基于智能分类算法的音乐和弦识别与分析

——基于PCP模型的实现与改进

摘 要： 和弦的辨别能力是在音乐学习当中必不可少的技能。但是，为进行过专业训练的人很难快速准确的对一首音乐中的和弦进行快速准确的听辨。随着计算机技术的发展，利用快速傅里叶变换与各种智能算法，本文拟使用上述算法，对音乐的和弦进行识别与分析。拟采用由Takuya FUJISHIMA于1999年提出的PCP（Pitch Class Profile）算法为基础，结合DNN等智能算法加以改进，增加音乐的节拍跟踪，实现对于音乐中和弦的智能识别系统。

关键词：音乐，和弦识别，智能算法，PCP，节拍跟踪

Music chord recognition and analysis based on intelligent classification algorithm

—— Implementation and improvement based on PCP model

**Abstract：** The ability to distinguish chords is an essential skill in music learning. However, it is difficult for a trained person to listen quickly and accurately to the chords in a piece of music. With the development of computer technology, we use Fast Fourier transform and various intelligent algorithms to identify and analyze the chords of music. Based on the algorithm of Pitch Class Profile (PCP) proposed by Takuya Fujishima in 1999, and combined with intelligent algorithms such as DNN, this paper proposes to improve it by adding the beat tracking of music and realizing the intelligent recognition system of chords in music.

**Key words：Music, Chord recognition, intelligent algorithms, PCP, beat tracking**

1 研究背景及其意义

**1.1 研究背景及其意义**

自动和弦识别算法是指能够从音频信号中提取和弦信息的计算机程序。和弦是音乐中最基本的元素之一，它能够表达音乐的调性、节奏、风格和情感。自动和弦识别算法的研究具有重要的理论意义和实际价值。从理论上来说，自动和弦识别算法可以揭示音乐的结构和规律，为音乐分析、理解和创作提供有力的工具。从实际上来说，自动和弦识别算法可以应用于多种场景，如音乐教育、音乐制作、音乐检索、音乐推荐等。因此，设计一个高效、准确、鲁棒的自动和弦识别算法是音乐信息检索领域的一个重要课题，也是本文的研究目标。

**1.2当前研究现状**

1999年由斯坦福大学的Takuya FUJISHIMA提出的PCP（Pitch Class Profile）模型，为后续的所有和弦识别算法提供了理论基础。随着计算机科学与人工智能模型的发展，后续提出了诸多以神经网络为基础的更加优秀的算法，如深度色度提取与谐波-冲击分离算法。

然而，当前的各种研究都将重点集中于旋律组乐器的频谱分析上，忽略了节奏在音乐中的实际作用。本文拟将和弦识别算法与节拍跟踪算法进行结合，拟构建一个更具有乐理性与稳健性的和弦识别与分析系统。同时，本次实验拟提出一个节拍跟踪算法。下文将详细描述本次实验中拟采用的实验方法。

2 律法

现代音乐音阶中的音以一定的频率点位进行定义，这种定义方式称作律法。历史中出现过很多不同的律法，例如纯律、十二平均律、五度相生律等。现代乐器调律基本按照十二平均律进行调律，在数学上，十二平均律的计算也更加方便。故本次实验仅探讨在十二平均律下的情况。

**2.1十二平均律**

十二平均律将两个频率比为2：1的两个音称作八度（Octave）。在乐理上，这两个音有着同样的音名。

十二平均律将一个八度内的音平均分为十二份，相邻的音的频率比相同。

十二平均律将标准音定为440Hz（标准A），以这个音为标准，计算出其他音的频率。

根据以上特点，可以得出在十二平均律下各个音的计算方法：

将一个八度内的音分为十二份，假设每个音是前一个音的x倍，标准音的频率为F0,易得方程：

解得

设十二平均律中十二个音的下标分别F0、F1……F11,已知Fx的频率，由上述公式易得求x的公式为：

**2.2 音阶**

**2.2.1 半音阶**

上文已经提到了在十二平均律中各个音频率的算法。在音乐理论中，一个八度内的十二个音构成的音阶称之为半音阶，每个音分别被记作：

C、#C/bD、D、#D/bE、E、F、#F/bG、G、#G/bA、A、#A/bB、B

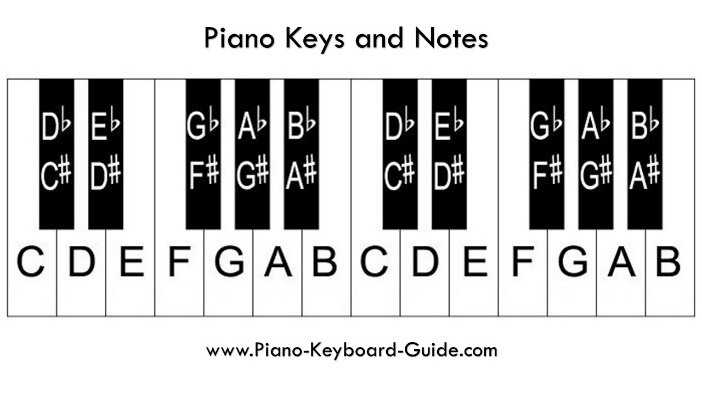
再以八度循环。

图1.1 钢琴键盘

**2.2.2自然大调音阶、自然小调音阶与调式体系中的功能和弦**

十二平均律中的某几个音的组合，组成调式音阶。

在音乐中存在许多各式各样的音阶，如大小调音阶、中古调式音阶、中国五声音阶、日本音阶等。但是在本项目中，我们着重探讨常用的自然大小调音阶的情况。

自然大调的一个八度由七个音组成，音阶之间按照全音、全音、半音、全音、全音、全音、半音的音程关系组成。以C大调为例，此音阶由C、D、E、F、G、A、B七个音组成。C称为音阶的主音（Tonic）。

自然小调与自然大调相似，以半音、半音、全音、全音、半音、全音、全音的音程关系构成。以a自然小调为例，构成音阶的音为a、b、c、d、e、f、g。

以其中的一个音为根音，向上三度与五度构成三和弦，称为调内和弦。以自然大调音阶中的C大三和弦为例，由C、E、G三个音组成，以主音C为根音的和弦被定义为主和弦。同理，音阶中的每一个音都可以作为根音构成和弦，分别记错C、Dm、Em、F、G、Am、Bdim。音阶中的音构成的和弦共同组成功能和弦组。和弦按照一定的乐理规律进行连接，构成和弦进行。

和弦的连接一般遵从和弦功能。最简单的和弦进行及其逻辑基础是：在主和弦之后引入一个或者几个不稳定的和弦，形成明显的紧张性，这种紧张性在进行到或者回到主和弦时得到解决：

*T（稳定）-> 非T（不稳定，紧张）-> T（稳定，紧张的解决）*

下面给出一个具体的例子：

以C自然大调为例，从主和弦开始，在主和弦之后加入几个不稳定的和弦：

*C（主和弦，稳定）->G（属和弦，不稳定）-> C（主和弦，稳定）*

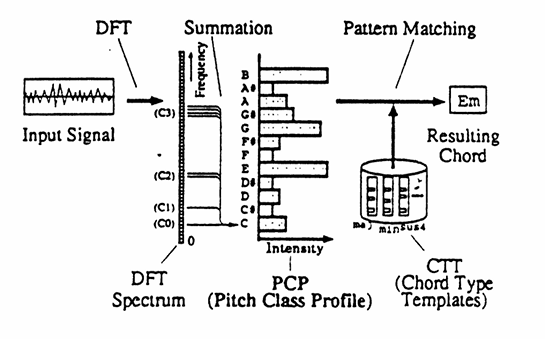
这个进行在古典和声学中被成为正格进行。它揭示了和弦进行中一个最本质的规律：构建紧张感，然后进行解决。在实际的音乐创作中，和弦的进行并不会像这样简单，其中将涉及到附属和弦、同功能组和弦替代等手法，但是这样普遍的规律时后续进行和弦识别与预测的重要参考。

在后续的算法中，将对和弦进行的上下文进行判断，并根据和弦进行的一般规律对后续的和弦进行预测。

以十二平均律中不同的音作为主音，可以按照相同的规律推出所有十二个自然大调音阶与自然小调音阶。不同调式的识别是后续对和弦识别的重要前提条件。

3 利用PCP(Pitch Class Profile)进行和弦识别

PCP(Pitch Class Profile)由斯坦福大学的Takuya FUJISHIMA于1999年提出。这是一个十二维的向量，每个维度表示某个音频片段在十二平均律上每个音的强度。

图3.1使用 PCP进行和弦识别的求解流程

PCP对音频的DFT（离散傅里叶变换）结果进行考察。考察DFT结果中的频率点位在十二平均律中每个音所占比重的大小。在1999年发布于ICMC Proceedings的原始论文中，Takuya FUJISHIMA给出了PCP向量的求解过程：

**3.1 PCP的求解流程**

**3.1.1 音频采样与DCT离散傅里叶变换**

傅里叶变换在信号处理中的作用，一般是将一个时域函数转换为频域函数。一个周期性连续函数的傅里叶函数的复数形式如下所示：

在现实生活中，声音由物体的振动产生。一个质点的简单受迫振动方程一般可以由一个连续的周期性函数表示。但是在数字音频领域中，一个连续的音频信号需要进行采样，转换成一个离散的函数进行储存。对音频的采样速率使用采样率表示，单位为Hz。根据奈奎斯特-香浓采样定理，如果周期函数 x(t) 不包含高于 B cps（次/秒）的频率，那么，一系列小于 1/(2B) 秒的x(t)函数值将会受到前一个周期的x(t)函数值影响。

因此 2B 样本/秒或更高的采样频率将能使函数不受干扰。相对的，对于一个给定的采样频率 fs，完全重构的频带限制为 B ≤ fs/2。为了平衡计算量与信号失真，本次实验拟将音频的采样率统一定为44100Hz。

由于在计算机中存储的音频信号为离散形式，故对其进行的傅里叶变换也应该是离散的。下面给出离散傅里叶变化的数学表达式：

其中，数组为DFT频谱，下标k为频率点位。N为FFT大小，为输入信号。在为实数的情况下，由于每个频率点位存在一堆共轭的复数结果，至的结果即为完整的DFT频谱。由于仅对信号在各个频率点位的强度（振幅）进行考察，故DFT频谱的结果应对复数结果取模。

**3.1.2 Pitch Class Profile**

在得到了DFT频谱之后，可以对音频的PCP向量进行计算：

定义PCP向量

其中，为一个将频率点位映射至PCP下标的矩阵，其定于如下：

其中，为标准音频率，记为. 为的频率。

**3.2 模式匹配**

PCP向量用于进行模式匹配。和弦类型模板（Chord Type Templates，CTT）为一个十二维的向量，用于判断和弦的种类。若一个和弦是以C为根音的和弦，则

否则为0.

一个M7（大七和弦）和弦的CTT向量如下所示：

**3.3 评分方法**

此文章中使用两种不同的评分方式来确定和弦模式匹配得分：

**3.3.1最邻近算法**

最邻近算法用于求出和弦模式匹配的最小得分：

**3.3.2加权求和法**

加权求和法用于求出和弦模式匹配的最大得分：

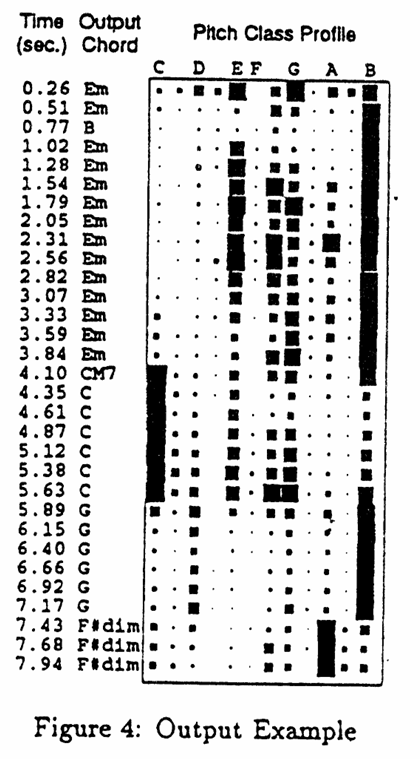
**3.4 结果样例**

图3.2 原始PCP与CTT算法的输出结果样例

使用加权求和法计算出的结果如上图所示。黑色方块为PCP元素的值大小。

**3.5 算法的缺陷与思考**

此算法与1999年被提出，为后续和弦识别方向的研究提供了理论基础。但是此算法也存在诸多明显的缺陷。在此实验中，仅使用钢琴为识别的对象。在现代流行音乐中，配器日渐复杂，和声设计也更加丰富。如何在配器更多的音乐中做到准确的和声识别，成为当今研究的新方向。同时，由于配器日渐丰富，频谱的谐波影响成为了不可忽略的影响因素。同时，节奏组乐器对和弦分析的影响也是不可忽略的因素。

其次，PCP与CTT模型根据PCP向量的突变来判断和弦的改变。但是在日渐复杂的和声对位创作在流行歌曲中的运用，和弦的改变对PCP向量的影响或许并不会很强（例如循环的旋律动机的运用）。故本文将拟提出一个算法，检测音乐的节拍特征，并根据音乐的节拍特征作为可能的和弦转换参考。

4 深度色度提取

基于DNN网络的深度色度提取算法由开普勒大学计算机视觉系的Filip Korzeniowski等人提出。色度图（Chromagram）与PCP类似，是一个带时间序列信息的色度向量，表示了在特定时间中音频的谐波内容。

**4.1 算法流程**

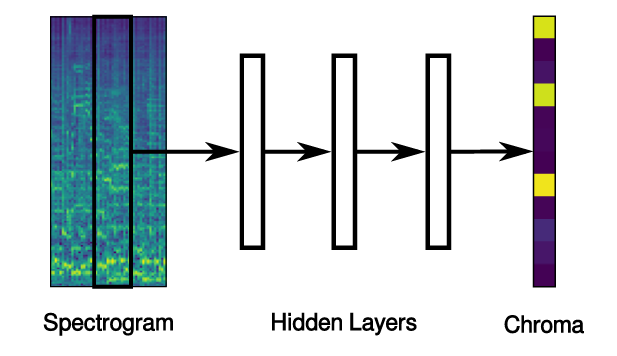


图4.1 利用深度色度提取进行和弦识别的流程

**4.2 输入数据处理**

与PCP的工作流程类似，在对音频信号进行处理前，都需要进行傅里叶变换。但是在本算法中，额外使用了另外的算法对傅里叶变换后的数据进行了进一步的处理：

**4.2.1 四微分频谱（quarter-tone spectrogram）与对数压缩（logarithmic compression）**

传统的短时傅里叶变换在色度提取算法上对比连续Q变换（Constant-q transform）往往给出更差的结果。但是在对短时傅里叶变换的结果进行四微分滤波与对数压缩处理之后，算法对短时傅里叶变换对结果的影响并不敏感。

四微分频谱使用一个滤波器对短时傅里叶变换进行滤波：

其中为滤波乘子。四微分频谱将原频谱压缩至30Hz至5500Hz的范围内，每个八度内被划分为24个块。在进行四微分滤波之后，原来的傅里叶变换结果将被压缩至178个频率节点。

对四微分频谱再应用对数压缩：

将变换后的频谱作为DNN网络的输入。

**4.3 DNN网络**

DNN网络包含L个隐含层hl，每个隐含层包含Ul个计算单元（神经元）。每个神经元的输入由上一层的神经元的输出决定：

其中为神经元输出，为连接权重，为上一层神经元输出，为偏移值，为一个非线性的激活函数。

定义模型预测色度向量

与目标色度向量t

定义损失函数

对模型进行训练。

5 节拍跟踪

当前对计算机算法实现和弦的算法，大多集中于对音频信号的频率进行分析，而忽略了节拍在音乐中发挥的重要作用。在本次实验中，拟将节拍在时间上的特征纳入和弦识别与分析的考量范围之内。在本次实验中，拟提出一个对音乐信号中的节拍进行检测的算法。

**5.1节拍与音乐中节拍的特点**

**5.1.1 拍（Beats）**

A picture containing screenshot, text, font, number

Description automatically generated拍是音乐中组织音符的基本时间单位，所有的音符都基于这一个具有周期性的时间概念来组织。这个概念可以类比计算机科学中“时钟”的概念，它为计算机中所有的操作提供同步的标准。

图4.1 Beat Level表示一拍在小节中的组织方式，Division levels表示节拍的均分

节拍层（Beat level）：指示乐段中节拍的概念，在常见的4/4拍音乐中，通常以四分音符为一拍，每小节的第一拍为重拍，第三拍为次重拍，二、四拍为轻拍。

分拍层（Division levels）：将四分音符进行二次、四次或者更加多次的均分，构成分拍层。换句话说，分拍层由八分音符、十六分音符等组成。

合拍层（Multiple levels）：节拍层的音符的组合称为合拍层。

在本实验中，将分拍层、合拍层的音符与节拍层的音符进行区分，尽可能减小分拍层与合拍层的音符对节拍层音符的判断是本次实验的关键所在。

**5.1.2 小节（Measure）**

拍按照一定的数量与形式组织成小节，组织方式通常用拍号（Signature）表示。以现代歌曲中常见的4/4拍为例子，它表示以四分音符为一拍，每小节由4拍构成。小节是乐句的基本组成。每小节的第一拍被称为下拍（Downbeat）。在本次实验中，下拍时间戳是判断和弦变换点位的主要参考标准。

**5.1.3曲速（Tempo）**

曲速是指每分钟的拍数（Beats per Minute，BPM），这是衡量音乐快慢的量化数据。在本次实验中，将提出一种检测歌曲曲速的算法。

**5.2节拍跟踪算法**

**5.2.1音头**

音头，是指一个音符开始发出声音的瞬间。在实际演奏中，表现为按下琴键（钢琴）的瞬间，或者鼓棒接触到鼓面的瞬间（打击乐器），也就是在谱面上显示出的音符的位置。上文提到，音符一般按照节拍为时间基础进行组织，也就是说，音头一般出现在拍上或者分拍层（Division Levels）上。

A picture containing text, font, line, screenshot

Description automatically generated由于音头是一个音符开始发出声音的瞬间，而一个音符的延音，响度一般会缓慢衰减，即声音的能量会逐渐衰减。故我们可以对频谱的能量在时间维度上的变化来观察音头的大致位置，即能量的极大值点可以认为是一个可能的音头。

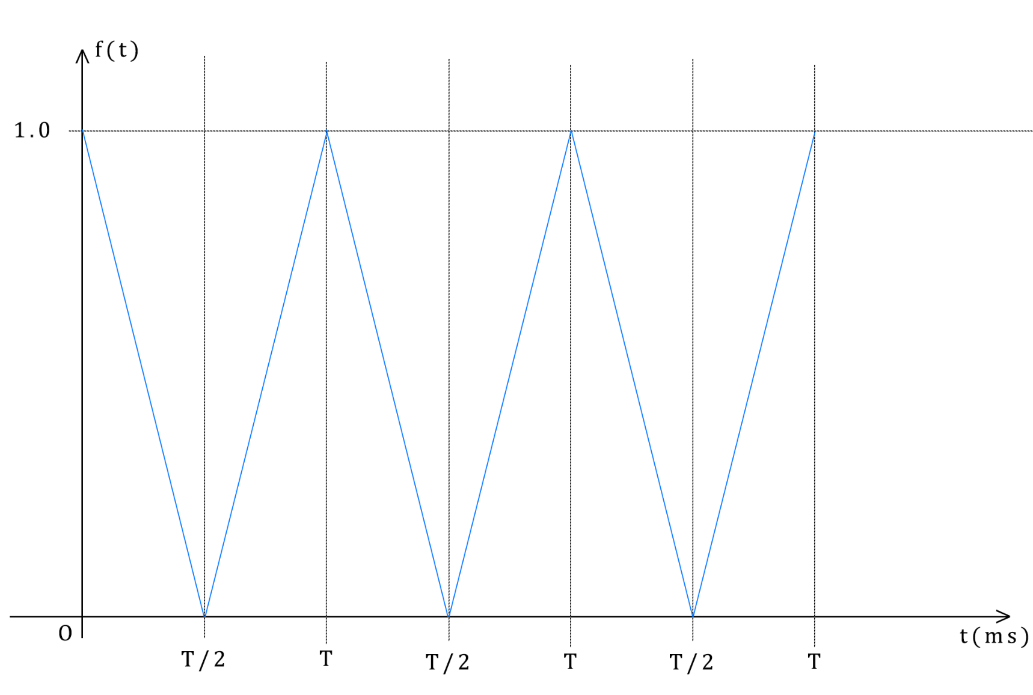
图4.1 一个音乐片段随时间的响度变化

图4显示了一个音乐片段的响度变化。可以看出，音频的响度符合上述音头的响度衰减特征。在本次实验中，拟提出一个算法，通过音频的周期性的响度，检测音频的BPM。

**5.2.2周期乘数法**

本节拟提出一个可能的BPM检测算法：周期乘数法。

拟构建一个简单周期乘数函数：

图5.2 一个最简单的周期乘数函数

其中周期T为当前预测的节拍长度，若要判断歌曲的BPM是否为

则令周期乘数函数的周期

与数字音频采样相同，周期乘数函数也应该是离散的。根据上述公式生成离散周期乘数函数

需要注意的是，周期乘数函数的离散化应与音频分帧的离散化相关，而不是与音频采样点的离散化相关。也就是说，下标与音频帧下标相同，而不是音频采样点。

设每一帧的离散傅里叶变换（DCT）结果的能量谱线

A screen shot of a computer

Description automatically generated with low confidence其中为一个均衡器（Equalizer，EQ）函数，它的数学意义是DCT变换结果的频率点位在能量计算中的权重，物理意义是能量谱线对某个频率点位的敏感程度。在实际音乐中，节奏组乐器（打击乐器）对节奏的指示性相比于旋律组乐器更加显著。而节奏组乐器的频率特征与旋律组乐器有着明显的区别：

图5.3 一个典型的节奏组乐器的频率响应特征

上图显示了一个典型的节奏组乐器（架子鼓）中各乐器的频率响应特征：

手击镲（HiHat）：图中蓝色框部分：集中在高频部分，中心频率10KHz，频率范围从5KHz至20KHz。

小军鼓（Snare）：图中绿色框部分，在中频与高频部分均对频谱有着较大的冲击。

合理设计均衡器函数，可以使得能量谱线对节奏组乐器的频率响应更为敏感，提高识别的准确率。

计算待测试乐段能量谱线在当前预测BPM下的评分：

其中为乐段的帧总数。即为周期乘数。

**5.2.3带下拍检测权重的周期乘数法与低频振荡器（LFO）**

上文提到的周期乘数发检测乐段的BPM存在一个致命的缺陷：容易受到分拍层的影响。假设一个乐段中在分拍层有着密集的音符（常见于歌曲的副歌部分），简单的周期乘数函数将受到分拍层的音符较大的影响。于是在本节提出一个周期乘数法的改进形式：带下拍权重的、以低频振荡器生成的非线性周期乘数函数：

A picture containing text, line, diagram, plot

Description automatically generated上文提到，在一个常见的4/4拍的乐段中，每小节的第一拍为重拍，第三拍为次重拍，二、四拍为轻拍。那么可以考虑将周期乘数的函数的下拍、重拍与轻拍赋予不同的权值，并引入一个由低频振荡器生成的非线性的周期乘数函数，来提高系统的稳健性（Robustness）。

一个带下拍检测权重的简单周期乘数函数形如下图所示：

图5.4 一个带下拍检测权重的简单周期乘数函数

带下拍检测权重的周期乘数函数赋予下拍（通常为能量更高的拍）更高的权重，从而降低分拍层音符对乐段节拍跟踪的影响。

低频振荡器（Low Frequency Oscillator，LFO）是一个能生成低频周期脉冲的电子元件，将不同的周期函数进行叠加，可以生成复杂的周期性的非线性函数。它一般被用作电子音乐制作中的合成器乐器设计上。受到合成器设计的启发，利用低频振荡器的波形叠加，可以生成复杂的非线性的周期乘数函数。利用代码生成一个LFO的算法，相比于直接生成一个非线性的周期乘数函数要更加便捷，也更加容易调试。利用低频振荡器的这个特性，可以将一个复杂的周期非线性函数转换为几个简单的周期非线性函数之和，而不需要考虑复杂的数学表达式。

图5.5 电子合成器中的一个低频振荡器波形

**5.2.4谐波-冲击分离算法**

2008年由东京大学的Nobutaka Ono等人提出的谐波-冲击分离（Harmonic-percussive source separation）算法利用支持向量机（SVM）实现了节奏组乐器与旋律组乐器的频谱分离。这个算法通常被用作色度提取的数据预处理环节。本次实验也将尝试利用此算法对音频信号进行预处理，同时，也将此算法运用于节拍跟踪算法，将谐波-冲击分离算法与均衡器函数结合，有望提高系统的稳健性。

参考资料：

[1] 李重光. 音乐理论基础 [M].北京：人民音乐出版社,1962:253.

[2] 伊·杜波夫斯基，斯·叶甫谢耶夫，伊·斯波索宾，符·索科洛夫. 和声学教程 [M].北京：人民音乐出版社，2008:435

[3] 阿诺德·勋伯格. 作曲基本原理 [M]. 上海：上海音乐出版社，1984

[4] 同济大学数学系. 高等数学 下册 第七版 [M]. 北京：高等教育出版社，2014：325-326

[5] 王万良. 人工智能导论 第五版[M]. 北京：高等教育出版社，2020：211-250

[6] Takuya FUJISHIMA，Realtime Chod Recognition of Music Sound: a System Using Common Lisp Music[J]. ICMC Proceedings,1999,464-467

[7] Filip Korzeniowski and Gerhard Widmer. “Feature Learning for Chord Recognition: the Deep Chroma Extractor”[J],17th International Society for Music Information Retrieval Conference,2016

[8] Siddharth Sigtia, Nicolas Boulanger-Lewandowski, Simon Dixon. “Audio Chord Recognition with a Hybrid Recurrent Neural Network” [J], 16th International Society for Music Information Retrieval Conference, 2015.

[9] Ono, Nobutaka, et al. "Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram." [J]，2008 16th European Signal Processing Conference. IEEE, 2008.