JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Floral Classification of Honey Using Mid-Infrared Spectroscopy and Surface Acoustic Wave Based z-Nose Sensor

JAGDISH C. TEWARI*,† AND JOSEPH M. K. IRUDAYARAJ§

Department of Biological Sciences and Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, Indiana 46907

Fourier transform infrared spectroscopy (FTIR) and z-Nose were used as screening tools for the identification and classification of honey from different floral sources. Honey samples were scanned using microattenuated total reflectance spectroscopy in the region of 600−4000 cm^{-1}. Spectral data were analyzed by principal component analysis, canonical variate analysis, and artificial neural network for classification of the different honey samples from a range of floral sources. Classification accuracy near 100% was achieved for clover (South Dakota), buckwheat (Missouri), basswood (New York), wildflower (Pennsylvania), orange blossom (California), carrot (Louisiana), and alfalfa (California) honey. The same honey samples were also analyzed using a surface acoustic wave based z-Nose technology via a chromatogram and a spectral approach, corrected for time shift and baseline shifts. On the basis of the volatile components of honey, the seven different floral honeys previously mentioned were successfully discriminated using the z-Nose approach. Classification models for FTIR and z-Nose were successfully validated (near 100% correct classification) using 20 samples of unknown honey from various floral sources. The developed FTIR and z-Nose methods were able to detect the floral origin of the seven different honey samples within 2−3 min based on the developed calibrations.

KEYWORDS: FTIR spectroscopy; z-Nose; multivariate statistics; honey; neural networks

INTRODUCTION

Components of honey comprise carbohydrates, water, traces of organic acids, enzymes, amino acids, pigments, pollens, and wax in various proportions depending upon the floral and geographical origin and the harvest season. The average composition from a survey of about 490 samples of honey from the different regions in the United States presented in **Table 1** indicates that besides glucose and fructose, there are about 25 oligosaccharides present in minor (<2%) to trace (<0.1%) quantities with various concentrations depending on the variety and the source of the nectar. Sucrose present in honey is ∼1−3%; however, this level can increase if the bees are overfed with sugar (sucrose) during the spring season (*3*).

Various analytical methods such as GC-MS and HPLC are available to determine the floral source of honey; however, the experimental method to determine this could be involved. Floral identification of six of the most commonly used honeys was investigated, and their chemical characteristics were reported and compared with the Saudi Arabian standards via microscopic examination (*4*). Phenolic compounds of heather, lavender, acacia, rape, sunflower, rosemary, citrus, rhododendron, thyme, chestnut tree, and calluna honey samples were examined by capillary zone electrophoresis, and correlations between the phenolic profile and the botanical origin of the honey were established (*5*). The physicochemical properties of the honeys such as color, moisture, pH and acidity, lugol test, diastase index, reducing and nonreducing sugars contents, and hydroxymethylfurfural contents were also determined using a colorimetric method (*6*). GC-MS could also be used as an effective method for honey authentification studies based on aroma profiling (*7*). Near- and mid-infrared and spectroscopies have already shown good promise as powerful tools for the assessment of sugar contents in various food and agricultural products (*8*−*10*).

Table 1. Average Composition of Honey

component	av	SD	range
moisture (%)	17.2	1.46	13.4–22.9
fructose (%)	38.19	2.07	27.25–44.26
glucose (%)	31.28	3.03	22.03–40.73
sucrose (%)	1.31	0.95	0.25–7.57
maltose (%)	7.31	2.09	2.74–15.98
higher sugar (%)	1.50	1.03	0.13–8.49
lactone, mequiv/kg	0.335	0.13	0.0–0.95
ash (%)	0.169	0.15	0.02–1.028

It has been suggested that the ratios between the concentrations of amino acids could be used to determine the geographical/floral source of 98 honey samples (*11*, *12*). Examination of the ratio between the acids and amino acids reveals that

* Corresponding author (e-mail jtewari@purdue.edu).
† Department of Biological Sciences.
§ Department of Agricultural and Biological Engineering.

10.1021/jf050139z CCC: $30.25
Published on Web 08/12/2005

variations in the ratios between the same types of honey from different geographical locations exist; however, the variation between honeys from different floral sources was much greater. Honey from different botanical sources (acacia, citrus, chestnut, rhododendron, rosemary, and lime) have been analyzed by gas chromatography and the data evaluated statistically for possible development of amino acid profiling for classification (*13*). Chromatography has been used to determine the total amount of proline, leucine, and phenylalanine and their enantiomeric ratios in different honey samples. Significant amounts of D-leucine and D-phenylalanine have been found in honeys from different botanical and geographical origins (*14*).

The functional properties of honey very much depend on the volatile and semivolatile organic compounds, which vary on the basis of the floral origin and the method of handling. An elucidation of the origin of aroma compounds should lead to a better understanding of factors causing flavor differences between honeys. Traditional analytical techniques for quality evaluation include HPLC, GC with headspace sampling, and GC-MS analysis with solid-phase microextraction (*15, 16*). Other methods such as the mass spectrometry based e-nose (MS E-nose) have also been introduced as fast and sensitive techniques, but they are expensive (*17*). More rapid and simpler methods, however, are being pursued as attractive alternatives primarily because of their potential in reducing the analysis cost and time (*18*−*21*).

Limited availability and the increased price of honey have provided major incentives for falsification with other carbohydrates that could mimic the constituents of the authenticated product. Saccharides can be determined by a number of different methods based on the use of their physical and chemical properties (*22, 23*). Fourier transform infrared (FTIR) spectroscopic methods have been shown to detect the presence of beet invert, cane invert, and corn syrup in honey (*24*−*29*).

Arboleda and Loppnow have applied Raman spectroscopy to characterize a mixture of carbohydrates present in unknown sugar samples (*30*). FT-Raman spectra of several commercial samples of honey from different U.S. states showed that the relative intensities of the vibrational bands in the C−H stretch region of the FT-Raman spectra are sensitive to the observed physical states of the specimen (*31*). Several vibrational bands in the region between 500 and 1800 cm^{-1} could also be used as indicators of the two major components, fructose and glucose. The near-infrared spectroscopy has also been used for composition assessment (*32*). More recently, artificial neural networks (ANN) have been applied for food classification and authenticity evaluation (*33*).

The most recent study on the potential of a surface acoustic wave based sensor (*34*), the z-Nose, showed that complementary information related to flavor compounds could also be obtained rapidly and accurately. Hence, if a sensor fusion approach could be undertaken, where information on the volatile components obtained from z-Nose and the nonvolatile components from FTIR could be integrated, a robust protocol could be developed for honey quality analysis. The main objective of the proposed work was to evaluate the potential of FTIR spectroscopy and the GC-based z-Nose technology with relevant normalization algorithms in conjunction with multivariate statistics and ANN analysis for the classification of honeys based on their floral origin.

MATERIALS AND METHOD

Honey Samples. Clover [South Dakota (SD)], buckwheat [Missouri (MO)], basswood [New York (NY)], wildflower [Pennsylvania (PA)], orange blossom [California (CA)], carrot [Louisiana (LA)], and alfalfa [California (CA)] honeys were obtained from the National Honey Board (Longmont, CO) directly from the beekeepers (**Table 2**). Fifty samples from each floral honey were used for calibration; therefore, the calibration data set comprised spectra from 350 samples and the validation data set consisted of data from 20 different samples (**Table 4**) for each floral source (the total number of spectra in the validation set is 140).

Table 2. Classification Groups Based on Floral Origin

floral source of honey	group	floral source of honey	group
clover (SD)	1	orange blossom (CA)	5
buckwheat (MO)	2	carrot (LA)	6
basswood (NY)	3	alfalfa (CA)	7
wildflower (PA)	4		

Table 3. Vibrational Modes in the FTIR Spectra of Honey

wavenumber (cm^{-1})	vibrational group	vibrational mode
927	C—H (carbohydrates)	bending
991	C—O (C—OH)	stretching
1042	C—O (C—OH)	stretching
	C—O (C—OH)	stretching
1110	C═O of ketones	stretching/bending
	C—O (C—O—C) bond	stretching/bending
1259	C—O (C—OH)	stretching
1327	O—H (C—OH)	stretching/bending
1419	O—H (C—OH groups)	stretching/bending
	C—H alkenes	
2929	C—H (carbohydrate)	stretching
	O—H (carboxylic acids)	stretching
	NH_3^+ (free amino acids)	stretching

FTIR Spectroscopy Experiment. FTIR measurements were carried out using a Bio-Rad 3000 Excalibur spectrometer with a microattenuated total reflectance spectroscopy (m-ATR) sampling accessory (Pike Technologies, Madison, WI). The spectrometer was equipped with a deuterated triglycine sulfate (DTGS) detector, operating at 4 cm^{-1} resolution and 0.32 cm/s mirror velocity. Two hundred and fifty-six interferograms were co-added before Fourier transformation. The instrument was allowed to purge for 5 min with nitrogen gas (grade1) prior to acquisition of the spectra to minimize the spectral noise due to atmospheric carbon dioxide and water vapor. The sample station was equipped with an overhead m-ATR accessory that contains transfer optics, through which infrared radiation can be directed to a detachable ATR crystal. The m-ATR cell used as a sampling accessory has a single-reflection horizontal ATR (HATR) crystal (1.5 mm in diameter) made of zinc selenide (ZnSe) with a refractive index of 2.34 and a depth of penetration of 1.46 μm for small sample analysis. This accessory requires smaller sample volumes compared to the horizontal ATR; however, because of the single-bounce acquisition, the spectral clarity is not expected to be as high as with the multiple-bounce ATR crystal.

The reference spectrum was first recorded using a blank m-ATR cell. Single-beam spectra of all samples were obtained and rationed against the background spectrum of air to present the spectra in absorbance units. After every measurement, the m-ATR crystal was thoroughly washed with distilled water and dried. The cleaned crystal was examined by repeated blank measurement to ensure that no sample residue from the previous sample was retained on the crystal surface. All experiments were done in triplicate.

z-Nose Experiment. All z-Nose experiments were carried out using the model 7100/4100 vapor analysis system from Electronic Sensor Technology (Newbury Park, CA). The z-Nose operates at the speed of an electronic nose while delivering the precision and accuracy of a GC. The z-Nose consists of a sensor head, a support chassis, and a system controller housed within a small carrying case. The sensor head contains the hardware necessary to separate and detect the compounds in the analyte. The support chassis includes a small helium gas tank, power supply, and electronics to run the system using appropriate control systems. The analyzer is based on a single, uncoated quartz

Table 4. Discriminant Analysis of the FTIR Spectra of Honey Using PCA–CVA

			classification (%)			
			calibration		validatio validation	
analysis method	analysis region (FTIR and z-Nose)	no. of factors	total no. of samples	classification accuracy (%)	total no. of samples	classification accuracy (%)
FTIR PCA	950–1500 cm^{-1}	7	350	100	140	100
		6		99		98
FTIR PCA	2200–3200 cm^{-1}	7	350	98	140	96.2
		6		97		97.3
z-Nose PCA	1–10 s (full chromatogram)	7	350	100	140	100
		6		99		100

based surface acoustic wave sensor (SAW) with an uncoated piezoelectric quartz crystal that vibrates at a fundamental frequency. The crystal is in contact with the thermoelectric element, which controls the temperature for cooling during vapor adsorption and for heating during cleaning and operates by maintaining a highly focused and resonant surface acoustic wave at 500 MHz on its surface. Upon adsorption of the mass, the frequency of the surface acoustic wave will change in proportion to the adsorbant (*35*).

For z-Nose measurements, 8 g of honey was transferred into a 40 mL vial (98 mm length and 28 mm o.d.) and sealed with a screw cap containing a septum. The vials were then transferred to a water bath at 50 °C for 120 min to ensure that all of the sugar crystals were melted and to allow the aroma to equilibrate within the headspace of the vial for a minimum of 2 h. To prevent any leakage during this equilibration period, the screw cap with septum was covered with an extra plastic cap. An analysis temperature of 50 °C was chosen after the initial experiments, whereby the profiles of the pure honey samples were compared over five repetitions at room temperature and 50 and 70 °C. At room temperature the profiles were less concentrated and more sensitive to changes in ambient temperature. At 70 °C the profiles were very intense but more susceptible to noise, possibly due to reactions occurring in the honey at high temperature. At 50 °C the profiles were eventually both intense and stable and equilibration of the headspace was relatively fast. After equilibration, the samples were measured one by one with the z-Nose. The z-Nose consists of a 5 cm needle at the inlet, which was used for sampling through the septa of the vials. The sampling mode was set to 5 s, after which the system switched to a 10 s data acquisition mode. During this time period the gas sample was released from the trap inside the system and carried over the column (DB-5) by a helium flow of 3.00 cm^3/s. The different chemical components in the gas sample were separated on the basis of their molecular weights and sequentially detected by the SAW detector through their frequency shift; the data were collected every 0.02 s. The inlet temperature was 150 °C, the valve temperature was 120 °C, and the initial column temperature was 70 °C. During analysis the column temperature was ramped at the rate of 10 °C/s to a final column temperature of 100 °C. The SAW sensor was operated at a temperature of 40 °C. After each data sampling period, the system needed a 15 s baking period, in which the sensor was shortly heated to 125 °C and after which the temperature conditions of the inlet, column, and sensor were reset to their initial conditions. Between each sample measurement at least one blank was run to ensure proper cleaning of the system and a stable baseline.

Discriminant Analysis. Discriminant analysis is a procedure used to classify unknown samples into groups based on similarities to the characteristics of the training group (*36*). In this study all FTIR spectra in the region from 900 to 3600 cm^{-1} were exported from the WIN-IR Pro software (Bio-Rad), and the z-Nose data in the full range of time from 1 to 10 s were exported from the Micro sense software (Electronic Sensor Technology, Newbury Park, CA) as a "txt" file and imported directly into the WIN-DAS (Wiley, Chichester, U.K.) software for discriminant analysis. To develop the discriminant calibration model, honey samples were classified into seven groups to represent the respective clover (SD), buckwheat (MO), basswood (NY), wildflower (PA), orange blossom (CA), carrot (LA), and alfalfa (CA) sources (**Table 2**). Principal component analysis (PCA) and canonical variate analysis (CVA) were used to develop discriminant models.

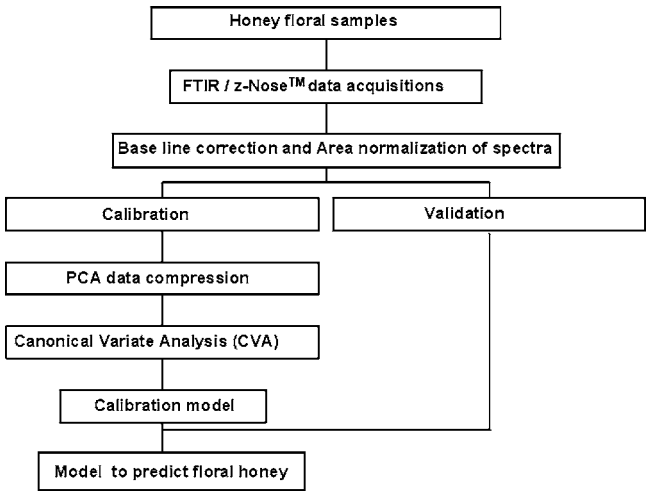


Figure 1. Schematic of the discriminant analysis procedure.

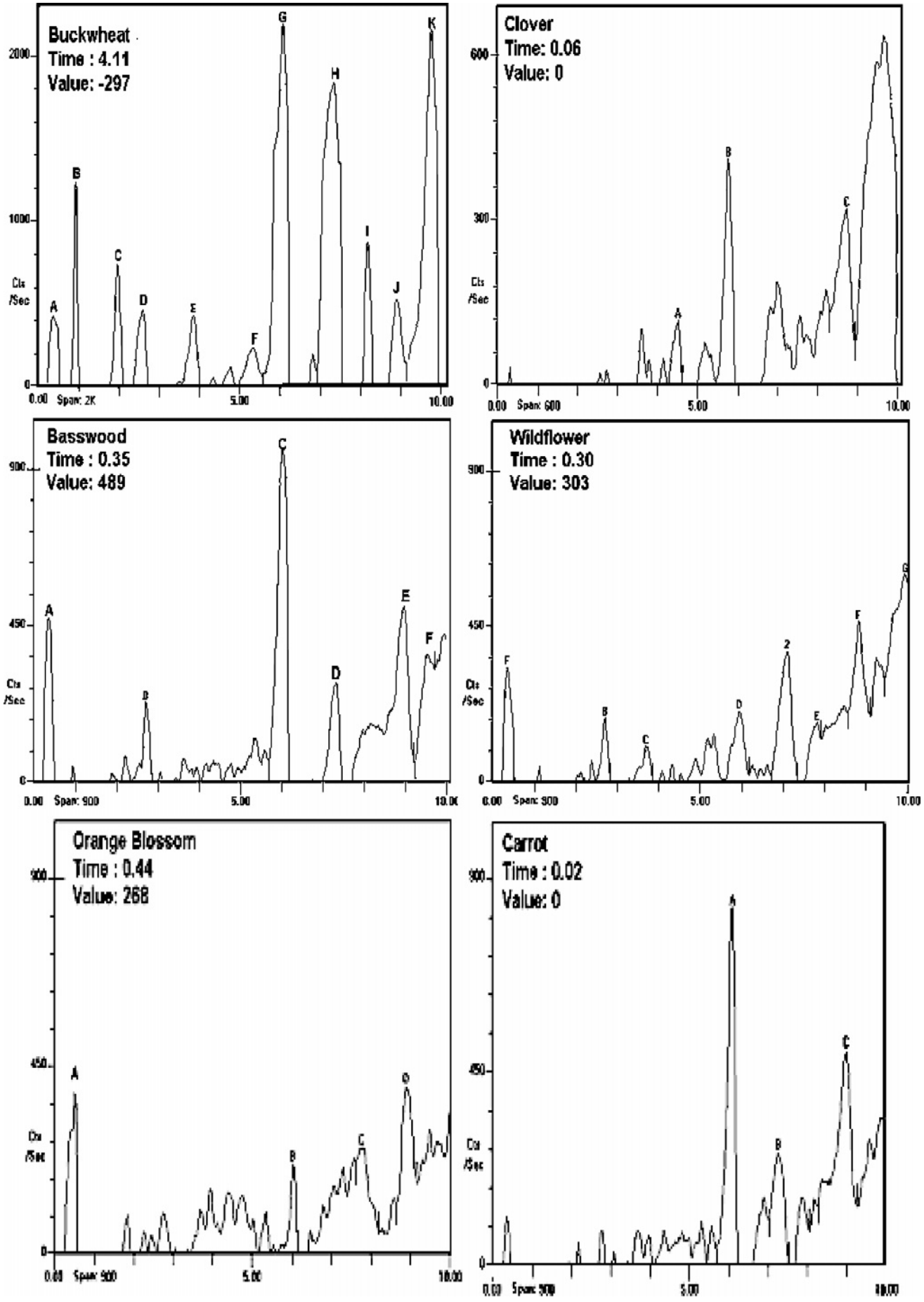
Principal Component Analysis. PCA is a very popular and powerful technique for the identification, classification, and other aspects of data evaluation. PCA decomposes the original matrix into several products of multiplication corresponding to the loadings and scores that indicate the variation of the data as well as the degree of fit. In this study PCA was performed in the spectral ranges between 900 and 1500 cm^{-1} and between 2200 and 3600 cm^{-1} to obtain a correct classification (**Table 4**).

The discriminant analysis procedure consists of three steps. In the first step, the spectral data were subjected to principal component (PC) data compression. In the second step, each spectrum is reconstructed by a linear combination of the product of PC scores and their weights (loading). In the third step, the PC scores are used for multiple group classification. The PC scores were then used in CVA.

Canonical Variate Analysis. The second method used for discriminating between groups of observations is the CVA. Canonical variate scores have successively maximized between-group variance/within-group variance, and the CV loadings are obtained as eigenvectors of a matrix given by

$$[W - 1][B]$$

where W is the within-group covariance matrix and B is the between-group covariance matrix (*37*). The objective of this procedure is to minimize the within-group variance and maximize the between-group variance. The goodness of fit is indicated by the percentage of correct classification. Discriminant models were based on the calibration data and evaluated separately using the validation data set. The correctly classified samples are expressed as a percentage of the total number of samples in the specific groups. A schematic of the discriminant analysis procedure is given in **Figure 1**. The spectra were normalized by dividing the intensity values corresponding to each wavenumber in the spectrum by its standard deviation before analysis.


Buckwheat
Time : 4.11
Value: -297
Clover
Time: 0.06
Value: 0
Basswood
Time : 0.35
Value: 489
Wildflower
Time : 0.30
Value: 303
Orange Blossom
Time : 0.44
Value: 268
Carrot
Time : 0.02
Value: 0

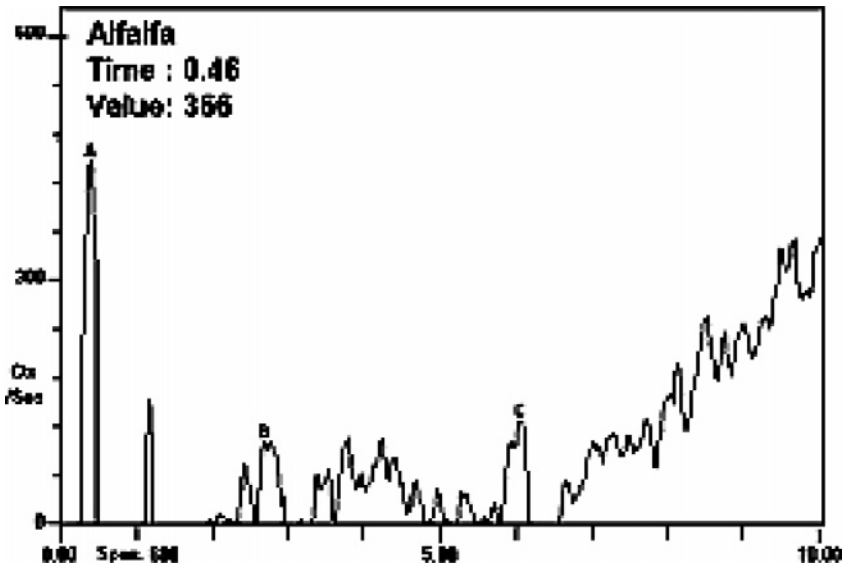


Figure 2. Original z-Nose chromatogram of honey from seven (buckwheat, clover, basswood, wildflower, orange blossom, carrot, and alfalfa) different floral sources.

Data preparation methods adopted for z-Nose data prior to PCA and CVA were slightly different; however, the entire spectrum of z-Nose within the first 10 s was used for improved classification. Because the z-Nose data can be looked upon as a chromatograph due to of its relevance to GC-based sensing and also as spectra primarily because of the signal from a sensor, two different data preparation methods could be pursued (*34*). In the GC-based data analysis approach, comparison of different peaks and peak areas was attempted. This was possible through the software of the instrument, which automatically transforms the frequency profile read from the SAW sensor to its first derivative. When only the positive values of this first derivative plot were considered, a chromatogram, which is similar to a regular GC chromatogram, resulted (**Figure 2A**,**B**). Each peak found in this plot corresponds to a specific volatile compound and retention time specific to the column and analysis temperature. The area under the peak was correlated to the compound concentration and was expressed in counts (cts). Relative peak areas were calculated as the absolute peak area (in counts) of each peak divided by the sum of all peak areas. When a peak was not present in a certain chromatogram, its relative area was set to zero.

In a spectroscopic data interpretation approach, the first-derivative profile (positive and negative values) was considered and treated as spectral data (**Figure 3**). In this case the full frequency spectrum of every sample was analyzed. Vertical baseline shifts in the frequency profiles were automatically filtered out by taking the first derivative. Next to the vertical shifts the horizontal shifts are a very common phenomenon in all types of chromatography. Small fluctuations in the injection time and operational variability can cause the different volatile components to be released and detected at slightly different retention times or within a "time window". In normal chromatographic analysis this is not a major hurdle because only a limited number of selected peaks are compared, each within its own window. However, when the full spectra are compared, this shift leads to misinterpretation, because a relevant peak could be compared with noise when the two spectra are not perfectly aligned. To correct for the time shifts along the horizontal axis, an algorithm developed and applied in one of our previous studies (*34*) using MATLAB version 6.1 (Mathworks, Inc.) was adopted. Hence the corrected frequency versus time spectrum with respect to a chosen reference spectra is given by

$$t_{\text{new},i} = a + bt_{\text{old},i} + ct_{\text{old},i}$$

where $t_{\text{new},i}$ is the new corrected time assigned to the ith data point in the z-Nose spectra, $t_{\text{old},i}$ is the original x-coordinate (i.e., time) for the ith frequency reading, and $a-c$ are the regression coefficients applied

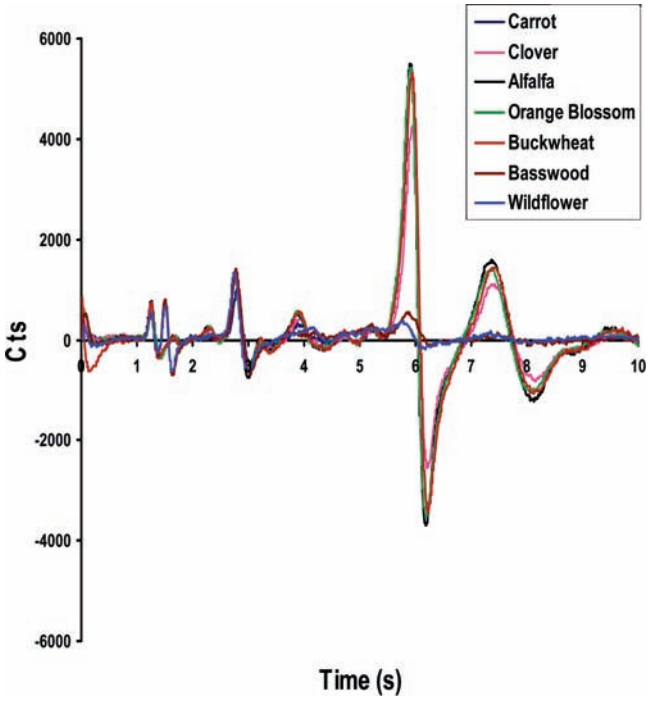


Figure 3. Baseline and time shift corrected z-Nose spectra of honey with buckwheat honey spectrum as reference.

to transform the old time value into a new value. For $a = 0$, $b = 1$, and $c = 0$, a horizontal shift correction is 0. For $a = 1$, $b = 1$, and $c = 0$, the spectrum shifts over a constant value a. For a positive and negative value of a, the shift occurs to the right or left, respectively. This approach was found to work satisfactorily in a z-Nose approach to detect adulteration in honey.

For data analysis in this study, a spectrum of buckwheat honey was selected as a reference spectrum because buckwheat is the product with the most complex aroma profile. On the basis of this reference, the remaining spectra were shifted horizontally to provide the best overlap with this reference spectrum. The three parameters ($a-c$) were adjusted manually to shift and stretch the spectra linearly or nonlinearly, depending on the alignment of a known frequency (or frequencies) with the reference spectrum. The spectra thus obtained are ready for further statistical analysis.

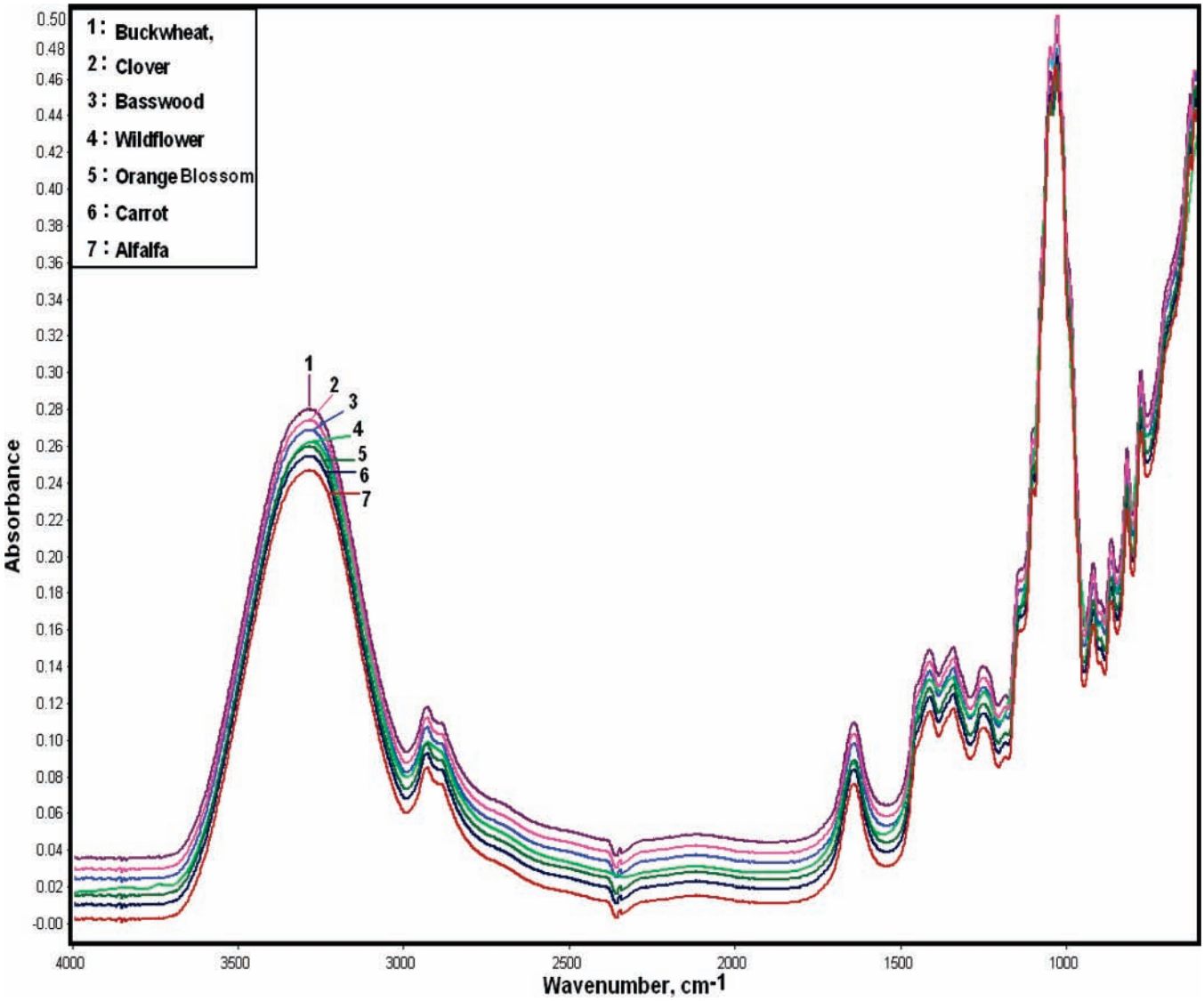


Figure 4. FTIR spectra of honey from seven different floral sources.

Artificial Neural Network Analysis. Baseline-corrected and area-normalized FTIR spectra of honey were converted as a txt file in "wavenumber versus absorbance" format. Similarly baseline-corrected z-Nose data were also converted as a txt file in "time versus area" format. The corrected data sets of FTIR and z-Nose were then used for neural network analysis for classification and floral similarities between seven different floral honey samples.

Three types of ANN were applied: a quick back-propagation network (BPN), a radial basis function network (RBFN), and a probabilistic neural network (PNN). All 350 honey samples were used for BPN, RBFN, and PNN to obtain the best classification network. BPN and RBFN were developed using the ANN tool box of Matlab 6.5 (Matlab, 2003), and PNN was computed using the neural net software from Ward Systems Group, Inc. (Baltimore, MD).

BPN is the most commonly used model, and RBFN is a type of neural network used for classification analysis (*10*). A three-layer [input (*i*), hidden (*j*), and output layer (*k*)] feed-forward network was used. The intensity at specific frequencies in the form of an *i*th-dimensional vector was used as input. For the BPN and RBFN methods the spectral region from 900 to 1500 cm^{-1} corresponding to the vibrational modes of most carbohydrates was used. For the PNN method the spectral range from 900 to 2200 cm^{-1}, which includes the vibrational modes due to carbohydrates, organic acids, amino acids, and other components of honey, was used. The calculations for the number of neurons were automatically done by the software according to the size of the data. The main intent is to assess the variation based upon the components of honey.

The region between 900 and 1500 cm^{-1} contains 46 data points representing intensities at various frequencies in this region. Thus, there are 46 neurons in the input layer of the network. The number of hidden neurons was optimized on a trial basis. The output layer consists of seven neurons representing the seven honey types based on floral origin. If the assignment due to ANN is correct, the output corresponding to the assignment group is set to 1 and the other outputs are set to 0. Before training, all weights were randomly chosen to be in the range from −0.5 to 1. A learning cycle (epoch) during which all of the training patterns are presented randomly to the network was used. The performance of the ANNs used in this study was tested with a validation set and monitored during the learning session to determine the terminating phase of the network so as to prevent overtraining. Overtraining occurs when an ANN is trained with an excessive number of learning epochs and thus loses its capacity to identify unknown patterns.

Quick BPN adopted two strategies of momentum and self-adaptation adjustment of learning rate to increase the learning rate and the reliability of the BP network. The activation function of all neurons was sigmoidal, a logistic function described by $\varphi(v) = 1/1 + e^{-a}v$, where v is the output signal of neurons, and the slope a is set to 1.0. The parameters chosen for BPN included a learning rate (lr) $\eta = 0.02$,

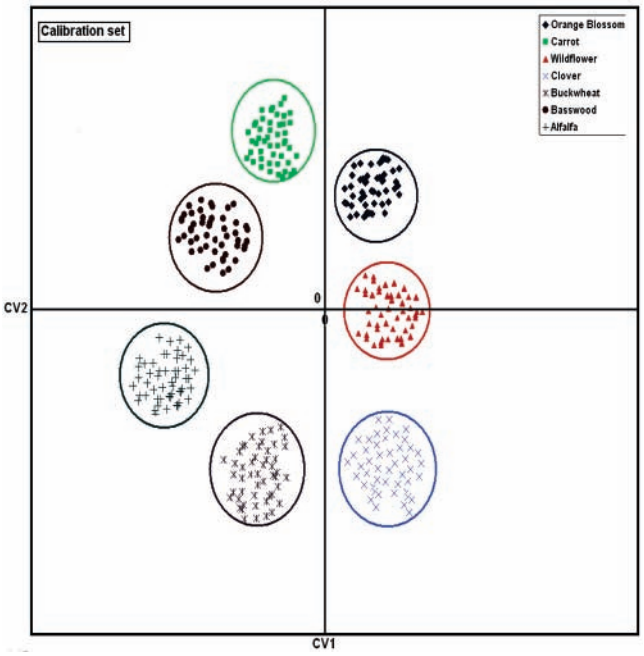


Figure 5. FTIR spectroscopy based classification of honey using PCA–CVA.

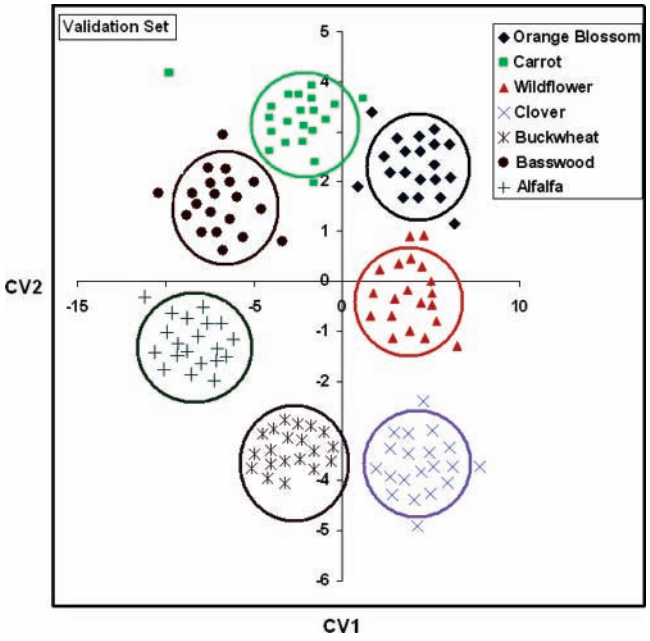


Figure 6. Validation of the calibrated PCA–CVA model with unknown samples.

a momentum constant (mc) = 0.9, and increase learning ratio (lir) = 1.05, a decrease learning ratio (ldr) = 0.7, and a maximum error ratio of 1.04.

PNN was also deigned for classification and to assess similarities between different floral honeys. Results from this network could be used to classify honey with similar compositional characteristics. This method is highly useful to identify the similarities between the floral honeys.

RESULTS AND DISCUSSION

Characterization of Honey by FTIR Spectroscopy. The numerous molecular motions of most polyatomic molecules afford a unique set of multiple and overlapping absorption bands in the fingerprint region of the infrared spectrum below 1500 cm^{-1} wavenumbers. **Figure 4** shows the overlay spectra (600−3800 cm^{-1}) of clover (SD), buckwheat (MO), basswood (NY), wildflower (PA), orange blossom (CA), carrot (LA), and alfalfa (CA) honey samples. The spectral region between 750 and 1500 cm^{-1} corresponds to the absorption region of monosaccharides such as glucose and fructose and disaccharides such as sucrose. The region of 750−900 cm^{-1} corresponds to the anomeric region characteristic of the saccharide configuration (*38*). The bands in the 904−1153 cm^{-1} region are assigned to C−O and C−C stretching modes (*39*), and those around 1199−1474 cm^{-1} are due to the bending modes of O−C−H, C−C−H, and C−O−H. Negative bands observed around 1618 and 3635 cm^{-1}

Table 5. Neural Network Analysis Using the Quick BPN Approach[a]

		calibration sample set			validation sample set		
neuron no. of hidden layer	optimum training epochs	right assignment no.	wrong assignment no.	classified correctly (%)	right assignment no.	wrong assignment no.	classified correctly (%)
10	5000	240	10	68.57	96	44	68.57
20	15000	265	85	75.71	114	26	81.42
40	22000	328	22	93.71	135	5	96.42
60	28000	315	35	90	122	18	87.14

[a] Learning rate (lr) η = 0.02; momentum constant (mc) = 0.9; learning increase ratio (lir) = 1.05; learning decrease ratio (ldr) = 0.7; maximum error ratio = 1.04.

Table 6. Neural Network Analysis Using RBFN

	calibration sample set			validation sample set		
group	right assignment no.	wrong assignment no.	classified correctly (%)	right assignment no.	wrong assignment no.	classified correctly (%)
1	50	0	100	20	0	100
2	50	0	100	12	8	60
3	50	0	100	16	4	80
4	50	0	100	20	0	100
5	50	0	100	20	0	100
6	50	0	100	19	1	95
7	50	0	100	20	0	100
overall results	81	0	100	35	3	95.72

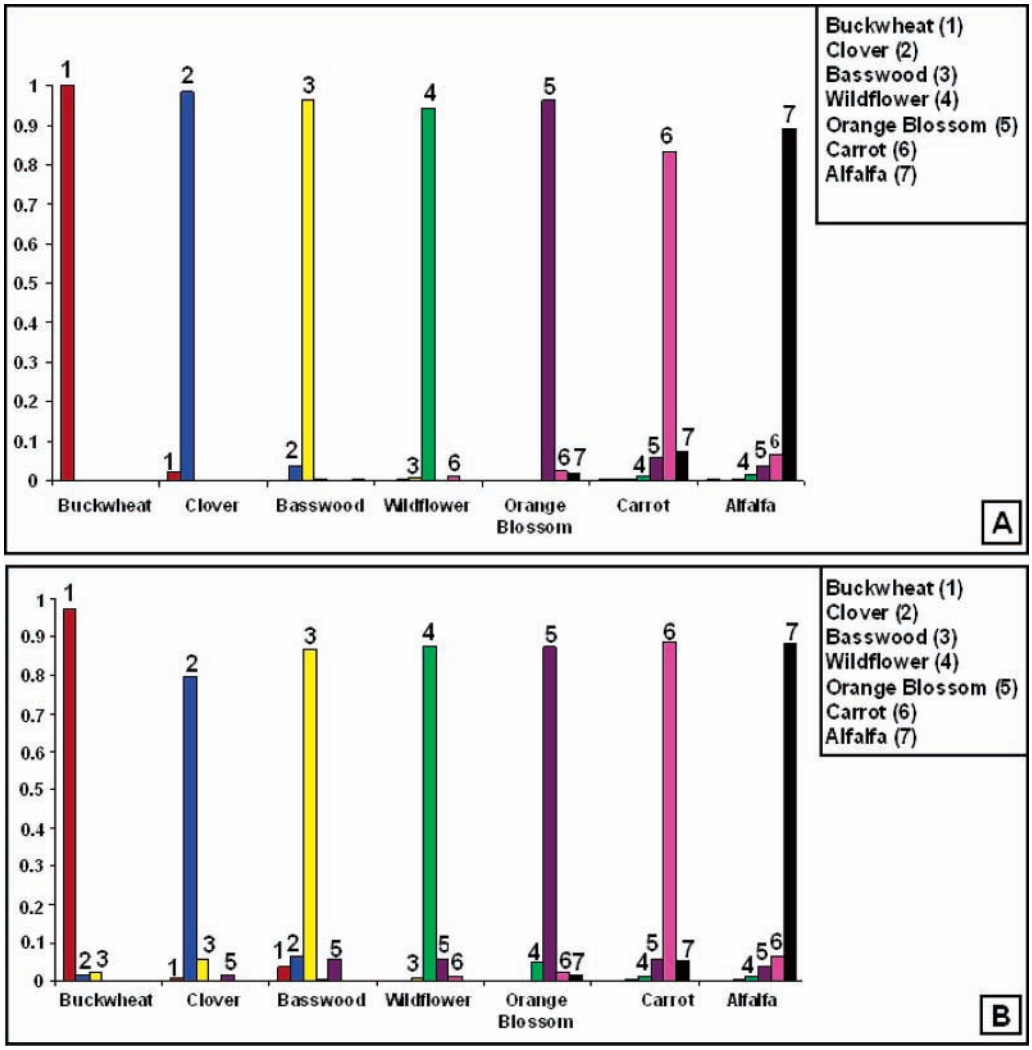


Figure 7. PNN for classification of honey based on FTIR spectra [(**A**) calibration set; (**B**) prediction set].

are due to lower water concentration in honey compared to the reference, because water presents an O−H stretch at these regions (*40*). The broad, strong NH_3^+ stretch in the 2600−3100 cm^{-1} region can be related to primary amino acids, which are present in honey at very low concentrations (*41*).

Table 3 presents the bands and functional group assignments along with the corresponding modes of vibrations in the FTIR spectrum of honey. The 927 cm^{-1} peak may be due to the C−H bending of carbohydrate, whereas the peaks observed at 991, 1042, 1106, and 1259 cm^{-1} may be due to the C−O stretch in the C−OH group as well as the C−C stretch in the carbohydrate structure. The peak at 1110 cm^{-1} could be related to the stretching of the C−O bond of the C−O−C linkage. The C−O−C is present here in sucrose as a glycosidic bond, a band linking monosaccharides such as glucose and fructose. The peak around 1327 cm^{-1} may be due to O−H bending of the C−OH group, and the band at 1419 cm^{-1} may be due to a combination of O−H bending of the C−O−H group and C−H bending of alkenes. The structure of organic acids such as fumaric acid has a CH═C bond, which may contribute to the peak at 1419 (*42*). The peak at 2929 cm^{-1} could denote the C−H stretching of carboxylic acids and NH_3^+ of free amino acids. Hence, the region from 800 to 1200 cm^{-1} could be characterized as the carbohydrate region, whereas the 1200−1800 and 2800−3200 cm^{-1} regions would indicate absorption due to organic and amino acids.

Classification of Floral Honey Based on Carbohydrate and Organic and Amino Acid Patterns from the FTIR Spectra. Chemometric models for the discrimination of honey based on floral origin are listed in **Table 3**. **Table 4** shows the classification results by data compression using PCA and discrimination using CVA (i.e. PCA−CVA). As stated earlier, regions from 950 to 1500 cm^{-1} and from 2200 to 3200 cm^{-1} represent the carbohydrate, organic acids, amino acids, and vitamins in honey that correspond to specific constituents such as glucose, fructose, sucrose, maltose, gluconic acid, citric acid, lactic acid, fumaric acid, succinic acid, malic acids, thiamin, riboflavin, niacin, pyridoxine (vitamin B_6), ascorbic acid (vitamin C), L-cystine, dimethylxanthine, and acetylcholine. Classification of honey from different sources using discriminant analysis is based on similarity of the characteristics of members in a cluster (*36*). The analysis pursued contained 50 data points (each data point denotes a spectrum) for each floral source, but from different locations. Two regions in the mid-infrared range, 950−1500 and 2200−2300 cm^{-1}, were used for calibration and validation. **Figures 5** and **6** present the respective PCA−CVA plots for calibration and validation based on the FTIR spectra. **Figure 5** presents the 96% confidence elliptical cluster of the different floral honeys using WINDAS. Their centers can be calculated from the means of the group coordinates, and their axes represent the values of the confidence regions for each dimension calculated by multiple regression. Well-separated groups

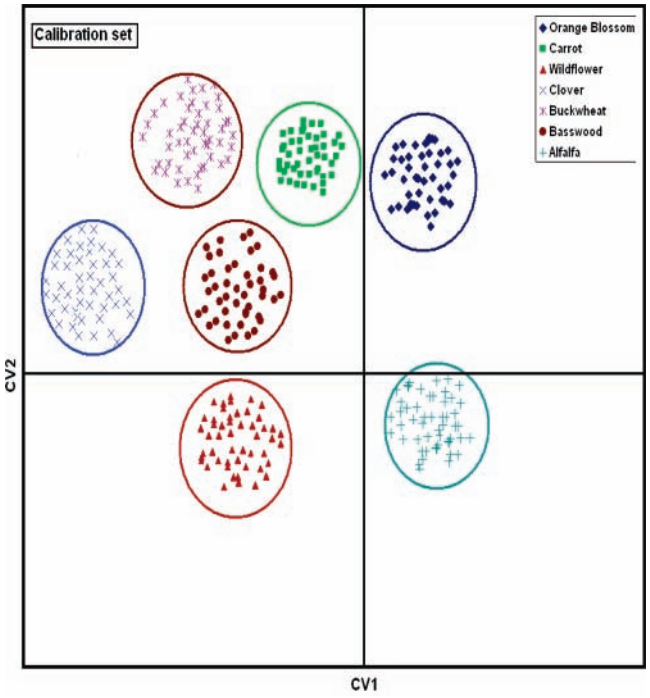


Figure 8. Classification of floral honey using z-Nose (calibration set) with PCA–CVA.

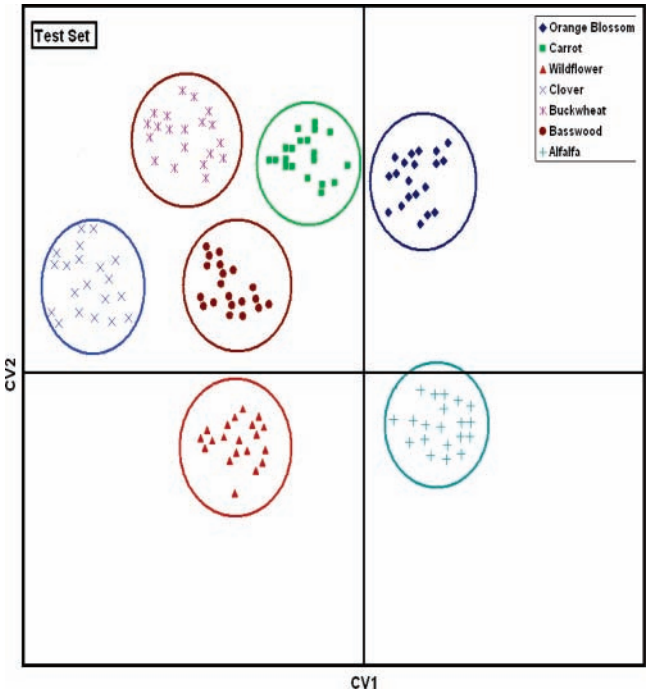


Figure 9. Validation of the calibrated PCA–CVA model using the corrected z-Nose spectra with unknown honey samples.

for all of the different floral honey can be observed from the CVA plot. **Figure 6** shows an excellent prediction of the unknown honey using the developed calibration model. FTIR results of the seven different types of honey show that the developed model could classify the floral honeys to nearly 100% accuracy.

FTIR spectra using information from the 950−1500 cm^{-1} region resulted in a 100% correct classification in the calibration set with seven PCA scores, and a 99% correct classification was achieved using six PCA scores. When the region of 2200−3200 cm^{-1} in the FTIR spectra was used, a 98% correct classification was achieved with seven PCA factors. Validation of the calibrated models using 20 spectra for each floral honey considered gave consistent prediction accuracies in the range between 97 and 100% as shown in **Table 4**.

Classification of Honey Using FTIR Spectra by Artificial Neural Networks. Neural networks look for patterns in the training data sets and learn to make accurate forecasts and predictions. The training data contain several sets of input variables corresponding to its output. The inputs are often called independent variables, and the output (classification) is called the dependent variable. Each set of the corresponding independent variables and dependent variable is called an observation or example. Examples in the training data should include a representative set of the problems likely to be encountered in real honey samples (*43*).

Table 5 shows the overall results for sample sets using quick BPN. As the number of neurons in the hidden layer increased, especially in the beginning stages, the assignment success rates also increased. When the number of neurons reached 60, the assignment success rates began to decrease. It was also shown that the greater the number of neurons in the hidden layer, the longer the training epochs. For the quick BPN, the optimum neuron numbers in the hidden layer and epochs were 40 and 22000, respectively. Using this approach, the final optimum architecture of BPN was 46−40−3, and the assignment success rate using BPN was 93.75% for both the calibration and validation sets. **Table 6** shows the overall results for sample sets using RBFN. For RBFN, the network automatically accounted for the number of neurons in the hidden layer to meet the training targets based on the training set. The final assignment success rates of RBFN were 100 and 95.72% for the calibration and validation sets, respectively. Although the RBFN's fit to the training data was 100%, the capacity to identify unknown patterns was not as high, possibly due to overtraining. Another probable reason could be the overlap between groups and the sensitivity and specificity of the sensor itself.

The PNN analysis used all of the 350 honey samples. Using an appropriate number of hidden neurons, the network classified all of the honey types with nearly 100% accuracy. **Figure 7A** shows the classification plots of the seven floral types of honey using PNN. The probability graph displays the prediction of the network for each selected honey types as a bar chart. The bar chart depicts the probability that a given set of inputs will lead to a correct classification in each of the output categories with the probability values at the output categories adding up to 1. The PNN study can be very useful for the identification of similarities of the floral honeys. Hence, when two output categories have probabilities that are fairly close, for example, with values such as 0.98 and 0.97, the implication is that the network did not have enough information to make an unequivocal classification; hence, additional inputs to the network were considered to obtain a better discriminatory model. **Figures 7** and **10** show that it is possible to classify honeys that have similar characteristics using the PNN. Thus, FTIR fingerprints could serve as a rapid means to identify honeys that have similar characteristics based on their carbohydrates as well as their amino acids, enzymes, and volatile and nonvolatile constituent contents. A separate set containing 20 independent samples from the honey types considered were used as a test set to further validate the PNN models. **Figure 7B** shows that the PNN models are validated with very high accuracy using the unknown sample set.

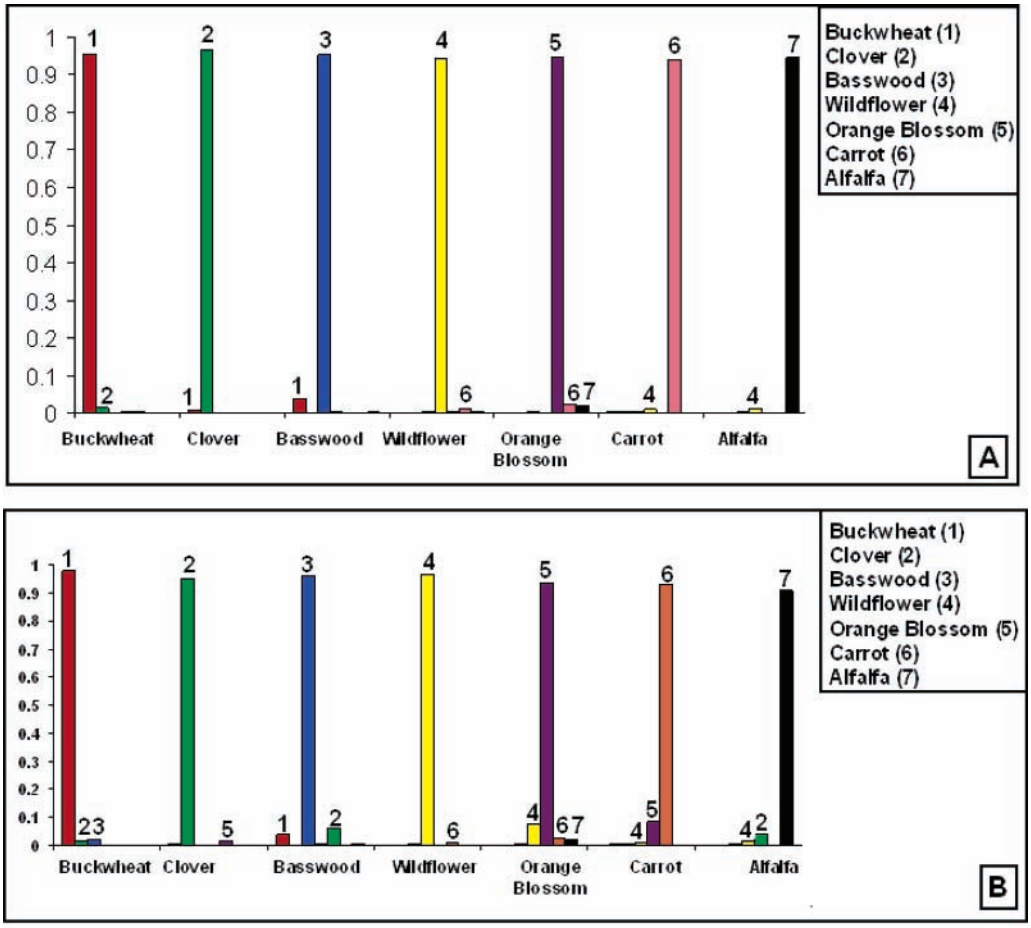


Figure 10. PNN for classification of honey based on the corrected z-Nose data [(**A**) calibration set; (**B**) prediction set].

Classification of Honey Using z-Nose Data by Multivariate Data Analysis. Figure 2A,**B** presents the original z-Nose chromatogram of the different types of honey. It has been clearly shown that a distinct chromatogram can be obtained for the different honeys due to their qualitative and quantitative differences in volatile components. The data analysis procedure is similar to FTIR classification except that the z-Nose data with all of the peaks were used. Fifty chromatograms for each floral honey were used to develop the calibration model. Hence, for the seven different floral honey types a total of 350 z-Nose chromatograms were used. **Table 4** presents the results of the statistical analysis using the z-Nose data including the calibration and validation accuracy along with the number of factors. In the chromatogram approach, the 12 most abundant honey volatiles were selected and used as variables in the discriminant analysis. PCA was used to reduce and compress the z-Nose data before CVA could be used.

The calibration model was successfully used to classify unknown honey samples up to an accuracy of 100% using six PCA factors and up to 99% accuracy using seven PCA factors (**Table 4**). **Figure 8** presents the classification of seven types of floral honey using the calibration model with excellent accuracy. In a two-dimensional canonical variate plot, all honeys can be visually discriminated. A nearly 100% classification accuracy was achieved using the z-Nose calibration set. A validation accuracy of 100% was achieved using the z-Nose data. Similar to the FTIR analysis, the validation set for z-Nose analysis consisted of 20 samples for each floral honey and is presented in **Figure 9**. Upon comparison, the z-Nose (**Figure 9**) predictions were better than the FTIR predictions (**Figure 6**).

Classification of Honey by z-Nose Data Using Probabilistic Neural Networks. Baseline-corrected chromatograms of all 350 samples were used to build PNN models. **Figure 10A** illustrates the probabilistic classification of seven types of honey using z-Nose. All 350 samples were clearly discriminated, and the classification accuracy of this model was 100%. An independent set of 20 samples for each floral honey also resulted in a nearly 100% prediction accuracy (**Figure 10B**) using the developed PNN model.

FTIR spectroscopy and z-Nose methods were successfully used to classify honeys from seven different floral sources [clover (SD), buckwheat (MO), basswood (NY), wildflower (PA), orange blossom (CA), carrot (LA), and alfalfa (CA)]. Chemometric and ANN models developed using FTIR and z-Nose yielded R^2 values >0.98. Fingerprints of honey from seven floral sources were sufficiently specific to discriminate on the basis of their nonvolatile as well as aroma composition. The z-Nose approach in conjunction with PCA−CVA and ANN using a chromatogram approach with relative peak areas or a spectral approach following a time-dependent peak identification process could be successfully used for honey authenticity studies. A key advantage of the z-Nose approach is that it is portable; however, suitable data correction procedures were necessary beyond the standard baseline and intensity normalization. Honey quality sensing could be further improved if the volatile constituent information from z-Nose and nonvolatile

constituent information from FTIR spectroscopy could be used to provide complementary information.

ACKNOWLEDGMENT

Drs. Jeroen Lammertyn and Dr. Els Veraverbeke are acknowledged for algorithm development and for their generous assistance in z-Nose data analysis.

LITERATURE CITED

(1) White, J. W.; Riethof, M. L.; Subers, M. H.; Kushnir, I. Composition of American honey. *USDA Tech. Bull.* **1962**, *No. 1261*, 12.

(2) Tewari, J.; Irudayaraj, J. Quantification of saccharides in multiple floral honeys using Fourier transform infrared mirco-attenuated total reflectance spectroscopy. *J. Agric. Food Chem.* **2004**, *52*, 3237−3243.

(3) White, J. W. Honey. *Adv. Food Res.* **1978**, *4*, 287−374.

(4) Abu-Tarboush, H. M.; Al-Kahtani, H. A.; El-Sarrage, M. S. Floral type identification and quality evaluation of some honey types. *Food Chem.* **1993**, *46* (1), 13−17.

(5) Andrade, P.; Ferreres, F.; Isabel, G.; Francisco, A.; Tomás-B. Determination of phenolic compounds in honeys with different floral origin by capillary zone electrophoresis. *Food Chem.* **1997**, *60* (1), 79−84.

(6) Azeredo, L. D. C.; Azeredo, M. A. A.; Souza de, S. R.; Dutra, V. M. L. Protein contents and physicochemical properties in honey samples of *Apis mellifera* of different floral origins. *Food Chem*. **2003**, *80*, 249−254.

(7) Radovic, B. S.; Careri, M.; Mangia, A.; Musci, M.; Gerboles, M.; Anklam, E. Contribution of dynamic headspace GC-MS analysis of aroma compounds to authenticity testing of honey. *Food Chem.* **2001**, *72*, 511−520.

(8) Tewari, J.; Mehrotra, R.; Irudayaraj, J. Direct near infrared analysis of sugarcane clear juice using fiber optic transmittance probe. *J. Near Infrared Spectrosc*. **2003**, *11* (5), 351−356.

(9) Tewari, J.; Mehrotra, R.; Gupta, A.; Varma, S. P. Determination of sugar content in sugarcane juice using infrared reflectance spectroscopy. *Mapan* **2001**, *No. 1*, 301−304.

(10) Irudayaraj, J.; Fujiyan, X.; Tewari, J. Rapid determination of beet sugar and cane sugar adulterants in honey using FT-IR and neural network. *J. Food Sci*. **2003**, *68* (6), 2040−2045.

(11) Davies, A. M. C. Amino acid analysis of honey from eleven countries. *J. Apic. Res.* **1975**, *14*, 29−39.

(12) Davies, A. M. C. The application of amino acid analysis to determination of geographical origin of honey. *J Food Technol.* **1975**, *11*, 515−523.

(13) Pirini, A.; Conte, L. S. Capillary gas chromatography determination of free amino acids in honey as means of discrimination between different botanical sources. *J. High-Resolut. Chromatogr.* **1992**, 165−170.

(14) Pawlowska, M.; Armstrong, D. W. Evaluation of enantiometric purity of selected amino acids in honey. *Chirality* **1994**, *6*, 270−276.

(15) Perez, R. A.; Sanchez-Brunete, C.; Calvo, R. M.; Tadeo, J. L. Analysis of volatiles from Spanish honeys by solid-phase microextraction and gas chromatography−mass spectrometry. *J. Agric. Food Chem.* **2002**, *50*, 2633−2637.

(16) Zhou, Q. X. Identification and quantification of aroma-active components that contribute to the distinct malt flavor of buckwheat honey. *J. Agric. Food Chem.* **2002**, *50*, 2016−2021.

(17) Stijn, S.; Jeroen, L.; Amalia, Z.; Berna, Els, V.; Corrado, D.; Bart, M. An electronic nose and a mass spectrometry-based electronic nose for assessing apple quality during shelf life. *Postharvest Biol. Technol.* **2004**, *31* (1), 9−19.

(18) Staples, E. The z-Nose™, a new electronic nose technology for analytical exploration of the chemical world which surrounds us. In *Proceedings of the On-site Conference (Geo-Odyssey Meeting)*, Boston, MA, Nov 8, 2001.

(19) Staples, E. *The Chemistry of Black Tea Aroma*; EST Internal Report; April 2002.

(20) Staples, E. Detecting 2,4,6 TCA in corks and wine using the z-Nose™; http://www.estcal.com/TechPapers/TCA in Wine.doc, 2000.

(21) Kunert, M.; Biedermann, A.; Koch, T.; Boland, W. Ultrafast sampling and analysis of plant volatiles by a handheld miniaturised GC with pre-concentration unit. *J. Sep. Sci.* **2002**, *25*, 677−684.

(22) White, J. W. Detection of honey adulteration by carbohydrate analysis. *J. Assoc. Off. Anal. Chem*. **1980**, *63*, 11−18.

(23) Peris-Tortajada, M.; Puchades, R.; Maquieira, A. Determination of reducing sugars by the neocuproine method using flow injection analysis. *Food Chem.* **1992**, *43*, 65−69.

(24) Paradkar, M, M.; Irudayaraj, J. Discrimination and classification of beet and cane inverts in honey by FT-Raman spectroscopy. *J. Food Chem.* **2001**, *76*, 231−239.

(25) Sivakesava, S.; Irudayaraj, J. Detection of inverted beet sugar adulteration of honey by FTIR spectroscopy. *J. Sci. Food Agric.* **2001**, *81*, 683−690.

(26) Paradkar, M. M.; Irudayaraj, J.; Sakhamuri, S. Discrimination and classification of beet and cane sugars and their inverts in maple syrup by FT-Raman. *Appl. Eng. Agric.* **2002**, *18* (3), 379−383.

(27) Sivakesava, S,; Irudayaraj, J. Prediction of inverted cane sugar adulteration of honey by Fourier transform infrared spectroscopy. *J. Food Sci.* **2001**, *66* (7), 972−978.

(28) Kelly, J. D.; Downey, G.; Fouratier, V., Initial study of honey adulteration by sugar solutions using mid-infrared (MIR) spectroscopy and chemometrics. *J. Agric. Food Chem.* **2004**, *52*, 33−39.

(29) Downey, G.; Fouratier, V.; Kelly, J. D., Detection of honey adulteration by addition of fructose and glucose using near infrared trans reflectance spectroscopy. *J. Near Infrared Spectrosc.* **2003**, *11* (6), 447−456.

(30) Arboleda, P. H.; Loppnow, G. R. Raman spectroscopy as a discovery tool in carbohydrate chemistry. *Anal. Chem.* **2000**, *72*, 2093−2098.

(31) Oliveira, L.; Colombara, R.; Edwards, H. Fourier transform Raman spectroscopy of honey. *Appl. Spectrsoc.* **2002**, *56* (3), 306−310.

(32) Ding, P. Y. Q.; Tang, H. B. Y. K.; Xu, R. Determination of chemical composition of commercial honey by near-infrared spectroscopy. *J. Agric. Food Chem.* **1999**, *47*, 2760−2765.

(33) Fidencio, P. H.; Ruisanchez, I.; Poppi, R. J. Application of artificial neural networks to the classification of soils from Sao Paulo state using near-infrared spectroscopy. *Analyst* **2001**, *126*, 2194−2200.

(34) Lammertyn, J.; Veraverbeke, E.; Irudayaraj, J. Surface acoustic wave based z-Nose™ technology for rapid aroma profiling of honey. *J. Sensors Actuat.* **2004**, *98* (1), 54−62.

(35) Edward, J. S. Real time characterization of food & beverages using an electronic nose with 500 orthogonal sensors and VaporPrint™ imaging. Electronic Sensor Technology, www. estcal. com.

(36) Friedman, J. H. Regularized discriminant analysis. *J. Am. Stat. Assoc.* **1989**, *84* (405), 165−175.

(37) Kemsley, E. K. Chemometric methods for classification problems. In *Discriminant Analysis and Class Modeling of Spectroscopic Data*; Kemsley, E. K., Ed.; Wiley: Chichester, U.K., 1998; pp 16−18.

(38) Tul'chinsky, V. M.; Zurabian, S. F.; Asankozhoev, K. A.; Kogan, G. A.; Khorlin, A. Y. Study of infrared spectra of oligosaccharides in the region 1000−400 cm^{-1}. *Carbohydr. Res*. **1976**, *51*, 1−15.

(39) Hineno, M. Infrared spectra and normal vibrations of β-D-glucopyronose. *Carbohydr. Res.* **1977**, *56*, 219−223.

(40) Tewari, J.; Joshi, M.; Gupta, A.; Mehrotra, R.; Chandra, S. Determination of sugars and organic acids concentration in apple juice using infrared spectroscopy. *J. Sci. Ind. Res.* **1999**, *58*, 19−24.

(41) Chen, M.; Irudayaraj, J. Sampling technique for cheese analysis by FTIR spectroscopy. *J. Food Sci.* **1998**, *63*, 96−99.

(42) Weston, R. J.; Brocklebank, L. K.; Lu, Y. Identification of quantitative levels of antibacterial components of some New Zealand honey. *Food Chem.* **2000**, *70*, 427−435.

(43) Specht, D. Probabilistic neural networks. *Neural Networks* **1990**, *3*, 109−118.

Received for review January 19, 2005. Revised manuscript received June 27, 2005. Accepted June 29, 2005. The National Honey Board (Longmont, CO) is acknowledged for partial funding of this work.

JF050139Z