音乐信号的时频分析

谢秀琴，刘若伦

（山东大学威海分校信息工程学院，山东威海 264209）

摘要：对已用于音乐处理上的主要时频(TF)分析方法进行了综述，并描述了它们的应用范围。在 Cohen 类框架下对技术方法进行检验，有助于新方法的设计和比较。所述内容涵盖了基音同步分析，短时傅立叶变换(STFT)，相位声码器，恒 Q 变换和小波变换，Wigner 分布和模式分布。详细描述了窗函数法的局限性和对稳态假设、无限持续正弦曲线的依赖性。与窗函数法相反，Wigner 分布用解析信号定义瞬时频率和功率参数，而模式分布是对优化的 Wigner 分布的线性变换。应用领域涉及分析、再合成、记谱、分类和识别、可视化。

关键词：时频分析；Cohen 类；Wigner 分布；模式分布；记谱

中图法分类号：O423　　**文献标识码**：A　　**文章编号**：1000-3630(2008)-04-0543-04

Time-frequency analysis of musical signal

XIE Xiu-qin，LIU Ruo-lun

(*Electronic System Engineering Department*, *Shandong University at Weihai*, *Weihai* 264209, *Shandong*, *China*)

Abstract: The major time-frequency analysis methods that have been applied to music processing are traced and the application areas described. Techniques are examined in the context of Cohen´ s class, facilitating comparison and the design of new approaches. Analyses spanning pitch synchronous analysis, short-time Fourier transform (STFT), phase vocoder, constant-Q and wavelet transforms, the Wigner distribution, and the modal distribution are all covered. The limitations of Windowing methods and their reliance on steady-state assumptions and infinite duration sinusoids to define frequency and amplitude are detailed. The Wigner distribution in contrast, uses the analytic signal to define instantaneous frequency and power parameters, and the modal distribution is a linear transformation of Wigner distribution optimized. Application areas consider analysis, resynthesis, transcription, classification and recognition, visualization.

Key words: time-frequency analysis; cohen's class; wigner distribution; modal distribution; transcription

1 引　言

长期以来信号处理的对象局限于确定性信号或是统计量不随时间变化的平稳信号，其有效的分析工具就是傅立叶数学，它是一种全局性的变换，无法表达信号的时频局部特性。但对非平稳信号的分析而言，过去人们一直是沿用平稳信号的处理方法来做近似，效果当然不够理想。随着研究的深入和科技实践的需要，针对非平稳信号的时频(TF)分析理论就孕育而生了。TF 分析实际上是将一维时间信号映射到 TF(有的是时间尺度)二维，很好地表示出信号的频率成分随时间的变化规律。在音乐声学中，傅立叶数学只能充分地表示能听见的单频信号的音高，不能简洁地表示音乐信号的感知现象，从而促进了 TF 表示理论在音乐信号分析中的发展。

这样一种理论在音乐信号处理上的应用和改进，反映了人们对量化音乐信号特性的兴趣，从而可以更好地理解有关音乐产生和欣赏的过程。事实上，许多数学物理学家已经创造了一些系统，从 Pyt-hagoras 建立音长、振动和音高的基本关系，到 Helmholtz、Rayleigh 等人发展的偏微分方程，音乐已经为当今的标准工程数学和物理学提供了一笔灵感财富，并开拓

收稿日期：2007-09-30；修回日期：2007-12-06
作者简介：谢秀琴(1985-)，女，内蒙古宁城人，硕士研究生，主要研究乐音信号的音色分离。
通讯作者：刘若伦，E-mail：ruolun.liu@sdu.edu.cn

了当前对TF分布的理解和需要。通过对技术、理论、教育和经济因素的汇合,研究人员一定能够在音乐所提供的基础上增加新的数学层次。

2 TF分布的发展

为表示音乐信号的时间和谱结构，将介绍一个通用的数学框架,使之能直接比较各种表示方式。

2.1 一般的数学形式

为了比较音乐信号的表示方法,所用的一般形式是信号能量在时间和频率上的联合分布。一个音乐信号是作为一个时间函数给出的，同时频率和能量也提供了有用的信息。许多分布的系数都可以通过计算平方幅度转换成能量函数,如傅立叶级数(FS)、傅立叶变换(FT)和恒Q变换。其它一些分布的系数则有能量单元,如Wigner分布(WD)和模式分布(MD)。因此,一个能量函数为各种分布的比较提供了公共依据。

对一个频率变量,可以考虑用多种方法加以分析。将频率变量与时间和能量结合定义一个基本框架,在这个框架下可以为音乐信号比较和设计合适的表示方式，后来在这个框架下发展了TF分布。Wigner和Ville都先后得到了TF分布，且两者在形式上很相似。一个信号$s(t)$的Wigner-Ville分布(WVD)如下所示：

$$W_s(t,\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}s(t+\frac{\tau}{2})s^*(t-\frac{\tau}{2})\exp(-j\omega\tau)d\tau \quad (1)$$

Cohen经过进一步的研究,将WVD进行线性变换后得到了Cohen类。这种线性关系的一种表达式为：

$$C_S(t,\omega,\varphi)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}W_S(\tau,\xi)\varphi(t-\tau,\omega-\xi;t,\omega)d\tau d\xi \quad (2)$$

其中$C(t,\omega;\varphi)$是Cohen类的一个成员,$\varphi(\tau,\xi;t,\omega)$叫做分布的核。选择Cohen类作为各种TF分析的框架，是因为它为音乐信号处理中已存在分布的定量分析和一些具有理想特性的新分布的设计提供了基础。

仿射类是时间和尺度的能量函数,其中恒Q变换和加窗的指数小波变换是少数展示了尺度特性的音乐信号分析技术,也可以将其认为是频率的函数。

2.2 傅立叶级数和傅立叶变换

虽然Helmholtz知道乐音是随时间变化的,但他还是认为从一个感性观点出发,音乐中最突出的部分几乎是周期的,并且把它叫做稳定状态[1]。Helmholtz认为是信号中谐波分量的幅度决定了响度、音高和音色,并且这些特性也是单个乐音的主要感知属性。当Helmholtz分析技术直接估计分量的幅度和频率时，此时最接近的数学技术是FS。将FS功率谱密度放在Cohen类框架中,它只是一个频率的函数,与$P_s(\omega)$很相似。对一个分布的时间变量求积分,得到了频率边缘分布：

$$P_s(\omega)=\int_{-\infty}^{\infty}C_s(t,\omega;\varphi)dt \quad (3)$$

如果一个分布的频率边缘分布与功率谱密度对应,那么就说这个分布满足频率边缘分布。

虽然FT比FS提供了更多频率上的信息和更长的时间间隔,但它还是缺少关于任何额外谱内容的时间信息。由于FS和FT都缺少关于信号能量在时间分布上的信息,所以提出了瞬时功率的概念。瞬时功率更直接地对应于一个二维Cohen类分布的一维时间边缘分布,通过对频率变量进行积分可以得到：

$$P_s(t)=\int_{-\infty}^{\infty}C_s(t,\omega;\varphi)d\omega \quad (4)$$

如果该函数等于瞬时功率,就说明一个分布满足它的时间边缘分布。不过通常用于音乐信号分析的大多数分布都不满足频率和时间边缘分布，但由于WD在Cohen类中的特殊地位,这两个条件它都满足。

3 音乐声学中TF方法

第2节描述的分析观点给出了关于单一时间或频率的音乐信号能量分布，但是为了提供一个完整的描述还需要这两个变量的一个联合函数。

3.1 时变FS和FT扩展

FS和FT是通过将信号分解成具有恒定频率和振幅的正弦曲线和的形式来分析信号。时变修正假设，信号的频率和振幅在有限的时间间隔上近似为常量,然后在频谱上对每个间隔进行分析。

扩展静态频率分析方法包含FS技术的时变扩展。Mathews和他的同事发展了语音的基音同步分析技术,该技术应用了一个连续的FS分析,它的周期随着基频估计的变化而改变。Luce用一个类似于Mathews的连续FS分析方法对14个管弦乐器进行了分析,Risset和Mathews将这种技术应用到小号的分析和再合成上，他们观测到在击打期间频率上的一些移动，以及在高频分量的起始阶段的一个延迟。这些研究都证实了时变信号内容和显著的击打在乐器识别和音色中的重要性。

对于非周期的时变信号，刚刚所描述的基音同步分析技术存在泄漏问题。Beauchamp对基音同步分析进行了一次扩展，使其能部分地处理以上未能解决的泄漏问题。外差滤波器是在Beauchamp方法上的一个变形，在进行基音同步分析时用了一个

单周期的矩形窗，并在输出部分增加了后延滤波。

另一项扩展的静态谱分析技术是 STFT 和相位声码器。STFT 和相位声码器可以用较长的、较窄带宽的窗减小泄漏，但是要求持续时间相对于听觉时间变化要长一些。不过较长的窗在持续时间上对击打进行了平均，而击打对于人类听觉的时间分辨率和信号的变化率非常重要，这就在估计幅度和频率时产生了偏差。

为了对语音信号进行编码，Flanagan 和Golden 在贝尔实验室首先开发了相位声码器。Dolson 开发了一个跟踪一个音乐音调基频的形式，还分析了用声码器计算的参数之间的关系，该声码器有解析信号的瞬时幅度和频率。这是一个重要的对比，因为解析信号给出了有关时变频率和振幅的良好定义。

STFT 的一个变形是恒 Q 分布。Brown 将恒 Q 变换应用到音乐上，是受到这样一个事实的启发：恒 Q 变换所提供的呈几何分布的频率分解反映了音阶的结构。计算恒 Q 变换的公式为：

$$S^{CQ}(t,\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}s(\xi)h((t-\xi)\frac{\omega}{\omega_0})\exp(-\mathrm{j}\omega\xi)\mathrm{d}\xi \quad (5)$$

方程(5)也是一类小波变换。Daubechies 的研究为小波变换提供了许多理论基础，他是通过将信号分解成一个母小波的扩展和平移。对于一个特殊的频率 ω_m，如果将公式(5)中的窗和复指数联合起来，就可以得到一个对音乐信号分析有用的小波：

$$w(\frac{t}{a})=h(\frac{t}{a})\exp(-\mathrm{j}\frac{t}{a}\omega_m) \quad (6)$$

与 STFT 不同，小波变换能较好地解决时间和频率分辨力的矛盾[2]：小波变换的窗是可调时频窗，在高频时用短窗，在低频时则用宽窗，即以不同的尺度观察信号，以不同的分辨力分析信号，充分体现了多分辨率分析的思想，这与时变、非平稳信号的特性一致。

3.2　TF 分布的推广

以上描述的 TF 分布都是使时间独立的变换适合于时变信号，其具体实现是用滑动窗并假设信号在窗长范围内是稳定的，将信号分解成无限持续的、时变的正弦曲线。但这些正弦曲线并没有反应一个时变信号的真实特性。Gabor 提出了一个瞬时频率的概念，以及信号的时变振幅或包络，他提议用 Hilbert 变换将一个任意信号分解成时变包络和频率。

WD 有几个特性都涉及到了 Gabor 的解析信号概念。因为 WD 满足它的边缘分布，所以将公式(4)用于 WD 中来表示解析信号的瞬时功率和复包络的平方模，并且 WD 的平均频率等于解析信号的瞬时频率。虽然 WD 有许多可取的特性，但是在音乐信号分析上也存在问题。对 Cohen 类中的所有成员来说，WD 的双线性特征反映了能量计算的非线性特性，这些计算在不同的信号分量之间产生了叉积，叉积的出现极大地干扰了 TF 分布，同时也抑制了二次型 TF 分布的推广。但是，这些叉积可以用时间窗核来消除。

对于音乐信号，Cohen 类的灵活性允许设计一个模式核，该核可以在 STFT 和 WD 之间产生分布，既减少了 STFT 的平滑偏差，又在 WD 中没有产生叉积。MD 关注的是，估计一个具有最小偏差的多分量音乐信号模型的瞬时频率和振幅。对于 Cohen 类，模式核由下式定义[1]：

$$\varphi_M(\tau,\xi)=h_{LP}(\tau)G_{LP}(\xi) \quad (7)$$

其中，$h_{LP}(\tau)$和 $G_{LP}(\xi)$是低通滤波函数，它们的截止频率与信号特性有直接的关系。模式核是一个可分离的核，因此 MD 拥有所有其它方法的优点，如不会引起分量间的混叠或仅在谐波频率处采样的问题，没有额外的时间-平均代价，也没有交叉项干扰。

4　TF 分析在音乐处理问题上的应用

音乐处理与语音处理有许多相同的应用，如分析、合成和压缩，但与其它应用方面相比，这些应用方面的发展较少。接下来关注几个其它的应用方面。

4.1　分析和再合成

作曲家及研究者 J.-C. Risset 对小号音调进行了一次广泛的分析[1]。Risset 的研究表明，小号的频谱几乎是谐波的，与低谐波相比，高谐波的振幅包络有一缓慢的增长时间和快速的衰退时间，振幅和频率中的小的伪随机偏差表现出了声音击打部分的特点。通过软件合成，Risset 证实了他的分析。再合成过程不仅证实了在他的研究中发现的性质，还证实了这样一种观点：一个时变频谱的特定性质在音色感知中起到了关键的作用。

对于 TF 分析，已经提出了一项用于测量乐调重音程度的技术，该技术对于不同的声学材料是鲁棒的[3]。同时，还有许多理论合成方法，如调频(FM)、整形、附加合成、物理建模和共振峰合成等。

4.2　记　谱

在过去 20 年，人们在记谱上取得的成就日益增加。音乐记谱情况的复杂性在于，它与当前对语音处理的要求不同，演奏的细微差别与正在演奏的乐谱一样重要。而造成这种复杂性的原因是：信号是由许多声音或音色组成的，而不是一个，并且制作唱片通

常是在相当大的多通道失真情况下进行。

一个最基本的自动音乐记谱系统是用音乐信号作为输入,并确定了基音。几种解决该问题的方法已经得到了成功的单音色记谱系统。在 Moorer 方法中,在特定的限制条件下已经对两个具有相似音色和有限时间变化的声音记谱。同时,还有对两个乐器记谱的方法。但是目前,多音音乐的自动记谱一直是个难题,主要是因为不存在适合乐音中的宽范围特征检测的单个尺度的 TF 表示。但是,多分辨率傅立叶变换(MFT)可以解决该问题[4],它是利用一组尺度来提供几个 TF 表示,该方法对于比较复杂的音乐表演可以得到很好的记谱结果。多音音乐记谱中的另一个问题是八度音检测问题。由于缺少有关音乐中每个乐器音色的信息,用于多音音乐的音高检测器很难检测八度音。在文献[5]中介绍了一个解决该问题的音乐记谱系统。

一个比较完善的记谱系统不仅能确定由多音色源演奏的乐谱,而且还能记谱音乐家在他们的表演中的微妙(以及不太微妙的)表现。正如 Brown 所讨论的,表演的这些方面以及它的记谱不能很容易地用标准的傅立叶技术证明,而需要一批更强大的 TF 工具。

4.3 分类和识别

在多媒体技术中,越来越需要对音频信号进行有效的分类,使从大型数据库中进行基于内容的检索过程更加准确和容易。该任务的重点在于如何从信号中提取适合于分类的准确特征。音频信号本质上是非稳定的,因此 TF 方法是适合对其进行分析的方法。

有一种方法[6]利用自适应 TF 分解算法来分解信号,再用基于八度音阶的信号分解参数产生特征。自适应 TF 算法是处理音频信号的非平稳行为中一种有效的方法,在基于方法的分解中,信号近似为小的 TF 函数。Mark[7]提出了一种系统,该系统通过在一个特定歌曲中识别类似于合唱或重复的特性,来搜索结构的冗余。为执行该基于结构的模式识别而分离这些有用的特征类,引入了一种色度图的方法,这是在传统 TF 分布上的一个变形,该方法寻找表示循环的音高感知特性,即色度。文献[8]给出一个新的音乐类型识别实时算法。该算法依照音乐特性,利用小波变换和 DTFT 在音乐音高频率上分析频谱,这不仅增加了计算效率,而且还抵制了噪音。以上介绍的系统对音乐数据库的浏览,多媒体搜索都很有用。

4.4 可视化

对音乐处理领域来说,一个相对独特的新兴应用正在出现,就是我们所谓的可视化。其通常与语音通信上的视觉训练设备有很多的相似特征。可视化就是将一个听觉信号转换成可视化形式,比起其它方法能更好地突出信号的理想特点。在文献[9]中,介绍了一种利用颜色和三维空间显示音乐的方法。为了包含视觉和听知觉而扩展反馈机制,可能会从根本上改变音乐表演的教学方式,得到更好的学习率,提高实践中的效率,而且能更准确地控制音乐的制作过程。

5 结 语

本文总结并概括了时频分析在音乐信号上的发展及应用,可以帮助对此方向感兴趣的读者获取相关信息,了解其发展状况。

参 考 文 献

[1] William J. Pielemeier, Gregory H. Wakefield, and Mary H. Simoni. Time-frequency analysis of musical signals[J]. Proceedings of the IEEE, 1996, **84**(9): 1216-1230.

[2] Peter De Gersem, Bart De Moor, and Marc Moonen. Applications of the continuous wavelet transform in the processing of musical[J]. Proc. of the 13th International Conference on Digital Signal Processing (DSP97), 1997, 563-566.

[3] Anssi P. Klapuri, Antti J. Eronen, and Jaakko T. Astola. Analysis of the meter of acoustic musical signals[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2006, **14**(1): 342-355.

[4] Keren R. Zeevi Y Y, Chazan D. Multiresolution timefrequency analysis of polyphonic music[J]. IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis (TFTS'98). 1998, 565-568.

[5] Chien Yu-Ren, Jeng Shyh-Kang. An automatic transcription system with octave detection[J]. IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP), 2002, 1865-1868.

[6] Karthikeyan Umapathy, Sridhar Krishnan, Shihab Jimaa. Multigroup classification of audio signals using time-frequency parameters[J]. IEEE Transactions on Multimedia. 2005, **7**(2): 308-315.

[7] Mark A. Bartsch, Gregory H. Wakefield. Audio thumbnailing of popular music using chroma-based representations[J]. IEEE Transactions on Multimedia. 2005, **7**(1): 96-104.

[8] CAO Zheng, GUAN Shengxiao, WANG Zengfu. A real-time algorithm for music recognition based on wavelet transform[J]. Proceedings of the 6th World Congress on Intelligent Control and Automation, 2006, 9926-9929.

[9] Sean M. Smith, Glen N. Williams. A visualization of music [J]. Proceedings of the 8th IEEE Visualization '97 Conference, 1997, 499-503.