复调音乐主旋律提取方法综述

张维维[1,2],陈 喆[1],殷福亮[1],张俊星[2]
(1. 大连理工大学信息与通信工程学院,辽宁大连 116024;2. 大连民族大学信息与通信工程学院,辽宁大连 116600)

摘 要: 主旋律提取在音乐检索、乐谱抄录、翻唱识别等领域具有广泛应用.复调音乐频谱结构复杂多样,音高变化范围广,因此复调音乐的主旋律提取较为困难.本文综述了复调音乐主旋律提取研究进展,对复调音乐主旋律提取方法进行分类,阐述了典型方法,介绍了主旋律提取的评价指标,给出了最新 MIREX 主旋律提取评测结果,说明了主旋律提取面临的主要挑战,并对主旋律提取技术发展方向进行了展望.

关键词: 主旋律提取; 复调音乐; 多声部音乐; 音乐信号处理; 音乐信息检索
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2017)04-1000-12
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2017.04.032

Review on Melody Extraction from Polyphonic Music

ZHANG Wei-wei[1,2], CHEN Zhe[1], YIN Fu-liang[1], ZHANG Jun-xing[2]
(1. *School of Information and Communication Engineering, Dalian University of Technology, Dalian* 116024, *Liaoning, China*;
2. *School of Information and Communication Engineering, Dalian Minzu University, Dalian* 116024, *Liaoning, China*)

Abstract: Melody extraction is to automatically extract the melody from a given piece of polyphonic music, and has been widely applied in music retrieval, score transcription, cover song identification and so on. Melody extraction from digital polyphonic music is reviewed in this paper. Firstly, the methods used for melody extraction are classified. Then, the representative methods are elaborated in detail, evaluation methodologies are presented, and the most recently MIREX melody extraction competition results are shown. Finally, the existing challenges are summarized, and the future research directions are provided.

Key words: melody extraction; polyphonic music; multi-part music; music signal processing; music information retrieval

1 引言

随着计算机网络和多媒体技术的发展,数字音乐的创作和传播不断加快.数字音乐产业的蓬勃发展及海量音乐数据的涌现,使得音乐的分类、检索、推荐、内容分析等技术越来越重要,并成为数字音频处理领域的研究热点.

主旋律提取是基于内容的音乐信号处理领域一项重要研究课题,它根据给定的一段音乐,由计算机自动分析音乐音频内容,提取出该段音乐的主旋律.通常情况下,未经专业训练的普通人也具有理解、处理复杂音乐信息的能力,如在聆听一段具有多个声源(可能包括管弦乐器、打击乐器、歌唱音等)的音乐后,多数人能顺利辨识歌唱音或某个主要乐器演奏的声音而忽略其它声源的影响,也能重复哼唱出该段乐曲的旋律,但由计算机进行主旋律提取却非常困难.

旋律(Melody)是一种基于人类听觉判断的音乐学概念,是音乐信号的基本特征,但并没有严格的定义.目前,音乐信号处理学术界普遍认同由 Poliner 等[1]给出的定义:主旋律是指听者根据一段复调音乐(Polyphonic Music)感知的,并被听者识别作为音乐本质的单音高序列.旋律是音乐的灵魂和基础,在音乐表现中具有重要意义[2].根据音乐学,按照音乐作品旋律线的数量,音乐分为单声部音乐和多声部音乐.多声部音乐是指音乐不单纯依靠单一旋律,还有以各种方式加入的其他织体,包括复调音乐(Polyphony)、主调音乐(Homopho-

收稿日期:2016-01-29;修回日期:2016-07-28;责任编辑:李勇锋
基金项目:国家自然科学基金(No.61172107,No.61172110);国家863高技术研究发展计划(No.2015AA016306);辽宁省教育厅科学研究一般项目(No.L2015135);中央高校基本科研业务费(No.DUT13LAB06,No.DC201502060404)

ny)和支声音乐(Heterophony)三种类型. 在音乐信号处理领域,常用"复调音乐"指代音乐学领域中的"多声部音乐"[3],即泛指音乐中含有两条或两条以上的旋律,允许有两个或者更多的音符同时发声,这些音符可以来源于不同的声源(如歌唱音、吉它等),也可以来源于能同时发两个以上音的声源(如钢琴). 如无其他特殊说明,本文所述"音乐"均指"复调音乐".

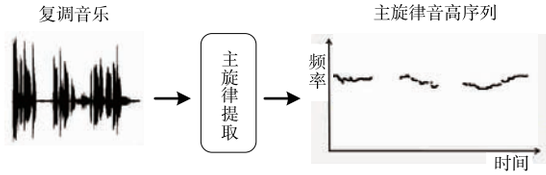

图1　复调音乐主旋律提取[3]

复调音乐主旋律提取就是生成与音乐片段的主旋律音高相对应的频率序列值,如图 1 所示,它在音乐结构分析、音乐检索、音乐风格分类、翻唱识别等领域具有广泛应用. 音乐结构分析是理解作品内涵的关键环节,通常先用主旋律提取方法来检测主旋律相似的区域,然后将相似区域归类,进而完成音乐结构分析[4]. 基于主旋律的音乐检索系统,首先提取所给定音乐片段的主旋律音高轮廓,然后采用序列模式匹配算法在歌曲库中查找与其相似的歌曲[5,6],与传统基于文本的音乐检索技术相比,该方法不需要人工标注,也不要求用户知道乐曲的准确信息,从而可实现智能检索,提高检索效率和准确性. 基于主旋律的音乐风格分类技术先从音频文件中提取出主旋律,然后从中获得颤音、拓扑形式等更高级语义特征,用于音乐风格分类,以管理大型个人与商业音乐库[7]. 此外,利用主旋律来度量两个曲目的相似性,能有效地进行翻唱识别,可用于版权保护、音乐组织与搜索等应用场景[8,9].

根据文件类型,音乐主旋律提取主要分为两类:MIDI 文件的旋律提取[10~14]和音频文件的旋律提取. MIDI 文件的乐曲相对真实音乐失真度较大,每个通道存放一种乐器的演奏信息,因此旋律提取比较简单,且准确率较高. 随着数字音频技术的发展,MIDI 音乐已逐步被音频文件类型的数字音乐所取代,为了阐述旋律提取的最新研究进展,本文不再对 MIDI 类型音乐的主旋律提取作进一步的归纳和总结.

音频文件的音乐听觉效果逼真,风格多样,应用普遍,但其记录的是各个声源的混合信号,因此音频文件的主旋律提取具有一定的挑战性. 2004 年,Goto 首次提出了针对音频文件的复调音乐主旋律提取方法,此后该课题得到了广泛关注,并成为近十年来音乐信息处理领域的研究热点. 目前,音乐信息处理领域所提到的主旋律提取,一般是指基于音频文件的提取方法[3].

本文综述十年来基于音频文件的复调音乐主旋律提取研究的最新进展与成果,详细阐述了基于显著性、基于源分离和基于机器学习这三类主旋律提取方法的基本原理、主要特点以及彼此间的区别与联系,给出了主旋律提取方法的性能指标和评价结果,指出了主旋律提取技术面临的挑战与未来发展方向.

2　主旋律提取方法的分类

根据音乐中是否含有歌唱音,主旋律提取方法分为声乐主旋律提取[15~26]和通用主旋律提取[27~33]两类. 如果复调音乐中包括歌唱音,则歌唱音的音高序列就认为是声乐主旋律;如果不存在歌唱音,则能量占主导地位的乐器演奏音的音高序列作为器乐主旋律[3]. 通用主旋律提取方法适用于声乐主旋律和器乐主旋律.

根据方法原理,主旋律提取主要分为基于显著性的方法[16,22,31,34~41]、基于源分离的方法[19,21,27,42]和基于机器学习的方法[43~45]. 本文第 3 节将详细阐述此三种类型的代表性方法.

根据处理的音乐类型,主旋律提取方法分为通用音乐旋律提取[22,31,38,46]和特定类型音乐旋律提取[15,47~50]两类,目前研究大多是通用音乐的主旋律提取.

根据方法的实时性,主旋律提取方法分为在线旋律提取[41]和离线旋律提取[22,27,42]两类.

3　主旋律提取主要方法

如前所述,根据各方法的基本思想,主旋律提取技术主要分为基于显著性的方法、基于源分离的方法和基于机器学习的方法. 这三类方法既有明显的区别,也存在一定的内在联系. 基于显著性的方法首先估计音乐片段中的多个音高,然后根据主旋律较伴奏音能量更显著的特点,选择兼有能量显著性和时间平滑性的音高序列作为旋律输出,即先进行多音高估计,然后进行主旋律跟踪. 该类方法在器乐主旋律和声乐主旋律提取中均有应用. 基于源分离的方法主要思想在于,从混合信号中分离出具有主旋律特征的音源分量,抑制伴奏音分量的干扰,然后根据单声部音高估计与跟踪算法输出主旋律音高序列,即先源分离,后音高估计与跟踪. 该类方法本质上是利用主旋律分量与伴奏音分量在表示域特征的差异来进行源分离,主要应用于声乐主旋律提取. 基于机器学习的方法,其主要思想是在训练集基础上,提取能够区别主旋律和伴奏音的特征,这些特征可以是人为指定的,也可以是算法自动学习产生的,然后借助模式分类算法将两者区分开,即先训练后测试,先提取特征后分类. 该类方法需要先验知识较少,但当测试集较小时易出现过拟合现象.

尽管三类方法在原理上有较大差异,但彼此间还存在一定的有机联系.(1)这三类方法都以能量显著性和时间平滑性为依据;(2)基于显著性的复调音乐主旋律提取方法常将各次谐波分量的幅度或能量加权和作为显著性量度,而忽略其他的谱分量,这与源分离的思想也有异曲同工之妙;(3)基于源分离的方法常利用某些特征(如抖音、各向异性等)辅助分离出主旋律分量,也有部分方法融合了聚类分析、统计建模等思想;(4)机器学习类方法所采用的能量、时间周期性、频谱谐波性等特征,在基于显著性的方法和基于源分离的方法中都有所应用.

基于显著性、基于源分离和基于机器学习这三类主旋律提取方法的比较如表 1 所示. 本节将对这三类方法的研究进展进行详细阐述.

表 1 三类主旋律提取方法比较

方法	影响性能的主要因素	参数	训练集	与律制关系	应用场合
基于显著性的方法	(1)显著函数构建 (2)旋律轮廓跟踪	人为设定	不需要	无关	(1)声乐主旋律 (2)器乐主旋律
基于源分离的方法	(1)音源分离 (2)单音高估计 (3)旋律轮廓跟踪	人为设定	不需要	无关	声乐主旋律
基于机器学习的方法	(1)训练集多样性 (2)训练集与测试集相似性	机器自动学习	需要	有关	(1)声乐主旋律 (2)器乐主旋律

3.1 基于显著性的主旋律提取方法

基于显著性的主旋律提取方法以主旋律的能量显著性和频率谐波性为主要依据,结合音高时间连续性约束,应用音频信号预处理、频谱分析、多音高表示、多音高跟踪和旋律活动检测等步骤进行主旋律提取,如图 2 所示[3]. 自 2004 年 Goto 率先提出主旋律提取问题后,基于显著性的主旋律提取方法研究取得重要进展,涌现出多种有效方法,如表 2 所示. 鉴于频谱分析、多音高表示、多音高跟踪是基于显著性的主旋律提取方法的核心内容,各方法在此三部分的差异较大,因此本节先分步阐述这三种关键技术,然后再对基于显著性的主旋律提取方法作总结归纳.

表 2 基于显著性的主旋律提取典型方法

序号	第一作者,年	预处理	频谱分析	多音高表示(显著函数)	多音高跟踪	旋律活动检测
1	Goto[35]2004	带通滤波器	多分辨率滤波器组 + 瞬时频率	音调模型显著性	多 agent	无
2	Paiva[38]2006	无	STFT + cochleagram + correlogram	谐波群检测	谐波群轨迹 + 音符检测	谐波能量
3	Ryynänen[40]2008	谱白化	STFT	谐波求和 + 谱能量	Viterbi	音高轮廓特征
4	Cancela[36]2008	无	常 Q 变换 + 倒谱 + 高通滤波	子谐波求和	轮廓跟踪 + 加权 + 平均旋律音高线	谐波能量
5	Rao[22]2010	无	高分辨率 STFT + 主瓣幅度匹配	双向失配算法	动态规划	音高轮廓特征
6	Joo[37]2010	无	STFT	谐波结构强度	基于规则的方法	无
7	Salamon[31]2012	等响度滤波器	STFT + 瞬时频率	谐波求和	音高连续性	音高轮廓特征
8	Dressler[51]2014	无	多分辨率 STFT + 瞬时频率	谱峰对分析 + 乐音模型	计算听觉场景分析 + 听觉流显著性	谐波能量
9	Ikemiya[52]2014	无	STFT + Robust PCA + 样条差值	子谐波求和	Viterbi	音高轮廓特征
10	Degani[53]2014	无	STFT + SMS(Sinusoidal Modelling Synthesis framework)	谐波频率偏差	无	无
11	Chien[54]2015	无	常 Q 变换	声学语音模型	音高连续性	谐波能量

3.1.1 频谱分析

主旋律音高估计误差要求在半个半音范围内. 为了提高低频段正弦分量频率的估计精度,同时又不牺牲时间分辨率,提出了多分辨率 STFT 方法[37,55,56]、瞬时频率法[31]、多分辨率 STFT 加瞬时频率法[35,51]以及谱峰插值法[22,52]等谱峰频率校正算法.

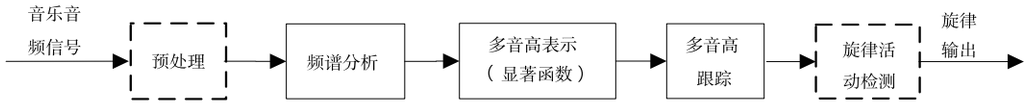

图2　基于显著性的主旋律提取框图

Joo 等[37]和 Park 等[56]根据幅度谱自相关系数选择数据帧长度,得到随音乐音频内容自适应变化的时间窗长度;而 Cancela 等[55]用时变 IIR 滤波器计算常 Q 变换来实现多分辨率分析,多分辨率 STFT 方法对旋律线动态变化具有良好的鲁棒性. 为了克服 STFT 频率精度的限制,Charpentier[57]对声码器算法[58]进行改进,提出了利用相邻帧相位谱计算瞬时频率的算法(简称为 Charpentier 算法),这两种瞬时频率法提高了正弦分量频率估计的精度[31,35]. Salamon 等[59]研究了固定分辨率 STFT、多分辨率 STFT、声码器等算法对音乐信号谱峰估计精度的影响,研究结果表明,对频谱进行频率和幅度校正能提高音高估计的精度. 为了在实时处理中获得适当的时-频分辨率,Goto[35]和 Dressler[51]均对信号进行多分辨率分析,然后再计算瞬时频率,其谱峰估计精度较瞬时频率法有所提高,但增加了计算复杂度. 此外,还有研究人员采用谱峰样条插值法[52]和主瓣匹配法[22]提高谱峰估计精度.

在上述频谱分析方法中,多分辨率 STFT 方法对谱估计准确率提升效果有限;瞬时频率法存在相位解码不唯一问题,故须与谱峰搜索相结合才能估计出较准确的正弦分量频率;多分辨率 STFT 与瞬时频率法相结合,可取得更好的效果,但计算复杂度有所增加;谱峰插值法和谱峰主瓣匹配法计算量较小,但不能提供额外的信息.

3.1.2　多音高表示

经预处理和频谱分析后,可获得每帧音频信号频谱中幅度相对较大的谱峰频率,这些谱峰表示主旋律基频及其各次谐波分量、乐音伴奏的基频及其各次谐波分量、打击乐器某些分量等. 多音高表示也称为音高估计,它是通过构建显著函数来量度各可能音高的显著性,然后按显著性从大到小排序,选出前 N 个最显著的可能音高值进行多音高跟踪处理. 泛音谐波性与能量显著性是多音高表示的主要理论依据. 多音高表示方法可分为基于谱峰频率的方法和基于谱峰能量的方法.

(1)基于谱峰频率的多音高表示方法

鉴于各次泛音分量的谐波性,Goto[35]提出了主-基频估计(Predominant-F0 Estimation,PreFEst)方法,该方法用若干个具有谐波结构的音调模型加权混合来表示观测概率密度函数,用最大后验概率密度方法估计音高,并用期望最大化方法估计音调模型参数,但在音符起点处会出现明显短时误差. Fujihara 等[18]在此基础上,用线性预测梅尔倒谱系数等特征来区分歌唱音与乐器音,进而实现声乐主旋律提取.

与 Goto 用高斯混合模型构建音调模型不同,Rao 和 Rao 将双向失配方法用于提取声乐主旋律[22],由于歌唱音谐波幅度衰减速度低于乐器音,故双向失配方法能明显提高歌唱音音高显著性.

文献[39]假设每对谱峰都是连续(奇次)谐波分量组合,由每对谱峰求出两个或四个音高候选,提出了基于谱峰对的多音高估计方法. 该方法在低次谐波分量受到伴奏音及其他乐音分量干扰时具有良好的鲁棒性,但其错误候选音高滤除方法较复杂.

文献[53]构建了基于谐波频率与十二平均律音符偏差的音高显著性度量函数,由于短时傅里叶变换存在频谱泄露,故该方法音高虚警率较高.

(2)基于谱峰能量的多音高表示方法

考虑到乐音包括基音与泛音分量,其含有的噪音具有宽频谱特性,故谐波加权和在各音符基频位置会出现较大峰值,据此,Klapuri 将各次谐波分量幅度加权求和作为基频显著性量度以检测基频,提出了谐波幅度求和方法[60,61]. 之后,该方法也应用于其他基于显著性的主旋律提取方法中[31,40]. 考虑到对数频率域各次谐波分量相对基频具有固定的偏移量,Hermes 提出了基于子谐波求和的语音信号音高确定方法[62],后来,这一方法也用于音乐信号音高估计[36,52,63]. 这两种方法都选择具有最大加权求和能量的频率作为候选旋律音高,符合旋律的显著性约束.

3.1.3　多音高跟踪

多音高跟踪方法,都用音高显著值结合平滑约束来识别主旋律线轨迹. 多音高跟踪方法主要分为三类:(1)结合动态音高显著函数值和平滑约束条件,寻找音高空间的最优演变路径,例如 Viterbi 方法[18,20]和动态规划方法[19,22]均属于此类;(2)跟踪正弦模型中的主要正弦分量,生成随时间变化的音高轨迹轮廓,并将具有最大累计显著能量的轨迹线作为最终的主旋律线,例如多代理(Multi-agent)跟踪方法[35]和多音高轨迹跟踪方法[38];(3)谐波聚类跟踪方法[41].

值得一提的是,在多音高跟踪阶段,利用嗓音的颤音、抖音以及人类的音域范围等信息来增强或辅助辨识主旋律,可提高声乐主旋律提取的准确性. Rao 和 Rao[22]根据嗓音的非平稳性检测嗓音分量,提高歌唱音

音高跟踪优先级,用动态规划方法跟踪声乐主旋律.这样,即使有强伴奏音存在,也能取得较好的主旋律提取效果.Arora 和 Behera[41]将音高轮廓的方差作为歌唱音的判断准则,用于音频源识别.Chien 等[64]建立了不同性别、风格、嗓音类型及元音的音色库,基于音色拟合(即观测音高轮廓与人声音色库内实例音色相似程度的量度)和响度两个参数来构建音高的似然函数.

3.1.4 基于显著性的典型主旋律提取方法及小结

上面综述了基于显著性的主旋律提取方法各主要功能模块的研究进展,应用这些技术,已经构建出多种主旋律提取方法.Salamon 用旋律音高均值分布和标准差分布、轮廓平均显著性分布和标准差分布、轮廓显著性标准差分布、轮廓整体显著性分布、轮廓长度分布等特征[31],从候选轮廓中逐渐删除最不可能轮廓,最终筛选出主旋律.

Rao 和 Rao 提出同时跟踪两条具有较大能量的旋律线,借助颤音和抖音等特征辅助识别歌唱音音高[22],该方法即使在歌唱音能量与器乐伴奏相近时,也能准确提取出声乐主旋律.

Dressler 用频谱图中的频率乘以幅度构建加权谱图,使幅度谱每八度提升 6dB,补偿音乐信号自身频谱幅度的自然衰落[39].由于加权谱图抵消了共振峰对歌唱音中频段谐波分量的增强效果,因此对声乐主旋律提取效果要差于器乐主旋律.

Chien 等构建大规模含歌唱音的主旋律数据库,提出了基于共振峰信息的声学-语音模型,并用该模型提取音乐音频中的歌唱音[54],在声乐主旋律提取方面取得了较好的效果.

基于显著性的主旋律提取,以信号处理技术为理论基础,结合主旋律的能量显著性和时间连续性等约束,构建显著性函数来估计帧级音高,并通过音高跟踪策略输出旋律线.该类方法具有思路清晰、理论明确、显著函数构建灵活等优点,是目前主旋律提取的最主要方法,并取得了良好的提取效果.

3.2 基于源分离的主旋律提取方法

基于源分离的主旋律提取方法框图如图 3 所示.该类方法先从音乐信号中分离出占主导地位的音频分量,然后跟踪主音频分量的音高.与基于显著性的主旋律提取方法不同,此处音高跟踪可看成是单声部音乐的音高跟踪问题.基于源分离的主旋律提取可分为基于谱分解和基于谐波聚类的方法.

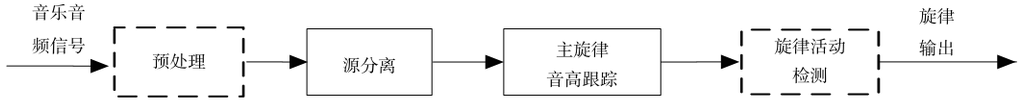

图3 基于源分离的主旋律提取框图

3.2.1 基于信号谱分解的源分离方法

基于信号谱分解的源分离方法主要包括:基于谐波音/敲击音声源分离模型的方法、基于子带互相关法和基于非负矩阵分解的方法.Tachibana 等[42]根据谐波音和敲击音在时间和频率域的“各向异性”,用谐波音/敲击音声源分离(HPSS)模型,通过优化方法求解谐波音分离问题,并跟踪最可能的旋律线.由于乐音伴奏或和弦会严重影响整体音高估计性能,且固定的音高范围不能自适应体现各曲目的音高动态变化,为此,Hsu 等引入趋势估计方法[65],用谐波音/敲击音声源分离模型[42]增强歌唱音,有效地抑制了同时存在的其它伴奏音,且音高范围根据输入音频估计得到,明显提高了旋律跟踪准确率.趋势估计方法在无旋律段进行音高跟踪时会出现紊乱,导致在歌唱音的起始段与结束段很难准确跟踪歌唱音.为此,Yeh 等提出了融合前向趋势估计、后向趋势估计和 HMM 模型的音高轮廓估计方法[21],并在 MIREX 2010 年主旋律提取评测中获得最高的原始音高准确率.龚君才和刘刚[66]也将 HPSS 模型和隐马尔科夫模型相结合来进行基频提取.

Li 和 Wang[67]根据听觉外周模型,将频谱分为多个子带,然后计算子带互相关谱,设计通道-峰选择算法检测歌唱音音高,借助 HMM 跟踪歌唱音音高,该方法用乡村乐、流行乐和摇滚乐三种风格数据库进行了测试,其性能优于 Klapuri 提出的多基频估计方法[61].

Durrieu 等[27]用非负矩阵分解方法来分离旋律和伴奏,并用 Viterbi 平滑方法进行主旋律音高跟踪.该方法对不同数据库具有良好的鲁棒性,但对中低频旋律音高其提取准确率较低,容易产生高八度错误.针对该问题,Wang 和 Ou[68]提出了 HMM 与非负矩阵分解相结合的主旋律提取方法,借助 Hsu 提出的感兴趣半音能量特征[63]辅助选择音高,从而抑制了八度错误,提高了运算速度.

3.2.2 基于谐波聚类的源分离方法

通常,主旋律提取要同时考虑每个乐音的谐波结构性和连续时间帧的动态演进性,以及低次谐波分量受低音伴奏干扰可能导致失真或被掩盖,为此,Arora 和 Behera 提出了两级谐波聚类在线主旋律提取方法[41].该方法基于卡尔曼滤波框架,同时跟踪多个声源,对每个声源仅跟踪高次谐波能量,尤其适于低次谐波分量被污染的情形.

宋岳阳根据人类发元音时声带和声道特征连续变化的特点以及相邻语音帧谐波能量互相关值较大的假设,提出了基于语音谐波能量互相关的基音频率跟踪方法[69],实现了基于单源欠定音频分离的主旋律提取.

基于源分离的主旋律提取方法,将音乐信号映射到时域之外的表示空间,通过映射空间中主旋律和伴奏表现出的差异将两者分离,从而抑制伴奏音并增强主旋律分量. 对于声乐旋律提取情形,基于源分离的方法可利用人声与乐器伴奏音的音域、速度、力度、声音抖动等特性分离出歌唱音,但对器乐旋律和强乐音伴奏情况,则很难将主乐器分量与其他伴奏分开,故基于源分离的方法主要适用于声乐旋律提取.

3.3　基于机器学习的主旋律提取方法

基于机器学习的主旋律提取方法是将主旋律提取看作模式分类问题,借助频域或其他表示域的旋律统计特征,用分类器分离出旋律音高,并通过动态规划类方法进行优化,提取出旋律线. Poliner 和 Ellis[70] 提出了基于分类器的旋律抄录方法,该方法以 STFT 幅度谱或对数幅度谱的傅里叶逆变换、MFCC 倒谱等为特征,以支持向量机为分类器,将旋律映射到 MIDI 音符,然后将 MIDI 音符的频率作为旋律音高. 该方法不需要先验知识,但对歌剧类的音频文件,其主旋律提取准确率较低. 之后,他们继续进行改进[43],用 HMM 模型平滑整体旋律,然后输出音高序列,该方法不需要谐波结构和周期结构假设,整体准确率提升了约 5%. Ryynänen 和 Klapuri 根据声学和音乐学模型,判断歌唱音音符存在与否,实现了基于 HMM 的声乐主旋律抄录系统[71]. Jo 等[45]用状态空间分析法构建序贯贝叶斯模型,以表示旋律音高、谐波幅度和复调音频的概率关系,用音乐统计特征计算旋律音高的转移概率,用序贯蒙特卡洛方法估计 Rao-Blackwellized 粒子滤波器参数,实现了主旋律提取. Song 和 Li[44] 在贝叶斯框架下建立音高演化模型和声学模型,利用音高频率差、谐波能量向量互相关系数、音高周期性、谐波形状等信息计算每个音高的概率,根据 MFCC 位移差分倒谱、谱对比度特征、谐波特征等来训练嗓音和非嗓音模型,最后用 Viterbi 方法搜索最佳音高序列.

近年来,深度神经网络在语音处理领域得到广泛应用. 由于深度神经网络能够建立主旋律与伴奏音之间的深层联系,因此可得到更精准的模型. Fan 等用深度神经网络从混合信号谱中分离出歌唱音的频谱,并用动态规划实现音高跟踪[72],该方法参加了 MIREX2015 主旋律提取评测,取得了令人满意的效果,详见 4.2 节表 4 中的“FYJ1”.

基于机器学习的主旋律提取方法,符合认知学习规律,可借助成熟的机器学习理论框架,但音乐信号种类繁多,风格各异,获得针对某乐曲的足够且准确的先验信息非常困难;音乐音频演奏的音高是连续变量,而用机器学习的方法会将连续变化的音高量化成音符,从而引入量化误差;此外,人类对于音乐的听觉认知(如音色等)目前还无法用某些具体特征准确表达,这些因素都限制了机器学习类主旋律提取方法的性能.

4　音乐主旋律提取方法的性能评价与最新成果

4.1　主旋律提取方法的性能评价

主旋律提取方法需完成两个任务:(1)准确估计旋律的音高. 当估计值与参考值之差在半个半音范围之内时,则认为本帧估计的旋律音高正确;若超过该范围,则认为本帧旋律提取错误;(2)旋律活动检测,即判别旋律存在与否,当有旋律帧时,输出主旋律音高值;无旋律帧时,输出音高值为 0.

主旋律提取方法的性能需通过技术指标进行评价,常用的评价指标[73]如表 3 所示.

表 3　主旋律提取性能评价指标

评价指标	定义
召回率(VR)	VR = #TP/#GV
误检率(VFA)	VFA = #FP/#GU
原始音高准确率(RPA)	RPA = (#TPC + #FNC)/#GV
原始音度准确率(RCA)	RCA = (#TPCch + #FNCch)/#GV
整体准确率(OA)	OA = (#TPC + #TN)/#TO

上述评价指标中部分参数的关系如图 4 所示,其中 GU 代表参考结果中无旋律帧,GV 参考结果中有旋律帧,DU 检测结果中无旋律帧,DV 检测结果中有旋律帧,TF 无旋律帧被正确检测,FN 有旋律帧被错误检测,TP 有旋律帧被正确检测,FP 无旋律帧被错误检测.

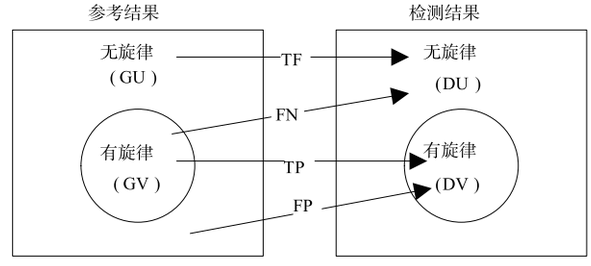

图4　主旋律提取的参考结果与检测结果的转移关系

4.2　主旋律提取的最新成果

为了推动主旋律提取技术的研究发展,自 2005 年起,国际音乐信息检索系统评估实验室组织的音乐信息检索评价交流评测(Music Information Retrieval Evaluation eXchange,MIREX[73])设置了主旋律提取项目,且评测结果在 MIREX 网站公布. 该领域最著名的国际会

议 ISMIR(International Conference on Music Information Retrieval)每年都会发表有关主旋律提取的最新研究进展和成果.

主旋律提取测试数据库包括 ADC2004、MIREX05、MIREX09、INDIAN08 等[73]. 表 4 列出了最近两年 MIREX 的主旋律提取项目在 ADC2004、MIREX05、MIREX09、INDIAN08 等数据库上取得较好效果的几种算法性能评价平均值,其中每个方法用各作者姓氏的首字母标识,数字代表提交评测的最终版本号,如"CWJ3"表示"Yu-RenChien, Hsin-Min Wang, Shyh-Kang Jeng"三位作者提交的第 3 版代码提取主旋律的结果.

KD3 方法[51]是典型的基于显著性的主旋律提取方法,它将声码器和 Charpentier 方法的谱峰瞬时频率均值作为谱峰频率估计值,从而提高了谱峰频率估计的精度,其平均原始音高准确率(RPA)最高. 在主旋律能量占主导地位时,其性能明显优于其他方法,但对主旋律分量没有明显能量优势的数据集(如 MIREX09(-5dB))[73],其准确率明显降低.

CWJ3 方法[64,74]是一种基于声学-语音学表示的主旋律提取方法,它利用人声似然音高模型估计人声旋律音高,用 40-phon 等响度曲线计算人耳感知音高响度,并将其作为旋律活动检测的依据. 该方法利用了人类嗓音统计特性和人耳音高感知特性,因此在旋律音强较低情况下,也能得到比其他方法稍高的准确率(如 MIREX09(-5dB))[73].

SL1 方法[44]利用旋律谐波能量向量的互相关性、旋律时域周期性、音高变化分布、人耳听觉特性等,结合 MFCC-SDC、SCF 和 HF 特征,实现基于贝叶斯框架的主旋律提取. 虽然 CWJ3 和 SL1 方法均考虑了歌唱音特点和人耳听觉特性,但两种方法基本思路有一定差异,SL1 方法是从音频中提取出具体特征作为音高估计和旋律活动检测的依据,而 CWJ3 方法则直接将音频数据作为处理对象,没有信息损失,故在评测数据库上性能优于 SL1 方法.

IYI1 方法[52]利用旋律显著性和时间连续性约束信息,用鲁棒主成分分析(RPCA)和二进制时-频掩蔽方法从音乐信号频谱图中分离出歌唱音频谱[75],然后用子谐波求和法构建显著函数,并用 Viterbi 方法估计音高轨迹. 该方法未考虑人耳听觉特性和强伴奏音的干扰,其整体准确率比 KD3 和 CWJ3 等方法稍低. 在此基础上,该课题组在 MIREX2015 提交了 IIY2 方法,利用估计的歌唱音音高轨迹辅助分离歌唱音,双向利用旋律提取和歌唱音分离互依赖信息,使得性能提升约 7%[76].

BG1 方法利用 Durrieu 的非负矩阵分解技术[27]来分离旋律和伴奏,并用 Salamon 的音高轮廓特征法[31],从各候选音高序列中筛选出最终的旋律. 该方法在近两年的评测中总体效果并不突出,但在 ORCHSET15 数据集上取得了最好的效果,可见该方法比较适合复调性较高的音乐,如交响乐[77].

表 4 最近两年 MIREX 旋律提取评测典型方法的评价结果

方法	OA	RPA	RCA	VR	VFA
SL1	0.57	0.52	0.55	0.73	0.23
DD1	0.71	0.72	0.75	0.86	0.30
KD3	0.73	0.81	0.83	0.91	0.41
IYI1	0.60	0.69	0.73	0.85	0.39
CWJ3	0.68	0.73	0.75	0.74	0.20
IIY2	0.67	0.76	0.78	0.94	0.53
FYJ1	0.68	0.75	0.78	0.71	0.12
BG1	0.61	0.69	0.75	0.85	0.53

西方音乐包括流行、爵士、乡村、歌剧、交响乐、摇滚、节奏布鲁斯等类型;中国音乐包括民族、古典、流行、摇滚等. 目前的主旋律提取方法大多都是通用的,可用于所有音乐类型,但对不同的音乐类型,其性能有所差异. 有少量已有研究是针对印度古典音乐、交响乐、西方歌剧等特定音乐类型的音乐. 我们对近两年 MIREX 主旋律提取评测典型方法在中西方音乐数据库上的效果进行了统计,结果表明,KD3 和 BG1 两种方法在西方音乐数据库上的效果明显优于在中国音乐数据库上(MIREX09)的效果,而 CWJ3 和 FYJ1 方法在中国音乐上的测试效果明显优于在西方音乐数据库上的测试效果. 由此可见,中西方音乐类型对主旋律提取结果有一定影响,有待进一步深入研究.

5 挑战与展望

主旋律提取是音乐信息处理领域的一项重要研究课题. 由于复调音乐的频谱结构复杂,准确提取主旋律具有一定的难度. 尽管主旋律提取技术已经取得了一定进展,但仍面临很多挑战.

(1) 音乐的基本要素包括旋律、节奏、音色、和声、速度等,这些要素间有一定的关联性. 目前,大多数方法在提取旋律时没有考虑旋律之外的其他要素,虽然有些研究考虑了音色等要素,但仍未找到充分描述音色的确切特征表示. 实际上,人类听觉系统在跟踪旋律时充分利用了音色等信息. 因此,应从信号处理角度挖掘音乐信号的深层次特征,寻找音乐各要素与旋律的联系,探索相应的主旋律提取方法,则会明显提高旋律提取的准确率.

(2) 目前的主旋律提取方法对二声部、三声部等音乐可以取得较好的主旋律提取效果,但处理复杂的

多声部音乐(如交响乐)时还很困难,因此对复杂的多声部音乐信号进行主旋律提取,还有待进一步研究.

(3) 近年来,主旋律提取方法研究主要围绕文中阐述的典型方法展开,很多衍生方法都是在此基础上做些改进,性能没有明显提高. 深度神经网络在语音处理领域已经取得巨大成就,近期刚引入到主旋律提取中,并取得较好效果. 因此,探索深度学习等新理论与新方法进行主旋律提取,这是未来发展的重要研究方向之一.

(4)旋律活动检测是制约旋律提取准确率的一个重要因素. 因此,自动学习每段音乐旋律的高级语义特征,并将其用于旋律活动检测,这会提高主旋律提取方法的准确率.

(5) 近年提出利用 HMM 与非负矩阵分解相结合、感兴趣半音能量等方法来减小主旋律提取八度错误,但目前已有主旋律提取算法的原始音度准确率与原始音高准确率间还有约 3% 的差异. 因此探索八度错误的去除方案仍是主旋律提取的重要课题之一.

(6) 目前的各种方法都是在指定的音乐数据库上进行实验,而数据库中的音乐风格种类和各风格的音乐条目数量有限,且强打击乐器干扰的音频片段较少,不能全面体现各不同风格乐曲的本质. 因此,丰富测试数据库是提高方法性能评价客观性的重要条件,也是主旋律提取领域要完成的一项重要任务.

6　结束语

主旋律提取是音乐信号处理领域的一项重要研究课题,在音乐检索、音乐推荐、翻唱识别等领域具有广阔应用前景,研究多声部音乐主旋律提取具有重要意义. 本文综述了复调音乐主旋律提取的研究进展,阐述了基于显著性的主旋律提取方法、基于源分离的主旋律提取方法及基于机器学习的主旋律提取方法,简要介绍了常用的主旋律提取测试数据库及评价指标,展示了 MIREX 主旋律提取评测最新结果,并对性能较好的方法进行了说明. 最后,介绍了复调音乐主旋律提取现有研究成果及其局限性,指出了主旋律提取研究所面临的主要挑战和未来发展方向.

参考文献

[1] Poliner G E, Ellis D P, Ehmann A F, et al. Melody transcription from music audio: approaches and evaluation [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(4): 1247 – 1256.

[2] 李重光. 基本乐理通用教材[M]. 北京:高等教育出版社,2004.
Li Chong-Guang. Basic Music Theory Textbook [M]. Beijing: Higher Education Press, 2004 (in Chinese)

[3] Salamon J, Gomez E, Ellis D P, et al. Melody extraction from polyphonic music signals: approaches, applications, and challenges [J]. IEEE Signal Processing Magazine, 2014, 31(2): 118 – 134.

[4] Maddage N C, Xu C, Kankanhalli M S, et al. Content-based music structure analysis with applications to music semantics understanding [A]. 12th Annual ACM International Conference on Multimedia [C]. New York, USA, 1999. 112 – 119.

[5] Rolland P Y, Raškinis G, Ganascia J G. Musical content-based retrieval: an overview of the Melodiscov approach and system [A]. 7th ACM International Conference on Multimedia [C]. Florida: The Special Interest Group on Multimedia, 2004. 81 – 84.

[6] Salamon J, Serra J, Gómez E. Tonal representations for music retrieval: from version identification to query-by-humming [J]. International Journal of Multimedia Information Retrieval, 2013, 2(1): 45 – 58.

[7] Salamon J, Rocha B, Gómez E. Musical genre classification using melody features extracted from polyphonic music signals [A]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [C]. Toyoto: IEEE Signal Processing Society, 2012. 81 – 84.

[8] Tsai W H, Yu H M, Wang H M. Using the similarity of main melodies to identify cover versions of popular songs for music document retrieval [J]. Journal of Information Science and Engineering, 2008, 24(6): 1669 – 1687.

[9] Foucard R, Durrieu J L, Lagrange M, et al. Multimodal similarity between musical streams for cover version detection [A]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [C]. Dallas: IEEE Signal Processing Society, 2010. 5514 – 5517.

[10] Wu Y D, Li Y, Liu B L. A new method for approximate melody matching [A]. International Conference on Machine Learning and Cybernetics [C]. Xi´an: IEEE Systems, Man, and Cybernetics Society, 2003. 2687 – 2691.

[11] Chang C W, Jiau H C. Extracting significant repeating figures in music by using quantized melody contour [A]. IEEE International Symposium on Computers and Communication [C]. Antalya: IEEE Communications Society, 2003. 1061 – 1066.

[12] Ozcan G, Isikhan C, Alpkocak A. Melody extraction on MIDI music files [A]. 7th IEEE International Symposium on Multimedia [C]. California: IEEE Computer Society, 2005. 1 – 8.

[13] Li J, Yang X, Chen Q. MIDI melody extraction based on improved neural network [A]. International Conference on Machine Learning and Cybernetics [C]. Baoding:

IEEE Systems, Man, and Cybernetics Society, 2009. 1133 – 1138.

[14] Shih H H, Narayanan S S, Kuo C C. Automatic main melody extraction from MIDI files with a modified Lempel-Ziv algorithm [A]. International Symposium on Intelligent Multimedia, Video and Speech Processing [C]. Hongkong: IEEE Tainan Section, 2001. 9 – 12.

[15] Akant K, Limaye S. Pitch contour extraction of singing voice in polyphonic recordings of Indian classical music [A]. International Conference on Electronic Systems, Signal Processing and Computing Technologies [C]. Maharashtra: IEEE Computer Society, 2014. 123 – 128.

[16] Cao C, Li M, Liu J, et al. Singing melody extraction in polyphonic music by harmonic tracking [A]. 8th International Society for Music Information Retrieval Conference [C]. Vienna: Music Information Retrieval Society, 2007. 373 – 374.

[17] Dong M, Chan P, Cen L, et al. Aligning singing voice with MIDI melody using synthesized audio signal [A]. 7th International Symposium on Chinese Spoken Language Processing [C]. Nantou: International Speech Communication Association, 2010. 95 – 98.

[18] Fujihara H, Kitahara T, Goto M, et al. F0 estimation method for singing voice in polyphonic audio signal based on statistical vocal model and Viterbi search [A]. IEEE International Conference on Acoustics, Speech and Signal Processing [C]. Quebec: IEEE Signal Processing Society, 2006. 253 – 256.

[19] Hsu C L, Jang J S R. Singing pitch extraction by voice vibrato/tremolo estimation and instrument partial deletion [A]. 11th International Society for Music Information Retrieval [C]. Florida: Music Information Retrieval Society, 2010. 525 – 530.

[20] Li M, Li J, Han J, et al. Singing melody extraction from pop songs using a novel feature and Viterbi search [A]. International Conference on Computational Intelligence and Software Engineering [C]. Wuhan: Engineering Information Institute, 2010. 1 – 4.

[21] Yeh T C, Wu M J, Jang J, et al. A hybrid approach to singing pitch extraction based on trend estimation and hidden Markov models [A]. IEEE International Conference on Acoustics, Speech and Signal Processing [C]. Kyoto: IEEE Signal Processing Society, 2012. 457 – 460.

[22] Rao V, Rao P. Vocal melody extraction in the presence of pitched accompaniment in polyphonic music [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 18(8): 2145 – 2154.

[23] Sutton C, Vincent E, Plumbley M D, et al. Transcription of vocal melodies using voice characteristics and algorithm fusion [J]. Music Information Retrieval Evaluation eXchange, 2006, 7(1): 1 – 4.

[24] Vidwans A, Ganguli K K, Rao P. Classification of Indian classical vocal styles from melodic contours [A]. 2nd CompMusic Workshop [C]. Istanbul: Music Technology Group of the Universitat Pompeu Fabra in Barcelona, 2012. 139 – 146.

[25] Shao X, Xu C, Kankanhalli M S. Predominant vocal pitch detection in polyphonic music [A]. IEEE International Conference on Multimedia and Expo [C]. Ontario: IEEE Circuits and Systems Society, 2006. 897 – 900.

[26] Yao G, Zheng Y, Xiao L, et al. Efficient vocal melody extraction from polyphonic music signals [J]. Elektronika ir Elektrotechnika, 2013, 19(6): 103 – 108.

[27] Durrieu J L, Richard G, David B, et al. Source/filter model for unsupervised main melody extraction from polyphonic audio signals [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 18(3): 564 – 575.

[28] Gómez E, Klapuri A, Meudic B. Melody description and extraction in the context of music content processing [J]. Journal of New Music Research, 2003, 32(1): 23 – 40.

[29] Joo S, Park S, Jo S, et al. Melody extraction based on harmonic coded structure [A]. 12th International Society for Music Information Retrieval Conference (ISMIR) [C]. Florida: Music Information Retrieval Society, 2011. 227 – 232.

[30] Marolt M. Audio melody extraction based on timbral similarity of melodic fragments [A]. International Conference on Computer as a Tool [C]. Belgrade: IEEE Vancouver Section, 2005. 1288 – 1291.

[31] Salamon J, Gómez E. Melody extraction from polyphonic music signals using pitch contour characteristics [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(6): 1759 – 1770.

[32] Thornburg H, Leistikow R J, Berger J. Melody extraction and musical onset detection via probabilistic models of framewise STFT peak data [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(4): 1257 – 1272.

[33] Han J, Chen C W. Improving melody extraction using probabilistic latent component analysis [A]. IEEE International Conference on Acoustics, Speech and Signal Processing [C]. Prague: IEEE Signal Processing Society, 2011. 33 – 36.

[34] Goto M. A robust predominant-F0 estimation method for real-time detection of melody and bass lines in CD recordings [A]. IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. Istanbul: IEEE Signal

Processing Society,2000. 757 – 760.

[35] Goto M. A real-time music-scene-description system:Predominant-F0 estimation for detecting melody and bass lines in real-world audio signals [J]. Speech Communication,2004,43(4):311 – 329.

[36] Cancela P. Tracking melody in polyphonic audio[A]. 9th Music Information Retrieval Evaluation eXchange (MIREX 2008) [C]. Illinois:The International Music Information Retrieval Systems Evaluation Laboratory,2008. 9(1):1 – 3.

[37] Joo S,Jo S,Yoo C D. Melody extraction from polyphonic audio signal [A]. 11th Music Information Retrieval Evaluation eXchange (MIREX) [C]. Illinois: The International Music Information Retrieval Systems Evaluation Laboratory,2010. 11(1):1 – 4.

[38] Paiva R P, Mendes T, Cardoso A. Melody detection in polyphonic musical signals: Exploiting perceptual rules, note salience, and melodic smoothness [J]. Computer Music Journal,2006,30(4):80 – 98.

[39] Dressler K. Pitch estimation by the pair-wise evaluation of spectral peaks [A]. AES 42nd International Conference [C]. Ilmenau: Australasian Evaluation Society, 2011. 1 – 10.

[40] Ryynänen M P, Klapuri A P. Automatic transcription of melody,bass line,and chords in polyphonic music [J]. Computer Music Journal,2008,32(3):72 – 86.

[41] Arora V,Behera L. On-line melody extraction from polyphonic audio using harmonic cluster tracking [J]. IEEE Transactions on Audio,Speech,and Language Processing, 2013,21(3):520 – 530.

[42] Tachibana H,Ono T,Ono N,et al. Melody line estimation in homophonic music audio signals based on temporal-variability of melodic source [A]. IEEE International Conference on Acoustics, Speech and Signal Processing [C]. Dallas:IEEE Signal Processing Society,2010. 425 – 428.

[43] Ellis D P,Poliner G E. Classification-based melody transcription [J]. Machine Learning, 2006, 65 (2 – 3): 439 – 456.

[44] Song L, Li M. Bayesian framework-based vocal melody extraction for MIREX2014 [A]. Music Information Retrieval Evaluation eXchange (MIREX) [C]. Illinois:The International Music Information Retrieval Systems Evaluation Laboratory,2014. 15(1):1 – 2.

[45] Jo S,Yoo C D,Doucet A. Melody tracking based on sequential bayesian model [J]. IEEE Journal of Selected Topics in Signal Processing,2011,5(6):1216 – 1227.

[46] Yoon J Y, Song C J, Lee S P, et al. Extracting predominant melody of polyphonic music based on harmonic structure [A]. 7th Music Information Retrieval Evaluation eXchange (MIREX) [C]. Illinois:The International Music Information Retrieval Systems Evaluation Laboratory, 2011. 7(1):1 – 2.

[47] Ryynänen M, Virtanen T, Paulus J, et al. Accompaniment separation and Karaoke application based on automatic melody transcription [A]. IEEE International Conference on Multimedia and Expo [C]. Hannover: IEEE Circuits and Systems Society,2008. 1417 – 1420.

[48] Zhu Y, Gao S. Extracting vocal melody from Karaoke music audio [A]. IEEE International Conference on Multimedia and Expo [C]. Amsterdam:IEEE Circuits and Systems Society,2005. 1 – 5.

[49] Bosch J J, Gómez E. Melody extraction in symphonic classical music:a comparative study of mutual agreement between humans and algorithms [A]. 9th Conference on Interdisciplinary Musicology [C]. Berlin: World Academy of Science,Engineering and Technology,2014. 1 – 6.

[50] Tang Z, Black D A. Melody extraction from polyphonic audio of western opera:a method based on detection of the singer's formant[A]. 7th International Society for Music Information Retrieval Conference (ISMIR)[C]. Taipei: Music Information Retrieval Society, 2014. 161 – 166.

[51] Dressler K. Audio melody extraction for MIREX2014 [A]. Music Information Retrieval Evaluation eXchange (MIREX) [C]. Illinois:The International Music Information Retrieval Systems Evaluation Laboratory,2014. 15 (1):1 – 3.

[52] Ikemiya Y, Yoshii K, Itoyama K. MIREX2014:audio melody extraction [A]. Music Information Retrieval Evaluation eXchange (MIREX) [C]. Illinois:The International Music Information Retrieval Systems Evaluation Laboratory,2014. 15(1):1 – 2.

[53] Degani A,Leonardi R,Migliorati P,et al. A pitch salience function derived from harmonic frequency deviations for polyphonic music analysis [A]. 17th International Conference on Digital Audio Effects [C]. Erlangen: Fraunhofer Institut für Integrierte Schaltungen,2014. 1 – 7.

[54] Chien Y R, Wang H M, Jeng S K. An acoustic-phonetic model of F0 likelihood for vocal melody extraction [J]. IEEE/ACM Transactions on Audio,Speech and Language Processing,2015,23(9):1457 – 1468.

[55] Cancela P,Rocamora M,López E. An efficient multi-resolution spectral transform for music analysis [A]. 10th International Society for Music Information Retrieval [C]. Kobe: Music Information Retrieval Society, 2009. 309 – 314.

[56] Park S,Jo S,Yoo C D. Melody extraction from polyphon-

ic audio signal MIREX 2011 [A]. 7th Music Info Retrieval Evaluation eX-change (MIREX) [C]. Illinois: The International Music Information Retrieval Systems Evaluation Laboratory, 2011. 102(1): 1 – 2.

[57] Charpentier F. Pitch detection using the short-term phase spectrum [A]. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) [C]. Tokyo: IEEE Signal Processing Society, 1986. 113 – 116.

[58] Flanagan J L, Golden R. Phase vocoder [J]. Bell System Technical Journal, 1966, 45(9): 1493 – 1509.

[59] Salamon J, Gómez E, Bonada J. Sinusoid extraction and salience function design for predominant melody estimation [A]. 14th International Conference on Digital Audio Effects[C]. Paris: The National Center for Scientific Research, 2011. 73 – 80.

[60] Klapuri A P. Multiple fundamental frequency estimation by summing harmonic amplitudes [A]. 7th International Society for Music Information Retrieval Conference (ISMIR) [C]. Victoria: Music Information Retrieval Society, 2006. 216 – 221.

[61] Klapuri A P. Multiple fundamental frequency estimation based on harmonicity and spectral smoothness [J]. IEEE Transactions on Speech and Audio Processing, 2003, 11(6): 804 – 816.

[62] Hermes D J. Measurement of pitch by subharmonic summation [J]. Journal of the Acoustical Society of America, 1988, 83(1): 257 – 264.

[63] Hsu C L, Chen LY, Jang J S R, et al. Singing pitch extraction from monaural polyphonic songs by contextual audio modeling and singing harmonic enhancement [A]. 10th International Society for Music Information Retrieval Conference[C]. Kobe: Music Information Retrieval Society, 2009. 201 – 206.

[64] Chien Y R, Wang H M, Jeng S K. Vocal melody extraction based on an acoustic-phonetic model of pitch likelihood [A]. Music Information Retrieval Evaluation eX-change (MIREX) [C]. Illinois: The International Music Information Retrieval Systems Evaluation Laboratory, 2014. 15(1): 1 – 2.

[65] Hsu C L, Wang D, Jang J S. A trend estimation algorithm for singing pitch detection in musical recordings [A]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [C]. Prague: IEEE Signal Processing Society, 2011. 393 – 396

[66] 龚君才,刘刚. 一种基于隐马尔科夫模型的波形文件主旋律基频提取算法[J]. 软件,2013,34(12):152 – 155.
Gong Jun-Cai, Liu Gang. A melody pitch extraction algorithm for waveform file based on hidden Markov mode [J]. Software, 2013, 34(12): 152 – 155. (in Chinese)

[67] Li Y, Wang D. Separation of singing voice from music accompaniment for monaural recordings [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(4): 1475 – 1487.

[68] Wang Y, Ou Z. Combining HMM-based melody extraction and NMF-based soft masking for separating voice and accompaniment from monaural audio [A]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [C]. Prague: IEEE Signal Processing Society, 2011. 1 – 4.

[69] 宋岳阳. 基于单源欠定语音分离的音乐主旋律提取方法研究[D]. 北京:北京邮电大学,2012.
Song Yue-Yang. An approach for music melody extraction based on underdetermined single-source speech separation [D]. Beijing: Beijing University of Posts and Telecommunications, 2012. (in Chinese)

[70] Poliner G E, Ellis D P. A classification approach to melody transcription [A]. 6th International Conference on Music Information Retrieval (ISMIR) [C]. London: Music Information Retrieval Society, 2005. 161 – 166.

[71] Ryynänen M, Klapuri A. Transcription of the singing melody in polyphonic music [A]. International Society for Music Information Retrieval (ISMIR) [C]. Victoria: Music Information Retrieval Society, 2006. 222 – 227.

[72] Fan Z, Jang J, Lu C. Singing voice separation and pitch extraction from monaural polyphonic audio music via DNN and ddaptive pitch tracking [A]. 2nd IEEE International Conference on Multimedia Big Data [C]. Taipei: Academia Sinica, 2016. 1 – 8.

[73] http://www.music-ir.org/mirex/wiki/MIREX_HOME.

[74] Chien Y R, Wang H M, Jeng S K. Simulated formant modeling of accompanied singing signals for vocal melody extraction [A]. 9th Sound and Music Computing Conference (SMC) [C]. Copenhagen: Logos Verlag Berlin GmbH, 2012. 33 – 40.

[75] Huang P S, Chen S D, P Smaragdis, et al. Singing voice separation from monaural recordings using robust principal component analysis [A]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [C]. Toyoto: IEEE Signal Processing Society, 2012. 57 – 60.

[76] Ikemiya Y, Yoshii K, Itoyama K. Singing voice analysis and editing based on mutually dependent F0 estimation and source separation[A]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [C]. Brisbane: IEEE Signal Processing Society, 2015. 574 – 578.

[77] Bosch J, Gómez E. Melody extraction by means of a source-filter model and pitch contour characterization

[A]. Music Informormation Retrieval Evaluation eXchange (MIREX) [C]. Illinois: The International Music Information Retrieval Systems Evaluation Laboratory, 2015. 16(1):1 - 3.

作者简介

张维维　女,辽宁大连人,1981 年 5 月生,大连理工大学在读博士生,大连民族大学讲师. 主要研究方向为音乐信号处理,音乐信息检索.
Email:zhangww@ dlnu. edu. cn

陈　喆　男,黑龙江省泰来人,1975 年 11 月生,大连理工大学副教授. 主要研究方向为数字信号处理、语音处理、图像处理、宽带无线通信技术.
E-mail:zhechen@ dlut. edu. cn