

STANDARDS AND INFORMATION DOCUMENTS

AES41-3-2012
reaffirmed 2017



AES standard for digital audio — Audio-embedded metadata — Part 3: AAC & HE-AAC

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.

Document preview:
for full document, go to
www.aes.org/publications/standards

AES standard for digital audio - Audio-embedded metadata - Part 3: AAC & HE-AAC

Published by
Audio Engineering Society, Inc.
Copyright © 2012, 2017 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 describes the format for the data to be transmitted with audio to signal downmix coefficients, loudness and dynamic range control metadata as used in ISO/IEC 14496-3:2009 Subpart 4 (AAC and HE-AAC). A method of carrying this data is described in Part 1 of this Standard.

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.

Contents

0 Introduction	4
0.1 Rationale for part 3 of this standard.....	4
1 Scope	4
2 Normative references	4
3 Definitions and abbreviations	5
3.1 Definitions.....	5
3.2 Symbols.....	5
3.3 Abbreviations.....	6
4 ISO/IEC 14496-3:2009 minimal metadata set	7
4.1 AES41 metadata type.....	7
4.2 Syntax.....	7
4.3 Semantics.....	7
5 Sample alignment of data with respect to audio signal	8
5.1 General.....	8
5.2 ISO/IEC 14496-3:2009.....	8

Preview only
www.aes.org/standards

Foreword

This foreword is not part of AES41-3-2012 *AES standard for digital audio - Audio-embedded metadata - Part 3: AAC & HE-AAC*

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.

As predicted in the foreword to AES41-2000, 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 "downmix 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 part extends AES41 to include a data format for carrying this loudness and downmix 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.

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 3: AAC & HE-AAC

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 to signal downmix coefficients, loudness and dynamic range control metadata as used in ISO/IEC 14496-3:2009 Subpart 4 (AAC and HE-AAC).

0.1 Rationale for part 3 of this standard.

Unwanted loudness variations in broadcast audio have historically been the source of many audience complaints. Broadcasters have adopted numerous techniques to address the problem - one of which involves indicating the long-term average loudness of the audio.

MPEG-4 AAC and HE-AAC are increasingly being used for broadcast sound services associated with standard- and high-definition television services, in stereo and surround, and the MPEG-4 standard makes provision for metadata to control loudness and downmix. This part of the AES41 standard enable dynamic metadata to be carried to an MPEG-4 audio encoder as part of the PCM audio signal.

1 Scope

This document describes a format for the data to be transmitted to convey a subset of the data in ISO/IEC 14496-3, which included downmix and loudness metadata. This Part assumes that the transmission mechanism according to Part 1 of this Standard is used.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

AES41-1-2012, *AES standard for digital audio - Audio-embedded metadata - Part 1: General*, Audio Engineering Society, New York, NY., US.

ISO/IEC 14496-3:2005(E), *Information technology – Coding of audio-visual objects - Part 3: Audio*. International Electrotechnical Commission, Geneva, Switzerland.