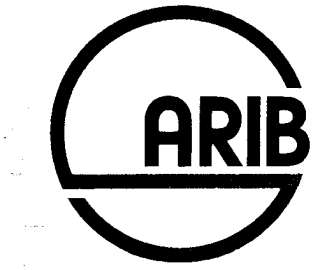


ENGLISH TRANSLATION

METADATA TO MONITOR ERRORS OF VIDEO AND AUDIO SIGNALS ON A BROADCASTING CHAIN

ARIB TECHNICAL REPORT

ARIB TR-B29 Version 1. 1

Established on July 29th 2009 Version 1.0
Revised on July 15th 2010 Version 1.1

Association of Radio Industries and Businesses

General Notes to the English translation of ARIB Standards and Technical Reports

1. The copyright of this document is ascribed to the Association of Radio Industries and Businesses (ARIB).
2. All rights reserved. No part of this document may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior written permission of ARIB.
3. The ARIB Standards and ARIB Technical Reports are usually written in Japanese and approved by the ARIB Standard Assembly. This document is a translation into English of the approved document for the purpose of convenience of users. If there are any discrepancies in the content, expressions, etc., between the Japanese original and this translated document, the Japanese original shall prevail.
4. The establishment, revision and abolishment of ARIB Standards and Technical Reports are approved at the ARIB Standard Assembly, which meets several times a year. Approved ARIB Standards and Technical Reports, in their original language, are made publicly available in hard copy, CDs or through web posting, generally in about one month after the date of approval. The original document of this translation may have been further revised and therefore users are encouraged to check the latest version at an appropriate page under the following

   URL:

   http://www.arib.or.jp/english/index.html

Introduction

With participation of radio communication equipment manufacturers, broadcasting equipment manufacturers, telecommunication operators, broadcasters and general equipment users, Association of Radio Industries and Businesses (ARIB) defines basic technical requirements for standard specifications of radio equipment, etc. as an "ARIB STANDARD" or “ARIB TECHNICAL REPORT” in the field of various radio systems.

An ARIB TECHNICAL REPORT provides the industry with specifications designed to ensure the quality and compatibility of radio equipment, particularly with regards to measurement and operational methods, based on “ARIB STANDARD” derived from technical standards released by the national government as well as voluntary standards used in the industry.

An ARIB TRECHNICAL REPORT herein is published as "Metadata to monitor errors of audio and video signals on a broadcasting chain." In order to ensure fairness and transparency in the defining stage, the technical report was set by consensus of the standard council with participation of interested parties including radio equipment manufacturers, telecommunication operators, broadcasters, testing organizations, general users, etc. with impartiality.

It is our sincere hope that the technical report would be widely used by radio equipment manufacturers, testing organizations, general users, etc.

Contents

Chapter 1 General Descriptions

1.1 Objectives

This technical report defines metadata, transport methods and operational guidelines to monitor errors of audio and video signals at arbitrary monitoring points on a broadcasting chain.

1.2 Scope

This technical report applies to monitoring of errors of audio and video signals at arbitrary monitoring points on a broadcasting chain.

1.3 References

1.3.1 Normative

(1) Recommendation ITU-R BT.1364 “Format of ancillary data signals carried in digital component studio interfaces”

(2) Recommendation ITU-R BT.1865”Metadata to monitor errors of SDTV and HDTV signals in the broadcasting chain”

(3) 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”

(4) ARIB BTA S-005C Ver. 1.0 “Ancillary data packet and space formatting of bit-serial digital interface for 1125/60 HDTV systems” (July 2009)

(5) ARIB STD-B6 Ver. 1.2 “Ancillary data packet and space formatting of bit-serial digital interface for 525/60 television system“(July 2009)

(6) ARIB STD-B40 Ver. 1.0 ”PES packet transport mechanism for ancillary data” (July 2002)

(7) ISO 3166-1:2006 “Codes for the representation of names of countries and their subdivisions – Part 1: Country codes”

1.3.2 Informative

(1) Recommendation ITU-R BT.1790 “Requirements for monitoring of broadcasting chains during operation”

(2) ITU-T Recommendation J.243 “Requirements for operational monitoring in television programme transmission chains”

(3) ITU-T Recommendation P.911 “Subjective audiovisual quality assessment methods for multimedia applications”

1.4 Terms and definitions

1.4.1 Abbreviation

bslbf	bit string, left bit first.
uimsb	unsigned integer, most significant bit first.

<Intentionally blank>

Chapter 2 Metadata for monitoring

2.1 Overview of the metadata for monitoring

Metadata describing features of audiovisual signals are transported synchronously with the audiovisual signals and are extracted for monitoring at arbitrary monitoring points on a broadcasting chain. Video and audio features are defined as monitoring information that can be used to detect trouble such as blackouts, freezing, and muting which are likely to occur due to malfunction of equipment and transmission errors. This will contribute to constructing sophisticated and highly accurate monitoring systems, and to automation and labor-saving of monitoring.

A schematic diagram of an operational monitoring process using metadata on a broadcasting chain is shown in Figure 2-1 where operational monitoring is assumed to be conducted at arbitrary monitoring points on a broadcasting chain. The metadata are extracted and inserted, at each monitoring point. Details on the monitoring process are given at the bottom of Figure 2-1. 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 whether there are experiencing any problems, and
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.

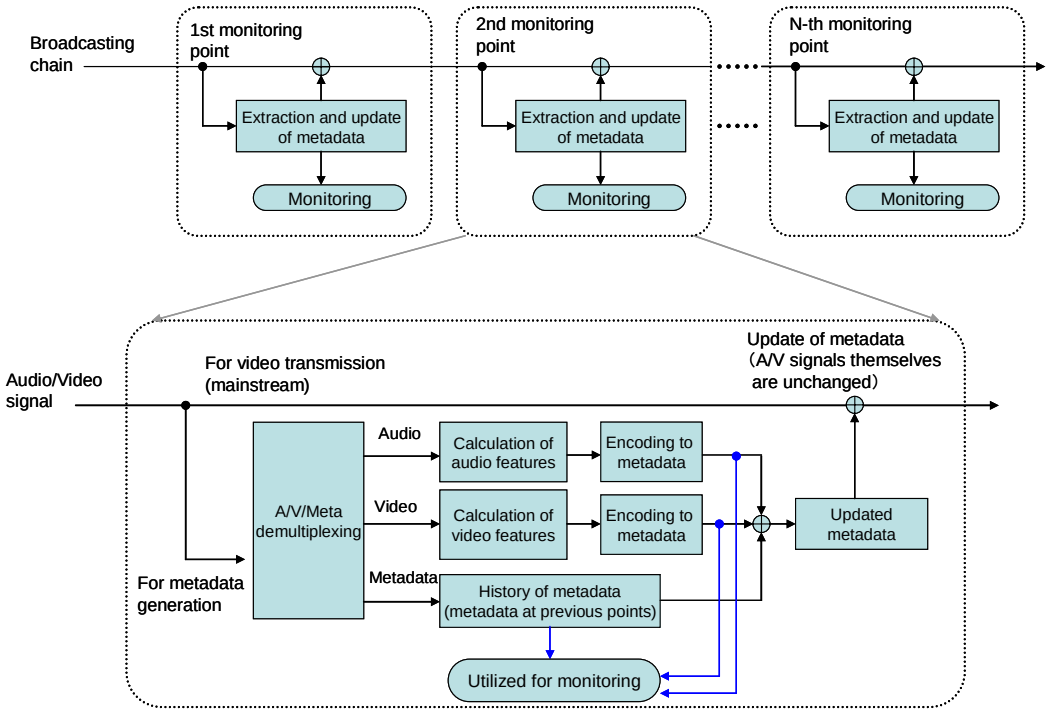


Figure 2-1 Configuration of monitoring points on broadcasting chain and monitoring process for metadata

2.2 Metadata for operational monitoring

2.2.1 Configuration

The basic configuration for the metadata for operational monitoring is outlined in Figure 2-2. The configuration complies with Recommendation ITU-R BT.1865. The metadata type identifier appears first, followed by the metadata at the monitoring point on the upper end and then the metadata at the current monitoring point. Other metadata at the previous monitoring points follow in order of those most recently monitored on the broadcasting chain. The same type of metadata is to be used for the series of metadata in a 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.

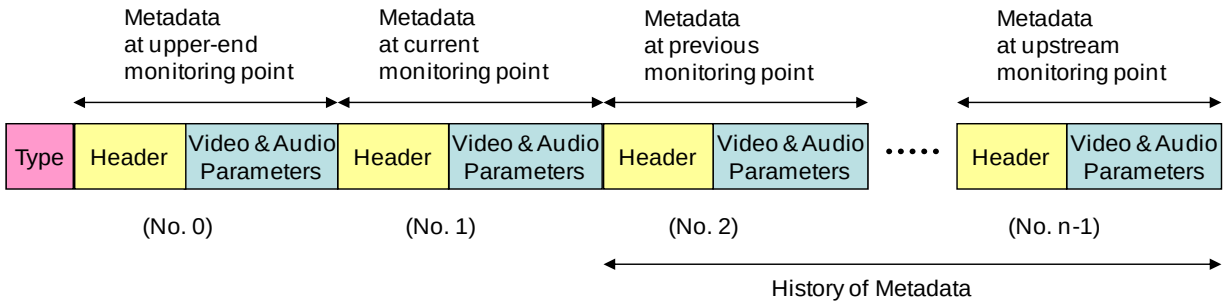


Figure 2-2 Configuration for metadata

The syntactical definitions of the metadata are listed in Table 2-1. The definitions of the header, video parameters, and audio parameters of the metadata are given in the following sections.

Table 2-1 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	Mnemonic
monitoring_metadata{		
header( )	11	
video_parameters( )	10	
audio_parameters( )	21	
}		

metadata_type indicates the type of metadata used and shall be set to 0x01.

2.2.2 Header

Header information precedes the video and audio feature parameters to identify the types of video and audio signals, the location of the monitoring point and the organization that generated the metadata. The syntactical definition of the header is listed in Table 2-2.

Table 2-2 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 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, six sets of metadata, data numbers 0 through 5, can be used at maximum as the history. Figure 2-3 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.

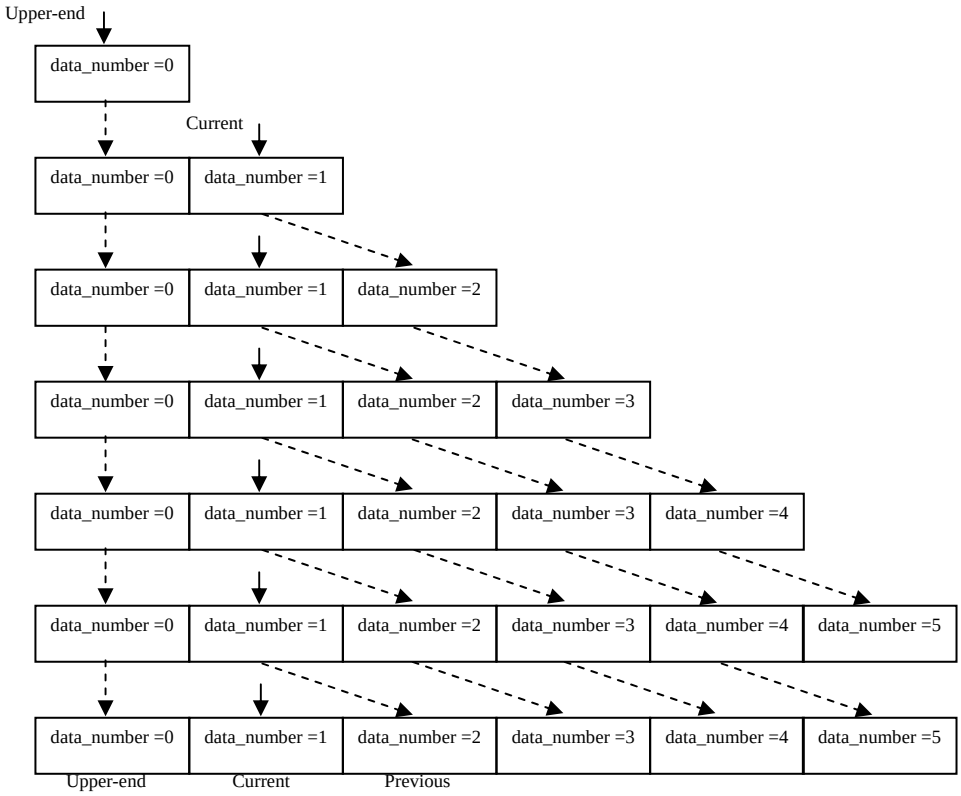


Figure 2-3 Management metadata history and data number

video_signal_type indicates the type of video signal. Composite signals are not 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.

Type	Value
Uncompressed	00
Compressed	01
No audio	10
Reserved	11

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.

2.2.3 Video parameters

Table 2-3 lists the syntactical definition of video parameters.

Table 2-3 Definition of 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 § 2.2.3.1.

y_ti, **cb_ti**, and **cr_ti** indicate the video temporal features of Y, Cb, and Cr signals calculated

as per § 2.2.3.2.

2.2.3.1 Video spatial feature

Video spatial feature is the Spatial Perceptual 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:

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

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

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

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

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.

2.2.3.2 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:

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

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.

2.2.4 Audio parameters

Table 2-4 lists the syntactical definition of audio parameters.

Table 2-4 Definition of 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. superimpose 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 § 2.2.4.2.

audio_oi indicates the audio out-phase information (AOI) between two channels in an AES

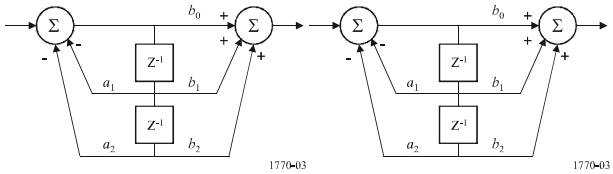
stream calculated as per § 2.4.2.

audio_rms_1 indicates the audio magnitude information (AMI) of audio channel 1 in an AES stream calculated as per § 2.4.3.

audio_rms_2 indicates the audio magnitude information (AMI) of audio channel 2 in an AES stream calculated as per § 2.4.3.

2.2.4.1 Pre-processing

Pre-filter having a cut-off frequency of 20 Hz is applied to audio signals before calculating audio feature information. The pre-filter is defined by the filter in Figure 2-4 with the coefficients specified in Table 2-5. Floating point operation should be used. The response is shown in Figure 2-5.



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

Figure 2-4 Signal flow diagram as 4th order filter

Table 2-5 Filter coefficients for pre-filter

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

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

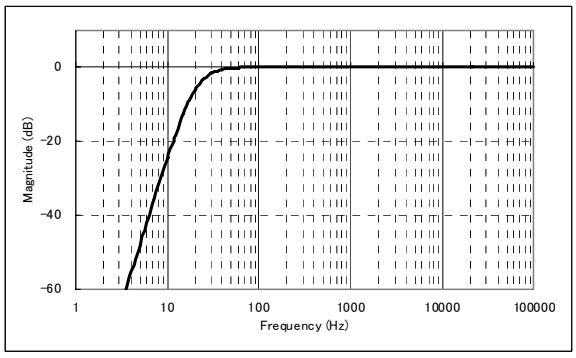


Figure 2-5 Frequency response of pre-filter

2.2.4.2 Audio inter-channel features

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

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

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

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 $2^{10} - 1$.

2.2.4.3 Audio magnitude feature

Audio magnitude information (AMI) is defined as:

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

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 $2^{10} - 1$.

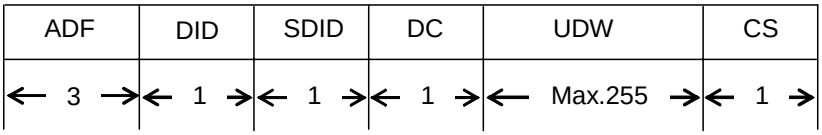
<Intentionally blank>

Chapter 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 (normative reference (1)), ARIB BTA S-005C (normative reference (3)) and ARIB STD-B6 (normative reference (4)), and one word consists of 10 bits in this format. The format for the data packets is given in Figure 3-1.



ADF: Ancillary data flag (see the normative references (1), (3) and (4))
DID: Data identification word. DID of this data packet is set to 0x143.
SDID: Second data identification word. SDID of this data packet is set to 0x104.
DC: Data count word.
UDW: User data word.
CS: Check sum word.
Note: Numbers indicate number of words.

Figure 3-1 Format for ancillary data packets of metadata

3.2 User data word (UDW) format

User data words comprise the metadata defined in Chapter 2. The format for the UDW is listed in Table 3-1. The metadata are to be byte aligned, where the first bit of a byte occupies b7 and the last bit occupies b0.

Table 3-1 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.

3.3 Transport of the ancillary data packets

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

Ancillary data packets containing the metadata are transported by the following methods. (See Appendix 1)

(1) The ancillary data packets are multiplexed into the ancillary data space of the serial digital interface (SDI). (See the normative references (1), (3) and (4)). 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. (See the normative references (2) and (5)).

(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 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, and (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 S1-1 and S1-2 describe typical configurations of programme transmission between master controls. SDI is used in Figure S1-1 and DVB-ASI in Figure S1-2.

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 on 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.

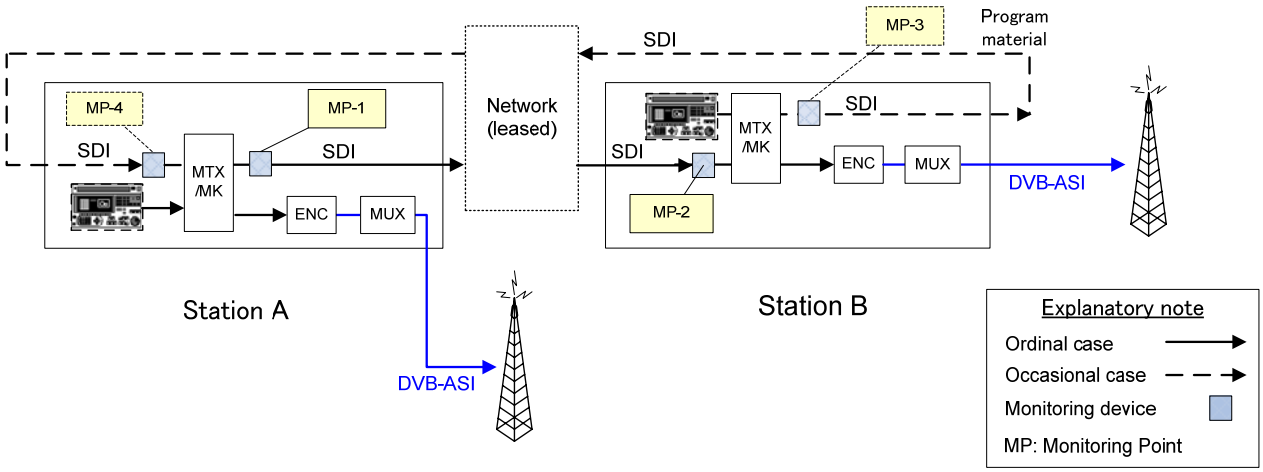


Figure S1-1 Programme transmission between master controls using SDI baseband signals

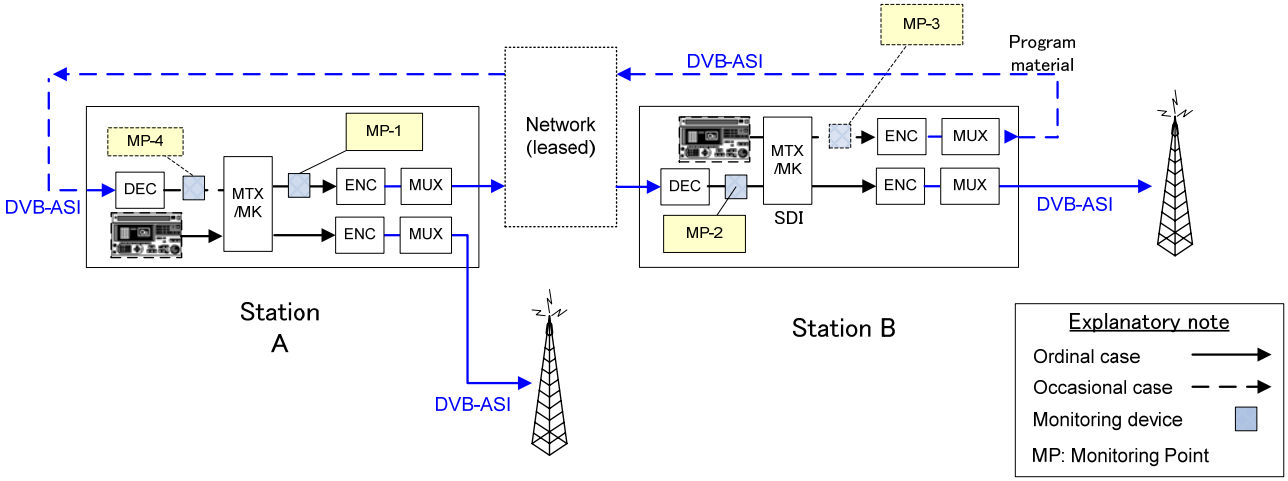


Figure S1-2 Programme transmission between master controls using DVB-ASI compressed signals

2.2 Programme material transmission

Figure S1-3 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.

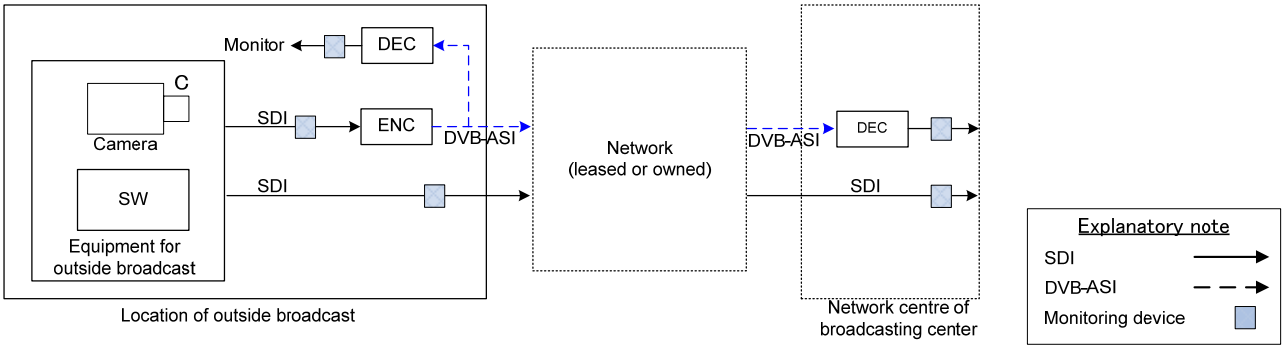


Figure S1-3 Programme material transmission in outside broadcasting

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 Figure S1-4.

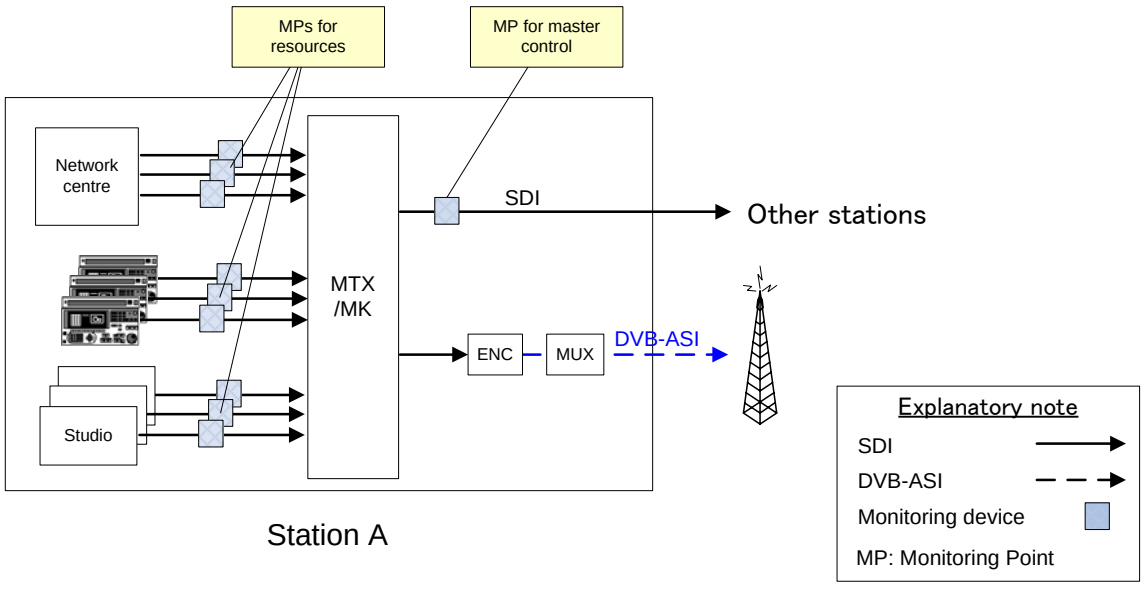


Figure S1-4 Monitoring points at output of broadcast resources

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 S1-5 shows their location in a master control and Figure S1-6 in outside broadcasting.

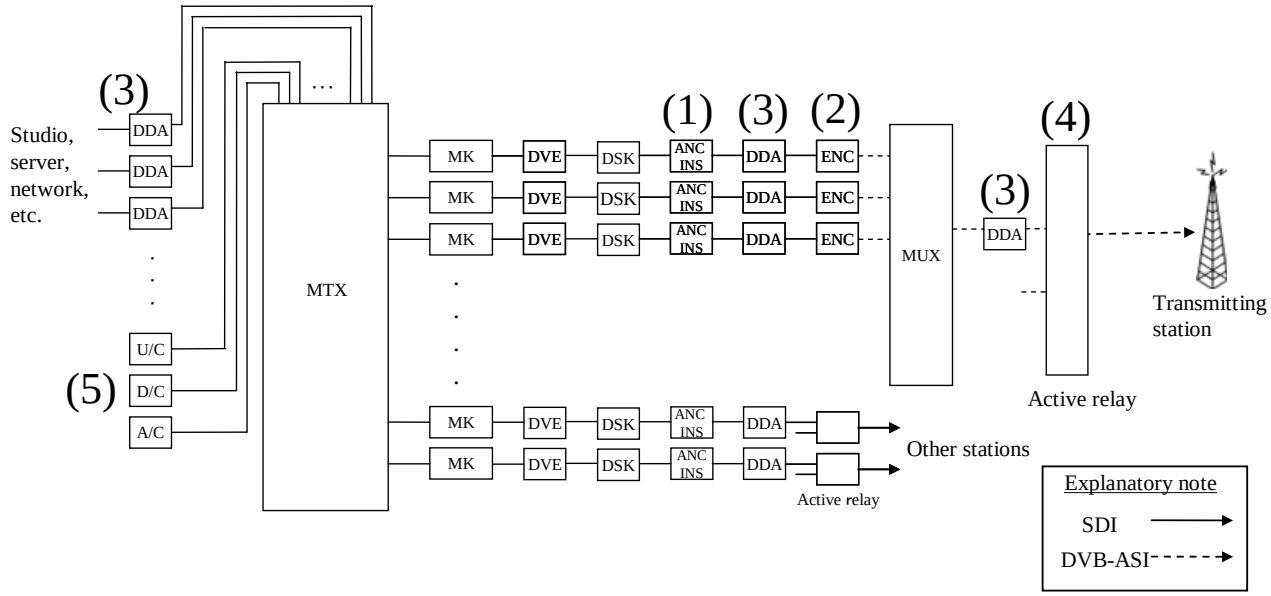


Figure S1-5 Possible devices to install monitoring points in master control

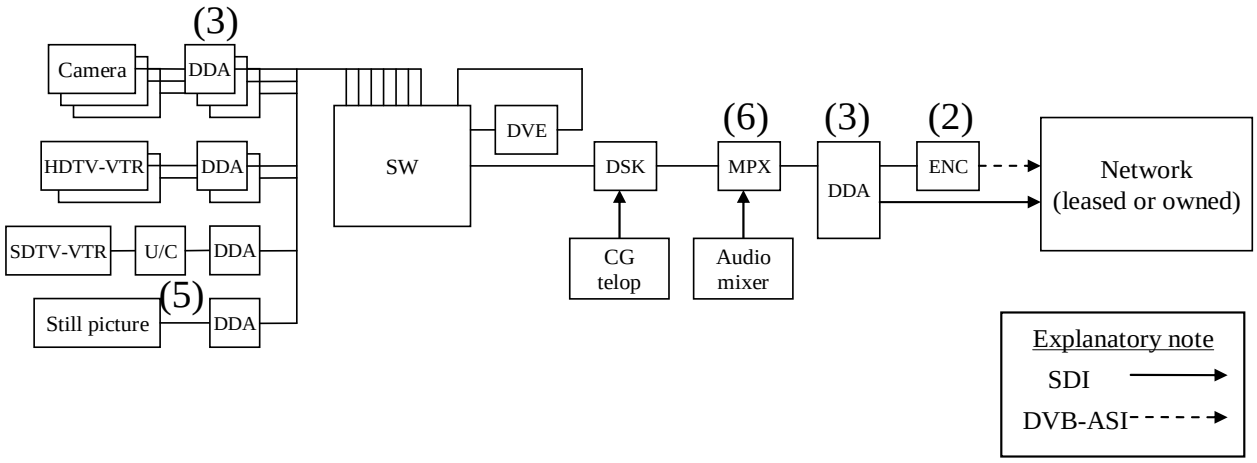


Figure S1-6 Possible devices to install monitoring points in outside broadcasting

(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.

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 Figure S1-7 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.

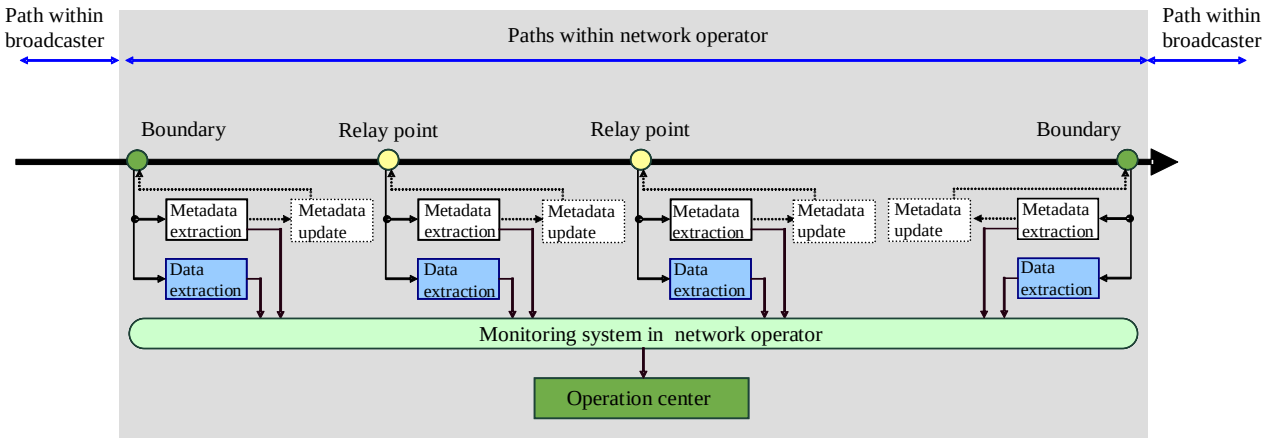


Figure S1-7 Operation example by network operators

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 Figure S1-7 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 metadata

Figures S1-8 and S1-9 illustrate examples of monitoring assuming possible cases where a programme is delivered between broadcasting stations using DVB-ASI compressed signals.

Figure S1-8 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.

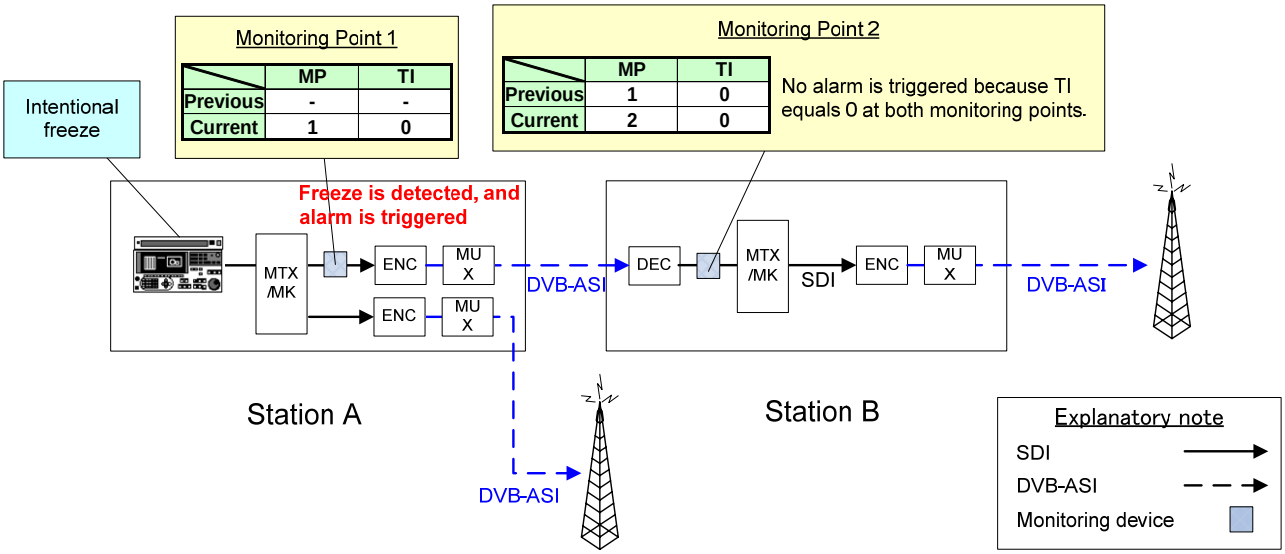


Figure S1-8 When intentional "freezing" is inserted

Figure S1-9 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.

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, and (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.

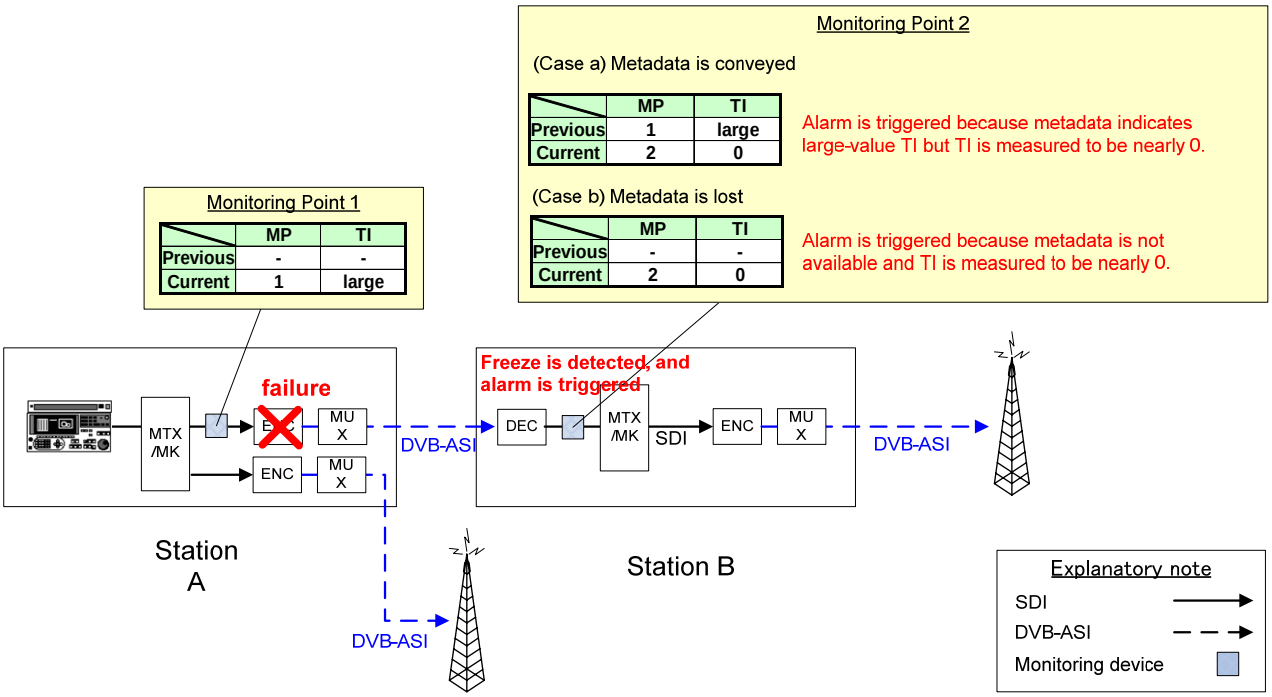


Figure S1-9 When failure occurs in transmission path

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 2 Experimental results by measuring video spatial feature and video temporal feature

This Appendix presents the experimental results obtained by measuring the video spatial feature (SI) and video temporal feature (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 (Macross 7)

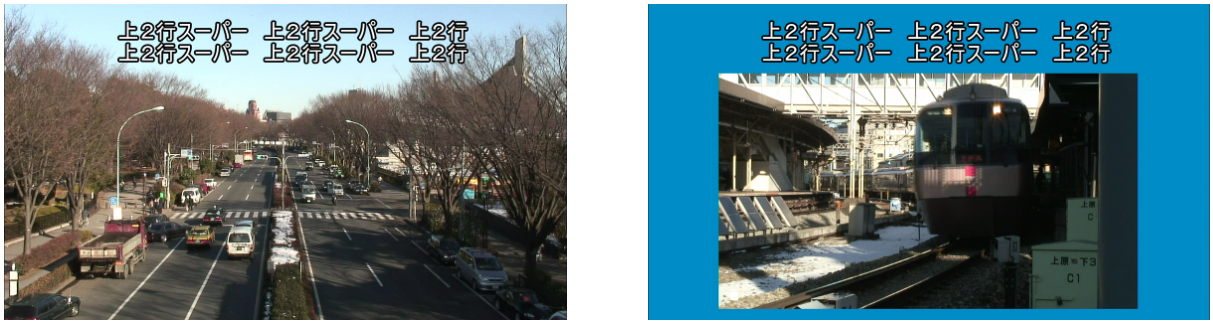
1.9 Superimpose and wipe

1.9.1 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 Figure S2-1 (a) for the image with slight motion.)

1.9.2 Wipe

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 Figure S2-1 (b) for the image with significant motion.)



a) Slight motion　　b) Significant motion

Figure S2-1 Test sequences with superimposed text and wiped with text

2 Measurement results

2.1 Blackout I

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

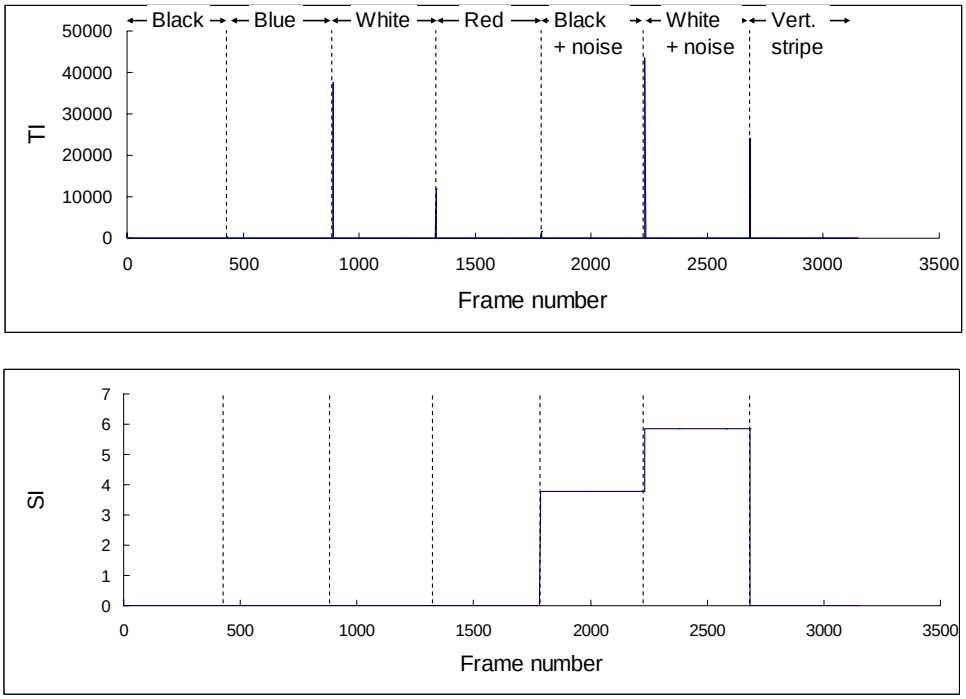


Figure S2-2 Blackout I

2.2 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.

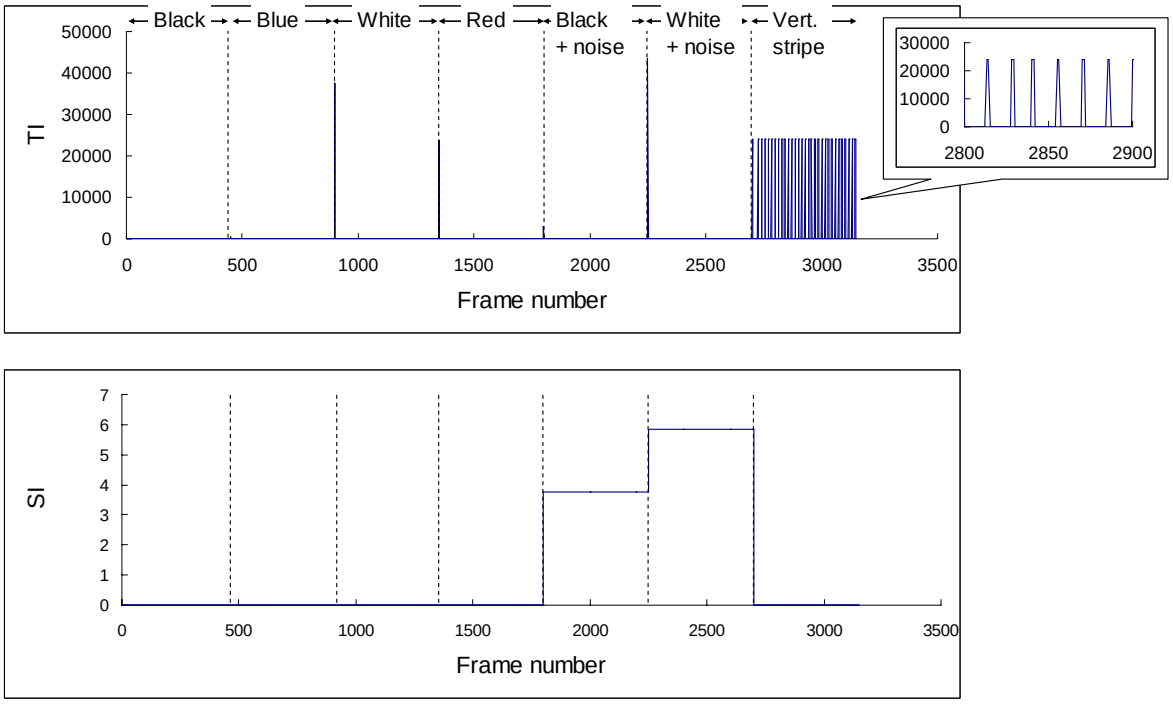
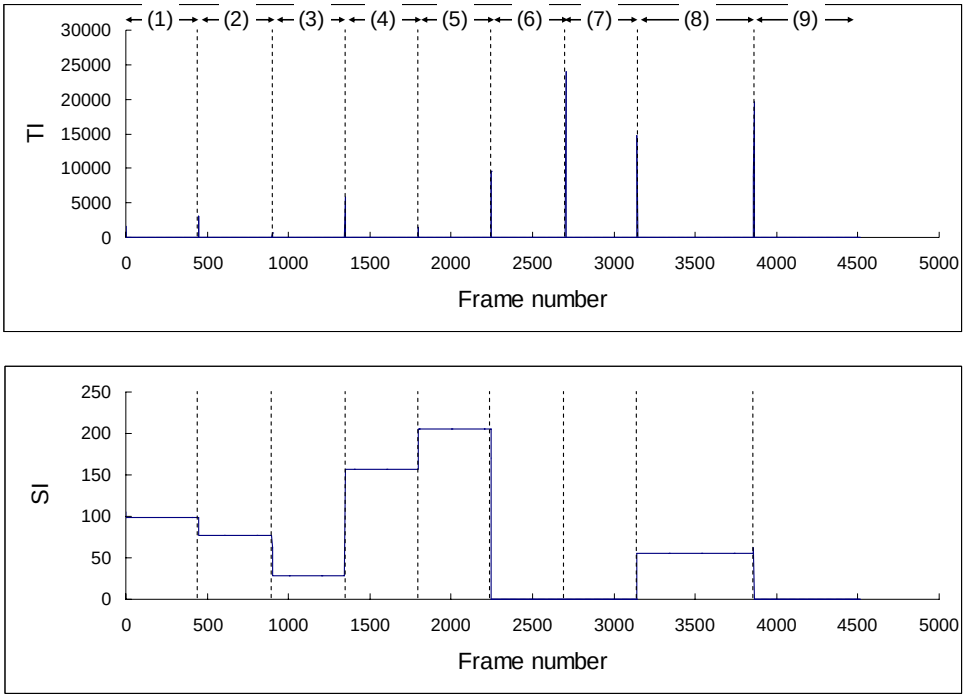


Figure S2-3 Blackout II

2.3 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.

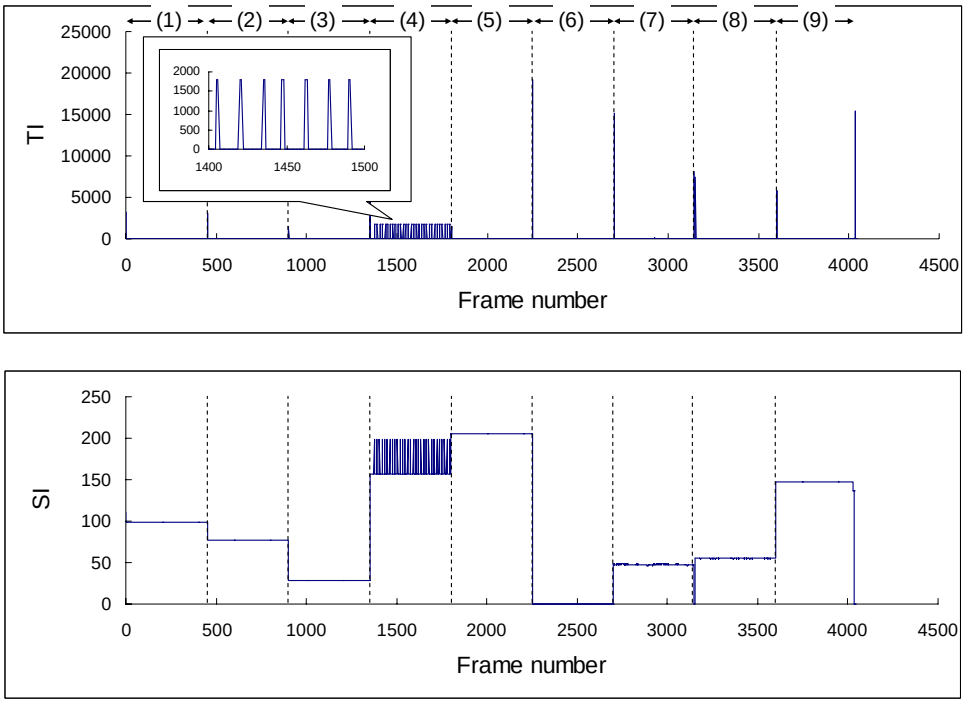


(1) Black with noise, (2) Red with noise 1, (3) Blue with noise 1, (4) Red with noise 2, (5) Blue with noise 2, (6) Grey, (7) Vertical stripe, (8) Flower Basket, (9) Rustling Leaves
(Noise 1: noise is black, Noise 2: noise is white)

Figure S2-4 Freeze I

2.4 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.



(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) Gray, (7) Animation, (8) Flower Basket, (9) Rustling Leaves
(Noise 1: noise is black, Noise 2: noise is white)

Figure S2-5 Freeze II

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.

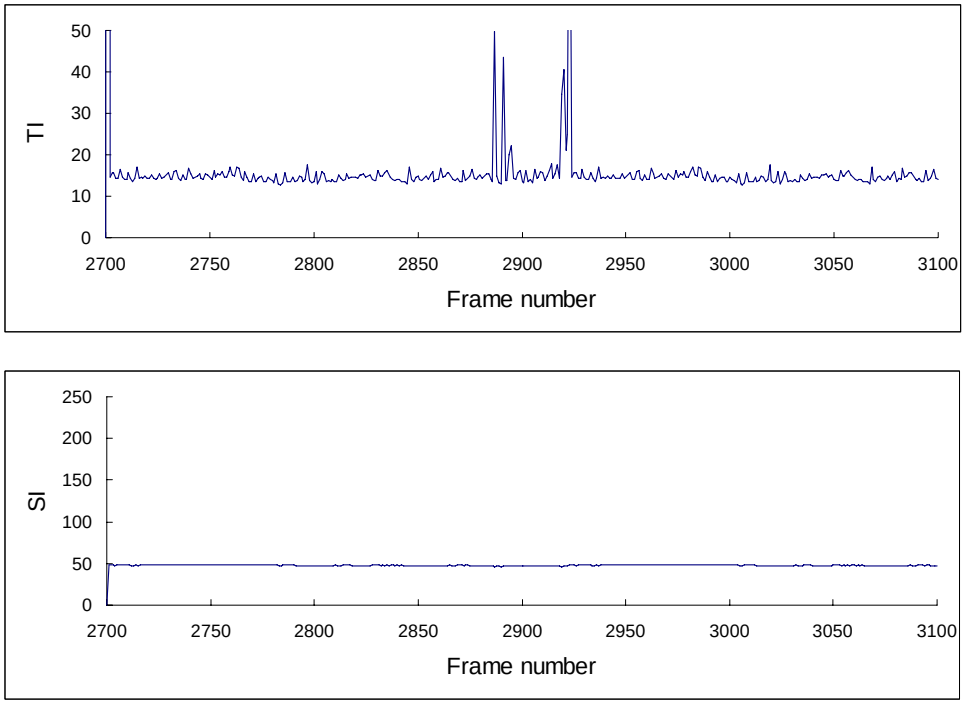


Figure S2-6 Scene 7 in Freeze II

2.5 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 #1150 - 1200), 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.

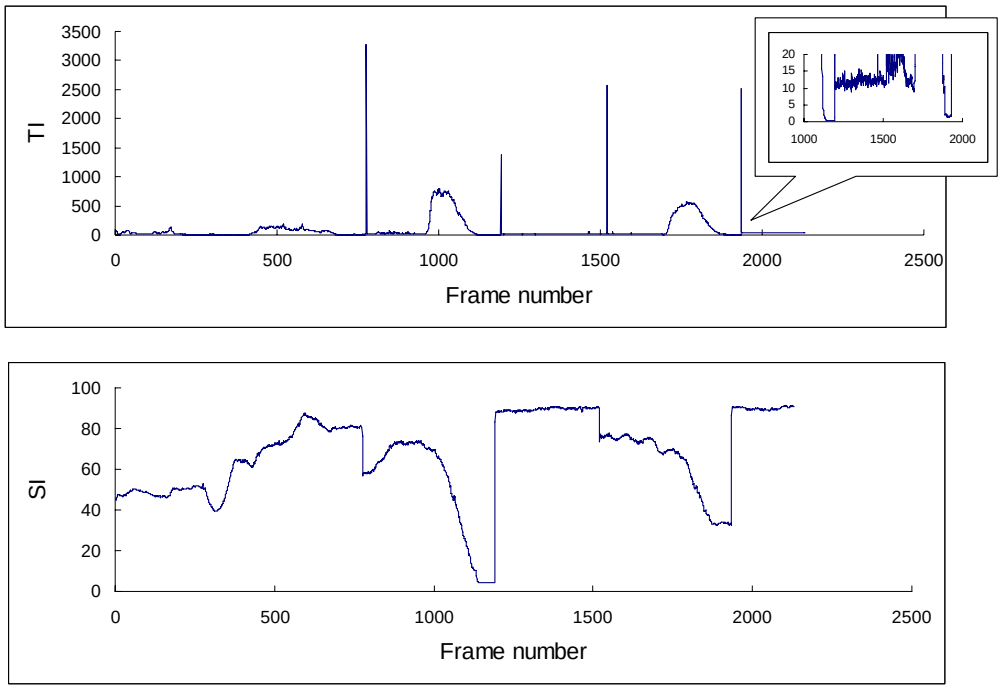


Figure S2-7 Natural image sequence (a summer day)

2.6 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.

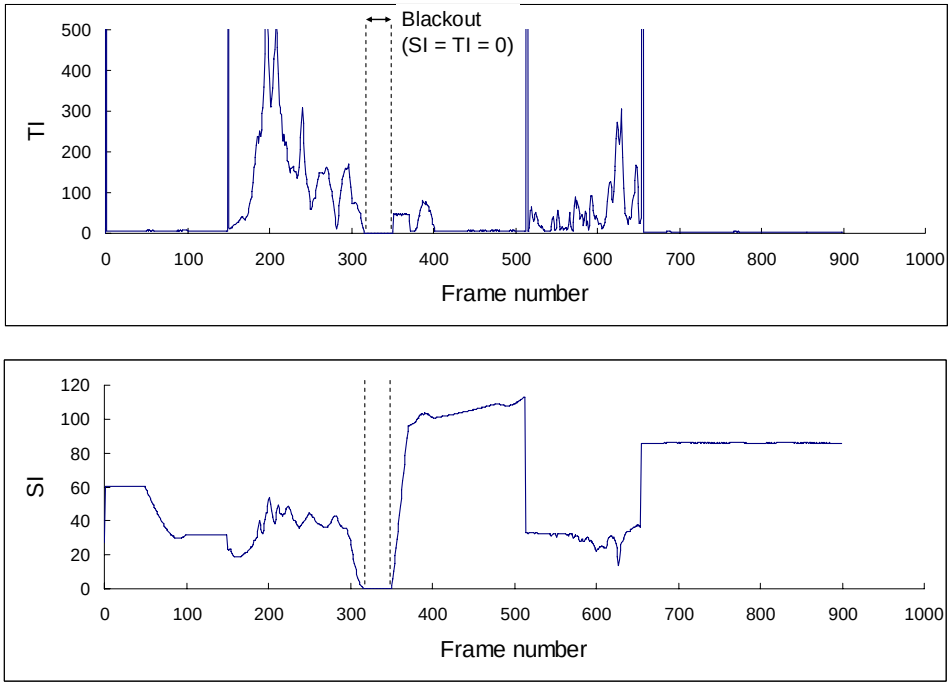


Figure S2-8 Natural image sequence (drama)

2.7 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.

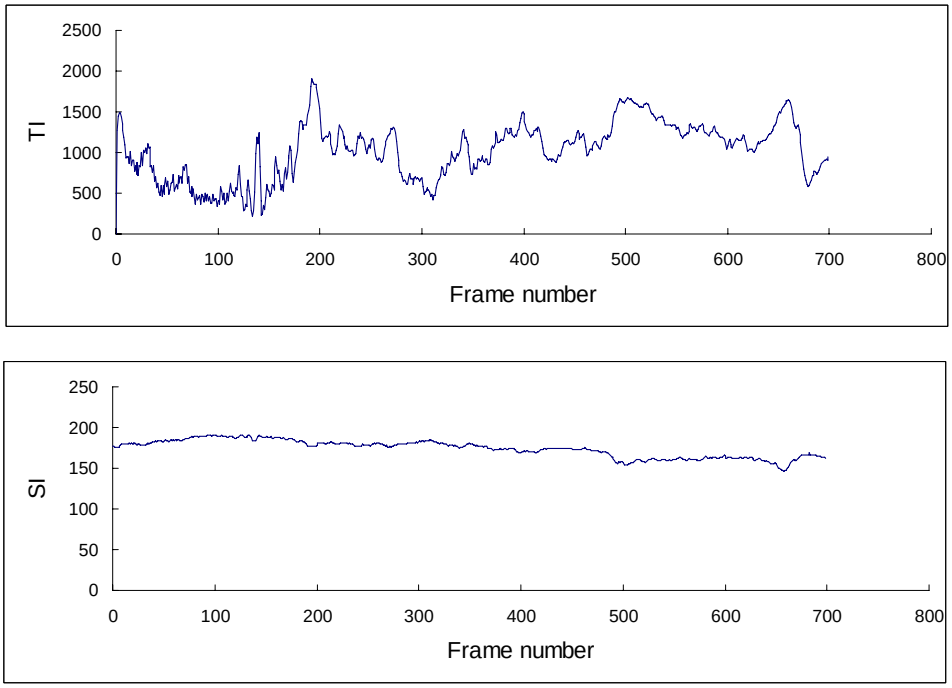


Figure S2-9 Natural image sequence (mobile and calendar)

2.8 Animation (Macross 7)

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.

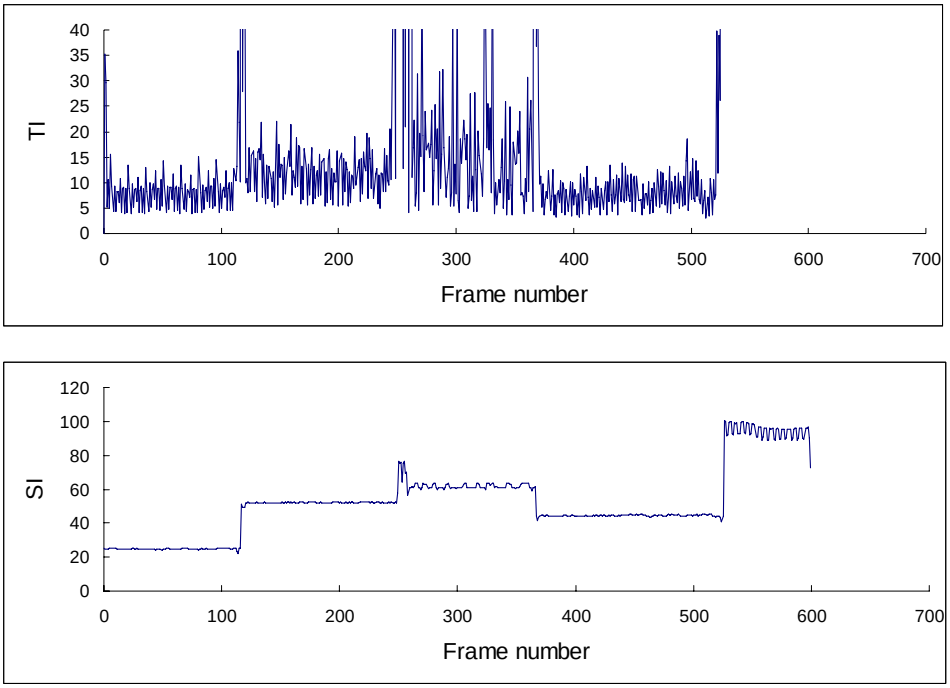


Figure S2-10 Animation (Macross 7)

2.9 Superimpose and wipe

TI and SI are compared for three sequences, without superimpose and wipe, with superimpose, and with wipe, and are shown in Figures S2-11 and S2-12. 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 that that of “without” and the TI of “with superimpose” is the largest.

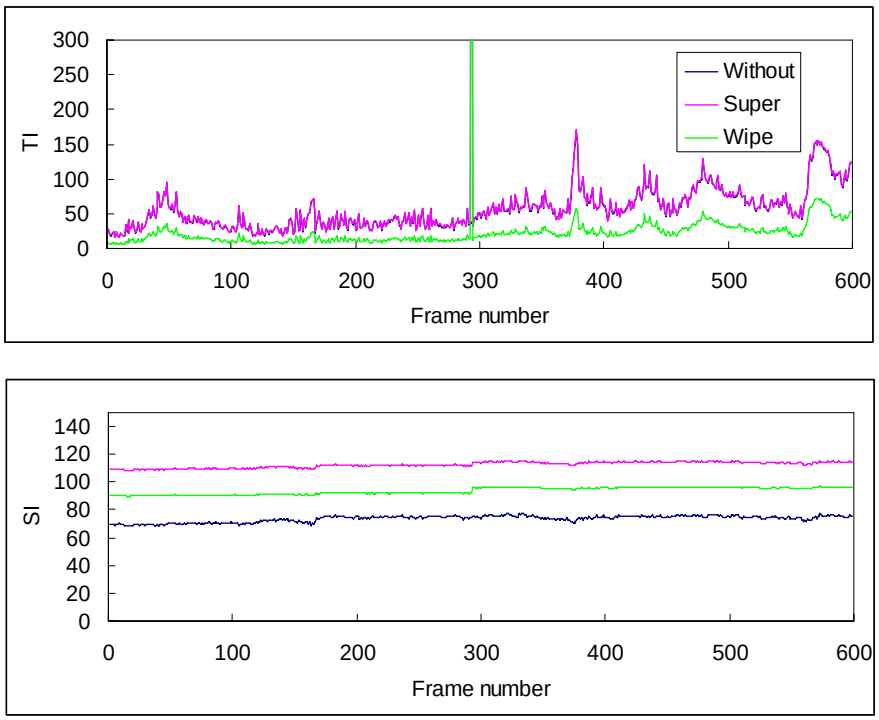


Figure S2-11 Superimpose and wipe for sequence with slight motion

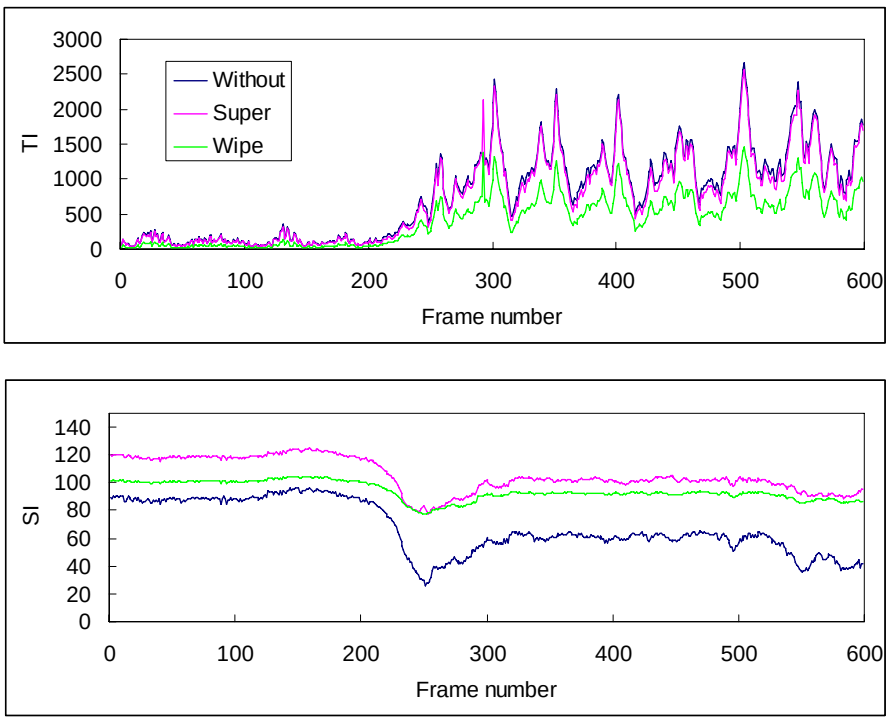


Figure S2-12 Superimpose and wipe for sequence with significant motion

Appendix 3 Experimental results by measuring audio features

This Appendix presents the experimental results obtained by measuring audio features 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.

1 Overview

Figure S3-1 and Table S3-1 show the configuration for the experiment. Figure S3-2 shows how impairments were added to the audio test signals. Table S3-2 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.

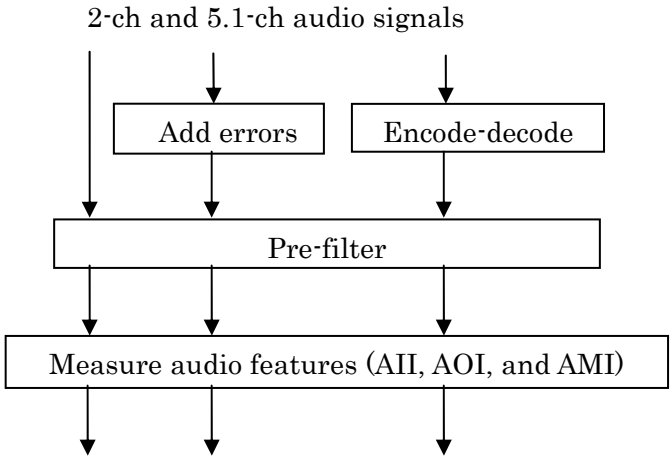


Figure S3-1 Configuration of experiment on audio features

Table S3-1 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 Figure S3-2)		Random noise: First two samples of each frame are replaced by random noise for 4 seconds.
		Mute: First 50 samples and last 50 samples of each frame are replaced by value 0x0000 for 4 seconds.
Encode-decode		AAC, 256kbit/s (2-ch)

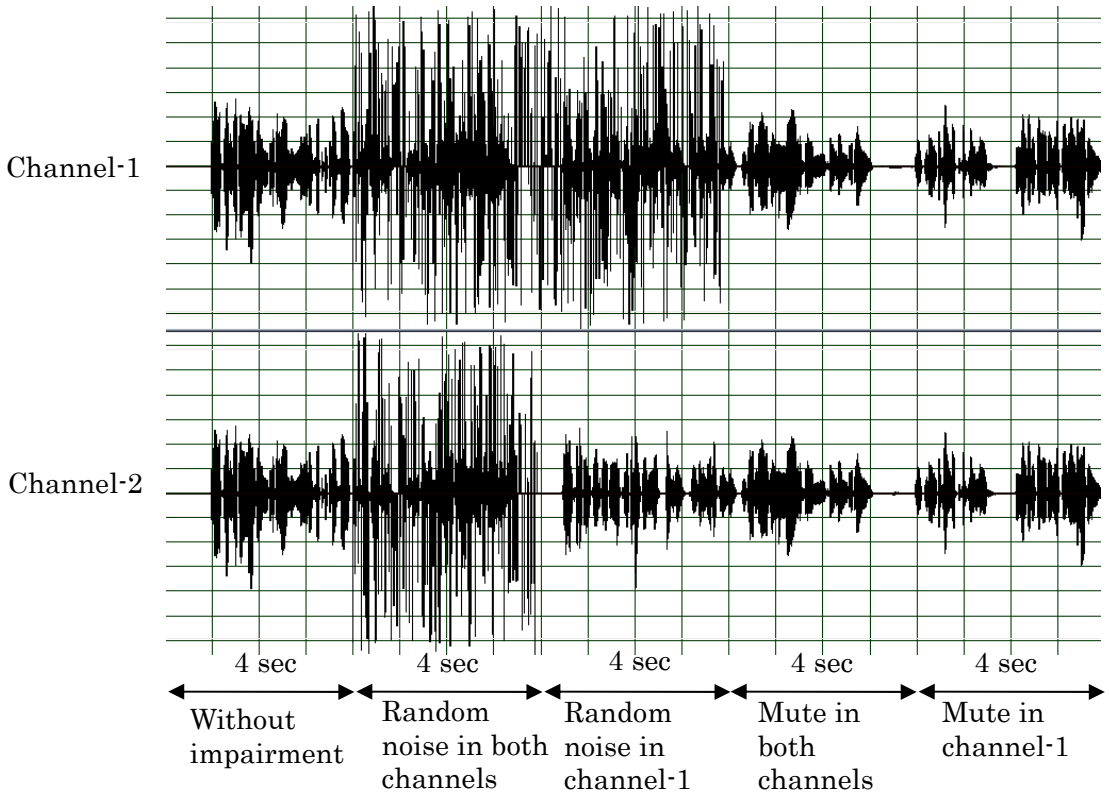


Figure S3-2 Example of impaired signal

Table S3-2 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	S3-3	S3-10	S3-15
AII and AOI (original)	S3-4	S3-11	S3-16
Difference in AII and AOI (original – impaired)	S3-5	S3-12	S3-17
Difference in AII and AOI (original – coded)	S3-6		
AMI-1 and AMI-2 (original)	S3-7	S3-13	S3-18
Difference in AMI-1 and AMI-2 (original – impaired)	S3-8	S3-14	S3-19
Difference in AMI-1 and AMI-2 (original – coded)	S3-9		

2 Results

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

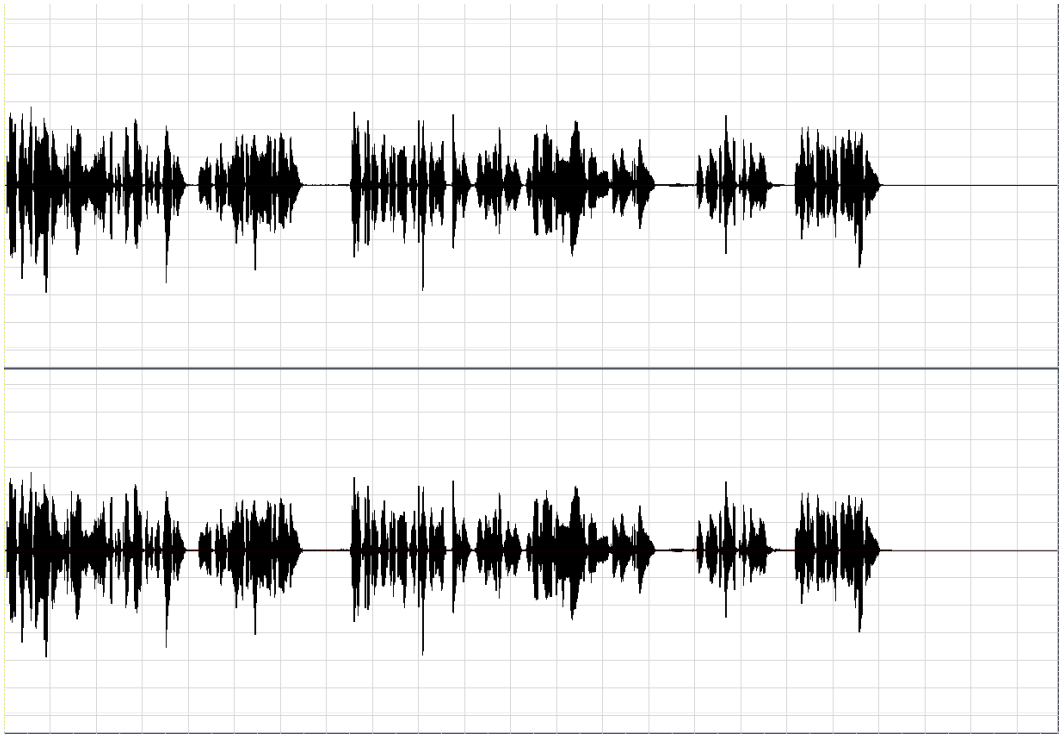
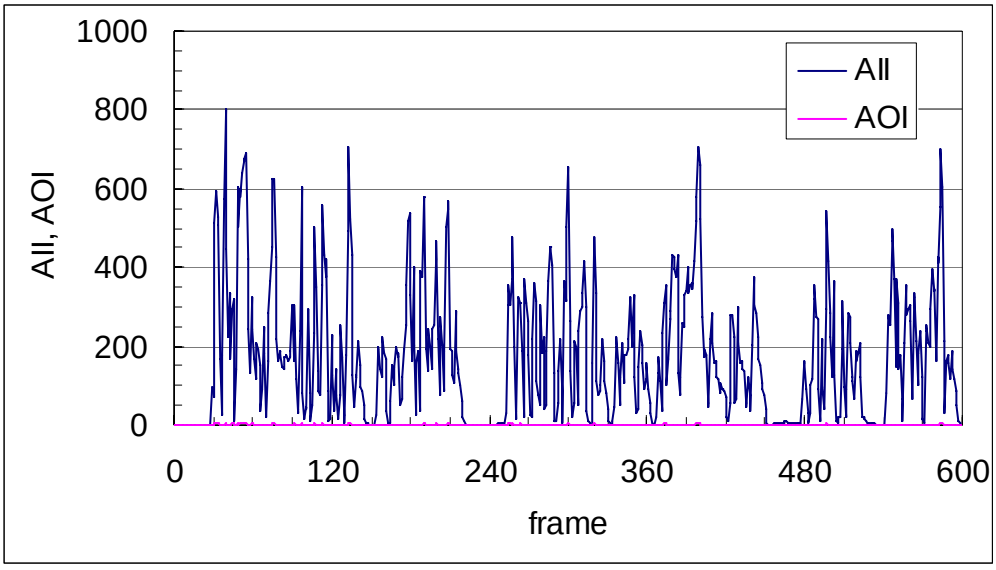


Figure S3-3 Original sound source wave forms



NOTE – AOI was almost zero over the duration.

Figure S3-4 AII and AOI (original)

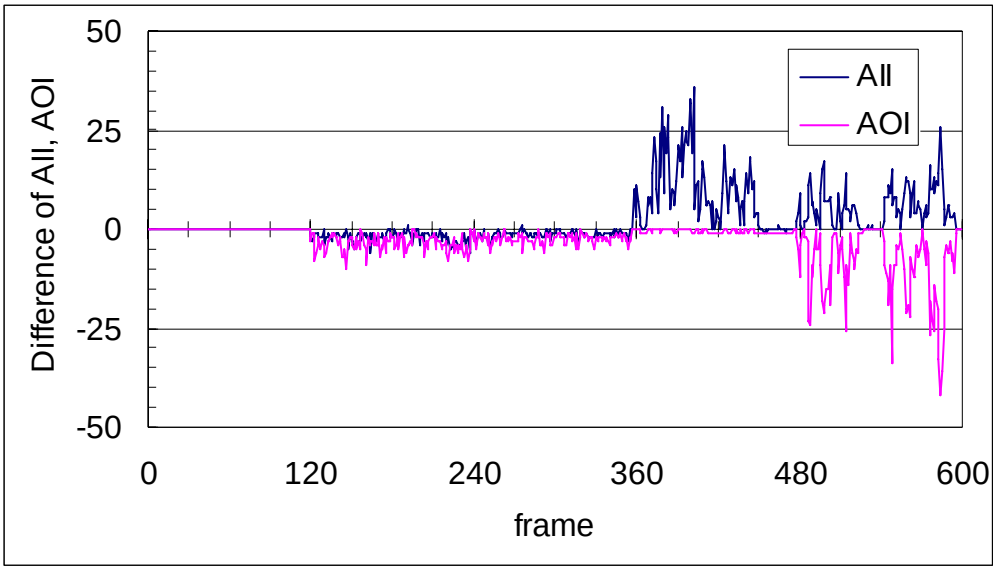


Figure S3-5 Difference in AII and AOI between original and impaired signals

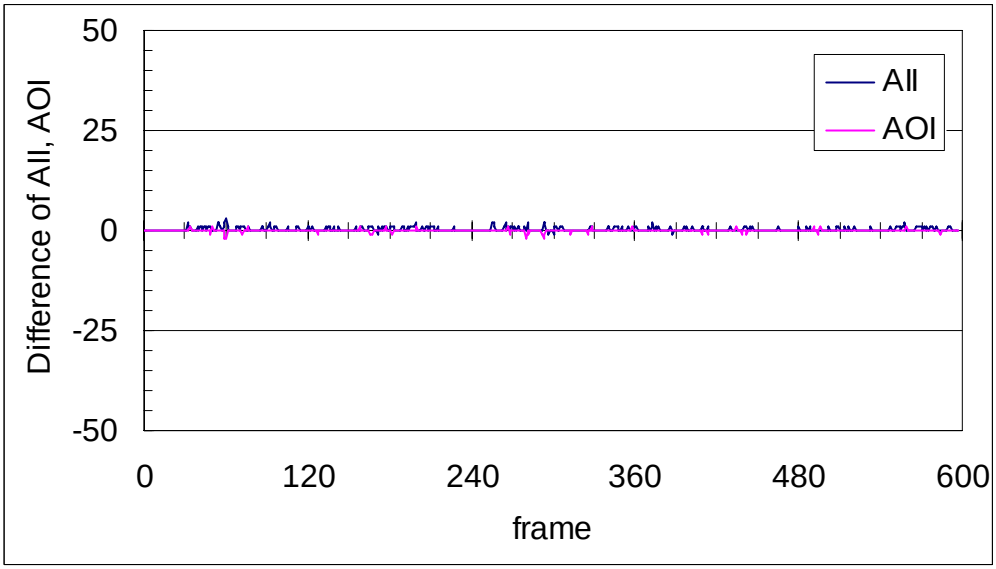
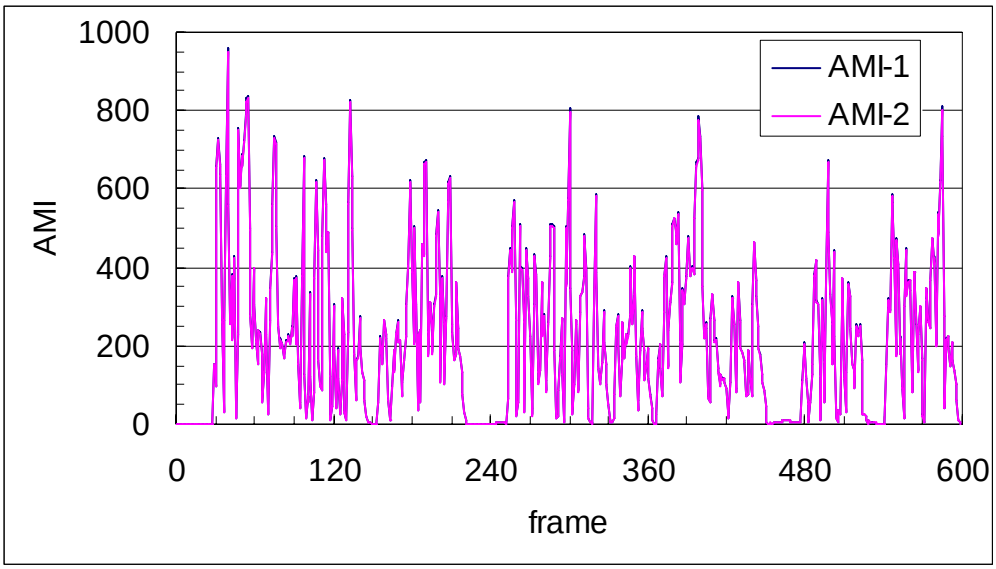


Figure S3-6 Difference in AII and AOI between original and coded signals



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

Figure S3-7 AMI-1 (Left) and AMI-2 (Right) (original)

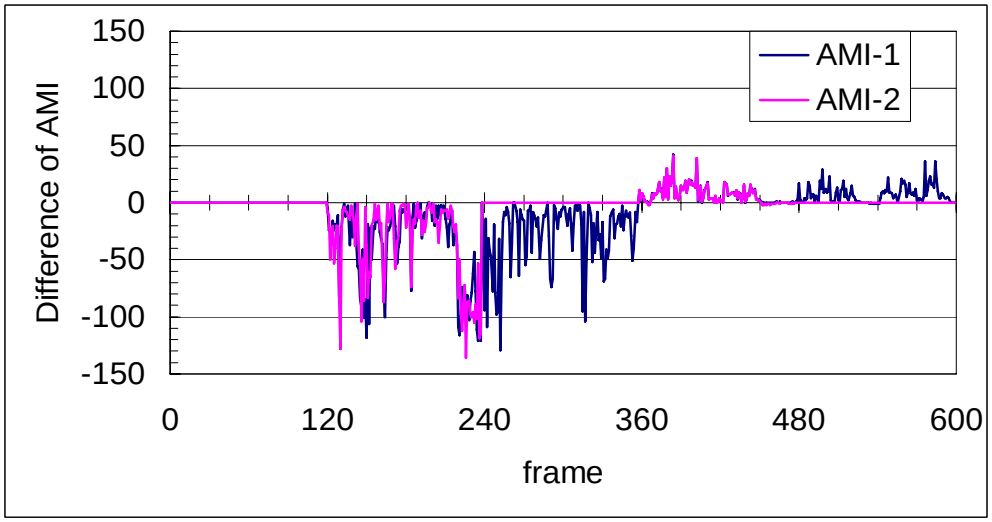


Figure S3-8 Difference in AMI-1 and AMI-2 between original and impaired signals

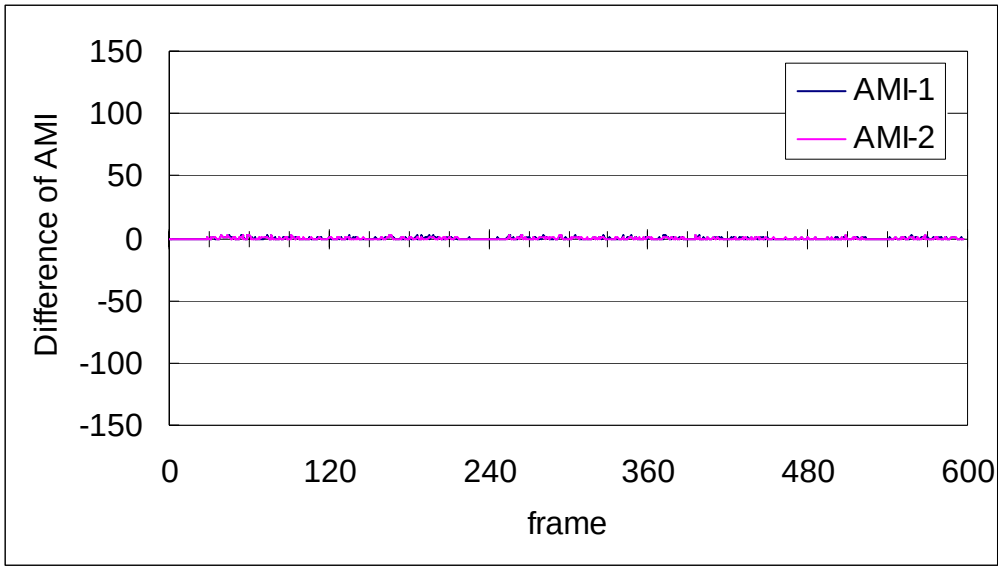


Figure S3-9 Difference in AMI-1 and AMI-2 between original and coded signals

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

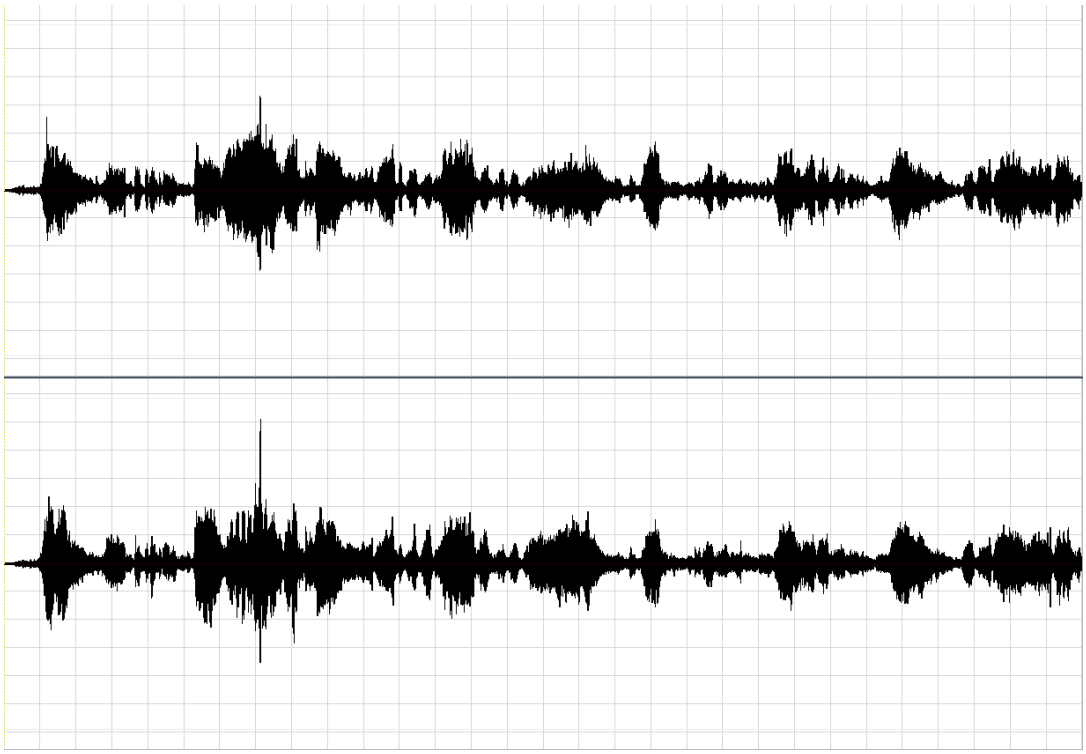


Figure S3-10 Original sound source wave forms

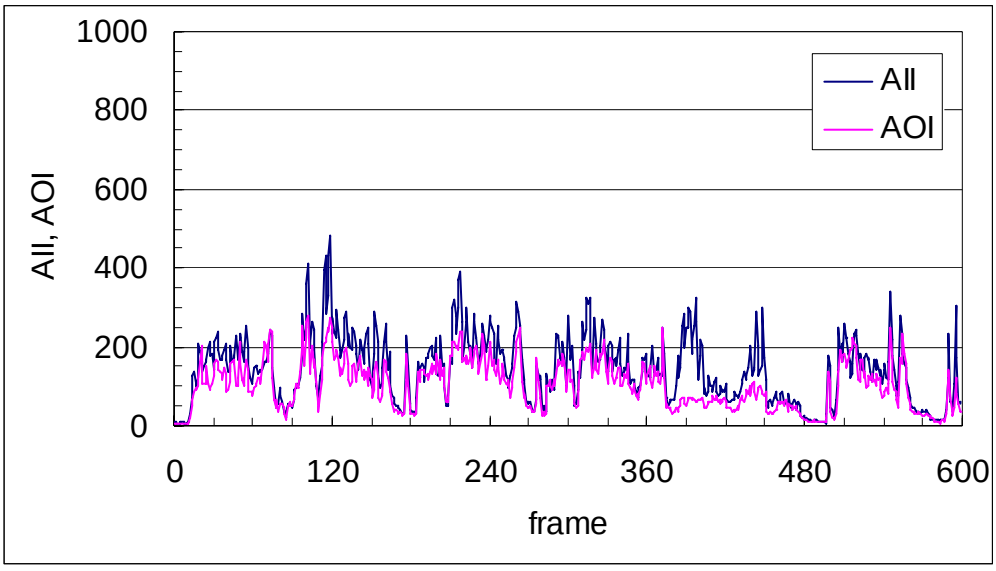


Figure S3-11 AII and AOI (original)

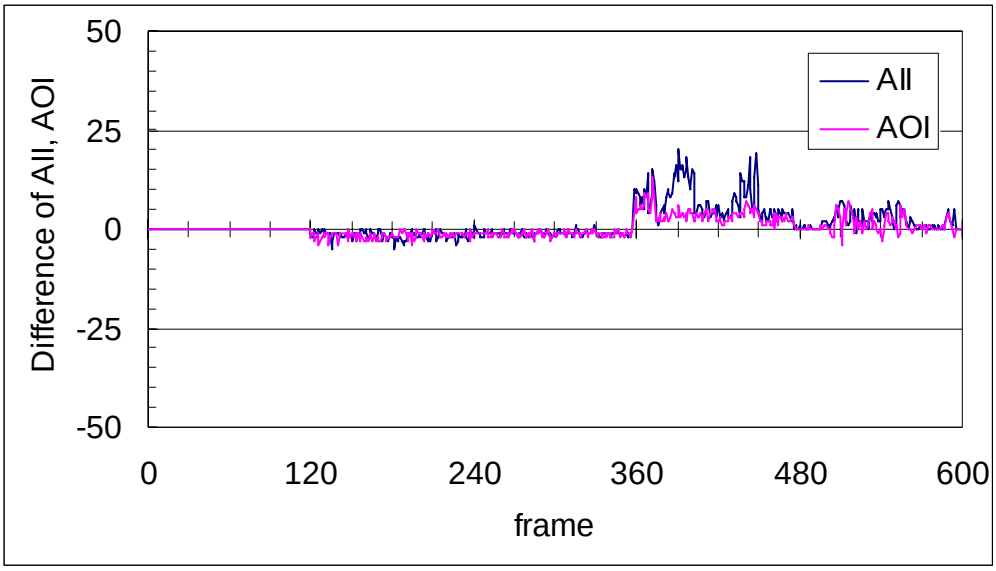


Figure S3-12 Difference in AII and AOI between original and impaired signals

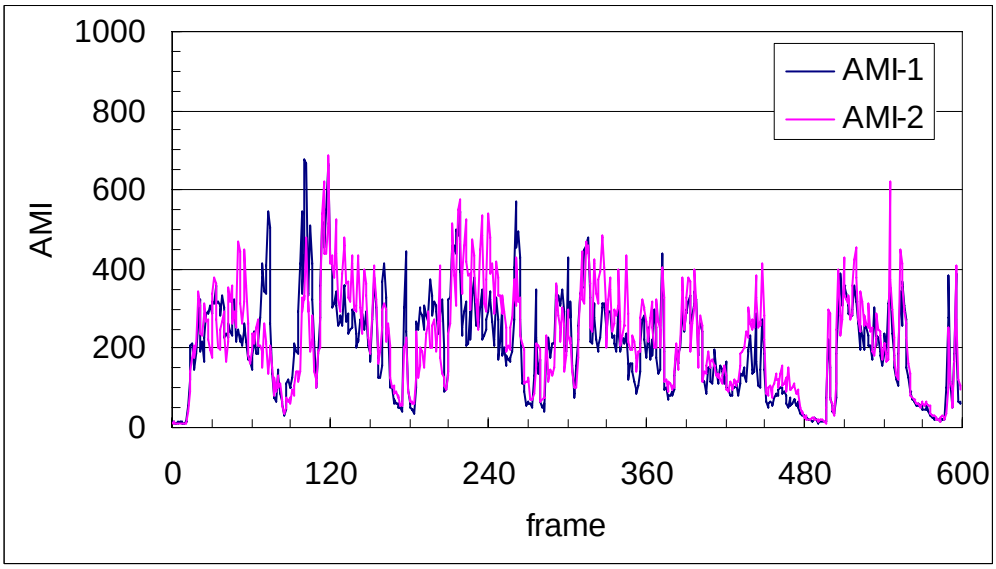


Figure S3-13 AMI-1 and AMI-2 (original)

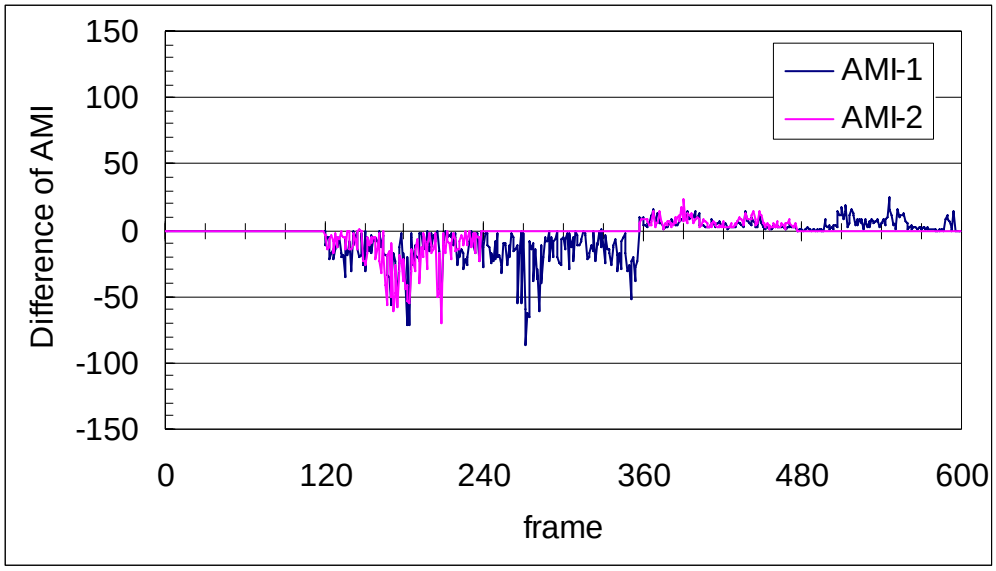


Figure S3-14 Difference in AMI-1 and AMI-2 between original and impaired signals

2.3 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.

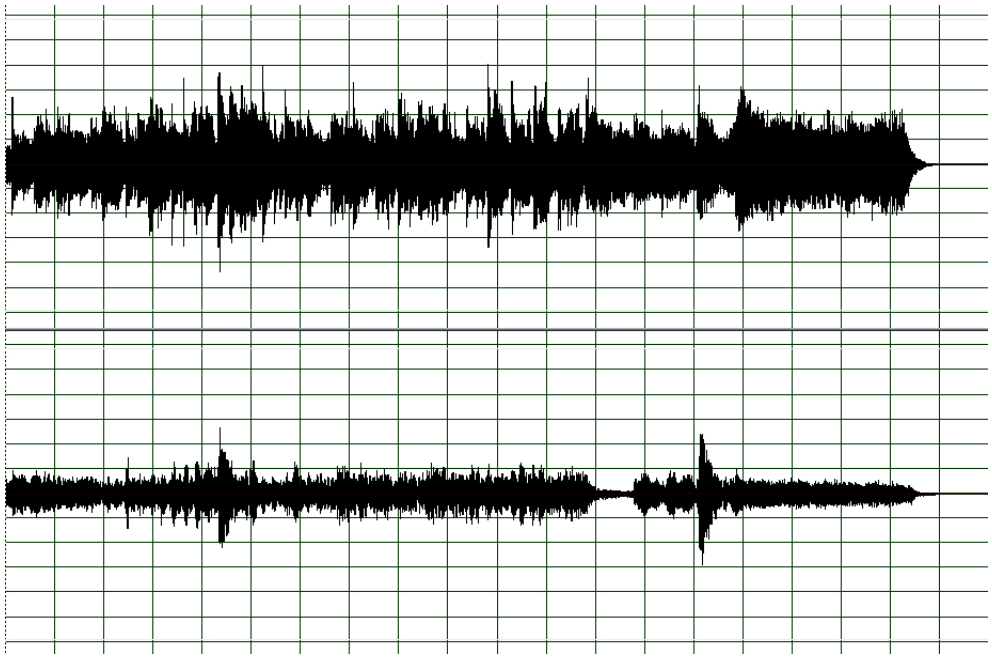


Figure S3-15 Original sound source wave forms

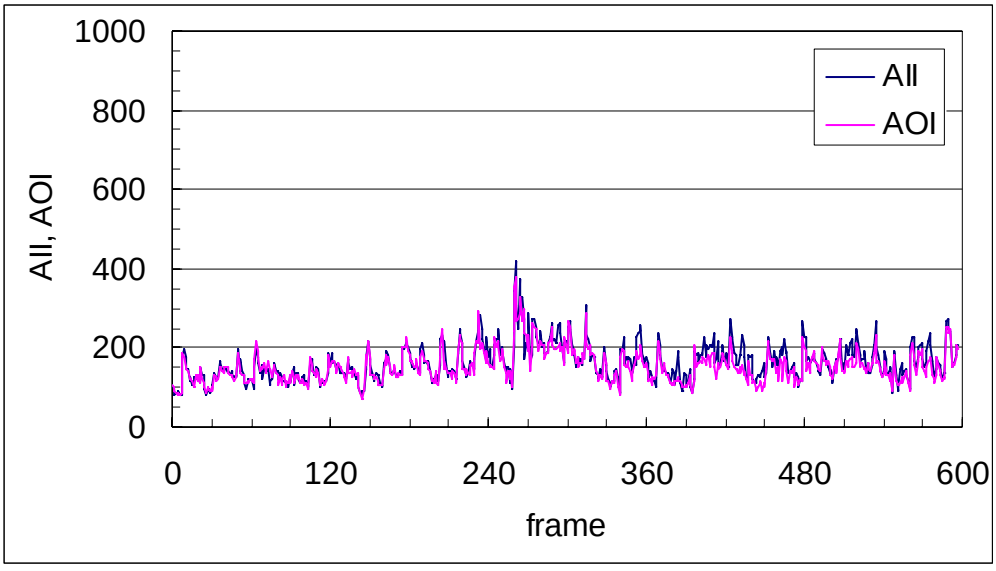


Figure S3-16 AII and AOI (original)

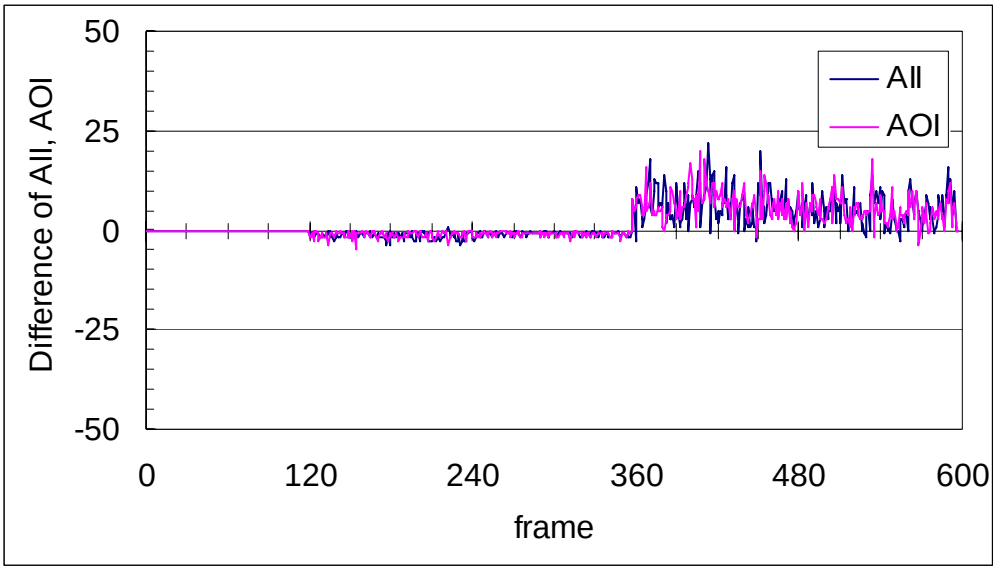


Figure S3-17 Difference in AII and AOI between original and impaired signals

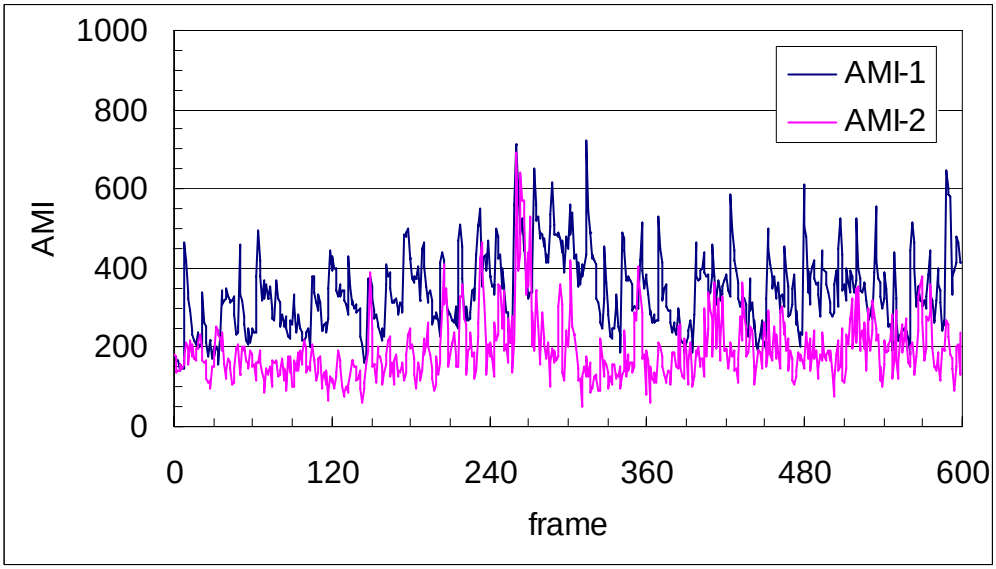


Figure S3-18 AMI-1 and AMI-2 (original)

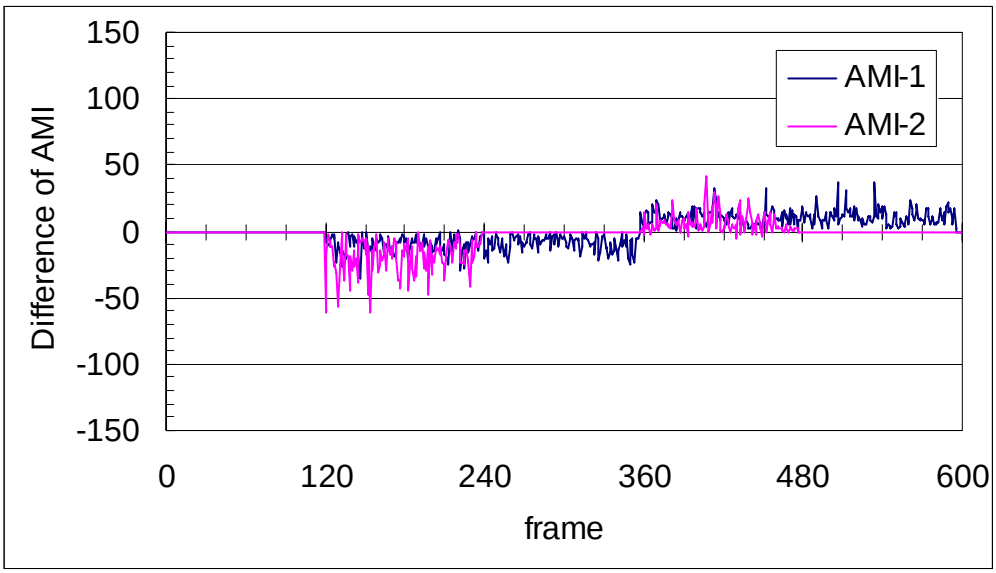


Figure S3-19 Difference in AMI-1 and AMI-2 between original and impaired signals

METADATA TO MONITOR ERRORS OF VIDEO AND AUDIO SIGNALS ON A BROADCASTING CHAIN

ARIB TECHNICAL REPORT

ARIB TR-B29　Version 1.1-E1
(July 15th 2010)

This Document is based on ARIB technical report of "Metadata to Monitor Errors of Video and Audio Signals on a Broadcasting Chain" in Japanese edition and translated into English in August 2010

Published by

Association of Radio Industries and Businesses

Nittochi Building,11F
1-4-1 Kasumigaseki, Chiyoda-ku, Tokyo 100-0013,Japan
TEL　81-3-5510-8590
FAX　81-3-3592-1103

Printed in Japan