International Telecommunication Union

ITU-R

Radiocommunication Sector of ITU

Recommendation ITU-R BT.1865

(03/2010)

Metadata to monitor errors of SDTV and HDTV signals in the broadcasting chain

BT Series

Broadcasting service (television)

International Telecommunication Union

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Annex 1 of Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from http://www.itu.int/ITU-R/go/patents/en where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Recommendations

(Also available online at http://www.itu.int/publ/R-REC/en)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	**Broadcasting service (television)**
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management
SNG	Satellite news gathering
TF	Time signals and frequency standards emissions
V	Vocabulary and related subjects

Note: *This ITU-R Recommendation was approved in English under the procedure detailed in Resolution ITU-R 1.*

Electronic Publication
Geneva, 2010

RECOMMENDATION ITU-R BT.1865

Metadata to monitor errors of SDTV and HDTV signals in the broadcasting chain

(Questions ITU-R 44/6, ITU-R 48/6, ITU-R 109/6 and ITU-R 130/6)

(2010)

Scope

This Recommendation defines metadata to monitor errors of audio, video and data signals at arbitrary monitoring points in an SDTV/HDTV broadcasting chain[1]. The metadata is packetized into ancillary data packets. This mechanism may also be applied to other types of metadata for picture quality measurement.

The ITU Radiocommunication Assembly,

considering

a) that digital broadcasting has been introduced in many countries, and monitoring of the digital broadcasting chain by broadcasters and network operators has become increasingly important;

b) that in broadcast operational monitoring, an objective method of automatically detecting malfunctioning equipment and quality degradation in audiovisual content is needed;

c) that the monitoring method should be reliable, efficient, and cost-effective;

d) that the cause of degradation such as signal loss, blackouts, freezing, and muting may be transmission paths, equipment, human error, or intentional audiovisual effects, and operators have to accurately assess the cause of the alarms and rectify the problems;

e) that if appropriate metadata were to be added to the audiovisual content at an early stage of the broadcasting chain, it would make the monitoring system at later stages of the chain more reliable and help operators to analyse the cause of malfunctions or degraded quality,

recognizing

a) that ITU-R has established Recommendation ITU-R BT.1364 – Format of ancillary data signals carried in digital component studio interfaces, which defines the data structure for packetized data that may be carried as part of the payload in serial digital interfaces as defined in Recommendations ITU-R BT.656 and ITU-R BT.1120,

recommends

1 that for the use of metadata to monitor errors of audio and video signals in television broadcasting chains, the specifications in Annex 1 should be used.

NOTE 1 – Appendix 1, attached to this Recommendation, provides additional explanations of metadata for operational monitoring.

---

[1] Monitoring on a broadcasting chain from news gathering, programme production and postproduction through master control is the primary subject of this Recommendation.

Annex 1

Reference (informative)

– Recommendation ITU-R BT.1790 – Requirements for monitoring of broadcasting chains during operation.

– ITU-T Recommendation J.243 – Requirements for operational monitoring in television programme transmission chains.

– ITU-T Recommendation P.911 (1998) – Subjective audiovisual quality assessment methods for multimedia applications.

– ITU-T Recommendation J.240 (2004) – Framework for remote monitoring of transmitted picture signal-to-noise ratio using spread-spectrum and orthogonal transform.

– ITU-T Recommendation J.249 (2009) – Perceptual visual quality measurement techniques for digital cable television in the presence of a reduced reference.

Reference (normative)

– Recommendation ITU-R BT.1364-1 – Format of ancillary data signals carried in digital component studio interfaces.

– ITU-T Recommendation J.187 – Transport mechanism for component-coded digital high-definition television signals using MPEG-2 video coding including all service elements for contribution and primary distribution.

– ISO 3166-1:2006 – Codes for the representation of names of countries and their subdivisions – Part 1: Country codes.

1 Overview

A schematic diagram of an operational monitoring process using metadata in a broadcasting chain is shown in Fig. 1 where operational monitoring is assumed to be conducted at arbitrary monitoring points in a broadcasting chain. Details on the monitoring process are given at the bottom of Fig. 1. This can be summarized as:

1 The metadata inserted at the upstream monitoring points are extracted.

2 Audio and video signals are analysed at monitoring points to generate metadata.

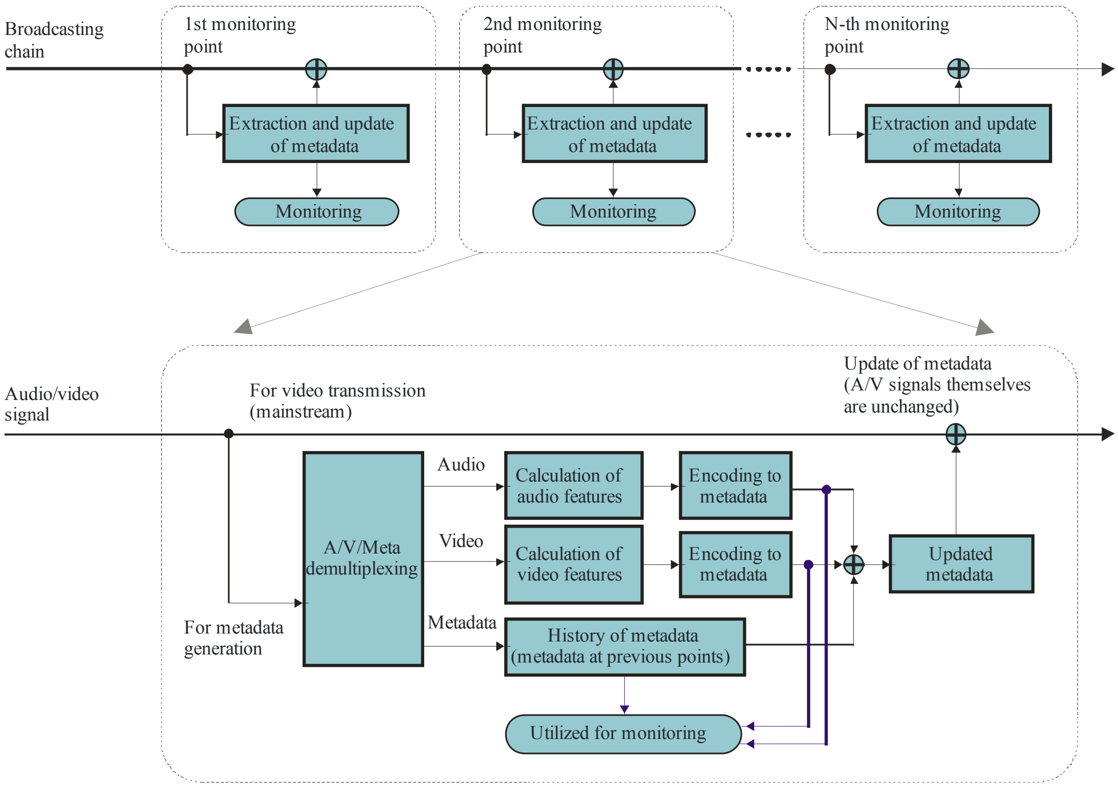
3 By comparing the current and upstream metadata, the audio and video signals are monitored to determine if any problems have occurred.

4 The metadata generated at the current monitoring point are added to the metadata history.

Only metadata that are used for operational monitoring are updated within the process, and audio, video and any other auxiliary signals are left unchanged.

FIGURE 1

Configuration of monitoring points in broadcasting chain and monitoring process for metadata



BT.1865-01

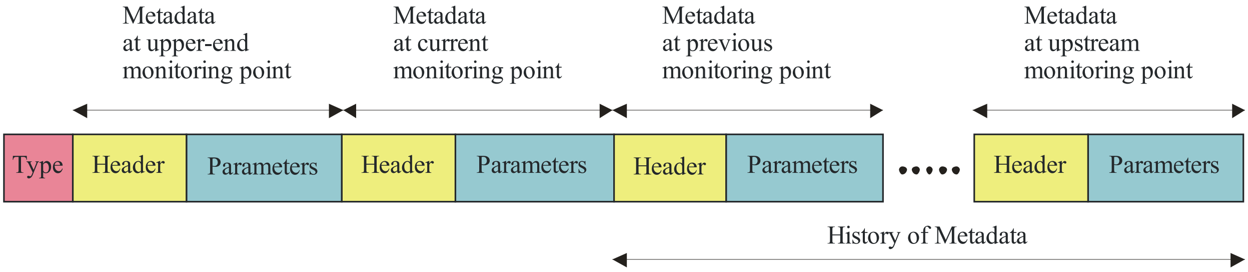
2 Metadata for operational monitoring

2.1 Configuration

The basic configuration for the metadata for operational monitoring is outlined in Fig. 2. The configuration allows the use of different types of metadata, but the same type of metadata is to be used for the series of metadata in a broadcasting chain. Metadata such as those defined in ITU-T Recommendation J.249 for perceptual picture quality measurement and ITU-T Recommendation J.240 for remote monitoring of transmitted picture signal-to-noise ratio may also be used.

FIGURE 2

Configuration for metadata



BT.1865-02

The metadata at the monitoring point on the upper end always appear first, followed by the metadata at the current monitoring point. Other metadata at the previous monitoring points follow in order of those most recently monitored in the broadcasting chain. The number of history items of metadata depends on the capacity of the data area; however, the first two sets of metadata, i.e. the metadata at the monitoring point on the upper end and the current monitoring point, should be retained. When the history items are no longer valid for comparison with the metadata at the down stream monitoring points, the metadata are to be reset.

The syntactical definitions of the metadata are listed in Table 1. The definitions of the header, video parameters, audio parameters, and data parameters of the metadata are given in the following sections.

TABLE 1

Definition of metadata for monitoring

Syntax	No. of bytes	Mnemonic
metadata_type	1	bslbf
for(i=0; i<N; i++){		
monitoring_metadata( )		
}		

Syntax
monitoring_metadata{
header( )
video_parameters( )
audio_parameters( )
data_parameters( )
}

metadata _type indicates the type of metadata used.

The metadata may not have to contain all three parameters.

Value	Assignment			
	Type	Video parameter	Audio parameter	Data parameter
0x00	Reserved	Not applicable		
0x01 to 0xFE	To be specified	To be specified		
0xFF	Reserved	Not applicable		

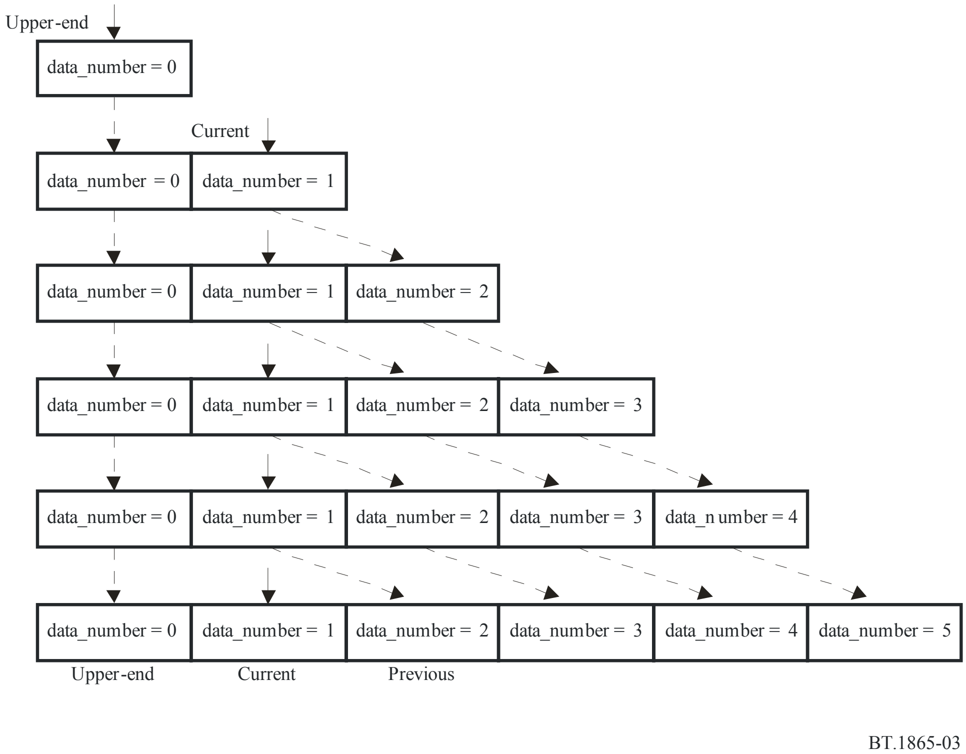
2.2 Header

The header information precedes the video, audio, and data parameters to identify the data number, types of video, audio, and data signals, the location of the monitoring point and the organization that generated the metadata. The header may contain **data_number**, **video signal_type**, **audio_sygnal_type**, **data_signal_type**, **country_code**, **organization_code**, and **user_code**. An example of the header is provided in Appendix 1 (to Annex 1).

data_number indicates the data number within the history of metadata. The first metadata inserted at the upper-end monitoring point has a data number of 0, the current metadata has a data number of 1, and the previous metadata has a data number of 2. Due to the restricted size of the user data word (UDW) of an ancillary data packet, the number of metadata sets as the history is limited. Figure 3 shows the flow chart with respect to managing the metadata history and the data number. When the history is reset, the metadata are to be inserted starting from the data number of 0.

FIGURE 3

Management of metadata history and data number



video_signal_type indicates the type of video signal (e.g. uncompressed, compressed).

audio_signal_type indicates the type of audio signal (e.g. uncompressed, compressed). In some cases, audio signals may not be embedded with video signals.

country_code indicates the country where the monitoring point is, specified by the two-letter country code as per ISO 3166-1.

organization_code indicates the organization that operates the monitoring point designated by four ASCII characters.

user_code indicates the monitoring point in an organization designated by four ASCII characters.

2.3 Video parameters

The video parameters need to be able to detect video errors between two monitoring points. An example of the video data parameters is provided in Appendix 1.

2.4 Audio parameters

The audio parameters need to be able to detect audio errors between two monitoring points. An example of the audio parameters is provided in Appendix 1.

2.5 Data parameters

The data parameters need to be able to detect data errors between two monitoring points.

3 Transport of metadata

The metadata for operational monitoring is conveyed by being packetized into the ancillary data packets which are multiplexed with the video and audio signals.

3.1 Format for ancillary data packets for metadata

The format for the ancillary data packets of the metadata conforms to Type 2 ancillary data packets as defined in Recommendation ITU-R BT.1364, and one word consists of 10 bits in this format. The format for the data packets is given in Fig. 4. DID and SDID need to be assigned.

FIGURE 4

Format for ancillary data packets of metadata

ADF	DID	SDID	DC	UDW	CS
3	1	1	1	Max.255	1

ADF: Ancillary data flag ($0 \times 000_{(10)}$, $0 \times 3FF_{(10)}$, $0 \times 3FF_{(10)}$)
DID: Data identification word.
SDID: Second data identification word.
DC: Data count word.
UDW: User data word. Maximum of 255 data words.
CS: Check sum word.
Note: Numbers indicate number of words.

BT.1865-04

3.2 User data word format

User data words (UDW) comprise the metadata defined in § 2. The format for the UDW is listed in Fig. 5. The metadata are to be byte aligned, where the first bit of a byte occupies b7 and the last bit occupies b0.

FIGURE 5

Bit assignment for data words

Bit number	Description
B9(MSB)	Not b8
B8	Even parity for b0 through b7
B7	Metadata as per Section 2.
B6	
B5	
B4	
B3	
B2	
B1	
B0(LSB)	

LSB: least significant bit.
MSB: most significant bit.

BT.1865-05

3.3 Transport of ancillary data packets

The metadata generated for the video, audio, data signals in a given video frame are to be transported by an ancillary data packet attached to the following frame. Synchronization between the video frame and the metadata needs to be ensured.

Ancillary data packets containing the metadata are transported by one of the following methods:

1 The ancillary data packets are multiplexed into the ancillary data space of the serial digital interface, preferably in the vertical ancillary space. It is necessary to identify the available data area for this purpose by taking into consideration the current usage by broadcasters of the ancillary data space.

2 The ancillary data packets are multiplexed into the MPEG-2 Transport Stream. This method is specified in ITU-T Recommendation J.187.

3 The ancillary data packets are transported through a path different from that for video and audio signals. This method requires a means for synchronization between the ancillary data packets and video and audio signals.

Appendix 1
(to Annex 1)

Example of metadata for operational monitoring

This appendix describes the Type-1 metadata for operational monitoring as defined in ARIB TR-B29.

Reference

– ARIB Technical Report TR-B29 – Metadata to monitor errors of video and audio signals on a broadcasting chain.

Terminology

– **AES stream**: digital audio stream as defined in Recommendation ITU-R BS.647. One AES stream contains two audio channels.

– **bslbf**: bit string, left bit first, where “left” is the order in which bit strings are written in this Recommendation.

– **uimsb**: unsigned integer, most significant bit first.

1 Configuration

The syntactical definitions of the metadata are listed in Table 2. The definitions of the header, video parameters and audio parameters of the metadata are given in the following sections.

TABLE 2

Definition of metadata for monitoring

Syntax	No. of bytes	Mnemonic
metadata_type	1	bslbf
for(i=0; i<N; i++){		
monitoring_metadata( )	42	
}		

Syntax	No. of bytes
monitoring_metadata{	
header( )	11
video_parameters( )	10
audio_parameters( )	21
data_parameters	
}	

metadata _type indicates the type of metadata used.

Value	Assignment		
	Type	Video parameter	Audio parameter
0x01	Type-1	As per § 3	As per § 4

2 Header

The syntactical definition of the header is listed in Table 3.

TABLE 3

Definition of header

Syntax	No. of bits	Mnemonic
header( ){		
data_number	3	uimsbf
video_signal_type	1	bslbf
audio_signal_type	2	bslbf
reserved	2	
country_code	16	bslbf
organization_code	32	bslbf
user_code	32	bslbf
}		

data_number as defined in § 2.2 and Fig. 3 of Annex 1. Due to the restricted size of the UDW of an ancillary data packet, six sets of the Type-1 metadata can be used at maximum as the history.

video_signal_type indicates the type of video signal. Digital component video signals at the 4:2:2 level are assumed.

Type	Value
Uncompressed	0
Compressed	1

audio_signal_type indicates the type of audio signal. In some cases, audio signals may not be embedded with video signals. Digital audio signals are assumed.

Type	Value
Uncompressed	00
Compressed	01
No audio	10
Reserved	11

country_code as defined in § 2.2 of Annex 1.

organization_code as defined in § 2.2 of Annex 1.

user_code as defined in § 2.2 of Annex 1.

3 Type-1 video parameters

Table 4 lists the syntactical definition of the Type-1 video parameters. A total of 10 bytes (i.e. 80 bits) are allocated for the video parameters.

TABLE 4

Definition of Type-1 video parameters

Syntax	No. of bits	Mnemonic
video_parameters( ){		
video_input_error	1	bslbf
video_processing	3	bslbf
reserved	4	
y_si	8	uimsbf
y_ti	16	uimsbf
cb_si	8	uimsbf
cb_ti	16	uimsbf
cr_si	8	uimsbf
cr_ti	16	uimsbf
}		

video_input_error indicates whether an error is detected by the diagnostic in the physical layer of the video interface (e.g. errors detected by cyclic redundancy check codes (CRCC) of serial digital interfaces). When the diagnostic is not available at the monitoring point, this is set to 0.

Status	Value
Normal/unavailable	0
Error	1

video_processing indicates whether any video processing is conducted at the monitoring point. When such information is not available, this is set to 000.

Status	Value
Normal/unavailable	000
Frame repeat	001
Freeze[*1]	010
Frame skip	011
Special effects (e.g. wipe and superimpose)	100
Reserved	101-111

[*1] Freeze means repeats of a frame for more than two successive frame periods.

y_si, **cb_si**, and **cr_si** indicate the video spatial features of Y, Cb, and Cr signals calculated as per § 3.1.

y_ti, **cb_ti**, and **cr_ti** indicate the video temporal features of Y, Cb, and Cr signals calculated as per § 3.2.

3.1 Video spatial feature

Video spatial feature is the video spatial information (SI) defined by ITU-T Recommendation P.911. Sobel horizontal and vertical direction filters are applied to each Y/Cb/Cr component of the video signals frame-by-frame and the degree of edge sharpness is derived as:

$$SI = INT\left[\sqrt{\frac{1}{N}\sum_{i,j}\left\{SI_h(i,j)^2 + SI_v(i,j)^2\right\} - SI_m{}^2}\right] \tag{1}$$

$$\begin{aligned} SI_h(i,j) = &\{X(i+1,j-1) - X(i-1,j-1)\} \\ &+ 2\{X(i+1,j) - X(i-1,j)\} \\ &+ \{X(i+1,j+1) - X(i-1,j+1)\} \end{aligned}$$

$$\begin{aligned} SI_v(i,j) = &\{X(i-1,j+1) - X(i-1,j-1)\} \\ &+ 2\{X(i,j+1) - X(i,j-1)\} \\ &+ \{X(i+1,j+1) - X(i+1,j-1)\} \end{aligned}$$

$$SI_m = \frac{1}{N}\sum_{i,j}\sqrt{SI_h(i,j)^2 + SI_v(i,j)^2}$$

where $X(i, j)$ denotes the level of each component of the video signal at the i-th line and the j-th active sample of a frame, and N denotes the total number of active samples in a frame. In calculating the SI for interlace signals, a frame is composed by "field-merge". When $X(a, b)$ corresponds to inactive samples, active samples adjacent to $X(a, b)$ are to be used instead. $INT[x]$ returns the nearest integer value of x by rounding off fractional values below 0.5 or by rounding up fractional values above or equal to 0.5. The most significant eight bits of all the active samples of the video signal (0 to $2^8 - 1$) are used for this calculation, and the SI value is presented in the eight-bit unsigned integer notation.

3.2 Video temporal feature

Video temporal feature is the video temporal information (TI) defined by ITU-T Recommendation P.911. The power of the frame difference is calculated for each Y/Cb/Cr component of the video signal as:

$$TI = INT\left[\frac{1}{N}\sum_{j,j}\{X(i,j,n) - X(i,j,n-1)\}^2\right] \tag{2}$$

where $X(i, j, n)$ denotes the level of each component of the video signal at the i-th line, j-th active sample of a frame, and n-th frame ,and N denotes the total number of active samples in a frame. The most significant eight bits of all the active samples of the video signal (0 to $2^8 - 1$) are used for this calculation, and the TI value is presented in the 16-bit unsigned integer notation.

4 Type-1 audio parameters

Table 5 lists the syntactical definition of the Type-1 audio parameters. A total of 21 bytes (i.e. 168 bits) are allocated for the audio parameters.

TABLE 5

Definition of Type-1 audio parameters

Syntax	No. of bits	Mnemonic
audio_parameters( ){		
audio_input_error	1	bslbf
audio_processing	3	bslbf
audio_aes_channels_minus1	2	uimsbf
reserved	2	
for(i=0;i<4; i++){		
audio_ii	10	uimsbf
audio_oi	10	uimsbf
audio_rms_1	10	uimsbf
audio_rms_2	10	uimsbf
}		
}		

audio_input_error indicates whether an error is detected by the diagnostic in the physical layer of the video interface (e.g. errors detected by CRCC of serial digital interfaces). When the diagnostic is not available at the monitoring point, this is set to 0.

Status	Value
Normal/unavailable	0
Error	1

audio_processing indicates whether any audio processing is conducted at the monitoring point. When such information is not available, this is set to 000.

Status	Value
Normal/unavailable	000
Mute	001
Limiter	010
Special effects (e.g. super-impose and fade-in/out)	011
Reserved	100-111

audio_aes_channels_minus1 plus 1 indicates the number of AES streams. There are a maximum of four AES streams. One AES stream contains two audio channels.

audio_ii indicates the audio in-phase information (AII) between two channels in an AES stream calculated as per § 4.2.

audio_oi indicates the audio out-phase information (AOI) between two channels in an AES stream calculated as per § 4.2.

audio_rms_1 indicates the audio magnitude information (AMI) of audio channel 1 in an AES stream calculated as per § 4.3.

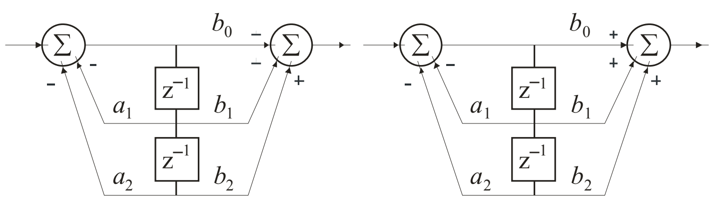
audio_rms_2 indicates the audio magnitude information (AMI) of audio channel 2 in an AES stream calculated as per § 4.3.

4.1 Preprocessing

Prefilter having a cut-off frequency of 20 Hz is applied to audio signals before calculating audio feature information. The prefilter is defined by the filter in Fig. 6 with the coefficients specified in Table 6. Floating point operation should be used. The response is shown in Fig. 7.

FIGURE 6

Signal flow diagram as 4th order filter



Note 1 – The filter is generally structured as a cascade of 2nd order filters.

BT.1865-6

TABLE 6

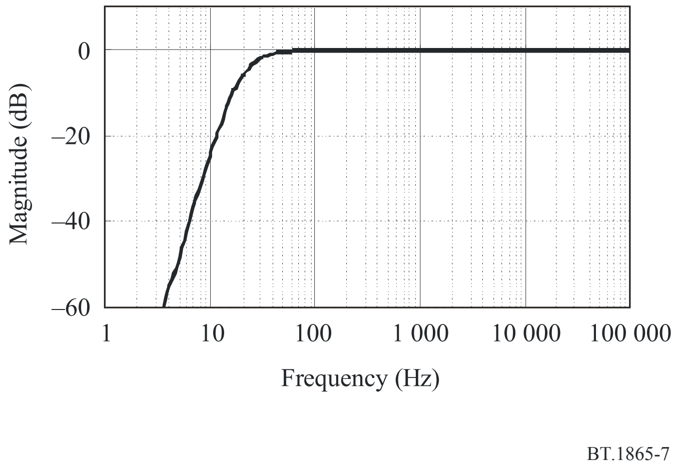
Filter coefficients for prefilter

		b_0	0.9981318
a_1	−1.9962602	b_1	−1.9962636
a_2	0.996267	b_2	0.9981318

NOTE 1 – These filter coefficients are for a sampling rate of 48 kHz and should be processed as single-precision real numbers.

FIGURE 7

Frequency response of prefilter



4.2 Audio inter-channel features

Audio in-phase information (AII) and audio out -phase information (AOI) are defined as:

$$AII = INT\left[\frac{1}{8}\left(\frac{1}{2N}\sum_{i=0}^{N-1} abs\big(X(i)+Y(i)\big)\right)\right] \tag{3}$$

$$AOI = INT\left[\frac{1}{8}\left(\frac{1}{2N}\sum_{i=0}^{N-1} abs\big(X(i)-Y(i)\big)\right)\right] \tag{4}$$

where $X(i)$ and $Y(i)$ denote the i-th sample value of the X- and Y-channels, and N denotes the number of audio samples within the duration of a video frame. Function $abs(x)$ returns the absolute value of x. The X- and Y-channels correspond to a pair of channels in an AES stream. A scaling factor of 1/8 has been adopted to represent the feature value. The most significant 16 bits of the audio signal (-2^{15} to $2^{15}-1$) are used for this calculation, and the AII and AOI values are represented in the 10-bit unsigned integer notation. When the calculated value is above $2^{10}-1$, the value is to be clipped at $2^{10}-1$.

4.3 Audio magnitude feature

Audio magnitude information (AMI) is defined as:

$$AMI = INT\left[\frac{1}{8}\sqrt{\frac{1}{N}\sum_{i=0}^{N-1} X^2(i)}\right] \qquad (5)$$

where $X(i)$ denotes the i-th sample value of an audio channel, and N denotes the number of audio samples within the duration of a video frame. A scaling factor of 1/8 has been adopted to represent the feature value. The most significant 16 bits of the audio signal (-2^{15} to $2^{15} - 1$) are used for this calculation, and the AMI values are represented in the 10-bit unsigned integer notation. When the calculated value is above $2^{10} - 1$, the value is to be clipped at $2^{10} - 1$.

5 Transport of metadata

The metadata for operational monitoring is conveyed by being packetized into the ancillary data packets as specified in § 3 of Annex 1.

The format for the ancillary data packets of the metadata is as specified in § 3.1 of Annex 1. DID and SDID of the data packets for the Type-1 metadata are as follows:

DID	$0x143_{(10)}$
SDID	$0x104_{(10)}$

Appendix 2
(to Annex 1)

Operational guidelines for metadata

This appendix describes the operational guidelines for the metadata.

1 Signals to be monitored at monitoring points

Baseband signals over SDI (serial digital interface) are generally used for the inter-connection between equipment and studios in a broadcasting station. Compressed signals over DVB-ASI (asynchronous interface) are also used for transmitting the signals between broadcasting stations. Compressed signals are usually used for television outside broadcasting (TVOB) to reduce transmission bandwidth.

Metadata are multiplexed in baseband signals into ancillary data space within SDI. Ancillary data in compressed signals need to be multiplexed into a transport stream. ITU-T Recommendation J.89 offers a transport mechanism for ancillary data. State-of-the art encoders may support this mechanism.

2 Operation cases

Three cases are assumed where metadata are used for monitoring:

1 programme transmission from the master control of a broadcasting station to that of a network-linked broadcasting station;

2 transmission of programme material from a network-linked station or an outside broadcasting location to the broadcasting centre;

3 interconnection from VTRs or studios to a master control.

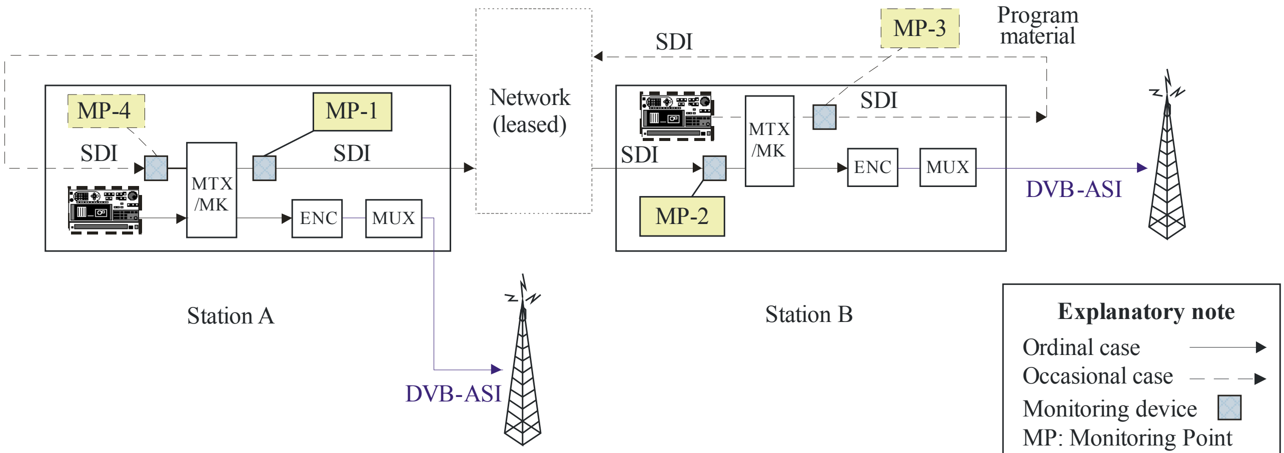
Monitoring may be conducted by broadcasters and by network operators.

2.1 Programme transmission between master controls

Figures 8 and 9 describe typical configurations of programme transmission between master controls. SDI is used in Fig. 8 and DVB-ASI in Fig. 9.

FIGURE 8

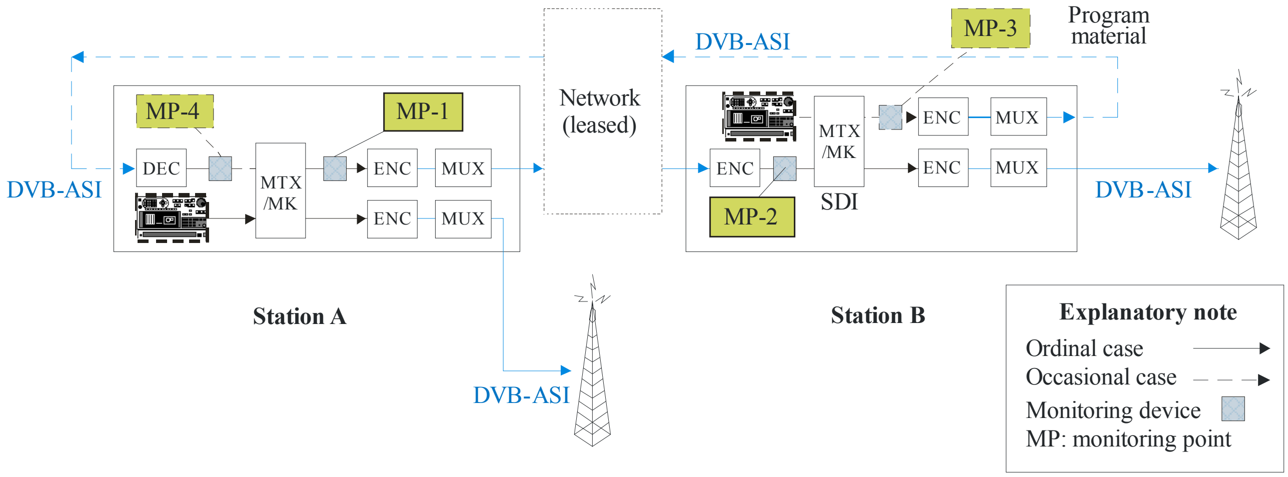
Programme transmission between master controls using SDI baseband signals



BT.1865-8

FIGURE 9

Programme transmission between master controls using DVB-ASI compressed signals



BT.1865-9

Sending station A installs a "monitoring point 1" at the back end of the master output to insert metadata indicating the status of the sending signals. Receiving station B installs a "monitoring point 2" at the front end of signal input and monitors the received signals. If an unusual status is detected in the received signals, the metadata are fully utilized to identify at which portion in the transmission chain the cause is. If the metadata indicate the same status as that detected at "monitoring point 2", the receiving station can determine that there are no problems in the transmission path.

These figures also indicate a possible case where Station B transmits its programme material to Station A, and Station A delivers the programme to its network-linked stations including Station B. In this case, "monitoring points 3 and 4" are additionally installed.

The networks provided by network operators are usually employed for this type of transmission. Monitoring by network operators is described in a later section.

2.2 Programme material transmission

Figure 10 describes a typical configuration for programme material transmission from an outside broadcasting location to the broadcasting centre. The transmission lines are those provided by network operators or owned-operated networks. In the former, broadcasters and network operators are usually interfaced with SDI baseband signals.

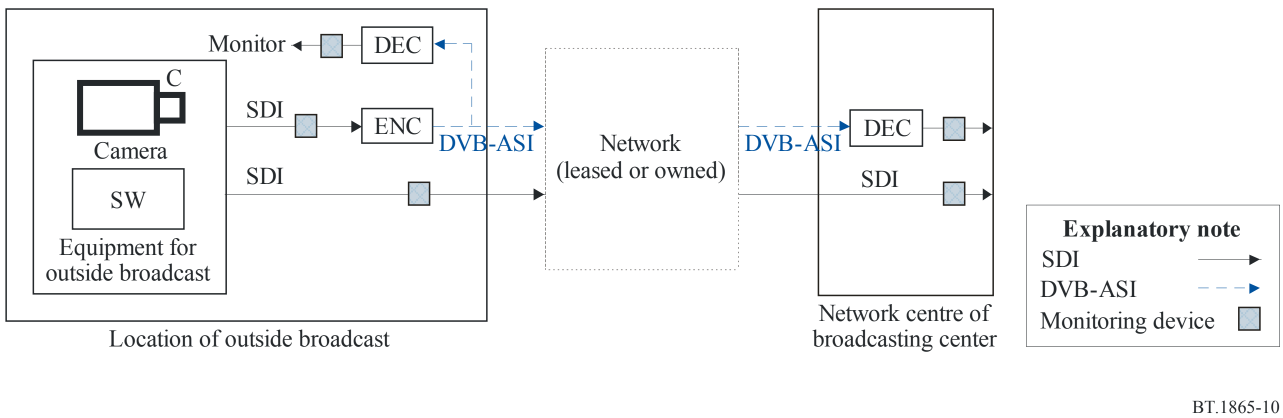
On the sending side, i.e. the location of outside broadcasting, a monitoring device is installed at the back end of SDI baseband signal output or the front end of encoder input. When compressed signals are used for transmission, the installation of an additional monitoring point at the output of a local decoder enables encoder-related problems to be monitored on the sending side.

On the receiving side, a monitoring device is installed at the front end of SDI baseband signal input or the back end of decoder output.

The metadata are inserted at the sending location to indicate the status of the sending signals. The received signals are monitored at the broadcasting centre and compared with the status indicated by the metadata.

FIGURE 10

Programme material transmission in outside broadcasting

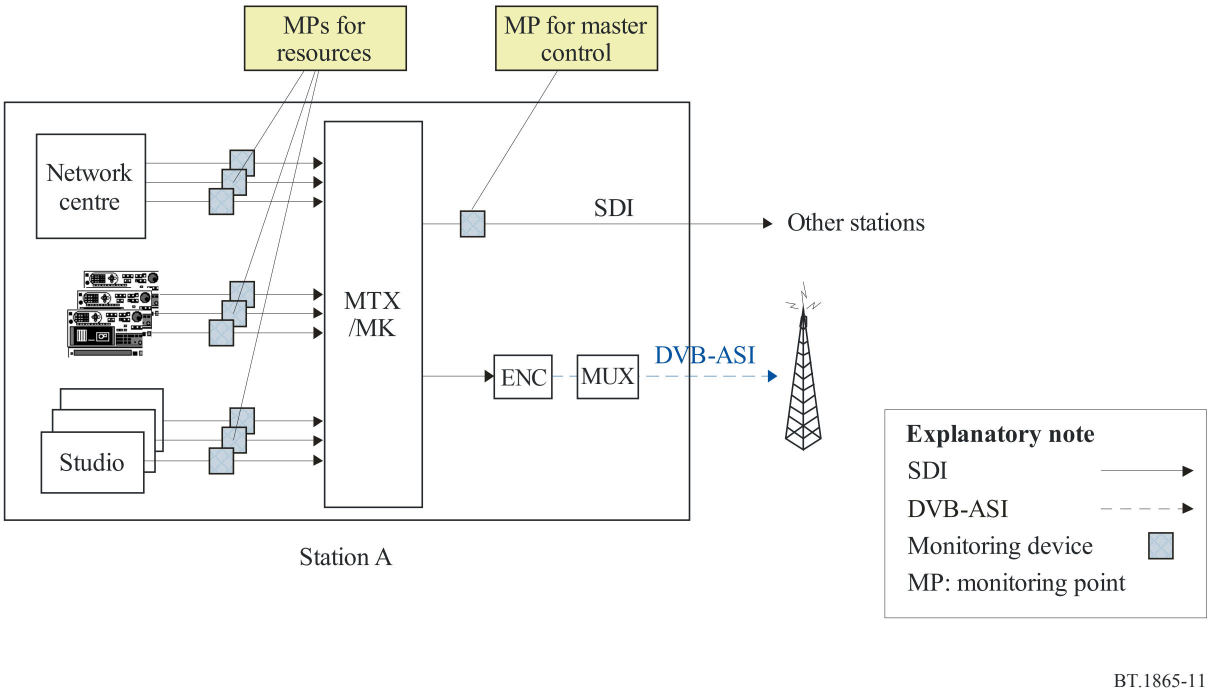


2.3 Transmission and signal processing in broadcasting centre

MK switchers and devices for video-signal processing like the DVE used in studios, editing rooms, and network centres do not usually convey ancillary data, and delete the metadata attached to the input programme materials. It is therefore unfeasible to retain metadata from outside broadcasting locations or VTRs in studios in the master control. As an alternative, monitoring points should be installed at the output of those broadcast resources inserting the metadata as shown in Fig. 11.

FIGURE 11

Monitoring points at output of broadcast resources



3 Monitoring points in broadcasting centre

The additional installation of monitoring devices along the signal chain should be avoided as this decreases the reliability of broadcasting operations. It thus seems practical to implement monitoring functions by adding functions to existing devices. We also need to be able to bypass monitoring functions in cases of emergencies.

The following devices are assumed to implement monitoring functions. Figure 12 shows their location in a master control and Fig. 13 in outside broadcasting.

1) *ANC Inserter*

The ANC Inserter multiplexes data such as closed captions and controls into an ancillary data space of SDI. This may be a potential cause that decreases the reliability of equipment by adding new functions and complexity to the existing equipment. However, the ANC Inserter intrinsically manages the ANC space, and there would therefore be less risk in installing monitoring functions in the ANC Inserter.

2) *Encoder*

The encoder compresses the amount of audio and video information and is considered to be a converter to change SDI baseband signals to DVB-ASI compressed signals. The input of an encoder is the back end of a baseband signal chain, and an encoder is thus most appropriate for installing monitoring functions.

3) *DDA (digital distribution amplifier)*

DDA distributes SDI baseband signals and DVB-ASI compressed signals. There are a number of DDAs used in a master control, and DDAs with monitoring functions would make it easier to detect failure points.

4) *Active relay*

The Active Relay seamlessly switches SDI baseband signals and DVB-ASI signals. Monitoring functions may be effectively installed to monitor input signals and insert metadata at the output. Automatic switching may also be possible in conjunction with monitoring functions.

5) *U/C (up converter), D/C (down converter), A/C (aspect-ratio converter)*

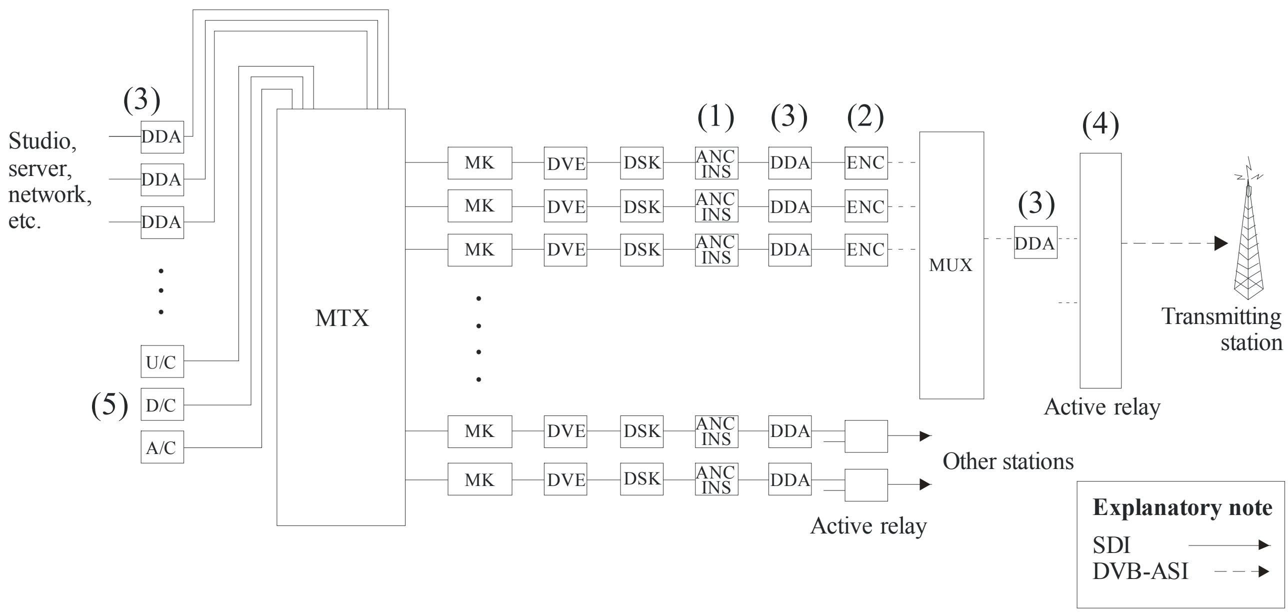
These converters alter the signal formats, and by recording the status of video signals as metadata in these processes, accurate monitoring would become possible at the following monitoring points.

6) *MPX (audio multiplexer)*

The MPX multiplexes audio signals into SDI baseband signals. The status of video and audio signals when multiplexing is done is recorded as metadata.

FIGURE 12

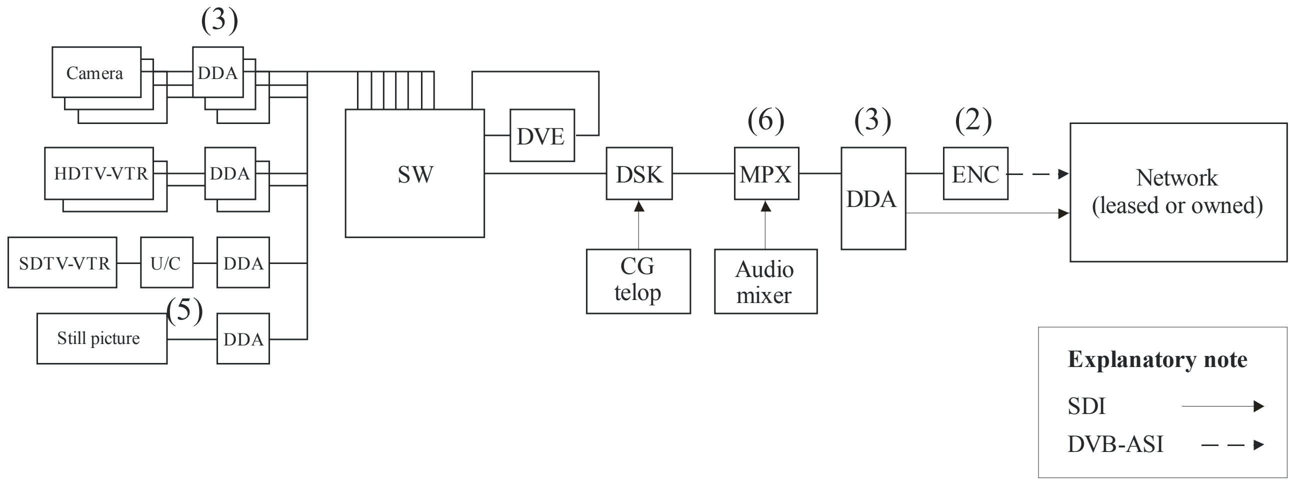
Possible devices to install monitoring points in master control



BT.1865-12

FIGURE 13

Possible devices to install monitoring points in outside broadcasting



BT.1865-13

4 Monitoring by network operators

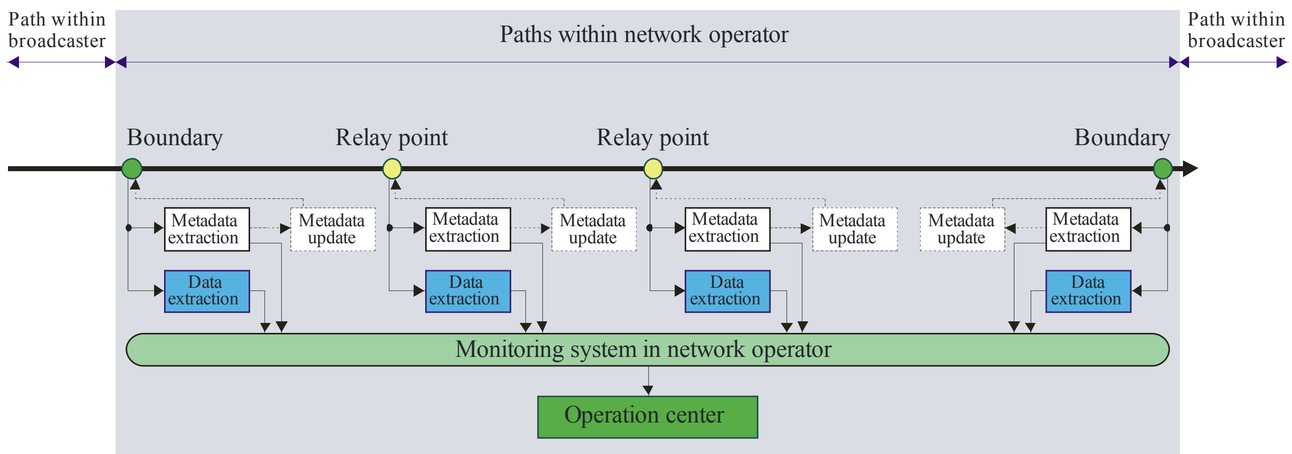
The utilization of metadata for monitoring added to signals is expected to enhance monitoring operations by broadcasters and network operators. However, in the present circumstances, network operators scarcely manage the ancillary data space that is potentially used by broadcasters. Considering this fact, two cases are assumed for network operators to utilize the metadata, i.e. where a) network operators update the metadata, and where b) they do not update the metadata.

4.1 When metadata can be updated

When metadata can be updated within the paths governed by network operators, i.e. when the dashed line in Fig. 14 is valid, monitoring at all points is possible. The organization code to be indicated in the header is that for the network operator. In addition to monitoring using metadata, independent monitoring by the network operator itself may also be conducted, where additional information may be extracted from the signals and utilized.

FIGURE 14

Operation example by network operators



BT.1865-14

4.2 When metadata cannot be updated

When the metadata cannot be updated within the paths governed by network operators, i.e. when the dashed line in Fig. 14 is invalid, the metadata inserted by the sending broadcaster are conveyed without any changes. The network operator conducts its own monitoring using a domestic monitoring system. Nevertheless, the metadata inserted by the broadcaster can be utilized.

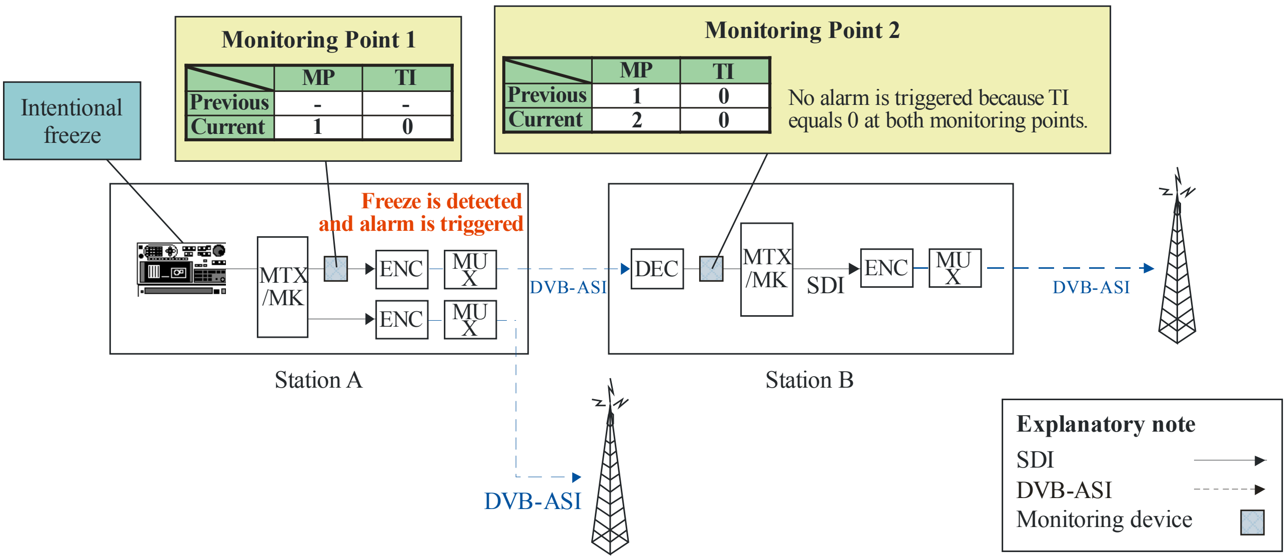
5 Monitoring examples using Type-1 metadata

Figures 15 and 16 illustrate examples of monitoring assuming possible cases where a programme is delivered between broadcasting stations using DVB-ASI compressed signals.

Figure 15 shows a case where intentional freezing is included in a programme played back by a VTR. At monitoring point 1 of the sending station, no metadata are available in the signal played back by the VTR, TI is measured to be nearly zero for the frozen picture and an alarm is triggered. The TI ≈ 0 is then signalled in the metadata. At monitoring point 2 of the receiving station, TI is measured to be nearly zero and the metadata also indicates TI ≈ 0 at the sending station. Consequently, no alarm is triggered.

FIGURE 15

When intentional "freezing" is inserted

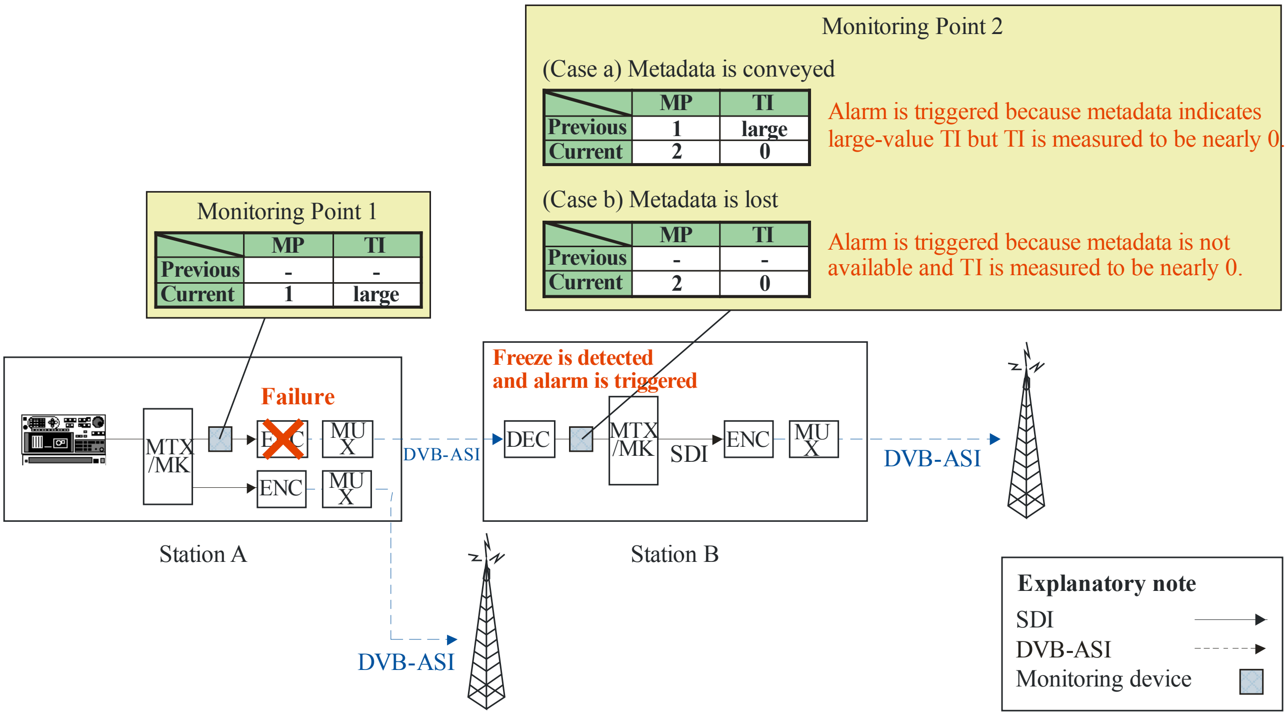


BT.1865-15

Figure 16 shows a case where some failure has occurred at the encoder of the sending station and the decoded picture at the receiving station is frozen.

FIGURE 16

When failure occurs in transmission path



BT.1865-16

At monitoring point 1 of the sending station, video features are measured and inserted as metadata. For a normal picture, TI would be significantly large and no alarm would be triggered. Depending on the failure of the encoder, two cases may be assumed:

a) the added metadata are retained, i.e. the coded bit-stream contains some failure causing freezing but the ancillary data are transported without loss,

b) the metadata are lost, i.e. the ancillary data are also lost:

   a) At monitoring point 2 of the receiving station, TI is measured to be nearly zero for the frozen picture but the metadata indicate significantly large TI. Consequently, a failure can be determined to have occurred between the monitoring points.

   b) At monitoring point 2 of the receiving station, TI is measured to be nearly zero for the frozen picture and the metadata from the previous monitoring point are not available. Consequently, an alarm is triggered.

In both cases, alarms can effectively be triggered using the monitoring system associated with the metadata.

6 Glossary (see also § 3)

MTX (matrix switcher)	Highly sophisticated switcher to select sending resources and used for master control. ANC data can be conveyed through MTX
SW (switcher)	General switcher (video mixer) used for studios and OB vans, and distinguished from switcher used for master control. ANC data usually cannot be conveyed through SW
DVE (Digital video effect)	Equipment for electronically generating special video effects
MK (Mixer and keyer)	Equipment for mixing and superimposing video or audio
DSK (Down stream keyer)	Equipment for superimposing video

Appendix 3
(to Annex 1)

Experimental results of Type-1 video parameters for test sequences

This appendix presents the experimental results of measuring the Type-1 video parameters (SI and TI) for test sequences. This experiment was conducted to test and verify the usability of SI and TI for operational monitoring.

1 Set of test sequences

The following test sequences were used for taking the measurements of SI and TI. Each sequence consisted of several scenes, which are typical in cases of transmission failure or malfunctions in transmission equipment.

1.1 Blackout I (almost flat texture and monochrome)

Black → Blue → White → Red → Black with noise → White with noise → Vertical stripe

1.2 Blackout II (almost flat texture and monochrome)[*1]

Black → Blue → White → Red → Black with noise → White with noise → Vertical stripe II[*2]

[*1] All scenes except for "Vertical stripe II" are the same as for Blackout I.

[*2] Different textures were inserted every 15 frames.

1.3 Freeze I

Black with noise I → Red with noise I → Blue with noise → Red with noise → Blue with noise → Grey → Vertical stripe → Flower Basket[*1] → Rustling Leaves[*1]

[*1] Cutouts from HDTV standard test sequences.

1.4 Freeze II

Black with noise I → Red with noise I → Blue with noise → Red with noise → Blue with noise → Grey → Animation (with less inter-frame difference) → Flower Basket[*1] → Rustling Leaves[*1]

[*1] Cutouts from HDTV standard test sequences.

1.5 Natural image sequence (a summer day)

Fade-out to black is included. Night scenes are also included.

1.6 Natural image sequence (drama)

Intentional blackout is inserted.

1.7 Natural image sequence (mobile and calendar)

1.8 Animation

1.9 Superimpose and wipe

1.10 Superimpose

Text was superimposed at the top of an urban cityscape, and the text was changed during the sequence. Two types of urban cityscapes with different magnitudes of motion were used. (See Fig. 17 a) for the image with slight motion.)

2 Wipe with text

The images were scaled down and wiped to show some text at the top, and the text was changed during the sequence. Two types of urban cityscapes having different magnitudes of motion were used. (See Fig. 17 b) for the image with significant motion.)

FIGURE 17

Test sequences with superimposed text and wiped with text

a) Slight motion b) Significant motion

BT.1865-17

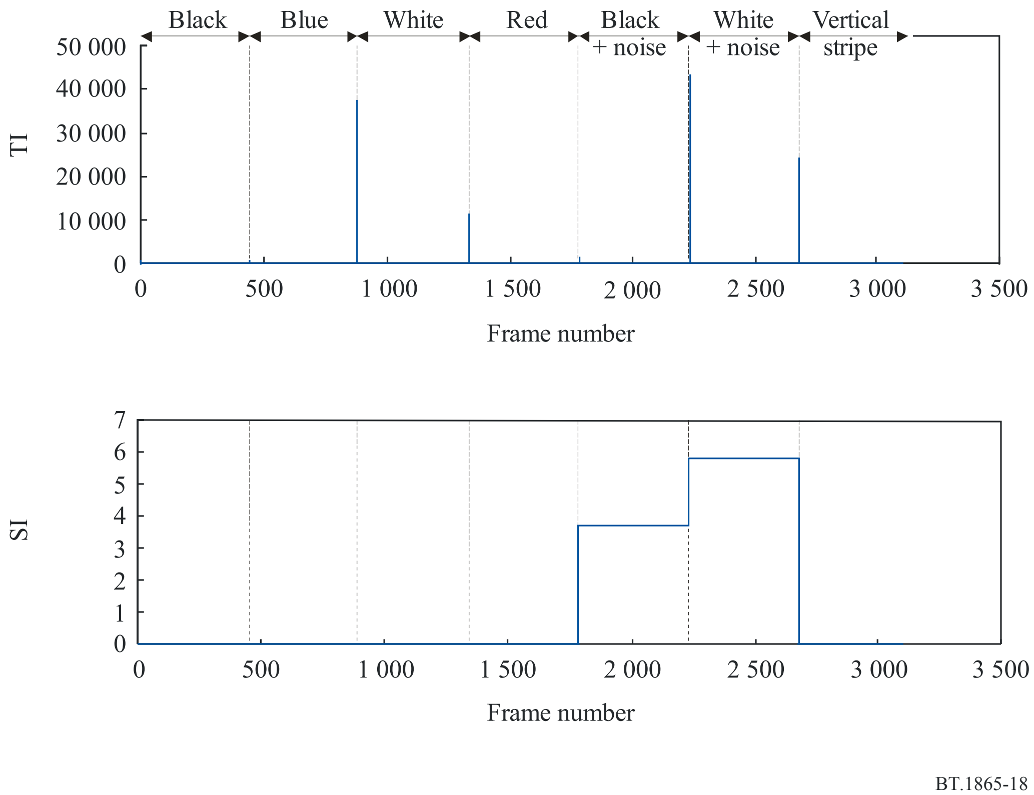
3 Measurement results

3.1 Blackout I

In all the frames except for those after a scene change, TI = 0. For the scenes without noise, SI = 0.

FIGURE 18

Blackout I



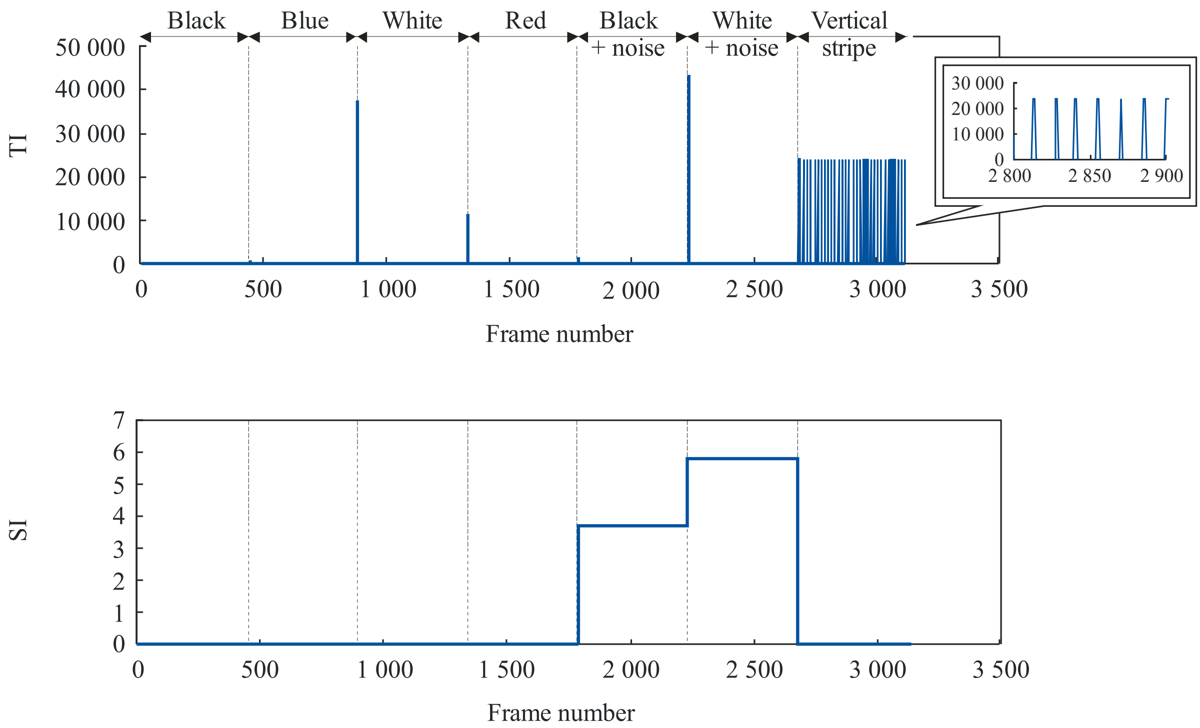
3.2 Blackout II

In terms of TI, the results are the same as those for Blackout I, except for "Vertical stripe II" in which different textures are inserted every 15 frames.

In terms of SI, all the scenes have the same results as Blackout I.

FIGURE 19

Blackout II



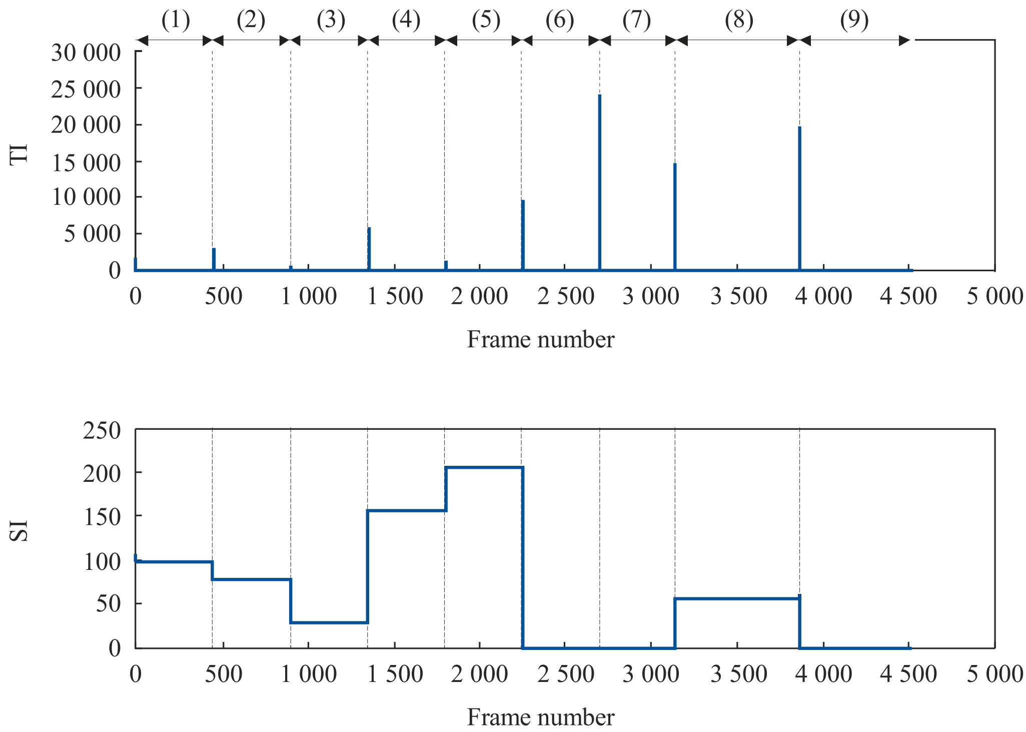
BT.1865-19

3.3 Freeze I

In the scenes with noise, TI is not exactly zero but is less than 1.0. This may have been caused during the authoring process of the test sequence.

FIGURE 20

Freeze I



(1) Black with noise, (2) Red with noise 1, (3) Blue with noise 1, (4) Red with noise 2,
(5) Blue with noise 2, (6) Grey, (7) Vertical stripe, (8) Flower Basket, (9) Rustling Leaves.
(Noise 1: noise is black, Noise 2: noise is white.)

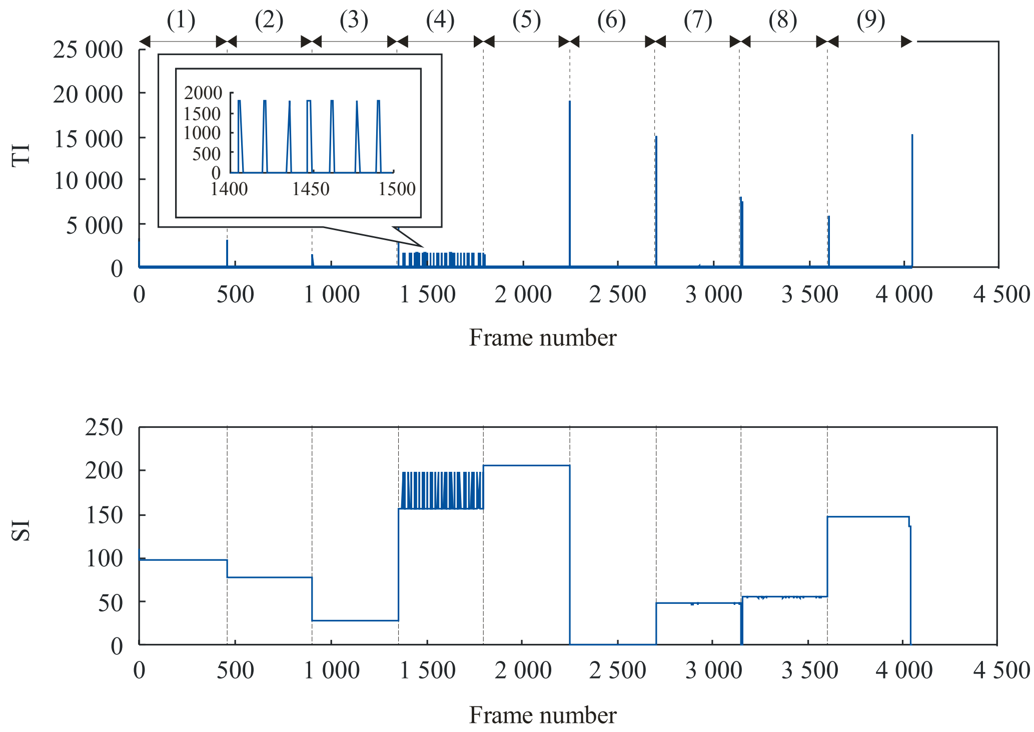
BT.1865-20

3.4 Freeze II

The TI and SI values are different to those in the Freeze I sequence because the noise pattern in scene 4 changes frame-by-frame, which is the case in the Freeze I sequence.

FIGURE 21

Freeze II



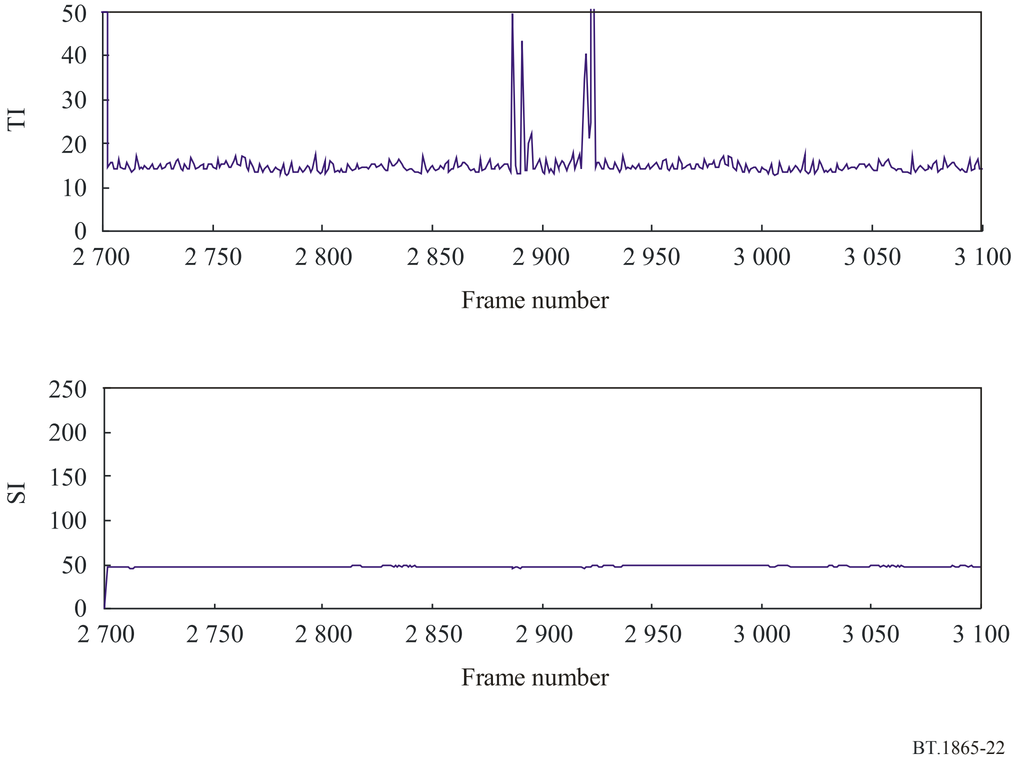
(1) Black with noise, (2) Red with noise 1, (3) Blue with noise 1, (4) Red with noise 2, (5) Blue with noise 2, (6) Grey, (7) Animation, (8) Flower Basket, (9) Rustling Leaves. (Noise 1: noise is black, Noise 2: noise is white.)

BT.1865-21

Scene 7 (animation) is a moving picture (not a still image), but this scene may be falsely recognized as a still picture because only a small region in the frame is in motion. In this scene, TI is always more than 10 and the data series is different to that of a still picture. Even animation images, which tend to have fewer inter-frame differences, can be distinguished from still pictures by monitoring the TI.

FIGURE 22

Scene 7 in Freeze II



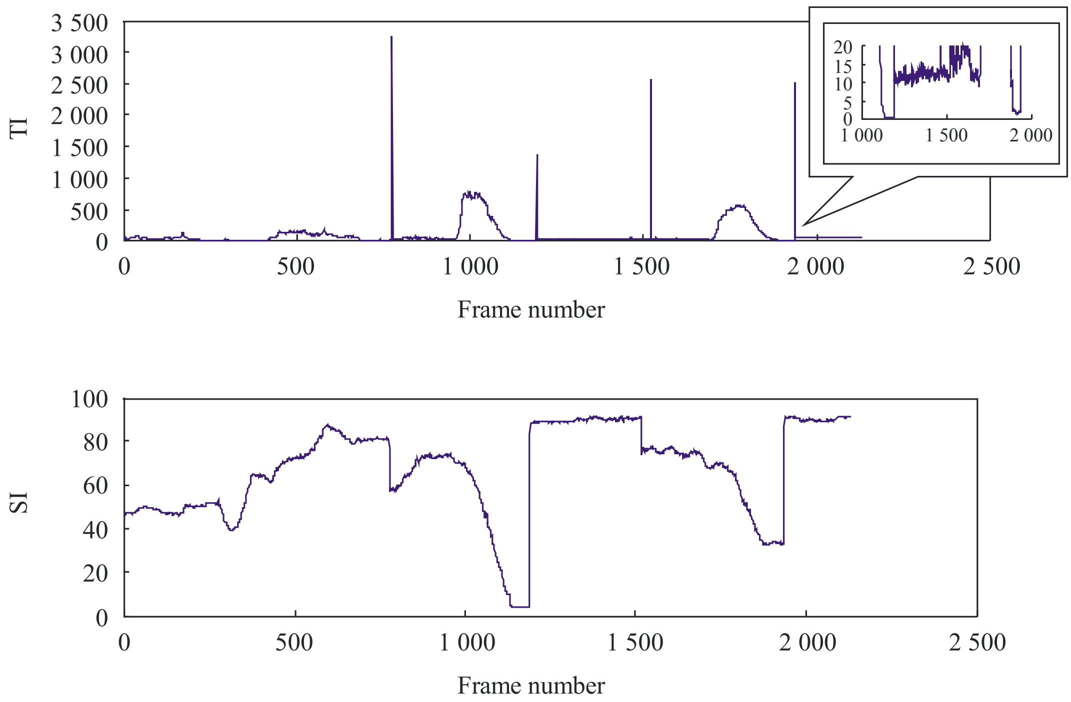
3.5 Natural image sequence (a summer day)

This sequence includes a transition (fade out) of scenes. During the transition, some black frames are inserted and this may be falsely recognized as a blackout. The minimum TI and SI during the transition correspond to 0.45 and 4.5 (at frames 1 150-1 200), and this scene is very similar to blackout in terms of SI and TI.

Around frame 1 900, there is a night canal scene, which is very dark. However, SI in this scene is between 33 and 34 and this scene can thus be distinguished from blackout.

FIGURE 23

Natural image sequence (a summer day)



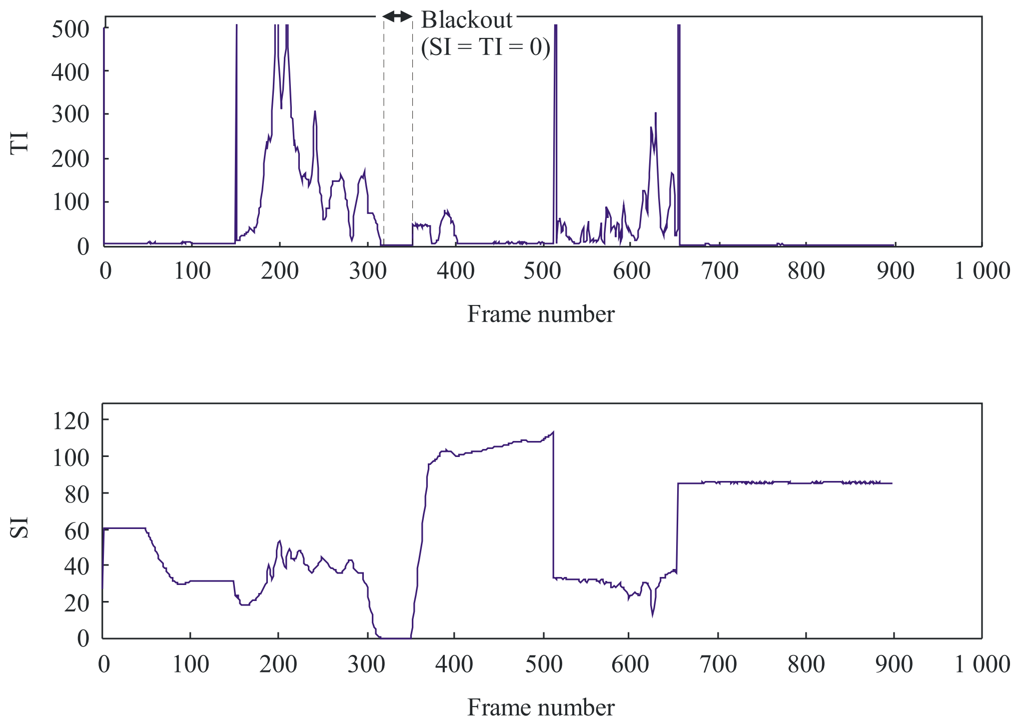
BT.1865-23

3.6 Natural image sequence (drama)

In this sequence, there is a transition of scenes, during which completely black frames (0 SI and TI) are intentionally inserted for about one second. This intentional blackout can be distinguished from blackouts caused by transmission failure or the malfunction of video equipment by using metadata history.

FIGURE 24

Natural image sequence (drama)



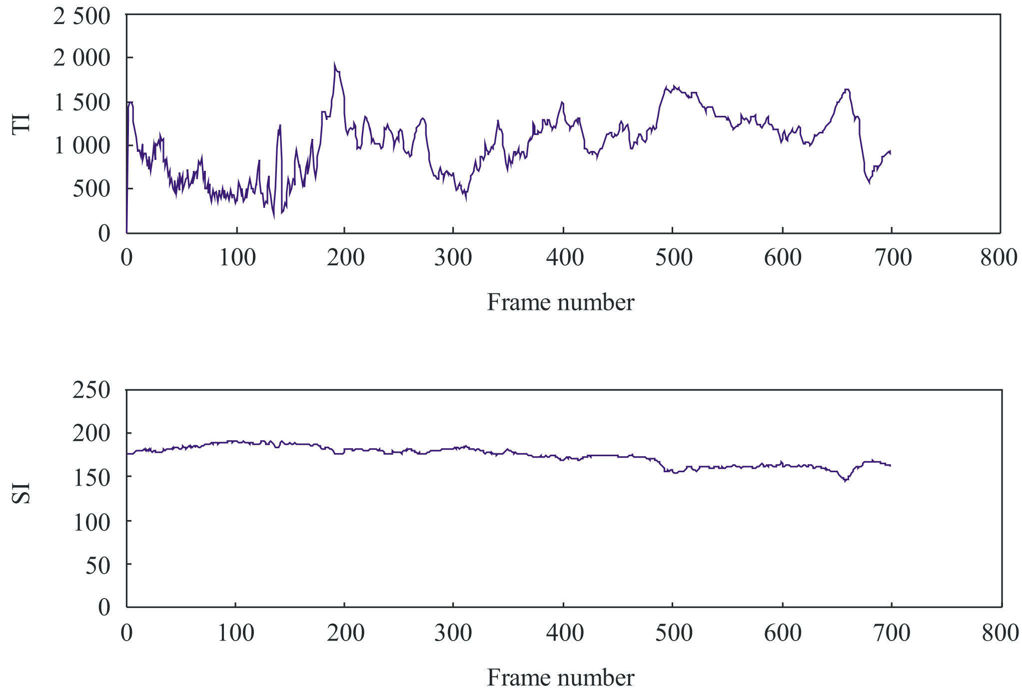
BT.1865-24

3.7 Natural image sequence (mobile and calendar)

This is a standard sequence, which has a complex texture and several motions. As both TI and SI are very high, it is easily distinguished from blackouts and freezing.

FIGURE 25

Natural image sequence (mobile and calendar)



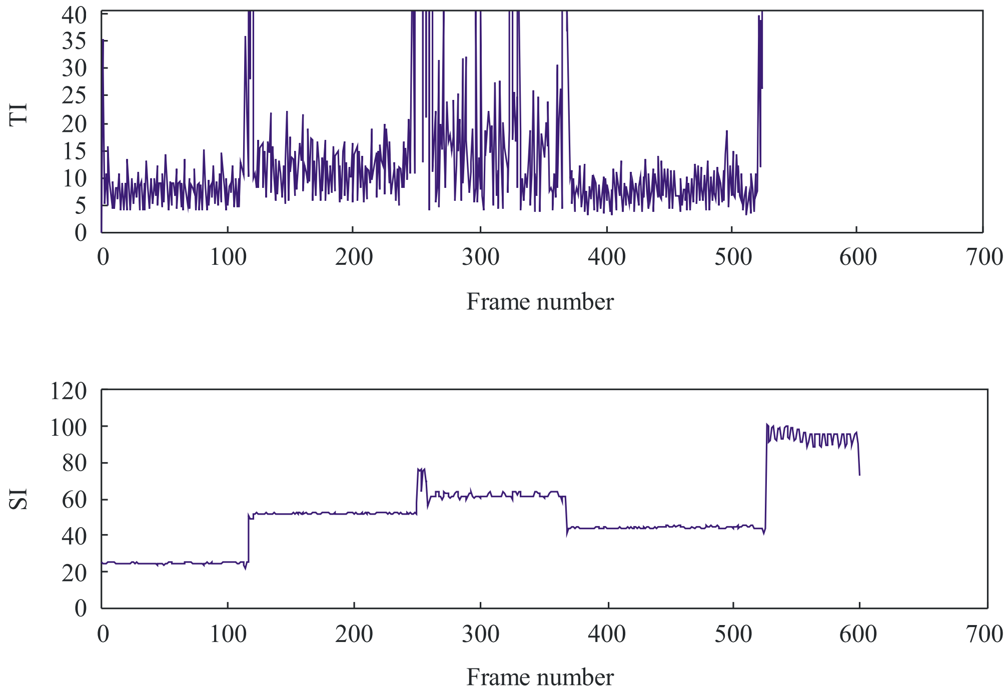
BT.1865-25

3.8 Animation

This sequence is animation footage. However, the series of TI has different characteristics from that of scene 7 in Freeze II; TI changes are large every 1-2 frames. This may be because this sequence was originally in 24-fps film format and then dubbed into Digital VCR with telecine conversion, whereas scene 7 in Freeze II is in NTSC format.

FIGURE 26

Animation



BT.1865-26

3.9 Superimpose and wipe

TI and SI are compared for three sequences, without superimpose and wipe, with superimpose, and with wipe, and are shown in Figs 27 and 28. At the timing the superimposed text is changed (at around the frame number 300), TI is significantly increased. The TI of "without superimpose and wipe" (blue) and that of "with superimpose" (red) are very comparable, but that of "with wipe" (green) is lower than that of the others. The SIs of "with superimpose" and "with wipe" are larger than that of "without" and the TI of "with superimpose" is the largest.

FIGURE 27

Superimpose and wipe for sequence with slight motion

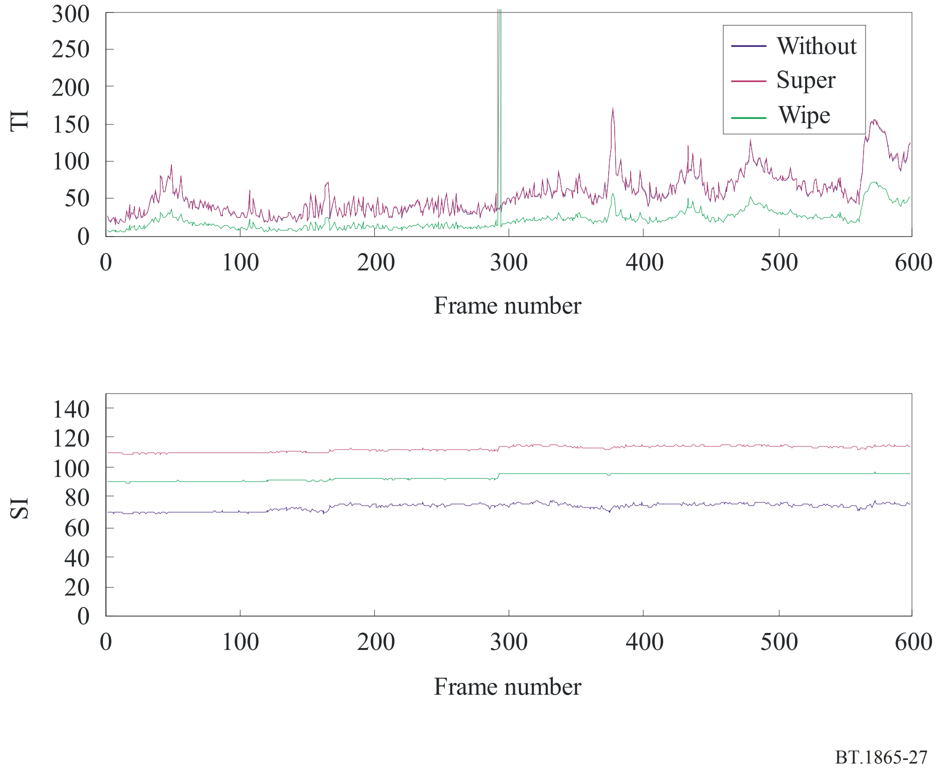
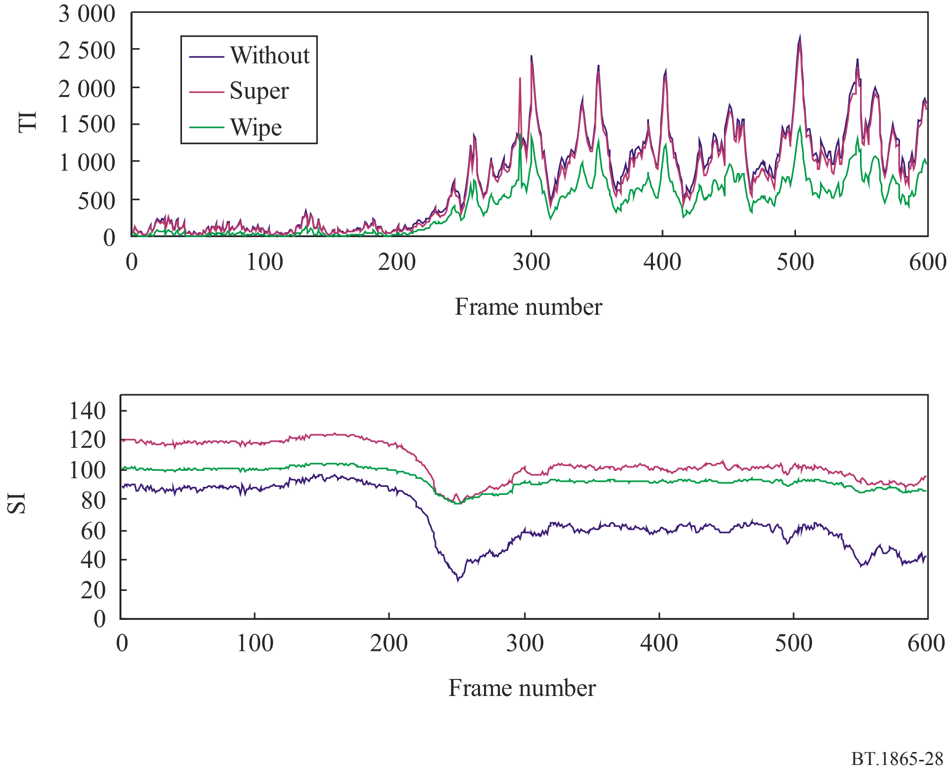


FIGURE 28

Superimpose and wipe for sequence with significant motion



Appendix 4
(to Annex 1)

Experimental results by measuring type-1 audio parameters

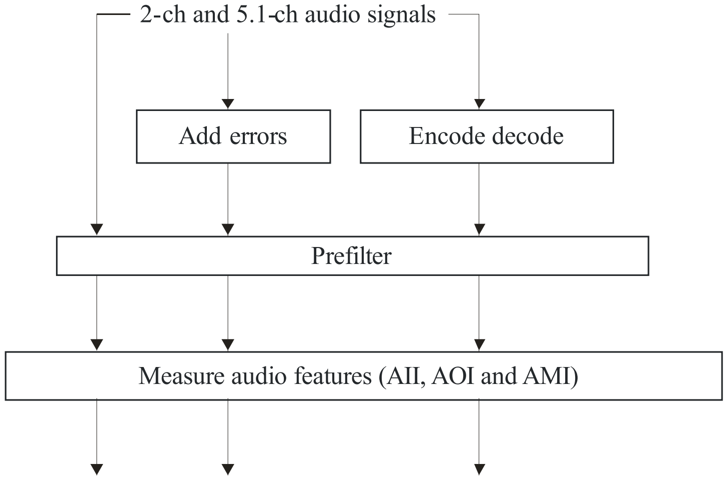
This appendix presents the experimental results obtained by measuring Type-1 audio parameters (AII, AOI, and AMI) for test sequences. This experiment was conducted to test and verify the usability of AII, AOI, and AMI to detect audio signal errors in operational monitoring.

Figure 29 and Table 7 show the configuration for the experiment. Figure 30 shows how impairments were added to the audio test signals. Table 8 gives a list of Figures shown in this appendix. They are source audio wave forms, extracted audio features for the source, and differences in audio features between the source and impaired signals for each test material.

It has been confirmed that the proposed audio features could effectively detect audio errors and were insensitive to distortions caused by low bit-rate audio coding.

FIGURE 29

Configuration of experiment on audio features



BT.1865-29

TABLE 7

Configuration of experiment on audio features

Audio source	2-ch	1 SQAM Tr.49 Female speech in English 2 SQAM Tr.61 Soprano and orchestra (Sound quality assessment material (SQAM), EBU Tech 3253)
	5.1-ch	3 Pines of Rome for Symphonic Poem/Ottorino Respighi (Surround Sound Reference Disc/Surround Study Group of AES Japan Section (AESSJ001-2), Disc 2, Tr.3-4, 7:13” 00~7:43”00)
		48 kHz sampling, N = 1602 (Number of samples per frame)
Audio errors (See Fig. 30)		Random noise: First two samples of each frame are replaced by random noise for 4 s
		Mute: First 50 samples and last 50 samples of each frame are replaced by value 0x0000 for 4 s.
Encode-decode		AAC, 256 kbit/s (2-ch)

FIGURE 30

Impaired signal

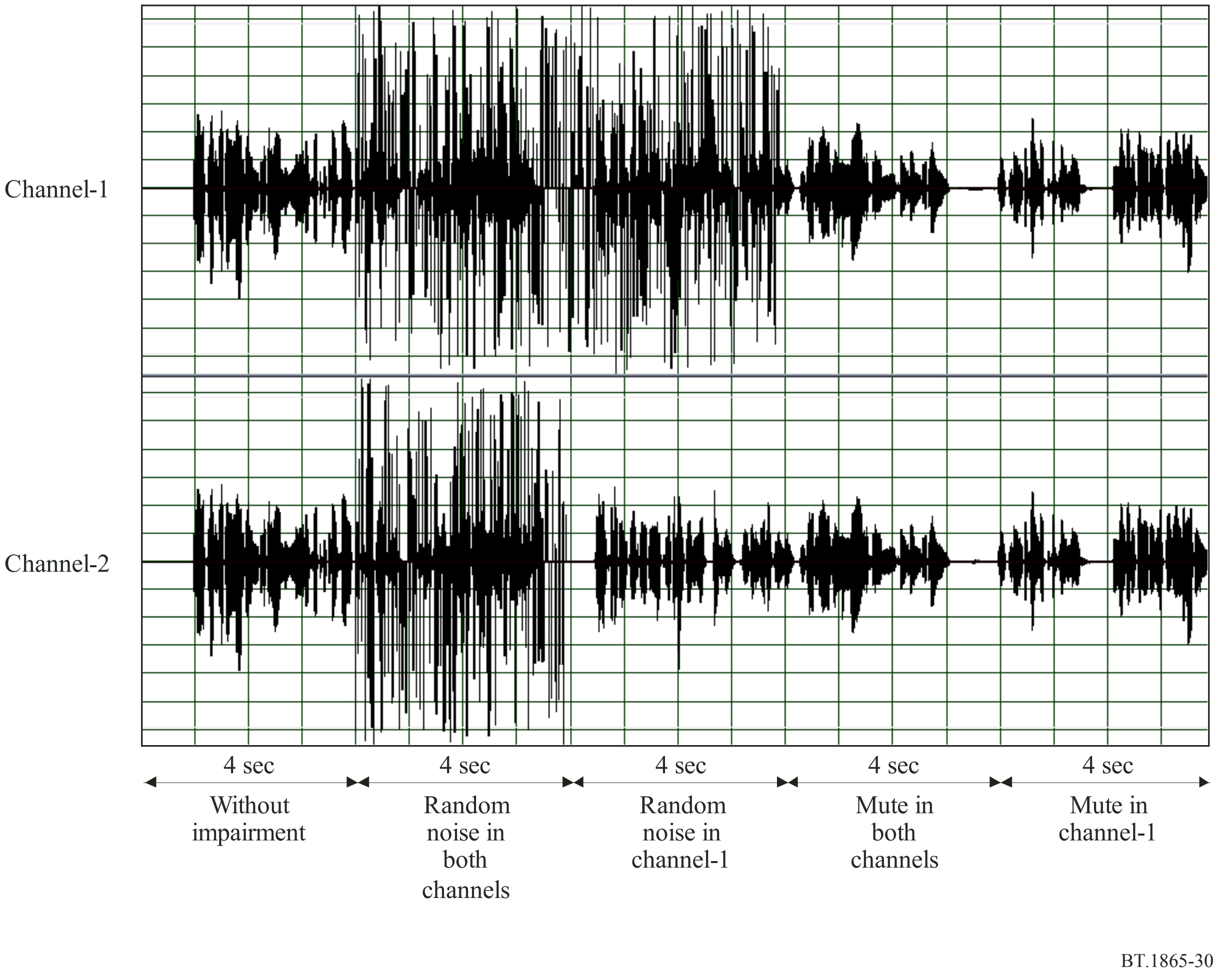


TABLE 8

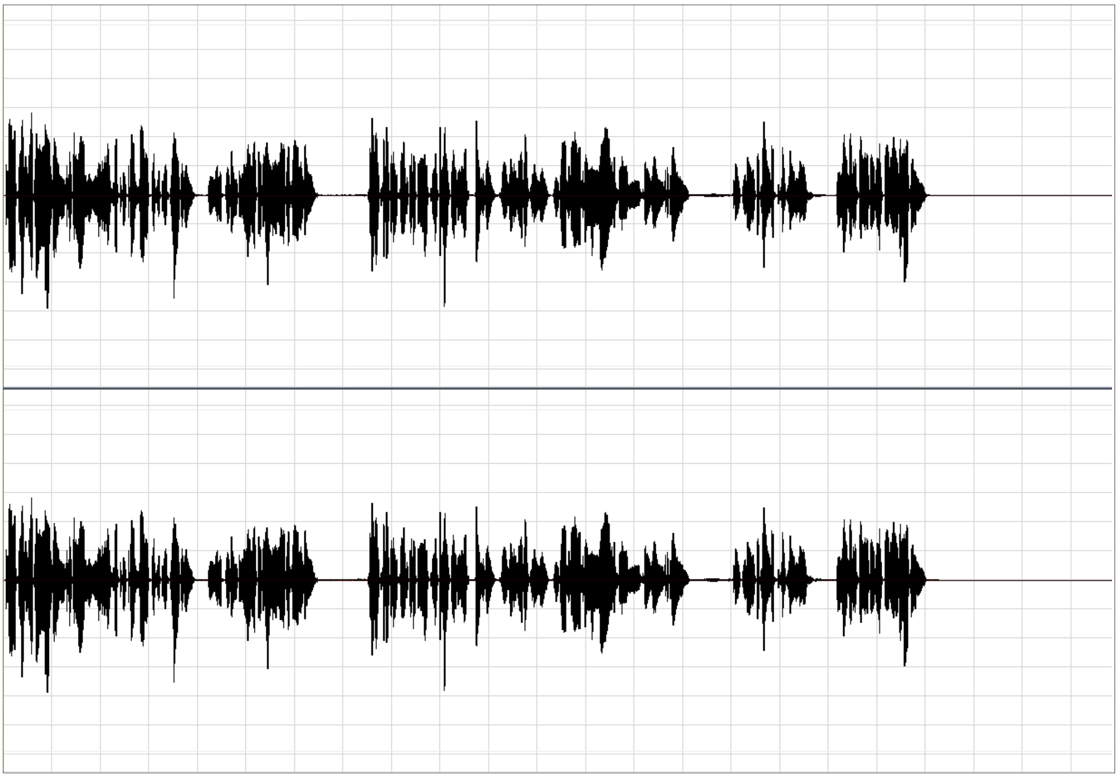
List of figures

Sound source	**SQAM Tr.49 Female speech in English (2-ch)**	**SQAM Tr.61 Soprano and orchestra (2-ch)**	**Pines of Rome for Symphonic Poem/Ottorino Respighi (5.1-ch)**
Wave form of original sound source	A4-3	A4-10	A4-15
AII and AOI (original)	A4-4	A4-11	A4-16
Difference in AII and AOI (original – impaired)	A4-5	A4-12	A4-17
Difference in AII and AOI (original – coded)	A4-6		
AMI-1 and AMI-2 (original)	A4-7	A4-13	A4-18
Difference in AMI-1 and AMI-2 (original – impaired)	A4-8	A4-14	A4-19
Difference in AMI-1 and AMI-2 (original – coded)	A4-9		

1 SQAM Tr.49 Female speech in English (2-ch)

FIGURE 31

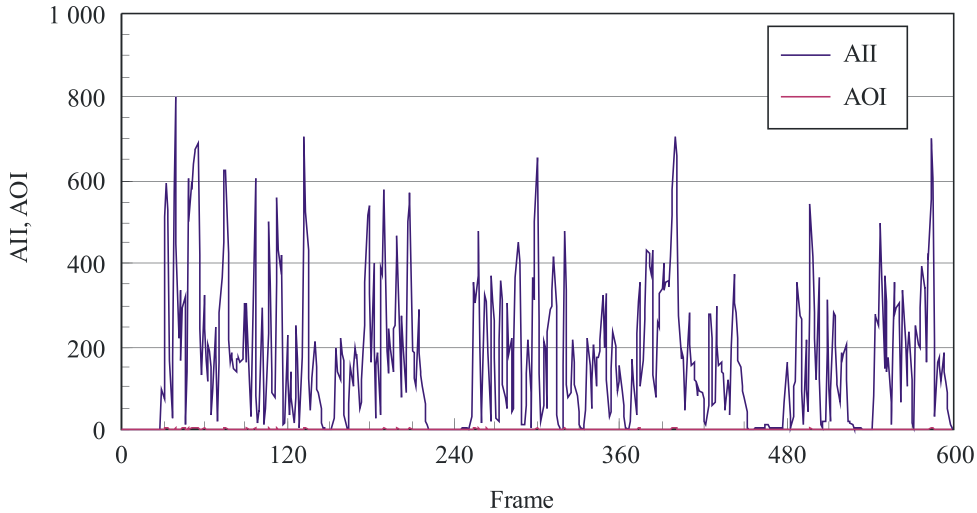
Original sound source wave forms



BT.1865-31

FIGURE 32

AII and AOI (original)

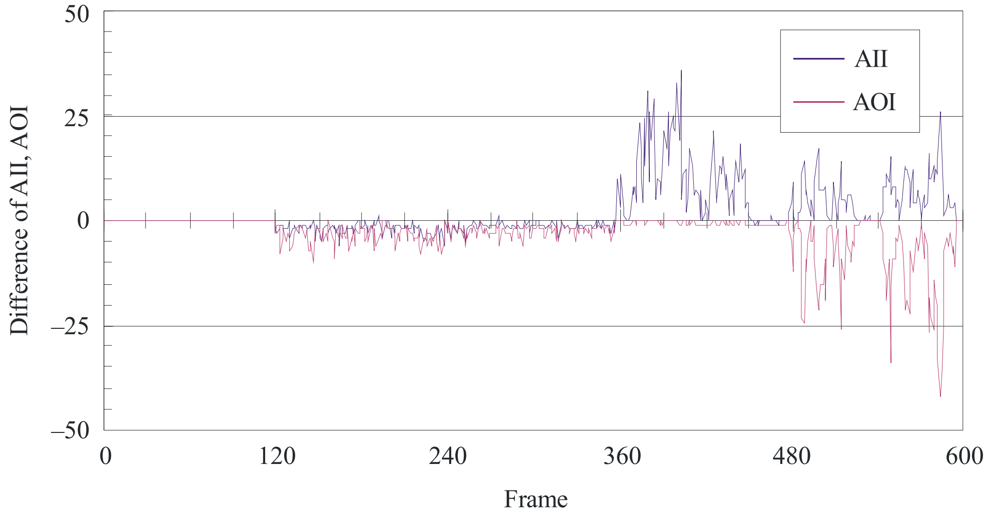


Note 1 – AOI was almost zero over the duration.

BT.1865-32

FIGURE 33

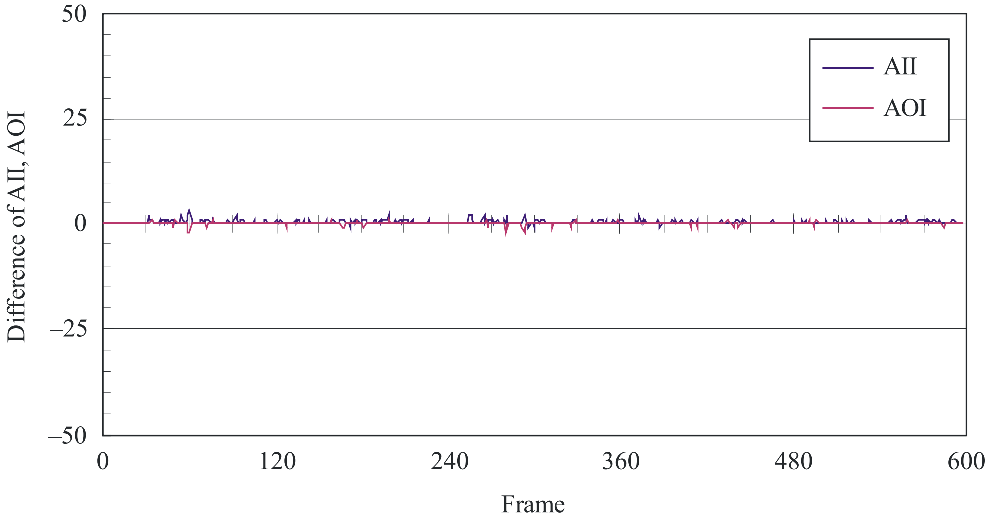
Difference in AII and AOI between original and impaired signals



BT.1865-33

FIGURE 34

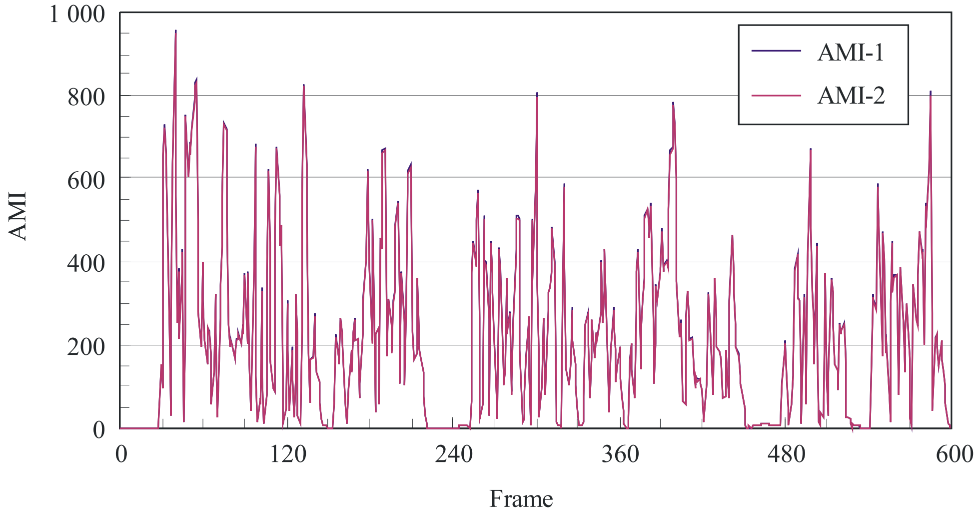
Difference in AII and AOI between original and coded signals



BT.1865-34

FIGURE 35

AMI-1 (left) and AMI-2 (right) (original)

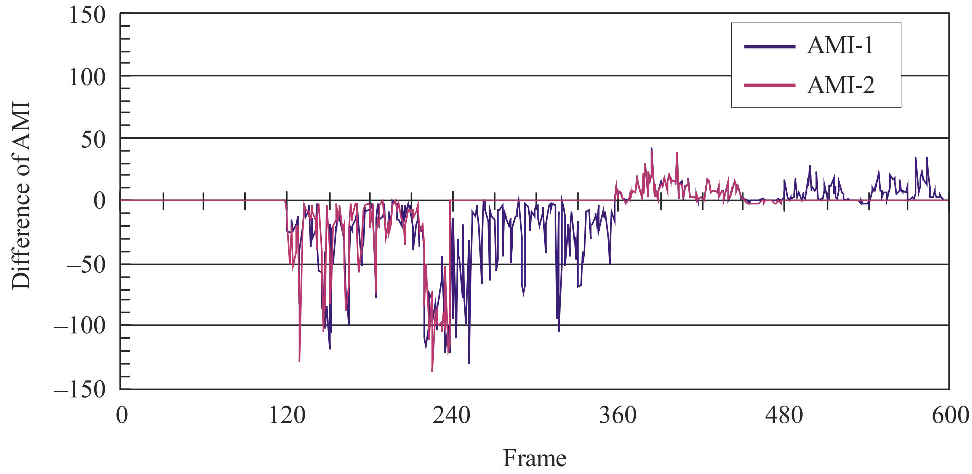


Note 1 – AMI-1 was almost hidden by AMI-2.

BT.1865-35

FIGURE 36

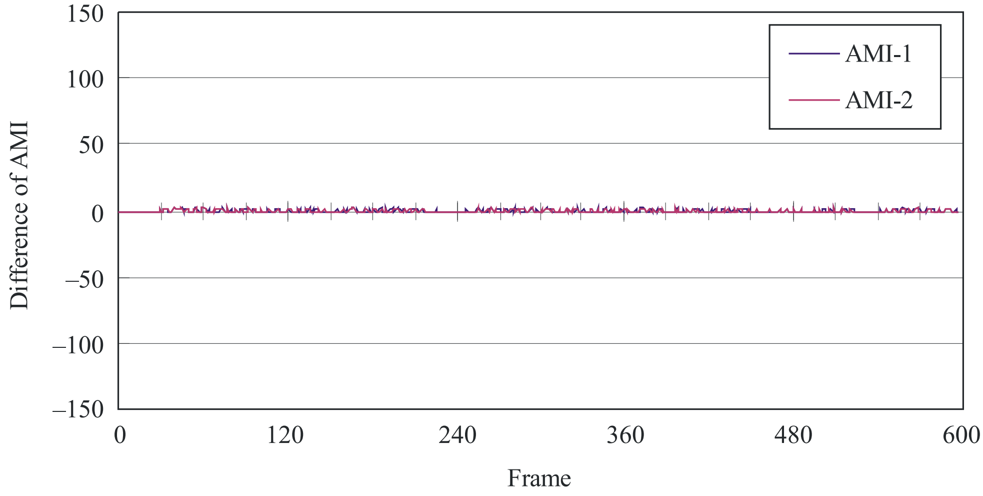
Difference in AMI-1 and AMI-2 between original and impaired signals



BT.1865-36

FIGURE 37

Difference in AMI-1 and AMI-2 between original and coded signals

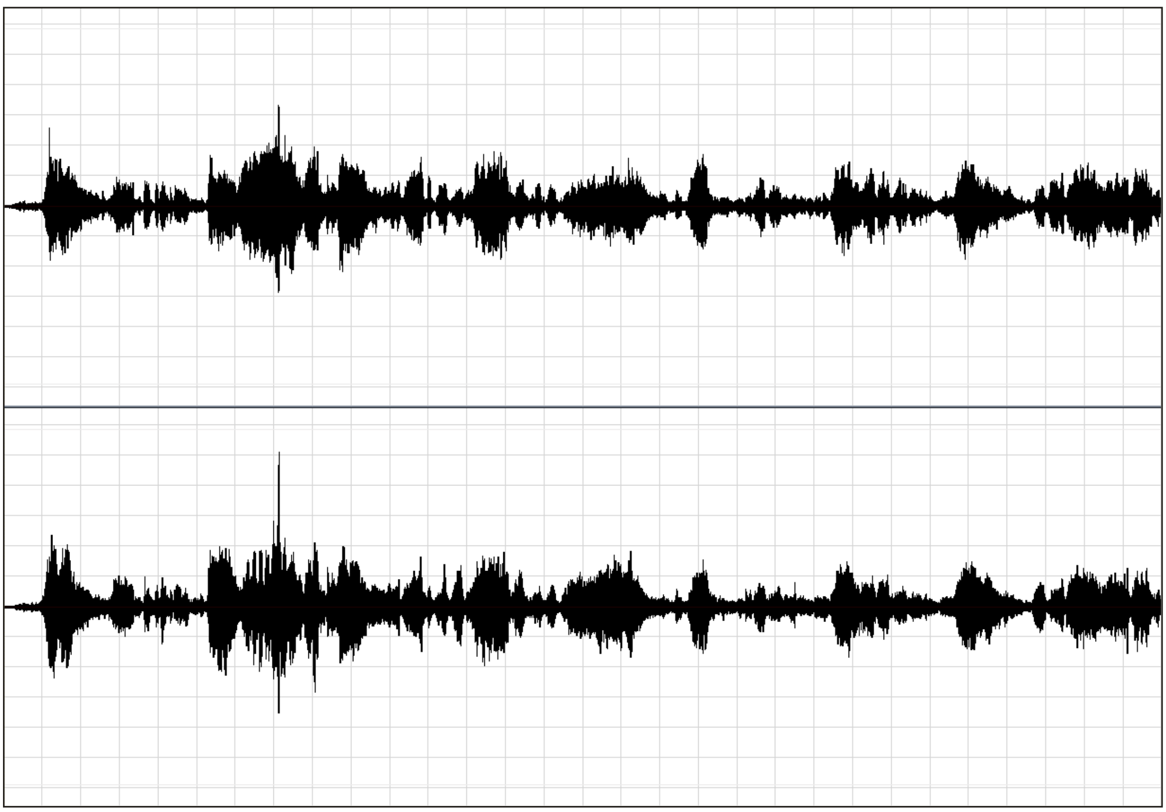


BT.1865-37

2 SQAM Tr.61 Soprano and orchestra (2-ch)

FIGURE 38

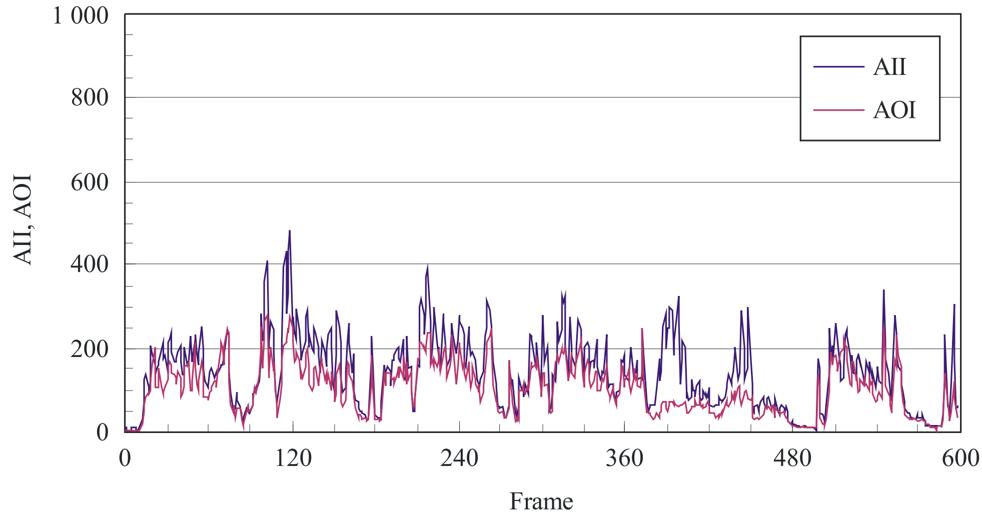
Original sound source wave forms



BT.1865-38

FIGURE 39

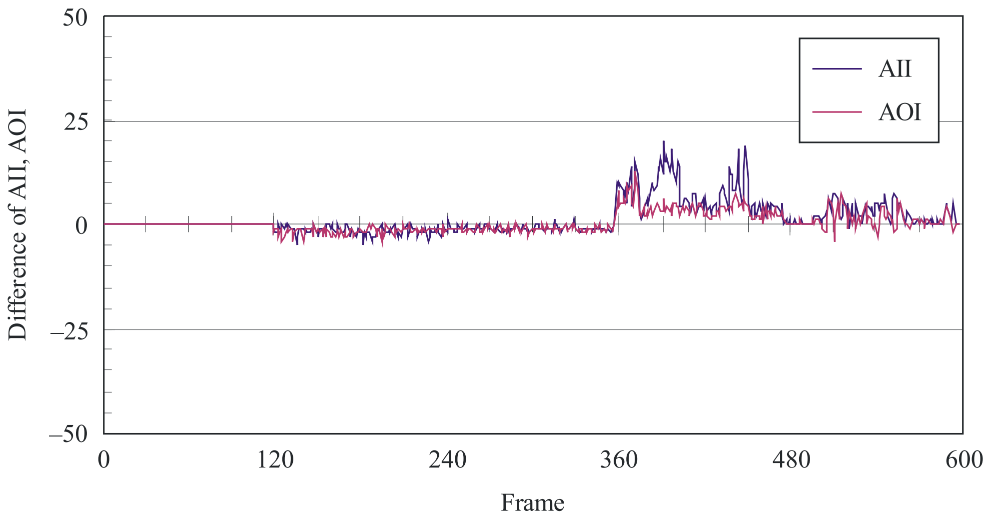
AII and AOI (original)



BT.1865-39

FIGURE 40

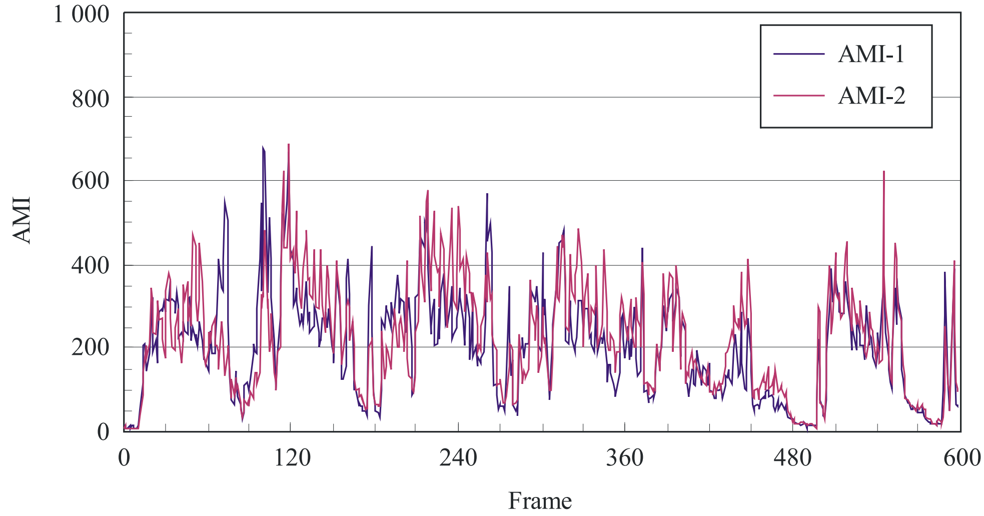
Difference in AII and AOI between original and impaired signals



BT.1865-40

FIGURE 41

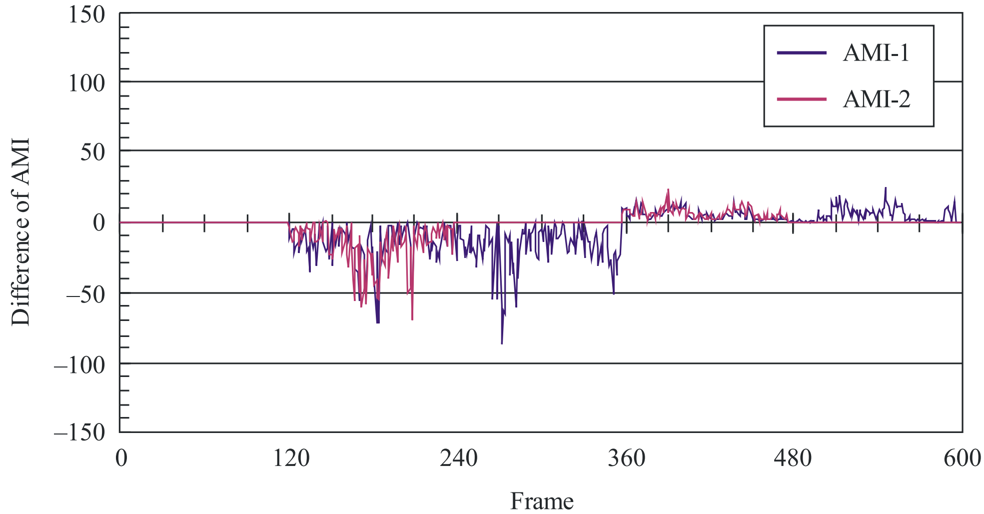
AMI-1 and AMI-2 (original)



BT.1865-41

FIGURE 42

Difference in AMI-1 and AMI-2 between original and impaired signals



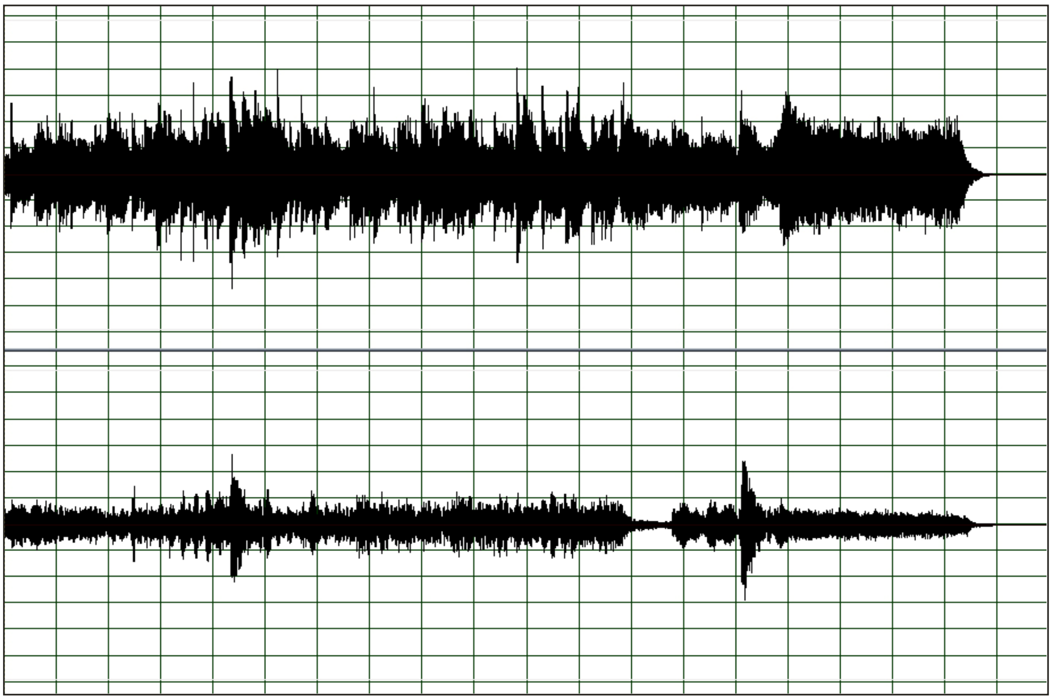
BT.1865-42

3 Pines of Rome for Symphonic Poem/Ottorino Respighi (5.1-ch)

This sound segment is a part of track 3-4 of Disc 2, from 7:13″ 00 to 7:43″00, of the “Surround Sound Reference Disc”/Surround Study Group of AES Japan Section (AESSJ001-2). Graphs have only been shown for the channel pair of Centre and LFE.

FIGURE 43

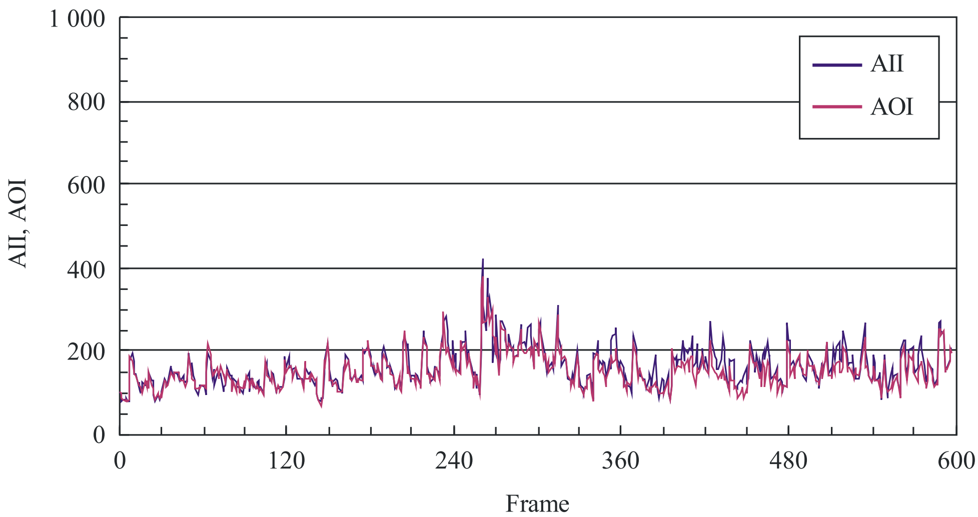
Original sound source wave forms



BT.1865-43

FIGURE 44

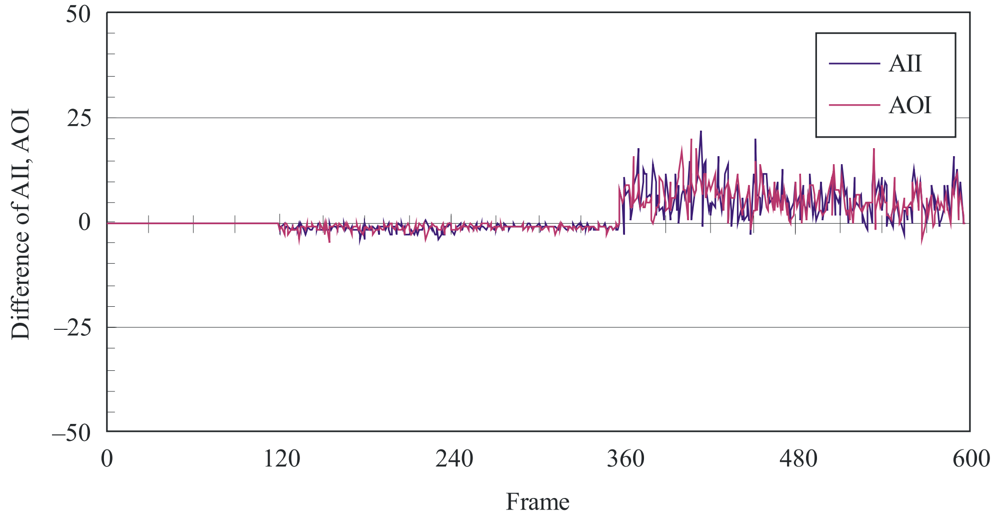
AII and AOI (original)



BT.1865-44

FIGURE 45

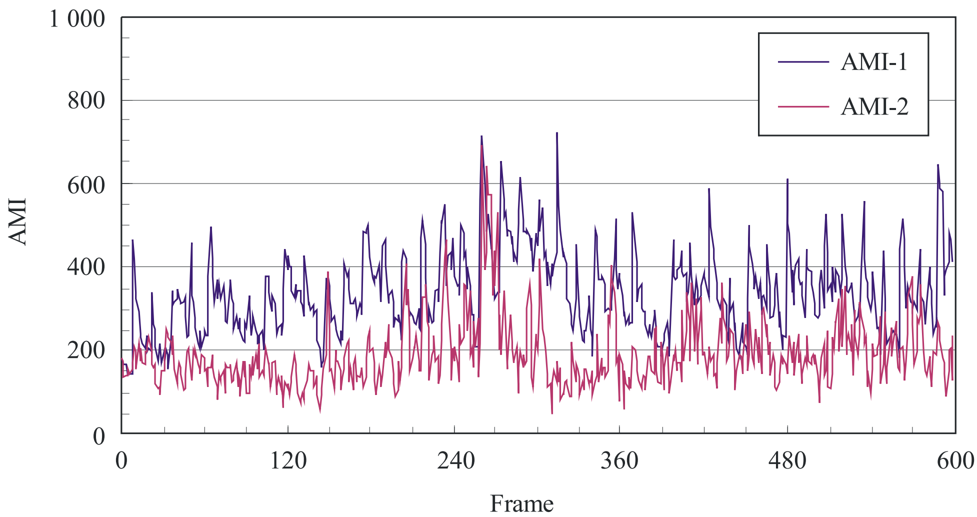
Difference in AII and AOI between original and impaired signals



BT.1865-45

FIGURE 46

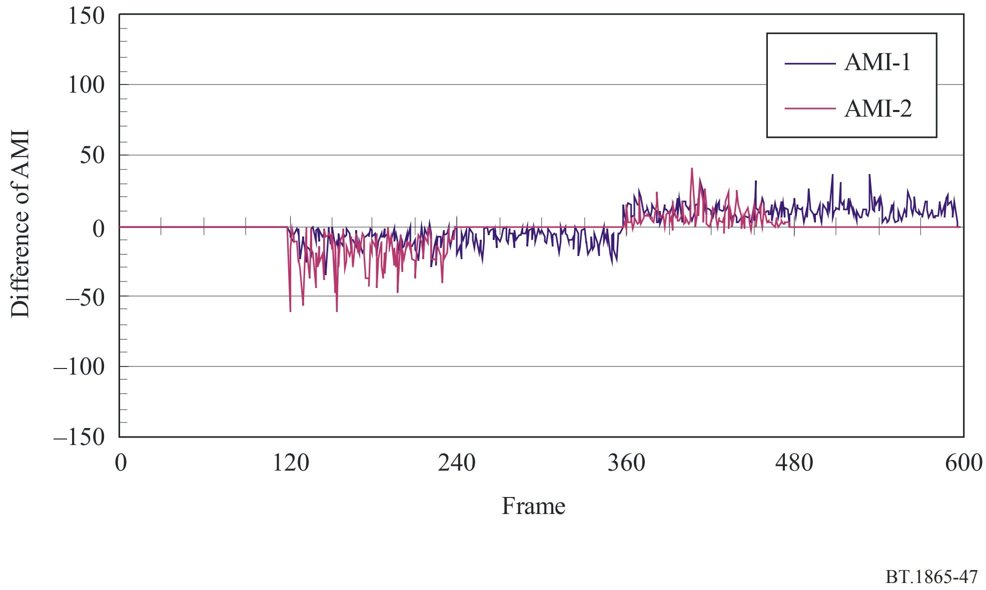
AMI-1 and AMI-2 (original)



BT.1865-46

FIGURE 47

Difference in AMI-1 and AMI-2 between original and impaired signals



Appendix 5
(to Annex 1)

Additional explanation of metadata for broadcast operational monitoring

1 Example usage of metadata for broadcast operational monitoring

Metadata may be used for several purposes in broadcast operational monitoring. Three of these are discussed below:

1 *Metadata for quality management*

One possible method of monitoring the degree of quality degradation in audiovisual content is to derive some feature information on the audiovisual signal and multiplex it to the signal in the upper broadcast chain, and compare the original feature information with the feature information derived from the received audiovisual signal in the lower broadcast chain.

2 *Metadata for discriminating*

Between intentional audiovisual effects and trouble as the cause of blackouts, freezing, muting, and noise are usually regarded as errors in audiovisual signals. However, such signals are occasionally used intentionally for audiovisual effects. If information indicating that an unusual signal is intentionally used is flagged by metadata at the stage a programme is produced, a monitoring system at a later monitoring point in the broadcast chain will not need to issue an alarm when it detects these unusual signals.

3 *Metadata for checking lip-sync*

By adding timing information to both video and audio signal frame-by-frame beforehand, the relative timing at a later process could very easily be checked.

2 Examples of possible metadata for broadcast operational monitoring

A number of metadata as described below may be used for broadcast operational monitoring.

1 *Metadata related to signal format*

Video	Format (horizontal and vertical samples, field/frame rate, luminance/chrominance sampling structure, scanning format) Bit length Aspect ratio
Audio	Channel mode (e.g. mono, stereo, and multichannel) Sampling frequency Bit length Alignment level

2 *Metadata related to quality of audiovisual signals*

Feature information derived from audiovisual signals for checking/measuring degree of degradation (e.g. motion and activity) Video/audio quality metrics (VQM/AQM) of transmitted video and audio signals

3 *Metadata related to status of audiovisual signals*

Spatial/temporal shift from original signals Relative timing between audio and video signals Field information for 2-3 pull-down images Time code Indication of unusual signals such as freezing, blackouts, and muting

4 *Metadata related to intentional audiovisual effects*

Indication of intentional audiovisual effects

5 *Metadata related to broadcast programming*

Event ID Service type Ancillary data such as teletext and control signals

6 *Metadata related to network operations*

Name of sender and recipient Start and end time of transmission Name of relay site and OB van Type and name of transmission line Bit rate Received field strength Bit error rate

7 *Metadata related to troubleshooting*

Type of error Origin of error History of error

3 Store and transfer of metadata for broadcast operational monitoring

The format and means of storing and transferring metadata should be specified to store and transfer them for broadcast operational monitoring in the broadcasting chain.