17 Sensory Evaluation in Fruit Product Development

Deborah dos Santos Garruti, Heliofâbia Virginia de Vasconcelos Facundo, Janice Ribeiro Lima, and Andréa Cardoso de Aquino

CONTENTS

17.1 Introduction ........................................................................................................ 416
17.2 Classification of Sensory Methods ................................................................. 417
  17.2.1 Discrimination Tests .................................................................................. 417
  17.2.2 Descriptive Tests ...................................................................................... 419
  17.2.3 Affective Tests .......................................................................................... 421
    17.2.3.1 Preference Tests ............................................................................... 422
    17.2.3.2 Acceptance Tests ............................................................................ 422
    17.2.3.3 Qualitative Affective Tests ............................................................. 425
17.3 Planning Sensory Tests in a Product Development Program ...................... 425
17.4 Developing a Tropical Fruit Nectar: A Case Study ........................................... 430
17.5 Trends in Consumer Research ........................................................................ 432
  17.5.1 Get the Consumer to Describe ................................................................. 432
    17.5.1.1 Flash Profile .................................................................................... 432
    17.5.1.2 Check-All-That-Apply (CATA) ...................................................... 432
    17.5.1.3 Free Listing ..................................................................................... 433
    17.5.1.4 Projective Mapping ......................................................................... 433
    17.5.1.5 Ideal Profiles .................................................................................. 434
  17.5.2 Understanding Consumers ........................................................................ 434
Acknowledgments ................................................................................................. 435
References ................................................................................................................ 435
17.1 INTRODUCTION

Let us not forget that the millions of dollars invested in our businesses depend on that small feeling that our products evoke in our customers’ mouth

(Platt)

We don’t sell products, we sell sensory properties

(Alejandra Muñoz)

Based on these thoughts, we have to agree with Meilgaard et al. (1999) that the primary function of sensory testing is to provide reliable data on which sound decisions may be made. It is an integrated, multidimensional measure with three important advantages: it identifies the presence of notable differences, identifies and quantifies important sensory characteristics in a fast way, and identifies specific problems that cannot be detected by other analytical procedures, as consumer preference, for instance (Nakayama and Wessman, 1979). Comprising a set of techniques for accurate measurement of human responses to foods under minimum potentially biasing effects on consumer perception, sensory analysis attempts to isolate the sensory properties of foods themselves and provides important and useful information to product developers, food scientists, and managers about the sensory characteristics and acceptability of their products (Lawless and Heymann, 1999).

Demands for sensory methodology and technology have grown tremendously around the world, due mainly to the advent of total quality. In addition, the need for understanding people as consumers is something that has been constantly growing and becoming a target of all food industry. Sensory analysis fits into this context as an analytical tool used to translate the link between food products and the consumer, expressing numerical values that can be analyzed and verifying its accuracy through statistical support. Nowadays, most large consumer food companies have departments dedicated to sensory evaluation.

The importance of sensory analysis in the fruit-based food sector is unquestionable, given the variety of applications in food science research, product development, and quality control:

• Improvement of plant varieties and production systems, selection of sources of supply
• Improvement/development of new products and processes
• Product modifications derived from substitution of ingredients and suppliers, changes in processing and packaging, and cost reduction
• Formulation of a product similar to a market leader
• Nutritional enrichment
• Determination of shelf life
• Development of quality standards
• Quality control (raw material and suppliers, processing, end product, packaging)
• Market control (determination of product’s acceptability and consumer preferences, determination of market segmentation)
However, a question arises: why do we have to do sensory analysis? Why cannot we monitor those changes by analytical means? The problem is that sensory quality is not an intrinsic property of a food item. It is the result of an interaction between food and the human being. A particular food has its structural, physical, and chemical properties that determine its sensory characteristics, while man carries its culture and food habits. It is also important to consider his psychological condition when he is analyzing the product, which is influenced by his emotional state and physiological and socioeconomic conditions such as age, sex, education, income, and degree of urbanization among others.

In sum, the sensory quality of a product is the way humans perceive them. And human perceptions are the results of complex processes that involve sensory organs and the brain. It now becomes clear that sensory quality must be measured by sensory techniques. Only human sensory data provide information on how consumers perceive or react to food products in real life. Instrumental measurements are useful only when they show good correlation with sensory data (Schiffman, 1996).

However, when man is used as a measuring instrument, strict control of the conditions of tests application and methodology to be used is required, in order to avoid errors of psychological or physiological nature. The principles and practices of sensory evaluation give strict rules for the preparation, coding, and serving of samples under controlled conditions so that the biasing factors are minimized and use techniques drawn from behavioral science that allow numerical data to be collected and statistically treated, establishing lawful relationships between product characteristics and human perception.

17.2 CLASSIFICATION OF SENSORY METHODS

In sensory evaluation, scientific methods are usually classified according to their primary objective (Table 17.1). Two types of methods are generally recognized by the sensory scientists, analytic methods and affective methods, which comprise three classes of tests: discriminative, descriptive, and affective tests. More detailed discussions and explanations on how to conduct, analyze, and interpret each method are given by Amerine et al. (1965), Moskowitz (1983), Stone and Sidel (1993), Lawless and Heymann (1999), Meilgaard et al. (1999).

17.2.1 DISCRIMINATION TESTS

Discrimination or Discriminative tests answer whether any noticeable difference exists among products. It is possible for two or more samples to be physically or chemically different, but this difference may not be perceived by humans. If the difference among samples is very large and obvious, then discriminative tests are not necessary. For example, use these tests if products resulting from a change in ingredients, processing, packaging, or storage show subtle differences and you want to know if they will be perceptible to people.

Discrimination tests are also called difference tests. Meilgaard et al. (1999) subdivide them into overall and directional difference tests.
1. **Overall difference tests**: Used to check if a significant sensory difference exists between two samples and not in what or how much they are different. Used when no specific attribute(s) can be identified as having been affected. High statistical significant levels do not indicate that the difference is large but only that there is a big chance of a real difference existing. Some of the most used tests in sensory laboratories include the triangle test, duo-trio test, simple difference test, and similarity test, among others.

2. **Directional difference tests**: Will reveal the direction of the difference and which sample has the highest intensity of a particular sensory characteristic. Note that we cannot determine the quantitative measure of those intensities. For instance, we can identify which sample is sweeter, but we do not know if it is a little sweeter or much sweeter. Be aware that a lack of a difference among samples with regard to one attribute does not imply that no overall difference exists. Directional tests can be subdivided according to the number of samples under analysis:
   a. Directional difference between two samples: Paired comparison test or 2-alternative forced choice (2-AFC)
   b. Directional difference among more than two samples: n-alternative forced choice test (n-AFC); ranking test (Friedman analysis); difference-from-control test

---

**TABLE 17.1**

Classification of Traditional Test Methods in Sensory Evaluation

<table>
<thead>
<tr>
<th>Analytic Laboratory Tests</th>
<th>Descriptive</th>
<th>Affective Consumer Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discrimination</strong></td>
<td><strong>How do products differ in specific characteristics?</strong></td>
<td><strong>Which product is preferred?</strong></td>
</tr>
<tr>
<td>Are products different in any way?</td>
<td>Attribute rating (scales)</td>
<td>How well are products liked?</td>
</tr>
<tr>
<td>Simple difference</td>
<td>Time-intensity</td>
<td>How is the product supposed to be?</td>
</tr>
<tr>
<td>Triangle</td>
<td>Quantitative descriptive analysis</td>
<td>Preference</td>
</tr>
<tr>
<td>Duo-trio</td>
<td></td>
<td>Paired preference</td>
</tr>
<tr>
<td>Two-out-of-five</td>
<td>Spectrum</td>
<td>Ranking preference</td>
</tr>
<tr>
<td>A-not-A</td>
<td></td>
<td>Acceptance</td>
</tr>
<tr>
<td>Difference from control</td>
<td>Free-choice profiling</td>
<td>Preference diagnosis (rating)</td>
</tr>
<tr>
<td>Similarity</td>
<td></td>
<td>Just-about-right</td>
</tr>
<tr>
<td>Paired comparisons</td>
<td></td>
<td>Food action scale (FACT)</td>
</tr>
<tr>
<td>n-Alternative forced choice</td>
<td></td>
<td>Purchase intent</td>
</tr>
<tr>
<td>Ranking</td>
<td></td>
<td>Qualitative</td>
</tr>
</tbody>
</table>

17.2.2 Descriptive Tests

The second major class of sensory tests comprises methods that quantify the perceived intensities of the sensory characteristics of a product. These procedures are known as descriptive analyses. All descriptive analysis methods involve the detection and the description of both the qualitative and quantitative sensory aspects of a product. Trained panelists describe the sensory attributes of a sample, often called descriptors. In addition, they rate the intensity of each descriptor to define to what degree it is present in that sample. Meilgaard et al. (1999) explains how two products may contain the same qualitative descriptors, but may differ markedly in the intensity of each, thus resulting in quite different and easily distinctive sensory profiles.

Descriptive analyses are the most sophisticated, comprehensive, and informative sensory evaluation tool. These techniques allow the sensory scientist to obtain complete sensory description of products and help identify underlying ingredient and process variables and other research questions in food product development. The information can be related to consumer acceptance and to instrumental measures by means of statistical techniques such as multivariate regression and correlation (Murray et al., 2001).

Quantitative descriptive analysis or QDA, developed by Stone et al. (1974), is still the most used descriptive method. During several training sessions, the sensory panel is exposed to many possible variations of the product and has the task of generating a set of terms (descriptors) that describe differences among samples. Then, through consensus judges establish definitions for each term and reference standards that should be used to calibrate the intensity scales. However, the actual product evaluation is performed by each judge individually, in booths. Unstructured line scales, anchored with intensity terms, also generated by the panel (e.g., weak–strong) are used, allowing QDA data to be analyzed by both univariate and multivariate statistical techniques: ANOVA of each descriptor, multivariate analysis of variance, principal component analysis (PCA), factor analysis, cluster analysis, and many others. Graphical representation of the data is usually done by radar plots, also known as “cobweb graph” or “star diagram” (Figure 17.1).

Today many product development groups use variations of QDA. The relative simplicity of this technique allows it to be adapted in many different ways; however, any adaptation invalidates the use of the name QDA to describe the procedure.

In free-choice profiling, developed by Williams and Arnold (1984), there is little or no training at all. It allows the panelists to use as many terms as they want to describe the sensory characteristics of a set of samples. The data are analyzed by generalized procrustes analysis (GPA) (Gower, 1975), a multivariate technique that adjusts for the use of different parts of the scale by different panelists and then manipulates the data to combine terms that appear to measure the same characteristic. This technique is very useful in stability studies, where we do not know a priori what sensory characteristics will be developed in the samples, and so we cannot train judges to recognize and measure them. It is also helpful in consumer studies where the objective is to investigate how consumers perceive the products.

Measuring a single descriptor of interest, using scales to express the intensity of a perceived attribute (sweetness, hardness, smoothness, etc.), is also a
The most known scales are category scales, line scales, and magnitude estimation scales, but methods of scaling are under intensive study around the world: cross-modality matching (Stevens and Marks, 1980), labeled magnitude scale (Green et al., 1993, 1996), indirect scaling using Thurstonian model (Baird and Norma, 1978; Frijters et al., 1980; Brockoff and Christensen, 2010), among others.

However, perception of tastes, flavors, and texture in foods is a dynamic phenomenon due to the dynamic nature of processes of breathing, chewing, salivation, swallowing, temperature changes, and tongue movements (Dijksterhuis and Piggott, 2001). By means of conventional scaling methods, panelists can only make a static measurement, which can be a function of an integral of the perception over time, or, more often, a response to the highest intensity perceived. In many cases, this may be the only information required but in other situations it is important to know when the sensation starts, when it reaches the maximum intensity, and how long its duration is. Typical examples are chewing gums and extruded snacks. In the first one, the flavor has to remain as long as possible, and in the other one the flavor needs to “explode” in the mouth and extinguish quickly. Time-intensity (T-I) sensory evaluation provides the opportunity to scale the perceived sensations over time. Today, several commercial sensory analysis software bring T-I scales, but it is possible to develop your own software to collect T-I data. An example is the SCDTI—Sistema de Coleta de Dados Tempo-Intensidade (in Portuguese), which means T-I data collecting system, developed at the State University of Campinas, Brazil.

T-I analyses have been widely used in studies of sensory response to sugar-free foods, since they have to mimic all the sucrose sweetness sensations. Bitterness and astringency in several products have also been investigated by time-intensity analysis.
17.2.3 Affective Tests

Nowadays, it is crucial that the industry understands the consumers’ needs and their desires and expectations about the products. However, to make the consumer describe is not always easy, as people generally have difficulty in clearly describing what they want. It is very important, then, to have clear goals and use simple methods to match the products’ characteristics to the consumer’s expectation.

The sensory tests that assess subjective personal responses of customers toward a product are called affective tests. Affective tests measure attitudes such as acceptance and preference. Preference tests determine the customer’s preference of a product over the other(s). Acceptance tests quantify the degree of liking or disliking.

Whenever an affective test is conducted, a group of subjects must be selected as a sample of the target population, that is, the population to whom the product is intended. There is no sense in testing the acceptability of a product with people who do not like or do not use that kind of product. In addition, formulating a product for elderly people is different from formulating products for teenagers, for example. In the same respect, developing products for consumers in a highly urbanized area may be different from doing it for people in a rural zone.

Affective tests may be designed as in-house panels (in the lab) or as hall tests (conducted at central locations like fairs, supermarkets, etc.), also called consumer tests. As a general rule, one can use in-house panels for most jobs and then calibrate against consumer tests as soon as possible. Some kind of products require more than that—requiring that the product be tested under its normal conditions of use at the consumer’s home (home use tests or home placement tests), where he is an active agent, preparing, serving, proving, and evaluating all aspects of products: package, preparation instructions, sensory attributes, quantity of the portion, and other relevant questions.

Typically, an affective test may involve from 50 to 100 consumers in lab tests, until 300 to 500 in central location and home use tests. The larger size of an affective test arises due to the high variability of individual preferences and thus a need to compensate with increased numbers of people to ensure statistical power and test sensitivity (Lawless and Heymann, 1999).

It is crucial to note that finding no significant preference/acceptance for one sample over another does not mean that there are no perceptible differences among samples. One can equally like both orange and mango juices, but still the orange will be different from mango!

Other usual mistakes in the interpretation of results are, when one sample rates or scores higher preference/acceptance than another, to conclude that “product X was better” than the other. Affective tests do not measure quality. As it was already mentioned, by affective tests we may assess subjective personal responses of a specific group of customers toward a product. Think about a product you like a lot. Let us say it is a fruit candy. Probably it is not the best candy in the world, maybe it is too hard, or too sticky, but it is the one you like best. Then, the right conclusion would be “product X was preferred” or “product X was most accepted” and not “was better” because “better” implies quality judgment.
Common to both preference and acceptance tests is a problem with the univariate analyses of data. There is an implicit assumption that all subjects exhibit the same behavior and that a single value is representative of all subjects. However, individuals’ opinions vary or cluster into similar groups, and if they show opposite opinions about the products, mean values will be similar for all products. Although this would suggest that there was no difference in acceptability among the products, this would clearly not be true. One solution, when working with a large number of products (minimum 6), is to use a multivariate statistical analysis called preference mapping (MacFie and Thomson, 1988). The basic data are collected by a larger number of subjects and then individual differences are not averaged, but are built into the model and play an integral role in the fitting algorithm. In the case described earlier, preference mapping would show each individual response and indicate the different opinions of the two groups very clearly.

Greenhoff and MacFie (1994) explain the two distinct ways of dealing with the data, including case studies: internal preference map (MDPREF), which achieves a multidimensional representation of the stimuli, based only on the acceptance/preference data; and external preference map (PREFMAP), which relates product acceptability to a multidimensional representation of stimuli derived from descriptive analysis or instrumental data.

### 17.2.3.1 Preference Tests

When the objective is to look for the preference of one product or formulation against another, as in product improvement or comparison with a competitive brand, the technique used is the paired preference test, similar to the paired comparison test. Judges receive two coded samples and must choose the sample that is preferred. It is a simple test. Choice is an every-day task for consumers. When the research requires assessing the preference among more than two samples, one can do series of paired preferences, but a ranking preference test is time saving and easier to interpret. This method is also a forced choice, since the participant has to rank several products in either descending or ascending order of preference and is not allowed to have ties in the ranking. The problem with choice tests is that they are not very informative about how well the products were liked by the consumers. If all products are bad, participants will choose the least bad product.

### 17.2.3.2 Acceptance Tests

When it becomes necessary to determine how well the product is liked by consumers, we will have to collect hedonic or attitude responses from consumers using scales. From relative acceptance scores, one can infer preference; the sample with the higher score is preferred. However, not always a single measure of liking and disliking when a food is tasted in isolation represents the real feeling about it. People’s historical preferences may fail to predict their acceptance for certain foods or beverages in an actual tasting (Cardello and Maller, 1982). Context and expectations can affect simple hedonic judgments (Deliza and MacFie, 1996). For this reason, hedonic tests can be complemented by other tests like the ones with appropriateness approach (what judge thinks is a good product) and other behaviorally oriented tests.
17.2.3.2.1 Hedonic Approach

The most popular scale among sensory analysts is the 9-point hedonic scale (Figure 17.2A), developed at the U.S. Army food and Container Institute (Jones et al., 1955; Peryam and Pilgrim, 1957). This method provides a balanced scale for liking with categories labeled with adverbs that represent psychologically equal intervals, with a centered neutral point. Thus, this scale has ruler-like properties, whose equal intervals favor the assignment of numerical values to the responses and the statistical treatment of the data.

However, it has been criticized for a long time for presenting a series of limitations. The main problems are related to the lack of requirements demanded by parametric statistical analyses that are often applied to the data. The 9-point hedonic scale is a category scale, and as such, generates discrete data. It also frequently fails to satisfy the statistical assumptions of normality, homoscedasticity, and additivity required by the Analysis of Variance models (McPherson and Randall, 1985; Pearce et al., 1986; Vie et al., 1991). With a view to overcoming these problems, various authors have proposed alternative methods to generate and statistically analyze sensory data (Miller, 1987; Gay and Mead, 1992; Wilkinson and Yuksel, 1997).

Villanueva et al. (2000) compared the performance of the 9-point hedonic scale and self-adjusting scale (Figure 17.2B), with reference to the discriminative power and statistical assumptions required by the usual ANOVA models, under real consumer test conditions. In a posterior work (Villanueva et al., 2005), the same authors proposed a hybrid hedonic scale (Figure 17.2C). Although the mean values derived from each scale, for each sample, were very similar, the 9-point hedonic scale showed the smallest standard deviation values. However, this scale presented problems with the inequality of sample’s variances (lack of homoscedasticity). The self-adjusting scale presented problems with nonnormality of the residuals. For this reason, the significance levels associated with the $F_{\text{samples}}$ values, for both scales, are only approximate. The authors

![FIGURE 17.2 Example of hedonic scales: (A) 9-point hedonic scale; (B) self-adjusting hedonic scale; (C) hybrid hedonic scale.](image)
suggest that these data should not be analyzed by statistical methods based on assumptions of normality and homoscedasticity but by means of alternative procedures such as generalized linear models (GLM), analysis of categorical data, nonparametric tests, or data transformation for normality. In fact, when proceeding analysis of variance, one should always carry out first a checking of the residuals in order to find out any model inadequacies or violations of the ANOVA model’s assumptions. Thus, when the ANOVA model is appropriate, advantages are taken of the greater simplicity and clarity of interpretation that it provides. By its turn, the hybrid hedonic scale data were shown to be adequate in this diagnosis of the ANOVA model. This scale also presented a slightly superior discriminative power than the other two.

Villanueva and Da Silva (2009) studied the performance of these three scales also regarding the generation of the internal preference map (MDPREF). There were strong similarities among the product spaces, but the MDPREF obtained from the hybrid scale showed a slight superiority over the 9-point hedonic and self-adjusting scales with respect to sample segmentation, consumer segmentation, number of significant dimensions, and the proportion of significantly fitted consumers.

However, in those works, authors mentioned that additional experiments must be carried out to confirm the results so far. In our opinion, differences in the scales’ performance are not big enough to justify any strong endorsement of the hybrid scale nor any condemnation of the traditional 9-point scale. This is also the opinion of Lawless (2010). In fact, we have found the hybrid scale more difficult to use for both consumers and technicians. If the lab facilities do not have software to collect data, it will be necessary to use a ruler to measure each respondent’s evaluation sheet.

Hedonic scaling can also be applied to children and illiterate people. It has used face scales, cartoons characters, or realistic pictures of children or adult faces, representing each one of the verbal points. In many cases, these scales do not perform well. Kroll (1990) showed that verbal descriptors of “good” and “bad,” the so-called P&K scale, worked better for children. Below 4–5 years old, acceptance must be inferred from behaviors, such as oral interviews and ad libidum situations or by the amount of ingestion from a standardized sample.

17.2.3.2.2 Appropriateness Approach
As part of a consumer test, researchers often desire to determine the reasons for any preference or rejection by asking additional questions about the sensory attributes. Meilgaard et al. (1999) used category and line scales to assess the intensity of specific attributes and called it attribute diagnosis. For example, the question would be “How intense is the sweetness of this juice?” And the scale would vary from “very week” to “very strong.” Differently, just-right scales (Vickers, 1988), also known as just-about-right, assess the intensity of an attribute relative to some mental criterion of the subjects for the product that is under analysis. Example: for the same question given earlier about the sweetness of the juice, the scale would vary from “not at all sweet enough” to “much too sweet,” with a middle point corresponding to the ideal sweetness (“just right”).

17.2.3.2.3 Behavioral Approach
The food action scale (FACT) developed by Schutz (1965) is an example of a behavioral approach to assessing food acceptability. It is based on consumers’ attitudes
in relation to the frequency at which they would be willing to consume the product in a given period. Examples of category labels are as follows: “I would eat it every opportunity I had” and “I would eat it only if I were forced to.” FACT scale is recommended for testing products with which consumers are not familiar.

Purchase intent scales are very similar to FACT scales, based on consumers’ attitudes in relation to their willingness to buy the product if it was for sale. Examples of category labels include “I certainly would not buy it”; “I probably would buy it”; and “I certainly would buy it.”

17.2.3.3 Qualitative Affective Tests
Qualitative methods are used to study consumer habits and attitudes that may be useful in predicting the behavior of consumers and to develop the terminology used by them to describe the sensory attributes of the concept or prototype of a product. The main interest of these tests would generate the most varied and possible ideas and reactions on a given product. It is quite useful in product development. In small groups or individual interviews, consumers verbalize their opinions and expectations about the product. The most used techniques so far are focus groups, focus teams, and one-on-one interviews (McQuarrie and McIntyre, 1986; McNeill et al., 2000; Bruseberg and McDonagh-Philp, 2002). In the focus group technique, a group of participants, usually 6–8, sit together for a more or less open-ended discussion about a product or a specific topic. The discussion moderator lets participants introduce themselves and feel comfortable and makes sure that the topics of significance are brought up. To help participants verbalize their needs, interaction among group members is encouraged. The products may or may not be served. The report summarizes what was said and perhaps draws inferences from what was said and left unsaid in the discussion. Generally, the sessions are videotaped.

17.3 PLANNING SENSORY TESTS IN A PRODUCT DEVELOPMENT PROGRAM
Sensory results are only useful when interpreted in the context of hypotheses, background knowledge, and implications for decisions and actions to be taken. Defining the needs of the project is the most important requirement for conducting the right test. Thus, sensory specialists should be full partners with their clients, taking an active role in developing the research program, collaborating on the choosing of the attributes to be analyzed, and setting the experimental designs, which ultimately will be used to answer the questions posed. Only through a process of total involvement they can be in a position to select the most appropriate tests necessary at every point of a research project. They will also be able to contribute interpretations and suggest actions, since they best understand the limitations of tests and what their risks and liabilities may be.

Lawless and Heymann (1999) and Meilgaard et al. (1999) present some guidelines and general steps for choice of techniques (Figure 17.3), reminding us that those rules are generalizations and sometimes the goals, requirements, and resources available in a particular situation will dictate deviations of those principles. Figure 17.3A shows a sensory evaluation flowchart, with an overview of the tasks and decisions in setting up and conducting a sensory test.
1. **Problem definition**: First of all, define the problem. We must define what we want to measure. Start analyzing the nature of the project and what is expected from the samples. For example, is it a new product development; product improvement; ingredients, equipment, or process change; or product matching? Are differences desirable or is the accent on proving that no difference exists between the sample and another or former product? Does the product vary only in one or several attributes? Is there any cue like color, consistency, or others that may introduce sensory biases?

2. **Definition of test objective**: Once the objective of the project is clearly stated, the sensory analyst and the project leader can determine the test objective: overall difference or attribute difference? determine a complete sensory profile? and relative preference or acceptability, etc.?

   However, many projects need a sequence of tests for achieving the goal rather than a single test. It is a sequential decision process as in any other problem-solving activity. Figure 17.3B shows the flow of a sensory evaluating testing program during product development/product improvement projects. First, define exactly what sensory characteristics need improvement or need to be evaluated, then determine that the experimental product is indeed different, and finally confirm that the experimental product is liked better than the control. If working with ingredients, equipment,
Sensory Evaluation in Fruit Product Development

or process changes confirm that no difference exists, and if a difference does exist, determine how consumers view the difference.

a. **Bench-top testing:** This test is mandatory to get familiar with product attributes; choose the ones that will be evaluated and check if sensory variations are obvious or a difference test is needed. All sensory properties should be examined: appearance, aroma, basic tastes, flavor, texture, mouth feelings, aftertastes, and after mouth feelings. Products vary in which attributes? One attribute may influence the analysis of other attributes?

b. **Difference tests:** Although our ultimate interest may lie in whether consumers will like or dislike a new product variation, we must conduct a simple difference test first to see whether any change is perceivable at all. The logic in this sequence is the following: If a screened and experienced discrimination panel cannot tell the difference under carefully controlled conditions in the sensory lab, then a more heterogeneous group of consumers is unlikely to see a difference in their less controlled and more variable world. If no difference is perceived, there can logically be no systematic preference. So a more time-consuming and costly consumer test can sometimes be avoided by conducting a simpler but more sensitive discrimination test first.

c. **Descriptive tests:** If a difference exists, it is usually necessary to know the sensory profile of prototypes or monitoring specific attributes in order to evaluate the effect of process or ingredient variations. In this case, descriptive tests should be applied. This sensory profile will also be useful in determining the specific reasons as to why a product is preferred over the others.

d. **Affective tests: Pilot consumer panels.** Consumer sensory evaluation is usually performed toward the end of the product development or reformulation cycle. At this time, the alternative product prototypes have usually been narrowed down to a manageable subset through the use of analytical sensory tests. With in-house panels explore degrees of liking/disliking and identify potential problems for rework. In general, a product tested under true conditions will not give the same result as a product tested by pilot panels. However, testing in pilot panels will help to select samples that should be evaluated in real conditions tests, which are time consuming and involve higher cost and logistics.

e. **Affective tests: Central location panels.** Central location evaluation is used to answer the question: Is what we find in pilot panels tests the same as what would be found by real-world consumers? The test also explores degrees of liking/disliking but requires time and expenses and is usually performed by companies that will actually introduce the product in the market. For scientific purposes, answers from pilot consumer panels are found to be sufficient.

f. **Determination of shelf life:** The sensory testing sequence usually ends with a stability study, performed to know how product quality changes during storage and allows establishing shelf life. Shelf life
can be defined as the period between manufacture and retail purchase of a food product during which the product is of satisfactory quality (Dethmers, 1979); in other words, the length of time for the product to become unacceptable for sale. A practical use for stability studies is open dating of foods. Open dates are placed on the labels of food products to help consumers in the purchase decision (Labuza, 1982).

Sensory analysis is as important as chemical, physical, and microbiological analyses because it indicates whether the storage products still conform to standards for appearance, aroma, flavor, taste, texture, and functionality. Besides, sensory analysis indicates whether the food is still acceptable to the consumers.

3. Methods selection
   a. Discriminative testing: The issue here is the kind of difference you want to investigate and the number of samples, as stated earlier. If the objective is to check the overall difference between two samples, triangle test is the most sensitive test, because the probability the panelist has to check the odd sample by chance is one-third while in paired comparison and duo-trio tests it is one-half. However, if the samples are complex in any way use duo-trio test, since the task of comparing samples to a standard is easier to panelists. Lawless and Heymann (1999) say that duo-trio test is also more sensitive than the triangle when subjects are familiar with the reference material. For example, if the reference is a product with a long company history and a great deal of ongoing evaluation, deviations from this familiar item may be readily noticed. If the problem relies on a single attribute, forced-choice tests will be more sensitive, since panelists will focus in the intensity of that specific attribute, even if there are other sources of variation.

   b. Descriptive testing: When differences in the target attribute are obvious among samples, or you need to measure the difference in more than one attribute, use direct scaling. However, you must follow the descriptive test methodology, since humans are not good absolute measuring instruments and need reference samples of what is a low intensity and what is a high intensity of those attributes in that specific product. Panelists will calibrate themselves according to that frame of references and will be able to become a good relative measuring instrument.

   Choose QDA or one of its free adaptations when a complete sensory profile of the samples is needed, in order to specify the nature of all sensory changes or differences in a set of samples. Sensory profiles can also be correlated to affective data to investigate the reasons for people’s likings and disliking about a set of products.

   c. Affective testing: Affective tests bring the consumer into the product development. Choose acceptance test to determine how well the product is liked by consumers. For example, how nectar blends from various fruits will be acceptable to consumers who are used to drinking
traditional one-fruit nectar. Choose preference tests to determine preference of one product against another. For example, to decide which sweetener is preferable when formulating a sugar-free fruit jam.

d. **Shelf-life tests:** The selection of a particular sensory evaluation procedure for evaluating products in storage is determined by the test purpose. Acceptability assessment by untrained panels is essential for open dating. Discriminative tests with trained panels are useful when one characteristic is more important than others or when some degradation problems are already expected. Descriptive analysis can be used with new products, when there is no information on the behavior of the product under storage.

Depending on the food characteristics, several failure criteria can be used to terminate a shelf-life study (Labuza and Schmidl, 1988). Besides microbiological growth and physical changes, there are many sensory criteria to determine the end of the test (Gacula, 1975):

i. An increase or decrease in x number of units in a mean panel score. For example, in shelf-life dating of oils, an increase in 2 units of oxidation flavor by a trained panel may determine the end of the test.

ii. Failure time, when a sample reaches an average panel score. This criterion is useful when overall acceptance is used in the stability test. For example, when a 4.5 score is obtained in a 9-point scale for a stored juice.

iii. Just noticeable difference, when the difference between the quality of samples in test and control samples can be detected by trained panels. In this case, a control that can be maintained with only negligible change over time is essential.

iv. Results of a profile descriptive analysis. Comparison between profiles before and after storage may be used to indicate changes in important characteristics.

A question in stability experiments is whether we really need to know the true end of shelf life of a product. Usually, the answer is negative. Most of the time, the producers need the assurance that the product will be acceptable if it is held in the distribution system for a given period of time at certain temperature and humidity. For those situations, stability tests can be planned for a specific time, shorter than that necessary for establishing shelf life. Later, as a complementary data, shelf life can be determined by regression techniques or any other statistical approach.

Another question is whether storing of a food product leads only to deterioration. Actually, some food products require controlled aging to develop characteristic aroma, flavor, and texture. The most known examples are the aging process for wine and cheese. In those products, sensory analysis can also be used for monitoring the product’s changes.
4. **Panel selection:** Each kind of test will require different types of panelists.
   a. **Discriminative tests:** It is mandatory that subjects are screened for normal sensory acuity, especially regarding vision, smell, and taste, and for their ability to detect differences among similar products with ingredient or processing variables. Discrimination panels can receive a little training, but they are usually only oriented in the test method.
   
b. **Descriptive tests:** Panelists are screened for normal sensory acuity, discriminative ability, and motivation. Descriptive panels are trained, except in free-choice profiling. They are asked to put personal hedonic reactions aside, as their job is only to specify what attributes are present in the product and at what levels of intensity, extent, amount, or duration. It has been a central principle in sensory analysis that you should not rely on consumers for accurate descriptive information. Consumers not only act in a nonanalytical frame of mind but also often have fuzzy concepts about specific attributes, confusing, sour, and bitter, for example. However, in recent years, some researchers are breaking this paradigm, as we are going to see ahead in this chapter.

c. **Affective tests:** Participants must be chosen carefully to ensure that the results generalize to the population of interest. They can be recruited among users or at least potential users of the new product, who are frequent users of similar products. They possess reasonable expectations and a frame of reference within which they can form an opinion relative to other similar products they have tried. Never use a panelist who has been trained to evaluate that particular product as a consumer, even though he or she uses to consume that product. During training, these panelists have been asked to assume an analytical frame of mind and will not be able anymore to look at that product in an integrative form. Consumers’ reactions, on the other hand, are often immediate and based on the integrated pattern, although their attention is sometimes captured by a specific aspect.

5. **Experimental design**
6. **Conducting the experiment**
7. **Statistical analyses**
8. **Reporting results**

These last four steps involve many details and particularities that are not covered in this book, though they must be discussed with the sensory specialists.

### 17.4 DEVELOPING A TROPICAL FRUIT NECTAR: A CASE STUDY

Next, we see how sensory tests were conducted in a new-product developing project, helping to make important decisions.

Bacuri, a Brazilian tropical fruit, has a pulp with a distinct, strong, acid-sweet, and agreeable flavor (Clement and Venturieri, 1990). However, it shows very high consistency, which hampers the industrial processing steps such as filtration and concentration. In addition, it has not been possible yet to elaborate bacuri nectars within
Brazilian legislation requirements: 30% fruit pulp or at least 20% pulp for fruits with high consistency (Brasil, 2003). Nazaré (2000) and Silva et al. (2007) observed good acceptability only for the nectar formulated with 10%–12% pulp. One way to reduce the pulp consistency is through the technology of enzymatic maceration.

At first, pulp was macerated with a pectinase (P1), and nectars with 20% and 30% macerated pulp were formulated and compared with the control nectar (10% pulp without maceration) in a bench-top testing. The formulation with 30% macerated pulp proved to be too consistent and was discarded. Nectar with 20% macerated pulp was still more consistent than the control and also too acidic.

In the next step, we tried to correct the high acidity adding sugar. We used an in-house panel to make hedonic evaluations, since in this case it mattered to know how sweet people liked the nectar to be. We found that it was necessary to add too much sugar to bring the nectar to a good level of acceptability.

With those results, we turned to test other enzyme preparations, searching for one that could reduce consistency and did not yield so much acid. We tested two pectinases combined (P1 + P2) in two different proportions, and each one of these combined to a cellulase (P1 + C and P2 + C). A difference-from-control test was performed with nectars elaborated with 20% macerated pulps, where the control sample was nectar from P1 macerated pulp, in order to see if those formulations could reduce the consistency and acidity observed for this nectar on the bench-top testing. Only formulations containing cellulase showed reduced consistency and acidity.

A central composite design was used to determine the best combination and concentrations of pectinases and cellulase. Dependent variables were total acidity, consistency, and sugar content, all determined by means of chemical and instrumental analyses. The best performances were observed for P1 + C (1:2) and P2 + C (1:2).

In order to investigate whether the maceration process could modify sensory properties of bacuri nectars, a descriptive profile was determined for products formulated by the optimized enzyme preparations and compared with the control sample (10% nonmacerated pulp). Figure 17.1 presents the star diagram. Both enzyme preparations produced nectars with similar sensory profiles, but very different from control sample. Macerated nectars reached the same consistency of 10% nonmacerated sample but were still more acid. The characteristic aroma of bacuri was enhanced, and panelists perceived stronger green and pungent aromas. However, they were darker in color and showed more dark spots and lumps. Between macerated nectars, enzyme P1 caused less lump formation, but the nectar was more acid and astringent than nectar macerated by P2.

The final step was to evaluate the new products’ acceptance by consumers and compare it with the well-established 10% pulp nectar. Hedonic tests were performed in two Brazilian regions: the North region, where bacuri is well known and appreciated, and the Northeastern region, where this fruit is almost unknown. Global acceptance and attribute acceptance were evaluated. Consumers from both regions accepted the macerated samples as much as the control nectar, but in the North region, the scores were higher, as expected. Sensory results indicated that this bacuri nectar has a great potential in the market, even in regions where consumers are not used to this fruit, and the enzymatic maceration studies should continue until an optimized product is made.
17.5 TRENDS IN CONSUMER RESEARCH

In order to put the best possible product on the market or even achieve the “ideal product,” it is essential to understand consumer product perception and preferences and relate hedonic responses to sensory product specifications (Worsch et al., 2010). New trends in sensory analysis have been related to two great challenges for consumer researchers: (1) get the consumer to describe food properties and (2) understand consumers’ reactions (Kleef et al., 2005; Hough, 2010).

17.5.1 GET THE CONSUMER TO DESCRIBE

In sensory analysis, the classical approach has been to use a trained panel for sensory description of products and consumers only for hedonic evaluations. However, descriptive methods can be expensive and time consuming (both in terms of panelist training and testing time). In addition, it is hard to maintain a trained panel, especially in the industry, where there is a high turnover of employees. In recent years, researchers have been breaking the paradigm and advocating the use of consumers to generate sensory profiling to lead product development. Husson and Pàges (2003) showed that consumers meet the requirements of discrimination, consensus, and reproducibility, Worch et al. (2009) found no significant differences between products profiled by trained or consumer panels. Already known methods like Kelly repertory grid (Kelly, 1955) and more recent methods have been presented like flash profile (FP), check-all-that-apply (CATA), projective mapping techniques, free listing, and ideal profile, among others.

17.5.1.1 Flash Profile

Flash profile (Dairou and Sieffermann, 2002) is adapted from free-choice profiling, where untrained subjects select their own terms to describe and evaluate a set of products simultaneously. The difference is that subjects rank the products on an ordinal scale for each term they individually created, instead of rating. They are asked to focus on the descriptive terms, not on the hedonic terms. Flash profiling (FP) can also be used at the initial stage of a project to create the sensory attributes for the conventional descriptive analyses (Delarue and Sieffermann, 2004) and shows practical feasibility in the evaluation of a large food product set. The individual sensory maps are treated with general procrustes analysis (GPA) to create a consensus configuration. Cluster analysis can also be performed after the GPA on the descriptive terms to assist in the interpretation.

17.5.1.2 Check-All-That-Apply (CATA)

In the CATA method (Lancaster and Foley, 2007), consumers are asked to check all perceived attributes in a specific product, from a list of prechosen terms (Figure 17.4). The actual generation of terms can be performed in many ways: the consumers can choose words to describe the product during the test (like in free-choice profiling), terms can be generated by consumers not testing the product (i.e., a focus group), or terms can be given by a trained panel (descriptive analysis). CATA method requires minimal instruction; it is relatively easy to perform
and is completed quickly. It is different from scaling since no intensities are given to the attributes. The data are formed by the counts or percentages of consumers that checked each term for each sample. Frequently listed descriptors would be more relevant than those less frequently listed. Data are analyzed by multivariate statistical tools such as correspondence analysis or multiple factor analysis, generating a sensory space similar to PCA. CATA data can also be used for the creation of preference maps correlating hedonic judgment with sensory attributes (Dooley et al., 2010).

### 17.5.1.3 Free Listing

Free listing (Ares et al., 2010; Hough and Ferraris, 2010) is a variation of CATA and is also called open-ended question. Instead of using a list of terms, consumers are asked to use their own terms to describe the samples. Similar words are grouped, and the matrix of frequencies is also analyzed by correspondence or multiple factor analysis. Cluster analysis can reveal associated descriptors, because they are similar stimuli, for example crispy and crunchy, or because they belong to similar categories in the mind of the consumer, like apple and fruity.

### 17.5.1.4 Projective Mapping

Projective mapping (Risvik et al., 1994) studies do not involve numerical judgments. They are holistic approaches of characterizing similarities and differences in sensory attributes of products or assessing their preference or liking. Among the techniques that have been gaining popularity are napping (Pagès, 2005), partial napping (Pfeiffer and Gilbert, 2008), free sorting (Abdi et al., 2007), and a combination of them called sorted napping (Lê et al., 2009). Assessors position the products on a two-dimensional surface (e.g., large sheet of paper) according to overall sensory similarities and differences being free to choose the various criteria used to separate the products (Figure 17.5). They are often asked to enhance the map with descriptive terms for each product or drivers for liking and disliking. Multiple factor analysis provides a quick profile showing relationship between products and descriptors, similar to PCA results from conventional profiling.
17.5.1.5 Ideal Profiles

While the just-about-right (JAR) scale only asks the deviations from ideal for each attribute and product combination, in the ideal profile method (Punter, 2008; Worch et al., 2009), both perceived and ideal intensities are asked directly (the JAR question “is it just right, too much, or too little” is replaced by the question “how strong is it and what would be the ideal strength”). Worch et al. (2010) compared two different methodologies to analyze JAR and ideal profile data and suggested PLS on dummy variables (Xiong and Meullenet, 2006) for the analysis of JAR data, and the Fishbone method for the analysis of Ideal Profile data.

17.5.2 Understanding Consumers

The psychophysical approach of sensory evaluation (sensory scales) is based on the idea that people are rational and can give explicit reasons for their behavior. However, according to Koster (2009), paradigms like uniformity, consistency, objectivity, and conscious choice are fallacies in human behavior. In other words, people are different from one another and you cannot average their behavior; people change their own behavior, evaluations are subjective, and choices are not always rational and conscious. The author considers that the use of scales is efficient but they do not contain all the necessary information.

Sensory science has appropriate psychophysics and marketing techniques to determine consumers’ drivers of liking and preference. They allow us not only to hear what people say they like or they do but to see their real behavior, allowing us to obtain a snapshot of people’s lives, their experiences, and their relationships. It captures the subject’s inner thoughts, feelings and emotions, their values, and rules that guide them, because it is carried out in his or her own environment (Deliza, 2009). Some studies are also reaching, invading, and exploring the domestic environment. To handle this kind of data, some methods have been used like category appraisal, conjoint analysis, experimental design, focus group, free listing, Bayesian networks, laddering empathic design, and information acceleration. Some of these are briefly discussed in the following.

Free listing can also be used to explore behaviors and habits, foods considered appropriate for certain uses or occasions, and feelings related to food consumption. For example, a subject can be asked to list all the things he or she feels while eating a bar of their favorite chocolate. Frequently listed feelings can be useful from a concept development perspective. Associated feelings like, for example, “fattening” associated to “guilty” can help discover why chocolate has a negative image in the minds of some consumers.
Bayesian networks (Craignou and Jouffe, 2008; Craignou and Bezault, 2009), also referred to as belief networks, Bayes nets or causal probabilistic networks are a modern data analysis tool that can handle variability and uncertainty using probability distributions. These techniques can be used for explanation, exploration of information, and prediction of system behaviors and for decision making. Despite their popularity in various fields such as finance, medical diagnosis, robotics, and genetics, their application to food-related problems have only recently emerged. Most current Bayesian network algorithms require discrete variables. Modeling with Bayes nets enables the use of expert knowledge as well as the combination of data from different studies.

Empathic design (Polanyi, 1966; Ulwick, 2002). A multifunctional team is created to observe the actual behavior and environment of consumers. A visual record is made of consumers interacting with their environment. Photographs, videotape, sketches, and notes are tools that make a record of behavior. Data can as well be gathered through responses to questions like “why are you doing that?” Team members have a brainstorming session to transform observations into graphic, visual representations of possible solutions. A nonfunctional, two- or three-dimensional model of a product concept provides a vehicle for further testing among potential consumers.

Information acceleration (Urban et al., 1997). The researcher constructs a virtual buying environment that simulates the information that is available to consumers at the time they make a purchase decision. Respondents are “accelerated” into the future by providing them alternative future environments that are favorable, neutral, or unfavorable toward the new product. In this virtual buying environment, they are allowed to search for information or shop. Measures are taken of respondents’ likelihood of purchase, perceptions, and preferences. Based on these measures, a model is developed to forecast sales and simulate strategy alternatives.

ACKNOWLEDGMENTS

The authors are very grateful to Victor Costa Castro Alves and Idila Araujo for figure designs, Maria Aparecida Azvedo Pereira da Silva, for valuable suggestions and ideas for the text, Gustavo Saavedra Pinto, for making available bacuri nectar results.

REFERENCES

Advances in Fruit Processing Technologies


