INTERNATIONAL STANDARD

ISO 13301

First edition
2002-09-15

Sensory analysis — Methodology — General guidance for measuring odour, flavour and taste detection thresholds by a three-alternative forced-choice (3-AFC) procedure

Analyse sensorielle — Méthodologie — Lignes directrices générales pour la mesure des seuils de détection d'odeur, de flaveur et de goût par une technique à choix forcé de 1 parmi 3 (3-AFC)

Reference number
ISO 13301:2002(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.ch
Web www.iso.ch

Printed in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this International Standard may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

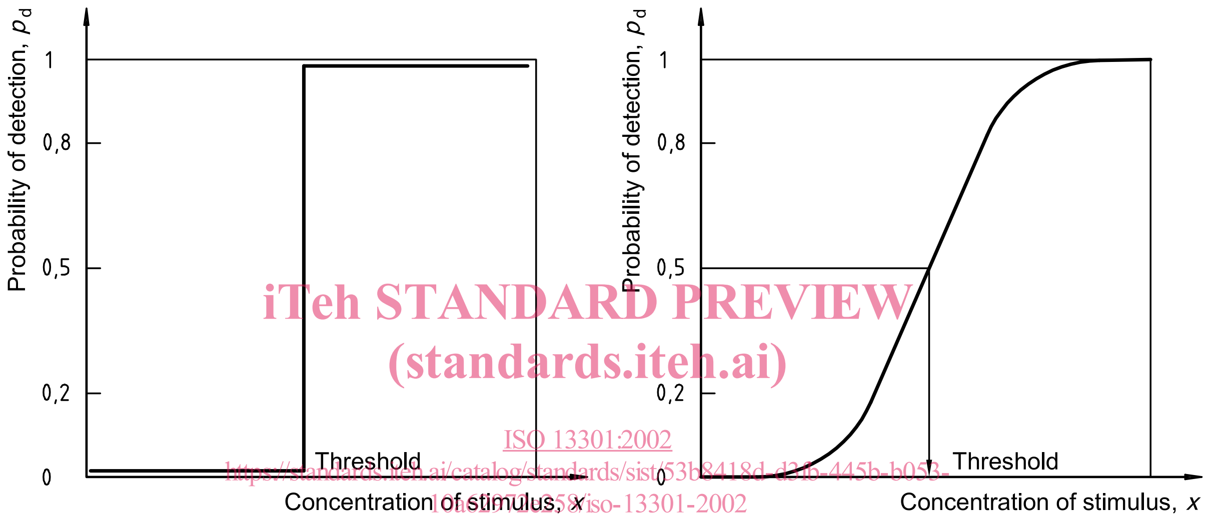
ISO 13301 was prepared by Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 12, *Sensory analysis*.

Annexes A and B of this International Standard are for information only.

Introduction

The concept of "threshold" has wide use in sensory analysis and is often used in the literature on sensory studies of food and drink. Data on sensory thresholds to chemical stimuli are used in sensory studies in two main ways: as measures of the sensitivity of assessors or groups of assessors to specific stimuli; as measures of the ability of chemical substances to evoke sensory responses in assessors. In the first, the value of the threshold is taken as a description of an assessor's performance; in the latter, as a measure of a property of the substance.

The term "threshold" was introduced by 19th century psychophysicists and used to denote a stimulus concentration above which the stimulus could be detected, and below which it could not (see Figure 1a).



a) Traditional notion of threshold **b) Probabilistic nature of threshold**

Figure 1 — Traditional notion and probabilistic nature of threshold

However, in practice the graph of the probability of detection1) against the intensity of the stimulus is always an ogive (see Figure 1b), and it is convenient to assume, for purposes of calculation, that the threshold fluctuates so that a particular stimulus concentration exceeds it on some occasions but not on others. The threshold can then be obtained as an estimate of the median of these momentary values, i.e. as the stimulus concentration for which the probability of detection is 0,5. The threshold defined in this way has analogies with median effect levels used in branches of biology such as pharmacology and toxicology, which are concerned with the effect of chemicals on organisms.

1) This International Standard is based on the use of the 3-AFC method of presenting the stimuli, and the probability of detection, p_d, is modeled as $p_d = 1,5 \times p_c - 0,5$, where p_c is the probability of a correct selection. This is strictly a "guessing model" of the assessor's behaviour. It is not a psychometric model of the assessor's decision process, such as a Signal Detection model, which could also be applied, see Macmillan and Creelman [13].

Where detection thresholds of a particular substance in air or water have been measured in more than one laboratory, the reported values often span two or three orders of magnitude or more (Devos *et al.* [6], Fazzalari [10], van Gemert *et al.* [14]). This range is greater than can be expected from experimental errors alone or from differences in the processing of data; but it probably can be accounted for by difference in concepts of thresholds between laboratories, and differences in experimental procedure. Devos *et al.* [6] suggest a procedure for standardizing detection thresholds in air.

The user needs to be aware that the determination of detection thresholds requires more experimental effort than is at first apparent from this description. Experimental results demonstrate that on repeated testing, the observed individual thresholds tend to decrease, and the difference between individuals likewise tends to decrease. Threshold testing is often an unfamiliar activity, and assessors will improve their sensitivity as they become accustomed to the substance and the mechanics of the test. The 3-AFC procedure requires that assessors can recognize the stimulus. Training programmes require effort but will in turn yield needed information about each assessor's range of partial detection. Results improve as the experimenter learns to tailor the concentrations presented to each assessor's range, see 6.3.

Sensory analysis — Methodology — General guidance for measuring odour, flavour and taste detection thresholds by a three-alternative forced-choice (3-AFC) procedure

1 Scope

This International Standard provides guidance on:

— obtaining data on the detection of chemical stimuli that evoke responses to odour, flavour and taste by a 3-AFC (three-alternative forced-choice) procedure;

— the processing of the data to estimate the value of a threshold and its error bounds, and other statistics related to the detection of the stimulus.

Typically, the procedures will be used in one of the following two modes:

— investigation of the sensitivity of assessors to specific stimuli;

— investigation of the ability of a chemical substance to stimulate the chemoreceptive senses.

(Although experiments may encompass both modes.)

Examples of the first mode would include studies of the differences among individuals or specified populations of individuals in sensitivities and of the effects of age, gender, physiological condition, disease, administration of drugs and ambient conditions on sensitivity. Examples of the latter mode would include:

— studies in flavour chemistry and the impact of specified chemicals on the flavour of foods;

— classification of chemicals for their impact on humans, if present in the environment;

— studies on the relationship of molecular structure to capacity of a chemical to act as a stimulant;

— quality assurance of gaseous effluents and of water, foods and beverages;

— studies in the mechanism of olfaction.

In both modes the way in which probability of a correct response changes with intensity of stimulus, i.e. the slope of the dose/response curve, could be an important aspect of the study as well as the threshold value, and the data processing procedures described here provide this information.

The focus of this International Standard is on data requirements and on computational procedures. Regarding the validity of the data, the text is restricted to general rules and precautions. It does not differentiate between detection and difference thresholds; fundamentally, the procedures measure a difference threshold because a test sample is compared with a reference sample. Typically, the reference sample is not intended to contain the stimulus under investigation, but the Guidelines do not exclude experimental design in which the reference could contain the stimulus, or it might not be known if the reference contains the stimulus. The Guidelines do not measure a recognition threshold as defined in ISO 5492. They do not address the standardization of methods of determining air quality as a European Standard is in preparation [9].

2 Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this International Standard. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid International Standards.

ISO 5492:1992, *Sensory analysis — Vocabulary*

ISO 6658:1985, *Sensory analysis — Methodology — General guidance*

ISO 8586-1:1993, *Sensory analysis — General guidance for the selection, training and monitoring of assessors — Part 1: Selected assessors*

ISO 8586-2:1994, *Sensory analysis — General guidance for the selection, training and monitoring of assessors — Part 2: Experts*

ISO 8589:1988, *Sensory analysis — General guidance for the design of test rooms*

3 Terms and definitions

For the purposes of this International Standard, the terms and definitions given in ISO 5492, as well as the following, apply.

3.1
stimulus
substance that may or may not cause a sensation, detectable by one or more of the senses, depending on the amount present

3.2
medium
any material used to dissolve, dilute, disperse or sorb a stimulus whose threshold is to be measured

3.3
reference sample
quantity of the medium containing no added stimulus

3.4
test sample
quantity of the medium to which a stimulus has been added at a known concentration

3.5
three-alternative forced-choice (3-AFC) test
test of discrimination in which the assessor is presented with three samples, one of which is a test sample containing a nominated stimulus familiar to the assessor, the other two being references, and where the assessor is instructed to indicate the test sample

3.6
presentation
set of three samples forming a 3-AFC test

3.7
threshold model
model of sensory detection where a stimulus presented on a particular trial is either detected (resulting in a correct response) or is not detected (resulting in a response being made at random)

3.8
signal-detection model
model of sensory detection where a stimulus presented on a particular trial provides some level of evidence of its presence

NOTE The evidence contributes to a decision by the assessor about the presence or absence of the stimulus.

3.9
detection threshold
the lowest intensity of a sensory stimulus that has a probability of detection of 0,5 under the conditions of the test, as calculated from the threshold model

3.10
individual threshold
detection threshold of a single assessor

3.11
average threshold
average (whose type must be specified, e.g. arithmetic mean, geometric mean, or median) of individual thresholds

3.12
group threshold from pooled data
estimate obtained by using the sum of outcomes for a particular group of assessors at each concentration of the stimulus as input when fitting the statistical model

4 Principles

4.1 Experimental procedures

The stimulus is formulated in the medium at a specified concentration and is presented along with a pair of reference samples to the assessor. The assessor is required to select one of the samples as containing the stimulus or having the stimulus at a greater concentration. The assessor must make a selection. It is a requirement of the 3-AFC test that the assessor be able to recognize the stimulus.

Typically the stimulus is dissolved in air or water. It is unlikely that a gas other than air will be used as a gaseous medium in tests with human assessors, but solvents other than water, solutions in water or other solvents, or solids, e.g. foods, can be used as liquid or solid medium to dilute the stimulus as the experiment dictates. It is essential that the medium be homogeneous so that the members of the pair of references are identical, and the same in all presentations.

The stimulus is presented at several concentrations. The presentations are replicated, at each concentration, a sufficient number of times to achieve a desired precision of the threshold and parameters of the mathematical model. The nature of the replications within assessors, across assessors, and combinations of the two are set by the experimental design of the study.

4.2 Data processing

The outcome of a presentation is a binary result – the sample nominated by the assessor is the test sample (a correct selection) or is one of the references (an incorrect selection). The number of correct selections is summed over the number of presentations at each stimulus concentration and forms, along with the total number of presentations and the stimulus concentration, the data to be processed for obtaining the derived statistics. The statistical model is that the number of correct selections at a particular concentration comes from a binomial distribution.

For the 3-AFC test, the threshold is the concentration of the stimulus at which the proportion of correct selections is equal to 2/3, i.e. 50 % above chance. The data, as proportions of correct selections, can simply be inspected and interpolated to derive this point, but a more accurate estimate of the threshold, and its bounds, can be obtained by fitting a mathematical model to the data. A logistic model is used in these guidelines, and the model is fitted by a maximum likelihood procedure, or alternatively, by a least squares procedure. The fitting estimates the two parameters of the model, one a location parameter, the other a shape parameter. The former locates the fitted curve on the stimulus continuum, the latter determines the steepness of the curve. The fitted curve allows estimates of proportions of detection other than 50 % to be derived.

The simplest model to fit is one in which the distribution of proportion of correct selections comes from a single, approximately normal, distribution. This would typically be the case where the data come from replications within a single assessor. A single logistic function can then be adequately fitted, that is, one with a single pair of values for

the parameters of the curve. It is not uncommon for the sensitivities to chemicals to be not normally distributed, or even symmetrically distributed, among assessors. For some chemical stimuli the distributions are distinctly bimodal, but deviations from a normal distribution are difficult to demonstrate unless measurements are made with a large sample of assessors, typically more than 100. A single logistic function will not be an adequate fit to data that come from a distribution which deviates significantly from a single, normal distribution, but the mathematical model can be extended to accommodate these cases.

5 Experimental procedures

5.1 Preparation of samples

5.1.1 General precautions

See ISO 6658. Ascertain that stimulus and medium are stable over the duration of the study and are non-toxic and nonallergenic. Ascertain that they are representative of the purpose of the study, e.g. exhaust gases may vary with the process generating them, and chemical substances may require purification to remove off-flavours or irritants from the molecule to be studied. Prepare a large enough homogeneous quantity of both stimulus and medium to ensure that assessors receive identical presentations with exception of the concentration of stimulus and its position in the set. Prepare the samples in a facility that complies with ISO 8589. Use containers that do not adsorb the test chemical or contribute odour or taste. Make certain that the presence or absence of the stimulus cannot be detected visually or by any means available to an assessor other than the chemical senses. Store samples away from light and heat when not in use.

5.1.2 Gases

Collect or prepare stimulus and medium in vessels such as teflon- (PTFE) coated bottles or balloons. If the stimulus is an inodorous gas containing an odorous impurity, flush the vessel and associated tubing and valves several times with a fresh sample in order to saturate the walls. For the same reason, and to avoid volume changes, maintain a constant temperature near that to be used when presenting the gases to the assessors. Use smoothbore PTFE-coated tubing and valves free from points of sudden pressure change.

5.1.3 Liquids

For stimuli to be presented in an aqueous medium, make certain that complete dissolution can be obtained and maintained for the duration of the experiment. For partially hydrophobic substances, prepare the first dilution stage in ethanol or ethylene glycol purified with activated carbon to remove off-odours. Note that distilled water and absolute alcohol often contain strong odours; use food grade product instead and purify with activated carbon if required. Present fully hydrophobic substances in a nonaqueous solvent such as odourless liquid paraffin or dinonyl phthalate and avoid plastic containers as the substance may dissolve in the polymer. When preparing sequential dilutions, be aware that the higher the dilution, the larger the proportion of the stimulus that may be lost by adsorption to the vessel wall. As far as is possible, prepare each dilution by microsyringe or equivalent, directly from a stock solution, and avoid sequences of preparing each dilution from the preceding sample.

5.1.4 Solids

The medium of interest is typically a food such as cheese, fish or meat. Unless a technique exists whereby the solid can be dissolved and reconstituted, finely divide or comminute it before adding the stimulus in a suitable solvent, then mix well and allow time for the chemical to diffuse within the matrix before preparing the samples for presentation to the assessors. Code each aliquot, e.g. with a random, 3-digit number.

5.2 Selection of concentrations of the stimulus

Present a series of 3-AFC presentations of which each concentration is greater than the preceding one by approximately a factor denoted by X. Be guided by the acceptable size of the error of the threshold estimate: typically choose $X \approx 3$-5 for approximate studies and $X \approx 2$ for higher precision. For each assessor, choose a strategy of experimentation that will result in defining the ogive of the logistic model at points distributed over his or her range of partial detection. The most effective data points are those corresponding to 45 % to 90 % correct selection in the test, i.e. $p_d = 0{,}18$ to $0{,}85$.

For economy of sample and assessor's time, begin by locating the concentration range of interest for each assessor using a large factor X. Observe that these initial tests also serve to demonstrate the mechanics of the test and to teach the assessors how to recognize the stimulus when it is above their range of partial detection.

Proceed with the definitive set of 3-AFC presentations at concentrations tailored to each assessor using a low factor X. If on completion it is found that the data do not adequately define an assessor's ogive, administer additional concentration levels until this is the case. Regularly ask an assessor to describe the nature of the detected stimulus so as to guard against lapses of memory for it. Interrogation may also uncover an unintended sequence of correct replies caused by chance and not by detection; e.g. a series of 3 chance hits will occur once in 27 tests.

5.3 Presentation of samples

5.3.1 Preparation

Present samples with assessors seated in booths (see ISO 8589) and observe the rules of good sensory practice as described in ISO 6658. Code samples with three-digit random numbers, or place samples in a prearranged pattern, e.g. side-by-side in front of the assessor with the first sample on the left, using the identical pattern on the response sheet. To avoid positional bias, balance the three combinations of orders of presentation, AAB, ABA, BAA, across the assessors. Instruct assessors to minimize sensory fatigue by ingesting a minimum quantity of any sample that exhibits above-threshold concentration and by allowing sufficient time for sensory recovery between samples.

5.3.2 Gases

Present samples using an olfactometer such as those described in [8] and [12].

5.3.3 Liquids

Present non-volatile chemicals dissolved in purified water or in a flavourless solvent. Use containers that do not absorb the chemical, e.g. 100 ml glass beakers one quarter full. Present volatile chemicals in stoppered, wide mouthed containers suitable for sniffing or sipping, or in flexible closed containers, e.g. 250 ml squeeze bottles suitable for delivering a measured volume of headspace or liquid into the nostrils or mouth, see [4], [7] and [11]. If the medium is a beverage, use the type of container that is customary for sensory evaluation of the product.

5.3.4 Solids

If the medium is a food, present the samples in the form that is customary for sensory evaluation of the product.

5.4 Training of assessors

For most purposes, the threshold of interest is that of an informed observer, trained by repeated exposure to detect the substance in question whenever its presence is perceivable, e.g. as a pollutant in air or water, or as a component or taint of the flavour of a food or beverage. Familiarity with the substance is also a requirement in the 3-AFC test. Inadequate training may artificially extend the observed range of thresholds upwards by 1-2 orders of magnitude. An artificial extension downwards can result from overtraining, when assessors become adept at discovering the treated sample by means other than its flavour. If the threshold sought is that of a casual observer, e.g. for a warning agent in household gas, untrained assessors and mild distraction (e.g. noise) may be used and the triangle test or paired comparison substituted for the 3-AFC test.

A training programme can be by presentation of the stimulus monadically at high concentrations, then at two or more concentrations with the assessor requiring to rank them, then as 3-AFCs while locating the assessor's range of partial detection. Observe that initial thresholds decrease with practice and should tend to stabilize after 3 to 5 tests and that individual assessors may differ in their basic sensitivity to the substance in question by a factor of two or three orders of magnitude, or more.

5.5 Selection of assessors

5.5.1 General

Select assessors to meet the objectives of the investigation, following the guidelines given in ISO 8586-1 and ISO 8586-2.

5.5.2 Individual threshold

The test may be made, e.g. to compare an individual's threshold with a literature value, with a previously determined value under different circumstances, or with his or her thresholds for other substances. The test may be made to diagnose anosmia or hyperosmia, ageusia or hypergeusia.

5.5.3 Distribution of thresholds

The experimenter may wish to know the distribution of thresholds within a population. The group tested might itself be a sample drawn from a larger population, or it may be all members of a selected population, e.g. members of a testing panel. Selection of populations is outside the scope of this International Standard, but the experimenter should carefully define the population, or the sample of the population, under study. For the presentation of the results, see 6.7.

5.5.4 Measurement of thresholds of stimuli

The value of a group or average threshold for a stimulus is valid only for the panel of assessors used in the trials and the experimenter should be cautious in extrapolating the results outside of this panel. The experimenter should select the panel to meet the objectives and purposes of the measurements. For example, a study of the relative organoleptic properties of members of a set of chemicals could be carried out using a small panel of selected assessors, whereas a study of the properties of potential flavouring compounds in foods might require a larger panel which is representative of a particular population.

The number of assessors and the number of presentations to achieve a required precision of estimates are matters to be considered together. When small numbers of assessors are being used, it will be necessary to replicate presentations over assessors to generate sufficient data, whereas single presentations at each, or perhaps just some, concentrations to each assessor might be adequate for large panels.

5.6 Design of the experiment

5.6.1 Individual threshold

The most effective range of concentrations for estimating the parameters of the logistic is between 45 % and 90 % correct selections. Within this range the main determinant of precision of the estimates is the total number of presentations assuming they are roughly balanced around the threshold. Table 1 shows factors for approximate error bounds relative to the estimate of the threshold, in original concentration units. See also annex A.

Table 1 — Guide for determining the number of presentations required for a desired precision of an estimate of the threshold

Total number of presentations	40	60	80	100	120	160	200
Error bound relative to threshold	2,5	2,2	2,0	1,8	1,7	1,6	1,5

The bounds are obtained by both dividing and multiplying the estimate of the threshold by the factors in Table 1; e.g. if the threshold obtained with 80 presentations was 2,4 ppm (2,4ml/m^3), the bounds would be 1,2 ppm to 4,8 ppm. Precision increases only slowly above 200 presentations and the improvement is probably not worth the extra effort. A sequential strategy is effective. After a few replicate presentations at each concentration, fit the logistic and calculate the threshold and error bounds. Carry out more replicates at concentrations within the most effective range determined from the fitted logistic, and repeat until the desired precision is obtained.

