Refining of Methods for Subjective–Objective Evaluation of Flavor

John J. Powers* and Mary C. Quinlan

An experimental blueberry-whey beverage containing different amounts of blueberry pulp, five brands of bourbon, and four brands of canned freestone peaches were evaluated subjectively and by measurement of chemical or physical attributes. A new computer program was applied to the sensory scores to construct contingency tables to estimate the likelihood that the panel was truly showing a preference in its scoring rather than that the assignment of scores was approaching randomness, to determine the distance between score levels, and to seek out reversal of score levels. Correlations between each judge and all other judges were calculated as to acceptability judgments, and the panel was divided into subsets according to judges having like preferences. For each of the products, the multiple and simple correlations between acceptability and color, appearance, flavor, and texture (or mouthfeel, as appropriate) were calculated. Application of discriminant analysis to gas–liquid chromatographic (glc) peak areas showed a correlation between the sensory scores for flavor and the glc patterns. Objective measurements for color, flavor, and texture could be correlated with overall acceptability only by circuitous routes. One procedure was to differentiate among some of the products with one formula, then use a second formula to effect further differentiation. A second procedure was to form ratios between some of the measurement values or to form sums of them, and then to use these ratios and sums in place of the original measurement values in the discriminant analysis. Neither of these trial-and-error procedures was really satisfactory, and they point out the need for a better understanding of the fundamental principles or relations among sensory factors and between sensory and physicochemical factors.

Within the past 8 to 10 years, there has been intensification of effort to develop multivariate means of correlating flavor with chemical composition. The literature has been cited by Powers and Keith (1968), Dravnieks and Krotoszynski (1968), Powers (1968, 1970), Pattee and Singleton (1972), Persson and von Sydow (1973), Kosaric *et al.* (1973), Dravnieks *et al.* (1973), Powers and Quinlan (1974), and Quinlan *et al.* (1974). Programs for computer analysis have likewise been developed or cited by the authors listed above.

Methods for subjective–objective correlation of color or texture are much older, having been in practical use in the food industry for at least 30 years, some for as many as 50 years. Objective procedures for color and texture have one great advantage as compared with flavor. In general, the same force or property is measured by the objective means as humans use in sensing. The eye responds to wavelength and intensity—and such other factors as surface character—and the brain integrates the various signals. In part, spectrophotometers do the same. In chewing foods, we make judgments about shear force, compression, particle size and hardness, melting points, viscosity, and other physical attributes. Again, there are instruments to measure the same force or property. Comparisons cease when one comes to flavor. With rare exceptions, we do not know the properties of a compound which make a particular substance a flavor or taste material. As a consequence, we have to deal with flavor substances indirectly by measuring some property other than flavor. The best we can expect to derive is a correlation between flavor and flavor compounds.

From a practical point of view, statistical significance is not enough. A correlation has to be exceptionally great if it is to be usable. In determining sensory quality and in carrying on chemical determinations, there is error on both sides. In essence, the two errors must be as low as possible if practical correlations are to be obtained.

Before the advent of computers, only relatively simple correlations were generally sought for because of the difficulties of computation. Today computations no longer hinder us as much as does our own lack of knowledge about the interplay of chemicals as they affect sensory response.

One purpose of this study was to apply a new method of analysis to sensory data to determine whether error in sensory analysis could be reduced. The second was to ascertain the importance of flavor relative to color and texture in determining acceptability, and the third was to apply statistical methods previously used to correlate flavor with glc patterns to a wider range of objective values to learn if acceptability could be correlated with chemical or physical measurement values for color, flavor, and texture.

EXPERIMENTAL SECTION

Sensory Trials. Three different products were used. One was an experimental blueberry-whey beverage containing 19, 26, 33, and 40% blueberry pulp. The panel consisted of 32 judges who evaluated the products hedonically for overall acceptability, appearance, flavor, and mouthfeel. The trials were replicated six times.

Five commercial brands of bourbon whiskey were used for the second set of trials. The brands sold at the retail level for 15, 17, 21, 26, and 30 cents/oz. To simulate "sipping" conditions, the bourbons were diluted 1:1 with distilled water to compensate for dilution by melting ice. The five diluted samples were packed in vials containing 15 ml. The panelists judged the samples at home, using disposable shot glasses as containers, and the bourbons were evaluated both at room temperature and cold. On the days they were to be evaluated cold, the panelists were instructed to submerge the vials in ice-water for at least 0.5 hr. Because whiskey is normally drunk chilled, we wished to observe the effect of temperature on acceptability and discrimination. There were five replications each for the warm and cold sampling trials. The products were evaluated hedonically for acceptability, odor, flavor, and mouthfeel.

There were 13 panelists, ranging in age from 24 to 55, with about one-third of the panel being women. Except for one panelist who had experience judging brandy in

Department of Food Science, University of Georgia, Athens, Georgia 30602.

Europe, none had any particular experience judging alcoholic beverages except as occasional social drinkers.

The third product consisted of commercially canned peaches, again selected according to price. The selling prices per No. 2½ can were 33, 37, 49, and 53 cents. The products were likewise evaluated hedonically for acceptability, color, appearance, flavor, and texture. The trials were replicated six times. There were 30 judges.

For all of the trials, the samples were randomly color coded each day. The panelists were instructed to sample from left to right. The position of any given sample was randomly determined for each session. Color coding facilitated coding, de-coding, and record keeping. The various hedonic descriptions were converted into numerical scores of 5–1, 5 being the best and 1 being the poorest.

Objective Tests. The blueberry-whey is part of another project and details as to objective tests will be given elsewhere. The bourbons and peaches were analyzed gas chromatographically. Other objective measurements were also made on the peaches. They were: pH, titratable acidity, relative conductivity of the packing syrup, drained weight, relative viscosity of the packing syrup, absorbance of a 1-butanol extract of the peaches at 400 and 660 nm, force required to shear the peaches (g/100 g), and work required to shear the peaches (area of the time–force curve).

The bourbon was prepared for glc analysis by mixing 50 ml of bourbon with 50 ml of H_2O and 2 drops of H_2SO_4. The mixture was stirred vigorously for 5 min in a 250-ml erlenmeyer flask, using a magnetic stirrer. To the mixture was then added 40 ml of anhydrous ether and 10 ml of *n*-pentane. Stirring was continued for 15 min. The mixture was next poured into a separatory funnel and allowed to stand for 5–7 min. The lower layer was drawn off into a 200-ml erlenmeyer flask containing 26 ml of anhydrous ether and 7 ml of *n*-pentane. After stirring vigorously for 15 min, the mixture was poured into a separatory funnel, the phases were allowed to separate, and the lower layer was discarded. The upper layer was combined with the upper layer from the first extraction. The extract was then stored in a freezer at −34.4° to freeze out water. The organic phase was decanted while still at −34.4° to eliminate the water (ice).

The peaches were extracted by the Likens–Nickerson procedure as modified by Young *et al.* (1970). Two hundred grams of peaches were used.

Both the bourbon extract and the peach extracts were concentrated in a Kuderna–Danish condensor (Milutinovic *et al.,* 1970). To concentrate the peach extracts, the water bath was held at 42° and for the bourbon extracts it was held at 75°. The high temperature for concentration of the bourbon extracts was to eliminate much of the ethanol. Lower boiling components were naturally also lost.

The glc column for the bourbon and peach analyses was a 2.74 m × 6.25 mm o.d. stainless steel tubing packed with 5% SP-1000 on 60–80 mesh Supelcoport. The chromatograph was a dual-column instrument, programmed from 50 to 200° at 6°/min. A second instrument with a nonpolar column was also used to permit tentative identification of some of the bourbon peaks by comparison of the retention times for the bourbon components on the two columns. The column was a 2.29 m × 6.25 mm o.d. stainless steel tubing packed with 7% Dexsil on Chromosorb acid-washed dimethylchlorosilane, 60–80 mesh. Programming was from 80 to 225° at 9°/min.

Peak areas were electronically integrated on the first instrument and trigometrically calculated for the second to avoid conflicts in the use of the integrator. The peak areas were converted into percentage areas by dividing the total area under the chromatogram into each peak area (Powers and Keith, 1968). For the bourbon samples, 2-octanol was added as a reference compound to the concentrate. Peak areas were also then calculated relative to the area of the 2-octanol peak.

Statistical Analysis. The sensory data and the objective measurements were subjected to statistical analysis, using either the MUDAID program (Applebaum and Bargmann, 1967) or the PREPRO program (Kundert and Bargmann, 1972).

The normal order of analysis was to use the MUDAID program to run a one-way analysis of variance with the judges as variables to detect those who were nondiscriminating or inconsistent. Once the acceptable judges had been selected, the PREPRO program was then used to construct a contingency table for the scores *vs.* the products. The program feeds back the same values to construct a second contingency table with the scores assigned at random to the products but with the column totals kept the same. The correlation coefficients are printed out for each contingency table. If the panel assignment is only slightly better correlated than the random assignment, there is no justification for further analysis. One has already learned that the panelists have no preference or are inconsistent judges. If the probability level for the correlation coefficient is significant, then one knows that the panelists were almost certainly selecting the products by preference.

The PREPRO program performs a second function. It endeavors to transform the scores so that there is a maximum difference between products.

Once the scores had been re-scaled, the transformed values were then subjected to analysis of variance (using the MUDAID program) to calculate product means, the correlation coefficients between judges, and the correlation coefficients between the various sensory categories.

The next thing that was done was to examine the correlation coefficients between judges to partition them into sets of judges whose preferences were the same or closely alike.

The glc and the other objective measurements were analyzed by the MUDAID program as previously described (Young *et al.,* 1970; Powers *et al.,* 1971; Powers and Quinlan, 1974; Quinlan *et al.,* 1974).

RESULTS AND DISCUSSION

In each of the three trials, there were negative judge–judge correlations which had to be taken into account. Establishment of differences in foods is comparatively simple. Within the limits of their sensitivities, the judges should agree as to whether there is or is not a difference. When acceptability is involved, a second factor comes into play. Judges may agree there is a difference, but they may not necessarily prefer the same sample. In acceptability testing, significant differences among samples may be overlooked because of failure to realize that the panelists often fall into subsets. Normally, once a panelist has been checked out as to discriminating power and sensitivity, the results of that judge and all others are pooled. If an appreciable number of panelists disagree as to which sample is preferred, their scores may cancel out each other, and the conclusion might be that there is no difference.

Table I shows the correlation coefficients among some of the 32 judges involved in the blueberry-whey trials. Of the 32 judges, 13 were acceptable. The complete set of correlation coefficients is not shown so as to simplify the table. Note that the judges fit into two subsets. Judges 5, 9, 10, 11, and 28 had like preferences. Judges 3, 12, and 25 had different preferences. The means for the two sets are listed in Table II. If judges 3, 12, and 25 are pooled with the other judges, sight is lost of the fact that an appreciable percentage of the panelists had preferences which differed from the grand means. There is naturally at least one more subset, those who cannot discriminate or, in practical terms, have no preference. There may be more than three subsets. Unlike Table I where only some of the judges were listed, three or more subsets having like pref-

Table I. Judge–Judge Correlation Coefficients and Subsets with Like Preferences[a]

Judges	5	9	10	11	12	25	28
3	−0.36	−0.33	0.05	−0.37	0.35	0.30	−0.17
5		0.36	0.42	0.23	−0.58	−0.53	0.67
9			0.00	0.21	−0.17	−0.43	0.34
10				0.16	−0.15	−0.12	0.29
11					−0.26	−0.28	0.25
12						0.47	−0.55
25							−0.70

[a] Subset A: 5, 9, 10, 11, 28. Subset B: 3, 12, 25.

Table II. Acceptability of Blueberry-Whey Beverage to Subsets A and B

Judges	Level of blueberry pulp 19%	26%	33%	40%
5	1.67	1.50	3.33	3.50
9	2.50	2.83	3.67	4.00
10	3.33	3.83	4.50	3.67
11	1.67	2.50	3.33	3.00
28	2.33	2.50	3.33	4.33
Mean	2.30	2.63	3.63	3.70
3	2.67	2.50	2.00	1.67
12	2.83	3.33	1.83	1.67
25	2.67	2.50	1.83	1.50
Mean	2.72	2.78	1.89	1.61
Grand mean	2.46	2.67	2.98	2.92

erences within each set but different between sets may become evident if the number of judges is large.

The results showed that judges should not necessarily be treated *en masse*. Judge–judge correlations should be inspected and, if necessary, the judges should be subdivided into sets. The judges who cannot discriminate are not without value. They give a rough estimate of the per cent of the population which would find the difference below their levels of detectability.

A judge who was acceptable for one sensory task was not necessarily equally satisfactory for another. This observation of course is nothing new; we merely wish to reaffirm that judges should be selected on the basis of every task they have to perform. Checking them out for their suitability to evaluate acceptability, or flavor, or some one of the other qualities is not enough. Their performance of each task should be evaluated in selecting them.

A second problem in analyzing sensory results is that judges often do not maintain the same interval between score levels. In hedonic testing, the fault may be in the formulations of the descriptive terms. When direct scoring is used, judges sometimes still do not maintain the same degree of difference between score levels.

Table III shows a contingency table for the flavor scores of the blueberry-whey beverage. Listed below the table are the transformed scores for the products and the intervals between scores. The transformation alerts one to the fact that differences in scores should not all be treated alike. On the original basis, the difference between score levels 2 and 3 has a different meaning than between 3 and 4.

In subsequent analyses, the transformed values should be used to correct for fault in formulation of the descriptions or the panelists' memory-estimation faculties.

Sometimes an out-and-out reversal of score levels takes place (Table IV). The score levels should be compressed by combining the two levels reversed, levels 2 and 3 in this case, because reversal is an indication that the panel cannot discriminate in this particular range.

A third thing one obtains from the PREPRO program is a comparison of panel assignment of scores and a random assignment. Note the "body" category in Table V. The correlation coefficient for the panel's assignment of scores is only slightly better than a random assignment. This means the panelists had no real preference. For acceptability, appearance, and flavor, the panel had clearcut preferences.

The refinements attained as compared with the ordinary handling of sensory data are: (1) subsectioning of panelists into compatible sets to determine the judges whose preferences do not agree with the majority, and, with larger panels, to estimate the percentage of the population whose preferences might be different from the panel means; (2) estimating the distances between scoring levels or the even greater aberration of reversal of score levels; (3) interpreting the degree of randomness in the panel's assignment of score levels.

Each of the analyses is preliminary to permit treatment of the data according to the outcome of the MUDAID program for judge–judge correlations and the PREPRO program for randomness, unequal scale differences, or reversal of score levels.

The bourbon and the peach trials were carried on so as to try the methods on commercial samples. The real test is not whether one can detect significant correlation over any group of samples but whether one can do so within the range of commercial acceptability or a part of that range. The narrower the range, the more difficult it is to get significant correlations. Bourbon was used for this rea-

Table III. Contingency Table for Flavor Scores, Blueberry-Whey Beverage

Level blueberry, %	Flavor score levels 1	2	3	4	5	Total
19	6	12	13	4	1	35
26	3	9	13	11	0	36
33	0	3	12	17	4	36
40	0	3	11	15	7	36
	9	27	49	47	12	144

Original level	1		2		3		4		5
Transformed values	−2.22		−1.17		−0.12		0.82		1.60
Interval between		1.05		1.05		0.94		0.78	

Table IV. Reversal of Score Position, Blueberry-Whey Acceptability Scores

Original score	Transformed score
1	−0.959
2	−0.445
3	−0.800
4	1.096
5	1.970

Table V. Blueberry-Whey Beverage, Contingency Table Correlation

	Panel		Random	
	r	Probability level	r	Probability Level
Acceptability	0.427	1.26×10^{-3}	0.176	0.893
Appearance	0.533	7.93×10^{-6}	0.164	0.959
Flavor	0.492	1.21×10^{-4}	0.284	0.323
Body	0.268	0.313	0.220	0.663

son and to determine whether there was a relation between price and preference. Out of the 13 panelists, only 6 were statistically acceptable. Table VI shows Duncan's multiple range test applied to the five brands. The two highest priced products were preferred but in some categories or under some conditions, the cheapest product was judged equally desirable. The correlation coefficients for the various multiple and simple correlations are listed in Table VII. When the bourbon was cold, odor and flavor were negatively correlated with acceptability. When the bourbon was warm, flavor and mouthfeel were just as useful for predicting acceptability as odor, flavor, and mouthfeel.

Four of the six judges had like preferences; thus the results of Table VI were determined largely by their preferences. The chief observation made was that price was not highly correlated with preference for the five brands selected, and, for the 13 judges, there was no relationship at all. The correlation coefficients of Table VII indicate that smoothness (mouthfeel) was not as important when the bourbon was cold as when it was warm.

Figure 1 shows the relation of the brands to the weighted means for the most discriminating glc peaks. Twelve peaks were necessary to effect discrimination among the bourbons, but even then the HH and KB brands could not be separated. A second discriminant equation was calculated between brands HH and KB only, using a different selection of 12 peaks and the reference area (2-octanol) as a base. Discrimination then could easily be effected between brand HH and KB. Brand HH had a weighted mean of 1.92 whereas the mean for sample KB was 0.16.

Of the five bourbons examined, casual inspection of the chromatograms suggested that two or three of the bourbons could be identified from a few peaks. Statistical analysis indicated otherwise. We sense certain chemicals by olfaction and gustation and we integrate our response to them. Not until the cumulative difference of 12 or more peaks had been summed did organoleptic judgment and glc data correlate.

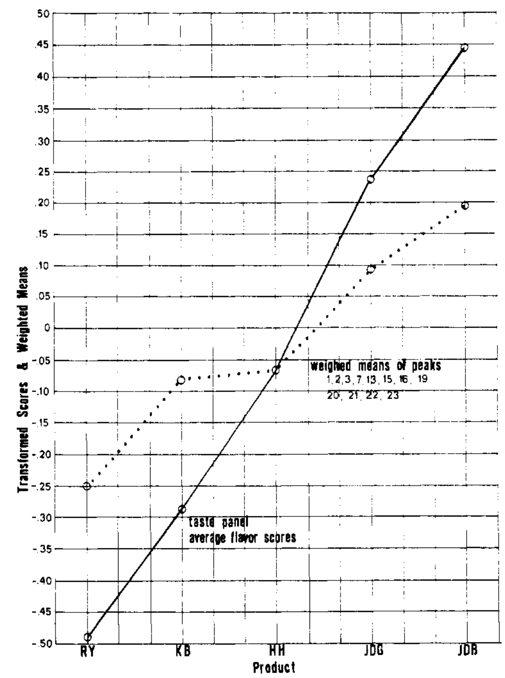

Figure 1. Relation of panel scores for the five brands of bourbon and the weighted means of 12 glc peak areas. The solid line shows the flavor scores; the dotted line, the weighted means.

After it was seen that the panelists had no strong preference, identification studies were discontinued. About 80 compounds have been identified in whiskies of various types (Jones and Wills, 1966; deBecze *et al.*, 1967; Kahn *et al.*, 1968; Nykanen *et al.*, 1968; Kahn, 1969; Kahn *et al.*, 1969; Suomalainen and Nykanen, 1966; Schoeneman *et al.*, 1971; Nishimura and Masuda, 1971; Jennings *et al.*, 1972.

The trials with the canned peaches were the most consistent on the organoleptic side. The panel clearly differentiated the various brands of peaches in all categories. This is shown in Table VIII. No individual objective measurement was well correlated with sensory quality. Viscosity is determined chiefly by the strength of the packing syrup and drained weight is influenced by it. The maturity of the fruit also affects the drained weight. The color and the shear force measurements are likewise affected by the maturity. Conductivity was run purely on the chance that there might be some relationship to electrolyte materials. None of the objective measurements lined up in the same order as the organoleptic scores, primarily because the OS brand was over-mature and showed poor workmanship. It was downgraded in all categories. The MG brand was somewhat green, so the panel downgraded it in appearance and texture but for different reasons.

With respect to the glc phase, the organoleptic ratings for flavor correlated with the weighed sum of three of the glc peaks (Figure 2). The organoleptic difference between the second-ranked and the third-ranked brand was not great. As with the bourbon, a second formula could have been calculated to widen the interval between brands D

Table VI. Duncan's Multiple Range Test Applied to Bourbon Samples[a]

	Cold					Warm				
Acceptability	JDG	JDB	HH	RY	KB	JDB	JDG	HH	KB	RY
Flavor	JDB	JDG	HH	KB	RY	JDB	JDG	HH	KB	RY
Odor	JDG	JDB	RY	KB	HH	JDG	JDB	HH	KB	RY
Mouthfeel	JDG	JDB	HH	KB	RY	JDB	JDG	RY	KB	HH

[a] The prices per ounce for brands JDB, JDG, HH, RY, and KB were, respectively, 30, 26, 15, 21, and 17 cents.

Table VII. Correlation Coefficients for Sensory Properties of Bourbon

	Sampled warm Simple corr coeff		
	Odor	Flavor	Mouthfeel
Acceptability	0.59	0.67	0.62
Odor		0.67	0.28
Flavor			0.30
	Sampled cold		
	Odor	Flavor	Mouthfeel
Acceptability	−0.39	−0.58	0.48
Odor		0.18	−0.07
Flavor			−0.47

	Multiple corr coeff	
	Warm	Cold
Acceptability *vs.* odor–flavor–mouthfeel	0.78	0.79
Acceptability *vs.* odor–flavor	0.65	0.63
Acceptability *vs.* odor–mouthfeel	0.72	0.61
Acceptability *vs.* flavor–mouthfeel	0.77	0.61

Table VIII. Duncan's Multiple Range Test Applied to Peach Measurements

	Sensory evaluation[a]			
Acceptability	K	D	MG	OS
Color	K	D	MG	OS
Appearance	K	D	OS	MG
Flavor	K	D	MG	OS
Texture	K	D	OS	MG
	Objective measurements			
Viscosity	MG	OS	K	D
Drained weight	D	MG	K	OS
Soluble solids	K	OS	D	MG
Absorbance at 660 nm	K	D	OS	MG
Absorbance at 400 nm	MG	D	OS	K
Shear force, peak height	MG	K	OS	D
Shear force, work	K	OS	D	MG
Conductivity	MG	D	K	OS

[a] The prices for the K, D, MG, and OS brands were, respectively, 49, 53, 37, and 33 cents per No. $2^1/_2$ can.

and MG. Since the sensory scores were fairly close together, there was no need to expand the objective scale.

The chief effort was devoted to attempting to develop an equation which correlated with acceptability rather than flavor alone. None of the chemical or physical measurements for color or texture, coupled with the glc peak areas, yielded good correlation for all four samples. Depending upon the particular combination of measurement used, either the MG or the OS brand was out of order. As explained above, both brands were rated low in texture, but for different reasons, and the OS brand was rated especially low in appearance. The shear force measurement accounted for texture in part, but there was no objective test for appearance. Its role in acceptability was thus unaccounted for.

Two approaches were taken to solving the problem. One was the same as that used for bourbon. First, a general equation was calculated for the four samples; then a second equation was calculated using a different set of variables. The first equation permitted one of the samples to be placed in a class by itself; the second equation permitted the order of the other three samples to be determined. In that way, the weighted means of objective tests for color, flavor, and texture could be correlated with the panel scores for acceptability.

Table IX. Correlation Coefficients for Peaches

	Color	Appearance	Flavor	Texture
Acceptability	0.419	0.388	0.734	0.544
Color		0.582	0.362	0.379
Appearance			0.253	0.370
Flavor				0.441

Acceptability *vs.* flavor and texture	0.787
Acceptability *vs.* flavor and color	0.752
Acceptability *vs.* flavor and appearance	0.763
Acceptability *vs.* texture and color	0.590
Acceptability *vs.* texture and appearance	0.580
Acceptability *vs.* appearance and color	0.455
Acceptability *vs.* appearance, flavor, and texture	0.787
Acceptability *vs.* color, flavor, and texture	0.781
Acceptability *vs.* color, appearance, flavor, and texture	0.788

Table X. Transformed Values and Correlation Coefficients for Peach Products

	Products			
	MG	OS	K	D
Acceptability	−0.192	−1.671	0.924	0.562
Color	−0.718	−1.066	1.509	0.286
Appearance	−1.048	−0.791	1.460	0.391
Flavor	−0.294	−1.688	0.901	0.499
Texture	−1.041	−0.871	1.357	0.567
	Panel		**Random**	
	r	Prob. level	r	Prob. level
Acceptability	0.64	1×10^{-15}	0.15	0.384
Color	0.33	3.7×10^{-10}	0.10	0.675
Appearance	0.29	1.04×10^{-5}	0.14	0.178
Flavor	0.64	1×10^{-15}	0.11	0.821
Texture	0.36	2.1×10^{-11}	0.12	0.496

Another mode of attack was to form the ratios of some of the measurement values (the two absorbancy values, for example) and the sum of certain glc peak areas. By combining ratios and sums, differentiation could likewise be made.

The fact that empirical means had to be resorted to, points out the major problem involved in correlating subjective with objective estimates of quality. If the underlying principles or relations are known, applications are more likely to be logical than mere trial-and-error methods, as were used to correlate acceptability with the objective measurements. The objective methods chosen to measure color and texture of the peaches apparently were insufficient, and no objective value was available for appearance; yet the latter was important. It had approximately the same weight as color and texture (see Tables IX and X). Until recently, flavor could not be included in objective grading systems because almost always flavor could be judged only by subjective means. Quinlan *et al.* (1974) pointed out that flavor should be given greater attention in grading systems because it is so important organoleptically. Procedures such as those used here and by others (see the citations in the introductory paragraph) go part way toward solving the problem of being able to use objective measurement values.

Most of the procedures have not yet resolved the problem of the objective values being weighed in approximately the same way as humans subconsciously do in assessing acceptability. Flavor, color, and all the other attributes are appraised in arriving at a decision as to acceptability. The relations of each sensory quality to acceptability and to each other need to be more rigorously defined. Tables VII and IX list these relationships for the bourbon and the peaches. If the relative contribution of each sensory quality were known, then it would be easier to develop objective means which approximate sensory judgment.

Table X shows the distance between the various score levels for each of the quality factors of the canned peaches and the correlation coefficients for panel and random assignment. The correlation coefficients of Table X could merely mean that the panel was most proficient at judging flavor, but those of Table IX indicate that flavor strongly influenced acceptability. Of the various multiple correlations, they were higher when flavor was one of the components than when it was absent.

Studies such as this one provide in a numerical fashion evidence of the importance of flavor, texture, and the other attributes to overall acceptability. Further studies are needed to establish the correlations between each sensory factor and the weight each factor has in determining acceptability. This will have to be done on a commodity basis because the relative importance of each sensory quality varies according to the commodity. Until the correlations and weights are known, objective means of assessing quality will be handicapped. Correlation of objective measurements with quality is already practical, but for major advances to be made, greater fundamental knowledge has to be obtained on both sides of the coin: sensory and objective.

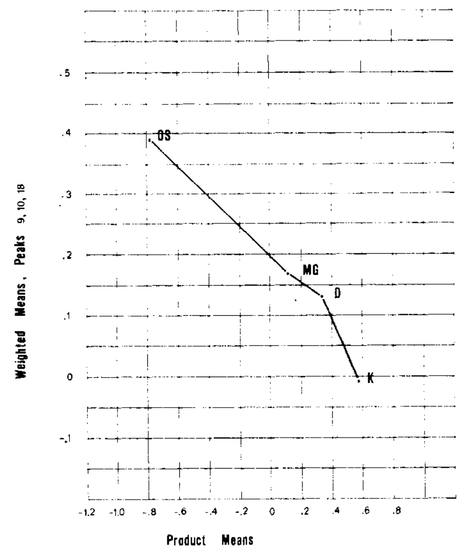

Figure 2. Relation of scores for the four brands and the weighted means for glc peak areas. The product scores are shown on the ordinate.

LITERATURE CITED

Applebaum, M., Bargmann, R. E., Technical Report No. NONR 1834 (39), University of Illinois, Urbana (1967).
deBecze, G. I., Smith, H. F., Vaugh, T. E., *J. Ass. Offic. Anal. Chem.* **50**, 311 (1967).
Dravnieks, A., Krotoszynski, B. K., *J. Gas Chromatogr.* **6**, 144 (1968).
Dravnieks, A., Reilich, H. G., Whitfield, J., Watson, C. A., *J. Food Sci.* **38**, 34 (1973).
Jennings, W. G., Wohleb, R., Lewis, M. J., *J. Food Sci.* **37**, 69 (1972).
Jones, K., Wills, R., *J. Inst. Brew., London* **72**, 196 (1966).
Kahn, J. H., *J. Ass. Offic. Anal. Chem.* **52**, 1166 (1969).
Kahn, J. H., Shipley, P. A., LaRoe, E. G., Conner, H. A., *J. Food Sci.* **34**, 587 (1969).
Kahn, J. H., Trent, F. M., Shipley, P. A., Vordenberg, R. A., *J. Ass. Offic. Anal. Chem.* **51**, 1330 (1968).
Kosaric, N., Duong, T. B., Surcek, W. J., *J. Food Sci.* **38**, 369 (1973).
Kundert, K. R., Bargmann, R. E., Themis Report No. 22, Department of Statistics and Computer Science, University of Georgia, Athens, Ga., 1972.
Milutinovic, L., Bargmann, R. E., Chang, K. Y., Chastain, M., Powers, J. J., *J. Food Sci.* **35**, 224 (1970).
Nishimura, K., Masuda, M., *J. Food Sci.* **36**, 819 (1971).
Nykanen, L., Puputti, E., Suomalainen, H., *J. Food Sci.* **33**, 88 (1968).
Pattee, H. E., Singleton, J. A., *J. Agr. Food Chem.* **20**, 1119 (1972).
Persson, T., von Sydow, E., *J. Food Sci.* **38**, 377 (1973).
Powers, J. J., *Food Technol.* **22**, 383 (1968).
Powers, J. J., *Proc., Int. Congr. Food Sci. Technol., 3rd*, 395 (1970).
Powers, J. J., Cox, R. J., Eiland, M. C., Ackroyd-Kelly, P., *Flavor Ind.* **2**, 87 (1971).
Powers, J. J., Keith, E. S., *J. Food Sci.* **33**, 207 (1968).
Powers, J. J., Quinlan, M. C., *Lebensm. Wiss. Technol.* in press (1974).
Quinlan, M. C., Bargmann, R. E., El-Galalli, Y. M., Powers, J. J., *J. Food Sci.* in press (1974).
Schoeneman, R. L., Dyer, R. H., Earl, E. M., *J. Ass. Offic. Anal. Chem.* **54**, 1247 (1971).
Suomalainen, H., Nykanen, L., *J. Inst. Brew. London*, **72**, 469 (1966).
Young, L. L., Bargmann, R. E., Powers, J. J., *J. Food Sci.* **35**, 219 (1970).

Received for review October 15, 1973. Accepted December 5, 1973. Paper presented at the Symposium Computers in Flavor Chemistry, 166th National Meeting of the American Chemical Society, Chicago, Ill., Aug 27–31, 1973.