STANDARDS AND INFORMATION DOCUMENTS

Call for comment on DRAFT REVISED AES standard for digital audio - Digital input-output interfacing - Sample-accurate timing in AES47

This document was developed by a writing group of the Audio Engineering Society Standards Committee (AESSC) and has been prepared for comment according to AES policies and procedures. It has been brought to the attention of International Electrotechnical Commission Technical Committee 100. Existing international standards relating to the subject of this document were used and referenced throughout its development.

Address comments by E-mail to standards@aes.org, or by mail to the AESSC Secretariat, Audio Engineering Society, 551 Fifth Ave., New York NY 10176. **Only comments so addressed will be considered**. E-mail is preferred. **Comments that suggest changes must include proposed wording.** Comments shall be restricted to this document only. Send comments to other documents separately. 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 will be approved by the AES after any adverse comment received within **six weeks** of the publication of this call on http://www.aes.org/standards/comments/, **2018-06-21**, has been resolved. Any person receiving this call first through the *JAES* distribution may inform the Secretariat immediately of an intention to comment within a month of this distribution.

Because this document is a draft and is subject to change, no portion of it shall be quoted in any publication without the written permission of the AES, and all published references to it must include a prominent warning that the draft will be changed and must not be used as a standard.

The AES Standards Committee is the organization responsible for the standards program of the Audio Engineering Society. It publishes technical standards, information documents and technical reports. Working groups and task groups with a fully international membership are engaged in writing standards covering fields that include topics of specific relevance to professional audio. Membership of any AES standards working group is open to all individuals who are materially and directly affected by the documents that may be issued under the scope of that working group.

Complete information, including working group scopes and project status is available at http://www.aes.org/standards. Enquiries may be addressed to standards@aes.org

The AES Standards Committee is supported in part by those listed below who, as Standards Sustainers, make significant financial contribution to its operation.

This list is current as of 2018/6/11

AES standard for digital audio - Digital input-output interfacing - Sample-accurate timing in AES47

Published by
Audio Engineering Society, Inc.

Abstract
This document specifies how the timing markers specified in AES47 may be used to associate an absolute timestamp with individual audio samples. AES47 specifies a format for the transmission of digital audio over asynchronous transfer mode (ATM) networks. A recommendation is made to refer these timestamps to the SMPTE epoch which in turn provides a reference to UTC and GPS time.

Contents

Foreword

This foreword is not part of AES53-2006 *AES standard for digital audio - Digital input-output interfacing - Sample-accurate timing in AES47*

This document was prepared by J. Grant for working group SC-02-02, under project AES-X148, “Sample-accurate timing in AES47”. The task group members included K. Brown, R. Caine, C. Chambers, P. Eastty, C. Gaunt, K. Gross, U. Horbach, H. Lemcke, M. Page, M. Schindele, J. Strawn, T. Thompson, J. Waas, and M. Yonge.

Foreword to 2018 edition

AES53-2006 uses the SMPTE epoch which was specified in SMPTE 404M at the time. That has been superseded by a different epoch specified in SMPTE ST 2059-1. This edition uses the current SMPTE epoch, and also includes some improvements to the text. Contributors include J. Fletcher, B. Harris, I. Rudd, and S. Scott.

J Grant, chair, SC-02-02

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 - Digital input-output interfacing - Sample-accurate timing in AES47

0 Introduction

0.1 General

This document provides a standard way of identifying the time to which each sample in an AES47 audio stream relates.

It thus provides a way of aligning streams from disparate sources, including synchronizing audio to video, and also allows the total delay across a network to be controlled when the transit time of individual cells is unknown.

This is most effective in systems where the audio is aligned with an absolute time reference such as GPS, but can also be used with a local reference.

The identification repeats every 16 seconds if sequence numbers are used in the AES47 stream, 1 second if only the "ATM user to ATM user indication" (UI) bit in the cell header is used. If absolute time can be distributed by a means outside the scope of this standard (see, for example, 4.4 and A.4.4 of AES51-2006), absolute timestamps can be associated with all audio samples.

0.2 Patents

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 standard specifies how the timing markers specified in 4.1.4.1.1 and 4.5 of AES47 may be used to associate an absolute timestamp with individual audio samples.

It does not specify how the recipient of a call is informed whether the timing markers will conform to this standard or merely meet the minimum specifications laid down in AES47.

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.

AES47: *AES standard for digital audio — Digital input-output interfacing — Transmission of digital audio over asynchronous transfer mode (ATM) networks*: Audio Engineering Society, New York, NY., US. 2002, 2006

SMPTE ST 2059-1:2015: Generation and Alignment of Interface Signals to the SMPTE Epoch, Society of Motion Picture and Television Engineers, White Plains, NY., US.

3 Definitions and abbreviations

For the purposes of this document, the following terms and definitions apply.

3.1
asynchronous transfer mode
ATM
networking technology in which data are carried in 48-octet cells

3.2
Subframe
data structure that bundles an audio sample with ancillary data and protocol overhead, as defined in AES47, 4.1.1

3.3
Global positioning system
GPS
satellite-based navigation system that makes accurate timing information globally available.

4 Numbering of samples

PCM samples in an audio stream shall be numbered consecutively as if the sample clock had been running continuously at its nominal frequency since the "sample-numbering epoch", such that the first sample occurring at or after the sample-numbering epoch shall be sample number 0.

The sample-numbering epoch should be the SMPTE epoch specified in SMPTE ST 2059-1. However, a different epoch may be used in applications that require it; how the use of a different epoch is communicated to the audio source and destination is outside the scope of this standard.

> Note: The SMPTE Epoch is 1970-01-01T00:00:00TAI, which is the same as the PTP Epoch specified in IEEE Standard 1588-2008 and is 63 072 010 seconds before 1972-01-01T00:00:00Z (UTC).
> The sample at GPS time w weeks + s seconds is number (315 964 819 + (w × 604 800) + s) × [sample rate], remembering that w is now more than 1024 and only the least significant 10 bits of it are conveyed in the GPS time format.

5 Numbering of cells

Samples shall be aligned such that sample 0 of channel 1 is in the first subframe in a cell.

> NOTE 1: Where the "multi-channel" packing scheme specified in 4.2.3 of AES47 is used, every sample of channel 1 is the first subframe in a cell, so this requirement is only relevant for the other packing schemes.

Cells shall be numbered consecutively such that cell 0 contains sample 0 of channel 1.

> NOTE 2: Where the "multi-channel" packing scheme specified in 4.2.3 of AES47 is used, sample s is in cells ns to $ns+n-1$, where n is the number of cells per sampling interval.

> NOTE 3: Where one of the packing schemes specified in 4.2.2 and 4.2.4 of AES47 is used, cell c contains samples nc to $nc+n-1$, where n is the number of sampling intervals per cell.

6 Sequence numbers and blocks

The sequence number specified in 4.1.4.1.1 of AES47 shall be set to the least significant four bits of the binary representation of the cell number.

The grouping of cells into blocks as specified in 4.2 and 4.5 of AES47 shall be such that cell 0 is the first of a block.

For each t such that a new GPS second starts at sample t or between samples t-1 and t, the "local clock" specified in 4.5.2 of AES47 shall tick between samples t'-1 and t', where t' is the smallest multiple of 3072 not less than t.

When the "second number" is incremented as specified in 4.1.4.1.3(a) of AES47, the new value shall be set equal (modulo 16) to the number of seconds that have elapsed since the sample-numbering epoch.

Annex A: (Informative) Background information

A.1 Choice of origin of sample numbering

The SMPTE epoch was chosen as the origin because it is the time to which many television signals are aligned.

In some applications, timing is reset periodically (for example, daily at midnight). In such applications, it will be appropriate to use the instant when the reset occurs as the sample-numbering epoch, and this is permitted but discouraged. If calls which are connected at the time of the reset switch to the new numbering, there may be one second where the number of samples between ticks is not a multiple of 3072 and this might cause the audio to skip or mute. If, on the other hand, they continue to use the previous epoch then they may not be correctly aligned with other streams which use the new epoch; also, special provision is needed when adding a new destination to a call which has existed since before the current epoch.

A.2 Calculation of sample number

In the note to clause 4, the numerical coefficients are derived as follows.

Although the GPS epoch is defined as a time point on the UTC timescale, the offset from the SMPTE epoch (1970-01-01T00:00:00 TAI) to the GPS epoch (1980-01-06T00:00:19 TAI) is 10 years, 2 of which are leap years, plus 5 days and 19 seconds. The number of days is thus (10 x 365) + 2 + 5 = 3657, and the number of seconds (3657 x 24 x 60 x 60) + 19 = 315 964 819. The number of seconds in a week is 7 x 24 x 60 x 60 = 604 800.

Note that the sample number is only required modulo 3072, so can be calculated as ((403 + 2688w + s) x [sample rate]) modulo 3072, where w is the week number modulo 1024, as conveyed in the GPS message.

Since 1972-01-01 00:00:00 (UTC) UTC has differed from TAI by an integral number of seconds, with an initial calibration offset of 10 seconds. A "Leap Second" is occasionally introduced to the UTC-TAI offset value to compensate for the irregular rotation of the Earth and maintain an approximate alignment of calendar days to observed solar days.

The SMPTE epoch lies two years and ten seconds before 1972-01-01 00:00:00 (UTC), that is 730 days plus 10 seconds, or 63 072 010 s.

A.3 Local clock ticks

AES47 states that the ticks occur once per second, but with no specified accuracy. This standard adjusts the length of each "second" to be a multiple of 3072 samples, for instance at 48 kHz each "second" will be either 46 080 or 49 152 samples, the average over any 8 consecutive seconds being 48 000.

The number 3072 was chosen as follows; in the description this number of samples is called a "quantum".

The smallest subframe permitted by AES47 is 8 bits, so a single-channel call using this format has 48 samples per cell so $48 \times 8 = 384$ samples per block. The smallest subframe that includes a sequence number is 16 bits (e.g. 12 audio bits + protocol overhead; 4 audio bits + protocol overhead is permitted by 4.1 but is not supported by 6.2 and is anyway unlikely to be useful); a single-channel call using this format has 24 samples per cell so $24 \times 16 = 384$ samples between one cell with sequence number 0 and the next. Thus if the quantum is a multiple of 384, the tick will always occur just before the start of a block, and if sequence numbers are used it will also be just before a cell with sequence number zero.

It is likely to be helpful for receiving equipment to be able to put the incoming audio into a circular buffer whose size is equal to a quantum or a fraction of a quantum; then it can align the incoming audio simply by ensuring that the first sample after a "new second" marker goes at the start of the buffer. This buffer needs to be large enough to absorb all the jitter that may be introduced from all sources (not just the network) so should be as large as conveniently possible; with the chosen quantum size it is 64 ms at 48 kHz and 16 ms at 192 kHz. Note that the quantum sets an upper limit to the size of this buffer but not a lower limit.

However, a quantum needs to be significantly less than 1 second. The lowest sample rate supported by 6.4 of AES47 is 8 kHz, at which a quantum is 384 ms.