
**Information technology — Coding of
audio-visual objects —**

**Part 3:
Audio**

Technologies de l'information — Codage des objets audiovisuels —

Partie 3: Codage audio

iTech Standards
(<https://standards.iteh.ai>)
Document Preview

[ISO/IEC 14496-3:2019](https://standards.iteh.ai/catalog/standards/iso/337e9c6d-bb07-485b-aacb-b62fba1f13c3/iso-iec-14496-3-2019)

<https://standards.iteh.ai/catalog/standards/iso/337e9c6d-bb07-485b-aacb-b62fba1f13c3/iso-iec-14496-3-2019>



iTeh Standards
(<https://standards.iteh.ai>)
Document Preview

[ISO/IEC 14496-3:2019](https://standards.iteh.ai/catalog/standards/iso/337e9c6d-bb07-485b-aacb-b62fba1f13c3/iso-iec-14496-3-2019)

<https://standards.iteh.ai/catalog/standards/iso/337e9c6d-bb07-485b-aacb-b62fba1f13c3/iso-iec-14496-3-2019>



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2019

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Foreword	iv
0 Introduction	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	2
4 Abbreviated terms	21
5 Structure of this document	23
Subpart 1: Main	24
Subpart 2: Speech coding — HVXC	169
Subpart 3: Speech coding — CELP	320
Subpart 4: General Audio coding (GA) — AAC, TwinVQ, BSAC	483
Subpart 5: Structured Audio (SA)	895
Subpart 6: Text To Speech Interface (TTSI)	1043
Subpart 7: Parametric Audio Coding — HILN	1053
Subpart 8: Parametric coding for high quality audio — SSC	1112
Subpart 9: MPEG-1/2 Audio in MPEG-4	1231
Subpart 10: Lossless coding of oversampled audio — DST	1244
Subpart 11: Audio lossless coding — ALS	1281
Subpart 12: Scalable lossless coding— SLS	1355
Subpart 13: Audio Synchronization	1429

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see <http://patents.iec.ch>).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, Coding of audio, picture, multimedia and hypermedia information.

This fifth edition cancels and replaces the fourth edition (ISO/IEC 14496-3:2009), which has been technically revised. It incorporates the Amendments ISO/IEC 14496-3:2009/Amd.1:2009, ISO/IEC 14496-3:2009/Amd.2:2010, ISO/IEC 14496-3:2009/Amd.3:2012, ISO/IEC 14496-3:2009/Amd.4:2013, ISO/IEC 14496-3:2009/Amd.4:2013/Cor.1:2015, ISO/IEC 14496-3:2009/Amd.5:2015, ISO/IEC 14496-3:2009/Amd.6:2017 and ISO/IEC 14496-3:2009/Amd.7:2018 as well as Technical Corrigenda ISO/IEC 14496-3:2009/Cor.1:2009, ISO/IEC 14496-3:2009/Cor.2:2011, ISO/IEC 14496-3:2009/Cor.3:2012, ISO/IEC 14496-3:2009/Cor.4:2012, ISO/IEC 14496-3:2009/Cor.5:2015, ISO/IEC 14496-3:2009/Cor.6:2015, ISO/IEC 14496-3:2009/Cor.7:2015.

A list of all parts in the ISO/IEC 14496 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

0 Introduction

0.1 Overview

ISO/IEC 14496-3 (MPEG-4 Audio) is a new kind of audio standard that integrates many different types of audio coding: natural sound with synthetic sound, low bitrate delivery with high-quality delivery, speech with music, complex soundtracks with simple ones, and traditional content with interactive and virtual-reality content. By standardizing individually sophisticated coding tools as well as a novel, flexible framework for audio synchronization, mixing, and downloaded post-production, the developers of the MPEG-4 Audio standard have created new technology for a new, interactive world of digital audio.

MPEG-4, unlike previous audio standards created by ISO/IEC and other groups, does not target a single application such as real-time telephony or high-quality audio compression. Rather, MPEG-4 Audio is a standard that applies to every application requiring the use of advanced sound compression, synthesis, manipulation, or playback. The subparts that follow specify the state-of-the-art coding tools in several domains; however, MPEG-4 Audio is more than just the sum of its parts. As the tools described here are integrated with the rest of the MPEG-4 standard, exciting new possibilities for object-based audio coding, interactive presentation, dynamic soundtracks, and other sorts of new media, are enabled.

Since a single set of tools is used to cover the needs of a broad range of applications, *interoperability* is a natural feature of systems that depend on the MPEG-4 Audio standard. A system that uses a particular coder — for example a real-time voice communication system making use of the MPEG-4 speech coding toolset — can easily share data and development tools with other systems, even in different domains, that use the same tool — for example a voicemail indexing and retrieval system making use of MPEG-4 speech coding.

The remainder of this clause gives a more detailed overview of the capabilities and functioning of MPEG-4 Audio. First a discussion of concepts, that have changed since the MPEG-2 audio standards, is presented. Then the MPEG-4 Audio toolset is outlined.

0.2 Concepts of MPEG-4 Audio

0.2.1 General

As with previous MPEG standards, MPEG-4 does not standardize methods for encoding sound. Thus, content authors are left to their own decisions as to the best method of creating bitstream payloads. At the present time, methods to automatically convert natural sound into synthetic or multi-object descriptions are not mature; therefore, most immediate solutions will involve interactively-authoring the content stream in some way. This process is similar to current schemes for MIDI-based and multi-channel mixdown authoring of soundtracks.

Many concepts in MPEG-4 Audio are different than those in previous MPEG Audio standards. For the benefit of readers who are familiar with MPEG-1 and MPEG-2 we provide a brief overview here.

0.2.2 Audio storage and transport facilities

In all of the MPEG-4 tools for audio coding, the coding standard ends at the point of constructing access units that contain the compressed data. The MPEG-4 Systems (ISO/IEC 14496-1) specification describes how to convert these individually coded access units into elementary streams.

There is no standard transport mechanism of these elementary streams over a channel. This is because the broad range of applications that can make use of MPEG-4 technology have delivery requirements that are too wide to easily characterize with a single solution. Rather, what is standardized is an interface (the Delivery Multimedia Interface Format, or DMIF, specified in ISO/IEC 14496-6) that describes the capabilities of a transport layer and the communication between transport, multiplex, and demultiplex functions in encoders and decoders. The use of DMIF and the MPEG-4 Systems specification allows transmission functions that are much more sophisticated than are possible with previous MPEG standards.

However, LATM and LOAS were defined to provide a low overhead audio multiplex and transport mechanism for natural audio applications, which do not require sophisticated object-based coding or other functions provided by MPEG-4 Systems.

Table 0.1 gives an overview about the multiplex, storage and transmission formats currently available for MPEG-4 Audio within the MPEG-4 framework:

Table 0.1 – MPEG-4 Audio multiplex, storage and transmission formats

	Format	Functionality defined in MPEG-4:	Functionality originally defined in:	Description
Multiplex	M4Mux	ISO/IEC 14496-1 (normative)	-	MPEG-4 Multiplex scheme
	LATM	ISO/IEC 14496-3 (normative)	-	Low Overhead Audio Transport Multiplex
Storage	ADIF	ISO/IEC 14496-3 (informative)	ISO/IEC 13818-7 (normative)	Audio Data Interchange Format, (AAC only)
	MP4FF	ISO/IEC 14496-12 (normative)	-	MPEG-4 File Format
Transmission	ADTS	ISO/IEC 14496-3 (informative)	ISO/IEC 13818-7 (normative, exemplarily)	Audio Data Transport Stream, (AAC only)
	LOAS	ISO/IEC 14496-3 (normative, exemplarily)	-	Low Overhead Audio Stream, based on LATM, three versions are available: AudioSyncStream() EPAudioSyncStream() AudioPointerStream()

To allow for a user on the remote side of a channel to dynamically control a server streaming MPEG-4 content, MPEG-4 defines backchannel streams that can carry user interaction information.

0.2.3 MPEG-4 Audio supports low-bitrate coding

Previous MPEG Audio standards have focused primarily on transparent (undetectable) or nearly transparent coding of high-quality audio at whatever bitrate was required to provide it. MPEG-4 provides new and improved tools for this purpose, but also standardizes (and has tested) tools that can be used for transmitting audio at the low bitrates suitable for Internet, digital radio, or other bandwidth-limited delivery. The new tools specified in MPEG-4 are the state-of-the-art tools that support low-bitrate coding of speech and other audio.

0.2.4 MPEG-4 Audio is an object-based coding standard with multiple tools

Previous MPEG Audio standards provided a single toolset, with different configurations of that toolset specified for use in various applications. MPEG-4 provides several toolsets that have no particular relationship to each other, each with a different target function. The profiles of MPEG-4 Audio specify which of these tools are used together for various applications.

Further, in previous MPEG standards, a single (perhaps multi-channel or multi-language) piece of content was transmitted. In contrast, MPEG-4 supports a much more flexible concept of a *soundtrack*. Multiple tools may be used to transmit several *audio objects*, and when using multiple tools together an *audio composition* system is provided to create a single soundtrack from the several audio substreams. User interaction, terminal capability, and speaker configuration may be used when determining how to produce a single soundtrack from the component objects. This capability gives MPEG-4 significant advantages in quality and flexibility when compared to previous audio standards.

0.2.5 MPEG-4 Audio provides capabilities for synthetic sound

In natural sound coding, an existing sound is compressed by a server, transmitted and decompressed at the receiver. This type of coding is the subject of many existing standards for sound compression. In contrast, MPEG-4 standardizes a novel paradigm in which synthetic sound descriptions, including synthetic speech and synthetic music, are transmitted and then *synthesized* into sound at the receiver. Such capabilities open up new areas of very-low-bitrate but still very-high-quality coding.

0.2.6 MPEG-4 Audio provides capabilities for error robustness

Improved error robustness capabilities for all coding tools are provided through the error resilient bitstream payload syntax. This tool supports advanced channel coding techniques, which can be adapted to the special needs of given coding tools and a given communications channel. This error resilient bitstream payload syntax is mandatory for all error resilient object types.

The error protection tool (EP tool) provides unequal error protection (UEP) for MPEG-4 Audio in conjunction with the error resilient bitstream payload. UEP is an efficient method to improve the error robustness of source coding schemes. It is used by various speech and audio coding systems operating over error-prone channels such as mobile telephone networks or Digital Audio Broadcasting (DAB). The bits of the coded signal representation are first grouped into different classes according to their error sensitivity. Then error protection is individually applied to the different classes, giving better protection to more sensitive bits.

Improved error robustness for AAC is provided by a set of error resilience tools. These tools reduce the perceived degradation of the decoded audio signal that is caused by corrupted bits in the bitstream payload.

0.2.7 MPEG-4 Audio provides capabilities for scalability

Previous MPEG Audio standards provided a single bitrate, single bandwidth toolset, with different configurations of that toolset specified for use in various applications. MPEG-4 provides several bitrate and bandwidth options within a single stream, providing a scalability functionality that permits a given stream to scale to the requirement of different channels and applications or to be responsive to a given channel that has dynamic throughput characteristics. The tools specified in MPEG-4 are the state-of-the-art tools providing scalable compression of speech and audio signals.

0.3 The MPEG-4 Audio tool set

0.3.1 Speech coding tools

0.3.1.1 Overview

Speech coding tools are designed for the transmission and decoding of synthetic and natural speech.

Two types of speech coding tools are provided in MPEG-4. The *natural* speech tools allow the compression, transmission, and decoding of human speech, for use in telephony, personal communication, and surveillance applications. The *synthetic* speech tool provides an interface to text-to-speech synthesis systems; using synthetic speech provides very-low-bitrate operation and built-in connection with facial animation for use in low-bitrate video teleconferencing applications.

0.3.1.2 Natural speech coding

The MPEG-4 speech coding toolset covers the compression and decoding of natural speech sound at bitrates ranging between 2 and 24 kbit/s. When variable bitrate coding is allowed, coding at even less than 2 kbit/s, for example an average bitrate of 1.2 kbit/s, is also supported. Two basic speech coding techniques are used: One is a parametric speech coding algorithm, HVXC (Harmonic Vector eXcitation Coding), for very low bit rates; and the other is a CELP (Code Excited Linear Prediction) coding technique. The MPEG-4 speech coders target applications range from mobile and satellite communications, to Internet telephony, to packaged media and speech databases. It meets a wide range of requirements encompassing bitrate, functionality and sound quality.

MPEG-4 HVXC operates at fixed bitrates between 2.0 kbit/s and 4.0 kbit/s using a bitrate scalability technique. It also operates at lower bitrates, typically 1.2 - 1.7 kbit/s, using a variable bitrate technique. HVXC provides communications-quality to near-toll-quality speech in the 100 Hz – 3800 Hz band at 8 kHz sampling rate. HVXC also allows independent change of speed and pitch during decoding, which is a powerful functionality for fast access to speech databases. HVXC functionalities including 2.0 - 4.0 kbit/s fixed bitrate modes and a 2.0 kbit/s maximum variable bitrate mode.

Error Resilient (ER) HVXC extends operation of the variable bitrate mode to 4.0 kbit/s to allow higher quality variable rate coding. The ER HVXC therefore provides fixed bitrate modes of 2.0 - 4.0 kbit/s and a variable bitrate of either less than 2.0 kbit/s or less than 4.0 kbit/s, both in scalable and non-scalable modes. In the