



普通高等教育“十一五”国家级规划教材
电子信息科学与工程类专业精品教材

*Electronic Information
Science and Engineering*

多媒体通信技术基础

Fundamentals of Multimedia Communication

Fourth Edition

(第4版)

© 蔡安妮 等编著



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>



普通高等教育“十一五”国家级规划教材

电子信息科学与工程类专业精品教材

多媒体通信技术基础

(第4版)

蔡安妮 等编著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书是普通高等教育“十一五”国家级规划教材。作者紧密跟踪国际上多媒体技术发展的动向和研究成果,在分析本领域内大量具有代表性的文献及书籍的基础上,结合多年的科研和教学经验,综合提炼出本书的大纲和内容。全书比较全面地介绍了这一新领域内的主要理论与技术,内容包括:多媒体技术的特征、视觉特性与彩色电视信号、数据压缩的基本技术、视频数据的压缩编码、音频数据的压缩编码、多媒体传输网络、多媒体同步与数据封装、多媒体通信终端与系统、视频数据的分组传输、视频在异构环境中的传输等。全书在理论上力求严谨、叙述上尽量深入浅出。

本书选材兼顾到研究生及本科生教学两个方面的教学需要,同时可以作为从事通信、信息及相关行业的科研和工程技术人员的参考。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

多媒体通信技术基础 / 蔡安妮等编著. — 4 版. — 北京: 电子工业出版社, 2017. 8
ISBN 978-7-121-31957-0

I. ①多… II. ①蔡… III. ①多媒体通信—通信技术—高等学校—教材 IV. ①TN919.85

中国版本图书馆 CIP 数据核字(2017)第 139695 号

策划编辑: 陈晓莉

责任编辑: 凌 毅 文字编辑: 陈晓莉

印 刷: 北京季蜂印刷有限公司

装 订: 北京季蜂印刷有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 787×1092 1/16 印张: 21.75 字数: 750 千字

版 次: 2000 年 8 月第 1 版

2017 年 8 月第 4 版

印 次: 2017 年 8 月第 1 次印刷

定 价: 49.00 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010)88254888,(010)88258888。

质量投诉请发邮件至 zlts@phei.com.cn,盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式:(010)88254540,chenxl@phei.com.cn。

第4版前言

本书自2000年出版第一版以来,已经是第4版了。在这十几年中,作者一直是按照下述方向努力的:首先阅读这个新技术领域中绝大多数有代表性的论文、书籍,并根据自己的科研和教学经验,经过消化、提炼,分门别类,使其系统化;然后用自己的语言,尽量严谨、通俗、简繁适当地写出来。

本书选材时尽量兼顾了本科生教学和研究生教学两方面的需要,以及在通信、信息及相关领域中从事研究开发的工程技术人员掌握多媒体技术的需要。在高等院校使用本教材时,可根据本校、本系的特点,选择书中的不同章节、或不同章节中的部分内容分别作为本科生、研究生讲课使用。为了加深理解和拓展各章中的内容,每章之后都附有一定数量的习题和参考文献。

本书第4版的修订主要沿着如下三条主线进行:

第一条主线是视频压缩编码。由于近年来技术的发展,H.264/AVC和H.265/HEVC正在逐步取代以往的压缩编码标准,因此我们在第2章删去了模拟彩色电视的细节,增加了高清晰度和超高清晰度数字电视信号。在第4章大大简化了对早期编码标准的介绍,而将H.264和H.265分别列为单独的一节;在率失真优化的模式选择中,突出了在H.265中的应用;在速率控制中删去了某些具体算法的细节,而将其总结为Q域、 ρ 域和 λ 域三类控制算法,其中 λ 域方法是在H.265速率控制中提出的。此外,在第10章增加了一节对H.265可伸缩性编码(SVC)的介绍。H.265的SVC采用了与以前的SVC颇不同的结构,可以期望这种改变能够促进SVC的市场化推广。

第二条主线是将多媒体同步和传输层协议合并成新的一章——多媒体同步与数据封装。这一章删去了一些不实用的同步方法,分别讨论了用于广播应用的MPEG 2 TS、用于分组网传输应用的RTP/RTCP和用于存储应用的多媒体文件等数据封装方式以及同步方法,从而自然地引出MPEG最新提出的MMT标准。该标准适用于以上三类应用的数据封装,是下一代的多媒体传输协议。

第三条主线是多媒体终端与系统。原先对各种系统的讲述有些凌乱,这次从技术的角度将其梳理为两大类,分别为会话与会议系统和流媒体系统;后者又分为基于RTP/UDP的系统、基于HTTP/TCP的系统,以及复盖广阔地域的系统(CDN、P2P和基于云的CDN)。此外,在第9章增加了讨论流媒体系统自适应速率控制的一节。

其余各章也做了部分修改,并增删了少量习题和参考文献。

本书第5章由苏菲编写;第3.8节、第10.2.3节和第10.3节由庄伯金编写;其余由蔡安妮编写。

本书取材于众多的文献,作者在此对这些推动多媒体技术发展的人们表示敬意;同时,对给本书提出意见和建议的老师和学生表示感谢。最后,作者仅以本书缅怀北京邮电大学多媒体通信与模式识别实验室创建人、本书第一版作者之一孙景鳌教授,由于他在实验室对多媒体方向科研的倡导和坚持,才有今天这本书。

作者

2017年7月

目 录

第1章 概论——多媒体技术的特征	1	2.4.4 亮度信号与色差信号	24
1.1 概述	1	习题二	25
1.2 多媒体的概念与含义	1	参考文献	26
1.3 多媒体产生的技术背景	3	第3章 数据压缩的基本技术	27
1.3.1 图像压缩编码技术的成熟	3	3.1 概述	27
1.3.2 大规模集成电路技术的发展	4	3.2 数据压缩的理论依据	27
1.3.3 大容量数字存储技术的发展	4	3.2.1 离散信源的信息熵	27
1.4 多媒体系统的基本类型	5	3.2.2 信源的概率分布与熵的关系	28
1.4.1 独立商亭式系统	5	3.2.3 信源的相关性与序列熵的关系	29
1.4.2 多媒体信息检索与查询	5	3.3 信息率—失真理论	30
1.4.3 多媒体会议与协同工作	7	3.3.1 通信系统的一般模型	30
1.4.4 多媒体即时通信	8	3.3.2 信息率—失真函数	31
1.4.5 视频点播与直播	9	3.3.3 限失真信源编码定理	32
1.5 网络融合与业务融合	10	3.4 取样频率的转换	33
1.5.1 网络的融合	11	3.4.1 下取样	34
1.5.2 多媒体业务的融合	12	3.4.2 上取样	35
习题一	12	3.4.3 分数比率转换	37
参考文献	12	3.5 预测编码	37
第2章 视觉特性和彩色电视信号	13	3.5.1 差分脉冲编码调制(DPCM)	38
2.1 人的视觉特性	13	3.5.2 序列图像中运动矢量的估值	40
2.1.1 图像对比度与视觉的对比度 灵敏度特性	13	3.5.3 具有运动补偿的帧间预测	44
2.1.2 空间频率与视觉的空间频率响应	14	3.6 正交变换编码	46
2.1.3 视觉的时间域响应	15	3.6.1 最佳线性正交变换	46
2.1.4 彩色的计量和彩色视觉	16	3.6.2 离散余弦变换	48
2.2 扫描	18	3.7 子带编码	50
2.2.1 空间频率到时间频率的转换	18	3.7.1 子带编码的工作原理	50
2.2.2 隔行扫描与逐行扫描	18	3.7.2 正交镜像滤波器组	52
2.3 模拟彩色电视信号	19	3.7.3 时域混叠消除	53
2.3.1 电视信号的带宽	19	3.8 小波变换编码	55
2.3.2 彩色空间的处理	20	3.8.1 多尺度分析	56
2.3.3 模拟彩色电视制式	21	3.8.2 二进小波变换	57
2.4 数字电视信号	22	3.8.3 变换系数的排序和编码	58
2.4.1 取样结构	22	3.9 量化	61
2.4.2 彩色空间	23	3.9.1 均匀量化器	61
2.4.3 主要参数	23	3.9.2 最小均方误差量化器	62
		3.9.3 最小熵量化器	63

3.9.4	自适应量化	64	4.6.3	Q 域速率控制	115
3.9.5	预测误差和 DCT 系数的量化	64	4.6.4	ρ 域速率控制	116
3.10	熵编码	65	4.6.5	λ 域速率控制	117
3.10.1	熵编码的基本概念	65	4.7	变速率视频编码	118
3.10.2	霍夫曼编码	66	4.8	压缩编码算法性能的评价	120
3.10.3	算术编码	68	习题四		122
习题三		70	参考文献		124
参考文献		72	第 5 章 音频数据的压缩编码		125
第 4 章 视频数据的压缩编码		74	5.1	概述	125
4.1	基于帧的视频编码	74	5.2	人的听觉特性	125
4.1.1	典型的编码器与解码器	74	5.2.1	响度级和响度	125
4.1.2	视频序列的编码	79	5.2.2	听觉灵敏度	126
4.1.3	帧内预测编码	79	5.2.3	听觉掩蔽	127
4.1.4	帧间预测编码的优化	81	5.2.4	临界带宽	128
4.1.5	码流结构	85	5.3	音频信号的数字化	128
4.2	视频压缩编码的国际标准	86	5.4	音频自适应差分脉冲编码调制	129
4.2.1	国际标准的制定	86	5.5	音频子带编码	129
4.2.2	几个术语	87	5.6	线性预测编码	131
4.2.3	早期的国际标准	87	5.7	码激励线性预测编码(CELP)	131
4.2.4	其他编码标准	89	5.7.1	感知加权滤波器	132
4.3	H.264/AVC	90	5.7.2	合成分析法	132
4.3.1	帧间预测编码	90	5.7.3	CELP 编解码原理	133
4.3.2	图像的帧内预测编码	92	5.7.4	G.729 编解码器	133
4.3.3	低计算复杂度的变换编码与量化	92	5.8	感知编码	135
4.3.4	基于上下文的自适应变长编码	95	5.9	MPEG-1 音频编码	136
4.3.5	基于上下文的自适应算术编码	96	5.9.1	概述	136
4.3.6	去块效应滤波器	97	5.9.2	MPEG-1 心理声学模型	137
4.3.7	其他特点	97	5.9.3	编码层次	137
4.4	H.265/HEVC	98	5.10	MPEG-2 音频编码	140
4.4.1	编码器与图像的划分	98	5.10.1	MPEG-2 BC	141
4.4.2	帧间预测编码	101	5.10.2	MPEG-2 AAC	142
4.4.3	图像的帧内预测编码	104	5.11	杜比(AC-3)编码	143
4.4.4	变换编码与量化	104	5.12	音频压缩编码的国际标准	145
4.4.5	熵编码	106	习题五		147
4.4.6	环路滤波器	106	参考文献		147
4.4.7	其他特点	108	第 6 章 多媒体传输网络		148
4.5	基于率—失真优化的编码模式选择	108	6.1	概述	148
4.5.1	率—失真优化的基本方法	108	6.2	多媒体信息传输对网络的要求	148
4.5.2	H.265 的率—失真优化方法	110	6.2.1	性能指标	148
4.6	恒定速率编码器的速率控制	112	6.2.2	网络功能	151
4.6.1	速率控制的例子	112	6.2.3	服务质量(QoS)和用户体验质量(QoE)的保障	152
4.6.2	速率控制的基本步骤	114			

6.3 网络类别	153	7.3 多媒体数据时域特征表示	198
6.3.1 电路交换网络和分组交换网络	153	7.3.1 时域场景和时域定义方案	198
6.3.2 面向连接方式和无连接方式	154	7.3.2 时域参考框架	199
6.3.3 资源预留、资源分配和资源独享	155	7.3.3 同步容限	200
6.4 电路交换广域网对多媒体信息传输的支持	156	7.4 分布式多媒体系统中的同步	200
6.4.1 电路交换广域网	156	7.4.1 分布式多媒体系统的结构	201
6.4.2 多点控制单元	156	7.4.2 影响多媒体同步的因素	201
6.5 局域网对多媒体信息传输的支持	157	7.4.3 多级同步机制	203
6.5.1 共享介质的局域网	157	7.5 连续媒体同步的基本方法	203
6.5.2 吉比特以太网	159	7.6 同步算法举例	205
6.5.3 以太网 QoS 保障	161	7.6.1 基于播放时限的流内同步算法	205
6.6 IP 网对多媒体信息传输的支持	161	7.6.2 基于缓存数据量的流内同步算法	206
6.6.1 传统的 IP 网(IPv4)	162	7.7 广播应用的传输层协议	208
6.6.2 IP 多播	163	7.7.1 MPEG2 TS 数据封装与复接/分接	208
6.6.3 新一代 IP(IPv6)	165	7.7.2 MPEG2 TS 定时信息及同步	211
6.7 IP QoS 保障机制	166	7.8 宽带应用的传输层协议	213
6.7.1 综合服务模型与资源预留协议 RSVP	166	7.8.1 应用层分帧和集成层次处理	213
6.7.2 区别服务	168	7.8.2 传统的因特网传输层协议 TCP 与 UDP	214
6.7.3 多协议标记交换(MPLS)	170	7.8.3 实时传输层协议 RTP 与 RTCP	215
6.7.4 流量工程和基于约束的寻径	171	7.8.4 RTP/RTCP 定时信息及同步	218
6.8 无线网络对多媒体信息传输的支持	172	7.8.5 包长的选择与包头压缩	219
6.8.1 无线传输的特点	172	7.9 多媒体文件	221
6.8.2 无线局域网(WiFi)	173	7.9.1 支持多媒体流式应用的文件	221
6.8.3 蜂窝移动通信网	178	7.9.2 MP4 文件格式	222
6.8.4 无线城域网(WiMAX)	181	7.9.3 MP4 的同步信息	223
6.8.5 无线网络中多媒体传输的特殊问题	184	7.10 编码层与传输层的适配	224
6.9 宽带用户接入网	185	7.11 下一代多媒体传输协议	225
6.9.1 数字用户线路	185	7.11.1 对下一代传输协议的需求	225
6.9.2 光缆接入	187	7.11.2 MPEG 多媒体传输协议(MMT)	226
6.9.3 光缆-同轴电缆混合接入	188	7.11.3 MMT 定时信息及同步	231
6.10 网络融合的水平模型	193	习题七	232
习题六	193	参考文献	233
参考文献	194	第 7 章 多媒体同步与数据封装	195
第 7 章 多媒体同步与数据封装	195	7.1 概述	195
7.1 概述	195	7.2 多媒体数据	195
7.2 多媒体数据	195	7.2.1 连续媒体数据与静态媒体数据	195
7.2.1 连续媒体数据与静态媒体数据	195	7.2.2 多媒体数据内部的约束关系	196
7.2.2 多媒体数据内部的约束关系	196	7.2.3 多媒体数据的构成	198
7.2.3 多媒体数据的构成	198	第 8 章 多媒体通信终端与系统	234
7.3 多媒体数据时域特征表示	198	8.1 概述	234
7.3.1 时域场景和时域定义方案	198	8.2 国际电联 H 系列的会话与会议系统	234
7.3.2 时域参考框架	199	8.2.1 视听通信终端的一般框架	234
7.3.3 同步容限	200	8.2.2 同步电路交换网视听业务标准(H.320)	234
7.4 分布式多媒体系统中的同步	200		
7.4.1 分布式多媒体系统的结构	201		
7.4.2 影响多媒体同步的因素	201		
7.4.3 多级同步机制	203		
7.5 连续媒体同步的基本方法	203		
7.6 同步算法举例	205		
7.6.1 基于播放时限的流内同步算法	205		
7.6.2 基于缓存数据量的流内同步算法	206		
7.7 广播应用的传输层协议	208		
7.7.1 MPEG2 TS 数据封装与复接/分接	208		
7.7.2 MPEG2 TS 定时信息及同步	211		
7.8 宽带应用的传输层协议	213		
7.8.1 应用层分帧和集成层次处理	213		
7.8.2 传统的因特网传输层协议 TCP 与 UDP	214		
7.8.3 实时传输层协议 RTP 与 RTCP	215		
7.8.4 RTP/RTCP 定时信息及同步	218		
7.8.5 包长的选择与包头压缩	219		
7.9 多媒体文件	221		
7.9.1 支持多媒体流式应用的文件	221		
7.9.2 MP4 文件格式	222		
7.9.3 MP4 的同步信息	223		
7.10 编码层与传输层的适配	224		
7.11 下一代多媒体传输协议	225		
7.11.1 对下一代传输协议的需求	225		
7.11.2 MPEG 多媒体传输协议(MMT)	226		
7.11.3 MMT 定时信息及同步	231		

8.2.3	分组交换网视听业务标准 (H.323)	236	9.2.3	非 TCP 流的拥塞控制	280
8.2.4	公用电话网视听业务标准 (H.324)	239	9.2.4	无线网络中的拥塞控制	283
8.2.5	移动电话网视听业务标准(H.324M 和 3G-324M)	240	9.2.5	DASH 自适应速率控制	284
8.2.6	视听系统的通信控制协议	241	9.3	传输层差错控制	288
8.2.7	H 系列系统间的互通	244	9.3.1	自动重传技术(ARQ)	288
8.3	基于 SIP 的会话与会议系统	245	9.3.2	前向纠错码	289
8.3.1	会话发起协议(SIP)	245	9.3.3	前向删除复原码	293
8.3.2	SIP 消息格式和会话描述协议 (SDP)	247	9.3.4	数据交错	298
8.3.3	IP 电话(VoIP)	248	9.3.5	前向差错控制与 ARQ 的结合	298
8.4	协同计算与组通信	249	9.4	编码层差错控制	299
8.4.1	协同计算的概念	249	9.4.1	误码和包丢失对已压缩视频 信号的影响	299
8.4.2	组通信	250	9.4.2	抗误码编码	300
8.4.3	应用共享控制	251	9.4.3	错误掩盖	307
8.4.4	协同工作的发展	253	9.4.4	编解码器交互差错控制	311
8.5	基于 RTP/UDP 的流媒体系统	253	9.5	差错控制小结	313
8.5.1	流媒体的概念	253	习题九		314
8.5.2	系统结构与工作过程	254	参考文献		315
8.5.3	实时流协议(RTSP)	255	第 10 章 视频在异构环境中的传输		316
8.5.4	流服务器	256	10.1	概述	316
8.5.5	IPTV	258	10.2	可伸缩性编码	317
8.6	基于 HTTP/TCP 的流媒体系统	259	10.2.1	可伸缩性编码的概念	317
8.6.1	HTTP/TCP 流传输技术的发展	259	10.2.2	空间域可伸缩性编码	318
8.6.2	顺序式流传输	261	10.2.3	时间域可伸缩性编码	319
8.6.3	动态自适应流传输	262	10.2.4	质量可伸缩性编码	321
8.6.4	MPEG DASH 标准	264	10.2.5	可伸缩性编码的联合应用	321
8.7	覆盖广阔地域的流媒体系统	265	10.2.6	基于小波变换的可伸缩性编码	323
8.7.1	内容分发网络(CDN)	265	10.2.7	H.265 的可伸缩性编码	323
8.7.2	对等(P-to-P)网络	270	10.3	视频转码	326
8.7.3	基于云的内容分发网络	275	10.3.1	视频转码框架	326
习题八		276	10.3.2	运动向量的再利用	329
参考文献		277	10.3.3	DCT 域的运动补偿	331
第 9 章 视频数据的分组传输		278	10.3.4	转码的速率控制	332
9.1	概述	278	10.4	多描述视频编码	333
9.2	应用层拥塞控制	278	10.4.1	多描述编码的概念	333
9.2.1	TCP 拥塞控制	278	10.4.2	多描述的生成	335
9.2.2	TCP 流量模型和 TCP 友好性	279	10.4.3	视频多描述编码	338
			习题十		339
			参考文献		339

第 1 章 概论——多媒体技术的特征

1.1 概述

在技术发展史上,计算机、通信和广播电视一直是三个互相独立的技术领域,各自有着互不相同的技术特征和服务范围。但是,近几十年来,随着数字技术的发展,这三个原本各自独立的领域相互渗透、相互融合,形成了一门崭新的技术——多媒体。多媒体技术最初体现的是配之以声卡、视卡的多媒体计算机。它一出现立即在世界范围内,在家庭教育和娱乐方面得以广泛的应用,并且迅速地向网络化的远程应用发展。多媒体技术的应用与发展,又反过来进一步加速了这三个领域的融合,使多媒体通信成为通信技术今后发展的主要方向之一。

有许多技术,从它们开始出现时就给人以清楚明了的概念。例如电话技术,从最初用两根电线把两部简陋的电话机连接起来实现远距离通话时起,它就被称为电话技术;后来经历了人工交换、步进制交换、程控交换,以至于发展到数字式移动电话,仍然是电话技术。多媒体技术所遇到的情况则有所不同。使人们不容易清楚地建立起“什么是多媒体”的概念的因素很多。首先,通信、计算机与彩色电视本来都是技术面宽而复杂的技术,由它们融合在一起而产生的多媒体技术,其技术覆盖面自然就更宽,技术的交叉更为复杂。这就使得多媒体不能像其他诸如电话、电影、电视、汽车、马车等事物那样一目了然。另外,为了经济上或商业上的利益,某些商家把本来不属于多媒体的技术说成是多媒体技术,人为地造成了概念上的混乱。此外,新闻报道中某些不准确的用词也产生了概念上的误导。由于上述种种原因,造成了这样的局面:“如果你向 10 个不同的人请教多媒体一词的含义,你至少会得到 10 种不同的答案”^[1]。

鉴于上述情况,我们力图在本章中,使读者对多媒体和多媒体技术建立起一个比较完整和全面的概念。

1.2 多媒体的概念与含义

1984 年美国 RCA 公司在普林斯顿的 David Sarnoff 实验室,组织了包括计算机、广播电视和信号处理三个方面的 40 余名专家,综合了前人已经取得的科研成果,经过 4 年的研究,将彩色电视技术与计算机技术融合在一起,于 1987 年在国际第二届 CD-ROM 年会上展出了世界上第一台多媒体计算机。这项技术后来定名为 DVI(Digital Video Interactive)。这便是**多媒体**(Multimedia, 由 Multiple Media 两个词组合而成)技术的雏形。

多媒体技术一出现,在世界范围内立即引起巨大的反响,因为它清楚地展现出信息处理与传输(即通信)技术的革命性的发展方向。国际上在同一年内立即成立了交互声像工业协会,当该组织 1991 年更名为交互多媒体协会 IMA(Interactive Multimedia Association)时,已经有 15 个国家的 200 多个公司加入了。

多媒体计算机与普通计算机有什么不同,它的出现何以如此引人瞩目?早期的计算机只能进行数学运算,后来又具有了文字处理能力,经过若干年的发展之后又增加了图形与动画的功能。而多媒体计算机则是增加了对包括伴音在内的活动图像(即动作连续的电视图像)的处理、存储和显示的能力。这在技术上是一个质变性的飞跃。

为了比较深入地理解这一飞跃,首先让我们来做简单的计算。20世纪后期存在于我国的模拟电视制式在一帧图像内有625行,去掉在扫描逆程(不显示图像的部分)中占去的约50行,出现在一幅画面上有效的扫描行数是576行左右。电视画面的宽与高尺寸之比是4:3,要保证图像在水平方向上单位距离内可分辨的像素数与垂直方向上相等,那么在图像水平方向上的像素数应为 $576 \times 4/3 = 768$ 。将这样一幅单色的电视图像数字化,其取样点数应不少于 768×576 。

根据三基色原理,一幅彩色图像是由红、绿、蓝3幅单色图像组成的,每秒钟要传送25幅彩色图像才能保证电视图像的连续性。如果每一个像素采用8比特量化,1秒钟时间内需要传送的数据量则为

$$768 \times 576 \times 3 \times 8 \times 25 \approx 265 \text{ (Mb)} = 33 \text{ (MB)}$$

一部2小时电视电影的总数据量为

$$33 \times 60 \times 60 \times 2 = 238 \text{ (GB)}$$

在20世纪80年代,计算机硬盘的容量是40MB,总线传输能力在3Mb/s左右,由此可见,普通微机不具有存储未经压缩的活动图像信号的能力,也无法通过总线将其传送到显示器上进行连续显示。

多媒体计算机的出现,标志着人们已经在电视图像信号(或称视频信号)的实时压缩、存储、传送、解压缩和显示等技术上取得了突破性的进展,具有划时代的意义。

多媒体计算机区别于普通计算机的第二个技术特征,是解决了同时存储、读取和显示两个在时间上紧密相关的数字信号(即伴音信号和图像信号)时,如何在时间上保持同步的问题。同步的必要性很容易理解,屏幕上说话的人嘴形与声音配不上就是不同步。在电视系统中,图像信号与伴音信号是组合成一个信号传送的,二者之间总是保持着同步。在多媒体计算机中,数字形式的伴音与视频图像信号是可以作为两个信号分别处理的,要保证同步就需要考虑二者在读取、传送和显示过程中的正确时间关系。在单纯处理数字、文字和图形的早期计算机中并不需要维护这种严格的时间约束关系。当多媒体扩展到远程应用时,服务器与客户端(或两个对等的终端)通过网络连接,由于网络延时和延时抖动的存在,保证声/像信号在传输过程中的同步也是有别于传统数据传输的新问题。

根据以上分析我们看到,多媒体计算机所增加的处理包括伴音在内的电视图像的功能,不仅使计算机在功能数量上多了一项,而是一个质变的过程。声音和电视图像在传统上是属于通信技术和电视技术的研究对象的,多媒体计算机已经将计算机延伸到电视技术和通信技术领域了。

前面的简单计算表明,包括伴音在内的电视信号,是各种媒体中数据量最大的一种。既然技术已经发展到了能用计算机处理这种数据量最大的业务,其他数据量小的,如信息查询、电子邮件、电话、传真、可视电话等业务,都将可以融合到一起,在一台多媒体终端上实现。这是多媒体中“多”字的另一层含意。也有人用集成一词来描述多媒体将多种媒体、多项业务集合在一起的这一特点。

有的学者从数据本身的特点来理解多媒体技术与其他技术的区别。多媒体数据是由内容上相互关联的文本、图形、图像、声音、动画、活动图像等媒体的数据所形成的**复合数据**。这一数据合成的过程是在计算机控制下完成的^[2]。在这里首先说明了多媒体所涉及的信号是数字化的,而不是模拟的;其次应当注意“内容上相互关联”和“合成”这两个要点,内容上毫不相干的文字、图形、声音、活动图像等数据的集合并不是多媒体数据。通常对时间敏感的声音、活动图像的数据称为**实时数据**,其他类型的数据则是**非实时数据**。如果说某个数据是多媒体数据,则意味着该复合数据中至少包含有一种实时数据和一种非实时数据^[3]。

能够通过计算机对视频数据进行操作,形成了多媒体技术的另一个基本特征,即**交互性**(Interactive)。人一机交互是计算机固有的技术特征。通过键盘或鼠标打开一个文件,就会显示你所要的内容,经过修改、补充之后,再写入到给定的文件名下,或者通过一条打印指令则可在打印机上将文章打印出来。这种人与计算机之间通过指令“对话”式的操作就是交互操作,这已经是人们几十

年来所熟知的常识。而多媒体强调的是以交互的方式对活动图像(包括伴音)进行操作,例如,对其进行“快进”、“暂停”等录像机式的操作。

当人一机交互中的“人”和“机”处于不同的地域时,二者之间的交互需要通过双向的信道来完成,例如,北京的通信终端自动地从位于上海的数据库内调取电视节目时的交互工作(如图 1-1 所示)。这项业务使用了双向的非对称信道。所谓非对称是指两条线路的信道带宽或数据率等参数不相同。上行线(由用户向信息中心)传送指令,是低速率的;下行线传送读取的电视信号,是高速率的。需要强调的是,这项业务中的上、下行信道中传输的信息是相互关联的,体现了人一机交互操作。

除了人一机交互之外,一些多媒体系统支持人与人之间的交互,例如,可视电话、多媒体会议和协同工作就是典型的例子。传统的电话也是实现人与人交互的通信系统;支持人与人之间交互的多媒体系统与电话的最大区别在于,它使用了更多的媒体(除相互听得见外还相互看得见)和允许同时进行更多的业务(如除通话外还可以共同观看、修改文件等)。

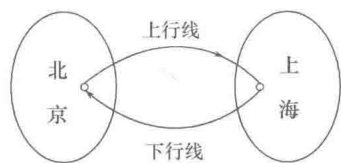


图 1-1 人一机交互中的双向通信

通过上面的讨论,我们可以对多媒体的特征作一个简要的总结,即它具有**集成性**、**同步性**和**交互性**。其中集成性包括多种媒体的集成和多种业务的集成;交互性包括人与人的交互和人一机交互。

1.3 多媒体产生的技术背景

一种新技术的产生与发展往往是与其特定的技术背景相联系的,是以其他有关技术的发展作为基础的。实际上,多媒体技术之所以能够在 20 世纪 80 年代末期出现,主要得益于下述几个方面的技术成果。

1.3.1 图像压缩编码技术的成熟

在通信领域中,人人都知道数字通信具有模拟信号通信所无法比拟的优越性。模拟信号在传输过程中产生了失真或者混进去噪声,在接收端难以使其恢复原形。数字信号则不同,因为发送的脉冲信号形状是已知的,如果在传输中产生失真或叠加上噪声,在接收端经过放大、幅度切割等整形处理,失真和噪声可以在很大程度上被消除。

数字通信的缺点是将模拟信号变为数字信号以后,对信道带宽的要求大幅度增加。以电话为例,一个模拟话路只需要 3.4 kHz 的带宽。变成数字信号时,取样频率取 8 kHz,每个取样点采用 8 比特量化,一路数字电话的数据率则为 64 kb/s。当用二进制码传输时,每赫兹带宽最高只能传送 2 b/s(采用多进制码传输时,这个数字可以高一些)。可见一路电话从模拟传送改为数字传送,对信道带宽的要求提高了很多。彩色电视所遇到的情况则更为困难。我们在 1.2 节里已粗略地估算了彩色电视信号的数据率。按照国际标准,一路按分量进行编码的彩色电视信号(不包括伴音)的数据率 $R=216 \text{ Mb/s}$,而一路模拟彩色电视信号的带宽只有 6 MHz。正是由于这个原因,虽然早在 1937 年 A. Reuees 就发明了 PCM(脉冲编码调制),但数字通信得到广泛的应用还是在 20 世纪 70 年代之后。

要以数字方式传输电视信号,必须解决数据率的压缩问题,这也称为信源编码问题。人们对信源压缩编码已进行了几十年的深入研究,进入 20 世纪 80 年代,这项技术已经较为成熟,能够将数字电视信号数据率实时地压缩到 34 Mb/s 左右。这里所讲的技术的成熟是指压缩方法,而实时是指压缩与解压缩的速度跟得上 25 帧/秒、或 30 帧/秒的图像显示要求。处理速度的高低取决于用

以实现压缩和解压缩的电子电路的集成化水平。

人们研究数字电视信号的压缩编码问题的最初出发点,是要解决电视信号长距离传输中的抗干扰问题,也就是说,将电视台发出的信号数字化,然后压缩编码以求用较低的数据率传送,传到目的地以后(如从北京传到上海),数字信号经过切割整形,再还原为模拟信号,送到本地发射机发射出去。因此在多媒体出现之前的几十年中,电视信号的压缩编码一直是针对通信领域中的应用的。

将图像压缩编码的研究成果应用到计算机领域则导致了新技术的产生。当 DVI 技术与世人见面时,它已经能够将图像信号和伴音信号实时地压缩 100 倍以上(包括适当地将电视信号的图像分辨率降低),其速率为 1.2~1.4 Mb/s,这使得活动图像数据能够在当时的计算机总线上传输,从而成为计算机可以处理的数据类型之一。同时也使得一张 CD-ROM 能记录 74 分钟的电视节目,如果数字电视信号没有经过压缩,则只能记录 30 秒钟的电视节目。

1.3.2 大规模集成电路技术的发展

代表大规模集成电路技术水平的主要参数之一是制作在芯片上的线的宽度。线宽做得越窄,一块芯片上能容纳的元件便越多,集成度也便越高。至 20 世纪 80 年代末,已经能在芯片上制作线宽小于 $0.5\ \mu\text{m}$ 的线了。在多媒体技术发展初期,CPU(如 80286)的处理能力还比较低,那时数据的压缩和解压缩运算要靠专用的芯片来完成。在 Intel 公司的 DVI 技术中,图像的压缩、解压缩是用 2 个芯片来完成的,其中每个包含有 26 万多个晶体管^[4]。这个数字清楚地表明,要实时地将彩色电视信号的数据率压缩到几个 Mb/s 以下,电路的集成度不高是无法实现的。

让我们看一下这几个数字,286 CPU 只集成了 13.4 万只晶体管,386 CPU 则有 27.5 万只,发展到 486、586(P5)和 686(P6)时,CPU 内集成的晶体管数分别为 120 万、300 万和 350 万只。这些数字充分说明了大规模集成电路技术发展之迅速,从而也为多媒体技术的发展提供了良好的条件。使早期的多媒体终端的成本降低到普通家庭的购买力能接受的水平,在很大程度上也有赖于大规模集成电路技术的发展。

1.3.3 大容量数字存储技术的发展

激光视盘(LVD,后称 LD)是 20 世纪 70 年代研究成功的,能够在 1 张直径 12 英寸的大盘上记录大约 30 分钟的电视节目。LVD 的出现最初并没有引起太大的重视,人们的注意力还集中在探讨究竟光盘机与磁带录像机哪一种技术更有发展前途。光盘记录技术于 1982 年被用来记录音乐、流行歌曲,1 张 5 英寸直径的小盘 CD(Compact Disc)能够记录超过 70 分钟的数字化的、高质量的音乐节目。CD 的出现与迅速发展提醒了人们用它来记录计算机程序与数据。

用来记录计算机数据的光盘与记录音乐的光盘有不同的技术要求。首先,音乐的播放通常是顺序进行的,即从头至尾地播放;当需要从这个曲子跳到另一个时,跳跃的间隔也比较大。计算机数据的读取则不同,它要求可以从光盘的任何一点取出数据,这通常称为随机访问(Random Access)。其次,个别数据发生错误将降低音乐播放的质量,但是 CD 的误码率在不高于 10^{-8} 时,人们并不容易察觉到播放质量的降低。而对于计算机数据而言,这个错误率是不能容忍的。

随着光盘技术的发展,随机访问问题的解决,并且能够将误码率降低至 10^{-12} ,于 1984 年出现了记录计算机数据的 CD-ROM(Read Only Memory)。最初 1 张 CD-ROM 的存储容量为 660~1080 MB,读取速率是 150 kB/s,寻道时间(即找到文件的起始位置的时间)为几百毫秒。至此,光盘的容量已经满足存储一个电影节目(1 小时左右)的要求,而读出速率已经足够支持实时地提取已压缩的活动图像数据流的需要(1.2 Mb/s),这就为多媒体技术的诞生提供了另一个必要条件。

差不多与 CD-ROM 迅速得到广泛应用的同时,只写一次光盘(Write Once)研究成功。这种 12 英寸的光盘最初能记录 1 GB 的数据,后来其容量也迅速扩展到 5 GB 和 10 GB。

近年来可读/写光盘和计算机硬盘的容量迅速增大,并出现了磁盘阵列、光盘阵列等大型数据存储设备,为多媒体技术的实际应用和全面发展提供了充分的条件。

1.4 多媒体系统的基本类型

多媒体计算机是多媒体技术的最直接、最简单的表现形式。因其本身具有存储、运算、处理和显示的能力,具有独立的功能,如动画显示、视频播放等,因此,多媒体计算机一出现便立即在家庭、教育和娱乐方面得到广泛的应用。但是,多媒体技术真正的意义在于与网络的结合,在于通过网络(局域网和广域网)为用户以多媒体的方式提供信息服务。

基本的多媒体系统除了以多媒体计算机为基础的独立(Stand-Alone)商亭式系统之外,通过网络提供业务的系统可以分为两大类:一类是人与人之间交互的系统,如多媒体会议与协同工作、多媒体即时通信等;另一类是人机交互的系统,如多媒体信息检索与查询、视频点播等,本节中将分别对这些系统及其技术特点进行介绍,而其中所涉及的关键技术将在以后的章节中加以讨论。

1.4.1 独立商亭式系统

凡是以一台多媒体计算机为核心的应用系统,例如商场的导购系统、展览馆的导游系统等,我们都称为独立商亭式系统。在这类系统中,除了各种媒体的采集、表示、压缩存储和解压缩播放之外,如何组织素材,并运用多媒体手段将信息有效地、具有感染力(或艺术性)和方便地提供给用户是制作应用软件时应考虑的重要问题。这里涉及的不仅有技术、艺术,甚至还有社会、心理学等方面的问题。多媒体制作软件(如 Authoring Tool、Authorware 等),或者原有操作系统的多媒体扩展(如 Video for Windows),是为制作应用软件而提供开发环境的软件。它不仅向应用程序的开发者提供多媒体输入/输出设备的接口,更重要的是,还提供建立媒体数据之间的空间布局和播放时间顺序等关系的手段。因此,开发优秀的制作软件本身远比开发应用软件困难。

在这类系统中,操作系统的实时性是值得重视的另一个问题。在嵌入式系统或工业控制机中常常涉及实时操作系统,在那里强调的是对事件中断的实时响应。而在多媒体系统中,由于视频和音频数据需要在一定时间约束条件下(如每秒 25 幅图像)连续不断地送到输出设备上供用户聆听和观看,因此这里操作系统的实时性强调的是,处理这些有时间要求的连续媒体流的能力。

提供更友好的人—机接口是商亭式系统技术发展的一个方向。除了使用键盘和鼠标,触摸式输入也很普遍,人们还试图通过声音、手势,甚至表情等多种模态的接口对系统进行控制,从而构成更人性化的多媒体交互环境。

如 1.2 节所述,实时数据引入进计算机导致了多媒体的诞生,因此本书更关注音、视频的表达、压缩/解压缩和传输的问题。至于其他媒体(文字、图形、图像等)的表达和处理(包括压缩),以及多媒体的制作、人—机接口等,读者可以参考从计算机角度来阐述多媒体的书籍。

除了独立商亭式系统外,下面将要介绍的 4 类系统都是在多媒体终端与终端之间、终端与应用服务器之间有网络相连接的多媒体通信系统。

1.4.2 多媒体信息检索与查询

通过因特网进行信息查询已是十分普及的应用。多媒体信息检索与查询 MIS(Multimedia Information Service)系统除可以根据关键字等对文本资料进行查询之外,也同时具有对活动图像和声音的查询能力。从通信方式而言,MIS 是点对点(信息中心对一个用户),或一点对多点(信息中心对多个用户)的双向非对称系统。从用户到信息源只传送查询命令,要求的传输带宽较小,而从信息源传送到用户的信息则是大量的、宽带的(见图 1-1)。

MIS 所涉及的两个重要技术问题是:① 如何向用户提供丰富的信息和如何让用户快速、有效地查询与浏览这些信息;② 如何合理、有效地组织多媒体数据的存储和检索。

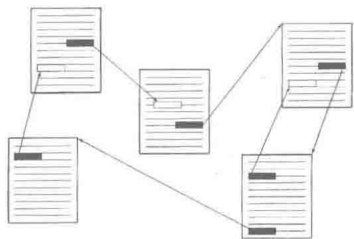


图 1-2 超文本文件结构

为了对第一个问题有所认识,首先让我们回顾一下人人都熟悉的读书过程。对于阅读一本小说来说,人们通常是从头至尾逐页阅读的,或者说是按顺序阅读的。但在有些情况下,特别是在技术或社会科学领域,在阅读某本书的过程中,经常需要从另一本书或论文查找某个论点,或者说,在几本书之间需要交叉参考的情况常常发生。图 1-2 表示出用电子的方法来实现交叉参考的情况,这实际上已经是大家在因特网的查询中十分熟悉的过程:用鼠标点击黑框所标的地方,就会显示出箭头所示的有关参考信息,看完该信息后可以回到原

来的页面,或者再进入其他页面,……箭头指向的页面(信息单元)可能与原来的页面在同一个文件中,也可能在其他文件里。这种信息的非顺序(或称为非线性)的组织结构称为**超文本**,超文本中信息单元之间的链接称为**超链**。当上述信息组织方式不仅用于文本,还包括其他媒体数据,特别是音频和视频数据时则称为**超媒体**。超文本和超媒体这两个词在很多文献中也常常被混用。

超媒体为用户提供了一种在文件内部和文件之间迅速查找和浏览多媒体信息的方法,但是人们希望在更大的范围内迅速、有效地获取信息,这就不能不提到推动因特网突飞猛进发展的 **WWW 技术**。**WWW**(World-Wide Web)最初是 1989 年在日内瓦 CERN 启动的一个研究项目的名称,由于它的巨大成功,现在 WWW 已经意味着在超媒体原理下发展起来的一系列概念和通信协议。**Web** 这个词也代表了世界范围内由因特网相互连接起来的众多的信息服务器所构成的巨大的数字化的信息空间,也有的学者将之称为**超空间**。

WWW 的基本思想和它所解决的问题主要体现在如下几个方面:

(1) 在超空间中没有一个统一的管理者。任何人都可以创建超文本文件、将其与其他文件链接,并放入超空间中去。标准的超文本文件采用 **HTML**(Hyper Text Markup Language)格式。超文本的数据交换和传递使用 **HTTP**(Hypertext Transfer Protocol)协议。HTTP 定义了客户端和服务端之间请求和应答的语言和规则。

(2) 定义了一种在超空间中寻找所需要的文件的机制,称为**统一资源定位器 URL**(Universal Resource Locator)。通过 URL 可以知道每个文件处于哪一台机器,叫什么名字,以及以何种机制可以将该文件传输到需要链接它的地方去。

(3) 具有一个统一的、简单的用户界面,无论查询到的信息来自本机,还是来自远方的服务器,用户从界面上看起来都是一样的。实现 WWW 用户端功能的软件称为**浏览器**。通过浏览器不仅能够调取 HTML 格式的文件,还可以调取以任何形式存储在已有的数据库,或信息库中的信息(虽然此时不具备超链接功能)。

以上 3 个问题的解决,使得世界上使用不同硬件和软件的分离的信息系统,通过因特网构成了一个庞大的统一的信息系统,从而为用户打开了通往一个大得难以想象的信息库的大门。这正是 WWW 取得巨大成功的原因。为了使用户不至于面对浩瀚的信息而不知所措,人们又进一步设计了帮助用户过滤掉无用信息、尽快找到所需要的信息的专门软件,这就是所谓的**搜索引擎**。

随着声音和活动图像等实时信息的逐步增加,因特网正在演变成世界范围内最大的 MIS 系统。由于以上所介绍的如何向用户有效地提供和查找信息的技术首先是从文字信息查询与检索发展起来的,关于这些技术的书籍已经很多,所以本书将不准备讨论这方面的内容,但是从后面的章节我们将会看到,Web 技术对多媒体技术的发展有着重要的影响。

MIS 系统涉及的第二个重要技术问题是多媒体数据的存储和检索。与存储传统的数据不同,多媒体数据需要有适当的数据结构,以表达不同媒体数据之间在空间上与时间上的相互关系;对不

同媒体要有合理的存储方式;对于数据量大而在时间上又有严格要求的音频和视频数据流,要有实时的提取算法;等等。目前多媒体数据一般是以特定格式的文件存储的。

此外,传统的、利用关键字或属性描述等来进行信息查询的方式,比较适用于文字信息,用来对声音、图像等多媒体信息的查询则有不方便之处。**基于内容的检索**是伴随着视频和音频查询而发展起来的新技术。利用这种技术,给出(或从查找对象中自动提取出)所要求的特征,例如图像中物体的形状、颜色等,就能找出具有同样、或类似特征的物体的图像来。更高级的查询方式则是给出“概念”或“事件”,如国旗、山脉、骑自行车的人等,找出具有同样概念或事件的图像或视频来。这种方式也称为**基于语义的检索**。基于内容和基于语义的检索涉及图像和视频的分析与理解、语义提取、模式识别、机器学习与人工智能等,是当前多媒体领域中的一个重要研究方向。由于本书侧重于多媒体通信,因此将不准备讨论这方面的内容。

1.4.3 多媒体会议与协同工作

可视电话和会议电视是早在多媒体出现之前就已经存在的人与人之间进行通信的手段。**计算机支持的协同工作 CSCW**(Computer Supported Co-operative Work)也是早在 20 世纪 80 年代初在计算机领域内提出的概念。它是指用来支持多个用户共同参与一件工作(如共同编辑文件、修改设计图等)的计算机系统及其相关的技术,但合作者之间不能“见面”与交谈。多媒体的出现为这两种交流形式提供了结合的基础,合作者既能看得见、听得到,又能一起处理事务,使他们真正像聚集在同一个房间里面对面地交流与工作。这种通信系统和业务称为**多媒体协同工作 MMC**(Multimedia Collaboration)。多媒体远程医疗诊断系统、多媒体远程教育系统等都是融入了一定 MMC 功能的应用。

1. 会议室会议电视系统

传统的电视会议需要在专门的会议室内进行。由电视摄像机对着主会场、主席等拍摄,通过电缆、光缆、微波或卫星信道送到分会场收看。如果要求主会场也能看到、听到分会场发言的情况,传输信道则是双向的,以将分会场的信号送到主会场。主会场(或者通信网的某个节点上)有信号切换设备,用来选取某一分会场的信号,并将该信号送至其他分会场;或者将几个分会场的信号综合起来,以分画面的形式送给各个会场。在有的系统中,主会场还可以对分会场摄像机的摄取方向等进行控制。

这类系统的一个重要特点是,需要像电视台的演播室一样,对被拍摄的景物(人、黑板、会场的全景等)给以专门的照明(普通室内照明设施不能满足要求)。由于会议电视系统拍摄的景物没有什么剧烈的运动,主要是讲话人面部和形体的运动,而广播电视要传送包括诸如运动员的快速动作在内的高速运动的图像,所以会议电视的摄像机、信道设备等相对于广播电视所用设备而言比较简单,而且在同样的图像分辨率下,会议电视的数据率可以被压缩更大的倍数。另外,由于同样的原因,为了保证动作的连续,电视图像每秒钟需要传送 25 帧,而会议电视每秒传送 10~15 帧即可以被接受。数据率为 384 kb/s 的系统所给出的图像质量已经可以令人满意了。

在会议室会议电视系统中,通过电子白板等辅助人机交互设备的使用,可以多媒体方式呈现、修改、记录和存储计算机中的文件,实现与会各方对信息的共享。

2. 桌面或手持终端会议电视系统

用计算机或手持智能设备取代会议室会议电视系统中的编解码设备和显示设备,是这类会议电视系统的基本特征。

在会议室会议电视系统中,摄像机不仅要拍摄讲话者还要能够对整个会场进行拍照,这要求摄像机有较大的视野和较高的灵敏度,因此其照明条件必须达到演播室的标准。而在桌面或手持系

统中,摄像机只需要对准讲话的人,这不仅降低了照明要求,也降低了对摄像机的视野和灵敏度要求,摄像机大为简化。

在这类系统中,音、视频的处理与文字等其他媒体的处理被集成在一个系统中,这使得实现与会各方的信息共享和协同工作更为方便,有利于系统向多媒体协同工作的方向发展。

3. 多媒体协同工作

MMC的最终目标是希望使身处异地的人们,能够像处于同一房间内面对面一样地交谈、协商工作,下面列举的是人们向着这一目标所正在作的努力。

教师从显示器屏幕的3个窗口分别看到在3个地方听课的学生,与在一个教室中面对全体学生的感觉是不一样的。利用计算机的图形功能可以生成类似真实图像的虚拟图像,例如具有天花板、窗户、灯具的教室,并将从3个地方传送来的学生的现场图像与计算机生成的虚拟教室图像结合在一起,构成一个全体学生在内的完整的教室全貌,将会给人以更真实的感觉。

在现实生活中举行会议时,某个与会者有时需要和邻座说一些不愿意让别人听到的悄悄话,或拿出一份文件与其小声商量;有时与会者要边讨论、边对一个文件或一份设计图纸同时进行修改,甚至需要共同操纵一台仪器进行实验;如此等等。在多媒体会议中,要实现类似现实生活中的这些行为要涉及许多技术问题。

显示器的屏幕是平面的,无论屏幕上显示的景象是多么的有立体感,人们仍然是身在其外,而不是身在其中。如何将虚拟现实与协同工作结合起来,使人们在虚拟的三维环境之中协同工作是一个值得研究的课题。

人们会面时的第一个动作往往是一边握手,一边说“你好”。如果MMC终端可以用语言(不是键盘)输入、并配有机械手,可能使你感受到远方合作者向你握手问好的真实感觉。除了听觉和视觉之外,将其他的感觉,如触觉、嗅觉等结合进协同工作环境;或者将多媒体协同工作与机器人技术结合起来,使合作者能够共同进行除了屏幕上的工作(如编辑文件之类)以外的事情,这些都是研究者在探索的问题。

总之,多媒体协同工作将从各种不同的方面,向着能够使得被空间距离分开的人,在必要的时候可以像已经聚在一起,有面对面地一起工作的条件与自我感觉的方向发展。但要真正达到这一目标,要走的路途还相当遥远。这里包括的不仅是技术问题,还有许多为社会学和心理学家们所感兴趣和值得研究的问题。

从通信的角度来看,MMC系统是对通信系统要求最高的应用。它要求一点对多点,或者多点对多点的双向信息传输。另外,在MMC系统中,声像信号是实时产生的,需要实时地压缩、传送,整个系统的时延要足够小,才能满足人们对话时自然应答的时间要求。在复杂的协同工作系统中,要实现“开小会”和进行共同操作等,还要能够随时建立、撤销某些专有信道。当涉及视、听之外的其他形式的传感器时,通信机制的复杂程度则会更高。

1.4.4 多媒体即时通信

即时通信系统更完整的表述是出席与即时消息系统(Present and Instant Messaging System, IMS)。它允许用户相互之间了解各自的状态和状态的改变,如在线、离线、繁忙、隐身等,并允许用户相互之间传递即时的短消息。第一个即时通信系统于1996年在以色列诞生。人们通过IM系统发送文本型的短消息,由于消息传送的即时性,对方可以立即给予回应,一来一往如同“聊天”;可以多个人一起聊,仿佛在一个聊天室,也可以两个人进行“私聊”,等等。由于这种交流方式的方便和快捷,IM在世界范围内得到了迅速的发展,成为最流行的网络应用之一。现在,IM从最初的个人聊天应用,逐步扩展到成为企业内部进行工作交流的有力工具,企业可以随时查看各部门在线人员情况,沟通各分支机构等。同时,IM从原来支持简单的文本短消息交流,发展到加入图片和文

件传输、视/音频消息的即时传送,以及双方和多方的视/音频通话,等等。因此我们在本节标题中将它称为多媒体即时通信,它是一项极具潜力的业务,微信的广泛使用就是一个例子。

加入了视、音频的 IM 系统从功能上讲与可视电话或会议系统类似,但实现方法并不相同。可视电话系统由通话双方通过呼叫协议直接建立双向的连接;而经典的 IM 系统采用客户端/服务器(C/S)结构,“聊天”双方的信息需要通过服务器进行中间转接。当进行视/音频通话时,由于数据量大,服务器中转可能引起响应的不及时,此时可以在“聊天”双方建立直接连接,但这个连接的建立通常也需要在服务器的帮助下完成。由于服务器是 IM 系统的核心,用户必须先登录服务器才能接受各种服务,因此服务器了解各用户的状态及状态的变化,从而能够向一个用户提供其他用户的状态信息,让他了解其他人的在线情况。这就是“出席”(present)服务。而在可视电话系统中,被叫方并不需要事先“出席”,主叫方可以通过一定的通信协议呼叫对方(如振铃),对方应答就直接建立连接了。

如上所述,一个典型 IM 系统包含两种基本服务:出席服务和即时消息服务。图 1-3(a)为出席服务的基本框图。出席服务有两类客户,一类称为出席者(presentity),另一类称为观察者(watcher)。出席者向服务器上的出席服务提供自己的出席信息。观察者可以定期或不定期地向出席服务请求得到某些出席者的当前出席信息;也可以订阅出席信息,此时出席服务会在出席者的出席信息发生变动时主动告知订阅者。

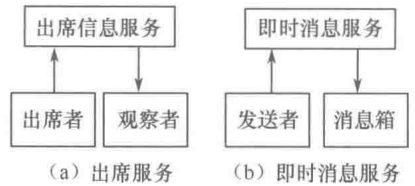


图 1-3 即时通信

图 1-3(b)为即时消息服务的基本框图,其中发送消息的一方称为发送者,接收消息的一方称为即时消息箱。发送者向服务器上的即时消息服务发送消息,消息中包含目的即时消息箱的地址;即时消息服务则根据目的地址向该即时消息箱转发消息。当用户之间需要进行视、音频通话时,发送者从服务器获得接收方的地址和状态信息,并通过一定的协议在服务器的帮助下建立起二者之间的直接连接。然后在此连接上视、音频数据可以采用与可视电话和会议系统中类似的方式进行传输。

由于即时通信主要传递的是文字、图片和短文件等非实时数据;在进行视频通话时又采用了与可视电话类似的技术(除了建立连接的方法不同外),因此本书将不专门讨论即时通信系统。

1.4.5 视频点播与直播

多媒体计算机出现以后,由于它具有以交互操作的方式调取包括伴音在内的活动图像的功能,立即导致了这样的构思:在节目中心,将节目以压缩后的数据形式存入视频节目库;用户在家里可以按照菜单调取任何一套节目,或者一套节目中的任何一段,并能实现录像机式的功能,即快进、快退、重放、慢动作以及播放静止画面等,这种系统与业务便是视频点播(Video on Demand, VOD)。

从使用功能上讲,VOD 与多媒体信息检索与查询系统是类似的,但是二者的业务特点却有很大不同,因而技术的侧重点也有所不同。在多媒体信息查询系统中,信息的主要部分通常是通过文字、图片表达的,其中需要显示的视频片断一般时间不长,通常是完全下载后才播放。在 VOD 系统中,视频节目的长度是以小时计算的,完全下载播放将使用户的等待时间过长,因此节目是一边传输、用户一边播放的。电视信号的数据率很高,其中标准清晰度的电视信号达 2~4 Mb/s。这意味着在几个小时中,每一秒钟都需要传送几兆比特的数据,才能使用户正常、不中断地收看节目。如何保证在同一时间内向成千上万、甚至更多的用户提供内容不同,而又连续不间断的高速数据流是这类系统需要解决的关键技术问题。

图 1-4 是一个 VOD 系统示意图。图中负责按用户提出的要求向用户传送节目数据流的设备