Food Chemistry **47** (1993) 235–245

Application of pattern recognition techniques to sensory and gas chromatographic flavor profiles of natural orange aroma†

James C. C. Lin

Science and Technology Center, Kellogg Company, Battle Creek, Michigan 49016, USA

&

Steven Nagy & Martin Klim

Florida Department of Citrus, 700 Experiment Station Road, Lake Alfred, Florida 33850, USA

(Received 27 April 1992; revised version received and accepted 15 September 1992)

Meaningful analysis of citrus aroma required a multidimensional approach. First, organoleptic evaluation provided information on the product that could be equated to consumer acceptability. Secondly, gas chromatographic–mass spectral analysis yielded both identity and quantitation of the aroma constituents. Thirdly, computational profiling of organoleptic data and instrumental data were used to construct multidimensional plots that separated samples with different compositions and flavor attributes into several categories. Quantitative descriptive analysis (QDA) utilizing linear numerical scores were used to quantify 12 aromatic descriptors, and various pattern recognition routines of ARTHUR were applied to evaluate the data. A panel separated samples into categories described as fruity, fresh orange-like, flowery herbal-like, butter oil-like, pepper-like and fruity tea-like. Data treatment, depicted by nonlinear mapping (NLM), resolved these samples into categories with differing off-flavor and non-off-flavor attributes. Feature weightings (variance and Fisher) indicated that the descriptor 'overall quality' had the most discriminating power for sensory evaluation of good quality aroma samples. A separate ARTHUR analysis of gas chromatographic data clearly separated aroma samples into groupings that were related to the categories previously separated by an aroma panel and depicted by NLM. Ethyl acetate, acetal and ethyl butyrate were high variance- and Fisher-weighted GC features.

INTRODUCTION

Considerable effort has been expended in the past to develop a methodology capable of utilizing instrumental data to supplement or replace organoleptic sensory analysis for performing quality analysis of foods and beverages. Because of the complex nature of most beverages, simple mathematical routines as regression analysis have little chance of success. With the advent of powerful pattern recognition programs (Harper *et al.*, 1977), a closer look at the complex matrix of a fruit juice beverage using instrumental data is more realistic. A pattern recognition program consists of mathematical analysis of the data and mathematical modeling. Most programs can be 'taught' to recognize a pattern or patterns by analyzing a 'training' or previously classified set of data. These programs can also analyze data independently, to determine if any patterns exist.

Significant results have been achieved in applying pattern recognition to gas chromatographic data for classification of cheese varieties (Aishima & Nakai, 1987), stress-induced changes in coffee aroma (Roberts & Bertsch, 1987), and essential oils from different varieties of hops (Stenroos & Siebert, 1984). Applications in the citrus industry include the detection of adulteration of orange juice (Nikdel *et al.*, 1988), the determination of geographical origins of frozen concentrated orange juice (Bayer *et al.*, 1980), the classification of navel and Valencia orange essence oils (Mayfield *et al.*, 1986), and the identification of quality factors in orange juice (Rouseff & Nagy, 1987).

---

† Florida Agricultural Experiment Station Journal Series No. R-02379

The quality of natural orange aroma (NOA), also known as aqueous phase orange essence, is commonly accepted as a very important component in the production of good frozen concentrated orange juice (FCOJ) and in orange juices reconstituted from concentrates. Quality NOA is also used as an additive to impart a natural citrus flavor to synthetic drinks, as well as other food products. Factors such as fruit maturity, orange cultivar, and processing methods affect the quality and strength of NOA (Redd *et al.*, 1992). The objectives of this study were to establish a method to define commercially produced NOA by application of a pattern recognition program, and to investigate the relationship and the classification performance between subjective sensory evaluation and objective GC analysis. Many of the statistical routines of pattern recognition used in our study are defined in a book by Burgard and Kuznicki (1990).

MATERIALS AND METHODS

Sample preparation

Sixteen batches of natural orange aroma were collected from six flavor houses in central Florida and labeled A through P (Table 1). The samples were stored in 200-ml amber flasks at 2°C and warmed to ambient temperature when used. Basic compositional information was supplied by the flavor houses for their respective products. Fifteen components identified in NOA were selected for preparing solutions for sensory panel use (Table 2). Selection of compounds was based on previous studies by Johnson and Vora (1983) and Moshonas and Shaw (1984) and by our own extensive evaluation of over 100 different commercial NOA samples. All components were dissolved in a water base containing 14% ethanol.

Table 1. Commercial natural orange aroma: source and label designations

Commercial flavor house	Sample number	Sample code	Category	Test set
I	1	A	1	
I	2	B		X
I	3	C	2	
I	4	D		X
I	5	E	3	
I	6	F	4	
II	1	G	5	
III	1	H	6	
IV	1	I		X
IV	2	J		X
IV	3	K	7[a]	
IV	4	L	7[a]	
IV	5	M	8	
IV	6	N		X
V	1	O	9	
VI	1	P		X

[a] Same commercial product code but of different lot numbers from the same flavor house (IV).

Table 2. Flavor components of natural orange aroma: reported literature concentration levels and prepared concentration levels for the sensory panel

Flavor component	Reported concentration		Prepared concentration
	A[a]	B[b]	
Ethanol (base)	13%	NC[c]	14%
Acetaldehyde	600 ppm	NC	600 ppm
Methanol	0·5%	NC	0·5%
1-Propanol	100 ppm	40 ppm	100 ppm
Ethyl acetate	100 ppm	100 ppm	100 ppm
Acetal	70 ppm	45 ppm	70 ppm
Hexanal	trace	40 ppm	40 ppm
Ethyl butyrate	50 ppm	65 ppm	60 ppm
trans-2-Hexenal	trace	40 ppm	40 ppm
D-Limonene	200 ppm	10 ppm	200 ppm
Linalool	—	85 ppm	80 ppm
cis-3-Hexenol	—	10 ppm	10 ppm
Methyl butyrate	—	5 ppm	5 ppm
4-Terpinenol	—	20 ppm	20 ppm
α-Terpineol	—	20 ppm	20 ppm

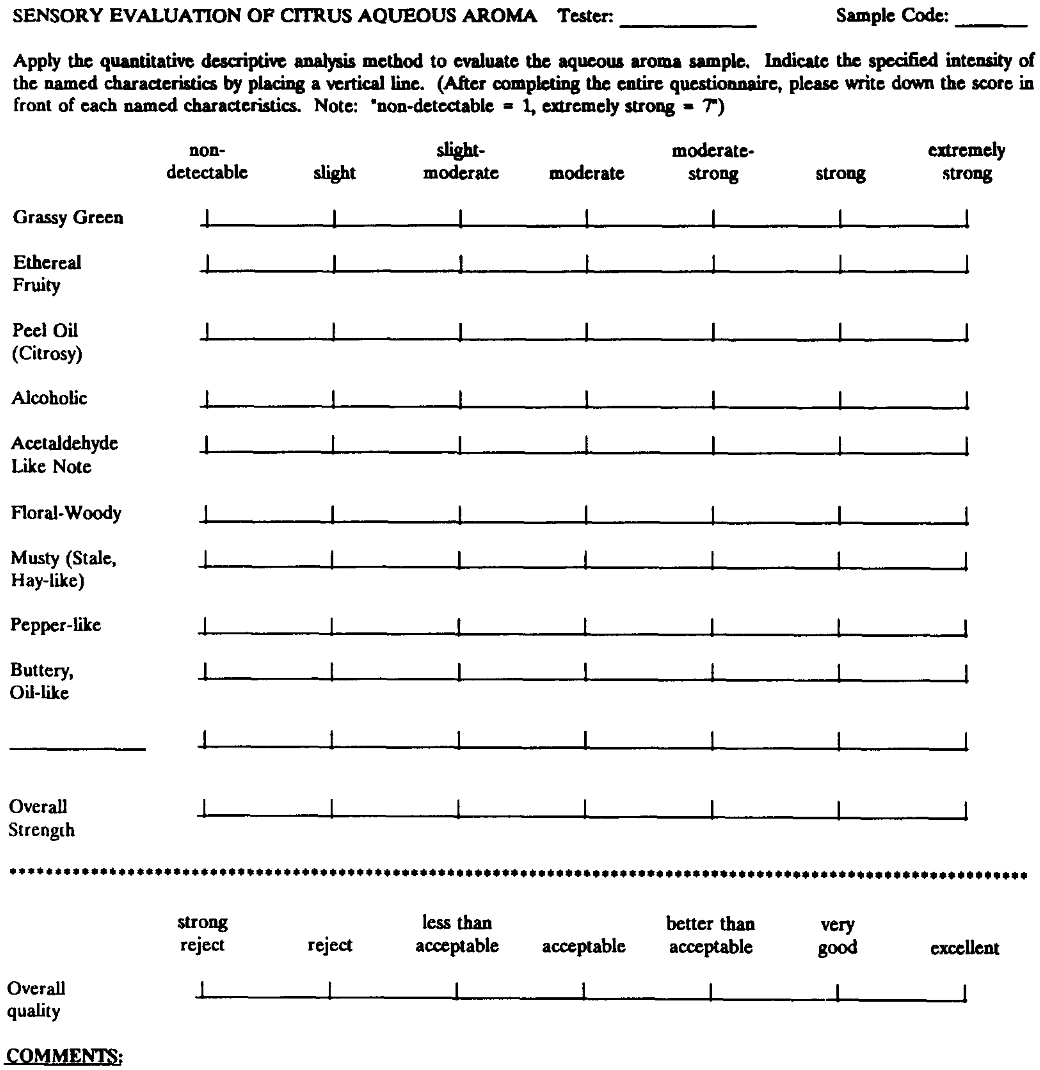
[a] Johnson and Vora, 1983.
[b] Moshonas and Shaw, 1984.
[c] NC-identified but not calculated.

Sensory evaluation

A sensory panel consisting of eight members was selected from the staff at the Citrus Research and Education Center and the training of panellists was conducted for two hours each week for three months prior to evaluating the NOA samples. The training consisted of recognition of selected aroma components in the 14% ethanol–water base, selection of specific aroma descriptors, development of a report form, and practice sessions with NOA samples. Emphasis was placed on reproducibility of perception for strength and quality. The quantitative Descriptive Analysis (QDA) report form (Fig. 1) included the aroma descriptors developed during training plus a blank line for any unusual aromatic characteristic, such as flowery herbal or fruity tea-like, that were used to describe qualities of some samples. Panellists were asked to draw a vertical line in any of the six segments to indicate the intensity of each characteristic. The left-most tick mark was assigned a score of 1 (equal of nondetectable), whereas the right-most tick mark was assigned a score of 7 (equal to an extremely strong aroma impact).

Gas chromatographic analysis

The NOA samples were analyzed with a Hewlett-Packard 5890A gas chromatograph (Avondale, PA) equipped with a flame ionization detector. The column was a 30-m fused silica (0·32mm i.d.) coated with RTX-5 (Restek, Bellefonte, PA) with a 1·0 μm thickness. The oven temperature was initially held at 32°C for 3 min and then programmed at 7·5°C/min to 185°C and held for 2 min. The injector and detector temperatures were 220°C and 260°C, respectively. The carrier gas was hydrogen with a linear velocity of 53·5 cm/s,



SENSORY EVALUATION OF CITRUS AQUEOUS AROMA **Tester:** ____________ **Sample Code:** ______

Apply the quantitative descriptive analysis method to evaluate the aqueous aroma sample. Indicate the specified intensity of the named characteristics by placing a vertical line. (After completing the entire questionnaire, please write down the score in front of each named characteristics. Note: 'non-detectable = 1, extremely strong = 7')

	non-detectable	slight	slight-moderate	moderate	moderate-strong	strong	extremely strong
Grassy Green							
Ethereal Fruity							
Peel Oil (Citrosy)							
Alcoholic							
Acetaldehyde Like Note							
Floral-Woody							
Musty (Stale, Hay-like)							
Pepper-like							
Buttery, Oil-like							
________							
Overall Strength							

* * *

	strong reject	reject	less than acceptable	acceptable	better than acceptable	very good	excellent
Overall quality							

COMMENTS:

Fig. 1. Quantitative descriptive analysis report form.

and makeup gas flow to the detector was dry air at 28 ml/min. Samples of 1 μl were manually injected and split at 50:1 ratio. An internal standard, 2-heptanol, was used for quantitation. The detector output was processed with a Hewlett-Packard 3393A computing integrator and transferred to an AT clone computer utilizing the Hewlett-Packard File Server program. Final data processing was done using a spreadsheet (Twin, Mosaic Software, Cambridge, MA) to prepare the data for input to a pattern recognition program.

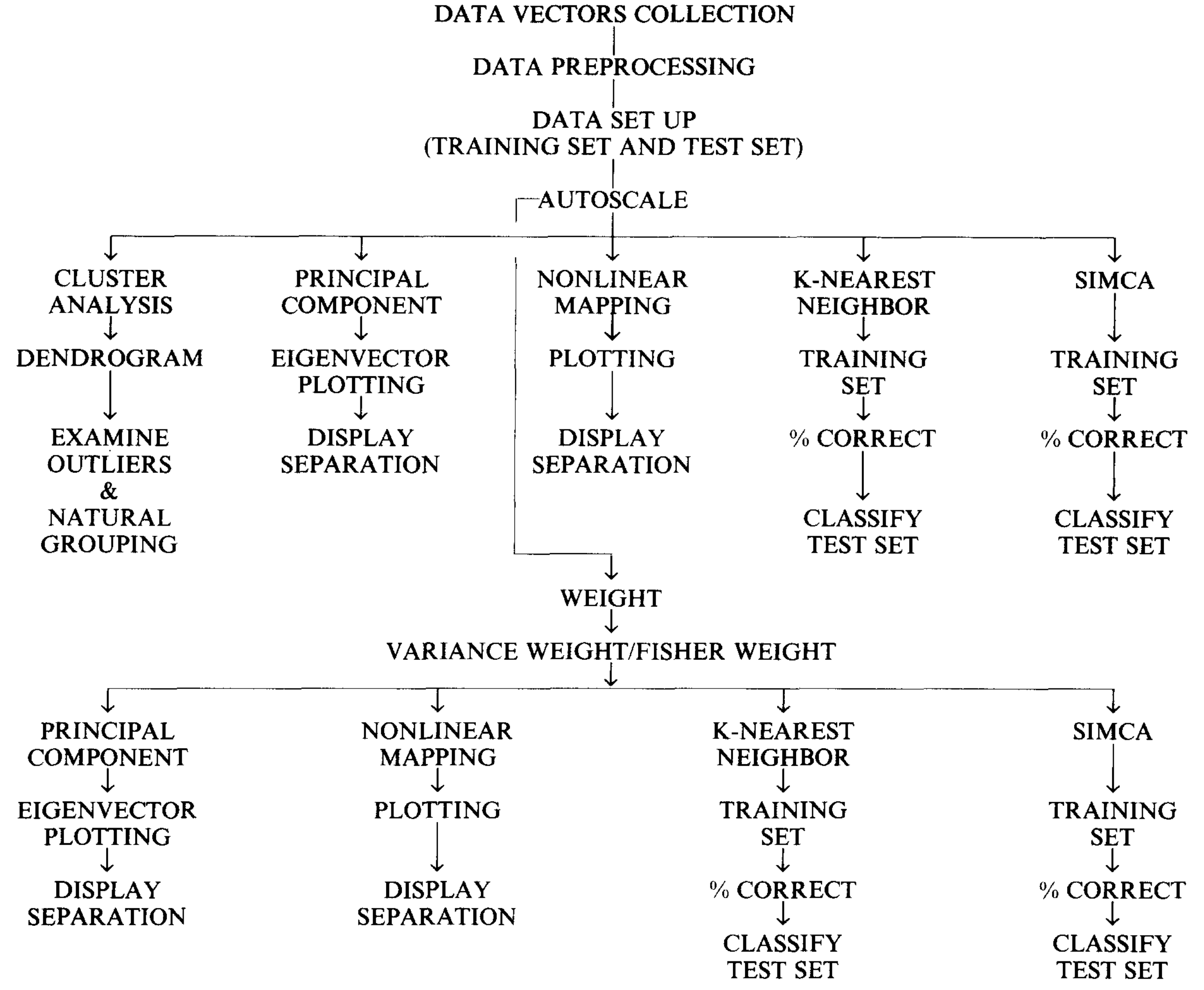
Gas chromatography–mass spectral identification

Mass spectral work was conducted in a gas chromatograph–mass spectrometer system consisting of (a) Carlo Erba Fractovap 4200 gas chromatograph retrofitted with a J & W on-column injector (J & W Scientific, Folsom, CA) and a Hewlett-Packard split–splitless injector operating at a 10:1 split ratio, (b) a Kratos MS 25 magnetic sector, double focusing, mass spectrometer, and (c) a Taknivent Vector 1 data acquisition system (Teknivent, St Louis, MO). Mass spectral searches were performed with both the Wiley/NBS and National Institutes of Science and Technology (NIST) mass spectral data bases. The capillary column was a Restek RTX-5 capillary column (0·25 mm i.d. × 30 m long). The gas chromatographic programming was as follows: initial temperature at 25°C for 3 min, temperature program from 25°C to 220°C at 6°C/min, and hold at 220°C for 5 min. Carrier gas (hydrogen) flow was 24 cm/s. Mass spectral scanning was conducted at 0·7 s/dec under the following parameters: 2 KV accelerating voltage, 70 eV ionization voltage, 200°C source temperature and GC–MS transfer line was at 210°C.

Pattern recognition program

The sensory QDA and GC data were analyzed using ARTHUR (version 1981, Infometrix Seattle, WA) running on a VAX 360 computer. Sample classification was analyzed by K-nearest neighbor (KNN) and by statistical isolinear multicategory analysis (SIMCA). Nonlinear mapping (NLM) and principal component analysis (PCA) were used for data display and dimensional analyses (Burgard & Kuznicki, 1990). SAS/GRAPH (SAS Institute, Cary, NC) was used for G-3D (three-dimensional) plots. More descriptive of the techniques of discriminant analyses may be found in the book by Burgard and Kuznicki (1990).

Table 3. Data processing flow chart of ARTHUR analysis



RESULTS AND DISCUSSION

The application of full-featured pattern recognition program to a set of data required a logical approach. First, an unsupervised learning mode was used to determine the existence of natural class characteristics. Secondly, the supervised learning mode was used to process a training set of data (data for which some of the characteristics to be studied were known and quantifiable), followed by a test set of data to determine the validity of the training. Finally, the set of data to be analyzed was processed and the results

Table 4. Principal component analysis (PCA) of the autoscaled variance- and Fisher-weighted sensory QDA data[a]

Eigenvector number	Variance-weighted PCA			Fisher-weighted PCA		
		Variance preserved			Variance preserved	
	Eigenvalue	Each	Total	Eigenvalue	Each	Total
1	9·765	26·3	26·3	33·980	50·4	50·4
2	6·194	16·7	43·0	13·490	20·0	70·4
3	5·566	15·0	58·1	10·530	15·6	86·0
4	4·498	12·1	70·2	6·751	10·0	96·0
5	3·804	10·3	80·4	1·222	1·8	97·8
6	1·638	4·4	84·9	0·727	1·1	98·9
7	1·538	4·1	89·0	0·393	0·6	99·5
8	1·133	3·1	92·1	0·147	0·2	99·7
9	0·801	2·2	94·2	0·065	0·1	99·8
10	0·717	1·9	96·2	0·045	0·1	99·9
11	0·555	1·5	97·7	0·032 5	0·0	99·9
12	0·502	1·4	99·0	0·029	0·0	100·0
13	0·386	1·0	100·0	0·015	0·0	100·0

[a] Both variance- and Fisher-weighted PCA were previously autoscaled.

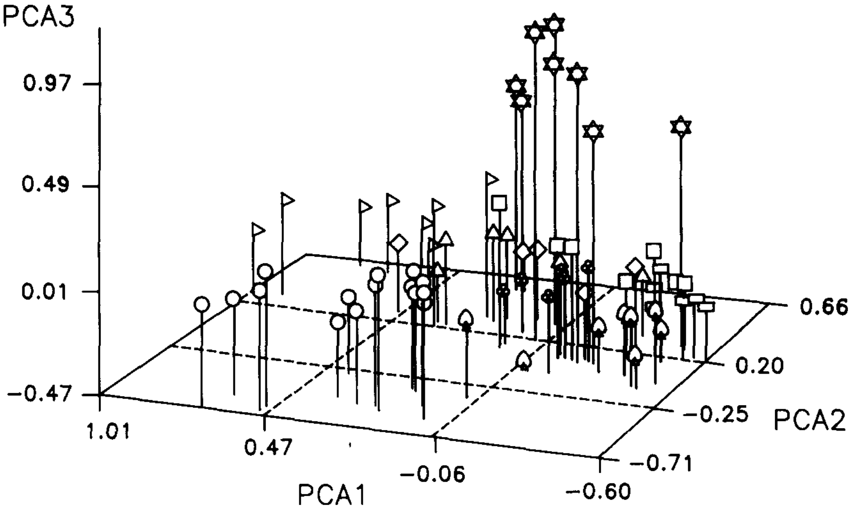


Fig. 2. Principal component analysis (PCA) of autoscaled variance-weighted QDA data; three-dimensional plot of first three eigenvectors. Sample symbols are as follows: club = sample A, diamond = sample C, cube = sample E, flag = sample F, square = sample G, spade = sample H, balloon = samples L and K, triangle = sample M, star = sample O.

reviewed. Table 3, a data processing flow chart, was developed on the basis of our data structure and working experience.

Sensory QDA

The purpose of feature (transformed variable) weighting in sensory analysis is to provide a measure of the discriminating ability of a variable (specific aroma note, see Table 5 below) to separate NOA samples into categories. Practical consequences of feature selection include a potential reduction in the number of measurements that need to be made for sample characterization, an indirect evaluation of experimental design, and an improved understanding of the basic chemical and sensory differences between NOA samples. Two statistical procedures by which this may be accomplished are variance weighting and Fisher weighting. The variance weight for categories is calculated as the ratio of the intercategory variances to the sum of the intracategory variances, whereas the Fisher weight for categories is calculated as the ratio of the difference of category means to the sum of the intracategory variances (Sharaf *et al.*, 1986). The statistical routine which best differentiates categories is determined by experimentation.

Table 4 reveals the results of principal component analysis (PCA) of the autoscaled variance- and Fisher-weighted sensory data. The first three eigenvectors explained 58·1% and 86·0% of the data set, respectively, Figures 2 and 3 are three-dimensional (G-3D) plots of variance- and Fisher-weighted sensory data. The variance-weighted plot (Fig. 2) reveals more scattering among categories than the Fisher-weighted plot (Fig. 3). However, Fisher-weighted plots of samples E, G, and D were tightly grouped but lacked separation of the respective groups. A review of the feature weights (Table 5) reveals that off-flavor notes such as buttery oil-like, flowery herbal, and fruity tea-like have heavier weightings than other aroma characters. Since unusual off-flavor notes evoke a stronger sensory response, the heavier weightings are to be expected. Overall quality,

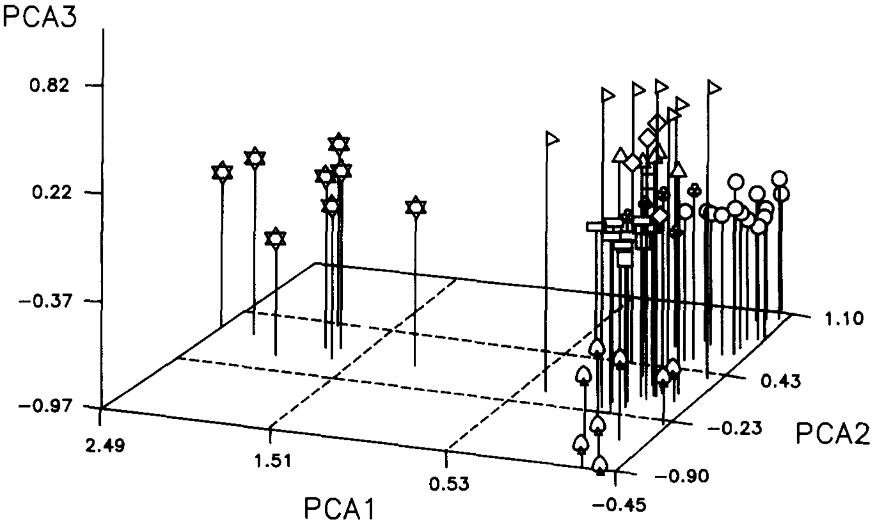


Fig. 3. Principal component analysis (PCA) of autoscaled Fisher-weighted QDA data; three-dimensional plot of first three eigenvectors. Sample symbols are as follows: club = sample A, diamond = sample C, cube = sample E, flag = sample F, square = sample G, spade = sample H, balloon = samples L and K, triangle = sample M, star = sample O.

Table 5. Results of feature weighting programs on sensory ODA data

		Variance weights			Fisher weight		
Old feature		New feature			New feature		
No.	Description	No.	Description	Weight	No.	Description	Weight
1	Green grassy	9	Buttery, oil	2·545	10	Herbal flowery	5·529
2	Ethereal fruity	10	Herbal flowery	2·239	11	Fruity, tea	3·058
3	Peel oil	11	Fruity, tea	2·075	9	Buttery, oil	2·885
4	Alcoholic	13	Overall quality	2·004	8	Pepper-like	2·384
5	Acetaldehyde-like	8	Pepper-like	1·950	13	Overall quality	1·313
6	Floral, woody	2	Ethereal fruity	1·567	2	Ethereal fruity	0·671
7	Musty, hay	6	Floral woody	1·460	6	Floral woody	0·561
8	Pepper-like	7	Musty, hay	1·419	7	Musty, hay	0·468
9	Buttery, oil	12	Overall strength	1·253	12	Overall strength	0·256
10	Herbal, flowery[a]	1	Green grassy	1·198	1	Green grassy	0·195
11	Fruity, tea[a]	3	Peel oil	1·175	3	Peel oil	0·173
12	Overall strength	5	Acetaldehyde-like	1·131	5	Acetaldehyde-like	0·126
13	Overall quality	4	Alcoholic	1·094	4	Alcoholic	0·087

[a] Unusual aroma characteristics incorporated on the blank line of Fig. 1.

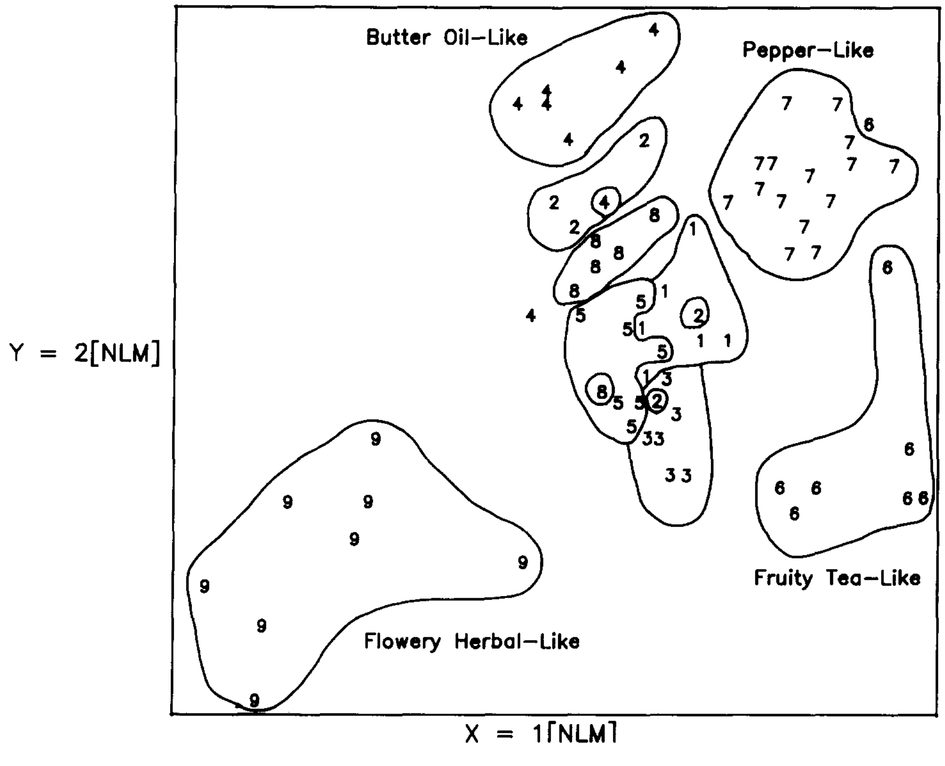


Fig. 4. Nonlinear mapping program of Fisher-weighted data vectors.

Table 6. The summary of K-Nearest Neighbor (KNN) on the autoscaled variance-weighted and Fisher-weighted sensory QDA data vectors of the training set.

Data vectors	K-NN	1-NN	3-NN	4-NN	5-NN	6-NN	7-NN	8-NN	9-NN	10-NN
Variance-weighted	Total missed	17	17	16	15	15	12	14	14	14
	% Correct	75·7	75·7	77·1	78·6	78·6	87·9	80·0	80·0	80·0
Fisher-weighted	Total missed	10	13	15	12	14	15	11	11	13
	% Correct	85·7	81·4	78·6	82·9	80·0	78·6	84·3	84·3	81·4

Table 7. The SIMCA Misclassification Matrix of the variance-weighted sensory QDA data vectors of the training set[a]

True class	Computed Class 1	2	3	4	5	6	7	8	9
1	6	0	0	0	0	0	0	0	0
	100·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
2	0	6	0	0	0	0	0	0	0
	0·0	100·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
3	0	0	6	0	0	0	0	0	0
	0·0	0·0	100·0	0·0	0·0	0·0	0·0	0·0	0·0
4	0	0	0	8	0	0	0	0	0
	0·0	0·0	0·0	100·0	0·0	0·0	0·0	0·0	0·0
5	0	0	0	0	7	0	0	0	0
	0·0	0·0	0·0	0·0	100·0	0·0	0·0	0·0	0·0
6	0	0	0	0	0	8	0	0	0
	0·0	0·0	0·0	0·0	0·0	100·0	0·0	0·0	0·0
7	0	0	0	0	0	0	15	0	0
	0·0	0·0	0·0	0·0	0·0	0·0	100·0	0·0	0·0
8	0	0	0	0	0	0	0	6	0
	0·0	0·0	0·0	0·0	0·0	0·0	0·0	100·0	0·0
9	0	0	0	0	0	0	0	0	8
	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	100·0

[a] Percent correct: total 100·0; average 100·0.

as noted, appears to be an important discriminating power, whereas peel oil, acetaldehyde-like and alcoholic are seen as less important qualities.

The non-linear mapping program (NLM) of the Fisher-weighted data vectors is shown in Fig. 4. NLM is another display technique which preserves interpoint distances to the extent possible when projecting into a two-dimensional space for plotting. By encircling the various nine categories, Fig. 4 reveals good separation for off-aroma categories, as witnessed near the outside perimeters, but considerable overlap for the more pleasant, non-offensive categories near the center. This observation suggests that by applying NLM programs, good quality natural orange aroma will more likely be found in the central region of the plot, whereas aromas with off-flavour, more offensive notes will be found near the perimeters of the NLM plot.

KNN classification of the sensory QDA data vectors of the training sets (see also Table 1) is summarized in Table 6. The best results were for the 7-NN result (87·9% correct) for variance-weighted data and 1-NN result for Fisher-weighted data (85·7% correct). However, the SIMCA results of the variance-weighted data vectors for the training sets resulted in '100% correct' responses (Table 7). Similar application of SIMCA analysis to the Fisher-weighted data vectors yielded a '96·8% correct' (one class missed) response.

The superior performance of SIMCA over KNN in sample classification was obvious; similar observations have been reported elsewhere (Stenroos & Siebert, 1984; Mayfield *et al.*, 1986). Those authors indicated that SIMCA analysis was the most consistent; it yielded satisfactory results for sample classification when compared to other discrete classification methods. Table 8 reveals overall classification by SIMCA on the test sample sets. Samples D, I, J, and P (see also Table 1) were effectively classified into categories 2, 7, 7, and 3, respectively. Sample B was not accurately classified. Since variance-weighted data vectors demonstrated 100% success with the training sets (Table 7), it was a logical assumption that, based upon its prediction, sample B had a flavor property intermediate between categories 1 and 2 (i.e., samples A and C). This was a strong indication that samples A, B, and C (categories 1, 2, and 3) might have a continuous (non-differentiating) flavor property rather than belong to distinct categories.

Table 8. The SIMCA classification results on the autoscaled variance-weighted and Fisher-weighted sensory QDA data vectors

Test sample	Variance-weighted	Fisher-weighted
B	4/8 CATEGORY #1 4/8 CATEGORY #2	2/8 CATEGORY #1 4/8 CATEGORY #2 1/8 CATEGORY #3 1/8 CATEGORY #4
D	1/8 CATEGORY #1 7/8 CATEGORY #2	6/8 CATEGORY #2 1/8 CATEGORY #3 1/8 CATEGORY #5
I	2/8 CATEGORY #5 6/8 CATEGORY #7	1/8 CATEGORY #1 1/8 CATEGORY #5 6/8 CATEGORY #7
J	1/8 CATEGORY #1 1/8 CATEGORY #2 6/8 CATEGORY #7	2/8 CATEGORY #1 1/8 CATEGORY #5 5/8 CATEGORY #7
N	1/8 CATEGORY #2 1/8 CATEGORY #5 6/8 CATEGORY #8	4/8 CATEGORY #2 1/8 CATEGORY #5 3/8 CATEGORY #8
P	7/8 CATEGORY #3 1/8 CATEGORY #8	1/8 CATEGORY #2 7/8 CATEGORY #3

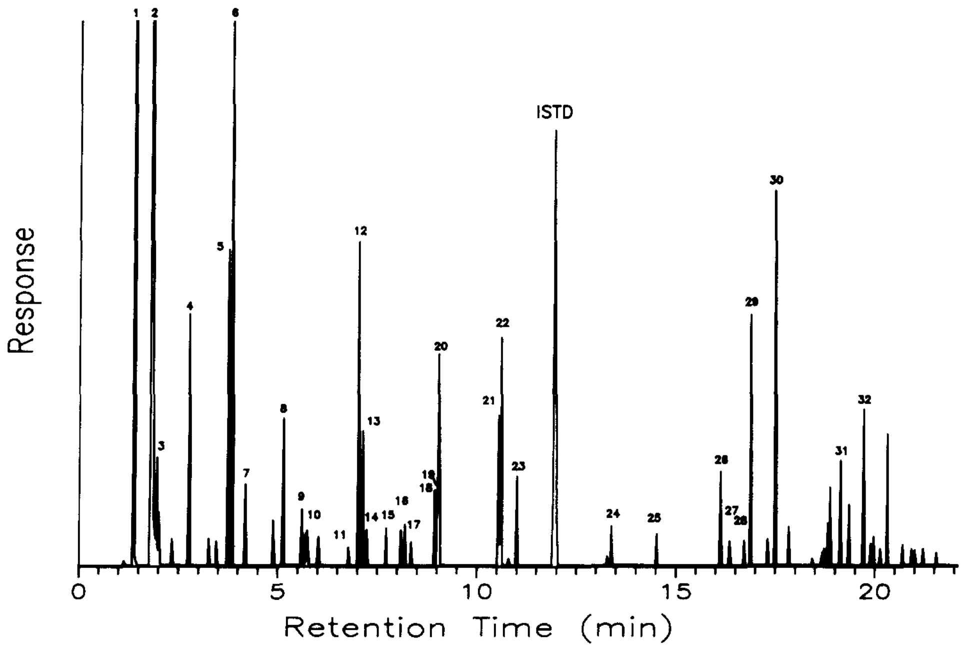


Fig. 5. A typical chromatographic profile of natural orange aroma. Peak identifications: (1) acetaldehyde and methanol, (2) ethanol, (3) acetone, (4) 1-propanol, (5) 2-methyl-3-buten-2-ol, (6) ethyl acetate, (7) 2-methyl-1-propanol, (8) 1-butanol, (9) 1-penten-3-ol, (10) ethyl vinyl ketone, (11) methyl butyrate, (12) acetal, (13) 3-methyl-1-butanol, (14) 2-methyl-1-butanol, (15) not identified, (16) 1-pentanol, (17) 2-penten-1-ol, (18) *cis*-3-hexenal, (19) hexanal, (20) ethyl butyrate, (21) *trans*-2-hexenal, (22) *cis*-3-hexen-1-ol, (23) 1-hexanol, (ISTD) 2-heptanol, (24) not identified, (25) octanal, (26) 1-octanol, (27) *cis*-linalool oxide, (28) *trans*-linalool oxide, (29) linalool, (30) ethyl-3-hydroxyhexanoate, (31) 4-terpineol, (32) α-terpineol. ISTD = Internal standard.

Gas chromatographic analysis

A typical gas chromatogram of natural orange aroma and peak identifications are noted in Fig. 5. Thirty peaks were identified by GC–MS. The first peak was not resolved by capillary gas chromatography but was previously resolved and identified with a 5% Carbowax 20M packed column and was shown to contain acetaldehyde and methanol. Cluster analysis of the 32 flavor components of the NOA training set samples is shown in the dendrogram of Fig. 6. The groups were clearly separated (less than 0·65 similarity), except for G and E at 0·82 similarity. Table 9 lists the eigenvalues of the PCAs of the autoscaled GC data vectors. The first 20 eigenvectors explained 100% of the total variance, whereas the first three explained 82·5% of the total variance. A three-dimensional plot of the first three eigenvectors is given in Fig. 7. Samples I, J, K, and L

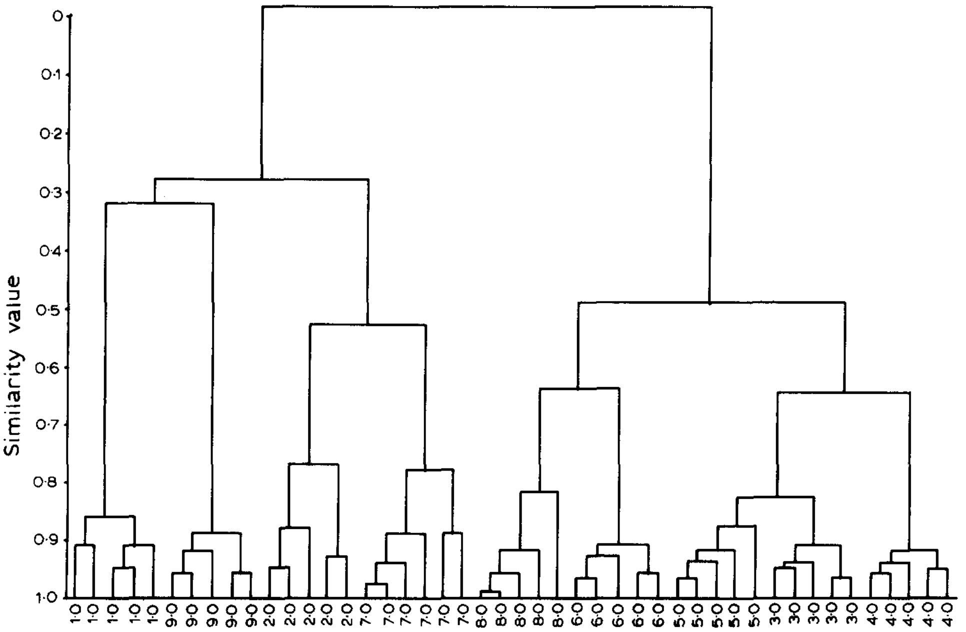


Fig. 6. Dendrogram of cluster analysis of 32 flavor components of natural orange aroma training samples.

(see also Table 1) defined a specific group. Groups A and B, H and M, E and F were not clearly separated based upon the first three principal components. However, samples C, D, O and P had sufficient separation for classification. Plotting the principal components in two or three dimensions is generally considered a display technique but can be used as a classification tool if the first two or three eigenvalues contain a sufficiently high percentage of the total variance to produce unambiguous separation. Headley and Hardt (1989) applied the first two principal components for classification of whiskies using Ein*Sight, a personal computer version of a pattern recognition program (Infometrix, Seattle, WA). KNN, SIMCA and other classification techniques were not available.

Results of NLM processing of autoscaled GC data are shown in Fig. 8. Unlike the sensory QDA data, a clear and distinct separation of the individual categories was observed. This might be due to the objective nature of the GC data when compared to the subjective aspect of sensory QDA data. The results of feature weighting (Table 10) indicated that acetal (P12) and ethyl acetate (P6) are the two compounds most responsible for categorizing samples. Methyl butyrate (P11), α-terpineol (P32), 1-propanol (P4), acetone (P3), and *cis*-3-hexenal (P18) were among the least important for category placement.

Table 11 shows the classification of variance-weighted GC data vectors. Classification of test samples I, J, and N are in agreement with sensory QDA results. However, SIMCA placed sample B in category 1 and KNN placed 3 out of the 5-B samples into category 7. Reviewing the sensory QDA data, it was noted that samples placed in category 7 had a pronounced peppery note, whereas B samples belonged in categories with no noticeable off-flavor notes. From this we conclude that SIMCA provided better classification results than KNN for analysis of GC data vectors. Conflicting classification results were also observed for samples D and P. SIMCA analysis of the autoscaled GC data vectors placed sample D in category 3 and sample P in category 2. The same analysis of sensory QDA data vectors placed samples D in category 2 and sample P in category 3 (the exact reverse). Further review of SIMCA analyses of sensory QDA data vectors demonstrated that samples B, D, I, J, N, and P belong to overlapping or adjacent categories. Therefore, we concluded that they belong to a flavor property that is continuous. Quantitation by gas chromatography for specific individual components provided sufficient precision to subdivide this continuous flavor property into discrete subsections for classification (see Table 11).

Table 9. Principal component analysis of the autoscaled gas chromatographic data vectors[a]

		Variance preserved	
Vector no.	Eigenvalue	Each	Total
1	15·080	47·1	47·1
2	7·036	22·0	69·1
3	4·293	13·4	82·5
4	2·069	6·5	89·0
5	1·179	3·7	92·7
6	0·872	2·7	95·4
7	0·570	1·8	97·2
8	0·301	0·9	98·1
9	0·238	0·7	98·9
10	0·111	0·3	99·2
11	0·076	0·2	99·5
12	0·070	0·2	99·7
13	0·025	0·1	99·7
14	0·017	0·1	99·8
15	0·013	0·0	99·8
16	0·012	0·0	99·9
17	0·008	0·0	99·9
18	0·006	0·0	99·9
19	0·005	0·0	99·9
20	0·004	0·0	100·0

[a] Only the first 20 eigenvectors are shown.

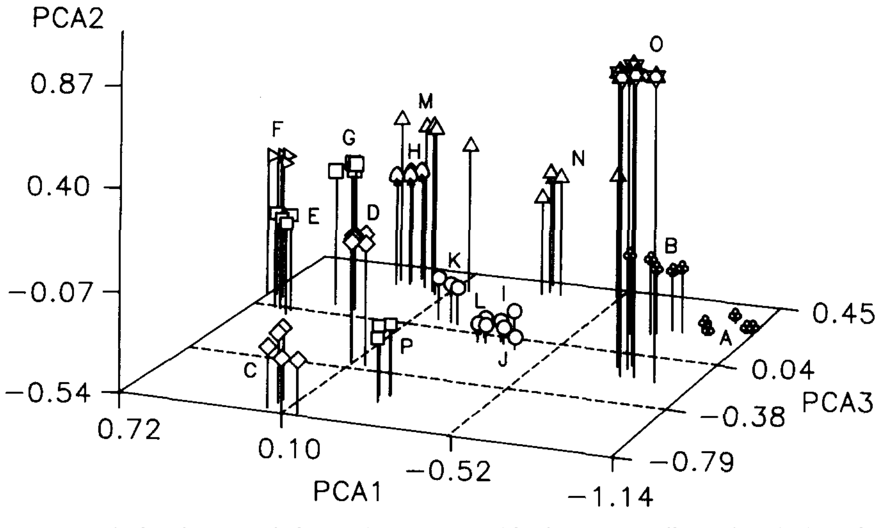


Fig. 7. Principal component analysis of autoscaled gas chromatographic data; three-dimensional plot of first three eigenvectors. Sample codes labeled from A through P (see also Table 1).

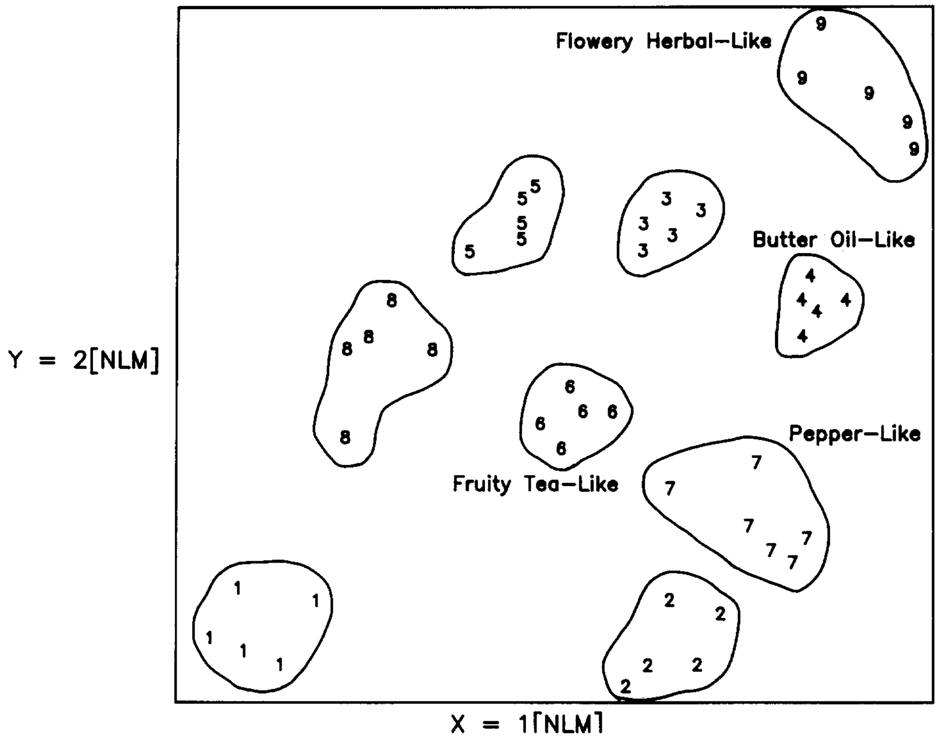


Fig. 8. Nonlinear mapping profile of autoscaled GC data.

Table 10. Results of feature-weighting programs on gas chromatographic data[a]

Variance weights			Fisher weights		
New feature			New feature		
Peak no.	Identity	Weight	Peak no.	Identity	Weight
12	acetal	168·300	6	ethyl acetate	660·300
6	ethyl acetate	154·500	12	acetal	524·600
10	ethyl vinyl ketone	98·910	29	linalool	288·500
19	hexanal	77·780	20	ethyl butyrate	287·200
20	ethyl butyrate	76·620	8	1-butanol	250·600
8	1-butanol	73·450	21	*trans*-2-hexenal	238·400
22	*cis*-hexen-1-ol	65·850	22	*cis*-3-hexen-1-ol	204·300
9	1-penten-3-ol	64·670	10	ethyl vinyl ketone	198·700
21	*trans*-2-hexenal	63·440	26	1-octanol	167·300
5	2-methyl-3-buten-2-ol	52·910	9	1-penten-3-ol	151·500
17	2-penten-1-ol	49·010	5	2-methyl-3-buten-2-ol	136·900
16	1-pentanol	41·950	19	hexenal	134·600
25	octanal	38·580	16	1-pentanol	120·900
26	1-octanol	35·230	17	2-penten-1-ol	113·600
13	3-methyl-1-butanol	33·070	15	NI	103·500
14	2-methyl-1-butanol	26·940	25	octanol	94·020
27	*cis*-linalool oxide	25·630	13	3-methyl-1-butanol	76·850
31	4-terpineol	25·530	31	4-terpineol	76·150
30	ethyl-3-hydroxyhexanoate	25·020	14	2-methyl-1-butanol	68·740
15	NI[b]	23·790	27	*cis*-linalool oxide	66·050
7	2-methyl-1-propanol	23·330	23	1-hexenol	60·870
24	NI	23·300	7	2-methyl-1-propanol	51·820
29	linalool	20·560	2	ethanol	40·620
23	1-hexanol	19·710	24	NI	37·220
2	ethanol	19·160	30	ethyl-3-hydroxyhexanoate	36·220
28	*trans*-linalool oxide	18·420	1	acetaldehyde + MeOH	33·060
1	acetaldehyde + MeOH	15·380	28	*trans*-linalool oxide	31·520
18	*cis*-3-hexenal	12·710	3	acetone	30·070
4	1-propanol	10·730	4	1-propanol	19·180
32	α-terpineol	9·284	18	*cis*-3-hexanol	17·770
3	acetone	7·711	32	α-terpineol	15·500
11	methyl butyrate	5·367	11	methyl butyrate	7·169

[a] See also Fig. 5.

[b] Not identified.

Table 11. The SIMCA and KNN classification results on test set autoscaled GC data vectors with variance-weighted program

Test sample	SIMCA	KNN classification	
B	5/5 CATEGORY #1	2/5 CATEGORY #1	3/5 CATEGORY #7
D	5/5 CATEGORY #3	2/5 CATEGORY #2	3/5 CATEGORY #5
I	3/3 CATEGORY #7	3/3 CATEGORY #7	
J	3/3 CATEGORY #7	3/3 CATEGORY #7	
N	5/5 CATEGORY #8	5/5 CATEGORY #8	
P	5/5 CATEGORY #2	5/5 CATEGORY #2	

CONCLUSION

The results of the sensory QDA study of natural orange aroma demonstrated that a well-trained sensory panel could effectively distinguish good quality from objectionable orange aromas. Secondly, whereas off-aroma samples were easily differentiated and placed into distinct categories, those samples with pleasant or nondescript aromas were difficult to subdivide, and to place into distinct 'good quality' categories. With respect to sample classification of QDA aroma data, the superior performance of SIMCA over KNN was noted.

The best differentiation of good quality orange aromas was accomplished by the use of SIMCA analysis of the gas chromatographic data. Whereas the continuous flavor properties of good quality NOAs could not be separated by QDA data, precise GC analytical data provided the means for this type of resolution.

REFERENCES

Aishima, T. & Nakai, S. (1987). Pattern recognition of GC profiles for classification of cheese variety. *J. Food Sci*, **52**, 939–42.

Bayer, S., McHard, J. A. & Winefordner, J. D. (1980). Determination of geographical origins of frozen concentrated orange juice via pattern recognition. *J. Agric. Food Chem.*, **28**, 1306–7.

Burgard, D. R. & Kuznicki, J. T. (1990). *Chemometrics: Chemical and Sensory Data.* CRC Press, Boca Raton, FL.

Harper, A. M., Duewer, D. L., Kowalski, B. R. & Fasching, J. L. (1977). ARTHUR and experimental data analysis: the heuristic use of a polyalgorithm. In *Chemometrics: Theory and Applications.* ACS Symposium Series 52, ACS, Washington, DC.

Headley, L. M. & Hardt, J. K. (1989). Classification of whiskies by principal component analysis. *J. Food Sci.*, **54**, 1351–4, 1358.

Johnson, J. D. & Vora, J. D. (1983). Natural citrus essences. *Food Technol.*, **37**(12), 92–3.

Mayfield, H. T., Bertsch, W. & Mar, T. (1986). Application of chemometrics to the classification of orange essence oil varieties by GLC. *J. High Resol. Chromatogr. & Chromatogr. Commun.*, **9**(2), 78–83.

Moshonas, M. G. & Shaw, P. E. (1984). Direct gas chromatographic analysis of aqueous and other fruit essences. *J. Agric. Food Chem.*, **32**, 526–30.

Nikdel, S., Nagy, S. & Attaway, J. A. (1988). Trace metals: defining geographical origin and detecting adulteration of orange juice. In *Adulteration of Fruit Juice Beverages*, ed. S. Nagy, J. A. Attaway & M. E. Rhodes. Marcel Dekker, New York, pp. 81–105.

Redd, J. B., Hendrix, D. L. & Hendrix, C. M., Jr (1992). *Quality Control Manual for Citrus Processing Plants*, Vol. 2. Ag Science, Inc., Auburndale, FL.

Roberts, D. & Bertsch, W. (1987). Use of computerized pattern recognition in the analysis of stress induced changes in coffee aroma. *J. High Resol. Chromatogr. & Chromatogr. Commun.*, **10**(5), 246–51.

Rouseff, R. & Nagy, S. (1987). A multivariate pattern recognition approach for identifying quality factors in Florida orange juice. In *Flavour Science and Technology*, ed. M. Martens, G. A. Dalen & H. Russwurm, Jr. Wiley, New York, pp. 481–8.

Sharaf, M. A., Illman, D. L. & Kowalski, B. R. (1986). *Chemometrics.* Wiley, New York.

Stenroos, L. E. & Siebert, K. J. (1984). Application of pattern recognition techniques to the essential oil of hops. *Am. Soc. Brew. Chem.*, **42**, 54–8.