

· 数据库技术及应用 ·

## 语音合成系统中语音库的设计与实现

朱亚喆 柴佩琪

(同济大学计算机科学与工程系 上海 200092)

**摘要** 在基于拼接方法的语音合成系统中, 建立一个合适的语音数据库是非常重要的, 语音数据库的好坏对合成语音的质量有很大影响。该文从实际应用出发, 介绍了一个语音数据库的设计及其实现。

**关键词** 语音库 语音处理 质量

### The Design and Implementation of Voice Database in Speech Synthesis System

Zhu Yazhe Chai Peiqi

(Dept. of Computer Science, Tongji University Shanghai 200092)

**【 Abstract 】** It is very important to establish an appropriate voice database in speech synthesis system based on concatenation. The quality of voice database has direct affection to synthetic speech quality. In this paper, a method of how to design and implement a voice database is introduced.

**【 Key words 】** Voice database; Speech signal processing; Quality

语音合成技术经过几十年的发展, 已经取得了很多的研究成果。早期的合成方法以发音器官模拟合成法和规则合成法为主, 但采用这些方法合成的语音自然度和可懂度都不高, 因此不能付诸实用。随着计算机硬件水平的不断提高, 越来越多的研究人员已将其研究重点转向拼接合成法。

#### 1 语音库的设计

语音数据库不同于一般的关系数据库, 它有着自身的许多特点。比如要对语音进行查询将是一件非常困难的工作, 它涉及到语音识别问题; 对语音段进行索引也是较为困难的。但从数据管理来说, 语音库与一般的数据库有许多相似之处。在系统中, 对语音库的利用主要是为合成某段已知的文本提供语音波形数据, 而不对语音进行查询, 因此在设计语音库时选择一个现有的数据库管理系统来做数据管理工作。在建立语音数据库的过程中, 通常要考虑以下 3 个问题: (1) 数据库系统的选择; (2) 语音单元的选择; (3) 语音单元的分割和提取。\*

##### 1.1 数据库系统的选择

现在, 有许多科研人员在进行多媒体数据库理论方面的研究, 多媒体数据库的理论基础还不十分完

善。当前一些多媒体应用基本上也都是利用现有的具有管理大规模变长数据能力的关系数据库系统来作为其数据管理的工具。因此在设计语音库时, 首先也是选择一种数据库管理系统作为建立语音库的基础。

在语音库中, 由于需要存储大量的语音波形数据, 其特点是数据量大, 长度不固定, 所以在选择数据库系统时主要考虑: (1) 该数据库能对数据量大, 不定长的数据进行管理; (2) 易于实现, 通用性好。由于语音库与一般的用于事务处理的数据库有些区别, 它对事务处理和恢复、安全性和对网络的支持等方面要求较低, 因此, 在比较了多个数据库管理系统后, 最后决定采用 FoxPro 作为建立语音库的基础。

##### 1.2 语音单元的选择

在建立语音库之前, 必须要选择好基本语音单元, 以便进行语音合成时最为方便和高效。由于受到数据库数据存储能力和硬件设备的限制, 因此语音单元的数目应该越少越好; 从另一个角度说, 每一个语音单元大一些能减少拼接时的工作, 并且能提高合成

\* 朱亚喆 男, 24 岁, 研究生, 主要研究方向: 语音识别与语言处理

收稿日期: 1997-09-01

语音的质量。

在选择语音单元的过程中,主要考虑到应该保证实际应用中的语音质量,因此在选择语音单元时,以单字、二字词、三字词和四字词作为基本单元,在每个语音单元中,存放相应的语音波形数据。这些数据都是 11kHz 采样频率下,单声道的 8 位 PCM 编码数据。将这些数据直接存放在 FoxPro 的 MEMO 字段中。

### 1.3 语音单元的分割和提取

在语音录入时,在一个这安静的环境中录制。根据对一批实验用录入语音进行分析,发现录入的语音

中基本无噪音,但语音单元的首尾都有一定数量的静音数据。这些静音主要是因为是在录音过程中录入者按下操作按钮后并未马上开始录音而产生的。这种情况在实际操作过程中很难避免,而这些静音数据若与语音单元一起存放在语音库中,对于以后进行语音合成是非常不利的,因此必须将这些静音数据从原始语音数据中剔除掉,从而得到需要的语音单元。

我们的任务实际上是一个简化了的语音起始点的检测问题,图 1 是一段录入的原始语音文件的波形图:

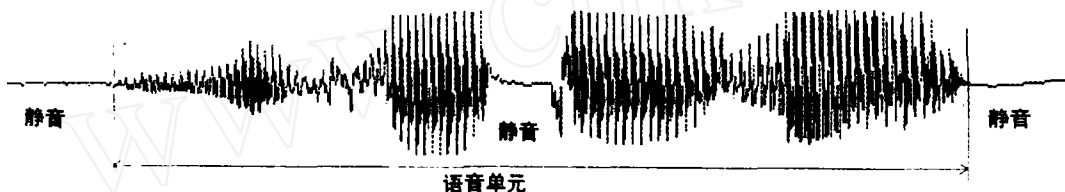


图 1 原始语音文件的波形图

由图中可以看到,除了在语音单元的开始和末尾处,在语音单元的中部也还存在着一些静音段,这些静音段的存在在合成时对语音的自然度有着较大的影响,因此这些静音段应该予以保留。

能量的方法来判断静音段,并取出所需的语音单元。

离散时间信号的短时能量定义为:

$$E_n = \sum_{m=-\infty}^{\infty} [X(m) \cdot h(n-m)]^2 \quad (1)$$

在计算过程中,选取的窗函数为矩形窗,其长度定为 51,即相当于 5ms 的持续时间。另外为了便于计算,用样本幅度的绝对值来代替平方值,即:

$$E_n = \sum_{m=-\infty}^{\infty} |X(m)| \cdot h(n-m) \quad (2)$$

$$h(n) = \begin{cases} 1 & 0 \leq n \leq 50 \\ 0 & \text{其它} \end{cases} \quad (3)$$

由上可得:

$$E_n = \sum_{m=n-50}^n |X(m)| \quad (4)$$

根据式(3),对原始音段的样本数据逐点计算短时能量值,可得到该音段的短时能量序列。为了判断某音段是否属于静音段,设定一个阈值 Voice\_Energy,用计算得出的  $E_n$  ( $n=1, \dots, N$ ;  $N$  为语段数据长度)与该阈值比较。若从某采样点起,样点的短时能量值小于该阈值,则视为静音段的开始;若大于该阈值,则视为静音段的结束。根据以下算法可得出语音单元的起点和终点。

因为在实际应用的过程中,录入的语音信噪比高,因此采用这种方法进行语段分割是非常有效的。根据设定的录音参数,阈值 Voice\_Energy 取 150-200 较为合适。

(下转第 64 页)

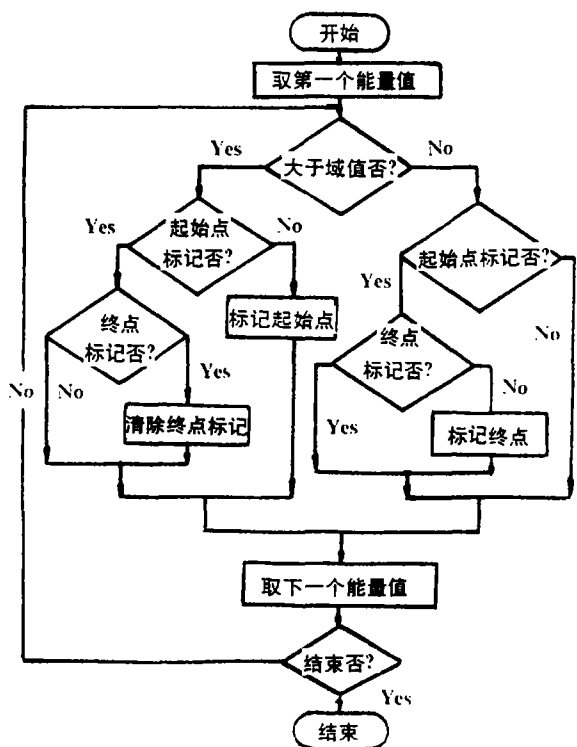


图 2 确定语音单元起点、终点的流程图

对于语音段起点和终点的检测问题,现在已经有许多方法来实现。在此,通过计算原始语音段的短时

### 2.3.2 我们的 OLAP 工具

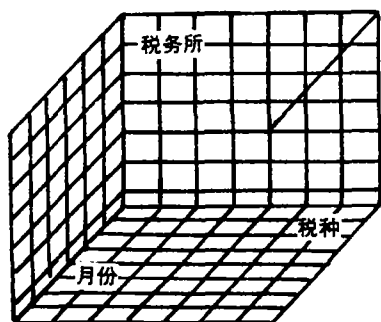


图 6 关于税收的 3 维数据空间

我们的 OLAP 工具建立在 4 维数据空间之上，每一个 4 维实体都面向一个主题，这 4 维是：行、列、部门、日期，图 6 是一个 3 维实体的例子。4 维实体是特定格式和内容的行列 2 维表在日期和部门轴上的连续移动生成得到的，它通过我们的 OLAP（在

~~~~~  
(上接第 46 页)~~~~~ 绑定, 而 CString 只接受字符串, 对于二进制数据的

## 2 语音库的实现

用 Microsoft Visual C++4.0 作为开发工具来实现语音库的管理。VC 作为一种常用的软件开发平台,其多媒体处理和数据库操作功能是十分强大的。通过 VC 提供的 MFC 类库和 Windows 的 API,可以方便地进行数据组织和处理。

## 2.1 语音数据处理

通过 Windows 的 API 可方便地进行语音的录入和输出操作。在 VC 中可以通过两种方式来操纵语音：第一种方式是通过 MCI 函数，发送 MCI 命令消息或命令字符串给多媒体设备驱动程序，再由驱动程序去驱动多媒体设备。MCI 是一个高层的多媒体控制接口，当只进行语音录入和输出时，MCI 使用起来十分方便。第二种方式是通过 WAVE 低阶函数来控制语音。用低阶函数可控制语音处理的进程，使语音处理更加高效。在系统中采用了第二种方式对语音数据进行处理。

## 2.2 数据库操作

在用 VC 实现数据库操作之前，首先要通过 ODBC 管理程序来建立一个与语音库的连接，然后通过 CRecordset 类来建立变量与数据库字段的绑定。在通过 Wizard 建立与数据库连接时，VC 会自动把 FoxPro 的 MEMO 字段与 CString 的变量进行

线 4 维) 数据生成器——即数据引擎得到的。基于这些面向主题的 4 维数据实体, 可对它的任意断面进行查询、旋转、图形显示, 可进行四维数据的相关分析、趋势分析等。

### 3 结论

我们将 DSS 用于税务系统，基本上是成功的，但还有大量工作要做。我们的实践表明基于数据仓库技术、数据挖掘技术和 OLAP 技术的 DSS 系统，具有广泛的应用前景。我们采用的 4 维数据空间结构及其建立在 4 维空间之上的 OLAP 分析方法是一个灵活实用的系统。我们在数据挖掘方面应用神经网络进行税务稽查选案也是比较成功的尝试。

## 参考文献

- [illegible]

### 3 结束语

本文介绍了建立语音库的一般方法,利用该语音库进行语音合成时取得了较好的效果。但在该语音库中还缺少语音的参数信息,因此对于该语音的利用还是十分有限的。若需利用语音库实现其它的语音处理功能时,还必须对该语音库进行扩展,根据需要增加一些字段,使语音库的辅助功能更加完善。

## 参考文献

- 1 Lynch J F Jr.Speech/ Silence Segmentation for Real-time Coding via Rule Based Adaptive Endpoint Detection.ICASSP-87, 1987, 3(4): 1348-1351
- 2 李国辉.多媒体数据库系统的设计考虑.NCMT'97论文集, 1997-10: 272-275
- 3 艾克拜尔.语音合成语料库管理系统的研究和设计.NCMT'97论文集, 1997-10: 292-296
- 4 [美]Corporation Q, 孟小峰译.ODBC2.0使用大全.北京: 清华大学出版社, 1995
- 5 [美]Shammas N C著, 王国印译.Visual C++使用指南.北京: 清华大学出版社, 1995