第六届
全国声音与音乐技术会议

6th Conference on
Sound and Music Technology

大会手册
Conference Manual

2018 年 11 月
中国 · 厦门

Nov 2018
Xiamen, China

目录

第一部分　会议介绍

声音与音乐计算是一个多学科交叉领域。在科技方面涉及声学、信号处理、计算机科学、电子工程、数学等学科，在音乐方面则涉及作曲、音乐制作与设计、音响工程、声音设计、装置设计等等。由于各种原因，我国的相关研究较为分散，不同领域的学术交流也相对较少，这直接制约了该领域的进一步发展。

2013 年，由复旦大学、清华大学主办的第一届全国声音与音乐计算研讨会（CSMCW 2013）在复旦大学召开。国内众多重点高校包括浙江大学、上海交通大学、天津大学以及中国科学院声学所等相关机构的计算机、通信、声学学科的研究者参加这次会议。

2014 年，由清华大学承办，并与复旦大学、上海音乐学院、上海市计算机音乐协会联合主办的第二届会议在清华大学召开。在继续广泛邀请各学科各高校研究者的基础上，增加百度公司、杜比公司等企业的研究者加入，进一步增加了各界人士的交流。我们也欣喜的看到，在两次研讨会召开之后，与会机构已经逐步开始在科研、学术、教学上互相合作。

2015 年第三届研讨会由上海音乐学院承办。此次研讨会与 2015 年国际电子音乐周同期合并举行，众多的国内外音乐家、科学家齐聚上海，包括美国斯坦福大学音乐与声学计算机研究中心（CCRMA）、加州大学伯克利分校新音乐与音频技术中心（CNMAT）、法国里昂国立音乐创作中心（GRAME）等。他们在此期间以工作坊及音乐会的形式展示他们的研究成果，国内主要音乐学院的学者也带来各自的成果和作品。

2016 年第四届研讨会更名为声音与音乐技术会议（Conference on Sound and Music Technology, CSMT2016），会议由南京邮电大学承办。此次会议我们邀请了更广泛领域的包括各大高校、音乐学院、科技企业、研究所等不同领域的专家、学者、同学等参与。自此次会议开始，大会开始征集论文并发表。

2017 年第五届会议由苏州大学与 UCLA 苏州先进技术研究院联合承办，同时作为第 18 届国际音乐信息检索会议（ISMIR 2017）的卫星会议。来自包括清华大学、北京大学、复旦大学、上海交通大学、浙江大学、天津大学、哈尔滨工业大学、吉林大学、中国音乐学院、上海音乐学院、上海电影学院、中国科学院、英国伦敦大学玛丽女王学院数字音乐中心（C4DM, QMUL）、英国萨里大学视频语音信号处理中心（CVSSP, UoS）、西班牙庞贝发布拉大学音乐技术研究组（MTG, UPF）等数十所国内外高校和研究机构的教授和学生，以及百度、腾讯、阿里巴巴、杜比、酷狗、唱吧等多家公司的企业研究者参加了会议。

为了更好地促进我国声音与音乐技术领域的学术与技术交流，为全球在相关领域工作的华人学生学者及相关企业创造交流机会，开拓国内同行视野，建立国内相关领域与国际研究人员的联系，第 6 届中国声音与音乐技术会议将于 2018 年 11 月 24 日至 11 月 26 日在厦门理工学院，信息中心学术报告厅召开。

第二部分　第六届全国声音与音乐技术会议组织委员会名单

大会主席

孟卫东	厦门理工学院

大会共同主席

李　伟	复旦大学
陈强斌	上海音乐学院
朱瑞元	厦门理工学院

组委会共同主席

宁佐良	上海计算机音乐协会
康长河	厦门理工学院

顾问委员会主席

蔡莲红	清华大学
王 晔	新加坡国立大学
韩宝强	中国音乐学院

程序委员会共同主席

邵 曦	南京邮电大学
李圣辰	北京邮电大学
李子晋	中国音乐学院
肖仲喆	苏州大学

课程讲座及会议展示共同主席

施正珊	斯坦福大学
张克俊	浙江大学
朱梦尧	上海大学

赞助委员会

施正珊	斯坦福大学

网站及系统管理员

张添一	上海大学
陈 珂	复旦大学

面向普通音频计算机听觉（特殊议程）共同主席

李圣辰	北京邮电大学
唐　刚	北京化工大学

中国传统音乐技术研讨会（特殊议程）共同主席

李荣锋	北京邮电大学
李子晋	中国音乐学院
龚　嵘	庞贝法布拉大学（西班牙）

大会会刊编辑

李　伟	复旦大学
李圣辰	北京邮电大学
邵　曦	南京邮电大学
李子晋	中国音乐学院

大会学生协调委员会共同主席

张　暄	南京邮电大学
刘威良	北京邮电大学
高彬航	厦门理工学院
刘益雨	厦门理工学院

审稿人

Baoqiang Han	Beici Liang	Chen Qiao
Deshun Yang	Emmanouil Chourdakis	Emmanouil Benetos
Gang Tang	Gus Xia	Gyorgy Fazekas
Haifeng Li	Haojun Ai	Jiajie Dai
Jordan Smith	Kejun Zhang	Ken O'Hanlon
Li Zhou	Lingyun Xie	Meijun Liu
Mengyao Zhu	Mingxing Xu	Qichao Han
Qiuqiang Kong	Rong Gong	Rongfeng Li
Ru Zhang	Shengchen Li	Tao Jiang
Wei Li	Weicong Li	Wenlin Ban
Wenwu Wang	Xi Shao	Xiao Hu
Xiaodong Li	Xiaolin Hu	Xiaoou Chen
Xiaorui Wang	Yong Xu	Yujing Guan
Yuping Ren	Zhiyao Duan	Zhiyong Cheng
Zhongzhe Xiao	Shijia Zhu	Zijin Li
Yinbing Cheng	Shenglong Han	Ning Chen
Qinglin Meng	Jianfen Ma	Minwei Gu
Ming Li	Xianfeng Tang	Zhengshan Shi
Yufeng Hao	Yulong Wan	Bilei Zhu
Qiyong Lu	Ya Wang	

会议志愿者

古　典	汪梦晗	孙琦英	盛夏钰清
彭佳丽	李依纯	吴泓锦	黄锦煌
徐明喆	周　溢	陈可辉	高俊杰
刘少燕	陈嘉悦	陈　佳	张澜卉雅
吴晨瑜	董文慧	刘益雨	曾杰昌
陈诗崟	许　嘉	高彬航	吴东洋
房　邱	肖一帆	陈潇婧	傅胜楠
王博为	杨怀泽		

第六届全国声音与音乐技术会议组织委员会对辛勤付出的组织委员会成员和大会志愿者们表示真诚地感谢！

第三部分　欢迎辞

主办方欢迎辞

尊敬的各位领导、专家：

您们好！

“厦庇五洲客、门纳万倾涛”，在这全国多处已滴水成冰，鹭岛厦门却依然金风玉露的时节，我们非常高兴地欢迎前来参加“2018 年全国声音与音乐技术会议”的各位朋友。此次全国声音与音乐技术会议在我校举办，是我校发展史上的一项盛事，也是对我校艺术与科技（音乐工程）专业办学成效的一次检验。我们深深地感到了各位专家对我校工作无比信任和殷切期望。在此，我们向各位领导和专家的莅临表示热烈的欢迎！

厦门理工学院是伴随厦门经济特区诞生而成立的一所省属公立本科大学，建校于 1981 年。现有集美、思明、厦软三个校区，占地面积 1927 亩，建筑面积 50 余万平方米；教学科研仪器设备总值 5.2 亿元，馆藏纸质图书 190.7 万余册，电子图书 190.6 万册。全日制在校生 20081 人（含研究生 401 人，留学生 137 人）；专任教师 1116 人，高级职称的教师占比 51.1%。本科专业 58 个，涵盖理、工、经、管、文、艺、法等学科门类，并有 2 个一级学科硕士学位授权点和 1 个专业硕士学位授权点。

目前，我校正处于发展建设的关键时刻，各位领导和专家的光临指导，必将给我校的发展建设带来不可估量的推动力量。“一城如花半倚石，万点青山拥海来”，希望大家与会之余也可在这座海上花园城市观光徜徉。最后，祝各位领导、专家厦门之行吉祥平安、万事如意！

厦门理工学院

2018 年 11 月 22 日

组委会欢迎辞

各位尊敬的与会代表：

又逢一年冬月，一年一度的全国声音与音乐技术会议（以下简称“大会”）如约而至。在此，第六届全国声音与音乐技术会议组织委员会（以下简称“大会组委会”）隆重欢迎各位与会代表莅临大会现场，共襄盛举。

今年是大会走过的第六个年头。六年以来，在各届大会组委会的辛勤耕耘之下，大会从无到有，由小到大，逐步走上了正轨，成为一项在国内业界颇具影响力的盛会。随着会议参与人数的增加与会议组织委员会的逐步壮大，大会的各项组织工作逐步完善，已经基本具备了一项严谨学术会议的各个要素。截至本年度，会议参与人数稳步增加，在全球业界内，全国声音与音乐技术会议的参与人数也已名列前茅。

大会所取得的成功，离不开各位计算机音乐业界前辈的努力奋斗，他们披荆斩棘，克服种种困难，使得大会茁壮成长，国内计算机音乐业界得以发展壮大。令整个业界欣喜的是，近年来，计算机音乐的青年人才已崭露头角，在业界占有了一席之地。在这些青年才俊的带动下，越来越多的年轻人了解到了计算机音乐这一领域的存在，并投身于计算机音乐领域的建设之中。大会的发展壮大，一方面起到了播种机的作用，吸引更多的新生力量投身我国计算机音乐行业的建设之中；另一方面也成为了业界发展的晴雨表：大会参与度的提升，成为了我国计算机音乐业界发展的直接体现。

尽管大会乃至全国计算机音乐业界的发展拥有一片光明的前途，但是发展的路途注定坎坷，金玉之下有隐忧：作为新兴学科，学科从业人员的工作尚不被国内主流学术评价体系所接纳，极易打击从业人员的积极性；业界从业人员虽逐年增长，就总体数量而言仍然偏少，学界与企业界从业人员的平衡脆弱，不易保持；作为一个高度跨学科的领域，音乐从业人员与技术从业人员的融合工作仍然差强人意。作为全国计算机音乐行业的标杆，大会任重而道远：除需努力提升自身影响，带领计算机音乐业界在国内继续蓬勃发展，成为企业界与学界的沟通桥梁，还要警惕相近学科中学界与

企业界界限过于模糊所带来的种种不良后果。除此以外，作为国内相关领域为数不多的学术会议，大会还要紧追国际学界潮流，扶植如普通音频事件检测等一些更为新兴的近邻学科，不可谓不是任重而道远。

我们坚信，在广大计算机音乐同行的努力下，全国声音与音乐技术会议会越办越好，影响力会越来越大。我们也坚信，国内计算机音乐业界的明天会更加辉煌灿烂。

全国声音与音乐技术会议组织委员会

2018 年 11 月 22 日

第四部分　会议日程

11 月 23 日星期五

全天报到
地点：厦门杏林湾大酒店

11 月 24 日星期六

地点：厦门理工大学信息中心学术报告厅

培训讲座 主持人：施正珊	
08:30–09:30	音乐插件中的“中间件”
09:45–10:45	音乐学习对语言加工的影响
11:00–12:00	音乐推荐系统的现状与未来
12:00–13:00	午餐
13:00–14:00	声音技术在专业音乐教学与科研中的应用
14:15–15:15	深度学习简介
15:30–16:30	深度学习思考及听觉错觉
16:30–18:00	腾讯音乐技术分享

11 月 25 日星期日

地点：厦门理工大学信息中心学术报告厅
（具体安排请见下页表格）

08:30–09:40	开幕式
09:40–10:30	**主旨演讲** 虚拟人里的语音技术 主讲人：俞栋 主持人：李伟
10:30–10:50	茶歇
第一议程：音乐处理 主持人：杨德顺	
10:50–11:10	A Novel Singer Identification using GMM-UBM
11:10–11:30	Multimodal Music Emotion Recognition using Unsupervised Deep Neural Networks
11:30–11:50	A Practical Singing Voice Detection System based on GRU-RNN
11:50–12:10	Using Multiple Impulse Responses In 5.1 ch Production Work
12:10–12:30	Music Summary Detection with Feature Embedding
12:30–13:30	午餐
13:30–14:30	声音与音乐技术发展研讨会及颁奖礼
14:30–15:00	选举 2020 年中国声音与音乐技术大会举办地
第二（特别）议程：中国传统音乐技术研讨会 主持人：李荣锋	
15:00–15:20	古琴艺术数字化保护概述与琴律智能分析
15:20–15:40	古琴历史录音数字修复研究
15:40–16:00	Constructing a Multimedia Chinese Musical Instruments Database
16:00–16:20	CCMusic: 用于 MIR 研究的中国音乐数据库建设
16:20–16:40	茶歇
第三议程：音乐隐写术 主持人：陈宁	
16:40–17:00	一种基于弦乐配器的音乐隐写方法
17:00–17:20	A Standard MIDI File Steganography based on Music Perception
第四议程：字典学习 主持人：邵曦	
17:20–17:40	An Adaptive Consistent Dictionary Learning for Audio Declipping
17:40–18:00	Speech Enhancement Combining Nonnegative Dictionary Training and Robust Principal Component Analysis
18:30–20:00	会议晚宴（18:00 从会议场馆出发返回酒店）

11 月 26 日星期一

地点：厦门理工大学信息中心学术报告厅

第五（特别）议程：面向普通音频的机器听觉 主持人：侯丽敏	
09:00–09:50	**主旨演讲** Deep Learning for Audio Scene Classification, Event Detection and Audio Tagging 主讲人：王文武
09:50–10:10	茶歇
10:10–10:50	理解数字声音-基于普通音频的计算机听觉综述
10:50–11:10	Bird Sound Detection Based on Binarized Convolutional Neural Networks
11:10–11:30	A Comparison of Attention Mechanisms of Convolutional Neural Network in Weakly Labelled Audio Tagging
11:30–11:50	基于 MCKD 与 CEEMDAN 的声信号故障特征提取方法
11:50–12:10	基于人工神经网络的鼾声相关信号分类
12:10–12:30	Data Augmentation based Convolutional Neural Network for Auscultation
12:30–13:30	午餐
第六议程：计算机音乐生成 主持人：周莉	
13:30–14:20	**主旨演讲** 情境自动作曲经验谈 主讲人：黄志方
14:20–14:40	茶歇
14:40–15:00	基于动态规划的自适应和弦编配算法研究
15:00–15:20	基于最小二乘法和高斯混合模型的语音转音乐算法
15:30	会议参观

第五部分　会议演讲及摘要

5.1　培训讲座

音乐插件中的“中间件”

11 月 24 日星期六　08:30–09:30

主讲人：吴洲（星海音乐学院）

众所周知，目前市面上用于音乐制作的插件大致分两种：音色插件和效果器插件，比如一个长笛音色后面挂一个混响效果，这就完成了一个插件的常规衔接，有没有可能在这两种插件中间加入用于其它需求或效果处理的“中间件”呢？本文主要讨论的就是对音乐插件的“中间件”制作的一些想法。

音乐学习对语言加工的影响

11 月 24 日星期六　09:45–10:45

主讲人：南云（北京师范大学）

音乐与语言依赖于共同声音加工机制，音乐学习对声音表征的加工更为精细，因而促进了相关的语音表征的加工。

音乐推荐系统的现状与未来

11 月 24 日星期六　11:00–12:00

主讲人：顾旻玮（腾讯音乐）

自从 2000 年 pandora 开始 music genome project 以来，音乐中的智能数据理解和个性化推荐就在不断发展演进。本讲座将带你回顾这十几年来的音乐推荐领域的一些代表技术，并提出目前国内推荐应用中的问题和挑战与大家一起思考。

声音技术在专业音乐教学与科研中的应用

11 月 24 日星期六 13:00–14:00

主讲人：杨健（上海音乐学院）

演讲者将从教师、家长、学者与开发者等多重角度结合实例，对声音技术在音乐专业领域中的应用理念进行剖析。

深度学习简介

11 月 24 日星期六 14:15–15:15

主讲人：施正珊（斯坦福大学）

深度学习（Deep Learning）是一种结构化神经网络学习模型，试图从原始数据构建高级抽象概念。本课程将从零开始介绍深度学习的基础理论，及其在音乐信息检索中的应用。

深度学习思考及听觉错觉

11 月 24 日星期六 15:30–16:30

主讲人：张军平（复旦大学）

深度学习形成了端到端的革命，使得对行业知识的依赖显著下降。我将介绍深度学习进展及在图像检索、步态识别等的应用，以及其在视听觉错觉方面存在的不足和思考。

5.2 主旨演讲

虚拟人里的语音技术

11 月 25 日星期日 09:40–10:30

主讲人：俞栋（腾讯西雅图人工智能实验室）

我们认为下一代的人机交互界面将是多模态的，而虚拟人是多模态交互的一个很重要的载体。在这个演讲中，我将介绍理想虚拟人的一些重要特性和功能、以及为达到这些功能我们在语音技术上的一些探索。

Bio: Dr. Dong Yu is a distinguished scientist and vice general manager at Tencent AI Lab, an IEEE Fellow and an ACM Distinguished Scientist. Prior to joining Tencent in 2017, he was a principal researcher at Microsoft Research, where he joined in 1998. His research has been focusing on speech recognition and other applications of machine learning techniques with two monographs and 170+ papers. His works have been cited for over 20,000 times per Google Scholar and have been recognized by the prestigious IEEE Signal Processing Society 2013 and 2016 best paper award. Dr. Dong Yu currently is serving as

a member of the IEEE Speech and Language Processing Technical Committee (2013-2018) and a distinguished lecturer of APSIPA (2017-2018), and will serve as a technical co-chair of ICASSP 2021. He has served as an associate editor of the IEEE/ACM transactions on audio, speech, and language processing (2011-2015), an associate editor of the IEEE signal processing magazine (2008-2011), and members of organization and technical committees of many conferences and workshops.

Presenter' s website: https://sites.google.com/site/dongyu888/

Deep Learning for Audio Scene Classification, Event Detection and Audio Tagging

11 月 26 日星期一 09:00–09:50

主讲人：王文武（萨里大学）

Audio scene classification, event detection and audio tagging have attracted increasing interest recently, with a variety of potential applications in security surveillance, and acoustic sensing for smart homes and cities. This talk will present some recent and new development for several challenges related to this topic, including data challenges (e.g. DCASE 2016-2018), acoustic modelling, feature learning, and dealing with weakly labelled data using deep learning techniques. We will show some latest results including the results of our proposed algorithms and some benchmark methods. We will also use some sound demos to show the potential applications of these algorithms.

Biography: Wenwu Wang is a Reader in Signal Processing within the Centre for Vision Speech and Signal Processing, University of Surrey, Guildford, U.K.. His current research interests include blind signal processing, sparse signal processing, audio-visual signal processing, machine learning and perception, machine audition (listening), and statistical anomaly detection. He has (co)-authored over 200 publications in these areas. Dr. Wang is an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He is a Member of the Ministry of Defence University Defence Research Collaboration in Signal Processing (since 2009), a Member of the BBC Audio Research Partnership (since 2011), and a Member of the BBC Data Science Research Partnership (since 2017). He was a Tutorial Speaker for ICASSP 2013 and UDRC Summer School 2014, 2015, 2016, 2017, and SpaRTan/MacSeNet Spring School 2016, and London Intelligent Sensing Summer School 2017.

For more details, check his webpage: http://personal.ee.surrey.ac.uk/Personal/W.Wang/

情境自动作曲经验谈

11 月 26 日星期一　13:30–14:20

主讲人：黄志方（台湾清华大学）

基于情境的算法音乐方法是我基于音乐情感映射技术自动生成音乐的想法。视频，图像，文本等可以与音乐心理学相关联，并且算法组合可以用于与从情感到音乐的关系整合。本研究分析了多媒体作品的音乐创作方法。其他一些方法也适用于中国歌词之生成。在未来，我们将继续致力于 AI 歌词和算法音乐作品之间的整合，以生成适当情况的歌曲。最终，该研究表明基于情境的算法音乐方法最终成功地产生了所需之音乐。

主讲人简介：电子音乐作曲家与音乐科技研究人员，同时具有工程硕博士与作曲硕士雙學位，擅长音乐科技跨领域之研究与创作。师事吴丁连教授，电子音乐向 *Prof. Phil Winsor* 学习。作品曾获选受邀于欧、美、日、亚洲、中国大陆、中南美洲等国际音乐节演出，作品在国内外获奖与演出，如台湾省交响乐团作曲比赛佳作、台湾艺术教育馆之全国艺术创意作品在线竞赛音乐创意类比赛「特优」等。*2004* 年获选 *ACL* 亚洲作曲家联盟大会暨音乐节以色列演出；*2004* 年「澎湖乡土音乐创作甄选暨作品发表」以所创作之三乐章钢琴组曲「菊岛风情」获得「首奖」，并担任钢琴独奏演出。*2006* 年在台北国家音乐厅实验剧场发表数件室内乐创作并指挥。作品曾获选受邀于美国 *ICMC* 国际电子音乐会议、美国北德州大学 *CEMI* 实验音乐与跨媒介中心、德国柏林 *Randspiele Zepernic 2011*、科隆 *Alte Feuerwache 2011*、意大利 *Musica Insieme Panicale 2012* 现代音乐节作品、联合国 *UNESCO* 国际作曲评议会发表等。现为开南大学资传系专任副教授，国立交通大学声音与音乐创意科技硕士学程与国立清华大学音乐系兼任副教授。任桃园新爱乐管弦乐团指挥，并荣获 *2012* 年纽约 *OMI* 国际艺术家驻村 *Fellow*。曾获选 *2013* 年捷克国际暑期指挥大师班受教于 *Kirk Trevor* 大师与 *Donald Schleicher* 教授，指挥马替奴爱乐管弦乐团演出 *Debussy, Brahms*…等经典曲目，*2015* 年获邀担任美国大迈阿密青年管弦乐团客席指挥颇受好评，*2017-2018* 年担任 *ICMC* 音乐、论文评委。共计期刊论文 *30*、研讨会论文 *90* 篇以上、专书 *4* 册；其中 *SCI: 9* 篇、*SSCI: 2* 篇、*A&HCI: 1* 篇、*EI: 6* 篇、*THCI: 3* 篇，为极少数涵跨设计艺术、科技、人文领域在 *SCI/SSCI/A&HCA* 均有发表的跨领域研究学者。重要跨领域电声、互动、管弦乐等作品创作展演超过 *100* 场次，个人电声 *CD* 专辑发行。

5.3 第一议程：音乐处理
11 月 25 日周日 10:50–12:30

A Novel Singer Identification Using GMM-UBM

Zhang Xulong, Jiang Yiliang, Deng Jin, Li Juanjuan, Tian Mi, Li Wei

This paper presents a novel method for singer identification from polyphonic music audio signals. It is based on the Universal background model (UBM) which is a singer-independent Gaussian mixture model (GMM) trained on a large number of songs to model the singer characteristics. For our model, singing voice separation on polyphonic signal is used to cope with the negative influences caused by background accompaniment. Then we construct UBM for each singer trained with the Mel-frequency Cepstral Coefficients (MFCCs) feature using the Maximum a posterior (MAP) estimation. Singer indentation is realized by matching test samples to obtained UBMs for individual singers. Another major contribution of our work is to present two new large singer identification databases with over 100 singers. The proposed system is evaluated on two public datasets and the two new ones. Results indicate that UBM can build more accurate statistical models of the singer voice than conventional methods. Evaluation carried out on the public dataset shows that our method achieves 16% improvement of accuracy compared with the state-of-the-art singer identification system.

Multimodal Music Emotion Recognition Using Unsupervised Deep Neural Networks

Zhou Jianchao, Chen Xiaoou, Yang Deshun

In most studies on multimodal music emotion recognition, different modalities are generally combined in a simple way and used for supervised training. The improvement of the experiment results illustrates the correlations between different modalities. However, few studies focus on modeling the relationships between different modal data in music emotion recognition field. In this paper, we are interested in modeling the relationships between different modalities (i.e., lyric and audio data) by deep learning method. Several deep networks are first applied to perform unsupervised feature learning over multiple modalities. We then design a series of music emotion recognition experiments to evaluate the learned features. The experiment results show that the deep networks perform well on unsupervised feature learning for multimodal data and can model the relationships effectively. In addition, we demonstrate an unimodal enhancement experiment, where

better features for one modality (e.g., lyric) can be learned by the proposed deep network if the other modality (e.g., audio) is also present at unsupervised feature learning time.

A Practical Singing Voice Detection System Based on GRU-RNN

Chen Zhigao, Zhang Xulong, Deng Jin, Li Juanjuan, Jiang Yiliang, Li Wei

In this paper, we present a practical three-step approach for Singing Voice Detection based on a Gated Recurrent Unit (GRU) Recurrent Neural Network (RNN) and it achieves comparable results to state-of-the-art method. We combine four classic features, namely Mel Frequency Cepstral Coefficients (MFCC), Mel-filter Bank, Linear Predictive Cepstral Coefficients (LPCC) and Chroma, then the mixed signal is first preprocessed by Singing Voice Separation (SVS) with the Deep U-Net Convolutional Networks. Long Short-Term Memory (LSTM) and GRU are both proposed to solve the Gradient Vanish problem in RNN. In our experiments, we set the block duration as 120ms and 720ms respectively, and we get comparable or better results than state-of-the-art, while results on Jamendo are not as good as which on RWC-Pop.

Using Multiple Impulse Responses In 5.1 ch Production Work

Li Yingzi

Impulse responses have been used regularly across myriad production works to simulate the room acoustic. In order to achieve the room acoustic, only the impulse responses from one certain position of that location are typically used. However, when sound is generated in different places, the actual impulse responses of each spots are different. This research will investigate whether multiple impulse responses should be used in 5.1 ch production works.

Music Summary Detection with Feature Embedding

Gao Yongwei

Automatic music summary detection is a task that identifies the most representative part in a song, facilitating users to retrieve the songs they want. In this paper, we propose a novel method based on state space embedding and recurrence plot. Firstly, an extended audio feature with state space embedding, instead of raw audio features used by majority of music summary detection methods, is extracted to construct similarity matrix. Compared with the raw features, this extended feature is more robust against noise. Then recurrence plot based on global strategy is adopted to detect similar segment

pairs within a song. Finally, the most repeated part as summary of the processed song will be extracted by two principles we proposed. Experimental results show that the performance of our posed algorithm is more powerful than the other two competitive baseline methods.

5.4 第二（特别）议程：中国传统音乐技术研讨会 11 月 25 日周日 15:00–16:20

古琴艺术数字化保护概述与琴律智能分析

陈根方，黄晓东，张建国

古琴艺术是我国最古老的音乐艺术之一，作为世界非物质文化遗产，它在我国艺术史上具有不可替代的历史地位。古琴的演奏是三维空间的立体运动，古琴的乐谱是由独有的减字谱记录演奏技法。在数字信息时代，古琴艺术的跨媒体传播面临新的挑战，本文从记谱法、信息检索、琴律、音乐标注、风格分析、智能打谱和算法作曲等等二十四个方面来简述了古琴艺术的数字化保护内容。通过对古琴三分损益律与纯律两种琴律的建模，利用最近邻法和 K-Means 聚类算法，参照十二平均律的音律，对古琴的泛音和按音的音律进行分类，实验结果表明，最近邻法和 K-Means 聚类算法对按音和泛音分类结果相同。这也预示着人工智能在音律研究领域具有更广泛的应用前景。

古琴历史录音数字修复研究

王芳

本文是对以开盘磁带录音载体为主的古琴历史音响进行降噪、修复、还原为主要目标的音频技术研究。研究对象为我国上世纪应用磁质开盘带采录或转录，并遗存下来的的琴曲音响档案，这些琴曲为我国著名古琴演奏家弹奏或弹唱。研究内容为古琴历史录音修复方法。

Constructing a Multimedia Chinese Musical Instruments Database

Liang Xiaojing, Li Zhijin, Liu Jingyu, Li Wei, Zhu Jiaxing, Han Baoqiang

Throughout the history, more than 2,000 Chinese musical instruments have existed or been recorded, they are of non-negligible importance in Chinese musicology. However, the public knows little about them. In this work, we present a multimedia database of Chinese musical instruments. This database includes, for each instrument, text descriptions, images, audio clips of playing techniques, music clips, videos of craft process and

recording process, and acoustic analysis materials. Motivation and selecting criteria of the database are introduced in detail. Potential applications based on this database are discussed, and we take the research on subjective auditory attributes of Chinese musical instruments as an example.

CCMusic: 用于 MIR 研究的中国音乐数据库建设

李子晋，于帅，肖畅，耿余曼，钱文琪，高永伟，李伟

随着计算机技术的发展，大量的音乐资源需要被检索、分类、理解及分析。数据库是音乐信息检索研究的基础，丰富的数据库能够提高音乐信息检索领域算法的准确性，对于算法改进具有重要的意义。本文提出了一个新的音乐数据库——CCMusic Database。该数据库对录音环境、录音设备以及录音人员、流程等方面进行专业的限定。数据库将歌声与伴奏分离，对音乐信息检索的研究有重要的意义。本数据库搜集流行音乐、民族音乐及数百种民族乐器的音响素材，并进行全面的标注，构成一个 MIR 领域研究者使用的多用途的音乐数据库。数据库由音乐学院专业学生录制，版权清晰，并方便大规模扩展。

5.5 第三议程：音乐隐写术
11 月 25 日周日 16:40–17:20

一种基于弦乐配器的音乐隐写方法

朱照华，王健宗，肖京

音乐隐写术是一种由古老的音乐加密术发展而来的一种信息安全技术，其目的在于以一种不可感知的方式将秘密信息隐藏在形如旋律，伴奏等音乐内容当中。音乐隐写术在近些年来的发展较为迅速，但是就目前而言，尚未有在音乐配器当中进行隐藏的相关算法。基于这个现状，本文提出了一种基于弦乐配器的音乐隐写方法，首先我们通过音乐中每个音符的时值和节拍力度计算出音符的表现张力，并以此选择可以用于嵌入秘密信息的音符位置；然后在嵌入过程中，我们以音乐小节为单位，通过秘密信息调制其中不同类别音符的比例。此外，本文通过盲听测试对模型进行相关的评价，理论分析和实验表明本文的方法可以在保证嵌入透明性的同时实现不错的隐藏容量和很好的安全性。

A Standard MIDI File Steganography Based on Music Perception

Guan Lei, Jing Yinji, Li Shengchen

This paper proposes a Steganography method that embeds information in MIDI files based on music perception. As human auditory system has a threshold of note duration,

the lengths of notes can be changed under the threshold to code information. This method sits between the covered and coverless Steganography methods by synthesizing a revised version of original media to code information, which introduces no expansion of file size but has efficient capacity. Working under music perception principles, the result of listening tests reveals that the perceptual difference between Stego-media and original media is not perceivable in general. The proposed method gives an example that how the principle of musicology is used in Steganography which introduces a new way to code information.

5.6 第四议程：字典学习
11 月 25 日周日 17:20–18:00

An adaptive consistent Dictionary Learning for Audio Declipping

Wu Penglong, Zou Xia, Sun Meng, Li Li, Zhang Xingyu

Clipping is a common problem in audio processing. Clipping distortion can be solved by the recently proposed consistent Dictionary Learning (cDL), but the performance of restoration will decrease when the clipping degree is large. In order to solve this problem, a method based on adaptive threshold is proposed. The method automatically estimates the clipping degree, and the factor of the clipping degree is adjusted in the algorithm according to the degree of clipping. Experiments show the superior performance of the proposed algorithm with respect to cDL on audio signal restoration.

Speech Enhancement Combining Nonnegative Dictionary Training and Robust Principal Component Analysis

Xia Zou

In this paper, an unsupervised single channel speech enhancement algorithm is proposed. It combines both the nonnegative dictionary training and robust principal component analysis (RPCA) so that we name it as NRPCA in short. The combination accomplishes by incorporating the nonnegative speech dictionary into the RPCA model, which can be learned via nonnegative matrix factorization (NMF). With the NRPCA model, the method of alternating direction method of multipliers (ADMM) is applied for optimized solutions. Objective evaluations using perceptual evaluation of speech quality (PESQ) on TIMIT with 20 noise types at various signal-to-noise ratio (SNR) levels demonstrate that the proposed NRPCA model yields superior results over the conventional NMF and RPCA methods.

5.7 第五（特别）议程：面向普通音频的机器听觉
11 月 26 日周一 09:00–12:30

理解数字声音-基于普通音频的计算机听觉综述

李伟

声音是人类获取信息的重要来源，对声音内容进行自动分析和理解，具有重要意义。本文介绍声音的基本知识，从信号、听觉感受、声音特性等三个角度对声音进行分类，阐明各个分类之间的关系，明确基于普通音频的计算机听觉技术的研究对象和学科位置。之后，介绍计算机听觉技术的基本概念，原理，研究课题，和技术框架。作者全面总结了计算机听觉技术在各个领域中的典型应用，包括音频场景/声景分类识别，声音事件检测，医学领域，安全监控领域，交通领域，航空领域，声品质领域，机械工程领域，军事领域，管道传输领域，电力领域，容器领域，农业领域，林业领域，日常生活领域，身份识别领域，机器人领域，植物领域，养殖及加工领域，生态领域，江河海洋领域，冶金领域，材料领域，矿业、气象、地震等其它应用领域。分类总结了各领域计算机听觉应用中现有典型文献的基本原理、技术路线。最后总结计算机听觉领域存在的各方面问题，并展望未来发展趋势。

Bird Sound Detection Based on Binarized Convolutional Neural Networks

Song Jia'nan, Li Shengchen

Bird Sound Detection (BSD) is helpful for monitoring biodiversity and deep learning has shown good performance in BSD in recent years [1]. But the expensive calculations and resources of the complex network structure make it difficult to implement the hardware of BSD. Therefore, we design an audio classification method for BSD using Binarized Convolutional Neural Networks (BCNNs). The convolutional layers and fully connected layers of the original Convolutional Neural Network are binarized to two values. This paper has designed two networks (CNNs and BCNNs) for the BSD task of the IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events (DCASE2018). The Area Under ROC Curve (AUC) score of BCNN has achieved comparable results with CNN on the unseen evaluation data. More importantly, BCNN can reduce the memory requirement and the hardware loss unit which are of great significance to the hardware implementation of the bird sound detection system.

A Comparison of Attention Mechanisms of Convolutional Neural Network in Weakly Labelled Audio Tagging

Hou Yuanbo, Kong Qiuqiang, Li Shengchen

Audio tagging aims to predict the types of sound events occurring in the audio clips. Recently, Convolutional Recurrent Neural Network (CRNN) achieves the state-of-the-art performance in audio tagging. In CRNN, convolutional layers are applied on the input audio features to extract high level representations followed by recurrent layers. To better learn high level representations of acoustic feature, attention mechanisms were introduced to the convolutional layers of CRNN. Attention is a learning technique that could steer the model to information important to the task to obtain better performance. There are two attention mechanisms in CRNN: Squeeze-and-Excitation (SE) block and Gated Linear Unit (GLU), both of them are based on gating mechanism, but their concerns are different. To compare the performance of the two different attention mechanisms, SE block and GLU, we propose to use CRNN with SE block (SE-CRNN) and CRNN with GLU (GLU-CRNN) in weakly labelled audio tagging and compare these results with the CRNN baseline. The experiments show that the GLU-CRNN achieves an Area Under Curve score of 0.877 in polyphonic audio tagging outperforming the SE-CRNN of 0.865 and the CRNN baseline of 0.838. That is, the attention based on GLU is better than the attention based on SE block in CRNN for weakly labelled polyphonic audio tagging.

基于 MCKD 与 CEEMDAN 的声信号故障特征提取方法

申博文，宋浏阳，唐刚，王华庆

在双转子轴承状态监测与故障诊断中，信号传递路径复杂，很难通过加速度传感器直接获得信号，而声音信号有非接触式测量的优势，包含大量特征信息。为了能够准确、有效地通过声音信号实现滚动轴承故障诊断，检测出轴承故障，提出了基于自适应噪声完备集合经验模态分解 (CEEMDAN) 和最大相关峭度反褶积 (MCKD) 的滚动轴承故障诊断方法。首先运用最大相关峭度反褶积方法增强轴承故障声音信号中的冲击，然后对处理后信号进行 CEEMDAN 处理，计算每个经验模态分量的峭度值，根据峭度值选取最优分量并求 Hilbert 包络谱，以准确提取故障特征频率。文中采用该方法基于声音信号实现滚动轴承故障诊断，为提取最优分量提供了理想筛选标准，一定程度上降低了故障诊断复杂程度，具有良好自适应性。

基于人工神经网络的鼾声相关信号分类

侯丽敏，刘焕成，施晓宇，张新鹏

本文提出了基于人工神经网络 (Artificial Neural Networks，ANN) 对鼾声、呼吸声和其它噪声分类的方法。提取每个声音片段的频谱相关特征集作为 ANN 的输入特征，用小批量训练以及 Adam 学习率自适应等策略加快了模型有效的训练过程，使用了丢弃法优化了 ANN 的结构，区分鼾声、呼吸声和其它噪声的正确率分别为 98.88%、97.36%、95.15%。

Data Augmentation based Convolutional Neural Network for Auscultation

Jiang Yiliang

Acoustic analysis has great potential for clinical application because of its objective, non-invasive and low-cost nature. Auscultation is an important part of traditional Chinese medicine (TCM). By analyzing a voice signal, we attempt to diagnose the syndrome of the subject by labelling them normal or deficient. In this paper, we explore a Data Augmentation based Convolutional Neural Network (DACNN) for auscultation. The idea behind this method is the use of CNN on imbalanced data with data augmentation for automatic feature extraction and classification. We conduct experiments on our auscultation dataset containing voice segments of 959 speakers (346 males and 613 females), which were labeled by two experienced TCM physicians. We demonstrate the effectiveness of data augmentation to overcome the imbalanced dataset problem. We also compare its performance with traditional machine learning methods. By using DACNN, we achieve 97.25% diagnosis accuracy for females and 95.12% diagnosis accuracy for males, with 1% 10% improvement in accuracy and slight improvements in other indicators over traditional machine learning methods. The experimental results demonstrate that the proposed approach is helpful for objective auscultation diagnosis.

5.8 第六议程：计算机音乐生成

11 月 26 日周一 13:30–15:20

基于动态规划的自适应和弦编配算法研究

邓阳，周莉，许多，岳诚成，游梦琪

和弦的编配是作曲过程中耗时较长的一个重要步骤，传统作曲中的和弦编配主要采用人工完成，尚无成熟的自动和弦编配技术。本文针对以上问题，根据和弦构成规律与进行逻辑，提出了 CFCS 和弦体系构造函数，设计出一种自动和弦编配的动态规划算法，以此

来实现机器自动和弦编配。通过对多样算例进行实验，检测结果验证本算法是有效可行的。本算法可用于人工智能作曲及现代算法作曲技术等音乐创作和音乐教学领域，为推进机器音乐创作提供了切实有效的技术参考。

基于最小二乘法和高斯混合模型的语音转音乐算法

段伟博

本文描述了一种基于语音的音乐合成系统。给定一段朗读的歌词语音文件和乐谱的信息，系统能够合成音乐。系统根据乐谱的信息，依照机器学习下获得的转换函数，自动调整说话人语音的基频、每个字的持续时间和频谱包络。根据歌声和普通语音声学特征的差异，我们选择最小二乘法去学习修改语音文件的基频，生成歌声带有波动的基频轮廓；利用高斯混合模型 (GMM) 学习音乐频谱和说话人频谱的映射关系，将说话声音的频谱包络转换为音乐带有特定共振峰的频谱包络；根据节拍信息来修改说话语音中每个字的时长。最终得到的三个参数合成音乐歌声，实现机器学习下的语音转音乐。实验结果表明，该系统能够将说话声音转换为较好的歌唱声音。

5.9 海报展示

11 月 25 日周日

基于变异字典的中国工尺谱即兴演奏研究

李荣锋，李学明，柳杨

工尺谱是现存数量最多的中国传统乐谱。今人解读工尺谱的最大难点，在于处理不确定的节奏型，即工尺谱只规定了节拍的起始位置，而未分配音符的具体时值。本文的研究对象是昆曲、京剧以及古乐器演奏中的工尺谱，重点研究工尺谱符号，包括音高、节拍、歌词、及其读音的音乐学和语言学的量化语义。本文针对演唱者的即兴演唱问题，将利用数据驱动方法，自动生成基于每一拍实际演唱音符的变异字典。希望能够通过本研究，将千百年来中国人在使用工尺谱中凝结的智慧，以数据的形式保存下来，并通过统计模型与计算机，更好地传承、传播和发扬中国传统文化。

A Framework for Automated Pop-song Melody Generation and Piano Accompaniment Arrangement

Wang Ziyu, Xia Gus

Most popular songs consist of three parts: a lead (vocal) melody, a chord progression, and an accompaniment. Automated music generation and automated arrangement, which

usually involve the generation of one or two of these three parts, are considered difficult tasks due to their complex interrelationship. In this study, we contribute a computational framework, which 1) reveals the complex dependencies between the three parts explicitly, 2) decomposes and simplify the accompaniment texture into a set of melody lines, and 3) generate all parts in a natural order that composers usually follow. Specifically, harmony alternation model, melody generation model, and melody integration model are proposed. To verify the framework, an automated composition and accompaniment arrangement system is created as a simple implementation, which takes a raw chord progression as the input and performs harmony optimization, multi-task melody generation, and polyphonic music integration.

基于虚拟现实模型的实时混响生成方案设计

Xueting Dong

在虚拟现实空间中，混响的真实程度是实现沉浸感的关键要素。根据建筑声学理论，一个空间的混响取决于容积、吸声材料等因素，并且当听者处于空间内不同位置时，由于直达声与混响声的比例不同，听感效果会发生改变。本文阐述了在虚拟现实模型中，根据建模和所选材料等因素实时生成混响的方案，根据听者的走动生成更真实的听感效果，为虚拟现实声音制作提供参考，同时为建声仿真技术在虚拟现实中的应用开辟思路。

Binaural Rendering based on Linear Differential Microphone Array and Ambisonic Reproduction

Dou Shanshan

Binaural recording is normally deployed on a human listening subject or on a dummy head with recording devices of a similar size to the human head. However, with popularity of augmented reality systems based on mobile devices, PDA and portable visualization devices, the problem is how to achieve binaural rendering using small-sized recording devices. In this work, we propose a binaural rendering system for augmented reality based on linear differential microphone array, which is compact and has small array structure. We combine differential beamforming and Ambisonic encoding for sound field decomposition, and then use the concept of virtual loudspeakers and Ambisonic decoding for rendering. From the experimental results, comparing to the standard binaural recording device Sabine panorama microphone, the proposed method achieves better results on sense of orientation and out-of-head experience.

Towards Detecting Short Segments in Music Borrowing

Gao Yongwei

Music borrowing such as medley and quotation has been ubiquitous from past to present. To detect the similar segments between two musical signals is very beneficial in music plagiarism detection, medley and quotation analysis. Such segments are generally short with locations and length unknown, which makes the task interesting yet very challenging. To the best of our knowledge, we are the first to tackle this problem. As the first attempt, we propose a new algorithm based on the transposition-invariant cross-similarity matrix and the improved sparse cross recurrence plot to detect short similar segments in music borrowing. The center-weighted extended feature, which is able to resist noise and inaccuracy in single frame, is adopted to construct the similarity matrix. In order to obtain more accurate position of the similar segments, we build a combined sparse cross recurrence plot by integrating the strength of globally and locally sparse threshold strategies. Experimental results demonstrate the effectiveness of our proposed method, particularly for shorter similar segments.

浙江大学 Next Lab：设计智能与创新创业

张克俊等

人工智与设计的进一步融合，带来了设计领域的不断发展。当前，设计智能中视觉与听觉的创意设计越来越得到重视，衍生了一些很有前景的创新创业成果。但相关研究和探索还处于起步阶段，也遇到了不少困难。我们团队一直在研究设计智能中的关键技术和方法，并也开始了若干创新创业探索。

数字音频处理技术

朱梦尧

上海大学通信与信息工程学院“数字视音频处理与多媒体传输研究方向”是教育部“新型显示技术与应用”重点实验室四大研究中心之一。数字音频研究内容得到上海市一流学科、高峰高原学科和高水平大学重点建设的支持。数字音频方向具有多位长期从事音频信号处理、阵列信号处理、模式识别等研究的教师。研究小组承担了数十项国家级、省部级课题与企业合作项目，发表论文百余篇。项目组近期主要开展了“复杂场景下双麦克风语音增强研究”、“球谐域全景音频关键技术研究”、“基于听觉与认知的复杂音频质量客观评价方法研究”、“3DTV 音频再现理论研究”、“球面阵列的多声源空间定位方法研究”等内容，并取得了较多的研究成果。项目小组将学术成果应用于解决企业的实际需求，开展了多项校企合作以及产学研转化项目。

第六部分　特别致谢

今年是全国声音与音乐技术会议走过的第六个年头，大会喜人的发展局面，离不开行业领头人的带领，更离不开业界初创阶段各位同事的辛勤耕耘与努力。在大会的崭露头角的几年间，历届主办方为大会的召开与组织作出了开创性的贡献，为整个业界的发展与开拓起到了领头羊的作用，大会在此对以下诸位同事致以最高的敬意和衷心地感谢：

李　伟	复旦大学
蔡莲红	清华大学
徐明星	清华大学
陈强斌	上海音乐学院
陈世哲	上海音乐学院
宁佐良	上海计算机音乐协会

学术质量是全国声音与音乐技术会议的立会之本，下列同事在历届会议中，为提升大会来稿的学术质量作出了卓越贡献，特颁发全国声音与音乐技术会议“学术贡献奖”：

邵 曦	南京邮电大学
肖仲喆	苏州大学
李圣辰	北京邮电大学

优良的赞助资源是全国声音与音乐技术会议的立会之源，下列同事在本届会议中，为大会赞助工作作出了卓越贡献，特颁发全国声音与音乐技术会议“赞助贡献奖”：

施正珊	斯坦福大学
段淑菲	太原理工大学

良好的宣传渠道是全国声音与音乐技术会议扩大影响力的保证，下列同事在本届会议中，为大会宣传工作作出了卓越贡献，特颁发全国声音与音乐技术会议“宣传贡献奖”：

李子晋	中国音乐学院
张添一	上海大学

全国声音与音乐技术会议的宗旨之一是促进音乐学科与工程学科的交流与融合，下列同事在历届会议中，为学科间的融合工作作出了卓越贡献，特颁发“学科交流贡献奖”：

李子晋	中国音乐学院

大会对本届会议赞助商的大力支持表示感谢：

金牌赞助商

腾讯音乐娱乐

VIP陪练 PEILIAN·COM

银牌赞助商

唱吧

华睿宝 FREEBER.CN

铜牌赞助商

抖音 短视频APP

主办单位

上海计算机音乐协会

复旦大学

清华大学

上海音乐学院

南京邮电大学

苏州大学

厦门理工学院