

6

Sensory evaluation of milk

K. W. Chapman, Cornell University, USA

Abstract: Powerful results will be achieved when using appropriate sensory tests in the accurate manner, with proper analysis and interpretation of results. This chapter reviews common methods used in the sensory evaluation of milk such as discrimination tests, descriptive analysis, acceptance (hedonistic), preference and threshold testing.

Key words: sensory evaluation, milk, dairy, descriptive analysis, discrimination, acceptance, preference, threshold.

6.1 Introduction: key issues in the sensory evaluation of milk

Quality of processed milk can be evaluated by trained scientists in a variety of ways, such as chemical and microbiological analyses. Consumers, however, will make their own analysis, based on their perceptions of whether this milk tastes good. Therefore, the ultimate evaluation is sensorial and is done by the consumer. Systematic sensory analysis is needed to provide useful information about the human perception of milk. Such sensory evaluation is key to the improvement of product quality and shelf-life, enabling milk to compete with innovative new introductions, as well as with currently popular shelf-stable products. Whenever new procedures are developed to reduce the microbiological growth and biochemical changes, sensory evaluation is important in determining whether these will meet with consumer acceptance. Sensory evaluation is also important for new product development, such as increasing the nutritional value of milk, e.g., by iron (Rodriguez, 2007) or folic acid (Achanta *et al.*, 2007) fortification, or formulating low-fat milk so that it tastes more like full-fat milk (Phillips and Barbano, 1997). From a consumer perspective,

sensory characteristics most directly influence product acceptability. If the milk does not taste good, there will not be repeat purchases.

This chapter will provide some historical perspective on human perceptual evaluations of milk and a description of common methods currently in use to conduct sensory evaluation. This is followed by discussion of the major sensory analysis tools employed, along with examples, and additional sources of information about sensory evaluation.

6.2 Historical perspective

In the latter part of the nineteenth century the grading of dairy products first received national and international attention. The establishment of product grades such as the American Dairy Science Association (ADSA) score cards, as well as standards for various dairy products, closely paralleled the growth of the dairy industry and development of dairy product markets (Bodyfelt *et al.*, 1988). The ADSA score cards have also been used by the International Collegiate Dairy Products Evaluation Contest since 1916 (Trout *et al.*, 1939). The score card is a tabulated list of the various factors used to assess the quality of dairy products. A numerical value is assigned to each factor (Bodyfelt, 1981).

The beginning of the twentieth century marked the establishment of major brands and trade names for dairy products. Some brands of dairy products have become widely known for their high quality. For instance, Sealtest, Inc. established the 'Sealtest brand' recognizing their high quality milk. Brand development necessitated recognition of set standards of finished products by an experienced, competent judge (Bodyfelt *et al.*, 1988).

American agricultural writers, from early on, recognized that the consumption of dairy products depended primarily upon their flavor characteristics. These writers cautioned dairymen concerning feeding and milk handling practices to obtain a high quality dairy product. For example, Deane (1797) advised: 'In feeding milch cows, the flavour of the milk should be attended to ... Feeding them with turnips is said to give an ill taste to the butter made of the milk.'

During the last 25 years, sensory science has advanced tremendously (Goff and Griffiths, 2005; Barbano and Lynch, 2006). 'Sensory Science is enjoying a period of strong growth, both at the intellectual and at the practical levels' (Moskowitz *et al.*, 2003, p. 1). Sensory evaluation of milk has evolved from a defect identification oriented system involving the use of the score card by a single expert, the dairy judge, to an attribute intensity scaling approach that uses trained panelists for quantitatively describing many dimensions of a product's characteristics. Ten to 12 trained panelists evaluate the intensities of the attributes to profile the aroma, flavor, aftertaste, and texture of fluid milk (Stone and Sidel, 2004b). These specific test methods take the form of planned experiments.

The traditional dairy judging approach for evaluating fluid milk for sensory characteristics is based on scoring a product against a specified list of defects commonly found in conventionally pasteurized milk. Traditional dairy judging

has been criticized for failure to predict consumer acceptance, lack of objectivity in quality assessments, difficulty in assignment of quantitative scores, and lack of utility for combining analytically oriented attribute ratings with affectively oriented quality scores (Claassen and Lawless, 1992). In addition to these shortcomings, application of traditional judging strategies to new products, such as extended shelf-life ultra-pasteurized (UP) milk, raises further analytical challenges, as changes in UP product characteristics are more subtle and occur over a longer period of time than changes typically encountered in conventionally pasteurized milk (Boor and Nakimbugwe, 1998; Shipe, 1980). Thus, as the dairy industry moves toward production of extended shelf-life products, the need emerges for development of appropriate sensory tools that are sensitive and specific for these products.

6.3 Principles of sensory evaluation

Sensory evaluation has been defined as a scientific method used to evoke, measure, analyze, and interpret those responses to products as perceived through the senses of sight, smell, touch, taste, and hearing (Stone and Sidel, 2004a). The best raw materials and ingredients yield the best products; hence, sensory quality is a crucial consideration for finished product ingredients such as fluid and dried milk. When dealing with dairy foods, sensory quality is always involved (Clark *et al.*, 2009).

The degree of liking is not the only question answered by sensory analysis. For example, trained panelists can be used to generate data that are objective and analogous to instrumental data; threshold tests can be used to estimate sensory thresholds and qualitative tests can be used to determine consumer emotional responses to products. Shelf-life studies can be used to determine how long a product will taste good, which is critical with dairy products, because of their fragile nature. Consumer perception, the impact of storage, ingredient substitution, packaging, and process variability can be quantified; and correlation can be established between instrumental tests and sensory perception. To meet a specific objective, the correct sensory test must be chosen.

6.4 Examples of evaluation methods, their application and effectiveness

Sensory evaluation comprises a set of techniques for product presentation and well-defined response tasks, statistical methods, and guidelines for interpretation of results. There are many tools in the sensory toolbox, but these tools basically fit into three broad categories, each with a different goal and using different criteria for panelist selection. These three primary kinds of sensory tests focus on the existence of overall differences among products (discrimination tests), specification of attributes (descriptive analysis), and measuring consumer likes

and dislikes (affective or hedonic testing). Once the test objective has been determined, an appropriate test method must be selected along with panel setup, experimental design, test setup, ballot decisions, and scale/question decisions. After the test is conducted, the results are analyzed and interpreted.

No matter which method is chosen, standard sensory practices should be used in order to assure consistent, actionable data (Lawless and Heymann, 1998). Good practices include a controlled sensory testing environment, test protocol considerations (sample serving procedures, sample size, sample serving temperature, serving containers, carriers, palate cleansing), experimental design, and panelist qualifications, as well as accurate tabulation and appropriate statistical analysis.

6.4.1 Discrimination tests

If the objective is to determine whether two samples are perceptibly different, discrimination tests should be used (Peryam, 1958; Amerine *et al.*, 1965; Meilgaard *et al.*, 2006; Stone and Sidel, 2004b). Product developers often need to reformulate a product by using different ingredients. They do not want the consumer to detect a difference, and it is possible for two samples to be chemically different in formulation but for humans not to perceive this difference. For example, a flavored milk manufacturer wants to substitute a less expensive vanilla extract in its chocolate milk. The objective would be to determine whether consumers can detect a difference between the control and the new chocolate milk. The triangle test is an example of a discrimination test that could be used for the above situation. In the triangle test, panelists are given three samples, two the same and one different. The panelist is asked to identify the different sample. The number of correct responses is recorded. Tables are used to determine if the results are statistically significant, i.e. not just due to chance performance or correct guessing. Discrimination testing can also be used if the manager wants to know whether, for instance, a new procedure for pasteurizing milk will be noticeable. There are many different discrimination tests. Some of these tests involve sorting different samples from a set of alternatives such as the triangle test. Others involve matching test samples to standards, such as the duo-trio, ABX and dual standard tests. Others involve choosing the weakest or strongest sample on a specified attribute such as the paired comparison or *n*-alternative forced choice test. All of these have the following in common: an answer is forced ('don't know' is not an option) and there is a chance guessing level of expected correct choices when there is no real difference. Straightforward statistical tests can then test against this chance level of correct choices.

Another type of discrimination test that is useful for the dairy industry in determining the minimum concentration of an off-flavor compound involves sensory threshold measurement. Several issues must be addressed to determine an accurate and reproducible threshold value. Thresholds are impacted by several factors, and perhaps the most significant element is a proper and

consistent testing procedure. This includes an appropriate threshold method, an appropriate number of panelists, and consistent methodology. Examples of threshold methods are the following: ascending forced-choice or 3-AFC method of limits, R-index method (signal detection) and CHARM analysis (yes/no response) (ASTM, 1992; Lawless and Heymann, 1998). With CHARM analysis a person sniffs the effluent of a gas chromatograph to separate volatile flavors and measure several thresholds at once.

An example of discrimination testing on milk is the Matak *et al.* (2007) study where triangle tests revealed differences between the odors of raw milk and ultraviolet (UV) irradiated milk. Triangle tests were also used by Prescott *et al.* (2005) to determine the detection threshold for 2,4,6-trichloroanisole (TCA) in wine. Cork taint in wine produced by TCA is characterized by generally unacceptable musty or earthy odors.

6.4.2 Descriptive analysis

Descriptive analysis consists of training a group (usually 6–12) of individuals to identify and quantify specific sensory attributes or all of the attributes of a food. The extent of the training is dependent upon the complexity of the sensory attributes that are to be profiled.

Descriptive analysis is ideal for shelf-life testing (the length of time during which a food product performs satisfactorily), especially if the panelists are well trained and available, and give consistent results over time. However, descriptive panels are not asked about likes and dislikes, i.e. they are used like an analytical instrument. The results can be correlated with chemical and/or microbiological analysis (Hedegaard *et al.*, 2006). These methods can also be used to determine where a niche lies in the prototype products on the market. In quality control, descriptive techniques can be invaluable when a problem must be defined. Some examples of these tests are Quantitative Descriptive Analysis (Stone and Sidel, 2004b), Spectrum, Flavor Profile, and Texture Profile (Meilgaard *et al.*, 2006).

6.4.3 Acceptance and preference tests

Acceptance (hedonic) and preference testing with consumers is usually performed toward the end of the product development process. When testing the preference of one product directly against a second product, the paired preference test can be used, while with more than two products, a preference ranking test can be used. The most commonly used scale to determine the degree of liking is the 9-point hedonic scale. The scale is anchored with words ranging from 'like extremely' to 'dislike extremely' as in Fig. 6.1 (Lawless and Heymann, 1998). There are similar scales for the testing of acceptance by children that use a smaller number of points with kid-friendly words or smiley faces (Kroll, 1990). Hedonic data can be used in preference mapping. This is a very valuable procedure that allows visualization of the directions of product preferences in spatial models of a product set.

The 9-point hedonic scale

Please evaluate sample ____ and indicate how much you like or dislike it OVERALL?

Dislike	Dislike	Dislike	Dislike	Neither	Like	Like	Like	Like
Extremely	Very	Moderately	Slightly	Like	Slightly	Moderately	Very	Extremely
	Much			Nor			Much	
				Dislike				

The 'Just About Right' scale

Please indicate your opinion about the sweetness of sample ____.

Much too	A little	Just About	A little	Much too
weak	too weak	Right	too strong	strong

Fig. 6.1 Examples of hedonic scales: 9-point hedonic scales (for measuring intensity of liking) and 'just about right' (JAR) for measuring desirability of a specific attribute.

Just Right scales are variations of acceptance testing. The Just Right scale, also called the Just About Right (JAR) scale, measures the desirability of a specific attribute and can be used to determine the optimum level of attributes. For example, in testing chocolate milk, the JAR scale can be anchored with 'not sweet enough' on the left, 'just about right' in the center, and 'much too sweet' on the right. Directional information for product reformulation or optimization can be gleaned from JAR scales, which combines intensity and acceptance judgments (Meilgaard *et al.*, 2006). The Labeled Affective Magnitude scale can be used for assessing food likes/dislikes. It has an increased ability to differentiate among extremely well liked or extremely disliked foods (Schutz and Cardello, 2000) (Fig. 6.2).

When new procedures are developed for processing milk, the most important question to answer is how much consumers like the taste of the milk. A study from the Cornell Sensory Testing Facility (Chapman *et al.*, 2001) utilized sensory evaluation procedures to measure children's acceptance of three types of milk. Children, 6 to 11 years of age, were asked hedonic questions about how they felt about milk in general and also their degree of liking. A 7-point facial hedonic scale with verbal descriptors for affective testing with children, anchored by 'Super bad' and 'Super good' (Kroll, 1990; Resurreccion, 1998) was used to evaluate three types of milk: milk conventionally pasteurized by high temperature short time (HTST), ultra-pasteurized milk (UP), and ultra-high temperature (UHT) milk. The children preferred HTST to UHT milk, which was preferred over UP milk. Children's opinions of the sensory quality of milk should be taken seriously. As UP milks are often distributed in fast food establishments, a place where children commonly drink milk, attention should be directed toward making these products more appealing. According to Dr Joseph Hotchkiss, packaging specialist from Cornell University, Ithaca, NY, 'If you want to increase milk sales, make sure the milk appeals to children. If they like milk when they were young, they are more apt to drink it as an adult' (personal communication, 1998).

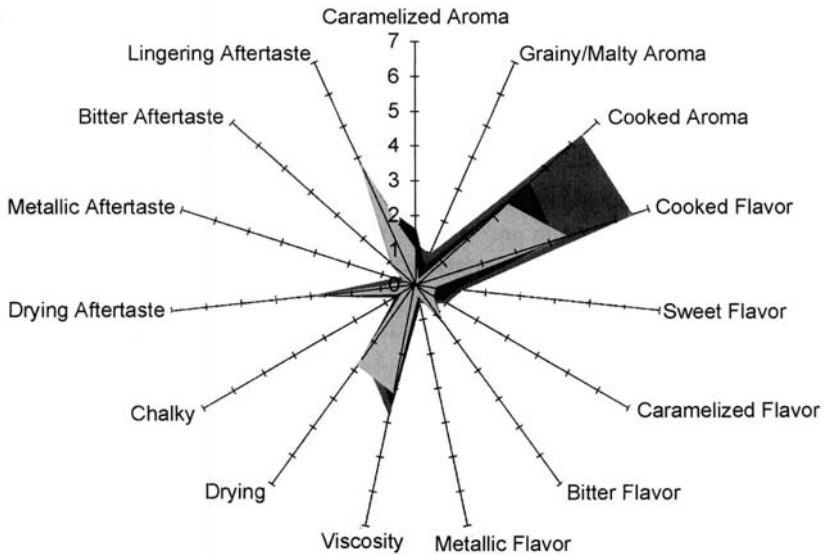


Fig. 6.2 Sensory profiles of reduced fat milk stored at 6°C. Individual attributes are positioned like the spokes of a wheel around a center (zero, or not detected) point, with the spokes representing attribute intensity scales, with higher (more intense) values radiating outward. Legend: dark gray area is day 2, black area is day 29, and light gray area is day 61. (Taken from Chapman *et al.*, 2001, with permission.)

Milk is a perishable product and has a relatively short shelf-life. Therefore, extending the shelf-life of milk will contribute to the competitiveness of the dairy industry in the beverage market. The shelf-life of milk can be influenced by many interacting factors, such as raw milk quality, processing conditions, storage temperature, oxygen pressure, light, and package configuration. Experiments have been designed to determine the shelf-life of milk, which consist of subjecting several samples of a product to tests and observing failure rates over time. Shelf-life testing of milk involves sensory testing for observing and tasting changes in the quality of the product. This may involve the length of time it takes for either the quality to deteriorate to the extent that the product has an off-flavor or until it becomes unfit for human consumption, depending upon the criterion agreed upon for judging the product's failure. Correlating consumer acceptability to trained panelists' data to define sensory failure is an improvement over more arbitrary criteria (Hough *et al.*, 2002).

The Milk Quality Improvement Program (MQIP) at Cornell University follows a strict protocol when performing shelf-life studies. The test objective is to determine the shelf-life of representative milk samples of dairy plants in and around New York State. Quantitative descriptive analysis was used to identify the attributes for the ultrapasteurized-milk ballot. About twelve panelists are trained to reliably identify and rate product attributes and appropriate intensities with refresher training sessions given at least biannually. Six panelists are then used during each shelf-life study. At each testing period, sample containers are

mixed by inversion. Then, in dim light, 60 ml of sample are poured into 148 ml plastic cups (with blind three-digit codes) and capped with the appropriate plastic lid, and placed in a cardboard box. The box of samples is presented to panel members seated in individual booths when the samples reached 15°C. Panelists evaluate the milk according to the following method:

1. Each cup is swirled and the odor is noted by removing the lid and placing the nose directly over the cup and sniffing the headspace.
2. The flavor is noted by taking a generous sip, rolling the milk around in the mouth and then expectorating.
3. After-taste is enhanced by drawing a breath of air slowly through the mouth and then exhaling slowly through the nose (Chapman *et al.*, 1998).

Panelists are provided water and unsalted saltine crackers for rinsing and palate cleansing. Each panelist uses an individual booth equipped with a computer that leads them through the tests and collects data using Compusense[®] *five* software (ver. 4.6; Guelph, ON, Canada). The ballot includes intensity rating scales for each attribute and an overall quality rating. The panelists perform independent rating observations on randomly ordered samples of milk. They only rate the attributes that are perceived in the milk sample (Fig. 6.3). When two or more panelists note an attribute, that attribute and its intensity are reported to the dairy plant.

As product quality drives consumer acceptance and demand, the ability to measure sensory attributes characteristic of high quality products is necessary for development and production of products that meet consumer expectations. Whited *et al.* (2002) used a trained panel to determine the effects of light exposure on vitamin A degradation and on light-oxidized flavor development of whole, reduced fat, and nonfat milk for 2 to 16 hours and found that even brief moderate light exposure can reduce the nutritional value and flavor quality of milk.

After a failure criterion has been agreed upon, the time to failure can be observed. If the criterion is the point in time at which observers first respond to a stimulus, threshold testing will have a key role in the shelf-life determination. Sensory evaluation has been successfully used to measure the threshold of odor or taste in milk. Thresholds are the limits of sensory capacities, the point at which a sensory signal becomes minimally detectable (Meilgaard *et al.*, 2006). For example, the semi-ascending paired difference method was used to evaluate the sensory threshold for changes in milk flavor resulting from oxidizing effects of light (Chapman *et al.*, 2002). Trained panelists were able to detect flavor defects following as little as 15 to 30 minutes of light exposure, while consumers detected defects between 54 minutes and 2 hours.

Since product quality drives consumer acceptance and demand, the ability to sensitively measure sensory attributes characteristic of high quality products is necessary for the development and production of products that meet consumer expectations. An ascending two-alternative forced choice test with 12 panelists was used, for example, to establish the threshold for certain dairy-related flavor

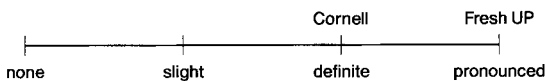
Milk Quality Improvement Project Presents this Shelf-life Study on Ultra-Pasteurized Milk Box 1

Panelist Code: _____

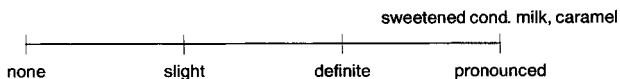
- A. Compare your samples with the references.
 1. Use the **2% Cornell milk** reference for '**cooked**' and '**2% fat**'
 2. Use the **100% lactose reduced** reference for **pronounced 'sweet.'**
- B. When rating the attributes, you may click anywhere on the intensity line.
- C. Start with sample %01 on left.
- D. You only have to rate those attributes that are relevant.

Question #1 Please evaluate the following Aroma attributes for Sample __

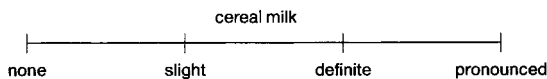
Cooked Aroma



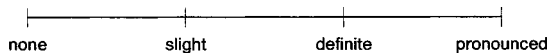
Caramelized Aroma



Grainy/malty aroma



Other Aroma

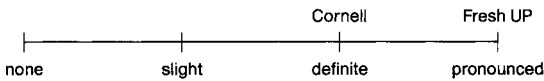


Question #2 If you detect 'other aroma', type in the description below.
Sample __

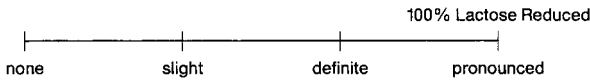
Fig. 6.3 Ballot used for ultra-pasteurized milk shelf-life studies.

Question # 3 Please evaluate the following Flavor attributes for Sample __

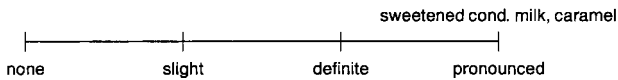
Cooked Flavor



Sweet Flavor



Caramelized Flavor



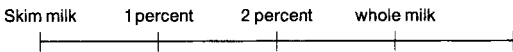
Other Flavor



Question # 4 If you detect 'other flavor', type in the description below.
Sample __

Question # 5 Please evaluate the following Texture attributes for Sample __

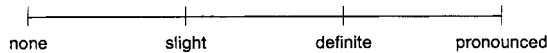
Viscosity



Chalky

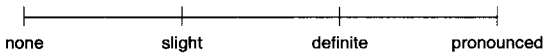
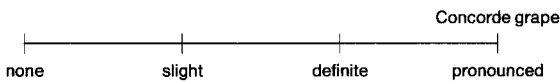
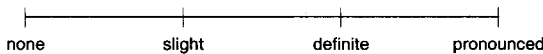


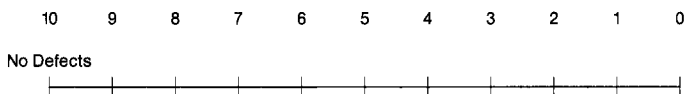
Other Texture



If you detect 'other texture', type in the description below.

Fig. 6.3 continued

Question # 6 Please evaluate the following Aftertaste attributes for Sample __**Bitter****Drying****Lingering****Other Aftertaste****If you detect 'other aftertaste', type in the description below.**

Overall Quality Rating**Check scoring guide for off-flavors of Ultra-pasteurized milk posted in the booth.****OVERALL QUALITY RATING****Fig. 6.3** *continued*

compounds (diacetyl, hexanol and δ -decalactone) (Adhikari *et al.*, 2006). Prescott *et al.* (2005) used a method that combined a paired preference test with a threshold procedure to determine an average rejection point by consumers.

New understanding of sensory judgment has been obtained through comparison with physical and chemical data. Hotchkiss *et al.* (1999) studied the effect of dissolved CO₂ and barrier films on the shelf-life extension of milk. Once the sensory threshold was determined, the relationship between microbial

growth and package barrier properties to which CO₂ had been added at concentrations near the flavor threshold could be established. In a study by Hedegaard *et al.* (2006), descriptive sensory analysis was used to analyze the oxidative changes in milk. Sensory analysis and chemical analysis showed high correlation between the typical descriptors for oxidation, such as cardboard, metallic taste, and boiled milk and specific chemical markers for oxidation such as hexanal. Notably, primary oxidation products (i.e., lipid hydroperoxides), and even the tendency to form radicals as measured by electron spin resonance spectroscopy, showed high correlation to the sensory descriptors for oxidation.

An objective of many sensory evaluation studies is 'to provide useful chemical benchmarks of when off-flavors will be detected' (Santos *et al.*, 2003b, p. 2492). The following are two milk quality studies of this type. Santos *et al.* (2003a) used an ascending forced choice method (a series of triangle tests with ascending difference) to assess the sensory threshold for off-flavors in milk due to lipolysis and proteolysis resulting from the action of native milk enzymes. In 2% milk, lipolysis at a free fatty acid concentration of 0.25 meq/kg of milk was detected by more than one third of the panelists. For the detection of activity of native milk proteases in skim milk, the researchers found a detection sensory threshold at a 4.76% decline in casein as a percentage of true protein (CN/TP). Similarly, an off-flavor sensory detection threshold related to proteolysis was determined by Ma *et al.* (2000) to be about a 4% decline in CN/TP.

However, it must be remembered that detection does not necessarily mean rejection. Prescott *et al.* (2005) conducted two studies using a paired preference test with a method of constant stimuli threshold procedure to determine a consumer rejection threshold. Estimates of TCA thresholds in wine had been reported, but the level at which TCA became unacceptable had not. This study found that the detection threshold was 2.1 ppt, while the consumer rejection threshold was 3.1 ppt.

Often, several sensory techniques are applied in one study. Croissant *et al.* (2007) studied the impact of feed on the flavor of milk using the following methods: Spectrum descriptive analysis (to describe the flavor), a triangle difference test (to determine whether consumers could detect a difference between milk from conventional and pasture-based systems), a point hedonic scale (to determine consumer acceptance), and principal component analysis (PCA) (to determine how the treatments and individual sample collections differentiated from each other across sensory and instrumental measurements).

Descriptive sensory analyses are probably the most sophisticated tools in the arsenal of the sensory scientist. These techniques allow the sensory scientist to obtain complete sensory descriptions of products. They help identify underlying ingredient and process variables and/or determine which sensory attributes are important to acceptance. Usually, descriptive analyses yield objective descriptions of samples in terms of the perceived sensory attributes. One example of descriptive analysis is quantitative descriptive analysis (QDA). The principle of QDA is based on a panelist's ability to measure specific attributes of a product in a reproducible manner to yield a comprehensive quantitative

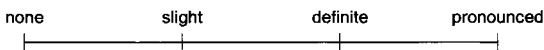
product description amenable to statistical analyses. In a QDA approach, panelists recruited from the general public work together in a focus group to identify key product attributes and appropriate intensity scales specific to a product. See Fig. 6.4 for examples of intensity line scales that are used in descriptive analysis. This group of panelists is then trained to reliably identify and score product attributes. As panelists generate the attribute terms, the resulting descriptions are meaningful to consumers.

Thus, analyses provide information amenable to modeling predictions of consumer acceptability. Physical reference standards determined by panel consensus are used to develop the proper descriptive language to reduce the amount of time required to train the panelists and to calibrate the panel in the use of the intensity scale. QDA results can be analyzed statistically and then represented graphically. The analytic technique of quantitative descriptive analysis has gained acceptance for sensory evaluation of various food and dairy products (Stone and Sidel, 2004b) and their shelf-life. Some examples of such studies include conventionally pasteurized milk (Phillips *et al.*, 1995; Quiñones *et al.*, 1998a,b) and microwave-treated milk vs. UHT milk (Clare *et al.*, 2005).

Chapman *et al.* (2001) used QDA to characterize the key attributes of nine UP milk products throughout their 60-day product shelf-life. The attribute terms were as follows: aroma attributes (cooked, caramelized, grainy/malty), flavor attributes (cooked, sweet, caramelized, bitter, metallic), texture attributes (viscosity, drying, chalky, lingering), and aftertaste attributes (drying, metallic, bitter). The 12 trained panelists evaluated the milk according to the method

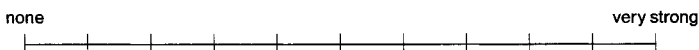
Examples of Intensity Line scales

1. Intensity labeled

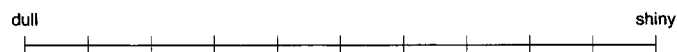


2. End anchors labeled

a. vanilla aroma



b. gloss of surface



3. Intensity line scale with a reference

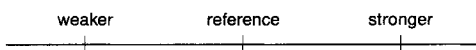


Fig. 6.4 Examples of intensity line scales for measuring intensities of attributes.

described in Chapman *et al.* (1998). The mean intensity of each attribute was graphically displayed in a 'spider plot' to provide a visual profile or 'fingerprint' of the important product attributes. Figure 6.2 shows a UP milk high in cooked aroma and cooked flavor, and medium in sweetness, caramelized flavor, drying, drying aftertaste, and lingering aftertaste. (Note that such a plot also shows that some taste attributes were *not* present, e.g., bitter flavor, chalky texture, bitter and metallic aftertaste.) Superimposing of spider plots can be used to compare products with each other or to compare attribute intensities of a single product tested at different points in time.

Lawless and Heymann (1998) indicated that data from sensory methods such as QDA, when combined with chemical or physical characteristics of a product detected, for example, by gas chromatographic olfactometry (GCO) or dynamic mechanical analysis (rheological measurements), respectively, can be used to define and optimize dairy product characteristics for different market segments. The combined sensory and instrumental data can then be analyzed with more advanced statistical protocols such as principal component analysis (PCA).

6.5 Application of advanced statistical methods

There are many new statistical methods available to the sensory scientist that can apply to dairy research. Some methods extend to marketing techniques such as conjoint analysis. Recently conjoint analysis has been used to simulate real-world consumer purchasing behavior. It also has been used to survey general knowledge and perception of health benefits of dairy products (Moskowitz *et al.*, 2005; Jones *et al.*, 2008). Conjoint analysis is also ideal for optimizing new product designs by identifying the most appealing sets of features. Conjoint analysis, sometimes referred to as trade-off analysis, is a multivariate technique that quantitatively measures the relative importance of different marketing variables, attributes, or product features related to a brand, product, or service. The distinguishing feature of this technique is that each variable's importance is determined implicitly or indirectly. That is, the respondent is not consciously aware of what is being measured.

The following papers illustrate some of the advanced methods of logistic regression, principal component analysis, and modeling. Logistic regression was used for the prediction of food shelf-life (Hough *et al.*, 2002, 2003; Gambaro *et al.*, 2004; Salvador and Fiszman, 2004). Chapman *et al.* (2006) used logistic regression for analyzing preference milk data. Because preference options can be reduced to a dichotomous response (preference/no preference), logistic regression is a useful tool for examining the impact of experimental manipulations on preferences (Hosmer and Lemeshow, 2000). Binomial (or binary) logistic regression is a form of regression that is used when the dependent variable is dichotomous and the independent variables are continuous or categorical. This procedure can be used to predict the probability of each response based on values of the independent variables, to rank the relative importance of independent

variables, to assess interaction effects, and to understand the impact of covariate control variables.

With descriptive sensory data, several dependent variables may be correlated with one another. To illustrate, following ANOVA, several sensory descriptors may appear to significantly discriminate among the samples, but multiple descriptors may be driven by the same underlying causes. Principal component analysis (PCA) is a multivariate technique that provides a method of extracting structure from a variance–covariance or correlation matrix. PCA identifies patterns of correlation among dependent variables and substitutes a new variable, called a factor, for the group of original attributes that were correlated. The analysis then identifies a second and a third group of attributes and derives a factor for each, based on the residual variance (that which is left after the variance accounted for by the previous factor has been removed). The attributes will have a correlation with the new dimensions, called a factor loading, and the products will have values on the new dimensions, called factor scores. The factor loadings are useful in interpreting the dimensions, and the factor scores show the relative positions among the products in a map (Lawless and Heymann, 1998). Thus, PCA transforms original dependent variables into new uncorrelated dimensions to simplify the data structure, eliminate descriptor redundancies, and indicate potential latent causal variables. That is, the factors may suggest underlying causes or processes that give rise to the resulting pattern of correlated changes in sensory attributes.

In a study with UP milk (Chapman *et al.*, 2001), PCA was applied to the mean attribute ratings listed to simplify interpretation of data from 15 attributes measured on nine products. PCA was applied with factor analysis (Lawless and Heymann, 1998). PCA was applied to the mean attribute ratings. The four principal components (PCs) generated from this analysis accounted for 94.4% of the total variance in the data set. The attributes selected each had a consistently high value, indicating that the attribute was often present, had low standard deviation, and was not highly correlated with another attribute. The analysis extracted the most significant variables with minimum loss of information. Varimax rotation was performed on these four PCs to bring them into closer alignment with the original variables. PC1 was entirely related to the following ‘cooked’ attributes: cooked aroma, grainy/malty aroma, and cooked flavor. PC2 had large negative loadings for dry and lingering. PC3 was largely negatively related to ‘sweet’ attributes: caramel and sweet. PC4 was almost entirely influenced by bitterness. Finally, the samples (products) were plotted into the data space described by the retained principal components. Factor scores were calculated for each sample to determine its location on the retained components. Samples farther apart on the principal component map are perceptually more different from samples found closer together (Coxon, 1982; Schiffman *et al.*, 1981).

After PCA analysis, a regression model can be used to estimate the overall product quality rating based on measurements of its attributes (Siebert, 1999). The overall quality ratings were modeled as a function of the four Varimax

rotated PCs for cooked, dry lingering, sweet and bitter scores for the UP products. Models were constructed using ordinary least squares, principal components regression, and partial least squares regression. The best fit equation for prediction of overall quality at day 60 was obtained (Chapman *et al.*, 2001):

$$\begin{aligned}\text{Overall quality rating} = & 7.01 + 0.127 \text{ cooked} + 0.013 \text{ dry/lingering} \\ & + 0.154 \text{ sweet} - 0.424 \text{ bitter}\end{aligned}$$

In general, as shown by the model, perception of bitter flavor had the most dramatic effect on overall quality perception.

Generation of quantitative descriptive sensory data can contribute to a well-defined competitive marketing strategy. Product positioning can assist a firm's target customers to understand and appreciate a specific product's characteristics in relation to those of its competitors' products. In this strategy, each brand within a set of competitive products is thought to occupy a certain position in a customer's 'perceptual space' (Urban *et al.*, 1987). In general, marketers have two broad objectives in mind when undertaking perceptual mapping. One objective is to determine where a target brand is positioned versus the competition. The other objective is to help identify determinant product attributes that influence customer choice within the product class (Kohli and Leuthesser, 1993). These determinant attributes must be important to customers and must also exhibit differences across brands. No matter how important a product attribute is, if brands are not perceived to differ in that attribute, then the attribute will not be influential in customers' decisions. Perceptual mapping can contribute to strategic product positioning for development and marketing of new products.

Sensory tests provide useful information about the human perception of product changes due to ingredients, processing packaging or shelf-life. Sensory results reduces decision risks. A well functioning sensory program will be useful to a company in meeting consumer expectations and ensuring a greater chance of marketing success. The utility of the information provided is directly proportional to the quality of the sensory measurement.

Statistics have always been an essential part of sensory evaluation. All quantitative sensory data should be analyzed statistically to be interpreted properly and to give actionable results. Even though statistics are necessary, they are only as good as the sensory data used.

6.6 Sources of further information and advice

- A 1974 survey conducted by Cornell University revealed a direct correlation between milk flavor and levels of milk consumption by school-age children (Boor, 2001). This clear link between milk quality and product consumption provided the initial impetus for the development of the Milk Quality Improvement Program (MQIP). Current objectives of the ongoing MQIP are

to monitor initial and keeping quality of commercially processed and packaged fluid milk products in New York State; to monitor raw milk quality; and to assist dairy plants in identifying and correcting handling and processing problems affecting dairy product quality. Sensory evaluation was used in monitoring HTST milk acceptability. Through the research and expertise of MQIP, nine of New York State's dairy plants have increased their shelf-life from 14 to 17 days. For more information contact Dr Kathryn J. Boor at Cornell University, Ithaca, NY.

- For practical procedures on conducting a taste panel, consult a good sensory evaluation text, like *Sensory Evaluation Techniques* (Meilgaard *et al.*, 2006). *Sensory Evaluation of Food: Principles and Practices* (Lawless and Heymann, 1998) is also helpful, and includes philosophical and psychological background. For statistical methods and procedures, consult *Sensory Evaluation of Food, Statistical Methods and Procedures* (O'Mahony, 1986). *Consumer Sensory Testing for Product Development* (Resurreccion, 1998) is excellent for consumer testing. *Sensory Evaluation of Dairy Products*, 2nd Edition (Clark *et al.*, 2009) is (1) an overview of the history, art, and science behind the sensory evaluation of dairy products; (2) a guide to assist in tracing the origins of identifiable sensory defects in dairy products with strategies for their correction; (3) a practical guide to the preparation of samples for sensory evaluation; and (4) a training tool for personnel in the evaluation of dairy products.
- ASTM International publishes standard specifications, tests, practices, guides, and definitions relating to materials, such as sensory methods for milk. See ASTM org or ASTM journal provides technical papers in print and online.
- The International Organization for Standardization (ISO), Geneva, Switzerland. ISO Standard 8586. 1. Sensory Analysis – General guidance for the selection, training, and monitoring of assessors.
- The Sensory User Group (Internet): visit sensory@yahoogroups.com.
- Institute of Food Technology (IFT).
- The Student Product Development Competition competes at the IFT annual meeting. They often develop new dairy products. IFT offers short courses. At the annual meeting are presenters from industry and academia (poster and oral presentations). The sensory evaluation division of IFT publishes a newsletter, called *Sensory Forum*, which tells about upcoming events and courses, employment opportunities, recent literature and more. Visit www.IFT.org.
- The Society of Sensory Professionals. Visit www.sensory.org/SSP/.
- Symposia: *Pangborn Sensory Symposium* (www.pangborn2011.com), *Sensometrics Meeting* (www.sensometric.org).
- Journals in the field include *Journal of Sensory Studies* (published by Food & Nutrition Press, Inc.; view online resource from Blackwell Synergy), *Food Quality and Preference* (published by Elsevier Ltd, <http://www.sciencedirect.com/science/journal/09503293>), *Journal of Dairy Science* (published

by the American Dairy Science Association), *Journal of Food Science* (published by the Institute of Food Technology; view online resource from Blackwell Synergy), and *Dairy Foods* (published by BNP Media II, LLC, <http://www.dairyfoods.com/>).

- Some multivariate approaches to mapping product similarities are Drivers of Liking[®] and Landscape Sementation Analysis[®], The Institute for Perception (ifpress@cs.com).
- Short courses are given by universities such as Cornell University (<http://www.foodscience.cornell.edu/cals/foodsci/research/sensory/index.cfm>) and UC Davis (www.extension.ucdavis.edu/sensory) and by companies such as International Resources for Insights and Solutions, The Institute for Perception (www.ifpress.com), Sensory Spectrum, Inc. (www.sensoryspectrum.com), Insights Now (Greg.Stucky@InsightsNow.com), and IRIS: International Resources for Insights and Solutions, LLC (amunoz@IRIS-consulting.net).
- Organizations that are involved in assessing, regulating, or promoting dairy products include:

Agricultural Marketing Service (AMS)
Ams.usda.gov

American Dairy Science Association
1111 B, Dunlap Avenue
Savoy, IL 61874
217-356-5146
Adsa.org

Center for Food Safety and Applied Nutrition
5100 Paint Branch Parkway
College Park, MD 20740-3835
Sfsan.fda.gov/list.html

Centers for Disease Control and Prevention (CDC)
1600 Clifton Road
Atlanta, GA 20222
1-800-311-3435
cdc.gov

Code of Federal Regulations (CFR)
Access.gpo.gov/cgi-bin/cfrassemble.cgi?title=200021
Collegiate Dairy Products Evaluation Contest
Ams.usda/gpv/dairy/cdpec/coach_corner.htm

Collegiate Dairy Products Evaluation Contest
Ams.usda.gov/dairy/cdpec/coach_corner.htm

Department of Health and Human Services (HHS)
200 Independence Avenue, S.W.
Washington, DC 20201
hhs.gov

Future Farmers of America (FFA) represents a large diversity of over 300 careers in the food, fiber and natural resources industry. FFA (www.FFA.org) teaches sensory evaluation of milk.

International Dairy Foods Association (IDFA)
1250 W. Van Buren, Suite 900
Washington, DC 20005
202-737-4332
IDFA.org/index.cfm

National Dairy Council
Nationaldairycouncil.org/nationaldairycouncil/sitemap

Pasteurized Milk Ordinance (PMO)
Cfsan.fda.gov/~ear/pmo03toc.html

U.S. Food and Drug Administration (FDA)
5600 Fishers Lane
Rockville, MD 20857-0001
1-800-463-6332
Fda.gov

U.S. Public Health Service (PHS)
Phs.gov/

6.7 References and further reading

- ACHANTA K, BOENEKE CA and ARYANA KJ (2007), 'Characteristics of reduced fat milks as influenced by the incorporation of folic acid', *J Dairy Sci*, 90, 90–98.
- ADHIKARI K, HEIN K, ELMORE J, HEYMAN H and WILLMOTT A (2006), 'Flavor threshold as affected by interaction among three dairy-related flavor compounds', *J Sensory Studies*, 21, 626–643.
- AMERINE MA, PANGBORN RM and ROESSLER EB (1965), *Principles of Sensory Evaluation*, New York, Academic Press.
- ASTM (1992), Standard practice for determination of odor and taste threshold by a forced choice method of limits E-679-91. In: *Annual Book of Standards*. 15.07 Philadelphia, PA, American Society for Testing and Materials, pp. 35–39.
- BARBANO D and LYNCH J (2006), 'Major advances in testing of dairy products: milk component and dairy product attribute testing', *J Dairy Sci*, 89, 1189–1194.
- BODYFELT FW (1981), 'Dairy product score cards: are they consistent with the principles of sensory evaluation?', *J Dairy Sci*, 6, 2303–2308.
- BODYFELT FW, TOBIAS J and TROUT GM (1988), *Sensory Evaluation of Dairy Products*, New York, Van Nostrand/AVI.
- BOOR KJ (2001), 'ADSA Foundation Scholar Award Fluid Dairy Product Quality and Safety: Looking to the future', *J Dairy Sci*, 84, 1–11.
- BOOR KJ and NAKIMBUGWE DN (1998), 'Quality and stability of 2% fat ultrapasteurized fluid milk products', *Dairy, Food Environ San*, 18, 78–82.
- CHAPMAN KW and BOOR KJ (2001), 'Acceptance of 2% ultra-pasteurized milk by

- consumers, 6 to 11 years old', *J Dairy Sci*, 8, 951–954.
- CHAPMAN KW, ROSENBERRY LC, BANDLER DK and BOOR KJ (1998), 'Light-oxidized flavor development and vitamin A degradation in chocolate milk', *J Food Sci*, 63, 930–934.
- CHAPMAN KW, LAWLESS HT and BOOR KJ (2001), 'Quantitative descriptive analysis and principal component analysis for sensory characterization of ultrapasteurized milk', *J Dairy Sci*, 84, 12–20.
- CHAPMAN KW, WHITED LJ and BOOR KJ (2002), 'Sensory threshold of light-oxidized flavor defects in milk', *J Food Sci*, 67, 2770–2773.
- CHAPMAN KW, GRACE-MARTIN K and LAWLESS HT (2006), 'Expectations and stability of preference choice', *J Sensory Studies*, 21, 441–455.
- CLAASSEN MR and LAWLESS HT (1992), 'A comparison of descriptive terminology systems for the sensory analysis of flavor defects in milk', *J. Food Sci*, 57, 596–621.
- CLARE DA, BANG WS, CARTWRIGHT G, DRAKE MA, CORONEL P and SIMUNOVIC J (2005), 'Comparison of sensory, microbiological, and biochemical parameters of microwave versus indirect UHT fluid skim milk during storage', *J Dairy Sci*, 88, 4172–4182.
- CLARK, S, COSTELLO, M, DRAKE M and BODYFELT FW (2009), *The Sensory Evaluation of Dairy Products*, New York, Springer Science.
- COXON APM (1982), 'Three-way and further extensions of the basic model', in *The User's Guide to Multidimensional Scaling*, London, Heinemann Educational Books, pp. 186–241.
- CROISSANT AE, WASHBURN SP, DEAN LL and DRAKE MA (2007), 'Chemical properties and consumer perception of fluid milk from conventional and pasture-based production systems', *J Dairy Sci*, 90, 4942–4953.
- DEANE S (1797), *The New-England Farmer*, Worcester, MA, Isaiah Thomas, p. 78.
- GAMBARO A, FISZMAN S, GIMENEZ A, VARELA P and SALVADOR A (2004), 'Consumer acceptability compared with sensory and instrumental measures of white pan bread: Sensory shelf-life estimation by survival analysis', *J Food Sci*, 69, 401–405.
- GOFF HD and GRIFFITHS MW (2005), 'Major advances in fresh milk and milk products: fluid milk products and frozen desserts', *J Dairy Sci*, 89, 1163–1173.
- HEDEGAARD RV, KRISTENSEN D, NIELSEN JH, FRØST MB, ØSTDAL H, HERMANSEN JE, KRÖGER-OHLSSEN M and SKIBSTED LH (2006), 'Comparison of descriptive sensory analysis and chemical analysis for oxidative changes in milk', *J Dairy Sci*, 89, 495–504.
- HOSMER DW and LEMESHOW S (2000), *Applied Logistic Regression*, 2nd edn, New York, John Wiley & Sons.
- HOTCHKISS JH, CHEN JH and LAWLESS HT (1999), 'Combined effects of carbon dioxide addition and barrier films on microbial and sensory changes in pasteurized milk', *J Dairy Sci*, 82, 690–695.
- HOUGH G, SANCHEZ RH, GARBARINI DE PABLO G, SANCHEZ RG, CALDERON VILLAPLANA S, GIMENEZ AM and GAMBARO A (2002), 'Consumer acceptability versus trained sensory panel scores of powdered milk shelf-life defects', *J Dairy Sci*, 85, 2075–2080.
- HOUGH G, LANGOHR K, GOMEZ G and CURIA A (2003), 'Survival analysis applied to sensory shelf life of foods', *J Food Sci*, 68, 359–362.
- JONES VS, DRAKE MA, HARDING R and KUHN-SHERLOCK B (2008), 'Consumer perception of soy and dairy products: a cross-cultural study', *J Sensory Studies*, 23:65–79.
- KOHLI CS and LEUTHESSER L (1993), 'Product positioning: a comparison of perceptual mapping techniques', *J Prod Brand Mgt*, 2, 10–19.

- KROLL BJ (1990), 'Evaluating rating scales for sensory testing with children', *Food Technol*, 44, 78–86.
- LAWLESS HT and CLAASSEN MR (1993), 'Validity of descriptive and defect-oriented terminology systems for sensory analysis of fluid milk', *J Food Sci*, 58, 108–112.
- LAWLESS HT and HEYMANN H (1998), *Sensory Evaluation of Food: Principles and Practices*, New York, Chapman & Hall.
- MA Y, RYAN C, BARABANO DM, GALTON DM, RUDAN MA and BOOR KJ (2000), 'Effects of somatic cell count on quality and shelf-life of pasteurized fluid milk', *J Dairy Sci*, 83, 264–274.
- MATAK KE, SUMNER SS, DUNCAN SE, HOVINGH E, WOROBO RW, HACKNEY CR and PIERSON MD (2007), 'Effects of ultraviolet irradiation on chemical and sensory properties of goat milk', *J Dairy Sci*, 90, 3178–3186.
- MEILGAARD M, CIVILLE GV and CARR BT (2006), *Sensory Evaluation Techniques*, 4th edn, Boca Raton, FL, CRC Press.
- MOSKOWITZ HR, MUÑOZ AM and GACULA MC (2003), *Viewpoints and Controversies in Sensory Science and Consumer Product Testing*, Trumbull, CT, Food and Nutrition Press.
- MOSKOWITZ HR, GERMAN JB and SAGUY IS (2005), 'Unveiling health attitudes and creating good-for-you foods: the genomics metaphor, consumer innovative web-based technologies', *Crit Rev Food Sci Nutr*, 45, 165–191.
- O'MAHONY M (1986), *Sensory Evaluation of Food, Statistical Methods and Procedures*, New York, Marcel Dekker.
- PERYAM DR (1958), 'Sensory difference tests', *J Food Tech*, 12, 231–236.
- PHILLIPS LG and BARBANO DM (1997), 'The influence of fat substitutes based on protein and titanium dioxide on the sensory properties of lowfat milks', *J Dairy Sci*, 80, 2726–2731.
- PHILLIPS LG, MCGIFF ML, BARBANO DM and LAWLESS HT (1995), 'The influence of nonfat dry milk on the sensory properties, viscosity, and color of lowfat milks', *J Dairy Sci*, 78, 1258–1266.
- PRESCOTT J, NORRIS L, KUNST M and KIM K (2005), 'Estimating a "consumer rejection threshold" for cork taint in white wine', *Food Qual Pref*, 16, 345–349.
- QUIÑONES HJ, BARBANO DM and PHILLIPS LG (1998a), 'Influence of protein standardization by ultrafiltration on the viscosity, color, and sensory properties of skim and 1% milk', *J Dairy Sci*, 80, 3142–3151.
- QUIÑONES HJ, BARBANO DM and PHILLIPS LG (1998b), 'Influence of protein standardization by ultrafiltration on the viscosity, color, and sensory properties of 2 and 3.3% milks', *J Dairy Sci*, 81, 884–894.
- RESURRECCION AVA (1998), *Consumer Sensory Testing for Product Development. Affective Testing with Children*, Gaithersburg, MD, Aspen Publishers.
- RODRIGUEZ R (2007), 'Sensory and Nutritional Quality of Iron Fortified Milk', MS thesis, Cornell University.
- ROLAND AM, PHILLIPS LG and BOOR KJ (1999), 'Effects of fat replacers on the sensory properties, color, melting, and hardness of ice cream', *J Dairy Sci*, 82, 2094–2100.
- SALVADOR A and FISZMAN SM (2004), 'Textural and sensory characteristics of whole and skimmed flavored set-type yogurt during long storage', *J Dairy Sci*, 87, 4033–4041.
- SANTOS MV, MA Y, CAPLAN Z and BARBANO DM (2003a), 'Sensory threshold of off-flavors caused by proteolysis and lipolysis in milk', *J Dairy Sci*, 86, 1601–1607.
- SANTOS MV, MA Y and BARBANO DM (2003b), 'Effect of somatic cell count on proteolysis

and lipolysis in pasteurized fluid milk during shelf-life storage', *J Dairy Sci*, 86, 2491–2503.

SCHIFFMAN SS, REYNOLDS ML and YOUNG FW (1981), *Introduction to Dimensional Scaling*, New York, Academic Press.

SCHUTZ HG and CARDELLO AV (2000), 'A labeled affective magnitude (LAM) scale for assessing food liking/disliking', *J Sensory Studies*, 16, 117–156.

SHIPE WF (1980), 'Analysis and control of milk flavors', in *The Analysis and Control of Less Desirable Flavors in Foods and Beverages*, G. Charalambous (ed.), Orlando, FL, Academic Press, pp. 201–239.

SIEBERT KJ (1999), 'Modeling the flavor threshold of organic acids in beer as a function of their molecular properties', *Food Qual Pref*, 10, 129–137.

STONE H and SIDEL JL (2004a), *Sensory Evaluation Practices*, 2nd edn, San Diego, Academic Press.

STONE H and SIDEL JL (2004b), 'Quantitative descriptive analysis: developments, applications, and the future', *Food Technol*, 52, 48–52.

TROUT GM, WHITE W, MACK MJ, DOWNS PA and FOUTS EL (1939), 'History and development of the Students' National Contest in the judging of dairy products', *J Dairy Sci*, 22, 375–387.

URBAN GL, HAUSER JR and DHOLAKIA N (1987), 'Mapping consumers' product perception', in *Essentials of New Product Management*, Englewood Cliffs, NJ, Prentice-Hall, pp. 103–120.

WHITED LJ, HAMMOND BH, CHAPMAN KW and BOOR KJ (2002), 'Vitamin A degradation and light-oxidized flavor defects in milk', *J Dairy Sci*, 85, 351–354.