《语音信号处理》整理

reference

《spoken language processing》

1 语音基础

- ## 名词解释

pitch 音调 以HZ为单位
基频又称音高
能量又称音强

说话的声音（声带震动）和其他声音相比，有独特的时域和频域模式。声带的震动产生基频(fundamental frequency)，口腔共振(the pharyngeal and ora cavities)等产生高频谐波

- ## 基频

就是声带的闭-开频率

- ## 声道模型

- ## 语谱图、共振峰

语图纵坐标是Frequency (Hz)，横坐标是Time (s)。语图上还有第三个维度，颜色的深浅，就是表示振幅的大小，即音强（sound intensity）。每隔5m里叶变换。

> 语图某段频率相对于周围较黑，就是说这里振幅较大，音强较大。我们把这一段称为「共振峰」。相对周围较黑的有几处，就有几个共振峰。一般以的条纹的中间位置作为共振峰的频率值，是一个声音区别于其他声音的主要特征，观察共振峰和它们的转变可以更好的识别声音。也就是说，共振峰主要特征。人耳就像一个滤波器组一样，它只关注某些特定的频率分量，所以人的听觉系统是一个特殊的非线性系统，它响应不同频率信号的灵敏度的。
> form https://www.zhihu.com/question/24190826/answer/32315664

- ## 辅音和元音的区别

• 1、辅音发音时，气流在通过咽头、口腔的过程中， 要受到某部位的阻碍；元音发音时，气流在咽头、 口腔不受阻碍。这是元音和辅音最主要的区别。
• 2、辅音发音时，发音器官成阻的部位特别紧张； 元音发音时发音器官各部位保持均衡的紧张状态。
• 3、辅音发音时，气流较强；元音发音时，气流较 弱。
• 4、辅音发音时，声带不一定振动，声音一般不响 亮；元音发音时，声带振动，声音比辅音响亮。

> 一般只有元音（一些介于元音辅音中间分类不明的音暂不讨论）才会有共振峰，而元音的音质由声道的形状决定，而声道的形状又通过[illegible]的动作来（articulatory+movements）。
>
> 语谱图上频率能量峰值按照时间延伸形成带状
>
> from 安时

- **清音和浊音**
  - 清音：声带不振动
  - 浊音：声带振动而发音
  - 元音都是浊音、辅音有清音也有浊音。

Discrete-time model for speech production.

4 语音编码 Speech Coding

语音编码技术的目的：为了减少传输码率或存储量，以提高传 输或存储的效率。经过这样的编码之后，同样的信道容量能传 输更多路的信号，如用于存较小容量的存储器。因 而这类编码又称为压缩编码。需要在保持可懂度与音质、降低数码率 和降低编码过程的计算代价三方面折衷。

- 波形编码：波形编码器没有使用模型，而是试图使重构的语 音和原始语音之间的误差最小化。波形编码的方法简单，数码率较高，在64kbit/s至32k质优良，当数码率低于 32kbit/s的时候音质明显降低，16 kbit/s时音质非常差。
- 参数编码：基于参数或模型的编码器提供了一种可用来模拟 语音产生的模型，并从原始语音中提取可用来描述此模型的 参数，然后随着语音信号特周期地更新模型参数 。 声码器编码后的码率可以做得很低，如1.2kbit/s、2.4kbit/s， 但是也有其缺点。首先是合成语音质量较差，往往清晰度可以有，难于辨认说话人是谁，其次是复杂度比较高
- 混合编码：混合编码是将波形编码和声码器的原理结合起来，数码率约在4kbit/s—16kbit/s之间，音质比较好，最近有个别 算法所取得的音质可与波当，复杂程度介乎与波形编码器和声码器之间

电话的语音采样频率为8khz. 评价分辨率好坏的标准：the Mean Opinion Score (MOS)
解码延迟： Coder delay is the sum of different types of delay. The first is the algorithmic delay arising because speech coders usually operate on a blo samples, called a frame, which needs to be accumulated before processing can begin. Often the speech coder requires some additional look-ahead b frame to be encoded. The computational delay is the time that the speech coder takes to process the frame. For realtime operation, the computationa be smaller than the algorithmic delay. A block of bits is generally assembled by the encoder prior to transmission, possibly to add error-correction prop bit stream, which cause multiplexing delay. Finally, there is the transmission delay, due to the time it takes for the frame to traverse the channel. The d incur a decoder delay to reconstruct the signal. In practice, the total delay of many speech coders is at least three frames.
编码纠错[1]

2 语音识别

Fundamental Equation of Statistical Speech Recognition

解码(decoding)：把直接的观测结果看作是源码的编码，那么根据编码推测源码就是解码过程，是根本目的。解码可以是直接在可行解空间进行搜索。-搜索是不可行的，因为解空间是巨大的，甚至是无穷大的，普遍采用的是启发式搜索(即生成式搜索，另一种搜索思路是进化搜索)。

- **声学模型(Acoustic Modeling)**

决定语音分布的因素（因此在生成训练样本需要下面因素的变化才能拟合正式环境下的语音分布）：

1. 上下文
2. 说话风格（情绪、语速、重音等）

3. 说话人的习惯
4. 说话环境

- ## 测量识别模型的正确率

- ## 语音采样

端点检测：

- 过零率[2]
- 谱熵分布
- 频带方差
- 二分类器
  the EM algorithm can iteratively estimate the Gaussian parameters without having a precise segmentation between speech and noise segments.

- ## 短时分析

决定短时能量特性有两个条件：不同的窗口的形状和长度。
窗长越长，频率分辨率越高，而时间分辨率越低。如果很大，它等效于很窄的低通滤波器，此时随时间的 变化很小，不能反映语音信号的幅度变化，信节 就看不出来；反之，窗长太小时，滤波器的通带变宽，随时 间有急剧的变化，不能得到平滑的能量函数。
矩形窗谱平滑性能好，但损失高频成分，波形细节丢失， 海明窗与之相反

- ## MFCC

提取MFCC特征的过程：
1）先对语音进行预加重[3]、分帧[4]和加窗[5]；
2）对每一个短时分析窗，通过FFT得到对应的频谱[6]；
3）将上面的频谱通过Mel滤波器组[7]得到Mel频谱；
4）在Mel频谱上面进行倒谱分析（取对数，做逆变换，实际逆变换一般是通过DCT离散余弦变换来实现，取DCT后的第2个到第13个系数作为MFCC系数Mel频率倒谱系数MFCC，这个MFCC就是这帧语音的特征；

Mel三角滤波器组

MFCC参数提取

![特征普遍采用的语音特征[8]](http://upload-images.jianshu.io/upload_images/3444195-0882821befe50ddc.png?imageMogr2/auto-orient/strip%7CimageView2/2/w/1240)

- ## 模板匹配法(传统)

模板匹配语音识别系统基本构成

语音识别模式匹配的问题：时间对准

- 同一个人在不同时刻说同一句话、发同一个音，也不可能具有完全相同的时间长度；
- 语音的持续时间随机改变，相对时长也随机改变；
- 端点检测不准确；

方法1：线性时间规整，均匀伸长或缩短 – 依赖于端点检测； – 仅扩展时间轴无法精确对准；
方法2：动态时间规整 – DTW－Dynamic Time Warping； – 60年代Itakura提出来的；其思想是：由于语音信号是一种具有相当大随机性的[illegible]，即使相同的词，每一次发音的结果都是不同 的，也不可能具有完全相同的时间长度。因此在与已存储 模型相匹配时，未知单词的时间轴要不均匀地扭曲或弯曲，以使其特征与模板特征对正。 动态时间规整DTW是一个典型的优化问题，它用满足一定条 件的时间规整函数描述输入模板和参考模板的时间对应关系，求解两者时累计距离最小所对应的规整函数。

DTW的问题：

- 运算量大；
- 识别性能过分依赖于端点检测；
- 太依赖于说话人的原来发音；
- 不能对样本作动态训练；
- 没有充分利用语音信号的时序动态特性；

DTW适合于特定人基元较小的场合，多用于孤立 词识别；

- ## 语音的识别单元

phoneme是用于区别词汇的最小单元，音节(Syllables)介于音素和单词的中间，说话时一次发出的， 具有一个响亮的中心，并被明显感觉的语音片断。词语作为识别单元？词汇太多；无法应对新产生的词。声学单元越小，其数量也就越少，训 练模型的工作量也就越小；但另一方面，单元越 小，对上下文越大，越容易受到前后相 邻的影响而产生变异，因此其类型设计和训练样 本的采集更困难。不过phone是一个相邻无关的单元，而triphone是考虑到相邻前phone的影响，于是认为只有当前后及本身的phone都相同时才认为是同样的triphone. 每个词的发音可能有多种变化方式，在子词串接时，必须有所修
替换：即词中的某个音子可能被用其它相似 而略有差异的子词单元所替换。
插入和删除：词中有时增加了一个不是本词 成分的子词单元，有时又将本词成分中的某个 子词删除。

声学模型选择---声学单元如何组成词

声学模型

- ## GMM-HMM声学模型

我们认为语音是由许多状态组成的一个HMM序列所生成出来的：每一个时刻t到达某个状态s，s按照自己的分布产生一个采样(观测)，这个采样就是MFCC，是一段时间内产生了一个MFCC参数序列，即是特征提取后的语音。生成一段语音的GMM-HMM模型不是固定的，而是很多building block组合起来的，block可以是一个状态，也可以是三个状态(triphone)。我们需要确定的模型参数就是所有这些building block的观测分布(GMM参数)以及它们之间的相互转移概率(HMM参数)。另外，根据一段语音的MFCC参数，在已知GMM、HMM参数的情况下，计算可能的状态序列概率，以找出最大可能的状态序列(deco

对于一个给定的观测序列(O1,O2,O3)，计算某状态序列(1->2->2)的概率。可以看到每隔状态对应一个分布，而观测是分布的一个采样

根据HMM的分布观测样本空间的是否离散，HMM分为离散HMM和连续HMM. 由于原始输入的信号是连续空间的，转化为离散HMM需要进行"采样"，也即将样本空间划分成M块，用块值代替原始的样本。
半连续HMM（SCHMM）：相当于离散HMM和连续HMM的混合。状态输出的特征向量是连 续的，也是用多个高斯分布的加权和来近似概率分布函数，

加权和的高斯函数的集合是固定的，类似于对高斯密度函 数建立了“码本”，各个状态输出概率密度之间不同的是对“码 本”中各个高斯密度函数的加权系
训练过程分为两个部分：GMM、HMM

- GMM参数训练
- HMM参数训练

GMM没有利用帧的上下文信息 • GMM不能学习深层非线性特征变换

- **DNN-HMM**

- **CTC**
  - 不要需要输入与输出帧级别的对齐信息，不用和HMM模型结合
  - 约90%的帧其对应的输出为空（blank），可以采取跳帧，加快解码速 度
  - 因解码速度快，识别性能也较优，所以工业界大多采用这种模型

连续语音识别的声学模型和语言模型

大词汇量连续语音识别技术

3 语音合成

文本分析的主要功能是使计算机知道要发什么音、怎么发音，并将发 音的方式告诉计算机。对于汉语来说，还要让计算机知道文本中的词 边界、短语边界，以便发音时设置不同长度的停顿。文本 分析还应将汉字、符号、数字等转换成适当的拼音。
• 文本分析的结果既要告诉计算机发什么音，也要告诉计算机以什么方 式发音。如：发音的声调；音节是长还是短；是重还是轻；是高还是 低；到哪儿停顿的长短。TTS系统要给出代表这些韵律特 征的声学参数，这就是韵律生成模块的功能。
• 计算机知道要说什么以及有了韵律控制参数后，计算机通过声学模块 产生语音输出。在系统中，声学模块负责产生合成语音。声学模块从 语音数据库的语音基元，拼接成语句， 再经过韵律修饰， 就可以输出自然连续的语声流。
• 文本分析、韵律生成可以采用基于规则或基于数据驱动的方法。韵律 修饰可以直接改变波形或进行参数变换。

主要分为三个步骤：

1. 音素分析(phonetic analysis)：
   将文本转化为对应的音素序列，主要依靠查表。句子切分、句子分词、POS，非标准词处理，同形字辨别(Homograph Disambiguation)
2. 韵律分析(prosodic analysis)：对音素序列添加适当的停顿和延迟信息，这也是prosody与phoneme之间的区别。停顿添加的训练是用二分类器

   利用人工标注的韵律分解，使用决策树作二为分类器

- 基频建模

  - 基于规则的方法 – 通常规则系统包括两个方面，一是汉语的通用 规则，比如汉语的4个调的基本形状，上声连接 的变调规则，时长变化，语气变化 等；二是目标说话人的特定韵律特征规则，比 如个人的基本调型、调域、语速停顿规则。
  - 基于数据驱动的方法
    数据驱动模型通常考虑哪些上下文信息
    短语信息：短语中音节的个数、词的个数 ，短语在句子中的位置
    词信息：词长，词性，词在短语中的位置 • 音节信息：声韵母类型，声调，在词中位 置，在短语中位置，前音节信息和后音节 信

1. 语音合成(waveform synthesis)：将上述音素序列转化为波形信号
   有两种方式：

- 双音合成 diphone

非拼接方式，对于每一个phone都需要一个duration 和 F0值

两个缺点：
1. 必须对原音频进行信号处理，导致听上去不自然
2. 仅考虑一个相邻音素的影响

- 单元选择合成 (unit selection)
  拼接(concatenative)方式，没有对拼接单元进行信号加工。diphone有多样本，每一个都以长句子形式存储，因此每一个diphone样本[illegible]环境非
  在需要合成的时候，我们挑选最合适的diphone样本，而无须对diphone进行信号加工
  在选择时我们考虑的合适指标有两个方面：

  对应于两个数值指标：

  得到综合指标：

基于HMM的语音合成

In the HMM-based speech synthesis,一个语音单元的语音参数：语谱spectrum, 基频fundamental frequency (F0)，音长 phoneme duration are statistic and generated by using HMMs based on maximum likelihood criterion.

一个状态的输出是MFCC参数向量，但是观测样本仅与当前状态相关，与相邻的观测样本没有直接相关，这样和i出现不平滑。为了保证平滑，并不是直分布函数进行采样，而是输出分布(GMM参数)，在T时刻后，对每一维的参数进行平滑采样。

HMM-based speech synthesis system.

基于深度神经网络的语音合成

- 评价

mean opinion score (MOS) tests were conducted. In the paired comparison tests, after listening to each pair of samples, the subjects were asked to cl they preferred, though they could choose “neutral” if they did not have any preference. In the MOS tests, after listening to each stimulus, the subjects v to rate the naturalness of the stimulus in a five-point Likert scale score (1: Bad, 2: Poor, 3: Fair, 4: Good, 5: Excellent).

6 说话人识别/自适应

说话人之间的差异对非特定人语音识别系统造成的影响主 要有两方面原因:
(1) 当某一使用该系统的说话人语音与训练语音库中的所有说话人 的语音都有较大的差别时,对该使用者的语音系统的识别性能会有严 重的
(2) 训练一个较好的识别系统需要采集数量很大的说话人的语音用 于训练,让训练语音库覆盖更为广泛的语音空间,这样虽然可以减低 样本[illegible]布中影响
成识别系统参数分布较广,而不是 较为尖锐的分布,造成识别性能的下降

文本相关的语音转换：相当于文本中的平行语料：对同一句话，不同人进行语音实现，得到的平行语音作为训练样本

TTS system can generate synthetic speech which closely resembles an arbitrarily given speaker's voice using a small amount of target er's spe applying speaker adaptation techniques such as MLLR (Maximum Likelihood Linear Regression) algorithm

文本无关的语音转换：没有平行语料。使用场景可以大大拓展，也可以用于跨语言语音转换。从音素的角度出发进行建模

基于Pitch Target模型的基频转换框架

GMM的基频转换方法

7 语音系统

对话系统的特点

1. 口语对话系统都有比较明确的领域限制，一般说来它 只需要关心领域相关的内容，对于超出领域限制的用户输入 可以不加理会；
2. 不同于语音命令系统中的孤立词和听写机系统中的朗 读语音，对话系统面对的是自发语音（Spontaneous Speech），发音比较随意；
3. 对话系统的输入是人们日常生活中的口语，语句中常 常包括不流利、不合语法、内容不完整等口语现象；
4. 口语对话系统的应用环境比较多样化，可能是非常安 静的实验室环境，可能是充满噪音的正在行驶的汽车中，更 有可能是人声嘈杂的商场。

语音理解过程都是分两步完成的：

1. 语音识别器对输入语音进行识别，输 出 N-best 或者词图（Word Graph）形式的识别 结果；
2. 语言理解器对识别器的输出进行分析 和理解，得到对话管理模块所需要的语义表示 形式。

对话管理

- 对话管理系统要做到能够在与用户多次交 互的情况下保持回答的连续性和合理性, 并 且能够处理用户在交互过程中转变提问目 的的情况。
- 在已经实现并应用的对话管理的设计中, 主 要有：基于状态图的结构、填充槽结构和 基于任务的结构。
  基于**状态图**的结构采用有限状态机来控制对话的进行：
  - 每个对话片段的情况可以看成是一个一个的状态, 将对话 过程的每一次交互都看作是一次状态的跳转, 即每一个状 态节点都表示着当时对话的
    统动作, 每一个 连接弧表示用户的每次操作。因此, 整个对话的过程, 从开 始到结束可以看成是在状态图中的一个连接开始节点和结 束节点的
    径。这种对话管理结构要求设计者要在设计时预计出所有可 能的对话状态和用户可能的操作, 即所有状态之间的转 移条件。
  - 从工程实现的角度来讲, 由于此种结构要求对于每一个 状态用户的任何操作都要有一个跳转的规定, 因此这种 结构在对话清晰明确的时候有着
  - 如果领域的内容复杂则状态图很难保证没有任何的纰漏, 实现起来要耗费大量的人力。
  - 有限状态的结构有着其必然的缺点, 即难以应付没有预 测到的情况, 如果用户的反应完全超乎设计师的预计, 则对话必然不能正常地进行, 并且
    个以用户 为对话主导的互动系统中会更加突显出来。

  填充槽结构采用一个多维特征向量来表示对话 的情况, 并且在对话的过程中不断地修改向量的值。特征向量通常是由从用户接收到的信息和一些 状
  根据特征向量的值来决定下一 步的操作。
  这种方法与上一种基于状态图的方 法的最大区别在于: **对于操作的顺序没有严格的限制**, 即只关心当前对话的状态信息, 根据现在的状态作出反应,
  的回答或系统的反应修改特征向量。
  因为这种结构不考虑整个对话的顺序, 所以 比基于状态图的结构适应更多的对话类型。 同样, 这种结构也有着自己的适应范围。
  – ①与基于状态图的结构一样, 也要列出所有的可 能状态, 即所有可能的特征向量。
  – ②由于填充槽的结构要求列出所有的槽来表示 状态, 所以槽的数目要有一定的限制, 这也是对 其可以实现的系统范围的一个约束; 并 于它 只记
  的状态, 所以对于多提问目标的 情况就难以应对

  基于任务的结构是一种目前最受瞩目的结构, 并且适应 的范围也最为广泛。-

- 任务是指用户为达到某种目的而采 取的一系列的操作或对话
- 一般来讲, 任务包括进度表(Plan) 和目标。目标就是用 户想要达到的目的。
- 通常来讲, 系统要通过一系列的步骤与用户交互才能完 成特定的任务, 这些交互的步骤就构成进度表。例如上 例中, 为了达到上面 务, 系统要 确定电影 的时间、要求的电影院等, 用户可能要求系统提供影片 介绍, 系统要根据用户提出的新的要求不断地修改进度 表, 最终 务。同时 够支持在对话过程 中任务的突然跳转。
- 对于一个应用的领域, 通常采用树型结构来描述任务。在表示领 域的根节点下面的第一层子节点是任务节点, 任务节点的子节点 决这个任 到的信息要素, 一个信息要素节点的 子节点表示这个信息要素的子要素。要素之间的关系, 如 “与”、“或”等, 在节点关系中体现出 两个节点 的关系, 则表示两个节点要同时满足才能完成这两个节点 的父节点; 若节点之间是“或”的关系, 则表示两个节点只要满 足一个即可 域包含 的关系比较复杂, 则还会 包括其他节点间关系。当用户与系统交互的时候, 系统首先要判 断用户的任务是什么, 即要达到什么样的 目的; 然后 应的任务树, 将用户提供的信息填进各个信息要素的节点中。 根据节点间的逻辑关系判断目前所拥有的信息量是否足够完成 该任 果不能, 息的节点, 根据节点所定义的提问 方法对用户进行提问, 要求用户对该节点的信息进行补充, 即根 据任务树来不断地制定修改进度

语音检索

语音检索就是在语音数据库中搜索查询其中出现的关键词。 语音检索需要使用自动语音识别(ASR)技术分析语音数据的 内容。
在语音检索中，首先采用ASR技术为语音数据库建立索引， 然后在检索时，先从查询中提取关键词，接着从索引数据 库中搜索这些关键词，并对搜索到 置信度计算 以判别其有效性。最后根据搜索到的文档与查询间的相关 程度对查询结果进行排序输出。

用于语音检索的常用技术有关键词检出技术、 连续语音识别技术和说话人识别技术等

8 语音增强

语音增强是指当语音信号被不同噪声干扰、甚至淹没 后，从噪声背景中提取有用的语音信号，抑制噪声干 扰的技术。语音增强在语音识别、语音编码等 要的应用，是语音交互 系统中最前端的预处理模块。
噪音类型：1. 混响 2. 背景噪声 3. 人声干扰 4. 回声

- 单通道语音增强
  - 谱减法(原理简单，算法计算复杂度低)
    将含噪语音信号和VAD判别(Voice Activity Detection （语音激活检测）)得到的纯噪声信号进行DFT变化 ，从含噪语音谱幅度特征中减掉纯噪声 征，得 到增强的幅度谱特征，再借用含噪语音的相位进行IDFT变 化，得到增强的语音。 谱减法假设
    语音和噪声信号是线性叠加的 噪声是平稳的(指的是频谱固定)，噪声与语音信号不相关(指的是噪音在语音频率上能量小)。
    谱减法相当于对带噪语音的每一个频谱分量乘以一个 系数。信噪比高时，含有语音的可能性大，衰减系数 小；反之衰减系数大。
  - 维纳滤波
    在最小均方准则下用维纳滤波器实现对语音信号的估 计，即对带噪语音信号y(t)=s(t)+n(t),确定滤波器的 冲击响应h(t),使得带噪语音信号经过该 出 能够与s(t)的均方误差最小。
    计算复杂度低，满足实时性要求
    算法要求输入信号具有平稳特性
    算法要求带噪语音和安静语音存在线性关系
    在处理非平稳噪声时，降噪效果会变差 在复杂环境下难以跟踪非平稳噪声变化轨迹
  - 矩阵分解
    增强的谱参数通过语音参数基矢量加权得到，可以抑制过 平滑问题
    建立的基矩阵可以通过扩帧来考虑相邻帧的特征，从而捕 获噪声变化轨迹
    相对于其它数据驱动方法，不需要大数据进行训练
    算法计算复杂度高，实时性难以满足要求
  - 基于分析-合成框架语音增强
    语音增强问题进行分解
    准确提取语音参数
    增强处理语音参数
    声码器合成语音

- 数据驱动（例如深层神经网络）

- 多通道语音增强
  - 波束形成
    通过波束形成方法：建立空间滤波器模型，它的作用包括：
    - 将多个麦克风采集的信号进行同步，生成单通道信号
    - 只增强目标方向的信号，对其它方向的信号进行抑制

---

1. 差错控制编码：想在一个带宽确定而存在噪声的信道里可靠地传送信号，无非有两种途径：加大信噪比或在信号编码中加入附加的纠…。
   from http://fiber.ofweek.com/2016-10/ART-210007-8500-30059906.html ↩
2. 在离散时间语音信号情况下，如果相邻的采样具有不同的代数符号就称为发生了过零。单位时间内过零的次数就称为过零率。清音的…率明显高…零率 ↩
3. 预加重的目的是提升高频部分，使信号的频谱变得平坦，保持在低频到高频的整个频带中，能用同样的信噪比求频谱。同时，也是为了消除发生过…嘴唇的效应，来补偿语音信号受到发音系统所抑制的高频部分，也为了突出高频的共振峰。 ↩
4. 分帧： 先将N个采样点集合成一个观测单位，称为帧。通常情况下N的值为256或512，涵盖的时间约为20~30ms左右。为了避免相邻两帧的变化过…让两相邻帧之间有一段重叠区域，此重叠区域包含了M个取样点，通常M的值约为N的1/2或1/3。通常语音识别所采用语音信号的采样频率为8KHz或…以8KHz来说，若帧长度为256个采样点，则对应的时间长度是256/8000×1000=32ms。 ↩
5. 加窗（Hamming Window）：将每一帧乘以汉明窗，以增加帧左端和右端的连续性。 ↩
6. FFT：由于信号在时域上的变换通常很难看出信号的特性，所以通常将它转换为频域上的能量分布来观察，不同的能量分布，就能代表不同语音的…在乘上汉明窗后，每帧还必须再经过快速傅里叶变换以得到在频谱上的能量分布。对分帧加窗后的各帧信号进行快速傅里叶变换得到各帧的频谱。…号的频谱取模平方得到语音信号的功率谱。 ↩
7. Mel三角带通滤波器:对频谱进行平滑化，并消除谐波的作用，突显原先语音的共振峰。（因此一段语音的音调或音高，是不会呈现在 MFCC 参数…说，以 MFCC 为特征的语音辨识系统，并不会受到输入语音的音调不同而有所影响） 此外，还可以降低运算量。 ↩
8. 一帧的音量（即能量），也是语音的重要特征，而且非常容易计算。因此，通常再加上一帧的对数能量（定义：一帧内信号的平方和，再取以10为…值，再乘以10）使得每一帧基本的语音特征就多了一维，包括一个对数能量和剩下的倒频谱参数。 ↩

语音信号处理入门书籍 阅读数 1167

语音信号处理一般包括以下几个部分：（1）语音信号的声学基础及产生模型（2）语音信号的特征分析… 博文 | 来自： qq_33874667…

语音信号处理 阅读数 937

文章目录语音信号处理第一章 绪论第二章 语音信号处理基础知识语音和语言汉语语音学汉语的声母和… 博文 | 来自： Heart_Sea的…

前端语音信号处理 阅读数 891

1、语音活动检测语音活动检测（Voice Activity Detection， VAD）用于检测出语音信号的起始位置，分… 博文 | 来自： xinshuwei的博客

语音信号处理1：Introduction 阅读数 85

参考An introduction to signal processing for speech，极好的入门引导，摘录+补充。This chapter aims… 博文 | 来自： lyrich的博客

语音识别学习资料入门整理_mandagod的博客-CSDN博客

语音信号处理笔记(未整理) - 努力奋斗的小菜鸟的博客 - CSDN博客

语音信号处理基础(一) 阅读数 346

语音信号处理基础(一)文章目录语音信号处理基础(一)1.绪论1.1概述1.2语音信号处理的三个主要分支1.… 博文 | 来自： 张亚楠的博客

语音信号处理 - Heart_Sea的博客 - CSDN博客

语音信号处理_u010384318的专栏-CSDN博客

基于MATLAB的语音信号处理 阅读数 3万+

基于MATLAB的语音信号处理摘要：语音信号处理是目前发展最为迅速的信息科学研究领域中的一个，… 博文 | 来自： 小哲的博客

兔子与花
10篇文章
关注 排名:千里之外

by21010
12篇文章
关注 排名:千里之外

Audio_Wang
169篇文章
关注 排名:7000+

weixin_34262482
4568篇文章
关注 排名:千里之外

语音信号处理入门书籍_qq_33874667的博客-CSDN博客

语音信号处理基础(一)_张亚楠的博客-CSDN博客

语音信号处理领域国内外高手homepage分享 阅读数 3501

详细内容见群文件，欢迎大家加入音频/识别/合成算法群（696554058）交流学习，谢谢！本内容原创… 博文 | 来自： king_audio_vid…

语音信号预处理及特征参数提取 阅读数 2万+

1.WAVE文件格式在进行语音信号处理时，基本上会采用WAVE文件进行处理。WAVE文件格式有什么… 博文 | 来自： hugua专栏

- **语音信号处理1:Introduction_lyrich的博客-CSDN博客**

- **【转】语音信号处理由浅入深_Hibiscus_Jin的博客-CSDN博客**

语音信号处理基础（二） 阅读数 1688

语音信号处理基础（二）语音信号的特性主要是指它的声学特性、时域波形、频谱特性以及语音信号的… 博文 | 来自： Yoc Lu

- **前端语音信号处理_xinshuwei的博客-CSDN博客**

聊聊C语言和指针的本质 阅读数 3万+

坐着绿皮车上海到杭州，24块钱，很宽敞，在火车上非正式地聊几句。很多编程语言都以“没有指针”… 博文 | 来自： Netfilter,iptable…

程序员必须掌握的核心算法有哪些? 阅读数 39万+

由于我之前一直强调数据结构以及算法学习的重要性，所以就有一些读者经常问我，数据结构与算法应… 博文 | 来自： 帅地

Python 基础（一）：入门必备知识 阅读数 14万+

Python 入门必备知识，你都掌握了吗? 博文 | 来自： 程序之间

分享靠写代码赚钱的一些门路 阅读数 6万+

作者 mezod，译者 josephchang10如今，通过自己的代码去赚钱变得越来越简单，不过对很多人来说… 博文 | 来自： qq_33570092…

史上最全的mysql基础教程 阅读数 6万+

启动与停止启动mysql服务sudo /usr/local/mysql/support-files/mysql.server start停止mysql服务sudo /us… 博文 | 来自： 智障小鲁班

C++知识点 —— 整合（持续更新中） 阅读数 2万+

本文记录自己在自学C++过程中不同于C的一些知识点，适合于有C语言基础的同学阅读。如果纰漏，… 博文 | 来自： 逆流而尚 的博客

python学习方法总结(内附python全套学习资料) 阅读数 5万+

不要再问我python好不好学了我之前做过半年少儿编程老师，一个小学四年级的小孩子都能在我的教学… 博文 | 来自： 一行数据

程序员求助：腾讯面试题，64匹马8个跑道，多少轮选出最快的四匹 阅读数 4万+

昨天，有网友私信我，说去阿里面试，彻底的被打击到了。问了为什么网上大量使用ThreadLocal的源… 博文 | 来自： web前端学习…

数据库优化 - SQL优化 阅读数 16万+

以实际SQL入手，带你一步一步走上SQL优化之路！ 博文 | 来自： 飘渺Jam的博客

有哪些让程序员受益终生的建议 阅读数 15万+

从业五年多，辗转两个大厂，出过书，创过业，从技术小白成长为基层管理，联合几个业内大牛回答下… 博文 | 来自： 启舰

语音信号处理基础（三） 阅读数 282

语音信号处理基础（三）倒谱分析（Cepstrum Analysis）下面是一个语音的频谱图。峰值表示语音的… 博文 | 来自： Yoc Lu

linux系列之常用运维命令整理笔录 阅读数 22万+

本博客记录工作中需要的linux运维命令，大学时候开始接触linux，会一些基本操作，可是都没有整理起… 博文 | 来自： Nicky's blog

2019年10月中国编程语言排行榜 阅读数 2万+

2019年10月2日，我统计了某招聘网站，获得有效程序员招聘数据9万条。针对招聘信息，提取编程语… 博文 | 来自： 毛毛虫

20190510 语音识别资源整理 阅读数 1762

语音处理课程推荐|Speech Processing（2019） 台师大Speech Processing。国立台湾师范大学的陈柏… 博文 | 来自： Grace_yan的…

字节跳动面经 阅读数 5455

操作系统1.进程间的通信方式？匿名管道 有名管道 共享内存 socke通信 信号 信号量2.管道间如何具体… 博文 | 来自： weixin_438464…

计算机组成原理 存储器 阅读数 2553

一、概述1.存储器分类按存储介质分：半导体器件：分双极型（TTL）和 MOS 管两种 磁表面存储器：… 博文 | 来自： lintong的博客

比特币原理详解 阅读数 16万+

一、什么是比特币比特币是一种电子货币，是一种基于密码学的货币，在2008年11月1日由中本聪发表… 博文 | 来自： zcg_74145489…

通俗易懂地给女朋友讲：线程池的内部原理 阅读数 10万+

餐盘在灯光的照耀下格外晶莹洁白，女朋友拿起红酒杯轻轻地抿了一小口，对我说：“经常听你说线程… 博文 | 来自： 万猫学社

了不起的前端性能优化 阅读数 2933

前言说到前端性能优化，绝对是对一个前端攻城狮的综合考量 ~ 作为一个前端，在功能ok的前提下，最… 博文 | 来自： HiSen的博客

语音信号处理-2-----语音信号处理的常用算法1(HMM) 阅读数 802

这个Blog主要介绍语音信号处理中隐马尔科夫模型。一些小常识HMM在语音识别中的地位一直很高，… 博文 | 来自： zhangming041…

H.264解码过程剖析-4 阅读数 1万+

x264开源工程实现H.264的视频编码，但没有提供对应的解码器。ffmpeg开源多媒体编解码集合汇集了… 博文 | 来自： mandagod的…

经典算法（5）杨辉三角 阅读数 8万+

杨辉三角 是经典算法，这篇博客对它的算法思想进行了讲解，并有完整的代码实现。... 博文 | 来自： 扬帆向海的博客

小白都能看得懂的java虚拟机内存模型 阅读数 4万+

目录一、虚拟机二、虚拟机组成1.栈栈帧2.程序计数器3.方法区对象组成4.本地方法栈5.堆GCGC案例一… 博文 | 来自： 我爱吃土豆

程序员接私活怎样防止做完了不给钱? 阅读数 17万+

首先跟大家说明一点，我们做 IT 类的外包开发，是非标品开发，所以很有可能在开发过程中会有这样… 博文 | 来自： DavidGoGo_…

了解了这些，轻松拿offer——Java面试之道 阅读数 1万+

本篇文章将讲述Java基础中的面试题及其解答，希望对大家有所帮助。synchronized和Lock的区别sync… 博文 | 来自： 我本将心向明…

爬虫福利二 之 妹子图网MM批量下载 阅读数 19万+

爬虫福利一：27报网MM批量下载 点击看了本文，相信大家对爬虫一定会产生强烈的兴趣，激励自己去… 博文 | 来自： Nick.Peng 的…

python json java mysql pycharm android linux json格式 c#提取字符串中的数字 c# 扩展无返回类 c#网站如何做预览功能 c# 异步 返回值 c#chart实时曲线图 c# 窗体 隐藏 c#实现终端上下滑动 c# 数组类型 泛型约束 c#的赛狗日程序 c# 传递数组 可变参数

mandagod

TA的个人主页 >

原创	粉丝	获赞	评论	访问
227	323	153	109	75万+

等级: 博客 周排名: 4389

积分: 1万+ 总排名: 3147

勋章:

关注 私信

最新文章

Core ML API

Reducing the Size of Your Core ML App

将经过训练的模型转换为 Core ML

Integrating a Core ML Model into Your App

Core ML概览

分类专栏

多媒体技术 41篇

多媒体封装格式 32篇

多媒体协议 30篇

视频编解码 13篇

举报

音频编解码 30篇

展开

归档

2020年3月 7篇
2020年2月 13篇
2019年12月 13篇
2019年11月 3篇
2019年10月 5篇
2019年9月 3篇
2019年8月 3篇
2019年7月 12篇

展开

热门文章

资源|5本深度学习和10本机器学习书籍（免费下载）
阅读数 24027

向上取整⌈ ⌉和向下取整⌊ ⌋符号
阅读数 20371

ubuntu14 apt-get 简单 安装 ffmpeg
阅读数 19721

最新xcode打包IPA（完整详细图文）
阅读数 13737

如何实现1080P延迟低于500ms的实时超清直播传输技术
阅读数 13068

最新评论

语音识别学习资料入门整理
ZONGDAOFU：这是招聘吗?

程序员经典电子书下载(超全)
wangyongshuai88：楼主大大，链接都失效了，方便分享一个网盘吗

iOS 11 删除 Main.st...
mandagod：别的方法都已经过时，请参考本文...

解决adbd cannot run...
qq_28903151：你们这些抄袭的真的死码

螺钉螺母的匹配问题
qq_41256099：看不懂