

学校的理想装备

电子图书·学校专集

校园网上的最佳资源

实验心理学

 **eBOOK**
网络资源 免费下载

自序

心理学是当代社会与行为科学中发展最迅速的学科之一。关于心理学的探索研究，目前可说是正处在“著人滋味，真个浓如酒”的时期。在心理学科的百花园中，实验心理学的发展尤为突出。实验法是心理学研究的主要方法，它不仅促使心理学成为一门独立科学，而且其日新月异的演进，也推动着心理学各领域的迅猛发展。诚然，由美国心理学家斯珀灵（George Sperling）1960年首次公布的部分报告法，证实了感觉记忆的存在。此外诸如信号检测论的应用，使对内部心理的分析更上层楼，反应时新法使认知心理学大展宏图，间接测量法导致了内隐记忆的新发现。由此等实验心理学的发展，在在有力地说明了实验方法是揭露心理和行为的规律性的重要途径和手段。一位心理学家可以对心理学任一领域任一分支感兴趣，可以专门从事工业心理、医学心理、教育心理或知觉心理、记忆心理、思维心理，以至社会心理的研究，但是他们必定有一个共同的特点，即是确切地掌握了实验心理学的研究方法，了解应当如何科学地考察心理和行为的规律。这就足以证明实验心理学的重要性。

本人有幸受邀参加《世纪心理学丛书》中《实验心理学》一书的编写工作，获得了一次珍贵的学习锻炼机会。本书是专为大学和师范院校学生撰写的一本教科书，一般读过《普通心理学》和《心理统计学》者，都可读懂此书。本书采用教科书的形式，宗旨是将实验心理学中已经确立起来的最基本的、最可靠的科学方法介绍给同学和读者，使同学们初步掌握实验心理学的重要理论，学会基本的实验设计，并具有相当的实验技能。

本书的成功出版，首先要归功于张春兴教授和卓鑫森先生。台湾师范大学张春兴教授费五年多心力，邀集两岸心理学者合作撰写《世纪心理学丛书》，对未来中国心理科学的发展，意义极为重大。台湾东华书局负责人卓鑫森先生，鼎力资助由繁简两种字体在两岸发行此《世纪心理学丛书》，对两岸文化学术交流，贡献卓著。本书的编写过程中，又得到主编张春兴教授的关怀和悉心协助，以及东华书局编辑们在体例、图表乃至内容上的诸多帮助。没有这些鼓励和帮助是无法写出本书的。还要说明的是，本书在写作过程中参考了国内外有关心理学专著、论文和文献资料，吸取了许多学者的实验成果也引用了我本人过去编写的著作和与人合著的著作，在此一并向原作者致深深的谢意。

最后要说明的是，由于本书编者自身水平和时间的限制，疏漏和缺失在所难免，恳切地期望广大读者批评指正。

杨治良 谨识

一九九七年一月于华东师范大学

实验心理学

第一章

绪论

本章内容细目

第一节 实验心理学的性质

- 一、实验心理学的含义 3
- 二、实验中的几种变量 4
- 三、实验中主试者与被试者的关系 8
 - (一) 按实验程序进行的主-被试者的相互作用
 - (二) 干扰实验程序的主-被试者的相互作用
- 四、额外变量的控制 11
- 五、实验效度 13
 - (一) 影响实验内部效度的因素
 - (二) 影响实验外部效度的因素

第二节 心理实验的程序

- 一、课题的确定 17
 - (一) 课题的来源
 - (二) 实验类型的确定
 - (三) 问题的陈述
- 二、被试者的选择 19
 - (一) 依问题的性质选择被试者
 - (二) 依研究结果的概括程度来选择被试者
- 三、实验因素的控制 20
 - (一) 自变量的控制
 - (二) 反应的控制
 - (三) 反应指标的选择
 - (四) 仪器和材料的使用
- 四、实验资料的收集与分析 and 综合 24
 - (一) 心理学研究的资料类别
 - (二) 研究资料的整理与解释
- 五、撰写实验报告 26

第三节 实验心理学简史

- 一、心理实验发展的第一阶段 0
- 二、心理实验发展的第二阶段 1

本章实验

- 一、自变量和因变量的确定 6

本章摘要

建议参考资料

实验心理学一词，是 1862 年由德国著名心理学家、实验心理学的先驱之一冯特（Wilhelm Wundt，1832～1920）在他的《感官知觉理论贡献》论文集的导言里最早提出来的。冯特提出“实验心理学”是作为他创建的新心理学（new psychology）的代名词。为什么会出现创建新心理学的愿望呢？这是因为心理学在成为一门独立的科学之前始终作为哲学的一个附属部分，是由哲学家们用思辨的方法来进行研究的。冯特所说的实验心理学是指用实验的方法来进行研究的心理学，亦即科学的新心理学，它是为了与以前的心理学相区别而被提出的。但这时他所说的实验心理学只是指用实验的方法来研究心理学中的感知觉问题，而对于思维、想象、情感等高级的心理过程，他认为不能用实验的方法来进行研究的，他称用非实验方法研究的这部分心理学为民族心理学（folk psychology）。尽管冯特的观点有一定的局限性，但在当时乃至今日仍不失为心理学中比较好的研究方法，为科学心理学的创建立下了功劳。我们把 1879 年冯特在德国莱比锡大学建立第一个研究心理学的实验室这一事件作为心理学从哲学中分化出来，成为一门独立学科的标志，并称他为实验心理学的创始人。

虽然心理学的实验方法都是在加以控制的条件下进行研究的，但由于不同的心理学流派所主张的心理学研究的对象不同，在控制条件下观察什么则不尽相同。例如，构造主义（或结构主义）（structuralism）者就主张用内省法（introspection）观察自己的直接经验，并企图把意识分为最简单、最基本的心理元素；行为主义（behaviorism）者则主张在控制条件下观察人和动物的行为；等等。

在 20 世纪 50 年代，随着电子学和工程技术的发展，心理学的实验装置和测量仪器日益精密，从而使心理学的实验研究工作在客观性和准确性方面都得到了提高，并逐渐摆脱了不同心理学流派对实验方法的束缚和影响。心理学发展到今天，争论的问题已不再是实验方法对心理学的研究是否适用，而是如何使心理学的实验方法更加完善、更加自动化，以及如何用在实验室中发现的心理学规律来解决实际问题。

第一章的绪论，是为了使读者在进入以后各章的专题讨论前，先对实验心理学有一个概括性的认识。因此，本章的撰写旨在回答以下六个问题：

1. 实验心理学是什么样的学科。什么叫实验。
2. 什么叫自变量？什么叫因变量？并举例说明。
3. 什么叫额外变量？对额外变量的控制通常采用哪些方法。
4. 心理学研究的课题主要来自哪些方面。
5. 一个完整的实验报告应包括哪几项内容。
6. 从实验心理学简史的学习中，获得了哪些启示。

第一节 实验心理学的性质

一、实验心理学的含义

实验心理学 (experimental psychology) 就是在实验控制条件下对心理和行为进行研究工作心理学。这一心理学分支是以研究方法来定义的。根据这个定义, 只要是用实验法 (experimental method) 来研究的心理学问题, 都可以包括在实验心理学的范围之内。不过, 实验心理学所讲的实验法一般是指在严密控制条件下的实验室实验。实验心理学的内容通常包括两个部分, 一部分是阐述实验方法和实验设计, 另一部分则是阐述实验法在一些专门领域中的应用, 如心理物理学、知觉、学习、思维、情绪等专题。前者论述心理学实验的一般原理, 后者论述在具体研究领域如何应用实验方法等问题。

实验法不等于自然观察法 (naturalistic observation)。在使用自然观察法时, 研究者只能被动地仔细观察和记录研究对象在自然状态下所发生的情况, 而不能有任何干预。虽然长期的、系统的观察也可以发现事物之间的规律, 例如天文学中的规律大部分是这样发现的。但这种观察只能等待所要观察的事物出现时才能进行, 受自然条件的限制。而实验 (experiment) 则是人为地去干预、控制所研究的对象。实验者可以创造条件, 引发所需要的事件来观察其变化; 为了验证, 可以创造同样的条件进行重复观察。与自然观察法相比较, 实验法是探寻事物之间的规律的一种更加有效的办法。

现代心理学是一个非常庞大的学科体系, 包含有许多心理学分支, 例如普通心理学 (general psychology)、生理心理学 (physiological psychology)、学习心理学 (psychology of learning)、认知心理学 (cognitive psychology)、发展心理学 (developmental psychology)、社会心理学 (social psychology)、人格心理学 (personality psychology)、工业心理学 (industrial psychology)、教育心理学 (educational psychology)、管理心理学 (managerial psychology) 等等。在现代心理学的各分支学科中几乎都在应用实验法, 并且应用实验法的范围正在日益扩大; 即使在过去曾被认为难以进行实验研究的人格心理学和社会心理学也都在进行实验研究。心理学家把那些用实验法所得到的科学资料概括为实验心理学的综合成果, 从而发展了实验心理学; 同时, 实验心理学也成了现代心理学各分支学科的发展基础。实验心理学在现代心理学中的地位及其与其他心理学分支学科的关系如图 1-1 所示。

二、实验中的几种变量

从以上叙述中, 我们可以知道实验是一种控制的观察。与自然的或偶然的观察不同, 实验是实验者人为地使现象发生, 对产生现象的情景或影响现象的条件加以操纵、变化以及控制的观察。然而, 自然的或偶然的观察是研究者在自然的情境中任现象自然的发生, 对现象及其发生的情境不加入人为的干预而进行的。与其他研究方法相比, 实验法的主要特点是能较好地控制额外变量。伍德沃斯和施洛斯贝格 (Woodworth & Schlosberg, 1954) 指出了实验具有三个优点: (1) 在实验的方法中, 实验者可以在他愿意时, 使事件产生, 可以充分地进行精密的观察。(2) 实验在同样条件下是可以重复的, 别人可以验证它。(3) 系统地变化条件, 可以追究与此相随的事件的变化。在这三个优点中, 第 2 个优点是特别重要的。只有按照同样的手续, 能被别

人重复验证的事实，才能构成科学的知识。当然，每个实验的复杂程度是不同的。在简单、单一的实验里，只让一个条件变化而控制其余的一切条件。在妥善安排的复杂设计里，可以让两个或两个以上的条件同时变化，分析每个条件的单独影响以及它们之间可能的相互作用。在我们明确了“实验”的含义后，我们就可以来讨论什么是变量？变量主要有几种？

变量（或变项）（variable）是指在数量上或质量上可变的事物的属性。例如，光的强度可以由弱变强，呈现时间可以由短变长，智力的 IQ 可以由小到大，这些都属于量的变量。又如，人的性别有男有女，人的宗教信仰有佛教、道教、基督教、天主教和伊斯兰教等等，这些是质的变量。质的变量有时也可以用数字代替类别，以便于统计分析。

在自然科学中，常用数学方程式来描述一些现象。用实验法研究心理学问题时，也用数学方程式来探明变量与变量的关系。在实验中实验者所操纵的、对被试者的反应产生影响的变量称为自变量（或自变项）（independent variable，简称 I.V.）；由操纵自变量而引起的被试者的某种特定反应称为因变量（或依变项）（dependent variable，简称 D.V.）。例如，要研究灯光亮度对阅读速度的影响时，实验者所操纵的灯光亮度（变化烛光数）就是自变量，而阅读速度（以每分钟多少字来表示）就是因变量。因此，自变量和因变量是相互依存的，没有自变量就无所谓因变量，没有因变量也无所谓自变量。

除了自变量之外，还有其他许多因素都会影响因变量的变化。上例中，文字的形状、大小、排列方式、清晰度、熟悉度以及材料的呈现位置和呈现先后等都会影响阅读速度。此外，像大气压、被试者的高矮、胖瘦和头发长短等因素对阅读速度就没有什么影响。凡是对因变量产生影响的实验条件都称为相关变量（relevant variable，简称 R.V.），而对因变量不产生影响的实验条件称为无关变量（或无关变项）（irrelevant variable，简称 I.V.）。在相关变量中，实验者用以研究的变量称为自变量，实验者不用于研究的那些相关变量称为额外相关变量（extraneous relevant variable），或简称为额外变量（或外扰变项）（extraneous variable）。在实验中，额外变量是必须加以控制的。如果不控制额外变量，就会弄不清因变量的变化是由自变量的影响引起的，还是由因变量的变化引起的。因而就无法得出明确的结论。由于在实验中额外变量是必须加以控制的，所以额外变量也被称为控制变量（controlled variable）。评价一项实验设计的好坏的一个重要依据就是看研究者能否成功地控制那些额外变量。

自变量的种类很多，大致可以分为三类：

1. 作业 作业（task）是指实验中要求被试者作出特定反应的某种呈现刺激，例如，字母串、配对联想词表、Müller-Lyer 错觉图、字谜等等。如果把这些作业的任何特性作为自变量来操纵，则这种自变量即为一种作业变量（task variable）。

2. 环境 当实验呈现某种作业时，如果改变实验环境（environment）的任何特性，则改变了的环境特性即为环境自变量（environmental independent variable）。例如，我们可以改变实验室内的亮度或噪声（噪音）强度，也可以改变呈现刺激的时间间隔等等。

3. 被试者 被试者的特性因素如年龄、性别、健康状况、智力、教育水平、人格特性、动机、态度、内驱力等都可能影响对某种刺激的反应，这些因素

统称为被试者变量（或受试者变项）（subject variable）。在这些被试者变量中，有的是实验者可以主动操纵加以改变的，例如内驱力强度可以用禁食或禁饮的时间来加以操纵，而有的则是不能主动操纵的，只能进行测量，例如智力、教育水平、自我强度等。被试者本身固有的、实验者不能加以操纵使其改变的特性称为本性变量（或属性变项）（attribute variable）。

在某些情况下，研究者把几个不同的自变量当作一个复合自变量（complex independent variable）来操纵，以确定它们的综合效应。例如，有些学校为提高学生的学习成绩进行实验，采取了一系列的教学改革措施，如使用新的教材、加强课堂管理、奖励成就等。假定这些措施确实提高了学生的学习成绩，那我们是不大可能鉴别出哪一个自变量在起什么作用的，或许其中只有一种措施才真正起作用，但设计这种实验的目的并非要鉴别出某个变量的作用，而是考察其综合效应。因此，只要研究者不对复合自变量作出分析性结论，这类研究无可厚非。由于复合自变量更接近日常生活实际，并能解决某些实际问题，因而常被研究者所采用。

因为自变量的变化而产生的现象变化、或结果，称为因变量，也就是反应变量（response variable）、或反应测量（response measurement）。对于被试者的反应可以从下列几方面来测量：（1）反应速度，例如，简单反应时间、或潜伏期，走完一个迷津所需的时间，在一定时间内完成某项作业的数量等。（2）反应的正确性，例如，选择反应正确的次数、走迷宫入盲巷的次数、跟追踪盘离靶的次数或距离等。（3）反应的难度，有些作业可以定出一个难易等级，看被试者或动物能达到什么水平，如斯金纳箱就有三个难度等级。（4）反应的次数或机率，是指在一定时间内被试者能做出某种反应的次数，例如在心理物理学实验中，根据机率来规定阈限。（5）反应的强度，例如，情绪实验中的出汗量，皮肤电反射电阻变化的大小，等等。

除了上述反应指标外，被试者的口语报告内容（即口语记录）也是一项重要的反应变量。口语记录（protocol）是指被试者在实验时对自己心理活动进程所作叙述的记录，或在实验之后，他对主试者提出问题所作回答的记录。在心理学实验中，口语记录是很重要的参考资料，有助于我们分析被试者的内部心理活动。目前口语记录分析已被用于问题解决的实验中，借以了解被试者解决问题时所使用的思维策略。

对于人的心理活动，我们无法直接观察到。但他的行为反应是受心理活动支配的，是我们可以观察到的。心理学家之所以对反应指标的测量感到兴趣，是因为他们相信从由刺激引起的反应指标可以推知某些更基本的心理过程。例如我们让被试者解一道数学题： $6+7+8+9=?$ 要求他们在解题时说出头脑中所进行的一切活动，并记录其得出的答案和使用了多少时间。结果发现，虽然被试者的答案都是对的，但解题所用的时间却有显著差异，有的解得快；有的解得慢。进一步分析他们的口语记录，发现解得慢的被试者是用累加法，解得快的被试者是用另一种策略，先做 $7+8=15$ ， $6+9=15$ ，再做 $15+15=30$ 。这就是说，在这个实验中，一个变量（一道数学题）导致另一个变量（解题所用的时间多少）的变化是由某个中介变量（intervening variable）（解题所用的策略）所引起的。现代实验心理学的目的就是要说明和解释人在完成某种活动时的心理活动是如何进行的，即通过刺激和反应（即 S—R）之间的关系来推断心理活动的方式。

然而，在没有控制的条件下，一个刺激可能会引起多种心理和行为的反

应；相反，一种反应也可能是由多种刺激复合作用的结果。因此，如果要建立 S—R 的对应关系，就必须发展一系列的有效措施，进行严密的实验设计。这是心理学实验不同于物理、化学等实验的第一个特点。

三、实验中主试者与被试者的关系

心理实验的对象统称为被试者 (subjects)，包括人和动物。在实验中，特别是以人为对象的实验中，如何处理好实验者（即主试者）和被试者的关系是实验取得成功的一个重要条件。在心理实验中主试者 (experimenter) 和被试者的关系如图 1-2 所示，包括两类性质不同的相互作用：(1) 按实验程序进行的相互作用，(2) 干扰实验程序的相互作用。

(一) 按实验程序进行的主-被试者的相互作用

在心理学实验中涉及到的人就是主试者和被试者。主试者对被试者的干预及被试者对主试者的实验态度都会对实验结果产生影响。由于心理实验都是通过被试者完成任务的方式进行的，所以主试者对被试者最直接的干预是向被试者交代任务。主试者为交代任务向被试者所讲的话，在心理实验中就称为指示语 (instruction)。

在以人作被试者时，指示语在实验中不仅是对被试者说明实验，更重要的是给被试者设定课题，这也是控制被试者这一有机体变量的一种手段。指示语不同，所得的结果也不相同。因此，主试者在给出指示语时应注意以下几点：

1. 要严格确定给被试者什么样的指示语 不同的实验会有不同的要求，有的要求被试者尽量做得准确，有的要求尽量做得快，还有的要求又准又快。此外，是让被试者按特殊的方式完成某任务，还是让他随使用什么方式去完成任务。类似这样的问题，主试者都要事先确定，写到指示语中去。

2. 在指示语中，要把被试者应当知道的事交代完全 主试者要求被试者所做的事，可能是他从未做过的，要说明将要给他呈现什么，要他怎样做等等。

3. 要保证被试者确实懂得了指示语 指示语要写得简单明确。切忌模棱两可，也不要专门术语。为防止被试者误解指示语，可以让被试者用自己的话重述让他做什么、怎么做等等。

4. 指示语要标准化 事先要把指示语写下来念给被试者听。应做到所给的指示语前后一致。不要任意改变同一指示语中的有关词句。对有些实验最好能使用录音机给出指示语。

在指示语不能充分控制反应时，就要很好地考虑刺激条件和实验装置，使刺激条件、实验装置与指示语配合起来，使被试者只能做出主试者所要求的反应。

(二) 干扰实验程序的主-被试者的相互作用

在心理实验中，除了主试者给出指示语及被试者按指示语完成任务的相互作用外，他们之间还可能存在着某种干扰实验、使实验结果发生混淆的相互作用。例如，主试者在实验中可能以某种方式（如他的表情、手势、语气等）有意无意地影响被试者，使他们的反应附和实验者的期望。这种现象称为实验者效应 (experimenter effect)。实验者效应往往会以一种颇为微妙的方式在起作用。例如，当主试者了解到他的少数几位被试者有某种反应倾向时，他随后观察到的被试者的资料也往往有某种反应倾向。这里，除主试者无意识地以某种方式影响被试者的反应外，还可能故意地对被试者暗示、

提醒或鼓动，或不能耐心地等待被试者的真实反应的出现就进行记录，或仅仅记录自己所期望的行为反应。这种现象甚至在用动物做实验时也可能出现。在罗森塔尔 (Rosenthal, 1966) 的一个研究中，他让选修心理实验课的学生做白鼠走迷津实验，一组主试者用来做实验的白鼠的笼子上贴有“走迷津伶俐”的标签，另一组主试者用来做实验的鼠笼子上贴有“走迷津呆笨”的标签。其实，这些白鼠是随机地被分到这些笼子里的。结果，标记为“走迷津伶俐”的白鼠比标记为“走迷津呆笨”的白鼠学习得快些。实验者的期望莫名其妙地影响了白鼠的行为，动物的行为居然附和于错误的标签。

人类被试者参加实验并不是消极被动的。他们总是以某种动机、态度来对待实验的。因而实验结果在很大程度上也依存于被试者对待实验的态度，我们必须高度警惕实验中的霍桑效应和安慰剂效应 (placebo effect)。所谓霍桑效应 (Hawthorne effect) 是指 1924 年美国芝加哥西部电力公司霍桑工厂所进行的一项实验。研究者系统地改变照明强度，目的是确定工厂的最佳照明条件，借以提高工作效率。结果，不论照明增加或是减少，工人的工作效率都在逐渐提高。后来才发现，这是由于工人们觉得参加实验是厂里在关心他们，从而提高了工作效率。这里，很明显被试者的态度成了自变量，从而影响了反应变量。与霍桑效应相类似的是安慰剂效应。有时，医生开给病人的“药物”实际上并非是药物 (如维生素片)，但当病人相信那是有效的药物，服用后也产生效果。这是病人心理作用的结果。可以这样说，凡是在以人作被试者的任何研究中都可能出现霍桑效应或安慰剂效应现象。被试者对指示语的理解、参与实验的动机、焦虑水平、有关经验、以及当时的心理、生理状态等，都会影响他们完成任务的质量和数量。而被试者的反应成绩又会影响主试者的行为。这种相互作用有的是不知不觉地在进行，主试者往往没有察觉到。因此，主试者给予被试者的某种处理，所获得的不一定就完全是此处理所引起的反应，自变量也不一定只是主试者加以操纵的那个自变量。

总之，在心理学实验中主试者和被试者 (人类被试者) 都是具有主观能动性的。主试者用指示语规定被试者的反应，试图控制额外变量，使因变量的变化成为自变量的一种效应；但与此同时，主试者与被试者又可能以某些干扰实验的方式不知不觉地相互作用，使额外变量成了实验中的自变量，从而混淆了实验结果。这是心理学实验不同于物理、化学实验的另一个特点。

四、额外变量的控制

额外变量是使实验结果发生混淆的主要根源。要提高研究的科学水平，就要采取一定的方法来控制额外变量。对额外变量的控制，通常采用以下几种方法：

1. 排除法 排除法 (elimination method) 是把额外变量从实验中排除出去。如果外界的噪音和光线影响实验，最好的办法是进入隔音室或暗室，这样可把它们排除掉。霍桑效应和实验者效应会影响实验结果，最佳的办法是采用双盲实验 (double blind experiment)。从控制变量的观点来看，排除法确实有效。但用排除法所得到的研究结果却缺乏推论的普遍性。例如，如果顾虑主试者与被试者的彼此接触会影响实验结果，而采用自动呈现刺激及自动记录实验结果的方法，则所得结果便不能对人们日常生活中的同类行为作出推论和解释。

2. 恒定法 恒定法 (constant method) 就是使额外变量在实验的过程中保

持固定不变。如果消除额外变量有困难，就可以采用恒定法。不同的实验场所、不同的实验者、不同的实验时间都是额外变量。有效的控制方法是在同一实验室、由同一实验者、在同一个时间对实验组和控制组使用同样的实验程序进行实验。如果实验时强度变化的噪音无法消除，则可以用噪音发生器发生恒定的噪音来加以掩蔽。除上述实验条件保持恒定外，实验者和控制组被试者的特性（如年龄、性别、自我强度、成就、动机等）也是实验结果发生混淆的主要根源，也应保持恒定。只有这样，两个组在作业上的差异才可归于自变量的效果。用恒定法控制额外变量也有缺点：（1）实验结果不能推广到额外变量的其他水平上去。例如，如果只用男性成人作为被试者进行实验，其结果不能推广到女性成人。（2）操纵的自变量和保持恒定的额外变量可能产生交互作用。例如，如果被试者是男性，实验者是富有魅力的女性，实验时，实验者可能使被试者分心。这是交互作用产生的额外变量。

3. 匹配法匹配法（matching method）是使实验组和控制组中的被试者的特点相等的一种方法。使用匹配法时，先要测量所有被试者和实验中要完成的作业具有高相关的特点；然后根据测得的结果把实验组和控制组的被试者的特点匹配成相等的。若要做“练习对学习射击成功影响”的实验，先预测一下被试者打靶的成绩，然后把两个预测成绩相等（击中环数相等）的被试者分别分到实验组和控制组，匹配成条件相等的两组被试者参加实验。这种方法在理论上虽然可取，但在实际上很难行得通。因为，如果超过一个特性（或因素）以上时，实验者常感到顾此失彼，甚至无法匹配。例如，实验者要同时考虑年龄、性别、起始成绩、智力等因素，力图使所有因素均匹配成相等而编为两组就很困难了。即使能解决此困难，也将使很多被试者不能参加这个实验。更何况，属于中介变量的诸因素，如动机、态度等，更是无法找到可靠的依据进行匹配。因此，匹配法在实际上并不常用。

4. 随机化法随机化（randomization）是根据概率理论，把被试者随机地分派到各处理组中。从界定的被试者总体中用抽签法或随机数字法抽取被试者样本，由于随机取样使总体中的各个成员有同等机会被抽取，因而有相当大的可能性使样本保持与总体有相同的结构。随机取样后，再把随机抽出的被试者样本随机地分到各种处理中去。例如，有三种处理组：实验一组、实验二组、控制组。给每一处理组指定一个数字，如 0、1、2，并定好先分给样本 A，再分给样本 B、C。如果在随机表上遇到“2”，就把样本组 A 定为控制组；再遇到“0”，就把样本组 B 定为实验一，依此类推。从理论上讲，随机法是控制额外变量的最佳方法，因为根据概率理论，各组被试者所具备的各种条件和机会是均等的，不会导致系统性偏差。它不仅能克服匹配法顾此失彼的缺点，还能控制难以观察的中介变量（如动机、情感、疲劳、注意等）。随机法不仅能应用于被试者，也能应用于呈现刺激的安排。例如，如果有许多处理施加于被试者，为了消除系列效应（即前面的处理对后面的处理的影响），可以用随机法安排各种处理出现的顺序。

5. 抵消平衡法抵消平衡法（counterbalancing method）是通过采用某些综合平衡的方式使额外变量的效果互相抵消以达到控制额外变量的目的的方法。这种方法的主要作用是控制序列效应（sequence effect）。如果给被试者施加一系列以固定顺序出现的不同处理，被试者的反应将会受到时序先后的影响。如果先后两种处理在性质上无关，就会产生疲劳的影响。这两种影响都可以使实验发生混淆，因而要加以抵消。如果只有 A、B 两种处理，最常

用的抵消序列效应的方法是用 ABBA 的安排。即对同一组被试者先给予 A 处理，再给予 B 处理；然后倒过来，先给予 B 处理，再给予 A 处理。如果对几组被试者给予两种以上的处理，为了抵消序列效应则可采用拉丁方实验 (Latin square experiment)。

6. 统计控制法上述各种方法都是在实验设计时可以采用的。这些方法统称为实验控制 (experimental control)。但有时候由于条件限制，上述的各种方法不能使用，明知有因素将会影响实验结果，却无法在实验中加以排除或控制。在这种情形下，只有做完实验后采用协方差分析 (或共变数分析) (analysis of covariance)，把影响结果的因素分析出来，以达到对额外变量的控制。这种事后用统计技术来达到控制额外变量的方法，称为统计控制 (statistical control)。例如，在对两班学生进行实验以比较两种教学方法的好坏时，虽然实验者事先知道此两班学生的智力不等，但限于条件，实验前却无法对智力因素加以控制使两班学生的智力水平相当。显然智力是影响实验结果的重要因素。实验后，使用协方差分析将智力因素所产生的影响排除后，就可以比较两种教学方法的优劣了。统计控制法除协方差分析外，还可用偏相关等方法。

五、实验效度

实验效度 (experimental validity) 是指实验方法能达到实验目的的程度。实验目的是验证假设，验证自变量和因变量之间的关系，使实验结果的推论可用以解释和预测其他同类现象。由于不同的实验者在设计上和在对额外变量的控制程度上极不相同，实验的效度也会有很大的不同。此外，每种实验都有几个不同的组成部分，其中每一部分也会影响整个实验的效度。了解影响实验效度的诸因素，将有助于我们评价实验设计的质量，提高实验设计的科学性。实验效度主要包括内部效度和外部效度。下面让我们先讨论影响实验内部效度 (internal validity) 的诸因素，然后再讨论影响实验外部效度 (external validity) 的诸因素。

(一) 影响实验内部效度的因素

实验的内部效度是指实验中的自变量与因变量之间的因果关系的明确程度。一项实验的内部效度高，就意味着因变量的变化确系由特定的自变量引起的。由于除了自变量以外，任何额外变量都可能对因变量产生影响，导致实验结果的混淆。这样我们就难以判定实验中自变量与因变量之间的关系的确定性。因此，要使实验具有较高的内部效度，就必须控制各种额外变量。在设计实验时，如果能考虑到以下六个方面的因素，将有助于提高实验的内部效度。

1. 生长和成熟 除了实验中的自变量可能使个体行为发生变化外，个体本身的生长和成熟也是使其行为变化的重要因素。特别是在以幼小的儿童为被试者而又采用单组前测后测实验的情况下，生长和成熟因素的影响就更大。单组前测后测实验通常是实验处理之前先对被试者的某种行为作一次测量，实验处理后再以同样方法测量一次，两次测量之差即表示实验变量 (即自变量) 产生的效果。很明显，这种设计忽略了前后两次测量之间被试者的生长和成熟因素，其实验效果易受生长和成熟因素的混淆，从而降低了内部效度。解决的主要办法是增设同样条件的控制组进行比较。

2. 前测的影响 在一般正常情况下，前后两次测量的结果会有一些的差异，后测的分数将比前测的高。这中间包括练习因素、临场经验、以及对实

验目的的敏感程度，从而提高了后测的成绩。特别是前后两次测量时间较近，这一因素的影响就更显著。

3. 被试者的选择偏性 在对被试者进行分组时，如果没有用随机取样和随机分配的方法，在实验处理之前，他们在各方面并不相等或有偏性，从而造成实验结果的混淆，降低了内部效度。

4. 被试者的缺失 如果是一项长期的实验，要保持原实验被试者的人数不变是相当困难的。即使开始参加实验的被试者样本是经过随机取样和随机分配的，但由于被试者的中途缺失，常常使缺失后的被试者样本难以代表原来的样本。这就降低了内部效度。

5. 实验程序的不一致 在实验过程中，实验仪器、控制方式的不一致，测量程度的变化，实验处理的扩散和交流等都可能混淆实验变量（即自变量）的效果。实验者知道实验目的所产生的“实验者效应”以及被试者知道实验目的或其自己正被研究所产生的“霍桑效应”和“安慰剂效应”等都将混淆实验变量（即自变量）的效果，从而降低了内部效度。

6. 统计回归 将实验结果进行统计回归（statistical regression）后，可能使实验者对实验变量（即自变量）的效果产生误解。如果选择具有极端特性的个体作为被试者（如高焦虑组和低焦虑组）进行实验时，尤需注意这种现象。统计回归现象是，第一次测量平均值偏高者，第二次测量平均值有趋低的倾向（向常态分布的平均数回归）；第一次测量平均值偏低者，第二次测量平均值有趋高的倾向（也向常态分布的平均数回归）。因此第二次测量虽在实验处理之后，其升高或降低只是受统计回归的影响，可能并非是实验变量（即自变量）所产生的效果。

（二）影响实验外部效度的因素

实验的外部效度是指实验结果能够普遍推论到样本的总体和其他同类现象中去的程度，即实验结果的普遍代表性和适用性。以人的行为为对象所获得的实验结果，其推论法往往有相当的局限性。实验的外部效度主要受下列三方面的影响：

1. 实验环境的人为性 实验是在控制条件下进行的，实验环境的人为性可能使某些实验结果难以用来解释日常生活中的行为现象。实验室中的仪器设备会影响被试者的典型行为。被试者参与实验的动机也会影响其行为表现。而在实验室之外的日常生活中，就不会有这些因素的影响。因此，实验结果还不能完全等同于实验室之外的日常行为现象。

2. 被试者样本缺乏代表性 从理论上讲，从事于实验的被试者必须具有代表性、必须从将来预期推论、解释同类行为现象的总体中进行随机取样。但实际上这是很难做到的。因为，如果总体很大，即使能够随机取样，但心理学实验的被试者通常是自愿的，所以也很难把被随机选上的人全都请来做实验。如果总体是无限的（例如，“七岁儿童”就是一个无限的总体，其包括过去的、现在的、将来的所有七岁儿童），随机取样实际上是行不通的。这样的实验结果自然会降低其外部效度。

3. 测量工具的局限性 实验者对实验变量（即自变量）和反应变量的操作性定义往往以所使用的测量工具的测量结果来加以考虑的。例如，把成就动机作为一个因变量，实验者常以某种成就动机量表所测得的分数来界定并评定其强度。但成就动机的测量工具有各种不同的形式，所测量出的分数并不代表同一种成就动机及其强度。如果在实验时采用的是某一种成就动机的

量表，那么所得出的实验结果便不能推论到采用其他成就动机的量表的情况中去。

实验的内部效度和外部效度是相互联系、相互影响的。提高实验内部效度的措施可能会降低其外部效度，而提高实验外部效度的措施又可能会降低其内部效度。这两种效度的相对重要性，主要取决于实验的目的和实验的要求。一般而言，在实验中控制额外变量的程度越大，则对因果关系的测量就越有效。因此，可以在保证实验内部效度的前提下，采取适当措施以提高外部效度。

第二节 心理实验的程序

实验的程序 (experimental procedure) 就是实验的进程。它是指实验在各个阶段应做的事。怎样合理地安排实验程序是心理学实验的重要内容。心理学实验研究是一种创造性活动,它没有一成不变的法则,不能把科学研究看成是科学家简单地在奉行某种常规活动。尽管如此,从大量的心理学文献中,我们仍然能够分析出心理学实验通常要遵循的基本程序。这基本程序是:(1)课题的确定;(2)被试者的选择;(3)实验因素的控制;(4)实验资料的收集、分析和综合;(5)撰写实验报告。

本节将概要地阐明上述五个步骤中所遇到的问题和各种可能的选择。

一、课题的确定

(一) 课题的来源

一项科学研究总是从发现问题开始的。选择课题是科学研究的第一步。学生和有经验的研究者一样,可以从不同的来源得到启发,提出研究课题。那么,课题从何而来呢?研究课题的来源通常有以下四个方面:

1. 实际需要 实际的工作中存在着许多问题需要实验研究来解决。例如,从事数学教学改革时,发现小学生学四则运算题有困难,就提出了小学生能不能学代数?小学生学代数是否比学四则题容易?学四则题对儿童的思维发展有什么影响?诸如此类的实际问题,都可以通过实验研究获得明确的回答。

2. 理论需要 从理论或学说中推演出的某个假设是否符合实际,这就需要实验来检验。例如,学习以后不再复习,遗忘量通常随时间的增加而增加。为了说明遗忘的原因曾提出过一种干扰说,根据这个理论来推论,在学习以后,回忆以前,如其他条件相同,则插入学习的材料越多,对原来学习材料的记忆影响就越大。根据这个推论就可设计《学习额外材料的数量对原来材料记忆影响的实验研究》等课题。

3. 个人经验 在学习、工作和日常生活中,经常会遇到一些心理学问题,例如,为什么“人芝兰之室,久而不闻其香”?50赫兹的电灯光为什么看不见其闪烁?高压水银灯下的红花为什么看上去是黑色的?为什么人能看到秒针走,却看不出分针在移动?“江山易改,禀性难移”的说法对吗?千千万万的人有千千万万的心理现象,所以人的心理现象就像万花筒一样,纷繁复杂而又绚丽多彩。针对这些实际问题就能设计出种种实验。

4. 前人的研究与文献资料 实验研究虽然不一定要从文献出发,但是,在确立课题前,系统地查阅有关文献是必要的。阅读文献可以发现什么问题已经解决、什么问题尚待研究。例如,心理学家(许淑莲、杨治良等,1984)看到国内三例有关先天性无痛症的报导,于是就检索国内外有关研究和文献,并未发现心理学上有这项研究,这样,想到了从心理学的角度研究无痛儿。其中有一位女孩,当时她只有十岁,是个三年级的学生。她看上去模样端正、活泼可爱。对她进行智力测验,结果属于正常偏低。对她情绪特点进行调查,发现她对一般恐惧的反应与正常的儿童一样,如怕看电影里的紧张恐怖镜头,看见小狗、小毛虫也会害怕。但对与身体损伤有关的恐惧与正常儿童不同。这些研究进一步揭示了先天性无痛儿的心理特点。

(二) 实验类型的确定

课题虽然有不同的来源，但是，它们都是从提出问题开始的。对于“为什么”科学的探索，大致可分为两阶段或两个类型。第一阶段是探明规定某个行为的条件，第二个阶段是探明哪些条件与行为之间的函数关系。与这两个阶段相对应，可以把实验分为两种类型。

第一种类型是因素型实验（factorial type experiment），即探求规定行为的条件“是什么”的“什么型实验”，或是探明行为的规定要因的实验。在因素型实验里，逐个地除去、破坏或变化被看作是行为规定要因的几个条件，根据有无相应的行为变化，探明它是否是行为的规定要因。毫无疑问，这时候，对于被操作的条件之外的条件，都应当进行严密的控制。武德沃斯和施洛斯贝格根据他们的分类方法，把因素型实验看作是定性实验。

第二种类型是函数型实验（functional type experiment），即探求各种条件是“怎样”规定行为的“怎样型实验”，或是探明条件和行为之间的函数关系的实验。在函数型实验里，根据因素型实验的结果，系统地、分阶段地变化规定要因的条件，以进行确定条件和行为之间的函数关系的函数型实验，以找出行为的法则。武德沃斯和施洛斯贝格分类中的定量实验就相当于函数型实验。

如果将因素型实验和函数型实验做一比较，可以认为，因素型实验是函数型实验的前一阶段，具有函数型实验的预备实验的性质。当然，在不少实际的研究中，是将因素型实验和函数型实验作为一个实验来进行研究的。一般地说，根据以前的研究，在规定要因已被探明时，多半是直接进行函数型实验，这也是许多实验的基本过程。

由此可见，实验在提出课题的同时，还应明确所探求的问题是属于这两个阶段中的哪个阶段，即是属于因素型实验呢，还是属于函数型实验。

（三）问题的陈述

在课题确定及其所属的实验类型均明确之后，若能以假设的形式提出，那就更符合科学原则。假设是关于条件和行为之间的关系陈述。如果把对条件的叙述记为 a，把对行为的叙述记为 b，一般取“如果 a，那么 b”这样的形式。一切科学定律、法则虽然表面上不一定都符合这个形式，但实际上却包含先行条件（自变量）和后继条件（因变量）这样的逻辑关系。例如，对于缪勒-莱尔错觉（Müller-Lyer illusion）来说，“若变化夹角的大小，则视错觉就有变化”这一假设，是确定夹角是否是视错觉的要因这一因素型实验的假设。这是假设陈述的第一种方式。

假设的另一种陈述方式是用函数关系来表示。它用方程式 $b=f(a)$ 来表示自变量 a 与因变量 b 共变的函数关系，这个方程式读作 b 为 a 的函数，或 b 数量地依存于 a。应用这个模型，就可将上述关系改为：“视错觉的量与夹角的余弦成正比”。这就是函数型的假设。

二、被试者的选择

心理学实验的被试者是多种多样的，有正常人，也有疾病患者，有各种年龄段的人，也有动物，如鼠、狗、猴等。不同的实验对象常常要求不同的实验操作。因此，在制定研究计划之前，应确定研究中所用的被试者。涉及被试者选择的问题有：（1）是使用人类被试者还是使用非人类被试者？（2）被试者应具备哪些机体特征？（3）用哪一种取样方法才能使被试者样本代表总体？这些问题的解答主要取决于二个因素：即课题的性质及研究结果的概括程度。

（一）依问题的性质选择被试者

选用人类被试者还是非人类被试者，依据课题的性质而定。许多心理学研究选用人类被试者，因为它关心的是人类的心理和行为。有损被试者身心健康的实验就不能选用人类被试者，而应当考虑选用适当的非人类被试者。例如，拥挤度对本能行为的影响实验，这样的实验就不能选用人类被试者，而只能选用动物被试者。有的心理学实验使用非人类被试者是因为考虑到要严密地控制无关变量。例如，用同胎的幼小个体或限制它们的活动以削弱个体差异，然后进行实验研究。在人类被试者身上就不可能做到这样的控制，也有不少研究是先进行动物实验，然后在此基础上再进行人类实验。

（二）依研究结果的概括程度来选择被试者

在从事一项研究时必须依据研究结果的概括程度来选择被试者。心理学研究的群体可能是一个小群体，或仅具有某种特性的成员。例如，所有的生物、世上所有的人、某个国家的所有的人，某个民族的所有的人、所有的大学生、某个年龄组的所有成员、某个学校的学生……都可以作为心理学研究的总体。因此，选定什么样的被试者样本，要依研究的问题和据此而推论的全体而定。如果被试者的选择出现偏差，就会影响实验效度，用什么方法能减少这种偏差呢？简言之，可用如下二种方法：（1）随机抽样法。这是最基本的方法，实验用的被试者是随机抽选出来的。每个个体从总体中被选出的机会是均等的，任何个体的选择与其他个体的选择没有牵连，彼此之间的选择都是独立的。（2）分层随机取样法。当总体有不同大小的小组和层次组成时，分层方法最适用。例如，研究中国成人的听力与外国成人的听力是否有所差异，就必须用一定数量的被试者，他们最好是来自成人中不同的年龄段、不同的性别、不同的职业、来自国内不同的地区、甚至不同的民族。这样得到的结果才能代表中国成人的听力。

三、实验因素的控制

在实验过程中，对自变量、因变量和控制变量都要进行控制。有关控制变量的控制，已在第一节中作了讨论，这里讨论对自变量和因变量的控制。

（一）自变量的控制

在实验中对自变量的操纵、变化称为自变量的控制。对自变量控制的好坏，直接影响实验的成败。

对自变量的控制，首先要对自变量进行严格的规定，对心理学中一些含混不清的变量必须使之操作定义化，只有这样才能进行实验。那么什么叫操作定义呢？操作定义（operational definition）是由美国物理学家布里奇曼（Bridgman, 1972）提出的，他主张一个概念应由测定它的程序来下定义。操作定义的提出受到心理学界的欢迎。在心理学上，对一个心理现象根据测定它们的程序下定义就叫操作定义。例如，把“刚刚感受到”定义为“50%次感受到”，就可测定感觉阈限了。又如，疲倦（fatigue）没有一个共同的起点和尺度，怎么测量呢？如果定义为“工作效率的下降”，那么就可以进行测量和比较了。实验者可据此操纵这个变量了。因此，对一些含混不清的变量，一定要有操作定义。

其次，对于在刺激维度上连续变化的自变量，要做好三项工作：（1）要选一定数量的检查点，以找出自变量和因变量的函数关系。如果两者是线性关系，一般三至五个点就可以了。如果函数关系比较复杂，则至少要选五个检查点。（2）要确定好自变量的范围，对自变量范围的确定，有时前人的研

究可以提供线索，如在暗适应的研究中，一般过程在 0~60 分钟范围内。再如在两点阈的研究中，别人已对人体各部位的阈值做过测定，可作借鉴。若在无前人的经验时，就要通过预备实验来确定。(3) 要确定好各检查点之间的间距。间距的大小虽然和自变量的范围和检查点的数目有关，但还得根据实际情况而定。如果自变量和因变量的关系是接近于对数函数，则间距应按对数单位变化。

(二) 反应的控制

在实验目的明确、被试者已选定、刺激变量及操作也规定好了之后，如果把控制变量的控制方法也确定了，那么接下来的问题就是反应的控制了。可以设想，对于一个刺激，被试者个体所进行的、或能形成的反应种类是无限的。例如，被试者在注视一个视错觉图形时，有时把这一刺激作为整体而观察，也有时只注视特定的一部分。显然，这就会导致不同的实验结果。于是，把实验中的被试者的反应控制在主试者所设想的方向上，这就是反应的控制问题。

以人作被试者，往往用指示语来控制被试者的反应，指示语乃是心理实验中主试者给被试者交代任务时说的话。使用指导语时，应注意在允许的范围内做到引起动机，激发兴趣。被试者来到实验室时，不一定对参加实验感兴趣。因此主试者必须利用言词来引起他们的兴趣。在可能的范围内，告诉他们实验目的与应用价值，使他们认识到参与和合作的意义。

总之，指示语的内容与语调都可能影响实验结果。近年来，指示语已成为一个重要的自变量，广泛地得到心理学家们的重视。内隐学习的创始人雷伯(Reber, 1976)及其同事们进行坚持不懈的研究，积累了丰富的经验，获得了许多成果。他们的做法是：在内隐学习实验的学习阶段，向被试者呈现一系列由一套特殊规则产生的字母串，这些规则构成了指定各字母顺序的一种人工语法。给被试者的指示语，或是“学习并记住这些字串”(内隐学习)，或是“试发现这些字串的内在规则，以便更好地记住这些字串”(外显学习)。换句话说，记忆指示语产生内隐学习条件，而规则发现指示语产生外显学习条件。在实验的测试阶段，向被试者呈现一些新的字母串，让被试者判断有哪些符合学习阶段所呈现的刺激的规则，又有哪些不符合。这种用指示语的新方法导致了记忆研究的新发现，即除有意识的外显记忆(explicit memory)之外，还存在着另一个相对独立的记忆系统——内隐记忆(implicit memory)。

(三) 反应指标的选择

在心理学实验中，一般常用的指标有：绝对阈限、差别阈限(以上见第四章第一节)、反应时(见第三章第一节)、反应持续时间、反应程度、完成量、错误率(上见第三章第一节)、完成一定的作业所需要的时间、达到一定基准所需要的次数以及口头报告等等。这么多常用指标，如何作出选择是由选择指标的条件决定的。选择指标的条件很多，主要有：

1. 有效性 即是指标充分代表当时的现象或过程的程度，也称为效度(validity)。哪一个指标最能充分代表当时的现象或过程，那么这个指标的有效性最高。选择任何指标首先要考虑其有效性，如果效度不高，指标就无用。反应指标的效度直接关系到实验的效度。为了使所用的指标具有较高的效度，应了解指标本身的意义是什么、此指标的变化意味着什么、利用此指标对所研究的现象最多能了解到什么程度、有何局限性、如何补救。只有

这样才能全面考虑，才能选择好指标。例如，大多数读者会注意到，地平线附近的月亮比天顶的月亮看上去大；落日时的太阳比正午时的太阳看上去大。其实，月亮和太阳的大小无论在何时都是一样的。因此，这一现象是一种视错觉，心理学上称之为月亮错觉(moon illusion)。雷曼(Reiman, 1920)发表了月亮错觉的实验结果。即落日时所看到太阳的大小是正午时看到太阳大小的3.32倍，而波林(Edwin Garrigues Boring, 1886~1968)和赫威(A.H. Holway, 1940)得出的错觉量为1.7~1.9倍，考夫曼和拉克(Kaufman & Rock, 1962)得出的错觉量为1.5倍。相比之下，雷曼实验的错觉量要大得多。雷曼测定月亮错觉的方法是，在正对太阳的方向上呈现一定大小的圆板，挪动圆板的距离，使之看上去与太阳的大小相等，分别求出落日时和正午时圆板的观察距离，根据这两个距离之比，算出落日时和正午时的太阳外观的大小之比。即，圆板与落日时的太阳的大小相等时，圆板的观察距离为16.6米；圆板与正午时的太阳的大小相等时，圆板的观察距离为5米，则正午时和落日时的太阳外观大小之比为1:3.32。但是，外观的大小并不是观察距离简单的一次函数，这从过去大小恒常性的许多研究结果来看是很明确的。因此，雷曼实验中所选择的指标的有效性不高。另外，如果用望远镜式的纸筒，挡去一切背景来看太阳，这时月亮错觉不存在了，相反会觉得落日时的太阳其大小看上去与当空时的太阳是一样的。这一例子说明了在选择反应指标时考虑其有效性的必要。

2. 客观性 是指此指标是客观存在的，是可以通过一定的方法观察到的。反应时、反应频率、完成量等都是客观存在的指标，是可以客观的方法测量和记录下来的。一个客观的指标一定能在一定的条件下重现。这样的指标能经得起检验，并能够重复进行实验，验证结果。

3. 数量化 指标能数量化，也就便于记录、便于统计，并且量化的指标就能进行比较。

除了上述三个条件之外，还有指标的可靠性和可辨别性等。但是，在这几个选择指标的条件中，有效性是最重要的。

在选择指标时，还要考虑技术设备的条件和可能性。如用脑电波来研究高级神经活动，这是有效的。但是，没有脑电波设备就无法以此作为指标。另外要注意：在测定、记录、观察反应的时候，常常会有仪器误差，操作者的记录误差等，这都要尽力防止。

(四) 仪器和材料的使用

心理学研究使用仪器和材料的主要目的是：呈现实验处理、控制环境、观察行为、及处理观察资料。可把心理学中的仪器材料分为以下四种类型：

1. 呈现刺激的仪器和材料 经常使用的有速示器，时距控制器，记忆鼓，色轮，闪光融合仪，棒框仪，镜画器，深度刺激仪，及用于各种测验的量表和罗夏墨迹测验图(以上见第十一章)等等。

2. 侦察或记录反应的仪器或材料 常用的有记数器，反应时仪，示波器，多道扫描仪，眼动仪(以上见第十一章)，以及各种形式的记录纸等。

3. 控制环境的装置 经常使用的如暗室、隔音室、屏蔽室、单向玻璃装置、闭路电视、眼罩等。

4. 电子计算机 其在心理学上是一种具有多种用途的仪器设备，既可作呈现刺激和侦察或记录反应的装置，也可进行模拟实验和统计处理等。

研究者可根据课题的需要和实际的可能选择使用仪器和材料。在选择使

用仪器和材料时，应以经济实用为原则，不要片面地追求高精尖。实际上，心理学实验，特别是演示实验，常常用简单的材料就可以进行。这就犹如医学上的对症下药一样。

心理学研究必须在良好的环境下进行。这就应当做到：实验室内空气新鲜，温度正常，光线自然而充足，闲人免入，实验时不得闲谈和吸烟等。这些都是最基本的条件。

四、实验资料的收集、分析和综合

实验开始后，研究者的任务是收集资料。由于心理现象的复杂性，在心理学研究中准确无误地收集资料往往是十分困难的。这就更需要研究者具备敏锐的洞察力，严谨的科学态度，对观察资料进行全面而细致的记录。

（一）心理学研究的资料类别

我国心理学家黄希庭等人将心理学研究所能收集到的资料大致分为如下四类：

1. 计数资料 (enumeration data) 就是按个体的某一属性或某一反应属性进行分类计数的资料。这种资料只反映个体间有质的不同，而没有量的差别。例如，被试者的男或女、成年或未成年；反应的有或无、对或错；等等。

2. 计量资料 (measurement data) 就是用测量所得到的数值的大小来表示的资料。例如，被试者的年龄(岁)、体重(斤)、脉搏(次/分)、反应频率(次/秒)、智商(IQ)，等等。

3. 等级资料 (ranked data) 介于计数资料和计量资料之间，可称为半计量资料。心理量表法所取得的资料都属于这类资料。例如，被试者的领导能力可划分为强、中、弱；某个团体对某项改革措施的态度可分为强烈赞成、赞成、无意见、反对、强烈反对，等等。

上述三种资料统称为数据。根据研究需要，上述三种资料在一定条件下是可以互相转化的。

4. 描述性资料 (descriptive data) 即非数量化的资料。在心理学研究中，数量化的资料固然重要，但描述性资料也同样重要。少数初学心理学实验的学生有一种片面的观点，他们只重视数据，实验时只记几个数目字，而对于实验的条件、过程、观察到的现象、及各种影响都不作记录。为了说明问题，不仅要有数量化资料，而且还要有描述性资料。描述性资料可以补充说明数据，使数据更有说服力。但描述性资料由于没有数量指标作为客观尺度，在对它进行解释时容易产生主观片面的错误。因此对它的解释务必更加小心谨慎。

（二）研究资料的整理与解释

实验和观察到的数据应当用统计方法加以整理并制成图表。在统计表中则要列出均值、极差、标准差和标准误等数值。如果研究是在实验组和对照组、实验组和实验组之间进行对比，则还需要注明其 t 值、F 值、和可信限的 P 值，等等。实验结果的分析，主要是应用统计法、通过计算和检验来求得问题的答案或求得对假设的证实或否定。在应用统计技术时，有两点要特别注意：

1. 选用某种统计方法都有其前提条件，不能不分青红皂白，随手拿来就用。例如，t 检验与 F 检验都假定资料的数值具有常态分布。因此，在选用这些方法之前应当先检查资料是否能满足其前提条件。如果不能满足，就应

改用其他合适的方法，或不勉强做定量分析。现代数学为我们提供了十分丰富的统计技术。

2. 根据实验所得到的资料，在验证实验的假设时，不外有两种结果：(1) 如果实验结果与假设相一致，则进一步演绎而找出更深的假设，推进实验；(2) 如果实验结果与假设不一致，则修正假设，反复实验；或放弃它；或建立新的假设，再进行实验。

统计分析之后的工作是根据结果的分析对问题做出结论，判断实验前作的假设能否成立。在说明结果时，可以将前人有关的研究与本实验的发现进行比较，以发现其异同。在作结论时，要实事求是，只能根据在本实验的条件下就所得到的资料和事实作出判断，不能做过分的推论。如果测验结果具有广泛的含义，要谨慎地指出。即使观察到两个变量有高度的相关性，也不能过于肯定它们之间有因果关系。因此，在得出结论时，必须时刻牢记要“敏于事而慎于言”，客观而真实地反映研究结果。

五、撰写实验报告

心理学实验完成之后，就必须把实验报告写好，实验报告是总结科研成果的一种形式。与其他学科一样，心理学的实验报告是对过去工作的总结，更重要的是能为进一步的研究提供线索。学习写好实验报告也是实验课的教学任务之一。写心理实验报告，一方面要完整全面地阐述实验进行的情况；另一方面又要写得简洁明了。

心理学实验课中所写的实验报告与科学研究的实验报告基本项目是相同的，但也有不同之处。科学研究的目的是要解决新问题，在科学研究的实验报告中必须提供新的研究成果，而在实验课中做实验是为了学习，常常是重复一个经典实验或验证某个已有定论的问题，因而在实验课的实验报告中往往只能提供前人研究成果的补充材料。但因为实验是在新情况下的重复，在结果中包含有新的因素，因此也需要整理、分析，写出高水平的实验报告。当然，实验研究的问题的范围很广，解决问题的方法各不相同。因此，在写实验报告时可能会有些差别，但基本的形式和要求是各类实验报告都必须具备的。

一个完整的实验报告，必须包括以下几项内容：摘要、题目、引言、方法、结果、讨论、结论、参考文献及附录。现具体分述如下：

1. 摘要 正式发表的科研报告，一般应写出论文摘要（abstract），把它放在正文的前面。论文提要应当以最概括、最简洁的语言写出，内容包括本课题所要解决的问题、方法、以及获得的结果和结论。如果论文在供国际交流的《心理学报》、《本土心理学研究》、《心理科学》等高水平的杂志上发表，则还应写出外文摘要附于论文的最后。近几年来，随着国际上联机检索文献的发展，目前一般在论文摘要的后面，还要列出论文中的关键词（key words），以便于学术交流。

2. 题目 题目（title）是说明要做的实验研究是属于哪方面的问题。一般要求在题目中既要指出自变量，也要指出因变量。例如，“照度对视觉敏度的影响”，在这一题目中，“照度”是自变量，“视觉敏度”是因变量。这样，只要一看题目就知道这个实验的总轮廓了。

3. 引言 在引言（foreword）中一般要求说明此实验的意义以及题目产生的过程，提出问题的背景材料或提出问题的假设，最好能引经据典，把这类实验的来龙去脉指出来。当然，语句要简练，一般不超过一千字。

一般来说，问题来源可从如下三个途径指明：

(1) 为了扩展以前的工作，或探讨过去尚未解决的问题，在引言中要把以前的工作简要地作介绍，以便与本实验进行衔接。

(2) 若题目来自对以某一理论为根据而提出的假设的论证，则在引言中对这一理论的内容和背景、及假设的由来要解释清楚。

(3) 若题目来自实际部门提出的问题，则在引言中就要对此实际问题进行介绍。

4. 方法 方法 (method) 主要在说明取得实验结果的实验设计，如实验设计有独到之处，或有创意，则应多花些笔墨，介绍这些对传统方法有突破的新方法。另外，在方法中对技术路线的介绍是必不可少的，其中包括：

(1) 被试者：要说明被试者选择的方式，被试者的年龄、性别，及其他有关方面的情况，被试者的数目、及如何进行分组的，等等。

(2) 仪器、材料：实验所用的仪器及材料的名称要一一写上，必要时注明仪器的型号。有时同类的仪器，型号不同，实验结果也可能不同。

(3) 实验程序：即具体说明实验是如何进行的，进行实验的原则、方法步骤、指示语是什么、要控制什么条件，等等。这部分要写得清楚、确切，以便他人可随时照此重复实验、验证这个实验。

5. 结果 结果 (results) 主要是指统计结果，也就是原始材料经过统计后，以图、表形式表示出此实验的结果。另外，把观察结果的记录及被试者的口头报告也列在这部分中。必须指出，在结果中所列的全部内容，必须都是来自本实验的，既不能任意修改或增减，也不要加入自己的主观见解，要使读者清楚地了解这个实验的客观结果。有不少实验报告撰写者把结果看作仅仅是图表的罗列，这也是片面的。对每一幅图、每一张表，一般都应配以文字解释。例如，说明曲线的走向，P 值的显著性水平。当然，大篇幅的分析和讨论应放在“讨论”这一栏目下。

6. 讨论 讨论 (discussion) 是根据实验结果对所要解决的问题给予回答，并指出假设是否可靠。如果结果不能充分地说明问题或者各部分有矛盾时，就要进行分析，找出原因。如果结果与别人的结果不一致时，可以进行讨论，提出自己的见解。当实验得到意外的结果时，也要进行分析，不能弃之不管。因为意外的结果有时会有意外的发现。在讨论中，还可以对本实验的程序、所用的仪器、及进一步的研究提出修改意见和建议。对于方法上有突破的实验，可以将方法单列一点，进行方法学上的专论。

7. 结论 结论 (summary) 是说明本实验证实了或否定了什么假设。结论一般以条文形式、用简短的语句表达出来。但是，结果必须恰如其分，不可夸大，也不必缩小，一定要以本实验所得的结果为依据，确切地反映整个实验的收获，切忌把话说过头。

8. 参考文献 要把参考文献 (references) 的题目、出处、作者、出版日期都写明，以便查找。文献的顺序一般在文章中出现的先后为序，也可按出版社的规定。

9. 附录 学生型的实验报告一般要把全部的原始资料都列入附录 (appendix)。因为同一结果，不同的人，在不同的时候都可以进行不同的分析和处理。另外，重要的实验材料、指示语等也应列入附录。

上述各项仅是为了便于阐述和讨论才把一篇科研报告解剖成几个部分。实际上，由于研究的题材和内容的不同，上述各项完全可以、而且应当根据

需要有所增删，也可以把其中的某些项合在一起写。但实验报告的基本思路 and 结构，一般都包括上述各项内容，同时也要按上述顺序行文。

第三节 实验心理学简史

心理学成为一门独立的科学不是一蹴而就的，它和物理学、生理学等学科一样，也有漫长而艰难的发展历程。心理学的形成和发展无不与实验心理学联系在一起。实验心理学在整个科学史上，尤其在心理学史上占有很重要的地位。它建立于十九世纪下半叶，开始于应用自然科学的方法研究心理学问题。它的诞生不仅使心理学获得了收集材料的新手段，而且使心理学建立在精确可靠的实验基础上。从实验心理学的建立到目前仅一百年的时间，它的发展是十分迅速的。

心理学是一门古老而又年轻的科学，也就是说，心理学自古有之，《论语》、《孟子》等古书中就曾论及心理学问题，荀子就是中国古代一位伟大的心理学家，他论及到身心关系、心物关系等一系列的心理问题。古希腊亚里斯多德对心理学贡献较大，他当时就提出诸如感觉、记忆、认识、情感、联想、统觉等问题。

实验心理学的发展，归纳起来大致经历了如下两个阶段：

一、心理实验发展的第一阶段

大约从 18 世纪中叶到 19 世纪中叶是心理实验发展的第一阶段，也可说是实验心理学的预备阶段。从 18 世纪中叶开始，生理学家、物理学家和天文学家研究了一些感觉和知觉范围内的问题。例如，反应时的测定、后象的延续时间、某些差别阈限的测定，等等。他们把自然科学的研究方法逐渐引入心理学的研究中，虽然当时应用的方法、技术比较简单，类似某些物理实验和生理实验。被试者的自我观察和陈述都带有十分初级的性质。例如，对某一声音听见或听不见，对某一强度的刺激觉察或不觉察，等等。尽管如此，长期被唯理论禁锢的心理学已被冲开了一道缝隙，心理学的研究活跃了许多，科学的心理学正被孕育着，这阶段前后大约经历了一个世纪。在这一段时间里，尽管收集了大量的在心理学范围内非常有趣的实验材料，但对大多数哲学家和自然科学家来说，人的心理是永远也不能用实验的方法来加以研究的。至 19 世纪中叶，随着社会实践、自然科学的迅速发展，与心理学关系密切的生理学有了一系列新的发展成果，例如，缪勒（Georg Elias Müller, 1850 ~ 1934）感官特殊能力的研究，赫尔姆霍茨（Hermann Helmholtz, 1821 ~ 1894）和韦伯等的感官生理的研究，以及人们对人类自身某些规律的认识又深入了一步。所有这一切都为心理学准备了科学的基础知识和研究方法。这时哲学也已为心理学积累了不少心理事实的材料，确定了研究对象。当时在正理学问题的研究上，莱比锡大学解剖学和生理学教授韦伯（Ernst Heinrich Weber, 1795 ~ 1878）首创了实验测量和用数学公式表示感觉的差别阈限（differential threshold），韦伯根据多年的研究结果，总结出关于人的感觉方面的某些规律性的知识，后人称之为韦伯定律（Weber's law）。莱比锡大学物理学教授费希纳（Gustav Theodor Fechner, 1801 ~ 1887）发展了韦伯的工作，他根据当时物理学和数学知识，对感觉阈限问题进行了深入的实验研究和精密的数学论证，提出一个感觉强度与刺激强度的所谓心物关系的对数定律，同时制定了心理物理学（psychophysics）的基本方法。费希纳的心理物理学实验具有两个明显的特点：（1）利用了专门为研究心理物理学制定的实验方法，即极限法、恒定刺激法、及平均差误法（见第四章第二节）。（2）对实验结果作数学处理。费希纳的心理物理学实验的这两个特

点，在心理实验的发展中标志着心理学的研究方法向前迈了重要的一步，这意味着心理学从利用物理学和生理学的实验方法过渡到利用自己的特殊的心理学方法的开始，即标志着心理实验的“心理学化”的开端。多次重复同一试验、并对所得结论作统计处理，意味着心理实验从比较草率的、不精确的实验逐渐向专门的实验心理学过渡。费希纳的《心理物理学纲要》的出版，对以后心理学实验方法的建立起了重要的作用。

心理实验在这第一阶段具有以下特点：（1）应用的方法类似于某些简单的物理学和生理学的实验方法，研究的问题只限于某些简单的心理现象的量的测量，如视觉敏度、正后象的延续时间、差别阈限的测定、反应时间的测量，等等。（2）实验技术简单，对条件控制、及实验结果的数据处理都还有欠缺。（3）被试者的自我观察与陈述都带有十分初级的性质，例如，听见或听不见某一声音、两个刺激的强度差别觉察或不觉察，等等。

二、心理实验发展的第二阶段

心理实验发展的第二阶段始于19世纪60年代，一直延续到今天。开始阶段是以冯特的工作为标志的，这一阶段也可称为实验心理学的创立、传播和发展阶段。1879年，冯特在莱比锡大学建立了第一个正式的心理实验室，还培养了一批各国的心理学家。这对实验心理学的传播起了很重要的作用。对于冯特来说，心理实验就是系统的自我观察，而一切实验手段只是自我观察的辅助手段。在冯特看来，心理学的研究对象是所谓纯粹的“直接经验”，冯特所采用的实验方法仅仅是进行“科学的内省”。冯特还认为，“经验”是由许多心理元素构成的，他希望通过“内省”把“经验”分解为简单的心理元素，诸如感觉和感情，犹如化学中把水分解为氢和氧一样。冯特认为实验心理学的主要任务是在严格控制的自我观察的帮助下精确地分析个体经验。这就意味着，心理功能只有分解成简单的感觉成分才可以放到实验室中去研究。所以，冯特的实验法又称作内省实验法（或内省分析法）（introspective experimental method）。冯特的学生铁钦纳（Edward Bradford Titchener, 1867~1927）发展了其师的“内省实验法”，在内省实验法上，铁钦纳为提高自我观察的效度设定了更多的限制，要求被试者接受更严格的训练，使研究的范围越来越窄。虽然铁钦纳的极端化倾向限制了其体系的发展，但他严格的科学态度对心理学基础研究所作的贡献已被载入史册。

这一阶段心理实验的特点是：在实验中提出了被试者做系统的自我观察计划，被试者起着“观察者”的作用。冯特的心理实验理论也决定了其实验方法在心理学中应用的范围。冯特的兴趣主要在感觉方面，对于思维等“高级的”心理过程，冯特认为是不能用实验方法进行精确研究的，对这些“高级”的心理过程的研究，他提出了特殊的非实验的方法，称之为“民族心理学”的方法。他的学生屈尔佩（Oswald Külpe, 1862~1915）着重进行高级心理研究，并提出了“无表象思维”的理论，1904年在符茨堡大学建立了符茨堡学派（Würzburg School）的实验室，符茨堡学派是以自我观察的实验方法研究思维过程。冯特对此表示坚决反对。

除冯特式心理实验之外还有不少别的种类的心理实验，因素型实验或函数型实验就是其中之一。这些实验的任务不在于精确地分析意识过程，而是企图找出一定现象产生的原因，或是阐明两个变数之间的函数关系。艾宾浩斯（Hermann Ebbinghaus, 1850~1909）所做的许多记忆实验就属于这一类。艾宾浩斯运用严格的实验方法研究记忆，这就突破了心理实验局限于感觉、

知觉的范围，为研究“高级心理过程”的基本实验方法与材料提供了基础。同时，艾宾浩斯的兴趣从主观行为转向客观行为，他不用记忆的主观经验，取而代之的是客观指标——回忆量。艾宾浩斯的记忆研究开拓了高级心理过程实验研究的新时代。

此外，在 19 世纪 90 年代兴起的“测验式实验”或“心理测验”以及“动物心理实验”、“儿童心理实验”等新类型的心理实验，都有别于冯特式的心理实验。这个变化过程有以下几个特点：（1）制定和应用了实验研究的一般方法，注重对实验条件的严格控制；（2）制定和应用了实验研究的特殊方法，如记忆研究法、情绪研究法，等等；（3）广泛应用最新科学技术成就和统计学方法，这类新类型的心理实验的目的是力图得到精确、可靠和客观的实验结果。

到了 20 世纪初，在美国兴起了一个新的心理学学派——行为主义（behaviorism），创始人是华生（John Broadus Watson, 1878 ~ 1958）。华生于 1913 年发表的《一个行为主义者看来的心理学》的文章，是这一学派诞生的标志。行为主义反对冯特把意识当作心理学研究的对象、把内省当作心理学研究的方法，主张心理学研究的对象为行为，认为心理学要走生物科学的道路。华生拒绝把自我观察作为心理学的一种完全的研究方法，因为自我观察至少不能用于研究婴儿和动物，婴儿和动物是不具备报告自我观察的能力的。行为主义认为，人和动物可观察的活动是行为，因而，心理学研究的对象只能是行为。对行为的研究包括刺激和反应两个方面，所以，华生的行为主义也称刺激-反应（S-R）心理学（参见图 1-3）。刺激（stimulus）是指外界环境和身体内部的变化，如光、声、渴、饥等。反应（reaction）是指有机体所做的任何外部动作（外部反应）和腺体分泌（内部反应）。反应有遗传反应和习得反应两种。复杂反应和动作技能是通过建立条件反射学会的。

刺激 反应

信号（输入） 信息加工（讯息处理） 行为（输出）

在中国，行为主义的代表人物是郭任远（1898 ~ 1970），他和华生都否认本能的存生，郭任远毕生从事动物个体行为发展的研究，1921 年在美国发表《取消心理学上的本能说》，批判麦独孤（William McDougall, 1871 ~ 1938）的本能理论，震惊了美国心理学界。他坚信动物的理论只能从解剖生理的现实生活环境和个体发育的历史研究中探索。他对鸡胚胎行为发展的研究和小鸡啄食行为发展的研究证明，小鸡啄食的动作是从鸡胚胎的心脏跳动之际开始发展起来的。他训练猫不吃老鼠的实验研究表明，通过发育时期生活条件的控制，可以使猫亲鼠、护鼠、甚至怕鼠。这项实验也受到国际心理学界的重视。郭任远是国际上有影响力的心理学家，他的著作有《人类的行为》（1923）、《行为学的基础》（1927）、《行为主义》（1934）。在 1946 ~ 1970 年期间，他定居香港，著有《行为发展的动力形成论》。

行为主义在美国统治了四、五十年，到了 50 年代后期，美国心理学中出现了—个新兴的理论方向和研究领域，到 70 年代已成为美国心理学的一个主要方向，这就是认知心理学（cognitive psychology）。认知心理学反对局限于研究孤立的、外观的、可观察的反应，而致力于了解心理活动的过程。这种活动就是传统心理学中称之为认识活动的东西，也就是全部的信息加工（或讯息处理）（information processing）活动。认知心理学把研究的重

点转到内部心理过程上去了。图 1-3 (参见上页) 用模型的方式表示行为主义和认知主义的区别。在 70 年代, 曾任美国心理学会主席的麦基奇 (W.J.McKeachie, 1921 ~) 在一篇报告中说, 心理学是什么, 这一概念已发生改变, 在詹姆斯的经典教科书里, 把心理学定义为“对意识状态的描述和解释”, 后来改为“研究行为的科学”, 今天我们的定义又改变了, 心理学又回到意识上来了。但是, 更多的教科书和心理学家把心理学定义为研究行为和心理过程的科学。这就既包括外观的、可观察的动作, 又包括内部的心理活动。在研究方法上, 行为主义强调严格的实验方法, 排斥一切主观经验的报告, 认知心理学则既注重实验室研究、又重视主观经验的报告。目前, 认知心理学已经成为心理学中的一个重要潮流。可以这样说, 心理实验方法发展的第二阶段, 前期主要受到行为主义的影响, 后期主要受到认知心理学的影响。

从以上心理学实验发展的历史我们可以看出, 由于在心理学中采用了实验法, 才有了对心理现象进行客观研究的手段, 才能从对心理现象的一般哲学推论, 进入到具体心理过程及其物质基础的分析研究, 从而能越来越深入地揭示各种心理活动的规律, 大大丰富了心理科学。从某种意义上说, 正是因为实验心理学和心理物理法的诞生和发展, 才使心理学完全从哲学中分离出来, 成为一门独立的科学。

实验方法的建立, 特别是从 20 世纪 50 年代以来, 心理学已广泛应用了现代科学和工程技术的最新成就。近年来, 随着科学技术的日益发展, 国际上的一些新技术、新概念和新方法已被引进心理学领域中, 如控制论、信息论、微电子技术、分子生物学、数理分析等, 使心理学的实验研究水平有了显著的提高。心理学家从信息加工的角度去看待知觉、学习、记忆及思维等心理过程。心理学与邻近学科有了越来越多的共同语言, 从而更有利于其发展。这使我们认识到, 心理学应注意从邻近学科中汲取有益成果, 这样才有助于逐步揭示心理现象的奥秘。

实验方法的应用, 使心理学的发展大为加快。从心理学实验方法的建立到现在才一百多年, 而心理学的发展却大大超过以往许多世纪。虽然科学的心理学不能归结为实验心理学, 实验方法也不是心理学研究的唯一方法。但是, 任何当代心理学的教科书都以大量的篇幅证明, 现代心理学中的大量事实大都来自先前的实验研究, 心理学实验的诞生和发展有力地说明了实验方法是揭露心理和行为的规律的重要途径和手段。就以武德沃斯 (Robert Sessions Woodworth, 1869 ~ 1962) 等所著的《实验心理学》一书中所列举的参考文献来说, 1938 年的第一版列举了 1770 项, 到 1955 年的第二版就列举了 2480 项, 其中有 50% 是旧版中未曾引用过的。1971 年的第三版的参考文献目录长达 100 页, 不下于 4000 项, 其中 80% 以上是旧版未曾引用过的。它所引用的资料主要来源于美国的材料, 主要也只是涉及心理学基础领域的实验研究。

一位心理学工作者可以对心理学的任一领域或任一分支进行特别研究, 可以专门从事儿童心理、教育心理、医学心理、或知觉心理、思维心理, 以至社会心理的研究。但是, 如果他想成为一个真正严格的、科学的心理工作者, 他就必须要很好地掌握实验心理学的研究内容和方法, 了解应当如何科学地考察心理和行为的规律。

实验研究和任何一种研究方法一样, 有它一定的局限性。这是因为实验

研究是在一定控制条件下进行的，实验结果有时会与现实生活中人们的心理活动不完全一样。这种现象的发生在科学研究中并不少见，更不是不能解决的。只要研究课题来自实际，研究结果又不断到社会实际中去检验；同时，在研究过程中又不仅仅注意心理活动的单因素的实验研究，还注意多因素的交互作用的研究，那么，其研究结果是会与现实生活相一致的。

综上所述，综观心理学研究的历史，可以这样说，实验法是心理学的主要研究方法，实验心理学的诞生意味着科学心理学的建立。它是使心理学成为一门独立科学的基石。当然，它也不是心理学的唯一的研究方法，它和观察法(observation method)、临床法(clinical method)、问卷法(questionnaire method) 等研究方法相辅相成，构成完整的心理学方法学。

本章实验

一、自变量和因变量的确定

(一) 目的：

- (1) 通过动作学习的过程了解心理实验中确定自变量和因变量的方法；
- (2) 学习使用触棒迷津。

(二) 材料：触棒迷津、小棒、遮眼罩、秒表、记录纸。

(三) 程序：

(1) 三人一组，被试者带上遮眼罩，用小棒走迷津（实验前被试者勿看迷津，也勿用手触摸迷津）。主试者对被试者的指示语必须这样说明：“在排除视觉的条件下，尽快地学会用小棒走迷津，中间不要停顿，要积极运用动觉、记忆和思维，争取早些学会”。主试者把小棒放在迷津的入口处，然后让被试者用优势手拿住小棒，手臂悬空。

(2) 被试者手执小棒静候。主试者在每次开始前 2 秒钟，先发出“预备”口令，主试者再说“开始”时，被试者才用小棒走动。在发出“开始”口令的同时，主试者开动秒手表。

(3) 被试者在走迷津的过程中，凡进入盲巷一次就算出错一次，主试者记下错误次数。

(4) 当被试者的小棒进入迷津终点，主试者立刻说“到了”，同时停秒表，记录走一遍迷津所用的时间（秒）。再做下一次的准备工作。

(5) 学习遍数因被试者而异，均以连续三次不出错为学会的标准。

(6) 若被试者在学习途中感到疲劳，可在某次走到终点后休息几分钟。

(四) 结果：将每次走迷津所用的时间和错误次数列成表格（见下页）。

本实验记录用纸

自 变 量 因 变 量	学 习 遍 数															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	...
所需时间 (秒)																
错误次数 (次)																

注：学习遍数因被试者而异，均以连续三次不出错为学会的标准，学习遍数超过 15 次者可自行加行。

(五) 讨论：

(1) 本实验中自变量是什么？为什么在实验前要对所用的自变量提出一个操作上的定义？

(2) 本实验用什么作因变量的指标？它的作用是什么？

(3) 本实验控制了那些变量？

二、自变量的范围和间距

(一) 目的：通过对皮肤两点阈的测定，学习确定心理实验中变量范围

和间距的方法。

(二) 材料：两点阈量规、遮眼罩、记录纸。

(三) 程序：

(1) 主试者选定被试者的左手背(A区)与前臂背面为测量区，前臂背面又分为二个区(B、C区)，本试验取A、C两个区。测量前，在各区内标上蓝印或划出圆圈，刺激在圈内进行。

(2) 在使用两点阈量规时，必须垂直接触皮肤，对两个尖点施力均匀，接触时间不超过2秒钟，先在自己手上练习几次，再在被试者的非测验区练习几次。

(3) 实验序列的长度和起点，可根据初步测验后确定，大致在7~15毫米的范围内取5个点，每步变化在1毫米左右。每种间距做20次，顺序随机安排。

本实验记录用纸一

刺激代号相对应的实际刺激值

代号	1	2	3	4	5	...
刺激值(m)						

(4) 每隔100次，休息5分钟。被试者有时产生一种持久的两点后象(即使只有一点或距离很近的两点刺激，也有感到两点的印象)，此时应给予休息，使其恢复两点的标准。每隔数十次，可插入一次一点刺激，测验被试者是否有这种后象。

(5) 主试者对被试者的指示语必须这样说明：“在排除视觉的条件下，当被试者感觉到是两点时说‘二’；当被试者感觉到一点时说‘一’；当被试者不能确定感觉到的是几点时说‘不’”。

本实验记录用纸二

被试者反应记录表

	A区							C区							
	1	2	3	4	5	...	1	2	3	4	5	...			
被试者报告「两点」或「一点」															

(6) 主试者在每次刺激前先发出预备信号：“注意”。在被试者作出反应后，要记录下来。

(7) 在全部测量完毕后，在记录中挑出有“不”的序列，再重复那些序列，获得确定的判断。

(四) 结果：分别求出左手A区和C区的两点阈限，并对两均数的差别进行考验。

(五) 讨论：

(1) 说明身体部位不同，触觉有差异，两点阈值从一个侧面揭示了触觉感受性的大小。

(2) 如果要测定皮肤某部位的两点阈限，自变量的范围和间距应如何安排？

本章摘要

1. 实验心理学就是在实验控制条件下对心理和行为进行研究工作的心理学。这一心理学分支是以研究方法来定义的。也就是说，只要是用实验法来研究的心理学问题，都可以包括在实验心理学的范围内。

2. 实验是一种控制的观察，与自然的或偶然的观察不同，实验是实验者人为地使现象发生，对产生现象的情景或影响现象的条件加以操纵、变化与控制的观察。

3. 实验具有三个明显的特征：实验者掌握有主动权；实验具有可核对性、或验证性；实验者可以系统地变化条件，观察因这些条件的变化而引起的现象上的变化。

4. 变量是指在数量上或者质量上可以改变的事物的属性。在心理学的实验中，研究者必须考虑三类变量：(1) 已知的、对有机体反应发生影响的变量，叫作自变量，它是由实验者操作、掌握的变量；(2) 由操纵自变量而引起被试者的某种特定的反应，叫因变量，它是由实验者观察或记录的变量；(3) 在实验中应保持恒定的变量，叫控制变量。

5. 心理实验的对象统称为被试者，包括人和动物。在实验中，特别是在以人为被试者的实验中，处理好实验者（即主试者）与被试者的关系是实验取得成功的一个重要条件。

6. 额外变量（即控制变量）是使实验结果发生混淆的主要根源。对额外变量的控制常用的方法有：排除法、恒定法、匹配法、随机化法、抵消平衡法和统计控制法。

7. 实验效度是指实验方法达到实验目的的程度。换言之，实验效度就是实验结果符合客观实际的程度。实验效度主要包括内部效度和外部效度。

8. 实验的程序是实验的进程，它是指实验在各个阶段应做的事。心理学实验通常遵循的基本程序是：(1) 课题的确定；(2) 被试者的选择；(3) 实验因素的控制；(4) 实验资料的收集、分析和综合；(5) 撰写实验报告。

9. 课题的确定是科学研究的第一步，研究课题的来源通常有四个方面：实际的需要，理论，个人的经验，过去的研究与文献。

10. 心理学实验的被试者是多种多样的。有正常人，也有疾病患者；有各种年龄阶段的人，也有动物。不同的实验对象，通常要求有不同的实验操作。

11. 操作定义的提出，受到心理学家们的欢迎。在心理学上，对一个心理现象根据测定其程序下的定义称之为操作定义。

12. 一个完整的实验报告必须包括以下几项内容：摘要、引言、方法、结果、讨论、参考文献及附录。

13. 实验心理学的发展，大约经历了两个阶段。从 18 世纪中叶到 19 世纪中叶是心理实验的第一阶段，也可说是实验心理学的准备阶段。在这一阶段具有以下特点：应用的方法类似于某些简单的物理学和生理学的实验法；研究的问题只限于某些简单心理现象的量的测量；被试者的陈述带有十分初级的性质。

14. 心理实验发展的第二阶段始于 19 世纪 60 年代，一直延续到今天。开始阶段是与冯特的工作相联系的，这一阶段也可称为实验心理学的诞生、传播和发展阶段。心理实验发展的第二阶段，前期主要受到行为主义的影响；后期主要受到认知心理学的影响。

15. 综观心理学研究的历史，可以这样说，实验法是心理学的主要研究方法。实验心理学的诞生意味着科学心理学的建立，它是使心理学成为一门独立科学的基石。当然，它也不是心理学的唯一的研究方法，它和观察法、临床法、问卷等方法相辅相成，构成完整的心理学方法学。

建议参考资料

1. 王甦等（1993）：当代心理学研究。北京市：北京大学出版社。
2. 杨治良（1988）：心理物理学。兰州市：甘肃人民出版社。
3. 张春兴（1991）：现代心理学。台北市：东华书局（繁体字版）。上海市：上海人民出社（1994）（简体字版）。
4. 波林（高觉敷译，1981）：实验心理学史。上海市：商务印书馆。
5. 黄希庭（主编）（1988）：心理实验指导。北京市：人民教育出版社。
6. Atkinson, R.C. et al. (1988). Stevens' handbook of experimental psychology (2nd ed). New York: A Wiley-Interscience Publication.
7. Kantowitz, B.H., & Roediger, H.L. (1984). Experimental psychology (2nd ed.). New York: West Publishing Co.

第二章 实验设计

本章内容细目

第一节 实验设计的基本类型

- 一、被试者内设计 46
 - (一) 实验前后设计
 - (二) 定时系列设计
 - (三) 抵消实验条件的设计
- 二、被试者间设计 51
 - (一) 随机组设计
 - (二) 配对组设计
- 三、混合设计 56

第二节 多变量实验技术

- 一、多自变量实验的优点 58
- 二、多因素实验设计 63
- 三、拉丁方设计 65

第三节 实验数据的统计分析

- 一、统计表和统计图 66
 - (一) 统计表
 - (二) 统计图
- 二、实验数据的初步整理 73
 - (一) 偶然误差与系统误差
 - (二) 集中量
 - (三) 差异量
- 三、显著性检验 82
 - (一) 显著性检验的含义
 - (二) t 检验
 - (三) F 检验
 - (四) X^2 检验

本章摘要建议 参考资料

前面谈到，心理学是一门实验科学。为了完成好一个实验，并使其在方法上能站得住脚，必须首先做好实验设计，这是实验成功的关键。一个好的实验设计能够用最少的人力、物力来获得最多和最有效的实验数据以达到最佳的实验目的。否则，如果实验设计不当、对非实验条件控制不严、实验方法和程序安排不合理，即使实验的规模较大、实验次数较多，但是所得到的实验数据却不能为最初所提出的问题提供合理的答案，达不到原来的实验目的。这样的实验设计不是好的设计。

一个良好的实验设计是实验研究能够作出成果的重要因素，实验设计是

整个实验研究的重要组成部分。有人以为只有实验終了得出数据之后再考虑统计处理，这样想法是片面的。事实上，只有在实验开始之前就有精心的实验设计，才能事半功倍。心理是一种复杂的现象，它跟自然现象和社会现象一样，也有着自身本身的规律性。实验就是揭示这些规律性。但心理科学的实验设计要比其他科学更为困难，这是因为无论是实验对象，还是环境条件，反应测定都要受到很大的限制。因此，掌握好实验设计技术也就更加重要。科学的探索要求不断地发展自己的手段和方法，也就是要不断地创造出新的“工具”。这些新的方法和“工具”，理所当然地能反映在实验设计之中。历史事实表明，方法上的创新能使我们产生意想不到的效果。

顾名思义，实验设计就是对实验的设计。在长期的科学实验中，科学工作者已经总结出一些在实验设计时所必须遵守的共同原则、具体的实验方法和技术上应注意的细节。本章内容主要讨论下面六个问题：

1. 何谓实验设计，它有哪些基本类型。
2. 多自变量实验设计有哪些优点。
3. 统计表有何功能。
4. 统计图有何功能。
5. 标准差的主要用途有哪些。
6. 什么叫显著性检验？在使用显著性检验时应注意什么。

第一节 实验设计的基本类型

实验设计 (experimental design) 乃是进行科学实验前做的具体计划。它主要是控制实验条件和安排实验程序的计划。它的目的在于找出实验条件和实验结果之间的关系, 做出正确的结论, 来检验解决问题的假设。心理实验设计的内容包括: 刺激变量 (或刺激变项) (stimulus variable) 的确定及其呈现的方式, 反应变量 (或反应变项) (response variable) 的指标及其测量方法, 对一切有关变量 (或变项) (variable) 的控制措施, 确定被试者人数和选择被试者的方法, 拟定主试在实验开始前对被试者要说的指示语, 规定实验次数, 安排实验程序, 规定使用仪器的型号, 处理实验数据的方法等等。实验设计要对实验结果有预见性, 要保证严格按照实验设计进行才能取得有效的实验数据。

每一个实验设计都必须回答三个基本问题: (1) 实验采用多少自变量? 例如在一个阅读速度的研究中取“照明强度”为自变量。(2) 各自变量内又采用多少处理水平? 例如照明强度又分为强、中、弱等处理水平。(3) 在各自变量和各处理水平中用相同的被试者, 还是用不同的被试者? 根据这三个条件的组合, 就可构成许多不同类型的实验设计。

一般根据对上述三个问题的回答, 就可把实验设计的类型大体上分为三种: 被试者内设计、被试者间设计以及同时包括被试者内与被试者间的混合设计 (mixed design)。

被试者如果在自变量发生变化的所有情况下接受实验, 则是被试者内设计。例如, 自变量是照明强度, 假如有强、中、弱等三种照度供操纵; 因变量是观测一个在标准条件下视角为一分的 C 型视标的距离。采用被试者内设计时, 同一被试者需要在三种照明条件下都接受测试。

如被试者只接受多个自变量情况中的一个, 即不同的被试者接受不同自变量的处理, 这类设计便是被试者间设计。在上例中, 如实验者令一组被试者在强照度条件下测试, 令另一组被试者接受中等照度条件下的测试, 令第三组被试者接受弱照度条件下的测试, 如此类推, 直到各组都分别接受某一种照度条件下的测试。这种实验设计就是被试者间设计。

为什么在被试者内设计和被试者间设计中, 实验都可以同时操作两个或两个以上的自变量呢? 乍看之下, 这种处理似乎违背了实验法“单一变量”的原则。其实, 由于实验者采用了统计方法, 可以通过平衡多变量的关系, 通过控制其他一些自变量, 使某一阶段变化特性显露出来, 这一问题就迎刃而解了。因此, 可以说统计技术是实验设计的基础。

所谓混合设计, 是指在一项实验中, 有些自变量是被试者内的, 而有些自变量是被试者间的。例如, 一个被试者接受甲变量的每一种情况, 但只接受乙变量的一种情况。根据被试者内设计和被试者间设计的特征, 这时的甲变量是被试者内自变量, 乙变量则是被试者间自变量了。这类实验设计即是混合设计。

一、被试者内设计

被试者内设计 (或受试者内设计、单组实验设计) (within-subjects design) 是每个被试者须接受自变量的所有情况的处理。其基本原理是: 每个被试者参与所有的实验处理, 然后比较相同被试者在不同处理下的行为变化。这种实验设计下的同一被试者既为实验组提供数据, 也为控制

组提供数据。因此，被试者内设计无需另找控制组的被试者。

在实验研究中，如果实验者主要想研究每一个被试者对实验处理所引起的行为上的变化，一般可考虑采用被试者内设计。被试者内设计又可分为三种子类型：

（一）实验前后设计

实验前后设计（experimental before-after design）是指在实验条件处理前对被试者进行观测的结果与实验条件处理后所做的同样观测的结果加以对比的设计。也就是说，这种设计类型是实验（处理）前后的比较设计。我们可用表 2-1 来表示这种设计的模式。

实验（处理）前、后设计，具有二个优点：（1）能较明显地检查出实验处理的效果如何。因为在这种实验中，前、后被试者是同一的，如果能够控制好无关变量的话，那么实验处理前后的差异，就是实验处理的结果。（2）

表 2-1 实验前与实验后设计模式

组数	前观测	处理	后观测
1	Y_1	X	Y_2

说明： Y_1 表示实验处理前对被试者观测所得值

Y_2 表示实验处理后对被试者观测所得值

X 表示实验条件处理

对被试者的需要量较少，一组被试者当二组被试者用，无须再增设被试者控制组。不仅效率提高了，而且被试者变量也得到了较好的控制。

然而，我们应该看到不足之处。各类被试者内设计的共同缺点是，这类设计需要每个被试者在实验中花费很多时间。通常一个被试者被要求执行几项任务，这势必会产生疲劳，影响实验结果。而前、后设计还有二个不足之处：（1）由于前、后两次观测之间存在时间间隔，这就会带来外来影响。一般来说，前、后间隔时间愈长，则影响就愈大。反之，如果前、后两次观测时间很接近，倒可以认为这是实验处理引起的行为差异。（2）容易产生顺序误差。即前面的观测影响后面观测的结果，从而影响实验结果。例如实验中第一次观测会产生学习、疲劳、情绪等效应，从而影响第二次观测的结果。

（二）定时系列设计

定时系列设计（time serial design）是指实验处理前对一组被试者作一系列的定时重复观测，然后实施实验处理。再对被试者作一系列的定时重复观测，分析自变量（实验处理）对因变量的关系。定时系列设计可以看作是前、后设计的扩展形式。定时系列设计，一般先分别求出实验处理前和实验处理后的平均数，也有分别取众数或中数的。我们用表 2-2 来表示这种设计的模式。

表 2-2 定时系列设计模式

组数	前观测	处理	后观测
1	Y_{1a} 、 Y_{1b} 、 Y_{1c}	X	Y_{2a} 、 Y_{2b} 、 Y_{2c}

说明： Y_{1a} 、 Y_{1b} 、 Y_{1c} 分别表示实验处理前之观测值

Y_{2a} 、 Y_{2b} 、 Y_{2c} 分别表示实验处理后之观测值

X 表示实验条件处理

定时系列设计除了具有上述前、后设计的二个优点之外，还具有以下二个优点：（1）降低由于一次观测而得到被试者不正常行为的机率。实验处理前和实验处理后分别对每一位被试者作一系列测定，这就能降低只作一次观测而得到不正确结果的机率。（2）提供测量过程中的信息。无论是在实验处理前，或是在实验处理后，均对被试者进行一系列的观测，这样就可以使我们看出发展趋向。

但是，问题还具有两面性，定时系列设计的这些优点，也加重了前、后设计所存在的弱势：（1）由于更多次的观测，势必延长实验时间，从而会有更多的外来影响。（2）也正是更多次的观测，更易引起顺序误差，更易导致练习、疲劳、紧张或厌烦等效应，影响实验结果。

（三）抵消实验条件的设计

抵消实验条件的设计（reversal experimental condition design）是指抵消实验过程中无关变量的一种设计。前面讲到，有些无关变量在某些实验情况下既不能被消除，又不能保持恒定。例如，单组实验往往由于前一处理影响后一处理的效果，产生顺序误差。为了抵消顺序误差，最简单的方法就是用 ABBA 的排列顺序来安排实验顺序。其模式可用表 2-3 表示。

从表 2-3 中可见，首先给被试者作第一种处理（ X_A ）和第一种处理后的观测（ Y_{2a} ），其次作第二种处理（ X_B ）和处理后的第二次观测（ Y_{2b} ），再重复第二种、最后重复第一种的处理和观测。整个实验程序的安排，就是抵消实验条件的设计的基本形式。

这里不妨举一个例子。有甲乙两种操作方式，我们要比较这两种操作方式哪一种更正确，更快些。我们知道，以人作为被试者的实验，在不同时间所得结果必然有所不同，甚至同一人以同一操作方式作一系列实验，要使其其中两次结果完全相同也是一件不容易的事。定时系列规则是解决此类问题的方法之一。但是定时系列规则不能消除和顺序有关的实验误差。

表 2-3 抵消实验条件设计的模式

处 理	X_A	X_B	X_B	X_A
后观测	Y_{2a}	Y_{2b}	Y_{2c}	Y_{2d}

为了消除这类影响，方法之一就是使用抵消实验条件的设计。本例就可用表 2-4 表示其模式。

表 2-4 抵消实验条件设计的实例

处理	操作方式甲	操作方式乙	操作方式乙	操作方式甲
结果	试验一之观测值	试验二之观测值	试验三之观测值	试验四之观测值

这种实验设计在处理 X_A 的情况下得到两个数值，在处理 X_B 的情况下也得到两个数值。统计时就通过分别求平均数后，加以比较。这样的实验设计，既安排了二种操作方式在程序上的均等，同时又提供了估计随机误差的可能

性。

抵消实验条件的设计具有三个明显的优点：(1)能较好地控制被试者变量，这是单组实验设计的共同优点。单组实验设计不但比两组实验设计用的被试者数量少，而且由于两种实验条件使用同一被试者，从而较好地控制了实验变量。(2)能较好地控制顺序误差。在心理实验中有许多变量与时间有关，例如学习、练习、迁移、挫折、疲劳等等。这些变量特别需要用这类实验设计来消除顺序误差。如果一半被试者所作的顺序为 ABBA，另一半被试者所作的顺序为 BAAB，那就抵消得更彻底。(3)时间上比较经济。这类实验一般观测次数不多，上面操作方式甲乙的比较，也只要通过四次观测就能完成某一被试者的实验。因此，被试者较易配合，以顺利完成实验，同时也排除了某一些实验误差。

抵消实验条件设计虽有上述优点，但也有其局限性。主要有两点：(1)反应变量在时间维度(轴)上的关系是线性时才能使用。抵消实验条件的设计的前提条件是反应变量在时间维度(轴)上的关系是线性关系。例如单纯的前影响作用，无论是从试验一到试验二，还是从试验三到试验四，都是后者受前者的影响。但是，当这些变量的实际效果与行为效果的关系是非线性时，则采用这类设计也不能达到抵消实验的顺序误差的目的。(2)对有些实验不适用。例如用两种学习方法学习同一实验材料就不适用。若一定要用这种设计，就必须改用同一被试用两种学习方法，学习两种难度相等的不同材料，才能消除练习误差，同时还应用相同的计量单位，否则无法进行比较。以上我们介绍了三类单组实验设计，清楚了这些设计所具有的优点。在心理实验中确实有不少运用这种设计方法达到较好效果的例子。例如：杨治良等(1979)曾用单组实验前后设计，对痛的成分进行心理学研究。采用这种设计的出发点是愿接受痛刺激的被试者不可多得，为了弥补单组实验和被试者数量较少的不足，实验采用单组实验前后设计。在这样的设计中，被试者变量能得到较好的控制，因为能消除由被试差异引起的误差，从而对自变量的效果作出更精确的估计。同时，为尽可能减少单组实验中的被试者在实验中花费很多时间，实验采用了心理量表法记分，通过被试者一次反应，记录较多的反应指标，而这些指标之间又存在着很大的一致性(见表2-5)。从表2-5可见，众多的生理和心理指标增加了对照程度，在一定程度上弥补了被试者较少的缺陷。

表 2-5 痛反应量表

级 别	—	二	三	四	五	六			
刺 激 量	t 值	1	2	5.6	8.3	13.9	19.5	2'	
	mA	0.37*	0.74	2.06	3.06	5.13	7.22	1C	
动 作 单 位	兴奋神经 纤维类型	无或少量 小幅度 A A	增大的 A	接近最大的 A	最大 A 小幅度传 导快的 A	最大 A 增大的 A	最大的 A 较大幅度传 导慢的 A		
	传导速度范围 (35)(m/s)	42.8 ~ 55.9	40.0 ~ 55.9	41.5 ~ 56.8	34 ~ 62.4 20 ~ 33.4	35 ~ 56.8 21 ~ 32.9	34.5 ~ 59.4 13 ~ 38.8		
感觉成分(主诉)		触	轻麻	重麻	轻痛	中痛	重痛	极	
情 绪 成 分	主 诉	无所谓	无所谓	无所为	稍有不适	难过、害怕	非常难过 非常害怕	恐 惧	
	外 表 行 为	语言	自如	自如	自如	语调稍变	轻声 叫痛	叫“喔唷”	连 大
		动作	协调	协调	欲动 能控制	头动、 手动	上身动、握 拳、转身	屈肌反射 全身动	
		表情	安详	安详	注意会神	轻皱眉 眨眼	争眉 紧闭目	咬牙 重皱眉	
出汗状况		无	无	无	身体发热	略有	手汗	大	

相当于各级刺激的强度为被试的平均值(采自杨治良, 1979)

应看到单组实验设计有其局限性,同时,正如前面所述,对有些实验不适用。单组实验设计在不能避免某种系统误差时,就要用被试者间设计或混合设计来完善实验设计。

二、被试者间设计

被试者内设计要求每一个被试者接受所有自变量处理,关键是如何安排各处理的先后次序。被试者间设计(或组间法)(between - subjects design)是要求每个被试者(组)只接受一个自变量处理,对另一被试者(组)进行另一种处理,故又称这种设计为独立组设计(independent - groups design)。这种设计的主要问题是决定哪一个被试者(组)接受哪一个实验处理。若有两种以上的处理,有多少种处理就采用多少个被试者(组)。被试者间设计或独立组设计有两类设计技术:随机组设计和配对组设计。

(一)随机组设计

随机组设计(或随机分组设计)(random - groups design)是将被试者随机地分配在不同的组内接受不同的自变量处理。这一假设是将被试者随机分配到不同的组,若对各组用一样的课题,在相同的条件下进行测量,其结果就成为相等组(或等组)(equivalent groups),则它们的成绩在统计上应是相等的。换句话说,假设各组在与实验课题有关的特性上(如年龄、智力、性格特征等)没有差别(在统计允许的限度以内),而实验结果却出现了差别,这差别就是由于处理的不同而引起的。

从统计理论来讲,特别是从抽样理论来说,此设计的各被试者样本(随机组)在未受不同的实验处理前,他们的作业平均数在统计上是没有显著差

异的，如果有，也仅仅是抽样变动。各自样本的平均数 \bar{X} 都是总体平均数 μ 的无偏估计值。换言之，各随机组在未经受不同处理之前是相等的（在统计允许的限度以内）。若这些随机组经受不同的实验处理后，经 t 检验和 F 检验后发现作业的平均数有显著性差异，那么这些差别是实验处理的不同引起的。各实验处理的平均数不是来自总体的平均数 μ ，而是来自各实验处理总体 $\mu_1, \mu_2, \dots, \mu_k$ 。这就是根据随机理论所作的逻辑推理。

怎样才能做到随机分组呢？常用的方法有两种：同时分配法和次第分配法。这两种方法各有其特点，可择宜采用。

1. 同时分配法同时分配的条件是，被试者同时等候，而实验者可随意调派其中任何一个被试者。同时分配法通常有三种技术。

(1) 抽签法：先将所有被试者编号，记入纸片，每一纸片号码代表一个被试者，然后将纸片放入容器内搅匀，按组抽取。若要将 40 个被试者分为四组时，第一次抽 10 片，代表第一组，第二次也抽 10 片，代表第二组，依次类推。假定 40 个被试者中有 16 个女生，则每组应各有 4 名女生。所以可以在 16 名女生中先随机选 4 名，再在 24 名男生中随机选 6 名，归入第一组，依次类推。

(2) 笔划法：若要将 40 个被试者分为四组时，首先将被试者依其姓氏笔划数进行次序排列，再查随机数表每一数列的第一位数，只取第一个数为 1、2、3、和 4 的数字，分别归属四个组，各查 10 个，共查满 40 个以后，按姓氏笔划先后对入 1、2、3、4 所表示的组别。用此法时只要注意在查第一位数为 1、2、3、4 的随机数时各查 10 个即可。

(3) 报数法：若要将 40 个被试者分为四组时，采用类似于体育课上的报数，假定被试者都坐在教室内，实验者令其从第一排报数，报 1 的被试者都被分在第一组、报 2 的被试者都分在第二组，依此类推，只是要注意原有顺序的影响。若每排报数的方向随机改变，例如用 1234, 4321, 2341 等不同的顺序报数，则随机分组的效果将更好。

2. 次第分配法 次第分配法的条件是，由于实验持续时间较长或其他原因，实验者知道有一群被试者，但不知道究竟那位被试者什么时间会来，只能根据预先拟好的原则进行分派，而且当实验结束时，各组要符合随机组的要求。这里介绍两种技术。

(1) 简便法：按被试者出现在实验的先后分派，第一名属第一组，第二名属第二组，第三名属第三组，依此类推等。这有点像同时分配法中的第三例。一般而言，使用本法能满足随机的条件，然而它取决于被试者报到的次序是否符合随机原则。

(2) 区内随机法：为了避免被试者非随机出现的可能性，可按照被试者来到实验室的先后，使用区内随机次序分派被试者归属各组。例如，可根据随机数表来分配被试者。

随机组设计有优点也有缺点。其优点是：(1) 用随机分配被试者的方法可控制两组被试者变量的差异，分组方法简单可行。(2) 由于对每一被试者只作一次观测，可消除某些实验误差，如消除学习误差的影响。这种设计的缺点是：(1) 分成等组的方法仍欠精密。(2) 若两组在不同时期观测，就有可能插入实验以外的偶发事件，影响因变量的观测结果。

(二) 配对组设计

配对组设计（或对等组设计）(matched - groups design) 是随机组设

计的一种逻辑扩展。配对组设计的目的是使各组的特性更加相同。这种设计可控制组内变异和组间变异。在心理学研究中常常会遇到某些变量，特别是机体内部的变量，例如智力、态度……。它们对实验变量（自变量），例如对教材、教法等来说是额外的、无关的，而对因变量来说却是有关的。因此，它们是共变量。共变量若不受到实验控制，则进行 F 检验时作为实验误差的组内误差往往就不合理地扩大了。配对组设计就能解决此类问题。这一设计要求把共变量分成几个等级，经过测量，然后把具有同一等级特征的 K 个被试者加以配对，此时，每一个配对组便是一个层级。配对组的 K 个被试者每人只接受一种实验处理，至于谁应接受哪一种实验处理，则用随机分派的方法来决定。表 2 - 6 乃是配对组设计的基本模式。

表 2 - 6 配对组设计的基本模式

层级	实验变量					层级平均数
	A ₁	A ₂	A ₃	A _K	
1	X ₁₁	X ₁₂	X ₁₃	X _{1K}	$\overline{X_1}$
2	X ₂₁	X ₂₂	X ₂₃	X _{2K}	$\overline{X_2}$
3	X ₃₁	X ₃₂	X ₃₃	X _{3K}	$\overline{X_3}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	X _{n1}	X _{n2}	X _{n3}	X _{nK}	$\overline{X_n}$
实验处理 平均数	$\overline{X_{\cdot 1}}$	$\overline{X_{\cdot 2}}$	$\overline{X_{\cdot 3}}$	$\overline{X_{\cdot K}}$	$\overline{X_{\cdot \cdot}}$

层级	实验变量					层级平均数
	A ₁	A ₂	A ₃	A _K	
1	X ₁₁	X ₁₂	X ₁₃	X _{1K}	$\overline{X_1}$
2	X ₂₁	X ₂₂	X ₂₃	X _{2K}	$\overline{X_2}$
3	X ₃₁	X ₃₂	X ₃₃	X _{3K}	$\overline{X_3}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	X _{n1}	X _{n2}	X _{n3}	X _{nK}	$\overline{X_n}$
实验处理 平均数	$\overline{X_{\cdot 1}}$	$\overline{X_{\cdot 2}}$	$\overline{X_{\cdot 3}}$	$\overline{X_{\cdot K}}$	$\overline{X_{\cdot \cdot}}$

（采自杨治良、乐竟泓，1990）

在表 2 - 6 上，每一等级的每一处理只是观察一个被试者的做法，有所观察的基本单位不是一个被试者，而是一个团体或一个子集合。

配对法常有两个步骤：第一，令所有被试者做“共同作业”，即接受预备测验，获得作业分数；第二，根据作业分数形成配对组。

第一步：共同作业

共同作业（common task）亦称为先检验（或前测）（pretest）。配对组设计的优劣完全依赖实验作业是否与共同作业有高度的相关。先检验与实验作业的相关越高，组间差异越小，则接受不同处理后的实验结果越能反映实验处理的差异。先检验作业有两种：一种是和实验作业有高度相关的其他作业；另一种是利用被试实验作业的初期表现。通常，同一作业两个阶段的行为表现是相关的。很多学习实验以智力测验为配对分组的依据，但在不能

找出和实验作业相关的作业时，对有些无练习效应的实验可使用实验练习的成绩作配对依据。

第二步：配对分组

配对分组 (matched groups) 乃是取得先检验测试分数后，再用两种方法形成配对组。

方法一：将被试者按先检验作业分数的高低排列 (见表 2-7)。现设有三种自变量处理 (情况)，将前三个被试者依区内随机方式派到 A、B、C 三组，接着将后三个被试者如法分配，直到分配完毕。从表 2-8 可以看到，

表 2-7 30 名被试者先检验分数的高低排列表

被试者	分数	被试者	分数	被试者	分数
1	75	11	40	21	27
2	65	12	39	22	25
3	64	13	37	23	21
4	63	14	37	24	19
5	60	15	34	25	19
6	45	16	32	26	16
7	45	17	31	27	15
8	42	18	30	28	13
9	41	19	30	29	10
10	41	20	30	30	2

(采自杨治良、乐竟泓，1990)

按这种方法分配的各组在平均数和标准差上都很接近。

方法二：取先检验分数中相同或非常接近的，以三人为一单位，然后按区内随机法逐次分派到 A、B、C 各组。每三个被试者的分数接近的程度完全是任意的，如表 2-9 的分派限定三个被试者分数差落入 4 的范围内。因为有了这样的限制，超过标准的被试者被剔除。如表中每组被试者只有 8 人，各组分数相差甚微。

表 2-8 30 名被试者按方法一配对分组的结果

被测者 分组	1	2	3	4	5	6	7	8	9	10	\bar{X}	S
A	65	63	42	41	34	31	30	19	19	2	34.6	18.5
B	64	60	41	39	37	32	27	25	15	11	35.0	16.5
C	75	45	45	40	37	30	30	21	16	13	35.2	17.1

说明： \bar{X} 为平均数，S为标准差 (采自杨治良、乐竟泓，1990)

表 2-9 30 名被试者按方法二配对分组的结果

被试者 分组	1	2	3	4	5	6	7	8			\bar{X}	S
A	63	45	41	37	34	30	19	13			35.6	15.7
B	64	45	41	39	31	30	21	15			35.8	15.2
C	65	42	40	37	32	30	19	16			35.1	15.3

(采自杨治良、乐竟泓，1990)

二种方法相比，方法二中各被试者的平均值比方法一更相近，但由于有随机分配的程序因素参与其中，只比较平均数还不能说明问题，应再比较所分各组的标准差。若被试者在方法二中的标准差比在方法一中的更接近，这就有力地说明了方法二的各组较为接近，而根据方法二的处理，各组的标准差总会小一些。

配对组设计的作用在于控制组内变异与组间变异。它的优点是在实验处理之前，就把组间变异缩到最小和要求两组组内变异比单独的随机分配更接近相等。因此，这种设计能对被试者个别差异给予很多的控制，小型实验用配对设计，其效果比用随机分组的效果更为显著。但这种设计的缺点是，实验者因分配被试者而大大增加其工作量。

三、混合设计

混合设计 (mixed design) 乃是在一个实验中同时采用两种基本设计的实验设计。它要求一个自变量用一种设计处理，如被试者内设计处理，而另一个自变量用不同种类的设计处理，如被试者间设计处理。这时，实际上是在进行两个实验。如果某个实验要求处理三个自变量，可按同样的原理对三个自变量采用不同的设计。注意，这里谈及的三个自变量不同于一个自变量的同类的三种水平。这里举一个例子，若将实验限定在两个自变量的范围以内，在最简单的双变量实验设计 (2×2) 中，每个自变量在二种水平 (A 变量有 A_1 、 A_2 ，B 变量有 B_1 、 B_2) 上的混合设计有如下几种可能：

1. 表 2 - 10 中的 1、6、11、16 组是非混合设计，而 2、5 组代表同种混合设计，3、9 组也代表同种混合设计，等等。真正的混合设计有十二组。其实只要用对角线的一半编号组就可以说明问题，真正的混合组有二个六组，要么是 2、3、4、7、8、12，要么是 5、9、10、13、14、15。这两个六组只在接受处理的方向上不同，而接受处理的内容基本一致。

2. 现再以 2 组为例，展开随机组和配对组 2×2 混合设计。步骤一：将全部被试者分成二个随机组，一组接受变量 A_1 ，另一组接受 A_2 ；步骤二：将 A_1 组被试者分解为二个配对组 (G_1 、 G_2)，将 A_2 组也分解为配对

表 2 - 10 混合设计一例

		自变量 A			
		随机组	匹配组	不完全的被试者内设计	完全的被试者内设计
自变量 B	随机组	1	2	3*	4
	匹配组	5	6	7	8
	不完全的被试者内设计	9*	10	11	12
	完全的被试者内设计	13	14	15	16

说明：表中数字编号是任意的，和*代表同种混合。

(采自杨治良、乐竟泓，1990)

组 (G'_1 、 G'_2)；步骤三： A_1 组中的两个配对组接受变量 B_1 情况；另一组接受 B_2 情况，在 A_2 组中的二个配对小组也分别接受变量的 B_1 、 B_2 情况。上述情况可归纳为表 2 - 11。

3.再以 9 组为例，第一部是构成两个随机组，一组是接受变量 A_1 情况，另一组接受 A_2 情况。另一变项 B 的二个情况 B_1 、 B_2 以任何方式给 A 变项的二组中所有的被试者，构成被试者间和被试者内混合设计，从而把全部被试者的材料总合起来的时候，渐近误差将被抵消。

2×2 实验仅是混合设计的简要形式，若有需要，完全可以实现 2×4 、 3×3 、 3×6 等等的设计。例如，作 2×4 的随机组与被试者内混合设计可用表 2 - 12 的矩阵来表示，表中的数据是假想观察值。在决定一个实验是否用混合设计时，我们有必要在两种水平考虑：首先看看所涉及的有效变量是

表 2 - 11 2×2 混合设计

		自变量 B		
			B_1	B_2
自变量 A	A_1	G_1	接受	
		G_2		接受
	A_2	G_1	接受	
		G_2		接受

(采自杨治良、乐竟泓，1990)

表 2-12 2×4 混合设计

			自变量 B				X
			B ₁	B ₂	B ₃	B ₄	
自 变 量 A	A ₁	S ₁	0	0	5	3	8
		S ₂	3	1	5	4	13
		S ₃	4	3	6	2	15
	A ₂	S ₄	4	2	7	8	21
		S ₅	5	4	6	6	21
		S ₆	7	5	8	9	29
X		6/S	23	15	37	32	X=107

(采自杨治良，乐竟泓，1990)

否要求特殊设计，如指导语变量要求被试者间设计；在这一点上没问题时，就得从方便、经济和注意统计正确性的角度作进一步的考虑。

第二节 多变量实验技术

在上一节中，我们讨论了实验设计的基本类型，在以人或动物为研究对象的心理学实验中，首先要解决的问题是：实验处理中用了相同的被试者还是用了不同的被试者。根据这一原则，上一节中我们讨论了被试者内设计、被试者间设计和混合设计。尽管这样的分类也穷尽了各种实验，但在心理学实验中，大多数比较重要的实验都是多自变量实验。所以在这一节里，我们对多自变量的实验设计作进一步介绍。

一、多自变量实验的优点

多自变量（或多自变项）（multiple independent variable）是指在一个实验中包含有两个或两个以上的自变量。它不是指同一自变量的多个水平。在心理学的刊物上较少发现只用一个自变量的实验研究报告。一个复杂的心理现象很少是单一刺激因素决定的，往往是多因素的结果。例如，想研究不同的教学效果，可能有教学方法的影响，也可能有教师水平的影响，还有学生自身水平的影响等。这种教学效果，实际上是受多因素交互影响的结果，要使实验更接近实际，就应采取多因素，即多自变量的实验设计。典型的实验是同时操纵两到四个自变量。

那么为什么在心理学实验中在较多的情况下要采用多自变量实验呢？这是因为多自变量实验具有三个明显的优点：

1. 效率高 有两个自变量的实验要比分别做两个只有单一自变量的实验效率要高，也就是说事半功倍，花同样的时间，做了一倍，甚至二倍、三倍的工作。

2. 实验控制较好 做一个实验时某些控制变量比进行两个实验时更易于控制和恒定。如被试者条件是一样的，同一个被试者，同样的身心状况。再如时间条件也是一样的，日期相同、时间相同。其他条件就更不用说了，如相同的外界环境、相同的温度、相同的湿度等等。这就在很大程度上排除了许多实验误差，减少了实验污染。

3. 实验结果更有价值 有多种自变量的实验所得的结果，由于在多种情况下都证明是确实的，这样就比多个单独实验所概括的结果更有价值。例如，我们想要知道两种奖励办法中哪一种促进了中学生学习体育。第一种奖励办法是对正确完成规定的运动项目的学生给予物质奖励；第二种是给予提前下课，即每正确完成一个规定的运动项目允许学生提前下课，哪种更好些需要用实验证实。而且，在把它作为一项中学的规定之前，我们还应把这两种办法放到语文、数学等课程中去实验。在这里，不同课程的被试者是第二个自变量。显然，把奖励办法与不同的课程这两个自变量结合在一项实验中，要比进行两个连续的单自变量的实验为好。特别是，当一个自变量产生的效果在第二个自变量的每一水平上不一样时，交互作用（interaction）就发生了。在有交互作用的情况下，分别讨论每一自变量的效应就不够了。因为一个自变量的效应依赖于另一个自变量的水平。这里，我们举个例子。杨治良等人（1981）做了一个实验，目的是想了解年龄对再认能力的影响。第一个自变量是年龄，这里我们选取初中生年龄组和大学生年龄组。第二个自变量是实验材料，这里我们选取具体实物图形组和词组两个组。因变量取再认能力 d' 作指标（信号检测论方法用于再认实验，采用 d' 作为再认能力的指标）这个实验的方法是采用再认法，把被试者识记过的材料和没有识记过的材料混在

一起，要求被试者把两种材料区分开来。实验结果如图 2-1 所示（见下页）。

从图 2-1 可见：第一，在哪个实验条件下，初中生的再认能力较强；第二，对具体图画的材料，再认能力较强。然而，此图是将两个变量（年龄大小和材料性质）分开绘制的。在左图，初中生的再认能力指标中，既包括了对具体图画的材料，又包括了对词的材料。大学生的情况也是一样。

如果把同样一个实验结果，换另一种方式绘图（如图 2-2），我们就可以看到，实际情况并非这样简单。图 2-2 将两个变量画在一起，显示了交互作用的发生。它告诉我们：第一，在此实验条件下，再认能力较低的情况仅仅是在以具体图画作材料的大学生组发生；第二，无论是对具体图画材料，还是词材料，初中学生都表现出较高的再认能力，比较图 2-1 和图 2-2，我们看到了区别。具体地说，大学生对词材料的再认能力，在二张图上有不同的分析。图 2-2 能明确指出一个自变量的各水平受到另一个自变量水平的不同影响的交互作用就代表了这种不同的影响。

为进一步分析多自变量的优越性，我们也可以设想，两个自变量没有显出交互作用。图 2-3 就是虚构的情况。从图 2-3 上可以看到，任何一个自变量的效果对于另一个自变量的两个水平来说都是相同的。也就是说，第一，初中学生总是表现出较高的再认能力，不论实验材料是图画还是词。第二，对于图画材料，不管是初中学生还是大学生都表现出较高的再认能力。这就是图 2-3 说明的问题。平行线意味着没有交互作用的发生。

实验的实际情况并不是这样。实际情况如图 2-2 所示，再认能力的高低既依赖于年龄大小，也依赖于实验材料；一个自变量的效应依赖于另一个自变量水平。

上述三种情况的图解分析揭示了多自变量实验能揭示变量间是否存在交互作用。这是单自变量实验所不能完成的，我们想象这个实验是分两部分进行的。第一部分实验中，只有年龄大小被当作唯一的自变量。那么，实验材料就成了控制变量。我们知道，控制变量在实验过程保持恒定。如果选择的是图画作实验材料，实验结果将表明，初中学生比大学生表现出较高的再认能力。但是，实验者将不会知道，如果用词作材料会产生什么结果？在第二部分实验中，只有实验材料被当作唯一的自变量。年龄大小就成了控制变量。若用初中生做被试者，将揭示出不同的结果。我们可以看到，实验虽然做了两个，但得到的信息反而比一个双自变量实验少。二个单自变量实验所获得的信息反而比一个双自变量实验少。

图 2-2 是实际的实验结果，它表明了交互作用。一个自变量的效果依赖于另一个自变量的水平。具体地说：第一，若用词作材料，初中生和大学生的年龄因素对再认能力的影响较少；第二，若用图画作材料，初中生和大学生的年龄因素对再认能力的影响较大；第三，若用初中生作被试者，实验材料的因素关系不大；第四，若用大学生作被试者，实验材料的因素关系就很大。这里我们不难发现，一个双自变量实验实际上等于四个单自变量的实验，而且还要看单自变量实验是否做得好。

综上所述，当一个自变量的水平受到另一个自变量水平的不同影响时，交互作用就发生了。在有交互作用的情况下，分别讨论每一个自变量的效应就不够了，因为一个自变量的效应依赖于另一个自变量的水平，因而在实验结果的分析讨论中，除了对单独因素的分析之外，还必须分析讨论出现交互作用的原因和后果。

多自变量的实验是效率高、价值高的实验。在这些实验中，每一种因变量的测量至少能提供两个变量信息，从而有可能研究变量之间所有可能发生的交互影响。在实验中自变量数目增加时，能发生交互影响的数目也迅速增加。两个自变量只有一个交互作用影响的可能。用三个自变量（自变量 A、B、C）就有四种交互作用影响的可能，如 AB、AC、BC、ABC 四种。假定每一自变量有两个水平，三种自变量的交互影响由四条线生动地表现出来；两条表示变量 B、两条表示变量 C，与横坐标变量 A 起交互作用，见图 2-4，它们的交互作用由四条非平行线表示出来。

二、多因素实验设计

多因素实验设计 (multifactors experimental design) 是指在同一个实验里可以同时观测两个或两个以上的自变量的影响，以及自变量与自变量交互作用的效果的实验设计。它与上述所讨论的只用一个自变量的实验设计是不同的。包括 X_a 和 X_b 两个自变量的设计，叫做双向（或二向）析因设计，简称为 $A \times B$ 因素设计。包括 X_a 、 X_b 和 X_c 三个自变量的设计叫作三向析因设计，简称为 $A \times B \times C$ 因素设计。在因素设计中，每个因素（自变量）又可以包括几个水平。若两个自变量 X_a 和 X_b 各有两个水平，则可称为 2×2 设计；若自变量 X_a 有两个水平， X_b 有三个水平，则称为 2×3 设计。例如，有人想研究大学生对红、绿、黄三种灯光的反应是否与灯光的强度有关，在这项实验中，A 代表灯光强度，是一种因素（自变量 X_a ），包括强和弱两种水平；B 代表灯光频率，是另一个因素（自变量 X_b ），包括红、绿、黄三种水平。这种二向析因设计就是 2×3 因素设计。若它还想同时研究反应时是否有性别差别，则 C 代表性别（自变量 X_c ），分男女两个水平。这种三向析因设计就是 $2 \times 3 \times 2$ 因素设计。当然也可以有多于三个因素的设计。不过，因素多，解释结果的困难就大，交互作用也就多，因而研究者一般是将研究限于 2 至 4 个变量。在析因设计中，研究者要操作的实验处理的个数就是各自变量的水平个数的乘积。例如，在一项二因素实验设计中，设 A 因素有 p 个水平，B 因素有 q 个水平，则研究者就应操作 $p \times q$ 个实验处理。在确定了实验中所要操作的实验处理后，就要决定怎样分配被试者了。下面我们举一个研究实例。

杨治良等人（1993）在研究汉字内隐记忆的实验研究中，通过四个 $2 \times 2 \times 2$ 方差设计实验，试图寻找汉字认知范畴里内隐记忆存在的条件。此方差设计的三个因素为：（1）A 因素，即注意程度因素，分目标字和非目标字两个水平；（2）B 因素，分直接测量和间接测量两个水平，直接测量为再认判断任务，间接测量则是汉字偏好判断任务；（3）C 因素，即时程因素，分单元一和单元二两个水平，参见图 2-5。结果发现了内隐记忆存在的三个必要条件是：非目标汉字、汉字的整体加工和偏好判断任务。此例说明了，正是这些多因素实验设计，才发现了汉字内隐记忆存在的条件。

多因素实验设计可以概括成某一个基本形式。假设实验因子有 A、B 这两个，A 因子有三个水平，B 因子有四个水平，那么就有 $3 \times 4 = 12$ 种组合，每一个组合称作一种处理，现随机抽取被试者 1、2、3、……24，共 24 个分别安排在各种处理中接受实验，实验设计列表如表 2-13。

表 2-13 多因子设计表

实验因子		B 因子			
		B ₁	B ₂	B ₃	B ₄
A 因 子	A ₁	1,2	3,4	5,6	7,8
	A ₂	1,10	11,12	13,14	15,16
	A ₃	17,18	19,20	21,22	23,24

(采自杨治良、乐竟泓, 1990)

实验所得的数据,可表明哪种处理(即水平的结合)效果比较好。但到底多种处理的效果是否有显著差别,还要作进一步的统计检验。

三、拉丁方设计

拉丁方设计(或拉丁方格设计)(Latin - square design)是多变量实验设计中一种较为常用的设计方案。心理实验中采用循环法平衡实验顺序对实验结果的影响,就使实验顺序、被试者差异都作为一个自变量来处理。只要是实验中自变量的个数(因素)与实验处理水平数相同,而且这些自变量之间没有交互作用的存在时,都可采用拉丁方设计方案。这里对于这些因素之间没有交互作用的假设是很重要的。否则,若按没有交互作用的统计方法处理实验结果,只能是准实验设计。此设计的基本模式如表 2-14。

表 2-14 拉丁方设计

被试者	实验顺序		
	B ₁	B ₂	B ₃
C ₁	A ₁	A ₂	A ₃
C ₂	A ₂	A ₃	A ₁
C ₃	A ₃	A ₁	A ₂

(采自金志成, 1991)

表 2-14 中 A₁、A₂、A₃ 为实验处理的三个水平。C₁、C₂、C₃ 是被试者的三种不同类型,存在个体差异,被试者可为一人,也可为多人。对于这种拉丁方设计的实验结果,可用独立样本拉丁方变异数分析法,可分析出被试者间、实验顺序间、实验处理间的差异情形。就是说实验顺序、被试者差异对实验结果的影响都可分析出来。

假如,被试者是随机抽取的,且随机分派,并保证各组被试者相等。这时的结果分析可用相关样本拉丁方分析法来处理。可分析出被试者内的顺序(B因素)、处理(A因素)及部分 A×B 的交互作用。

拉丁方设计能够抵消实验中因实验顺序、被试者差异等造成的无关变量效果,因此在心理实验中经常被应用。随着数理统计的发展,多因素设计的方案越来越多,多因素的实验设计执行起来很费劲,被试者的数量要多,结果的统计处理也较复杂,因此,选用哪一种实验设计,要因地制宜,以做到恰到好处。

第三节 实验数据的统计分析

一、统计表和统计图

在心理学实验研究中，一般都是先获得一大堆原始数据和观测材料，虽然这些数据乍看起来十分凌乱，但它们却是实验所获得的最宝贵的财富。实验做完以后的分析和立论，都将以这些数据为基础。因此，为使实验获得成功，数据的整理十分重要。在数据整理的过程中，第一步是对数据的特点和种类加以分析，绘制出简单的统计图和统计表。统计图和统计表的一个共同优点在于一目了然，它所表示的信息容易被人们理解和接受。有人曾给予一个不甚恰当的比喻，如果文字的信息量是“1”，那么统计表的信息量就是“10”，统计图的信息量则是“100”。艾宾浩斯(Ebbinghaus Hermann, 1850~1909)著名的遗忘曲线就是用一个统计图概述了他实验的全貌。下面我们分别来介绍统计表和统计图的制作方法。

(一) 统计表

1. 统计表的功能和结构 在心理学研究报告中免不了要用几张统计表。这是因为，统计表(statistical table)是对被研究的心理现象和过程的数字资料加以合理叙述的形式。它在叙述统计资料方面有着重要的作用，有人称之为统计的速记。设计良好的统计表使统计资料表现得充分、明显而又深刻、有力，可以避免冗长的叙述。统计表由标题、横行和纵栏、数字资料等要素组成。统计表的构造一般包括如下几个项目：

(1) 序号：序号就是表的编号，要写在表的上方，表题的左方。序号一般在文章中出现的先后顺序编列。

(2) 名称：又称表题，是一个表的名称，应写在表的上方序号之后。表题的用语要简洁扼要，使人一望可知表的内容。如果用语过简，可在下面附加说明，但这种情况不宜多用。

(3) 标目：又分横标目和纵标目。横行标目写在表的左方；纵列标目写在表的上方，分别说明横行和纵栏的内容。

(4) 数字：数字是统计表的语言，又称统计指标。它占据统计表的大部分空间，书写一定要整齐，位数要上下对齐，小数点后缺位的要补零，缺数字的项要用“—”符号表示，不能空白。

(5) 表注：写于表的下方。它不是统计表的必要组成部分。如果需要可对标目补充说明。数据性质、数据来源、附记等都可作为表注的内容，文字可长可短，参见表 2-15。

表 2 - 15 29 名被试康复治疗前后体重差别阈值的變化

康复治疗	差别阈限(克)
前	175 ± 37.8*
后	84.4 ± 35.2

为了使统计表能对所研究的心理现象以鲜明的数字叙述，制表时应注意以下几点：

(1) 每一张统计表都必须有名称，统计表的各种标题，特别是表题的表述，应十分确切明了。内容应紧凑而富有表现力，避免过分庞大和琐碎。

(2) 表的各纵列之间要用线条隔开，表的两个纵线可以省去，上下两边

须有横线，标目与数字之间、数字和总计之间、两个总标目之间都须用线条隔开。表的上下二横线条要粗些。

(3) 表中各栏，通常是按照由局部到全部的原则编列的。

(4) 统计表应有计量单位名称。计量单位名称，通常加用圆括号并置于表头的右上方，或者置于标题或标目的下面。

2. 统计表的种类 统计表可以以形式及内容的不同来作为分类标志 将其划分成不同的类型。不同类型的统计表的具体功能不同。下面简述几种常用统计表：

(1) 简单表：只列出调查名称、地址时序或统计指标名称的统计表。例如表 2 - 15。

(2) 复合表：统计分组的标志有两个或两个以上的表。若只有两个分组指标的称两项表；若分组指标有三个的称为三项表，依此类推。表 2 - 16 就是一个复合表，且分类的标目有五个：间隔时间、实验数、节省百分数、节省百分数的中数、中数机误。

表 2-16 不同时间间隔后的记忆成绩

间隔时间	实验数	节省百分数的范围	节省百分数的中数	中数机误
1/3 小时	12	45 — 64	58.2	1
1 小时	16	34 — 54	44.2	1
8 - 9 小时	12	28 — 48	35.8	1
24 小时	26	15 — 46	33.7	1.2
6 天	26	12 — 46	27.8	1.4
31 天	26	3 — 40	25.4	1.3
	45	7 — 44	21.1	0.8

(采自 Ebbinghaus, 1885)

表 2 - 16 就是艾宾浩斯 (Ebbinghaus, 1885)

实验研究的主要结果，这个统计表集中了他对 1,300 个字表，反复测定了学习和回忆间隔时间的长短同遗忘的关系，分类统计而成。有了这个表，他就可以绘制成著名的遗忘曲线 (forgetting curve) (见 70 页图 2 - 6)。

(二) 统计图

为了写好实验报告，还需要利用统计图来表明心理现象的数量关系，这样就不需要作很多解释就可以让读者看懂。统计图有明显的优点，它不仅对统计资料和实验结果做出具体、明确的表达，易为读者所理解和获得深刻印象，而且由于统计图的表现生动活泼、醒目动人，具有很强的说服力。所以统计图是分析统计资料的重要工具，也是实验报告的重要内容。通过作图，可帮助我们揭示心理规律。

但是统计图也有不足之处，它不能获得确切数字，如果作图不当反而会掩盖事实真相，因而我们不能用统计图来代替统计表。在论文中，常将统计图和统计表一并列出。下面我们分别来讨论统计图的功用、结构和种类。

1. 统计图的功用和结构 统计图 (statistical figure) 乃是依据数字资料，应用点、线、面、体、色彩等来绘制成整齐而又规律、简明而又数量化的图形。统计图在数据整理中占有很重要的地位。一图知万言，一张简单的

图形，就可以把一大堆数据中 useful 信息概括地表现出来。

统计图一般多采用直角坐标系，横坐标用来表示事物的组成或自变量 X ，纵坐标常用来表示事物出现的次数或因变量 Y ，除直角坐标外还有角度坐标等。

2. 统计图的结构与制图要点

(1) 图号及图题：统计图的名称为图题。图题的文字应简要，只要求能扼要地叙述统计图的内容，使人一见就能知道图所要显示的是何事、何物，发生于何时、何地。图号是图的序号，图题与图号一般写在图的下方。图题的字体应是图中所用文字中最大的，但也不能过大，要与整个图形的大小相称，一般与图目文字的顺序一致，从左至右书写，放在居中的位置上。

(2) 图目：是写在图形基线上的各种不同类别、名称，或时间、空间的统计数量，即坐标上所有的各种单位名称。在统计图的横坐标和纵坐标上都要用一定的距离表示各种单位，这些单位称为图尺，有算术单位，亦有对数单位，百分单位等等，这要根据资料的情况加以选用，图尺分点要清楚，整个图尺大小要包括所有的数据值，如果数据值大小相差悬殊，图尺可用断尺或对数法，进行技术处理，减少图幅，增强图形效果。

(3) 图形：是图的主要部分，图形曲线要清晰，一般除图形线外，应尽量避免书写文字。为表示不同的结果，要用不同的图形线以示区别，各种图形线的含义用文字标明，选图中或图外一适当位置表示，目的是使整个图和谐美观且醒目。

(4) 图注：凡图形或其局部或某一点，需要借助文字或数字加以补充说明的，均称为图注。图注部分的文字要少，印刷字型一般要小，它可以帮助读者理解图形所示资料，提高统计图的使用价值，却又不破坏图的协调性。

此外，一个图形要使用各种线条，这些线条因在图中的位置不同而有不同的名称，如图形基线（横坐标）、尺度线（纵坐标）、指导线、边框线等。

3. 统计图的种类 常用的统计图有曲线图、条形图、直方图、点图、圆形图等等。通常表示事物各组成部分的构成情况的资料可用圆形图；频数分布资料可用直方图；资料内容各个独立者可用直条图；表示事物数量发展过程的连续性资料可用曲线图；表示两种事物的相关性的趋势可用点图。下面我们分别介绍这几种图。

(1) 曲线图：适用于连续性资料，表示事物数量在时间上的发展变动情况。因为借助于连续曲线，能够最恰当地描绘出心理现象在时间上的不断变化过程，见图 2-6。

绘制曲线图时，以横轴尺度表示时间、年龄等等，纵轴表示频数。纵、横轴尺度必须等距或有一定规则。纵轴尺度一般要从零开始，图线应按实际数字绘制，切勿任意描改为光滑曲线。若有几根曲线，应用不同形式的线条（实线、断线、点线等）区别开来，并用文字说明。

对数尺度曲线图是动态曲线图的一种特殊图式。图形的一轴制订出对数尺度，另一轴则按原样。图 2-7 是一个以分计的对数时间作横轴尺度，这样，我们从短短的横轴上看到了二年时间内的发展趋势，同时还看到了与相对量有关的绝对量的大致变动情况。

(2) 条形图：以相同宽度的条形长短来比较图形指标的大小，它是比较图中最常用的图形。条形图绘制方法简单，同时形式明确，图示效果好。绘图时，须先绘制一水平线作为条形的共同基线，依此基线为起点所绘制的条

形的长度，视图示的指标数值的大小而定。因此，必须定出一个比例尺度，作为绘制条形图的依据，同时各个条形的宽度要相等，各（组）条形间要有相当的间隙。

图 2-8 就是条形图。条形图有好几种。以同一水平线为基线的纵式条形图，也可制作成横式条形图。

（3）圆形图：一般用来表示事物各组成部分的构成情况。以一个圆的总面积代表总数，把面积按比例分成若干部分，以圆心角的角度大小来表示各组成部分的数量（如百分比）。代表圆面积中 1% 面积的扇形有 3.6 度的弧，各扇形面积中要标明百分数，并加文字说明。图 2-9 是一种表示被试来源的圆形图。不过一般被试来源用不着作图，作了图就有强调被试构成成分的含义。

（4）点图：表示两种事物相关性的趋势多采用点图。图的纵轴尺度和横轴尺度代表一种变量值的大小。习惯上自变量（X）的尺度放在横轴，因变量（Y）的尺度放在纵轴。不论纵轴、横轴的尺度都不必从零点起。在自变量与因变量的交叉点绘一个点，我们依据点的情况可以推测两种事物的相关情况。参见图 2-10。

二、实验数据的初步整理

心理学研究的结果一般都有数量记录，这些原始的数字材料往往是复杂而分散的，使人读了难得要领。因此必须经过分析整理，把材料有系统地组织起来。经统计处理能使复杂的材料变得简单扼要，把事实要点表示出来。下面我们从统计学的角度，概括地介绍如何对实验所得的大量原始资料进行科学的加工整理。

使用图表概括大量实验结果，这仅仅是对数据整理的第一步。实验所得资料只是表明每个个体的详细资料，为了能够反映综合特征，集中量和差异量是两个特别重要的参数。在次数分布上有两个重要特征：即重心位置和分布范围。用一定量数概括、规定重心位置的数字叫集中量，而用一定量数概括、规定分布范围的数字叫差异量。在介绍集中量和差异量之前，我们先讨论有关误差的问题。

（一）偶然误差与系统误差

大家都有这样的经验，无论实验做得多么精密，获得的观测数据总不完全一致，表现为数据的波动。产生数据波动的原因是因为有许多偶然因素影响实验结果。

1. 偶然误差 偶然误差（fortuitous error）或机误（chance error）是指实验中无法控制的偶然因素所引起的误差。例如测量仪器的灵敏度的有限性。又如，在有些测量中并未把温度、湿度看成影响因素，但是，温度和湿度时刻都在变化，这些都是偶然因素。无数的偶然因素影响实验或观测结果，使得测出的数据范围绕真值有一些上下波动。

例如，用天平称某物体的重量，进行 10 次观测得到 10 个数据，记在表 2-17 的第二列（栏）。

如果此物体重量的真值为 $a=150.6$ 克，表 2-17 中第三列（栏）记下的是各观测值与真值之差。

设： a 为某变量的真值， X_1, X_2, \dots, X_i 为其各次的观测值，则数 $X_i - a$ （ $i=1, 2, 3, \dots, n$ ）叫做 X_i 的偶然误差。

表 2-17 10 次观测所得的原始数据

序号	观测值 X (克)	Xi-a (偶然误差)
1	150.68	0.08
2	150.53	-0.07
3	150.54	-0.06
4	150.66	0.06
5	150.62	0.02
6	150.57	-0.03
7	150.56	-0.04
8	150.53	-0.07
9	150.61	0.01
10	150.58	-0.02

(采自华东师大数学系, 1980)

从表 2-17 第三列看出偶然误差可正可负, 可大可小, 因为它是由无法控制的偶然因素引起的。

2. 系统误差 有时在实验中还会出现另一种类型的误差, 它的观测值不是分散在真值的两侧, 而是有方向性和系统性的。所有重复实验的观测值大部分都会比真值偏高、或者偏低, 其原因是存在有系统误差 (systematic error)。产生系统误差的原因有很多, 如上例中可有仪器的故障, 有时也要考虑实验环境如照明、温度、压力、湿度的变化对实验结果的影响, 这时照明、温度、压力、湿度的变化就不能视为偶然因素了, 而是系统误差因素。另外, 在心理实验中, 观测者本身的一些因素 (如位置、练习、疲劳、时间等), 也能产生系统误差。排除系统误差是实验成败的关键。

(二) 集中量

对数据的概括了解, 在统计学上常由二种量数来表示: 一为表示集中趋势 (central tendency) 的集中量 (或集中量数) (measure of central tendency); 一为表示离中趋势 (variation) 的差异量 (measure of variation)。常用的集中量有平均数、中数和众数。下面我们分别进行讨论。

1. 平均数 一个物理量的真值是客观存在的, 通常我们无法知道它的真值, 而是通过测量或实验观测算出它的近似值。平均数 (mean) 或称算术平均数 (mathematic mean) 就是把一组数据加起来再用次数去除。它是刻画数据集中位置的极为重要的数。因此, 平均数有两个意义: (1) 对一组数据获得一个总的印象; (2) 将此组数据和另一组数据进行比较。平均数是一个主要量数。平均数用符号 M 表示, 其计算公式为:

$$M = \frac{\sum X}{n} \quad [\text{公式 2-1}]$$

X: 每一个度量
: 总和

n: 度量的总次数

假如所测原始数据较多, 可以进行归组计算, 则求平均数的公式为

$$M = \frac{\sum fX}{n} \quad [\text{公式 2-2}]$$

f: 每组的次数

X' : 各组的均数

2. 中数 中数 (或中位数、中点数) (median, 简称 M_d) 常用符号 M_d 表示。中数是在按大小顺序排列的一组数据中, 占中间位置的那个数。这个数可能是数据中的某一个, 也可能不是原有数。中数是集中量数的一种, 它能描述一组数据的典型情况, 其特点是极少受极端数据影响。

单列数据的中数的计算方法十分简单。如果个数为奇数, 则取序列为第 $(n+1)/2$ 的那个数据为中数。如果数据为偶数, 则取序列为第 $(n/2)$ 或第 $(n/2+1)$ 这二个数的平均数为中数。例如有下列八个数, 大小排列为: 3, 5, 6, 9, 10, 11, 13, 16。序列为 $n/2$ (即第 4 个数) 的数为 9; 序列为 $n/2+1$ (即第 5 个数) 的数为 10, 则中数为第 4 个和第 5 个二个数的平均数, 即 $(9+10)/2=9.5$ 。

假如数据较多, 也可以进行归组计算。则求中数的公式为:

$$M_d = L + \frac{(1/2 \times n - F) \times i}{f} \quad [\text{公式 2-3}]$$

L : 含有中数那一组的真实下限

n : 度量总数

F : 低于含有中数那一组的度量数

i : 组距

f : 含中数那一组里的度量数

3. 众数 众数 (或密集数、通常数、范数) (mode, 简称 M_o) 通常用符号 M_o 表示。众数是在整个分数里次数最多的一个度量, 在分组的次数分配上便是次数最多的一个组的中点。它也是一个集中量数, 也可用来代表一组数据的集中趋势。

众数计算起来很快, 不论是分组的数据还是未分组的数据, 都可用观察法来求众数。例如有一组数据为 4, 5, 6, 5, 7, 5, 3, 6, 不难看出 5 出现次数最多, 因此众数为 5。

在数据整理成数据分布的过程中, 同一数据由于分组组距的大小可变动, 因此组距中点的数值也必随之而有改变, 致使众数也有相当的移动。所以众数是不够稳定的, 在比较结果时它只能用作约略的参考而已, 因为众数受分组情况的不同而有所不同。

在心理学上, 众数和平均数的差别能反映实验的难度。如果平均数大于众数, 说明大多数人的度量结果低于平均数, 可见在此实验中多数被试者存在低估的情况。反之, 如果平均数小于众数, 说明大多数人的度量结果高于平均数, 可见在此实验中多数被试存在高估的情况。在统计学上, 众数和平均数之差可作为分配偏态 (skewness distribution) 的指标之一, 如平均数大于众数, 称为正偏态 (positive skewness); 相反, 则称为负偏态 (negative skewness)。

以上我们讨论了三种集中量。统计分析时可选择使用一种、二种或全使用。一般而言, 平均数和中数用得较多些。当没有极端数字影响, 数据分布比较对称, 此后的运算需要平均数时, 应使用平均数。当数据中有极端数据, 数据分布不对称时, 应使用中数。当需要很快估计出集中趋势或需要知道最多的典型情况时, 应使用众数。另外, 我们在日常体育和艺术比赛中, 也广泛地使用这些集中量数。例如“去掉一个最高分, 去掉一个最低分”等等, 都是为了能更好地反映集中趋势。

(三) 差异量

前面讲到的集中量，只描述数据的集中趋势和典型情况，它不能说明一组数据的全貌。一组数据除典型情况之外，还有变异性的特点。对于数据变异性即离中趋势进行度量的一组统计量，称之为差异量(或变异量数)(measures of variation)。这些差异量主要有全距、平均差、四分差、百分位差等，它们被称为低效差异量；标准差或方差被称为高效差异量。

1. 低效差异量

(1) 全距(或两极差)(range)：常用符号 R 表示。它是一组数据离散程度最简单的度量。计算起来也十分简便，可用如下公式求得：

$$R=U-L \quad [\text{公式 } 2-4]$$

R：全距

U：一组数据中的最大值

L：一组数据中的最小值

全距的计算比较简单，而且能回答我们直觉地提出的关于变量范围和间距等诸如此类的问题。但是全距与下面将介绍的其他差异量相比较，是比较不稳定的，因为，它仅仅是从分配中的两个个案的数值计算得来的，所以随机遇变化的幅度很大。

(2) 四分差(quartile deviation)：是指在一个次数分配中，中间 50% 的次数的全距的一半。四分差常用符号 Q 表示。其计算公式为：

$$Q = \frac{Q_3 - Q_1}{2} \quad [\text{公式 } 2-5]$$

Q：四分差

Q_3 ：第三个四分位数

Q_1 ：第一个四分位数

从以上公式可见四分差的计算也很简单，然而意义却十分明了。这就是说，在全分配上第一个四分位数与第三个四分位数之间包含着全体项数之半。次数分配越集中，离中趋势越小，则这二者的距离也越小。因此，根据这两个四分位数的关系，观测次数分配的离散程度，也可以得到相当高的准确性。可见，四分差可说明某系列数据中间部分的离散程度，并可避免两端值的影响。

(3) 百分位数(percentile)：百分位数的度量在心理学中也常用以表示度量的变异性。例如关于感受性的实验，要使刺激能被某组被试中百分之九十的人清晰的感受到，那就用到第九十个百分位数了。

百分位数的求法与中数相同。实际上中数本身也是一个百分位数，它是第五十个百分位数。

另外，百分位数也可以相当准确地用作图法求出，就是在绘成的累积次数曲线上进行简单的内插处理。

(4) 平均差(简称均差)(average deviation)：一般多用符号 AD 来表示。这也是一种检验离散程度通用的计算。尤其在阅读早年的心理学研究报告时，时常遇到用此度量表示离中趋势。它能告诉我们一组数据里所有的各量度与平均数的差数平均是多少。其计算公式为：

$$Ad = \frac{\sum(X - M)}{n} \quad [\text{公式 } 2-6]$$

Ad : 平均差

M : 平均数

X : 每一量数

n : 总量数之和

等式里两条垂直线表示两线段之间的数字只计其绝对值，而不计其正负号。因为我们感兴趣的是各个量度距离平均数有多远，而不管各个量度是比平均数大，还是比平均数小。从公式上可以看到，平均差的求法就是先算出各量度与平均数之差，不计正负号，加在一起，除以总次数，其商数就是平均差。平均差有其独特的功能，下一章将讲到的平均差误法（一种心理物理法）就是由平均差引伸而出的。但是平均差也有欠缺之处，即它易受极端数值的影响。

2. 高效差异量 高效差异量，顾名思义是指这些差异量能效率较高地反映分布范围。高效差异量有二个：标准差和方差。它们的具体优点很多。与全距相比，标准差和方差大大减少了两极端值的影响；与四分差相比，它们在计算过程中考虑到全部的离差；与平均差相比，它们在离差测定中避免了绝对值，因而有利于代数处理，从总体上看，与低效差异量相比，它们既能用于小样组，又能用于大样组。鉴于高效差异量的种种优点，在整理资料中常用标准差和方差。下面我们分别讨论这两个差异量。

(1) 方差（或变异数、变差、均方）(variance)：方差是每个数据与此组数据的平均数之差乘方后的均值，也就是离均差 X_d 平方后的平均数，它是度量数据分散程度的一个很重要的量数。方差作为统计量时，常用符号 S^2 表示。方差的计算公式为：

$$S^2 = \frac{(X_i - \bar{X})^2}{N} = \frac{X_d^2}{N} \quad [\text{公式 } 2-7]$$

(2) 标准差 (standard deviation) 是方差的平方根，通常用 S 或 SD 来表示。标准差的计算公式为：

$$S = \sqrt{\frac{(X_i - \bar{X})^2}{n}} \quad [\text{公式 } 2-8A]$$

S : 标准差

\bar{X} : 平均数

X_i : 个别分数

n : 总量数

当观测次数 $n < 25$ 时，亦即样本较小时，若除数用 n 算出来的数值用来估计总体标准差时往往会偏低，因此可用 $n-1$ 作为除数。上述公式 2-8A 就变为：

$$S = \sqrt{\frac{(X_i - \bar{X})^2}{n-1}} \quad [\text{公式 } 2-8B]$$

实际运算时，为简化计算可将分子 $(X_i - \bar{X})^2$ 演算成 $X_i^2 - \frac{(X_i)^2}{n}$ ，这样公式写成：

$$S = \sqrt{\frac{X_i^2 - \frac{(\sum X_i)^2}{n}}{n-1}} \quad [\text{公式 } 2-8C]$$

兹举一个工业心理学中的例子来说明离中趋势和平均数代表性之间的关系。设有两个生产小组各有工人 11 人，生产同样数量的零件，每人每天生产零件数如下：甲组：3、4、5、8、10、15、17、18、22、30、33；乙组：10、11、12、13、14、15、16、17、18、19、20。为计算方便，这里用公式 2-8A 对 81 页列表 2-18 进行运算。

两组工人平均日产零件数都为 15 件，它们的标准差却彼此不同。

甲组工人日产零件数的标准差为 $S_{\text{甲}} = \sqrt{1050/11} = 9.8$ （件）

乙组工人日产零件数的标准差为 $S_{\text{乙}} = \sqrt{110/11} = 3.2$ （件）

标准差是描写数据围绕其算术平均值离散程度的一个很重要的数据，具有重要的理论意义和实际意义：（1）首先说明平均数代表性的高低。上例告诉我们，虽然两组工人的平均日产零件数相等，但对两组工人的代表性来说，就不一样了。对甲组的代表性较小，而对乙组的代表性则相对大多了。可见把平均数和离中趋势结合起来应用，对反映现象的典型特征来说，具有一定的意义；（2）其次，在确定现象水平的基础上，进一步测定现象发生的节奏性或稳定程度。例如，工业生产中就可以通过离中趋势来看该企业执行计划的节奏性，变动程度很大的，就说明生产中存在着突击现象，前松后紧，时作时辍，还可以推测工作效率。

标准差用途很多，常用的主要有：（1）表示变量频数分配的离散程度：
表 2-18 甲、乙生产小组工人日产零件数及其计算标准差过程

甲组			乙组		
日产零件数 X	$(X - \bar{X})$	$(X - \bar{X})^2$	日产零件数 X	$(X - \bar{X})$	$(X - \bar{X})^2$
3	-12	144	10	-5	25
4	-11	121	11	-4	16
5	-10	100	12	-3	9
8	-7	49	13	-2	4
10	-5	25	14	-1	1
15	0	0	15	0	0
17	+2	4	16	+1	1
18	+3	9	17	+2	4
22	+7	49	18	+3	9
30	+15	225	19	+4	16
33	+18	324	20	+5	25
合计	0	1050	合计	0	110

（采自中科院心理所，1980）

在前面讨论过的例子中可以看到，在均数相同的情况下，标准差大，表示变量值分布得较散；标准差小表示变量值在平均数附近分布密集。（2）对变量

频数分配作出概括性的估计：统计学发现大多数的测量资料在数量很大时，其变量频数分配是靠中间近的比较多，离开中间远的比较少，且越远的越少，这种分配称为常态分配。常态分配是有一定规律可循的。这就是：总体内约有 68% 左右的个体变量值在平均数 ± 1 个标准差范围内；总体内约有 95% 左右的个体变量值在平均数 ± 2 个标准差范围内；总体内约有 99.7% 左右的个体变量值在平均数 ± 3 个标准差范围内。根据这个规律，只要算出平均数和标准差之后，就可以通过一批实际样本测量资料对所研究的总体做出概括的估计。（3）应用标准差计算平均数的标准误。同时它还是许多其他统计指标如正态曲线、相关系数、统计检验等的计算公式的要素。正因为如此，它在统计分析中占有极其重要的地位。目前，连普及型的电子计算器都可一按按钮就得出这个数据，并由此计算出其他统计量。

标准差用途中的第三条，即用来计算均数的标准误（用符号 $S_{\bar{x}}$ 表示），

计算标准误常常是显著性检验的最主要参数。标准误可用下列公式计算：

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} = \sqrt{\frac{X_i^2 - \frac{(\sum X_i)^2}{n}}{n(n-1)}} \quad [\text{公式 } 2-9]$$

从公式 2-9 中可看到，标准误大小与研究现象本身变异量的大小成正比，与样本个除数的平方根成反比。

三、显著性检验

（一）显著性检验的含义

表示样组上各种特性的常数叫做统计数 (statistic)，如平均数、中数、变差、差数、比值等等；表示全域特征的常数叫做参数 (parameter)，如全域的平均数、全域的中数、全域的差数、全域的比值等等。依照成套的、有系统的方法，借助样组去对全域参数作出某些表达，叫作统计推理（或统计推论）(statistical inference)。使用这类方法的目的在于检验统计假设，从而解决研究中的问题。所以统计假设 (statistical hypothesis) 一般是指关于全域参数的假定。统计假设 (H) 可用下列算式符号表示：

$$H = \mu_0$$

：全域参数

。：假设上规定的某一数值

通常决定是否拒绝假设，取决于检验样组指标与假定的全域指标差异是否显著。故统计检验 (statistical test) 又称显著性检验 (test of significance)。显著性检验的主要用途是检验两个或两个以上样本的统计量是否有显著差别。一般按三个步骤进行检验。第一步：提出假说或假定样组的平均数是从全域中取出来的。第二步：通过实际计算，求出 t、F 或 χ^2 等值。第三步：对假设做出取舍的决定。

显著性检验是统计整理资料的必然继续，也是统计分析的前提条件，它的积极意义是不言而喻的。但是在使用过程中，亦不可过分夸大它的作用。陈舒永 (1983) 曾对此作过精辟的分析，提出在使用显著性检验时应把握好以下四点：

(1) 显著性水平的高低并不代表差异的大小，只表明这种差别因抽样误差引起的可能性小于某个水平。有人把差异显著性考验后 $P > 0.05$ 当做两个

实验结果无差异的指标来使用。例如考察两个同类的实验材料难易程度是否一致时，往往把两个实验结果考验一下，如果 $P > 0.05$ ，就宣布这两个实验材料难易相等，可以作为同一实验材料的复式使用。这个方法似乎中外心理学家都用过。但它是否符合实际，是否真有理，还是值得商榷的。上面已经指出，当一个作者只列出 $P > 0.05$ 时， P 值可变动的范围是很大的。如 $P = 0.95$ ，也就是说两个实验材料的差异有 95% 的可能性是由机遇造成的。在这种情况下，说它们基本上无差异，无疑是对的。可是如果 $P = 0.06$ 或 $P = 0.10$ ，那就意味着两个实验材料的差异只有 6% 或 10% 的可能性是由机遇造成的，也就是它们的差别有 94% 或 90% 的可能性是真实可靠的。因此，显著性水平仅指差别的可能性不大。

(2) 显著不显著，并不代表实验设计的正确与否。经考验差异显著，只能说明这个差异由机遇造成的可能性很小，并不能保证实验设计就一定正确。有些作者把差异显著性考验的结果 $P < 0.01$ 或 $P < 0.001$ 当做王牌，好像有了它就可以保证一切结论的正确性。例如我们常常看到：由于 $P < 0.001$ ，所以第一个学习方法比第二个学习方法更有效；由于 $P < 0.01$ ，所以记忆广度是随年龄不同而变化的等等。实际上这是对差异显著性检验的要求太多了些，超出了检验的性能。再从另一个方面来分析，如果实验条件没能控制好，即使是 $P < 10^{-10}$ 也不能弥补实验设计的缺陷。例如有一个研究刺激的可编码性 (codability) 或可命名性 (namability) 对“异”“同”判断的反应时间影响，所用的实验材料为纯音-图形组合，纯音为 100 和 1060 赫，图形为三角和方块。实验程序是先呈现一个纯音-图形组合，再呈现另一个纯音-图形组合。要求一组被试只判断在两次呈现的纯音-图形组合中，两个纯音的异同；要求另一组被试只判断图形的异同。结果是对纯音来说， $RT_{\text{相同}} > RT_{\text{不同}}$ ，而几何图形则是 $RT_{\text{相同}} < RT_{\text{不同}}$ （在这里 $RT_{\text{相同}}$ 和 $RT_{\text{不同}}$ 分别代表着“相同”和“不同”判断的反应时间）。原作者假设纯音是不易编码的刺激，而几何图形则是容易编码的，并且经显著性考验证明，对异同判断的反应时间因实验材料（纯音或图形）不同而各异，它们的相互作用是显著的（ $P < 0.01$ ），从而得出结论：对不易编码的材料判断异同时， $RT_{\text{相同}} > RT_{\text{不同}}$ ；而容易编码的材料则相反， $RT_{\text{相同}} < RT_{\text{不同}}$ 。在这里 $P < 0.01$ 能不能保证上述结论正确呢？细细分析，这里有几个变量是混淆在一起的，可编码性、可辨别性和感觉到的不同，都可以成为判断异同的反应时间有差异的原因，为什么把这个差异只归因于可编码性呢？在这里 $P < 0.01$ 并不能改变实验设计中自变量混淆的情况，虽然差异非常显著（ $P < 0.01$ ），但得出上述结论仍然是错误的。由此可见，差异显著性检验并不代表实验设计正确与否。这是显著性检验应把握好的第二点。

(3) 显著或不显著只是相对的，不是绝对的。在心理学实验中，两种实验条件下得到的结果有差异时，常常要进行显著性检验。根据统计学的惯例，如果考验的结果为 $P < 0.05$ ，则两个结果差异显著；如果 $P < 0.01$ ，则差异非常显著；如 $P > 0.05$ ，则差异不显著。所谓 $P < 0.05$ 意味着得出的两个结果有差异的结论所冒的犯错误的风险等于或不到 5%。用统计学的术语来说，就是犯第一类错误的概率等于或不到 5%。如果 $P > 0.05$ ， P 值可以是 0.06 ~ 0.99 之间的任何值。但 $P < 0.05$ 和 $P < 0.06$ 的意义很不相同。如 $P < 0.05$ ，两个结果差异由机遇造成的可能性 5%；如果 $P < 0.06$ ，则两个结果差异由

机遇造成的可能性 6%。实际上 P 是一个连续变量，把差异显著和不显著的界线划在哪里完全是人为的。这就像高考的录取分数线一样，可以定在 380 分以上，也可以定在 400 分以上。总之，要定一个界线，否则不好办事。依统计学的惯例把显著和不显著的界线定在 0.05 而不定在 0.06 处，这是无可厚非的。于使用显著考验的过程中，有人把这个分界线看得过重，好像 P 0.05 和 P > 0.05 有天壤之别。P 0.05 和 P 0.50 固然差异很大，但 P 0.05 和 P 0.06 相比，其差异则是微不足道的，我们不能把数字过于绝对化。

(4) 当检验结果相差不显著时，不能马上得出结论说没有差别，要考虑假不显著的可能，即两个样本来自不同的总体，但检验却得出差异不显著的结果即犯了第二类错误 (type II error)。这可能由于样本所包含的例数太少或其他原因致使误差偏大等，必要时可加大样本重复实验。当然，当所得结果没有实际意义时，则不必进行显著性检验。

综上所述，我们应以科学的严谨态度看待显著性检验。既要看到它的巨大的积极作用，又要看到它并不是包医百病的万应灵药，不能只根据它来判断一篇论文的结论是否正确。另外，显著性检验还有一些前提条件，这些在专门的统计学书籍中有详细的叙述。

(二) t 检验

在心理学实验研究中，两项实验结果之差，有时是随机引起的差异，有时则是由自变量所造成的，t 检验 (或 t 检定) (t test) 就是分辨随机差异与自变量引起的差异的手段之一。当总体 (或母体) (population) 指标 X 服从常态分布时，测统计量 t 为：

$$t = \frac{\bar{X} - a}{S_x} \quad [\text{公式 } 2-10]$$

\bar{X} ：容量为 n 的子样平均值

S_x ：子样均数的标准差，即标准误

a：母体指标 X 的平均值

t 分布 (或 t 分配) (t distribution) 的概率密度函数的图形是对称于直线 t=0 的曲线。当 n 较小时，t 分布较标准常态分布的分散程度大些，当 n 无限增大时，t 分布则趋于标准常态分布。图 2-11 为 t 分布图解。样本 (或子样) (sample) 平均数和总体平均数的差数用标准误的倍数来表示，这就是 t 值。若 t (统计量) = 0，则表示两个小样本来自同一母体。t 进入危机领域，说明不来自同一个母体。

t 检验用来确定两个平均数的差别是否显著。t 检验因具体情况有所不同，检验方法也稍有差别。一般有以下三种情况：

1. 比较样本平均数与总体平均数差异的显著性 这里我们通过一个具体的实例进行解释。

假设：已知我国六岁儿童记忆能力的平均数是 65 分 (假定单位，可从大量调查和测量中获得)，现从患某病的六岁儿童 16 名中测得子样组的平均数 X=55 分，标准差为 12，试问患这种病的儿童与正常儿童的记忆能力有无本质区别。

分析：当我们根据大量调查的结果，或以往的多次实验 (或经验)，已知某事物的平均数 (例如生理、心理的正常值)，可将其当总体的平均数看待。此时可用公式 2-9 来检验样本平均数与总体平均数的差异的显著性。具

体计算如下：

$$t = \frac{65-55}{\frac{12}{\sqrt{16}}} = 3.33$$

查 t 值表，此处自由度 df 为 $n-1=16-1=15$ 时，t 值的 $t_{0.005(15)}=2.131$ ， $t_{0.01(15)} = 2.947$ ， $t_{0.001(15)}=4.073$ 。

现 $t_{0.01} < t < t_{0.001}$ ，故 $P < 0.01$ 。这说明患此病的儿童的记忆能力与正常儿童相比，在统计学上有非常明显的意义。

2. 比较同一批对象实验前、后差异的显著性 比较这类资料时，要先求出各个体实验前、后的差数，然后求出各差数的平均数及标准误。和所有统计推理一样，第一步作无效假设，即假设实验处理是什么作用。依据这一假设，实验处理前，后差数应等于0，而现在的实际观测差数的平均数为 \bar{X} ，检验这个 \bar{X} 与0之间有无显著性差异。第二步实际计算。第三步站在n%意义层级上讲话，看是否拒绝无效假设。

比较同一批实验对象实验前、后差异是否有显著性的计算公式是：

$$t = \frac{\bar{X}-0}{S_i} \quad [\text{公式 2-11}]$$

我们用一个例子来说明这一类型的具体计算过程。例如，时蓉华等对针灸的镇痛效应进行了研究。先对被试者进行一次痛阈测定，具体指标是产生痛阈的钾离子致痛仪上的电流值（单位为 mA），然后对某一穴位进行针灸。继而再测定一次痛阈。比较同一被试者针灸前后痛阈的变化，实验获得了表 2-19 上的结果。

表 2-19 针灸前后痛阈的变化

被试姓名	针 灸		差数 X	χ^2
	前	后		
赵 × ×	0.92	1.10	0.18	0.0324
杨 × ×	0.89	1.05	0.16	0.0256
王 × ×	0.94	0.98	0.04	0.0016
钱 × ×	0.88	0.96	0.08	0.0064
张 × ×	0.86	0.89	0.03	0.0009
李 × ×	1.01	0.97	-0.04	0.0016
宋 × ×	1.03	1.14	0.11	0.0121
丁 × ×	1.02	1.19	0.17	0.0289
薛 × ×	0.95	1.03	0.08	0.0064
黄 × ×	0.88	0.91	0.03	0.0009

*钾离子致痛的电流毫安数（采自时蓉华等，1980）
根据公式 2-9 和 2-11，获得如下具体计算：

$$S_{\bar{x}} = \sqrt{\frac{x^2 - \frac{(\sum x)^2}{n}}{n(n-1)}} = \sqrt{\frac{0.1168 - \frac{(0.84)^2}{10}}{10 \times 9}} = 0.022667$$

$$t = \frac{\bar{x}}{S_{\bar{x}}} = \frac{0.084}{0.022667} = 3.71$$

查 t 表, 当 $df=10-1=9$ 时, $t_{0.05(9)}=2.262$, $t_{0.01(9)}=3.25$, 现 $t > t_{0.01(9)}$, 故 $P < 0.01$ 。由此得出, 被试者针灸后, 对痛阈有显著性影响。

3. 比较二个样本的平均差异的显著性 设两个母样都是常态分布 标准误差相等, 各自抽取一个子样, 子样的容量为 n_1 和 n_2 , 子样平均值为 X_1 和 X_2 , 子样标准差为 S_1 和 S_2 , 可以证明统计量:

$$t = \frac{X_1 - X_2}{\sqrt{\frac{(n_1 - 1) \times S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \times \frac{n_1 + n_2}{n_1 \times n_2}} \quad [\text{公式 } 2-12A]$$

上述 t 值计算公式也可写为:

$$T = \frac{X_1 - X_2}{\sqrt{\frac{[x_1^2 - \frac{(\sum x_1)^2}{n_1}] + [x_2^2 - \frac{(\sum x_2)^2}{n_2}]}{n_1 + n_2 - 2}} \times \frac{n_1 + n_2}{n_1 \times n_2}} \quad [\text{公式 } 2-12B]$$

检验二个子样是否来自母体平均数相等的常态母体的步骤亦可分为三步: 第一步, 先作无效假设; 第二步, 按公式算出 t 值; 第三步, 按照采取的信度 (如 5%) 查 t 分布表, 自由度为 $(n_1 + n_2 - 2)$ 。如果 t 值大于信度水平, 则可认为两个母体平均数是有差异的。

这里结合具体实验结果进行运算分析。杨治良 (1988) 为研究中国人和外国人对汉字和英文在概念形成过程中的某些特点, 对二组被试者进行比较。一组是中国人掌握汉字假设检验模型, 另一组是美国人掌握英文假设检验模型 (见表 2-20)。

根据公式 (2-12), 可作如下计算:

$$t = \frac{11.42 - 7.67}{\sqrt{\frac{76.67 + 78.92}{22 \times 6}}} = 3.4543$$

查 t 表, 当 $df=22$ 时, $t_{0.001}=2.819$, 现 $t > t_{0.01}$, 故 $P < 0.01$ 。这说明二组间有显著差异, 即中国人掌握汉字假设检验模型优于外国人掌握英文假设检验模型。

表 2-20 中、外二组被试概念形成速度比较

中国人汉字组		美国人英文组	
4	16	9	81
5	25	9	81
6	36	11	121
6	36	12	144
6	36	12	144
7	49	14	196
7	49	14	196
8	64	15	225
9	81	16	256
10	100	17	289
11	121	33	1089
13	169	33	1089
=92	782	195	3911
$\bar{X}=7.67$		1625	

*概念形成所需的学习单位数（采自杨治良，1988）

（三）F 检验

F 检验（或 F 检定）（F test）是以数据的方差（变异数）分析为基础，故又称方差分析（或变异数分析）。上面讲到的 t 检验法只能对两组的平均数加以比较，而方差分析法却能对二组和二组以上的平均数加以比较。这在研究工作中是常遇到的情况。把实验个体完全随机地分配到几组中，各组分别用不同的处理法进行实验，所得到的数据是单因素的，可是在这个因素中却包含好几个水准，每种处理代表一个水准。这些水准有时是选择型的（固定的），有时候是随机型的（非固定的）。

F 检验的功能在于分析实验数据中不同来源的变异对总变异的贡献大小，从而确定实验中的自变量是否对因变量有重要影响。

1. 方差分析的基本原理 方差分析（或变异数分析）（analysis of variance，简称 ANOVA）是一种应用非常广泛的变量分析方法，它乃是用试验结果的观察值与其平均值之差的平方和，来分析某些因素对试验结果是否有显著影响。设所考察的因素为 A，把 A 的变异分成 b 个等级，每一等级重复 a 次试验，以 X_{ij} 表示第 j 个等级第 i 次试验的观测值（指标），于是得到单因素分析的一个子样，容量 $n = ab$ ，见表 2-21：

$$\text{表中 } T_j = \sum_{i=1}^a X_{ij} \quad \bar{X}_j = T_j / a \quad (j = 1, 2, \dots, b)$$

$$T = \sum_{j=1}^b T_j \quad \bar{X} = \left(\sum_{j=1}^b \bar{X}_j \right) / b$$

实际上 \bar{X} 是 $n = a \times b$ 个 X_{ij} 的总平均数。

$$\bar{X} = \frac{\sum_{i=1}^b \sum_{j=1}^a X_{ij}}{ab} \quad [\text{公式 } 2-13]$$

若 $\sum_{j=1}^b \sum_{i=1}^a (X_{ij} - \bar{X})^2$ 称为离差平方和，可以证明下列分解式成立：

$$\sum_{j=1}^b \sum_{i=1}^a (X_{ij} - \bar{X})^2 = \sum_{j=1}^b \sum_{i=1}^a (X_{ij} - \bar{X}_j)^2 + a \sum_{j=1}^b (\bar{X}_j - \bar{X})^2$$

上式等号右端第一项为各组（同一等级的数据构成一组）内部离差平方和。第二项为组与组间的离差平方和，即：

$$S_{\text{总}} = \sum_{j=1}^b \sum_{i=1}^a (X_{ij} - \bar{X})^2 \quad [\text{公式 2-14}]$$

$$\text{而 } S_{\text{误}} = \sum_{j=1}^b \sum_{i=1}^a (X_{ij} - \bar{X}_j)^2 \quad [\text{公式 2-15}]$$

$$S_A = a \sum_{j=1}^b (\bar{X}_j - \bar{X})^2 \quad [\text{公式 2-16}]$$

则上式分解式就是 $S_{\text{总}} = S_{\text{误}} + S_A$ 。此式说明围绕总共平均值的波动值 $S_{\text{总}}$ 由两部分组成，一部分表示偶然误差引起的数据波动值 $S_{\text{误}}$ ，另一部分为因素取不同等级引起的数据波动值 S_A 。

有了上述各等式，我们就可进行 F 检验。为了检验因素 A 的不同等级对试验结果的影响是否显著，我们只要比较 $S_{\text{误}}$ 和 S_A 的大小就行了。设所考察的指标的母体服从常态分布，可以证明变量：

$$F_A = \frac{\frac{S_A}{b-1}}{\frac{S_{\text{误}}}{b \times (a-1)}} \quad [\text{公式 2-17}]$$

这样，服从自由度 $n_A = b-1$ ， $n_{\text{误}} = b(a-1)$ 的 F 分布。显著性检验方法是先用表 2-21 的数据按上式算出 F_A 的值，然后取一定的信度 d ，例如取信度 $d=5\%$ （或者 1% ），查 F 分布表，找出信度为 d 、自由度为 n_A ， $n_{\text{误}}$ 的 F 值为： $F_d(n_A, n_{\text{误}})$ ，若是 $F_A > F_d(n_A, n_{\text{误}})$ ，就以 $1-d$ （95% 或者 99%）的把握断定因素 A 是显著的；若是 $F_A \leq F_d(n_A, n_{\text{误}})$ ，就不能认为因素 A 等级的变异对试验结果有显著影响。

以上我们介绍了 F 检验的基本原理，下面我们就可以讨论 F 检验的几种情况了。

2. 单因素方差分析 只考虑一个因素的变异对试验结果是否有显著影响的问题，就是单因素方差分析（simple factor analysis of variance）的问题。进行 F 检验法时，常将实验数据列成下列方差分析表，计算起来比较方便（见表 2-22）。

表 2-22 单因素方差分析表

变差来源	离差平方和 SS	自由度 df	$MS = \frac{SS}{df}$	F	P
总的	$X^2 - \frac{(\sum X)^2}{n}$	n-1			
组与组间	$\sum_{i=1}^b \frac{(\sum_{j=1}^a X_{ij})^2}{a_i} - \frac{(\sum X)^2}{n}$	b-1			
组内	$X^2 - \sum_{i=1}^b \frac{(\sum_{j=1}^a X_{ij})^2}{a_i}$	$\sum_{i=1}^b (d_i - 1)$			

(采自杨纪柯 , 1965)

在计算平方和 SS 中比较复杂的第一项可具体分解为 :

$$\begin{aligned} \sum_{i=1}^b \frac{(\sum_{j=1}^{a_i} X_{ij})^2}{a_i} &= \sum_{i=1}^b \frac{(X_{i1} + \dots + X_{ia_i})^2}{a_i} \\ &= \frac{(X_{11} + \dots + X_{1a_1})^2}{a_1} + \frac{(X_{21} + \dots + X_{2a_2})^2}{a_2} + \dots \\ &\quad + \frac{(X_{b1} + \dots + \dots X_{bab})}{a_b} \end{aligned}$$

在计算自由度 df 中的末一项可具体分解为 :

$$\sum_{i=1}^b (a_i - 1) = (a_1 - 1) + (a_2 - 1) + \dots + (a_b - 1)$$

从表 2-22 可以看出组与组间变差的平方之和与组内变差的平方之和相加得总的平方之和。这个总的平方之和相当于以前在未加分组的情况下所算得的 $(X - \bar{X})^2$ 相同, 现在却可以划分为两部分了。自由度照样依此划分为两部分。

下面用实例来分析具体计算过程。时蓉华等 (1980) 为比较针刺与暗示对痛阈的影响, 设立四种实验处理以考察其效应 (见表 2-23)。

根据上述有关公式, 可作如下计算 :

$$= 8.75 + 3.80 + 10.80 - 2.75 = 20.$$

表 2-23 四种实验处理对痛域的影响

组类	针刺组 (A)		暗示组 (B)		结合组 (C)		对照组 (D)	
	d	d ²						
	0.5	0.25	0.85	0.72	0.15	0.02	0.7	0.49
	2.45	6.00	0.15	0.02	1.0	1.0	-0.6	0.36
	0.75	0.56	-.95	0.90	0.9	0.81	0	0
	2.0	4.00	3.0	9.00	1.05	1.10	-1.15	1.32
	3.4	11.56	-0.25	0.06	1.0	1.0	-1.8	3.24
	-0.25	0.06	0.5	0.25	2.3	5.29	1.15	1.32
	1.15	1.32	0.3	0.09	0.7	0.49	0	0
	-0.05	0	0.4	0.16	1.7	2.89	0.75	0.56
	-1.1	1.21	0.4	0.16	0.7	0.49	0.4	0.16
	-0.15	0.02	1.35	1.82	0.8	0.64	0.95	0.90
	-0.05	0	-1.05	1.10	0.5	0.25	0.05	0.00
	0.1	0.01	0.9	0.81	0	0	-.9	0.81
	8.75	24.99	3.80	15.09	10.80	13.98	2.75	9.16
ai	12		12		12		12	

(采自时蓉华等 , 1980)

n=48

$$\sum d^2 = 24.99 + 15.09 + 13.98 + 9.16 = 63.22$$

$$C = \frac{(\sum X)^2}{n} = \frac{(20.6)^2}{48} = 8.84$$

$$\sum x^2 = \sum X^2 - C = 63.22 - 8.84 = 54.38$$

$$\frac{(\sum X)^2}{a_i} = \frac{(8.75)^2}{12} + \frac{(3.8)^2}{12} + \frac{(10.8)^2}{12} + \frac{(2.75)^2}{12} = 17.93$$

$$\frac{(\sum X)^2}{a_i} - C = 17.93 - 8.84 = 9.09 \text{ (组与组间项的平方和)}$$

$$\sum x^2 - \frac{(\sum X)^2}{a_i} = 54.38 - 17.93 = 36.45 \text{ (组内项的平方和)}$$

各项自由度为：

“ 总的 ” 项 : df=n-1=47

“ 组与组间 ” 项 : df=b-1=3

“ 组内 ” 项 : df= (ai - 1) = 44

有了这些 , 就可填方差分析表 , 如表 2-24。

表 2 - 24 方差分析表

变差来源	SS	df	MS	F	p
总的	54.38	47			
组与组间	9.09	3	3.03	2.9417	< 0.05
组内	45.29	44	1.03		

检表得： $F_{0.05.3.40}$, $F_{0.01.3.40}=4.31$ （采自时蓉华等，1980）

取信度 $d=0.05$ ，得到 $F_A > F_{0.05.3.40}$ ，按 F 检验法得出，实验处理对试验结果的影响是显著的。进一步的配合 t 检验就可发现各组间的差异。见表 2-25。

表 2 - 25 实验各组痛域、耐痛域变化比较

项目	痛域	耐痛域		
	F 检验	各组比较—— t 检验	F 检验	各组比较—— t 检验
针刺组 暗示组 结合组 空白组	$p < 0.05$	1. 结合组与对照组： $p < 0.01$ 2. 针刺组与对照组： $p < 0.05$ 3. 暗示组与对照组： $p < 0.05$ 4. 结合组与暗示组： $p < 0.05$	$p = 0.05$	结合组与对照组比较： $P < 0.01$

（采自时蓉华等，1980）

3. 双因素方差分析 在前面第二节中，我们曾论述了多自变量的优越性。方差分析中，最常见双因素方差分析（two factors analysis of variance）。下面举实例分析双因素方差分析的具体计算方法。杨治良（1988）采用人工概念探索概念形成的过程，实验处理是对被试随机等分成二大组，一组被试阅读要求问题解决的指导语，另一组被试阅读要求记忆的指导语。在每组被试中，再随机分成五种实验条件，即：

- 条件一；主试给被试的反馈答案正确率为 100%；
- 条件二；主试给被试的反馈答案正确率为 90%；
- 条件三；主试给被试的反馈答案正确率为 80%；
- 条件四；主试给被试的反馈答案正确率为 70%；
- 条件五；主试给被试的反馈答案正确率为 60%。

这样，问题解决组和记忆组各有被试 60 名，而每种实验条件各有被试 12 名（见表 2-26）。

表 2 - 26 对全部被试掌握概念的观测结果

条件 A a=5	条件一		条件二		条件三		条件四		条件五	
因素 B b=2	问题 解决	死记	问题 解决	死记	问题 解决	死记	问题 解决	死记	问题 解决	死记
k = 12	7	14	8	17	8	19	9	23	10	24
	8	16	9	18	9	22	10	24	13	27
	9	17	9	20	10	24	11	26	13	29
	9	18	9	21	10	25	12	29	14	33
	10	20	10	24	11	27	13	31	14	33
	11	22	11	27	11	29	14	33	15	33
	12	23	11	28	12	31	14	33	15	33
	12	25	12	30	12	33	15	33	15	33
	12	29	13	32	14	33	15	33	16	33
	12	31	14	33	14	33	16	33	16	33
	12	33	14	33	15	33	16	33	19	33
15	33	16	33	20	33	17	33	19	33	
129	281	136	316	146	342	162	364	179	377	

(采自杨治良, 1985)

根据上述有关公式, 可作如下计算:

$$n = 5 \times 2 \times 12 = 120, \quad T = 2420, \quad T^2 = 58932$$

$$C = \frac{(2420)^2}{120} = 48803.3$$

$$SS_T = 58932 - 48803.33 = 10128.67$$

$$SS_A = \frac{(389)^2 + (452)^2 + (488)^2 + (526)^2 + (556)^2}{2 \times 12} - 48803.33 = 641$$

$$SS_B = \frac{(740)^2}{5 \times 12} + \frac{(1680)^2}{5 \times 12} - 48803.33 = 7863.3$$

$$SS_{AB} = \frac{(117)^2 + (281)^2 + (136)^2 + (316)^2 + (146)^2 + (342)^2 + (162)^2 + (364)^2 + (179)^2 + (377)^2}{12}$$

$$- 48803.3 - 641 - 7363.3 = 41.67$$

$$SS_D = 10128.67 - (641 + 7363.33 + 41.67) = 2082.67$$

$$df_T = 120 - 1 = 119, \quad df_A = 4, \quad df_B = 1, \quad df_{AB} = 4, \quad df_D = 119 - 9 = 110$$

将上面求得的结果, 列成方差分析表(见表 2-27)。

表 2-27 方差分析表

变差来源	SS	df	MS	F	p
总的	10128.67	119			
条件, A	641	4	160.25		
指导语, B	7363.33	1	7363.33	8.465	< 0.001
交互影响, A, B	41.67	4	10.42	388.977	< 0.001
抽样误差	2082.67	110	13.93	0.550	> 0.05

(采自杨治良 ' 1985)

根据表中的 F 值和 p 值, 就可做出实验推论: (1) A 因素对概念的形成的速度有显著的影响; (2) B 因素对概念的形成的速度有更为显著的影响; (3) A 因素和 B 因素的结合, 对概念的形成的速度并无显著的交互影响存在。

(四) X^2 检验

X^2 检验 (或 X^2 考验, 卡方检定) (chi-square test) 是比较观察次数与理论次数之间的差异的统计方法。这里的 X 是希腊字母, 读作 [hai], 不按拉丁系字母 X 读音。在统计学中, 检验分为参数检验 (parametric test) 和非参数检验 (nonparametric test)。前者如某一总体指标是否等于某一数值; 后者如某一随机变量是否服从常态分布 (normal distribution)。在以实用常态分布和 t 分布作为准则尺度去检验统计假设时, 这些假设都是有关参数的假设。为了使得检验结果有效, 它们都需要在事件假定与检验对象相适应的特种分布形态。另一些统计检验是用来检验分配, 而不是用来检验参数的。它们所检验的分配在先验的假定上并不要求具有一定的形态, 故称为自由分配' 而用来作假设检验的准则量数就叫非参数的统计量数。 X^2 即属于这种非参数统计 (nonparametric statistics) 之列。

X^2 的定义可用下式表达之:

$$X^2 = \sum_{i=1}^n \frac{(\text{实计数} - \text{预计数})^2}{\text{预计数}} = \frac{(A - T)^2}{T} \quad [\text{公式 } 2-18]$$

A: 实际值 (或实计数)

T: 理论值 (或预计值)

: 总和

n: 计数组

上式标志着实际进行计算 X^2 的定义。至于在理论函数上的 X^2 恰如 t 分布一样, 亦是随着自由度的变化形成一簇理论上的分配形态。自由度越大, 其分配形态便越接近于常态。

在进行 X^2 测验时应注意以下几点:

1. 计算 X^2 值过程中, 必须用绝对值, 切不可用相对数, 因 X^2 值的大小与频数有关。

2. 做 X^2 检测时, 应先检查每一格的理论值是否够大, 如理论值小于 5 时, 应将附近两组或几组合并使用数值增大后, 再进行 X^2 测验, 否则易导致错误结论。当只有两项对比 (4 格) 而不能合并时, 如理论值小于 5, 则应进行校正。校正公式为 (其他公式可参考统计专著):

$$X^2 = \frac{(A - T - 0.5)^2}{T} \quad [\text{公式 } 2-29]$$

3. 这只列举了 X^2 测验的基本公式。从此基本公式还可根据不同需要演变成许多公式，使计算更为简捷。

下面通过实例来分析具体计算过程。假如有某课题组对 213 名工人的操作效果进行了观测，比较新、老二种操作方法的优劣。表 2-28 为完成某一工作程序所犯的动作错误的情况。从表上可见，有 116 例新法错误数小于老法（用符号“+”表示），28 例老法错误数小于新法（用符号“-”表示）。其他 69 例两者计数相同（用符号“0”表示）。这一统计资料因此出现许多 0 和 1 计数，显然总体分布不是正态。而且在统计资料中（未列出），老法的错误数各人变动在 0~12 之间，新法则在 0~30 之间，方差差别也大，因此 t 检验不大合适，这时可以应用非参数的符号检验法检验两种方法差别有无显著意义。统计检验的无效假设有两种方法操作效果相同，即应出现“+”与“-”的概率相同。

表 2-28 两种操作方法完成工序中的动作错误数

序	老法	新法	符号
1	5	7	-
2	5	0	+
3	6	1	+
4	0	0	0
...
211	1	0	+
212	0	0	0
213	1	1	0

（采自杨治良，1985）

检验时可用符号表示，差为“0”者可不列入计算。这样，共有 144 例两者错误数不等，按无效假设应正负号各有一半，即 72 例。故理论值 T 为：72 例。

“+”号理论值一经求出，则“-”号理论值也同时被确定，故自由度为 1。本例应用校正的 X^2 公式（2-19），获得如下计算：

$$X^2 = \frac{(|A - T| - 0.5)^2}{T} = \frac{(|116 - 72| - 0.5)^2}{72} + \frac{(|28 - 72| - 0.5)^2}{72} = 52.56$$

当 $n=1$ ， X^2 ，因为 $P < 0.01$ ，故差别有极其显著意义。

本例在实际计算时还可以用以下简便公式：

以“+”号数为 a，“-”号数为 b，则：

$$X^2 = \frac{(|a - b| - 1)^2}{a + b} = \frac{(|116 - 28| - 1)^2}{116 + 28} = 52.56$$

二法所得结果相同。从这一例可以看到符号检验法极为简便。若用 t 检验法处理本例，则就十分麻烦。符号检验法仅是非参数检验中的一种方法，

主要用于成对资料的显著性检验。另外，还有一种符号等级检验法，此法是上述方法的改进，主要用于配对资料的检验，由于用了差数的大小，故效果比符号检验法更好些。

非参数统计方法有许多优点，除了可以应用于许多总体分布不明确的情况外，由于非参数统计方法在收集资料时可用“等级”或“符号”来评定观察结果，因而收集资料也十分方便，在分析时也可以应用“等级”或差异的“正负号”，因而一般都比较简便而易于掌握，但如果资料的总体分布接近某一有标准理论的分佈（如常态分佈），或资料可以转换成这种分佈，那么非参数方法效果较差。此时如无效假设是正确的，非参数法与参数法一样好；但如无效假设是错误的则非参数法效果较差，如需检验出同样大小的差异往往要较多的资料。

第二章到此为止，我们对实验设计和统计处理作了梗概的介绍。实验设计和统计分析都已构成了独立的学科，而且正日新月异地飞快发展着。因此，以上的介绍只能是非常初步的，其目的只是为学员们在撰写实验报告时提供基础知识。

本章摘要

1. 实验设计乃是进行科学实验前做的具体计划。它主要是指控制实验条件和安排实验程序的计划。它的目的在于找出实验条件和实验结果之间的关系，做出正确的结论，来检验解决问题的假设。

2. 实验设计根据自变量的多少，各自变量内处理水平的多少，和被试情况的不同，而构成不同类型的实验设计。

3. 实验设计大体上分为三类：被试者内设计，被试者间设计，混合设计。被试者内设计乃是指被试者在自变量发生变化的所有情况下接受实验。被试者只接受多个自变量情况中的一个，即不同的被试者接近不同自变量的情况，则称为被试者间设计。兼有被试者内设计和被试者间设计的实验设计为混合设计。

4. 多自变量是指一个实验中包含有两个或两个以上的自变量。它不是指同一自变量的多个水平。多自变量实验具有三个明显的优点：（1）工作效率高；（2）实验的控制较好；（3）实验结果更有价值。

5. 当一个自变量的水平受到另一个自变量的水平的不同影响时，交互作用就发生了。在有交互作用的情况下，分别讨论每一自变量的效应就不够了。此情况下还必须分析讨论出现交互作用的原因和后果。

6. 多因素实验设计是指在同一实验里可以同时观测两个或两个以上自变量的影响，以及自变量与自变量交互作用效果的实验设计。在心理学实验中，居多的是多因素实验设计。

7. 拉丁方设计是多自变量实验设计中较为常用的设计方案。只要是实验中自变量的个数与实验处理水平数相同，而且这些自变量之间没有交互作用存在时，都可采用拉丁方设计方案。拉丁方设计能抵消实验中因实验顺序、被试差异等所造成的无关变量的效果。

8. 统计表是对被研究的心理现象和过程的数字资料加以合理叙述的形式。它在叙述统计资料方面有着重要作用，有人称之为统计的速记。统计表是由表题、横行和纵栏、数字资料等要素组成。统计表可以以形式及内

9. 统计图乃是依据数字资料，应用点、线、面、体、色彩等绘制成整齐而又规律，简明而又数量化的图形。常用的统计图形有曲线图、条形图、直

方图、点图、圆形图等等。

10.偶然误差是指实验中无法控制的偶然因素所引起的误差。有时在实验中还会出现另一种类型的误差，它的观测值不是分散在真值的两侧，而是有方向性或系统性的，这就是系统误差。

11.对数据的概括了解，在统计学上常由二种趋势来度量，一为集中趋势，一为离中趋势。度量集中趋势的统计量称集中量，度量离中趋势的统计量称差异量。集中量有平均数、中数和众数。差异量有全距、平均差、四分差、百分位差、标准差和方差等。

12.显著性检验的主要用途是检验两个或两个以上样本的统计量是否有显著差别。一般可按三步进行检验。第一步，提出假设或假定样组的平均数是从全域中取出来的。第二步，通过实际计算，求出 t 、 F 、或 χ^2 等值。第三步，对假设作出取舍的决定。

13.在心理学实验中，两项实验结果之差，有时是随机引起的差异，有时则是由自变量所造成的。 t 检验就是分辨随机差异与自变量引起的差异的常用手段之一。

14. F 检验是以数据的方差分析为基础的，故又称方差分析。 t 检验只能对两组的平均数加以比较，而方差分析能对二组或二组以上的平均数加以比较。

15.在统计学上，检验分参数检验和非参数检验。 χ^2 检验属于非参数的统计量数。它所检验的分配在先验的假定上并不要求具有一定的形态。

建议参考资料

1.杨治良等(1981):再认能力最佳年龄的研究。心理学报,第13卷第1期,42~50页。

2.张春兴(1989):张氏心理学辞典。台北市:东华书局(繁体字版)。上海市:上海辞书出版社(1992)(简体字版)。

3.张厚粲(主编)(1988):心理与教育统计学。北京市:北京师范大学出版社。

4.陈立(1985):习见统计方法中的误用与滥用。心理科学通讯,第8卷第3期,1~6页。

5.陈舒永(1983):关于使用差异显著性考验的几个问题。心理科学通讯,第6卷第3期,35~36页。

6.黄希庭(主编)(1988):心理学实验指导。北京市:人民教育出版社。

7.Christensen, L.B. (1991). Experimental methodology (5th ed.). Needham Heights: Allyn and Bacon.

8.Conrad, E. & Maul, T. (1981). Introduction to experimental psychology. New York: John Wiley & Sons.

9.McGuigan F.G.(1990). Experimental psychology method of research (5th ed.). New Jersey: Prentice Hall.

第三章 反应时间

本章内容细目

第一节 反应时间的性质

- 一、反应时间研究的简史 105
- 二、简单反应时间和选择反应时间 107
- 三、反应时间实验的要求 109
- 四、反应时间的因变量 110
- 五、反应时间的实际应用 113
 - (一) 相关的研究
 - (二) 应用的研究

第二节 测量反应时的仪器和方法

- 一、刺激与反应键 119
- 二、自由落体直尺计时器 120
- 三、单摆微差计时器 121
- 四、时间描记器 123
- 五、机械钟表计时器 123
- 六、电子计时器 124
- 七、特殊摄影 125

第三节 影响反应时间的因素

- 一、反应时间受刺激变量影响 126
 - (一) 因刺激的不同类型而异
 - (二) 因刺激的强度不同而异
 - (三) 因刺激的复杂程度而异
 - (四) 因刺激呈现方式不同而异
- 二、反应时间受机体变量影响 134
 - (一) 适应水平
 - (二) 准备状态
 - (三) 练习次数
 - (四) 动机
 - (五) 年龄因素和个体差异
 - (六) 酒精及药物的作用

第四节 用反应时间分析信息加工的方法

- 一、减数法 141
 - (一) 什么叫减数法
 - (二) 减数法的典型实验
- 二、加因素法 151
 - (一) 什么叫加因素法
 - (二) 加因素法的典型实验
 - (三) 开窗实验

本章实验

- 一、简单反应时间实验 157
- 二、选择反应时间实验 159

本章摘要

建议参考资料

反应时间研究的历史比实验心理学还要早，它曾有过一段有趣的记事。早期的天文、生理和心理学家对反应时间的研究都曾有过贡献。18世纪末至19世纪初，天文学家已注意到不同观察者观测星体运行的时间存在着个体间的差异。当时天文学界盛行布雷德利（Francis Herbert Bradley, 1846~1924）的“眼耳”法观测星体通过望远镜铜线位置的时间。观察者掌握此法要有眼耳的协作和准确的空间判断，这无疑是一件困难的事。1796年，英国格林威治天文台台长马斯基林在观察星辰经过望远镜中的铜线时，多次发现其助手金尼布克比他自己观察的时间慢约半秒钟，台长认为这是重大的错误，因而辞退了助手。这位天文学家觉察到了反应时间这一心理现象，可惜他并没有进行深入的研究。二十六年之后，德国天文学家贝塞尔（Friedrich Wilhelm Bessel, 1784~1846）见到此事的报导，发觉这个现象的意义，才正式加以科学研究。他比较了自己和其他天文学家观察同一星体的通过时间，也发现有明显的差别。1823年贝塞尔与另一位天文学家阿格兰德共同观察七颗星，B是贝塞尔的反应时间，A是阿格兰德的反应时间，二人反应时间的差别如下：

$$B-A=1.233 \text{ (秒)}$$

这个等式即著名的人差方程式（或个人方程式）（personal equation），它反映着两个观察者之间的个体差异。这一发现引起了天文学家经久不衰的兴趣。他们确定了不同观测者的人差方程式及其校正方法。

反应时间直接作为心理学研究的课题开始于冯特。在1879年，他在莱比锡首创了心理实验室，从此以后，对简单和复杂反应时间进行了一系列研究，为心理学作出了贡献。从冯特至今天，实际上这个变量这样普及，以致反应时间的研究已成为一个专门的研究领域。当心理学的研究人员和实验心理学家们聚在一起的时候，他们会异口同声地说：“我擅长的是反应时间的研究”。反应时间这样重要，因为它的研究不仅是一种工具，而且它自身也是一个研究课题。

以上简单说明，反应时间是实验心理学上的重要课题。本章之内容，即将对此具有浓厚兴趣的主题范围，探讨以下六个重要问题：

1. 什么叫反应时间？反应时间在认知心理学研究中的重要性如何。
2. 减数法的原理是什么？并引用实验加以说明之。
3. 加因素法的原理是什么？并引用实验加以说明之。
4. 简述测定反应时间的仪器种类，及其不断更新情况。
5. 影响反应时间的因素主要有哪些。
6. 什么叫速度——准确性权衡。

第一节 反应时间的性质

一、反应时间研究的简史

反应时间（或反应时）（reaction time，简称 RT），它是一个专门的术语，不是指执行反应的时间，而是指刺激施于有机体之后到明显反应开始所需要的时间。反应时间是实验心理学常用的反应变量之一。反应时间这一术语，最先是由生理学家提出的，当时称之为心理过程的“生理时间”。按实验心理学传统的理解，反应时间是指从刺激作用发生到引起机体外部反应开始动作之间的时距，它包含以下几个时段（phase）：第一时段，刺激使感受器产生了兴奋，其冲动传递到感觉神经元的时间；第二时段，神经冲动经感觉神经传至大脑皮质的感觉中枢和运动中枢，从那里经运动神经到效应器官的时间；第三时段，效应器官接受冲动后开始效应活动的时间，以上三个时段的总和即是反应时间。可见，反应时间可由多种因素合成，它在心理学实验中可以作为测定反应变量的一种指标。虽然反应时间这一问题，最早是天文学家提出来的。早期天文学家曾对反应时间有不少讨论和研究，但作为反应时间的实验，一般地说，是 1850 年由著名生理学家赫尔姆霍兹（Hermann Von Helmholtz，1821~1894）发明的。赫尔姆霍兹是实验心理学的奠基人之一。他成功地测定了蛙的运动神经的传导速度（约为 26 米/秒）。其后，他又测定了人的神经的传导速度约为 60 米/秒。这与后来穆乃奇（Munnich，1915）测定的每秒 66~69 米颇为接近。根据神经传导的大致速度，他认为神经传导所占据的时间是很短的，而整个反应时间却比较长且变动很大。

在 1865~1868 年间，荷兰生理学家唐德斯（Franciscus Cornelis Donders，1818~1889）第一次企图研究心理因素如何影响一切简单的与复杂的反应。他的作法就是把上述心理过程交织在刺激和反应中间，从而考察其结果，并比较简单与复杂的反应。在简单反应的实验中，实验者先告知被试者将有什么样的刺激出现，比如一种颜色光或一种声音，要他觉察到哪种光或哪种声音出现时，就用一个手指按一个反应键。在辨别反应实验中，他们用两种不同的刺激，如两种不同颜色的光或两个高低不同的乐音，要求被试者只反应其中之一，而对另一种不反应；还有一种方法是使被试者对于甲乙两种刺激，准备两种不同的反应，比如看见甲刺激（红光）时用右手反应，看见乙刺激（非红光）时用左手反应。总而言之，他创造了选择反应时间的实验，发现这种反应时间比简单反应时间约长 100 毫秒。唐德斯认为一个复杂反应只是在一个简单的反应上加一些别的动作，这些动作所需时间可用反应的全部时间减去简单反应所需时间来求得。这个时间差就是上述心理过程所需要的时间。奥地利生理学家埃克斯纳（Exner，1873）指出被试者在反应时间实验中定势（或心向）（set）的重要性。

反应时间直接作为心理学研究的课题开始于冯特（Wilhelm Wundt，1832~1920）。1879 年，冯特在莱比锡大学首创了心理实验室。当时他便认为唐德斯指出了实验心理学的一条重要途径，即心理活动的时间测定工作。冯特的学生对简单和复杂的反应时间进行了一系列工作，但他们在注意、知觉、联想和选择等过程上却未测出确切的反应时间。在冯特的早期学生中后来有两位学生建立了研究反应时间实验室。冯特的学生卡特尔（James McKeen Cattell，1860~1944）作了许多关于反应时间的实验。他认为被试者在做

简单反应测验时，其注意力完全集中于那个将出现的刺激和那个将动作的手指。当刺激来到时，眼睛 大脑 手指之间的神经通路早已准备好了，反应的时间就快。在辨别和选择反应的实验中，需要有更多的神经通路接通的准备，这时被试的心理状况比较复杂，会产生焦虑、怀疑等复杂的心理状态，所以反应时间就会延长。卡特尔先在德国然后在美国对反应时间作了广泛而系统的实验研究，其中不少材料至今仍为人们所引用。20 世纪初，德国心理学家屈尔佩 (Oswald Kulpe, 1862 ~ 1913) 在符兹堡对简单与复杂的反应发展一种内省的研究，他的学生证明了准备定势的选择影响。法国心理学家皮耶隆 (Henri Pieron, 1881 ~ 1964) 对反应时间的研究也作了贡献。但总的说来，自卡特尔之后心理学家对反应时间研究的兴趣已不在于分析它的原因，而转向测量技术的改进方面，以及深入到应用的实际领域中去了。

现代心理学家在总结反应时研究的这段历史时，把自 1850 年赫尔姆霍兹的研究至 1969 年长达一百多年的时间称之为唐德斯反应时 ABC 时期。这是反应时研究的第一阶段，这一阶段的方法学的核心是减数法（见本章第四节）。1969 年心理学家斯顿伯格 (Sternberg, 1969) 提出了加因素法（见本章第四节）之后，反应时研究便进入第二阶段，开始了反应时间研究的新时期。

二、简单反应时间和选择反应时间

反应时间的种类很多，分类方法也很不一致，但最常见的反应时间有简单反应时间和选择反应时间两类。它们是以刺激与反应的不同数目进行分类的。

简单反应时间 (simple reaction time) 是给予被试者以单一的刺激，要求他作同样的反应。被试的任务很简单，他预先已知道将有什么样的刺激出现并需要作出什么样的反应。如以对光简单反应时间的测量为例，在一弱光照明的室内，被试端坐在桌前，面对一个屏幕，注视屏幕上的一个孔（通过这个孔可以呈现灯光）。事先呈现灯光给被试看看，让他熟悉这个刺激。桌上放一电键，指示被试当他听到预备信号时即将手指放在电键上，当灯光一呈现就立即按下电键。屏幕后是主试操纵仪器的地方，使用计时器来测量刺激到反应的时间。最初测得的反应时间可能长达 0.5 秒，多次测定之后很快会降至 0.2 ~ 0.25 秒，再后可能会降至 0.2 秒以下，但无论如何练习不能减至 0.15 秒以下。在经过一定练习之后，对听的简单反应可能到 0.12 秒，触觉则可能更短些。这是经过大脑的随意运动反应的潜伏期。有些非条件反射，特别是膝跳反射和眨眼反射却特别快，其潜伏期约为 0.04 秒。总之，简单反应的时间是比较短的。

选择反应时间 (choice reaction time) 是根据不同的刺激物，在各种可能性中选择一种符合要求的反应。如安排红和绿两种不同的色光刺激，有两个反应键放在被试面前，规定其用左右手指各放在一个键上，并用右手反应红光，用左手反应绿光，这是选择反应测量的典型示例。显然，在这一选择时间里包括了被试的辨别和选择活动所花费的时间，他必须在作出反应之前对不同刺激有所辨别，并作出不同反应的选择。一般说来，选择反应时间总比简单反应要长，就是由于选择性反应的中枢活动比较复杂，需要进行一定的思维活动，作出选择，执行正确的反应动作。生活中的动作大多属于这一类反应。最初级的选择反应只是在两种可能性中选择一种反应，对于视觉

刺激物的反应时间就在 0.25 秒以上。反应的选择余地愈大，反应速度就愈缓慢。

关于选择（或辨别）的数目与反应时间的关系，早期心理学家就作过不少研究。默克尔（Merkel，1885）曾揭示了二者之间的函数关系，并用下列公式来表示，式中 N 为分辨信号的数目。

$$RT = \lg N$$

于 20 世纪 50 年代，心理学家卡克尼（R.M.Gaue）用公认的选择数目反应时曲线图示了选择数目与反应时间的关系（见图 3-1）。

从图 3-1 可以看到：在有 4 种选择时，反应时间已超过简单反应（1 种刺激选择）一倍；在有 8 种选择时，反应时间超过简单反应二倍以上。

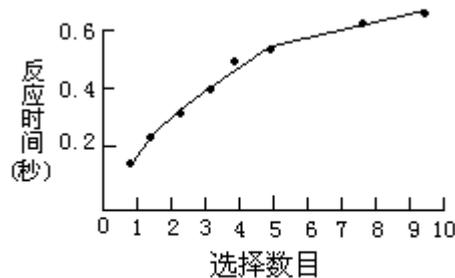


图 3-1 选择数目与反应时间的关系
（采自曹日昌等，1979）

由于信息论的广泛应用，在反应时间的研究方面也采用了新的理论和技术。希克（Hick，1952）关于信息量（注 3-1）与反应时的工作成为这方面研究的一个转折点。他采用了申农等（Shannon et al.，1948）等所提出的计算信息的方法来分析刺激的不确定性，发现在刺激所负荷的信息与反应时间之间存在着恒定的关系。他把这关系概括为以下的公式：

$$\text{反应时} = a \lg(n+1)$$

注：3-1：信息量（或讯息量）（information）乃是对各种各样消息的一种统一称谓，其单位比（bit）。它表示消息所具有的概率价值，可用公式表示如下：

$$I = PB/PA$$

I：代表收到的信息量

PB：代表消息传到后收信者所视某一事件的事后概率

PA：代表消息传到前收信者所视同一事件的先验概率

式中 a 为一个常数，n 为等概率出现的选择对象的数目。所以用 n+1 是因为在实验中刺激呈现的间隔是不固定的，增加了信源的不确定性。在这种情况下，被试不仅要考虑出现哪一个刺激（ S_1, S_2, \dots, S_n ），而且还要考虑是否出现刺激，故多一个选择对象（ $S_0, S_1, S_2, S_3, \dots, S_n$ ）。

希克用这一公式去分析 80 年前默克尔所进行的类似研究，同样得到了相符的结果。

三、反应时间实验的要求

进行反应时间实验，除了和其他心理实验一样，都要很严格地控制实验条件，像刺激的强度和刺激持续的时间等等。除此之外，进行反应时间实验研究，要讲求策略思想，应严格遵循一定的程序和要求，方能取得预期的效果。反应时间作为定量化实验的反应变量，必须要测量得准确可靠，并要有足够的精确度，这除了在下一节将讲到对实验仪器的要求外，研究者必须把握下列两个要点：

1. 对被试反应的要求 首先从被试反应要求上看，应避免出现过早反应或其他错误的反应。由于被试总是希望尽快反应，他可能约束不住自己的手，而在刺激呈现之前就作出“反应”，尤其是在刺激与预备信号之间的时距保持恒定的情况下，这种现象更是常见。因此，测定反应时间的标准程序就是每次呈现刺激前要略为改换一下这个时距的长度，以避免被试对“恒定时间”这一额外变量作出了错误的反应。可是要区分快速的正确反应与抢先的“假反应”往往是不太容易的。防止出现“假反应”的有效措施是在实验中插入侦察试验（detection test），即给预备信号之后并不呈现刺激。例如实验安排 20 次反应时间测定为一组，主试便可在这 20 个刺激中插一或两个“空白刺激”，如果被试对此仍作出反应，那就向他宣布这 20 次实验的结果全部无效，被试需按规定要求重做一组实验。至于选择反应时间就不必插入这个“侦察试验”，因为被试倘若出现“抢步”的话，他就作出错误的反应，主试可据仪器反馈信号立即发现问题予以警告。

2. 选择反应数目需要与辨别的刺激数目相等 其次，选择反应时间实验的要点是务必使选择反应数目和要辨别的刺激数目相等，即每一种反应必须是针对同它相应的预定刺激而作出的。如刺激 A 引起反应 a，刺激 B 引起反应 b 等。倘若忽视这一点，实验中只有一个反应键，而刺激却不止有一个，可能是红光，也可能是黄光，要被试在辨别出是红光还是黄光时尽快按压反应键，而不要在辨清刺激前作出反应。从表面上看，这样的安排似乎切实可行，被试在辨清色光刺激之前不作反应，而所作出的反应也不是对刺激的简单反应。但在预试之后，会很快发现结果是混乱的，这样的步骤并不能保证避免被试只作简单反应，也无法防止被试为了不致过早反应而不恰当地延长了反应时间，问题在于主试难以了解被试对刺激的辨别已到何等程度。解决的办法是有几种刺激，就安排几种反应，每种反应只对应于其中某一特定的刺激。

四、反应时间的因变量

反应时实验中有二个基本因变量（或依变项）（dependent variable），即速度（speed）和准确性（accuracy）。反应时间实验中的一个突出问题就是权衡反应速度和反应准确性的相互关系。人们都有这样的常识，当一个人很快去完成某件事时，他会比慢慢地做某件事犯更多的错误。反之，如果某人很正确地做某件事时，速度上就会变慢。心理学家称这种关系为速度-准确性权衡（speed - accuracy trade off）。这是反应时实验中因变量上一个十分重要的问题。

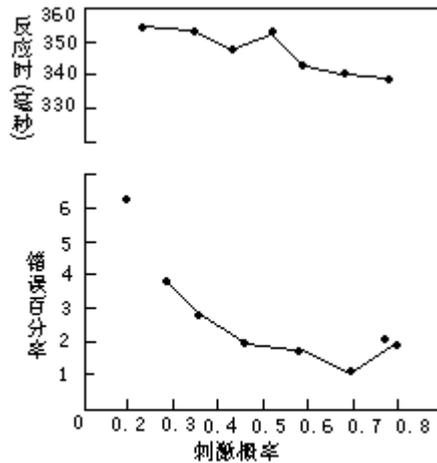


图 3-2 反应时和错误百分率作为刺激概率的函数
当反应时基本恒定时,错误百分率随刺激
的降低而升高。(采自 Theios, 1957)

不少心理学家通过实验来说明速度-准确性权衡的道理。这里先举泰奥斯 (Theios, 1975) 的实验为例。泰奥斯对刺激呈现的概率和反应时的关系进行了研究。实验中,呈现的是视觉刺激,每次出现一个数目字,被试的任务是对某一个特定的数字(例如个位数 4)作出反应,而对其他数字则不作反应。实验的自变量是这个数字出现的概率,概率变化范围是 0.2~0.8。换句话说,特定的数字(例如个位数字 4)在一系列呈现的刺激中出现的百分率是 20%至 80%。参见图 3-2。泰奥斯的结论是,刺激(指特定的信号刺激)的概率不影响反应时间,如果丢开反应准确率这个情况,这个结论是颇有道理的,但是,我们只要看一下图 3-2 的下半部分就不难发现,实验结果中还存在着如何看待错误百分率的问题。虽然这一问题表面上看来很简单,但实验上却关系到实验的基本结论。从表面上看,如果把各种概率的错误率平均起来,则平均错误率为 3%,这个数字不能说是高的。但是进一步分析我们就会注意到:第一,在这个实验中被试的任务是相当简单的;第二,实验对象是具有较高文化水准的大学生;第三,更重要的是,错误率的变化是有规律的,最高的错误率(6%)发生在最低的刺激概率上,并且反应的错误率随刺激概率的增加而降低。由此可见,速度和准确性之间存在着内在的联系。

我们再举一个实验以说明如果所有不同概率刺激的反应错误做到相等的话,那么反应时又将出现怎样的变化。帕彻拉 (Pachella, 1974) 的实验结果表明,为了使错误率降低到 2%,对 0.2 概率刺激条件的反应时将增加 100 毫秒。这说明根据速度-准确性权衡,在低刺激概率条件下,为了减少错误率,反应时间就会延长。可见,在西奥斯的实验结果中,一旦考虑了反应错误率,那么单说刺激概率不影响反应时,是有问题的,是不全面的。在此情况下,二个因变量(速度和准确性)都应考虑,这就是速度-准确性权衡技术。

奈特和凯托威茨 (Knight & Kantowitz, 1974) 的实验更明白地阐述了速度-准确性权衡的道理。他们用两个间隔很短的刺激作为自变量呈现给被试,这种刺激称之为间隔刺激。因变量取二个,一个是反应时间,一个是反应错误百分比(见图 3-3)。起先,奈特和凯托威茨发现,对于不同的刺激间隔反应时间是恒定的。如同上例泰奥斯发现刺激概率不影响反应时一样。但是当奈特和凯托威茨绘制刺激间隔和错误百分比的关系图时,一种特定的关系被发现了:刺激间隔愈短,错误率就愈高。这里我们再次要问,如果在

各种不同的刺激间隔条件下,要保持相同的错误率,那么反应将是怎样的呢?根据速度-准确性权衡的原则,在短的刺激间隔情况下,要保持与长的刺激间隔相同的反应错误率,反应时将会延长。因而,起先得出反应时不受刺激间隔长短的影响的结论乃是片面的,甚至是不对的。

上述实验都清楚地说明了这一点。在进行反应时间实验时,应考虑速度和准确性两个指标,当然,有时可以同时选择二个指标,有时可以选择其中一个指标。但是在只选择其中一个指标时,应对另一个指标有所交代,说明其可以忽略不计的原因。这是反应时实验中的一种重要技术。

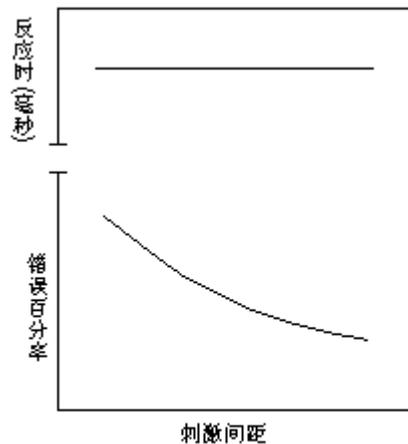


图 3-3 反应时和错误百分率作为刺激概率的函数当反应时基本恒定时,刺激间距越长,错误百分率越低。(采自Knight & Kantowitz, 1974)

五、反应时间的实际应用

(一) 相关的研究

反应时间作为反应变量,不仅在各类反应时间之间存在着一定的关系,而且也常常同非反应时间的反应变量之间存在着多样的关系。对反应时间相关法的研究,可对有关的反应变量作出相互预测,还有助于增进人们对反应时间的可能影响因素的了解。席斯克(Sisk, 1926),拉尼尔(Lanier, 1934)以及福尔贝斯(Forbes, 1945)等早年在这方面做了许多研究工作,取得了相近的实验结果。如以同一手指对光、声和触觉的简单反应时间,两两间的相关系数是在+0.43 到+0.85 之间。对声音的反应,两手之间的相关系数是+0.92,两足之间是+0.93,手与足之间是+0.81。一般认为,对某个人来说,他的各种反应时间变量之间的高相关有很大的一致性。但是我们不能过分看重这些相关,如画字、敲击、目标追踪等许多快速反应是由一系列刺激所引起的,这些作业的时间并不是完全取决于简单反应时间的长短,因此还不能把简单反应时间当作连续性作业速度的有效指标。赖伊(Lahy, 1939)曾经设计了一个系列反应实验,要求被试对单一的听觉刺激作出二个反应。从刺激到按下动作开始为反应时间;从按下动作开始到再释放这个键的动作开始为彭左恢复时间(Ponzo's recovery time),结果发现,反应时间和恢复时间的相关差不多是零。因此反应时间显然并不是一个人连续工作的速度的一种满意指标。

近年来,我国心理学家对这一课题也作了不少研究。陈舒永等(1986)研究了被试为大学生业余运动员和非运动员共 80 名,以声音刺激分别测定简

单反应时间和运动时间(从开始运动到运动完成所经历的时间)。结果表明,反应时间、运动时间及两者的总和时间在右手和右脚之间的相关系数分别为+0.72、+0.68和+0.79,且都在0.01水平上显著。说明这三种反应指标在手与脚(同侧)之间都可以相互预测。实验还发现简单反应时间和运动时间两者在同一反应肢体虽有显著的正相关($r_{手}=0.21, p<0.05; r_{脚}=0.29, p<0.01$),由于相关系数都很小,如对两者作相互预测,意义是不大的。

在运动心理学领域,洪伟等(1986)对有文化的成人共98名被试,作视听简单反应时间和短时记忆能力的测定。结果表明,反应时间和短时记忆这两种心理功能指标之间有较强的负相关,且都有极显著意义($r_{视、记}=-0.46, r_{听、记}=-0.36; p<0.001$)。这说明有了其中一个指标,在某些条件下,就可估计另一个指标的情况。

(二) 应用的研究

上面讲到,反应时间是有关人的活动、动作的速度和准确性特征最基本的内容。反应时间的实验研究,的确有助于人们从不同方面对自身更为复杂的心理与行为的特点作出定量分析,从而获得规律性知识。心理学研究和实际领域中应用反应时间的场合是很多的,下面分别作简要介绍。

1. 应用心理现象的研究 任何心理活动都需要一定时间。几乎所有心理学的实际问题的研究,都可以应用于反应时间的原理和方法。传统心理学中许多问题的探索,往往把反应时间作为一项客观指标,如辨别、联想、注意的分配、动机的引起、知觉的范围、定向运动、感觉强度、心理量表、情绪的差异及机制、条件反射、学习与词的替代等等。学者已积累了大量资料。现代心理学则对此进一步作出探索,取得了一个又一个学术成果。

我国心理学家曹传咏等(1963)曾以反应时间作指标,对小學生辨认汉字字形的特点进行了实验研究,刺激字有五个单元,每个单元包括一个单字及含有这个单字的词语和字组各三个,词语和字组分别由二、三、及四个字组成。五个单元共35张卡片,按随机次序在呈现器上呈现,见表3-1。

表 3-1 五个实验用的字、词语和字组的单元

实验单元	1	2	3	4	5
单字	方	理	海	满	放
词语	前方	理想	海洋	自满	放哨
	东方红	讲道理	海岸线	满天飞	解放军
	四面八方	理直气壮	移山倒海	心满意足	心花怒放
字组	方常	理提	海肯	满视	领放
	成旗方	问都理	山挂海	满摇产	放造界
	铁经方板	能技理要	海最为理	本满树谁	把学放双

(采自曹传咏等, 1963)

呈现时间从5毫秒起始,以2.5毫秒一级累增。在每一呈现时间内由被试观察三次,直到连续三次正确辨读为止,并以其中第一次正确读出所需的呈现时间作为时间阈值,所得结果列于表3-2。

实验表明,单字与2~4字词语之间辨认所需的时间没有显著差别,而

表 3-2 对不同刺激字正确辨认的时阈平均数

字类	单字	词语			字组		
		二字词	三字词	四字词	二字组	三字组	四字词
时间 (毫秒)	9.0	9.8	9.7	11.0	12.2	13.1	15.9

(采自曹传咏等, 1963)

单字与 2~4 字组之间的差异是十分显著的。这可能是由于中枢已形成的联系系统, 对字词的辨认产生了有利的影响。另外, 他们还研究了生字和熟字的反应时间, 见表 3 - 30

表 3 - 3 正确写出呈现字的时阈平均数

刺激字	熟 字							生 字						
	南	盖	闾	棉	照	幸	得	走	闾	侃	悖	煦	肇	紫
反应时间 (毫秒)	10.4	9.6	10.0	10.0	10.0	11.7	8.3	33.3	50.0	13.3	20.4	30.4	49.2	36.7
平均 反应时间	10.1							34.3						

(采自曹传咏等, 1963)

彭瑞祥等 (1983) 应用视听反应时间方法, 对由于掩盖而造成不完整的汉字进行再认的研究实验, 材料为 39 个常用汉字, 分整体字、左右分隔字及上分隔字三类, 每类 13 个。笔划 8~10, 有以横竖为主, 也有以撇捺为主的。10 名被试具有中等以上文化程度, 分成甲、乙两组。在速示条件下要被试尽可能迅速、准确地说出所呈现的字。字的大小为 50×50 毫米, 观察距离为 2.5 米。甲组再认左上角被掩的字, 乙组再认右下角被掩的字。实验结果列于表 3 - 4。由表可见, 对三类汉字的再认, 甲组的平均反应时间都长于乙组。表明字的左上角的特征或信息比右上角的较重要, 这一现象在可分隔型的字, 特别是左右分隔型的字表现尤其明显。甲组被试反应时间长, 再认字困难, 原因就在于此。这类研究对于识字教学、计算机识别汉字等工作无疑是有意义的。

自 70 年代起, 反应时间法也成功地应用于现代记旧心理的研究。斯

表 3 - 4 两组被试再认不完整汉字的反应时间

	左右分隔字		上下分隔字		整集结构字	
	\bar{X}	SD	\bar{X}	SD	\bar{X}	SD
甲组	0.84	0.330	0.72	0.81	0.72	0.228
乙组	0.71	0.219	0.66	0.161	0.66	0.163

(采自彭瑞祥等, 1983)

顿伯格 (Sternberg, 1970) 用反应时间作指标, 对短时记忆中信息提取的搜寻过程特点作了研究。先向被试呈现不同系列长度 (在记忆广度之内) 的数字, 而后再呈现一个检验数字, 要他将这个检验数字与记忆中的系列数字的各个项目作比较并尽快作出 “ 是 ” 或 “ 否 ” 的回答。结果表明, 短时记忆中存储信息的提取时间随系列项目数的增加而延长, 两者呈线性关系。大约每增加一个记忆项目, 从记忆中搜寻和提取有关信息的时间就要多花 38 毫秒。说明短时记忆信息的提取需要搜寻, 而搜寻过程的型式是逐个的 “ 系列比较 ”, 并不是同时的 “ 平行比较 ”。我们在第四节还将详细介绍这个著

名实验。

2. 应用于实际领域 反应时间是心理学上最重要的反应变量和指标。在现实生活中，有许多领域需要研究人的反应时间，尤其是在工业生产、交通驾驶、体育运动、设备设计和科学管理等实际部门，反应时间的应用更有其重要意义。

在工业生产中，要能提高工作效率，节省人力，就需要研究操作动作分析的科学工作方法。例如，纺织工人生产定额的制度、劳动指标的测量以及各项操作动作的分析使之正确、合理化，在这些场合上，都用得着反应时间指标。

在交通运输行业中，反应时间对于交通驾驶员来说是非常重要的，过慢或有时过快的反应，都可能导致事故的发生。例如有实验表明，汽车驾驶员需要快煞车时，平均要花 500 毫秒，其中约有一半的时间是属于反应时间本身，另一半是将脚从加速器移到煞车踏板的活动消耗了。如果司机的反应延长了 0.1 秒，车速仅为 30 公里/时，汽车还是要往前多行近 1 米；如果反应延长 0.2 秒，那就要多行 2 米，在这种条件下反应时间就变成严重事故的原因了。

现代科学技术的突飞猛进，使自动化设备的设计需要反应时间研究提供必要的参数和要求。如飞机驾驶舱里的表盘技术模型的制作，可以通过反应时间的测定结果加以鉴定。研究表明，各种显示器、操纵的设计，要适合人的身心特点，使操作者能迅速、准确地对信息作出辨读和应答，而且两者的空间位置的安排结构要合理，以保证操作的顺利进行。为此，有人设计了最佳表盘和弧形仪表盘，这类研究对于提高人机工作效率，确保飞行安全是必不可少的。

在体育运动上反应时间的应用更是大家所熟悉的事。有经验的教练员和教师常研制一套速度功能测验器，用来训练运动员的各项动作的反应时间，以不断提高运动员对感觉运动时间和动作时间的领会和控制能力，争取创纪录的成绩。这也是体育运动的宗旨之一。

由此见，反应时间的研究是社会生产和科学发展需要的产物，它对心理学的理论和实际应用都有重要意义。随着生产自动化的发展和人们生活节奏的加快，对反应时间的研究越将显现出它的重要性。

第二节 测量反应时的仪器和方法

从反应时间研究的简史来看，随着近代科学和实验技术的不断发展，实验仪器不断得到更新，使这类研究不断得到提高。因此，19世纪冯特时代的心理学仪器不可避免地带有深刻的机械时代的烙印（见图3-4），而现代心理学仪器则反映了电子时代和计算机时代的特征。

任何一种有效的测量反应时间的仪器都包括三个部分：刺激呈现装置，反应装置和记时装置。这三部分装置都会影响到实验结果的精度。因此，反应时间研究的仪器和手段要根据研究课题的要求选定。为了粗略测量一个继续几秒钟之久的反应，一个有0.2秒刻度的停表就够精确了。当然，为了测量小于1秒钟的反应时间，就需要精密的测量和某种自动记录的方法，此时就不能用人工停表，因为这种人工的机械停表本身就有0.2秒左右的反应时，从而会影响实验结果。要使测量值达到一定的精细程度，仪器的读数就要更精细些，一般说来，比较理想的记时器要能够达到精确和自动记录两项要求。下面我们简单介绍几种有关测定反应时间的器材和仪器。

一、刺激与反应键

反应时间实验，首先是如何设计刺激和反应同时开闭的电路装置。因而选择合适的刺激和反应键就十分重要。一般地说，简单线路和电键大半有一定的缺点和操作不方便的毛病。例如，用普通电键作为刺激键作视觉反应实验时，如主试、被试在同一间实验室里，刺激键在操作时发出的音响会成为一种不应有的额外刺激，影响实验结果。如用电子仪器，除对记时要求准确外，对刺激与反应装置的构造与注能的要求也不容忽视。因此，一个好的电路设计，对接通和断开记时装置的刺激键与反应键的型式应加以选择。

对刺激键的基本要求是使其适合所刺激的感官。一般机械的琴键式、电报键式或拨动式的刺激键，难免会发出一定的噪音，这种额外刺激将对视觉反应时间有影响。假如是一个简单反应，视觉刺激器又发出声音，那么反应可能是对声音而不是对光发动的，因为对声音的反应要比光快些。氖灯开亮或关熄不需要多大的潜伏期，同时它也没有太明显的声音。因此它为对光的简单反应提供了良好的刺激。听觉的刺激，可以使用音响键或蜂鸣器。

对反应键的要求则更严，至少要考虑到键的机械阻力和被试操作的方便与习惯。如一般琴键式反应键有一定的弹簧阻力而延长反应时间，这对幼儿被试尤其不适用，可换用一种轻触式微动开关，以利获得准确的结果。至于向下按压或向上松开反应键这两种不同反应方式的优劣问题，尚缺乏比较。国内有关的仪器多采用按压式并增设自锁继电装置，这既可减少被试的疲劳又能使主试避免产生混乱。

不同类型反应时间的测定对刺激和反应键都有特殊的要求。如测试言语反应时间，除要设计增益高、失真小的电路系统，还需要性能完善的言语反应键。我国心理学者杨治良（1979）曾设计此类仪器，其基本原理就类似于玩具市场上的声控开关。使用此仪器时，只需在话筒插口处插上话筒，让被试坐于话筒前，在距离话筒0.5米的距离内，用通常两人交谈的语音强度所产生语言信号即可使继电器工作。若继电器输出端接上控制电路，就可使语音信号起电路开关作用，计时精度达到实验要求。

旧式的唇键、声键灵敏度低，现代电声学所提供的声音开关和传声器便可克服这类缺点。研究者还可借用电生理方法，把导出的发声部位生物电直接或经磁带记录输入电信号记录装置和电子计算机进行数据分析，以获得更准确研究资料。随着电子元件广泛应用，我们可以用光电开关借助被试某部分的反应运动遮断或开通一个光源射到光源管上的光线作为反应键。遮断光线引起电流的变化，经过放大即可推动灵敏继电器来开关计时电路了。

在反应时间实验中，选择了合适的刺激和反应键之后，接下去就是挑选最佳的计时器，最后组合成一个刺激和反应组合的电路装置，保证实验的顺利进行。

二、自由落体直尺计时器

一般地说，计时器多是精密的仪器，而且价格昂贵。但是早期，皮耶隆 (Piéron, 1928) 想出一种几乎不用花钱的既简便、无噪音而又相当准确的计时方法。这是一种最简单的计时器，严格地说，这还不是一种仪器而仅仅是一种测量工具。这一方法在于视觉实验时，主试用拇指将一根米尺笔直地按在墙上，尺的零点朝下，尺的上端在墙上跟被试者的平视相齐的记号处；被试者将拇指摆在尺的下端，准备阻止尺自由下落。实验时主试说“预备”，1、2 秒钟后突然松开拇指，放开直尺，被试一接受刺激就立即用拇指再将尺按住。直尺下落的距离，就可折算成被试者的反应时间，计算公式如下：

$$T = \sqrt{\frac{2S}{G}}$$

T：代表以秒作单位的反应时间

S：代表尺子下坠的距离（以厘米或尺为单位均可）

G：代表地心引力的重力加速度（980 厘米/秒²或 32 尺/秒²）

为省略计算手续，可直接在尺上标明上与下落距离相应的反应时间刻度，b 距离可由公式： $S=1/2(GT^2)$ 求得。反应时间在 0.5 秒之内的一些近似对应值，可从表 3-5 上查得。

表 3-5 自由下落距离与反应时间测试换算表

反应时间 (秒)	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
下落距离 (公制：厘米)	4.9	11.0	19.6	30.6	44.1	60.0	78.4	99.2	122.5
下落距离 (英制：吋)	1.9	4.3	7.7	12.0	17.3	23.5	30.7	39.2	43.2

(采自 Woodworth & Schlosberg, 1954)

这种测定方法使用了自由落体的原理，既简易又经济，适合于测定反应时间的演示实验。这种方法也可借助一些附加装置，用来测定听觉、触觉以及其他比较复杂的选择反应时间。在高科技发达的今日，仍不失其为一种简便易行的测时方法。

三、单摆微差计时器

这一测时装置最先是由恺泽 (Kaiser, 1859) 运用于天文学上对人差的测量, 以观察两个摆长不同的摆锤视觉的重合来推算人造星体通过的时间, 其后由桑福德 (Sanford, 1889) 加以改进, 用两摆的视觉合一来测量心理反应时间。根据单摆振动定律 (law of simple pendulum vibration), 单摆的振动周期 (T) 跟摆长 (L) 与重力加速度 (g) 有以下的关系:

$T = 2\pi\sqrt{L/g}$ 。其中n代表两摆的振动周期有一个微小的差别, 根据这个微小

差别, 便可求得反应时间量值。

测量前, 主试先调整两摆的摆长, 例如使长摆每分钟摆动 75 次, 短摆每分钟摆动 77 次, 则两摆每周的时差为 $0.8 - 0.78 = 0.02$ (秒) (0.8 秒是长摆摆动一周的时间, 0.78 秒是短摆摆动一周的时间)。实验程序是这样的: 长摆和短摆均放在释放处上面, 等待释放。主试掌握长摆键, 被试掌握短摆键。在主试发出“预备”信号的 2 秒钟之内, 主试按动长摆键, 此时长摆开始摆动。被试的任务是看到长摆启动后, 就立即按下短摆键释放短摆。由于总是长摆先动, 然后短摆渐渐的追上, 每摆一次长摆被短摆追上 0.02 秒, 那么短摆追上长摆时的摆动次数乘以 0.02 秒就是短摆与长摆发动的时间间隔。设短摆追上长摆所需要的摆动次数为 n, 则 $0.02 \text{ 秒} \times n$ 就是反应时间。例如长摆摆动后经 11 次, 短摆才与长摆重合, 这次的反应时间为 $0.02 \times 11 = 0.22$ (秒)。如果被试者反应太慢, 比如长摆经过一个完全的摆动后, 短摆才开始摆动, 又经过 5 次, 二摆重合, 则可在正常计算外, 另加长摆摆动完整一次的时间, 即 $0.02 \times 5 + 0.8 = 0.9$ (秒)。在记数时, 我们应注意摆一去一回才算完整的一次; 另外在观察二摆重合时, 常常有三次不易分辨谁先谁后, 那么就以中间的一次为准。微差计时器乃是一种“手动”的仪器, 图 3-5 为其原理图。

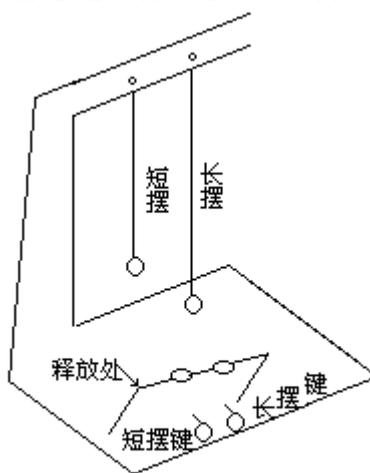


图3-5 微差计时器原理图

四、时间描记器

时间描记器 (chronographic) 又称记纹鼓 (kymograph drum), 是早期生理心理研究中常用的仪器, 当今精密的生理记录仪仍采用其工作原理。时

间描记器利用一个等速运动的表面，如记纹鼓或摄影胶片，来记录刺激和反应的痕迹，并根据两个痕迹间的距离算出反应时间。只要控制运动的表面速度不变，同时在运动表面上画出清晰的时间标记，就能容易而且准确地求得反应时间。我们一般用每秒振动 100 次的电音叉，在其臂上安装一支画笔，连续在运动表面上（纸带）画出时间标记，然后找出刺激与反应间的振动数就能算出反应时间。图 3-6 上二条垂直虚线间的距离就代表反应时间，图 3-6 为早期的记纹鼓。

五、机械钟表计时器

早期使用的钟表式计时器有希普计时器和邓拉普计时器。钟表匠希普（Hipp, 1843）应用摆的原理创制出一种可测量 1 毫秒的计时器，这是早期心理实验室精密计时器的雏形。这种计时器是由两个基本部分构成：（1）一个迅速运动的时钟机构，在刺激没有发出时，就让它先开动达到常态速度，反应后就停止；（2）用一个轻齿和连串的装置来移动实际测量反应时间的指针。刺激和反应都在齿轮上有记录。随着生产和制钟工艺的发展，希普计时器不断得到改进。邓拉普计时器的构造略优于希普计时器，它以电磁离合器代替希普计时器上的弹簧装置。邓拉普（Dunlap, 1918）计时器虽然也很笨重，但为设计更新的钟表式计时器作出了贡献。今天的钟表式计时器虽然日新月异，但是基本部件仍是由一个电动机和一个离合器的改进和更新。电动秒表，电子钟表也均是在其基础上发展而成的，从数量和质量上，为反应时间研究提供了可靠工具。

六、电子计时器

自 20 世纪 50 年代以来，各种新颖的电子计时器日益繁多，给反应时间研究提供了极为有利的条件。现代电子计时器已摆脱了传统的机械发条及摆锤结构，其基本部件一般采用晶振元件、集成电路（积体电路）及萤光数码管（或液晶）时间显示器。与上述微差计时器、时间描记器、机械钟表相比，电子计时器乃是更为精确的计时仪器。精密的电子毫秒计，其精确度可达到 1 毫秒。

现代心理实验室常用精密的电子仪器作为测时工具，如电子节拍器、电子毫秒器、电子定时器、电子时间描记器、数字式频率仪、示波器等，刺激的呈现和反应的启动须通过电子开关控制以获得更加可靠的实验数据。现代化的电子计时器，其构造大致如下：借一频率恒定的时间间隔脉冲发生器产生时间脉冲，通过一个控制器，输送至电子计数器；控制器在被测事件开始的瞬间接通时间脉冲信号，输入至计数器，计数器开始计数；而在被测事件结束的瞬间，切断脉冲发生器与电子计数器的通路，使计数器停止计数。因为脉冲频率固定，所以根据所记脉冲数即可折算成客观时间。一般我们可借十进位氖灯或计数管直接显示时间。这类仪器虽然价格高昂，但使用便利，精确度高，适合于实验室使用。

一般地说，电子计时器不仅计时精确而且功能多样。例如有一种多用途计时器，至少就有五种功能。这种仪器体积很小，只有 $15 \times 8 \times 10$ (cm^3)，仪器当中是一个定时钟，定时范围为 1 ~ 60 秒。左侧输入端有五对输入，具

体运作是：（1）给信号，定时钟启动；（2）给信号后，定时钟停；（3）给信号后，定时钟从头（回零）开始重新走；（4）给信号后，定时钟回零；（5）给信号后，暂停；信号解除，定时钟又走。这种仪器还有输出信号，输出信号可控制开关，并可控制别的仪器，作开关继电器用。多用途计时器，它还可以像体育馆的计时钟那样扣除暂停时间，因此可用于做学习记忆实验和其他计时实验。这种仪器在心理学上用途很广，台湾生产的多功能计时器，远销世界各地，是比较理想的计时器。

七、特殊摄影

当需要记录快速动作的反应时间或用很短的时间间隔连续记录反应时间时，可以把刺激的呈现、反应的动作及计时器的走时一齐拍入电影（录像或录影），必要时还可改变摄影画面的速度和距离，将反应时间加以“定格”或“放大”，使研究者更准确地获知反应时间。如在1981年莫斯科奥运会男子100米决赛中，对前三名运动员的反应时间和比赛成绩就是用高速摄影机和其他电子测时设备测定的。

从以上反应时间测定仪器的介绍中，我们也不难看到，反应时间研究的水平也同测量仪器息息相关。在科学研究发展的每一阶段，除了和学者所能正确运用的科学研究方法有关之外，尤其取决于当时生产和科学技术发展的水平。仪器设备与实验技巧的不断改进对反应时间的研究起了重要的作用，它帮助人们在这一领域内获得更加客观、精细、定量的知识。由于对时间测量的精确程度是反应时间测量水平的重要标志之一，一般来说，要使测值达到一定的精度，仪器的计量标准应当力求准确、精细并具有自动记录功能。因而，在设计反应时间实验时，也应选择好所需的仪器。

第三节 影响反应时间的因素

反应时间可因若干因素的影响而有变异，这是因为，反应时间作为反应变量，它要随着刺激变量和机体变量的不同而有变化。影响反应时间的因素有许多，我们主要可从刺激变量和机体变量二大因素进行分析。

一、反应时间受刺激变量影响

在刺激变量中，对反应时间影响比较大的因素有：刺激的不同类型、强度、复杂程度及刺激呈现的方式等。

(一) 因刺激的不同类型而异

不同类型的刺激通过特定的通道作用于各个感官，它们的反应时间是不同的。如果从赫希(Hirsch, 1862)较为成功的研究算起，学者对光、声、触等不同刺激的反应时间研究已积累了大量资料，可约略综合如表 3-6。在同一感觉道里，刺激的部位不同，反应时间也有差异。如刺激感觉灵

表 3-6 不同感觉道的反应时间(被试为有训练的成人)

感觉道	反应时间(毫秒)
触觉	117 ~ 182
听觉	120 ~ 182
视觉	150 ~ 225
冷觉	150 ~ 230
温觉	180 ~ 240
嗅觉	210 ~ 300
味觉	308 ~ 1082
痛觉	400 ~ 1000

(采自赫葆源、张厚粲、陈舒永等, 1983)

敏的部位、离大脑较近的地方，反应自然就比较快。另外，神经纤维粗细也影响着传导速度和感觉性质。由于传导痛觉的神经纤维细弱，传导速度也比较慢的原因，痛觉反应时间很慢。杨治良等人(1979)的研究发现，当兴奋 A 神经纤维兴奋时，被试者有“触”或“麻”的感觉，但无痛觉，只有当 A 纤维兴奋时，才有痛觉。柯林斯(Collins, 1979)报告指出：“刺激 A 纤维是不痛的，在脊髓切开手术之前，所有病人对刺激 C 纤维是不能忍受的”。所以，在不把问题过于简单化的前提下，每一种感觉有着各自的传入神经，因此反应时间上的差异也就不足为奇了。

(二) 因刺激的强度不同而异

刺激强度含义很广，既包括一般意义上的物理强度，也包括其他类似的因素，如视刺激的光强、空间面积、时间久暂及两个以上刺激的累积等。受刺激强度影响一般的情况是，当刺激强度很弱时，反应时间就长，而当刺激增至中等或较高强度时，反应时间便缩短了。绰绰里(Chocholle, 1945)用 1000 赫兹强度不等的纯音作反应时间测定，两名熟练的被试者的平均成绩列于表 3-7。从表中数据可知，反应时间随刺激强度的增加而减少，强度每增加一个对数单位，反应时间便有一定的缩减，但达到中等强度后，减小量减少，当刺激强度为 10 个对数单位(即在阈上 100 个分贝)时，被试者的反应时间不可能少于 110 毫秒，而达到一个极限。

表 3-7 对不同强度的 1000 赫兹纯音的反应时间

Igl	RT (毫秒)	Igl	RT (毫秒)
0	402	4	139
0.2	316	5	130
0.4	281	6	124
0.6	249	7	118
0.8	218	8	112
1	193	9	111
2	161	10	110
3	148		

(采自赫葆源、张厚燊、陈舒永等, 1983)

(三) 因刺激的复杂程度而异

刺激的复杂程度可以从两个方面来分析：(1) 刺激选择数目越多，愈复杂；(2) 刺激相似程度越低，则愈复杂。显然，刺激越复杂，反应时间必然越长。

从可供选择的刺激数目上看，选择的数目越多，反应时间必然越长，心理学家默克尔 (Merkel, 1885) 最早用实验证明了这一点。这种实验要求被试者对阿拉伯数字 1 到 5，按次序由右手 5 个手指作反应；而对罗马数字 1 到 5，则由左手 5 个手指作反应；每一系列试验所使用的刺激选择数目各不相同。研究者事先让被试了解选择的数目，实验结果如表 3-8 所示。

表 3-8 可供选择的刺激数目对反应时间的影响

刺激选择数目	1	2	3	4	5	6	7	8	9	10
反应时间(毫秒)	187	316	364	343	487	532	570	603	619	622

(采自曹日昌等, 1979)

谢帕德和梅茨勒 (Shepard & Metzler, 1973) 曾经做了另一个实验来说明反应时间受刺激复杂程度不同而异。实验是这样进行的：在银幕上显示图形，要求被试者回答银幕上的一对图形中物体的形状是“相同的”，还是“不同的”。实验共 8 名被试者，1600 对材料。在这个实验中，共有三种实验条件 (见图 3-7)：

A 条件：即左右两个图形中的物体是相同的，它们平放在一张纸上，只是角度有所不同，即从上到下旋转的角度不同。我们把这种旋转叫两维旋转，即在平面上的旋转。

B 条件：即左右两个图形中的物体是相同的，但它们之间的角度不相同，而且是从前到后旋转。我们把这种旋转叫做三维旋转，即因角度的不同而引起投射的大小和相互关系的不同。

C 条件：左右两个图形中的物体虽然有点相似，但实际上是不同的。

被试者通过几次练习，取反应时间为因变量。条件 A 和条件 B 的旋转角度范围为 $0^{\circ} \sim 180^{\circ}$ ，图 3-8 为这个实验所取得的结果。从图 3-8 上可见：(1) 两维旋转和三维旋转有相似的斜率；(2) 心理移转 (或心理旋转) (详见下一节) 呈线性规则，大约是 $60^{\circ}/\text{秒}$ ，也就是说，约 3 秒钟转半圈 (180°)。这里告诉我们，心理移转的弧度代表着刺激复杂性的程度，其移转的弧度越大，则越复杂，那么反应时间就越长。

反应依赖于刺激的复杂程度的原理也适用于社会心理方面的研究，达希

尔和希浦利 (Dashieil & Shipley, 1945) 等人在美学领域中也发现了相似的结果。他们给被试者呈现两种颜色, 被试喜欢哪边的颜色就用哪边的手去反应。对每个被试者来说, 颜色的美感价值越大, 选择反应时间也越快; 反之, 若两边颜色美感程度相差无几, 则选择时间就越长。若把选择反应时间作指标, 制作量表就能显示被试者对各种颜色的喜爱度。

我国心理学工作者 (詹美莎, 1987) 结合道路标志实践课题, 研究了反应时间受路标易读性而影响, 此实验被试者为 40 名职业汽车司机, 每人都填写过上述交通标志调查表。年龄最小为 20 岁, 最大为 52 岁, 平均 33 岁。教育程度小学至高中的均有。实验材料是将未达到理解度标准的 15 种标志 (见图 3-9) 分别制成带有各种街景、广告及人物等背景的幻灯片, 共三套, 每套 15 张。各标志在各幻灯片画面上出现的位置是随机的, 三套幻灯片中每种标志在画面上的位置又各不相同, 用幻灯机将刺激物呈现于 $30 \times 20\text{cm}^2$ 的屏幕上。用数字式电子停钟及自动控制装置记录被试者回答标志意义的声音反应时。被试者坐椅距离屏幕 2 米。话筒放在坐椅前的桌上。

在主试者向被试说明本实验的目的在于测定 15 种标志的辨认时间后, 接着将 15 种标志纸型依次呈现给被试并解释每种标志的意义及用途。要

表 3-9 再认 15 种道路标志的平均正确反应时和正确率

标志号	意 义	平面正确 反应时 (秒)	标准差 (秒)	正确率 (%)
1	立交路口直行和右转弯	1.81	0.67	87.5
2	单行路只准自行车上下行	1.82	0.70	100.0
3	只准两种车通行	1.61	0.50	95.0
4	单行路	2.20	1.22	92.5
5	快速路	1.48	0.44	95.0
6	快速路终点	2.15	0.13	90.0
7	让干路车先行	1.88	1.02	90.0
8	准试煞车	1.75	1.00	97.5
9	解除禁止超车	2.38	1.27	82.5
10	解除时速限制	1.88	0.94	92.5
11	山旁险路	2.07	0.95	97.5
12	落石	1.82	0.69	97.5
13	村镇	1.75	0.74	100.0
14	学校	1.52	0.48	100.0
15	施工	1.55	0.34	100.0

(采自詹美莎, 1987)

求被试者在听到主试者预备口令后注视屏幕, 当画面出现时迅速从中寻找交通标志并回头回答其意义, 尽量做到又快又准。

实验获得了表 3-9 的结果。从表 3-9 可以看出, 除 9 号“解除禁止超车”标志外, 其他标志的再认正确率均超过 85%, 9 号标志的再认正确率最低为 82.5%。结果表明各标志平均反应时间差异很显著。其中 9 号标志反应时最长, 为 2.38 秒, 它除了和 4 号 (单行路)、6 号 (快速路终点)、11 号 (山旁险路) 标志的反应时差异不显著外, 和其他 11 种标志的反应时差异都相当显著, 4 号、6 号、11 号标志反应时也相当长, 均超过了 2 秒。另外, 5 号 (快

速路)、14号(学校)、15号(施工)、3号(只准两种车通行)、8号(准试煞车)和13号(村镇)标志的反应时比较短,是在1.48~1.75秒之间,它们之间基本上无显著差异。这实验告诉我们,作为刺激材料复杂性的一种表现——易读性,对反应时间的影响是极为明显的。

(四) 因刺激呈现方式不同而异

在反应时间研究中,刺激呈现方式也归属于刺激变量,刺激的呈现方式不同,反应的难度也不同。这方面的实验研究古今都有不少,这里举一例。我国心理学家(张铁忠和沙建英,1987)对双眼视野不同位置的视觉运动反应时间进行了测定,获得了表3-10的实验结果。根据表3-10的八个不同

表3-10 六种距离、八种方向的视觉运动反应时间(单位:毫秒)

信号离视野中心的距离*	信号的方向								\bar{X}	S
	上	右上	右	右下	下	左下	左	左上		
5.73°	820.77	832.69	840.73	849.72	717.25	833.91	790.71	842.68	816.06	41.04
10.05°	815.82	824.42	791.54	814.07	699.80	790.90	775.34	814.67	790.82	37.73
20.10°	827.73	862.90	819.23	810.31	711.54	791.86	798.56	848.45	808.82	43.04
30.15°	832.47	892.17	836.20	831.89	707.28	839.88	804.91	861.82	825.83	50.12
40.21°	837.87	900.28	879.18	819.57	785.57	808.54	819.77	909.62	845.06	42.69
50.26°	905.19	941.32	874.96	832.24	792.68	871.97	909.94	893.23	877.69	43.90
\bar{X}	839.98	875.63	840.31	826.30	735.71	822.84	816.54	861.75		
S	32.91	44.32	33.32	14.60	41.87	31.66	48.08	34.76		

*信号离双眼视野中心的距离以角度表示

(采自张铁忠、沙建英,1987)

方向的视觉运动反应时间的平均值与标准差画成图3-10。从图上可以看出反应时间随着信号在视野中的方向不同而有差异,方向“下”反应最快,反应时最短,其次是方向“左”、“左下”与“右下”,再其次是方向“上”与“右”,“右上”与“左上”反应最长,反应时间的值最大。各方向反应时间的方差分析表明,差异非常显著。反应时曲线成草写“V”字形。如果以垂直与水平线上诸方向的反应时分别与按顺时针相邻方向的反应时进行比较,除“右”方向外,其他各方向的反应时均比斜方向快,整个曲线有成锯齿形趋势。

对图3-10和表3-10作双眼视野相反方向反应时的比较表明:方向“下”较“上”,“右下”较“左上”,“左下”较“右上”的反应时短,t检验表明差异均达显著水平;方向“左”与“右”反应的差异不显著。由此可见,偏离双眼视野中心的距离是影响反应时的重要因素。距离不同,反应时间也有明显的差异,反应时间随偏离视野中心距离的增大而增长。

二、反应时间受机体变量影响

影响反应时间的主要刺激变量如上所述,但外在原因需通过内在条件而起作用,和其他心理现象一样,反应时间也是以机体的内部状态为中介而对

外界刺激作出反映的。影响反应时间的机体变量为数众多，主要有：机体适应水平、准备状态、练习次数、动机、年龄因素和个别差异、酒精和药物作用等。

(一) 适应水平

适应 (adaptation) 在此处意指在刺激物的持续作用下，感受器发生的变化。感受器的适应水平对反应时间有着明显的影响。霍弗兰 (Hovland, 1936) 的实验中，当双眼对不同的照明适应后，测量了对 254 公尺烛光的反应时间。刺激是一直径为 20 毫米 (0.8 吋) 照明为 250 支烛光的白色圆块，观察距离为 30 厘米 (12 吋)，五名被试者的平均成绩 (每人每种条件各测试 100 次) 列于表 3 - 11。

这个材料表明，感受器适应水平的变化对反应时间的影响，这也相当于上面讲到的不同强度光刺激的影响。在反应时间实验中，不仅是视觉，而且听觉和其他感觉，都要考虑到被试者的感觉适应问题。

表 3-11 不同光适应水平的反应时间

适应水平 (呎·烛)	200	150	100	50	0
反应时间 (毫秒)	154	146	144	140	131

(采自 Hovland, 1936)

(二) 准备状态

准备状态 (readiness) 是指机体对于某种行为作出的准备情况。被试者在主试者发出“预备”口令到刺激呈现这段预备时间内的准备状态也是影响反应时间的因素之一。在这时距内，被试者处于积极准备状态，力求尽快对刺激作出反应。如果预备时间太短，被试者可能来不及准备；如果太长，被试者的准备状态又可能出现衰退而延误反应。许多研究说明，尽管存在个体差异，但最有利的预备时间大约是 1.5 秒。中村在研究蹲踞式起跑时，把从预备口令到鸣枪的时间间隔分别定为 1、1.5 和 2 秒三种，结果发现间隔时间为 1.5 秒时，运动员从听鸣枪到手离地的反应时间最短，见表 3 - 12。

表 3 - 12 准备状态和反应时间

间隔 (秒)	反应时平均数 (M) (毫秒)	标准差 (SD) (毫秒)
1	198.92	19.750
1.5	170.75	11.348
2	210.87	18.714

(采自藤田厚, 1982)

心理学家通过对肌电和脑电图形的分析来解释准备状态与反应时间的关系。当预备信号发出 200 ~ 400 毫秒时，执行反应的肌肉部位开始紧张起来，这种肌肉紧张不断增强直至预备时间之末而达到峰值；且当预备时间稳定在一个最适合于作出最快反应时，紧张度最高。此时，在相应脑部的波被抑制，在信号发出后 1000 ~ 1500 毫秒，波被抑制量为最大，大脑处于高水平的被激活状态。在这种条件下，被试者就可能作出快速反应。可见，反应时间有赖于机体准备状态的水平也有其生理基础。

(三) 练习次数

在心理实验中，练习 (exercise) 是一个控制变量，即一个潜在的自变量。练习与反应时间的关系相当密切。在一定范围内，练习次数越多 (上百次)，反应会越快，反应时间减少的趋势是逼近一个极限而稳定下来。兹以洛莫夫 (Lomov, 1965) 的实验加以说明。六名被试者对光刺激作顺序运动反应 (对光作一系列简单画图动作)，共练习 26 天，每天练习 10 次共 20 分钟。平均结果见图 3-11。由图可见，由于被试者起先不习惯于反应的方式，当他选定适合自己的操作方法后，练习因素开始起作用了。变异系数的值由大到小，表明反应时间逐渐减少到一定的水平并且稳定下来，接近极限水平。

关于练习对反应时间影响的机制，不少学者提出了一些解释。吉列齐娜 (Telekina, 1957)、格莱斯金 (Gerrelstein, 1958) 认为由于言语—思维的参与，不仅使感觉过程的调节机制得到改造，机体还能有意识地调节自己准确地掌握反应时间。

(四) 动机

动机 (motive) 是由于人的某种需要所引起的有意识或无意识的行为指向。反应时间实验中被试者易受某种额外动机的影响。约翰逊 (Johnson, 1922) 的实验发现主试者所设计的一些附加动机能使听觉反应加快，结果见图 3-12。图中激励系列是将每次反应结果告诉被试者，因而被试者得到一种略微的奖或罚的激励；惩罚系列是当被试者的反应慢于规定时间时，其手指就受到一次电击，而反应达到这允许时限便不受电击；常态系列是指没有附加动机的影响。这三种条件是循环使用的，以免产生练习效应。三名有训练的被试虽然都一直在尽快地作出反应，但在不同水平的奖惩动机影响下，反应还是有明显的变化。每条曲线都是 3600 次单一反应的分布曲线。

(五) 年龄因素和个体差异

人的心理发展变化存在着很大的个体差异，不同的人的发展速度，达到最高水平以及开始衰退的年龄都有所不同，年龄差异是指心理功能随年龄变化的总趋势。个体的年龄对反应时间的影响曾有不少研究，都得到相似的结果。一般认为，自发育阶段至 25 岁前 (青少年阶段)，反应时间随年龄增长而减少，起初减得快，以后较慢。学前期儿童由于难以控制的肌肉紧张、情绪激动与注意涣散的缘故，反应时间很不稳定，且不易获得较快的反应。7~8 岁学龄期儿童反应时间减少趋势比较明显，也开始稳定下来。

进入生理成熟期后，人的感觉-运动和短时记忆功能随着年龄的增长，有一个逐渐衰退的趋势。进入老年期后，这两个方面的功能可以有较显著的退化。戈茨达克 (Gottsdamker, 1968) 在对一组年龄为 18~93 岁的成人被试者的研究中注意到，随着年龄的增长，感觉-运动反应时间逐渐延长，但反应时间 (RT) 的增长量在相近的年龄组中是很缓慢的 (小于 2 毫秒/10 年)。此外，还有人发现在 25 岁以后到 60 岁的一段时间内，反应时间的增长极为缓慢，但 60 岁以后反应时间开始有了较大增加。图 3-13 是迈尔斯 (Miles, 1942) 的听觉简单反应的实验结果。

个体差异（或个别差异）（individual differences）指不同个体之间在品质和属性上存在的任何差别。无疑在反应时间实验中，常会遇到个别差异的问题。即使研究者能将上述来自内外的实验变量严加控制，但由于人们的心理特点不同，各人的反应时间是各不相同的，这就表现为在反应时间上的个别差异。菲萨尔（Fessard, 1926）曾对 1000 名男性成人被试者作听觉简单反应时间的测定，结果发现反应时间的均数分配大致上呈常态。为了测量各年龄阶段反应时间的离散程度或离中趋势（参见第二章），依据表 3-13 上

表 3-13 147 名成年被试者视听反应时成绩

项 目	视反应时		听反应时	
	平均数 (M)	标准差 (SD)	平均数 (M)	标准差 (SD)
青年组 (n = 57 人) 20 ~ 39 岁	195.64	29.10	140.27	27.26
中年组 (n = 42 人) 40 ~ 59 岁	220.22	29.47	160.32	31.35
老年组 (n = 48 人) 60 岁以上	246.29	48.90	82.78	45.10

（采自洪炜、梁宝勇，1986）

的数值算出了变异系数（variation coefficient）（注 3-2），结果发现，老年组的变异系数较显著地高于青年和中年组，说明老年组在感觉-运动反应速度方面存在着较大的个别差异。

注 3-2：

$$V = \frac{SD}{M} \times 100\%$$

说明：凡凭借算术均数来表示两个或两个以上标准差的相对差别的统计量数，就叫做变异系数。式中 V 表示变异系数；SD 表示标准差；M 表示平均数。

经过较长的练习，有些人比另一些人进步得快些，这就改变了它们在分配图上的相对地位，但由于各人的进步幅度不同，总的时间分配仍是对称的形态，个别差异仍然存在。

不仅不同被试者之间反应时间有差异，同一被试的反应时间也是有起伏的。主试者难以预测某个被试者在某个特定时刻的反应时间应是多少，这一变异也受制于他总的心理特点，而不是任何一、两个单独因素所造成的。

应当指出，年龄和个体差异还有交互作用。如有些老年人衰老得快些，有些人由于注意老年的保健与心身锻炼而延缓其病理性衰老。我国心理学工作者洪炜等对年龄的个体差异进行了研究，147 名有文化的健康成人参加了实验。正式测定前，进行 10~20 次练习，直到成绩比较稳定时为止。稍事休息后，再进行视觉和听觉的反应时间的正式测定。各记录 10 次结果。表 3-13 为三组被试者的平均数和标准差。图 3-14 为年龄与视听反应时间的关系。

（六）酒精及药物的作用

影响反应时间的因素可以列出许多，在实际领域一个引人注目的问题是酒精和药物也能影响反应时间。各类酒中都含有不等量的酒精。酒精在脑神经系统达到一定浓度时，中枢神经系统逐渐迟钝，对周围情况变化的反应速度大大下降。如果是酗酒，其反应时间将延长 2~3 倍，甚至更长，往往紧急情况已到眼前，还未发现或采取任何措施便已肇事。例如车速每小时 40 公里，未饮酒的驾驶员对道路复杂情况作出反应只需 0.6 秒，而饮酒后的驾驶员却要 1.8 秒。这样，汽车在反应时间内所行驶的距离就从 0.6 秒的 7 公尺增加到 1.8 秒的 21 公尺，所需的安全距离增加了 14 公尺，即从 27 公尺增加到 41 公尺。由于酒精对大脑皮层的抑制过程会产生破坏作用，使驾驶员很难正确估算车速、距离和自己的能力，以致驾驶动作不准确、不适当。据测试，当血液中的酒精达到 0.5~0.7 毫克时，驾驶员不仅选择反应时间增长，而且错误反应也增加 46%。

目前，刺激中枢神经的心理杀伤性药剂，大致分三类，即镇静剂、兴奋剂和致幻剂。镇静剂虽能消除情绪紧张、焦躁、恐惧感，但会使驾驶员肌肉活动能力下降，并出现睡意，以致反应迟钝。兴奋剂对中枢神经系统的作用与镇静剂作用相反，它可以消除疲劳，驱逐睡意，改善思维活动力，提高反应速度，但会使各类职业人员思想麻痹，过高估计自己的能力。致幻剂会使人有时产生幻觉。如伤风感冒服用过量解热镇痛剂，会使人的注意力、精力、反应能力下降。其他诸如安眠药、麻醉药品、止咳药、止痛片，咖啡因、非那明、利血平等类药品，要遵医嘱，不应随意服用。另外有些工程人员不应服用抗胺类药品，因为它会使人昏昏欲睡，反应能力下降。因病住院，或在治疗过程中，最好主动向医生介绍自己的职业，以及近期是否有出车等任务，以便医生合理用药。还有一种情况，即不能靠喝浓茶、浓咖啡来增加兴奋，因为这种兴奋是短暂的，随之而来的则是疲劳、困倦。

统观本节所述，影响反应时间的诸多因素可来自外界和机体内部各个方面。一般说来，对一些肌体变量，实验者是难以控制的，这也许是反应时间研究的一大难题。可是反应时间的探索曾给个别差异测量提供了机会，随着个别差异研究的深入，必将给反应时间研究的进一步发展创造条件，在相辅相成中，得到共同的发展。

第四节 用反应时间分析信息加工的方法

反应时间测量有两个用途：(1) 作为成就的指标，因为你对一件工作越精通，你就完成得越快；(2) 也可作为借以产生一种行为结果的内部过程复杂性的指标，因为内部过程越复杂，它所消耗的时间便越长。过去传统心理学更多地使用第一种用途，而今现代心理学则更多地使用第二种用途。

二十多年来，随着认知心理学 (cognitive psychology) 理论的发展，反应时间的测量受到特别的重视，认知心理学家对反应时间的关注是与信息加工 (或讯息处理) (information processing) 理论的基本思想分不开的。信息加工理论用抽象的方法分析人的心理过程，认为信息在脑内要经过一系列操作处理，这些操作处理有明确的先后次序，在不同的处理阶段信息处于不同的状态。心理处理是要时间的，处理信息量的多少，处理阶段的不同，都会直接反映在处理的反应时间特征上。因此，特定的心理操作能按它所需要的时间来测量，不同的心理操作按它们在时间分布上的不同而区分开来。

可见，反应时实验受到认知心理学家的重视并不是偶然，而是有着深刻原因的。认知心理学家强调研究内部心理过程，并将它看作一系列连续阶段的信息加工过程。要把握这个过程，则需要将输入和输出联系起来进行推论。反应时实验恰好可以为此提供极为重要的客观材料。应用反应时间分析人类信息加工过程的两种基本方法是减数法和加因素法。

一、减数法

(一) 什么叫减数法

一些有影响力的实验心理学教科书，把心理学上反应时间的研究划分为两个时期。第一时期称减数法时期，它是由唐德斯 (Franciscus Cornelis Donders, 1818 ~ 1889) 奠定基础的。又称唐德斯反应时 ABC (Donders ABC of reaction time) 或唐德斯三成分说 (Donders three components)。第二个时期称加因素法时期，它是由斯顿伯格 (Sternberg, 1969) 奠定基础的。减数法 (subtractive method) 乃是一种用减数方法将反应时间分解成各个成分，然后来分析信息加工过程的方法。它是由唐德斯首先提出的，故又称唐德斯减数法 (Donders subtractive method)。

唐德斯最初应用这种实验来测量包含在复杂反应中的辨别、选择等心理过程所需要的时间。在这种实验里，通常需要安排两种不同的反应时作业，其中一种作业包含另一种作业所没有的某个特定的心理过程，此特定心理过程即为所要测量的心理过程，而二者在其余方面均相同，这两种反应时的差即为此过程所需的时间。举例来说，当被试者觉察一个灯光刺激后立即用右手按键作出反应，这样就测到一个简单的视觉反应时 (RT_1)。如果实验安排红绿两个色光刺激，并要求被试者看到红光后立即用右手按键来反应，而看到绿光时则不作出反应，这样测到的复杂反应时 (RT_2) 要长于前面的简单反应时。这两种反应时作业的区别仅仅在于后者需要将红绿两个色光刺激区分开来，所以这两种反应之差就是辨别过程所需的时间，即 $RT_2 - RT_1 =$ 辨别过程时间。减数法反应时实验的名称也由此而来。同理，如果实验仍安排红绿两个色光刺激，但要求被试者在看到光刺激时，不但要作出辨别，而且还要对反应作出选择。将这样测得的反应时 (RT_3) 减去含有辨别过程的反应时 (RT_2)，就得到选择过程所需的时间，即 $RT_3 - RT_2 =$ 选择过程时间。

唐德斯的减数法把反应分为三类，即 A、B、C 三种反应：

第一类反应称 A 反应 (A-reaction)，又称简单反应 (simple reaction) (见图 3-15)。A 反应一般只有一个刺激和一个反应，例如被试者对一个灯亮，作一个按键的反应。唐德斯称这种反应为基线时间 (baseline time)，一方面它是最简单的反应，另一方面它又是更复杂反应的成分或基本因素。

第二类反应称为 B 反应 (B-reaction)，又称选择反应 (choice reaction) (见图 3-16)。它是复杂反应中的一种，在这类反应中，有二个或二个以上的刺激和相应于刺激的反应数。也就是说，每一个刺激都有它相应的反应。如在一个实验中，对红光信号灯按 A 键，对绿光信号灯按 B 键，对蓝光信号灯按 C 键，对白光信号灯按 D 键，共四个刺激和四个对应的反应，在这样的选择反应中，不仅要区别信号，而且还有选择反应。因而在这样的

反应中包括了刺激辨认的心理操作和反应的选择。

第三类反应称为 C 反应 (C - reaction)，又称辨别反应 (identification reaction)。它是另一种形式的复杂反应(见图 3 - 17)。C 反应和 B 反应有相同点：具有二个或二个以上的刺激。C 反应和 B 反应也有区别之处：C 反应中只有一个刺激是要求有反应的，而其余刺激是不要求作出反应的，即不要求做任何事。那么是不是 C 反应不需要心理操作呢？不是的。B 反应既有刺激的辨别，还有反应的选择。而 C 反应仅有刺激的辨别，但没有反应的选择。

通过以上叙述，我们可以分析 A、B、C 三种反应的辨别和选择心理操作所要求的时间。这种分析，称之为唐德斯的减数法。减数法顾名思义，其 A、B、C 三种反应是按减法规律进行的。

根据减数法法则来分析 A、B、C 三种反应。C 反应就等于辨别时间加同类的基线时间（神经传导时间等）。这里的“同类”意指刺激信号上相同。这样，从 C 中减去 A，就告诉我们辨别时间有多长。与此相似，从 B 中减去 C，就告诉我们选择时间有多长。因为 B 包括了辨别、选择和基线等三个时间。而 C 只包括辨别和基线两个时间。它们之间的关系可从图 3 - 18 上看到。

针对三种反应的区别，我们就可对 A、B、C 三种反应进行时间上的排队：B 类反应时间最长，C 类次之，A 类最短。和减数法原理相同，唐德斯明确提出了反应时间的成分说。即：反应时包括了三个成分——基线时间 (baseline time)、辨别时间 (identification time) 和选择时间 (selection time)。

以上就是唐德斯减数法的原理，尽管很长时间以来，也有不少人反对唐德斯的三成分说，其主要理由是：B 反应决不是在 C 反应上简单地加上些什么；C 反应也决不是在 A 反应上简单地加上些什么。但是，唐得斯的减数法在 19 世纪 60 年代到 20 世纪 60 年代这样长的时间内，在实验心理学中占有很高的地位，他的三成分说和减数法法则曾广泛应用于许多领域。今天，认知心理学正是这样应用减法反应时实验提供的数据来推论其现象背后的信息加工过程。

(二) 减数法的典型实验

在当今涉及快速的信息加工过程中，例如识别、注意、表象和短时记忆等，常常应用减数反应时实验。这种实验既可用于研究信息加工的某个特定阶段或其操作，也可用来研究一系列连续的信息加工阶段，证明某一心理过程的存在。下面我们介绍减数法的三个典型实验。

1. 证明心理移转存在的实验 70 年代初，库柏和谢波德 (Cooper & Shepard, 1973) 用减数法反应时实验证明了心理移转 (注 3 - 3) 的存在。怎样证明它的存在呢？

注 3-3：心理移转(或心理旋转)(mental rotation)指单凭心理运作(不靠实际操作)，

将所知觉之对象予以移转，从而获得正确知觉经验的心理历程。

库柏等取非对称性的字母或数字 (J、G、2、5、7 等) 为实验材料，根据“正”、“反”以及不同的倾斜度，构成 12 种情况 (见图 3-19)。从图上可见，由于 R 字母在垂直或水平方向均是不对称的，所以正、反也是不相

同的。在这个实验里，被试的任务是不管倾斜度如何，只要判明正或反，并作出反应。

实验根据：有没有提示？单项提示还是双项提示？先后提示还是同时提示？由这三个变量而分成五种实验条件（见图 3 - 20）。具体地说，这五种实验条件是：（1）完全没有提示，即测验前 2 秒钟后呈现空白信号，持续 2 秒钟；（2）提示正或反，即测验前 2 秒钟呈现正或反的提示信号，持续 2 秒钟；（3）提示倾斜度，即测验前 2 秒钟用箭头提示倾斜度数，也是持续 2 秒钟；（4）分别提示正、反和倾斜度，即用两个信号分别提示正、反和倾斜度，而倾斜度的提示时间是可变化的，共有距测验 100 毫秒、400 毫秒、700 毫秒和 1000 毫秒等四种情况；（5）同时提示正、反和倾斜度，即在测验前用二个信号同时提示正、反和倾斜度，都持续 2 秒钟。实验获得了可喜结果（见图 3 - 21）。我们从图上可见，反映五种实验条件的曲线是有规律的。图中数值为“正”和“反”二种反应的平均值。有趣的是“正”反应和“反”反应二者相差总是 0.1 秒，也就是说正、反移转相差 0.1 秒。

那么，如何分析实验所获得的实验结果呢？也就是说图上的几条“V”形曲线说明了什么呢？在减数法反应时实验里，通常需要安排两种不同的反应时作业，其中一种作业包含另一种作业所没有的某个特定的心理过程，即所要测量的心理过程，而二者在其余方面均相同，这两种反应时的差即为这个过程所需时间。这样，此实验中，倾斜度 0° 和倾斜度 60° 二种情况，由于其他条件均一样，故：

$RT_{60^\circ} - RT_{0^\circ} = \text{从 } 0^\circ \text{ 至 } 60^\circ \text{ 心理移转所需时间}$

RT_{60° ：表示倾斜度 60° 时的反应时间

RT_{0° ：表示倾斜度 0° 时的反应时间

其余类推。同理：

$RT_{120^\circ} - RT_{60^\circ} = \text{从 } 60^\circ \text{ 至 } 120^\circ \text{ 心理移转所需时间}$

有趣的是，从图 3 - 22 上可见，这条曲线呈倒“V”形，以中间 180° 相对称。由此就可推论，在 0° ~ 180° 这段范围内，心理移转的方向是顺时针的；在 180° ~ 360° 这段范围内，心理移转的方向是逆时针的，即：

$\text{心理移转所需时间} = RT_{180^\circ} - RT_{240^\circ}$

当然，我们从库柏和谢帕德的实验中还看到，每种反应中都包括有一个最基本的选择反应时间，这就是在 0° 或 360° 时的反应时间。这个反应是在这个实验中最基本的成分。

这里，我们可以从这个实验中，通过反应时减数法则，说明心理移转的存在。也就是说，在我们每人的头脑里，有事物的映象。人们都相信自己有心理的眼睛（mind's eye）。

2. 句子-图画匹配实验 在 80 年代，心理学家克拉克和蔡斯（Clark & Chase, 1972）作了句子-图画匹配实验。我国心理学家王甦撰文，将此实验推崇为减数法的范例。在这种实验里，给被试看一个句子，紧接着再看一幅图画，如“星形在十字之上，✳”，要求被试尽快地判定，此句子是否真实地说明了图画，作出是或否的反应，记录反应时。实验用的介词有“之上”和“之下”，主语有“星形”和“十字”。句子的陈述有肯定的（在）和否定的（不在），共有 8 个不同的句子。克拉克和蔡斯设想，当句子出现在图画之前时，这种句子和图画匹配任务的完成要经过几个加工阶段，并提出了度量某些加工持续时间的参数。照他们的看法，第一个阶段是将句子转换为

其深层结构，而且对“之下”的加工要长于对“之上”的加工（参数 a），对否定句的加工要长于对肯定句的加工（参数 b）；第二个阶段是将图画转换为命题，并常有前面句子中所应用的介词（“之上”或“之下”）；第三个阶段是将句子和图画两者的命题表征进行比较，如果两个表征的第一个名词相同，则比较所需的时间比不同时为少（参数 c），如果两个命题都不含有否定，则比较所需的时间比任一命题含有否定时为少（参数 d）；最后的阶段为作出反应，其所需的时间被认为是恒定的（参数 t）。这样，对句子和图画匹配的任务来说，减法反应时实验就在于将依赖所呈现的句子和图画的诸反应时加以比较。例如，如果“星形在十字之下”这个句子真实地说明了图画，那它就有参数 a 和 t_0 ；如果“星形在十字之上”，这个句子真实地说明了图画，那它只有参数 t_0 ；这两反应时之差就是反应参数 a 的持续时间。但参数 b 和 d 只出现在否定句中，所以是无法分别测量的。克拉克和蔡斯用这种办法计算一个实验的结果，得到如下的结果：参数 a 为 93 毫秒，参数 b 和 d 为 685 毫秒，参数 c 为 187 毫秒，参数 t_0 为 1763 毫秒。

克拉克和蔡斯的上述实验得到肯定的评价。

3. 证明某些短时记忆信息可以有视觉编码和听觉编码两个连续阶段的实验 20 世纪 60 年代以来，根据记忆实验中对错误回忆的分析，通常认为人的短时记忆信息（如字母）是以听觉形式来编码的。但是 70 年代初波斯纳等人（Posner et al., 1990）的实验却表明，这种信息可以有视觉编码。他们所依据的就是减数法反应时实验的结果。实验是这样进行的：给被试并排呈现两个字母，这两个字母可以同时给被试者看，或者中间插进短暂的时间间隔，要被试者指出这一对字母是否相同并按键作出反应，记下反应时。所用的字母对有两种：一种是两个字母的读音和书写方法都一样，即为同一字母（AA）；另一种是两个字母的读音相同而写法不同（Aa）。在这两种情况下，正确的反应均为“相同”。在两个字母相继呈现时，其间隔为 0.5 秒和 1 秒，或 1 秒和 2 秒等。他们得到的结果见图 3-22。从图 3-22 可以看出，在两个字母同时呈现时，AA 对的反应时小于 Aa 对；随着两个字母的时间间隔增加，AA 对的反应时急剧增加，但 Aa 对的反应时则没有发生大的变化。并且 AA 对和 Aa 对的反应时的差也逐渐缩小，当时间间隔达到 2 秒，这个差别就很小，在图上看到二条曲线趋于靠拢。

针对这有趣的结果，波斯纳等认为，既然 AA 对与 Aa 区的区别只在于前者的两个字母有一样的写法，而后者没有，那么，当两个字母同时呈现给被试，Aa 对的反应时大于 AA 对是由于 AA 对字母可以直接按写法来比较，但 Aa 对却不能按写法而必须按读音来比较。这意味着 AA 对匹配是在视觉编码的基础上进行的，而 Aa 对匹配只能在听觉编码的基础上进行。必须从视觉编码过渡到听觉编码，因此需时也较多。Aa 对与 AA 对的反应时之差反映内部编码过程的差别。可以说，先出现视觉编码，它保持一个短暂的瞬间，然后出现听觉编码。这样，即使是 AA 对，随着两个字母之间插入时间间隔及其增大，视觉编码的效应逐渐消失，听觉编码的作用增大，其反应时间也逐渐增大，从而缩小与 Aa 对的反应时的差别。AA 对的同时呈现和继时呈现的反应时之差也反映信息编码的这种特点。

从以上的实验结果可以看到，波斯纳等人应用减数法反应时实验清楚地确定，某些短时记忆信息可以有视觉编码和听觉编码两个连续的阶段，这也

是认知心理学上的重大发现。

以上介绍的“心理移转”实验、“句子-图画匹配”实验、和“视觉和听觉编码两个连续阶段”实验，乃是减数法及反应时间的典型实验范例。

减数法的反应时间实验的逻辑是安排两种反应作业，其中一个作业包含另一个作业所没有的一个处理（加工）阶段，并在其他方面均相同，从这两个反应时间之差来判定此加工阶段。这种实验在原则上是合理的，在实践上是可行的。认知心理学也正是应用减数法反应时间实验提供的数据来推论其背后的信息加工过程的。

减数法也有其弱点：使用这种方法要求实验者对实验任务引起的刺激与反应之间的一系列心理过程有精确的认识，并且要求两个相减的任务中共有的心理过程要严格匹配，这一般是很困难的。这些弱点大大限制了减数法的广泛使用。二、加因素法

（一）什么叫加因素法

在 20 世纪，斯顿伯格（Sternberg, 1969）发展了唐德斯的减数法反应时间，提出了加法法则，称之为加因素法（additive factors method）。这种实验并不是对减数法反应时间的否定，而是减数法的发展和延伸。加因素法反应时间实验认为完成一个作业所需的时间是一系列信息加工阶段分别需要的时间的总和，如果发现可以影响完成作业所需时间的一些因素，那么单独地或成对地应用这些因素进行实验，就可以观察到完成作业时间的变化。加因素法反应时间实验的逻辑是：如果两个因素的效应是互相制约的，即一个因素的效应可以改变另一因素的效应，那么这两个因素只作用于同一个信息加工阶段；如果两个因素的效应是分别独立的，即可以相加，那么这两个因素各自作用于不同的加工阶段。这样，通过单变量和多变量的实验，从完成作业的时间变化来确定这一信息加工过程的各个阶段。因此，重要的不是区分出每个阶段的加工时间，而是辨别认知加工的顺序，并证实不同加工阶段的存在。加因素法假定，当两个实验因素影响两个不同的阶段时，它们将对总反应时间产生独立的效应，即不管一个因素的水平变化如何，另一个因素对反应时间的影响是恒定的（模式如图 3-23，图中 g_1 、 g_2 为 g 因素的两个水平， f_1 、 f_2 为 f 因素的两个水平）。这样称两个因素的影响效应是相加的。加因素法的基本手段是探索有相加效应的因素，以区分不同的加工阶段。

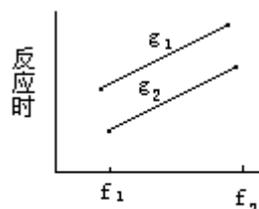


图3-23 独立因素相加效应的图解

（二）加因素法的典型实验

使用加因素法分析心理过程的一个典型实验是斯顿伯格的“短时记忆信息提取任务”。斯顿伯格在《反应时间实验揭示的心理过程》一文（1969）中，系统地阐述了加因素法反应时间实验和用此方法所作的短时记忆信息提取实验。在他的实验里，先给被试看 1~6 个数字（识记项目），然后看一个数字（测试项目）并同时开始计时，要求被试判定此测试数字是否是刚才识记过的，按键作出是或否的反应，计时也随即停止。这样就可以确定被试者

能否提取以及所需要的时间（反应时间）。通过一系列的实验，斯顿伯格从反应时的变化上确定了对提取过程有独立作用的四个因素，即测试项目的质量（优质的或低劣的）、识记项目的数量、反应类型（肯定的或否定的）和每个反应类型的相对频率。因此，他认为短时记忆信息提取过程包含相应的四个独立的加工阶段，即刺激编码阶段、顺序比较阶段、二择一的决策阶段和反应组织阶段（先前曾将最后的两个阶段合并为一个阶段）。照斯顿伯格的看法，测试数目的质量对刺激编码阶段起作用，识记项目的数量对顺序比较阶段起作用，反应类型对决策阶段起作用，反应类型的相对频率对反应组织阶段起作用。现在可以将以上所说的用图表示出来（图 3-24）。图中箭头表明信息流动的方向，虚线连接起作用的因素。从图中可以看到，从短时记忆中提取信息的过程包括测试项目的编码、顺序比较、决策和反应组织等四个依次进行的加工阶段，下面我们不妨分四步，每步作成一张图，进行分析：

第一步，改变检验刺激（probe）的质量，发现对一个残缺、模糊的刺激进行编码比对一个完整、清晰的刺激花的时间较长，而且对不同大小的记忆表影响相似，即记忆表的大小仅改变 Y 截距，而不改变直线的斜率，表明系列比较阶段之前存在一个独立的编码阶段（见图 3-25）。

第二步，改变记忆表中项目的数量，得出记忆表大小与反应时间之间的线性关系，证实了系列比较阶段的存在。他假定余下的三个阶段的反应时间对不同的记忆表是不变的。将斜线向左延伸至 Y 轴，Y 轴上的点提供了系列比较为 0 时的反应时间，实际上就是编码刺激做两分决定和组织反应共花的时间（见图 3-26）。

第三步，分别计算 Y 与 N 反应，发现对不同大小的记忆表，都是 N 反应时长于 Y 反应时，表明了两种决定阶段的存在，即在系列比较之后，有一个被试选择反应种类的阶段，而且产生 N 反应比产生 Y 反应难（见图 3-27）。

第四步，改变某一类反应（Y 或 N）的出现频率，发现对两类反应产生同样的影响，提高任一类反应的出现频率，都会使这类反应的组织更为容易，从而使反应时下降。表明反应选择之后存在一个独立的反应组织阶段（见图 3-28）。

斯顿伯格在完成了上述实验后提出，加因素反应时有二个特点：（1）实验者可以通过操作变更阶段的持续时间，完成这项工作的自变量就称之为因素，当然，因素可以不只是一个；（2）在这些因素中又可分为二类：一类为影响反应时间的附加因素，亦为非交互作用的因素，这类因素称之为影响反应的附加因素，另一类因素为影响同一阶段的因素，这类因素为交互作用的因素。因此，斯顿伯格认为，一旦找到交互作用和附加因素的模型，心理学家也就揭示了加工阶段是怎样相关的。

斯顿伯格的模型和方法引起了很多心理学家的兴趣，也激起了大量的研究。虽然，斯顿伯格的实验被许多心理学家看做加法反应时典型的实验，但它也引起一些批评和疑问。加因素法的弱点是，它的基本前提是人的信息加工是系列加工，这一点受到很多心理学家的质疑。因为加因素法反应时实验是以信息的系列加工而不是平行加工为前提的，因而有人认为其应用会有很多限制。其次减数法反应时实验也同样存在这个问题。这涉及认知心理学的一个基本原则，应当予以重视。然而更为直接的问题是关于加因素法反应时实验的逻辑，即能否应用可相加和相互制约的效应来确认信息加工的阶段。佩奇拉（Pachella, 1974）曾经指出，两个因素也许能以相加的方式对同一

个加工阶段起作用，也许能对不同的加工阶段起作用并且相互发生影响。虽然这两种可能性目前还不能排除，但也没有根据现在就否定加因素法反应时实验。还有人指出，加因素法反应时实验本身并不能指明一些加工阶段的顺序，在这个方面，它极大地依赖于一定的理论模型。这个意见也值得重视。但斯顿伯格首创的加因素法反应时实验，终究将反应时实验向前推进了一大步，并在很大程度上，对认知心理学的发展起着积极的影响。

（三）开窗实验

开窗（open window）实验，既和减数法反应时实验有相同之处，又和加因素法反应时实验有相同之处。由于“开窗”实验在反应时研究历史上是发展较晚的一种方法，因此较多的教科书上都把这种实验作为加因素法反应时的一种变式，加以分类，而从属于加因素法反应时实验。前面谈到的减数法和加因素法反应时实验难以直接得到某个特定加工阶段所需的时间，并且还要通过严密的推理才能被确认。如果能够比较直接地测量每个加工阶段的时间，而且也能比较明显地看出这些加工阶段，那就好像打开窗户一览无遗了。这种实验技术称为“开窗”实验，它是反应时实验的一种新形式，现在以汉米尔顿（Hamilton, 1977）和霍克基（Hockey, 1981）的字母转换实验为例加以说明。在这种实验里，给被试呈现1~4个英文字母并在字母后面标上一个数字，如“F+3”、“KENC+4”等，当呈现“F+3”时，要求被试说出英文字母表中F后面第三个位置的字母即“I”。换句话说，“F+3”即将F转换为I，而“KENC+4”的正确回答是“OIRG”，但这四个转换结果要一起说出来，凡刺激字母在一个以上时，都应如此，只作出一次反应。现以“KENC+4”为例说明实验的具体过程。四个字母一个一个地继时呈现，由被试者按键自行控制，被试者第一次按键就可以看见第一个字母K，同时开始计时，接着被试者按照要求作出声转换，说出LMNO，然后再按键看第二个字母E，再作转换。如此循环直至四个字母全部呈现并作出回答，计时也随之停止。出声转换的开始和结束均在时间记录中标出来。根据这种实验的反应时数据，可以明显地看出完成字母转换作业的三个加工阶段：（1）从被试者按键看到一个字母到开始出声转换的时间为编码阶段，被试对所看到的字母进行编码并在记忆中找到此字母在字母表中的位置；（2）被试按规定进行转换所用的时间即为转换阶段；（3）从出声转换结束到按键看下一个字母的时间为贮存阶段，被试将转换的结果贮存到记忆中，有时还须将前面的转换结束加以复述和归并。这三个阶段可用图3-29来表示。在四个刺激字母实验里，可以获得12个数据，通过对数据的归类处理就可得到总的实验结果。从中可以看到字母转换的整个过程和经过的所有加工阶段。

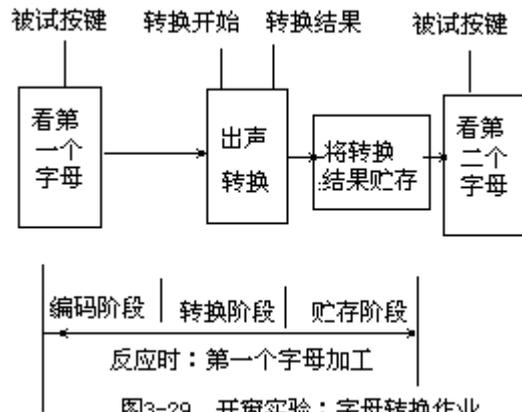


图3-29 开窗实验：字母转换作业

不难看到，这种“开窗”实验的优点是引人注目的，但也存在着一些问题。例如，可能在后一个加工阶段出现对前一个阶段的复查，贮存阶段有时还包含对前面字母的转换结果的提取和整合，并且它难以与反应组织分开来。然而经过细心安排，有些问题还是可以避免的。这也促使认知心理学对实验设计的严格要求。

我们从以上二节的介绍中看到，反应时法的发展对心理学研究带来了新的成果，其中，特别是减数法反应时实验，是目前认知心理学应用得最广泛的研究方法。这些实验被应用到具体课题研究时还会有具体的形式和特点。事实证明，它们是可以有效地发挥作用的，当然也存在一些问题和困难。

应当看到，这些情况与当前整个认知心理学的理论原则和现状是分不开的。随着认知心理学的发展，这些反应时实验也将发生变化。而对于这些实验中存在的具体困难和弱点，则要通过精心的实验设计加以避免或减少其不利的影响。一个带有普遍意义的重要条件是在进行实验之前作出相应的理论模型，或作出一些假设，使实验有明确的规划，以保证实验结果的可靠性，并得到有意义的结论。这也是认知心理学有效地应用反应时法的前提。

本章实验

一、简单反应时间实验

(一) 目的：学习对视觉与听觉简单反应时间的测定方法；比较两种简单反应时间的差别。

(二) 材料：简单反应时间测定装置。

(三) 程序：

1. 预备实验

(1) 接通仪器电源，主试者拨动信号发生开关，在光或声刺激呈现的同时，计时器应立即进行计时。

(2) 练习操作，刺激呈现器放在被试 1 公尺处。被试以右手食指轻触电键。主试者在发出“预备”口令后约 2 秒呈现刺激。被试者当感觉到刺激出现时，立即按压电键。计时器停止计时，主试者记下成绩。练习实验可作 2~3 次。

(3) 为防止无关刺激的干扰，主试者与被试者可分隔在两个操作室中进行实验。

2. 正式实验

(1) 刺激呈现按视——听——听——视方式安排，每单元各作 20 次，

总次数为 80 次。

次数	光	声	声	光
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
平均数				

(2) 同预备实验。

(3) 为了检查被试者有无超前反应现象，在每 20 次中插入一次“检查实验”。如被试对“空白刺激”作出反应，主试者根据反馈信号灯提供的信息须宣布此单元实验结果无效，重做 20 次。

(4) 做完 20 次，休息 1 分钟。一被试者测完 80 次后，换一被试者进行实验。

(四) 结果：

(1) 计算个人视觉与听觉反应时的平均数及标准差。

(2) 检验全体被试两种简单反应时是否有明显差别。

(五) 讨论：

(1) 根据实验结果说明视与听感觉道简单反应时的差别及其原因。

(2) 根据实验结果说明简单反应时是否受练习的影响。

二、选择反应时间实验

(一) 目的：学习测定视觉选择反应时间的方法，了解选择反应时间与简单反应时间的区别。

(二) 材料：选择反应时测定装置。

(三) 程序：

1. 预备实验

(1) 接通仪器电源。主试者按预先列出的程序操作信号呈现开关，发出“红”、“黄”、“绿”、“白”四种不同光刺激。

(2) 被试者以右手食指作按键状，当感觉到某种色光时，立即按压相应的反应键（即被试对四种不同的刺激相应作出四种不同的反应）。计时器记下时间，练习实验可作4~5次。

2. 正式实验

(1) 四种色光刺激各呈现20次，随机排列。

(2) 主试者呈现刺激与被试者反应方式同预备实验。如果反应错了，计时器不计时间，主试者根据反馈信号灯提供的信息，安排被试者重作一次。

(3) 每做完20次休息1分钟。一被试者测完80次，换另一被试者进行实验。

(四) 结果：

(1) 计算个人不同色光的选择反应时的平均数、标准差。

(2) 比较全体被试者对白光的简单与选择反应时的平均数差异。

(五) 讨论：

(1) 本实验结果是否与前人实验的数据一致？原因何在？

(2) 举例说明反应时实验的实际应用意义。

表 3-15 对四种色光的选择反应时实验记录

(反应时单位：毫秒)

次数	红光	黄光	绿光	白光
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
平均数				

被试姓名： 主试姓名： 实验时间：

本章摘要

1. 关于人的反应时间的差异。1795 年英国格林威治天文台台长马斯基林的助手，因观测星球通过子午线的时间总比台长观测的晚半秒钟而被辞退。后来德国的天文学家贝塞尔与另一位天文学家阿格兰德共同观察七颗星，发现时间也不相同，于是把两个观察者观察上的时间差异称为人差方程式。

2. 反应时间，又称反应时，它是一个专门的术语，不是执行反应的时间，而是指刺激施于有机体之后到明显反应开始所需要的时间。

3. 如果以刺激与反应的不同型式或复杂程度为标准，一般可把反应时间划分为简单反应时间和复杂反应时间（又称选择反应时间）两类。

4. 速度和准确性是反应时实验中的两个基本因变量。反应时间实验中的一个突出问题就是权衡反应速度和准确性之间的相互关系。

5. 反应时间是心理学上最重要的反应变量和指标。现实生活中有许多领域需要研究人的反应时间，尤其是在工业生产、车辆驾驶、体育运动、设备设计及科学管理等实际部门，反应时间的应用更有其重要意义。

6. 反应时间研究的水平同测量仪器息息相关。仪器设备的不断改进，对反应时间的研究起着重要的作用，它可帮助人们在这一领域内获得更加客观、精确、定量的知识。

7. 反应时间可因若干因素的影响而有变异。这是因为，反应时间作为反应变量，它要随着刺激变量和机体变量的不同而有变化。

8. 在刺激变量中，对反应时间影响较大的因素主要有：刺激的不同类型、强度、复杂程序及刺激呈现的方式等。

9. 影响反应时间的机体变量为数众多，主要有：机体的适应水平、准备状态、练习次数、动机、年龄因素和个体差异、酒精和药物作用等。

10. 认知心理学家对反应时间的关心，是与信息加工理论的基本思想分不开的。因此，特定的心理操作能按它所需要的时间来测量。

11. 实验心理学上把反应时间的研究划分为二个时期或两个阶段。第一时期称减数法时期，它是由唐德斯奠定基础的。第二时期称加因素法时期，它是由斯顿伯格奠定基础的。

12. 减数法乃是一种用减数方法将反应时间分解成各个成分，然后来分析信息加工过程的方法。

13. 加因素法认为完成一个作业所需的时间是一系列信息加工阶段分别需要的时间的总和，如果发现可以影响完成作业所需时间的一些因素，那么简单地或成对地应用这些因素进行实验，就可以观察到完成作业时间的变化。

14. 加因素法不是对减数法的否定，而是减数法反应时实验的发展和延伸。

建议参考资料

1. 王甦（1987）：认知过程的反应时实验。《心理学杂志》，第2卷，1期，37~42页。

2. 杨治良（1979）：几种简易心理学电子仪器。《心理学报》，第11卷，2期，249~251页。

3. 杨治良（1988）：心理物理学。兰州市：甘肃人民出版社。

4. 杨治良（主编）（1990）：实验心理学。上海市：华东师范大学出版社。

5. 张春兴（1989）：张氏心理学辞典。台北市：东华书局（繁体字版）。上海市：上海辞书出版社（1992）（简体字版）。

6. 张铁忠、沙建英（1987）：双眼视野不同位置的视觉运动反应时间的测定。《心理学杂志》，第2卷，1期，7~11页。

7. 陈舒永、杨博民等（1986）：不同肢体的反应时间和运动时间。《心理学报》，第18卷，1期，1~7页。

8. 洪炜、梁宝勇（1986）：感觉运动反应时间与短时记忆的年龄变化。《心理科学通讯》，第1期，41~43页。

9. 曹传咏等（1963）：在速示条件下儿童辨认汉字字形的试探性研究。《心理学报》，第8卷，3期，203~211页。

10. 彭瑞祥、喻柏林（1983）：普通心理学与实验心理学论文集：不同结构的汉字再认的研究。兰州市：甘肃人民出版社。

11. 赫葆源、张厚粲、陈舒永（1983）：实验心理学。北京市：北京大学出版社。

12. 藤田厚等（丁雪琴译，1982）：心理学研究法。济南市：济南心理学会。

13. Atkinson, R.C. et al. (1988). Stevens' handbook of experimental psychology. (2nd ed.). New York: A Wiley-Interscience Publication.
14. Best, J.B. (1986). Cognitive psychology. New York: West Publishing Co.
15. Kantowitz, B.H., & Roediger, H.L. (1984). Experimental psychology. (2nd ed.). New York: West Publishing Co.
16. Scripture, E.W. (1901). The new psychology. New York: Charles Scribner's Sons.
17. Yang, Z.L. et al. (1984). Psychological aspects of components of pain. *The Journal of Psychology*, 118 (2), 135 ~ 146.

第四章 传统心理物理法

本章内容细目

第一节 阈限的性质

一、心理物理法的诞生 167

二、阈限及其操作定义 168

第二节 测定阈限的三种基本方法

一、极限法 171

(一) 用极限法测定绝对阈限

(二) 用极限法测定差别阈限

二、平均差误法 177

(一) 用平均差误法测定绝对阈限

(二) 用平均差误法测定差别阈限

(三) 误差及其控制

三、恒定刺激法 180

(一) 用恒定刺激法测定绝对阈限

(二) 用恒定刺激法测定差别阈限

四、三种心理物理法的比较 185

第三节 心理量表法

一、顺序量表 188

(一) 等级排列法

(二) 对偶比较法

二、等距量表 193

(一) 感觉等距法

(二) 差别阈限法

三、比例量表 196

(一) 分段法

(二) 数量估计法

四、三种量表法的比较 200

本章实验

一、用极限法测定几种频率的听觉阈限 201

二、用平均差误法测定长度差别阈限 203

三、用恒定刺激法测定重量差别阈限 205

本章摘要

建议参考资料

费希纳 (Fechner, 1860) 在《心理物理学纲要》一书中提到了心理物理学有三种基本方法。这些方法称之为心理物理学方法 (或心理物理法) (psychophysical method)。心理物理法, 顾名思义, 乃是研究心理量和物理量之间的对应关系的方法。心理物理法是运用数学方法和测量技术, 研究心理现象和物理现象之间的数量关系。例如, 在光和视觉范围内, 辐射能的强度与光的明度相对应; 在声和听觉范围内, 声音的频率与音高相对应等等。在上述物理现象和心理现象中, 心理物理学把辐射能的强度、声音的频率看

作是刺激，把明度、音高看作是感觉。而感觉又必然是与之相应的刺激的一种函数。如光的辐射能强度越大，与之相应的感觉就越强。但是，光感觉如何随辐射能的强度增加而增加，这类问题也就是韦伯—费希纳律所研究的课题，只有借助于心理物理学方法，才能寻出各种感觉道中各自的与共同的规律。

心理物理法经过一百多年的发展、修改和补充，出现了许多变式。自坦纳和斯韦茨 (Tanner & Swets, 1954) 把信号检测论引入心理学领域以来，又为心理物理学研究提供了一个新的有力工具。后人为了有所区别，将信号检测论称之为新的心理物理法，而费希纳最早提出的三种基本的心理物理法称之为传统心理物理法。

本章内容主要讨论以下六个问题：

1. 阈限的操作定义是什么。
2. 何谓极限法？如何使用它。
3. 何谓平均差误法？如何使用它。
4. 何谓恒定刺激法？如何使用它。
5. 三种传统心理物理法的利弊是什么。
6. 三种心理量表法的制作方法是怎样的。

第一节 阈限的性质

一、心理物理法的诞生

心理物理法是与费希纳的名字联系在一起。实验心理学的主要先驱之一费希纳 (Gustav Theodor Fechner, 1801 ~ 1887) 在 1860 年发表了巨著《心理物理学纲要》，他在这部著作中探讨了心理量和物理量之间的函数关系。费希纳是德国心理学家，1801 年 4 月 19 日生于德国，1817 年入德国莱比锡大学学医，1822 年毕业，从事物理学的研究和译作。1834 年担任莱比锡大学物理学教授，不久又担任哲学教授。费希纳因研究后象，用有色镜注视太阳，以致伤害他的眼睛，加上身体虚弱，1839 年辞去物理学讲席。在病中他备受痛苦，三年不与任何人来往，罹患此病影响了他朝气蓬勃的事业。但他后来却又奇迹般地恢复了健康，社会上称这一时期是费希纳一生中的紧要关头，对于他的思想和后来的生活都有深刻的影响。费希纳喜欢谢林的自然科学，怀有宗教灵学的神秘思想，长期致力于寻求一种科学的方法，以使精神与物质两方面的范畴统一于灵魂之中。他在德国生理学家韦伯研究的基础上，假设最小可觉差 (或最小可觉差异) (just noticeable difference' 简称 j.n.d.) 是感觉单位，经过许多实验和推导，最后得出感觉与刺激的对数成比例的这一公式： $S=K \lg R$ (S 代表感觉，R 代表刺激，K 为常数)。他于 1860 年发表著名的《心理物理学纲要》一书，标志着心理物理学的诞生。

心理物理学 (psychophysics) 这个名词是由两个希腊字根 psyche 和 physike 所组成。费希纳的巨著《心理物理学纲要》震动了当时沉睡的世界。在《心理物理学纲要》公布后不久，费希纳的朋友沃尔克曼 (Alfred Wilhelm Volkmann, 1800 ~ 1877) 就发表了其关于心理物理学的论文。德尔帕夫 (Joseph Leopold Delboeuf, 1831 ~ 1896) 于 1885 年受到费希纳的鼓舞，开始其对视觉的实验，后来对心理学的发展颇有贡献。威洛特 (Karl Von Vierordt, 1818 ~ 1884) 于 1868 年对光成分的时间感觉进行研究。特别是艾

宾浩斯 (Hermann Ebbinghaus, 1850 ~ 1909) 受费希纳著作的鼓舞, 着手对较高级的心理过程“记忆”进行了测量。艾宾浩斯以自身为被试, 系统地测量了他自己的记忆能力以及记忆的保持过程。他不用记忆的主观经验, 取而代之的是客观指标——回忆量。在艾宾浩斯之后不久, 比内 (Alfred Binet 1857 ~ 1911) 继而把心理物理法应用于智力测量。当然, 一门自然科学的成熟程度与它是否能够测量有关, 而测量是要将带有一定规则的对象和事件以数的形式表现出来, 这就是说, 它在多大程度上可利用数字。可以认为, 经历了许多年的努力后, 心理物理学已经解决了它的核心问题: 测量心理量和物理量间的函数关系以及制定达到正确测量的实验方法。

二、 阈限及其操作定义

上面谈到, 心理物理法主要用于对阈限的测量。因此有必要先解释一下阈限 (threshold) 这个重要概念。阈限可分为二种, 即绝对阈限 (absolute limen, 简称 AL) 和差别阈限 (difference limen, 简称 DL), 可用心理物理法来测定。在一些普通心理学的教科书上, 把绝对阈限定义为刚刚能引起感觉的最小刺激量; 把差别阈限定义为刚刚能引起感觉的最小差别, 按照这样的定义, 还不能对这两种阈限进行测定。根据实验我们知道, 某一特定数值的刺激, 在重复作用于感官时, 被试有时报告为“无感觉”, 有时报告为“有感觉”, 有时还报告为“刚刚有一点点感觉”。人的感受性的这种随机性变化, 在每一种感觉道中都能发现。这种随机变化往往与 (1) 如何测量阈限有关; (2) 被试对任务的注意程度及其态度有关; (3) 被试的感受性暂时出现不稳定的摇摆有关。不过, 这种随机变化的感觉, 其次数分配基本上呈常态分布。因此, 根据统计学, 可以把那个可以刚刚引起感觉的最小刺激强度以其算术平均数来表示。而这个平均数恰好有 50% 的实验次数报告为“有感觉”的刺激强度, 由此可见, 阈限是个统计值。因而, 我们把阈限定义为: 有 50% 的实验次数能引起积极反应的刺激之值; 同理, 把差别阈限定义为有 50% 的实验次数能引起差别感觉的那个刺激强度之差。此定义要比普通心理学的定义具体, 也便于操作, 故称为阈限的操作定义。操作定义是由物理学家布里奇曼 (Bridgman, 1927) 提出的, 他主张一个概念应由测定它的程序来下定义。操作定义的提出受到心理学家的欢迎。在心理学上, 对一个心理现象根据它们的程序下定义就叫操作定义 (operational definition)。这样, 把“刚刚感受到”定义为“50% 次感受到”, 就可测定感觉阈限了, 参见图 4-2。

我们明确了阈限的操作定义, 就能够用心理物理学方法来测定阈限。关于测定阈限的实验, 一般具有下列三个明显的特点:

1. 选择好刺激系列和反应系列 用心理物理法求阈限的实验一般都经过预备实验。预备实验的首要目的就是确定和选择刺激系列和反应系列。例如, 要测定响度的绝对阈限, 可用音频发生器发出一系列强弱不同的声波强度。这一系列强弱不同的声波强度, 称为刺激系列。当这一系列强弱不同的声波强度作用于人耳会引起一系列大小不同的感觉 (响度), 这些所引起一系列大小不同的感觉, 成为反应系列。一般而言, 刺激系列可从零开始一直延伸到无穷大, 而反应系列则比刺激系列要短些。这是因为, 有些刺激太弱, 不能引起感觉; 有些刺激太强, 相应的感官受不了。因此, 刺激系列的范围可在对反应系列的预测基础上作合理的选择, 一般情况下, 刺激系列的范围要大于反应系列的范围。在图 4-2 上以物理量 (刺激) 为横座标, 以心理

量（反应概率）为纵座标，说明了这个道理。

2. 要尽量简化被试对刺激所作的反应 一般说来，在测定绝对阈限时，只要求被试探察刺激是否存在即可，在测定差别阈限时，只要求被试判定两个刺激是否相同还是相异。若被试对刺激所作的反应较复杂，那么这种复杂的反应将影响测定阈值的准确性。假定 A、B、C、D 代表四种不同的物理刺激，那么，在心理物理学实验中被试者一般作下列几种反应：

- (1) 有 A、或无 A；
- (2) A 等于 B、或 A 不等于 B；
- (3) A 大于 B、或 A 小于 B；
- (4) A 是 B 的两倍、或 C 等于 D 的一半；
- (5) A 与 B 的差和 C 与 D 的差相同。

以上五类选择和判断都很简单，实验时，采取哪类选择，视情况而定。

3. 测定的次数要多些 阈限的定义明确指出，阈限值不是一个突然的分界点，而是一个逐渐的过渡区，必须反复加以测量。因此测定阈限时，一定要经过测验才能确定某一刺激强度是否是阈限值，一般都需要上百次，特别是恒定刺激法，所需次数更多。当然，过多的测量容易引起被试者的调感和疲劳，这也是需要防止的。

第二节 测定阈限的三种基本方法

传统心理物理学 (classical psychophysics) 所处理的问题大体分两大类。第一类是感觉阈限的测量。测量感觉阈限的基本方法有：(1) 极限法；(2) 平均差误法；(3) 恒定刺激法。第二类是阈上感觉的测量，即心理量表的制作问题。本节讨论第一类问题，第二类问题在下节中论述。

一、极限法

极限法 (limit method) 又称最小变化法 (minimal-change method)、序列探索法 (method of serial exploration)、最小可觉差法 (或最小差异法) (method of least difference) 等，是测量阈限的直接方法。极限法的特点是：将刺激按递增或递减系列的方式，以间隔相等的小步变化，寻求从一种反应到另一种反应的瞬时转换点或阈限的位置。极限法既可用于测定绝对阈限，也可用于测定差别阈限。

(一) 用极限法测定绝对阈限

1. 自变量 用极限法测定绝对阈限，自变量是刺激系列。刺激系列要按递增或递减系列交替呈现。递增时，刺激要从阈限以下的某个强度开始；递减时，刺激系列的起点要大于阈限的某个强度，一般应选 10 到 20 个强度水平。为了使测定的阈限准确，并使每一刺激系列的阈限能相对稳定，一般递增和递减刺激系列要分别测定 50 次左右 (共 100 次左右)、刺激应由实验者操纵。为了避免被试者形成定势，每次呈现刺激的起点不应固定不变，而应随机变化。

2. 反应变量 用极限法测定绝对阈限的反应变量时，要求被试以口头报告方式表示。当刺激呈现之后，被试感觉到有刺激，就报告“有”，当被试没感觉到有刺激，就报告“无”，其依据是被试的内省，而不是刺激是否呈现。被试报告后，主试以“有”“无”或“+”“-”记录被试的反应，每个系列都需要被试作“有”到“无”、或“无”到“有”这两种报告，亦即，递增时直到第一次报告“有”之后，这一系列才停止；递减时直到第一次报

告“无”之后，这一系列才停止，然后再进行下一个系列。若被试者在这二类判断中有“说不准”的情况，则可以猜，但不可放弃。

3. 阈限的确定 在一个刺激系列中，被试者报告“有”和“无”这两个报告相应的两个刺激强度的中点就是这个系列的阈限。表 4-1 是以极限法测定音高绝对阈限的实例。表的下部显示：把各刺激系列的每一个阈限计算出来，然后分别求出递增系列或递减系列的平均数，之后再求出此两系列的平均数，最后求出的平均 11.7（赫兹）就是阈限值。

4. 误差及其控制 用极限法求绝对阈限经常会产生一些误差。在这些误差中，有些是由直接对感觉产生干扰的因素引起的；还有些是非感觉方面的因素引起的，如习惯和期望、练习和疲劳、时间和空间等等。这些因素在测定阈限的过程中经常起作用，以致使测定结果产生一定倾向的误差。这类误差叫做常误（constant error）。检查常误和消除常误的方法很多，要根据引起常误的原因不同而采用不同的方法。

极限法测定绝对阈限产生的误差主要有四种：习惯误差和期望误差、练习误差和疲劳误差。这四种误差恰好组成两对：习惯误差和期望误差相对，练习误差和疲劳误差相对。我们依然用表 4-1 的实例来分析这四种误差的情况。

在极限法实验中，由于刺激是按一定的顺序呈现的，被试在长序列中有继续给同一种判断的倾向，如在下降序列中继续说“有”或“是”，在上升序列中继续说“无”或“否”，这种被试习惯于前面几次刺激所引起的感觉叫作习惯误差（error of habituation）。由于习惯误差在递增法序列中，即使刺激强度早已超出阈限，被试仍报告感觉不到，这就会使测得阈值偏高；相反，在递减法序列中，即使刺激强度早已小于阈限，被试仍报告有感觉，这就会使测得的阈值偏低。与习惯误差相反的是另一种误差叫期望误差（error of anticipation）。它表现为被试在长的序列中给予相反判断的倾

表 4-1 以极限法测定音高绝对阈限的记录

次数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
增减系列 刺激 值—乐 音·频 率(赫)																
	21	+								+						
	20	+								+						
	19	+				+				+			+		+	
	18	+		+		+				+			+		+	
	17	+		+		+		+		+			+		+	
	16	+		+		+		+		+			+		+	
	15	+		+		+		+		+			+		+	
	14	+		+		+		+		+	+		+		+	
	13	+	+	+	+	+		+	+	+	-		+		+	
	12	-	-	+	-	-		+	-	+	-	-	+	+	+	+
	11		-	-	-	-	+	-	-	-	-	+	-	+	-	+
	10		-		-	-	-	-		-	-	-		-	-	-
	9		-		-	-	-	-		-	-	-		-		-
	8		-					-				-		-		-
	7							-				-		-		-
	6								-				-			
5								-				-				
阈限值	125	125	115	125	125	105	115	125	115	135	105	125	115	105	115	105
总平均值	M = 11.7		=0.87			=0.20										

(采自赫葆源、张厚粲、陈舒永，1983)

向，期望转折点的尽快到来。用递增法测定时，阈值就会偏低；用递减法测定时，阈值就会偏高。

为了使习惯误差和期望误差彼此抵消，可交替使用递增法序列和递减法序列，但可能其中的一种倾向较大。阈值用符号 T (threshold 的第一字母) 来表示。如果在上升序列中求得阈值 $T_{上}$ 大于在下降序列中求得阈值 $T_{下}$ ，即表示有习惯误差；反之，如果在上升序列中求得阈值 $T_{上}$ 小于在下降序列中求得阈值 $T_{下}$ ，则表示有期望误差。表 4-2 分别求出了上升与下降序列音高的绝对阈限的均数的差别(0.4)，经过 t 检验，表示序列误差是不显著的(本例 $p > 0.05$)。上升序列中求得阈值 $T_{上}$ 与下降序列中求

表 4-2 用不同系列测得阈限的比较

序列 比较项目	上升序列	下降序列	前 10 个系列	后 10 个系列	全部材料
绝对阈限	11.9	11.5	12.1	11.3	11.7
标准差	0.92	0.77	0.80	0.75	0.84
标准误	0.29	0.24	0.24	0.25	0.19
差异	0.4		0.8		
自由度	9		9		
t	1.1		2.28		
p	> 0.05		< 0.05		

(采自赫葆源、张厚燊、陈舒永, 1983)

得阈值 T_{\downarrow} 无显著性差异。

本实验还有练习和疲劳的影响。练习误差 (error of practice) 是由于实验的多次重复, 被试逐渐熟悉了实验情景, 对实验产生了兴趣和学习效果, 而导致反应速度加快和准确性逐步提高的一种系统误差。与此相反, 由于实验多次重复, 随着实验进程而发展的疲倦或厌烦情绪的影响, 而导致被试反应速度减慢和准确性逐步降低的一种系统误差, 称之为疲劳误差 (error of fatigue)。随着时间的进展, 练习可能使阈限降低, 而疲劳可能使阈限升高。为了检查有无这两种误差就要分别计算出前一半实验中测定的阈限与后一半实验中测定的阈限, 若前一半实验中测定的阈限比后一半实验中测定的阈限大, 并差别显著时, 就可以认为在测定过程中有练习因素的作用, 若前一半实验中测定的阈限比后一半实验中测定的阈限小, 并差别显著时, 就可以认为在测定过程中有疲劳因素的作用。为了消除练习和疲劳的影响, 就需作以下安排: 如以“ ”代表递增, 以“ ”代表递减, 并以四次为一轮, 就可以按照“ ”或“ ”排列。实验按此原理, 一般 10 次为一轮。总之, 递增和递减各自所用的次数要相等, 整个序列中在前在后的机会也要均等。练习误差和疲劳误差两种影响可能互相抵消, 也可能一种影响比另一种影响要大。为了比较两种影响的大小, 在表 4-2 中分别求出前 10 个系列与后 10 个系列的平均阈限值, 经比较, 两者差异为 0.8, 统计检验, $p < 0.05$ 。就是说, 在这个实验中, 练习比疲劳的影响要大得多, 练习显著地降低了阈值。

(二) 用极限法测定差别阈限

1. 自变量 用极限法测定差别阈限时, 每次要呈现两个刺激, 让被试比较, 其中一个为标准刺激 (standard stimulus, 简称 St), 即刺激是固定的, 其强度大小不变; 另一个是比较刺激 (comparison stimulus) 又称变异刺激 (variance stimulus, 简称 Sv), 即刺激的强度按由小而大或由大而小顺序排列。标准刺激和比较刺激可同时呈现, 标准刺激在每次比较时都出现, 比较刺激按递增或递减系列, 以测定绝对阈限的同样方法与标准刺激匹配呈现。

2. 反应变量 用极限法测定差别阈限的反应变量要求被试以口头报告方式表示, 一般用三类反应, 将比较刺激与标准刺激加以比较, 当比较刺激大于标准刺激时, 主试记录“+”; 当比较刺激等于标准刺激时, 主试记录“=”; 当比较刺激小于标准刺激时, 主试记录“-”。当被试在比较时表示怀疑, 可

记作“？”

3. 差别阈限的确定 确定差别阈限时先要求得一系列的数据，这些数据有：(1) 在递减系列中最后依次“+”到非“+”（即“=”或“-”或“？”）之间的中点为差别阈限的上限（upper limit，用符号 L_u 表示）；第一次非“-”到“-”（即“=”或“+”或“？”）之间的中点为差别阈限的下限（lower limit，用符号 L_l 表示）。(2) 在递增系列中最后依次“-”到非“-”（即“=”或“+”或“？”）之间的中点为差别阈限的下限（ L_l ）；第一次非“+”到“+”（即“=”或“-”或“？”）之间的中点为差别阈限的上限（ L_u ）。(3) 在上限与下限之间的距离为不肯定间距（ I_u ）。(4) 不肯定间距的中点是主观相等点。在理论上主观相等点（或主观等点）（point of subjective equality，简称 PSE）应与标准刺激（ S_t ）相等，但实际上两者有一定的差距，这个差距称为常误（constant error，简称 CE）。(5) 取不肯定间距的一半或者取上差别阈（ $DL_u=L_u-S_t$ ）和下差别阈（ $DL_l=S_t-L_l$ ）之和的一半为差别阈限。以上阐述可用公式表示为：

$$DL_u=L_u-S_t$$

$$DL_l=S_t-L_l$$

$$CE=S_t-PSE$$

$$DL = \frac{(L_u + S_t) + (S_t - L_l)}{2} = \frac{DL_u + DL_l}{2} = \frac{L_u - L_l}{2}$$

4. 误差及其控制 与用极限法求绝对阈限一样，在测定差别阈限时，也必须想方设法控制常误。除了要控制习惯和期望误差外，还要控制因标准刺激和比较刺激同时呈现所造成的误差。控制方法可采用多层次的 ABBA 法，如下形式为一种多层次的 ABBA 法的控制：

ABBA 法的控制形式

比较刺激系列呈现顺序	
标准刺激呈现位置	左右 右左 右左 左右
相继呈现的先后顺序	前后 后前 前后 后前

下面我们以时间辨别实验说明用极限法如何测定差别阈限。此实验采用 AB 法。实验结果见表 4-3。本实验以 0.40 秒的持续时间为标准刺激，以 0.20 秒至 0.56 秒，间隔 0.04 秒为比较刺激。主试先呈现标准刺激，然后呈现比较刺激，以小等间距逐渐增加，直至被试报告“=”，并第一次以长于标准刺激的反应为止。以递减法进行实验时，比较刺激的起点应长于标准刺激的持续时间，以小等间距逐渐减少，直至被试报告“=”，并第一次以短于标准刺激的反应为止。为了控制常误，可用比较刺激先于标准刺激的次数占全部实验次数的一半，并比较刺激的递增系列和递减系列各占一半。在比例中：

$$L_u=0.412 \text{ (秒)}$$

$$L_l=0.350 \text{ (秒)}$$

$$S_t=0.400 \text{ (秒)}$$

$$DL_u=0.412-0.400=0.012 \text{ (秒)}$$

$$DL_1 = 0.400 - 0.350 = 0.050 \text{ (秒)}$$

$$PSE = \frac{0.412 + 0.350}{2} = 0.381 \text{ (秒)}$$

$$CE = 0.400 - 0.381 = 0.019 \text{ (秒)}$$

$$DL = \frac{0.012 + 0.050}{2} = 0.031 \text{ (秒)}$$

表 4-3 以极限法测定时间的差别阈限的记录

次数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
增减系列																
变异刺激	.56		+										+			
	.52		+	+				+			+		+		+	
	.48		+	+			+	+	+		+		+		+	
	.44	+	+	+			+	+	+		+		=	+	+	
	.40	=	=	=	=	+	=	+	=	=	+	=	+	=	=	=
持续时间 (秒)	.36	-	-	-	=	=		=	=	-	=	-	=	=	=	-
	.32	-		-	-	=	-	-	-		=		-	-	-	
	.28	-		-												
	.24	-														
	.20	-														
上限	.42	.42	.42	.42	.38	.42	.38	.42	.42	.38	.42	.38	.46	.42	.42	
下限	.38	.38	.38	.34	.30	.34	.34	.34	.38	.303	.38	.34	.34	.34	.38	
M _{上限} = .412 M _{下限} = .350 DL = .031 PSE = .381																

(采自赫葆源、张厚燊、陈舒永, 1983)

二、平均差误法

平均差误法 (或均误法) (method of average error) 又称调整法 (method of adjustment)、再造法 (method of reproduction)、均等法 (method of equation), 是最古老且基本的心理物理学方法之一。虽然它最适用于测量绝对阈限和等值, 但也可用以测量差别阈限。平均差误法的特点是呈现一个标准刺激, 令被试再造、复制或调节一个比较刺激, 使它与标准刺激相等。比较刺激也可由实验者调节, 由被试判断。调整法曾是过去在天文学上常用的一个方法。那时让观察者调整一个光表面的强度, 使之与某星体的亮度相等, 这样就可以比较各星体间的亮度。因此, 这个方法是要求被试判断什么时候比较刺激和标准刺激相等。被试判断为与标准刺激相等的比较刺激, 并不总是一个固定的数值, 而是围绕着一个平均数变化的一个数。这个变化的范围就是不肯定间距。不肯定间距的中点就是主观相等点。通过对主观相等点和不肯定间距的测量, 就可以估计差别阈限。把平均误差作为差别阈限, 与差别阈限的定义并不相符, 但因为平均误差和差别阈限成正比, 所以也作

为测量差别感受性的指标。

(一) 用平均差误法测定绝对阈限

1. 自变量 用平均差误法测定绝对阈限，是让被试者调整一个比较刺激与一个标准刺激相等。不过，此时的标准刺激假设为零，即让被试者每次将比较刺激与“零”相比较。

2. 反应变量 用平均差误法测定绝对阈限的反应变量是被试者每次调整比较刺激与标准刺激相等的那个数值。

3. 绝对阈限的确定 让被试者每次调到刚刚感觉不到（即与“零”标准刺激等值），然后把各次测定数值加以平均即为绝对阈限。

(二) 用平均差误法测定差别阈限

1. 自变量 用平均差误法测定差别阈限，是向被试者呈现一个标准刺激，让其调整比较刺激。比较刺激是一种连续的量。在被试认为接近时，可反复调整，直到其认为满意为止。

2. 反应变量 用平均差误法测定差别阈限的反应变量是被试每次调整的数值，即其认为与标准刺激相等的数值。由于被试反复测试，每次的结果并不是一个固定的数值，它们是围绕着一个平均数变化的数值。这个变化范围就是不肯定间距。不肯定间距的中点，即多次调整结果的平均数，就是主观相等点，主观相等点与标准刺激的差就是常误。

3. 差别阈限的估计值的计算 用平均差误法求差别阈限，所得差别阈限只是一个估计值，平均差误（average error，用符号 AE 表示）有两种计算方法：

(1) 把每次的调整结果（X）与主观相等点（用 M 或 PSE 表示）的差的绝对值加以平均，作为差别阈限的估计，这个差别阈限的估计值用符号 AE 表示：

$$AE_M = \frac{\sum |X - PSE|}{N}$$

(2) 把每次调整结果（X）与标准刺激（S_t）的差的绝对值加以平均，作为差别阈限的估计，用符号 AE_{st} 表示：

$$AE_{st} = \frac{\sum |X - S_t|}{N}$$

(三) 误差及其控制

在平均差误法实验中，一般要被试自己操纵实验仪器来调整比较刺激，使其与标准刺激相等。这就要产生动作误差，亦即因被试所采用的方式不同而产生误差。若标准刺激和比较刺激是相继呈现的，又易产生时间误差。因此，在实验中应加以控制，控制方法依具体实验不同而不同，一般可采用多层次的 ABBA 法，还可使比较刺激从小到大，从大到小两方面来进行调整，以便控制动作误差等。

这里以长度差别阈限为例，来说明如何以平均差误法测定差别阈限。实验所用仪器是高尔顿（Galton, 1883）长度分辨尺。长尺中央有一分界线，分界线两侧各有一游标，尺的背面有刻度，可向主试显示被试在比较标准刺激与比较刺激时的差异。若标准刺激是 150 毫米，则被试的任务是调节比较刺激，使之与标准刺激相等。因长度分辨尺是视觉的，所以标准刺激设置的位置不同（或左侧或右侧），易产生空间误差。又因比较刺激的初始状态不

同（或长于标准刺激或短于标准刺激），因而被试者调整时或向里或向外移动游标的动作方式不同，易产生动作误差。为了控制这些误差，在整个实验中，标准刺激要有一半的次数在左边，另一半的次数在右边。实验结果见表 4-4。在表 4-4 中， AE_{st} 是表示用标准刺激来估计的平均误差，而 AE 是表示用平均数或 PSE 来估计的平均误差。两者都可作为差别阈限的估计值。

表 4-4 用平均差误法测定长度差别阈限的结果

X	$X-S_t$	$X-M$	计算
148	-2	-0.2	
143	-7	-5.2	$S_t=150$ 毫米
145	-5	-3.2	$M=PSE=148.2$ 毫米
147	-3	-1.2	$Ae_{st} = X-St /n=3.8$ 毫米
151	1	2.8	$AE_M = X-M /n=3.4$ 毫米
154	4	5.8	$CE=148.2$ 毫米-150 毫米=-1.8 毫米
144	-6	-4.2	
145	-5	-3.2	
154	4	5.8	
151	1	2.8	

（采自赫葆源、张厚粲、陈舒永，1983）

三、恒定刺激法

恒定刺激法（或固定刺激法）（method of constant stimulus）又叫正误法（true-false method）、次数法（frequency method），它是心理物理学中最准确、应用最广的方法，可用于测定绝对阈限、差别阈限和等值，还可用于确定其他很多种心理值。此法的特点是：根据出现次数来定阈限，即以次数的整个分布求阈限。具体作法如下：（1）主试从预备实验中选出少数刺激，一般是 5 到 7 个，这几个刺激值在整个测定过程中是固定不变的；（2）选定的每种刺激要向被试呈现多次，一般每种刺激呈现 50 到 200 次；（3）呈现刺激的次序事先经随机安排，不让被试知道。用以测量绝对阈限，即无需标准值，如用以确定差别阈限或等值，则需包括一个标准值；（4）此法在统计结果时必须求出各个刺激变量引起某种反应（有、无或大、小）的次数。

特别要注意的是，此法在实验之前需要选定刺激。所选定的刺激最大强度应为每次呈现几乎都能为被试感觉到的强度，它被感觉到的可能性应不低于 95%。所选刺激的最小强度应为每次呈现几乎都不能感觉到的强度，它被感觉到的可能性应不高于 5%。选定呈现刺激范围之后，再在这个范围内取距离相等的刺激，每种刺激强度呈现不得少于 50 次。此法虽然要求较大量的实验次数，但每次实验只用很短的时间，因此也比最小变化法节省时间，并且测得的阈限也更准确。

（一）用恒定刺激法测定绝对阈限

1. 自变量 用恒定刺激法测定绝对阈限，是从略高于感觉到略低于感觉这一范围内选 5 到 7 个等距的刺激强度。

2. 反应变量 用恒定刺激法测定绝对阈限的反应变量要求被试者以口

口头报告方式表示，在实验中每呈现一次刺激后，被试者若感觉到了，则报告“有”，主试者记录“+”；被试者若感觉不到，则报告“无”，主试者就记录“-”。然后根据被试者对不同刺激所报告的“有”或“无”的次数来求出百分数，以此来计算阈限。

3. 绝对阈限的计算 用恒定刺激法计算绝对阈限的方法很多，我们用测定两点阈的实验为例来说明这些方法。首先选出最大刺激和最小刺激，分别为 12 毫米和 8 毫米，在这个范围内选定间隔为 1 毫米的刺激 5 个：各刺激的两点距离分别为 8、9、10、11、12 毫米，每个刺激共呈现 200 次，所有刺激共呈现 1000 次，刺激按随机顺序呈现，每次呈现后，要求被试报告是“两点”还是“一点”。实验结果见表 4-5。

表 4-5 用恒定刺激法测定两点阈的实验记录

刺激（毫米）	8	9	10	11	12
报告“两点”的次数	2	10	58	132	186
报告“两点”的百分数	1 %	5 %	29 %	66 %	93 %

（采自 Woodworth & Schlosberg, 1954）

根据阈限的操作定义，两点阈应为 50% 的次数被感觉到的那个刺激的大小，可是在表 4-5 中，没有一个刺激的判断次数恰好有 50% 为感觉到的。可以看到，当刺激为 10 毫米时，其正确判断率为 29%；当刺激为 11 毫米时，其正确判断率为 66%。因此，可以断定 50% 的次数被感觉到的那个刺激值必在 10~11 毫米之间。为了求出这个刺激值，最常用的方法是直线内插法。直线内插法（linear interpolation）是将刺激作为横座标，以正确判断的百分数作为纵座标，画出曲线。然后再从纵轴的 50% 处画出与横轴平行的直线，与曲线相交于点 a，从点 a 向横轴画垂线，垂线与横轴相交处就是两点阈，其值等于 10.57 毫米（见图 4-3）。当然也可以用其他方法——诸如平均 Z 分数法（averaged Z scores），最小二乘法（least squares），斯皮尔曼分配法（Spearman distribution method）等——来求出阈限值。当实验次数很多，实验结果接近正态分布时，用这种直线内插法求出的阈限值与其他方法求出的阈限值是接近的。直线内插法是求阈限值的一种简便方法。

（二）用恒定刺激法测定差别阈限

1. 自变量 用恒定刺激法测定差别阈限，是让被试者将比较刺激与标准刺激加以比较，标准刺激是能被感觉到的某一刺激强度，比较刺激可在标准刺激上下一段距离内确定，一般从完全没被感觉出差别到完全感觉出差别的范围内选定 5 到 7 个刺激强度作为比较刺激。比较刺激要随机呈现，每个比较刺激与标准刺激至少要比 100 次。

2. 反应变量 用恒定刺激法测定差别阈限的反应变量要求被试者以口头报告方式表示，较早的方法是让被试作三类反应，即“大于”、“等于”和“小于”，分别记为“+”、“=”和“-”。后来有人提出，让被试作三类反应时，因其中有“等于”的反应，若被试较为自信，则作出“等于”的反应就较少；若被试较为谨慎，则作出“等于”的反应就较多。这样会直接影响到差别阈限的大小。所以，有的心理学家将“等于”的反应单列，并以此来分析被试的内在心理偏向。由于它会受到被试的态度的影响，为了消除这类影响，建议只让被试者作“大于”和“小于”两种判断，即使分不清时，

也要作出其中的一种判断。可见，三类反应和二类反应各有利弊，实验时可根据情况选定其中的一种。

3. 差别阈限的估计值的计算

(1) 三类反应的差别阈限的计算：这里以重量辨别为例。以 200 克的重量作为标准刺激，以从 185 至 215 克中的 7 个重量作为比较刺激，7 个重量间的间隔各为 5 克，要求被试作“重”、“相等”和“轻”这三类反应。实验结果见表 4-6。

表 4-6 用恒定刺激法测定重量差别阈限的结果

比较刺激 (克)	比较的结果(次数的%)			
	“+”	“=”	“-”	“+”+“=”
185	5	4	91	9
190	12	18	70	30
195	15	25	60	40
200	30	42	28	72
205	55	35	10	90
210	70	18	12	88
215	85	9	6	94

(采自赫葆源、张厚燊、陈舒永，1983)

根据表 4-6 的数据用直线内插法求得差别阈限。表 4-6 中的、
三个纵列的数据画出三条曲线，用直线内插法求得 50% 的次数被判断比标准刺激重的重量为 204.5 克和 50% 的次数被判断比标准刺激轻的重量为 196.6 克(见图 4-4)，这两个数值分别为上限和下限(即 $L_u=204.5$ 克， $L_l=196.6$ 克)。根据上限和下限，就可计算如下：

上差别阈 $=DL_u=204.5$ 克 - 200 克 = 4.5 克

下差别阈 $=DL_l=200$ 克 - 196.6 克 = 3.4 克

差别阈限 $=DL=(4.5+3.4)/2=3.95$ 克

(2) 二类反应的差别阈限的计算：这里以重量辨别为例。以 200 克的重量为标准刺激，以从 185 到 215 克中的间隔为 5 克的 7 个重量为比较刺激，要求被试者只作“重”和“轻”这两类反应。实验结果见表 4-7。用这种方法取得的实验数据，以重于标准刺激的比例和轻于标准刺激的比例分别画出两条曲线，这两条曲线恰好在 50% 处相交。如图 4-5 所示，两条曲线恰好在 50% 处相交，这说明 199 克有一半的次数被判断为重于标

表 4-7 两种判断的实验结果

比较刺激(克)	“+”的比例	“-”的比例
185	0.07	0.93
190	0.21	0.79
195	0.28	0.72
200	0.55	0.45
205	0.73	0.27
210	0.79	0.21
215	0.90	0.10

(采自赫葆源、张厚燊、陈舒永, 1983)

准刺激, 另一半的次数被判断为轻于标准刺激。也就是说, 它被判断为与标准刺激主观上相等。在这种特例的情况下, 就不可能取与 50% 点相对应的比较刺激作为上限和下限了, 与标准刺激完全能辨别的重量是在 100% 处, 与标准刺激不能辨别的重量则在 50% 处, 因此, 取此两者的中点 (亦即在 75% 处) 作直线内插与两条曲线相交, 以求得上限和下限。从图 4-5 可见, $L_0=206.6$ 克, $L_1=192.8$ 克, 差别阈限为

$$DL=[(206.6-200)+(200-192.8)]/2=6.9 \text{ 克}$$

这样求得的差别阈限与前面所提到的操作定义不相符合, 因为上限和下限与标准刺激比较都有 75% 的次数可辨别, 所以常把这种差别阈限称之为 75% 的差别阈限, 它是差别阈限中的一种。

四、三种心理物理法的比较

以上我们分别介绍了三种基本心理物理法的特点、用途、和计算方法, 虽然极限法、平均差误法和恒定刺激法这三种基本心理物理法都是用来测定感觉阈限 (sensation threshold), 但各有千秋。究竟使用哪种具体方法最为合适, 则要根据研究的对象、要求而定。下面我们从四个方面来比较这三种方法:

1. 从感觉阈限的含义上比较 根据感觉阈限的操作定义, 极限法求得的阈限值较能符合感觉阈限的操作定义。因为此法的操作是有系统地探查感觉的转折点, 它的计算过程就具体准确地说明了感觉阈限的含义。同时, 因被试知道刺激呈现的次序, 他可把注意力集中到特别需要集中的地方, 从而取得较好的实验效果。但也正是由于被试知道刺激呈现的次序, 从而容易产生期望误差和习惯误差。只要刺激序列可以按小阶梯变化, 都能用此法测定阈限, 它曾被用来测定声音、气味、味道、颜色、温度、痛、光及触觉的阈限。但递减法不宜用于测定味觉和嗅觉阈限, 因为嗅觉适应快, 味觉的后作用不易被消除。因此对于有一定后作用的刺激, 就不易采用极限法。

2. 从被试者方面比较 通常, 自变量 (刺激) 要通过被试者产生反应变化 (因变量)。测定的次数过多容易使被试者感觉单调而产生疲劳。若让被试者主动参与测定工作, 就可改变这种单调厌倦的情绪, 从而使测定结果更准确。平均差误法的实验程序, 对被试来说比较自然, 让被试自己主动调整比较刺激, 也可引起他的兴趣, 因此, 在这方面平均差误法优于极限法和恒定刺激法。

3. 从误差方面比较 在心理实验中, 除自变量外, 其他变量都应予以控制, 以减少误差。极限法中各刺激是按刺激量的大小依次改变的, 因此,

被试者回答较有把握，每次测定结果间的差异也就较小。但是，正是由于刺激量大小是依次改变的，因而易产生期望误差和习惯误差。恒定刺激法所用的刺激数目较少，且不需随时调整刺激的强度，因此，用这个方法测量那些不易随时改变强度的刺激（诸如重量之类）则较为方便。又因为刺激是随机呈现的，不会像极限法那样产生期望误差和习惯误差。在用三类判断测定差别阈限时，不肯定间距的大小随被试的态度而变化，从而对所测的差别阈限产生影响。用二类判断测差别阈限，虽然可以避免这一问题，但迫使被试作出肯定回答是不自然的，因为被试确实有难于判断比较刺激与标准刺激哪一个强或弱的情况。此外，在恒定刺激法中，刺激的变化是没有什么方向的，被试者在回答时猜的成分较多，因而产生较大的误差。在平均差误法中，由于它获得数据的标准和计算的方法有所不同，测得的结果可以说只是一个近似值，因此它测定的阈限不能直接与由其他方法求得的阈限进行比较。

4. 从效率上比较 效率是心理实验的质量指标。平均差误法中刺激是由被试者自己调节，回答和记录的次数较少，因而能较快地得到测量结果。而极限法中要回答和记录的次数较多，其效率不及平均差误法。恒定刺激法可不必临时改变刺激，这是其优越之处。在测重量阈限时，通常用恒定刺激法。而且恒定刺激法中的每一个记下的数据都可以利用上，不像在极限法中用来计算阈限值的数据是少数，因此，恒定刺激法的效率较高。

应当指出，上述三种方法的比较是相对的，在确定用哪一种方法时要同时考虑研究对象、要求和各种方法的优缺点。但是，在测定要进行比较的两种阈限时，必须用同一种方法，因为，不同的方法所用的指标不同。

第三节 心理量表法

上一节讨论了心理物理学中感觉阈限的测量。现在来讨论阈上感觉如何度量的问题。心理量表法构成了传统心理物理学的另一个重要组成部分。

物理刺激可以用物理量来表示，例如用“尺”来量长度，用“斤”来称重量等，有了像“尺”、“斤”等单位所表示的物理量表，我们就可以根据需要来改变刺激的强度，但是，要度量“响度”，例如“一个声音比另一个声音响多少？”这样的问题，单有物理量表是不够的。假设一个无线电工程师要设计一架收音机，使它的响度是另一架的两倍，如果他只把物理的输出加倍，那么，他会发现所增加的响度根本不是另一架的两倍，而是稍增加了点。这例子说明只用物理量表是不能解决的，这是因为刺激的物理值的等量增减并不引起感觉上等量的变化。为了要了解刺激的变化和感觉的变化之间的关系，就需要一个测量心理变化的方法，这种方法就是能够度量阈上感觉的心理量表（psychological scales）。

从量表有无相等单位和有无绝对零点来分，心理量表可分为顺序量表、等距量表和比例量表三类。下面我们分别论述。

一、顺序量表

顺序量表（或序级量表）（ordinal scale）是一种较粗略的量表，它既无相等单位又无绝对零点，只是把事物按某种标准排出一个次序。例如赛跑时不用秒表测速，按先到终点的算第一名，第二个到的算第二名等等。这样，我们可以按快慢排出一个顺序，这个顺序在一定程度上也可以表示跑的速度量。因为它没有相等单位，所以就是知道了第二名比第一名慢5秒，也不能

以此推算出第三名比第二名慢几秒；又因为它没有绝对零点，所以不能推知第二名的速度是第三名的几倍。所以，顺序量表只是在一个分类基础上对事物进行分类，每一类别只具有序列性，并不表示数与数之间的差别是相等的。

顺序量表的制作方法比较简单，一般用等级排列法和对偶比较法来制作顺序量表。

（一）等级排列法

等级排列法（或等级法）（rank-order method）是一种制作顺序量表的直接方法。这个方法是把许多刺激同时呈现，让许多被试者按照一定标准，把这些刺激排成一个顺序，然后把许多人对同一刺激评定的等级加以平均，这样，就能求出每一刺激的各自平均等级，最后，把各刺激按平均等级排出的顺序就是一个顺序量表。

下面我们举一个例子。某广告公司要对 10 张广告的优劣作评比，请来了 17 名评判者（即被试者）。评判的方法是让被试者将 10 张广告排成美到丑的一个顺序。通过众多被试者的比较，就可求出全体被试者对同一广告评判的等级的平均值，这个值就是广告的平均等级。各广告按平均等级排出的顺序就是一个顺序量表。

表 4-8 上我们可以看清楚等级排列法制作顺序量表的具体过程。表中 A、B、C 等英文字母，代表广告代号。纵列上的 1、2、3 等为被试者代号。广告代号下的数字，表示被试者对此广告的评分等级。如果被试者把它评为 1，这就表示此广告比其余 9 个好，列第一位，余类推。把某一广告

表 4-8 等级排列法的实验结果和具体运算过程

广告 评判者	A	B	C	D	E	F	G	H	I	J
1	5	3	4	7	1	9	2	8	6	10
2	4	2	5	6	1	10	3	7	8	9
3	6	4	3	5	1	9	2	8	7	10
4	6	3	4	7	1	10	2	9	5	8
5	5	1	3	6	2	8	4	0	7	9
6	6	4	3	5	1	9	2	7	10	8
7	6	3	4	5	2	9	1	8	7	10
8	5	2	4	6	1	8	3	9	7	10
9	3	5	2	4	1	9	6	10	8	7
10	4	7	3	5	1	10	2	8	6	9
11	5	3	2	6	1	9	4	8	7	10
12	4	3	1	6	2	10	5	8	9	7
13	5	2	4	7	1	8	3	6	9	10
14	5	4	3	7	1	9	2	6	10	8
15	5	2	3	4	1	10	9	7	8	6
16	6	2	5	4	1	8	3	9	7	10
17	6	3	4	8	1	7	2	5	10	9
等级总和	86	53	57	98	20	152	55	131	133	150
平均等级 (MR)	5.06	3.12	3.35	5.76	1.18	8.94	3.24	7.71	7.82	8.82
平均选择 分数 ($M_c = n - MR$)	11.94	13.88	13.65	11.24	15.82	8.06	13.76	9.29	9.18	8.18
$P = M_c / (n - 1)$	0.75	0.87	0.85	0.70	0.99	0.50	0.86	0.58	0.57	0.51
$M'c = M_c + 0.5$	12.44	14.38	14.15	11.74	16.32	8.56	14.26	9.79	9.68	8.68
$P' = M'c / n$	0.73	0.85	0.83	0.69	0.96	0.50	0.84	0.58	0.57	0.51
Z	0.61	1.04	0.95	0.50	1.75	0	1.00	0.20	0.18	0.03
Z'	0.61	1.04	0.95	0.50	1.75	0	1.00	0.20	0.18	0.03
名次	5	2	4	6	1	10	3	7	8	9

(采自杨治良, 1983)

的等级全部加起来, 被 n (被试者人数) 除, 就是平均等级。C 是选择分数, 相应的 P 分数说明广告成功的百分数。通过左列公式, 可先后计算出 M_c 、 P 、 $M'c$ 和 P' 值。有了 P 值, 我们就可以从 PZO 转换表 (见第五章表 5-6) 上查出 Z 值。最后, 以最小值作 0, 消除负值算出 Z' 值, 并在 Z 轴上表示出结果, 见图 4-6。

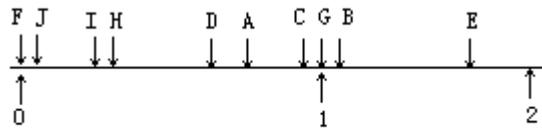


图4-6 17人对10张广告爱好的程度

(二) 对偶比较法

对偶比较法（或配对比较法）（method of paired comparison）是把所有要比较的刺激配成对，然后一对一对地呈现，让被试者对于刺激的某一特性进行比较，并做出判断：这种特性的两个刺激中哪一个更为明显。因为每一刺激都要分别和其他刺激比较，假如以 n 代表刺激的总数，所以配成对的个数是 $n(n-1)/2$ 。如共有 10 个刺激则可配成 45 对。最后依它们各自更明显于其他刺激的百分比的大小排列成序，即可制成一个顺序量表。

如果有五种样品，A、B、C、D、E，则可配成 $5 \times (5-1)/2=10$ 对。这 10 对呈现次序如表 4-9 所示。

表 4-9 10 对样品呈现次序

样品	A	B	C	D	E
A	-				
B	1	-			
C	5	2	-		
D	8	6	3	-	
E	10	9	7	4	-

（采自赫葆源、张厚燊、陈舒永，1983）

如果各对样品同时呈现，则要消除空间误差（space error）——即样品在空间中不同方位呈现，于判断时产生的误差现象。若第一轮以 AB 形式呈现，则第二轮中以 BA 形式呈现即左右颠倒。如果是相继呈现，则要注意消除时间误差（time error）——即相等的二个样品在先后不同时间出现，于判断时产生的误差。若第一轮以先 A 后 B 次序相继呈现，则第二轮要按先 B 后 A 次序相继呈现。比较时，把比较的结果填入事先准备好的空表中，例如横 A 与纵 B 比较，被试者认为“A”更好，则在横 A 与纵 B 交叉处写上“A”；如果横 E 与纵 D 相比较，被试者认为“D”更好，则在它们交叉处写上“D”，按此比较 10 次后，再倒过来比较，如纵 A 与横 B 比较，被试者认为 A 更好，则在它们交叉处写上“(A)”，余类推。参见表 4-10。

从表 4-10 上可见对偶比较法数据处理的顺序。在全部实验做完以后，先把 A 行、A 列中“A”或“(A)”出现的总次数记在 A 列下面，余类推，并把第一轮与第二轮的次数加起来作为被选中的分数 C。因为每个刺激要和 $n-1$ 个另外的刺激相比，所以每个刺激和其他刺激相比的次数为 $n-1$ 。为了消除空间或时间误差，又要倒过来重比 $n-1$ 次，因此每个刺

表 4-10 10 对样品对偶比较数据整理

	A	B	C	D	E
A	-				
B	A (A)	-			
C	A (C)	B (B)	-		
D	A (A)	B (D)	C (D)	-	
E	A (A)	B (B)	C (C)	D (E)	-
第一轮	4	3	2	1	0
第二轮	3	2	2	2	1
总计 C	7	5	4	3	1
$P = \frac{C}{2(n-1)}$	0.88	0.63	0.50	0.38	0.13
$C = C + 1$	8	6	5	4	2
$p = C / 2n$	0.80	0.60	0.50	0.40	0.20
Z	+0.84	+0.25	0.00	-0.25	-0.84
Z	1.68	1.09	0.84	0.59	0.00
顺序	1	2	3	4	5

(采自杨治良等, 1988) 激比较的总次数为 $2(n-1)$, 所以, 在计算选中比例 (P) 时, 要把选中分数 C 除以 $2(n-1)$, 但严格地说, 每种刺激和它本身也应比较, 只不过是分胜负, 故在每个 C 分数上都加 0.5 , 因其比较两次, 所以要加 $(0.5 + 0.5) = 1$, 把增加后的选中分数称作 C' , C' 的比例 P' , 可用 $C' / 2n$ 计算出来。根据选中分数或选中比例, 只能得出被试者对刺激爱好的顺序, 即根据数据可排出喜爱的顺序为 ABCDE。如果把 P 转换成 Z 分数, 就可得出对各样品爱好的程度的相对大小。如表中 Z 行所示, 因为这个程度本来就是相对的, 为了消除负值, 可以把每个 Z 分数加上 0.84 , 使起点恰好从零点开始, 结果如表 4-10 中 Z' 行所示。

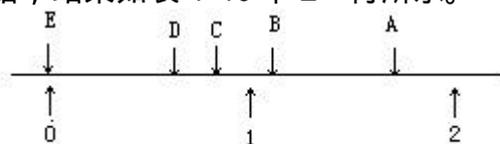


图4-7 某被试者对五种样品爱好的顺序

(采自杨治良, 1988)

从上面分析可以看出, 这个被试者最喜欢“ A ”, 最不喜欢“ E ”。根据 Z 行数据, 可画出此被试者对这几种样品喜爱的程度, 如图 4-7 所示。

根据我国心理学家陈立等 (1965) 的研究, 不同年龄的人都有类似的颜色爱好的心理量表, 就红、蓝、绿、黄四种颜色而言, 红最喜爱, 其后的顺序是蓝、绿、黄 (见图 4-8)。

二、等距量表

等距量表 (equal interval scale) 是一种有相等单位但没有绝对零点的量表。等距量表除了具有类别量表和等级量表的性质外, 它要求整个量表的各个部分的单位是相等的, 但没有绝对的零点。例如在普通温度上 60° 和 70° 的差别与 70° 和 80° 的差别是一样的, 都是 10° 。这在量表上单位是等距的。但它的参照点是人定的, 这不是绝对的零点。而温度的绝对零点是

摄氏零下 273 。因此我们不能说 80° 是 40° 的二倍。因为这是不合理的，等距量表没有绝对的零点，只能做加减的运算，不能作乘除的运算。

心理学家们对等距量表有更多的兴趣，这是因为：（1）根据观测值间一致变化效应的规律，在一组资料中，我们能对每一个观测值用同一常数作加减乘除运算，并不破坏这些数值的关系，因此在一个等距量表上所得到的观测值可以转换到另一个与此组资料计算的参照点和单位不同的等距量表上去；（2）与顺序量表相比，它能较广泛地应用统计方法，如计算均数、变差、相关系数等统计量，以及应用 t、F 检验等。

一般用感觉等距法和差别阈限法来制作等距量表。

（一）感觉等距法

要制作一个等距量表，最直接的方法是采用感觉等距法（equal sensedistance method），它是通过把一个感觉分成主观上相等距离来制作。它要求被试者将某种感觉上的一段心理量分成两个或两个以上的等分。最简单的是二分法（bisection）。例如 R_1 和 R_5 是两个不同响度的声音，且 $R_5 > R_1$ 。被试者的任务是：（1）找出 R_3 ，使其响度正好在 R_1 和 R_5 之间也即 $R_5 - R_3 = R_3 - R_1$ 。（2）再找出 R_4 ，使其响度正好在 R_3 和 R_5 之间，也即 $R_5 - R_4 = R_4 - R_3$ 。（3）再找出 R_2 ，使其响度正好在 R_1 和 R_3 之间，也即 $R_3 - R_2 = R_2 - R_1$ 。这样，利用三次二分法把 R_1 和 R_5 之间在响度上分成四分，即 $R_5 - R_4 = R_4 - R_3 = R_3 - R_2 = R_2 - R_1$ 。这就得到了按等距变化的一系列刺激。把这一系列的刺激作为横座标，把等响单位作为纵座标画出一条曲线，这就是响度的等距量表。

用感觉等距法制作等距量表，当然不限于二分法，也可用调整法同时分出几个等分来，例如在史蒂文斯和沃尔克曼（Stevens & Volkman, 1944）的实验中，要求被试把 200 ~ 6500 赫兹的纯音，按音高分出四个等分，即在这个范围内，同时找出三个频率的声音，使五个纯音依次在音高差别的感觉上相等。多人多次调整结果的平均值如表 4-11。根据表中所列的数据，以声音频率为横坐标，以音高单位为纵坐标，画出的音高等距量表如图 4-9 所示。

表 4-11 调整纯音的频率使其在音高上等距的实验结果

声音的频率（赫兹）	200	867	2022	3393	6500
音高单位	0	1	2	3	4

（采自 Stevens & Volkman, 1944）

（二）差别阈限法

差别阈限法（或差异阈限法）（differential threshold method）是制作等距量表的一种间接方法，通过在不同强度的基础上测量最小可觉差来实现。具体地说，用任何一种古典的心理物理法测出感觉的绝对阈，并以此为起点，产生第一个最小可觉差的刺激强度，以第一个最小可觉差为基准，再测量第二个最小可觉差……。这样测得许多最小可觉差以后，以刺激强度为横坐标，以绝对阈以上的最小可觉差数为纵坐标，画出的心理物理关系图就是等距量表。实验者就可从图上找出产生某一感觉水平所需要的刺激值。以表 4-12 的实验数据为例，这些结果在图 4-10 中很清楚地反映出来；随着最小可觉差的梯级增高，在刺激值方面需要的增加量越来越大。用刺激值来

作等距量表，虽然我们选择的绝对阈限值和韦伯分数都尽量少带小数，但到第五个梯级数字就显得累赘了。在小数增加到太多以前，我们必须开始设法削减小数。用刺激值的对数就方便得多，从表 4-12 的第三列相应的对数值中，我们可以看出对数的增加量（表 4-12 最后一列）是不变的。当以刺激值的对数为横坐标画图时，我们就得到一条直线而不是一条曲线，参见图 4-11。我们发现，当只涉及到强和弱的感觉的相对强度时，用对数就显得特别方便。我们不需要把中间的梯级加起来，就可计算出任何感觉水平所需
（采自 Woodworth & Schlosberg, 1954）195

表 4-12 某一实验的实验数据

最小觉察的梯级	刺激值	刺激的对数	对数的增加量
0 (绝对阈限)	8.0	0.903	
1	12.0	1.079	0.176
2	18.0	1.255	0.176
3	27.0	1.431	0.176
4	40.5	1.607	0.176
5	60.75	1.784	0.177

（采自 Woodworth & Schlosberg, 1954）

要的刺激值的对数。把梯级的数目和增加量的对数（即韦伯分数加 1 的对数）相乘，再把它和绝对阈限值的对数相加，就可得到所需要的刺激值的对数。费希纳为进行这种演算推导了几个公式，其中最著名的公式是：

$$S=KlgR$$

我们可将这个公式读为：感觉的强度的变化和刺激的对数变化成正比。我们把这种以差别阈限为根据所制成的量表的方法，叫作差别阈限法。费希纳定律就是一种等距量表。

三、比例量表

比例量表（或比率量表）（ratio scale）既有真正的零点，也是等距的。一个比例量表除含有名称、等级这些等距量表的特征外，还有一个具有实际意义的绝对零点。它是测量的最高水平，也是科学家认为理想的量表。若一项测量结果在比例量表上是零，那么，我们可以说某个事物并不具有被测量的属性或特征。由于它具有绝对的零点，且量表上的单位相等，因此就可进行加、减、乘、除四则运算。在物理学中，我们所用的绝对温度量表就属于这类量表。更为熟悉的是公制的尺所采用的量表制度。在心理测量中，智力的绝对零点是难以决定的，同时，由于等距量表对于大多数心理测量已经足够，所以，我们就不考虑用比例量表作为心理测验的标准化记分量表。但比例量表所适用的统计方法，除上面在等距量表上可采用的统计方法外，还可用几何平均数和相对差异量等计算，所以，对于心理物理学来说，应尽可能使用比例量表。

比例量表的制作方法有分段法和数量的估计法，现分述如下：

（一）分段法

分段法（fractionation method）是制作感觉比例量表的一种最直接的方法。这个方法是通过把一个感觉量加倍或减半或取任何其他比例来建立心

理量表的。具体作法是呈现一个固定的阈上刺激作为标准，让被试者调整比较刺激，使它所引起的感觉为标准刺激的一定比例，例如，2 倍、3 倍、1/2 倍、或 1/3 倍等等。每个实验只选定同一个比例进行比较，同一个标准刺激比较若干次后，再换另外几个标准刺激进行比较。如果所选定的比例是 2 : 1 的话，就可找出哪些刺激所引起的感觉为标准刺激所引起的感觉的一半。当把所有的标准刺激都比较完之后，便可用与各标准刺激在感觉上成一定比例的相应的物理量值制成一个感觉比例量表。

下面以史蒂文斯和戴维斯 (Stevens & Davis, 1936) 的听觉响度量表为例。史蒂文斯采用半分法 (即分段法的一种)，给被试者一个响度的音，作为标准刺激，让他调节另一个音直到他感觉到比标准刺激的音低一半。用这个方法，以不同的强度音作为标准刺激，让被试调节另一个音直到他感觉到比标准刺激的音低一半。

为了制定一个 (sone, 响度单位) 量表，必须先确定 的单位量。1936 年，史蒂文斯确定一个 等于一个在绝对阈限以上 40dB 的 1000 赫的音。当一个 被判断为 47dB 强度音的一半响度时，我们就可肯定 47dB 的响度就是 2 个 。同样，我们可以得出 55dB 的响度为 4 个 ，一个声音的响度被被试者判断为 N 倍，则这个声音的响度就是 N ，以此类推即可制成一个 的 量表。

史蒂文斯用这种方法编制响度量表，发现在高强度水平时， 单位增加得特别快。例如：40dB 只有一个 ，但到 80dB 时就有 25 个 ，而 100dB 就有 80 的响度了。

用半分法所得响度量表与费希纳的感觉和刺激对数关系的直线图是不符合的。在图 4-12 中，实线代表被判断的响度的增加量，它是以分贝 (decibel, 简称 dB) 为单位的刺激强度函数；虚线代表按费希纳定律所预期的响度变化，因为纵坐标是直线的，基线是对数的，这个函数就必然是一条直线。

(二) 数量估计法

数量估计法 (method of magnitude estimation) 也是制作比例量表的一种直接方法。此法的具体步骤是主试者先呈现一个标准刺激，例如，一个重量，并赋予标准刺激一个主观值，例如为 10，然后让被试者以这个主观值为标准，把其他不同强度比较刺激的主观值，放在这个标准刺激的主观值的关系中进行判断，并用数字表示出来。然后计算出每组被试者对每个比较刺激量估计的几何平均数或中数，再以刺激值为横坐标，感觉值为纵坐标，即可制成感觉比例量表。

在心理量和物理量关系的实验中，常会出现特别大的数字，所以数量估计法采用的数据处理通常是几何平均值。几何平均值定义为 n 个数值相乘之积的 n 次方根。例如，有三个数据：4、8、16，则几何平均数 X 为：

$$X = \sqrt[3]{4 \times 8 \times 16} = 8$$

也可用下列方法计算：

$$X = (\lg 4 + \lg 8 + \lg 16) \div 3 \text{ 的反对数} = 8$$

贝克和杜德克 (Baker & Dudek, 1957) 的实验可进一步说明数量估计法。贝克和杜德克用假定长度单位 l 作为标准刺激，用其他 10 种比它较长的刺激作为变异刺激。刺激用幻灯片呈现，时间为 30 秒；每对刺激共随机呈现 20 次，即要求被试者回答 20 次；为排除空间位置的误差，其中 10 次为标准刺激在变异刺激之上，10 次为标准刺激在变异刺激之下。图 4-13 为 49 个被试

者对 10 个变异刺激数量估计的结果。在图上，实线为理论值，虚线为实验的实际结果。实验显示了心理长度的增长明显低于物理长度的增长。斜率的夹角小于 45° 。

(采自 Baker & Dudek , 1957)

四、三种量表法的比较

为了便于比较上述三种量表，我们将它们的基本特征、基本制作方法、功能、统计方法和典型经验列成表 4-13。

世界上很少有事物与数字是对应得十全十美或天衣无缝的，就是在物理学中也是如此。基于此，我们不可过分追求绝对的测量，但也不能把归属于某种量表的资料当作另一种量表，进而运用适合后者的统计方法；也不能极其轻率地把顺序量表资料当作等距量表，运用适合等距量表的统计方法。为了使这些量表达达到更好的效果，心理学家常常用许多变通的办法。例如，为了使顺序量表升格成等距量表，最通常的方法就是把原始测验分数转换成

特点 项目	量表 水平	顺序	等距	比例
基本特征的描述		等级、位次、大于或小于	量表上单位相等	有绝对的零点
基本经验运算		确定大于或小于的关系	确定距离或差别相等	确定比例相等
数学组成结构		单调组， $X=f(x)$ 其中 $f(x)$ 表示任意单调上升函数	线性成近似组 $X=ax+ba > 0$	相似性组 $X=cxc > 0$
功能		能分类，有等级	能分类，有等级，还能比较差异	能分类，有等级，能比较差异，还能比较比值
统计方法		中位数，百分位数，等级相关，秩次检验	平均数，标准差，积距相关，t 检验与 F 检验	除上述外，还有几何均数，相对差异数，等比量数的测定
典型经验		矿物硬度、皮革、木材、羊毛的等级，智力测验的原始分数	温度、日历时间，智力测验标准分数	长度、重量、密度等音高量表 ()、响度量表 ()

(采自戴忠恒，1987) 标准分数，而标准分数作单位显然是一个等距量表。这里看到了量表之间可以转换的一面。

由此可见，各种量表之间，既有联系又有区别，它们也有各自的长处和短处，在进行研究工作时，可视情况选择适当的心理量表方法。

本章实验

一、用极限法测定几种频率的听觉阈限

(一) 目的：本实验的目的在于熟习极限法的应用，并初步理解纯音听觉阈限与不同频率的关系。

声音响度与声波的振幅相对应，音高则是与声波的频率相对应的。但是

这种对应关系并不是简单的直线性的。对不同频率的纯音进行听觉阈限的测定，可以揭露这种对应关系，而且也是一切与听觉有关的研究的基础工作之一。同时，这种工作对于通讯器材的设计、医用测听器的校准和聋症的诊断等有很大的助益。

作为心理物理学方法之一的极限法，是测定阈限的直接方法，它能形象地表明阈限这一概念。也就是说，在记录纸上可以直接看出这一类与哪一类（例如感觉得到和感觉不到）反应的界限。极限法一般交替地使用递增和递减系列，这样既能抵消习惯误差，又能抵消期待误差。

（二）仪器：音频信号发生器（附衰减器）、刺激键、耳机、信号灯、信号键、反应灯、反应键。声音刺激由音频信号发生器供给，音频信号发生器上附有衰减器，可以直接读出增减的分贝数。接通刺激键，声音电流输入耳机，也就是给出了刺激。本实验也可用听觉实验仪或听力计代替，而且使用起来更方便、简单。

（三）程序：试验之前让被试者面对仪器坐下，戴上耳机，注意信号灯的指示，此灯一亮就是即将通过耳机给出声音刺激的预备信号。刺激的呈现有时从强到弱，有时从弱到强，听到声音时就闭合反应键，听不到声音时就放开反应键。

正式测听时随机使用 250，500，1000，2000，4000，8000（周/秒）六种频率。每种频率都交替进行递增、递减各 4 个系列的测试。递减系列从远超过听觉阈限的声音强度开始，每次衰减 5 分贝，直到被试者听不见时，记下这时的衰减数值。递增系列从远在阈限以下的声音强度开始，每次增加 5 分贝，直到被试者刚刚听到声音时，记下这时的衰减数值。不论递增或递减系列，主试者必须随机改变相继系列的开始点。每次刺激延续 2~6 秒，两次刺激间间歇 2~4 秒。开始时较快，到接近阈限时，刺激时间和间歇时间要约略延长。正式开始实验之前，可对递增、递减系列各做几次练习，借此使被试者熟习测听的步骤。

表 4-14 响度绝对阈限实验记录表

频率	250 ~				500 ~				1000 ~				2000 ~				4000 ~				8000 ~			
顺序																								
各次阈 限值																								
平均数																								

为了避免被试者疲劳，可组成主试者、被试者各两组，每完成两种频率测听，主试者、被试者轮换一次。主试者二人，一人改变声音频率和强度，一人给预备信号、发出刺激并记录被试者反应。主试者应事先准备好随机使用的频率和递增、递减系列开始点的安排，列入表格，以便据以进行实验，并作记录。

实验完成后问被试者：你认为各次判断的标准是否一致？可信度如何？

（四）结果：计算每次测定的不同频率的纯音听觉阈限，用平均数作为听觉阈限，用标准差表示原始测定数据的集中趋势。计算后将声级（以微巴为基准的分贝数）为纵坐标，以频率（周/秒）为横坐标，画出被试者六种不

同频率声音刺激的纯音听觉阈限。

(五) 讨论：

(1) 在本实验中，为什么要随机改变相继系列的开始点？

(2) 在本实验中，被试者的练习和疲劳对实验结果有无影响？还有什么主客观因素影响了实验的结果？

(3) 极限法用于本实验中听觉阈限的测定还有什么缺点？如何改进？

二、用平均差误法测定长度差别阈限

(一) 目的：通过测长度差别阈限学习平均差误法。

(二) 材料：长度估计测量器。

(三) 程序：

1. 实验的基本方法

用长度估计测量器呈现白背景、黑线条，分左右两半。右半、左半分别用两个活动套子套住，背面有以毫米为单位的刻度（见第十一章仪器图）。主试者移动一个套子，使这边的直线露出 10 厘米作为标准刺激，又用同样的办法使另一边的直线露出一段明显短于或长于标准刺激的长度作为变异刺激，被试者借助于移动套子调节变异刺激，直到他认为与标准刺激长度相等为止。主试记下被试者调好的长度。

2. 在安排实验顺序时，要注意几个控制变量

(1) 为了消除动作误差，在全部实验中应有一半的次数呈现的变异刺激长于标准刺激（套子向内移动，简称“内”）；另一半的次数呈现的变异刺激短于标准刺激（套子向外移动，简称“外”）。

(2) 为了消除空间误差，在全部实验中有一半的次数呈现的变异刺激应在标准刺激的左边，即于中线的左边（简称“左”）；另一半的次数呈现的变异刺激应在标准刺激的右边（简称“右”）。这样，变异刺激的呈现方式可以有“左外”、“左内”、“右外”和“右内”四种方式。

(3) 为了消除系列顺序的影响，实验可按这样的顺序进行：“右外”、“右内”、“左内”、“左外”、“左外”、“左内”、“右内”、“右外”。每种条件需做 5 次，共做两个循环，合计 80 次。每做完 20 次，休息 2 分钟。

3. 在实验过程中，主试者不要告诉被试者调整出的变异刺激的长度是否和标准刺激相符，也不要有任何有关的暗示。4. 换被试者，再按上述程序进行实验。（四）结果：（1）分别计算两个被试者的长度估计的平均误（AE）。

$$AE = \frac{\sum |X - S|}{N}$$

X：每次测定所得数据，S：标准刺激的长度，N：测定的总次数。

(2) 考验两个被试者的平均误的差别有无显著性意义。

(五) 讨论：

(1) 比较两个被试者的长度差别阈限。

(2) 本实验在排除误差上作了哪些考虑？

表 4-15 长度差别阈限实验记录表单位：（毫米）

次数	右				左			
	外		内		外		内	
	1	2	1	2	1	2	1	2
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
平均数								

三、用恒定刺激法测定重量差别阈限

(一) 目的：通过测定重量差别阈限学习恒定刺激法。

(二) 材料：

(1) 高 2cm，直径 4cm 的圆柱体两套 8 个。这二套圆柱体的重量要求如下：

第一套：100 克两个， 100 ± 3 克、 100 ± 6 克、 100 ± 9 克各一个。

第二套：100 克两个， 100 ± 4 克、 100 ± 8 克、 100 ± 12 克各一个。

(2) 遮眼罩。

(三) 程序：

(1) 取第二套材料中重 100 克的圆柱体作为标准刺激。重 88 克、92 克、96 克、100 克、104 克、108 克和 112 克的圆柱体作为变异刺激，让被试把变异刺激和标准刺激各比较 20 次，共 140 次，每对刺激呈现的顺序按随机原则排列。为消除时间误差，前 70 次先给标准刺激，后 70 次先给变异刺激。

(2) 被试者坐在桌边，戴上遮眼罩，主试者把标准刺激放在桌边约 20 厘米处正对被试者的地方，并用粉笔标出圆柱体在桌面上的位置；被试者用右手的拇指和食指拿住圆柱体慢慢向上举，轻轻上下掂两下，约两秒钟后就放下（这时胳膊肘不要离开桌面）。

(3) 在进行前 70 次比较时，主试者先呈现标准刺激（即 100 克）后呈现变异刺激，被试者用单手先提举后报告第二个重量比第一个重量轻、重、还是相等，主试者分别用“-”、“+”和“=”三种符号进行记录。后 70 次比较时，因先呈现变异刺激，故记录时要注意选准三种符号。主试者呈现重量时要保持同一位置，第二个重量呈现的时间不可太迟，当被试者放下第一个重量后，主试者立刻换上第二个重量，间隔最好不要超过 1 秒，以免被试对第一个重量的感觉消退。两次比较至少要间隔 5 秒，以免各次感觉相互干扰。

(4) 分别计算在 20 次比较中被试者判断各变异刺激比标准刺激轻、重和相等的次数。

(5) 更换被试者，按上述程序再进行新一轮实验。(四) 结果：

(1) 分别统计每个变异刺激较标准刺激轻、重和相等的次数，并列出的百分数，制成表格。(2) 以变异刺激的重量为横座标，以反应各变异刺激重于、轻于和等于标准刺激的次数的百分数为纵座标，把所得结果画成三条曲线。(3) 用直线内插法分别根据图表求出重量差别阈限(DL)。

(五) 讨论：

(1) 用恒定刺激法测定差别阈限有什么特点？(2) 实验结果是否符合韦伯定律？为什么？

表 4-16 重量判断记录表 (标准刺激为 100 克)

变异刺激 先后 顺序	88 克		92 克		96 克		100 克		104 克		108 克		112 克	
	1	2	1	2	1	2	1	2	1	2	1	2	1	2
1														
2														
3														
4														
5														
6														
7														
8														
9														
10														
判断的%	重													
	轻													
	等													

本章摘要

1. 心理物理法是研究心理量与物理量之间对应关系的方法，它是与费希纳的名字紧密相联的，从诞生到现在已有一百多年的历史。

2. 在心理学上，对一个心理现象根据测定它们的程序下的定义就叫操作定义。例如，把“刚刚感受到”定义为“50%次感觉到”，就可测定感觉阈限了。

3. 传统心理物理学所处理的问题大体分为两类，一类是感觉阈限的测量，包括绝对阈限、差别阈限和等值的测量；另一类是阈上感觉的测量，即心理量表的制作。

4. 测量感觉阈限的基本方法有：(1) 极限法；(2) 平均误差法；(3) 恒定刺激法。

5. 极限法是测定阈限的直接方法，其特点是：将刺激按递增或递减的方式沿着一定的维，以间隔相等的小步变化，寻求从一种反应到另一种反应的过渡点。

6. 平均误差法的基本特征是：呈现一个标准刺激，令被试者再造或复制它的大小；或要求他调整一个变异刺激，使它和标准刺激相等。

7. 恒定刺激法的基本特点是：根据次数来定阈限，即以次数的整个分布求阈限。

8. 从量表有无相等单位和有无绝对零点来分，心理量表可分为顺序量表、等距量表和比例量表这三类。

9. 顺序量表是一种较粗略的量表，它既无相等的单位，又无绝对零点，只是把事物按某种标准排出一个次序。一般用等级排列法和对偶比较法来制作顺序量表。

10. 等距量表是一种有相等单位，但没有绝对零点的量表。一般用感觉等距法和差别阈限法来制作等距量表。

11. 比例量表既有真正的零点，也是等距的，是科学家们认为理想的量表。比例量表的制作方法主要有分段法和数量估计法。

建议参考资料

1. 杨治良（1988）：心理物理学。兰州市：甘肃人民出版社。
2. 陈立、汪安圣（1965）：色、形爱好的差异。心理学报，第十卷，三期，265~269页。
3. 郑昭明（1993）：认知心理学。台北市：桂冠图书有限公司。
4. 波林（高觉敷译，1981）：实验心理学史。北京市：商务印书馆。
5. 赫葆源、张厚粲、陈舒永（1983）：实验心理学。北京市：北京大学出版社。
6. Atkinson, R.C. et al. (1988). *Steven's handbook of experimental psychology* (2nd ed.). New York: A Wiley Interscience Publication.
7. Christensen, L.B. (1991). *Experimental methodology* (5th ed.). Needham Heights: Allyn and Bacon.
8. Underwood, B.J. (1966). *Experimental psychology*. New York: Appleton.

第五章 现代心理物理法

本章内容细目

第一节 阈限概念和理论的发展

一、阈限概念及其发展 211

(一) 传统的阈限概念

(二) 对传统阈限概念的异议

二、神经量子理论 216

三、史蒂文斯的幂定律 217

四、物理相关论 225

第二节 信号检测论的原理

一、信号检测论的由来 227

二、信号检测论的统计学原理 229

三、信号检测论的二个独立指标 233

(一) 反应偏向

(二) 敏感性指标

四、接受者操作特性曲线 243

第三节 信号检测论的应用

一、在医学心理学中的应用 250

二、在工程心理学中的应用 254

三、在认知研究中的应用 259

四、评价 260

本章实验

一、信号检测论用于记忆的实验 263

二、接受者操作特性曲线的实验 265

本章摘要

建议参考资料

上一章我们介绍了传统心理物理学的三种基本方法。传统心理物理学方法用极限法、恒定刺激法和调整法来测定绝对阈限和差别阈限，并以阈限的倒数来表示感受性的大小，阈限值越小，感受性越高。用传统心理物理学方法测定阈限时，常有一些非感受的因素对阈限的估计产生影响。因此，传统心理物理法测得的数据，往往是感受性和被试反应的主观因素相混合的。例如痛阈，因各人的主观因素不同，痛阈因人而异。所以，传统心理物理法的科学性和可靠性就受到一定的影响。为了有效地测定感受性，传统心理物理学中就想设法通过各种手段来消除如动机、态度等因素的影响。正是由于采用种种手段，使传统心理物理法在感受性测量中能得到较科学和较可靠的结果。但上述这种种手段也不能从根本上排除被试的主观态度的一类因素的影响。所以我们常发现，用传统方法获得的结果，有时甚至是相互对立的。例如，精神分裂症患者和正常人关于大小常性的对照研究：伯雷斯等人 (Pezer et al., 1961) 认为精神分裂症患者比正常人的大小常性要高；汉米尔顿等人 (Hamilton et al., 1963) 则认为精神分裂症患者比正常人的大小常性要低；而皮什克姆-史密斯等人 (Pishkm-Smith et al., 1962) 的看法认为他们彼此并没有差别。这些情况，以前总认为是实验条件和方法的不同造成的，实际上，最根本的问题是无法把感受性的测量和被试的动机、态度等主

观因素所造成的反应偏向区分开。而信号检测论，在测定感受性上，能把被试的主观态度区分出来。所以，我们把费希纳倡导的心理物理学三种基本方法称之为传统心理物理学，而把信号检测论这一心理物理学的新发展称为现代心理物理学。

但是必须注意，不能因此否定传统心理物理学在测定感觉阈限方面的作用，一般情况下测定感受性，传统心理物理学是可用的。而信号检测论的实验更宜用于人的主观因素对实验结果影响较大的实验条件中，它解决了传统心理物理学所不能解决的问题。本章内容所要讨论的是，希望能回答下面的问题：

1. 如何理解感觉阈限这个概念。
2. 如何解释史蒂文斯的幂定律。
3. 信号检测论的统计学原理是什么。
4. 信号检测论的二个独立指标是什么，如何进行计算，如何绘制接受者操作特性曲线（ROC）。
5. 信号检测论在心理学中有哪些主要应用领域。
6. 如何评价信号检测论。

第一节 阈限概念和理论的发展

一、阈限概念及其发展

（一）传统的阈限概念

传统的阈限概念起源于费希纳。波林在《实验心理学史》一书中指出，心理物理学的古典问题有五个：（1）绝对阈限：观察者对个别刺激的感受；（2）差别阈限：观察者对刺激增量的感受性；（3）等量：被判断为相等的刺激，一般指主观判断的强度方面；（4）感觉距离：被判断为相等的两对刺激间的差别；（5）感觉比例：彼此判断为有特定比率的那些刺激。在有影响的“数学、测量和心理物理学”一章中又提到：史蒂文斯（Stevens, 1951）又补充了两个问题。（1）刺激次序：观察者将某些组的刺激排成等级或次序的测定；（2）刺激等级评定：确定观察者评定刺激的真正物理值的准确性。由此可见，一百多年前，费希纳在创建心理物理学时，就把注意力集中在感觉阈限的测量上。经过长期来的研究，形成了许多阈限理论，包括柏拉克韦尔（Blackwell, 1953）的高阈限理论，路司（Luce, 1954）的低阈限理论，格林（Green）的高-低两种阈限理论，以及史蒂文斯的神经量子理论和斯韦茨等人（Swets et al., 1953）的信号检测理论。

从测量上考虑传统阈限的概念，后经卡特尔（Cattell, 1893）、贾斯特罗（Jastrow, 1888）和乌尔班（Urban, 1910）发展，最后形成了费-伽马（Phi-gamma）假设的传统觉察理论基础。这种觉察论假设有三个连续量：刺激（stimulus, 简称S）、内部反应（response, 简称R）和判断（judgment, 简称J）。从图5-1中可见，刺激 S_2 被定义为绝对阈限（T），因为它是引出“有”反应的一半次数。

图5-1上的三种连续量是相互联系的。假设实验与实验间刺激的物理变量是固定不变的，而内部反应则被认为是可变的，每个恒定的物理刺激所引起的反应可看作是常态分配。这样刺激的连续将引起一系列重叠的分布。在这些重叠的分布中，阈限T就是一个固定点（参见中间一条线）。通常，被

试判断的连续是被假设准确地对应于它的反应连续，当刺激强度超过阈限 T 时，被试判断有刺激出现，低于阈限 T 时，则判断无信号出现（参见上面一条线）。

根据阈限的传统定义，刺激 S_2 是这个被试的绝对阈限。由于反应分布是常态的，所以阈值对应于 S_2 所引起的反应平均值。常态分布的对称性，使得刺激 S_2 所引起的反应有一半次数是超过阈限。一半次数不超过阈限。 S_1 有时可以超过； S_3 则一般会超过； S_4 超过的概率就更高了。

（二）对传统阈限概念的异议

起源于费希纳的传统阈限概念，在 20 世纪 50 年代前，没有受到多大挑战。可是，近四十年来，关于阈限概念的理解上已成为古典的和现代的心理物理学争论的焦点。50 年代以后，许多心理物理学家愈来愈感到，经典的阈限测量最大的问题在于没有能够把被试的辨别能力（感受性）和他们做出判断时的倾向性（反应标准）区别开来。也就是说，经典的阈限测量没有考虑许多非感觉变量对被试的影响。在这些非感觉变量中，主要有下列二种最为主要：

1. 刺激出现的概率 传统的心理物理学实验中，刺激出现的概率（或机率）（probability）以 $P(S)$ 表示，其值总是 1.0，这样被试做出“有”或“无”的判断时，即使凭机遇也有 50% 的精确性，难以避免地存在着期望误差。实验证明，当 $P(S)$ 系统变化时，做出“有”判断的概率以 $P(\text{yes})$ 表示，也会发生变化，测出的阈限值亦发生变化。一个察觉指端震动的实验（Gescheider et al., 1971）表明，改变 $P(S)$ 对于传统的阈限测量具有强烈的影响。当 $P(S)$ 等于 0.7 时，阈限是 1.3 微米； $P(S)$ 为 0.3 时，阈限则是 2.3 微米。而且，当刺激不出现时， $P(\text{yes})$ 也明显地高于零。而根据传统的阈限理论，在不呈现刺激时，训练有素的、忠实可靠的被试者只是偶然地报告“有”感觉， $P(\text{yes})$ 不可能明显地高于零。

2. 反应代价 反应代价影响反应的结果，即反应受到奖赏和惩罚的影响。传统的阈限理论是对此忽略不顾的。事实上，如果正确地报告了“有”刺激（称为“击中”）得到的奖赏多，而且错报时受到的惩罚少，则被试倾向于降低反应标准，在这种情况下，所测量的阈限相对值便低。反之，如果击中时所得奖赏少，而错报时所受惩罚多，则被试者倾向于提高反应标准，不会轻易做出“有”反应，这时，所测量的阈限值便高。这一显而易见的事实，传统的阈限理论恰不能给予解释。

从上述二种非感觉变量的讨论中看到，现代心理物理学从对阈限概念的不同理解出发，把被试者的辨别力和他的反应倾向区分开处理，而传统心理物理学测得的感觉阈限是二者不可分割的混合物。这是现代心理物理学对阈限概念和感觉理论的新发展。

现代心理物理学不仅对心理研究方法带来新的突破，而且也对心理学理论带来了巨大影响。这些变化不仅明显地表现在对阈限概念的理解上，而且也反应在心理物理学所研究的课题上。对研究的课题，现代心理物理学从系

表 5-1 心理物理学问题的分类、选用方法和尺度准绳

类别	研究问题	选用方法	尺度准绳
感觉能力机 体感受性和 可辨性的测 量)	绝对感受 性：辨别某 一刺激存在 与否的能力	1.有 m 个选择的 强迫选择法(空间 或时间) 2.可做 双重决定(是或 否)的单一刺激法 或 对可信度的评 价法	正确百分数 (d)或其 他觉察指标 P (C) Max
	差别感受 性：在一些 刺激之间进 行辨别的能力	1.对偶比较法 2. 评价法或对各种 刺激 AA , AB 或 BB 的相似性或相 异性的双重决定 法 3.在 AB 序列中 识别第三个刺激 的方法	正确百分数 (d)或其 他辨别指标
	刺激识别： 从一组几个 刺激中，对 每个刺激给 予不同反应 的能力	识别法	1.常误 2.标 准错误或其 他离散尺度
	有效刺激 量：这是来 自感觉能 力测量单 位进行标 度的刺激 量	通过等距单位测 量有效刺激强度 的方法	在指定的判 断百分数上， 称为相等的 刺激值(范 围)
反应倾向 (刺激间及 刺激和周围 背景间可清 除辨别的情 况下，对一 些刺激或刺 激序列特有 反应的测 量)	等距：与刺 激间辨别能 力测定不相 同，在指导 语规定的 连续体上称 为相等刺 激的测定	调整法，允许做 “相等”判断的 对偶比较法，恒 定刺激法，极 限法	在指定的判 断百分数上， 称为相等的 刺激平均值 和离散值
	等差：与差 别之间的辨 别	调整法，允许做 “相等”	在指定的判 断百分数
	能力测定 不相同，在	判断的对偶比 较法，恒	上，称为相 等的刺激

(采自马谋超，1978)

统论出发，将一个事物看成是有层次阶段和多水平的结构，因此，它所面临

要解决的问题也是有层次和不同水平的事件。具体地说，现代心理物理学课题可分割为检测、认知、分辨、量表等四个基本问题，叙述如下：

1. 检测问题 检测（或觉察）（detection）就是察觉一个事物是否存在的问题。这是一个低水平、低层次的感知觉问题，它所要解决的问题是，人检测到了信号，报告为“有”，没有检测到信号，报告为“无”的简单问题。在使用方法时，如果研究者承认“阈限”的存在，就可用传统的极限法和恒定刺激法来测定阈限；如果不承认“阈限”的存在，就可用信号检测论来分析检测信号的辨别力和反应倾向。

2. 认知问题 认知（cognitive）就是辨认一个事物的问题。这里已包括较高水平的问题，如当一个雷达操作员在屏幕上观察到了有信号以后，就要进一步向自己提出这样的问题：“这是什么？”。这就包括了复杂的知觉过程。这里包括可以采用信息论的方法来认知对象。应当指出，检测和认知阶段有时是不可分的，相互联结的，所以有时也称为检测-认知阶段。虽然认知问题既重要又有趣，涉及到信息量的问题，但这方面的问题一般实验心理学和心理物理学中不作详细论述，主要在认知心理学中进行阐述和讨论。

3. 辨别差异问题 辨别差异（discriminative difference）是叙述某些事物与标准的区别问题。这是一个更高的梯级。例如雷达操作员要正确无误地分辨出这一刺激不同于别的刺激。心理物理学中的韦伯定律（或韦氏定律）（Weber's law），分辨中的信号检测论（见第二节），以及用反应时来测定分辨等，都属于辨别差异问题。

4. 量表问题 量表（scale）这方面的研究方法和内容，已构成了一个专门的学科——心理测量学（psychometrics）。测量心理活动并转换成某种尺度（yardstick）的问题。

通过以上的讨论，我们就可将问题进行新的归类，把讨论心理物理学的一些问题组织到那些感觉和反应倾向的范畴内。参见表 5-1。

二、神经量子理论

神经量子理论（neural quantum theory）是由史蒂文斯等人（Stevens et al., 1941）提出用来解释阈限的一种理论。他们在响度和音高的辨别实验中，推论其基本神经过程是按全或无定律（或全有全无律）（all-or-none law）进行的。神经量子理论假设反应刺激变化过程的神经结构在机能上被分为各个单元或量子。具体地说，被试者只有在此增量大到足以兴奋一个附加的神经量子单位时，才能够察觉到刺激增量。必要的刺激增量的大小将取决于某一个刺激高于上一个兴奋了的神经单位的阈限多少。超过上一单位的阈限越多，兴奋下一个单位所需要的刺激量则越少（见图 5-2）。图 5-2 表明两种连续：（1）刺激连续（stimulus continuum）（任意度标）。（2）神经单元阶梯式的感觉连续（sensory continuum）。在刺激连续上， S_t 是标准刺激值； a 是肯定够兴奋附加量子的刺激增量； p 只能部分兴奋神经量子所需要的刺激增量。

在感觉连续上， P 是 S 所产生的“剩余”兴奋量，如果说假定 17 个能量单位的刺激量足以兴奋神经量子 a 、 b 、 c 、 d ，而且还能部分兴奋“ d ”量子，这个剩余量只是由超出 20 能量单位的那点能量所引起的。由此可见，剩余量和感受性的波动紧密相关。只有 P 和 S 剩余量总合达到等于和大于兴奋一个附加量子所需的能量时，才能产生可觉的感觉反应。因此剩余和能量是有关的，即剩余大，要求增量便小；反之亦然。用数学式表述如下：

$=Q-P$ [公式 5-1]

：使附加量子活动所需要的刺激增量

Q：肯定能够兴奋一个量子的增量的大小

P：标准刺激 S 剩余能量引出的部分兴奋量

上式表明当 $Q-P$ 时，给定的 S 就完全可以兴奋附加量子。增量愈大，辨认的数量愈增加。诚然，这也取决于不同剩余量出现的相对频率。

继续加强增量，务必达到兴奋一个附加量子才会有一个最小可觉差。这样所得到的理论上的刺激反应间关系曲线应是梯形跳跃式的。但是，在实际测量阈限的实验研究中，所得到的总是一条递加的拱形曲线（ogive curve）。原因在哪里？史蒂文斯认为在于缺少对于被试的动机、注意疲劳等这些随机的波动因素的全面控制。未控制因素的波动可能是常态分配，因而是拱形曲线。史蒂文斯指出，如果满足下面四个条件，便能够得到理想的梯形曲线。

1. 必须仔细地控制刺激，避免噪音干扰。

2. 被试者在整个实验过程中必须保持恒定的判断标准，最好由动机高度明确、训练有素的实验者充当被试者。

3. 如果神经量子的大小在实验期间改变了，那么曲线将变为拱形。

4. 从标准刺激向比较刺激的转换必须迅速。

三、史蒂文斯的幂定律

费希纳（Fechner, 1860）从韦伯定律 $I/I=K$ 出发，提出二个假设：（1）每一个最小可觉差（jnd）可看作感觉上的一个最小变化，它的主观量是相等的，是感觉的单位；（2）任何阈上感觉的大小都可用感觉随机刺激变化而发生的总和表示，亦即可用最小可觉差作为感觉单位，对阈上感觉量进行间接测量。

那么可用下列公式表示函数关系：

$$S = K \lg R$$

S：感觉大小（以 jnd 为单位）

R：刺激强度

K：常数

这就是费希纳定律（或费氏定律）（Fechner's Law），也可称为对数定律（law of logarithmic function）。

于 20 世纪中叶，史蒂文斯对费希纳的对数定律进行了批评。他说费希纳是把分辨能力作为测量单位的，根据这种间接测量的哲学，可以证明同样经常察觉到的差别并不相等。他还指出，一百年来之所以没有打破费希纳的对数定律，是因为批评家们只罗列了一些相反的事实，而没有提出一种理论去取代这种有缺陷的理论。于是，他于 1957 年根据多年的研究结果，提出了刺激强度和感觉量之间关系的幂定律（the power law）：

$$S = bI^a$$
 [公式 5-2]

S：物理量 I 的幂函数

b：量表单位决定的常数

a：感觉道和刺激强度决定的幂指数

幂函数的指数值决定着按此公式所画曲线的形状。例如，当指数值为 1.0，便是一条直线，即刺激和感觉之间为简单的正比关系；指数大于 1，则为正加速曲线；小于 1，便为负加速曲线。

史蒂文斯认为存在着两种感觉连续体，即量的连续体和质的连续体，幂定律函数关系适用于量的连续体（prothetic continuum），这是一些对它们做出多少的判断的连续体。在这个连续体上的感觉的变化，是以刺激引起的神经兴奋多少为依据的。例如重量、响度、亮度等形成的连续体，都是量的连续体。史蒂文斯还认为，其他函数则可能来自与那种或那种的判断有关的质的连续体（metathetic continuum）。在这个连续体上感觉的变化，是以刺激引起的神经活动的部位为依据的。也可以说，在生理上它是以一个相互代替的机制为依据的。例如声音的音高、彩色的色调等形成的连续体，都是质的连续体。对质的连续体来说，阈值是个不变的截面或称分界点。

史蒂文斯用数量估计法（method of magnitude estimation）获得了大量的实验数据。数量估计法是制作感觉比例量表的一种直接方法。具体的步骤是实验者先呈现一个标准刺激，例如一个重量（或某一明度），并规定它的出现值为一个数字，例如 1.00，然后让被试以这个主观值为标准，把其他同类强度不同的主观值，放在这个标准刺激的主观值的关系中进行判断，并用一个数字表示出来。表 5-2 就是三种感觉道所获得的实验结果。表 5-2 是 22 种感觉道（连续体）的幂函数情况。由表可见，史蒂文斯对多种感觉道作了研究，并求出它们的指数，发现在同一感觉道内，其指数是相同的。

表 5 - 2 上的实验结果，以物理量为横坐标，以心理量为纵坐标，就可绘成图 5 - 3。如果把这三个感觉道的实验结果画在双对数坐标上，就形成了三条斜率不同的直线，如图 5 - 4 所示。我们从图 5 - 3 和图 5-4 上可看到，电击的感觉强度比产生出的电击的物理强度增长快得多（ $a=3.5$ ），明度比光能的增长却慢得多（ $a=0.34$ ），线段的主观长度和线段的物理长度则有同样的增长率（ $a=1$ ）。

表 5-2 三种感觉道的心理强度

物理量	心理量		
	明度	长度	电击
1	1.00	1.00	1.0
2	1.26	2.14	11.3
3	1.44	3.35	46.8
4	1.59	4.60	128
5	1.71	5.87	280
6	1.82	7.18	529
7	1.91	8.50	908
8	2.00	9.85	1450
9	2.08	11.2	2190
10	2.15	12.6	3160

（采自 Stevens, 1961a）

和对数定律一样，幂函数对于靠近阈限的微弱刺激就变得很不明确了。于是，史蒂文斯等人在 60 年代初又提出了修正的幂函数，即从刺激中减去一个常数：

$$S=b(1-I_0)^a \quad \text{〔公式 5-3〕}$$

这样，幂定律便可适用于全部可知觉的刺激范围。在某些研究者看来 I_0 就是绝对阈限值。从 I 中减去 I_0 ，意味着以阈限上有效单位而不是以物

表 5-3 在量的连续体上心理量和刺激量的幂函数关系

连续体	指数	刺激条件
响度	0.6	双耳
响度	0.54	单耳
明度（指自我发光体）	0.33	5° 靶面——暗适应眼
明度（指自我发光体）	0.5	点光源——暗适应眼
明度（指表面色）	1.2	灰纸反射率
嗅觉	0.55	咖啡气味
嗅觉	0.6	庚烷
味觉	0.8	糖精
味觉	1.3	蔗糖
味觉	1.3	盐
温度觉	1.0	冷——臂上
温度觉	1.6	暖——臂上
振动	0.95	60 赫——手指上
振动	0.6	250 赫——手指上
持续	1.1	白噪声刺激
重复率	1.0	光、声、触、电击
手指跨度	1.3	积木厚度
手心压力	1.1	皮肤上的静态力
重量	1.45	举重
握力	1.7	精确的手握力计
自动语音水平	1.1	发声的声压
电击	3.5	60 赫——通过手指

理表上的零点以上单位去说明刺激。

幂定律在对数定律的基础上前进了一大步。但是，幂定律的有效性有赖于被试正确使用数字去表示其真正的感觉量。与此同时，表 5-3 上列出的量是不同感觉通道的主观量。那么，这里要问：不同感觉通道之间的主观量能否比较，能否调节一个感觉通道中的刺激强度使其主观上感到好像同另一感觉通道中的刺激一样强？为克服这一局限，史蒂文斯于 1959 年研究了跨感觉道的匹配技术，它无需被试产生数字判断，被试的任务是把两个不同感觉道产生的感觉量相等起来。例如，可以要求被试调整施加于指端的震动强度，以便使震动的感觉印象和一爆破噪音的响度相匹配。这样，在不同的刺激水平上获得跨感觉道的匹配，一条称为等感觉函数的曲线便产生了，它表示出一感觉道的刺激值与造成相等感觉量的判断的另一感觉道刺激的关系。这种方法称为等感觉匹配法（equal-sensation functions obtained by matches）。

史蒂文斯又将心理物理法技术推进一步，用实验证明了不同感觉通道的感觉量是可以匹配的。其原理为：设有一个感觉道的主观值为：

$$S_1 = I_1^{m_1}$$

另一个感觉通道的主观值为：

$$S_2 = I_2^n$$

表 5-4 不同连续体的匹配

比例量表		以握力为量表		
连续体	幂函数的指数	刺激范围	主试比例量表的指数	匹配的握力指数
电击 (60 赫)	3.5	0.29 ~ 0.72 毫安	2.06	2.13
温度刺激 (热)	1.6	高于自然温度 2.0 ~ 14.5	0.94	0.96
举重	1.45	28 ~ 480 克	0.85	0.79
手掌上的压力	1.1	0.5 ~ 5.0 磅	0.65	0.67
温度刺激 (冷)	1.0	低于自然温度 3.0 ~ 30.6	0.59	0.60
振动 (60 赫)	高出阈值 17 ~ 47dB	0.56	0.56	
白噪音响度	0.6	高出 0.0002 达因 / cm ² 59 ~ 95dB	0.35	0.41
1000 赫音乐响度	0.6	高出 0.0002 达因 / cm ² 47 ~ 85dB	0.35	0.35
白光亮度	0.33	高出 10 ⁻¹⁰ 朗伯值 59 ~ 96dB	0.20	0.21

(采自 Stevens , 1961a)

如果主观值 S_1 和 S_2 相等 , 则最后的相等感觉函数将有以下形式 :

$$I_1^m = I_2^n \quad [\text{公式 5-4}]$$

$$\lg I_1 = n/m \lg I_2$$

这样 , 在双对数坐标中相等感觉函数将是直线 , 而其斜率将由二个指数决定。图 5-5 为史蒂文斯用于匹配实验的握力计。表 5-4 为实验所得的结果。图 5-6 为根据表 5-4 材料绘制的匹配图。

史蒂文斯出色的研究工作 , 得到实验心理学家的充分肯定 , 各种版本的教科书争相引用。很多研究者认为 , 史蒂文斯幂定律对数量估计材料是一份有效的总结 , 指数定律说明了感觉传导者的操作特征。或者称为将刺激能量转换为神经能以及脑形成感觉的数学描述。不同感觉通道的指数不同 , 说明了不同感觉传导者是以能量的不同形式转换的 , 即具有不同的转换特征。其后 , 一些实验心理学家对幂定律作了补充解释。例如 , 艾克曼 (Ekman , 1964) 认为 , 分辨反应基本为边缘过程 , 其转换特征服从费希纳对数定律 , 而主观反应为指数相关。边缘反应变成主观反应只要乘上一个转换数 , 一般为反对数转换。

近些年来，大量实验说明，分辨反应与指数定律有一定的相关，一般说来，小韦伯分数（高分辨力）同大指数相联系，如重量和电击；而大韦伯分数同小指数相联系，如亮度和味觉（见表 5-5）。这里也使我们看到了韦伯定律和史蒂文斯定律的一致性。

总之，幂定律的重要性在于其相等的刺激比例产生相等的感觉比例这一含义。由此可以认为，如果所有的刺激强度都按百分比增加或是减少，那么感觉变化的比例则保持恒定。它与费希纳的对数定律不同的地方还在于对感觉直接测量，因而有心理学家称它为“新心理物理学”的开始，这里所说的“新心理物理学”，就是指现代心理物理学。

表 5-5 I/I 和 n 的关联表

连续体	I/I	n
亮度	0.079	0.33
响度	0.048	0.60
指痛	0.022	1.30
重量	0.020	1.45
长度	0.029	1.04
味觉 (NaCl)	0.083	0.41
电击	0.13	3.50

(采自 Stevens , 1961a)

四、物理相关论

在史蒂文斯幂定律得到充分肯定的时候，同时也受到了一些实验心理学家的挑战。其中较有影响的是物理相关论。物理相关论 (physical correlatetheory) 是由瓦伦 (Warren , 1959) 在 20 世纪 50 年代末至 60 年代初提出来的，它与史蒂文斯的心理物理定律是针锋相对的。这种理论认为，被试者作出感觉量的判断时，实际上是通过过去的经验对与刺激相联系的某种物理属性作出判断。由不同的感觉通道得到的各种变化的幂函数的指数并不反映不同的生物转换器的操作特点，相反是由被试者对于不同刺激的特定物理属性的反应所决定。例如，对于刺激持续时间的判断，史蒂文斯的心理物理幂函数的指数是 1.0，这只是指示出，经过多年体验，被试者已有能力去作出与刺激持续时间有线性关系的判断。用工具性条件反射的语言来说，就是经过多年的强化与消退，被试者已学会对一特定的持续时间的刺激作出正确的反应。因此，被试者在心理物理实验中的判断是由刺激的物理属性决定而不是由感觉的持续时间决定。得到测量的，是被试在各种持续时间的刺激中作出分辨的能力。根据这一理论，响度应由听者和声源之间的典型相互作用来确定。由于多数声源发出一个固定强度，我们对强度差别的大多数经验是和声源移近或移远有关。瓦伦指出，当我们走近声源两倍时，正常体验到的强度变化被报告为两倍响度；一半响度则是与距离加倍相联系的强度变化等等。瓦伦认为，观察者戴上耳机在实验室内作的响度判断不是作为正常环境中判断声音时的经验影响的充分根据。物理相关论就是在此基础上提了出来。

为了进一步阐述物理相关论，瓦伦曾用明度判断的实验结果去解释这一理论。许多对测量光的体验较少的人不能对光的强度作出精确的判断，但他

们却经常能对物体的明度作出判断。瓦伦认为个体熟悉当物体和照亮它的光源之间的距离改变时物体的形象 (appearance) 也变化的情形。这样,在要求被试者在实验情形中作出判断时,他的反应便依赖于这种熟悉的效果。在刺激强度减少量确切地相等于在一个物体和光源之间的距离加倍时的刺激强度减少量时,便应出现“一半亮”的判断。根据反平方律 (inverse square law),如果物体 A 离照射它的光源距离是物体 B 的两倍,那么物体 A 受到的照射是物体 B 的 25% 那么多。因此比例为 1/2 的知觉明度比例需要比例 1/4 的刺激能量,所以明度函数应是:

$$B=KI^{0.5} \quad \text{〔公式 5-5〕}$$

B: 判断明度

K: 常数

I: 光能量

瓦伦认为,史蒂文斯所得到的 0.5 以外的其他指数只是在很严格的和非自然的观察条件下获得的。例如,史蒂文斯提出的数目为 0.33 的明度指数,是通过暗适应了的眼睛,被试者观察一黑暗背景上一个发光圆点 (约 5°) 而得到,而瓦伦自己的实验得到的明度指数为 0.5,则是在自然条件下得到的——即大面积刺激,使用适应于这种刺激水平的眼睛。因此,瓦伦认为他的物理相关论是一个预测感觉量增长的系统根据,更符合客观情形。根据上面的讨论和分析,对于心理物理函数服从幂定律还是把它看作对两个刺激变量之间的关系的描述,似乎现在还不宜作出一个绝对肯定或绝对否定的回答,它服从哪个定律常因具体条件不同而各异。史蒂文斯提出的质和量的两类连续体,也可能是区别心理物理函数服从哪个定律的一个条件,虽然现在看来还不是一个十分明确的条件。我们相信在进一步研究的基础上,将会有新的、可以概括更多事实的心理物理定律和理论被不断提出来。今天,我们在评价史蒂文斯的幂定律时,正是由于它在对感觉的直接测量基础上提出新的理论和定律,因此,不少心理学家把史蒂文斯定律以及围绕阈限问题所展开的争论看作是现代心理物理学的开始。在这些争论中,信号检测论独占鳌头,成为一种新理论。

第二节 信号检测论的原理

一、信号检测论的由来

随着阈限理论和近代科学技术的发展,一种新的心理物理法——信号检测论诞生了。信号检测论 (或讯号侦察论、讯号觉察论) (signal detection theory, 简称 SDT) 乃是信息论的一个分支,研究的对象是信息传输系统中信号的接受部分。它最早用于通讯工程中,即借助于数学的形式描述“接受者”在某一观察时间将掺有噪音的信号从噪音中辨别出来。

信号检测论的形成有一个发展过程。早在 20 世纪 20 年代末,就有人对信息传输的理论进行了讨论,引进信息量的概念,并取得初步的结果。到了 40 年代初,人们便清楚地认识到,由于接受的信息带有某种随机的性质,因此,系统本身的结构也必须适应于它所接收和处理的信息这种统计性质。1941~1942 年,人们开始将统计方法应用于通讯系统研究中,从而建立了最佳线性滤波理论——维纳滤波理论 (Wiener's filter theory)。从最小均方差准则出发,得出对线性滤波器最佳传输函数的要求。1943 年,人们在

雷达技术发展需要的推动下，在研究如何提高雷达检测能力时，提出了一种最佳线性滤波理论。人们在同噪音进行斗争中总结出来的各种方法，实质上都是有意识地利用信号与噪音的统计特性来尽可能抑制噪音，从而提取信号的。1946~1948年建立了基础信息论和潜在抗干扰理论。后者是用概率方法研究高斯噪音中接收信号的理想接收机问题，将那种能够使错误判断概率为最小的接收机称为理想接收机。申农(Shannon, 1948)便认识到对消息的事先确定性这一点恰恰是在通信的对象的基础上建立起来了信息论的基础理论。几年以后，于1950年人们开始把信息量概念引用于雷达信号检测中来，提出一系列综合最佳雷达系统的新观念。其基本特点在于，理想接收机应当能从信号与噪音混合波形中提取最多的有用信号。从50年代起，人们在广泛运用现代数学工具基础上，建立了比较系统的信号检测理论。

信号检测理论除了对雷达、声纳、通讯、自动控制等技术的发展奠定了理论基础外，目前还在心理学、地震学、天文学、生物物理学以及其他科学领域里获得了广泛地应用和发展。同时，信号检测论在这些学科中的应用，又反过来推动了信号检测论不断完善和发展。

那么信号检测论为什么能用于心理学中呢？这是由于人的感官、中枢分析综合过程可看作一个信息处理系统(或讯息处理系统)，因此有可能应用信号检测论中的一些概念和方法对它进行分析。信号检测论还可以从另一个侧面加深人们对感受系统的理解。通常把刺激变量看作是信号，把刺激中的随机物理变化或感知处理信息中的随机变化看作是噪音。这样，人作为一个接收者对刺激的辨别问题便可等效于一个在噪音中检测信号的问题。显然噪音的统计特性确定后，便可应用信号检测论处理心理学实验结果。于是，坦纳和斯韦茨(Tanner & Swets, 1954)等人最早在密西根大学的心理学研究中把信号检测论应用于人的感知过程，并取得了可喜的结果。

信号检测论引入心理学，确实解决了一些传统心理学方法所不能解决的问题。例如，关于精神病患者与正常人的大小常性，用传统心理物理学研究的结果很不一致。有的实验结果表明，精神分裂症患者和正常人比较，前者趋于超常性；有的实验结果恰好与此相反；也有实验结果表明，在大小常性的问题上二者并无差别，后来用信号检测实验得到的结果表明，在排除反应的倾向性后，非妄想型精神分裂症患者比正常人大小常性的程度要低得多。又如用传统心理物理学测定感觉阈限时，如果主试的指导语改变了，感觉阈限会随之发生变化。究竟是指导语影响了被试的辨别力，还是影响了他的反应倾向呢？传统心理物理学就回答不了这个问题。用信号检测实验得到的结果表明，在不同的指导语的影响下测得的感觉阈限所以不同，不是由于被试的辨别力发生了变化，而是由于改变了判断标准造成的结果。另外还有些信号检测实验表明，用传统心理物理学测得的痛阈提高了，并不一定意味着痛觉感受性的下降，而常常是由于改变了极痛标准而造成的。随着信号检测论的引入，确实把心理物理学的研究向前推进了一大步。目前，信号检测论已经成为一种新的心理物理学，并广泛应用到研究心理现象的各个领域，在感觉、知觉、记忆、工程心理、医学心理等领域都获得了有重要价值的成果。

二、信号检测论的统计学原理

上面谈到，信号检测论是人们在同噪音干扰进行斗争中总结出的方法，实质上是有意识地利用信号和噪音的统计特性来尽可能地抑制噪音，从而提取信号的。信号检测论是在多学科基础上形成的。统计决策理论是信号检测

论的数学基础。

我们知道，统计学是关于经验数据的一种数学推理，它的主要工具是概率论。通常它的目的是对一大堆数据进行一种简单的描述或使之精炼化，使得人们易于理解，并使之与研究的已知情况符合。大家知道，这种数据的精炼化可以用均值、方差和置信度等量来表达。根据统计学原理，可以把从噪音干扰中接收信号的过程看作为一个统计判断过程，即用统计判断方法，根据接收的混合波形作出信号存在与否的判断。从 1953 年起，人们开始将统计检测、参量估计、统计判断以及序列分析等统计学工具用于信号检测问题，并建立起一整套信号检测的统计理论。

下面我们来分析信号检测论的统计学原理。心理学上的信号检测实验一般是在信号和背景不易分清的情况下进行的。对信号检测起干扰作用的背景叫噪音 (noise)，这“噪音”不仅是指纯音信号出现时其他的噪音而言的；在视觉实验中，伴随着亮点信号出现时的照度均匀的背景也叫做“噪音”。总之，对信号起着干扰作用的因素都可当作“噪音”。一般的心理物理的辨别实验，其中包含着刺激 A 和刺激 B。在这种情况下，可将其中一个刺激作为噪音，另一个作为信号。主试呈现的刺激，有时只呈现“噪音”刺激（以 N 表示）；有时在信号刺激加噪音刺激同时呈现（以 SN 表示），让被试对信号刺激做出反应。在呈现刺激前，主试要先告诉被试者 N 和 SN 各自出现的概率。这个概率称为先定概率（或先验概率）(prior probability)。同时对被试者说明判定结果的奖惩办法。因为先定概率和奖惩办法都将影响被试者的判定标准（见本章三节），每次实验呈现的是 N 还是 SN 是随机安排的。主试在呈现刺激之前（约 2 秒前）要先给被试者一个预备信号。

在信号检测实验中，被试者对有无信号出现的判定，可以有四种结果：

1. 击中 当信号出现时 (SN)，被试报告为“有”，这称为击中（或中的）(hit)，以 Y/SN 表示。我们把这个判定的概率称为击中的条件概率，以 $P(H)$ 或 $P(Y/SN)$ 表示。
2. 虚惊 当只有噪音出现时 (N)，被试报告“有”，这称为虚惊（或误报）(false alarm)，以 Y/N 表示。我们把这个判定的概率称为虚惊条件概率，以 $P(FA)$ 或 $P(Y/N)$ 表示。
3. 漏报 当有信号出现时，被试报告为“无”，这称为漏报（或失察）(miss)，以 n/SN 表示。把这种判定概率称为漏报条件概率，以 $P(M)$ 或 $P(n/SN)$ 表示。
4. 正确拒斥 当无信号而只有噪音出现时，被试报告为“无”，这称为正确拒斥 (correct rejection) 或正确 (correct)，以 n/N 表示。我们把这个判定的条件概率称为正确拒斥的条件概率，以 $P(CR)$ 或 $P(n/N)$ 来表示。

这样，噪音背景下的信号检测实验，在每种刺激状态下都存在二种反应可能，其组合就构成一个两择一判决矩阵（见表 5-6），其中 H 和 CR 是正确反应，M 和 FA 是错误反应。如果用概率表示，则显然有

$$P(H) + P(M) = 1$$

$$P(FA) + P(CR) = 1$$

从式中可见，其他两个条件概率是这两个条件概率的补数，即知道其中一个数，就可求出互补的另一个数：

$$P(H) = 1 - P(M)$$

$$P(FA) = 1 - P(CR)$$

因此，被试的判定，虽然有四种结果，但判定的条件概率一般只用击中的条件概率和虚惊的条件概率两种，即 $P(H)$ 和 $P(FA)$ 。

以上这四种判定结果，往往用一矩阵表示，见表 5-6。

表 5-6 两择一判决矩阵

刺激 \ 反应	有信号	无信号
有信号	击中 HP (H)	漏报 MP (M)
无信号	虚惊 FAP (FA)	正确拒斥 CRP (CR)

(采自 Green & Swets , 1966)

从统计学观点来看，信号检测即是要检验两个统计假设 H_0 (无信号) 和 H_1 (有信号) 的真伪。设想检测者测量单一变量 X ，并以此为根据选择 H_0 或 H_1 。在无噪音条件下，当 $X=A_0$ 时， H_0 为真；当 $X=A_1$ 时， H_1 为真 (图 5-7)。但在噪音背景下，无信号时 X 并不总是等于 A_0 ，有信号时 X 也并不总是等于 A_1 ，而是分别形成两个概率分布 $P_0(X)$ 和 $P_1(X)$ 。这时，检测者需要确定一个反应标准 X_c ，将 X 分成二个值域，当 $X < X_c$ 时，判定 H_0 为真；当 $X > X_c$ 时，判定 H_1 为真 (图见 5-8)。

在噪音背景下，无论将 X_c 确定在哪一位置，都存在有错误的可能，即虚惊错误 FA 和漏检错误 M 。如图 5-8 所示，曲线 $P_0(X)$ 在 X_c 右面部分所包含面积为虚惊率 Q_{FA} ，曲线 $P_1(X)$ 在 X_c 左面部分所包面积为漏检率 Q_M ，两者均可用积分方法求得：

$$Q_{FA} = \int_{X_c}^{\infty} P_0(X)dX \quad [公式5-5]$$

$$Q_M = \int_{-\infty}^{X_c} P_1(X)dX \quad [公式5-6]$$

因此，在信号分布和噪音分布不变的情况下，检测者选择的反应标准 X_c 将影响 $P(H)$ 、 $P(M)$ 、 $P(FA)$ 和 $P(CR)$ 。反应标准的选择，称为检测者的反应偏向，它是信号检测论 (SDT) 的两个独立指标之一。

为了形象地理解上述原理，我们可想象日常生活中这样一个例子。假设有一个气象观察员，每天要在两个选择中进行判决。 H_0 表示明天是雨天； H_1 表示明天是晴天。这个判断根据单一量进行，这个量就是在过去 24 小时内气压表压力的平均变化率 X 。从多年积累下来的记录计算出能描述在雨天前一天的变化率分布的概率密度函数 $P_0(X)$ ，以及能描述在晴天前一天的变化率分布密度函数 $P_1(X)$ 。例如，变化率 X 可能是正态分布的，其中 $A_0 < C < A_1$ ，这些密度函数画在图 5-8 中。第一个概率密度函数意味着在雨天前一天气压表指针下降，平均下降率为 A_0 ，但下降率并不总是相同的，有时高一点，有时低一点，高低的原因观察者不可能完全知道。同理，第二个曲线表示在晴天的前一天，气压表通常是上升的，平均上升率为 A_1 。

利用上面的这些信息，这个气象观察员将怎样来确定他的策略呢？从两个概率密度函数的结构可以看出，把 X 取值的范围作一个简单的两段划分就

已够了，区域 R_0 由 $X < X_c$ 组成，当 X 在 R_0 取值时，就选择假设 H_0 ，而区域 R_1 由 $X > X_c$ 组成。但是不论观察员把 X_0 放在什么地方，他有时仍会作出错误的判断。事实上，当实际是 H_0 而选择的是 H_1 的概率，即所谓第一类错误 (type error)，为 Q_{FA} 。曲线 $P_0(X)$ 在 X_c 的右边所包面积就代表这个积分 (见图 5-8)，亦即为虚惊率 (probability of false alarm)。当实际是 H_1 而选择的是 H_0 的概率时，即第二类错误 (type error)' 为 Q_M 。曲线 $P_1(X)$ 在 X_c 左边所包的面积就代表这个积分，亦即为漏报率 (probability of miss)。

观察员所用的值 X_c 依赖于这些错误要付出多少代价。为了使问题更确定一些，我们设想这个气象观察员对于每种错误要付出一定量的罚金：对于第一类错误，他要付出 C_0 ，对于第二类错误他要付出 C_1 ，乘积 $C_0 Q_{FA}$ 称为与 H_0 相联系的风险；同理，乘积 $C_1 Q_M$ 是与假设 H_1 相联系的风险。

由上可见，统计决策理论是信号检测论的数学基础。

三、信号检测论的二个独立指标

前面讲到，信号检测论者认为阈限不是一个纯净的感知能力的指标，它受利害得失、动机、态度、意志等因素的影响。而且，一般纠偏和校正猜测机遇的公式都不能对被试者的真实感知能力作正确的估计。因为被试者在报告他对某个刺激或信号的感受时，总是受到动机等因素的影响，考虑其得失利害，采用某种策略，做出某一抉择，使被试者自己的报告符合他最大的利益。根据信号检测论把刺激的判断看成对信号的侦察和做出抉择的过程。这样，在同一过程中就出现了两个独立指标：一个是感觉辨别力 (sensory discriminability) 指标 (d')，表示感知能力，又称为感觉敏感性 (sensitivity)；另一个是反应偏向 (response bias)，可用似然比值 ()，有时也用报告标准 (C) 来表示，它包括利益得失、动机、态度等因素。

下面我们分别来讨论信号检测论的二个独立指标及其计算方法。

(一) 反应偏向

反应偏向可由二种方法计算：一种是似然比值，另一种是报告标准。

1. 似然比值——

信号检测论中用以表示反应偏向的指标之一是似然比值 (likelihood ratio)，意指信号加噪音引起的特定感觉的条件概率与噪音引起的条件概率的比值，其数学定义为给定 X_c 水平上信号分布的纵轴与噪音分布的纵轴之比，即：

$$= \frac{\text{纵轴}_{\text{击中}}}{\text{纵轴}_{\text{虚惊}}} = \frac{Q_{\text{击中}}}{Q_{\text{虚惊}}} \quad [\text{公式5-8}]$$

式中“0”代表纵轴。从中可以看到，因先定概率不同和判定结果的奖惩办法不同而不同。

当 值偏高时， X_c 右移， $P(H)$ 和 $P(FA)$ 均下降，表示检测者的反应标准较严；当 值偏低时， X_c 左移， $P(H)$ 和 $P(FA)$ 均上升，表示检测者的反应标准较宽。

信号检测论可以给出在给定信号强度和敏感性水平下能得到的最大收益

的值 (β_{OPT})。这里，英文字母 OPT 为 OPTIMAL 的缩写，其意是最佳的意思。就物理环境而言， β_{OPT} 的主要影响因素为信号概率 (probability of signal) 和两择一判决矩阵中四种可能结果的奖惩 (支付) 情况亦即支付矩阵 (pay-off matrix)。支付矩阵是指在一定的信号和噪音出现的先验概率条件下，对被试判断结果的奖惩办法。可表示为：

$$= \frac{P(N)}{P(S)} \times \frac{V(CR) + C(FA)}{C(M)}$$

式中 $P(N)$ 和 $P(S)$ 分别是噪音概率和信号概率， V 代表由正确反应得到的价值， C 代表错误反应得到的代价 (负值)。当 $V(CR) + C(FA) = V(H) + C(M)$ 时，最大收益将从最小错误而得到。此时 β_{OPT} 随信号概率 $P(S)$ 的变化而变化 ($P(S) + P(N) = 1$)，当 $P(S)$ 上升时，放宽反应标准，即降低 β 是有利的反应策略；当 $P(S)$ 降低时，严格的反应标准使提高 β 更有利。但当从正确反应中得到的收益与错误反应付出的代价不等时， β_{OPT} 将发生相应的变化。

由于检测者实际反应偏向 β 可以根据已知的 $P(H)$ 和 $P(FA)$ 求得，因此，人们自然会问，当信号概率和支付矩阵发生变化时，检测者实际上是如何确定的。格林和斯韦茨 (Green & Swets, 1966) 通过实验室研究发现，反应者确实在根据信号概率和支付矩阵的变化调节 β ，但 β 的调节没有达到理想水平。这个现象称为 β 惰性 (sluggish beta)，其关系如图 5-9 所示：如果 β_{OPT} 较低，实际 β 值倾向于偏高；而当 β_{OPT} 较高时，实际 β 值又倾向于偏低。此外，格林和斯韦茨还发现，支付矩阵对检测者实际反应倾向的影响大于信号概率的影响，换句话说，当信号概率发生变化时， β 惰性现象更加明显。

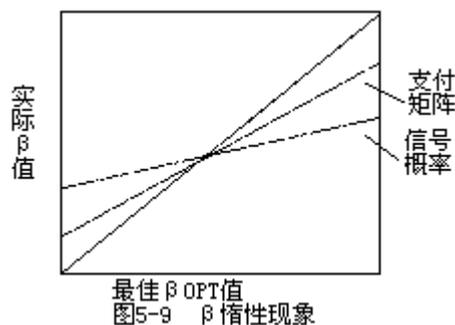


图5-9 β 惰性现象

以上分析使我们明确了反应偏向 (β) 的含义。可见 β 值虽被认为是反应阈限，但这个阈限和传统阈限概念不同，它并非恒定不变，而是随 SN 和 N 两种先定概率和判定结果的奖惩办法而变动的。 β 值和先定概率 $P(S)$ 、 $P(N)$ 以及奖惩办法的关系可用如下公式表示：

$$= \frac{P(N)}{P(SN)} = \frac{V_n N + (V_y N)}{V_y S + (V_n S)}$$

$V_n N$ ：正确否定的奖励数

$V_y N$ ：虚报的惩罚数 (往往是负数)

$V_y S$ ：击中的奖励数

$V_n S$ ：漏报的惩罚数 (往往是负数)

在明确了 β 含义后，我们举图 5-10 上 A、B、C 三种情况为例，说明 β 的

具体计算方法。

计算过程中必不可少地要使用 PZO 转换表（见表 5-7），表中 P 代表概率，Z 代表标准分数，O 代表概率密度，PZO 转换表就是这三个数值之间的对应互换关系。图 5-10 的情况 A，击中概率为 0.28，虚惊概率则是 0.06，通过查表 5-7，求得 $O_{\text{击中}}$ 的纵轴值为 0.3368， $O_{\text{虚惊}}$ 的纵轴值为 0.1192。则

$$= \frac{Q_{\text{击中}}}{Q_{\text{虚惊}}} = \frac{0.3368}{0.1192} \approx 3$$

一般认为， > 1 说明被试者掌握的标准较严。

图 5-10 的情况 B，击中概率为 0.70，虚惊概率为 0.30，通过查转换表 5-7，求得 $O_{\text{击中}}$ 的纵轴值为 0.3478， $O_{\text{虚惊}}$ 的纵轴值为 0.3478。那么，

表 5-7 PZO 转换表

P	Z	O	P	Z	O	P	Z	O	P	Z
.01	-2.326	.0267	.26	-0.643	.3245	.51	.025	.3988	.76	.706
.02	-2.053	.0484	.27	-0.612	.3308	.52	.050	.3984	.77	.738
.03	-1.881	.0681	.28	-0.582	.3368	.53	.075	.3978	.78	.772
.04	-1.750	.0862	.29	-0.553	.3424	.54	.100	.3970	.79	.806
.05	-1.645	.1032	.30	-0.524	.3478	.55	.125	.3958	.80	.841
.06	-1.555	.1192	.31	-0.495	.3529	.56	.150	.3945	.81	.877
.07	-1.476	.1343	.32	-0.467	.3577	.57	.176	.3928	.82	.915
.08	-1.405	.1487	.33	-0.439	.3623	.58	.201	.3909	.83	.954
.09	-1.340	.1625	.34	-0.412	.3665	.59	.227	.3888	.84	.994
.10	-1.281	.1756	.35	-0.385	.3705	.60	.253	.3864	.85	1.036
.11	-1.226	.1881	.36	-0.358	.3742	.61	.279	.3838	.86	1.080
.12	-1.175	.2001	.37	-0.331	.3777	.62	.305	.3808	.87	1.126
.13	-1.126	.2116	.38	-0.305	.3808	.63	.331	.3777	.88	1.175
.14	-1.080	.2227	.39	-0.279	.3838	.64	.358	.3742	.89	1.226
.15	-1.036	.2333	.40	-0.253	.3864	.65	.385	.3705	.90	1.281
.16	-0.994	.2434	.41	-0.227	.3888	.66	.412	.3665	.91	1.340
.17	-0.954	.2532	.42	-0.201	.3909	.67	.439	.3623	.92	1.406
.18	-0.915	.2625	.43	-0.176	.3928	.68	.467	.3577	.93	1.476
.19	-0.877	.2715	.44	-0.150	.3945	.69	.495	.3529	.94	1.555
.20	-0.841	.2801	.45	-0.125	.3958	.70	.524	.3478	.95	1.645
.21	-0.806	.2883	.46	-0.100	.3970	.71	.553	.3424	.96	1.750
.22	-0.772	.2962	.47	-0.075	.3978	.72	.582	.3368	.97	1.881
.23	-0.738	.3038	.48	-0.050	.3984	.73	.612	.3308	.98	2.053
.24	-0.706	.3110	.49	-0.025	.3988	.74	.643	.3245	.99	2.326
.25	-0.674	.3179	.50	0.000	.3989	.75	.674	.3179		

（采自马谋超，1978）这时的 值为

$$= \frac{0.3478}{0.3478} = 1$$

值接近或等于 1，说明被试者掌握的标准不严也不松。

图 5-10 的情况 C，击中概率为 0.94，虚惊概率为 0.72，通过查表，求得 $0_{\text{击中}}$ 的纵轴值为 0.1192， $0_{\text{虚惊}}$ 的纵轴值为 0.3368。故

$$= \frac{0.1192}{0.3368} \approx \frac{1}{3}$$

值 < 1 ，说明被试者掌握的标准较松。

以上实例，不仅告诉我们值的计算方法，且使我们看到值的三种情况：情况 A 为严标准、情况 B 为居中、情况 C 为宽松标准。

2. 报告标准——C

信号检测论中用以表示反应偏向的另一个指标是报告标准 (report criterion)，又称判断标准 (或判定标准) (judgment criterion)。在数学上，反应标准的另一种表示方法是感受经验强度，用符号 C 表示。C 是横轴上的判定标准位置。在数学上，C 的单位要转换成刺激强度单位，它的计算公式是：

$$C = \frac{I_2 - I_1}{d'} \times Z_1 + I_1 \quad [\text{公式 5-9}]$$

I_2 ：为高强度刺激

I_1 ：为低强度刺激

Z_1 ：为低强度刺激时的正确拒斥概率的 Z 值

我们以心理学上再认实验来说明报告标准会出现的种种情况。再认实验一般有两组图片，一组是“新的”（即未见过的），一组是“旧的”（即已看过的）。在实验过程中，先让被试看一组图片，然后将其与另一组图片混合，被试在再认过程中，根据自己确定的标准，回答“新的”或“旧的”。这里包括两个指标，其中一个感觉辨别力指标，亦称感觉敏感性，以 d' 值表示，不受情绪、期望、动机等变数的影响。 d' 值低表示被试对新、旧刺激不易辨别。当刺激很接近或被试者再认不敏感时， d' 值就低。 d' 降低表示再认能力减弱。另一个值是判断标准，即个体反应偏向，以 C 值表示。C 值高表示被试者判断的标准严格，不轻易报告“旧的”图片；反之，C 值低表示被试者判断旧刺激的标准宽松，易把一些新刺激说成旧刺激。

信号检测论指标 d' 比较稳定，不受实验条件不同的影响。信号检测论指标 C 受被试者的动机、态度、利害得失等心理因素影响。我们可从指标 C 的变化中，分析被试者的心理因素。

参看图 5-11，将 d' 固定，则 C 会出现三种情况：（1）宽松的报告标准：旧刺激呈现时，报告“旧的”概率接近 1.0；新刺激呈现时，报告“旧的”概率是高的。（2）中等的报告标准：旧刺激呈现时，报告“旧的”概率是较高；新刺激呈现时，报告“旧的”概率适中。（3）严格的报告标准：旧刺激呈现时，报告“旧的”概率是低的；新刺激呈现时，报告“旧的”概率接近 0.0。

表 5 - 8 再认实验中某被试者的刺激-反应矩阵

反应 \ 刺激	报告“旧的”	报告“新的”
旧刺激	42	8
新刺激	4	46

(采自杨治良, 1983)

根据公式 5 - 9, 我们就可求出某被试者在再认图片中的报告标准。假如在这个实验中, 新、旧图片各 50 张, 并获得了表 5 - 8 上的结果。在此实验中, 在横轴上设新刺激强度为 0, 旧刺激强度为 1, 就可先求得 d' 值。

$$d' = Z_{\text{击中}} - Z_{\text{虚惊}}$$

$$= 0.994 - (-1.405) = 2.399$$

上式中的数值是通过 PZ0 转换表所查得的, 数值 0.994 是击中概率 42/50 的 Z 值, 数值 (-1.405) 是虚惊概率 4/50 的 Z 值。这样, 求出了 d' 值就可按公式 5 - 8 算出 C 值:

$$C = \frac{I_2 - I_1}{d'} \times Z_1 + I_1 \quad [\text{公式5-9}]$$

$$= \frac{1 - 0}{2.399} \times 1.405 + 0 = 0.59$$

C 值 0.59 是在判定轴上的位置, 前面曾讲到, I_1 为 0, I_2 为 1, 因此, C 略靠近 I_2 , 所以可以认为被试掌握的标准略严。

(二) 敏感性指标

信号检测论的最主要贡献是在反应偏向与反应敏感性之间能作出区分。击中概率 $P(N)$ 可以因为检测者持严格的反应标准(高)而减低; 另一方面, 即使观测者持宽松的反应标准(低), $P(N)$ 也可能因为敏感性的降低而降低。

敏感性可以表现为内部噪音分布 $f_N(X)$ 与信号加噪音分布 $f_{SN}(X)$ 之间的分离程度。两者的分离程度越大, 敏感性越高; 分离程度越小, 敏感性越低。图 5 - 12 给出了反应偏向() 相同的情况下, 两种敏感性情况。

内部噪音分布 $f_N(X)$ 与信号分布 $f_{SN}(X)$ 的分离程度既受信号的物理性质影响, 也受被试者(测验者)特性的影响。因此, $f_N(X)$ 与 $f_{SN}(X)$ 之间的距离就可作为敏感性的指标, 称为辨别力 d' :

$$d' = \frac{M_{SN} - M_N}{\sigma_M}$$

即辨别力 (d') 等于两个分布的均数之差除之 N 分布的标准差。当 IUM) 等于两个分布的均数之差除之 N 分布的标准差。当 N 分布与 SN 分布均为常态分布时, 其变异数类同, 则有:

$$d' = \frac{M_{SN} - M_N}{\sigma} = Z_{SN} - Z_N = Z_{\text{击中}} - Z_{\text{虚惊}} \quad [\text{公式5-10}]$$

d' 越大, 表示敏感性越高, d' 越小, 表示敏感性越低。

与反应偏向一样, 反应敏感性同样存在最佳水平。被试者(检测者)能否达到最佳水平, 与能否对 N 条件和 SN 条件下物理能量的统计特征作出精确估计有关。格林和斯韦茨 (Green & Swets, 1966) 曾对白噪音背景下听觉信号的检测问题进行了实验室研究, 结果指出, 敏感性对最佳水平的偏离(下

降)与被试者缺乏对信号物理特征的精确记忆有关。这些资料具有极其重要的实用价值。

下面我们通过固定反应偏向()来看辨别力(d')的情况。本例设 $\beta=1$,则 d' 可能出现三类情况。我们知道,“击中”概率落在纵轴右方的高强度分布(或信号加噪音分布)曲线,“虚惊”概率落在纵轴右方的低强度分布(或噪音分布)曲线。

(1)第一类情况是:在输入感觉刺激非常敏感的情况下,当信号加噪音存在时,常常出现“肯定”,击中率为93%;当噪音单独存在时,很少做出“肯定”,虚惊率为7%,这时:

$$d' = Z_{\text{击中}} - Z_{\text{虚惊}} = 1.476 - (-1.476) = 3$$

以上是根据“击中”和“虚惊”概率,再通过PZO转换表求得的。图5-13(a)就是本例的情况。图上横轴的单位是噪音 $f_N(X)$ 时的Z值,纵轴单位是概率密度。

(2)第二类情况是:当感觉鉴别能力降低时,“击中”和“虚惊”分布二者相互接近。例如当中等情况时,击中率为84%;虚惊率为16%,则:

$$d' = Z_{\text{击中}} - Z_{\text{虚惊}} = 0.994 - (-0.994) = 3$$

图5-13(b)就是本例的情况。

(3)第三类情况是:被试者相对不敏感,或刺激相对比较弱,击中率下降为70%;虚惊率增为30%,则

四、接受者操作特性曲线

接受者操作特性曲线(receiver operating characteristic curve),简称ROC曲线(ROC curve)在心理学上又称为感受性曲线(sensitivity curve),这就是说,曲线上各点反映着相同的感受性,它们都是对同一信号刺激的反应,不过是在几种不同的判定标准下所得的结果就是了。接受者操作特性曲线以虚惊概率为横轴,击中概率为纵轴所组成的坐标图和被试者在特定刺激条件下由于采用不同的判断标准得出的不同结果画出的曲线。

前面讲到,判定过程需要给定一个标准,超过标准时就作出“有目标”的判定。例如,超过标准时,接受机接通继电器,发出警报,但是,若阈值(指给定的判定标准)太高时,只有强目标信号才能检测到;若阈值太低,会有许多“虚惊”。在给定的阈值较高时,目标检测概率和虚惊概率二者都较低;在给定的阈值较低时,这两个概率都变高了。阈值给定的这种效应如图5-14所示。

上图(a)表示三个目标信号的波幅;中图(b)是噪音的波幅,下图(c)是信号加噪音的波幅。从图5-14的下图可见,在阈值为 T_1 时只有第二个目标被检测到,在图中所示的时间区间内没有虚惊。对于固定的输出信噪比(signal-to-noise ratio,指通讯系统中,所要求的信号和背景噪音的功率之比),各种阈值下,有不同的虚惊概率($P_{SN}(A)$)和击中概率($P_N(A)$)。一个接受者的操作特性曲线(ROC曲线)的纵轴表示击中概率(定义为检测到信号的概率),横轴表示虚惊概率(定义为由于噪音而超出阈值的概率)。给定阈值改变时,在击中概率和虚惊概率图上可以画出曲线。图5-15中阈值从 T_1 变到 T_2 时所得的ROC曲线就是一个例子。

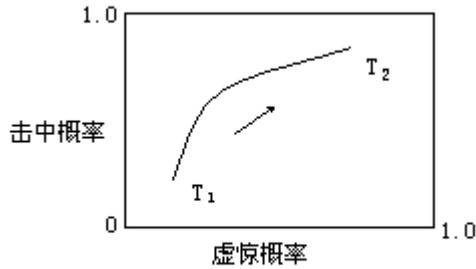


图 5-15 由图5-14构成的一条ROC曲线

接受者操作特性曲线是被试者在特定刺激条件下由于采用不同的判断标准得出的不同的结果所画成的曲线，这两者的形式表示下列各种函数关系：

(1) 当信号呈现的概率对 $P(y/SN)$ 和 $P(y/N)$ 的影响随信号呈现的概率增加时， $P(y/SN)$ 增加，同时 $P(y/N)$ 也增加，此时图中的弓形弯曲度也增加。(2) λ 值的改变对 $P(y/SN)$ 和 $P(y/S)$ 的影响：当 $\lambda=0$ 时，击中概率几乎为 0，即信号全当成噪音接受；当 λ 接近无穷大时，虚惊概率几乎为 0，即噪音全当成信号接受，故最佳的标准 λ 应选一定的 $P(y/SN)$ 和 $P(y/N)$ 的比值。(3) 曲线的曲率反应出敏感性指标 d' ：在图 5-16 中有一条对角线，代表 $p(y/SN) = P(y/N)$ ，即被试者的辨别力 d' 为 0，ROC 曲线离这条线愈远，表示被试者辨别力愈强， d' 的值当然就愈大。信号检测论的任务在于寻找最佳 ROC 曲线。

通过以上分析，可以看到接受者操作特性曲线是严格地由给定阈的信号和噪音的概率密度函数决定的（见图 5-8）。此噪音的平均幅度是 M_N ，横轴上用 A_0 表示；信号加噪音的平均幅度是 M_{SN} ，用 A_1 表示。假定噪音和信号加噪音都是正态分布，方差都是 σ^2 ，比值 d' 可由公式 5-10 计算，从这里可看出一些变量的关系。一般有两种情况：

(1) 第一种情况： d' 固定，从公式上可看到，当 λ 变大时，则 $(M_{SN}-M_N)$ 也成比例变大。此时，若 T 从右方向左方移，从图 5-8 上可看到 $P_0(X)$ 和 $P_1(X)$ 概率变大，这整个变化过程就构成了一条 ROC 曲线。

(2) 第二种情况： T 固定，又有两种情况：若 λ 变大，而 $(M_{SN}-M_N)$ 不变，则 d' 就会变小，此时二概率密度分布曲线就会靠近，其结果 $P_0(X)$ 就大，从图 5-16 上，我们就可见到 d' 值小的 ROC 曲线。若 λ 值不变，而 $(M_{SN}-M_N)$ 值变大，则 d' 就会变大，此时二概率密度分布曲线之间的距离就拉大，因为 T 固定，其结果 $P_0(X)$ 就小，此时，从图 5-16 上看到 d' 值大的 ROC 曲线。

下面我们通过一个实例来叙述 ROC 曲线的具体绘制过程。笔者曾做过这样一个实验：选图画页 500 页，分成五个组，每组 100 张。五组画页的先定概率分别是 0.1、0.3、0.5、0.7 和 0.9。对于每一组画页，主试者使用一种信号的先定概率，然后按此先定概率呈现给被试者一定数量的画页，要求被试者把它们当做“信号”记住。例如，先定概率为 0.1 时，则当作“信号”的画页为 10 张；当做“噪音”的画页为 90 张。作为信号的画页呈现完毕之后，与此组作为噪音的画页混合，然后随机地逐张呈现给被试。这时，每呈现一张画页，即要求被试判断此画页是“信号”还是“噪音”，并要求被试把结果记录在实验纸上。

根据五种先定概率得到的实验结果，就可计算击中概率和虚惊概率。其一般的计算格式见表 5 - 9。

表 5-9 刺激-反应矩阵

R S	是	否
信号	击 中 f_1	漏 检 f_2
噪音	虚 惊 f_3	正确拒斥 f_4

(采自赫葆源等, 1983)

感觉敏感性是 $d' = Z_{\text{击中}} - Z_{\text{虚惊}}$ ，反应偏向为 $b = 0_{\text{击中}} / 0_{\text{虚惊}}$ ，式中 0 代表纵轴，根据 d' 和 b 的公式可计算如下，见表 5 - 10。

根据上面所得的击中概率和虚惊概率，就可求出不同先定概率下的 d' 值和 b 值 (见表 5 - 11)。

表 5-10 五种先定概率的实验结果

当先定概率 P = 0.1			当先定概率 P=0.3			当先定概率 P = 0.5		
R S	是	否	R S	是	否	R S	是	否
信号	3	7	信号	16	14	信号	35	15
噪音	4	86	噪音	9	61	噪音	11	39
P _{击中} =0.30 P _{虚惊} =0.14			P _{击中} = 0.53 P _{虚惊} =0.13			P _{击中} =0.70 P _{虚惊} =0.		

表 5 - 10

五种先定概率的实验结果 (续)

当先定概率 P = 0.9			当先定概率 P=0.7		
R S	是	否	R S	是	否
信号	59	11	信号	83	7
噪音	13	17	噪音	6	4
P _{击中} =0.84 P _{虚惊} =0.43			P _{击中} = 0.92 P _{虚惊} = 0.60		

表 5.11 不同先定概率下的 d' 值和 b 值

项目		y/sN (击中)	y/N (虚惊)	d	
0.1*	P	0.30	0.04	1.227	4.035
	Z	-0.524	-1.751		
	O	0.3478	0.0862		
0.3	P	0.53	0.13	1.201	1.880
	Z	0.075	-1.126		
	O	0.3928	0.2116		
0.5	P	0.70	0.22	1.296	1.174
	Z	0.524	-0.772		
	O	0.3478	0.2962		
0.7	P	0.84	0.43	1.171	0.620
	Z	0.995	-0.176		
	O	0.2434	0.3928		
0.9	P	0.92	0.60	1.152	0.383
	Z	1.405	0.253		
	O	0.1481	0.3864		

* (采自杨治良, 1983)

最后, 根据不同先定概率下的击中概率和虚惊概率, 就可在图上确定各点的位置, 把五点联接起来就绘成一条 ROC 曲线 (见图 5 - 17)。图上各点, 均可通过 PZ0 转换表查得。

第三节 信号检测论的应用

作为一种新的心理物理学方法，信号检测论对实验心理学领域产生了巨大影响，它对应用心理学各个领域中的许多问题也具有广泛的应用价值。由于信号检测论在感觉敏感性 (d') 与反应偏向 (β) 之间作出区分，因此，它能够分析不同被试、不同操作条件下的反应敏感性；同时，信号检测论能够分析操作的恶化是因为敏感性下降还是因为反应偏向的变化，并根据这些分析的资料对操作进行改进。下面我们讨论信号检测论在心理学中的几个主要应用领域。

一、在医学心理学中的应用

在应用领域中，医学诊断是信号检测论大有作为的领域。异常症状既可出现于病人也可出现于正常人，医生最初的任务就是作出“是”或“不是”的决断。在这里，信号强度（将影响 d' ）与异常症状的显著程度、异常症状的多少、医生对有关线索的关注（取决于医生所受的训练）有关。反应偏向的影响因素包括信号概率和支付矩阵，具体地说，前者指疾病的发病率；后者主要是诊断的可能后果。例如，如果虚惊，就诊者将接受不必要的治疗（服药还是手术？治疗有无副作用？）；若漏检，病人将延误治疗（是否会产生严重后果？）（Lusted, 1976）。斯韦茨等人（Swets & Pickett, 1982）曾指出，医生的诊断操作可以用 ROC 曲线进行数量化，他们在进一步的研究中还详细考虑了应用信号检测论分析医学诊断的方法学问题。

一部分研究者关注更具体的诊断问题，例如放射学家根据 X 光片所作的诊断，这类诊断通常距最佳操作甚远。据估计这类诊断中异常症状的漏检率可高达 20% 到 40%，运用信号检测论进行的分析得出了一些有用的结果。例如，史文森（Swensen, 1977）的研究指出，让操作者的注意力集中于 X 光片上的异常症状可能发生的区域能够提高检测到症状的可能性。但提高的原因是操作者降低了反应偏向 β ，而不是增加了感觉敏感性 (d')。史文森等人（Swensen et al., 1979）还比较了两种条件下进行的异常症状检测，第一种条件是很多 X 光片中有一部分包含了一种异常症状。第二种条件是需检测的关键异常症状同包括有其他病理表现的 X 光片混合在一起。比较结果表明，第一种条件下的击中概率较高，但这一提高是由降低反应偏向 (β) 造成的，事实上第一种条件下的敏感性还低于第二种条件。

帕拉苏拉曼（Parasuraman, 1980）对医生和住院医师（实习医生）的诊断进行了信号检测论分析，结果表明，正式医生的感觉敏感性 (d') 一般高于住院医师，但总的说来正式医生的反应偏向 (β) 较高（更加保守）。此外，正式医生对发病率的反应更加敏感（调整 β ）。显然，这些研究所得出的结果对改善医生的诊断操作是较为有用的资料。

下面我们举一实例来说明信号检测论的实际应用价值。

根据传统心理物理学，痛阈（threshold of pain）乃是被试报告痛觉时的刺激强度，也就是对 50% 次数的刺激报告痛的强度。报告痛的次数比例高，表示痛敏感性大，即被试者对有害刺激的痛阈低；反之，报告痛的比例少，即对痛不敏感，表示个体具有高阈值。但是，信号检测论认为，一般常用的阈值测定是一种极不可靠的测痛方法。因为阈值并不单纯是感觉敏感度的指标，它还受被试者反应偏差的影响，即他愿意或不愿意报告是否有刺激存在。

这里引用克拉克 (Clack, 1974) 关于痛感受性和报痛的研究。克拉克采用信号检测论, 研究了提示对热辨别力 (d' 值) 和反应判定标准 (β 或 C) 的影响, 并与传统方法测得的阈值客观地进行了比较分析。研究结果不仅说明了信号检测论的理论意义, 且还使我们看到了实际价值。

由枪式热辐射器提供热刺激, 每一个被试者在提示前后各接受 72 次刺激, 分为 6 个强度, 各重复 12 次。被试者共 10 人, 试验前, 实验者对被试者说明实验的目的, 并罗列一些可能出现的反应, 以帮助被试者前、后回答一致。罗列的热感觉强度有 11 个等级, 它们是: 没有感觉、略有感觉、微温、温热、烫、很烫、极微痛、轻痛、中痛、重痛、回缩动作。

实验结束后, 休息 5 分钟, 继续进行实验。此时, 向被试者说“马上开始的实验要说明, 重复进行刺激可改变你的耐痛能力。因为, 原先的刺激已使你的皮肤感受器疲劳而不那么敏感。现在你可能更能耐受痛刺激, 将延迟投射器从皮肤上移开的时间。现在开始就要测定你的耐受剧痛的能力。”实验者在痛阈和个体反应标准的研究中, 将口头报告“极微痛”定为痛阈, 并将“回缩反应”定为耐痛阈 (threshold of bearable pain) 或痛觉耐力 (pain endurance)。

为了进行方法上的比较, 采用传统心理物理法分析实验结果所得的结论是: (1) 提示能取得提高回缩阈值 (即耐痛阈) 的预期效果, 当刺激强度为 0.435 卡/秒/平方厘米时, 回缩反应的概率从 0.75 下降到 0.53 (配对 $t=3.20$, 自由度=9, $P<0.025$), 提示前、后回缩阈值 (50% 阈值) 分别为 0.385 和 0.430 卡/秒/平方厘米。由此可见, 提示能够明显提高回缩阈值, 约为 0.045 卡/秒/平方厘米。(2) 提示时口头报告的痛阈影响极小, 累计的极微痛反应主要发生在 0.370 和 0.350 卡/秒/平方厘米强度刺激。由于在这类反应之间的差异, 在统计学上不显著, 因此提示不可能对口头痛报告的阈值有任何影响。

表 5-12 提示前、后被试者各类反应的累积条件概率

反应刺激		没有感觉	略有感觉	微温	温热	烫	很烫	极微痛	微痛	痛	很痛	回缩动作
提示前	-0.435*									1.0	0.85	0.80
	0.370						1.0	0.92	0.67	0.50	0.42	0.36
	0.305					1.0	0.92	0.67	0.36	0.25	0.08	0.08
	0.240		1.0	0.92	0.75	0.67	0.67	0.33	0.14	0.14	0	0
	0.120	1.0	0.92	0.75	0.67	0.25	0	0	0	0	0	0
	0	1.0	0.36	0.25	0.14	0.08	0.08	0.08	0	0	0	0
提示后	0.435								1.0	0.92	0.80	0.62
	0.370						1.0	0.75	0.75	0.50	0.33	0.14
	0.705				1.0	0.92	0.80	0.25	0.25	0.08	0	0
	0.240	1.0	0.83	0.75	0.75	0.67	0.42	0.08	0	0	0	0
	0.120	1.0	0.83	0.67	0.36	0.25	0	0	0	0	0	0
	0	1.0	0.42	0.25	0.08	0.08	0	0	0	0	0	0

*单位: 卡/秒/平方厘米 (采自 Clark & Yang, 1974)

若用信号检测论处理此实验结果又将会得到什么结果呢? 表 5-12 为某

一被试者的反应的累积条件概率。由此可算出感觉敏感性(d')和判定标准(C)。信号检测论假定,横轴代表主观经验不同强度,纵轴代表其发生的概率。随着温度刺激逐步加强,被试可以从无感觉到感到微温、烫、很烫、极微痛、轻痛、中痛、重痛等等级不同、但分布上互相连接的感觉。由于感觉在心理上是连续的,相邻的两个刺激引起的感觉分布曲线可以有部分重叠(见图5-18)。因此,沿判定轴分布的每一感受经验有两种概率表示,一种是低强度(噪音)的分布概率,另一是高强度(信号加噪音)分布概率。被试者所要判定的是:某一感觉是由于低强度还是高强度刺激引起的,因而必须作出统计学决定。被试者为了前、后一致地做出判定,就必须选择一个判定反应标准的点。在本例, C_w 表示回缩反应的判定标准,若被试者对高强度刺激出现回缩反应,表示“击中”;若对低强度刺激也出现回缩反应,表示“虚惊”。图5-18中, C_w 右方的高强度刺激分布下的面积(80%)表示击中概率,而同一标准右方的低强度刺激分布下的面积(36%)表示虚惊概率。根据“击中”和“虚惊”概率就可推算出被试者的二种反应指标:感觉敏感性(d')和判定标准 C 的位置。

根据公式 $d' = Z_{\text{击中}} - Z_{\text{虚惊}}$, 求得:

$$\text{提示前 } d' = 0.841 - (-0.358) = 1.2$$

$$\text{提示后 } d' = 0.305 - (-1.08) = 1.385$$

根据公式

$$C = \frac{I_2 - I_1}{d'} \times Z_{\text{正确拒斥}} + I_1$$

求得:

$$\text{提示前 } C = \frac{0.435 - 0.370}{1.2} \times 0.358 + 0.370 = 0.3894$$

$$\text{提示后 } C = \frac{0.435 - 0.370}{1.385} \times 1.08 + 0.370 = 0.4206$$

以上我们用传统及现代心理物理法处理了实验数据,分析表明,提示的唯一作用是使被试者提高其回缩的判定标准,即提示能使被试者耐受较高强度的热量,而不会引起躯体的回缩动作,使被试者提高其反应的判定标准。可见,传统的心理物理法都公认,提示能明显地提高回缩反应的痛阈,痛阈的提高是由于痛感受的减轻所致。但是,用新的心理物理法(即信号检测论)对同一资料进行分析,则完全否定了这一解释。事实上,被试者的感觉辨别力(d')始终没多大改变,所改变的仅仅是他的痛阈报告的标准而已。这正是信号检测论优越性所显示的应用价值。

二、在工程心理学中的应用

工程心理学(engineering psychology) 研究人与机器相互作用的过程中,人机功能相互适应与配合的心理学问题。随着生产技术的日益发展,人与机器间的相互关系也愈显重要。在复杂的人机关系中,警戒操作是工程心理学中的一个重要问题:操作者需要在相当长的一段时间里检测信号,而信号是间歇性的,无法预测的,发生频率很低。诸如雷达操作员监测偶然出现的飞机信号,复杂工业系统的操作人员监测偶然发生的系统故障,流水线上的质量检测员监测次品等操作都可以看作是警戒操作。日常生活中,文字的校对也和人的持续注意有关。

警戒 (vigilance) 是指操作者在相当长的一段时间内, 对环境偶然出现的某种信号的察觉并做出反应的持续准备状态。它是心理学领域中持续注意概念在工程心理学中的应用。对警戒问题的研究兴趣起源于第二次世界大战。当时, 军事上需要高水平的雷达操作员, 理想的雷达操作员应始终保持对显示屏的注意。可是, 实际情况常常是, 尽管重要的战备任务使操作员保持高度良好的动机和责任心, 可是他们的工作成绩总是不能令人满意, 或是漏报信号, 或是虚报, 而且往往随着工作时间的延续, 检测出的信号越来越少。很显然, 不是操作员的主观愿望或其他主观因素直接决定他们的工作成绩, 而是由于雷达监测工作的特点降低了操作员的警戒能力, 从而使他们的工作成绩随时间的持续而下降。

为了探明其中的原因, 心理学家们曾经做了大量的研究工作。麦克沃斯 (Machworth, 1948) 以经典的“钟表任务”来观测人的警戒行为。在这个任务中, 让被试监视一个钟表的指针, 规定指针跳动某一角度为标准角度, 而把偶尔跳动标准角度的两倍的刺激称为信号事件。实验结果发现, 被试正确检测信号的数量随着时间呈下降趋势, 在连续注意钟表达 30 分钟后, 就只能检测到很少的信号了。以后, 研究者们又设计了多种警戒任务, 以此来研究人的警戒行为。这些任务具有如下一些共性: (1) 被试的任务长, 且有连续性; (2) 被试的反应与信号发生的可能性无关; (3) 信号的清晰度较高; (4) 信号是随机发生的, 且概率较低。

根据这些任务所作的警戒操作研究的结论是: (1) 人们的操作远非“最佳操作”, 尤其是漏检率 $P(M)$ 很高 (或者是很长的延迟后才反应); (2) 以击中概率衡量, 某些情况下警戒操作在半小时内就会开始恶化, 这种情况称为警戒衰退 (vigilance decrement, 指操作随时间延续而衰退)。它不同于警戒水平 (vigilance level), 警戒水平是指警戒操作的稳定水平, 通常用 $P(H)$ 作为指标。警戒衰退与警戒水平的影响因素有所不同。一般说来, 当信号较短、能量较低、复杂程度高、信号概率降低时, 警戒水平将降低; 但当信号概率稳定而仅仅事件频率 (单位时间里的信号和噪音事件) 发生变化时, 警戒水平不受影响。对警戒衰退来说, 信号较短、能量较低、信号概率下降或信号概率稳定但事件频率上升都能导致警戒衰退的增加。

进一步的研究将警戒操作粗略地分成两类: 自由反应式 (free response paradigm) 和检验式 (inspection paradigm)。在自由反应的警戒操作中, 目标事件 (信号) 可能在任何时候发生, 但非目标事件 (噪音) 没有明确定义, 它的信号概率就是单位时间内的目标数 $P(S)$ 。由于自由反应式警戒操作中没有明确的非目标事件, 因此在用信号检测论进行分析时, 对虚惊必须另外作定义。在实验室实验中, 通常在每个信号后规定一个恰当的反应间隔, 操作者在此间隔内作出的反应指定为击中, 剩余的时间划分为几个与反应间隔相等的“虚惊间隔” $P(FA)$, 就指定为虚惊数除以虚惊间隔数。对检验式警戒操作来说, 事件通常以稳定的概率发生, 事件中有一部分是目标事件 (信号), 其余为非目标事件 (噪音)。这样, 检验式警戒操作中, 信号频率可以指单位时间内发生的目标事件数, 也可以指目标事件与总事件数之比例, 这两者可以独立变化。

对警戒衰退所作的信号检测论分析表明, 击中概率 $P(H)$ 降低的原因既有敏感性下降 (d' 降低), 也有反应偏向的变化 (增高), 或者两者兼而有之。因此, 在评价操作警戒能力时, 不能单一地采用击中概率为指标, 而

应把击中概率和虚报概率两者结合起来，还应把感觉敏感性和反应偏向分开处理，只有这样才能说明警戒下降的真正原因。

就敏感性下降而言，早期的研究集中于视觉信号的检测，并认为敏感性 (d') 降低的原因是因为疲劳 (Broadbent, 1971)。到了近期，柏拉苏拉曼等人 (Parasuraman et al., 1979) 的研究开始注意听觉信号检测任务，并指出记忆因素在敏感性 (d') 降低中的作用，因为信号的检测通常需要将当前刺激同保持在工作记忆中的标准刺激或“刺激模板”进行比较。例如，在一个听觉测验实验中，目标信号是音高、强度或时程与标准音调不同的声音，这时，非目标事件 (标准音调) 需要作为比较参照维持在记忆中，当时间延长，观察到了敏感性 (d') 降低。但当实验条件改为声音成对出现时，测验任务无需记忆参与，因为此时比较参照与当前刺激同时呈现。结果，未观察到敏感性 (d') 降低的现象。

在大部分警戒操作中，当击中概率 $P(H)$ 降低时，虚报概率 $P(FA)$ 经常也同时降低。在另一些情况下，虚报概率 $P(FA)$ 的降低幅度足以提醒人们，操作者的反应敏感性 (d') 并未变化，而且是反应偏向 (β) 发生了变化 (Broadbent & Gregory, 1965)。据此，韦尔福得 (Welford, 1968) 首先提出了警戒衰退的唤醒理论。他认为，在持续长时间的事件环境中，唤醒水平的降低将要导致神经系统整体活动的减少。这样，信号和噪音两者都偏向于减少，而噪音分布和信号加噪音分布的分离程度保持不变，其结果是既降低击中概率 $P(H)$ ，同时也降低了虚报概率 $P(FA)$ 。这样，操作者之敏感性 (d') 不变，只是反应偏向 (β) 增高。

韦尔福得的唤醒理论受到一些事实的支持。例如，随时间延续而发生的唤醒降低的生理表现与反应偏向 (β) 变化相一致；提高唤醒水平的药物能降低警戒衰退现象；提供背景噪音或任何新异的刺激都能降低警戒衰退。然而，唤醒理论无法解释所有的反应偏向 (β) 变化，某些变量显然影响了唤醒水平，但并未对反应偏向 (β) 产生预期的影响。例如，提高事件频率，能提高唤醒水平，但结果反而增加了警戒衰退。

贝克 (Baker, 1961) 提出的期望理论，将警戒衰退归因于操作者对觉察到的 (由此产生期望) 信号概率的反应——调整反应偏向 (增加反应偏向 (β))。实验室中观察到的警戒衰退至少可以部分归因于实验前的训练：实验开始前通常以较高的频率呈现信号刺激以便让操作者熟悉需检测的信号。然后，信号概率降低到警戒状态下的低频条件。这样，操作者的警戒衰退仅仅反应了操作者试图建立最佳反应偏向 (β) 的努力：当信号概率低时，反应偏向 (β) 应提高。

但是，实验前训练的影响也同样无法解释所有随时间而发生的反应偏向 (β) 变化。因此，期望理论进一步提出，操作者是根据觉察到的主观信号概率 $P_s(S)$ 来确定反应偏向 (β) 的。这样，只要信号因任何原因发生漏检，主观信号概率 $P_s(S)$ 将会降低。因为操作者相信，刚才发生的是非目标事件；主观信号概率 $P_s(S)$ 又引起反应偏向 (β) 的上升，也将进一步降低主观信号概率 $P_s(S)$ ，如此循环。柏拉苏拉曼 (Parasuraman, 1971) 称此种循环为“恶性循环”，即反应偏向 (β) 螺旋上升，击中概率 $P(H)$ 螺旋下降。在实际操作中，由于其他因素的影响，反应偏向 (β) 将大致维持在一个较高的水平之上。由于“恶性循环”开始于第一个信号漏检，也就

解释了降低反应敏感性的变量（如短、低强度信号）能增加警戒中期望效应这个事实。

借助信号检测论可以对警戒衰退现象进行更深入的分析，同时，它也能据此提出降低警戒衰退的措施。这些措施的作用在于提高反应敏感性，改善反应偏向（ β ），从而改善了警戒操作。

柏拉苏拉曼对警戒衰退作了归因分析，提出两类措施：改善操作者对信号特征的记忆和降低其记忆负荷的措施，将减少警戒衰退。同时，这将进一步有助于减少因期望因素产生的警戒衰退。有关研究者据此提出了许多可以改善操作者对信号特征记忆的措施。例如，反复训练操作者对目标信号的反应将大大降低警戒衰退。因为反复训练能使操作者形成自动化的刺激加工，而自动化的反应较少依赖于工作记忆。研究发现，当操作者的任务是检测一个目标而不是几个目标中的一个时，警戒操作得到了改善。此外，告诉操作者目标刺激“是什么”的效果，要优于告诉他们目标刺激“不是什么”。例如，对流水线上的质量检验人员来说，提供次品明确的视觉形象要比仅仅要求他们提出“与正品不同的产品”效果好得多。第二类措施为降低记忆负荷的措施，又称为信号增强技术（signal enhancement techniques）。若仅仅简单地增强信号事件的物理能量，往往难以收到改善操作的效果。因为，信号增强的同时，噪音强度也增加了，结果并未得到改善。因此，信号增强技术经常赋予信号事件以区别于噪音事件的特征，这通常可以通过改进信号呈现系统来达到。例如，在雷达监测任务中，若信号事件的呈现方式与噪音事件的呈现方式非常相似，监测任务将非常困难，但若雷达系统在接收到目标信号时的呈现方式有显著区别（如亮点有独特的运动方式），则操作将大为改善。有些信号增强技术甚至采用转换感觉通道的方法，例如，在工业系统的故障监测中，一旦系统发生故障，原先稳定的视觉呈现（信号灯）改为闪烁方式，同时信号转变为听觉呈现（警铃），这将大大提高信号被检测的概率。

从反应偏向这一角度同样可以提出改善警戒操作的措施。一般来说，能够提高唤醒水平的措施有助于降低警戒衰退。这些技术甚至包括生物反馈技术。贝蒂等人（Beattg et al., 1974）曾报告说，训练操作者抑制 θ 波（the-ta wave）（频率为 3~7 赫的脑电波，指示低唤醒水平）亦将减少警戒衰退。

贝克指出，在可能的情况下引入假信号将使反应偏向（ β ）维持在较低水平。因为引入假信号能将主观信号概率 $P_S(S)$ 提高，同时也能提高唤醒水平。更进一步，若假信号与真实信号物理上非常相似，它还能恢复操作者对信号特征的记忆，这将减少由 d' 降低引起的警戒衰退。例如，在流水线质量检测任务中，混入若干“预定的”次品（即使未被检验员发现最后也会被检出）将使检验员的主观信号概率 $P_S(S)$ 维持在较高水平，从而降低反应偏向（ β ）。当然，引入假信号的方法存在一定的危险，首先，操作者绝不能意识到存在假信号，若人们意识到存在假信号，他们将试图避免检测这些假信号，结果使反应偏向（ β ）反而提高。其次，在操作者检验到信号后可能导致严重后果的场合，例如，在工业系统故障监测任务中，虚假的故障信号可能导致操作者将整个系统关机，这将会对生产带来一定的损失。

从上述介绍中可见，信号检测论不仅使我们弄清了警戒下降的原因，而且也使我们找到了防止警戒下降的某些方法，原因和预防措施是紧密相联的。对此，我们可作出如下归纳：

第一，警戒下降可单纯地是由反应标准的提高而引起，表现为操作者报告某个刺激是否为信号变得愈来愈慎重。在上述信号检测论模型图上，报告轴向右移，即反应偏向（ β ）值变大，这时，操作者的正确检测概率和虚惊概率都随时间而下降。众多的研究表明，警戒的下降大多数表现为反应偏向的提高。

第二，警戒下降也可单纯地由感觉敏感性的下降而引起。这种下降具体表现为操作者愈来愈不能把信号从噪音中分辨出来。在信号检测论模型图上，可看到信号分布和噪音分布之间距离变小，即敏感性（ d' ）值变小。在此情况下，操作者的击中概率随操作时间而不断下降，而且虚惊概率则保持原水平。

第三，警戒下降还可由感觉敏感性和反应指标的双重变化而引起。这种情况一般表现为操作者愈来愈不能把信号从噪音中分辨出来。同时操作者报告一个刺激是否是信号也变得愈来愈慎重。在信号检测论模型图上，一方面报告标准轴向右移，即反应偏向（ β ）值变大；另一方面信号和噪音二分布之间的距离缩小，即敏感性（ d' ）值变小。当然，这两个因素所起的作用是不同的而多变的。

三、在认知研究中的应用

认知领域是信号检测论的重要应用领域。就感知研究而言，由古典心理物理学测得的阈限是被试辨别力和反应偏向两者的混合。尽管古典心理物理学采取了许多措施试图消除动机、态度等因素的影响，但仍然无法从根本上解决问题。信号检测论对辨别力和反应偏向作出明确的区分，从而可以深化有关的研究。

除了前面提到的感知觉问题外，信号检测论亦可应用于再记忆研究。在再记忆中，被试所面临的操作，实际上是检测当前刺激（可能识记过，也可能未识记过，即既可能是信号，也可能是噪音），将它同记忆痕迹进行“匹配”，作出“是”或“不是”的反应，这一操作可以看作典型的信号检测问题。信号检测论应用于再记忆研究同样具有很多优点。例如，在再认能力最佳年龄的一项研究（杨治良，1980）中，信号检测论分析指出，在不同指导语条件下再认分数的变化不是因为被试者辨别力的改变，而是由于反应偏向的变化所致，这是用传统方法分析难以得到的；其次，信号检测论将错误反应区分为漏报与虚警两类，有利于进一步分析错误原因。此外，信号检测论的敏感性（ d' ）指标比再认法的指标（对—错）/N 更加灵敏。

信号检测论对实际生活中的记忆问题也可提供很多帮助。一个有趣的例子是见证人提供证词中的记忆问题。在这一任务中，嫌疑者（可能参与犯罪即为信号，也可能未参与犯罪即为噪音）由见证人进行辨认，作出“是”或“不是”判断。这既是一个再认问题，同时也可以看作是信号检测问题。

我们不妨对这一实例进行分析。在见证人问题中，由于维护法律的公正和维护社会治安等动机因素的参与，见证人的反应偏向值往往偏高（漏检很少，罪犯很少逃脱，但清白者无故受冤的可能就大了），或偏低（漏检很多，罪犯就有可能逃脱，但清白者无故受冤的可能就小）。信号检测论对见证人问题的直接应用是见证人在若干名嫌疑者中进行辨认这一情况：见证人在5到6个人组成的“队列”中进行辨认，其中一人被警察认为是嫌疑犯。这时见证人的决策实际上分成了两个阶段：（1）嫌疑者是否包括在这些当

中？(2)若包括在其中，是哪一个？见证人在回答第一个问题时，所持的反应标准较宽，通常作出肯定回答。许多研究证实，对一个短暂的突发事件来说，见证人的视觉再认记忆可靠性相当低，因此，操纵见证人辨认的“队列”，将能影响他们的反应偏向。

从信号检测论分析，有如下三点可供实际工作者参考：(1)在肯定回答中表现出很大信心的见证人（“我可肯定是他”）实际上比那些不太肯定的见证人敏感性更低；(2)告诉见证人嫌疑者有可能不在“队列”中。这一简单措施能使反应偏向（ ）接近最佳反应偏向；(3)“队列”应由相似的人组成，尽管这样相似性将略微降低击中概率 $P(H)$ ，但它能在更大程度上降低虚报概率，其结果在敏感性上得到净收益。

综上所述，信号检测论的最大优点是它可以把操作者的感觉敏感性和反应偏向分开，这一区分使它不仅为实验心理学研究提供了新的研究方法，且为应用心理学中的许多领域提供了更加有效的分析工具，它能使分析更加深入，结论更可靠。今天，信号检测论无论在理论上还是在实践应用上，都显示了它的优越性。

四、评价

信号检测论能把被试者的辨别力（敏感性）和他的反应偏向分开，而以传统心理物理法测得的感觉阈限则是两者不可分割的混合物。信号检测论的引入，确实解决了一些传统心理物理法所不能解决的问题。

在全面评价新、旧心理物理法时，我们也应看到传统心理物理法为了有效地测量感受性，曾试图通过各种手段以消除动机、态度一类因素的影响。这些手段主要有：

1. 通过实验设计和操作抵消可能造成的误差 例如：在极限法的测量中，由于上升和下降的序列的影响，被试可能引起“期待”和“习惯”误差。因此，实验者在实验操作程序上，安排交叉的递增或递减法，以抵消序列效应。

2. 采用合理的数据处理尺度 例如：在上升和下降序列中，得到递增组和递减组的辨别极限 L_x 和 L_1 ，避免它们任何一个作为阈限的尺度，而是采用它们之差的一半。此外，还确定常误的概念，以表示误差的方向和量值。

3. 发挥指导语的控制作用 传统心理物理测量对于统一恒定的判断标准是十分重视的。为了减少某些不肯定性所造成的误差，在判断标准方面必须有一个统一明确的要求，尽量删去那些模糊的标准。例如，在常定刺激法中，和标准刺激等量就是属于难以一致的判断标准。因此，人们常常去掉它，仅保持大于或小于标准两类反应。另一方面，三类判断中的“相等”判断的次数多少，在一定程度上反应被试者的反应偏向和态度。

4. 进行严格而有效的训练，淘汰不合格的被试者。

5. 插入一些捕捉实验，以考察被试的态度 例如，在一个测量声音信号的响度绝对阈限的实验中，捕捉实验可以是一个没有声音信号出现的测试。若被试者对此回答说听到声音，则主试就判定此被试者没有很好地注意任务，故对这部分结果另行处理。反应时实验中的“抢先”反应也是类似的情况。

其实，传统心理物理法所用的手段远不止以上五种。总之，通过上述手段，使传统心理物理学的测量，在某种程度上获得一定的成功。许多资料如

视觉和听觉领域的心理测量函数，迄今还保持着相当的生命力。甚至，信号检测论也起异曲同工作用。

然而，传统心理物理学对于人的主观态度一类的影响依然没有得到根本的解决。在它们的测量中，共同的特点是只出现有效的信号刺激和作用相应的百分比，而无噪音的测量，这样就给它的科学性和可靠性带来了很大的限制。在动物的心理物理研究中，常遇到一个困难是关于基线的评定问题。从阈限的确定来考虑，基线需要固定，但实际上，这是不可能的。所以，对基线变量的评定反应出信号检测论对传统方法的一个发展，克服了传统方法的不足之处。

同时，我们常常看到，用传统方法获得的结果，有时令人误解，甚至是相反对立。例如：曾经对正常男女和胖的男女对饥饿感觉阈限进行了测定。结论说，胖男子阈限最低，正常人居中，而胖女人最高。但是，有研究者用信号检测论方法重新对此精心研究。他们让男女胖子吞下胃球，并记录胃的收缩情形。把胃的收缩看作信号，无收缩看作噪音。在整个过程的4小时里，不断询问被试是否感到饥饿。结果发现，正常男女在胃收缩时，报告饥饿，无收缩时则报告不饿。可是，男女胖子只表明相反的偏好。女胖子倾向于说“不饿”，尽管她的胃已收缩；男胖子则倾向于说“饿”。在感受性曲线(ROC)上，前者表现为左边更低，后者表现为右边更高。可见不同的方法获得的结果不同。

传统方法把被试者的反应偏好混为阈限的一部分，这种情形并不少见。又如：在个性和知觉实验中，被试的反应标准和他的动机是密切相关的，使用有关性行为作刺激时，可能系统地低估观察者的感受性。因为，人们不愿意报告这样的感觉。同样，精神病患者也常常采取较严的反应标准。因为，他可能产生把错误反应误认作是妄想的论据。在疼痛的研究中，也可能把哀伤叫苦的人误定为低阈限，而坚韧的人定为高阈限。所有这些测量明显地反映出传统方法的局限性和信号检测论的优越性。

传统方法得到的一些混乱的结果，有时直接表现为结果的对立。人们习惯地把它归结为实验条件、方法的不同。实际上，根本的原因是无法把感受性的测量和被试的动机、态度所造成的反应偏向区分开来。

在方法学上，传统方法的缺点是缺乏对噪音的测量。实际上，人的辨别力乃是信号及噪音的函数，也就是说，信号固然是信息，但噪音也是信息，这绝不可忽视。信号检测论很好地解决了这两类变量的测量，它明确地建立起反映噪音变量的虚惊概率，这是十分重要的。它不仅对于辨别力的估量不可缺少，而且对于被试主观态度或反应偏好的测量更具有重要意义。因此，信号检测论由于建立击中和虚惊两个经验尺度，在感性的测量上，把主观态度区分开来，以获得成功，这是心理物理法的一个重大突破。

传统方法所得结果，人们归结于实验条件和方法。而信号检测论认为，实验条件和方法的不同所以会造成不同的结果，多半是由于主观的反应偏向造成的。若去掉反应偏向的影响，人的纯感受性或辨别力就恒定了。

可见，信号检测论在重视对信号的测量的同时也强调了对“噪音”的测量的重要性。它借着辨别力和反应标准（或反应偏好）的指数，独立地处理了感受性和主观态度两类变量，无疑对传统心理物理法是一个发展，显示了其优越性。

归结到一点，信号检测论对心理物理法的最大贡献是它可以把被试者的

辨别力和他的反应偏向性分开，而以传统心理物理法测得的感觉阈限则是二者不可分割的混合物。不过，我们一方面要看到信号检测论对心理物理学研究的巨大贡献，要利用这个有利的工具去研究解决一些传统心理物理法所不能解决的问题；另一方面，也要看到传统心理物理法在一定条件下仍不失为测定感受性的一种可靠的方法，并且用起来比信号检测论要简便些。

本章实验

一、信号检测论用于记忆的实验

(一) 目的：了解信号检测论可用于再认实验，学会计算信号检测论指标 (d' , C , β) 的方法。

(二) 材料：实验用三组再认材料，第一组材料为具体实物图形新、旧图片各 25 张，并且有一定的对称性。如旧图片有一张“萝卜”，新图片就有一张“青菜”与它对应；第二组材料为抽象图形，新、旧图片各 25 张，旧图片如“”，新图片则为“”；第三组材料为词，新、旧词也是各 25 张，如旧词有“台北”，新词有“台南”。

(三) 程序：

(1) 实验前先给被试者讲解指导语二遍，指导语内容为：“今天请大家来做一个记忆实验。我们先给你们看一组图片（词），请大家记住它。看过一遍之后，我们把这看过的旧图片（词）和另外一组没有看过的新图片（词）混在一起，以后，我们每呈示一张图片，就要你们在纸上填上‘新’或‘旧’。实验时要求大家专心、安静、独立思考，不要讨论。答不出时，可猜，但不能不答”。

(2) 讲完指导语之后，然后用 6 张图片（新、旧各 3 张，此图片不用于正式实验）进行实验，使被试者明确实验要求和方法。练习做完后，就开始正式实验。先做第一组材料，把 25 张旧图片逐一呈现，每张呈现时间为 5 秒，间隔 2 秒。呈现一遍后，就把这看过的旧图片（词）和另外一组没有看过的新图片（词）混在一起，共 50 张，再向被试者一一呈示，每呈示一张，要求被试者在记录纸上填上“新”或“旧”。第一组材料做毕，被试者休息 2 到 3 分钟后，再开始做第二组材料。做完第二组材料后，再按同样方式做第三组材料。

(四) 结果：

整理具体图形、抽象符号和词这三种材料的结果，分别根据书中公式计算其 d' 、 C 、 β 。

(五) 分析和讨论：

(1) 比较具体图形、抽象符号和词三种材料的不同结果，并用信号检测论分析之。

(2) 评论信号检测论用于再认实验的优缺点。实验用纸

一、记录表

1			11			21			31			41		
2			12			22			32			42		
3			13			23			33			43		
4			14			24			34			44		
5			15			25			35			45		
6			16			26			36			46		
7			17			27			37			47		
8			18			28			38			48		
9			19			29			39			49		
10			20			30			40			50		

二、统计表

S \ R	新		旧	
	次数	%	次数	%
新				
旧				

二、接受者操作特性曲线的实验

(一) 目的：掌握信号检测论的基本理论，学会计算信号检测论指标 d' 、 C 、 β ；学习绘制接受者操作特性曲线 (ROC)，了解信号检测论的用途。

(二) 材料：(1) 从《三国演义》(或《水浒传》) 连环画中选出画页 500 张，分成五组，每组 100 张。

(2) 实物投影仪。

(三) 程序：

实验时，5 组画页的先定概率分别是 0.1, 0.3, 0.5, 0.7 和 0.9。对于每一组画页，主试使用一种信号的先定概率，然后按此先定概率在实物投影仪上呈示给被试者一定数量的画页，要求被试者把它们当信号记住。例如，先定概率为 0.1 时，当作信号的画页是 100×0.1 ，即 10 张，当作噪音的画页为 90 张。作为信号的画页呈现完毕后，再与此组作为噪音的画页混合，然后随机地逐张呈示给被试者。这时，每呈示一张画页，即要求被试者判断此画页是信号还是噪音，并要求被试者把答案记录在实验记录纸上。一组画页实验完成后，稍休息一会，随后，改变信号的先定概率，用另一组画页按上述方法继续实验。

(四) 结果：

(1) 整理 5 组连环画页的实验结果，将结果列出在各种先定概率下的矩阵图，计算击中概率和虚惊概率。

(2) 根据击中概率和虚惊概率，计算其 d' 、 C 、 β 。

(3) 根据上述计算结果，绘制 ROC 曲线图。

(五) 分析和讨论：

(1) 分析信号检测论的两个独立指标。

(2) 分析 ROC 曲线。

(3) 评论信号检测论用于再认实验的优缺点。实验用纸

序	0.1*	0.3	0.5	0.7	0.9	序	0.1	0.3	0.5	0.7	0.9
1						51					
2						52					
3						53					
4						54					
5						55					
6						56					
7						57					
8						58					
9						59					
10						60					
11						61					
12						62					
13						63					
14						64					
15						65					
16						66					
17						67					
18						68					
19						69					
20						70					
21						71					
22						72					
23						73					
24						74					
25						75					

续表

序	0.1*	0.3	0.5	0.7	0.9	序	0.1	0.3	0.5	0.7	0.9
26						76					
27						77					
28						78					
29						79					
30						80					
31						81					
32						82					
33						83					
34						84					
35						85					
36						86					
37						87					
38						88					
39						89					
40						90					
41						91					
42						92					
43						93					
44						94					
45						95					
46						96					
47						97					
48						98					
49						99					
50						100					

*先定概率

本章摘要

1. 现代心理物理学从对阈限概念的不同理解出发，把被试者的辨别力和其反应偏向分开处理，而传统心理物理学测得的阈限值是二者不可分割的混合物。这是现代心理物理学对阈限概念和感觉理论的发展。

2. 神经量子理论是一种由史蒂文斯等人提出的理解阈限的理论。他们在响度和音高的辨别实验中，推论其基本神经过程是按全或无规律进行的。神经量子理论假定反应刺激变化过程的神经结构在机能上能被分为各个单元或量子。

3. 史蒂文斯根据多年的研究，于 1957 年提出了刺激强度与感觉量之间关系的幂定律，即 $S = bI^a$ ，表示心理量 S 是物理量 I 的幂函数。 b 是由量表单位决定的常数， a 是由感觉道和刺激强度决定的幂指数。

4. 物理相关论是由瓦伦在 50 年代末提出来的。这种理论认为，被试者作出感觉量的判断时，实际上是通过过去的经验对与刺激相联系的某种物理属性作出判断。

5. 信号检测论简称 SDT，是信息论的一个分支，研究的对象是信息传输系统中信号的接收部分，它最早用于通信工程，即借助于数学的形式描述接受者在某一观察时间将混有噪音的信号从噪音中辨别出来。

6. 由于人的感官、中枢分析综合过程可看作一个信息处理系统，通常把刺激变量看作信号，从刺激中的随机物理变化或感知处理信息中随机变化看作噪音，这样，人作为一个接收者对刺激的分辨问题便等效于一个在噪音中检测信号的问题。

7. 在信号检测实验中，被试者对有无信号出现的判定可以有四种结果：击中、虚惊、漏报和正确拒斥。

8. 信号检测论认为阈限不是一个纯净感知能力的指标，它受利害得失、动机、态度、意志等因素的影响。这样，在同一过程中就出现了两个独立的指标：一个是感觉敏感性，一个是反应偏向。

9. 用以表示反应偏向的指标之一是似然比值，其定义为给定 X_c 水平上信号分布的纵轴与噪音分布的纵轴之比。

10. 感觉敏感性表现为内部噪音分布与信号分布的分离程度。两者的分离程度越大，敏感性越高；分离程度越小，敏感性越低。

11. 接受者操作特性曲线（ROC 曲线）在心理学上称为感受性曲线。它是以虚惊概率为横轴，击中概率为纵轴所组成的座标图和被试者在特定刺激条件下由于采用不同的判断标准得出的不同结果画出的曲线。

12. 传统心理物理学公认，提示能明显提高痛阈。但是，用信号检测论对同一资料进行分析，则完全否定了这一解释。事实上，被试者感觉辨别力（ d' ）始终没有多大改变，所改变的仅仅是他的痛报告标准而已。

13. 警戒是指操作者在相对较长的一段时间内，对环境偶然出现的某种信号的警觉与做出反应的持续准备状态。信号检测论把感觉敏感性和反应偏向分开处理，有助于说明警戒下降的真正原因。

14. 信号检测论应用于再认研究同样具有很多优点。分析指出，在不同指导语条件下再认分数的变化不是因为被试者辨别力的改变，而是由于反应偏向的变化所致。其次，信号检测论将错误区分为漏报和虚惊二类，有利于进一步分析错误。此外，信号检测论的指标，比传统再认法指标灵敏。

15. 信号检测论对心理物理法的最大贡献是它可以把被试者的辨别力和他的反应偏向分开，而传统心理物理法测得的感觉阈限是二者不可分割的混合物。

建议参考资料

1. 马谋超 (1980) : 感受性的测量。北京市 : 中研院心理研究所 : 《心理学参改资料》第 16 期。
2. 杨国枢、文崇一、吴聪贤、李亦园 (1978) : 社会及行为科学研究法。台北市 : 东华书局。
3. 杨治良 (1983) : 心理实验教学内容更新的初步尝试。心理科学通讯, 第 6 卷第三期, 37 ~ 45 页。
4. 杨治良 (1988) : 心理物理学。兰州市 : 甘肃人民出版社。
5. 杨治良 (1989) : 信号检测论的应用。心理科学通讯, 第 12 卷第三期, 37 ~ 45 页。
6. 张春兴 (1991) : 现代心理学。台北市 : 东华书局 (繁体字版)。上海市 : 上海人民出版社 (1994) (简体字版)。
7. 陈舒永、马谋超 (1979) : 信号检测论与古典的心理物理学方法。心理学报, 第 11 卷第三期, 331 ~ 335 页。
8. 郑昭明 (1993) : 认知心理学。台北市 : 桂冠图书公司。
9. Atkinson, R.C. et al. (1988) .Steven's handbook of experimental psychology (2nd.ed.) . New York: A Wiley - Interscience.
10. Egan, J.P. (1975) .Signal detection theory and ROC analysis. New York : Academic Press.
11. Green ,D.M., & Swets, J.A.(1974) .Signal detection theory and psycho-physics. New York: Robert E. Krieger.
12. Stagner, R. (1988) .A History of psychological theories. New York : Mac-millan.

第六章 视觉实验

本章内容细目

第一节 视觉研究中的变量

一、视觉的物理刺激及其测量	273
(一) 可见光谱	
(二) 单色光的单位	
(三) 光度学	
二、视觉实验中的变量	282
(一) 自变量	
(二) 控制变量	
(三) 因变量	

第二节 视觉研究中的基本实验

一、视觉适应的研究	285
(一) 暗适应	
(二) 明适应	
二、视敏度	290
(一) 视角和视网膜映象	
(二) 视敏度测定及其特征	
(三) 影响视敏度的若干因素	
三、闪光临界融合频率	294
(一) 测量闪光临界融合频率的方法	
(二) 影响闪光临界融合频率的因素	

第三节 颜色视觉

一、视觉的颜色现象	298
(一) 颜色的基本特征	
(二) 颜色混合和混合定律	
(三) 颜色视野和光谱敏感性	
二、颜色的视觉现象	307
(一) 颜色对比	
(二) 颜色适应	
(三) 颜色常性	
三、颜色的标定	311
(一) 孟塞尔颜色系统	
(二) 光源与颜色标定	
(三) 标准色度系统	

第四节 颜色的心理效应

一、色调的心理效应	321
(一) 色调的冷暖感	
(二) 色调的情感	
(三) 色调的环境心理效应	
二、明度的心理效应	323
(一) 明度和谐的心理效应	
(二) 明度对比的心理效应	
三、饱和度的心理效应	325
四、色彩爱好的民族差异	326

本章实验

一、颜色混合实验	328
二、闪光融合实验	330
三、视觉后象实验	331

本章摘要

建议参考资料

在心理学中，感觉 (sensation) 是研究最早和研究较多的领域。我们通过眼睛，可以感觉外界事物的各种颜色，可以辨别明暗，这叫视觉 (visual)。我们通过耳朵可以感觉外界的各种声音，这叫听觉 (auditory)。因此，感觉是直接作用于感官的事物的个别属性的反映。

感觉的种类很多，人们可以按若干种不同的依据对感觉进行分类。在众多的感觉中，视觉和听觉是最为重要的。

在人类的视觉系统中，视觉明显地占主导地位。倘若两个彼此矛盾的信息，一个用视觉接受，另一个用不同的感觉器官接受，这时被试者所反应的—定是视觉信息。这一事实在科拉维塔 (Colavita, 1974) 对视觉和听觉的研究、吉布森 (Gibson, 1933) 关于视觉和触觉反馈的研究以及罗克和维克托 (Rock & Harris, 1967; Rock & Victor, 1964) 的研究中都已得到了证明。这些研究资料表明，被试者不能有效地反应同时经两种不同感觉通道传入的相矛盾的信息，当这种矛盾的信息同时作用于视觉和其他任一感觉时，视觉的反应总是占优势的。因此，研究感觉和知觉现象总是从研究视觉开始的。视觉在人类的感知过程中担负着重要的任务。人类接受的环境信息绝大部分是经视觉通道传入脑中，再作出反应的。人类在生命的早期，视觉系统就开始执行探索世界的重要任务。怀特 (White, 1975) 报告，八个月到三岁的婴幼儿除睡眠外，视觉的探索是最经常的感觉活动。婴幼儿在清醒时，仅注视眼前物象的时间就占 20%，视觉执行的重要任务伴随人的一生。

心理学研究一般把视觉研究的内容分成“视觉”和“颜色视觉”两个系列，它们是相辅相成的。通过本章的学习，旨在回答如下七个问题：

1. 何谓可见光谱，光具有哪三个维度，与此物理相对应的心理特征是什么

么。

2. 单色光的单位有哪些。
3. 实验心理学为什么重视亮度测量。
4. 颜色混合的三条定律是什么。
5. 如何运用 CIE 标准色度图来标定颜色。
6. 暗适应曲线有哪些特征。
7. 何谓视敏度。

第一节 视觉研究中的变量

在生物漫长的进化过程中，动物发展出来各种类型的光感受细胞。光感受细胞先是局限在上皮的一定区域，然后内陷成单眼。此后从单眼进化到复眼，再进化到脊椎动物（包括人类）特有的折射眼，形成了高级视觉机能。视觉的形成，首先与视觉的刺激和人眼的光学系统有关。人眼的适宜刺激是光，光的物理特征以及人眼的生理机制使得心理学研究的视觉现象充满了特异性。在这一节中，我们从介绍刺激的性质开始，再阐述光度学上的有关概念，最后简单地讨论作为视觉实验研究的变量。

一、视觉的物理刺激及其测量

（一）可见光谱

特定的光刺激人眼而引起视觉。光是一种电磁波，引起人类视觉的电磁波称为可见光谱。光是放射性的电磁能。现在认为，光是由特殊的物质粒子（光子）所组成，而光波的特征则用单个光子的波动来表示，光的强度可根据发射源产生的光子数来表示。由于放射性电磁能是以起伏的形式传递的，故光又可用波长（wave length）来表示。波长是两个相继的波的相应点之间的距离。

可见光的波长范围只占整个光谱的极小一部分，电磁波的波长的差异极大，如图 6 - 1。整个光谱范围从最短的宇宙射线（波长仅 1 米的百万亿分之一）到无线电和电力波（其波长可达数十公里）。整个电磁光谱中，人眼所能接受的光波只占一小部分——约七十分之一。在正常情况下，可见光谱的波长范围约 380 毫微米到 780 毫微米之间。毫微米（manometer，缩写 nm）是计算光之波长的单位，等于十亿分之一米。若进入两眼的波是 400 毫微米，我们就感觉到紫色；若光波的波长是在 700 到 780 毫微米之间，我们感觉到的是红色。在可见光谱中段的光波波长以及相应的感觉到的颜色是：470 毫微米是蓝色，610 毫微米是绿色，580 毫微米是黄色。将不同波长的可见光混合起来，可以产生各种不同的色光，而将所有可见的色光混合起来则产生白色光。比 380 毫微米短的电磁波称紫外线，比 780 毫微米长的一端称红外线。

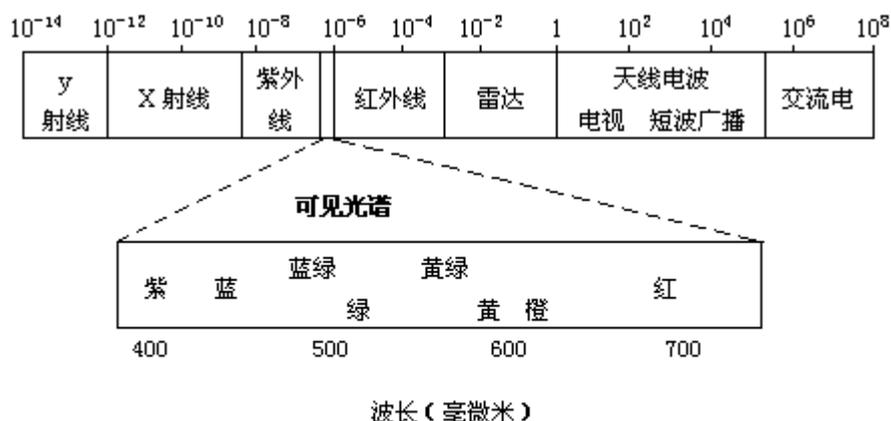


图 6-1 全部电磁光谱

可见光部分是全波长中的一小段，在图的下方予以放大

（采自 Geldard，1972）

光具有三维特征：波长、纯度和振幅。与此相对应的有三维心理特征：

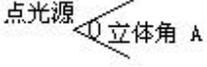
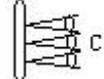
不同波长引起人们不同的色调的感觉；纯度是光波成分的复杂程度，它引起的视觉反应是饱和度；振幅是光的强度或能量单位，它引起的视觉维度是明度。

（二）单色光的单位

光不仅可用波长来度量，也可用纯度、强度来度量。我们把能发光的物体叫做光源（light source）。光的来源很多，有日光、灯光等等。前者是天然光源，后者是人造光源。在日常生活中，我们所看到的光大多是不止一种光波组成的。若投射到我们眼中的光是由单一波长的光组成，那么，这种光称为单色光。由两种以上波长组成的复合光波，称为多色光。在实验室中要获得单色光，我们可以利用三棱镜将白光折射分离，用其中某一种单色光。我们也可以用滤光片来获得某波段的色光。滤光片一般是一片透明的玻璃片或塑料片，它可以吸收一些波长的光而使另一些波长的光透过。透过的波段在视觉上就呈现为此波段的色光。现在有一种仪器叫单色仪（monochromator），可以用它产生单色光。

视觉研究中为了进行度量，常常要碰到一些光的单位。光的单位名目繁多，为了避免引起混乱，我们仅介绍一种国际单位制（SIE 单位制）的光度单位。见表 6 - 1。

表 6-1 国际单位制（SIE 单位制）的光度单位

几何学	光 度 量			
	名称	符号	单位	英文名
 光源	光强度	I	坎德拉	Candela (cd)
 点光源 立体角 A	光照度	E	勒克斯	Lux (lx)
 B	光亮度	L	尼特	Nit
 C	光通量		流明	Lumen (lm)

与心理学关系密切的单色光单位有：

1. 光强度 光强度（luminous intensity）是光源在单位立体角内辐射的光通量，以 I 表示，单位为坎德拉（candela，简称 cd）。1 坎德拉表示在单位立体角内辐射出 1 流明的光通量。

2. 光通量 光通量（luminous flux）是由光源向各个方向射出的光功率，也即每一单位时间射出的光能量，以 Φ 表示，单位为流明（lumen，简称 lm）。

3. 光照度 光照度（illuminance）是从光源照射到单位面积上的光通量，以 E 表示，照度的单位为勒克斯（Lux，简称 lx）。

4. 反射系数 人们观看物体时，总是要借助于反射光，所以要经常用到“反射系数”的概念。反射系数（reflectance factor）是某物体表面的流明数与入射到此表面的流明数之比，以 R 表示。

5. 光亮度 光亮度 (luminance) 是指一个表面的明亮程度, 以 L 表示, 即从一个表面反射出来的光通量。不同物体对光有不同的反射系数或吸收系数。光的强度可用照在平面上的光的总量来度量, 这叫入射光 (incident light) 或照度 (illuminance)。若用从平面反射到眼球中的光量来度量光的强度, 这种光称为反射光 (reflection light) 或亮度 (brightness)。例如, 一般白纸大约吸收入射光量的 20%, 反射光量为 80%; 黑纸只反射入射光量的 3%。所以, 白纸和黑纸在亮度上差异很大。

亮度和照度的关系如图 6-2 (a) 所示, 最常用的照度单位是呎烛光 (footcandle)。1 呎烛光是在距离标准烛光一英尺远的一平方英尺平面上接受的光通量。如果按公制单位, 则以米为标准, 照度就用米烛光 (metrecandle) 来表示, 即 1 米烛光是距离标准烛光一米远的一平方米面积上的照度。1 米烛光等于 0.0929 呎烛光。

从图 6-2 上, 我们不难理解亮度和照度之间的关系, 其关系为:

$$L=R \times E \quad [\text{公式 6-1}]$$

式中 L 为亮度, R 为反射系数, E 为照度。

因此, 当我们知道一个物体表面的反射系数及其表面的照度时, 便可推算出它的亮度。

亮度也有几种度量单位。亮度的单位是用一种理想化了的标准状态来定义的 (如图 6-2b)。以一支标准蜡烛当作光源, 放在一个半径为 1 公尺的球体的中心位置。假设这个蜡烛会均匀发散它的全部光线, 则落在球体内表面一平方公尺表面积上的所有光量为 1 个流明 (lumen)。实际应用中, 亮度单位用流明太小了, 所以通常取其十倍的单位——毫朗伯 (millilambert) 来表示。比毫朗伯稍大的单位是呎朗伯 (footlambert), 1 毫朗伯等于 0.929 呎朗伯。英国标准的呎朗伯是用光源的烛光数、从光源到表面积英尺数和表面的反射率来规定的。在有些国家, 普遍使用的是米制单位, 是以毫朗伯为基础的 [1 毫朗伯 (mL) = 0.929 呎朗伯 (ftL) = 3.183 烛光/平方米 (c/m^2) = 10 阿普熙提 (apostilbs)]。光亮度的单位还有: 坎德拉/平方米 (即尼特, $Nit = 1cd/m^2$) 等。

(三) 光度学

光度学 (photometry) 是一门研究可见光计算和测量的科学。光度学和心理学有特殊的关系。光度学的测量和计算是以心理学对人们的视觉系统感觉特征的研究为基础的, 视觉心理研究则要应用光度学的知识和测量技术。光度测量是一门专门的学问, 它不仅度量单位名称多, 各单位相互换算关系复杂, 而且受测量器械和技术的限制, 不易为人们所掌握。

我们之所以比较详细地描述亮度测量, 是因为对于实验心理学家来说, 规定一个视觉刺激或视觉环境的强度是视觉研究的前提。应该注意到照明值要受从光源到照明表面的距离的影响 (如图 6-2)。但是, 亮度则与从表面到观察者的距离无关。这就是为什么电影银幕从电影院后排看和从电影院的前排看一样亮的原因。因此, 在仪器视野的大小所能包括的范围之内, 可以在距离刺激任何方便的地方对光进行测量。这种亮度与距离无关的情况当然对摄影机和人眼都是适用的。在摄影时, 曝光时间和光圈大小必须以被拍摄物体的亮度而加以调整, 但却不必因物体到摄影机的距离而调整, 这是同样的道理。

光度学度量的原始标准是以特定规格制成的一支蜡烛作光源的。图 6-2

所表明的就是这样的一支蜡烛被放在离屏幕 (s) 的一段距离 (d) 处。屏幕以一定比例 (R) 反射投射到其表面的光。入射光被称作是照射到屏幕上的照度 (E)。由于它是以一支特定的蜡烛为光源，又距离屏幕一定的远近，所以是以呎烛光为单位 (当然在使用公制时也可以用公尺标定距离)，这个值等于光源的烛光数 (c) 除以从光源到屏幕的距离 (d，以英尺或公尺为单位) 的平方。将这几个参数列入一个公式，可以表示如下：

$$E = \frac{c}{d^2} \text{ (呎烛光)} \quad [\text{公式6-2}]$$

如果一个光源 c 移近一个表面 s，则后者受到光源较强的照明；但如果光源远离屏幕，照明便减弱了。从规范上讲，图中对 s 的照明直接随烛光以及距离的平方的倒数而改变，即 $E=c/d^2$ 。一个被试在 c 处所看到的 s 屏幕上的亮度 L，依赖于屏幕的反射率 R，因而 $L=cR/d^2$ 。可见不论距离表面远近，屏幕上的亮度维持不变。

照明工程师确定了不同视觉工作所需要的标准照明水平。例如：一个图书馆的设备应为读者提供较高的照明水平，而街道则只需较低的光照明水平即可。这些可用照度计进行测量。

制定工作场所的照明标准既科学，又具有经济效益。若标准订高了，将导致国家电力的浪费；反之，若订得过低，必然会导致工人视力受损。照明的标准与工作物和背景亮度对比有直接关系。计算对比的公式如下：

$$C = \frac{L_{\text{高}} - L_{\text{低}}}{L_{\text{高}}} \quad [\text{公式6-3}]$$

式中 C 为对比值，L 为亮度。

照明的标准还与对象在眼睛里所形成的视角有关。喻柏林等 (1979) 曾求得视觉对象的视角、对比和照度的关系，大对比如图 6-3，小对比如图 6-4 所示。从两图中的视角与照度的关系曲线来看，无论大对比或小对比都具有共同的变化趋势。视角越大，要求照度越低；视角越小，要求照度越高。视角递减时，照度发生递增变化。而且视角递减速度低于照度递增速度。换言之，照度的一个大的正增量才能适应视角的一个小的负增量。而且，若小视角和大视角在纵坐标上发生同样的变化，则照度在横坐标上分别引起的增值量相差越大，前者大大高于后者，这现象可称为照明收效递减律 (law of diminishing returns)。

表 6 - 2 大对比实验结果 (黑环/白背景视标，观察者 20 人)

照度 Lux 辨认结果	1.7	10	60	360	2160
平均视角(分)	1.42	0.81	0.69	0.58	0.53
(标准误差)	0.11	0.03	0.02	0.02	0.01
视敏度	(0.70)	(1.23)	(1.45)	(1.72)	(1.89)

(采自喻柏林，1979)

表 6-3 小对比实验结果（黑环/白背景视标，观察者 20 人）

照度 Lux	1.7	10	60	360	2160
辨认结果					
平均视角（分）	9.10	3.80	2.70	1.32	0.93
（标准误差）	(0.48)	(0.13)	(0.28)	(0.21)	(0.06)

（采自喻柏林，1979）

在上述大量的实验研究的基础上，我国视觉心理学工作者与建筑工程等专业的研究者协作制定了工业照明标准，可以约略综合在表 6 - 4 中。

根据图 6 - 3 和图 6 - 4 所示原则，我们可以看出视角与照度的关系，从而了解到，在视觉工作物细节大小的分级上，小尺寸的分级应较细，大尺寸的分级可较粗。采用这样的等级划分法，可以有针对性地对小物体细节提高照明，而达到节约能源的目的。表 6 - 4 很清楚地说明了这一点。

与亮度相关的心理量是明度（brightness），但是不要错误地认为明度只依赖于亮度。明度要受包括眼睛的适应状态、对比效应、曝光时间以及知觉恒常性在内的许多因素的影响。心理量表法可用于对主观明度的直接估计。表 6 - 5 列出了各种视觉刺激的亮度值，从这里可以看到眼睛所能感受的亮度的范围很大。表中混合视觉，即间视觉（mesopic vision）。勒格兰德等人（LeGrand et al.，1957）认为在明视和暗视之间应当有间视这一阶段。

表 6 - 4 视觉工作类别、物体细节大小与所需要照明的关系

视觉工作类别	物体细节大小 (毫米)	照度 (勒克斯)
	0.10	784
	0.11	637
	0.12	505
	0.13	377
	0.14	326
	0.15	258
	0.16	234
	0.18	200
	0.20	173
	0.22	152
	0.24	135
	0.26	122
	0.28	110
	0.30	100
	0.32	92
	0.36	79
	0.40	69
	0.44	61
	0.48	54
	0.52	49
	0.56	45
	0.60	41
	0.65	37
	0.75	31
	0.85	27
	0.95	24
	1.05	21
	1.15	19
	1.25	18
	1.35	16
	1.45	15
	1.60	15
	10.00	15

(采自喻柏林, 1979)

表 6 - 5 典型视觉刺激的亮度值

视觉刺激	亮度值 (mL)	
中午的太阳表面	10^{10}	损伤
	10^9	
	10^8	
钨丝灯泡的灯丝	10^7	明视觉
	10^6	
	10^5	
日光下的白纸	10^4	混合视觉
舒适的阅读	10^3	
	10^2	
	10	
	1	
月光下的白纸	10^{-1}	暗视觉
星光下的白纸	10^{-2}	
绝对刺激阈	10^{-3}	
	10^{-4}	
	10^{-5}	
	10^{-6}	

(采自 Riggs , 1965b)

二、视觉实验中的变量

在心理实验的情景中，实验者必须考虑三类变量，即自变量、控制变量和因变量。

(一) 自变量

在第一章，我们曾经介绍过自变量（或自变项）的一般类型。按照这种分类，视觉实验研究中的自变量可以大致归为如下几类：

1. 刺激变量 刺激变量（或刺激变项）（stimulus variable）是指刺激因其本身的不同特胜而产生不同的视觉效果。视觉实验中作为刺激的自变量是多项目多维度的。在项目上，实验者可以选择的刺激有：系列亮度，从 10^{-6} mL 到 10^7 mL；可见光谱中各段的色光，从 380nm 到 780nm。实验者在刺激维度上可以选择的内容有：刺激的时间维度——一组具有不同频率或不同波长的刺激；刺激的空间维度——一组等能光的面积或不同的观视距离。在研究颜色视觉现象时，实验者可根据光和颜色的组成成分对明度、色调、纯度（饱和度）实施三维操纵。但是，作为刺激，研究者必须设法使这些刺激能保持精确的定量。视觉实验对仪器的要求甚高。

2. 背景条件 背景条件也可以作视觉实验的一类自变量，目的在于使标准刺激恒定。背景条件在视觉实验中很明显，它使得可被任意操纵变更的项目内容超出标准刺激范围。例如，前面曾提到过的，研究者可设置大对比和小对比条件等等。

3. 被试者特点 被试变量是指在某一方面，被试者间彼此不同的个体差异。视敏度、辨色力是常见的个体特征。其他如年龄、机体状态在视觉实验中作为自变量都有获得显著意义的证据。此外，还可人为地变更同一被试的某些特征。例如，令被试者用单眼、或遮住鼻侧或颞侧视野，也是视觉实验中常用的自变量内容。

事实上，研究者确定自变量的范围远远超过以上论述的范围。在具体实验中，任何企图选用单一纯光的设想往往都较难做到，视觉研究的特点决定

它的刺激大都以复合形式存在，针对诸如此类的问题，研究者只能结合课题灵活应付，控制一些暂不研究的潜在变量。

（二）控制变量

在心理实验中，控制变量（或控制变项）（controlled variable）是指因为预定的自变量的变化与另一个已知的或潜在的自变量的变化伴随发生，而造成两个甚至两个以上的自变量混淆。在视觉实验中，这样的现象同样可能发生。因此，我们须注意控制变量，避免对自变量的混淆。

在视觉变量中，一个容易忽视的特殊因素是确定刺激变量属于哪类“连续体”。我们已经了解到光亮在感觉上是等级连续体（量的连续体），而色光是非等级连续体（质的连续体），实验时应当注意采用不同的方法，否则实验结果就会发生偏差。

在视觉研究中，有一个较为突出的问题是累积效应的预防。累积效应（accumulative effect）是对人体（或生物）有影响的环境条件或有关因素（如药物等），多次暴露所造成的生物效应的累积或叠加。累积效应通常有三种情况：一种是多次暴露的效应形式的简单相加；一种是形成比简单相加更为重的效应；还有一种是比简单相加更为轻的效果。视觉机能的特点提醒我们要注意刺激时间、面积在视网膜作用点产生的累积效应。此外，视觉研究还要注意刺激强度的变化是属于明视还是暗视阶段，刺激的投射点是在视网膜的中央凹还是边缘。在颜色视觉的研究项目中必须控制的变量似乎更多、更复杂。视觉实验的各项研究，在整个心理实验的范畴中以细微、精巧著称。由于视觉系统在功能上的敏感性加上心理的复合成分，如果各种变量稍有失控，其结果是使研究者试图验明的过程面目全非，而这样的混淆通常是不易被察觉的，结论就难免张冠李戴。

（三）因变量

一般在视觉实验中，刺激变量借助于仪器，因而较为严密。然而，视觉实验在如何定量因变量（或依变项）方面，就不如自变量定量那么精密。在大多数情况下，实验研究者记录到的因变量值相对于严格定量的刺激量值而言，一般是很粗糙的，有的甚至只是一些非量化的判断反应。因此，研究者能否根据因变量推断某种视觉现象发生的过程和特点，首先取决于为保证自变量和因变量之间的对应关系而采取的措施是否有效；其次涉及处理因变量的技术是否适用及合理。刺激—反应的保证措施的关键在于“控制”额外变量，而后的处理工作按照因变量的性质，有的是直接的，有的是间接的。当一项实验的因变量值是定量时，后期处理就能直接进行。例如，有相当部分的视觉实验采用调整法程序，被试者调定的量值一般都能够直接读出，但有的实验只能通过被试作口头报告来表达他的比较判断结果。例如，在颜色后象的研究中，由于被试者根据主观经验得到的现象报告结果，实验者只能将这些报告汇总后进行处理，间接地获得某些数据资料。

由于因变量是由实验者观察或记录的变量，因此，不管实验者用什么方式处理因变量，都必须先作好因变量的记录工作。有时候，实验者要借助计量仪器才能读出被试的反应量，为了防止度量时产生偏差，记录时就要注意操作技术。如用照度计测量入射光，硒光板与光源之间的距离和相对角度都必须限定好，否则，差之毫厘就会导致严重误差。实验者对记录口头报告的结果也要作出限定，一些限定有时候可以明确告诉被试者用规定范围内的词汇报告自己的感觉经验，有时候要防止暗示。限制标准只能由实验者掌握，

其目的是使被试者报告中的各种表达都可能被纳入相应的等级范围以内。这样，一个统一的指导语就十分重要了。

总之，视觉实验的因变量要能达到灵敏反应视觉过程或现象的程度，实验者必须很仔细地设计实验，严格地把握好自变量、控制变量和因变量。

第二节 视觉研究中的基本实验

视觉研究十分重要，而且包括的范围很广，本节介绍视觉研究中的一些基本实验，作为进一步视觉研究工作的基础。

一、视觉适应的研究

视觉适应一般涉及“暗适应”和“明适应”。前面表 6-5 中已介绍了明视觉、暗视觉和间视觉。心理、生理学家将明暗视觉的特征整理成表 6-6。

(一) 暗适应

当我们走进电影院，最初什么也看不见，需经过一段时间才逐渐适应，并能区分周围物体的轮廓。这种对低亮度环境的感受性缓慢提高的过程，称为暗适应 (dark adaptation)。图 6-5 是暗适应曲线。如果用白光测定，会发现暗适应曲线由两部分组成，表示人眼有两套适应机制。若用红光照射中央凹检查感光阈，只得到图中上部的均匀曲线，没有下部的曲线，这是视锥细胞的暗适应，这个过程约五分钟就基本完成。人眼对白光经过七分钟的适应以后，出现进一步的感光阈的减低，这是视杆细胞 (或杆状细胞) (rod)

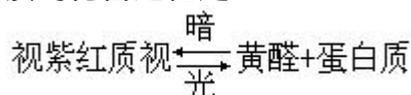
表 6 - 6 人眼睛的明暗视觉特征差异

特征	明视觉	暗视觉
感觉细胞	锥体细胞	杆体细胞
光化学物质	锥体色素	视紫红质
色觉	正常的三色	无色
所在视网膜区域	中心	外围
暗适应速度	快 (8 分钟或更少)	慢 (30 分钟或更多)
空间分辨能力	高	低
时间辨别	反应快	反应慢
照明水平	昼光 (1 到 10^7 毫朗伯)	夜光 (10^{-6} 到 1 毫朗伯)
空间总合	小	大光谱
灵敏峰值	555nm*	505nm

*nm : nanometer , 毫微米 (10^{-9} 米) (采自林仲贤 , 1987)

的暗适应，约 20 分钟基本完成。视杆细胞的暗适应出现慢，但适应程度很高。关于暗适应机制的解释，主要是化学反应说。克劳福德 (Crawford, 1947) 用分光描记法确定视杆细胞中的视紫红质的化学反应过程是暗适应过程的机制。视紫红质 (visual purple) 是一种化学感光物质，在曝光时被破坏变色，在暗适应中又重新合成而恢复活性。

克劳福德表述视紫红质的化合过程是：



在光刺激时，视紫红质发生了分解而退色，变为视黄质（视黄醛+蛋白质）。光刺激继续作用，视黄质再分解退色，变为视白质（维生素 A+蛋白质）。因此，眼睛受到的光刺激时间愈长，视紫红质分解就愈彻底，反之，暗适应时视紫红质循原路线重新合成的时间愈长，完成暗适应时间就较慢。

外界条件会影响暗适应过程。布兰查德（Blanchard, 1966）研究了适应前的照明作用，图 6-6 是他所得的结果，可以看出，暗适应前的视野亮度愈高，视觉感受性就愈低，但经过相当长的时间之后，视觉感受性都基本趋于一致。

伴随暗适应的还有瞳孔大小的变化。从明亮处进入黑暗处，在 10 秒钟瞳孔扩大到最大直径的三分之二，达到完全扩大约需要 5 分钟。这个过程中，瞳孔的直径由 2 毫米扩大到 8 毫米，进入眼球的光线增加到 10 到 20 倍。

除了上述因素外，影响暗适应的因素还与机体有关。例如，在视网膜（retina）（即眼球最内一层）的不同部位测试暗适应过程，它们反映出的结果有所不同，缺乏维生素 A 会引起暗适应机制紊乱。另外，年龄也是影响暗适应的因素，以 30 岁为界，30 岁后暗适应的感受性就逐渐降低。在 20 世纪 50 年代，克拉夫科夫（Kpabkob, 1954）研究认为，人长期禁食缺乏营养后，光感受性只达到正常状态时的十分之一到十五分之一。他还发现一只眼睛对光的感受性可以因另一只眼睛受到各种有关刺激而发生变化，如用白光刺激一只眼睛可使另一只眼睛的敏感性提高，而如用红光刺激一只眼睛可使另一只眼睛的敏感性降低，这也是同一感受器内部的相互作用。

暗适应机制使人类有能力在更大范围内适应环境的明暗变化。同时，暗适应机制的认识能使人类运用这些规律更好地适应环境。根据贾德（Judd, 1951）和沃尔德（Wald, 1945）的研究资料（图 6-7）表明，在几种不同的适应色光中，只有红色光暗适应保持得最好。而且，尤其在波长为 600 毫微米以上的视场中，视杆细胞的感受性比视锥细胞（或锥体细胞）（cone）的感受性高很多。二次世界大战期间，迈尔斯（Miles, 1944）曾经特制一种红色护目镜，戴上这种红色护目镜后既能使视锥细胞在明暗视场中都有较高的感受性，又能使视杆细胞相对地不受视场光线的变化而保持暗适应状态。之后，一些暗室用红灯照明，也是运用同样的道理。

（二）明适应

人从暗处到亮处，眼睛大约经过一分钟就能适应，这是明适应（或亮适应）（light adaptation）。明适应时，眼的感受性不是提高，而是降低，与暗适应正好相反。布兰查德（Blanchard 1931）用阈限法揭示：视杆细胞在极端黑暗转入极亮的条件下，其感受性下降 100 万倍。赖特（Wright, 1934）用间接方法求得光适应曲线，发现中央凹的光适应过程很快，它暴露在光线中一分钟后就几乎全部完成。

眼睛适应光强度变化的范围很大，这个范围约达到 13 个对数单位，大约要比最弱的绝对阈限的光强一万亿倍，见表 6-7。

在光适应过程中，眼睛首先通过调节瞳孔大小来适应光线刺激的强弱变化。光量的增加，瞳孔在 3 至 4 秒钟内就能迅速缩小以保护视网膜，免使过强光线对它的损伤。与此同时，视杆细胞作用转到视锥细胞作用。瞳孔

表 6 - 7 不同亮度适应时瞳孔直径、面积的变化

适应的视场亮度 (坎德拉/平方米)	瞳孔直径 (毫米)	瞳孔面积 (平方毫米)
10^{-5}	8.17	52.2
10^{-3}	7.80	47.8
10^{-2}	7.44	43.4
10^{-1}	6.72	35.4
1	5.66	25.1
10	4.32	14.6
10^2	3.04	7.25
10^3	3.32	4.23
2×10^4	2.24	3.94

(采自林仲贤, 1987)的放大和缩小是调节的第一道关口,它的大小根据进入眼睛的光线强度的变化,表 6-7 列出不同视场亮度下,眼睛进行适应时的瞳孔直径、瞳孔面积的平均值。

但是仅仅凭借缩小了的瞳孔还无法适应高强度的光。研究者发现,在视网膜的外层还有许多黑色颗粒,它们是一些具有保护作用的物质,能减少直接作用于感光细胞的光能量。所以,遇上强光刺激,人们会保护性地闭上眼睛,或戴上太阳镜,使眼睛逐渐适应光照水平的变化。

间视觉是介于暗视觉和明视觉之间的一个视觉阶段。研究者一般认为,当光亮达到 10^{-3} 烛光/平方米以上时,视锥细胞便被激发,这是间视觉的表现。间视的上限是视杆细胞的饱和,它随着视野(visual field)(指在眼不转头不摇的情形下目光所见及的广阔面)的大小而改变,但还不能真正地确定下来。现在的研究表明,间视阶段是视锥细胞和视杆细胞相互作用的阶段,不能简单地理解为两种细胞的简单混合。

二、视敏度

视敏度(visual acuity)是指分辨物体细节和轮廓的能力,是人眼正确分辨物体的最小维度。视敏度通常以找出两个物体之间的最小间隔来表示,它受物体的网膜映象、照明等因素的制约。下面我们就分别阐述与此有关的四个问题。

(一) 视角和网膜映象

人眼好比一架照相机,从图 6-8 可见,物体 AB 通过节点 n,形成物体的视网膜映象。对象与眼睛所成的张角,叫做视角。视角的大小决定映象在视网膜上投射的大小。图 6 - 8 是视角和网膜映象的关系图。An 和 Bn 在节点所成的夹角(AnB)称为视角(visual angle),它与眼到物体的距离成反比。视网膜映象是倒置的,但经大脑皮层的调整,我们感受到的还是正的像。

计算视角时,可以把物体大小 AB 近似作为圆周的弧,把从节点至物体的距离 Bn 作为圆周的半径,用下面简便的方法计算,即

$$= \frac{A}{D} \times 57.3 (\text{度})$$

: 视角大小(单位是度、分、秒,本式 的单位是度)

(1 弧度=57.3 度)

A：物体的大小

D：距离

例如，某人的身高为 1.8 米，在 6 米远处，我们看到他时所形成的视角为：

$$= \frac{1.8}{6} \times 57.3 = 17.19^\circ$$

由于 AnB 和 anb 相等，利用相似三角形原理，计算视网膜映象的大小，可有等比公式：

$$\frac{AB}{ab} = \frac{An}{an}$$

代入上例得出的数据：

因为： $AB=1.8$ ， $An=6.0$ ， $an=0.015$

所以： $x = ab = 0.0045$ 米=4.5 毫米

即：此人在观察者的视网膜映象上的大小为 4.5 毫米。

(二) 视敏度测定及其特征

医学界用视力表测定视敏度，它是以视角的倒数来表达的，其公式为：

$$V(\text{视敏度}) = \frac{1}{(\text{视角})}$$

但是，检查视敏度的方法有好几种。下面的图示列出了检查视敏度的几种常用的刺激项目，它们可分别测试受测者对物象的觉察 (detection)、再认 (recognition)、解象 (resolution) 和定位 (localization) 能力 (图 6-9)。

医学上用的视力表属于认知测试，通常以 5 米为标准观视距离。在图 6-9 中，认知类测试的“ ”型视标是国际通用的视标，称为蓝道环 (Landolt ring) 视标。它的规格结构标准是：黑线条宽度为直径的五分之一，环的开口也是直径的五分之一，即等于线条宽度。我国目前采用“E”型视表来检查视敏度，它的黑线条宽度和空白区宽度也是全字的五分之一。当视标“C”的直径为 7.5 毫米时，环的开口是 1.5 毫米 (见图 6-10)。在 5 米距离处，被试者如能指出缺口方向，他的视角约等于 1 分。1 分视角时的视网膜映象约等于 0.004 毫米，这个数字接近单个视锥细胞的直径。医学上一般认为 1.0 的视敏度是正常的。医学上正是根据这一原理制成视力检查表，能看清第一行为 0.1，看清第二行为 0.2，看清第十行正是视角等于 1 分。在日常生活中，有的人具有很高的视敏度，而有的人视敏度较低，有的国家规定视敏度为 0.05 就是“法定盲人”，即残疾人标准。

从图 6-9 上我们除了看到属于认知的材料之外，还有觉察、解象和定位三种材料。觉察测试采用一条线或一个点，见图 6-9，只需让被试者判断在一个视野里刺激是否存在即可。解象采用等宽黑白条纹图案进行测试，以可分辨的最佳栅条为视敏度指标，一般以空间频率来表示。在最佳条件下成人最小可辨敏度是 45 到 60 (周/度)，对应于二分之一到三分之二的条纹宽度。定位用游标测定视敏度，是让被试者辨别某一根线对应于另一根线的位移，被试者须区别上面的线是下面一条线的左边或右边，在最佳条件下成人刚好区别的位移约为 2 弧秒，相当于中央凹 (或中央窝) (fovea) 中光感受细胞直径的十分之一。

（三）影响视敏度的若干因素

在上面的讨论中，我们已经了解到距离对视敏度的影响呈现一种反比关系，即距离增加，视敏度下降。影响视敏度的因素除了上述的物体大小和距离之外，主要还有：

1. 不同亮度会影响视敏度 亮度增加，则视敏度增加，两者关系是对数关系。这在第一节中已作了讨论。

2. 物体与背景之间的对比度不同，视敏度将受到影响 当物体与背景之间的对比度加大时，则视敏度提高；反之，视敏度降低。

3. 视网膜不同部位的视敏度也不同 因为锥体细胞对细节分辨起主要作用，所以，在视网膜中央凹处（即锥体细胞集中之处）视敏度最大。

4. 视觉的适应影响视敏度 暗适应时眼睛的视敏度不如明适应时的视敏度高，这是因为视杆细胞与视锥细胞在功能上不同的结果。克雷克(Craik, 1939)让被试者被适应有一定照明的空旷视野，然后转向另一视野，并要求被试者报告看到的是两条平行黑线还是一条黑线，结果发现，在眼睛适应的空旷视野和试验视野照明条件大致相同的水平时视敏度最高。

5. 闪光盲会降低视敏度 在明适应的条件下，突然的强光刺激会暂时降低视敏度，这种现象称为闪光盲(flash blindness)。闪光盲持续的时间长短与闪光强度、曝光时间、照射的视网膜部位、目标大小、瞳孔和眼的适应状态都有关系。闪光盲也许是视觉功能的保护性抑制，但是过强的闪光可能造成永久性损伤。在不同的工作场合，闪光盲会危及机体安全或导致事故发生。

6. 练习可以大大提高对目标物的视敏度。

此外，影响视敏度的因素还有不少，这里就不一一列举了。

三、闪光临界融合频率

在18世纪，有人发现了视觉图象的暂留现象。当时把一块炽热的燃料绑在带子的一端，在黑暗中加速转圈，当达到一定转速时，光点就变成连续的光圈。如果用电筒来试，也能达到同样的结果。若要计算视觉图像暂留时间，只要知道看到光圈时转动速度就行。这样，一个间歇频率较低的光刺激作用于我们眼睛时，就会产生一种一亮一暗的闪烁感觉，随着光的刺激的间歇频率逐渐增大，闪烁现象就会消失。由粗闪变成细闪，当每分钟闪光的次数增加到一定程度时，人眼就不再感到是闪光而感到是一个完全稳定的或连续的光。这一现象称闪光的融合(flicker fusion)。闪烁刚刚达到融合时的光刺激间歇的频率称为闪光临界融合频率(critical flicker frequency, 简称CFF)。闪光临界融合频率是人眼对光刺激时间分辨能力的指标，是物理刺激与生理心理机能相互作用的结果，是受刺激的时空因素以及机体状态制约的感觉过程。不同人的CFF的差异相当大，但一般人的临界频率为30~55赫，这个数据也告诉我们差异达一倍左右。

这种时滞的存在对于我们知觉物体是一种优点。若我们的眼睛在时间上具有完全的分辨能力，那么我们在现代的交流电的灯光下，任何物体都将显得闪烁了。例如，电影画面每秒钟放映24幅，这个频率对许多人而言远达不到临界，为了避免闪烁，就得通过一幅画面连续闪烁三次。这样，虽然每秒仅24幅画面，但人们受到的刺激速率却是每秒72次的闪光，因而我们看到的就不再是闪烁的光。

（一）测量闪光临界融合频率的方法

闪光临界融合频率最早是用制成扇形的圆盘在光源前旋转来测定的。顾名思义，称之为转盘闪烁方法（rotation disc flicker method）（参见图 6 - 11）。

转盘闪烁方法在测试时由被试者控制转速，旋转慢时，可以看到间断的闪光，但是达到一定速度就可以感到连续的光亮，即闪光临界融合频率。这种闪烁还可以测量闪光强度。用转盘闪烁方法测量 CFF 的缺点是，由于光源来自外部，光源即使照射到黑的部分也会有光反射出来，因此，亮度控制较差，转速的频率测量有时也不太准确。

电子技术的发展已使闪光临界融合频率的测定有了更完善的仪器。用现代的电子仪器，实验者可以随意呈现由不同电脉冲组成的刺激。这些刺激的波形可以是脉冲波、方波、正弦波、锯齿波和三角波等。通过改变电信号的波幅，就可以改变电信号对光信号进行调制时的图形亮度，改变电信号的周期就可以获得图形呈现的不同频率，改变电信号相位就可以改变图形中黑白部分的比例。实际证明，用这些方法能够精确控制亮度、频率和亮度间隔，而且结果是稳定的，功能也多样化。例如，光源也可用不同的色光等等。

（二）影响闪光临界融合频率的因素

闪光临界融合频率是很复杂的现象，目前其生理机制还在探索之中，但它是一种较有用的视觉生理指标。我国一些研究者曾采用闪光临界融合频率作为视觉疲劳指标进行研究，指出，在海拔 2000 米 ~ 4000 米高的路上驾车，经 6 小时工作后，闪光临界融合频率将明显下降。同时它也是电影、电视放映的一个重要参数。当其他条件相同时，若闪光临界融合频率越高，就表明眼睛对于时间上的明暗变化的分析能力越强；也可以说，对时间的视敏度越好。

影响闪光临界融合频率的因素很多，主要有以下几种：

1. 闪光临界融合频率随光相的强度增高而增高 闪光在时间和强度上可分为二相，一为暗相，一为光相。假如暗相强度为零，则闪光临界融合频率和光相强度的对数成正比，其数学式为：

$$n = a \lg I + b$$

n：闪光临界融合频率

I：光相的强度

a, b：参数

此式称为费里-波特律（Ferry - Porter law）。此律和费希纳定律在形式上有相同之处，这种相似性表明闪光临界融合频率也是一种强度关系，但也只适合于中等强度，当光相的强度太大或太小，此公式就不适用了。闪光临界融合频率在低光强度时，可低至 5Hz；在高强度时，可高至 50 ~ 55Hz。

2. 刺激面积 小面积的闪光临界融合频率比大面积的闪光临界融合频率来得低。闪光临界融合频率随闪光照射的区域面积的增大而增大。和上述随强度的增加而提高一样，二者也是有同样的对数关系：

$$n = C \lg A + d$$

A：面积

C, d：参数

大面积具有较高的闪光临界融合频率，这一事实也是空间累积的进一步证明。空间累积的效果可以用这样一个方法来表明：用四个小点同时闪亮，测定融合频率，然后再测定一个点闪光时的融合频率。四点比一点的闪光临

界融合频率要高，这是由于四个小点同时闪亮增大了刺激面积。

3. 在视网膜中，杆体细胞和锥体细胞的闪光临界融合频率是不同的。总的说来，当刺激区域小时，闪光临界融合频率在中央凹比边缘高，这表明锥状细胞比杆状细胞有较高的空间视觉敏感度。

另外，一些附加刺激的作用，如声音、味觉、嗅觉等刺激都可以改变闪光临界融合频率。一些材料表明，年龄、疲劳、缺氧等因素都影响到闪光临界融合频率。55岁以上的人的闪光临界融合频率相对较低，视觉疲劳及缺氧也会降低闪光临界融合频率。

第三节 颜色视觉

我们生活在颜色的海洋中，昂首可望蔚蓝的天空，环顾四周可见草绿花红。颜色不仅装扮了大自然，也极大地丰富了人眼对客观世界的认识能力，色彩成为人类生活的必需。彩色电影、彩色电视能再现大自然的无限美妙的色彩，画家画出大自然奇妙的景象都需要颜色。色觉是视觉的基本机能。但人眼很少看见单纯的只有一种波长的光波，绝大多数情况下都是不同波长的光波混合起来的色光。

颜色是物体的一种属性，是由于光投射到物体，根据物质的性质，反射出没有被吸收的光的特性，并作用于我们的视觉而引起感觉的结果。人类认识颜色的本质最早是由牛顿（Sir Isaac Newton，1642～1727）的研究开始的。牛顿于1704年发表了《光学》以后，研究色度的工作就从浅到深由表及里地发展，终于形成了完整的理论体系。下面我们先从视觉的颜色现象入手展开讨论。

一、视觉的颜色现象

颜色的基本特征是认识颜色现象的基础，尽管颜色现象包含的内容非常广泛，但心理学家正是从研究颜色的基本特征入手，开始研究五彩缤纷的颜色世界的。

（一）颜色的基本特征

颜色可分为两大类：非彩色和彩色。非彩色是指从黑色到白色，由深浅不同的灰色组成的系列，这个系列的梯度可以用一条垂直线来表示，见图6-12。非彩色系列是无色系列，基本特征主要是明度。非彩色系列各梯度色没有绝对的纯度指标，系列中的各梯度色的非彩色反射率代表物体的明度，反射率越高越接近白色；反射率越低，则越接近黑色。一般地说，白色表面的反射率达80%左右，而黑色表面的反射率小于10%。由于人的视觉在明亮的白天和昏暗的夜晚是由两种不同的细胞进行工作的，这样二种感光细胞对明暗光的敏感程度不同，所以选择视觉刺激要考虑这些因素。

视觉感受一种颜色取决于三个特性，即亮度、色调和饱和度。任何一种颜色都是由三者总效果的结果。亮度（brightness）是彩色和非彩色所共有的属性，它是指作用于物体的光线的反射系数，它同光能的强度密切相关。强度越大，反射系数越大，颜色就越亮，最后成白色；反之，强度越小，反射系数越小，颜色就越暗，最后成黑色。色调和饱和度是彩色独有的特征。色调（或色别）（hue）是由物体表面反射的光线中什么波长占优势所决定的。饱和度（saturation）是色调的表现程度，它是指同一色调的两种颜色，哪一种含颜色较多或较少，它决定于物体所发射出来的光线中规定其色调的波

长占多少优势。实验证明：三个特征中若其中之一发生改变，颜色就起了变化。若两个颜色的三个特征相同，那么不论它们的分光组成如何，在视觉上总是产生同样的色感觉。

在物体反射的光线中，占优势的光波波长决定颜色感觉，这是最本质的颜色属性，见表 6 - 8。颜色的饱和度是指一个颜色的纯洁性，它取决于表面反射光波波长范围的大小，即光波的“纯度”。光谱上的各种颜色是最饱和的颜色。颜色中掺入白、灰或黑色越多，它就越不饱和。

表 6-8 光的波长，频率对人眼所产生的色觉

颜色	波长 (毫微米)	频率 (赫 × 10 ¹⁴)
紫	400	7.5
	450	6.7
蓝	480	6.2
蓝绿	500	6.0
绿	540	5.6
黄绿	570	5.3
黄	600	5.0
橙	630	4.8
红	750	4.0

(采自林仲贤，1987)

颜色的三个基本特征，可用一个锥体来形象地表示这三个维度的关系。锥体内的各点表示色调，半径表示饱和度，纵轴表示明度。颜色锥体的垂直部分，表示单一色调在饱和度和明度上的区别，见图 (6-12)。

了解了颜色的基本特征，我们能进一步去认识外部的颜色世界。白光来自于太阳中所有光谱色的混合。当然把阳光分解成彩虹时，我们实际上能够从其中看到可见光谱中的所有色彩，光谱是按红、橙、黄、绿、青、蓝、紫的固定顺序排列的。这种不变的顺序可以解释波长的不同长度，红光的波长最长，而紫光波长最短。这种差别也可以说明这样的一个事实，红光比其他色彩的光的传递距离更远。

当光源变化时，色彩也发生变化。在拂晓和黄昏，太阳的位置离地球表面很低，因此，它的光线必然更多地穿透地球表面的大气层。大气中的灰尘颗粒和水汽阻挡了波长较短的蓝光，在它们能够到达地面之前把它们驱散。结果这时候的自然光缺乏蓝色光线，因此天色显得较红。反之，晴朗天空的直射光可以使更多的蓝色光线到达地面。就人造光来说，白炽灯泡散发更多的黄色射线，而荧光灯则使橙色变为蓝色。水银灯用于公路照明，因此，减少红光，增强蓝光和绿光。

在我们的环境里物体能够显示颜色，这是因为它们吸收一些光线，反射另一些光线。例如，成熟的西红柿除了红色之外吸收所有的光线，它把红光波长反射给我们的眼睛。西红柿的藤蔓主要吸收红色、橙色和紫色射线，反射蓝色、黄色和大量的绿色射线，因此，呈现绿色。物体的表面色彩随同它的视觉色彩形成一种综合色调，它保留在光束之中反射给我们的眼睛，这取决于物体吸收或消除某种波长的能力。

任何量的色彩都能够通过物体表面反射出来，主要是根据光源中这种色

彩的比例。在白炽光下，由于具有大量的黄光，西红柿呈橙红色。在晴朗的白天，西红柿的红色偏蓝，是因为有大量的蓝色射线到达它的表面。在水银灯光下，由于极其缺乏红光，可怜西红柿从光束中吸收了所有其他颜色的射线，结果，几乎变成了黑色。服装色彩在光源的变化中特别受到影响。白天购买的一件淡紫色上衣在黄色的人造光下可能会变为淡灰色。

所有色彩系统都是根据彩虹来分段。也就是说，一系列颜色是按固定顺序排列的。各种色彩系统之间的基本差别在于每种系统所指定的原色稍有不同。普遍使用的原色包括红、黄、蓝三色。彩色印刷程序使用品红、黄、青三原色，而彩色电视的原色是红、绿、蓝。在我们的色彩研究中我们将使用孟塞尔颜色系统，因为它对限定颜色特性提供了精确的尺度。孟塞尔颜色系统将在本节加以讨论（参见图 6 - 19 和图 6 - 20）。

（二）颜色混合和混合定律

颜色混合（或混色）（color mixture）涉及两大法则，一是满足色光混合的加色法，二是符合颜色混合的减色法。导致这两种混合方向相反的原因主要是由于材料的物理属性不同。

1. 光谱中的色光混合是一种加色法 加色法（或相加混色）（additive mixture）的原色是红、绿、蓝，它们的波长分布属于可见光谱的两端和中部。用三架幻灯机同时投射这三色光并使之重叠在一起时，我们看到的是白色光。但除非是在光学实验室里，一般情况下产生的白色光并不是纯粹的三原色。事实上，我们见到的许多颜色大都是不同光波混合的结果，所谓同色异谱的现象说的就是这个道理。人眼不是非常精细的感觉器官。例如，当我们面对一个波长是 570 毫微米的黄色光，同时有将 650 毫微米的红光和 530 毫微米的绿光按一定比例合成另一个黄光，我们人眼是感觉不出这两种黄光有什么差别的。于是我们认识到光谱中的每一种色光，都有另一种按比例与它混合得到一种白色的色光，它们都能在色环和色三角上找到大致的关系，这种关系构成了补色对。

加色法的结果可用简单的式子表示如下：

红色+绿色=黄色
红色+蓝色=紫色
蓝色+绿色=青色
红色+蓝色+绿色=白色

红、绿、蓝色称为基色，而青、紫、黄色分别称为它们的相应的补色，这是因为，每一补色以适当的比例加上其相应的基色即可得到白色，例如黄色和蓝色混合即得白色（参见图 6 - 13）。

2. 用颜料、油漆等的混合配色是一种方法 它与色光混合不一样，例如混合黄色和蓝色是颜料吸收了一定波长的光线后所余下的光线的色调。例如黄色颜料是从入射的白光中吸收蓝光而反射红光及绿光，而这两种光合在一起引起黄色的感觉。

减色法（或相减混色）（subtractive mixture）的三原色是黄、青、

紫，它们是加色法三原色的补色（complementary color）。彩色电视主要是应用加色法，即彩色光在显象管光屏上组合是相加混合的结果。而彩色电影的画面则由黄、青、品红三种影片染料按减色法处理构成的。参见图 6 - 14 可得如下结果：

黄色=白色-绿色
紫色=白色-红色
黄色+紫色=白色-蓝色-红色=绿色
紫色+青色=白色-绿色-红色=蓝色

相加混色和相减混色得出的明度也不一样，在加色法混合后产生的颜色其明度是增加的，等于其投射的光束的明度的总和；而在减色法中混合后得出的颜色，其明度是减少的。

3. 混色定律 各混色光对视觉器官的作用是相加的过程，而且，遵循以下三条规律，即相加混色三定律。这三条定律具体内容是：

（1）补色律：补色律（law of complementary colors）是指每一种颜色都有另一种与它相混合而产生白色或灰色，这两种颜色称为互补色。牛顿将光谱色按照它们自然的秩序排成圆周，在红与紫的两端之间留一个弧度给非光谱色的绛色，再回到光谱中的红色，这样就排成一个色圈（见图 6 - 13）。若把各种色之间的间隔排列适当，如圆周上任何直径相对的两端颜色以适当的比例用色轮转动来混合，它们会消失各自的色彩而变成灰色。我们把直径相对的两色称之为补色。

（2）居间律：居间律（law of intermediary）是指混合色圈上两个非互补的颜色产生介于这两种颜色之间的中间色。例如，我们将红色与黄色进行混合便可得到介于这两色之间的橙色。中间色的色彩取决于两者的比例，若红与绿混合，按混合的比例不同，可以得到介于它们之间的橙、黄、黄绿等各种颜色。中间色的饱和度一般是较低的，它的饱和度与两色之间在色圈上的距离成反比，与两色原来的饱和度成正比。根据居间律的原理，我们只要确定出三种基本的颜色（即红、绿、蓝），逐一将其中两者进行混合，即可得出光谱上的各种颜色。

（3）代替律：代替律（law of substitution）是一条很主要的定律，混合色的颜色混合不随被混合的颜色的光谱的光谱成分而转移。不同颜色混合后产生相同的颜色可以彼此相互代替。例如：

颜色 A=颜色 B；颜色 C=颜色 D
则 $A + C = B + D$

代替律表明，只要在感觉上颜色是相似的，便可以互相代替而得到同样的视觉效果，尽管它们二者的光谱成分是不同的。例如：

$A + B = C$

若 $X + Y = B$ ，则 $A + X + Y = C$

根据这个定律，可以充分利用各种颜色混合，进行代数式的相加和相减得出所需要的各种颜色。

以上是颜色混合的三条定律，其他较复杂的问题也都是从这里出发的。对于混合规律的研究，不仅揭示了颜色混合的定律，而且在鉴定各类色盲、色弱患者上有临床意义。

4. 混合颜色的方法 混合颜色的方法很多，这里介绍二种常用的简易方法。一种采用一套已知透光率的良好滤色片，透光率不同的滤色片可以得到光谱中各种单色，然后，把它们同时投射在不反光的白色屏幕或视网膜的同一部位上，这是一种方便又对精度没有什么影响的做法。另一种混合方法更简单，用色轮（见图 6 - 15）就可以。色轮（或混色轮）（color wheel）乃是一个由不同颜色扇形所组成的圆盘，套在旋转器的轴上，在转速超过闪光临界频率时（约 30 转/秒），即产生一种均匀的混合色。混合色的性质决定于每种被混合的色纸显露部分的比例。但这种方法用颜色纸来配色，颜色纸反射的往往不是一种单色，因此，混合出来的颜色是很不饱和的。例如混合普通的红色和绿色得到黄色。但用这种方法所得到的的是灰黄色（棕色、黄褐色）。另外，用色轮混合时，混合色的亮度不像色光混合时的亮度相加，而是介于混合成分亮度的中间。除此之外，用色轮混色也是遵循颜色混合的三条规律。

（三）颜色视野和光谱敏感性

人眼注视外界景物，大约有近似 2° 的直径范围能获得最清晰的感觉，称为中央凹视觉（foveal visual）；在中央视觉周围的区域只能获得模糊的视觉景象，这是边缘视觉（peripheral visual）；人眼的中央凹视觉和边缘视觉用视野计（perimeter）能够准确地测量。正常人双眼同时注视一景物时，视野大约有 120° 左右是重叠的，双眼视野比单眼视野的范围要大。正常人的视野大约每只眼睛上下垂直视野可达 135° 到 140° 左右，水平视野可达 150° 到 160° 左右。各人的视野也有一定的个体差异。

有色视野与无色视野的范围是不同的。有色视野中不同的颜色所占的范围也不同。在同一光亮条件下，白色视野的范围最大，其次是黄蓝色，再次是红色，绿色视野最小（图 6-16）。我国研究人员测定中国人的有色视野所得的结果是：白色视野分别为上- 45.4° ，下- 63.6° ，向内- 61.7° ，向外 71.6° ；绿色的视野分别为 30.6° ， 42.9° ， 41.4° ， 51.9° 。参见图 6-16，图中心附近的小圆圈为盲点（blind spot）。

为什么不同颜色有不同的视野呢？这是由于视网膜的中央凹部位和边缘的结构不同。视网膜中央区能分别出各种颜色，由中央区向外周部分过渡，对颜色的分辨能力减弱，人眼感觉到的颜色的饱和度降低，最后直到色觉消失。通常，对于中等亮度的刺激，任何人的网膜边缘都看不到颜色，一切有颜色的物体在视网膜边缘区域呈现不同明暗的灰色。另外，视网膜中央部位有一层黄色素，它能降低光谱短波（如蓝色）的感受性。不同人种的黄色素的密度有所不同，并且黄色随年龄的增加而变化，年龄大的人水晶体（lens）变黄，黄色素增多，因此，不同年龄的人的颜色感受性会有所不同。在光亮处，等能光谱最亮的部分在 556 毫微米（黄绿光带）处。逐渐减少光亮度，光谱颜色越来越暗，以至看不出颜色，全部光谱呈不同程度的灰色。随着环境亮度的逐步降低，光谱最亮部分向左移，最亮部分为 510 毫微米（蓝绿光）处，而原来的红色部分则看不见了。所以在白天，黄绿色的物体显得最亮，而在黄昏时，蓝绿色显得较亮，红色不明显。这种现象称为浦肯野现象（Purkinje phenomenon）。这种现象只有当光照射视网膜边缘部分时才会出现。由此可见，中央凹视觉（视锥细胞）有辨色能力，边缘视觉（视杆细胞）

不能辨色。浦肯野现象是在光线转暗时，色觉由视锥细胞向视杆细胞转移的结果。这就进一步解释了为什么颜色感受视野范围比较狭小（参见图 6 - 17）。

二、颜色的视觉现象

要了解正常人的眼睛在光亮条件下是如何感觉颜色的，有些什么特点，掌握比较典型的视觉现象是十分必要的。下面介绍三种视觉现象。

（一）颜色对比

颜色对比（color contrast）是两种不同的色光同时作用于视网膜的相邻区域，或者相继作用于视网膜的同一区域时，颜色视觉所发生的变化。前者是同时对比现象，后者是继时对比现象。例如，注视黄色背景上的一小块灰色纸片几分钟，你就会感觉到灰色的纸片呈蓝色；同理，在绿色背景上灰色纸片会呈红色，这是同时对比（simultaneous contrast）现象。若在灰色背景上放一块颜色纸片，注视短时间后再撤走纸片；或先注视颜色纸片，再插入灰色背景，你就会在背景上看到原来颜色的补色。这是继时对比（或连续对比）（successive contrast）。无论是同时对比还是继时对比，它们主要是色调对比。

颜色对比现象还有其他方面的表现，例如，同一种颜色在亮的背景上看起来不如在暗的背景上看起来亮些，这是明度对比现象。依黑林（EwaldHering，1834~1918）的看法，明度对比产生的现象是一个视网膜区的活动引起临近区域相反活动的结果。他仿用谢灵顿（Sherrington，1897）“对比盘”（见图 6 - 18）证明，当黑白盘的转速达到闪光临界融合频率时，黑白明度对比效果就提高了。然而，色彩明暗对比的程度有助于整体外表的有效表现，其道理是很清楚的。

由于我们很少孤立地观察一种颜色，因此理解各种颜色接近时如何相互影响就非常重要。例如，当红色被黄色所包围时，红色会偏于光谱中的紫色的一侧；若被蓝色所包围，红色就会偏于橙色。这一简单的现象被称为同时对比：当两种不同的颜色开始直接接触，它们之间的差别就会增强。它们往往相互排斥。它们的明度、色调和饱和度的变化程度将会更大。

在实际生活中，常有这样的例子，邻接的色调在色谱中相互排斥。金发的人穿上鲜艳的黄色上衣会使头带绿色；天蓝色靠近绿色会呈现蓝色，就像红色靠近紫色会呈蓝色。

（二）颜色适应

在黑暗中经过较长的时间，视网膜的感受性会发生变化，这是一种适应现象。注视一个红色纸片半分钟，然后注视灰色背景，色觉会发生逆转，这就是一种适应。色适应与色对比有时很难划分，但研究者通常把先看到的色光对后看到的色光的影响叫作色适应。我们用实验的方法来表征这种现象。以每秒钟一次的黄色小闪光投射在注视点上，观察者注视红色强光视野，待适应后再回头看原来的黄色闪光。开始闪光变成了绿色，经过一段时间后

逐渐增加了黄色的感觉成分，几分钟后，才完全恢复黄色闪光的感觉。这就是典型的颜色适应 (vision adaptation) 现象。

黑尔森 (Helson, 1948) 曾做过一系列颜色适应实验，其中有一项是这样的：实验在暗室进行，照明是红色的，几分钟适应后，实验者请被试者判断一套从黑到白的 19 件标本，并要求他根据熟悉的评定标准为这些标本的色调、明度和饱和度作等级排列。结果，凡与墙背景的反射率相近的标本都被判断为红色，反射率愈高则被认为饱和度愈高。而比墙颜色深的标本被认为是绿色或蓝绿色，即红色照明的后象 (注 6-1) 补色；反射率愈低，蓝绿色显得愈饱和。

其后，麦克洛 (McCullough, 1965) 的研究也证明了这一视觉现象。他首先把色适应与图形方位结合在一起研究，他发现在让眼睛交替适应蓝背景上的水平栅条和橙色背景上的垂直栅条之后，紧接着让被试看无色背景上的同样黑白栅条，被试者便把水平栅条看成橙色，而把垂直栅条看成蓝色。这是一项典型的例子。

在生活中也有许多这样的例子，若你凝视红色，然而把视线移到白色，你就会看到淡淡的蓝绿色 (红色的互补色)。若你在任何其他鲜艳的颜色上重复这个试验，你都会发现一种互补色的残留影象。当视觉从鲜艳的橙色毛衣上移向淡灰色的裤子，就会附加上蓝色的残留影象，使裤子呈现蓝灰色。鲜艳的品红色上衣会使白色的皮肤产生绿色。

当一种强烈的色调和它最不相同的颜色——它的互补色 (complementary color) 结合时，就会附加上一系列的残留影象，这种增加的强度使人难以接受。这可以说明为什么我们的眼睛难以接受强烈的互补色的组合。当注 6-1：后象 (afterimage) 指刺激消失后而感觉暂留的现象。后象按性质不同而分为两种：一为正后象 (positive afterimage)，其特征是刺激消失后所遗留的后象与原刺激的色彩或明度相似，仍有短暂光色的视觉现象留存即属之。另一种为负后象 (negative afterimage)，其特征是后象的明度与原刺激相反，而色彩与原刺激互补 (见张春兴, 1989)。一种或两种互补色减弱时，振动的感觉也会减少。在服饰中，对应互补色的强化难以获得美观的效果。

目前，色适应的成因机制尚待研究。有一种解释是，这种现象的出现是因为人类眼睛的神经末梢对强烈的颜色很快产生疲劳；对于这种颜色的敏感减退了，视觉中接受的互补色倒成为主要的了。换言之，眼睛从强烈的颜色中为自己提供救援。有研究者发现，对一只眼睛建立强色光适应后会影响到另一只眼睛色觉的现象，使人们相信颜色适应是在较高的视觉系统水平上产生的效应。

(三) 颜色常性

人眼对物体颜色的感知，在外界条件变化的时候，仍能保持相对不变，表现出颜色常性 (或颜色恒常性) (color constancy)。这对人类适应环境的过程具有重要意义。试想，若缺乏这种能力，我们的感觉完全遵循实际的物理刺激，那么一块在阳光下的煤就应是白亮的，一支白色粉笔在阴影里是灰黑的了，因为实测发现在阳光下煤的明度 (反射系数) 比阴影里的粉笔的明度大得多。若是这样，我们的日常生活将陷于混乱，甚至无法生存。

究竟是什么原因使我们的视觉系统具有颜色常性的呢？历史上有过两种主要的解释，赫尔姆霍茨（Hermann Von Helmholtz, 1821 ~ 1894）认为这是一种“建立在无意识感觉上的理智判断”。而黑林不同意这种观点，他提出“颜色记忆”的概念。迄今为止，两种观点中哪一种更好些还尚不知晓，事实上，它们都有经不起检验的地方。然而自卡茨（Katz, 1911）把颜色常性问题引入心理实验室，并集中研究照明变化对于常性知觉的影响以来，对颜色常性的研究前进了一步。

布伦斯维克（Brunswik, 1929）提出一个简便的方法计算常性：

$$K = (R - S) / (A - S)$$

K：布伦斯维克比率（以百分数表示）
A：标准制刺激的反射率
S：按标准刺激计算的配对反射率
R：实验匹配的反射率

式中 K 表示布伦斯维克比率，比率为 0 时就表示没有常性；比率为 1 表示有完全的常性，一般用百分数表示。颜色常性实验可选定反射率、色调、明度和饱和度中的任一项指标。以反射率为例，现在令被试者用两组灰色进行一次颜色匹配，其中一块作为标准刺激，它的反射率 A=40%，照度条件 10 呎烛光。若我们要求被试者在 50 呎烛光的照度条件下，从比较刺激中选出一块灰色与标准刺激相匹配，可能存在的匹配方式有三种：（1）按照亮度匹配，被试者应当选择只有标准刺激照度的五分之一的比较刺激，即反射率是 8% 的灰色。（2）按刺激的反射率配对，那就应当选 R=40% 的灰色。（3）但实际上，被试既不按第一种方式选择也没有按第二种方式选择，而是选出反射率 R=24% 的灰色。因此，在本例中被试的常性应为：

$$K = (24 - 8) / (40 - 8) = 50\%$$

邵勒斯（Thouless, 1931）使用反射率的对数值进行计算，其效果更好些，他改进了布伦斯维克的公式。

$$\text{邵勒斯} = (\lg 24 - \lg 8) \times (\lg 40 - \lg 8) = 68\%$$

比例用布伦斯维克公式计算，得出常性系数为 50%，用邵勒斯公式计算，常性系数为 68%。常性作为知觉的一个特性，我们在第八章中还将进行专门讨论。

三、颜色的标定

标定颜色乃是一项专业性很强的工作，我们在此只能介绍有关颜色标定（color notation）的最基础的知识。

（一）孟塞尔颜色系统

孟塞尔颜色系统（Munsell color system）是根据美国美术教师和画家孟塞尔（Albert H. Munsell, 1858 ~ 1918）提出的颜色排列方案标色的方法。它用色调、明度、饱和度的测量标度来标色，这三个量分别对应于主波长、明度和强度（或纯度）。

自 1900 年，孟塞尔开始研究“色标测色法”，1915 年由美国出版的《孟塞尔颜色图谱》（Munsell Book of Color），就是孟塞尔长期研究的成果。这个图谱是从心理学出发自成体系的测色标准。经美国国家标准局和美国光学学会修订后的《孟塞尔颜色图谱》的最新版本包括两套样品：一套是有光泽样品，包括 1450 块颜色样品，37 块由黑到中性样品；另外一套是无光泽样品，包括 1150 块颜色样品，32 块中性色样品，每一块样品的大小为 1.8×2.1 平方厘米。国际染色或涂色表面的不透色就是用孟塞尔颜色系统来标定的。

孟塞尔标色系统用颜色立体的架构来表示（图 6 - 19）。立体的中央轴代表无彩色系列中性明度等级，从白到黑有 11 个感觉等距级，中间（第五级）正好是中灰。孟塞尔颜色系统的饱和度也按感觉分成许多相等的等级，从中央轴的 0 饱和度出发，离开中央轴越远，彩色饱和度越高，个别最饱和颜色的饱和度可达 20。围绕中央轴的立体平面中，色调可以用 10 种方向表示。根据图 6 - 19 所示，10 种色调中有 5 种主要色调，即红 R，黄 Y，绿 G，蓝 B，和紫 P；另外五种是中间色调：黄红 YR、绿黄 GY、蓝绿 BG、紫蓝 PB、红紫 RP。再细分，对每种色调又有 2.5、5、7.5、10 四个等级，其中 5 是主要色调。

用孟塞尔颜色系统测色时，测色结果用三个符号来表示：色调（符号 H）、明度（符号 V）、饱和度（符号 C）。现有一样品测色的结果为 $H=10P, V=5, C=8$ ，就写成 $10P5/8$ 。先写色调，然后明度，明度后画一斜线再写饱和度：
 $HV/C = \text{色调} \times \text{明度值} / \text{饱和度}$

孟塞尔色环（Munsell color circle）（见图 6 - 20）分成十种色调。十种主要色调的每一种的标度为 5。每种色彩再细分成中间色，标度为 1 到 10。以黄色为例，色环中黄色靠近红色的一侧标度较低（1Y, 2Y, 3Y），靠绿色的一侧标度较高（8Y, 9Y, 10Y）。更细划分可以用小数表示（2.5Y, 7.5Y, 9.5Y 等等）。色调划分几乎是无限的。

非彩色的黑白系列的中性色（neutral color）用 N 表示，N 后写明度值，斜线画在 NV 后，线后不写彩色。例如，明度值等于 5 的中性灰可以写作 $N5/0$ ，这时，对于彩色低于 0.3 的黑、灰、白色通常标定为中性色。若考虑标定，亦可采用如下形式：

$$NV / (H, C) = \text{中性色明度值} / (\text{色调}, \text{饱和度})$$

但是，在这种情况下，色调只用 5 种中间色中的一种，不作更细的标定。例如，对于一个带点黄色的浅灰色写成 $N8 / (Y.022)$ 。

（二）光源与颜色标定

度量颜色首先要弄清楚是在什么样的光源照明下进行的。这是因为测量颜色涉及人的观察，而不同光源对颜色视觉有不同的影响。人们体验到，在暗室红灯照明下，只有明暗感觉，看不出颜色。公路上用钠灯照明，虽然很亮，但人们看到的颜色失真了。为此，国际照明委员会规定了标准照明体和标准光源，用色温为度量单位，以统一测色标准。

色温（color temperature）是表达一个光源颜色的方法。它是光源颜色的简称。我们曾谈到形容颜色的词语并不能对颜色，特别是对饱和度加以清楚的描绘。而“色温”这个概念就可以使我们省去许多语言来描述颜色的

特征。色温并不是指颜色的温度，而是指什么样的物体加热到什么样的温度条件下，产生什么样的色光。我们日常接触到的颜色大量是表面色，即非自发光的物体色。另外还有自发光体色，称光源色（color of light source）。光源色主要是由光源的光谱能量来决定的。光谱能量的分布与色温有关，色温高则能量分布偏于短波一端。光源的发光强度常为大家所重视，而光源的色温却常被忽视。

那么，色温能不能用数值来表示呢？色温是黑体在不同温度上辐射光的颜色，用绝对温度（absolute temperature）（在摄氏温度上加上 273° 来表示）代表其数值，在此数值后附以 °K 的记号。例如，一块黑铁慢慢地加热，黑铁的颜色由原来的黑逐渐变成红、黄、白，最后变成青蓝色。若其光源辐射光谱分布与绝对温度 $T^{\circ}\text{K}$ 的绝对黑体辐射光谱相同，则绝对黑体的绝对温度 $T^{\circ}\text{K}$ 就是这个光源的色温，此时绝对黑体的颜色与光源的颜色相同。人们用黑体加热后所发出的不同颜色表示一个光源的颜色。

有了色温值就能客观表达光源的颜色了。“光源”是指能发光的物理辐射体，如太阳、灯光等。“照明体”是指特定的光谱功率分布，这一分布并不是一定得由一个光源直接提供，甚至不一定能用光源来实现。国际照明委员会规定 A、B、C、D 四种标准光源用于测色，见表 6 - 9。

表 6 - 9 标准照明体 A, B, C, D_{65} 的色度坐标和相应色温

标准照明体		A	B	C	D_{65}
色 度 坐 标	x	0.4476	0.3484	0.3101	0.3127
	y	0.4074	0.3516	0.3162	0.3290
	x_{10}	0.4512	0.3498	0.3104	0.3138
	y_{10}	0.4059	0.3527	0.3191	0.3310
	u	0.2560	0.2137	0.2009	0.1978
	v	0.3495	0.3234	0.3073	0.3122
	u_{10}	0.2590	0.2142	0.2000	0.1979
	v_{10}	0.3495	0.3239	0.3084	0.3130
相应色温		2856K	4870K	6770K	6500K

（采自林仲贤，1987）

表 6 - 9 四种标准光源中，A 光源可用色温为 2856K 的充气钨丝灯泡为代表；B 光源是在 A 光源基础上加上一组特定的滤光器产生相关色温 4870K 的辐射，代表中午平均直射阳光（黄色成分稍多的日光）；C 光源是在 A 光源基础上加上另一组特定的滤光器，产生相关色温 6770K 辐射的光源，它近似淡云天空的反射光。过去人们认为 C 光源是较好的光源，但近年来认为 B 光源和 C 光源都不是日光的最好代表，而推出 D_{65} 光源代表日光。在实验室模拟出 D_{65} 的光谱能量分布代表 300 ~ 830 毫微米光谱范围内的天然日光。

（三）标准色度系统

1. CIE 色度图

色度（chromaticness）是用三原色匹配可见光谱中各波长而产生某种颜色时的三原色比例系数。现代色度学采用国际照明委员会

(International Commission on Illumination, 简称 ICI, 后统一简称 CIE) 所规定的一套原理、数据和公式来测量颜色, 这套系统称为 CIE 标准色度系统。它是根据三原色原理, 用红 (R)、绿 (G)、蓝 (B) 三色按不同比例来匹配任一颜色。

CIE 色度系统在发展完善过程中有好几种变式, 修正工作主要围绕如何确定三刺激的光谱值, 而其基本的方程式仍是以 1931 年 CIE—RGB 系统的颜色方程为基础。此方程为:

$$(C) = r(R) + g(G) + b(B)$$

(C): 一种特定的颜色

(R)、(G)、(B): 红、绿、蓝三原色

r、g、b: 每种原色的比例系数

: 表匹配意思, 即视觉上颜色相等

因为三原色的总光量必须与被表示的颜色相等, 所以

$$r = \frac{R}{R+G+B}; g = \frac{G}{R+G+B}; b = \frac{B}{R+G+B}$$

$$r+g+b=1$$

例如, 对某一颜色表达为:

$$(C) = 0.06(R) + 0.63(G) + 0.31(B)$$

显然, 由于绿的成分占比例最大, (C) 主要表示为绿颜色。用三原色匹配某一颜色时, r、g、b 中的有些值可能会取负值, 而且往往须将其加到被配色上去。例如, 求橙色的等式:

$$(C) = r(R) + g(G) - b(B)$$

实际是:

$$(C) + b(B) = r(R) + g(G)$$

根据三原色的混合原理, 国际照明委员会在 1931 年制定了一个色度图 (chromaticity diagram), 又称 CIE 1931 色度图 (CIE 1931 chromaticity diagram)。在色度图上, 不仅可以了解组成某一颜色的三原色比例, 还能看到此颜色的补色、明度及饱和度。参见图 6-21。CIE 色度图以 X、Y 对颜色进行定义。图中 X 色度坐标相当于红原色的比例, Y 色度坐标相当于绿色比例。图中没有标出 Z 色度坐标 (蓝原色) 的比例, 但根据 $X+Y+Z=1$ 可以推算出来。图 6-21 呈马蹄形, 马蹄形上的各点代表 380 毫微米 (紫色) 到 780 毫微米 (红色) 之间所有单色光。其中图的右下端是红色光谱段, 左上角是绿色段, 左下部是蓝色段。马蹄形上各波长的连线叫光谱轨迹 (spectral locus), 从紫端到红端的连接直线是光谱上没有的由紫到红的颜色, 而色三角内包括一切物理上能实现的颜色。在色度图中, $Y=0$ 的直线即是无亮度线, 光谱的短波紧靠这条线, 这意味着 380~420 毫微米波长的能量的亮度感觉很弱。

图中三角形内的任何一种颜色, 只要已知匹配的三刺激就可以定出 X 和 Y 的色度坐标, 并且能进一步确定此颜色的主波长 (色调)、纯度 (饱和度), 但是, 要完全确定一种颜色还应加上它的亮度值。图中轨迹上的各点代表纯光谱色 (饱和度最大)。C 点代表 C 光源, 由三原色各三分之一产生, 可理解为一种标准的白光。

2. 实例计算

以上我们讨论 CIE 色度图的原理, CIE 1931 色度图只是色度图中的一

种。下面我们通过一实例来了解如何通过色度图上的坐标位置，推算出某一被测物的颜色特征。

前面讲到，任何颜色在色度图中都占一确定位置。假如有 Q、S 两个颜色，Q 色度坐标 $X=0.16, Y=0.55$ ，S 色度坐标为 $X=0.50, Y=0.38$ ，由 C 通过 Q 作一直线至光谱轨迹，在 511.3nm 处与光谱轨迹相交，Q 颜色的主波长即为 511.3nm，相当于绿色。饱和度由某一颜色离开 C 点接近光谱轨迹的程度来表示。某颜色愈靠近 C 愈不纯，愈靠近光谱轨迹愈纯。从色度图还可以推算出由两种颜色混合所得的各种中间色。如 Q 和 S 相加，得出 Q 到 S 直线的各种颜色。以这一直线上的 T 点为例，由 C 通过 T 抵达 572nm 的光谱色，则可由 572nm 波长的颜色看出 T 颜色的主波长，并可由 T 在 C 与 572nm 光谱之间所占的位置看出它的饱和度。

CIE 色度图三角形的底边为非光谱色。假若有个色点 U 处在 C 点之下，则颜色特征要从 C 点引直线通过 U 点与底边相交，表示它的主色为非光谱波长的绛色。若向上延长线与光谱边界相交，可以鉴定这个绛色的补色为 510nm。这个实例说明了求得补色的原则：通过 C 点的直线与三角形边界相交的两色互为补色。在标定颜色时，为了区别主波长，可以在补色波长后面加“C”表示。如 510C 就是表明 U 点的补色波长为 510nm。

我们不仅可从 CIE 色度图上，求出一种颜色的主波长，而且还可求出它的饱和度。一种颜色的饱和度是在色度图上用光源色度点至样品色度点的距离与光源色度点到光谱色度点的距离的比率来表示。饱和度 (P_t) 可用如下公式计算得出：

$$P_t = \frac{X - X_0}{X_\lambda - X_0} = \frac{Y - Y_0}{Y_\lambda - Y_0}$$

$X、Y$ ：被测物的色度坐标值

$X_0、Y_0$ ：白点，即光源色度坐标值

$X_\lambda、Y_\lambda$ ：主波长坐标值

若某一被测物坐标位置分别是 $X=0.1720, Y=0.5333$ ；其主波长为 510nm，坐标 $X_\lambda=0.0139, Y_\lambda=0.7502$ ；白点(CIE 光源 C 位置) $X_0=0.3101, Y_0=0.3163$ 。将上述各数值代入公式得：

$$\text{饱和度 } (P_t) = 0.5333 - 0.3163 / 0.7502 - 0.3163 = 50\%$$

从以上所述可见，有了色度图就可使我们很方便地推算出来某一被测物的主波长和饱和度。例如，我们用色度计测出两块样品，其色度坐标分别如表 6 - 10 所示，经作图计算，即可求出被测物主波长和纯度，见表 6 - 11。

表 6 - 10 某被测物的色度坐标值

被测物	X 色度坐标值	Y 色度坐标值
1	0.3	0.6
2	0.6	0.3

表 6 - 11 被测物的主波长和饱和度

被测物	主波长 (nm)	饱和度 (%)	颜色估计
1	550	75	绿
2	630	75	橙

尽管从色度坐标的比较也能粗略地估计颜色特征，然而知道了主波长和饱和度就使我们更进一步了解被测物的颜色特征。应当指出的是，本例仅是掌握色度系统的原理，用作图法大致估计了被测物体的主波长、饱和度和颜色。当今随着电子技术和计算机的迅速发展，精密的色度计（colorimeter）能十分准确地用数字显示被测物体的颜色特征。

第四节 颜色的心理效应

颜色不仅与人的视觉有关，而且可以影响人的情绪和行为，甚至还可以影响人的工作效率。我们把颜色对人们心理上的各种影响，称之为颜色的心理效应。

实验证明：在绿光的照明下，人的听觉感受性会提高；而在红光的照明下，则听觉感受性会下降；在橙黄色的灯光照射下，手的握力比平常要大，而在红光下，手的握力更大；在红光照明下观察物体大小比在蓝光照明下观察的要大些，等等。

根据色光波长的长短，色温也会相应低些或高些。在低色温光源照射之下，人有温暖的感觉，暖色调有黄、橙、红等色。反之，青、蓝、绿等色以短波占优势，为冷色调，使人有清凉或寒冷的感觉。

颜色还可以给人以轻、重、远、近的感觉，甚至可以影响人的味觉。有人做过这样的研究：将同样的咖啡倒入形状相同，只是颜色不一样的杯子，然后请别人分别一一品尝。结果大部分人认为，黄色杯子的咖啡淡；绿色杯子的咖啡带酸味；红色杯子的咖啡味道香浓。国外有些研究者专门研究用什么颜色的光照射时，肉类、蔬菜、水果等食物的颜色显得最新鲜。例如，用红橙色光照明肉类使人有新鲜红润的感觉；若用蓝绿色光照明就会使人觉得肉已变质腐烂。

颜色的心理效应可以从颜色的三个特征来具体说明。当然，在这三个特征中，色调是主要的。另外，世界上各个民族由于受到长期环境、历史、文化条件的制约，对颜色也有不同的心理效应。

一、色调的心理效应

色调是颜色家族的名称，是我们辨认红、橙、黄、绿、青、蓝、紫的术语。若我们横切一段彩虹，再把它弯成一个圆环，我们就能看到图 6-20 的色环。某种色彩的位置直接相对于其他色彩称为互补色（complementary color）。若把互补色的光混合在一起，它们就会呈现淡灰色。若互补色不是混合而是换个排列，视觉就会产生强烈对比，它们看上去要比实际的更强。生活中有这样的例子，棕红色头发者穿鲜艳的天蓝色的毛衣，会增加红的强度。然而，密集的互补色小斑点在一定距离之外可以使视线模糊，呈现出一片淡灰色。这就是许多花呢和印花布在我们的手中时比较鲜艳，而在一定距离外就失去了光彩的原因。

（一）色调的冷暖感

由于人们对客观事物的某些联想作用，另外还根据对光的吸收与反射程

度，而使不同的色调产生不同的冷暖感。

红、橙、黄长期与太阳这样的热源相联系，而蓝、绿、紫则暗示着清凉的河水、宁静的草地、遮阴的树木和遥远的山坡。以科学的观点来看，光谱中红色确实偏暖而蓝色偏冷。红色的波长较长，属远感色（或远色）（distant color），给人以温暖的感觉，故又称暖色（warm color）。蓝色和紫色波长较短，属近感色（或近色）（closed color），产生阴冷的感觉，故又称冷色（cool color）。在服饰中常能观察到这些现象，穿暖色调的服装的人体似乎进入最突出的位置，显得更大更重要，而穿冷色调的服装的人体似乎使距离拉远，因此显得小而不要。

处于红色和紫色中间的色彩是绿色，因此它通常作为光谱中冷暖色调的分界点。偏于黄色一侧的绿色，诸如苹果绿则比偏于蓝色一侧的绿色要偏暖。然而，所有色彩可以认为都具有冷暖的性质。即使是蓝色，只有稍加点绿色，也要比略含紫色的蓝色偏暖。

一般说来，研究显示不同的个人对于特殊的颜色的反应，其特征具有显著的共同之处。暖色调（黄、橙和红）通常被认为是欢乐的、积极的、刺激的和兴奋的，而冷色调则暗示着宁静、冷淡、镇定和肃穆的。在判断或组合色彩时，了解色彩的特征是十分重要的。

（二）色调的情感

色调的表现力是明显的，在我们的日常生活中，色调和感情或情绪相联系。人们常常用颜色来表示某种心态。例如在英语中，“feeling blue”，blue是蓝色，又可解释为感觉悲哀；“in the pink”，pink是粉红色，又可解释为情况良好；“green with envy”，green是绿色，又可解释为受强烈感情影响的心态，如极为羡慕；“rosy disposition”，rosy是玫瑰色，又可解释为天性乐观的。每个人对于颜色的感受是不相同的，他们给色彩附加的意义通常取决于他们过去与色彩的联系和经验。然而，有些实验表明，人类对于色彩的反应有某些共同的特征。

例如，佛罗里达大学教学医院儿科病房所做的一项实验，记录了九个月到五岁之间的幼儿首次接触穿白大褂的护士的反应，然后让护士再换上色彩更为柔和的上衣。幼儿对于浅色上衣的反应比对于白大褂的反应更为宁静、松弛和驯服。相同的研究表明，墙壁的颜色对于病人的精神状况有着明显的效果。

曾有一位诗人写过一首题名为《色彩》的诗，诗中写道：

生命是没有价值的白纸，自从绿给了我以发展，红给了我热情，黄教我以忠义，蓝教我以高洁，粉红赐我以希望。灰白赠我以悲哀，再完成这帧彩图。黑还要加以死。从此以后，我便溺爱于我的生命，因为我爱它的色彩。（闻一多，1940）

诗中写出了色彩的象征意义。诗人以简洁的语言写出了强烈的感情色彩和对生活的美的追求，使本无生命的色彩带上了一定的社会性和强烈的感情因素。

（三）色调的环境心理效应

不同的工作环境，可根据不同的情况采用暖色调或冷色调，这样会使工作者获得良好的心理效应。在餐馆中，为了让顾客有一种温暖的感觉，照明光采用黄、红的暖色调。毫不奇怪，生活在“热色”装饰的屋内，可使人们感觉到比实际温度暖和。俄国科学家证明：在红灯下工作的人的反应比其他人敏捷，不过他们完成任务的效率下降很快。日本和美国的科学家都认为红光可以改变人们大脑里正常的生物电，即使是红光短暂照射也能使人类皮肤内的生物电发生改变。另外，还有些很有趣的现象，如选花布，青少年比较喜欢跳跃而明朗的色彩，妇女比较喜欢鲜艳轻柔的颜色，农民喜欢大红大绿的颜色，特别是新春佳节更是如此，这是因为色彩变化与人们的情感是协调一致的。明快、暖和的色彩能调和冬日萧索的外界景象，使单调乏味的生活丰富起来，满足人们在有限条件下美的享受和愿望。又如，很多人感到粉红色明亮、温暖、幸福，这就是粉红色对人们的刺激而产生的抽象心理感受，但这还是从生活中抽象而来的，有着生活的依据。

由于色调对人们具有环境心理效应，因此，商业广告业也十分重视对色调搭配的研究。不少研究认为，商业广告的颜色要避免同时使用互补色。有实验证明，为了取得良好的观视效果，背景色和字体色的良好匹配的优劣等级如下：

黄色背景，黑字；白色背景，绿字；白色背景，红字；白色背景，青字；
青色背景，白字；白色背景，黑字；黑色背景，黄字；红色背景，白字；
绿色背景，白字；黑色背景，白字；黄色背景，红字；红色背景，绿字；
绿色背景，红字。

二、明度的心理效应

明度为一种颜色的明暗程度，它取决于吸收或反射光线的总体比例。白色物体可以吸收不到 10% 的光线，而黑色物体则可以吸收 95% 以上的光线。难怪夏季穿着黑色的服装似乎显得特别热。任何深色织物吸收光能的比例都较高，它比同样结构的浅色织物所感觉的实际温度要高。

明度用 1 到 10 的标度来表示(参见图 6-19)，接近黑色的一端标度为 1，接近白色的一端标度为 9。一种物体所显示的明度在黑白中间，其色值标度为 5。一种颜色色度较亮，其标度大于 5，一种颜色色度较暗，其标度小于 5。

一种颜色的明度终极是色彩在明度上达到最充分饱和的等级。色调按明度的自然顺序从黄色（其饱和等级为 8）两侧过渡，到达黄色的互补色紫蓝色处是最暗的标度。绿色的明度极值为 5，红色的为 4，紫色的为 3。这些标准中，饱和的最大数字将出现变化，也就是说，黄色的最大数字将在更亮的标度中发现，而紫蓝色的最大数字将在较暗的标度中。

（一）明度和谐的心理效应

明度是颜色的一个方面，它可以独立于色彩和饱和度而被观察。明度对比在任何图案中都是必需的。相同明度的临近色彩之间的界限是模糊难于分辨的。因此，明度对比的一个重要功能是确定形式和形状，使颜色之间的界

限更为分明。相同的明度产生柔和模糊的界限，而差别较大的明度则产生严格的清晰的界限。明度对比可以把我们的眼睛从明度相等的单调图案中解放出来。

明度的自然顺序提供了色调和谐的基本原理之一。当颜色组合时，若能够保持自然的明度关系，它们可能就是美观的。例如，黄、绿黄和绿色组合时，若黄色在配置中能保持最亮的明度，而绿色保持最暗的明度，则结果会更为和谐。在有些情况下，颠倒明度的自然顺序来设计不和谐的色调，会产生不寻常的令人兴奋的效果。色彩中的不和谐与音乐中的不和谐相似，在音乐中，一个小的和音会产生明显的振动效果。服装中不和谐的明度可以产生意想不到的效果，但是技术不娴熟，也会导致不愉快的情绪。

（二）明度对比的心理效应

在明度对比方面，中等明度与暗色对比就会显得较亮，与明色对比就会显得明度偏低。红色被黄色包围时就会比被蓝色包围时显得更暗，这就是为什么明色和黑色组合、暗色和白色组合对比似乎显得更亮的原因。

这些对比的效应常常在人们的生活中获得应用。一个图案的明度基调可以是明亮的、中性的、或暗淡的。相近明度间隔中单调的图案显示女性的柔弱特质。特别强调黑色的明亮色值往往显得更为有力。

图 6-22 就是以中性明度为基调的对比实例。一条黑色的天鹅绒带和银白色的发套形成鲜明对比。同样，一双黑鞋为白袜所强调。上衣的明度属于中性，处于黑白之间，丰富的刺绣图案使明暗发生有趣的变化，使明暗之间的对比转移。在黑色和白色有力的强调下，可以产生富于男子气质的特质。

缺乏明度对比常常会导致单调，而且位置和距离也影响着心理效果。因为位置首先引起我们眼睛的注意。当我们在远处观察物体时，明度对比效果比较弱，它们的轮廓接近我们的眼睛时，就会显得鲜明清晰。因此，强烈的鲜明的明度对比使它们的位置后缩。这个原理在服饰中的运用非常重要。若人们希望增加或减少自己体形的外在尺度，可用较高的明度，由于反射的光更多，通常使物体和体形突出，显得比低明度的相同的物体的尺寸更大。然而，这一现象还取决于物体和它的背景之间的对比关系，例如，对于一名滑雪运动员来说，在雪坡上穿黑色服装比穿白色服装更显眼。可见，对比原则运用是很复杂的。

三、饱和度的心理效应

饱和度是颜色的第三个维度，它表示色彩的强度或纯度。强烈的、鲜艳的、饱和的色彩说明强度高；柔弱的、灰暗的、中性的色彩说明强度低。在孟塞尔颜色系统中，饱和度是以淡灰色为起点的标度来度量的。从图 6-19 的中心轴开始向外延伸，随着强度的增加，标度从 1 达到最强的饱和点。图 6-19 的三维投射显示了色彩、明度和饱和度的关系。

一种色调的强度由于它和它的互补色的混合而降低。特定色调的最大饱和和只有在其明度达到终极的时候才能获得，明度的减少可以自动降低饱和度的标度。相同程度的色调的结合通常创造美观的效果，而柔弱的色彩组合比两种以上最大强度的不同色彩的组合更为和谐。鲜明的颜色往往产生强烈的

戏剧性结果，而灰暗柔弱的色调则产生柔和精巧的效果。

与颜色的其他特性一样，饱和度对于整体外观也有很大的影响。鲜艳的色彩也可以产生与暖色调一样的效果，它们往往突出、醒目，使外表尺度增加。柔弱的色彩和冷色调一样，平淡、缓和、较不显眼。

不同饱和度之间的相互作用也能产生一定的视觉效果。强度较弱的蓝色和饱和的蓝色并列就会变成灰色。强度中等的红色与玫瑰红对比就会增加饱和度。例如，生活中穿暗棕色的上衣会使深棕色的头发色彩显得更强烈。

颜色对比效果还随面积大小而有所差别，当刺激面积小到难以区分不同颜色的界限时，就产生色融合现象，产生混色感觉。

以上我们从颜色的三个特征上分别说明了它们的心理效应。应该看到，颜色的三个特征的作用常常是统一的，不能分割的。由于这些特征的协同作用，产生了不同的心理效应。例如，由于色彩的明度有高低不同，产生了一种或轻或重的感觉。影响轻重感觉的主要彩色属性是明度，以中等程度的明度为中心，在它以下为重感型，以上为轻感型。色调对它也有些影响，彩度高的色属于轻型。高明度，或较低的中彩度色彩使我们有一种柔和的情趣。低明度而高彩度或较低彩度色彩，会使人产生硬的感觉。无彩色的黑和白是硬色，而灰色是软色。明浊色有柔软的感觉。可见，明度和色调是决定柔软和坚硬的重要条件。低明度高彩度的颜色产生强的感觉，高明度低彩度的颜色有弱的感觉。

四、色彩爱好的民族差异

各国人民对色彩的爱好既有共性又有差异。这些差异是由于社会意识形态、文化历史、人种肤色、民族、地区、气候和风俗习惯等因素造成的。根据我国学者对多年的资料的整理，获得一些国家和地区人民对颜色的嗜好一览表，参见表 6—12。表上所列的喜爱的颜色定有疏漏，而且人们的喜爱也并非是一成不变的，因此，此表也仅供参考而已。

表 6-12 一些国家和地区对颜色的嗜好一览表

国家和地区	喜爱颜色	国家和地区	喜爱颜色
比利时	女孩爱蓝色、男孩爱粉红色	奥地利	绿色
爱尔兰	绿色及鲜明色	法国	粉红色（少女服）、蓝色（男孩服），一般人爱黄色
美国	鲜艳颜色	马来西亚	绿色（象征于宗教，亦用于商业）
荷兰	橙色、蓝色代表国家并多用于节日	希腊	白、蓝、黄
瑞士	红、橙、黄、蓝、绿、紫、红、白相间及浓淡相同色组	摩洛哥	喜爱稍暗的色彩
厄瓜多尔	凉爽高原喜暗色、炎热沿海喜白色及明暗间色	突尼斯	绿、红、白色组，居住该地犹太人喜爱白色
巴拉圭	明朗色彩	西印度群岛	明朗色彩
伊朗蓝教地区	绿色	夏威夷	蓝、黄、绿
埃及	绿色	西班牙	黑色
伊拉克	深蓝、红色	新加坡	红、绿、蓝
叙利亚	青、蓝、绿、红色	东南亚	各种鲜明色彩
秘鲁	红、紫、黄	巴基斯坦	绿、金、银、翡翠色及其他鲜艳色
委内瑞拉	黄色（医务标志）	土耳其	绿、白、绛红色及鲜明色彩
古巴	鲜明色彩，色彩感情上同美国相似	日本	黑色、红色、黑白相间
墨西哥	红、白、绿色组（国旗色）	伊朗	绿色
加拿大	同法国、美国差不多	意大利	食品、玩具喜欢用醒目鲜艳色、服装高级包装用淡雅色，绿色
挪威	红、蓝、绿等，鲜明色彩与当地冬季长有关	德国	深浅奶黄色、咖啡色，南方喜爱鲜明色彩
保加利亚	深绿色及不鲜艳的绿色和茶色		
非洲	当地人爱好鲜艳单色，大陆地区受欧洲影响，偏远地区保持民族色彩	中国港澳地区	红、绿

（采自上海纺织工专（编），1986）

本章实验

一、颜色混合实验

(一) 目的：颜色混合可归结为三条规律：

(1) 每一种颜色都有另一种颜色与它按适当比例相混合而产生白色或灰色(非彩色)或第三种颜色，凡混合能成非彩色的成对的颜色叫互补色。例如红与浅绿为互补色，黄与蓝互为补色。在混合时如比例不对，则成不饱和的彩色，色调偏于过多的一色。

(2) 在混合两种非补色时，便产生一种新的介于它们之间的中间色(蓝与红混合产生紫色、红与黄混合产生橙色)，中间色的色调偏于成分较多的一色；饱和度决定于二色在光谱色序中的远近，越近越饱和。

(3) 混合色的颜色不依被混合的颜色成分为转移，即每一种被混合的颜色本身也可以由其他颜色混合而得(黄与蓝混合，并不依赖于黄是纯光谱或其他颜色混合的结果，同样都产生灰色)。

本实验的目的在于表征颜色混合及其规则，并对简单的配色公式作初步的理解。

(二) 材料：混色轮(附刻度盘)甲乙两个。大纸盘(以直径为 195 毫米为最佳)：红、绿、黄、蓝等色各一种。小纸盘(以直径为 140 毫米为最佳)：白、黑、紫、橙等色各一种。

(三) 程序：

1. 预备实验

把两个混合轮安放在桌子上，相距约 15 厘米，能在同一垂直平面上旋转。把大纸盘放入甲混色轮中，并把小纸盘放入乙混色轮中作比较用。被试者坐在离混色轮 2 米远的地方进行观察，而主试者则操纵混色轮的旋转速度，调整纸盘上各种颜色的比例。当混色轮转动稳定后，被试者要仔细观察大小纸盘的色彩是否相配(指大纸盘与小纸盘的色调、明度与饱和度相同)或有所不同，然后主试有计划地调整大纸盘或小纸盘的颜色比例，直至两者相配为止。相配后，把纸盘上各种颜色的比例(即刻度)记录下来。

2. 补色的混合

(1) 把黄色与蓝色混合成为灰色。把白色与黑色的小纸盘放在乙混色轮上，使白色露出 70%，黑色露出 30%。当混色轮转动稳定后，要求被试观察两色混合后是否成为灰色，否则须继续调整两色的比例，直至产生灰色为止。然后把黄色与蓝色的大纸盘放在甲混色轮上混合，调整两色的比例，使与小纸盘的灰色相配。

(2) 把绿色与紫色混合成为灰色。用同样的方法混合绿色和紫色，使之与灰色相配。

3. 非补色的混合

(1) 按上述方法混合红色与蓝色，使之与紫色相配。

(2) 按上述方法混合红色与蓝色，使之与橙色相配。

(四) 结果：

写出如下配色公式：

1. 把补色混合的结果列成下式：

____% 的黄色 + ____% 的蓝色 = ____% 的白色 + ____% 的黑色
____% 的绿色 + ____% 的紫色 = ____% 的白色 + ____% 的黑色

2. 把非互补色混合的结果列成下式：

____% 的红色 + ____% 的蓝色 = 紫色
____% 的红色 + ____% 的蓝色 = 橙色

(五) 讨论：对下列参考问题进行讨论。

1. 实验结果是否能证实颜色混合的三项规律？
2. 画家配色与本实验的颜色混合是否相同？
3. 对本实验所用的仪器、材料以及实验方法有何改进意见？

二、闪光融合实验

(一) 目的：学习用闪光频率仪测定闪光临界融合频率 (CFF) 的方法。

(二) 材料：数字式闪光频率仪，单眼罩，头部固定支架。

(三) 程序：

1. 预备实验

(1) 主试者按仪器使用要求接通电路进行操作。

(2) 在微光中，被试在眼部距离仪器 40 厘米处端坐，支架固定头部，眼与仪器上的红光点呈一水平线。实验前，被试者暗适应约 10 分钟，分别用左眼和右眼进行观察。

(3) 主试者以小步调整闪光频率，数码管显示的数字为每分钟的闪光次数。按递增法先做二、三次练习实验，使被试了解实验要求。指示语：“当你感到光不闪时，立即报告”

2. 正式实验

(1) 主试者每次都按递增法调整闪光频率，直到被试者报告闪光融合为止，主试记录每次成绩。

(2) 先测左眼 10 次，休息 1 分钟，再测右眼 10 次。

(四) 结果：

分别求出个人和全组的左右眼 CFF 平均值。单位可化作闪次/秒。

(五) 讨论：

1. 左右眼 CFF 是否有差异，原因何在？

2. 试与另一被试者的成绩作比较，如有差异，原因何在？

用递增法测定闪光临界融合频率 (CFF) 的记录表

序次	1	2	3	4	5	6	7	8	9	10
CFF 值										
眼别										
左眼										
右眼										

三、视觉后象实验

(一) 目的：观察负后象，研究影响负后象的若干条件。

(二) 材料：

1. 投射卡片一组，25×25 厘米，灰色、白色。

2. 刺激卡片一组，在 10×10 厘米的颜色卡片中央，贴一张 2.5×2.5 厘米的颜色纸：蓝上贴黄，黄上贴蓝，蓝上贴绿，蓝绿上贴红，黑上贴白，白上贴黑，黄上贴红，白上贴红，白上贴绿，白上贴黄，白上贴蓝。

3. 一张黑卡片，上面贴着两张 2×2 厘米的纸块，一灰一白，相距约 1.5 厘米。

4. 一张黑纸片，上面贴着两张白纸块，其中一张是 2×2 厘米，另一张是 0.5×0.5 厘米，相距 1.5 厘米。

5. 双针停表。

(三) 程序与结果：

1. 刺激的颜色和负后象的颜色

(1) 被试者自己兼任主试者。在投射卡的顶上，放一块蓝上有黄的刺激卡片。在黄色方块中心作一小点，有助于视线的固定。注视黄色方块约 20 秒钟，双眼不要左顾右盼，然后移去刺激卡片，注视灰色投射卡片至少 10 秒钟。当后象出现时，要注意后象是什么色调与明度。明度可用 5 (白)、(很亮)、3 (亮)、2 (暗)、1 (很暗)、0 (黑) 表示。把观察结果记入表 (1)。

(2) 用下列刺激图片依照上述方法进行观察：黄上贴蓝、蓝绿上贴红、黑上贴白、白上贴黑。把每次观察的结果记入表 (1)。

(3) 试从上表所列的结果概括出一条尝试性的法则：_____。

(4) 如果你所概括出的法则是正确的，那么你就能预测表 (2) 中的各种颜色后象。在每次观察之前，要先把预测结果记在表中。

(5) 若表 (2) 所载结果，对你首次概括的法则有所扩展或改正，可重述于下：_____。

2. 刺激时间与后象的潜伏期及延续现象

(1) 用任何一种颜色刺激卡片与灰色投射卡片。每次注视刺激卡片的时间不同，事先安排如表 (3) 所示。用双针停表记录时间。

(2) 从刺激停止到后象出现这个时距称为“潜伏期”，从后象出现到后象消失这个时距称为“延续期”。在每次观察过程中，被试者对刺激停止、后象出现与后象消失都要口头报告，主试者先后分别记录时间。每种刺激时间均观察 3 次。

(3) 被试者观看刺激的时间不同是否影响后象的色调、明度和饱和度。

(四) 讨论：

1. 产生负后象的原因是什么？

2. 为什么在日常生活中很少看到负后象？
3. 负后象是“连续对比”的例子吗？举例说明“同时对比”。
4. 投射卡的颜色对后象有什么影响？试拟研究这个问题的实验方法。

表（1）在灰色背景上后象的色调与明度

后象	刺激	蓝上黄	黄上蓝	蓝绿上红	黑上白	白上黑
	色调					
	明度					

表（2）在白色背景上后象的色调与明度

色调	刺激	黄上红	蓝上绿	白上红	白上黄	白上绿
	预测					
	观察					
明度	刺激					
	预测					
	观察					

表（3）后象时间与刺激时间的依存关系

刺激时间（秒）		1	2	5	10	20	40
潜伏期 （秒）	1次						
	2次						
	3次						
	平均						
延续期 （秒）	1次						
	2次						
	3次						
	平均						

本章摘要

1. 特定的光刺激人眼而引起视觉。光是一种电磁波，引起人类视觉的电磁波称为可见光谱。可见光谱的波长范围大约是 380 毫微米到 780 毫微米之间。

2. 光的单位名目繁多，与实验心理学关系密切的有：光强度、光照度、光亮度、光通量、反射系数等。

3. 在视觉研究中，对于自变量、控制变量和因变量的精心考虑是十分重要的。

4. 在微光条件下，视觉感受性逐渐提高的过程叫暗适应。从暗处到亮处，眼睛大约经过一分钟就能适应，这是明适应。

5. 视敏度是指分辨物体细节的能力，是人眼正确分辨物体的最小维度。

用视力表测定视敏度，它以视角的倒数来表示。

6. 闪光临界融合频率是一个亮暗交替的光刺激，其交替速度正好达到被人感觉为闪光和稳定光之间的频率。闪光临界融合频率与刺激的物理特性和被试的生理心理因素有关。

7. 视觉感受一种颜色取决于三个基本特征：即亮度、色调和饱和度。任何一种颜色都是三者的总合的结果。

8. 光谱中的色光混合是一种加色法，加色法的原色是红、绿、蓝，它们分属于可见光谱的两端和中部。用颜料、油漆等的混合配色是减色法。减色法的三原色是黄、青、紫，它们是加色法三原色的补色。

9. 色光的混合遵循三条件定律，即补色律，居间律和代替律。

10. 人眼注视外界景物，大约有近似 2° 的直径范围能获得最清晰的感觉，称为中央凹视觉。在中央凹视觉周围的区域只能获得模糊的视觉景象，这是边缘视觉。人眼的中央视觉和边缘视觉用视野计能够准确地测量。

11. 颜色对比是两种不同的色光同时作用于视网膜的相邻区域，或者相继作用于视网膜的同一区域时，颜色视觉所发生的变化。前者是同时对比，后者是继时对比。

12. 人眼对物体颜色的感知，在外界条件变化的时候，仍能保持相对不变，表现出颜色常性。

13. 现代色度学系采用国际照明委员会所规定的一套原理、数据和公式来测量颜色，这套系统称为 CIE 标准色度系统。它是根据三原色原理，用红、绿、蓝三色的不同比例匹配成任何一种颜色。

14. 颜色不仅与人的视觉有关，而且可以影响人的情绪和行为，还可影响人的工作效率。

建议参考资料

1. 杨治良（1989）：服装心理学。兰州市：甘肃人民出版社。
2. 张春兴（1989）：张氏心理学辞典。台北市：东华书局（繁体字版）。上海市：辞书出版社（1992）（简体字版）。
3. 林仲贤等（1987）：视觉及测色应用。北京市：科学出版社。
4. 喻柏林（1979）：照明变化对视觉辨认的影响。心理学报，11 卷，3 期，319~325。
5. 赫葆源、张厚粲、陈舒永等（1983）：实验心理学。北京市：北京大学出版社。
6. Atkinson, R.C. et al. (1988). Stevens' handbook of experimental psychology (2nd.ed.). New York: A Wiley - Interscience.
7. Cohen, J. (1969). Eyewitness series in psychology. Chicago: Rand McNally.
8. Philipchalk, R.P. (1995). Invitation to social psychology. Orlando: Harcourt Brace & Company.
9. Roediger, H.L., & Weldon, M.S. et al. (1992). Direct comparison of two implicit memory tests: Word fragment and word stem completion. Journal of Experimental Psychology, 18 (6), 1251~1269.

第七章 听觉实验

本章内容细目

第一节 听觉实验中的变量

一、听觉的物理刺激和音感 340

- (一) 声波及其特征
- (二) 频率和音高
- (三) 振幅和响度
- (四) 波形和音色
- (五) 声波的相互作用

二、听觉实验中的各种变量 352

- (一) 自变量
- (二) 因变量
- (三) 控制变量

第二节 听觉的二个基本属性

一、声的测量 356

- (一) 声的记录
- (二) 声的测量

二、听觉属性——音高 358

- (一) 音高量表
- (二) 音高和强度的关系
- (三) 音高和周相的关系
- (四) 音高和复合音

三、听觉属性——响度 363

- (一) 响度量表
- (二) 声强的差别阈限
- (三) 等响曲线
- (四) 响度与时间

第三节 听觉的基本实验

一、人的听力 368

- (一) 听力图
- (二) 听觉模式

二、听觉掩蔽现象 370

- (一) 纯音掩蔽
- (二) 噪音掩蔽
- (三) 噪音与纯音对语言的掩蔽

三、听觉的疲劳和损伤以及适应 374

- (一) 听觉疲劳

- (二) 听力损伤
- (三) 听觉适应
- 四、听觉定位 377
 - (一) 双耳强度差
 - (二) 双耳时间差
 - (三) 双耳周相差
 - (四) 立体声听觉
- 五、可见语言 380
- 六、汉语语音 384
 - (一) 汉语语音特点
 - (二) 汉语语音结构特点

本章实验

- 一、听觉定位实验 387
- 二、可懂度实验 389

本章摘要

建议参考资料

上一章我们讨论了视觉。视觉和听觉是人们接受外界刺激的二个最主要通道。人类生活在充满声音的物质世界里，我们几乎每时每刻都在接受外界的声音刺激。听觉使我们能够享受到美妙的音乐和小鸟的歌唱。它使我们能与家人和朋友们交谈。电话铃声、敲门声和汽车的喇叭声能对我们进行提醒和告诫，火车轮子的吱吱作声和心脏的杂音使我们能作出质量的评价和临床的诊断。所以，通过听觉人们可获得声音所传递的各式各样的信息，得以通讯往来，欣赏音乐，传授知识，交流思想。听觉影响到人们实际生活的许多方面，是认识世界的仅次于视觉的重要信息源。

然而，在当今社会上，声音也常给我们带来烦恼。有许多声音是不悦耳的，或者是不受欢迎的，通常我们称之为噪音。噪音已成为当今环境心理学（environmental psychology）的三大主要研究课题之一。当然，噪音能引起多大烦恼不仅取决于声音的性质，而且还取决于听者的主观态度。例如，对于设计工程师，他所设计的喷气式飞机起飞时的声音，对他来说或许如同音乐般美妙，而对于居住在机场附近的居民来说，却是震耳欲聋的痛苦。声音甚至还有它的破坏性。超音速飞机的音爆可以粉碎附近的玻璃窗，震落墙皮。而最不幸的莫过于火箭、大炮等的巨响，有时会震聋人耳，给人带来终身残疾。当然，声音的这些负面比起它的正面来，那只是次要的。不论是正面，还是负面，都告诉我们掌握声学知识的重要性。

听觉是声波作用于听分析器所产生的感觉。听觉的适宜刺激是频率在 16 赫（Hz）至 2 万赫的声波。音高、响度、音色等音感都是由声波的频率、振幅、波形等物理特征决定的。

听觉也是心理学研究较多的领域之一。众多的研究，从各个侧面提示了听觉现象的奥秘。在本章里，将讨论如下六个问题：

1. 声波有哪三个基本特征，它和视觉刺激有何类同之处。
2. 试谈音高、响度二个听觉属性在听觉实验中充当变量的情况。
3. 何谓等响度曲线。
4. 什么叫听觉掩蔽现象，有何用处。
5. 语图仪的基本原理是什么，有何用处。
6. 谈谈汉语语音上的特点。

第一节 听觉实验中的变量

听觉是人类和许多动物共同具有的感觉现象之一。人的听觉是借助于耳来实现的，是对一定频率范围内声音刺激的感觉。听觉现象的产生受许多因素作用的影响，我们只有全面了解这些因素及其性质，才有可能把握住它，更好地处理好听觉实验中的各种变量及其关系，从而揭示纷繁的听觉现象和规律。

一、听觉的物理刺激和音感

(一) 声波及其特征

听觉 (audition) 是个体对声波物理特征的反映。频率为 16 赫 ~ 20000 赫的机械波为声波。听觉的适宜刺激是声波。声波是弹性媒质中物体振动所激起的纵波。当空气振动波到达我们耳内时，就产生听觉 (参见图 7-1)。

图 7-2 是一个机械振荡器振动扬声器产生声波直至画出正弦曲线的模拟装置。这个装置类似于蒸汽机火车头的汽缸动力装置。图中的圆轮即为机械振荡器，与它相联的为两个活塞，左边的活塞带动画笔，描绘运动曲线；右边的活塞振动膜片，使扬声器发音，再由麦克风传到示波器中，显示出声波的波形图。在实验操作中，当我们改变圆轮的转速，如使其逐渐加快，则扬声器中发出的声音会由“粗”变为“细”，同时，记录纸和示波器上的图形也会发生相应的变化——波纹由疏变密，频率增加但振幅不变；而当我们改变圆轮的半径，使其不断增加，则扬声器中的声音逐渐加大，波形记录上的频率不发生变化而振幅却相应加大。通过这个演示实验，我们可以知道，声源振动频率可以改变声音的高低，而振幅可以使声音的强度发生变化 (参见图 7-4)。

从图 7-2 上看到，扬声器所发出的声音的频率、振幅都是由圆轮及其转动所决定的。由于声音的波形为正弦曲线，所以这种音为纯音。在日常生活中，纯音一般是很难听到的，通常只有在实验室中由音叉或机械 (电子) 振荡器才能产生。在现实生活中，我们听到的是由若干波形混合而成的复合音，如各种汽车产生的噪音、电子琴发出的乐音和人的言语等等。纯音和复合音区分的根据是声波的线性特征，即是否可以进行波形的分解；而复合音中的乐音和噪音区分的根据是声波是否有周期性即规律性振动，当然，人的主观舒适感也是一个十分重要的指标。例如，受许多青年人欢迎的迪斯科音乐，却被某些老年人称作噪音。言语音是复合音中的一种特殊现象，它不可以简

单地视作乐音或噪音。

听觉刺激声波和视觉刺激电磁波均有三个主要特征，即频率、振幅和波形。与此相应，听觉有音高、响度和音色的区别。如果把声音和颜色类比的话，可发现二者的对应关系，见表（7-1）。

表 7-1 视觉和听觉的类同之处

感觉道		特征产生的原因
视觉	听觉	
彩色	音高	波频率的作用
亮度	响度	波振幅的作用
饱和度	音色	波混合的作用

（采自张春兴，1991）

在明白了声波及其特征之后，接下去我们分别讨论声波的三个主要特征以及与此相对应的音感。

（二）频率和音高

音高（或音调）（pitch）是人对于声波频率的主观属性，它首先和声波频率有关。频率（frequency）是物理量，指每秒振动的次数，单位为赫（或赫兹）（Hz）。声波的振动频率高，我们听到的声音就高；相反，振动频率低，听起来就低，但它们之间并非线性关系。当声波振动数大约在 16 ~ 20,000 赫时，是人所能感受到的音域。

让我们再回到图 7-2 的演示实验中，当改变圆轮的转速时，声音的高低发生了变化，这就是说频率决定了声音的音高。这里的频率是指声源每秒机械振动的周数，图 7-3 中罗列了乐器、人和多种动物发声的频率范围。从图中可知，女性的声音频率值稍高于男性，而且频率范围也要大些；动物和人的情况相差较大，最突出的是蝙蝠和海豚。蝙蝠和海豚所发出和被感知的声音频率可达 120,000 赫以上，它们的下限值与人类声音的倍音上限接近。通常，人耳接受的声音频率范围为 20 ~ 20,000 赫，40 岁以上的成人听力上限还会下降到 12,000 赫左右，甚至更低，而敏感的声音频率范围为 1,000 ~ 3,000 赫。我们所提到的声音的频率只是一种物理量，而人对它的感知则是音高。心理上的主观音高主要与声音刺激的频率大小有关，但并不是完全由刺激频率决定的，它也取决于声音刺激的强度。实验研究表明，对于同一声音刺激的音高感知，不同的人之间有着巨大的个体差异，这充分表现了音高这个心理量的主观性。在心理学中，我们根据实验研究，规定音高的单位为“美”（mel），确定 1000 美的音高为 1000 赫（声压级为 40 分贝）的声音刺激的主观感觉。这样，在声音强度不变的情况下，我们便可以对不同频率的声音刺激进行音高判断，并可探求出音高和频率间的关系。而且纯音音高和频率的相关可借助于心理物理法直接求得，即可在可听范围内把音高从低到高地分成等级，制成一种音量表。图 7-13 是一个音高量表，从图上可以明显地看出音高随频率而变的函数关系。

（三）振幅和响度

响度（或音强）（loudness）是声波振幅的一种主观属性，它是由声波的振幅所引起，振幅越大则响度越大。图 7-4 下面二条曲线虽然频率相同，但振幅不同。中间一条曲线振幅较小，即声音响度较弱；下面一条曲线振幅较大，即声音响度较强。

图 7-2 的演示实验表明，声音刺激的振幅变化，导致扬声器中声音响度的改变。声波的振幅是指示波器中显示出的纯音正弦曲线的最大高度，声波振幅的大小决定于作用在声源上的力的大小。通常讲，声源的振动频率主要决定于声音自身的属性，而声源的振幅则决定于外界施加的力——传递的能量。对于振幅，公认的测量方法是对声波的压力测量，它可用声压、声压级、声强、声强级、声功率级来度量。实际上，听觉响度常用比例表示，而不用绝对量表示。

声波造成的压力变化用分贝（decibel，简称 dB）量来测量。分贝量表是一种对数量表，它将人所能感受的巨大范围的振幅变化值压缩在较小的范围内（见表 7-2）。卡尔索（Corso，1967）通过多种实验研究指出，人耳所能探测的最强音比其所能听到的最弱音大约强 1000 亿倍。当然，不同的人对最强音和最弱音的察觉范围存在着个体差异，而且个体对各频率的声音的敏感度也不完全相同。对于分贝量，我们规定以人类能听到的平均绝对阈值，即 1000 赫附近的压力变化为 0 分贝参考点，此时的压力为 0.0002 达因/平方厘米。这样，20 分贝的声音，人的主观感觉约相当于三米远处柔和的低语声的声响，根据分贝量表其压力为我们能听到的最柔和声音的 10 倍。平常，人们的讲话声音约为 60 分贝的水平，这就比参照压力大 1000 倍。人耳能对高达 125 ~ 130 分贝的声压作出反应，如从身旁经过的火车、响雷或机枪射击时所发出的声压，但 130 分贝的声压也会使耳产生痛感，若长时间保持这样高的声压水平，人耳的听力机制就会受到损伤，这也就是纺织女工、爆破工等工种听力敏度下降的一个原因。

与声波振幅这个物理量相对应的心理量是响度。响度是振幅的一种主观属性，它主要与声波的振幅有关，但同样亦受频率的影响。测量声音响度的

表 7-2 分贝量表（dB 量表）

能量比率	分贝数	声音	能量比率	分贝数	声音
10^{18}	180	阿波罗火箭推进器	10^7	70	发动的汽车
10^{17}	170				火车站
10^{16}	160		10^6	60	谈话
10^{15}	150				棒球运动
10^{14}	140	防空警报器	10^5	50	百货商店
10^{13}	130	机枪*			集会会场
		气钻	10^4	40	住宅区交通
10^{12}	120	地下发射室			安静的办公室
10^{11}	110	雷	10^3	30	背景音乐
		锅炉厂			郊区住宅区
10^{10}	100	铆钉枪	10^2	20	轻吹的口哨
10^9	90	地铁			安静的声学实验室
		城市交通	10^1	10	树叶的沙沙声
		艇推进机	10^0	0	回音室
10^8	80	响亮的古典音乐 一般工厂			刚能听到的声音&

*痛阈；&可闻阈（采自Cohen，1969）

国际标准单位是宋（sone），一个宋为40分贝时所听到的1000赫的音调的响度。这样，在频率不变的情况下，我们就可以得到声压与响度之间的关系，以便进一步对频率、振幅和响度间的关系进行探索研究。人耳所感受到的响度大小，首先依声音的强度为转移，与强度的对数成正比；其次，不同频率的声音，若在我们主观感觉上听起来一样响，它们所要求的强度是不一样的。在实际工作中可参照等响曲线查出它的响度（参见图7-17）。

（四）波形和音色

音色（timbre）是声波波形的一种主观属性，也就是指我们听到声音在波形上的差异。不同的发音体所发出的音波都有自己的特异性。图7-5为四种不同的发音体所发出的波形，虽然它们的频率和振幅大体相同，但是它们的波形相差很大，因而我们听起来这四种音是极不相同的。这就是音色上的差异。

声波的类型是多种多样的，一般可分为纯音和复合音两大类。

1. 纯音

根据测听器的分析，波形呈正弦曲线的声音叫纯音（Pure tone），如音叉的声音和用音频信号发生器发出的声音。在自然环境中我们所能听到的声音极少是单一的纯音，而是不同频率和振幅混合而成的复合音。

事实上绝大多数发声体的振动波都是比较复杂的，都是由若干正弦声波合成的复合声波。如各种乐音、噪音和言语音都属于复合音。由于一般振幅的声波具有线性的特征，因而可以把任何复合音分解成若干不同频率的纯音（见图7-6）。

2. 复合音

复合音 (complex tone) 乃是不同频率和振幅的纯音相混合而成的音，按组成它的各纯音频率之间的不同关系可分为乐音和噪音两种：

(1) 乐音具有周期性的振动，给人以舒适的感觉：乐音 (musical tone) 乃是波形呈周期性变化的声音。乐音除包括一个频率最低、振幅最大的基音外，还包括许多与基音成整倍数的较高频率的倍音（又叫泛音或谐音）(over-tone)。基音 (fundamental tone) 是复合音中频率最低、振幅最大的成分。实际的发声体发出的声音，绝大多数是由许多频率成分组成的复合音。倍音的多少和强弱影响到声音的音色，根据不同的音色我们才能分辨各个人不同的口音和不同的乐器音。倍音越多，听者愈感到声音丰富和圆润。电子乐器比一般乐器能发出更加悦耳动听的声音，就是这个缘故。

音色是一个复杂的概念，具有多维的品质，它不但依存于倍音的结构，还受频率与振幅的影响。实验用的滤波器把各种乐器的倍音逐步滤掉，它们的声音就逐渐丧失其原有的音色，最后只剩下同一频率的基音，听者就不能分辨这些基音是从什么乐器发出来的。如果倍音逐步增加，各种乐器的音色又逐渐显示出来。音响设备的效果也与此有关。

(2) 噪音 (noise) 具有非周期性的特征，组成它的各纯音频率之间没有整倍数的关系。噪音不仅使人感到厌烦，而且还会引起听觉功能的障碍。事实上，噪音的振幅、频率和持续的时间等因素都能给人造成身心的影响，甚至导致死亡。但噪音也有它的实用价值，如临床上常用一种混合的噪音去掩蔽对侧耳的听力来测量被测耳的听力。这种噪声又称白噪声 (whitenoise)，它是由各种频率组成，各成分之间具有同样振幅在相位上无系统关系的非周期性复合音，因为白噪声像白光一样，把全部频率成分都结合在一起（见图 7-7）。

应当指出乐音和噪音只具相对意义。一方面，同一声音在不同场合可以是噪音，也可以不是噪音；另一方面，平常我们听到的乐音中总不免有噪音的成分，钢琴、无线电或留声机等发出乐音的同时也发出一定量的噪音，而一般噪音中也含有一定周期性的乐音。因此，鉴别某种声音是乐音还是噪音，既要看哪种音占优势，还要看听者的主观评价。

(3) 语音是特殊的复合音。语音是由元音和辅音构成的。元音 (或母音) (vowel) 是一种能连续发出的乐音。辅音 (或子音) (consonant) 是由声道的封闭或开放而发出的语音，包括塞音、擦音、流音、鼻音等，它主要是不能连续发出的短促的噪音。元音与辅音合成汉语音节。特殊的语音 (如汉语) 还带声调。元音是周期性的。不同的元音乃由不同的声道形式而形成，主要是通过舌和唇的动作，发元音时声带是振动的，声道是比较畅通无阻的。辅音一般是非周期性的，兼有乐音和噪音两个成分。辅音可分为有声和无声两种。有声辅音也称浊辅音，无声辅音也称清辅音。各种辅音都是由气流经过口腔或鼻腔，受到了发音器官不同部位的阻碍而形成的。要划分元音和辅音有时不太容易，为此语言学家提出了区别元音和辅音的三个标准：(1) 发元音时，气流不遇到什么阻碍，而发辅音要受到声道的阻碍；(2) 发元音时发音器官是均衡地保持着紧张的，而发辅音时，发音器官有张有弛，受阻部分紧张，其余部分松弛；(3) 元音的气流较弱，而辅音 (尤其是不带声的辅音) 气流则较强。

语音知觉是人类特有的听觉。语音虽然属于复合声，但带有它自己的特点。作为声源的言语器官是人类千万年前社会劳动的产物。我们对语音的知

觉也是在社会生活中形成和发展的，也带有它自己的特点。

(五) 声波的相互作用

一个声波同时与另一个声波相遇或与邻近的物体发生作用会产生声波的相互作用现象，下面介绍常见的声波相互作用现象。

1. 共 鸣

一个振动的物体产生的声波使邻近的其他物体也发生振动的现象，叫做共鸣（或和振）（resonance）。起共鸣的振动叫受迫振动。实验演示可将两个频率相同的音叉 A 与 B 相距数尺，用小木锤敲击 A，随即用手制止它振动，可以听到 B 在振动发音。再如，若把振动的音叉放在箱子上，会增强音叉的声音，把音叉从箱上移开，音叉的声音就微弱得多。各种乐器由于不同的结构而具有不同的共鸣的特性，不同的共鸣腔体为某些倍音提供共鸣的条件。歌唱家发出的高、中、低的每一个音都有头腔、口咽腔和胸腔这三个共鸣腔体不同比例的振动，虽然音色基本上应该是一致的，但由于三个腔体混合共鸣的作用不同，便获得优美的、富有音乐感的艺术效果。可见心理声学原理可以提高歌唱家的艺术效果。

2. 强化与干涉

二个声波间的相互作用，受声波的特征所制约。当两个振幅相差不大、频率相同、相位相反的声波合成后便相互抵消产生寂静，参见图 7-8 (c) 的情况。如果频率相同、相位一致、振幅近似相等的二个声波合成后，便相互强化产生近似两倍振幅的声波，参见图 7-8 (b) 的情况。如果频率相近，其合成波仍可看作是正弦波，只是交替地发生强化与干涉，参见图 7-8 (a) 的情况。合成波的振幅随时间发生周期性的变化，这时听到的声音叫做拍音（beat sound）。一秒钟内产生一次完全的强化或完全的干涉为一拍。每秒的拍数 V 等于两音频率的差数：

$$V = |V_1 - V_2|$$

假如将频率相近的两个正弦波同时输入示波器，可以观察到呈现周期性的强化与干涉的声音混合的合成图形。例如，两声波分别为 136 赫与 146 赫，一秒钟内会有 10 次完全的强化，10 次完全的干涉，也就是每秒 10 拍，参见图 7-9。

音调的和谐与否与拍数有密切的关系。当拍数达到 70 左右时（中音阶的 C 与 E 音）效果是和谐的。

3. 差音与和音

差音（difference tone）是两个频率适当的纯音同时发声时，可听到一个音高为两个原始纯音频率之差的第三音。差音是一组音，若 l 表示较低的原始纯音频率， h 表示较高的原始纯音频率， D 表示差音频率，则一组差音频率为： $D_1 = |h - l|$ ， $D_2 = |2l - h|$ ， $D_3 = |3l - 2h|$ ， $D_4 = |4l - 3h|$ 。和音（summation）是两个振幅大致相等，频率相差 50 赫以上的纯音同时作用于听觉器官，可以听到频率分别等于两原始音的和，即 $h + l$ ， $2l + h$ ， $3l + 2h$ 等一组音。例如，当混合两个振幅大致相等频率相差 30 赫以上的声音时，我们不仅可以听

到拍音，还可以听到另外二个音：一个等于两个频率差的差音，另一个等于两个频率和的和音。差音与拍音不同，因为两音虽然是同时听到的，但当听不到拍音以后，还可以听到差音。当听到一个比被混合的一个高频音更高的音调时，这就是和音的效果。当然混合不同的音阶的倍音同样都可以产生差音与和音，但要使它们从总的音响效果中分辨出来是很困难的，这要具备一定的实验条件和经过特殊的听觉训练才行。和视觉一样，人类听觉的潜能是巨大的。

二、听觉实验中的各种变量

第二节我们将介绍听觉现象中的主要属性及其各自的性质，在此之前，先了解它们各自在心理学实验中担任的角色，以及它们充当变量的情况，这是很有必要的。

（一）自变量

自变量（或自变项）是实验过程中主试有意加以改变的因素，通过它的变化来探寻因变量的变化。自然，因变量即是主试在实验中要观察的现象和测定的指标。任何一个自变量或因变量都不能单独存在，同时某一确定的自变量和因变量又不能离开具体的实验。某一指标在一个实验中可以作为自变量，而在另一实验中却可以成为因变量。下面，我们通过具体实验来看看听觉实验中的自变量。

听觉定位（或声源定位）（auditory localization）是听觉的方向定位。双耳是辨别声音方向的重要器官。图 7-10 是用声笼法来研究听觉定位的装置。实验中，主试将被试双眼蒙住，以去除掉视线的作用，然后用固定器固定好被试头部位置，防止实验进行中双耳位置的变化。此时，被试的头部周围有三个圆环，一个是矢状切面的圆环（即图中的深色环），与矢状在顶中相交的为冠状环，而与这两个环水平相交的为横向环，这三个环是以被试的双耳联线中心为圆心的圆球上的三个圆环。实验中，将刺激呈现器随机地在事先选择好的位置上呈现纯音、滴答声或噪音，让被试进行声音方位的判断，并作相应的记录。

听觉定位实验是一个典型的听觉实验。在实验中，一般取不同种类的声音刺激（纯音、滴答声和噪音）和不同呈现位置作为自变量，主试通过控制刺激呈现的位置变化和呈现不同的刺激内容来探寻人类在双盲情况下依据固定的双耳进行声音刺激位置判断的规律。任何一个听觉实验，都离不开声音刺激，声音刺激可以作为自变量，也可以不作为自变量而作为控制变量。在这个实验中，当其他条件不变时，我们选用声音刺激作为自变量。只有主试能控制声音刺激时，才有可能使被试进行声音定位判断，而当声音刺激成为控制变量而不起变化时，也就谈不上进行定位判断了。此外，声音刺激能由主试进行良好的操纵，这主要体现在以下几个方面：

（1）声音刺激的恒定性：在声音刺激呈现的时候，只要确定了振幅、频率（如纯音），就可以使音高、响度不变，条件均等。

（2）辨别性：我们选用自变量的声音刺激，使被试利用双耳进行方位判断成为可能。因为根据声波传导的时间特性及几何学原理，人类可以利用双

耳进行方位辨别。

(3) 可控性：对于声音刺激作自变量，主试能较严密地进行操作控制。我们既可以利用仪器设备保证自变量的恒定性，同时又能在恰当的范围内随意变化自变量。自变量的变化主要有两个方面，一为量的变化，如某一声音持续时间的长短；一为质方面的变化，声音的三个特性中任一个发生变化，就会导致声音刺激的性质发生变化。对于任何一个实验，在自变量挑选确定以后，其操作——控制呈现亦是一个至关重要的问题。同一个自变量，不同的呈现方式就有可能导致因变量产生影响。呈现位置一般取自三个圆环上以 45° 为间隔的各个位置，倘若我们在呈现位置选择不合适（如不对称），则会使我们在进行结果分析时产生差错；倘若我们对刺激的呈现序列不科学——不随机化，则被试就会掌握其规律或产生其他错觉，从而使得整个实验失去意义。从这个例子中可见，自变量的选择、控制在听觉实验中是非常重要的。

(二) 因变量

在心理实验中，主试在确定好自变量以后，接下来就是考虑因变量了，因为了解因变量（或依变项）的变化规律是实验的目的所在。就上述实验而言，根据声音刺激出现在不同的十七个位置点时，被试对声音定位的正误判断，实验者就可进行统计分析，可以得出被试对位于左右方向（即冠状面）声音刺激的辨别正确率明显高于前后方向（即矢状面）；而当分别用纯音和噪音作刺激呈现时，则噪音作自变量时的判断正确率（任何位置）都要高于纯音。当然，对这一实验结果还可以进行深入的研究。如我们可以对第一个结果的因变量记录进行数量化，即不仅只用正误判断，而改用偏离正确定位的角度记录，于是得到表 7-3。表中的数据反映了声音刺激偏离矢状面的角度与相应位置判断偏离的平均角度数。从这不同的实验结果中，我们可以看出，不同的自变量变化可以产生不同的因变量变化，而相同的自变量变化亦产生不同的因变量结果。

表 7-3 声音位置与声音定向的关系

刺激偏离矢状面的角度数	0°	15°	30°	45°	60°	75°	90°
平均错误角度数	4.6°	13.0°	15.6°	16.3°	16.2°	15.6°	16.0°

（采自张翔等，1990）

针对任一确定的自变量，如何确定因变量和利用其变化来产生所需的各种变化，完全在于主试如何确定相应的因变量指标。首先，主试要挑选能反映实验目的和要求的指标。就上例而言，主试若以反应时的长短为指标，则不能说明被试在判断方位准确性方面的情况，而仅能反映速度方面的规律。其次，主试所挑选的指标在用指导语进行解释时，要言简意赅，前后统一，使得任何一名被试能有一致的客观标准。如果不同的被试对反应方式或实验方法不能很好掌握，无疑会使实验结果产生各种误差，从而得不到真实的实验结果。再次，对于因变量的客观指标，要力求能进行明确的辨别和准确的测量。对于能够数量化的因变量，要力求用数字表示，以便于对自变量与因变量之间的关系进行定量研究。对于性质辨别（如是否判断）也应力求客观

化，以避免主观因素的影响作用。这些心理学研究的共同原则，都适用于听觉实验。

（三）控制变量

听觉实验的控制变量，主要有机体变量和操作变量。对于机体变量的控制，如被试的性别、年龄、智力等特征，可以通过实验设计的安排来进行控制，而情绪状况、健康状况等指标则可在对被试的选择时进行适当的剔除，以保证实验结果的可靠。

对于操作变量的控制，主试要适当地运用娴熟的实验技术来对实验条件和实验过程进行良好的调节。听觉实验多为室内实验，这就对实验环境提出了一些特殊要求，尤其是进行听力测量等实验，必须具备相对无噪音的环境——隔音室。隔音室亦称作消声室，它的主要用途是建立可以控制的自由声场，以便进行语言、听觉有关的测试、研究工作；测定机器和其他声源的发声特性，或其他需要避免反射声、外来声干扰和模拟自由声场的工作。例如声波的衍射、干涉等现象。隔音室内的情况接近自由声场，它除了采用必要的隔音措施抑制外来噪音的干扰外，室内的六个表面（即四壁、地板、天花板）上都敷以吸音系数特别大的结构。通常尖劈结构用得较为普遍，相对地说，此构造比较简单，效果亦比较好。通常使用的尖劈为长 1m 底部 $20 \times 20\text{cm}^2$ ，尖劈内装松乱的玻璃纤维，为了达到最佳的消音效果，玻璃纤维的密度也有一定的规定。

在听觉实验中，仪器也是一个十分重要的因素。例如在人耳听力的测量中，是用听力计发出频率、振幅通过电压加在耳机上来进行刺激的呈现。由于电子仪器能对声音刺激的频率、振幅进行随意选择，且能保持先后呈现的一致性，因而是一种较好的呈现装置，便于主试进行控制。对于一些较难重复的刺激，如言语、机械声、噪音等等，则可借助于录音机，进行再现和控制。对于实验中如何确定被试接受的刺激方式，则必须根据具体实验而定。如果是听觉定向实验，被试必然要运用双耳听觉，否则实验就无法完成，而且用扬声器和用耳机的效果是不同的。若是用单耳进行距离判断实验，则单耳的选择亦须十分小心，因为人的左右两耳在听力方面是有差别的，犹如双眼在视力上也常常是有差异的。

第二节 听觉的二个基本属性

第一节介绍声音刺激的特征时，我们就涉及到听觉具有音高和响度二个基本属性（attribute），在本节中我们再来具体地讨论一下音高、响度与声音刺激的频率、振幅、周相之间的关系，以及对音高和响度的感知规律。

一、声的测量

在讨论听觉的基本属性之前，先熟悉一下声的记录和测量的原理是很有必要的。

（一）声的记录

人类很早就开始研究声音，并利用当时已掌握了的声音的某些规律来制造乐器、进行设计建筑，使发出的声音传得更远。可是几千年来，人类只能凭耳朵来辨别声音的高低、强弱，而不能把声音记录和储存起来。所以与其他研究领域相比，声学的研究要算落后的了。直到 19 世纪爱迪生发明了留声机，人们才能用机械的方法把各种声音记录在唱片上，可是声音、机械振动不容易传递，也不容易放大，机械方法很不方便。到了 20 世纪，电学，特别是电子学的高度发展，才把声的振动先转换成电能，使声音的记录成为轻而易举的事。这也影响到人类生活的许多方面，如今人们已离不开录音机和电视机。

图 7-11 是基本的录音系统的方框图。传声器主要是用来接受空气介质中的声波，它输入的是声波，输出的是相应的电压。放大器的作用就是把从传声器来的微弱电压放大成与它成正比的较高电压。由于掺杂的干扰与声波振动信号一起被放大，所以滤波器就像过滤用的纱布一样，将有用的信号放过，而把不需要的噪音留下。虽然滤波器的输出仍是电压，但其功率还不够大，不足以能推动记录器，所以还需功率放大器来补充能量。

现代科学技术已经能够制造出理想的各种部件，也就是说这些部件能够完成人们所要求的功能。简单地说，它们是线性的，即输出量与输入量成正比。这种部件的组成系统在传递信号时不致产生失真，记录下来的声音能保存原来信号的特性，只是强度不同，并可以重新放出来。如果组成系统的部件不是线性的，输出与输入不成正比，则记录就有不同程度的失真了。

（二）声的测量

声波的测量与分析按过去传统的方法是先用电换能器把声波转换为相应的电信号，然后用电子仪表放大到一定的电压级进行测量与分析。有关声学的实验室包括消音室、混响室和行波管等，以提供测量用的行波声场和混响声场。声源可以是普通的扬声器、特殊的点声源或无指向性声源。当需要大的声功率时，可以用气流扬声器。

由于计算技术的发展，使许多计算和测量工作都使用了电子计算机或程序运算器。这些带计算机处理系统的高级声学测量仪器，能完成下列一些测量工作：

1. 评价值的测量 响度和响度级，感觉噪音级，清晰度指数，语言干扰级，噪音评价数。
2. 声源的测量 频谱的时间变化，声功率，指向性，效率，共振峰的跟踪，频谱特征和比较，平均和积分，求极大谱，幅值分布，相关性。
3. 音质的测量 混响时间，隔音量，吸音量。
4. 其他的测量 计权声级等。

声测量的基本仪器是声级计。声级计是一种能对声音作出类似人耳的反应的仪器，同时，它能进行客观而可重复的声压和声级测量。声压测量的好处很多：它能帮助音乐厅提高音响效果；能对烦扰声音进行精密的、科学的分析。声级测量还能明确地告诉我们什么声音会引起听力损害，并提醒人们

采用适当的听力保护措施。因此，声测量是不可少的。图 7-12 为声测量示意图，图中间的人正手持声级计对环境噪音进行测量。

二、听觉属性——音高

音高是听觉的首要属性，它是一种心理量，音高的单位是美 (mels)，纯音的音高依赖于声音振动的频率。下面先介绍音高是如何确定的，然后再说明音高的特征。

(一) 音高量表

纯音音高和频率的相关可借助于心理物理法直接求得，即在可听范围内把音高从低到高地分成等级制成一种音高量表。图 7.13 是一个音高量表，横坐标表示频率，纵坐标表示相应的音高，显示了音高随频率而变的函数关系。音高量表的制定，可以采用二分法和多分法两种方法。二分法是让被试者将一可变纯音的音高调到标准音高的一半，求得相应的频率。为便于音高量表的建立，一般指定 40 分贝的 1000 赫纯音的音高 1000 美作为参照点。按照二分法计算，被判断是参照音高减半的乐音为 500 美时，与其相应的频率约 558 赫，被判断是参照音高加倍的乐音为 2000 美时与其相应的频率约 2100 赫，以此推测，便可求得整个可听范围的音高量表。多分法（这里以四分法为例）是给被试者一个高频声 S_1 和一个低频声 S_5 ，让他在两者之间调出三个音，使各个相邻两音的音高距离相等，即 $S_1 - S_2 = S_2 - S_3 = S_3 - S_4 = S_4 - S_5$ ，而求得各点相应的频率值。以上两种方法所制成的量表基本相同。

(二) 音高和强度的关系

多年以来，许多学者已经注意到声音的频率并不是决定音高的唯一的因素，一个乐音的音高也随强度的改变而改变。传统实验采用可连续变换强度（可听范围）的从 150 到 12,000 赫之间的 11 个频率的声音，使两个频率交替呈现，让被试者调节其中一个声音的强度，使两音的音高相等，结果如图 7-14 所示。这个音高随强度改变的曲线又叫等高线 (equal pitch curve)。当强度改变时，各频率的音高随之发生变化。对于低音来说，音高随强度增加而降低，对于高音来说，音高随强度增加而升高，对于中等频率的声音来说，两种影响都有轻微程度的表现。例如 2000 赫的声音，当强度初增加时，音高略有升高，但强度再增加时，其音高略微降低了。从图中可以看出音高受强度改变影响最小的那些频率都是人耳对它们最敏感的声音，也是耳感受性最高的部分。

综上所述，音高主要决定于声音频率的高低，但声音强度对于音高也有一定的影响。音高的实验研究应该将声音的频率和强度结合起来考虑。当说明一个声音的音高时，应选择一个标准的强度水平作为参考，一般常用的参考强度是 40 分贝，上述音高美量表就是依此来制定的。

(三) 音高和周相的关系

音高与周相的关系同音高与强度的关系相似。非常短促的纯音音高是很难鉴别的。布鲁克、卡特沃斯基和利基特 (Burck, Kotowski & Lichte) 三人试验性地建立一个人们能进行音高鉴别的持续时间极限。他们发现,对低频纯音的区别要求有较长声音刺激呈现持续时间,在 1000 ~ 4000 赫范围内的中高频纯音,只要呈现时间持续 10 秒人们就能进行较正确的音高鉴别和归类。布鲁克在 1968 年的实验中,要求受过音乐训练的被试,对突然出现的短音音高寻找相等的纯音。实验采用匹配调整法。结果表明,当低频音的持续时间较短时,音高会明显地被判断得偏高;而缩减高频音的持续时间,则会使音高的报告值偏低;在 1000 ~ 4000 赫 (1000 ~ 3000 赫) 之间时,周相对音高的影响作用最小。这个频率范围正好是主要语言区。

(四) 音高与复合音

以上,我们涉及的均是纯音,而在日常生活中,人耳接受的却是大量纯音同时出现时合成的复合音。这些复合音包括基音和许多泛音及倍音。人能同时对许多同时呈现的纯音进行加工,也就是说,人的听觉能区别出复合音中各个独立的成分。

在 20 世纪初期,斯图姆夫 (Carl Stumpf, 1848 ~ 1936) 研究了人对泛音的鉴别能力,并得出了较低的泛音 (第二、三级的) 比更高次序上的谐音难以感知的结论。他报告说,对有些全音他能听出甚至高达 27 级部分的倍音。波罗姆 (Plomp, 1964) 报告,他系统地研究了正常听觉的被试在这方面的能力,并确立了可听范围内相当大部分的临界频率分隔点。在研究听者听到泛音的能力时,波罗姆发现,这时能听到的泛音是以最低的泛音的频率数为函数的,参见图 7 - 15。这一研究结果与早期斯图姆夫的研究完全相反。图 7 - 16 是一个被试从 250 赫的基音选择泛音的例子,整个操作持续得较好,一直进行到了第 4 级倍音,但偶然也有一下子降低到第 10 级的。

弗兰切 (Fletcher, 1953) 的许多实验证实了基音在确定一个复合音时所担当的关键作用。一个由 300、700、900 赫组成的复合音,被试报告的基音相当于 200 赫纯音的音高。这儿的基音不是由所有组成的纯音的平均值来确定的 (如这儿为 700 赫),而是由各个纯音的频率差 (200 赫) 决定的,如果在上面三个纯音中再加上 600 赫、800 赫、1000 赫,那么基音就从 200 赫降到 100 赫。

贝肯姆 (Bachem, 1959) 在本世纪中期的研究中发现,复合音的绝对音高与纯音的绝对音高不全相同,复合音依赖于声音的频率和音色。音色是指所给予声音的倍音特点的独特组合所产生的听觉效果。卡贝尔 (Korpell, 1965) 试图对音色进行评定,并观察是否它的确是由倍音的组合而产生的。他重新录制了乐音音符,将录音再以比原来录音时略慢的速度放给被试听,对他们进行进一步的训练。卡贝尔推断,如果音色确实是由倍音的组合而产生的,那么,这些重新录制的音调应当产生同样的基音,因为倍音间的关系应该是保持不变的;如果音色依赖于听觉系统内部的变化,那么被试会听到有新的频率产生的音调,因为由于播放速度的加快或减慢使得原来音符的频率发生了变化。事实上,大部分受过音乐训练的被试曾判断出有新的频率的音符,这样,音色就不仅是倍音特性的作用了。例如,平时我们发现唱机的转速太慢,这就是由于我们对歌曲的音符和速度方面的经验作用,而不是由

于我们对绝对音高变化的鉴别。

三、听觉属性——响度

响度（或音强）（loudness）是听觉的第二属性，声音响度是人耳对于声音强度反应的主观量，响度单位是宋（sone）。声音的强度可用声压、声压级与声强、声强级来量度。那么响度又是如何确定的呢？它有哪些特征？

（一）响度量表

响度又是如何确定的呢？上面讲到，以赫兹为单位的频率数不能简单地作为音高的指标，同样，以每平方厘米的达因数的声压强度也不能简单地作为响度的指标。听觉的响度，作为人的一种听觉经验，不仅与声音刺激的强度有关，也与刺激的频率有关。当刺激的频率数保持恒定时，便可以得出一个随声音刺激的强度变化而变化的响度量表——宋量表。宋量表的建立方法很多，主要有以下几种：

1. 二分法 让被试调节一个可变音，直至其响度等于两个连续音响度的中间值。例如，如果相邻两响度值代号为 2 和 4，则调节后的变量值代号应为 3。

2. 多分法 这个方法在建立音高量表时已经用过。被试按要求来调一个可变音的物理强度，直至它听起来与标准音响度的几分之一相当。这样的过程持续进行，使连续的响度值分为许多段，直至主试获得足够的数据来建立主观响度量表为止。

3. 单耳和双耳平衡法 这种方法的特点是，一个纯音同时传给双耳听，使其听起来是这个音单独传给单耳听时响度的 $1/2$ 倍、1 倍或 2 倍。因此，如果一个被试对单耳听到的一个可变音的强度判断为双耳听到的相同频率音的响度相等，这样，我们便得到了一种相对自由的判断方法。单耳和双耳平衡法是二分法和多分法的补充。

有了上述方法和原理，就不难制定出量表，图 7 - 17 即是史蒂文斯通过实验建立的响度量表。响度用宋作单位，一个宋是指 40 分贝时 1000 赫的纯音声音刺激的响度感觉。被试对与标准刺激频率相同但强度不同的许多纯音刺激进行与标准刺激的比较和判断。2 宋的值确定为 1 宋音的二倍响度而 0.5 宋则为 1 宋音的响度一半。实验者对其他由听觉刺激引起的响度感觉都作同样的规定，并让被试进行相当数量的这种不同值判断。这样，如果响度与刺激强度呈线性关系，则相关线上每点的宋值必与频率值相当，但实验结果显露出明显的非线性关系。例如一个 70 分贝刺激产生 10 宋的响度感觉，这样，70 分贝的刺激所产生的响度是 40 分贝的音的 10 倍，而不是线性相关的 1.75 倍。

（二）声强的差别阈限

用心理物理学的次数法可以测定声强的差别阈限，即呈现两个声音，让听者判断哪一个较强。有人用此法直接测量了不同频率的声音在各种感觉水平时能区分出不同响度所需要的压力变化。实验结果如表 7 - 4 所示。本实验

用从 64 赫开始相隔一个音程的八种频率，每一种频率都用低、中、高三种强度，分别为 5 分贝、20 分贝和 60 分贝。表中列出了 20 名被试 E/E 的平均值。从表中可以看出：

表 7 - 4 各频率的声音在不同感觉水平时的强度差别阈限

频率 (赫)	感 觉 水 平		
	5 分贝 E/E	20 分贝 E/E	60 分贝 E/E
64	3.6	1.1	0.20
128	2.3	0.7	0.18
256	2.3	0.6	0.16
512	1.4	0.5	0.14
1024	1.0	0.4	0.12
2048	0.9	0.35	0.10
4096	1.0	0.4	0.14
8192	1.7	0.5	0.18

(采自陈舒永，1980)

(1) 当感觉水平较弱 (5 分贝) 时被试对于频率低的声音 (如 64 赫) 比对于频率高的声音 (1024 赫) 差别感受性小得多，当频率高于 2048 赫时，差别感受性又逐渐减小。

(2) 当感觉水平增强时差别感受性变化减少，感觉水平为 60 分贝时，各种频率声音的差别感受性几乎相同，并且感觉水平高比感觉水平低时，强度差别感受性大。可见，声强的差别阈限，受到频率和强度二个属性影响。

(三) 等响曲线

等响曲线 (equal loudness curve) 乃是把响度水平相同的各种频率的纯音的声压级连成的曲线。如前所述，频率不同的声音响度，不能单纯地用声强级大小来衡量声音的响度。例如对两个频率为 1000 赫和 100 赫的声音，声压级虽然都是 40 分贝，但响度感觉却大不相同，1000 赫声音要比 100 赫的声音响得多。要使两者响度一样，就要把 100 赫声音的声强级增加 11 个分贝。因此需要确定一种响度级来量度各频率声音响度的大小。

从等响曲线上可以看出，在低强度时曲线呈 V 字型；在高强度时，响度曲线趋向平化，即在相同强度时有近似的响度。

对于响度级的研究可采用间接对比的方法进行。先选定一定强度的 1000 赫纯音作为标准刺激，用频率的声音为比较刺激，由听者调节比较纯音的强度，直至和标准纯音响度感觉相等，于是得出图 7 - 18 所示的一组曲线，图中每条曲线上各种频率的声音的响度感觉是相等的，所以称为等响曲线，这个等响曲线具体表明了响度级的特征。为了便于说明和区别各条等响曲线，同样采用分级的办法。取图中参考音调 1000 赫的垂直线与等响曲线相交点的强度级为各等响曲线的级别，叫做等响级 (equal loudness level)，单位是 方 (phon)，就是说任一条曲线上的响度级相当于同样响的参考音调 1000 赫声音的强度级 (1000 赫纯音的强度级就是它的响度级)。例如在 20 方的等响曲线上，一个强度级为 20 分贝的 1000 赫声音，与一个强度级为 37 分贝

的 100 赫声音听起来是一样响的，它们的响度级都是 20 分贝。

（四）响度与时间

响度与时间有关，对一个声音的响度进行估计必须要达到一定的持续时间。在这个最短时间值以上时，随着持续时间的增加（减少），响度也会增加（降低）。若一个人要对一个持续时间太长的音进行感知的话，则其响度也会变化。密拉贝勒和泰钦纳等人（Mirabella & Teichner et al., 1967）对这个问题进行了研究。他们要求被试对一个持续几分钟的音连续进行响度调整。结果表明，对高强度的音，为了要获得同样的主观响度，被试逐渐随时间的加长而增加音强；而对低强度的音，则结果相反：随时间的推移，音的知觉响度不断加大。

另外，响度在时间上还有总和作用。当两个不同频率的音同时呈现时，其响度要比任何一个单独呈现的响度大。根据前面讨论过的音高知觉中基音与复音的关系，我们就不会对以下的现象感到惊讶：即复音的响度感觉既不等于所有纯音响度的总和，也不等于它们响度的均数。史蒂文斯（Stevens, 1956）通过实验指出，纯音响度总和的程度，不仅与组成音的强度有关，而且还与各音之间的频率差有关，并提出了最小频率间隔的概念。所谓最小频率间隔，可以定义为产生响度总和所必须的组成音之间的频率差距（frequency spread），一旦达到这个值，总和的水平便会迅速提高。当各组成音一样响时，总和的响度值最大，当各个单音间的间距为同样的临界频宽时，亦能达到最大的总和作用。

第三节 听觉的基本实验

一、人的听力

（一）听力图

听力图（audiogram）乃是记录听力测验的图表。图 7 - 19 是不同比例成人的可感受耳（即正常）所听到的声音范围，最低的一条曲线是最小可听阈，即可闻阈限（或听觉阈限）（auditory threshold），而最上的一条则为听觉上限，即最大可听阈，亦称作痛阈。这些曲线是通过以下技术来获得的：被试在听功能较好的一只耳上戴上耳机，接受某个纯音刺激。例如，一个刺激 1000 赫并且是在可闻阈以下的音，被试不可能听到它。主试慢慢地提高声音的强度，当被试第一次报告“听到了”，主试记下这个听觉阈限水平值，并作为听力图上可闻阈曲线上 1000 赫时的一点。然后，主试继续缓慢地提高声音强度水平，当被试第一次报告耳朵感到搔痒或产生疼痛时，主试记下这时的感觉阈限值，并将其作为痛阈点。对于其他各个频率，重复进行同样的实验，直至可闻阈曲线与痛阈曲线完全形成为止。

曲线的下界是最小可闻阈，曲线的上界是最大可闻阈。常人无法听到 20 赫以下和 20,000 赫以上的声音，但对 3000 赫的声音刺激却是最敏感的，这部分也正是人类语言频率密集的区域。

图 7 - 19 所示听力图是 1969 年美国贝尔电话实验室发表的听觉阈限曲线

簇。这里的曲线系列是从大样本 (large sample) 美国成年人抽样测得的听力曲线。其中, “ 50% ” 那条曲线是最令人感兴趣的。在全城中一半的人具有高于这这阈限曲线的听觉敏锐度, 而另一半的人员则相反。由于不良听觉敏锐度能用来较好地定义失听 (或耳聋), 所以 50% 的这一曲线亦常用来作为听觉缺失的参照标准。

(二) 听觉模式

在实验室中, 我们可以将声音分解出确定的频率、强度和周相, 但是我们极少将较简单的声音进行同样的分析和处理。平时, 我所听到的是音高和响度不断变化着的一连串声音, 其中有些声音掩盖或淹没了具有相类似结构的其他声音, 同时, 具有不同结构的声音却仍旧存在。与视觉模式相类似, 声音总是被感知为一个个有意义的模式, 遵循类似于图形知觉的规律。当我们进入一个挤满人的房间, 大家在相互交谈, 你便会立刻感知到一定的噪音水平, 但如果你自己专注地加入到交谈中, 便不可能听清你交谈对象以外的其他人所说的每个字词。但有趣的是, 若有一个人在你附近同别人交谈时提到你的名字, 你便会立刻听到, 因为在这时这个刺激就像背景中的一个图形一样从许多交谈背景中突显出来。这些现象称之为听觉组织 (auditory organization)。

听觉组织现象在音乐中也可见到。如不同的乐音音阶总是以特殊的音调关系形成组群, 不属于一个特殊音阶的音能立刻被鉴别出来, 而且呈现时间愈短、愈接近的音总是作为整体而被感知。这种组织即是通常所指的节奏。

弗雷塞 (Fraisse , 1966) 研究了如何改变两个连续拍音的周相来影响听觉组织。当两个拍音间的短期间隔不超过 2 秒时, 被试总是报告节奏的印象或是相续的音符模式的组合。

近期的听觉研究工作, 主要是对环境各物体产生的声音, 利用摄谱仪来进行声音的成分测量、分析。每个声音都具有一个特征声谱, 并能通过产生各相同频率和相同强度的声音的组合来再生模仿。对人的言语亦是如此, 每个不同的个体, 每个声音都有一个特征结构。这些研究的目的, 在于揭示听觉的信息加工 (或讯息处理) 过程。

二、听觉掩蔽现象

听觉掩蔽 (auditory masking) 是两个声音同时呈现时, 一个声音因受到另一个声音影响而减弱的现象。在日常生活中经常可以遇到声音的掩蔽现象。一个可听声由于其他声音的干扰而使听觉发生困难, 前者必须增加强度才能重新听到, 这种阈限强度增加的过程和强度增加的量就叫声音的掩蔽效应。要听的声音叫做被掩蔽音, 起干扰作用的声音叫掩蔽音, 影响掩蔽效果的有频率、强度等因素。掩蔽现象大约有三种情况, 下面分别进行阐述。

(一) 纯音掩蔽

对听觉掩蔽的研究是从纯音掩蔽 (pure tone masking) 开始的, 即以某个定额频率的纯音来掩蔽其他不同频率的纯音, 再来观察后者阈值提高的情

况。图 7-20 是佛莱奇尔 (Fletcher, 1953) 的一个实验结果, 从图上可以看到以下几种情况: (1) 掩蔽音强度提高, 掩蔽效果随之增加, 当 400 赫的掩蔽音是 40 分贝时, 800 赫的纯音要达到 13 分贝时才能听到; 当该掩蔽音提高到 80 分贝时, 800 赫的纯音须增加到 60 分贝才能听到, 而且掩蔽音愈强, 它的影响范围也愈大。例如 20 分贝的 400 赫, 掩蔽音只影响到 200~800 赫的频率范围, 而 80~100 分贝的 400 赫掩蔽音可影响到 4000 赫以上的频率范围。(2) 掩蔽音对于频率相近声音的影响最大。例如 3500 赫掩蔽音对于 330~4000 赫纯音的影响明显大于 3000 赫以下纯音的影响。(3) 低频对高频的掩蔽效果大于高频对低频的掩蔽。例如 400 赫掩蔽音对高频音的影响范围和效果相当大, 而 3500 赫掩蔽音对低频音的影响范围和效果就相当小。所以在生产劳动与无线电通讯中, 应当着重考虑排除低频音的干扰作用。

(二) 噪音掩蔽

上述是纯音对纯音的掩蔽作用。在实际生活中, 更常见的是噪音的掩蔽作用。图 7-21 是一种白噪声对纯音掩蔽的实验结果。从图上可以看到, 当噪音强度低的时候, 各种纯音的阈限差别很大, 当噪音强度提高时, 各种纯音的阈限差别缩小。在制订克服声音掩蔽的最佳方案时, 这类参数对通讯有很高的实用价值。

依根等人 (Egan et al., 1965) 曾使用窄频带噪音作为一种掩蔽刺激。图 7-22 表示依根等人的一组实验结果。使用的噪音带宽为 90 赫, 中心在 410 赫。图中表明对一些被掩蔽的频率 (横轴) 所产生掩蔽总量。当噪音在低水平 (40 分贝) 和中水平 (60 分贝) 时, 所产生的掩蔽总量相当对称。但是, 在 80 分贝水平时, 掩蔽噪音便对较高频率比对较低频率产生的掩蔽作用范围更大些。

(三) 噪音与纯音对语言的掩蔽

噪音和纯音都可对人们的语言产生干扰, 使语言变得不可理解。语言通讯一方面决定于听者耳朵的敏感度, 另一方面决定于说话者的语言特性, 所以了解语音的频率和强度分布是很重要的。图 7-23 是贝尔电话实验室的佛莱奇尔 (Fletcher, 1953) 的测试结果。在距离说话者的口唇 30 厘米的地方测量相连的每个音阶 (在低频段) 和每半个音阶 (在高频段) 的声强, 实线是男声, 虚线是女声, 均表示在每一频带中语音的强度比所有各频率的语音的平均强度低多少分贝。从图上可以看到, 在语音频率范围内各个频带的强度是不同的, 强度最高的频带在 300~500 赫左右, 600 赫以上强度逐渐减低, 超过 5000 赫强度就减到非常小了。我们知道, 人耳对 3000 赫左右的声音最敏感, 可是强度最高的语音频带处于 300~500 赫, 因此, 在语言通讯中使用的最重要的频率并不就是耳朵最敏感的频率。对于一般的语言通讯来说, 去掉 2000 赫以下的频率, 就会无法听懂对方话的意思。相反, 去掉 2000 赫以上的频率, 却不会影响对语言的理解。应注意到图 7-23 中两条曲线的尾端男声和女声强度的差别。为了设计好一个语言通讯系统, 就需要有各种专门的数据。

掌握语言音谱的特点, 有助于我们了解噪音和纯音对语言的掩蔽效果。

噪音的掩蔽效果比纯音的掩蔽效果好，但并不是所有强度的噪音都能造成同样干扰。噪音要相当大，大到叫人厌烦的程度，才会降低语言的清晰度或可懂度。在纯音的掩蔽效果中，300 赫比 1000 赫掩蔽作用大，所以要注意排除低频音对语言的最重要的频率干扰，以保证语言的清晰度。

三、听觉的疲劳和损伤以及适应

疲劳和适应是所有感觉系统所共有的特性，听觉系统在足够长和足够强的刺激作用下，其感受性将发生变化。

（一）听觉疲劳

听觉疲劳 (auditory fatigue) 乃是声音刺激强度大大超过听觉感受器的正常生理反应限度，或声音刺激长时间作用于听觉器官而引起的听觉阈限暂时提高的现象。听觉疲劳测量方法可先测定被试对某种频率声音的阈值，而后让他听一段时间引起疲劳的特定频率和强度的纯音，再测定他的听阈，所得阈值的改变量，即暂时阈移 (temporary-threshold shift, 简称 TTS)，就是听觉疲劳的指标。

暂时阈移 (TTS) 的大小受多种因素的影响，主要有：

(1) 暂时阈移的大小，和引起疲劳的声音停止多少时间有关。引起疲劳的声音停止作用的时间愈长久，即恢复时间愈长久，则暂时阈移愈小。

(2) 暂时阈移一般随疲劳声强度的增加而加大，当疲劳声在低强度时，阈移变化相对小些；当疲劳声强很高时，阈移增加很快。1400 赫的暂时阈移大于 1000 赫的暂时阈移，当两者在 90 分贝处，暂时阈移急剧增加，这一转折点可能就是能恢复的疲劳和不能恢复的听力损伤的一个分界线。目前许多国家将允许噪音标准定在 80 ~ 90 分贝之间，其根据就是部分来自对暂时阈移的大量测试的结果。

(3) 暂时阈移和疲劳声作用时间的久暂有关，据研究，暂时阈移的大小和作用时间的对数成正比例关系。

(4) 频率在 4000 ~ 6000 赫的高频高强度的疲劳声对暂时阈移的影响最大，不可恢复的听力损失也最为厉害。

（二）听力损伤

听力损伤 (hearing damage) 乃是声强超过听觉系统正常生理反应程度的声音，持续作用于听觉器官造成的听力下降。通常听力平均损失大于 25 分贝，即为听力损伤。在职业性听力影响的情况下，防止听力损伤应当视作劳动保护的目标。

听力损伤主要有两种类型：一为传导性耳聋 (conduction deafness)，即由于听觉系统传导机能的缺陷所致；另一种为神经性耳聋 (或中枢性聋) (central deafness)，即由听神经系统的损伤所致。传导性耳聋，其实质为外耳或中耳的机能紊乱，阻止或妨碍了声波适当地传导到内耳的感受细胞。神经性耳聋则为毛细胞及其神经联系受损伤的结果，它不同于传导性耳聋，而是一种无法恢复的听力缺失。

传导性耳聋是一种常见的病症，典型的症状是外耳道或鼓膜或中耳的听小骨丧失机能。对于正常人，若我们用蜡堵住耳朵，听力的敏度便会丧失。但若鼓膜被轻扯或稍刺一下，听觉丧失通常不算严重。这两种情况一般容易矫正，蜡可以移去，鼓膜将会痊愈。如果鼓膜穿孔过大，形成疤痕，其结果是听力敏锐度永远丧失，这主要是由于鼓膜将听觉信号转变为使听小骨产生运动振动的能力丧失了。由于听小骨损伤而造成的传导性耳聋，会产生重大的长期的缺陷，最严重的缺陷叫做耳硬化症。有这种疾病时，小骨链失去正常活动的灵活性，镫骨的底板紧粘在卵圆窗上，最终的结果是听小骨传导声音振动的能力受损或完全丧失。如果仅是能力受损，则助听器可以使声音刺激加大，仍能经过正常的管通道传导，使问题得到缓解。在更为严重的病例中，如镫骨底板固定在卵圆窗上，则需要手术治疗。有种最新发明的治疗耳硬化症的手术叫做底板切除术，它去掉整个的镫骨，以人造的支柱代替，这种手术经过一千多例临床验证，大约可使 80% 的患者听力敏锐度基本恢复正常水平，但仍可能丧失 15 分贝的听力。

神经性耳聋乃是听力损伤中最严重的类型。当听觉的神经机制受到破坏时，这种听觉损伤是不能恢复的，助听器也不能强使不复存在的细胞发放冲动。产生神经性耳聋的原因各不相同，可以是长期处于过度噪音水平中的结果（刺激性耳聋），这种病例可在长期听摇摆乐曲和常乘履带式车辆的人中发现；某些药物，包括人们熟知的链霉素等，当使用过量和持续时间过长，亦会导致神经性耳聋。此外，老年人会产生一种称作老年性耳聋的神经性耳聋。老年性耳聋是一种正常现象，它主要是对高频音敏感性的丧失，并随年龄增长逐年加剧。贝克赛（Bekesy, 1969）经过调查和实验研究发现，事实上，当人在 40 岁以后，他的听力上限每半年降低 80 赫。具体的实验结果见图 7-24。

（三）听觉适应

听觉适应（auditory adaptation）乃是持续的声音刺激引起听觉感受性下降的现象。听觉系统一般对一个稳定声的感受性在最初 1~2 分钟内有所下降，而后很快稳定在一个水平上，听觉适应的特点就在于它是一个平衡过程。听觉适应的研究法是响度平衡法。以一定声强（如 80 分贝）的纯音作用于左耳，用另一频率相同但声级可变的的声音同时作用于右耳，使两者等响（对一个正常听者，两者平衡的声级可能相等）。然后，将右耳的声音停止，让左耳继续听 3 分钟。在这一适应期后，重新使左右耳等响，这时右耳的等响级常下降，如降到 60 分贝，适应量为 $80-60=20$ 分贝。

实验表明，高声级和低声级都可产生适应，且适应范围有随响度加大而扩大的趋势。适应效果对频率带有选择性：在 125~1000 赫之间，适应随频率增加而轻微扩大；但在 1000~8000 赫，听觉系统保持恒定。最大的适应发生在和适应声相同和相近的频率。

四、听觉定位

在第一节谈到自变量时，我们列举了听觉定位的实验，对此，大家已有了初步认识。听觉定位（auditory localization）是指利用听觉器官判断发

声体的空间方位。一般说来，听觉的定向与定位不如视觉准确。在多数情况下，我们往往先听到刺激物的声音，然后转动头用耳朵与眼睛去寻找声源，最后利用视觉对刺激作出更准确的定位。例如，在战场上，如果敌人躲在隐蔽的地方射击，战士往往先闻枪声，然后用眼睛去寻找敌人，进行反击。

这里，不妨把视觉和听觉作一比较，在视觉中，双眼往往比单眼能更完善地知觉对象的深度，而单眼能如双眼一样准确地判断对象的方向。在听觉中，情况恰恰相反，要从听觉线索中判断发声体的方向，必须利用双耳，而判断距离或深度，单耳如同双耳一样好，见表 7-5。

表 7-5 单、双眼和单、双耳定位能力比较

判断感觉道	方 向	深 度
视 觉	单眼、双眼一样好	双眼好
听 觉	双耳好	单耳、双耳一样好

(采自杨治良，1980)

对声源方向的判断，主要有三种双耳（利用两只耳朵）线索：强度差、时间差和周相差。下面分别予以叙述。

（一）双耳强度差

当双耳离声源的距离不同时，会产生强度上的差异。声源很少发自人体的正中面，这样它与双耳的距离之差就产生双耳声强差（intensity difference of the two ears）。向头部投影一个声影（sound shadow）（类似于光的影子），与声源方向相反的一耳处在声影之中，从侧面来的声音必须绕过头部才能到达另一耳。在声音到达之前，许多声波已被头部与其周围物体吸收，因此到达另一耳的声音强度相对比较弱。

图 7-25 是水平面上不同方位声源所引起的双耳强度差。从图上可见，当耳轴水平面上，和前方成 60° 和 120° 时，双耳强度差最大。强度差不仅和方向有关，还和波长有关。波长愈短（即频率愈高），双耳强度差亦愈大。这种情况类似湖面水波，短波易为障碍所阻，而长波可以顺利地绕过障碍物。

（二）双耳时间差

来自非正中面的声音，不仅会产生双耳强度差，而且还会产生双耳时间差。这一原理易被人理解。双耳时间差（time difference of the two ears）是辨别声音方向的重要线索。人体头部近似球形，两耳间的半圆周约为 27.6 厘米，声音到达两耳的时差的最大值（即与人体正中面成 90° 时）约为 0.5 毫秒。假如声源位于正中面上（如正前方、正后方），声波同时到达两耳，时差为零。其他情况则介于零和极大值之间。听分析器正是利用这时间上的差别，来确定声源的方位。

（三）双耳周相差

如果声源不在正中面上，不仅会产生双耳强度差和双耳时间差，而且还会出现双耳周相差（phase difference of the two ears）。这是因为声波

是由一系列的正压和负压组成的，因此任何瞬间，最大的正压到达两耳的时间不同，声调在两耳可能产生周相差。在日常生活中，双耳周相差线索在低频上较为有效，因为两耳的距离有 27.5 厘米，距离相当大，容易显出周相差；相反，高频的波长短，周相差出现在高频的机率就较小，可靠性差些。

不难理解，上述三种双耳线索总是融汇在一起的，它们之间各自提供着既有联系又各不相同的信息。例如双耳时间差和周相差既有区别又有联系，1000 赫的连续音的音调每周时间为 1 毫秒，因此时差 0.5 毫秒为半周，周相差为 180° 。假如频率为 500 赫，同样时差为 0.5 毫秒，周相差就变为 90° 了。周相差的单位是度，时差的单位是毫秒。总之，三种双耳线索的协同作用，使我们能正确地进行听觉定位。

（四）立体声听觉

立体声听觉是利用双耳强度、时间和周相差异的原理产生的。立体声广播和立体声电影等乃是采用特定的技术建立听觉透视 (auditory perspective) 的错觉效果。这种特定的技术主要有三种：双耳记录法、立体声记录以及半立体声记录。我们知道，在任何复杂的情境中，总是有些声音出现在远方，有些声音在近处，有些在左面，有些在右面。因此，声音存在着多维关系。所谓立体声效果，就是充分利用双耳决定声源的位置。

影视中的“听觉透视”乃是利用两个或两个以上的扬声器来产生的。如图 7-26 所示，a 与 b 为摄影时声音发出的位置，x 与 y 为当时记录的录音机放置的位置。x 受 a 的声音的影响大些，y 受 b 的影响大些。在放映电影时，两个扬声器仍置于 x 与 y 处，观众就重建原来的情境，好像听到从银幕上的 a 与 b 发出的声音。这样，听觉错觉与视觉错觉同步配合，在二个感觉道上协同活动，产生特定的听觉立体声效果，即“听觉透视”效果。通过建立在心理声学原理基础上的艺术上加工，给人类产生听觉上美的享受。

五、可见言语

顾名思义，语图仪 (language pictorial display apparatus) 是能将复合音或语言分析为组成成分频率，显示频率-强度-时间型式变化的仪器，它能形象地显示言语听觉特征。通过语图仪可以使人清晰地看到言语的形象，正像听到了言语的声音一样，故称可见言语 (visible speech)。语图仪的诞生具有很重要的理论和实践的意义。它既能形象地分析出言语听觉所依据的各个特征，使人可以探索语言的哪些因素影响言语清晰度，这便有助于设计通讯系统，来提高言语清晰度，以及如何用人工改造言语和隐蔽通讯语言等等。可见言语在训练与说话有关的专业人员时，有助于改正言语的质量，另外还可以训练聋人学习这种可见言语，用以辨别他人的讲话，并能通过把自己的语音图与正常人的语音图比较，来改进自己的言语。可见言语还为人工合成言语创造了条件。

可见言语的转换原理并不复杂，大致是：说话者对一传声器说话，传声器连接在精密的电子仪器上。它通过 12 个不同的带通滤波器，把言语声音的全部频率 (30 ~ 20,000 赫)，分成 12 个波带，参见图 7 - 27。从不同滤波器通过的能量，经过放大后，分别地控制着排成一行的 12 个小灯的光强。在

这些灯光前面有一块萤光屏，以一种恒定速度向一个方向移动着。在萤光屏移过各灯光时，便在与小灯相应的水平上发出萤光，其萤光的强度也与各小灯的亮度相应。荧光的余辉较长，在移过小灯位置之后并不立即熄灭，而是延续一段时间。因此，语音的图象可以呈现 2 到 3 秒之久，从而使人能看一个词或短句的确切形象，也可以拍成连续或单张的照片。

因此，语图仪可以使人们清晰地看出言语。图 7 - 28 是由语图仪描绘出来的“阴阳上去无”五个汉字的语图。我们从图 7 - 28 上可以看出：（1）语音高低的不同，主要是在基频上。语音图中上下各水平代表音高或频率的不同；（2）黑线密集的程度代表各不同频带的分布情况；（3）图形的横轴代表时间，从左向右进展。

通过对不同人的语音图的分析，发现带有不同方言说话者的语音图，虽然有着不同频带强度上的差别，但言语的总模式却是相似的。任何高低不同的方言说同样一句话，都表现出一种共同的可以认出的模式。

如将传声器收得的言语电能通过各带通滤波器，计算并平均各滤波器的输出，总在一起，便可绘出言语声谱曲线。有人绘制了英语平均言语谱如图 7-29。图上曲线表明，不论是男还是女说话者所得的数据，在一段长时间之内，言语能量大部分都落在 1000 赫的频率之下。在较短时间之内，言语能量的分布可能很不相同。正是这些言语能量分布的变化给我们提供了知觉的线索。图形的分析还可以提供作为时间函数的言语能量分布变化的更详细的形象。图 7 - 30 表示英语单词“out”中元音在连续时间之内一系列的谐音的分析。纵立粗线的高度表示在各种频率振动成分的幅度。虚线表示共振的范围，或者在整个发音过程中连续变动的共振峰。当然，整个句子也可以使用这类动态分析，不过这样要复杂得多。

六、汉语语音

近几年来，海峡两岸的心理学家在“心理学本土化”的鼓舞下，对汉字心理的研究有了很大的发展，在汉语语音方面也取得了很大成绩，揭示了不少汉语语音的特点，现介绍如下：

（一）汉语语音特点

和世界上许多语种的语音相比较，汉语语音最明显的特点之一就是，在汉语音节中元音占优势，而在许多外国语中则不然。因此，研究汉语中元音的物理特性，无论对于汉语语言规律的探索，还是为我国通讯工程作参考，都是十分重要的。根据现有的一些材料，已可看出汉语和英语在元音频率特征上有一定的差别。在汉语的标准发音中，每个单元音的吐字都十分清晰明朗，英语则含混得多。这方面的差别可能与两种语言频率特征上的不同有一定的关系。其次，在基频的范围方面，汉语与一些欧洲语言也有区别。例如汉语在声调不变化时，基频的范围无论男女都比瑞典语及英语高。这与我们平常感觉外国人说话声音比较低这一事实是相符合的。

（二）汉语语音结构特点

汉语语音的结构上有它独特的性质，这主要表现在有语音的结合和声调（四声）上，以下分别说明。

1. 语音的结合

在汉语里一个方块字就是一个音节。音节是由一个或几个音素组成的最小的语音片断。无音和辅音是就单个的音说的，是语音学分析的结果，而不是听觉的直接感受。音和音连结起来形成的音节，才是可以理解并能凭听觉或肌肉感觉觉察到的最小单位。在一个音节的组成上，具体到汉语单字上，一个字起始的音称为声母，其后的音素则称之为韵母。例如汉语说“东北”[dong bei]，[d]和[b]都是声母，其余的则是韵母。声母一般都是辅音（零声母——[i][u][y]等起头的字除外）。韵母有的由元音组成，如“北”字中的[ei]；有的由元音加辅音组成，如“东”字的[ong]。声母和韵母拼在一起组成单个的字。从声母到韵母的收尾整个字音及其中辅音和元音的延续时间，随说话者和具体的字音而有差别。即使同一个受过训练的说话者，在说不同的字音时，延续时间也不相同。表 7 - 6 举出同一个人发几个不同字音时的延续时间（毫秒）为例。表中字音的声调都是阴平（第一声）。从表中可以约略看出辅音和元音在拼成的字音中的时间关系。从表中还可以看到，当字音较长时可以明显地听出音渡（tonal transition）（又称过渡音或过渡成分）的延续时间。当然，这里所举的是孤立地发出这些字音的情况。在整个句子里，由于上下文、意义、语气等的影响，字音及其组成部分的延续时间必将还有复杂的变化，那必须进一步作动态的分析。例如通过上述的语图仪，就可观察到种种变化。

表 7 - 6 汉语一些音节中辅音、元音和过渡成分各自长度举样（单位毫秒）*

音节	全长	辅音	音渡	元音	音节	全长	辅音	音渡	元音
ba	407	6	6	395	Jia	450	31	31	388
ma	459	27	9	423	xia	494	144	22	328
fa	516	62	28	426	zha	494	85	62	347
da	391	17	11	363	cha	447	64	41	337
ta	372	18	24	330	sha	483	99	42	342
na	455	79	36	340	ca	416	51	48	317
la	564	51	131	382	sa	650	94	49	507

*为了方便，这里使用了汉语拼音；不同人或不同情境中发音长度可能不同。

（采自李文生、郭念锋，1980）

2. 声 调

世界上有许多语言（如英语、俄语）特别注意重音，汉语则特别注意声调。因此，声调是汉语的重要特征之一。例如“妈”、“麻”、“马”、“骂”四个字的声母和韵母的组成完全相同，仅因为声调不同，意义也就各不相同了。汉语发音有阴平、阳平、上声、去声四个调，如表 7 - 7 所示。表中同一横行声母韵母相同，声调不同；同一纵列声调相同，声韵不同。声调既然是音高的变化，它便是由发音时声带的松紧决定的。

发音时声带松弛，声调就低；声带紧张，声调就高。声带先松后紧，声调由低变高；声带先紧后松，声调就由高降低。声调可分为“低”、“半低”、

表 7.7 汉语四个声调及其例字

字首	阴平	阳平	上声	去声
pao	抛	袍	跑	炮
fang	方	房	纺	放
qi	欺	旗	起	气
hui	灰	回	毁	惠

(采自李文生等， 1980)

“中”、“半高”、“高”五个水平。汉语阴平从头到尾都是高的，不升不降，名曰高平调。阳平从中升到高，是高升调。上声从半低降到低，再升到半高，是降升调。去声从高降到低，是全降调。

汉语的调号有二种表示方法。一种是用国际音标方法，右上用附加数目字代表调号。例如“高”[gao1]，“肥”[fei2]，“矮”[ai3]和“瘦”[shou4]。另一种则在元音字母上加“-”“’”“ˇ”“`”分别表示四声，如“高”(g o)，“肥”(féi)，“矮”(i)，“瘦”(shòu)。这里必须提出，声调的高低乃是相对的，即同一个人说同一个音节时前后相对的声音高低。因此，声调是不受男、女、老、幼的发声影响的。男人说的阴平字，女子说来也是阴平声调；大人说的上声字，小孩说来也是上声调。总之，发声尖或粗(绝对音高)因人而异，但声调的高低升降(相对音高)则是应该共同遵守的。

声调的不同可导致韵母的长短有所差别。一般说来，汉语上声、去声字比阴平、阳平字长些。例如“可”(k)字的元音比“科”(k)字长，动程也宽些。用语图仪分析可以进一步揭露调性的长度特点。例如 a、i、u 三个元音表现出声调的阈限音长值是不同的，u 为 76 毫秒，i 为 39 毫秒，a 为 14 毫秒。各元音声调的阈限音长依次排列为第一声最短，第四声次短，第二声略长，第三声最长。第一声前后没有声调高低的变化。第二声和第四声是同一类型但调式相反的一对音，一是升调，一是降调。实验结果表明，升调要求的阈限音长稍长些。第三声是先降后升，变化比较复杂，阈限音长也最长，从语图仪分析来看，它并不等于第四声和第二声的机械相加，因为它的阈限值或大于第四声和第二声阈限值相加，或小于这两个阈限值之和。可见它的调型是独特的。汉语的四声构成了汉语的一大特征。

汉语所据有的声调的重要性，要求说话者在学习汉语的过程中在神经系统的听觉、视觉和动觉(喉头、口腔肌肉感觉)间建立起相互辨别的暂时联系。这种联系与日常使用外国语言的人所建立的有所不同。根据语言是思维的外壳这一原则，可以推想，从言语反映到思维上的特点方面看，说汉语的人也必然有其特点，这是有待于语言学和心理学协作进一步探索的广阔而重要的领域之一。随着认知心理学的发展，一门崭新的学科——心理语言学也就油然而生了。

本章实验

一、听觉定位实验

- (一) 目的：了解听觉定位现象；学习测定方法。
- (二) 材料：音笼及附件，遮眼罩，讯响器，米尺。
- (三) 程序：

1. 准备工作：

- (1) 学习使用音笼呈现不同方位的声音的方法。
- (2) 声源所处的方位，在本实验中分为两大类：第一类声源是在通过额头的中心和鼻尖与地面垂直的平面上；第二类声源是在通过两耳的中心与地面平行的平面上。第一类声源方位包括七种：前下、前、前上、上、后上、后、后下。第二类声源方位则包括八种：左、左前、前、右前、右、右后、后、左后等。
- (3) 使两类声源中的每种方位出现 20 次，第一类声源共作 140 次，第二类声源共作 160 次。每一平面上各刺激位置呈现的顺序要随机排列。

2. 对垂直平面声源的定向实验：

(1) 被试者戴上遮眼罩（或蒙上眼睛），坐在音笼的凳子上，把被试者的头用头夹固定起来。在实验过程中，当被试者听见声音时就报告它是从什么

声音定位记录用纸（每次在方格中作一记号）

声源 定向	右	右后	后	左后	左	左前	前	右前	下前	上前	上	上后	下后
右													
右后													
后													
左后													
左													
左前													
前													
右前													
下前													
上前													
上													
上后													
下后													
总计													
正确数													

方向来的，即报告：“左”、“右”或“右上’等等，主试记下来。如当时被试者是借额外的暗号（如主试者移动的声音等）来判断声源方位，须立刻记录，出现了这样的情况，最后把这一次补作一下。

- (2) 被试者只对本实验用的信号反应，实验开始前让他熟悉这个信号。
- (3) 每做完 20 次，休息 1 分钟。

3.对平行声源的定向实验:用同法按已排好的顺序,对8种声源作160次测定。

(四) 结果:

- 1.计算各被试对刺激呈现的每一位置判断正确的次数的百分数。
- 2.把在两个平面上所得结果分别列出表格,并画出直方图。
- 3.整理错误判断超过50%的方位的结果。

(五) 分析:

- 1.根据本实验结果说明听觉定向准确性与声源方位的关系。
- 2.分析不同被试者错误判断的结果,试解释其可能原因。
- 3.试说明本实验的实际意义。

二、可懂度实验

(一) 目的:

- 1.测定保持语音可懂度的语音强度范围。
- 2.学习采用“语音平衡词表”测定语音可懂度的方法。

(二) 材料:印好的语音平衡词表、记录用纸、铅笔、录音机及用汉语灌录的“语音平衡词表”磁带、声级计。

(三) 程序:

1.准备工作

(1)选择被试:经听力检查无听力障碍者,并要求懂汉语,参加实验的被试以20~30名为宜。

(2)选择适当的测听室,尽量保持室内简单、整齐、清洁,室内噪音低于45分贝。如有条件可在隔音室内进行。

2.被试坐位,距录音机3~5米为宜,实验时要求仔细倾听放送的单音节词并尽快把听懂的词记录下来,听不清的,在记录纸上画一“0”。

3.主试记下每一次放送时的语音强度。可选择轻(约30分贝)、中(约50分贝)、强(约70分贝)三个强度等级。

4.每测完1组词表后(每组词表为25个单音节词),让被试将记录与印好的词表核对,听不清和听错的均算错,同音字可以按正确计算。计算出每组词表的可懂度数值(即听懂的词数占总词数的百分率)。

5.休息片刻后,继续测听。

(四) 结果:

- 1.分别统计出各音强组的可懂度。
- 2.找出可懂度在50%以上的语音强度,按可懂度数值,由低至高进行排列。
- 3.找出可懂度最高一组的语音强度。

(五) 讨论:

- 1.确定可懂度与语音强度的关系。
- 2.对本实验用的言语测听材料进行评论。

可懂度实验记录用纸

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
字表一																										
字表二																										
字表三																										
字表四																										
字表五																										

注意：请按顺序记录，听不清的，在记录纸上画一“0”。

附录：

本实验用的测听字表，可请汉语语音教师或播音员灌制。制作过程要注意音量控制，最好能在隔音室内进行。发音音量由发音人自行调节。磁带经过复制加工，使字与字之间的声压级相差不超过正负2分贝。字与字之间的间距为3秒。磁带开始部分录1000赫纯音约4秒钟用作测听器上的零级校准。纯音后面录有10个字（草 cǎo、棉 mián、秀 xiù、宗 zōng、停 tíng、勇 yǒng、裤 kù、抓 zhuā、钩 gōu、觉 jué）作为测听练习用。练习字表后是录制的正式测听字表。

发音人录音用的五组字表

字表一	字表二	字表三	字表四	字表五
风 f ng	分 f n	定 d ì ng	方 f ng	党 d ng
假 ji	华 hu á	别 bi é	生 sh ng	挤 j
脚 ji o	倍 b è i	灯 d ng	过 gu ò	才 c á l
办 b à n	地 d ì	给 g i	点 di n	不 b ù
读 d ú	洗 x	放 f à ng	席 x í	确 qu è
亲 q n	苦 k	信 x ì ng	决 ju é	流 li ú
使 sh	权 qu á n	和 h é	借 ji è	多 du
快 ku à i	见 ji à n	麻 m á	众 zh é ng	干 g à n
助 zh ù	井 j ng	起 q	五 w	能 n é n
位 w è i	打 d	穿 chu n	吨 d n	发 f
国 gu ó	完 w á n	山 sh n	床 chu á ng	好 h o
双 shu ng	艾 à i	肿 Zh ng	夜 y è	破 p ò
红 h ó ng	张 zhang	自 Z ì	米 m	向 xi à ng
照 zh à o	追 zhu	交 ji o	在 z à i	雨 y
有 y u	拾 sh í	温 w n	手 sh u	如 r ú
香 xi ng	灭 mi è	云 y n	炮 p à o	吃 chi
派 p à i	油 y ó u	煮 zh	言 y á n	加 ji
四 s ì	尺 ch	是 sh ì	病 b ì ng	同 t ó ng
门 m é n	个 g è	右 y ò u	他 t	死 S
英 y ng	扑 p	土 t	强 qi á ng	镇 zh è n
甜 ti á n	层 c é ng	靠 k à o	内 n è i	受 sh ò u
断 du à n	晒 sh à i	来 l á i	今 j n	围 w é l
厂 Ch ng	师 Sh	错 cu ò	算 su à n	妙 mi à o
寸 c ù n	造 z à o	耳 r	啃 k n	蒸 zh ng
茶 ch á	绿 l	片 pi à n	粮 li á ng	按 à n

(采自沈晔、王书鑫，1983)

本章摘要

1. 听觉是声波作用于听分析器所产生的感觉。听觉的适宜刺激是 16 ~ 20000 赫兹的声波。音高、响度、音色等音感都是由声波的频率、振幅与波形等物理特征决定的。

2. 音高是人对声波频率的主观属性，它首先与声波的频率有关。频率是物理量，指每秒振动的次数，单位为赫兹。声波的振动频率高，我们听到的声音就高，反之亦然，但它们之间并非线性关系。

3. 响度是声波振幅的主观属性，它是由声波的振幅引起的。振幅越大则响度越大，但它们之间也不是线性关系。

4. 音色是声波波形的主观属性。不同的发音体所发出的音波都有自己的特异性。声波的类型是多种多样的，一般可分为纯音和复合音两大类。

5. 语音是特殊的复合音。语音由元音和辅音所构成。元音是一种能连续发出的乐音，辅音主要是不能连续发出的短促的噪音，元音与辅音合成汉语音节。

6. 一个声波与另一个声波相遇，或与邻近的物体发生作用会产生声波的相互作用现象。一个声波使邻近的其他物体产生振动的现象，叫做共鸣。当两个振幅相差不大，频率相同，相位相反的声波合成后便相互抵消。如果频率相同，相位一致的声波合成后便相互强化。

7. 听觉实验多为室内实验，这就对实验环境提出了一些特殊要求，尤其是进行听和测量等实验必须具备相对无噪声环境——隔音室。

8. 声测量的基本仪器是声级计。声级计是一种能对声音作出类似人耳的反应的仪器，同时，它能进行客观而可重复的声压测量。声压测量具有很高的应用价值。

9. 纯音音高和频率的相关可借助于心理物理法直接求得，即在可听范围内把音高从低到高地分成等级制成音高量表。为便于音高量表的建立，一般指定 40 分贝的 1000 赫纯音的音高作 1000 呔为参照点。

10. 等响曲线乃是把响度水平相同的各种频率的纯音的声压级连成的曲线。

11. 听力图乃是记录听力测验的图表。曲线的下界是最小可听阈，曲线的上界是最大可听阈。

12. 听觉掩蔽是两个声音同时呈现时，一个声音受到另一个声音影响而减弱的现象。要听的声音叫做被掩蔽音，起干扰作用的声音叫掩蔽音。

13. 听觉定位是听觉的方向定位。它主要通过三种双耳线索（强度差、时间差和周相差）来实现。

14. 语图仪是能将复合音或语言分析为组成成分频率，显示频率——强度——时间变化的仪器，它能形象地显示言语听觉特征。

15. 汉语语音有它一定的特点，其中最明显的特点之一是，在汉语音节中元音占优势，而在许多外国语中则不然。同时，汉语语音在结构上有它独特的性质，这主要是表现在语音的结合和声调上。

建议参考资料

1. 朱川（1986）：实验语音学基础。上海市：华东师范大学出版社。
2. 杨治良（主编）（1990）：实验心理学。上海市：华东师范大学出版社。
3. 杨国枢（主编）（1993）：本土心理学研究——本土心理学的开展。台北市：桂冠图书公司。
4. 沈 晔、王书鑫（1983）：一个言语测听材料的编创。心理学报，15卷，3期，316~326页。
5. 张春兴（1989）：张氏心理学辞典。台北市：东华书局（繁体字版）。上海市：辞书出版社（1992）简体字版）。
6. 赫葆源等（编）（1983）：实验心理学。北京市：北京大学出版社。
7. Allen, L., & Santrock, J.W. (1993). Psychology. Iowa: Wm C. Brown Communications, Inc.
8. Atkinson, R.C., et al. (1988). Steven's handbook of experimental psychology (2nd ed.). New York: Wiley-Interscience.
9. Cohen, J. (1969). Sensation and perception. Audition and the minor senses. Chicago: Rand McNally & Co.

10. Kling ,J.W. , & Riggs ,L.A. (1972). Woodworth & Schlosberg ' s experimental psychology (3rd ed.). New York :Holt ,Rinehart &Winston.

第八章 知觉实验

本章内容细目

第一节 知觉研究的基本变量

- 一、自变量 399
 - (一) 刺激的定量变化
 - (二) 刺激的定性变化
- 二、因变量 401
 - (一) 常用的因变量
 - (二) 因变量的控制
 - (三) 因变量与自变量的关系
- 三、控制变量 403
 - (一) 一般的控制变量
 - (二) 控制变量的方法

第二节 知觉现象的研究

- 一、知觉的组织 406
 - (一) 接近法则
 - (二) 相似法则
 - (三) 好图形的法则
 - (四) 过去经验和定势
- 二、错觉 411
 - (一) 不可能图形
 - (二) 视错觉
- 三、知觉的恒常性 420
 - (一) 经验和知觉恒常性
 - (二) 大小恒常性
 - (三) 形状恒常性

第三节 空间知觉和运动知觉

- 一、空间知觉 427
 - (一) 生理调节线索
 - (二) 单眼线索
 - (三) 双眼线索
 - (四) 深度视锐
- 二、运动知觉 441
 - (一) 真动知觉
 - (二) 似动现象
 - (三) 诱动现象

本章实验

- 一、大小知觉恒常性 450
- 二、深度知觉 452
- 三、似动现象 454
- 四、棒框测验 455

本章摘要

建议参考资料

知觉 (perception) 是当前的客观事物的各个部分和属性在人脑中的综合反映。因此, 知觉的事物是复合刺激物, 知觉一般是由多种分析器的联合活动产生的。

知觉形成过程中, 分析器的活动起着极大作用。人们在研究经典条件反射的时候, 曾做过一个有意义的实验。如以频率 500 赫的纯音来形成狗对食物的条件反射, 起先, 在其他频率的纯音作用下, 狗也有流唾液的反应。但是, 当进一步对这个 500 赫的频率给以食物强化, 也就是说, 只有在这种音频后才给予食物, 几次之后, 狗对其他的频率便不再反应, 甚至极相近的频率' 譬如 498 赫的纯音, 也引不起它们的反应了。从这两个实验中, 可见分析器对刺激物的辨识能力是很高的, 而这种辨识能力的精确度是从后天训练获得的。

知觉的形式不仅与分析器的活动有关, 而且依赖于过去的知识和经验。这两者是相互联系的。当所感知的事物同过去的知识经验没有联系时, 就不能立刻把它确认为一定的对象。例如, 当我们看到一个芒果时, 虽然它的颜色、形状、大小等已经通过了我们的眼睛, 刺激了视觉神经和视觉分析器, 但这只是一件实物反映于我们的大脑。如果事实仅限于此, 那么, 这种感觉是没有多大意义的。因为我们还不能就此认出芒果。如果我们过去曾吃过或见过芒果, 情形就不同了。当我们一眼看到它那黄色的果皮、类似球形的外表时, 由于过去经验的作用, 当然晓得它是可以吃的芒果了。不仅如此, 就是当我们看到芒果的图画时, 也能认出它。可见, 正是由于人在实践活动中积累了关于一定对象的知识 and 经验, 才能够借助于这些知识和经验把当前的刺激认知为现实界的确定事物。

知觉的分类方法很多。我们知道, 周围世界的一切对象和现象都存在于空间和时间中, 空间和时间是物质存在的基本形式。因此, 知觉有空间知觉和时间知觉之分。知觉不仅可以根据起主导作用的分析器来划分, 而且可以根据知觉所反映的客体来划分。例如, 不符合事物的客观情况的错误的知觉叫错觉。考虑到人们的大部分信息是通过视觉获得的, 所以在本章内主要讨论和视觉有关的知觉现象。为寻求此问题的答案为取向, 本章将讨论以下七个问题:

1. 通过具体实例, 分析知觉实验中的自变量、因变量和控制变量。
2. 说明知觉组织的一些主要因素。
3. 从对不可能图形的识别中, 告诉了我们什么原理。
4. 几何错觉的研究和应用。

5. 知觉恒常性的计算方法。
6. 对判断距离起作用的主要有哪三类线索。
7. 视觉的运动知觉包括哪三种现象，具体分析说明之。

第一节 知觉研究的基本变量

为了探索知觉的过程，现代认知心理学家提出了知觉的信息加工模型（或讯息处理论）（information - processing model）。他们用信息加工过程来比拟人的行为，用输入（input）和输出（output）等变量来描绘知觉过程。虽然这些模型有一定的局限性，但这些模型有助于分析知觉研究中的基本变量。应用于知觉的信息加工模型，一般具有下列四点基本假设：

- 第一，知觉乃是一种过程的结果，而不是刺激的直接的、瞬时的产物；
- 第二，研究者能够根据这个模型设计出探索知觉过程各个环节的实验；
- 第三，在过程的各个环节上，所保持的或在一定时间内通过的信息量是有限的；
- 第四，这种有限性通常导致选择性：即当呈现的信息量超过容量时，有些信息能完全通过，而有些信息就不能通过。

图 8 - 1 就是哈伯和赫谢桑（Haber & Hershense，1973）提出的视知觉信息加工模型。在此模型中有四个阶段。

第一阶段称之为短暂的视觉储存（或瞬时形象、感觉记录）（sensory memory）。它是从光波投射到视网膜上的信息在内部的复现。这是因为平时眼睛是不随意运动的，投射在视网膜的物体并非就被意识到。短暂的视觉存储的信息过程历时四分之一秒，形成了视觉映象之后才被察觉和意识到。第二阶段是短时记忆（或短期记忆）（short - term memory，简称 STM）作用，它发生在视觉映象构成物的同时或之后，这一阶段包括视觉将信息编码（coding）成概念和词。

第三阶段是长时记忆（或长期记忆）（long - term memory，简称 LTM）的反馈（feedback）作用，有些（但不是全部）短时记忆内容被送入长时记忆，有些长时记忆的信息天天用到它，有些长时记忆信息则几年或更长时间也用不到它。在第二和第三阶段，来自长时记忆和短时记忆的生物信息反馈不断分析、归类和鉴定着视觉映象。

最后阶段是输出过程。在这一阶段借助于输出过程确定一个反应。如果是口头反应，就制定出语言器官的运动程序；如果是要求书写反应，那么另一种反应就被制定出来。

在确定反应程序和作出反应之前，来自长时记忆和短时记忆的生物反馈需要去分析、归类和鉴定视觉映象（Haber & Hershensen，1973）。图 8 - 1 指明了知觉过程也同样包含着输入、中枢加工和输出等主要环节。下面我们讨论知觉研究中的基本变量。

一、自变量

知觉研究中可作为自变量的刺激特别多，刺激的时间间隔、空间间隔、

持续时间和空间三维变化等都可以作为自变量。我们可以把某种刺激的组成部分分开，以一定的时间间隔呈现给被试，从而研究被试的知觉变化。知觉研究中最常用的有二类自变量，一类是刺激在量上的变化，即刺激的定量变化，另一类是刺激的定性变化，采用这类自变量变化的实验，往往是为了研究人或动物于特殊环境下的知觉变化状况。以下将分别讨论之。

（一）刺激的定量变化

埃里克森和柯林斯（Erikson & Collins, 1973）作了一项实验，他们用散点图作为刺激物（见图 8-2）。图中（a）和（b）是两张无意义的散点排列图。假如把（a）和（b）重叠起来，则可以看到一个由两图合并构成的无意义音节“VOH”的散点图（c）。实验中分别将（a）图和（b）图以 6 毫秒的呈现时间和 25~100 毫秒的时间间隔连续呈现出来。结果发现，随着两张散点图（a）和（b）之间的呈现时间间隔的逐渐缩短，被试对无意义音节“VOH”识别的次数逐渐增加。

用刺激的持续时间来作为自变量，常见于时间知觉的实验中，主试把声和光的刺激持续一定时间呈现给被试，要求被试用同样方法将这一持续时间复制出来。

和其他领域的研究一样，知觉研究中的自变量不仅可用单一的变量，也常用组合变量，即由几个单一变量组合成一个较大的变量来作为自变量。自变量的选择要根据实验的目的和要求而定。

（二）刺激的定性变化

采用定性变化的自变量实验，往往是为了研究人或动物在特殊环境下的知觉变化情况，例如将人或动物置于完全黑暗的环境中，剥夺正常的知觉信息的输入或者将正常的知觉进行严重的歪曲等等。斯特拉顿和柯勒（Stratton & Kohler, 1964）曾设计了一种变形眼镜，他们利用包括特制望远镜在内的透镜及反射镜系统将网膜像反转或变位，戴上这种眼镜所看到的世界是上下颠倒和左右反转的（参见图 8-3）。柯勒和其他被试的实验结果发现，戴上这种倒视护目镜，起先视觉对象全部倒置，但过了几周后，触觉和视觉好像重新变得协调起来，并且能从事正常活动。上图为柯勒经过几周后，能戴上倒视护目镜骑自行车。斯特拉顿自己戴上这种眼镜进行了一周的实验来研究知觉的变化和适应问题。

在具体的实验中，将什么样的刺激作为自变量是由实验的目的、仪器设备和条件等因素决定的。确定知觉实验的自变量后，研究者对于自变量的操作和控制必须注意严密性，不要与其他变量发生混淆，尽可能地做到使实验中自变量产生的变异最大，其他无关变量产生的变异最小。另外，研究者还应确定自变量在什么水平、什么等级和什么阶段上发生变化，这样才能使实验最有效。

二、因变量

因变量是实验中由刺激所引起的反应变化，以下我们将讨论知觉实验中

常用的因变量，如何对因变量进行控制，以及因变量与自变量的关系。

（一）常用的因变量

1. 语言描述 知觉实验较感觉实验更为复杂，因而反应形式更多地采用被试的语言描述。采用这一变量非常方便。如主试把各种各样的形状呈现给被试，要求被试在知觉到形状后给出这一形状的名称。但是，客观上的形状是不计其数的，而只有一小部分具有特定的名称，如正方形、梯形、圆形和三角形等，形状愈复杂，被试的语言描述愈困难。所以言语描述具有一些不利之处，没有经过专门训练的被试，其语言描述的精确性不高，而受过训练的被试的语言描述，又很难说是由实验中的刺激所引起而不是由训练本身所引起的。

语言描述带有很大的主观性，各人的言语描述因过去经验、情感动机、身心状态等一系列因素的不同而迥异。在具体的研究中，往往要参考其他因素或者把言语描述和其他反应变量结合起来。格兰泽(Glanze, 1964)在形状识别的研究中发现，被试对于具体的形状如正方形、三角形等，其描述的言语很短；而愈是复杂的、不规则的形状，被试的描述语言愈长。因此，格兰泽提出被试描述某一形状时所用的语言长度也是一种对于刺激量大小的测量。在一般情况下，被试描述某一具体形状时所用的言语愈长，则识别它的准确性就愈小。在这类实验中，言语描述的长度本身就成为一种因变量。

在这类实验中，我们必须注意知觉现象与感觉现象、记忆现象的混淆。例如，有时被试的言语描述不是对他知觉到的刺激的描述，而是对于刺激的记忆进行描述。在实验中做到对它们的真正区别常常是很困难的，这要靠实验的良好设计和主试的经验、技术等，同时被试的一定训练也是需要的。

我们知道，一切变量都必须是可以证实的。言语描述似乎难以直接地被证实，但并非说它无法被证实，我们可以通过其他间接的办法来验证，如量表测量等统计方法。除了这一类不可直接证实的因变量外，在知觉实验中还有一类可以直接证实的因变量，如反应时和某些口头报告等。

可以直接证实的因变量比较客观。例如用速示器以 200 毫秒的速度呈现一组字母，要求被试把它们所看到的字母报告出来。主试可以把被试的报告直接拿来与原刺激比较一下，就可以检验被试报告的正确性如何了。

2. 时间测量、反应时和其他 时间测量(time measure)是心理学实验中最常用的方法之一。反应时同样可以用作知觉实验的一个因变量。如将一系列复杂程度不同、受掩蔽程度不同的图形呈现给被试，同时记录被试对图形辨认的正确性或错误量以及从刺激呈现到图形被辨认出的时间等。在知觉实验中反应时作为一个测量指标通常是与其他指标结合起来使用的。

知觉实验中的因变量大都是从反应的准确性和速度等方面来考虑的。在某些错觉实验中也考虑量的变化和自我估计等方面。例如海门斯(Gheymans, 1896)发现对一个错觉图形的连续观察会使错觉量减少。他用调整法进行研究，让被试多次观看缪勒-莱尔错觉(Müller-Lyer illusion)图(见图 8-15)并进行调整，使图上的两个线段看起来主观上相等。虽然主试并不告诉被试他每次所调整的结果，但在多次观察后错觉量也逐渐减少以至最后趋于零。使用恒定刺激法时，主试用缪勒-莱尔错觉图中的一条线段作为标准刺激，保持其长度恒定；以另一条作为变量，每次呈现的长度不等，按随机

顺序连续呈现。被试在每一次呈现时将它与标准刺激进行比较并做出它是等于、长于或短于标准刺激的主观判断。

（二）因变量的控制

确定了因变量之后，还需要对因变量进行必要的控制，因为对于一个或一组刺激，被试的知觉反应往往是无限的。如何把被试的反应控制在实验要求的方向上，是实验之前应该解决的问题。我们可以通过两个途径来达到对因变量的控制，一是确定一定的指示语，二是选择适当的反应指示。

1. 指示语的一致 指示语在实验中相当重要，它是主试向被试说明实验的途径，也是控制被试反应方向的手段。指示语可对实验结果造成很大的影响。不同的指示语往往得到不同的实验结果。在实验中，指示语要前后一致，明白易懂，不产生歧义，并且要标准化，最好能使用录音机给出指示语或者以文字形式打印给被试。

2. 适当的反应指标 反应指标的选择是控制因变量的主要方法。知觉实验中常用的指标有反应时、错误量、正确率等。选择反应指标要考虑到以下几点：

（1）有效性（effectiveness）：即所选择的指标能够充分反映实验中的知觉现象，指标的变化确实反应了被试知觉上的变化。同时也要认清指标的适用范围，了解它的局限性。对实验中反应指标的选择要作全面的考虑。

（2）客观性（objectivity）：所选择的指标必须是客观的。它可以用客观的方法观察、记录下来。在相同条件下它可以再现，具有可重复性。

（3）数量化（quantization）：数量化是任何科学实验的基本要求。把指标数量化不仅便于观察和记录，也有利于对实验结果的统计处理和提高实验的精确度。

（三）因变量与自变量的关系

自变量与因变量的区分具有绝对意义，又有相对意义，两者实际上很难单独分开来讲。在一个实验中刺激必然引起一定的反应，任何反应必然有引起它的刺激存在。因变量同自变量一样，因实验的目的和方法不同而具有无数种，这在知觉实验中尤为突出。

三、控制变量

在知觉实验中，与实验无关的变量都属于控制变量，它包括除自变量以外的一切条件。哪些无关变量该控制，首先要由实验本身来确定，实验的目的、方法不同，控制变量也不同。在实验之前，实验者可以研究过去的文献或用因素型实验进行预测，确定哪些是该控制的变量。以下将分为二部分来讨论控制变量，一是一般要求该控制的变量，二是控制变量的方法。

（一）一般的控制变量

1. 非自变量刺激的物理量 一般说来，不作为自变量的刺激的物理量都

需要控制，如刺激的持续时间、强度、照明、对比等。例如在一个关于图形的正置与倒置对图形辨认的影响实验中，实验者先呈现给被试 40 张分别有正置和倒置的图片，然后呈现给被试一对一对的正置图片。每一对图片中有一张是被试先前看过的，另一张则是新的。主试要求被试指出每两张图片中哪一张是他在前 40 张里见过的。在这一实验中控制变量首先是图片呈现的速度，如果 40 张图片呈现的时间不同，那么呈现时间长的必然有较多的可能性被认出来。由于实验不是要考察呈现时间的长短对图形辨认的影响，所以这一因素需要加以控制，在实验中 40 张图片按 3 秒一张的速度连续呈现。图形的难易程度或复杂性也是一个变量，在这一实验中也需要加以控制。另外，实验时的照明条件、图片的先后次序、大小等都是需要加以控制的变量。

2. 被试的机体因素 被试的机体因素如情绪、动机、遗传、经验、年龄等也是知觉实验中的控制变量。例如在上一实验中被试的年龄和经验都需要进行控制。不同的年龄和经验本身就是一个自变量。如果上述实验中的被试有小学生、中学生又有大学生，那么实验结果的差异就很难说是由正置和倒置因素引起的。

（二）控制变量的方法

1. 排除法 在知觉实验中对那些需要加以控制的变量的控制方法主要有排除法，即把那些与实验无关的变量消除掉。例如为了消除无关因素的影响，实验可以在特殊环境下或实验室中进行。在距离知觉实验中，为了消除作为距离知觉线索的双眼视差和双眼视轴辐合，可以令被试用单眼观察；为了消除眼睛的调节机能，可使用人工瞳孔等等。但是排除法是有限度的，实验中有些变量是无法消除的，如被试的情绪、动机、遗传、经验、年龄等，以及刺激的物理特性如形状、大小、复杂程度、呈现时间等。

2. 恒定法 在某些变量无法消除时，我们可以用恒定法使这些变量在实验中保持恒定。我们可以把实验安排在同一时间、同一地点、使用同一实验仪器，照明、室温、刺激物的形状、大小、复杂程度、呈现时间以及被试的机体因素等都保持恒定。

3. 实验设计法 除以上二种方法之外，我们还可以通过一定的实验设计来排除某些无关变量对实验的影响。如有人做了一个实验，实验材料用两个社会禁忌词和两个中性词。主试将这些词分别呈现给被试，要求被试以言语报告出来。结果发现，社会禁忌词的呈现到被试报告之间的时间长于中性词的呈现到被试报告之间的时间。这种差别到底是由于知觉系统造成的呢还是由于反应系统造成的呢？或者说这到底意味着对社会禁忌词的知觉时间长于对中性词的知觉时间呢？还是由于被试受传统习惯和社会道德的影响而造成的反应滞后呢？光凭这一个实验我们很难得出结论。我们可以再把实验这样安排一下：刺激仍用原来的四个词，只是当出现社会禁忌词时被试报告中性词，出现中性词时被试报告社会禁忌词。这样如果实验结果表明对出现社会禁忌词的反应时间仍长于对出现中性词的反应时间，那么就可以证实对社会禁忌词的知觉需要较长的时间。这一实验结果是被试报告社会禁忌词时需要的时间较长，从而说明这种差别是由于反应系统而不是由于知觉系统造成的。当然，这只是实验设计上的考虑，在执行实验中，很可能还有更复杂的情况需要考虑。

第二节 知觉现象的研究

知觉是心理学研究的最多最早的领域之一。许多知觉现象的规律性已经得到了共识。但争论的问题仍然很多，尤其有关知觉的理论庞杂繁复更是争论纷纷。各个心理学派别都从不同的方面来探究这一问题，如传统的心理物理学、格式塔心理学（或完形心理学）和现代的认知心理学以及信息论等都对知觉的研究作出了一定的贡献。本节仅选取知觉研究中的知觉组织、视错觉、知觉恒常性等三个问题来进行讨论。

一、知觉的组织

在知觉心理学研究中，形成了不少学派。各学派的中心问题是关于知觉理论，特别是感觉映象是怎样组织的。这些理论涉及较多的形状知觉问题。

在许多学派中，比较有影响力的是格式塔学派（或完形学派）（Gestaltschool）。格式塔就是完形的意思。一般认为韦特海默（Max Wertheimer, 1880~1943）关于似动现象（apparent motion）的研究奠定了格式塔学派的基础。格式塔学派的主要的概念是“整体性”。应当承认在心理学的研究中采用整体性的概念是一个进步，在以前的一些联想学派都认识不到心理过程中由量变到质变的过渡，而格式塔正是强调整体活动的质的特点。他们强调整体具有的规律，它不但是整体的各成分所缺乏的，就是这些成分的总和也得不到这些特点。这一问题在近年来有了新的发展，许多著作从信息论的角度论证了知觉整体性的特点。

知觉组织（perceptual organization），从理论上讲涉及的是知觉理论学派的问题，从实践上讲，更多的是图形和背景（或形象与背景）（figure-and-ground）方面的具体研究。因而我们可以从对象和背景的关系，讨论知觉的组织问题。

平时，我们看到的几何图形，总是在一个背景上的形象，并且以轮廓和界限来形成物体。对象和背景相辅相成，组成了最基本的刺激图形。然而，知觉的对象和背景是可逆的，那些曾经是知觉的对象，可以由于没有价值或者完成了任务而变成为背景，背景中的某些东西在一定时间内可以成为知觉的对象。许多研究发现，天生的盲人经过外科手术重见光明之后，许多知觉的特性都有缺失，唯独从背景中区分出对象的能力依然存在。他们过去虽然无法看见图形，但第一次重见光明的人，都毫无困难地在背景中看出对象。

早期专门研究这个问题的是丹麦心理学家鲁宾（Edgar Rubin, 1886~1951），他确定了图形从背景中分出的一些原则；第一，图形有形状，而背景相对来说没有形状，如果背景被知觉为有形状的话，那也是由于其他完形的作用；第二，图形具有一般物件的性质，而背景看起来像是一种无形的东西；第三，图形似乎是向前突出，而背景似乎是向后退；第四，图形可以引起更深刻的印象，也比较容易记住。

继鲁宾之后，考夫卡（Kurt Koffka, 1886~1941）进而发展了他的学说，并补充了一些新的原则：第一，对象颜色的刚柔对图形来说要比背景更重要；第二，组成图形的定向因素，例如，垂直和水平的方向更容易组成图形；第三，内包和外围的关系，内包的部分较易成为图形；第四，能力的密

度因素，图形中能力的密度较背景的能力密度要高；第五，组织的简单性和均匀性，图形与背景的配置造成的形状，愈简单效果愈好。

近几十年来，在这方面的研究更是丰富多彩了，所发现的因素不下几十种，概括起来，知觉组织的理论解释，归纳出很多法则，称为组织完形法则（Gestalt laws of organization），下列法则尤为重要；

（一）接近法则

接近法则（law of proximity）是指视野中的接近（即空间位置相近）容易合成一组，构成轮廓，当然，接近不限于空间视觉方面，也可以在时间和听觉等方面。例如按不同规则的时间间隔发生的一系列轻拍声中，在时间上接近的响声倾向于组合在一起。图 8-4 是视野中的接近容易合成一组的实例，我们常常把图 A 知觉成正方形。图 B 上由于横向圆点靠得紧，我们把它知觉为四个横排。图 C 上由于纵向圆点靠得紧，我们把它知觉为四个纵排。

（二）相似法则

在形状方面相同或相似的，以及在亮度和色彩方面相同或相似的图形倾向于合成一组构成一个图形是为相似法则（law of similarity）。我们从图 8-5 上可以看到，形状上相同（或相似）的点子，易组成图形。

（三）好图形的法则

形成一个好图形（或完形）（Gestalt）的刺激将具有组合的倾向。好图形一般是同一刺激显示的各种可能的组合中最有意义的图形，此即为好图形法则（law of good figure）。构成好图形的具体因素有四：（1）连续（或连续法则）（law of continuity）。视野中有延续倾向或连续的刺激往往被看成为一条直线与一条波浪形的曲线（如图 8-6A），而以曲线为界分开为两半（如图 8-6B），却很困难。这里包含有“自然”的因素。（2）对称（balance）。对称或平衡的整体，有利于组合，图 8-7 上，凡是对称的，不论是白色还是黑色，都组合成图形，看起来舒服顺眼。（3）趋合（或闭合法则）（law of closure）。轮廓闭合的对象比轮廓不全的对象易被看成一个整体，但我们对自己十分熟悉的对象，即使轮廓缺少一部分，仍然将它知觉为一个整体。例如图 8-8 上尽管均缺少了一部分轮廓线，但仍被我们看成一个圆（左上方）和四边形（右上方）；图 8-8 下方为五个英文字母“HELLO”。据研究，一般而言，只要对象的轮廓线达到 68~72%，就能被知觉为一个整体。（4）共同的变化（common fate）。这个原则的相似组合在物体上的应用，艺术家在设计舞蹈动作时常利用此原则。特别是在大型集体舞的情况下，将循同样路线动作的人组合在一起，使纷繁的变化成为一种迷人的和复杂的活动整体。

（四）过去经验和定势

上面讲到的三点均为图形组织的外界刺激因素。除此之外，还有依个人

主观条件而改变的因素，称之为非刺激性因素。形形色色的可逆图形和双关图形就是这方面的例子，图 8-9 是画家埃斯切尔的木刻画，当你注视黑色图形时所见到的是魔鬼，当你注视白色图形时所见到的却是天使。图 8-10 是少女和老妇二可图。二可图（又称双关图或暧昧图）（ambiguous figure）是指对象和背景可以不时加以转换的图形，即在一个图形上，一部分被知觉为对象，其余的就成了背景，背景和对象可以变换。这里列举的几张图都是典型的二可图。图 8-11 是一张拉丁美洲萨尔瓦多的奴隶市场图，但也能看为一个老人头。图的左边还有一个半身塑像。同一图形可以产生两种知觉现象，这说明了知觉的选择性（selectivity）。这也是知觉的一个特性。

除视觉之外，其他感觉也有知觉的组织以及对象和背景的关系。例如，日常生活中我们可以在户外的嘈杂声中听到鸟的歌声；我们也可以从交响乐团的演奏中听出小提琴、大提琴和小号的旋律。

二、错觉

我们的感觉器官不一定总是能对客观事物作出正确的反映。知觉过程也不是总有一个十分容易而又自动化的程序。知觉有时是非常复杂的。知觉像是解谜一般，知觉者必须把外界的许多线索综合起来。在大多数情况下，我们都能把那些线索正确地进行组合从而很快地解开了谜，这就是为什么我们总是把知觉看得如此容易的缘故。但是在某些情况下，我们会被一些线索所迷惑，我们所感知到的现象并不反映或者符合外部刺激，这就产生了通常所称的错觉（illusion）。

（一）不可能图形

不可能图形（或不合理图形）（impossible figure）是一种无法获得整体和知觉经验的图形，也可说是一种特殊的错觉。图 8-12 是一版画瀑布，由荷兰画家埃斯切尔（Escher, 1961）所作。瀑布一泻而下，汇集到池子中，然后顺着水渠往下流去。可是拐了几道弯之后突然又回到了瀑布口！真是不可思议。可是在画面上却表现得明明白白，天衣无缝。埃斯切尔有效地利用了知觉中的某些线索，如形状和大小的单眼线索、深度知觉、线条透视和视角等，使这个图乍看之下似乎没有什么不正常的地方，图的每一组成部分本身都是完美无缺的。但是如果我们再仔细一看，便会发现它有点问题：图的各个部件组合不妥当，这一瀑布既无始也无终。这就是怪图，在心理学上称之为不可能图形。这一图形之所以产生这种不准确的知觉，是由于它具有一定的线索模糊性和不连贯性。我们能够看出它的错误是由于经验告诉我们它的一些知觉线索自相矛盾，所以有些心理学家认为知觉过程实质是假设的产生和检验的过程，知觉线索的自相矛盾和模棱两可起因于冲突性和不确定性。我们的知觉系统要受到理智的修正，这一点是显而易见的。在皓月当空的夜晚，我们看头顶上的月亮最多不过几百米远，而物理学知识告诉我们月球距离我们约有三十九万公里之遥。

图 8-13 所示也是些不可能图形，右边的图形是心理学上著名的无尽头楼梯图，由于相互冲突的深度线索而导致的矛盾在此图中反映得非常明显。

虽然这些错觉是人为的，但是我们必须认识到，在一定情境中，如果某

些线索具有模棱两可的性质，那么就会有有可能产生知觉上的错误，研究这些人为的错觉现象是为了把知觉的那些基本组成部分抽象出来进行分析，以便帮助我们了解知觉的功能。

（二）视错觉

在心理学上研究的错觉现象，多属视错觉。视错觉 (visual illusion) 是指凭眼睛所见而构成失真的或扭曲事实的知觉经验。视错觉的种类很多，下面我们来讨论一下几种基本的几何错觉形式：

1. 线条错觉 (linear illusion) 表现在线条的长度，方向或弯曲的错觉是相当普遍的。最为大家熟知的是图 8-14 中横竖两条线的物理长度是相等的，但是在我们知觉上它们却不相等，此为横竖错觉 (horizontal vertical illusion)。以及由于两条直线上的两端箭头方向不同，看起来下边的直线长得多的缪勒-莱尔错觉。(参见图 8-15)。

错觉现象有很大的趣味性和很高的应用价值。这里不妨举一个日常生活中的例子。图 8-16 是服装设计中应用缪勒-莱尔错觉原理，产生身高变化的视觉效果。图 A，由于向外伸长的箭头使人们的眼球运动超过了主要线段的长度。相反，在图 B 上，由于向内收缩的箭头使人们的眼球运动的距离缩短。这样产生图 A 比图 B 显得更高些的视觉效果。

迪尤尔 (Dewar, 1967) 曾对缪勒-莱尔错觉中所含变量进行了研究。他用 160 名被试做实验，要求被试调整两条线段中的一条，使之看起来与另一条长度相等。表 8-1 显示了箭头张合角度和箭头长度的每种组合所产生的错觉强度，即一条线段的长度需调整多少才可以使两条线段看起来相等。如表 8-1 中所示，这两种因素都对错觉产生影响。有趣的是，变异数分析表明箭头张合角度和箭头长度两者之间并无相互作用。这意味着两个变量对知觉产生的影响是各自独立的。

另一种线条错觉是我们常常感觉同样长度的垂直线要比水平线为长 (图 8-14)。这种现象可以用这种理论来加以解释，眼睛沿着水平线运动更为容易一些，由于眼睛沿着垂直线作纵向运动需要更多的努力，距离就似乎显得更长。菲克 (Fick, 1852) 和冯特 (Wundt, 1862) 等人最早发现了这种错觉。但对这一错觉形式进行深入研究的是库纳帕斯 (Kunnapas, 1955)。他发现标准的这种错觉 (图 8-17) 实际上包含了两种错觉：垂直线的长度相对于等长的水平线的长度而言，更易于被高估；对被分割的线段而产生低估。图 8-

17 是库纳帕斯发展的一个图形，它展示了这两种错觉。线段 a 相对于线段 b 造成了对垂直线高估的错觉。线段 c 相对于线段 b 造成了未被分割的线段显得长于被分割的线段的错觉。库纳帕斯还发现，对于垂直线长度的错觉，如果把这一线段从水平线的中央移至水平线的一端，则错觉量大大地减少。

表 8-1 缪勒-莱尔错觉的箭头张合角度和箭头长度与错觉量的关系

箭头长度 (毫米)	箭头张合角度			
	30 °	60 °	90 °	120 °
10	6.2*	6.1	5.0	3.0
20	7.9	6.7	5.6	5.9
30	10.2	8.4	9.4	5.8
40	11.9	8.4	8.2	6.9

*单位：毫米

(采自 Dewar , 1967)

图 8-18 中，水平线还被短线所分割。按照视觉运动的理论，沿着水平线的视觉运动在到达另一端点之前被打断而产生线段长度上的错觉。天真的观察者总是认为 B 线要比 A 线短，同时还相信 C 线的中段要比 B 线中分的任何一个线段为长。这种错觉也发生在 D 线和 E 线中。D 线的中段显然要比 E 线的中段为长。与长线相邻的线段似乎显得较短，而与短线段相邻的线段似乎显得较长。

另一种对于线条长度判断的错误是由于我们倾向于以透视的方法观察图像。图 8-19 中的平行四边形通常被当作一个斜面，因为知道左边的平行四边形大于右边的平行四边形，我们就会假定对角线 AC 长于对角线 AB，此称为桑氏错觉 (Sander's illusion)。

邻近的线条和角度所引起的变形不仅表现在线条的表面长度中，而且还表现在线条的方向和弯曲中。例如，图 8-20 是所谓的左氏错觉 (Zollner illusion)，其中的数条平行线被不同方向之斜线所截时，显然不再平行了。而且斜线的方向不同时，看起来斜线的黑色深浅也不相同。在图 8-21 中则是赫氏错觉 (Hering illusion)，A 图的两条平行线显然在中间部分向外弯曲；B 图的两条平行线显然在中间部分向里弯曲。这种错觉可以用图形的后效应理论来解释，它与颜色的残留影像现象十分相似。

2. 大小错觉 大小错觉 (size illusion) 也就是知觉对比 (perceptual contract)。在某些错觉中，物体的物理大小与知觉大小并不一致。图 8-22 是几个大小错觉的例子。A 是庞氏错觉 (Ponzo's illusion)，等长的两条粗黑平行线，因受两边斜线的影响，使上方的粗线看起来比较长；B 是庞氏错觉的变式，常指为透视错觉；C 是戴氏错觉 (Delboeuf illusion)，左图的内圆和右图的圆实际上大小相等，但看起左图的内圆较大；b 是最早由铁钦纳 (Titchener, 1906) 描述过的大小错觉。

西克尔斯 (Sickels, 1942) 曾对庞氏错觉进行了研究。他发现两条斜线的倾斜角有一个最佳角度。一般来说，两条线向水平方向倾斜时错觉增加，一旦超过最佳角度，错觉又显著减少。温特劳布等人 (Weintraub et al., 1969) 对戴氏错觉做了大量的研究。他们发现，当内、外圆环之比为 2/3 时产生的错觉最为显著。他们还发现如果用虚线或者用外圆的一部分把测验圆环围起来也可以产生错觉。而且，变化内外圆的明度对比使内圆的明度略低于外圆，可以增强错觉。

3. 形状错觉 在几何错觉中形状错觉 (shape illusion) 尤为显著。图 8-23 是埃氏错觉 (Ehrenstein illusion)，它是由庞氏错觉变化而来；图 8-24 是奥氏错觉 (Orbison illusion)。奥尔比逊 (Orbison, 1939) 曾以线条为背景把不同的几何形状显示在上面。他发现所有的形状 (如圆形、方

形、三角形等)都由于受到斜线的影响而变形。华莱士(Wallace, 1966)曾研究了各种斜线和观察距离对左氏错觉的影响。一般来说,对图形增加的线条越多,错觉越强;观察距离增远(4.6米)时,仅几条斜线就可以产生弯曲的错觉。

几何错觉的图形是多种多样的,按照它们所引起的错觉形式基本可以分为两类:一类是关于数量上的错觉,包括在大小、长短方面引起的错觉,如缪勒-莱尔错觉、铁钦纳错觉等;另一类是关于变形或方向上的错觉,如埃氏错觉、左氏错觉等。

关于几何错觉的问题,早期人们做了大量的关于各种错觉的错觉量与年龄之间关系的研究。比较典型的是莱博维茨和古兹德克(Leibowitz & Gwozdecki, 1967)的研究。他们以波氏错觉(Poggendorf illusion)为实验材料,发现年龄因素对错觉的判断有如图8-25所示的关系:早期,错觉强度随着年龄的增长而呈下降趋势;大约过18岁以后达到较稳定状态,其曲线如一渐近线。但也有人指出,由于观察条件和观察次数的限制不同,某些研究的结果并不一致。一些研究者,例如贾德(Judd, 1902)和贝努西(Benussi, 1904),曾对错觉呈现的次数对错觉效果的作用提出疑问。一般而言,错觉强度随呈现次数的增加而减弱,但是其他一些因素也会在实验中产生影响。由于被试知道自己所面对的是错觉问题,他们在反应时会相应地对自己的反应作一定的修正。

4. 自然错觉 上面讲到的线条、大小、形状等三种错觉,均存在于自然界之中,我们仅举人为的例子加以分析。这一段中,我们将讨论自然界里有许多可引起我们错觉的现象。我们清晨看朝阳从东方渐渐升起,傍晚目送夕阳在西边慢慢沉落,我们觉得脚下的大地(地球)是不动的,而是太阳在运转,所以千万年来人们都视太阳为东升西沉。事实上,地球不仅以每秒20哩的速度绕地轴自转,还以每秒18.5哩的速度绕太阳旋转。这些自然界里的错觉现象,我们称之为自然错觉(natural illusion)。

我们还在某些地方会看到似乎是倾斜的房子或倒流的河水。这些错觉通常依赖于周围的环境,如环绕的树木和地质结构所造成的与实际物理倾斜相反的地形倾斜。一辆汽车停在这样一种地形上,一旦放开刹车,车子就会向“上”滑去。还有,如果登上起伏的群山上,总有“这山看着那山高”的感觉。其实到了那山,发现怎么又是这山高了呢。这些都是自然错觉。

最典型的自然错觉是月亮错觉(moon illusion)。月亮在水平线上时,比月亮到正顶上时看起来显得大些。对于这一现象比较流行的解释是:对月亮大小的判断依赖于一定的参照物。当月亮在水平线上时,月亮前面的地形给人一种距离遥远的感觉,使人产生月亮较大的印象。这种大小与距离的关系是这样的:如果网膜象的大小保持恒定,感知到的空间距离的增加相应地产生知觉大小的增加,即对物体大小的判断依赖于对物体距离的知觉。当月到中天时,没有已知大小的物体来作为参照物,月亮的大小只能依赖于天空的感觉距离。金和格鲁伯(King & Gruber, 1962)曾有实验证明在水平线上知觉到的天空距离大于在正顶上知觉到的天空距离。他们要求被试把后象投射在蓝天上判断后象的大小,结果证明在水平线上的后象大于正顶上的后象。这表明,天空看起来是一个圆顶形状,且正顶部分最近,水平部分或正前方最远。考夫曼和拉克(Kaufman & Rock, 1962)也做过月亮错觉的研究,他们要求被试观看位于水平线上或正顶上的人造月亮。他们发现当月亮

在水平线上时被试估计的月亮大小增长 1.2 至 1.6 倍。他们还通过一系列反光镜把实际上在正顶上的月亮反射到水平线上，同样被试对其大小的估计增加；而当倒过来时，即把水平线上的月亮反射到正顶上，被试感知的大小则相应减小。

月亮错觉除了这一因素作用外，波林（Boring, 1946）又提出了前庭机制说（vestibule function theory）。他认为，在头部或身体倾斜的情况下，对物体的感知大小趋于减小。其后有人做了一系列的实验证明了前庭刺激确实对大小判断有影响。前面提到的金和格鲁伯关于后象大小判断的实验结果，实际上也可用前庭机制来解释。所以，月亮错觉至少受地形和前庭刺激两种因素的影响。

三、知觉的恒常性

知觉的恒常性（perceptual constancy）是指当距离、缩影比、照明改变的时候，知觉对象的大小、形状和颜色的相对固定性。例如，对一只挂在墙上的挂钟，当我们在房间里走动的时候，我们总把它感知为同一大小和同一形状，虽然它们在我们网膜上的映象是各不相同的。

我们知道，人眼的构造好比一架照相机，人眼中的水晶体（lens）就是一个双面凸起的透镜，眼睛的视网膜起着投影成象的屏幕作用。如果被感知的对象移远些，那么眼网膜上的像就会缩小。镜子反映的映象就是这样。然而人的知觉特点却不然。知觉时，不管实际的光线如何，我们认为一件东西的颜色是相同的，这种倾向称为颜色恒常性；不管看的角度如何，我们认为一件东西的形状不变，则称为形状恒常性；不管距离的远近如何，我们认为物体大小相同，称为大小恒常性；而即使我们到处走动，物体看起来仍然在老地方，则称为位置恒常性（或方向恒常性）（orientation constancy）。恒常性一词有点夸大，它是指一定范围内而言，但它说明了我们对于物体知觉的一种稳定特性。

知觉的恒常性对生活有很大的作用。假如，知觉不是恒常性，那么每走一步、转弯、进行其他活动以及外界光线改变时，便会使我们仿佛碰到了新的天地和新的对象，也就无法辨认以前已经知道的东西，从而使我们无法适应新的环境。

（一）经验和知觉恒常性

于 20 世纪 50 年代后期开始兴起了认知心理学理论，提出知觉不单纯是客观世界的映象，而且还包含着对客体的解释。知觉不可能只单纯从刺激一方描述，人的过去经验在知觉中起着重要作用。认知心理学认为知觉是在直接作用于人的感觉信息的基础上进行推理而产生的，它的内容比感觉基础更为丰富。知觉的恒常性作为知觉的特性，正说明了过去经验在知觉中的作用。

图 8-26 是玛格利特所画的怪画。玛格利特是超现实主义画家，他用违反知觉恒常性的手法，达到知觉冲突，以产生独特的艺术效果。从图上我们可见到大得出奇的梳子和极为微小的床铺。当知觉恒常被否定的时候，现实也同时被否定。这幅图和埃斯切尔的瀑布版画（图 8-12）都反证了过去经验在知觉中的作用，达到异曲同工的效果。

1. 埃默特定律 埃默特 (Emmert, 1881) 发现, 知觉到的后象的大小与眼睛和后象所投射的平面之间的距离成正比, 见图 8-27。后人把埃默特发现的这条规律称之为埃默特定律 (Emmert's Law)。假如把一个后象投射到比原先的刺激物远 10 倍的平面上的话, 那么它的后象看起来变成比原来的刺激物大 10 倍, 但网膜上的映象大小还是相同的。

几何学上的欧几里德定律可以帮助我们理解埃默特的原理。欧几里德定律认为, 客体的大小可用物理量去度量, 网膜象的大小是按光学原理变化的, 可用如下的公式来计算:

$$a = \frac{A}{D}$$

a: 实际物体在网膜上成像的大小

A: 物体的大小

D: 人眼和物体之间的距离

根据这一公式, 如果客体的大小不变, 那么 a 的大小随距离 D 成反比例关系。这个定律可以用简单的实验演示出来。如果我们注视着一个距眼睛 1m 远的黑色背景上的 10cm² 的白色方纸块, 观察约 1~2 分钟以后这一方块就会形成后象, 这时若把眼睛转向一个光亮的、距离约为 1m 的平面, 则可以看到一个与原来白色方块大小相同的黑色方块; 若把后象投射到一个距离为 2m 远的平面, 则看见的后象大小是原来白色方块的两倍, 则后象的大小为 20 × 20cm²; 如果投射面的距离是 4m, 则后象的大小为 40 × 40cm²。

2. 布伦斯维克比率 我们在知觉物体时存在着恒常性, 即不完全依赖于视角规律来判断物体的大小, 但是常常我们的知觉大小与物体的物理大小也不尽一致。例如, 观察者在 5m 处观察一个身高 1.8m 的人, 如果按视角计算, 观察者看到的人只有 1.8 × 1/5 (m), 即 0.36m, 但是实际上我们是不会这样来知觉的, 我们很有可能把他知觉为 1.76m, 也可能把他知觉为 1.84m。布伦斯维克 (Brunswik, 1929) 提出了一个测量恒常性程度的公式, 即布伦斯维克比率 (Brunswik ratio, 简称 BR):

$$BR = \frac{R - S}{A - S}$$

BR: 布伦斯维克比率, 一般用百分数表示

R: 被试知觉到的物体大小, 亦即被试对大小判断的结果

S: 根据视角计算的物体映象大小

A: 物体的实际大小

当知觉到的大小与物体的实际大小很接近时, 布伦斯维克比率趋于 1, 这表示趋于完全恒常性; 当知觉到的大小与按视角计算的大小很接近时, 则表示基本上没有恒常性。前面讲的身高 1.8m 的人在 5m 处观察时被知觉为 1.76m, 那么 $BR = (1.76 - 0.36) / (1.8 - 0.36) = 0.97$, 即大小知觉的恒常性保持了 97%。

3. 邵勒斯比率 邵勒斯 (Thouless, 1931) 提出的计算恒常性系数的公式与布伦斯维克的公式基本相同, 只是取了 R、S、A 三个数的对数:

邵勒斯比率 (Thouless ratio, 简称 TR) 的计算结果也在 0 到 1 的范围内。但是用邵勒斯公式计算出的恒常性系数有时比用布氏计算的要大些。这

两种方法都可以用于研究大小、形状、颜色及亮度等知觉的恒常性，但是在大小、形状知觉中更常用布氏比率，而在亮度知觉中则常用邵氏比率。这是因为物理亮度与知觉亮度成一定的对数关系。

知觉恒常性在大小、形状、颜色、亮度等方面都存在。关于颜色和亮度的恒常性，我们在第六章中已有讨论。下面讨论大小和形状二种恒常性。

（二）大小恒常性

前面讲到，我们所知觉到的物体大小一般至少要受到两个因素的影响，即物体在网膜映象的大小和物体的表现距离。在同一距离上的两个物体，网膜映象大的我们知觉较大些；若在不同距离上的两个物体，其网膜映象大小相同，则距离较远的物体我们知觉为较大些。网膜映象大小和距离的这种关系是大小恒常性保持的基础。

不因对象距离的改变，把它知觉为平常大小的倾向，叫做大小恒常性（size constancy）。我们从视角的概念中已经知道，网膜映象的大小取决于对象的距离的远近。但是当对象距离在变化时，我们视觉的大小或主观大小并不因网膜上大小的变化而感到它的大小在变化。如果深度知觉的线索具备，知觉有很高的大小恒常性。如果把这种线索消除，知觉的对象又不熟悉，那么它的大小知觉便会接近透视大小。图 8-28 显示，上下四扇窗的网膜映像差别如此之大，但我们仍都把它知觉为长方形的窗。

这里还须指出，在一定的距离范围内和有深度线索支持的条件下，我们具有着良好的大小恒常性。但是遥远距离超过 1 公里的对象在我们主观上看来仍然是很小的。例如，我们站在二、三十层的高楼房顶上，下面的汽车像玩具一般。真可谓：“会登临绝顶，一览众山小”。图 8-29 显示了远距离物体大小知觉的变化。

（三）形状恒常性

形状恒常性（shape constancy）是指从不同角度观看一个熟悉的物体时，虽然这个物体在视网膜上的映象都不相同，但是我们仍把它知觉为一个恒常的形状。一扇门只有从下面看时，它在网膜上的映象才是长方形的，从其他任何角度看，都是不同的梯形，然而我们并不把它知觉为梯形。同样，一个倾斜的圆盘在网膜上的映象是一个椭圆形，而我们仍把它知觉为圆形。我们走在马路上，很少从正面看到商店招牌，但从未怀疑它不是方块字。

但是我们的知觉也不尽是完美到与客体保持完全一致的程度，知觉形状与客体某一状态下的形状有时会有不同。在形状恒常性的研究中，人们做了大量关于知觉形状与知觉倾斜关系的实验，因为物体倾斜程度的信息是我们正确判断一个物体形状的基础之一。邵勒斯（Thouless, 1931）发现，观察者为判断测验物体的判断形状大多在物体的真实形状与倾斜形状之间。判断的条件越少，判断形状与真实形状的差异越大，但即使在正常的视觉条件下，观察者也很少表现出完全的恒常性。

莱博维茨（Leibowitz, 1967）研究了年龄与形状恒常性的关系。在他们所用的实验仪器中，首先呈现给被试一个物体，这是一个可以倾斜成各种角度的图形。实验让被试在四种倾斜角度的比较刺激中选择一个与观察刺激

看起来同样的形状。这些比较刺激是一系列从圆形到逐渐拉长的椭圆形。被试年龄范围从4~21岁，实验中许多条件可以利用。图8-30是观察刺激在两种倾斜角度下不同年龄被试的形状恒常性保持程度曲线。

实验结果出乎人们的意料：形状恒常性随年龄的增长而呈下降趋势。他们对此提出了两点假设：（1）形状恒常性的关系或许在儿童早期即已习得，并且这种关系随着个体年龄的增长而变得越来越不重要；（2）形状恒常性也许完全不是经验的结果，而是人的一种先天能力。鲍尔（Bower，1966）曾证明仅两个月的婴儿就表现出形状恒常性。但这两种假设尚待进一步验证。

针对恒常性受认知经验的影响，爱波斯坦等（Epstein et al.，1963）曾研究了不同指示语对形状恒常性的影响问题。这一实验所用的材料为简单的几何图形，图形以不同的倾斜角度呈现。指示语分为三类：一组被试的指示语是要求他们根据客观的或已知的形状来判断观察图形的形状；另一组则要求他们根据物体的网膜象的形状来判断；第三组要求根据观察图形的表面形状或现象上的形状来判断。实验结果：第一组的布氏比率平均为0.58；第二组为0.21；第三组为0.30。这表明要求被试按客观形状来判断的指示语所产生的形状恒常性最小。利希特和博雷森（Lichte & Borreson，1967）用无意义图形为测验材料重复了这一实验，所得结果如下：第一组布氏比率为0.93；第二组为0.45；第三组为0.62。客观组的判断接近物体的真实形状；网膜象组受到形状恒常性趋势的极大影响；现象组介于两者之间。

知觉恒常性作为知觉的特性之一，已被肯定下来。但人们对它的实验研究还是很不够的。

第三节 空间知觉和运动知觉

一、空间知觉

空间知觉（space perception）是三维知觉。我们知道，人眼的网膜是一个二维空间的表面，但是在这个二维空间的网膜上却能看出一个三维的视觉空间。也就是说，人眼能够在只有高和宽的二维空间视象的基础上看出深度。这是因为人在空间知觉中依靠许多客观条件和机体内部条件来判断物体的空间位置。这些条件称为线索（cues）。人在知觉对象的空间关系时，并不完全意识到这些主客观条件的作用。

据研究，先天失明的人，在经过医治复明的头几天内，是分不清形状、大小和远近的。可见，空间知觉不是生来就有的，而是后天学习的结果。

判断距离起作用的条件主要有三类：生理调节线索、单眼线索和双眼线索。生理调节线索也叫肌肉线索（muscle cues），包含眼睛的调节和双眼视轴辐合；单眼线索也叫物理线索（physical cues），包括大小、遮挡、线条透视、注射角、空气透视、光亮与阴影、纹理梯度和运动视差等；双眼线索主要是双眼视差。下面我们分别加以讨论。

（一）生理调节线索

生理调节线索（physiological accommodation cues）仅指纯生理上的调节线索，包括眼睛的调节和双眼视轴辐合。

1. 眼睛的调节 眼睛调节 (eyes accommodation) 是指人们在观察物体时, 眼睛的睫状肌 (或毛状肌) (ciliary muscle) 可以对水晶体进行调节, 以保证网膜视象的清晰。看远物时水晶体较扁平, 看近物时较凸起。这样, 眼睛肌肉紧张度的变化所传递给大脑的信号就成为估计物体间距离的线索之一。但是眼睛的调节对深度知觉所起的作用并不大, 一般, 这种线索所提供的信息只限于距眼球 10 米范围内才是有效的。

彼得 (Peter, 1915) 最早研究了眼睛的调节作用对深度知觉的影响, 他给被试在不远的距离上呈现两个圆盘。圆盘的大小可以改变, 从而使它对被试所形成的视角保持不变。他在实验中排除了其他深度线索。结果发现, 当标准刺激物在 130 厘米处, 比较刺激物在 70 厘米处时, 被试才能判断后者比较近些, 而当两个刺激物的相对距离小于此值时, 被试便不能判断哪个在前, 哪个在后。这一实验表明了眼睛的调节在空间距离知觉中的独立作用。

人们极少单独利用眼睛的调节作用来作为距离知觉的线索。一个好的空间知觉一般都不自觉地同时利用了多种线索。

2. 双眼视轴辐合 生理调节的另一个线索是双眼视轴辐合 (binocular convergence)。双眼视轴辐合也是由于眼肌的调节而产生的深度线索。双眼视觉所提供的深度知觉线索, 主要包括双眼辐合和双眼视差。双眼辐合是指在两眼注视远物时视轴分散超于平行, 辐合程度减小; 注视近物时两眼视轴交叉, 辐合程度增大。由于辐合的角度不同, 提供了物体的深度线索。双眼视差是指双眼注视一点后, 近于或远于此点的物体, 将投射至两眼视网膜的非对称点而造成视差。因此, 双眼辐合不同于双眼视差。当我们看一个物体时, 为使物体的映象落在网膜感受性最高的区域里, 以获得清晰的视象, 视轴就必须完成一定的辐合运动。在看近距离物体时, 眼球外部肌肉紧张度增加, 两个眼球转向鼻侧, 视轴趋于集中; 看远距离物体时, 眼球外部肌肉紧张度减少, 视轴趋于平行。控制两眼视轴辐合的眼肌运动提供了关于距离的信号, 但是, 由视轴辐合而产生的距离线索只是在物体距离眼球约几十米以内才有效。观察距离更远的物体时, 双眼视轴接近平行, 对于距离的判断就不起作用了。

图 8-31 表明辐合角与注视物体之间距离的关系。P 为注视的物体, L 和 R 分别为左右眼的位置。假定二目间距为 65 毫米, 当视轴向 P 点辐合时, 左眼向内侧转动的角度为 L , 右眼向内侧转动的角度为 R 。两个角度之和就是整个辐合角的度数, 等于 C 。若已知距离 D , 可求出 C , 或已知 C , 可求出距离 D 。

计算可用公式为 $C/2=32.5/D$ 。然而, 若把目间距 LR 看作圆弧, 则计算更为方便也较精确。这样, D 为圆的半径, $C=65/D$ (用弧长表示)。1 弧度约等于 57.3° 或等于 206, 265 秒, 表 8-2 是几个常用的对应值。我们可得下列二种计算方式:

(1) 已知 D , 求 C

$$C = \frac{65}{D} \times 206, 265 = \frac{134007225}{D} \text{ (秒)}$$

(2) 已知 C , 求 D

表 8-2 不同距离下的双眼辐合角 (目间距: 65 毫米)

距 离 D (毫 米)	辐 合 角	
	秒	度
100	134 , 072	36
300	44 , 691	12
600	22 , 345	6
1 , 000	13 , 407	3.7
10 , 000	1 , 341	0.37
50 , 000	268	0.07

(采自 Kling&Riggs , 1972)

$$D = \frac{13407225}{C} \text{ (毫米)}$$

例如, 当眼睛辐合 12° 时获得单一视像, 则眼睛与物体的距离为:

$$D = \frac{13407225}{44691} = 300 \text{ (毫米)}$$

(二) 单眼线索

许多深度线索只需要一只眼睛就能感受到, 刺激物所具的此类特征, 称为单眼线索 (monocular cues)。这些线索一般是空间视觉的物理条件, 由于人的经验作用, 这些物理条件也可以提供环境中物体的相对距离的信息。艺术家便常利用这些线索在二维平面上创造出能够表达深度经验的画幅。下面列举几种主要的单眼线索:

1. 遮挡 依靠物体的遮挡判断对象的前后关系完全取决于物理因素。两种或多种物体在同一平面上, 就会产生遮挡 (或重叠) (superposition) 现象。当观察者运动时或对象在运动时, 遮挡的改变使我们很容易判断物体的前后关系。物体的互相遮挡是单眼或双眼判断物体的前后关系的重要条件。如果一个物体部分地掩盖了另一个物体, 那么前面的物体被知觉得近些。图 8-32 遮挡 (采自 Darley et al., 1988) 8-32 表明了这种关系。

艾姆斯 (Ames, 1951) 曾做了一个有趣的实验。这个实验首先给被试观看两张放在不同距离处的扑克牌, K 放在 5 英尺处, Q 放在 10 英尺处。两张牌的放置位置部分重叠, K 遮住了 Q 的一角。实验中控制其他距离线索。由于这种遮挡关系, 被试能够正确判断两张牌的前后关系。第二步, 再次呈现这两张牌, 但把两张牌的位置互换, 并且将 Q 的一角 (先前被 K 遮掉的那只角) 剪掉。结果被试会知觉到放在远处的 K 仍在 Q 的前面。

2. 直线透视 直线透视 (linear perspective) 是指平面上的刺激物, 根据视角原理, 近处的对象面积大, 占的视角大, 看起来较大; 远处的对象占的视角小, 看起来较小。透视原理在古代就已经被发现了, 所以中世纪及后来文艺复兴时期的画家都利用这个原理在平面上表现出空间关系。德国著名画家达莱 (Dürer, 1525) 最早提出了这个原理, 他确定了客观对象和艺术创造力之间的数学透视关系。17 世纪画家杜布尔 (Jean Dubreul) 根据达莱的数学分析透视原理, 用蚀刻画制成样板画, 以此来教授他的学生。图 8-33 就是杜布尔的蚀刻画。

3. 单眼运动视差 单眼运动视差 (monocular movement parallax) 是指视觉对象不动, 而头部与眼睛移动时, 所给出一种强有力的线索。这是由于三维空间的各物体分布在离观看者不同的距离上, 头部运动确实地改变了网膜上的刺激模式, 因为这些物体是相继地从不同角度被观看的。比如, 如果头向左运动, 较近的物体好像移位到右边, 而较远的物体好像移位到较近的另一个的左边。在日常生活可以经常找到说明这种情况的例子。例如, 你可手持圆珠笔置于一臂之远, 并使这支圆珠笔和房间任何距离上的物体(例如, 一只吊灯)成一条直线。此时, 闭上一只眼睛, 虽然你明明知道圆珠笔是在近处, 但看起来几乎并不如此。然而, 仍闭上一只眼睛, 沿着水平方向摇动你的头, 哪怕是一次微笑的运动, 就会产生明显的深度知觉效果。再如一团乱线, 如用一个眼睛静止地去看的时候, 不容易分辨出哪根在前哪根在后, 但是当我们头一动或者身体一动的时候, 就能看出哪根在前哪根在后, 提供深度线索。

因此, 单眼运动视差, 也就是观察者通过观察角度的变化, 以达到从另外一个方向去观看视野中的一个物体。这种运动过程作为提高单眼空间知觉的条件, 在电影中和医学中得到广泛利用。我们知道了在电影中缺乏双眼空间知觉的条件, 所以拍摄景物常是在运动过程中进行的, 如在汽车中拍摄路上的景物。在医疗事业中, 有人发明利用运动的照相方法, 如使病人慢慢地转动再进行 X 射线银屏记录。用这种方法可以使深度印象突显出来, 从而更精确地判断出病变的位置。显然, 没有差异的比较, 是很难诊断的。

4. 高度 对象在水平面上的高度也是深度知觉的线索。如果我们把同样大小的对象置于不同的水平面的高度上, 高水平面上的看起来较远, 低水平面上的看起来较近(参见图 8-34)。

5. 纹理梯度 最早将纹理梯度列为深度线索的是心理学家吉布生。吉布生在《视知觉》一书中写道: “某个维度上某种东西的递增或递减称为纹理梯度”(Gibson, 1950)。视野中对象重复而众多的成分, 构成一种视觉表面纹理, 距离愈远, 纹理愈细愈密。如生活中常见的铺石、地毯图案、湖水或草原表面都呈一种纹理梯度(或纹路梯度)(texture gradient), 在这些表面上, 随着距离的增加都产生近处稀疏和远处密集的纹理梯度。如果一个人站在一条砖石路上向远处眺望, 由于在网膜上的远外部分每一单位面积上的砖石映象的数量较多, 故远处的砖石块越显得小, 产生深度知觉。

由于物体和距离任何一种空间安排在不同的照明和纹理形状下, 都会产生结构密度梯度(texture density gradients)(见图 8-35B), 所以吉布生假定, 重要的刺激变量是梯度, 而不是视网膜上的刺激点和物体的线索。因此他指出, 纹理梯度如网膜表象上的波长和明度一样, 都是视觉系统能够作出反应的真实而适宜的刺激。在视网膜上的这些梯度一方面与客观的安排直接有关, 另一方面又与相应的主观知觉有关。因此, 吉布生假定, 对当前环境及其他包含的物体的全方位的知觉, 可能在対这种环境所包含的每一部分作出详细的分析之前就获得。正像图 8-35 所表明的那样, 纹理梯度携带着: 关于一个平面上的物体的大小和距离的信息(图 8-35C), 关于两面所成夹角的信息(图 8-35E), 甚至有关形状的知觉也可以据此加以解释。例如, 在图 8-35F 中, 由于图形的末端和前端具有相同数量和宽度的纹理单位, 所以就被觉察为一个“倾斜的长方形”。在图 8-35G 中, 图形的顶端双底端具有较窄的结构单位, 所以就被觉察为“正前方平面上的梯形”。

图 8-35 是一组图形，总的表明，纹理结构的梯度乃是判断空间轮廓的信息资源。吉布生指出，具有均匀的纹理梯度的平面（图 8-35B）为眼睛提供了一种与该平面的倾斜度有关的梯级。图 8-35C 表明，即使物体 i 和物体 ii 形成了同样的视角，但物体 ii 比物体 i 覆盖了更多的纹理单位，所以观察者就认为 ii 比 i 大。图 8-35D 表明，平面倾斜度的改变导致了梯度的改变。图 8-35E 表明，距离的突然改变导致了密度的突然改变。图 8-35F 上看到，X 和 Y 覆盖了同样的纹理单位，所以上面的图形看起来像是一个长方形。图 8-35G 上看到，边 X 覆盖了比边 Y 更宽的纹理单位，所以图形看起来像是一个梯形。

以上讨论的是观察者和观察对象处于静止状态的情况，当观察者与周围环境有相对运动的时候，纹理梯度就更为显著。例如，以飞机朝向地平线飞行的时候，地面及天空层都发生连续的梯度变化。当飞机朝向地平线降落的时候，周围的物体看起来都从这一中心向外扩散。这种表面纹理扩散速度的差别就形成了不同的距离知觉。

心理学家还研究了上述各种线索的交互作用。显然在实验室之外的这些线索并不是独立作用于人们的视觉系统的。相反，它们常常同时影响人们的距离和深度判断。有时它们是一致的，有时它们又相互矛盾。那么它们究竟是如何相互影响的呢？

有许多情况下，线索会产生一种联合效果。詹姆森和赫维奇（Jameson & Hurvich, 1959）报告，被试者对距离差异的感受性，在多种线索共同作用的情况下，则与每一种线索所产生的感受性的算术总和密切相关，即

$$\frac{1}{\Delta D} = \frac{1}{\Delta D_1} + \frac{1}{\Delta D_2} + \frac{1}{\Delta D_3} \dots\dots$$

D：观察距离，

D：视差距离，

1/ D：感受性，

D1：某一种线索的视差距离

然而，这种推论的精确程度如何，至今仍有异议。

还有不少研究表明，线索的交互作用依赖于我们选择什么样的线索以及在什么条件下联合使用它们。另外也有研究表明，人们在使用线索时具有很大的选择性，有人喜欢使用这样的线索；有人喜欢使用那样的线索，在这种情况下，线索的交互作用就削弱了。在格式塔心理学（或完形心理学）家们看来，线索是不存在的，对空间知觉起作用的是人们所获得的关于外部世界的规则以及人们与生俱来的组织知觉的方式。吉布生也说过，线索是无关紧要的，主要的变量是刺激的梯度和作用于观察者的其他不变的刺激特征。

（三）双眼线索

双眼线索（binocular cues）主要是指双眼视差（binocular disparity），双眼视差是知觉立体物体和两个物体前后相对距离的重要线索。借助于双眼视差比借助上述各种线索更能精细地知觉相对距离。特别是在缺乏其他线索来估计对象距离的时候，双眼视差更为重要。距离和深度视觉主要是双眼的机能。

在正常的知觉情况下，人都会利用双眼来观察环境和物体，双眼线索给空间知觉的单眼线索和肌肉线索提供了必要的补充。在三维空间的深度知觉中，双眼线索起了重要的作用。由于人的两只眼睛相距约 65 毫米，两眼的左、

右视野是略有不同的，但在双眼视野中，左右视野有大部分重合在一起。处于重合部分之内的物体是双眼都能看到的。不重合的部分叫颞侧新月（temporal crescent）（见图 8-36），在这部分视野内的物体是对侧眼睛所看不到的。故人在观察空间中的立体对象时，两只眼睛所看到的部分是略有不同的，左眼看到物体的左边多一些，右眼看到物体的右边多些。两只眼睛把各自所接收到的视觉信息传递到大脑皮层的视觉中枢，在这里经过一定的整合，产生一个单一的具有深度感的视觉映象。

物体同时刺激双眼形成两个独立的网膜视象，而人们仍然把它知觉为单一的物体。早期，缪勒（Müller, 1912）认为这是由于物体的同一部分落在两个视网膜上相应点的缘故。相应点是两个视网膜上对应的各点，每一对相应点在视网膜上都与中央窝同一距离同一方向。

由此可见，我们用双眼看东西时，左眼和右眼所看到的映象并不相同。这种稍有差别的映象合而为一，就产生了立体效果。我们有二只眼睛，并非是因为可以在一只眼睛受伤时有备用之物，犹如一个人有二个肾一样，少了一肾也能同样工作。有两只眼睛的人其视觉比只具单眼的人优越。这不只是视野广阔，能见到更多的东西，而更重要的是具有深度知觉（或立体知觉）（depth perception）的优点。我们可做一个简单的实验，闭上一只眼睛，右手拿钢笔杆左手拿钢笔套，然后把钢笔杆插到钢笔套里。这时你会发现，在单眼视觉条件下，第一次往往套不上，往往要套多次才能套上。假如你的优势是右手，一般优势眼也为右眼，那么右眼单眼视觉条件下，只要套几次就能套上。但是当左眼单眼视觉工作时，有时要套十几次才能成功。这个生动的例子可以说明双眼视差在立体知觉中的作用。

我们在第六章里讲到，网膜的中央窝比其他区域敏感得多。当我们观察一个对象时，双眼注视它，使对象的映象尽可能地落在中央窝上。由于两眼彼此相隔约 2.5 英寸，因此对同一立体对象得到不完全相同的映象。而且距离愈近，差别愈大。从这些差别中，我们得到深度的线索。这种从网膜视差中再造出来的立体感，就称之为立体视觉（stereoscopic vision）。

根据这一原理，惠斯通（Wheatstone, 1883）发明了实体镜（stereoscope）。实体镜的原理即是先把从每只眼睛的角度所看到的画面制作出来，然后再把这两张略有不同的画面分别呈现给左右眼，从而形成一个立体的图像。图 8-37 就是一架透镜式立体镜。

立体电影就是应用了双眼视差的原理。立体电影是把左右两部影片用红与蓝绿两色叠印在同一部影片上，观众戴上左蓝绿右红的滤光眼镜，左眼的蓝绿滤光镜阻碍观众看到右眼的蓝绿色影象，但能看见红色影象；同样，右眼的红色滤光镜阻止观众看到左眼的红色影象，但能看见蓝绿色影象。这样，左右眼分别得到不同的映象，从而产生深度知觉。应用双放映机与偏振光系统把左右两部影片两个不同的平面投射在银幕上，观众戴上偏振片眼镜后使左眼看不到右眼看到的影象，右眼看不到左眼的影象。这种方法不仅产生了立体知觉，而且还可以伴有彩色，增加深度知觉效果。

美国贝尔电话公司的朱里兹（Julesz, 1964）曾做了一个实验，成功地将双眼视差与其他深度线索分离开来。他用计算机制成一对随机点子图（8-38）。两张图除了右图中央一小块比左图的中央一小块略向左移动一些外，其余相同。当把两张图中任何一张呈现给被试时，或把两张图呈现给被试的一只眼睛看，被试均不产生深度知觉。但是若把它们放在实体镜上分别单独

地同时呈现给被试的两只眼睛看时，被试产生了深度知觉。图中央的一小块突出地浮现在周围的点子背景之上。这是由双眼视差引起的立体视觉实例，说明得如此明白，许多教科书争相引用。

以上我们分别讨论了各种线索在形成立体视觉中的作用。但是，在正常的视觉经验中，我们对空间深度的判断中更多地依赖于生理调节线索、单眼线索和双眼线索的综合作用。库纳帕斯 (Kunnapas, 1968) 曾做了五个系列实验来评价每一种线索对深度知觉的作用。实验是让被试判断一个置于 0.25 ~ 3.95 米处的圆形物体的距离。他发现实验中被试可利用的深度线索愈多，判断愈准确 (参见表 8-3)。

表 8-3 可利用线索的比较研究

实 验	观察方式	可 利 用 的 线 索
一	单眼	眼睛的调节
二	单眼	眼睛的调节、相对大小
三	双眼	眼睛的调节、双眼视轴辐合、双眼视差
四	双眼	眼睛的调节、相对大小、双眼视轴辐合、双眼视差
五	双眼	所有深度线索

(采自 Kunnapas , 1968)

这五个实验中，在实验五的条件下被试判断的准确性达到最高。而在实验一条件下，被试由于仅能通过眼睛的调节来判断距离，他们几乎把所有处于不同的距离上的目标都判断为在同一个距离上。从而，库纳帕斯认为，眼睛的调节对深度知觉所提供的信息作用是不大的。而实验四条件下，由于包括了一个单眼线索，被试判断的准确性几乎达到与实验五条件下同样的准确程度。在实验二和实验三条件下，被试对短距离的判断是相当准确的。但是随着观察距离的增加，被试的判断准确性明显下降。这表明单眼线索对短距离的深度准确性判断所起的作用不大，生理调节线索和双眼线索在短距离判断中作用较大。然而，随着观察距离的增加 (1000 英尺以外) 人们就越来越多地依赖于单眼线索。但是，单眼线索一般对较远距离的判断是不太准确的。

(四) 深度视锐

深度视锐 (depth visual acuity) 是指能够辨别两个处于不同距离上物体之间距离的能力。深度视锐是双眼视差对距离或深度的最小辨别阈限。

对深度视锐的测定一般用霍瓦-多尔曼知觉仪 (见第十一章) 或称深度知觉仪 (depth perception apparatus)。这个仪器上有一固定的立柱，在它旁边还有一个可以前后移动的立柱。仪器上标有刻度，可读出活动立柱与固定立柱的距离。被试在 2 米处通过仪器上的一个观察窗观察这两根立柱，并对活动立柱进行调节，使之与固定立柱看起来在同一距离上。实验中要排除其他深度线索，只让双眼视差起作用。

设 P 和 Q 为仪器上两个立柱，它们距离观察者的距离分别为 y 和 $y-x$ ， a 是双目间距 (如图 8-39 所示)。双眼深度视锐可用象差表示，象差角 n 定义为对近物体的辐合角 c_1 减去对远物体的辐合角 c_2 ，用公式表示即为：

$$n = \frac{a}{y-x} - \frac{q}{y} = \frac{ax}{y(y-x)} \times 1 \text{弧度}$$

若一被试在 2 米距离处调节活动立柱，当两立柱之间距离为 2 毫米时，被试开始不能辨别两者前后差别，即认为此时两者处于同一距离上。假定被试的目间距为 65 毫米，则深度视锐为：

$$\begin{aligned} n &= \left(\frac{65}{2000-2} - \frac{65}{2000} \right) \times 206265 \\ &= \frac{130}{2000(2000-2)} \times 206265 = 6.71 \text{秒} \end{aligned}$$

双眼深度视锐受照明条件的影响，良好的照明条件可以提高深度视锐。例如在良好的条件下，视力良好的观察者可以辨别 $n=2$ 秒的深度距离。根据虞积生（1980）的资料表明，人眼在观察距离为 6 米时，深度阈限的平均值为 2.94 秒，标准差为 1.79 秒。

二、运动知觉

运动知觉（或移动知觉）（motion perception）是对于物体在空间位移的知觉，它是多种感觉器官的协同活动的结果。运动知觉的产生一般至少有两个原因：一是物体在空间的位置变化而视网膜上留下轨迹；二是观察者自身的运动（如身体运动、眼球运动等）所提供的动觉信息。当物体改变空间位置，而我们又能够察觉到这种变化时，我们便产生了该物体运动的知觉。运动知觉是一个复杂的过程。参与运动知觉的感官有视觉、动觉、平衡觉，有时还有听觉和肤觉。例如，在人眼和头部不动时，运动物体连续刺激网膜各点，视象在网膜上的移动，通过视觉的信息，我们便知觉到物体在运动。

但是产生运动知觉的情况却是多种多样的。有时，在某种情况下，虽然没有同一物体实际的空间位移，也能产生物体的运动知觉。概括起来，视觉的运动知觉包括三种现象。一种是真动知觉：观察者处于静止状态，运动的物体以一定的速度作空间位移。例如当我们看见飞机在天空飞行、火车在铁轨上奔驰时，就产生了对飞机、火车的运动知觉。第二种现象是似动现象：它是连续的静止刺激在视野的不同地点出现，而使观察者产生的运动知觉。例如电影、霓虹灯广告等所引起的知觉。第三种是诱动现象：这种现象是观察者本身在运动，他与客观对象的相对空间关系的改变，或者两个以上的对象彼此互换的空间关系发生变化而引起的诱动现象。下面我们分别讨论这三种现象。

（一）真动知觉

物体在空间的移动都有一定的速度，它们在空间的位置变化反映到我们的视网膜上，便产生了关于它们运动的知觉。真动知觉（或真实移动知觉）（real motion perception）是指我们所见到的物体确实在移动，而且其速度达到知觉阈限。运动知觉依赖于许多主客观条件，其中最基本的条件是同一物体以一定的速度作空间位移。在一般情况下，当刺激的映象在网膜的不同部位上运动（包括位移、缩小和扩大）时，我们便知觉到运动。引起这种知觉的主要变量是刺激映象在网膜上运动的速度。那么在空间中位移的物体

在什么条件下才能被我们知觉到它的运动呢？这就是运动知觉的阈限问题。我们所说的运动知觉阈限实际上是关于真动知觉的阈限。

当物体位移速度过于缓慢时，我们便不能察觉它是在移动。只有当它的位移速度加快到某种程度，我们才能对它产生运动知觉。例如，钟表上的分针和时针，虽然我们可以根据间隔一段时间后它们的位移来推测它们是在运动，但我们不能直接感知它们的移动。刚刚可以辨认出的最慢的运动速度，称为运动知觉下阈（lower threshold of motion perception）。研究表明，运动知觉的下阈为 2 分 ~ 6 分/秒左右（运动知觉阈限用视角/秒表示）。大略地说，10 呎远的对象必须至少每秒钟运动 0.06 吋，才能使我们知觉到它的运动。

当物体位移速度过于快速时，我们同样不能觉察它是在动。例如，我们无法看清射出枪膛的子弹。运动速度大到看不清时，这种运动速度称为运动知觉上阈（upper threshold of motion perception）。运动知觉上阈为 35 度/秒。运动的物体低于这个速度，才能被我们知觉到它的运动。

我们可从运动知觉阈限的单位上知道，决定运动知觉的变量是角速度，而不是线速度。例如，在同样速度下，飞机在 5000 米的高空上看来飞得较慢，而在 500 米低空上看来飞得较快。飞机在 5000 米的高度上比在 500 米的高度上需要 10 倍的时间才能形成同样的视角。假若观察者维持眼睛不动，在网膜映象上低空飞机要比高空飞机快 10 倍，飞机在雷达屏幕上的情况也是如此。这些原理具有着较高的军事应用价值。

布朗（Brown，1958）的研究表明，运动物体的知觉上阈一般约为每秒钟 35°。布朗曾报告了几个有关运动物体的物理速度与运动知觉之间关系的实验。他用一组小黑方块为实验材料，每一方块都可以由主试控制产生运动。他的实验结果与奥伯特的实验结果相符：可觉察的物体运动最低速度约为每秒钟 2 分。他在实验中把黑方块的运动速度不断加以变化，从而得出八种不同的主观经验（见表 8 - 4）。

表 8-4 运动知觉与速度的关系

知 觉 经 验	物体运动速度[视角（度、分、秒）\秒]
1. 无运动	2 分\秒以下
2. 在视野某些部位可见方块运	2 分 ~ 6 分\秒
3. 在视野所有部位可见方块运动	10 分\秒以下
4. 从慢到快逐渐加速	10 分 ~ 4 度\秒
5. 方块似向后运动	3 度 ~ 9 度\秒
6. 知觉到两个方块在运动	10 度 ~ 19 度\秒
7. 带有一些深暗部分的淡灰线	12 度 ~ 20 度\秒
8. 平稳的灰线	13 度 ~ 32 度\秒

（采自 Brown，1958）

运动知觉还依赖于许多主客条件。谢弗等人（Shaffer et al.，1966）曾用信号测验论的方法对运动知觉阈限做了研究。他们在实验中用常定刺激法比较了两种条件下运动知觉的阈限。一种条件为：背景静止，物体运动；另一种条件为：背景运动，物体静止。他们发现，在这两种条件下所引起的运动知觉经验是相同的，但觉察其运动的阈限却不相同。当运动速度为每秒

钟 164 分时，物体只需位移 1.8 分便可被知觉运动，而背景需位移 4.8 分才可被知觉到运动。

布朗的研究中还分析了阈限上运动物体的物理速度的主观估计问题，其基本设计是要求观察者把一个比较刺激的速度与标准刺激的速度进行匹配，使两者速度在主观上相等。在一个实验中，被试观察一个大于比较刺激一倍的标准刺激，并且，标准刺激所在的背景两倍于比较刺激所在的背景。布朗发现被试为了获得较满意的匹配，往往把比较刺激的速度调到大约为标准刺激速度的一半。所以，他得出结论说，现象的速度依赖于物体的相对大小及其背景。

若呈现大小不同的物体分别作为测验刺激和标准刺激，而使背景的大小保持一致，一般观察者是对较小物体的运动速度作出高估。所以运动速度的判断还依赖于这一速度所在的参照系统。如果在完全不知道物体的真实大小情况下，例如在一个完全黑暗的环境中，观察者一般都不能正确判断物体的运动速度。

运动知觉的研究成果在工程心理学和军事心理学中得到广泛应用。在许多工种的选拔中，都会涉及到运动知觉能力的测验。例如，轮船和汽车驾驶员必须正确估计速度，使自己的轮船或汽车不致与另一对象相撞。这种判断是很复杂的，各个人在判断能力上的差别很大。我们可以编制测验来测量这种能力，采用适当的训练方法来培养与提高这种能力。再如，在军事心理学中，低空飞行的敌机难以观察与防御，因此就需经常变换雷达屏的方位以及采取其他应变措施。这些都无不与运动知觉有关。

（二）似动现象

似动现象（apparent motion）是我们对实际上没有空间位移的物体所产生的运动知觉现象。例如在一个黑暗的房间内两个相距一定距离的光点相继一明一灭时，观察者会知觉到一个单一的光点来回晃动。似动现象的应用很广，如霓虹灯广告的制作，电影和电视的摄制等，都是利用似动的原理而产生出一种视觉上连续、自然的运动效果。

似动现象的发现已有一百六十多年的历史。普拉梯（Plateau, 1833）制造了第一个动景盘（stroboscope）。直到现在，动景盘还是心理实验室演示似动现象的常用教学工具。图 8-40 就是一个动景盘的图案，会产生对象连续运动的效果。转动过慢时，只看到画面（图案）的转动，而不是连续的舞蹈动作。转速过快时，会出现多余动作。若从动景盘顶上观察，只看到一片模糊的印象。从这个演示里我们可以看出，决定似动现象产生的主要因素是前后图案出现的时间间隔与空间间距。也就是说，似动现象受到两图案呈现的时间和空间条件制约。

对似动现象的研究，早期集中在对它的分类上，曾提出似动现象的多种类型和形式，其后，研究较多地阐明其机理。最早对似动现象进行系统和细致研究的是韦特海默（Max Wertheimer, 1880~1943），他探索了形成似动知觉所需要的最适条件。他把那种没有运动对象，纯粹的运动称为似动现象（或飞现象）（phi phenomenon）运动，并认为它是不能再分解的一种图 8-40 动景盘的图案（采自 Cohen, 1969）基本感知现象，并以一种纯粹的形式被感知。当时，根据这种解释，发现似动现象至少有如下四种形式：

1. α -运动 它是贝努西 (Benussi, 1912) 报告的一种似动现象, 他相继呈现缪勒-莱耶错觉图形的两部分, 先呈现箭头朝里的, 紧接着呈现箭头朝外的图形, 几次反复之后便产生了水平线扩张与缩短的现象。在 α -运动中观察者有时还可以知觉到 β -运动, 当相继呈现的速度较慢时箭头似乎前后翻动; 但是当相继呈现速度很快时, 便产生枢轴运动, 线条似乎在绕着与水平线的接头处旋转, 同时产生侧向运动和在深度上的运动。这些缪勒-莱耶错觉图形引起的似动现象, 都称之为 α -运动 (alpha motion)。

2. β -运动 韦特海默 (Wertheimer, 1912) 最早对 β -运动进行了研究。当把两个有一定空间间隔的静止目标先后连续呈现时, 如果时间间隔和空间距离恰当的话, 观察者便可以看到一个单一的目标从一个位置向另一个位置运动。 β -运动 (beta motion) 的产生依赖于两个刺激连续呈现的时间间隔、空间距离和明度水平。

柯尔特 (Korte, 1915) 研究时间间隔、空间距离和刺激明度三者与最适 β -运动的关系。他发现, 这三个变量以某种组合产生了 β -运动之时, 如果其中一个变量的值发生变化, 那么只要在其余两个变量中给予适当的补偿变化, 其似动知觉可以保持不变。这就是柯尔特定律 (Korte's Law)。可用公式表示为:

$$= f(s/ig)$$

：最适似动

s : 两个刺激的空间距离

i : 明度 (刺激强度)

g : 时间间隔

由这一定律可见: 如果时间间隔恒定, 那么在增加空间距离的同时以一定的比例增加明度, 最适 β -运动保持不变; 如果空间距离恒定, 那么在增加明度的同时以一定的比例降低时间间隔, 最适 β -运动保持不变; 如果明度恒定, 那么在增加时间间隔的同时以一定的比例增加空间距离, 最适 β -运动保持不变。柯尔特似动定律, 虽然未具体指明各函数加数情况, 但它所揭示的函数关系至今仍然有效。

3. γ -运动 γ -运动 (gamma motion) 是由于增强或减弱一个刺激的明度水平而产生的刺激深度大小的变化。如果一个刺激的明度增加, 它似乎在向近处运动逐渐变大; 如果明度降低, 则觉得它在收缩或远去。倘若用速示器相继呈现一个亮圆, 观察者看到中心先出现高点, 然后扩大而充满全圆。照明停止后, 光亮由边缘缩小至中心, 而后消失。亮圆相继出现, 则产生刺激深度大小的变化。

4. δ -运动 如果把 β -运动和 γ -运动的刺激条件结合起来, 则产生另一种形式的似动, 观察者可以看到刺激从一个位置向另一个位置的侧向运动和在三维上的运动。如果第一个刺激的亮度远大于第二个刺激的亮度, 那么这种似动看来是从第一个刺激向一旁和后侧运动; 如果第二个刺激的亮度远大于第一个刺激的亮度, 则运动方向相反, 即由第二个刺激向第一个刺激方向运动; 如果在相继出现的两个刺激之间存在障碍物, 似动可能迂回绕过障碍, 看起来似乎超越所在的平面而在三维上运动 δ -运动 (delta motion) 一般是指三维似动现象。典型的图案是《和》, 相继呈现这二个刺激物, 就可见到三维似动现象。

似动知觉产生的原因曾有种种不同的解释。现在的理论解释为, 由于刺激所引起的感官兴奋有一个短暂的持续时间, 当第一个刺激所产生的兴奋尚未消失时便接着出现第二个刺激, 两个刺激印象便发生融合 (synthesis),

这样观察者就产生了似动的知觉。似动现象的机理还有待进一步的研究。

（三）诱动现象

诱动现象（或运动幻觉、游动错觉）（induced motion）即观察者与客体的相对空间关系的改变，或者两个以上的彼此空间关系的变化，就可以引起诱动现象。在有的心理学教科书中，把多种形式的诱动现象归属于似动知觉。常见的诱动现象是由人体的移动或眼动引起的。图 8 - 41 就是一个诱动现象的例子。当注视这张图片时，由于经常的不随意的眼动，导致图片上的曲线会有运动的幻觉。

诱动现象又有很多种。例如，当一个人站在一片漆黑的房间里，室内仅有一个静止的微弱光点，当人长久地注视它，他会开始觉察到光点在游荡。这种现象又称自动现象（autokinetic phenomenon）。对自动现象的研究，有着一定的实用价值，在航空飞行中，驾驶员的这种幻觉有时会成为夜间飞机失事的原因。用无线电仪表导航，就能预防这类事故。

两个对象彼此互换空间关系，也会引起诱动现象。例如，我们坐在行驶着的火车窗口看窗外，远的对象看来与我们同方向运动，而近的对象朝相反的方向运动，参见图 8 - 42A。还有一种类似的情况，假如我们坐在停靠车站的车厢内，另一辆火车刚开出车站时，我们会感到移动的是站台，这种现象又称站台错觉（station illusion）。这种错觉也常见于屋顶或塔上的旋转餐厅，坐在旋转餐厅里起初总感到移动的不是自己，这两种错觉又称相对移动（relative motion）。

心理学家们寻求以理论去解释这些现象，并试图通过经典的例子去说明它。研究表明，运动视差（motion parallax）和运动透视（motion perspective）依赖于视域组成部分的相对运动。它不仅在物体移动时出现，而且在观察者本身移动时也会产生。观察者本身的移动对于空间知觉提供了重要的视觉信息资源。用赫尔姆霍茨的话说：

当观看一个静止的远物时作出的眼睛运动引起了视象的运动。通常我们都频繁地移动我们的眼睛，但静止的物体仍然是静止的。然而眼球运动是有意识的。被动的眼动，例如由于眼睑被动受压所产生的眼动，的确使得视域似乎在运动。（Helmholtz, 1866）

视差（parallax）一般是由观察者位置的改变所引起的视域的改变造成的。当你的头部和身体移动时，画面上物体的投影全都在移动。如果在你向左移动时，把一个物体固定在画面的中间（图 8-42A 上的点 iv），那么所有比它近的物体图像在画面上将会向右移动（i），比它远的物体图像将会随着你向左移动（v）。画面上图像的运动速度随着它们和固定的物体实际距离的增大而增加。

不仅平面上不同强度距离的物体有不同的运动速度，而且每一个平面上的点相互之间也有不同的流速。这种点的流动包含了大量的有关观察者位置和运动的有用信息。这些信息的实用性已经被吉布生等所证明。图 8-42 就是一个很好的具体说明：观察者移动的方向导致了视域中点为核心的有规律变化，改变的方向和空间中物体的轮廓及运动密切相关。如果观察者移动他的眼睛、头部和身体，整个画面便会动起来。在图 A 中，观察者若向左移动，并注意观察视域中间的点 iv。请注意图中箭头的长短表示画面上各点之间的

相对流动速度，可以看到从 i 到 iv 的分布呈均匀的梯度形式。图 B 表示观察者正在朝 iv 急速移动。图 C 表示观察者正在朝着被墙隔开的中心点 iii 移动。在图 D 中，观察者正在 iv 所示的地面上向左移动，物体 i 正在靠近观察者，物体 ii 在地面上是垂直静止的，物体 iii 正在向左移动，但不像观察者移动得那么快。

本章实验

一、大小知觉恒常性

(一) 目的

在不同角度下观察同一事物时，感官活动随着条件的变化也不断地变化着，但我们仍然将它视为同一事物，这个特点叫做知觉的恒常性。如当距离改变时，我们知觉到的物体的大小仍然不变，这叫做大小恒常性，简称大小常性。

透视定律告诉我们，物体离开我们越远，它在网膜上形成的象就越小。但对熟悉的事物，如一枚镍币，由于经验的作用，即使它离我们很远时，我们也不会产生与它放在手心时不同的知觉。不过，网膜成象和距离变化的关系并不呈现完全的反比关系。这些关系有哪些规律？就得通过实验取得。

常性的研究方法有描记法、比配法等多种。本实验用比配法验证视觉大小常性现象，比较单眼、双眼观察时大小常性的程度，以便学习测定常性的方法。

(二) 仪器和方法

1. 仪器 知觉大小常性测量器，白布幕（或黑布幕），等边三角形硬纸 12 片（高从 5cm 到 20cm 不等，每个依次相差 1cm）。

2. 方法 选择长度为 7m 以上的场地，以每 1m 为一个度量级，测定距离位置。在 6m 处悬挂白布幕（放幻灯用）或黑布幕一块，根据程序在幕上先钉上一个三角形硬纸片作标准刺激（被观视图形）。被试在不同距离处调节常性测量器，每名被试配备一名记录员。

(三) 程序

1. 实验时被试注视 6m 处的标准刺激，然后调节放置在明视距离 25cm 处的常性测量器，直到认为所调图形面积同幕上图形面积相同为止。记录下测量器背面的刻度。

2. 先用双眼比较不同距离、不同面积的三角形，即在 1m、2m、3m、4m、5m、6m 处。选定四个图形作标准图形，每个距离比较四次。每做完一种距离，休息 5 分钟。然后，分别用左右眼比较不同距离不同面积的三角形。

3. 绘制记录表，严格登记实验数据。记住，在实验中，记录员不得将测量数据告知被试。实验记录用纸形式如下：

大小常性实验记录表 单位： mm

	1m				2m				3m				*
双眼													
左眼													
右眼													

* 4m、5m、6m 可按此要求画表

4. 指导语：“请你注意正前方屏幕上三角形的大小，并照此大小调节你手边的测量器，直到你主观感知到一样大小为止。报告记录者记下你调节后图形的数值。注意测量器的观视距离应保持 25cm”。

(四) 结果

1. 计算各种情况下的大小常性系数；
2. 求出各个图形在各种情况下的透视值；
3. 制作在不同观察条件下双眼、单眼（左、右）的大小知觉曲线图，格式参照曹日昌主编的《普通心理学》第 176 页。

布伦施维克提出的计算公式如下：

$$\text{常性系数 } K_B = \frac{R - S}{A - S}$$

R：见到的形状

A：实际的形状值

S：透视形状值，完全没有常性时应有的结果

$$\text{透视值 } (S) = \frac{A \times B}{D}$$

D：标准刺激的观视距离

B：被试对测量器的观视距离（本实验定为 25cm）

(五) 分析与讨论

1. 试分析单、双眼大小常性的差别及其原因。
2. 分析在这个实验中，透视值固定，常性系数随标准刺激大小和距离远近变化趋势。

二、深度知觉

(一) 目的：比较双眼和单眼的辨别深度中的差异。学习使用霍瓦-多尔曼深度仪测量深度知觉阈限。

(二) 仪器和材料：霍瓦-多尔曼深度知觉仪、电源接线板。

(三) 程序：

1. 学习使用霍瓦-多尔曼深度仪；被试坐在仪器面前，通过观察孔进行观察。以仪器内部两根立柱中的一根为标准刺激，距离被试 2 米，位置固定。以另一根作为变异刺激。先由主试调到某一固定的位置，然后由被试根据观察自由调节到他认为两根立柱的距离相等为止。主试记录两根立柱的实际距离，即误差。

2. 在双眼视觉的情况下，进行 20 次试验。其中有 10 次是变异刺激在前，由近向远调整；有 10 次是变异刺激在后，由远向近调整，顺序及距离随机安排。求出 10 次的平均结果。

3. 按照上述程序，再以单眼观察试验 20 次，求出平均结果。

(四) 结果：

1. 计算在双眼观察情况下，表示深度阈限的视差角。

利用公式：

$$\text{视角差} = 206265 \times \frac{b \cdot D}{D(D + D)} \quad (\text{单位：弧秒})$$

b=目间距 65mm

D=观察距离。本实验为 2000mm

D=视差距离

2. 根据全班被试的平均结果，比较双眼和单眼深度视觉阈限的差异，并考验差异的显著性。

实验记录用纸：

条件	次序 误差	双眼观察		单眼观察	
		远 近	近 远	远 近	近 远
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
平均数					

(五) 讨论：

1. 说明双眼视差在深度知觉中的作用。

2. 单眼和双眼在辨别远近中是否有差异，原因何在？
3. 霍瓦-多尔曼深度仪有没有缺点？应如何改进？
4. 深度知觉研究的应用价值？

三、似动现象

(一) 目的

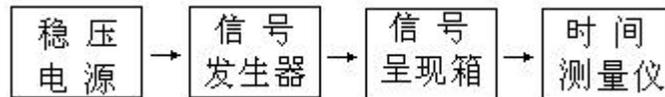
学习似动现象的研究方法，掌握似动现象的基本规律，分析影响似动现象产生的时间与空间条件。

(二) 材料

EP502 型似动现象呈现器、稳压电源、时间控制器、信号发生器、记录用纸。

(三) 程序

1. 按要求接好实验仪器电路。



2. 被试在暗室内熟悉实验条件。被试坐在似动现象呈现器前约 2m 的地方，眼睛注视着似动现象呈现器上的小灯。实验要求被试报告，观察到的两个光点是同时出现的呢，还是相继出现的，或者移动。重复多次，这个步骤主要是要求被试掌握似动现象的标准，以便开始正式实验。

3. 实验者事先安排好两个亮点呈现的时间间隔和空间距离，列出实验顺序表。一方面如规定几种两点间的空间距离 (cm)；另一方面规定，以下几种时间间隔 (ms)：50、100、150、200、250、300、350、400。每种时空条件要做 4 次，各刺激要随机呈现。

4. 正式实验开始后，按照已排好的顺序用上述方法进行实验，每 48 次作为一个单元系列，完成单元系列休息 3 分钟。

5. 每一单元系列开始前，实验者调整好两光亮点的空间距离，再按不同的时间间隔呈现刺激。

6. 做完上述系列实验后，要求被试回答是否已掌握了辨别似动现象的标准。

(四) 结果

1. 将所得结果分别统计为“同时”、“先后”、“似动”三种情况，计算出现次数的百分比，并填入记录用纸上。

似动现象记录表

时距 (ms)	50	100	150	200	250	300	350	400
同时 似动 先后								
同时 似动 先后								
同时 似动 先后								

2. 整理和计算出不同空间距离条件下阈值的范围，其中明确从观察到“同时”到“似动”为下限，而从“似动”到“先后”为上限。
3. 整理和计算出不同时间间隔条件的似动阈值（毫秒）的范围。
4. 分别作出同一被试在不同情况下所得的结果，以及不同被试结果的比较表。要求画出在各种时空条件下似动现象出现的数次的曲线图。

（五）讨论

1. 根据实验结果说明似动现象产生的时间与空间条件。
2. 分析似动现象产生的个体差异。
3. 实验结果与经典实验结果有何差异？这种差异如何从实验上加以分析？
4. 试述似动现象在实践中的意义。

四、棒框测验

（一）目的：掌握棒框仪的使用方法；通过实验，认识棒框测验在认知方式研究中占有很重要的地位。

（二）仪器：棒框仪。参见第十一章有关棒框仪的介绍。

（三）程序：

1. 令被试者端坐在仪器前，双眼紧贴观察孔。实验过程中始终保持这种情况（休息除外），暗适应 5 分钟后，开始施测。要求实验过程中头部始终保持正直，一定不能随框的倾斜而倾斜。为了做到这点，可用下额托将头部固定。待被试者坐好后，施测者陈述指导语：“注意看框和棒，并将棒调节与地面垂直，时间不限。当你认为棒已调至与地面垂直时就报告垂直了，并且每次判断垂直的标准要一致”。施测者在被试者报告调至垂直后，记下调节的误差数（所谓的角度与真正垂直之差数），不计正负号，只取误差的绝对值。正式实验前可让被测者练习 1~2 次，让其掌握方法，施测时还有二个应注意的问题：一是施测者调节框和棒的倾斜角度的过程不能让被试者看到，可采用两个方法：在施测者调节时，将仪器后面进光部分遮上；另一方法是可让被试者闭目，待施测者调好后再让被试者自己调节。二是仪器的光源要在—项研究中保持一致。因框棒的不同亮度，对垂直判断的误差大小是

有影响的。

2. 实验过程要求主试者调节框的度数。让框从 0° 开始，每 3° 为一个梯级，直至 45° 。即框安排的度数分别为 0° 、 3° 、 6° 、 9° 、 12° 、 15° 、 18° 、 21° 、 24° 、 27° 、 30° 、 33° 、 36° 、 39° 、 42° 、和 45° 。框倾斜的度数，随机呈现。每个度数上，随机测验四次，二次要求被试将棒从顺时针方向调至垂直，二次由逆时针方向调至垂直。取四次的平均数为该倾斜度的平均误差数。

(四) 结果

1. 分别整理框的各倾斜度的平均误差数。
2. 以框的倾斜角度为横坐标，以被试调节棒的平均误差数为纵坐标，绘制棒框测验曲线图。

(五) 讨论

1. 比较框的各倾斜度的平均误差大小，根据自己实验结果，看哪个倾斜度的平均误差数较大。并和前人实验结果进行比较。
2. 根据本实验结果，并结合平时自己的认知方式特点，分析自己是属于场依存性还是属于场独立性认知方式。

棒框测验记录用纸

本章摘要

1. 知觉是当前的客观事物的各个部分和属性在人脑中的综合反映。因此，知觉的对象是一个复合刺激物。
2. 在知觉研究中，刺激的时间间隔、空间间隔、持续时间和空间的三维变化等都可以作为自变量。知觉实验较感觉实验更为复杂，因而反应形式更多地采用被试的言语描述。
3. 在知觉心理学研究中，形成了不少学派。其中较有影响的学派是格式塔学派。格式塔学派的最主要的概念是“整体性”。
4. 在知觉的组织中，重要因素有：接近组合，相似法则，好图形的法则，过去经验和定势。
5. 不可能图形（或不合理图形）是一种无法获得整体知觉经验的图形。
6. 线条、大小、形状等错觉，均存在于自然界中。自然界里有许多可以引起我们错觉的现象。最典型的自然错觉是月亮错觉。
7. 知觉的恒常性是指距离、缩影比、照明改变的时候，知觉对象的大小、形状和颜色的相对固定性。
8. 埃默特在 1881 年发现，知觉到的后象的大小与眼睛与后象所投射的平面之间的距离成正比。后人把这一规律称之为埃默特定律。
9. 空间知觉是三维知觉。人眼能够在只有高和宽的二维空间视象的基础上看出深度。这是因为人在空间知觉中依靠许多客观条件和机体内部

条件来判断物体的空间位置。

10. 对判断距离起作用的条件主要有三类：生理调节线索、单眼线索和双眼线索。
11. 生理调节线索也叫肌肉线索，包括眼睛的调节和双眼视轴辐合；单眼线索也叫物理线索，包括大小、遮挡、线条透视、注视角、空气透视、光亮阴影、纹理梯度和运动视差等；双眼线索主要是双眼视差。
12. 运动知觉是对于物体在空间位移的知觉，它是多种感觉器官的协同活动的结果。
13. 真动知觉是指我们见到的物体确实在移动，而且其速度达到知觉阈限。
14. 似动现象是我们对实际上没有空间位移的物体所产生的运动知觉现象。似动现象的种类很多。
15. 诱动现象又称运动幻觉或游动错觉。观察者与客体的相对空间关系的改变，或者两种以上的对象彼此空间关系的变化，可以引起诱动现象。

建议参考资料

1. 玛里琳·霍思（乐竟泓、杨治良等译，1991）：服饰——人的第二皮肤。上海市：上海人民出版社。
2. 杨治良（主编）（1990）：实验心理学。上海市：华东师范大学出版社。
3. 张春兴（1991）：现代心理学。台北市：东华书局（繁体字版）。上海市：上海人民出版社（1994）（简体字版）。
4. 郑昭明（1993）：认知心理学。台北市：桂冠图书公司。
5. 胡德辉等（1989）：现代心理学。郑州市：河南教育出版社。
6. Eysenck, M.W. (1994). The blackwell dictionary of cognitive psychology. Massachusetts: Blackwell Publisher.
7. Kantowitz, B.H., & Roediger, H.L. (1984). Experimental psychology (2nd.ed.). New York: West.
8. Levine, M.W. (1981). Fundamentals of sensation and perception. Massachusetts: Addison-Wesley.
9. Woodworth, & Schlosberg (1972). Experimental psychology (3rd.ed.). New York: Holt.

第九章

记忆实验

本章内容细目

第一节 记忆研究的变量和材料

一、记忆实验的基本变量 465

- (一) 自变量
- (二) 因变量
- (三) 控制变量

二、记忆研究的材料 468

- (一) 无意义材料
- (二) 有意义材料
- (三) 其他刺激材料

第二节 记忆研究的传统方法

一、回忆法 473

- (一) 系列回忆法
- (二) 对偶回忆法
- (三) 自由回忆法

二、再认法 481

三、再学法 483

四、重建法 484

五、部分报告法 485

六、记忆广度法 488

七、分散注意法 490

第三节 内隐记忆

一、内隐记忆的研究领域 493

- (一) 再学时的节省
- (二) 阈下编码刺激的作用
- (三) 无意识学习
- (四) 启动效应
- (五) 健忘症病人的内隐记忆

二、内隐记忆现象的理论解释 496

- (一) 什么是内隐记忆
- (二) 内隐记忆和外显记忆的区别
- (三) 内隐记忆的理论解释

第四节 内隐记忆的测验方法

一、经典的测验方法 501

- (一) 词干补笔
- (二) 知觉辨认

二、非语词信息的内隐测验 506

- (一) 熟悉的非言语信息的内隐测验
- (二) 新异非言语信息的内隐测验

三、内隐学习的测验方法 509

本章实验

一、短时记忆 510

二、长时记忆 513

本章摘要

建议参考资料

记忆 (memory) 是过去经验在人脑中的反映。汉语中的“记忆”一词, 简洁地表明了人对过去经验的反应, 总是先有“记”再有“忆”, 用现代信息加工 (或讯息处理) 的观点解释记忆, 把记忆看作是对输入信息的编码、储存, 以后在一定条件下提取的过程。例如, 我们熟读了一首李白的诗, 过几天还能把它背出来, 这就是通过记忆来实现的。

记忆是一个复杂的心理过程, 它包括识记、保持、再认或回忆三个基本环节。识记 (或认知) (cognition) 就是通常所说的“记住”, 保持 (或保留) (retention) 就是巩固已获得的知识、经验的过程, 再认或回忆就是在不同情况下恢复过去经验的过程。以一个曲调为例: 我们以前听过一个曲调, 若能不看乐谱把它哼出来, 便是回忆 (recall); 若在别人演奏时能听出是以前听过的, 便是再认 (recognition)。在记忆过程中, 这三个环节相互联系又相互制约, 没有识记就谈不上对经验的保持; 没有识记和保持就不可能对经历过的事物再认和回忆。因此, 识记和保持是再认和回忆的前提, 再认和回忆又是识记和保持的结果, 并能进一步巩固和加强识记和保持。

记忆是智力发展的必要条件。一切智慧的根源在于记忆。人们依靠记忆把过去经验保存在自己的头脑中, 然后在经验恢复的基础上, 进行思维和想象活动, 这些思维和想象的结果又作为经验保存在头脑中, 作为进一步思维和想象的基础。这样, 就使人的思维逐步深化、复杂化、抽象化, 促使其智力逐步向更高的水平发展。如果没有记忆, 以前所感知过的事物都会是陌生的, 人将永远处于新生儿状态, 在他的头脑中由于没有感知的材料, 就无法进行思维和想象活动, 更谈不上智力发展了。

自从德国心理学家艾宾浩斯 (Hermann Ebbinghaus, 1850 ~ 1909) 在 1885 年发表他的实验报告后, 记忆就成为心理学中实验研究最多的领域之一。其后, 记忆理论和研究方法即不断发展, 50 年代出现的信息加工理论 (或讯息处理论) 把记忆看作是对信息的输入、编码、贮存和提取的过程。因而按储存的时间而分为: 瞬时记忆、短时记忆和长时记忆。这期间尽管研究者们使用的具体方法、材料、仪器和实验条件不相同, 但这些研究有一个重要的共同点, 就是把记忆看作是人们对先前经验的有意识的外显恢复。

近些年来, 一些新的观点对这个看法提出了挑战。愈来愈多的研究者提出, 不应把记忆看作单一的实体, 而应把它看作是由不同的结构 (form)、系统 (system) 或类型 (type) 所组成, 它们之间具有性质不同的功能 (Schacter, 1985)。他们提出了多重记忆结构 (multiple memory form) 这一术语。记忆的多重结构意味着记忆的不同类别或组合, 它是指功能相互独立的过程所引发的性质不同的记忆效果。

主张存在多重记忆结构这一术语的根本原因在于愈来愈多的证据揭示, 在有意识的外显记忆 (explicit memory) 之外还存在着一个相对独立的记忆系统, 即内隐记忆 (implicit memory)。这种记忆特点是人们并不是有意识地知道自己拥有这种记忆, 它只是在对特定任务的操作上表现出来。据此,

在本章的撰写中旨在回答以下六个问题：

1. 记忆实验的自变量有哪些？记忆实验的材料有哪两类？
2. 在 20 世纪 50 年代中期以后，用信息加工（或讯息处理）的观点去解释记忆，将记忆分为哪三个系统？
3. 研究记忆的传统方法有哪些？
4. 何谓内隐记忆的研究领域？何谓内隐记忆？
5. 何谓任务比较方法学？何谓实验性分离？
6. 补笔和知觉辨认的方法特点是什么（从间接测量上分析）？

第一节 记忆研究的变量和材料

记忆这个曾经被披上神秘色彩的心理现象吸引了古今中外众多的专家学者。能够借助科学的方法对人类记忆进行实验研究是心理学发展到一个新阶段的标志，记忆研究首创者艾宾浩斯因此而声誉显赫。在艾宾浩斯以前，冯特（见第六章）曾建立第一个心理实验室，将心理学引入全新的科学领域，但他否认对记忆、思维等高级心理过程能够进行实验研究。在这种观点指导下，早期形成的心理物理学方法、内省法、反应时间等方法几乎完全被限制在感觉研究的领域中。

在 1875 年后，艾宾浩斯受费希纳《心理物理学纲要》一书的启发，开始了记忆研究的实验构思。他的目标是用精心控制的自然科学方法去研究英国经验主义所描述的联想过程。1885 年，艾宾浩斯的力作《论记忆》终于问世了。这部心理学名著主要有二方面的意义：第一，突破研究高级心理活动的方法学“禁区”，强化了心理学的科学性；第二，大大充实了实验心理学内容，为扩大心理学的研究领域创造了条件。

艾宾浩斯以后，自 20 世纪 20 年代起到 50 年代的三十余年，是行为主义的盛行期，行为主义坚持以动物研究为主，故以人为对象的记忆心理研究受到了影响。这种影响的结果不仅是研究内容方面受到限制，而且实际进行研究的数量也非常有限。尽管有这样的曲折，整个记忆研究工作并没有中断。在记忆心理研究工作演进的过程中，除这段缓慢发展的时期外，更多的研究在内容上表现出螺旋上升的特点，即从早期的“短时记忆”、“长时记忆”研究再回到更深入的“短时记忆”、“长时记忆”的研究阶段。在后一阶段中，由于研究者使用了新的方法，又开辟出“瞬时记忆”的研究领域，从而使当今关于人类记忆的研究工作呈现出新、旧方法并用，新、旧理论互补，研究成果日新月异，众多学科相互影响的崭新局面。图 9-1 概要地示意了记忆研究的历史进程。

记忆研究的成果和方法息息相关，巴甫洛夫（Ivan Petrovich Pavlov, 1849~1936）说过，方法学每前进一步，我们便仿佛上升了一级阶梯。50 年代，斯珀灵（Sperling, 1960）成功地研究了瞬时记忆，也是因为他首创了“部分报告”的方法。方法上的创新既要求创新者具有了解现有方法的基础，还要求对具体条件、设计思想等方面的问题作统筹考虑。为此，我们在这一节中先讨论实验的基本变量，再分析记忆研究的典型材料。

一、记忆实验的基本变量

（一）自变量

自变量（或自变项）（independent variable）是引起行为变化的因素。研究的自变量必须具备“定性”和“定量”两个条件。以实验的刺激材料为例，若各个刺激单位性质不稳定或者渗入其他性质的材料，那么实验就不可能获得准确的量值。艾宾浩斯设计“无意义音节”，目的就是寻求能获得准确的量值。所谓定性（qualitification），也就是保持刺激的性质，克服不确定性，这是实验有效性的又一个保证。在记忆实验中，设计无意义刺激材料为的是防止产生联想，由于被试者在知识、经验、理解上的不同，一旦产生联想，就会产生不同的记忆效果。设计有意义刺激材料，要求划分“相等的意义单位”即定量（quantification），也是为了使每个被试者在实验中接受性质相同的自变量。只有这样，实验才能实施定量，结果才能比较。

按统计原理，保持刺激材料性质的恒定，才能达到刺激的各个子单元可以互比或互换。满足实验随意变更操作要求，这是保证刺激材料作为自变量的基本条件。在记忆这个高级心理过程的实验中，如何确定刺激材料的自变量是一项十分关键且相当困难的任务，有关这方面的设计思想和技术方法在下面的内容中还要介绍。然而，记忆实验中的自变量并不限于刺激材料，其他方面的自变量大致还有：

1. 刺激呈现的速度、间隔时间的长短；刺激的空间位置和距离；学习的遍数和时间等。
2. 刺激作用的感觉通道或方式，实验的干扰条件。
3. 实验研究要求回忆还是再认，或其他方式。
4. 指导语。

（二）因变量

一项实验有了自变量，并有相应的实验程序，最后就是收集因变量的问题了。因变量（或依变项）（dependent variable）即受自变量影响的行为表现。传统记忆实验的因变量按照实验的内容和记忆过程的特点基本分成两大类：回忆类和再认类。回忆实验要求被试全部或部分再现先前所接受的刺激材料；而在再认实验中，被试是面对再次呈现的刺激材料，进行有选择地同一再认。这两类记忆实验在程序上、内容编排上、难易程度上都有差异。因此，记录因变量的方法各有特点。一般来说，回忆类实验的反应量值比较直接。例如记忆广度量，取绝对量就可说明问题。再认类实验的因变量大多数需要经过运算，相对来说是间接获得的。例如“节省分数”、“再认分数”就是百分数。

1. 回忆类实验 这类实验共有三种：系列回忆、自由回忆和对偶回忆（详见本章第二节）。这三种实验的计量方法不尽相同，但都有适宜各自特点的规则。这些规则包括：（1）系列回忆要求记录符合系列顺序的量值或错误量，通常还要根据因变量反映前后系列的回忆特征；（2）自由回忆不要求被试恪守刺激呈现的序列顺序，允许随意提取记忆项目。这种方法比较简单，但只要计量时注意各量值对应刺激项目（自变量）的关系，就可以比较深刻地揭示自由回忆的过程特征；（3）在对偶回忆程序中刺激是成对呈现的，如“苹果——茄子”成对，被试回忆时，主试者再现“苹果”一词，看被试者是否

能回忆出“茄子”这个反应词。对偶回忆的计量工作既可以是记录一次再现能够回忆的绝对量，也可以是记录被试者达到全部准确回忆水平一共需要多少轮刺激的量。这种实验的程序变式和特点，读者将会在下一节中了解到。

2. 再认类实验（详见本章第二节）这类实验有两种基本程序。第一是“是/否”式的实验，这是一种“二择一”程序。被试者的任务是从混入一定比例无关刺激的材料中，分辨出先前学习过的有关刺激来，反应报告只有“是”“否”两种。第二种是“迫选再认”程序，这是“多择一”程序。被试者必须在多个选择中作出一次反应，其猜中的概率小于“二择一”程序中的猜中概率。这两种实验的因变量既可以是简单的，也可以衍化成复杂的，犹如满足信号检测论（SDT）的计量方法。

以上分别讨论了回忆类和再认类的实验。回忆和再认是记忆过程的两种表现形式，是同一过程的两个方面，由于研究的需要，将两者分开来讨论，但绝对不是割裂。事实上，将这两个方面分离研究也只能是相对的，因此当实验者在根据因变量值解释过程的时候要审慎。最近一、二十年来，记忆研究中新理论、新方法不断出现，如何记录因变量的难度更大了，因而因变量的内涵也就更深了。

（三）控制变量

控制变量（或控制变项）（controlled variable）是在实验过程中会对因变量发生影响、应被实验者控制的因素和条件，对它控制得好坏是心理实验成功与否的关键。这个问题在高级心理过程的实验研究中表现得更加突出。记忆实验中要求实施控制变量有的是外显的，有的却是内隐的。控制外显的变量相对容易些。从构思实验起，就要充分考虑刺激材料、呈现方式、实验环境、仪器性能、间隔时间、计量精度、感觉通道、样本大小、被试者分布以及指导语等环节的编排准备，一般都要有效地防止干扰，保证实验顺利地进行。

实验设计中存在的问题是关于控制那些无从确切把握的干扰因素。这种干扰包括被试的动机、态度、情绪等社会性的额外变量，以及机体内部的变量，如饥饿、病症、疲劳等。研究表明，记忆过程中被试者机体内部的波动是十分活跃的，其中的奥秘目前还远远不能得到全面揭示。认识到有这一干扰，无非是使实验研究者更加重视控制额外变量。有经验的研究者往往采取间接的步骤，以把这种干扰尽可能地限制住。例如，完善实验前期的各项设计准备，巧妙地处理全部实验数据，然后是注重样组间以及实验间的比较印证工作。历史上那么许多著名的实验就是这样靠研究者的努力逐步积累完成的。可以这样说，记忆研究领域是当今对变量分析最为丰富的领域之一。

二、记忆研究的材料

记忆实验选用的材料是相当广泛的。一般说来，凡是人们在生活中能感受到的东西都可以作为实验材料，但是实际上由于实验室的条件限制和为了进行实验的方便，在实验中所使用的材料一般要符合以下标准：

1. 材料对每个被试者来说，在熟悉程度上应是相等的。这样才能保证所有被试者从同一基线开始学习，测出的效果才能归之于实验条件的作用。

2. 材料在数量上或意义上可以分为相等的单位。这样既便于记分，又便于对实验结果作数量化的分析。当然，这也不是绝对的，倘若以掌握全部材料所需要的时间或尝试的次数作为记分标准的话，就无需把材料分成相等的单位。

3. 材料应为被试者在较短的时间内就可以学会和掌握的。若一次使用的材料过多、费时过长，势必会影响实验的效果，同时让被试这么长时间留在实验室中也是不可能的。

根据上述标准，在记忆实验中可选用的材料还是很多的。这些材料概括起来有二大类：言语材料和非言语材料。言语材料又可分为无意义的和有意义的两类。

(一) 无意义材料

无意义音节是言语材料中的一种，言语材料是指音节、字母、单词、句子、诗歌、散文和数字等。其中音节、字母和某种条件下的数字都属于无意义的言语材料。在这些无意义的言语材料中，无意义音节被使用得较多。艾宾浩斯首创了“无意义音节”，目的是为他的记忆实验提供大量难度相等的材料，控制刺激项目可能引起的联想。他设计的无意义音节字共有 2300 个。每个由一个元音和首尾两个辅音组成。它的实验以自己为被试者。具体运用时，他把一切可能的拼合都一一单独写在一张纸片上，把它们洗均，然后从中抽出每一字表所需要的数目，直到每个音节都用过为止，然后把所有的纸片重新洗一遍。采取这种方法可能会碰到熟悉的组合字，但艾宾浩斯认为，对一个稳定的被试者来说，只要学习的字充分多，那些偶然出现的误差就能得到平衡。

下面，我们列出一张无意义音节的字母表（表 9-1）。

表 9-1 无意义音节字母表举例

1.	TAJ	YIC	HUZ	CEX	YAD	MEP
2.	ZIN	QOM	GOK	MOQ	FEP	GAW
3.	VEC	GEP	YIN	RUY	SUH	KOJ
4.	YOX	DUZ	TEV	GAF	JIK	CIB
5.	FUQ	RU	ZAD	LIQ	WOZ	ZUR
6.	BIP	NAW	XUR	KOC	LEQ	TEY
7.	DAK	XOL	QIG	QUZ	XAF	WOQ
8.	XEW	HUQ	LOJ	DEJ	MUJ	XIG
9.	CUG	TEF	DEH	TAH	RIY	NAH
10.	JOF	ZIK	BUP	WOG	KEB	JEC
11.	QID	VOB	WIX	FIK	QON	QUF
12.	LEH	PAH	KAQ	VUS	GUW	YOF
联想值	26.7	27.3	27.3	26.2	24.5	22.7

(采自 Ebbinghaus , 1885)

设计无意义音节的主要目的是要控制联想，但由于各语种的语言构成不同，一些规则并非都一样。例如，在讲英语的国家，由于英语中有很多单音

节字，若在遵循音节不能是熟悉的字和只许是三个字母构成的两个假定，就大大限制了无意义音节字的数目。但是，这无关紧要，重要的是无意义音节为我们提供了一条重要的原则。

（二）有意义材料

有意义的记忆研究材料的内容和种类十分广泛，可以是句子、诗歌、散文、词汇到图形、音乐等，很难尽述。研究者使用有意义材料从事记忆研究时，一点也不比使用无意义材料容易。由于使用有意义材料会使反应变量变得错综复杂，研究者不得不采用更多的控制措施，以保证实验顺利进行。这些措施还得以安排合理的刺激材料。例如，将有意义材料分解成容易归类的或便于定量的子单位，这些子单位可以是“意群”、“韵律”、“音节”、“词性”、“长度”等等。亨德森（Henderson, 1903）成功地选择一段能分成差不多相等的意义单位的材料，这种思想是有参考价值的。

巴特利特（Bartlett, 1932）认为一般的记忆实验由于采用无意义刺激材料，对于日常生活的记忆缺乏现实意义。事实上，一个人总有很大的潜力从材料中得出符合个人偏好的意义，剥夺这种机会就很难看到被试者如何把材料加以组织和吸收。奥尔伯特和普斯特曼（Allport & Postman, 1974）的一项研究结果与巴特利特的假说基本吻合。他们有意编造几个战时的谣言故事，并使这些故事相当自由地发展起来，然后注意它们在传播中的变化。结果证明，在传播过程中，每一个人都加上了一些他们自己在知觉和回忆中的错误，而这种变化与传说中的人对所在地区的恐惧和愿望相一致。

（三）其他刺激材料

其他刺激材料，如图形材料，设计时要考虑“复杂性”、“抽象性”和“相似性”这三个因素。前人的经验表明：被试者面对要求其学习、记忆的图形刺激所做的“意义搜索”过程，要比我们通常认为的“留下感觉印象”的过程复杂得多。由于这个原因，实验者为了取得有效的结论，选择刺激材料的任务就显得很艰巨。哈纳瓦尔特（Hanawalt, 1937）采用形状基本保持一致的刺激材料，研究被试者根据模画和记忆所画的图画，揭示出知觉和画图的综合作用常常是记忆错误的原因。尽管现在关于图形记忆的研究很多，发展水平也大有提高，可是如何控制诸如知觉、动作之类的因素对记忆过程的影响问题仍然需要从材料的“复杂性”、“抽象性”和“相似性”入手。

除了“图片材料”外，其他比较特殊并具有典型性的材料还有许多，如听觉记忆材料、嗅觉记忆材料、动作记忆材料。一般说来，这些材料的设计思想都没有超出上述材料的范围，研究者只要把握住如何使材料更加适宜研究目的就可以了。

在人类记忆研究实验中，上述言语和非言语材料的区分只是相对的，因为人的第二信号系统（即语言）是与第一信号系统协同活动的，即使在学习和记忆的一些非言语材料中，也离不开言语的活动。例如，在迷宫的学习活动中，被试者的整个操作过程显然主要依靠触觉和动觉，但是也伴随有言语的活动，常常是用言语来支配和调解自己的动作。因此，在使用非言语材料进行实验时，应考虑到言语活动的因素，不能将二者对立起来。

第二节 记忆研究的传统方法

在 50 年代中期以后，许多心理学家倾向于用信息加工的观点去解释记忆，根据记忆过程中信息输入到提取所经的时间间隔不同，对信息的编码方式有不同的特点，一般把记忆分为三种系统（见图 9 - 2），这三种记忆系统，又称记忆的三个阶段——感觉记忆、短时记忆和长时记忆。记忆研究的传统方法就是针对这三种记忆系统而言的，在介绍传统方法前，先简单介绍一下这三种记忆系统。

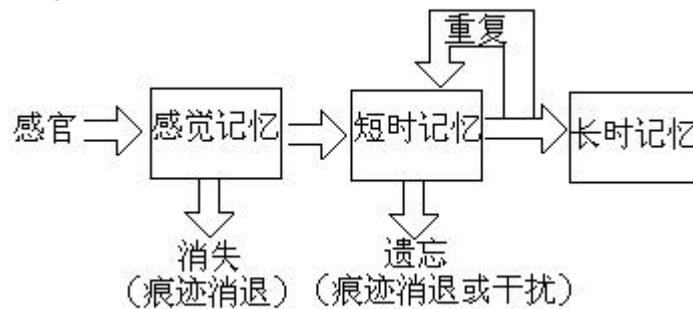


图 9-2 三种记忆系统图

感觉记忆（或感官记忆）（sensory memory，简称 SR）又称为瞬时记忆（immediate memory）。刺激直接作用于感觉器官便产生感觉知觉，当刺激停止作用后，感觉信息仍按原来的形式保持一个极短暂的瞬间，犹如按原先的样子登记了下来，这种短暂的保持或登记就是感觉记忆。在感觉记忆中材料保持的时间很短，大约 0.25 ~ 2 秒。例如，听人唱完歌后，好像耳朵里还有一点声音。看一个光点后，好像在眼睛里还有一点影子。在感觉记忆中信息完全依据它具有的物理性编码，有鲜明的形象性。例如，在网膜上的映像与刺激非常一致，基本上是外界刺激的复制品。感觉记忆中保持的材料如受到注意，它就转入短时记忆，如没有受到注意，它就会很快消失。

短时记忆（short - term memory，简称 STM）是指在一分钟以内的记忆。当我们从电话簿上查到一个需要的电话号码时，立刻就能根据记忆去拨号，但事过之后，就记不清了，这就是短时记忆的例证。另外，我们在上课时，边听边记，也是依靠短时记忆。短时记忆的广度（记忆广度指材料呈现一次后被试者能记住的最大值）大约为 7 ± 2 个组块。但如何组织好材料，在记忆储存的能量方面有重要影响。例如，电话号码 2824713492，可以把它分解为 282（地区号），471（电话局），和 3492（办公室）这三个部分，就能把这个 10 位数号码保持在短时记忆中。

多数人认为短时记忆是以言语形式编码的。英国心理学家康拉德（Conrad，1964）做了实验，他用两组容易混淆的字母 BCPTV 和 为实验材料，发现记忆混淆经常发生在声音相似的项目间（如 S 和 X）。但在短时记忆中还具有其他方式的编码。

随着近代科学技术的进步，自动化程度的提高，依照仪表的数字显示进行操纵和控制的门类越来越多。因此，短时记忆的研究越来越引起人们的重视。过去那种认为短时记忆只是为生活琐事服务的观点，显然已过时了。在现代化工业生产中，在军事上，若短时记忆出差错，往往会误大事。更重要

的是，短时记忆经过重复后就会进入长时记忆。

长时记忆 (long-term memory, 简称 LTM) 是从一分钟以上直到许多年甚至终身都能保持的记忆。从信息来源来说，它是对短时记忆加工复述的结果，但也有些长时记忆是由于印象深刻，如富有情感的事物，便能一次形成。在长时记忆里，大多是以意义的方式对信息进行编码的，如呈现一个词单：铅笔—橘子—床—狼—桌子—猫—苹果—毛笔—钢笔—梨—椅子—狗。当回忆这张词单时，我们往往打乱原来的顺序，把铅笔—钢笔—毛笔，苹果—橘子—梨，桌子—椅子—床，狗—狼—猫分别联系在一起，即按意义加以整理、归类储存和提取。人类的长时记忆大多数是要经过言语加工的。材料的组合依赖于概念的分类。但在长时记忆中编码方式有听觉的，也有视觉的等等。

从艾宾浩斯开始，记忆研究的实验方法究竟有多少种已经无法计数了。老的实验方法不断被取代，而新的实验方法名目繁多，我们选择其中最一般的实验方法归类，加以介绍。这里提到的传统方法，只是区别于内隐记忆方法而言。

一、回忆法

回忆法 (recall method) 是研究记忆的主要方法之一，它是当原来的识记材料不在面前时，要求被试者再现出原来识记材料的方法，故而也称再现法 (或复现法) (reproduction method)。回忆法是从记忆中将识记过的材料提取出来的过程。研究记忆过程的材料和方法很多，由于人们对语言材料比较熟悉，实验效果也比较明显，所以这方面的研究实例也较多，各类书中也常见这方面的介绍。另一方面，不管采用何种材料，实验的方法都有普遍意义。回忆的实验程序设计自艾宾浩斯起，是一个继承发展的过程，新旧方法在运用时孰优孰劣要视研究的目的而定。回忆法又可分为三种主要形式：系列回忆法、对偶回忆法、自由回忆法。

(一) 系列回忆法

系列回忆法 (或依序回忆法) (serial recall method) 是后人对艾宾浩斯以及早期关于回忆研究程序的概括，这种方法的主旨是要测量被试者达到某种记忆标准所需要的学习时间或学习遍数，所以又称为学习时间 (遍数) 法 (method of learning time)。

系列回忆法的基本程序比较简单：实验者根据材料特点先确定“熟练的标准”，要求被试者背诵实验材料，直到符合标准为止。为了达到所订“标准”，被试者需要一定的时间或者经过多次“尝试”，实验者再根据时间或遍数确定被试者的回忆水平。1885年，艾宾浩斯运用这种方法研究了“音节表的长度对学习难度的影响”，其结果通过图 9-3 反映出来。此后，艾宾浩斯又研究了有义意材料的回忆问题，他选用英国诗人拜伦的诗《唐璜》中的六节作为实验材料，进行了七次实验。结果发现，对于长度约为 80 个字音的诗段，被试者只需诵读 8 次就可以正确背诵；而对于同样数量的无意义音节，差不多需要 80 次诵读才能达到“标准”。也就是说，需要用 10 倍的努力才能背出一首与诗歌长度相同的无意义音节。早期的系列回忆法应用范围

很广，为后来的研究提供了良好的基础。例如，有人在系列回忆法的基础上发展了节省法程序等等。

但早期的这种方法也有不少地方使人发生困难。例如，当被试者很快报告完毕，而检验又不能通过时，计量工作怎么做？又如，等被试者经过6次“尝试”学会了95%的材料，余下的5%可能再需要6次或更多次的“尝试”，计量工作又如何做？针对这些问题，艾宾浩斯（Ebbinghaus, 1902）、罗宾逊和布朗（Robinson & Brown, 1926）对原先的系列回忆法进行修正，目的有两个：（1）解决被试者在什么时候已把字表学习好的不确定性；（2）探索学习和回忆的进程。修正后的方法定名为提示法和预料法。所谓提示法（anticipation method）就是在限定的一次或几次呈现刺激材料之后，要求被试者背诵，主试在被试者背诵发生迟疑时进行提示，在发生错误时予以纠正，直到全部背出。实验者根据记下的提示次数或校正次数，计算得分。“提示”量的反面就是“预料”量，即哪些项目经提示就能“预料”。提示法一问世，衍生的方法就不断出现。如，采用逐个提示的程序，即要求被试回忆时，不管他是否能说出，主试者都及时予以提示，给予反馈或诱发的方法。这个方法直到今天仍为许多研究者采用。

运用系列回忆法可以很容易地研究回忆的系列材料的特点。通常的做法是一遍一遍地呈现系列刺激，直到被试者能正确无误地预先说出下一个项目为止，实验者统计被试者预报每一个项目出现的错误或错误率。麦凯里和亨特（McCary & Hunter, 1953）曾证明，用错误量和错误率计量的总趋势相同，但采用错误率指标似乎更利于排除难度变化对刺激材料的干扰。图9-4就反映了用错误量和错误率为指标，并且用二种速度呈现刺激的系列位置效应的曲线。

这项研究表明，系列的开始部分，或最初学的项目较容易记忆，末尾部分或最后（时间最近）学习的材料也容易记忆，而中间部分是最难记忆的。这就是系列材料效应（或序位效应）（serial - position effect）。从信息加工理论（或讯息处理理论）来看，系列材料效应则与注意和信息加工的策略有关。这些在以后还将论述到[见本节一之（三）]。

（二）对偶回忆法

对偶回忆法（paired recall method）又可称为成对联合法（或联对法）（paired associates, 简称PA），是由卡尔金斯（Calkins, 1896）于19世纪末提出。对偶回忆法的设计思想得益于艾宾浩斯等人的早期实验，到20世纪初期又促进艾宾浩斯及其他研究者的实验。运用这种实验程序，可以避免刺激项目的双重作用。因为在系列“提示”程序中（例如A、B、C三项），每一次提示项目（假定B项），既可能是前者（假定A项）的反应，也可能是后者（假定C项）的刺激。实验者要确定项目的实际作用很难，这一困难给理论分析带来了混乱。相对而言，对偶回忆法的刺激项目仅对相应的反应项目提供线索，意义明确。在功用上，对偶回忆法除了检查系列材料效应不甚方便外，兼容了“提示法”所能度量的所有变量。对偶回忆法的优点很明显，但直到20世纪50年代它才逐步胜过早期的系列回忆法，在记忆实验中广为应用。

对偶回忆法实验一般有两种程序，一是预期法，二是检验法。

1. 预期法 预期法 (anticipative method) 的程序分成两步：第一步先单独显示刺激项目，要被试者努力预想对应的反应项目；第二步将刺激项目和反应项目成对呈现。刺激材料全部呈现完毕后，实验者改变顺序做第二轮。每出现一次刺激项目就要求被试者尝试报告反应项目，不管被试者能不能报告，间隔一过就同时呈现刺激项目和相对应的反应项目作为强化或反强化，如此，一轮一轮做到被试全部记住为止。图 9 - 5 是预期法采用的 10 对词组成的词表。

预期法的实验程序使人联想到“系列回忆”中的提示程序。预期法汲取了提示法的养分并作了改进；首先改变了序列呈现的顺序，实验者每做完一轮就改变呈现排列，克服了上一提示对下一项目的刺激效应。其次改变不规则的插入式提示，排除了实验者提示操作中的额外干扰，有利于建立稳定的刺激-反应程序。

2. 检验法 检验法 (test method) 程序不同于预期法的地方是，被试者先学习材料的所有刺激和对应的反应项目。检验一般不规定速度，允许有足够的时间让被试者回忆，但具体操作时实验者仍可限定时间，比如 30 秒。检验程序也是一轮一轮地进行的，一被试者全部通过为终结，而每实验一轮要改变呈现顺序。根据这样的安排，检验阶段的刺激出示顺序必须不同于预期程序 (见图 9 - 6)。

检验法本质上更适宜对人类学习记忆的研究，因为在条件反射式的学习中如果延迟强化，其效果是有很大差别的；在检验程序中动物的条件反射同人类的言语学习间的差别就更明显了。



对偶回忆法对材料的选择和如何组对 (配对) 有很高的要求。安德伍德

等人 (Underwood et al., 1960) 的研究表明, 学习材料越有意义, 就越容易记住, 并且与刺激项目配对的反应项目有意义比刺激项目有意义更为重要。佩威奥等人 (Paivio et al., 1975) 证明, 词的具体性是衡量记忆难度

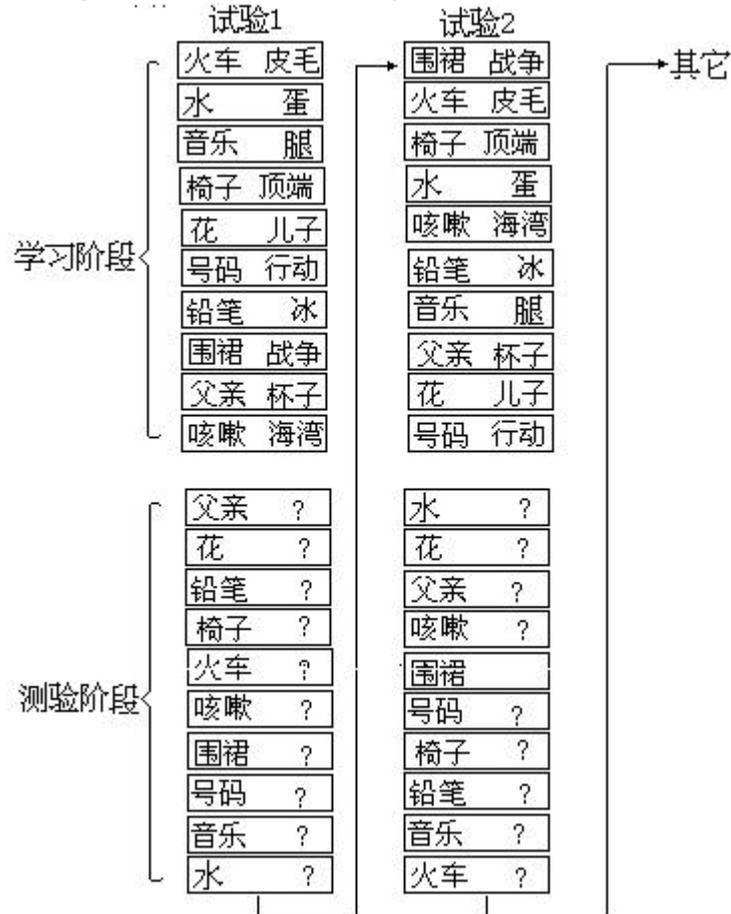


图 9-6 对偶回忆的检验刺激项目顺序
(采自 Hintzman, 1978)

的一个重要因素, 而流行的解释是, 具体的词汇能唤起生动的心理映象 (或意象), 使得联想刺激物更容易记住。事实上, 只要能够控制住最初学习的条件, 研究者是能揭示在后来的学习进程中意义性和具体性学习是如何引起旧知识向新任务迁移的。另一方面, 研究者还发现, 能够引起对偶联想学习的最强有力的一个变因是刺激物的相似性。以欣茨曼 (Hintzman, 1972) 实验为例 (见图 9 - 7), 他以大中学生为被试者, 分别学习相似性高或低的两种材料, 每组八对。相似性高的刺激项目由五个字母组合成四个三辅音, 相似性低的四个刺激项目无重复辅音字母。两种刺激材料的各项都分别与从 1 到 8 的数字组对:

	JCQ—7		RXL—1
相似性低	TFZ—5	相似性高	RNW—4
	KGP—2		RNL—3
	SBV—6		RXW—8

结果很明显, 当刺激相似性低时, 记忆容易; 相似性高时, 记忆受到干扰。由此可见, 在对偶回忆法实验中, 刺激材料和配对是极其重要的变量。

（三）自由回忆法

自由回忆法（free recall method）对所回忆的资料在顺序上不加限制。柯克帕特里克（Kirkpatrick, 1984）第一次采用我们现在称作“自由回忆”的方法，此法允许被试者可以随意回忆而不必遵循刺激顺序。这种方法看上去十分简单，当时的大多数研究者对之不屑一顾。艾宾浩斯（Ebbinghaus, 1891）对这种方法的评价是“粗糙、肤浅”，原因之一是它不符合联想主义的传统。20世纪60年代情况开始逆转了，自由回忆法大受欢迎，并成为最常用的回忆法。

自由回忆法的程序比较简单，然而一旦有了比较适宜的控制程序，用它进行研究的自由度就会很大。经过若干年的研究，人们普遍认识到自由回忆也遵循某些规律。被试者无顺序回忆实际上表现出三条规律：

1. 自由回忆法能体现系列材料效应 典型的数据出自迪斯（Deese, 1957）的实验。他们以每秒钟一个的速度向被试者念刺激词，刺激词的长度为10个或32个词不等，获得的结果可画成图9-8所示的曲线。这一实验揭示了二种效应：近因效应（或时近效应）（recency effect），表示末尾刺激的词记忆最好；首因效应（或初始效应）（primacy effect），表示最先呈现的刺激回忆效果次之。在一个长的刺激程序中，刺激的中间部分回忆时相对难一些，曲线中间部分起伏不大，处于最低部位，如图中的渐近线。

2. 自由回忆的顺序表现出一定的结构 我们可以通过两个实验来认识。其一为鲍斯菲尔德（Bousfield, 1953）的实验内容是回忆的分类倾向性。他以完全随机的方式向被试者呈现四类名词（动物、人名、职业名称、蔬菜名称）共60个。结果证明被试的回忆是分类“群集”的。其二为图尔文（Tulving, 1962）做的实验，他选了一系列假设是互不相关的16个单词作为刺激，并以不同的顺序呈现16次。分析被试者每次自由回忆的结果，发现每次反应的顺序与最后一次的反应顺序都有联系。图尔文把这种现象归结为主体组织（subject organization）。

3. 自由回忆并没有耗尽全部有关记忆 19世纪末柯克帕特里克就证明，自由回忆之后残存的记忆用再认实验还能提取检测。更简单的方法是进行提示回忆，一个不能自由回忆的刺激项目一经提示常常能回忆成功。根据这一特点，在有些实验中，研究者较多地选用再认法。

二、再认法

再认法（recognition method）是检查记忆保持量的方法，它是把识记过的材料（亦称旧材料）和未识记过的材料（亦称新材料）混合在一起，要求被试者把两种材料区分开来。再认记忆的研究已有百年历史，在行为主义高涨时期它受到冷落，当认知主义再度兴盛时，它又被广泛使用。自1960年以来，再认记忆研究在内容和技术的广度和深度上都是早期的这类研究不能相比的。再认和回忆两种方法构成了传统记忆研究的两种基本方法。

早期的再认记忆实验是在心理物理学实验的基础上形成的。当时有人将重量或声音进行前后比较的时距加长，用这种方法揭示被试者对刺激的保持究竟能维持多久。大约在20世纪初，有人开始应用再认法从事记忆项目判断的记忆研究。其程序分成两个阶段进行，先向被试者呈现一系列刺激项目，

接着将含有先呈现的刺激称为有关刺激 (relevance stimulus)，即识记过的材料，后呈现的包括有关刺激和无关刺激 (irrelevance stimulus) 即未识记过的材料。如有关刺激有 20 项，无关刺激有 30 项，合起来后呈现的就有 50 项刺激了。刺激呈现后，要求被试者逐项判断并作出“是”或“否”的反应，反应“是”的意思表示他认识这项刺激或判定这项刺激在第一次呈现中曾经看到过。反之，则反应“否”。实验程序要求被试者做的是“二择一”选择，体现了“再认法”的基本程序。采用这种程序时，有两种计算方法：

(1) 当“有关刺激数 > 无关刺激数”时，
再认分数 = (正确再认的百分数) - (错误再认的百分数)

(2) 当“有关刺激数 = 无关刺激数”时，
再认分数 = $\frac{(\text{正确再认数}) - (\text{错误再认数})}{\text{有关刺激数} + \text{无关刺激数}}$

由于在一个简单的“二择一”程序中，被试者猜中的概率就会很高，所以实验者应当尽可能地使用“多选”程序。对于一个“五择一”的程序，被试者完全猜对的概率只有 20%，实际使用效果会相当不错。但是，上述两种计算公式毕竟受“猜测”因素的干扰，即使用了“五择一”程序，理论误差也是很大的，下式能较明白地说明关系，校正误差：

再认分数 = $\frac{(\text{认对有关、无关刺激量}) - (\text{认错有关、无关刺激量})}{\text{有关刺激数} + \text{无关刺激数}}$

欣茨曼 (Hintzman, 1977) 认为，再认记忆研究中最引人注目的成果是关于记忆容量的研究，他例举谢波德 (Shepard, 1967) 的实验。谢波德以 68 对“二择一”刺激项目按一定的速度呈现给被试者，要求他们作强迫选择。三组被试者接受的刺激分为三类：第一组是单词；第二组是句子；第三组是广告图片。结果被试者对这三种不同类型的材料做出正确再认判断的平均百分比是：单词 88%，句子 89%，图片 87%。假如采用“再现法”实验程序，被试者能够达到的保存量大约只有 50%，即随机水平。

心理学自从引入了信号检测论 (或讯号侦察论) (signal detection theory, 简称 SDT) 后，再认记忆的研究工作如鱼得水，难以处理的问题也容易解决了。这主要是因为信号检测论有二个指标，用指标 d' 来表示再认能力，就不受心理偏向影响了。

三、再学法

再学法 (relearning) 也称节省法 (savings method)。此法要求被试者学习一种材料，达一定标准后，经过一段时间，再以同样的程序重新学习这些学过的材料，达到初次学习的标准为止。再学习与初学习两次学习所需要的练习次数之差，即代表初学习之后所保持的记忆。再学法是艾宾浩斯早期创立且颇为看重的方法。保持 (或保留) (retention) 是记忆以后，保留印象的过程。对于印象的记忆深度和印象的消失的速度，艾宾浩斯企图找到一个客观的度量指标，并把这种函数关系像物理定律一样表述出来。他发现，印象的深度可以用计算初学时反复诵读音节表的次数来表示。确定印象消失的速度或程度，只要在保持的不同时间后，重新学习这个音节表，通过比较重新掌握这个音节表的诵读次数 (时间) 和初次掌握它的诵读次数 (时

间)，便可明了。于是就有了再学法实验程序。

这种比较通常看成是节省的百分数：

$$\text{节省的百分数} = \frac{\text{初学时的诵读次数} - \text{再学时的诵读次数}}{\text{初学时的诵读次数}} \times 100$$

节省的百分数是测量保持量的一个有效指标。如某材料不需要再学就能回忆出来，则节省 100%；再学遍数同初学遍数一样，节省等于 0，识记的材料一点也没保持住。实际上节省为 0 的情况几乎没有，即使一点也回忆不出来的材料，再学仍能表现出节省的迹象。节省的百分数直接反映保持的深度，同时也间接表示“遗忘”的程度，这是一个极灵敏的尺度。

当然严格地说，我们不应把再学的遍数（时间）与初学的遍数（时间）相比，因为两次学习的个体条件可能会有差异。一种较好的设计是再学的遍数（时间）要与现行学习一个等价的材料所用的遍数（时间）相比较，虽然所谓“等价”的材料编制困难，但也不是不可能的。为研究学习程度对保持的影响，艾宾浩斯（Ebbinghaus, 1885）用再学法做了一项典型的实验，他把 60 张 16 个音节的音节表分别读了 8、16、24、32、42、53、64 遍，并在 24 小时后重新学习，直到背出。用节省分数所表示的保持量，表示初学时练习遍数的函数，如图 9-9。与重学节省的百分数间有着线性关系。图 9-9 上可见到一条拟合直线。

四、重建法

重建法（或重构法）（reconstruction method）。它的程序简单地说就是请被试者复现刺激项目的次序或排列。闵斯特伯格和比格浩姆（Munster-berg & Bigham, 1894）第一次做的实验，用 3.5 厘米见方的纸块涂上不同的颜色或写上数字，形成长度为 10 个和 20 个项目的两个刺激系列，用视、听和两者兼用三种方式呈现。简伯（Gamble, 1909）用气味、颜色和无意音节又做了重构的记忆实验。

重建法的实验程序一般都分两步：第一步，先由实验者向被试者呈现有一定次序或位置的刺激系列；第二步，将原刺激系列打乱后，交给被试，要求按照刺激呈现的次序或位置复原。当然，适合重建法程序的刺激材料并不限于文字言语，并且刺激项目的呈现方式可以是二维甚至更复杂的。实验者要做的工作中最主要的也许是如何记分。通常用的指标是以重建顺序的错误数或正确数衡量回忆水平，例如在闵斯特伯格的实验中，记分标准均以位置为准则，如果被试者将“黄—绿”材料误构成“绿—黄”，则累记为两个错误。实验中达到正确顺序的最大量是：

$$(\text{单个系列的项目数} - 1) \times \text{系列数} = \text{正确顺序数}$$

由于一系列中可能发生的最大错误数与最大正确数相同，即：

$$\text{错误的百分数} = \frac{\text{实际错误数}}{\text{正确顺序的总数}} \times 100\%$$

现有一实验的刺激系列由 20 项目组成，实验共做 10 次，每次在施测前先调整项目的排列顺序。根据公式，得正确顺序总数是 190，即 $(20-1) \times 10=190$ 经计算，实有错误数 42 个，则错误的百分比 = $42/190 \times 100\% = 22.1\%$ 。这是一种比较简单的方法。

重建法还有一种较为复杂的、且精确的统计方法。在上述简单的系列次

序的情况下，还需要运用斯皮尔曼等级相关（Spearman rank - order correlation）公式进行统计处理。这种方法对需要二维乃至多维的刺激构造可以获得分析数据。不过计算较为复杂。如果重建是一维的，即只有先后顺序的差别，其保持量（retention value）用等级相关系数表示，公式为：

$$r_R = 1 - \frac{6 \sum D^2}{N \times (N^2 - 1)}$$

r_R ：等级相关系数

D：重建顺序与原来顺序的差别

N：记忆材料的项目数

重建法除运用于语言文字材料外，也适用于其他非语言材料。

五、部分报告法

上面介绍的四种方法多见于长时记忆的研究中。近 20 年间，短时记忆是研究者们关心的内容，在这个领域中研究者们也取得了令人瞩目的成果。早期的研究者（Oberly et al., 1928）也研究过短时记忆问题，记忆广度是一个典型的例子，虽然现在人们还不时采纳传统的记忆广度法来开展工作，但已不能以此说明后来的整个短时记忆研究状况了。这是因为，当后来的研究者再次热衷于这项研究时，方法程序、内容深度都远不是早期的研究所能比喻的。不仅如此，更有研究者，如斯珀灵（Sperling, 1960）将研究的触角伸进瞬时记忆领域，开创了与以往完全不同的局面。

斯珀灵针对传统的短时记忆方法的某些缺点，设计了部分报告法（partial-report procedure），以便弥补以前那些用全部报告法（或整体报告法）（whole-report procedure）研究短时记忆的缺陷。全部报告法测量被试者依次识记的全部刺激，但实际上，即使被试者记住了某个量值的刺激，在报告时也总会遗忘一部分，实验者最后检测到的量值总是偏低的。部分报告法运用随机原理，使报告分段进行，有效地控制住报告时滞引起的误差。

部分报告法的刺激呈现与全部报告法相同。斯珀灵的研究将全部刺激项目排成 3×4 的矩阵，刺激以 50 毫秒的时程呈现，在刺激消失的同时，某一信号随即出现，被试者根据约定信号的种类（例如音调：高音——第一行；中音——第二行；低音——第三行）报告相应的刺激行项目。根据被试者报告的准确率很容易计算出被试者的记忆量。例如，对每行有 4 个字符刺激，被试者平均报告量是 3.04 个，则准确率为 76%，又知刺激项目的总字符是 $3 \times 4 = 12$ 个，以 76% 计准确率，则被试者的瞬时记忆量为 9.12 个。参见图 9 - 10。

X	G	O	K
J	M	R	I
C	U	T	S

图9-10
斯珀灵的部分报告实验材料法
(来自Sperling, 1960)

斯珀灵使用的刺激材料一般分为英文字母和非字音符号两类，部分报告法实验刺激的总项和项目含有的字符量都可变化，与此同时都施以传统的全部报告程序，对同一刺激进行比较研究。从图 9 - 11 上看到了两条曲线的差

异。那么，测量出来的感觉记忆的广度为什么会有那么大的差别呢？这个差别主要是实验方法造成的。有实验证明，若把回忆与呈现终止之间的间隔延长，回忆的成绩就会明显下降。若把声音信号在刺激终止后延迟 0.15 秒出现，回忆成绩就降为立即回忆的 60%，若延迟 0.3 秒，则降为 55%，延迟 1 秒，就降到接近传统说法的数量，约为 12 个字母中的 4 到 5 个了。在这 1 秒钟内究竟是什么原因使回忆受到如此重大的影响呢？斯珀灵认为，人的记忆系统中有一个视觉记忆系统。感觉信息以映象的形式保存在其中，但消退得很快，在 1 秒钟内几乎会完全消退。他还认为，传统所谓记忆、知觉或注意的范围是 4 至 5 个项目的说法，反映的并非最初的那个信息存储系统的容量，而是在映象消退之前能够提取出来转入下一个记忆阶段的信息数量。

图 9 - 12 是斯珀灵证明残留影象（或余象）（icon）假设实验的结果被试者可以报告的平均数字，随着延迟时间逐渐下降，但达到 1 秒钟后，曲线趋于水平，与全部报告法所得的平均数一致。这个结果表示视觉影象的贮存不超过 1 秒钟。

六、记忆广度法

记忆广度法（memory span method）乃是研究记忆材料呈现一次后所能记忆最大量的方法。记忆广度的研究最早是由贾克布斯（Jakobs, 1887）创用的，它是根据艾宾浩斯发明的系列回忆法稍加改动形成的。研究者事先准备好一系列若干项目的刺激材料，各项目分别有 3 到 12 个左右的数字符号。实验时，主试者口述或用速示器向被试者呈现某个刺激项目，刺激消失即请被试者按照同样的次序说出刺激内容。实验的目的是根据被试者的反应，度量他能正确记忆多少项目。刺激内容可以是数字、符号、单词等等。为了避免误差，研究者一般得准备多套等价的材料，每个刺激只对被试者使用一次。表 9.2 是一套材料的样本。

实验一般从一个短的刺激开始，逐步增加长度，直到被试回答发生错误为止。关于记分方法主要有两种：

第一种方法设每种刺激长度为三个项目，每一个完全记住的项目得三分之一分。假如被试者通过 6 个和 6 个以下长度的全部刺激，记 6 分。他还通过一次 7 个数目的刺激，一次 9 个数目的刺激，但 8 个数目的刺激

表 9-2 研究记忆广度的数字表

972	641
1406	2730
39418	85943
067285	706294
3516927	1538796
58391204	29081357
764580129	042865129
2164089573	4790386215
45382170369	39428107536
870932614280	541962836702

（采自 Woodworth et al., 1955）

没有通过，那么总分是 $6 + (2/3)$ 。有人认为这种方法在理论上有问题，于是引出第二种方法。

第二种方法若被试者做对了 8 个项目的刺激，但没有通过 9 位数的刺激。我们知道他的“广度”至少为 8，但不够 9，于是取两数的中点记分。实验可以进行几轮，算出每轮的平均值再平均计算得分。这种方法就像心理物理学计算阈值那样在理论上有所依托，因而被认为是一种改良的方法。

对此，奥伯利 (Oberly, 1928) 的一项研究常常被人引用。他的刺激系列的长度以 2 到 14 个数字不等，测量被试正确回忆的数目，结果如图 9 - 13 所示。图中 50% 的正确率对应的实际广度量约为 8.7。

由于记忆广度的实验程序比较简单，因此一经使用就开拓出许多方面的研究，从较早的文献中可以见到一些很有价值的研究。例如，记忆广度因年龄而变化；经过训练能获得较大幅度的提高；超广度呈现刺激材料，会使被试者降低广度。

七、分散注意法

关于分散注意对短时记忆的作用，早在 1900 年以前就由比格汉姆 (Big - ham, 1894)、史密斯 (w.G.Smith, 1895) 和史密斯 (T.L.Smith, 1896) 分别进行过研究，然而现今人们比较了解的是彼得森和彼得森 (L.R.Peterson & M.J. Peterson, 1959) 所做的实验。

分散注意法 (distractor method) 程序的关键是在记忆的保持阶段。为防止被试者利用刺激间隙进行重复练习，必须把他的注意从记忆的材料上转移开来，1959 年彼得森他们设计被试者对三个字母的组合或三个辅音字母的组合 (例如 X—J—R) 进行记忆测验。刺激以三个音串的方式呈现 1 秒钟，紧接着给被试者一个三位数字，让他从这个数字开始每次减去 3 地倒着计算。实验者在适当的间隔时给予一个信号，这时被试者必须立即对原刺激作出回忆。当时他们确定的回忆间隔分别定在 3, 6, 9, 12, 15 和 18 秒以后。从图 9-14 可见，仅间隔 3 秒钟，被试者显然就有了遗忘 (回忆的正确率是 90% 左右)，而到 18 秒钟时，回忆的正确率猛降到了 10% 以下，其遗忘的发展和艾宾浩斯曲线的趋势十分相近。

默多克 (Murdock, 1961) 用几乎一样的程序，只是选择了不同的材料重复彼得森的实验，结果除单词的保持时间较长些外，其他都相同。图 9-15 上可见到三条相近的曲线，1、2 为三辅音连串，4 为三词组合，3 是一个单词的曲线。

第三节 内隐记忆

在当今心理学的研究中，记忆课题已成为最为活跃的领域，尤其是在研究方法上的日新月异，引人注目。传统的记忆研究，常把注意力集中于对先前经验的有意识的、外显的恢复。这些研究中常用的方法是标准的记忆测验方法，如前面讲到的自由回忆、再认、全部报告和部分报告等方法。此类研究曾在现代认知心理学的记忆研究中占了绝大部分篇幅。自艾宾浩斯以来，已经完成了大量的有关记忆的实验研究，使我们对记忆有了一个较深入的了解。从信息加工的观点出发，记忆可简明地描述为刺激的编码、储存和提取

过程。基于记忆持续时间及其他性质的不同，记忆被分为感觉记忆、短时记忆和长时记忆三个阶段，或称三个类型的记忆。根据加工深度理论，它们只属于一种记忆类型，不同的只是它们的处理水平，尽管研究者们使用的具体方法、材料和实验条件各不相同，但这些研究间的一个重要的共同点是把记忆仅看作人们对先前经验的有意识的恢复。

近十年来，认知心理学、神经心理学和神经学的理论和大量实验研究已开始集中于一个共同的题目上。心理学和神经生物学的研究者们主张的一个共同点是认为存在着“多重记忆结构”。沙克特 (Schacter, 1985) 认为，多重记忆结构 (multiple memory form) 这一术语与其说是解释性的，不如说是描述性的，它是指功能相互独立的过程所引发的性质不同的记忆结果。说不同记忆任务上的不同成绩由不同的记忆结构所引起，并非指内部过程的本质，它只提供对有关资料的一个有用的描述。记忆的多重结构意味着记忆的不同类别或组合，无疑，这些不同类别的记忆具有许多共同特征，且常在记忆任务的操作上发生相互作用。

导致人们主张存在着多重记忆结构的根本原因在于有越来越多的证据表明，在有意识的外显恢复——外显记忆 (explicit memory) ——之外，还存在着一个相对独立的记忆系统，即内隐记忆 (implicit memory)，其根本特征是被试者并非有意识地知道自己拥有这种记忆，它只在特定任务的操作中能自然地表现出来，这种任务的操作不依赖于被试者对先前经验的有意识恢复。故这种记忆也被称作无意识记忆 (或潜意识记忆) (unconscious memory) 或无察觉记忆 (unaware memory)。

一、内隐记忆的研究领域

自 20 世纪初 50 年代以来，关于内隐记忆的实验研究已经积累了不少资料。概括起来，内隐记忆主要来自五个不同但又相互交叉的研究领域：再学时的节省，阈下编码刺激的作用，无意识学习，启动效应以及健忘症病人的残余学习。

(一) 再学时的节省

斯莱梅卡等人 (Slamecka et al., 1985) 认为，可以把再学时的节省当作内隐记忆的指标，再学一个先前学过的词表并不依赖于先前学习片断的外显记忆，外显记忆在这里是指能完全再认或回忆学习过的材料。而对那些不能再认和回忆的材料，由于曾经学习过，再学时就会缩短时间——表现为节省，这就可以认为是内隐记忆的作用。纳尔逊 (Nelson, 1978) 发现，既不能回忆又不能再认的项目也表现出节省。

(二) 阈下编码刺激的作用

尽管支持阈下知觉的早期实验遭到了严厉的批评 (Eriksen, 1960)，但最近利用新的实验技术所做的实验令人信服地证明：不为被试主观知觉到的刺激，能被知觉系统作出高水平的加工 (Fowler, et al., 1981; Marcel, 1983)。一些实验证明，不能被有意识地知觉到，因而也不能被外显地记住的刺激，能够对不需要有意识地恢复阈下刺激的任务行为成绩发生影响。这些任务包括自由联想 (free association) 及创造想象性故事和幻想 (imaginative story and fantasy production) 等。最近的研究清楚地表明，在被试很少或没有对阈下刺激的外显记忆的条件下，存在着对这些刺激的内

隐记忆。威尔逊和扎琼 (Wilson & Zajone, 1980) 给被试者呈现几何图形, 由于呈现时间仅 1 毫秒, 因此, 被试者无法有意识地知觉到这些图形。由强迫再认作指标的外显记忆仅在随机水平上, 然而, 当给被试者呈现两个图形——其一是呈现过的, 另一是新出现的——并要求他们选择较喜欢的一个图形时, 却表现出明显的内隐记忆, 被试者明显地偏向喜欢已呈现过的那个图形。谢蒙 (Seamon, 1983) 也报告了类似的结果。曼德勒等人 (Mandler et al., 1987) 的研究表明, 产生随机水平外显记忆的短暂刺激呈现能对非情感判断产生影响。研究表明, 给被试者短暂呈现一个令人讨厌的形容词能影响他对目标词的判断, 尽管事实上他缺乏对这些词的外显再认 (Bargh, 1986)。刘易基 (Lewicki, 1985) 发现, 在给被试者阈下呈现一个形容词和名词组合 (如 old-tree) 后, 再问及有关名词的问题时 (如 Is a tree big or old tree), 被试者趋于选择呈现过的形容词。

(三) 无意识学习

雷伯等人 (Reber et al., 1976) 关于内隐学习 (implicit learning) 的一系列实验中, 呈现根据不同人工语法规则产生的字母串, 被试者分别在无意识指导条件或外显指导条件下学习这些字母串, 然后再给被试者一些未学过的, 根据相同规则产生的字母串, 让其识别这些字母串是否符合语法规则, 他们的研究证明, 即使在不能有意地、外显地知道这些规则时, 被试者还是能够学会确定符合语法的字母串, 在语法规则较为复杂时, 记忆性的内隐学习比规则发现性的外显学习更有效。贝里 (Berry, 1987) 等研究者使用不同的方法同样观察到了两种不同的学习方式, 并对内隐记忆所获得知识的意识性、抽象性以及两种学习方式的相互作用等问题进行了研究, 发现无意识的学习机制比有意识思维更能检测微妙和复杂的关系。

(四) 启动效应

绝大部分的内隐记忆研究工作都是关于启动效应这一现象的。启动效应 (priming effect) 是指由于近期与某一刺激的接触而使对这一刺激的加工得到易化。启动效应一般又可分为重复启动效应和间接启动效应二种。重复启动 (repetition priming) 是指前后呈现的刺激是完全相同的, 即后呈现的测验刺激完全相同于前呈现的启动刺激。在启动研究中, 最常应用的测验方法有词汇确定 (lexical decision)、词的确认 (word identification)、以及词根或词段完词 (word stem or fragment completion)。在词汇确定测验中 (Scarborough, Gerard & Cortese, 1979), 要求被试者说明某特定的字母串是否构成一个合法的词, 由某一字母串在第二次呈现时被试者作出词的决定所用的时滞减少来反映启动。在词的确认测验中 (Feustel, Shiffrin & Salasoo, 1983), 给被试者短暂呈现某一刺激, 然后要求他们确认它, 通过与新项目的比较, 由被试者对新呈现项目确认精度的提高或确定新近呈现项目所需呈现时间的下降来反映启动, 在完词测验中 (Graf, Squire & Mandler 1984), 给予被试者一个词根 (例如 tab__, table) 或词段 (例如 __ss__ss__, assassin), 要求他们用想到的第一个合适的词来完成它。由对在先前学习词表上的词的使用的增加来反映启动。

在经典的词汇确定测验中, 要求被试说明某特定的字母串是否构成一个合法的词, 由某一字母串第二次呈现时, 被试者做出词的决定所用的时滞减少来反映启动。在词的确定测验中 (Feustel et al., 1983) 给被试短暂呈现某一刺激, 然后要求他们确认它。通过与新项目的比较, 由被试对新近呈

现项目确认精度的提高或确定新近呈现项目所需呈现时间的下降来反映启动。这两类内隐测验都可以归入重复启动的范畴。词汇确定测验也可用于间接启动 (indirect priming) 的研究中,进行单词识别的有关规律的探索。精确地说,重复启动范式与间接启动范式是有差别的,重复启动要求前后两次呈现的刺激是完全相同的,间接启动除包含重复启动之外,还允许两刺激有所差别 (Roediger et al., 1990)。

内隐记忆的研究主要使用重复 (直接) 启动范式,包括前面介绍的四类测验以及偏好判断、单词提取等。使用间接启动范式也可以进行内隐记忆研究。比如在音似和形似启动中,学习阶段的启动物和测验阶段的目标词是在某一特征上相似。由于这种相似,先前对启动物 (prime) 的接触也可以导致对目标词 (target) 作业的反应时或正确率提高或其他影响。

(五) 健忘症病人的内隐记忆

有关健忘症病人的内隐记忆研究可分为两大类:技能学习和重复启动。其中,技能学习的研究是由米尔纳和科金等 (Milner & Corkin et al., 1968) 开始的。他们发现,健忘症病人虽不能外显地记得他曾完成过某一任务,但能获得诸如追赶旋转体和镜形跟踪等运动技能。健全的运动技能学习和大致正常的知觉和认知技能在其他健忘症病人身上也得到了证明。

健忘症病人的启动效应研究是由沃林顿和韦斯克朗茨 (Warrington & Weiskrantz, 1968, 1970, 1974, 1978) 的一系列重要实验开始的。他们二人发现,虽然健忘症病人的自由回忆和再认测验成绩遭受很大破坏,但当用词根或词段作为线索时,他们能表现出对一系列熟悉词的正常的保持。但随后使用词段线索的研究发现,病人的成绩有时受到破坏。现已清楚,当使用词段或其他线索时病人是否表现出正常的保持取决于测验的内隐和外显性质,当给予被试外显指导时,其成绩就会受到破坏,而当给予被试内隐指令时,病人就表现出与正常被试同样数量的启动。上述观点已由大量实验所证实。

最近,考察健忘症病人是否对记忆中无预先存在表象的新信息 (如非词和无义词对等) 具有正常的启动或内隐记忆,获得的结果并不一致。塞马克等 (Cermak et al., 1985) 发现,健忘症病人在非词的知觉确认任务上无启动效应。戴蒙得和罗津 (Diamond & Rozin, 1984) 用三个字母的词根来测定内隐记忆,获得了类似的结果。格拉夫和沙克特 (Graf & Schacter, 1987) 使用完词测验发现,一些健忘症病人——其记忆障碍相当轻——对无义词对间接获得的联系具有正常的内隐记忆,而重度健忘症病人却没有对新联系的内隐记忆。莫斯科维塔等人 (Moscoviten et al., 1986) 用对降级的无义词对的阅读任务来评价内隐记忆,在有严重记忆障碍的病人身上观察到正常的对新联系的内隐记忆。总之,研究表明,健忘症病人能表现出对新获得信息的启动效应,但这种效应依赖于所使用的内隐记忆测验类型以及病人病情的轻重和程度。

除技能学习和重复启动现象之外,健忘症病人在其他情景下还表现出内隐、外显的分离。沙克特 (Schacter et al., 1987) 等人证明,健忘症病人能学习一些虚构的关于人们的信息,但不能外显地记得他们刚被告知过此信息。且健忘症病人对先前呈现的旋律有偏好的现象。他们还发现,在一个短暂的接触之后,健忘症病人能较快地发现隐藏的图形。在上述情形下,被试均缺乏对实验材料和学习片段本身的外显恢复。

二、内隐记忆现象的理论解释

(一) 什么是内隐记忆

上面，我们已对记忆中的新生领域——内隐记忆——作了一些介绍，但什么是内隐记忆呢？就目前的研究状况来看，对内隐记忆可从以下几个方面来理解：

1. 从现象上看，内隐记忆是被试在操作某任务时，不经有意识地回忆而存贮在大脑中的信息却会在操作中自动起作用的现象。这就反映出了先前所学内容的存在和作用。其特征是，被试对信息的提取是无意识的。

2. 从研究模式看，内隐记忆是启动效应的一种，这在内隐记忆与启动效应的关系中已有论述。

3. 从测量上看，内隐记忆是另一类记忆任务，这类任务不要求被试有意识地去回忆所学习的内容，而是要求被试去完成某项操作，在被试的操作中反映出其所学习内容的作用。这也是一类测量方法，即间接测验（indirect test），或可称为内隐记忆测验（implicit memory test），也有人称之为不自觉记忆测验（unaware memory test）。在记忆的实验性分离研究中，内隐记忆和外显记忆就是相分离的两种测量结果。

4. 在心理学研究理论上，研究者在对记忆的实验性分离现象进行深入研究后，提出一种理论假设——多重记忆说，推测记忆系统可划分为内隐记忆和外显记忆这两个在机能上相对独立的记忆系统（Tulving & Squire, 1967）。这就可以认为内隐记忆是通过间接测量和推理得出的，是一种假设的记忆系统的名称。

(二) 内隐记忆和外显记忆的区别

关于内隐记忆和外显记忆的区别，现有的结果至少在四个方面显示出外显和内隐记忆的不同：

1. 学习加工的水平对两种记忆具有不同的影响（Winnick & Daniel, 1970；Jacoby & Dallas, 1981）。对新形成的联系的内隐和外显记忆都需要一定程度的意义加工。但意义加工的水平（如用一词或造一句将两个无义词对联系起来）对外显记忆的效果影响较大，而对内隐记忆则影响较小（Graf & Schacter, 1985；Schacter & Graf, 1986）。

2. 学习和测验呈现方式的变化，对两种记忆具有不同的影响。雅克比和达拉斯（Jacoby & Dallas, 1981）发现，学习阶段所用听觉方式呈现材料，测验阶段用视觉方式呈现材料可减弱启动效应，使以单词辨识为指标的內隐记忆成绩显著降低，而对以再认指标的外显记忆则没有影响。格拉夫等人（Graf et al, 1985）用单词补全和线索回忆测验也得出相似的结果。张厚粟等（1989）的实验表明汉字音似的启动效应显著，而形似的启动效应则不显著。

3. 两种记忆保持的时间不同。这一研究目前具有不同的结果：有实验表明启动效应可持续几天或几周，而在相同的时间间隔下，再认能力则显著消退（Tulving, Schacter & Corkin, 1984）但也有实验表明：某些启动效应

非常短暂，几小时甚至几分钟就消退。而相同时间里的外显记忆却可保持（Graf, Mandler & Squire, 1984）。用汉字材料的实验表明，保持时间与词干补笔的可能性大小有关沫澄等' 1989）。

4. 格拉夫和沙克特（Graf & Schacter, 1987）的研究表明，某些干扰因素可以显著影响外显记忆，而对内隐的单词补全的启动效应则没有影响。

（三）内隐记忆的理论解释

由于内隐记忆现象的多样性和复杂性，目前还无一种理论能够成功地解释所有的内隐记忆现象。目前影响最大、争议最热烈的是多重记忆系统说和加工说这二种理论。下面我们简要地介绍这二种理论。

1. 多重记忆系统说 多重记忆系统说（multiple memory systems view）的核心是认为记忆的实验性分离现象反映了记忆系统存在着不同的子系统。内隐记忆和外显记忆现象分别代表了记忆的两种不同的子系统。图尔温等人（Tulving et al, 1985）提出启动效应（内隐记忆现象）代表一种新的记忆系统，即知觉表征系统（Perceptual representation system），且用随机独立性的概念来论证这一观点。按照这一论证逻辑，可推出‘内隐记忆的各种测验间应是随机相关的’结论。但威瑟斯庞和莫斯科维达（Witherspoon & Moscovitch, 1989）用实验证明：在两种内隐记忆测验间发现了随机独立性。这就出现了矛盾。究其原因可能有三：（1）论证逻辑正确，知觉表征系统独立存在，并可继续划分。（2）论证逻辑不正确，不能从实验性分离现象推出机能上独立的记忆系统。（3）记忆的确由不同的系统组成，但并不是以此逻辑论证所得。图尔温等人坚持此一论证逻辑，继续将知觉表征系统再细分为字词系统、结构描述系统和概念语意系统。

图尔温等（Tulving et al., 1990）设计了新的实验，从另一个角度论证了多重记忆说的观点。实验发现：在外显记忆测验中，针对同一目标的两种不同线索，结果表现出中等程度的相关；而在内隐记忆测验中，针对同一目标的两种不同线索，结果却是无关的。由此证明内隐记忆的表征和外显记忆的表征是不同的，内隐记忆代表知觉表征系统，这种系统没有单字的固定痕迹，是无痕迹的记忆系统（traceless memory system），但它可能包含单字的多重分布特征，每一重表征的唤起是通过一个特殊的线索来实现的。

神经心理学家在另一个研究层次上（认知神经学）提出了多重记忆说。有的还利用先进的正电子发射 X 射线断层照相术，以不同的推理逻辑论证了多重记忆说的观点。

在对健忘症病人的研究中，斯夸尔（L.R.Squire）把记忆分为两类：陈述记忆（declarative memory）和程序记忆（Procedural memory）。陈述记忆是唯一能进入意识中的可以觉察的记忆，而程序记忆只能通过记忆系统的操作来实现。另有一些神经心理学家把记忆分为知觉表征系统（perceptual representation systems, 简称 PRS）和语意记忆系统（semantic memory systems），其核心与斯夸尔的观点是一致的。斯夸尔的观点代表着神经心理学家对此类记忆现象的看法，但对于陈述记忆与程序记忆的神经机制目前研究的还不多。一般说来，陈述记忆涉及脑边缘系统的神经结构，它必须依赖大脑皮质及一些特殊脑区；而程序记忆中的启动效应则可能是普遍存在于神经系统（包括脊髓）所有主要神经节中的一种特性。

多重记忆说能较好地解释健忘症病人的记忆分离现象，因为各个记忆系统是独立的，当陈述记忆系统受损伤时，程序记忆系统仍保持完好。对于正常人的记忆分离现象表现为，某变量只影响陈述记忆而不影响程序记忆，或某变量只影响程序记忆但不影响陈述记忆。

2. 加工说 解释内隐记忆现象的另一种理论为加工说，加工说 (processing view) 认为，记忆的实验性分离现象反映了两类测验所要求的加工过程不同，并不说明记忆系统存在着在机能上相独立的两个不同的子系统。加工说是适当传输加工 (transfer - appropriate processing) 的简称，它与多重记忆系统的观点是对立的。

持此观点的代表人物罗迪格 (Roediger, 1990) 提出传输适当认知程序 (transfer - appropriate procedures approach) 的观点，认为外显记忆测验要求概念驱动过程 (conceptually driven processing)。概念驱动过程要求有意义的加工、精细编码和心理印象等加工过程，正是句子的精细编码以及材料的有意义组织等因素，提高了直接测量 (即外显记忆测验) 的成绩。因此，外显记忆测验要求概念驱动过程。与此相反，大多数内隐记忆的测验非常依赖于学习时与测验时的知觉过程的匹配程度。许多内隐记忆测验几乎是提取过去经验中的知觉成分。因此，认为内隐记忆测验要求材料驱动过程 (data-driven processing)。所以，影响概念加工的各种变量对内隐记忆测验就没有影响，而行为特征方面的变量对内隐记忆就会有很大影响；反过来也有这种分离。

布拉克斯顿 (Blaxton, 1989) 设计了一组构思新颖的实验，发现实验性分离现象有规律地依赖于加工方式，支持了加工说。有些研究者也获得了同样的结果。但沙克特最近在其系列研究中发现：某些线加工水平的变量对内隐记忆没有影响，这与加工说不相符。

加工说从另一个侧面成功地说明正常被试的两类测验间的实验性分离现象。材料驱动测验与概念驱动测验各自所要求的认知过程不同，若测验与学习的认知过程相匹配，则测验成绩就好；若不匹配，则测验成绩就差。这就会导致实验性分离现象。但是，此观点在对健忘症病人的实验性分离现象进行解释时却不理想。对于健忘症病人有意识记忆效果差、而无意识记忆效果好的现象的解释是，其概念驱动过程有障碍而材料驱动过程保持完好。但有些病人在那些要求概念驱动过程的无意识记忆中也表现出与正常人一样的启动效应，这就难以解释了。

综上所述，从某种意义上说，多重记忆系统说和加工水平说相辅相成，学术争论推动了科学的发展，揭示内隐记忆机理的日子已为时不远了。

第四节 内隐记忆的测验方法

心理学家为了研究复杂的心理现象，曾经用过形形色色的测量方法，但归纳起来，不外乎两种：一种是直接测量 (direct measures)，另一种是间接测量 (indirect measures)。

传统的记忆测量，主要有再认和回忆两种。再认是判断当前的刺激物是否为被感知过的某种信息。回忆是要求如实再现以前感知过的信息。这些方法在指导语言方面却是明确要求被试者有意识指向，回想他经历过的某事件，即具体发生在一定地方、时间的事件。这些方法测量的是被试关于该事件的

知识，如学过的单字是否回忆起来。这些传统的测量方法包括再认和回忆就是直接的测量方法。

近十年来，认知心理学家们正采用间接测量方法来研究记忆。与传统方法不同，补笔或词干补笔在指导语方面不要求被试专注于眼前的任务，也不是被试关于过去发生的某次事件的具体知识，而是被试的一些永久性的不依赖于时间、地点的知识。这样，根据指导语与测量的知识不同，把词干补全等这类方法称为间接测量方法。常见的间接测量方法还有知觉辨认法。

一、经典的测验方法

罗迪格等人 (Roediger et al., 1993) 归纳了 13 种内隐记忆测验的类型，包括：残画辨认，词干补笔，残词补全，词汇决定，词的确认，知觉辨认，同音词拼写，相关信息的自由联想，新联系的完词，偏好判断，条件作用，习惯化，单词提取，人脸辨识等。归纳起来，主要可分为词干补笔和知觉辨认两大类。

(一) 词干补笔

词干补笔 (word-stem completion) 是指被试学习一系列单字后，测验时提供单字的头三个字母，让被试补写其余二或三个字母构成一个有意义的单字。例如 jui__ 填成 juice，补笔的另一种形式是残词补全 (word-fragment completion)，是让被试学习一系列单字后，把缺一些字母的缺笔字填上适当的字母成为有意义的单字 (如 a_a__in 填成 assassin，__ys__ry 填成 mystery 等等)。虽然词干补全测量的是一些永久性的知识，但这些知识的提取也依赖于实验中发生的事件。

这里以沃林顿和韦斯克朗茨 (Warrington & Weiskrantz, 1970) 对健忘症病人所作的实验为例。他们取四名健忘症病人为实验组，16 名没有脑损伤的病人为控制组，两组被试在学习单字后进行四项测验。两项为直接测量：自由回忆和再认；两项为间接测量：词干补笔和模糊字辨认。实验结果见图 9-16。从图上看见：(1) 图 A 和图 B 显示，健忘症病人组在直接测量方法中成绩都比控制组差，且具有统计学意义；(2) 图 C 和图 D 显示，健忘症病人组在间接测量方法中成绩与控制组相比，虽略有高低，但无统计学意义，这说明健忘症病人也能以一定方式保持语词的信息。但是当时人们解释为健忘症病人不能从短时记忆中把词语信息传到长时记忆中。然而，这种间接测量方法和直接测量方法的比较使用，显示了方法学上的优越性。

(二) 知觉辨认

知觉辨认 (perceptual identification) 乃是在实验中，被试首先学习一系列单字，然后要求他在速示条件下 (30 毫秒) 辨认学过的单字以及另外一些未学过的单字。通常的结果是，学过的单字辨认率显著高于未学过的。在词汇决策实验中，被试首先学习系列单字，然后要求他对呈现的单字与无意义的字母串作出判断。通常反应时间是因变量 (或依变项)。

模糊字辨认 (word fragment identification) 是指在测验时呈现的单

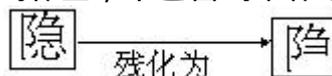
字的字母不大清楚，要求被试辨认是什么字。将单字模糊呈现的方法有很多，其中有一种是利用幻灯机投射焦距变化，焦点未集中时，屏幕上的投射是模糊不清的，由此造成模糊字。

这里以雅各比（Jacoby，1983）对正常人的实验为例。他让被试在三种条件下学习反义词。条件一为无上下文，被试看到“xxx——冷”而大声说出“冷”字，这是要求被试加工：“冷”字的字形。条件二为有上下文，被试看到“热——冷”的目标字并大声读出“冷”字，这是要求被试加工冷字的意义。条件三为产生（generation）条件，被试看到“热——???”，而要求被试说出“冷”字，这是条件一和条件二的结合，却要求被试既加工字形，也加工意义。三种条件下，被试的一半用直接测量方法测验——即再认；另一半用间接测量方法测验——即知觉辨认。实验结果见图9-17。

从图9-17上可以看到，用直接测量方法（再认法）来测量，结果显示：产生条件下击中概率高，有上下关系条件次之，无上下关系最差。与此恰恰相反的是用间接测量方法（知觉辨认）来测量，三种实验条件的结果显示：无上下关系条件下击中概率最高，有上下关系条件次之，产生条件最差。这样雅各比在正常被试身上做的实验结果竟和沃林顿和韦斯克朗茨在病人身上做的实验结果何等相似。由此可知，直接测量方法和间接测量方法所得的实验结果正好相反。

雅各比通过运用知觉辨认和再认测验的比较，发现了交叉双重任务分离现象，这是在正常人身上用任务分离式发现的第一个交叉双重分离现象。再认测验中，产生条件击中概率最高，有上下文条件次之，无上下文最差，这说明意义变量可促进再认测验的成绩。在知觉辨认测验中，三种实验条件的结果显示：无上下文关系条件下击中概率最高，有上下文次之，产生条件最差，这说明意义变量抑制了知觉辨认的成绩。同一变量在两类测验中表现了不同的结果，说明这两类测验彼此不相同，可能是两种底层过程（外显和内隐记忆）作用的结果，也可能是两种加工方式的差异造成的。

朱滢等（1991）运用类似于残词补全测验的方法，将汉字随机删去若干笔画，形成残字。测验时要求被试用首先想到的汉字，填出来，称为补笔测验，这种补笔实际上指残词补全，不包含词干补笔，例如：



补笔和知觉辨认是使用最广泛的内隐记忆测验方法。多重记忆系统说和迁移适当加工说都是在以比较内隐测验和外显测验的基础上，运用任务分离的范式，揭示两种测验表现出来的差异。前者认为内隐测验触及到的是一类不同于情节记忆、语义记忆的知觉表征系统。正是由知觉表征系统支持了内隐测验，情节记忆或语义记忆支配了外显测验，才使得两种测验表现出功能上的分离，否则，两种测验就应该有相同的功能表现。这样，如同再认测验揭示记忆一样，内隐测验就是揭示新的记忆系统的方法，人们就称之为任务比较方法学的技术。

任务比较方法学（task-comparison methodology）就是用不同的测验任务来揭示同一自变量（或自变项）的不同作用。它的基本假设是，完成不同的测验任务所需要提取的信息是不同的，因而参与的心理加工过程也不相同。这样，用不同的测验任务可以揭示不同的心理机能。如果同一自变量使不同测验任务有不一致或类似结果，就可以据此推测完成这些不同测验任务

的心理状态和过程之间存在差异或类似。

从方法学上分析，新的测验方法引入实验，常常会产生一些预想不到的结果，而这些结果可能会带来新的观点，特别是不同测验产生相反结果的情况，情形更是如此。人们把不同测验任务产生相反结果的情形称为实验性分离（experimental dissociation），它是任务比较中最有意义的部分。图尔温曾对实验性分离作了如下的描述：

符合实验性分离逻辑的实验是这样的：控制单一的变量而比较在两种不同的任务中变量的效应……，如果变量影响被试在一种任务中的操作，但不影响另一种任务的操作，或者变量对两种任务的操作的影响有不同的方向，我们就说分离产生了。（Tulving, 1985）

实验性分离包括几种类型的实验设计：单一的分离（a single dissociation），非交叉双重分离（uncrossed double dissociation），交叉双重分离（crossed double dissociation）和双向关联（reversed association）。单一的分离是指实验中的一个自变量（符号 V_1 ）影响一测验任务（符号 A），但不影响另一测验任务（符号 B）。上面我们介绍的沃林顿等人的实验就是典型的单一的分离实验设计。非交叉双重分离指的是，实验中有二个自变量，自变量 V_1 影响任务 A，但不影响任务 B；而自变量 V_2 影响任务 B，但不影响任务 A。交叉双重分离指的是，实验中一个自变量对两个任务有相反的作用。雅各比的实验就是典型的交叉双重分离实验设计。双向关联指的是，在同一实验中，两个任务的结果既是正相关的，又是负相关的。也就是说，超出了单向性的表现。

随着科学研究的深入，人们使用越来越多的内隐记忆测量，或者探讨新的记忆系统的特征等，或者对功能上的任务分离进行更具体和细致的客观描述，在介绍了内隐记忆基本的二种方法之后，下面我们介绍近期发展起来的一些内隐记忆测验。

二、非语词信息的内隐测验

传输适当加工说的支持者所做的工作多是选用语言材料，着重比较几种典型的内隐记忆测验的异同，考察出现的任务分离现象以及论证和发展加工理论的基本原则等。作为佐证，也有使用非语言材料的，如残图补全测验用作残词补全测验的参照，图片优势效应在内隐测验中的变化等。

多重记忆系统的支持者一部分来自神经心理学家，另一部分来自认知心理学家，认知心理学家近期的工作重点在于运用非语言材料，考察内隐记忆测验触及的底层记忆系统以及非语言信息的启动效应。这些非语言信息包括图片、形状和面孔等，对于非言语信息，又可分为熟悉和新异两种。接下来分两点介绍熟悉和不熟悉非言语信息的内隐测验方法。

（一）熟悉的非言语信息的内隐测验

这种类型的内隐记忆研究中，首先呈现图片或线条画给被试，这些图片或线条画都是熟悉的三维物体的二维特征。或者是有生命，如狗和牛；或是无生命的，如桌子或汽车。内隐测验时，要求被试识别知觉上已不完全的刺激，比如不完全的残图或简短呈现的物体的图片等非言语项目。为了考察图

片和词之间的转换问题，测验时还辅以呈现残词或快速呈现的单词等言语项目。下面举残图范式来说明。

实验时最初呈现的是最不完全的形式，接下来逐渐呈现稍微完好一点点的图片，直到被试能够识别。在 1 小时或 2 周间隔后，呈现新、旧残图，结果表明：年龄越小的儿童比年龄大的和成年人在达到识别时，需要更多的尝试，启动效应用第二次呈现的识别成绩与第一次呈现的识别成绩的比例来表示。比例分析表明存在显著的启动效应，重要的是，在启动效应的大小上没有年龄效应。与启动效应的结果相对照，在 yes/no 再认测验上则有很大的年龄效应，从 3 岁到成人，外显记忆稳定地增加，而启动效应量在 1 小时和 2 周保持间隔之间出现了衰退，但每个年龄组在长时延迟后仍显示出相当大的启动效应。分析实验的结果，年龄变量的控制导致了内隐测验和外显测验的分离，外显测验上存在着年龄效应而内隐测验则没有。如果两类测验触及的是同样的记忆系统，那么就表现出相同的结果而不是出现分离（dissociation），因此，实验性分离支持着一种新的记忆形成——内隐记忆的存在。

（二）新异非言语信息的内隐测验

至今已有一系列研究考察过新异非言语信息的活动效应，所谓“新异”的含意是相对于上述“熟悉的”而言的，其意是非熟悉。这类研究使用的内隐记忆测验在某种程度上类同于广泛使用言语领域的识别、完词和词汇决定等作业。两个近期的研究着重于点阵的启动效应，在对严重健忘症病人的神经心理学研究中，使用的目标材料是 3×3 点阵中 5 点的空间排列，它是由 4 条绒线连接组成的特殊因素。在给健忘症病人和一组控制被试呈现一系列这种点阵之后，用“点完成”测验来评估启动效应，在这个测验中呈现的是未被连接的 5 个点，要求被试用直线连接这 5 个点组成任意图形，每个图形都有许多可能的连法，主要是看被试是否倾向于把这些点连接起来构成他们曾经学过的图形。结果健忘症病人组和控制组都有启动效应，并且还观察到启动效应和外显记忆的分离，在要求被试外显地记忆前后启动过的点阵的再认测验中，健忘症病人的图形启动效应却未受损坏。

点完成测验（point construction test）是一种新型的内隐记忆测验。健忘症病人再认成绩为随机水平，而启动效应十分明显，这直接证明了内隐记忆的存在。同时看到，被试编码对象的特殊图形点阵，在被试头脑中不存在预存的表征，结果仍然出现了启动效应，即使在健忘症病人身上也有启动效应。

非启动信息内隐记忆测验使用的另一种间接测量方法是物体决定任务，由沙克特及其同事（Schacter, Bowers & Booker, 1989）首先使用。在他们做的一系列实验中，目标材料是新异的不熟悉的三维物体的两维图像。其中一半描绘的是结构上可能的物体，它们的表面和边界连接起来可以以三维形式存在。另一半图像与此相反，代表的是不可能的物体，它们的表面边界或轮廓违反常规，使得它们不可能以三维形式存在，详见图 9-18。在物体决定测验中，给被试快速呈现（100 毫秒）可能和不可能图形，要求被试决定每个物体结构上是可能的还是不可能的，一半的被测物体是事先已学习过的，另一半是新的，主要的问题是新异物体的启动效应是否能观察到，亦即

对可能与不可能物体的学习接触是否会提高判断正确率，用传统的直接测量方法之一（再认测量）评估外显记忆。

在这个实验中，先是学习任务，之后进行物体决定测验，诱使被试按照物体的全三维结构信息进行编码（要求被试指出每一个物体是朝左还是朝右的），结果出现了启动效应。相反，学习时要求被试按照物体的局部特征编码（指出每一物体的水平线多还是垂直线多），则未观察到启动效应。

这里，我们看到，运用随机独立性法则（law of stochastic independence）判定外显记忆和内隐记忆（启动效应）的独立性是多重记忆系统说的支持者们常用的方法；随机独立性的存在是支持知觉表征系统假设的。随机独立性法则来自于概率论关于相互独立随机事件的规定，若事件 A 与 B 满足 $P(AB) = P(A) \cdot P(B)$ ，则称 A 和 B 相互独立。该规定又是在乘法原则的基础上建立的。后者对 A 事件或 B 事件的要求是两事件依次进行才可能由始点到达终点，单独进行 A 事件或 B 事件就不能完成任务，因此，这就要求被试要经历内隐测验和外显测验，亦即测量（任务）变量应该是组内变量。如果定义 $P(A)$ 、 $P(B)$ 为简单概率、 $P(A \cdot B)$ 为联合概率的话，则简单概率指内隐测验或外显测验的正确率（或击中概率），联合概率是两类测验同时正确的项目与总测验项目的比例（朱滢等，1993）。

在左/右（即上例中要求被试指出某一物体是朝左还是朝右）情景中，两类测验的简单概率之积与联合概率无显著差异，因此，内隐记忆与外显记忆是相互独立的。

继这类实验之后，除了继续使用随机独立性法则证明两类测验的独立性外，研究的重点在于形成满足提取意识性标准的任务分离。实验者操纵加工深度变量与左/右编码作业的成绩。这是情理之中的，外显测验中有显著的加工水平的效应已为众多实验证实。但是，令人吃惊的是精细编码作业之后在内隐测验中未观察到任何启动效应。也就是说，随着加工水平的提高，启动效应下降至随机水平，任务分离出来了。加工深度（编码水平）变量的操纵导致在再认测验（外显记忆）和物体命名测验（内隐记忆）中表现出不同的结果，前者出现加工水平的效应，而后者则没有。

前面已介绍了词干补笔、残词补全、知觉辨认、词汇决定、词的确认、残图识别、点完成、物体决定等内隐记忆测验，这些方法的名称反映了它们的特点，它们在对记忆现象的揭示中，各显神通，表现了一定的生命力。

三、内隐学习的测验方法

内隐记忆侧重于知识的提取，内隐记忆测验指向的都是记忆的提取过程。指导语也是测验开始前给出的。在编码阶段，给出的指导语目的只是在于控制实验变量以形成分离而不是形成两种测验类型。内隐学习则不同，虽然有人认为内隐学习是内隐研究领域之一，但内隐学习范式与内隐记忆范式是有差别的，内隐学习更着重于知识的获得和编码，其指导语针对的是被试的学习方式，因此，内隐学习的测验方法与外显测验基本上是相同的，两类学习的差别只是在于被试以何种策略去编码知识。

这里举一例来说明，在研究内隐学习特征的实验中（杨治良、叶阁蔚，1993），学习阶段要求外显与内隐被试无意识或有意识地寻找相片分类的规律。此实验选取二个维度：一个为正侧维度，定义为具有显著特征的维度，

被试很容易通过观察相片的排列找出该维度的规则；另一个是容貌维度，定义为具有非显著性质特征的维度，极少有被试发现该维度及其规则。具体做法是：选取 50 张人物相片，根据明显的和不明显的两个维度把相片分成 4 类，每类相片有 4 张，练习和测验相片由共同的底层规则支配，在学习时，要求外显被试努力寻找相片分类的规则，对内隐被试只要求他（她）们

	丑	美
侧		
正		

把相片正确归类。两类被试进行相同的测验，要求被试对每张（共 16 张）相片作 4 次回答。第一判断选择的组号表示非常确信相片属于哪一类，第二判断表示如果第一判断失误，该相片还可能属于哪一类，第三、四判断依此类推。经过相同的测验后，不同的学习方式（内隐和外显）的学习效果有很大的差别，结果出现了内隐学习效应，并且发现了内隐知识和内隐学习的高选择力，高潜力和高密度贮存三大特征。

此一通过对内隐学习测验方法的分析，不难看到内隐学习的研究如果要与内隐记忆的任务分离范式联系起来的话，内隐学习范式符合提取意识性标准的要求，只是任务指导语出现的位置不同罢了，内隐记忆出现在学习后，测验前；内隐学习再现在学习前。

本章实验

一、短时记忆

（一）目的：

1. 测定短时记忆的广度；
2. 测定再认能力；
3. 学习回忆法和再认法。

（二）材料：

1. 记忆鼓，节拍器，数字盘：5~13 位数字各三组。
2. 图形卡片四套：每套 10 张，其中第三套和第四套卡片上的图形相似。

（三）程序：

第一部分：用回忆法测定数字的记忆广度

1. 主试和被试共同学习使用记忆鼓。实验开始时，被试坐在记忆鼓前的适当位置，要能清晰地看见记忆鼓的显示部分。

2. 主试先呈现 5 位数字，每个数字呈现时间为 1 秒钟，呈现完毕，被试须立刻把呈现过的数字按原来顺序背诵出来。主试用数字表核对，并记下

“对”或“错”。

3. 用上述方法呈现其他二组 5 位数字。如果在三组 5 位数字中，答对两组以上就算通过。

4. 按 2、3 的方法，再用 6 位、7 位、8 位等数字组进行实验，直到某种长度的数字系列有两次不能通过时为止。

第二部分：用再认法检查短时记忆的结果

1. 检查对两套差别较大的图形的再认：

(1) 主试把第一套 10 张卡片按每张 1 秒的速度呈现给被试者一遍。

(2) 主试把第一套和第二套图形卡片掺和在一起，然后把掺和了的 20 张图形卡片，仍按每张 1 秒的呈现速度依次呈现给被试看。要求被试对每张图形看后立即报告这张图形刚才是否看过，并记下答案“旧”或“新”。

2. 检查对两套差别较小的图形的再认：

(1) 把第三套 10 张图形卡片按每张 1 秒的速度呈现给被试看一遍。

(2) 主试把第三套和第四套图形卡片混合起来，按上述(1)、(2)的方法进行实验。

(四) 结果：

1. 统计被试的数字记忆广度（即最后通过的数字系列的长度）。

2. 按下列公式分别计算对图形的保持量：

认对的项目 - 认错的项目

$$\text{保持量} = \frac{\text{认对的项目} - \text{认错的项目}}{\text{原识记的项目} + \text{新的项目}} \times 100\%$$

原识记的项目+新的项目上式中，“认对”包括：对原识记过的图形反应“旧”，对新图形反应“新”。“认错”包括：对原识记的图形反应“新”，对新图形反应“旧”。

(五) 讨论：

1. 根据本实验的结果，分析说明短时记忆的特点。

2. 记忆材料的相似程度对再认有哪些影响？

本实验记录用纸：

记忆广度实验记忆用纸

数字位数	第一次呈现	第二次呈现	第三次呈现
5			
6			
7			
8			
9			
10			
11			
12			
13			

* 记下“对”或“错”

对材料再议结果记忆用纸

回答“旧” 或“新” 材料呈现序	材料名称	混合第一套 和第二套图形 卡片	混合第三套 和第四套图 形卡片	备 用
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				

二、长时记忆

(一) 目的：

比较有意义和无意义材料对识记和保持的影响；学习提示法和重学法。

(二) 材料：

1. 节拍器、马表或记忆鼓。
2. 词—数字 10 组，分写在 10 张卡片上；
无意义音节—数字 10 组，也分写在 10 张卡片上。

(三) 程序：

第一部分：用提示法识记对偶材料

1. 识记“ 汉语—数字 ”卡片 10 张：

(1) 主试按每张卡片 2 秒的呈现速度，把 10 张卡片依次向被试呈现一遍，并要被试记住卡片上的词对。

(2) 主试第二遍呈现卡片时，先呈现汉字，要求被试说出对应的数字。如被试说对了，就在记录表上记下符号“ ”，否则记下符号“ × ”。但不论被试能否说出，或说的正确与否，经 3 秒即呈现有关的字对。10 个字对依次以同法呈现。

(3) 随机地改变 10 个字对的呈现顺序，再按上述第二遍的方法继续做下去，直到被试能对每一字对看着汉字把对应的数字连续两遍无误的背出来为止。

(4) 记下这次实验结束时的时间。

2. 识记“ 无意义音节—数字 ”卡片 10 张。实验进行的方法同上。

第二部分：用重学法检查保持量。

1. 检查“ 汉语—数字 ”的记忆保持量：

(1) 记下这次开始实验时的时间，并注明与第一次实验结束时的间隔时间。

(2) 一开始就用提示法进行再学习，直至达到原来识记的标准，即连续两遍无误为止。记下第二次达到原标准的学习遍数。

2. 检查“ 无意义音节—数字 ”的记忆保持量：

(1) 这次实验开始时的时间与第一次实验结束时的间隔时间，要与再学“ 汉语—数字 ”的实验相同。

(2) 实验进行的方法同第一部分的 (3)。

(四) 结果：

1. 分别整理两种材料初学时达到学会标准所需学习的遍数，和每遍记住材料的百分数，用表和图比较识记两种材料的进度。

2. 分别整理两种材料在重学后达到学会标准所需学习遍数，和每遍记住材料的百分数，分别计算对两种材料的保持量。计算公式如下：

初学遍数-再学遍数

$$\text{保持量} = \frac{\text{初学遍数} - \text{再学遍数}}{\text{初学遍数}} \times 100\%$$

(五) 讨论：

1. 根据本实验结果说明不同材料对识记和保持的影响。
2. 如何用回忆法检查对偶材料的识记速度和保持量？比较提示法、重学法和回忆法的异同。

长时记忆实验记忆用纸（一）

遍 数		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
无 意 义 音 节	初	1															
		2															
		3															
		4															
		5															
	学	6															
		7															
		8															
	习	9															
		10															
 数 字 材 料	再	1															
		2															
		3															
		4															
		5															
	学	6															
		7															
		8															
	习	9															
		10															

长时记忆实验记忆用纸（二）

遍 数		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
无 意 义 音 节	初 学	1															
		2															
		3															
		4															
		5															
	习	6															
		7															
		8															
		9															
		10															
 数 字 材 料	再 学	1															
		2															
		3															
		4															
		5															
	习	6															
		7															
		8															
		9															
		10															

本 章 摘 要

1. 自从艾宾浩斯在 1885 年发表他的实验报告后，记忆就成为心理学中实验研究最多的领域之一。20 世纪 50 年代出现的信息加工理论（或讯息处理理论），把记忆看作是信息的输入、编码、储存和提取的过程。因而按储存的时间将记忆分为：瞬时记忆（或感官记忆）、短时记忆和长时记忆。

2. 近年来一些新观点对传统看法提出挑战，越来越多的证据揭示，在有意义的外显记忆之外，还存在着一个相对独立的记忆系统，即内隐记忆。

3. 记忆研究成果和研究方法是息息相关的。记忆研究的自变量，必须具备定性和定量二个条件。

4. 记忆实验中可选用的材料是很多的。这些材料概括起来有两大类：言语材料和非言语材料。

5. 回忆法又称再现法，是研究记忆的主要方法之一。它是当原来识记材料不在面前时，要被试再现出原识记材料的方法。回忆法又可分为系列回忆法，对偶回忆法和自由回忆法三种。

6. 再认法乃是检查记忆保持量的方法。它是把识记过的材料（称旧材料）和未识记过的材料（称新材料）混合在一起，要求被试把两种材料区别开来。

回忆和再认两种方法构成了传统记忆研究的二种最基本的方法。

7. 斯珀灵针对传统的短时记忆方法的某些不足, 设计了部分报告法, 将研究的触角伸进感觉记忆的领域。开创了与以往不同的局面。

8. 内隐记忆主要来自五个不同但又相互交叉的研究领域: 再学时的节省、阈下编码刺激的作用、无意识学习、重复启动效应以及健忘症病人的残余学习。

9. 词干补笔和知觉辨认是使用最广泛的内隐记忆测验。多重记忆系统说和传输适当加工说都是在比较内隐测验和外显测验的基础上, 运用任务分离的范式, 揭示两种测验表现出来的差异。

10. 任务比较方法乃是用不同的测验任务来揭示同一自变量的不同的作用。它的基本假设是: 完成不同的测验任务所需要提取的信息是不同的, 因而涉及到的心理机能也不相同。这样用不同的测验任务可以揭示不同的心理机能。

11. 内隐记忆侧重于知识的提取, 而内隐记忆测验指向的都是记忆的提取过程。指导语也是在测验开始前给出。内隐学习则不同, 内隐学习更加侧重于知识的获得和编码, 其指导语则是针对被试的学习方式。内隐和外显两类学习的差别只在于被试以何种策略去编码知识。

建议参考资料

1. 王甦、朱滢等(1993): 当代心理学研究。北京市: 北京大学出版社, 37~67页。

2. 吴敏、杨治良(1994): 试论内隐记忆的性质和理论解释。心理学动态, 2卷1期。

3. 杨治良(1990): 实验心理学。上海市: 华东师范大学出版社。

4. 杨治良、叶阁蔚(1993): 内隐学习“三高”特征的实验研究。心理科学, 16卷3期。

5. 张春兴(1989): 张氏心理学辞典。台北市: 东华书局(繁体字版)。上海市: 辞书出版社(1992)(简体字版)。

6. 彭聃龄、张必隐(1997): 认知心理学。台北市: 东华书局。

7. Greenwald, A.G., & Banaji, M.R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, Vol.102, No.1, 4~27.

8. Merikle, P.M., & Reingold, E.M. (1991). Comparing direct (explicit) and indirect (implicit) measures to study unconscious memory. *Journal of Experimental Psychology*, 17(2), 224~233.

9. Schacter, D. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology*, 13(3), 501~518.

10. Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University.

第十章 情绪的实验研究

本章内容细目

第一节 情绪实验研究概述

- 一、情绪实验研究的传统方法 521
 - (一) 印象法
 - (二) 表现法
- 二、情绪实验研究的变量 522
 - (一) 认知变量
 - (二) 行为变量
 - (三) 生理变量

第二节 情绪的生理指标

- 一、皮肤电反应 532
- 二、循环系统的指标 534
 - (一) 脉搏
 - (二) 血管容积
 - (三) 血压三、呼吸 537
- 四、语图分析法 539
- 五、脑电波 540
- 六、生化指标 541
- 七、指标的综合使用 543

第三节 表情研究方法

- 一、表情研究的理论根据 545
- 二、面部表情的早期研究 546
 - (一) 面部表情的直线量表
 - (二) 面部表情的圆形量表
 - (三) 面部表情的三维模式图
 - (四) 表情判断的早期方法
- 三、面部表情的现代测量技术 552
 - (一) 面部动作编码系统
 - (二) 两个测量系统

第四节 主观体验测量方法

- 一、形容词检表 558
 - (一) 心境形容词检表
 - (二) 情绪-心境测查量表
- 二、维量等级量表和分化情绪量表 564

- (一) 维量等级量表
- 三、应用性情绪量表 567

本章实验

- 一、广告悦目测定 570
- 二、情绪对动作稳定的影响 572

本章摘要

建议参考资料

传统心理学把情绪列为心理现象的三大方面之一。情绪也是心理学理论体系中一个不可缺少的研究环节。

情绪(emotion)是体验,又是反应;是冲动,又是行为;它是有机体的一种复合状态。情绪的表现有和缓的和激动的,细微的和强烈的,轻松的和紧张的等诸多形式,广泛地同其他心理过程相联系。

自古以来,科学家们十分注意探讨情绪之奥妙,但与情绪的重要性不相适应的是,长期以来情绪研究一直是心理学尤其是实验心理学研究中的一个薄弱环节。造成这一现象的最主要原因是情绪所特有的复杂性以及由此衍生出来的情绪研究方法学上的困难。

情绪研究的第一个困难是因为它发生在特定的情境中,具有外部的行为表现,同时又伴随有内部的生理变化。情绪与环境、认知和行为存在着复杂的相互作用,它在有机体生理和心理的多种水平上广泛地同其他心理过程相联系。当人们使用“情绪”一词时,可以是指完整意义上的情绪,也可以是指它的某一侧面。而研究者在具体从事研究工作时,其工作范围往往只能是情绪的某一局部。这样,不同研究者之间在出发点、研究方向、所用方法以及所得结果乃至结论上存在着种种差异,这就造成了情绪研究和情绪理论建设中的特殊困难。难怪有人夸张地说,除了心理学家之外,人人都知道情绪是什么,然而心理学家却不能准确地描述它。

情绪研究的另一个困难是:在严格的实验情境中引起的情绪带有明显的人为性质,难以为科学研究提供正确的依据;而在复杂现实情境中的观察则又太笼统,难以进行定量分析。一个典型的例子是,在森林中知觉到一头老虎能引起恐惧,但在动物园里知觉到关在笼子里的老虎就不产生恐惧了。在实验室情境中,由于被试知道这是一项实验,本来在现实中会引起某种反应的情绪刺激,这时可能引不起或者只引起强度较低的反应,甚至可能改变其性质而成为另一种反应。

值得欣慰的是,所有这些困难都没有阻止情绪研究的发展。尤其是在过去的二十多年中,心理学的这一重要领域在研究方法和理论建设上都有了质的变化。据此,在本章的写作旨在于回答以下五个问题:

1. 举例说明情绪研究中的几种变量。
2. 情绪的生理指标有哪些。
3. 面部表情的现代测量技术有哪些。
4. 为什么主观测量方法在情绪研究中占有重要地位。

5. 什么叫形容词检表，它有哪些。

第一节 情绪实验研究概述

在情绪研究中，伊扎德（Izard, 1980）总结了过去情绪研究中的问题，并提出了情绪研究的一些主要因素，其中最为突出的是情绪的多维量性和多成分性，这就决定了在研究方法上不能把它规定为单一变量的现实。本节正是从这个原则出发，讨论情绪的研究方法和实验变量。

一、情绪实验研究的传统方法

传统心理学研究情绪的基本方法可概括为两种：印象法和表现法。

（一）印象法

印象法（impression method）乃是通过谈话或问答来了解被试的情绪体验。其典型方法是：给被试一个情绪刺激，然后要求被试报告他的内部体验。例如，给被试听一段音乐，然后让他报告是否产生了情绪体验，是愉快的还是不愉快的，这些体验是何时以及怎样发生的，强度如何等等。早期的实验心理学家试图用印象法来得到对内省情绪体验的系统描述。例如冯特（Wilhelm Wundt, 1832~1920）曾在印象法的基础上，提出了情绪的三维体系：愉快——厌恶；兴奋——抑郁；紧张——轻松。现代情绪心理学家在使用印象法时还采用了问卷方式。例如伊扎德（Izard, 1972, 1974, 1977）以分化情绪理论为基础，建立了分别适用于儿童、青少年和成人的分化情绪量表。

现代的问卷法（questionnaire method）在一定程度上还可以对此进行量化，这是它的优点，也是其他方法所不能替代的。印象法的主要缺点是：（1）带有较多的主观色彩；（2）在某种程度上说，从不同被试得到的有关情绪体验的内省报告在比较时发生困难，例如，实验者很难判断一个被试所说的情绪形容词，非等于另一个被试所说的形容词；（3）情绪在内省时趋于消散，回忆时情绪又会降低强度。因此，在情绪研究中印象法一般不宜单独使用，需同时配以其他方法或指标。

（二）表现法

情绪过程既有内部体验又伴有一定的生理变化和外部行为，因此，可以将有机体的生理和行为变化作为情绪的指标加以测量和记录，这就是表现法（expression method）。表现法的主要优点是比较客观。早期研究者所关注的重点尽管不是情绪的生理过程和行为表现，但他们在情绪研究中仍然采用了表现法。他们在记录皮肤电反应、呼吸、脉搏和血压变化的仪器上作出了很多创造性努力。随着科学技术的发展，今天的研究者已经可以使用各种电子仪器，以各种方法测量并记录伴随情绪而发生的生理变化，例如呼吸、心跳、血管收缩、皮肤电反应、心电、脑电等等。与此同时，他们也可以采用照相、录像（或录影）等方法记录表情动作等行为变化，并借助计算机对测

量结果进行处理，以达到对情绪比较客观的测量。

但表现法也遇到不少困难，由于伴随情绪的生理变化本身并不等同于情绪，从严格的意义上说，对生理变化的测量只是对有机体“唤醒”水平的测量。至于表情，也能测量，也有人类的共同规律，但有时离开了具体背景，就很难判断（见图 10-13）。

情绪研究者一般认为，比较适宜的做法是两者并用，既用表现法取得对情绪的客观记录，又用印象法记录被试的主观体验。在情绪状态发生时，有机体产生各种不同的变化。对情绪反应的各种表现，实验心理学研究较多的是：（1）情绪的生理变化（第二节将讨论）；（2）情绪的表情动作（第三节将讨论）；（3）情绪的主观体验（第四节将讨论）。

二、情绪实验研究的变量

一般而言，心理过程愈复杂，则它对方法上的要求就愈高，对变量的控制也愈严格。从现有的情绪研究中，我们可以归纳出两种基本的情绪研究策略：（1）根据一系列已知的情绪反应寻找引起这种反应的各种刺激；（2）选择一种具体的刺激模式作用于有机体，确定所导致的情绪变化。前一类研究通常属于准实验研究，后一类研究则是真正的实验研究。

由于情绪实验条件中所涉及的变量极其繁杂，在所有这些变量中要严格地区分出固定的自变量、因变量和控制变量是比较困难的，因为所有可操纵的潜在变量在一个具体的情绪实验中究竟担任什么角色，取决于该实验的目的。因此，在叙述情绪研究中的变量时，将采用另一种分类方法，即把变量分为认知变量、行为变量和生理变量，下面我们分别予以讨论。

（一）认知变量

情绪实验中的认知变量（cognitive variable）包括情境变量、由实验者操纵的对情境或其他情绪刺激的认识解释、以及被试的自我报告。它们有些属于自变量，有些则属于因变量。

1. 情境变量 情境变量（或情境变项）（situational variable）指一种外部的、非物理和生理性的情绪刺激，它通常由一个事件以及事件发生的环境所构成。情境变量一般作为自变量，它需要通过认知才能产生作用。在大多数实验中，它与其他自变量共同发挥作用，或与其他因素共同构成一个自变量。

情境变量是情绪研究中应用较多的一种自变量。除了沙克特（Schachter）、瓦林斯和拉扎勒斯的工作外，其他的情绪研究也频繁地使用情绪作为一种情绪刺激。例如在情感移入（empathy）研究中，克雷格和洛厄里等人（Craig & Lowery et al., 1969）的实验均使用了情境变量。他们的基本方法是让被试观看一个人（通常是实验助手）正处在危险情境中，或者正在遭受电击（通常是虚假的），这些情境布置得非常逼真，使被试不自觉地作出相应的情绪反应。

情境变量作为一种变量，尤其是作为自变量时，就需要对其进行操纵或控制，情绪研究中操纵情境变量的具体方法有：（1）让被试直接处在某个情境中，例如战场上，飞行跳伞现场等；（2）被试并不直接处在某个情境中，

但让他（她）观看一个实际发生的情景，内容可以是悦目的人像，也可以是紧张性电影等；（3）控制认知所依据的因素，如期望等。

所有这些方法，一般都要求对被试隐瞒真实实验目的。由于情境刺激通常由多种因素复合而成，实验者对情境刺激中可能影响被试认知或情绪反应的种种因素事先应详加考虑，并谨慎地予以控制，否则极易发生“污染”实验结果。

2. 认知解释 认知解释 (cognitive explanation) 是指由实验者操纵的对情境或其他情绪刺激的认知解释。它可以单独构成一种自变量。根据特定实验的需要，这种认知解释可以是真实的，也可以是虚假的。

例如在沙克特 (Schachter, 1962) 的一个实验中，将被试分为三组。第一组，告知肾上腺素的效果。即向被试说明注射药剂后将产生长达 20 分钟的副作用，对这些副作用的描述与注射肾上腺素后产生的主观体验相同。第二组，不告知肾上腺素的效果。注射时告诉被试，药剂是温和无害的，而且没有任何副作用。第三组，歪曲肾上腺素的效果。告诉被试注射药剂后将产生双脚麻木、发痒和头痛等现象，这与肾上腺素的真实效果完全不同。通过这三种不同的认知解释，就可从一侧面了解认知解释在情绪中的作用。

3. 自我报告 由于情绪是一种体验，因而运用自我报告也就是理所当然了。早期的自我报告 (self report)，一般是由被试报告其情绪体验，近来的研究虽然仍然保留了这一方法，但更多的是以情绪量表或问卷的形式出现。情绪自我报告量表可分作两大类，一类是情绪状态量表 (Emotional State Scale)，一类是情绪特质量表 (Emotional Trait Scale)。前一类测量被试的情绪体验，在实验中用作因变量；后一类情绪量表测量的是被试的一般情绪反应倾向，即情绪特质，是常用作辅助手段选择或操纵被试变量的。我们知道，被试变量可以是自变量，也可以是因变量。

从形式上区分，情绪量表的形式主要有三种：

形式之一：情绪体验的直接提问，如“你感觉怎样？”或“你是否感到____？””。例如费斯巴哈和洛 (Feshbach & Roe, 1968) 的感情移入问卷即取这种形式。费斯巴哈和洛的这套问卷由一组幻灯片、解说词（作为情绪刺激）和问题“ How do you feel ? ” 组成。主试逐字记录被试的内省报告，然后根据被试共鸣性情感与预定引起的情感的吻合程度进行评分。

形式之二：列出描述情绪状态的若干形容词，由被试找出符合其当前情绪状态的形容词，例如形容词核对表 (Adjective Check List, 简称 ACL) 就被广泛用作心境测量的工具。

形式之三：运用现成的等级评定量表 (rating scale)。量表的通常形式是列出若干种基本情绪，然后由被试按其情绪体验在一个五点或七点量表上进行等级评定。如果所提问题是“你是否经常感到____？”，则量表又可作情绪特质测量之用。例如艾波斯坦 (Epstein, 1972) 用以测量成人情感移入倾向的问卷和伊扎德的分化情绪量表 (DES) 均取这种形式。在情绪等级量表上列出的，可以是情绪名称，如愉快、愤怒、恐惧等，也可以是对各种情绪体验的具体描述，或者是列出各种典型的情绪行为（见表 10-1）。

表 10-1 分化情绪量表情绪分类词

兴 趣	注意的 集中的 警觉的	厌 恶	不喜欢的 厌恶的 恶心的
愉 快	高兴的 幸福的 快乐的	轻 蔑	轻蔑的 鄙视的 嘲弄的
惊 奇	惊奇的 惊愕的 大吃一惊的	恐 惧	惊吓的 害怕的 恐惧的
痛 苦	沮丧的 悲伤的 消沉的	害 羞	忸怩的 害羞的 羞愧的
愤 怒	激怒的 生气的 狂怒的	内 疚	悔悟的 自罪的 该受责备的

(采自 Izard , 1977)

尽管自我报告在情绪测量中存在着明显的弊端，但由于它具有其他方法所不具备的优点，因此在各类情绪研究中仍被广泛使用，而且是最具有心理学特色的方法。这些，我们在第四节中还将给予讨论。

(二) 行为变量

行为变量 (behavioral variable) 在多数情况下是动物情绪研究和人类情绪研究中常用的因变量，它通常是指情绪行为 (明显的反应模式和显著的行为)。在少数情况下，行为变量也可能是自变量，例如前动作对后动作的影响是因果关系；又如反应时和正确率的权衡则是共变关系。显然，动物研究和人类情绪研究中所使用的行为变量存在着很大的区别。在动物情绪的研究中，行为变量可以分为二类。

第一类是典型的情绪行为，所谓测量也就是对典型情绪行为的直接描述和记录。例如，对猫而言，惊栗和冲撞表示恐惧，甩尾、弓腰、嗥叫、瞋目表示狂怒，等等。在动物情绪性 (emotionality) 的测量中，排尿和排便是最常用的指标。这种简单的描述方法有时也可以发展成颇为复杂的系统。例如布雷迪和诺达 (Brady & Nauta , 1953) 曾提出了情绪性的六种成分，它包括反抗触摸、叫声、惊跳、逃跑、遗尿、遗便，并以此为基础建立了评定量表。

第二类主要来自条件性情绪反应 (conditional emotional response) 和回避反应 (abient response) 的研究技术。简单地说，这种技术就是当某一自变量影响了某些正在进行中的操作时，实验者间接地记录下这个刺激产生的情绪影响。积极回避反应 (positive abient response) 是在信号刺激出现时作出的一种反应以避免某种即将出现的厌恶刺激。消极回避反应 (nega-tive abient response) 可解释为经过学习靠不作出任何反应而避免某种厌恶刺激。在这种技术中，测量的直接行为指标通常是动物的完成行为

(consummatory behavior)，即某种操作性行为。根据实验安排的不同，具体测量的完成行为可以有很大差异。例如，它可以是动物接近食物的潜伏期和摄食(水)量；可以在特定跑道上的奔跑速度；甚至可以是明度辨别任务。一种典型的方法叫做旷场反应(open field reaction)，又称户外行为(open field behavior)，它常被心理学家用作测量动物活动量和情绪性的指标。刘素珍(1995)发现：在突然的高强度噪音条件下，老鼠出现恐惧(phobia)的行为反应，动物在旷场反应中，爬格子数，由控制组的32.9格降至恐惧症组的16.1格。图10-1为旷场反应格子图。这些技术在自变量及因变量的设置和测量上有独到之处。

行为变量在人类情绪研究中也常用作因变量。除了对外显行为的直接描述外，还可采用照相、录像(影)等手段，或者用专门的符号系统记录被试的动作和身体姿势，并借助计算机进行分析。人类情绪研究中的行为变量还细分为三类。

1. 情绪性行为 情绪性行为(emotional behavior)是在特定情绪状态下的典型情绪性行为。由于没有一种单一的行为可以充分而必要地指向一种特定情绪(Lewis, 1982)。例如哭泣通常代表悲伤，但也有可能是因为愤怒，甚至是快乐。因此，作为因变量测量的应该是特定情境中具有内在联系的反应模式，而不是某种单一的行为。

2. 情绪表现 情绪表现(emotional expression)是指个体在情绪状态时，在生理上、心理上以及外显行为上的一切变化或活动，包括身段表情、面部表情和言语表情。身段表情包括头面部、躯体和四肢的姿势、位置、运动方向、速度以及幅度。人类的身体姿势不仅是一种重要的人际交往信号，而且在社会情境中常常表征着某种情绪。有关表情的研究，特别是面部表情的研究，我们在第三节将有专门的讨论。

3. 完成行为 完成行为(或完结活动)(consummatory act)即将被试作了一连串预备反应后，最终完成某项作业的情况以此作为因变量。这一变量在研究情绪的影响时经常采用。例如我国心理学工作者在研究情绪状态对幼儿认知操作的影响时，就使用了这种因变量(孟昭兰, 1985)。

行为变量在使用时，首先应注意具体情境，即同一行为在不同情境中可以代表不同的情绪；其次，还应注意文化差异，不同文化背景(譬如不同民族)的人在情绪的行为表现上存在一定的差异；第三，应注意个体差异，不同的个体在行为强度上可以有很大区别。这些原则在情绪实验中尤为重要。

(三) 生理变量

生理变量(physiological variable)包含的内容极为广泛。首先，我们可以把它分成两类：自变量的生理变量和作为因变量的生理变量。

生理变量作为自变量使用时，其方法和技术大部分直接来源于生理心理学的研究方法。最普通的方法有损伤法、电刺激法和化学刺激法。

1. 损伤法 损伤法(injury method)即损伤神经系统的某些部位，或者切断神经系统的某些部分之间的联系。这种损伤可以局限在神经系统的某一部分或整个断面，也可能是切除脑的某些部位。

有关情绪中枢机制的大量动物实验研究使用了这种自变量。例如，巴伦奈(Barenne, 1920)的经典研究就是切除动物皮质而使丘脑和下丘脑保持完

好；克里弗 (Kluver, 1937, 1938, 1939) 对动物颞叶中的海马、杏仁核、梨状区及额、颞皮层进行损伤；布雷迪 (Brady, 1953, 1955, 1958, 1960) 的系列研究是损伤动物脑的隔区；富尔顿 (Fulton, 1951) 切除猴和黑猩猩的额叶，损伤法对情绪机制和传导通路的探索都不失为一种好方法。

但科学的损伤法也存在着问题：首先是技术上的困难，即如何将皮层或皮层下系统区分为明确的区域，以及如何确定一个特定损伤或切除的位置和范围；其次是机能定位问题，涉及中枢神经系统的情绪研究都以机能定位为前提，但事实上很难确定它的确切含义，这就会导致自变量的某种混淆。再次，由于伦理上的考虑，这种自变量显然只能局限于动物被试，对于人类被试中只可寻找病理个案进行分析：但对病理个案进行分析时这种方法显然已不再是实验了，而是名副其实的个案分析法。

2. 电刺激法 电刺激法 (method of electrical stimulation) 乃是多种技术手段，达到电流刺激被试者中枢神经系统某些部位的目的。对动物被试，电刺激可直接针对脑部进行。脑部电刺激的最著名工作是由奥尔兹和米尔纳 (Olds & Milner, 1954) 完成的，他们在最初工作的基础上还进一步发现了电极的位置、电信号的频率和强度与刺激效应间的关系。图 10-2 是奥尔兹 (Olds, 1958) 所做的老鼠自我刺激的实验。A 是实验的照片，B 是实验的模式图。

奥尔兹用如图 10-2B 所示的斯金纳箱进行实验。在老鼠的脑子里装入电极，老鼠只要一按作为开关用的杠杆，电路就接通，装有电极的脑部位就会受到微弱的电刺激。老鼠经过一定时间的学习，就学会通过按压杠杆来控制电流对脑的刺激，即建立操作性条件反应。实验表明，如果在鼠脑的某些部位装入电极（如图 B 鼠脑中的 a 点），老鼠会无休止地、连续按压杠杆以进行自我刺激，如果在鼠脑的另一一些部位装入电极（如图 B 鼠脑中的 b 点），老鼠则会在按压一、二次杠杆后，就不再按压杠杆以避免刺激。

那些能引起老鼠连续按压杠杆的脑的部位在受刺激时给老鼠带来快感，因此就被称为“快乐”中枢或“奖励”中枢，它包括从大脑腹侧的嗅脑到脑干的很多部位。在这些部位装入电极，特别是在下丘脑的视索前核、前部、腹内核等处装入电极的话，老鼠甚至可以以每小时 8000 次的频率连续按压杠杆以进行自我刺激，直至精疲力尽进入睡眠状态为止。那些不能引起老鼠按压杠杆的部位在受到刺激时给老鼠带来痛苦，因此就被称为“痛苦”中枢或“惩罚”中枢，如中脑的内侧及其附近部位等等。

戴尔加多 (Delgado, 1970) 叙述了这一领域内一个重要技术的发展，这就是用无线电波对脑进行刺激，同时遥控记录大脑的活动。实验心理学工作者从这一基本技术中已建立了两种基本研究策略：(1) 用猫、猴、黑猩猩作被试，发现一个程序化的脑部电刺激可以改变其行为；(2) 对大脑颞叶障碍人实行双向交流 (two way's communication)，通过对脑电活动的连续监测，可以视察脑电活动与行为的关系。

电刺激也可针对外周神经（或周围神经系统）(peripheral nervous system) 进行，这就是常用的电击方法。在动物实验中，电击经常与其他因素相结合共同构成一个自变量，用来引起某种情绪状态（通常是焦虑）。这种方法在条件性情绪反应的研究中常用。例如布瑞迪 (Brady, 1970) 在实验中使被限制的猴子每隔 20 秒钟受到一次电击，但是，如果猴子按动一个特设的杠杆就可以使电击推迟 20 秒。这一自变量引起了猴子持续的焦虑。在人类

情绪研究中有时也用到这种自变量，例如，杨治良等（1979）曾用不同强度的安全电流，刺激腓肠神经（sural nerve），产生不同的痛情绪。

3. 化学刺激法 化学刺激法（method of chemical stimulation）一般也是以动物作被试，实验者可以使药物直接作用于被试的中枢神经，这是情绪生理研究中的一项新技术。在大部分情况中，这个过程对机能的影响是暂时的，不会产生永久性的组织损害，因此，它在阐述中枢神经系统与行为的一般关系是有效的。化学刺激所使用的药物一般是某种激素，如乙酰胆碱（其作用类似于副交感神经系统的活动）、阿托品（与肾上腺素效果类似）以及某种神经化学介质，如去甲肾上腺素、五羟色胺（存在于联系下丘脑和边缘系统的神经轴突的末端，并且在中枢神经系统中起着递质的作用）和麦角酸二乙基酰胺。化学刺激的一项新技术是在动物颅内植入一根套管，这样可以将少量的化学药品直接注入大脑皮层表面，从而影响皮层的机能活动。

使用化学刺激作为自变量所存在的问题是：（1）难以精确确定药物在脑内扩散的范围和速度；（2）药物发生作用的方式尚不明确，例如它可能直接影响行为而不是通过情绪影响行为；（3）不同种类动物对药物的反应可能存在差异；（4）缺乏有关的、可供比较的控制条件。其中前二个问题可能引起自变量的混淆，而后二个问题则使我们对实验资料的解释发生困难。尽管如此，化学刺激法仍然是情绪研究的重要方法之一。

化学刺激法也可通过注射等方式直接作用于整个有机体。药物的作用通常是使有机体处于一种激活或唤醒状态，如前面介绍过的沙克特和辛格的著名实验。某些药物可直接影响情绪状态，例如在精神病临床治疗中常用的抗抑郁药物和抗躁狂药物，此外，其他药物，如抗高血压药物利血平对情绪也可产生抑制性作用。与此相反，体育运动中形形色色的兴奋剂，则是运用科学成果达到不可告人的目的，这是为大家所唾弃的。

综上所述，作为自变量（或自变项）的生理变量一般通过直接作用于生理唤醒状态或作用于生理唤醒机制中的某个环节而影响情绪状态。但是，单一的生理刺激本身并不必然能引起情绪反应（尤其是在人类实验中），因此在作为自变量使用时，生理变量需要同其他变量，在人类被试中通常是与认知变量共同发挥作用。

作为因变量的生理变量（包括生理测量和生化测量），可以作为情绪状态或情绪反应的测量指标，即用作情绪研究中的因变量。下一节我们将进行详细讨论。

第二节 情绪的生理指标

上一节我们讲到，情绪有内心体验和外部行为表现，同时也有其生理机制。由于自主神经系统的活动，当有机体处于某种情绪状态时，其内部会发生一系列的生理变化，测量这些变化的指标就是生理指标（physiological index）。人们可能会认为自主神经系统的外周变化即生理指标的变化将有规律地表现出符合于各种情绪的不同模式，但是，大量的研究表明，除了少数外，生理指标的测量并未为特定的情绪提供明确的模式。例如，加拿大生理学家塞尔耶（Selye, 1974）的研究认为，不管引起紧张是由于生理的或心理的原因，也不管紧张是与愉快或不愉快有关，紧张的生化反应是一样的。相反，这种测量所提供的只是关于有机体所处的特定唤醒

水平的信息。在测量有机体唤醒水平时，我们能够采用的生理指标有几十种，但其中只有一部分与情绪有较大的关系，下面我们介绍最为常用的生理指标。

一、皮肤电反应

皮肤电反应是较早应用的生理指标。皮肤电反应（或肤电反应）（galvanic skin response，简称GSR）最早的名称是心理电反射（psychogalvanic reflex），它是由费利（Fere，1888）和泰赫诺夫（Tarchanoff，1890）发现的。费利将两个电极接到前臂上，并把它同弱电源和一个电流计串联。他发现当用光或声音刺激时，皮肤表面的电阻降低，电流增加。当时，用这种方法所测量的皮肤电反应称为费里效应（Fere effect）。传统的肤电反应测量方法以电阻为定量单位，最通用的电路是惠斯通电桥，使用时可以调节有刻度的电阻器到零点，来抵消未知的电阻，然后从电阻器上读出被试的电阻。费里的方法能够测量皮肤电的绝对水平及其变化，而且比较可靠，因此近代的这类仪器都应用这种原理。目前，电路中加入精密放大电路，创造出高水平的皮肤电反射仪。

费利和泰赫诺夫发现的现象依赖于同一基本生理过程，即由自主神经活动引起的皮肤内血管的收缩或舒张，以及受交感神经节前纤维支配的汗腺活动变化。应该注意的是，皮肤电反应所反映的是汗腺分泌反应，而不同于出汗量。

影响皮肤电基础水平的主要因素有三个：

1. **觉醒水平** 在正常温度范围内，手掌和足掌特别能反映唤醒水平，因此这二个区域是测量皮肤电反应的适宜部位。有证据表明，睡眠时皮肤电水平较低，但一旦觉醒，它就会很快升高（Farmer et al.，1925；Freeman et al.，1935）。在催眠条件下，皮肤电水平也倾向于降低（Davis et al.，1935）。瓦勒（Waller，1919）和韦克斯勒（Wechsler，1925）的研究表明，皮肤电水平在早晨较低，到中午达到顶点，而在晚上又降低，其变化似乎对应于一天内工作效率的变化。

2. **温度** 身体皮肤电主要反映身体的温度调节机制。因此，当气温很高，身体需要散热时，皮肤因出汗电水平就高；而气温较低，身体需要保存热量时，皮肤电水平就低。人的手掌和足掌也参与温度调节，但主要是在极端的气温情况下才参与。

3. **活动** 当被试正准备着某项任务时，皮肤电水平会逐渐上升；开始从事某一活动时，皮肤电水平将相应地升高到一个较高水平；而在休息时皮肤电水平降低。如果长时间从事某项难度不大的工作，皮肤电水平会缓慢地下降，但对难度较大的工作，这种变化就不明显。

许多心理现象与皮肤电水平有密切关系，情绪反应会引起皮肤电水平的急剧变化。韦克斯勒（Wechsler，1925）和塞兹（Syz，1926）的研究表明，带有情绪色彩的词能引起皮肤电反应，而重复刺激则能降低这种反应（见表10-2）。

表 10-2 情绪词引起的皮肤电反应（GSR）

刺 激	表现 GSR 变化的被试百分数		报告有情绪体验的百分数
	第一次听见	第二次听见	
被试的名子	84	68	36
被试的姓	74	40	28
接 吻	72	40	34
芳华虚度	48	20	4
未付的帐单	36	24	2
欺 骗	28	22	10
母 亲	26	12	30
笨蛋	18	24	0

(采自 Wechsler & Syz , 1925 , 1926)

愉快和不愉快的情绪刺激也能引起不同的皮肤电反应。戴星格 (Dysinger, 1931) 曾用能引起愉快感受的词 (如: 亲爱的、休假)、引起不愉快感受的词 (如: 呕吐、自杀) 和无关的词 (如: 筐子、作用) 作为情绪刺激进行实验。与无关刺激相比, 愉快和不愉快刺激均能引起皮肤电反应提高; 而不愉快刺激引起的皮肤电反应更为明显。参见表 10-3。

兰笛斯和亨特 (Landis & Hunt, 1935) 的实验是向被试呈现各种刺激物以试图引起惧怕、欢乐、性唤起、愉快和不愉快以及其他一些情绪状态, 并要求被试报告自己的主观体验。表 10-4 是实验结果。从表上可见到, 不同的主观状态, 引起不同的皮肤电反应。而且从反应值上看, 其差异也是很大的。其后, 许多实验心理学家作过皮肤电反应的研究, 他们的结果是一致表 10-

3 愉快和不愉快对皮肤电反应 (GSR) 的影响

刺激的作用	GSR (假定单位)
最愉快	129
愉快	98
无关	79
不愉快	101
最不愉快	147

(采自 Dysinger , 1931)

表 10-4 不同主观状态时的皮肤电反应

报告的主观状态	平均 GSR (假定单位)
紧张	1248
惊恐、惊异、惧怕	846
混乱	740
欢乐	514
预期	401
抑制、不确定	319
不愉快	260
努力	169
愉快	105

(采自 Landis et al. , 1935)

的，即：皮肤电反应可以作为情绪的生理指标。

二、循环系统的指标

自主神经系统不仅控制着皮肤电反应，而且也控制着循环系统活动。因此，循环系统活动的指标也可以成为情绪测量指标。经常用到的指标是：脉搏、血管容积和血压。

(一) 脉 搏

循环系统的动力来源是心脏跳动。可用几种不同的方式测量心脏活动，最直接的方式是记录心跳，另一种方式是记录脉搏率。因为心脏每一次收缩都发出一个“波浪”，通过动脉，所产生的脉搏可以直接感觉到，也可以借助脉搏描记器加以记录。脉搏率是情绪反应的良好生理指标之一。人们在日常生活中都有这样的体会，满意或愉快时，心跳正常；而处于紧张、恐惧或暴怒状态时，心跳就加速。伍德和霍根森（Wood & Hokanson, 1965）曾用脉搏率作为紧张的指标，研究操作（或作业）（performance）和激动水平的关系。实验用握力计来诱发紧张，亦即被试者的紧张程度随握力计压力的增加而加强。从图 10-3 的右图上可见，心率就是作为紧张度的指标。操作曲线呈倒 V 字型。图中实线表示心率，虚线表示操作或作业成绩。随着心率的不断增加，操作水平亦随之提高；但到一定程度之后，心率的增加反使操作水平下降。在这个著名的实验中，脉搏率就是情绪反应的生理指标。图中纵轴操作效率单位为假定单位，心率单位为与正常水平相比较之差，横轴为激动紧张水平。

(二) 血管容积

血管容积的变化是由自主神经系统控制动脉壁平滑肌收缩和舒张所造成的，即由局部血管收缩和舒张引起的。一些实验研究清楚地表明，人的某些情绪状态如紧张的脑力工作、生气、害怕、新异刺激等，都可以引起皮肤血管的收缩，这是由升压中枢的作用引起的。这种情绪刺激引起的反射作用使动脉压升高，从而使更多的血液进入脑中。当人感到为难或羞耻时，由于降压中枢的反射作用，会引起皮肤血管的舒张，更多的血液进入表面，从而表现出面红耳赤等情况。

北京大学心理系等（1973）曾研究了在紧张害怕以及痛刺激作用时的血管容积反应的特点（注 10-1）。他们的研究表明，紧张害怕与致痛刺激所引起的血管容积反应均表现为明显的收缩，但两者在动态过程上存在一定的差异：紧张害怕引起的血管容积振幅缩小、比较恒定，而且反应延续期较长；一般痛刺激作用下的血管容积振幅缩小、比较波动，且反应持续时间短。这两种情况所引起的血管容积反应，经过仔细的研究，可以找到它们之间的区别点。

(三) 血 压

血压变化与血管容积变化是相关联的，它们都反映了循环系统的活动情况有着相近的生理机制。早期，司科特（Scott，1930）做了一个著名实验，曾以 100 名医学院二年级学生为被试，研究血压变化与情绪状态的关系。他采用的方法是放映三段不同的影片，第一段内容是关于爱情的，第二段表现主角受到虐待，第三段描写城市被地震毁灭，主角处于危险中。他希望这三段影片能分别引起被试的性、愤怒和恐惧情绪。实验是单独进行的，在各段影片之间插入十分钟的无关影片，主试在被试们观看影片的过程中记录他们的血压变化。实验结束后，主试要求被试报告出自己观看影片时的情绪体验。从内省报告看，预期会产生愤怒和恐惧等情绪反应的影片，效果不够理想，但引起性情绪的影片是成功的。图 10-4 为观看三段影片时收缩压的变化情况，由图中可见，第一段影片（关于爱情的内容）引起的血压变化十分明显，粗线均处于图的右侧，100 名被试中有 88 名有明显变化。纵轴表示反应分配频率，单位为百分率。

三、呼 吸

在情绪状态时，呼吸系统的活动在速度和深度上会有所改变。对剧痛的情绪反应往往会使呼吸加深加快；突然惊恐时，呼吸会发生临时中断；狂喜或悲痛时，会发生呼吸痉挛现象。当人体受到某种刺激时，呼吸反应的“ \dot{V}_E ”值可大于正常，表现为呼气相对缩短，呼吸频率加快；“ \dot{V}_E ”值也可以小于正常，表现为呼气相对延长，出现憋气的情况。测量呼吸的方法一般有三种：

1. 吸气呼气比率法 吸气呼气比率法（method of inspiration-expiration）是测量呼吸的一种方法，它通过吸气和呼气的的时间之比求得，即求出 I/E 值，其中 I 是吸气的的时间，E 是呼气的的时间。有些研究者在测量 I 或 E 时略去停顿的时间，另一些人又把它们包括进去，使用不同的计算方法会产生不同的结果。心理学家斯特林（Storring，1906）在研究吸呼时间比率的改变与情绪的关系时发现，欢笑者吸气慢呼气快，呼吸的比率低，约为 0.30。瑞沃耳特等（RehwoIdt，1911）发现，人在恐惧时，吸气和呼气的比率从一般状态的 0.70 上升到 3.00 或 4.00，人在吃惊时，吸气是呼气的二到三倍等等。呼吸描记器可以记录到这些情况。图 10-5 为呼吸周期的记录。从图上可清楚地看到呼吸包括着两个时相，即吸入和呼出。

2. I 分数法 I 分数法（I-fraction）又称吸气相对时间表示法（method-of relative duration of inspiration and expiration），它是以整个呼吸周期的时间来除吸气的的时间而得到的，它表示吸气所占时间的比例。因为吸气通常是呼吸周期中积极的肌肉活动时期，所以 I 分数表示着供应空气到底要花多少时间。根据佛斯勒（Fossler）早期的研究，在说话时所有被试的平均分数是 0.163，范围在 0.090 到 0.258 之间。这是说，我们在说话时，为了供应所需要的氧气而平均失去了六分之一的说话时间。在平常安静地呼吸时，I 分数为 0.4~0.45，约占呼吸总时间的一半。

3. 次数法 次数法（frequency method）因呼吸的频率和深度变化一样，同样也反映情绪的唤醒水平。例如，在一般平静状态下呼吸频率为每分钟 16~20 次，非激动情绪下呼吸率变化不大，然而在愤怒和惊恐情绪下，呼吸频率可增至每分钟 40~60 次（见图 10-6）。

但是，呼吸次数和情绪状态的关系也十分复杂。例如，暂时的注意状态常常需要呼吸的部分或全部抑制（射手在精确瞄准时常常摒住呼吸）；而突然的刺激会使呼吸“摒息”。所以，呼吸作为一项生理指标，并不是简单地表示情绪特异性。这里也使我们看到情绪研究的复杂性。

四、语图分析法

语图分析法 (method of language pictorial analysis) 是新近发展起来的一种情绪分析指标。它是借助于语图仪中的声音应激分析器 (soundstress analysis apparatus) 来实现的。声音应激分析器可以测量出人耳所不能直接听到的语音的某些变化。它所以具有这种作用，是由于人的发音器官（包括声带）在气流通过时，因气流对发声管道的碰撞而发出轻微的颤动。人处于紧张情绪状态时，发音器官的正常颤动被抑制。在使用该仪器记录人的语声时，其记录曲线的描记速度比正常录音慢四倍，并以声音图像显示出来。

声音应激分析器也用于测谎。它的优点在于：人说话时，其发音的颤动不能人为地随意加以控制。多导生理记录仪的描记，由于机体的某种运动，例如肌肉的随意性紧张，有可能被干扰而影响描述的准确性；而声音应激分析器的效应则没有这个缺点。声音应激分析器的另一个优点在于，记录时仪器与被记录者之间不必用导线联接，只要通过收音或者录音，再通过该仪器的转换，就可以得到声音图像。目前，仪器市场上的语图仪和声音应激分析器，已成为动态记录声音反应的理想仪器（可参见第七章中有关语图仪的介绍）。

五、脑电波

脑电波（或脑波）(brain wave) 又称脑电图 (electroencephalogram, 简称 EEG)，有多种类型，如 α 波 (alpha wave)、 β 波 (beta wave) 和 δ 波 (delta wave) 等，它同心理状态有极为密切的联系。脑电波型主要与频率有关。

在情绪发生变化时，不仅有外周的变化，而且也会有中枢的变化，所以我们可以通过脑电活动的变化来测定中枢的变化。因此，利用脑电记录技术就可以测定出在一定情绪状态下所引起的大脑不同部位电位差的变化。这种研究方法主要用在研究下丘脑、丘脑与皮层的相互关系的比较中，据以推论人的情绪状态。

根据目前对脑电波的研究，人们初步探知，在强烈的情绪状态下，人的脑电波活动与正常状态下脑电波活动不同。此时， α 波消失，脑电波振幅降低。在焦虑状态下枕叶的 α 波消失，脑电波振幅降低；此外，在额叶区、运动区、颞叶区的脑电波振幅都较正常状态时小。在疼痛刺激时，脑电图上会出现 α 波阻断或不完全抑制的现象，并且快波增多；疼痛减轻时，快波逐渐减少，同时 α 波也开始恢复。近年来，国内外常用脑电图检查精神病等疾病，例如，尽管脑电图检查对老年性痴呆诊断的异常率存在不同看法，但所见脑电图改变是基本接近的，即认为脑电图改变为一般老年期特点的加重。综览近期文献，张明岛（1991）曾归纳出以下诸点：（1） α 波频率明显减慢，指

数减少，甚至每秒 8 次以下，其中优势 波慢波化为最重要的特征；(2) 波波幅偏低，以枕叶明显；(3) 出现弥漫性低波幅 波和 波活动；(4) 严重进展性的病例中甚至快波完全消失；(5) 对声光刺激的反应降低或消失；(6) 脑电图结果与痴呆程度有近似的关系。贝克 (Beck, 1991) 认为老年性痴呆脑电图虽尚未概括出特异指标，但脑电图对不同性质痴呆的诊断及鉴别是有临床价值的。

六、生化指标

生化指标 (index of physiological chemistry) 即在情绪状态时，生化系统、中枢神经介质也发生一系列变化。许淑莲等 (1979) 曾以病房配血时病员血浆 11-羟皮质类固醇含量为基础，与入手术室后进行比较，变化值 (以绝对值和变化百分率计算) 和入手术室情绪状态的关系见表 10-5。从表上可见，入手术室后血浆 11-羟皮质类固醇含量较病房配血时普遍有所增加，但镇静组患者增加较少，紧张组患者增加较多，其差异达显著性水平。可见，血浆 11-羟皮质类固醇含量可作为情绪的生化指标之一。

表 10-5 11-羟皮质类固醇变化和情绪状态的关系

情绪状态	例数	皮质激素变化值 (平均数 ± 标准误) 微克/100 毫升	变化百分率 ($\frac{\text{入室后含量}}{\text{配血时含量}} \times 100\%$)	与镇静组比较	
				t	p
镇静	19	+5.5 ± 1.5	153.3 %	-	-
一般	12	+10.9 ± 2.1	215.7 %	2.3739	< 0.05
紧张	16	+10.3 ± 1.3	209 %	2.0896	< 0.05

(采自许淑莲等, 1979)

多数研究表明，在情绪紧张状态下，体内神经化学物质的分泌量或排出量的变化可作为情绪研究的客观指标。例如，汤慈美等 (1985) 对运动员比赛和训练前后尿内肾上腺素和去甲肾上腺素的排出量作为指标，并以所得部分分析结果来说明神经化学物质测定和情绪的关系，从而认定神经生物化学测定，可作为情绪研究的方法之一。由于体力活动和情绪紧张均可引起肾上腺素和去甲肾上腺素分泌量的增加，因此，尿内儿茶酚胺的排出量可以作为测定的客观指标。研究结果表明，男子组与女子组虽稍有差别，但均显示比赛后肾上腺素和去甲肾上腺素的排出量较比赛前有显著的增长。同时在比较了比赛后和训练后的测定后发现，在比赛和训练处于同等体力活动量的情况下，肾上腺素和去甲肾上腺素的排出量也有不同。结果为，训练后比比赛前所测定的排出量虽有增长，但与比赛后相比，排出量要小。这似乎证明肾上腺素和去甲肾上腺素排出量的增加，可以作为情绪紧张的客观指标。参见图 10-7。

七、指标的综合使用

以上虽然介绍了不少生理生化指标，但远远没有穷尽。例如重要的生理

生化指标还有：眨眼、瞳孔反应、皮肤温度、血糖、血液的化学成分（如血氧含量）、外部腺体（泪腺、汗腺）、内分泌功能（如胰岛素、抗利尿激素）、肌电、肌肉（如表情肌）等等。心理学家和生理学家在长期的实践中发现，所有这些指标都可以不同程度地反映出有机体的唤醒水平和活动情况，但是单一的指标在使用时存在着很大缺陷。

许淑莲等（1979）在研究针刺麻醉原理时，对皮电、血压等生理指标进行了研究，结果见表 10-6。从表中我们可以看到，血压、脉搏率、自发皮电三者与情绪状态的关系均不显著；而脉幅波动和呼吸波动则与情绪状态有显著关系，即镇静者呼吸比较均匀，脉搏比较平稳。如果将脉幅波动、呼吸波动和自发皮电三项综合指标与情绪状态相比较，则两者有极显著的关系（ $p < 0.001$ ）。这说明在测量情绪状态时，采用多项生理指标综合使用具有很大优点。

由于对生理指标的综合使用的需要，形形色色的多项生理记录仪应运而生。典型的多项生理记录仪（polygraph）一般包括心率、血压、呼吸和皮肤电的测定（见图 10-8）。

多项生理指标的综合使用也给心理学应用带来广阔前景。多项生理记录

表 10-6 临床观察情绪状态和生理波动的关系

生理指标	病例人数	波动人数	和情绪状态关系	
			t	p
1. 血压	64	35	4.353	$p > 0.05$
2. 脉搏率	66	22	5.487	$p > 0.05$
3. 自发皮电	49	25	6.948	$p > 0.05$
4. 脉幅波动	50	26	9.890	$p > 0.05$
5. 呼吸波动	51	35	16.508	$p > 0.01$
6. 皮电、脉搏、呼吸综合	51	35	23.068	$p > 0.001$

（采自许淑莲等，1979）

仪的另一个名称叫测谎仪，正说明了这一点。心理学上测谎的基本方法之一是联想实验。刺激词一部分与案情有关，称为关键词；另一部分则与案情无关。测量的指标是：（1）对关键词的有意义或不正常反应；（2）对关键词的反应时间延长。这种联想实验方法的缺点是比较繁琐，同时也不十分可靠。因此，实验心理学家试图在其中加进各种情绪指标。他们的假设是：任何说谎的企图或实施都会提高其唤醒水平；为掩盖谎言，个体通常需要较多的智力努力，这也会在情绪指标上有所反映。测谎常用的生理指标是皮电反应、呼吸、血压和手指运动。这些指标同说谎时的情绪变化有较大的联系。但需说明的一点是，测谎仪的结果不能作为法庭上的证据。当然，作为破案的辅助办法是无可非议的。

第三节 表情研究方法

情绪是一种多成分、多维量、多种类和多水平整合的复杂心理过程。这一事实对情绪研究方法学提出了严重的挑战。在过去，行为主义把情绪仅看作为反应这个单一变量，条件反射和操作条件反射的研究方法相应出现。内省学派则把情绪归结为主观体验（subjective experience）这个变量，根据口

头报告的自陈法被广泛采用。当詹姆士把情绪确定为内脏反馈时，生理测量就成为情绪研究的唯一方法。承认情绪可分化为多种各不相同的具体情绪，一般来说，现在已不成问题，许多学者正在试图把情绪进行分类。分化情绪理论把情绪的分化和进化放置在表情的分化和进化的基础上，这就为研究具体情绪找到了一个可靠的方法。它就是面部表情测量。

一、表情研究的理论根据

半个世纪来，对面部表情测量方法及其理论根据的专门研究已经使表情测量成为符合客观化原则的方法。我国心理学家孟昭兰（1987）将理论根据归纳为如下三点：

1. 面部表情是传递具体信息的外显行为 面部表情是提供人们在感情上互相了解的鲜明标记。情绪过程既包括情绪体验，也包括情绪表现，而表情既是情绪的外部表现，也是情绪体验的发生机制；既是最敏锐的情绪发生器，也是最有效的情绪显示器。这就从机制上说明了以面部肌肉运动模式作为情绪标志的根据。

2. 面部表情具有全人类共同性 盖达塞克（Gajdusek，1957）研究发现，前文化民族的面部表情与文明社会民族的面部表情十分相似。从所拍摄非洲新几内亚前文化民族影片中的表情，都是曾在文化民族中见过的，参照作出这些表情的人们的社会行为，能够了解这些表情的含义亦与文化人无异。诚如艾克曼（Ekman，1976，1978，1984）进行了更多的研究后指出，外国人的表情不同于外国语，不经学习就可作为有效的信息在人们之间交流。

3. 新生婴儿具有不经学习所显露的基本情绪的面部表情模式 在前言阶段，新生婴儿基本情绪的面部表情模式是婴儿表达愿望与要求、传递生理和心理信息的唯一手段，也是建立母—婴之间感情联结、保证婴儿健康成长的关键。婴儿时期基本情绪的面部表情是随着生理成熟而实现的先天预成模式的显露。在到达成人时期，一般在不受社会习俗或文化影响的场合下，与婴儿相似的人类的普遍表情模式仍然存在。

在表情研究的理论根据的指引下，一些具体的研究方法就应运而生。面部表情（facial expression）的发生是有其客观的物质基础的：表情按面部不同部位的肌肉运动而模式化，面部反应模式携带着心理学的意义，那就是或快乐、或悲伤等具体情绪。但是，对表情进行测量的原则在于：所要测量的是面孔各部位的肌肉运动本身，而不是面部所给予观察者的情绪信息。因此，测量的最终目的是揭示其显露的情绪信息，而测量本身必须是产生情绪信息的物质过程。为此，测量方法必须严格遵循解剖学基础。整个情绪表情产生的过程，要受严格的解剖生理过程所支配。它包括精细分化的肌肉组织系统的活动和灵敏传导的躯体神经系统的活动。在测量中，使用肌肉刺激，把面孔上一块块肌肉的单一活动和互相有牵连的肌肉的组织活动通过肌电记录和照相记录保存下来，从而得到了面部各个分别独立的部位的变化模型和观察到的、由这些肌肉活动所支配的面容。艾克曼（Ekman，1976，1978）的面部肌肉活动编码系统，伊扎德（Izard，1980）的表情辨别整体判断系统等，是目前主要被采用的测量标准。为了阐明这些现代的表情测量方法，我们还得从早期的传统方法说起。

二、面部表情的早期研究

半个多世纪以前，许多研究者就曾试图用量表和模式图来描述人类的复杂情感。最早出现的是武德沃斯的直线量表。

（一）面部表情的直线量表

费勒基 (Feleky, 1922) 发表了武德沃斯经过 100 名被试判断 86 张照片的成绩分配制成的一个单维直线量表。

：喜爱、幸福、快乐；	：惊讶；
：恐惧、痛苦；	：愤怒、决心；
：厌恶；	：蔑视。

这个量表是单维的，所以也叫直线量表 (Linear Scale)。这个量表有一点很令人满意，就是大多数被试者认为一张照片是恐惧，其他被试者最差也会把它看作是邻近的梯级 (step)。例如惊讶或者愤怒，很少人认为是距离很远的喜爱或是厌恶。这个量表对使用其他照片所搜集的数据的整理起了量表的作用。

（二）面部表情的圆形量表

武德沃斯的助手施洛斯贝格 (Schlosberg, 1952) 在直线量表的基础上提出了面部表情的圆形量表 (Circular Scale) (见图 10-9)。它有两个轴，主轴是愉快和不愉快，由梯级 9(喜爱、幸福、快乐) 到梯级 1(愤怒、决心)，它通常被认为是情绪的基本因素。另一个轴称为注意——厌弃。注意可以惊讶为例，惊讶的时候，双眼、鼻孔、有时包括口部都是张开的，好像准备接受刺激；相反的一端厌弃以厌恶和蔑视为例，这时，双眼、鼻孔和嘴唇都是紧闭的，似乎要把刺激拒之门外。两个轴的交叉点是处于两个极端的中间状态，即介于厌弃和注意之间的表情。

有了圆形量表的坐标轴，就可以利用它们来决定某一表情照片在圆面上的方位。令被试用一个九点量表来评定某张照片，量表纵轴上的九个点代表愉快——不愉快的九个梯级；然后用一类似量表再在注意——厌弃轴上进行评定，这样，我们就可以用每张照片在愉快——不愉快及注意——厌弃两维度上的平均评定来描述它。如某张照片的评定结果是，它在愉快——不愉快轴上的数值是 7，在注意——厌弃轴上数值也是 7，这样，它在量表中的位置就是 (7.7)。从两轴交点 (5.5) 向点 (7.7) 作一射线伸出边外，得到一交点，找出交点的度数，除以 60 后就得到预测的圆形量表数值 (1.75)。这张照片可以叫作“愉快的惊讶”。类似于“东的北”在坐标轴上的位置。

（三）面部表情的三维模式图

在上述研究的基础上，施洛斯贝格 (Schlosberg, 1954) 提出，面部表情可分为三个维度：(1) 愉快——不愉快维度；(2) 注意——拒绝维度；(3) 睡眠——紧张维度 (见图 10-10)。他认为，通过这三个维度就能够把各种面部表情合情合理地区分开来。面部表情的三维模式图 (three -

dimensional pattern) 是在圆形量表的基础上增加了一个睡眠——紧张维度，紧张是紧张性水平和兴奋的标志；另一端睡眠也是极端情绪所引起的表情。

(四) 表情判断的早期方法

德国解剖学家皮德里特 (Piderit, 1859)，在达尔文发表《人类和动物的表情》(Darwin, 1872) 之前就发表了关于面部表情的著作，他提出：全部表情可分解成脸面几个部分的元素性的表情。他相信颜面肌是感觉器官的附属物，有帮助和阻碍接受刺激物的作用。正像眼的睁闭可以辅助或是阻止接受视觉刺激一样，鼻子的某些位置可以帮助嗅气味，而另一些位置则可以使难闻的气味不致侵入。波林和铁钦纳 (Boring & Titchener, 1923) 根据皮德里特的图形设计了能相互调换的脸面部分，用成批不同的眉毛、眼睛、鼻子和嘴拼成了 360 个综合象。奇怪的是他们发现了几乎每一个综合象，即使在部分上有矛盾，许多被试仍认为是一种真正的表情。当被试体会了这个实验的精神，他在个别脸面部分里见到了皮德里特的元素，从而见到了可以辨认的表情，最后拼成一个完整的面部表情，见图 10-11 和图 10-12。

由于演员不仅有表演才能而且有一定的训练，所以一般典型的表情由演员来完成。但是，表情毕竟是十分复杂的。用模拟表情照片进行实验存在以下几个问题：首先，照片摄取的只是演员认为可以代表某种情绪的表情；其次，同义词问题，例如，在暴怒、愤怒、激怒和愤慨这些词中，由于难于区分，被试较难选定其中哪一个是暴怒；再次，存在着其他计分标准，例如，与另一些被试的判断相一致的标准，可能也是一个合理的计分标准。

日常生活中的表情判断是在一定情景下进行的。人们都有这样的体会，离开了情景或现场环境，常常就很难区分和辨别某些表情，请你看一下图 10-13 和图 10-14，这是二张剪去全部现场情景或部分情景的照片，你能正确判断吗？(答案为本章图 10-20，图 10-21。)

三、面部表情的现代测量技术

上一节我们介绍了表情辨别早期研究所使用的典型方法，这些方法可概括为：由未经训练的被试对情绪的面部表情进行整体判断。对这些研究所获得的资料目前仍存在着争论，争论的核心是这些研究多少带有主观性。

从 70 年代初开始，出现了大量关于面部表情和面部动作编码系统的研究。这些研究可分为两大类，一类专门涉及与情绪活动有直接关系的面部动作，如艾克曼 (Ekman, 1971)、伊扎德 (Izard, 1979, 1980) 等人的工作；另一类则涉及了所有可观察到的面部动作，如艾克曼等人 (Ekman et al., 1978) 的工作。这些研究导致了面部表情测量技术的问世，在表情研究中具有极为重要的意义。与传统的方法相比较，现代面部表情测量技术具有四个突出的优点：(1) 这些技术所测量的是面部肌肉运动本身，而不是面部所给予观察者的信息；(2) 严格遵循神经解剖学原则。所有这些测量技术的共同特征是，它们均以面部肌肉的神经解剖学特点和肌肉活动所造成的面容变化 (appearance change) 作为测量的基础和目标；(3) 较准确而客观，便于使用。按上述原则确定的测量技术将情绪功能与颜面解剖学结合起来，任何

人只要经过有关的训练，就能通过肉眼的观察去识别情绪，而不必使用像肌电图机那样的专门设备；(4) 某些测量技术将表情的反应时和持续时间引进了表情测量，因此，它能够测量面部表情的动态过程。

下面就以面部表情测量技术中应用较广、较有代表性的面部动作编码系统、最大限度辨别面部肌肉运动编码系统和表情辨别整体判断系统为例，对这些方法作一简单介绍。

(一) 面部动作编码系统

面部动作编码系统 (facial action coding system, 简称 FACS)，它是艾克曼等人 (Ekman et al., 1978) 在总结过去对面部表情评定工作的基础上制定出的一个尽最大可能区分面部运动的综合系统，它是迄今为止最为详尽、最为精细的面部运动测量技术，它能够测量和记录所有可观察到的面部行为。

面部动作编码系统在制订过程中详细地研究了面部肌肉运动与面容变化的关系。在实际测量时，它以面容活动为单位，称为活动单位。一个单一的活动单位可以包括一块或几块肌肉组织。由于多数面容变化是几个活动单位叠加发生的，从而又可以把那些可以明显辨认的叠加的活动单位列成复合活动单位。面部动作编码系统共列出了二十八种单一活动单位 (参见表 10 - 7) 和十九种复合活动单位。可根据各个活动单位之间的主导或次要、竞争或对抗的关系，从而规定这种情形下的测量规则和方法。面部动作编码系统的使用手册内容有：单一和复合活动单位列表，针对面容变化的详细描述，以及供对照用的照片和影片以及具体的使用指导等。

表 10 - 7 面部动作编码系统 (FACS) 的单一活动单位

AU 编号	FAC 名称	肌 肉
1	额眉心上抬	额肌，内侧
2	额眉梢上抬	额肌，外侧
4	额眉低垂	眉间降肌，降眉肌，皱眉肌
5	上眼睑上抬	提眼睑肌
6	面颊上抬	眼环肌
7	眼睑紧凑	眼环肌
9	鼻纵起	提唇肌，提鼻肌
10	上眼睑上抬	提唇肌
11	鼻唇褶加深	嘴小肌
12	口角后拉	口角迁缩肌
13	面颊鼓胀	口角上提肌
14	唇颊微凹（酒窝）	
15	唇角下压	口角降肌
16	下唇下压	下唇降肌
17	下巴上抬	上提肌
18	口唇缩拢	上翻唇肌，内翻唇肌
20	口唇前伸	口角收缩肌
22	口唇呈筒形	口环肌
23	口唇紧闭	口环肌
24	口唇压紧	口环肌
25	两唇张开	唇压肌、额提肌放松
26	下唇下垂	咬肌、翼状肌放松
27	口前伸	翼状肌、二腹肌
28	口唇啞吸（吮吸）	口环肌

（采自 Ekman，1976，孟昭兰整理）

表 10 - 8 最大限度辨别面部肌肉运动编码系统（MAX）
面部运动分区记录及编号

编 号	眉	额	鼻根
No.20	上抬、弧状或不变	长横纹或增厚	变窄
No.21	一条眉比另一条眉抬高		
No.22	上抬、聚拢	短横纹	变窄
No.23	内角上抬、内角下呈三角形	眉角上部额中心有皱纹	变窄
No.24	聚拢、眉间呈竖直纹		
No.25	下降、聚拢	眉间呈竖纹或突起	增宽

编号	眼	颊
No. 30	上眼睑与眉之间皮肤拉紧、眼睁大而圆，上眼睑不抬高	上抬
No. 31	眼沟展宽，上眼睑上抬	
No. 32	眉下降使眼变窄	
No. 33	双眼斜视或变窄	
No. 36	向下注视、斜视	
No. 37	紧闭	
No. 38		
No. 39	向下注视，头后倒	上抬
No. 42	鼻梁皱起（可作为 54 和 59B 的附加线索）	

编号	口-唇
No. 50	张大、张凹
No. 51	张大、放松
No. 52	口角后收、微上抬
No. 53	张开、紧张、口角向两侧平展
No. 54	张开、呈矩形
No. 55	张开、紧张
No. 56	口角向下方外拉，下颊将下唇中部上抬
No. 59A = 51/66	张开、放松、舌前伸过齿
No. 58B=54/66	张开、呈矩形、舌前伸过齿
No. 61	上唇向一方上抬
No. 63	下唇下降、前伸
No. 64	下唇内卷
No. 65	口唇缩拢
No. 66	舌前伸、过齿

（采自孟昭兰整理，1987）

（二）两个测量系统

前面讲到的面部动作编码系统只是利用解剖学原理对面部各部位进行测量，还不是对情绪的测量。为了对情绪进行解释，伊扎德等人（Izard et al., 1979, 1980）曾经提出了两个互为补充的测量系统，即最大限度辨别面部肌肉运动编码系统（maximally discriminative facial movement coding system, 简称 MAX）和表情辨别整体判断系统（system for identifying affect expression by holistic judgments, 简称 AFFEX）。最大限度辨别面部肌肉运动编码系统是为保证客观性和精确性的微观分析系统，它以面部肌肉运动为单位，是用以测量区域性的面部肌肉运动的精确图式。表情辨别整体判断系统是保证有效性的客观分析系统，它提供的是关于面部表情模式的总概貌。

最大限度辨别面部肌肉运动编码系统将人的面部划分为（1）额眉—鼻根

区；(2)眼—鼻—颊区和(3)口—唇—下巴区三部分，并包括二十九个相对独立的外貌变化的运动单元。这些单位分别编为号码(见表10-8)，通过对三个部分外貌变化的评分及综合，最大限度辨别面部肌肉运动编码系统可以辨别出兴趣、愉快、惊奇、悲伤、愤怒、厌恶、轻蔑、惧怕和生理不适引起的痛苦等多种基本情绪。最大限度辨别面部肌肉运动编码系统的具体使用分为二步。第一步，评分者三次观看面部表情的录像(影)，每次辨认面部一个部位的肌肉运动，并记下相当区域的面容变化及出现时间。例如表10-8中的25号为额眉区的双眉下压、聚拢；33号为眼鼻区的眼变窄；54号为口唇区的口张大呈矩形。第二步，将记录下来面容变化同可观察到的活动单位的组织相对照，辨别出独立的情绪或几种情绪的组合。例如，这三个区域的肌肉活动组合起来，就表示了愤怒的表情(见图10-15)。

最大限度辨别面部肌肉运动编码系统的材料包括一个手册和一套录像。手册包括面部肌肉的详细分类、肌肉组织的位置分布、肌肉活动编号列表和详细描述以及练习使用最大限度辨别面部肌肉运动编码系统的方法、步骤及达到学会标准的要求。之后在学会使用最大限度辨别面部肌肉运动编码系统的基础上，使用者可以进一步学习表情辨别整体判断系统的使用。表情辨别整体判断系统是以最大限度辨别面部肌肉运动编码系统为基础，组合面部运动，从整体上描述基本情绪。使用者根据最大限度辨别面部肌肉运动编码系统，整合面部不同部分的信息，从整体上直接判断面部表情类别。此二系统都已证明具有较高的一致性和可靠性。

第四节 主观体验测量方法

主观体验测量方法一般要求被试报告其直接感受到的经验，其测验方法是用标准化的量表来测量被试者的情绪体验。主观体验测量方法和上述生理指标方法和表情研究方法组成了一个方法体系，它们从不同的侧面揭示情绪的机理。

那么，为什么主观测量方法在情绪研究中占有重要地位呢？主要有二点理由。

第一，出于理论上的考虑。这一考虑立足于一定的理论出发点之上。心理现象或情绪现象都是脑的功能，是客观存在着的一种现象。客观存在的东西总是可以被认识的。从这一认识出发，建立了各种以脑的活动和外显行为为指标的研究方法，包括近年来发展的模拟法在内。尽管这些方法还不尽完善，但已在探索的道路上有所进展。因此从理论上的或方法论上的考虑促使人们探索建立情绪体验的研究方法。

第二，出于应用上的考虑。以情绪而论，无论是生理测量或表情测量，均需采用复杂的仪器设备。而且，利用仪器测量情绪现象或心理现象，一般来说，不是一次测量就能得出所需要的结果。因此，这类技术和方法更适用于研究的目的，而对应用是很不方便的。因而在生活、教育，特别是临床诊断方面，要求尽可能以简便方法去进行测量。例如，对情绪性精神障碍(诸如躁狂抑郁症、焦虑症等)的诊断或治疗，需要提供可为医学用的，甚至为个人自我评定所能采用的测量方法。这样的考虑也促使人们尝试探索建立情绪体验测查的有效方法。

在本节中，我们将介绍一些具体测量方法。主要有形容词检表、维量等

级量表和分化情绪量表、应用性情绪量表等。

一、形容词检表

(一) 心境形容词检表

自 50、60 年代以来，为了解成人的情绪体验，建立了一种简便、通用的技术，称之为形容词检表 (Adjective Check List, 简称 ACL)。用这类检表可测查被试即时存在和体验着的情绪。这种方法被称为主观体验自我报告测量 (Subjective Experience Self Report Measure)，用于检测心境和用于临床诊断。

心境 (mood) 是一种比较微弱、持久具有渲染性的情绪。换句话说，可把心境确定为：延续一定时期 (几天或更长些) 的淡薄而朦胧的情绪状态。它在意识里存在，但并不一定经常在语词意识中出现。它可以是正性的，如沉静而有活力的状态；也可以是负性的，如郁闷而怠倦的状态。心境似乎为行为抹上一层色彩，这层色彩不论是明亮的或阴暗的，都能为心理操作提供一种特定的背景。诺利斯 (Nowlis, 1956) 早在 50 年代就曾提出，心境可被考虑为一种中间介入变量，它影响其他心理活动的发生和可靠性。因此心境似乎担任着一种自我监测的角色，提示自身的心理效应和状态。心境无论

表 10-9 一个简短的心境形容词检表

发怒的	懊悔的	过度高兴的
被占据的	犹豫不决的	沉思的
不经心的	夸口的	懒散的
得意的	积极的	和气的
集中的	对抗的	悲伤的
困倦的	害怕的	忧愁的
充满感情的	玩笑的	自我中心的
精力旺盛的	愉快的	抱歉的
反抗的	专心的	怀疑的
紧张不安的	拘束的	集中于己的
机智的	热情的	有魄力的

是正性的还是负性的，都可以是正常的；然而在某种情况下，它可以成为病态的。那就是人们所熟知的情绪性精神障碍了。形容词检表正是用于测查心境或情绪性障碍。它既可为医生、教师所使用，也可由被试个人用来作自我评定。

形容词检表是选用一系列描述情绪的形容词，如镇静的、神经质的、害怕的、忧郁的等等并列为检表。被试者通过内省，从检表中选出符合自身当时情绪状态的词汇用来确认自身的情绪体验。各种检表选用的形容词数目不同，有多达 300 个词汇的，也有少至几十个或十几个的，其数量视检测目的的多寡而定。各种检表的测查内容也有所不同，有的用来测查一般心境，另一些则用来测定专门项目，如测量应激 (或压力) 或焦虑等特定情绪。表 10-9 是诺利斯在他的拥有多达 200 个词汇、并被命名为心境形容词检表 (Mood Adjective Check List, 简称 MACL) 的基础上抽取了 33 个词汇而

构成的简化的检表。表 10 - 10 则是伯肯 (Berkun, 1967) 等仅用 14 个词汇, 以等距法编制的主观性应激量表 (Subjective Stress Scale)。

形容词是检表的基础。由于形容词术语数目繁多, 如何归类是一个需要解决的问题。有些学者采用因素分析的方法, 借鉴情绪维量的观点, 进行了归纳的工作。这为人们使用量表提供了极大的方便。诺利斯通过 400 名大学生用检表的 100 个词汇进行的自我评定, 并通过因素分析, 总结出 8 种因素。

诺利斯的工作之后, 又不断有人对不同数目的检表进行因素分析, 得出的因素有的是 6 个, 有的是 9 个或 12 个。鉴于结果如此不一致, 且情况同人格特质研究出现的问题相类似, 从而认为采用循环丛 (circumplex) 的办法来描述心境可能更为适宜。劳尔等人 (Lorr et al., 1967) 采用四点记分方法, 通过对 339 名大学生的评定, 对 52 个词汇进行了相关和因素分析, 设计出心境八因素环形丛, 参见表 10-11。

(二) 情绪-心境测量量表

普卢特奇克 (Plutchik, 1969) 以他的情绪维度理论为基础, 利用上述形容词检表测量方法, 编制一个情绪-心境测量量表 (Emotion-Mood Measurement Scale)。情绪-心境调查标准中的前八项相当于在他的情绪三维模式中列入的八种基本情绪。把这八种基本情绪扩充为相似类别的情绪群, 并用相当的形容词术语标出。另外, 在这个检表中加入了标示激活量, 作为第九项, 参见表 10-12。

普卢特奇克后期研究工作都致力于临床应用。“情绪-心境测量量表”曾用于临床。在被诊断为躁狂抑郁症的患者中, 对 26 名处于躁狂状态、42 名处于抑郁状态、75 名暂时处于正常状态的患者, 进行了逐项测查。结果发现, 在合作、再生、再整合三项的每一词汇比较中, 躁狂同抑郁患者之间均有显著性差异。其他项的许多词汇在这两种患者之间也有显著性差异。此外, 在三种患者的两两比较中, 也在许多项目的词汇上显示出差异。

正是由于临床实践的需要, 此后, 形容词检表编制的发展越来越趋向简化。在经过标准化检验之后, 有的检表只留用少量词汇, 在同他种量表比较中就能达到相当高的相关。例如卢宾 (Lubin, 1966) 的几种忧郁检表, 从 34 个词汇减少为 17 个, 这个量表同明尼苏达多项人格测验 (Minnesota Multiphasic Personality Inventory, 简称 MMPI) 中忧郁量表之间有显著性相关。

普卢特奇克对八种基本情绪只以一个单词来代表, 制订了一个五点记分量表 (见表 10-13)。他把这个量表用于一项大学生心境测查: 在 40 名学生中按量表进行自我评定。在一周内的星期一、三、五各天的课堂内进行, 作为控制组。在此之后, 于考试前、考试后各一天重复评定。在考试前的评定中, 惧怕与兴趣评分显著提高, 愉快和接受两项记分明显下降。在学生拿回考试试卷和分数之后, 再次的自我评定表明, 愉快、接受、兴趣分数明显下降; 相反愤怒、厌恶、悲伤和惊奇分数明显提高。这类单词汇心境量表在一个很广的感情状态范围内测量应激状态十分灵敏, 而且具有理论上的相关。

经过几年的实践, 普卢特奇克不但充实了这测查表, 而且把测查情绪和

人格特质结合起来，编制了情绪人格测查（Emotion Personality Inventory 简称 EPI）（1974）。这份检表按八种基本情绪提出一套数目很大的情绪人格特质，并通过因素分析筛选出如下 12 种特质：

冒险的	郁闷的	怨恨的
富于感情的	冲动的	自我意识的
沉思的	服从的	害羞的
好争吵的	谨慎的	社交性的

在测查表上为上述每个特质作一简要注解。按照这 12 种特质，配对组合为 62 对特质词汇，这样就成了最后的情绪人格测查。情绪人格测查是一种强迫选择测验。即被测试者必须从每一对词中选择一个作为符合本身情况的。例如，从“好争吵的——害羞的”这一对词中只能选择其中之一。情绪人格测查的 62 对特质包含着一个按照八种基本情绪编制的情绪系统记录表，也就是从 62 对特质中归纳出八种情绪的系统。被试填写的是 62 对词汇表，但用来记分的是情绪表。这样既测查了人格情绪特质，又测查了情绪。而且，由于被试并不知道这 12 种特质中包含的情绪系统是什么，这样就避免了主观性。情绪人格测查还以八种基本情绪建立了分别以功能语言、特质语言、主观语言和为行为语言来描述的对应术语（见表 10-14），这样使用起来就方便得多。

二、维量等级量表和分化情绪量表

在第五章讲到斯宾塞提出了心理连续量的概念。如愉快——不愉快是一个连续的变化量。冯特的情感三维理论（或感觉三元论）（three-dimensional theory of feeling）具体地在情绪三维（快乐——厌恶、紧张——轻松、兴奋——抑郁）中提示了连续变化量的模式。伊扎德在情绪体验主观评定方法的制订中，运用连续变化量的概念，编制了两个对应使用的量表：一个是测量各情绪维量的等级量表，另一个是测量各情绪成分的分化情绪等级量表。下面分别给予讨论。

（一）维量等级量表

维量等级量表（Dimensional Rating Scale，简称 DRS）是一个四维量表。伊扎德最初提出的八种维量是从众多的对情绪情境作自我评估的数据中得出的，后经筛选，确定了维量等级量表的四维是：愉快度、紧张度、冲动度和确信度。筛选掉的四个维是：活跃度、精细度、可控度和外向度。作为维量等级量表的四个维度，伊扎德作了如下解释：

1. **愉快维** 这表示评估主观体验最突出的享乐色调方面。
2. **紧张维** 这是表示情绪的神经生理激活水平方面的，它包括肌肉紧张和动作抑制等诸成分水平。
3. **冲动维** 此维量涉及对情绪情境的出现的突然性，以致个体缺少准备的程度。
4. **确信维** 此维量表达个体胜任、承受感情的程度。在认知水平上，个体能报告出对情绪的理解程度；在行动水平上，能报告出自身动作对情境适宜的程度。

维量等级量表假定为量表应包括情绪体验、认知和行为三方面。因此，维量等级量表实际上包括着三个分量表，每个分量表由四个维量所组成。每个维量作五级记分，例如：

你感受到愉快吗？——情绪体验分量表愉快维；

你认识到自己的紧张吗？——认知分量表紧张维；

你有冲动的行为表现吗？——行为分量表冲动维。

在使用维量等级量表时，每种情绪都可得到其维量的等级分析。也就是说，在 12 个维量上分别选择五个等级。

（二）分化情绪量表

分化情绪量表 (Differential Emotions Scale, 简称 DES) 和维量等级量表一样也是以形容词检表为基础所建立的一种量表。它是当个体处在特定情绪情境下时，用来测量个体情绪中的分化成分。分化情绪量表包括十种基本情绪，每种情绪有三个描述它的形容词，共 30 个形容词。分化情绪量表被发展为用来测量两种情绪情况：一种为测量情绪强度，作五级记分，称 DES₁；第二种为测量情绪出现的频率。因此它可用来测量心境或情绪特质，也作五级记分，称 DES₂。

分化情绪量表要求被试描述（想象或回忆）某一情绪发生的具体情境，填写分化情绪量表和维量等级量表两个量表。按维量等级量表填写某种具体情绪的四维量强度，得出五级强度分数。按照分化情绪量表填写具体情绪的形容词检表，得出各种情绪成分的五级强度记分或频率记分。两个量表同时使用。例如，让被试想象某一引起惊奇情绪的情境，用维量等级量表测量其惊奇情绪的维量。图 10-16 显示了这一结果，表明惊奇情绪的愉快维得最高分数，自信承受维和冲动维的分数也很高，而只有紧张维较低。用分化情绪量表测量这一情景产生的情绪组成成分，表明了极高的惊奇、快乐和兴趣的分数，和极低的惧怕和害羞的分数（见图 10-17）。然而，在想象某一引起羞怯情绪情境时，维量等级量表测量得到的结果恰恰与惊奇情况相反，这时只有紧张维得了高分，其余三维量均得低分数（见图 10-18）。分化情绪量表测量中，羞怯、惧怕、兴趣、痛苦等组成成分从高分依次下降，而愉快、惊奇、愤怒等情绪得低分（见图 10-19）。

分化情绪量表和维量等级量表测量得出的标准图形达到一定的信度和效度，在表明情绪体验的性质和测量方法上具有理论和应用的意义。

三、应用性情绪量表

情绪和认知、情绪和操作均有着极为密切的关系，情绪对认知、操作的影响，已为人们所公认。这里，我们看到了情绪研究的实用价值。为此，心理学家们编制了形形色色的量表，以便在了解的基础上，最终能达到认知和操作水平的提高。限于篇幅，我们仅介绍一种应用性情绪量表。

运动竞赛状态焦虑量表 (Competitive State Anxiety Inventory-2), 简称 CSAI-2 问卷 (CSAI-2 questionnaire), 是由美国伊利诺斯大学的马腾斯等人 (Martens et al., 1990) 以多维的竞赛状态焦虑理论为指导，而编制成的一种对运动员具有特殊测定价值的状态焦虑问卷。在编制此问卷之

前，马腾斯等人已用了六年时间在交互作用理论的指导下编制出了专门用于运动情境的竞赛特质焦虑表(Sport Competition Anxiety Text,简称 SCAT)。这一量表已由祝蓓里(1993)修订出了中国常模。国内外的许多研究也已经证明，马腾斯等人(1977)编制的竞赛特质焦虑表比美国佛罗里达州大学斯比尔伯格等人(Spielberger et al., 1970)编制的状态-特质焦虑问卷(State-trait Anxiety Inventory,简称 STAI)更能预测运动成绩。

马腾斯等人(Martens et al., 1990)编制运动竞赛状态焦虑量表是通过系统的心理测量过程而编制成的一种专门测定运动竞赛情况下的多维状态焦虑问卷表。在最初的 A 形式中包含有四个分量表：(1) 认知焦虑；(2) 躯体状态焦虑；(3) 怕身体受伤；(4) 一般性焦虑。通过专家的评定，评定和分析了此量表的内容效度。为选定测题，又分析了各题目与其分量表间的相关系数(要求达到 0.50)。此外，还对测题进行了因素分析和分辨力的分析(要求至少有 0.5 分辨率和有最大的权重相关)。通过对 A 形式问卷的分析，主要发现了状态自信心这个分量表。于是把 A 形式修改为 B 形式。后几经修改，将 B 形式修改成 C 形式又修改成 D 形式。最后，E 形式就是被人们所接受的最后确定下来的 CSAI-2 问卷。此量表包括了在 D 形式问卷中的三个分量表，共 27 道测题。刊头有指导语，表 10-15 是祝蓓里修订后的问卷。

至此，我们对情绪的研究方法作了介绍。侧重于方法学上的介绍，正是本书的撰写宗旨。有关情绪和认知的关系等等，这里就不一一列述了。

本章实验

一、广告悦目测定

(一) 目的：学习对偶比较法和等级排列法，判断哪种广告较美，分析其原因，并为实际广告部门提供资料。

(二) 材料：选定和制作 10 张典型广告幻灯片。

(三) 方法和程序 使用对偶比较法和等级排列法来测定哪张广告较美。

1. 对偶比较法。此法最早由寇恩(Cohn, 1894)在他的颜色爱好的研究中介绍出来。这个方法是把所有要比较的刺激配成对，然后一对一对地呈现，直到所有的广告都被比较过为止。如果每一个广告和另外的广告逐一配对；则配出对的数目是 $n(n-1)/2$ ，10 个广告则为 45 对。为尽量排除空间误差，在同一个广告组成的配对中，该广告在左和右面的位置各半。被试者的任务就是在两个广告中进行选择，确定哪个广告较美。

2. 等级排列法。这个方法是 10 个广告同时呈现，让被试者按他的判断标准，把 10 个广告排列成一个顺序。被试者可以反复比较，只要最后排出一个美丑的等级顺序就可以了。然后把许多人对同一广告评判的等级加以平均。

3. 在记录上述结果的同时，务必详细记录被试评定广告美 丑的理由，以供分析时参考。

(四) 结果：整理结果填入记录纸(表 10-16，表 10-17，表 10-18)

(五) 讨论：

1. 对本实验所得量表进行分析讨论，并提出心理学依据。
2. 在本实验中，为什么要随机改变左、右的位置，为什么对每对广告要有间隔？

3. 对偶比较法和等级排列法用于实验中还存在什么问题？

表 10-16 对偶比较法记录纸

A	A	B	C	D	E	F	G	H	I	J
B										
C										
D										
E										
F										
G										
H										
I										
J										
选择分数 (C)	第一轮									
	第二轮									
	总计									
$P = c\sqrt{2(n-1)}$										
$C = C + 1$										
$P = C \sqrt{2n}$										
Z										
Z										
顺序										

表 10-17 等级排列法

广告幻灯片	A	B	C	D	E	F	G	H	I	J
评判名次										

二、情绪对动作稳定的影响

(一) 目的：学习使用测定动作稳定的仪器，测量简单动作的稳定性，检验情绪对动作稳定的影响。

(二) 材料：九洞动作稳定器，计数器，记时器

(三) 程序：

1. 测定动作稳定的程度

(1) 先将动作稳定器、计数器、记时器等串联成一电路。被试坐时，面向动作稳定器，使仪器的边缘与桌边齐平，并与被试的右肩相对。然后被试手拿电笔，手臂悬空，电笔与动作稳定器表面垂直。

(2) 当主试发出“预备”口令时，被试用电笔尖端插入最大的洞孔内，深度为1~2毫米。然后主试发出“开始”口令，同时记录时间。被试用电笔插入最大洞中到15毫米处即取出，做到不碰洞边才可继续做下一个洞；如果一次碰边可做第二次，假如第三次又碰边，就不再往下做。本实验的指标是，被试能通过的最小洞的直径。

(3) 左手与右手各试验10次，在每次左右手轮换时，可休息一分钟。

2. 测定比赛时紧张情绪对动作稳定的影响

两组进行比赛，共分成四对，每对比三次，三战二胜。比赛进行时主试们要在旁边分别报告进行情况，造成竞赛时的紧张气氛。

(四) 结果：

1. 把测量动作稳定性的成绩填入表10-19，并计算其平均数和标准差。
2. 比较右手与左手的动作稳定性。
3. 比较各被试在安静时和在比赛动作稳定的情况。

(五) 讨论：本实验在研究体育教学和训练问题中的意义。

本章摘要

1. 传统心理学研究情绪的基本方法可概括为两种：印象法和表现法。

2. 本世纪上半世纪，在心理学上对情绪研究最有影响的理论有二个：行为学习理论和心理分析理论。

3. 伊扎德总结了过去情绪研究中的问题，提出了多维度和多成分性。这就决定了在研究方法上不能把它规定为单一变量的现实。情绪研究的变量可以从认知变量、行为变量和生理变量等三方面考虑。

4. 情绪有内心体验，也有外部行为表现，同时也有其生理机制。由于自主神经系统的活动，当有机体处于某种情绪状态时，其内部会发生一系列的生理变化，测量这些变化的指标就是生理指标。

5. 皮肤电反应又称心理电反射，它是古老的、常用的和较简易的情绪生理指标。影响皮肤电基础水平的主要因素有三个，即觉醒水平，温度和活动。

6. 由自主神经系统控制的循环系统活动与情绪状态有关。因此，循环系统活动的指标也可以成为情绪测量指标。经常用到的指标是脉搏，血管容积和血压。

7. 在情绪状态不稳定时，呼吸系统的活动会有所变化：或加速或减慢，或变浅或变深，对剧痛的情绪反应往往使呼吸加深加快；突然惊恐时，呼吸会临时中断；狂喜或悲痛时，会有呼吸痉挛现象发生。

8. 孟昭兰将表情研究的理论根据归纳为三条：(1) 面部表情是传递具体情绪信息的外显行为，是提供人们在感情上互相了解的鲜明标记；(2) 面部表情具有全人类普遍性；(3) 新生婴儿具有不经学习所显露的基本情绪的面部表情模式。

9. 与传统的方法相比较，现代面部表情测量技术具有三个突出的优点：(1) 这些技术所测量的是面部肌肉运动本身，而不是面部所给予观察者的信息；(2) 严格遵循神经解剖学原则；(3) 较准确而客观，便于使用。

10. 主观体验测量方法，一般要求被试报告其直接感受到的经验，其测量方法是用标准化的量表来测量被试者的情绪体验。主观体验测量方法和生理指标及表情研究方法组成了一个方法体系，它们从不同侧面指示情绪的机理。

11. 形容词检表是选用一系列描述情绪的形容词，将其列为检表。被试者通过内省，从检表中选出符合自身当时情绪状态的词汇来确认自身的情绪体验。

12. 分化情绪量表和维量等级量表，均是以形容词检表为基础所建立的一种量表，它是当个体处在特定情绪情境下，用来测量个体情绪中的分化成分或维量等级。

建议参考资料

1. 张燕云（译）（1986）：情绪心理学。沈阳市：辽宁人民出版社。
2. 孟昭兰（1987）：为什么面部表情可以作为情绪研究的客观指标。心理学报，19卷，二期。
3. 孟昭兰（1989）：人类情绪。上海市：上海人民出版社。
4. 曹日昌等（译）（1965）：实验心理学。北京市：科学出版社。
5. Izard, C. (1977). Human emotions. New York: Plenum Press.
6. Izard, C. (1979). The maximally discrimination facial movement codingsystem. Newark: Instructional Resources Center.
7. Izard, C. (1980). A system for identifying affect expression by holistic judgments. Newark: Instructional Resources Center.
8. Pluthik, R. (1980). Emotion: A psychoevolutionary synthesis. New York: Harper.

第十一章

心理实验常用仪器

本章内容细目

第一节 仪器在心理实验中的作用	
一、心理实验仪器的诞生和发展	577
二、仪器对心理实验的作用	578
三、目前心理实验仪器发展的趋向	580
第二节 感觉类实验仪器	
一、色轮	582
二、听力计	583
三、声级计	584
四、语图仪	584
五、两点阈测量器	585
六、测痛仪	586
(一) 热辐射仪	
(二) 钾离子测痛仪	
(三) 弹簧棒	
(四) 止血带致痛仪	
七、动觉计	588
第三节 知觉类实验仪器	
一、眼动仪	590
二、实体镜	592
三、长度和面积估计器	592
四、动景器	593
五、似动现象研究装置	594
六、深度知觉仪	595
七、大小常性测量器	596
第四节 记忆类实验仪器	
一、注意分配仪	597
二、棒框仪	598
三、迷宫	600
四、速示器	601
五、记忆鼓	602
六、多重选择器	603
第五节 情绪和技能类实验仪器	
一、皮肤电反应仪	604
二、动作稳定器	605
三、反应时测定仪	
(一) 简单反应时测定仪	
(二) 选择反应时测定仪	
四、手指灵活性测验仪	606

五、双手调节器	607
六、镜画	608
第六节 计算机在心理学上的应用	
一、计算机与心理学的关系	609
二、人-计算机系统的研究	610
三、心理活动的计算机模拟	612
(一) 感知模拟	
(二) 记忆模拟	
(三) 思维模拟	
四、计算机在心理实验中的应用	616
(一) 计算机在心理实验的应用领域	
(二) 心理实验运用计算机的优缺点	
本章摘要	
建议参考资料	

在前面十章的讨论中，明显地看到，实验心理学的产生和发展离不开仪器。一般地说，自然科学的发展和仪器是息息相关的。实验心理学的前进和发展也不能违背这条规律。德国是现代心理学的诞生地，是 19 世纪心理实验仪器研究的中心。20 世纪后，美国逐步成为心理学乃至心理实验仪器的研制的大本营。仪器对心理实验研究之所以重要，这是因为它能使我们测量到一些不能直接观察到的事件；它能使我们获得标准化的数据材料；它还能使实验结果自动记录下来。当今受认知心理学的发展和现代科学技术的影响，一个突出的趋向是计算机已广泛被采用。

可见，仪器是心理实验的重要手段和可靠保证。一个高明的实验者，总善于思索选用适当而先进的仪器来达到实验的目的。例如，时间是心理实验中最常用的反应指标，在感知觉、注意、记忆、思维乃至个性实验中都有时间因素的记录，因而选用合适的记时器就至关重要了。为了粗略测量一个继续 10 秒以上的反应，一个有 1/5 秒的准确度的停表就够准确了，这时实验者可以自己来掌握停表。然而为了测量一个小于 1 秒钟的反应时间，此时就不能使用人工停表了，因为人工停表本身就有 1/5 秒的误差。此刻就需要精度到毫秒级的测量和采用自动记录，于是，形形色色用途的计时器就发展起来了。

心理实验仪器的种类繁多，美国印第安那州的拉斐特仪器公司就生产 2000 种心理实验仪器。本章仅介绍心理实验上的常用仪器和基础性仪器。本章内容之讨论，主要希望回答以下五方面的问题：

1. 心理实验仪器的发展经历哪三个时期。
2. 仪器在心理实验研究中的作用。
3. 当今心理实验仪器发展的新趋向。
4. 以一个仪器为实例，说明其在心理实验中的作用。
5. 心理实验运用计算机进行有哪些优缺点。

第一节 仪器在心理实验中的作用

一、心理实验仪器的诞生和发展

心理学是一门实验科学，实验常常离不开仪器。心理实验早期的研究就有使用仪器的记录。例如，惠斯通（Wheastone，1938）曾创造一种立体镜，应用于心理物理实验。心理实验仪器的系统使用的标识是1879年莱比锡大学心理实验室的建立。从那时到现在的一百余年中，心理实验仪器的发展突飞猛进，然而也表现出明显的阶段性。它表明了心理实验仪器的发展，一方面受到心理学科本身的制约和影响，如行为主义和认知主义都给心理实验仪器打上烙印。另一方面心理实验仪器的发展又受到时代科学技术水平的制约和影响，也就是说，机械时代、电子时代和计算机时代都给心理实验仪器带来影响。

在心理学科自身水平和时代技术水平的双重作用下，心理实验仪器的发展大致经历了三个时期。

1. 第一时期 从19世纪70年代到20世纪初是心理实验仪器发展的第一时期。在这段时期，心理学家冯特和詹姆士、皮尔斯等，先后创建了小型心理实验室。由于心理学实验室建设和对心理实验仪器的需要，1886年美国芝加哥成立了世界上第一个生产心理实验仪器的专业公司。由于此时期心理学研究中心在德国，最新水平的仪器是由德国所生产，例如，以利普曼（Lipmann，1908）命名的记忆仪，具有比较完善的传递装置，呈现的刺激信号的窗口，由定时的齿轮控制，达到那时期较精密的程度。

2. 第二时期 自20世纪20年代开始至60、70年代，是心理实验仪器发展的第二个时期。这个时期的特点是行为主义占统治地位，行为主义在美国兴盛了四、五十年之久，与行为主义的发展相适应，心理实验仪器亦得到迅速发展。这时期，美国仪器公司生产的心理实验仪器，远销许多国家和地区。例如，斯托尔汀仪器公司生产的计时装置就被写入了许多实验心理学和普通心理学的教科书中，该公司生产的仪器斯金纳箱，见于众多的动物心理实验室。

3. 第三时期 自20世纪60、70年代开始至现在，认知心理学盛行起来，这是心理实验仪器发展的第三个时期。认知心理学的发展给美国心理学带来了重大变化，这主要是较多地强调对意识过程的研究。和这个时期的特点相适应，一方面仪器的数量和质量逐年提高；另一方面在心理研究中不仅使用仪器而且使用材料的实验增多，例如，美国拉斐特仪器公司不仅生产心理实验仪器，而且供应多种实验和测量材料。这些仪器和材料销售到世界许多国家和地区。

二、仪器对心理实验的作用

心理学研究常常要做实验，仪器就显得特别重要。仪器对心理学研究之所以重要，概括地说，主要原因有：（1）仪器能使我们在已知的和控制的条件下获得标准化的数据和资料；（2）仪器能使信息永久记录下来，以供进一步的分析；（3）仪器能使我们测量到我们感觉器官不能直接观察到的事件，从而极大地扩展了我们的观察范围。

为说明仪器在心理实验研究的作用，杨治良（1984）曾将美国的四种杂志（《实验心理学杂志》、《美国心理学杂志》、《心理学杂志》和《普通（采自杨治良，1984）（采自杨治良，1984）心理学杂志》）上发表的文章，

作统计分析。若将 1960 年以来在这四种杂志上发表的论文分为三类：一类为用仪器的研究，即研究论文中明确提到使用某种仪器；另一类为使用材料的研究，即研究论文中没有明确提到使用仪器，但明确提到使用某种材料；还有一类是仪器和材料均未使用的论文。最后得出图 11-1 和图 11-2 的结果。从这二图上我们可看到下列一些情况：

(1) 使用仪器的总的情况是：四种美国心理学杂志，自 1960 年以来共发表的研究论文 9987 篇，其中明确提到使用仪器的有 4966 篇，占总数的 49.7%。

(2) 从图 11-2 中似乎可见，使用仪器的研究有减少的趋向。

(3) 从图中似乎还可看到，使用材料的研究有增长的趋向。

再从我国最近几年心理学科研究情况来看，实验仪器在其中也占有相当的比例。以 1980~1984 这五年发表在《心理学报》上的文章为例，在总计 335 篇文章中，实验研究占总数的 51%，其中不同程度地使用了实验仪器的研究又占实验研究总数的 59.6%，使用实验材料的研究占实验研究总数的 40.4%（参见图 11-3）。

综上所述，我们可以概括为二点：(1) 仪器在心理学研究中起着巨大作用，国内和国外近乎半数的普通心理学和实验心理学研究使用仪器；(2) 使用材料的研究有增长的趋向，这可能和认知心理学的发展有关，也可能和问卷、测量技术的广泛应用有关。

三、目前心理实验仪器发展的趋向

从以上的介绍中，我们可以说，仪器在心理学研究起着重要的作用。从对美国的心理实验仪器情况的研究（杨治良，1984）观之，今日心理实验仪器的发展有下列一些新的趋向。

1. 不断向标准化努力 标准化（standardization）是实验仪器的核心所在。举一个例子来说，斯莫尔（Small，1900）设计了本世纪首次使用的迷津。随后，不同大小和式样的迷津层出不穷。大家知道，大小和式样对于老鼠的学习很有影响。斯金纳箱（Skinner box）的情况也很类似。今天，拉斐特仪器公司生产的几种斯金纳箱，就是根据标准化原则设计的。使用同类的仪器，各研究者所得的研究结果就可相互比较。

2. 系列化 随着心理学实验的多样化，仪器公司在设计仪器时，尽力使规格大小统一和一机多用，并把许多仪器统一成一个整体或单元，使用起来十分方便。同一个单元的仪器，既能单独使用，又能组合成多种形式。如积木块组成多种图案一样（参见图 11-4）。

3. 应用于教学 美国几家较大的心理实验仪器公司，除了生产精密的研究仪器之外，还生产相当数量的基础实验仪器。这些基础实验仪器更适用于学生实验和教学演示，展示仪器的广泛用途和在教学上的功能。

4. 心理测验材料的增多 不少仪器公司开始成批生产各种心理测验材料。材料有别于仪器，但是材料也有一个标准化的问题，斯托尔汀仪器公司生产的各类心理测验材料多达数百种，涉及从感觉到个性等各个领域，有适用于正常人的，也有适用于异常人的，有适用于儿童的，也有知用于成人的，应用范围之广，都是前所未有的。

5. 电子计算机的普及 电子计算机的广泛采用，也使心理学方法发生了

变化，它对实验设计、数据处理、储存信息等起着十分重要的积极作用。许多由计算机装备的实验室纷纷建立起来，甚至可以说，心理实验室离不开计算机。

从以上的趋势可见，心理实验仪器正向着更高的水平发展。心理实验仪器有较强的特殊性。同其他学科的实验仪器相比，心理学实验仪器有特殊的要求。因此，心理实验仪器的研制更强调有心理学工作者参加，仪器技术人员也应具有心理学的基础。这些年来，海峡两岸一些高等院校的心理学教师和工厂的技术人员协作，已经研制出数量可观的心理实验仪器，在相当程度上满足了心理学教学的需要，也在一定程度上为科研提供了条件。从今后发展看，心理实验仪器的研制和生产，还应重视大型的、复杂的心理实验仪器的研制，以求尽快形成多层次的、较为完整的、具有本土化的心理实验仪器的研制和生产体系。

综上所述，可知心理实验仪器的研究和制造已构成了一个广阔而专门的领域。在心理学研究中，仪器所起的作用是引人注目的。我们可以这样说，心理实验仪器的研究已经成为一门专门的学问，且愈来愈受到人们的重视，并已构成实验心理学的重要组成部分。

第二节 感觉类实验仪器

一、色轮

色轮（或混色轮）（color wheel），它是一种演示混合颜色和其他视觉现象的装置。色轮一般由色盘和动力装置（多为变速马达）二个部分组成（见图 11-5）。使用时用两个或更多的不同颜色的纸盘，镶嵌混合成 360° 整圆。当转速超过闪光融合临界频率时，就可以产生一种平稳的混合色。混合色是色光混合，从色纸反射到眼睛的二（或三）色光实际上并没有真正混合，而是交替地刺激网膜，由于一扇色光引起的网膜兴奋在另一扇色光到来时尚未停止，因此产生融合过程，使观察者看到一均匀色的表面，感觉到色混合的同样效果。色轮除了可作颜色混合实验外，还可作颜色饱和度和明度实验、视觉螺旋后效实验、闪光融合实验、主观色实验和闪光融合实验等。

二、听力计

听力计（audiometer）是测定个体对各种频率感受性大小的仪器，通过与正常听觉相比，就可确定被测者听力损失情况。最早的听力计是由西肖尔（Seashore, 1919）设计的。可让被试从耳机中听到 40 种强度不同的咔嚓声。最大和最小的声音强度之比为 1079 : 1，从 1 到 40，每增加一个刻度，增加的响度都尽量使之相等。正常听觉的响度绝对阈限大约在刻度的中部。这种仪器只能对噪音测定响度的绝对阈限。后来的听力计又发展成测量各种频率纯音的响度绝对阈限。这种听力计可以产生各种频率及各种强度的可听纯音。

听力计种类很多，心理学上的听力计通常是指纯音听力计。使用时，仪器部件（见图 11-6 中央部分）自动提供由弱到强的各种频率刺激，自动变换频率，测听时被试戴上双封闭隔音的耳机，当听到声音时，即按键，仪器可

根据被试反应直接绘出可听度曲线。另外，还可能附有骨传导器等附件。在医学上检查听力用的听力计，其读数就是听力的损失，即低于正常听力水平的程度。将正常的阈限定为 0 分贝，若被试在某个频率上的听力损失为 10 分贝，则说明该被试在该频率上的听阈限比正常值高 10 分贝，用听力

计测量的结果可以在听力图上画成听力曲线。听力测定能评定一个人的听觉灵敏度，因此，它在听力保护工作中是必不可少的仪器。

三、声级计

声级计 (sound level meter) 是一种对声音的物理强度进行测量和分析的仪器，它能对声音作出类似人耳的反应，同时，它还能对声源进行客观而可重复的声级测量。我们知道，声信号可以用高质传声器转换成一个相等的电信号。声级计的计量单位是分贝 (decibel, 简称 dB)，这个分贝的参考点是 0.0002 微巴 (即 0.0002 达因/厘米²)。同时还可利用声级计所附滤波器对所测声源的频率成分进行分析。由于这种信号很微弱，必须经多级放大，才能在电表上指示，电表上读出的是用分贝表示的声级。由于声级计是高度精密的仪器，因此需要经常校准，以保持准确的测量 (见图 11-7)。

四、语图仪

语图仪 (language pictorial display apparatus) 是将复合音或语言分析为组成成分频率，显示频率—强度—时间型式变化的仪器。顾名思义，它是一种能形象地显示言语听觉特征的仪器。我们知道，波形图是将声波幅值的变化作为时间的函数来表现的。语图仪克服了波形图和声谱图二种方法的不足之处，以三维表示出声波的幅值作为时间和频率两者的函数关系。语图也叫“可见语言”，从声学所录制的语图上，可以看到频率、强度、四声、共振峰、时间等方面的不同。这就使人们有可能据以探索语音的哪些因素影响着言语的清晰度，并有助于设计通讯系统，以提高清晰程度，训练有关专业人员，改正言语质量。

语图仪是比较复杂的仪器。如用横坐标表示时间，纵坐标表示频率，高频在上端，低频在下端，感光纸上黑暗程度表示振幅大小，这样就把时间、频率与振幅合并为一个记录。

五、两点阈测量器

两点阈测量器 (或两点阈量规) (two-point aesthesiometer) 是测定人体上各部位两点阈值的仪器。两点阈测量器有多种形式，有的近似圆规，参见图 11-8，使用两点阈测量器，必须垂直地降落，并使两个尖点重力均匀地接触皮肤，接触时间不能超过两秒钟。当同时刺激非常邻近的两个皮肤点时，被试往往仅能感觉到一个点：然后逐渐加大两点之间的距离，达到一定距离后，被试就能感觉到两个点了。两点阈值就是感觉到两个点而不是一个点的两点之间的最小距离。威洛特 (Vierordt, 1870) 最早使用两点阈测量器，并发现，从肩部到指尖，两点阈愈来愈小。这种身体部位的触觉的空间感受性随其运动能力的增高而增高，后人称之为威洛特运动律 (Vierordt ')

sLaw)。

两点阈是触觉定位的一种测量方式。人体各部位触觉定位的准确性和各部位的触觉感受性的大小是一致的。也就是说两点阈值高，说明这部位触觉感受性低。因此，通过皮肤两点阈的测定，能从一个侧面反映出触觉的感觉能力。一般说来，手指和面部的两点阈最小，肩背和大腿、小腿部位的两点阈最大。

六、测痛仪

测痛仪 (pain threshold tester) 乃是一种测定痛觉感受性的仪器。测痛具有很高的生理心理学意义和临床医学价值。由于引起痛觉的刺激都是伤害性的，它们都在不同程度上损伤组织使其释放出化学致痛物质引起局部组织反应。多次重复这样的刺激可使组织对痛刺激越来越敏感，因此给实验工作带来很大困难。理想的测痛方法：(1) 刺激强度的增加与痛觉强度的增加相平行，可分级定量，易确定终点；(2) 对组织无害，可重复测量；(3) 强或弱止痛剂的作用都能反映出来。常用的痛刺激有温度、电、机械、化学等四类，在心理学上常用的测痛仪有如下四种：

(一) 热辐射仪

热辐射仪 (radiant heat tester) 乃是由李等人 (Lee et al., 1951) 发明的痛阈测定方法。先前是用由电流加温的金属等刺激被试者的皮肤，后来改用一种辐射热作用于机体皮肤，产生疼痛刺激，以研究痛觉感受性。仪器由一个可调的光源和一个横切面约 3.5cm^2 的管道组成。形状相似于理发用的电吹风。用辐射热引起痛，可避免触觉干扰，当皮肤温达 44.6 左右即随时间的积累而产生疼痛，有关痛测定的文献中约半数系用此法进行。使用此法时，一般先将皮肤涂黑，将光聚焦照射于 3.5cm^2 左右的范围内。光的强度可用电位器控制。测定方法可分：(1) 固定时间，改变强度：单位用毫卡/ cm^2 /秒表示。实验时要尽可能减少重复测定的次数。从痛阈到最强痛可分 21 个最小可觉差异 (just noticeable difference, 简称 j.n.d.)。(2) 固定强度，改变时间：一般是选定几种光强度，分别测定所需的照射时间。

(二) 钾离子测痛仪

钾离子测痛仪 (K^+ pain threshold tester) 乃是采用直流钾离子透入法致痛，以研究痛觉感受性的仪器。用直流钾离子透入法进行刺激，主要要使用好仪器输出的二根导线电极，一根为正极，一根为负极。实验前先选定人体的测痛部位，如选定右小腿部位，可在其周围涂以导电膏，并加以磨擦，使皮肤阻值降低，以增加皮肤的渗透性，然后取 10×5 厘米的镀银镍片作无关电极 (即负极)，下垫盐水纱布，裹扎于小腿上部，另用有机玻璃制成的测痛头作为有效电极 (即正极)，在有机玻璃测痛头中心的小孔内，注以氯化钾导电液，将其置于被试者测痛部位上 (如图 11-9 中右踝，即三阴交穴位)。当仪器开启后，痛刺激从无到有，从弱到强，逐渐增加，当被试者感到微痛时，立即报告，记录这时的电流毫安数，作为痛阈；当被试者报告“受

不了”时,记录这时的毫安数,作为耐痛阈(或痛觉限度)(pain tolerance)。由于钾离子测痛仪给出的最大电流也不会超过5毫安,所以对人体是绝对安全的。

(三) 弹簧棒

弹簧棒(spring-stick)是一种用机械压力作用于机体皮肤致痛的工具。使用时,以固定的探头对人的皮肤加压,在棒上有指示压力大小的刻度。测试点应选骨头表面的皮肤,如前额、胫骨表面等。加压时应注意加压速度,即压力变化率。同一部位重复测定的间隔不宜过短,弹簧棒是一种使用方便又便于携带的仪器。缺点是误差较大,测试结果比较粗糙。

(四) 止血带致痛仪

刘易斯(Lewis, 1942)用血压带缚在被试的上臂加压,造成上肢肌肉缺血而产生疼痛,此法称为止血带致痛法(pain caused by tourniquet)。后来,根据这种方法设计制造了专门的止血带致痛仪。通常,仪器输出的压力和持续的时间作为刺激强度的指标。止血带致痛法的优点是痛点明确,缺点是重复测试花费时间较长,可应用的测痛的部位较少,被试者对仪器在心理上有一种害怕感。

七、动觉计

动觉计(kinesthesiometer)是一种用于测定动觉感受性的仪器,也可测定关节活动方位控制能力。常用的动觉计一般用于测定前臂位移的动觉感受性,它备有放前臂的鞍座,以便使被试的前臂能保持在适宜的位置上。鞍座一端放在一个半圆仪圆心处的轴上。鞍座可随意地沿着一个半圆的轨迹摆动,摆动的幅度可从半圆仪上的刻度读出来。仪器还备有制止器,制止器可以随时安放和移去,使用制止器的目的是使排除视觉的被试能感觉到前臂摆动的幅度,以便复制。动觉计仪对于主动和被动的的前臂摆动都是适用的。也可以用来测量练习对动觉感受性提高的作用。实验时,主试者任选一高度做标准刺激,让被试认真地体会各种关节所处部位,接着要求被试复制,尽量做到每次误差角度越小越好。然后重复多次,按心理物理法处理实验结果,求出动觉感受性(见图11-10)。

第三节 知觉类实验仪器

一、眼动仪

眼动仪(eye view monitoring system)是测试人眼活动情况和研究有关心理过程的专用仪器。人在进行无意注意或者有意注意时,眼睛的视线总是对准知觉对象的某一点,这一点成为眼睛的注视点,并且不断地转动着视线以转换注意的目标。眼睛是以跳动的方式将视线转换到新的目标上去的,

在每次转换目标以后，眼睛稍许停顿片刻，注视这一目标，然后再跳动到新的注视点去，在注意对象时，眼睛就是这样不断地以注视、跳动、再注视……来达到对知觉对象的整体观察。眼动仪就是用来研究注意眼球运动的轨迹，以确定注意时对不同部分所起的作用，分析信息加工（或讯息处理）过程的某些特点的。

眼动仪是测量眼睛注意目标位置及眼瞳直径的完整系统。它通过红外线摄像机测出瞳孔中心位置与角膜的相对位置，来推算出注意目标的 x 与 y 坐标值，一般每秒输出 50 组数据。研究者通过分析这些数据，来推断内部的心理活动。眼动仪在阅读心理学、工程心理学和实验心理学的研究中占有极为重要的地位。新型的眼动仪配有计算机处理系统，而且记录器能自动追踪人眼瞳孔中心位置（限在 2 米距离范围内）。图 11 - 11 为被试正在眼动仪上做实验的情形。图 11 - 12 是观看一张照片的眼动轨迹图。

二、实体镜

实体镜（stereoscope）乃是一种证实双眼视差和产生立体知觉的仪器。早年惠斯通（Wheatstone, 1828）发现了双眼视差在实体视觉中的作用，开创了现代空间知觉实验的新领域。实体镜的作用在于它可把左眼和右眼的视线分开，使左眼只见左图，右眼只见右图，并帮助两只眼睛辐合，产生立体效果。常见的实体镜有三棱镜式实体镜、反射式实体镜和栅栏式实体镜。图 11-13 为三棱镜式实体镜。

三、长度和面积估计器

长度和面积估计器（length and area estimator）是用平均误差法（或平均误差法）（average error method）测定线段长度的感觉阈限和进行面积估计的仪器。使用时，根据实验目的在较硬的纸上画好线条或其他图形，然后插入槽中，将一边的活动盖放在一定位置，使线条呈现出一定的长度或使图形呈现出一定的面积，让被试调节另一活动盖复制出这一长度或面积，或者该长度（或面积）的几分之一或几倍。刺激的长度可在仪器反面的刻度上读出。经过多次测定后，就可用所得数据（一般采用平均误差法）计算线段长度的感觉阈限。图 11 - 14 为我国生产的长度和面积估计器。

四、动景器

动景器（或动景盘）（stroboscope），它是演示在一定速度条件下运动物体产生视觉效果，说明似动现象的仪器。普拉梯（Plateau, 1833）制造了第一个动景器。直到现在，动景器还是心理学演示似动现象的常用仪器。演示时，先将实验用的卡通片放入景盘，使之紧贴内壁，同时用手慢慢转动手轮，随着景盘的旋转，观察者从窗口向内能看到卡通片上的静止的图形随之活动起来，当景盘的转速在每秒 16 格左右时，动景效果最佳。重复观察几次，即可体会似动现象的基本原理（见图 11 - 15）。

五、似动现象研究装置

似动现象研究装置 (apparent movement phenomenon apparatus) 是产生似动知觉的实验仪器。所谓似动现象 (apparent motion) 是一种知觉现象, 它是在一定条件下刺激物实际上没有动而感知到了运动。这种装置功能大致可分为二类:

1. 可演示似动现象 如有的仪器在呈现箱前安上各种插片, 常见的图片有:

(1) 相继呈现两个飞鸟或线条, 均可产生相当于真动的似动, 看到鸟飞来飞去, 或线条作 90 度转动等。

(2) 相继呈现一正一反的“<”或“>”, “ ”或“ ”的图形, 观察者可看到翻转现象, 运动好像是在三维空间中进行。

2. 可用于显示似动现象产生的时间和空间条件的实验 仪器呈现箱上一般有两个小圆孔, 一个固定, 一个可通过旋转往复移动, 实验时可选定 3~5 种间距使用, 为了做好实验, 尚需选定 5~8 种频率作为时间条件, 与空间条件一起编排实验顺序。

似动现象实验一般需在暗室内进行, 被试者坐在仪器桌前 2 米远处, 眼睛注视着呈现箱上两圆孔的小灯, 被试者观察两个亮点并报告两个亮点是同时出现, 或者先后出现, 或者向一个方向移动, 重复几次, 等到被试掌握了似动现象的标准后开始正式做实验。将实验设计的“同时”、“先后”、“运动”三种情况分别填入实验数据表上, 最后整理出出现次数的百分数, 画出各种时间、空间条件下似动现象出现的区域。图 11-16 为似动现象研究装置。

六、深度知觉仪

深度知觉仪 (depth perception apparatus) 是由霍瓦-多尔曼 (Howard-Dolman, 1919) 设计的, 故又称霍瓦-多尔曼知觉仪 (Howard-Dolman perception apparatus), 是用于研究深度知觉的仪器。当时主要用来作为一种选择测验以淘汰那些不符合深度知觉要求的航空候选人员。它是根据黑尔姆霍兹三针实验 (three-needle experiment) 原理而制成的。这种仪器上有一个固定的直棒, 在它旁边有另一个同样大小的, 可以前后移动的直棒。固定的直棒为标准刺激, 可以前后移动的直棒为比较刺激。被试在 2 米距离外通过一个长方形窗孔观察这两根直棒, 并遥控来调节可移动的直棒, 使二者看起来在同一距离上。在这种条件下, 除了双眼视差起作用外, 排除了其他深度的知觉线索。图 11-17 为深度知觉仪。

七、大小常性测量器

大小常性测量器 (perception constancy tester) 乃是用比配法 (comparative and matching method) 原理测定大小常性的仪器, 也可用于制作心理量表。此仪器一般有两件, 每一件都有一个可调面积的等边三角形放在立柱上, 立柱的高矮可以调节。三角形的高可以从仪器背面的刻度上读出来, 调节范围在 65~165 毫米之间。测定大小常性时, 主试把一个三角形放在远处, 把三角形的高调到一定的尺寸, 要求被试调节近处三角形面积, 使之与远处三角形的面积看起来相等, 然后根据二个三角形的高之差, 计算

大小常性系统。图 11 - 18 为大小常性测量器。

第四节 记忆类实验仪器

一、注意分配仪

注意分配仪 (distribution of attention apparatus) 是测量注意分配能力大小的一种仪器, 也可检验被试同时进行两项工作的能力, 还可用于研究动作、学习的进程和疲劳现象。仪器可呈现二类刺激, 要求左、右手分别尽快地作出回答反应: (1) 声音刺激分高音、中音、低音三种, 要求被试者对仪器发出的连续的、随机的、不同声调的声音刺激作出判断和反应; 用左手按下与不同音调相应的按键, 由仪器记录下正确的反应次数; (2) 光刺激由 8 个灯泡组成, 要求被试者对仪器发出的连续的、随机的、不同位置的灯光刺激作出判断和反应, 并用右手按下与不同位置灯光相应的按键, 关掉灯光, 尽快地工作一个单位时间, 由仪器记录下正确的反应次数。以上两类刺激可分别出现, 也可同时出现, 两类刺激的出现是随机的, 并按规定的时间自动地、连续地出现, 实验结果由仪器自动显示, 根据这些数据就能计算出注意分配分数。图 11 - 19 为简易的注意分配仪。它通过被试左手和右手的协调, 记录注意分配能力。

二、棒框仪

棒框仪 (rod and frame apparatus) 乃是用于棒框测验的仪器, 它的主要特点是在一个暗视场背景上提供一个亮度均匀的亮的框和棒。棒在框的内部, 二者都可单独作顺时针或逆时针调节, 并且有读数盘随时将框和棒的倾斜角度用指针显示出来 (图 11 - 20)。近年来心理学家对认知方式给予很大关注, 成为当代研究个体差异和人格问题的一大趋势, 棒框测验在认知方式研究中占有很重要的地位。早期, 它与身体顺应测验、转屋测验都是垂直知觉研究中经常使用的测验。它们主要用来研究, 当外在的视野线索 (框) 与内在线索 (身体的垂直知觉) 发生矛盾时, 究竟参照哪个线索为主进行垂直判断。美国心理学家威特金等人 (Witkin et al., 1981) 进一步研究证明: 当外在线索与内在线索发生矛盾时, 依内在线索为主进行垂直判断的称为场独立性 (field independence), 把偏于以外在线索为主进行垂直判断的称为场依存性 (或场依赖性) (field dependent), 并且发现垂直判断上的这种个体差异, 具有一定的稳定性。这样他们就可把垂直知觉的问题, 引申到个体差异——人格理论 (theories of personality), 即认知方式的新领域中去。所谓认知方式主要是指在心理机能中主要依靠外在参照, 还是依靠自我对比的倾向。

早期的棒框测验是在暗室中做。后来奥尔特曼 (Oltman, 1977) 设计了手提式测验器, 做起来较为方便。实验时, 令被试者端坐在仪器前, 双眼紧贴观察孔, 实验过程中始终保持这种情况。暗适应 5 分钟后, 开始施测。为了使被测者在实验过程中头部始终保持正直, 不随框倾斜, 因此采用下颞托将头部固定。待被试者准备就绪后, 主试者讲指导语: “注意看框和棒, 并将棒调节与地面垂直, 时间不限, 当你认为棒已调至与地面垂直时就报告垂

直了，并且每次判断垂直的标准要一致。”主试者在被试者报告调至垂直后，记下调节的误差数（所调的角度与真正垂直的差数），不计正负号，只取误差的绝对值。然后用心理物理方法之调整法和极限法安排实验程序和处理实验结果，但较多是采用调整法，因为实验程序和调整法特征十分接近。

三、迷宫

迷宫（或迷津）（maze）乃是研究动物知觉-动作学习的仪器，它的种类很多，结构方式也不一样，有心理的与实物的迷宫两大类，心理的迷宫有听觉的与纸上（视觉）的迷宫，实物的迷宫主要靠触觉和动觉来进行，又可分为T形的、Y形的、平面的、立体的、凹进的、凸出的等等，不下百余种。一切迷宫有一个共同的特征，这就是有一条从起点到终点的正确途径与从此分出的若干盲路。被试者的任务是寻找与巩固地掌握这条正确的途径。学习的量度是用达到一定的标准所需的尝试次数、时间或错误数为指标。一般说来，动物和人类的实验常常以三次完全无错作为完成学习的标准，常见的人类迷宫为触棒迷宫（见图 11 - 21）。

触棒迷宫（或铁笔迷宫）（pencil maze）是在手指迷宫的基础上发展起来的最简便、最常用的迷宫。珀金斯（Perkins，1927）最早使用了这种迷宫，后经心理学家们不断改善。触棒迷宫是在排除视觉的条件下，被试用小棒从迷宫的起点沿通路移动，直到走到终点，小棒进入一次盲巷（一般用声音表示）就算一次错误。学习的进程以从起点到终点每走完一遍所花的时间或以所犯的错误次数表示。

四、速示器

速示器（tachistoscope）是一种短时呈现视觉刺激的仪器，在知觉、记忆和学习等方面的研究中，经常要用适当的仪器来把刺激呈现给被试，以记录他们的反应。沃尔克曼（Volkman，1895）最先把短时呈现视觉刺激的仪器叫速示器。早期常用的是冯特实验室中所用的利用自由落体原理制成的速示器。戈德沙德与缪勒（Goldsheider & Muiler，1893）把转盘原理用于速示器的装置，在电子技术年代之前，这些机械控制方法是制造速示器的主要原则。一百多年来，速示器经历了多次重大的改进。早期的速示器一般为“落体式”速示器，后来发展为“快门式”速示器，继后又出现电子控制的速示器。随着认知心理学的迅速发展，80年代出现了多视野速示器（见图 11 - 22）。美国杰布兰茨公司生产的四视野速示器（four - field tachistoscope），可在四个视野分别单独呈现刺激也可同时呈现刺激，刺激可以是标准卡片也可以是投影片或实物。实验过程中既可半自动换片，也可手动换片。刺激呈现的时间范围为 1 毫秒至 100 秒。由于还可以由计算机与反应时系统联机使用，这样就把认知过程的研究提高到一个崭新的水平。

五、记忆鼓

记忆鼓（memory drum）乃是一种研究记忆的仪器。最初由缪勒和舒曼（Müller & Schumann，1894）设计，后来由李普曼等不断改进，早期的李

普曼记忆鼓 (Lipmann, 1904), 用一个定速马达作动力, 全部装置以一个铝片屏蔽起来, 使被试者不能看到。从二个窗口中, 以固定的间歇不断向被试者呈现刺激材料, 要求被试者识记。由于材料附着在鼓形体的周围, 所以称记忆鼓。记忆鼓具有比较完善的传动装置, 呈现窗口的刺激信号由定时齿轮控制, 因而适用于多种记忆研究, 如提示法记忆研究、系列学习和成对联想研究等, 图 11 - 23 为新型记忆鼓。

六、多重选择器

多重选择器 (或耶基斯选择器) (multiple - choice apparatus) 是由耶基斯 (Yerkes, 1921) 设计定型的。多重选择器通过对被试者简单和复杂的空间位置的概念形成过程的观察, 研究思维问题。仪器一般有 10 余个活动的电键, 分别与声音信号相连, 主试者可任意推出几个电键, 并按已定原则 (例如推出的几个电键中的当中一个等), 使有关的电键与声音信号连接, 让被试者选出和声音信号连接的那个电键来, 直到被试者经多次尝试后, 第一次按键就能选对, 并将选择的原因也说对为止, 此外, 被试者还须说明他是怎样掌握这个原则的。图 11 - 24 为耶基斯设计的多重选择器。

在做实验时, 主试者的任务是在屏障后操作, 按照事先定好的顺序将几个活动键推到屏障前, 并将其中一个键与蜂鸣器连接 (注意: 不要让被试者看到)。然后, 主试者宣读指导语: “请你找出和声音连结的键在什么位置。如果先按一个键, 没有发声, 再继续按其他几个, 直到按对为止。”这时主试者再更换另外几个键, 做法和前面相同, 直到连续三次选对并能说出声音键所处空间关系的位置为止。

第五节 情绪和技能类实验仪器

一、皮肤电反应仪

皮肤电反应仪 (或心理肤电反应计) (psychogalvanometer) 乃是显示或以曲线波纹形式记录皮肤电反应变化的仪器。皮肤电反应最早叫心理电反射 (或心理肤电反应) (psychogalvanic reflex, 简称 PGR), 但这一词不甚确切, 以后一般都采用皮肤电反应 (galvanic skin response, 简称 GSR) 一词。皮肤电反应的机理是: 在情绪状态时, 皮肤内血管的舒张和收缩以及汗腺分泌等变化, 能引起皮肤电阻的变化。皮肤电反应仪就是以此来测定植物性神经系统的情绪反应。最早研究这一现象的是费利 (Fere, 1888), 他把两个电极接到前臂上, 并把它与弱电源和一个电流计串联。他发现当被试接受音叉、气味等刺激时, 电流计就迅速偏转, 后人称之为费利现象 (phenom - enon of Féré), 以后皮肤电反应即被广泛应用。传统的皮肤电反应测量方法多以电阻为定量单位, 也有用假定单位表示的。图 11 - 25 为新型皮肤电反应仪。

二、动作稳定器

动作稳定器 (steadiness tester) 用于测定手的动作稳定程度, 也可

间接测量情绪波动的程度。仪器由一块多孔的金属画板和一根插入洞孔的金属棒组成。这种仪器的画板上一般有九个洞，故又称九洞仪（nine-holes appar - atus），见图 11 - 26。九洞的直径分别为 2.5、3、3.5、4、4.5、5、6、8 和 12 毫米，实验时要求被试按大洞到小洞的顺序，将笔尖伸入洞内一定的长度后抽出。当笔尖碰到洞边时，指示灯就亮，每次伸入和抽出洞时不碰洞边才算通过，最后以三次通过的最小洞直径的平均数之倒数作为动作稳定程度的指标。

三、反应时测定仪

反应时是心理学研究中最常用的因变量，又分简单反应时间和选择反应时间两大类，现分述如下：

（一）简单反应时测定仪

简单反应时测定仪（simple reaction time tester）是呈现单一刺激，要求被试者作单一反应的仪器，主要用于测试人的视听道的反应速度。通过测定，可比较在声光的不同条件下，反应速度的快慢。

使用时，简单反应时测定仪器可分别测定视觉和听觉的反应时间。实验过程，主试者操作开关，选择声或光为约定的刺激信号，当被试者感觉到信号（声或光）时就随即按下反应键，这时计时器就记录了从“刺激”到“反应”之间的时间，如此重复多次取得可靠的数据。从实验获得的数据在需要时还可以与选择反应时测得的数据进行比较。

（二）选择反应时测定仪

选择反应时测定仪（complex reaction time tester）呈现不同的刺激，要求被试者作出不同反应的测时仪器，主要用于测定人对复杂刺激的反应速度和准确性。

教学用的选择反应时测定仪一般能呈现四种颜色光的视觉刺激。做实验时，主试者按预先编排好的程序，操作开关分别发出“红”、“黄”、“绿”、“白”各色刺激信号，要求被试见到某一种颜色信号灯亮时，就随即准确按下手中某一个特定的反应键，如此重复多次记下各次数据。从本实验获得的数据，若有需要还可以与简单反应时测得的数据进行匹配比较。

四、手指灵活性测验仪

手指灵活性测验仪（finger dexterity tester）乃是测定手指尖、手、手腕、手臂的灵活性，以及手和眼协调能力的仪器。应用心理学测定方法进行能力方面的动态研究，能够弥补和纠正用快速法进行职业咨询和职业选择时的不足，通过长期动态的对个人进行研究，可以取得向被试提出选择职业建议的宝贵资料。这种测验方法在就业指导 and 咨询上正得到越来越广泛的应用。

手指灵活性测验仪种类很多，归纳起来主要有二类。一类用于测定手指

灵活性，另一类用于测定手指尖的灵活性。

手指灵活性测定要求被试者用一把镊子钳，把 100 根金属针分别插在实验板上相应的 100 个洞里，金属针和洞的大小规格不等，最小的洞直径仅 1.6mm。

手指尖灵活性的测定，实际上是一个安装螺母的操作；在实验板上固定着数十个螺栓，被测者的任务是尽快地先把垫片放在螺栓上，然后再把螺母旋在螺栓上。

以上二项操作的指标是完成任务的速度，仪器有计时器，只要被试把第一根针（或垫片）放在洞里（或螺栓上），计时器就自动开始计时，当被试者把最后一个针（或螺母）放到最后一个洞里（螺栓上），计时器即自动停止计时。这时计时器上的数码就显示整个实验所需的时间。图 11-27 为手指灵活性测验仪的操作板。

五、双手调节器

双手调节器（two arm coordination tester）乃是一种主要研究注意分配和双手协调技能的仪器。这类仪器的种类很多，多数仪器的主件由一块金属板构成，板上有一个形状不规则的沟槽，有一根金属针可以沿这个沟槽移动，这根针由双手调节器的两个转轮柄来推动。转动其中的一个轮柄，针就会前后移动；转动另一个轮柄，针就会上下移动。被试者的任务是把注意分配到转动两个轮柄的两种动作上，尽可能快移动这根针，使它到达尽头，但又要使它不要碰到沟槽的边缘，否则，电流会接通，记下一次错误。由于整个操作过程，要求对活动进行特殊的组织，这种组织也就是注意的分配。

六、镜画

镜画（或镜画板）（mirror tracer）主要用于练习效果的研究和技能迁移作用的研究。镜画是由一面 6 寸×8 寸大小的镜子，一块画有多角星图案的底板，一块遮挡板和一支金属笔所组成。实验时遮挡板支撑在底板上，挡住被试者直接看到底板上的多角星等图案。被试者只能从镜子里看到颠倒的多角星，并操作金属笔尽可能地沿着图案轨迹移动。主试者记录被试者的实验时间和错误（金属笔跑到轨迹外面）的次数。实验的结果是，初学者对这件工作感到困难很大，但进步很快。通过学习进程的研究和分析，就可了解练习和技能迁移的效果。图 11 - 29 为被试者正在作镜画操作实验。

第六节 计算机在心理学上的应用

一、计算机与心理学的关系

从世界上第一台计算机在美国普林斯顿的一所高级研究所诞生以来，短短的半个世纪中，电子计算机已进入第四代、第五代，至今出现的算法语言已不下几百种。微处理机和微型计算机问世以后，由于其体积小、价格低和推广方便，故它的发展速度和影响远远超过了它的前几代。数字（位）计算机目前已不仅用于数学题目的计算上，而且广泛地用于数据处理，例如

图形显示、测量，设计工程、电路以及实验、模拟等。在科学研究中计算机已成为强有力的助手。可以预见，它对科学技术的发展和人类生活将愈来愈显示出不可忽视的影响和作用。今天，计算机科学已经渗透到各个领域，计算机作为实验工具是否被充分利用，以及计算机和心理学结合程度，标志着心理学学科发展的水平。

从前面几节的介绍中，我们已经知道，心理学像许多其他自然科学一样，特别重视采用先进的技术作为工具来武装和发展自己，因此，将电子计算机作为一门技术广泛地在心理学各领域的实验中加以应用，毫无疑问地，这将是现代科学心理学的任务。除此而外，心理科学和计算机科学这二门学科之间存在着许多共同需要解决的问题，如图形识别、人工智能等。这类问题的解决既利于计算机技术及学科的发展，又有利于心理科学的发展，相辅相成，这种应用和结合带来了共同的发展。

计算机在心理学的应用从 60 年代初开始到现在已有一段成功的历史。计算机的基本操作方法是体现在它的程序算法，这一特点使得它们在作为实验工具时非常有用。现在，计算机在心理实验室中已被用于研究许多不同类型的心理学问题。例如，认知科学中的许多实验已经通过使用计算机来完成。在感知方面，计算机已经广泛地应用在视觉和听觉研究中，在诸如控制心理物理实验过程，进行心理生理量的测定，以及帮助产生听觉和视觉刺激等方面发挥很大的作用。计算机还补充用于模拟人的感知、记忆、学习，以及行为等等。同时，纷繁复杂的各种计算机程序在心理实验中发挥了刺激呈现、过程控制和数据处理的作用，充分显示了用计算机进行心理实验所具有的准确、方便等特点。

概括地说，计算机在心理学中的应用主要有三个方面：

1. 人-计算机系统的研究；
2. 心理和行为的计算机模拟；
3. 计算机在心理实验的控制和数据处理中的应用。

二、人-计算机系统的研究

人-计算机系统乃是人-机系统的一部分。人-机系统（或人机合一系统）（man-machine system）是人和机器组成，通过人机相互作用以实现一定功能的系统，它被看作是控制系统的特殊情况，是工程心理学研究的对象。在这种系统中，人和机器形成一个统一的调节回路。为研究人-机系统中进行的一些控制过程，以及它们的结构和功能，工程心理学把人、操作者看作是这种系统的一个环节。图 11-30 是闭环人-机系统单回路图的基本结构。图中 X_1 、 X_2 、 X_3 、 X_4 代表从系统的一个环节到另一环节的作用的转换。



图 11-30 人-机系统的基本结构图
（采自洛莫夫，1958）

事实上，人-计算机系统也是人-机系统。但是此系统不同于一般的人-机系统或人-自动机系统，这是由于在人-计算机系统不是去获得新的信息，而是根据人或其他系统的需要去处理或应用已有的信息。在此，人-计算机系

统中人和计算机相互作用的实质是利用程式，计算机利用人给予的程式进行运算，而人要再从计算机中获得对人最合适和可靠形式的信息。这样，在人-机系统中用计算机实现信号响应能力的测定。

这里用一个例子加以说明。为了使计算机实现信号响应能力，一个突出的例子是挑选最佳的字符。国内外心理学家在这方面做了大量工作。我国崔代革（1986）为达此目的，使用计算机控制主要仪器，用以模拟高度字符动态显示的情景。它与人（被试者）组成一个人-机系统，通过人的信号响应能力来评定高度字符显示的优劣，以手动方式作出响应。

图 11 - 31 为信号响应示意图，实验过程中，被试通过接受显示器上所呈现的动态画面，准确、迅速以手动方式（按键）对指令作出反应。微型计算机记录反应时、正确响应次数、提前响应次数及延迟响应次数（包括漏反应次数）。每人在连续响应 30 次后，微型计算机显示并打印统计数据，还能将每人响应情况用直方图形式显示出来。

由于实验模拟了平显动态情景，并以四种指标进行探讨，结果发现，具备最优显示方式的字符应具备直观性、字符不过简也不过繁等特征。

应该看到，电子计算机仍然是按人所编好的程式执行解题。所以说，在人-计算机系统中，人仍然起着控制机能，并作为主要的决策部分。在问题解决与决策的过程中，人仍然承担问题识别、问题诊断及程式的制定和选择等创造性工作。在现代概率信息处理系统中，计算机也正在利用人的一套决策规则来运算。

由此可见，人-计算机系统中人和计算机处于同一个系统中执行着同一个任务，这是人和机器的同一性的一面，在这统一的系统中由人和机器共同完成最优控制。

三、心理活动的计算机模拟

人类具有感觉、知觉乃至高级的思维和意识能力。这些能力能否完全地传递给机器人，即由计算机或机器人来执行这些功能？最终计算机会不会代替人的脑力劳动？对这样一个问题不能作简单的肯定或否定的回答。绝大多数的科学家认为，机器只是劳动工具，人是劳动的主体，劳动是人有意意识的活动。从这种观点看，即使是最先进的电子计算机也是由人创造的，计算机中的程式，运算的目的性都是由人来决定的。当然，在肯定机器不会最终代替人的前提下，也不能否定可以不断扩大计算机模拟（simulation）人的高级功能的可能性。计算机模拟人的心理和行为的前景是很广阔的，心理学也可以从这种模型研究中，深入探索和验证已有的理论并探索新的心理和行为活动的规律性。

人的心理和行为的计算机模拟是可行的，心理和行为的计算机模拟这项工作具有双重意义。一方面，人脑是产生心理和行为的生理基础，但是人脑极为复杂，而且不易直接进行实验和观察，于是运用信息处理和控制论原理，设计一个电子计算机系统作为脑的原型，将是分析人脑工作原理的重要方法；另一方面，人脑工作原理的阐明，又将为计算机对心理和行为的模拟提供新结构与新原理。

计算机可以从不同的角度来模拟人的心理和行为。从方式上看，可分为结构模拟和功能模拟等，目前在感知模拟、记忆模拟和思维模拟等方面已经

取得显著的成绩。

（一）感知模拟

早期的感知模拟（sensation and perception simulation）企图从感觉的神经结构模拟着手，可称为结构模拟阶段（form simulation period），其特征是从人的神经解剖上模拟，显著成果是 60 年代初期盛行于世的感知机（perceptron）。感知机的实验系统是由一个主机（感知机）、一个刺激环境和一个强化控制装置三大件组成。其中感受单元把刺激的某一物理量转换成相应的信号；联系单元把此信号进行加权总和和阈限运算；再由反应单元对结果进行分类显示。随着时间的推移，人们发现感知机的局限性，因为图像中包含的信息量是巨大的，即使同一类图形其变化也是繁多的，而且在文字符号识别工作中，许多类别是人为规定的。所以感知机再复杂，如果不考虑预先处理工作，特征抽取工作以及各种特征之间的关系问题，也是不容易解决问题的。

科学家从现代心理学的研究得到启发，把注意力转向心理功能的模拟。举一个例子来看：计算机图形识别的心理研究表明，人们识别图形时，不仅要辨别出某些特征是否存在，而且要看特征之间的关系，即图形结构；特征相同，结构不同其含义可能大相径庭。根据这种方法，认为图形的基本单元是原始特征，或叫元素，一定关系的一些元素构成子图形，一定关系的子图形构成图形。图形的识别，就是使计算机把这些简单元素，按照一定法则来分析理解，图 11-32 为计算机识别的脸图。构成它的成分有：黑点、方块、线段、圆。

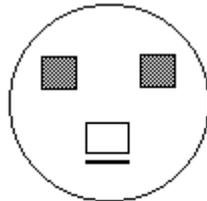


图 11-32 计算机识别的脸图

至今，计算机对图形的识别，仍有很大的局限性。尽管计算机的模式识别研究取得了不少成绩，但比起人的知觉还相当幼稚，极不成熟。可是从中却揭示出一条规律单纯的感觉，纯粹的知觉是可能存在的，只有理解性的知觉，即高级认识过程与低级认识过程的结合，才是人们反映、认识客观世界的真实途径。

（二）记忆模拟

自艾宾浩斯以来，对记忆已完成了大量的实验研究，20 世纪 50 年代出现的信息处理（信息加工）理论，把记忆看作是对信息的输入、编码、储存和提取的过程。记忆的心理学研究也大都集中在揭示遗忘的规律上，但是计算机具有“过目不忘”的特长，不可能也不必要用它来模拟遗忘。然而，人们却同样期望从计算机科学中得到启发，说明记忆信息在人们记忆系统中是如何输入、编码、储存、提取的。于是各种记忆系统的理论相继出现。所谓瞬时记忆、短时记忆、长时记忆理论已是心理学的常识。目前人们根据计算

机模拟提出了短时记忆、长时记忆和工作记忆系统理论。认为除短时记忆和长时记忆之外，还有一个工作记忆系统的存在。工作记忆系统具有控制功能，它可以随工作需要临时把某些信息从长时记忆系统中调出，用过以后或是原封不动，或是以经过进一步加工的新的形式重新送回长时记忆系统。这种假设显然类似于计算机语言中的分层组织程序，即每程序可能包括一个或多个子程序，当程序的执行到子程序时就执行子程序，子程序执行之后，控制又返回继续执行程序。

通过扩大计算机储存量的模拟研究，也进一步逼近于人的记忆规律。计算机是以它巨大的储存量而著称的，现在的大型机储存容量接近 10⁸ 位，一台磁盘的容量就超过 10⁹ 位，估计光存储器的容量可达到 10¹² 位左右。但对于大批需要存入计算机的信息，计算机的容量还显得不够。一张高质量的空中摄影照片所包含的信息量就超过了 10¹³ 位。因此，人们可以发现，若不经过加工，即使是计算机，它们的记忆储存都是相当有限的，而是以经过加工之后的组块（或意元）（chunk）的形式单位存储的。

（三）思维模拟

目前思维模拟（thinking simulation）是人工智能最活跃的领域。计算机的思维模拟，其方法大体分为两类：一是定程法，二是启发法。

1. 定程法 定程法（或算法）（algorithm）乃是问题解决的一种策略或方法，它含有一组规则，如果得到遵循，就将自动地得出正确的答案。例如，乘法规则就构成一个算法，只要恰当地应用，总可得出正确答案。所以，定程法就是为了使问题得到一个解，而规定在一切可能的条件下所采取的步骤。一个定程法就是确定一个计算过程的一连串步骤，应满足两个要求：（1）应用定程法于一定的原始资料上，如有答案，则必能在有限的步骤内得出；（2）定程法适用解决一类问题，而不是某一个特定问题。

虽然，计算机穷尽一切可能性的定程法能解决问题，与心理学上称之为试误法（trail-and-error method）相类似。实际上人们解决问题时运用尝试法的场合并不多，原因之一是人脑比计算机慢得多，于是不得不要绕过组合爆炸，尽管人脑如何绕过“组合爆炸”的机制尚不完全清楚，毕竟还是可以从它的工作方式中得到启发。所以，对许多复杂的问题，至今尚未发现其适当的算法，也许根本就不存在这些算法，而有些定程法又过于费时，于是，需要发展一种启发式程序设计，这时就转向启发法。

2. 启发法 启发法（式直断法）（heuristic）乃是一种凭借经验解决问题的策略或方法。按照“heuristic”一词的原意，是有助发现的或发明的艺术。人在生活中积累了行之有效的解决问题的方法。在新问题出现时，可以利用这些经验去解决问题，而不必进行大量的尝试或系统的探求。在西蒙（Simon, 1975）的信息处理（或信息加工）心理学理论中，启发法是问题解决的主要途径。科学上的许多重要发现往往是利用了极为简单的启发法规则。相反，用定程法解决问题时，要把所有的可能性都尝试一遍，以求最后找到答案。因此，用定程法解决问题要花费大量时间和精力；启发法可以减少尝试的数量，迅速地解决问题，但不排除失败的可能性。启发法的程序的设计，就是从人的智能行为中得到启发，找出人解决实际问题的规律性，把一些策略、技巧、经验法则、简化规则以及其他有助于求解复杂问题的方法

编进程序。

近些年来，思维模拟的研究也正朝向应用化方面发展，这方面取引人注目的是各种类型的专家咨询系统，所谓专家咨询系统（expert consulting system），就是一个计算机系统事先将有关专家的知识总结起来，分成事实及规则，以适当的形式存入计算机，即建立起知识库，根据这样的数据库，采用合适的控制系统，按输入的原始数据选择合适的规则进行推理、演绎，做出判断和决策，能起到专家的作用，所以称专家咨询系统。我国也曾报导过中医临床诊断的专家咨询系统。如 1984 年，南开大学和天津医学院合作研制的小儿咳喘的计算机软体系统，它是一个小型中医诊疗专家系统，已在电脑上运行，经过 100 个实际病例的考验，结果正确无误。经过 234 个典型病例检验分析，证明可进行临床诊疗工作，并已开始门诊。

上面分别叙述了计算机的各种心理行为、功能模拟，从 60 年代起，这种模拟已经进入综合化阶段，具有感知、记忆、思维功能的机器人已经在各行各业大显身手，科学家们正在进一步设想建立一个通用的智能机器人，可以相信这种机器人将日臻完美。

四、计算机在心理实验中的应用

前面讲到的计算机功能，对心理实验来说还是比较间接的。计算机还能直接用于心理实验，我们知道，微型计算机和微处理机在过程控制中得到广泛的应用。事实上，心理实验的控制是一般过程控制的变形。在这种情况下，控制装置是计算机，过程是实验。图 11 - 33 所示为微型计算机控制心理实验的方框图。在心理实验中，由微型计算机产生刺激，被试对刺激作出响应，然后将判断信息反馈到计算机，对刺激作出调整。

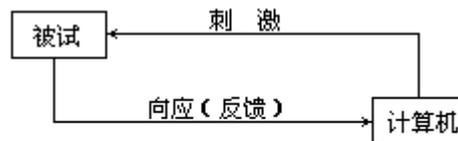


图 11-33 计算机控制的心理实验
（采自王 甦，1986）

在心理学领域中，利用计算机做实验过程的控制设备，不但能够达到实验过程的自动化，而已还能够改善实验条件，保证实验质量，提高实验效率。我国心理学者王甦（1986）曾做过利用微型计算机控制被动触觉的一个实验，由被试在排除视觉的条件下，判断以不同速度经过优势手食指的实验材料，检验被动触觉锐度。图 11 - 34 为实验系统的结构图。我们从这二个图上可以看到，从刺激的产生与呈现，到响应的记录，最后到数据的处理和分析，在这心理实验的主要步骤上，都由计算机加以控制。

（一）计算机在心理实验的应用领域

计算机已被应用于许多心理学领域，其主要领域有：

1. 认知 认知科学有许多实验是使用计算机完成的。几乎所有的计算机系统能够做到：显示口头或其他刺激材料，呈现需解决的问题，模仿真实

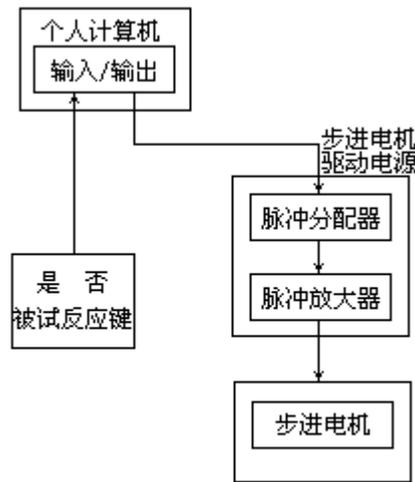


图 11-34 实验系统的结构图
(采自王 甦, 1986)

或想象的情景，以及提供测定认知能力的复杂比赛。特别是在视觉和听觉研究中，计算机广泛地使用于控制心理物理过程，使用于对眼动和瞳孔扩大参数作心理生理测量，使用于帮助产生听觉和视觉刺激。

2. 技术和运动操作 计算机能同时控制大量刺激和监视被试反应的能力，使它成为连续运动知觉任务中研究人类操作的一个出色装置。例如，模拟驱力行为，操作工业机器和测量紧张和疲劳程度。

3. 计算机帮助治疗 目前使用于行为治疗和行为调整的认知治疗中的许多方法，均特别适宜采用计算机控制技术。例如，监视一个病人的生理和其他反应，根据确定的标准呈现刺激，测量肌肉活动松弛程度和皮肤电阻的变化，或病人的反应输入到一个简单的键盘上等等。

4. 心理生理学 联机计算机技术频繁地使用于测量或控制由电测量派生的心理生理参数。心电图数据、肌电记录和许多指标都可借助于联机计算机系统而获得。因此，所获得的测量可为以后的分析而储存起来。脑电图的研究也属于这个范畴。

5. 心理学教学实验 心理学上的许多教学实验可以在计算机上进行，例如北京大学电子仪器厂开发了心理学实验软件。其中有做呈现刺激的心理测试，还有统计计算软件。呈现刺激的心理测试软件包括测量记忆广度，测量选择反应时、测量感觉阈限、研究学习进程、研究人在概念形成过程中头脑里所进行的内部活动等等，这些都适用于教学实验。

(二) 心理实验运用计算机的优缺点

前面我们已经讲到了心理学中使用计算机的种种优点，这里不防再加以系统地归纳。一般说来，计算机应用于心理实验中有如下一些优点：

1. 较好地控制变数 实验是一种控制的观察。实验的主要特点是较好地控制额外变量。计算机在这方面显示其明显的优越性。与其他实验装置不同，计算机能够向被试呈现随机变化的刺激。以选择反应为例，依据被试选择的反应，计算机程序能够从多个刺激中选择一个向被试呈现，或者只允许向被试呈现一个刺激。然后，极快地形成新的刺激装置。计算机能做到的这些事，是人工实验所做不到的，因为人类不能这样快地作出决定。

许多实验（特别是在视知觉领域）不仅要求刺激者有精确性，而且要求有复杂性。计算机控制技术乃是产生复杂刺激模型的最好方法。例如，谢波德和塞马克（Shepard & Cermak, 1973）创立了一套研究知觉的自由形式刺激。他们设计了八个刺激图形，通过采用“菜单”的形式供被试或被试者选择。计算机还可以控制刺激图形中一些参数的变化。实验者或被试也可以根据需要，通过键盘或旋钮改变刺激图形中的一些参数，从而使刺激图形的调整变得十分方便。

2. 计时精确 由于时间是心理实验的常用反应变量，因此对时间的控制和记录就成为实验的基本要求。使用计算机控制心理学实验的一个重要优点是能精确地控制实验所要求的限定时间。计算机应用于呈现刺激和测定反应两个方面。对于计算机来说，1 毫秒的精确度是很容易达到的。如果需要的话，100 微秒的精确度也是相当容易达到的。这意味着，反应时的模式就可以被制成一个较好的量表，刺激呈现的定时精度也能得到改进。在测量双耳间注意分配的实验中，特别需要尽可能地严格控制刺激的突然开始。在早期的实验中，精确度要高于 100 毫秒是很困难的。计算机技术的出现，就能以大约 10 毫秒的精确度匹配突然出现的刺激。技术上这些改进的结果是，实验的输出起了变化。以前延期较长才能产生的效应，现在能产生突发且稳定的新效应。例如，在对缪勒-莱尔错觉（Müller-Lyer illusion）研究的实验中，通过对刺激呈现时间和刺激消失时间的控制，使刺激图形一会儿出现，一会儿消失，这可以在一定程度上避免由于长时间同一错觉图而错觉量降低这一因素对实验结果的影响。

随着时间精确度的提高，计算机还能使实验者有机会增加对刺激材料其他维度的时间控制。例如，在听觉实验中运用不同的声音频率，或在视觉显示实验中对刺激物的位置、速度或加速的控制。

3. 可减少实验者对被试的影响 计算机控制心理实验的另一个重要优点还在于可以减少实验者对被试的影响。一般在心理实验进行过程中，实验者和被试者之间无意的相互作用有时会使实验结果产生人为的影响，使得结果不很客观。在许多心理实验中，特别是在视错觉的研究中，言语的指导作用会使被试的判断结果大不相同，应用计算机的实验程序能够做到无需实验者的参与，被试也能照常进行，只要被试在实验前预先做几次练习，便可以独立地完成整个实验。而且，在实验进行过程中，实验者可以去做其他的工作，这样既节省了许多时间，还避免了无关暗示对被试者的影响。

4. 记录方便 由于计算机是一台理想的记录器，许多精密的心理实验仪器（如眼动仪等）都和计算机一起组成一个系统。由计算机控制心理实验，可以方便记录和测量响应。被试者的响应主要有二类：离散变量响应（discontinuous variable response）和连续变量响应（continuous variable response）。前一种响应的例子如人或动物被试的按键；后一类型的响应如在整个实验期间需要监控的心理生理变量。无论是哪一类被试反应，计算机都能较好地记录。这里再以波根多夫错觉实验为例，记录的是离散响应，被试者有五个按键可以选择。假如它们是 U、I、D、F 和 Y。当被试要调整图形时，可按前四个键。当被试觉得两条截断线共线时，可按 Y 键，计算机便可求出错觉量。可见，计算机不仅可以对实验进行控制，而且还具有强大的数据处理能力，在这个实验中，只要采用简单的迭加技术，得到的错觉量就具有统计性质。此外，计算机是一种多用途的机器，因为依据的程

序如何，它能应用于任何一种巨大的作业，这种可塑性促使它能适用于许多种心理实验。有些实验需要几种类型（例如，动物实验中，不同的强化日程表），就可以通过一定的程序，来达到实验目的。

计算机在心理实验中显示了共同性的优点，在心理测验中也显示了特殊性的优点。但是用呈现刺激的软件在计算机上做心理实验也并非万能的，有时它有一定的局限性。概括起来，主要有以下四个方面：

1. 有些心理实验难于在微型计算机上实现。把它用于简单操作反而带来了技术困惑。例如，金恩（King, 1973）借助于混合未煮过的黑白稻米，用照相拍下其结构图，以产生随机的黑、白点的视觉模型。正如他们所指出的，这种方法要比用计算机在萤光屏上产生的随机模型的方法好。

2. 有些心理实验只单纯靠软件来实现不能达到预期的效果，测试结果也很不准确、不可靠，因为该软件的执行时间相对某些测试的反应量也还是太长了。

3. 一般地说，编制计算机程式是一项不易的工作，而且有时计算机程式是一种花费时间的操作，因为付出的努力中包括学习一种新的程式语言。如果要做大量有关实验，或者这里没有其他方法可以达到目的，那么是值得花费时间的。如果这些实验无须重复它的原始形式，如果这程式对其他程式关系不大，或者掌握它需要学习很长时期，或者使用一种你可能今后不会再用的语言工具，在这些情况下，改用其他方法就更好些。

4. 应用计算机的一种真正严重缺陷是，它们剥夺了主试者去观察被试者的机会。实验（有别于论证）是一个过程，其结果是不能预先确定的。在可能出现的结果中，也许有些行为已被人类观察者注意到，但是不能被他们所设计的测量系统所记录到。像意外的反应经常为以后的实验工作提供材料，给人以启示。在计算机执行的实验中，这个主试者必须自始至终遵守他所写下的程式算法，而且主试者也可能不愿去花时间来修改实验过程中的细小法则，相反，这些修改在主试者控制的实验中是较容易做到的。因为人既能掌握原则，又具有灵活性。

本章摘要

1. 心理实验仪器是心理学研究的工具。它既包括心理学特有的仪器，也包括一部分属于其他学科，但也可应用于心理学的仪器。

2. 心理实验仪器的发展大致经历了三个时期：从 19 世纪 70 年代到 20 世纪初是第一时期；从 20 世纪 20 年代开始到 60、70 年代是第二时期；自 20 世纪 60、70 年代开始到现在是心理实验仪器发展的第三时期。

3. 仪器对心理学研究之所以重要，主要原因有：仪器能使我们在已知的和控制的条件下获得标准化的数据和资料；仪器能使信息进一步记录下来，提供进一步的分析；仪器能使我们测量到我们感觉器官不能直接观察到的事件，从而大大地扩展了我们的观察范围。

4. 当今心理实验仪器的发展有下列一些新趋向：不断向标准化努力；系列化；应用于教学；心理学研究材料的增多；电子计算机的普及。

5. 用于研究感觉的基础性仪器有：色轮，听力计，声级计，语图仪，两点阈量规，测痛仪，动觉计。

6. 用于研究知觉的基础性仪器有：眼动仪，实体镜，长度和面积估计器，动景盘，似动现象研究装置，深度知觉仪，大小常性测量器。

7.用于研究记忆的基础性仪器有：注意分配仪，棒框仪，迷宫，速示器，记忆鼓，多重选择器。

8.用于研究情绪和技能的基础性仪器有：皮肤电反射仪，动作稳定器，反应时测定仪，手指灵活性测验仪，双手调节器，镜画。

9.概括地说，计算机在心理学中的应用主要有三个方面：人-计算机系统的研究；心理和行为的计算机模拟；计算机在心理实验的控制和数据处理中的应用。

10.人-计算机系统是人-机系统的一部分。其特点是，人-计算机系统也是人-机系统但不是去获得新信息，而是根据人或其他系统的需要去处理或应用已有的信息。

11.计算机可以从不同角度来模拟人的心理和行为。从方式上看，可分为结构模拟和功能模拟；从内容上看，可分为感知模拟、记忆模拟和功能模拟等。

12.计算机应用于心理实验有如下一些优点：较好地控制变量，计时精确，可减少实验者对被试的影响，记录方便等。

建议参考资料

1.王 甦（1986）：利用微机控制触觉锐敏度实验。北京市：中国首届微机在心理实验中的应用讨论会专题论文。

2.孙 晔（1979）：美国心理学漫谈。心理学报，11卷，2期。

3.杨治良（1984）：仪器在心理学研究中的作用——兼评介美国心理实验仪器。心理科学通讯，7卷，2期。

4.杨治良（1988）：心理物理学。兰州市：甘肃人民出版社。

5.崔代革（1986）：用微机实现动态画面显示——平显高度字符信号响应的实验研究。北京市：中国首届微机在心理实验中的应用讨论会专题论文。

6.虞积生等（1980）：中国正常男青年的深度视觉阈值的测定。心理学报，12卷，3期。

7. Lafayette Instrument（1992~1993）. Psychology & Biology Catalog. Lafayette: Lafayette Instrument Company.

8. Stoelting Co.（1990）. Products From Volume 1~4. Chicago.

9. Takei & Company（1988）. T.K.K. Psychological & Physiological Apparatus. Tokyo.

10. Witkin, H.A, & Goodenough, D.R.（1977）. Field dependence and interpersonal behavior. Psychology Bulletin, 84, 4, 661~689.

