

International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

FG AVA TR

Version 1.0
(10/2013)

Focus Group on Audiovisual
Media Accessibility
Technical Report

**Part 13: Audio characteristics for audio
descriptions and/or spoken subtitles**



FOREWORD

The procedures for establishment of focus groups are defined in Recommendation ITU-T A.7. The ITU-T Focus Group on Audiovisual Media Accessibility (FG AVA) was proposed by ITU-T Study Group 16 for creation in-between TSAG meetings and it was established on 22 May 2011. The Focus Group was successfully concluded in October 2013.

Even though focus groups have a parent organization, they are organized independently from the usual operating procedures of ITU, and are financially independent. Texts approved by focus groups (including Technical Reports) do not have the same status as ITU-T Recommendations.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Technical Report may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU-T Focus Group participants or others outside of the Technical Report development process.

© ITU 2014

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Table of Contents

1 Channels 3

2 Loudness and peak levels 3

3 Soundtrack vs audio description/described video (AD/DV) mix levels..... 3

4 Equalization 3

5 Sync 4

6 Recording quality..... 4

7 Output parameters..... 4

8 Other considerations 4

Summary

This Technical Report of FG AVA was prepared by Working Group B "Audio/video description and spoken captions" ¹ that FG AVA had foreseen as a future work item for ITU-T Study Group 16 (SG16) "Multimedia" and ITU-R SG6 "Broadcasting service".

The ITU-T Focus Group on Audiovisual media Accessibility (FG AVA) agreed to recommend the characteristics of audio descriptions provided in this document for audio descriptions and/or spoken subtitles for assisting content accessibility of audiovisual programme material.

¹ *Audio descriptions* are also called *described video* in North America.

Audio characteristics for audio descriptions and/or spoken subtitles

1 Channels

If the audio description and spoken subtitles² output is stereo or surround sound, the sound image should be centred on the TV screen. If the audio description and spoken subtitles are supplied at the same time, the user should be able to choose both or each service separately

2 Loudness and peak levels

2.1 Audio description and spoken subtitle loudness should approximately track or slightly exceed soundtrack loudness, adjusting levels as needed. Peak levels should not exceed soundtrack peaks by more than 6 dB to avoid unnecessarily triggering the broadcasters' compression/limiting gear, which can negatively affect mix quality.

2.2 Depending on national practice for audio, audio description and spoken subtitle loudness should fall within 2 (LKFS)³ (or 1 LUFS⁴) of -24 LKFS (or -23 LUFS) over the course of the programme. (Pls see footnotes. 3 and 4)

2.3 The audio description and spoken subtitle audio levels should 'ride' the loudness dynamics of the original audio tracks, and this should be achieved either by matching the energy of the delivery during recording, or by adjusting the tracks themselves during the mix, neither method being preferable to the other.

3 Soundtrack vs audio description/described video (AD/DV) mix levels

3.1 Audio description and spoken subtitles should be noticeably louder and more present than the coinciding original soundtrack audio. Unless deemed necessary by the producer/mixer, the soundtrack should not be completely suppressed; it should be possible to hear the original soundtrack under audio description and spoken subtitles, with no more reduction than judged necessary to ensure that audio description and spoken subtitles can be clearly understood. If possible, the user should be able to adjust the volume of the audio description and spoken subtitles separately from the soundtrack.

4 Equalization

4.1 Audio description and spoken subtitles should be adjusted to emphasize clarity of speech (a small boost in the 1-6 kHz range as needed to enhance consonants, and a corresponding reduction/roll-off below 200 Hz; however, frequency ranges may differ due to differing recording and mixing circumstances).

4.2 Audio description and spoken subtitles are intended to be complementary to a programme. Approximating soundtrack equalization (is allowable if the original audio is of a significantly reduced frequency range than AD, or is an effort to maintain the aural atmosphere of a soundtrack; however, AD/DV clarity should not be compromised and should take priority.

² *Audio descriptions* are also called *described video* in North America.

³ Loudness, K-weighted, relative to Full Scale (LKFS) is a loudness standard designed to enable normalization of audio levels for delivery of broadcast TV and other video. LKFS is standardized in ITU-R BS.1770.

⁴ Loudness units relative to Full Scale (or LUFS) is used in EBU R128.

5 Sync

Resulting audio description and spoken subtitles mix should be in sync with the original soundtrack, preferably sample-accurate, no worse than within a frame (1/50, 1/60 s) of the original soundtrack. In cases where the original soundtrack is out of sync with the picture, the audio description and spoken subtitles mix should sync with the audio.

6 Recording quality

6.1 Description recordings should, at a minimum, have the same bit depth and sample rate as the original soundtrack.

6.2 Signal-to-noise ratio should be of broadcast-quality standard. Recording environments should be as deadened and quiet as possible; mixer should not need to attempt to filter out reverberation and external noise. Voice over speaker should be familiar with basic microphone set-up and technique to avoid varied voice quality over the course of a recording.

7 Output parameters

Bit depth/sample rate of the audio description and spoken subtitles mix should match the original soundtrack material mixing.

8 Other considerations

Audio description and/or spoken subtitles should be available at no additional cost as a separate file to all subsequent users and purchasers of programmes.
