

Potential of Semiconductor Sensor Arrays for the Origin Authentication of Pure Valencia Orange Juices

Christophe Steine,^{*,†,‡} Frédéric Beaucousin,[†] Chhan Siv,[‡] and Gilbert Peiffer[‡]

Centre de Recherche et de Valorisation des Produits de la Consommation (SGS-CERVAC), Parc Club du Golf, Bât. 8, 13856 Aix-en-provence Cedex 3, France, and Laboratoire des Organo-Phosphorés, Faculté des Sciences et Techniques de Saint Jérôme, Avenue Escadrille Normandie Niemen, 13397 Marseille Cedex 13, France

Sensor array is a new method used in aroma analysis. This technique was employed for the differentiation of 49 pure Valencia orange juices from 5 different origins representative of the main culture areas of citrus according to their volatile organic fractions. An experimental design was used to set analytical factors, which control the generation of the static headspace; therefore, the preliminary optimization of analytical conditions allows the evaluation of the potential of such an apparatus in this kind of survey. Results obtained were then statistically treated by multivariate analyses such as principal component analysis and factorial discriminant analysis. Sensor arrays performed a good discrimination of the whole juices in classing them according to the origin of Valencia oranges used. A calibration of the sensors was then performed to build a database aiming to classify the juices according to the origin of oranges involved. New supplementary samples were then analyzed to assess the efficiency of the database. This technique coupled with other ones such as high-pressure liquid chromatography and gas chromatography may be a new tool for the investigation of adulteration detection due to the authentication of the origin of raw materials employed in the orange juice processes.

Keywords: Orange juice; origin; adulteration, authentication, sensor array; volatile organic compounds; static headspace; optimization of analysis; multivariate analysis; pattern recognition; calibration of the sensors

INTRODUCTION

The association of one or several arrays of electronic chemical sensors with partial specificity to an appropriate system of pattern recognition based on statistical treatments enabled the conception of instruments known as "electronic noses" (1). The main technologies of sensors that equip this kind of apparatus are the following: metal oxide semiconductor sensors (Mox), gas sensitive field effect transistors (GasFETs or MosFETs), conducting polymers, acoustic wave devices such as the buck acoustic waves (BAW) often referred to as quartz crystal microbalance (QCM), and surface acoustic wave transducers (SAW) (2). Contrarily to the classical techniques used in aroma analysis such as headspace gas chromatography combined or not with mass spectrometry, this new method does not identify the composition of volatile compounds but gives a fast comparative measure of patterns of odors, representative of compounds disengaged by a substratum (3). In the case of Mox sensors, interactions with these kinds of compounds induce mechanisms of adsorption and desorption, taking place on the surface of the sensors and provoking the modification of its electrically measurable properties by a variation of resistance versus time (4).

The domains of utilization of electronic noses concern principally the food industries (5, 6) but also the

industries of tobacco, some cosmetics industries, and the automotive industry (7). In the case of beverages, this technology has already been applied for the authentication of wines and the discrimination of fruit juice and coffees. It permits the detection of contaminants such as the presence of diacetyl in orange juices or trichloroanisoles in wines (8). A few studies relative to orange juices were found in the literature (9). Bazemore differentiated some orange juices from Florida according to the harvest times of oranges and of their pasteurization level with a system equipped with 12 polymer sensors (10).

The aromatic fraction of orange juices is essentially composed of alcohols, aldehydes, esters, and hydrocarbons, one of which is limonene, the major compound (11). During these past years, industries have developed new types of products in the domain of the fruit juice. The consumer of today has available a range of innovative products divided into three main branches: beverages known for their specific health advantages, fruit juices of superior quality (not from concentrate, from certified origin fruits), and ready-to-drink products with all-point distribution (12). As for orange juices, the notion of quality is closely bound to the geographical origin of the oranges involved. It is therefore necessary to create new analytical methods revealing the origin of the processed raw products.

A recent survey showed that high-performance liquid chromatography (HPLC) allows the geographical recognition of some orange juices due to the detection of some chemical compounds in their composition (13).

* Author to whom correspondence should be addressed [telephone +33 4 95 05 00 00; fax + 33 4 95 05 00 48].

[†] Centre de Recherche et de Valorisation des Produits de la Consommation (SGS-CERVAC).

[‡] Laboratoire des Organo-Phosphorés.

Table 1. Sensors of the Fox3000 Electronic Nose^a

sensor	sensor class	P type	T type	description
1	Mox		T30/1	polar compounds, ethanol
2	Mox	P10/1		hydrocarbon
3	Mox	P10/2		methane and propane
4	Mox	P40/1		fluoride, chloride, aldehydes
5	Mox		T70/2	aromatic compounds
6	Mox	PA2		alcohol, solvents
7	Mox	P30/1		polar compounds, ethanol
8	Mox	P40/2		fluoride and chloride
9	Mox	P30/2		fluoride, chloride, aldehydes
10	Mox	P70/1		volatiles from carbonized compounds
11	Mox		T40/1	fluoride, chloride, aldehydes
12	Mox		TA2	organic solvent

^a These 12 sensors are not selective but known to be more sensitive to chemical products belonging to chemical families or kinds of compounds cited in the table.

Some encouraging results were also given by some isotopic analyses, thanks to the SNIFNMR method (nuclear magnetic resonance applied to specific natural isotopic fractions) (14). These techniques are appreciable but do not always permit an accurate analysis of the aroma, which often results from mechanisms of complex interaction between the chemical compounds, which compose it. The analysis by sensors avoids the need for long presampling contrary to the other techniques and consequently preserves the chemical, physical, and organoleptic properties of the orange juices practically intact. The objective of this paper is to show the potential of a sensor array for origin recognition. We have to constitute new kinds of databases on pure Valencia oranges juice, the most widely grown variety in the world (15). The pure Valencia orange juices analyzed came from five different origins, representative of different producer countries or states: Israel, Spain, Belize, Cuba, and Florida (USA). These regions/countries belong to the two main climatic areas of culture of citrus fruits: the Mediterranean and tropical areas.

MATERIALS AND METHODS

Samples. Studies were carried out on 49 pure Valencia orange juices of 5 different origins; the samples were distributed as follows: Israel, 12; Spain, 11; Belize, 8; Florida, 9; Cuba, 9. All of these juices came from oranges that were harvested during two periods: 1996–1997 and 1997–1998. These pure industrial juices worked out according to processes of manufacture conforming with the standard norms: washing, selection of the fruits, calibration, extraction in line, FMC process (Food Machinery Corp., Chicago, IL), pasteurization, and conditioning. The FMC process allows limited contact between the orange juice and the skin of the fruit during the extraction, so the presence of essential oils such as limonin is limited in the juice. Furthermore, oranges from different sizes can be used in this process.

Reagents. The carrier gas used is synthetic dry air of FID quality. The deionized water introduced in the air-conditioning unit, an ACU 500 (Alpha Mos, Toulouse, France), is of HPLC grade. The following compounds have been used as standards for the calibration of sensors: guaiacol, α -pinene, furfural, limonene, ethyl butyrate, and γ -butyrolactone. Some of these products were diluted in propylene glycol. All of these chemicals were of analytical grade (Sigma-Aldrich).

Sensor Array. The analytical system is composed of the following modules: a Fox3000 electronic nose (Alpha Mos), which includes 2 rooms of 12 Mox sensors (Table 1) (each room is a pack of 6 sensors allowing one to update the system up to 3 rooms with other kinds of sensors); electronic devices and a mass flow meter; an ACU humidifier 500 (Alpha Mos), an automatic headspace sampler HS-50 (CTC Analytics), and a

Table 2. Summary of Dilutions of Pure Standards for the Calibration of Sensors

compound ^a	solvent	conc'n (% v/v)	injection vol (μ L)
limonene	none	pure (10 μ L)	200
ethyl butyrate	propylene glycol	0.02	200
α -pinene	none	pure (10 μ L)	200
guaiacol	none	pure (100 μ L)	200
γ -butyrolactone	none	pure (500 μ L)	200
furfural	none	pure (20 μ L)	200

^a Chemical compounds used as standards for the calibration of the sensors.

Table 3. Range of Influent Factors

factor ^a	level -1 ^b	level 0 ^b	level +1 ^b
X_1 , temp ($^{\circ}$ C)	40	50	60
X_2 , generation time (min)	5	10	15
X_3 , injection vol (μ L)	150	225	300

^a The three influent factors studied concerning the headspace generation. ^b Range of factors set at three levels: -1, 0, +1.

personal computer (Compaq Deskpro) used for the acquisition data realized with the Fox3000 5.0 software (Alpha Mos).

Experimental Procedure. The volatile organic fractions of pure Valencia orange juices were analyzed with the Fox3000 electronic nose. For each sample, 2 mL of pure juice was introduced to ambient temperature and sealed in vials of 10 mL with a septum. The static headspaces were generated by the automatic sampler HS-50 (CTC Analytics): pure juices were heated at a temperature of 50 $^{\circ}$ C for 15 min. The amount (150 μ L) of the headspace is automatically taken from the sample vial by a microsyringe Hamilton, thermostatically controlled at 55 $^{\circ}$ C in order to avoid condensation phenomena. The gas phase composed of volatile organic components is then injected and routed into the two rooms of the sensors by a flux of synthetic air humidified with the ACU 500 (RH = 20 \pm 2%, T = 36 $^{\circ}$ C) at the constant flow level of 300 mL/min controlled by the mass flow meter and set at the pressure of 5 psi. It is necessary to work at a constant level of hygrometry due to the known sensitivity of Mox sensors to the water (16, 17). The acquisition time is set at 180 s: for each of the 12 sensors, the Fox3000 software records the variation of resistance versus time divided by the original sensor resistance (R_0) in the carrier gas: $\Delta R/R_0 = f(t)$. The default run time between injections is set at 20 min: this is the time required for the return to baseline of each sensor. The samples are analyzed according to a definite uncertain sequence after randomization, to limit all memory effect. Standards used for the sensors' calibration are analyzed in the same way except that they are heated at a temperature of 45 $^{\circ}$ C for 3 min. Preparation of these standards is summarized in Table 2.

Optimization of Analysis. The optimum analytical conditions for the generation and analysis of headspace by the HS 50 sampler were determined by a series of 13 experiments, from an experimental design in which three factors were varied simultaneously (oven temperature, headspace generation time, and injection volume). The levels used for each of the parameters, which were varied, are shown in Table 3. The experiments were performed using pure juices from Spain, Israel, and Florida, taken at random in our sampling. We limited the analyses to these three origins because of the number of requisite tests (experiments and repetitions) in relation to the analytical limits of the sampler HS 50 (maximal number of analysable samples = 50).

DATA ANALYSIS

Mathematical Algorithm and Data Preprocessing. A sensor i in the presence of a complex mixture of volatile compounds could be at the origin of an odor j , which produces an electric time-dependent signal: $U_{ij}(t)$. Different mathematical algorithms could be used to

disengage the analytical information produced by the sensor array of Mox sensors (1).

$$x_{ij} = \frac{R_{ij}^{\max} - R_{ij}^{\min}}{R_{ij}^{\min}} = \max\left(\frac{\Delta R}{R_0}\right)$$

R_{ij} is the resistor of sensor i in the presence of a mixture of volatile j , R_{ij}^{\max} is the maximal measured value, $R_{ij}^{\min} = R_0$ is the initial resistor of i sensor balances in the synthetic air, and x_{ij} is the response of sensor i recorded.

In the case of n analyzed samples, we therefore get a matrix of n lines and 12 columns (sensors). Normalization of the data contributes to decreasing the influence of the volatile compound concentration related to the sensor responses (18): we have therefore chosen to use this normalization to differentiate the pure orange juice qualitatively.

Similarity Index. The similarity index gives an analytical measure of the difference between two samples analyzed by the Fox3000 electronic nose (19). After statistical treatment, samples are classified by groups or classes of samples. To be classified in the same group, two compared samples must have an index of <10 units, an empirical value fixed for the use of the Fox3000 electronic nose. This index is the so-called intragroup similarity index. Reciprocally, samples differently classified have an index >10 units, called the intergroup index, which thus assesses the discrimination power of the analysis (20).

$$\text{similarity index} = \frac{\sqrt{\sum_{i=1}^N (x_{1ij} - x_{2ij})^2}}{\sqrt{\sum_{i=1}^N \left(\frac{x_{1ij} - x_{2ij}}{2}\right)^2}} \times 100$$

x_{1ij} is the fractional difference measured for sample 1, x_{2ij} is the fractional difference measured for sample 2, and N is the number of sensors.

Multivariate Statistical Analysis. The data obtained are treated by multivariate statistical analysis, more specifically, by principal component analysis (PCA) and factorial discriminant analysis (FDA). These analyses were done on 42 Valencia pure orange juices taken from the sampling, which contained 49 samples. The 7 remaining pure juices, representative of the whole of the countries of origin studied, constitute the supplementary samples: they have been randomly chosen among the whole juices (2 for Spain and Israel, 1 for Belize, Florida, and Cuba). These tools enable the reduction of the dimension of a data set consigned in a matrix while conserving the maximum information loaded by the sensors.

PCA is a nonparametric technique often used by analysts notably in the data treatment in mass spectrometry (21). Its utilization concerning semiconductor sensors is found in the literature (22). PCA enables one to find some linear combinations of sensor responses expressing the maximum of the representative variance of the data. Each linear combination is called a principal component or an eigenvector. Generally, the maximum of the variance appears therefore thanks to the two or three first components: we can thus get from the 12th

Table 4. Optimization of Analysis

expt ^a	X_1^b	X_1^c	X_1^d	X_2^c	X_2^d	X_3^d	X_1X_2	X_2X_3
A	40	1600	64000	5	25	225	200	1125
E	40	1600	64000	10	100	150	400	1500
F	40	1600	64000	10	100	300	400	3000
B	40	1600	64000	15	225	225	600	3375
G	50	2500	125000	5	25	150	250	750
K	50	2500	125000	5	25	300	250	1500
M	50	2500	125000	10	100	225	500	2250
J	50	2500	125000	15	225	150	750	2250
L	50	2500	125000	15	225	300	750	4500
C	60	3600	216000	5	25	225	300	1125
H	60	3600	216000	10	100	150	600	1300
I	60	3600	216000	10	100	300	600	3000
D	60	3600	216000	15	225	225	900	3375

^a The 13 experiments of this three-level experimental design. ^b X_1 , temperature (°C). ^c X_1 , time (min). ^d X_3 , injected quantity (μL). (X_1)ⁿ, X_1X_2 , and X_2X_3 are the interaction factors.

dimension to the 2nd or 3rd ones and thus give a graphic representation of the information contained in the initial matrix.

FDA enables one, in addition to the PCA, to take into account the existence of populations of known samples, in our case, samples coming from five different countries. FDA regroups in populations the 42 pure Valencia orange juices and discriminates them. These populations could be then used as new references: we can thus test some new observations of the 7 remaining pure Valencia orange juices that were not used in the discrimination (23). The use of these supplementary samples is a way to test the validity of the classification of the whole samples with FDA.

Pure juice data were processed with UNISTAT 3.0 software (Megalon and UNISTAT) for statistical treatments. The raw acquisition files from the Fox3000 5.0 software (Alpha-Mos) were computed with Excel97 (Microsoft Corp.) and then exported into UNISTAT 3.0. PCA and FDA figures were made up with the Power-Point97 program (Microsoft Corp.).

RESULTS AND DISCUSSION

Developing the Method. Analysis of volatile compounds via sensor arrays depends greatly on the analytical technique, which enables one to generate them, the static headspace. Often used in gas chromatography, this process depends on many parameters to be fixed. We have first to find the analytical optimal conditions, the more repeatable headspaces and thus the most discriminating possible. Few works concerning aroma sensors relate the use of experimental devices (24) to optimize analytical conditions. This methodology is a powerful tool, which considers interaction factors and allows one to evaluate the potential of sensor arrays in the discrimination of single-variety pure orange juice according to the origin of the oranges involved in the process. The sequence of the experiment device (A–M) was random in order to limit any memory effects bound to the constancy of some factors (Table 4).

The intra- and intergroup similarity indices (Si) were calculated to give a measure of the repeatability and also the sensor's powers of discrimination. Values of the similarity indices (Table 5) show that conditions are more interesting: we have to choose experiments with the smallest intragroup similarity indices and the highest intergroup ones. We fit quadratic models to the experimental results, the mean intra- and intergroup similarity indices: We obtain two quadratic equations

Table 5. Intergroup and Intragroup Similarity Indices

expt	Si ^a			Si(intra) ^d exptl	Si(intra) ^e calcd	Si ^b			Si(inter) ^f exptl	Si(enter) ^g calcd
	(S,S) ^c	(F,F) ^c	(I,I) ^c			(S,F) ^c	(S,I) ^c	(F,I) ^c		
A	4	3	5	4	4	9	10	4	8	8
B	1	2	3	2	1	8	9	3	7	8
C	2	3	2	2	3	9	6	4	6	6
D	2	2	1	2	2	7	5	4	5	6
E	3	3	3	3	4	22	24	4	17	16
F	0	3	1	1	1	15	16	3	11	10
G	6	8	7	7	7	25	22	11	19	19
H	1	6	8	5	4	22	14	11	16	15
I	1	3	2	2	2	11	10	4	8	8
J	3	3	2	3	3	29	15	15	20	19
K	1	3	1	2	2	21	10	12	14	14
L	2	2	2	2	2	19	10	11	13	12
M	2	4	5	4	4	26	10	19	18	21

^a Intragroup similarity index (Si) giving a measure of the repeatability for pure Valencia orange juices originating from Spain Si(S,S), Florida Si(F,F), and Israel Si(I,I). ^b Intergroup similarity index (Si) measuring the discrimination of pure Valencia orange juices from the three different origins studied Si(X, Y): X and Y = S (Spain), F (Florida), I (Israel). ^c Each sample is analyzed three times. ^d Si(intra) is the mean of the experimental Si^a. ^e The calculated value is obtained thanks the quadratic equations. ^f Si(inter) is the mean of the experimental Si^b. ^g The calculated value is obtained via quadratic equations.

Table 6. Sensor Responses for Pure Valencia Orange Juices from Tropical and Subtropical Origins and from Mediterranean Origin [Values of Maximal Fractional Resistance Change Measured: max($\Delta R/R_0$)]

sensor ^a	Belize max($\Delta R/R_0$)			Florida max($\Delta R/R_0$)			Cuba max($\Delta R/R_0$)			Spain max($\Delta R/R_0$)			Israel max($\Delta R/R_0$)		
	mean ^b	SD ^c	CV ^d (%)	mean ^e	SD ^c	CV ^d (%)	mean ^f	SD ^c	CV ^d (%)	mean ^h	SD ^c	CV ^d (%)	mean ⁱ	SD ^c	CV ^d (%)
T30/1	0.426	0.023	5.418	0.135	0.099	73.276	0.005	0.013	282.843	0.264	0.105	39.581	0.117	0.086	73.729
P10/1	0.399	0.012	2.918	0.401	0.015	3.779	0.416	0.015	3.483	0.481	0.012	2.532	0.516	0.022	4.312
P10/2	0.355	0.009	2.566	0.353	0.010	2.777	0.358	0.010	2.733	0.381	0.010	2.621	0.405	0.015	3.696
P40/1	0.227	0.007	3.165	0.219	0.009	3.912	0.226	0.008	3.528	0.262	0.007	2.542	0.278	0.012	4.476
T70/2	0.513	0.013	2.553	0.483	0.028	5.757	0.438	0.019	4.343	0.493	0.018	3.609	0.483	0.015	3.039
PA2	0.669	0.018	2.713	0.663	0.015	2.260	0.658	0.021	3.164	0.675	0.018	2.687	0.712	0.025	3.478
P30/1	0.996	0.010	0.989	0.891	0.054	6.070	0.600	0.127	21.239	0.999	0.004	0.447	0.934	0.075	7.987
P40/2	0.584	0.010	1.734	0.583	0.019	3.228	0.540	0.015	2.833	0.569	0.020	3.478	0.574	0.014	2.452
P30/2 ^g	0.940	0.029	3.063	1.000		1.000				0.941	0.049	5.194	0.998	0.006	0.648
P70/1	0.005	0.001	11.376	0.007	0.003	45.967	0.004	0.003	82.200	0.005	0.001	8.531	0.001	0.011	274.331
T40/1	0.084	0.004	4.379	0.070	0.003	4.683	0.058	0.004	6.964	0.070	0.004	5.035	0.063	0.004	6.917
TA2	0.087	0.004	4.050	0.080	0.003	3.125	0.072	0.003	3.653	0.074	0.002	3.205	0.069	0.004	5.611

^a Names of sensors. ^b Means of 7 determinations. ^c Standard deviation. ^d Coefficient of variation. ^e Means of 8 determinations. ^f Means of 8 determinations. ^g P30/2 given the highest responses to the organic volatile; some of its values equal 1.000 because of the normalization of the data. ^h Means of 9 determinations. ⁱ Means of 10 determinations.

to calculate values of similarity indices versus the experimental factors set.

$$\text{Si(intra)} = 21.1838256 - 0.87622888X_1 + 0.2676839X_1^2 - 0.00024767X_1^3 - 0.84938838X_2 - 0.01617737X_2^2 - 0.04234964X_3 + 0.01X_1X_2 + 0.00232416X_2X_3$$

$$\text{Si(inter)} = 68.8289425 - 8.25964335X_1 + 0.25360059X_1^2 - 0.0022638X_1^3 + 4.35795107X_2 - 0.20030581X_2^2 - 0.0248787X_3 - 0.00178593X_2X_3$$

Si(inter) is the calculated intragroup similarity index, and Si(intra) is the intragroup one. X_i ($1 \leq i \leq 3$) are the influent factors (see Table 3), and X_1^2 , X_2^2 , X_1X_2 , and X_2X_3 are the interaction factors (see Table 4).

We notice that all of the conditions give good intragroup Si except for experiments G and H. In the case of experiment G, the generation time is certainly too low (5 min) for the injected quantity (150 μ L). Concerning experiment H, repeatability is very good for orange juices from Spain [Si(S,S) = 1] but not for juices from Israel and Florida. These remarks are closely bound to

the complex reaching of equilibrium of generated static headspaces. The best intergroup similarity indices are obtained for the medium temperature of 50 °C in the case of experiments G and J–M, with Si values >10 for the whole countries. The average of Si (S,F), Si(S,I), and Si(F,I) is higher for experiment J: By comparing experimental values to the calculated ones, we conclude the validity of the experimental design. We have then chosen this one to set the analytical conditions and to do all of the analyses: generation time, 15 min; temperature, 50 °C; injected quantity, 150 μ L.

Sensor Responses. The kinetics obtained for the whole of the pure juice is in sigmoid form. This shape results from complex phenomena of transfers of matters and interactions between volatile compounds and the sensors: these interactions result from mechanisms of adsorption, desorption, competition, and catalysis on the surface of sensors. The T30/1 and P70/1 sensors have the largest coefficients of variation concerning the studied origins (Table 6): they do not respond regularly to the volatile compounds generated in the headspace, and we do not take them into account in the results. The Mox sensors have very good sensitivity but an average selectivity: it is difficult to correlate the nature of the volatile compounds to the sensor responses. Nevertheless, a comparative study of the curves (Fig-

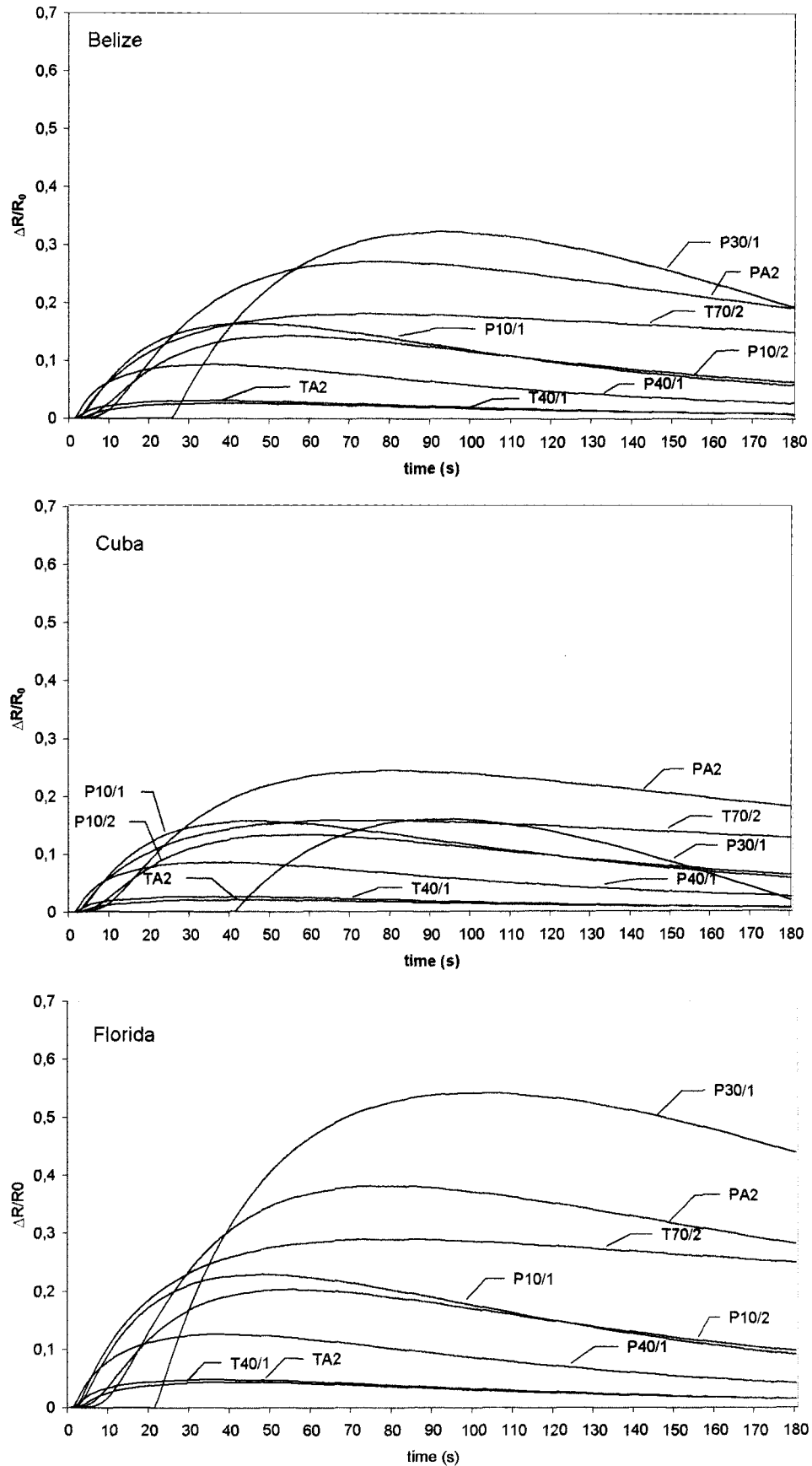


Figure 1. Curves of sensor responses in their original acquisition format versus time: pure Valenica orange juices of tropical and subtropical origins.

ures 1 and 2) shows different kinetics, which can lead one to think there is a real discrimination of the pure juice. The kinetics of the P10/1 and T70/2 sensors never

cross except for the Florida pure orange juices. We note also that the curves for sensors P10/2 and T70/2 do not join for the pure juices of Israel and Spain (Figure 2),

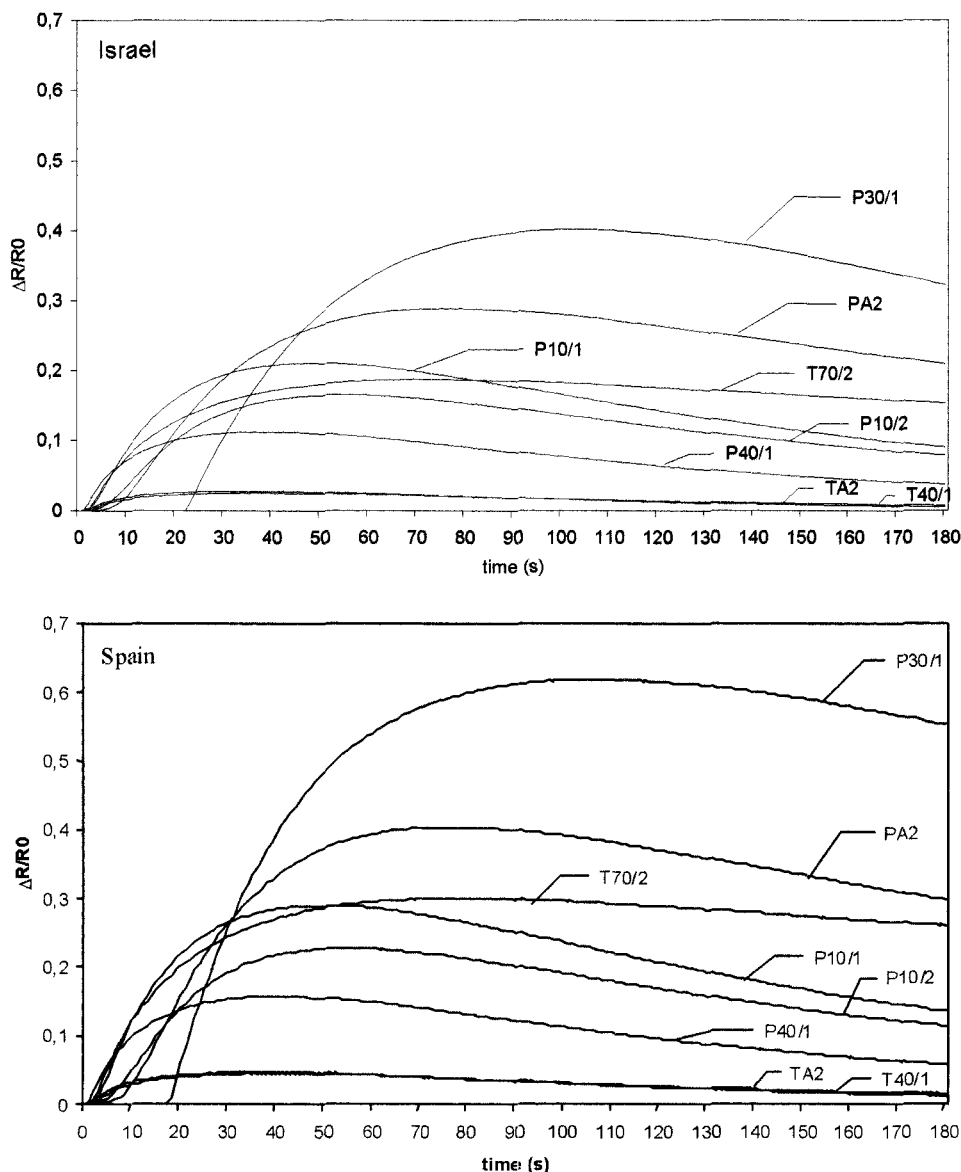


Figure 2. Curves of sensor responses in their original acquisition format versus time: pure Valencia orange juices of Mediterranean origin.

contrary to those coming from Cuba, Belize, and Florida (Figure 1). Intersection occurs in a time interval between 100 and 125 s: We also note here that these qualitative differences can induce characteristics which, by hypothesis, can lead to an ordering of the pure juices as a function of their geographical origins.

Analysis of Variance. Analysis of variance to one factor (ANOVA-1) has been performed to evaluate the power discrimination of the 12 sensors used; for each of them, the mean responses measured have been studied on the five following groups: Cuba, Spain, Israel, Belize, and Florida. Table 7 contains information on the degrees of freedom, sum of squares, and mean squares. The total variance of the dependent variable is partitioned into the explained part (intergroup variance) and the unexplained one, expressed in the residuals (intragroup variance).

Analysis of Table 6 allows us to eliminate some sensors: P70/1, because of a probability value of 0.14; and P40/2, for which the intragroup sum of squares is more important than the intergroup one. The T30/1 sensor has the largest coefficients of variation concerning the studied origins (Table 6): it does not respond

regularly to the volatile compounds generated in the headspace, and we do not take it into account in the results. Furthermore, sensor T30/1 has a large intragroup value, 0.227145, calculated in Table 7.

Principal Component Analysis. After the first process of PCA, we have eliminated sensor P30/2, which gave redundant information, measured by its correlation coefficient, to improve the discrimination of the samples during the second process PCA. Therefore, the PCA of the data was done with the eight remaining sensors. The first two main components PC1 and PC2 give 93.0% of the total expressed variance, with 53.5 and 39.5%, respectively (Figure 3). We consequently hypothesize that descending from the eighth dimension to the graphic one (second dimension), almost all of the information contained in the raw data matrix is preserved and that the graph is representative of the sample set. The matrix of eigenvectors in Table 8 show that the PC1 axis is mainly built by sensors PA2, P40/1, P10/1, and P10/2 and that the axis PC2 is mainly formed by T40/1, T70/2, P30/1, and TA2. The variance expressed on these axes is almost equidistributed (about 53 and 40%), and we thus admit that we have to

Table 7. ANOVA

sensor	variance	$\sum x^2$ ^a	df ^b	MS ^c	F ^d	probability
P10/1	intergroup	0.099920	4	0.024980	94.01571	<0.05
	intragroup	0.009831	37	0.000266		
	total	0.109751	41			
P10/2	intergroup	0.018396	4	0.004599	35.56981	<0.05
	intragroup	0.004784	37	0.000129		
	total	0.023180	41			
P40/1	intergroup	0.024414	4	0.006104	72.76861	<0.05
	intragroup	0.003103	37	0.000084		
	total	0.027517	41			
T70/2	intergroup	0.021330	4	0.005332	14.86961	<0.05
	intragroup	0.013269	37	0.000359		
	total	0.034599	41			
PA2	intergroup	0.017395	4	0.004349	10.81053	<0.05
	intragroup	0.014884	37	0.000402		
	total	0.032279	41			
T40/1	intergroup	0.002761	4	0.000690	47.34899	<0.05
	intragroup	0.000540	37	0.000014		
	total	0.003301	41			
P30/1	intergroup	0.820665	4	0.205166	41.18096	<0.05
	intragroup	0.184337	37	0.004982		
	total	1.005002	41			
TA2	intergroup	0.001643	4	0.000411	43.94914	<0.05
	intragroup	0.000346	37	0.000009		
	total	0.001989	41			
T30/1	intergroup	0.790797	4	0.197699	32.20359	<0.05
	intragroup	0.227145	37	0.006139		
	total	1.017942	41			
P70/1	intergroup	0.000260	4	0.000065	1.83797	0.14
	intragroup	0.001315	37	0.000036		
	total	0.001575	41			
P30/2	intergroup	0.034463	4	0.008616	13.03925	<0.05
	intragroup	0.024448	37	0.000661		
	total	0.058911	41			
P40/2	intergroup	0.008729	4	0.002182	8.49438	<0.05
	intragroup	0.009505	37	0.000257		
	total	0.018234	41			

^a Sum of squares. ^b Degree of freedom. ^c Mean square. ^d Fisher value.

Table 8. Matrix of Eigenvectors: Coefficients of Correlation

sensor	principal components	
	PC1	PC2
P10/1	0.46	-0.16
P10/2	0.47	-0.11
P40/1	0.47	-0.10
T70/2	0.24	0.47
PA2	0.44	0.00
P30/1	0.28	0.41
T40/1	0.03	0.55
TA2	-0.14	0.52
variance ^a (%)	53.5	39.5
cumulative ^b (%)	53.5	93.0

^a Percent of the variance expressed on the two principal component axes. ^b Cumulative variance.

consider the distribution of the samples in the same way according to PC1 and PC2 axes.

PCA allowed us to constitute five groups of samples. We propose to explain this constitution thanks to their relative position to the main components and initial variables. We studied the two following distances: the distance between samples and the distance samples/variables. The five trained groups are representative of

the juice orange origin countries: Belize, Florida, Cuba, Spain, and Israel.

There is at least a value of $\max(\Delta R/R_0)$ that characterizes a juice according to its country of origin: concerning Cuba, the sensor value of P30/1, 0.600 unit, is distinctly below those registered for the other countries (0.890, 0.934, 0.996, and 0.999). For the pure juices of Florida, the value 0.890 of sensor P30/1 differentiates also from the other values. Concerning the pure juices of Spain and Israel, the sensor P10/1 responds more strongly (0.481 and 0.516, respectively). Note that for Israel, the middle value of PA2 is the only one >0.700. These characteristics may therefore explain a regrouping of the samples according to their country of origin thanks to the sensor array. The PC1 axis induced a separation between the juices of Spain and Israel, positively correlated, and Belize, Cuba, and Florida, negatively correlated. The pure juices from Israel and Spain are therefore bound to the sensors PA2, P40/1, P10/1, and P10/2. Concerning juices from Belize, Florida, and Cuba, the P10/1 values measured are near 0.400 (0.399, 0.401, and 0.406), whereas those measured for juices from Spain and Israel are around 0.500 (0.481 and 0.516, respectively). Concerning sensors P40/1, P10/2, and PA2 the raised values are always more elevated for the countries of the Mediterranean area. Although not selective, sensors P10/1 and P10/2 preferentially respond to hydrocarbons compounds; the volatile fraction, including this chemical family, must differ qualitatively and quantitatively according to the oranges' climatic area, Mediterranean or not.

Factorial Discriminant Analysis. FDA allowed us to classify the samples already known according to their origin. The trained groups concern the Mediterranean countries of Spain and Israel as well as the tropical/subtropical countries of Belize, Cuba, and Florida. All of the pure orange juices were classified correctly in their respective group. The whole of the variance expressed by the two factorial axes is 98.3% (Figure 4): this value, almost 100%, indicates that all of the information is conserved and that the graphic representation is an accurate portrayal of the data relative to the pure juice of orange. The expressed variance is not equally distributed on the two first factorial axes: axis 1 carries the maximum of the variance, 87.2%, whereas the FD2 axis carries 11.1% of the variance. Nine-tenths of the information is therefore reached by the FD1 axis: the FD1 axis is mainly formed by the variables PA2, P40/1, P10/1, and P10/2. These sensors permit a very good separation of the groups according to their geographical adherence: The countries of Mediterranean origin, Spain and Israel, are regrouped and negatively correlated with the FD1 axis, whereas those of tropical and subtropical origins are positively correlated. The FD2 axis induced a separation of the pure orange juice country according to the sensors P30/1, T70/2, T40/1, and TA2. Table 9 gives the Mahalanobis distances between the five groups. Sensor P10/1 permits good discrimination of the pure orange juice: its response is ~0.400 for the samples of tropical and subtropical origin, whereas the response reaches 0.500 for those of Mediterranean origin.

The classification of the seven supplementary samples is correct: all were placed in the groups corresponding to their origin. The relaxing pattern about the classification is therefore robust and could be extended to the classification of other supplementary samples.

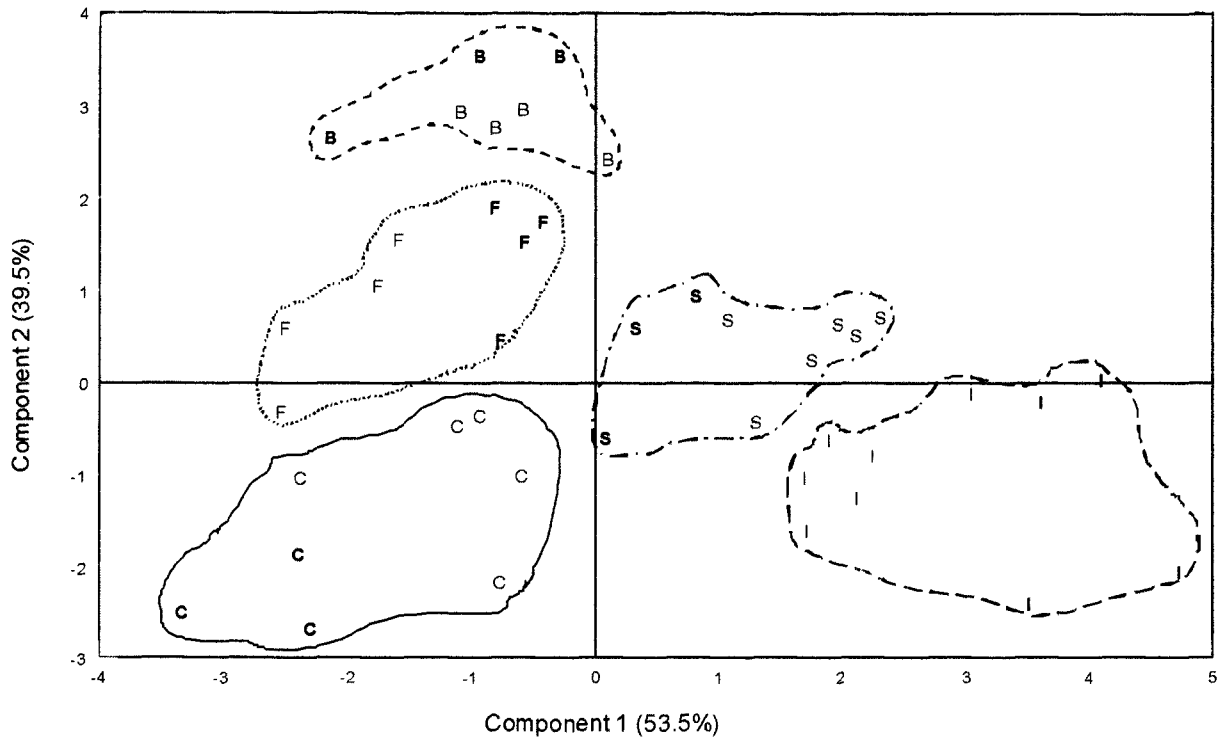


Figure 3. PCA of the data set counting 42 samples of pure Valencia orange juices. B, Belize; F, Florida; S, Spain; C, Cuba; I, Israel. Oranges were harvested in 1997–1998. Boldface letters indicate samples made from oranges harvested in 1996–1997. Variance of the two first principal components: PC1, 53.5%; PC2, 39.5%.

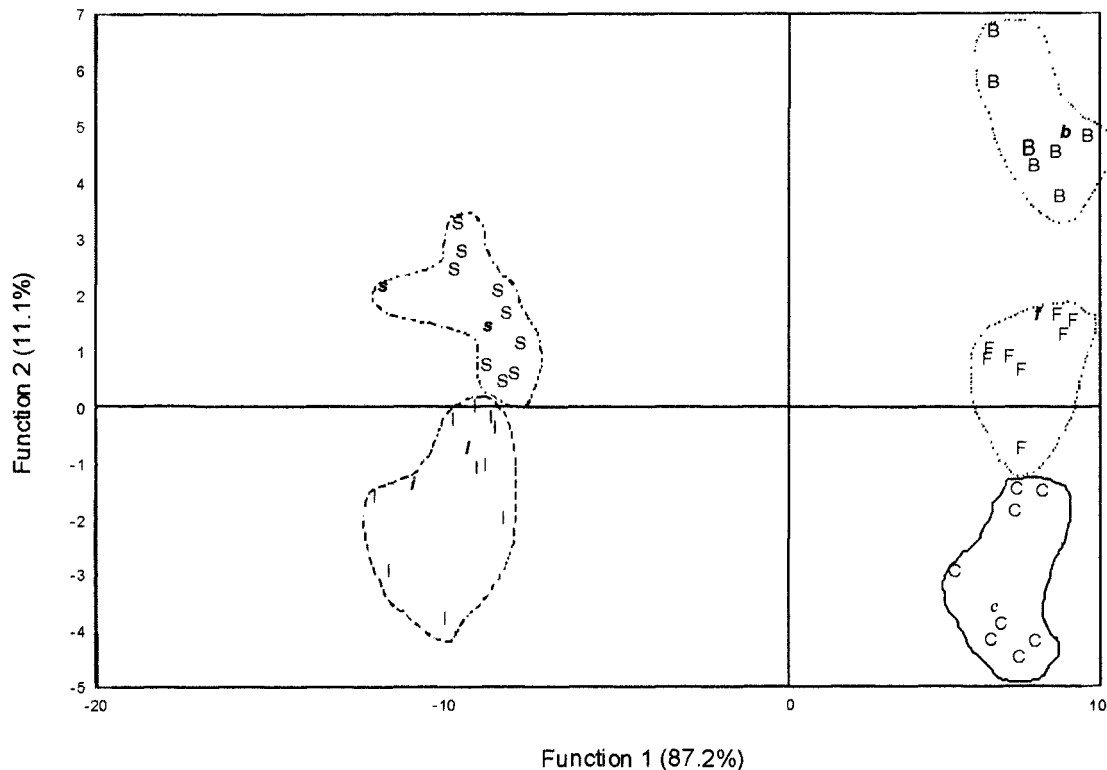


Figure 4. FDA of the data set counting 49 samples of pure Valencia orange juices. B, Belize; F, Florida; S, Spain; C, Cuba; I, Israel. Lower case italic letters indicate the seven supplementary samples from these countries. Variance of the two first factorial components: FD1, 87.2%; FD2, 11.1%.

Calibration. The analytical conditions for the discrimination of pure orange juices from different origins have been determined. Sensor drift is one of the main criticisms in the use of electronic noses at this time. Calibration must be now employed for long-term confidence in results. Each of the electronic noses propose

different ways to calibrate the sensor arrays equipping their apparatus: one of the difficulties in performing the calibration comes from the simultaneous use of different kinds of sensors, which generally drift nonlinearly. Alpha Mos proposes a calibration performed with mathematical algorithms, which compensate automati-

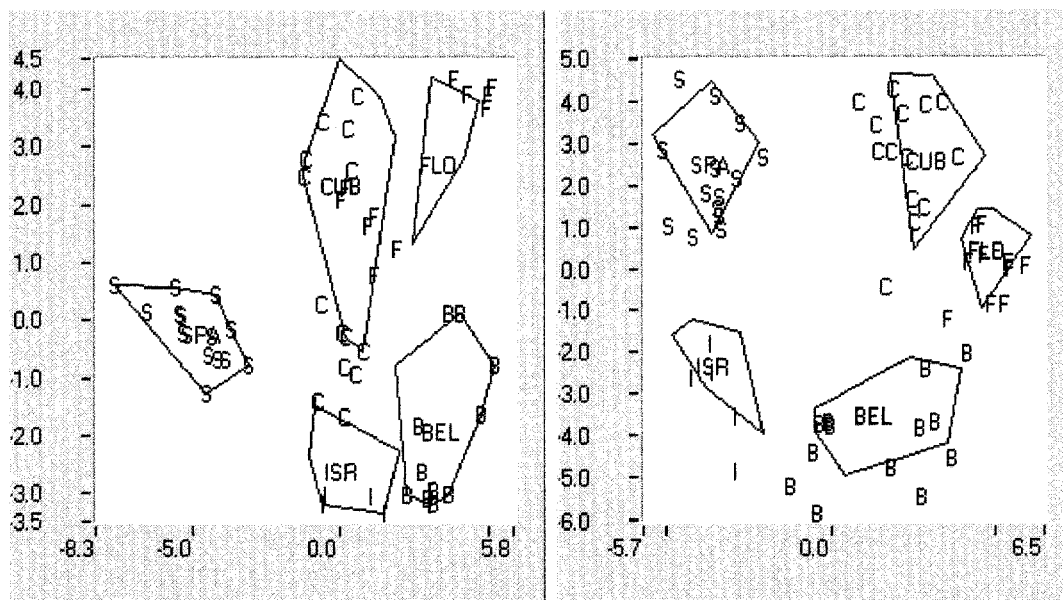


Figure 5. Calibration of pure orange juices from Spain, Israel, Cuba, Florida, and Belize. The left-hand side shows the uncalibrated results. The right-hand side gives the calibrated results from the Fox3000 software calibration process. The clusters represent the five groups of origin: SPA, Spain; ISR, Israel; CUB, Cuba; FLO, Florida; BEL, Belize. Single capital letters are the cross-validating samples clustered around the groups before and after the calibration of the sensors.

Table 9. Mahalanobis Distances

	Cuba	Israel	Spain	Belize	Florida
Cuba	0				
Israel	16.94	0			
Spain	16.72	2.93	0		
Belize	7.94	18.57	17.07	0	
Florida	4.01	17.66	16.76	3.96	0

^a Matrix of Mahalanobis distances between the five groups represented on the FDA diagram (see Figure 4).

Table 10. Supplementary Samples

sample ^a	recognized	group	% of recognition ^b
B	yes	BEL	100.00
I1	yes	ISR	100.00
I2	yes	ISR	100.00
C	yes	CUB	93.44
F	yes	FLO	100.00
S1	yes	SPA	99.09
S2	yes	SPA	98.07

^a Supplementary samples labeled as B for Belize, I1 and I2 for Israel, C for Cuba, F for Florida, and S1 and S2 for Spain. ^b Percent of recognition calculated by Fox3000 software. BEL, ISR, CUB, FLO, and SPA are the names of territories represented on Figures 5 and 6, meaning, respectively, Belize, Israel, Cuba, Florida, and Spain.

cally for the sensor drift. This process is done by the Fox3000 software. Chemical standards are used for this calibration: These compounds are associated with the sample matrix being analyzed. They are generally found among major chromatographic peaks detected in the aroma of orange juices, given by literature references; we have chosen six compounds: guaiacol, α -pinene, furfural, limonene, ethyl butyrate, and γ -butyrolactone. Pure products are used, as calibration standards must react similarly to the samples: the sensor responses get the same response curve with the next speed of response. Analytical conditions have then been set to fit these conditions. To build the calibration model, we have to train the system. Analyses are repeated 10 times: the repeatability of the measurements over time was assessed for a 6-week period. The combined repeat mea-

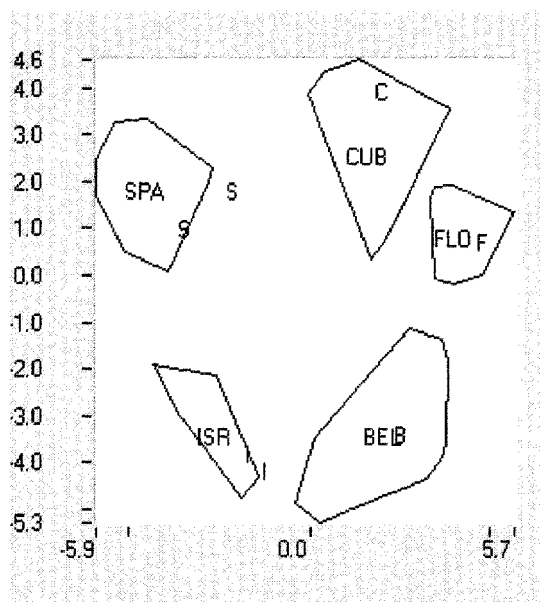


Figure 6. Test of the calibration with the supplementary samples. The territories of the diagram represent the five groups of origin after the calibration of the sensors. Country codes are the same as in Figure 5. Supplementary samples are tested for the validation of the database: two from Israel labeled I, two from Spain labeled S, and one each from Belize (B), Cuba (C), and Florida (F). These samples are positioned around the five groups of countries or state.

surements were used to develop and cross-validate the models used to define different qualitative territories in a discriminant function analysis plot (Figure 5). When 10 analyses of samples and calibration standards have been run, the database and the calibration model are built: The results obtained confirm the efficiency of the calibration model; the right-hand diagram shows the calibration results, where territories are further defined with reference to repeat measurements on the chemical standards. In this diagram, the cross-validating samples were clustered around the correct groups.

The selection of the optimal calibration standard is automatically calculated: limonene and α -pinene have been selected; these standards are then used, 1 week later, to analyze the seven unknown samples that have been applied to the calibration mode. During these blind tests, all of them have been recognized, with a percentage of recognition close to 100% in Table 10; Figure 6.

Conclusion. The use of the sensor array showed a real potential for the discrimination of pure orange juices according to the origins of the oranges involved in the process. The optimization of the headspace generation produced the best conditions to analyze pure orange juices with the electronic nose. There are, therefore, in the generated volatile fraction, compounds that can be used as origin tracers relative to the array sensor responses. Recent results showed that carotenoid profiles of Valencia orange juice by HPLC allow one to differentiate them as a function of the origin of oranges. However, the sensor array permitted a better discrimination of the Belize and Cuba groups. After the calibration of the sensors, we then built a database aiming to classify pure Valencia orange juices from five different origins. The advantage of the electronic nose is the absence of sample preparation, which allows all of the samples' properties to remain intact. Thus, we have showed in this paper the interest of a nonseparative method compared to classical analyses.

ACKNOWLEDGMENT

We thank L. Lapiere, Couecou Society, and M. Bouyer, Fruival Society, for authentic orange juice samples. We gratefully acknowledge the ANRT for providing facilities (Dos. 556/97). We also thank T. O'Neill for assistance in English corrections, M. Sergent for help in the choice of the experimental design, and P. Mouly for sharing his knowledge about the orange juices.

LITERATURE CITED

- Gardner, J. W.; Bartlett, P. N. A brief history of electronic noses. *Sens. Actuators, B, Chem.* **1994**, *18–19*, 211–220.
- Strike, D. J.; Meijerink, M. G. H.; Koudelka-Hep, M. Electronic noses—A mini-review. *Fresenius' J. Anal. Chem.* **1999**, *364*, 499–505.
- Mielle, P. Une technique de pointe au service du controle de la qualite aromatique. *Technoscope Biofutur.* **1998**, *174*, 2–10.
- Aishima, T. Aroma discrimination by pattern recognition analysis of responses from semiconductor gas sensor array. *J. Agric. Food Chem.* **1991**, *39*, 752–756.
- Bartlett, P. N.; Elliott, J. M.; Gardner, J. W. Electronic noses and their application in the food industry. *Food Technol.* **1997**, *51*, 44–48.
- Schaller, E.; Bosset, J. O.; Escher, F. "Electronic Noses" and their application to food. *Lebensm.-Wiss. -Technol.* **1998**, *31*, 305–316.
- Mielle, P. "Electronic Noses"—toward the objective instrumental characterization of food aroma. Special issue on flavour perception. *Trends Food Sci. Technol.* **1996**, *7* (12), 432–438.
- Barnett, D. The electronic nose and food assessment. *Food Aust.* **1999**, *51* (6), 226.
- Bazemore, R.; Rouseff, R.; Sims, C. Comparison of fresh squeezed orange juice analysis by "Electronic nose", sensory taste panel and GC FID. Presented at the Olfaction and Electronic Nose, 3rd International Symposium, Toulouse, France, 1996.
- Bazemore, R.; Rouseff, R. Discrimination of thermally treated orange juices by an electronic nose equipped with organic polymer sensors. *Semin. Food Anal.* **1998**, *3*, 59–66.
- Shaw, P. E.; Moshonas, M. G. Quantification of volatile constituents in orange juice drinks and its use for comparison with pure juices by multivariate analysis. *Lebensm.-Wiss. -Technol.* **1997**, *30*, 497–501.
- Tillotson, J. E. What do consumers want in fruit juice beverage quality? In *Report of Congress IFU*; XII International Congress of Fruit Juice, Interlaken, Switzerland, May 20–24; International Fruchtsaft Union: Paris, France, 1996; pp 57–64.
- Mouly, P.; Gaydou, M.; Lapiere, L.; Corsetti, J. Differentiation of several geographical origins in single-strength Valencia orange juices using quantitative comparison of carotenoid profiles. *J. Agric. Food Chem.* **1999**, *47*, 4038–4045.
- Fournier, J. B.; Martin, Y.-L.; Fourel, F.-P.; Martin, G.-B.; McManus, H. J. D. Characterization of the geographic origin of orange juice from Florida, Brazil and Israel—A combined isotopic and trace element method. *Report of Congress IFU*; XII International Congress of Fruit Juice, Interlaken, Switzerland, May 20–24; International Fruchtsaft Union: Paris, France, 1996; pp 215–223.
- Saunt, J. *Citrus Varieties of the World*; Sinclair International Limited: Norwich, U.K., 1990; pp 15–31.
- Clifford, P. K.; Tuma, D. T. Characteristics of semiconductor gas sensors. I. Steady-state gas response. *Sens. Actuators* **1983**, *3*, 233–255.
- Yamazoe, N.; Fuchigami, J.; Kishikawa, M.; Seiyama, T. Interactions of tin oxide surface with oxygen, water and hydrogen. *Surf. Sci.* **1979**, *86*, 335–344.
- Gardner, J. W. Detection of vapours and odours from a multisensor array using pattern recognition. Part 1. Principal component and cluster analysis. *Sens. Actuators, B, Chem.* **1991**, *4*, 108–116.
- Alpha Mos. *Fox2000–4000 Software Manual 5.01*; Toulouse, France, 1996; pp 78–79.
- Gower, J. C.; Legendre, P. Metric and Euclidean properties of dissimilarity coefficients. *J. Classification* **1986**, *3*, 5–48.
- Shurmer, H. V.; Gardner, J. W. Odour discrimination with an electronic nose. *Sens. Actuators, B, Chem.* **1992**, *8*, 1–11.
- Hibbert, B. Data analysis of multi-sensor arrays. *Electroanalysis* **1998**, *10*, 1077–1080.
- Tomassone, R. *Comment Interpréter les Résultats d'une Analyse Factorielle Discriminante?* INA-PG, ITCF: Paris, France, 1988.
- Roussel, S. *CEMAGREF. Optimisation des mesures: l'influence determinante des paramètres, Nez électroniques et capteurs d'arômes*; ADRIA Formation Interentreprise: Paris, France, 1998.

Received for review December 8, 2000. Revised manuscript received May 2, 2001. Accepted May 7, 2001.

JF0014664