



浙江大学 计算机学院
数字媒体与网络技术



群名称：数字视音频处理-数2019媒
群 号：398550295

数字语音处理

浙江大学计算机学院

杨莹春

yyc@zju.edu.cn, QQ:1169244241

QQ群：数字视音频处理-数2019媒（398550295）

验证信息/群名片：姓名学号口音



浙江大学外经贸楼520

2019年9月11日



讲述提纲

- 语音研究简介
- 课程内容安排
- 语音技术基础





语言是衡量人类进化与智力水平高低的标志

语言是人类特有的一种交流工具,不仅承载了人类对物质世界的反映,而且促进了人类对物质世界的再创造.人类的发展离不开语言,语言的进化与人类的进化发展历程相伴随。





人类关于语言的千年梦想

IBM近日宣布，有八所大学与 IBM 研究人员正在合作推动沃森计算机系统背后“问答”技术的开发，该系统在2011年2月14-16日播出的智力竞赛节目Jeopardy!《危险边缘》中与人类一较高下。麻省理工学院（MIT）、德州大学奥斯汀分校、南加州大学（USC）、伦斯勒理工学院（RPI）、纽约州立大学阿尔巴尼分校（UAlbany）、特兰托大学（意大利）、马萨诸塞大学安姆斯特分校以及卡内基梅隆大学。





人类关于语言的千年梦想

The implications of progress in A.I. are being brought into sharp relief now by the broadcasting of a recorded competition pitting the I.B.M. computing system named Watson against the two best human Jeopardy players, Ken Jennings and Brad Rutter.

The real value of Watson may ultimately be in forcing society to consider where the line between human and machine should be drawn.





语言是衡量人类进化与智力水平高低的标志

人类关于语言的千年梦想

现实社会的发展需求

DEMO

Universal Speech Interface CMU

C-Star Travel agent CMU

MonkeyBiz 2006 MIT MEDIA LAB



经典书籍



1. Xuedong Huang, Alex Acero , Hsiao-Wuen Hon, Spoken Language Processing : A Guide to Theory, Algorithm and System Development, 1008 pages Prentice Hall; 1 edition (May 5, 2001)
2. Daniel Jurafsky , James H. Martin, Speech and Language Processing (2nd Edition) , 1024 pages , Pearson Prentice Hall; 2 edition (May 26, 2008)
3. Lawrence R. Rabiner, Ronald W. Schafer, Theory and Applications of Digital Speech Processing , 1056 pages, Pearson, 2010



重要会议



- ICASSP (International Conference on Acoustic, Speech and Signal Processing)
每年一届，10月截稿，次年5月开会。
- Interspeech
 - ICSLP (International Conference on Spoken Language Processing)
偶数年举办，4月截稿，9月开会
 - EuroSpeech (European Conference on Speech Communication and Technology)
奇数年举办，4月截稿，9月开会



主要杂志



- Speech Communication
- Computer Speech and Language (CSL)
- IEEE Transactions on Audio, Speech and Language Processing (IEEE ASLP)



技术评测



- NIST Spoken Language Technology Evaluations Benchmark Tests
<http://www.nist.gov/speech/tests/index.htm>
- 中国863语音技术评测
<http://www.863data.org.cn/2005eval.php>





讲述提纲

- 语音研究简介
- 课程内容安排
- 语音技术基础



教学内容



数字语音处理

(1) 语音分析技术:

主要包括语音产生基础知识、语音产生模型、时域分析技术、频域分析等;

(2) 语音识别、编码与合成技术:

研究探索主题之一

语音识别与合成技术(NIST评测)





数字语音处理课程安排

- 讲授内容
 - (9月11日) 秋1: 课程简介+语音技术引言
 - (9月18日) 秋2: 语音分析 (I)
 - (9月25日) 秋3: 语音分析 (II)、语音识别 (I)
 - (10月9日) 秋5: 语音识别 (II) 语音编码与合成
 - (12月30日) 冬8: 复习及项目成果展示 (周一实验课)
- 实验内容
 - 1. PRAAT 语音分析 (9月16日) 秋2
 - 2. VOICEBOX说话人识别 (9月30日) 秋4

考试: 2020年1月16日08:00-10:00





考核方法

三重考核

(1) 期末考试 (40分)

期末统一闭卷考试，内容涉及语音、视频和音乐处理三块内容，每块内容比重相当。

(2) 平时表现 (30分)

根据同学在课程学习过程中的主动提问、讨论、测验、反馈等互动环节中的不同表现进行考核。

语音、音乐、视频3部分独立考核，每部分均为10分

(3) 实验考核 (30分)

语音、音乐、视频3部分独立考核，每部分均为10分

所有未及时提交（迟交），则在得分基础上降2分；提交的报告不符合要求或报告为抄袭，相应的报告得分为“不及格”。未提交报告者给0分。





讲述提纲

- 语音研究简介
- 课程内容安排
- 语音技术基础



语音技术基础

浙江大学计算机学院
数字媒体与网络技术



- 语言交际过程
- 语音产生过程



语音技术基础



[https://en.wikipedia.org/wiki/2001:_A_Space_Odyssey_\(film\)](https://en.wikipedia.org/wiki/2001:_A_Space_Odyssey_(film))



2001: A Space Odyssey is widely regarded as one of [the greatest and most influential films ever made](#).^[10] In 1991, it was deemed "culturally, historically, or aesthetically significant" by the United States [Library of Congress](#) and selected for preservation in the [National Film Registry](#).^[11] [Sight & Sound](#) magazine ranked *2001: A Space Odyssey* sixth in the top ten films of all time in its 2002^[12] and [2012 critics' polls](#) editions; it also tied for second place in the [magazine's 2012 directors' poll](#). In 2010, it was named the greatest film of all time by [The Moving Arts Film Journal](#).^[13]



语言处理中的知识



语音学（phonetics）和音系学（phonology）的知识：

帮助我们建立词如何在话语中发音的模型



语言处理中的知识



形态学（morphologic）方面的知识：

能够产生并识别单词的这样或那样的变体，需要形态学方面的知识，这些知识能够反映关于上下文中词的形态和行为的有关信息。

语言处理中的知识



句法 (syntax)：关于组词成句的知识。

语言处理中的知识



词汇语义学 (lexical semantics) :

为了理解Dave的请求事实上是关于要求关闭分离舱门的一个命令，而不是讲关于当天中饭的菜单的事情，就要有复合词的语义知识、词汇语义学的知识。



语言处理中的知识



这种礼貌和委婉语言的用法属于语用学（pragmatics）的研究领域。

语言处理中的知识



正确地把这样的会话组织成结构，需要**话语规约**（discourse convention）的知识。





语言处理中的知识

- 语音学与音系学—研究语言的语音
- 形态学—研究词的有意义的组合
- 句法学—研究词与词之间的结构关系
- 语义学—研究意义
- 语用学—研究如何用语言来达成一定的目的
- 话语学—研究大于段的语言单位



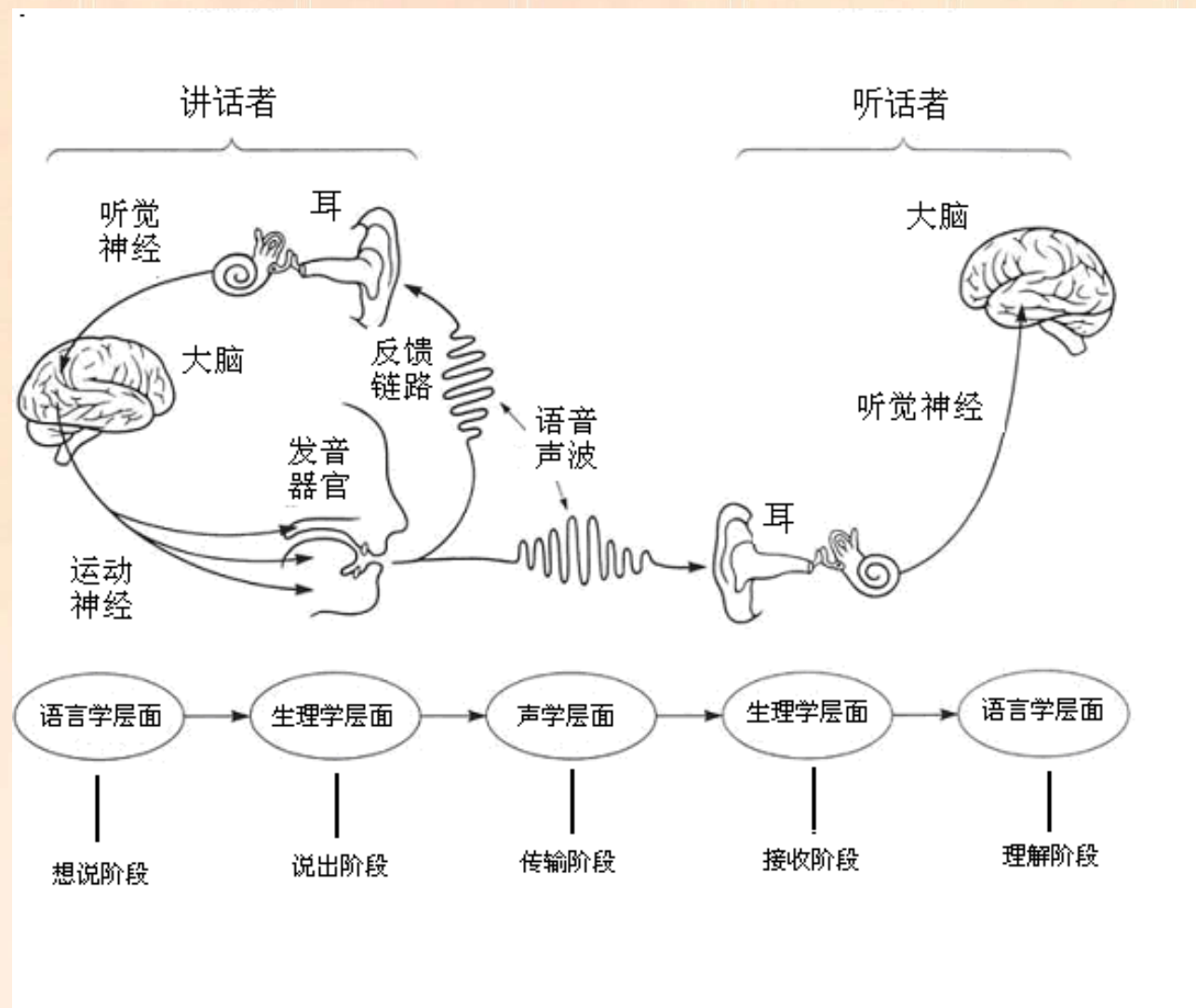


模型和算法

- 状态机 (state machine)
 - 包括状态、状态之间的转移、输入表示等
- 形式规则系统 (formal rule system)
 - 正则语法、正则关系、上下文无关语法
- 逻辑 (logic)
 - 逻辑表达方法是处理语义学、语用学和话语分析等方面知识的选择工具
- 概率论 (probability theory)
 - 其他的各种模型都可以使用概率得到进一步提高
 - 也是一种机器学习 (machine learning) 的模型



语音链





- “发音-传递-感知”三阶段
- 语音学 (Phonetics)
 - 发音语音学(Articulatory Phonetics)
 - 声学语音学(Acoustic Phonetics)
 - 感知语音学(Auditory Phonetics)



参考文献



1. 冯志伟, 孙乐 译, 自然语言处理综论, 电子工业出版社, 2005 年6月

2. 吴朝晖, 杨莹春, 说话人识别模型与方法, 清华大学出版社, 2009, 2

3. 杨莹春, 陈华, 吴飞, 视音频信号处理, 浙江大学出版社, 待出版



课后任务I



- 下载安装实验软件

- PRAAT 语音分析

- [*doing Phonetics by Computer*](#)

- VOICEBOX说话人识别

- [Speech Processing Toolbox for MATLAB](#)

- <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>



课后任务II



- 阅读文献

- L. R. Rabiner and R. W. Schafer, Introduction to Digital Speech Processing
Ch1_Introduction

