

# **AES standard for digital audio - Audio-embedded metadata - Part 1: General**

Published by  
**Audio Engineering Society, Inc.**  
Copyright © 2012 by the Audio Engineering Society

## **Abstract**

AES41 provides for the carriage of audio metadata by embedding it in the audio samples themselves. This tightly associates the metadata with the audio, yet makes it fragile so that changes to the audio will invalidate the metadata. Several metadata sets have been defined, covering applications such as cascaded compression (bit-rate reduction), and loudness control.

This Part specifies the method by which a metadata set is embedded in the audio samples. Other parts define the content and format of the metadata sets.

An AES standard implies a consensus of those directly and materially affected by its scope and provisions and is intended as a guide to aid the manufacturer, the consumer, and the general public. The existence of an AES standard does not in any respect preclude anyone, whether or not he or she has approved the document, from manufacturing, marketing, purchasing, or using products, processes, or procedures not in agreement with the standard. Prior to approval, all parties were provided opportunities to comment or object to any provision. Attention is drawn to the possibility that some of the elements of this AES standard may be the subject of patent rights other than those identified herein. AES shall not be held responsible for identifying any or all such patent rights. Approval does not assume any liability to any patent owner, nor does it assume any obligation whatever to parties adopting the standards document. Recipients of this document are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation. This document is subject to periodic review and users are cautioned to obtain the latest edition.

Document preview:  
for full document, go to  
[www.aes.org/publications/standards](http://www.aes.org/publications/standards)

## Contents

<b>0 Introduction</b> .....	<b>5</b>
0.1 Rationale for this standard .....	5
0.2 Conventions used in this standard .....	5
0.3 Patents .....	5
<b>1 Scope</b> .....	<b>6</b>
<b>2 Normative references</b> .....	<b>6</b>
<b>3 Definitions, symbols, and abbreviations</b> .....	<b>6</b>
3.1 Definitions .....	6
3.2 Symbols .....	6
3.3 Abbreviations .....	7
<b>4 Method for describing the bit-stream syntax</b> .....	<b>8</b>
<b>5 Specification of the audio coder control data bit-stream syntax</b> .....	<b>9</b>
<b>6 Semantics of the bit-stream syntax</b> .....	<b>10</b>
6.1 Header .....	10
6.2 AES41 metadata type registry .....	10
6.3 Error_check .....	11
6.4 Ancillary_data .....	11
<b>7 Method of signaling the audio coder control data bit stream on the AES3 interface</b> .....	<b>11</b>
<b>Annex A (informative) AES41 metadata types</b> .....	<b>13</b>
<b>Annex B (Informative) Informative References</b> .....	<b>14</b>

## Foreword

[This foreword is not a part of *AES standard for digital audio — Recoding data set for audio bit-rate reduction*, AES41-2000.]

This document describes a detailed specification of a set of coding-decision data for MPEG Layer II with a generic transport mechanism developed as project AES-X78 in the AESSC SC-02-02 Working Group on Digital Input-Output Interfacing of the SC-02 Subcommittee on Digital Audio. The task group was lead by A. Mason. The intent has been to produce a standard in cooperation with other bodies. Individual experts associated with potential end-users, equipment developers, and manufacturers who are directly and materially affected by this project were encouraged to participate in the process of standardization by joining SC-02-02.

In the near future, digital compression techniques are expected to dominate the broadcast television environment. Digital compression will be used for a significant amount of program acquisition, for program production and storage, and for distribution and broadcasting.

Compression systems add impairments to sound. The audibility of these impairments depends on the degree of compression and the techniques used. It has been proved that re-encoding with minimal further impairment can be achieved by using information describing the previous encoding process. The information that is preserved varies according to the encoding scheme that has been used. For example, for transparent cascaded coding of MPEG-1 Layer II or MPEG-2 Layer II audio, the information required by the downstream encoder includes:

- . the positions of the frame boundaries;
- . the bit rate of the compressed bit stream before decoding;
- . the coding mode (monophonic, joint stereo, and so on);
- . the bit allocation for each sub-band within a frame;
- . the scale factors for each sub-band within each sub-block of a frame.

In order to help with the process of editing decoded audio, it is useful to send additional coding-decision data to describe any timing offset that may be introduced by editing only at points corresponding to block boundaries (for example, 24-ms boundaries for MPEG-1 Layer II at 48 kHz). These data are particularly relevant for editing sound associated with co-timed video. The bit rate of these additional data is less than 30 kbit/s. A cyclic-redundancy check word must be appended to frames of data to provide a means to detect errors in the data.

This data signal can be used by a downstream encoder to guide its coding decisions. For example, if the audio signal is changed in any way (as it would be by mixing it with another signal) then it could be inappropriate for the downstream encoder to re-use any previous coding decisions. However, in such a case, when a suitable transport mechanism is used, the coding-decision data would be corrupted so the coder could detect that the data are no longer valid.

This specification describes in detail the content of the coding-decision information, such as MPEG-header information, bit allocations, and scale factors, plus its representation in terms of the number, order, and meaning of its bits. A mechanism by which these data can be conveyed in a pulse-code-modulation audio signal is also described. The mechanism modifies one of the least significant bits of the audio word to signal the data. The precise time relationship of the data to the audio is also defined.

R. Finger, chair, SC-02-02  
J. Dunn, vice-chair, SC-02-02  
2000-04-08

**Foreword to the second edition, 2009**

As predicted in the foreword to this document in 2000, digital compression techniques now dominate the broadcast television environment. In addition to the problems foreseen relating to cascaded compression, new problems have arisen because of the use of loudness control and surround sound with those digital compression techniques.

Metadata within the compressed audio bit-stream is used to control loudness and the mixing down of multi-channel surround sound to two-channel stereo. These metadata are usually known by terms such as "dialnorm", "prog\_ref\_level", and "down-mix coefficients".

Whilst this might seem unrelated to the original scope of AES41, dealing with bit allocations and scale factors, it is simply another form of data that can affect a later encoding of the audio: this time it is more macroscopic than microscopic.

The metadata is lost when the bit-stream is uncompressed unless provision is made to transport it or store it somewhere. Existing methods rely on non-audio mechanisms to convey the metadata alongside the audio, for example a serial data link like RS-422 and serial digital video SMPTE 259M, or a "chunk" in an audio file (for metadata that does not change).

This revision extends AES41 to include data formats for carrying this loudness and down-mix metadata with the uncompressed PCM using the same transport mechanism as before. The metadata can therefore be carried in the audio to which it relates.

J. Grant  
Chair, SC-02-02 Working Group on digital input/output interfacing  
2009-12

**Foreword to the third edition, 2012**

The proposal to add a further data type to this standard led the working group to conclude that this should be done in conjunction with a partitioning of the standard. The resulting multi-part structure provides a clear separation between the specification of the mechanism by which data is carried and the specification of the semantics of the data that is carried. This simplifies the use of the standard since a reader might not need all the parts, and it simplifies the addition of further data types in the future.

Please note that clause numbers in this edition of the standard will, in general, no longer correspond to clause numbers in the first or second editions.

The draft of this document was developed by a writing group whose primary author was Andrew Mason.

J. Grant  
Chair, SC-02-02 Working Group on digital input/output interfacing  
2012-03

**Note on normative language**

In AES standards documents, sentences containing the word "shall" are requirements for compliance with the document. Sentences containing the verb "should" are strong suggestions (recommendations). Sentences giving permission use the verb "may". Sentences expressing a possibility use the verb "can".

Document preview:  
for full document, go to  
[www.aes.org/publications/standards](http://www.aes.org/publications/standards)

# AES standard for digital audio - Audio-embedded metadata - Part 1: General

## 0 Introduction

### 0.1 Rationale for this standard

It has become apparent that there are requirements for the transport of audio metadata together with the related audio, in a way such that the two are not separated. Further, there are requirements where any alteration of the audio should render the related metadata invalid. Metadata held in files, or transported in associated video signals, or in AES3 user bits, for example, does not meet these requirements: The metadata may become dissociated from the audio, or may continue to be associated with the audio even after substantial changes have been made.

Initially, this standard was developed to satisfy these requirements, of tight association and fragility, for an application in cascaded digital compression (bit rate reduction). Other applications have been identified that have these same two requirements, including the carriage of metadata to describe loudness.

### 0.2 Conventions used in this standard

#### 0.2.1 Decimal points

According to IEC directives, the comma is used in all text to indicate the decimal point. However, in the specified coding, including the examples shown, the full stop is used as in IEC programming language standards.

#### 0.2.2 Data representation

In this standard, all coding and data representations are printed in an equally spaced font.

#### 0.2.3 Non-printing ASCII characters

Non-printing characters are delimited by angle brackets, as in <CR> for carriage return.

### 0.3 Patents

The Audio Engineering Society draws attention to the fact that it is claimed that compliance with this AES standard may involve the use of an application for patents concerning "Lip sync" and "Lip-sync with sub-frame error feedback."

The AES holds no position concerning the evidence, validity and scope of this patent right.

The holder of this patent right has assured the AES that it is willing to negotiate licenses under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statement of the holder of this patent right is archived with the AES.

Information may be obtained from:

British Broadcasting Corporation  
Attention of Daniel Pike  
BBC Research and Development  
Centre House  
56 Wood Lane  
London, W12 7SB  
UK

Document preview:  
for full document, go to  
[www.aes.org/publications/standards](http://www.aes.org/publications/standards)

Attention is drawn to the possibility that some of the elements of this AES standard may be the subject of patent rights other than those identified herein. AES shall not be held responsible for identifying any or all such patent rights.

## 1 Scope

This document describes a mechanism for embedding a set of data in an audio signal. The syntax and semantics of a number of sets of data are defined in other parts of this standard. Provision is also made for the transmission of additional ancillary information. The data is embedded in the audio so that it is transported with the audio so that the data may be recovered and used to control subsequent processing of the audio, for example by low bit-rate coders or dynamics processors.

## 2 Normative references

The following standards contain provisions that, through reference in this text, constitute provisions of this document. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this document are encouraged to investigate the possibility of applying the most recent editions of the indicated standards.

This standard has no normative references

## 3 Definitions, symbols, and abbreviations

### 3.1 Definitions

#### 3.1.1

##### **CRC**

cyclic redundancy check

#### 3.1.2

##### **nextbits function**

##### **nextbits ( )**

function that permits comparison of a bit string with the next bits to be decoded in the bit stream

### 3.2 Symbols

The mathematical operators used in this standard are similar to those used in the C programming language. The bitwise operators are defined assuming 2's-complement representation of integers. Numbering and counting loops generally begin from zero.

#### 3.2.1 Arithmetic operators

##### 3.2.1.1

+

addition

##### 3.2.1.2

++

increment

##### 3.2.1.3

--

decrement