

计算机语音集成原理 技术和应用

张延平 林博文 编著



人民邮电出版社
PEOPLE'S POSTS &
TELECOMMUNICATIONS
PUBLISHING HOUSE

TP3
117

计算机语音集成原理、技术和应用

张延平 林博文 编著

人民邮电出版社

内 容 提 要

本书是一本专门介绍计算机语音集成技术和应用的图书,书中内容涉及到了计算机技术和通信技术,是目前信息技术发展的一个热点。

本书共 7 章,内容包括:呼叫处理、通信基础知识、计算机语音系统集成技术、计算机电话语音卡、CTI 技术标准简介、系统编程、应用系统设计等。

本书具有较强的应用性,可供从事计算机语音集成开发应用方面的技术人员和大专院校相关专业的师生阅读参考。

计算机语音集成原理、技术和应用

◆ 编 著 张延平 林博文

责任编辑 王晓明

◆ 人民邮电出版社出版发行 北京崇文区夕照寺街 14 号

北京顺义振华印刷厂印刷

新华书店总店北京发行所经销

◆ 开本:880×1230 1/32

印张:8.375

字数:242 千字

印数:1-5 050 册

1998 年 2 月第 1 版

1998 年 2 月北京第 1 次印刷

ISBN7-115-06969-7/TP·620

定价:29.00 元

序 言

当今我们正在进入到一个伟大的信息时代,通信、计算机与信息工业融为一体,它推动着人类社会以历史上不曾有过的高速度向前发展。电话网和计算机网已成为遍布全球的两个最大的通信网络。众所周知,在发展的初期它们是平行的,信息时代的到来使人们意识到我们不仅需要语音通信,更需要数据通信和图像通信。十几年前,一些有远见的学者提出了将计算机(Computer)和电话机(Telephone)通过某些硬件和软件集成为一体(即 CTI: Computer Telephony Integration)的技术,它使语音和数据服务融为一体,并在一个终端(计算机或电话)上得以实现。当然 CTI 技术的意义还远远不只这一点。近年来人们正在努力研究和发展的多媒体及其通信技术,以便实现语音、数据及图像的综合通信。如果我们再把光纤通信系统、卫星通信系统、无线通信系统、微波系统等各种信息传输系统构成一体,则人类在通信方面就真正地实现所追求的目标:任何人无论任何时间、在任何地方,都可与任何人实现任何类型(语音、数据、图像)的通信,即未来的个人通信。可见 CTI 技术将在该领域中起到十分重要的作用。

CTI 技术是一门新兴的科学技术,虽然它的发展历史不长,但是发展速度很快、应用很广。

十几年来,我国的邮电通信事业取得了惊人的成就,到 1997 年 7 月全国局用程控交换机总线数已超过 1 亿,而用户安装话机数已超过 7000 万台,使我国电话网已跃居世界第二。

我国的计算机网同样发展迅速,形势喜人。现在各类 PC 机正以每年增加近二百万台的速度发展,各种数据通信网纷纷建立并投入使用,其中中国公众数据通信网,主要由中国邮电部运营;它包括中国公用分组交换网(Chinapac)、中国计算机互联网(Chinanet)、中国公用数据网(ChinaDDN)等等。目前这些网的用户已近百万。此外各部委、各省

市也根据部门、行业和地方需要建起了不少计算机网络和信息服务系统,如中国教育和科研计算机网(Cernet)、中国科学院网(CASnet)、中国“金”字系列网(如金桥、金关、金穗)等。据不完全统计,用户数已超过百万。可想而知,电信和计算机事业的发展为 CTI 技术的发展提供了十分广阔的领域。近几年来世界上 CTI 技术和产业正在以每年 30% ~ 50% 的速度增长,我国的 CTI 产业虽然刚刚开始,但每年已出现成倍增长的好势头。当前国内许多大学、研究机构、工厂和公司都在组织人力,积极开展有关 CTI 的研究和开发工作,为我国 CTI 事业的发展创造了良好的环境。为了适应这一发展的需要,作者在较短时间内编写了这本介绍 CTI 技术原理和应用的书供读者参考。

本书主要介绍计算机语音卡,它是构成计算机语音系统的核心部件,语音卡的性能直接影响 CTI 系统集成功能。书中以业界知名的 Dialogic 语音卡为例,介绍了语音卡的工作原理和使用方法,并列举了大量语音卡(包括模拟/数字语音卡、传真卡、人工座席卡、电话会议卡、调制解调器卡、七号信令卡和开放的 DSP 平台等)的实例,以开阔视野。

与通信和计算行业一样,CTI 领域也需要标准,以利于实现各厂商产品的兼容性和互操作性。数据信号计算系统结构(SCSA)赢得了广泛的信任,脱颖而出,并得到了美国国家标准局(ANSI)和企业计算机语音论坛(ECTF)等标准化组织的认可和采纳,因此本书介绍了与 SCSA 有关的 SC 总线和 SCapi 等内容。

CTI 技术涉及多种周边的技术,如数字信号处理、计算机软件编程和通信等技术。为此,本书介绍了呼叫进程分析、状态机编程、脉冲识别和传真通信等知识,尤其对语音识别、文语转换等需求迫切并逐渐走向应用的新技术予以重视。

CTI 应用系统的集成需要很多技巧,包括语音卡的选配、用户界面的设计、系统规模的确定以及软件平台的选择等。本书对如何设计一个好的 CTI 系统,避免一些常见的错误,给出了一些方法和建议。

CTI 系统应用灵活、广泛、生动,在信息化社会的今天,每时每刻都在潜移默化地影响着我们的生活。本书给出一些 CTI 应用系统设计的实例,包括语音信箱、传真、电话银行、Internet 电话和 Internet 传真等。

这些实例,对提高应用系统集成能力、开拓应用领域、扩展视野大有裨益。

此外,为方便非通信工程专业的读者阅读本书,书中还扼要地介绍了一些通信基础知识,包括通信网、信令等。

我们看到,人类的需求推动着 CTI 技术正向着更深、更广的方向发展。而要实现技术上的飞跃,关键是要有人才,要培养成千上万的 CTI 技术专家。为此,北京邮电大学已决定在研究生教学计划中开设该内容的课程和实验,并对相关工程技术人员进行培训。几年来我们与 Dialogic 公司的研究人员紧密合作,开展了多项 CTI 领域的研究、开发和推广应用工作,并希望能为我国 CTI 事业的发展作出贡献。

本书写作风格轻松,叙述简明、扼要、易懂,较全面地讨论了与 CTI 技术相关的一些基本概念、原理、软件设计、硬件结构、CTI 标准等,也讨论了它的典型应用以及开发方法,它集原理、技术、应用于一体,因此它不仅可以作为高等学校本科生、研究生学习 CTI 科学技术时的参考书,也适合一切工作在 CTI 领域的科研、教学和工程技术人员阅读和使用。同时它也是从事 CTI 技术开发工作的工程师必备的参考书。

宋俊德
于北京邮电大学
1997.12

目 录

第一章 呼叫处理	1
1-1 什么是呼叫处理	1
1-2 呼叫处理的功能	2
1-3 呼叫处理的主要应用	3
第二章 通信基础知识	6
2-1 基本概念	6
2-2 通信信令	18
2-3 呼叫进程	24
2-4 E-1 数字中继	29
第三章 计算机语音系统集成技术	43
3-1 系统配置	43
3-2 用户接口	51
3-3 特定短语的发音	59
3-4 音频信号的数字化	69
3-5 声音文件格式	75
3-6 传真集成	79
3-7 语音/数据集成协议	89
3-8 脉冲识别	93
3-9 语音识别和文语转换	96
第四章 计算机电话语音卡	115
4-1 工作原理	115
4-2 语音卡的安装	124
4-3 Dialogic 语音卡	131

第五章 CTI 技术标准简介	154
5-1 SCSA	154
5-2 TAPI	161
5-3 TSAPI	167
第六章 系统编程	172
6-1 操作系统	172
6-2 驱动程序	176
6-3 状态机编程	191
6-4 SC 总线编程	203
6-5 TAPI 编程	208
第七章 应用系统设计	215
7-1 系统规模	215
7-2 高级 CTI 模块	223
7-3 CTI 应用系统设计实例	239
参考文献	254
缩写词对照表	255

第一章 呼叫处理

1-1 什么是呼叫处理

呼叫处理,是融合通信技术和计算机数据处理技术的新技术,它能使电话用户通过电话机终端访问计算机存储的信息,就像键盘和屏幕终端一样方便。

呼叫处理在人们的日常生活中随处可见,例如,把要说的话留在语音信箱里;查询银行的帐务等等。

通信领域把电话线上传递的信息分为二类:话音和数据。话音是由人的发音器官发出,由人的听觉器官—耳朵接收的信息;数据是编码信息,在计算机之间传递。从信息传递的角度而言,话音和数据对信息的“损失”量要求不同。人的耳朵和大脑具有很强的模式识别能力,即使传输中存在“喀喀”声、短的停顿、高频或低频畸变等失真,也能辨别出对方说话的内容。而计算机处理的数字数据却不然,即使一个简单的疏漏或二进制数字的传输错误,都可能对传输信息产生很大的影响。数据传输必需是高可靠性的,换言之,语音的传输一定程度上能容忍语音质量的变化。

呼叫处理为语音信息的传递提供了智能特性。带有呼叫处理(或语音处理)功能的系统单元,称之为语音应答单元(Voice Response Unit,或 VRU)。

1-2 呼叫处理的功能

VRU 所能提供的一些主要功能如下:

1. 语音的存储转发(Store & Forward)

所谓存储转发,也就是语音的录制和回放。具备语音存储转发功能的语音处理系统,可以录制音频信号,并把它重放出来。家庭用的应答电话就是原始的、具有存储转发功能的 VRU。语音信箱是典型的存储转发应用实例,人们可以把要表达的意思说出来并录音,而被叫用户可迟些时候听取这些消息。

2. 识别键盘上的数字

用户可按话机上的 0~9 十个数字键,语音处理系统能接收、鉴别并按要求做出相应的响应。

3. 文语转换(Text-to-Speech)

文语转换(Text-to-Speech,或 TTS)指的是 VRU 在没有预录语音的情况下,能合成出音频消息。例如:123 可分为五部分读出:

“一”“百”“二”“十”“三”

如果能预存 40 个字的语音,构成词汇表,那么能读出任何一个小于十亿的数字(对英语而言)。

复杂情况下,TTS 能读出任何文本,如报纸、文章等。TTS 技术现已取得飞速发展,并得到应用,但仍存在一些有待解决的问题。

4. 语音识别(Voice Recognition)

语音识别也称作语音—文本转换,即把发音器官发出的声音转化为字、词等文本。在一般情况下,语音识别系统仅能识别有限的字词集合,比如“是”、“否”及 0~9 十个数字。语音识别系统主要分为二类:不依赖于人的识别(Speaker-independent)和依赖于人的识别(Speaker-dependent)。不依赖于人的识别可识别任何人的声音,而与发音的口音、方言以及电话线路的特性(如背景噪声、带宽等参数,它们随本地电话、长途电话以及蜂窝电话的不同而不同)关系不大。依赖于人的识别系统,根据使用者个人的发音进行训练,但可以做到大词汇量的识别。

5. 电话呼出

具有电话拨号特性的 VRU 能拨打电话号码,并接续呼叫。这类 VRU 要求能进行呼叫进程分析。比如确定拨号是否成功:接续是否成功;线路忙否等等。电话呼出的一个重要应用是电话系统里的呼叫转移,即 VRU 转接呼叫至某一分机,并检测电话是否被摘起。

6. 传真的存储与转发(Fax Store and Forward)

传真数据与语音信息一样,可存储起来并重发出去。语音处理系统可包括传真处理功能。例如,一个传真文件在语音信箱里可视为一条消息。

7. 传真合成

我们经常能见到,传真信息发自计算机,即通过扫描,把字处理文件、传单页、数据库转化为传真的格式并发送出去。传真合成指的就是 VRU 的一种功能,把一定格式的数据转化为传真的格式,无需打印等中间步骤直接地发送出去。

1-3 呼叫处理的主要应用

目前为止,呼叫处理技术的应用涉及以下方面:

1. 自动总机(Automated attendant)

具有自动总机功能的 VRU,能自动处理电话呼叫,并完成类似于接线员所完成的基本任务:把来话转接至本电话系统内的某一分机。一般情况下,自动总机会拨放一段预录的话音消息以应答呼叫用户,如:“您好,这里是 XYZ 公司。查询分机号请拨‘*’键;使用按键话机请直拨分机号;使用号盘话机请稍等……”。

2. 语音信箱(Voice mail)

前已提及,语音信箱的主要功能是存放语音消息,供接收者在方便时听取。一般情况下,呼叫用户会首先听到类似如下的问候语:

“您好,这里是某某某。很抱歉现在不能听您的电话。在听到‘嘟’的一声后请留言。”

每一个用户有一个属于自己的信箱,用来存放消息。信箱的主人

可通过拨打语音信箱系统的电话号码,并输入口令来存取信箱内的消息。在一般情况下,系统能说出消息录制、转发时涉及的日期和时间。系统能够把消息转发给另一信箱,或者转发至一组信箱,给每人一份消息拷贝。

许多语音信箱系统集成自动总机功能。当然有些公司更愿意设专人负责接电话,并向主叫提供选择,如果要找的人不能听电话,则转至语音信箱。

3. 消息传递(Message delivery)

当被叫不是本系统内的用户时,可通过消息传递系统把话音信息存储并转发给被叫。比如,一个在机场准备登机的公司经理,急于呼叫某个人,但被叫话机忙。这时可按一个“#”号,把留言录制下来。此后,系统会每隔一定时间间隔(如十五分钟)呼叫该用户,并把消息传递给他。

4. 可闻文本(Audiotex)

“可闻文本”指的是通过自动电话系统向用户提供语音形式的信息,如在美国流行的自动银行账务查询系统。通过拨打一个服务号码,输入银行帐号和密码,或税务 ID 号码,用户会听到系统对该查询的声音应答输出,包括:钱款总额、开具的支票、存款等内容。有些人也把这类系统称作交互式语音应答系统(Interactive Voice Response,或 IVR)。泛指用来从数据库中提取需要的信息的系统。

这类系统一般按给定的选择菜单来操作,如:

“如果您希望…请按 1;如果您希望…请按 2;如果您希望…请按 3”。

5. 按呼叫次数付费(pay-per-call)

在美国,呼叫以 976 为前缀的本地电话号码,或者长途电话呼叫 900 号,计费是只收一个基本费用。这些呼叫产生的附加利益归拥有这些号码的信息业务提供者。这类业务一般称作“按呼叫次数付费”。具有这种功能的某些业务号码由话务员来应答。许多“按呼叫次数付费”业务属于能提供预录信息或自动数据库存取的可闻文本系统。典型的例子是天气预报和体育报导。目前这类业务已在许多国家开展。

6. 自动订购系统(Automated order entry)

许多公司在商业运作中要向熟练的订购人员,如分销商、代理商和

大宗用户提供丰富的用户订货手段。建立在计算机基础上的 VRU 能受理这些用户的呼叫,并进行一项或几项订购处理,由提供的菜单选项辅助完成。

7. 呼叫中心(Call center)

无论是呼入还是呼出,都将从语音应答处理系统中受益。处理呼入时,VRU 能在无人受理时应答呼叫,提供娱乐或呼叫队列中的信息,或者是提供某产品信息的菜单选择。当然 VRU 还能接收来自电信运营公司的诸如 ANI(主叫用户电话号码)之类的重要信息,以便建立客户数据库。

处理呼出时,VRU 承担拨号器的任务,能根据数据库中存储的电话号码,试着完成每一个呼叫,并在被叫应答时转至业务受理人。

由于呼叫过程是静态进行的,所以会出现叫通时无人受理的情况,这时 VRU 可以请主叫用户保持接续,并播放一段消息。这个特性在社会调查中被广泛使用。

第二章 通信基础知识

2-1 基本概念

一、电话线

一般情况下电话机都是通过电缆中的二条导线连接至交换机上。电话交换机所在的交换局在国外也叫做中心局(Central Office, 或 CO)。功能与 CO 交换机类似的商业电话系统叫做专用小交换机(PBX 或 PABX)。

普通电话机通过二根导线与电话局相连,一般称作“二线连接”方式。也叫做模拟连接,这时传送的语音信号是用连续变化的电流表示的,而不是二进制比特流。通过公共交换电话网(PSTN)连接到电话局的电话可呼叫世界各地的电话用户。

有些商业电话系统(PBX)借助多条线(而不是二条线)把电话连接到 PBX 上的电话接线盒上。其中额外的线可用来实现一些附加的功能,如 LED 显示、会议等。一般“标准”的语音卡并不支持这些额外的信号。也有些 PBX 采用二线连接,但可能与商业电话的接口不兼容,原因是控制按钮和指示灯的信号是高频信号。CO 会为二条模拟线提供一个直流电压。

图 2-1 所示为 CO 交换机与电话机连接的示意图。

CO 交换机向模拟电话线提供电压,在交换机和电话机之间形成环路。电话机摘机时,会形成回路,在电话线间产生电流。电流的变化代表声音信号。电话机的二条线哪条接“+”极,哪条接“-”极关系不大。

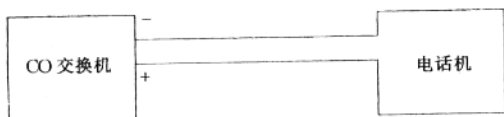


图 2-1 电话与交换机相连

二、开始——听拨号音

当话机上的话筒提起时,称作话机处在摘机(off-hook)状态。摘机相当于话机的叉簧开关关闭,这样沿着模拟线形成完整的回路,也叫本地环路(local loop)。其中流过的电流也叫本地环流(loop current)。当 CO 检测到环流时,会送出一个拨号音(dial tone),表示用户可以拨号了。

以上这种方式也叫“环流启动(loop start)”,大多数话机摘机属于这种情况;另一种是“接地启动(ground start)”,当检测到二条线中的一条接地时,表明有呼叫请求。

三、振铃——电话呼入

当话机上的话筒未被提起时,称话机处于挂机(on-hook)状态。挂机时,主电路拆线,但仍有一个电流流过一个电容器,如图 2-2 所示。该电容的阻抗非常大。当有呼叫时,CO 在电容的二端加上大约 150V 的交流电压(铃流电压),这样产生振铃。

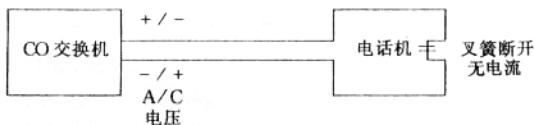


图 2-2 直流话机挂起情况

当话机摘起时,叉簧开关关闭,电路闭合,CO 停止供应铃流电压,而只保持正常的直流电压,这时环路电流流过电路,电流的变化代表讲话时话音信号的变化。

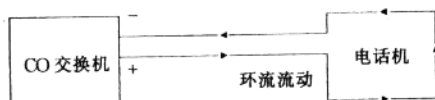


图 2-3 应答时话机提起情况

四、拨号

拨号有二种基本方式:双音频拨号和脉冲拨号。

双音频拨号使用音频信号来代表键盘上的数字(0~9, # 和 *)。每个数字一般采用二个频率的组合来表示,所以叫双音多频(Dual Tone Multi Frequency, DTMF)信号。

实际上有二种音频数字的编码标准:DTMF 和 MF(Multi Frequency)。MF 由电话局内部使用,有时也用于一些特殊业务,如自动号码证实(Automatic Number Identification, ANI)。ANI 功能也称为主叫用户证实(Caller Identification),主叫用户的号码会在被叫用户的话机上自动显示。在本质上, MF 与 DTMF 非常相似,差别是表示键盘数字时,采用的频率组合不同。有些语音卡能同时检测出 DTMF 和 MF 信号,而有些只能识别 DTMF。

脉冲拨号使用环流本身来发送数字。当话机极性反转时,产生 AB 线的打开/关闭变化,从而产生环流信号的通/断变化,表示环流的脉冲。对脉冲进行计数,即可知道所拨的数字是什么。所以,当一个数字拨完时,要有一个时间上的停顿,然后拨下一个数字。

脉冲拨号有二点不足。首先,脉冲拨号的速度比双音频拨号慢;其次,许多交换机不能传送这些脉冲。用脉冲话机拨号,并且在谈话期间拨脉冲数字,会听到“咯咯”声,但不会破坏环流。这意味着:如果对方想知道拨的是什么数字,唯一的方法是对声音样本进行分析,来“猜测”拨的是什么。举例来说,区分数字“1”和线路上的静态“咯咯”声并不容易。根据声音信号识别脉冲数字的功能模块叫“脉冲识别器”。由于还能把这些数字转为 DTMF,所以也叫作“脉冲—音频转换器”。

五、呼叫接续

当拨完号码后,拨号人听电话线上的信号就能判断出呼叫是否断了,什么时候断的。呼叫时会产生一系列呼叫进程信号,指明该过程的进展。多数呼叫进程信号是音频信号,由交换机产生,或通过环流下降产生。

铃流信号表示被叫用户线已加上了铃流电压。回铃音(ring-back)由 CO 交换机产生,而不是被叫电话机产生的。

如果接续时被叫用户的话机处于摘机状态,会产生一个忙音信号(busy)。如果电话局的网路忙,本地 CO 交换机不能把呼叫接续至被叫话机,会产生一个网路拥塞信号。一般情况下,网路拥塞信号很少听到,它与忙音信号类似,只是音的停顿稍短些。

如果拨打一个根本不存在的号码,或线路未连通的号码,会听到一个话务员中断信号(operator intercept)和一个录音通知:“很抱歉,您拨打的号码不存在。”

另外一种不太常见,但很重要的信号是环流下降信号(drops in loop current),一般在长途呼叫中产生,电话局把它当作接续至远端 CO 的证实信号,或者是被叫话机摘机的信号。

六、呼叫进程分析

识别呼叫进程中出现的各种信号叫作呼叫进程分析。多数语音卡能检测这些呼叫进程信号。呼叫进程分析是一个相当复杂的过程,有些环境下很难达到高的准确率。不同的环境,信号会有较大的特性偏差。例如:当 PBX 与电话局的交换机不同时,PBX 所产生的“振铃”和“忙音”信号也不相同。如果使用呼叫进程分析,语音卡在配置时必须能识别这些信号音。

呼叫进程分析一般采用呼叫处理算法,并给出可能的分析结果:

- 久振不在(Ring no answer)。已检测到振铃信号,但若干次振铃后没有应答。
- 忙(Busy)。检测到忙音信号。
- 网路拥塞(Fast busy)。检测到网路拥塞信号,表明无法接续至