



# 语音信号处理

---

李军锋

中国科学院声学研究所



# 提纲

---

- 学科介绍
- 课程概览
  - 语音产生与听感知
  - 语音表示
  - 语音估计算法
  - 语音信号处理的应用
- 课程安排



# 学科介绍：语言声学

声 学						
声波频率	0Hz	20	200	8k	20k	1G
波段划分	次声	语音			超声	特超声或微波超声
应用领域	气象预测 军事用途	语音通信、信息检索、人机交互等			工业探测 医疗保健	声学研究 微电子学 医学
特点		物理学、数学、信息、心理 声学高度交叉融合, 人类自然 交流的方式！				



# 学科介绍：语音

---

- **语音是人类日常使用的最有趣的信号之一，语音也是一个多学科交叉的研究主题**
  - **语音是人类交流的最自然的方式**
  - **语音跟语言有关，而语言学是社会科学的一个分支**
  - **语音跟人类的生理能力有关，而生理学是医学科学的一个分支**
  - **语音跟声音有关，而声学是物理科学的一个分支**
  - **语音跟信号和信息有关，信号和信息处理则是电子、通讯和计算机等领域的重要主题**



# 学科介绍：声音(语音、乐音、音频)

---

- 物理学中，声音是由物体振动产生的，通过介质（空气或固体、液体）传播，能被人或动物的听觉器官所感知的波动现象。人耳听觉范围：20 Hz~20 kHz。
- 音频泛指各种声音：语音，乐音，噪音
- 音频信号是声音的一种电的表示
- 课程内容覆盖
  - 重点是语音信号处理
  - 不失一般性时，音频信号处理
  - 未覆盖：语音编码



# 学科介绍：信号与处理

---

## □ 信号

- 可以是物理的；可以是符号或抽象的信息，如基因序列；也可以是序列信息的抽象属性，如冷、热，高、低
- 常见的信号：音频，视频，语音，语言，图像，多媒体，传感器，通讯，声纳，雷达，地球物理的，生物的，化学的，分子的，基因的，医学的，音乐的，数据，属性序列，数量等

## □ 处理

- 各种操作：表示、过滤、编码、传输、估计、检测、推理、发现、识别、合成、录制或产生
- 形式：硬件、软件或固件
- 手段：数字或模拟的设备、技术或算法



# 学科介绍：信号处理

---

- 是一项包含了对多种不同的物理的、符号的或抽象形式的被广泛称作信号进行处理或信息变换的基础理论、应用、算法和实现的通用支撑性技术
- 采用数学的、统计的、计算的、启发式的和/或语言学的表示、形式和技术，来进行各种处理：表示、建模、分析、合成、发现、恢复、感知、获取、学习等



# 学科介绍：语音信号处理

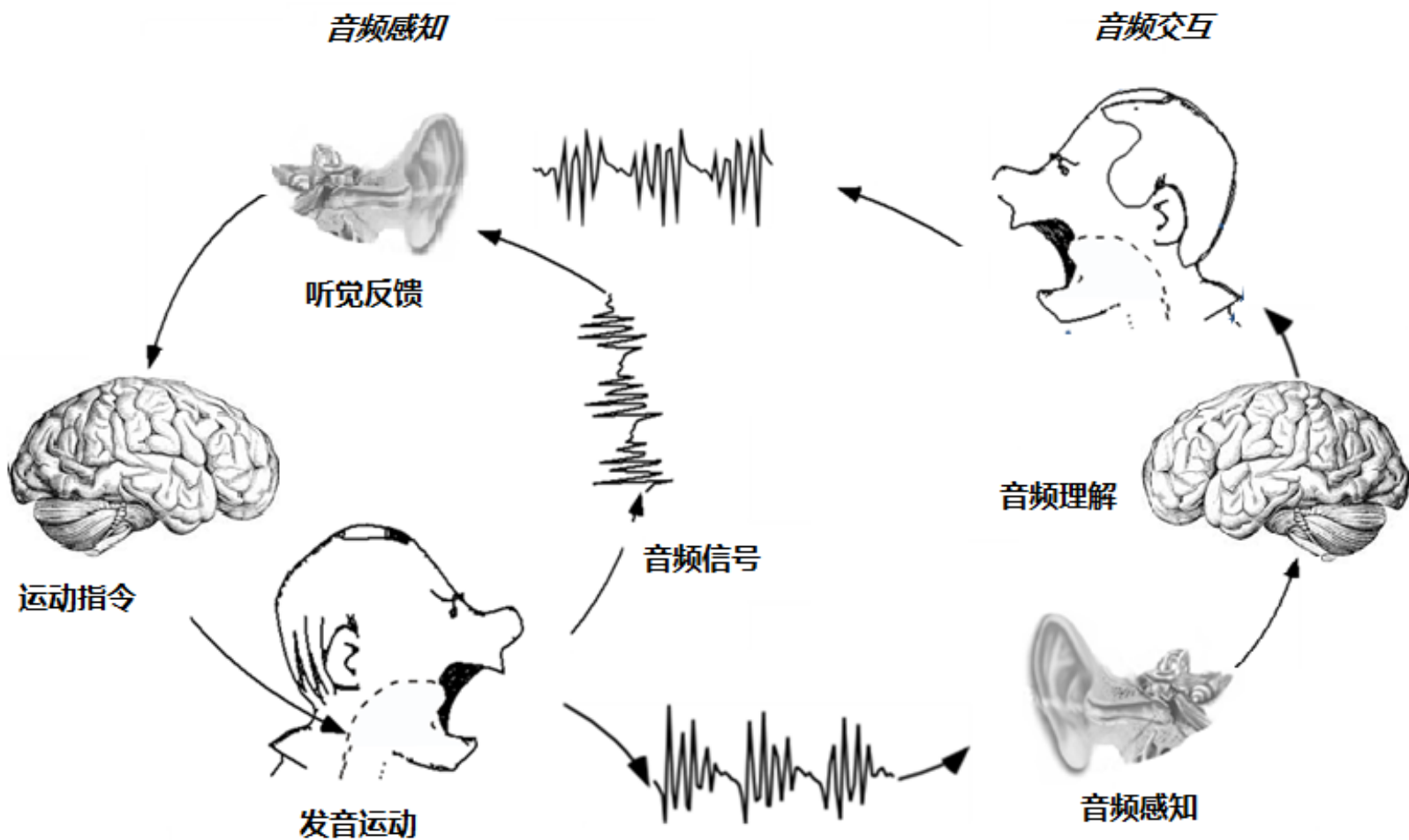
---

- 将一种语音信号表示转换为另一种表示，以便揭示语音信号的各种数学的或实际的属性，或通过合适的处理来帮助解决基本和深层次的问题
- 主要目的包括：
  - 理解作为一种交流方式的语音
  - 为变换和复制而表示语音
  - 为自动识别和提取信息而分析语音
  - 为了发现说话者的生理特点等





# 学科介绍：语言声学

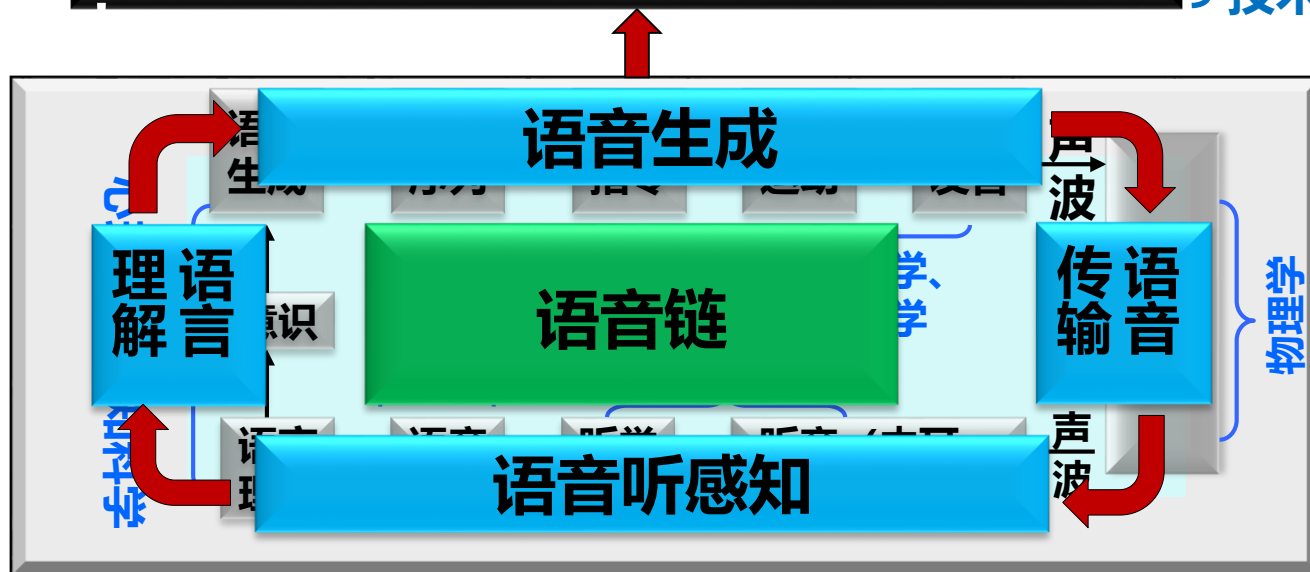




# 学科介绍：语言声学

新发展：让计算机能听会说、能理解会思考

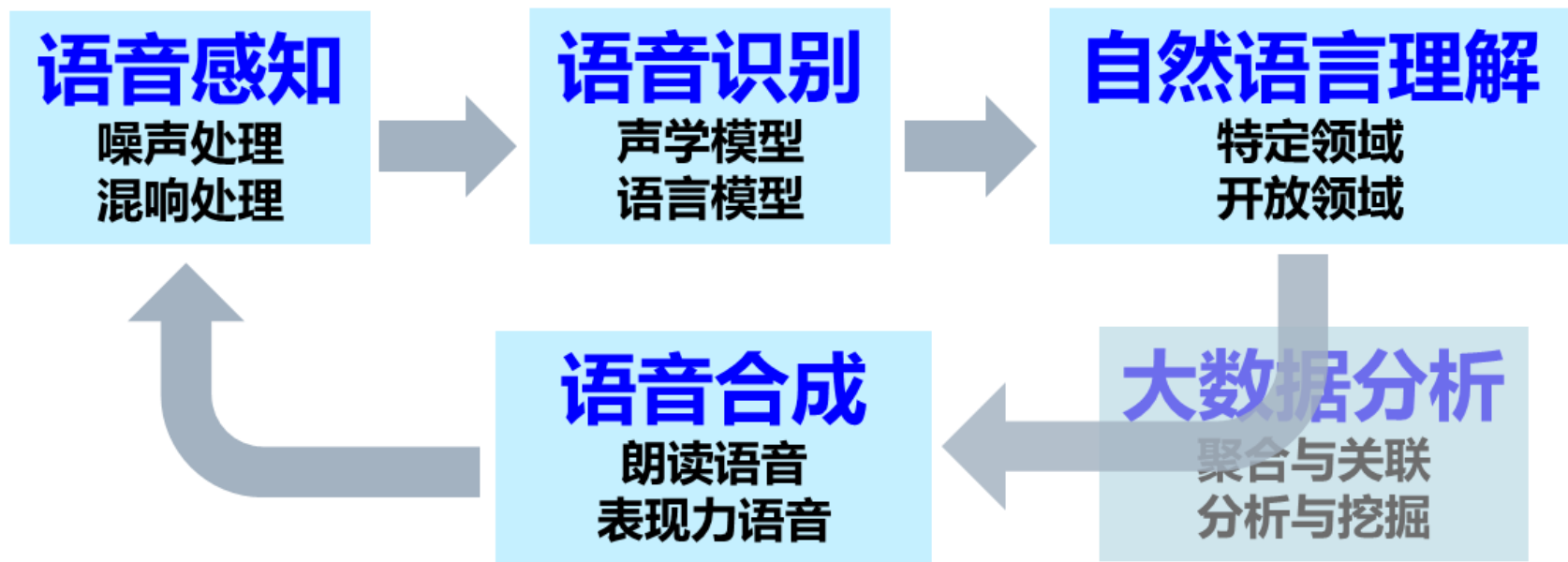
信息技术





# 学科介绍：语言声学

## 智能语音交互系统及关键技术





# 学科介绍：语言声学

## 行业发展需求



skype™



**智能语音处理技术是行业发展的迫切需求**

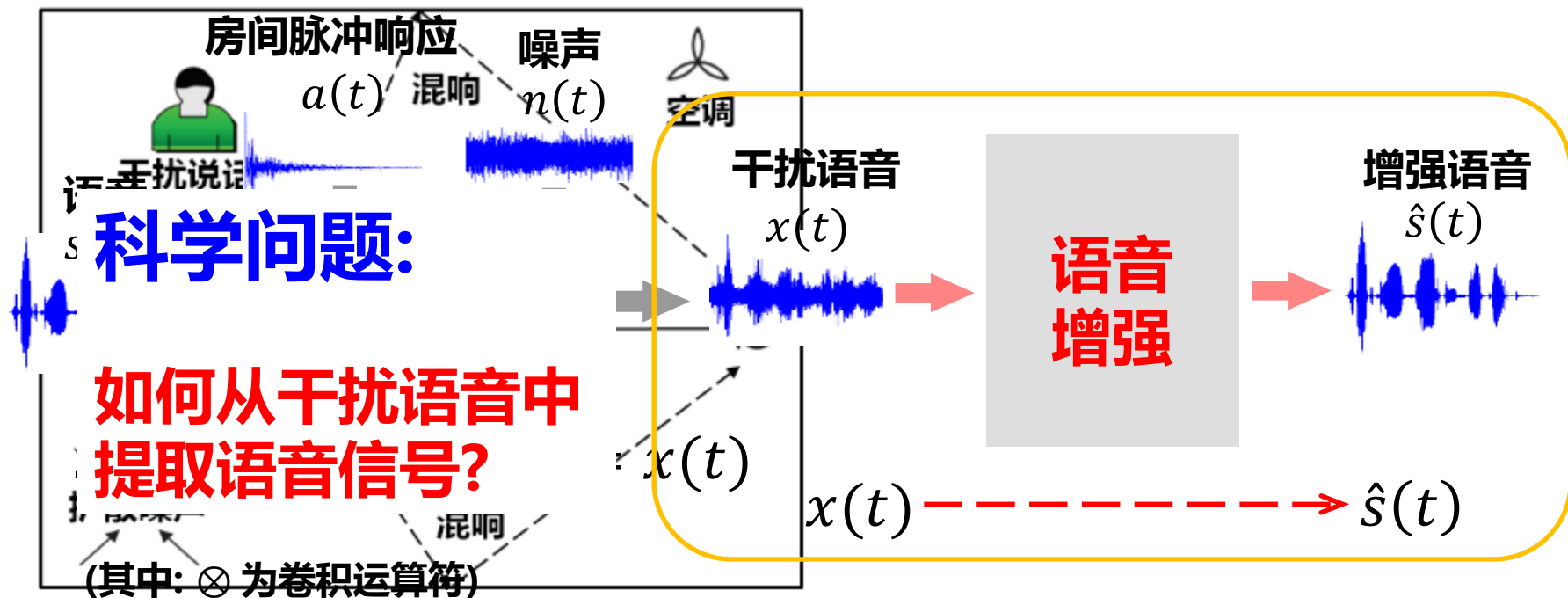
## 国家安全需求



**海量语音数据的智能化处理是国家安全中亟待解决的问题**

# 学科介绍：语音信号处理当前研究热点

噪声混响降低了语音质量和可懂度, 影响了语音应用的性能



语音增强是语言声学领域的核心问题和研究热点之一



# 学科介绍：语音信号处理当前研究热点

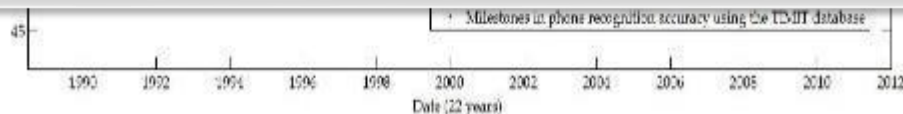
## 近年语音识别性能进步的两大驱动

深



音素识别率提升有限，说明未来识别提升的空间已经有限！

网  
络



Google 上万小时言语数据的实验表明现有框架依赖数据量的提升也即将遭遇天花板！





# 学科介绍：典型应用场景

## 大数据时代的音视频



在Netflix上每天观看2200万小时电视电影节目

在Youtube上每天上传86.4万小时视频



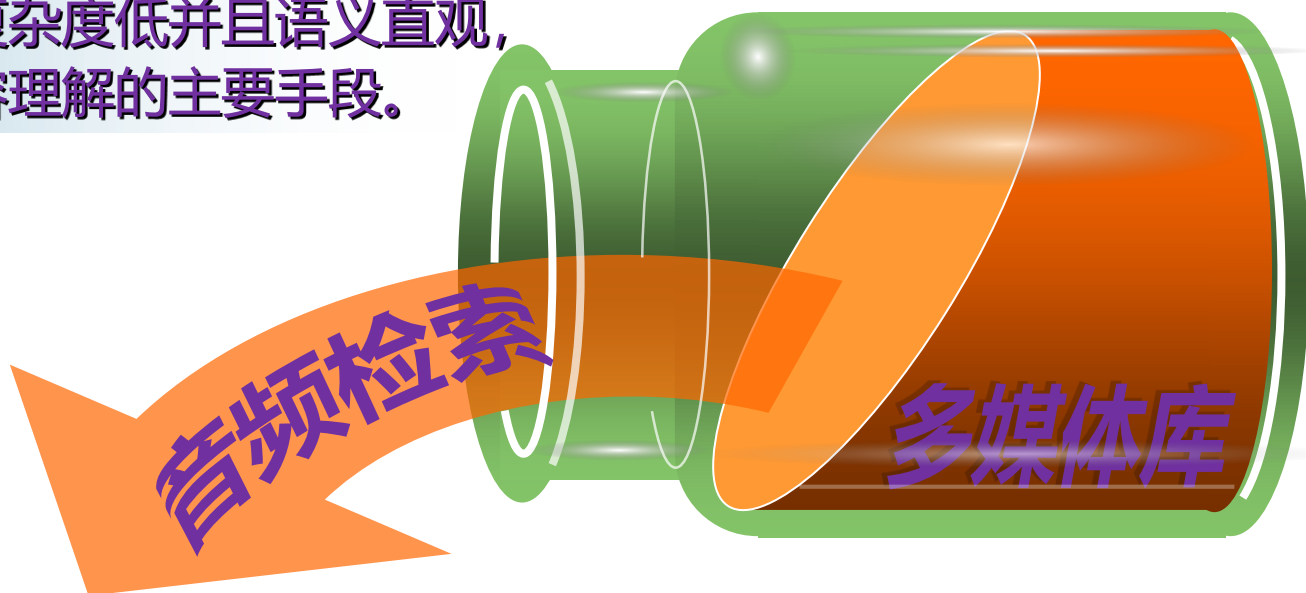




# 学科介绍：典型应用场景

文字浏览，可以一目十行，但音视频不行。

语音数据量小，计算复杂度低并且语义直观，  
将是多媒体检索和内容理解的主要手段。

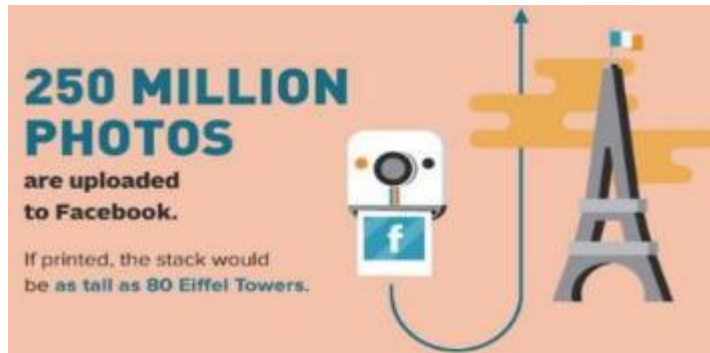






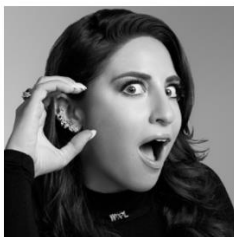
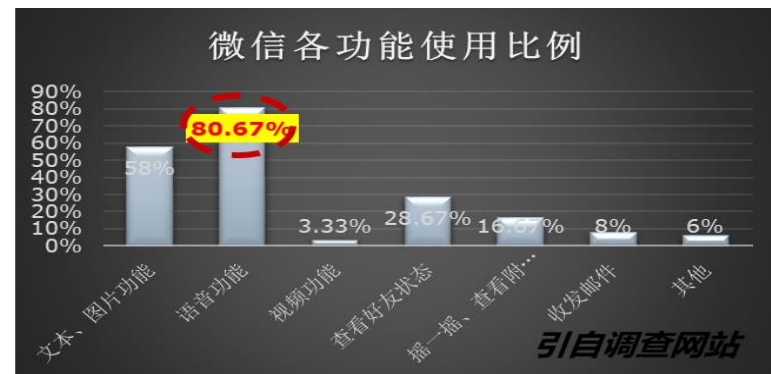
# 学科介绍：典型应用场景

## 社交分析



在facebook上每天  
上传2.5亿张图片

在照片分享应用日趋饱和后，语音分享将成为移动社交革命性的应用



### Clubhouse: The Social Audio App

Alpha Exploration Co. Social

★★★★★ 338,065

Teen

⚠ You don't have any devices

Add to Wishlist

Install



# 学科介绍：典型应用场景

## 语音数据挖掘

- 与文字相比，语音中包含更好的信息
- 语音中的信息能更直观反映说话人
- 语音是有感情的





# 学科介绍：典型应用场景

## 免提场景下的语音技术



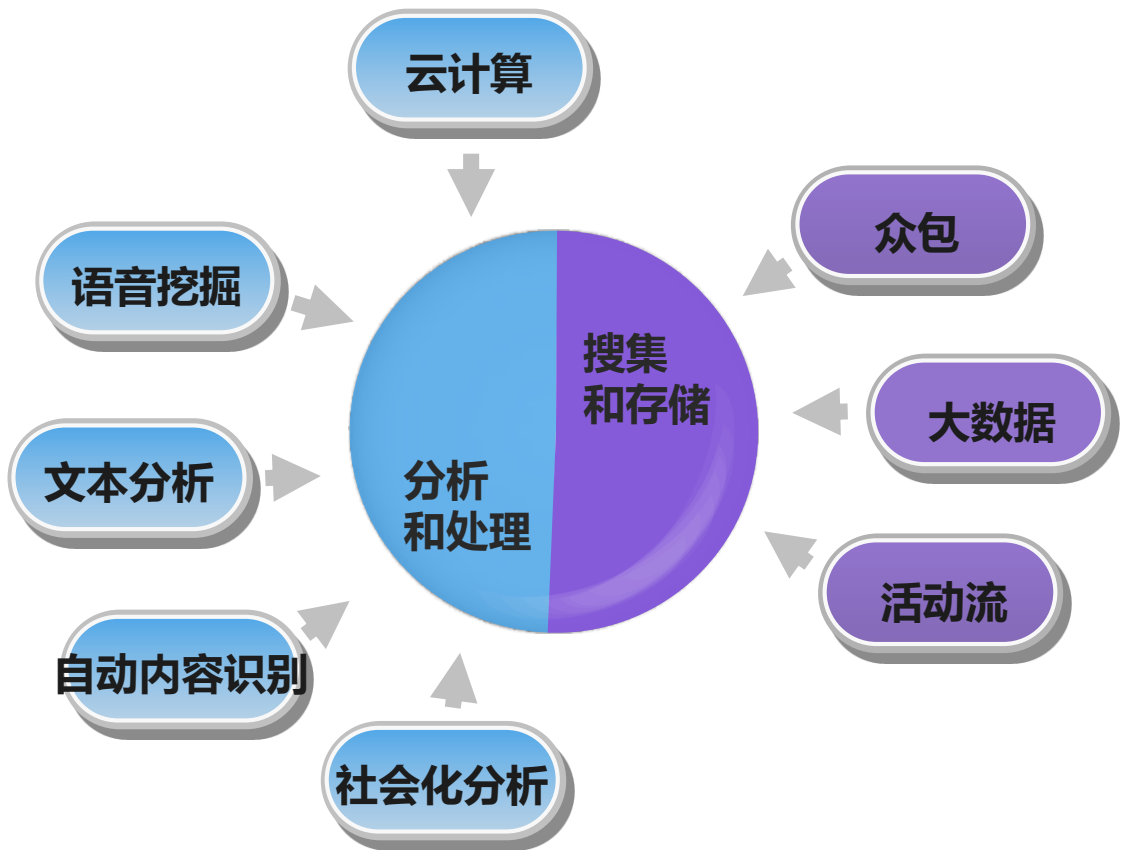
**智能家电、智能家居、可穿戴设备、车联网终端等可计算终端已成为自然人机交互的入口硬件设备**



# 学科介绍：典型应用场景

## 客户语音将被全部存档

人类天生就是社会化的，因此分享是必须的。“客户声音”被完整保留的，这些语音记录会被存放到云端，可以被存取、分析以便更好地剖析。





# 学科介绍：典型应用场景

## 智能语音交互的当前热点





# 学科介绍：语音技术发展史

## Speech Milestones



**1971**

DARPA starts work on first speech project



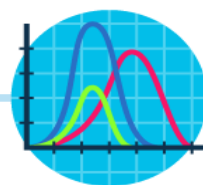
**1975**

Tech later known as Hidden Markov Models proposed; IBM begins speech work



**1978**

Texas Instruments releases Speak & Spell



**1982**

Gaussian Mixture Models start being used for speech recognition



**1986**

Researchers begin experimenting with neural networks



**2007**

Microsoft releases Mobile Voice Search for Live Search (now Bing)



**2001**

Microsoft introduces Windows and Office XP Speech using SAPI 5.0



**1996**

SRI International deploys interactive voice response software



**1995**

Microsoft introduces the first speech tools in Windows 95



**1994**

Dragon Naturally Speaking, IBM Via Voice introduced



**2010**

Microsoft introduces Kinect for Xbox



**2011**

Apple introduces Siri for iPhones



**2012**

Microsoft makes deep learning breakthrough; Google Now introduced



**2014**

Microsoft introduces Cortana personal assistant



**2015**

Microsoft makes Skype Translator available in German, Mandarin and other languages



# 课程内容

---

- **基础：**语音生成和听觉感知、声学、语言学
- **语音表示：**时域、频域、线性预测等
- **语音估计算法：**语音活动度检测、清音和浊音的分割和分类、基频检测、共振峰估计等
- **语音应用：**语音增强、三维音频、语音识别、说话人识别、语种识别等



# 各讲主题

---

## □ 基础知识

- 第二讲：语音生成与听觉感知

## □ 语音信号处理基础

- 第三讲：语音信号短时分析
- 第四讲：语音信号线性预测

## □ 语音信号处理算法

- 第五讲：语音增强原理与应用
- 第六讲：多源语音信号分离
- 第七讲：多源语音定位
- 第八讲：语音活动性检测
- 第九讲：三维音频

## □ 语音/音频应用

- 第十讲：深度神经网络在语音处理中的应用
- 第十一讲：语音识别基础原理及应用
- 第十二讲：鲁棒语音识别
- 第十三讲：语音合成技术
- 第十四讲：说话人识别与语种识别
- 第十五讲：复习与答疑
- 第十六讲：考试



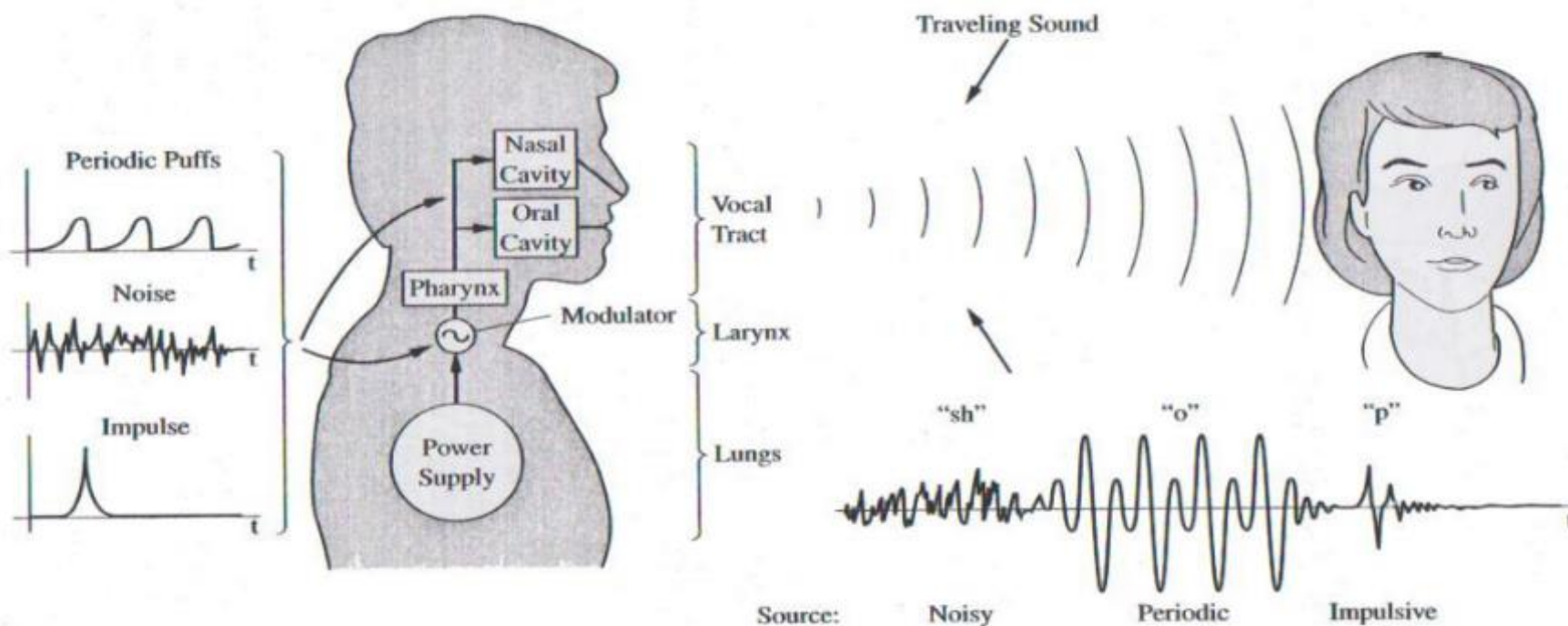


# 课程概览

# 第二讲：语音生成与听觉感知

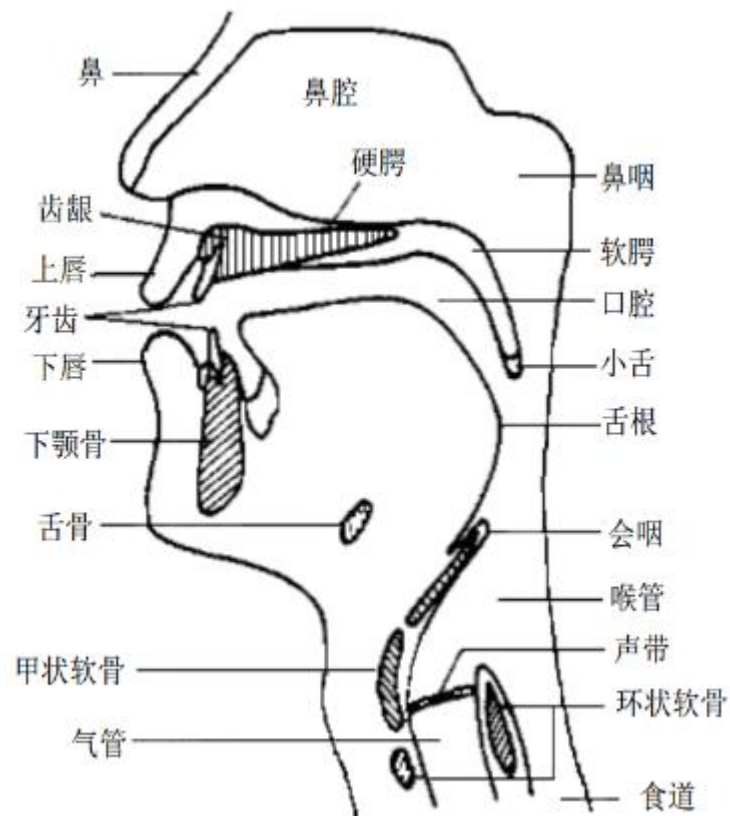
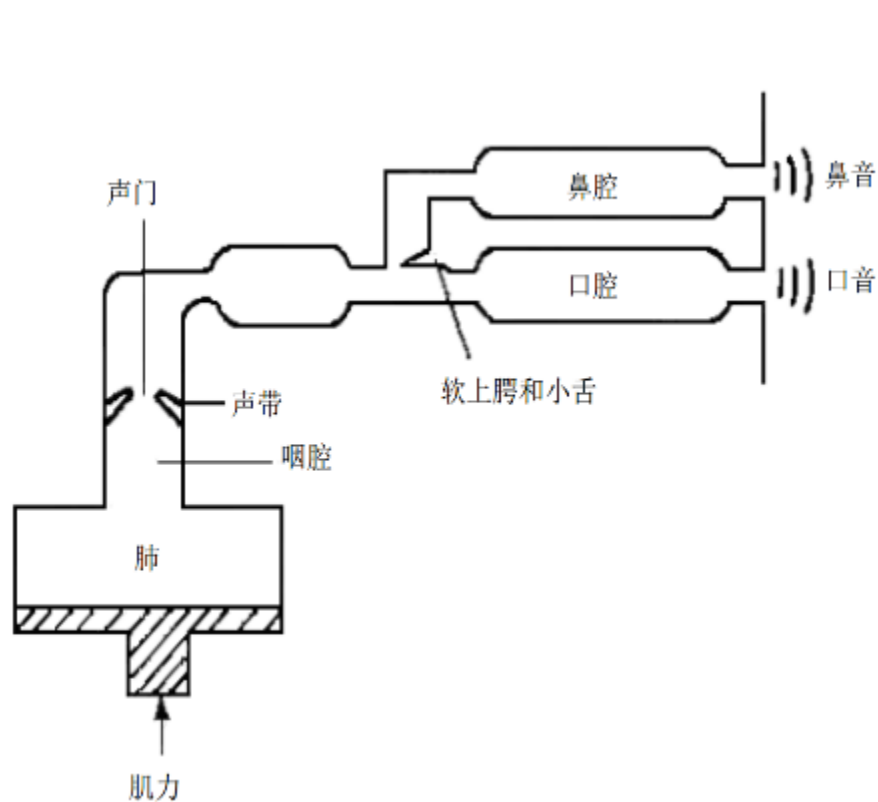
## □ 生理产生过程

■ 能量供给 - > 音源生成 - > 调音发声



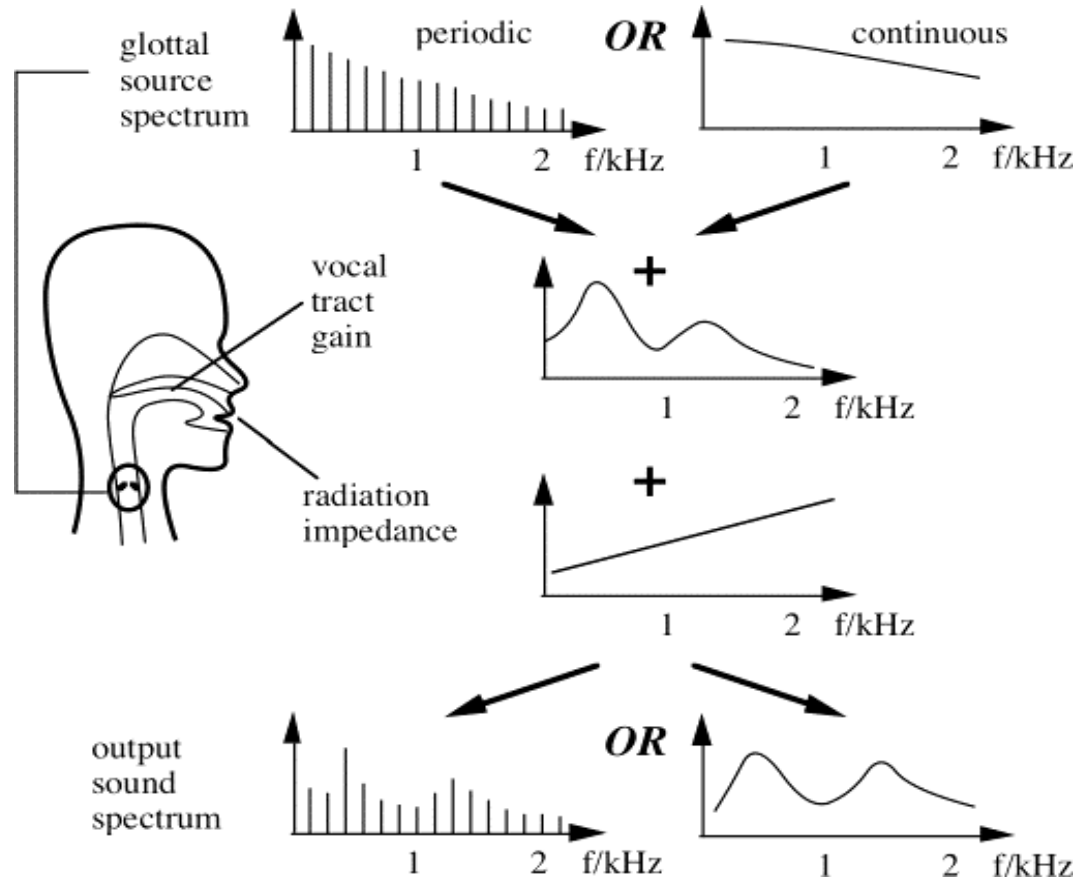
# 第二讲：语音生成与听觉感知

## □ 发音器官的机理模型



# 第二讲：语音生成与听觉感知

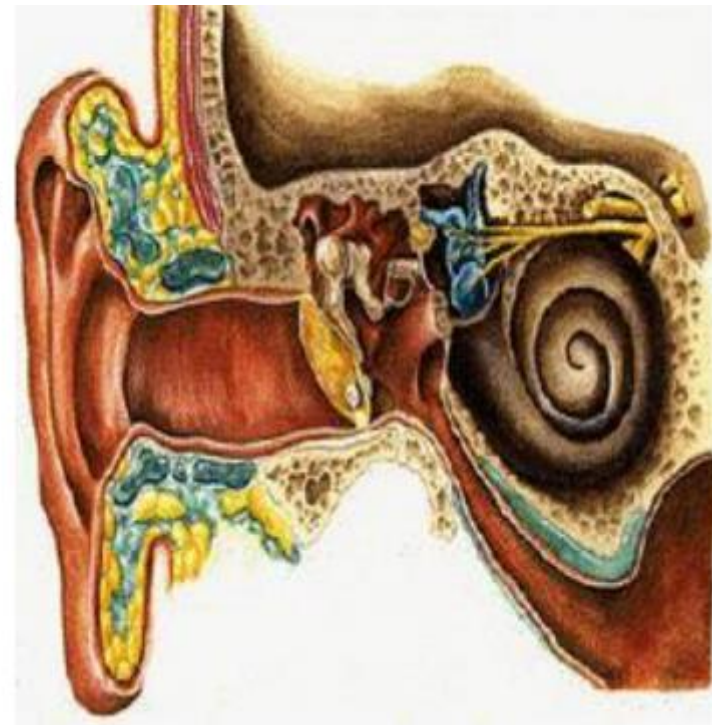
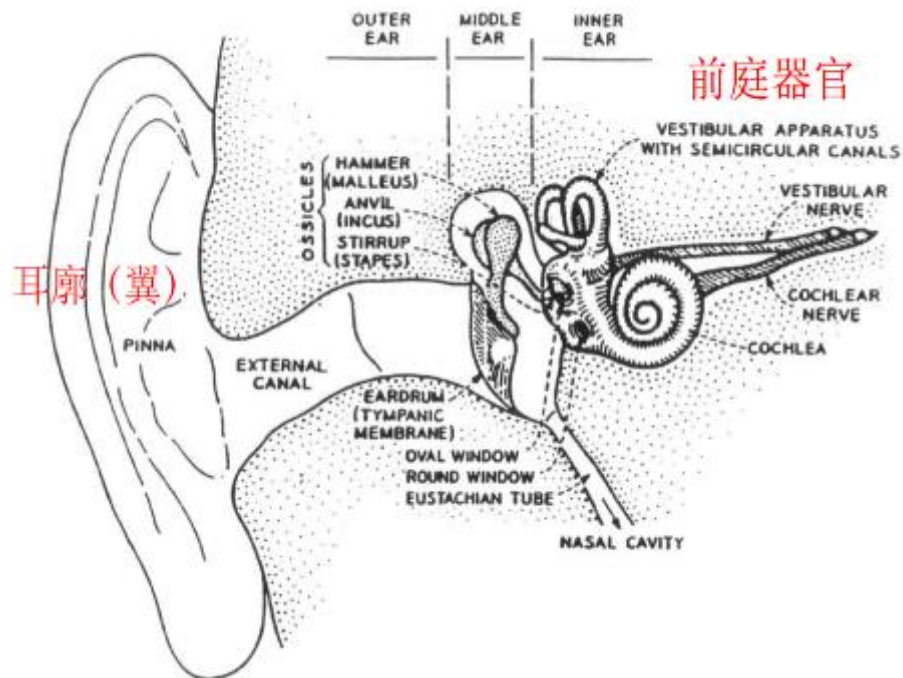
## □ 语音生成与数字化建模



# 第二讲：语音生成与听觉感知

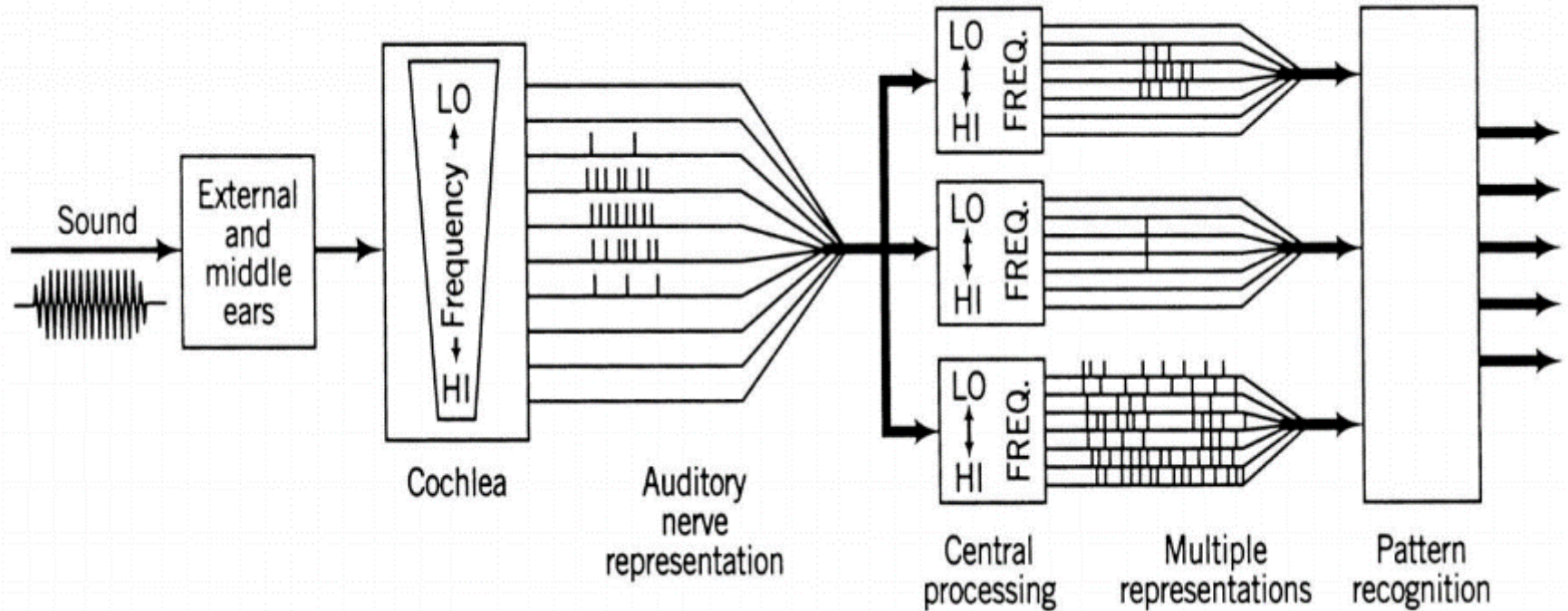
## □ 听觉系统的生理基础

### ■ 外耳、中耳、内耳



# 第二讲：语音生成与听觉感知

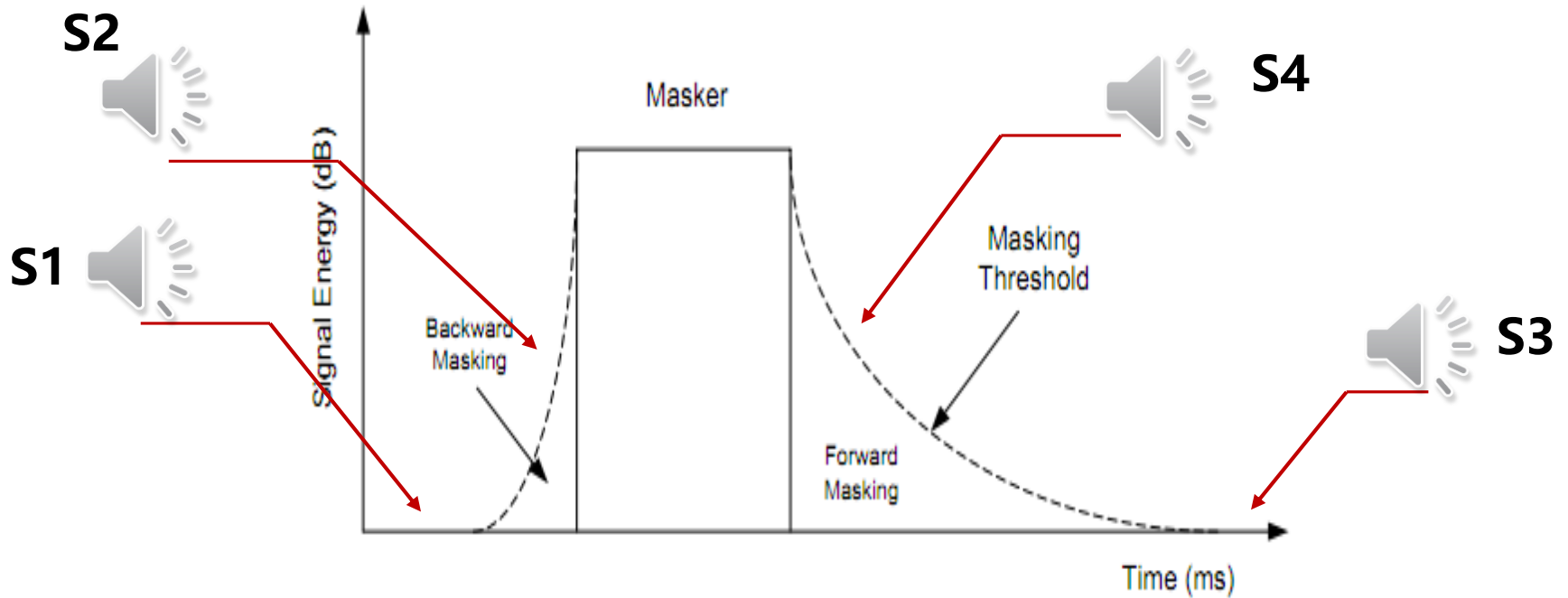
## □ 听觉感知过程的机理模型





# 第二讲：语音生成与听觉感知

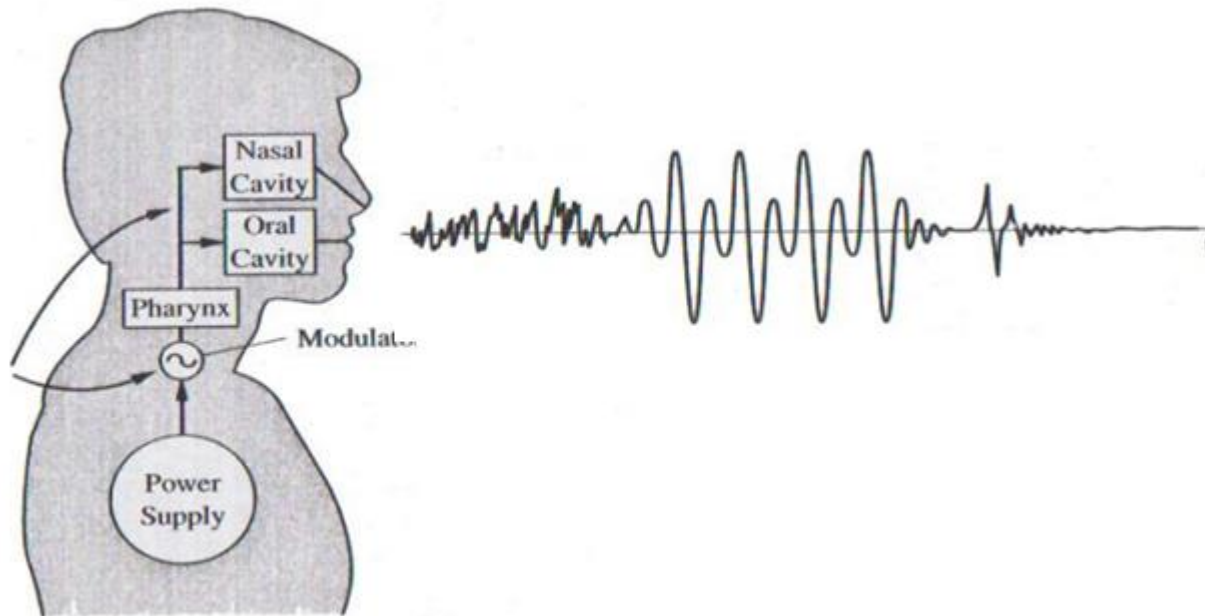
## □ 时域掩蔽 (temporal masking)





# 第三讲：语音信号短时分析

## □ 必要性：短时平稳性







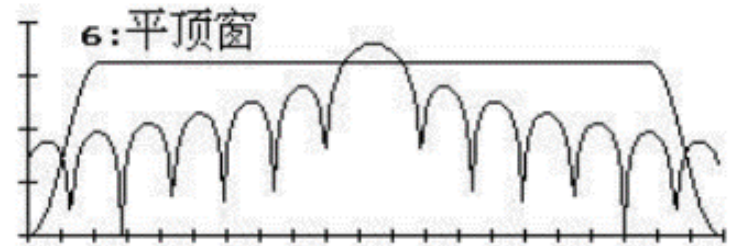
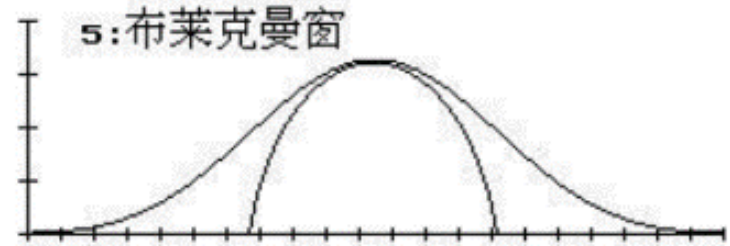
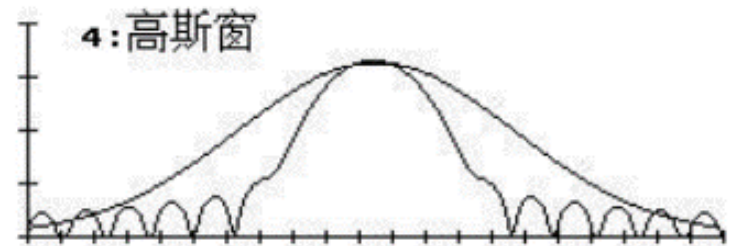
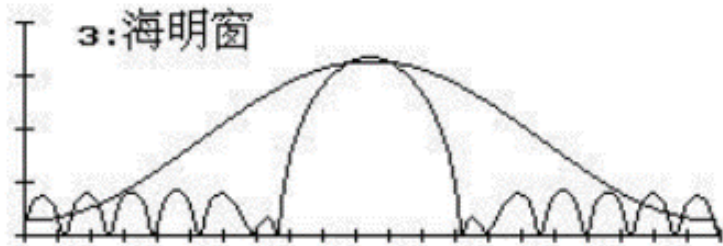
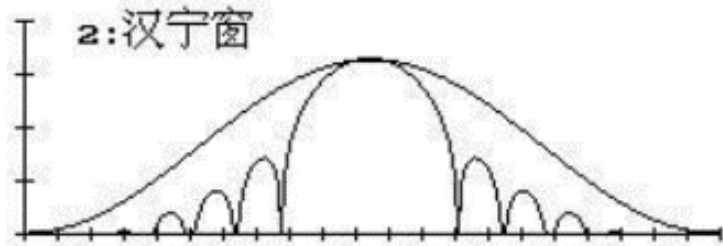
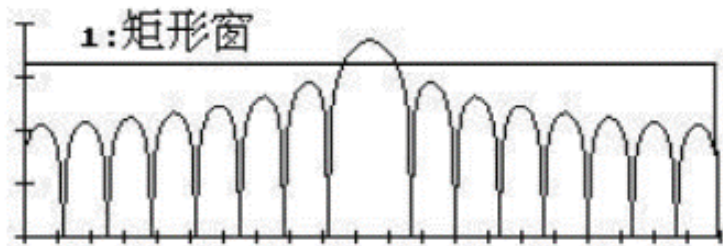
# 第三讲：语音信号短时分析

---

- 语音信号的特征参数
  - 短时平均幅度和能量
  - 短时平均过零率
  - 短时自相关函数
  - 短时平均幅度差

# 第三讲：语音信号短时分析

## □ 短时处理中的窗函数



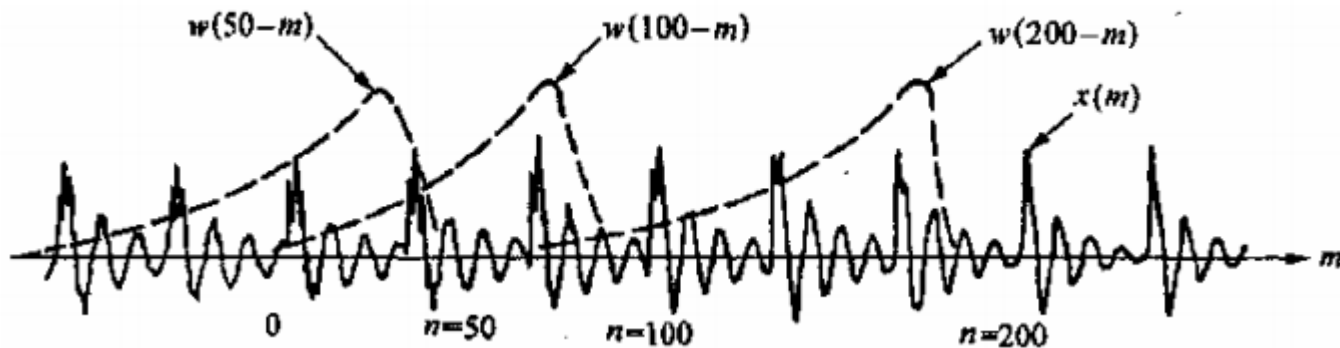


# 第三讲：语音信号短时分析

## □ 短时傅里叶变化定义

$$X_n(e^{j\omega}) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)e^{-j\omega m}$$

短时傅里叶变换有两个自变量： $n$  和  $\omega$ ；它既是时间  $n$  的离散函数，又是频率  $\omega$  的连续函数。





# 第四讲：线性预测模型

---

## □ 线性预测—语音生成的数字化建模

- 语音是由声道中的激励信号产生，它受声道、鼻腔和咽喉的形状而形成的共振峰调节。
  - 三种激励信号
    - 声带周期性开合而产生的声门波，声门波的开合周期决定了语音信号的基音频率。
    - 肺部收缩而产生的空气流。
    - 上述二者的组合。
  - 声道、鼻腔和咽喉中的声波共振由共振峰描述。
-



## 第四讲：线性预测模型

---

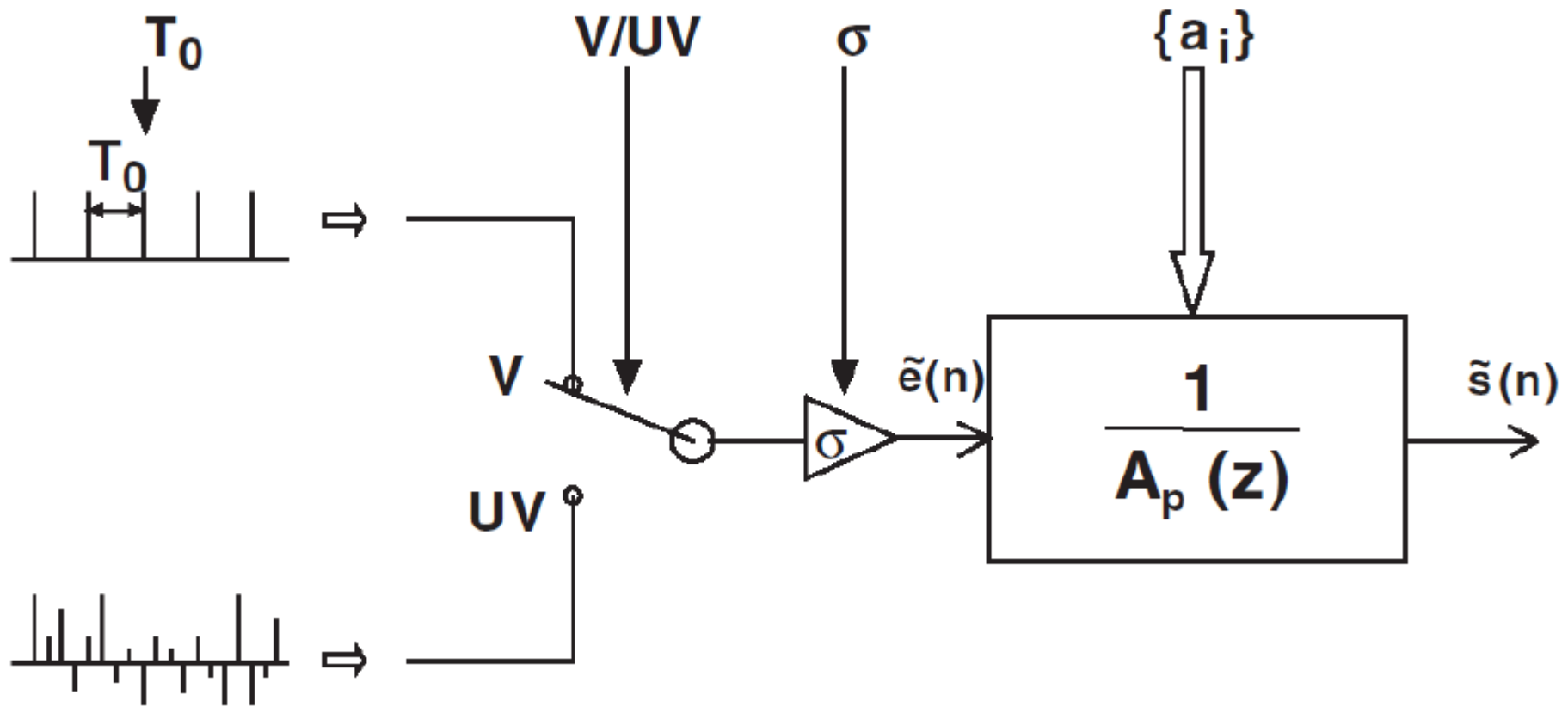
- 最初由Gunnar Fant于1960年代提出，采用一个线性模型描述语音生成的过程，这里声门波与声道形状完全分开考虑。
- 语音信号  $y[n]$  是全极点滤波器  $1/A(z)$  在激励信  $x[n]$  作用下的输出，传递函数描述如下：

$$Y(z) = X(z) \frac{1}{1 - \sum_{k=1}^p a_k z^{-k}} = X(z) \frac{1}{A_p(z)}$$

- 滤波器  $1/A_p(z)$  称为合成滤波器  $A_p(z)$  称为逆滤波器。
  - 激励信号要么是周期性的声门波，要么是类似于高斯白噪声的信号。
-



# 第四讲：线性预测模型





# 第五讲：语音增强

---

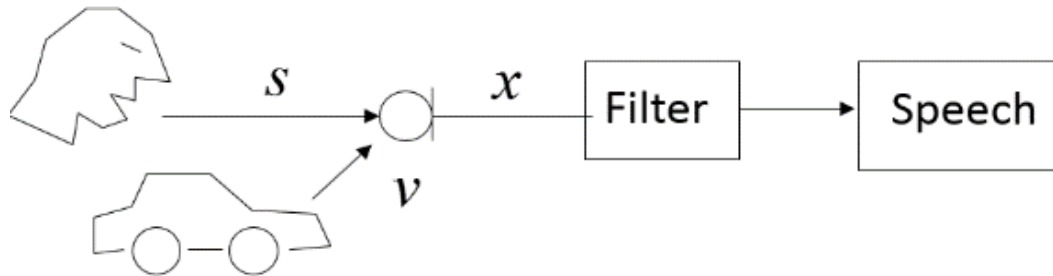
- **语音增强是指当语音信号被各种各样的噪声干扰、甚至淹没后，从噪声背景中提取有用的语音信号，抑制、降低噪声干扰的技术。简而言之，从带噪语音中提取尽可能纯净的原始语音。**
  
- **应用：**
  - 语音通信**
  - 人机交互**
  - 语音识别**
  - 助听器具**



# 第五讲：语音增强

## □ 单通道语音增强

- 单通道环境下，带噪语音的产生



- 信号模型

$$x(n) = s(n) + v(n)$$

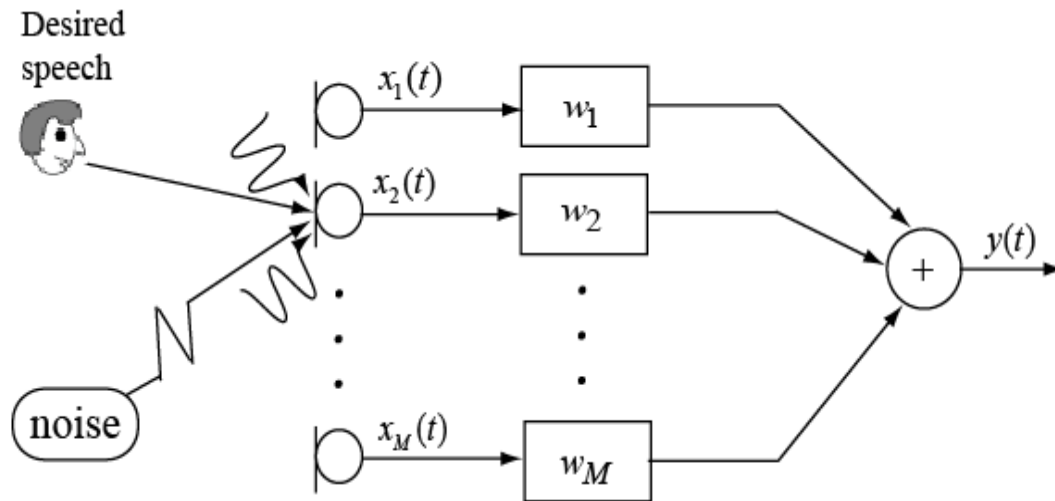
- 单通道语音增强方法:

- 谱减法
- 统计模型法 (LogMMSE, )
- 子空间法



# 第五讲：语音增强

## □ 多通道语音增强系统框架

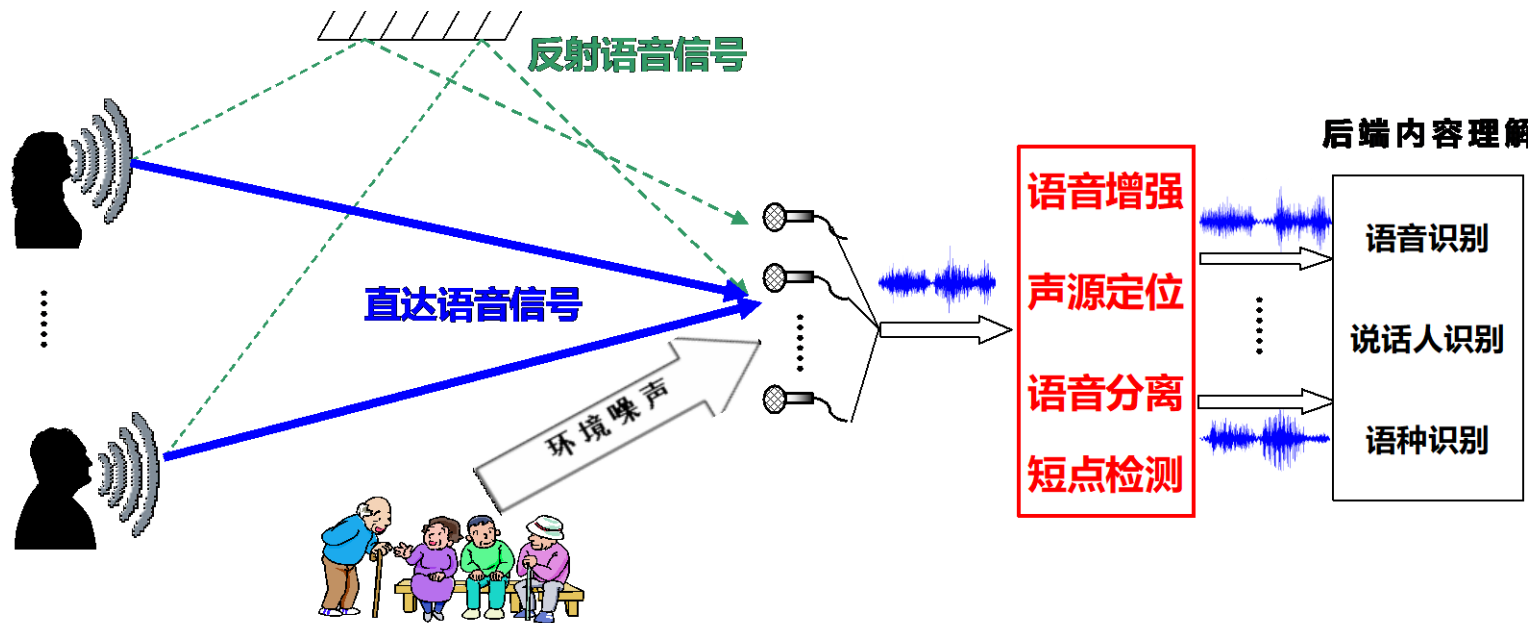


$$Y(k, \ell) = \sum_{i=1}^M W_i(k, \ell) \times X_i(k, \ell) = \underbrace{\sum_{i=1}^M W_i(k, \ell) \times A_i(k, \ell) \times S(k, \ell)}_{f(k, \ell)} + \underbrace{\sum_{i=1}^M W_i(k, \ell) \times N_i(k, \ell)}_{Y_v(k, \ell)}$$

- $f(k, \ell)$  = 语音成分的传递函数
- $Y_v(k, \ell)$  = 噪声成分

# 第六、七讲：多源语音场景分析

## □ 经典的多源语音场景



排除复杂声学环境对语音内容理解的干扰，降低这些干扰对性能的影响。



# 第六、七讲：多源语音场景分析

---

## □ 多源语音场景的若干要素

### ■ 直达麦克风的语音信号的传播

- 对相位产生一个加性的延迟，延迟由传播时间决定；
- 接收声源的幅度随着传播距离增加而衰减；

### ■ 早期混响

- 声波遇到障碍物后的前几次反射，与直达混响声的延迟比较短；
- 人耳可以辨别，很难将它与直达声波区分开来。

### ■ 晚期混响

- 声源声波在最初几次反射后，声波经由多次反射而逐渐直至消失的一段信号；
- 人耳不可辨别内容，由声学环境决定

### ■ 环境噪声

- 环境产生的嘈杂声，与目标信号加性叠加。

### ■ 多源语音信号的叠加混响

- 多个语音源信号的叠加
-



# 第六、七讲：多源语音场景分析

---

## □ 多源语音场景：问题与求解

### ■ 多路可观察信号——求解的出发点

□ 环境噪声 + 早期混响 + 晚期混响 + 多源语音叠加

### ■ 求解问题

□ 声源的绝对/相对传播时间（声源的空间位置）

□ 求解语音源出现的时间段（语音短点检测）

□ 求解语音源的数目（声源数目估计）

□ 还原各语音源信号（语音分离）

---



# 第六、七讲：多源语音场景分析

---

## □ 多源语音场景：难点与挑战

### ■ 混响的未知性

- 环境很难预知，如房间大小、形状、固体表面反射系数、声源与麦克几何位置等不可预知；
- 多路混响传递函数的盲估计，增加不确定性。

### ■ 环境噪声

- 环境噪声对部分语音频谱的掩蔽作用；
- 环境噪声的非稳定性，统计特征难以确定。

### ■ 声源位置的确定

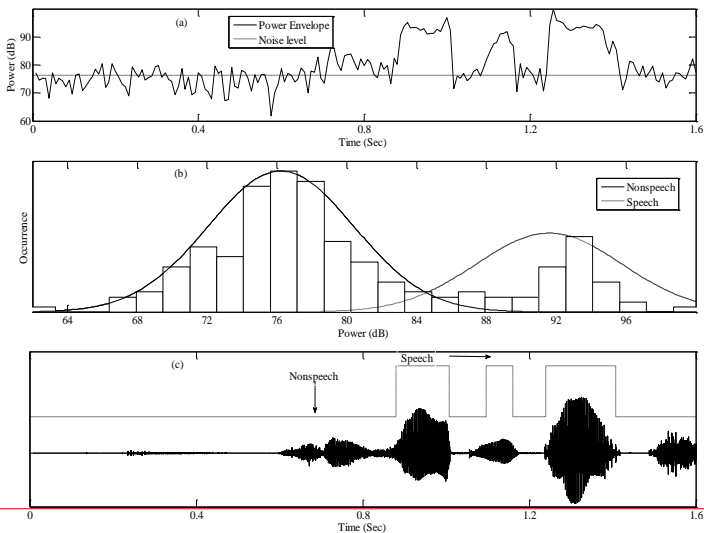
- 声源位置接近而难以区分
- 声源的运动导致位置的连续变化



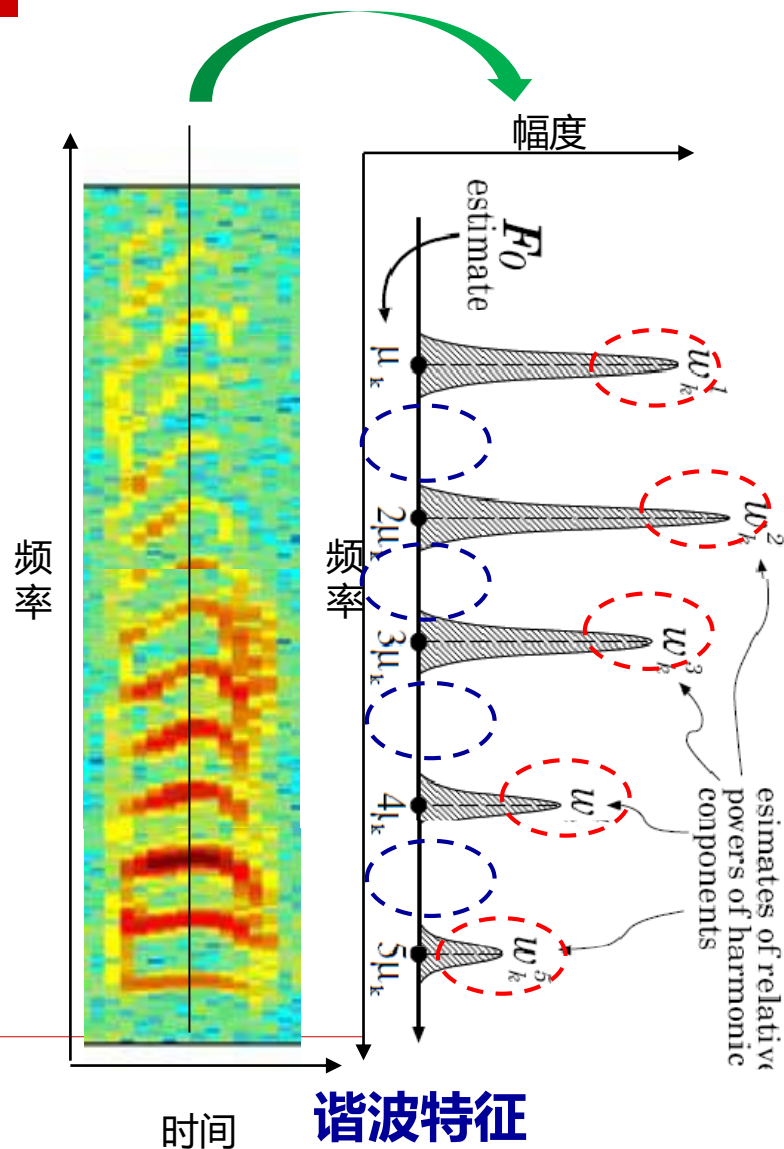
# 第八讲：语音活动性检测

## 基于短时声学特征的方法

- ◆ 谱熵
- ◆ 过零率
- ◆ 能量
- ◆ 谐波结构



短时能量特征

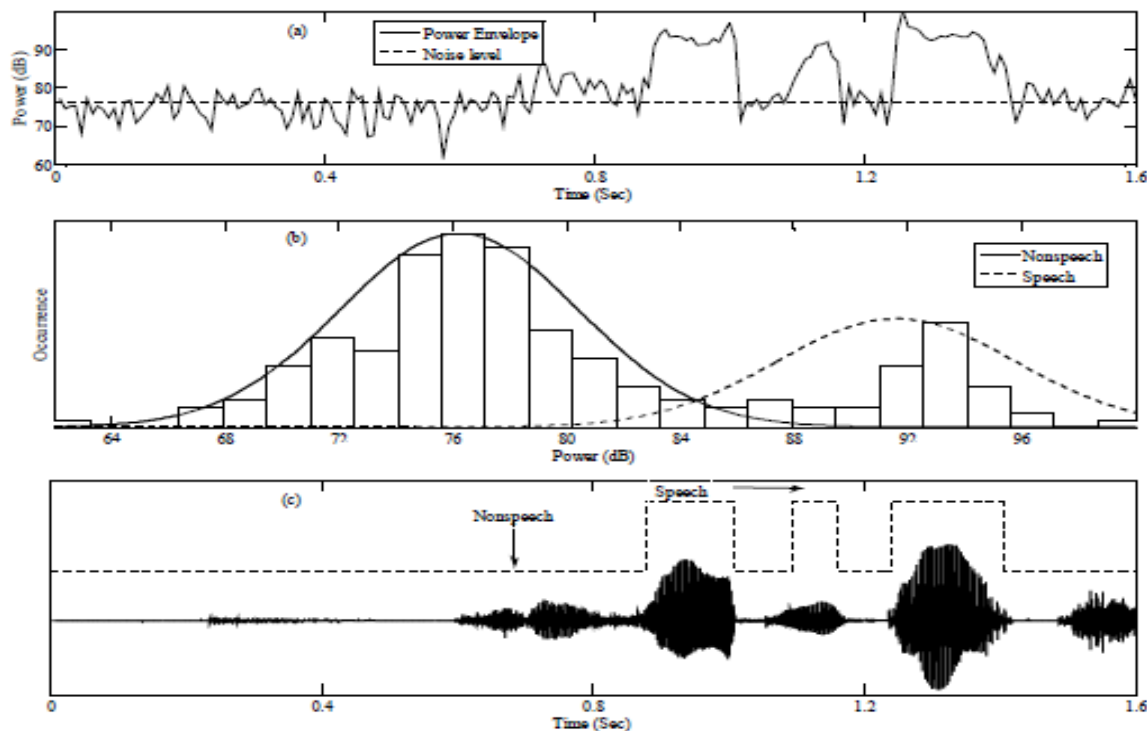


谐波特征

# 第八讲：语音活动性检测

## □ 基于子带能量的方法

- 由于语音是纯净语音和噪声的叠加，所以语音均值大于非语音均值
- 由于通常假设噪声比语音信号更加稳定，所以语音方差大于非语音方差





# 第九讲：三维音频

---

- **三维音频技术**通过计算机、信号处理等技术实现对真实世界中**声音事件**及其**三维声场信息**的**获取、处理与重放**。使声音具有强烈的空间感、包围感和沉浸感，让人产生“身临其境”的感受。
  
  - **研究内容：**
    - 三维音频采集
    - 三维音频传输
    - 三维音频重放
-





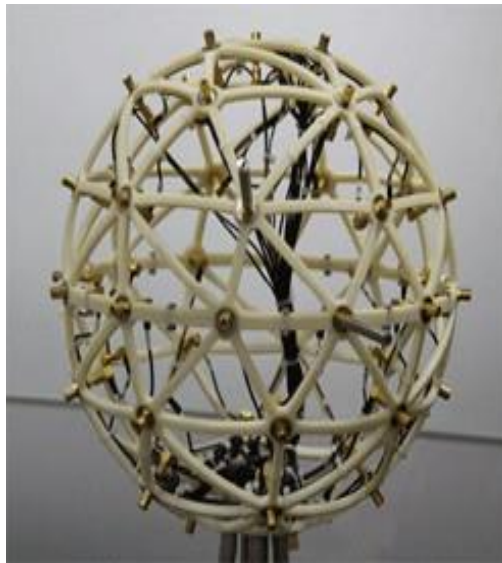
# 第九讲：三维音频

---

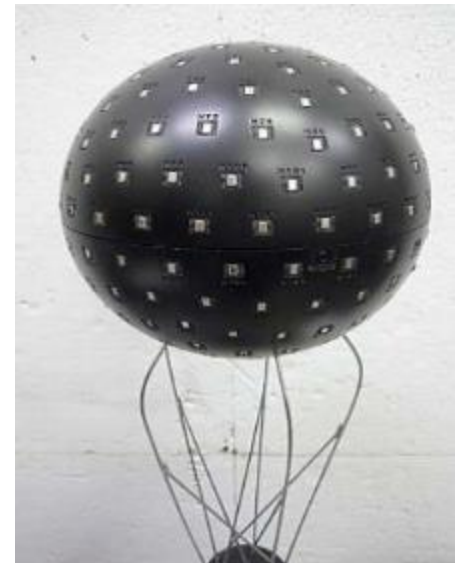
## 三维音频采集



双耳录音



球型阵列录音



球型阵列录音



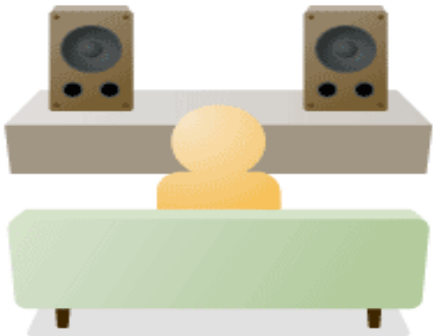
# 第九讲：三维音频

---

## 三维音频重放



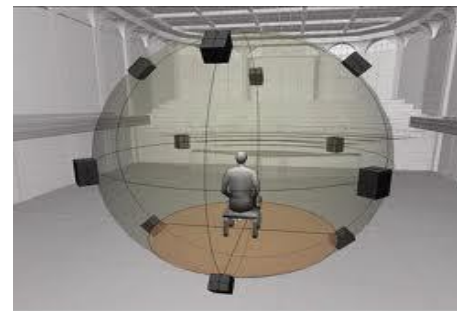
耳机重放



双扬声器重放



多通道重放



Ambisonics重放



---

# 语音信号处理的应用



# 各讲主题

---

## □ 基础知识

- 第二讲：语音生成与听觉感知

## □ 语音信号处理基础

- 第三讲：语音信号短时分析
- 第四讲：语音信号线性预测

## □ 语音信号处理算法

- 第五讲：语音增强原理与应用
- 第六讲：多源语音信号分离
- 第七讲：多源语音定位
- 第八讲：语音活动性检测
- 第九讲：三维音频

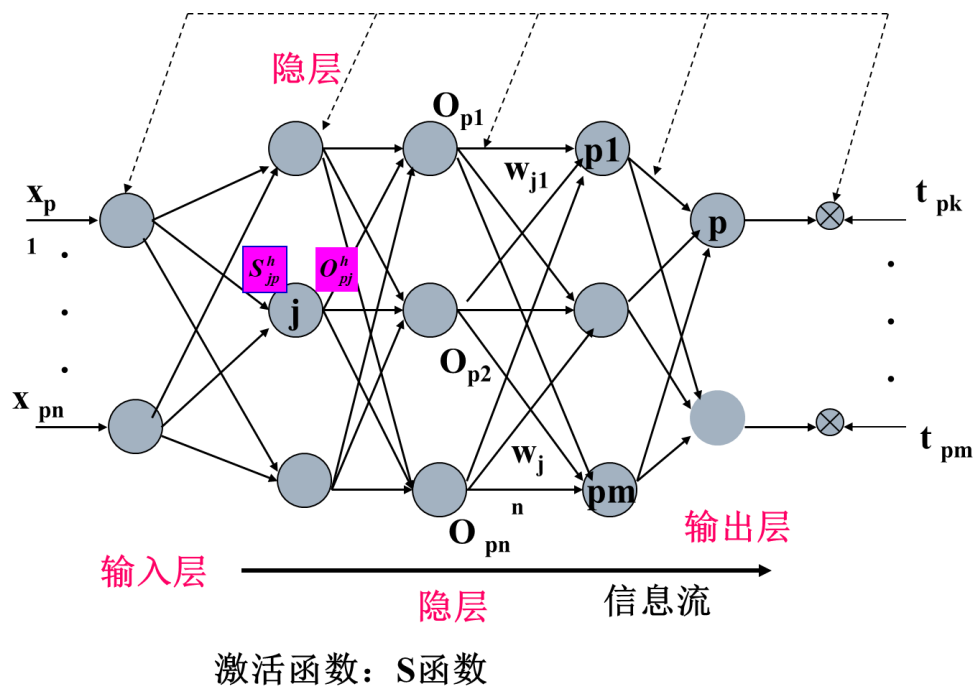
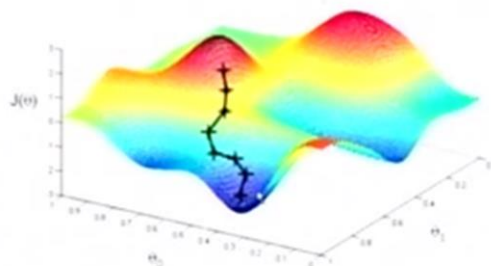
## □ 语音/音频应用

- 第十讲：深度神经网络在语音处理中的应用
- 第十一讲：语音识别基础原理及应用
- 第十二讲：鲁棒语音识别
- 第十三讲：说话人识别与语种识别
- 第十四讲：语音合成技术
- 第十五讲：复习与答疑
- 第十六讲：考试

# 第十讲：深度神经网络在语音处理中的应用

- 基本概念
- 梯度下降算法
- BP神经网络
- 递归神经网络
- 卷积神经网络

Gradient Descent





# 第十讲：深度神经网络在语音处理中的应用

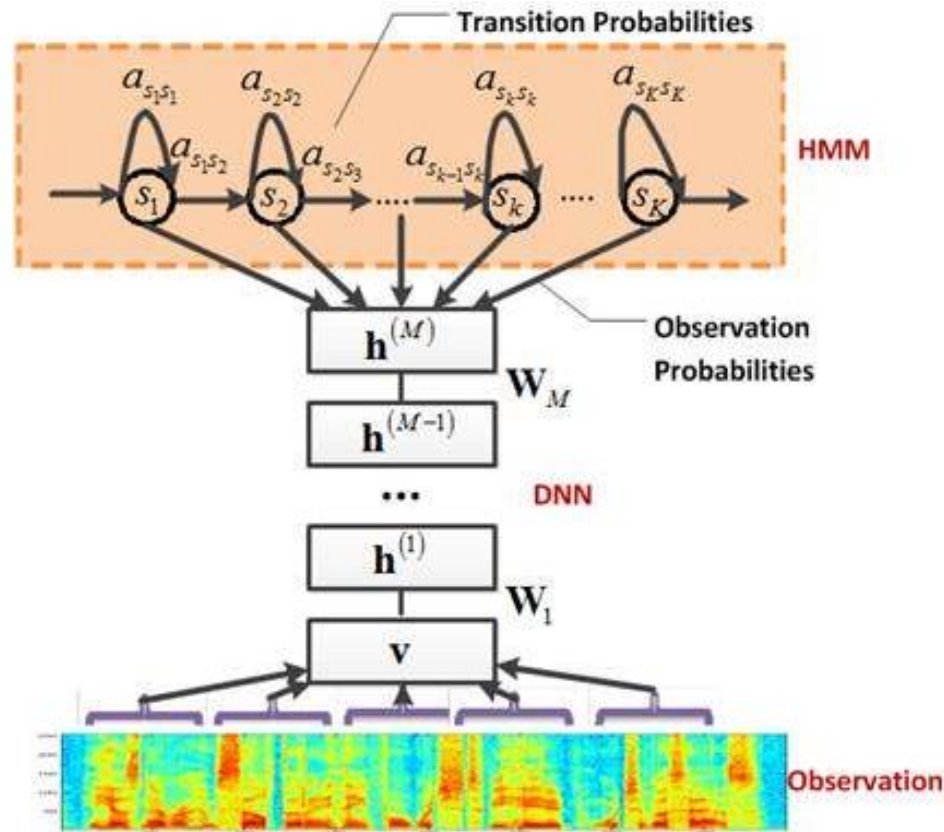
---

## □ 神经网络 (Neural Network, NN)

- **生物神经网络**主要是指人脑的神经网络，它是人工神经网络的技术原型。人脑是人类思维的物质基础，思维的功能定位在大脑皮层，后者含有大约 $10^{11}$ 个神经元，每个神经元又通过神经突触与大约 $10^3$ 个其它神经元相连，形成一个高度复杂高度灵活的动态网络。作为一门学科，生物神经网络主要研究人脑神经网络的结构、功能及其工作机制，意在探索人脑思维和智能活动的规律。
- **人工神经网络**是生物神经网络在某种简化意义下的技术复现，作为一门学科，它的主要任务是根据生物神经网络的原理和实际应用的需要建造实用的人工神经网络模型，设计相应的学习算法，模拟人脑的某种智能活动，然后在技术上实现出来用以解决实际问题。因此，生物神经网络主要研究智能的机理；人工神经网络主要研究智能机理的实现，两者相辅相成。

# 第十讲：深度神经网络在语音处理中的应用

## □ 深度神经网络架构





# 第十讲：深度神经网络在语音处理中的应用

---

## □ 传统神经网络的学习算法

- 前馈型神经网络学习算法
- 反馈型神经网络学习算法
- 自组织神经网络学习算法





# 第十讲：深度神经网络在语音处理中的应用

---

## □ 几种典型的神经网络

- NN (前向神经网络)
- RNN (递归神经网络)
- BRNN (双向递归神经网络)
- LSTM/LSTM-DBRNN (长短时记忆-双向递归神经网络)
- CNN (卷积神经网络)



# 第十讲：深度神经网络在语音处理中的应用

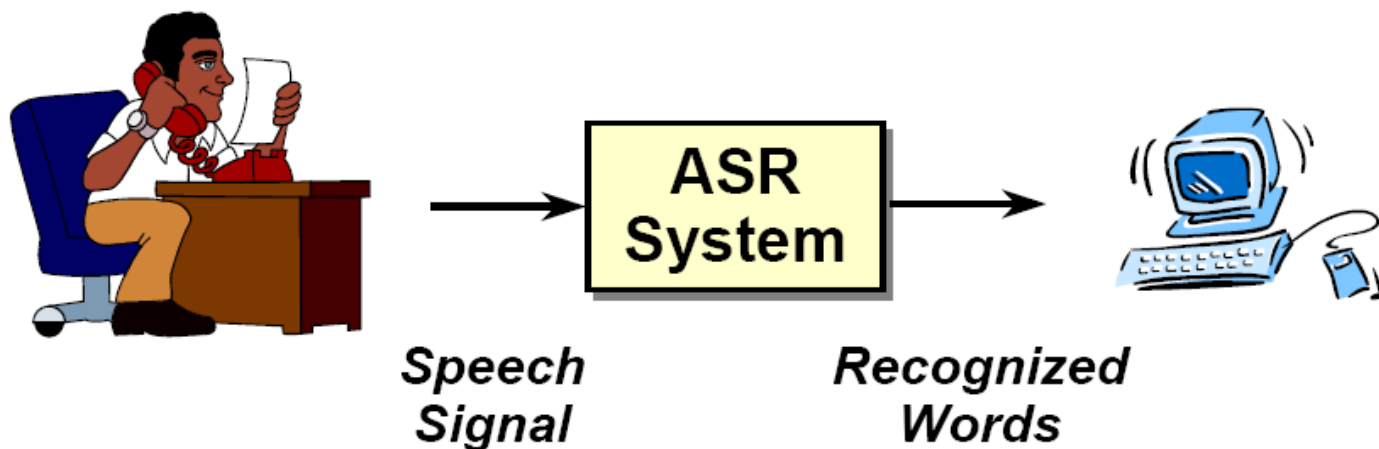
---

## □ 在语音处理中的应用

- 基于DNN的声学模型建模
- 基于DNN的语言模型建模
- 基于DNN的语音端点检测
- 基于DNN的情感识别
- 基于DNN的叠音检测
- 基于DNN的语音增强
- .....

# 第十一讲：语音识别基础原理及应用

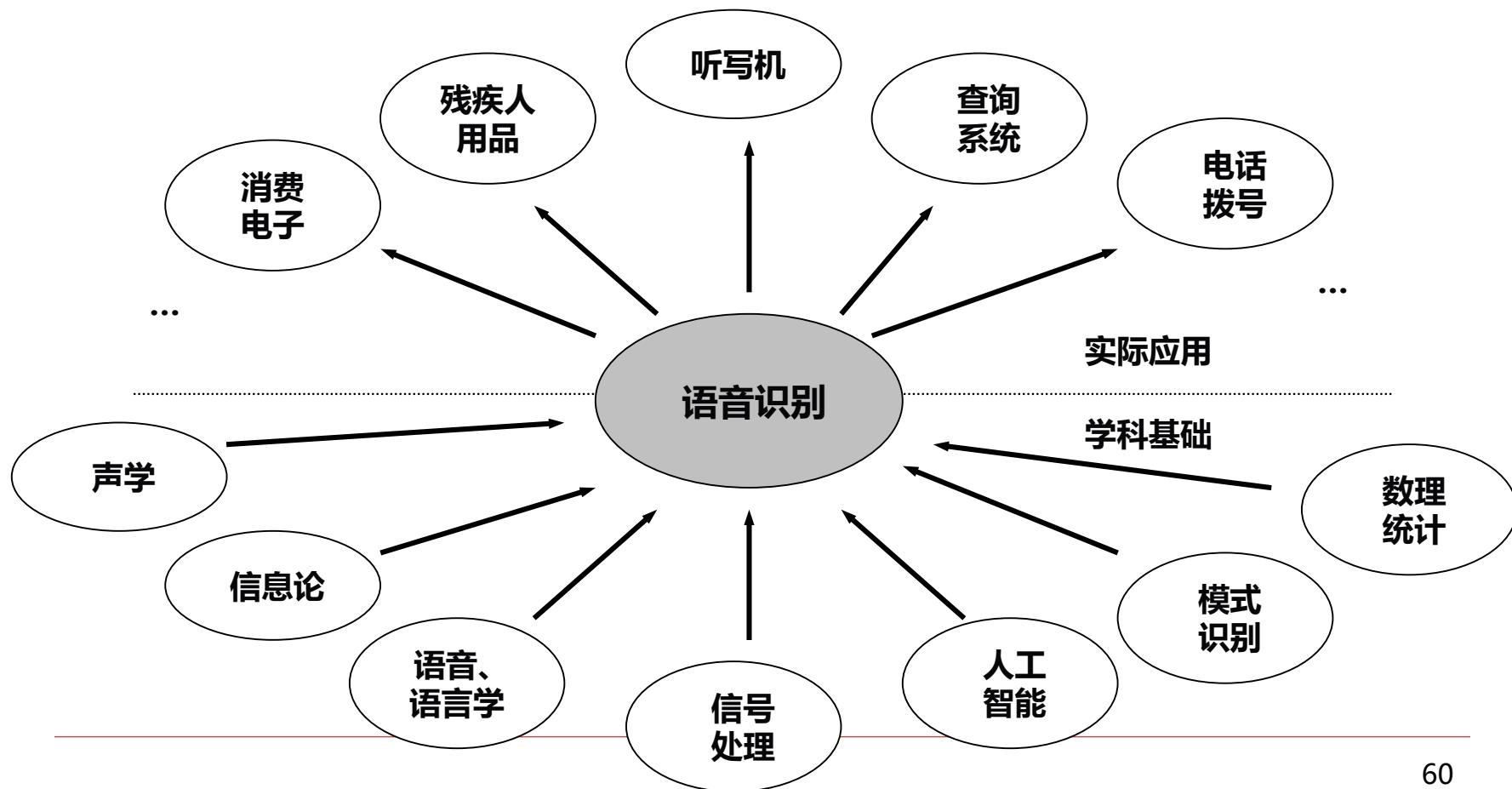
## □ 语音识别 (Automatic Speech Recognition)





# 第十一讲：语音识别基础原理及应用

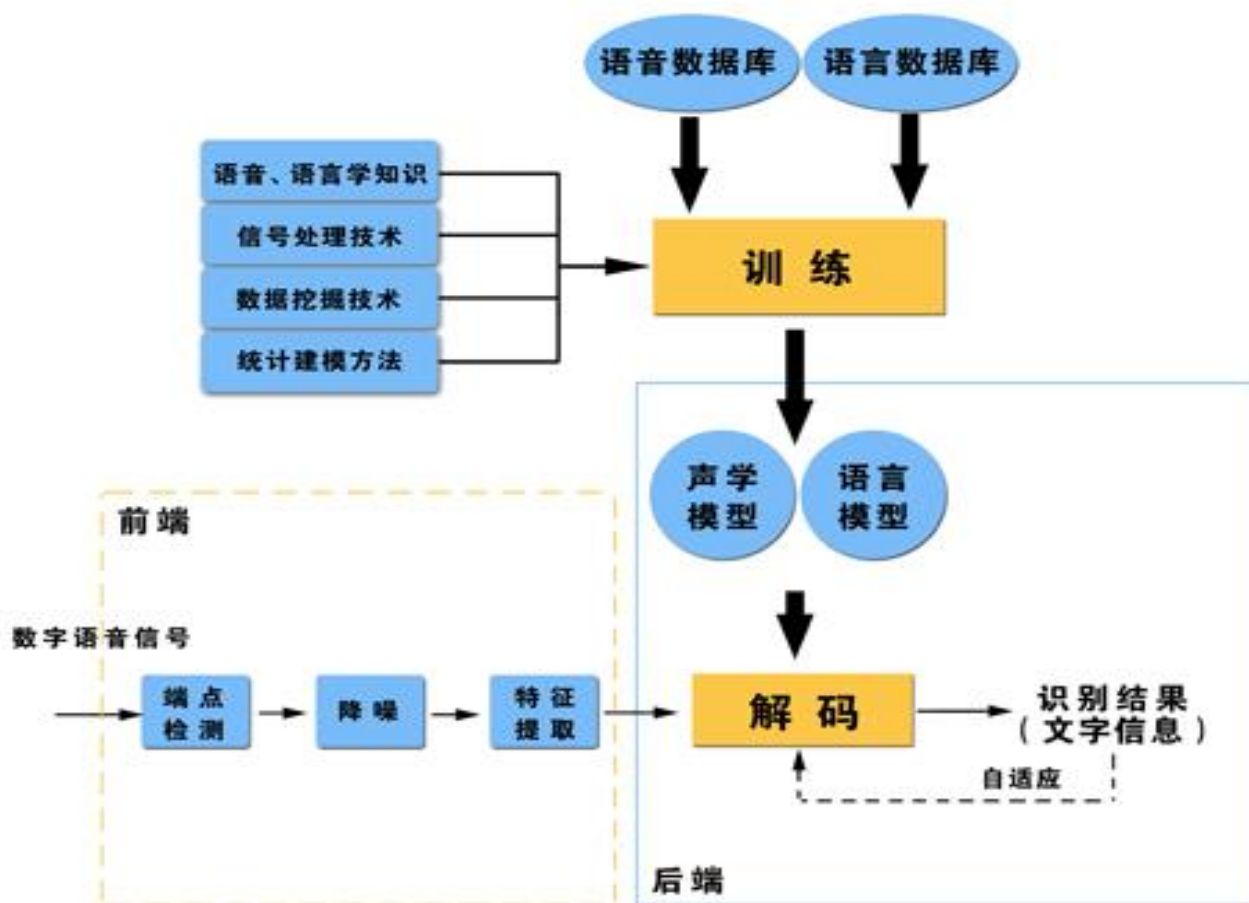
## □ 语音识别的实际应用和学科基础





# 第十一讲：语音识别基础原理及应用

## □ 语音识别：基本框架





# 第十一讲：语音识别基础原理及应用

## □ 现有识别系统

识别过程是基于声学模型、语言模型、发音词典、语音信号找到一个使后验概率最大化的词串作为识别结果

$$= \sum_{k=0}^i c_k \frac{1}{\sqrt{2\pi}\sigma} e^{-(o_t - \mu)^2 / 2\sigma^2}$$

$$= \prod_{k=0}^{M-1} P(w_k | W_{k-N+1} \dots W_{k-1})$$

$$F = \arg \max_i \frac{P_\lambda(O(u) | w_i(u)) P(w_i(u))}{P(O(u))}$$

建模单元的选择  
概率模型的选择  
模型的训练方法



用户语音

前端处理

声学模型

语言模型

解码器

识别结果

搜索空间的构建  
快速算法的研究

发音词典

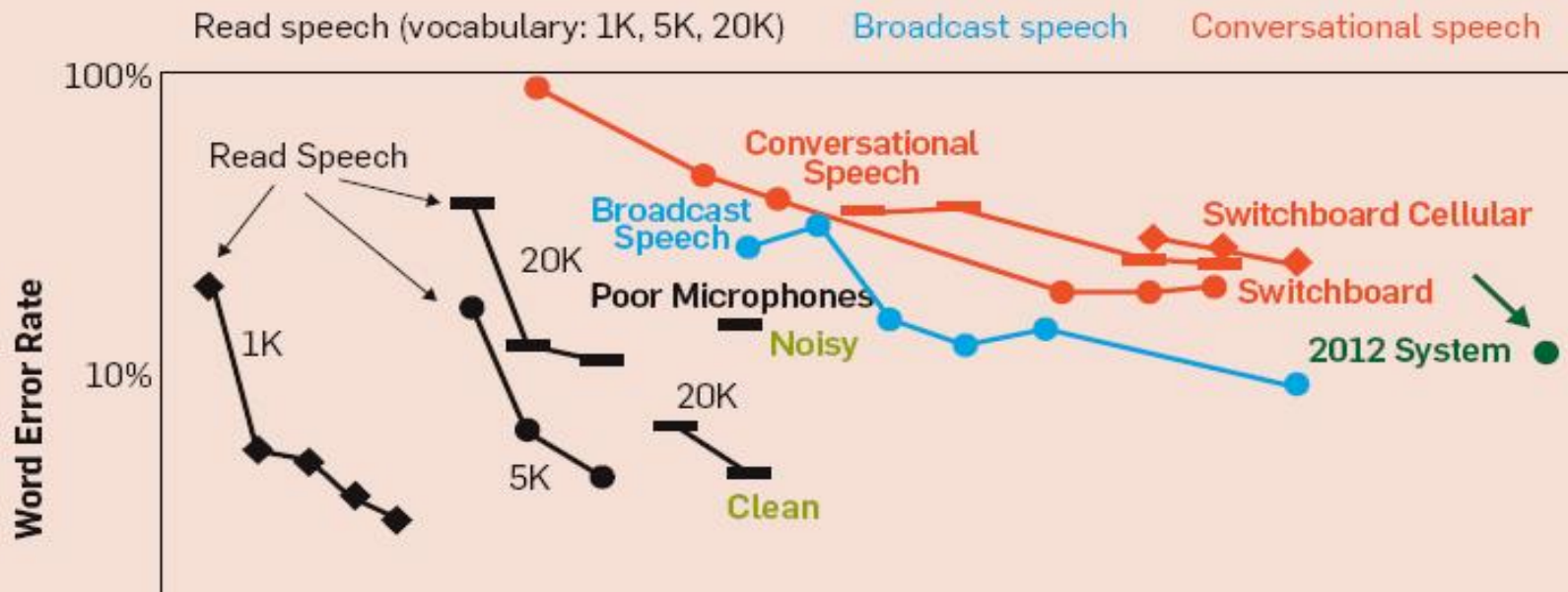
中国:  
zh ong1 g  
guo2  
人民:  
r en2 m in2

特征矢量:  $O(u) = \{O_1, O_2, \dots, O_T\}$





# 语音技术的进展与识别性能



**错误率下降一定程度后遭遇瓶颈！**  
**与人的能力相比（2%到4%）差距很大！**

数据来源: 美国National Institute of Science & Technology



# 第十一讲：语音识别基础原理及应用

---

## □ 主要挑战

- 鲁棒性 (Robustness)
- 自适应 (Adaptation)
- 语言模型 (Language Modeling)
- 可信度度量 (Confidence Measuring)
- 集外词 (Out-of-Vocabulary Words)
- 韵律信息 (Prosody) 的利用





# 第十二讲：鲁棒语音识别

---

- 影响语音识别性能的环境变化因素
- 噪声环境下的鲁棒语音识别技术（前端）
- 自适应技术（后端）

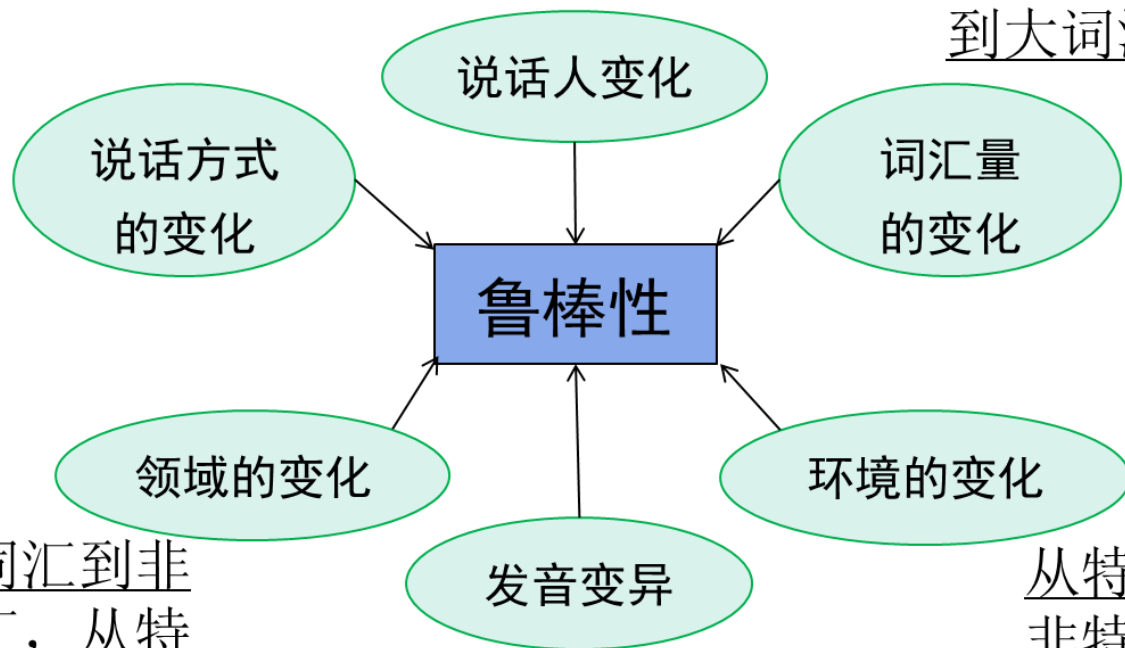


# 第十二讲：鲁棒语音识别

从孤立词识别到连续语音识别

从特定说话人到非特定说话人

从小词汇量任务到大词汇量任务



从特定词汇到非特定词汇，从特定领域文法到非特定领域文法

话者由于受到生理、心理、情感等影响而产生的发音变化

从特定环境到非特定环境



# 第十二讲：鲁棒语音识别

---

## □ 鲁棒语音识别技术

### ■ 基于语音增强方法

- ✓ 谱减的方法
- ✓ 维纳滤波

### ■ 通道畸变的抑制方法

- ✓ 倒谱均值减CMS方法
- ✓ 二级CMS的方法
- ✓ RASTA-PLP技术

### ■ 基于模型的补偿方法

- ✓ 多重风格训练
- ✓ HMM分解
- ✓ 并行模型融合PMC
- ✓ 矢量泰勒级数VTS方法
- ✓ 基于最小分类错误准则的补偿
- ✓ 基于最小分类错误的环境特征学习



# 第十二讲：鲁棒语音识别

---

## □ 自适应技术

### ■ 基于最大后验概率的方法 (Maximum a posteriori, MAP)

基本准则是后验概率最大化，利用Bayes学习理论将SI模型的先验信息与目标说话人的信息相结合实现自适应

### ■ 基于变换的方法

估计SI系统模型与目标说话人之间的空间影射关系，利用SI模型参数和变换矩阵实现模型自适应

### ■ 基于聚类的方法

用一组HMMs进行自适应的方法



# 第十三讲：说话人识别与语种识别

---

## □ 说话人识别

是一种根据语音中反映的代表说话人生理和行为特征的语音参数，来自动识别说话人身份的技术。

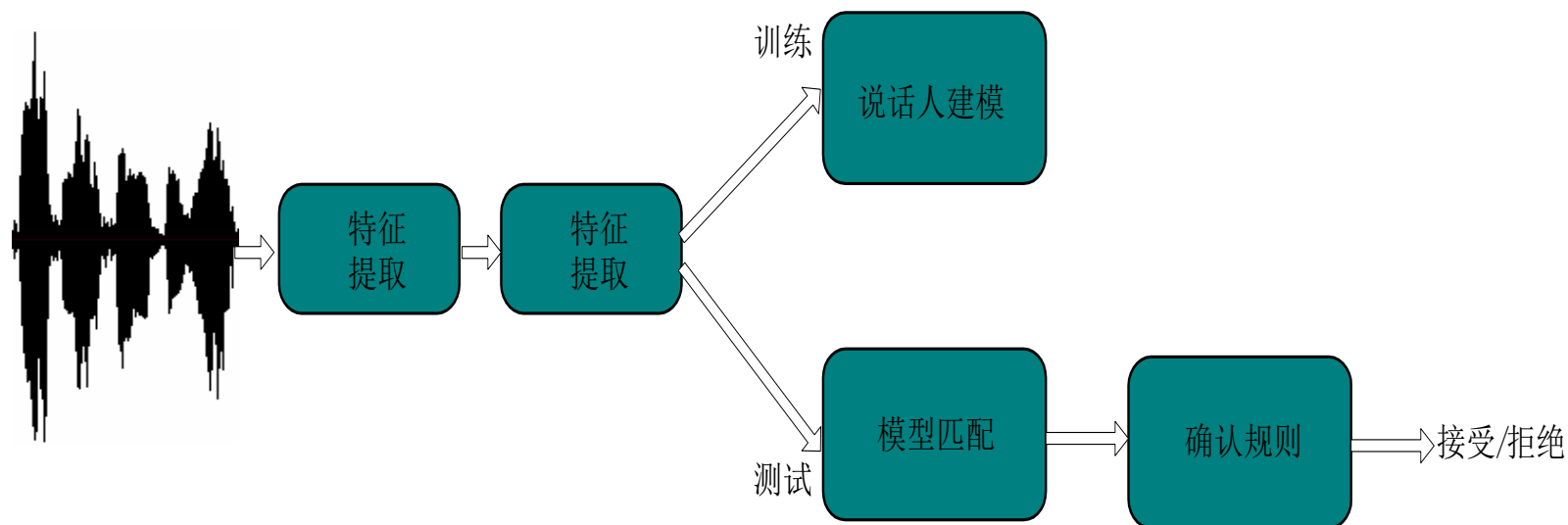
## □ 说话人识别步骤：注册、测试

## □ 说话人识别的应用

- 司法鉴定
- 军事应用
- 信息安全
- 机密场所的门禁系统
- 语音检索

# 第十三讲：说话人识别与语种识别

## □ 说话人识别系统原理



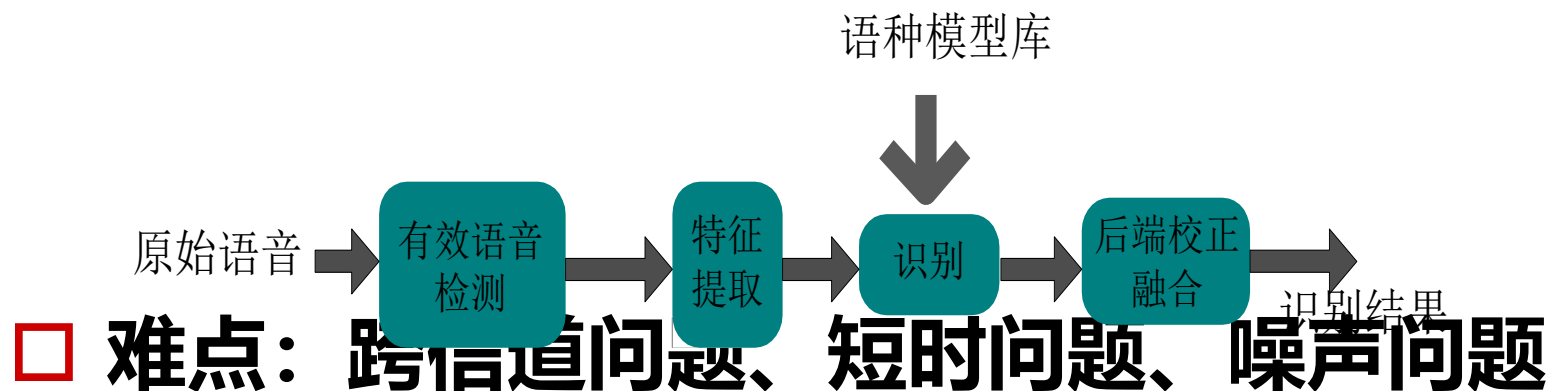
## □ 难点：跨信道、短时、噪声、合录语音



# 第十三讲：说话人识别与语种识别

□ **语种识别**：又称为语种辨识，它是通过分析处理一个语音片段，来判断其属于某个语言种类。其本质是语音识别的一个方面。

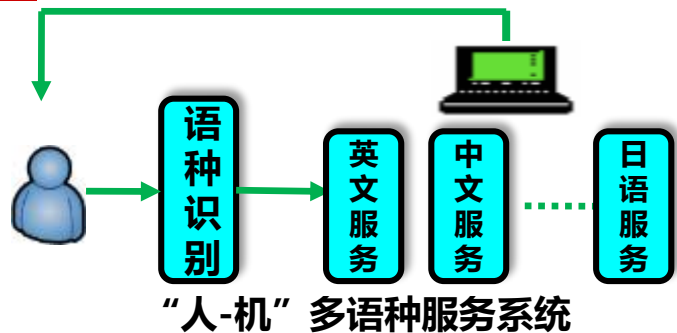
□ **基本原理**



# 第十三讲：说话人识别与语种识别

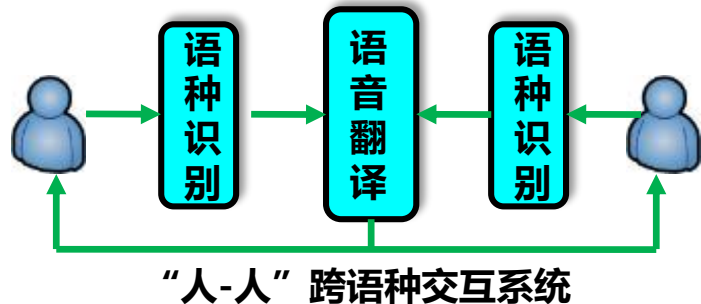
## ■ “人-机” 多语种服务系统

- 自动识别用户语种信息，启动相应自动服务系统



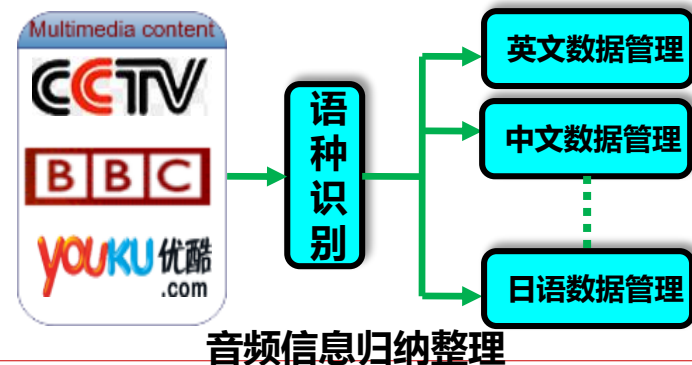
## ■ “人-人” 跨语种交互系统

- 识别对话人语种信息，启动自动语音翻译系统



## ■ 音频信息检索

- 识别音频文件语种信息，进行音频文件分类整合







# 第十四讲：语音合成技术

---

## □ 波形合成法一般分为三种：

- 波形编码合成
- 波形编辑合成
- 端到端语音合成

## □ 他们有相同前端：

- 文本规整、词的切分



# 第十四讲：语音合成技术

系统	优点	缺点
波形拼接	音质好，自然度高	对数据要求高，系统构建复杂，磁盘占用高，合成语音不连贯
统计参数	流畅度高，易懂度高，灵活度高，系统尺寸小，构建系统迅速	自然度不够，过于平滑
端到端	自然度高，流畅度高	技术不成熟，训练过程复杂，对训练数据要求高，合成速度较慢



---

# 课程安排



# 授课

---

## □ 团队组成

- 教师：李军锋
- 助教：邱泽林

## □ 时间：周四，13:30-16:20 （第5-7节）



# 考核

---

- ☐ 期末考试
- ☐ 平时作业
- ☐ 出勤



# 联系

---

## □ 课程联系与讨论微信群





# 参考书目（无指定教科书）

---

- ❑ Jacob Benesty, M. Mohan Sondhi, and Yiteng (Arden) Huang. *Springer Handbook of Speech Processing*. Springer-Verlag New York, Inc., 2007.
- ❑ Lawrence R. Rabiner and Ronald W. Schafer. *Theory and Applications of Digital Speech Processing*. 电子工业出版社, 2011.
- ❑ Thomas F. Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practices*, Prentice Hall, 2001.
- ❑ Mark Tatham and Katherine Morton, *A Guide to Speech Production and Perception*, Edinburgh University Press Ltd. 2011.
- ❑ 韩纪庆, 张磊, 郑铁然。语音信号处理(第2版)。清华大学出版社, 2013年。
- ❑ 赵力。语音信号处理(第2版)。机械工业出版社, 2011年。
- ❑ 杨行峻, 迟惠生。语音信号数字处理。电子工业出版社。1995年。



**谢谢！**