



“十二五”国家重点图书出版规划项目  
航空航天精品系列

DIGITAL AUDIO AND VIDEO TECHNOLOGY AND APPLICATIONS

# 数字音视频技术及应用

(第2版)

● 吴韶波 顾奕 李林隼 主编



哈尔滨工业大学出版社  
HARBIN INSTITUTE OF TECHNOLOGY PRESS



“十二五”国家科技计划项目  
航空航天精品系列

DIGITAL AUDIO AND VIDEO TECHNOLOGY AND APPLICATIONS

# 数字音视频技术及应用

(第2版)

● 吴韶波 顾奕 李林隽 主编



哈尔滨工业大学出版社  
HARBIN INSTITUTE OF TECHNOLOGY PRESS

## 内容简介

本书从实际应用出发,以数字音视频技术为主线,系统地介绍了数字音视频技术的基本原理、相关先进技术和具体实际应用。全书共分为9章,对数字音视频的编码技术、传输技术、存储技术进行了系统的阐述;对数字音视频技术在广播电视领域、多媒体通信领域、智能监控领域、机器视觉领域、消费数码产品领域的实际应用进行了全面介绍。本书内容详尽,重点突出,注重理论和实践的结合。

本书可作为电子信息、通信、计算机、自动化、物联网等专业高年级本科生教材或教学参考书,也可供相关领域的技术人员参考使用。

## 图书在版编目(CIP)数据

数字音视频技术及应用/吴韶波,顾奕,李林隽主编.2版.—哈尔滨:哈尔滨工业大学出版社,2016.3

ISBN 978-7-5603-5895-6

I. ①数… II. ①吴…②顾…③李… III. ①数字技术-应用-音频设备-高等学校-教材②数字技术-应用-视频信号-高等学校-教材 IV. ①TN912.27②TN941.3

中国版本图书馆 CIP 数据核字(2016)第 053166 号

策划编辑 王桂芝

责任编辑 李长波

出版发行 哈尔滨工业大学出版社

社 址 哈尔滨市南岗区复华四道街10号 邮编 150006

传 真 0451-86414749

网 址 <http://hitpress.hit.edu.cn>

印 刷 哈尔滨市工大节能印刷厂

开 本 787mm×1092mm 1/16 印张 16.75 字数 408 千字

版 次 2014年9月第1版 2016年3月第2版

2016年3月第1次印刷

书 号 ISBN 978-7-5603-5895-6

定 价 38.00 元

(如因印装质量问题影响阅读,我社负责调换)

# ◎ 再版前言

Preface

随着计算机与通信技术的迅猛发展,数字音视频技术已在数字娱乐、多媒体通信、高清数字电视、宽带网络视频传输等领域得到广泛应用。由于音视频信息,尤其是视频信息十分丰富且信息量大,故对于音视频信号的处理、传输、存储和显示等都提出了新的要求。因此,音视频技术的研究和应用也是目前信息技术领域最热门的问题之一。

本书是2014年出版的《数字音视频技术及应用》的修订版。考虑到网络技术与数字音视频技术的发展,本次修订在保持教材内容先进性的基础上,强调理论与实践的结合,充实了数字视频在视频监控及机器视觉方面应用的新内容,并调整了部分章节的安排,进一步深入体现科学而合理的认知规律,引导读者进行高效率的学习。本书内容可以满足32~64学时的教学。

本书是编者在“数字视频技术”课程教学实践的基础上,结合多年课程教学心得与体会,并参考了近年来数字音视频技术的最新进展编写而成的。

全书内容共分为9章:第1章初步介绍数字音视频技术的基本概念、特点、关键技术与发展情况;第2章为数字音频的研究基础知识,对音频的数字化、音频格式和质量评价方法进行介绍;第3章重点讨论数字音频的压缩编码技术及音频编码的国际编码标准;第4章为数字视频的研究基础,对彩色模型、视频信号采集与表示及质量评价方法进行讨论;第5章针对数字视频的冗余,讨论视频压缩与编码的方法;第6章重点讨论数字视频的编码标准;第7章根据数字音视频对通信网络的要求,讨论传输所涉及的网络技术和调制解调技术;第8章主要介绍基于内容的视频检索标准与应用;第9章对数字音视频技术在广播电视领域、多媒体通信领域、智能监控领域、机器视觉领域、消费数码产品领域的实际应用进行全面介绍。

本书由吴韶波、顾奕、李林隽、龙民共同编写,具体编写分工如下:第1、2、7、8章由吴韶波编写;第3、5、6章由顾奕编写,第4章由龙民编写,第9章由李林隽编写,全书由吴韶波统稿完成。在编写过程中,李红莲、冷俊敏老师在资料查询中给予了帮助和支持,参与了本书的辅助性工作,周金和教授和马牧燕老师对本书中存在的问题给予了指导和建议,在此一并表示感谢。本书在编写过程中参考了大量的文献和资料,书后仅列出主要参考文献,在此特向本书所引用资料的有关作者致以衷心感谢,如有疏漏敬请原谅。

由于数字音视频技术处于快速发展当中,新理论、新技术、新应用层出不穷,相关设备及标准推陈出新,加上编者水平有限,书中疏漏之处在所难免,敬请各位同行专家、读者批评指正。

本书配有电子教案,需要者可联系作者,邮箱 16550942@qq.com。

编者

2016年3月

# ◎ 目录

## Contents

第1章 数字音视频技术概述	1
1.1 数字音视频技术的基本概念	1
1.2 数字音视频技术的特点	2
1.3 数字音视频系统及其关键技术	3
1.4 数字音视频技术的应用	6
1.5 数字音视频技术的发展	8
习题	10
第2章 数字音频基础	11
2.1 声学原理	11
2.2 音频的数字化	16
2.3 声卡的组成与工作原理	20
2.4 数字音频格式	23
2.5 数字音频质量评价方法	26
习题	29
第3章 数字音频压缩及标准	30
3.1 音频冗余	30
3.2 音频压缩编码技术	32
3.3 MPEG 音频编码	45
3.4 杜比 AC-3 音频压缩算法	54
习题	59
第4章 数字视频基础	60
4.1 人类视觉系统	60
4.2 彩色模型	68
4.3 视频表示	77
4.4 电视图像采集与显示	78
4.5 数字视频格式	90
4.6 数字视频质量评价	96
习题	99
第5章 数字视频压缩与编码	100
5.1 视频冗余	100
5.2 预测编码	101

5.3	变换编码 .....	113
5.4	分形编码 .....	122
	习题 .....	126
<b>第6章</b>	<b>数字视频编码标准 .....</b>	<b>127</b>
6.1	视频编码标准简介 .....	127
6.2	JPEG 标准 .....	129
6.3	MPEG 标准 .....	136
6.4	H.26X 标准 .....	165
6.5	AVS 标准 .....	179
	习题 .....	180
<b>第7章</b>	<b>数字音视频信号的传输 .....</b>	<b>182</b>
7.1	数字音视频对通信网的要求 .....	182
7.2	音视频传输网络技术 .....	184
7.3	数字音视频传输常用调制技术 .....	185
7.4	数字电视及其标准 .....	188
	习题 .....	197
<b>第8章</b>	<b>视频信息检索 .....</b>	<b>198</b>
8.1	视频检索技术综述 .....	198
8.2	基于内容的视频信息检索 .....	201
8.3	视频信息检索标准 .....	210
8.4	视频信息检索的应用系统 .....	213
	习题 .....	216
<b>第9章</b>	<b>数字音视频技术应用 .....</b>	<b>217</b>
9.1	数字音视频的未来应用 .....	217
9.2	数字音频技术在数字电视中的应用 .....	225
9.3	手机电视 .....	227
9.4	数字视频监控系统 .....	229
9.5	机器视觉 .....	236
9.6	VoIP .....	242
9.7	IPTV .....	249
9.8	数字音频工作站 .....	258
	习题 .....	261
	<b>参考文献 .....</b>	<b>262</b>

# 第 1 章

## 数字音视频技术概述

### 本章要点:

- ☑数字音视频技术的基本概念
- ☑数字音视频技术的特点
- ☑数字音视频系统的组成
- ☑数字音视频技术的应用与发展趋势

随着计算机技术、网络技术和现代通信技术的逐步渗透与结合,数字音视频技术不再是某个系统具体所指的编码记录格式,而成为有着更加丰富内涵和外延的新概念,在现代信息技术环境下无处不在。因特网传送的流媒体,广电网传送的数字电视,电信网综合数字业务的视频会议和可视电话,移动网的手机电视和视频通话,数码摄像机,VCD,DVD,MP4 等都离不开数字音视频技术。

### 1.1 数字音视频技术的基本概念

数字音视频技术是对音视频信息(文本、图形、图像、声音、动画、视频等)进行采集、获取、压缩、解压缩、编辑、存储、传输及重现等环节全部采用数字化的技术。

声音是通过空气传播的一种连续的波,由空气振动引起耳膜的振动,被人们所感知。人类从外部世界获取信息的 10% 从听觉获得。

音频(Audio)通常指正常人耳所能听到的,相应于正弦声波的任何频率。正常人耳的音频范围一般为 20 Hz ~ 20 kHz,人说话的语音范围常取 300 ~ 3 400 Hz 的频率范围。音频信号可分为两类:语音信号和非语音信号。其中,语音信号是语言的物质载体,是社会交际工具的符号,包含了丰富的语言内涵,是人类进行信息交流所特有的形式;非语音信号主要包括音乐和自然界存在的其他声音形式,不具有复杂的语义和语法信息,信息量低,识别简单。

光辐射刺激人眼时,人眼将接收的光波转换成生物电信号,经过汇聚与处理后通过视觉神经通路送给大脑的视觉感知区域,刺激大脑形成感知,引起复杂的生理和心理变化,这称为视觉。视觉是人类感知信息最重要的途径,从外部世界获取的 70% ~ 80% 信息来源于视觉。

一般而言,图像是由摄影空间中物体对光源的反射和折射在摄影平面上的投影所产生的。视频(Video)泛指将一系列的静态影像以电信号方式加以捕捉、记录、处理、储存、传送及重现的各种技术。视频是按一定时间间隔获取的图像序列,序列中的一幅图像也被称为

一帧(Frame)图像。连续的图像变化每秒超过24帧画面以上时,根据视觉暂留原理,人眼将无法辨别单幅的静态画面,得到平滑连续的视觉效果。

视频分为模拟视频和数字视频。模拟视频是以模拟电信号的形式记录的视频,依靠模拟调幅的手段在空间传播,目前的电视系统仍属于模拟视频系统。数字视频则是依据人的视觉暂留特性,基于数字技术记录,借助计算机或微处理器芯片的高速运算,利用编解码技术、传输存储技术等来实现的以比特流为特征的能按照某种时基规律和标准在显示终端上再现“活动影音”的信息媒介。相比模拟视频,数字视频可以无限次复制也不产生失真,并且可以通过计算机进行编辑与再创作。

## 1.2 数字音视频技术的特点

随着大规模集成电路、计算机数字技术的发展,传统的影视媒体、消费类电子以及通信行业几乎全部实现数字化。数字音视频技术已成为当前最流行、使用最频繁、应用范围最广的新技术,日益深刻地影响着人们的生活方式。与传统模拟技术相比,数字音视频技术具有以下特点:

### (1) 数据量大

声音、图像以及视频和动画的数据量都十分庞大。1 min 立体声音乐采样频率为44.1 kHz,16 位量化精度的数据量大约为10 MB,存储一首4 min 的歌曲约需40 MB;一幅640×480的RGB彩色图像的存储量为900 kB;1 s(25 帧/秒)的视频数据量为22 MB,1张650 MB的CD-ROM光盘只能存储约30 s的视频。

### (2) 数据存在大量冗余

声音、图像以及视频和动画的大量数据中存在着大量的冗余。图像相邻像素之间、视频序列前后帧之间具有很大的相关性,人耳与人眼具有掩蔽效应等听觉和视觉特性,因此,可根据数据的内在联系将数据中的冗余信息去除,通过压缩编码减少数据量。

### (3) 数据存储容量大,传输效率较高

数字音视频数据量大,在存储与传输的过程中必须进行压缩编码。音视频数字信号经过压缩后,可以在6~8 MHz的传输信道传输2~4套标准清晰度电视(SDTV)节目或一套高清晰度电视(HDTV)节目,而一张压缩格式的DVD存储容量可达7~8 GB。

### (4) 便于进行编辑加工

传统磁带重复听某段音乐或观看某段画面时需不停地倒带、快进,编辑过程也是顺序的线性。数字音视频则不同,它可以瞬时定位,非线性逻辑组织,还可以利用非线性编辑软件做特效。

### (5) 信息传输存储的可靠性高

数字信号不会产生噪声和失真的积累,便于存储、控制、修改。数字音视频可以不失真地进行无数次复制,而模拟音视频信号每转录一次,就会有一次误差积累,产生信号失真。模拟音视频长时间存放后质量会降低,而数字音视频可以长时间存放而没有任何失真。

### (6) 有效保护信息和进行版权管理

数字音视频可以方便地与密码及认证技术相结合,便于实现信息加密/解密以及加扰/解扰,适用于专业应用(军用、商用、民用)或条件接收、视频点播、双向互动传送等应用。



(7)具有可扩展性,便于与其他数字设备融合

数字音视频易于与其他系统配合使用,与其他数字设备融合,在各类通信信道和网络上进行传输。易于集成化和大规模生产,其性能一致性好,且成本低。

### 1.3 数字音视频系统及其关键技术

数字音视频系统的传输模型如图 1.1 所示。

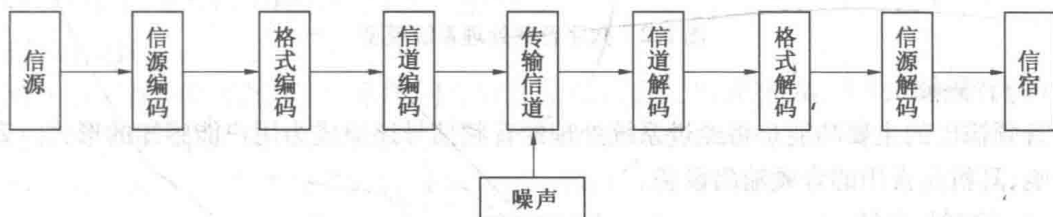


图 1.1 数字音视频系统的传输模型

其中,信源为数字化的语音或视频信号;信源编码旨在通过对信源的压缩、加密、扰乱等处理,用最少的编码传递最大的信息量,即提高通信的有效性,使信号更有效地传输和存储;格式编码提供对数字音频或视频信号在各个格式之间进行转换,使之符合编码的要求;信道编码主要用于提高可靠性,旨在保证信号在传输或存储的过程中尽量不出错,或出错后能够检错甚至纠错,传输信道指信号传输的通道或介质,包括由光缆或电缆构成的有线信道,以及无线通路、微波线路和卫星中继等构成的无线信道。信宿指数字音频或视频重放的终端设备。

数字音视频系统涉及的方面很多,其关键技术包括:

- ①数字音视频信号的获取技术。
- ②数字音视频压缩编码和解码技术。
- ③数字音视频数据的实时处理和特技。
- ④视频和音频数据的输出与存储技术。

#### 1.3.1 数字音频处理系统

数字音频处理系统模型如图 1.2 所示,它是把人耳所能听到的声音信号进行数字化并记录、存储或传输、重放以及其他加工处理等一整套技术,以物理声学、生理声学、心理学、语言学、语音学为基础,涉及电声技术、数字信号处理技术、计算机科学、模式识别和人工智能等相结合的多学科领域。

数字音频处理系统重点研究音频信息的获取、表示、传输与处理(编码、变换、识别、综合、理解、存储)的方法、规律及其利用,包括输入模块、输出模块、存取/通信模块、控制与存储模块以及作为核心的音频处理模块。

##### (1) 音频输入

音频输入是将待处理的声音信号输入到系统中,可通过录音机、录像机、电视机、麦克风等采集声音信号。

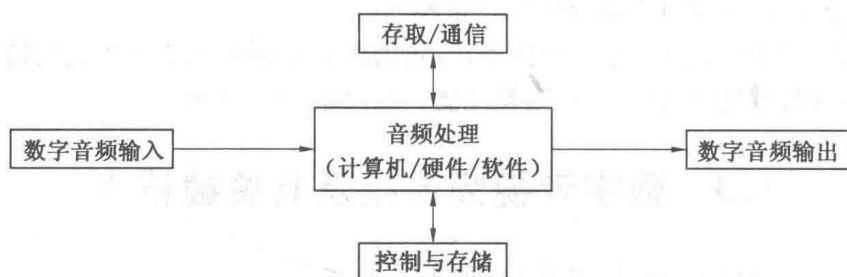


图 1.2 数字音频处理系统模型

### (2) 音频输出

音频输出的主要功能是将经过系统处理的音频信号还原成为用户能感知的形式。音箱和音响、耳机是常用的音频输出设备。

### (3) 控制与存储

控制设备主要用于在处理过程中对音频处理设备进行控制,如键盘、鼠标、各种开关等;存储设备主要用于在处理过程中对音频信号本身以及相关信息进行暂时或永久的保留,如各种 RAM、ROM、硬盘、闪存、光盘等。

### (4) 存取/通信

存取和通信的操作是使用户能够按需将已处理好的,或还需进一步处理的音频信号取出或送入音频处理设备中。存取指本地操作,如光盘或硬盘;通信则是指远端的存取操作,如基于局域网或 Internet 等。

### (5) 音频处理

音频处理分为软件和硬件两类。在计算机中通过软件或硬件来实现对音频的时间处理(混响器和延时器)、音色处理(均衡器和激励器)、动态处理(噪声门和压缩器)。常用音频处理软件如 GoldWave, Adobe Audition 等,硬件如声卡、专用 DSP 等。

## 1.3.2 数字视频图像处理系统

数字视频/图像处理系统的构成与数字音频处理系统类似,也包含五个功能模块,即输入模块、输出模块、存取/通信模块、控制与存储模块和视频/图像处理模块。

### (1) 视频输入设备

视频输入设备主要功能是获取待处理的视频/图像信号。根据应用的不同需求,通常采用摄像头、数字摄像录像机、数码照相机、视频采集卡、扫描仪、红外/X 光摄像机、电视调谐器等不同的设备。

### (2) 视频输出设备

视频输出设备主要功能是将经过系统处理后的视频/图像信号以用户能感知的形式显示出来。目前,最常用的视频/图像输出设备包括阴极射线荧光屏(CRT)、液晶显示屏(LCD)、等离子体显示屏(PDP)、电致发光显示器(ELD)和荧光显示器(VFD)等电子显示设备,以及打印机、绘图仪等硬拷贝设备。

### (3) 视频处理设备

实际的视频处理设备是一个复杂的软、硬件系统。大到分布式计算机组、大型计算机、

工作站,小至一台个人计算机、一块 DSP 芯片。目前,视频处理系统也分为软件型和硬件型。

### 1.3.3 数字音视频系统的性能指标

数字音视频系统的根本任务是存储和传播音视频信息。用户总是希望音视频系统中存储和传输的音视频信息越多越好,数据量越少越好,与原始音视频相比其失真越小越好,操作越方便越好。也就是说,衡量一个数字音视频系统的优劣程度,可以从存储和传输信息的有效性、可靠性、安全性和便利性四个方面进行比较。

#### (1) 有效性

对于数字音视频系统来说,系统的有效性是指信息传输速率  $R_b$  或码元传输速率  $R_s$ 。信息传输速率指单位时间内传输的信息量,单位是比特/秒(bit/s);码元传输速率指单位时间内传输的码元数目,其单位是波特(Baud);二者都反映了在给定信道内所传输的信息量。

根据信息量定义,一个二进制码元含有 1 bit 信息量。在二进制情况下,码元速率和信息速率在数值上相等,但含义和单位不同。一个  $M$  进制码元所含的信息量为

$$1 \text{ Baud} = (\log_2 M) \text{ bit} = k \text{ bit}$$

例 1.1 4 进制的 1 个码元含有 2 bit 的信息量。在多进制情况下,信息速率和码元速率存在以下关系:

$$R_b = R_s \log_2 M$$

另外,为了利用有限的信道带宽支持信源信息量大的通信业务传输,根据信息理论可以采用信源压缩编码,即消除源信息中的冗余部分,如电视信号中只含有大约 4% 的有效信息,采用无失真压缩编码,可能达到 30 多倍的压缩率。更进一步,根据不同应用要求的精度,由香农率失真理论,还可以去掉一些次要信息,即有损压缩编码,压缩率可以达到 100 倍以上,如多媒体会议电视及可视电话可以分别利用 2 Mbit/s 速率及 PCM 系统和 3 kHz 带宽的 PSTN(公用交换电话网)进行传输,便能够满足一般需要。

#### (2) 可靠性

数字音视频系统的可靠性是指接收端所接收到信息的准确程度,通常用误码率或误比特率来进行衡量。误码率是指所传输的信息码元总数中发生差错的码元数目所占的比值,记为  $P_s$ ;误比特率是指所传输的信息比特总数中发生差错的比特数目所占的比值,记为  $P_b$ 。

在二进制情况下, $P_s$  与  $P_b$  在数值上相等,但含义不同。

为提高系统的可靠性,可以采用选择合适的调制技术、改善信道及存储介质以及采用差错控制编码的方法。可见,通信系统的有效性和可靠性是一对矛盾。一般情况下,要增加系统的有效性,就要降低可靠性;反之亦然。在实际中,常常依据实际系统的要求采取相对统一的办法,即在满足一定可靠性指标下,尽量提高消息的传输速率,即有效性;或者,在维持一定有效性的条件下,尽可能提高系统的可靠性。

#### (3) 安全性

数字音视频系统的安全性是指系统对所传信号的加密措施,通过对音视频数据使用的授权,防范非合法授权的接收端使用系统中传输或存储的信息。这点对军用或商用系统尤为重要。

#### (4) 便利性

所谓便利性是指用户在数字音视频系统中按需检索出目标信息的简洁程度。

## 1.4 数字音视频技术的应用

对于不同的应用,人们对数字音视频信息的要求是不同的,并且在选择数字音视频信息编码所采用的技术时也需要了解音视频信息的各种应用。

### 1.4.1 数字音频处理技术的应用

目前,数字音频处理技术主要应用于:

#### (1) 消费电子类数字音响设备

CD唱机、数字磁带录音机(DAT)、MP3播放机以及MD(Mini Disc)唱机已经广泛地应用了数字音频技术。

#### (2) 广播节目制作系统

在声音节目制作系统,如录音、声音处理加工、记录存储、非线性编辑等环节使用了数字调音台、数字音频工作站等数字音频设备。

#### (3) 多媒体应用

数字音频技术在多媒体上的应用体现在VCD、DVD、多媒体计算机以及Internet。VCD采用MPEG-1(活动图像专家组(Moving Picture Experts Group, MPEG)制定了MPEG音频编码标准)编码格式记录声音和图像;DVD-Audio格式支持多种不同的编码方式和记录参数,可选的编码方式包括无损的MLP, DSD, Dolby AC-3, MPEG2-layer2 Audio等,而且是可扩充的、开放的,并可以应用未来的编码技术:Internet上采用MP3的音频格式传输声音,以提高下载能力。

#### (4) 广播电视数字化

在广播电视和数字音频广播系统中,声音编码采用MUSICAM编码方法,符合MPEG-1 Layer 1高级音频编码。如当今的数字电视采用的音频标准就是Dolby AC-3和MPEG-layer2。

#### (5) 通信系统

为了提高通信系统的有效性,在通信系统中传输音频信息必须先对音频进行压缩。传统的PSTN电话中采用的是G.711和G.726的标准;2G移动通信技术的GSM采用的是GSM HR/FR/EFR标准,CDMA采用的是3GPP2 EVRC, QCELP8k, QCELP16k, 4GV标准;3G中的WCDMA采用的是3GPP AMR-NB, AMR-WB标准。另外,在IPTV和移动流媒体中,采用的是AMR-WB+和AAC的标准。

总之,根据应用场合的不同,可以将数字音频编码分为如下两种编码:

①语音编码:针对语音信号进行的编码压缩,主要应用于实时语音通信中减少语音信号的数据量。典型的编码标准有ITU-T(International Telecommunication Union-Telecommunication) G.711, G.722, G.723.1, G.729; GSM HR, FR, EFR; 3GPP AMR-NB, AMR-WB; 3GPP2 QCELP8k, QCELP 13k, EVRC, 4GV-NB等。

②音频编码:针对频率范围较宽的音频信号进行的编码,主要应用于数字广播和数字电

视广播、消费电子产品、音频信息的存储、下载等。典型的编码标准有 MPEG-1/MPEG-2 的 layer 1、2、3 和 MPEG-4 AAC 的音频编码。还有最新的 ITU-T G. 722.1、3GPP AMR-WB+ 和 3GPP 2.4G V-WB, 它们在低码率上的音频表现也很不错。

### 1.4.2 数字视频处理技术的应用

自 20 世纪 90 年代以来,以计算机和软件为核心的数字化技术取得了迅猛发展,文字、图像、声音和视频都被进行数字化处理,而数字化的视频无疑是其中最具有挑战性的部分,且应用领域宽广,目前的主要应用体现在以下几个方面:

#### (1) 数字电视

数字电视是指从节目采集、节目制作、节目传输直到用户端都以数字方式处理信号的电视类型。基于 DVB 技术标准的广播式和“交互式”数字电视,采用先进用户管理技术能将节目内容的质量和数量做得尽善尽美并为用户带来更多的节目选择和更好的节目质量效果。

数字电视可以通过卫星、有线电视电缆、地面无线广播等途径进行传输。与模拟电视相比,数字电视具有图像质量高、节目容量大(是模拟电视传输通道节目容量的 10 倍以上)和伴音效果好、节目更丰富等特点,市场前景非常广阔。

#### (2) 可视通信

数字可视通信是一种创新的、全业务的 IP 数据业务网,它利用已有的宽带网络作为数据传输平台,以话音通信、视频通信和数据通信为基本手段,以信息存储、转发、应用、共享为可选手段,通过各种可视通信终端,向用户提供双向视频和双向音频的网络 IP 电话、视频家庭监控、大楼和小区的可视对讲、个人信息存储转发、视频点播等多种信息服务,适用于军事、国防、公安、会议及个人或家庭娱乐等多个应用领域。

#### (3) 电子出版

电子出版是指在整个出版过程中,从编辑、制作到发行,所有信息都以统一的二进制代码的数字化形式存储于磁、光、电等介质中,信息的处理与传递借助计算机或类似的设备来进行的一种出版形式。

电子出版物具有体积小、容量大、保存方便、检索容易等优点。一张普通的容量为 650 MB 的 CD-ROM 可以存储约 3 亿汉字(按每个汉字 2 个字节计算)的书籍,可以存储约 50 min 采用 MPEG-1 标准压缩的通用中间格式(Common Intermediate Format, CIF)的视频及其伴音。随着更大容量的光盘问世以及压缩技术的进步,电子出版必将成为出版的主要方式。目前,DVD 已经成为电影电视的重要出版方式。

#### (4) 多媒体咨询服务

使用多媒体咨询系统,用户可以方便地找到自己需要的信息,例如新闻、金融资讯、天气预报、交通、旅游、购物,以及自己感兴趣的电影电视节目等。通常,查询终端采用触摸屏操作系统,无须使用键盘,支持手写输入;用户可将自己选择的内容通过蓝牙和多种接口下载到手机或自备 U 盘等存储设备上,以便于随时使用。

#### (5) 多媒体家用电器

现在 VCD、DVD、数码照相机、数码摄像机、MP3 播放机等已经进入并影响着人们的生活。

### (6) 手机视频技术

随着 3G 时代的到来,手机视频业务越来越多地受到人们的关注,其发展拥有广阔的未来。手机视频技术主要应用于两个方面:手机电视和手机监控。所谓手机电视业务,从用户的角度来看,就是利用手机终端观看电视的一种业务。这种业务最初是通过采用传统移动流媒体的方式来实现的。随着移动数据业务的普及、手机性能的提高以及数字电视技术和网络的迅速发展,目前采用在手机上实现以广播的形式接收广播质量的音视频内容。手机视频监控技术突破了网络带宽限制的瓶颈后,已具备了良好的视频清晰度和实时性,不但具备了固定线路视频监控的图像功能,还从根本上满足了用户随时随地远程观看实时监控视频的需求;与 PC 相比手机价格较低,具有良好的成本优势和安全优势,便于用户管理。

## 1.5 数字音视频技术的发展

### 1.5.1 语音编码标准的发展

国际电信联盟(ITU)主要负责研究和制定与通信相关的标准,作为主要通信业务的电话通信业务中使用的语音编码标准均是由 ITU 负责完成的。其中用于固定网络电话业务使用的语音编码标准如 ITU-T G.711 等主要在 ITU-T SG 15 完成,并广泛应用于全球的电话通信系统之中。随着 Internet 网络的快速发展,ITU-T 将研究和制定变速率语音编码标准的工作转移到主要负责研究和制定多媒体通信系统、终端标准的 SG 16 中进行。

在欧洲、北美、中国和日本的电话网络中通用的语音编码器是 8 位对数量化器(相应于 64 kbit/s 的比特率),该量化器所采用的技术在 1972 年由 CCITT(ITU-T 的前身)标准化为 G.711。1983 年,CCITT 规定了 32 kbit/s 的语音编码标准 G.721,其目标是在通用电话网络上应用(标准修正后称为 G.726)。这个编码器价格虽低但却提供了高质量的语音。

至于数字蜂窝电话的语音编码标准,在欧洲,TCH-HS 是欧洲电信标准研究所(ETSI)的一部分,负责制定数字蜂窝标准。在北美,这项工作是由电信工业联盟(TIA)负责执行。在日本,由无线系统开发和研究中心(称为 RCR)组织这些标准化的工作。此外,国际海事卫星协会(Inmarsat)是管理地球上同步通信卫星的组织,也已经制定了一系列的卫星电话应用标准。

目前,业界在语音编码领域已取得了很多重要的进展,现在的研究焦点一方面是在保证语音质量的前提下,降低比特率,主要的应用目标是蜂窝电话和应答机;另一方面是对传统的语音编码器进行全频带扩展,使其适应音频的应用。除此之外,为适应在 Internet 上传送语音的需要,ITU-T SG 16 组研究和制定了可变速率的语音编码标准,变速率的语音编码将是近期语音编码发展的一个趋势。

### 1.5.2 音频编码标准的发展

音频编码标准主要由 ISO 的 MPEG 组来完成。MPEG-1 是世界上第一个高保真音频数据压缩标准,是针对最多两声道的音频而开发的。但随着技术的不断进步和生活水准的不断提高,立体声形式已经不能满足听众对声音节目的欣赏要求,具有更强定位能力和空间效果的三维声音技术得到蓬勃发展。而在三维声音技术中最具代表性的就是多声道环绕声技

术。目前有两种主要的多声道编码方案: MUSICAM 环绕声和杜比 AC-3。MPEG-2 音频编码标准采用的就是 MUSICAM 环绕声方案,它是 MPEG-2 音频编码的核心,是基于人耳听觉感知特性的子带编码算法。而美国的 HDTV 伴音采用的是杜比 AC-3 方案。MPEG-2 规定了两种音频压缩编码算法,一种称为 MPEG-2 后向兼容多声道音频编码标准,简称 MPEG 2BC;另一种是高级音频编码标准,简称 MPEG-2 AAC,由于其与 MPEG-1 不兼容,也称 MPEG NBC。

MPEG-4 的目标是提供未来的交互多媒体应用,它具有高度的灵活性和可扩展性。与以前的音频标准相比,MPEG-4 增加了许多新的关于合成内容及场景描述等领域的工作。MPEG-4 将以前发展良好但相互独立的高质量音频编码、计算机音乐及合成语音等第一次合并在一起,并在诸多领域内给予高度的灵活性。

随着以 IPTV 业务为代表的信息检索业务的开展,适合于在 IP 网络上传输的音频信号编码技术,用于制作、检索和存储音频信息的技术将成为数字音频的发展方向。

从发展的角度看,多媒体技术将继续以一个整体的形象出现在广大消费者的面前。数字音频编解码技术作为其中的一部分也将越来越受到整体化和网络化的影响,各种算法将进一步相互融合,取长补短,趋于统一化和工具化。在技术方面,音频信号压缩将进一步在模式认知与声音信号的分类、合成、算法优化、技术构造等方面进行深入研究。

### 1.5.3 数字视频处理技术的发展

随着计算机技术和网络技术的发展,信息高速公路的建设,以及多媒体的推广应用,各种视频资料源源不断地产生,随之建立起了越来越多的视频数据库,出现了数字图书馆、数字博物馆、数字电视、视频点播、远程教育、远程医疗等许多新的服务形式和信息交流手段。

视频技术是多媒体应用的核心技术,数字视频由于数据量可压缩的信息量最多,压缩处理后的视频质量决定了多媒体服务质量的好坏。视频压缩编码技术已制定了一些国际标准,如 ITU-T H. 261, H. 263, 及 ISO/IEC 的 MPEG-1 和 MPEG-2,覆盖了很大的视频速率范围和应用领域,支持不同速率、不同的图像质量要求等条件的视频业务,能够满足包括电视会议、视频电子邮件、可视电话、广播级视频应用等不同要求的服务。随着视频应用需求的不断发展,视频压缩技术也有了很大的提高,新出现的压缩标准有了更高的压缩效率(在相同的图像质量下需要更低的传送码率或在相同的传输速率下提供质量更好的图像),同时支持不同的传输速率以适应不同的传送网络。

在目前已经大量应用的压缩标准中,MPEG-1 和 MPEG-2 是面向广播级或准广播级应用的。MPEG-1 标准主要是为了视频存储媒体如 VCD 而制定,该标准能够适应变码流的处理,其主要目的是在 1~1.5 Mbit/s 的情况下,提供 30 帧 CIF(352×288) VHS 质量的图像。MPEG-1 的实时编码通常需要硬件才能完成,解码可以用软件来完成。MPEG-1 不能提供分级图像编码,也不能在丢包率高的情况下应用。

MPEG-2 标准扩展了 MPEG-1 标准,能够支持高分辨率图像和声音。目标码率是在 3~15 Mbit/s 传输速率条件下提供广播级的图像,而且能够提供 SNR、时间、空间三种分级编码。该标准应用于卫星广播时,在当前的一个模拟信道中,不牺牲质量的情况下能提供五路数字的编码节目。

H. 261 与 H. 263 标准主要面向于低码率的视频应用,如可视电话和会议电视。H. 261

是最早出现的视频编码标准,它的输出码率是 64 kbit/s 的倍数。H. 261 主要是为了 ISDN 的会议电视和可视电话的应用,它所需要的计算量能够显著下降。这种算法通过均衡图像质量和运动来优化带宽,所以图像快速运动时质量会下降。H. 261 的输出速率是恒定的,而图像质量非恒定。

H. 263 是为了支持低速率的通信而制定的标准,但希望能够适应较大的动态范围,而不仅限于低码率,而且能取代 H. 261。H. 263 能适应误码率高的信道,具备容错的能力。

由于公用电话网和无线网络上的传输速率仍然很有限,而且误码率高,上述标准不能满足高压缩率和强信道容错能力的应用要求。新的压缩标准应运而生,如 H263+和 MPEG-4 标准。

H263+以及后来的 H263++,H26L 能很好地解决低码率视频应用问题,在提高编码压缩效率的同时,提高码流对高误码率信道的容错能力,方便灵活,且能够兼容本标准的以前版本,由于实现成本较低,H263+标准已经越来越多地被采用。

MPEG-4 标准是目前压缩标准的主流,既能够支持码率低于 64 kbit/s 的视频应用,也能够支持广播级的视频应用。与其他压缩标准相比,MPEG-4 标准在 DCT 的基础上引入了图像模型的概念,从而具有更高的压缩效率。

发展中的视频处理技术领域在不断扩大,未来的研究主要集中在以下几个方面:

- ①对视频流中预期目标的实时检测。
- ②对数字视频中特定感兴趣特征的检索。
- ③将场景视频流中的自然人与人工制作的物体合成。
- ④视频数据库管理中的摘要生成。
- ⑤将视频流分割成有代表性的镜头。

## 习 题

1. 数字音视频技术有哪些特点?
2. 简述数字化视频处理系统的基本组成。
3. 数字音视频系统有哪些性能指标?
4. 简述数字音视频技术未来的发展方向。



## 第 2 章

# 数字音频基础

### 本章要点:

- ☑ 数字音频的声学原理与人耳的听觉特性
- ☑ 模拟音频数字化过程
- ☑ 影响音频质量的技术参数
- ☑ 主要的数字音频格式
- ☑ 数字音频质量评价方法

在生活中,声音被分为无规则的噪声和有规则的音频信号;有规则的音频信号是一种连续变化、周期性的模拟信号。通常所说的声音就是指有规则的音频信号。

## 2.1 声学原理

### 2.1.1 声音的物理特性

任意时刻,模拟声波信号都可以分解为一系列正弦波的线性叠加。声波信号由基音和泛音组成,其中频率最低的声波称为基频或基音,其他声波称为泛音,其频率是基频的整数倍。

从物理角度看,描述声音信号参数有频率、声速和波长;从听觉角度看,声音具有音调、响度和音色三个要素。

#### (1) 音调(Pitch,也称为声调)

音调是指声音的高低。音调与声音的频率有关,声源振动的频率越高,声音的音调就越高;声源振动的频率越低,声音的音调就越低。通常把音调高的声音称为高音,反之称为低音。

#### (2) 响度(Loudness)

响度即声音的响亮程度,用声压描述大气压变化的幅度,单位是帕斯卡(Pa),与声音的振幅有关,取决于声波信号的强弱程度。由于人的听觉响应与声音信号的强度不是线性关系,一般采用声音信号幅度取对数后再乘以 20 来描述,称为声强,单位是分贝(dB)。声强与声压的关系如图 2.1 所示。

#### (3) 音色(Timbre)

音色是指人耳对各种频率和响度的声波的综合反映,与声音波形无关,取决于声波的频谱,由混入基音的泛音决定。各阶谐波即泛音的幅度比例不同,随时间衰减的程度不同,则