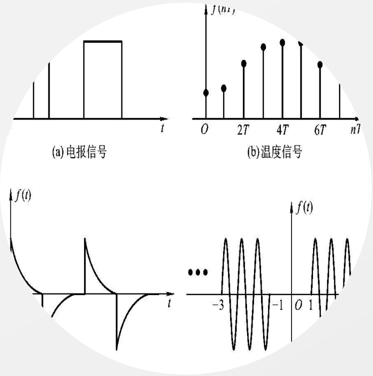
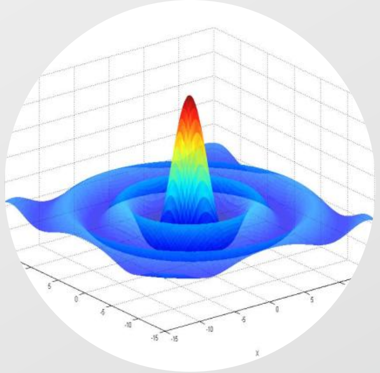
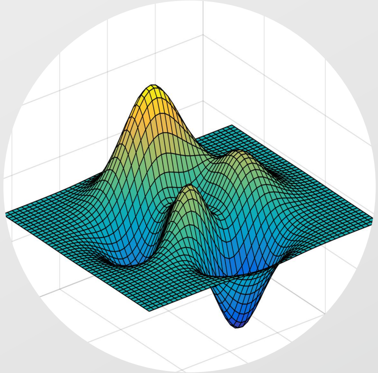
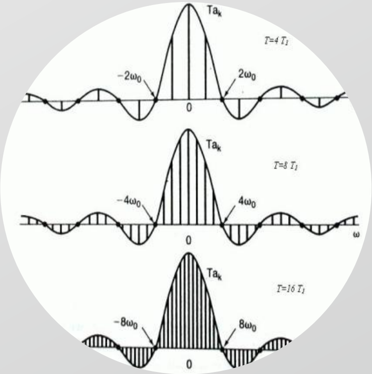








Project1：Speech Synthesis And Perception With Envelope Cue

主讲老师：王小静
办公地点：慧园2栋411

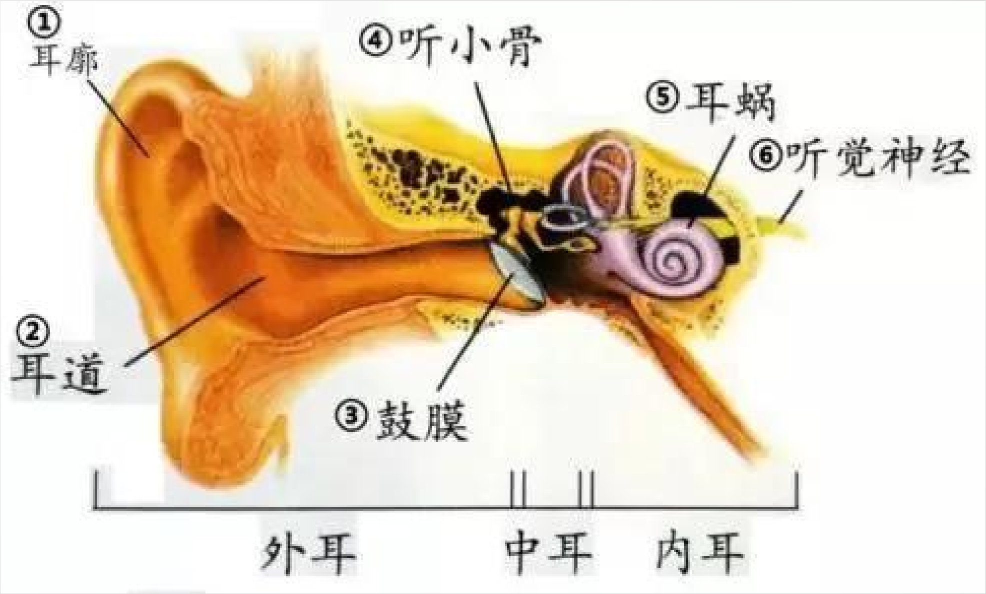
Overview

- In this tutorial, you will learn to synthesize a speech signal based on ***multi-band envelope cues***.
- In lab 5, you've learned:
  - how to design a low-pass/band-pass filter
  - how to extract envelope
  - how to generate a speech-spectrum shaped noise (SSN)
  - how to do energy normalization
  - how to read/save a '*.wav' file

Background: The principle of hearing

acoustic pressure waves → outer ear → eardrum(鼓膜)

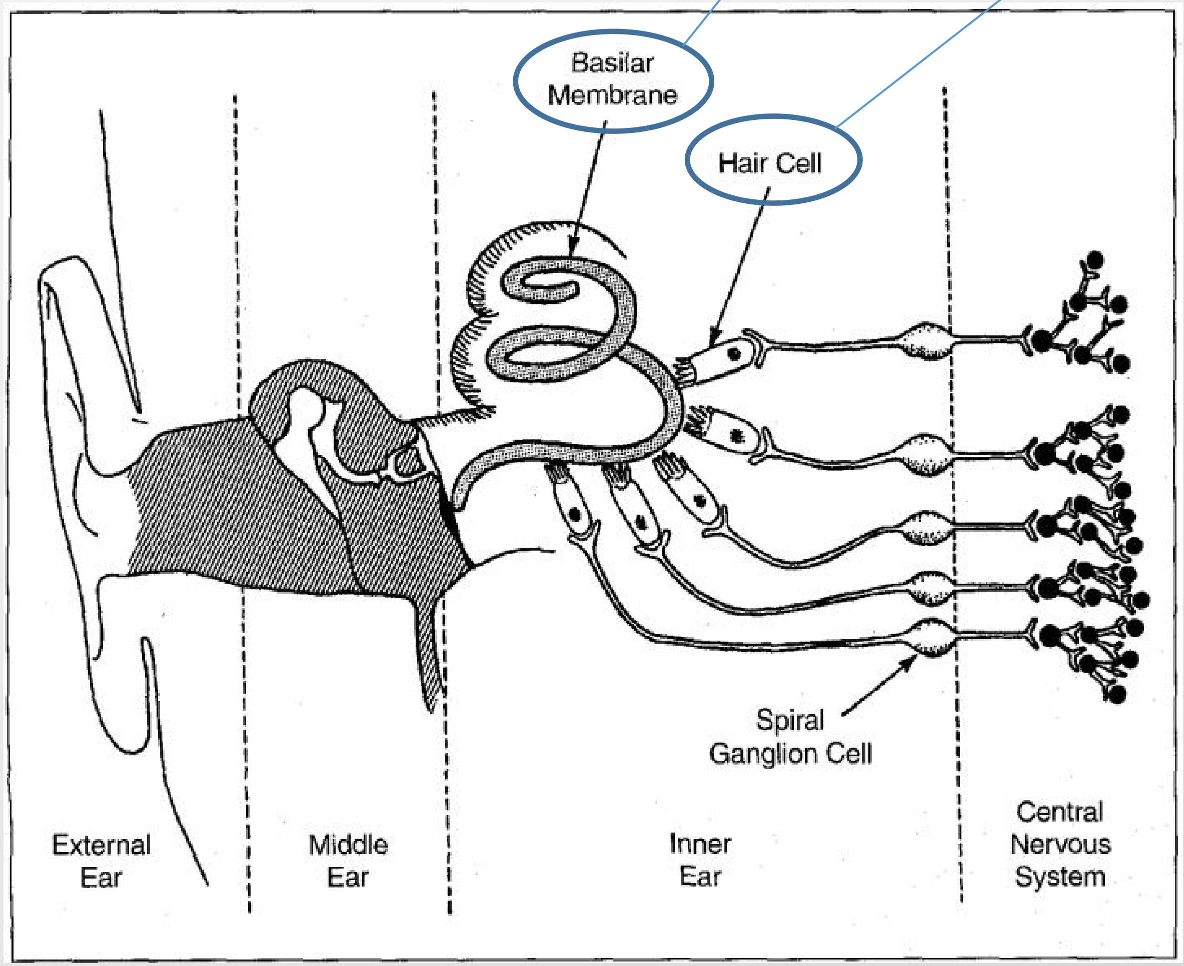
→ ossicles(听小骨) → cochlea(耳蜗) → auditory nerve → brain



Loizou (1999) "Introduction to cochlear implants," IEEE Eng. in Med. and Bio. Mag., 18, 32-42.

The function of cochlea

基底膜 毛细胞



mechanical vibrations

↓

vibrations in fluid

↓

pressure variations

↓

displacements of basilar membrane

↓

bending of hair cells

↓

electrochemical substance

↓

neurons fire

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Loizou (1999) "Introduction to cochlear implants," IEEE Eng. in Med. and Bio. Mag., 18, 32-42.

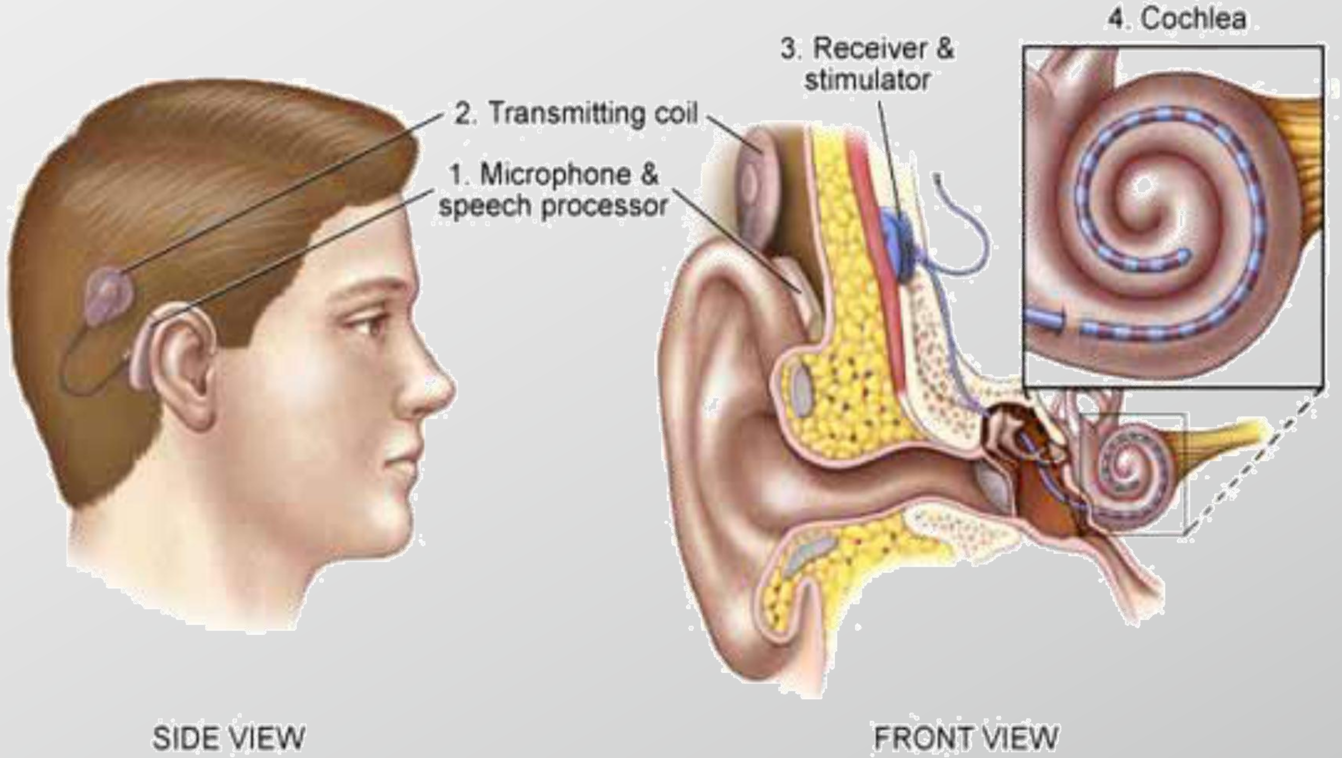
Deafness & cochlear implants

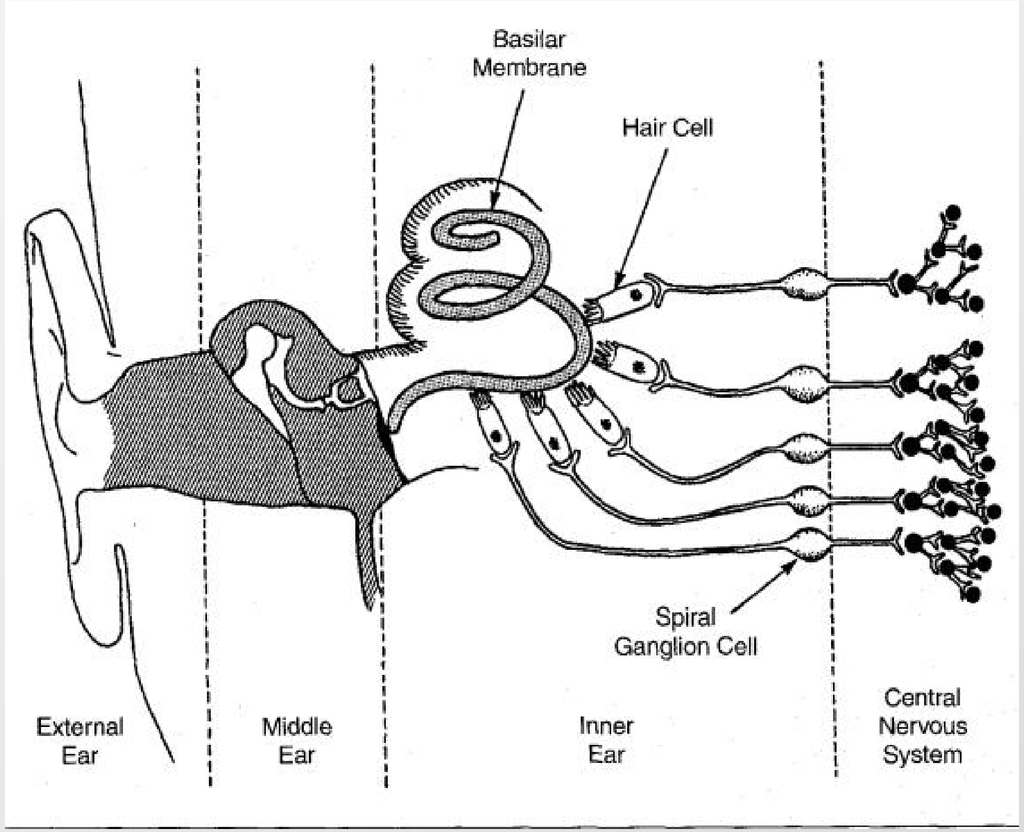
The most common cause of deafness is the loss of hair cells, rather than the loss of auditory neurons.

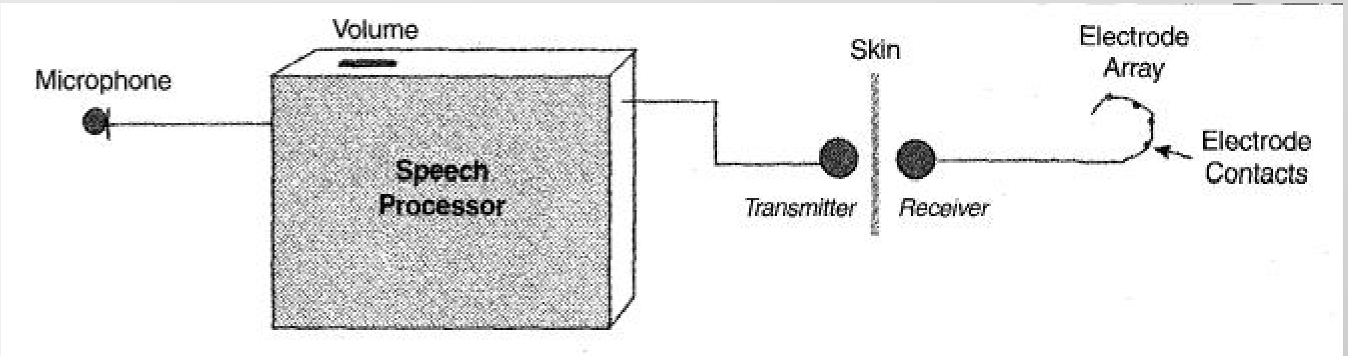
Belal, A., Maher, A., & Talaat, M. (1976). Histopathology of profound deafness. *Indian Journal of Otolaryngology*, 28(3), 108-111.

The principle of cochlear implants

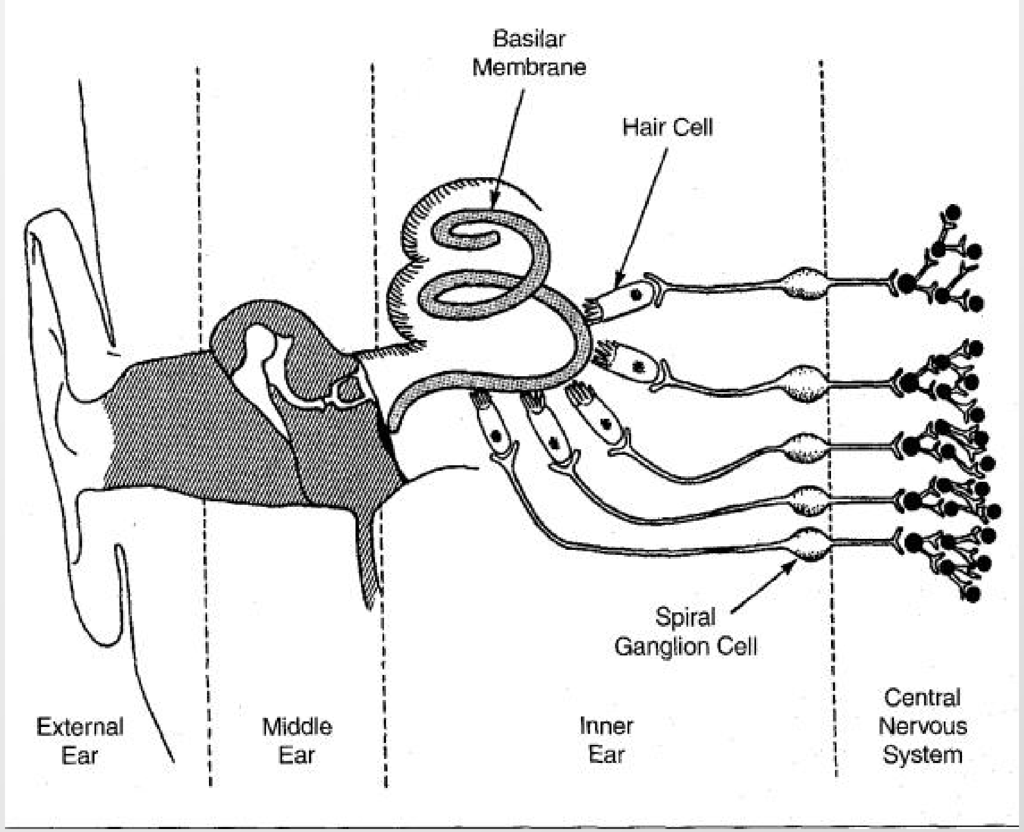
- The microphone collects signals, and the speech processor processes the collected signals.
- The wireless transmitter sends the processed signal to the inner side of the cochlea implants.
- Wireless receiver receives signals.
- Received signals stimulate the auditory nerve through electrodes to reconstruct signals.

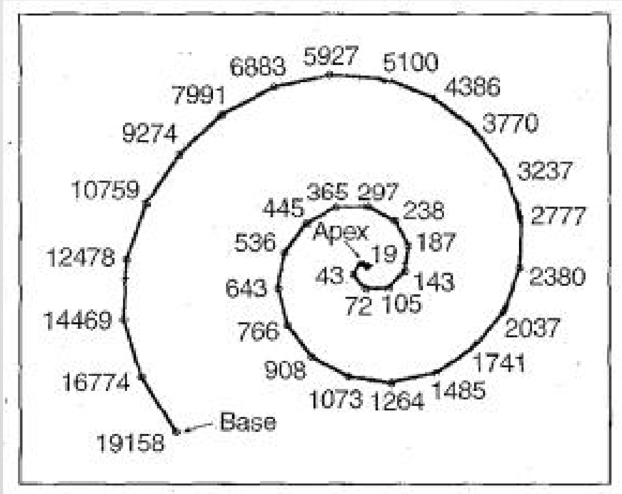






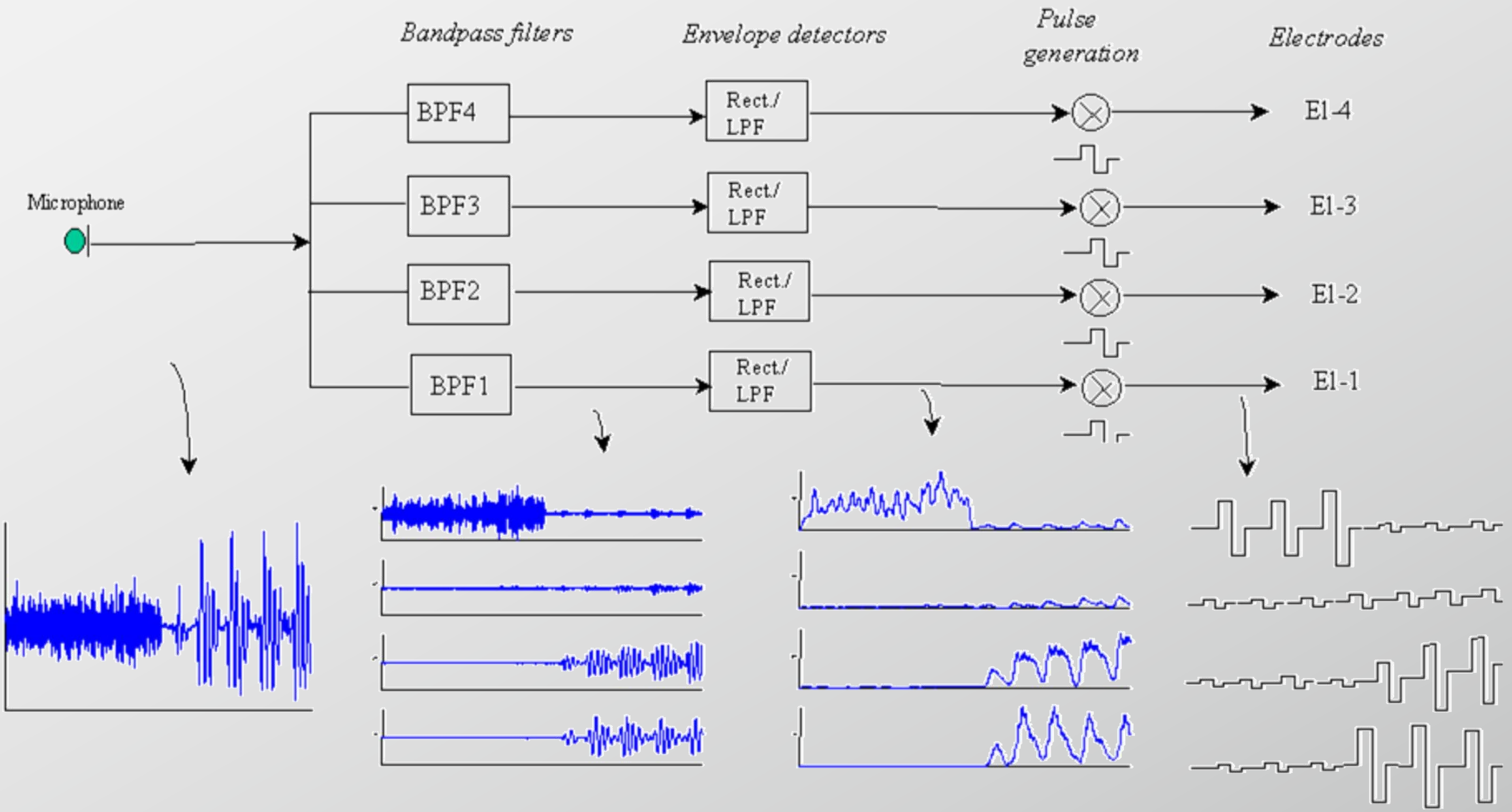
Loizou (1999) “Introduction to cochlear implants,” IEEE Eng. in Med. and Bio. Mag., 18, 32-42.





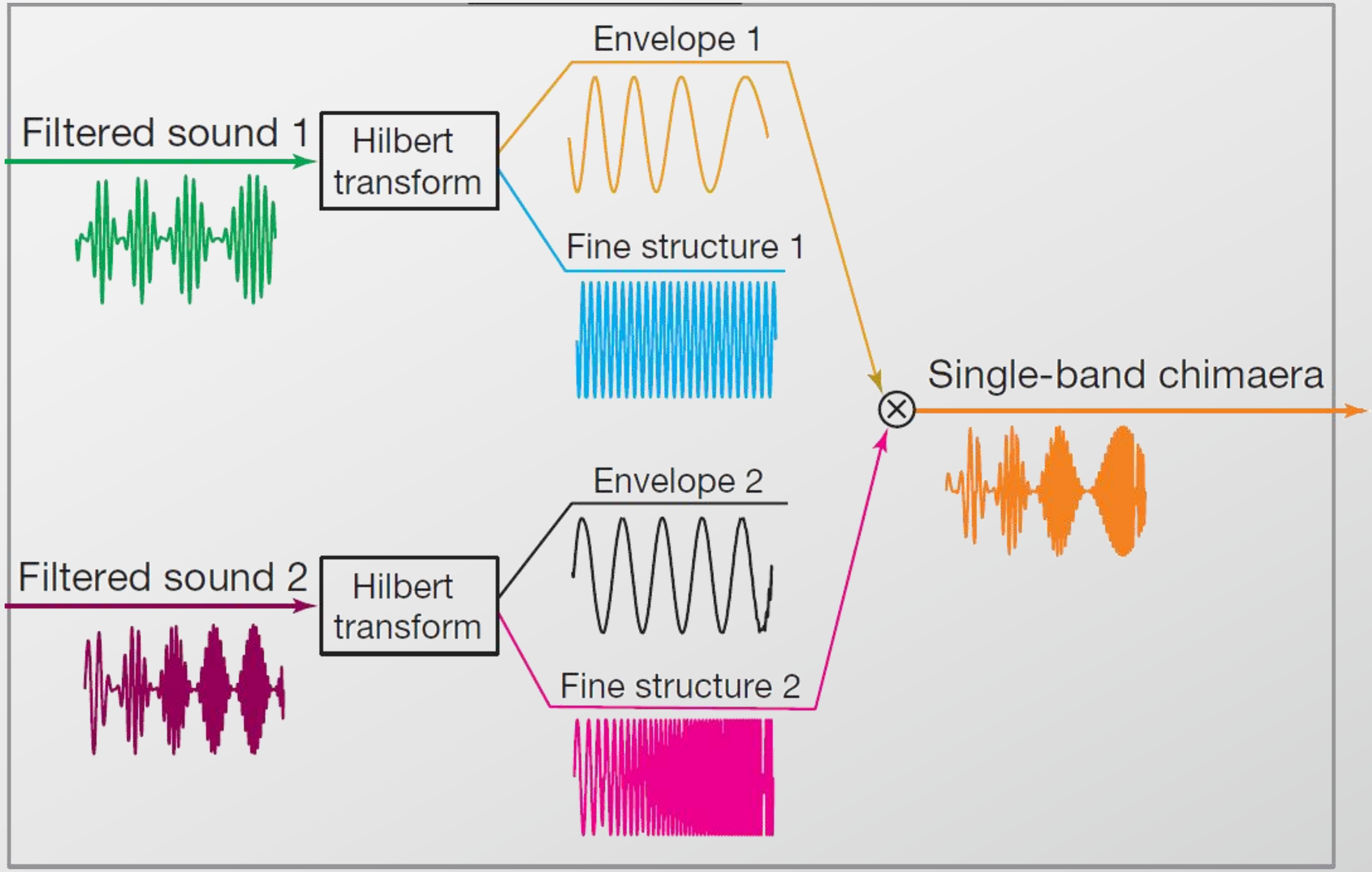
Loizou (1999) “Introduction to cochlear implants,” IEEE Eng. in Med. and Bio. Mag., 18, 32-42.

Speech processing in cochlear implants



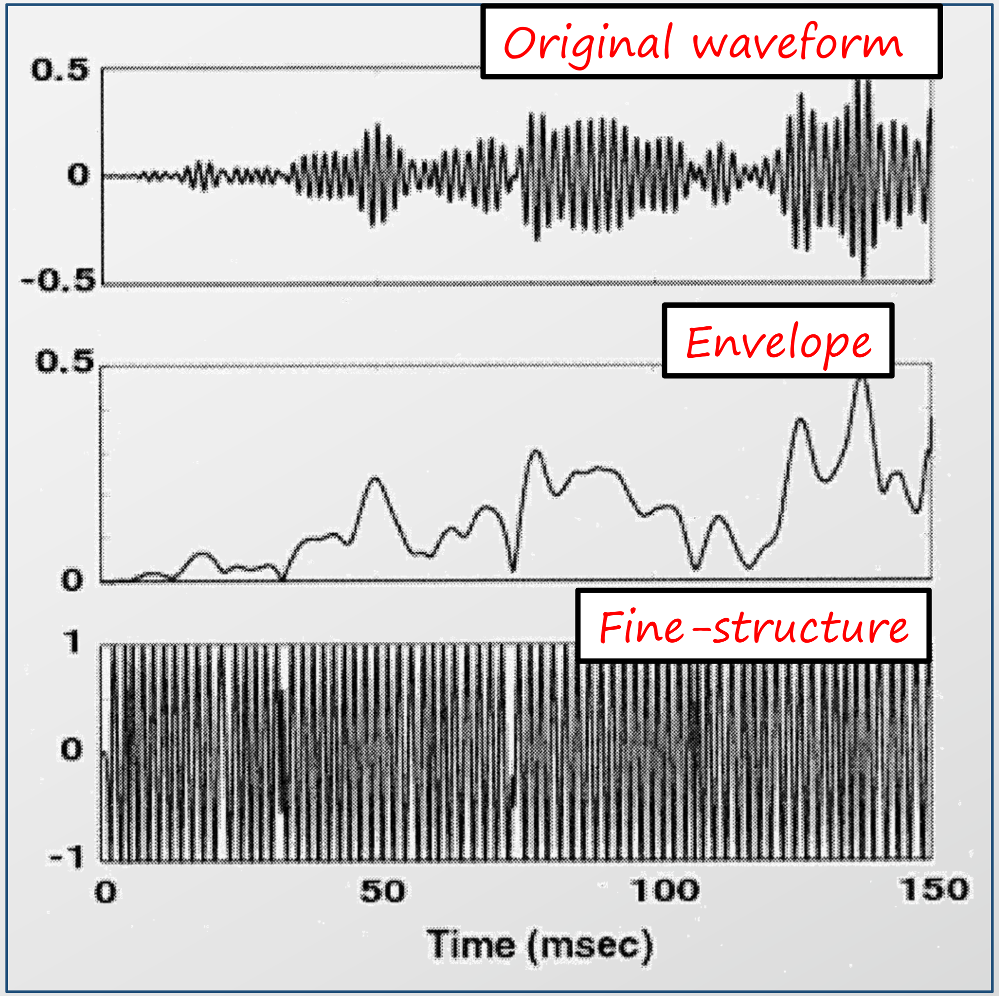
Loizou (1999) “Introduction to cochlear implants,” IEEE Eng. in Med. and Bio. Mag., 18, 32-42.

Acoustic cues of speech signal



- Envelope and fine- structure
  - **Envelope:**
    amplitude modulation,
    low- frequency
  - **Fine-structure:**
    frequency modulation,
    high- frequency

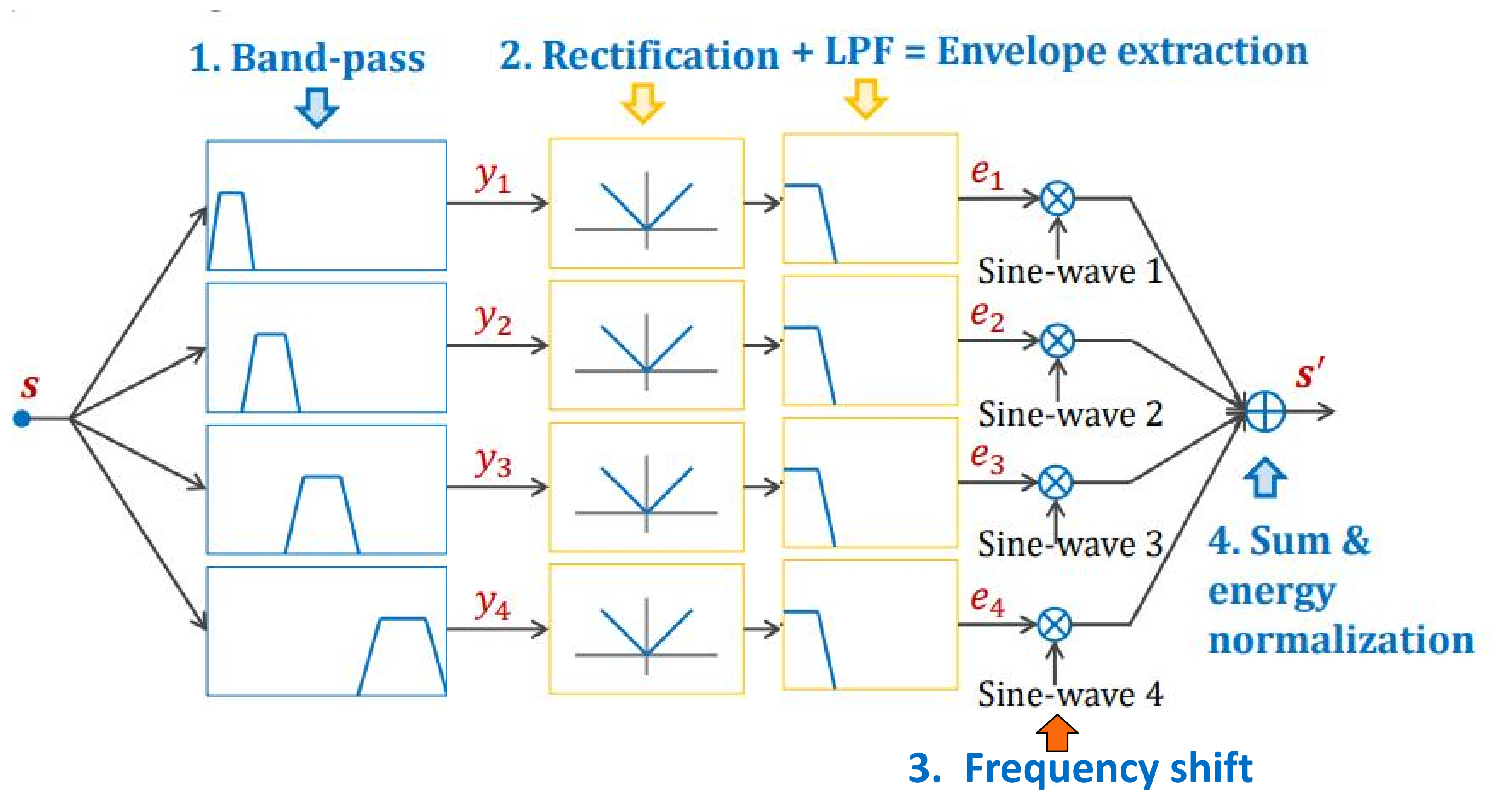
Z. M. Smith et al. (2002), "Chimaeric sounds reveal dichotomies in auditory perception ," Science.



- **Envelope:**
  The envelope is most important for speech reception
- **Fine-structure:**
  The fine structure is most important for pitch perception and sound localization.

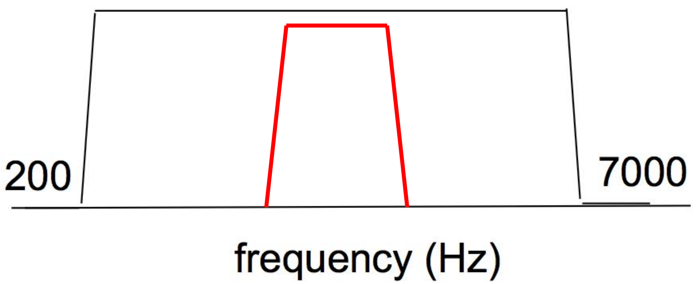
B.S. Wilson et al. (2005), "Two new directions in speech processor design for cochlear implants," Ear Hear.

Speech synthesis with envelope cue: Tone-vocoder



Example

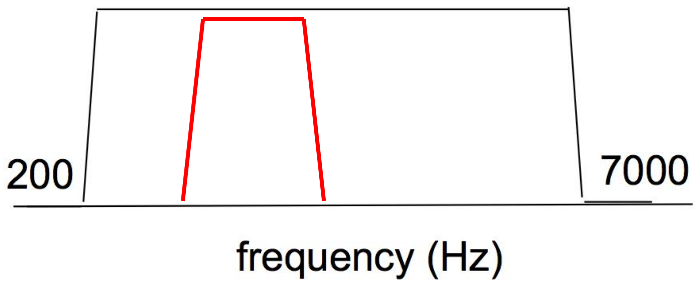
- Set the number of bands to N.
- For the i^{th} band



1) Design band-pass (e.g., 400-1000 Hz) filter at band i
   >> fs=16000; %sampling frequency
   >> [b, a]=butter(4, [400 1000]/(fs/2)); %band-pass filter
2) Do band-pass filtering at band i
   >> y= filter(b, a, s); % s is speech signal, and y is the band-passed signal at band i
3) Do full-wave rectification, and low-pass filtering to get the envelope at band i
4) Generate a sinewave, whose frequency equals to the center frequency of the i^{th} band-pass filter
5) Multiply the envelope signal in 3) and sinewave in 4)

6) Repeat for all N bands
7) Sum up the outputs from all bands (denoting the summed outputs as s')
8) Do energy normalization, i.e., letting the energy of s' equals to that of s
9) Save the wavefile for signal s'

- Set the number of bands to N.
- For the i^{th} band



1) Design band-pass (e.g., 400-1000 Hz) filter at band i
   >> fs=16000; %sampling frequency
   >> [b, a]=butter(4, [400 1000]/(fs/2)); %band-pass filter
2) Do band-pass filtering at band i
   >> y= filter(b, a, s); % s is speech signal, and y is the band-passed signal at band i
3) Do full-wave rectification, and low-pass filtering to get the envelope at band i
4) Generate a sinewave, whose frequency equals to the center frequency of the i^{th} band-pass filter
5) Multiply the envelope signal in 3) and sinewave in 4)

6) Repeat for all N bands
7) Sum up the outputs from all bands (denoting the summed outputs as s')
8) Do energy normalization, i.e., letting the energy of s' equals to that of s
9) Save the wavefile for signal s'

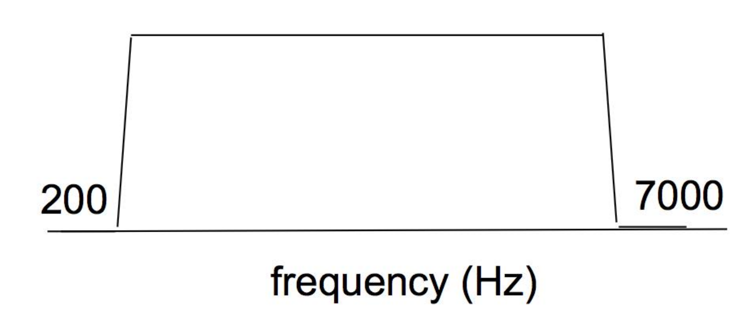
Question? How to divide pass-band from 200 Hz to 7000 Hz?

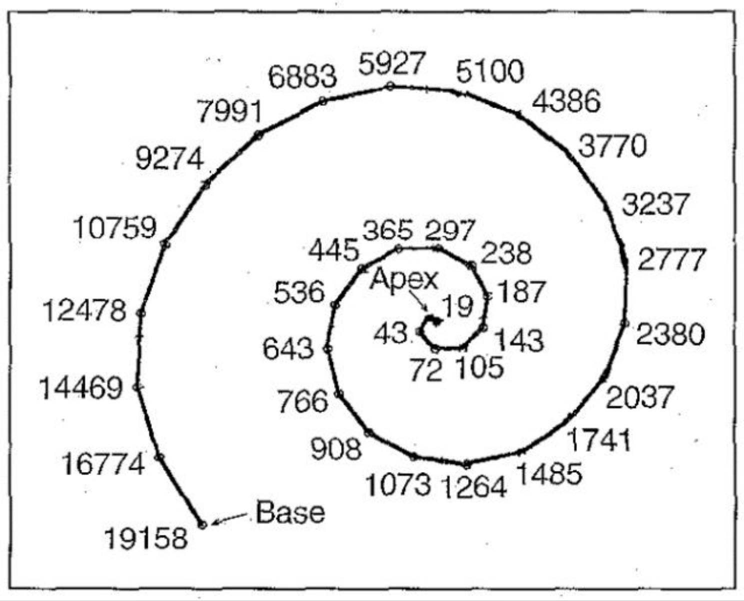
Determine the frequency range of the i^{th} band :

- Frequency-to-place mapping as:

$$f = 165.4 \times (10^{0.06 \cdot d} - 1)$$

where f is -3 dB cut-off frequency, and d is the distance (in millimeter) along the cochlea.





1) Equally divide the cochlea length.

2) Equal frequency interval division

Greenwood, D.D., 1990. A cochlear frequency-position function for several species—29 years later. J. Acoust. Soc. Am. 87, 2592–2605.

Project tasks -1

- **Sentences for pro 1: ‘C_01_01.wav’ & ‘C_01_02.wav’**
- **Task 1**
  - Set LPF cut-off frequency to 50 Hz.
  - Implement tone-vocoder by changing the number of bands to N=1, N=2, N=4, N=6, and N=8.
  - Save the wave files for these conditions, and describe how the number of bands affects the intelligibility (i.e., how many words can be understood) of synthesized sentence.

Project tasks -2

- **Task 2**
  - Set the number of bands N=4.
  - Implement tone-vocoder by changing the LPF cut-off frequency to 20 Hz, 50 Hz, 100 Hz, and 400 Hz.
  - Describe how the LPF cut-off frequency affects the intelligibility of synthesized sentence.

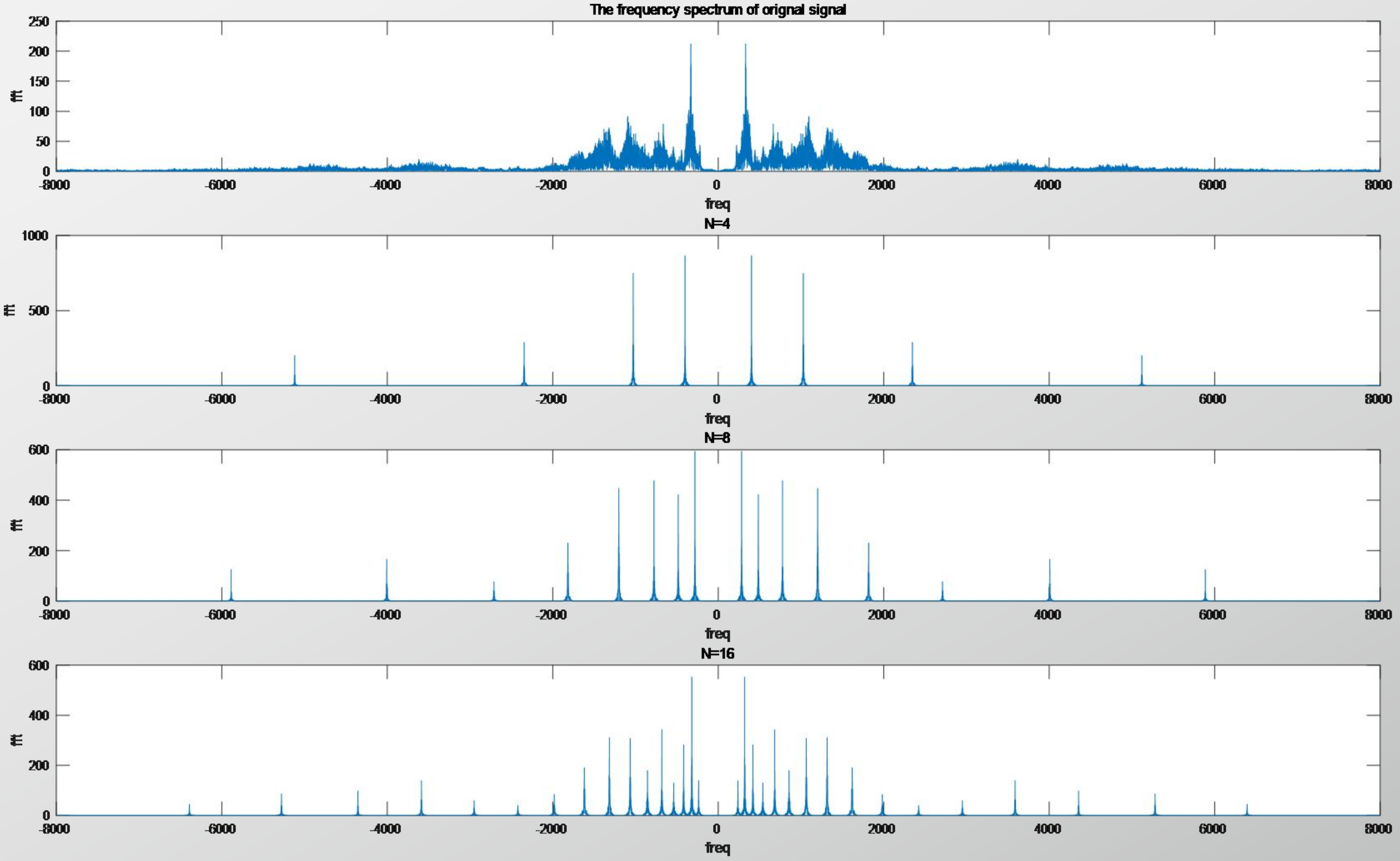
Project tasks -3

- **Task 3**
  - Generate a noisy signal (summing clean sentence and SSN) at SNR -5 dB.
  - Set LPF cut-off frequency to 50 Hz.
  - Implement tone-vocoder by changing the number of bands to N=2, N=4, N=6, N=8, and N=16.
  - Describe how the number of bands affects the intelligibility of synthesized sentence, and compare findings with those obtained in task 1.

Project tasks -4

- **Task 4**
  - Generate a noisy signal (summing clean sentence and SSN) at SNR -5 dB.
  - Set the number of bands to N=6.
  - Implement tone-vocoder by changing the LPF cut-off frequency to 20 Hz, 50 Hz, 100 Hz, and 400 Hz.
  - Describe how the LPF cut-off frequency affects the intelligibility of synthesized sentence.

Results（How do we draw an conclusion？）



Organization

- **Each group consists of three or four students.**
- **Each group need present one Lab project (submit reports for both projects):**
- The presentation date is May. 31 for project 1， and June 7 for Project 2.
- **Each presentation is 15 minutes (including Q & A)**
  - All team members need to contribute to the presentation.
  - Presentation in English or Chinese.

The presentation should...

- Introduce
  - team
  - objective of the project
  - background review (search more additional information)
  - methodology
- Present
  - relevant data, figure, etc.
  - the results for project tasks (e.g., with demo, Figure, etc.)
  - interpretation of project findings
- Discuss
  - what you have learned from this study?
  - problems during this project and your solution
  - investigation beyond project tasks
  - critical thinking
- Appendix (if any)
- Team effort (e.g., individual contribution)
- Reference
- Q & A (answer questions raised from audience)

project1作业

实验目标： 语音合成

实验报告内容：

Introduction（20%）：

1、语音合成器的基本模型，

2、巴特沃斯低通/带通滤波器原理。

3、滤波器函数设计、验证、使用步骤，

4、包络提取原理；

5、言语谱噪声产生原理；

6、能量归一化原理 ，

注：原理不要直接从PPT截图。

Result and Analysis：

1、 题目（Task1，Task2，Task3，Task4），注：题目不要直接从PPT截图。（10%）

注意：（1）低通/带通滤波器系统必须验证；（2） 包络 图要画出；（3）信噪比-5dB必须验证；（4）能量归一化验证必须体现

2、 结果图（时域波形【重要：所有图横坐标以时间秒为单位，纵坐标为幅度，所有坐标轴需要标签】，频谱【重要：所有图横坐标以频率Hz为单位，纵坐标为幅度谱， 所有坐标轴需要标签 】）（20%）

3、 对结果的分析和解释（对结果必须有分析、解释和结论）（20%）

4、扩展（20%）：

1）录制一段不同的WAV文件，进行语音合成。

2）完成项目基础内容，鼓励自己增加尝试，研究、讨论，内容自拟，加分项需要在Pre中明确讲出

5、总结（10%）：对所做工作进行总结，得到概括性结论。整体报告保证结构完整，逻辑清晰。

Score：自我评分（0-100, 小组成员分别给出）

Code：报告的最后需要附上所有代码。

提交截止日期：6.2日24:00点。

Questions