Designation: E 1432 – 91 (Reapproved 1997)

Standard Practice for Defining and Calculating Individual and Group Sensory Thresholds from Forced-Choice Data Sets of Intermediate Size[1]

This standard is issued under the fixed designation E 1432; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ϵ) indicates an editorial change since the last revision or reapproval.

INTRODUCTION

The purpose of this practice is to determine individual sensory thresholds for odor, taste, and other modalities and, when appropriate, calculate group thresholds. The practice takes as its starting point any sensory threshold data set of more than 100 presentations, collected by a forced-choice procedure. The usual procedure is the Three-Alternative Forced-Choice (3-AFC), as exemplified by Dynamic Triangle Olfactometry. (Practice E 679 is suitable as a rapid method of determining an approximate group threshold for limited-size data sets of 50 to 100 3-AFC presentations.)

Collection of the data is not a part of this practice. The data are assumed to be valid; for example, it is assumed that the stimulus is defined properly, that each subject has been fully trained to recognize the stimulus and did indeed perceive it when it was present above his or her momentary threshold, and that the quality of dilution air did not vary. Assurance that the data are valid must be obtained (*1*) by repeating the threshold determination several times and (*2*) by having the experimenter interview each subject about the certainty of his or her perception.

It is recognized that precise threshold values for a given substance do not exist in the same sense that values of vapor pressure exist. A panelist's ability to detect a stimulus varies as a result of random variations in factors such as alertness, attention, fatigue, events at the molecular level, etc., the effects of which can usually be described in terms of a probability function. At low concentrations of an odorant or tastant, the probability of detection by a given individual is typically 0.0 and at high concentrations it is 1.0, and there is a range of concentrations in which the probability of detection is between these limits. By definition, the threshold is the concentration for which the probability of detection of the stimulus is 0.5 (that is, 50 % above chance, by a given individual, under the conditions of the test).

Thresholds may be determined (*1*) for an individual (or for individuals one by one), and (*2*) for a group (panel). While the determination of an individual threshold is a definable task, in which precision of the result is mainly a question of letting the individual take enough tests at enough concentrations, careful consideration for a group needs to be given to the additional question of what the group is meant to represent.

The user must keep in mind the large degree of random error associated with estimating the probability of detection from less than approximately 500 3-AFC presentations. (A presentation is a set of samples, usually two, three, or five, judged as a unit by a single individual.) The reliance that may be placed on the results can be increased greatly by enlarging the panel and by replicating the tests.

1. Scope

1.1 The definitions and procedures of this practice apply to the calculation of individual thresholds for any stimulus in any medium, from data sets of intermediate size, that is, consisting

[1] This practice is under the jurisdiction of ASTM Committee E-18 on Sensory Evaluation of Materials and Products and is the direct responsibility of Subcommittee E18.04 on Fundamentals of Sensory.

Current edition approved Aug. 15, 1991. Published October 1991.

of more than 20 to 40 3-AFC presentations per individual. A group threshold may be calculated using 5 to 15 individual thresholds.

1.2 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Principles

2.1 The 3-AFC procedure is one of the set of n-AFC procedures, any of which could be used, in principle, for the measurement of sensory thresholds, as could the duo-trio, the triangular, and the two-out-of-five procedures.

2.2 For calculation of the threshold of one individual, this practice requires data sets taken at five or more concentration scale steps, typically six or seven steps, with each step differing from the previous step by a factor usually between 2 and 4, typically 3.0. The practice presupposes that the range of concentrations has been selected by pretesting, in order to ensure that the individual's threshold falls neither outside nor near the ends of the range, but well within it. At each concentration step, the individual must be tested several times, typically five or more times.

2.3 Individual thresholds, as determined in 2.2, may be used for calculation of a group (or panel) threshold. The size and composition of the panel (usually 5 to 15 members, preferably more) is determined according to the purpose for which the threshold is required and the limitations of the testing situation (see 7.2).

2.4 Pooling of the data sets from panel members to produce a single step calculation of the panel threshold is not permitted.

3. Referenced Documents

3.1 *ASTM Standards:*

E 122 Practice for Choice of Sample Size to Estimate a Measure of Quality for a Lot in Process[2]

E 679 Practice for Determination of Odor and Taste Thresholds by a Forced Choice Ascending Concentration Series Method of Limits[3]

3.2 *French Standard:*

NF X43-101, Air Quality; Method of Measuring the Odor Intensity of Gaseous Effluents; Determination of the Dilution Factor to Perception Threshold[4]

4. Terminology

4.1 *Definitions of Terms Specific to This Standard:*

4.1.1 *Three-Alternative Forced-Choice (3-AFC) test procedure*—a test presentation used in many threshold tests. For example, in odor testing by Dynamic Triangle Olfactometry, the panelist is presented with three gas streams, only one of which contains the diluted odorant, while the other two contain odorless carrier gas. The panelist must indicate the one containing the added substance. (The 3-AFC procedure is different from the classical Triangle test, in which either one or two of the three samples may contain the added substance.)

4.1.2 *model*—an abstract or concrete analogy, usually mathematical, which represents in a useful way the functional elements of a system or process. In short, the experimenter's theory of what is guiding the results observed.

4.1.3 *statistical model*—a model assuming that the principal factor causing the results to deviate from the true value is a random error process. This can usually be described in terms of a probability function, for example, a bell-shaped curve, symmetrical or skewed. Errors are binomially distributed in the 3-AFC test procedure.

4.1.4 *threshold, detection*—the intensity of the stimulus that has a probability of 0.5 of being detected under the conditions of the test. The probability of detection at any intensity is not a fixed attribute of the observer, but rather a value which assumes that sensitivity varies as a result of random fluctuation in factors such as alertness, attention, fatigue, and events at the molecular level, the effects of which can be modeled by a probability function.

4.1.5 *individual threshold*—a threshold based on a series of judgments by a single panelist.

4.1.6 *group threshold*—the average, median, geometric mean or other agreed measure (or an experimentally determined measure) of central tendency of the individual thresholds of the members of a group (panel). The meaning and significance of the term depends on what the group is selected to represent (see 7.2.2).

4.1.7 *scale step factor*—for a scale of dilutions presented to a panel, the factor by which each step differs from adjacent steps.

4.1.8 *dilution factor*—the following applies to flow olfactometry: If F_1 represents the flow of odorless gas which serves to dilute the flow of odorant, F_2, the dilution factor, Z, is given by:

$$Z = \frac{F_1 + F_2}{F_2} \tag{1}$$

where Z is dimensionless. F_1 and F_2 may be expressed, both in units of mass, or (preferably) both in units of volume; the report should state which. The term Z_{50} represents the dilution factor to threshold. The letter Z is used in honor of H. Zwaardemaker, a Dutch scientist and early investigator of olfactometry. Alternate terminology in use is as follows: dilution-to-threshold ratio (D/T or D-T); odor unit (OU); and effective dose (ED).

5. Summary of Practice

5.1 From a data set according to 2.2, calculate the threshold for one individual graphically or by linear regression according to 5.2, or by using a model fitting computer program according to 5.3.

5.2 Obtain the threshold in 5.1 by first calculating the proportion correct above chance for each concentration step. This is accomplished by deducting, from the proportion of correct choices, the proportion that would have been selected by chance in the absence of the stimulus (see 8.1.2). Then, for each individual calculate that concentration which has a

[2] *Annual Book of ASTM Standards*, Vol 14.02.

[3] *Annual Book of ASTM Standards*, Vol 15.07.

[4] English translation available from AFNOR, Tour Europe, Cedex 7, 92080 Paris La Défense, December 1986.

probability of 0.5 of being detected under the conditions of the test. This is the individual threshold.

5.3 Alternatively obtain the threshold in 5.1 directly from the proportion of correct choices by non-linear regression using a computer program, as described in 8.2.2.

5.4 Always report the individual thresholds of the panelists. Depending on the purpose for which a threshold is required (see 7.2), and on the distribution found, a group threshold may be calculated as the arithmetic or geometric mean, the median, or another measure of central tendency, or it may be concluded that no group threshold can be calculated (see 7.4).

6. Significance and Use

6.1 Sensory thresholds are used to determine the potential of substances at low concentrations to impart odor, taste, skinfeel, etc. to some form of matter.

6.2 Thresholds are used, for example, in setting limits in air pollution, in noise abatement, in water treatment, and in food science and technology.

6.3 Thresholds are used to characterize and compare the sensitivity of individuals or groups to given stimuli, for example, in medicine, ethnic studies, and the study of animal species.

7. Panel Size and Composition Versus Purpose of Test

7.1 *Panel Size and Composition*—Panel variables should be chosen as a function of the purpose for which the resulting threshold is needed. The important panel variables are as follows:

7.1.1 Number of tests per panelist,

7.1.2 Number of panelists,

7.1.3 Selection of panelists to represent a given population, and

7.1.4 Degree of training.

7.2 *Purpose of Test*—It is useful to distinguish the following three categories:

7.2.1 *Comparing an Individual's Threshold With a Literature Value*—The test may be conducted, for example, to diagnose anosmia or ageusia, or to study sensitivity to pain, noise, or odor. This is the simplest category requiring a minimum of 20 to 40 3-AFC presentations to the individual in question (see 2.2). A number of training sessions may be required to establish the range of concentrations that will be used and to make certain that the individual is fully familiar with the stimulus to be detected as well as the mechanics of the test.

7.2.2 *A Population Threshold is Required*, for example, the odor threshold of a population exposed to a given pollutant, or the flavor threshold of consumers of a beverage for a given contaminant. In this case, recourse must be had to the rules of sampling from a population (see Ref **(1)**[5] and Practice E 122), which require the following:

(1) That the population be accurately defined and delimited,

(2) That the sample drawn be truly random, that is, that every member of the population has a known chance of being selected, and

(3) That knowledge of the degree of variation occurring within the population exists or can be acquired in the course of formulating the plan of sampling.

7.2.2.1 In practice, the cost and availability of panelists places serious limitations on the degree to which population factors affecting thresholds, for example, age groups, gender, ethnic origin, well versus ill, smoker versus nonsmoker, trained versus casual observers, etc., can be covered. The experimenter is typically limited to panels of 5 to 15, with each receiving 20 to 40 3-AFC presentations, for a total of 100 to 600 presentations. If the resulting thresholds are to have validity for the population, the experimenter should include the following steps:

(1) Calculate and tabulate the thresholds for each individual;

(2) Repeat the test for those individuals (outliers) falling well beyond the range of the rest of the panel;

(3) For any individuals whose threshold at first did not fall well within the range of samples presented to them, adjust the range and repeat the test; and

(4) If needed to obtain a desired level of precision, repeat the test series with a second or third panel sampled from the same population of interest.

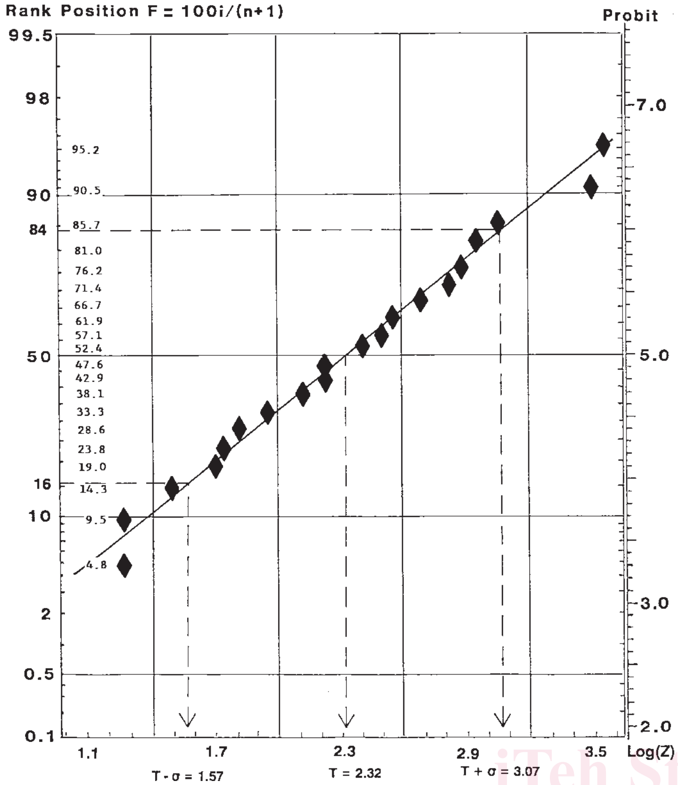
7.2.2.2 Thresholds vary with age, and one approach to the question of a generalizable population value is to attempt to estimate the threshold of healthy 20-year olds. According to Amoore **(2)**, between the ages of 20 to 65, odor threshold concentrations double for approximately each 22 years of age.

7.2.3 *The Distribution of Thresholds in the Population is Required*, for example, to determine what proportion of the population is affected by a given level of a pollutant, or, conversely, to determine which concentrations of a pollutant will affect a given percent of a population. The requirements for testing are the same as in 7.2.2, except that it is even more important to cover the range well, for example, to repeat the tests for those individuals whose thresholds fall in thinly populated parts of the panel range. Consideration should be given to increasing the number of presentations per concentration from 5-7 to 7-10 for such panel members. If the individual thresholds are plotted as in Fig. 1, any sector requiring study will be apparent from the graph.

7.3 *Trained Versus Casual Observers*—Thresholds should normally be determined for observers trained by repeated exposure to detect the stimulus in question whenever it is present; however, if the threshold sought is that of a casual observer (for example, for a warning agent in household gas), naive panelists and mild distraction (for example, noise) may be used (see Ref. **(3)**).

7.4 *Choice of the Measure of Central Tendency*—The report should contain a table or graph providing the individual thresholds of each observer. If a group threshold is required, the measure of central tendency chosen should be that which best represents the distribution obtained. In a few cases (Fig. 2), the results form a symmetrical, bell-shaped distribution, and the arithmetic mean, or median may be used. With sensory

[5] The boldface numbers in parentheses refer to the list of references at the end of this standard.



NOTE 1—This probability graph shows 20 panelists sorted by rank as described in 9.3.2. Data are adapted from French Standard X 43-101. Group threshold = T = 50 % point = log(Z_{50}) = 2.32. Group standard deviation from % and 84 % points = σ = (3.07 − 1.57)/2 = 0.75 in log(Z) units. The 99 % point is off the graph but can be calculated as 2.32 + (0.75 × 2.327) = 4.07, where 2.327 is the % point on the abscissa of the normal curve of error.

FIG. 1 Group Threshold by Rank-Probability Graph

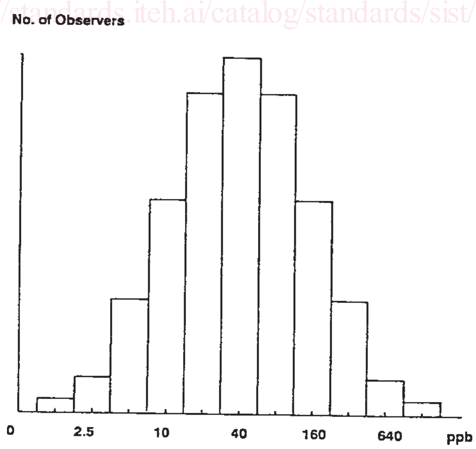


FIG. 2 Symmetrical, Bell-Shaped Distribution

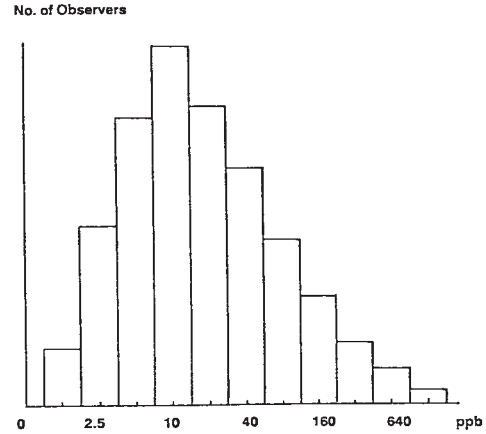


FIG. 3 Skewed Distribution

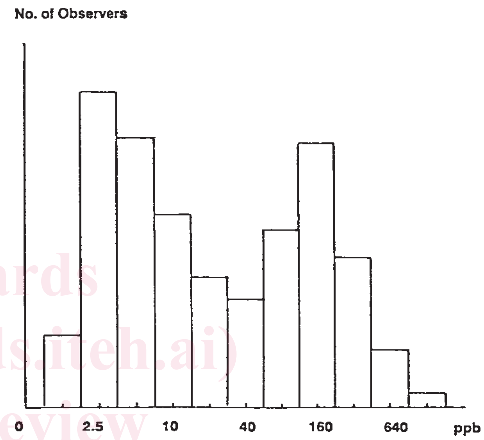


FIG. 4 Bi-Modal Distribution

data, the distribution is typically skewed (Fig. 3); however, it may be normalized by converting the concentration units to log concentration, which is equivalent to converting the arithmetic mean into the geometric mean. If, as is often the case, the distribution is irregular but approaches normality, the 50 % point of a log-probability graph (see Fig. 1) is the appropriate measure. Conversion of the concentration scale into double logarithms (log of log) is occasionally needed to normalize a distribution. However, if the data show a bi-modal (Fig. 4) or multi-modal distribution, indicating the existence of sub-populations with different thresholds, the distribution cannot be normalized; instead, an attempt may be made to estimate the size and group threshold of each sub-population.

7.5 *Group Standard Deviation*—To characterize the dispersion of thresholds around the population mean, a group standard deviation, σ, may be estimated as shown in the examples, Figs. 5 and 6, and Fig. 1. This is permissible only if the distribution is normal or near-normal, or has been normalized.

8. Procedure

8.1 *Tabulation and Transformation of Data*—See Table 1.

8.1.1 *Example 1: Threshold of Substance X in Purified Water*—Six observers took part; each was tested at five or more concentrations chosen in advance[6] to bracket each person's threshold; each took six tests per concentration, for a total of 30 to 36 presentations per observer (Table 1):

[6] For example, by giving the person a single test (or a few tests) of the concentrations 640, 160, 40, 10, and 2.5 ppb.