RECOMMENDATION ITU-R BS.562-3*,**

Subjective assessment of sound quality

(1978-1982-1986-1990)

The ITU Radiocommunication Assembly,

considering

a)   that subjective listening tests permit assessment of the degree of annoyance caused to the listener by any impairment of the wanted signal during its transmission between the originating source and the listener;

b)   that such an assessment implies that a programme sequence which has been subjected to impairment should be compared with the original sequence, which should be of "excellent quality" or with "imperceptible impairment";

c)   that to make these assessments comparable one with the other, the conditions of listening, the composition of the team of assessors and the programme sequences should, as far as possible, be standardized;

d)   that it would be desirable for a single scale of assessment to be available for both sound and television programmes,

recommends

1   that the grading scales given below should be used for the subjective assessment of the quality or of the impairment, of the quality of sound in broadcasting (for television pictures, see Recommendation ITU-R BT.500). The nature and object of the tests will determine which of the two scales is the more appropriate.

---

*   This Recommendation is of interest to the Telecommunication Standardization Study Group 9.

**   Radiocommunication Study Group 6 made editorial amendments to this Recommendation in 2002 in accordance with Resolution ITU-R 44.

1.1 Five-grade quality and impairment scale*

TABLE 1

Quality	Impairment
5 Excellent	5 Imperceptible
4 Good	4 Perceptible, but not annoying
3 Fair	3 Slightly annoying
2 Poor	2 Annoying
1 Bad	1 Very annoying

1.2 Seven-grade comparison scale

For certain types of subjective tests it may be more convenient to use a comparison scale, in which case the following seven-grade scale should be used:

TABLE 2

3 Much better
2 Better
1 Slightly better
0 The same
−1 Slightly worse
−2 Worse
−3 Much worse

2 Presentation of results

The results obtained by the use of expert listening panels should be presented separately from those provided by non-expert panels. Details should be given of listening conditions and sound levels; any statistical methods used to analyse the test results should be described.

NOTE 1 – The general considerations governing the assessment procedure, the listening conditions, the selection of assessors, etc., are given in Annex 1.

_____

* In view of the large number of documented results which have been obtained using a six-grade scale, it is desirable to have a means of converting these results to the above five-grade scales so that the data can still be used. Uncertainties arise in attempting to convert results obtained with one scale into another. However, as a first approximation, the following linear relationship can be used to convert a grade, A_6, obtained in an experiment using a six-grade scale into a grade, A_5, in the corresponding five-grade scale:

$$A_5 = 5.8 - 0.8\,A_6$$

When results which have been converted by means of the above equation are presented, it should be stated that this conversion has been carried out.

ANNEX 1

1 General

Programme sequences used for testing should include silent intervals so that, in the absence of the wanted signal, the subjective assessment of the impairment caused by noise in the system is not excluded. On the other hand, the tests should exclude any assessment of defects, the audible effects of which might, in certain cases, not be objectionable and which might even give a subjective impression of improved quality. The programme sequences should, therefore, be free of any audible defects similar to those produced in the system under test, but where this is impracticable the consequent limitations on the validity of the results should be clearly indicated.

For tests using the five-point grading scale mentioned in § 1.1 of the Recommendation, a system of lights should be used to indicate to the listener the source (impaired or unimpaired) of the programme he is hearing. To test the listener's attention and consistency, some tests in which the impaired condition would be replaced by the unimpaired condition should be included randomly, without informing the listener. For tests involving the use of the seven-grade comparison scale, no indication should be given which may bias the judgement of the listener. However, for the comparison tests, it could be useful from time to time to give a reference condition which may be the unimpaired source, and this reference condition may be indicated by a light.

The amount of data which needs to be collected depends upon such interrelated factors as the degree of statistical confidence which is needed in the result, the standard deviation of the measurements, and the relative magnitude of the effect which it is required to detect. The following suggestions are intended as guide-lines to assist in formulating a considered experimental design.

2 Selection of listening panel

Although in a normal listening audience there will be some expert listeners*, the proportion of them is likely to be so small that it is proper to concentrate the objective of laboratory tests on the opinions of non-experts, because the use of experts could lead to results which are much more critical than would be obtained with non-expert listeners. The choice of test listening conditions should be more critical than average, but not unduly so. As tests with non-expert listeners tend to be lengthy, it is often desirable that a quick test should be carried out by experts. In this case, a smaller number of listeners can be used. However, it should be noted that in certain circumstances tests carried out with expert listeners may not be a satisfactory substitute for tests carried out by non-experts. In cases of doubt, the relationship between expert and non-expert opinion should be investigated.

The minimum number of non-expert listeners should normally be twenty whilst the minimum number of expert listeners should normally be ten. In all cases, the number and category of listeners and the duration of the tests should be stated. Whenever the system is intended for high-quality sound broadcasting or reproduction, expert listeners should be used exclusively.

---

* The term "expert listeners" is considered to apply to listeners who have had recent extensive experience of assessing sound quality or impairment, particularly of the type being studied in the subjective tests.

3   Test procedure and duration

Because of the extreme unreliability of the long or medium-term aural memory, the instantaneous comparison method should always be used.

For tests using the five-grade quality or impairment scales each process involves the repetition, four times consecutively, of the same programme sequence in the following order:

1. original sequence,

2. same sequence, impaired

3. original sequence (repeated),

4. same sequence, impaired (repeated).

Each programme sequence should not last longer than 15 to 20 s; it may be very short (a few seconds) for some tests. In the case where the sequence is a musical item, the phrase should not appear to be interrupted. The interval between presentation 1 and 2 and between 3 and 4 should be about 0.5 to 1 s, while the interval between 2 and 3 should be somewhat longer, for example 1.5 s. The exact time should depend upon the type of programme. The switching device should not introduce audible interference.

The programme sequences and impairments should be presented in random order subject to the condition that the same sequence should never be presented on two successive occasions with the same or different levels of impairment.

No session with any one listener should last for more than about 15 to 20 min without interruption. If the sessions must be consecutive, they should be separated by rest periods of roughly the same length.

For tests with the seven-grade comparison scale involving two impaired conditions, a similar set of presentations can be used, the order being:

1. Condition 1,

2. Condition 2,

3. Condition 1 (repeated),

4. Condition 2 (repeated).

Conditions 1 and 2 should be interchanged on a random basis. In addition, a reference condition may be presented at the beginning of each four presentations and, in this case, a definite indication (such as the use of a light signal) should be given, that this item is the reference condition.

4   Choice of programme sequences

Depending on the precise objective fixed and in particular on the category of the sound-programme transmission or reproduction system tested, the following programme sequences should be used:

–   either a representative selection of typical programme material,

–   or, a selection of a few sequences picked deliberately for their highly critical behaviour with respect to the impairments introduced by the system being tested. For example, when assessing protection ratios, a suitably critical test sequence would be speech on the wanted programme impaired by "pop" music on the unwanted programme.

Whenever the system is intended to carry high-quality sound, the second type of programme sequence should be used. To ensure the comparability of test data obtained in different places and at different times, preferably the same programme sequences should be used. The subjective quality assessment material (SQAM) compact disk adopted and published by the EBU provides an appropriate source of high-quality digital programme material from which suitable items may be chosen for this purpose.

In any event, the artistic or intellectual content of a programme sequence should be neither so attractive nor so disagreeable or wearisome that the listener is distracted from his purpose.

5 Choice of reproduction device

Depending on the category of impairment to be assessed, either headphones or loudspeakers may be used.

It has been shown that certain quality shortcomings are more clearly perceptible in the case of headphone reproduction than in the case of loudspeaker reproduction. For example, the signal-to-noise ratio required for noiseless listening using headphones exceeds the figure obtained using loudspeakers at the same sound intensity by as much as 10 dB. Similar differences occur in the case of the quality losses caused by clicks (caused by bit errors in digital transmission), by quantizing distortions, non-linearity distortions, phase distortions, etc.

However, other quality shortcomings are more clearly perceptible in the case of loudspeaker reproduction. Especially those influences which affect the characteristics of the stereophonic sound-image between the loudspeakers should be assessed by means of loudspeaker reproduction. For example, this is due to the quality losses caused by any difference between the A and B channels.

In order to make the assessments as far as possible comparable with one another, it may be advisable to use headphones. Because headphone reproduction is independent of the geometric and acoustic properties of listening and control rooms, it can, in principle, be defined with great accuracy and can easily be reproduced without systematic error. This does not apply to loudspeaker reproduction.

In addition, in the case of headphone reproduction, assessment tests can be carried out with a great number of listeners at the same time and under identical listening conditions.

6 Sound level

6.1 Loudspeaker reproduction

When using a wanted signal of high peak level, the sound level should be measured with a sound level meter with no weighting and the "slow" time constant standardized by the IEC (Publication 123). For other signals, and for measuring room noise, the level should be measured with a sound level meter with weighting *A* and the "slow" time constant standardized by the IEC (Publication 123). For measurement of the sound level of a programme sequence in the special conditions of the test and at a given position in the listening room, the sound level will be taken by definition as equal to the maximum value shown by the sound level meter during each sequence. In case of assessments of high-quality high-level signals, a listening sound level of 80 to 90 dB should be used.

The sound level considered in defining exactly the conditions in which the tests have been carried out will be the mean of the sound levels measured at the various positions occupied by the listeners. The difference from this mean value for any position must be as small as possible. A value of ±4 dB might be reasonable. All measurements should be made with the listeners present.

6.2   Headphone reproduction

In order to avoid measuring the sound level in the ear canal in the case of headphone reproduction, the sound level should be adjusted in such a way that loudness equal to a reference sound field is achieved. To determine equal loudness, the listener should be positioned in a reference sound field according to § 6.1.

When comparing the loudness of the headphones with that of the reference sound field, the signals are presented to the listeners alternatively (not simultaneously). The headphones are supplied with an input signal of the same nature as that of the reference sound field and are adjusted to the same loudness according to the judgements of the listeners.

The mean value of all loudness comparison judgements should be used to ensure that the correct headphone sound level is used for the tests.

7   Listening conditions

Generally speaking, an effort should be made to minimize the masking effect due to room noise, particularly when establishing tolerances for high-quality sound transmission.

The mean level of the room noise should always be indicated and, when it is manifestly likely to have a noticeable masking effect, the mean spectrum should also be indicated.

Furthermore, precautions should be taken to prevent as far as possible the listener(s) from being annoyed or distracted by certain features of the surroundings (temperature, light, moving objects or persons, etc.).

7.1   Loudspeaker reproduction

Whenever the tests are conducted with loudspeakers, all the essential information concerning the dimensions and the reverberation time of the listening room*, the arrangement of listeners in the room and their distance from the loudspeaker or loudspeakers should be given.

Technical requirements for the loudspeaker characteristics are in use in the Russian Federation.

7.2   Headphone reproduction

Whenever the tests are conducted with headphones, all the essential information concerning the type designation of the headphones used should be given.

Technical requirements for the headphone characteristics have to be defined. A current EBU text proposes an action programme aimed at drawing up an international standard applicable to high-quality headphones.

---

*   For acoustical properties of listening rooms, see Report ITU-R BS.797.

8 Assessment of special characteristics of equipment*, programmes, studios, etc.

8.1 Protection ratios

The assessment of protection ratio requires a slightly different testing procedure. In this case, the unimpaired programme sequence used for comparison should be such that the sound quality reproduced by the receiver is appropriate to the broadcasting system for which the receiver is designed.

8.2 Recorded programmes, studios

For the assessment of recordings no uniform method exists. The OIRT suggests special working methods for assessing recordings intended for the international exchange of programmes (OIRT, Recommendations Nos. 63/1; 91), and methods of assessing the acoustical properties of studios and concert halls (OIRT, Recommendation No. 68). Information is available on requirements for high-quality subjective assessment which are applied in the Russian Federation (listening conditions, choice of method, number of listeners and their selection).

8.3 Applications of subjective assessment of sound quality

Studies in the Russian Federation have attempted to identify requirements for subjective tests in broadcasting. The applications of subjective assessment were divided into three areas:

– sound recordings for programme exchange;

– studios, halls and other listening rooms;

– equipment.

These three areas were further broken down into groups and presented in a table.

The assessment requirements are based on international practice reflected in ISO, IEC and OIRT texts (i.e. for noise level, lighting, instructions, positions of loudspeakers and subjects, etc. in IEC Publication 543; for protocols in OIRT Recommendation No. 68/1.

9 Subjective assessments of multi-dimensional sound systems

It is argued that subjective assessments of signal distortion caused by non-linearities, interference or noise can appropriately be measured by the methods given in the body of this Recommendation.

However, in certain fields such as sound "surround-sound" or high definition television, the design problem is more complex. The subjective assessments are needed to design or choose a multi-channel sound system, and there are in this case new considerations which go beyond distortion of a sound channel alone. They include the extent to which localization is possible or the effectiveness of the reproduction of a multi-dimensional sound field at a given point. In this case, new assessment methods are needed, which go beyond those considered currently in this Recommendation.

---

* For example, apparatus such as compressors, compandors, recorders, etc.

As an example, in the study by Oghusi of NHK using multi-dimensional scaling, the following attributes were examined:

– apparent sound stage width,

– surround effect,

– apparent room size,

– horizontal and vertical localization,

– naturalness,

– sense of reality,

– agreeableness,

– correspondence of sound and image,

– appropriateness of sound image for pictures.

Assessors were asked to grade a number of alternative systems for these attributes. A table of dissimilarities was prepared, from which two orthogonal perceptual axes seem to be associated with the perceived realism (strongest link to the number of channels) and the coincidence of sound and picture (strongest link to the provision of a central source). The choice of system in the NHK case was based on overall quality and Japan would advocate such an approach in such cases.