

通信声学的发展及研究现状

陈小平[†]

(中国传媒大学录音系 北京 100024)

摘要 人类作为信息交流的主要信息产生源和信息接收槽,语言声学、心理声学和生理声学在信息技术中发挥着越来越重要的作用。作为声学信息技术相互渗透的产物-通信声学,近十几年来随着数字信号处理技术的发展,其研究和应用领域得到了很大的扩展。本文在简述通信声学产生和发展的必然性和必要性之后,主要介绍了通信声学目前的研究领域及其应用。

关键词 信息技术, 听觉信号处理, 虚拟声环境, 听力学, 声音质量评估

The development and research areas of communication acoustics

CHEN Xiao-Ping

(Department of Sound Recording, Communication University of China, Beijing 100024)

Abstract Because speaking and hearing are the most common ways for information exchange as information source and information sink, speech acoustics, psychoacoustics and physiological acoustics are playing a more and more important role in information technology. Communication acoustics, the result of interpenetration between information technology and acoustics, is now getting a big extension in its researches and applications, due to the development of DSP technology. This paper will mainly introduce the present research areas and the applications of communication acoustics, after briefly introducing its history of development.

Key words Information technology, Auditory signal processing, Auditory virtual environment, Audiology, Sound quality assessment

1 引言

通信声学(Communication acoustics)这一名词是由德国声学学会创始人之一 J. Blauert 教授提出的,1974年在德国鲁尔大学的电子工程和信息技术系建立了以此词命名的声学研究所。该研究所几十年来一直致力于通信声学领域的科学研究,取得了较为显著的研究

成果,其中相当一部分在工业生产中得到了应用。在几十年国际上众多方面的探索和实践的基础上,Blauert 教授在2002年 AES年会上对通信声学这一研究领域做了一个专题演讲^[2],进一步强调了通信声学已成为应用声学的一个重要分支以及其存在的必然性和必要性。

通信声学是指与信息技术和计算机科学

2007-06-04 收稿; 2007-10-08 定稿

作者简介:陈小平(1963-),女,副教授,硕士生导师,研究方向:心理声学及听觉传输技术。

[†] 通信联系人 E-mail: xiaopingchen@21cn.com

密切相关的声学分支^[1],可以认为是由电声学产生的,它是语言声学、心理声学和生理声学在通信领域和声音信号处理中的应用。它主要研究与听觉密切相关的信号以及语音信号的分析 and 处理问题,例如听觉模型的建立、听觉信号合成、音质评价、语音信号合成与识别等。

2 从电声学到通信声学

声学是一门古老的学科,但它得到迅速发展,并涉足不同的工程技术和应用领域,衍生多个分支,是从一百多年前开始的。声学的发展存在两个高峰期^[2]。第一个高峰期始于19世纪末,由于电子管的发明和低频信号放大的实现,使得早期的一些发明,如声音信号磁性录音和光学录音技术,能够进入实用阶段,由此促进了录音技术的极大发展。1920年以来,随着无线电广播这一公众传播媒体的出现以及高保真直接辐射式扬声器的发明,公共场所的广播扩声系统又得到了广泛应用,电影工业也从无声时代进入了有声时代。在这一时期,声学开始从纯粹的物理学研究变成为涉足多个应用领域的学科,其研究领域得到了很大的扩展。电声学这一专业名词就是从这个时候开始使用的。1965年以来,随着计算机技术的发展,数字信号处理技术开始渗透到声学的各个应用领域,带来了新的声音信号记录和存储方式,使声学在各个领域的应用得到了进一步的飞速发展,形成声学发展的第二个高峰期。

与此同时,计算机技术的发展也给电气工程这一传统学科所涵盖的内容带来了变化。在上世纪50年代以前,人们普遍认为电气工程是由强电和弱电两大部分共同组成,但是,50年代以后,人们逐步将强电部分和弱电部分进行更明确的划分,称前者为机电工程,后者为电子信息工程。现在,电气工程在大多数人的概念里是指机电工程,而原来所涵盖的弱

电部分已被完全分离出来,称之为信息技术。这种变化从现在各大学院系的命名中可以看出。

从电声学的应用来看,电声学与信息技术中通信的关系最为密切。随着声学和技术之间的相互渗透越来越深入,电声学的研究领域也不断扩大。由此,逐渐将电声学中与信息技术关系较为密切的一部分研究领域称为通信声学。而现在的电声学更多地是指电能与机械能或声能之间的能量转换原理、技术和应用。

3 通信声学的研究领域及现状

3.1 语音信号处理

语言是人与人之间进行通信交流的主要工具。因此,语音信号处理在现代信息技术中起着非常重要的作用。从较简单的语言信息传输到较复杂的人机对话接口系统,语音信号处理模块是构成系统的重要组成部分。

语音信号处理主要包括语音传输、语音识别、语音合成和语音降噪等课题。在现代通信技术中,语音信号处理主要应用于人机对话接口系统(Human-machine interfaces),而人机对话接口系统在人机交互系统中又是不可缺少的一个组成部分。图1所示为语音信号处理在现代通信技术中应用的示意图^[3]。在图1中,中心模块为语音信号处理的核心部分,主要完成对拾音传声器指向性的控制、语音声源定位、降噪、回声消除、语音信号压缩和均衡以及声场合成等工作。语音信号的输入可以由多个传声器组成的传声器阵来完成,以达到控制传声器指向性的目的,使传声器总是指向声源。降噪处理主要是根据语音信号和干扰噪声信号的统计特性不同将携带信息的语音信号和噪声信号进行分离。中心模块处理后的语音信号既可以经过编码后进行传输,也可以直接送给人机对话接口系统进行处理,实现人机对话功能。

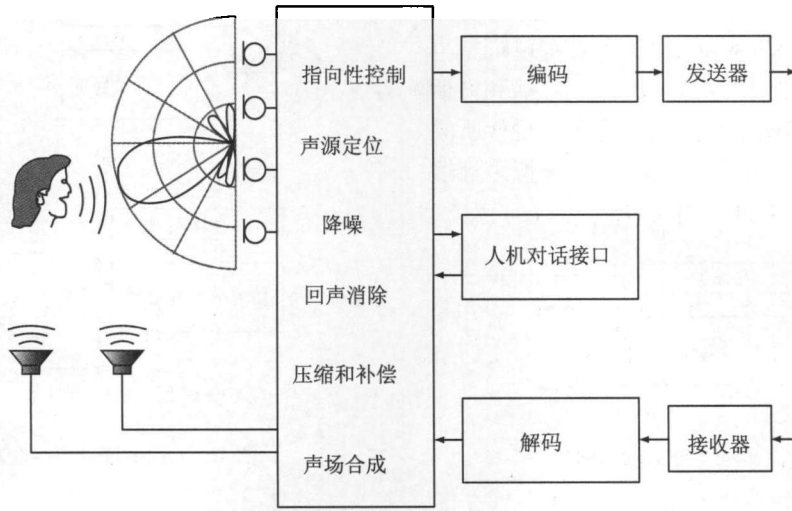


图1 语音信号处理在现代通信技术中的应用

人机对话接口模块的工作原理如图2所示^[3],它主要由语音识别、语句理解、对话管理、响应生成和语音合成等几个部分组成。经过中心模块处理后的语音信息首先进入语音

识别,然后进入语句理解后输出可理解的语句,进入对话管理系统,根据不同的应用来产生响应,最后进入语音合成模块,将响应生成模块的输出转换为听觉能够接收的语声。

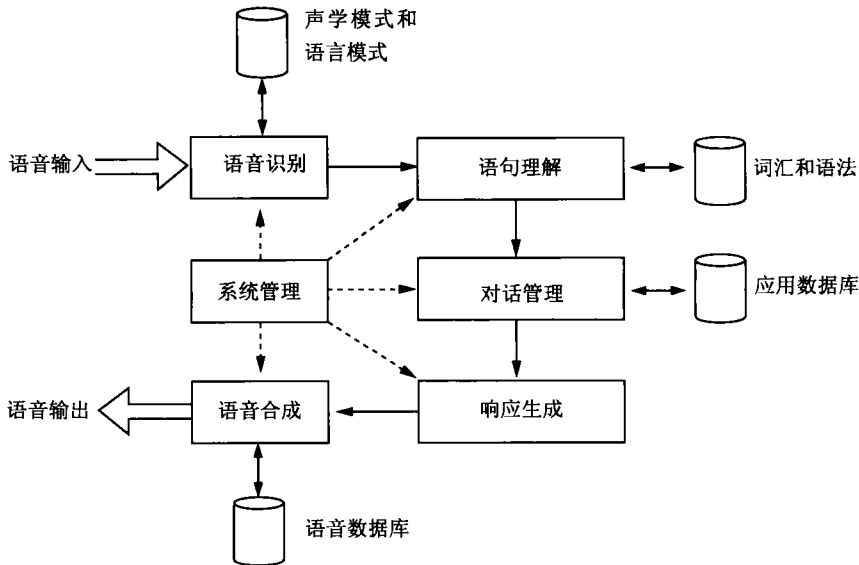


图2 人机对话接口模块工作原理

3.2 双耳技术和听觉信号处理

通信声学的主要研究对象之一是听觉传输系统(Audio-transmission systems),它是双耳技术(Binaural technology)的应用。双耳技术(有

时也叫听觉传输技术^[4])是基于以下假设,即人的双耳鼓膜处的声压信号包含有整个听觉事件的所有信息,包括音质信息和空间信息。如果能在听音者双耳鼓膜处准确再现双耳信号,那

么听音者会产生身临其境的听觉感受。听觉传输系统简图如图3所示,主要用来实现听觉事件在不同时间和不同地点的传送。广播和电话可以说是听觉传输系统的初级形态。现代听觉传输系统在录音、声场听觉化模拟、虚拟声环境等方面都有着广阔的应用前景。

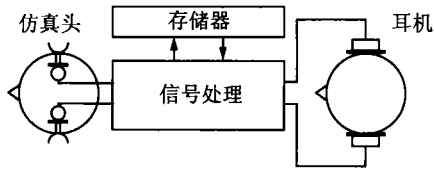


图3 听觉传输系统简图

听觉传输系统又可以分解成两大部分,如图4所示^[2],分别对应于通信声学的两大研究方向,一是听觉事件分析(Analysis of auditory scenes),有时也称为听觉信号处理(Auditory signal processing);二是听觉事件合成(Synthesis of auditory scenes)。在图4(a)中,听音者双耳像两只压强式传声器一样接收声音信号,声音信号经由外耳、中耳、内耳后到达听觉中枢。在这一过程中,听音者根据听音的目的,如声源定位、在众多声源中分辨所要听的声音或是对某种声音进行主观评价等,可以指挥听觉对声音信号进行一系列分析和处理,并在大脑产生响应。听觉信号处理或听觉事件分析就是了解在某个听觉事件中听觉对声音信号的分析和处理过程,并把这一过程用算法表示出来,通过计算机对这一信号处理过程进行模拟,用输出的一系列参数表示听觉对这一听觉事件所产生的响应。听觉事件合成则是一个相反的过程。在图4(b)中,输入某一听觉事件的特性参数,经过计算后产生双耳信号,在输出端用耳机或扬声器重放时,能在听音者大脑重现这一听觉事件。听觉事件合成实际上是用计算机来合成某一听觉事件的双耳信号。当听音者需要与系统进行互动来改变听觉事件时,系统还要接收来自听音者的反馈信号,这种动态系统称为虚拟声环境(Auditory virtual environments)。虚拟声环境将在随后的小节做进一步介绍。

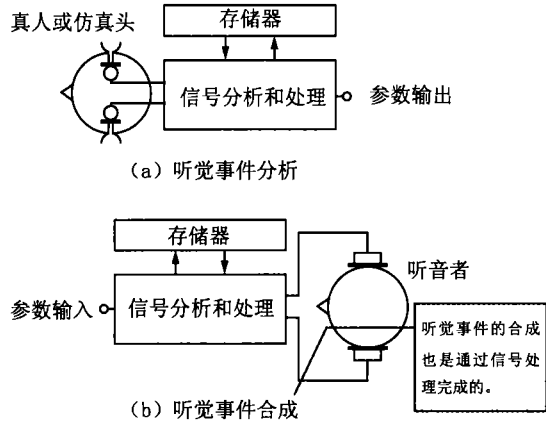


图4 听觉事件分析与合成

从双耳效应出发,听觉信号处理的研究课题有很多^[3],例如,研究语音信号在实际听音环境中直达声和反射声之间的掩蔽效应和反射声对直达声的加强作用。一旦了解了听觉这一信号处理方式,其成果可应用于自动语音识别器和助听器中的语音增强。再如,通过对鸡尾酒会效应的研究,可以模拟听觉这一信号处理过程,得到所谓的“鸡尾酒会处理器”。“鸡尾酒会处理器”可用于在噪声环境中或多个声源同时存在的不利听音环境中提取有用的语音信号,或用在基于声音信号的监视系统和导航系统,也可以用在助听器和机器人语音识别器的前端。此外,双耳测量技术(用仿真头测量)也是建立在听觉信号处理的研究基础之上。

听觉信号处理或计算机听觉事件分析一般是以人类听觉的信号处理为原型,其信号处理方框图如图5所示^[2,5]。人类听觉信号处理的基本过程大体上分为两大部分。第一部分是自下而上的信号走向,代表外耳、中耳、内耳以至听觉中枢对声音信号的处理,属于心理物理声学部分。其信号处理过程如下:首先利用真人头或仿真头拾取双耳信号,这相当于声音信号经过了外耳模块处理。然后经过中耳模块后,两路信号分别送入耳蜗模块,该模块主要完成的工作是,将信号按照听觉的频率分辨率(临界频带)进行频谱分析,然后进入自动音量控制单元,对较高的音量进行压缩,以模

拟耳蜗的非线性,然后各频率分量被转换成神经电脉冲形式,用脉冲的频度表示信号的强弱,这个处理过程可看成一种特殊的 A/D 变换,这样便产生了一对脉冲密度随时间变化的信号。随后,这对耳蜗输出信号被送到双耳信号处理模块,由双耳信号处理模块通过双耳互相关函数计算出双耳信号在不同临界频带的时间差。在有多个声源同时存在的情况下,分析计算结果会出现多个峰值。这些峰值再通过“增强程序”进行加强,以模拟听觉的“双边抑制”效应。然后这些峰值所在的位置及形状等特征可用来识别不同的声源及其空间位置。同时,双耳信号处理模块还可以计算双耳信号的声级差,用来修正互相关函数的分析结果,使声源定位更加准确。由于用一只耳朵也能

听到声音,所以还需要增加单耳信号处理模块。最后,由下至上的听觉信号处理的结果可用一个随时间变化的三维图像来表示,称为双耳工作样式(Binaural - activity pattern),三个变量分别是频率、强度和声源侧向方位角。第二部分是由上向下的信号走向,代表人的思想状态、认知水平以及其他感知器官(如视觉、触觉等)对听觉事件形成的影响,属于感觉心理学部分。同时,心理学部分又会对心理物理声学部分形成反馈即产生影响,例如,听音者的注意力是否集中会影响到听觉的敏感度。最后,综合上述多种因素,在大脑产生了听觉事件的响应。在这里听觉事件的响应用听觉事件的特性参数来表示,如声源方位、声源个数、音色、响度、房间大小、混响时间等。

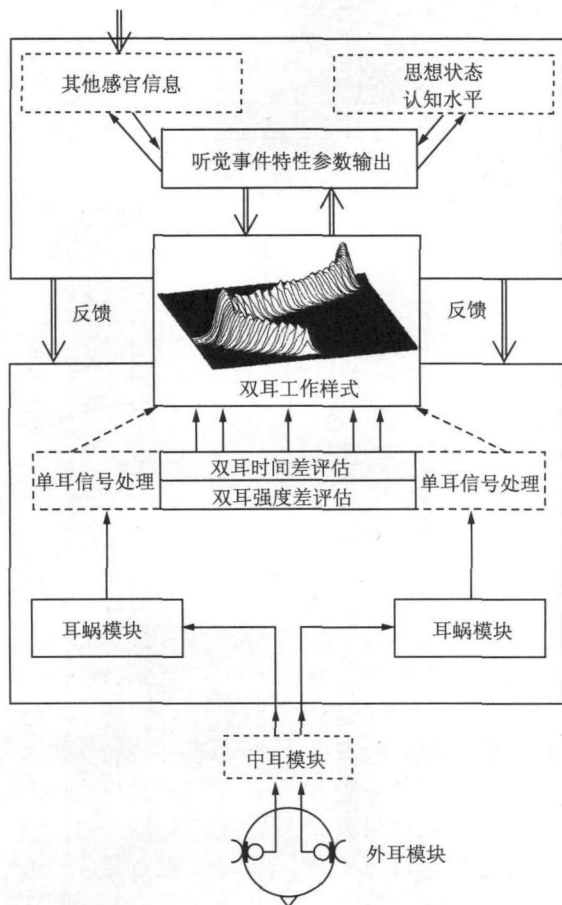


图 5 听觉信号处理方框图

3.3 虚拟声环境和声场听觉化模拟

虚拟声环境是多元虚拟现实(Multi-modal virtual reality)系统的一个重要组成部分。虚拟声环境往往是一个动态系统,即使用者可以与系统进行交互作用,例如用手自由移动声源或在虚拟空间中走动等,会在听觉产生与之对应的响应。虚拟声环境主要应用有远程电话会议、虚拟演播室和听音室、声音效果器、交通工具模拟装置的伴音系统以及科学研究等。

为了说明虚拟声环境通常是多元虚拟现实系统的一个组成部分,图6所示为一个多元虚拟现实系统的结构简图^[2],其系统结构主要由两大部分组成。一是系统的核心部分,称之为事物模块(World model),其本质上是一个数据库,用来存储所要描述的客观对象的相关数据。其中,应用模块存储着与不同应用相关联的各种算法,控制模块的主要功能是收集来

自使用者的反馈信息并使系统及时地产生响应。跟踪手、手指和头部位置的信号通过位置跟踪系统送到控制模块。二是响应生成部分,主要进行数字信号处理,产生最终馈送给使用者的信号。其中声音生成模块主要进行声场模拟计算,并依据外耳传输函数(HRTFs,即头部相关函数),产生所谓的双耳脉冲响应(Binaural impulse responses)。双耳脉冲响应代表听音者所在声环境和外耳的传输特性,将它们分别与无混响的声音信号卷积后,就得到所需的双耳信号。头部相关函数一般是事先测量好并储存在数据库中,而且最好采用使用者本人的外耳测量数据。虚拟声环境系统对数字信号处理的硬件要求是每秒钟能进行至少30次的双耳脉冲响应计算,而且系统的延迟时间应小于50ms。只有这样,听觉才不会产生跳跃或不连续的感觉。

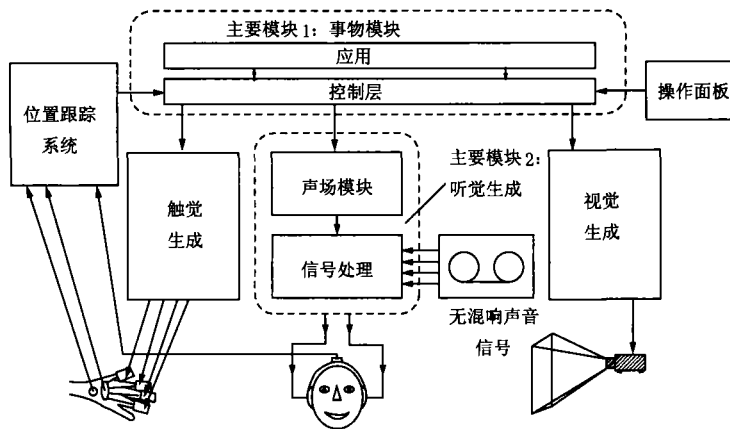


图6 虚拟现实系统结构简图

由于虚拟现实系统需要对声音信号进行实时处理,因此对系统硬件和计算能力有一定的要求。然而,大部分的虚拟现实系统并不要求产生十分逼真的效果,其设计目标仅仅是让使用者产生置身于某一虚拟环境的感受,只要结果令人可信即可,这使得系统进行信号处理的负担有所减轻。尽管如此,对听觉信号处理过程的了解和心理声学方面的知识对虚拟声环境技术是十分重要的。只有了解了听觉信号处理的详细过程,才能确定在某一时刻哪些信号属性与听觉的关系最为密切,需要进行精

确和快速的计算,而哪些信号属性与听觉的关系不大,可以稍后进行计算或完全予以忽略。

在听觉特性的研究过程中,往往要进行大量的听音实验。在进行听音实验时,通常要对听音事件进行一些参数的改变。为了简化实验过程和减少实验进行的工作量,最好能应用虚拟声环境,这样可以简便地通过计算机改变实验参数,达到对输入听觉事件进行调整的目的。因此,科学研究是目前虚拟声环境的重要应用之一。

声场听觉化模拟(Binaural room simula-

tions)是指用计算机合成现实中可能存在也可能不存在的某一听音环境的双耳信号,重放后使听音者产生身临其境的听觉感受。它通常是指静态模拟,不要求使用者与系统之间的交互性,但是对虚拟结果的逼真性要求较高,这是其与虚拟声环境的主要不同之处。由于不需要实时跟随头部移动等,因此容许算法十分精确复杂,也就可能得到十分逼真的模拟效果。声场听觉化模拟技术较多地应用于建筑声学设计,使在工程实施前能够听到厅堂的声学效果。

3.4 听力学

听力学是以正常的和听力有障碍的听觉功能为主要研究对象的一门科学^[6]。人的听觉器官的解剖结构和听觉生理非常复杂,听力减退的原因和机理也是多种多样。听力学是在许多学科发展的基础上,应人类对解决听觉障碍的困扰的需要而发展形成的一门学科,它与许多学科如生理学、病理学、耳科学、其他临床医学、心理学、电声学等有着密切联系。

听力学可分为实验听力学(experimental audiology)和临床听力学(clinical audiology)。实验听力学研究听觉生理、听觉病理以及听觉心理。临床听力学分诊断听力学和康复听力学两大部分。诊断听力学包括听功能的测试、测试结果的分析、听功能状况的评价、判断听觉障碍程度等,协助耳科、神经内、外科发现病变,提供诊断和处理的建议。

通信声学在这一领域主要涉足对听觉外围部分即外耳、中耳和内耳的生理机能的研究^[3],借助于声学和力学的测量手段,找到并弥补缺失的数据,深入了解听觉器官的工作机理,从而建立其物理模型。听觉外围工作机理的认识可以实现和优化各种不同的应用技术,如耳机和听觉保护器设计、气导和骨导助听器设计、中耳移植和耳蜗移植技术、听力诊断技术等。

3.5 系统声音质量评估和预测

虽然通过电信网络进行语音传输已有一百多年的历史,但是,由于近年来新的编码和传输技术的出现,给传统的信号传输网络带来了巨大的变革,又由于许多国家电信市场的开放,带来了网络的互联和更经济的传输带宽,使得通信系统的声音质量越来越成为值得关注的问

题。过去的通信网络是较为单一的,主要采用模拟或数字传输,带宽都是300~3400Hz,终端都是普通固定电话,通信质量主要受到信号衰减、电路噪声和脉码调制产生的量化噪声的影响,因此通信系统的声音质量较为一致,容易在用户思想上形成对通信系统质量的评判标准,因此供应商也容易以此为准则来满足用户的使用要求。现在的情况却不同。由于大规模的移动通信网络和IP网络的建立,声音的质量有可能受到低码率编码技术、信号处理带来的较大延时的影响,使用的终端设备可能是传统的固定电话、移动电话、电脑耳机或语音电话系统等多种多样的形式,因此不存在一个声音质量相对稳定的参考系统,很容易在用户对系统质量的期望值与供应商的标准之间产生隔阂,使用户对服务质量感到不满意。

不论是在通信方式还是在家用电器上,随着社会经济的发展和科学技术的进步,当今的市场是一个多样化的市场,涌现出了许多基本功能相同的产品。在这种情况下,声音便成为影响产品质量不可忽视的因素,分析和评估声音质量以至工业产品声音设计便成为科学研究的一个大课题^[1]。

这里所说的系统声音质量主要包括通信系统声音质量、厅堂声音质量、重放声质量和工业产品声音质量等四个方面,广义上又可以统称为产品的声音质量。不论从哪个方面来讨论声音质量,就质量本身而言,它始终不仅与客观因素有关,而且与使用者的主观因素以及事件发生的背景有关。因此,质量可以看成类似于“听觉事件”的“质量事件”,发生在使用者对服务进行感受并做出判断的过程中。在研究声音质量时,要同时考虑到影响“质量事件”的客观因素和主观因素,通常把前者称为“质量要素”(Quality elements),后者称为“质量特征”(Quality features)^[7]。质量要素可以理解为描述系统物理特性和服务特性的客观参数,是由设计者提供的,设计者可以通过优化设计提高系统质量;质量特征则是指使用者对系统各方面性能的主观感受,并由此产生一个系统质量的印象。以语音通信系统为例,质量要素有传输通道不同点的语音和噪

声信号、传输通道的频率响应、延时、噪声电平、质量特征有可懂度、听音费力程度、噪音恼人程度、声音质量感受等。

在研究质量时,更重要的是要找到质量特征和引起质量特征的质量要素之间的关系,以及它们与系统总体质量水平之间的关系。而它们之间并不存在一种固定不变的关系,而是因具体应用、使用者的不同而不同。可以针对某一特定的应用建立它们之间的关系,这对系统的优化设计有着重要意义。影响系统质量的各个因素之间的关系以及它们与系统总体质量水平之间的关系一般用分类图的形式来

描述,称为质量分类图(Quality taxonomy)^[1],其实质是质量评估模型(Quality assessment model),用来描述质量的产生过程。作为一个例子,图7所示为虚拟声环境系统质量(声音质量)分类图^[8]。因此,当设计一个系统或某个音频软件并进行质量评估时,首先要找到研究对象的质量要素和质量特征,然后建立描述系统质量特性的模型即质量分类图。在建立了这种分类图以后,需要通过听音实验找到各个相互关联的元素之间的权重关系。只有这样才能最终完成对系统质量的评估,并使得对系统质量的预测成为可能。

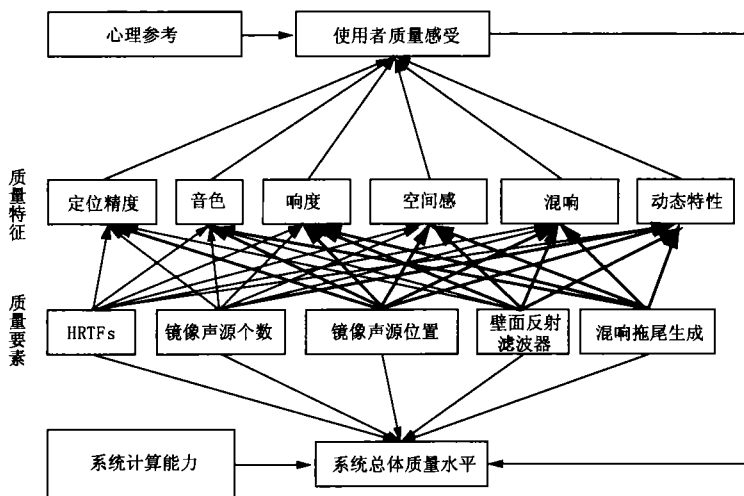


图7 虚拟声环境系统质量分类图^[8]

进行系统声音质量评估和预测是信息技术不断发展的必然结果。由于计算机信息处理的能力是有限的,而高端的信息系统如虚拟声环境系统往往要求计算机有很强的信号处理能力,因此,只有通过系统质量评估和预测,才能了解在不同的应用场合哪些质量要素和质量特征对系统的总体质量起主导作用,从而对这些关键的因素投入较大的计算成本,合理地分配计算资源。这样不仅能够优化设计,而且能够降低成本,使一些难以实现的系统设计能够得到实现。

参 考 文 献

[1] J,Blauert (Ed). Communication Acoustics. Springer Berlin Heidelberg New York, 2005.

[2] J,Blauert. Instrumental Analysis and Synthesis of Auditory Scenes: Communication Acoustics, AES 22nd International Conference on Virtual Synthesis, Entertainment and Audio, 2002.

[3] <http://www.ika.ruhr-uni-bochum.de/>

[4] 谢波荪. 听觉传输技术及其应用. 电声技术, 1997, 20(12): 2-8.

[5] J,Blauert. Spatial Hearing-The Psychophysics of Human Sound Localization. Revised Edition. MIT Press, 1997.

[6] <http://www.digiton.com.cn/>

[7] U,Jekosch. Basic Concepts and Terms of "Quality", Reconsidered in the Context of Product-Sound Quality. Acustica-acta acustica, 2005, 90(6): 999-1006.

[8] A,Silzle. Quality Taxonomies for Auditory Virtual Environments. 122nd AES Convention, Vienna, Austria, 5-8 May 2007.