

A chemometric study of human odor perception

James R. McGill, Fein-Marquart Associates, Baltimore, MD

The sensation of smell has been a topic of scholarly discussion since the time of ancient Greece. Two general areas of investigation have developed over the years; the problems of empirical odor classification and the process of odor reception. A number of related empirical classification schemes have been proposed. Perhaps the most complete odor classification is that of Zwaardemaker, which is listed in the first two columns of Table I.¹ While there is a fair degree of unanimity among such empirical schemes, theories of how odors are sensed are exceedingly diverse. There are presently four prominent theories of odor reception: shape sensors, vibration sensors, neural penetration, and acid-base interaction.

Amoore's *shape sensor theory* has been the most widely discussed of the four. It postulates that there are templates present in the olfactory area which sense the shape of the molecules in the passing air by fitting the molecules in a lock-and-key manner.² This theory developed from the observation that for certain categories compounds of similar odor have similar shapes when constructed from space-filling models. Seven categories have been postulated (see column 3 in Table I), but no unifying shapes have been found for pungent or putrid odors, so interactions other than shape must be involved.

The *vibration sensor theory* stems from the observation of Dyson,³ expanded by Wright,⁴ that compounds of similar odor absorb in similar Raman spectral regions. This led to the postulation of vibrational coupling between the molecules and the neurons in the nose. Recently, this theory has lost support because of the difficulty of postulating vibrational coupling in aqueous medium.

The *neural penetration theory* derives from experimentation which showed strong correla-

tions between the heat of absorption and odor of molecules,⁵ which may imply that the odorant disrupts the neural wall on absorption, causing the nerve to sense its presence. The *acid-base interaction theory* stems from the correlation between the Lewis acidity of molecules and their odor, although no mechanism has been proposed.⁶

None of the above theories are exclusive of any of the others, and none are complete explanations of odor sensation. What is needed is an empirically-derived data set of odor similarities which will allow comparison of these various theories. Such a data set was developed by Wright and Michaels, who presented 45 odor compounds to 84 volunteers and had them classify them on a six point scale as to the similarity

of the test compound's odor to nine different standard odorants.⁷ These standards are shown in column 4 of Table I. As one can see, the standards were chosen to span the empirical odor spectrum as well as possible. Five of the test compounds were presented twice to the volunteers so that an estimate of their error could be made. Thus one has a nine by 50 similarity matrix with each of the nine columns representing the similarity of the 50 compounds to a single standard and the nine values associated with each compound being a measure of the similarity of properties between that compound and a standard.

TABLE I
Comparison of Three Odor Classification Schemes

Zwaardemaker	BOOCE	Wright
Ethereal	Fruity Waxy Ethers Camphor Clove Cinnamon Aniseed	Hexyl Acetate(9) Camphor Eugenol(1) Benzothiozot(6)
Aromatic	Minty Thyme Rosy Citrous Almond Jasmine Orange Blossom	Minty Citral(7)
Fragrant	Lily Violet Vanilla Amber	S-8001(Lily of the Valley)(8)
Ambrosial	Musk Leek	Turpentine(2) Allyl Disulphide(4)
Alliaceous	Fish Bromine	Hydrogen Sulphide(5)
Empyreumatic	Burnt Phenolic	
Caprylic	Caproic Cat Urine	
Repulsive	Narcotic Bad Bug Carrion	
Nauseating	Fecal	Putrid Pungent Propionic Acid(3)

This data set was initially attractive because of the opportunities it provided to study both multiple property data and continuous property data from the viewpoint of the analytical chemist, and to see if these nine similarity continua could be related to spectral and physical information about the molecules. The ultraviolet, infrared, and nuclear magnetic resonance spectra were obtained from the literature, as were various physical parameters such as molecular weight and melting and boiling point. Chemometric methods were used to relate these data to the properties measured by Wright and Michaels.

Chemometrics is the application of data analysis techniques to chemical problems. The field is very new as an independent discipline, although chemists have been doing data analysis for many years. The most widely known section of the field is statistical analysis, which has been used extensively to aid in chemical inference. But classical statistical methodology needs large data sets, and this need severely limits its applications. For that reason, many methods have been developed to allow inference from smaller sets of data. Several different names exist for this area, but many prefer to call it pattern recognition. As the results from research areas as

dissimilar as electrical engineering and sociology are included in pattern recognition, there is very little agreement on the limits of the field, but generally one can define the three major sections as methods for classification of unknown data; methods for data display and determination of data structure; and methods for modifying data to aid the first two.

Several good reviews of the subject are available so details will not be given here.^{8,9} Some general introduction to the underlying concepts will be helpful though. If one has a compound which has had 43 measurements (also called features) taken on it, one can consider each of the measurements as a dimension. The value of each measurement positions the compound in question along that measurement axis. Thus the compound can be considered a point in a 43-dimensional space. The methods of pattern recognition attempt to determine the structure of this 43-dimensional space or use the structure to aid in understanding the problem in question. If one is trying to explain something about a particular point, that something is called the property of the point. In this application the property is the similarity of a particular molecule to a particular standard. The entire set of spectral and physical information about a molecule and its property form a data vector (sometimes called a pattern in older literature).

A number of methods from each of the three divisions of pattern recognition were applied to Wright and Michaels data set. The first question was whether the spectral and physical information collected from the literature was at all pertinent to the problem. In order to determine this, each property continuum was studied to determine how well its trends could be fit using the spectral and physical data. The error in the nine property continua was first determined by using the twice tested compounds. This error

TABLE II
Comparison of Fit Results

Category	Step	Features	Least(5)	Category
	Error	Used	Error	Error
1	1.3	21	4.5	6.6
2	2.3	18	3.5	5.5
3	0.9	27	3.6	8.2
4	2.8	20	6.4	10.7
5	3.2	20	8.0	10.1
6	1.7	24	4.8	9.5
7	1.1	23	4.2	4.4
8	1.8	26	6.6	3.2
9	1.7	22	4.4	5.1
average	1.8	22	5.1	7.0

was 7.0% on the average (Table II). All of the continua could be fit with considerably less error using all the spectral and physical features and stepwise regression (Table II), and that a least squares fit using the best five features most

closely reproduced the error in the original data (Table II). This was encouraging, so we investigated what each of the five best features was (fig. 1). As one can see, substructural information is important in fitting the odor similarities. But several of the variables, most notably molecular weight, also seem to be very important for fitting eight of the properties. This is worrisome because the nine standards were

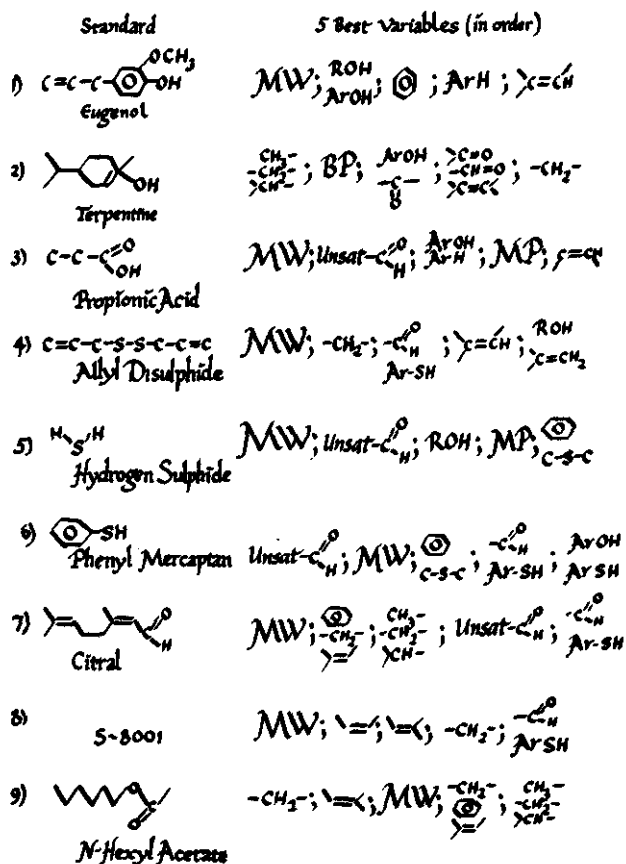


Figure 1. Five most highly correlated features to each standard.

chosen to be as dissimilar as the empirical odor classification scheme allows. Therefore, one would not expect eight of them to correlate. This observation led to the calculation of a correlation matrix for the nine continua (Table III), which showed that eight of the standards are correlated at a moderate to high degree.

TABLE 3
 Correlation of 9 Standards

	1	2	3	4	5	6	7	8	9
1	1.00	-.16	-.54	-.49	-.57	-.57	.51	.52	.43
2		1.00	-.06	-.27	-.27	-.11	-.05	-.17	.03
3			1.00	.53	.66	.50	-.63	-.72	-.59
4				1.00	.88	.73	-.64	-.65	-.65
5					1.00	.88	-.73	-.75	-.73
6						1.00	-.73	-.74	-.70
7							1.00	.74	.68
8								1.00	.63
9									1.00

This in turn led to the investigation of how many independent phenomena were actually being measured by the nine standards. Using principal component analysis (or the eigenvector or Karhunen-Loeve transform) it was found that there were only two significant factors. As these can be used as axes for plotting the data space, a two-dimensional plot (fig. 2) can be constructed which represents 73.4% of the information in the original nine-dimensional data space. One can see that there are chemical trends associated with the axes, but they are hard to define. Fitting the axes with the spectral and physical data (fig. 3) used before does not clarify matters much. The horizontal axis seems related to molecular weight and electron distribution, while the vertical axis seems related to polarizing atoms in various bonding situations, but the results are unclear.

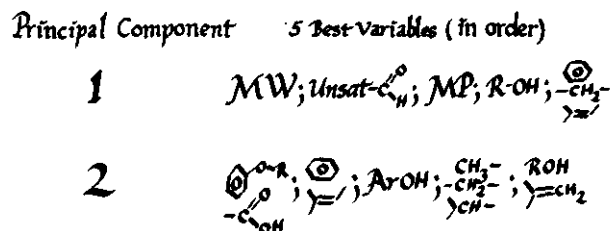


Figure 3. Five most highly correlated features to each principal component.

A new source of more fundamental information about the molecules then being studied seemed necessary. The semiempirical CNDO method of quantum chemical calculation¹⁰ was applied to as many of the members of the data set as possible (the solid points in fig. 2 were calculated) and the resulting information used to elucidate the two axes. The information used included the energy of the top filled and bottom unfilled molecular orbitals, the dipole along

each axis of the molecule, the dimensions of the molecules, and the molecular weight. The differences between the top filled and bottom unfilled molecular orbitals, the ratios of the dipoles along different axes, and the ratios of the dimensions were also calculated and used in fitting the axes. The results were startlingly simple. The horizontal axis is well fit by the molecular weight of the molecules and by information about the dipole of the molecule along various axes, particularly the ratio of the dipole along the longest axis to that along the shortest axis. The vertical axis is fit by information about the energy of the highest occupied and lowest unoccupied molecular orbitals, particularly the difference between them. One would expect this latter value to correlate to the Lewis acidity of a molecule, a quantity not possible to calculate. Note that the actual dimensions of the molecules were never used by the regression analysis to fit either axis.

These results are interesting, but what do they mean and how do they relate to the four major theories of odor sensation? They seem to indicate that the former two theories are focused on the effects rather than the causes of odors, while the latter two correspond well with the results found here. But there are several causes for reservations.

The category of odors with which the shape sensor theory has had the greatest success is the ethereal group. These are nearly spherical atoms with little or no directed dipole which lie in the center of figure 2. The shape theory does acceptably well with camphorous, minty, floral, and musky which lie to the right of center and are all molecules in which the strongest directed dipole of the molecule lies approximately along the long molecular axis. However, no shape sensors have been found for the pungent and putrid odors which, in figure 2, are seen to form the left half of the plot as compounds with strong dipoles perpendicular to the long axis of the molecule. From this it seems that the shape sensor theory succeeds because of the correlation between the directed dipole and the long axis of the molecules rather than because the shape of the molecules is important. This supposition is supported by the fact that the actual dimensions were not used in fitting either of the principal components.

The success of the vibrational coupling theory is quite understandable since Raman absorptions are directly related to the internal dipole of a molecule. The infrared spectral information used as part of the fitting of the nine initial continua was very important to the final fits.

This leaves the neural penetration and acid-base theories, each of which can be rationalized with one of the principal components found in the data space. The acid-base theory is fairly

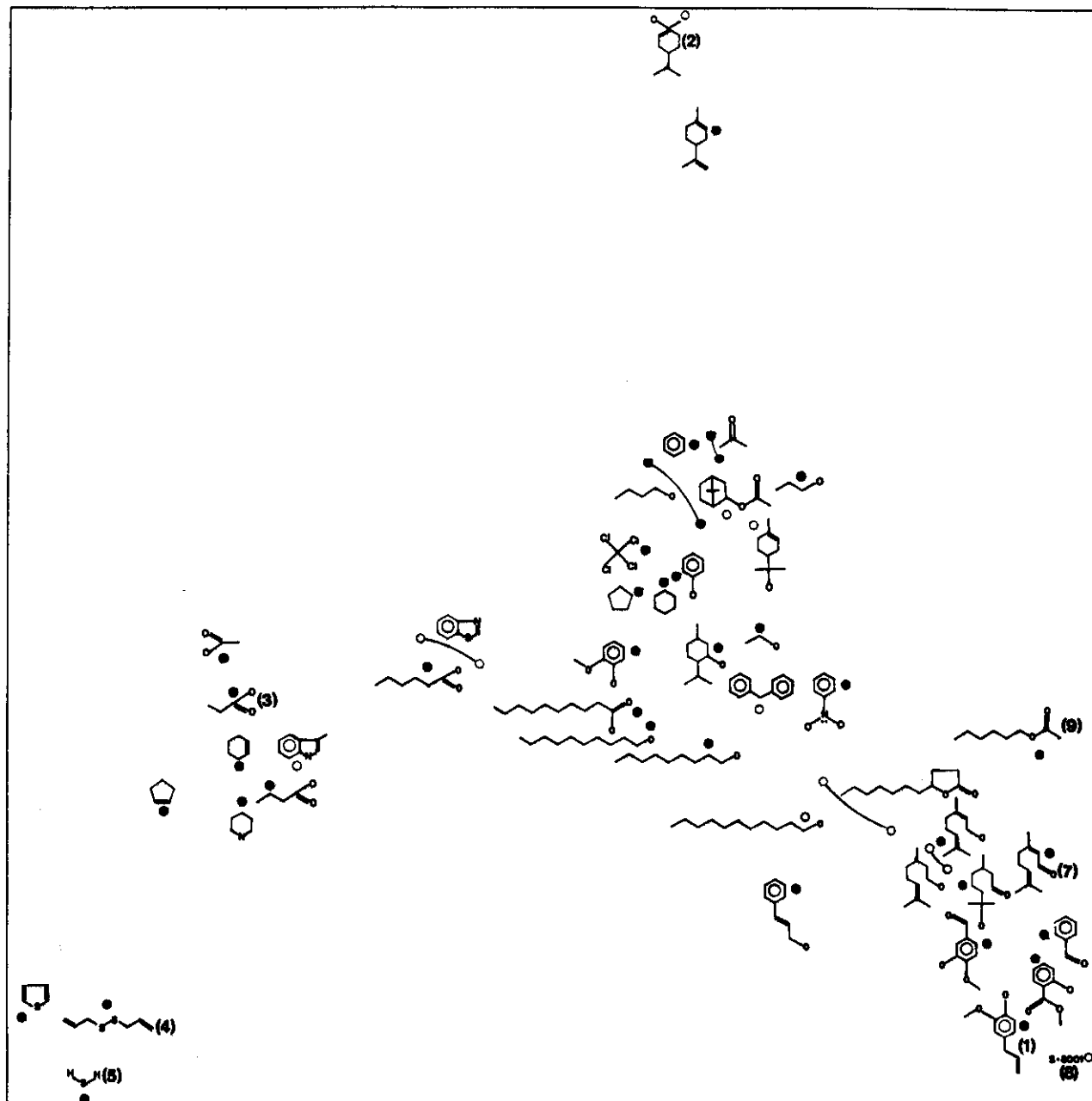


Figure 2. Plot of principal component 1 against 2.

easily connected to the second component (ordinate in fig. 2) which represents the difference between the highest occupied and lowest unoccupied orbitals in the molecule. This difference and Lewis acidity should be highly correlated. Whether the nose is sensing the electron donor-acceptor ability of the molecule or the first excited state is not clear from this work, and remains to be investigated. The neural penetration theory is a bit less easy to rationalize with the first component, but it is possible when one considers that a neural wall is composed of a highly ordered double layer of dipoles with a considerable charge difference across it. One can then theorize that adsorption of a molecule

onto the surface would cause a reordering of the dipoles and a consequent reduction in the wall's insulating ability, causing the neuron to fire. The extent of the breakdown would be determined by how much the adsorbed molecule reordered the surface, which would relate to its dipole strength. Again, this is a speculation and must be studied experimentally.

Finally, it should be noted that the results of this study are no better than the initial information, and that there are several problems both with Wright and Michaels' study and with the information used in this study. The importance of molecular weight in fitting eight of the nine standards and one of the principal components

is curious and, perhaps, an artifact of the set of odorants used. One can see that there are a number of large aromatics that cluster together in the lower right corner of figure 2, thus biasing the data set. Another problem is that eight of the nine standards were highly correlated, so that some of the test compounds did not smell particularly like any of them. These compounds tend to lie in the center of figure 2, which may be why acetone and benzene lie so close together. Also, there was little allowance for the effect of the previous odorant on the next one in the initial study. The series of compounds was presented to the volunteers in one of two orders and no account of the possible interactions seems to have been made. (This may not be an important point, as the variation in the twice tested compounds is at most 11%, quite good for a subjective testing scheme.) Finally, both the spectral and physical measurements used in the fitting of the nine continua were from the literature, rather than taken directly on the compounds used, which increases the possibility of error.

Thus this work suggests that, from this data set, the human odor sensory spectrum can be explained by the existence of a two-dimensional continuum with the coordinates (odor) of a molecule determined by its directed dipole and electron donor-acceptor ability. This result is pleasantly straightforward and makes a trained expert's ability to distinguish thousands of different odors reasonable. The simplicity is really not that disconcerting when one considers that all of color vision is performed by only three sensors.

Acknowledgements

The author would like to thank Dr. V. Schoemaker for his many helpful discussions, and M. daKoven for his aid.

Tables I, II, and III, and figure 2 are reprinted with permission from *Analytical Chemistry* **49**, 596, 1977, and are copyrighted by the American Chemical Society.

References

1. Zwaardemaker, H., *Die Psychologie des Geruchs*. Englemann, Leipzig, 1895.
2. Amoore, J. E., *Molecular Basis of Odor*. C. C Thomas Co., Springfield, IL, 1970.
3. Dyson, G. M., *Chemistry and industry*, 1938, **647**, 1938.
4. Wright, R. H., *The Science of Smell*. Basic Books, New York, 1964.
5. Davies, J. T., *Olfactory theories in Handbook of Sensory Phenomena IV Chemical Senses I. Olfaction*, L. M. Beidler, ed. Springer-Verlag, New York, 1971.
6. Brower, K. R., and Schafer, R., *J. Chem. Ed.* **52**, 538, 1975.
7. Wright, R. H., and K. M. Michaels, *Annals N.Y. Acad. Sci.*, **116**, 535, 1964.
8. Kowalski, B. R., *Pattern recognition in chemical research*, in *Computers in Chemical and Biochemical Research Vol 2*, C. E. Klopfenstein and C. L. Wilkins, eds. Acad. Press, New York, 1974.
9. McGill, J. R., and Kowalski, B. R., *Anal. Chem.*, **49**, 596, 1977.
10. Pople, J. A., and Beverige, D. L., *Approximate Molecular Orbital Theory*. McGraw-Hill, New York, 1970.