

# 语音控制计算机

俞铁城 鲍玉珍 王俊生 荣美玲 张恩耀

(中国科学院声学研究所)

## 一、引言

用语音直接控制机器(计算机)从而实现真正的人与机器对话,这是人们早就想实现的。我们平常说的某种计算机语言(例如 BASIC 语言),是人机对话式语言,那是通过键盘、打印机进行文字对话,用户要敲键符,将信息告诉计算机,不是语音对话。用语音控制而不是键控制计算机的好处是显而易见的。人们将可以用语音方便地直接控制计算机,通过计算机再去控制其它机器工作,从而实现更程度的自动化。这在机器操作员忙碌时尤感需要。例如,显微镜的用户可以不停止观察而用语音不断地将观察结果报给计算机。

语音识别的原理可以从语言识别要解决的三个主要问题来了解:

1. 决定采用什么特征参数来表示语音。应该选择对区分语音特别有用、稳定性好而又容易提取的参数。我们用  $1/3$  倍频程实时频谱分析仪给出随时间变化的频谱,作为特征参数。每 10 毫秒采一个频谱,每个谱中至多只包含 16 个通道的数据。

2. 如何训练机器,建立参考样板。机器是死的,得先受训练,存贮好参考信息后,才能识别。就语音识别来说,训练机器就是按既定顺序唸一遍(或几遍)字表 1 中的项目,让机器逐个记住它们的特征。由于发音有轻、重、快、慢的变化,在建立参考样板时要对这些因素进行自适应处理,否则,在识别比较时将遇到麻烦或容易出错。对每次发音,我们总是选出固定数目的频谱来表征它,这样就可实现时间域规正<sup>[1]</sup>。

我们用二值频谱[即每个谱分量只有两种值(0或1)]来解决音量的规正。

3. 决定采用什么计量技术来计算待识别音与参考样板音之间的相似性或差距。目前,我们采用两个二值谱之间的“异或”(模 2 加)距离来计算它们的差距<sup>[2]</sup>。挑选总的相似性最好者或总差距最小者为识别结果。

本文将介绍一个用语音直接控制计算机算题的软件系统(代号为 VCCV 620)。人机对话的方式是,人说话,计算机用文字。它可以进行加、减、乘、除和乘方等五种运算。系统的关键部分是要计算机能听对用户的口呼指令。这方面的工作已陆续发表<sup>[1,2]</sup>。所取得的结果是:实时识别 10 个数字和 20 句话,精度 99.7%;能以 99% 以上的精度实时识别多至 150 个短语;能实时识别 200 个短语(精度 98.85%);接近实时地识别 400 个短语(精度 97.7%)。

## 二、语音控制计算机系统简介

语音控制计算机系统如图 1 所示。其中计算机是通用机 Varian 620/L,有内存 8 K 字(计算机领域中  $1 K = 2^{10} = 1024$ ),存取周期 1.8 微秒,缺外存设备(可接包括外存设备如磁盘、磁带机在内的 64 种外围设备),字长 16 位。这是一台小型通用电子计算机。进行语音识别时,用实时频谱分析仪对经由传声器拾进的话音进行  $1/3$  倍频程的实时频谱分析,这便是计算机外面的预处理。模拟式的频谱数据,经 A/D(模数转换器)转换成数字式频谱,送入计算机进行处理,识别结果可在电传打字机上输出。

所用的  $1/3$  倍频程实时频谱分析仪是丹麦

表1 VCCV 620 用的 31 条命令<sup>1)</sup>

| 代号 | 反应字符(串)    | 汉语呼音   | 英语呼音           |
|----|------------|--------|----------------|
| 0  | 0          | 0      | Zero           |
| 1  | 1          | 1      | One            |
| 2  | 2          | 2      | Two            |
| 3  | 3          | 3      | Three          |
| 4  | 4          | 4      | Four           |
| 5  | 5          | 5      | Five           |
| 6  | 6          | 6      | Six            |
| 7  | 7          | 7      | Seven          |
| 8  | 8          | 8      | Eight          |
| 9  | 9          | 9      | Nine           |
| 10 | ●          | 点      | Point          |
| 11 | +          | 加上     | And            |
| 12 | -          | 减去     | Sub            |
| 13 | -          | 负的     | Minus          |
| 14 | *          | 乘以     | Multiply       |
| 15 | /          | 除以     | Devide         |
| 16 | ↑          | 方次     | Exponential    |
| 17 | =          | 等于     | Equals         |
| 18 | 1 次 bell   | 请等一等   | Wait           |
| 19 | OK bb      | 请 算 题  | Begin          |
| 20 | READY @    | 开始     | Are you ready? |
| 21 | ### b      | 接着往下算题 | Continue       |
| 22 | WELL b     | 请暂时休息  | Pause          |
| 23 | bBACK @    | 中断休息   | Start again    |
| 24 | THANKS@    | 工作结束   | Stop           |
| 25 | bNEXTb     | 再算一道题  | Next one       |
| 26 | bb {F12.4} | 打印结果   | Type           |
| 27 | @          | 回 车    | Return         |
| 28 | ←          | 听 错 啦  | Delete         |
| 29 | @          | 全部重来   | Clear          |
| 30 | 5 次 bell   | 继续工作   | Go on          |

1) 表中@代表回车与进行, b 表示空格, {F12.4} 是打印格式,表示占用 12 格,保留四位小数。

生产的 B/K 3347。附带说明一下,我国南京无线电仪器厂有同类产品,名为 BP-7 型数字式动态频谱分析仪,只要配上接口板即能连机使用。

### 三、题目内容的格式与运算规则

题目的格式如下:

其中 SS...S 为由数字及小数点构成的一个数据,例如 3.1416。

表2 31 条命令分类表<sup>1)</sup>

| (一) 算题前的控制命令 (2 条)  |              | (三) 报题目过程中的控制命令 (5 条) |           |
|---------------------|--------------|-----------------------|-----------|
| 呼 音                 | 电传打印         | 呼 音                   | 电传打印      |
| 请算题                 | OK bb        | 请等一等                  | { 1 次铃声 } |
| 开始                  | READY        | 继续工作                  | { 5 次铃声 } |
| (二) 报题控制目的命令 (19 条) |              | 回 车                   | @         |
| 呼音                  | 电传打印         | 听 错 啦                 | ←         |
| 0, 1, ..., 9        | 0, 1, ..., 9 | 全部重来                  | \@        |
| 点                   | ●            | (四) 算完一道题后的控制命令 (5 条) |           |
| 加上                  | +            |                       |           |
| 减去                  | -            |                       |           |
| 负的                  | -            | 呼音                    | 电传打印      |
| 乘以                  | *            | 接着往下算题                | ###b      |
| 除以                  | /            | 请暂时休息                 | WELLb     |
| 方次                  | ↑            | 中断休息                  | bBACK@    |
| 等于                  | =            | 工作结束                  | THANKS@   |
| 打印结果                | bb{F12.4}    | 再算一道题                 | bNEXTb    |

1) 表中 @ 代表回车与进行, b 表示空格, {F12.4} 是打印格式,表示占用 12 格,保留四位小数。

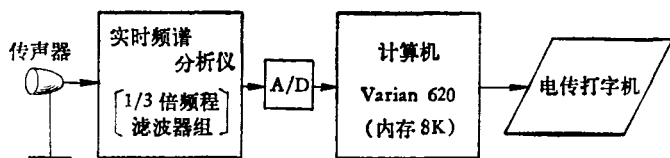
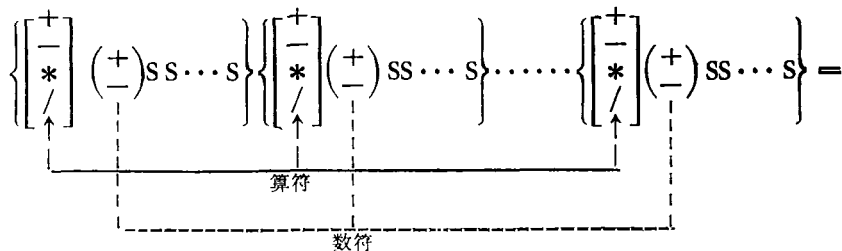


图1 语音控制计算机系统框图

运算规则是自左向右运算。五种算符 (+, -, \*, /, ←) 为同一等级算符,即不进行先乘除后加减这样的优先运算。作这样的规定,完全是为了简化解释程序和节省程序。只要计





均无效;(3)“回车”是供电传打印机打字头需要回车时使用的,跟题意无关,只是让打字头回车与进行而已;(4)“听错啦”是告诉计算机,它听错了最后一条命令,应予消去,用户接下去重报命令,显然这是用来纠错的。这条命令可以连呼多次,以便将尾部的多个字符消去;(5)“全部重来”是用来清除整个题目的已呼部分,用户重新从头报题目内容。当发现差错较多或错了的字符已在好几个字符之前时,则用本命令。

算完一道题后,有四种状态可供选择:(1)在已算出的结果上继续往下算,则呼“接着往下算题”;(2)算另外一道题,则呼“再算一道题”;(3)让计算机等一会,用户要查询一下什么或休息片刻,则呼“请暂时休息”。计算机一旦处于“暂时休息”状态,则只有呼“中断休息”才能叫计算机返回来工作;(4)工作完毕,要计算机停机,则呼“工作结束”,此时电传打印机上敲出 THANKS, 并走一节纸,计算机由运行状态进入停机状态。

由上述可知,有一些命令,如两组控制计算机休息及继续工作的命令,是多余的。之所以这样设计字表,无非是有意增加识别难度而已。

## 五、程序框图说明和口呼算题例子

程序框图(见图2)。SPEECH 为识别语音的子程序,识别语音时用到它,训练机器时也用到它。当 SPEECH 用来识别语音时,则由它返回时即知道识别结果(字表1中某项目的顺序号)。这个关键性的子程序极为精炼,光指令部分仅占用不到500个计算机字。

训练计算机就是按字表1呼一遍。

当计算机面板上的一个开关(S1)使用时,则可单纯地检查已存进的样板音的可靠性。此时,呼一个音,电传打印机上即打印出一个代号。如果呼音比较纯熟,就不必检查样板音。计算机一旦训练完毕,即可开始算题。附录是一

次口呼算题结果,均正确无误。

## 六、讨 论

1. 本文介绍的话控计算机算题系统,只不过是计算机实时识别语音的一个应用示范而已。读者必定能想像到计算机受话音控制后将能干很多很多的工作。例如,自动邮件分类;产品检验;生产线的自动控制;密语通信;资料查询;话音数控机床的程序编制;话音编辑器;等等。

2. 由于 VCCV 620 的关键部分是实时识别语音子程序<sup>[2]</sup>,它并不对固定的字表工作,即命令表可以任意更换。呼音可以用方言,甚至可以用外语(近来,我们已多次用英语字表向国内外来宾公开表演);它要求的存贮量特别小,而又具有高精度的识别率;当字表小于200时均能实时识别;因此,可望有相当宽的适用面。

3. 解决换人的办法是,新用户先训练一遍计算机。在字表不大的情况下,尚不麻烦。而且如果有外存设备的话,不必每次训练一番,只需将易错的命令更新一下便行了。笔者本人发现自己一年前的样板音基本上还能用。

4. 随着组件集成度的提高和计算机的迅速推广使用,话控机器(包括计算机本身)的前景是光明的。可望在不久的将来,我们的语音识别研究结果,会对自动化作出应有的贡献。

5. 迄今为止,识别程序已精炼到完全具备制作硬件的条件。如果制出专用硬件,识别速度还将大大提高。

对语音识别有兴趣的读者可参考本文所引的文章及其中提到的文献。

## 参 考 文 献

- [1] 俞铁城,物理学报,26(1977),389.
- [2] 俞铁城,物理学报,27(1978),508.

(下转第313页)