A System for Classifying Sensory Attributes

Castura, J.C., Findlay, C.J. Compusense Inc.
(jcastura@compusense.com)

Data from a descriptive analysis panel sometimes fails to detect differences between products for one or more sensory attributes. Results might nonetheless be consistent with the best possible data; lack of discrimination could be meaningful information if no true sensory difference exists between products for the attribute, which might occur when all products fall within one just-noticeable-difference (JND) interval (Castura *et al.*, 2006).

A panel leader sometimes has prior knowledge of a product group and what level of performance is possible for particular sensory attributes within the product context. This information can help the panel leader to focus training in a relevant manner.

Systematic classification of sensory attributes based on the difficulty associated with identifying and scaling the attribute in specific product contexts can make this information available to the broader community of sensory practitioners.

Attribute Examples

The following system for defining attributes was proposed by Findlay *et al.* (2005):

Specific Standard – a primary reference exists that defines the attribute completely (*e.g.* sugar, salt)

Group of Attributes – a number of examples provide the definition for several related attributes (e.g. fruit, floral)

Verbal or Evocative – no specific reference can be used, but the concept can be communicated (e.g. barnyard, diesel)

Note that evocative attributes are not necessarily more difficult to scale than specific standards (Sulmont *et al.*, 1999).

Scaling Difficulty

The following system for defining the scaling difficulty was proposed by Findlay *et al.* (2005):

Full Scaling - the attribute can be measured across the full range with precision of <10% of the scale range (e.g. sucrose in juices)

Rankable - the attribute can be measured across the range for the product, may be detected at 2, 3 or 4 levels (e.g. bitterness in black coffee)

Off/On - the absence or presence of an attribute can be detected in the product, but it does not lend itself to scaling (e.g. TCA in wine)

The system is illustrated pictorially in Fig. 1.

Context Effects

The JND and both the detection and saturation threshold values will be influenced by the other components of any system (Findlay *et al.*, 2005).

Materials + Methods

Two panels were recruited and trained to evaluate white wine; one panel was composed of experienced red wine panelists (Panel T), the other of panelists with no experience in sensory analysis (Panel U). Each panel used the Wine Aroma Wheel (Noble *et al.*, 1984, 1987) to develop their own white wine lexicon over 5 days of training sessions of 2.5h each. Panels T and U used 110 and 76 line scale attributes, respectively. Four additional training sessions were used to apply best practices from conventional training and computerized feedback. At the conclusion of training, each panel evaluated the same 20 white wines in triplicate. The experiment is discussed in greater detail in Findlay *et al.* (2006).

Classification strategy

Begin with the assumption that no panel can identify or scale any sensory attributes. H_0: the panel will generate random data. Measurements will lack precision and products will not be discriminated.

For an attribute, a momentary psychometric function with threshold and saturation levels can be conceptualized for each panelist, even if measurement is elusive due to such factors as variation of response (Lawless & Heymann, 1988).

Identification

(i) Counting the number of repeatable panelists is a useful approach to determining how difficult an attribute is to identify.

Repeatability refers to the ability of individual panelists to repeat previous assessments, success in which normally involves the panelist's understanding of the sensory attribute being measure.

(ii) The size of the mean squares error (MSE) from the one-way anova is related to the repeatability of the panelist. Taking the square root of MSE and dividing by the panel mean for the attribute provides a unitless indicator of repeatability. Counting the number and proportion of repeatable and nonrepeatable panelists on a panel provides an indication of their ability to identify attributes.

Scaling

(i) A two-way mixed model anova will act as a gateway to post hoc tests.

(ii) When panel data are submitted to post hoc tests, quotients (range between the minimum and maximum product mean score, divided by Fisher's LSD) will be greater than 2 for any attributes that lend themselves better to scaling.

(iii) To validate quotients, the number and proportion of homogeneous subsets will be calculated. Products will be considered well discriminated for the *j*-th attribute if wines are separated into four or more homogeneous subsets by Tukey's HSD. Note that no post hoc tests will be run if p-wine > 0.1.

Results + Discussion

Selected attribute classifications for scalability are shown in Table 2. Selected examples are discussed below.

Example 1 — Sweet was a taste attribute evaluated by both Panel T and Panel U by mouth. It was demonstrated by a specific standard. Mean scores for wines ranged from 24.68 to 52.38 for Panel T, and 11.89 to 50.80 for Panel U. Calculations related to this attribute appear in Table 1. It was well discriminated by both the panel and individual panelists. Data for the wines evaluated by Panel T in this study provide sufficient evidence to reject H. This attribute can be classified as easy to identify and easy to scale (fully scalable), as indicated in Fig. 2.

Note that the preliminary classification of sweet taste as easy to identify and scale is intended to provide general guidance to the panel leader. While consistent with expectations (Findlay *et al.*, 2005), the classification should nonetheless be validated with data from other studies to ensure the accuracy and robustness of the classification. Exceptions, perhaps resulting from context effects in another group of wines, could be such that scalability and identification are quite difficult.

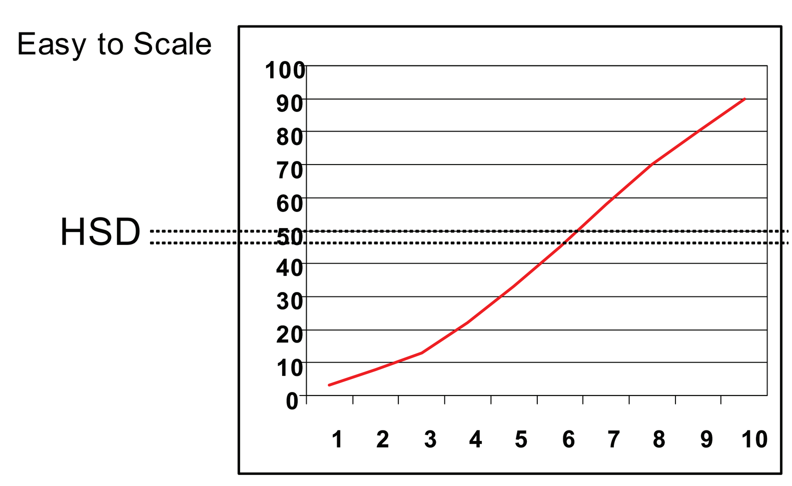

Fig. 2. Conceptual psychometric function for sweet taste.

Example 2 — Nail Polish Remover was an aroma attribute that Panel T evaluated by nose before stirring the wine. It was demonstrated by a commercial product. Mean scores for wines ranged from 3.97 to 7.94. The panel did not discriminate wines using this attribute (p=0.45). According to criteria discussed in footnotes to Table 1, there were no discriminators, 6 non-discriminators, and 2 non-detectors for this attribute. For the wines evaluated by Panel T in this study, there is no evidence to indicate that Nail Polish Remover (evaluated before stirring the wine) can be precisely identified or scaled.

Note that the preliminary classification of nail polish remover aroma before stirring as difficult to identify and scale might have resulted from the absence of specific defects in the 20 wines submitted to evaluation. As with all attribute classifications, results from other wines in the product category could lead to a reclassification of the attribute.

Example 3 — Asparagus was an aroma attribute that Panel T evaluated by nose before stirring the wine. It was demonstrated by natural examples. Mean scores for wines ranged from 0.94 to 9.21. The panel differentiated wines using this attribute (p=0.00); 3 homogeneous subsets were identified and quotient=2.58. Individually, 5/12 panelists discriminated wines at p=0.1, but 12/12 panelists had relatively low repeatability. Results indicate that this attribute is relatively difficult to identify, but that it is nonetheless possible to scale in this attribute in the aroma before stirring sensory modality for the product group.

The preliminary classification of hard to identify, yet possible to scale (rankable), might seem counter-intuitive. Should identification not be a pre-requisite for scalability? All wines had relatively low values on the scale, a near-threshold condition that made consistent identification more difficult, with a corresponding negative impact on panelist repeatability.

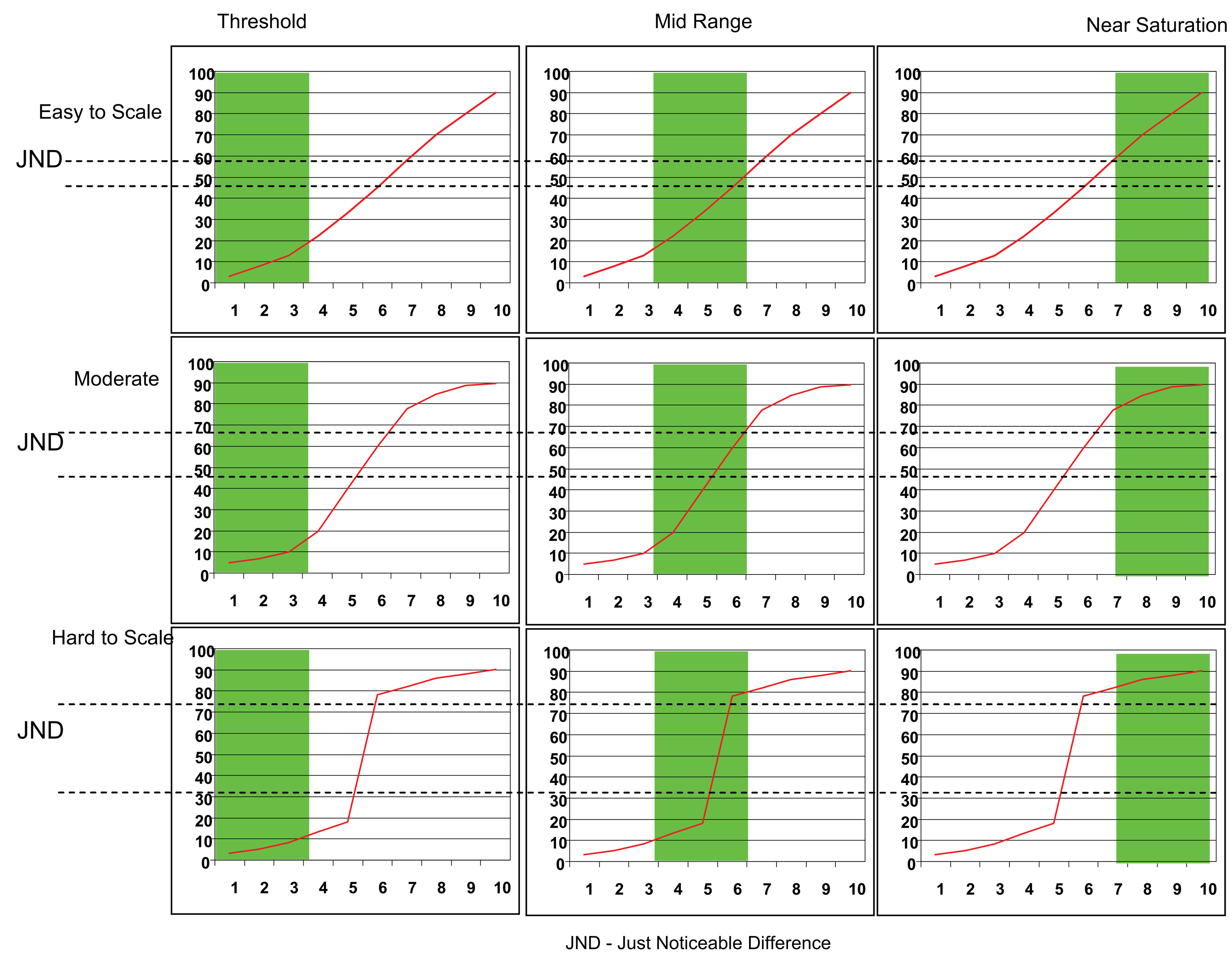

Fig. 1. Cases that describe attribute scaling properies (adapted from Findlay *et al.*, 2005).

Table 1. Calculated values for the sensory attribute Sweet (taste).

	p-value[a]	Quotient[b]	Subsets[c]	Discriminators[d]	Non-Discriminators[e]	Non-Detectors[f]
Panel T	0.0000	4.87	7	8 / 12	1 / 12	0 / 12
Panel U	0.0000	6.01	7	9 / 10	1 / 10	1 / 10

[a] p-value (wines) was calculated in SPSS 9.0 using a two-way mixed-model anova with interaction.
[b] Quotient was calculated by dividing the product mean range by Fisher's Least Significant Difference, where range is the distance between the maximum and minimum product mean scores and Fisher's LSD was calculated as a post hoc test to the anova[a] at p=0.05. Higher quotients reflect a greater ability to discriminate wines using the attribute. Pearson's Coefficient of Concordance indicated that quotients were well correlated with the range [b] (r=0.90 for Panel U, r=0.92 for Panel T).
[c] Subsets is a count of homogeneous subsets determined by Tukey's HSD at p=0.05. Tukey's HSD was calculated as a post hoc test to the anova[a]; one homogeneous subset was recorded when p>0.05. When p(wine)>0.05, Homogeneous subsets were used to validate quotients[b] and to assist in the identification of "Off-On" attributes.
[d] Discriminators were the number of panelists who discriminated wines according to one-way anova (p<0.1)
[e] Non-Discriminators were the number of panelists that satisfied the following conditions: (i) did not discriminate wines according to one-way anova (p<0.1) and (ii) the panelist's MSE divided by the panel average MSE > 2.0. Non-discriminators represented large sources of experimental error due to lack of session-to-session repeatability.
[f] Non-Detectors were the number of panelists that scored the attribute zero for all products.

Example 4 — Rose was a flavor attribute that both Panel T and Panel U evaluated by mouth. It was demonstrated by natural examples. Mean scores for wines ranged from to 14.15 to 20.51 for Panel T, and from 9.86 to 15.05 for Panel U. Neither panel discriminated wines using this attribute (p=0.75 for Panel T, p=0.17 for Panel U). According to criteria discussed in footnotes to Table 1, there were Panel T had 3 discriminators and 1 non-discriminator, while Panel U had no discriminators and no non-discriminators. For the wines evaluated by in this study, there is no evidence to indicate that Rose flavor can be identified or scaled. Since Rose flavor was discriminated by Panel U both before and after stirring, it is possible that the background of flavors against which Rose flavor was evaluated made it difficult to precisely identify or scale in this sensory modality.

Table 2. Preliminary classification of a selection of a subset of attributes for relative scalability.

Attribute	Quotient Panel T	Quotient Panel U	Scaling
Sweet [t]	4.87	6.01	scalable
Burn [m]	4.41	4.38	scalable
Alcohol [f]	4.39	3.88	scalable
Pungent [f]	4.16	3.44	scalable
Sour [t]	3.99	2.74	rankable
Smooth [m]	3.71	3.56	rankable
Bitter [t]	3.15	3.08	rankable
Astringent [m]	2.89	2.91	rankable
Earthy [f]	2.84	2.58	rankable
Oak [f]	2.53	3.13	rankable
Vinegar [f]	2.23	3.81	rankable
Melon [f]	2.08	2.98	rankable
Honey [a]	1.90	2.46	rankable
Apple [f]	1.83	3.07	rankable
Pineapple [f]	1.80	3.32	rankable
Pear [b]	1.73	1.75	on-off
Smoky [f]	-	3.56	scalable
Grape [f]	-	3.32	scalable
Black Pepper [f]	-	3.14	scalable
Caramel [f]	-	2.91	rankable
Cloves [f]	-	2.90	rankable
Medicinal [f]	-	2.53	rankable
Cedar [b]	-	2.51	rankable
Sulphur [b]	3.20	-	rankable
Elderflower [b]	3.01	-	rankable
Sulphur [f]	2.90	-	rankable
Vanilla [a]	2.79	-	rankable
Sulphur [a]	2.71	-	rankable
Musty [f]	2.61	-	rankable
Asparagus [b]	2.58	-	rankable
Pear [f]	2.58	-	rankable
Prickling [m]	2.52	-	rankable

Products were evaluated in the following modalities: [b]aroma before stirring, [a]aroma after stirring, [t]taste, [m]mouthfeel, and [f]flavor.

A system for classifying sensory attributes based on sensory criteria requires further research. Classifications developed for two white wine panels must be validated or updated using addition data from this and other product categories. Further study is needed to determine how attributes might be affected by different contexts. Automation and refinement of the classification methodology could assist in the classification process.

References

Castura, J.C., Findlay, C.J., & Lesschaeve, I. (2005). Monitoring calibration of descriptive sensory panels using distance from target measurements. *Food Quality & Preference*, 16, 682-690

Findlay, C.J., Castura, J.C., Schlich, P., & Lesschaeve, I. (2006). Use of Feedback Calibration to reduce the training time for wine panels. *Food Quality & Preference*, 17, 266-276.

Findlay, C.J., Castura, J.C., & Phipps, K. (2005). Setting meaningful attribute targets for feedback training of descriptive panelists (poster). In The 6th Pangborn Sensory Symposium, 7-11 August, Harrogate, Yorkshire, UK.

Lawless, H.T., & Heymann, H. (1998). Sensory Evaluation of Food, Principles and Practices. (1999). Gaithersburg, MD: Aspen Publishers, Inc. (a Chapman & Hall Food Science Book).

Noble, A. C., Arnold, R. A., Masuda, B. M., Pecore, S. D., Schmidt, J. O., & Stern, P. M. (1984). Progress towards a standardized system of wine aroma terminology. *American Journal of Enology and Viticulture*, 35, 107-109.

Noble, A.C., Arnold, R.A., Buechsenstein, J., Leach, E.J. Schmidt, J.O., and Stern, P.M. (1987). Modification of standardized system of wine aroma terminology. *American Journal of Enology and Viticulture*, 38, 143-146

Sulmont, C., Lesschaeve, I., Sauvageot, F., Issanchou, S. (1999.) Comparative training procedures to learn odor descriptors: effects on profiling performance. *Journal of Sensory Studies*, 14, 4, 467-490.

Acknowledgements

National Research Council – IRAP … Vincor International … Liquor Control Board of Ontario
Karen Phipps … Isabelle Lesschaeve … Amanda Bartel … Pascal Schlich … Compusense Panelists