
**Information technology — MPEG audio
technologies —**

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

Technologies de l'information — Technologies audio MPEG —

Partie 3: Discours unifié et codage audio

This document is a preview generated by EVS



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2012

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword	iv
Introduction.....	v
1 Scope	1
2 Normative references	1
3 Terms, definitions, symbols and abbreviated terms	1
3.1 Terms and definitions	1
3.2 Symbols and abbreviated terms	2
4 Technical Overview	2
4.1 Decoder block diagram.....	2
4.2 Overview of the decoder tools	5
4.3 Combination of USAC with MPEG Surround and SAOC.....	8
4.4 Interface between USAC and Systems.....	9
4.5 USAC Profiles and Levels.....	9
5 Syntax.....	12
5.1 General	12
5.2 Decoder configuration (UsacConfig).....	12
5.3 USAC bitstream payloads	17
6 Data Structure	50
6.1 USAC configuration	50
6.2 USAC payload.....	63
7 Tool Descriptions	81
7.1 Quantization	81
7.2 Noise Filling	82
7.3 Scalefactors	84
7.4 Spectral Noiseless coding.....	84
7.5 enhanced SBR Tool (eSBR).....	90
7.6 Inter-subband-sample Temporal Envelope Shaping (inter-TES).....	139
7.7 Joint Stereo Coding	142
7.8 TNS.....	149
7.9 Filterbank and block switching.....	151
7.10 Time-Warped Filterbank and Blockswitching	159
7.11 MPEG Surround for Mono to Stereo upmixing	167
7.12 AVQ decoding.....	180
7.13 LPC-filter	186
7.14 ACELP.....	193
7.15 MDCT based TCX.....	202
7.16 Forward Aliasing Cancellation (FAC) tool	206
7.17 Post-processing of the synthesis signal	208
Annex A (normative) Tables	211
Annex B (informative) Encoder Tools	216
Annex C (normative) Tables for Arithmetic Decoder	254
Annex D (normative) Tables for Predictive Vector Coding	260
Annex E (informative) Adaptive Time / Frequency Post-Processing	269
Annex F (informative) Audio/Systems Interaction.....	275
Annex G (informative) Patent Statements	277
Bibliography.....	278

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. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

ISO/IEC 23003-3 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

ISO/IEC 23003 consists of the following parts, under the general title *Information technology — MPEG audio technologies*:

- *Part 1: MPEG Surround*
- *Part 2: Spatial Audio Object Coding (SAOC)*
- *Part 3: Unified speech and audio coding*

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 part of ISO/IEC 23003 Unified Speech and Audio Coding (USAC) is a new audio coding standard that 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, multi-media 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 patents.

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

The holder of this patent right has assured ISO and the IEC that he 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 the IEC. Information may be obtained from the companies listed in Annex G.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights other than those identified in Annex G. ISO and the 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 part of ISO/IEC 23003 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 part of ISO/IEC 23003 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 referenced documents are indispensable for the application of this document. For undated references, the latest edition of the referenced document (including any amendments) applies.

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

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

3 Terms, definitions, symbols and abbreviated terms

3.1 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 14496-3, ISO/IEC 23003-1 and the following apply.

3.1.1

algebraic codebook

fixed codebook where an algebraic code is used to populate the excitation vectors (innovation vectors)

NOTE The excitation contains a small number of nonzero pulses with predefined interlaced sets of potential positions. The amplitudes and positions of the pulses of the k th excitation codevector can be derived from its index k through a rule requiring no or minimal physical storage, in contrast with stochastic codebooks whereby the path from the index to the associated codevector involves look-up tables.