

文章编号:1000-2073(2001)01-0040-07

一个汉字语音库的实现^{*}

吕 强,吴 嫻,杨季文

(苏州大学 工学院计算机工程系,江苏 苏州 215006;江苏省计算机信息处理重点实验室,江苏 苏州 215006)

摘 要:从多媒体计算机走进人们的生活开始,语音服务成为当今计算机界应用的热点之一。介绍了一个汉字语音库的实现,并且制定了基于这个语音库的由汉字机内码到汉字读音的访问方法。考虑到这种访问可能被普遍应用,所以将该汉字语音库制作成 DLL,以便被更多的应用程序调用。在这个语音库的基础上还给出了两个应用程序,作为对本成果的测试和应用示范。

关键词:汉字语音库;动态链接库;语音文件

中图分类号:TP311 **文献标识码:**A

0 引言

当多媒体计算机步入人们的生活,人们开始热衷于以计算机为中心的语音处理和服务。现在,人们只要在计算机上安装声卡、配置耳机或音箱就可以利用计算机欣赏音乐、观看电影。当我们用计算机处理汉字的时候,也自然希望能够听到该汉字的读音。

在计算机中汉字的机内码是唯一的,而由汉字输入码到汉字机内码的转换可以通过查阅输入码对照表实现。由此联想到,可以在得到汉字的机内码后,通过查阅某种表格,经过某种转换得到该汉字的声码。汉字语音库的设想由此引发。

这是一个为上层应用程序提供服务的底层环境。实现这个汉字语音库主要完成两项工作:第一,建立一个语音库,记录在普通话中所有汉字的读音;第二,建立汉字机内码到汉字读音的访问方法,实现由输入的汉字机内码得到该汉字的读音。在研究的过程中,发现这种转换可能被许多应用程序普遍应用到。为了使其他的程序方便地调用,实现机内码到语音的转换,又联想到了将它制作成 DLL(dynamic-link library,动态链接库)。

当汉字语音库制作完毕后,在此基础上编写了两个具体的应用程序。它们分别实现从键盘和电话按键获取汉字的区位码,通过汉字语音库的转换,经过语音文件的拼接,形成一段教学语音信息,然后,通过计算机上的媒体播放器或电话听筒将这些信息播放给用户。这

^{*} 收稿日期:2000-07-10

作者简介:吕 强(1965-),男,江苏苏州人,副教授,主要从事中文信息处理方面的研究。

两个具体的应用,可以作为对 DLL 的测试,也可以说是对这项新技术的应用。

1 语音库的实现

1.1 汉字语音库的实现

在普通话中,实际存在的发音有 1333 种^[1]。首先我们要将这 1333 个语音分别录制成 1333 个独立的 .WAV 文件,每一个读音都有它自己的序号。之后,再制作一个小的连接程序将它们按照读音的序号连接成一个大文件。应该注意的是,这种连接保留了每个 .WAV 文件的首部结构。这个大文件就相当于一个语音库了,取名为 YYK.WAV(其实它并非是真正意义的 .WAV 文件,但为了明了起见,仍然将它叫作 YYK.WAV)。至此,语音库的建立工作完成了,下面就是如何实现转换的问题了。

为了可以在 YYK.WAV 中截取对应的一段语音文件,需要建立两张索引表(HZINDEX.txt 和 YYINDEX.txt),分别记录汉字对应的读音序号和该读音在 YYK.WAV 中的偏移。以上这些是实现机内码到语音转换的基础。

当我们输入一个汉字后,就可以得到该汉字的机内码,取出它的高字节赋给变量 high 和低字节赋给变量 low,利用公式:

$$\text{start} = ((\text{high} - \text{B0h}) * 94 + (\text{low} - \text{A0h}) - 1) * 2;$$

得到输入的汉字在 HZINDEX.txt 索引文件中的偏移,然后从该偏移处取出一个字(WORD)。这是该汉字对应的读音序号,将它赋给 offset。由于在截取 YYK.WAV 文件时,不但要有偏移,还要有该汉字对应的读音文件的文件长度,因此需要得到下一个读音的偏移,赋给变量 nextoffset。此外,要对 offset 和 nextoffset 作如下处理:

$$\text{offset} = (\text{offset} - 1000) * 4;$$

$$\text{nextoffset} = ((\text{offset} - 1000) + 1) * 4;$$

(注:减 1000 是因为在建立 HZINDEX.txt 文件时将读音序号加上了 1000 之后存入其中,乘 4 是因为在 YYINDEX.txt 中以双字(DWORD)为单位存放汉字在语音库中的偏移)。其中 offset 是该语音在 YYINDEX.txt 中的偏移,nextoffset 是紧接着的下一个读音在 YYINDEX.txt 中的偏移。得到这两个参数后,就可以从 YYINDEX.txt 中的 offset 处得到一个双字(DWORD),赋给变量 yyoffset,在 nextoffset 处得到另一个双字(DWORD),赋给变量 nextyyoffset,用这两个变量就可以在 YYK.WAV 的 yyoffset 处截取 filelen =

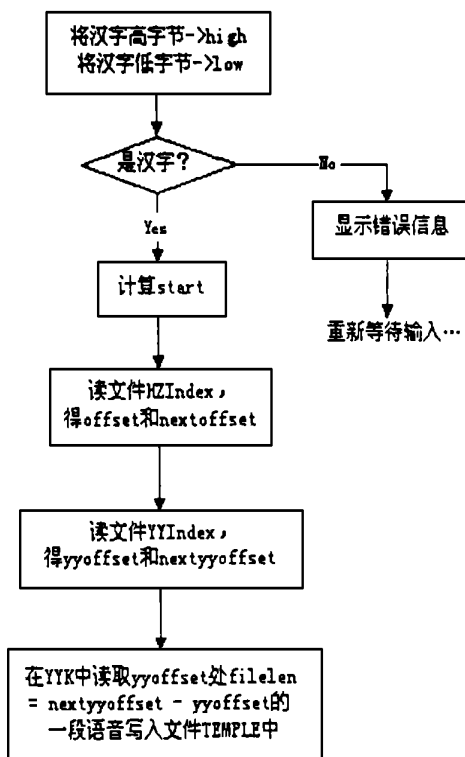


图 1 汉字机内码到语音的转换

(nex - tyoffset - yyoffset) 长的一段文件,存在临时文件 TEMPLE 中,这就形成了单个汉字的语音文件. 汉字机内码到语音的转换流程如图 1 所示.

至此,汉字机内码到语音文件的转换工作完成了. 这时应该说这个汉字语音库也实现了,通过标准的 .WAV 资源访问器就可访问语音库.

1.2 制作 DLL

从汉字机内码到语音文件的转换是一个可能被普遍应用到的过程,如果可以制作成 DLL 将会使其他一些应用程序很方便地调用这里的转换过程. 这一阶段,初步实现了 DLL 的制作,但仍有一些技术问题尚未解决.

1.2.1 理解 DLL

DLL 是一个包含了函数和数据的模块. 一个 DLL 库一般是由它的装入模块(一个 EXE 或 DLL)在运行时加载. 当运行一个调用了 DLL 库函数的可执行文件时,DLL 库就被映射到装入过程的地址空间.

DLL 提供了一种使用应用程序模块化的方法,这种方法使得功能模块能够很容易地更改和重复使用. 并且在几个应用程序同时使用同一个功能模块时,可以有效地减少系统的开销. 当允许共享代码时,通过给每个应用程序的数据做备份,可以实现这一功能.

1.2.2 制作难点及功能说明

制作 DLL 并非是一件困难的事,特别在有了完整的功能模块实现代码后,只要将程序的格式改成 DLL 格式即可. 当然为了更好的封装,对原先的功能模块也作了小的修改,主要将转移和合成 WAVE 文件的过程封装在 DLL 中. 在没有制作 DLL 时,很多问题只局限在单个应用程序中. 可是制作成 DLL 后,就需要考虑到方便多个应用程序的使用和有关资源共享等问题.

第一,在单个程序中,文件的存储路径是事先确定好的,然而制作 DLL 的目的就是让其他应用程序方便的使用,这就不可能事先规定好文件的存储路径. 在 Windows API 中有查找 Windows 目录的函数,可以在需要用到这个 DLL 的应用程序安装时将要用到的文件复制到 Windows 目录,在 DLL 中首先调用 GetWindowsDirectory() 函数,得到本机的 Windows 目录,再进行文件的读写操作.

第二,DLL 是一段可供多个应用程序调用的函数库,在这个 DLL 中需要对某些文件进行写操作,这样的文件是临界资源. 那么在多个应用程序同时调用这个 DLL 时就存在多个应用程序共享这些临界资源的问题,对这一问题可以用操作系统中对临界资源的加锁问题解决. 在 Windows API 中提供了 GetTempFileName() 函数,可以为不同的应用程序生成临时文件. 采用这种方法可以避免加锁的问题,从而实现多个应用程序的共享. 加锁的实现,是一个值得继续研究、开发的问题.

下面举例说明这个 DLL 的 Delphi 原型的使用^[5].

可调用的函数为:Make_Teach_File(qwcode:string;var isRight; integer);

输入参数:汉字的区位码(string 类型).

输出参数:对于错误的输入或操作出错,isRight 的返回值不同,分别是:

- 0 输入为空或者含有非数字字符;
- 1 输入的区位码非法;
- 2 打开文件出错;

- 3 创建文件出错;
- 4 输入正确,返回在特定目录下的一段语音文件,名为~1~2~3~.wav.

2 汉字语音库的示范应用

为了对这项新技术作一个全面的测试,开发了两个具体的应用程序,它们都调用了上文介绍的 DLL. 这两个应用程序可以看作是对 DLL 和汉字语音库的测试。

2.1 简单文字语音合成应用

功能:接受键盘输入的区位码,合成播放该字的纵横码教学信息。

这个应用程序实现在计算机上用键盘输入汉字的区位码,通过汉字语音库的转换和 .WAV 文件的合成制作成一段可以播放的语音文件。这段语音文件的内容是纵横码的教学信息,告诉用户这个汉字的纵横码是什么。该应用程序的总体流程如图 2 所示:

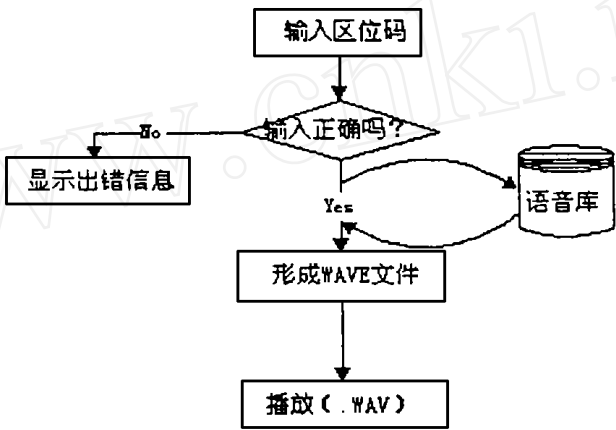


图 2 应用一总体流程

在该程序中,为了得到汉字的纵横码,需要查两张数据库表,它们分别是 ZH1.DBF 和 HZ.MDB. 这两张表的结构如表 1、表 2 所示:

表 1 纵横码码表(ZH1.DBF)

字段名称	字段类型	字段宽度	字段说明
Flag	数值型	1	汉字纵横码级别
hz	字符型	2	汉 字
Srm	字符型	6	汉字的纵横码

表 2 汉字区位码表(HZ.MDB)

字段名称	字段类型	字段大小	字段说明
ID	自动编号	长整型	按照区位码的顺序编号
HZ	字符型	2	汉 字

在形成最终的教学信息时,用到了 .WAV 文件的拼接,因此首先要对 .WAV 的结构作一个了解,表 3 是 .WAV 文件首部的格式^[2]。

在实际的编程过程中发现,上述格式是早期使用的一种形式,现在使用的是一种扩展型的格式。其文件首部共有 58 个字节,在说明子块中加入了一些其他的参数指标。但是这种改动并未影响到课题的实现,因此在这里就不作多余的说明了。

.WAV 文件的拼接必须对文件首部和音频数据块分别进行。对于连接好的 .WAV 文件

要修改文件首部中的文件长度和数据长度(分别是偏移为 4 和 53)两个字段. 将通过汉字语音库转换得到的单个汉字的语音文件和诸如“的纵横码是”、“或者是”的 .WAV 文件进行拼接,存放在指定目录下供播放设备调用. 拼接流程如图 3 所示.

表 3 .WAV 文件头格式

偏移量(字节)	内 容
0 ~ 3	资源交换文件标志(RIFF)
4 ~ 7	文件长度
8 ~ 11	WAVE 声音文件标志(WAVE)
12 ~ 15	波形格式标志(fmt)
16 ~ 19	过渡字节(一般为 10H00H00H00H)
20 ~ 21	格式种类(值若为 1,则表示数据为 PCM μ 律编码的数据)
22 ~ 23	通道数,单声道为 1,双声道为 2
24 ~ 27	采样频率,表示每一个通道应以此速度播放
28 ~ 31	波型音频数据传输速率,其值为通道数 * 每秒的数据位数 * 样本数据位数/8,播放软件根据此值可以估计缓冲区的大小
32 ~ 33	数据块的调整数(按字节算),其值为通道数 * 样本数据位数/8,播放软件需要一次处理多个该值大小的字节数据,以便将其值用于缓冲区的调整
34 ~ 35	样本数据位数,表示每个声道中各个样本的数位数,如果是多声道,则每个声道样本大小一样
36 ~ 39	数据标志符(DATA)
40 ~ 43	声音数据长度

2.2 电话声讯服务

功能:电话按键输入汉字区位码,电话听筒输出该字的纵横码教学信息.

这个应用程序通过电话按键输入汉字的区位码,仍然是利用汉字语音库和 .WAV 文件的拼接形成教学信息,从电话的听筒中播放出来. 该程序的总体流程如图 4 所示.

该程序只是对上一个应用程序稍作修改,使其成为“基于 Web 技术的电话语音服务系统”中的远程教学模块. 它的实现基本上与上一个应用程序的实现相同. 这里将不同点作一个说明.

首先,在计算机端,用户的输入是在 Edit 框的一串字符串,对它通过类型转换(转换成整型)后可以进行不同的处理. 在电话端,用户通过电话按键输入的是一种叫作 DTMF(Dual Tone Multi-Frequency,双音多频)码,通过电话卡提供的 GetDTMFStr() 函数接收到汉字的区位码(这时已经转换成字符串类型),之后进行的处理与计算机端相同.

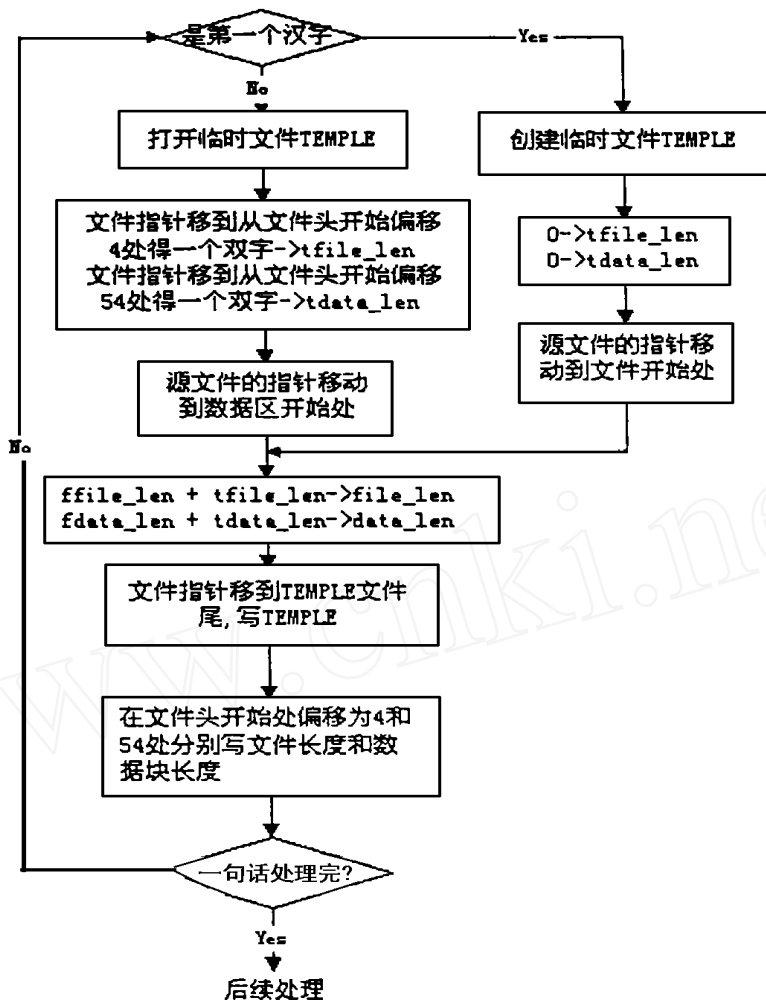


图3 .WAV文件的拼接

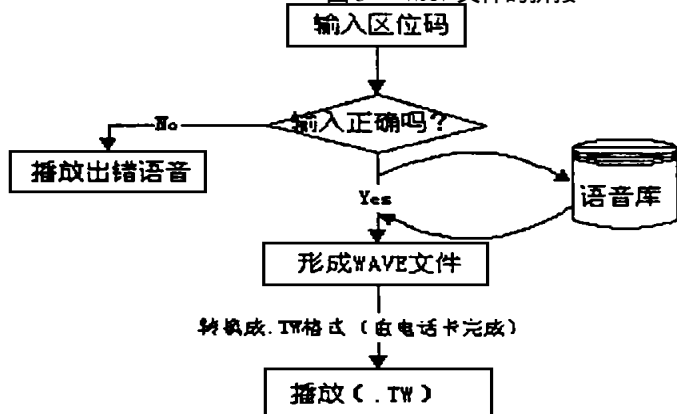


图4 应用二总体流程

其次,涉及出错信息的播放和最终教学信息的播放.在计算机端实现的时候,出错信息通过显示器可以显示在用户面前;教学语音通过媒体播放器播放.电话端的出错信息和教学语音都要通过听筒播放给用户.这里需要将程序中合成的 WAVE 文件转换成特定电话卡可以接受的格式,如 .TW 格式.一般来说,电话卡中提供这样的转换工具.

3 结束语

本文建立了一个汉字语音库并制定了由汉字机内码到汉字读音的访问方法.基于这个语音库,提供了由汉字机内码到汉字读音的转换,并且将这个汉字语音库制成 DLL 供更多的应用程序调用.在此基础上,我们演示了两个具体的应用程序,作为对本文成果的测试.

参考文献:

- [1] 赵珀璋,徐力.计算机中文信息处理[M].北京:宇航出版社,1987.
- [2] 周敬利,余胜生.多媒体计算机声卡技术及应用[M].北京:电子工业出版社,1998.
- [3] 李莹.Web 语音信箱系统的设计与实现[D].苏州大学,1999.
- [4] 霍晓军.基于 Web 技术的电话语音服务系统的设计与实现[D].苏州大学,1999.
- [5] WALLACE N, TENDON S. Delphi2 开发人员指南[M].北京:机械工业出版社,1997.

An implementation of Chinese character voice library

LU Qiang, WU Xian, YANG Ji-wen

(Dept. of Compu. Engin., School of Engin., Suzhou Univ., Suzhou 215006, China; Key Lab of Compu. Information Processing Tech. of Jiangsu Province, Suzhou 215006, China)

Abstract: Multimedia computer makes voice service a popular hot point in computer application. This paper introduces an implementation of Chinese character voice library, and provides the access method to every of Chinese character pronunciation based on this library. We distribute this library as a DLL due to the vast amount of Chinese voice application. Based on this voice library, two sample applications are presented.

Key words: Chinese characters voice library; dynamic-link library; voice file

(责任编辑:周建兰)