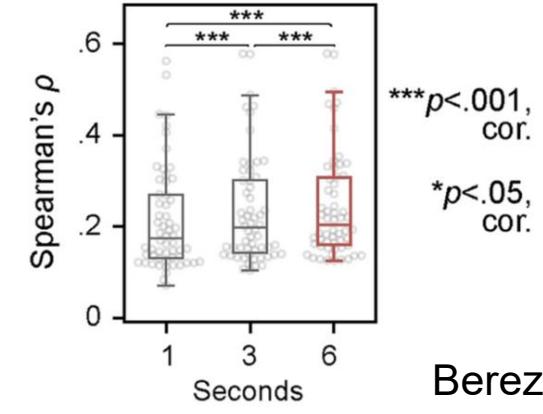
Type of input to RCNN



Depth of convolutional block in RCNN



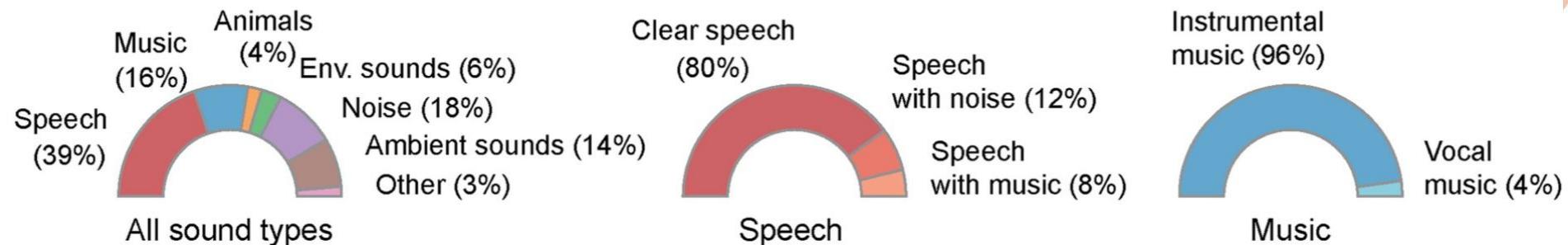
Size of temporal window for RCNN



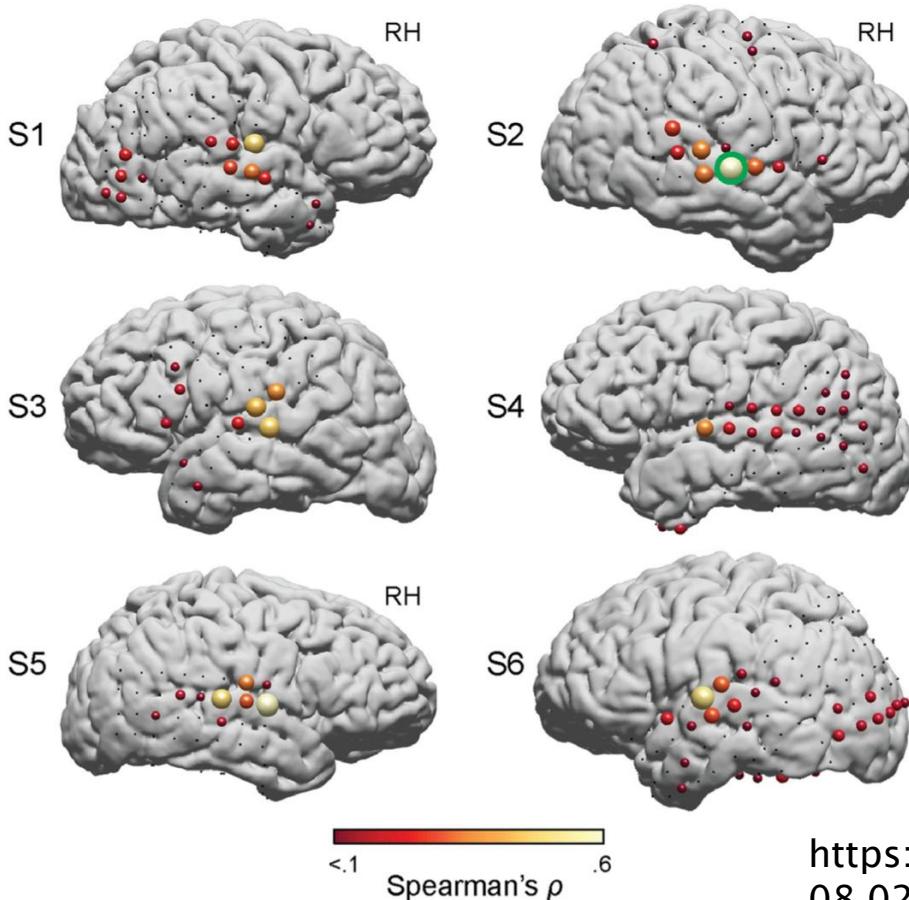
\*\*\* $p<.001$ , cor.  
\* $p<.05$ , cor.

**a**

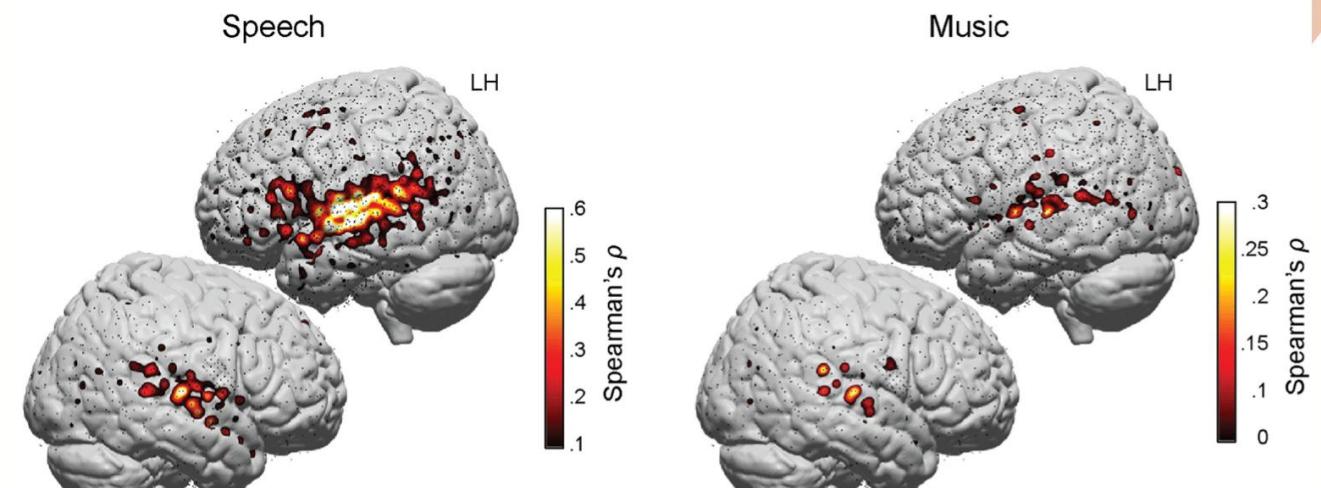
### Distribution of sound types in soundtrack of Movie I



### Cortical maps of BO-NN prediction accuracy



### BO-NN performance in Movie II



Berezutskaya et al. (2020) *PLoS Comp. Biol.*

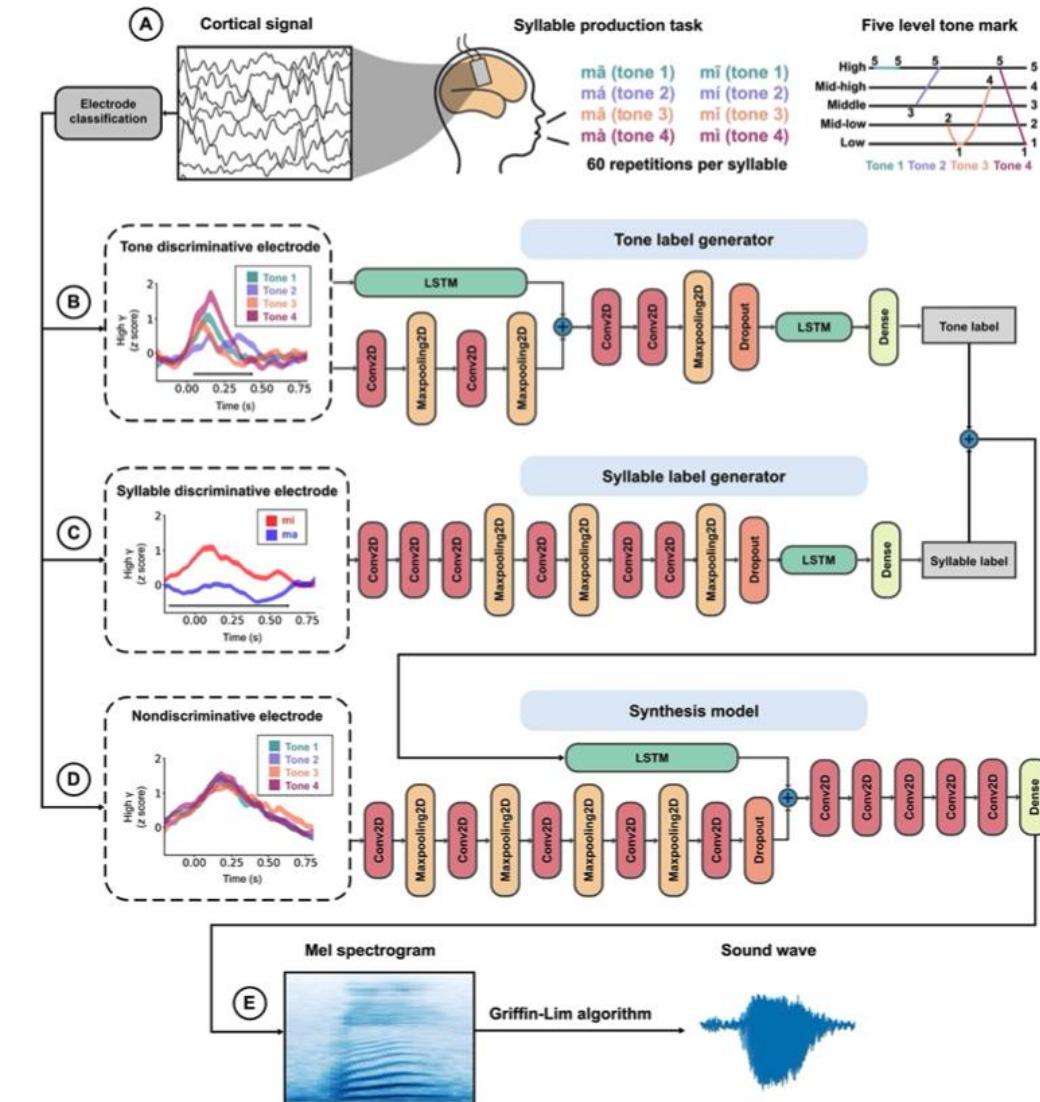
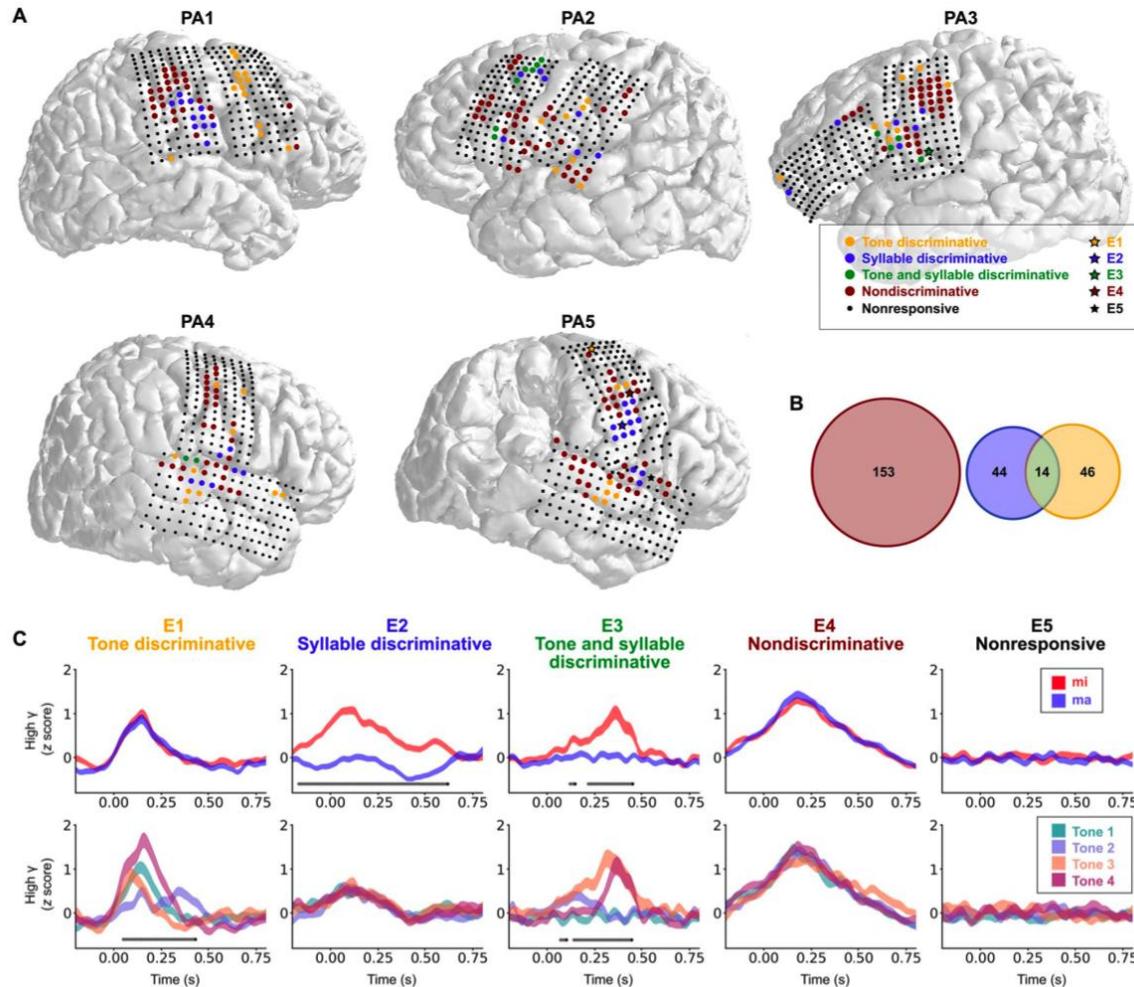
**Q2:**

**Can we reconstruct the sound  
based on neural signals?**

neural decoding:  
neural activity → sound

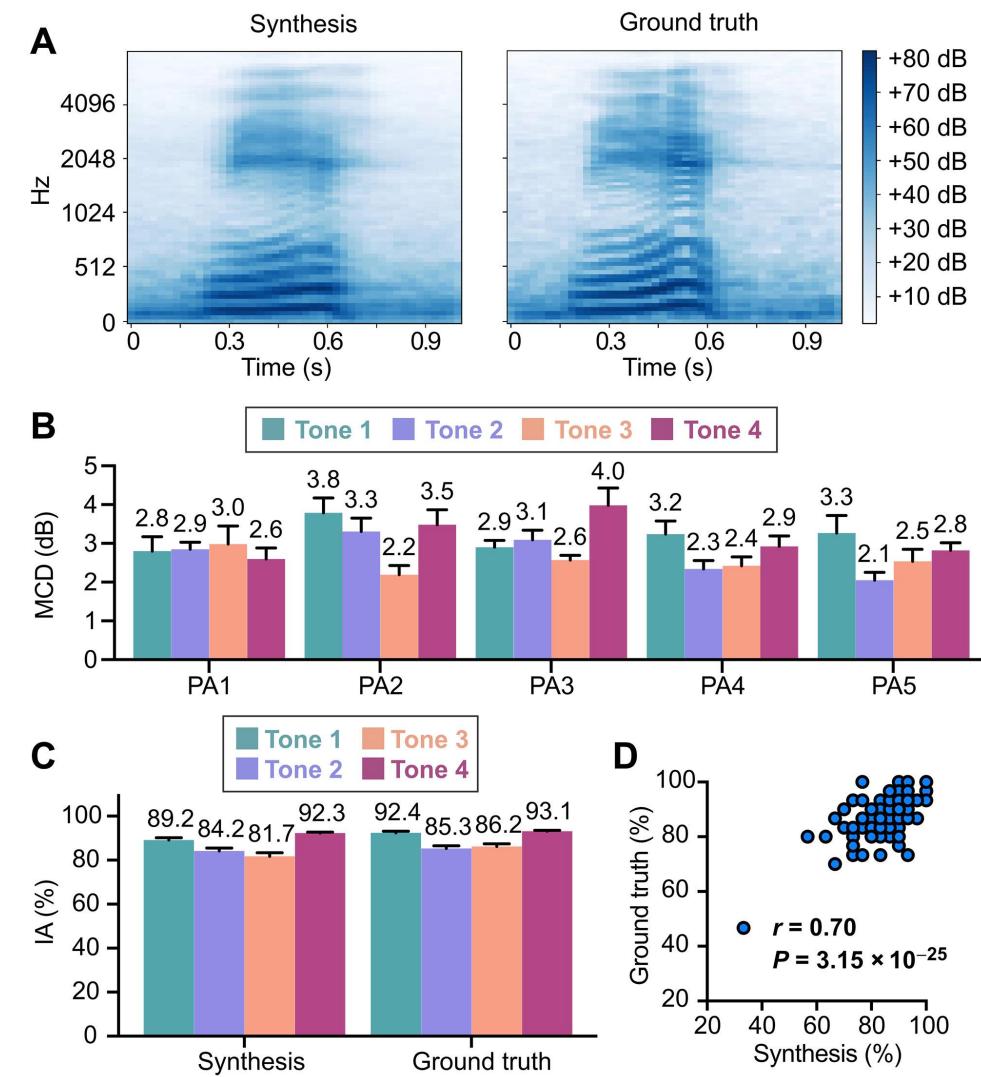
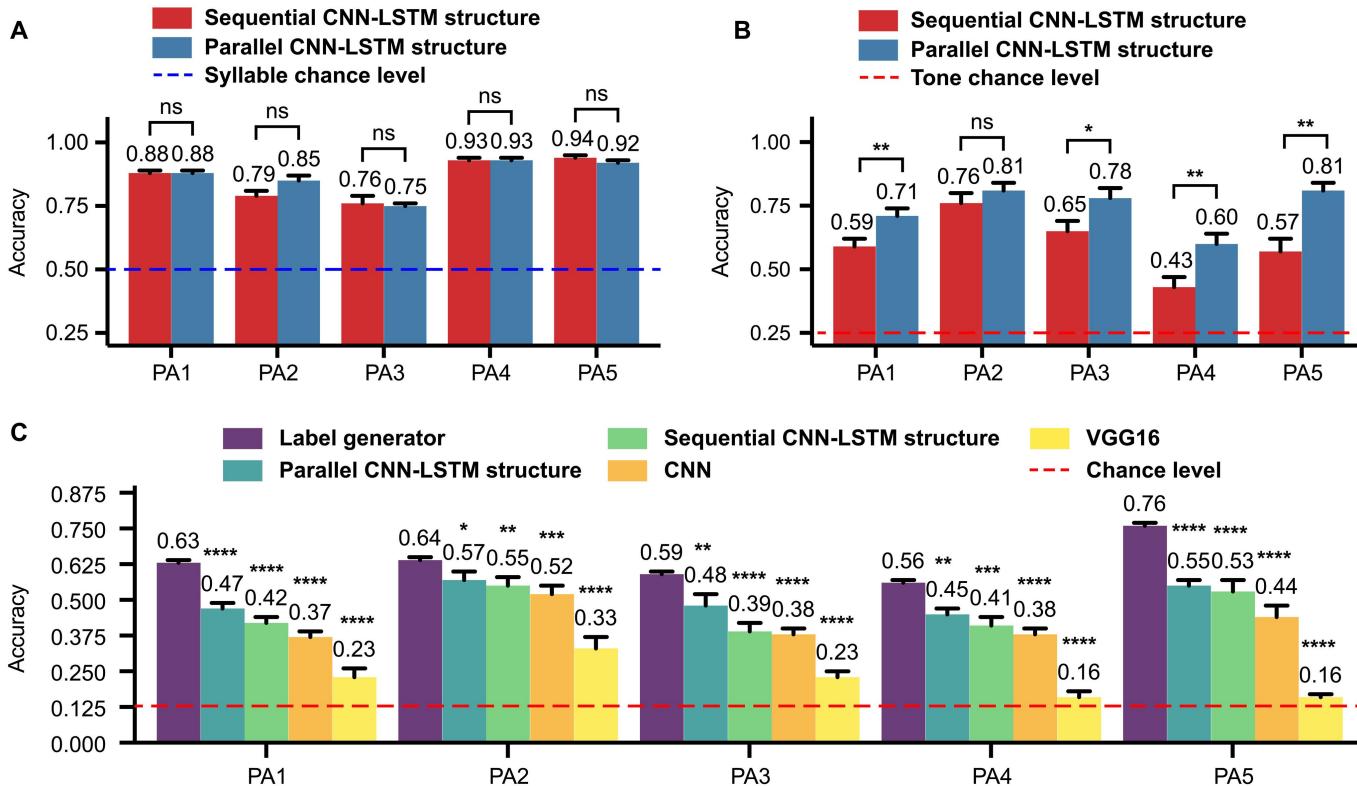
# Deep learning for decoding tonal language speech

mi: 咪、迷、米、密  
 ma: 妈、麻、马、骂



# Deep learning for decoding tonal language speech

从大脑信号解码语音

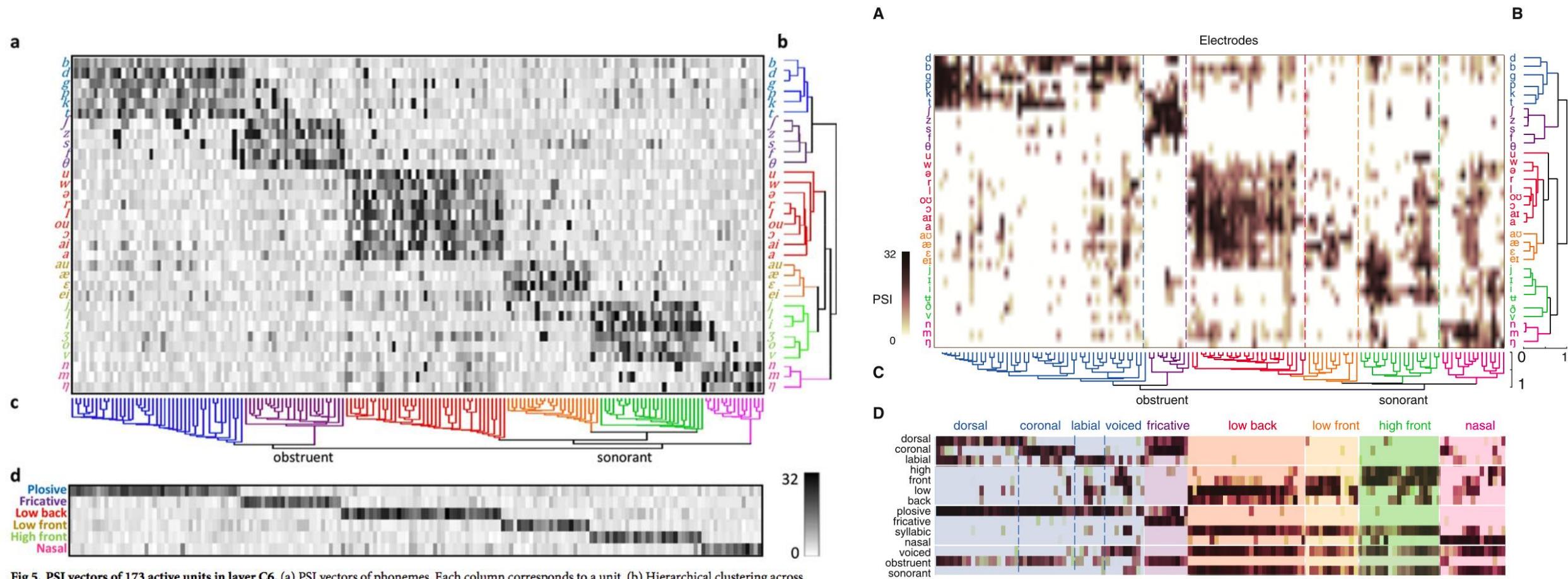


**Q3:**

**Can AI models help us understand  
the neural mechanism of audition?**

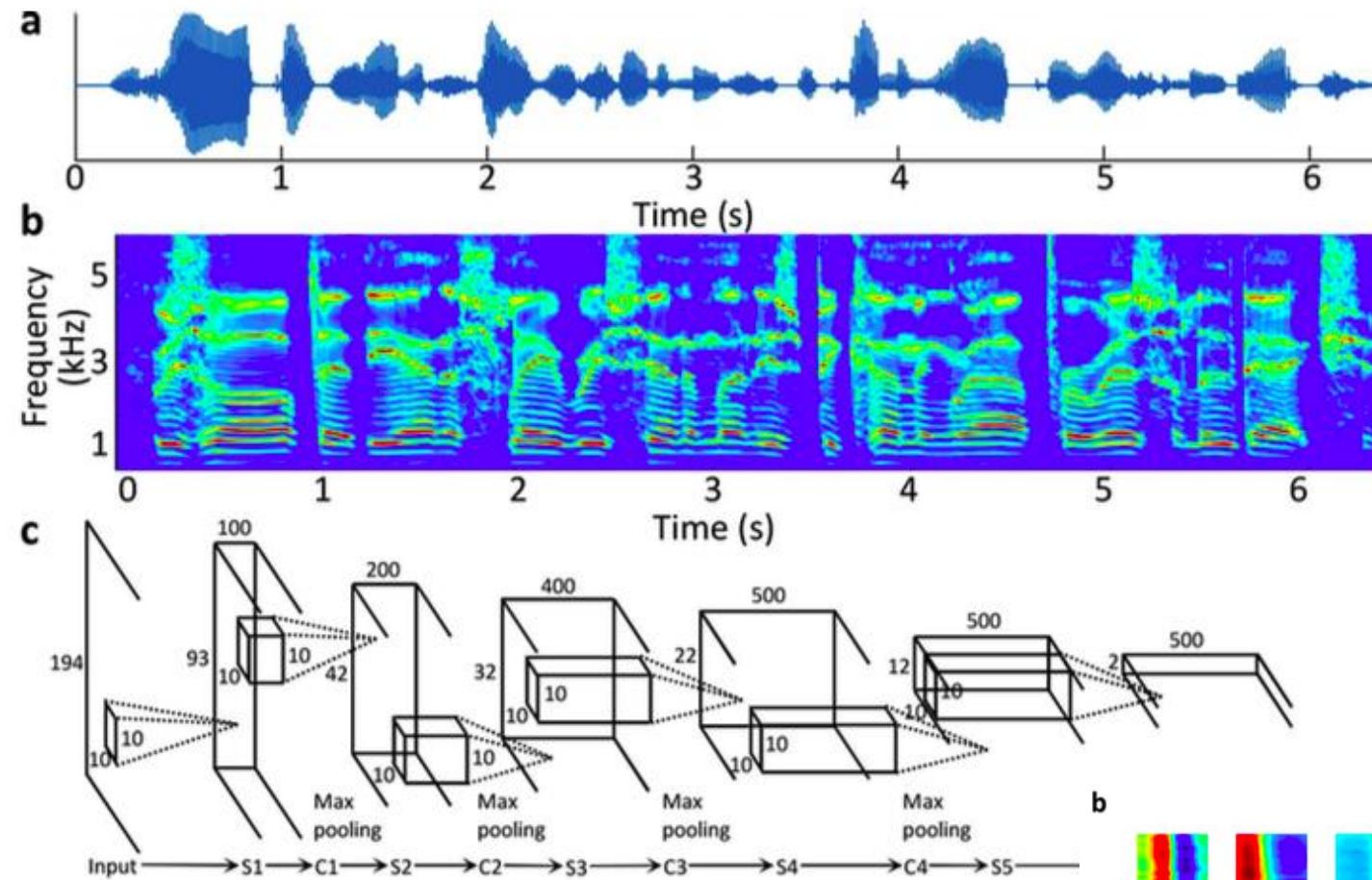
model interpretability

# Deep learning for studying the **hierarchy** of human audition

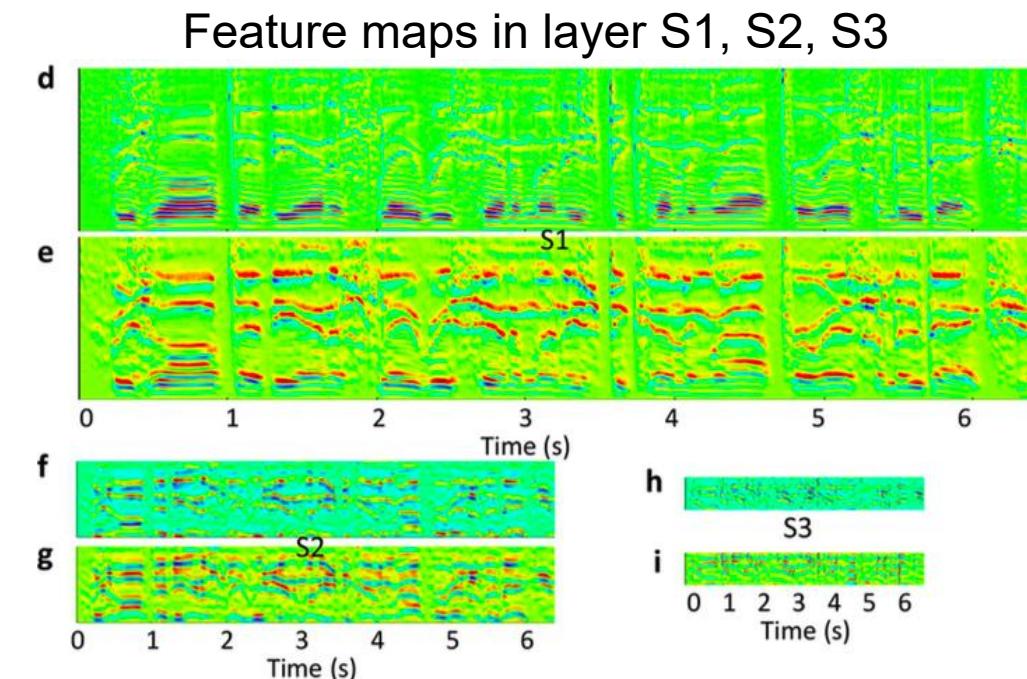


**Fig 5. PSI vectors of 173 active units in layer C6.** (a) PSI vectors of phonemes. Each column corresponds to a unit. (b) Hierarchical clustering across phonemes. (c) Hierarchical clustering across units. (d) PSI vectors of six phonetic features.

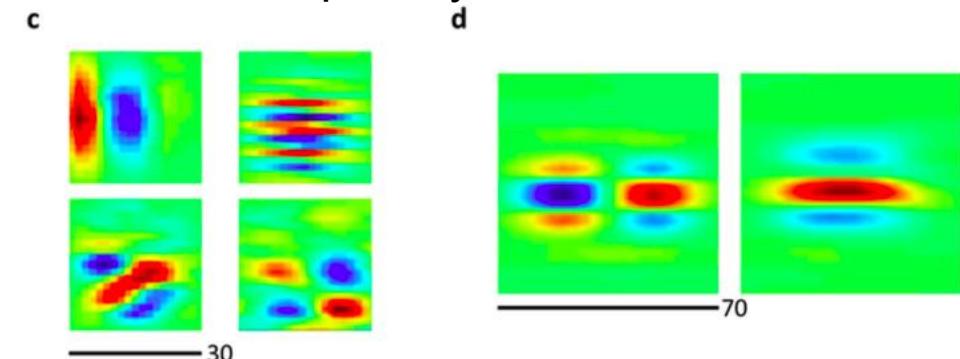
# Deep learning to uncover the **sparse** neural representation



Zhang et al (2019), PLoS Comp. Biol.



STRF map in layer S1, S2, S3



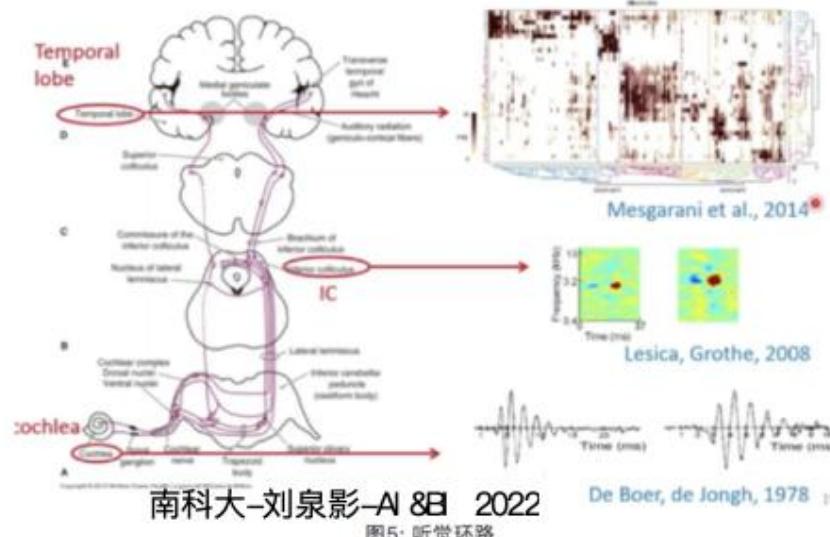
# 讲座总结 | 利用深度学习模型研究神经元稀疏发放放在视听觉通路的作用

原创 NCC lab 神经计算与控制实验室 2020-05-21 10:36



“受南方科技大学生物医学工程系刘泉影助理教授邀请，来自清华大学计算机系胡晓林副教授 (<http://www.xlhu.cn/>) 于腾讯会议平台在线讲座，通过深度学习模型对脑的视听觉环路建立计算模型这两个案例，阐释了：1) 利用层次化结构的深度学习计算模型能够很好地展现视觉、听觉的高层皮层神经元的功能特点，2) 神经元的稀疏发放与神经元表征的分层特异性密不可分。同时，本讲座还回顾了脑智能 (Brain Intelligence) 和人工智能 (Artificial intelligence) 之间是如何相互促进发展的。”

## Response properties of auditory neurons



## Lower layer units resemble auditory midbrain neurons

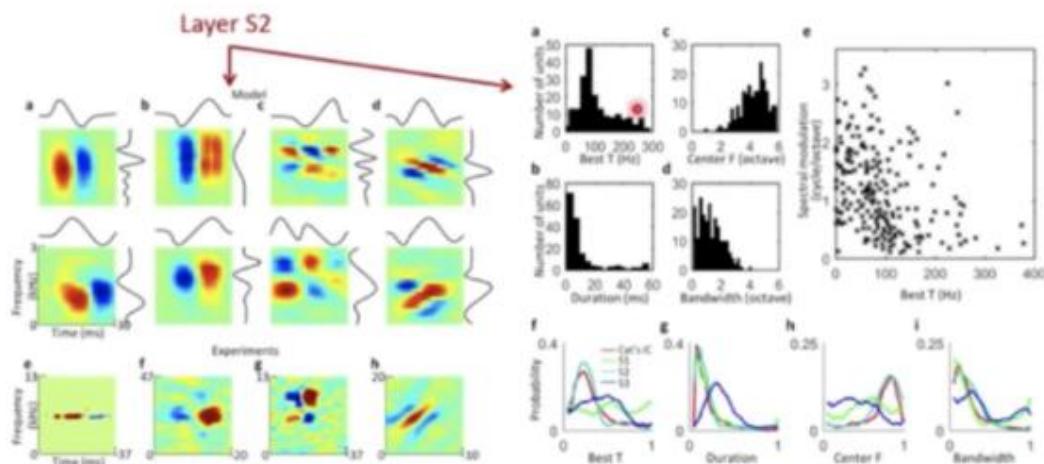


图7: 听觉深度学习计算模型 S2层的神经元和猫下丘的神经元的统计结果

微信扫一扫  
关注该公众号

[https://mp.weixin.qq.com/s?\\_biz=MzI1MTc1MjQxMg==&mid=2247484492&idx=1&sn=638474cd85c0f4a0fbb2df8bc01d97f0&chksm=e9ef7effde98f7e925dc5ab48a1cb034220ee27c85867498a5a27e6cfab6a69e8f3f187c396&token=80294869&lang=zh\\_CN#rd](https://mp.weixin.qq.com/s?_biz=MzI1MTc1MjQxMg==&mid=2247484492&idx=1&sn=638474cd85c0f4a0fbb2df8bc01d97f0&chksm=e9ef7effde98f7e925dc5ab48a1cb034220ee27c85867498a5a27e6cfab6a69e8f3f187c396&token=80294869&lang=zh_CN#rd)

# Prompt to chatGPT 4o: How to use deep learning to understand the auditory processing in brain?

Using deep learning to understand auditory processing in the brain involves a combination of neural signal analysis, model architecture selection, and interpretation of biological data. Here's a step-by-step approach:

## ### 1. \*\*Collect and Preprocess Auditory Data:\*\*

- **Neuroimaging Data**: The primary sources of data include fMRI, EEG, or MEG, which capture brain activity while subjects are exposed to different auditory stimuli (e.g., speech, music, tones).
- **Preprocessing**: Preprocess the data to remove noise and artifacts. For EEG or MEG, this includes filtering, epoching, and artifact rejection, while for fMRI, preprocessing involves motion correction, normalization, and smoothing.

## ### 2. \*\*Feature Extraction:\*\*

- **Auditory Stimuli Features**: Extract meaningful features from the auditory input (e.g., frequency, amplitude, pitch). Time-frequency analysis methods like Fourier or wavelet transforms are commonly used.
- **Neural Features**: Use methods like PCA or ICA for dimensionality reduction in neuroimaging data, and time-series analysis techniques to capture neural dynamics related to auditory processing.

## ### 3. \*\*Select a Model Architecture:\*\*

- **Recurrent Neural Networks (RNNs)**: Since auditory processing is sequential and time-dependent, RNNs, especially LSTM or GRU, are suited for modeling how the brain processes auditory stimuli over time.
- **Convolutional Neural Networks (CNNs)**: For spatial data, such as fMRI, or for frequency analysis, CNNs can be effective in detecting patterns across brain regions.
- **Transformer Models**: These can be used for more advanced models that capture long-range dependencies in neural time series or auditory inputs.

## ### 4. \*\*Train the Model:\*\*

- **Supervised Learning**: Use labeled datasets where the auditory stimuli and corresponding brain responses are paired, allowing the model to learn associations between them.
- **Unsupervised Learning**: Use techniques like autoencoders to learn latent representations of neural activity related to auditory processing without specific labels.

## ### 5. \*\*Model Interpretability:\*\*

- **Attention Mechanisms**: Apply attention layers to highlight the time points or brain regions most relevant to auditory stimuli.
- **Neural Activation Maps**: For CNNs, use techniques like Grad-CAM to visualize which parts of the brain are activated by specific auditory inputs.
- **Latent Representation Analysis**: Examine the learned embeddings from unsupervised learning models to understand the structure of auditory processing in the brain.

## ### 6. \*\*Relate Model Outputs to Neuroscience:\*\*

- **Compare to Neural Circuits**: Analyze whether the model's learned patterns correspond to known auditory pathways in the brain, such as the auditory cortex or brainstem.
- **Temporal Dynamics**: Investigate whether the model captures the temporal dynamics of auditory processing, such as phase locking or frequency tuning.
- **Cognitive Insights**: Use the model to explore higher-level auditory processes, like speech comprehension, music appreciation, or auditory scene analysis.

## ### 7. \*\*Apply to Cognitive or Clinical Studies:\*\*

- **Speech Processing**: Use deep learning models to study how the brain processes speech and language, linking brain activity to specific linguistic features.
- **Hearing Disorders**: Apply these models to understand abnormalities in auditory processing in conditions like hearing loss, tinnitus, or auditory processing disorders.

This approach combines neural data analysis with deep learning techniques to enhance the understanding of how the brain processes auditory stimuli, potentially leading to insights into both basic neuroscience and clinical applications.

If I want to publish in *Nature Neuroscience*, what scientific questions can be studied along this line?

Publishing in *Nature Neuroscience* requires addressing significant, cutting-edge scientific questions that advance our understanding of brain function. Along the line of using deep learning to understand auditory processing in the brain, here are several potential research questions that would likely attract attention in the field:

### 1. How does the brain encode complex auditory environments?

- **Scientific Question:** How does the brain process multiple overlapping auditory stimuli, such as speech in a noisy environment, and how can deep learning models help disentangle this processing at the neural level?
- **Potential Contribution:** Develop deep learning models that simulate auditory scene analysis in real-world conditions and investigate how these models relate to neural mechanisms in the auditory cortex. Insights could lead to better understanding of how the brain separates signals in complex auditory environments.
- **Relevance:** Understanding how the brain handles competing auditory signals is crucial for fields like hearing aid development and speech recognition technologies.

### 2. What are the neural representations of speech and music in the brain?

- **Scientific Question:** How does the brain differentiate and represent speech versus music at the neural level, and can deep learning help decode these neural representations?
- **Potential Contribution:** Use deep learning models to analyze neural recordings (fMRI, EEG) while participants listen to both speech and music, identifying key neural patterns or regions responsible for each type of auditory processing. This could reveal insights into specialized neural pathways for different types of complex auditory stimuli.
- **Relevance:** This research could help elucidate the evolution of language and music processing in the human brain and has potential applications in auditory prosthetics and speech-language therapies.

### 4. What are the neural mechanisms of auditory attention?

- **Scientific Question:** How does the brain selectively attend to specific auditory stimuli in complex environments, and can deep learning models simulate this selective attention process?
- **Potential Contribution:** Build models that mimic the brain's ability to focus on certain sounds (e.g., a person's voice) while ignoring others (e.g., background noise). Use these models to understand neural circuits involved in auditory attention, particularly in the prefrontal cortex and auditory cortex.
- **Relevance:** This work would have profound implications for understanding how attention is modulated in the brain and could inform clinical interventions for disorders like ADHD or schizophrenia, where auditory attention is often impaired.

### 5. How does auditory learning reshape neural circuits?

- **Scientific Question:** How does learning new auditory skills (e.g., a new language or musical instrument) reorganize the neural circuits involved in auditory processing, and can deep learning models predict such neuroplastic changes?
- **Potential Contribution:** Apply deep learning models to study brain plasticity in auditory regions during auditory learning tasks. Analyze how learning changes the representation of sounds in the brain over time and whether deep learning models can track and predict these changes.
- **Relevance:** Understanding neuroplasticity in auditory learning could inform education strategies and rehabilitation approaches for individuals with auditory processing deficits.

## **8. How does the brain process auditory stimuli during sleep or unconscious states?**

- **Scientific Question:** How is auditory information processed during sleep or states of unconsciousness, and what role does it play in memory consolidation or other cognitive functions?
- **Potential Contribution:** Use deep learning to decode neural activity in response to auditory stimuli during sleep, exploring whether specific sounds influence memory consolidation or other brain processes during different sleep stages.
- **Relevance:** This research could lead to new insights into sleep-related learning and memory mechanisms, as well as potential clinical applications for sleep disorders or auditory training during sleep.

## **9. Can we build brain-machine interfaces (BMIs) for auditory prosthetics based on neural decoding models?**

- **Scientific Question:** How can deep learning models improve auditory brain-machine interfaces, such as cochlear implants, by decoding neural signals more accurately and facilitating better auditory perception?
- **Potential Contribution:** Develop and test deep learning models for decoding neural responses to auditory stimuli in real-time, with the goal of improving brain-machine interfaces for individuals with hearing loss. This research could explore the feasibility of directly interfacing with the brain's auditory regions to restore or enhance auditory perception.
- **Relevance:** Advancing BMI technology would significantly impact the quality of life for individuals with severe auditory impairments.

Each of these questions taps into the core areas of neuroscience—perception, cognition, neuroplasticity, brain connectivity, and disorders—while integrating advanced deep learning methods to offer new perspectives on auditory processing. Given *Nature Neuroscience*'s focus on groundbreaking research, addressing one of these questions with novel techniques and robust data would be highly relevant for publication.

Can the mechanism of auditory system improve AI models? If I want to publish in nature machine intelligence, what research question should I study along this line?

Yes, the mechanisms of the auditory system can inspire and improve AI models, particularly in areas related to processing complex, noisy, and sequential data, which are challenges both in biological and artificial systems. Drawing insights from how the brain's auditory system processes sounds could lead to innovations in AI architectures, particularly for tasks such as speech recognition, sound classification, auditory scene analysis, and more.

Here are several potential research questions that could be explored for publication in *Nature Machine Intelligence*:

## 1. Can biologically-inspired auditory processing models outperform traditional deep learning models in real-world noisy environments?

- **Scientific Question:** How can incorporating mechanisms from the human auditory system, such as auditory filtering, adaptive noise suppression, and temporal coding, improve AI models' robustness to noise in speech or sound recognition tasks?
- **Potential Contribution:** Develop AI models that mimic key aspects of the auditory system, such as cochlear filtering, the tonotopic organization of the auditory cortex, or temporal precision in spike timing. Compare the performance of these models with traditional AI models under real-world noisy conditions.
- **Relevance:** This could lead to significant advances in robust AI systems for speech recognition, hearing aids, or noise-resistant audio-based devices.

## 2. Can AI models replicate the hierarchical organization of auditory processing in the brain for multi-scale feature extraction?

- **Scientific Question:** How can AI architectures simulate the hierarchical nature of auditory processing (from early auditory processing of basic features to higher-order cortical processing) to improve multi-scale feature learning?
- **Potential Contribution:** Design hierarchical deep learning models that replicate the multi-stage processing observed in the auditory system, where low-level features (such as frequency and amplitude) are processed first, followed by the extraction of more abstract patterns (like speech or music).
- **Relevance:** Such models could provide better insights into multi-scale representation learning and significantly improve AI applications in fields like music information retrieval, speech processing, and sound event detection.

## ## 3. \*\*Can biologically-inspired temporal dynamics of the auditory system enhance AI's ability to process sequential data?\*\*

- **Scientific Question:** How can the brain's temporal encoding and phase-locking mechanisms be adapted to improve AI models for processing sequential auditory data (such as speech or music)?
- **Potential Contribution:** Develop AI models that incorporate temporal coding strategies, such as phase locking, observed in the auditory system. This could lead to models that handle temporal dependencies more effectively, especially in sequential tasks like speech-to-text, audio generation, or music composition.
- **Relevance:** Time-based models inspired by auditory mechanisms could outperform traditional models (e.g., RNNs, LSTMs) in handling long-range dependencies, potentially contributing to breakthroughs in AI's ability to process sequential data.

## ## 4. \*\*How can the auditory system's attention mechanisms inspire AI models for auditory stream segregation?\*\*

- **Scientific Question:** How can insights from the brain's auditory attention system (which enables selective attention to specific sounds in complex environments) improve AI models for separating and focusing on individual sound sources?
- **Potential Contribution:** Design AI models that incorporate selective attention mechanisms, inspired by the brain's ability to focus on a single auditory stream in a crowded or noisy environment. These models could outperform traditional sound separation techniques, such as blind source separation, by better mimicking how humans track sounds in challenging acoustic environments.
- **Relevance:** Advances in this area could improve applications in fields such as auditory scene analysis, speech enhancement, or audio-visual integration in robotics.

## ## 5. \*\*Can auditory brain-inspired sparse coding improve efficiency and interpretability in AI models?\*\*

- **Scientific Question:** How can the sparse coding strategies employed by the brain's auditory cortex (where only a small number of neurons are activated in response to sounds) inspire more efficient and interpretable AI models?
- **Potential Contribution:** Develop AI models that adopt biologically-inspired sparse coding techniques, reducing redundancy and enhancing efficiency. This approach could lead to AI models that are not only more efficient in terms of computational cost but also more interpretable, as the representations they learn might be more aligned with meaningful auditory features.
- **Relevance:** Sparse coding models could have far-reaching implications for energy-efficient AI systems, particularly for edge computing and AI deployed in low-power devices such as smartphones or hearing aids.

## ## 6. \*\*Can the auditory system's plasticity mechanisms inspire adaptive AI models for dynamic environments?\*\*

- **Scientific Question:** How can neural plasticity in the auditory cortex (the brain's ability to adapt to new auditory experiences or recover after hearing loss) inspire AI models that can adapt to changing environments or learn from few examples?
- **Potential Contribution:** Develop AI models that incorporate mechanisms akin to auditory system plasticity, such as dynamic rewiring or adaptive learning rates. These models could adapt more quickly to new environments, handle non-stationary data, or achieve better performance with limited data.
- **Relevance:** Plasticity-inspired models could lead to breakthroughs in AI for dynamic environments, where the ability to adapt rapidly to new or evolving conditions is critical (e.g., autonomous systems or personalized AI).

## ## 7. \*\*How can insights from auditory perception of emotions improve affective computing systems?\*\*

- **Scientific Question:** How can the brain's mechanisms for perceiving emotional cues in speech (prosody, tone) be used to improve AI models for emotion recognition from auditory data?
- **Potential Contribution:** Design deep learning models that simulate how the brain decodes emotional information from auditory features such as pitch, tone, and rhythm. Such models could enhance the performance of AI in emotion

### **### 8. \*\*Can auditory perceptual invariance inspire AI models for cross-domain sound recognition?\*\***

- **Scientific Question**: How does the auditory system achieve perceptual invariance (recognizing the same sound across varying contexts or distortions), and how can this inspire AI models for cross-domain or robust sound recognition?
- **Potential Contribution**: Develop AI models inspired by the auditory system's ability to maintain robust sound recognition across different environments, distortions, or speaker variations. These models could better handle real-world variability and achieve improved performance in cross-domain sound recognition tasks (e.g., recognizing speech across different languages or accents).
- **Relevance**: Perceptual invariance-inspired models could lead to more generalized and robust AI systems, overcoming challenges in cross-domain adaptation, transfer learning, and domain generalization.

### **### 9. \*\*How can the interplay between auditory and visual systems inspire multi-modal AI models?\*\***

- **Scientific Question**: How does the brain integrate auditory and visual information to form coherent perceptions of the environment, and can this inspire AI models for multi-modal processing?
- **Potential Contribution**: Develop multi-modal AI models that integrate auditory and visual data in a way that mimics the brain's ability to combine these sensory modalities for improved perception and decision-making. These models could surpass current AI systems in tasks that require understanding both sound and vision, such as audio-visual speech recognition or scene understanding.
- **Relevance**: Multi-modal AI models could have broad applications in fields like autonomous driving, robotics, and assistive technologies, where integrating multiple sensory inputs is critical for real-world performance.

### **### 10. \*\*How can deep learning models of auditory learning inform unsupervised learning in AI?\*\***

- **Scientific Question**: How does the brain learn to recognize new sounds without explicit supervision, and how can this process inspire more effective unsupervised learning techniques in AI?
- **Potential Contribution**: Investigate how unsupervised or self-supervised learning in AI can be enhanced by mimicking auditory learning mechanisms, such as the brain's ability to cluster and categorize new sounds based on exposure without explicit labels.
- **Relevance**: Insights from unsupervised auditory learning could lead to better AI algorithms for tasks where labeled data is scarce, improving systems in areas like speech recognition, sound classification, and even language modeling.

By focusing on these cutting-edge questions that intersect neuroscience and AI, you can contribute both to the understanding of biological auditory processing and to the advancement of AI models. Research along these lines is highly relevant to \*Nature Machine Intelligence\*, as it addresses both the theoretical understanding of intelligent systems and their practical applications.

I would like to study the optimized neurostimulation to improve auditory perception in a closed-loop manner, how can i possibly do it?

Studying optimized neurostimulation to improve auditory perception in a **closed-loop** manner is an exciting and complex challenge. Closed-loop systems offer the ability to adapt stimulation in real-time based on brain activity, making them ideal for applications like enhancing auditory perception. Here's how you could potentially approach this research:

### Step-by-Step Approach to Study Optimized Neurostimulation for Auditory Perception:

---

#### 1. Define the Scientific Problem and Hypothesis

- **Scientific Question:** How can real-time neurostimulation be optimized to enhance auditory perception in humans? Specifically, how can a closed-loop system adapt to neural responses to auditory stimuli to improve perception?
  - **Hypothesis:** Adaptive, real-time neurostimulation can improve auditory perception by modulating neural circuits in the auditory cortex and related brain regions based on immediate neural feedback.
- 

#### 2. Understand the Neural Mechanisms of Auditory Perception

- **Neural Targets:** Identify brain areas involved in auditory perception, particularly the **auditory cortex** (e.g., primary and secondary auditory areas), **superior temporal gyrus**, and potentially **prefrontal cortex** (for attention-related modulation).
  - **Neural Biomarkers:** Determine what neural signals (e.g., EEG oscillations, auditory evoked potentials, or brain rhythms like the theta or gamma band) correlate with auditory perception or attention to auditory stimuli.
  - **Key Mechanisms:** Understand how neuroplasticity can be influenced by electrical stimulation, targeting mechanisms like **phase synchronization**, **neural entrainment**, or modulation of specific brain rhythms linked to auditory processing.
- 

#### 3. Design a Closed-Loop Neurostimulation System

- **Closed-Loop Concept:** In a closed-loop neurostimulation system, the brain's response to auditory stimuli is continuously monitored in real-time, and stimulation is adjusted dynamically based on this feedback to optimize auditory perception.
  - **Real-time Neural Sensing:** Use **EEG**, **MEG**, or **ECoG** (for invasive settings) to monitor brain activity related to auditory perception. Choose a signal that provides **high temporal resolution** since auditory processing involves fast neural dynamics.
  - **Stimulation Device:** Use non-invasive methods like **tACS** (**transcranial Alternating Current Stimulation**), **tDCS** (**transcranial Direct Current Stimulation**), or more advanced techniques like **transcranial Temporal Interference Stimulation (tTI)** or **deep brain stimulation (DBS)** (if allowed in clinical contexts) to stimulate auditory-related brain regions.
    - **tACS** is particularly interesting because it can entrain brain oscillations to enhance synchrony during auditory perception tasks.
  - **Real-time Feedback Control:** Develop an algorithm that processes the neural feedback in real time and adjusts the stimulation parameters (frequency, intensity, location) to optimize auditory performance. Machine learning or adaptive control algorithms can be used to fine-tune the stimulation.
    - Example: If EEG signals indicate a drop in attentional focus during a listening task, the system could adjust stimulation to restore focus by modulating alpha or beta waves.
- 

#### 4. Develop the Experimental Protocol

- **Auditory Tasks:** Design auditory tasks that challenge perception and attention, such as:
    - **Speech in Noise Tasks:** Present speech in noisy environments and assess how well subjects can understand it with and without stimulation.
    - **Complex Sound Identification:** Use musical or environmental sounds to see how neurostimulation enhances recognition and discrimination abilities.
  - **Baseline Measurements:** Measure baseline performance in auditory tasks without stimulation and record the neural activity to establish a model of typical brain responses.
  - **Neurostimulation Protocol:** Implement neurostimulation during the tasks, adjusting the stimulation parameters based on real-time neural feedback.
    - **Stimulation Timing:** The timing of neurostimulation can be aligned with specific phases of auditory processing, such as the onset of a sound or during key brain rhythms (e.g., entraining gamma rhythms, which are linked to auditory perception).
- ↓
- 

#### 5. Machine Learning for Optimization

- **Learning from Feedback:** Use reinforcement learning or adaptive control algorithms that dynamically update the stimulation strategy based on brain signals and behavioral performance. The system should learn which stimulation patterns optimize the user's perception of sound.
    - **Reward Function:** Define a reward function that increases with improved auditory performance (e.g., better speech discrimination) and reduced error signals in the brain.
  - **Personalized Stimulation:** Each individual's brain responds differently to stimulation. Machine learning models can adapt to the neural patterns of each participant, personalizing stimulation strategies for optimal improvement.
- 

#### 6. Measure and Evaluate Auditory Performance

- **Behavioral Outcomes:** Quantify auditory perception improvements using standard tests, such as speech-in-noise recognition accuracy, reaction times, or subjective ratings of sound clarity.
  - **Neural Outcomes:** Analyze how neurostimulation affects neural markers of auditory processing. For example, look for changes in:
    - **Auditory Evoked Potentials (AEPs).**
    - **Oscillatory power** in specific frequency bands (e.g., enhanced theta or gamma band activity).
    - **Phase coherence** between brain regions involved in auditory processing.
  - **Comparative Studies:** Compare auditory performance with **no stimulation**, **open-loop stimulation**, and **closed-loop stimulation** to show the efficacy of the adaptive approach.
- 

#### 7. Address Clinical and Cognitive Applications

- **Hearing Loss:** Investigate whether optimized neurostimulation can improve auditory perception in individuals with **hearing impairments**. This could be particularly useful for those with **hearing aids** or **cochlear implants**, potentially serving as an adjunct to those devices.
- **Auditory Attention Disorders:** Explore applications in conditions where **auditory attention** is impaired, such as **ADHD**, **schizophrenia**, or **age-related auditory decline**.
- **Tinnitus:** Consider studying how neurostimulation could reduce or mask the perception of **tinnitus** by adjusting neural dynamics in auditory processing regions.

## 8. Ethical and Practical Considerations

- **Safety:** Ensure that the stimulation protocol is safe and does not lead to long-term changes in brain activity that could negatively impact cognitive or sensory functions.
  - **Invasiveness:** Use non-invasive techniques when possible. Invasive techniques may only be applicable in clinical settings (e.g., patients with epilepsy undergoing brain surgery).
  - **User Comfort:** Given that closed-loop systems operate in real time, ensure that the setup is comfortable and the auditory tasks/stimulation do not cause fatigue or discomfort.
- 

### Tools and Technologies:

- **Neural Recording:** High temporal resolution methods like **EEG** are ideal for capturing auditory-related neural dynamics.
- **Stimulation Techniques:** Non-invasive brain stimulation like **tACS** or **tDCS** for real-time modulation.
- **Closed-Loop Software:** Real-time processing platforms that integrate neural data acquisition, analysis, and stimulation control, such as **BCI** (Brain-Computer Interface) systems or custom-built control loops.
- **Machine Learning Frameworks:** Use libraries like **PyTorch** or **TensorFlow** to implement reinforcement learning or adaptive algorithms for stimulation control.

### Example Research Hypothesis for Publication:

"Optimized closed-loop neurostimulation enhances auditory perception by modulating auditory cortex oscillations in real-time, resulting in improved performance in speech recognition tasks under noisy conditions."

### Potential Outcomes and Impact:

- **Scientific Contribution:** Discover novel neural mechanisms that contribute to real-time modulation of auditory perception.
- **Practical Impact:** Develop adaptive stimulation techniques that could be used for clinical applications, such as improving hearing aids, enhancing cognitive focus in auditory tasks, or treating auditory attention disorders.

### Publication Strategy:

This research could be submitted to high-impact journals like **Nature Neuroscience** or **Nature Biomedical Engineering**, focusing on the neuroscience behind the closed-loop system. For engineering and AI-focused work, **Nature Machine Intelligence** could be appropriate if the focus is on developing adaptive algorithms for neurostimulation optimization.

I would like to study the optimized neurostimulation to improve auditory perception in a closed-loop manner, how can I possibly do it?

要研究优化神经刺激以在闭环系统中改善听觉感知，你可以从以下几个方面入手：

## 一、了解闭环神经刺激的基本原理

闭环神经刺激系统（如闭环迷走神经刺激术）能够监测患者的生理状态（如心率变化），并据此自动调节刺激的时间和强度。这种系统相较于传统的固定时间固定刺激模式，能更精准地把握最佳刺激时机，从而提高治疗效果。

## 二、查阅最新科研成果和临床案例

- 学术文献：通过PubMed、Google Scholar等学术搜索引擎，查找关于闭环神经刺激、听觉感知改善及大脑功能恢复的最新研究论文。特别是关注那些探讨了通过刺激特定脑区（如蓝斑位点LC）来改善听觉感知的研究。
- 临床案例：查找已成功实施闭环神经刺激术改善听觉感知的临床案例，了解手术过程、术后效果及患者反馈。这些案例可以为你提供宝贵的实践经验和参考依据。

## 三、选择适当的实验模型和方法

- 实验模型：在动物模型中（如大鼠）进行初步实验，以验证闭环神经刺激对听觉感知的改善效果。这些模型可以模拟人类的听觉障碍和神经刺激反应。
- 刺激参数优化：通过调整刺激强度、时间、频率等参数，观察不同参数组合对听觉感知改善效果的影响。利用先进的神经科学和神经工程技术，对刺激参数进行精细优化。
- 闭环控制算法：开发或优化闭环控制算法，以实时监测患者的生理状态（如心率、脑电活动等），并根据这些状态自动调节刺激参数。

## 四、收集和分析数据

- 行为学测试：通过行为学测试（如声音识别任务、音调辨别任务等）评估受试者的听觉感知能力。这些测试可以提供客观、量化的评估指标。
- 神经影像学分析：利用脑成像技术（如fMRI、EEG等）观察刺激过程中大脑活动的变化。这有助于揭示闭环神经刺激改善听觉感知的神经机制。
- 数据分析：对收集到的数据进行统计分析，比较不同刺激参数、不同时间点下的听觉感知改善效果。利用统计学方法（如ANOVA、t检验等）对数据进行分析和处理。

## 五、与多学科团队合作

- 神经科学家：与神经科学家合作，了解大脑听觉处理的机制和神经环路。
- 临床医生：与临床医生合作，了解听觉障碍患者的实际需求和治疗效果评估方法。
- 工程师：与生物医学工程师合作，开发或优化闭环神经刺激设备及其控制算法。

## 六、伦理和安全考虑

在研究过程中，必须严格遵守伦理规范和安全标准。确保所有受试者在充分知情同意的前提下参与研究，并采取措施保护受试者的隐私和安全。

综上所述，研究优化神经刺激以在闭环系统中改善听觉感知需要综合运用神经科学、生物医学工程、计算机科学等多学科知识，并通过严谨的实验设计和数据分析来验证假设和发现新的治疗方法。

沿着上面这个问题方向，如果我想发表在nature biomedical engineering，我应该怎么具体设计研究呢？

若您想在《Nature Biomedical Engineering》上发表关于优化神经刺激以在闭环系统中改善听觉感知的研究，您需要设计一个既创新又严谨的研究方案。以下是一些关键步骤和建议，以帮助您规划这一研究：

### 1. 确定研究目标和假设

- 明确研究目标：首先，您需要清晰地定义研究旨在解决的具体问题，例如“如何通过闭环神经刺激系统精确调节大脑活动，以改善特定人群的听觉感知能力？”
- 提出研究假设：基于现有文献和理论，提出一个或多个可验证的假设，如“在闭环控制下，特定频率和强度的电刺激能够显著提高听觉障碍患者的声音识别准确率。”

### 2. 文献综述与背景分析

- 全面回顾相关文献：梳理并总结前人在神经刺激、听觉感知、闭环控制等领域的研究成果，特别是与您的研究直接相关的最新进展。
- 分析研究空白与不足：指出当前研究中的不足之处或尚未解决的问题，明确您的研究将如何填补这些空白或改进现有方法。

### 3. 设计实验方案

- 选择实验模型：根据研究目标，选择合适的实验模型（如动物模型、细胞培养模型或人体临床试验）。如果可能，优先考虑使用更接近人类生理和病理状态的模型。
- 制定刺激方案：设计闭环神经刺激系统的具体参数（如刺激强度、频率、持续时间等），并确定如何根据实时生理反馈（如脑电活动、心率等）自动调节刺激参数。
- 设置对照组与实验组：确保实验设计具有足够的统计学效力，通过设立对照组来排除非特异性效应，并比较不同实验组之间的差异。
- 数据收集与分析方法：规划如何收集实验数据（如行为学测试、神经影像学数据等），并确定使用哪些统计方法进行分析。

### 4. 伦理与安全考虑

- 遵守伦理规范：确保研究符合国际和国内的伦理标准，特别是涉及人体试验时，必须获得伦理委员会的批准，并确保受试者的权益得到充分保护。
- 关注安全问题：在设计实验时，充分考虑可能的安全风险，并制定相应的应急预案以应对可能出现的意外情况。

### 5. 撰写研究计划与研究论文

- 研究计划：详细撰写研究计划，包括研究背景、目标、假设、方法、预期结果和伦理考虑等内容。这将有助于您在整个研究过程中保持清晰的方向和焦点。
- 研究论文：在研究完成后，根据期刊的投稿要求撰写研究论文。论文应结构清晰、逻辑严密、数据准确可靠，并充分讨论研究的局限性和未来研究方向。

### 6. 寻求合作与反馈

- 跨学科合作：由于该研究涉及神经科学、生物医学工程、计算机科学等多个领域，因此寻求跨学科合作将有助于提升研究的创新性和影响力。
- 专家咨询：在研究过程中，定期向相关领域的专家咨询，以获取宝贵的建议和指导。
- 同行评审：在投稿前，可以考虑将论文提交给同行进行评审，以获取反馈并改进论文质量。

请注意，《Nature Biomedical Engineering》作为一本高影响力的学术期刊，对论文的创新性、科学性和严谨性都有极高的要求。因此，在整个研究过程中，您需要保持高度的专注和严谨性，并不断努力提升研究的质量和水平。

# **Summary of Lecture 3 – Auditory System**

- **Sound:** frequency (Hz) & intensity (dB)
- **Ear**
  - Structure of human ear: outer ear, middle ear, inner ear
  - Cochlea; Organ of Corti; Hair cells
- **Central auditory process**
  - Auditory pathways: SG, (VCN, SO, IC, MGN), A1
  - Encoding sound intensity and frequency
  - Mechanisms of sound localization
  - Auditory cortex
- **Neural representation & plasticity in auditory cortex**
- **Deep learning for auditory processing:** encoding, decoding, interpreting