Proceedings of the IMProVe 2011
International conference on Innovative Methods in Product Design
June 15th – 17th, 2011, Venice, Italy

Understanding a particular semantic dimension: how selecting the products for evaluation tests?

Emmanuel Aliouat [a], Jean-François Petiot [b], David Blumenthal[c], Marc Danzart[c], Jean-Marc Sieffermann[c]

[a] Renault, Direction de la Recherche, Groupe Perception et Analyse Sensorielle DREAM/DTAA, Technocentre Renault, 1 avenue du Golf 78288 Guyancourt CEDEX, France.
[b] Institut de Recherche en Communications et Cybernétique de Nantes (UMR CNRS 6597), Ecole Centrale de Nantes, 1, Rue de la Noë, BP 92101, 44321 NANTES, France.
[c] Agroparistech, Laboratoire de perception Sensorielle et Sensométrie 1 Avenue des Olympiades, 91744 Massy Cedex, France

Article Information

Keywords:
Sensory Science,
Sensorial Design,
Product Space,
Subjective Evaluation,
Features Picking.

Corresponding author:
Emmanuel Aliouat
Tel.: +33 1 76832791
e-mail:
emmanuel.aliouat@renault.com
Address: 1 avenue du Golf 78288 Guyancourt CEDEX, France

Abstract

In the industrial field, manufacturers frequently aim to improve acceptation by targeting global trends of consumers' liking (to create a product that looks innovative, robust…). Sensory Science could support this task by evaluating perception and tastes (e.g. in sensory profiles or in hedonic tests) or even semantic dimensions regarding products of the marketplace. Nevertheless, the choice of these products may have significant consequences on the reliability of the results (and bring about biased conclusions). The control of this kind of effect is very tough because the choice of the products is imposed by constraints (costs, availability…). We attempt to develop a method to have a maximum of relevant information with a minimum of products. The present work deals with an experimental method, called "Features Picking", using consumers' feedback to define a product space. To illustrate the method, the proposed application concerns a specific semantic dimension: the fluidity of shape of car dashboards. In this case, 60 subjects selected "key products" according to the defined semantic dimension and attempted to verbalize which design features can explain it. Then, we assessed some statistical methods between hedonic and descriptive data to evaluate the efficiency of this approach. We finally highlighted the method giving the more relevant features with a reduced number of products. In this paper, we describe how we reduced a large set of 30 products to a subset of 11 products while keeping a maximum of information (87 % of the total information generated by the subjects).

1 Introduction

In a field such as the automotive industry, the consumers' feelings play a key role in determining his/her acceptation of a product. That is why manufacturers attempt to maximize this acceptation by giving some subjective specifications to their products (to create a product that looks innovative, robust…). Nevertheless, it remains difficult to traduce these global semantic dimensions in accurate requirements. Sensory Science could support this task by measuring sensory properties in order to predict the acceptation of products [1]. Using a set of products is usual to evaluate perception and tastes concerning concrete stimuli. Sensory studies are conventionally divided in two categories:
- descriptive analysis by experts (product experts but also sensory experts) (QDA, sensory profile…).
- preference analysis by consumers (give a global hedonic score on products).

Depending on the focus of the study, one or both assessments are used. In the case where we tend to understand which sensory properties can explain the preference, both tests are used and linked with methods such as the preference mapping method [2]. The separation of the two tests takes postulate that experts are the most suitable to give an analytic judgment and naïve consumers are the most suitable to give hedonic judgment (Stone & Sidel [1] mentioned in particular the influence of learning on hedonic judgments). However, in certain case such as the studying of a particular semantic dimension, it is tough to split the two aspects. For example in the case of the study of the comfort of a seat, sensory properties and hedonic dimension are undividable. On the one hand, subjects measure stimuli with few discrepancies (e.g. for the hardness of foam: the seat 1 is perceived as very hard by the whole panel) and on the other hand, subjects have very variable opinions (50% of subjects like sporting seat and 50% like wide seat). That is why we attempt to develop a method to get advantage of the complementary of the point of views. We attempted constructing a product set for evaluation tests integrating concrete stimuli influencing the consumers' perception. Nevertheless, we completed the consumer feedbacks with expert judgments to generate words, which are difficult to verbalize.

Our focus is to get the more relevant features with a minimum of products. It is particularly valuable in field such as the automotive industry, where selecting too many products can lead to very high costs (for example, by buying or even renting too many cars. To illustrate our proposal, we will focus on a specific semantic dimension: the fluidity of shape of car dashboards. This work deals with how we selected products to understand this semantic dimension. After this introduction, we will present some traditional methods and their limits. In a second part, we will present an innovative method: the features picking protocol. Then we will discuss the results and data analysis. Finally, the last part will treat some concluding remarks.

1.1.1 Background on the construction of a set of products

According to our reading of the literature, only few studies deal precisely with the construction of a set of product for evaluation tests. Some general recommendations are evocated but no accurate method.
For example, it is mentioned that the set of products should be chosen carefully, because it can lead to significant consequences on the results. In particular, this could occur when products are too different or too similar. In the first case, the difference brought about contrast errors (the perception of the difference between the products are exaggerated) and in the other case the difference brought about convergence errors (the difference between the products are masked or overshadowed)[1].
Traditionally, sensory studies use a set of products selected by experts [3]. In order to cover a huge range of sensations, they should exhaustively choose products with the main features estimated as interesting (a priori, about preliminary knowledge acquired by a specific training [4]. By the way, some literatures postulate that evaluating a product alone is very inefficient. In fact, the human being seems to be a better judge to evaluate products in a comparative way (e.g to evaluate the width of a product compared to another) than in absolute [5] (e.g to evaluate the width of a product without other references). This assumes that the experts are able to choose products that are interesting to bring relevant information from customers. Thus, some studies propose to avoid implicit choices by selecting products thanks to quantified data. For example, they propose to use data, coming from experts (e.g. via a sensory profile [6] , of an initial set of products. Then a HAC (Hierarchical Ascendant Classification) is used to segment the product space into groups and to select the representative prototype of each group. A second kind of studies uses kind of sensory data coming from consumers. For example,[7] use a "KJ method" (Kawakita Jiro Method) to classify products according to their main differences, then a HAC to select the more representative product of each group.
Considering the number of products, literature recommends to adapt the set of products relatively to the tiredness of assessors, to the task and to the product. (12-15 for sauces, 12 for candy…[5]). These recommendations are nevertheless more adapted for food fields and not really adapted for the automotive industry. Anyway, for statistical reasons, 7 products is the minimum number to operate preference mapping on the first two axes of a PCA (Principle Components Analysis) and 11 products for a PCA on the 3 first dimensions. That will be a first indication for our test.
Let us notice that these existing methods allow the construction of a product space after a first iteration. These methods need an initial exhaustive set of products or initial knowledge on these products. Consequently, the design of a product selection method meets two paradoxes:

- An initial set of products is necessary to be the basis from which a sub-set can be selected. Indeed, in order to propose a selection method, it is necessary to present a first selection to offer a range of possibilities of choices.

- Literature recommends not having a too heterogeneous set or a too homogenous set. But expectations on the differences between products can only be done after the selection of the set.

It leads us to believe that in our case, an initial set would be necessary too. Our main problematic is that in an area such as the automotive industry, products are constrained (expensive, complex, not easy to be evaluated…) and it is not realistic to rigorously evaluate an exhaustive set of products. Even if it is more reliable to check exhaustively the impact of all the design factors on the perception of consumer, a descriptive analysis on a huge set of products is not possible for time, tiredness and cost constraints (especially, with complex products having a high number of design factors). Moreover, the more complex the products are, the less easy the link between design features and responses will be. Therefore, it would be advantageous to focus on important factors and so to limit the size of the product space.

1.1.2 Proposal

Even if methods exist to estimate the effect of each design factors with a mathematical model (e.g. the conjoint analysis: building of a decompositional model between factors levels (the design features) and responses (the semantic scores)), it can only be done a posteriori. However, at first we think it would be wiser and richer to seek more broadly the important factors in key products rather than estimate them a posteriori on a restricted set. We consider that with complex products, it is more interesting, for finding relevant factors, to ask for it directly to the consumers. Thus, the main goal of this experiment is to make a reduced selection of products with the most influent factors according to the consumer perception. We propose to avoid laborious quantification by developing a screening methodology. Our data will be less accurate in a first time, but allow us to consider more products. We take the same approach than the study [7] (using data from consumers) but we try to avoid differences which do not affect the semantic dimension (i.e even a huge number of typologies of product exist, we attempted to avoid the products that do not allow us to deduce relevant features for our semantic dimension). We share the point of view (with this kind of approach) that it could be interesting to propose a product space, which avoids a possible discrepancy between experts and consumers. It is assumed that experts usually are more sensitive than consumers [8] but we concentrate on the difference felt by the consumers. We suppose that the experts are aware of the range of sensation, but are not always able to know which sensation is implied in hedonic or semantic dimension. Indeed, even if the experts develop specific competences, it does not imply that they are representative of the customers and their tastes[3].

2 Materials and Methods

2.1.1 Selection of the initial set

We attempted to understand the “fluidity” of the dashboard shape of a B segment car. The iterative framework previously described leads us to select a large initial set, in accordance with our industrial goal. Consequently, we selected a set of the main competitors of the B Segment car (in term of number of vehicle sold in 2009). Moreover, we completed the set with cars one segment around (C segment and A segment). We take postulate that these crossovers afford diversity without going to caricatured differences. Giving that we work on the visual aspect of dashboard, we use triptychs of photos to show at best the different shapes from different angles (fig. 1). While ordinary products sets are around a number of 6 to12 products, we conceive a protocol to consider a huge number of products (30 products). The triptychs were presented simultaneously to the subjects, on two desktops.

Fig. 1 An example of triptych with three views of the same dashboard (left, front, right).

We are aware that photos can be differently perceived than real products. Nevertheless, this is a way to asses a high number of products in a reasonable time (in average one hour). Using photos decreased the global cost of the study and allowed us to avoid certain bias. The choice of the pictures and the photo editing was done with great care. It assumes that we work specifically on intrinsic product attributes. Thus, we had neutralized some extrinsic factors that could create variability (such as the brand [9][10]): brand names have been masked, the body colors have been neutralized, environments have been homogenized …

2.1.2 Consumer testing

We have done the hypothesis that, in the case of the study of the fluidity of a dashboard, both kinds of judgment are used. Thus, we constituted variegated panel. We attempted to get spontaneous feedbacks on the key features of the fluidity of form for the consumers (without effect of the learning previously mentioned) and complement it with an expert point of view (few features may be difficult to put into words [11]).
The distribution between the two categories of subjects was not balanced: 6 experts and 54 “naïve” consumers. Consumers can be less sensitive than trained subjects, but in some cases and more particularly on the visual assessment, naïve and expert results are comparable (the number of people counterbalances their lack of specificity in descriptive accuracy [12]). The number of subjects meets the norm requirements [3]. The test included 60 subjects who met the following criteria: consumer with variable knowledge of the products, 25-55 years of age, males and females…
Because of limitations in terms of the available resources (i.e. time and money) due to the fact that this was a methodological study, the participants were recruited from the Staff of our company. We had no difficulty to recruit subjects with specific knowledge on the product. Nevertheless, to represent the naïve consumers, we selected subjects, which were not implied in the design of dashboards (Database administrator, Human resources assistant…).

2.1.3 The “Features Picking” protocol

We propose a protocol divided into three sequential stages. The subjects executed the instructions without knowing what would be the one after:

- Subjects should give a score of overall liking (from 0 to 10) for each 30 dashboards. We were not sure that subjects clearly understand the ambiguous term of fluidity. We ask them to score their liking in order to see if their perception of the fluidity was really independent.

- Subjects should give a score of the global fluidity of the dashboards (from 0 to 10) for each 30 dashboards. It was clearly explained that global fluidity score could be different than preference rating (i.e. their preferred dashboard can be the less fluid and vice versa).

- Subjects should define the main design features which contribute (in their opinion) to the fluidity of the shape of a dashboard. These features could influence positively, but also negatively the global impression of fluidity and should be non-hedonic. To illustrate their proposal (and avoid us misinterpretation), they finally should select at least one particular product, which is typical according to the verbalized feature (e.g. the consumer 1 suggest that the squareness of the control panel affects the fluidity of the dashboard. He mentioned that P268 is a good example of fluidity due to its rounded control panel, while the P018 is a bad example of fluidity due to its rectangular control panel). No constraint was given on the number of features or the number of products. We gave a relative freedom to formulate characteristics, even if it was beyond our scope (for example words describing the A-pillars, the color of dashboards…). It was a mean to not block the subjects and to help us to measure the holistic aspect of our problem. Then, subjects should draw an arrow up or an arrow down to explain the effect of each feature (e.g. the consumer 1 selects the product 018, because the squareness of the control panel is in opposite with the concept of fluidity…).

To help the subjects to do these tasks we proposed them different supports:

- In the light of previous pretests, the term of fluidity has not always been understood. To help the subjects to proceed the test, we showed them some examples of what can be the fluidity in the daily. A specific mood board (fig. 2) was proposed. Generally, human being is more creative when he/she is stimulated by images rather than words [13]. We tend to limit the bias of induction (pictures can induce specific definition of the fluidity) by creating this support in a collaborative work with five naïve subjects. This allows us to bring diversity of opinion on the semantic dimension. We nevertheless found that the low number of subjects could be a bias. For future work, it would be advisable to increase this number.

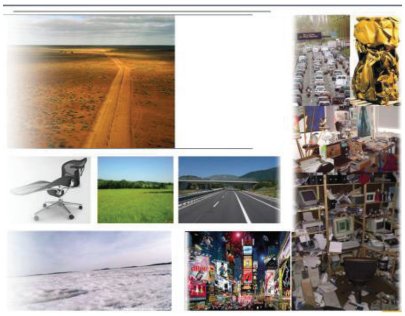


Fig. 2 the specific mood board was created by a collaborative work including few definitions of what can be the fluidity in the daily but too antonyms of fluidity.

- To avoid slowing the description for vocabulary reason. We proposed drafts on which subjects could scribble (circle areas, highlight a line...). This draft represented a CAD (Computer Aided Design) snapshot of a dashboard.

- We proposed to fill verbalization table (one per feature). In relation to their selections, the consumers verbalized each feature and each "level" they found (tab. 1). It should be noted that subjects had not to be exhaustive; they only had to give few examples.

Verbalization : *Squareness of the control panel*	Level *Rectangular*	Level *rounded*
Product …018	X	
Product …129		
Product …268		X
Product …n		

Tab. 1 Verbalization table

3 Results

The descriptive test done by the 60 subjects on the set of 30 dashboards generated 275 verbatim. In average, in one hour (the duration of the test was from 30 to 90 minutes), the subjects:

- Verbalized 4,7 verbatim in average (standard deviation of 1,8) and selected 16,3 products (standard deviation of 5,1) to explain their understanding of the semantic dimension.

- Selected 5,5 products in average (standard deviation of 2,9) to illustrate each verbatim.

- Selected 2,7 products in average (standard deviation of 2,1) to illustrate the positive values of the semantic dimension.

- Selected 2,7 products in average (standard deviation of 1,3) to illustrate the negative values of the semantic dimension.

On the whole, the 30 products were selected. It demonstrates that, for this case, every product brought a part of information. Initially, we thought that the protocol would lead to a consensus. However, this test showed that subjects had a very variable definition of the semantic dimension “fluidity”. This high number of products definitely prevents us to use this raw data to constitute a “reasonable” set. That is why we attempted to reduce this number while keeping the variety by applying a specific data analysis.

3.1.1 Analysis of the scores

As previously mentioned, we noticed in pretests that the term of fluidity has not always been understood. We were not sure if the subjects did not confuse the fluidity and the preference and consequently if our test was reliable. Thus, we wondered if subjects respond simply to the question by giving a hedonic judgment. We analyzed it with statistical analysis. Therefore, we wanted to asses if the “fluidity” score and the preference score were correlated, and if the scores are similar. Our hypothesis (H0) stated that the fluidity of a dashboard was not correlated to the preference scores. We use the Pearson’s correlation coefficient to measure the covariance between the two scores. Then X-Y Graphs are employed in order to evaluate the different mean scores.

Products	Correlation	P Value
018	**0,500**	**< 0,0001**
032	**0,340**	**□,□□8**
094	**0,313**	**0,015**
129	**0,605**	**< 0,0001**
191	**□,□□8**	**0,017**
198	**0,431**	**0,001**
217	**0,464**	**0,000**
219	0,211	0,106
230	**0,272**	**0,036**
257	**0,503**	**< 0,0001**
268	**□,□9□**	**< 0,0001**
363	**0,305**	**□,□□8**
374	**0,446**	**0,000**
410	0,157	0,230
418	**0,549**	**< 0,0001**
432	**0,607**	**< 0,0001**
459	0,149	0,255
548	**0,319**	**0,013**
569	**0,500**	**< 0,0001**

573	**0,289**	**0,025**
581	**0,510**	**< 0,0001**
625	**0,509**	**< 0,0001**
675	**0,487**	**< 0,0001**
708	**0,509**	**< 0,0001**
713	**0,524**	**< 0,0001**
755	0,248	0,056
788	**0,278**	**0,031**
790	**0,390**	**0,002**
799	**0,367**	**0,004**
808	**0,398**	**0,002**

Table 2 : Pearson's correlation coefficient

Tab.2 shows that 26 correlations among the 30 products are significant (P value <0.05 are in bold). Consequently our hypothesis was not true. Concerning the interpretation of the correlation coefficient, it was difficult to know which threshold was adapted.

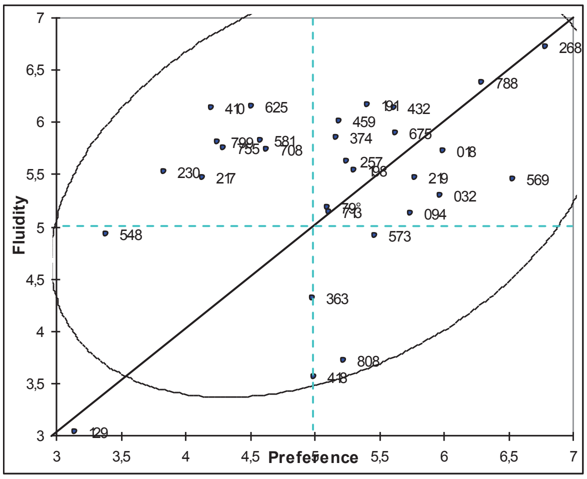


Figure 3: X-Y graph comparing preference rating and scores of fluidity. The graph is centered on the range of mean score (from 3 to 7 points over 10)

To complete our analysis, a X-Y graph has been constructed to plot data pairs. Fig.3 presents the spread of the different mean scores for the whole panel (of the 60 subjects). By taking in consideration the whole set of products, it seems that the most of products have two different mean scores. In addition, we noticed that few products were fluid and disliked (in the lower left quarter of the graph) and few products were non-fluid and liked (in the lower right quarter of the graph). In the light of these treatments, we can suppose that subjects did not replicate their preference score even the fluidity played a determining role in the product liking.

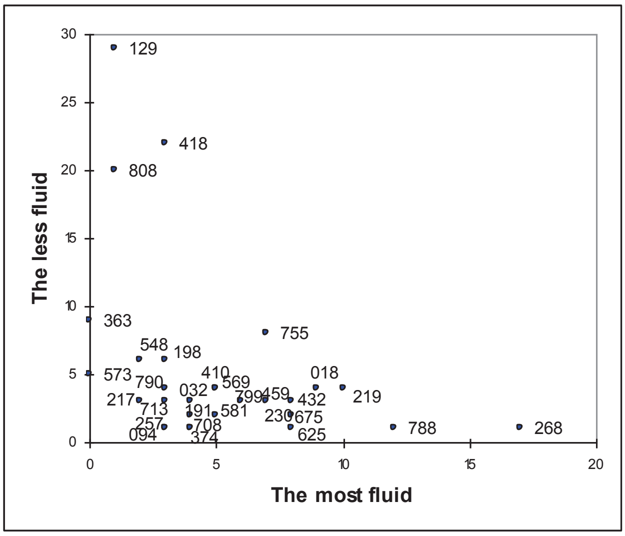
3.1.2 Analysis of the number of selections

As previously seen, many ways are possible to select a subset from an initial set of products. Given that there was no ambiguity on the fact that the products will be perceived as different, given that their number and their complexity prevented us to proceed reasonably to an exhaustive descriptive analysis (such as sensory profile), we tend to believe that our subset should have the following properties:

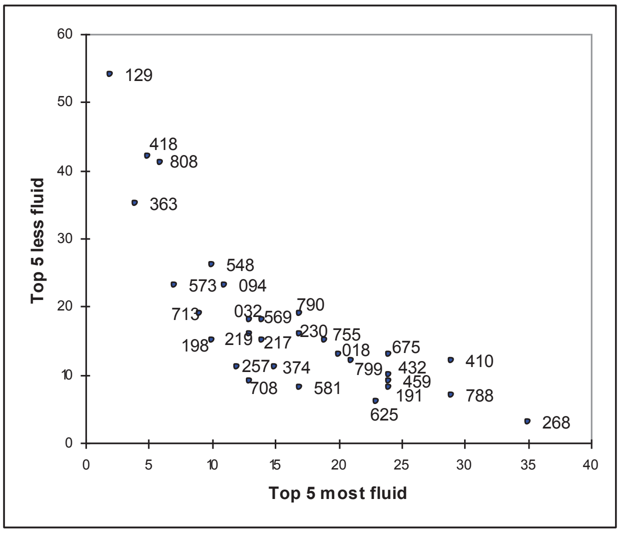
- Having products with high ratings on the particular semantic dimension. If no product possesses the specific dimension, we estimate that it's not possible to have relevant information on this dimension. Moreover, it seems reasonable that products considered as "having a fluidity of a shape" contain some ingredients or features representative of this semantic dimension.

- Having products with low ratings on the particular semantic dimension. We believe that certain unacceptable features can exist and should be avoided. Moreover, we noticed in previous fundamental research that the consensus between subjects is generally stronger on low ratings or negative values.

Given that subjects rated the products with different amplitudes (e.g. Subject 3 gives score between 0 to 10, Subject 27 gives scores between 3 to 7...), individual scores were transformed in individual ranking. In order to respect the evocated properties, we selected in the order of occurrence the products considered as "being the most fluid dashboard" and products considered as "being the worst fluid dashboard". The opinions were various and no product has been selected unanimously. Yet some products were more consensual than others. An empirical adjustment has been performed to determine the threshold of selection. Indeed, some alternatives have been considered: selecting Top 5 (with the 5 most fluid dashboards and the 5 worst fluid dashboards), selecting the "Top 4", "the Top 3"…. To determine the best alternative, we attempted to represent the consensus between rankings. Fig.4.a shows the discrepancy between the subjects with the fact that the most of products were both regarded as the most fluid and the less fluid. The number of times the products were selected as the most fluid is depicted along the x-axis and the number of times the products were selected as the less fluid is depicted along the y-axis. This effect is especially pronounced when the threshold becomes more tolerant e.g. Fig.4.b shows for example a higher confusion between the products with a threshold of "top 5" (e.g. P032 was selected 14 times among the top 5 of the most fluid and 19 times among the top 5 of the less fluid).



a. The number of times a product has be chosen as the most fluid and the less fluid (Threshold "Top 1")



b. The number of times a product has be chosen among the 5 most fluid and the 5 less fluid (Threshold "Top 5").

Figure 4 : Example of two thresholds considering the occurrences in product selections.

Product	*Occur*	*Product*	*Occur*
268	*17*	*129*	*29*
788	*12*	*418*	*22*
219	*10*	*808*	*20*
018	*9*	*363*	*9*
432	*8*		
625	*8*		
675	*8*		

Table 3: Set of products according to the strictest threshold. The products on the left was selected for their fluidity of dashboard and the product on the right was selected for their "anti-fluidity"

At the end of this step, we move towards the strictest threshold ("Top 1"). We selected the 11 products with the highest occurrences (tab. 3) to let us the possibility to use statistical models (such as Preference Mapping on the three first dimensions of the PCA). Unfortunately, it would be likely that all the illustrating products were selected for the same reason and do not illustrate the variety of what subject perceived. In consequence, we decided to analyze what the products enabled to deduce.

3.1.3 Analysis of the descriptive data

We underlined the fact that the subjects generated 275 verbatim. Nevertheless, it did not mean that these terms are not redundant. Therefore, we conducted an interpretation phase. Because of the design of the protocol, illustrating examples was not exhaustive. That is why statistical Clustering was not efficient. Consequently, we chose to perform a semantic analysis. This delicate work has been decomposed in three steps:

- First, the words were grouped in relation to the technical area they designated. Indeed, we consider that the risk of error was relatively insignificant on the misinterpretation of the name of dashboard areas.

Ex: How we construct the group "Buttons":

Verbatim n°A. "too many buttons" such as in products P418, P808 and contrary to products P94, P217, P191".

Verbatim n°B "the number of buttons is too high" such as in products P418, P808, P191 contrary to products P094, P217".

Verbatim n°C "the buttons are prominent" such as in the product P418, P219 contrary to the product P191".

Second, the words were aggregated in relation to the kind of descriptors they used. The reduction of terms was relatively light (from 275 to 224).

Ex: How we dissociate two groups:

"Number of Buttons":

Verbatim n°A. "too many buttons" such as in product P418, P808 and contrary to products P094, P217,P191".

Verbatim n°B "the number of buttons is too high" such as in product P418, P808, P191 contrary to product P094, P217".

"Prominence of the buttons":

Verbatim n°C "the buttons are prominent" such as in products P418, P219 contrary to products P191".

Third, to avoid misunderstanding, we looked on illustrating products of each terms. In addition, when they were similar, we aggregated them.

Ex : "Number of Buttons" :

Vnew. "Too many buttons" such as in product 418,808, 191 and contrary to products 94, 217,191".

At the end of this step, 112 new verbatim were obtained. Let us notice that we take a risk with this interpretation but we keep a trace of the inter-individual discrepancy by keeping eventual opposition such as the product P191 in the example. In this case, a subject considered that this product has too many buttons and a subject estimates the opposite. In order to keep this information, data was discretized in two matrices (a matrix of positive values and a matrix of negative values).

3.1.4 Assessment of the product selection

Finally, we attempted to create a link between the number of products and the descriptive data. Our goal was to have an overview of the progression of the information relative to the number of selected products. We wanted to see sequentially how many verbatim were taken into account as the number of products increased. We first estimated the possible combinations by calculating the number of subsets with "n" elements in a set of 30 elements (C_n^{30}) (fig. 5). In the light of this result, we understood that even our most powerful program with our most powerful computer was not able to calculate all the possible combinations. For this reason, we decided to calculate exhaustively the overall possibilities in the low and the high ranges (before 11 products and after 19 products) and to forecast a predictive curve.

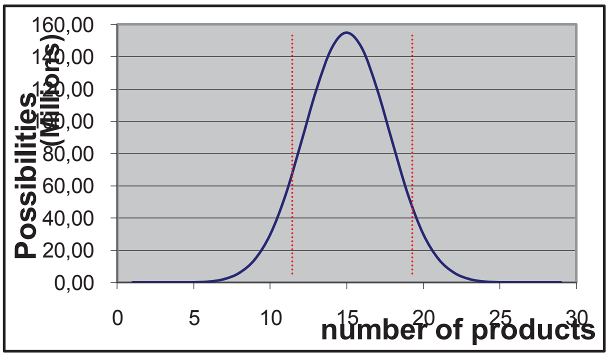


Figure 5: The possible combinations of products consist in estimating the number of possible couples, triads…

To draw a prediction curve we decided to estimate the average contribution per number of products. By analyzing the number of different verbatim generated by each product, we proceed to basic statistical treatments. We created an envelope of values by taking in consideration the better possible combination of products (max), the worst combination (min) and the standard deviation (fig. 6). As each point could not be calculated, we draw a prediction curve to fill the empty areas. We fitted a polynomial curve that gives the more satisfying adjustment (5 degrees, $R^2=1$, MCE=0,057). Note that we discretized verbatim by their polarity. In other words, we considered the 112 verbatim (the interpreted verbatim) and we distinguished the products cited for positive value and the products cited for negative value. We considered that the information increased as well with positive values (e.g "a poor number of buttons is a symbol of fluidity") than with negative values (e.g "a high number of buttons damage the fluidity"). We wanted to see for a determined number of products how many factors and how many level of factors were covered. Indeed, we considered that the two levels should be considered to illustrate the subjects' feedbacks … It is the same as saying we had 224 verbatim (112 verbatim with a positive value, 112 verbatim with a negative value).

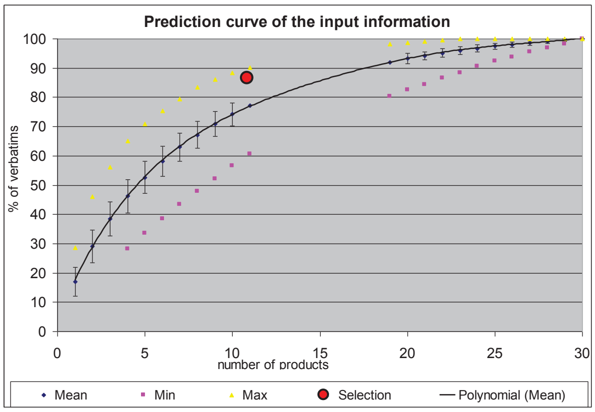


Figure 6: Prediction of the input information

This prediction curve has reinforced the fact that the method of selection by occurrences gives good results: the percentage of coverage of the global information is relatively near of the best combination of products. Moreover, the number of 11 products seems to deliver most of the time, substantive information (87% of information). In relation to the design of the protocol and the psychological errors, this prediction should nevertheless be tempered. We mean that the 100% of information probably represent only a restricted part of what the semantic dimension really is (Features probably missing and, products with specific feature was probably not declared). Then, we cannot help but hide the fact that information was born (even unconsciously) from "contrast" between different products.

4 Conclusion

The aim of the current study was to seek consumers' key features for a defined semantic dimension in order to select a set of products. Initially, we were not sure if the subjects could afford us reliable results on this dimension without confusing the fluidity and their liking. Thus, we first showed that even if fluidity appeared to play a role in the preference; the subjects understood that fluidity and preference could be different. Our method allows us to create a set containing some key products. We evaluated this set by creating a link between the number products and verbatim they can deduce. In our case we reduced a large set of 30 products to a subset of 11 products while keeping consequent information (87 % of the total of data generated by the subjects). This work brings benefits on different aspects:
- In a methodological perspective, by quantifying the benefit of a fundamental approach (selecting products with a high ratings and low ratings of the particular semantic dimension).
- In an industrial perspective, by using the consumers feedback as soon as possible in the evaluating process and avoiding products which are irrelevant. Without taking into account that we have conducted a long validation process, we propose a simple and fast way to construct a product space.

Nevertheless, this information was relative because it is possible that the 13% remaining was the most important verbatim. It is assumed that information come from a declarative task and is by nature non-exhaustive. There is obviously a need to extend this work to an exhaustive descriptive analysis that could improve precision of the prediction curve. It is not reasonable in all case because if products are too complex and too numerous (such as with car dashboards), the experiments would be too tiring. The complexity of the products is probably a key information to determine the order of the sensory evaluation. If products are too complex, they may be laborious quantify. Therefore, it would be preferable to begin with a hedonic test or "Features Picking" on large set. If products are easy to be sensory quantified, it would be preferable to begin with accurate descriptive analysis (such as sensory profile) and reduce the initial set via clustering.

Further research

The set of products we obtained could be the basis of different kind of studies allowing to build a statistical model of the fluidity. Sensory profile, napping, can be used in order to describe precisely the product sensory properties. Our future goal is to hierarchy the importance of the verbatim and the design factors behind these verbatim. Thus, we expect to proceed to an optimal design of experiments. The present work will be an input to recommend technical factors to control in order to reconstruct a new product space. The benefit would be to evaluate factors by avoiding covariance problems. This experiment will use a parameterized CAD model and will allow us to avoid noises coming from extrinsic factors. Indeed, even if we homogenized the photos, there is no doubt on the fact that some factors disturb our test (some subjects have certainly recognized the models).

Acknowledgement

The authors wish to acknowledge Guillaume Willemotte for his support in competitive analysis and Stephane Bouillot for his support in programming.

References

[1] Stone, H. and Sidel, *J.L. Sensory Evaluation practices* Academic Press, 1974.
[2] Schlich P., Preference mapping; relating consumer preferences to sensory or instrumental measurements. In Etiévant P. & Schereier P., Bioflavour : Analysis/precursor studies/biotechnology. Versailles. INRA Editions 1995, pp231-45.
[3] AFNOR XP V 09-51, Sensory Analysis. General guidance for sensory evaluation. Description differentiation and hedonic measurement, 1999.
[4] ISO 8586-2, Sensory analysis. General guidance for the selection, training and monitoring of assessors. Part 2 : experts 1994.
[5] SSHA (Société Scientifique d'Hygiène Alimentaire) *Evaluation sensorielle, manuel méthodologique* (2 ed.). Collection STAA. Techniques et documentations, Paris. 1998.
[6] Taréa S., Danzart M., Cuvelier G., Sieffermann J.M. An original procedure to identify the key instrumental parameters regarding the sensory texture of fruit purees. In: 1st International Workshop on Materials and Sensations, October 27-29, 2004, Pau.
[7] Lin Y.C., Lai H-H., Yeh C-H. *Consumer oriented product form design based on fuzzy logic : A case of mobile phones* International Journal of Industrial Ergonomics – Vol. 37(2007) pp531-543.
[8] Ishii R., Hawaguchi, O'Mahony M., Rousseau B., *Relating consumer and trained panels discriminative sensitivities using vanilla flavoured ice cream as medium.* Food Quality and Preference Vol 18 (2007) pp89-96.
[9] Pronko N. H., & Bowles J. W. *Identification of cola beverages I: a first study Journal of Applied Psychology*, 32 (1948) pp559-664.
[10] Pronko N. H., & Bowles J. W. *Identification of cola beverages II: a further study Journal of Applied Psychology*, 33 (1948) pp605-608.
[11] Aliouat E., Blumenthal D., Petiot J-F., Danzart M., Sieffermann J-M., *Contribution to the selection of products for evaluation tests: How to select products for the study of a particular semantic dimension ?* IDMME - Virtual Concept, October 20 – 22, 2010, Bordeaux, France.
[12] Faye P., Brémaud D., Teillethttp://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6T6T-4KHC34H-1&_user=10&_coverDate=12%2F31%2F2006&_rdoc=1&_fmt=high&_orig=search&_origin=search&_sort=d&_docanchor=&view=c&_searchStrId=1624405453&_rerunOrigin=google&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=893fd6d3dfce6d3fba0c9bc457e49efa&searchtype=a - aff3#aff3 E., Courcouxhttp://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6T6T-4KHC34H-1&_user=10&_coverDate=12%2F31%2F2006&_rdoc=1&_fmt=high&_orig=search&_origin=search&_sort=d&_docanchor=&view=c&_searchStrId=1624405453&_rerunOrigin=google&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=893fd6d3dfce6d3fba0c9bc457e49efa&searchtype=a - aff3#aff3 P., Giboreau A., Nicod H., *An alternative to external preference mapping based on consumer perceptive mapping* Food Quality and Preference Volume 17, Issues 7-8, October-December (2006), pp 604-614.
[13] Malaga R. A. *The effect of stimulus modes and associative distance in individual creativity support systems*. Decision Support Systems 29 (2) (2000) pp125‐141.