2005年，随着戴石麟大师的顶功输入方案开山之作“声笔码”及“声笔飞码”横空出世，输入法界的新时代悄然来临。越来越多的人从质疑到了解再到恍然大悟、为己所用，一场轰轰烈烈的革命随之爆发。

十余年后的今天，顶功已成为输入法界无人不知的名词，多种多样的顶功输入方案或带有顶功元素的输入方案被制作出并公开，给输入法界增添了一份别样的色彩。

下面，某渣渣将带领大家对这个新生的世界作惊鸿的一瞥，然而即使只有一瞥，也足以领略到它的神奇。

食用方式：黄色高亮是重要的内容；蓝色是不太重要、具有专题讨论性质的内容，第一遍可以先不看。

-------------------------------------我是萌萌哒分割线-------------------------------------------

1什么是顶功

1.1 基于码表的输入方案

在了解什么是顶功之前，让我们回顾一下基于码表的输入方案的特点。

Def 1我们称一个输入方案是码表方案，当且仅当它建立了一个编码串到字符的映射（即码表），并且当用户输入由码表中一定数量的编码串顺次连接构成的按键序列时，这个序列不可能用另一种方法切分成码表中的编码串。

举例：

1.全拼输入方案

设字符A对应的编码是bopo，字符B对应的编码是pomo，C是mo，D是bo，当输入bopomo的时候，根据码表，并不能确定是切分成bo和pomo还是bopo和mo，因此不是码表方案。（在以后的讨论中，我们用'代表切分标记。）

2.五笔字型输入方案（以下简称五笔）

五笔接受的按键消息主要有两类：第一类是a-y，第二类是“空格”（在以后的讨论中，记空格为\_）、2、3、4…等等。我们以a为第一类的代表，\_为第二类的代表，那么很显然，五笔的码表里存在4类不同编码串：

a\_（一简）

aa\_（二简字）

aaa\_（三简字）

aaaa（全码）

Def 2我们称所有按键的总和为“键集”，形如a-y的按键类别为“键组”。

Def 3我们称形如a\_这样的，表示键组顺次连接形成的编码串的类别为“键系”，键系的总和为“键群”。

下面我们试图说明：对于任意由4个键系构成的按键序列，五笔不可能用另一种方法切分成这些键系。

证明：

1若按键序列的前两个按键是一简（a\_），它只能是一简的编码而构不成任何其他编码或编码的前缀，故只能从a\_'这里切分。

2若前两个按键不是a\_，则是aa。

2.1若按键序列的前三个按键是二简（aa\_），同理只能从aa\_'这里切分。

2.2若前三个按键不是aa\_，则是aaa。

2.2.1若按键序列的前四个按键是三简（aaa\_），同理只能从aaa\_'这里切分。

2.2.2若前四个按键不是aaa\_，则是aaaa。它只能从aaaa'这里切分。

证毕。

由Def 1，五笔是码表方案。

通过证明的过程我们也可以发现，凡是键群中不存在两个键系，使得一个是另一个的前缀的方案，都是码表方案。

Def 4我们称一个码表方案是前缀方案，当且仅当键群中不存在两个键系，使得一个是另一个的前缀。特别的，当它的键群中所包含的键系长度最长为N且只由一种键组构成（形如aaaa），其余键系由该键组和一个另外一种键组构成（形如a\_ aa\_等等），此时我们称之为N码定长方案。

五笔是典型的四码定长方案，也是典型的前缀方案。

1.2 定义的给出

根据以上复习，我们暂且给出顶功的一个较合理且广泛的定义：

Def 5顶功方案是非前缀的码表方案。

顶功方案和前缀方案是对立的，也就是说，顶功方案的键群中至少存在两个键系，使得一个是另一个的前缀。

那么，顶功方案的键群具有这样的性质将带来什么好处呢？

简单地讲，由于某一个键系的前缀能够和它本身同时存在于键群中，提高了短码长空间的利用率，从而降低了码长，提高了输入速度。

1.3 根据定义进行顶功方案键群性质的推导

细心的读者一定注意到了，既然顶功方案一定首先是码表方案，那么我们需要有一定方法验证某一顶功方案具有码表方案的性质。刚才，我们通过验证前缀的方法确定了五笔是码表方案，现在这种方法行不通了，有没有其他办法呢？

设两个键系的形式分别为：

a1a2a3...ak

a1a2a3...aka(k+1)...an

自然地，我们想到：既然通过这一段前缀本身无法分辨它只是它本身还是属于较长的键系，那么一定是在它之后的某些编码决定了它的从属。不妨设这个编码是at，我们将其分为：

Def 6第一类顶功：k<t≤n。

Def 7第二类顶功：t>n。

为什么要作出这样的分类呢？我们知道，在码表方案中，键系划分的完成通常意味着相应字词的上屏，如果这个决定太久不作出，将推迟上屏，使得用户更容易打错。因此，第一类顶功显然更为易用。值得注意的是，现在所有的成型顶功方案都是第一类的，这也印证了分类的合理性。下面也将主要论证第一类顶功的特点。

那么，在第一类顶功中，at为什么能决定前缀的从属呢？我们发现，如果取键集的两个交集为空的子集，编码属于其中一个子集时表明前缀属于较长的键系，另一个子集时表明不属于较长的键系，就完成了切分。at被称为分流码。在具体操作中，我们将其分为：

Def 8一般的，当t-k=N时，我们称一个第一类顶功是N阶顶功。

大部分顶功都是一阶顶功。

2基础概念复习

2.1码元：为了区分字，需要不重不漏地将汉字的某个信息（如声母，字的第一笔等等）划分成多个类别，每个类别称为一个码元。如：信息“声母”可以分为【b p m f d t n l g k h j q x zh ch sh r z s c y w '】这24个码元；信息“字的第一笔”可以分为【一，丨，丿，丶，乙】这5个码元。

2.2码组：描述汉字的某个信息用到的所有可能的码元构成的集合称为一个码组，如{字的声母：24个声母}、{字的第一笔：5个笔画}都是码组。

2.3全码：按照编码规则，对字符能够编出的完整的编码串，它的每一码涉及到的码组构成的序列称为全码。如：小鹤双拼的单字全码为【声韵形形】。全码中的每一码组用到的键组构成的键系称为全码键。

2.4简码：全码的其中几个码组构成的序列。如：小鹤双拼的单字简码可能是【声】、【声韵】、【声韵形】。简码中的每一码组用到的键组构成的键系称为简码键。

3一阶顶功

假设键集为X，分流码的键组一个为Y，一个为Z，则Y∩Z＝空集，Y∪Z＝X。

当其为一阶顶功的时候，我们不妨令两个键系具有如下形式：

a1a2...ak

a1a2...aka(k+1)...an

这时，不管是哪一个编码接在a1a2...ak后，都需要a1和a(k+1)形成互斥关系来决定a1a2...ak是它本身还是属于更长的键系。不妨设a1=Y，a(k+1)=Z。

如果一组键系的前缀都是a1a2...ak，且这里面包含a1a2...ak本身，那我们认为这些键系构成了k码起顶群。

3.1连续顶功：百家争鸣

特别的，当a(k+2)、...、an包含于a(k+1)时，我们发现，如下键系亦可存在于键群中：（请读者自行证明）

a1a2...aka(k+1)

a1a2...aka(k+1)a(k+2)

...

a1a2...ak...a(n-1)

Def 9我们称一个一阶顶功是连续顶功，当且仅当它的键群是一个k码起顶群，并且对于其中任意一个键系a1a2...aka(k+1)...an，满足a(k+2)、...、an包含于a(k+1)。我们亦可称之为k码顶。

我们注意到，在k码顶的键群中，只对a1，a(k+1)~an作了要求，而对a2~ak并无要求。出于编码空间最大化的希求，a2~ak一般视为X。因此，键群中键系的通式为：

YX...XZ...Z

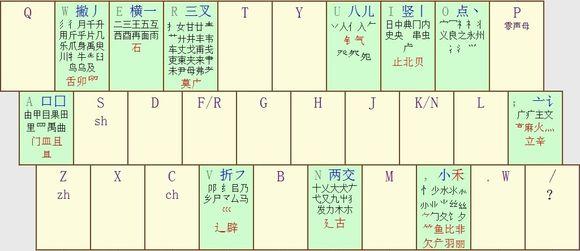
其中，有k-1个X，个数无限制的Z。

3.1.1一码顶类：左飞1811

一码顶键群为【Y YZ YZZ...】。

3.1.1.1左飞1811

相关：贴吧主贴（<http://tieba.baidu.com/p/3512095579>）



3.1.1.1.1编码结构分析

这是一款单字顶功输入法，单字全码为【声形形形形】；Y=【18个声母键】，布局了码元组y=【声母】；Z=【11个笔形键】，布局了顺序拆分的4个笔形的同系码元组。因此，全码键和简码键构成了一码起顶群。

3.1.1.1.2编码特点分析

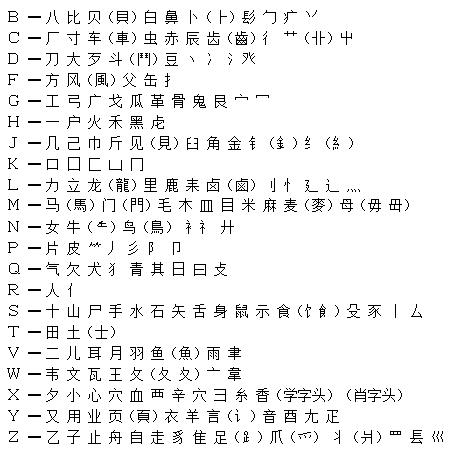
一键字空间18，二键字空间198，三键字空间2178。

众所周知，单字具有快饱和性，即常常是很少量的单字占了输入文本的很大部分。因此，这18个能一键上屏的最高频单字，198个两键上屏的高频单字对码长起到了很大的作用，将打单字的码长降低到了2.2-2.3。

3.1.2二码顶类：声笔飞码、西风瘦码、小兮码、灵形速影、Leo

二码顶键群为【YX YXZ YXZZ...】。

3.1.2.1声笔飞码



3.1.2.1.1编码结构分析

这是一款字词混顶输入法，单字全码为【声形笔笔笔……】；Y=【21个声母键】；X=【26个形码键】，布局了音托字根；Z=【5个笔画键】，布局了笔画码元组，取的分别是：字去掉第二码已取过的字根后的一二三笔。因此，全码键和简码键构成了二码起顶群。另外，有形如【Y\_】的传统一简。（今后如不作特殊说明，顶功都含有小于等于起顶码长的传统简码。）

3.1.2.1.2编码特点分析

【YX】提供了五百多个能两键自动上屏的顶屏码，而字频前500的字总频可达80%，剩余的大部分为【YXZ】三键自动上屏，因此声笔飞码单字平均码长仅为2.2-2.3。

但最长码长仍较长，高频长码字仍然较多。

值得注意的是，飞码的【X】26键，由于布局的是音托字根，aeiou键并没有利用，而飞码用这些键做了“规则简码”，即不同于主规则的、声母后直接加笔画的21\*5个第二规则编码。规则简码的使用在难度提高不大的情况下增加了不少效率，也是飞码中的一个亮点。

【补充阅读1】一码顶与二码顶的对比：

[http://tieba.baidu.com/p/3479211077#61968581826l](http://tieba.baidu.com/p/3479211077" \l "61968581826l)

这个贴子详细讨论了一码顶和二码顶的优劣。总体来说，一码顶的优势是超高频字一键上屏，二码顶的优势是二键上屏的字多，因此二者在码长上不相上下。但二码顶二简能做到70%左右，则打字的时候大多数都是二码，节奏感较强。一码顶则一简二简三简的占比差不多，在这一点上不如二码顶。

其实，如果将标点也作为字看，二码顶中也存在一码顶的成分（，。一键上屏），因此最高效的方式应当为一二码混顶，一码顶在3-4个（即除了，。外还有1-2个单字可以一键上屏）左右。

事实上，我们现在可以修改一下二码顶的表示方法，使其适用于一二码混顶：假设所有用于编码的码元键构成的集合为X，X的三个子集为Y、Z、W；Y、Z、W两两没有重叠的部分。单字全码为YXZZ……或W。YX、YXZ、YXZZ……等编码可以被W顶（因为W不和Z重合）；W也可以被YX、YXZ……顶（因为W不和Y重合）。在飞码中，W包含两个键【,.】，分别编码了逗号和句号。因为W的存在，二码顶的Y+Z一般至少比X小2，即至少要让【，。】两个字符一键上屏。但在后续的讨论中，简便起见，仍不提及W。【1结束】

【补充阅读2】顶功结构之间的复合

Def 10若一个顶功的键群能写成几个起顶群的并集，那么我们称之为复合顶功。比如【补充阅读1】中提到的一二码混顶就是一个复合顶功。但为方便起见，我们仍将其按照最主要的起顶群命名。

飞码的二字词全码是【声(笔2)声笔…】。“笔2”是另一套五笔画，布局在,./;'这5键上。那么，显然这样的编码与单字并不冲突（自行证明），并且是一个三码起顶群。从这个意义上看，飞码是二码顶和三码顶的复合。

三字词及更长词的编码是四码开顶系，实现方式类似，在此不作赘述。

由此可见，简单的顶功结合成复合顶功大大提高了顶功的多样性，可玩性。【2结束】

3.1.2.2西风瘦码（改编自飞码）

相关：贴吧（<http://tieba.baidu.com/f?kw=%E8%A5%BF%E9%A3%8E%E7%98%A6%E7%A0%81&ie=utf-8）；QQ（15571505>）；论坛（<http://shouma.six168.com/>）

西风瘦码是化外愚民（面壁）在飞码基础上改进、增强而得的一个效率更高的单字顶功。

3.1.2.2.1编码结构分析

这是一款单字二码顶，单字全码为【声形笔笔笔】；Y=【21个声母键】；X=【21个形码键+9个规则简码码元】，布局了音托字根；Z=【5个笔画键+2个规则简码码元】即【eioua;/】，布局了笔画码元组，取的分别是：字去掉第二码已取过的字根后的一二三笔。

3.1.2.2.2编码特点分析



相比飞码，改动主要有两点：

1. 取【X】上的字根时，规定被绑根、弱根让位，增强了字根的分辨能力，因此将最长编码压缩到了5码。
2. 使用了更多的规则简码（总数800字左右，YX上的规则简码为声母+笔画，其中eiuoa用于非左右结构，,;/.a用于左右结构；YXZ上的规则简码为声母+字根+结构，;为左右结构，/为非左右结构。），使得高频长码字大量减少。

得益于以上两点，西风瘦码的码长进一步降低（2.1左右），红楼梦全书2.09的码长，立于高山之上，俯视群马。

【补充阅读3】从单字角度看顶功与四码定长的比较

四码定长类方案一简打两键，二简打三键，三简和全码打四键……平均码长在3.0附近。

然而，单字的信息熵约为9bit/字，一个26键的输入法一个键是4.7bit(2^4.7=26)，那么打单字的码长似乎本该在2以下！多出来的码长，显然是由于一二三简要空格才能上屏；空格没承载任何编码信息，所以码长才长。顶功单字，就是省略了这一部分码长而达到提效的目标。【3结束】

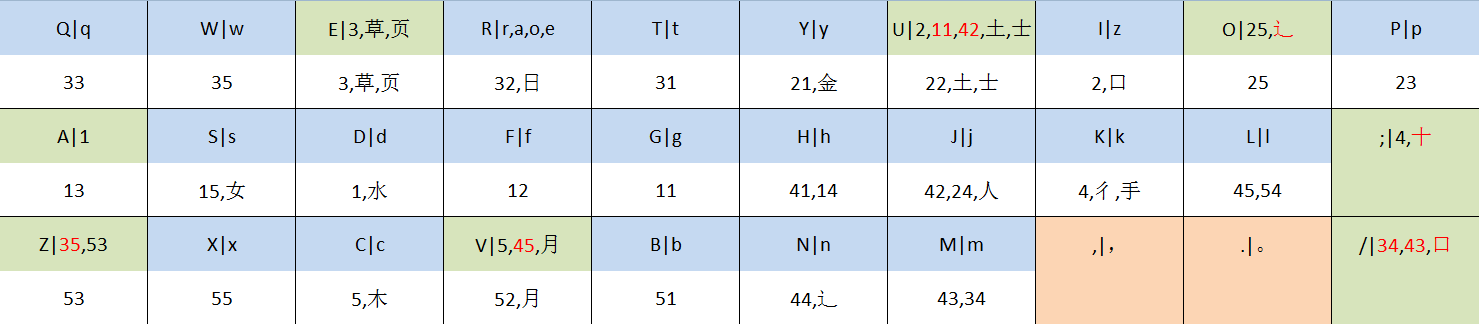
3.1.2.3小兮码

相关：贴吧主贴（http://tieba.baidu.com/p/3485999115）；QQ群（557085982）。

3.1.2.3.1编码结构分析

这是一款单字二码顶，单字全码为【声(形A)(形B)(形B)(形B)】；Y=【20个声母键（蓝色）】；X=【26个形A键（白色）+4个规则简码码元】，布局了哲哲豆的二笔键盘；Z=【8个形B键（绿色）】，布局了经合并的二笔键盘。

3.1.2.3.2编码特点分析



【补充阅读4】分流码与键盘的设计

一码顶，YZ两个键组，那么一个布局音，一个布局形是自然的选择。

二码顶，YXZ三个键组，似乎Y布局声，X布局韵，Z布局形是最自然的。然而，前两码是声韵的离散性能稍差，加之二码顶编码空间小，故难以做好。声韵不行，那声形呢？Y布局声，X布局形，Z上如果也布局形，则由于Z远小于X，必定会有很多字根在Y到Z的过程中发生移位，导致思维不顺畅。所以，飞码和瘦码都选择了Z直接用笔画，笔画的离散并不好，因此瘦码用二、三级规则简码字来增加Z的离散能力。而对于兮码来说，本身就是二笔，用来完整地描述【形】用到的码元数量较小且灵活，因此在Z的设计上有更大的发挥余地，用尽可能少的字根移位将Z扩大，增强离散。【4结束】

兮码的指导思想是“规则简单，高频无理，低频规范”，在前500字中使用了200个无理码，其他都是正常编码。得益于更大的Z，小兮没有使用二三级规则简码就将编码控制在5码内；当然，这样做导致二三键字覆盖率（96.5%）低于瘦码等（98%以上）。码长与瘦码持平。

3.1.2.4灵形速影（改编自092五笔）

相关：贴吧（<http://tieba.baidu.com/f?kw=%C1%E9%D0%CE%CB%D9%D3%B0&fr=ala0&tpl=5>）；QQ群（172465411）。

3.1.2.4.1编码结构分析

这是一款单字二码顶，单字全码为【(形A)(形B)笔笔笔】；Y=【21个形A键】即【abcdfghjklmnpqrstvwxy】；X=【26个形B键+5个规则简码码元】；Z=【6个笔画键+2个规则简码码元】即【zeiou/;’】（z和/都代表折，是飞键关系），布局笔画码元组，取的分别是：字去掉前二码已取过的字根后的一二三笔。

3.1.2.4.2编码特点分析

【补充阅读5】再谈键盘的设计

二码顶，YXZ三个键组，Y和X能不能都布局形呢？YX既然都布局形，那么这两个形码键盘应该是大体相近的，否则思维上说不过去。但相近就会产生问题：X常常比Y大，那么X比Y多出来的那些键布局什么好呢？具体到灵形速影上，X中含5个规则简码键，那么剩下的26个键都要布局形码，如何处理好这26键和Y的21键的不一致性？灵形速影给出了一个富有创意的尝试。



在这张字根图上，我们看到了两组笔画：一组在某个键的上部，位于HSTDN；一组在某个键的中部，位于EIUOZ。当首码要取的字根是单笔画的时候，打的是HSTDN（都在Y键盘上）。当次码要取的字根是单笔画的时候，如果首码在右手，则打ESTDZ；如果首码在左手，则打HIUON。这样一来，将浪费的空间转化为了异手率的提升。【5结束】

在瘦码的基础上，灵形速影使用了较多的字根（400+），提高了离散的性能，能在五码内容纳全GB+少量GBK的7000余字（瘦码只做了5000余字），并减少了规则简码及其他不规范编码（瘦码900+，灵形600+）。

3.1.2.5Leo乱序全形二顶

3.1.2.5.1编码结构分析

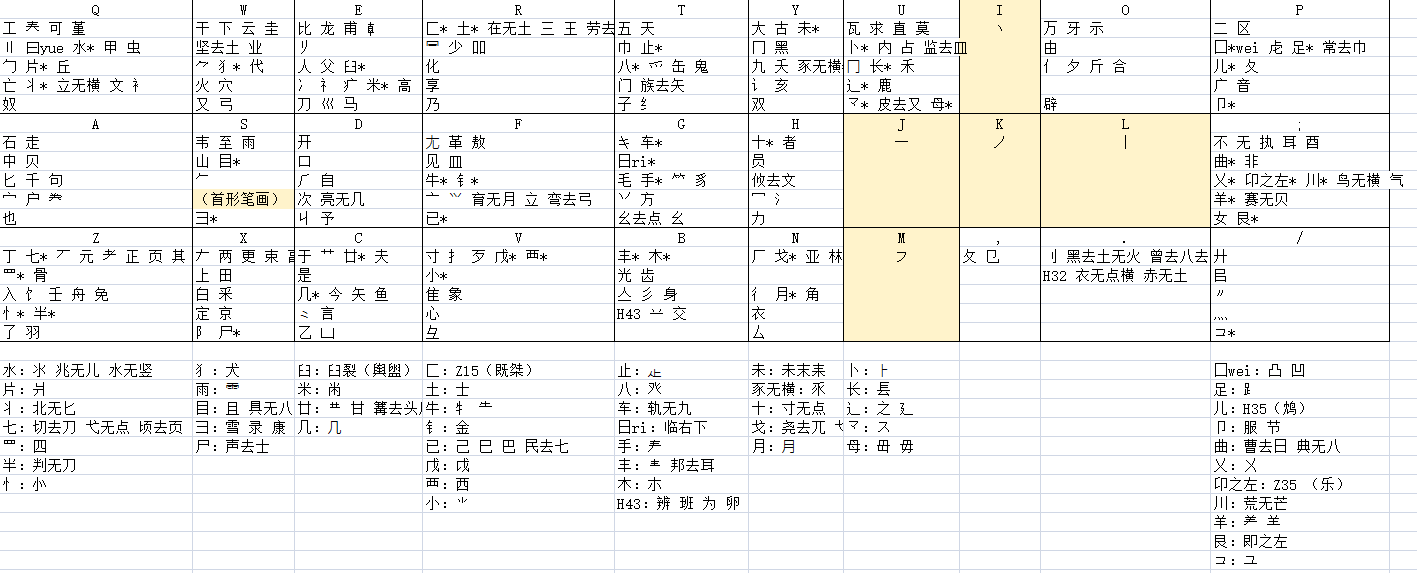
这是一款单字二码顶，单字全码为【(形A)(形B)笔笔】；Y=【22个形A键】；X=【30个形B键】； W=【,./】（即“，。的”都收为一码顶字符，的在/），Z=【5个笔画键】；布局了两个笔画码元组。全码为【YXZZ】。

另有形如【YX\_】的编码，可以看作“用空格选二简二重”，显然它不和其他编码冲突。也就是说，顶功在起顶码长及之后的编码时空格可以挪作他用，充分利用这一好按的键位，降低码长。

3.1.2.5.3编码特点分析

Leo使用了较多的字根（360个）且全乱序，使得离散性能更高，只用了四码，即Leo是第一个实现四码容纳所有字的二码顶（尽管只编了5000余字）。此外，编码较为规范，没有无理码和规则简码，但有二简二重400字和三简二重200字。码长比上面三个稍低。

Leo里面X比Y大8，这8键是怎么处理的呢？首先，jklim5个键在第二码的位置也代表单笔画，并规定如果单笔画出现在第一码，则统一归并到s（这也是合理的做法，因为第一码用到的单笔画较少）；其次，,./三个键放置的实际上是“并不在一个字的首位出现的字根”，因此也就并不用考虑重复布局的事。



3.1.3三码顶类：左飞三顶、听雨三顶、声笔码

三码顶键群为【YXX YXXZ YXXZZ...】。

3.1.3.1左飞三顶

相关：贴吧主贴（<http://tieba.baidu.com/p/4147070425>）

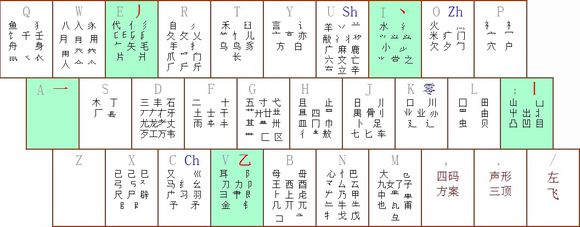
【补充阅读6】怎么又是左飞？没错，这是一位高产的输入方案作者。据不完全统计，他做过：左飞1811、小左双拼、左飞双拼顶、左飞三码顶、左飞四笔、左飞五笔……其中左飞四笔（<http://tieba.baidu.com/p/3845956487>）和左飞双拼顶（<http://tieba.baidu.com/p/3618242188>）也是顶功类方案，不过左飞四笔和声笔码结构类似，左飞双拼顶和瘦码等类似，且知名度并不高，故不再赘述。【6结束】

3.1.3.1.1编码结构分析

这是一款单字顶功输入法，单字全码为【声形形笔】；Y=【22个声母键】；X=【26个形码键】；Z=【5个笔画键】，布局了“末笔”。因此，单字简码和全码构成了三码起顶群。

3.1.3.1.2编码特点分析

三码顶相比二码顶，不顶二简，导致码长上升至2.6；但三码空间、四码空间大，可以较为容易地在四码内解决常用字；同时打字思维较顺畅（除一简外，基本上都是三键连打）；且可以没有无理码、规则简码等，记忆量小。



3.1.3.2听雨三顶

3.1.3.2.1编码结构分析

这是一款字词混顶输入法，单字全码为【声形形韵韵】；Y=【21个声母键】；X=【26个形码键】；Z=【euioz】，布局了“韵母首末码”。

词组全码为【声形声韵韵】，同时设【YXYZ】的简码。

相关：贴吧主贴（<http://tieba.baidu.com/p/4033406637>）



3.1.3.2.2编码特点分析

在三码顶单字的基础上，听雨给出了一种提高效率的思路。首先，不设二简字，除了一简，单字都是三码自动上屏，加强了思维的顺畅度；其次，在二简位放置二简词，在四、五码处放置四五码词，亦能够提高效率。

3.1.3.3声笔码

相关：贴吧主贴（http://tieba.baidu.com/p/2782533961）

3.1.3.3.1编码结构分析

这是一款字词混顶输入法，单字全码为【声笔笔笔笔笔……】；Y=【21个声母键】；Z=【5个笔画键】，布局了顺序取的多个笔画。因此，单字全、简码键构成了二码起顶群。（思考：为什么不是一码起顶群？看下面。）

另有二字词全码为【声声笔笔笔笔……】，三字及以上词全码为【声声声笔笔笔……】，它们合起来实际上构成了形如【YYXZZ…】的三码起顶群。

（显然，声笔码以词组输入为主，因此将其划归三码顶是合适的。）

声笔码是复合顶功。

3.1.3.3.2编码特点分析

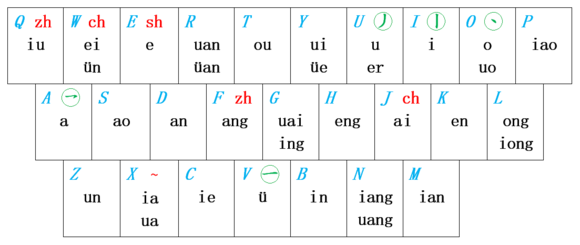
三键两字词、三键多字词的存在使得方案的词能力强，因此实际打字码长较短。加之编码只用到声母和笔画，极其简单，堪称顶屏输入法的经典之作。当然，单字能力并不强。

3.1.4四码顶类：星空键道

四码顶键群为【YXXX YXXXZ YXXXZZ...】。

3.1.4.1星空键道

相关：贴吧主贴（<http://tieba.baidu.com/p/4588150405）；QQ群（320053116>）



3.1.4.1.1编码结构分析

这是一款词组顶功输入法，词组全码为【声韵声韵笔笔】；Y=【21个声母键】（通过拆分、合并声母将声母压缩到了21键）；X=【26个韵母键】；Z=【5个笔画键】，布局了“词组首字及末字第一笔”。因此，词组全码键为【YXYXZZ】，简码键为【YXYX】、【YXYXZ】。单字全码【YXXX】不参与顶功（这可以看作另一种形式的复合，即四码定长和顶功的复合）。

3.1.4.1.3编码特点分析

这类词组顶功给出了一种解决传统四码定长的词重码（特别是对于双拼等词重码较多的方案）的解决办法：低频的重码词扩展到5、6码空间，在码长几乎不增加的情况下减少了重码。但如果要熟练使用则需要记忆词简码，较为困难。

3.2非连续顶功

一般的，当a(k+2)~an中存在a(k+l)不包含于a(k+1)时（l是大于2的正整数），那么a(k+l)不能起到分流码的作用，我们发现，相较于连续顶功，如下键系不可存在于键群中：

a1a2...aka(k+1)...a(k+l-1)（请读者自行证明）

Def 11我们称一个一阶顶功是非连续顶功，当且仅当它的键群是一个k码起顶群，且满足a(k+2)~an中存在a(k+l)不包含于a(k+1)。特别的，当l的最小值是p，次小值是q...时，我们称之为k(k+p)(k+q)...码顶。

我们注意到，在k(k+p)(k+q)...码顶的键系中，出于编码空间最大化的希求，a(k+l)既然不能起到分流的作用，那么一般视为X。因此，它的通式为：（[ZX]表示既可为Z也可为X）

YXX...XZ[ZX]...[ZX]

其中，第一个Z前有p-1个X，第一个Z后有个数无限制的Z和X。

3.2.1二四顶：希码顶功动动版

二四顶键群为【YX YXZX】

3.2.1.1编码结构分析

单字顶功输入，单字全码键为【YXZX】，简码键为【YX】。单字全、简码构成了二码起顶群。另有【YX\_】的简码。

3.2.1.2编码特点分析

相比于二码顶来说，二四顶在第四码改为使用【X】，放弃了三码字。那么这样做有什么好处呢？

显而易见的是，四码总空间大小变大了（X大于Z，故YXZX大于YXZZ），所以可以较为容易地在四码内放下大多数常用字；

另外还有隐藏的好处，只打一二简和全码，简化了单字简码的分级记忆。

能两键上屏的字是Y\*X个（其他都要四键），而Y与Z之和是X，因此应设法把Z做的足够小，来使Y\*X足够大。

在单字理论码长上，纯二四顶比不过二码顶，当然有了空格选择的二简二重就不一样了。

3.2.1.3希码顶功动动版

相关：贴吧主贴（http://tieba.baidu.com/p/4199731824）

单字全码为【形形A形】；Y=X=【26个形码键】；Z=【1个键】即【A】这一个键。因此，这个顶功实际上是放弃了第三码的离散，即是由一个三码解决能力的单字输入法改造而成的。动动本人作为打字快手，在各大跟打群跟打，也取得了不错的码长实战成绩。

但实际上，原版希码约有1000多字是四码，也就是说在这个顶功版里要变成选重码。这些记忆量是不可忽略的。

【补充阅读7】利用编码空间和待编码字数估算选重字数（部分引自现代五笔百度百科词条）

假设我们有N个编码空间，M个待编码的字，那么：

我们现在把这些字随机地扔到这些空间里。对任意一个空间来说，它某一次不被扔中的概率是(1-1/N)，因此全部扔完后仍然空着的概率是(1-1/N)^M。有N个空间，因此全部扔完之后空着的空间是N\*(1-1/N)^M个。换句话说，全部扔完之后有字的空间是N\*[1-(1-1/N)^M]个。

我们发现，我们实际上把M个字放到了N\*[1-(1-1/N)^M]个空间里，那么就要有M-N\*[1-(1-1/N)^M]个字是该空间的第二或第三或更靠后的字，它反映在输入法里就是选重字。

一般的，我们有牛顿二项式展开定理：

(a+b)^n=C(n,0)\*a^n+C(n,1)\*a^(n-1)\*b+C(n,2)\*a^(n-2)\*b^2+...+C(n,n)\*b^n

那么M-N\*[1-(1-1/N)^M]

=M-N\*[1-(1-M/N+M^2/2N^2-M^3/6N^3+M^4/24N^4...)]（由于M很大，我们近似的将M-1看做M）

=M-N\*(M/N-M^2/2N^2+M^3/6N^3-M^4/24N^4...)

=M-(M-M^2/2N+M^3/6N^2-M^4/24N^3...)

=M^2/2N+M^3/6N^2-M^4/24N^3...

=M^2/2N(1+M/3N-M^2/12N^2...)

显然，N至少是M的几倍，因此上式括号内第三项及以后的项远远小于1，可以忽略不计。更简单的，如果N是M的几十倍，第二项也可忽略不计。

我们认为，用N个编码空间，编M个待编码的字，从纯概率上讲会产生c=M^2/2N(1+M/3N)个重码字。在实际编码过程中，会有字根分部不均等多种情况阻止输入方案到达这一极限值。因此，用这个判断重码字的下限是合理的。（对于少量突破此下限的方案，如张码，我们的解释是它充分根据了字的特点，使得它的针对性高于“随机”，从而实现低重码。）【7结束】

一般的，对于27键的三码定长方案，我们利用重码的估算公式，得到GB6763有1040个需要选重的字；即使结合简码，使得有效的编码字数降低到6000字左右，也有828个重码。因此，用三码方案改的二四顶，实际上面临着重码的问题。

【补充阅读8】我们如何看待单字顶功的重码和简码

三码二四顶重码多，其他单字顶功难道就没有这样的问题吗？当然有。

一般的，k码定长键系可认为包含k个X，而全码长度是k的顶功的键系除了X外至少包含一个Y一个Z。根据多元均值不等式，我们有：

N=x^k/x^(i+j)\*y^i\*z^j=x^k/x^(i+j)\*i^i\*j^j\*(y/i)^i\*(z/j)^j≤x^k/x^(i+j)\*i^i\*j^j\*(x/(i+j))^(i+j)=N0\*i^i\*j^j/(i+j)^(i+j)

其中N是顶功的编码空间，N0是定长的编码空间，xyz分别是XYZ的大小，ij分别是YZ的数目。

显然，i^i\*j^j/(i+j)^(i+j)是个很小的数。特别的，我们令k=4，x=26，那么这个数对于一码顶是27/256，二码顶是4/27，三码顶、二四顶是1/4。利用重码的估算公式，一码顶、二码顶、三码顶的理论重码数分别是定长的9.1，6.5，3.9倍。

出于这个原因，一二三码顶大多选择了扩展到五码，降低重码；有人觉得不够，于是再使用“出简不出全”，进一步降低重码；有人觉得还不够，于是又使用“规则简码”，在三四码上放入更多字…

反观四码定长，记住一二简之后，剩下的知道有三简就打三简，不知道的直接打全码就可以了。因此，顶功获取低码长的代价也是不小的，不少非顶功单字派对这一点很发了些自己的看法。

然而，实际上问题并没有看起来的那么严重：既然顶功单字的高频字是完全确定的，而大多数人又没有完全不看候选框的需求，那么在那些不很确定的低频重码字、出简不出全字出现的时候，看一下候选框即可，这并没有影响顶功单字打常用文本时太多的确定性。顶功单字仍然是单字派很值得考虑的选择。【8结束】

3.2.2二五顶：自然光速

二五顶键群为【YX YXZXX】

3.2.2.1编码结构分析

单字顶功，单字全码键为【YXZXX】，简码键为【YX】。单字全、简码构成了二码起顶群。

3.2.1.2编码特点分析

二五顶比二四顶更进一步，在第四、五码都使用【X】，使得编码空间达到或超过了四码定长，四二顶的高重码也就随之降下来了。然而，真的有这样的好事吗？

能两键上屏的字是Y\*X个（其他都要五键），而Y与Z之和是X，因此应设法把Z做的足够小，来使Y\*X足够大。不过即使这样，码长也很长（相比四二顶，20%-25%的字多打了一键）。因此，二五顶在实际使用的时候是结合四码唯一自动上屏来用的。即：前4码常常能确定唯一一个字，如果这样，那么在4码的时候就上屏。

换句话说，为了知道前4码能不能确定一个唯一的字，还是要记住4码后有哪些重码…记忆量一点没少，顶多是选那些重码的时候手感好点罢了。

3.2.1.3自然光速

单字全码为【形形D声韵】；Y=X=【29个形码键】；Z=【1个键】即【D】这一个键。实际上，形码由自然码的30键形义辅助码改编而来。形义码+声母也是约有1000多字分辨不了，仍然变成5键字。

【补充阅读9】二五顶的另一用途，附谈二码顶加词

更常见的，二五顶用于复合在二码顶上作为词的编码。

一般的，词取字的前两码或一码组成共4码的编码，这在二码顶上显然与单字编码冲突。然而，如果在4码的中间插入一个另外的键（我们记为V），这个键既不属于二码顶第一码的Y，也不属于第三码的Z，形成如下键群：

YX

YXZ...Z

YXVXX

显然，这是可行的。这一类5键词通常用于单字打法码长太长的词，或连绵词、成语等等，降低了思维强度（但并不能降低太多码长）。

如果要在二码顶的基础上用词降低码长，那么需要使用码长更短的词。飞码的“三码起顶”词，就是很好的例子。小兮、瘦码采用第一码+打词键或声母+声母打高频两键词；灵形有仿092的一简、二简十重词；leo则将100个词和字统一规则编码，主要编入二简。【9结束】

3.2.3并击的天然顶功特性——及一例：TK

3.2.3.1编码结构分析

这是一款字词混顶输入法，单字全码为【形形音形】，词组全码为【形形形形】。为什么能顶呢？请看下面。

定义：单手不带空格击出的码元键集为a，单手带空格击出的码元键集为A；双手不带空格击出的码元键集为b（即击一次出两个b），双手带空格击出的码元键集为B（击一次出两个B）。即：并击通过单双手、是否带空格将码元四分。

一简字：a（一次单手不带空格并击）；

二简字：bb（一次双手不带空格并击）；

三简字：bbA（一次双手不带空格并击加一次单手带空格并击）；

全码字：bbAA（一次双手不带空格并击加两次单手带空格并击）；

全码词：bbBB（一次双手不带空格并击加一次双手带空格并击）。

3.2.3.2编码特点分析

在并击中，由于单手一击的组合数目远大于普通串击输入法的键的个数，因此码元空间可以非常的大。tk的a、A都包含了120个码元键，b、B也包含了60个码元键，这些键提供了120个一简及3600个二简的广大空间，使得一二简总频达到了惊人的93%，即只打单字击长在1.07以下。此外，在二简空位填的一击词及一简多重的词使得平均击长降的更低。

3.2.3.3 TK T－tick 如时间般独立；K－kick 自由需要努力。

相关：贴吧（http://tieba.baidu.com/f?kw=tk%CA%E4%C8%EB%B7%A8&fr=ala0&tpl=5）

4二阶及更高阶顶功

假设键集为X，分流码的键组一个为Y，一个为Z，则Y∩Z＝空集，Y∪Z＝X。

当其为二阶顶功的时候，我们不妨令两个键系具有如下形式：

a1a2...ak

a1a2...aka(k+1)...an

这时，不管是哪一个编码接在a1a2...ak后，都需要a2和a(k+2)形成互斥关系，不妨设a2=Y，a(k+2)=Z。

一般的，对于t阶顶功，我们令at=Y，a(k+t)=Z。它的键群是一个k码起顶群。

Def 12我们称键群是k码起顶群的t（t>1）阶顶功为(k+t)k码顶。

我们注意到，在(k+t)k码顶的键群中，出于编码空间最大化的希求，除at/a(k+t)外既然不能起到分流的作用，那么一般视为X。因此，它的通式为：

X...XYX...XZ...

其中，Y前有k-1个X，YZ间有t-1个X。

4.1二阶顶功

通式为X...XYXZ...

4.1.1四二顶：C输入法顶功版

四二顶键群为【XY XYXZ】。

4.1.1.1编码结构分析

单字顶功输入，单字全码键为【XYXZ】，简码键为【XY】。

4.1.1.2编码特点分析

基本同“二四顶”。但注意，由于延迟上屏（打完一个字还要再按两键才能上屏）的情况比其他顶功严重，所以更易出现回改等情况。

【补充阅读10】顶功的等价

我们发现，二四顶和四二顶有着密切的联系。这种联系背后是怎样的内在原因？

Def 13我们称两个顶功方案是等价的，当且仅当它们键群中的所有键系可以建立一个一一对应关系，且对应法则只能是交换键系中键组的顺序。

按照这个定义：

四二顶XYXZ——二四顶YXZX

四二顶XY——二四顶YX

它们是等价的。

由于键系产生了一一对应关系，那么显然这两个方案的重码、码长等等指标性能应该完全一样。

有人可能要问了：既然这样，那为什么还要分别研究两个顶功？

因为这个编码，首先是需要人去打出来的。一个键系里的键组换了换顺序，人的取码思维可能就会发生改变，上屏的机制也可能发生改变。比如，希码顶功中间那个A的插入是否有人觉得不太自然？如果使用某种变换使得三个形码在一块，是否更为易用？再比如，四二顶的延迟上屏是否有人觉得容易打错？如果使用某种变换使得它变成一阶顶功，是否更为易用？

一般的，对于任意一个高阶顶功，一定存在与它等价的一阶顶功，反之亦然。【10结束】

4.1.1.3 C输入法顶功版

单字全码为【形形形空】；Y=X=【27个形码键】；Z=【1个键】即【空格】这一个键。其余指标与希码顶功差不多，但C输入法原本四码字较少（700余字），故做成顶的重码也比较少，但仍显稍多。

4.1.2 四二-二码复合顶（概念顶功，无作品）

Def 14有的时候，两种顶功不能直接复合，但是，当我们将某些键系的键组取为它的子集时，却可以复合，我们称这种顶功为部分复合顶功。

4.1.2.1编码结构分析

我们有二码顶：YX YXZ...Z

我们有四二顶：XY XYXZ

让二码顶取子集X→Y，四二顶X→Y，则变为：

变二码顶：YY YYZ...Z

变四二顶：YY YYYZ

那么我们发现，它们可以被复合了。

4.1.2.2编码特点分析

拟从四二顶、二码顶两个角度分析特点：

从四二顶出发：放弃了X比Y多出来的部分（实际上，有些四二顶XY是一样的，在这种情况下根本没有损失），换来了YYZ...Z的三码起顶空间，弥补了四二顶无三键字的不足。

从二码顶出发：放弃了YZZ...Z的空间（实际上，像瘦码、灵形等等这些空间里几乎没有字，所以也没有损失），换来了YYYZ的四码大空间，降低单字重码。

总之，这样的复合去除了冗余，弥补了缺点，不失为一大创举。

4.2三阶顶功

通式为X...XYXXZ...

4.2.1五二顶（概念顶功，无作品）

五二顶键群为【XY XYXXZ】

同二五顶等价，它本身并没有什么可说之处，主要用于复合。它可以直接复合于二码顶：

YX

YXZ...Z

YXYXV

复合后也等价。不同的是，YXYXV比YXVYX更加连贯，但延迟的也更长。

4.2.2五二-四二复合顶（概念顶功，无作品）

4.2.2.1编码结构分析

我们有四二顶：XY XYXZ

我们有五二顶：XY XYXXZ

让四二顶取子集X→Y，五二顶X→Y，则变为：

变四二顶：YY YYYZ

变五二顶：YY YYYYZ

那么我们发现，它们可以被复合了。

4.2.2.2编码特点分析

拟从四二顶、五二顶两个角度分析特点：

从四二顶出发：放弃了X比Y多出来的部分，换来了YYYYZ的五码大空间，降低单字重码。

从五二顶出发：放弃了X比Y多出来的部分，换来了YYYZ的四码空间，弥补无四键字的不足。

这样，比单纯的“二五顶+四码自动上屏”少了不唯一四码首项的空格。

5总结

这些五花八门的顶功类方案令我们大开眼界。

总的来说，单字派想要低码长和易学，有三码顶单字；想要极低码长，能接受难度和不确定性，有二码顶（包括四二，二四等等一系列二码起顶）单字；

字词派想要低码长，有三码字词混顶；想要不改变整体习惯的情况下解决词重码，可以用四码词组顶；

无论什么派，如果想玩，有一码顶；

无论什么派，如果极其想玩，请玩并击顶功。

-------------------------------------我也是萌萌哒分割线----------------------------------------

此文供抛砖引玉，旨在激发更多人参与顶功类输入法设计的热情与想象，愿在大家的尝试与折腾中产生出更好的方案！

那么问题来了，有了点子，到哪里去讨论呢？

请看下面——

6一个安利

札记学园 QQ（417742289）

此群定位，大神云集、群英荟萃的高端输入法编码理论、编程开发交流群，汇集各种形码、顶功、二笔、双拼等输入法资深现役码老。最前沿的理论成果，最高端的开发工具，尽在此群，请理性交流、兼容并蓄，惜字如金、相互促进。

何谓札记？

一为汉字之载体，传承之器具。

二为灵感之流溢，思辩之源起。

三为五形之印记，编码之形骸。

何谓学园？

无人不为师，无人不成生，师生聚之而成园。

每天有水吹，日日有牛聊，火热程度堪比跟打群。

如果你对输入法有新的感悟和启发而苦于无人交流，如果你正在烦恼一大堆输入法群都死气沉沉，如果你想学习先进的编码理论和开发经验，就请加入本群！

想研究输入法？加这一个群就够了~

群内现有大神一览：



7 另一个安利

讨论也讨论完了，该动手了吧，我们应该在什么样的平台上实现顶功方案呢？

这里推荐Rime平台，因为它所支持的顶功形式最广泛。

Rime的table\_translator可以实现顶功，具体方法是auto\_select设为true即可。（不需要像多多、小小一样具体设置顶屏码长）凡是满足以下条件的顶功都可以在Rime上使用：

1.打完分流码且向短码分流时，根据顶功的定义，此时一定会出现切分，那么Rime只能切到最长的编码。

2.一次分流只能导致一次切分。

注意：当用到空格编码的时候，需要在speller声明use\_space=true，在码表里凡是遇到空格都用一个其它字符代替（推荐用\_），然后在speller的algebra里定义

- xform/\_/ /

至少在这篇文档里的所有顶功Rime都能支持，至于还未收入这篇文档里的，当然就要等着你们去创造啦~

蓝落萧

一稿 2016.8.23

二稿 2016.12.31