

# 人工智能程序设计

# python

```
import turtle  
turtle.setup(650,350,200,200)  
turtle.penup()  
turtle.fd(-250)  
turtle.pendown()  
turtle.pensize(25)  
turtle.color("purple")  
for i in range(4):  
    turtle.circle(40, 80)  
    turtle.circle(-40, 80)  
    turtle.circle(40, 80/2)  
    turtle.circle(-40, 80/2)  
    turtle.fd(40)  
    turtle.circle(16, 180)  
    turtle.circle(16 * 2/3)  
    turtle.fd(40 * 2/3)
```



# 人工智能程序设计

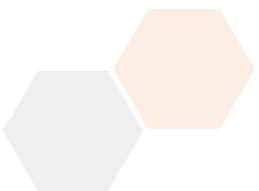
## 15.1 语音技术基础

北京石油化工学院 人工智能研究院

刘 强

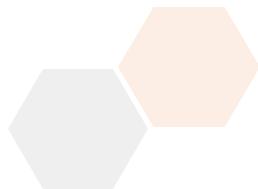
# 学习内容

- 语音技术的发展历程
- 现代语音技术的核心任务
- 语音技术的应用领域



# 语音技术概述

语音技术是人工智能的重要分支，从传统信号处理发展到深度学习时代，实现了质的飞跃。掌握语音技术的发展脉络、核心任务和应用场景，是构建现代语音处理系统的基础。



# 语音技术的发展历程

语音技术经历了三个重要发展阶段：

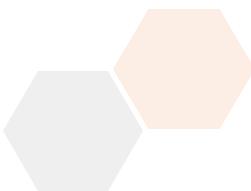
| 阶段     | 时期          | 特点                               |
|--------|-------------|----------------------------------|
| 早期阶段   | 1950s-1980s | 模板匹配、动态时间规整，只能处理特定说话人的有限词汇       |
| 统计模型时代 | 1980s-2000s | 隐马尔可夫模型、高斯混合模型，实现连续语音识别          |
| 深度学习时代 | 2010年至今     | DNN、RNN、LSTM、Transformer，性能突破性提升 |



# 深度学习带来的突破

## 为什么深度学习能带来突破性进展？

- 端到端学习：直接从原始数据学习特征，无需人工设计
- 表征能力强：深层网络可以捕获复杂的语音模式
- 大规模数据：利用海量语音数据进行训练
- 计算能力提升：GPU加速使复杂模型训练成为可能



# 现代语音技术的核心任务

现代语音技术包含四个核心任务：

| 任务    | 功能描述                        |
|-------|-----------------------------|
| 语音识别  | 将语音转换为文本，基于端到端深度学习模型        |
| 语音合成  | 将文本转换为自然语音，涉及文本分析、参数预测、波形生成 |
| 说话人识别 | 通过语音信号识别说话人身份，分为确认和辨认两种模式   |
| 语音增强  | 去除噪声干扰，提高语音质量               |



# 语音识别技术

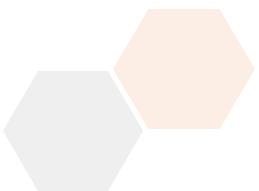
**语音识别** (Speech Recognition) 将语音转换为文本。

## 技术特点：

- 现代系统基于端到端深度学习模型
- 支持连续语音识别
- 高精度、低延迟

## 应用场景：

- 语音输入法
- 语音搜索
- 智能客服



# 语音合成技术

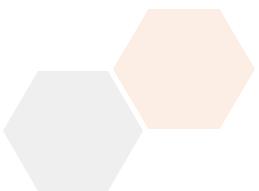
**语音合成** (Text-to-Speech, TTS) 将文本转换为自然语音。

**处理流程：**

1. 文本分析：分词、韵律预测
2. 参数预测：预测声学参数
3. 波形生成：生成语音波形

**应用场景：**

- 有声读物
- 导航播报
- 智能客服



# 说话人识别技术

**说话人识别**通过语音信号识别说话人身份。

**两种模式：**

- **确认模式**：验证声称的身份是否正确
- **辨认模式**：从多个候选人中识别说话人

**应用场景：**

- 声纹解锁
- 安全认证
- 电话银行



# 语音增强技术

语音增强去除噪声干扰，提高语音质量。

## 技术方法：

- 谱减法
- 维纳滤波
- 深度学习降噪

## 应用场景：

- 嘈杂环境通话
- 语音会议系统
- 助听设备



# 语音技术的应用领域

| 应用领域   | 具体应用            |
|--------|-----------------|
| 智能语音助手 | Siri、Alexa、小爱同学 |
| 语音转录服务 | 会议记录、客服质检、媒体字幕  |
| 语音翻译系统 | 跨语言实时交流         |
| 无障碍技术  | 语音导航、屏幕阅读、语音转文字 |



# 智能语音助手

**代表产品:** Siri、Alexa、Google Assistant、小爱同学

**核心功能:**

- 语音唤醒
- 自然语言理解
- 多轮对话
- 任务执行



# 语音转录与翻译

## 语音转录服务：

- 会议记录自动化
- 客服质检分析
- 媒体字幕生成

## 语音翻译系统：

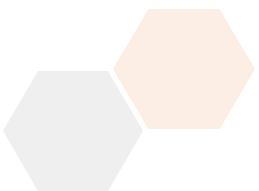
- 结合语音识别、机器翻译、语音合成
- 实现跨语言实时交流



# 无障碍技术应用

为视障和听障人士提供辅助功能：

- **视障辅助：**语音导航、屏幕阅读
- **听障辅助：**语音转文字、实时字幕
- **语音控制：**智能家居语音控制



# 实践练习

## 练习 15.1.1：语音技术发展理解

分析语音技术三个发展阶段的特点和局限性，说明深度学习为什么能够带来突破性进展。



# 实践练习

## 练习 15.1.2：核心任务对比分析

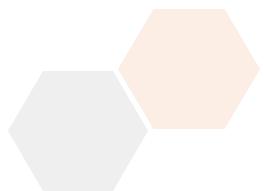
比较语音识别、语音合成、说话人识别、语音增强四个核心任务的技术特点和应用场景。



# 实践练习

## 练习 15.1.3：应用场景调研

选择一个语音技术应用领域，调研其技术实现方案和面临的挑战。



# 实践练习

## 练习 15.1.4: Python环境准备

安装语音处理相关的Python库（如librosa、pydub、pyaudio），为后续实践做准备。

