

RECOMMENDATION ITU-R BT.1359-1

RELATIVE TIMING OF SOUND AND VISION FOR BROADCASTING

(Question ITU-R 35/11)

(1998)

The ITU Radiocommunication Assembly,

considering

- a) that a perceptible time difference between the sound and vision components of a television signal impairs the viewers' reception of the programme;
- b) that separate picture and sound processing is becoming more and more widely used in broadcasting systems;
- c) that digital production and distribution equipment causes differential delay between the sound and vision signals;
- d) that programme production may involve tandem connected studios;
- e) that in studios the sound/vision relative timing should be the responsibility of the programme production directors;
- f) that the transmitting equipment and the receiver may introduce an additional, variable timing difference;
- g) that subjective evaluations show that detectability thresholds are about +45 ms to -125 ms and acceptability thresholds are about +90 ms to -185 ms on the average, a positive value indicates that sound is advanced with respect to vision,

recommends with reference to Figure 1

- 1** that the timing zero, as a reference for the subsequent measurement of the relative timing of sound and vision signals is defined at the point of the final programme source selection element*;
- 2** that the overall tolerance in sound/picture timing (between points 1' and 6') shall not exceed +90 ms or -185 ms;
- 3** that the timing tolerance between image source (point 1) and the zero reference point as defined in *recommends* 1 above be taken as falling within the limits of +25 ms and -100 ms (Note that this is the zone within which the programme producer may exercise control over the relative timing of sound and picture. It is not possible to identify the correct or intended timing within this range, first because of the "plateau of undetectability" as noted in Appendix 1 Figure 2 and secondly because the producer may have selected "for artistic effect" a non-zero relative timing.);
- 4** that the timing difference in the path from the output of the final programme source selection element* to the input to the transmitter for emission should be kept within the values +22.5 ms and -30 ms**;
- 5** if correction of errors is not possible then each downstream segment that is not under the control of the broadcaster shall not introduce any timing error in excess of ± 2 ms.

Annex 1 outlines the user requirements for the corrections of relative timing difference of sound and vision signals which were used in preparing this recommendation.

Appendix 1 is an explanation of the selection of the recommended timing difference values.

Appendix 2 details currently used conditions for subjective assessment of sound/vision delay difference testing.

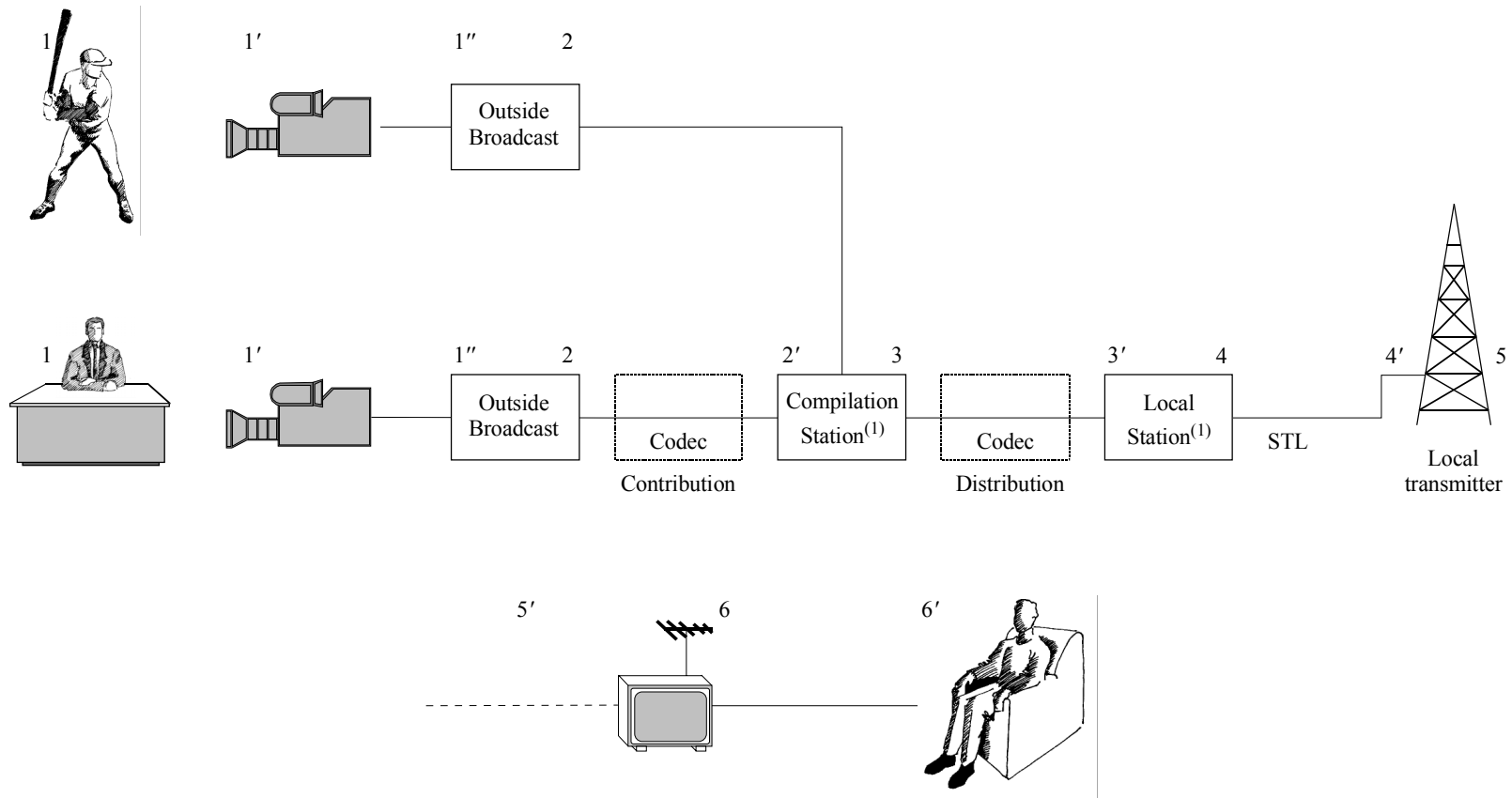
NOTE 1 – A positive value indicates that sound is advanced with respect to vision.

NOTE 2 – Studies should be made of time stamping in such a way as to facilitate the maintenance and correct timing at appropriate points in the broadcast chain.

* The definition of this point may vary depending on the particular broadcast organization and operating requirements. Typical examples are master control, network control, master switching or outside broadcast control.

** Where the path from the output of the final programme source selection element to the input of the transmitter is comprised of one or more digital codecs it should be noted that Recommendation ITU-R BT.1203 specifies that the delay error introduced by any single digital codec should be in the range ± 2 ms.

FIGURE 1
Simplified reference chain for television sound/vision timing



⁽¹⁾ This includes synchronization equipment needed such as frame synchronizer

ANNEX 1

User requirements for the correction of relative timing difference of sound and vision signals

When implementing the correction of relative timing difference or error of sound and vision signals, the following user requirements should be satisfied.

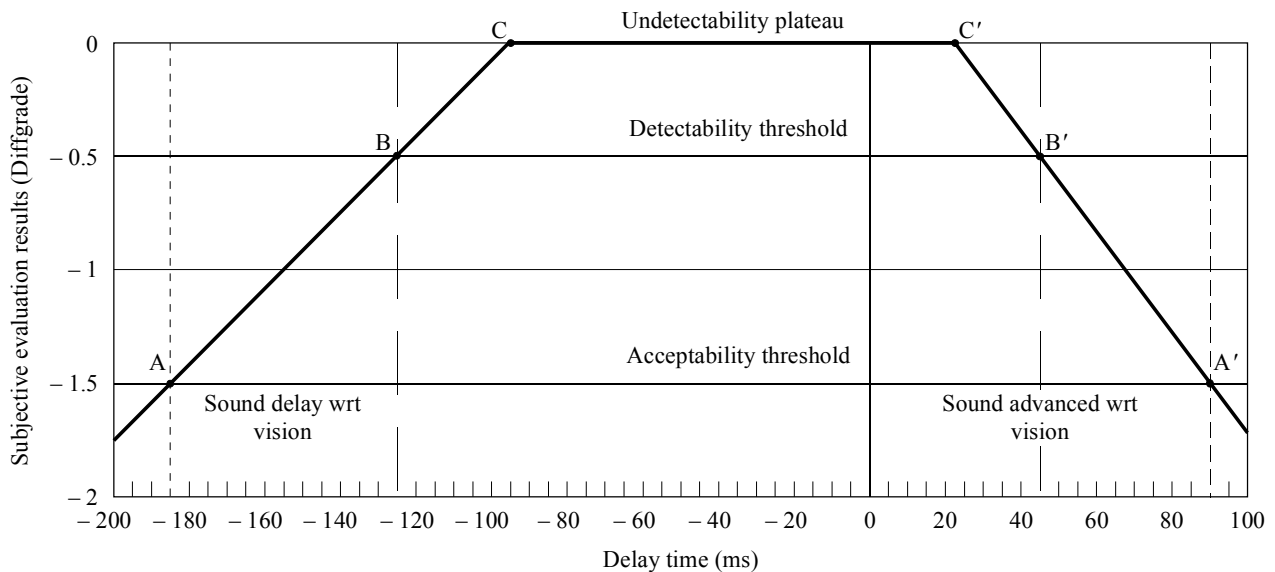
- 1** In the case of **on-line** correction of the timing error of sound and vision signals, the audio quality of the sound signal when observed at the output of the sound signal timing corrector, should be maintained at the start of, during and at the end of the correction, with the quality of 4.5 or higher, when evaluated using the subjective evaluation method based on the ITU-R five-grade impairment scale and the results are presented using the diffgrade.
- 2** The correction of the relative timing error of the sound and vision signals should be carried out within a responsibility boundary in the signal chain.
- 3** The standard reference signals intended for the **off-line** use for the measurement and/or correction of the error should be able to be observed by the eye and ear, while they should be measured by at least using a piece of equipment displaying the timing difference of the two signals.
- 4** The cost of the equipment to produce the reference signals and/or to measure the timing difference should be within a reasonable range.

APPENDIX 1

Explanation for the selection of the recommended value for sound/vision timing difference

- 1** It is known for many years from experience with film projection that the relative timing between picture and sound is very important and shows an identifiable point at which the timing error becomes objectionable to the viewer. Recommendation ITU-R BR.265 indicates that the precision of accuracy of location of sound and picture information should be within \pm half a frame. For 24 fps film, this is an acceptable variation of about ± 22 ms.
- 2** Differing imaging techniques generating source television signals appear to introduce unavoidable uncertainty of the actual sound/vision timing of about half a television field.
- 3** Subjective evaluations undertaken in Japan, Switzerland and Australia show a high degree of similarity in the sensitivity of viewers to errors in sound/vision timing in television material for NTSC and PAL systems. Tests conducted have shown that the thresholds of detectability are about +45 ms to -125 ms and thresholds of acceptability are about +90 ms to -185 ms on the average. Each set of test results indicates a broad area of acceptable timing covering "sound leading" through zero timing difference to "sound delayed". The range of timing between the "just detectable" limits of sound leading and sound delayed is about 170 ms. Each case also shows a clearly defined and rather consistent range of values for the difference (1 grade) between detectable and acceptable limits of about 45 ms for sound leading and about 60 ms for sound delayed as shown in Figure 2.

FIGURE 2
 Detectability and acceptability thresholds



1359-02

4 For the purpose of establishing a Recommendation concerning an agreed limit to sound/vision timing error in television, the range of values between the detectable limits is not relevant. The actual timing value is the province of the programme producer in the studio. Because we do not necessarily know, and have no recommended way of determining, the precise timing difference, we accept as being correct the relative timing that occurs at the studio output. An unsatisfactory situation may now exist because the studio output timing may be set to be very close to one of the limits of perceptibility and thus there is a limited margin of additional error before the timing error is such as to become unacceptable.

5 Because of the undetectable plateau (C-C', see Figure 2) the limit of allowable error should be constrained within 0.5 grade points (5 point impairment scale) above the subjectively evaluated detectable threshold (B-B'). The subjective evaluations have shown that a one grade point impairment results in a 60 ms change in delay which is shown on Figure 2 as the rising slope from A-B. The allowable delay should be constrained within a half grade point impairment which is 30 ms shown in Figure 2 on the rising slope from B-C. Likewise the advance limit is determined to be 22.5 ms from the rise on the slope from B'-C'.

APPENDIX 2

**Currently used conditions for subjective assessment
of sound/vision delay difference testing**

1	Distance between source and microphone	50 cm	
2	Distance between loudspeaker and assessor	200 cm	Assessors should be able to easily see lip movements of announcers. If a 22" diagonal monitor is used the display and loudspeaker will be in approximately the same position and this equates to a 6H viewing distance.
3	Parameter(s) for measurement	Acceptable thresholds (i.e. rating 3.5 on DSIS)	But it is also considered important to measure the "detectability" thresholds (i.e. rating 4.5 on DSIS).
4	Camera type	Tube camera	
5	Evaluation method	Double stimulus impairment scale method	
6	Viewing conditions	Recs. ITU-R BT.500 and 1128	
7	Range of sound/vision timing differences to be included on tape(s)	Must include full range of impairments, grade 1 through 5	Based on Swiss PTT work the range -200 to +300 ms should be assessed. To obtain precision in assessment of "detectable" thresholds a second tape with a concentration of values around the detectable threshold points may be required.
8	Test material	Female newsreader	To avoid viewer fatigue it is desirable that several different sequences and/or announcers be used.
9	Duration of test session	Less than 30 minutes	It may be necessary to hold two sessions if both "acceptable" and "detectable" thresholds are to be accurately assessed.
10	Type of assessors	Expert and non-expert	
11	Number of assessors	At least 15	
12	Age of assessors	To be stated	
13	Visual acuity of assessors	Normal (or corrected to normal) acuity assessed using a Snellen chart	