

# 汉字识别中的笔画数效应新探<sup>\* 1)</sup>

## ——兼论字频效应

喻柏林 曹河圻

中国科学院心理研究所, 北京 100012

### 摘 要

本研究采用命名作业测查二至十五画常用字和不常用字的反应时(RT), 结果发现(1)命名RT是笔画数的函数,但在几个小笔画范围内,RT基本无变化;(2)命名作业中有字频效应。本文使用整字加工与字形结构加工的并行假说,对笔画数效应和字频效应的机制进行了初步探讨。

**关键词:** 汉字、识别、笔画、字频

笔画是组成汉字的最小结构单位。笔画数的多少标志着字的繁简。因而,笔画数就成为汉字形码整体特性的一种描述,同时也是汉字形码重要特征之一。由于字的识别始于其形的识别,所以笔画数与识别的关系早就吸引着不少研究者的兴趣。但时至今日尚未见一致的认识。比如,蔡乐生的研究结论是,“识别力与字的复杂性,未见肯定关系”[参见高尚仁(1982)<sup>[1]</sup>]或者说,“汉字笔画的多少,对辨认难易没有多大关系”[参见沈德立(1982)<sup>[2]</sup>]。而曹传咏、沈晔(1963)<sup>[3]</sup>与叶重新、刘英茂(1972)<sup>[4]</sup>的实验结果是,多笔画的字比少笔画的字难于辨认,从中肯定了汉字识别中的笔画数效应。前人在认识上的不一致给我们留下了继续探讨的任务。这种不一致其中也许与研究选用的笔画范围有关。本研究在前人工作基础上,拟将笔画范围在今日实用简化汉字中尽可能地加以扩大,以求获得能更全面反映笔画数效应的结论。

此外,肯定笔画数效应的研究者在实验方法上使用的方法是速示法。而瞬间呈现刺激字的认知成绩,首先含有刺激字消失后的猜测、推理等知觉后的过程;其次,瞬间呈现下被试者抓住的字的特征与笔画数之间的关系问题,也还有待进一步考查。叶、刘(1972)在解释笔画数效应时说:“……笔画数多与笔画数少的字都占有同样大的空间,因此对于笔画数少的字来说比较容易辨认字的特征。”并进而提出,“假如将笔画数多的字按比率放大,则可以推测笔画数对认识阈可能没有影响。”这一基于笔画密度的视清晰度假说,似乎有理。但字的特征又是什么呢?该文没有任何说明。如果把在速示下最先,同时也是最易辨认出的字的组成成分叫做特征,那么,彭瑞祥、喻柏林(1983)<sup>[5]</sup>发现,字的各个结构层次如笔画、笔画组合甚至成字部件,都可成为字的特征。它们或是因其笔画密度相对小一些,提

\* 本研究得到国家高技术研究发展计划和国家自然科学基金的资助。

1) 本文于1991年8月16日收到。

供的视清晰度就高,如相对游离的笔画、轮廓等;或是它们本身即是字的突出特征,而和字的笔画数并无直接的关系,如拐折处、字的中轴部位等等因素。可见,视觉清晰度的说法在解释笔画数效应机制时,至少是不全面的。本研究采用不同于速示条件的另一种视知觉实验范式,即不限制刺激字呈现时间的命名识别,以向被试提供各种笔画下的清晰视觉条件,并保证被试的识别是在现实刺激的作用之下。我们试图在视知觉条件充分及尽可能广泛的笔画范围内考查笔画数效应,并从认知心理学观点探讨其可能的机制。

为了叙述的方便,本文分为四个小实验。它们使用同一批仪器、同一种实验程序并由同一批被试完成。各实验变量其细节详见各实验说明。

### 实验方法

**仪器:** 一台 AST/P286 微机和 EGA 彩色显示器,以及声音开关、话筒、控制键、打印机等外部设备。

**刺激字:** 独体字(二画至六画)和左右型合体字(七画至十五画)共 115 个,详见附录 1 和 2。每个字高 0.6cm、宽 0.55cm。

**被试:** 男女大学生共 32 名,裸视或矫正视力正常。

**程序:** 实验在半暗室内进行。被试坐在距显示器约 30cm 前方使用双眼观察。经过两个段落(每段 10 个字)的练习之后,被试能做到又快又准地识读显示器上呈现的汉字。每一个字试验程序如下: 每字呈现前有一注视点与铃声共同作用 200ms,然后空屏 400ms 呈现刺激字。一旦被试读出该字字音,此字即消失。计算机记录从刺激字出现到声音反应开始这段潜伏时间——反应时(RT)。主试操作控制键向计算机输入被试的错误反应。间隔 1 秒再进行下一个字的识别。115 个刺激字针对每名被试随机混合,并被分成四组,每组前有 3 个练习字,对其反应不计入正式结果。组与组试验之间被试有几分钟休息时间。

## 实 验 一

选用最常用的 I 级字作为刺激字<sup>[注]</sup>。笔画范围为 2—6 画,相邻两种笔画数差距为 1 画。每种笔画下有 5 个字。它们都是由一个部件组成的独体字,且又是组成合体字的常用成字部件。32 名被试平均识读成绩表明,识读每种笔画下 5 个字的错误率均小于 2%,故本文只考虑反应时 RT 一项指标。笔画数与命名 RT 的关系,详见表 1。可见,在 2—

表 1 笔画数与命名反应时(ms)的关系  
(常用 I 级字)

反应指标	笔 画 数 (画)								
	独 体 字					合 体 字			
	二	三	四	五	六	七	九	十一	十三
反应时	398.4	389.9	408.3	398.6	415.5	431.9	438.9	466.2	462.5

5 画内,相邻笔画数之间于 RT 上的差值甚微,且高低起伏波动。六画下增长较快。经 S-PSS 统计软件包中 Manova 重复测量的单因素变异数分析表明,2—6 画内笔画数效应

注: 本研究选用刺激字的频率等级全部取自《汉字信息处理词典》(上海交通大学汉字编码组和上海汉语拼音文字研究组编,科学出版社,1983年)。

显著 $[F(4, 124) = 7.67, p = .000]$ ; 相邻两平均数之差  $t$  检验显示, 除二与三画之差并不显著外, 其他都显著。若进行后一项与前面各项平均数之差的  $t$  检验, 则可发现: 四画与三画前平均数之差显著( $t_{31} = 2.659, p = .012$ ); 而五画与四画前平均数之差不显著( $t_{31} = 0.08 < 1$ ); 但六画与五画前平均数之差又非常显著( $t_{31} = 4.751, p = .000$ )。这表明, 在如此少笔画的 2—6 画范围内, 成年熟练阅读者的命名反应, 总的说来受笔画数增加的影响, 但在某些小笔画范围内不受影响, 或基本不受影响(二至五画内), 而在某些笔画数(或范围)之间却有明显的影响(如五画与六画, 或六画与五画前的一个范围)。命名 RT 随笔画数增加而发生的这种变化, 可以形象地用台阶状上升之势来描述: 即 RT 在组成台阶的笔画数内其差异不显著, 但在台阶之间其差异是显著的。下个实验将扩大笔画数范围, 以便对这种笔画数效应做进一步考查。

## 实 验 二

为了与少笔画字的字频等级保持可比性, 本实验仍选用最常用的 I 级字。选用七、九、十一、及十三画共 4 种笔画。每种笔画下有 5 个字。它们全部是左右型合体字, 其构字复杂程度有如下约定: 经一次切分产生两个一级部件后, 只允许至多有一个部件可进行第二次切分, 但不可再进行第三次切分。识读结果表明, 每种笔画下的平均错误率都小于 4%, 且任一对平均数之差的  $t$  检验都不显著, 故不予以讨论。

表 1 所列合体字的结果表明, 命名 RT 有随笔画数增加而增加之势, 其变异数分析也是显著的 $[F(3, 93) = 11.06, p = .000]$ 。但成对平均数  $t$  检验中, 七画与九画之差值、以及十一画与十三画之差值分别不显著( $t_{31} = .84; .55 < 1$ ), 但九画与十一画之差却是显著的( $t_{31} = 3.79, p = .001$ )。由此看来, RT 是笔画数的函数: 随着笔画数的增加, RT 呈现两级台阶状上升趋势。这种函数关系完全重现了独体字下的结果模式。为了再次验证这个结果模式, 我们又在常用 II 级字和次常用 III 级字中选字, 并把笔画数扩展至十五画。这就是实验三的用意。

## 实 验 三

选择七、九、十一、十三和十五画共 5 种笔画。每种笔画下有 II 级字与 III 级字各 5 个。选字原则同实验二。实验结果表明, 在任一条件下平均错误率都没有超过 4%, 且任一对平均数之差都不显著, 故表 2 只列出各字频级于多种笔画数下 32 人平均 RT 的数据。

首先考查 II 级字的结果。在七至十一画内, RT 随笔画数增加而增加, 其后十三画下似有下降, 但十五画又有较大回升。变异数分析表明, 七至十五画范围内笔画数效应也是非常显著的 $[F(4, 124) = 25.40, p = .000]$ ; 成对平均数之差  $t$  检验显示, 除十一画与十三画之差并不显著( $t_{31} = 1.44, p = .161$ ), 组成一级台阶外, 其他三对相邻两平均数之差都显著( $t_{31} = 3.61, 5.26, 2.68; p = .001, .000, .012$ )。可见, II 级字重现了 I 级字的结果模式。至于 III 级字下笔画数效应考查, 从表 2 可见, 七至十三画内显然无笔画数效应(任一对平均数之差的  $t$  检验都不显著), 它们组成一级台阶。但十五画下 RT 增值较大, 十三画与十五画之差显著( $t_{31} = 4.67, p = .000$ )。在七至十五画范围内, 变异数分析表明, 笔画数效应也非常显著 $[F(4, 124) = 10.56, p = .000]$ 。可见, III 级字的笔画数效应, 其具体形式也完全同于独体字, 且与 I、II 级合体字类似。

表2 笔画数与命名反应时(ms)的关系(常用Ⅰ级字和次常用Ⅲ级字)

字频级	笔画数(画)				
	七	九	十一	十三	十五
I*	431.9	438.9	466.2	462.5	—
II	430.8	451.6	484.3	475.7	503.3
III	468.2	456.8	464.9	464.1	502.3
平均	443.6	449.1	471.8	467.4	503.1

\* 引自表1。

总之,统观已有独体字与常用Ⅰ,Ⅱ级Ⅲ级合体字的实验结果,一致表明,随着笔画数的增加,RT呈现台阶式的上升趋势。如果跨出常用字而进入使用频度低的字,RT与笔画数间的这种函数关系是否仍然如此?叶重新、刘英茂(1972)的工作指出“在高意义度的字中,笔画数的多少对认识阈有很大的差异,而在低意义度的字中,笔画数对认识阈的影响较小”。他们这一结果是在不充分视觉加工,即速示条件下获得的。我们拟在不限时的充分视觉加工下验证字频与笔画数的交互作用。这就是实验四的任务。

## 实验四

本实验选用字频相对低的不常用Ⅳ级字作为重复考查笔画数效应的刺激字,并取表2中常用Ⅰ、Ⅱ与Ⅲ级字的平均RT与Ⅳ级字同笔画数下的RT作一番比较,以探讨笔画数效应与字频效应间是否存在交互作用。

Ⅳ级字只包括九、十一、十三和十五画共4种笔画数。每种笔画下仍取5个字。结果表明,每种笔画下的平均错误率不超过4%。同理,不予讨论。从表3中可见,不常用字

表3 笔画数与字频对字命名反应时(ms)的影响

字频级	笔画数(画)				
	九	十一	十三	十五	平均
不常用级	479.2	511.1	513.3	539.8	510.9
常用级	449.1	471.8	467.4	503.1	472.9
差值	30.1	39.3	45.8	36.7	38.0

下笔画数效应显著 $[F(3, 93) = 9.74, p = .000]$ ,而且RT随笔画数增加也呈现台阶状上升趋势。相邻两两平均数之差的t考验表明,只有十一画与十三画之差不显著( $t_{31} = 0.227 < 1$ ),组成一级台阶;而其他两对平均数之差都显著( $t_{31} = 2.903, 2.430; p = 0.007, .021$ ),可见,不常用字重现了常用字的结果模式。

其次,常用字与不常用字在四种笔画下的结果比较,经笔画数(4)×字频(2)两因素变异数分析表明,笔画数主效应非常显著 $[F(3, 93) = 27.10, p = 0.000]$ ;命名常用字快于不常用字的字频效应也非常显著 $[F(1, 31) = 99.59, p = 0.000]$ ;但两因素间无交互作用 $[F(3, 93) = 0.63, p = .599]$ 。这表明,笔画数主效应不受字频因素影响,它恒定地出现在高与低频字之中。

## 总的讨论

### 一、关于字频效应

我们所获笔画数与字频因素无交互作用的结果,显然与叶、刘(1972)的结果相反。我们暂且把这两项工作在结果上的差异归因于速示与非速示实验条件之差吧。但叶、刘一文就笔画数对低意义度字认识阈影响较小所做解释,似可商榷:“低意义度的字是较少见过的字,因此其字的特征亦可能很少接触过。因此,不管笔画数的多少,受试者皆不易靠发现字的特征来认明一个字”。可见该解释的前提条件是,字低频其结构特征也低频。而这一论断本身还有待理论的和实验的验证。且不说成其为字的特征的某些较复杂的结构成分,就是最简单的相对游离的笔画,它们既可存在于高频字之中,也可存在于等笔画的低频字之中。对它们所在整字辨认成绩上的差异,总不该归因于它们在使用频率上的差异吧。我们认为速示条件下的字频效应,或低频字的无笔画数效应,与其说是发生在字特征抽取(或辨认)的知觉阶段,还不如说是发生在刺激消失之后,据此特征进行部分与整体关系推断的知觉后阶段。在此阶段,对高频字的整体知觉比对低频字的占优势,因而就可能产生字频效应。而低频字的识别之所以不太受笔画数影响,也皆由于整字使用频率低。即使是少笔画字,当由其特征去推断整字时,也不会比等频的多笔画字显得容易。

其次,字频效应的机制究竟是什么? Seidenberg (1985)<sup>[6]</sup>在论证他的词再认时序模型时,也曾获得命名汉字高频字快于低频字的结果。他的解释是,熟练阅读者对绝大多数单词再认,是以词形为基础直接完成识别的,语音码在词再认之后才被激活。低频字再认速度慢,语音码有时间被激活,语音转录是它的必经之道。暂且不论高、低频字在语音转录上的这种差异,但这一解释并没有就低频字再认速度何以缓慢做出解答。我们认为,本研究不限时条件下的字频效应,一是不能归因于高频字比低频字有优越的视觉观察条件,二是不能归因于字形结构方式或笔画数,因为高低频字有着同等的条件。那么字频效应的机制究竟是什么呢? 喻柏林等人(1990)<sup>[7]</sup>曾就汉字整体知觉做过具体分析,指出汉字形、音、义三种特性,既有相对独立的整体性(即字形的完整性、字音的单一性和字义的独特性),又可相互结合为一整体。而人类信息加工系统有着相应的表征机制对它们进行相对分离的加工操作,或协同对整字的加工操作。有鉴于此,字形码中的结构方式和笔画数等因素尽管在高低频字中保持一致,但是由等笔画数构成的字形完整性却是各不相同的。人们在长期使用中对其字形整体性形成的整体认知之熟悉程度也有高、低之分。这样,高频字就比低频字占有认知优势,命名反应时就会快一些。其次,形码与音码间的联结是由人们通过长期学习和使用形成的。因而使用频率高的字就较频率低的字,其形音联结紧密度就要大些,也即由于形音码加性效应所致的字之整体知觉也要强一些,于是“见字读音”的速度也要快些,而不管由形码信息至读音反应之间是否存在如 Seidenberg 所言的语音中介。(确切地说,对于高频字的读音反应是以语义码的接通为中介的)。

### 二、关于笔画数效应

本研究考查的二一十五画这样一个相当广泛的笔画数范围,它囊括了简化和未简化的 11834 字中的 9934 个字,即占 11834 个字中的 84% [作者的计算值,请参看付永和(1985)<sup>[8]</sup>]。这么大的比例数说明,我们选取的笔画数范围具有很大的代表性。本研究在

此范围内获得的笔画数效应可能意味着,被试识别每个单字经历着两类平行加工的过程。第一类是认知任务规定的识别整字的加工。喻柏林等人(1991)<sup>[9]</sup>已指出,整字是正常阅读条件下的知觉单位。这是就人们没有对字的结构方式和表音特点做出知觉分析而言的,此时是将整字作为一个整的视觉模式表征的。现在,如果我们从已有整字识别的工作出发,则该承认不论字的复杂程度如何,整字既是知觉单位,它们就应有同样的识别反应时,但现有实验结果与此大相径庭。这意味着很可能存在着与整字加工并行的另一类加工。字的多个笔画及其组合体作为光刺激同时作用于视网膜,并获得感觉登记。这样,人们在工作记忆中加工整字时,同时也在加工笔画及其组合。于是,笔画及其组合的加工与整字加工发生分享有限注意(或知觉)资源的竞争。可以想象,笔画数多分享的注意就多,从而留给整字加工的注意就会减少。此时整字识别需要借助增加 RT 以求补偿。同理,笔画少占用的注意资源就少,整字加工就占有相对多的注意,于是识别 RT 就少。但是在相邻笔画数的某一小范围内,笔画因素占用的注意相差无几,甚或相同,这样,它们对应的整字也就享受着大体相同的注意。因而,在整字识别的 RT 随笔画数增加而增加的总趋势之下,还可见到 RT 出现一个或几个相对稳定的平坦台阶。如果只见 RT 随笔画数在某一较小范围内的增加而出现的直线上升之势,那么,就易误入以偏概全之迷途;如果只见某一台阶,则易得出识别时间与笔画数无多大关系的看法。最后可以认为,这种整字与笔画(成分)同时竞争有限注意资源的平行加工假说,比较符合平行分布加工模型[例如, McClell and Rumelhart(1981)<sup>[10]</sup>]的基本思想。

附录1 I级字频独体字

笔画数(画)				
二	三	四	五	六
厂人力 又刀	土才门 马大	天中火 日不	平东主 电用	产虫米 西农

附录2 左右型合体字\*

字频(级)	笔画数(画)				
	七	九	十一	十三	十五
I	坏报别 利(体)	活除标 相研	断接检 粗深	新解群 搞惯	/
II	听劲针 财乱	封脉独 轻秋	船够敢 斜探	跟献暗 鼓劲	横播稻 懂(慎)
III	吐狂妙 馥邮	剑砍秒 狠恒	蛇甜跃 欲惯	睡填碰 矮楼	颜躺醋 撞聪
IV	/	殃盼虹 拯(饮)	聊琐梳 (硕)(蚌)	聘酬颞 韵魂	撕嗜嘘 (磁)(碾)

\*表中( )内的字没参与实验结果的计算,因为32人中有5人以上发生读音错误

## 参 考 文 献

- [1] 高尚仁, 中国语文的心理与研究, 载高尚仁、郑昭明合编《中国语文的心理学研究》第4页, 文鹤出版有限公司, 1982年。
- [2] 沈德立, 语言的知觉理解与表达, 心理学杂志, 1988年, 1, 60—66。
- [8] 曹传咏 沈畴, 在速示条件下儿童辨认汉字字形的试探性研究, 心理学报, 1963年, 4, 271—279。
- [4] 叶重新、刘英茂, 影响本国文字认识网的因素, 台湾大学理学院心理学系研究报告, 1972年, 14, 113—117。
- [5] 彭瑞祥、喻柏林, 不同结构的汉字的再认研究, 普通心理学与实验心理学论文集, 甘肃人民出版社, 1983年, 182—194。
- [6] Seidenberg, M. S., The time-course of phonological code activation in two writing systems, *Cognition*, 1985, 19, 1—30。
- [7] 喻柏林、曹河圻等, 汉字形码和音码的整体性对部件识别的影响, 心理学报, 1990, 3, 232—239。
- [8] 付永和, 汉字结构及其构成成分的分析 and 统计, 中国语文, 1985, 4, 261—272。
- [9] 喻柏林 冯玲等, 汉字的视知觉, 心理学报, 1990, 2, 141—148。
- [10] McClelland J. L., Rumelhart D. E., An interactive activation model of context effects in letter perception; part 1. An account of basic findings, *Psychological Review*, 1981, 5, 375—405。

## A NEW EXPLORATION ON THE EFFECT OF STROKE-NUMBER IN THE IDENTIFICATION OF CHINESE CHARACTERS

Yu Bolin    Cao Heqi

*Institute of Psychology, Academia Sinica, Beijing*

### Abstract

Adopting the task of naming, the present research examined the response time(RT)of characters with high and low frequencies in the range from 2 to 15 strokes, it was discovered that (1) the RT of naming was the function of stroke-number, but basically there was no change in RT in a few small ranges of the stroke; (2) there was effect of frequency in the naming task. Finally, by using the hypothesis of parallel processing of the whole character and morphological structure, the present research preliminarily explored the mechanisms of the effects of the stroke number and the frequency.

**Key words:** Chinese, identification, stroke frequency