Sensory Analysis

Subcommittee Members: M. C. Meilgaard, *Chairman;* N. M. deBanchs, R. Berndt, L. Bernstein, E. F. Blaschke, A. M. Canales, B. J. Clarke, A. Dravnieks, T. J. Konis, J. W. Larson, C. Latoni, M. Lepage, M. Moll (*EBC*), J. E. Muller, R. M. Pangborn, F. Pastrana, H. B. Pfenninger, P. A. Prell, J. J. Powers, D. J. Schmitt, J. Seigel, D. Simpson, J. C. Sonnemann, P. H. Todd, Jr., R. S. Williams, K. Word, and S. R. Palamand (*ex officio*)

Key words: *Ascending method of limits test, Flavor reference standards, Paired comparison test, Threshold values.*

CONCLUSIONS

Although 21 flavor reference standards are now available, certain key flavor terms still require a standard before the subcommittee's work on flavor terminology can be terminated.

The subcommittee has determined that the original Paired Comparison Test is unsuitable for detecting whether a flavor difference exists between two samples. Two different forms of the Paired Comparison Test are now recommended: the Directional Difference Test and the Paired Preference Test. This study was done in collaboration with the International Standards Organization (ISO), the American Society for Testing and Materials (ASTM), the Institute of Food Technologists (IFT) and the European Brewery Convention (EBC). New tables were introduced to replace those of Bengtsson (4).

The ASTM Ascending Method of Limits is the recommended method for determining the threshold of added substances. The simpler method known as ISO Draft 3972 (6) may be used for certain well-defined conditions, but not in research meant for publication.

RECOMMENDATIONS

1. Adopt as international flavor reference standards the compounds listed in Table I, which includes the previously accepted standards.
2. Adopt the Paired Comparison Test and the Threshold of Added Substances: Ascending Method of Limits Test for inclusion in ASBC "Methods of Analysis."
3. Complete the study of flavor reference standards and publish the detailed procedure of preparation for each standard.
4. In collaboration with the organizations mentioned above, continue the study of methods of sensory analysis, notably the Triangular Test, and complete the "Methods of Analysis" section General Methodology and Choice of Methods.
5. As soon as the Triangular Test can be issued, discontinue the current BEER-30, Testing for Taste Difference between Two Beers.

0361-0470/80/03009909/$03.00/0

As in previous years, this subcommittee is charged with developing a common system of flavor terminology, with

TABLE I
Compounds Recommended for Use as Flavor Reference Standards

Term		Compound	Supplier	Method of Purification	Difference Threshold[a]	In Beer Containing
0110	Alcoholic	Ethanol	High-quality vodka[b]	None required	17 g/L	33–42 g/L
0111	Spicy	Eugenol	Aldrich	Solvent wash + fractional distillation + adsorption	40 μg/L	···
0131	Isoamyl acetate	Isoamyl acetate	Aldrich	Adsorption + GC[c]	0.5–1.7 mg/L	1–3 mg/L
0132	Ethyl hexanoate	Ethyl hexanoate	K & K Laboratories	Adsorption + GC	0.15–0.25 mg/L	0.2–0.4 mg/L
0133	Ethyl acetate	Ethyl acetate	Fluka	Adsorption	20–40 mg/L	10–30 mg/L
0145	Melony	Melonal®[c]	Givaudan	None required	1 μg/L	···
0150	Acetaldehyde	Acetaldehyde	Merck	Adsorption + distillation + adsorption	10–20 mg/L	2–10 mg/L
0173	Hop oil	Cluster hop oil[d]	S. S. Steiner	None required	0.1 mg/L	···
0224	Almond	Benzaldehyde	Aldrich	None required	1 mg/L	···
0611	Caprylic	Octanoic acid	Sigma	Recrystallization of calcium salt	5–10 mg/L	2–8 mg/L
0613	Isovaleric	Isovaleric acid	Sigma	None required	0.5–1.5 mg/L	0.5–1.5 mg/L
0614	Butyric	Butyric acid	Merck/Schuckard	2 × Fractional distillation	2–3 mg/L	0.5–1.5 mg/L
0620	Diacetyl	2,3-Butanedione	Aldrich	Fractional distillation + adsorption	0.07–0.15 mg/L	0.03–0.3 mg/L
0710	Sulfitic	Sodium meta-bisulfite	Fisher Scientific	None required	20 mg/L SO_2	1–10 mg/L SO_2
0732	DMS	Dimethyl sulfide	Matheson, Coleman and Bell	Adsorption	25–50 μg/L	30–100 μg/L
0841	Earthy	Geosmin	Nat'l Environment Research Center	None required	0.1 μg/L	···
0841	Earthy	2-Ethyl fenchol	PFW, Inc.	None required	5 μg/L	···
0910	Acetic	Acetic acid	J. T. Baker "Ultrex"	None required	60–120 mg/L	30–200 mg/L
1000	Sweet	Sucrose	Grocery	None required	2.6 g/L	···
1100	Salty	NaCl	Grocery	None required	0.6 g/L	···
1200	Bitter	Isohumulone	Kalsec "Isolone"[e]	None required	7–15 mg/L	0–30 mg/L
1340	Astringent	Quercitrin[f]	K & K Laboratories	Recrystallization from 50% ethanol	80 mg/L	···

[a] The standard recommended addition for reference purposes is three times the threshold.
[b] Smirnoff or equivalent. Strength varies with locality and the vodka must be analyzed before use. Addition to beer should be by weight, not volumetrically.
[c] Trade name for 2,6-Dimethyl-5-hepten-1-al. Store under refrigeration.
[d] Not a reference standard; recommended for demonstration purposes.
[e] A solution of varying strength, usually 17%.
[f] Quercitrin is both astringent and bitter.

TABLE II
Threshold of Added Substances: Methods under Consideration

Number	Method and Reference	Concentrations Tested (n)	Glasses Required[a] (n)	Individual Thresholds Determined
1	Powers Multiple Pairs Test (5)	6	4,800	No
2	Guadagni Multiple Pairs Test (5)	5	1,000	No
3	ASTM Ascending Method of Limits Test (2)	6	450	Yes
4	ISO Draft 3972 (6)	9	250	Yes
5	Difference Rating Test (5)	5	175	No

[a] For 25 assessors.

associated reference standards that can be added to beer to demonstrate the various terms. A second duty of the subcommittee is to develop a set of standard methods for sensory analysis of beer.

TERMINOLOGY AND REFERENCE STANDARDS

The system of terminology was published in 1979 (8), followed by a set of 17 reference standards (1). Many more standards would be desirable because the terminology has 122 terms, but the testing of a compound often requires months of study. Twenty-seven compounds are currently under consideration as standards; the four recommended for international adoption at this time are melonal (Givaudan) for 0145 Melonal, benzaldehyde (Aldrich) for 0224 Almond, octanoic acid (Sigma) for 0611 Caprylic, and 2-ethyl fenchol (PFW, Inc.) for 0841 Earthy, for a total of 21 (Table I). A fifth substance, Cluster hop oil (S. S. Steiner, Inc.), while not a standard, is acceptable for demonstration purposes. The compound 2-ethyl fenchol is an alternate standard for geosmin, which was accepted last year; both are difficult to obtain.

In 1980–1981, the subcommittee plans to complete work on approximately 10 standards, which must include the following: sodium sulfide for 0721 H_2S, ethyl mercaptan for 0722 Mercaptan, 2-trans-nonenal for 0820 Papery, vanillin for 1003 Vanilla, and ferrous ammonium sulfate for 1330 Metallic.

PAIRED COMPARISON TEST

In September 1979, the ISO issued its long-awaited "Second Draft" (7), which was distributed to the subcommittee members for comments in November along with the ASBC third draft. Both documents were also distributed to the ASTM Subcommittee E-18.05 on ISO matters, which generated many comments. The majority objected to two items:

1) Use of the question, "Do you perceive a difference?" In this method, if only the combinations AB and BA are served, assessors return too many "yes" verdicts because few people want to be known as poor discriminators. If all four combinations (AA, AB, BA, and BB) are served, the method wastes sample and the panelist's time.

2) The practice of combining different questions on a single form, which was found to introduce bias in several ways.

TABLE III

Ascending Method of Limits: Individual and Group Thresholds of Three Substances Added to Beer of Collaborator's Choice

BENZALDEHYDE

Collaborator	No. of Panelists	No. of Panelists with BET at Concentrations Shown (mg/L)									Threshold, mg/L		
		.0354	.0707	.141	.283	.566	1.13	2.26	4.53	9.05	G.M.	M.V.	EHRM
A	16			1	1	1	8	5			1.08	1.13	1.04
B	17					2	4	6	1	4	2.36	2.26	2.15
C	16					5	1	3	7		1.91	2.26	1.97
D1	16			2	1	4	3	4	2		1.05	1.13	0.92
D2	16					3	7	3	3		1.56	1.13	1.49
E	16		3	2	1	0	4	4	2		0.67	1.13	0.60
F1	16			1	1	3	1	2	8		1.75	3.40	2.01
F2	16		1	1	2	4	1	3	4		0.95	0.85	0.98
G	16		1	0	2	5	6	2			0.70	0.85	0.64
J1	19		2	3	2	1	2	4	5		0.85	1.13	0.88
J2	19		5	1	5	2	5	1			0.33	0.28	0.31
Total	183		12	11	15	30	42	37	32	4			
				.177	.354	.707	1.41	2.83	5.66				
D3	16			3	0	5	4	3	1		0.96	1.06	0.91
D4	16			2	2	2	4	4	2		1.19	1.41	1.15

GERANIOL

Collaborator	No. of Panelists	No. of Panelists with BET at Concentrations Shown (mg/L)												Threshold, mg/L		
		.0044	.0088	.0177	.0354	.0707	.141	.283	.566	1.13	2.26	4.53	9.05	G.M.	M.V.	EHRM
A	16			2	3	1	2	1	1	2	1	1	2	0.295	0.212	0.297
B	16		1	1	1	2	2	3	2	1	0	3		0.259	0.283	0.241
C	16			3	0	2	1	1	6	3				0.228	0.566	0.236
E	16			5	0	0	1	4	2	4				0.176	0.283	0.225
F1	16			7	1	1	1	2	2	2				0.085	0.053	0.072
F2	16			7	0	0	0	2	2	5				0.143	0.283	0.165
G	16			5	1	3	6	0	1					0.065	0.071	0.062
H1	22			8	1	1	4	3	4	1				0.091	0.141	0.086
H2	21	5	3	0	1	1	8	3						0.042	0.141	0.042
I	19			3	2	1	4	2	3	4				0.176	0.141	0.178
J1	18			6	2	1	2	0	4	3				0.112	0.106	0.098
J2	18			3	2	1	5	4	1	2				0.131	0.141	0.128
J3	18			5	1	2	2	2	2	4				0.136	0.141	0.135
Total	228	5	4	55	15	16	38	27	30	31	1	4	2			
						.088	.177	.354	.707	1.41	2.83	5.66				
D1	16					4	5	2	2	1	1	1		0.324	0.177	0.324
D2	16					10	1	1	0	1	2	1		0.239	0.088	0.169

SODIUM METABISULFITE

Collaborator	No. of Panelists	No. of Panelists with BET at Concentrations Shown (mg/L as SO_2)												Threshold, mg/L		
		.884	1.77	3.54	7.07	14.1	28.3	56.6	113	226	453	905	1810	G.M.	M.V.	EHRM
A	16			1	1	7	6	1						17.6	14.1	16.5
B	17	1	0	0	1	0	0	2	1	2	2	4	4	266	453	295
C	16			1	0	1	3	1	7	3				67	113	63.5
D1	16			1	1	4	2	6	2					29.5	42.4	32.9
D2	16				2	6	5	3						20.9	21.2	21.0
E	15						1	4	7	3				98	113	96.3
F1	16			3	1	1	2	6	2	1				29.5	56.6	27.2
F2	16			5	0	2	3	1	3	2				23.7	28.3	23.7
J1	17			3	2	3	5	2	2					18.8	28.3	18.5
J2	17			4	5	4	2	0	0	2				12.5	7.07	12.7
Total	162	1	0	18	13	28	29	26	24	13	2	4	4			
			1.41	2.83	5.66	11.3	22.6	45.3	90.5							
G1	12		1	0	0	1	4	4	2					26.9	33.9	24.3
G2	12				3	0	2	5	2					26.9	45.3	26.3
G3	12		1	0	2	1	3	3	2					20.2	22.6	19.4
		.953	1.91	3.81	7.63	15.2	30.5	61.0								
I	18	2	2	0	2	6	4	2						11.2	15.2	10.2

BET = Best Estimate Threshold. G.M. = Geometric Mean. M.V. = Median Value. EHRM = Ellis Hall Ranking Method.

The subcommittee accepted an IFT proposal to distinguish between two forms of the test: the Directional Difference Test, in which the question is "Which sample has the stronger intensity of (the characteristic studied)?" and the Paired Preference Test, in which the question is "Which sample do you prefer?"

In addition, all modern versions of these tests recognize that they can be used in this bilateral form, in which the assessor can reply "Sample A" or "Sample B," and also in the unilateral form, in which the questionnaire identifies one sample and asks whether the other is either "stronger" (directional test) or "preferred" (preference test).

THRESHOLD OF ADDED SUBSTANCES

The suggestion for this study came from the EBC. A preliminary report (5) was based on examination of 13 methods. Table II contains a list of five methods, of which methods 1, 2, and 5 were examined by three laboratories in 1977; method 3 was submitted to a collaborative test during the present year.

An unexpected conclusion emerged from the 1977 studies and from an investigation carried out in the chairman's laboratory (9): individual variations in sensitivity are so large that the *group threshold* and its confidence limits are of little interest. What needs to be determined is *the frequency distribution of the individual assessors' threshold values.* This means that methods 1 and 2, which determined the group threshold with good accuracy, and method 5, which determined the same number with poor accuracy, would not be useful as official methods.

ISO Draft 3972. The fourth method (6) was not tested because of a major flaw, the absence of coded samples for which the identities are not revealed until after the test. In this method, subjects receive a control (blank) and nine samples containing increasing concentrations of the test substance, with instructions to report for each sample whether the added flavor is absent, doubtful, present, or clearly present.

Ascending Method of Limits Test. Method 3, the ASTM method (2), avoids all bias by presenting each sample as part of a coded triangle containing two blanks, so that the probability of being right by chance is only 33%. In the collaborative test, three substances—benzaldehyde (Aldrich, 98%), geraniol (Aldrich, 99+%), and sodium metabisulfite (Fisher, 67.4% SO_2)—were distributed by the chairman to nine collaborators. Each used his or her own beer and panel, so the results (Table III) cannot be expected to be identical.

Essentially normal results are shown, eg, by collaborators A, C, and D in the sodium metabisulfite tests. Individual thresholds form a bell-shaped distribution around a central value, the group threshold. Approximately half the tests reported in Table III conform to this pattern, indicating a successful test; the rest show various anomalies.

An abnormality that could be called "crowding at the low end" is seen in the geraniol tests by collaborators G and H. A group of assessors was able to detect the added substance at all concentrations at which it was offered. The panel leader should conclude from such results that the range of concentrations has been too high. A new test, as in H2 or preferably at still lower concentrations, should be given until a level is reached where few can detect the lowest addition. At the same time, taking into consideration that, in each triangle, panelists can be right by chance in one case out of three, the panel leader should review the results with each assessor, using the extended form of the test (Appendix B), so as to eliminate chance as the operative factor.

Another variation, illustrated by collaborator F1 with benzaldehyde, could be called "crowding at the high end" and should cause the panel leader to demonstrate the test compound to the assessors and then, having ascertained that they are fully familiar with its flavor, offer the test again (eg, test F2).

Lack of familiarity with the test flavor was the most frequent cause of high individual thresholds, both in the earlier (5) and the present series. Thresholds were found to improve from test to test, sometimes dramatically, until a plateau was reached after three to six tests with a given substance. Examples were seen of panelists missing concentrations 10–100 times above their threshold because of unfamiliarity at their first or second encounter with the substance. The effect was compared to being in a dark room and looking for a flicker of light at threshold intensity without knowing the direction from which it might come. Not until a flavor is *known and retained in memory* can it be detected with confidence.

Collaborator B's test with metabisulfite (Table III) is of interest because panelists were unprepared and results were not reviewed with them. Under such conditions, a panel may return an almost meaningless set of verdicts, ranging from a chance result of 0.884 ppm to the responses of eight confused panelists who could not identify samples reeking of SO_2 at 640 and 1,280 ppm. The same panel returned a normal set of thresholds with its third compound, benzaldehyde; hence *unfamiliarity with the test setup* and the flavors caused the confusion in the SO_2 test.

Several collaborators reported that they needed a couple of tests to learn to approach the set of 18 glasses without blinding the senses of taste and smell. A useful procedure was to start at middle concentrations and to try to locate the highest concentrations by smell or by taking very small sips, thus saving one's discriminatory powers for those glasses containing concentrations near one's threshold. The objective is not to taste all 18 samples but rather to locate those glasses containing the addition at above threshold levels.

Motivation Factor. A factor in the success of the test is how much effort an assessor will make to identify the stimulus. The panel leader has the responsibility of generating the required atmosphere, and many factors intervene. Motivation is best in a well-understood, well-defined test situation and strongly improves as subjects become familiar with the test substance.

The ISO Draft 3972 method is open to motivational bias because no hidden controls are included. Bias is introduced partly because no one wants to be a poor discriminator and partly because of autosuggestion, to which even experts are exposed: a sample known to contain a given substance often seems to take on the corresponding flavor. The subcommittee feels that the ISO 3972 method may be useful under certain well-defined conditions: ie, each assessor is fully familiar with the substance and has sufficient experience and involvement with the study to be able to avoid bias and autosuggestion and the results are confirmed by subsequent blind tests. The method is unsuitable for studies destined for publication.

Calculation of Group Threshold. Although the group threshold is of less value than the distribution curve of personal thresholds, it is needed for comparisons between several substances or groups. In Table III, three methods were used for calculating group thresholds. The definition of threshold (5) calls for use of the *median,* or 50%-point, but this method is too inflexible; a single triangle by a single assessor can shift the group median by a factor of two. The *geometric mean* avoids this but gives too much weight to panelists who cannot detect the substance and too little to those who can (9). The *Ellis Hall ranking method* (4) is a method of normalizing data so that Probits or probability paper may be used, as for dose-response data in toxicology. Because the resulting values showed no systematic advantage over the geometric means, the extra effort required to calculate Ellis Hall values does not appear to be warranted. The geometric mean is therefore recommended, despite its shortcoming.

Threshold of Geraniol. Whereas the results in Table III show benzaldehyde and sodium metabisulfite to be suitable as flavor reference standards, a number of assessors apparently were quite insensitive (almost flavor blind) to geraniol and others could detect it at extremely low levels. Such a bimodal distribution, if confirmed by further tests, would make geraniol a candidate for a "primary" odor according to Amoore (3) but also makes it unsuitable as a reference standard.

In conclusion, the Ascending Method of Limits Test is recommended for inclusion in the ASBC "Methods of Analysis," with the understanding that no reliable result can be obtained with this or any other method until assessors are fully familiar with the flavor of the test substance and also with the mechanics of the method.

LITERATURE CITED

1. American Society of Brewing Chemists. Report of Subcommittee on Sensory Analysis. *Journal* 37:130, 1979.
2. American Society for Testing and Materials. E 679-79 Standard Practice for Determination of Odor and Taste Thresholds by a Forced-Choice Ascending Concentration Series Method of Limits. ASTM: Philadelphia, PA, 1979.
3. Amoore, J. E. *Chem. Senses Flavour* 2:267, 1977.
4. Bengtsson, K., and Helm, E. *Wallerstein Labs. Commun.* 9:171, 1946.
5. Brown, D. G. W., Clapperton, J. F., Meilgaard, M. C., and Moll, M. *J. Am. Soc. Brew. Chem.* 36:73, 1978.
6. International Organization for Standardization. Draft International Standard 3972-1977 (E). ISO: Paris.
7. International Organization for Standardization. Draft International Standard ISO/DIS 5495.2: Sensory Analysis—Methodology—Paired Comparison Test. ISO: Paris, 1979.
8. Meilgaard, M. C., Dalgliesh, C. E., and Clapperton, J. F. *J. Am. Soc. Brew. Chem.* 37:47, 1979.
9. Meilgaard, M. C., and Reid, D. R. *Inst. Brew. (Aust. N.Z. Sect.) Proc. 15th Conv., Christchurch, N.Z., 1978,* p. 97.

APPENDIX A
PAIRED COMPARISON TEST

Scope

Use this test for the following purposes:

(a) *Directional Difference Test:* to determine in what way a particular sensory characteristic differs between two samples (eg, more sweet or less sweet);

(b) *Paired Preference Test:* to establish whether a preference exists between two samples (eg, in consumer tests);

(c) *Assessor Training:* to select, train, and perfect assessors.

The Paired Comparison Test is one of the simplest and most-used sensory tests. It is often used first to determine whether other, more complicated tests should be applied. Different test forms and methods of interpretation are used, depending on the test objective, and distinguishing between bilateral (two-tail) and unilateral (one-tail) test situations is necessary.

Field of Application

Apply the test to any two samples of beer, wort, water, etc. whether or not a sensory difference is expected.

Principle of the Method

A pair of samples is presented to assessors; one sample may be a control. After tasting, assessors complete a written answer form and the replies are interpreted.

General Test Requirements

Test Room. Conduct the tests in a room that complies with the requirements in **SENSORY ANALYSIS-2**[1]; Test Room, Equipment, Conduct of Test.

Assessors. Choose the number of assessors with reference to the degree of difference expected between the samples, the aptitude of the assessors, and the risk of error which can be accepted, eg, 5, 1, or 0.1%.

As a general rule, the minimum required is seven assessors, but discrimination is much improved if 20, 30, or (for consumer tests) one or several hundred are employed.

[1] In preparation.

A

Directional Difference Test (Bilateral)

NAME ..... DATE ..... (year/month/day)
OBJECT OF TEST .....
TEST CRITERION .....

Test Pairs	Which sample is more ..... ?
Sample No. / Sample No.	
....	.....
....	.....
....	.....

Comments

Paired Preference Test (Bilateral)

NAME ..... DATE ..... (year/month/day)
OBJECT OF TEST .....
TEST CRITERION .....

Test Pairs	Which sample do you prefer?
Sample No. / Sample No.	
....	.....
....	.....
....	.....

Comments

B

Directional Difference Test (Unilateral)

NAME ..... DATE ..... (year/month/day)
OBJECT OF TEST .....
TEST CRITERION .....

Test Pairs	Is Sample L more ..... than Sample R?	
Sample L / Sample R		
....	Yes	No .....
....	Yes	No .....
....	Yes	No .....

Comments

Paired Preference Test (Unilateral)

NAME ..... DATE ..... (year/month/day)
OBJECT OF TEST .....
TEST CRITERION .....

Test Pairs	Do you prefer Sample L to Sample R?	
Sample L / Sample R		
....	Yes	No .....
....	Yes	No .....
....	Yes	No .....

Comments

Fig. 1. Specimen answer forms for the Directional Difference Test and the Paired Preference Test, showing the bilateral (**A**) and unilateral (**B**) forms of the test questions.

Preliminary Discussion and Test. Ascertain that assessors are fully familiar with characteristic studied and with the mechanics of the tests. Arrange, if necessary, a preliminary general discussion concerning the test problem and nature of samples. Direct such discussion in a manner that cannot influence subsequent judgments.

Present and discuss a few samples typical of the series to be analyzed. Limit the number to two or three. If the test concerns the detection of off-flavors, include in the preliminary test a sample free from any off-flavor, or, on the contrary, when possible, a demonstration of the off-flavor to be detected.

In general, inclusion of controls (reference substances) may be advisable.

TABLE I
Unilateral Test[a]

Number of Replies	Minimum Number of Positive Replies for Probability Level of		
	$P \geq 95\%$	$P \geq 99\%$	$P \geq 99.9\%$
7	7	7	...
8	7	8	...
9	8	9	...
10	9	10	10
11	9	10	11
12	10	11	12
13	10	12	13
14	11	12	13
15	12	13	14
16	12	14	15
17	13	14	16
18	13	15	16
19	14	15	17
20	15	16	18
21	15	17	18
22	16	17	19
23	16	18	20
24	17	19	20
25	18	19	21
30	20	22	24
35	23	25	27
40	26	28	31
45	29	31	34
50	32	34	37
60	37	40	43
70	43	46	49
80	48	51	55
90	54	57	61
100	59	63	66

[a] The values given in the tables were calculated from the exact formula: binomial law for parameter $P = 0.50$ with n repetitions (replies).

The risk of error connected to the statistical conclusion, denoted by α, is the complement to 100 of the probability level. For $P \geq 95\%$, $\alpha \leq 5\%$; $P \geq 99\%$, $\alpha \leq 1\%$; $P \geq 99.9\%$, $\alpha \leq 0.1\%$. Because the number of replies can only include whole values, probability levels exactly equal to 95%, 99%, or 99.9% cannot be obtained, but only lower limits for these levels (higher limits for the risk α). This explains the fact that, for a small number of replies, the values given in a table may be the same for two probability levels.

When the number of replies is higher than 100 ($n > 100$), it is necessary to use the following formula based on the approximation binomial law by the normal law and which gives the actual minimum given number of assessments to be obtained, with a maximum error equal to 1 unit. Minimum number of replies: Nearest whole value to

$$\frac{n+1}{2} + k\sqrt{n}$$

in which, in the unilateral test, for $P \geq 95\%$, $k = 0.82$; for $P \geq 99\%$, $k = 1.16$; for $P \geq 99.9\%$, $k = 1.55$; and in the bilateral test, for $P \geq 95\%$, $k = 0.98$; for $P \geq 99\%$, $k = 1.29$; for $P \geq 99.9\%$, $k = 1.65$.

[b] Tables for calculating other significance levels may be found in Roessler et al (3) and Ferdinandus et al (4).

Procedure

Preparation of Samples. Take care that assessors cannot draw conclusions about the nature of the samples from the way in which they are offered. Use colored glasses. Present the various pairs of a series in random order. Serve samples at uniform temperature; 12° C is suitable for full perception of flavor. Ascertain that the temperature of each sample pair is identical, also that of samples offered in succession. This applies not only at the time of presentation but also throughout the tests. Code glasses using random numbers or key the placement of glasses in advance.

Test Portion. A suitable quantity is 50–100 ml in a 250-ml (8 oz) glass.

Test Technique. Offer paired samples simultaneously or successively for tasting. Prepare equal numbers of the combinations AB and BA and allot them at random among the assessors. Series of pairs may be offered, but to avoid sensory fatigue, do not exceed 10 glasses per sitting. Instruct assessors to examine sets in a specified order, eg, always from left to right, leaving, however, the opportunity of making repeated tests of any sample while tasting of a set is in progress.

Test Question and Answer Forms. In most test situations (bilateral), the test question does not distinguish between samples and the reply may favor one or the other sample. The unilateral test situation is less frequently used and applied when the characteristic studied can vary in only one direction. For example, a beer "A" may suffer from lack of intensity of a certain characteristic, and the test supervisor wishes to know whether sample "B" has more of that characteristic, or sample "A" may be of inadequate quality and the question asks whether sample "B" is an improvement. The question must assume a directional difference or preference.

Specimen answer forms are shown in Fig. 1. Do not place more than one question on a given form, although space for several sample pairs may be provided. Recode and resubmit samples for each separate question.

Replies of the Assessors. The test supervisor may use one of the following two possibilities: 1) according to the "forced choice" technique, oblige the assessors to choose one sample or the other (by guessing, if no difference is perceived) or 2) allow the answer "no difference."

	Bilateral Test	Unilateral Test
Directional Difference Test	Question: Which sample has the stronger intensity of the characteristic studied? Count the number of replies citing one of the two samples the more frequently. Conclude that the intensity for this sample is significantly stronger than for the other if the number obtained is greater than or equal to that shown in Table II.	Question: In relation to sample "452", does sample "603" have a stronger intensity of the characteristic studied? Conclude that this stronger intensity is significantly apparent if the number of positive replies is greater than or equal to the number shown in Table I.
Paired Preference Test	Question: Which sample do you prefer? Count the number of replies citing one of the two samples the more frequently. Conclude that this sample is significantly preferred to the other if the number obtained is greater than or equal to that shown in Table II.	Question: Do you prefer sample "517" to sample "832"? Conclude that there is a preference for sample "517" if the number of positive replies is greater than or equal to the number shown in Table I.

Fig. 2. Questions and interpretations for the two forms of each test.

Expression and Interpretation of Results

"Forced Choice" Technique. Collate and interpret results as shown in Fig. 2 and refer to probabilities in Tables I or II to determine the signficance level.

"No Difference" Technique. Choose one of the following possibilities: 1) ignore the "no difference" replies, ie, subtract them from the total number of the panel (this technique increases the chance of obtaining a significant result); or 2) allocate half of the "no difference" replies to each of the two categories of replies (this reduces the chance of obtaining a significant result).

The "forced choice" technique is preferred but may not be usable because assessors object to it. If so, the proportion of "no difference" replies may be accepted as an additional piece of information that may help in planning subsequent tests, revealing, for instance, imperfections in experimental design, the existence of an important physiological variation in the assessors, or even a lack of motivation of certain assessors for the tests in which they are participating.

Example

Directional Difference Test. A beer "A" is deemed insufficiently bitter and a test brew "B" is made using a higher hop rate. The test supervisor wishes to know whether a significant increase in bitterness has been obtained. He accepts a 1% risk of error, ie, $P \geq 99\%$. The samples are coded "452" and "603," respectively, and offered to a panel of 30 assessors.

Bilateral Test

Question: Which sample is more bitter?
Replies: 22 opt for sample "603" and 8 opt for sample "452."

From Table II he can conclude that the two beers are not significantly different.

Unilateral Test

Question: Is sample "603" more bitter than "452"?
Replies: 22 Yes and 8 No.

From Table I he can conclude that beer "B" is significantly more bitter than beer "A."

TABLE II
Bilateral Test[a]

Number of Replies	Minimum Number of Replies Citing One Sample for Probability Level of		
	$P \geq 95\%$	$P \geq 99\%$	$P \geq 99.9\%$
7	7	...	...
8	8	8	...
9	8	9	...
10	9	10	...
11	10	11	11
12	10	11	12
13	11	12	13
14	12	13	14
15	12	13	14
16	13	14	15
17	13	15	16
18	14	15	17
19	15	16	17
20	15	17	18
21	16	17	19
22	17	18	19
23	17	19	20
24	18	19	21
25	18	20	21
30	21	23	25
35	24	26	28
40	27	29	31
45	30	32	34
50	33	35	37
60	39	41	44
70	44	47	50
80	50	52	56
90	55	58	61
100	61	64	67

[a] Refer to footnotes of Table I.

Paired Preference Test. The samples are re-coded "832" and "517," respectively, and submitted again to the panel. The test supervisor now accepts a 5% risk of error.

Bilateral Test

Question: Which sample do you prefer?
Replies: 21 prefer sample "517" and 9 prefer sample "832."

From Table II he can conclude that beer "B" is preferred over beer "A."

Unilateral Test

Question: Do you prefer sample "517" to sample "832"?
Replies: 21 Yes and 9 No.

From Table I he can conclude that beer "B" is preferred over beer "A."

Expression of Results

Make reference to this method and give the following information:

(a) Test problem
(b) Full identification of the sample or samples
(c) Test parameter adopted (ie, characteristic studied)
(d) Whether or not reference substances were used
(e) Any other recommendations given during the test (for example information relating to the type of beer)
(f) Number of tests and number of trained or untrained assessors
(g) Any test conditions differing from the recommendations given in this method
(h) The results, with their probability levels
(j) Date of test
(k) Name of test supervisor

References

1. AMERICAN SOCIETY OF BREWING CHEMISTS. Report of Subcommittee on Beer Flavor Analysis. *Proc. 1962,* p. 181; *Proc. 1963,* p. 203; *Proc. 1964,* p. 262; *Proc. 1965,* p. 253; *Proc. 1966,* p. 246; *Proc. 1967,* p. 247; *Proc. 1968,* p. 226; *Proc. 1969,* p. 212; *Proc. 1970,* p. 222.
2. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. Draft International Standard ISO/DIS 5495.2 Sensory Analysis—Methodology—Paired Comparison Test. ISO: Paris (1979).
3. ROESSLER, E. B., PANGBORN, R. M., SIDEL, J. L., and STONE, H. *J. Food Sci.* 43:940 (1978).
4. FERDINANDUS, A., OOSTEROM-KLEIJNGELD, I., and RUNNEBOOM, A. J. M. *Tech. Q. Master Brew. Assoc. Am.* 7:210 (1970).

APPENDIX B
SENSORY ANALYSIS: THRESHOLD OF ADDED SUBSTANCES—ASCENDING METHOD OF LIMITS TEST

Scope

The test determines the lowest concentration of an added substance that can be detected by odor or taste. The threshold may be determined in order to:

(a) Learn whether a substance present in beer is likely to exert an effect on its flavor;
(b) Provide an indication[1] of the magnitude of any such effect;
(c) Study the relative sensitivities of individuals to the test substance;
(d) Select, train, and perfect assessors.

The test determines the Difference Threshold as defined in Brown et al (3). If the recognition threshold is desired, assessors must describe and correctly identify the added flavor.

Field of Application

The test may be applied to any nonpoisonous substance or mixture that dissolves in beer, wort, or water. Substances that affect the appearance of the solution require evaluation in colored glasses and/or subdued or colored light.

[1] By application of the concept of "Flavor Unit" (FU). For a given substance, the number of FUs present (or added) equals the concentration of substance present (or added) divided by the threshold (6). FU is also called "Odor Unit" when only the odor is considered.

The simple form of the test is suitable for obtaining rough estimates under (a) and (b) above. The extended form is required for more accurate determinations and for the study of individual thresholds.

Principle of the Test

The experimental design used is known in psychophysics as the forced-choice modification (5) of the ascending method of limits (4).

Assessors (16 or more) receive sets of six triangles, each consisting of two controls and one test sample. Test samples increase in concentration by a constant factor, usually 2.0. Individual best estimate values of threshold are found separately for each assessor as the geometric mean of the highest concentration missed and the next higher (adjacent) concentration. Group thresholds are then derived from the individual values. In the extended form of the test, the individual values are refined by repeated testing before the group threshold is calculated.

General Considerations

A discussion of basic principles of sensory testing, choice of assessors, motivation, test rooms, apparatus, serving temperatures, etc., is found in **SENSORY ANALYSIS-1–5.**[2] Basic concepts regarding thresholds, such as definitions, variations among assessors, and flavor blindness are discussed by Brown et al (3) and Meilgaard and Reid (7).

Threshold determination by any method depends critically on prior training. Assessors cannot obtain reliable results unless 1) the flavor of the compound under test is known and retained in memory and 2) the mechanics of the test are fully familiar. A training schedule of three or four applications of the present method is a minimum requirement.

Published values (3,7) indicate that panels of 5–15 members are representative only of themselves. A panel must have 20–30 members for its threshold pattern to have any general validity.

[2] In preparation.

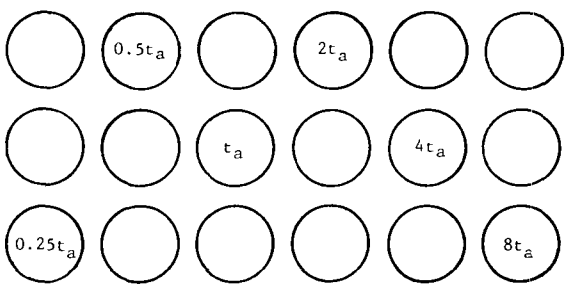

Fig. 1. Example of presentation of the six test triangles. t_a = approximate threshold.

ASCENDING METHOD OF LIMITS

Date _____ Assessor _____

You have received six sets of beer samples. Each set is a triangle consisting of two identical controls and one test sample containing an added substance. Concentrations of the added substance increase from left to right.

Please locate as many as you can of the test samples, indicating their position with a check mark in the corresponding box in each column. - Avoid sensory fatigue: locate strong samples by smell or by taking very small sips, conserving your discriminatory power for those triangles near your threshold. Review your results with the test supervisor.

Triangle No.: 1 2 3 4 5 6

Describe the flavor of the added substance _____

Fig. 2. Example of questionnaire.

Procedure

Determine the approximate threshold (t_a) using 5–10 assessors in a preliminary test in which test samples increase in concentration by a factor of 3.0.

Prepare the dilution series for the main test using a factor of 2.0. Arrange samples so that step 3 = t_a, step 2 = 0.5 t_a, step 1 = 0.25 t_a, and steps 4, 5, and 6 = 2 t_a, 4 t_a, and 8 t_a, respectively. Present the dilution series to the assessors in the manner shown in Fig. 1, using a questionnaire similar to that shown in Fig. 2. Each triangle consists of one test sample and two controls. Choose the position of test sample at random, taking care that it appears approximately an equal number of times in the bottom, middle, and top positions.

Instruct assessors to indicate the position of the test sample in each triangle, guessing when necessary. Allow assessors to optimize a strategy of locating test samples, eg, by avoiding overexposure and tasting a minimum number of the 18 glasses. At the end of the test, review results with each assessor. Because a finite possibility exists that a correct result will occur by chance, allow assessors to retest triangles where they indicate that this may have occurred.

Extended Form of Test. Repeat critical triangles two to four times with each assessor until assessor and test supervisor agree that they have successfully bracketed his or her threshold. For asssessors at the top or bottom of the range, add extra concentration steps as required, ie, 16 t_a, 32 t_a, etc.

Procedure: ASBC Sensory Analysis-13 (Ascending Method of Limits)
Sample: Dimethyl sulfide, MCB cat. MX 1445
Purification: 4-Adsorbent procedure
Medium used: Stroh Bohemian lager, 11.4°P, 14 BU, 50 µg/L DMS
Number of scale steps: 6 Dilution factor per step: 2.0
No. of assessors: 16 Form of test used: simple [√] extended [ ]

Panelist	Concentrations presented, µg/L 10	20	40	80	160	320	Best Estimate Threshold µg/L	Log_{10}
WA	0	0	+	0	0	+	226	2.35
LB	0	+	+	+	+	+	14	1.15
PB	0	0	0	+	+	+	57	1.75
JD	+	+	+	+	+	+	7.1	0.85
CF	+	+	0	+	+	+	57	1.75
EK	0	0	+	+	+	+	28	1.45
JM	+	0	0	+	+	+	57	1.75
MM	0	+	0	+	0	0	453	2.65
DR	0	0	+	+	+	+	28	1.45
MR	0	+	+	+	+	+	14	1.15
JS	0	+	0	+	+	+	57	1.75
KS	0	0	+	+	+	+	28	1.45
LS	0	+	+	+	+	+	14	1.15
JW	0	0	+	+	+	+	28	1.45
KW	+	+	+	+	+	+	7.1	0.85
TW	0	0	+	+	+	+	28	1.45

Sum ⟶ 24.40

Group BET, geometric mean, µg/L 33.5 ⟵ 1.525

Log standard deviation = 0.471

Histogram of Individual BE Thresholds:

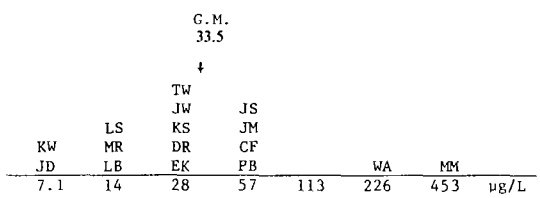

Fig. 3. Example of application. + = correct choice, 0 = incorrect choice.

Expression and Interpretation of Results

Collate and interpret results as shown in the example in Fig. 3. For each assessor, calculate best estimate threshold (BET) as the geometric mean of the highest concentration missed and next higher (adjacent) concentration. In the example, the BET for Assessor WA is $(160 \times 320)^{1/2} = 226\ \mu g/L$. Assessor MM missed the highest concentration; assume that he would be correct at next higher level, 640 $\mu g/L$. In the extended form of the test, supply MM extra triangles at 640, 1,280 $\mu g/L$ etc., until his true BET is found.

Report results as frequency distribution (histogram) of individual BETs. Obtain the group threshold as the geometric mean of BETs.

Test Report

Refer to this method and give, in addition to test results, the following minimum particulars:

(a) Test problem
(b) Full identification of the added substance and the medium used
(c) Other test variables as shown in the Example
(d) Number of repeat tests performed on the samples
(e) Number of experts or of trained or untrained assessors
(f) Any test conditions differing from the recommendations given in this method
(g) Date of test
(h) Name of test supervisor

References

1. AMERICAN SOCIETY OF BREWING CHEMISTS. Report of Subcommittee on Sensory Analysis. *Journal* 38:99 (1980).
2. AMERICAN SOCIETY FOR TESTING AND MATERIALS. E 679 79 Standard Practice for Determination of Odor and Taste Thresholds by a Forced-Choice Ascending Concentration Series Method of Limits. ASTM: Philadelphia, PA (1979).
3. BROWN, D. G. W., CLAPPERTON, J. F., MEILGAARD, M. C., and MOLL, M. *Am. Soc. Brew. Chem. J.* 36:73 (1978).
4. ENGEN, T. Psychophysics I. Discrimination and Detection. Pages 14 and 19 in: Kling, J. W., and Riggs, L. A., eds. Woodworth and Schlosberg's Experimental Psychology, 3rd ed. Holt, Rinehart, and Winston: New York (1971).
5. JONES, F. N. *Am. J. Psychol.* 69:672 (1956).
6. MEILGAARD, M. C. *Tech. Q. Master Brew. Assoc. Am.* 12:107, 151 (1975).
7. MEILGAARD, M. C., and REID, D. S. *Inst. Brew. (Aust. N. Z. Sect.) Proc. 15th Conv., Christchurch, N.Z. 1978,* p. 97.