Hyphenated Headspace-Gas Chromatography-Sniffing Technique: Screening of Impact Odorants and Quantitative Aromagram Comparisons

Philippe Pollien, Andreas Ott, Franck Montigon, Marcel Baumgartner, Rafael Muñoz-Box, and Alain Chaintreau*

Nestlé Research Centre, NESTEC Ltd., P.O. Box 44, Vers-chez-les-Blanc, 1000 Lausanne 26, Switzerland

A new headspace-GC-sniffing method is proposed. Using a recently developed headspace cell, the vapor phase is collected under conditions that mimic well those of an aroma above a food. Data treatment is based on detection frequency, rather than on perceived intensity or successive dilutions as used in other approaches. Repeatability appears satisfactory, and independent panels are even able to generate similar aromagrams, without training prior to the analysis. Using a minimum of six assessors, this technique seems to be more reliable than classical ones. To compare detection frequencies between two aromagrams, an estimation of the least significant difference is given. A theoretical justification of this approach is suggested, on the basis of determination of detection thresholds.

Keywords: *Headspace-GC-sniffing; impact odorants; detection frequency*

INTRODUCTION

In the past decades, many detection techniques have been hyphenated to gas chromatography. Less attention has been paid to GC-olfactometry (GC-O) in which the human nose plays the role of the detector. However, the human nose is often more sensitive than any physical detector, and GC-O exhibits powerful capabilities that can be applied to flavors and perfumes, as well as to any odoriferous product (e.g. pollutant).

Olfactometric (or "sniffing") techniques allow the determination of impact odorants in foods. They can be classified into two categories: *dilution methods*, which are based on successive dilutions of an aroma extract until no odor is perceived at the sniffing port of the chromatograph; and *intensity methods*, in which the aroma extract is only injected once but the smeller records the odor intensity as a function of time by moving the cursor of a variable resistor.

Dilution methods are the most often cited in the literature. Two variants are commonly applied: CHARM analysis (Acree et al., 1984) and aroma extract dilution analysis (AEDA) (Ullrich and Grosch, 1987). Both have proven their efficiency for screening of impact odor contributors of an aroma (Acree, 1993; Grosch, 1994; Blank, 1996). This qualitative aspect is now well established, and most recent publications show an increased interest for a more quantitative approach: can we compare two aromas by comparing the odor intensities of their aromagrams (Acree and Barnard, 1994; Guichard et al., 1995)? To achieve this objective, two questions must be answered: (1) What is the most repeatable/reproducible sampling method, representative of the aroma which is really perceived by the nose? (2) How can olfactometric data be treated to overcome the lack of repeatability and reproducibility inherent in a sensorial technique?

Since GC's capabilities are well-known, its suitability for quantitative analysis is beyond question.

* Author to whom correspondence should be addressed (e-mail alain.chaintreau@chlsnr.nestrd.ch).

Sampling. Used conventionally, all sniffing procedures are based on injection of an extract of food. Its preparation generally requires a steam distillation (e.g. SDE) or a direct solvent extraction. If quantitative recoveries are assumed, these two alternatives yield a solution which is more representative of the flavor composition in the matrix than that in the vapor phase surrounding the food. Since only headspace flavorings are perceived by the nose, an aromagram based on an extract will not quantitatively represent the odor profile of the product (Abbott et al., 1993b). In addition, if the extract must be concentrated prior to analysis, low-boiling volatiles will be lost. Also, the solvent required by successive dilutions will mask the first eluting peaks.

Data Treatment. Among dilution methods, AEDA does not require any sophisticated equipment. This simplicity also limits its applicability, since no recordable signal is generated during elution. Therefore, data cannot be computed into a continuous function of time like CHARM. Grosch's quantitative approach, based on odor activity values (OAV = concentration/threshold), has been criticized (Abbott et al., 1993a). It assumes the perceived intensity to be proportional to aroma concentration, instead of fitting Stevens' or Fechner's law (Stevens, 1961; Baird and Noma, 1978), which are currently accepted as the best representations of sensorial perceptions:

$$\text{Stevens:} \quad I = k(C - C_t)^n$$

$$\text{Fechner:} \quad I = k' \ln(C/C_t)$$

I is the aroma intensity, C is the aroma concentration, C_t is the perception threshold, and k', k, and n are constants relative to a given compound.

Stevens' law seems to be verified in the case of GC-sniffing experiments with direct recording of perceived intensity (da Silva et al., 1994).

In CHARM a signal is generated and recorded during the whole chromatographic run, allowing computer calculations. A critical evaluation of these two dilution methods has revealed reproducibility problems due to

S0021-8561(96)00885-0 CCC: $14.00

perception differences within and between individuals (Abbott et al., 1993a). In addition, a compound can be undetected at a given dilution, although it is smelled again at a higher dilution level. In other terms, specific anosmia, inattention, or simply the noncontinuous breathing process (no odor perception while breathing out) alters the perception.

CHARM analysis has proven its reliability to screen impact odorants. Its quantitative use requires replication of the sniffing runs by at least three different trained panelists (Guichard et al., 1995). After dilutions are performed, this represents >30 injections!

The OSME method was proposed (McDaniel et al., 1990) to obtain intensity information in a single run. A good reproducibility of peak intensities was claimed by da Silva et al. (1994), which seems in conflict with their 1990 paper: the 1994 work was based on a model mixture of 6 pure components eluting at about 50 index units from each other, whereas its application to a real system (Pinot noir wine) showed high within and between individual fluctuations, reflecting "day-to-day variations in sensitivity" (McDaniel et al., 1990). Recently, OSME was further improved (Guichard et al., 1995), but gaps mentioned in dilution methods also occur with intensity techniques. These authors observed large variations of peak area for a given sniffer: up to a 108% coefficient of variation was noticed for two replicates. They concluded that there was a very high variability within and between panelists for intensity evaluations. Unfortunately, they did not test the repeatability of the panel after summing individual responses, and no application to a real food was mentioned. In this last case, an odor description is often required to help peak identification. It will be very difficult for a panelist to simultaneously detect an odor, find a descriptor, and register an intensity from a previously memorized scale, especially as peaks may elute rapidly and close to each other from a capillary column.

From all of these observations, our objective was to design a method with the following characteristics: (1) no training of the panel, therefore no intensity measurement, since it requires learning a scale; (2) acceptable compromise between good repeatability/reproducibility and a restricted number of injections.

This paper describes a new headspace-GC-olfactometry method for which preliminary results were previously presented (Ott et al., 1996). It associates a headspace cell with an automated thermal desorber, itself coupled to a gas chromatograph and a sniffing port. Olfactometric signals are computerized using a detection frequency approach.

MATERIALS AND METHODS

Flavor Cocktail. A flavor cocktail diluted in water was prepared according to the composition described in Table 1. Isobutyl and butyl acetates occurred in isoamyl acetate as impurities.

The first solution of the simplified flavor cocktail only contained isoamyl acetate (0.148 mg/kg) and 2(*E*)-hexenal (0.143 mg/kg). In solutions 2, 3, and 4, these concentrations were multiplied by 2, 4, and 8, respectively.

Animal Dejecta. As an illustration of the method applicability to malodorants, dejecta of an author's pet were collected, homogenized, and stored at 4 °C until the GC-O analysis. Identification of impact odorants was not performed.

Headspace Sampling (Chaintreau et al., 1995). The odoriferous material (0.5 g of flavor cocktail solution or 9 g of animal dejecta) was placed into the sample space of the headspace cell and equilibrated at 30 °C in a water bath for 0.5 or 2 h for the cocktail solution and the dejecta, respectively. The headspace of the cell (30 or 240 mL for the flavor cocktail or the feces, respectively) was then passed through the trap containing 250 mg of Tenax at a flow rate of 40 mL/min.

To maintain a carrier flow compatible with the internal volume of the trap, without splitting it, a wide-bore column was chosen. Volatiles were thermally desorbed from the Tenax using an ATD400 thermal desorber (Perkin-Elmer Corp., Norwalk, CT) at 250 °C for 10 min. They were refocused in the instrument's internal Tenax cold trap (−30 °C) and desorbed (260 °C, 3 min) into an HP 5890 GC (Hewlett-Packard, Avondale, PA) equipped with a DB-Wax column (J&W Scientific, Folsom, CA; 60 m length, 0.53 mm i.d., 1.00 μm phase thickness). Helium was used as carrier gas at 8 mL/min. The oven was kept at 20 °C for 5 min, and then the temperature was increased to 200 °C at 4 °C/min. The final temperature was maintained for 10 min. For the simplified cocktail, the column was kept for 2 min at 60 °C and increased at 4 °C/min up to 150 °C, and this temperature was maintained for 2 min. The column oulet was either connected to an FID or to a sniffing port (Brechbühler SA, Switzerland) equipped with a humidified air makeup.

After each sampling, the cell was cleaned in a vacuum oven at 50 °C under 10 kPa for at least 1 h. Tenax sampling tubes were cleaned before use by heating for 1 h at 300 °C under a nitrogen flow.

Variability of the Headspace/GC System. The stability was evaluated by sampling the headspace of a 4 ppm standard solution of 2-butanol according to the procedure described above. GC peak areas were measured over 152 days using an FID detector (Figure 1).

Sniffing Procedure. In a typical experiment, six people were randomly selected among available colleagues. Sniffing of the chromatogram was divided into two parts of about 25 min. Each person participated in the sniffing of both parts, but during two distinct sessions to avoid lassitude. Elution of aroma relevant flavorings was recorded by pressing a button during the whole sensory impression. The square signal was registered by an HP Pascal workstation. When peak recognitions were needed, the assessor recorded the corresponding odor descriptors on a tape recorder.

Table 1. Aromagram Repeatability and Reproducibility (Six Panelists, Flavor Cocktail)

compd no.	compd name	intrapanel SD (B, series 1 and 2) NIF (%)	intrapanel SD (B, series 1 and 2) SNIF[a]	interpanel SD (A, mean of series 1 and 2 for B) NIF (%)	interpanel SD (A, mean of series 1 and 2 for B) SNIF[a]	cocktail compos (mg/L)
1	acetaldehyde	11.0	717	17.3	1081	8.6
2	ethyl acetate	11.7	562	5.8	1188	8.8
3	isobutyl acetate	0.0	371	11.8	472	?
4	butyl acetate	11.8	646	5.9	619	?
5	isoamyl acetate	12.5	1147	6.3	864	2.1
6	*trans*-2-hexenal	11.0	566	6.2	346	1.3
7	*cis*-3-hexenol	0.7	665	12.1	227	0.43
8	linalool	11.0	54	6.2	314	3.72
9	phenylethyl acetate	0.0	484	11.7	589	43.1
	av SD	9.4	641	11.8	811	
	av RSD (%)	14.1	13.8	15.7	14.8	

[a] See Materials and Methods for SNIF units.

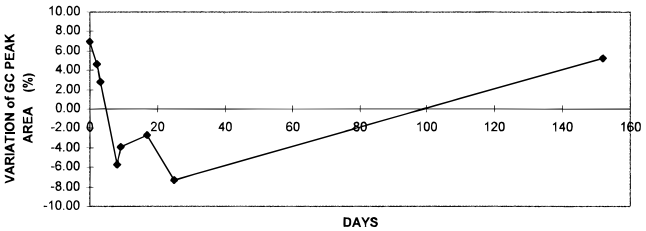

Figure 1. 1. Stability of GC signal over 152 days using a 4 ppm standard solution of 2-butanol sampled with the headspace cell.

The six individual aromagrams of a given sample were summed to one chromatogram and normalized with home-made software, yielding an averaged aromagram. Peak heights and areas are called NIF and SNIF, respectively (1000 SNIF units correspond to 100% NIF, over a duration of 1 s; nif means an unpleasant odor).

Linear retention indices were calculated (Van den Dool and Kratz, 1963) after injection under the same condition of an *n*-alkane series (C_5-C_{25}).

Computations. Variances between two values of the same peak (two repetitions or two panels) were computed as $d^2/2$ (d is the difference between the two values) and averaged over all the peaks. The square root of this average was considered as the average standard deviation (SD):

$$SD = \sqrt{\frac{1}{n}\sum_{i=1}^{i=m}\left(\frac{d^2_i}{d}\right)}$$

The parametric equations were adjusted to the experimental data using the curve fitting software TableCurve 2D for Windows (Jandel Scientific, Erkrath, Germany).

RESULTS AND DISCUSSION

Sampling. To overcome difficulties mentioned in the Introduction, when using an extract, Grosch's group proposed to sample the headspace surrounding the food (Guth and Grosch, 1993; Semmelroch and Grosch, 1995). In addition, its composition better represents the smell that is really perceived by the consumer. However, this group used a gas syringe: such a method does not allow injection of a large volume, which should be refocused before elution. On the other hand, sampling a large volume in a closed vessel would create a depression and modify the equilibrium between both phases. Moisture in the headspace should also be limited to ensure the best possible GC column performance.

Using a purge-and-trap technique would partially solve these drawbacks: the aroma is concentrated in the adsorbant and moisture is not retained on polymers, such as Tenax. However, stripping the liquid sample or flushing the sample surface with a gas leads to different recoveries of the flavorings following the individual solvent-to-matrix partition coefficients: the most volatile components will be more enriched than the others and the composition will not be representative of the gas phase at equilibrium as it is perceived by the nose.

To overcome this difficulty, a technique combining static and dynamic headspace was chosen, using a new headspace cell developed for partition coefficient measurements (Chaintreau et al., 1995). In this device, volatiles are allowed to equilibrate between the matrix and the headspace chamber. When pressing the piston of the headspace chamber, the volatiles are evacuated through a Tenax trap, without disturbing the equilibrium between the phases. These conditions are close to those existing in a food packaging, and large volumes can be collected. Air and moisture are not retained in the absorbent. Reproducibility of cell sampling was tested over a longer period of time (Figure 1) than we previously reported. Starting from a standard solution of 2-butanol, the headspace was regularly trapped after equilibration in the cell. The standard deviation of the GC peak area was <6%. As this value included variability of the FID detector (hydrogen and air flow fluctuations), the real reproducibility was even better.

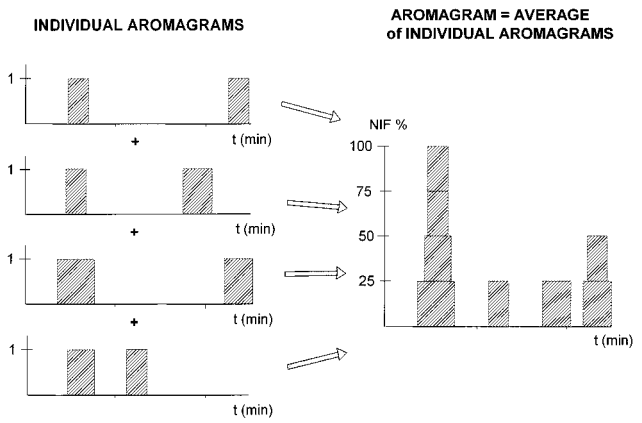

Figure 2. Data treatment procedure (example with four sniffing replications).

Data Treatment of Sniffing Signals. Due to the literature criticisms of dilution methods previously mentioned, and since the smallest peaks represent low contributions, only one concentration level was used, allowing detection of only the most intense odorants. The volume of the sampled headspace was chosen to provide us with a reasonable intensity (<30 odorants in the aromagram). Typically, the sniffing run was replicated by six to eight assessors under the same conditions, and individual aromagrams were averaged by the computer (Figure 2).

Such a treatment offers the advantage of "smoothing" differences within or between individuals, since each panelist only participates in $1/n$ of the final result, n being the number of panelists. "Gaps of the coincidence responses" (Abbott et al., 1993a) are thus considered as a normal phenomenon due to the probability of an odorant perception at a given retention. Consequently, peak intensities are not related to flavoring intensities, but to their detection frequencies. In this paper, peak heights and peak areas will be called NIF and SNIF (nasal impact frequency and surface of nasal impact frequency, respectively).

Under such conditions, a NIF of 100% means that the odorant was detected by all n panelists. In other terms, its concentration was above the odor threshold for everybody. A smaller peak corresponds to a flavoring that was below the detection threshold for one or more of the panelists. The smallest height ($NIF = 100/n$) indicated an odor found by only one panelist and occurring randomly ("odor noise"), due either to external odorants or to contributors below the detection threshold for all the other panelists.

In addition, and after normalization of the peak height scale to 100%, NIF and SNIF must be independent of the number of sniffers, within the confidence interval limits.

On the basis of the statistical concept of NIF and SNIF values, would it be possible to compare peaks of two aromagrams? To answer this question, repeatability and reproducibility tests were performed, as was an evaluation of the influence of increasing aroma concentration on the final profile.

Aromagram Repeatability and Reproducibility. *Simplified Flavor Cocktail.* At first, the reproducibility of the sniffing procedure was evaluated using a model mixture. Three sniffing runs were organized, using two different panels of six people, without training prior to the experiment: panel A, one series of sniffing; panel B, two series of sniffing.

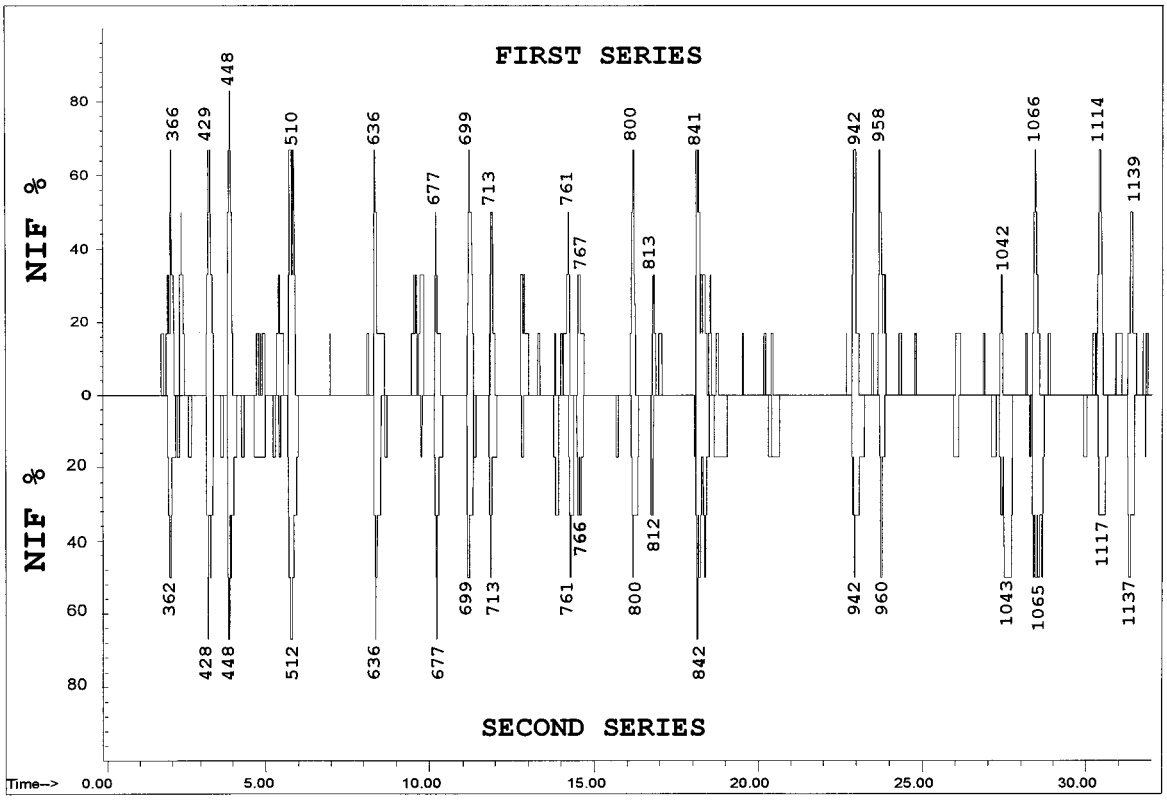

Figure 3. Repeatability of NIF profiles from animal dejecta (same panel, six judges).

Table 2. Reproducibility of Peak Intensities after Guichard et al. (1995) (D_1 = 100 mg/kg; D_2 = 50 mg/kg)

		panel A, panelists 1−5			panel B, panelists 6−10			comparison A/B		
		av	variance	SD (%)	av	variance	SD (%)	av	variance	SD (%)
thiophene	D_1	291.0	89888	103.0	114.0			202.5	15665	61.8
	D_2	130.0	613	15.3	73.0			101.5	1625	39.7
ethyl butyrate	D_1	751.4	110562	44.3	613.8	234237	78.8	682.6	9467	14.3
	D_2	500.2	16839	25.9	382.4	71093	69.7	441.3	6938	18.9
2,6-dimethylpyrazine	D_1	251.3	87894	118.0	186.7	29940	92.7	219.0	2086	20.9
	D_2	193.3	39823	103.3	172.5	4513	38.9	182.9	215	8.0
isovaleric acid	D_1	648.6	142510	58.2	407.6	99116	77.2	528.1	29040	32.3
	D_2	379.2	57832	63.4	284.8	48769	77.5	332.0	4456	20.1
trimethylpyrazine	D_1	467.0	154112	84.1	324.4	57740	74.1	395.7	10167	25.5
	D_2	372.6	220969	126.2	186.8	17363	70.5	279.7	17261	47.0
acetophenone	D_1	211.8	25637	75.6	129.4	13966	91.3	170.6	3391	34.1
	D_2	227.0	44359	92.8	97.3	3495	60.8	162.1	8418	56.6
av		368.6	82586		247.7	58023			9061	
SD			287.4			240.9		av RSD (%)		30.9

NIF values generally did not differ more than 16.6 units, corresponding to one panelist (100%/6 = 16.6%) (Table 1). For an easier comparison between heights and areas, standard deviations were calculated (Table 1). There was no evidence that NIF was more reproducible than SNIF to compare peaks or vice-versa. This was confirmed by the average relative standard deviations that did not significantly differ between NIF and SNIF.

More surprisingly, interpanel reproducibility appeared to be comparable to intrapanel repeatability, although no training of the panelists was required, contrary to intensity methods (da Silva, 1994). In other terms, **two independent panels were able to generate similar aromagrams** of a given product. To our knowledge, this has never been reported before.

As a comparison, Guichard et al. (1995) recently published detailed results obtained with 10 sniffers using an intensity method. From their individual values, we grouped assessors into two panels of five judges. Similarly to our experiments, standard deviations were calculated (Table 2). The average standard deviation (30.9%) seems to indicate a lower reproducibility between panels compared to NIF or SNIF. However, it must be taken into account that both types of information are different. In addition, the number of panelists (five) in each group of Table 2 was lower than in our case, but each intensity value already represented the average of two intensity evaluations.

Acree uses the logarithm of CHARM peak areas to achieve quantitative comparisons of aromagrams (Acree and Barnard, 1994). A CHARM peak is made of several fractions corresponding to discrete values of the dilution levels. Therefore, the peak area itself is not a continuous function of the odorant's concentration. As logarithms should not transform a scale consisting of only discrete values, the significance of CHARM area logarithms is difficult to relate to a physical meaning (odorant concentration or perceived odor intensity). The theoretical justification of NIFs will be given under Method Justification and Aromagram Comparison.

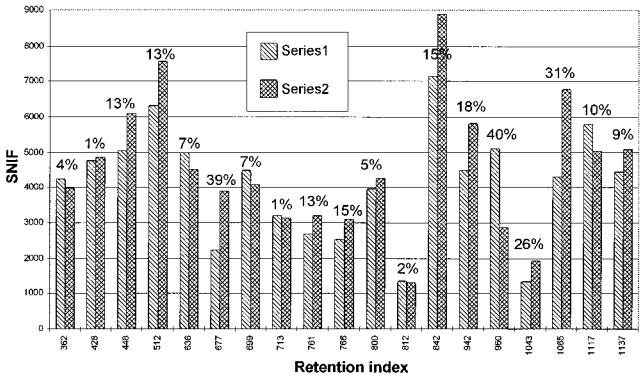

Figure 4. Repeatability of SNIFs from animal dejecta (same panel, RSD at the peak apex).

Animal Dejecta. As a more real case, and to illustrate the applicability of the method to malodorants, animal dejecta were analyzed by GC-olfactometry. Figure 3 shows two replicates of the same feces pool, for which their headspace was collected, desorbed into the GC, and smelled according to the proposed technique. For each peak, NIF standard deviations between series were <20%, except for the unresolved peak at index 1114−1117. The average relative standard deviation (RSD) of NIFs between aromagrams was 17%, in agreement with that found using the standard mixture. Neither profile differed by more than 1 panelist perception. In addition, peaks eluting with an index difference of 5−6 (e.g. at indices 761 and 766−767) were baseline resolved. Mean peakwidths in the dejecta aromagram were 0.11 and 0.14 min for series 1 and 2, respectively. This corresponded to 3−4 index units. These performances seem acceptable compared to those obtained with the same system connected to an FID (peakwidth = 0.07 min).

The average RSD of SNIFs (Figure 4) was 18%, indicating a similar repeatability when using peak heights or areas. Considering the very complex matrix of this "digest", which contains a great number of odorants, these performances were satisfactory using a restricted number of only six panelists.

Necessary Number of Assessors. To determine the necessary number of panelists required to establish an aromagram, a model mixture of two compounds [isoamyl acetate and 2(*E*)-hexenal] was sniffed by 21 assessors. Results from these 21 sniffers were randomly ordered to simulate 200 panels. Average aromagrams were then reconstructed (Figure 2) using 2, 3, 4, . . . assessors, for each generated "panel". For a given panel, the NIF (or SNIF) variation for a given peak was calculated by subtracting the NIF of the peak obtained with $n - 1$ sniffers from that obtained with n sniffers. Repetition of this calculation for the 20 differences ($n = 21$) in each of the 200 panels is represented in Figure 5. The maximum difference was found to vary from ±50% for 2 panelists to <±5% for 21 panelists. Curves were similar for other concentrations, for both peaks of the model mixture and for SNIF values.

These results show that working with one or two panelists, as usually observed in the literature, cannot lead to a reliable profile. Due to the drawbacks already mentioned in the Introduction, the first assessor may detect or miss a peak which is above his own detection threshold. Even when the second assessor correctly detects the same peak, a 50% difference from the asymptotic result can be observed. Figure 5 indicates that a minimum of ±20% must be expected if the panel consists of fewer than 6 people. A reasonable compromise should be 8−10 panelists.

Method Justification and Aromagram Comparisons. Although NIF profiles are reproducible, this does not mean that they are suitable for distinguishing between two flavor mixtures which differ only by their ingredient concentrations. To test the method's ability to evaluate quantitative differences, the previous model mixture of isoamyl acetate and 2(*E*)-hexenal was used to compare NIF of peaks at four different concentrations.

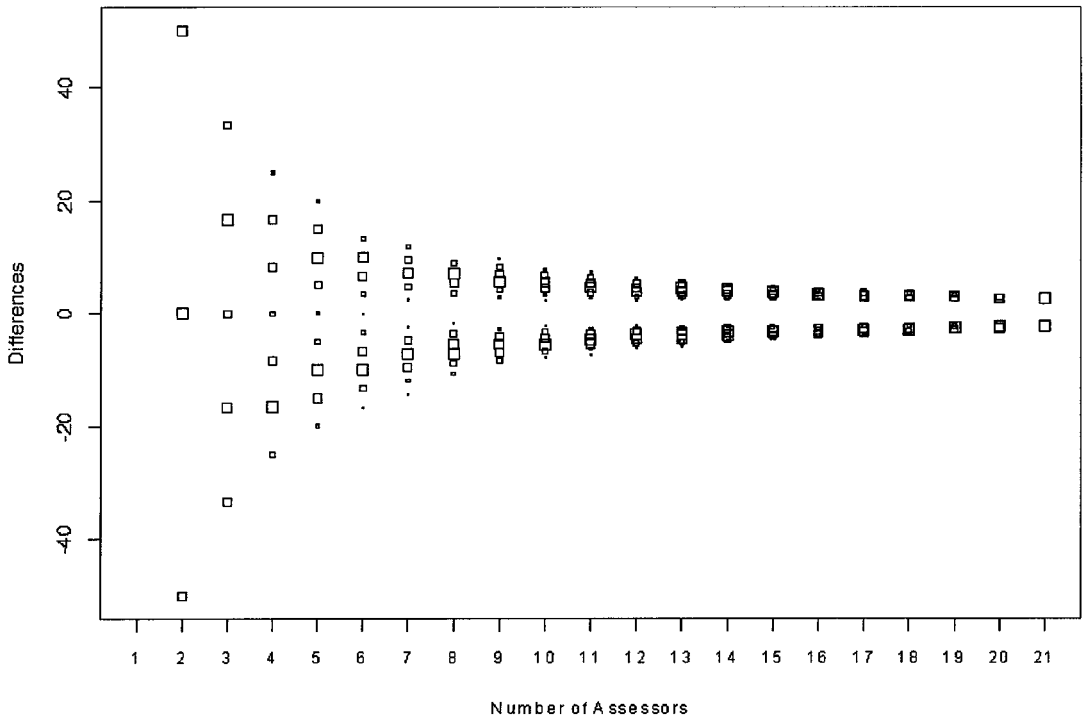

Figure 5. NIF differences between isoamyl acetate peaks reconstructed with $n - 1$ and n assessors (lowest concentration; areas of the square are proportional to the frequency of this difference).

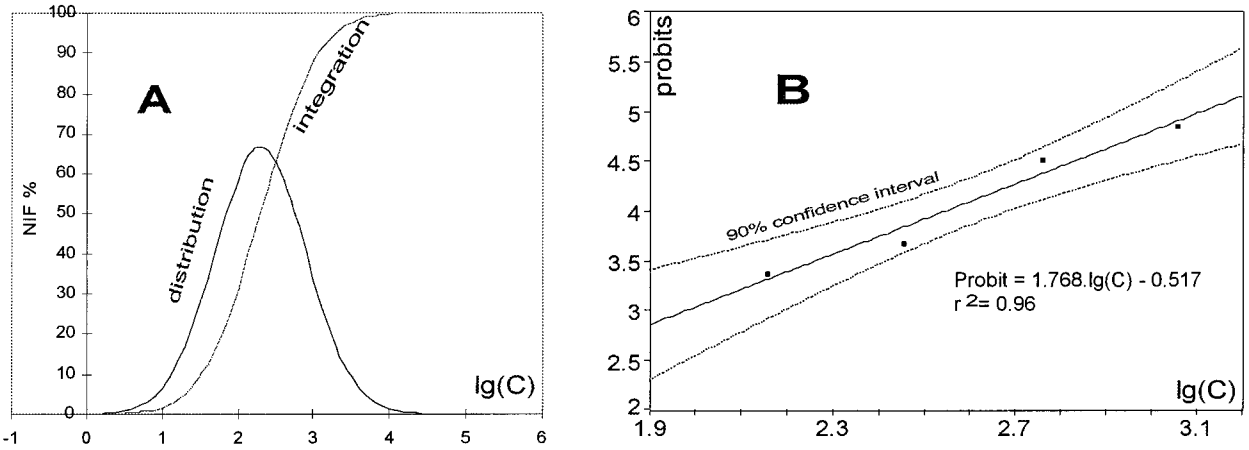

Figure 6. Unimodal distribution of a population detecting an odorant (21 panelists): (A) theoretical curve; (B) curve of 2(*E*)-hexenal after Bliss's linearization of NIF values with confidence intervals of the mean.

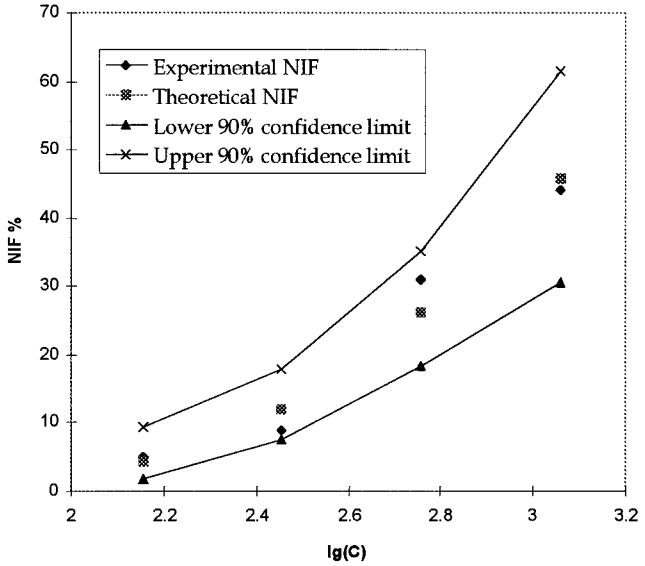

Figure 7. NIF 90% confidence interval of 2(*E*)-hexenal, with 21 assessors (concentrations are increased by a factor of 2 from one sample to the following sample).

As the present method is based on the detection limit of most potent odorants, this can be related to the measurement of flavorings' odor thresholds by headspace sniffing in glasses (Voirol and Daget, 1986; Ott et al., 1997). Assuming that the detection threshold of a panel is distributed according to a log-normal distribution as a function of concentration implies that, after integration, the total number of panelists above the threshold will be represented by a sigmoidal curve (Figure 6A).

On the basis of this assumption, NIFs of four different concentrations of 2(*E*)-hexenal, using 21 assessors, were transformed into probits according to the procedure of Bliss (1967). As expected, r^2 and the residuals indicated a better linear relationship between probits and lg(*C*) than between probits and *C*. The probit curve exhibited a low slope (1.768), which seems to indicate a moderate variation of NIFs when the concentration increases.

From the regression shown in Figure 6, 90% confidence intervals were calculated and converted back into NIFs confidence intervals (Figure 7). Differences between experimental and predicted NIF were equivalent to one assessor or less.

The variation of odor intensities of a given product at two different concentrations differing by a factor of 2 was not fully significant at the 90% confidence, as

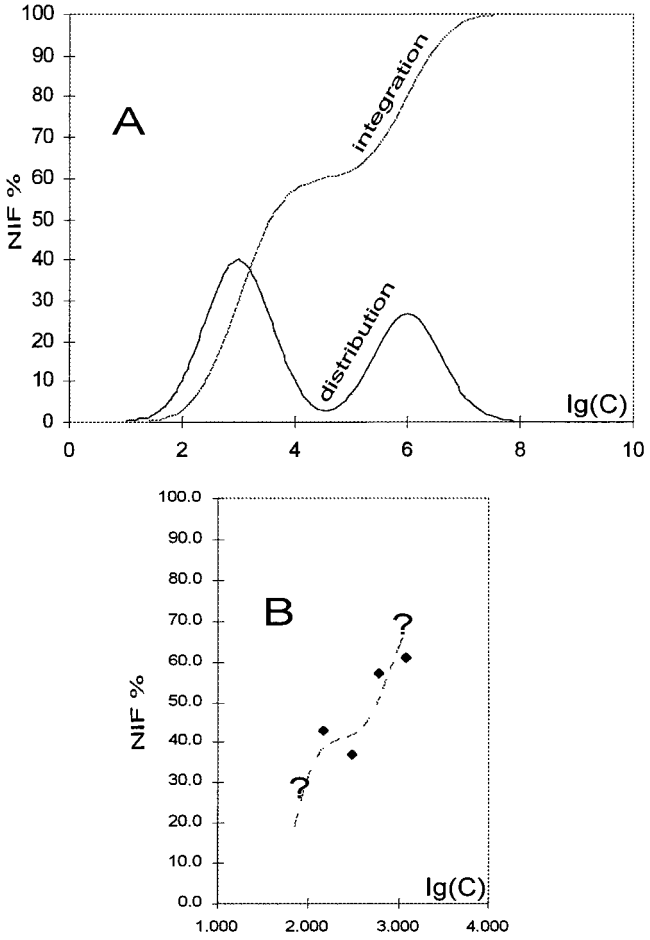

Figure 8. Bimodal distribution of a population detecting an odorant: (A) theoretical curve for a bimodal distribution; (B) experimental data for isoamyl acetate.

some overlap of the confidence intervals occurred (Figure 7). On the other hand, a factor of 4 led to clearly separated NIF confidence intervals.

Applying the same computation to isoamyl acetate led to a lower correlation coefficient between probits and lg(*C*) ($r^2 = 0.70$). This could be attributed either to a larger peakwidth of the panelist distribution around the threshold leading to a lower confidence with 21 panelists or to a bimodal distribution. In other terms, panelists could be divided into two groups with two distinct detection thresholds. This possibility has already been documented for taste perception: the threshold of phenylthiocarbamide exhibits a bimodal detection thresh-

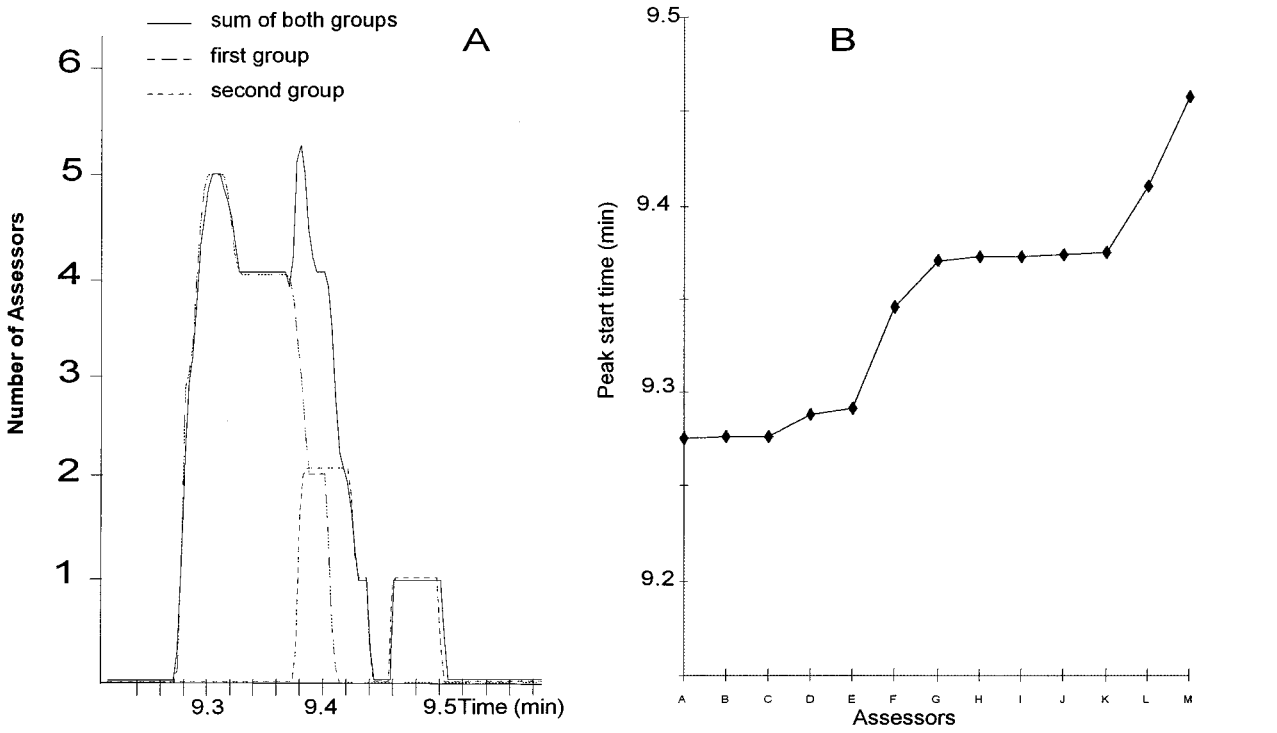

Figure 9. Peak shape of isoamyl acetate reconstructed with panelists detecting either exclusively in group 1 or in group 2 (A) and start time of the panelist's detection for the same peak (B).

old (Kalmus, 1971). As this phenomenon could be due to a compound coeluting with isoamyl acetate, the peak purity was checked by mass spectrometry. When all ion chromatograms were reconstructed, they exhibited an apex at the same retention time, indicating that the presence of an impurity was unlikely to be responsible for the bimodality.

Under these conditions, the NIF curve should not be a sigmoid, but a "double sigmoid" as shown in Figure 8. Due to the limited number of concentrations used, it was not possible to ascertain this hypothesis by fitting the curve to the experimental values. However, some elements seem to support this idea. If we consider the reconstructed sniffing peak of isoamyl acetate, for the three lowest concentrations, it appears to be fractionnated into two subpeaks which partially overlap (Figure 9A). This suggests that a fraction of the panelists detect this stimulus earlier than the others. Classifying panelists according to the start time detection of this peak clearly showed two distinct steps (Figure 9B). Early "clickers" perceived the odor at a retention time corresponding to the beginning of the FID peak, indicating a lower concentration in the carrier gas than for those detecting the peak later.

The higher intensity of the first subpeak (Figure 9A) seems in agreement with a better sensitivity of the early clickers compared to late clickers. Consequently, a bimodal distribution could be seriously suspected. More experienced sniffers were not more numerous in the first group than in the second one, and nobody clicked two times in the same GC-O run.

As a consequence of the panel's behavior with these two flavorings, it can be demonstrated that peak height (= NIF) increases as a function of concentration. With a unimodal distribution, this increase will be continuous. In case of a multimodal distribution, an inflection or plateau may be observed. Therefore, it is not possible to predict whether two concentrations, such as C1 < C2, will lead to increasing NIF or SNIF intensities, without knowing the distribution of panelists for these given compounds. On the other hand, intensities such as NIF1 < NIF2 mean that concentrations will be in the same order, i.e. C1 < C2.

Table 3. Estimation of Least Significant Differences (lsd) between NIFs or SNIFs of the Same Peak at Two Different Concentrations by a Panel of Six Assessors

	lsd of the cocktail (%)					
	within panel A		between panels A and B		lsd of animal dejecta (%)	
confidence level (%)	NIF	SNIF[a]	NIF	SNIF[a]	NIF	SNIF[a]
95	31	2085	33	2639	35	2351
90	25	1687	26	2134	29	1937

[a] See Materials and Methods for SNIF units.

As the distribution of panelist's perceptions cannot be known for all flavorings, we propose evaluating a least significant difference (lsd) between NIF or SNIF values for a given peak found in two different samples. This uses the average standard deviation for six assessors given in Table 1 and Figures 3 and 4 (about 18.5%). The lsd can be calculated from the standard deviation (SD) and Student's constant (t), which takes into account the number of degrees of freedom (9 for the cocktail, 19 for the dejecta):

$$\text{lsd} = t(\sqrt{2})\text{SD}$$

An estimation of its value is given in Table 3, on the basis of the results of the aromagram established with the flavor cocktail and animal dejecta.

Obviously, increasing the panel size to eight members, as previously proposed, will decrease these lsd's. Comparing aromagrams made by two different panels does not seem to significantly increase these values, but no repetition was made for panel B.

Practically speaking, it may be assumed that working with the same panel of eight assessors, a NIF difference of 30%, or a SNIF difference of 2000, will generally indicate a significant concentration difference.

CONCLUSIONS

Without repeating the sniffing at several dilution levels, as in usual methods, most potent contributors to an odor can be determined using NIF and SNIF concepts based on a detection frequency approach. A compromise between higher reliability than other techniques and a minimum number of panelists was found with 6−8, ideally 8−10, assessors. Under such conditions, two independent panels were able to generate similar aromagrams without training the panelists prior to the GC-O analysis.

Although they are not a direct measurement of the perceived odor intensity, NIFs and SNIFs increase with concentration and, consequently, with odor intensity. Therefore, they can be used to compare peak intensities between two aromagrams.

As the proposed sampling well simulates the conditions of the headspace aroma surrounding a food, this new headspace-GC-sniffing technique has recently been applied to determine impact flavorings in yogurt and to deduce their origin (Ott et al., 1997).

ACKNOWLEDGMENT

We gratefully acknowledge Mrs. A. Quintar for providing us with the flavor cocktail, Mr. Y. Krebs for the sniffing software, Drs. E. Prior, S. Bobillot, and H. Brevard for reviewing this paper, and all colleagues who participated in the sniffing panel.

LITERATURE CITED

Abbott, N.; Etievant, P.; Issanchou, S.; Langlois, D. Critical evaluation of two commonly used techniques for the treatment of data from extract dilution sniffing analysis. *J. Agric. Food Chem*. **1993a**, *41*, 1698−1703.

Abbott, N.; Etievant, P.; Langlois, D.; Lesschaeve, I.; Issanchou, I. Evaluation of the representativeness of the odor of beer extracts prior to analysis by GC eluate sniffing. *J. Agric. Food Chem*. **1993b**, *41*, 777−780.

Acree, T. E. Sensible principles and techniques. In *Flavor Science*; Acree, T. E., Teranishi, R., Eds.; American Chemical Society: Washington, DC, 1993.

Acree, T. E.; Barnard, J. Gas chromatography-olfactometry and Charm analysis. In *Trends in Flavour Research*; Maarse, H., van der Heij, D. G., Eds.; Elsevier Science: Amsterdam, 1994.

Acree, T. E.; Barnard, J.; Cunningham, D. G. The analysis of odor-active volatiles in gas chromatographic effluents. In *Analysis of Volatiles*; Schreier, P., Ed.; de Gruyter: New York, 1984.

Baird, J.; Noma, E. *Fundamentals of Scaling and Psychophysics*; McCorquodale, K., Ed.; Wiley: New York, 1978; p 23.

Blank, I. Gas chromatography-olfactometry in food aroma analysis. In *Techniques for Analyzing Food Aroma*; Marsili, R., Ed.; Dekker: New York, 1996.

Bliss, C. I. *Statistics in Biology*; McGraw-Hill: London, 1967.

Chaintreau, A.; Grade, A.; Muñoz-Box, R. Determination of partition coefficients and quantitation of headspace volatile compounds. *Anal. Chem*. **1995**, *67*, 3300−3304.

Da Silva, A. A. P.; Lundhal, D. S.; McDaniel, M. R. The capability and psychophysics of Osme: a new GC-olfactometry technique. In *Trends in Flavour Research*; Maarse, H., van der Heij, D. G., Eds.; Elsevier Science: Amsterdam, 1994.

Grosch, W. Determination of potent odorants in foods by aroma extract dilution analysis (AEDA) and calculation of odour activity values (AOVs). *Flavour Fragrance J*. **1994**, *9*, 147−158.

Guichard, H.; Guichard, E.; Langlois, D.; Issanchou, S.; Abbott, N. GC sniffing analysis: olfactive intensity measurement by two methods. *Z. Lebensm. Unters. Forsch*. **1995**, *201*, 344−350.

Guth, H.; Grosch, W. Identification of potent odourants in static headspace samples of green and black tea powders on the basis of aroma extract dilution analysis (AEDA). *Flavour Fragrance J*. **1993**, *8*, 173−178.

Kalmus, H. Genetics of taste. In *Taste*; Beidler, L. M., Ed.; Springer Verlag: Berlin, 1971.

McDaniel, M. R.; Miranda-Lopez, R.; Watson, B. T.; Michaels, N. J.; Libbey, L. M. Pinot noir aroma: a sensory/gas chromatographic approach. In *Flavors and Off-Flavors*; Charalambous, G., Ed.; Elsevier Science Publishers: Amsterdam, 1990.

Ott, A.; Pollien, P.; Montigon, F.; Chaintreau, A. Improved headspace-GC-sniffing technique: screening of aroma impact flavorings. Presented at the 4th International Symposium on Hyphenated Techniques in Chromatography, Bruges, Belgium, **1996.**

Ott, A.; Fay, L. B.; Chaintreau, A. Determination of the aroma impact compounds of yogurt flavor. *J. Agric. Food Chem*. **1997**, *45*, 850−858.

Semmelroch, P.; Grosch, W. Analysis of roasted coffee powders and brews by gas chromatography-olfactometry of headspace samples. *Lebensm. Wiss. Technol*. **1995**, *28*, 310−313.

Stevens, S. S. To honor Fechner and his law. *Science* **1961**, *133*, 80−86.

Ullrich, F.; Grosch, W. Identification of most intense volatile flavor compounds formed during autoauxidation of linoleic acid. *Z. Lebensm. Unters. Forsch*. **1987**, *184*, 277−282.

Van den Dool, H.; Kratz, P. D. A generalization of the retention index system including linear temperature programmed gas-liquid partition chromatography. *J. Chromatogr*. **1963**, *11*, 463−471.

Voirol, E.; Daget, N. Comparative study of nasal and retronasal olfactory perception. *Lebensm. Wiss. Technol*. **1986**, *19*, 316−319.

Received for review November 25, 1996. Revised manuscript received April 18, 1997. Accepted April 23, 1997.[⊗]

JF960885R

[⊗] Abstract published in *Advance ACS Abstracts,* June 15, 1997.