

【音乐理论】

中国传统乐器音响数据库构建研究

李子晋 韩宝强

【摘要】本文针对中国传统乐器音响数据库自身特点、构建方法以及需要解决的问题加以阐述。讨论了专门为中国传统乐器音响建立数据库的意义及其可持续利用的价值。提出利用前沿计算机声音分析技术实现自动标注体系的建立,基于内容的海量音响数据检索,以及其他有助于自动完善数据库功能、方便数据库管理的方法。在国内已有若干乐器音响数据库类的建设项目中摆脱了同质化窠臼,避免了重复性建设,将音响数据库的资源结合以现代音乐科技而走出一条新路,为音乐信息检索、音乐数据机器学习以及未来音乐人工智能的发展等方面进行基础性建设,做出了探索与阐述,并指明中国传统乐器音响数据库建设的意义并非仅在于保存已有数据,而在于昭示未来乐器以及乐器学学科整体发展的方向。

【关键词】人工智能;音乐信息检索;民族乐器;音响数据库

【作者简介】李子晋(1982-),女,博士,中国音乐学院副教授、硕士研究生导师;韩宝强(1956-),男,博士,中国音乐学院教授、博士研究生导师。

【原文出处】《中国音乐学》(京),2020.2.92~102

【基金项目】本文系国家艺术基金“中国传统乐器交互式数字博物馆建设与推广”项目(01020120180529564031)阶段性成果。感谢文化部科技项目“中国民族乐器音响标准库”、文化部科技提升重大项目“民族低音拉弦乐器改良”、科技部基础项目“中国传统乐器声学测量与频谱分析”、北京市科研专项基金“中国民族乐器多媒体数据库”,以及在课题中做出贡献的师生。

中国传统乐器是中国传统音乐文化的物质载体,是绚烂夺目的传统文化的重要组成部分,在国内外享有极高的声誉,拥有广泛的市场需求。众所周知,声音是乐器的灵魂,音响性能是乐器的核心。对于乐器音响的研究直接反映不同时期、不同民族的文化背景和审美积淀。继承中国传统乐器的音响数据不仅为中国音乐的创作、表演、理论、教育、推广、研究提供权威可靠的基础数据支持,而且可以为其他国家和民族提供了解中国传统乐器音响特色的平台。近年来,随着计算机前沿技术发展,人工智能技术已开始应用于数据库建设,在我们看来,将信息深入挖掘的概念和技术应用于传统乐器音响数据建设的路径也变得越来越清晰。

一、乐器音响数据库相关工作

中国传统乐器音响数据库是收录有关中国传统乐器音响信息的数据集。目前,对于中国传统乐器音响数据库的研究较多地出现在博物馆学中。它侧

重于对乐器收藏的数字化建设,主要目标是完善博物馆档案数据,是实物收藏的一种补充。如中央音乐学院与华中科技大学、华中师范大学联合创建的《中国国家级乐器数字博物馆》^①运用信息化手段,实现了中国乐器数字博物馆资源的数字化建设与共享。乐器博物馆在线 MIMO(Musical Instrument Museums Online)^②,也属于这种类型。这种类型的数据库大多侧重于图文并茂的生动展示,主要目的是满足群众的物质文化生活需求,而并不专注于音响本身的质量、声学特征、特性标注、检索以及相应音响学上的学术应用,仅仅是乐器实物信息的虚拟化补充。

在文化艺术领域,近年来用于音乐学学术研究与应用开发的中国传统乐器音响数据库逐步得到关注并有了少量实践,如文化部“中国民族乐器音响标准库”^③、“民族低音拉弦乐器改良”^④,科技部“中国传统乐器音响与频谱分析”,北京市“中国民

族乐器多媒体数据库”等。这类乐器音响数据库沿着音乐文化资料收藏的思路,专注于乐器学、演奏法等。内容一般包括乐器的单音、音阶、演奏技巧、乐曲片段,录音环境选择消声室或录音棚等优质环境,演奏多为专业演奏员为主,部分还进行了音响特征的初步分析。这些都从乐器本体的角度出发,为乐器的音质评价、性能评估、主观实验提供了丰富的基础数据。

在传统乐器音响数据库中,有很多涉及音响信号本身的应用,比如智能乐器识别、音乐搜索、乐器辅助教学等,都需要与相关的科学技术领域紧密结合。这样的学科叫做基于内容的音乐信息检索,即CBMIR(Content-based Music Information Retrieval),简称MIR。MIR技术是近二十年来,随着数字化音乐数量急剧增加而出现的一个新兴学科。MIR技术研究范围广泛,包括旋律提取、节奏分析、和弦识别、音乐搜索、流派分类、音乐情感计算、音乐推荐、音乐内容标注、歌手识别、歌唱评价、歌声合成、算法作曲、音频水印等数十个子领域,在娱乐、音乐教育及科研、日常生活、心理及辅助医疗、版权保护、电影、机器人等方面应用众多^⑤。MIR技术框架以音频信号处理及人工智能AI(Artificial Intelligence)—机器学习(Machine Learning)为基础,可看成人工智能(AI)在音乐上的分支,需要大量数据对机器学习模型进行训练。因此,在MIR领域中,数据库的建设占据极其重要的地位。

MIR领域的数据库内容大多数为带标注的纯乐器音响数据库或音乐数据库中的乐器音响片段等,用于音高、音色、旋律、节奏、和声等信息的提取。在这类数据库中,乐器种类多以西洋管弦乐队中的乐器为主,辅以建立数据库的国家特有的乐器,或以研究对象涉及地区的乐器为主,上述乐器库目前都来自于国外,并没有专门的中国传统乐器数据库。这类数据库的乐器音响内容以几秒钟的音响片段为主,大多无完整乐器音阶、乐器片段等信息,且录音环境良莠不齐,经常不能满足研究者的需求。以日本的RWC(Real World Computing)Music Database^⑥为例,该数据库是世界上第一个专门为音乐信息技术研究而研制的大型数据库,也是MIR领域使用频率

较高的数据库之一。它涵盖了流行音乐数据库、古典音乐数据库、爵士音乐数据库、音乐流派数据库、乐器音响数据库等五项内容。其中乐器音响数据库包括50种乐器,共150段乐器演奏样本。样本为单声道、16bit量化、44.1kHz采样、RIFF WAVE格式数据,内容以西方管弦乐器及日本传统乐器为主。但涉及的乐器内容以音频片段为主,不涉及从乐器本体出发的音阶、音色、音量等可用于音域、动态范围等测量的必要数据。除了RWC数据库,在MIR领域,国外还有数个涉及乐器的音乐数据库^⑦,组织架构与RWC大同小异,也同样都没有涉及中国传统乐器,这为中国传统乐器数据的深入挖掘带来了阻力。

二、中国传统乐器音响数据库(CTIS)的特征

中国传统乐器音响数据库(Chinese Traditional Instrument Sound Database)拟包括传统乐器、改良乐器、少数民族乐器在内的287件中国民族乐器,目前已完成40余件乐器的入库及数据库的基本构架,该数据库的主要特点体现在以下三个方面。

(一)完备性

主要体现在乐器数量、种类和内容的完备。乐器音响采集从气鸣、体鸣、膜鸣、弦鸣,涉及传统乐器、改良乐器、少数民族乐器200多种。除此之外,对于乐器的制作者、演奏者、弹弦位置、演奏音高、演奏技法做了全面的标注。由于中国民族乐器的制作工艺以及演奏者对乐器音质的影响很大,因此,对每件乐器进行详尽的标注是十分有必要的。对已有采集的部分乐器列表如下。

每种乐器的演奏技巧均有单独录音和标注,以其中的京胡为例,下级目录为。

(二)标准化

标准化是人类文明发展史上最重要标志之一,然而人们对音乐中的标准化重要性的认识还比较陌生。实际上,中外音乐家从两千多年前就开始探索标准化问题。众所周知的“三分损益律”和“五度相生律”,就是中国和古希腊先贤为解决音高标准而制定的计算方法。^⑧今天全世界都在遵循的国际标准音高(international standard pitch),即A4=440Hz,是1939年由国际标准协会在伦敦决定

表1 部分采录乐器列表

名称	总容量	文件数量	涵盖内容	文件格式
24簧键笙	63.9MB(67,055,297字节)	66	音阶、技法、乐曲片段	.wav(RIFF)
三弦	21.8MB(22,908,928字节)	38	音阶、技法、乐曲片段	.wav(RIFF)
中音板胡	16.3MB(17,160,688字节)	19	音阶、技法、乐曲片段	.wav(RIFF)
中音虎锣	3.55MB(3,731,456字节)	6	音阶、技法、乐曲片段	.wav(RIFF)
二十四簧键笙	32.8MB(34,447,261字节)	66	音阶、技法、乐曲片段	.wav(RIFF)
二胡	25.2MB(26,460,160字节)	30	音阶、技法、乐曲片段	.wav(RIFF)
京胡	39.5MB(41,435,509字节)	49	音阶、技法、乐曲片段	.wav(RIFF)
京胡二黄	6.72MB(7,053,312字节)	20	音阶、技法、乐曲片段	.wav(RIFF)
京胡西皮	32.8MB(34,418,688字节)	27	音阶、技法、乐曲片段	.wav(RIFF)
传统笙	22.7MB(23,831,898字节)	29	音阶、技法、乐曲片段	.wav(RIFF)
古琴	18.1MB(19,036,872字节)	19	音阶、技法、乐曲片段	.wav(RIFF)
唢呐	36.2MB(37,999,132字节)	36	音阶、技法、乐曲片段	.wav(RIFF)
大锣	7.69MB(8,065,024字节)	11	音阶、技法、乐曲片段	.wav(RIFF)
大鼓	15.7MB(16,547,582字节)	16	音阶、技法、乐曲片段	.wav(RIFF)
小锣	6.20MB(6,504,448字节)	9	音阶、技法、乐曲片段	.wav(RIFF)
扬琴	127MB(133,481,940字节)	30	音阶、技法、乐曲片段	.wav(RIFF)
排鼓	15.2MB(15,958,486字节)	5	音阶、技法、乐曲片段	.wav(RIFF)
曲笛	22.6MB(23,751,935字节)	32	音阶、技法、乐曲片段	.wav(RIFF)
柳琴	9.05MB(9,491,981字节)	19	音阶、技法、乐曲片段	.wav(RIFF)
梆笛	21.2MB(22,324,131字节)	35	音阶、技法、乐曲片段	.wav(RIFF)
琵琶	22.0MB(23,088,461字节)	45	音阶、技法、乐曲片段	.wav(RIFF)
筝	18.9MB(19,890,237字节)	20	音阶、技法、乐曲片段	.wav(RIFF)
高胡	15.2MB(15,991,094字节)	16	音阶、技法、乐曲片段	.wav(RIFF)
高音板胡	34.5MB(36,232,300字节)	46	音阶、技法、乐曲片段	.wav(RIFF)
高音虎锣	3.15MB(3,309,568字节)	6	音阶、技法、乐曲片段	.wav(RIFF)

表2 京胡演奏技巧及样本标注

京胡演奏技巧及样本标注	内容概述	大小(KB)
京胡颤弓西皮 1y.wav	西皮京二胡的颤弓技法	559,664
京胡抖弓西皮 1y.wav	西皮京二胡的抖弓技法	1,401,534
京胡快弓西皮 1y.wav	西皮京二胡的快弓技法	214,260
京胡连弓里弦西皮 1y.wav	西皮京二胡里弦的颤弓技法	786,028
京胡推弓西皮 1y.wav	西皮京二胡的推弓技法	1,467,800
京胡中弓西皮 1y.wav	西皮京二胡的中弓技法	178,158
京胡长弓西皮 1y.wav	西皮京二胡的长弓技法	489,790
京胡滑音二黄 1y.wav	二黄京二胡的滑音技法	359,792
京胡抹音二黄 1y.wav	二黄京二胡的抹音技法	680,250
京胡揉音二黄 1y.wav	二黄京二胡的揉音技法	493,264
京胡空弦西皮 1y.wav	西皮京二胡空弦音演奏	637,336
京胡揉音二黄 2y.wav	二黄京二胡的揉音技法	644,316
京胡示范乐曲 1y.wav	一般京二胡示范曲	1,531,502

续表 2

京胡演奏技巧及样本标注	内容概述	大小(KB)
京胡示范曲二黄 1y.wav	二黄京二胡的示范曲	1,214,554
京胡示范曲反二黄 1y.wav	反二黄京二胡的示范曲	2,277,658
京胡示范曲西皮 1y.wav	西皮京二胡的示范曲	19,401,412

的。中国于 1956 年 6 月决定加以采用。^⑨正是有了国际音高标准,现在中国音乐家才能和世界其他音乐家一起同台演出,为观众奉献中外精湛音乐作品。音乐的研究也需要这种标准化的准备,一直以来学术界对于数据库所选音响片段的质量要求逐渐提高,这要求在制作、收集数据库素材方面,需要专业音乐学习者对基础音乐片段进行筛选与判断,从而保证数据库音乐的严谨性与专业性,本数据库的标准化主要体现在录音环境、专业的演奏员及样本内容及规范的标注。

本数据库的应用场景大多数涉及乐器音响研究,且为体现中国民族乐器涉及许多细腻的演奏技法,如古琴的揉、箜篌等民族乐器的压颤等技法,因此,在全消声室中对乐器进行声学采录与测量。麦克风置于全消声室中间的方形平台,并放置于直径为 3m 的 3/4 球形支架,在其上均匀布放 30 只 B&K4190-L-001 型自由场传声器单元,其中底层布放 6 只,第二层布放 8 只,第三层布放 6 只,第四层布放 5 只,第五层布放 4 只,顶部布放 1 只。在乐器正前方采录位置布放 1 只进行频谱采录和监听,在乐器正下方的地格栅下和地网上的位置布放 1 只,总共采用 32 只自由场传声器单元进行全消声室中的乐器采

录。目前在消声室中已完成对中国传统乐器中的古筝、古琴、扬琴、二胡、唢呐和阮等 15 件民族乐器的采录。下图为大阮和古筝采录时的照片。

为保证乐器音响样本的可复制性,需体现该乐器的音高、音强、音色、时值等元素,从而对其进行声学测量与分析而得出量化数据;这些离散性数据的集合还要进一步整合,从而对乐器音域、动态范围、音区与音色、音区与音强的关系进行归类与定性化描述;乐器所发出的音响是人对其施加演奏行为的结果,因此还需要选择所有常规演奏技法生成的典型性音响;以上样本在很大程度上是脱离了音乐语境的静态样本,而作为构成音乐元素之一的乐器音响,还应选择音乐作品中的较为短小的代表性片段,此为活态样本。

此外,乐器由人操控,演奏员对乐器实施的演奏行为直接影响着乐器音响的结果,即使是同一种技术,其运用与把握的程度也会导致音响结果产生较大的差异。因此,要求演奏人员必须掌握娴熟的乐器演奏技术,并能够最大程度上以标准化而不是个性化的技法进行演奏,从而生成该乐器常规演奏过程中的典型音响。因此,本数据库对演奏员的选择,主要以专业演奏家或专业教师为主。同时演奏

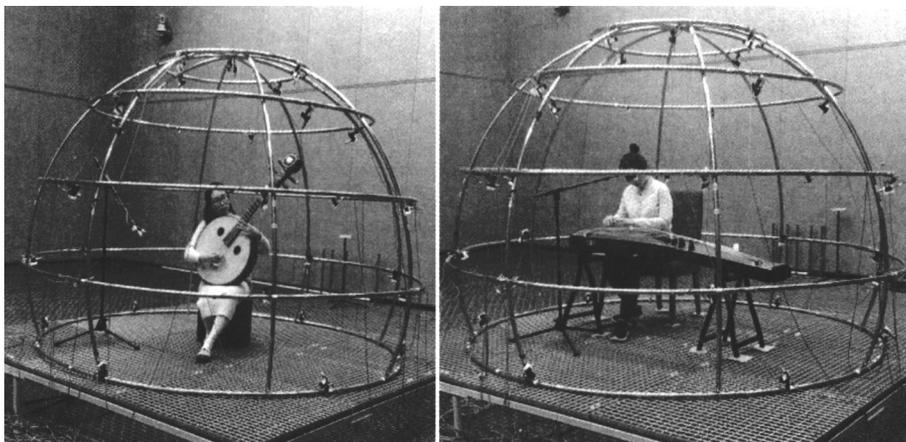


图 1 大阮与古筝采录照片

员在样本录制现场时需保持身体与情绪状态均为正常状态。^⑩

(三)多功能性

本数据库的建构主要目的之一是便于音乐学界所有研究者进行研究,因此利用已有的元数据,根据音响的物理属性对乐器音域、音高偏离度、乐器声压级、动态范围、乐器音色、音区划分、演奏技法的初步可视化做了初级的数据分析,这样既可以保证数据的可复制性,也可以使一些音乐学学者根据一些分析的基础数据做进一步的音乐学的研究。^⑪

1.音域和偏离度分析

由于中国民族乐器的多样性,每种乐器音域各不相同,考虑到乐器在乐队中与其他声部乐器进行配合使用为方便作曲家创作,对乐器的音域进行了测试。客观测试并参考演奏员的意见,选择乐器可获得良好的音色作为最后的音域范围,超出这个范围的音,使演奏员感觉演奏起来比较困难,发音不稳定。此外由于中国民族乐器的律制及演奏方法的多样化,还测量了与十二平均律中各音的音高偏离

度。这一定程度上也体现了民族乐器的性能以及演奏过程中的偏差,是显示中国民族乐器与西方乐器最重要的不同之一。以本次测试所用A调梆笛为例。如表3所示,此样本音阶与十二平均律相比较,个别音高偏离度较大,总体略有偏高趋势。其中,最大向上偏离47音分(A7);最大向下偏离35音分(A7),可能是由于在超高音区,演奏者不能有效控制音高所致。

例1 梆笛音阶



2.音响声压级及动态范围分析

乐器的表现力与乐器的音强及动态范围有较大关系,对于音强的分析涉及乐器声压级及动态范围的分析,数据库中要求测量绝对声压级,并对动态范围进行计算。这种对于音强的测量还有益于数据日后的复制。各乐器的声压级对音色有一定影响,未来对于乐器音响数据的再现可以以目前的声压级的

表3 梆笛各音的基音频率列表

序号	音名	基音频率(Hz)	偏离音分(cent)	序号	音名	基音频率(Hz)	偏离音分(cent)
1	E5	664.55	E5+13	10	[#] G6	1652.88	[#] G6-9
2	[#] F5	747.24	[#] F5+16	11	A6	1775.93	A6+15
3	[#] G5	825.48	[#] G5-11	12	B6	1986.52	B6+9
4	A5	888.1	A5+15	13	[#] C7	2241.97	[#] C7+19
5	B5	1000.54	B5+22	14	[#] D7	2464.19	[#] D7-18
6	[#] C6	1134.17	[#] C6+39	15	E7	2695.48	E7+37
7	D6	1193.64	D6+27	16	[#] F7	2986.26	[#] F7+15
8	E6	1335.22	E6+12	17	A7	3450.47	A7-35
9	[#] F6	1504.35	[#] F6+28	18	A7	3618.59	A7+47

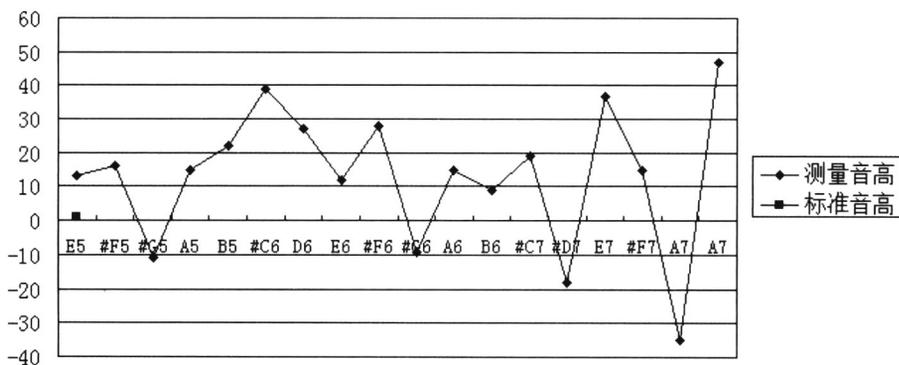


图2 测量音高与标准音高对比

量化为标准,再现或补充数据库中的数据,更有利于科学研究。以E调唢呐为例:

计算每个单音的动态范围,即强奏声压级P强和弱奏声压级P弱相减,两者差值 ΔP 为乐器各音的动态范围。根据上述数据,得出E调唢呐的动态范围图。

从表4和图3可以看出,E调唢呐强奏中声压级最强的位置位于常规音域的最低音B4的109.5dB,最弱的位置位于中音区B5的100.6dB。从折线图可知,唢呐强奏声压级随音的升高而渐次减弱,以中音区B5为转折点,继续升高时声压级则由渐次增强。E调唢呐弱奏中声压级最强的位置同样位于常规音域的最低音B4的106.1dB,最弱音位于中音区A5的91.6dB。出现显著变化的是在低音区出现一个仅次于A5的弱音 $\sharp C5$ (91.8dB)。其他声压级走势均与强奏时相类似。E调唢呐中,动态范围超过10dB的只有

低音区的 $\sharp C5$ 和中音区的A5,而超过7dB的5个单音也均位于中低音区。

3. 音色分析

音色是评价乐器的重要指标,对于乐器音响的频谱分析便于总结其特征,用于MIR中的基于音色的研究内容。音乐分析包括频谱分析、分音时间过程分析、特殊演奏技法分析等,用于说明每件乐器在不同音区的音色特征以及不同演奏技法的腔韵。频谱以及分音在时域上的特征是决定音色的重要因素,它可以通过表现泛音数量、泛音之间的音程关系以及泛音之间的强度关系来体现乐器声音特征。^⑩虽然不能单纯根据频谱对音色做出全面的且同主观感受完全一致的结论,但频谱分析至少可以对声音的属性做出大体的判断。频谱分析能提供量值性的信息,作为验证某种推断以及实施这种推断所采取

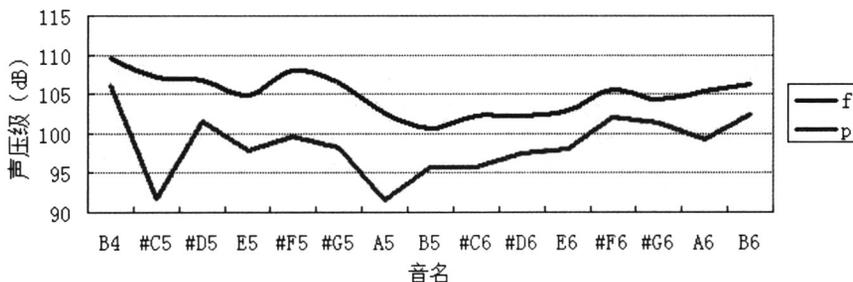


图3 E调唢呐动态范围图

表4 E调唢呐散音各音声压级表

序号	音名	弱奏声压级(dB)	强奏声压级(dB)	动态范围(dB)
1	B4	106.1	109.5	3.4
2	$\sharp C5$	91.8	107.1	15.3
3	$\sharp D5$	101.5	106.7	5.2
4	E5	97.8	104.8	7
5	$\sharp F5$	99.6	108	8.4
6	$\sharp G5$	98.3	106.4	8.1
7	A5	91.6	102.6	11
8	B5	95.7	100.6	4.9
9	$\sharp C6$	95.8	102.3	6.5
10	$\sharp D6$	97.6	102.2	4.6
11	E6	98.1	103	4.9
12	$\sharp F6$	102	105.5	3.5
13	$\sharp G6$	101.3	104.4	3.1
14	A6	99.2	105.3	6.1
15	B6	102.5	106.2	3.7

的技术性措施是否正确的依据。^⑬以梆笛音阶样本为例,从最低音开始每个八度取样E、A两音。梆笛共取样本6个,其中包含了最低音和最高音,基本保证了低音、中音、高区取样的平均性。乐谱中标记为“()”表示取样音。频谱图4所示,谐音数量较少,E5音频谱的谐音中,第一谐音(基音)峰值最高,第三谐音峰值为次高,第二到六号谐音均与第一号谐音构成简单频率比,因此音响特征较为纯净,有部分气流声。

例2 梆笛音阶样本

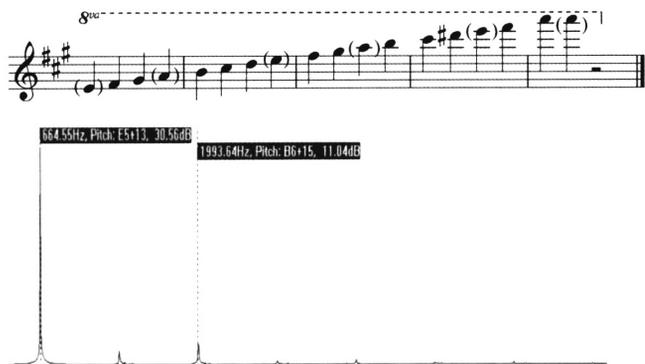


图4 梆笛E5音频谱图

4. 特殊技法分析

我国大部分的民族乐器,在民间俗乐中具有重

要地位,民间演奏家对于乐器技巧的发挥大大提高了乐器的表现力,但同时演奏人员往往对各种演奏技法在音高、音量、音色等音响声学参量的瞬间变化方面存在一定的误区,使其演奏技法缺乏标准化的认定。对于技法的研究以唢呐为例展开说明,唢呐的演奏技法十分丰富,包括双吐音、三吐音、花舌音、颤音、滑音等运用音乐声学、语音学等领域的测量方法进行音高、音量变化的对比分析,可以修正和补充多年以来我们对于唢呐演奏技法过程中存在的一些不全面的理解和认识。

颤音的演奏在唢呐中较为普遍,在吸足气的情况下,用小腹之力有节律,有弹性的颤动,借助小腹颤动对于腹中气流的挤压力以达到气流有规律的强弱变化,当气流进入口腔后,造成哨片有规律的振动,从而形成较为均匀的下波音效果。

图5为运用PRAAT语音学软件对A调唢呐演奏的小字一组E音的颤音样本生成的波形图。上部的波形图体现波浪形均匀分布的振幅效果,下部的蓝色旋律走势线和黄色音量线同样随之呈波浪形进行。即颤音振幅变高,音高升高,音量加大;颤音振幅变低,音高和音量也随之降低,即呈正比关系。由此可知,颤音的音高与波形图对应,呈波浪形循环均

表5 E5谐音列列表

谐音列序号	频率(Hz)/音组/音强(dB)	谐音列序号	频率(Hz)/音组/音强(dB)
第一谐音	664.55/E5+13/30.56	第五谐音	3328.45/G7+3/-2.69
第二谐音	1329.09/E6+13/7.14	第六谐音	3993.31/B7+15/-10.55
第三谐音	1993.64/B6+15/11.04	第七谐音	4657.16/D8-16/-7.3
第四谐音	2664.76/E7+16/-4.82		

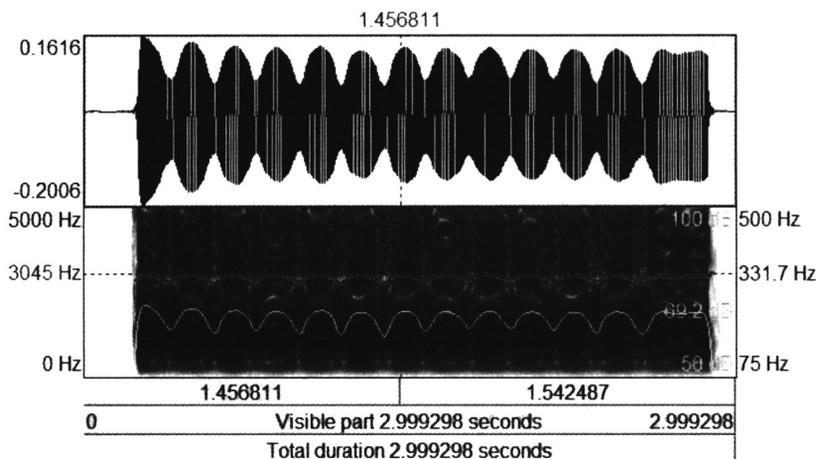


图5 唢呐颤音演奏波形图

匀变化,波峰频率为331.7Hz,接近小字一组E,波谷为317.7Hz,接近小字一组D,波峰音量为69.2dB,波谷音量为62.5dB。

此外还有滑音、双吐音、苦音等技法的可视化与分析的体现。

三、数据库建设过程中的问题

在数据库技术发展史上,检索技术和维护方法是两大永恒讨论主题。随着传统乐器数据库建设的不断完善,数据库音响数据规模的不断增加,数据库的应用和维护就出现了两个问题。第一,是检索方式单一造成数据库的使用不便。目前大部分的乐器数据库只有依据霍萨体系(将乐器分为弦鸣乐器、气鸣乐器、体鸣乐器、膜鸣乐器、电鸣乐器)以及地域分布等基于文本的乐器检索方式,需要用户提供准确的乐器名称。当使用者只有音响素材而对于所要搜寻的乐器名称、种类或使用区域不太了解时,就很难进行有效的搜索,局限性很大。第二,数据规模扩大后造成的维护不便,主要体现在对于标签的整理占据了大量的人力和时间资源。在数据库的建设中最为重要的任务就是为每个音响样本标注信息。待标注的标签内容可能包括流派、乐器、音色、声部、风格、民族、乐队、音阶、音色特征等无法穷尽的描述,具体选择哪些标签需依赖于具体应用场景确定^④。这些标注对于数据库数据的深入挖掘具有重要意义,如乐器演奏辅助教学、音乐音响分析、智能乐器识别、乐器辅助音质判断、辅助作曲等。但是人工标注效率低,费时费力,希望可以利用人工智能深度学习的方法学习已有的标签样本,为音响样本进行AI自动标注,人工矫正,以满足日益丰富的乐器样本的标注工作。

因此,建设一个带有标注功能的大规模中国传统乐器音响数据库,提供信息丰富、数量充足的基础数据,已成为音乐学尤其是乐器学领域进行学术研究与相关应用开发的一个迫切需求。在已有数据库的基础上,借鉴MIR领域数据库的建设经验,完善各种标签数据及附属基于内容的检索功能,有利于深入挖掘乐器音响数据的作用,在弘扬传统文化,科技与艺术融合方面都具有重要意义。

四、将MIR和AI用于中国传统乐器

音响数据库的方法

中国传统乐器音响数据库是在音乐声学分析指

导下提取音频特征,通过运用人工智能AI和MIR技术,使原有传统乐器数据库具有自动标注(Automatic Labelling)和基于乐器音响的检索(Retrieval)等功能,使得构建和管理更加方便有效,为民族音乐作品创作、民族演奏技法量化分析、传统乐器音响识别、中国民族管弦乐器音响研究、中国传统乐器配器法、传统乐器改良等多个研究领域提供数据基础。数据库的建设主要包括以下几个方面:

(一)基本数据库的采集

采集乐器音响数据首先要先确定50件以上最具地域和民族代表性的入库乐器。由于中国传统乐器流派众多,差异较大,对于入库乐器需进行标准化界定,选择具有代表性且个性特征相对不明显的一类乐器作为标准。并对已选乐器进行录音采样。运用专业录音设备在安静环境中对演奏员在常规状态下的演奏进行录音,录音内容包括单音演奏、音阶演奏、旋律演奏等,为保证录音样本获取的科学性,应以不同力度、速度演奏多遍。^⑤

(二)数据的声学分析

数据分析是数据挖掘的基础。首先将录音样本输入电脑,运用MATLAB或者Sonic Visualiser等声学测量系统进行音高(包括乐器定音、各音的频率、与十二平均律的偏离度、音域、律制等)、响度(包括各音峰值的最强与最弱强度、各音区音响强度、动态范围等)、音长(包括乐音的起振、稳态持续、衰减、终止时间、乐器混响时间等)的声学测量工作,以获取可靠数据。其次,对音响样本进行频谱分析,分析在乐音中基音的位置、基音频率、基音响度、乐音中泛音的数量、泛音的分布状态、泛音与基音间的响度关系、泛音与泛音间的响度关系等。最后进行主观评价,运用音色主观评价标准语汇,对不同乐器所发乐音、相同乐器不同音区所发乐音进行音响特征的描述。^⑥

在对每个乐器音响进行音乐声学分析的基础上,再对数据进行测量及相关性分析,寻找并总结每个乐器音响在频谱上的规律,在此指导下有目的地设计提取音频特征,作为后续人工智能辅助研究内容的基础。

(三)数据库自动标注

除了搜集足够多的音响信号样本并分类存储拍照展示,最核心的工作是数据库标注(Dataset Label-

ling),即对数据库中的每个音响信号样本的各个方面进行人工或自动标记,产生一系列的文字标签(Label)并存储到一个关系数据库表格或Excel表格中。标注是数据库建设的一个有机组成成分。只有音响数据的数据库相当于没有进行修饰,或利用价值不高的元数据建设,而进行标注后的数据库的用途会延展很多。标注越全面越准确后续的用途就越多。标注的目的就是以丰富的标签数据为基础,通过对其深入挖掘进行后续研究,并在辅助教学、辅助创作、识别等场景获得应用。例如,根据情感(Emotion)标签选择悲伤的马头琴、艾捷克等各种传统乐器的特定音域,可以辅助音乐创作。根据乐器的音域(Range)标签、音色(Timbre)标签选择适合民族管弦乐的低音声部,可以辅助民族管弦乐配器法的教学和实践。利用标注的音响信号样本音色,可以辅助音乐声学教学中的音乐音响分析课程等。利用乐器名称标签和演奏风格(Performance Style)标签可搜索“悠扬浪漫的唢呐独奏”这些研究和应用都依赖于详细的标注信息来完成。

在目前的数据库建设中,通常采用人工标注的方式,耗时耗力效率低下。而且,为保证标注的准确性和全面性,必须雇佣具有一定音乐学基础的专家进行标注,这使得标注成本大大提高。基于人工智能AI技术会利用一定数量的专家标注的乐器音响数据来训练机器学习模型,之后即可使用基于该模型实现的软件系统进行大量自动标注,辅以人工纠错,从而提高标注的效率和精度并降低成本。

建立数据库自动标注系统具体可分为以下三个步骤:

1. 确定标注的内容。事实上,待标注的标签内容本身无法确定,可能包括音乐情感、歌手、乐手、流派、乐器、语言、音色、声部、场景、风格、年代、唱片公司、歌词主题、流行度、民族、乐队、词曲作者等无法穷尽的描述,一般需按照具体情况进行约束。本项目针对中国传统民族乐器库的潜在应用场景,初步确定音响信号样本待标注的标签包括传统乐器名称、年代、民族、单音音名、音域、音色、演奏风格、演奏技巧、音乐情感等信息,并在项目执行过程中逐步补充完善。

2. 挑选少量音乐素养高的标注人员,建立人工标注的标准,并对一定数量的传统乐器音响信号样

本进行人工标注。以这些人工标注的标签和在音乐声学指导下提取的音响信号音频特征(Audio Feature)组成训练数据对的集合(音频特征,人工标签),然后输入机器学习模型学习音频特征与标签之间的复杂映射关系。常用的机器学习模型包括高斯混合模型[GMM(Gaussian Mixture Model)],支持向量机[SVM(Support Vector Machine)],及近年来流行的一系列深度学习模型如深度神经网络[DNN(Deep Neural Network)]、递归神经网络[RNN(Recurrent Neural Network)]等。虽然深度学习具有从音响信号中自动进行特征学习(Automatic Feature Learning)的能力,但这种黑匣子似的方式仍显盲目,预期不如在专家音乐声学知识指导下有目的提取的音频特征。这些具体效果均需在实验过程中比较验证。

3. 采用基于上述训练好的机器学习模型的软件系统,输入未标注乐器音响信号的音频特征,即可输出自动预测的标签。此类系统基于数据驱动方式运行,训练数据越多越全面,自动标注的结果便越准确。按照人工智能技术在AI医疗、AI交通、AI下棋等领域取得的成就预期,自动标注的结果通常可以达到具有较高音乐素质的专业人员的标注水平。若存在少量错误,可由专家级别标注人员人工纠错。与纯人工方式相比,会大大减少所花费的时间、精力、费用,同时取得不错的结果。

(四)基于内容的乐器音响检索

乐器音响数据库建立后,由于规模较大,已无法用逐个线性聆听的方式查找用户想要的音响数据。一个常用功能是建立搜索引擎,对存储标签数据的关系数据库表进行基于SQL(Structured Query Language)的结构化查询。此种方式基于文本,要求准确地输入乐器音响名称的全部或部分文字。在有错别字甚至忘记名字的时候,将无法查询。

与此不同,MIR技术提供了基于音频内容的乐器音响检索方式。在使用者无法准确写出具体的乐器名称的时候,仅需提供一段音响(比如5秒左右),即可对数据库中所有音响样本根据其音频特征进行匹配查询,并返回相似度最高的结果。在数据量特别大的时候,还可以采用局部敏感哈希[LSH(Local Sensitive Hashing)]技术加速查询。基于内容的检索为基于文字的检索提供多样化的补充,大大提高了数据库的检索效率。

五、信息技术应用于数据库的意义

从机械、电子、数字、互联网一直到今天的AI技术,人类的每一次进步都离不开科学的发展,科学技术的每一次进步都对各个学科的发展形成翻天覆地的革新,近年来,大数据、人工智能等信息技术的发展对音乐学也具有重要意义。从MIR到自动作曲,从消费端音乐到古典音乐无不渗透着信息技术对音乐的理论、创作、表演、教育带来的重大改变,作为音乐研究的基础的数据库建设更不例外。同时,人工智能技术高度发展的今天,更多的以专业知识内容为背景的数据成为了重要的基础,二者相互促进发展。

(一)对于数据库本身的意义

数据库建设是个长期不断完善的过程,随着数据量的增加,需要持续不断的努力、维护、改进与迭代。利用人工智能方法辅助数据库的建设既为采集的乐器音响数据提供全面丰富的信息描述,又为后续艺术领域工作者需要的乐器辅助教学、音质评价、辅助作曲、智能乐器识别与检索等实际日常应用提供可以深入挖掘的基础数据。对于数据的标注和音响的检索这两项重要的研究也有利于扩展到其他音乐数据库中,针对音乐数据库自身特点方便建设和使用。

(二)学科发展的意义

从学科角度来讲,音乐声学研究音乐信号的底层偏物理科学知识,而绝大多数人文艺术领域工作者需要的是与乐器相关的各种日常应用,这就需要跨艺术与科技领域,打破学科界限,进行跨领域的知识融合。结合音乐声学理论、音乐信息检索MIR技术、人工智能AI技术,在全面分析各种传统乐器音响特征的基础上,建立起可用于学术研究与应用开发及应用开发的、带有音响检索及自动标注功能的中国传统乐器音响数据库,为中国传统乐器改良、音乐创作、辅助教学等提供数据基础,并积极推进乐器学学科与信息科技的相互结合,共同发展。

结语

本文总结了中国传统乐器音响数据库的完备性、标准化、多功能性等特点,从数据库构建的工作实践中发现其可为音乐信息学提供帮助的巨大潜力,

并应期望利用音乐信息检索MIR以及人工智能[AI]等信息技术进行基于音响内容的检索和自动标注,使数据库的管理和使用更加便捷,提升数据库的效率。同时探索了数据库如何结合信息技术的方法,推动音乐技术的进步,以期对音乐类的其他数据库建设有借鉴作用,进一步推动我国民族音乐研究发展。

注释:

①<http://www.cmimo.org/about>.

②<http://www.mimo-international.com/MIMO/>.

③陈琴:《文化部科研项目“中国民族乐器音响标准库”项目研究通过结项验收》,《演艺科技》2011年第11期,第72页。

④李子晋:《“民族低音拉弦乐器改良”课题综述》,《演艺科技》2017年第5期。

⑤李伟、李子晋、高永伟:《“理解数字音乐——音乐信息检索技术综述”,第五届全国声音与音乐技术会议(CSMT)特约报告》,《复旦学报(自然科学版)》2018年第2期,第1-50页。

⑥<https://staff.aist.go.jp/m.goto/RWC-MDB/>.

⑦李子晋、于帅、肖畅、耿瑜曼、钱文琪、高永伟、李伟:《CCMusic:用于MIR研究的中国音乐数据库建设》,《复旦学报(自然科学版)》2019年第3期,第351-357页。

⑧参见缪天瑞《律学》(第三次修订版),人民音乐出版社,1996年。

⑨参见缪天瑞主编《音乐百科词典》“标准音高”条目,人民音乐出版社,1998年。

⑩付晓东:《〈中国民族乐器音响标准库〉项目的研制标准》,《演艺科技》2012年第3期,第48-51页。

⑪引用示例均来源于“中国传统乐器声学测量和频谱分析”“中国民族乐器多媒体数据库”课题实验报告。

⑫韩宝强:《现代音乐声学导论——音的历程》,中国文联出版社,2003年,第108页。

⑬William M. Hartmann: Principles of Musical Acoustics. New York: Springer-Verlag New York Inc, 2013.

⑭J. C. Wang, Y. H. Yang, H. M. Wang and S. K. Jeng, “The acoustic emotion Gaussians model for emotionbased music annotation and retrieval”, ACM International Conference on Multimedia(ACM MM), 2012, pp. 89-98.

⑮韩宝强:《“关于“音”的性质的讨论”》,《中国音乐学》2002年第3期,第27-36页。

⑯李子晋:《民族低音拉弦乐器改良中音质评价实验方法研究》,《中国音乐》2018年第3期,第157-163页。