Harmonizing Sensory Evaluation Internationally

A European sensory and consumer study on coffee demonstrates that sensory attributes used by a trained panel in one country can be used to understand consumer preferences in other countries

JEAN A. McEWAN

The author is Sensory and Consumer Research Manager, Dept. of Consumer and Sensory Sciences, Campden & Chorleywood Food Research Association, Chipping Campden, Gloucestershire GL55 6LD, United Kingdom

Within the European Union (EU), the market for consumer goods now extends to more than 350 million consumers. This is a huge potential market made even more accessible by the removal of barriers to trade within the member states of the EU. However, the market is diverse, with major economic and cultural differences among countries and regions. For food and drink manufacturers and retailers, understanding how products relate to this market is key to their success.

Sensory science is increasingly seen as providing the tools to aid companies to understand those product characteristics which are important in determining consumer likes and dislikes. A combination of sensory analysis with trained panels and market research with consumers is recognized by many companies as a valuable step in product optimization, to ensure that product development is targeted and focused.

In the context of sensory analysis and international trade, little work has been reported on comparisons of descriptive sensory profiling by different laboratories in different countries, and how this relates to consumer acceptability and preference. It is recognized that sensory analysts in different laboratories and countries have different approaches to descriptive profiling, yet all are considered experts within the field. It was not known whether some aspects were more important to control than others, and whether there would be consequences for sensory science if different laboratories produced different results for the same products. For food manufacturers selling products throughout the EU, it is clearly important to establish whether the description of products for development and control can be performed in one country even if products are distributed in other countries, and whether the profiling data obtained in a particular country can be related to consumer preferences in other countries.

An interlaboratory study was therefore carried out by the European Sensory Network (ESN), an organization whose members are committed to working together to develop and improve sensory methodology for the benefit of the European food industry.

Descriptive Analysis of Coffee

To establish whether panels operating in different organizations in Europe would characterize products in the same way, it was essential to ensure that all panels received identical products and that all preparation procedures were standardized. It was also thought important that the product selected for the study was universally consumed by all countries taking part in the exercise.

Products. Samples of high-quality roast ground coffee were selected to represent the range of coffees typically consumed in Europe. Advice on sample selection was provided by the International Coffee Organisation (ICO), which also carried out coffee bean roasting and sample preparation under controlled conditions. One coffee was roasted to three different levels to represent the degree of roast which could be encountered in the marketplace. All roasted beans were ground to the same degree. Measured amounts of coffee were vacuum packed to maintain freshness and to ensure controlled concentration of coffee in sample preparation. One coffee was packed at three different weights to ensure three different concentrations to cover the different strengths of coffee nor-

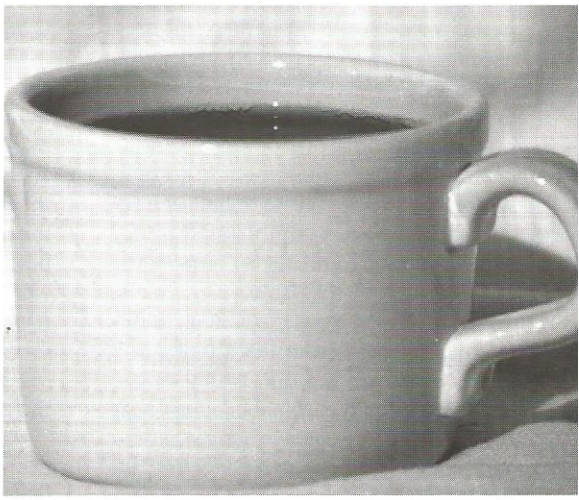

Table 1 The 16 coffee samples used for the sensory profiling, with the preferred consumer samples identified

Sample	Country	Roast	Preference
1	Brazil	Light-Medium	
2	Brazil	Light-Medium	Yes
3	Columbia	Medium	
4	Indonesia	Medium-Dark	Yes
5	Uganda	Medium	
6	Zaire	Medium-Dark	
7	Hawaii	Medium	
8	Ethiopia	Medium	Yes
9	Guatemala	Medium	
10	Various	(Decaffeinated)	Yes
11	Kenya	Light	Yes
12	Kenya	Medium	Yes
13	Kenya	Dark	Yes
14	Brazil/ Costa Rica	Medium (40 g/L)	
15	Brazil/ Costa Rica	Medium (55 g/L)	Yes

mally consumed. In all, 16 samples were used in the study (Table 1).

Sample Preparation. All samples used in the study were prepared in the same way. Each participant received four new brewing machines of the same type, and used these exclusively for the study. The procedure for brewing coffee was standardized by ICO and followed by all participants. Bottled Evian water was used to prepare coffee, thereby excluding the effect of regional water variability. Participants used different methods to hold samples after brewing and prior to tasting, with the requirement to serve samples at a temperature of between 50 and 60°C. Brewed samples were held for a maximum of 20 min and then discarded if not used by the panel.

Panels. Eleven panels from eight different countries (including three from France and two from the U.K.) took part in the study, with each panel comprising 7–12 judges. Most panels had no experience in descriptive analysis of coffee, although most had experience in descriptive analysis of other products. The panel from ICO had been exclusively used for descriptive analysis of coffee and had many years of experience.

Training of panels was left largely to individual panel leaders. Standards and training samples were provided by ICO, together with a list of 25 descriptors with definitions in English and French, although most panels developed their own vocabulary of coffee flavor and after-flavor characteristics. Some panels also developed vocabulary to describe coffee odor.

All evaluations were carried out in facilities designed for sensory analysis, and appropriate measures were taken to minimize sample color differences during the evaluations. The procedure for sample evaluation was not standardized, and although ICO advised spitting out samples after assessment, some panels chose to swallow or let it be optional. Most panel leaders required judges to clean their palates between samples.

Evaluations were carried out by all panels in sessions of four samples per session, although the number of sessions per day and the timing of the session varied from participant to participant. However, participants followed the same experimental design, ensuring that the same samples were grouped in sessions for all panels, and three replicate judgments were made on each sample by each judge.

How the Coffees Compared

Describing the differences and similarities between products is the main purpose of descriptive analysis. In this study, much was known about the characteristics of the products from previous work carried out by ICO. It was important, therefore, to establish two things in the analysis of the results:

1. To establish whether the panels characterized the samples in the same way, even though they were using different attributes and different languages.

2. To make sure that this characterization "made sense" with what was already known about the samples.

The multivariate statistical procedure, principal component analysis (PCA), was used to summarize the key information in the data while still explaining the differences between samples that are described by the data. Biplot representations of the data for each participant for dimensions 1 and 2 were produced. For example, Figs. 1 and 2 show the results for the Norway and U.K. panels; the numbers refer to the samples in Table 1. Samples which are close together on the biplot are perceived as similar by the panel, and those furthest apart are the most different. However, two samples can be close together on the first dimension, but far apart on the second dimension. To understand why, one needs to look at the direction and position of the attributes from the origin. The length of the attribute vector from the origin indicates the contribution to the structure of the sample map: the longer the line, the more important the attribute in contributing to the description of the product characteristics. The angle the vector makes with the dimension indicates its contribution to sample separation along the dimension: the smaller the angle, the greater the contribution.

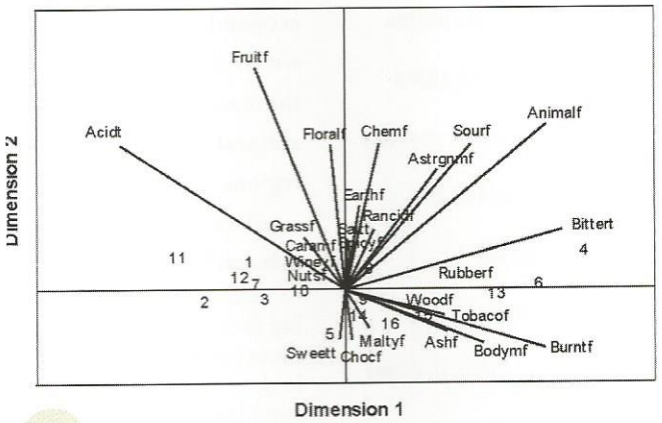

Fig. 1—Sample and attribute biplot derived from principal component analysis (PCA) of the sensory profile data from the Norwegian panel

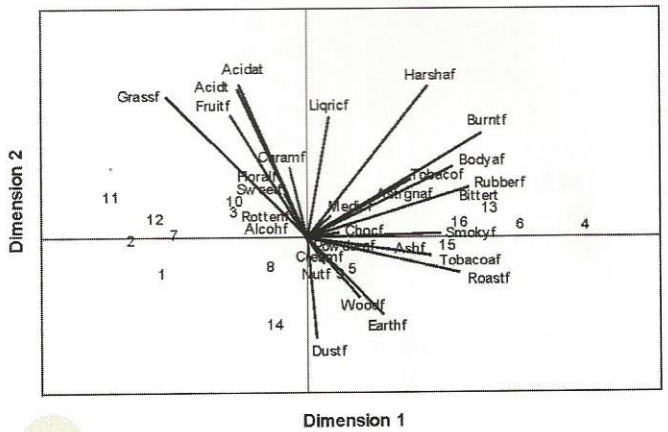

Fig. 2—Biplot for the U.K. panel

and 2 in terms of attributes was not possible, as each participant used different attributes, different numbers of attributes, and different languages to describe the coffee. However, if the panels were delivering the same "overall information" about the samples, then it should be possible to obtain similar sample structures from each panel. In this instance, the sample structure for each participant was very similar, with the main differences being in the degree of separation of the samples.

Looking at the individual panel plots in more detail, it was apparent that dimension 1 represented the "bitter vs non-bitter" of the coffees profiled. Thus, on the bitter side, it was always common to find the two Robusta coffees, the washed Java (4) from Indonesia and the unwashed Zaire (6). Also, because of its dark roast, the washed Arabica Kenyan (13) was always found on the bitter side. This is expected, since Robusta coffees, either washed or unwashed, develop a very characteristic profile, with bitter as one of the main attributes. Similarly, the dark-roasted, washed Arabicas will lose most of their acidity at that roasting level. Even if Robustas are roasted lighter than the roast used for this study, most of them would still develop a bitter character which will not be found in Arabica coffees roasted to the same degree.

The non-bitter (or less-bitter) coffees were the Brazilian Arabica (2) processed by the "semi-dry" field processing method, the light-roasted Kenyan (11), and the Colombian washed Arabica from the Medellín coffee-growing region (3). These coffees are typically characterized by a medium/high acidity. Acidity is an attribute which, in the coffee trade, is mostly considered to be independent of bitterness; i.e., they are two different dimensions. It does not necessarily follow that non-bitter coffees are acidic. For most panels in this study, dimension 1 was represented by the bitterness of coffee, along with burnt, smoky, ashy, tobacco, rubbery, chemical, roasted, strong, and woody, while acidity was a main attribute on dimension 2, together with green, sweet, floral, and fruity characters.

As expected, the unwashed Arabica from Brazil (1) was characterized by most panels as having low bitterness, either low or medium acidity, and some of the attributes caramel, sweet, spicy, green, herbal, nutty, and floral. In contrast, the other unwashed Arabica, Djimma 5 from Ethiopia (8), was apart from the unwashed Arabica from Brazil (1) on the biplots, indicating that it was characterized in a different way. This particular coffee is always highly sought after by gourmet traders.

The four other washed Arabica coffees, Uganda Bugisu (5), Hawaiian Kona (7), Guatemalan (9), and decaffeinated Colombian/Kenyan blend (10), were mostly positioned toward the center of the biplots, although there was always a clear separation between them for most panels. For example, for the ICO panel, the decaffeinated sample (10) was close to the attributes nutty and caramel, while sample 9 was closer to the chemical, cereal, earthy, and woody attributes. On the other hand, 7 was more fruity/winey and close to sweet and malty and 5 was more bitter and close to the attribute ashy. While these four coffees were roasted to the same degree, their sensory profiles are influenced, of course, by the climatic conditions and husbandry practices at origin.

Coffees 14, 15, and 16, which represented three different brewing concentrations of the same coffee, showed a clear increase in bitterness as the concentration increased by the panels from Denmark, Holland, Germany, and Sweden. The panels from France 2, ICO, and the U.K. differentiated these three coffees more clearly along dimension 2, which indicates that these panels detected more differences in more attributes than just an increase of intensity along the bitter/burnt/roasted dimension. The other panels, France 1, France 3, Norway, and Poland, did not clearly separate between the two highest concentrations, meaning that the panels did not find major differences between the samples.

For all panels, there was a clear picture of the changes in sensory profile of the Kenyan coffee as the degree of roast changed from light (11) to medium (12) to dark (13). With respect to bitterness, the dark roast was always clearly separated, whereas the differences between the light and medium roasts were less clear. For some panels, acidity was high for light roast, decreased at medium roast, and completely disappeared for dark roast. Other panels, however, tended to find a stronger developed acidity with medium roast, and slightly less acidity with more sweet, nutty, caramel, and malty notes with light roast.

A Common European Vocabulary

One objective of the study was to try to establish whether panels operating in different organizations in Europe would characterize products in the same way. The initial interpretation of the biplots outlined above confirms that panels are able to characterize the main variation within the sample set in a similar way. From what is known in the coffee trade about the quality characteristics of different types of coffee, they are able to describe the sample differences in a way which makes practical sense. From an industrial perspective, these results demonstrate the objective nature of descriptive analysis using trained panels, and that the description of products for development and control could be performed in one country even if products are distributed in other countries.

A second objective was to bring together attributes and profiles from each of the different European participants, and in so doing, propose a common vocabulary that would be understood by all panels. Detailed analyses of the data were carried out to establish whether a common meaning was a reasonable and jus-

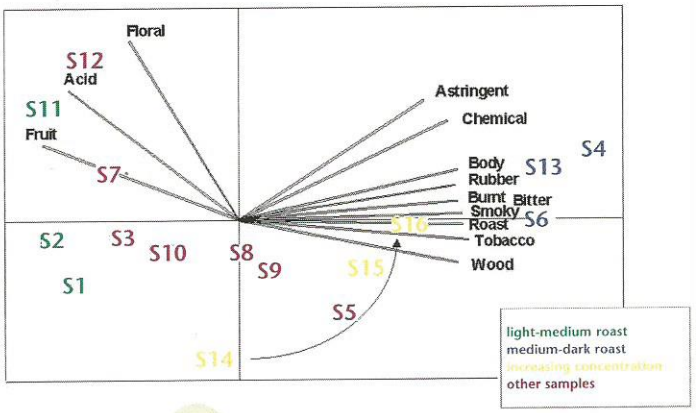

Fig. 3—Biplot for the sensory profile data associated with the devised European vocabulary

tified assumption for bringing attributes together, or whether different panels placed different emphasis on particular attributes. The details of these analyses have been published in the 1996 report, "A European Sensory and Consumer Study: A Case Study on Coffee," published by the European Sensory Network and available from Campden & Chorleywood Food Research Association, Chipping Campden, Glos., U.K.

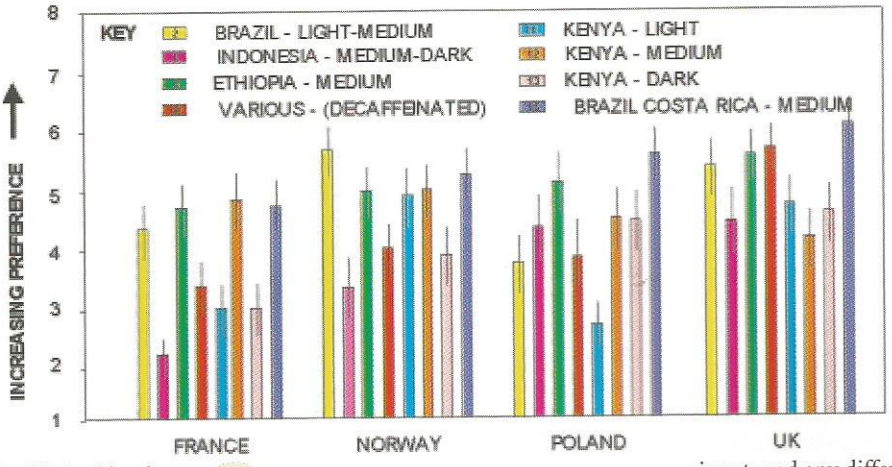

Fig. 4—Histogram of the sample hedonic mean scores for four of the eight consumer panels

It was possible to propose a common vocabulary of 13 attributes: bitter, burnt, rubber, roast, chemical, strength of coffee flavor, acid, fruity, citrus, sour, rancid, astringent, and floral. A PCA was performed on this selected European vocabulary of 13 attributes, and the results are presented as a biplot in Fig. 3. Dimension 1 was described as a bitter/burnt/roast axis, while dimension 2 separates samples on acid/fruity/floral characteristics. In terms of sample structure, this European biplot is visually similar to those obtained from individual panels, and the relationships between the attributes also reflected those of the individual panels.

Relating Consumer Preferences to Sensory Attributes

For food and beverage companies in Europe, understanding product differences is not enough. Companies recognize the economic importance of understanding the cultural differences between consumers across Europe and, more precisely, identifying which sensory characteristics of foods and beverages are most desirable for European consumers. The second part of the interlaboratory study was to investigate whether consumer preferences for coffee were the same for all countries, whether there were groups of European consumers who had the same likes and dislikes, and whether these could be explained in terms of sensory descriptions identified by the trained panels.

Following the descriptive analysis, it was clear that the coffee sample map could be adequately represented by fewer samples, so eight coffees were selected for the consumer trials based on the descriptive profile sample maps. Samples 4 and 13 were chosen to represent the group of three samples (4, 6, and 13) positioned at the far right of dimension 1. Samples 11 and 12 represented the samples at the far left of dimension 1 and were chosen because they were light and medium roasts of the same coffee. Sample 8 was chosen to represent the "middle" of dimension 1. The other three samples selected were 2, 10, and 15. Selection of a smaller number of samples had practical benefits in the organization of subsequent consumer trials.

Eight separate consumer trials were carried out in Poland, Sweden, Denmark, Norway, Germany, France 1, France 3, and the U.K. Essentially the same experimental procedures were used by each participant, and any differences in the way in which samples were presented to consumers, and the questions asked, were recorded. All participants used a 9-point hedonic scale and aimed to recruit 70–80 consumers (51 achieved in Denmark).

Simple histograms (such as that shown in Fig. 4 for four countries) illustrating the hedonic mean score for each sample made it easy to observe similarities and differences between panels. For example, the washed Indonesia Robusta (4) was scored highly by Polish and U.K. consumers, yet was least liked by consumers in other countries. Also, the Brazilian/Costa Rican/Mexican blend (15) was on the whole the most acceptable sample across all countries.

It was more interesting, however, to segment the consumers into groups with similar likes and dislikes, rather than by country. The data from all 605 consumers were analyzed by cluster analysis, and eight clusters were selected. The number of consumers falling within each cluster was 62–99, and it was possible to see clear differences in preference between the clusters, as well as the proportion of consumers from each country within each cluster.

For example, cluster 1 (Fig. 5) highlighted the decaffeinated coffee (10) as the most acceptable, with a clear indication of a U.K. influence—approximately 25% of the U.K. consumers were in this cluster, representing one

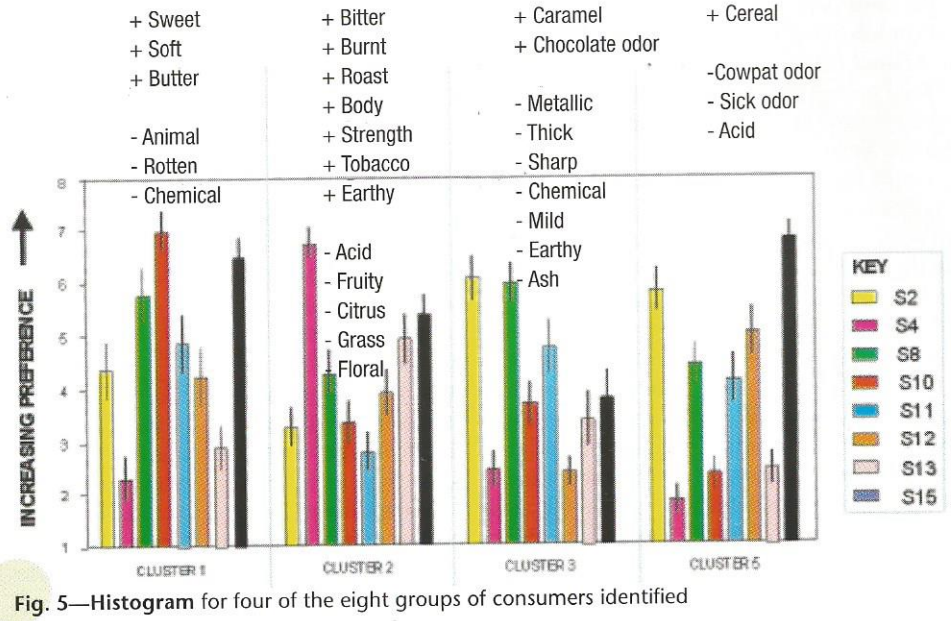

Fig. 5—Histogram for four of the eight groups of consumers identified

third of the total number of consumers in the cluster. Similarly, the Indonesian Robusta (4) was the most acceptable coffee for cluster 2, indicating that even though it was the least liked coffee overall, there was a group of consumers which had a positive preference for it (Poland 36% and U.K. 24%).

A PCA was performed on the mean sample scores, where the eight clusters each represented one group of consumers (Fig. 6). To interpret the biplot, it should be realized that the lines represent directions of increasing preference. To determine which samples are most liked for a given preference segment, projections from the preference line to each of the samples should be made, as illustrated for cluster 2 in Fig. 6. From this biplot, it can be seen that consumers in clusters 3 and 6 preferred the Brazilian Arabica medium/light roast (2) and the Kenyan Arabica light roast (11). Clusters 2 and 5 had preference vectors in the opposite direction, indicating that these consumers preferred the Kenyan Arabica dark roast (13), the Indonesian Robusta medium/dark roast (4), and the Brazilian/Costa Rican/Mexican blend medium roast (15), and that clusters 1 and 8 were very closely related.

One of the key objectives of undertaking the consumer research was to seek to establish whether consumer preference data collected in one country could be related to descriptive sensory attributes generated by a sensory panel in another country. Once again, this would have distinct advantages to industry who wished to develop and manufacture products in one country and be confident that they met the particular requirements of consumers in their target market.

In Fig. 6, the sensory attributes from the common vocabulary were projected onto the preference map, to represent directions of increasing attribute intensity. It can be seen that sample 4, the Indonesian medium-dark coffee, and sample 13, the Kenyan dark, had chemical and astringent characteristics, which were generally negative for preference. Consumers in clusters 3 and 6 preferred coffees with acid, fruity, and floral notes (e.g., sample 2, Brazilian light-medium), while consumers in clusters 2 and 5 liked bitter, burnt, roasted coffees with more body (e.g., sample 13, Kenyan dark).

Applicable to Other Products

We recognize that the results of this interlaboratory study are specific to roast and ground coffee, and that the interpretation of the data is specific to the samples within the trial. Nonetheless, this was an extensive study, and the methodological results have clear implications for the whole of the European food and drink industry, not to mention for studies in America, Australasia, etc.

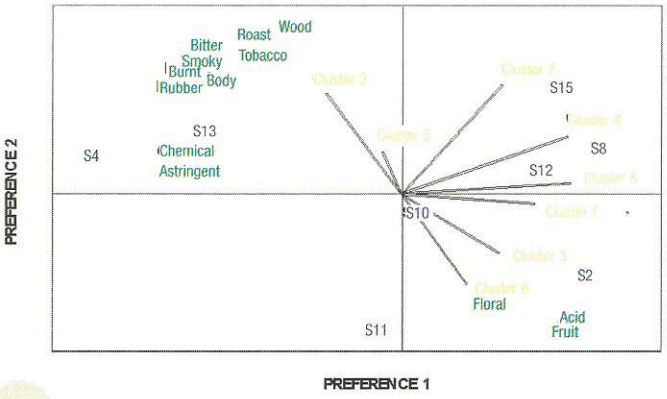

Fig. 6—PCA biplot for the eight samples selected for the consumer trial and the main directions of preference

Nevertheless, the methods used in this study are directly applicable to other products, and a number of important conclusions can be drawn:

- Descriptive profiling using trained sensory panels is an objective tool which can be used effectively in product development and control for definition of product characteristics.
- Panels operating "best practice" for the selection and training of sensory panels for descriptive analysis are able to discriminate and characterize sample differences in the same way. Therefore, it is possible to profile products in one country, confident that the interpretation of the results can be understood in other countries.
- It is possible to develop a common vocabulary for products which can be used by different sites throughout Europe, and which ensures a common understanding between panels at different sites.
- It is possible to establish consumer preferences for products in different countries, or groups of consumers with similar likes and dislikes, and relate these to sensory descriptive analysis in a way which means that directions of common preference can be obtained together with the attributes driving preference.

This work was always intended as an initial study which will raise as many questions as answers. For example, this study did not investigate in detail why consumers fall into particular segments and what common characteristics they share. Nor did it address the issue of communication of sensory claims. In what way, for example, should a company inform consumers about the sensory properties of its product, what kinds of expectations are raised by this information, does the product come up to these expectations, and if not, why not?

Furthermore, access to a large and complete data set has allowed some novel statistical techniques to be applied and further developed, giving further information on the development and application of sensory methodology.

Edited by Neil H. Mermelstein, Senior Editor