



[12] 发明专利申请审定说明书

[21] 申请号 86100123

[51] Int.Cl⁴

G06F 3/153

[44] 审定公告日 1989年2月1日

[22] 申请日 86.1.10

[30] 优先权

[32] 85.2.15 [33] DE [31] P3505276.7

[71] 申请人 西门子公司

地址 联邦德国慕尼黑维特尔斯巴赫广场2号

[72] 发明人 乔基姆·海因策尔

[74] 专利代理机构 中国专利代理有限公司

代理人 匡少波

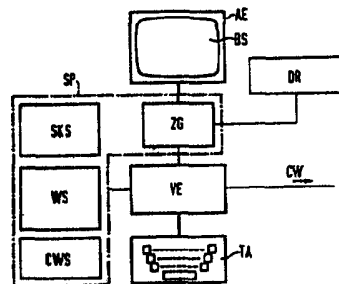
说明书页数: 9

附图页数: 3

[54] 发明名称 汉字输入终端的方法与装置

[57] 摘要

汉字的输入采用拉丁字母拼音。与汉字相应之每个音节对应一个音节标识码(SK)。由于在拼音中同一音节可以对应列出很多汉字(CHZ), 所以利用字存储器(WS)给与字对应的音节标识码(SK)组配属一组序号(OZ), 序号表明这指的是与同一音节标识码(SK)的音节相对应的第几个汉字(CHZ)。利用音节标识码(SK)和所属序号(OZ)在代码存储器(CWS)中找出相应汉字的国标码(CW)。



<39>

1. 一种汉字输入终端之方法, 这里汉字组成的词由拉丁字母拼音输入, 其中一个音节对应一个汉字 每一个以拼音输入的词的一个音节都有一音节识别码, 在一存储器内存有一音节顺序, 其特征是: 用拼音输入之词每个音节都对应一个音节标识码(SK), 在词存储器中对每组音节存贮一组音节标识码(SK), 各组音节标识码都对应有序号组(OZ), 序号给出与同一个音节标识码(SK)相对应的多个汉字中相应的汉字, 并且根据音节标识码(SK)及所属序号(OZ)找出与相应字对应的汉字(CHZ)和代码。

2. 根据权利要求1 所述的方法, 其特征是: 词存储器(WS)对出现的每个音节标识码(SK)都有一个自己的存贮区。

3. 根据权利要求2 所述的方法, 其特点是: 每组音节标识码(SK)的输入严格按照音节标识码(SK)排列, 并且首先按第一音节的音节标识码(SK), 然后才按第二音节的音节标识码(SK), 第三音节等等。

4. 根据权利要求3 中之一所述的方法, 其特征是: 用取出项在相应的音节标识码(SK)范围内查找一个字, 是用取出项在查找范围的中央进行, 并且在每个取出项都要查清, 如果有存入的话, 那么它是否包括在取出项之前或之后的范围内, 所以每个取出项都使要查找的范围缩小一半。

5. 根据权利要求4 中之一所述的方法, 其特征是: 如果在词存储器(WS)查找到一个以上与输入音节组相对应的项目, 那么首先显示第一组汉字(CHZ), 按了相应的键之后, 或确认这组或显示下一组。

6. 根据权利要求5 中之一所述的方法, 其特征是: 在较长词中作为汉字组存在的单音节或多音节词通过追加输入较长词格外加以标志。

7. 根据权利要求6 中之一所述的方法, 其特征是: 由多个词组成之成语通过输入分隔符代替各个字之间的空格, 如同多音节词一样处理,

从而使词组如同多音节词一样可以在词存储器(WS)中查找。

8. 根据权利要求7中之一所述的方法,其特征是:往词存储器(WS)中插字,可通过键盘(TA)与输出装置(AE, DR)对话的方式,将各个汉字组合一起,并按着存入词存储器(WS)中,插入完毕以后在词存储器中又自动重新排序。

9. 根据权利要求8中之一所述的方法,其特征是:在词存储器(WS)中删除字,字经过键盘(TA)输入,并在输入删除命令后即删除,同时在内存删除以后,词存储器(WS)中便又自动重新排序。

10. 根据权利要求1至9中之一所述的方法,其特征是:如果字在词存储器(WS)中未查找到,便在输出装置(AE, DR)做出推荐,看起来可能象汉字组(CHZ),由几个较短的字组组成,适合于所查找字的一部分,并且是用与其音节相对应的最常用的汉字。

11. 实施根据权利要求1的方法的装置,其特征是:一个音节标识码存贮器(SKS),对输入的每一个音节标识码(SK)存储;一个词存储器(WS),其特征是,通过音节标识码(SK)组及所属之序号(OZ)存贮很多字,序号说明指的是用同一个音节标识码(SK)标志的许多汉字(CHZ)中的第几个字,备有一处理器(VE),它给每个音节对应一个音节标识码(SK),通过词存储器(WS)找出与每个音节相对应的序号(OZ),并根据每个音节标识码(SK)及所属的序号(OZ)找出相应的汉字(CHZ),配备有一代码存贮器(CNS),存贮与音节标识码(SK)及所属序号(OZ)相对应的代码(CW),代码时时与汉字(CHZ)对应。

12. 根据权利要求11所述的装置,其特征是:规定配备一个符号发生器(汉字库ZG),存贮与汉字(CHZ)字形相对应的数据字,并根据代码(CW)将与汉字(CHZ)对应的数据字输送给输出装置(AE, DR)。

13. 根据权利要求11或12所述的装置,其特征是:规定配置一台有荧光屏的(BS)输出装置(AE)和/或一部按点阵打印原理工作的打印机。

汉字输入终端的方法与装置

本发明涉及汉字输入终端方法，是用拉丁字母拼音将汉字构成的词输入，这里一个音节对应一个汉字。此外，本发明还规定了实施这一方法之装置。

众所周知，汉字可以通过键盘用拉丁字母拼音输入。这里可以使用 410 个音节，以拼音方式对应 6764 个汉字。其中许多汉字读法相同（同音字），而各个音节与汉字的对应也往往是多义的。

此外大家还知道，这 6764 个汉字可以用代码表示，代码用四位数字组成，并且是单一地对应一个汉字。因此，譬如要传输中文可以输入和传输代码。但是，这里要求输入者给出每个汉字所属的代码只限于国标码表上所能查得到的。

因此，本发明之基本任务是说明汉字输入之方法与装置。用本发明时，在用拼音输入词以后，自动给出所属之汉字。本发明的目的在于给出一个中文字符的输入方法及其设备，在其使用中，在中文字符借助于拼音符号输入以后，其所属的中文字符及其相应的代码能自动地被判别。

按发明开始所提的这种方法，其任务的解决如下：与用拼音所输入词的每个音节都对应有一个音节标识码（SK），在词存贮器（WS）中对应每组音节存贮一组音节标识码，对应音节标识码有一组序号（OZ），这组序号在与同一音节标识码（SK）对应的许多汉字中指出相应的汉字，根据音节标识码（SK）及所属之序号（OZ）得出与相应字对应的汉字（CHZ）。

本发明出于以下认识，即与单音节对应之汉字往往是多义的，与此相反，与多字对应的多音节组大多数情况下可以单一地对应一组汉字。这里，对应拼音中的每个音节有个音节标识码，并且在词存贮器

中通过音节标识码组存贮很多汉字。此外，对每个音节标识码都存贮一个配属的序号，该序号表明这指的是用同一音节标识码标志的许多汉字中的第几个字。譬如，在词存贮器中存贮30000字的音节标识码及所属之序号。

词存贮器也是按其目的安排的，即词存贮器对出现的每个大于1的音节数都有一个自己的存贮区。用这种方法，根据音节标识码及所属序号可以加速查找的进程。

音节组是有目的地严格按照音节标识码排序的，而并且是如此排列的，即首先按第一音节的音节标识码，然后按第二音节标识码，第三音节，等等。

如果在相应音节标识码范围内通过抽项查找某字，便可尤其加快进行，如果有输入，那么每抽一个中间项都要查明在其前或其后范围内是否包括有这个输入。

如果几个多音节组不能单一地对应一组汉字，那么，可以首先在输出装置显示与输入音节组相应的第一组汉字，并且通过按第一或第二键认定这一音节组正确或显示下一音节组。

如果在较长的词中作为汉字组合有一个或多个音节，那么也可以通过附加输入较长的词来标识这些音节，并且由此来决定，倘使词存贮器中存有这些较长的词。

为了用简单方法也能输入成语或常用固定词组，若把多个词组成的成语如同多音节词一样处理，这是有益的，即在各个字之间的空隙位置输入一个分隔符，并从而使之表明，词组如同多音节字一样可在词存贮器中查找。

词存贮器的优点是构成上可以扩展，所以经键盘输入后，单个汉字可以组合在一起，按动相应的插入命令键插入完毕以后，词存贮器

中又自动重新排序。

词存贮器能够适合用户的词汇也是一大优点。如果要删除词存贮器中的字，可以把字用键盘输入，并在按动删除命令键以后字即删除，同时，在删除后词存贮器又自动重新排序。

当词存贮器中没有某个字时，也便于输入，如果输出装置显示一组汉字，字组较短，相应于所查找字的一部分，并且用最常用的音节表示其余音节，，并且给出一个信号，请用户确认这一组，或者要求逐段地显示其它字组。

本发明方法的实施及实施该方法的装置，下面用图进一步说明。所列各图的内容如下：

图 1：实施本方法的装置方框图，

图 2：取自音节与音节标识码、序号、代码及汉字相对应的对照表，

图 3：二音节词与音节标识码、序号、代码及汉字的对照表，

图 4：三音节词与音节标识码、序号、代码及汉字之对照表，

图 5：列出了对应“ba”音节的多个汉字，

图 6：列出了对应“ba”音节的两个汉字，

图 7：列出了对应带有角符“ge”和“ti”的“ba”音节的汉字，

图 8：列出了对应“ba”、“bao”和“jiao”音节的汉字，

图 9：列出的是汉语成语“哀兵必胜”。

在图 1 中所列的装置中，汉字通过拉丁字母国际键盘（TA）用拼音输入。这套装置具有一组存贮器，包括一个 6764 个编码汉字及其发音属性的符号发生器（ZO）。此外，存贮器还包括一个词存贮器（WS），譬如，其内存贮 30000 字的代码。另外，该装置还包括一

个处理器 (VE) 和一个输出装置, 这可以用荧光屏 (BS) 显示器 (AE) 和/或打印机 (DR), 特别是点阵打印机。处理器 (VE) 将代码 (CW) 送给计算机或传输段, 以便将所输入之汉字作进一步处理。

汉字是从键盘 (TA) 用拼音输入的。音节连续书写, 就象现代字典中那样。词的结尾用输入空格或标点表示。为了能够检查内存情况, 存入的内容可以在显示器荧光屏上 (BS) 和/或在打印机上显示, 而且是以汉字、用拼音输入的音节和/或与汉字对应的国际码 (CW) 的形式显示。

一个词的字母输入以后, 字母组成各个音节。这样就可每个音节输入之后, 可以输入一个分隔符, 譬如符号“/”或“-”, 或者自动分配到各个音节中, 如专利申请 No 中所描述的一样。存贮器 (SP) 包括一个音节标识码存贮器 (SKS), 并且可能的 410 个音节的每个音节都对应有一个音节标识码。接着便开始查找词存贮器 (WS) 中用同一组音节标识码的字。在字存贮器 (WS) 中用其音节标识码存贮 30000 字。词存贮器 (WS) 中存入的内容是按如下方式排序的, 首先是相应的汉字总数, 即相应的音节总数, 然后对应的是第一个汉字拼音的字母, 然后对应的是第二个汉字拼音的字母, 再后对应的是其后汉字拼音的字母及音节标识码, 等等。

查找与已输入音节组相应之汉字组, 可用第二音节之音节标识码的对照表进行。该表的首址和尾址起第一个音节的音节总数和音节标识码的作用。即使查长字情况下, 查找时间也要不到 100 毫秒。

当词存贮器 (WS) 中与所输入的音节组相应之汉字组超过一个时, 便显示第一组汉字, 并请用户选择, 确认这一组或向他提供后边可能适合的汉字组。在二音节词时, 输入的词大约有 2% 是多意的。三音

节或多音节词时，多义情况很少，因而查找可很快完成。

如果某个词在词存贮器（WS）中还没有时，词可以作为一组单个的字输入。这组词存贮在字存贮器（WS）未分类的一个小区内，可以存到词存贮器（WS）的下一实现性过程时再进行正确入序。用这种方法的词存贮器（WS）可适应用户的词汇。

存贮器（WS）中只存贮音节标识码及所属之序号。因此每个汉字都需要二个字节。平均由3个字符组成的30,000个字时，要求存贮容量为180 K字节。词存贮器（WS）为快速查找而做如下安排，即对出现的每个大于1的音节数规定一个自己的存贮区。存入的各个音节组是严格地按照音节标识码排序的，而且首先按第一个音节的音节标识码，而后按第二个音节的音节标识码，第三个音节等等。

查找词存贮器（WS）中的一个字可以通过抽项在相应音节标识码的范围内进行。此时，取出项在查找范围的中央，每试一项都要查清，如果有存入的话，它是否包括在中间项之前或之后的范围内，从而通过每个取出项便可使存在的查找范围减半，所以，用很少几个抽项便可找到。

音节标识码和序号与代码（CW）的对应由存贮器（SP）中包括的代码存贮器（CWS）来完成。代码（CW）也送给符号发生器（ZG），以便在输出装置显示汉字。

当用键盘拉丁字母输入汉字时，处理器（VE）可以自动将词分成各个音节。也可以用分隔符将字的音节分开，譬如输入“/”或“—”等符号。处理器（VE）用音节标识码存贮器（SKS）可以求得当时音节所属之音节标识码。为了能够输入多字组成的常用词组或成语，这些语组就可当作多音节词一样来处理，各个字之间的空位代之以其它符号，譬如符号“—”，用这种方法便可使词组如同多音节

词一样在词存贮器中来查找。

下表给出410个音节和相应的音节标识码之间的对照表。

对 照 表

	0	1	2	3	4	5	6	7	8	9
0		a	ai	an	ang	ao	ba	bai	ban	bang
10	bao	bei	ben	beng	bi	bian	biao	bie	bin	bing
20	bo	bu	ca	cai	can	cang	cao	ce	cen	ceng
:	:	:	:	:	:	:	:	:	:	:
400	zhui	zhun	zhuo	zi	zaig	zon	zu	zuan	zui	zun
410	zuo									

譬如，当输入“ba bao cai”这个词时，那么就把它分成“ba”、“bao”和“cai”几个音节，与其相对应的标识码则为6、10和23。

处理器（VE）只是在用词存贮器（WS）时，根据音节标识码组找出序号，序号说明用同一音节标识码表示的第几个汉字是与这个字相对应的。继而处理器（VE）用代码存贮器（CWS）得出与这些汉字对应的代码（CW），譬如为作下一步处理，这些代码可以输入一个计算机或传输给另一个终端。

图2取自词存贮器（WS）中存贮的代码（CW）和音节与汉字相对应的对照表。在整个对照表中共有6764个汉字对应410个音节和相应之代码（CW）。

图3取自在词存贮器（WS）中存贮的用拼音输入的二音节词及相

应之汉字的对照表。例如，当输入“cai”这个词时，处理器用音节标识码存贮器（SKS）求得与音节“a”和“ci”对应的音节标识码1和49。用存贮所有其它二音节词的词存贮器（WS）找出与那个音节对应的第几个汉字是在输入词中出现的。

序号（OZ）往往从0开始，如上所列，词存贮器（WS）中得出音节“a”的序号为1，序号1就说明这指的是音节“a”的第二个字，其代码为1602。图3同样列出了所属的汉字（CHZ）。同样，词存贮器（WS）通过序号（OZ）9说明，这指的是音节标识码为49的音节“ci”的第十个字，其代码（CW）为2044。同样，可以得出输入词“adou”，“afi”及“aiai”的音节标识码（SK）、序号（OZ）和代码（CW）以及汉字（CHZ）。以“aiai”这个字来说，在表中可以看出这是音节标识码为2的音节。从词存贮器中找出，序号为5的字，即列出的“ai”音节的第六个字，代码为1608并给出相应的汉字。

图4中给出词存贮器（WS）中存贮的几个三音节词的音节标识码（SK）、序号（OZ）、代码（CW）和汉字（CHZ）。譬如，如果输入“aihaozhe”这个字，处理器找出音节标识码（SK）2，115，及388，用词存贮器（WS）可找出对应这个字的序号（OZ）11，5及0，并且从代码存贮器（CWS）中找出代码1614，2635及5358。然后根据这些代码（CW）从符号发生器（ZG）便可在输出装置显示汉字（CHZ）。

如果在词存贮器（WS）中有几个项目相应于输入的音节标识码（SK），那么首先在输出装置显示第一组汉字，并请用户确认或者继续往下翻页。

如果只输入单音节，那便象图2所示，与单音节相对应的汉字往往

可能有多。图2中对音节“a”和“ai”只列出最常用的汉字(CHZ)。作为代码CW'还可以给出其它许多同样与音节“a”或“ai”相对应的代码,但这是较罕见的。根据拼音表示的音节得出所想要的汉字,有多种可能性。譬如在图5中所列,假设与音节“ba”相对应的某个汉字对输入一个字是需要的,若只输入拼音“ba”,那么在输出装置,譬如在显示器(AE)的荧光屏(BS)上首先显示代码(CW)为1649的汉字,这是与音节“ba”相对应的最常用的汉字。但由于图5中上方所列的是所希望的,所以显示的汉字是不可取的,而且用户按一下“继续翻页”的键,直到所需的代码(CW)为1648的汉字出现为止,这是个较罕见的汉字。这也通过按“确认”键接受。图5中汉字并排列出,但在荧光屏(BS)上是短时间先后逐个出现的。

图6所列表示能够很快找到所需与音节“ba”相对应之汉字的另一可能性。除“ba”这个音节外,还标上了数字3,说明这个音节读第三声,并且供用户选择的就仅仅还有代码(CW)为1649和1648的两个汉字了。

图7列出的是快速找到所需字的又一种方法。除“ba”音节外,用“井”字分开,角符“ge”在符号中出现。在这种情况下,立即显示代码(CW)为1648的所需之字。如果需要音节“ba”的字,上面要标角符“ti”时,那么这个角符用“井”字隔开输入,并在荧光屏上短时间先后显示代码(CW_s)为1638至1649的四个字。

如果要查找一个单音节或多音节词时,作为较长词中含有的字组,还有另一种方法,即通过附加输入已知的较长的词,并用分隔符,例如“<”。图8所列,要查找“ba”音节,在“baxin”这个词中含有,则请用户选择代码为1648和1649的汉字。若查找在

“bachang”这个词中也有的“ba”音节，则供用户选择的只有代码(CW)为1648的音节。同样，例如可以找出“babao”中出现的“bao”字，或得出“bajiaoshan.”这个词中含有的“jiao”字。

词存贮器(WS)可以适应用户之需要，即可以插入另外的字或删除不必要的字。用键盘输入这些字可通过与输出装置对话的方式进行，将单个(符号)字组合一起，接着给一插入命令，插入后字存贮器中便自动重新排序。

删除词存贮器(WS)中的字，通过以下方法进行：字同样用键盘(TA)输入，并在删除命令之后字就删除，删除之后词存贮器(WS)中便又自动重新排序。

如果一个字在词存贮器(WS)中未查找到，则在输出装置向用户作出推荐，象一个字(符)组，是由几个较短字(符)组组成，相应于所查找字的一部份，并且用最常用的同音词代替其它音节。接着请用户选定字组或逐段地继续翻页。

图9所示的是一多字组成的汉语成语“aibingbisheng”。同样，这可以象多音节中文词一样来处理。在字与字之间的空格输入分隔符，譬如“—”来替代，并用词存贮器(WS)找出各字所对应之音节，然后用代码存贮器(CWS)如同在各个字一样，将代码与音节对应起来。

图 1

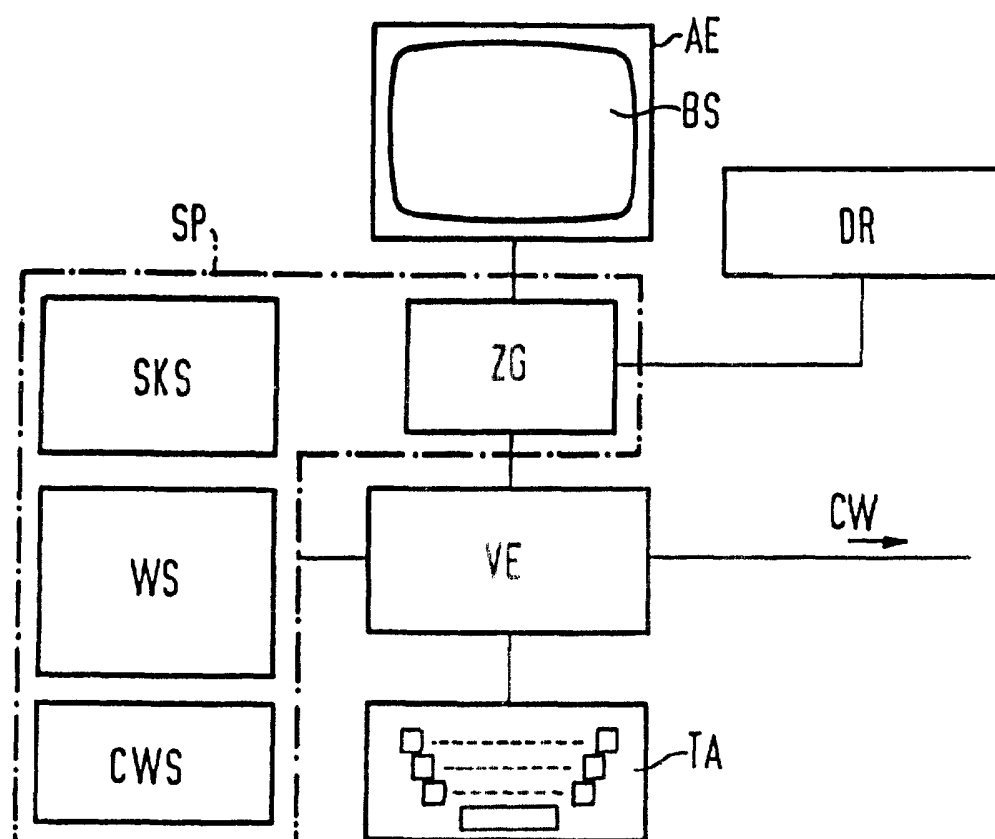


图 2

	a		ai							
SK	1		2							
OZ	0	1	0	1	2	3	4	5	6	
CW	1601	1602	1603	1604	1605	1606	1607	1608	1609	
CHZ	啊	阿	埃	挨	哎	唉	哀	皑	癌	
CW'	2639		6263							
	6436		7040							

图 3

Wort	a - ci		a - dou		a - fei		ai - ai		---
SK	1	49	1	73	1	86	2	2	---
OZ	1	9	1	2	1	3	5	5	---
CW	1602	2044	1602	2223	1602	2341	1608	1608	---
CHZ	阿刺		阿斗		阿飞		皑皑		---

图 4

Wort	ai - hao - zhe			ai - mao - zi			an - an - shi			---
SK	2	115	388	2	195	403	3	3	298	---
OZ	11	5	0	11	2	11	5	2	13	---
CW	1614	2635	5358	1614	3510	5551	1621	1618	4219	---
CHZ	爱好遮			爱锚子			啖安食			---

ba
靶

图 5

ba 1649	1653	1654	1638	1639	1640	1641	1643	1644
把	罢	爸	捌	扒	叭	吧	八	疤
1645	1646	1648						
巴	拔	靶						

ba3
1649 1648
把 靶

图 6

ba#ge 1648	ba#li 1638	1639	1646	1649
靶	捌	扒	拔	把

图 7

ba<baxin<baxin<bachang				bao<babao	jiao<bajiao	shan
1648	1649	1648	1706	2922		
靶	把	靶	宝	蕉		

图 8

aibing_bi_sheng			
1607	1788	1756	4204
哀	兵	必	胜

图 9