SENSORY ANALYSIS

Section 4.

Methods of Sensory Evaluation

Sensory evaluation methods may be divided into two broad classes: affective and analytical methods (Institute of Food Technologists, 1981). Affective methods use consumer panels or trained panelists to answer questions such as the following:

- Which product do you prefer?
- Which product do you like?
- How well do you like this product?
- How often would you buy/use this product?

Affective methods require a much larger panel size than do analytical methods in order to have greater confidence about the interpretation of the results. The most common analytical methods of sensory evaluation used in the wine industry are discrimination (or difference) and descriptive methods. Discrimination tests can be used to determine if products are different, if a given wine characteristic is different among samples, or if one product has more of a selected characteristic than another. Experienced panelists can complete discrimination tests.

Descriptive methods are used to provide more-comprehensive profiles of a product by asking panelists to identify the different characteristics within the product and quantify characteristics. Trained panelists must be used for descriptive methods (see Trained Panelists and Panelist Training).

Discrimination (Difference) Tests

Difference testing is used to determine if different winemaking processing techniques or operations have a sensory impact. As such, difference testing methods generally provide the winemaker with the practical information needed. They are the most feasible for use in a winery environment, and are simple and robust.

There are many other sensory methods available, including consumer preference and acceptance tests, and descriptive analysis. However, performing some of these more-elaborate tests may not be feasible in small- to medium-sized wineries. They are available through sensory service companies.

Difference testing is a way to determine if a sensory difference actually exists between samples. The degree or nature of the difference cannot be quantified, however. Descriptive tests are generally needed to truly define differences. There are four types of difference tests which can be used to answer some practical questions. The most common for use in the wine industry are the triangle difference test and the duo-trio difference test:

- triangle: “Is a particular lot made with rot-compromised fruit different from other lots?”
- duo-trio: “Is there a sensory difference among wines fermented with different yeasts?”

- paired comparison: “Does the high VA in this wine impact it sensorially?”

A brief description of the methodology of these procedures, including how to perform the tests, the number of tasters required, and the required result for concluding that a significant difference truly exists, is outlined in Table 1. Once a difference has been established, another more elaborate test, such as a preference test, can also be performed.

Difference tests are sometimes applied in a crude fashion at the winery where only one or two tasters perform the evaluation. While better than no testing at all, to achieve a statistically-significant sensory result, a slightly more advanced procedure should be carried out using the minimum number of tasters, as suggested in Table 1.

Generally, the larger the number of evaluators, the better. However, even a small panel of 5-7 will provide highly-valuable information that will greatly increase reliability and consistency of production decisions based on sensory assessment. Table 1 provides the number of correct responses for various tests to be statistically viable at the 95-percent confidence limit. This means that there is but a 5% chance that these results are simply due to random error.

Using a panel, as opposed to a single taster or two, reduces the risk of concluding there is no difference among wines, when one actually exists. Any number of panelists can be used, and the more tasters, the better. For an overview of wine sensory evaluation, and to determine the number of correct responses required for a significant result for any number of tasters, see Zoecklein et al. (1999 and 2005).

Table 1. Outline of Sensory Difference and Preference Tests

Test	**Min. tasters[1]**	**Use**	**Samples**	**Basic method**	**Results: are the wines significantly different?**
Triangle	5	Multi-purpose	Three coded test samples. Two are the same wine (A) (but are coded differently). One is a different wine (B).	Tasters assess all three samples, then pick the sample which is different from the other two, or the odd one out. Serving orders[3]: AAB, ABA, BAA, BBA, BAB, ABB	**Correct response** - taster picks the odd one out. **Significance** - Required no. of correct/total responses[2]: Single tasting 4/5 5/6 5/7 6/8 Repeated tasting 7/10 8/12 9/14 9/16
Duo-trio	7	Comparison to a reference wine	One reference sample (Ref). Two coded test samples (A,B). A is the same wine as the reference (control wine). B is the wine to test.	Tasters assess the reference (Ref), then the two test samples (A,B). Tasters are asked to indicate which test sample is the same as the reference. Serving orders: Ref AB, Ref BA	**Correct response** - taster picks A as the same as the reference. **Significance** - Required no. of correct/total responses[2]: Single tasting 7/7 7/8 8/9 9/10 Repeated tasting 10/12 11/14 12/16 13/18
Paired comparison	7	When a difference is known	Two coded test samples (A,B). One is known to be chemically higher in an attribute (e.g., sweetness).	Tasters are asked to identify which sample is higher in an attribute (e.g., identify which sample is sweeter). Serving orders: AB, BA	**Correct response** - taster picks the sample that is higher (e.g., the presumed sweeter sample). **Significance** - Required no. of correct/total responses[2]: As for *Duo-trio*
Same/ different	7	When a difference is unknown	Two coded test samples (A,B).	Tasters assess both samples and indicate whether they think samples are the same or are different. Serving orders: AB, AA, BA, BB (Note: two serving orders are presented to each taster)	**Correct response** - taster correctly picks the two samples as being the same or different, depending on the serving order. **Significance** - Required no. of correct/total resonses[2]: As for *Duo-trio*
Paired preference	7	Which wine is preferred	Two coded test samples (A,B).	Tasters assess both samples and indicate which one they prefer. A choice must be made; the taster can't say they prefer neither. Serving orders: AB, BA	Count the number of people who prefer one wine over another (e.g., A over B). **Significance** - Required no. preferred A/total[4]: Single tasting 7/7 8/8 8/9 9/10 Repeated tasting 12/14 13/16 14/18 15/20

[1] Indicates the minimum number of tasters required for testing to achieve a statistically significant result ($p \leq 0.05$).
[2] Figures denote minimum number of correct responses required out of the total number of responses to conclude the wines are significantly different ($p \leq 0.05$) from each other.
[3] *Serving orders* denotes possible arrangements of the samples to be presented randomly to tasters.
[4] Figures denote required minimum number of tasters who agree on preference for one wine, out of the total number of responses, to conclude that one wine is significantly preferred ($p \leq 0.05$) over the other.
Adapted from Cowey and Travis (2008).

Selection of the appropriate difference test depends on many factors, including the following:

- objectives
- number of available tasters
- volume of wine available

Again, the most common discrimination methods include the triangle test, the paired comparison test, and the duo-trio test (ASTM, 1968; Meilgaard et al., 1991; Stone and Sidel, 1985). Although discrimination tests may be completed by a small number of panelists (10 to 12), statistical determination of differences is more enhanced with a greater number of responses. Analysis of these methods is made easy by the use of statistical tables from which results of the test may be quickly analyzed.

These tests are relatively simple for the panelists if the panelists are knowledgeable about the product and characteristics of interest. In each method, the panelist is forced to make a decision or choice among the products. The amount of information drawn from these tests is limited to a detection of difference. It is not possible to know the degree of differences that exist among products, or if the change in the characteristic affects acceptability of, or preference for, the product.

Triangle Test. Triangle tests are useful as a multi-purpose test. The taster is required to select the sample which is different. Triangle tests are often preferred, as they require fewer tasters, and there is a greater likelihood that a result will be genuine and not due to a chance effect.

The triangle test uses three samples to determine if an overall difference exists between two products. The three samples include two that are identical, and one that is different. The samples must be coded with individual three-digit numbers (derived from a random numbers table) and presented at one time to the panelists. The panelist is requested to identify the code on the scorecard representative of the odd sample.

This method requires the panelist to make a choice among the samples; the panelist has a 33% chance of simply guessing correctly. This test method has good applications in determining if a process change affects the overall product character. Fatigue is a factor, as panelists usually must re-taste several times. Adaptation may also occur as a result of re-tasting. It is recommended that no more than two sets (six samples) be evaluated at one testing session.

Interpretation is based on the minimum number of correct responses required for significance at a predetermined significance level, given the total number of responses received. The minimum number of correct responses may be found in statistical tables provided in several publications (ASTM, 1968; Meilgaard et al., 1991; Stone and Sidel, 1985).

For example, two wines processed under identical conditions are fermented in the bottle for 6 months at two different temperatures, 7°C (45°F) and 13°C (55°F), in order to determine if the warmer fermentation temperature results in a different product. Twenty-four panelists evaluate the two wines, using a triangle test.

A difference is wanted at a 5% level of significance ($p \leq 0.05$). Fifteen of the 24 panelists correctly identify the odd sample. Is there an overall difference between the product fermented at 7°C, compared with the product fermented at 13°C?

At the 5% level of significance with 24 panelists, a minimum of 13 correct responses were needed to determine that there was a significant difference between the samples. There is a 95% or greater probability that the product fermented at 13°C is noticeably different from the product fermented at 7°C.

Duo-Trio Test. Duo-trio tests are sometimes used instead of triangle tests to compare unknown differences between wines. Tasters are presented with a reference wine, and then two test wines; one wine is the same as the reference, and the other is the wine to be tested. Evaluators are asked to identify the sample that is the same as the reference wine.

This test might be preferred, as the evaluator has a reference wine to compare. Generally, people find it easier to evaluate with a reference standard. It can also be better for assessing red wines by palate, as there is less palate fatigue. One disadvantage of this test, versus a triangle difference, is that more tasters are required (see Table 1).

The reference sample may always be set as the same product (constant reference) or may be randomly chosen so that each product is represented (balanced reference). This test method is less efficient than the triangle test, with a 50% chance of guessing correctly, and requires a large amount of sample, but is frequently used when a flavor is complex or intense. Interpretation is easily determined from a statistical table (ASTM, 1968; Meilgaard et al., 1991; Stone and Sidel, 1985).

Paired Comparison Test. Paired comparison tests can be used when there is a known difference in chemical composition of the wines (a simple difference test), which requires a sensory assessment. For example, a higher VA is present. But does the wine have a spoilage character? Is the wine more volatile?

They can also be used to see if there is a directional difference for a single characteristic (“Which product is more acidic? Product 738 or Product 429?”) between the samples. This test can be useful when assessing alternative wine blends. The test requires the same amount of wine and tasters as the duo-trio test. Products are compared such that each sample is placed in the first tasting position an equal number of times. This test causes less fatigue and is frequently used for strongly flavored or complex products.

Whereas the triangle test provides only a 33% chance of guessing correctly, there is a 50% chance with the paired comparison test. Therefore, more panelists (at least 20) are recommended to complete this test. Interpretation is easily determined via a statistical table (ASTM, 1968; Meilgaard et al., 1991; Stone and Sidel, 1985).

Same/Different Tests. A same/different test is similar to the paired comparison test, however, it is used when the difference between two wines is unknown. Evaluators are asked to identify whether they think the two samples presented are the same or different. These tests are easy to set up, but more panel members are required, and evaluators must perform the test at least twice, receiving a different randomized serving order each time.

Descriptive Tests

Frequently, it is important to know how a wine changes with a new vineyard site, how intense a characteristic is, etc. Discrimination testing, which is easy to use, easy to interpret, and easy for panelists to complete, is initially used to determine that a difference does exist. Such methods cannot provide information about the description of those differences, though.

Descriptive evaluation methods are more difficult to complete and interpret, but provide much more information. They provide a quantitative measure of wine characteristics that allows for comparison of intensity between products, and a means of interpretation of these results. Examples of descriptive test methods include quantitative descriptive analysis (QDA®), flavor profile analysis, time-intensity descriptive analysis, and free-choice profiling (Hootman, 1992; Meilgaard et al., 1991; Stone and Sidel, 1985). QDA® is frequently used because it requires less training time than several of the other methods.

Quantitative Descriptive Analysis (QDA®). QDA® was developed by the Tragon Corporation and the University of California, Davis. Proper QDA® techniques require following strict methodology.

Descriptive analysis of wines using QDA® requires products with relatively similar characteristics. If possible, 10 to 12 panelists are selected to participate based on ability to discriminate, communication skills, and task comprehension. Using fewer trained panelists is possible for descriptive analysis, but individual responses have a greater effect on the mean scores.

During the training, the group is provided with different wines that represent the range of characteristics that may be tested during the actual test sessions (see Panelist Training). Wine characteristics are identified, definitions or descriptions of the characteristics are determined, and references exemplifying the characteristics are established. The reference standards are used to further train the panelists to identify the desired characteristics and learn to rate intensity levels.

A scorecard is developed by consensus of the panel that includes all characteristics of interest in the order in which they are to be evaluated. Each characteristic is rated on a 6-inch (15.2-cm) line scale with descriptors of “weak” and “strong” as endpoint anchors. The anchor terms are located 1/2 inch (1.27 cm) from each end of the line; an alternate method places vertical anchors 2.5 cm from each end (Hootman, 1992). Either method is satisfactory. It is important that panelists know they may use the entire line to mark their perception of intensity, even those regions toward the ends of the line.

Panelists continue training until the sensory specialist determines that everyone is demonstrating comprehension of the task, and all panel members can identify and rate each characteristic. Preliminary testing should occur to determine the reliability and validity of the individual panelists before initiating the actual testing (see Performance Evaluation).

After the training is completed, the panelists function independently in the evaluation setting. No more than six or seven attributes should be evaluated at each setting to avoid fatigue. If intense aroma characteristics are evaluated, fewer than six wine samples should be evaluated in one testing.

Products are evaluated for intensity of the characteristics on the scorecard. Panelists rate the intensity of each attribute by marking a vertical mark across the appropriate horizontal rating line. These marks are converted to numerical data by measuring the distance from the origin ("weak") of the line to the vertical mark.

Sometimes only one or two characteristics are of interest, so an entire profile is not necessary. This reduces the effort required by the panelist and the time needed for data handling and interpretation by the sensory specialist. A line scale as is used for QDA® is appropriate for this, or a category scale with seven to nine verbal descriptors may be used. The category scale is easier to use by panelists with less training than a QDA® panelist.

Data analysis is completed using a mixed model analysis of variance for treatment by subject, with replication (Hootman, 1992). To determine individual panelists' abilities to perceive differences among products, a one-way analysis of variance is completed for each panelist. This analysis can also be used to determine if an attribute is helpful in differentiating among wines.

Subsequent analysis, using a two-way analysis of variance design, is needed to determine product differences and interactions by the panel (Hootman, 1992). An independent statistical analysis is completed for each characteristic that is measured.

A graphic presentation, called a spiderweb plot, of all characteristics may be made to illustrate the differences and similarities of the descriptive profiles of the wine samples evaluated. This is accomplished by plotting the mean score for a given characteristic on an axis that represents the 15.2-cm line scale used on the scorecard (Hootman, 1992). Each axis extends from a center point like spokes on a wheel, and represents a single characteristic. The center point is equivalent to the low-intensity origin of the line scale, and the highest intensity is equivalent to the end of the axis.