A Method for Sensory Data Collection and Analysis in Product Development

Flávio Sanson Fogliatto
UFRGS
Suzan L. Albin
Ana Maria dos Santos Silva
RUTGERS USA
E-mail: sffoglitto@ppgep.ufrgs.br

Abstract: The quality of several industrial products can only be fully measured through sensory evaluation of some of its properties. That is particularly true in the food and pharmaceutical industry, where product development technicians are often faced with the task of collecting and processing data from sensory evaluation panels. In this paper we propose a new method for sensory data collection and analysis. Our method presents two important features that may appeal to the product development practitioner: (*i*) reliable sensory panel data may be elicited from untrained panelists, and (*ii*) a consistency index which objectively measures how well panelists perform sensory evaluations is determined for each panel member. The method we propose is applied in a case study from the food industry.

Keywords: Indirect pairwise comparison method, sensory evaluation, product development.

1. Introduction

Sensory evaluation methods offer an organized way to collect information on sensory aspects of samples as perceived by the human senses. These methods are applied in product development and reformulation, on-line and off-line quality control, and for marketing purposes (monitoring competition, for example).

Samples evaluated in sensory panels correspond to different outcomes of an experiment, i.e., different formulations of a product, manufacturing setups, and so forth. Ideally, sensory evaluation results should allow the analyst to relate the factors varied when preparing samples to their corresponding sensory impact. In other words, results are transformed into mathematical models used to predict sensory outcomes according to the combination of factors chosen when preparing samples. Such models may only be determined when sensory evaluation of samples yields quantitative data, one or more measurement per sample.

In this paper, we propose Indirect Pairwise Comparison (IPC) method for sensory data collection and analysis. Our method is based on a family of psychophysical scaling methods introduced in the 50's (STEVENS, 1957), and usually denoted by magnitude estimation (ME). Originally, ME was used to determine mathematical relationships between physical intensities of an attribute and corresponding subjective intensities, as perceived by panelists. Although mostly applied in psychophysical studies, ME is also reportedly efficient when used in the sensory evaluation of products (see, LEIGHT & WARREN, 1988).

We use the central idea of ME in the IPC method: measuring the intensity of an attribute as perceived from different samples using ratios (i.e., describing the intensity of pairs of samples using a ratio of intensities). We also incorporate to our method the use of graphic rating scales to measure responses, as suggested in Quantitative Descriptive Analysis techniques (MEILGAARD *et al.*, 1991). Data analysis in our method uses analytic tools from SAATY's (1977) Analytic Hierarchy Process, a methodology used in decision making for selecting the best among a set of alternatives, given some criteria.

The key idea to the IPC method is to quantitatively evaluate the intensity of sensory attributes in samples by comparing them to a control element. We present the panelist to the entire group of N samples, one of which is identified as the control element. The panelist is asked to evaluate samples regarding the intensity of a given attribute, recording evaluation results on a printed scale. Intensities as perceived in the samples are marked on the scale according to their relation to the control element: the center of the scale corresponds to a sample with intensity identical to the control element and the extremes correspond to samples with intensities much weaker or much stronger than the control element; intermediate scale points denote compromise situations. We then change the control element and ask the panelist to perform the evaluations once again. Each sample in the group will be the control element at its turn. After the data collection is complete, N printed scales will be at hand, one per control element.

Scale marks are then converted into numerical values reflecting the results of comparing each sample against the control element. We create an (N x N) square matrix with rows labeled 1 to N, each corresponding to a control element, and entries a_{ij} giving the result of comparing sample j against control element i. We call this matrix the panelist's *judgement matrix*. The numerical results from each of the N scales are then written onto the judgement matrix, in their appropriate rows. There will be one judgement matrix per panelist.

Through algebraic manipulation we extract the following information from a judgement matrix: (*i*) a weight vector giving the intensity ranking of the samples, and (*ii*) a performance measure for the subject, the consistency ratio. The consistency ratio describes to what extent transitivity is respected when several samples are evaluated simultaneously by a subject. For example, samples a and b are compared in intensity with control element c; if $a = 2c$, and $b = 2c$, then transitivity is respected if $a = b$. These calculations are detailed in section 3.

Similar to Quantitative Descriptive Analysis and magnitude estimation techniques, the IPC method yields quantitative data on the samples, which can be used for model building purposes. Our method, however, has the advantage of being able to measure the efficiency of panelists through their consistency ratios. Using this information, the panel leader is able to assess the effectiveness of training practices on panelists, or to combine evaluations from different subjects using their consistency scores as weights.

The IPC method also presents a major advantage when compared to standard magnitude estimation procedures: samples are assigned to positions on a scale instead of evaluated using numbers. Panelists are known to perform better when asked to match intensities with positions on a scale, rather than numbers (STONE *et al.*, 1974). Panelists denote intensities by measuring distances in the scale, intuitively comparing ratios of distances even without being instructed to do it.

In the IPC method, we are likely to observe less inconsistency in evaluations, since all samples are available at once for comparison. In addition, our test procedure like those in magnitude estimation does not require intensive training of panelists. IPC's major drawback is the fatigue imposed on panelists by simultaneous presentation of samples, leading to more evaluation sessions and higher data collection costs. In fact, any sensory data collection method deals with this same problem. The gain in consistency, however, seems to outdo this limitation.

The rest of this paper contains the sensory data collection procedure we suggest (section 2) and how to analyze the collected data (section 3). A case example from the food industry where the method is applied is presented in section 4. A conclusion closes the paper in section 5.

2. Data Collection Procedure

The IPC method is applied as follows: intensity of an attribute is to be evaluated by a given number of panelists. Panelists must be able to identify the attribute under study and be trained to assign numbers to stimuli corresponding to their intensity (for training procedures, see AMERINE *et al.*, 1965; STONE *et al.*, 1974; and MOSKOWITZ, 1977).

Organize samples in a judgement matrix, like the one presented in Table 1. At each row of the matrix all samples are compared with respect to the sample corresponding to the row label n, $n = 1,...,N$, and each row constitutes a separate test. A total of N tests is to be performed, and more than one evaluation session is likely to be needed. At each test, all N samples are exposed at once, and the one corresponding to the row label is identified as the control element. Subjects are instructed as follows: "You will be presented with a group of samples, one of them identified as the control element. Your task is to tell how intense they appear to you in comparison with the control. The intensity of the control sample corresponds to the center of the scale in front of you. Samples that are more intense than the control must have their codes marked on the right-hand side of the scale accordingly, and those less intense than the control on the left-hand side. When two samples appear to be equally intense, write their codes at the same spot on the scale."

Samples must be coded appropriately (for coding procedures, see AMERINE *et al.*, 1965). The subject is given a printed scale, like the top scale in Fig. 1. We suggest the use of a 15 cm-long line, with three (or five) marks equally distanced from each other. These measures should be taken as suggestions. Empirical evidence gathered by STONE *et al.* (1974) points to a higher sensitivity in results when a 15 cm scale is used in evaluating attributes of a *single* sample. As the number of samples increases (> 6), a larger scale may be more convenient.

The use of printed marks as "anchors" to the evaluation is common practice in sensory tests, and tend to reduce the bias introduced by the use of numbers; STONE & SIEDEL (1993). However, if subjects are comfortable with numbers, they may be assigned to the center and end marks of the scale. We use descriptions instead of numbers. Subjects are instructed to place the code corresponding to each sample with a mark on the scale, according to the perceived intensity of the sample attribute (see bottom scale in Fig. 1). Their evaluations are later converted into numerical values, ranging from $^1/_9$ to 9, which are written in the rows of the judgement matrix. Two estimates of each entry a_{ij} of the judgement matrix will be available (a_{ij} and $1/a_{ji}$).

	1	2	Λ	*N*
1	1	a_{12}	Λ	a_{1N}
2	$1/a_{12}$	1	Λ	a_{2N}
Ν	Ν	Ν	O	Ν
N	$1/a_{1N}$	$1/a_{2N}$	Λ	1

Table 1. Judgement matrix.

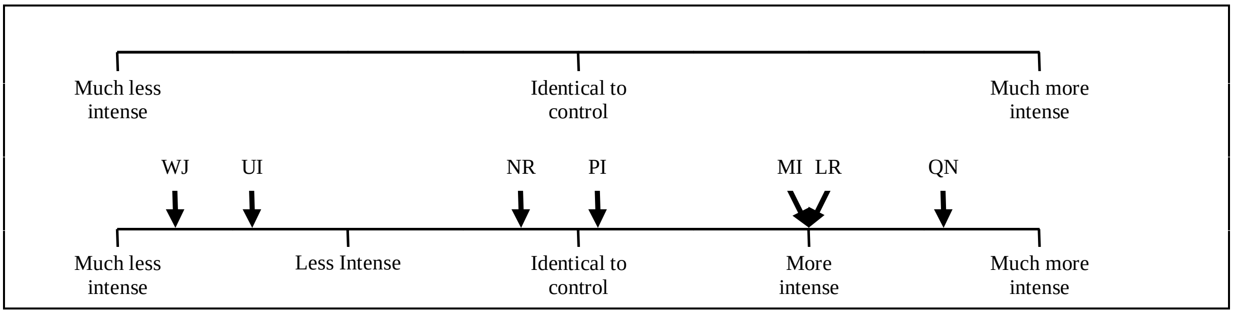

Figure 1. Examples of 15 cm scales (arrows refer to sample codes).

Each subject performs *N* tests, one printed scale resulting from each test. Consider subject *k*. After performing all *N* tests, we must convert the resulting printed scales into numerical values in a [$^1/_9$, 9]-scale, and write them onto the rows of subject *k*'s judgement matrix. This is how to convert marks on a printed scale into numerical values. For a point *z* units to the left of the center of the printed scale, the element in row *i* and column *j*, $j = 1\ldots N$, $i \neq j$, expressed in the [$^1/_9$, 9] scale is

$$a_{ij} = (8z/s)+1, \tag{1}$$

where s is half of the length of the printed scale. For a point z units to the right of the center of the printed scale, the [$^1/_9$, 9] scale value is

$$a_{ij} = [(8z/s)+1]^{-1} \tag{2}$$

Note that a judgement matrix will contain two comparisons for samples *i* and *j*, namely, a_{ij} and $1/a_{ji}$; these comparisons are likely to be non-identical. To overcome that problem, calculate midpoint values of a_{ij} and $1/a_{ji}$ and make the unique resulting value equals a_{ij}; the final corrected judgment matrix will be obtained by forcing reciprocity along the main diagonal.

By changing the control element at each test, there may be situations where all samples are more (or less) intense than the control. ENGEN & LEVY (1955) reported a very small influence of the position of the control element on test results when samples of various intensities are compared. A very moderate tendency to overestimate high intensity samples was detected when the control was the least intense sample; the reverse situation also was noticed. The overall effect, however, was minor.

The number of samples in a test should ideally be less than 10, but this number varies with the type of attribute being evaluated. A guide table is given in Tab. 2. In that table, absolute evaluations refer to individual (one at a time) evaluation of samples, while relative evaluations refer to tests where samples are evaluated against a control.

The data collection procedure above may be adapted to reduce the number of tests required to obtain judgement matrices. For an $N \times N$ judgement matrix, for example, N tests are required, which may be economically infeasible in many cases. The procedure below may reduce considerably the number of tests.

From a given judgement matrix row we may obtain the remaining matrix entries using two identities:

$$a_{ij} = 1/a_{ji} \qquad (3)$$
$$a_{ij} = a_{il} \times a_{lj}, \text{ for any samples } i, j \text{ and } l.$$

The resulting matrix is perfectly consistent. From the N tests needed to fill out a judgement matrix, choose M tests, where $M << N$. For example, from a total of (N =) 10 tests, choose (M =) 3 to be actually performed by panelists. We want to evaluate samples regarding their intensity of a given attribute. To properly choose M out of N possible tests, review samples to roughly identify their attribute intensity. Then choose samples with attribute intensities evenly distributed along the intensity scale to be the control samples in each test. For example, when M = 3 tests are to be performed out of N = 10, choose the first control sample to have a low intensity, the second control sample to have an intermediate intensity, and the third control sample to have a strong intensity of the attribute under study.

Sense Stimulated	Absolute Evaluations Number of Samples	Relative Evaluations Number of Samples
Taste	4-6	6-8
Sight	7-9	6-10
Smell	-	8-13
Hearing	4-8	6-10
Sources: ERIKSEN & HAKE (1955), GARNER (1953), JACOBS & MOSKOWITZ (1988), LEIGHT & WARREN (1988), MOSKOWITZ (1970, 1971, 1977), POLLACK (1953), POLLACK & FICKS (1954), REYNOLDS & STEVENS (1960).		

Table 2. Number of stimuli usually presented to subjects performing absolute and relative evaluations.

Suppose panelist k performs M (<< N) sensory tests. Write the results from each test in M separate matrices; each matrix will have a single row with numbers. To fill out the remaining matrices' rows, use the identities in eqn. (1). If k is perfectly consistent, the M resulting matrices will be identical. Otherwise, a matrix of intermediate judgements may be determined calculating the midpoint of the M outcomes obtained from each matrix. For example, suppose the comparison between samples 1 and 3 (a_{13}) yields three distinct numbers: 1.2, 2 and $^1/_{1.8}$; the resulting midpoint is $^1/_{1.1}$. The matrix of midpoints is then used for determining a weight vector giving the intensity ranking of the samples according to panelist k, as well as the panelist's consistency measure.

3. Data Analysis

We now introduce the basic notation and analytical tools to be used in the IPC method. We describe two analytical tools: calculation of (*i*) weight vectors and (*ii*) consistency ratios. These tools were originally conceived by SAATY (1977), and are explained next.

A weight vector is a vector of intensity weights, each corresponding to a sample in a judgement matrix; from the weights, a scoring of samples may be determined. Consistency ratios describe to what extent transitivity is respected when several samples are evaluated pairwise. Both weight vectors and consistency ratios are calculated from judgement matrices. As noted previously, after sensory data collection at least two estimates of each evaluation are available. When all N tests are performed, exactly two estimates are obtained, a_{ij} and $1/a_{ij}$; when M tests are performed, M estimates are obtained. Weight vectors and consistency ratios can only be determined from reciprocal matrices, i.e., matrices where $a_{ij} = 1/a_{ji}$ for all i and j. A judgement matrix can be made reciprocal by calculating the midpoint of entries $a_{ij}^{(1)}, a_{ij}^{(2)}, \ldots$, for all i and j, and forcing reciprocity.

We denote a reciprocal matrix of midpoints obtained from panelist k's evaluations by $\overline{\mathbf{A}}_{kp}$, where p denotes the response or sensory attribute under evaluation; the calculation of weight vectors and consistency ratios outlined next are performed on these matrices.

3.1 Weight vectors

We want to calculate a weight vector for each matrix of midpoints $\overline{\mathbf{A}}_{kp}$. The weight vector of $\overline{\mathbf{A}}_{kp}$ is denoted by $\mathbf{w}_{kp} = [w_{1p}, w_{2p}, ..., w_{Np}]$. SAATY (1977) shows that the weight vector corresponds to the matrix principal eigenvector. Whenever we evaluate a sample j against a control i we are indeed estimating the ratio of their intensity weights, i.e., $a_{ij} = w_i / w_j$. The eigenvector gives the values of w_i and w_j, given the estimates a_{ij}. Let λ_{max} denote the largest eigenvalue of a matrix $\overline{\mathbf{A}}_{kp}$. Its principal eigenvector $\mathbf{w}_k$ is given by (STRANG, 1988):

$$\overline{\mathbf{A}}_{kp}.\mathbf{w_{kp}} = \lambda_{\max}.\mathbf{w_{kp}}. \tag{4}$$

3.2 Consistency ratios

For a judgement matrix to be consistent $a_{ij} = w_i / w_j$, and $a_{ij} = 1 / a_{ji}$ must hold for all i, j. Also, for any elements i, j, and p in a matrix, $a_{ij} . a_{jp} = a_{ip}$. A measure of consistency in a matrix may be derived from λ_{max} in eqn. (2). Under perfect consistency, $\lambda_{max} = N$ for a matrix of order N. Otherwise, SAATY (1977) shows that $\lambda_{max} > N$ and suggests a consistency index given by:

$$\mathrm{CI} = (\lambda_{\max} - N) / (N - 1). \tag{5}$$

We compare CI with a random consistency index, RI, obtained from 500 randomly generated judgement matrices of order N where judgement values a_{ij} are randomly picked from a $[^1/_9, 9]$ scale. That gives us an idea of how far we are from the worst case situation. Values of RI in Table 3 (SAATY, 1980) represent the 95^{th} percentile of the randomly generated CIs. The consistency in a judgement matrix is measured by the consistency ratio CR, given by:

$$\mathrm{CR} = \mathrm{CI} / \mathrm{RI}. \tag{6}$$

The threshold value for CR suggested by Saaty is 0.1. A matrix with CR > 0.1 should have its evaluations reviewed (Saaty's choice of threshold value for CR is thoroughly justified in VARGAS, 1982). For sensory evaluation purposes however, such a low CR value may be unrealistic and larger values may be adopted (0.2 or 0.3, for example).

N	3	4	5	6	7	8
RI	0.58	0.90	1.12	1.24	1.32	1.41

Table 3. Random Index values for matrices of order N = 3,...,8.

4. Case Example

The case study deals with the development of a new formula for a well-known brand of a pet food product (dog biscuits). Samples are obtained from a mixture experiment with ten experimental runs; percentage of three ingredients and biscuit thickness are the control variables. A central composite design ran on independentized control variables (CORNELL, 1990; Ch.3) is the design chosen. Table 4 presents both the coded independent control variables (W's and *Thick.*) and the % of each ingredient tested in the runs (I's).

Two sensory attributes are measured through sensory panel: texture and general appearance. Evaluations are performed by five untrained company employees. Texture essentially measures the baked dough consistency and crackiness. General appearance encompasses aspects such as biscuit color and integrity. Panelists are familiar with the product and its desired sensory characteristics, being requested therefore to compare products with the control element regarding their compliance to target characteristics. Sensory tests followed the procedure in section 2. In view of cost and time constraints, only three samples were used as control elements in the sensory tests (i.e., N = 10 and M = 3).

Run	1	2	3	4	5	6	7	8	9	10
W1	-1	-1	1	1	0	-1,41	1,41	0	0	0
W2	-1	1	-1	1	0	0	0	-1,41	1,41	0
Thick	-1	-1	-1	-1	-1	1	1	1	1	1
I1	46,4	32,8	16,6	39,7	23,6	7,4	30,5	14,3	0,7	23,6
I2	16,6	32,8	46,4	7,4	23,6	39,7	0,7	14,3	30,5	23,6
I3	7,3	4,7	7,3	23,2	23,2	23,2	39,1	41,7	39,1	23,2

Table 4. Experimental runs in terms of coded independent variables (*W*'s and *Thick.*) and mixture ingredients (*I*'s).

Panelists are numbered 1 to 5. There are three judgement matrices for each panelist, each corresponding to a different control sample. Results for panelist 1 on texture are presented in Table 5; matrices are identified $\mathbf{A}_{11}^{(m)}$, m = 1,2,3; the matrix of midpoints is identified as $\overline{\mathbf{A}}_{11}$. Results for other panelists are given in form of their texture weight vectors ($\mathbf{w}_{k1}$, k = 1,…,5) and consistency measures (CR), obtained from their midpoint matrices. Note that panelist 1's CR is above the threshold value of 0.1; all other panelists, however, displayed very low CRs. We decided to keep panelist 1's evaluations in spite of their low consistency, since the CRs will be used latter to assign weights to evaluations from different panelists, which will penalize 1's lack of consistency. General appearance results are also given in Table 5; we restrict ourselves to present panelists' weight vectors ($\mathbf{w}_{k2}$, k = 1,…,5) and CRs, obtained from midpoint matrices.

The panel leader subjectively rated panelists regarding their excellence in performing sensory evaluations. Ratings were numbers in the [0,1]-interval. The resulting normalized vector of ratings is named *Exc*, and presented in the lower portion of Table 5. In that vector, panelists with good knowledge of the product under study and with past experience in sensory panels were given high ratings. In addition, the reciprocals of the consistency ratios were used as importance weights to panelists. The resulting normalized vector of reciprocals is named 1/*CR* and presented in the lower portion of Table 5.

$\mathbf{A}_{11}^{(1)}$

	1	2	3	4	5	6	7	8	9	10
1	1,00	8,03	1,08	0,97	1,05	0,93	0,96	4,21	0,90	67,21
2	0,12	1,00	0,14	0,12	0,13	0,12	0,12	0,53	0,11	8,37
3	0,92	7,40	1,00	0,90	0,97	0,86	0,88	3,89	0,83	61,97
4	1,03	8,23	1,11	1,00	1,08	0,95	0,98	4,32	0,92	68,96
5	0,95	7,61	1,03	0,92	1,00	0,88	0,91	4,00	0,85	63,71
6	1,08	8,65	1,17	1,05	1,14	1,00	1,03	4,54	0,97	72,45
7	1,04	8,37	1,13	1,02	1,10	0,97	1,00	4,40	0,94	70,12
8	0,24	1,90	0,26	0,23	0,25	0,22	0,23	1,00	0,21	15,95
9	1,11	8,93	1,21	1,08	1,17	1,03	1,07	4,69	1,00	74,78
10	0,01	0,12	0,02	0,01	0,02	0,01	0,01	0,06	0,01	1,00

$\mathbf{A}_{11}^{(2)}$

1	1,00	0,79	9,96	24,13	12,32	0,79	1,90	1,25	3,09	14,04
2	1,27	1,00	12,65	30,66	15,65	1,00	2,41	1,59	3,92	17,84
3	0,10	0,08	1,00	2,42	1,24	0,08	0,19	0,13	0,31	1,41
4	0,04	0,03	0,41	1,00	0,51	0,03	0,08	0,05	0,13	0,58
5	0,08	0,06	0,81	1,96	1,00	0,06	0,15	0,10	0,25	1,14
6	1,27	1,00	12,65	30,66	15,65	1,00	2,41	1,59	3,92	17,84
7	0,53	0,41	5,25	12,71	6,49	0,41	1,00	0,66	1,63	7,40
8	0,80	0,63	7,94	19,24	9,82	0,63	1,51	1,00	2,46	11,19
9	0,32	0,25	3,23	7,82	3,99	0,25	0,61	0,41	1,00	4,55
10	0,07	0,06	0,71	1,72	0,88	0,06	0,14	0,09	0,22	1,00

$\mathbf{A}_{11}^{(3)}$

1	1,00	1,50	2,56	0,05	0,61	0,14	0,05	0,26	0,05	0,37
2	0,67	1,00	1,71	0,03	0,41	0,09	0,03	0,18	0,03	0,25
3	0,39	0,58	1,00	0,02	0,24	0,05	0,02	0,10	0,02	0,15
4	19,94	29,81	51,12	1,00	12,15	2,73	0,91	5,27	0,91	7,47
5	1,64	2,45	4,21	0,08	1,00	0,22	0,07	0,43	0,08	0,61
6	7,31	10,93	18,75	0,37	4,45	1,00	0,33	1,93	0,34	2,74
7	21,98	32,87	56,35	1,10	13,39	3,01	1,00	5,81	1,01	8,23
8	3,78	5,66	9,70	0,19	2,30	0,52	0,17	1,00	0,17	1,42
9	21,80	32,59	55,88	1,09	13,28	2,98	0,99	5,76	1,00	8,17
10	2,67	3,99	6,84	0,13	1,63	0,37	0,12	0,71	0,12	1,00

$\overline{\mathbf{A}}_{11}$

	1	2	3	4	5	6	7	8	9	10
1	1,00	3,00	4,54	1,80	3,60	0,62	0,17	1,10	0,20	23,50
2	0,33	1,00	3,00	0,50	1,80	0,40	0,06	0,42	0,06	8,00
3	0,22	2,00	1,00	0,08	0,82	0,33	0,36	1,37	0,39	34,50
4	0,56	2,00	12,84	1,00	4,00	0,11	0,13	0,33	0,23	23,00
5	0,28	0,56	1,23	0,25	1,00	0,39	0,15	0,42	0,18	23,50
6	1,62	2,49	3,04	9,00	2,57	1,00	0,80	2,00	1,25	31,01
7	6,00	18,00	2,75	7,50	6,67	1,25	1,00	3,50	1,19	28,58
8	0,91	2,40	0,73	3,00	2,40	0,50	0,29	1,00	0,25	9,52
9	5,00	17,01	2,59	4,44	5,56	0,80	0,84	4,00	1,00	29,17
10	0,04	0,13	0,03	0,04	0,04	0,03	28,58	0,11	0,03	1,00

	1	2	3	4	5	6	7	8	9	10	*CR*
$\mathbf{w}_{11}$	0,1	0,04	0,05	0,1	0	0,2	0,26	0,1	0,2	0,01	0,20
$\mathbf{w}_{21}$	0	0,04	0,04	0,1	0	0,1	0,11	0,2	0,4	0,05	0,04
$\mathbf{w}_{31}$	0,1	0,08	0,06	0,1	0	0,1	0,12	0,1	0,2	0,11	0,02
$\mathbf{w}_{41}$	0	0,04	0,04	0,1	0,1	0	0,2	0,1	0,3	0,07	0,03
$\mathbf{w}_{51}$	0,1	0,08	0,08	0,1	0,1	0,1	0,15	0,1	0,2	0,05	0,01

	1	2	3	4	5	6	7	8	9	10	*CR*
$\mathbf{w}_{12}$	0,1	0,01	0,32	0	0,4	0	0,09	0,1	0	0,01	0,10
$\mathbf{w}_{22}$	0,1	0,06	0,15	0,1	0,1	0,1	0,1	0,1	0,1	0,02	0,02
$\mathbf{w}_{32}$	0,1	0,06	0,06	0,1	0,1	0,1	0,28	0,1	0,1	0,02	0,03
$\mathbf{w}_{42}$	0	0,04	0,05	0,2	0,1	0,1	0,18	0,1	0,2	0,03	0,02
$\mathbf{w}_{52}$	0,1	0,09	0,16	0,1	0,1	0,1	0,14	0,1	0,1	0,02	0,00

k	*Exc*	1/*CR*
1	0,04	0,02
2	0,12	0,09
3	0,20	0,20
4	0,28	0,13
5	0,36	0,57

Table 5. Panelist 1 judgement matrices on texture; panelist 1's matrix of midpoints on texture; panelists' weight vectors and CRs on texture and appearance; panelists' ratings on excellence and CR.

Our objective is to build regression models for texture and general appearance that would allow us to optimize the product regarding their sensory properties. We chose to build aggregate models for the attributes, instead of modeling panelists' individually. Therefore, we combine the texture and appearance weight vectors using the following procedure. Combine vectors *Exc* and 1/*CR* to obtain a single vector of ratings, say $\mathbf{r}$, for the panelists. We used a weighted sum of vectors, with weights 0.3 and 0.7 for *Exc* and 1/*CR*, respectively (i.e., $\mathbf{r} = (0.3 \times Exc) + (0.7 \times 1/CR)$). Entries in $\mathbf{r}$ are given by $[r_1,\ldots, r_5]'$. Arrange vectors $\mathbf{w}_{1p},\ldots, \mathbf{w}_{5p}$ as columns of a matrix $\mathbf{W}_p$. The combined weight vector for attribute p (p = 1,2) is given by $\mathbf{w}_p = \mathbf{W}_p\mathbf{r}$. Using the information in Table 5, we arrive at the following combined weight vectors:

$$\mathbf{w}_1 = [0.080;\ 0.069;\ 0.066;\ 0.079;\ 0.053;\ 0.101;\ 0.149;\ 0.122;\ 0.218;\ 0.062\,] \quad (7)$$

$$\mathbf{w}_2 = [0.086;\ 0.065;\ 0.124;\ 0.104;\ 0.102;\ 0.096;\ 0.161;\ 0.119;\ 0.119;\ 0.022\,] \quad (8)$$

Regression models are determined relating vectors in eqs. (7) and (8) with independent variables *W*'s and *Thick* in Table 4. The resulting models, including terms that are at least 95% significant, are (R^2 refer to the coefficient of determination):

$$\begin{aligned} &Texture = 0.075 + 0.03 Thick + 0.031(W_2 \times W_2) \quad (R^2 = 0.630) \\ &Appearance = 0.109 + 0.021 W_1 - 0.005 W_2 + 0.014 Thick + 0.005(Thick \times W_2) \quad (R^2 = 0.972) \end{aligned} \quad (9)$$

Note that the models above present reasonable fit, considering the variability inherent to sensory panel data.

Once the regression models for *Texture* and *Appearance* are at hand, we are able to perform different optimizations on the experimental results. To illustrate that, we determine the appropriate control factors settings such that the attributes *Texture* and *Appearance* are at their best. Recall that in the weight vectors derived for each panelist samples were given weights in the [0,1] interval, such that values near 1 denote good samples. Initially, we performed a nonlinear search for the best settings of control factors *W*'s and *Thick*, such that the attribute *Texture* was optimized. In our search, the regression model for *Texture* in eq. (9) was the objective function to be maximized, and the search was limited to the design region in Table 4. The best settings for the control factors, in terms of the mixture ingredients, was determined to be

I_1 = 3.07% I_2 = 10.85% I_3 = 56.38% *Thick* = +1.

We repeated the procedure above, using this time the regression model for *Appearance* in eq. (9) as the objective function to be maximized. The search was constrained by the design region, as previously. The best settings for the control factors was determined to be

I_1 = 6.65% I_2 = 6.65% I_3 = 57.0% *Thick* = +1.

As expected, the two sets of control factor settings above do not coincide. That is usually the case when optimizing experiments regarding more than one response variable. Although limiting ourselves to the two optimization exercises above, we direct the reader to the works of DERRINGER & SUICH (1980), PIGNATIELLO (1993) and FOGLIATTO *et al.* (1999) to carry the optimization a step further in search of a global optimum.

5. Conclusion

We propose a new method for sensory data collection and analysis based on psychophysical scaling methods. Our method yields quantitative data, which can be used for model building purposes, and generates an efficiency measure for subjects, the consistency ratio. Using the consistency ratio, the efficiency of different subjects may be taken into account when combining their evaluations into a single vector of evaluations.

The proposed method is applied in a case study from the food industry, where 5 panelists evaluate the intensity of two attributes in 10 samples obtained from a designed experiment. We generate intensity scorings of samples using evaluations from each panelist, and determine their consistency ratios. Evaluations from the panelists are then combined using their consistency scores and a subjective assessment of their efficiency as weights. Combined weights are modeled as function of the control factors and partial optimizations are carried out.

6. References

AMERINE, M.A.; PANGBORN, R.M.; ROESSLER, E.B. **Principles of Sensory Evaluation of Food.** New York, Academic Press, 1965.

CORNELL, J.A. **Experiments with mixtures – designs, models, and the analysis of mixture data**. New York, John Wiley, 1990.

DERRINGER, G.; SUICH, R. Simultaneous optimization of several response variables. **J. Quality Technology**, Vol. 12, n. 4, 1980, p. 214-219.

ENGEN, T.; LEVY, N. The Influence of Standards on Psychophysical Judgements. **Perceptual and Motor Skills**, Vol. 5, 1955, p. 193-197.

ERIKSEN, C.W.; HAKE, H.W. Absolute Judgments as a Function of the Stimulus Range and the Number of Stimulus and Response Categories. **J. Experimental Psychol.**, Vol. 49, 1955, p. 323-332.

FOGLIATTO, F.S.; ALBIN, S.L.; TEPPER, B.J. A hierarchical approach to optimizing descriptive analysis multiresponse experiments. **J. Sensory Studies**. Vol. 14, n. 4, 1999, p. 443-465.

GARNER, W.R. An Informational Analysis of Absolute Judgments of Loudness. **J. Exp. Psychol.** Vol. 46, 1953, p. 373-380.

JACOBS, B.E.; MOSKOWITZ, H.R. Magnitude Estimation: Scientific Background and Use in Sensory Analysis. *In* **Applied Sensory Analysis of Foods**, H. Moskowitz (Ed.), Boca Ratón, CRC Press, 1988.

LEIGHT, R.S.; WARREN, C.B. Standing Panels Using Magnitude Estimation for Research and Product Development. *In* **Applied Sensory Analysis of Foods**, H. Moskowitz (Ed.), Boca Ratón, CRC Press, 1988.

MEILGAARD, M.; CIVILLE, G.V.; CARR, B.T. **Sensory Evaluation Techniques**.2nd Ed., Boca Ratón, CRC Press, 1991.

MOSKOWITZ, H.R. Ratio Scales of Sugar Sweetness. **Perc. & Psychophysics**, Vol. 7, 1970, p. 315-320.

MOSKOWITZ, H.R. Intensity Scales for Pure Tastes and Taste Mixtures. **Perc. & Psychophysics**, Vol. 9, 1971, p. 51-56.

MOSKOWITZ, H.R. Magnitude Estimation: Notes on How, When, and Why to Use It. **J.Food Qual.**, Vol. 3, 1977, p. 95-227.

PIGNATIELLO Jr., J.J. Strategies for robust multiresponse quality engineering. **IIE Transactions**, Vol. 25, 1993, p. 5-15.

POLLACK, I.; FICKS, L. Information of Elementary Multidimensional Auditory Displays. **JASA**. Vol. 26, 1954, p. 155-158.

POLLACK, I. Information of Elementary Auditory Displays (II). **JASA**. Vol. 25, 1953, p. 765-769.

REYNOLDS, G.S.; STEVENS, S.S. Binaural Summation of Loudness. **JASA**, Vol. 32, 1960, p. 1337-1344.

SAATY, T.L. A Scaling Method for Priorities in Hierarchical Structures. **Journal of Math. Psychology**, Vol. 15, 1977, 234-281.

SAATY, T.L. **The Analytic Hierarchy Process**. New York, McGraw-Hill, 1980.

STEVENS, S.S. On the Psychological Law, **Psychological Rev.** Vol. 64, 1957, p.153-181.

STONE, H.; SIDEL, J.L. **Sensory Evaluation Practices**. 2nd Ed. New York, Academic Press, 1993.

STONE, H., SIDEL, J.; OLIVER, S.; WOOLSEY, A.; SINGLETON, R.C. Sensory Evaluation by Quantitative Descriptive Analysis. **Food Technology**. Vol. 28, n. 1, 1974, p. 24, 26, 28, 29, 32, 34.

STRANG, G. **Linear algebra and its applications**. 3rd Ed. Fort Worth, Saunders College, 1988.

VARGAS, L.G. Reciprocal Matrices with Random Coefficients. **Math. Model.**, Vol. 3, 1982, p. 69-81.