STANDARDS AND INFORMATION DOCUMENTS

AES41-2-2012
rev. AES41-2009
reaffirmed 2017

AUDIO
A
ES
STANDARDS

AES standard for digital audio - Audio-embedded metadata - Part 2: MPEG-1 Layer II or MPEG-2 LSF Layer II

Users of this standard are encouraged to determine if they are using the latest printing incorporating all current amendments and editorial corrections. Information on the latest status, edition, and printing of a standard can be found at: http://www.aes.org/standards

AUDIO ENGINEERING SOCIETY, INC.
551 Fifth Avenue, New York, New York, 10176, US.

AES41-2-2012
Revision of AES41-2009
reaffirmed 2017

AES standard for digital audio - Audio-embedded metadata - Part 2: MPEG-1 Layer II or MPEG-2 LSF Layer II

Published by
Audio Engineering Society, Inc.

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 describes the format for the data to be transmitted with audio decoded from MPEG-1 Layer II or MPEG-2 LSF Layer II encoding. The data may be used to aid a subsequent re-encoding process. A method of carrying this data is described in Part 1 of this Standard.

2017-12-31 printing

Contents

2017-12-31 printing

Foreword

This foreword is not part of the *AES41-2-2012 AES standard for digital audio - Audio-embedded metadata - Part 2: MPEG-1 Layer II or MPEG-2 LSF Layer II.*

This document describes a set of data that may be conveyed according to the method described in Part 1 of this Standard. It is one part of a multi-part revision of AES41-2009.

Digital compression is 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 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.

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

John Grant
Chair, working group SC-02-02
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”.

AES standard for digital audio - Audio-embedded metadata - Part 2: MPEG-1 Layer II or MPEG-2 LSF Layer II

0 Introduction

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 describes the format for the data to be transmitted with audio decoded from MPEG-1 Layer II or MPEG-2 LSF Layer II encoding. The data may be used to aid a subsequent re-encoding process. This part should be read in conjunction with part 1.

0.1 Rationale for part 2 of this standard

Digital compression techniques allow the use of lower bandwidth channels and less storage space or, conversely, allow faster transmission and the storage of more program material. 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 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, but might include framing and quantization resolution data.

0.2 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

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.

2017-12-31 printing

1 Scope

This document describes a format for the data to be transmitted to identify MPEG-1 Layer II and MPEG-2 LSF Layer II audio bit-rate reduction. This part assumes that the transmission mechanism according to part 1 of this Standard is used.

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.

AES41-1-xxxx, *AES standard for digital audio - Audio-embedded metadata - Part 1: General*, Audio Engineering Society, New York, NY., US.

ISO/IEC 11172-3 (1993-08), *Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s - Part 3: Audio*. International Electrotechnical Commission, Geneva, Switzerland.

ISO/IEC 13818-3 (1998-05), *Information technology - Generic coding of moving picture and associated audio information - Par 3: Audio*. International Electrotechnical Commission, Geneva, Switzerland.

3 Definitions, symbols, and abbreviations

3.1 Definitions

3.1.1
analysis filter bank
filter bank in the encoder that transforms a broadband pulse-code-modulation (PCM) audio signal into a set of subsampled sub-band samples

3.1.2
channel
means of carrying an audio signal (for example, a sequence of data forming part of a bit stream)

3.1.3
coded-audio bit stream
coded representation of an audio signal

3.1.4
compression
reduction in the number of bits used to represent an item of data

3.1.5
CRC
cyclic redundancy check

3.1.6
dual-channel mode
mode where two audio channels with independent program contents (for example, bilingual) are encoded within one bit stream

NOTE The coding process is the same as for the stereo mode.