

1 概 论

感官评价在工业中是一个新兴事物。它是伴随着消费品公司（主要是食品公司）的快速发展，在最近 40 年中产生的。感官评价的进一步发展取决于一些因素，其中最重要的一条是人员及其培训。

——Elaine Skinner (1978)

1.1 定义与焦点

20 世纪下半叶，随着加工食品和消费品工业的扩展，感官评价领域迅速地成长起来。感官评价包含一系列精确测定人对食品反应的技术，把对品牌中存在的潜在偏见效应和一些其他信息对消费者感觉的影响降低到最小限度。同时，它试图解析食品本身的感官特性，并向产品开发者、食品科学家和管理人员提供关于其产品感官性质的重要而有价值的信息。Amerine、Pangborn 和 Roessler 于 1965 年对该领域作了全面的回顾，Stone 和 Sidel (1993) 与 Merilgaard、Civille 和 Carr (1991) 的教科书也已于最近出版。最近的这两本著作有很强的实践性，针对工业上的感官评价专业人员，反映了著作组的哲学思想。本书的目的是对该领域提供一个较全面的概观，基于研究发现的一个均衡的观点，适合于相关的学生和从业者使用。

感官评价是用于唤起、测量、分析和解释产品通过视觉、嗅觉、触觉、味觉和听觉所引起反应的一种科学方法 (Stone 和 Sidel, 1993)。该定义已被各类专业组织中感官评价委员会所接受和认可，如食品技术专家协会和美国检验和材料学会等。感官评价的原理和实践包括定义中所提及的四种活动，请注意“唤起”一词。感官评价提出了应在一定的控制条件下制备和处理样品以使偏见因素最小这一原则。例如，感官检验者通常应在单独的检验室检验以便得出他们个人的判断，而不会反映周围的观点。样品应以随机数标记，以使人们得出的判断不是来自标记，而是来自其自身的感官体验。另一个例子是应如何使产品以不同的顺序提供给受试者，以帮助测量和平衡由于一个接一个地检验产品所引起的连续效应。对于样品的温度、体积和间隔时间应建立标准的操作程序，以控制意外的变化和提高了检验精度。

接下来是“测量”一词。感官评价是一门定量的科学，通过采集数据，在产品性质和人的感知之间建立起合理的、特定的联系。感官方法主要来自于行为研究的方法，是通过观察和测量人的反应的方式。例如，我们可以估计人们能分辨产品微小变化的次数的比例或者是一组受试者中偏爱一种产品的比例。又如可以使人们产生数量化的反应，以表示他们对于一种产品尝起来或闻起来个人的感受有多么强烈。对于应该采用多少种测量方法以及他们潜在的缺点和适用范围，行为研究的方法与实验心理学提供了方向。

感官评价的第三个过程是分析。适当的数据分析是感官检验的重要部分。通过人的

观察而产生的数据经常有很大的变动。造成人的反应变化有很多的原因，这些在感官检验中难以完全控制。比如参与者的情绪和动机，对感官刺激先天的生理敏感性以及过去的经历和对类似产品的熟悉程度。虽然针对这些因素有一些筛选方法，但也可能只是部分地控制了这些因素，一个检验小组的成员由于其自身的特性，就像是一个为了产生数据的来源不同的仪器。为了评估观察得到产品性质和感官反应间的联系，可能是真实的而不仅仅是不可控制的反应变化的结果，可以采用统计方法来分析评价数据。组合应用适当的统计分析是一个好的实验设计应考虑的问题，以便各种影响因素都能以一种可得到合理结论的方法予以考察。

感官评价的第四个过程是解释结果。感官评价练习是一项必要的实验。在实验中，数据和统计信息是在解释假设、背景知识、结论的含义和应采取的措施的过程中惟一有用的内容。所下的结论必须是基于数据、分析和实验结果而得到的合理判断。结论包括所采用的方法、实验的局限性以及研究的背景和前后框架。感官评价专家不仅仅只是得到实验结果的一条途径，专家们必须给出解释并根据数据提出合理的措施。为指导进一步的研究，他们与顾客——实验结果的最终使用者应该是真正的伙伴。感官评价的专业人员在认识实验结果的合理解释及广大消费者对产品的感受时处于最好的地位，而对于消费者，这些结果可能并无特殊意义。感官专家最好能理解实验操作的局限性及其所冒风险和不利条件可能是什么。

一名准备以研究作为职业的感官科学家，必须接受定义中所提到的所有四个阶段的训练。他或她必须了解产品、作为测量仪器的人，以及统计分析和研究过程中数据的解释。正如 Skinner 所指出的，该领域未来的发展取决于培训新的感官科学家的广度和深度。

感官评价也是一门测量的科学。像其他的分析检验过程一样，感官评价应考虑精度、准确度、敏感性，而避免错误的结果 (Meiselman, 1993)。精度与行为科学中的可靠性概念相似。在任何实验操作中，我们希望重复实验能得到相同的结果。通常，在得到的结果附近有一些误差波动，所以在重复实验中结果不会总是完全相同。这对由人的感知能力作为产生数据的必需部分的感官检验是非常正确的。然而，在许多感官检验操作中，总还是希望尽可能减小误差波动，使实验在重复测量时误差较低。这可以通过几种方式实现。如前面已经提到的，我们解析了对于影响因素的感官反应，尽量减小外来影响，以及控制样品的制备和展示。必要时，我们可以筛选和训练参与者。

其次，需考虑的是检验的准确性。在物理科学中，这被看做是检验仪器得到的测定值接近“真实”值的能力，而“真实”值是由已被适当校准的另一种或一套仪器独立测量得到的。这一原则在行为科学中有一相关概念，称作检验的正确性。这反映了检验程序对于所要测量内容的测量能力，正确性可通过许多方法得到。当检验结果对预知另一条件或另一测量方法所得结果有价值时，一个实用的标准是正确性的前提。例如在感官检验中，检验结果应反映可能购买产品的消费者的感受和看法。换句话说，感官检验结果也适用于其他更多的人。检验结果也可能与仪器测量方法、操作或系统因素、贮存条件、货架时间以及其他会影响感官品质的条件有关。在考虑正确性时，我们也必须注意检验所得信息的最终使用。一种感官检验方法可能在有些情况下是有效的，而在另一些情况下则不然 (Meiselman, 1993)。一项简单的差别检验可以辨别一种产品是否发生了

变化，但却不能说明人们是否喜欢这种新产品。

一项好的感官检验应尽可能减少测量误差和结论中的错误。在任何检验过程中，会有不同类型的误差存在。检验结果是否反映了真实的现象是一个重要问题，特别是在误差和不可控制的波动是测量过程所固有的情况下。感官检验最主要的是产品区别检验的敏感性，也就是说检验不会经常遗漏所存在的重要区别。“遗漏区别”就表明检验过程是不灵敏的。为保持高灵敏性，我们必须通过仔细控制实验以及选择和训练合适的参与者以使可能的误差波动最小。检验必须包括大量的测量数据以确保统计得到严密而可靠的统计值，如平均值或比率。用统计语言，发现真实的区别就是避免了类型Ⅰ误差和得到 β 风险的最小值。从统计学的观点来看检验力和敏感性将在统计学附录中进行讨论。

在检验中可能产生的另一项误差是发现一个在感官检验之外，大量的人或产品实际上并不存在的正结果。一旦发生，一个正结果通常意味着在检验产品间存在着统计学意义上的显著区别。这对于采用一种避免错误的正结果，或统计术语中的类型Ⅰ误差的检验方法显得很重要。应用于科学发现的基本统计训练和普通的统计检验可引导避免产生此类误差。不管检验结果反映了真实的区别还是结果可能是由于偶然因素产生，随机产生的背离结果也必须考虑在结果中。一般的统计学过程可确保我们限制发现一个并不真实存在的区别的可能性。统计过程将这一风险降低到了很低的水平，通常我们所做检验的最高限度为5%。

注意这种错误的正实验结果的误差在基本的科学研究中有潜在的破坏性：假如结果仅仅是由于随机因素引起的，整套理论和研究计划可能由于虚假的实验结果而偏离方向。因此，我们在适当应用统计检验时应警惕此类危险。然而在产品开发工作中，类型Ⅰ误差，也就是遗漏一项真实的区别，可能同样具有破坏性。这可能是使产品在某一种感官上变得更好或更糟的一种重要成分或工艺过程的改变，而这一改变未被发现。所以，感官检验同样应注意不要遗漏真实的差别，以避免错误的正结果。这类额外的统计学问题在感官专家的实验中，与许多其他科学研究领域的人相比，他们需要考虑得更多。

最后，绝大多数的感官检验都是在一套工业体系中进行的，商业利益和战略决策也包括在内。作为在决策时减小风险和不确定因素的手段，我们能够体会到感官检验所产生的效益。当一名产品开发经理要求进行感官检验时，通常是因为关于人们对于产品的感受如何存在一些不确定性。为了解该产品是否不同或相同于某种标准产品，是否可被推荐作为某种有竞争力的标准，是否具有特定的令人满意的品质，都需要数据来回答这些问题。有数据在手，最终用户在较低的不确定性或商业风险的情况下也能作出明智的选择。在大多数情况下，对于研究者和市场经理来说，感官检验的作用是一种减少风险的机制。

感官评价除了有产品开发等明显的用处外，还可给其他部门提供信息。包装的功能性和方便性可能要求进行产品检验，产品质量的感官标准可以成为质量控制体系的一个完整部分。盲标感官消费者检验结果可能需要与相关概念的市场研究结果相对照。感官评价部门甚至可以在对索赔的调查与广告质询方面与相关的法律部门之间产生相互影响，感官评价也在研究以外的条件下运作。关于食品和材料及其性质和处理的学术研究，经常要求感官检验来评价人们对于产品变化的感受 (Lawless 和 Klein, 1989)。在学术范

围内，感官科学家的一个重要作用是提供咨询和解决方法，以保证质量检验可以由那些试图理解所研究的其他各种因素的感官效应的研究者和学生所实施。在政府服务部门（如食品检验）感官评价起着关键作用（York, 1995）。为保证检验方法反映了目前对于感官功能和检验设计的了解，感官原理和适当的训练可能是至关重要的。

1.2 历史的里程碑及三类检验方法

人的感官用来评价食品的质量已经数个世纪了。对于我们随时享用的食品，我们都有判别标准（“每个人都有他自己的品味，无论他走到哪里，都会用它来使自己快乐。”——Henry Adams, 1918）。这并不意味着所有的判断都是有用的，或者任何人都有资格参与一项感官检验。高品质食品的生产经常取决于某一专家感官的敏锐性，他负责生产，或者有权改变生产工艺以确保产品具有满意的品质。这是啤酒酿造师、葡萄酒品尝师、乳品检验员和其他作为质量仲裁人员的食品检验人员的历史传统。现代感官评价代替了那些单一的权威方法，即评价小组参与的采用预定实验方式的特定的检验方法。有几个原因会导致这一改变的发生。首先是认识到品评小组的判断通常比个人的判断更可信，能承担更小的风险，因为某个专家可能会生病、旅行、退休、去世或是其他原因而不能作出判断。代替个人判断并不是个小问题。其次，专家并不一定能反映消费者或消费群体的一部分对产品的需要。因此，对于产品质量和整体要求，直接面向目标人群是较为安全的（尽管这经常要花费更多的时间和金钱）。虽然由于非正规传统的影响，在一些行业中仍采用像“一刀切”的定性检验，但已逐步被更正规的、定量的和控制观测方法所代替（Stone 和 Sidel, 1995）。

当前的感官评价方法包括一整套在工业和学术研究中采用的确定的痕量记录检测技术。不幸的是，对于不同方法的灵敏度和可信度的对比研究几乎没有开展（Meiselman, 1995）。然而，我们所考虑的许多标准操作是针对于感官评价专业人员近 50 年的食品和消费产品研究的实践中所遇到的缺陷和问题，这一经历是相当重要的。任何感官评价专业人员最关心的是确保检验方法适合于回答在检验产品中所提出的问题。基于这一原因，检验通常根据其主要目的和适当的用途进行分类。一般采用三种典型的感官检验，每一种都有不同的目标，选择参与者时采用不同的标准。这三种主要的检验类型见表 1.1。

表 1.1 感官评价检验方法分类

类别	关键问题	检验类型	小组成员特征
区别	产品是否在任何方面均不同？	“分析”	按感官敏锐性挑选，检验方法经指导，有时经过训练
描述	产品在一定的感官特性方面如何不同？	“分析”	按感官敏锐性和动机挑选，经训练或较高级训练
情感	对产品的喜爱程度或更喜爱何种产品？	“快感”	按产品用途挑选，未经训练

最简单的感官检验仅仅试图回答两种类型产品间是否存在不同。方法有区别检验法或简单的差别检验法。分析是基于频率和比率的统计学原理（计算正确和错误的答案）。从检验结果中，我们可以根据从一系列相似或对照产品中，能够正确挑选出检验产品的

准来表示这些特点的强度并排出顺序。这一进步在一些领域中是令人瞩目的。

描述分析技术中的一些变化和不断改进正不断涌现。通用食品技术中心的一个组织于 20 世纪 60 年代早期开发和精练了一种方法以量化食品质量，许多像风味剖面等方法已经能够量化风味特征（Brandt 等，1963；Szczesniak 等，1975）。质地剖面法这一技术采用一套固定的力相关和形相关的特性来表述食品的流变学和触觉特性以及咀嚼时时间是如何变化的。这些特性与食品磨碎和流体评价相对应。例如，硬的感觉与穿透样品所要求的物理力有关。流体或半固体的黏稠的感觉与物理黏度有部分关系。利用标准产品或者作为标准的规范的模拟食品，质地剖面品评小组成员也要经过训练以体会每一范围的特殊强度点。

其他用于描述分析问题的方法也得到发展。在 20 世纪 70 年代早期的斯坦福研究院（Stanford Research Institute），一个小组提出一种描述分析方法以弥补风味剖面法的一些明显的缺点，这一方法甚至对食品的所有感官特性有更广泛的应用性，不仅仅是口感和质地（Stone 等，1974）。这一方法称为定量描述分析法或缩写为 QDA[®]法（Stone 和 Sidel，1995）。QDA 法大部分借鉴了行为研究的传统，采用实验设计和统计分析，如变量分析等。相对于风味剖面法的小组讨论和集体意见，该方法保证了小组成员的独立判断和统计检验。其他一些描述方法也经尝试并得到一定的普及，如光谱法（Spectrum[™]，Meilgaard 等，1991），该方法对于小组成员的强度范围点有很高的标准，比较接近于质地剖面法。还有一些研究者运用了混合技术，包含了各种不同描述方法的一些特点（Einstein，1991）。由于各种方法的优点可能适用于一个特定公司的产品和资源，今天许多产品开发部门采用混合的方法。

描述分析法已被证明是最全面、信息量最大的感官评价工具，它适用于表述各种产品的变化和食品开发中的研究问题。它所得到的信息与消费者赞同的信息有关，与利用统计技术（如回归和相关等）的仪器测量有关。

一种小甜饼的质地评价描述调查表的样本如表 1.2 所示。该产品以不同的时间间隔，按统一的、可控制的方式，一种典型的分析感官检验法进行评价。例如，第一口可以定义为门牙的切割。进行类似分析的品评小组可以由 10~20 名经过良好训练的人员组成，他们已了解术语的含义并用样品练习过。借助于一些衡量点为例，用一些方法来表示强度。注意本例中所提供的详细信息量，记住这仅仅是考察了产品的质地。风味也可以构成一个同样详细的感官分析，也许只是用另一个训练过的品评小组。小组成员的人数相对较少（12 人左右）是有理由的，因为他们标准较高。由于他们经过训练以相似的方式确定品质标准，降低了误差波动，可以保持统计检验力和检验的敏感性，尽管观测次数较少（每个产品数据点较少）。Meilgaard 等人（1991）提供了质地、风味、香味和触觉分析的类似实例。

第三类感官检验方法主要是试图对产品的好恶程度量化的方法，称作快感或情感检验法。对于这一问题最直截了当的办法是提供给人们两种产品以供选择，然后从大多数反馈中判断是否存在一个明显的倾向。选择检验的问题是不能从反馈信息中得到很多关于好恶程度的信息。20 世纪 40 年代末期于美国陆军军需食品与容器研究所开发的快感准则是此类检验的一个历史性的里程碑（Jones 等，1955）。该方法对喜好进行均衡的 9 点

划分，中间是中性范围，尝试用代表在欢乐氛围中相同心理阶段或变化的副词作为各刻度点的标记。换句话说，它是一个带有标尺性质的刻度，有相等的间隔以适用于统计分析。

表 1.2 小甜饼的描述评价——质地特征

阶段	特征	词语定位
表面	粗糙	光滑—粗糙
	微粒感	无—多
	干燥	有油—干燥
第一口	脆性	松脆—硬脆
	硬度	软—硬
	微粒大小	小—大
初次咀嚼	黏稠度	稀薄—浓厚
	口感的均匀性	均匀—不均匀
咀嚼	吸湿性	无—强
	黏合性	松散—黏合
	黏牙性	无—多
	颗粒疏松性	无—强
残留	油质	干燥—油质
	微粒	无—多
	白垩状	无—多

9点标度的样例如图 1.2 所示。今天一项典型的快感检验包括 75~150 名消费者作为样本，他们是某种产品的经常性用户。该项检验包含可供选择的对产品的几种看法，以便于在集中场所或感官检验较方便的场所进行。较大规模的情感检验也可进行，由于个人偏爱具有较大的可变性，因此，需要增加检验人数来弥补，以确保统计检验力和检验的敏感性。这同时也提供了一个机会来寻找喜爱不同产品风格的人的比例，如不同的颜色或风味等，也提供了探明对某种产品好恶原因的重要信息的机会。

食品工业的工作者曾经很偶然地接触了感官研究而且开发了评估感官技术的心理学家 (Moskowitz, 1983)。9 点快感标度的开发可以作为一个能够认识到的实验心理学家和

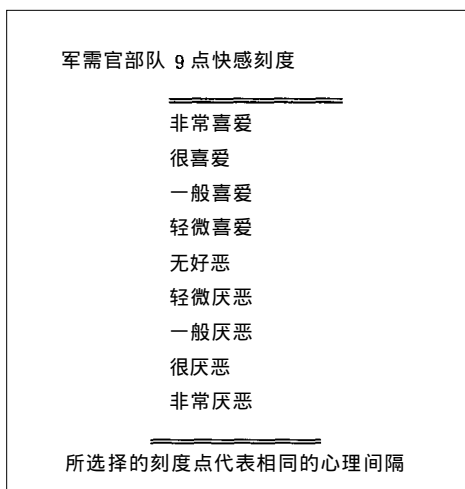


图 1.2 用于评估好恶的 9 点快感刻度
该刻度最初开发于美国陆军军需食品与容器研究所 (军需官部队)，现已广泛应用于食品的消费检验。

食品科学家之间相互影响的很好的例证。一项被称作塞斯通标度（量表衡量）（见第 5 章）的心理学测量技术被用来确认作为 9 点快感标度标记的副词，这一相互影响在本书作者身上也可以体现。本书一位作者从事食品科学和微生物学，而另一位作者是一名实验心理学家。相互影响的发生不应使我们惊奇，也许惟一使人迷惑的是为什么这样的交流没有进一步地持续和扩大。语言、目标和实验注重点的不同可能会产生一些困难。心理学家主要的研究对象是人的个体，而感官评价专业人员则主要对食品（刺激物）进行研究。然而，由于感官的感知包含了人与刺激物间必然的相互作用，类似的检验方法很明显对于表述这种人-产品的相互作用是必然的。

1.3 中心原则——分析与快感检验

对于所有感官评价，中心原则是检验方法应该与检验目的相适应。图 1.3 所表示的是如何从关于检验目的的问题提出到检验方法的选择这一流程。为实现这一目标，有必要在感官检验管理人员和客户或信息的最终用户间保持清楚的信息联络。对话经常是需要的。产品间是否存在根本的差别是重要问题吗？如果是，区别检验予以证明。与老产品相比，消费者是否更喜欢新产品？则需要进行消费者认同检验。我们要知道新产品的哪种感官特性发生了改变？需要进行描述分析检验。有时有多个检验目标，就需要有一系列的不同检验（Lawless 和 Claassen, 1993）。在有竞争力的产品开发期间，如果所有的问题都要求立即或者在一定的时间内回答，这种形式能够发现问题。食品工业中感官专业人员的最重要的工作之一是保证对最终用户所需要的信息类别有一个清晰的理解和详细的说明。检验设计可能要求与不同的人进行大量的对话、交谈，甚至要记录检验要求，以便说明信息被收集的原因以及结果如何被用于特定的决策及随后应采取的措施。感官专业人员处于最有利的地位，能够了解各种检验的用途和局限，并从数据中得到适当或不适当的结论。

对于这一原则有两个重要的推论。感官检验的设计不仅包括适当方法的选择，也包括对合适的参与者和统计方法的选择。三类统计方法可以分为两种形式：分析感官检验（包括区别和描述方法）和情感或快感检验 [如那些评估消费者喜好或偏爱的方法（Lawless 和 Claassen, 1993）]。对于分析检验，品评员的选择要求是对于所要评价的产品的主要特性（口味、香气、质地等）总体上要有很好的感官灵敏度。他们熟悉检验程序，根据检验方法可能或多或少要接受一些训练。在描述分析时，他们采用一种思路分析框架，主要针对问卷所限定范围的产品的一些特定方面。他们被要求将个人倾向和快感影响排除在外，因为他们的工作仅仅是详细描述产品表现何种特性以及感官的强度、广度、数量或持续时间达到何种水平。

与这种思路分析框架相比，消费者在情感检验中的行为表现得更为一致，他们将产品作为一个整体来感受。虽然他们的注意力有时被产品的某一特殊方面所吸引（特别是当产品腐败、令人失望或不满意时），他们对于产品的反映经常是直截了当的，根据产品对感官刺激的整体印象表达他们的好恶。这并不需要对产品特性情况进行很多思考和剖析，也就是说，消费者根据他们整体的感知有效地表达了他们的印象。在这样的消费者

检验中，必须仔细地选择参与者，以保证所得的结果概括了公众的爱好。参与者应该是产品的经常性用户，因为他们最可能形成目标市场，同时也熟悉类似的产品。他们有合理的期望值和一个参考的框架，在这个框架内他们对于其他所尝试的类似的产品能形成自己的看法。

分析/快感的特点对一些非常重要的、但又简单粗糙的方法和一些关于检验方法与反馈相称的问题会提出警告。向有经验的品评员询问他们对产品的偏爱或者是否喜欢一种产品都是不明智的做法。他们被要求用一种不同的、更具有分析性的思路框架，而将个人爱好排除在外。而且，他们没有必要作为产品的经常性用户而被选中，所以，他们不是用以概括快感检验结果的目标人群的一部分。一个普通的类比就是一架分析仪器。你不会去问一台气相色谱仪或是一个 pH 计它是否喜欢一个产品，所以，为什么要去问你的分析描述品评小组呢？（O'Mahony, 1979）。

相反，当消费者被要求提供对于产品品质非常专业的信息时会发生问题。消费者的行为不仅不会在一个分析的思路框架内，而且他们对于产品的特定品质经常会有非常模糊的概念，例如混淆了酸味和苦味。每个人对于问卷中感官属性词汇的解释经常会有显著的不同。一个有经验的质地剖面品评小组对于一种产品咀嚼后的黏着程度达成一致不会有问题，而我们不能期望消费者也能提供关于这种特定技术性品质的建议性的信息。总之，我们避免将有经验的品评员用于情感信息检验，也不要要求消费者做专业的品质分析。

分析/快感的特点涉及到一个问题，实验的控制和精度是否应达到最大值，或者对于现实世界的正确性和普遍性是否关键。这两者之间经常交替使用，很难同时达到最大值。实验室的分析检验利用特殊筛选和训练的评判员，与消费者检验相比结果更可信，随机误差更低。然而，由于使用人工条件和一个特殊的参与者小组，我们在一定程度上放弃

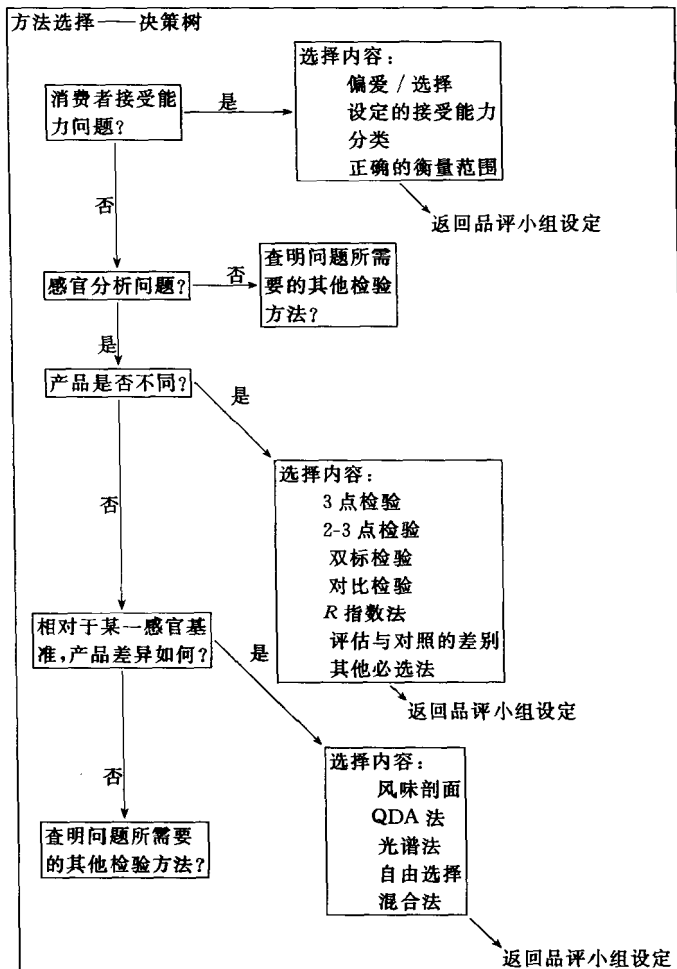


图 1.3 方法确定流程图

根据主要目的和研究问题，选择不同的感官检验方法。类似的决策程序也用于对品评员的选择、反应范围的设定、实验设计、统计方法的选择以及设计感官检验中的一些其他工作。

了现实世界结果的普遍性。相反，在消费者家中进行的消费者产品检验，我们有许多现实生活的正确性，但也有许多数据干扰。Brinberg 和 McGrath (1985) 把精确性与正确性之间的矛盾称为“不可兼得”。O'Mahony (1988) 将感官评价类型 I 和类型 II 进行了划分。感官评价类型 I 中，可靠性和灵敏性是关键因素，参与者更像是被看做一台分析仪器用来分析检测食品的变化。感官评价类型 II 中，挑选的参与者代表了消费群体，他们可以在自然条件下评价食品。这里要强调的是消费者感知的反馈。任何一种感官检验在有些场合会失败，在这些场合中可靠性对现实生活的推断存在一种潜在的交替侧重的关系。这一因素也必须与数据的最终用户讨论，以便了解他们的重点位于何处，他们可以接受何种水平的交替侧重。

统计分析的选择也必须着眼于数据的特点。由于区别检验包括对正确反馈数据的选择和计算，来源于二项式分配或以类似 χ^2 设计的比例统计是适当的。相反，对于大范围的数据，我们可以应用适用于类似一般分布和连续数据的参数统计，如平均数、标准偏差、*t* 检验、方差分析等。选择一个适当的统计检验并不总是直截了当的，所以，感官专业人员应该有足够的智力以接受详细的统计学训练，并且应该知道在一项复杂计划编排的早期阶段就将统计和设计专业人员安排在内。

偶尔也会违反这些原则，不应该为了节省开支或成本而经过一个合理的分析将它们排除在外。一个普通的例子是在消费者认同的接受性检验之前，进行一次区别检验。尽管我们最终关心的是消费者是否喜爱一种新产品，我们可以进行一次简单的差别检验以明确是否有变化被察觉到。这一次序的逻辑如下：如果一个经过筛选的有经验的区别评价小组在感官实验室仔细控制的条件下不能分辨出差别，那么一个更加不统一的消费者群体在他们缺乏控制、更加变化不定的世界里是不可能察觉这一差别的。如果没有察觉到差别，逻辑上不可能存在整体的偏爱。所以，先进行一次比较简单但又比较灵敏的区别检验有时可能避免进行一次耗时更多、花费更大的消费者检验。控制下的区别检验，额外的可靠性为得到关于消费者感受的结论提供了一张“安全网”。当然，这一逻辑也不是没有它的缺陷：一些消费者可能在家庭试用期间与产品有广泛的相互影响，可能形成一些在短期的实验室检验中无法得到的稳定而重要的看法，得到错误的或否定的结果的可能性也总是存在的（误差可能会遗漏某种差别）。MacCrae 和 Geelhoed (1992) 描述了一件有趣的事情，在一项 3 点检验中差别被遗漏，导致了在对比检验中两个水样品间得到了一个显著的偏爱。感官专业人员必须知道这些实验结果中的异常情况有时会发生以及它们发生的一些原因。

1.4 应用 为何收集感官数据

人类对于食品和消费产品的感知是复杂的感觉和解释过程的结果。在科学进程发展到这一阶段，对这种由人体神经系统平行传导的多方面刺激的感知是很难或者是不可能由仪器测量可以预知的。在许多情况下，仪器缺乏人的感官系统的灵敏度，嗅觉就是一个很好的例子。仪器很难模拟食品在品尝时的机械处理或是在感受器周围渗透的形式，这类渗透发生在如唾液或黏液等生物流体中，能引起风味物质的化学区分。最重要的是，仪

器评估得到的值忽视了一个重要的知觉过程：人脑在反应之前对感官体验的解释。人脑位于感官输入和反应产生之间，而产生的反应构成了我们所需的数据。这是一架巨大的平行分布处理器和计算引擎，具有模式快速识别的本领。它用于感官评价的工作完全以个人经历和体验的框架为参考。感官体验被解释，在参考的框架内给出含义，相应于期望值给以评价，能够包含多重同步的或连续的输入信息。最后，判断作为我们的数据而输出。这样，有一个“感受链”存在而不是简单的刺激与反馈（Meilgaard等，1991）。

对于消费者在现实生活中可能怎样对食品进行感知和反应，只有人的感官数据提供了最好的模型。我们收集、分析和解释感官数据以形成对于产品在开发计划中是如何改变的预测。在食品和消费产品工业中，这些改变的发生有三个重要因素：配料、工艺和包装。第四个需要考虑的经常是产品的存放方式，也就是它的货架期，但我们也可以考虑货架稳定性作为一种特殊的处理情况，虽然，通常这是一种非常被动的方式（但也应考虑到产品暴露在温度起伏、光催化氧化、微生物污染以及其他一些不良条件下）。配料的改变有许多原因，它们可能被用于提高产品质量或降低生产成本，或者仅仅是因为原料的一定量的供应难以实现。工艺的改变同样是为了试图在感官、营养、微生物稳定性等一系列因素方面提高质量，降低成本或者提高生产能力。包装的改变主要是出于产品稳定性或其他质量因素的考虑，如一定量氧的渗透可以保证新鲜牛肉制品保持鲜红的颜色以提高对消费者视觉上的吸引。包装的作用是作为产品信息和商标图案的载体，所以，产品信息如何由包装材料和印刷图案所携带和展示，感官特性和期望值都可能随之而改变。包装和印刷的油墨可能会在产品取出时使风味和香气发生变化，有时异味也会进入产品中。对于氧化反应、潜在的光催化反应的有害影响、微生物污染和其他一些麻烦，包装也是一道重要的屏障。

人们利用感官检验来研究对这些产品的处理，会使人对观察到的感知产生怎样的变化。在这样的认识下，感官评价是最传统的心理物理学、最古老的科学心理学的分支，它试图详细说明不同能量水平对感觉器官的冲击（心理物理学的物理学部分）与人体反应（心理学部分）之间的关系。我们经常不能准确预测由于在配料、工艺或包装作用下感官将会发生什么样的变化，或者说由于食品和消费产品通常是相当复杂的系统，这么做是非常困难的。风味和香气是由许多挥发性化学物质组成的复杂的混合物所决定的，实验室中非正式的检验可能不会得到解决感官问题的可靠或充分的答案。由于存在一些娱乐、竞争的氛围，以及非标准的照明条件等，开发实验室中的品评台对于判断潜在的感官效应不是一个好地方。最后，产品开发者的鼻子、眼睛和舌头并不能代表其他大多数将购买这种产品的人。所以，关于消费者如何看待一种产品存在一些不确定性，特别是在更加自然的条件下。

这里，不确定性是关键。如果感官检验的结果能被很好的理解，是可预见的话，就没有必要进行正式的感官评价。不幸的是，在工业环境下，一组感官检验中经常要进行一些无用的检验。由于对产品开发程序、公司传统或者仅仅是出于保护自己免受意外失败的责备这一愿望的过度保护，产生了无用的常规检验的负担。然而，感官检验又是对减少所发生的不确定性数量惟一有用的。·如果不存在不确定性，感官检验也就没有必要了。在工业环境中，感官评价提供了一条信息渠道，这对于指导产品开发和产品改变的

商务管理决策是非常有用的。一旦得到了感官信息，这些决策的做出就会基于一个较低的不确定性和较低的风险性。

感官评价也可用作其他目的。对于质量控制（QC）或品质保证中的感官分析，它可能是相当有用甚至是必需的。为适应品评小组这一小型团体，改变传统感官实践是需要的，而且在制造行业中，在线质量控制经常要求快速评估。由于组织品评小组、准备检验样品以及分析报告感官数据需要时间，将感官技术作为在线评估应用于质量控制可能是非常具有挑战性的。感官检验可更方便地适用于包括成品感官评估在内的品质保证，而品质保证可以与货架期评估或质量监控的常规程序结合在一起。通常，希望在感官反应和仪器测量之间建立联系。如果能够很好地实现这一愿望，仪器测量有时就有可能代替感官检验，特别是在需要快速转变的条件下，尤其适合采用这种方法。如果感官评价有可能引起感觉疲劳，且在一些重复评价时具有一定风险（如杀虫剂的香气）的话，那么用仪器来代替检验感官数据可能是非常有用的，而且在意外的感官问题发生和遗漏时，商业风险也不高。

除了这些围绕产品领域的检验，感官研究在较为广泛的相关领域也很有价值。感官检验可帮助我们理解产品的关键特性，及为什么消费者认为该产品被接受并获得成功。当我们注意到消费者使用语言的模糊方式时，消费者感官检验可能会提供一个关于产品优缺点的可供判断的信息。消费者感官评价可能会暗含对进一步调查的设想，比如发现一个开发新产品的机会。

感官分析科学中有一些重复出现的课题和一直未解决的问题。1989年，关于材料和产品感官评价的ASTM委员会E-18出版了感官方法的产生和委员会成立的回顾庆典纪念册（ASTM，1989）。在书中，Joe Kamen，军需食品和容器研究所的早期感官工作者简述了30年前就已很活跃的感官研究的9个领域。在20世纪即将结束之际重新认识感官科学的地位，我们发现这些领域仍然是研究活动的沃土，当前有许多感官实验室与这一活动相呼应。Kamen（1989）划分了以下一些类别：

（1）针对提高可靠性和效率的感官方法研究，包括程序上细节的研究（漂洗等）和不同实验设计的运用 Meiselman（1993）作为美国陆军食品实验室一名感官科学家，提出了许多在感官评价领域仍未解决的方法问题。Meiselman指出，对于诸如可靠性、灵敏度和正确性等测量质量的问题，缺乏方法的研究。许多感官技术都来源于解决实际问题的需要。方法已经成熟，达到了在工业痕量级基础上的标准操作水平，而不是依靠经验数据的其他方法。在一些期刊如《感官研究》杂志和《食品质量和偏爱》中实验性研究报告增加的比例主要在纯方法上的比较，这表明以感官检验为基础的知识的进展，但仍有许多工作要做。

（2）解决问题 Kamen 提出在产品间建立产品等价关系的简单例子，但在工业实践中仍有许多天天发生的与产品有关的问题。要求索赔调查（Gacula，1991）以及法律和广告的质询都是例子。另一个普通的例子是鉴别引起不良风味、“腐败”或其他不令人满意的感官特性的原因以及对分离和鉴别引起这些问题的原因的探查工作。

（3）建立检验规范 这对于厂商和卖方，在多个车间中制造的产品质量控制、具有国际销售能力的产品的开发以及进行复合感官检验的地点和评价小组问题上都可能是很重要的。

(4) 环境和生物化学因素 Kamen 认识到偏爱可能会随条件作用而改变（当你在户外或饥饿时，食品味道经常会好一些）。Meiselman (1993) 提出大量感官研究是否表现在对消费者反应可能更有预见性的现实饮食条件的研究这一问题。

(5) 解决实验室与野外研究间的差别 对于在感官实验室中可靠、详细而又精确的分析方法的研究中，在预测野外检验结果时可能会丧失一些准确性。如果一套完整的检验程序没有执行，或者说在新产品上市之前对检验程序作了一些简化，管理层必须当心得到错误的、被动的或否定结果的潜在可能性。工业中感官评价专业人员不是总有时间来研究实验室和野外检验间的相关水平，但一个谨慎的感官分析计划应该包括对这一问题的定期检查。

(6) 个体差异自 Kamen 的时代以来，越来越多的文献阐明了这一事实，品评成员不是同一的、可互换的测量仪器。每个人的生理特点不同，参照框架不同，集中和保持注意力的能力不同，以及动机源不同。比如在生理上的差异，我们不断有文献报道关于对特定化合物的嗅觉缺失症——“嗅盲”而其他人的其他嗅觉正常 (Wysocki 和 Labows, 1984)。即使对于进行感官评价的经过训练的品评员来说，一些嗅觉特征的差异也是令人惊奇的。

(7) 与产品变化相关的感官差异 这当然是感官科学在工业实践中的实质内容。然而，许多产品开发者并没有将足够的感官专业人员投入到这一根本性的研究问题上。他们也可能陷入没完没了的对比检验中，很少或者根本没有有计划的实验设计，也没有关于基本的物理变化（配料、工艺等）是如何产生感官动态范围变化的模型。物理变化对于感官反应的关系是精神物理学思考的精髓。

(8) 感官的相互作用 食品和消费产品涉及到许多方面。感官科学家对各种感官特性间的相互作用，如增强和掩蔽效应等理解得越来越多，他们除了报告数据和统计意义外，还能更好地解释感官检验结果，提供广泛的意见和合理的结论。

(9) 感官评价 感官数据的最终用户和需要感官检验的人经常希望有一种可以回答所有问题的工具。Kamen 在分析和快感检验间引用了简单的二分法（即差异与偏爱），如何解释这一差异是一项固定的工作。由于在感官科学领域缺乏普遍的训练，感官教育的工作我们今天仍然要做，感官专业人员必须能够解释检验方法背后的基本原理，并向非感官科学家和管理人员传达感官技术的重要性和逻辑性。

1.5 与其他研究方法的区别

对感官结果有效传达的另一挑战是关于其他研究方法产生的数据与感官数据的相似之处。由于一些感官消费者检验与市场研究机构所进行的研究有明显的相似性，问题可能会产生。然而，一些重要的区别还是存在的，如表 1.3 所示。感官检验几乎总是在盲标的基础上进行的。也就是说，产品的身份通常是模糊的，没有提供允许在一个特定的范畴内评价产品的最少信息（如冷的早餐谷物食品）。相反，市场研究检验经常明确表述关于产品的概念：商标、广告图案、营养信息或者可能包含在混合设计中，使产品在概念上具有吸引力的其他信息（例如将注意力吸引在制作的方便性上）。

表 1.3

感官评价消费者检验与市场研究检验的对比

对消费者的感官检验：	市场研究检验（概念—产品检验）：
参与者按不同产品类型的用户来筛选	以对概念有积极反应来挑选产品检验阶段的参与者
采用无标签样品一帶有最少概念信息的随机代码	概念的表述、信息和参考框架都是清晰的
如果感官特性和所有要求与目标一致才能确定	以概念 / 宣传和类似的产品用途确定期望值
根据同类别的相似产品确定期望值	离开概念和期望，不能测量感官要求
不需要对产品概念的反应或要求进行评估	

在感官检验中，所有这些潜在的偏见性的因素都被除去，以便仅仅在感官特性的基础上进行判断。在科学调查的传统中，我们需要分离出各种影响因素的变化（配料、工艺及包装的改变），将感官特性评估为这些变化的作用而不是观念影响的作用。这样做可以将由于复杂的观念而产生的信息的期望值所造成的较大的感知负担的影响降至最低。有许多潜在的反馈偏见和工作要求在“出卖”想法时被继承下来，就像出卖产品一样。参与者经常喜欢迎和检验人，给出他们认为与检验人希望得到的相一致的结果。有许多文献是关于一些影响因素的作用的，诸如商标对消费者反应的影响。产品信息以复杂的方式与消费者的态度和期望相互影响（Cardello和Sawyer, 1992; Shepard等, 1991/1992; Aaron等, 1994; Deliza和MacFie, 1996）。期望值可以引起感官反应向着在一定条件下所期望的方向同化，而在另外的条件下则会表现出相反的作用，当与期望值不相符时则增大差异（Deliza和MacFie, 1996）。包装和商标信息也会影响感官判断（Gacula等, 1986）所以，盲标感官检验与完全概念化的市场研究检验间表面上的相似相当程度上是一种错觉。公司管理层需要注意这一重要区别。

这两种类型的检验提供不同的信息，但都非常重要。感官评价可使产品开发者明白他们是否使感官在产品特性的感知方面与工作目标相一致。这一信息只有当检验方法尽可能地脱离概念上、地位上的影响时才能得到。产品开发者有权知道产品是否与其感官目标相符，正如市场商人需要知道产品在概念上、地位上及广告的综合方面是否与其吸引消费者的目的相一致。如果产品失败，没有这两类信息，那么发展的战略就不会很清楚。

有时，这两类检验也会给出表面上互相矛盾的结果。然而，从来没有发生过这样一种状况，即一种是“对”而另一种是“错”的。它们仅仅是不同的评价类型，分别由不同的参与者实施而已。例如，进行市场研究检验的产品评价可能仅仅是对于那些以前对所涉及的概念表现出积极反应的人群。由于他们很可能是购买者，因此这似乎是合理的。但必须注意到他们的产品评价是在他们已经表现出一定的积极态度之后得到的，而人们喜欢保持内在的一致性。然而，盲标感官消费者检验是以经常性产品用户作为样本的，没有经过概念上兴趣或态度的预先筛选。所以，他们没有必要有相同的样本群，得到不同的结果也不会使任何人惊奇。

第二个与感官评价有表面相似的是采用感官标准的传统的产品等级体系，农产品的等级对于保证消费者对其购买食品的质量标准的趋势有历史性的影响。此类技术已广泛地应用于简单产品如牛奶和黄油（Bodyfelt等, 1988），而一种理想的产品应该能被

广泛地认同，任何不良操作和处理而引起的缺陷会带来众所周知的感官效应。进一步的推动力来自这样一个事实，即竞争可以用于考验正在接受培训的新品评员的判断是否与专家的观点相一致。这在牲畜定级的传统中是很常见的：一个年轻人可以在州级展示会上评判一头牛，并因掌握了与专家所采用的相同的标准和有判断力的眼光而受到奖励。感官检验和质量评价的执行方式有显著的不同，表 1.4 列举了其中的一些方面。

表 1.4 感官评价检验与质量检验的对比

感官检验：	质量检验：
在各自检验中分离快感（喜爱 / 厌恶）与描述信息	用于失败后或在线的操作决策
利用代表性的用户来评估产品的吸引力（喜爱 / 厌恶）	提供检验的质量得分和关于缺陷的诊断性信息
利用经训练的品评员来详细说明品质，但不表明喜爱 / 厌恶	采用极有经验的感官专家的意见
适用于研究支撑	可能只利用一个或极少的几个专家
也适用于新的、工程化的、创新的产品	强调产品知识、潜在的问题和原因
强调决策的统计推论、适当的设计和样本大小	传统的衡量尺度是多方面的，难以适合统计分析
	决策的依据可能是定性的
	适用于标准日用品

日用品定级和检验的传统在当前深加工食品和市场分割的时期有严格的限制。与产品多样性相关的“标准产品”越来越少，如在风味、营养水平（如低脂肪）、方便制作以及陈列在超市货架上的其他可选择方面。一种产品的缺陷也可能会给另一种产品带来市场厚利，就如胶水并没有像带给我们到处都是的“Post-It”的标签一样成功。质量评判的方法并不适合于进行研究计划，这一技术受到了许多科学团体的广泛批评（Pangborn 和 Dunkley, 1964; O'Mahony, 1979; Sidel 等人, 1981; Claassen 和 Lawless, 1992），虽然在工业和农业中仍然有他们的支持者（Bodyfelt 等, 1988）。

质量定级中的缺陷鉴别强调发生的根本原因（如氧化风味），而描述方法则利用更加基本的单一的术语来描述感觉而不是推断原因。对于氧化的风味，描述分析的评价小组可能会运用许多术语（油腻、着色过度、鱼腥味），而氧化作用造成了许多质量上不同的感官效应。与主流感官评价另一个显著的不同是质量评判将所有的质量范畴（大致反映消费者厌恶的）与关于缺陷的诊断性信息结合起来，成为一种仅仅注重产品负面的描述分析。而在主流感官评价中，描述作用和消费者评价被明显区分为两种不同的检验，有不同的反应。单个专家的观点是否能有效地代表消费者的观点在这一历史时期是非常不确定的。

1.6 概要与结论

感官评价包含一系列的检验方法，对于产品的表现形式和很好定义反应、统计方法以及指导解释结果的工作提供了指导方针和确定的技术。三种主要的感官检验针对产品

间存在的所有差异（区别检验），品质的特殊性（描述分析）和对消费者好恶的测量（情感或快感检验）。感官技术的正确应用包括对于检验对象正确的选择方法，这也要求感官专业人员与检验结果的最终用户间有很好的交流。合理选择检验参与者和适当的统计分析方法是感官技术的一部分，就像在区别和描述分析检验中要求实验控制良好并且尽可能地提高检验精度。另一方面，情感检验要求有代表性的产品消费者参加，检验条件应该体现出在实际生活中消费者如何检验产品的质量。

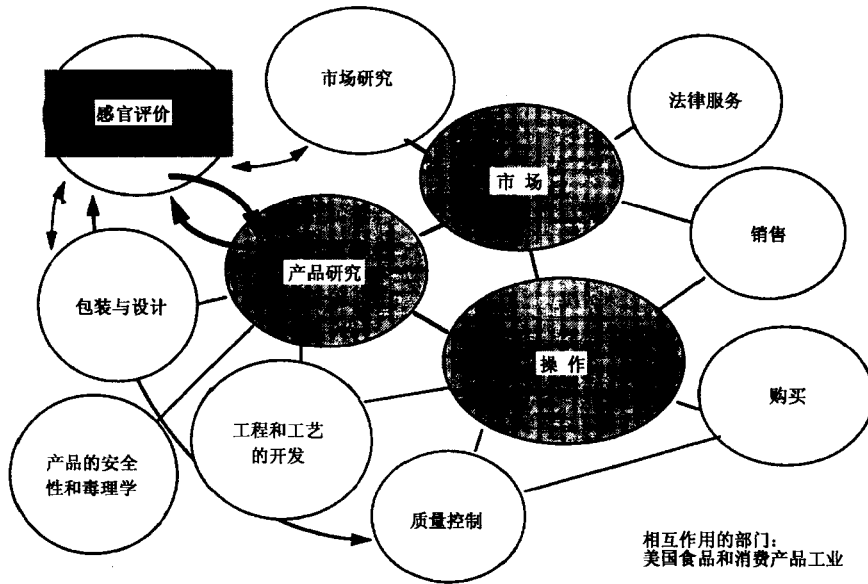


图 1.4 感官评价部门与其他部门的相互关系

一家食品或消费品公司的感官评价部门可能会与许多其他部门相互间发生关系。他们主要是在支持产品研究和开发中发生相互作用，比如市场研究有助于公司取得市场成就。然而，他们也可能与质量控制、市场研究、包装和设计部门、甚至与法律服务部门就一些诸如对索赔的调查与广告的质询等问题发生关系。

感官检验提供了人对于产品由于配料、工艺、包装或货架期的改变而感知的有用信息。感官评价部门与新产品开发部门间的相互影响非常大，对质量控制、市场研究和包装，以及间接的对整个公司的其他部门也提供信息（图 1.4）。感官信息降低了产品开发和满足消费者需要的战略决策的风险。一项功能完备的感官检验计划将对一个公司在满足消费者期望以及确保更大可能的市场成功方面是很有用的。所提供信息的效用直接与感官测量的质量有关。

度量量，量生数，数生称，称生胜。——孙子：《孙子兵法》（卷四 形篇）

参 考 文 献

Aaron, J. I., Mela, D. J., and Evans, R. E. 1994. The influence of attitudes, beliefs and

- label information on perceptions of reduced-fat spread. *Appetite*, 22 (1), 25–38.
- Adams, H. 1918. *The Education of Henry Adams*. Modern Library, New York.
- Amerine, M. A., Pangborn, R. M., and Roessler, E. B. 1965. *Principles of Sensory Evaluation of Food*. Academic, New York.
- ASTM. 1989. Sensory evaluation. *In Celebration of Our Beginnings*. Committee E-18 on Sensory Evaluation of Materials and Products. ASTM, Philadelphia.
- Bett, K. L., and Johnson, P. B. 1996. Challenges of evaluating sensory attributes in the presence of off-flavors. *Journal of Sensory Studies*, 11, 1–17.
- Bodyfelt, F. W., Tobias, J., and Trout, G. M. 1988. *Sensory Evaluation of Dairy Products*. Van Nostrand/AVI, New York.
- Brandt, M. A., Skinner, E. Z., and Coleman, J. A. 1963. Texture profile method. *Journal of Food Science*, 28, 404–409.
- Brinberg, D. and McGrath, J. E., 1985. *Validity and the Research Process*. Sage, Beverly Hills, CA.
- Cardello, A. V., and Sawyer, F. M. 1992. Effects of disconfirmed consumer expectations on food acceptability. *Journal of Sensory Studies*, 7, 253–277.
- Caul, J. F. 1957. The profile method of flavor analysis. *Advances in Food Research*, 7, 1–40.
- Claassen, M., and Lawless, H. T. 1992. Comparison of descriptive terminology systems for sensory evaluation of fluid milk. *Journal of Food Science*, 57, 596–600, 621.
- Deliza, R. and MacFie, H. J. H. 1996. The generation of sensory expectation by external cues and its effect on sensory perception and hedonic ratings: a review. *Journal of Sensory Studies*, 11, 103–128.
- Einstein, M. A. 1991. Descriptive techniques and their hybridization. *In* H. T. Lawless and B. P. Klein, eds. *Sensory Science Theory and Applications in Foods*. Dekker, New York, pp. 317–338.
- Gacula, M. C., Jr. 1991. Claim substantiation for sensory equivalence and superiority. *In* H. T. Lawless and B. P. Klein, eds. *Sensory Science Theory and Applications in Foods*, Dekker, New York, pp. 413–436.
- Rutenbeck, S. K., Campbell, J. F., Giovanni, M. E., Gardze, C. A., and Washam, R. W. 1986. Some sources of bias in consumer testing. *Journal of Sensory Studies*, 1, 175–182.
- Helm, E., and Trolle, B. 1946. Selection of a taste panel. *Wallerstein Laboratory Communications*, 9, 181–194.
- Jones, L. V., Peryam, D. R., and Thurstone, L. L. 1955. Development of a scale for measuring soldier's food preferences. *Food Research*, 20, 512–520.
- Kamen, J. 1989. Observations, reminiscences and chatter. Sensory evaluation. *In Celebration of Our Beginnings*. Committee E-18 on Sensory Evaluation of Materials and