Question on ipm: A case study on lemon juice

Minh-Tam Lê, H.D. Nguyen, Gaëlle Roudaut, Dominique Valentin

SPISE2012

Proceedings

INtegrating SENSORY EVALUATION to PRODUCT DEVELOPMENT

An Asian Perspective

Dominique Valentin
Christelle Pêcher
Dzung H. Nguyen
Delores Chambers
Hervé Abdi

Vietnam National University-
HoChiMinh City Publishing House

Proceedings of SPISE 2012

Summer Program in Sensory Evaluation 2012
3d International Symposium

INTEGRATING SENSORY EVALUATION INTO PRODUCT DEVELOPMENT
AN ASIAN PERSPECTIVE

Edited by

Dominique Valentin, Christelle Pêcher, Dzung Hoang Nguyen, Delores Chambers & Hervé Abdi

Vietnam, July 24-26, 2012

Preface

SPISE 2012, "**Taste and Think. Integrating sensory evaluation into product development: An Asian perspective,"** the third symposium on sensory evaluation in ASEAN, was held on July 24–26, 2012 at Ho Chi Minh-City University of Technology, Vietnam. We had the great pleasure to welcome more than one hundred scientists from Vietnam, Korea, Singapore, China, Thailand, Japan, Indonesia, the USA, France, New Zealand, the Netherlands, and Denmark. The focus of this meeting was on sensory evaluation and product development.

There is widespread agreement in the food and beverage sectors that competitiveness in food markets relies in developing new, differentiated products that are adapted to the different and specific preferences of consumers or consumer segments. Yet, developing products that are successful is a difficult and time-consuming process. As competition is high and demand ever changing the rate of success might be very low. Knowing both consumers demand and product sensory properties is now essential to success because in order to survive products need to meet consumers' expectations. Therefore, it is necessary, when developing new products, to investigate and to consider factors influencing quality and their effect on sensory properties. The use of sensory evaluation in product development programs by food manufacturers is becoming a major tool to reach this goal.

Sensory evaluation is a science that measures, analyses and interprets the responses of people to products as perceived by the senses (see *e.g*., Stone & Sidel, 1992; Lawless & Heymann, 2010). Human sensory data provide a better model of how consumers will react to food products than instrumental data as these data take into account both the product properties and the interpretation of these properties by consumers. The methods used in sensory evaluation can be divided into three categories: discriminative, descriptive, and hedonic tests.

Discriminative tests are used to evaluate whether any difference exists between two products. The most well known discriminative test is the triangle test. In this test, panelists receive three samples, two identical and one different. Panelists are asked to pick the odd sample among the three. Another well known test is the 2 or 3 alternative forced choice (2 or 3 AFC) which can be used when the main dimension of difference between the samples is known. Panelists receive 2 or 3 samples and are asked to indicate which one is the highest in the specified dimension (e.g., sweetness). Discriminative tests are generally conducted with about 30 untrained panelists who have been screened for sensory acuity and familiarized with the test procedure.

Descriptive tests aim at understanding product characteristics such as taste, texture, smell, and appearance in a controlled environment. The most well known descriptive test is the Quantitative Descriptive Analysis (QDA™, Stone *et al.*, 1974). QDA is performed by a small number of panelists (from 8 to 15) who provide intensity ratings for a set of selected attributes. It involves three main steps. The *first* step is product familiarization and development of a lexicon, which comprehensively and accurately describes the product space. The *second* step consists in training the panelists in order to align and standardize the sensory concepts of the panel. The *third* step is scoring of the products on the basis of each descriptive attribute on an intensity scale. The performance of the panel is monitored in terms of discrimination power, agreement between panelists, and reproducibility during training to achieve the most accurate, reliable, and consistent results as possible. But as training panelists is quite costly, some new methods will rely instead on untrained panelists (see, *e.g.*, Valentin *et al*. 2012, for a recent review).

Hedonic tests aims at understanding consumers. They are performed with large groups (> 100) of untrained panelists screened for product use. In product development, hedonic tests are usually used towards the end of the formulation process to evaluate which formulation is preferred. The easiest method to address this question is to ask panelists to rank the products in order of preference. This approach is known under the name of preference ranking and is quite useful to compare different formulations but does not provide information on the liking magnitude. An alternative is to ask panelists to indicate their liking of the products on a hedonic scale (hedonic scaling). Different types of hedonic scale can be used: semantic, numeric, unstructured line scale. Among those the most popular one is the so-called 9-point hedonic scale. This is a balanced semantic scale going from like extremely to dislike extremely with a central neutral category (Jones, Peryam, & Thurstone, 1955).

The chapters in this proceeding present some new developments in sensory testing or some applications in product development. They are organized into three topics which we used to organized this meeting sessions and themes:

1. New methods and research tools in consumer research
2. Food choices and consumer behaviour studies
3. Product development and food market
4. Cultural and social determinants of food choices

We would like to use this opportunity to express our gratitude to our two keynote speakers, Pr Harry T. Lawless (Cornell University, USA) and Mr Hajime Nagai (Product Development Center, Japan*)*, for their great contributions. Our special thanks are due to our partners who participated to the organization of this meeting: the HCMC University of Technology, AgroSup Dijon, CSGA, AgroCampus-Ouest and Groupe ISA Lille. We would also like to thank, for their generous help, our sponsors: Fizz-Biosystems, LogicStream, Masan Consumer, Agrosup Dijon and CSGA. We extend our special thanks to those who have helped us so much and worked so hard to make this event possible: Phan Thụy Xuân Uyên , Nguyễn Bá Thanh , Nguyễn Thị Hằng , Trần Thị Cúc Phương , Nguyễn Thị Thu Hà , Nguyễn Quốc Cường, Nguyễn Quốc Dũng, và Lê Minh Tâm.

References

Jones, L.V., Peryam, D.R., and Thurstone, L.L. 1955. Development of a scale for measuring soldiers' food preferences. *Food Research, 20*, 512-520.

Lawless, HT, & Heymann, H. (2010). *Sensory Evaluation of Food: Principles and Practices*. 2nd Edition, New York : Springer.

Stone, H., & Sidel, J.L. (1992). *Sensory evaluation practices, 2nd Edition*. San Diego: Academic Press.

Stone, H., Sidel, J., Oliver, S., Woosley, A., & Singleton, R.C.C. (1974). Sensory evaluation by Quntitqtive Descriptive Analysis. *Food Technology, 28*(11), 24-34.

Valentin, D., Chollet, S., Lelièvre, M., & Abdi, H. (2012). Quick and dirty but still pretty good: A review of new descriptive methods in food science. *International Journal of Food Science & Technology*, 47, 1563-1578.

Table of Contents

PART 1

New Methods and Research Tools in Consumer Research

QUESTION ON IPM: A CASE STUDY ON LEMON JUICE

M.T. Lê*[a,b], H.D. Nguyen[c], G. Roudaut[d] and D. Valentin[a]

(a) UMR CSGA 6265 CNRS, Inra, UB, France
(b) Agrocampus Ouest, France
(c) Ho Chi Minh City University of Technology, Vietnam
(d) UMR PAM AgroSup Dijon, UB, France

*Email: minh-tam.le@agrocampus-ouest.fr, Tel: +33.2.23485476

Abstract

In food product development, using consumers to obtain ideal products is becoming more frequent in recent years. Despite being an economical way of optimizing products, relying on consumers is, however, not yet widely accepted in the sensory community. The main objectives of this study were to evaluate (1) whether the stability of ideal intensities and ideal products could be improved by providing a frame of references when consumers perform the Ideal Profile Method (IPM) and (2) the effect of the order of the questions on consumers' responses.

In the present study, sixty participants conducted two tasks: (1) rating the perceived and ideal intensities of attributes and (2) reporting the overall liking. They were randomly divided into four groups. Group G1 and G2 used references in rating intensity, whereas group G3 and G4 did not. Concerning the order of question, group G1 and G3 answered the overall liking question before rating the intensity, whereas G2 and G4 followed the inverse order. The experimental design included six lemon juice samples, and four attributes *(pulp, lemon odor, sweetness, and sourness)*. A three-way ANOVA was conducted on the ideal intensities of attributes and overall liking with reference and order as a between-subject factor and product as a within-subject factor.

Results showed that providing a frame of references improved the stability of the ideal attribute *pulp*, but not of the other ideal attributes. Besides, the order effect was without influence on overall liking. For ideal intensities, an order effect was observed on the ideal attribute *pulp* between groups G3 and G4 (without reference), and was not observed between groups G1 and G2 (with references). Several possible modifications to improve the IPM are discussed.

Keywords: *Ideal Profile Method; order effect; food product development*

1. Introduction

Product development plays an important role for all food companies. There are two approaches to develop a food product: the process-oriented approach and the consumer-oriented approach.

The process-oriented approach originally refers to a process in which various ingredients are systematically varied to create a number of different products. These products are then rated by a sample of category users, with each respondent rating overall liking as well as various sensory attributes of the products. The resulting data are then analyzed by analysis of variance (ANOVA), regression and/or response surface analysis in order to obtain the optimal product formulation. The advantage of the process-oriented approach is to control the technical parameters and therefore to set up an optimal formulation. However, this approach has also some limitations. First, consumers' demand is not given special importance even though, after all, the consumers are the buyers (Moskowitz *et al.*, 2006). Second, this approach takes time and resources (Lewis *et al.*, 2010). To overcome these limitations, the consumer-oriented approach focuses on the expectations of potential buyers, and then uses this information to modify a product to satisfy consumers' demand. In the last decades, three consumer-oriented methods (preference mapping, just about right scale, and ideal profile) have been described in the literature.

Preference mapping includes internal preference mapping (Chang & Carroll, 1969) and external preference mapping (Schlich, 1995; Meullenet *et al.*, 2008; Carbonell *et al.*, 2008). Internal preference mapping analyzes overall liking data to give the directions of preferences and to identify consumer segments. External preference mapping relates the overall liking data to sensory attributes generated by trained panelists. Despite their current popularity among sensory scientists, preference mapping techniques have some limitations. The main limitation is that training a panel to describe products is time-consuming, as it takes normally about 4 to 6 months to perform. To overcome this limitation, there has been an increasing interest in the past years in obtaining both sensory and overall liking data directly from consumers. New methods such as Just About Right (JAR) scales and Ideal Profile Method (IPM) have been developed.

JAR scales are commonly used to determine the optimal level of products' attributes. The scales combine intensity and overall liking judgments (Rothman & Parker, 2009); they typically consist of five or seven points, ranging from *too little* to *too much* for a given attribute (Meullenet, Xiong, & Findlay, 2007). The center point can be labeled "just-right" or "just-about-right." Consumers are asked to indicate whether the intensity of each attribute of the food product is *too high*, *too low* or *just about right* in addition to their overall liking. Data analysis is performed by computing the percentage of consumers who evaluated the attributes as *too high*, *too low* or *just about right*. Limitations of JAR scales occur when they are used without asking any additional intensity-related questions. For example, two groups of consumers might both mark a product as "just about right." However, one group might think the product is very strong (the level they prefer), whereas the other group thinks the product is fairly mild (the level they prefer). Thus, the results might mislead product developers into thinking that the participants are from a homogeneous population, while they are really from two different consumer segments (Lawless & Heymann, 1998). Moreover, the task is rather complicated for the consumers as they have to evaluate the intensity of attributes of an actual product and subtract it from the intensity of the same attributes in their ideal product (Punter & Worch, 2009). According to Punter and Worch (2009), it might be simpler to explicitly ask consumers to evaluate both the perceived intensities of the product and the ideal intensities for this type of product as it is done in the Ideal Profile Method (IPM).

IPM has been proposed recently as an alternative to both preference mapping and JAR scales (van Trijp *et al.*, 2007). In addition to evaluate their overall liking of the product, consumers are asked to rate, for a number of attributes, both the perceived intensity of the attribute and the intensity of that attribute if it was ideal. The deviation from the ideal product is then computed and related to overall liking. The IPM presents several advantages. First, absolute intensity information is obtained as well as the position of the ideal product. Second, individual scores can be expressed as deviations from the ideal score. So the directional information that helps to adjust the concentrations of attributes can be obtained. However, like other methods, IPM has its own limitations.

Despite being an economical way of guiding product development, relying on consumers is not yet widely accepted in the sensory community. For instance, Lawless and Heymann (2010) question the ability of untrained consumers to act in an analytical way when they taste products and to precisely understand some specific attributes. Popper *et al.* (2004) pointed out that overall liking of consumers could be altered when analytical questions on specific attributes were involved. Finally, an "ideal product profile" can be constructed only if the samples of consumers are reasonably homogeneous (van Trijp *et al.*, 2007; Punter & Worch, 2009; Lawless & Heymann, 2010).

The main objective of this study was to evaluate whether the stability of ideal intensities and ideal product can be improved by providing a frame of references when consumers perform IPM. Besides, we were also interested in evaluating whether the order of overall liking and intensity questions had an influence on consumers' responses.

2. Research procedure

A three-step procedure *(see Fig.1)* was used in this study. The first step consisted of constructing a product space. The second step consisted of generating the attributes and selecting the samples for the main experiment using Flash Profile (Delarue & Sieffermann, 2004). Finally, the third step included preparing a frame of references and measuring consumers' ratings in IPM. In the third step, the *reference* factor effect was evaluated by comparing the rating data obtained in two experimental conditions: with references and without references. To evaluate the *order* factor effect, half of the participants began by the liking question and the other half by the intensity question.

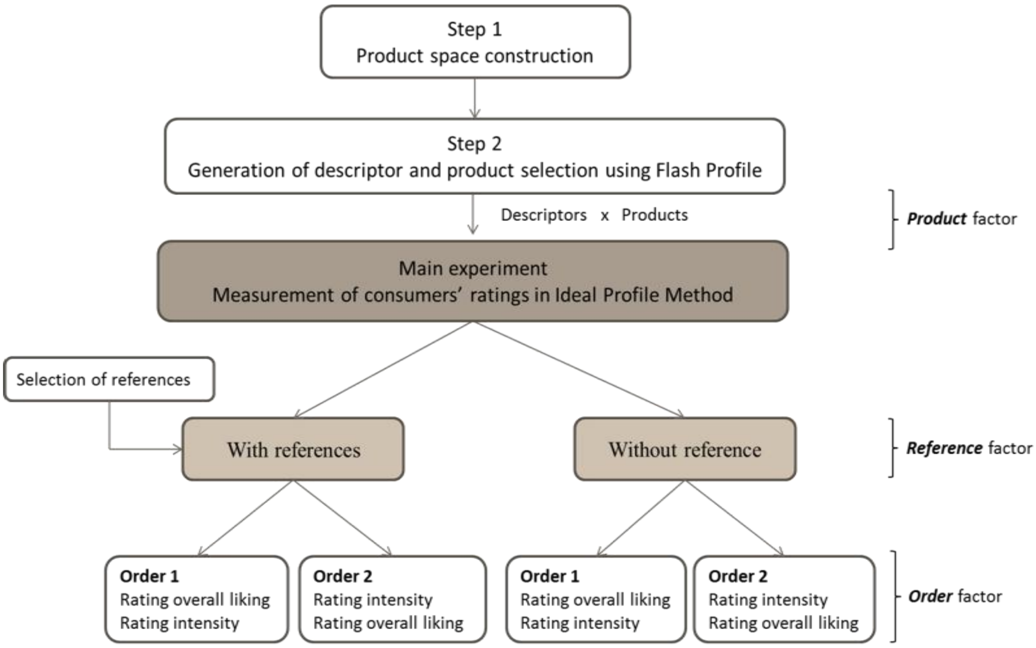

Figure 1: Three-step procedure used to evaluate the effect of *reference* and *order* on overall liking and intensity judgment in IPM.

The goal of the study was to test the two following hypotheses. *Hypothesis 1* – If providing a frame of references improves ideal product stability, we expect a significant *reference* effect and a *reference* by *product* interaction; *Hypothesis 2* – If the order of question (presentation of intensity and overall liking) has an effect on ideal product stability, we expect a significant *order* effect, a *reference* by *order* interaction, and an *order* by *product* interaction.

2.1. Step 1 - Product space construction

2.1.1. Samples

The base product was made by diluting one part of "Pulco Citron Vert" with six parts of water. Two series of samples were then made by adding either sucrose (sweet samples) or citric acid (sour samples) into the base product. The amounts of sucrose added in the base product were 5, 10, 15, 20, and 25 grams per one liter of base product, to make the sweet samples. The added amounts of citric acid were 0.7, 1.4, 2.1, 2.8, and 3.5 grams per one liter base product, to create the sour samples.

2.1.2. Participants

Twenty four staff members (12 men, 12 women, 24 to 40 years old) were recruited from the "*Centre des Science du Goût et de l'Alimentation"* (CSGA).

2.1.3. Procedure

Two series of triangle tests were carried out, one with the *sweet* samples and the other with the *sour* samples. Half of the participants received the *sweet* samples and the other half received the *sour* samples. Each participant was asked to do five triangle tests. For each triangle test, three samples were prepared from two products: the base product plus one of the products from *the sweet (or sour)* series. All samples were presented in plastic cups coded by 3-digit numbers and were served at ambient temperature (22^0C to 24^0C). Participants tasted the samples from left to right and were asked to chose the odd sample. They were requested to rinse their mouth with water after each test, but not between samples in each test. All sessions were conducted at the sensory laboratory in CSGA.

2.1.4. Data analysis

The number of correct answers for each triangle test was counted. It was then compared to the critical value of the binomial distribution at α risk 0.1%.

Table 1: The 8 samples were varied from 3 factors *product, sourness* and *sweetness.*

Sample	*Product*		*Sourness*		*Sweetness*	
	Pulco Vert	Pulco	C_0	C_1	C_0	C_1
F1	X		X		X	
F2	X		X			X
F3	X			X	X	
F4	X			X		X
F5		X	X		X	
F6		X	X			X
F7		X		X	X	
F8		X		X		X

Pulco citron vert: a lemon juice product of Orangina Schweppes France with the ingredient: water, lemon juice 40%, pulp of lemon 4%, and acidifying. Pulco citron: a lemon juice product of the same company with the ingredient: water, lemon juice (35.5%) and orange juice (4.5%), pulp of lemon 4%, acidifying.

C_0: the concentration of citric acid (and/or sucrose) in base product, C_1: the concentration of citric acid (and/or sucrose) chosen in step 1.

2.1.5. Results

We selected the minimum concentration at which the number of correct answers was more than the number corresponding to the critical value for and alpha level of .05 (equal to 10 in this case). Over a total of 12 answers, the selected concentrations used to vary *sweetness* and *sourness* were 15.0g/l and 3.5g/l respectively. These concentrations were noted as C_1. To formulate samples, a factorial design was constructed with three factors *product*, *sourness* and *sweetness*. Each factor had two levels. For the *product* factor, two lemon juice products were used. For the *sweetness* and *sourness* factors, two concentrations C_0 and C_1 were prepared. Table 1 presents the eight formulated samples.

2.2. Step 2 - Generation of attribute and product selection using Flash Profile

2.2.1. Samples

The eight samples prepared from step 1 were used.

Table 1: The eight samples were varied from three factors *product, sourness* and *sweetness.*

Sample	*Product*		*Sourness*		*Sweetness*	
	Pulco Vert	Pulco	C_0	C_1	C_0	C_1
F1	X		X		X	
F2	X		X			X
F3	X			X	X	
F4	X			X		X
F5		X	X		X	
F6		X	X			X
F7		X		X	X	
F8		X		X		X

2.2.2. Participants

A panel of seven judges (3 men, 4 women, 24÷40 years old) was recruited from the staff at CSGA. The judges were not trained to evaluate lemon juice products before, yet they had participated in descriptive tests.

2.2.3. Procedure

The Flash Profile method was used to generate attributes of lemon juice products. This method was carried out in a single session of one hour including three steps. In the first step, the whole set of products was presented simultaneously to the judges. Each judge observed, smelled, and tasted all samples in order to generate attributes. In the second step, each judge read the lists of other judges to update his/her own list. In the third step, judges ranked all eight samples from “least” to “most” according to their list of attributes (Figure 2).

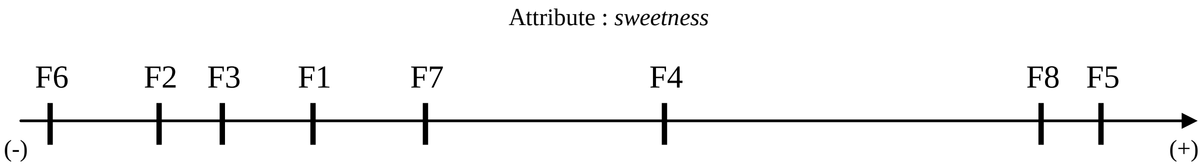

Figure 2: An illustration of ranking the *sweetness* of eight formulations using Flash Profile.

2.2.4. Data analysis

First, the number of citations for each attribute was counted. Second, the ranked data for each attribute were analyzed using Friedman test, with a *p*-value set to $\alpha = 0.10$. Third, Multiple Factor Analysis (MFA) (Pagès, 2005) was used to analyze the data matrix obtained from the samples scored on the individual attributes by the judges. The results of this analysis gave a map of the samples and a map of the attributes generated by each judge to describe the samples. The analysis was conducted by using the R software (R Development Core Team 2011) with the package SensoMineR 1.14 (Husson *et al.*, 2011).

2.2.5. Results

Each judge used between six and seventeen attributes to describe lemon juice products. Fourteen attributes were used by more than two judges (Table 2). The Friedman test showed that six attributes were discriminant at $\alpha = 0.10$. These attributes were *pulp, lemon odor, sourness, sweetness, astringency*, and *pungency*.

The first two dimensions of the MFA accounted for 37.36% and 25.16% of the total variance of the data set (Figure 3). Dimension1 was primarily a function of *pulp* and *lemon odor*, and Dimension 2 opposed *sourness, astringency and pungency* to *sweetness*. All samples were organized in six groups (F1,F4), (F5,F8), (F2), (F6), (F3), and (F7). The samples prepared from product Pulco (F5, F7, and F8) were characterized by *pulp* and *lemon odor*. The F3 and F7 were characterized by *sourness, astringency,* and *pungency*; whereas F2 and F6 were characterized by *sweetness*.

Table 2: Usage frequency and p-value of Friedman test of 14 attributes generated in Flash Profile.

Manner	Attribute		Frequency of citation	p-value
	French	English		
Appearance	*Pulpe*	*Pulp*	5	*0.04*
	Opaque	*Opaque*	5	0.4395
	Jaune	*Yellow*	4	0.7243
Odor	*Citron pressé/ naturel*	*Pressed/ natural lemon juice*	4	*0.07*
	Citron	*Lemon*	2	_
Taste.sensation	*Acide*	*Sour*	7	*8.19e-05*
	Sucré	*Sweet*	6	*8.62e-05*
	Astringent	*Astringent*	5	*0.06*
	Piquant	*Pungent*	4	*0.007*
	Citron	*Lemon*	3	_
	Amer	*Bitter*	2	_
	Salé	*Salty*	2	_
	Gratte la gorge	*Rasp the throat*	2	_
After taste	*Persistant*	*Persistent*	2	_

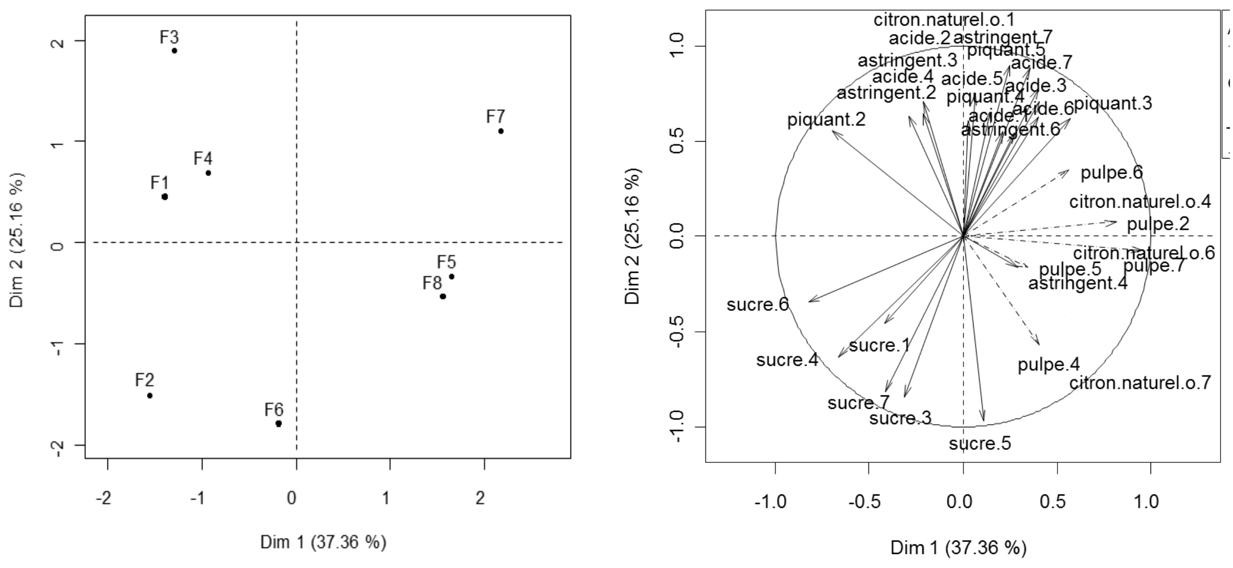

Figure 3: Representation of samples (left) and attributes (right) by the first two dimensions of MFA. Appearance is represented by dotted lines and Taste sensation by solid lines.

2.2.6. Product selection

The attributes used in the main experiment need to be understood by most of consumers. The first criterion to select the attributes for the main experiment was that the attributes were used by more than 50% of judges (3/7 judges). The second one was that the attributes should have the ability to discriminate the samples (Friedman test at p = .10). Six attributes met both criteria. Among these attributes, four were selected for the main experiment: *pulp*, *lemon odor*, *sweetness* and *sourness*. Moreover, to prevent consumers from adaptation and fatigue, the number of samples was set at six. Samples F1 and F5 were not used in the main experiment because their sensory characteristics were close to that of samples F4 and F8.

2.2.7. Selection of references

References for attributes pulp and lemon odor. As mentioned in the result of step 2, the samples prepared from product Pulco were characterized by the attributes *pulp* and *lemon odor*. Concerning the two references of these attributes, Ref.1 (minimum reference) was prepared from product Pulco Vert and Ref.2 (maximum reference) was prepared from product Pulco citron (Table 3). The same procedure as in Step 1 was used. Two series of dilution were prepared, and each series had five ratios. To make the *weak* series, Ref.1, the dilution ratios of water to product were 6.5, 7.0, 7.5, 8.0 and 8.5 parts of water to 1 part of product. To make the *strong* series, Ref.2, the ratios were 5.5, 5.0, 4.5, 4.0 and 3.5 parts of water to 1 part of product. Six participants received one series and performed five triangle tests.

The correct responses were counted and compared to the critical value of binomial distribution at the level of $p = .01$. With six correct responses, the dilution ratio of Ref.2 was 1:4, and the dilution ratio of Ref.1 was 1:8. The dilutions of Ref.1 and Ref.2 of attribute *odor* were determined in the same way.

References for attributes sweetness and sourness. Concerning the attribute *sweetness*, the sample F1 was used as Ref.1 for the attribute *sweetness*. Ref.2 was prepared from product Pulco from which a *sweet* series with five dilutions was prepared by adding concentrations of sucrose 20, 25, 30, 35, 40 g/l gradually. Three samples for the triangle tests were also prepared from two products: one product at concentration C_1 and one of 5 products from the *sweet* series. Based on the triangle test results, Ref.2 concentration was prepared by adding 30g/l sucrose to base product of Pulco.

Concerning the attribute *sourness,* the sample F5 was used as Ref.1. The pH value of product Pulco Vert was smaller than the pH value of product Pulco. Product Pulco Vert was then varied at 5 concentrations to make a *sour* series to determine concentration of Ref.2 by adding gradually 4.2, 4.9, 5.6, 6.3 and 7.0 g/l of citric acid into the base product. Three samples of triangle tests were also prepared from two products: one product at concentration C_1 and one of five products from the sour series. Based on triangle test results, Ref.2 was prepared by adding 7.0g/l citric acid into the base product of Pulco Vert.

Table 3: Concentration of the two references for attributes *pulp, lemon odor, sweetness*, and *sourness*.

Manner	Attributes	Minimum reference Ref.1	Maximum reference Ref2
Appearance	*Pulp*		
Odor	*Lemon*	Pulco vert (1:8)	Pulco (1:4)
	Sweetness	Pulco vert (1:6)	Pulco (1:6) + 30g/l sucrose
Taste	*Sourness*	Pulco (1:6)	Pulco vert (1:6) + 7.0g/l citric

3. Main experiment: Measurement of consumers' ratings in IPM

3.1. Materials and methods

3.1.1. Samples

The six samples F2, F3, F4, F6, F7, and F8 were used and noted as P1, P2, P3, P4, P5, and P6 in this experiment.

3.1.2. Participants

Sixty participants (27 men, 33 women, 19 to 60 years old) were recruited. To participate in this study, the participants had to consume lemon juice products at least once a month. They were informed that the study was on lemon juice.

3.2. Procedure

Participants were randomly divided into four groups of 15 participants each (Table 4). Group G1 and G2 received references for intensity ratings, whereas group G3 and G4 did not. Concerning the order of question, group G1 and G3 answered the overall liking question before rating the intensity of attributes, whereas G2 and G4 rated the intensity before answering the overall liking question.

Table 4: Experimental design.

Experimental condition	Overall liking, then intensity rating	Intensity rating, then overall liking
With references	G1	G2
Without reference	G3	G4

To rate the intensities of attributes, participants in groups G1 and G2 were asked to observe, smell, and/or taste the references (Figure 4*)*. They were instructed that the concentration of Ref.1 was anchored at 1/3 of the scale, and the concentration of Ref.2 was anchored at 2/3 of the scale; the endpoint labeled (-) indicates the weak intensity, and the endpoint labeled (+) indicates the strong intensity. After observing/smelling/tasting the references, participants rated the intensities of attributes. For each attribute, they were asked to rate the perceived intensity and then the ideal intensity. After rating the intensity of the four attributes and answering the overall liking question, they went to the next sample.

Figure 4: Illustration of the experimental set up.

Participants of G3 and G4 followed the same procedure as participants of G1 and G2 did; however, they were not provided with the references. Participants were instructed that the scale was marked at 1/3 and 2/3, the endpoint labeled (-) indicated the weak intensity, and the endpoint labeled (+) indicated the strong intensity.

A 150 mm unstructured line scale (marked at 50% and 100%) was used in rating intensities and a 9-point scale was used in rating overall liking (Figure 5)

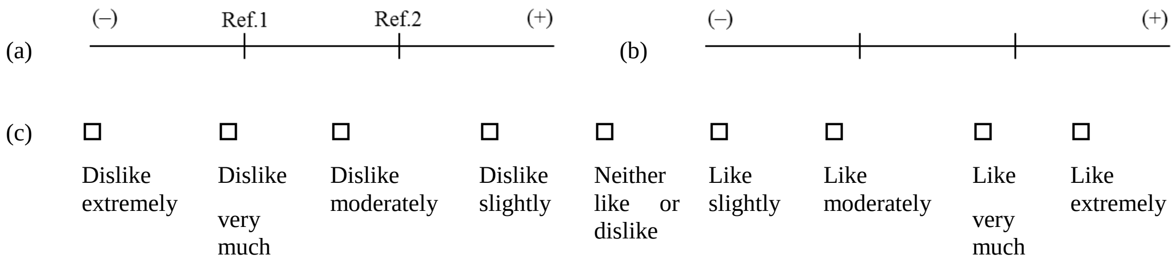

Figure 5: Illustration of the 150mm unstructured line scale used in the two experimental conditions with references (a), without reference (b); and of the overall liking 9 point scale (c).

3.3. Data analysis

3.3.1. Assumptions of IPM

First, the homogeneity of consumers' preference was verified by performing a PCA on the matrix of samples by overall liking score. Second, the differences of overall liking between the samples, as well as the differences of perceived intensities by attributes between the samples were tested using a three-way ANOVA with *reference* and *order* as between-subject variable, and *product* as within-subject variable. Third, if a significant *product* effect was observed, a mean comparison between the samples were performed using TukeyHSD *post hoc* test at p =.05.

3.3.2. Verifying the hypothesis

A three-way ANOVA was conducted on the ideal intensity of the four attributes. In this model, *reference* and *order* were defined as between-subject variables, and *product* as within-subject variable. Besides, the coefficient of variation (CV)[1] of each attribute in each experimental condition was calculated. The analysis was conducted with R software v.2.12.2 (R Development Core Team 2011), with package SensoMineR 1.14 (Husson *et al.*, 2011).

4. Results

4.1. Assumptions of IPM

4.1.1. Are participants taken from a homogeneous population?

Figure 6 presents the first two dimensions of the overall liking PCA. These two dimensions accounted, respectively, for 43.71% and 20.23% of the variance of the data set. Vectors toward the right space of the map represented the responses of participants who liked most samples P1, P3, P4, and P6; whereas vectors S4.10 (participant 10 in G4) and S1.11 (participant 11 in G1) in the lower-left quadrant represented two participants who preferred sample P2. In general, most participants had the same direction of preference with lemon juice samples.

4.1.2. Are all products different in overall liking?

The result of ANOVA showed a significant effect of *product* on overall liking *(F(df=5) = 22.08, MSe = 44.33, p < .001)*. A mean comparison showed that samples P1, P3, P4, and P6 were liked more than samples P2 and P5 (Table 5).

[1] The coefficient of variation (CV) is defined as the ratio of the standard deviation to the mean. This parameter represents the ratio of noise/signal. In the context of this study, the CV value characterizes the variation within each group. The smaller the CV value was, the more consensual the participants were.

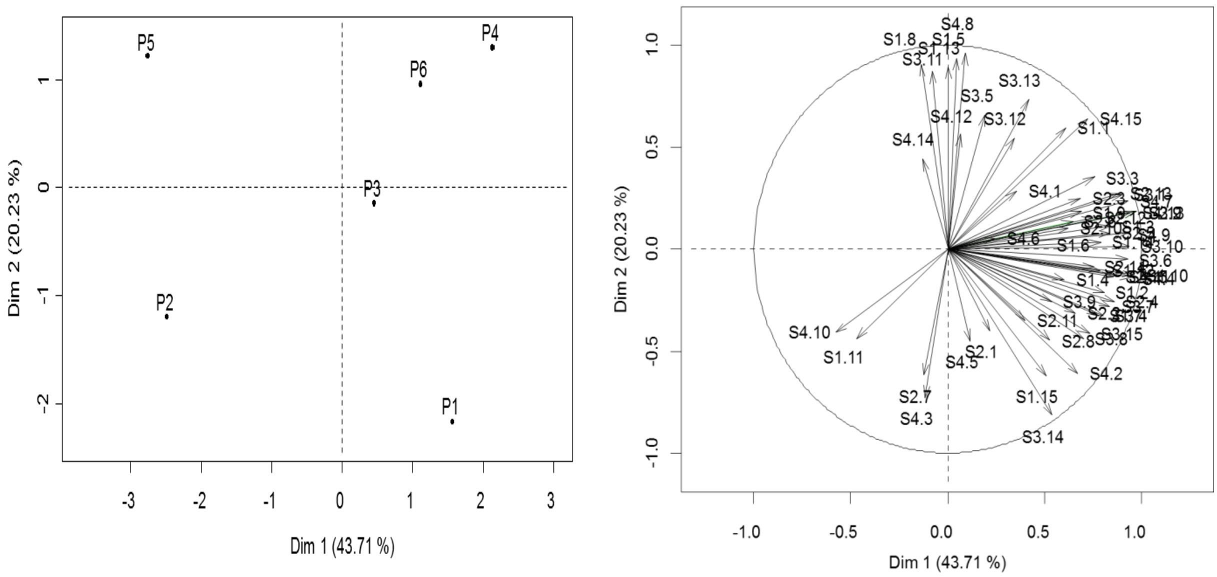

Figure 6: Direction of the preference for the six lemon juice samples.

Table 5: Average scores of overall liking and perceived intensities of the six samples.

Rating question	Samples					
	P1	P2	P3	P4	P5	P6
Overall liking	5.58^{b}	3.92^{a}	5.15^{b}	5.93^{b}	4.03^{a}	5.60^{b}
Pulp	48.15^{ab}	46.93^{ab}	44.58^{b}	70.13^{c}	61.13^{ac}	56.36^{bc}
Lemon odor	69.60^{a}	77.93^{a}	67.65^{a}	68.25^{a}	68.45^{a}	72.3^{a}
Sweetness	59.70^{b}	32.31^{a}	55.35^{b}	64.55^{b}	35.92^{a}	53.78^{b}
Sourness	67.21^{bc}	99.88^{a}	83.70^{ab}	64.85^{c}	97.21^{a}	77.97^{bc}

Correspond to each attribute, share a common letter do not differ significantly from one another (p<.05).

4.1.3. Are all products different in perceived intensities?

A significant effect of product was observed on attributes *pulp (F(df=5) = 11.79, MSe = 5853, p < .001))*, *sweetness (F(df=5) = 23.93, MSe = 10312, p < .001)*, and *sourness (F(df=5) = 18.49, MSe = 12996, p < .001)*. A mean comparison showed that the *pulp* in sample P4 was larger than that in samples P1, P2 and P3; the *sweetness* of samples P2 and P5 was weaker than the *sweetness* of other samples, whereas the *sourness* of samples P2 and P5 was stronger than that in other samples. No *product* effect was observed on the attribute *lemon odor (F(df=5) = 1.06, MSe =917.5, p = .38)*.

4.2. Stability of ideal product evaluation.

4.2.1. Does the frame of references influence on ideal intensity?

Table 6 presents the *p-value* of a three-way ANOVA performed with *reference*, *order*, and *product* on ideal intensity of attributes. There was no significant effect of *reference* on attributes *sweetness.id*[2] *(F(df=1) = 0.005, MSe = 11, p = .94)*. However, a significant effect of *reference* was observed on attributes *pulp.id (F(df=1) = 5.20, MSe = 14314, p = .026)*, *lemon odor.id (F(df=1) = 8.99, MSe = 16430, p = .004))*, and *sourness.id (F(df=1) = 3.58, MSe = 9996, p =.06)*. This result indicated that for the attributes *pulp.id*, *odor.id* and *sourness.id*, the ideal intensities obtained in the experimental condition with references were different from those obtained in the experimental condition without reference.

[2] *id*: ideal attribute. For example: *sweetness.id* is the ideal attribute *sweetness.*

Table 6: The *p-value* of three-ways ANOVA test on ideal intensity of four attributes.

Source of variance	Ideal Intensity			
	Pulp.id	*Lemon odor.id*	*Sweetness.id*	*Sourness.id*
Reference	*0.0263**	*0.0040***	0.9459	*0.0636.*
Order	0.2390	0.9589	0.6304	0.3725
Reference:Order	*0.0380**	0.9062	0.5152	0.6192
Product	0.3111	0.6262	*0.0085***	0.7185
Reference:Product	0.1609	0.8234	0.8214	0.8450
Order:Product	0.9081	0.8139	0.4650	0.2095
Reference:Order:Product	0.9613	*0.0475**	0.5163	0.5535

Significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Table 7: Coefficient of variation between the two experimental conditions.

Attribute	Experimental condition	
	With references	Without reference
Pulp.id	0.32	0.38
Lemon odor.id	0.29	0.24
Sweetness.id	0.30	0.28
Sourness.id	0.39	0.36

4.2.2. Does providing a frame of references make ideal intensity more consensual?

The CV of each attribute in each experimental condition was calculated and presented in Table 7. The smaller CV was, the larger the consensus between participants was. A difference in consensus was observed only for the attribute *pulp.id*. A higher consensus was observed in the frame of reference condition.

4.2.3. Is the ideal product unique?

A significant effect of *product* was observed for the attribute *sweetness.id (F(df=5) = 3.16, MSe = 669.3, p =.008)*. However, a mean comparison showed that there was no significant difference on attributes *sweetness.id* between the samples. Besides, the effect of *product* was not observed on ideal attributes *pulp.id*, *lemon odor.id*, and *sourness.id*. In general, the four attributes of the ideal profile of lemon juice product were not differently perceived between the products.

4.3. Order effect

4.3.1. Is there an order effect on overall liking?

The *order* effect was without influence on overall liking *(F(df=1) = 0.13, MSe = 0.62, p =.71)*. Besides, the result showed that there was no significant interaction of *reference* by *order (F(df=1) = 0.49, MSe = 2.33, p = .48),* and *order* by *product (F(df=5) = 0.97, MSe = 1.95, p = .43)* on overall liking.

4.3.2. Is there an order effect on ideal intensity?

A significant *reference* by *order* interaction was observed for the attribute *pulp.id (F(df=1) = 4.51, MSe = 12414, p = .038)* (Table 6). The difference between the four groups G1, G2, G3, and G4 was determined by TukeyHSD *post hoc* comparison and illustrated in Figure 7. Result showed a significant difference between the group G3 and G4 (without reference), but not between groups G1 and G2 (with references).

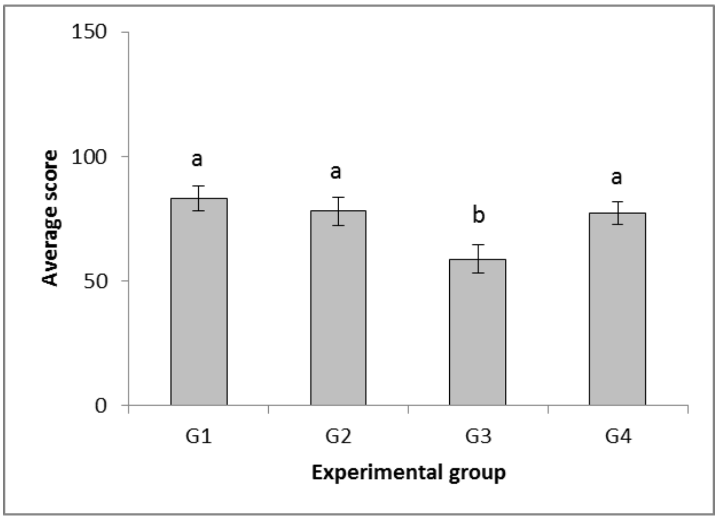

Figure 7: Illustration of *reference* by *order* interaction on attribute *pulp.id*. G1: with references, overall liking/intensity, G2: with references, intensity/overall liking, G3: without reference, overall liking/intensity, G4: without reference, intensity/overall liking.

4.4. Effect of product on reference by order interaction

A three-way interaction *reference* by *order* by *product* was also observed for the attribute *lemonodor.id [F(df=5) = 2.77, MSe = 717.4, p = .0475]* (Table 6). The three-way interaction implied that *reference* by *order* interaction was affected by *product*. However, a mean comparison failed to show any significant difference between groups by *product*. For other attributes, no significant three-way interaction was observed.

5. Conclusion

The issue of experts versus consumers is a strong debate in the sensory community. Some authors are sensitive of the reliable capacities of consumers' rating (Worch *et al., 2010;* Moskowitz, Muñoz, & Gacula, *2003*), whereas others express doubt in it (Lawless & Heymann, 2010). This study addressed a small part of this complex question in the framework of IPM.

The results showed that providing a frame of references could improve the stability of only one attribute *pulp.id*. Popper *et al.* (2004) and Earthy *et al.* (1997) pointed out the effect of question *order* on overall liking. However, in the present study, no effect of question *order* was observed. The data showed that, for ideal intensities, this effect was observed on the attribute *pulp.id* between groups G3 and G4 (without reference), but not between groups G1 and G2 (with references).

Although we attempted to offer a frame of references for all studied attributes, this frame of references proves useful for attribute pulp.id, but not for the other attributes. This suggests that, references might be more efficient for visual attributes than for chemical stimulus. This might be due to the fact that contrary to olfactory or gustatory references, visual references do not need to be memorized. For sensory modalities requiring some memory storage some training might be necessary for references to be efficient. However, additional work is needed to verify this hypothesis.

Concerning the methodological aspect, the Ideal Profile is a quick and easy method to meet industrial needs in food product development. However, it also has some limitations. The implication of this study is to propose two modifications to improve the Ideal Profile.

With references, participants can have a better understanding of attributes. However, using references is time consuming (i.e., to select and prepare the references), and required participants to taste more samples than in classical IPM. To overcome these disadvantages, experimenters might use some pre-tests to better understand their product space. This step could help in choosing attributes needing references.

Ideal attributes should be evaluated only once after the participants rate all samples for the two following reasons. First, if each ideal product is obtained for each formulation, and a final ideal product is obtained by computing the average of all ideal products, it may pose a risk. From a statistical point of view, the concentrations of ideal attributes of the final ideal product may be the average of all ideal products. Hence, we can get an imprecise lead in optimizing intensities/product. Second, from an empirical point of view, generating only one ideal product can decrease the demand on participants and, hence, save time for tasting additional samples.

Acknowledgements

Our special thanks go to Catherine Dacremont (Professor of AgroSup Dijon, France) for her help in this research. We would like to thank our colleagues Phan T.X.U (Kansas State University, USA) and Nguyen Q.D. (Agrocampus Ouest, France) for their comments on the manuscript.

References

Carbonell, L., Izquierdo, L., Carbonell, I., & Costell, E. (2008). Segmentation of food consumers according to their correlations with sensory attributes projected on preference spaces, *Food Quality and Preference*, 19, 71-78.

Chang, J.J., & Carroll, J.D. (1969). How to use MDPREF, A computer program for multidimensional analysis of preference data. *Computer manual*, Bell Labs, Murray Hill, NJ.

Delarue, J. & Sieffermann, J.M. (2004). Sensory mapping using flash profile. Comparison with a conventional descriptive method for the evaluation of the flavor of fruit dairy products, *Food Quality and Preference*, 15, 383-392.

Earthy, P.J., Macfie, J.H., & Hedderley, D. (1997). Effect of question order on sensory perception and preference in central location trials, *Sensory Studies*, 12, 215-237.

Husson, F., Lê, S., & Cadoret, M. (2011). SensoMineR: Sensory data analysis with R. R package version 1.14. http://cran.r-project.org/web/packages/SensoMineR/index.html (accessed May 17, 2011).

Lawless, H.T. & Heymann, H. (2010). *Sensory Evaluation of Food: Principles and Practices (Second edition)*, Springer, 620p.

Lawless, H.T. & Heymann, H. (1998). *Sensory Evaluation of Food: Principles and Practices (First edition)*, Chapman & Hall, 819p.

Lewis J.N., Teratanavat, R., Beckley, J., Jeltema, M.A. (2010). Using a consumer-driven rapid product navigation process to develop an optimal product, *Food Quality and Preference*, 21, 1052-1058.

Meullenet, J.F., Lovely, C., Threlfall, R., Morris, J.R., Striegler, R.K. (2008). An ideal point density plot method for determining an optimal sensory profile for Muscadine grape juice, *Food Quality and Preference*, 19, 210-219.

Meullenet, J.F., Xiong, R., & Findlay, C. (2007). *Multivariate and probabilistic analyses of sensory science problems*, Blackwell Publishing, 256p.

Moskowitz, H.R., Beckley, J.H., Resurreccion, A.V.A. (2006). *Sensory and consumer research in food product design and development*, Blackwell Publishing, 358p.

Moskowitz, H. R., Munoz, A. M., & Gacula, M. C. (2003). *Descriptive panels/experts versus consumers. In Viewpoints and controversies in sensory science*. Trumbull, CT: Food & Nutrition Press, Inc.

Pagès, J. (2005). Collection and analysis of perceived product inter-distances using multiple factor analysis: Application to the study of 10 white wines from the Loire Valley, *Food Quality and Preference*, 16, 642-649.

Popper, R., Rosenstock, W., Schraidt, M., & Kroll, B.J. (2004). The effect of attribute questions on overall liking ratings, *Food Quality and Preference*, 15, 853-858.

Punter, P.H., & Worch, T. (2009). The ideal profile method: combining classical profiling with JAR methodology, *Proceedings of SPISE2009 "Food Consumer Insights in ASIA: Current Issues & Future*, 5-18.

R Development Core Team (2011). R: A langue and environment for statistical computing, R foundation for statistical computing, Vienna, Austria. ISBN 3-900051-07. http://www.r-project.org/foundation/main.html (accessed May 17, 2011)

Rothman, L. & Parker, M.J. (Eds.) (2009). Just-About-Right (JAR) Scales: Design, Usage, Benefits and Risks, ASTM MNL63, 110 pg.

Schlich, P. (1995). Preference mapping: Relating consumer preferences to sensory or instrumental measurments. In P.Étiévant and P.Schreier (Eds.)*Bioflavour'95: Analysis/Prescursor Studies/Biotechnology*, pp.135-150, INRA, Dijon, France.

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.

Worch, T., Dooley, L., Meullenet, J.F., Punter, P.H. (2010). Comparison of PLS dummy vairables and Fishbone method to determine optimal product characteristics from ideal profiles, *Food Quality and Preference*, 21, 309-318.

ORGANISERS

SPONSORS

SPISE2012 SECRETARIAT
306B2 Ho Chi Minh City University of Technology
268 Ly Thuong Kiet street, District 10, Ho Chi Minh City, Vietnam
Tel: 84-8-38647256 ext. 5696
Website: *http://conferences.hcmut.edu.vn/spise2012*