Simple Methods of Odor Quality Evaluation of Essential Oils and Other Fragrant Substances

By Dr. Wladyslaw S. Brud, Chairman of IFEAT Scientific Subcommittee, Pollena-Aroma, 03-115 Warszawa, Poland

Odour evaluation is usually considered by nonprofessionals, especially trade people, to be a mysterious mixture of scientific magic tricks and perfumers' voodoo giving ambiguous results fomenting difficult and hopeless disputes on product quality and value. To some extent, such an opinion is correct. When looking for efficient and simple methods of odour evaluation one requires books, publications, recommendations and standards which need a superqualified staff, sophisticated equipment and a good computer to process data. They are definitely necessary in olfactory research where most exact and reproducible data are requested.

It should be said here that research on chemoreception which also covers sensory-olfactory analysis forms an important branch of science organized within the European Chemoreception Research Organization (ECRO). This kind of research is basic for the problem of odour quality evaluation discussed here.

One should understand that all works on standardization and the best instrumental methods of analysis will not answer the simple question: "Is that sample good or not?" It is well known that an essential oil may comply with all standards and data and be totally unacceptable for perfumery use, and that the opposite may also, although rarely, be true.

Minor compounds[2] which appear in oils in minute quantities are detectible olfactorally at the level on parts per billion and yet are, in most cases, difficult if not impossible to detect by analytical methods. They can basically change the odour profile of an oil although its physiochemical parameters and gas chromatography/mass spectral data meet the standard. Minor components can appear in oil as a result of weather, harvesting methods, or plant and oil treatment. Therefore, the importance of reliable methods of odour evaluation cannot be overestimated.

It should be stated, however, that there is no method for objective odour evaluation of essential oils without comparison to a specific standard. One cannot say "this is good oil" because being perfect for one purpose, the product can be useless for another. Hence, any quality description should be referred to a certain previously approved standard.

What we are looking for in everyday manufacturing are simple, cheap and reliable methods which quickly and without hesitation will answer the basic question, "Is that sample identical with the standard, and if not, is that diference small enough that the product can still be used?"

This paper describes our practical experiences in evaluation of odour quality against standards accepted by perfumers for specified fragrances.

There are many methods of differential analysis widely described in literature. The best review of the problem is in my opinion by Moskowitz and Fishken.[3] The basic methods used for comparison of two or more products, e.g., sample and standard, are: paired comparisons, triangle test, duo-trio test, dual standard, multiple standard, and multiple pairs. Most practical and at the same time exact and reliable is the triangle test. Properly applied, it can give a definite answer to the first part of our question, "Is the sample identical with the standard or not?"

Before we go into some details of the method, a few words should be said about the people who will make the test, in other words, "how to create a good reliable panel."

 0272-2666/83/0004-4701$04.00/00

Some managers used to say, "Okay, we have perfumers and it is their job, or let the customer say yes or no." But usually it is much better to be convinced of quality of your product before the customer judges it.

The best panel can be selected from non-professional company employees previously tested for ability in odour differentiation.

According to the size of company and number of samples to be tested, the number of panelists can vary, but most useful is to have about twenty selected and trained people from whom 6-9 are asked at random for each session. This system minimizes the routine boredom which affects results from people permanently smelling samples in a control department.

For purposes of simple differential evaluation, only one testing method is necessary. A few tests can easily show the ability of the panelist to differentiate two samples. The set of samples in Table I can be recommended.

Table I. Mixtures for Testing of Panelists

Sample "Y"	Sample "X"	
Sandalwood oil	Sandalwood oil Cedarwood oil virg.	8 parts 1 part
Pine needle oil	Pine needle oil Terpineol	6 parts 1 part
Bergamot oil	Bergamot oil Terpinyl acetate	10 parts 1 part
Alphaamylcinnamic aldehyde	Alphaamylcinnamic aldehyde Benzaldehyde	35 parts 1 part

The test is the simple triangle method. The subject is presented with eight sets of three samples for each of four combinations of materials.

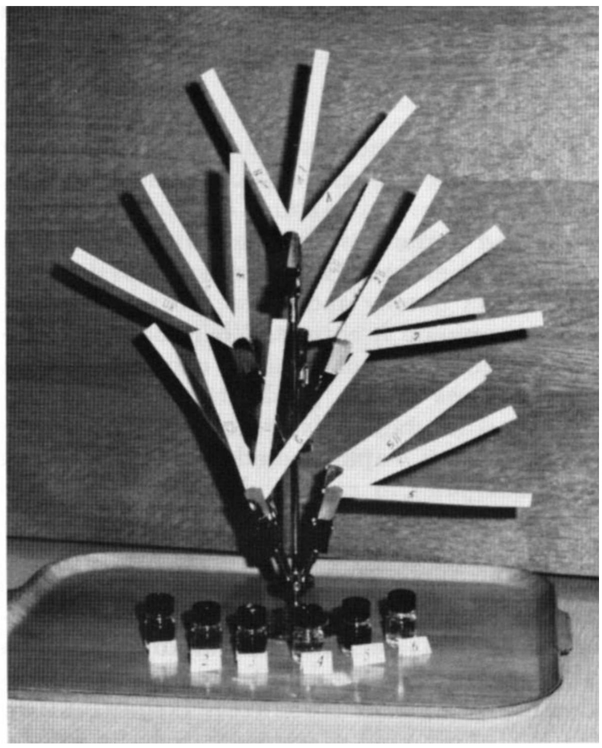
Example Combinations of Sample X and Standard Y

	A	B	C
1.	X	X	Y
2.	X	Y	X
3.	Y	X	X
4.	X	Y	Y
5.	Y	X	Y
6.	Y	Y	X
7.	X	X	X
8.	Y	Y	Y

After evaluation, about thirty minutes for each set, the panelist should specify which of the three samples in each set is different or state "no difference." A minimum of seven correct out of eight answers are necessary to accept the person for our panel.

Once the panel is selected, the most difficult part of the problem is solved. After essential training, regular tests can be started.

For each test, we need 6-9 panelists (preferably 7). Each of them is presented with coded paper strips dipped in appropriate products. There are two ways of preparation of the test according to the number of tests to be made and working time available for test preparation. The simple way is presentation of strips marked A, B and C to all panelists in the same order, e.g., each blotter represents the same material.



Sample Marking for Test made by 7 Panelists

I. For each product 1 set of paper strips.
Marking: 1A, 1B, 1C, Where 1 = Sample Number

II. For each product 7 sets of paper strips dipped in samples with the same sequence.
Marking: As above multipled 7 times

III. For each product 7 sets of paper strips dipped in samples with random sequence.
Marking: 1/1A, 1/1B, 1/1C, 1/2A, 1/2B, 1/2C, Etc. Where 1/ = Sample number and /1,2,3, ... = set of paper strips number

More complicated in preparation and calculation of results, but more objective (especially if panel is working in the same room) is the presentation strips dipped for each sample in different sequences so that each set of three strips represents a different combination. For very simple and quick work, if more samples are to be evaluated, only one set of strips is prepared for each product and panelists smell them in turn.

Panelists are requested to specify which of three strips has a different odour from two of the others (or which two strips are the same). With the questionnaire shown in figure 1, a quick record of results is collected.

One very important point should be mentioned here. The whole idea presented in this paper, which allows us to obtain reliable results with a minimum number of highly-qualified and expensive staff who should be more effectively used for creative work, requires one very efficient and responsible person, a panel secretary. This person should prepare all samples from production, suppliers, purchase, or others, code strips, organize panel meetings, keep all records and confidential documentation, and deliver results to appropriate departments in the company. Panelists should not know the sources of evaluated samples to avoid any preferences which may affect results. Therefore, the panel secretary should be the most confidential person available.

In my opinion, the triangle method as presented above is very useful for professional groups in control laboratories to avoid preferences and other factors which may affect objectivity of the evaluation tests.

Table II. Number or Percentage of Correct Answers in Triangle Test as Minimum for Reliable Result

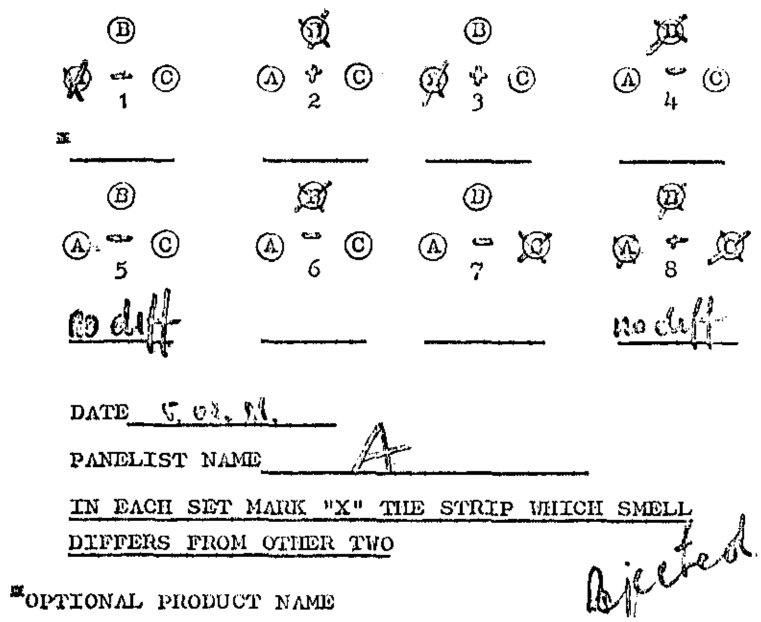
Number of Samples or Panelists (a)	Minimum Correct Answers (b)	True Percentage Correct Answers T (c)
5	4	min. 70
6	5	min. 70
7	5	min. 60
8	6	min. 60
9	6	min. 50
10	7	min. 50
20	11	min. 33
50	24	min. 22
100	43	min. 15

(a) For one panelist and one product number of samples. For group of panelists their number for each product.

(b) At significance level 0.05.

(c) As above according to formula for true percent good answers after correction for guessing $T = 1.5 (P - 33.3)$, where P = percent correct answers in the test.

Figure 1. *Triangle test questionnaire*



A final conclusion from results obtained from the panel should be made by the panel secretary. According to literature sources on statistical data evaluation, there is a direct relation between the number of panelists (or tested sample sets) and the minimum number (or percentage) of correct answers for a reliable result. These data are shown in Table II.

Obviously, the percentage of necessary correct answers decreases with an increasing number of panelists. Our experience, however, showed that seven panelists were quite enough to give reproducible and true results, and even five people with some training can be reliable on routine tests. The significance level mentioned in the table represents a probability of mistake in the form of a rejection of proper hypothesis. A 0.05 significance level is low enough for our purpose. In certain tests, for example, testing of panelists, lower values of 0.01, 0.001 may be necessary which require more correct answers for reliable results. Some authors suggest a higher reliability of triangle method with an additional sample of uncoded standard. The panelists should decide which of three samples (one or two) is identical with the standard. This method, however, is more complicated in preparation and calculation of results.

In a recently published paper, it was established that repetition of a triangle test improved reliability of results due to so-called "incidental training."[5] If certainly true for newly created panels, it is not so important with sufficiently trained groups.

It is important for regular checking of panelists and for stimulation of their attention to present occasionally a set of three identical samples coded as usual. The panelists should be aware of possibility of such a control, but should never know when it will happen. Although a little bit complicated at first glance, the triangle method is very simple for routine work and for the purpose discussed here the most reliable and reproducible.

The real problem starts when our sample of the product to be used or delivered is rejected by triangle test. In most situations, an oil which shows some differences in odour in comparison with a standard can be easily used in compounds without any influence on their quality. This is especially true for natural oils where minor differences are normal. What we are now looking for is the acceptable range of this difference; or in other words, the answer for our second question, "Is that difference small enough that a product still can be used?" This question is especially essential when we deal with natural products, compositions, and blending processes.

Physiochemical data can provide a range that is clear and acceptable. The problem is how we can get subjective odour evaluation data into figures reliable and comparable with other standard parameters. In other words, we have to create a scale where a range of odours can be described similarly without ambiguity.

There are many methods and scales in literature for odour description and evaluation which can be used for our purpose including odour profiles and multidimensional descriptions. Most of them are either too complicated for routine manufacturers' work or too simple to give a reliable answer for a basic question, "use or not use." For example, a popular difference scale[8] of slight difference, moderate difference, large difference and extreme difference is good for testing market products but useless for our purpose as difference between "slight" and "moderate" and the exact meaning of these words is rather ambiguous.

We found a solution of the problem with an eight-point scale.

1—TOTALLY DIFFERENT
2—DIFFERENT WITH SOME NOTE OF STANDARD
3—DIFFERENT TOP OR BASE NOTE
4—DISTINCT DIFFERENT NOTE
5—WEAK, STABLE DIFFERENT NOTE
6—WEAK, PASSING DIFFERENT NOTE
7—JUST NOTICEABLE DIFFERENCE
8—NO DIFFERENCE, IDENTICAL WITH STANDARD

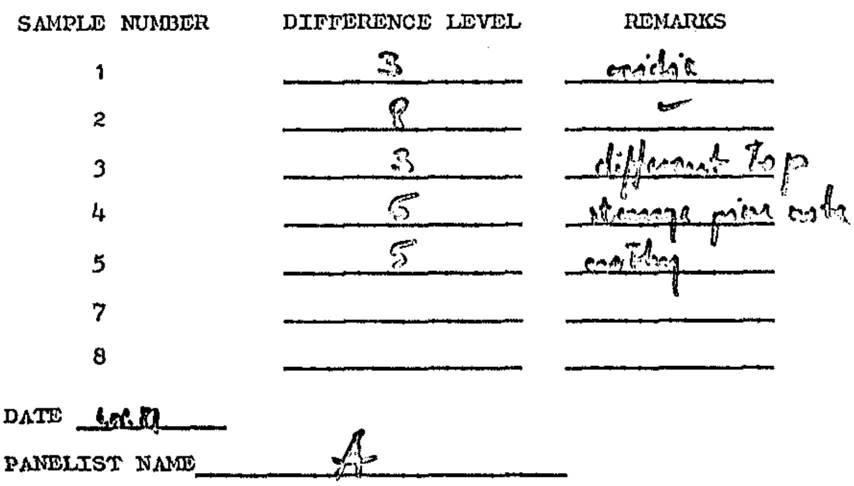
The main purpose of this scale was to let our panelists clearly distinguish the degree of difference between tested sample and standard, and to obtain results which allow us to make an objective decision on product quality.

Although the scale itself is rather clear for anybody dealing with fragrances, it needs some comments. 8 is obvious and this is the standard. 7 means situations when a panelist is not absolutely sure if there is or there is not any difference. As some people say, "there is something in it" which means that some nonrecognizable difference is felt or suspected but nothing clear can be said of its character or odour quality.

6 is used when the sample shows a clearer strange note temporarily, e.g., at the opening of the bottle, or just at first sniff on paper strip, or said note appears for awhile during evaporation, but disappears after a minute or two. When such a strange note remains detectable for up to one hour (usually during the test time), 5 shall be used. It is important to emphasize that within that range we speak of very weak, practically nonrecognizable foreign notes which can be detected only by well-trained panelists and usually without a clear description.

When a distinct, recognizably different note appears for a few hours, we use 4. A totally different note remaining during the first 8-12 hours

Figure 2. Questionnaire for odour difference level test



SAMPLE NUMBER	DIFFERENCE LEVEL	REMARKS
1	3	
2	8	
3	3	different Top
4	5	
5	5	
7		
8		

DATE

PANELIST NAME A

of evaporation will result in 3. When only some note or reminiscence of standard can be smelled in the sample, 2 is appropriate. 1 needs no comment.

Of course, users of this scale need some training if a nonprofessional panel is used, but two or three sessions of demonstrations can give enough knowledge for routine testing. It is useful to create some kind of scale sample for illustration of meaning of each of the levels.

Results over 7.0 from a group of 5-7 panelists means standard quality, e.g., product can be used for any purpose without any hesitation. Even in control testing of trained panelists when the standard is coded as tested sample, the mean result is usually 7.0-8.0. Limit 5.5 means generally acceptable quality. In most cases, use of the essential oil situated in this quality range should not cause any damage for fragrance composition. Of course, the percentage used and kind of final product involved are important. When such a doubt appears, the best solution is to prepare the fragrance composition with questioned oil and run the same comparison test with the composition standard. According to my experience, the use of a component marked 5.5 means the composition goes over 7.0. There are, of course, exceptions, but they are rare.

The test is prepared and run in the following way. The panelists are presented with coded paper strips of the standard (marked "S" with appropriate number of product) dipped fresh before the session, and 12 hours and 24 hours before it. The same set of blotters is made with tested sample. For volatile products (mainly aroma chemicals), only 12 hours or an even shorter time gap is appropriate (see fig. 2).

The panelists use a questionnaire where apart from marks they give comments, if any, concerning odour quality. This is especially significant when tested oil should be used for a special purpose where some specific odour quality cannot be accepted. Of course, to give such information, a panel should have some practice. Ten to twenty samples can be tested by the same panel during one session. Two sessions per day are possible with a minimum of a two-hour pause. Data from the questionnaires (minimum of five best of seven panelists should participate in each session) are collected in preprinted tables and the final result is calculated.

A simple, but according to our practice, accurate and reliable method is used for calculation of the results. First, the simple arithmetic mean from all marks is calculated and then all individual marks which differ from that mean more than 1.5 are rejected. A second and final mean value is calculated from remaining figures. This simple trick allowed us to avoid significant variations of the result caused by serious mistakes of one or two panelists. In Table III, one can easily see that sample 1 by first mean value should be rejected, and this was caused by a very low mark given by panelist A. A second mean allows the product to be used although the panelist's opinion suggests previous neutralization (acidic top note of the oil).

Table III. Calculation of Results for Odour Difference Level Test

Sample Number	Panelists A	B	C	D	E	F	G	Total	First Mean	Second Mean	Result
1	3	6	4	4	5	6	6	34	4.9	5.5	Accept
2	8	7	7	8	7	7	8	52	7.4	7.4	O.K.
3	3	3	4	5	4	3	6	28	4.0	3.7	Reject
4	5	6	5	3	5	7	5	36	5.1	5.2	Reject
5	5	8	6	7	4	6	6	42	6.0	6.0	Accept

As stated at the beginning, this paper does not contain anything new or totally original. The method presented here was developed on the basis of achievements of sensory analysis and adapted for practical use for everyday routine work. It is especially useful in production practice of small manufacturers who cannot afford expensive teams of perfumers and sophisticated equipment.

Once again it should be said that while much more refined methods are necessary in research works where highly experienced personnel and numerous panels and tests made with hundreds of subjects are the only source of trustworthy result in everyday practice of regular manufacture of oils, compounds and other fragrance products, the method presented here gives excellent results.

Acknowledgement

This paper was presented at the World Council Meeting of the International Federation of Essential Oil and Aroma Trades in London in October 1982.

References

1. W. S. Brud, Methods of evaluation of perfumery materials, TSPK "POLLENA", **17**,333, 1973
2. W. S. Brud, J. Gora, Progress in synthesis of sensory important minor components of essential oils and natural flavours, 1982, in print.
3. H. R. Moskowitz, D. Fishken, Sensory evaluation and psychophysical scaling for cosmetics, toiletries and perfumes, Allured Publishing, Wheaton, IL, USA, 1978
4. N. Barylko-Pikielna, Sensory analysis of food, in Polish, WNT, Warsaw, 1975
5. D. E. R. Frijters, Y. H. Blauw, S. H. Vermaat, Incidental training in the triangular method, Chemical Senses, **7**, 63, 1982.