

# 乘法模型下西洋乐器音色特征

刘若伦, 张家琦

(山东大学威海分校信息工程学院, 山东威海 264209)

**摘要:** 基于乐音的乘法模型分析研究了可用于西洋乐器音色识别的声学特征: 谐波能量比、倒谱与 Mel 倒谱系数 (MFCC) 和该文提出的机制声 MFCC。对西洋管弦乐器的单音进行了特征提取, 通过分析乐器特征的模式, 评价它们与音色的关联程度。音色由激励源和共振体共同决定, 与共振体特性相比, 激励源的特征更能反映音色的不同。研究表明: 强调激励源本质的机制声 MFCC 和乐音 MFCC 能更好地体现音色之间的不同; 机制声 MFCC 尤其适合对乐器大类的识别; 谐波能量比仅反映共振体特性, 可以在一定程度上反映音色的差别; 复倒谱与实倒谱区别不同乐器的能力有限。

**关键词:** 发声机制; 音色特征; 西洋乐器

**中图分类号:** O423

**文献标识码:** A

**文章编号:** 1000-3630(2009)-03-0269-07

**DOI 编码:** 10.3969/j.issn1000-3630.2009.03.016

## Timbre features of western instrument under multiplication model

LIU Ruo-lun, ZHANG Jia-qi

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

**Abstract:** The acoustic features related to western instrument timbre, including the distribution of harmonic energy, cepstrum and Mel-frequency cepstral coefficients (MFCC), and the mechanism sound MFCC (MSMFCC) which is proposed in this paper, are investigated under the multiplication model of musical sound. Features are extracted from pure tone of western orchestral instruments, and their influences on timbre are assessed through analyzing each feature pattern. Results show that: MFCC and MSMFCC emphasizing the nature of excitement perception can reflect timbre difference very well; MSMFCC is more capable for instrument classification; The distribution of harmonic energy, which reflects only the properties of resonator, can differ timbre to some degree; Complex and real cepstra have limited ability to distinguish the different instruments.

**Key words:** sounding mechanism; timbre feature; western instrument

## 1 引言

乐音是声音中的特殊一类, 与人的语音相比, 它具有明显的谐波特性。乐音有四个基本特征: 音高、响度、感觉上的持续时间和音色。其中音色作为区分不同乐器的最重要的特征, 也有许多应用。比如: 乐器识别、歌曲中的歌声与音乐的分离等。目前, 对音色的认识程度远比不上前三个特征。音色没有一个与之相对应的物理参数, 它通常被定义为: 听觉上区别具有相同响度和音高的两个声音之所以不同的属性。这个定义显然不足以用来进行定量的科学分析, 所以本文的目的就是寻找与音色相关程度较高的物理特征。

## 2 乐器的发声机理

乐器基本由三个部分组成: 振动体、激发体、共振体。振动体是主要声源, 如弦、膜、板等, 它与激发体共同决定基音频率和基本音色。激发体用以激发主要声源, 它决定激发的强度和频率。共振体则给乐器的音色以具体的感知特征<sup>[1]</sup>。以往的研究大多参照语音信号的激励源-滤波器模型: 激励源可看作是振动体与激发体共同作用的结果, 共振体则相当于语音信号中的声道滤波器。不过有人指出, 乐器发声模型的滤波器并不一定是线性的<sup>[2,3]</sup>, 所以基于激励源-滤波器模型的分析方法, 虽然能够得到一些令人鼓舞的结果, 但若要想揭示乐器发声机理, 还是有很大局限性的。而且我们的研究表明, 激励源与共振体的乘法模型更适合针对乐器发声机制的乐器音色分析。

在西洋乐器中, 铜管乐器的激发体可认为是口

收稿日期: 2008-04-04; 修回日期: 2008-08-03

作者简介: 刘若伦(1971-), 男, 河北昌黎人, 博士, 副教授, 研究方向为音频信号与信息处理。

通讯作者: 刘若伦, E-mail: ruolun.liu@sdu.edu.cn

腔内的气流,其振动体为演奏者的嘴唇。气流以驻波的形式在乐器管腔内共振,其发声机制与语音的发声机制最相似。在语音模型中,声门脉冲和声道间相互影响往往是可以忽略的,但演奏铜管乐时,激励源和共振体会相互影响。一方面共振体的谐振频率受演奏者嘴唇气压影响,另一方面管腔中气柱的振动会反作用于嘴唇,从而改变唇部肌肉的紧张度,进而影响到人们最终听到的乐音的基频。木管乐器的发声机制与铜管乐器类似,但振动体是薄膜或簧片。通过改变吹气的力度和侧面的开孔大小,改变管子内部气柱的谐振频率。弦乐器的激励源是琴弓与琴弦的摩擦引起的振动,共振体包括共鸣板和琴体空腔。弦乐器的激励源振动比较复杂,它与运弓的速度、琴弓施加的压力、摩擦点距琴马的距离密切相关。共振体的谐振频率通常具有较宽的范围,以放大琴弦上不同频率的振动。

乐音的时域变化主要可分为起始部分、稳定部分和衰减部分。Martin指出,起始部分和衰减部分的音色特征均不明显<sup>[3]</sup>,准确地讲,应该是共振体的音色特征不明显。倒谱以分离激励源和系统见长,所以可以根据乐音的稳定部分在一定程度上分析激励源对音色的贡献。但如果激发体瞬时状态在音色中起绝对作用时,如有音高的打击乐器,就一定要考察乐音的起始部分。

与音色有关的特征一般可划分为时域和频域的特征、稳态和暂态的特征。对这些特征的基本要求是与基音无关,好的特征应该在同一种乐器中表现出稳定的趋势,而在不同乐器中差别明显。在稳态特征中,许多人采用LPC倒谱<sup>[4]</sup>和具有非均匀频率刻度的WLPC<sup>[5]</sup>。其中WLPC与MFCC类似,采用了适合听觉分析的Bark刻度,有比较广泛的应用。另外,还有许多学者对暂态参数进行研究,如谱包络<sup>[6,7]</sup>和起始时间<sup>[8,9]</sup>等,都有一定的进展。

西洋乐器对铜管、弦乐和木管分类,主要考虑的是乐器的发声机制。根据前面对乐器发声机理的分析,不同乐器演奏相同音高时,谐波能量的不同分布主要取决于共振体。所以同一大类的两件乐器,其音色差别主要体现在共振体上;两大类乐器间的音色差别同时反应了激励源和共振体的不同,所以更加明显。

本文的分析对象是西洋乐器中的铜管、木管、弦乐的共振体和激励源的声学特征,以确定它们与音色之间的关联程度,故完全可以根据纯净单音信号的稳定部分来提取各种需要的音色特征。本文主要分析了谐波能量比、复倒谱与实倒谱、Mel倒谱、机制声Mel倒谱。试验选用的乐音样本来自于

McGill University Master Samples(MUMS) CD。

### 3 谐波能量比

乐音具有明显的谐波特性,能量大都集中在基频及低次谐波频率附近。1877年,Helmholtz<sup>[10]</sup>在对频谱的分析中,首次提出,不同谐波之间能量的对比分配是影响音色的一个重要特征。乐音的高次谐波越丰富、幅度越大,听起来就越明亮<sup>[3]</sup>。

令描述单音信号的乘法谐波模型为:

$$x_f(t) = v_f(t) \sum_{k=1}^K A_{k,f} \cos(k2\pi ft + \varphi_{k,f}) \quad (1)$$

其中, $x_f(t)$ 为某件乐器在基频为 $f$ 时的单音; $v_f(t)$ 为该基频下乐音时域波形的缓变包络,它主要取决于声源的发声机制; $A_{k,f}$ 是单音基频为 $f$ 的第 $k$ 个谐波幅度; $t$ 为乐音时间变量。那么谐波幅度矢量 $r_f$ 就可以表示某一音高单音的谐波能量比。

$$r_f = [A_{1,f}, A_{2,f}, \dots, A_{K,f}] \quad (2)$$

$r_f$ 可以通过对单音信号先进行傅里叶变换,然后检测各谐波幅度并进行归一化处理来计算。公式(3)代表某件乐器谐波能量比的平均值。其中, $N$ 代表该件乐器在音域范围内所能演奏的单音个数。

$$r = \frac{1}{N} \sum_{n=1}^N r_{f_n} \quad (3)$$

从乐器的物理结构、发声机制,以及听觉的生理特性考虑,对于任何一个单音,人耳可辨别的谐波(泛音)频率是有限的,那么基音越高,可感知的谐波就越少。所以如果一件乐器音域范围位于较低频段,则该乐器可感知的泛音就比较丰富。低音的谐波通常有8个以上,而高音的谐波一般至多4个。

图1是小号、单簧管、小提琴的归一化谐波能量的对比情况。每一条虚线代表一个单音,而粗实线则代表了该乐器在整个音域内谐波能量分布的平均趋势。三件乐器的结果集中反映出,一件乐器不同音高的谐波能量变化很大,能量分配与谐波次数的函数关系不明显。小号的谐波较丰富,谐波能量虽然也随谐波次数缓慢递减。单簧管高次谐波的能量分布趋势相对稳定,主要是由于木质管壁对高频成分的吸收。小提琴的发声机制与前两者完全不同,所以它们的谐波数量接近,但小提琴的谐波能量分布更分散,规律性不明显。

同一件乐器在不同音域范围上谐波能量比的差别,一方面是因为谐波个数的变化,另一方面也受到演奏方式的影响。当吹奏气流符合一定条件,管乐的基频可能被抑制,这时谐波幅度比例关系就会明显的偏离平均趋势,其他乐器也有类似现象<sup>[1]</sup>。

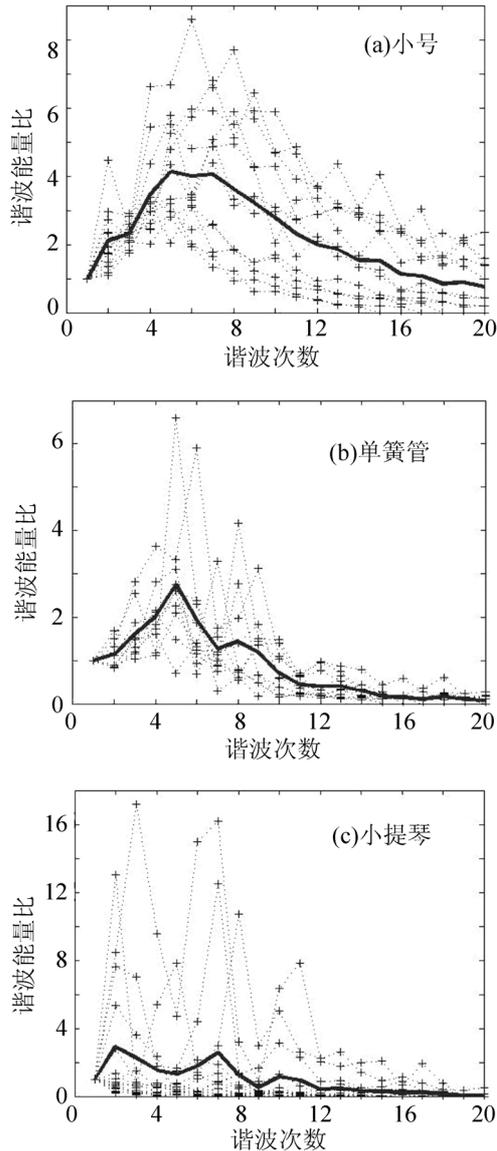


图1 归一化谐波能量比曲线。虚线对应某个单音;实线对应整个音域的平均值。

Fig.1 Normalized harmonics energy distribution. Dot line with cross is of single note; Solid line is of the average over all notes

### 4 倒谱与 Mel 倒谱

倒谱定义为信号频谱取对数后的傅立叶反变换,它是同态分析的有力工具,非常适于分析非线性信号,如乘性或卷积信号。两个信号在时域的卷积,对应于倒频域的相加。于是在时域和频域都无法分离的卷积信号,就可以在倒频域内进行分离。

在语音信号分析的“激励源-滤波器”模型中,激励源对应声带或喉的脉冲气流,滤波器对应声道的共振峰模型<sup>[11,12]</sup>。通常情况下,声道的共振峰相对于脉冲频谱变化缓慢,所以语音分析中可以根据低倒频分析共振峰,而根据高倒频研究脉冲激励源特性。但对于乐音信号,这种对应关系是值得怀疑的。

我们之所以能够在不同音高、不同声强下,区分不同乐器的音色,很重要的原因是乐器的发声机制不随基频和演奏方式变化,反映到公式(1)中,就是时域包络和谐波相位不随基频而变化。那么乐音信号就可以简化为:

$$x_f(t) = v(t) \sum_{k=1}^K A_{k,f} \cos(k2\pi ft + \phi_k) \tag{4}$$

若令式(4)中的求和部分为  $s_f(t)$ ,用  $m$  表示倒频,则按照倒谱定义,乐音信号的复倒谱为:

$$\tilde{x}_f(m) = \mathbb{F}^{-1} \{ \ln [V(\omega) * S_f(\omega)] \} \tag{5}$$

其中,  $V(\omega)$  是包络信号的频谱,  $S_f(\omega)$  是基频为  $f$  的谐波信号频谱,  $\mathbb{F}^{-1}$  代表傅立叶反变换,“\*”代表卷积。实倒谱则忽略相位信息,只对幅度谱取对数。

由于包络信号相对于谐波信号频率非常低,所以可以安全地假设,  $V(\omega)$  的频谱宽度远远小于谐波基频。根据对数及傅里叶变换的性质:

$$\tilde{x}_f(m) = 2B_f \delta(m) + 2\tilde{v}(m) \sum_{k=1}^K \cos(k2\pi fm) \tag{6}$$

其中  $B_f$  为常数,  $\tilde{v}(m)$  为包络信号的倒谱。式(6)的推导过程详见附录。因为反映声源发声机制的  $v(t)$  为缓变信号,所以  $\tilde{v}(m)$  在倒频域具有较宽的倒频带宽。在分析语音的声道共振峰特性时,通常取倒频  $m < 25$  的部分就足够了<sup>[10]</sup>。考虑到乐器共振体与人的声道之间在物理结构和声学特性上的差异,分析乐器发声机制时,考察了倒频  $m < 30$  的较宽范围。

Mel 倒谱系数采用了不均匀频率刻度,以适合听觉特性,因此更适于分析语音和乐音<sup>[13]</sup>。在人耳分辨各种音色的过程中,耳蜗起了非常重要的作用。耳蜗实际上相当于一个滤波器组,它的中心频率在 1000Hz 以下近似为线性尺度,而在 1000Hz 以上为对数尺度,所以人耳对低频信号比高频信号更为敏感。

Mel 倒谱与普通复倒谱最重要的区别就是,先用 Mel 滤波器组对信号进行滤波,实现频率轴的不均匀划分,继而对每个 Mel 滤波器输出的对数功率做离散余弦变换(DCT)而不是傅立叶反变换。即:

$$C_{mel}(m) = \sum_{k=1}^{R-1} P_{mel}(k) \cos \left[ \frac{\pi}{R} m \left( k + \frac{1}{2} \right) \right] \tag{7}$$

其中  $R$  为滤波器个数<sup>[14]</sup>,  $P_{mel}(k)$  为第  $k$  个 Mel 滤波器输出的对数功率:

$$P_{mel}(k) = \ln \left[ \sum_{\omega=\omega_l^k}^{\omega_h^k} |M_k(\omega) X(\omega)|^2 \right] \tag{8}$$

$\omega_l^k$  和  $\omega_h^k$  分别是第  $k$  个 Mel 滤波器频率响应  $M_k(\omega)$  的频率上下限。  $P_{mel}(k)$  还可以被看成时域信号

的对数 Mel 功率谱, 相当于对经过 Mel 滤波器组的时间信号按 Mel 频率  $k$  进行傅立叶变换再取对数。参考激励源-滤波器模型考虑乐音信号, 共振体的频率响应比激励源的频谱变化缓慢, 则 MFCC 的低倒频对应共振体, 高倒频对应激励源。

Mel 滤波器组是一组带通滤波器, 其中心频率按照 Mel 频率刻度是均匀排列的, 但不同中心频率对应的滤波器带宽并不相同。MFCC 最初是用于语音分析的, 后来 Beth Logan 验证了 MFCC 用于乐音信号分析的可行性<sup>[15]</sup>, 认为它至少不会是有害的, 因此本文专门验证了 MFCC 与音色的关联程度。

传统的 Mel 滤波器组采用的是三角窗, 在本文的实验中, 采用了更接近耳蜗特性的 Gammatone 滤波器。由于乐音信号的频率大都在 18kHz 以下, 所以本文采用了 0-18kHz 的由 25 个滤波器组成的 Mel 滤波器组, 得到 24 项 Mel 倒谱系数。Gammatone 滤波器采用了 Malcolm Slaney 的 Auditory Toolbox 中的函数。在对三类乐器的 MFCC 分析中, MFCC 的确显示出胜过谐波能量比和普通倒谱的优势。MFCC 不仅与基音无关, 而且在不同乐器之间表现了不同的趋势。

试验结果表明, 同一件乐器不同音高的倒谱差别较大, 所以稳定性和区分度都不好。图 2(a)为小号的复倒谱, 不同音高的复倒谱没有一个统一的表现, 有些是先升后降, 有些则是先降后升。而且大部分乐器的复倒谱稳定性都很差。但对个别乐器的某些特殊情况会有所不同, 如图 2(b)所示小提琴顿奏的复倒谱, 规律性明显优于小号的复倒谱。

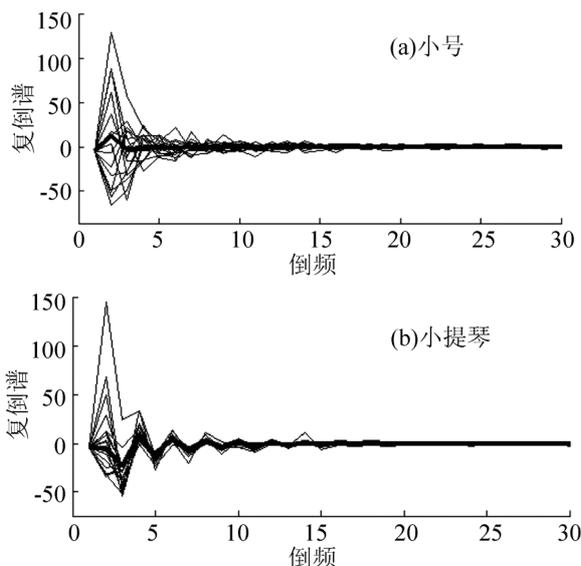


图2 小号(a)与小提琴顿奏(b)的复倒谱, 细实线: 一个单音; 粗实线: 全音域均值  
Fig.2 Complex cepstrum of trumpet(a) and violin(b). Thin line is of single note; thick line is of the average over all

对比分析了实倒谱。实倒谱的稳定性较强, 同一件乐器不同音高的实倒谱有明显的相似趋势, 但对不同乐器只是略有差别。图 3 所示的是小号和小提琴的实倒谱。一件乐器的实倒谱非常稳定, 但二者的差别却很小。此外, 由于实倒谱本身幅度小、衰减快, 区分度明显变差。即使用复倒谱容易区分的小提琴顿奏, 它的实倒谱与其它乐器区别也很不明显。

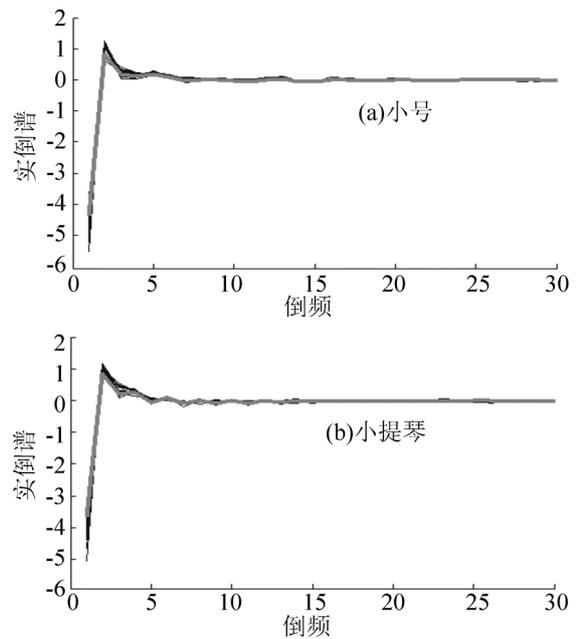


图3 小号(a)与小提琴顿奏(b)的实倒谱, 细实线: 一个单音; 粗实线: 全音域均值  
Fig.3 Real cepstrum of trumpet(a) and violin(b). Thin line is of single note; thick line is of the average over all notes

从综合倒谱的稳定度和对乐器的区分度来看, 它并不是一个很好的特征。其一是人耳对频率分辨的非线性: 人耳对低频成分的分辨能力要远大于对高频的分辨能力, 而复倒谱采用线性的频率刻度, 相当于强化了高频成分。其二是乐器共振体对不同频率成分的相频特性比较复杂, 因此考虑了相位的复倒谱规律性不明显, 所以在下面的 Mel 倒谱分析中没有考虑相位。

图 4(a)是圆号的 MFCC, 它反映出了明显的一致趋势: 先大幅下降到第 4 点然后上升, 从第 7 点开始保持稳定, 到第 11 点开始波动。双簧管(b)和小提琴(c)也有类似的结果。图 4 还表明, 不同音高的 MFCC 不仅具有相同的趋势, 在幅度分布上的方差也很小, 基本维持在均值附近的很小范围内。总的来说, 与前述特征相比, MFCC 的稳定性较好, 明显与基音无关。

图 4 也显示了 MFCC 较好的区分度, 由于三种乐器的共振体——圆号的螺旋铜管、双簧管的直木管、

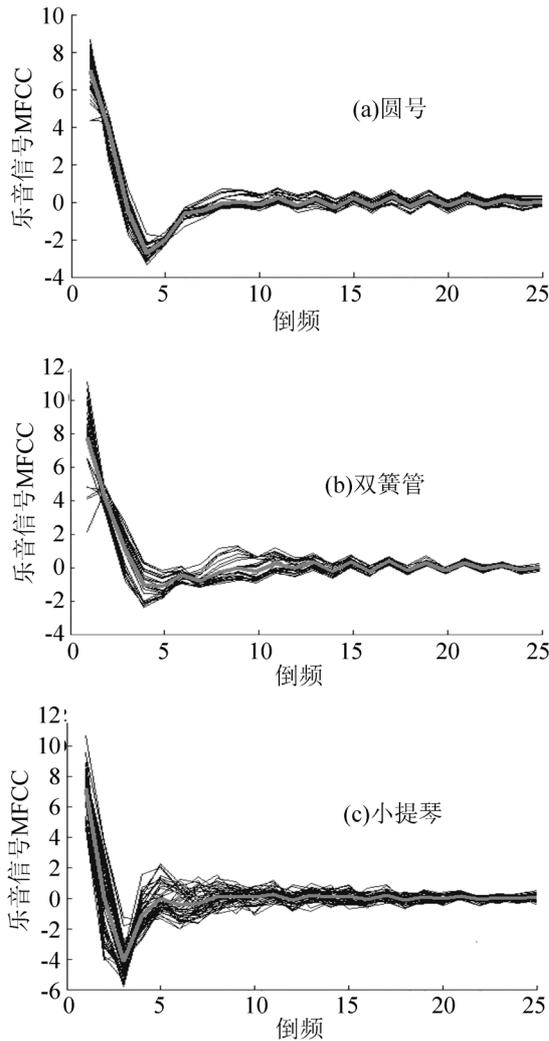


图 4 三种乐器在不同音高下的 Mel 倒谱系数, 细实线: 一个单音; 粗实线: 全音域均值  
 Fig.4 MFCC over all pitches of French horn(a), oboe(b), and violin(c)  
 Thin line is of single note; thick line is of the average over all notes

小提琴的木质琴箱相差迥异, 故而在低倒频系数上差别较大; 激励源方面, 唇的振动和具有指肚阻尼的弓弦摩擦振动, 有一定相似之处。但它们和双层簧片的振动差别非常大, 所以圆号和小提琴的高倒频系数的分布接近, 但双簧管的高倒频系数分布则完全不同。

限于篇幅, 这里只在每类乐器中挑选一件进行对比。其他乐器的结果与此类似, 所以 MFCC 在不同类别乐器的识别上有一定价值。从图 5 所示三类乐器的 MFCC 均值比较可以看出, 不同类的 MFCC 均值差别不太明显, 因此识别时不能单纯采用类均值来比较, 否则可能还不及对单个乐器的识别。但综合前述分析, 它仍是表现最好的特征, 不仅各个乐器之间的差别比较明显, 在一件乐器中的稳定度也很好, 而且如果综合分析低倒频和高倒频系数的分布规律, 大多数乐器都可通过 MFCC 得以区分。

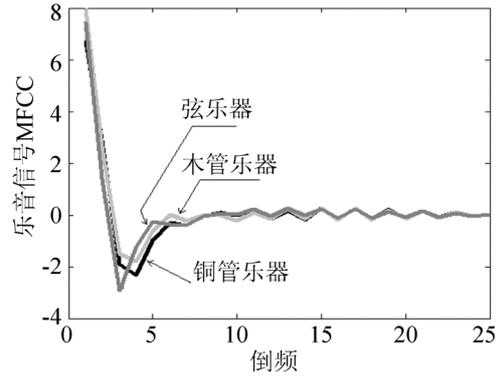


图 5 三类西洋乐器的 MFCC 均值  
 Fig.5 MFCCs averaged over three types of western instrument

也有 MFCC 区分起来比较困难的乐器, 比如木管类乐器不仅音色相近<sup>[2]</sup>, MFCC 的分布方差也较大, 这样两件乐器的 MFCC 表现可能会相差很大, 但仍难以区分。另外, 实验中选取了 0~18kHz 的 25 个滤波器, 分析的效果还不错。识别中还可以进一步检验其他阶数和频率范围的滤波器组的区分效果, 以确定最佳阶数和频率范围。

### 5 机制声 Mel 倒谱

前文所研究的特征主要还是侧重于共振体的。谐波能量比的变化以及复倒谱和 MFCC 低倒频部分的差别, 都取决于共振体的频率响应。传统倒谱的高倒频部分虽然能够反映激励源特征, 但往往幅度小、衰减快, 而且不同乐器还会出现相似的分布。因此前面三个特征都没有很好地体现激励源的差别, 但激励源对音色也是具有重要影响的。

观察乐音频谱可以发现, 乐音并不能简化成一个纯粹的谐波模型。在每个谐波周围, 都有一些分布接近的频率成份, 是它们丰富了乐器的音色, 而它们正是来自乐器激励源的不同发声机制。

要对发声机制进行分析, 首先要消除共振体的影响。因此用点阻滤波器<sup>[16]</sup>对各谐波进行调整, 使其幅度相同。我们把经过点阻滤波的乐音信号称为机制声信号, 分析该信号的 MFCC, 更有助于揭示不同乐器的音色本质。

由于乐音频谱各谐波的位置并不是严格的整数倍关系, 谐波的频谱宽度也不同, 因此在构造点阻滤波器时, 首先要检测各谐波位置以确定中心频率, 再以乐音频谱的等效矩形带宽幅度(频谱积分除以频率范围)衰减 3dB 的带宽为点阻带宽。点阻滤波器对各次谐波的抑制程度由各谐波幅度与等效矩形带宽幅度的差距来决定。如此, 各谐波幅度均被调整到等效幅度的水平。实验中, 对幅度低于等

效幅度的谐波暂不做调整。

机制声 MFCC 与乐音的 MFCC 相比, 虽有许多相似之处, 但由于抑制了共振体的影响, 又有些差别。图 6 为三种乐器在不同音高下的机制声 Mel 倒谱系数。将图 6 与图 4 对比可以发现: 机制声 MFCC 幅度明显小于乐音 MFCC 的幅度, 而且系数波动也不明显; 机制声 MFCC 仍与音高无关, 但它在不同音高下的分布更集中, 明显胜过谐波能量比和复倒谱。如果把机制声理解为冲激脉冲串通过一个反映发声机制的系统的話, 那么机制声 MFCC 的低倒频部分反映的就是该系统, 它可以更好地区分乐器大类发生机制的不同, 如吹、拉、弹、拨等。

和乐音 MFCC 一样, 同一类乐器的机制声 MFCC 也不尽相同。铜管乐器中, 各乐器机制声 MFCC 仅仅是有所差别。而木管乐器由于本身就有四个子类, 所以各乐器机制声 MFCC 则有很大差别。另外,

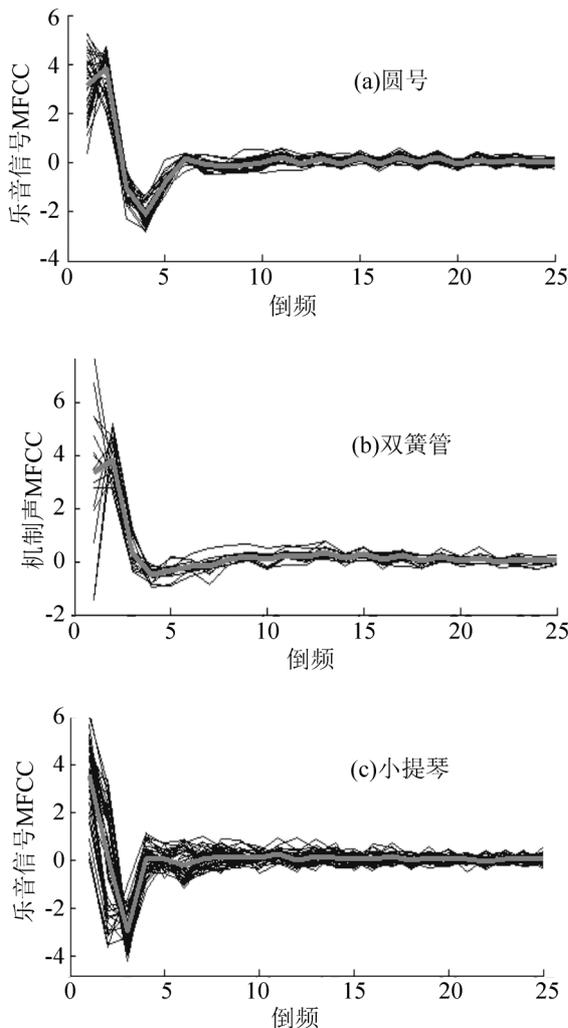


图 6 三种乐器在不同音高下的机制声 Mel 倒谱系数, 细实线: 一个单音; 粗实线: 全音域均值  
Fig.6 Mechanism sound MFCC over all pitches of French horn(a), oboe(b), and violin(c). Thin line is of single note; thick line is of the average over all notes

图 6 还反映出, 圆号和双簧管在不同音高下的机制声 MFCC 前两点, 有时上升有时下降, 趋势不稳定, 规律性也不明显。所以, 最低点(几乎总是  $m=3$ )之后的机制声 MFCC 的区分度更好, 可以考虑适当增加低倒频  $m \geq 3$  的识别权重。

图 7 是三类乐器的机制声 MFCC 均值。与图 5 比较不难发现, 机制声 MFCC 更利于区分不同类型的乐器。尤其是在最低点之后, 三种乐器的机制声 MFCC 曲线具有明显的差别, 而图 5 中三类乐器的 MFCC 几乎完全重合。总之, 机制声 MFCC 不但可以用于识别不同乐器, 而且稳定性及区分度都比较理想, 尤其是区分发声机制不同的乐器大类。

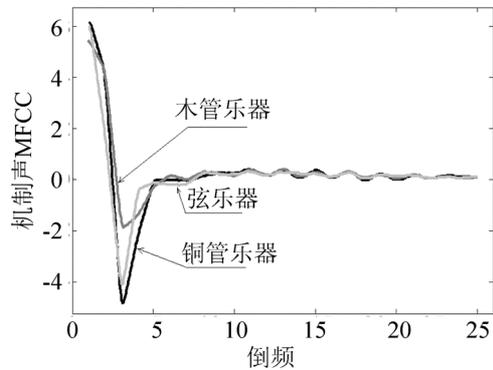


图 7 三类西洋乐器的机制声 MFCC 均值  
Fig.7 Mechanism sound MFCCs averaged over three types of western instrument.

### 5 结 论

乐器发声的全部过程, 都会影响人们对音色的感知。迄今为止, 音色到底与那些物理参数有关, 仍没有一个明确的答案。考虑到音色定义为使人在听觉上区分不同乐器的属性, 基于人耳听觉特性的特征必然更有利于音色的识别。本文详细分析了乐音信号的谐波能量比、倒谱、MFCC、机制声 MFCC 等特征对音色的影响程度, 得出以下结论:

- (1) 机制声 MFCC 是本文提出的新特征, 它对乐器识别的区分度和稳定性最好, 不论是各乐器之间, 还是各类之间, 该特征都有明显的区别。其中前几项系数由于反映发声机制的差别, 对区分不同乐器有重要作用。
- (2) 乐音的 MFCC 对音色的区分性能也比较好。稳定性虽然不及机制声 MFCC, 但也优于其他两个特征。识别中权重应仅次于机制声 MFCC。
- (3) 谐波能量比更适于管乐器的分析。对于同一件乐器, 谐波能量比在不同音高时方差较大, 并且受基音影响较大。谐波丰富程度以及谐波能量的变化趋势能体现共振峰的特点。虽然不同乐器的差

别明显,但在同类甚至同一件乐器没有稳定的表现,不能单独用于乐器的识别。

(4) 复倒谱是最不稳定的一个特征。仅对于特定乐器(如小提琴顿奏)有稳定的趋势,因此在识别这类乐器时,可以考虑适当加入复倒谱作为识别的特征之一。实倒谱虽然稳定性有所改进,但不同乐器的差别却不明显了。

以上特征有的侧重共振体,有的侧重激励源,只要采用合适的分析方法,都有比较好的表现。由于音色的复杂性,更进一步的深入分析需要对乐器的发声过程作更详尽的分析。乐器发声机制与机制声 MFCC 之间的内在联系也还有待进一步研究。此外,本文所研究特征对复杂的多乐件乐曲区分能力还有待进一步验证。乐音信号的相频特性在倒谱类音色特征中的作用,也是一个值得研究的问题。

参 考 文 献

[1] 马大猷,沈豪.声学手册[M].北京:科学出版社,2004.804-814.  
MA Dahyou, SHEN H. Acoustic manual [M]. Beijing: Science Press, 2004. 804-814.

[2] Eronen Antti. Master Thesis: Automatic Musical Instrument Recognition[D]. Department of Information Technology, Tampere University of Technology, 2001.

[3] Martin Keith D. Sound-source Recognition: A theory and computational model[D]. Department of Electrical Engineering and Computer Science, MIT, 1999.

[4] Luce D. Physical correlates of nonpercussive musical instrument tones[D]. MIT, 1963.

[5] J.E. Munoz-Exposito et al. Speech/Music discrimination using a single Warped LPC-based feature[A]. Proc. of ISMIR2005 [C]. London UK, 2005. 614-617.

[6] Rabiner L, Juang B. Fundamentals of Speech Recognition[M]. New York: Prentice-Hall. 1993.

[7] Kostek. Soft Computing in Acoustics: Applications of Neural Networks, Fuzzy Logic and Rough Sets to Musical Acoustics[M]. New York: Physica-Verlag. 1999.

[8] Poli Prandoni. Sonological Models for Timbre Characterization [J]. J. of New Music Research, 1997, 26: 170-197.

[9] McAdams, Beauchamp, Meneguzzi. Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters [J]. J. Acoust. Soc. Am., 1999, 105: 882-897.

[10] Helmholtz, Hermann L. F. On the Sensations of Tone[M]. NY USA: Dover Publications Incorporated, 1954.

[11] Oppenheim A V, Schaffer R W. Homomorphic Analysis of Speech[J]. IEEE Trans. Audio and electroacoustics, 1968, AU-16: 221-228.

[12] Quatieri T F 著, 赵胜辉, 刘家康, 谢湘译. 离散时间语音信号处理-原理与应用[M]. 北京: 电子工业出版社, 2004. 198-224.  
Quatieri T F. Discrete time speech signal processing: Principal and

Practice[M]. Prentice Hall, 2004. 198-224.

[13] 韩纪庆, 张磊, 郑铁然. 语音信号处理[M]. 北京: 清华大学出版社, 2004. 75-84.  
HAN J Q, ZHANG L, ZHENG T R. Speech Signal Processing [M]. Beijing: Tsinghua Univ. Press, 2004. 75-84.

[14] Davis Steve B, Mermelstein Paul. Comparison of Parametric Representation for Monosyllabic Word Recognition Continuously Spoken Sentences[J]. IEEE Trans. Acoustic, speech and signal processing, 1980, ASSP-28(4): 357-366.

[15] Beth Logan. Mel frequency Cepstral Coefficients for Musical Modeling[A]. In Proc. of ISMIR2000[C]. Plymouth, Massachusetts, USA, 2000.

[16] Wariar R, Eswaran C. A tracking digital notch filter[J]. IEEE International Symposium on Circuits and Systems, 1991, 5: 2810-2813.

附 录

设乐音信号  $x_f(t) = v(t) s_f(t)$ , 其中

$$s_f(t) = \sum_{k=1}^K a_{k,f} \cos(k\omega_0 t + \phi_k)$$

代表由共振体决定的谐波信号,  $v(t)$ 代表由发声机制决定的包络信号。于是有

$$X_f(\omega) = V(\omega) * S_f(\omega)$$

再假设  $V(\omega)$  的最高频率远远小于  $S_f(\omega)$  的基频, 由

$$S_f(\omega) = \frac{1}{2} \sum_{k=1}^K a_{k,f} [\delta(\omega - k\omega_0) e^{j\phi_k} + \delta(\omega + k\omega_0) e^{-j\phi_k}]$$

可得:

$$X_f(\omega) = \frac{1}{2} \sum_{k=1}^K a_{k,f} [V(\omega - k\omega_0) e^{j\phi_k} + V(\omega + k\omega_0) e^{-j\phi_k}]$$

上式求和各项频谱无重叠, 故可按频率分别取对数, 即:

$$\ln X_f(\omega) = \sum_{k=1}^K [b_{k,f} + \bar{b}_{k,f} + \ln V(\omega - k\omega_0) + \ln V(\omega + k\omega_0)]$$

其中:

$$b_{k,f} = \frac{1}{2} a_{k,f} e^{j\phi_k}$$

再根据倒谱定义可得:

$$\tilde{x}_f(m) = \delta(m) \sum_{k=1}^K 2 \operatorname{Re}(b_{k,f}) + \tilde{v}(m) \sum_{k=1}^K [e^{-jk\omega_0} + e^{jk\omega_0}]$$

若令:

$$c_f = \sum_{k=1}^K \operatorname{Re}(b_{k,f})$$

则:

$$\tilde{x}_f(m) = 2c_f \delta(m) + 2\tilde{v}(m) \sum_{k=1}^K \cos(k\omega_0 m)$$