

# 第6章 智能接口

---

## 6.1 人机对话

# 6.1 人机会话

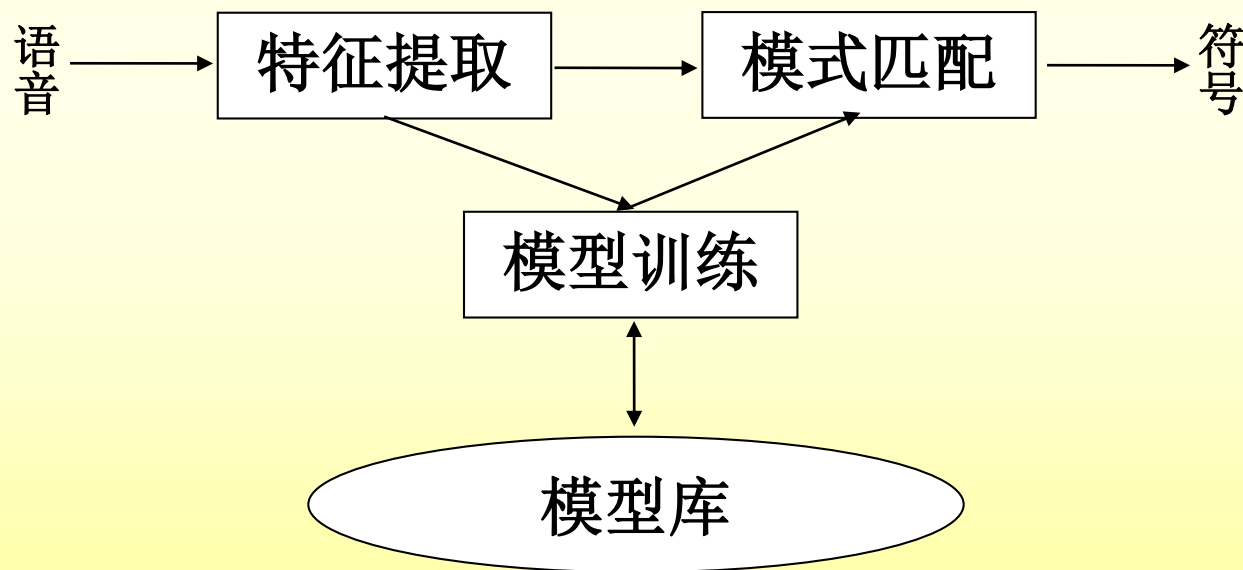
---

人机会话技术有着悠久的历史，是智能化人机接口最具代表性的智能化技术。

其目标：通过语音识别与生成方法与技术，来实现人机之间直接采用自然语言进行对话，并应用于智能机器人、聊天机以及各种智能系统的人机接口中。

# 1、语音识别环节

首先是语音识别。所谓语音识别就是要将语音信号转变为机器内部处理的文本符号。简单地讲，这一过程分为三个方面的内容，即特征提取、模式匹配以及模型训练，如图所示。



# 1、语音识别环节

---

应该说，针对特定人或小规模词汇量的语音识别技术基本上成熟，

但对于非特定人或者大规模词汇量的语音识别问题还是一个需要解决的科学难题。

目前针对非特定人的语音识别方法大致包括：

- （1）隐码尔柯夫模型(HMM)方法，也是语音识别的主流方法；
- （2）基于知识（利用构词、句法、语义、会话背景等方面的知识）的语音识别方法，并与大规模语料统计模型相结合；
- （3）神经网络、遗传算法、免疫算法、蚁群算法等自然计算方法。

# 1、语音识别环节

---

根据语音识别的目标不同，可以将语音识别任务大体可分为3类：

- (1) 孤立词识别 (isolated word recognition),
- (2) 关键词识别 (keyword spotting),
- (3) 连续语音识别。

根据针对发音对象的不同，还可以把语音识别分为：

- (4) 特定人语音识别
- (5) 非特定人语音识别。

# 1、语音识别环节

---

不管是什么具体的语音识别任务，大致上语音识别主要需要解决的5个问题包括：

（1）话语要素的分割问题：将连续的话语分解为词、音素等基本单位；

（2）确定语音模式区分标准：不同的说话人有不同的语音模式，即使同一个说话人，在不同的场合、不同的状态以及不同的时期，也会有不同的语音模式，这就为语音识别模式的分类带来了困难；

# 1、语音识别环节

---

(3) 模糊性问题：说话的含混现象、语言中普遍存在的同义词现象等，使得语音识别成为一个依赖于上下文与会话背景的复杂研究课题；

(4) 词语发音的动态性：单个字母或词、字的语音特性会受到上下文影响而变化，包括读音、重音、音调、音量和发音速度等方面的改变；

(5) 环境噪音干扰：人类具有鸡尾酒效应，可以在嘈杂环境下排除干扰，但这一问题对于机器而言却没有有效的解决方法。

# 1、语音识别环节

---

比如单单同音字现象，就是一个十分棘手的问题。我们知道，汉语有五万多汉字，却共用仅仅两千多个音节，因此同音字现象非常普遍。甚至会出现赵元任指出的《施氏食狮史》这种极端情况而使语音的机器识别研究陷于困境之中：

石室诗士施氏嗜狮，誓食十狮，氏时时适市视狮。十时，氏适市，适十狮适市。是时，氏视是十狮。恃十石矢势，使是十狮逝世，氏拾是十狮尸适石室。石室湿，使侍试拭石室。石室拭。氏始试食是十狮尸。食时，始识是十狮尸实石十狮尸。是时，氏始识是实事实。试释是事。



# 1、语音识别环节

赵元任的《熙戏犀》：西溪犀，喜嬉戏。嵇熙夕夕携犀徙，嵇熙细细习洗犀。犀吸溪，戏袭熙。嵇熙嘻嘻希息戏。惜犀嘶嘶喜袭熙。

杨富森的《于瑜与余欲渔遇雨》：于瑜欲渔，遇余于寓。语余：“余欲渔于渝淤，与余渔渝欤？”余语于瑜：“余欲鬻玉，俞禹欲玉，余欲遇俞于俞寓。”余与于瑜遇俞禹于俞寓，逾俞隅，欲鬻玉于俞，遇雨，雨逾俞宇。余语于瑜：“余欲渔于渝淤，遇雨俞寓，雨逾俞宇，欲渔欤？鬻玉欤？”于瑜与余御雨于俞寓，俞鬻玉于余禹，雨愈，余与于瑜踽踽逾俞宇，渔于渝淤。

《饥鸡集矶记》：唧唧鸡，鸡唧唧。几鸡挤挤集矶脊。机极疾，鸡饥极，鸡冀已技击及鲫。机既济蓊畿，鸡计疾机激几鲫。机疾极，鲫极悸，急急挤集矶级际。继即鲫迹极寂寂，继即几鸡既饥，即唧唧。

# 1、语音识别环节

---

从应用的角度看，根据语音识别应用设施的不同，语音识别则可以分为PC语音识别、电话语音识别和嵌入式设备（手机、PDA等）语音识别。

考虑到不同应用设施提供的采集信道会使人们的发音特性产生变形，因此在具体的应用系统开发中，还需要针对性地解决各种技术问题。

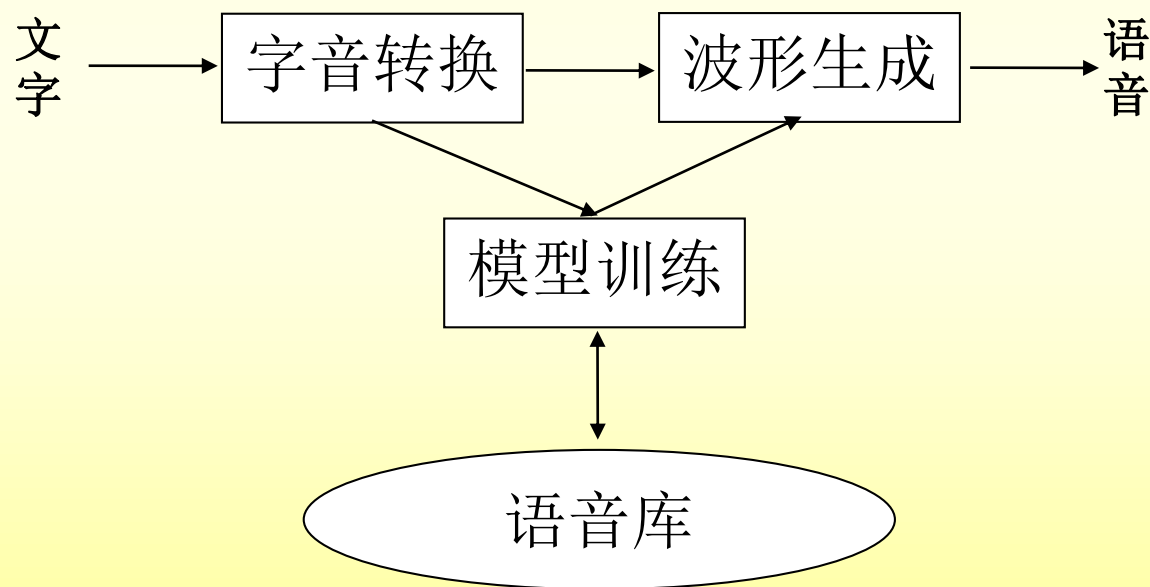
# 1、语音识别环节

---

我国开展语音识别技术研发的主要单位有中科院声学所、自动化所、清华大学、北京交通大学、哈尔滨工业大学、中国科技大学、四川大学、厦门大学、科大讯飞、百度等，开发了不少语音识别系统并应用到具体的智能手机、智能玩具以及其他应用系统中（如家庭服务、宾馆服务、旅行社服务系统、订票系统、医疗服务、银行服务、股票查询服务等等）。

## 2、语音合成环节

其次是语音合成。要实现人机对话系统除了语音识别外，还需要解决语音合成问题。与语音识别相反，语音合成是要将文字符号转换为连续声音形成的话语，因此，语音合成技术，有时也称为文语转换技术。



## 2、语音合成环节

---

从智能技术的应用角度看，文语转换系统实际上可以看作是人工智能的一个分支领域。为了获得高质量的合成语音，除了语音合成本身涉及到的技术外，还需要在理解内容的角度，给出富有情感表达的话语表达效果。

## 2、语音合成环节

---

当然，**语音合成**本身的技术主要有两个方面：

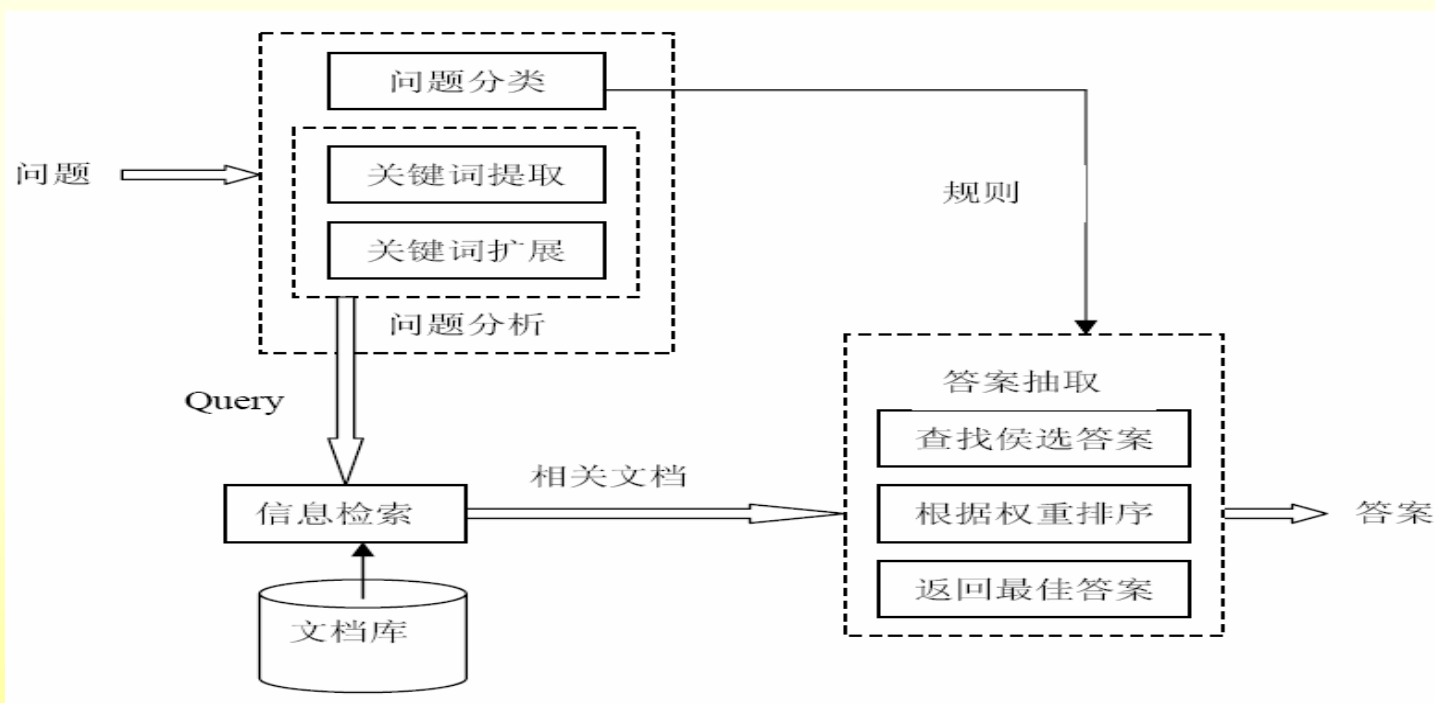
一是将**文字**序列转换为**音韵**序列，涉及到语言文本的处理技术。

二是再将**音韵**表征的文字转换为**语音波形**。涉及到声学处理技术。

End 2

# 3、对话管理环节

有了语音识别与合成，人机会话最后核心部分，就是对话管理机制的构建。如果将人机对话看作是一个问答过程，那么就可以采用如图所示方案来实现对话管理机制。



# 3、对话管理环节

1968年，美国麻省理工学院的J. Weizenbaum设计了一个心理医生咨询系统 ELIZA，**机器**可以**模仿心理医生**来与患者交谈。其典型的一次**会诊咨询**是这样的：

**患者：**男人都一个样。**机器：**怎么一个样法？

**患者：**他们总有许多事来烦我们。**机器：**你能举个例子吗？

**患者：**你瞧，我男友楞要叫我到这儿来。**机器：**您男友叫您到我这儿来？

**患者：**他总说我情绪消沉。**机器：**听说您情绪消沉，我很遗憾。

**患者：**是的，我心情不快乐。**机器：**你认为到我这儿来能使您高兴起来吗？

**患者：**那自然，我需要帮助。**机器：**如果你得到帮助，对您来说，这意味着什么？

**患者：**或许我会同妈妈相处得更好。**机器：**多讲一些您家里的情况，好吗？

.....



# 3、对话管理环节

---

很有意思的是，**ELIZA**仅使用了简单的关键词替换策略，就可以使这场心理咨询如此顺利进行。

不过细心的读者会发现，机器似乎只会换角度“复述性”提问来应付患者。

但作为心理咨询，这对于患者来说就不同了。因为有时患者们只是需要与一个人聊聊天，袒露一些心中的苦闷而已。此时，一台机器可能会比一位医生更“安全”，况且机器会有极大的耐心来倾听患者没完没了的唠叨、抱怨或倾诉。

因此，从这个意义上讲，这种交谈，与其说是**ELIZA**的成功应酬，倒不如说是人类患者本身的需要成全了这场交谈更确切。

# 3、对话管理环节

---

美国斯坦福大学的科尔比也设计了一个人机对话系统，与上述对话系统不同，该系统的任务主要是帮助训练精神病医生获得与重病患者交谈经验的，机器的角色是一名患迫害狂的病人。在机器的对话系统中存放有一些患迫害妄想症的病人资料，比如典型的病人病历资料是这样的：

佩里，28岁，未婚，在邮局工作；他极其脆弱，对涉及他的外貌、教育和信仰等所有问题都敏感；他的爱好是看电影和赛马；有一次，他竟朝赛马场赌注登记员猛扑过去，因为那个人没付给他赢得的彩金；从此，他便开始惧怕黑社会的报复。

# 3、对话管理环节

---

下面是一场精神病医生练习看病的对话（现在**机器**就是模仿**佩里**，一名情绪变化无常、一旦对话触及“敏感”话题，就会出现激动并妄想狂症状的精神病人）。

医生：您怕黑社会吗？**佩里**：黑手党是最可怕的暴徒。

医生：您跟黑手党有什么联系吗？**佩里**：您没看到吗，他们知道我。

医生：他们怎么会知道你？**佩里**：他们在监视我。

医生：为什么？**佩里**：他们想罗织罪名把我打入监狱。

医生：他们为什么要害您？**佩里**：他们想置我于死地。

医生：为什么呢？**佩里**：我掌握着黑手党的罪证。

# 3、对话管理环节

---

为了说明该系统的“智能”，研究者还开展了一场别开生面的图灵测验，具体方法就是安排精神病医生通过电传机与真病人或假病人（机器系统）进行交谈，结果发现，虽然这些医生知道其中有一位是假病人，但却无法判别哪一位是假的。

# 3、对话管理环节

---

不过即使这样，如果与人类交谈的流利机敏比较，机器的这种“应酬”简直是不能相提并论的。

如《三国演义》第八十六回“难张温秦宓逞天辩，破曹丕徐盛用火攻”中的一段对白：

次日，后主将金帛赐与张温，设宴于城南邮亭之上，命众官相送。孔明殷勤劝酒。正饮酒间，忽一人乘醉而入，昂然长揖，入席就坐。温怪之，乃问孔明曰：“此何人也？”孔明答曰：“姓秦，名宓，字子来力，见为益州学士。”温笑曰：“名称学士，未知胸中曾‘学事’否？”宓正色而言曰：“蜀中三尺小童尚皆就学，何况于我？”

### 3、对话管理环节

温曰：“且说公何所学？”宓对曰：“上至天文，下至地理，三教九流，诸子百家，无所不通，古今兴废，圣贤经传，无所不览。”温笑曰：“公既出大言，请即以天为问：天有头乎？”宓曰：“有头。”温曰：“头在何方？”宓曰：“在西方。《诗》云：‘乃眷西顾。’以此推之，头在西方也。”温又问：“天有耳乎？”宓答曰：“天处高而听卑。《诗》云：‘鹤鸣九皋，声闻于天。’无耳何能听？”温又问：“天有足乎？”宓曰：“有足。《诗》云：‘天步艰难。’无足何能步？”温又问：“天有姓乎？”宓曰：“岂得无姓！”温曰：“何姓？”宓答曰：“姓刘。”温曰：“何以知之？”宓曰：“天子姓刘，以故知之。”温又问曰：“日生于东乎？”宓对曰：“虽生于东，而没于西。”

# 3、对话管理环节

---

如何！语言清郎如流，人智机敏诙谐，焉机器所能攀比？！看来靠那种替换“关键词”的“复述”策略是无法从根本上解决语言生成问题，这里面起码还需要一种与“复述”相对的“情感态度主导”言说能力。

End 3