

---

---

**Information technology — MPEG  
audio technologies —**

**Part 3:  
Unified speech and audio coding**

*Technologies de l'information — Technologies audio MPEG —  
Partie 3: Codage unifié parole et audio*





**COPYRIGHT PROTECTED DOCUMENT**

© ISO/IEC 2020

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](mailto:copyright@iso.org)  
Website: [www.iso.org](http://www.iso.org)

Published in Switzerland

# Contents

Page

Foreword.....	vii
Introduction.....	viii
<b>1</b> <b>Scope</b> .....	<b>1</b>
<b>2</b> <b>Normative references</b> .....	<b>1</b>
<b>3</b> <b>Terms, definitions, symbols and abbreviated terms</b> .....	<b>1</b>
<b>3.1</b> <b>Terms and definitions</b> .....	<b>1</b>
<b>3.2</b> <b>Symbols and abbreviated terms</b> .....	<b>3</b>
<b>4</b> <b>Technical overview</b> .....	<b>4</b>
<b>4.1</b> <b>Decoder block diagram</b> .....	<b>4</b>
<b>4.2</b> <b>Overview of the decoder tools</b> .....	<b>5</b>
<b>4.3</b> <b>Combination of USAC with MPEG Surround and SAOC</b> .....	<b>9</b>
<b>4.4</b> <b>Interface between USAC and systems</b> .....	<b>9</b>
<b>4.4.1</b> <b>Decoder behaviour</b> .....	<b>10</b>
<b>4.5</b> <b>USAC profiles and levels</b> .....	<b>10</b>
<b>4.5.1</b> <b>General</b> .....	<b>10</b>
<b>4.5.2</b> <b>MPEG-4 HE AACv2 compatibility</b> .....	<b>11</b>
<b>4.5.3</b> <b>Baseline USAC profile</b> .....	<b>12</b>
<b>4.5.4</b> <b>Extended high efficiency AAC profile</b> .....	<b>13</b>
<b>4.6</b> <b>Combination of USAC with MPEG-D DRC</b> .....	<b>14</b>
<b>5</b> <b>Syntax</b> .....	<b>15</b>
<b>5.1</b> <b>General</b> .....	<b>15</b>
<b>5.2</b> <b>Decoder configuration (UsacConfig)</b> .....	<b>15</b>
<b>5.3</b> <b>USAC bitstream payloads</b> .....	<b>20</b>
<b>5.3.1</b> <b>Payloads for audio object type USAC</b> .....	<b>20</b>
<b>5.3.2</b> <b>Subsidiary payloads</b> .....	<b>23</b>
<b>5.3.3</b> <b>Payloads for enhanced SBR</b> .....	<b>34</b>
<b>5.3.4</b> <b>Payloads for MPEG Surround</b> .....	<b>43</b>
<b>5.3.5</b> <b>Payload of extension elements</b> .....	<b>53</b>
<b>6</b> <b>Data structure</b> .....	<b>53</b>
<b>6.1</b> <b>USAC configuration</b> .....	<b>53</b>
<b>6.1.1</b> <b>Definition of elements</b> .....	<b>53</b>
<b>6.1.2</b> <b>UsacConfig()</b> .....	<b>63</b>
<b>6.1.3</b> <b>Usac Output Sampling Frequency</b> .....	<b>63</b>
<b>6.1.4</b> <b>UsacChannelConfig()</b> .....	<b>63</b>
<b>6.1.5</b> <b>UsacDecoderConfig()</b> .....	<b>64</b>
<b>6.1.6</b> <b>UsacSingleChannelElementConfig()</b> .....	<b>64</b>
<b>6.1.7</b> <b>UsacChannelPairElementConfig()</b> .....	<b>64</b>
<b>6.1.8</b> <b>UsacLfeElementConfig()</b> .....	<b>64</b>
<b>6.1.9</b> <b>UsacCoreConfig()</b> .....	<b>64</b>
<b>6.1.10</b> <b>SbrConfig()</b> .....	<b>65</b>
<b>6.1.11</b> <b>SbrDfltHeader()</b> .....	<b>65</b>
<b>6.1.12</b> <b>Mps212Config()</b> .....	<b>65</b>
<b>6.1.13</b> <b>UsacExtElementConfig()</b> .....	<b>65</b>
<b>6.1.14</b> <b>UsacConfigExtension()</b> .....	<b>65</b>
<b>6.1.15</b> <b>Unique stream identifier (Stream ID)</b> .....	<b>66</b>
<b>6.2</b> <b>USAC payload</b> .....	<b>66</b>
<b>6.2.1</b> <b>Definition of elements</b> .....	<b>66</b>
<b>6.2.2</b> <b>UsacFrame()</b> .....	<b>68</b>
<b>6.2.3</b> <b>UsacSingleChannelElement()</b> .....	<b>69</b>
<b>6.2.4</b> <b>UsacExtElement()</b> .....	<b>69</b>
<b>6.2.5</b> <b>UsacChannelPairElement()</b> .....	<b>70</b>
<b>6.2.6</b> <b>Low frequency enhancement (LFE) channel element, UsacLfeElement()</b> .....	<b>70</b>
<b>6.2.7</b> <b>UsacCoreCoderData()</b> .....	<b>71</b>

6.2.8	StereoCoreToolInfo()	71
6.2.9	fd_channel_stream() and ics_info()	72
6.2.10	lpd_channel_stream()	76
6.2.11	Spectral noiseless coder	79
6.2.12	Enhanced SBR	80
6.2.13	Definition of MPEG Surround 2-1-2 payloads	82
6.2.14	Buffer requirements	84
7	Tool descriptions	85
7.1	Quantization	85
7.1.1	Tool description	85
7.1.2	Definition of elements	85
7.1.3	Decoding process	85
7.2	Noise filling	85
7.2.1	Tool description	85
7.2.2	Definition of elements	86
7.2.3	Decoding process	86
7.2.4	Generation of random signs for spectral noise filling	87
7.3	Scale factors	87
7.4	Spectral noiseless coding	87
7.4.1	Tool description	87
7.4.2	Definition of elements	88
7.4.3	Decoding process	89
7.5	enhanced SBR tool (eSBR)	93
7.5.1	Modifications to SBR tool	93
7.5.2	Additional pre-processing in the MPEG-4 SBR within USAC	108
7.5.3	DFT based harmonic transposer	110
7.5.4	QMF based harmonic transposer	120
7.5.5	4:1 Structure for SBR in USAC	128
7.5.6	Predictive vector coding (PVC) decoding process	138
7.6	Inter-subband-sample temporal envelope shaping (inter-TES)	141
7.6.1	Tool Description	141
7.6.2	Definition of elements	142
7.6.3	Inter-TES	142
7.7	Joint stereo coding	144
7.7.1	M/S stereo	144
7.7.2	Complex stereo prediction	144
7.8	TNS	151
7.8.1	General	151
7.8.2	Definition of elements	151
7.8.3	Decoding process	152
7.8.4	Maximum TNS bandwidth	152
7.9	Filterbank and block switching	153
7.9.1	Tool description	153
7.9.2	Definition of elements	153
7.9.3	Decoding process	153
7.10	Time-warped filterbank and blockswitching	161
7.10.1	Tools description	161
7.10.2	Definition of elements	161
7.10.3	Decoding process	163
7.11	MPEG Surround for mono to stereo upmixing	169
7.11.1	Tool description	169
7.11.2	Decoding process	170
7.12	AVQ decoding	182
7.13	LPC-filter	189
7.13.1	Tool description	189
7.13.2	Definition of elements	189
7.13.3	Number of LPC filters	189
7.13.4	General principle of the inverse quantizer	189
7.13.5	Decoding of the LPC quantization mode	190

7.13.6	First-stage approximation .....	191
7.13.7	AVQ refinement .....	191
7.13.8	Reordering of quantized LSFs .....	193
7.13.9	Conversion into LSP parameters.....	193
7.13.10	Interpolation of LSP parameters.....	194
7.13.11	LSP to LP conversion.....	194
7.13.12	LPC initialization at decoder start-up .....	195
7.14	ACELP.....	196
7.14.1	General.....	196
7.14.2	Definition of elements.....	196
7.14.3	ACELP initialization at USAC decoder start-up .....	197
7.14.4	Setting of the ACELP excitation buffer using the past FD synthesis and LPC0.....	197
7.14.5	Decoding of CELP excitation.....	197
7.14.6	Excitation postprocessing.....	203
7.14.7	Synthesis .....	204
7.14.8	Writing in the output buffer .....	204
7.15	MDCT based TCX.....	205
7.15.1	Tool description .....	205
7.15.2	Decoding process .....	205
7.16	Forward aliasing cancellation (FAC) tool.....	209
7.16.1	Tool description .....	209
7.16.2	Definition of elements.....	209
7.16.3	Decoding process .....	210
7.16.4	Writing in the output buffer.....	211
7.17	Post-processing of the synthesis signal .....	212
7.18	Audio pre-roll .....	214
7.18.1	General.....	214
7.18.2	Semantics .....	214
7.18.3	Decoding process .....	215
8	Conformance testing.....	217
8.1	General.....	217
8.2	USAC conformance testing.....	217
8.2.1	Profiles .....	217
8.2.2	Conformance tools and test procedure.....	218
8.3	USAC bitstreams .....	222
8.3.1	General.....	222
8.3.2	USAC configuration.....	222
8.3.3	Framework.....	225
8.3.4	Frequency domain coding (FD mode).....	226
8.3.5	Linear predictive domain coding (LPD mode).....	228
8.3.6	Common core coding tools .....	229
8.3.7	Enhanced spectral band replication (eSBR) .....	230
8.3.8	eSBR – Predictive vector coding (PVC).....	232
8.3.9	eSBR – Inter temporal envelope shaping (inter-TES).....	233
8.3.10	MPEG Surround 2-1-2 .....	233
8.3.11	Configuration Extensions.....	235
8.3.12	AudioPreRoll.....	235
8.3.13	DRC .....	236
8.3.14	Restrictions depending on profiles and levels.....	236
8.4	USAC decoders.....	238
8.4.1	General.....	238
8.4.2	FD core mode tests .....	238
8.4.3	LPD core mode tests.....	245
8.4.4	Combined core coding tests .....	250
8.4.5	eSBR tests.....	251
8.4.6	MPEG Surround 212 tests .....	260
8.4.7	Bitstream extensions.....	262
8.5	Decoder settings.....	265
8.5.1	General.....	265

8.5.2	Target loudness [Lou-<x>] .....	265
8.5.3	DRC effect type request [Eff-<x>] .....	265
8.6	Decoding of MPEG-4 file format parameters to support exact time alignment in file-to-file coding .....	265
9	Reference software .....	266
9.1	Reference software structure.....	266
9.1.1	General.....	266
9.1.2	Copyright disclaimer for software modules .....	266
9.2	Bitstream decoding software .....	268
9.2.1	General.....	268
9.2.2	USAC decoding software.....	268
Annex A (normative)	Tables.....	269
Annex B (informative)	Encoder tools.....	274
Annex C (normative)	Tables for arithmetic decoder.....	314
Annex D (normative)	Tables for predictive vector coding.....	320
Annex E (informative)	Adaptive time/frequency post-processing.....	329
Annex F (informative)	Audio/systems interaction.....	335
Annex G (informative)	Reference software .....	337
Annex H (normative)	Carriage of MPEG-D USAC in ISO base media file format .....	338
Bibliography	.....	339

## 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](http://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](http://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](http://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 second edition cancels and replaces the first edition (ISO/IEC 23003-3:2012), which has been technically revised. It also incorporates ISO/IEC 23003-3:2012/Cor.1:2012, ISO/IEC 23003-3:2012/Cor.2:2013, ISO/IEC 23003-3:2012/Cor.3:2015, ISO/IEC 23003-3:2012/Cor.4:2015, ISO/IEC 23003-3:2012/Amd.1:2014, ISO/IEC 23003-3:2012/Amd.1:2014/Cor.1:2015, ISO/IEC 23003-3:2012/Amd.2:2015, ISO/IEC 23003-3:2012/Amd.2:2015/Cor.1:2015 and ISO/IEC 23003-3:2012/Amd.3:2016.

A list of all parts in the ISO/IEC 23003 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](http://www.iso.org/members.html).

## **Introduction**

As mobile appliances become multi-functional, multiple devices converge into a single device. Typically, a wide variety of multimedia content is required to be played on or streamed to these mobile devices, including audio data that consists of a mix of speech and music.

This document specifies unified speech and audio coding (USAC), which allows for coding of speech, audio or any mixture of speech and audio with a consistent audio quality for all sound material over a wide range of bitrates. It supports single and multi-channel coding at high bitrates and provides perceptually transparent quality. At the same time, it enables very efficient coding at very low bitrates while retaining the full audio bandwidth.

Where previous audio codecs had specific strengths in coding either speech or audio content, USAC is able to encode all content equally well, regardless of the content type.

In order to achieve equally good quality for coding audio and speech, the developers of USAC employed the proven MDCT-based transform coding techniques known from MPEG-4 audio and combined them with specialized speech coder elements like ACELP. Parametric coding tools such as MPEG-4 spectral band replication (SBR) and MPEG-D MPEG surround were enhanced and tightly integrated into the codec. The result delivers highly efficient coding and operates down to the lowest bit rates.

The main focus of this codec are applications in the field of typical broadcast scenarios, multimedia download to mobile devices, user-generated content such as podcasts, digital radio, mobile TV, audio books, etc.

The International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) draw attention to the fact that it is claimed that compliance with this document may involve the use of a patent.

ISO and IEC take no position concerning the evidence, validity and scope of this patent right.

The holder of this patent right has assured ISO and IEC that he/she is willing to negotiate licences 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 registered with ISO and IEC. Information may be obtained from the patent database available at [www.iso.org/patents](http://www.iso.org/patents).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights other than those in the patent database. ISO and IEC shall not be held responsible for identifying any or all such patent rights.



# Information technology — MPEG audio technologies —

## Part 3: Unified speech and audio coding

### 1 Scope

This document specifies a unified speech and audio codec which is capable of coding signals having an arbitrary mix of speech and audio content. The codec has a performance comparable to, or better than, the best known coding technology that might be tailored specifically to coding of either speech or general audio content. The codec supports single and multi-channel coding at high bitrates and provides perceptually transparent quality. At the same time, it enables very efficient coding at very low bitrates while retaining the full audio bandwidth.

This document incorporates several perceptually-based compression techniques developed in previous MPEG standards: perceptually shaped quantization noise, parametric coding of the upper spectrum region and parametric coding of the stereo sound stage. However, it combines these well-known perceptual techniques with a source coding technique: a model of sound production, specifically that of human speech.

### 2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements 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.

ISO/IEC 14496-3:2019, *Information technology — Coding of audio-visual objects — Part 3: Audio*

ISO/IEC 14496-26:2010, *Information technology — Coding of audio-visual objects — Part 26: Audio conformance*

ISO/IEC 23003-1, *Information technology — MPEG audio technologies — Part 1: MPEG Surround*

ISO/IEC 23003-4, *Information technology — MPEG audio technologies — Part 4: Dynamic range control*