

媒介音视频教育部重点实验室 2016 年年度报告与 2017 年工作要点

实验室名称：媒介音视频教育部重点实验室

所属领域：信息领域

学科分类：通信与信息系统

依托单位：中国传媒大学

主管部门：教育部

实验室负责人：张勤

实验室联系人：叶龙

联系电话：010-65783425

传 真：010-65783425

电子邮件：klmav@cuc.edu.cn

通讯地址：北京市朝阳区定福庄东街一号

邮政编码：100024

填报时间：2016 年 12 月

内容提纲

一、实验室名称, 主管部门, 依托单位	1
二、实验室基本情况	1
三、实验室主要开展的研究工作	5
四、2016 年度实验室年度进展情况	8
(一) 科研项目进展情况.....	8
(二) 实验室建设进展情况.....	19
(三) 对外合作与交流.....	20
(四) 人才培养与团队建设.....	21
(五) 本年度新增成果列表.....	21
(六) 2017 年度实验室开放课题指南.....	26
五. 实验室 2017 年度工作要点	27

一、实验室名称，主管部门，依托单位

实验室名称：媒介音视频教育部重点实验室

Key Laboratory of Media Audio & Video (Communication
University of China), Ministry of Education

主管部门：教育部

依托单位：中国传媒大学

二、实验室基本情况

中国传媒大学围绕现代视听媒介中的前沿科学问题和关键技术、结合学校在广播与电视领域的优势与特色，开展了媒介音视频这一应用基础学科的研究。以此学科的研究积累为基础，中国传媒大学以学校原有新创信息技术研究所、传播声学研究所和高性能计算中心这三个科研特区为基础，组建了媒介音视频实验室。实验室于2010年通过教育部建设计划论证成为教育部重点实验室，并聘任张勤教授为实验室主任，从而正式进入实验室建设期。

媒介音视频教育部重点实验室主任张勤教授1982年毕业于天津大学，获得电子工程系学士学位；1984年毕业于北京广播学院，获得无线电工程系硕士学位；1990年毕业于加拿大不列颠哥伦比亚大学（UBC），获得电机工程系博士学位。1990到1995年：任加拿大UBC大学图像处理实验室研究工程师和项目负责人；1995到1996年：任Wavetek公司DSP集团的首席技术官；1996到2000年：任美国摩托罗拉公司DNS前端工程部门经理和高级技术顾问；2000年至今：历任美国思科与时代华纳公司的JV北京华诺公司CTO，TCL集团首任工业研究院院长，中国传媒大学（原北京广播学院）教授，硕士生、博士生导师。2010年5月受聘成为媒介音视频教育部重点实验室主任，期望经过五年的建设，在规模上将实验室建成在国内最完善的、在国际上相比具有良好的运行条件的媒介音视频技术基础研究基地。

经过五年的建设发展，媒介音视频教育部重点实验室现有三个研究方向，分别为传播声学 with 媒介音频方向、视觉感知与媒介视频方向以及融合媒体与智

能网络方向。实验室现在编员工 56 人，其中 48 人具有博士学位；科研岗 52 人，工程实验技术岗 1 人，产业岗 3 人；具有正高级职称 26 人，副高级职称 13 人；入选教育部“新世纪优秀人才支持计划”4 人、江苏省“双创人才”1 人，江苏省“高层次创新创业人才引进计划”1 人，北京市“青年英才计划”6 人。

在科学研究方面，实验室五年来承担了包括 2 项国家自然科学基金重点项目、23 项国家自然科学基金面上及青年项目、1 项国家科技重大专项、11 项国家科技支撑计划课题、1 项国家“863”计划等各类纵向研究课题 90 项，纵向科研经费 3654.85 万元，同时实验室建设期间获得横向项目 48 项，横向科研经费 994.87 万元；在《IEEE Trans. Multimedia》、《ACM Trans. Multimedia》、《IET Signal Processing》、《IEEE Signal Processing Letters》、《Electronics Letters》、《Chinese Journal of Electronics》等国内外重要期刊和会议上共发表学术论文 221 篇，其中被 SCI 检索 63 篇、EI 检索 158 篇；获授权发明专利 16 项，软件著作权 64 项；出版专著（教材）5 部；培养博士研究生三十余人，硕士研究生百余人；获得中国电影电视技术学会科学技术一等奖 1 项，中国电影电视青年科技奖 1 项，北京市科学技术奖三等奖 1 项。媒介音视频教育部重点实验室支撑着通信与信息系统、电磁场与微波技术两个二级学科北京市重点学科以及电子科学与技术、信息与通信工程两个一级学科博士点的发展，同时也是中国传媒大学“数字媒体技术”985 优势学科建设平台与高等学校学科创新引智计划（“111”计划）“数字媒体工程”平台的依托建设基地。

实验室现有用房面积 3740 平方米，建成有“DMS 动声技术实验室”、“高性能计算中心”、“沉浸式通信系统研究中心”、“Media-MAX 技术平台演示室”和 700 平方米“全消声室”等实验室、研究中心和演示场地。设备总价值超过 4000 万元，具体落实情况如下：

媒介音视频实验室建设概算和设备购置情况表

序号	建设项目	投入资金（万元）	落实时间
1	媒介声学实验室设备配套	200	2011 年 6 月
2	音频信号处理设备	110	2010 年 12 月

序号	建设项目	投入资金(万元)	落实时间
3	电生理测量系统	120	2010年12月
4	听力测量与诊断设备	50	2011年12月
5	视觉测量与诊断设备	100	2011年6月
6	视频信号处理设备	300	2011年12月
7	图形、图像采集、处理设备	100	2010年12月
8	智能网络系统	150	2011年6月
9	媒介材料环境设备	120	2011年12月
10	实验室常规电子电气设备	50	2011年12月
11	DMS+3D/4K 音视频实验平台 和声场测试实验室	200	2013年
12	沉浸式3D虚拟现实实验系统	600	2014年
13	超高清超媒体系统实验环境	150	2013年
14	互动业务与云计算系统实验环境	200	2012年
15	大数据分析与应用系统实验环境	450	2012年
16	高性能计算中心及相关设施环境	450	2012年
17	感知网络实验环境	200	2013年
18	传输覆盖网络实验与测试环境	450	2012年
19	实验室改、扩建	300	2011年12月
20	实验室运行费用	300	分五年配套完成
合计		4600	

科研实验场地、实验仪器设备等硬件条件的建设为实验室研究人员提供了良好的科研工作平台，加之制订了与实验室工作相适应的科研管理制度，切实保障了实验室各项科研工作的顺利开展。

实验室面向国家文化传播战略需求与音视频处理领域的重大科学问题，经深入建设和不断凝练，形成了视觉感知与媒介视频、传播声学及媒介音频以及融合媒体与智能网络三个特色研究方向。

1. 视觉感知与媒介视频技术研究方向

- 1) 研究大众媒介中的视觉生理与视觉心理，并通过建立反映视频内容的视频信息统计模型，来探索新的视觉信息处理与表达理论，研究基于感知的视频编码理论，视频处理与计算机图形学的融合理论与方法。
- 2) 通过研究人体运动的视频分解与综合，立体视频的基础理论，立体电视广播的基础理论，为下一代视频内容信息处理技术提供基础理论和关键技术储备。
- 3) 以高性能计算研发团队研发的“基于高性能的服务云平台”为基础，并通过植入游戏开发、电影特效、动漫渲染所需的各类专业应用软件，进行文化产业的在线虚拟工作室公共服务平台方面的研究。

2. 传播声学与媒介音频技术研究方向

- 1) 通过传播语言声学、传播环境声学、虚拟声环境技术、音乐传播声学等方面的研究建设传播声学理论研究体系。
- 2) 通过声场的散射与绕射计算技术，录音声场中的智能化解析理论，收音声场的自适应综合理论，声源的近场与远场模型以及相关理论，基于贝叶斯的交响乐声场的分轨算法与合成控制理论等研究建设多声道声场重建系统。

3. 融合媒体与智能网络研究方向

- 1) 面向根据媒体内容特性、媒体终端的能力、用户特性、交互性、接入网络特性包括网络带宽和信道误码特性等等智能的随时随地提供最佳媒体信息服务的目标，研究自治、自管理和自适应的智能网络和体系结构，可针对不同性质网络、网络实时状况和应用进行自组优化，研究媒体网络服务的安全性和可靠性。
- 2) 同时面对媒体融合的大数据环境下将给广电行业带来巨大的变化，打造“广电大数据挖掘与智能决策支持系统”，从而打破单向产业链的劣势，将各环节的数据和需求汇总并通过数据挖掘和决策支持反向服务

于各个广电系统产业链环节，最终达到指导科学运营的作用。

三、实验室主要开展的研究工作

在学校的大力支持和国家、地方、企业科研项目共同发挥作用下，媒介音视频教育部重点实验室在五年建设期间里，通过科研实践实现技术创新，取得了一系列成果。

标志性成果一：高新数字影像技术研究及产业化应用平台

高新数字影像技术研究及产业化应用平台由实验室“视觉感知与媒介视频技术”研究方向的媒介视频技术团队研发完成，该平台通过研究大众媒介中的视觉生理与视觉心理建立反映视频内容的视频信息统计模型，来探索新的视觉信息处理与表达理论，研究基于感知的视频编码理论，视频处理与计算机图形学的融合理论与方法，人体运动的视频分解与综合，立体视频的基础理论，立体电视广播的基础理论，为下一代视频内容信息处理技术提供基础理论和关键技术储备。

标志性成果二：“创意云”综合性文化创意产业公共服务云平台

“创意云”综合性文化创意产业公共服务云平台由实验室“视觉感知与媒介视频”研究方向的高性能计算研究团队研发完成，该平台以高性能计算研究团队研发的“基于高性能的服务云平台”为基础，通过植入游戏开发、电影特效、动漫渲染所需的各类专业应用软件，如 3D Max、MAYA、Mental Ray、RenderMan、Ogre、Unity 3D 等，建设了一个文化产业的在线虚拟工作室公共服务平台。

标志性成果三：DMS (Dynamic Matrix Sound) 多声道声场重建系统

DMS 多声道声场重建系统由实验室“传播声学及媒介音频”研究方向的音频信号处理与声场技术研究团队完成，系统中涵盖的主要技术模块包括：声场的散射与绕射计算技术，录音声场中的智能化解析理论，放音声场的自适应综合理论，声源的近场与远场模型以及相关性理论，声场的模式识别与补偿的最佳

控制原理，大型分轨演奏的视听心理，基于贝叶斯的交响乐声场的分轨算法与合成控制理论等。

标志性成果四：传播声学理论研究体系建设

传播声学理论研究体系由实验室“传播声学 with 媒介音频”研究方向的传播声学研究团队建设，主要通过对于传播语言声学、传播环境声学、音乐传播声学、听觉心理学、虚拟声环境技术、音频测量与录音技术等几个重点问题的研究，构建了传播声学理论体系。

标志性成果五：智能媒体网络分发理论与系统

智能媒体网络分发理论与系统的研究由实验室“融合媒体与智能网络”研究方向的智能媒介网络研究团队完成，智能媒体网络融合现有的不同网络、多样化媒介资源，为用户提供智能的媒体服务，主要研究内容包括媒体内容传输和分发、媒体大数据等相关内容。

标志性成果六：广电大数据挖掘与智能决策支持系统

广电大数据挖掘与智能决策支持系统的研究由实验室“融合媒体与智能网络”研究方向的媒体大数据研究团队完成，系统建设的意义在于打破单向广电产业链的劣势，将各环节的数据和需求汇总至智能决策支持系统，通过数据挖掘和决策支持，反向服务于广电产业链的各个环节，最终达到科学运营的作用。

实验室产业化成果

实验室研发的“天河-酷卡”动漫与影视超级渲染云计算平台已达到世界一流水平。该平台成功将渲染应用与超级计算机系统有机结合起来，使得“天河一号”的公共服务辐射能力借助互联网得到极大拓展，同时有力增强了我国三维制作行业的实力和影视产业的核心竞争力，大大提升渲染作业的质量和速度，

成为三维制作行业发展的新平台和新旗帜。在三个月的测试运行过程中，创造了采用 2,000 个物理节点（24,000 个处理器）同时对一个模型进行并行渲染的世界纪录。截至目前，世界上还没有第二款渲染管理软件能在这么大规模的集群系统上进行有效的渲染管理。

DVB/IPTV 双模互动电视系统在关键支撑技术上取得突破，给出了低成本、高性能的宽带、双模互动电视系统解决方案，实现了高清视频点播、IP 电话、网页浏览、视频游戏以及电子商务等多种功能。该成果推动了数字电视产业的发展，为我国广播电视行业的发展提供了新的业务增长点，促进了广电业务与互联网业务的融合，催生了新的业态，推动了现代服务业产业结构的进一步优化。该成果获得 2010 年度北京市科学技术三等奖和国家广播电影电视总局 2010 年度科技创新一等奖，并已成功应用于保加利亚电信、美国 intercom、美国 fyreTV、墨西哥 cablmas 等国内外媒体结构，产生巨大的经济效益。

“创意云”平台是高性能计算、网格技术、并行存储、云计算等当前主流 IT 技术在文化产业创作应用上协同创新的重要成果。它以“软件即服务”的方式为文化产业个人及企业提供了一个在线创作的虚拟工作室。目前，“创意云”在全国范围内建设 3 个分中心、7 个二级分中心。其中，华东分中心主要以苏州数据中心为基础，是整个平台的中枢系统，负责管理、协调各个分中心的运行，同时也负责覆盖华东区业务；华北分中心以“天河一号 A”超级计算机为基础，负责覆盖京津冀和东北地区业务，并通过大连、烟台、青岛间接覆盖韩国、日本市场；华南分中心以“天河二号”超级计算机为基础，负责覆盖珠三角和港澳台业务，并辐射东南亚，通过香港作为中转间接覆盖北美市场。从 2014 年初“创意云”正式上线到 4 月底，平台的个人用户数已经突破 10 万，注册企业超过 1000 家，中小创作团队接近 10000 家。

面向下一代声音处理与重放的多通道 DMS 动声系统是实验室自主研发的，继杜比、DTS 之后具有国际领先水平的声音处理、重放系统。与现有的多声道还音系统相比，其在技术指标和功能上具有如下优势：主声道自适应，有效弱化

声场相干波；3次谐波失真与动态声道数量的对数成反比；支持各种声场播放模式，播放模式可任意调整。

四、2016年度实验室年度进展情况

（一）科研项目进展情况

1. 国家自然科学基金青年项目《基于关键帧匹配与双向蒙特卡罗粒子滤波器的视频运动重定向研究》结题验收

本课题的研究内容为：对于给定视频，提取其关键帧并通过仿射变换与线性叠加对由关键帧形成的视频时空体参数进行状态初始化并形成观测模型；在每个时空体内部，基于所形成观测模型的双向马尔科夫性，提出一种新型的双向马尔科夫链蒙特卡罗采样粒子滤波器结构，通过充分利用参数状态与观测模型的前后帧关系提高运动重定向的准确性；同时，针对高维采样的效率问题，根据定向参数的相关性实现骨架参数-形态参数-物理参数的分级，并提出分等级遗传优化策略以有效提高算法效率。因此，本课题的研究计划要点可以做如下归纳：

首先，根据视频序列中前后帧的相关性，结合粒子滤波器框架，研究一种双向粒子滤波形式；其次，针对关键帧—时空体—视频序列的层级视频构成关系，研究适用于描述视频信息的特征表示方法，并使之适用于粒子滤波过程中的粒子状态的权衡；最后，在视频重构、运动重定向等领域做相关算法的整合与应用验证。

本课题在执行过程中按照上述计划要点执行，针对视频数据中的信息提取与建模技术，特别是粒子滤波器这种非线性贝叶斯信息提取方式进行研究，内容涉及粒子滤波器架构、视频深度信息提取、视频运动信息提取提取、视频内容表征与特征建模等方面，同时就研究的核心技术内容在视频运动重定向、视频重建、立体视频采集中的参数自适应调整等方面进行了应用验证。

2. 国家自然科学基金面上项目《基于广义建模理论的多原子库图像编码方法研究》第三年度研究完成

本年度，我们从基于多模式的图像/视频编码理论出发，研究广义建模理论。同时基于广义建模理论，进行基于多模式表征的图像编码、基于存储的图像编码以及编码过程中的特征匹配技术三个层面的研究。

● 图像/视频的广义建模理论层面的研究

在本项目的研究中，我们定义了一种模式冗余的概念，即在采用多模式情形下进一步去除的图像冗余称为模式冗余。模式冗余的去除是以增加模式索引信息为代价的，因此，使用多模式进行图像编码的过程是一个平衡模式冗余的减少与索引信息增加的过程。我们通过定义图像的广义建模性以衡量多模式情形下图像的信息量，从而形成一种新的指导图像与视频编码的理论。在本年度的研究中，我们通过理论分析证明了模式冗余存在的条件，并由此形成了利用模式冗余的去除进行进一步数据压缩的理论基础。具体来说，假设我们对一个长度为 N 的 n 元字符串进行编码， n 个元素的各自概率分布为 p_1, p_2, \dots, p_n 。该字符串又可以分成 m 段， m 段的长度依次 N_1, N_2, \dots, N_m ， m 段数据中每段数据内 n 个元素的分布为 $p_{i1}, p_{i2}, \dots, p_{in}$ ， $i=1,2,3,\dots,m$ 。整段数据的熵记为 E ，第 i 分段数据的熵记为 E_i ，且 $E'=E_1+E_2+\dots+E_m$ 。模式冗余MR则可以形式化定义为 $MR=E-E'$ ，根据Jensen不等式，我们可以证明 $MR \geq 0$ ，且等号成立的条件是每一个分段中 n 个元素的 p_1, p_2, \dots, p_n 是同分布的。

$MR \geq 0$ 意味着模式冗余的冗余是存在的，这就为我们进行数据压缩提供了一个新的方向，也建立了利用模式冗余进行数据压缩的基础。因此，为验证广义建模理论的正确性与实用性，本项目从映射多模式（混合稀疏表示）、扫描多模式、预测多模式等多个角度进行通过模式冗余的去除进一步进行了数据压缩的研究。

● 基于混合稀疏表示的图像压缩

在本项目的研究中，我们以广义建模理论为基础，通过增加映射空间的模式数进行图像数据的进一步压缩。本部分项目研究的目标是首先设计一个包含不同类图像特征的过完备字典。我们用三个不同类型的图片库：FERET_BMP 图片库、

Brodatz texture 图片库 和 images-view 图片库,并将这三个图片库中的图片分成 8×8 的块,组成三个新的图片库,即 FERET, Brodat 和 view。每一类图片库中,我们使用 KSV D 训练出针对各图片库特性的字典,并合并成具有混合稀疏特性的稀疏表示模式。在图像压缩过程中,我们采用正交匹配追踪算法得到图像稀疏表示稀疏矩阵,最后进行 8 比特量化并进行赫夫曼编码实现图像压缩。

在本部分的实验中,以 Lena 图像为例,Lena 的脸部、图片的背景以及图片的特定纹理区域能够在混合字典库中选择与所训练图片库相一致的原子进行表征,从而实现图像压缩。对于各种不同类型的压缩结果表明本文算法效果优于单独只用一个图像库训练出来的字典。

- 基于扫描多模式的图像压缩

在将图像数据映射到稀疏表示空间之后,一般的,我们需要对于系数进行扫描从而压缩掉系数间存在的统计冗余。本项目在研究具有方向性纹理信息的图像的系数能量分布规律的基础上,结合广义建模理论,训练得到多种系数扫描模板,通过进一步消除统计模式冗余来达到更高的压缩效率。

系数扫描模板的设定分为两步,统计每个位置非零系数出现的次数以及降序排列。统计次数的过程中,首先将图像分为 8×8 块,对每一个图像块进行 DCT 变换和量化,对于量化后的系数位置统计该位置出现非零系数的次数。然后对于每一类图像,对非零系数出现次数进行降序排列,为新模板的设计提供依据。本项目的研究中,我们用图片“Lena”进行了本项目对于方向性纹理信息分类合理性的验证。从“Lena”图像中取一些 8×8 块放大观察其方向性纹理信息,然后用 JPEG 算法对这些块压缩编码,在扫描这一步不只是同 Zigzag 扫描,而是用训练得到的多个代表不同方向信息的扫描模板依次扫描然后编码,选择最后码流最短的模板作为这个块的最佳模板。实验结果表明方向性纹理信息的分类与得出的模板具有一致性。同时,在本项目的研究中,我们针对不同的测试图片进行了实验,结果表明本项目提出的方法比使用单一模板进行扫描有更好的压缩效果,码数的节省率可以达到 5%~10%。

- 基于多线索融合深度生成算法的立体视频转换技术研究

由传统平面视频转制的立体视频，在目前的立体电视节目市场中仍然扮演着重要的角色。在本项目的研究中，我们提出了基于多线索融合深度生成算法的立体视频转换系统。首先，我们选择三个强健的深度感知线索来对二维视频进行深度估计，分别是空间线索，运动线索，及光照线索。每一个深度线索都对最终深度合成起到特殊作用：空间线索利用人类视觉感知对场景景深信息做出初步规划；运动线索用来区分前景及背景目标；光照线索考虑了纹理及辐射信息，可以使合成的深度图更加细致和接近真实。然后，利用本文提出的多深度线索融合策略进行深度融合，得到深度图。最后利用DIBR方法合成3D序列。具体过程如下：1) 在转换之前，对2D视频序列进行镜头分割，以保证深度信息及轮廓等标签信息可以在每个镜头序列中复制和传播。2) 提取空间深度线索（Linear Perspective, LP），经由边缘检测寻找边际线（Vanishing Line）或边际点（Vanishing Point），然后线性填充得到深度信息。LP规划了场景中的基本深度信息。3) 通过自动提取或交互式方法得到运动线索（Motion Cue, MC），以区分前景及背景物体。自动提取方法通常对场景序列有着苛刻的要求，如摄像机运动限制，前景目标有明显的运动视差，及背景静止等。因此对不满足要求的场景，可采用交互的方法提取运动目标，以保证提取的准确性。4) 利用我们提出的基于暗原色先验信息的改进算法提取光照深度线索（Aerial Perspective Cue, AP），其考虑纹理及辐射特性，同时可以减小或消除过亮/过暗区域对深度图合成的影响。可以应用于有烟雾或气泡的特殊场景。5) 利用深度线索融合算法整合得到的三个深度线索，生成深度信息图。6) 最后，利用DIBR方法合成立体视频序列。

实验结果表明本项目研究所生成的深度图接近于真实的深度图，证明我们制定的深度融合策略是非常有效的，其很好地整合了三个强健的，符合人类视觉的深度线索——空间线索，运动线索，光照线索。但是，我们也发现在深度估计过程中，对部分静止物体出现了一些错误，比如Poznan Street图中左侧的汽车及右边的树木，主要是因为其静止不动而被当做背景来处理。由于观众的注意力往往集中在运动的物体上面，所以这些小的错误不会对观众的观影体验带来大的影响。

3. 国家自然科学基金重点项目《三位音频基础理论与关键技术研究》（武汉大学合作项目）进展顺利

本年度该项目在声场一致性、音乐人声分离、中文歌词识别、语音情感识别四个方面展开研究。

- 200-1000Hz扫频信号的声场一致性研究

本项目研究的重点问题之一为线扬声器阵列声压级一致性分布的设计与实现，我们首先实现了 200-2000Hz 单频点的声场一致性设计。但是单频点的设计并不能在实际中得到应用，我们按照不同频率的辐射特性，针对 200-1000Hz 划分了频率子带，在每个子带上考虑到了该子带内所有频率的辐射特性，并使用非均匀滤波器组在 200-1000Hz 实现了双扫频信号的声场一致性的目标，双扫频信号为 200Hz-500Hz 的扫频信号与 400Hz-1000Hz 的扫频信号的叠加信号。这样我们在声场一致性的设计中，既考虑到了每个频点的一致性的实现精度，又尽量降低了非目标区域的能量的泄漏，同时还兼顾了高频和低频的辐射特性，在整个频段内实现了声场的一致性分布。

以前的研究中，我们在 200-1000Hz 的频段内基于自适应遗传算法，使用相控扬声器阵列实现了单频点声压级一致性分布的设计。但由扬声器阵列的辐射特性使得宽频带的声压级一致性设计设计有一定的难度。

在本项目的研究中，我们首先根据扬声器阵列在不同频率的辐射特性将整个频段分为了很多非均匀的子带，接着进行每个频带内不同频率点的参数统一优化方法，又考虑到了扬声器阵列的频响特性，在扬声器阵列的幅度上进行了优化补偿。输入信号被分为多个子带，幅度相位等信息经过升采样和降采样以后合成出新的信号。最后确定了数字滤波器组的采样率后设计出来了一组非均匀数字滤波器组以实现了整个频带的划分与参数优化设计。最终结果展示了成功的频带划分与数字滤波器组的优化效果。

- 基于RPCA的歌声伴奏分离改进算法研究

在对 RPCA 算法的实验研究中发现，由于伴奏与歌声并不是严格符合低秩与稀疏特性，对于仅有伴奏而没有歌声的部分，RPCA 在进行矩阵分解时，依然

把这部分的稀疏成分归类到了代表歌声的稀疏矩阵当中。而在基于概率模型的技术中，由于提取了歌声的基音特征，纯伴奏部分分离效果更好。因此，本文借鉴了基于概率模型技术的方法，在原始的 RPCA 算法中引入了基音特征，利用 CASA 系统在此提取信号，使得算法在纯伴奏部分的分离效果得到了提升。

本项目的研究借鉴了计算机听觉场景分析原理。首先混合信号经过 STFT，变换成时频小单元，然后利用 RPCA 把混合信号分解成两个矩阵，一个稀疏矩阵和一个低秩矩阵。低秩矩阵代表了伴奏音乐，而稀疏矩阵代表了歌声。其中，对分离出来的稀疏矩阵反傅里叶变换后，再利用对其进行信号标记重组。信号标记和重组使用的是 CASA 的处理方法。CASA 处理方法是利用听觉外围处理将 RPCA 分离出来的稀疏部分分成时频小单元，并且对其进行特征的提取。提取的特征有时频小单元自相关，通道互相关，包络自相关。然后通过比较小单元的响应周期与主基频周期，标记小单元，形成片段。如果一个片段中超过一半的帧与主基频的周期一致，那么把这个片段归至目标流。目标流即为初步改进的歌声信号。由于初步改进的歌声信号中还引入了一些新的噪音，因此利用 RPCA 中生成的掩模再对其进行二元掩膜分离，得到最终的歌声信号。得到最终的歌声信号后，再利用其产生一个新的二元掩模，利用这个新的二元掩模最后将背景音乐信号分离出来。

我们对 50 个歌曲片段的 RPCA、CASA，改进算法(RPCA+CASA)的分离结果进行了客观评价，并将客观评价指标进行了均值计算。本文的客观评价使用的是 BSS_EVAL 工具箱，对分离结果进行三种指标的比较，分别是信号偏差比 (Source-to-Distorition Ration, SDR)，信号干扰比 (Source-to-Interference Ratio, SIR)，系统误差比 (Source-to-Artifacts Ratio, SAR)。从客观评价指标数据可以看出，RPCA 算法的 SDR 值与 SAR 值最高，改进算法的 SIR 值最高。

● 基于SVM带伴奏的孤立中文歌词识别研究

本文以中文网络歌词作为研究对象，简化识别模型，以字为识别单元，从大量的网络歌曲中截取出现频率最高的字，自建具有代表性的识别训练库。基于人耳的听觉效应，选取目前在语音识别领域应用最广泛的 39 维 MFCC 作为识别特

征以及性能较好的 SVM 分类器构建识别模块。结合最终的识别便签，从主观听觉上分析，分类正确率高的字具有如下的特点：

- ① 字的发音更为清楚。
- ② 字完整发音的条件下截取时长更为接近。
- ③ 背景音乐单一且音量较小，不会淹没人声。
- ④ 相较于对应的语音，唱的字在变调、拉伸压缩等方面变化较小。

4. 国家自然科学基金青年项目《基于波内频率调制的音色模型研究以及在单通道音源分离中的应用》完成第一年研究

非线性现象往往引起波内调频和波间调频。波间调频比较常见，它的信号的频率随着波动逐渐变化。而波内调频是 Hilbert-Huang 变换特有的一种发现。波内调频是指在一个信号周期内，频率发生改变的现象。这在以往的基于 Fourier 理论的分析中，都把它当作谐波失真对待，在 FFT 频谱中，往往丢失了波内调频的信息，而且产生了误导性的谐波。波内调频提供了对非线性振动系统的新的认识，它深刻的反映了非线性振动的频率结构细节。若频率在波内时刻变化，则其波形不再是简单的正弦函数或余弦函数。因此，对简单正弦波形的畸变，都暗含着波内频率调制。在过去，这种现象被当作谐波失真来处理，其实这种波形畸变视为波内频率调制更合适，它更具有物理意义。波内频率调制需要瞬时频率才能描述，因为对于传统的傅里叶变换，必须要求基函数为无限长的正弦信号。

在语音和音频处理领域，正弦模型是应用最广泛的参数信号模型之一，语音和多种乐器的声音信号可以被这种模型精确的描述。正弦模型是用正弦分量简单的线性叠加来拟合原始信号，因此它无法表达信号中的非平稳分量和非线性分量，而自然音频信号中非平稳和非线性分量是常见的信号结构，因此这种模型是需要改进的。

通过对各种乐器瞬时频率谱的观察，我们发现，乐器的音色部分的瞬时频率谱在一个基频振动周期内往往会振动多次。这种在一个周期内频率多次改变的现象被称为波内频率调制。由于四种乐器的音色部分的波内频率调制都是有固定的周期，这使得我们可以用周期性频率调制信号来模拟波内频率调制。鉴于正弦调频信号

的特性，我们可以把乐器的音色建模为正弦调频信号(SFM)的组合。在这个基础上，我们只需要在调频信号模型的基础上提取少量参数就可以有效的描述乐器的音色，这为一系列的应用提供了新的方法和视角。

利用频率调制的公式可以用简单的参数产生复杂的频谱。频率调制音色合成的数字实现技术最早由 John Chowning 于 1967 年实现。频谱成分的时间变化特性可以很简单的被控制，这是这项技术最吸引人的优点。对于合成谐波声音，调制信号必须与原始载频具有谐波关系。随着频率调制程度的提升，声音信号逐渐的变得复杂。

一个正弦信号被另外一个正弦信号调制产生的谐波分布符合 Bessel 函数，这是 FM 合成的数学基础。FM 合成是“失真合成”或“非线性合成”的一种形式。它首先用振荡器产生音频载波波形，频率为 f_c ，然后用一个频率为 f_m 的音频波形来对载波进行调制。如果调制器的幅度为 0，则输出的信号频率为载波频率 f_c ，否则，调制信号的幅度会造成载波振荡器的频率在 f_c 上下波动。这个频率波动幅度被称为频偏。在简单的情况下，调制信号的幅度越大，载波的频率变动就越大。

通过实验发现，这种单纯的正弦频率调制信号对于波内频率调制的模拟效果较差，产生的声音音色与原始声音差别较大。通过对瞬时频率谱的分析发现，这种差异主要是在瞬时频率谱的音色部分的非线性机制与正弦频率调制信号有一些区别。为了更加准确的模拟这种非线性机制，可以使用更加复杂的调制模型。

$$X(t) = \sum_{i=0}^N A_i(t) \sin[\omega_i t + \sum_{j=1}^M \phi_{ij} \sin(M \omega_0 t)] + R(t)$$

公式中的调制基频函数不再是单纯的正弦波，而是多个不同调制系数的一组正弦波的叠加，并且每个正弦波的频率都是基频的整数倍。通过实验发现，这种复合调频信号可以较好的再现乐器信号音色部分的非线性波内调频特征。一般来说，M 值为 2~3 时已经可以产生足够复杂的波形，通过调节各模型参数，可以产生跟原始声音较为接近的信号。

至此，我们得到了两个音色表达模型，简化的模型具有形式简单，易于提取参数的优点，但产生的声音音质较差。这种模型可以用于对数据量较重视，而对音质要求不高的应用中。如慢速网络音频传输等应用上。而第二个模型形式较为

复杂,参数提取困难,但能产生较好的音质,可以用在对音质要求较高的环境中。

5. 国家科技支撑计划课题《立体电视技术体系架构及内容直播技术与应用研究》已递交结题申请

在立体视频的会聚式拍摄中,摄像机光心位于同一水平线上,摄像机光轴各自绕其光心向内旋转一定角度并相交于一点,该点即是会聚点。会聚式拍摄所拍摄得到的视差图像公共场景区域较大,且同时存在正视差、零视差及负视差,在立体显示器上显示时,会呈现出整个拍摄场景部分出屏部分入屏的立体效果。如果我们在拍摄过程中对于拍摄场景有一个基本的理解,对三维场景中前景以及后景目标有一个深度定位,知道了前景和后景相对于摄像机的距离,就可以在双机会聚式拍摄时对左右摄像机轴间距和轴夹角做出了计算,得到了合适而不失精确性的拍摄参数,从而帮助我们在立体拍摄中能以此参数为参考进行对立体摄像机进行参数初调,大大简化立体拍摄前期的计算工作,以技术辅佐立体拍摄艺术的完善。有鉴于此,本项目中研究了一种基于摄像机标定与 SIFT 特征点匹配的场景理解方案,其涵盖的关键技术主要包括如下几个方面:

- 双目立体视觉系统中摄像机标定理论与方法。在本项目的研究中,我们利用成像关系建立约束方程,基于单应矩阵线性求解摄像机内、外参数,通过最大似然估计进一步优化,对左右摄像机的相对位姿进行线性求解,构造目标函数再优化求解,完成了对不同位置的摄像机之间的相对位姿的标定。通过镜头畸变校正优化畸变参数,从而使标定结果更为精确和接近物理真实值。
- 基于像素与 SIFT 描述子相结合的匹配算法。采用自适应阈值的序贯相似性检测算法,得到目标的初始匹配区域,之后再进一步进行 SIFT 特征提取与匹配。同时结合摄像机不同位姿,考虑不同实际情况,在三角测量法和射线交汇法中选择合适方法,对目标场景中物体的深度信息进行感知,完成目标定位。
- 一种基于双向蒙特卡罗采样粒子滤波技术的人体运动跟踪技术,即在定义了视频关键帧并进行关键帧匹配后,在由关键帧所形成的每个时空体内部

产生前后帧的约束关系，从而将动力学模型与观测模型耦合起来了，也就形成了基于关键帧匹配与双向马尔科夫链蒙特卡罗采样的视频时空体滤波，并在每个时空体内部形成对于目标运动位置的估计。

- 在根据场景理解技术得到所拍摄场景中目标的深度位置，并能够根据粒子滤波技术预测场景中目标的运动趋势的前提下，我们可以有效的进行立体视频拍摄参数的调整。在双机会聚式拍摄的情况下对左右摄像机轴间距和轴夹角做出了计算，得到了合适而不失精确性的拍摄参数，从而帮助我们在立体拍摄中能以此参数为参考进行对立体摄像机进行参数初调，大大简化了立体拍摄前期的计算工作，以技术辅佐立体拍摄艺术的完善。

6. 广电总局科研项目《多通道智能动声系统关键技术及应用研究》顺利结题

2016年4月11日，国家新闻出版广电总局科技司对《多通道智能动声系统关键技术及应用研究》项目进行了验收。验收会由国家新闻出版广电总局科技司孙苏川副司长主持，验收委员会主任由国家新闻出版广电总局科学技术委员会杜百川副主任担任，委员由总局信息中心、电影质检所、北京电影学院、广科院、规划院、中广电设计院等单位的专家组成。

验收委员会听取了项目组有关项目的介绍，查阅了工作报告、技术报告、科技查新报告、检测报告、用户使用报告和经费决算表等相关材料，并观看了系统演示。经讨论认为：该项目提出了DMS (Dynamic Matrix System) 动声系统，它是一个基于有限音源的多声道播放系统，其通道数目可随意变化。系统提出了一个基于有限声源与声场结合进行综合分析的新方法，建立了多通道动态音频矩阵系统的多入多出 (MIMO, multi-channel input and multi-channel output) 模型，优化了不同声学环境之间的声场匹配；项目组完成了DMS动声系统的演示设计和实现，经测试指标符合项目合同书的要求；该项目对DMS动声系统中基于扬声器阵列的声场一致性优化方法、音乐信号分析及应用等关键技术进行了研究，发表学术论文30余篇，申请专利3项，软件著作权2项，完成了项目合同书考核指标；该项目搭建的DMS动声系统投入使用以来，运行稳定，在北京音乐厅成功演出DMS动声音乐会，在石家庄建立了DMS动声儿童教

育中心，用户反映良好。该项目推动了多声道技术发展，促进了音乐文化传播、带动了相关产业发展。

7. 广电总局科研项目《电影虚拟化制作中的虚拟资产管理》第一年进展顺利

国际上，随着计算机图形学、图像学、虚拟现实等新技术与方法的出现与发展，电影虚拟制作技术逐步成熟与完善，诞生了很多基于虚拟技术制作的优秀影片。电影虚拟化制作可以使电影拍摄时间缩短 30%、成本节约 20%，并且拍摄不受环境、气候、转场等因素的限制。电影虚拟化制作的目的是实现虚拟角色与真实场景、真实角色与虚拟场景的实时合成，以及多机位拍摄立体实时预览，从而让导演在片场实时监看虚拟与现实结合的最终效果。本年度该项目的研究主要集中在如下两个方面：

● 真实场景三维数字化重建方法

现代媒体技术体系中，媒体内容工具的开发是需要首先解决的问题。现有的三维模型的建立主要是基于计算机图形学的建模。本项目的首要任务即建立一种利用拍摄图像进行真实场景三维重建的工具，并在此基础上建立虚拟资产库。

自然场景中的景物多种多样，直接应用传统的三角网格法或泊松表面法无法实现各类景物的高精度重建。另外在图像中经常会出现景物之间的相互遮挡、景物纹理单调以及阴影等情况，使得通过稠密匹配得到的三维表面点云出现信息缺失的情况。本课题拟结合场景理解信息，参照已存储的点云模型进行修补，保证景物重建的完整性。对于基于单视图的景物三维重建，拟参考场景理解信息，结合数据库中的模型进行拟合，实现高度近似的景物重建。

● 多视角图像融合与阴影重建

真实场景与虚拟场景的融合是现代媒体技术中的重要一环，本课题即结合图像分割、光照模型、边缘融合等技术，研究一种多视角图像融合方法。为了实现逼真的 MR 环境，从复杂光照条件下已渲染视角得到图像深度估计是其中的关键。当前解决这个立体映射问题的方法是基于静态的照明、传递函数不随光照变化的，而现实的光照是随时间变化的。另一方面，从信号处理角度上看，局部图像特征的光照和尺度不变性已经成功应用于图像拼接，3D 建模和目标跟踪等。本项目

将采用尺度和光照不变的局部特征来映射立体摄像机的内容。从已映射的立体内容中恢复的深度和高度信息可以通过基于几何的算法进行阴影重建。

(二) 实验室建设进展情况

1. 本年度工作要点概述

1.1 实验室顺利通过教育部建设计划验收

1月19日，教育部组织专家组对我校“媒介音视频”教育部重点实验室进行了建设计划验收。经过实验室建设计划汇报、验收答辩、现场考察等验收环节，专家组一致认为实验室建设成果达到建设目标要求，同意通过实验室建设验收。实验室建设计划的顺利验收标志着我校首个教育部重点实验室筹建完成，“媒介音视频”教育部重点实验室正式进入运行。

1.2 实验室成功申请一项国家自然科学基金重点项目

由“媒介音视频”教育部重点实验室主任张勤教授牵头的国家自然科学基金重点项目《历史影音资料音频修复方法与关键技术研究》申请成功，获批经费250万。本项目在实验室已有声场重建技术领域与音频信号处理领域研究成果的基础上，结合目前我国珍贵历史音频文献数字化保存的需求，并基于我国现阶段音响及音像档案存档标准低于国际标准的事实，研究非接触读取模数转换技术与历史音频资料的噪声特征模型，建立基于音源分离的噪声消除和多种类噪声的自适应噪声消除方法，研究缺失数据插补与说话人转换理论与方法，研究基于深度学习的主客观评价模型。通过本项目的研究，“媒介音视频”教育部重点实验室将建立历史音频修复的理论技术体系，并提高历史音频修复的质量和效率，从而推动历史音频修复标准制定和产业发展。

1.3 CAVE 沉浸式交互显示系统投入运行

媒介音视频教育部重点实验室自2014年5月开始建设CAVE沉浸式交互显示

系统，两年时间内，经过建设方案规划、进口设备采购论证、设备采购招标等各个环节的严格掌控，系统最终于 2016 年 4 月建设完成并通过学校资产处与财务处等相关部门的验收，正式投入运行。CAVE 沉浸式交互显示系统的搭建将为媒介音视频教育部重点实验室开展虚拟现实技术研究提供有力的平台保障，实验室将以此平台为基础，全面发展三维建模、交互、多屏显示等多项虚拟现实技术领域所需技术层面的研究。

1.4 诺基亚全球数字媒体部销售总监到访实验室

诺基亚全球数字媒体部销售总监 Malachi Bierstein 和现场摄录产品部销售总监窦晓军到访中国传媒大学媒介音视频教育部重点实验室。结合实验室的科研优势和诺基亚的产业技术优势，双方将从科研、产业发展等展开合作，并筹划建立“中传-诺基亚虚拟现实联合实验室”。

1.5 科研成果进展

本年度，媒介音视频教育部重点实验室结题国家自然科学基金青年项目 1 项，国家科技支撑计划子课题 2 项，国家广电总局科技项目 2 项；新增科研项目 12 项，其中包括国家自然科学基金重点项目 1 项，横向项目 4 项，新增项目经费总额 610 万元；新增论文 41 篇，其中 SCI 检索论文 10 篇，EI 检索论文 28 篇；获批专利 1 项；颜金尧博士入选中国传媒大学首批“白杨学者”奖励计划。

（三）对外合作与交流

本年度，媒介音视频教育部重点实验室科研人员参加了 CCBN、BIRTV、ITTC、ISBT、电视日中国电视大会等会议和广播电视传媒业界进行技术交流。参加 CCF 互联网专委会、CCF 网络与数据通信专委会会议、CNCC2016、future Multimedia 等学术会议，并在全国高校软件定义网络 SDN 学术研讨会作大会报告（2016 年 8 月南京，江苏未来网络研究院），全国首届 SDN 培训暨专家学术研讨会圆桌讨论专家。

本年度实验室邀请了”111”国际大师，瑞士苏黎世联邦理工(ETH Zurich)的 Bernhard Plattner 教授来访 8 天，邀请美国 Kansas 大学的 James P.G. Sterbenz 教授顺访 2 次。

实验室支持了江苏省苏州市首届“天公苏作”杯非遗文化创意创客大赛在苏州国际博览中心开幕，为大赛提供技术和平台支持，大赛响应了“十三五”文化建设规划，深挖非遗文化艺术资源，汇聚非遗文化创新发展人才、产品、项目和机构，扩大非遗文化在当代社会中的影响力，推动文化产业经济发展；实验室高性能计算中心团队组织承办百万创新创业大赛，大赛覆盖影视、动漫、游戏等众多领域，旨在面向创意出众、想法超前、思维敏捷，具有创业梦想的年轻人以及拥有优秀文创项目、具备商业前景的企业、团队和工作室征集优秀原创项目。

(四) 人才培养与团队建设

本年度，媒介音视频教育部重点实验室共毕业博士 6 人，硕士 23 人。目前，实验室现有博硕士研究生共计 60 人，其中博士研究生 6 人，学术型硕士研究生 32 人，工程型硕士研究生 22 人。

(五) 本年度新增成果列表

新增项目列表：

- ✓ 国家自然科学基金重点项目：《历史影音资料音频修复方法与关键技术研究》，项目负责人：张勤，项目经费：250 万。
- ✓ 文化部 2016 年中央级文化产业引导资金：《支持数字创意内容在线创作和交易的双创服务平台建设》，项目负责人：鲁永泉，项目经费：150 万。
- ✓ 江苏省 2016 年江苏省文化产业引导资金：《数字技术与创意设计融合发展的线上众创空间建设》，项目负责人：鲁永泉，项目经费：120 万。
- ✓ 中国传媒大学工科规划项目：《视觉和听觉的感知关联、交互及其在媒体技术中的应用》，项目负责人：谢凌云，项目经费：30 万。
- ✓ 中国传媒大学工科规划项目：《虚拟 3D 声场通话与警示关键技术研究》，

项目负责人：齐娜，项目经费：8万。

- ✓ 中国传媒大学工科规划项目：《汉语语音清晰度的工程测试方法》，项目负责人：章斯宇，项目经费：7万。
- ✓ 中国传媒大学工科规划项目：《多通道滤波器组设计与规划》，项目负责人，钟微，项目经费：5万。
- ✓ 中国传媒大学工科规划项目：《基于融合策略的平面图像深度估计研究》，项目负责人，叶龙，项目经费：5万。
- ✓ 横向项目：《虚拟声场技术研究》，项目负责人：齐娜，项目经费20万。
- ✓ 横向项目：《移动式强指向性传声器阵列》，项目负责人：刘亚丽，项目经费：10万。
- ✓ 横向项目：《VR虚拟现实视频拍摄系统设计与流程测试》C，项目负责人，叶龙，项目经费：4万。
- ✓ 横向项目：《CAVE虚拟现实沉浸式系统交互插件开发》项目负责人，叶龙，项目经费：1万。

新增论文列表：

[1] Li Fang, Wei Zhong, and Qin Zhang. "Design of M-channel linear-phase non-uniform filter banks with arbitrary rational sampling factors." IET Signal Processing 10.2 (2016): 106-114.

[2] Wei Zhong, Li Fang, Qin Zhang, Long Ye. "Design of oversampled nonuniform filter banks with arbitrary rational frequency partitioning." Signal, Image and Video Processing.(Accepted)

[3] Haiyan Ma, Jinyao Yan, Panagiotis Georgopoulos, Bernhard Plattner, "Towards SDN Based Queuing Delay Estimation", China Communications, 13(3), pp 27-36, 1/3/2016

[4] Yan Hua, Shuhui Wang, Siyuan Liu, Anni Cai, Qingming Huang, "Cross-Modal Correlation Learning by Adaptive Hierarchical Semantic Aggregation", IEEE TRANSACTIONS ON MULTIMEDIA, ISSN/ISBN: 1520-9210, 2016.10.

- [5] Meimei Wu, Yongbin Wang, "Improved Markov song-requesting recommendation algorithm based on user's instant behavior", Journal of the Balkan Tribological Association, ISSN/ISBN: 1310-4772, 2016.7.
- [6] Jianglong Zhang, Xianglin Huang, Lifang Zhang, "Bridge the semantic gap between pop music acoustic feature and emotion: Build an interpretable model", Neurocomputing, ISSN/ISBN: 0925-2312, 2016.10.
- [7] Shang wenqian, Shi Minyong, Hong Zhiguo, Improved Feature Weight Algorithm and Its Application to Text Classification, Mathematical Problems in Engineering, ISSN/ISBN: 1024-123X, 2016.06
- [8] Jianglong Zhang, Xianglin Huang, Lifang Zhang, Feature selection and feature learning in arousal dimension of music emotion by using shrinkage methods, Multimedia Systems, ISSN/ISBN: 0942-4962, 2016.10
- [9] Ni Yepeng, Jianbo Liu, Shan Liu, Yaxin Bai, "An Indoor Pedestrian Positioning Method Using HMM with a Fuzzy Pattern Recognition Algorithm in a WLAN Fingerprint System", Sensors, ISSN/ISBN: 1424-8220, 2016.09
- [10] Ni Yepeng, Jianbo Liu, Shan Liu, Yaxin Bai, "A Context-Aware Adaptive Streaming Media Distribution System in a Heterogeneous Network with Multiple Terminals", MATHEMATICAL PROBLEMS IN ENGINEERING, ISSN/ISBN: 1024-123X, 2016.02
- [11] Baoping Li, Long Ye, Yun Tie, Qin Zhang, Stereo Matching Based on CF-EM Joint Algorithm, The 2016 Pacific-Rim Conference on Multimedia, Xi'an, 2016.
- [12] Qianhan Liu, Long Ye, Wei Zhong, Qin Zhang, A Novel Image Coding Framework Based on Content Similarity Analysis, The 13th International Forum of Digital TV and Wireless Multimedia Communication, Shang Hai, 2016.
- [13] Shengqian Zhang, Wei Zhong, Long Ye, Qin Zhang, A Modified Joint Trilateral Filter for Depth Image Super Resolution, The 13th International Forum of Digital TV and Wireless Multimedia Communication, Shang Hai, 2016.

- [14] Lin Yang, Long Ye, Wei Zhong, Qin Zhang, A real-time interactive system based on hand gesture recognition in virtual fitting, *Transactions on Edutainment*. (Accepted)
- [15] Yingying Zhang, Jingling Wang, Long Ye, Qin Zhang, A Virtual Music Control System Based on Dynamic Hand Gesture Recognition, *Transactions on Edutainment*. (Accepted)
- [16] Y. Zhang, J. Yan, X. Fu, Reservation-based Resource Scheduling and Code Partition in Mobile Cloud Computing, IEEE INFOCOM Workshop 2016, San Francisco, CA
- [17] Juanjuan Cai, Na Li, Hui Wang, Bing Zhu, Isolated Chinese Lyrics with Accompaniment Recognition Based on SVM. 2016 International Conference on Audio, Language and Image Processing (ICALIP 2016) July 11-12, 2016
- [18] Juanjuan Cai, NaNa Wang, Hui Wang, Bing Zhu, Research on the Recognition of Isolated Chinese Lyrics in Songs with Accompaniment Based on Deep Belief Networks. 2016 IEEE 13th International Conference on Signal Processing Proceedings (ICSP 2016) November 6-10, 2016, Chengdu China
- [19] Fei Y, Pengdong G, Yongquan L, et al. Bolt force prediction using simplified finite element model and back propagation neural networks[C]//Information Technology, Networking, Electronic and Automation Control Conference, IEEE. 520-523.
- [20] Yang, F. Gao, P. & Lu, Y, et al. Evolving Resilient Back-Propagation Algorithm for Energy Efficiency Problem[J] MATEC Web Conf., 2016, 77, 06016.
- [21] Yi Xiao, Pengdong Gao, Yongquan Lu, Yu Liu, A parallel Gaussian Elimination for Jacobian Calculation in Magnetotelluric Occam Inversion Algorithm, 2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC 2016), August 27-28, Hangzhou, China
- [22] 李戈, 刘亚丽, 孟子厚. 自然音节状态下韵母区别特征检测. 2016年中国声学

- 学会学术交流会, 2016年10月, 中国武汉.
- [23]刘亚丽, 杜劝劝, 孟子厚. 兰银官话区普通话声调学习的听说实验分析. 2016年中国声学学会学术交流会, 2016年10月, 中国武汉.
- [24]牛欢, 戴璐, 孟子厚. 早期反射声虚拟空间感效果分析. 2016年中国声学学会学术交流会, 2016年10月, 中国武汉.
- [25]罗敏, 孟子厚. 声学特征与数序关联编码对记忆加工过程的影响. 2016年中国声学学会学术交流会, 2016年10月, 中国武汉.
- [26]张婧颖, 孟子厚. 拉弦类民乐乐音的听感谐和性分析. 2016年中国声学学会学术交流会, 2016年10月, 中国武汉.
- [27]杜劝劝, 刘亚丽, 孟子厚. 兰银官话区普通话学习者声母发声与感知的关系. 第十二届全国语音学学术会议, 2016年7月, 中国通辽.
- [28]宋慧, 章斯宇, 孟子厚. 噪声环境中汉语音节与声韵调的清晰度感知关系. 第十二届全国语音学学术会议, 2016年7月, 中国通辽.
- [29]李戈, 刘亚丽, 孟子厚. 韵母区别特征声学参数匹配. 第十二届全国语音学学术会议, 2016年7月, 中国通辽.
- [30]刘晓蕾, 刘亚丽, 付京香. 大学生二语学习语音生态调查. 第十二届全国语音学学术会议, 2016年7月, 中国通辽.
- [31]宋亮, 张婧颖, 谢凌云. 小型扬声器感知低音的主客观评价分析. 演艺科技, 2016年4月第四期, 总第120期: 11-14.
- [32]牛欢, 戴璐, 孟子厚. 空间感主观评测中现场评测与录音评测的比较. 电声技术, 2016年4月, 第40卷第4期: 1:5.
- [33]罗敏, 吴帆, 孟子厚. 特征关联编码对相似声目标学习记忆的影响. 电声技术, 2016年4月, 第40卷第4期: 35:38.
- [34]ZHANG Siyu, ZHENG Xiaolin, MENG Zihou. The influence of dummy head on measuring speech intelligibility, 声学学报 (CHINESE JOURNAL OF ACOUSTICS), 2016, 35(2):178-192.
- [35]赵苗苗, 张婧颖, 谢凌云. 微型扬声器谐波失真的分析与调制. 电声技

术,2016年4月,第40卷第6期:28:32.

[36]陆赫,孟子厚.乐音音形的感知与谐和性分析.声学技术,2016年8月,第35卷第4期:349:354.

[37]仝欣,齐娜,孟子厚.录音重放与现场听音的方向定位信息比较.电声技术.2016年10月,第40卷第10期:36:42.

[38] Yali LIU, Zihou MENG. The examination of the relationship between perception and production of Mandarin tone of Kazak students, The 10th International Symposium on Chinese Spoken Language Processing (ISCSLP 2016), 2016年10月,中国天津.

[39]周颖慧,乔琼.孟子厚.“多边形膜片”传声器音质分析.电声技术,2016年11月,第40卷第11期:77-81.

[40]林凡瑞,赵志军,彭梦雅,谢凌云. Effect of Color on Pitch Discrimination of Pure Tone the 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI 2016) 2016.10.15-17 中国大同.

[41]段云鹏,谢凌云.音色感知不变的参数调节与分析.电声技术,2016年11月,第40卷第11期:82-85.

新增授权专利:

ZL201210516769.X, 单声道多音音乐信号的自动转录方法及装置,王晖,王雨田,朱兵,王颖,官妍竹

(六) 2017年度实验室开放课题指南

在2016年年初实验室建设计划验收会上,验收专家组对于实验室今后的发展提出了“顶天立地”的建设要求,即实验室既要做出一流的科研成果,也要充分服务于传媒产业的发展需求。2016年7月,广电总局发布了《关于进一步加快广播电视媒体与新兴媒体融合发展的意见》,加强加快媒体融合,特别包括了IP化和云化等媒体制播关键技术的研发与应用方面的研发成为了传媒行业内部

急需解决的技术性问题。因此，我们相应提出了媒介音视频教育部重点实验室2017年开放课题指南：

1. 视频直播平台中用户体验分析与内容分发研究
2. 边缘计算中的视频任务调度和资源分配优化
3. 软件定义网络（SDN）视频交换开放硬件平台
4. 视频浅压缩方法与实现
5. 基于SDN网络的视频可靠传输方法
6. 基于反馈控制的媒体直播数学模型
7. 互联网络传播模式研究
8. 数据驱动的影视内容制作模式研究

五. 实验室2017年度工作要点

本实验室2017年度建设指标可以概括为：第一，力争在重大课题项目申报、高水平论文发表、高层次人才培养方面实现突破；第二，科研项目和科研成果的稳步双增长，即科研项目的数量、质量以及科研经费总数稳步增长，发表科研论文、申请发明专利的数量、质量稳步增长；第三，以中国传媒大学苏州研究院和北京市协同创新研究院为平台，积极推动各项关键技术的产业化转移，即为已有的科研成果寻找产业化转移方向，为已集成的系统拓展盈利模式，增加利润增长点，为已示范的工程寻求更高的展示平台，扩大产业影响力。

1、 有针对性的建设思路

在科研团队建设方面，团队成员的职称结构、年龄层次要分布合理，在研究领域上也要做到既有分工又有合作。团队负责人除了自己负责的重点方向以外，还肩负着全局指导的责任，负责团队整体资源的调配。其他人至少要各自负责一项团队的重点研究方向，同时也要至少参与一项其他人负责的重点方向。团队还要注意和校内其他学院的合作，借助中国传媒大学在传媒学科方面的领先优势，开展科学与艺术的协作研究。团队要鼓励团队成员与其他学院的骨干教师形成稳定的合作关系，开拓新的研究思路和研究领域。同时，也要积极开展与外校的交流

流与合作，扩大本团队与研究人員个体在同行的知名度。在产学研转化平台建设方面，积极拓展高校科研团队和企业的合作，加大教师尤其是青年教师的实践能力，加大引入企业和行业专家作为兼职教授的力度，加大和业界的沟通交流，加大和国外先进科研团队的沟通和学习。

2、 保障举措

首先，强化管理考核。科学的管理与考核是各项研发工作不断推进的保障。本实验室在提供优厚科研条件的基础上，将实施科学的考核方法。在鼓励完成常规考核指标包括科研经费、科研成果产出的条件下，留出足够的自由发挥空间，鼓励创新的学术思想，有深度的学术研究。考核将融合学术委员会综合学术评议和行政考核，对有特长和特殊贡献的人员给予特殊奖励。考核期按不同情况分为短期、中期和长期，最短为一年。第二，稳定科研队伍。不仅要努力创造良好的工作条件和环境，还要让团队成员看到未来的发展空间，包括在实验室平台上的发展空间、在中国传媒大学的发展空间、在国内行业领域的发展空间以及在音视频学术界的发展空间等。第三，加强帮带培养。设立青年人才启动和培育基金，对刚进入实验室平台的青年人才配套 3-5 万元的科研经费；对有创意，但还没有实质经费资助渠道的学术研究设想，给予适当的培育经费支持。第四，浓厚学术氛围。在中国传媒大学多学科、小综合的特色基础上，努力营造一个宽松、包容、自由的学术氛围，消除行政管理对学术思想的干预，鼓励多学科交叉的创新性学术思想。在科学严谨、求实创新、淡泊名利、精益求精的前提下，提倡奇思、鼓励创意、包容异见。

3、 争取政府和社会支持的目标与对策

2017 年，媒介音视频教育部重点实验室将进一步完善以学术委员会为最高学术决策机构的主任负责制度，并充分发挥学校、学部的管理和监督作用，简化行政，形成以学生培养、学术研究、产业化转移为中心，自由、开明、团结、务实的学术氛围。加强国际学术交流，鼓励并支持科研能力出色的青年教师与博士、硕士研究生赴外交流，同时在实验室设立“大师工作室”，做好学术研究与艺术

创作的对外引智工作。构建社会资助体系，寻求多渠道的社会募捐和赞助，用于资助学术、奖教和奖学等。