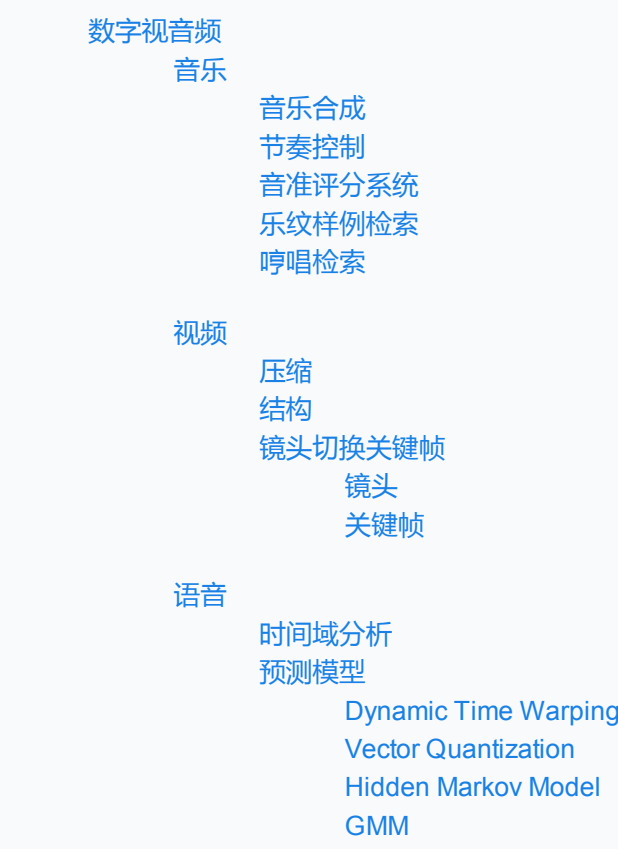


# 数字视音频



## 音乐

### 音乐合成

-	-
音高	频率
音强	振幅
音长	时值
音色	发声体谐波特性

Nyquist[www.cs.cmu.edu/~music/nyquist/](http://www.cs.cmu.edu/~music/nyquist/)) (`play (seq (osc 50) (osc 69))`)  
[OpenAL](#)  
[Matlab](#) ( 这都行？尽情地扯吧 )

### 节奏控制

节拍 < 节奏 < 旋律  
时域 频域 分析

### 音准评分系统

旋律 合拍 节奏 情绪 音色 歌词匹配

### 乐纹样例检索

代表一段音乐重要声学特征的数字签名  
鲁棒性 区分性  
时域 频域 分析  
特征点对索引 三重链表查询 Song Featruue ( Record )

## 哼唱检索

旋律三层表示  
声学层 基于帧的音高序列  
音符层 音符序列  
轮廓因子 乐句旋律  
规模逐渐减小  
  
轮廓因子索引  
Midomi 公司

## 视频

### 压缩

83:1  
帧间帧内信息冗余  
空间 降低编码分辨率  
时间 相邻帧间  
色彩 YUV 1.0 0.25 0.25编码 即UV按照Quarter SIF  
  
按时序相减，获得帧间差，对高速移动物体效果差  
所谓压缩 转化（可逆）-> 量化（信息丢失）-> 编码  
基函数

$$X_k = \sum_{n=0}^{N-1} x_n \cos \left[ \frac{\left(n + \frac{1}{2}\right) k}{N} \pi \right] \quad k = 0, \dots, N - 1.$$

运动检测  
  
首帧作参考 目标帧分块并与参考帧对比。找到 $N \times N$ 宏块间对应的运动向量。一般近邻搜索。线性查找太费时，2D对数查找。层级查找，不同分辨率。

H.261  
I/P Frame  
View and Encode IPPP IPPP IPPP  
参考帧 YCbCr 4:2:0 -> DCT -> Quantinization -> Entropy Coding  
预测帧 找到运动向量，编码差别矩阵

MPEG-1  
I/P/B Frame  
View:  
I BBP BBP BBI  
Encode:  
I PBB PBB IBB  
Different Intra Quantinization number

MPEG-2  
Interlaced  
Layered coding support  
Nonlinear Quatinization

MPEG-4  
Object Based Encoding  
Video Object-oriented

### 结构

帧  
镜头 用关键帧来表示镜头  
镜头组 场景中镜头可以分组，相似镜头成组  
场景 若干镜头组成的

镜头边缘检测 关键帧提取  
时（所有帧直方图均值）空（关键帧直方图的差）特征 五元组  
时间可适性成组 颜色 运动 相似性  
Elastic Matching  
Dynamic Time Wrap  
场景结构构造

足球比赛精彩场景 识别 概率融合

## 镜头切换关键帧

### 镜头

切 淡入淡出 溶解 抹去

帧间差  
亮度和差 变化点数目（单像素比较，宏块比较）直方图（相似度，分量分别取最小，然后除以本帧向量）  
DCT

图像矩不变量  
运动检测 光流 子块匹配

双阈值 渐变检测  
自适应阈值 阈值为滑动窗口中所有值的平均 阻尼函数

灰度均值曲线检测溶解

### 关键帧

先判别镜头  
特征差比较 运动向量  
（模糊？）聚类

## 语音

voiced浊音声带震动  
unvoiced清音

激励 -> 声道 -> 口唇（高频滤去）  
声管串联 共振峰（元音3个，辅音5个）

## 时间域分析

量化  
音量 振幅平方和取对数  
过零率  
端点检测 False Rejection/Acceptance  
音高基音周期 提取 倒谱法提取基频

梅尔倒谱系数  
预加重 >800Hz  
分帧 10-30ms  
加窗 FFT 三角带通滤波 DCT

时域到频域

梅尔倒谱 线性预测倒谱

# 预测模型

## Dynamic Time Warping

端点检测

## Vector Quantization

某一区域的矢量归为一类 划分向量空间（聚类）构成码本 寻找最小失真映射

## Hidden Markov Model

n个状态 一阶假设（仅与前一个相关）  
状态转移矩阵 初始概率  
只观察到了事件序列 选取的缸被隐藏

## GMM

参考别的，讲得实在太扯

第一部分选择题，7个题10个空，10分

- 1. 音乐制作的步骤
- 2. 音乐处理技术
- 3. 语音信号处理方法
- 4. 特征提取预处理的目的
- 5. 关键帧提取算法
- 6. 镜头相似度提取的特征
- 7. 视频目录构造的过程

第二部分判断题，3个题，错误的话需要说出错误理由，每题两分

- 1. Nyquist API的功能
- 2. MFCC
- 3. Group的定义

第三部分，简答题，6道题，每题7分

- 1. 音乐的四要素及具体定义
- 2. 三种音乐编程方法
- 3. PITCH提取
- 4. 端点检测算法
- 5. 运动向量搜索的三种方法
- 6. 时序结构图的构造

第四部分，论述题，三个题，共42分

- 1. 音乐检索的现状及未来发展预测
- 2. 说话人识别图，以及改进方法
- 3. I、P、B，视频压缩方法，主要是P帧

考题基本与去年一样，列出不同的或者补充  
选择题

（谱曲 作词 混音之类的）音乐制作的步骤  
音乐处理技术

（旋律检测还是节奏检测？）Midomi 公司用了什么技术  
（频域 时域 实频域 复频域）语音分析的方法有那三个？ACF属于什么技术？  
语音处理中高频部分需增益  
（扯淡题，算法名都是出题老师自己起的吧）哪些不是关键帧提取算法

判断题

（错，这个应该是在讲OpenAL）Nyquist具有什么什么功能，而且Nyquist根本没有API  
（错，顺序可能有误，考前记得把顺序背熟）MFCC  
（不确定）视频组的定义

简答题

说话人识别图  
语音的短时平稳理论  
写出检测运动向量的3种方法  
（和视频目录构造有什么区别呢？）构建视频时序结构的过程

论述题

音乐检索的发展现状以及趋势（材料给出了百度关键字检索，盛大哼唱检索，音乐雷达声纹检索）  
描述端点检测算法，不足及改进（给出了一张图，能量和过零，课程材料有）  
IBP帧是什么，详细解释P帧的压缩过程