

面向21世纪高等院校优秀教材

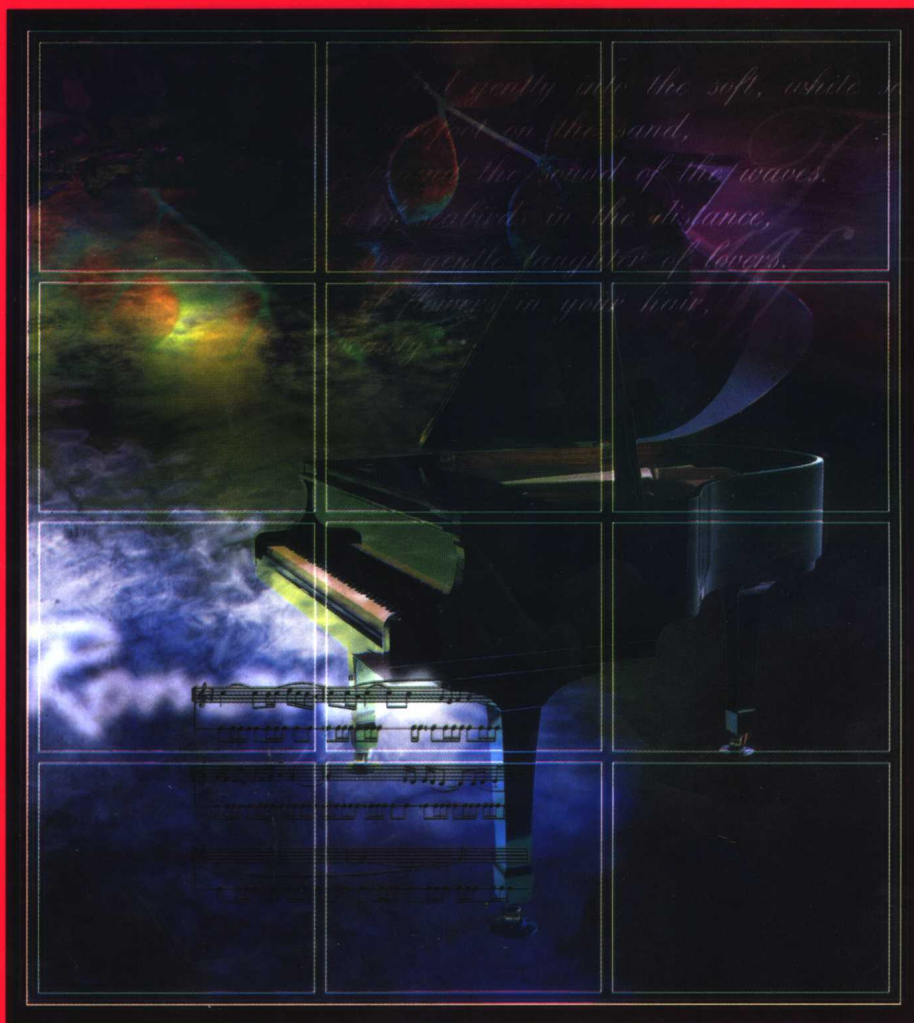
☆ 多媒体影音制作与传播丛书 ☆



主编 \ 刘毓敏

数字音频素材的制作与运用

刘毓敏 李剑琴 杨晓宏 张雪莉 编著



国防工业出版社

<http://www.ndip.cn>

面向 21 世纪高等院校优秀教材

多媒体影音制作与传播丛书

主编\刘毓敏

数字音频素材的 制作与运用

刘毓敏 李剑琴 杨晓宏 张雪莉 编著

国防工业出版社

·北京·

图书在版编目(CIP)数据

数字音频素材的制作与运用/刘毓敏等编著. —北京:
国防工业出版社, 2004. 7

(多媒体影音制作与传播丛书)

ISBN 7-118-03491-6

I. 数... II. 刘... III. 音频设备—数字技术
IV. TN912.2

中国版本图书馆 CIP 数据核字(2004)第 041025 号

国防工业出版社 出版发行

(北京市海淀区紫竹院南路 23 号)

(邮政编码 100044)

新艺印刷厂印刷

新华书店经售

*

开本 787×1092 1/16 印张 20¼ 464 千字

2004 年 7 月第 1 版 2004 年 7 月北京第 1 次印刷

印数:1—4000 册 定价:29.00 元

(本书如有印装错误,我社负责调换)



丛书总序

刘毓敏

在步入信息时代的今天，我们每个人都不应是信息的被动接收者，而更应是信息的积极运用者和制造者。也就是说，我们每个人都不应只满足于在电视机前看看新闻和影视节目、欣赏各种数码影音光盘，或在网上查查资料、收看各种直播或点播影音节目。今天的数字化、网络化的技术信息环境，已经为我们提供了优质、简便的多媒体信息制作与传播平台。各种质优价廉、使用简易的数字图声素材制作、采集工具，如数码相机、数码摄录像机、数码录音笔等；各种功能齐全、性能优良、使用傻瓜化的桌面数字影音制作、处理软件，如 Photoshop、Premiere、3DS MAX、Cakewalk 等；各种多媒体影音作品的存储和传输工具，如各种大容量的存储手段（如光盘、移动硬盘、优盘等）、支持多媒体影音网传输的流媒体技术等，所有这些，使得制作和传播影音信息已经不再仅仅是诸如电视台、各类影音制作中心等专业部门的事，每个影音爱好者都可以自己动手制作多媒体影音作品，如多媒体影音形式的 E-mail、个人简历、庆典纪念……

目前，多媒体影音制作与传播技术及工具不仅为广大影视爱好者所认识和广泛采用，而且已成为诸如中小学、企业、机关等社团自制非广播/发行用途多媒体影音作品的主要途径。但这些用户往往由于没有系统地接受过影音制作的培训，存在许多知识结构的缺陷，他们只是零星地懂得一点儿多媒体影音制作的技术知识，但对如何用好这些技术来制作一个完美的影音作品缺乏全面系统的知识，例如，拍和编好一段视频镜头所需的电视构图的知识、蒙太奇知识等。因此，有必要向广大影视爱好者、多媒体制作人员（如参与 CAI 制作的各级各类学科教师、各种社团的宣传干事）等系统传授多媒体影音制作的技术知识，这正是《多媒体影音制作与传播丛书》的目的。

本丛书在近期将与读者见面的第一批书目有《电视摄像与编辑》、《电视用光基础》、《数字图像素材的制作与运用》和《数字音频素材的制作与运用》等。这 4 种图书基本涵盖了目前多媒体影音制作与传播的主要方面，为读者完整掌握这方面的知识打下了初步的基础。其中：

《电视摄像与编辑》主要由刘毓敏、黄碧云和王首农编写，该书

通过电视画面语言的特点、电视摄像与编辑的基本方式、电视摄像的基本操作方法、桌面多媒体影音编辑制作的基本操作方法、电视画面造型基础、电视编辑基础等，全面系统地介绍了基于数字摄录像机和桌面多媒体影音制作系统的电视摄像与编辑的基本方法。

《电视用光基础》由刘毓敏编写，该书通过光度学与色度学基础知识、电视曝光与彩色还原控制的技术原理、光影与色彩构成的艺术原理、电视照明器材、电视采光与照明的方法与技巧等，全面系统地介绍了电视用光的技术、艺术原理和实际操作方法与技巧。

《数字图像素材的制作与运用》由刘毓敏等编写，该书通过数字图像技术基础、静态图像素材的制作与处理、视频素材的采制与处理、动画素材的制作等，全面系统地介绍了多媒体影音制作中常常用到的数字图像素材的制作、处理与运用的基本方法与技巧。

《数字音频素材的制作与运用》由李剑琴、刘毓敏编写，该书通过数字音频技术基础、各类数字声音素材的采制、数字声音素材的处理与运用等，全面系统地介绍了多媒体影音制作中常常用到的数字声音素材的制作、处理与运用的基本方法与技巧。

本丛书是上述参编人员的多年电视制作教学与实践经验的结晶，希望本丛书的出版能为电视制作知识的普及尽一份力，更企盼着广大读者、同行指正书中欠妥甚至谬误之处。

值得指出的是，为了更好地说明电视制作的艺术原理，本丛书广泛引用了影视作品、绘画、摄影、音乐等的成功范例，均已在文中注明出处，在此，谨向这些优秀作品的作者致以衷心感谢。如有错漏，更希望各有关作者原谅。在这些范例中，凡注明“DV 素材”者，均为刘毓敏创作。

2003 年 9 月于华南师范大学

前言

◆
刘毓敏

声音是多媒体影音作品意义建构中必不可少的媒体，它与图像、字幕等有机地结合在一起，共同承载着制作者所要表现的客观信息和所要表达的思想、感情。因此，声音素材的制作与运用是多媒体影音制作非常重要的一环。以往声音素材的制作技术是非常专业化的，无论是声音的拾取与记录，还是音频信号的调音和效果处理，均需要昂贵的专业设备和专业人员操作。更不用说，为了获得理想的音响效果，专业声音素材制作中还需要专业的乐队的演奏。所有这些，都是只有那些有设备、有人才的专业影音制作单位才“玩”得了和“玩”得起的！

而今，随着数字技术的广泛应用，不仅使得各种音频制作设备以其高性能、低价格而得以“飞入寻常百姓家”，而且随着 PC 的普及与性能的不不断提高，更使得原来许多只有价格昂贵、体积庞大的专业音频制作设备才具有的强大功能，可以通过软件而得以实现。而这些数字音频应用程序的用户界面又通常非常友好，不仅符合专业音响工程师的专业操作习惯，而且因为其直观易懂，一般多媒体开发人员也能很快掌握其操作使用的方法。正是这些数字音频技术的普及，使得今天的音频素材制作已经不再是专业影音制作单位的专营业务，也不再是音响工程师们垄断的职业。今天，音频素材制作已成了任何单位和个人都“玩”得起的家常“玩意儿”。本书的作者们正是数字音频“弄潮儿”，本书正是我们近年来在数字音频制作的教学和实践中的—些学习体会和经验总结。

为了让读者能较系统全面地了解数字音频技术的全貌，我们在第 1 章中从数字音频技术的基础知识入手，对数字音频技术的应用现状进行了概况性介绍。在此基础上，在第 2 章较全面地介绍了如何应用数字音响设备进行拾音、录音和调音的基本操作方法。在第 3 章和第 4 章中，则针对我国当前的数字音频应用情况，介绍了有代表性的几个数字音频制作与处理软件和 MIDI 制作软件的特点及其使用方法。最后，在第 5 章中，结合多媒体开发的实际和作者们的实践经验，介绍了数字音频素材运用中应注意的艺术和技术问题。

华南师范大学教育信息技术学院的刘毓敏、李剑琴，西北师范大学

教育传播与技术学院的杨晓宏、张雪莉参加了本书的编写，其中，刘毓敏编写了第1章和第2章，以及5.2.1节的部分内容；李剑琴编写了第3章和4.1、4.3、4.4、4.5、5.2.2节，以及5.2.1节的部分内容；张雪莉编写了4.2节和5.1节。全书由刘毓敏策划并统稿。囿于作者的理论水平和实践经验，书中错漏之处在所难免，恳请广大读者批评指正。

作 者

目 录



第 1 章	数字音频技术基础	1
1.1	数字音频技术概况	1
1.1.1	数字音频技术发展概况	1
1.1.2	数字音频技术基础知识	12
1.2	数字音频制作与处理设备	28
1.2.1	前期音频采制器材	28
1.2.2	后期数字音频制作器材	44
1.2.3	计算机音频制作系统	54
第 2 章	声音素材的拾取与数字化采集	70
2.1	声音素材的拾取与数字化采制系统	70
2.1.1	系统的基本形式	70
2.1.2	系统配接和使用中应注意的主要技术问题	72
2.2	录音场所	80
2.2.1	录音场所	80
2.2.2	专用录音场所	83
2.3	拾音技术	90
2.3.1	概述	90
2.3.2	单话筒拾音	104
2.3.3	双声道立体声拾音	107
2.3.4	多话筒拾音	110
2.4	调音技术	111
2.4.1	调音的典型结构	111
2.4.2	电平控制	112
2.4.3	频响控制	112
2.5	录音技术	116
2.5.1	录音技术概况	116
2.5.2	ENG 录音	118
2.5.3	EFP 录音	119
2.6	监听技术	125



第3章 数字音频制作与处理的基本方法	127
3.1 Sound Forge 的使用	127
3.1.1 概述	127
3.1.2 工作界面	128
3.1.3 录音	131
3.1.4 基本音频编辑	132
3.1.5 中级音频编辑	134
3.1.6 效果处理	136
3.2 Samplitude 的使用	140
3.2.1 概述	140
3.2.2 录音	140
3.2.3 音频编辑	145
3.2.4 调音台的使用	150
3.3 Cool Edit Pro 的使用	152
3.3.1 概述	152
3.3.2 音频效果器	154
3.3.3 操作实例	168
3.4 普及型数字音频采制与处理软件的使用	172
3.4.1 数字音频素材的网络下载	172
3.4.2 其他常用数字音频录制处理工具的使用	180
第4章 MIDI 音乐制作的基本方法	195
4.1 MIDI 音乐制作的基础知识	195
4.1.1 乐音的基本概念	195
4.1.2 MIDI 的发展简史	197
4.1.3 MIDI 的若干技术术语	198
4.1.4 MIDI 标准	200
4.1.5 MIDI 音乐制作的设备配置与应用	201
4.1.6 美化 MIDI 音频播放的软件	210
4.2 Cakewalk 的使用	214
4.2.1 Cakewalk 的特点与基本使用方法	214
4.2.2 Cakewalk 应用实务——用 Cakewalk 制作一首乐曲	232
4.3 Soner 的使用	239
4.3.1 概述	239
4.3.2 笔刷工具	239
4.3.3 DXi 软件合成器	240
4.3.4 Rewire 功能	243
4.4 TT 作曲家的使用	245
4.4.1 概述	245

4.4.2 基本设置	246
4.4.3 基本输入	247
4.4.4 其他功能的选择与应用	250
4.5 Storm 的使用	253
4.5.1 概述	253
4.5.2 工作界面	253
4.5.3 乐曲的编辑、保存与输出	257
4.5.4 MIDI 控制与 MIDI 同步	261
4.5.5 其他功能的选择与应用	262
第 5 章 数字音频素材的运用	266
5.1 多媒体作品中音频运用的艺术原理	266
5.1.1 媒体与多媒体	266
5.1.2 影视作品中音频运用的艺术原理	271
5.1.3 多媒体软件中音频运用的艺术原理	275
5.2 数字音频素材运用的技术问题	279
5.2.1 在桌面多媒体影视编辑软件中运用的技术问题	279
5.2.2 多媒体编辑软件中音频的导入与编辑方法	287
5.3 数字音频文件格式的转换	305
5.3.1 CDA 格式转换成 WAV、MP3 格式	305
5.3.2 WAV 格式转换成 MP3、MID 格式	306
5.3.3 WAV、MP3 格式转换成 RA 格式	307
5.3.4 WAV、MP3 格式转换成 WMA 格式	307
5.3.5 MID 格式转换成 WAV 格式	307
5.3.6 其他常用数字音频格式转换/处理软件简介	309
参考文献	312
内容简介	313



第1章 数字音频技术基础

1.1 数字音频技术概况

1.1.1 数字音频技术发展概况

1. 什么是音频

听觉是人类获取信息的主要渠道之一。人从外部世界所获取的信息中,有 20%是通过听觉获得的。人类听觉接收信息主要是通过对声音——频率范围为 20Hz~20kHz 的机械振动(其中人声大约分布在 100Hz~1kHz,而各类乐器声和其他自然响声则分布在全频率范围)的感知而实现的。

做机械振动时能发出声音的物体称为声源。声源振动时,带动其周围的诸如空气、水等介质分子振动,这些介质分子的振动又引起四面八方更远处的介质分子振动……如此这般,声音就能在介质中传播开去。我们把声源振动所导致的其周围介质的振动传播称为声波。声波的物理特性常用波长 λ 、频率 f 和速度 V 三个物理量表征,三者的关系: $\lambda = V/f$ 或 $\lambda = VT$ 。其中: T 是周期,是指声源完成一次全振动所需要的时间,单位为 s(秒);频率 f 是指单位时间内声源所完成的全振动的次数,单位是 Hz(赫兹), $f=1/T$;声速 V 是指声源的某一振动状态在单位时间所传播的距离,单位为 m/s(米/秒),介质不同声波的传播速度是不同的,其在空气中的传播速度约 340m/s;波长 λ 是指声源的某一振动状态在一个周期内所传播的距离,单位为 m(米)。

声音所携带的信息主要表现为机械波的频率(或波长)、振幅、相位等物理特性的变化。人耳正是通过辨别声波的这些状态的变化而获取声音信息的。声波的这些状态变化若转换为电子媒体的状态变化,则这些电子媒体就是携带着声音的上述听觉属性(信息)的信号,称为“音频信号”。

图 1-1 所示的声音电子信息系统较清晰地表达了音频信号运动变化的基本状况。从图可见,音频信号主要有两大来源:一是,通过对携带现实声音信息机械波的拾取(将机械信号转换为电信号)而形成的;另一是,通过机器(如计算机系统)对现实声音的模拟、合成而获得的,这是一种人造的声音信息。

但无论哪一类音频信号,都是通过处理、存储、传输、重发等环节,将电信号转换为适合于人的听觉系统接收的形式——声波,而服务于人。

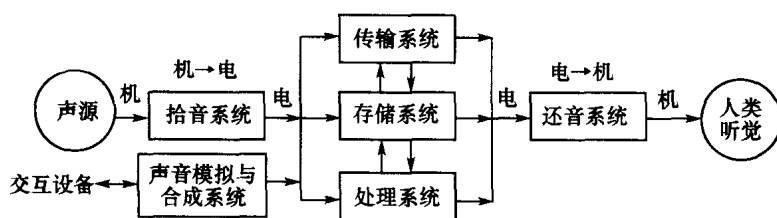


图 1-1 声音电子信息系统

2. 人的听觉特性及音质评价

音频信号的服务对象是人，是人的听觉系统，为了让音频信号发挥更大的效能，必须首先认识人的听觉特性。

1) 音调和响度

音调和响度是声音的重要属性。

(1) 音调与音色 音调是声音的频率这一物理量(客观量)给人的主观感觉，这是主观生理上的等效频率。频率越高，音调越高。具有某种音调的声音称为纯音(或单音)。纯音是单频的声音，其特点是单一音调，乐声包括某些附加的频率，称为复音。在复音中最低频率称为基音。比基音音调高的成分称为泛音。不同的乐器有不同的泛音。某些乐器比另一些乐器产生更多的泛音。泛音对音调增加了有特色的音质，称为音色。不同的人或乐器发声时，其音色不同，这是我们辨别不同人或乐器的听觉依据。

正常的人耳对声音的反应是在 20Hz~20kHz 的音频范围内，当然这一可听频率范围是因人、因年龄而异的。高于 10kHz 的频率对语言的可懂度和欣赏音乐无重大影响。因此，在专业录音或音乐声学设计中，常常以 63 Hz、125 Hz、250 Hz、500 Hz、1kHz、2kHz、4kHz 和 8kHz 为频率标准。

(2) 响度 响度是声音的强度这一物理量(客观量)给人的主观感觉。声音响度与声压(声强)有关，声压越大，响度越大。但人耳对不同频率的声音的响度感觉(灵敏度)是不同的，频率越高，灵敏度越差，而频率很高时，则灵敏度又会降低。声压级为 40dB 对人耳的这一灵敏度特性的响度感觉如图 1-2 所示的曲线。为了模仿人耳的这一灵敏度特性，在测量声压级的仪器中加入了对各种频率具有“计权”性质的网络，由此，可直接读出接近人耳响度感觉的计权声压级，称 A 档 dB 值或 A 声级 dB。

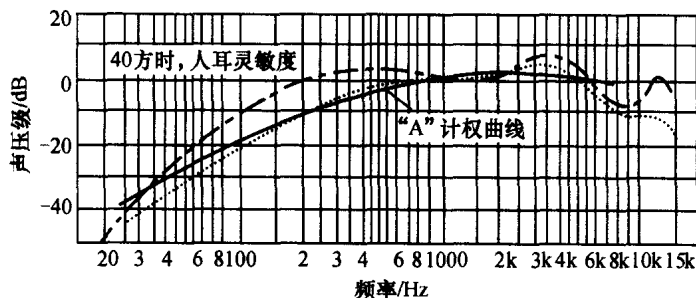


图 1-2 人耳对不同频率声音的灵敏度

此外，人耳所判断的声音响度，同声压级和频率两者都有关系。如，一个 40dB 的 1kHz 纯音，要比声压级相同的 100Hz 纯音响得多。要使 100Hz 声音听起来和 40dB 的 1kHz

声音有同样响度, 必须把声压级提高到 50dB; 反之, 要使 1kHz 声音听起来和 40dB 的 100Hz 声音有同样的响度, 则必须把声压级降低到 25dB。用这种对比试验方法, 可得出图 1-3 所示的一组等响曲线, 每条曲线代表不同频率和声压级的纯音听起来有相同的响度。为了便于说明和区别各等响曲线, 分别加上一个编号, 如图 1-3 中 1kHz 纯音垂直线旁所注明的从 0 到 130 所示。在声学上用“响级”表示各个编号, 并给出一个单位称为“方”。由图 1-3 还可看出, 频率为 1kHz 时, 20 方和 60 方等响曲线之间相差 40dB 的声压级。在 100Hz 时, 两条曲线只相差 40dB。即 1kHz 有 40dB 的变化所引起的响度差别, 在 100Hz 时只有 30dB 的声压级变化便可引起同样的响度差异。

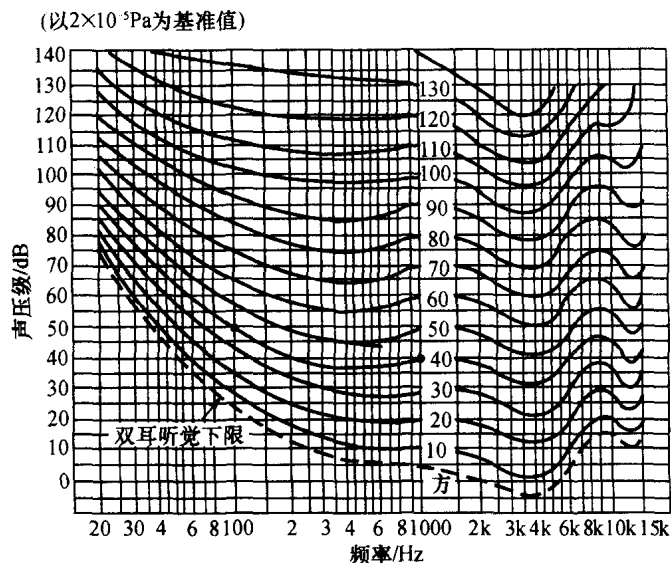


图 1-3 纯音等响曲线

2) 与音质评价有关的几个听觉效应

以下是人类听觉系统中与音质评价有关的几个主要听觉效应。

(1) 双耳听闻效应 人耳在头部的两侧, 约距 20cm, 由于到达双耳的声音有微小的时间差、强度差和相位差, 人们就能辨别声音的方向, 确定声源的位置。特别是左右水平方向的分辨方位能力要比上下垂直方向强得多, 因而声源左右移动时在两耳处引起的声压、时间和相位差别比较明显, 通常可以分辨出水平方向 5° 到 15° 的变动。但在垂直方向, 有时要大到 60° 才能分辨出来。听觉上具有方位感这一特性, 使人们有可能在嘈杂的噪声环境中分辨来自某个方向的一个比较特殊的声音。单耳听闻就不易辨别声音的方位。

双耳效应在电声技术和厅堂声学设计中都占有重要地位。我们今天所熟识的立体声系统, 正是根据人的双耳效应而发展起来的。在演播室、录音室等建筑的声学设计中, 很多声学参数需要考虑双耳效应。例如, 演播室、录音室的混响时间, 由于传声器是单耳听闻的。对此, 应短于同体积、同用途的大厅(音乐录音室应短于同体积音乐厅的混响时间; 对白录音室应短于同体积的报告厅的混响时间等); 噪声标准也是如此, 在录音建筑中噪声的允许值应低于同类型大厅。

(2) 掩蔽效应 演播现场的噪声对语言清晰度的干扰是必须考虑的。人耳具有一种不

寻常的能力,能在噪声环境下有选择地分出他所感兴趣的某些“信号”,而目前的精密仪器还做不到这一点,这是因为人耳对声音除了有方位感外,还有注意力集中的心理因素。例如,我们坐在播放着较响音乐的收音机旁,仍可用不大的声音进行交谈。当然,这时要求注意力集中才能听清对方的讲话,且还容易疲劳。这种排除部分噪声干扰(这里把音乐看作干扰噪声)的能力还和噪声的特性有关。

噪声对语言的妨害程度,在声学上称为“掩蔽效应”,它不仅取决于噪声的总声压级大小,而且还取决于它的频率成分和频率分布,通过实验,我们得到下述规律:

- ① 低音调的声音,特别当响度大时,会对高音调的声音产生较显著的掩蔽作用;
- ② 高音调的声音对低音调的声音只产生很小的掩蔽作用;
- ③ 掩蔽和被掩蔽声音的频率越接近,掩蔽作用越大,当它们的频率相同时,一个声音对另一个声音的掩蔽作用最大。

(3) 哈斯效应 哈斯效应(Haas Effect)是关于人耳对回声的感觉的规律。所谓回声感觉是指从声源直接传来的声音和经一次反射回来的声音,相继到达人耳,其延迟时间小于 30ms 时,一般人耳不能区分出来,仅能觉察到音色和响度的变化。但当两个相继到达的声音时差超过 50ms(相当于直达声与反射声之间的声程差小于 17m),人耳就能判别出它们来自不同方向的两个独立的声音。这后一个来自反射面的声音,有可能成为回声。回声的感觉会妨碍语言和音乐的良好听闻,因此,必须加以控制。

由哈斯通过大量主观评价实验提出了反射声延迟时间与感觉到回声百分率值的关系,见图 1-4。可见,当反射声与直达声具有相同声压级的情况(即 0dB 曲线),当延迟时间小于 30 ms 时,几乎没有人能感觉到回声;延迟时间大于 70 ms,差不多有一半人会感觉出有回声;延迟时间超过 100 ms 时,几乎所有人都能察觉到回声。当反射声强度减弱到直达声以下 10dB 时,即使延迟时间很长,也几乎没有人能感觉到有回声;若把两者的延迟时间压缩得很短,则反射声即使声压级高出直达声 10dB,也不会有回声感觉。

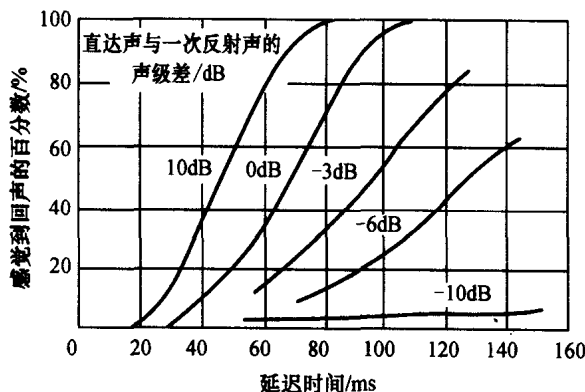


图 1-4 哈斯效应曲线

3) 声音质量的评价

声音质量的评价有客观和主观两大类。

(1) 音质的客观评价 音质评价的客观标准是根据各个物理量来比较的,若话筒前的声场与扬声器前的声场完全相同,则称这样的电声系统的还音质量为客观的真实。若只是根据听觉来比较,认为扬声器发出的声音和原声音有相同的感受,则称这样的电声系

统的还音质量为主观真实。显然,达到客观真实是困难的,但由于听觉的辨别力是有限的,因此在允许畸变条件下的主观真实是可以得到的。表 1-1 是人耳的辨别力和允许畸变的试验研究结果。

表 1-1 人耳辨别力和允许的畸变量

说 明	特 征
声压级辨别力	1dB
频率辨别力	1000Hz 以下为 3Hz, 1000Hz 以上 $\Delta f/f=0.003$
方位辨别力	3°
时差辨别力	0.02s
高频范围变窄	上限 8500Hz, 大约有 50% 听众感觉到
低频范围变窄	下限 100Hz, 30% 听众感觉到; 下限 150Hz, 50% 听众感觉到
频率范围变窄	带宽 50Hz~1000Hz, 约有 30%~35% 听众感觉到
频率范围的平衡感觉	$f_{\text{高}} \times f_{\text{低}} \approx 500000$
峰谷的频率失真	在 200Hz 以下, 1500Hz 以上基本感觉不出, 在 150Hz~200Hz, 6Hz~8kHz 范围内 5dB 的峰或谷有 50% 听众感觉到
失真辨别力	纯音为 0.5%, 音乐为 2%
允许非线性失真	对 40Hz~14000Hz, 7.5% 三次谐波有 50% 听众感觉到, 对 200Hz~4000Hz, 14% 三次谐波有 50% 听众感觉到
允许频率畸变	高低频同时衰减 10dB, 有 65% 听众感觉到

(2) 音质的主观评价 音质的评价也很难完全用客观测试的方法进行,除了目前对听觉本质的认识还没有完全弄清外,客观测量还无法区分音质评价的以下复杂因素:

- 听到的声音是大信号,而测量的多是小信号的结果;
- 测量常常用纯音,而音质评价是用音乐节目;
- 音质评价的结果包含了个人经验、文化修养和爱好,且可通过学习提高;
- 测量结果没有考虑听音环境(如房间)的影响,而听音时却不能不受听音环境的影响。

因此,目前的音质评价还是以主观评价为主。

① 电声系统的音质评价 对电声系统的音质评价通常是在具有规定声学特性的试听室进行的,试听室具有和典型生活住房相当的混响时间,IEC 建议 500Hz 频率的混响时间为 0.4s~0.6s,环境噪声级要低于 40dB,至少包含 7 名不同性别、年龄的试听人员,且以专业人员为主,非专业人员为辅,试听的节目要具有代表性,对音质优劣较易觉察,选择的节目应该既无很大吸引力,又不会令人感到厌倦,且具有语言、戏曲和音乐的特征。目前,常用的音质评价方法有直接评分法和成对比较法两类。其中表 1-2 是成对比较法所用的一些评价术语。而直接评分法对一些单一评语,如响度、音色、平衡、丰满、清晰度等采用 5 级评分制或 7 级评分制来判断。

表 1-2 成对评价术语

响	轻
清晰	模糊
丰满	干涩
亲切	发飘

② 语音音质评价 汉语是单音节的语言，一字一个音节，每个音节由元音(母音)和辅音(子音)组成。元音比辅音容易辨别，因为元音的能量比辅音大得多，持续时间也长得多，频谱特性有明显的特征。平常听报告时主要影响听懂因素是辅音听错，所以辅音对听懂语言有非常重要的作用。

讲话发出的声能大小用语言功率表示，单位为微瓦(μW)。语言功率通常是指某一时刻的瞬时峰值。对于决定清晰易懂的辅音来说，它比平均值或瞬时峰值低得多。一般人讲话时发出的能量极为微小，长时间平均的语言功率约为 $30\sim 50\mu\text{W}$ ，其中有 1% 的时间可达 $500\sim 750\mu\text{W}$ 左右，至于轻声耳语功率仅有 $0.01\mu\text{W}$ 。讲话中从最低到最高的语言功率的变动范围称为动态范围。根据统计分析，在一段时间内讲话的动态范围约比“平均语言声级”高出 12dB，其最弱声音约在平均语言声级以下 18dB。

语音主要是由声带震动产生的，男子声带长而厚，发声频率较低，基音约为 150Hz。女声的基音约为 230Hz。对于歌唱家，男低音基频可低至 55Hz，女高音的基频可高至 1000Hz。同时，发出的许多泛音(谐波)也要高得多，有的甚至超出 6000Hz。由于语音的频率范围并不宽，因此，以语音应用为主的电声设备只要求有 $300\sim 4000\text{Hz}$ 的平直频响范围。

人在讲话时，并非均匀地向四周辐射声音的，而是正面最响，背后最轻，也即沿着嘴唇前面有一定的方向性，频率愈高，方向性愈明显。因此，站在讲话者后面、后侧面的人，由于直射声中缺少很重要的高频成分，很难听懂。

语音音质的评定指标是对语言能够听清的程度，即语音清晰度。在试听室内，语言的清晰度，可用预先规定的字表念出一系列无连贯意义的单音节语言，然后由许多听者将其听到的语音尽可能正确地记录下来，其听得的正确音节的百分数即称为音节清晰度百分率，即：

$$\text{音节清晰度} = \frac{\text{听众正确听到的音节}}{\text{测定用全部音节数}} \times 100\%$$

通常音节清晰度在 85% 以上可以认为满意；60% 以下感到费力、难懂；60%~80% 之间，听众需要注意力集中才能听清。

③ 乐音音质的评价 音乐与语言的声学特性和评价标准有较大差异。

首先，在物理特性方面，音乐(包括声乐)与语言的差别如下：

- 音乐的频率范围比语言宽得多，约从 $40\text{Hz}\sim 15000\text{Hz}$ ，各种乐器也有较大差别；
- 音乐的泛音成分和结构比语言复杂，音色也丰富得多；
- 音乐的节奏变化比语言大得多；
- 乐器声功率一般比讲话大，演唱时能产生较大的声功率；

• 音乐的动态范围比语言大,通常有 50dB 左右,交响乐的动态范围则更大,有时可达 70dB。

其次,乐音音质评价不仅包含客观评价,更重要的是主观评价;不仅包含技术性评价,更重要的是艺术性评价。

最后,乐音评价涉及较多的领域,有些领域(如,音乐电声器材、电子乐器、某些传统乐器等)已经形成一整套专业化评价标准和方法,而许多领域还没有专门的评价标准和方法。

乐音的客观评价中,反映音高的客观量主要有:乐器的音域、音准,音响系统(或电声器材、元器件等)的频带宽度,CD 机或磁带录音机的转速和抖晃等,主要通过各种频率计或计算机辅助测试系统进行测量;反映声强的客观量主要有:电声器材的输出功率、动态范围、信噪比、各声道的均衡等,甚至还包括传声器的灵敏度、指向性等;反映音色的客观量主要有:频率响应特性、共振峰(频谱包络)、失真(如调制失真、谐波失真、交叉频响失真、串音、非线性失真、输入失真、阻抗不匹配引起的失真等)。

乐音的主观评价中,除了音高、声强、音色等分项指标评价外,更重要的是反映各种客观因素的综合效应的整体评价,其中主要是音乐审美的内容。因此,乐音的主观评价除了取决于人的生理心理因素(听觉特性)外,还与人的社会文化心理因素,如文化素质、受教育程度、专业修养、个人爱好、风俗习惯以及临场状态等有关。

3. 音频信号的数字化

1) 为什么要数字化

人类本来是生活在一个物理的、模拟的现实之中。然而,信息时代的信息处理和传输都要求将反映人类各种活动的多种信息形式,无论是数字或文字,也无论是声音、图像或影像,都以数字化的方式来进行。这是一个极大的矛盾。过去的近百年,我们的电话是模拟的,照片是模拟的,电视是模拟的,几乎一切都是模拟的,与现实生活是一致的,我们似乎过得很好,也很容易理解。现在为什么要数字化?从较高的层次看,是因为数字化更能反映信息的本质,因此,能更多、快、好、省地服务于人类,提高人类生活的品质。从具体的技术的角度来看,则是因为数字化的信息与模拟化的信息相比,在存储、检索、处理、传输和利用等各个方面都有着无可比拟的优越性,具体如下:

- ① 能实现高效编码,利于节省存储空间和传输带宽。
- ② 抗干扰能力强,特别是在中继传输时更为明显。
- ③ 可以进行差错控制,因而提高了信息传输存储的可靠性。
- ④ 便于使用现代计算机技术,对信号进行处理、存储和变换,从而提高了信息传输/存储的灵活性。
- ⑤ 便于加密,对所传输/存储的信息进行保密或版权保护。
- ⑥ 易于与其他系统配合使用,构成一个灵活、通用、多功能的综合业务信息传输网。
- ⑦ 易于集成化和大规模生产,其性能一致性好,且成本低。

2) 怎样数字化

PCM 是“数字化”的最基本的技术,模拟信号正是通过 PCM 而变换成数字信号的。PCM 是 Pulse Code Modulation(意为脉冲编码调制,简称脉码调制)的缩略语,其具体操作:通过取样、量化和编码三个步骤,用若干代码表示模拟形式的信息信号(如图像、声音信