Food Quality and Preference 21 (2010) 309–318

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

How reliable are the consumers? Comparison of sensory profiles from consumers and experts

Thierry Worch [a,b,*], Sébastien Lê [b], Pieter Punter [a]

[a] OP&P Product Research, Burgemeester Reigerstraat 89, NL-3581 KP Utrecht, The Netherlands
[b] Laboratoire de Mathématiques Appliquées, AGROCAMPUS OUEST, 65 Rue de Saint Brieuc CS 84215, 35042 Rennes Cedex, France

ARTICLE INFO

Article history:
Received 4 August 2008
Received in revised form 5 February 2009
Accepted 2 June 2009
Available online 6 June 2009

Keywords:
Experts versus consumers
Panels comparison
Multiple factor analysis
Confidence ellipses

ABSTRACT

This study compares expert and consumer sensory profiles for the same 12 perfumes in different ways: the discriminatory ability and reproducibility are analyzed through ANOVA and the panelists' consensus through the correlation coefficients. Next, the two product spaces are first analyzed separately for each panel, and then compared through multiple factor analysis. Finally, the two panels are compared using the confidence ellipses methodology. These analyses show that the two panels give similar results with respect to the important criteria for panels (discrimination, consensus, reproducibility). The comparison of the two products spaces shows high similarity. From the confidence ellipses, it can be concluded that no significant differences exist for a given product between the two panels. Hence, in this particular case, the use of consumers appears to be a good alternative to the classical sensory profile provided by a trained panel.

1. Introduction

In sensory analysis, one of the most important tools is the quantitative characterization of the perceivable product attributes. In the literature, this tool is referred to as "descriptive analysis", or "profiling" (two frequently used profiling methods are quantitative descriptive analysis (QDA®, Stone, Sidel, Oliver, Woosley, & Singleton, 1974) and Spectrum™ (Meilgaard, Civille, & Carr, 2006)). These methods use trained or expert panels. Because of their routinely use of the type of products in question, and because of dedicated training sessions, these panels seem to be more able to characterize products in an accurate way than naïve consumers. On the other hand, hedonic questions are also of great importance and most practitioners use consumers for hedonic tasks. So trained panels are required for sensory profiles and consumers are required for hedonic profiles. In the literature, many warnings are given concerning the use of consumers for profiling:

- "...as with any untrained panel, beyond the overall acceptance judgment there is no assurance that the responses are reliable or valid" (Stone & Sidel, 1993)
- "...consumers can only tell you what they like or dislike" (Lawless & Heymann, 1999)

According to these practitioners, profiling results from consumers lack two essential qualities: consensus between respondents and reproducibility.

Moreover, it has also been shown, that asking consumers liking and intensity questions in the same test can return an unwanted halo effects (Earthy, MacFie, & Hedderley, 1997). The potential impact of the attribute questions on the hedonic ratings is a key point in the objection of the use of getting sensory information from consumers. Since the aim of this paper is to compare experts' and consumers' sensory profiles, this point is not studied here.

In market research, most companies need quick answers about their products. Hence, they do not always have the possibility to train panels (which is time consuming). Profiles obtained with consumers can be a good alternative, depending on the type of tests one is interested in, especially in view of the fact that consumers' profiles also meet the requirements discrimination, panelists' consensus and reproducibility (Husson, Le Dien, & Pagès, 2001). Moskowitz also showed that consumers can be used to assess the sensory descriptions of sauces, and hence "refutes the notion that consumers are incapable of validly rating the sensory aspects of products" (Moskowitz, 1996).

Because of these two notions (training panels takes time, and consumers are not allowed to profile products), a number of faster methods for collecting sensory data have been developed. Among them is free-choice profiling (Williams & Langron, 1984), Flash profiling (Siefferemann, 2000, 2002), and ultra flash profiling (Perrin et al., 2008). These methods have in common that they avoid training sessions beforehand, and that they use naïve consumers (Gazano, Ballay, Eladan, & Sieffermann, 2005; Nestrud & Lawless,

* Corresponding author. Address: OP&P Product Research, Burgemeester Reigerstraat 89, NL-3581 KP Utrecht, The Netherlands. Fax: +31 (0) 30 254 0254.
E-mail address: thierry@opp.nl (T. Worch).

0950-3293/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodqual.2009.06.001

2008). Paradoxically, it is well accepted that consumers can be used for profiling products using these methods, but the use of consumers with standard QDA® type profiling using a fixed, predefined vocabulary is still subject to criticism. The question remains: How reliable are consumers? To answer this question, classical sensory profiles, obtained from an expert and a consumer panel on the same products, are compared.

2. Data

The datasets provided here concern 12 luxurious women perfumes. The list of the perfumes is given Table 1. These 12 perfumes were profiled by an expert and a consumer panel.

The expert panel was run in Agrocampus Ouest (Rennes, France) with 12 persons (eleven students and one teacher) from the Chantal Le Cozic School (esthetic and cosmetic school in Rennes). First, two focus groups, with two moderators, were conducted. Then, a summary discussion was conducted, and a list of 12 attributes was defined: "Vanille" (vanilla), "Notes Florales" (flower note), "Agrume" (citrus), "Boisé" (woody), "Vert" (green), "Epicé" (spicy), "Capiteux" (heady), "Fruité" (fruity), "Fraîcheur marine" (sea freshness), "Gourmand" (moreish), "Oriental" (oriental) and "Enveloppant" (wrapping).

Additionally, a specific training session for the most difficult attributes was performed (perfumed incense was used). The 12 products were then tested in duplicate, in two one-hour sessions. A 10 cm unstructured line scale was used for rating the products.

The consumer panel was run at OP&P Product Research (Utrecht, the Netherlands) with 103 naïve Dutch consumers (44 men and 59 women, 48 between 18 and 35 years old and 55 between 45 and 60 years old). The women all used luxurious perfume daily, and the men had a girlfriend or wife who used perfume regularly. Additionally, the men had to name at least two luxurious women perfumes. In this way, consumers who are (directly or indirectly) users of this type of product, were selected.

The vocabulary for consumers was based on the expert list of attributes. Since the consumers had no experience with specific perfume attributes, the list was adapted to make it understandable for naïve consumers by using descriptors found online (the perfume encyclopedia at www.osmoz.com). This resulted in a list of 21 attributes: "Odor intensity", "Freshness", "Jasmine", "Rose", "Camomile", "Fresh lemon", "Vanilla", "Mandarin/Orange", "Anis", "Sweet fruit/Melon", "Honey", "Caramel", "Spicy", "Woody", "Leather", "Nutty/Almond", "Musk", "Animal", "Earthy", "Incense" and "Green". Actually, this vocabulary has a large correspondence with the experts' one.

In order to measure the reproducibility of the consumer panel, two products (Shalimar and Pure Poison) were duplicated. The fourteen products (12 original and two replicates) were tested in two one-hour session (seven products in each session). A 100 mm unstructured line scale, with marks at 10% and 90% was used with consumers (EyeQuestion v2.2 software developed by Logic8).

In both tests, the experimental design was based on Latin Square balanced for first-order and carry-over effects (MacFie, Bratchell, Greenhoff, & Vallis, 1989).

Please note that the same line scales were used for experts and consumers, but for experts, the values were recorded between 0 and 10 while for consumers, they were recorded between 0 and 100.

Table 1
List of the products.

Product	Type
Angel	Eau de Parfum
Cinema	Eau de Parfum
Pleasures	Eau de Parfum
Aromatics Elixir	Eau de Parfum
Lolita Lempicka	Eau de Parfum
Chanel N5	Eau de Parfum
L'Instant	Eau de Parfum
J'Adore (EP)	Eau de Parfum
J'Adore (ET)	Eau de Toilette
Pure Poison	Eau de Parfum
Shalimar	Eau de Toilette
Coco Mademoiselle	Eau de Parfum

3. Unidimensional aspects

To measure the quality of the two panels, the following unidimensional measures have been computed: the product discrimination and the panel reproducibility through ANOVA and the panelists' consensus through the correlations between each panelist and the average of the panel without that panelist.

3.1. Expert panel

For each of the 12 attributes, an ANOVA is performed using the following model:

$$Y_{ijk} = \mu + \alpha_i + B_j + C_k + \alpha B_{ij} + \alpha C_{ik} + BC_{jk} + \varepsilon_{ijk} \quad (1)$$

where Y_{ijk} is the scores for the product i given by the consumer k at the session j, μ is the constant, α_i is the effect of product i, B_j is the effect of the session j (set as random), C_k is the effect of the panelist k (set as random), αB_{ij} is the effect of interaction between product i and session j, αC_{ik} is the interaction between the product i and the consumer k, BC_{jk} is the effect of interaction between the session j and the consumer k and ε_{ijk} is the residual.

In this case, the product effect expresses the discriminatory ability, while the interaction "product by session" expresses the reproducibility of the expert panel. The results of the F-tests for the discrimination and for the reproducibility are shown Table 2. The consensus between panelists is usually estimated through the interaction of panelist by product (Latreille et al., 2006). But as only two products were replicated for the consumers, a proper estimation of the interaction of consumers by product is not possible. Hence for both expert and consumer panels, the panelists' consensus is estimated through the correlation coefficients calculated between the data for each panelist and data averaged over the rest of the panel (11 experts or 102 consumers). These data are considered as vectors, by rearranging the scaled data of each matrix on a single column. For the expert data, the averaged table over the two sessions is taken into consideration (vectors of length 12 products × 12 attributes = 144). The distribution of these coefficients is presented Fig. 1a.

The expert panel results show:

- the panel can discriminate on all aspect except for "Agrume" and "Fraîcheur marine";
- the panel is reproducible on ten out of 12 attributes;
- there is a high consensus between the panelists, with an averaged correlation around 0.70 (the correlations range between 0.4 and 0.8).

3.2. Consumer panel

The discrimination ability of the consumer panel is measured on the 12 original products only. For each attribute, an ANOVA is performed using to the following model:

$$Y_{ik} = \mu + \alpha_i + C_k + \varepsilon_{ik} \quad (2)$$

Table 2
P-values from ANOVA model for the expert panel for the product effect and for the "product × session" interaction.

Effect	Attribute					
	Epice	Capiteux	Fruite	Vert	Vanille	Notes Florales
Product	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Product × session	0.213	0.353	0.904	0.876	0.579	0.038
Effect	Attribute					
	Boise	Agrume	FraTcheur Marine	Gourmand	Oriental	Enveloppant
Product	0.001	0.090	0.530	<0.001	<0.001	<0.001
Product × session	0.555	0.564	0.012	0.615	0.981	0.394

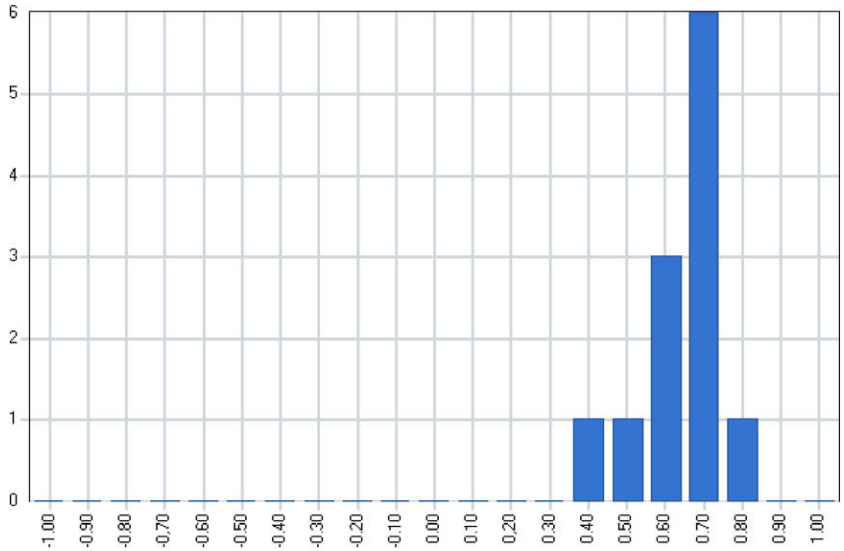

Fig. 1a. Distribution of the correlation coefficients showing the panelist's consensus for the expert panel.

where Y_{ik} is the scores for the product i given by the consumer k, μ is the constant, α_i is the effect of product i, C_k is the effect of the panelist k (set as random) and ε_{ik} is the residual. The results are shown Table 3.

The reproducibility of the consumers is measured on the two duplicated products. For each attribute, the ANOVA model from Eq. (1) is used. The results are also shown Table 3. It is completed with the results of the spider plots shown for the two duplicated products (Fig. 2a and b).

Finally, the consumer consensus is estimated through the correlations coefficients calculated between each consumer and the averaged of the 102 other consumers. Again, these two tables are considered as vectors, by rearranging the scaled data of each matrix in a single column (vectors of length 12 products × 21 attributes = 252). The distribution of these coefficients is presented Fig. 1b.

The consensus can also be seen in a multidimensional way with respect to the consonance analysis (Dijksterhuis, 1995). For each attribute, a PCA is run on a matrix where each row refers to a product and each column refers to the attribute considered given by one consumer. Hence, the matrix is of dimension $P \times C$, with P the number of products and C the number of consumers. The consonance analysis results for the attributes with the highest consensus (*woody*) and with the lowest consensus (*jasmine*) for the consumer panel are shown Fig. 3a and b.

Table 3
P-values from ANOVA model for the consumer panel for the product effect (12 products) and for the "product × session" interaction (two replicates).

Effect	Attribute						
	Intensity	Freshness	Jasmine	Rose	Camomile	Fresh Lemon	Vanilla
Product (n = 12)	<0.001	<0.001	0.016	<0.001	0.619	<0.001	<0.001
Product × session (n = 2)	0.987	0.646	0.571	0.587	0.484	0.431	0.612
Effect	Attribute						
	Citrus	Anis	Sweet Fruit	Honey	Caramel	Spicy	Woody
Product (n = 12)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Product × session (n = 2)	0.268	0.356	0.702	0.563	0.216	0.694	0.010
Effect	Attribute						
	Leather	Nutty	Musk	Animal	Earthy	Incense	Green
Product (n = 12)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Product × session (n = 2)	0.543	0.156	0.763	0.462	0.140	0.112	0.304

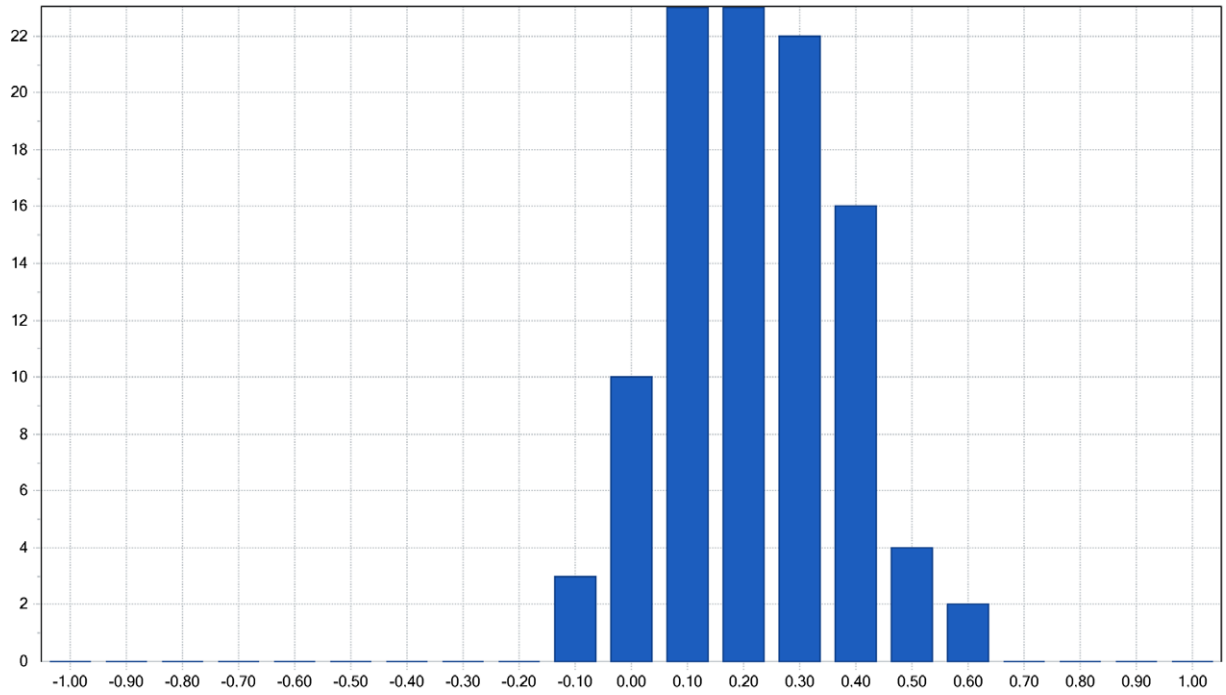

Fig. 1b. Distribution of the correlation coefficients showing the panelist's consensus for the consumer panel.

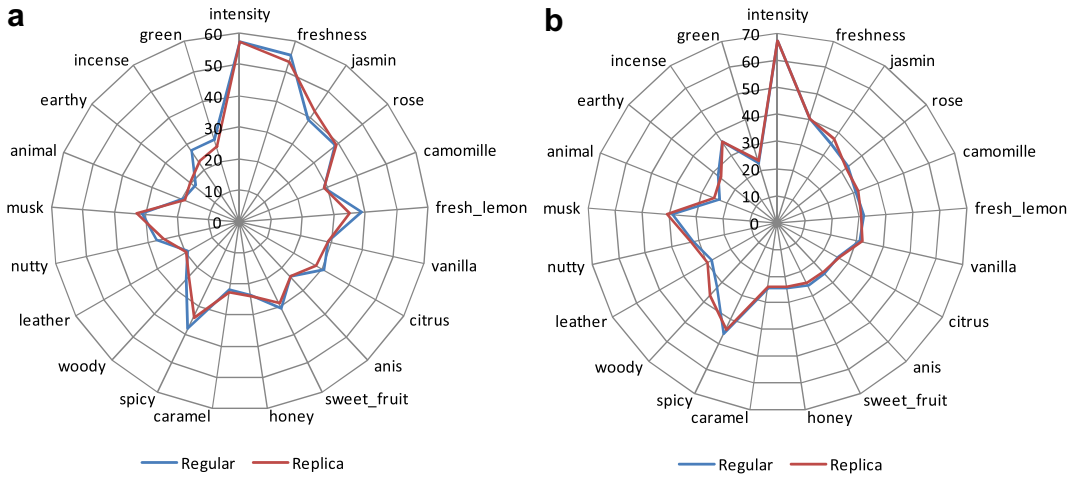

Fig. 2. Spider plots of the sensory profiles given by consumers for the products Pure Poison (left) regular and replica and for Shalimar (right) regular and replica.

The consumer panel results show:

- the products are discriminated on all attributes, except for "camomile";
- the consumers are reproducible on all attributes, except for "woody". The spider plots shown Fig. 2a and b confirm this finding;
- the correlation coefficients range between −0.1 and 0.60, with an average around 0.25;
- despite the high variability due to the use of consumers, the consonance analysis of the attribute with the higher panelists' consensus (*woody*) show a structure that the attribute with the lower (*jasmine*) panelists' consensus does not show (Fig. 3a and b).

3.3. Conclusion on the unidimensional analyses

The two panels show similar qualities with respect to discrimination and reproducibility. There is a difference between the two panels in the correlation coefficients: they are higher for the experts than for the consumers. This is most likely caused by the fact that the consumers were not trained. Moreover, for experts, we took the average over sessions, and looked at the correlation between 144 values, while for the consumers we looked at the correlations between 252 values. These reasons probably explain the difference observed between the two panels. But since the values are positive, we may conclude that both panels show consensus, as the averaged correlation coefficients for both panels are significant (with 100° of freedom (i.e. 102 observations), the correlation coefficients become significant at 5% when the values is higher than 0.195).

4. Multidimensional aspects

For each panel, the product space is computed by Principal Components Analysis on the products' profile, where one profile is a table crossing the products (i) in rows and the attributes

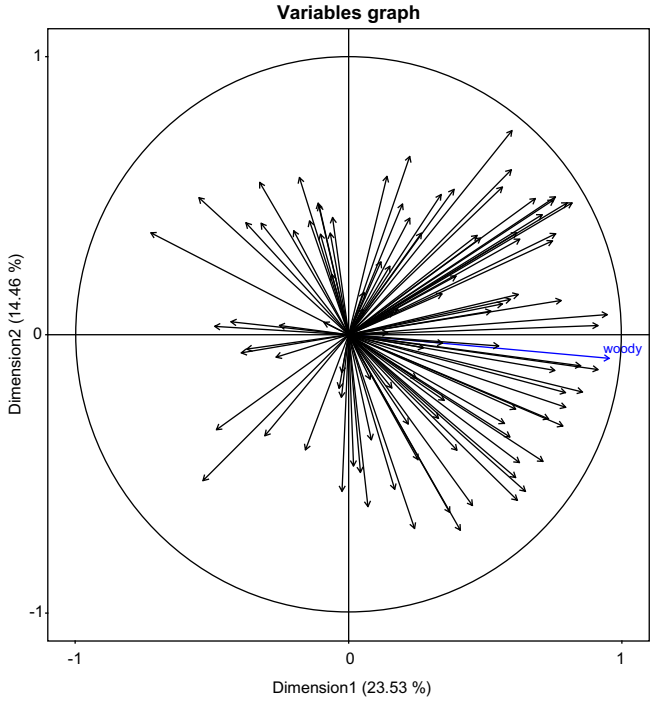

Fig. 3a. Consonance analysis showing the consensus between consumers for the attributes *woody* (one arrow represents the attributes *woody* given by one consumer). In blue, the average over the panel for this attribute is projected as illustrative. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

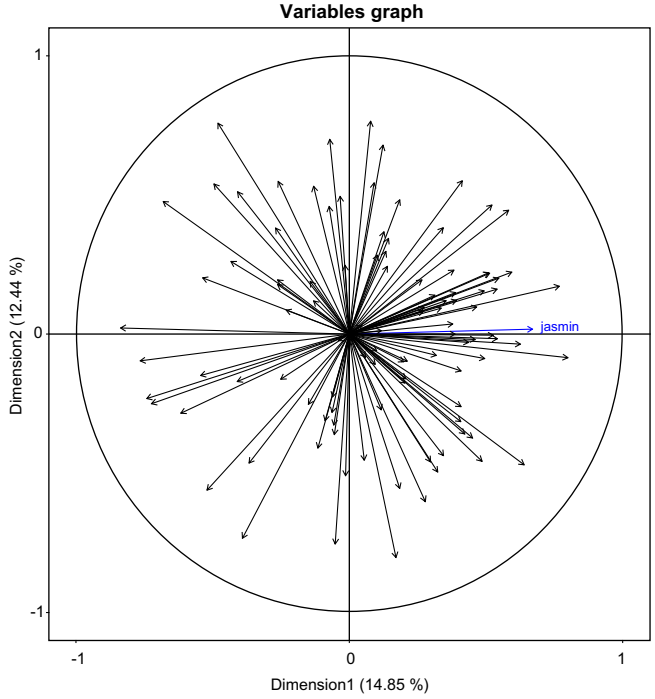

Fig. 3b. Consonance analysis showing the consensus between consumers for the attributes *jasmine* (one arrow represents the attributes *jasmine* given by one consumer). In blue, the average over the panel for this attribute is projected as illustrative. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

($k^{(\text{panel})}$, with $k^{(\text{expert})} = 12$ and $k^{(\text{consumer})} = 21$) in columns, and the cell (i, $k^{(\text{panel})}$) is the average score for the product i and the attributes $k^{(\text{panel})}$.

In order to compare the results given by the experts with those given by the consumers, the two products spaces are submitted to multiple factor analysis (Escofier & Pagès, 1998). As a complement, they are also submitted to Generalized Procrustes Analysis (Gower, 1975). To do these panel comparisons, a complete table, which is the juxtaposition of the two panel's sub tables, is created.

4.1. Expert products' space

This space shows different clusters of products (Fig. 4):

- the first dimension (64.21% of the total inertia), opposes the products Pleasures, J'Adore ET and EP (high intensity ratings in "Notes Florales", "Vert", "Agrume", "Fraîcheur Marine", "Fruité") to Aromatics Elixir, Shalimar and Angel (high intensity ratings in "Épicé", "Oriental", "Capiteux", "Enveloppant");
- the second dimension (21.86%) opposes Aromatics Elixir and Shalimar (high intensity ratings in "Boisé") to Lolita Lempicka and Angel (high intensity ratings in "Vanille", "Gourmand").

4.2. Consumer products' space

This space also shows different clusters of products (Fig. 5):

- the first dimension (68.29% of the total inertia), opposes the products J'adore ET and EP and Pleasures (high intensity ratings in "Citrus", "Sweet Fruit", "Freshness", "Green", "Jasmine", "Rose", "Fresh Lemon") to Angel, Shalimar and Aromatics Elixir (high intensity ratings in "Nutty", "Animal", "Musk", "Incense", "Leather", "Woody", "Earthy", "Spicy", "Intense");
- the second dimension (17.97%) opposes Aromatics Elixir (high intensity ratings in "Intensity") to Lolita Lempicka (high intensity ratings in "Vanilla", "Honey", "Caramel", "Anis").

4.3. Stability of the product spaces

The significance of the consumer product space is measured using a permutation test (Wakeling, Raats, & MacFie, 1993). For each consumer, the products are redistributed randomly: the configuration for each consumer is kept identical, but the names of the products are reorganized. Next, new average products table over consumers data are computed and submitted to a PCA. The percentage accounted for the two first dimensions is extracted. This methodology is repeated many times (in practice a hundred times) and the distribution of the percentage accounted for the first plan of the PCA is drawn. If the percentage of our first plan (68.29 + 17.97 = 86.26%) is in the 5% upper limit of this distribution, then we can conclude that the PCA map is significant for the consumers.

This simulation showed that the consumers product space is highly significant, as any simulation can get a percentage of inertia on the first plan higher than 57.04% (Fig. 6). In our case, the variance explained by the first dimension only is higher than this value (68.29%).

The permutation test done for the expert product space shows similar results.

4.4. Products' spaces comparisons

The comparison of the two products' spaces can be treated as a multi-block analysis, as for example Generalized Procrustes Analysis (GPA), multiple factor analysis (MFA), common components and specific weights analysis (CCSWA) (Qannari, Wakeling, Courcoux, & MacFie, 2000) or STATIS (Lavit, Escoufier, Sabatier, & Trais-

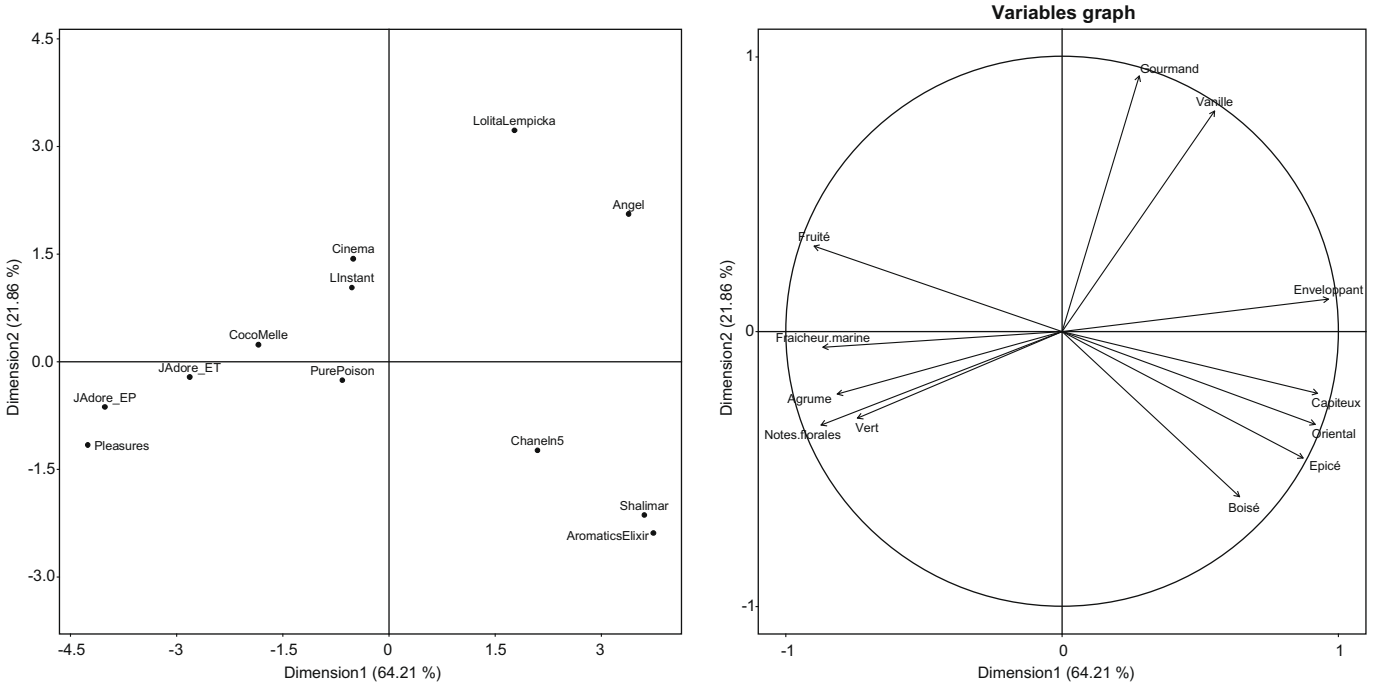

Fig. 4. product space (left) and variables' representation (right) for the experts.

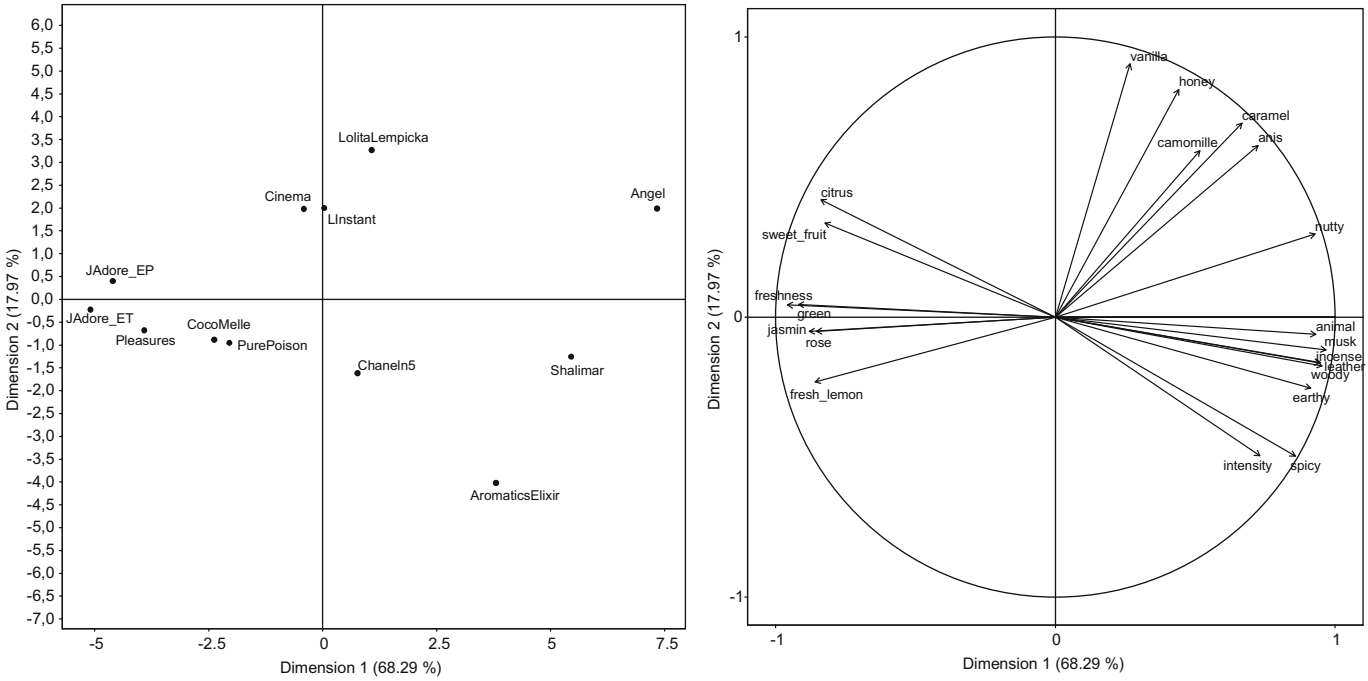

Fig. 5. Product space (left) and variables' representation (right) for the consumers.

sac, 1994). The choice was made to compare them through MFA. As a complement, they are also compared through GPA.

The comparison of the expert and consumer product spaces through MFA is shown Fig. 7. It shows that these two products' spaces are really close, not to say identical. The RV coefficient (Escoufier, 1973) calculated between the two configurations is 0.87. The high relation between these two spaces is confirmed by GPA (the similarity coefficient is equal to 0.93), which shows similar results as MFA (therefore the GPA results are not shown).

The MFA partial points' representation shows that the products with the more variability are Lolita Lempicka, Shalimar (the expert panel discriminates these products better in the second dimension) and Angel (the consumer panel shows a better discrimination for this product in the first dimension). Concerning the description of the products, the variables' representation shows high correlations between the equivalent attributes (i.e. "Fraîcheur Marine" and "Freshness"; "Épicé" and "Spicy"; "Vanille" and "Vanilla"). The comparison of the attributes *spicy* and *épicé* (Fig. 8a) confirms this statement (correlation = 0.97). Nevertheless, a disagreement between both panels for some attributes such as *boisé* and *woody* (Fig. 8b) is observed. In this particular case, the disagreement is clearly due to the product Angel which is described as intense by

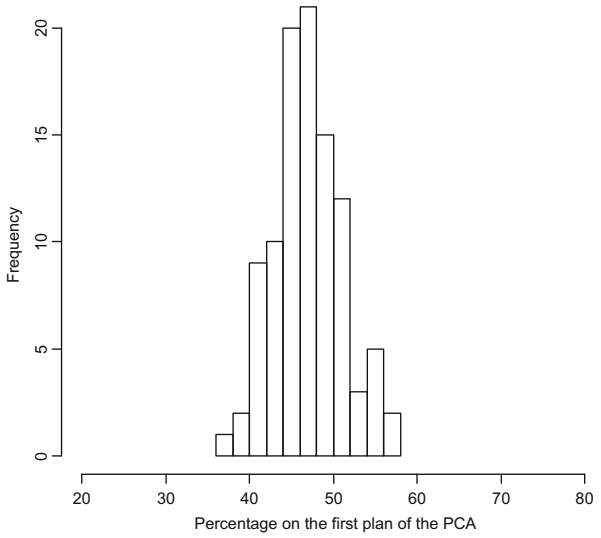

Fig. 6. Distribution of the percentage of inertia explained by the first PCA plan obtained from permutation tests for the consumer data.

the consumers, and as not intense by the experts. Despite this disagreement, the correlation between both panels is high (correlation = 0.67), which shows that the disagreement is minimal and concerns only one product. In conclusion, no clear disagreement, resulting in non significant or negative correlation, between the two panels is observed.

4.5. Panels' comparison through the confidence ellipses technique

The confidence ellipses technique (Husson, Le Dien, & Pagès, 2005; Pagès & Husson, 2005) enables creation of graphical confidence intervals around the products. It can also be used to compare the profiles provided by different panels (Lê, Pagès, & Husson, 2008). These confidence ellipses have two important properties:

- if two ellipses are superimposed, the two corresponding products are not significantly different;
- the size of the ellipses is related to the variability existing around the corresponding products: the bigger the size, the more variability.

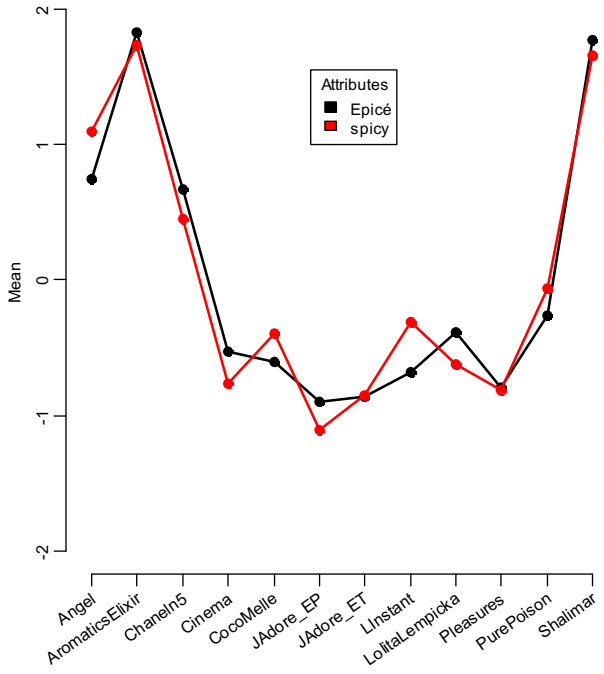

Fig. 8a. Comparison of the attribute *Epicé* (expert panel) and *Spicy* (consumer panel).

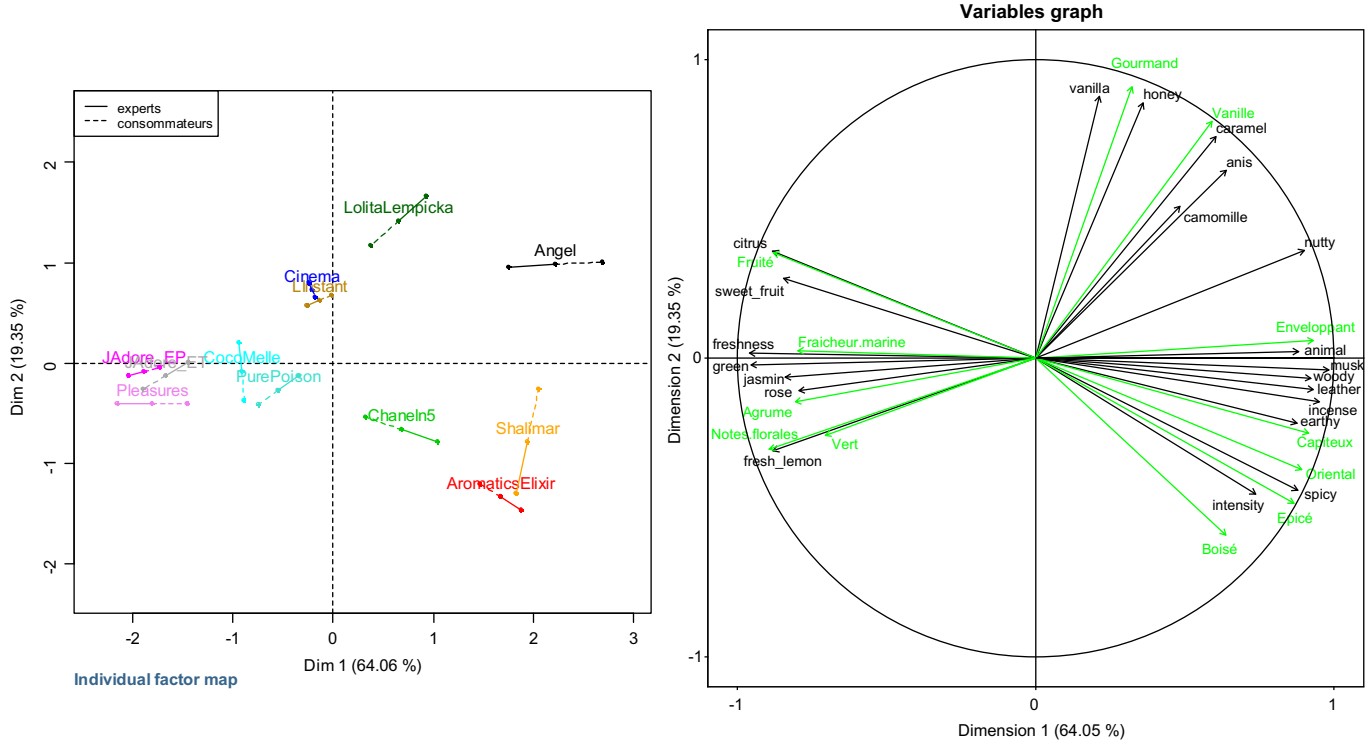

Fig. 7. Comparisons of the expert and the consumer product spaces through MFA: partial points representation (left) and variables' representation (right).

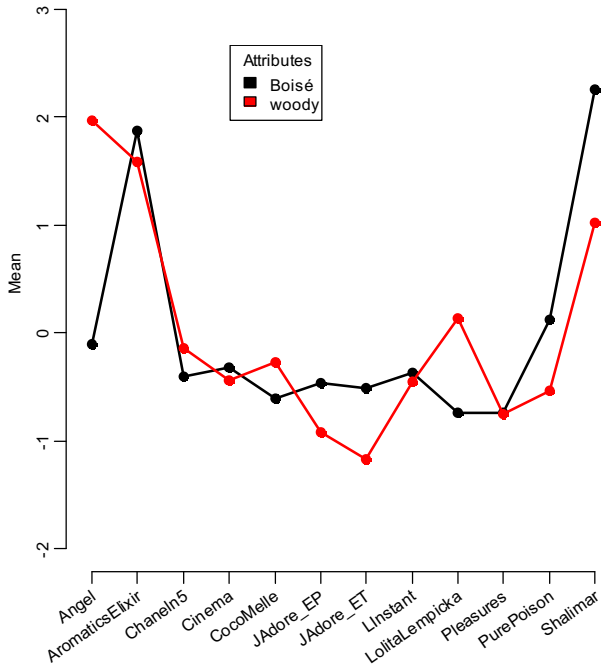

Fig. 8b. Comparison of the attribute *Boisé* (expert panel) and *Woody* (consumer panel).

With respect to the MFA partial points representation, one ellipse per product and per panel can be estimated. In this example, 24 ellipses are constructed (Fig. 9). It enables comparison of a given product for the two different panels (same color), or different products within a panel (same type of line). The confidence ellipses are accompanied with a Hotelling T^2 test (Table 4), which provides a *P*-value for each pair of products.

This analysis shows:

- overall, some products are clearly different (i.e. Angel and Pleasures) while some others are not (i.e. Coco Mademoiselle and Pure Poison);
- for each product, the two confidence ellipses related to the two panels are always superimposed: the products are not significantly different from one panel to another. This is confirmed by the *P*-values obtained from the Hotelling T^2 test shown Table 4;
- within a product, the two ellipses related to the two different panels have the same size: the higher amount of consumers compensates their higher variability;
- within the panels, pair comparisons of products show differences: the expert panel is able to discriminate 88% of the pairs while the consumers are able to discriminate 86% of the pairs. Moreover, some non significant pairs are common to both panels (i.e. Cinema and L'Instant), while some other are specific for one panel (i.e. J'Adore (ET) and Pleasures, or Aromatics Elixir and Shalimar).

5. Conclusions and comments

Both panels are similar in terms of discriminatory ability and reproducibility. By looking at the pair comparisons given by the confidence ellipses, the results are close, even though some specificity for each panel can be observed. In terms of panelists' consensus, the experts show more consistencies than the consumers: this might be due to the fact that they have a better knowledge about this type of product (through their experience and the training sessions). Moreover, they defined their own list of attributes (for the experts, the attributes they were not able to recognize were removed while for the consumers, all attributes were imposed), and the experts received additional training on the most difficult ones. Nevertheless, the consumers "created" a similar product space as the experts. A final difference between consumers and experts resides in the variability of the results: as the consumers are not trained, we can predict that their results show more variability. But the sizes of the confidence ellipses seem to show that the higher variability related to the consumers is compensated by the larger sample size.

In this particular experiment, the performance of naïve consumers does not differ from the performance of a trained expert panel. Consumers can describe the products in a reliable and repeatable way and do not differ from the trained experts. This supports the notion that consumers can also be used for classical profiling.

There are advantages and disadvantages in using consumers for classical profiling tasks. An advantage is that they use the same vocabulary which makes data analysis a lot easier. Another advantage is that they do not need training and can be employed at any moment in time. They can be recruited from a specific target group for the problem at hand and also provide hedonic information. Besides that, using consumers makes it possible to obtain addition information about the ideal intensities through Ideal Profiling or JAR scaling (Van Trijp, Punter, Mickartz, & Kruithof, 2007). The acquisition of sensory, ideal and hedonic information in one test from the target group can speed up product development considerably. Compared to the custom in market research where consumers are also asked hedonic and sensory questions in a Just

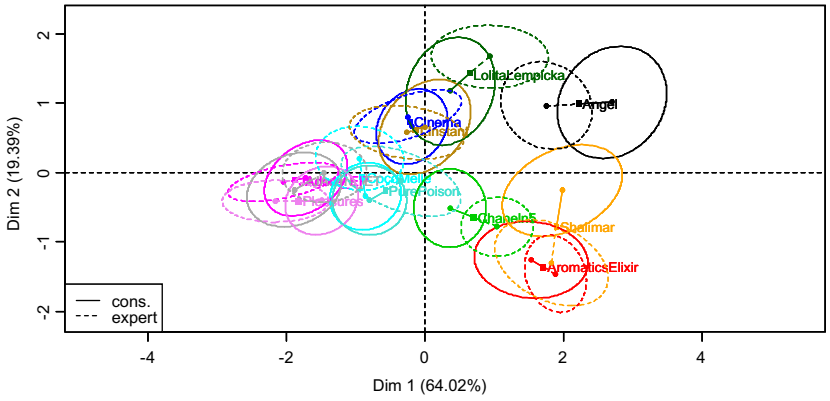

Fig. 9. Confidence ellipses for the comparison of the two different panels.

Table 4
P-values obtained from the Hotelling *T* test associated to the confidence ellipses.

	Angel	Aromatics Elixir	Chanel No 5	Cinema	Coco Melle	J'Adore (EP)	J'Adore (ET)	L'Instant	Lolita Lempicka	Pleasures	Pure Poison	Shalimar
By panel												
Expert												
Angel	1.00											
Aromatics Elixir	0.00	1.00										
Chanel No 5	0.00	0.00	1.00									
Cinema	0.00	0.00	0.00	1.00								
Coco Melle	0.00	0.00	0.00	0.03	1.00							
J'Adore (EP)	0.00	0.00	0.00	0.00	0.03	1.00						
J'Adore (ET)	0.00	0.00	0.00	0.00	0.28	0.36	1.00					
L'Instant	0.00	0.00	0.00	0.57	0.07	0.00	0.00	1.00				
Lolita Lempicka	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00			
Pleasures	0.00	0.00	0.00	0.00	0.00	0.25	0.05	0.00	0.00	1.00		
Pure Poison	0.00	0.00	0.00	0.00	0.29	0.00	0.05	0.02	0.00	0.00	1.00	
Shalimar	0.00	0.75	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
Consumer												
Angel	1.00											
Aromatics Elixir	0.00	1.00										
Chanel No 5	0.00	0.00	1.00									
Cinema	0.00	0.00	0.00	1.00								
Coco Melle	0.00	0.00	0.00	0.00	1.00							
J'Adore (EP)	0.00	0.00	0.00	0.00	0.00	1.00						
J'Adore (ET)	0.00	0.00	0.00	0.00	0.00	0.79	1.00					
L'Instant	0.00	0.00	0.00	0.82	0.00	0.00	0.00	1.00				
Lolita Lempicka	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.28	1.00			

By product								
Angel	Expert	Consumer	J'Adore (ET)	Expert	Consumer	Coco Mlle	Expert	Consumer
Expert	1.00		Expert	1.00		Expert	1.00	
Consumer	0.57	1.00	Consumer	0.80	1.00	Consumer	0.56	1.00
Aromatics Elixir	Expert	Consumer	L 'Instant	Expert	Consumer	J'Adore (EP)	Expert	Consumer
Expert	1.00		Expert	1.00		Expert	1.00	
Consumer	0.88	1.00	Consumer	0.94	1.00	Consumer	0.88	1.00
Chanel No 5	Expert	Consumer	Lolita Lempicka	Expert	Consumer	Pure Poison	Expert	Consumer
Expert	1.00		Expert	1.00		Expert	1.00	
Consumer	0.45	1.00	Consumer	0.64	1.00	Consumer	0.69	1.00
Cinema	Expert	Consumer	Pleasures	Expert	Consumer	Shalimar	Expert	Consumer
Expert	1.00		Expert	1.00		Expert	1.00	
Consumer	0.96	1.00	Consumer	0.52	1.00	Consumer	0.41	1.00

About Right format, the use of proper profiling has the considerable advantage that both perceived and ideal intensities can be obtained instead of a relative judgment.

Disadvantages of using consumers for classical profiling is the larger variability in the ratings due to the lack of training, for that reason the sample size of consumer panels should be much larger than for experts or trained panels. According to Moskowitz, the minimum base size to generate stable averages is 40–50 panelists per cluster of consumers (Moskowitz, 1997). He also states that beyond 80 panelists, the average is not particularly affected by the base size. Hence, where 10–12 trained panelists are required, we advise from our experience the use of 80–100 consumers. Consumers are also not suited for day-to-day tests within a company or for quality control. For these instances, a trained panel is much more efficient. There is also a limitation to the kind of attributes consumers can use, they are not able to use technical or highly specific attributes.

6. Software

The analyses were done with R 2.8.0 (R Development Core Team, 2008), with the packages SensoMineR v1.08 (Lê & Husson, 2006) and FactoMineR v1.10 (Husson, Lê, Josse, & Mazet, 2007), and with Senstools.

Acknowledgements

The authors would like to thank M. Cousin, M. Penven, M. Philippe and M. Toularhoat, students in applied statistics in Agrocampus Ouest (Rennes), who managed the expert study. They also would like to thank the reviewers for their constructive comments.

References

Dijksterhuis, G. (1995). Assessing panel consonance. *Food Quality and Preference, 6*, 7–14.
Earthy, P. J., MacFie, J. H., & Hedderley, D. (1997). Effect of question order on sensory perception and preference in central location trials. *Journal of Sensory Studies, 12*, 215–237.
Escofier, B., & Pagès, J. (1998). *Analyses factorielles simples et multiples*. Paris: Dunod.
Escoufier, Y. (1973). Le traitement des variables vectorielles. *Biometrics, 29*, 751–760.
Gazano, G., Ballay, S., Eladan, N., & Sieffermann, J. M. (2005). Flash profile and flagrance research: Using the words of the naive consumers to better grasp the perfume's universe. In *ESOMAR fragrance research conference, 15–17 May 2005*, New York, NY.
Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika, 20*, 33–51.
Husson, F., Lê, S., Josse, J., & Mazet, J. (2007). FactoMineR: Factor analysis and data mining with R. URL: <http://factominer.free.fr>.
Husson, F., Le Dien, S., & Pagès, J. (2001). Which value can be granted to sensory profiles given by consumers? Methodology and results. *Food Quality and Preference, 12*, 291–296.

Husson, F., Le Dien, S., & Pagès, J. (2005). Confidence ellipse for the sensory profiles obtained with principal component analysis. *Food Quality and Preference, 16*, 245–250.

Latreille, J., Mauger, E., Ambroisine, L., Tenehaus, M., Vincent, M., Navarro, S., et al. (2006). Measurement of the reliability of sensory panel performances. *Food Quality and Preference, 17*, 369–375.

Lavit, C., Escoufier, Y., Sabatier, R., & Traissac, P. (1994). The ACT (STATIS method). *Computational Statistics and Data Analysis, 18*, 97–119.

Lawless, H. T., & Heymann, H. (1999). *Sensory evaluation of food: Principles and practices*. New York: Kluwer Academic/Plenum Publishers.

Lê, S., & Husson, F. (2006). SensoMineR: Sensory data analysis with R. URL: <http://sensominer.free.fr>.

Lê, S., Pagès, J., & Husson, F. (2008). Methodology for the comparison of sensory profiles provided by several panels: Application to a cross-cultural study. *Food, Quality and Preference, 19*, 179–184.

Logic8. EyeQuestion: The web based software for sensory and consumer research. URL: <http://www.logic8.nl>.

MacFie, H. J., Bratchell, N., Greenhoff, K., & Vallis, L. V. (1989). Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *Journal of Sensory Studies, 4*, 129–148.

Meilgaard, M., Civille, G. V., & Carr, B. T. (2006). The SpectrumTM descriptive analysis method. In *Sensory evaluation techniques* (4th ed., pp. 189–253). CRC Press.

Moskowitz, H. R. (1996). Experts versus consumers: A comparison. *Journal of Sensory Studies, 11*, 19–37.

Moskowitz, H. R. (1997). Base size in product testing: A psychophysical viewpoint and analysis. *Food Quality and Preference, 8*, 247–255.

Nestrud, M. A., & Lawless, H. T. (2008). Perceptual mapping of citrus juices using projective mapping and profiling data from culinary professionals and consumers. *Food Quality and Preference, 19*, 431–438.

Pagès, J., & Husson, F. (2005). Multiple factor analysis with confidence ellipses: A methodology to study the relationships between sensory and instrumental data. *Journal of Chemometrics, 19*, 138–144.

Perrin, L., Symoneaux, R., Maître, I., Asselin, C., Jourjon, F., & Pagès, J. (2008). Comparison of three sensory methods for use with the Napping® procedure: Case of ten wines from Loire valley. *Food Quality and Preference, 19*, 1–11.

Qannari, E. M., Wakeling, I. N., Courcoux, P., & MacFie, H. J. (2000). Defining the underlying sensory dimensions. *Food Quality and Preference, 11*, 151–154.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. ISBN: 3-900051-07-0, URL: <http://www.R-project.org>.

Siefferemann, J. M. (2002). Flash profiling. A new method of sensory descriptive analysis. In *AIFST 35th convention*, July 21–24, Sidney, Australia.

Sieffermann, J. M. (2000). Le profil Flash-un outil rapide et innovant d'évaluation sensorielle descriptive. In *AGORAL 2000, XIIèmes rencontres "L'innovation: De l'idée au succès"* (pp. 335–340). Montpellier, France.

Stone, H., & Sidel, J. L. (1993). *Sensory evaluation practices*. California: Academic Press.

Stone, H., Sidel, J., Oliver, S., Woosley, A., & Singleton, R. C. (1974). Sensory evaluation by quantitative descriptive analysis. *Food Technology, 28*, 24–34.

Van Trijp, H. C. M., Punter, P. H., Mickartz, F., & Kruithof, L. (2007). The quest for the ideal product: Comparing different methods and approaches. *Food Quality and Preference, 18*, 729–740.

Wakeling, I. N., Raats, M. M., & MacFie, H. J. H. (1993). A new significance test for consensus in generalized procrustes analysis. *Journal of Sensory Studies, 7*, 91–96.

Williams, A. A., & Langron, S. P. (1984). The use of free-choice profiling for the evaluation of commercial ports. *Journal of the Science of Food and Agriculture, 35*, 558–568.