基于内容的音乐推荐系统研究
Content-Based Music Recommender Systems

孙学波

内容提要

1 课题的主要背景
2 国内外研究现状
3 基于内容的音乐推荐系统概述
4 音乐信号处理基础
5 课题的当前工作
6 问题与展望

1 课题的主要背景

- 随着网络技术的发展和应用的普及，人们寻找音乐和消费音乐的方式已经发生了根本性的变化，人们可以在任何时间和任何地点收听和欣赏音乐。人们可以从网络音乐系统在线欣赏或下载音乐。
- 为了提高音乐系统的服务质量，音乐系统不仅仅是帮助用户存储和传输它们喜欢的音乐，而更需要根据用户的欣赏习惯和嗜好，向用户推荐用户没有欣赏过的他们更喜欢的音乐。
- 好的音乐推荐系统应该能够帮助用户欣赏到他们可能更喜欢音乐，也应该能够帮助音乐人发布它们制作的音乐。还应该能够帮助音乐系统的持有者实现他们的目标。

1 课题的主要背景

- 传统音乐推荐系统的推荐方法都是基于分析用户欣赏习惯或协同过滤算法等多种方式。显然，一个好的基于内容的音乐推荐系统似乎应该是更好的音乐推荐系统。
- 其内容是用计算机分析和理解音乐。并以此为基础向用户推荐新的可能用户更喜欢的音乐。从而进一步提高音乐推荐系统的质量。

2 国内外研究现状

在基于内容的音乐推荐方面，相关的文献较少。

但应该说还是有数量相当可观的基础方面的研究，因为我们没有从事过这方面的研究。因此，对这方面的内容知之甚少。

现有的基于内容的音乐推荐方法均集中在“基于频谱分析的推荐方法”方面，即将数字音乐的波形信号转换成用使用某种形式表示的频率信号。再以此为基础定义音乐特征，进而进行音乐分类进行推荐。

这一基本的特点决定了这种系统的复杂度。

2 国内外研究现状

关于音乐推荐系统，我们在查阅到的文献中看到了下面三种方法。

- 傅里叶变换方法，包括傅里叶变换、快速傅里叶变换变换。
- 小波分析方法，可以解决傅里叶变换中频域信息中的时域信息丢失问题。
- 分形方法，将分形维数定义成音乐的特征。
- 其它方法？

3 基于内容的音乐推荐系统概述

- 系统目标：
  - 为用户推荐他们可能喜欢的音乐。
- 核心问题：
  1. 音乐相似性度量问题
     - 定义新的特征并提出一种新的特征提取框架或方法。
     - 这可以使用两种方法：一种是基于某种层级特征的相似性估计，另一种方法是在估计相似度之前对语义标记空间进行映射。最后，再将两种方法结合。
  2. 音乐推荐系统的结构问题。
     - 在更高的抽象层次上，音乐推荐系统可以被解释成一个推荐网络。
     - 基于推荐网络的分析表明，一个向前的Top-N推荐方法可以显著影响音乐推荐系统的可用性。

3 基于内容的音乐推荐系统概述

推荐系统的主要特点：

- 基本的推荐方法：是根据用户兴趣 I_{ui}，从可用的项目集I={ i_1，…，i_n}中为特定用户预测一个项目子集 I_{ua}。
- 音乐推荐的特定场景：活动项，即用户当前关注的特定项目，活动项通常可以定义成当前用户正在收听的一个曲目。推荐项目的数量应远小于项目的总数。

3 基于内容的音乐推荐系统概述

- 音乐推荐系统的推荐方法
  - 基于元数据的方法（Metadata-based Approaches）
  - 基于内容的方法 （Content-based Recommendation Approaches）
  - 混合方法（Hybrid Recommendation Approaches）

3.1 基于元数据的方法
Metadata-based Approaches

- 不使用直接从音频信号推导出来的音乐信息，而是使用与特定的音乐曲目相关的元数据（人文信息）进行的推荐。
- 这些元数据包括：音乐流派、标签、艺术家、歌曲名称、用户等级等元数据，也可以包括音乐的购买量、播放次数或跳过率等统计信息。
- 所有基于元数据的方法都是基于协作的方法，因为任何人都不能完整地诠释整个音乐世界。
- 元数据推荐方法的四种类型
  - 基于协同过滤的方法
  - Web挖掘方法
  - 基于标签的方法
  - 基于专家的方法

3.2 基于内容的方法
Content-based Recommendation Approaches

- 基于内容的推荐是一种直接分析数字对象内容的推荐。这种方法往往用于图像、视频、音乐或电子图书等数字对象的推荐。其基本方法是从数字对象中提取有意义的数字特征并产生可理解的对象表示。
- 基于内容的推荐源自于项目之间的相似性。因此，推荐算法可以分成两个步骤。
  - 首先，从数字对象的内容中提取特征，并将这些特征组合为有意义的对象表示。
  - 其次，为这个对象表示定义一个符合人类感知特点项对向项（item-to-item）的相似度函数。

3.2 基于内容的方法
Content-based Recommendation Approaches

- 基于内容的推荐方法的优点
  - 不需要元数据且完全可扩展的，可产生完全未知的和未评级的歌曲推荐。
  - 生成的推荐在某种程度上是客观的，不倾向于任何项目。
- 基于内容的方法的主要缺点。
  - 无法捕捉任何上下文信息或文化信息。只能从数字对象的内容中提取信息，并用于生成建议。
  - 在实践中，通常很难提取和充分地将数字对象内容内的信息映射到感兴趣的有意义描述。
- 因此，基于内容的推荐方法的主要问题是其推荐质量常常不令人满意。提高基于内容的音乐推荐系统的推荐质量是本课题的主要目标。

3.2 基于内容的方法
Content-based Recommendation Approaches

- 对于视频、文本、图像和音频等领域，如何定义一组合适的特征和相应的相似度函数仍是一个有待研究的问题。
- 本课题要关注的音乐相似度算法和音乐曲目的特征表示的定义以及相应的相似函数。

3.3 混合推荐方法
Hybrid Recommendation Approaches

- 混合推荐系统的思想是将两种或更多推荐技术结合起来，以获得更好的性能，同时消除底层推荐方法的特定问题。
- 在音乐推荐中，混合系统还没有得到广泛的研究。在最简单的情况下，将基于元数据的方法和基于内容的方法相结合，以提高推荐质量，同时解决基于元数据的方法的冷启动、稀疏性或偏好偏差问题。
- 利用基于内容的项目相似性来推断没有标签模型的歌曲的标签模型。另一个简单的混合方法已经被提出。已经提出了更高级的组合。虽然通过这些第一混合算法获得的结果看起来很有希望，但据笔者所知，目前还没有任何基于混合推荐方法的商业解决方案。不过，混合推荐系统似乎是音乐推荐的未来，因为它们可以缓解个别方法的弱点。

4 音乐信号处理基础
Music Signal Processing Fundamentals

- 1 数字音频信号
  - 采样频率：$f_s = 1 / s$ 是采样时间间隔的倒数，并以赫兹（Hz）给出。
  - 音频解码：在压缩编码音频（PCM）中重构原始数字音频信号，这个过程必要但也十分耗时。
  - 重采样：以确保特征提取阶段的所有输入信号具有相同的采样率。
    - 常见的采样频率有8KHZ，11kHz、22kHz、 44.1KHz、96KHZ或192KHz等。

4 音乐信号处理基础
Music Signal Processing Fundamentals

- 2 频谱分析Spectrum Analysis
  - 1）离散傅里叶变换：Discrete Fourier Transform (DFT)
    - 将波形信号转换成频率信号

$$X[k] = \mathbf{DFT}(x[n])] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi nk/N} \qquad k = 0, 1, \ldots, N-1. \qquad (2.1)$$

  - 2）逆离散傅立叶变换： inverse Discrete Fourier Transform （iDFT），
    -

$$x[n] = \mathbf{iDFT}(X[k]) = \frac{1}{N}\sum_{k=0}^{N-1} X[k]e^{j2\pi nk/N} \qquad n = 0, 1, \ldots, N-1. \qquad (2.2)$$

4 音乐信号处理基础
Music Signal Processing Fundamentals

- 2 频谱分析Spectrum Analysis
  - 3）频谱强度：magnitude spectrum

$$|X[k]| = \sqrt{\mathrm{Re}(X[k])^2 + \mathrm{Im}(X[k])^2} \tag{2.3}$$

  - 4）频谱相位：phase spectrum ϕ[k].

$$\varphi[k] = \arctan \frac{\mathrm{Im}(X[k])}{\mathrm{Re}(X[k])} \tag{2.4}$$

4 音乐信号处理基础
Music Signal Processing Fundamentals

$$X[k]_{dB} = 20\log_{10}(|X[k]|) \tag{2.5}$$

4 音乐信号处理基础
Music Signal Processing Fundamentals

- 图2.1展示了从瓦格纳歌剧中获取的真实世界音频信号x[n]及其幅度谱。
- 在最下面的图中，频率以赫兹表示。音频信号在22,000 Hz采样。幅度谱是围绕奈奎斯特频率（11 K Hz）进行镜像的。
- 对于音频信号，情况总是如此，因为它们总是被重估。因此，在MIR中，频谱通常只考虑到奈奎斯特频率。
- 关于MIR，DFT的线性频率分辨率不够理想，因为线性频率分辨率与人类对声音的感知并不对应。

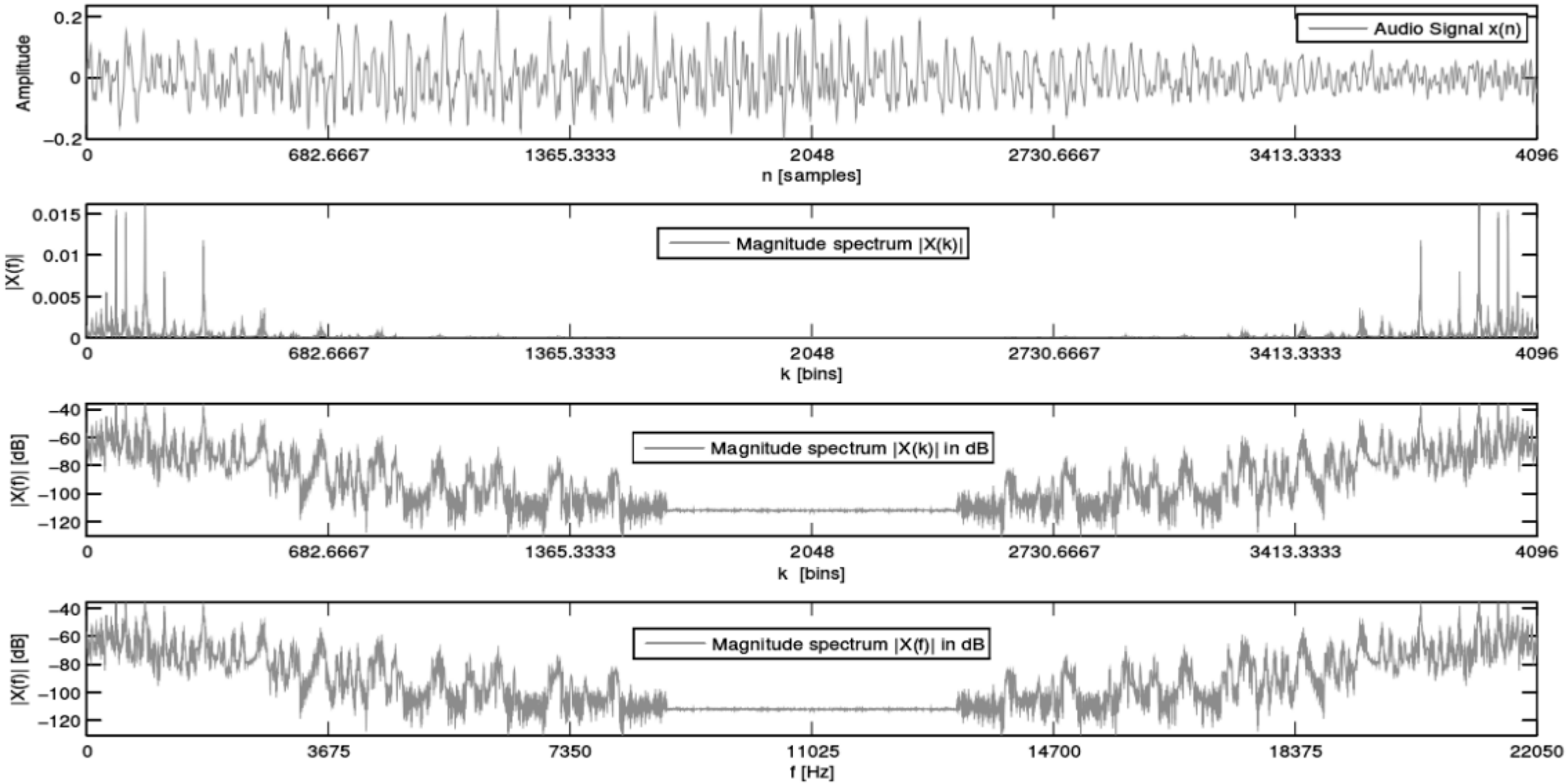


Figure 2.1: Visualization of an audio signal, the magnitude spectrum and the magnitude spectrum in dB.

4 音乐信号处理基础
Music Signal Processing Fundamentals

- 3 听觉尺度（Auditory Scales）
  - 音频信号的分析应考虑到人对声音的感知，只对音频信号的感知相关信息进行进一步处理。
  - 一方面，人类听觉系统的频率分辨率不是线性的，而是对数的。
  - 为了解释人类对声音的对数感知，DFT的线性频谱通常被压缩。最常用的从线性到对数频率分辨率映射的听觉音阶是Mel Scale、Boek、ERB和Cent四个尺度，线性频率尺度（FHz）与对数频率尺度之间的关系可由一个公式定义。

4 音乐信号处理基础
Music Signal Processing Fundamentals

- 3 听觉尺度（Auditory Scales）定义
  - 1）Mel-Scale

$$f_{\text{mel}} = 2595 \log_{10}\left(\frac{f_{\text{Hz}}}{700} + 1\right)$$

  - 2）Bark-Scale

$$f_{\text{bark}} = 13 \arctan(0.00076 f_{\text{Hz}}) + 3.5 \arctan((f_{\text{Hz}/7500})^2)$$

4 音乐信号处理基础
Music Signal Processing Fundamentals

- 3 听觉尺度（Auditory Scales）定义
  - 3）ERB-Scale

$$BW_{\mathrm{Hz}} = 24.7\,(0.00437 f_c + 1)$$

  - 4） Cent-Scale

$$\Delta f_{\mathrm{cent}} = 1200 \log_2 (\frac{f_a}{f_b}).$$

4 音乐信号处理基础
Music Signal Processing Fundamentals

- 4 听觉尺度的比较

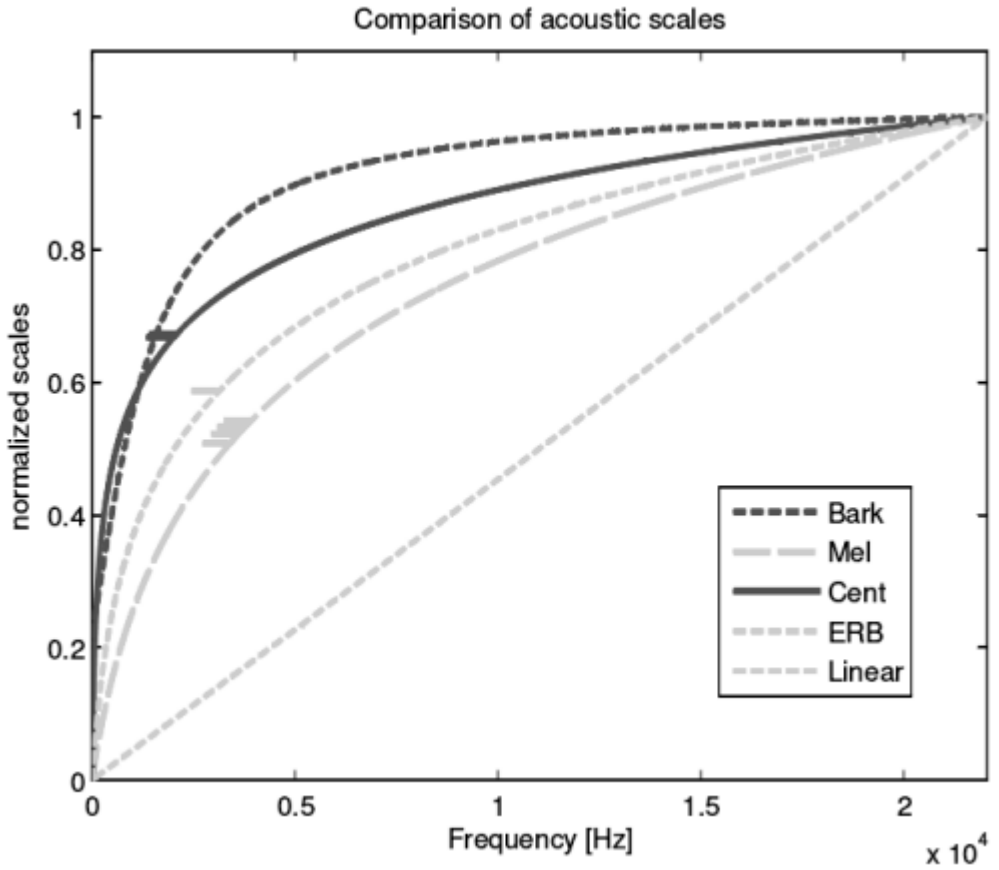


Figure 2.2: Comparison of auditory scales.

4 音乐信号处理基础
Music Signal Processing Fundamentals

- 5 时频表示
  - 傅立叶变换的一个假设是被分析的信号x(t)是一个平稳信号，这基本上意味着频率的内容随时间的变化是恒定的。不幸的是，音频信号的频率内容通常随时间而变化。因此，音频信号属于所谓的非平稳信号的范畴，但幸运的是属于所谓的准平稳信号的特定子类。准平稳信号是非平稳信号，但可以被建模为在局部时间帧内是静止的。为了能够捕获随时间变化的信号特性，使用时间频率表示（TFR）。

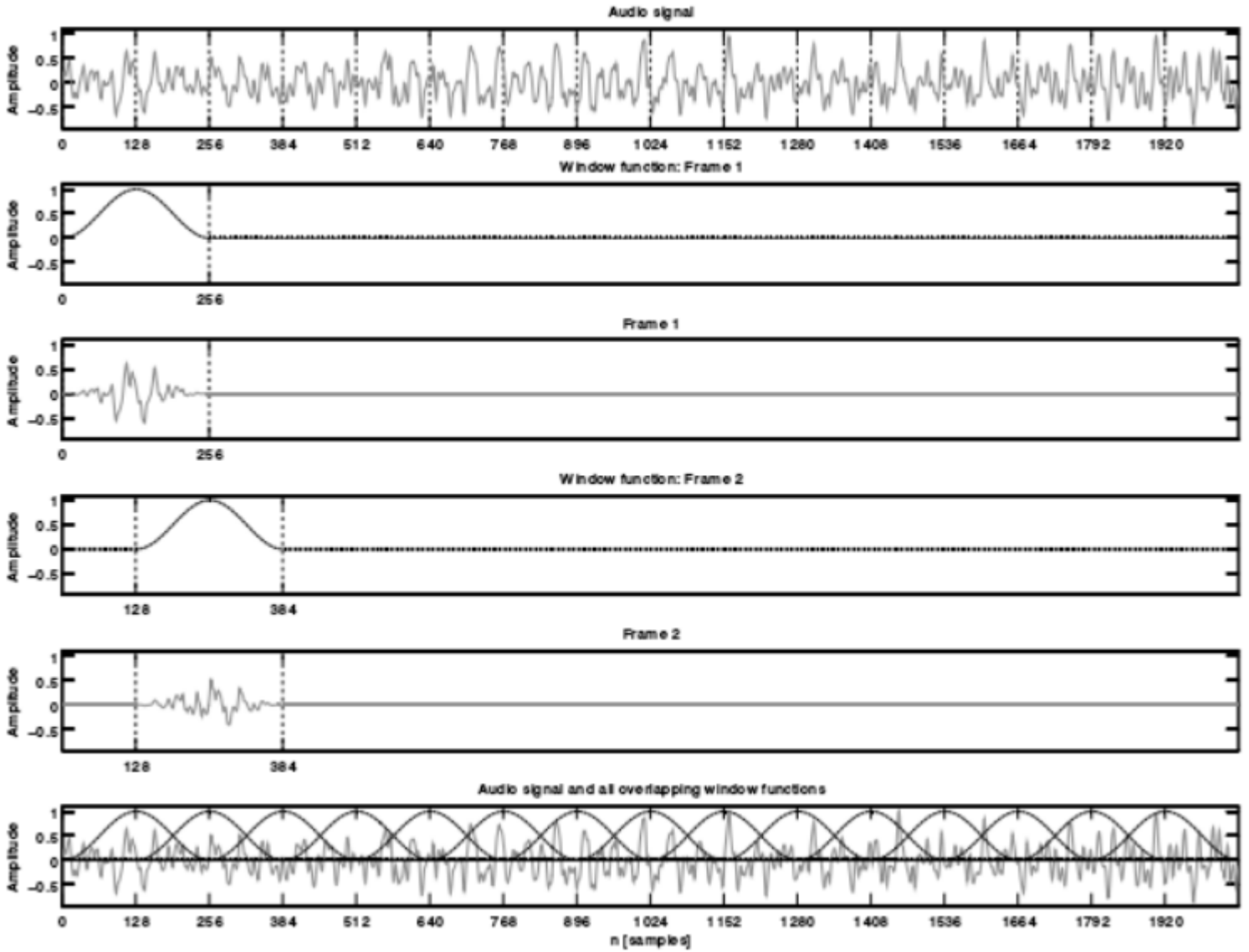


Figure 2.4: Windowing of an audio signal using a Hanning window function .

4 音乐信号处理基础
Music Signal Processing Fundamentals

- 数学上，STFT可以根据下面公式定义，其中W[n]是窗口函数，m定义窗口的位置。这里所用的STFT的定义在时间和频率上都是离散的。

$$\mathbf{STFT}\left[x[n]\right] \equiv X[m,k] = \sum_{n=0}^{N-1} x[n]w[n-m]e^{-j2\pi mk/N}$$

4 音乐信号处理基础
Music Signal Processing Fundamentals

- 2）窗口函数
- 众所周知，用于从音频信号中“切割”单个音频帧的窗函数W[n]对获得的离散傅立叶谱具有不可忽略的影响。根据卷积定理可以导出窗口函数对频谱的影响。假定信号x[n]的离散傅立叶变换是x[k]，窗口函数W[n]的离散傅立叶变换是W[k]，

$$x[n] \overset{\mathcal{F}}{\leftrightarrow} X[k]$$

$$w[n] \overset{\mathcal{F}}{\leftrightarrow} W[k]$$

- 音频帧y[n]是音频信号与窗口函数的乘积。

$$y[n] = x[n]w[n]$$

4 音乐信号处理基础
Music Signal Processing Fundamentals

- 然后，窗信号Y[n]的离散傅立叶谱是离散傅立叶变换的信号和窗函数的离散傅立叶谱的卷积。

$$Y[k] = X[k] \star W[k]$$

- 因此，原始频谱的每个频率分量与窗函数的傅立叶变换卷积。

4 音乐信号处理基础
Music Signal Processing Fundamentals

- 例如，频率ω的一个简单余弦信号的窗口使其傅立叶变换在ω以外的频率上具有非零值，尽管分析的信号不包含这些频率分量。这种现象称为泄漏现象。窗口函数的傅立叶变换定义了频谱如何涂抹这些频率分量。
- 最坏的情况是信号的频率分量由于泄露效应而不能得到解决。窗口函数的傅立叶变换的帧的宽度定义了两个频率分量的接近程度，直到一个不再能解决它们。存在各种窗口函数，例如Hanning、Kaiser或矩形窗口。
- 在MIR中，汉宁窗口是相当流行的，人们通常将其作为标准窗口函数。

$$w[n] = 0.5(1 - cos(\frac{2\pi(n-1)}{N})), \qquad 0 \leq n < N$$

4 音乐信号处理基础
Music Signal Processing Fundamentals

- 虽然这是可取的，但是时间频率表示不能具有任意的时间和频率分辨率。对于STFT，这种时间和频率分辨率之间的交易称为不确定性原理。不确定性原理指出，不能同时增加STFT的时间分辨率和频率分辨率。时间分辨率**Δ**T是由窗口大小定义的。假设音频帧包含K个样本，时间分辨率为**Δ**T =KTs，其中Ts是采样周期。STFT的频率分辨率**Δ** f是两个连续傅立叶系数之间的频率间隔，即**Δ**f＝Fs/K，因为离散频率箱是傅立叶谱的等距样本。

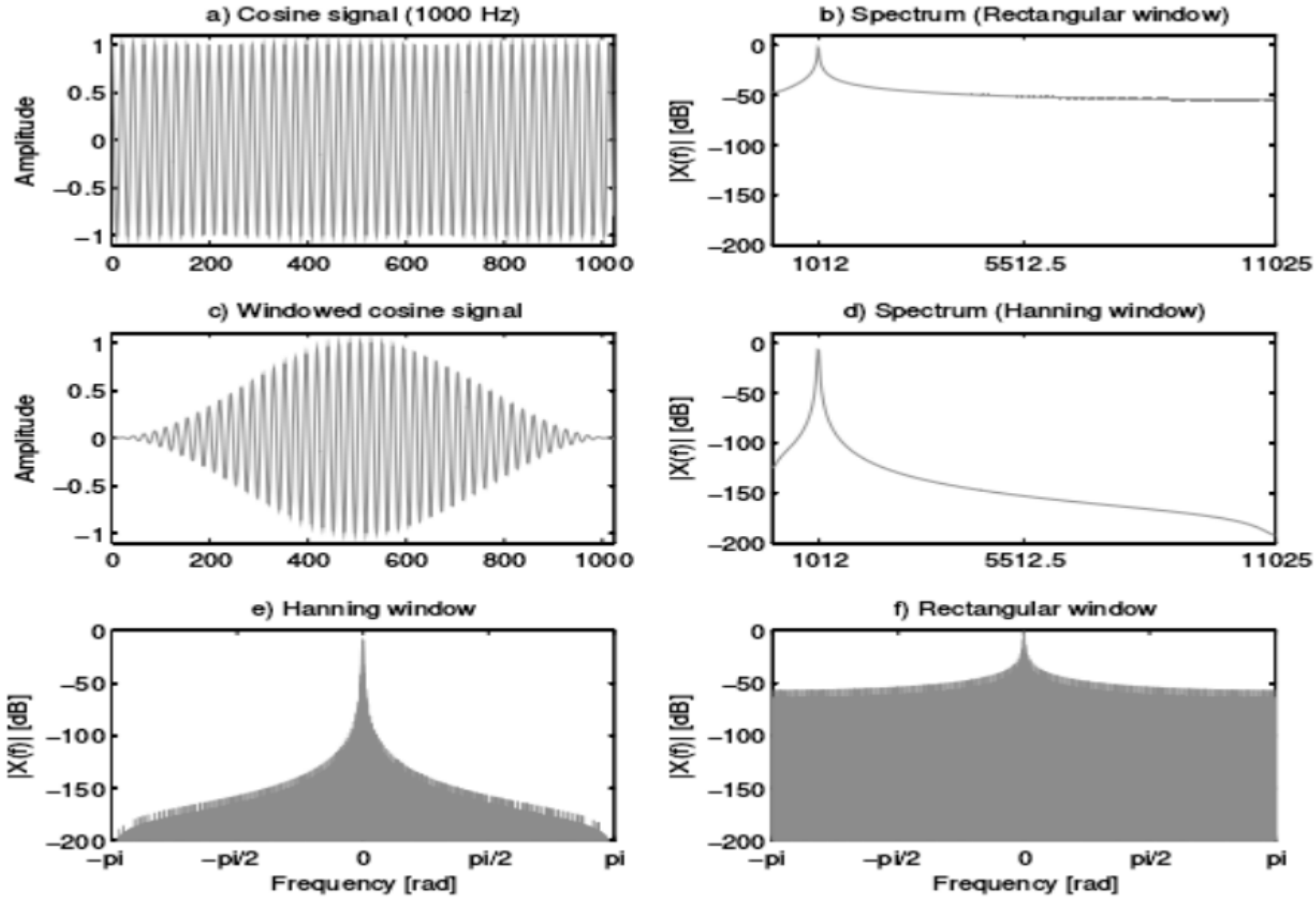


Figure 2.5: Windowing of an audio signal using a Hanning window function

.

4 音乐信号处理基础
Music Signal Processing Fundamentals

- 6 音频标准化(Audio Normalization)
- 音频标准化是一个重要的预处理步骤，因为音频文件经常以不同的音量水平记录。从技术的角度来看，这意味着整个音频信号x[n]是由一个常数因子A放大的。

$$\hat{x}[n] = ax[n]$$

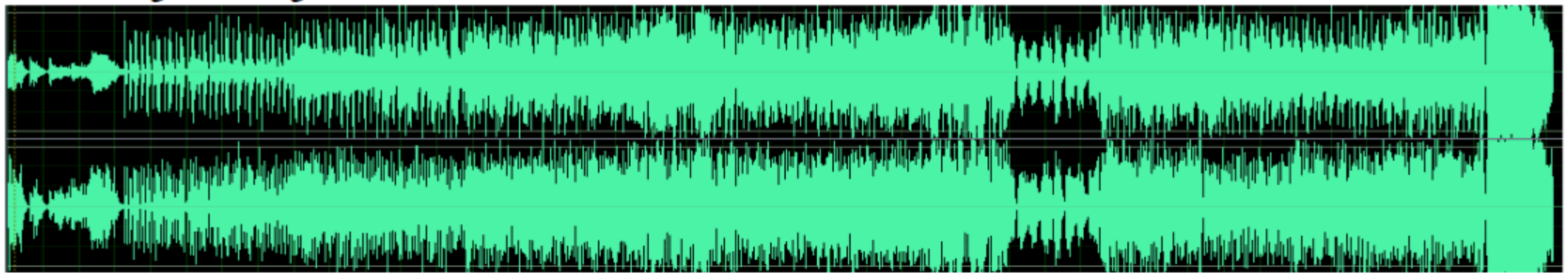
- 当傅立叶变换是线性变换时，放大系数的幅值谱x[k]被常数因子a缩放。

$$|\hat{X}[k]| = a|X[k]|$$

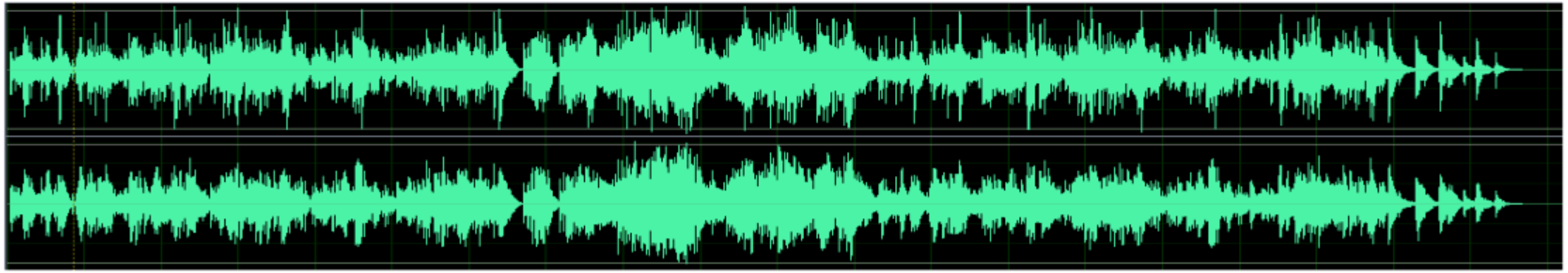
5 课题的当前工作

- 两首音乐的波形图：

Hotel of california



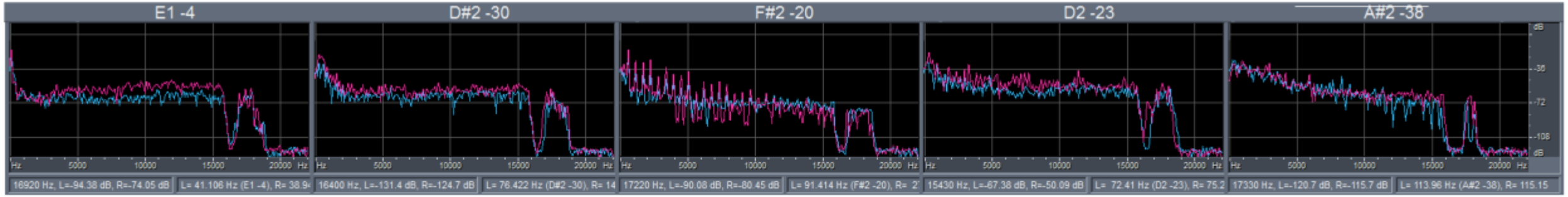
Deer Hunter



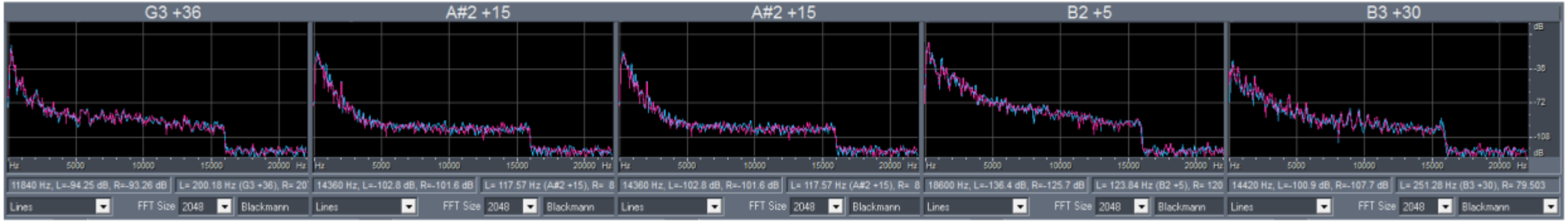
5 课题的当前工作

- 两首音乐的频谱分布比较：

Hotel of california



Deer Hunter



5 课题的当前工作

- 试图设计一个“基于频谱分析的音乐推荐方法”
  1. 音乐分析：将音乐波形划分成若干个帧，将每一帧转换成频谱信号，计算这些帧之间的相关系数。可得到一首音乐的自相关系数矩阵。通过实验分析音乐的频谱分布的自相关特性的稳定性。
  2. 定义音乐特征：将这些帧定义为音乐的特征。
  3. 音乐比较：
     1. 计算两首音乐的频谱帧的相关系数矩阵。
     2. 以矩阵为基础定义一个用于比较的阈值
  4. 讨论方法的评价

6 问题与展望

- 音乐分类是一个十分复杂的问题，尽管现在的音乐系统已经为数字音乐定义了大量的元数据并存储在音乐文件之中。但这些元数据并不足以描述音乐的所有特征，更不能有效或充分地支持令人满意的音乐推荐。
- 因此，基于内容的音乐推荐将是一个有着巨大发展前景和商业价值的研究领域。其复杂度（研究的难度）也决定了这一领域研究的理论意义。

谢谢各位同事！