INTERNATIONAL STANDARD ISO/IEC 14496-3:2001
TECHNICAL CORRIGENDUM 1

Published 2002-12-01

Information technology — Coding of audio-visual objects —

Part 3:
Audio

TECHNICAL CORRIGENDUM 1

Technologies de l'information — Codage des objets audiovisuels —

Partie 3: Codage audio

RECTIFICATIF TECHNIQUE 1

Technical Corrigendum 1 to ISO/IEC 14496-3:2001 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information.*

ICS 35.040

Ref. No. ISO/IEC 14496-3:2001/Cor.1:2002(E)

Published in Switzerland

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Information technology — Coding of audio-visual objects —

Part 3:
Audio

TECHNICAL CORRIGENDUM 1

Throughout the text, replace the term

„
bit stream
"

with

„
bitstream
".

Throughout the text replace

„
scaleable
"

with

„
scalable
".

Throughout the text replace

„
indexes
"

with

„
indices
".

In clause (Introduction), replace

1

"

MPEG-4 has no standard for transport. In all of the MPEG-4 tools for audio and visual coding, the coding standard ends at the point of constructing a sequence of access units that contain the compressed data. The MPEG-4 Systems (ISO/IEC 14496-1:2001) specification describes how to convert the individually coded objects into a bitstream that contains a number of multiplexed sub-streams.

There is no standard mechanism for transport of this stream 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 14496-6:1999) 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 bitstream specification allows transmission functions that are much more sophisticated than are possible with previous MPEG standards.

"

with

"

MPEG-4 has no standard for transport. 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.

".

In clause (Introduction), replace multiplex, storage and transmission formats table with:

"

	Format	Functionality defined in MPEG-4:	Functionality originally defined in:	Description
Multiplex	FlexMux	ISO/IEC 14496-1 (normative)	-	Flexible 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 14496-1 (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()

".

Remove subclause 2.2 (Definitions).

Remove subclause 3.2 (Definitions).

Remove subclause 4.3 (GA-specific definitions).

Remove subclause 5.3 (Definitions).

Remove subclause 6.2 (Definitions).

Remove subclause 7.2 (Definitions).

Replace subclause 1.3 (Terms and Definitions) with the following:
"

1.   **AAC:** Advanced Audio Coding.

2.   **Audio access unit**: An individually accessible portion of audio data within an elementary stream.

3.   **Audio composition unit**: An individually accessible portion of the output that an audio decoder produces from audio access units.

4.   **Absolute time:** The time at which sound corresponding to a particular event is really created; time in the real-world. Contrast score time.

5.   **Actual parameter:** The expression which, upon evaluation, is passed to an opcode as a parameter value.

6.   **A-cycle:** See audio cycle.

7.   **Adaptive codebook:** An approach to encode the long-term periodicity of the signal. The entries of the codebook consists of overlapping segments of past excitations.

8.   **Alias:** Mirrored spectral component resulting from sampling.

9.   **Analysis filterbank:** Filterbank in the encoder that transforms a broadband PCM audio signal into a set of spectral coefficients.

10.   **Ancillary data:** Part of the bitstream that might be used for transmission of ancillary data.

11.   **API:** Application Programming Interface.

12.   **A-rate:** See audio rate.

13.   **Asig:** The lexical tag indicating an a-rate variable.

14.   **Audio buffer:** A buffer in the system target decoder (see ISO/IEC 13818-1) for storage of compressed audio data.

15.   **Audio cycle:** The sequence of processing which computes new values for all a-rate expressions in a particular code block.

16.   **Audio rate:** The rate type associated with a variable, expression or statement which may generate new values as often as the sampling rate.

17.   **Audio sample:** A short snippet or clip of digitally represented sound. Typically used in wavetable synthesis.

18.   **AudioBIFS:** The set of tools specified in ISO/IEC 14496-1 (MPEG-4 Systems) for the composition of audio data in interactive scenes.

19. **Authoring:** In Structured Audio, the combined processes of creatively composing music and sound control scripts, creating instruments which generate and alter sound, and encoding the instruments, control scripts, and audio samples in MPEG-4 Structured Audio format.

20. **Backus-Naur Format:** (BNF) A format for describing the syntax of programming languages, used here to specify the SAOL and SASL syntax.

21. **Backward compatibility:** A newer coding standard is backward compatible with an older coding standard if decoders designed to operate with the older coding standard are able to continue to operate by decoding all or part of a bitstream produced according to the newer coding standard.

22. **Bandwidth scalability:** The possibility to change the bandwidth of the signal during transmission.

23. **Bank:** A set of samples used together to define a particular sound or class of sounds with wavetable synthesis.

24. **Bark:** The Bark is the standard unit corresponding to one critical band width of human hearing.

25. **Beat:** The unit in which score time is measured.

26. **Bitrate:** The rate at which the compressed bitstream is delivered to the input of a decoder.

27. **Bitrate scalability:** The possibility to transmit a subset of the bitstream and still decode the bitstream with the same decoder.

28. **Bitstream verifier:** A process by which it is possible to test and verify that all the requirements specified in ISO/IEC 14496-3 are met by the bitstream.

29. **Bitstream; stream:** An ordered series of bits that forms the coded representation of the data.

30. **Block companding:** Normalizing of the digital representation of an audio signal within a certain time period.

31. **BNF:** See Backus-Naur Format.

32. **BSAC:** Bit Sliced Arithmetic Coding.

33. **Bus:** An area in memory which is used to pass the output of one instrument into the input of another.

34. **Byte:** Sequence of 8-bits.

35. **Byte aligned:** A bit in a data function is byte-aligned if its position is a multiple of 8-bits from the first bit of this data function.

36. **CELP:** Code Excited Linear Prediction.

37. **Center channel:** An audio presentation channel used to stabilize the central component of the frontal stereo image.

38. **Channel:** A sequence of data representing an audio signal intended to be reproduced at one listening position.

39. **Coded audio bitstream:** A coded representation of an audio signal.

40. **Coded representation:** A data element as represented in its encoded form.

41. **Composition (compositing):** Using a scene description to mix and combine several separate audio tracks into a single presentation.

42. **Compression:** Reduction in the number of bits used to represent an item of data.

43. **Constant bitrate:** Operation where the bitrate is constant from start to finish of the coded bitstream.

44. **Context:** See state space.

45. **Control:** An instruction used to describe how to use a particular synthesis method to produce sound.

    EXAMPLES:
    "Using the piano instrument, play middle C at medium volume for 2 seconds."
    "Glissando the violin instrument up to middle C."
    "Turn off the reverberation for 8 seconds."

46. **Control cycle:** The sequence of processing which computes new values for all control-rate expressions in a particular code block.

47. **Control period:** The length of time (typically measured in audio samples) corresponding to the control rate.

48. **Control rate:** (1) The rate at which instantiation and termination of instruments, parametric control of running instrument instances, sharing of global variables, and other non-sample-by-sample computation occurs in a particular orchestra. (2) The rate type of variables, expressions, and statements that can generate new values as often as the control rate.

49. **Core coder:** The term core coder is used to denote a base layer coder in certain scalability configurations. A core coder does not code the spectral samples of the MDCT filterbank of the subsequent AAC coding layers, but operates on a time domain signal. The output of the core decoder has to be up-sampled and transformed into the spectral domain, before it can be combined with the output of the AAC coding layers. Within the MPEG-4 Audio standard only the MPEG-4 CELP coder is a valid core coder. However, in principal, also another AAC coding layer, operating at a lower sampling rate, could be used on the time domain signal, and then combined with the other coding layer in exactly the same way as described for the CELP coder, and would therefore be called a core coder.

50. **CRC:** The Cyclic Redundancy Check to verify the correctness of data.

51. **Critical band:** This unit of bandwidth represents the standard unit of bandwidth expressed in human auditory terms, corresponding to a fixed length on the human cochlea. It is approximately equal to 100 Hz at low frequencies and 1/3 octave at higher frequencies, above approximately 700 Hz.

52. **Data element:** An item of data as represented after encoding and before decoding.

53. **Data function:** An encapsulation of data elements forming a logic unit.

54. **Decoded stream:** The decoded reconstruction of a compressed bitstream.

55. **Decoder:** An embodiment of a decoding process.

56. **Decoding (process):** The process that reads an input coded bitstream and outputs decoded audio samples.

57. **Demultiplexing:** Splitting one bitstream into several.

58. **DFT:** Discrete Fourier Transform.

59. **Digital storage media; DSM:** A digital storage or transmission device or system.

60. **Dimension conversion:** A method to convert a dimension of a vector by a combination of low pass filtering and linear interpolation.

61. **Discrete cosine transform; DCT:** Either the forward discrete cosine transform or the inverse discrete cosine transform. The DCT is an invertible, discrete orthogonal transformation.

62. **Downmix:** A matrixing of n channels to obtain less than n channels.

63. **Duration:** The amount of time between instantiation and termination of an instrument instance.

64. **Editing:** The process by which one or more coded bitstreams are manipulated to produce a new coded bitstream. Conforming edited bitstreams must meet the requirements defined in part 3 of ISO/IEC 14496.

65. **Elementary stream (ES):** A sequence of data that originates from a single producer in the transmitting MPEG-4 Terminal and terminates at a single recipient, e.g. an AVObject or a Control Entity in the receiving MPEG-4 Terminal. It flows through one FlexMux Channel.

66. **Encoder:** An embodiment of an encoding process.

67. **Encoding (process):** A process, not specified in ISO/IEC 14496, that reads a stream of input audio samples and produces a valid coded bitstream as defined in part 3 of ISO/IEC 14496.

68. **Enhancement layer(s):** The part(s) of the bitstream that is possible to drop in a transmission and still decode the bitstream.

69. **Entropy coding:** Variable length lossless coding of the digital representation of a signal to reduce statistical redundancy.

70. **Envelope:** A loudness-shaping function applied to a sound, or more generally, any function controlling a parametric aspect of a sound

71. **EP:** Error Protection

72. **ER:** Error resilience or Error Resilient (as appropriate)

73. **Event:** One control instruction.

74. **Excitation:** The excitation signal represents the input to the LPC module. The signal consists of contributions that cannot be covered by the LPC model.

75. **Expression:** A mathematical or functional combination of variable values, symbolic constants, and opcode calls.

76. **FFT:** Fast Fourier Transformation. A fast algorithm for performing a discrete Fourier transform (an orthogonal transform).

77. **Filterbank:** A set of band-pass filters covering the entire audio frequency range.

78. **Fine rate control:** The possibility to change the bitrate by, under some circumstances, skipping transmission of the LPC indices.

79. **Fixed codebook:** The fixed codebook contains excitation vectors for the speech synthesis filter. The contents of the codebook are non-adaptive (i.e. fixed).

80. **Flag:** A variable which can take one of only the two values defined in this specification.

81. **Formal parameter:** The syntactic element that gives a name to one of the parameters of an opcode.

82. **Forward compatibility:** A newer coding standard is forward compatible with an older coding standard if decoders designed to operate with the newer coding standard are able to decode bitstreams of the older coding standard.

83. **Frame**: A part of the audio signal that corresponds to a certain number of audio PCM samples.

84. **Fs:** Sampling frequency.

85. **FSS:** Frequency Selective Switch. Module which selects one of two input signals independently in each scalefactor band.

86. **Fundamental frequency:** A parameter which represents signal periodicity in frequency domain.

87. **Future wavetable:** A wavetable that is declared but not defined in the SAOL orchestra; its definition must arrive in the bitstream before it is used.

88. **Global block:** The section of the orchestra that describes global variables, route and send statements, sequence rules, and global parameters.

89. **Global context:** The state space used to hold values of global variables and wavetables.

90. **Global parameters:** The sampling rate, control rate, and number of input and output channels of audio associated with a particular orchestra.

91. **Global variable:** A variable that can be accessed and/or changed by several different instruments.

92. **Grammar:** A set of rules that describes the set of allowable sequences of lexical elements comprising a particular language.

93. **Guard expression:** The expression standing at the front of an if, while, or else statement that determines whether or how many times a particular block of code is executed.

94. **Hann window:** A time function applied sample-by-sample to a block of audio samples before Fourier transformation.

95. **Harmonic lines:** A set of spectral components having a common fundamental frequency.

96. **Harmonic magnitude:** Magnitude of each harmonic.

97. **Harmonic synthesis:** A method to obtain a periodic excitation from harmonic magnitudes.

98. **Harmonics:** Samples of frequency spectrum at multiples of the fundamental frequency.

99. **HCR:** Huffman codebook reordering

100. **HILN:** Harmonic and Individual Lines plus Noise (parametric audio coding).

101. **Huffman coding:** A specific method for entropy coding.

102. **HVXC:** Harmonic Vector eXcitation Coding (parametric speech coding).

103. **Hybrid filterbank:** A serial combination of subband filterbank and MDCT. Used in MPEG-1 and MPEG-2 Audio.

104. **I-cycle:** See initialisation cycle.

105. **IDCT:** Inverse Discrete Cosine Transform.

106. **Identifier:** A sequence of characters in a textual SAOL program that denotes a symbol.

107. **IFFT:** Inverse Fast Fourier Transform.

108. **IMDCT:** Inverse Modified Discrete Cosine Transform.

109. **Index:** Number indicating the quantized value(s).

110. **Individual line:** A spectral component described by frequency, amplitude and phase.

111. **Informative:** Aspects of a standards document that are provided to assist implementers, but are not required to be implemented in order for a particular system to be compliant to the standard.

112. **Initial phase:** A phase value at the onset of voiced signal in harmonic synthesis.

113. **Initialisation cycle:** See initialisation pass.

114. **Initialisation pass:** The sequence of processing that computes new values for each i-rate expression in a particular code block.

115. **Initialisation rate:** The rate type of variables, expressions, and statements that are set once at instrument instantiation and then do not change.

116. **Instance:** See instrument instantiation.

117. **Instantiation:** The process of creating a new instrument instantiation based on an event in the score or statement in the orchestra.

118. **Instrument:** An algorithm for parametric sound synthesis, described using SAOL. An instrument encapsulates all of the algorithms needed for one sound-generation element to be controlled with a score.

    NOTE - An MPEG-4 Structured Audio instrument does not necessarily correspond to a real-world instrument. A single instrument might be used to represent an entire violin section, or an ambient sound such as the wind. On the other hand, a single real-world instrument that produces many different timbres over its performance range might be represented using several SAOL instruments.

119. **Instrument instantiation:** The state space created as the result of executing a note-creation event with respect to a SAOL orchestra.

120. **Intensity stereo:** A method of exploiting stereo irrelevance or redundancy in stereophonic audio programmes based on retaining at high frequencies only the energy envelope of the right and left channels.

121. **Interframe prediction:** A method to predict a value in the current frame from values in the previous frames. Interframe prediction is used in VQ of LSP.

122. **International Phonetic Alphabet; IPA :** The worldwide agreed symbol set to represent various phonemes appearing in human speech.

123. **I-pass:** See initialisation pass.

124. **IPQF:** inverse polyphase quadrature filter.

125. **I-rate:** See initialisation rate.

126. **Ivar:** The lexical tag indicating an i-rate variable.

127. **Joint stereo coding:** Any method that exploits stereophonic irrelevance or stereophonic redundancy.

128. **Joint stereo mode:** A mode of the audio coding algorithm using joint stereo coding.

129. **K-cycle:** See control cycle.

130. **K-rate:** See control rate.

131. **Ksig:** The lexical tag indicating a k-rate variable.

132. **Lexical element:** See token.

133. **Lip shape pattern :** A number that specifies a particular pattern of the preclassified lip shape.

134. **Lip synchronization :** A functionality that synchronizes speech with corresponding lip shapes.

135. **Looping:** A typical method of wavetable synthesis. Loop points in an audio sample are located and the sound between those endpoints is played repeatedly while being simultaneously modified by envelopes, modulators, etc.

136. **Low frequency enhancement (LFE) channel:** A limited bandwidth channel for low frequency audio effects in a multichannel system.

137. **LPC:** Linear Predictive Coding.

138. **LPC residual signal:** A signal filtered by the LPC inverse filter, which has a flattened frequency spectrum.

139. **LPC synthesis filter:** An IIR filter whose coefficients are LPC coefficients. This filter models the time varying vocal tract.

140. **LSP:** Line Spectral Pairs.

141. **LTP:** Long Term Prediction.

142. **M/S stereo:** A method of removing imaging artifacts as well as exploiting stereo irrelevance or redundancy in stereophonic audio programs based on coding the sum and difference signal instead of the left and right channels.

143. **Main audio channels:** All single_channel_elements or channel_pair_elements in one program.

144. **Mapping:** Conversion of an audio signal from time to frequency domain by subband filtering and/or by MDCT.

145. **Masking:** A property of the human auditory system by which an audio signal cannot be perceived in the presence of another audio signal.

146. **Masking threshold:** A function in frequency and time below which an audio signal cannot be perceived by the human auditory system.

147. **MIDI:** The Musical Instrument Digital Interface standards. Certain aspects of the MPEG-4 Structured Audio tools provide interoperability with MIDI standards.

148. **Mixed voiced frame:** A speech segment which has both voiced and unvoiced components.

149. **Modified discrete cosine transform (MDCT):** A transform which has the property of time domain aliasing cancellation.

150. **Moving picture dubbing :** A functionality that assigns synthetic speech to the corresponding moving picture while utilizing lip shape pattern information for synchronization.

151. **MPE:** Multi Pulse Excitation.

152. **MPEG-4 Audio Text-to-Speech Decoder :** A device that produces synthesized speech by utilizing the M-TTS bitstream while supporting all the M-TTS functionalities such as speech synthesis for FA and MP dubbing.

153. **M-TTS sentence :** This defines the information such as prosody, gender, and age for only the corresponding sentence to be synthesized.

154. **M-TTS sequence :** This defines the control information which affects all M-TTS sentences that follow this M-TTS sequence.

155. **Multichannel:** A combination of audio channels used to create a spatial sound field.

156. **Multilingual:** A presentation of dialogue in more than one language.

157. **Multiplexing:** Combining several bitstreams into one.

158. **Natural Sound:** A sound created through recording from a real acoustic space. Contrasted with synthetic sound.

159. **NCC:** Number of Considered Channels. In case of AAC, it is the number of channels represented by the elements SCE, independently switched CCE and CPE, i.e. once the number of SCEs plus once the number of independently switched CCEs plus twice the number of CPEs. With respect to the naming conventions of

the MPEG-AAC decoders and payloads, NCC=A+I. This number is used to derive the required decoder input buffer size (see subclause 4.5.3.1). In case of other codecs, it tis the total number of channels.

160. **Noise component:** A signal component modeled as noise.

161. **Non-tonal component:** A noise-like component of an audio signal.

162. **Normative:** Those aspects of a standard that must be implemented in order for a particular system to be compliant to the standard.

163. **Nyquist sampling:** Sampling at or above twice the maximum bandwidth of a signal.

164. **OD:** Object Descriptor.

165. **Opcode:** A parametric signal-processing function that encapsulates a certain functionality so that it may be used by several instruments.

166. **Orchestra:** The set of sound-generation and sound-processing algorithms included in an MPEG-4 bitstream. Includes instruments, opcodes, routing, and global parameters.

167. **Orchestra cycle:** A complete pass through the orchestra, during which new instrument instantiations are created, expired ones are terminated, each instance receives one k-cycle and one control period worth of a-cycles, and output is produced.

168. **Padding:** A method to adjust the average length of an audio frame in time to the duration of the corresponding PCM samples, by conditionally adding a slot to the audio frame.

169. **Parameter:** A variable within the syntax of this specification which may take one of a range of values. A variable which can take one of only two values is a flag or indicator and not a parameter.

170. **Parameter fields:** The names given to the parameters to an instrument.

171. **Parser:** Functional stage of a decoder which extracts from a coded bitstream a series of bits representing coded elements.

172. **P-fields:** See parameter fields.

173. **Phoneme/bookmark-to-FAP converter :** A device that converts phoneme and bookmark information to FAPs.

174. **Pi:** The constant π = 3.14159...

175. **Pitch:** A parameter which represents signal periodicity in the time domain. It is expressed in terms of the number of samples.

176. **Pitch control:** A functionality to control the pitch of the synthesized speech signal without changing its speed.

177. **PNS:** Perceptual Noise Substitution.

178. **Polyphase filterbank:** A set of equal bandwidth filters with special phase interrelationships, allowing an efficient implementation of the filterbank.

179. **Postfilter:** A filter to enhance the perceptual quality of the synthesized speech signal.

180. **PQF:** polyphase quadrature filter

181. **Prediction:** The use of a predictor to provide an estimate of the sample value or data element currently being decoded.

182. **Prediction error:** The difference between the actual value of a sample or data element and its predictor.

183. **Predictor:** A linear combination of previously decoded sample values or data elements.

184. **Presentation channel:** An audio channel at the output of the decoder.

185. **Production rule:** In Backus-Naur Form grammars, a rule that describes how one syntactic element may be expressed in terms of other lexical and syntactic elements.

186. **Program:** A set of main audio channels, coupling_channel_elements (see subclause 4.5.2.1), lfe_channel_elements (see subclause 4.5.2.1) and associated data streams intended to be decoded and played back simultaneously. A program may be defined by default, or specifically by a program_config_element (see subclause 4.5.1.2). A given single_channel_element (see subclause 4.5.2.1), channel_pair_element (see subclause 4.5.2.1), coupling_channel_element, lfe_channel_element or data channel may accompany one or more programs in any given bitstream.

187. **PSNR:** Peak Signal to Noise Ratio.

188. **Psychoacoustic model:** A mathematical model of the masking behaviour of the human auditory system.

189. **Random access:** The process of beginning to read and decode the coded bitstream at an arbitrary point.

190. **Rate semantics:** The set of rules describing how rate types are assigned to variables, expressions, statements, and opcodes, and the normative restrictions that apply to a bitstream regarding combining these elements based on their rate types.

191. **Rate type:** The "speed of execution" associated with a particular variable, expression, statement, or opcode.

192. **Rate-mismatch error:** The condition that results when the rate semantics rules are violated in a particular SAOL construction. A type of syntax error.

193. **Reserved:** The term "reserved" when used in the clauses defining the coded bitstream indicates that the value may be used in the future for ISO/IEC defined extensions.

194. **Route statement:** A statement in the global block that describes how to place the output of a certain set of instruments onto a bus.

195. **RPE:** Regular Pulse Excitation.

196. **Run-time error:** The condition that results from improper calculations or memory accesses during execution of a SAOL orchestra.

197. **RVLC:** Reversible Variable Length Coding

198. **Sample:** See Audio sample.

199. **Sample Bank Format:** A component format of MPEG-4 Structured Audio that allows the description of a set of samples for use in wavetable synthesis and processing methods to apply to them.

200. **Sampling Frequency (Fs):** Defines the rate in Hertz which is used to digitize an audio signal during the sampling process.

201. **SAOL:** The Structured Audio Orchestra Language, pronounced like the English word "sail". SAOL is a digital-signal processing language that allows for the description of arbitrary synthesis and control algorithms as part of the content bitstream.

202. **SAOL orchestra:** See orchestra.

203. **SASBF:** The MPEG-4 Structured Audio Sample Bank Format, an efficient format for the transmission of blocks of wavetable (sample data) compatible with the MIDI method for the same.

204. **SASL:** The Structured Audio Score Language. SASL is a simple format that allows for powerful and flexible control of music and sound synthesis.

205. **SBA:** Segmented Binary Arithmetic Coding which is the error resilient tool for BSAC

206. **Scalefactor:** Factor by which a set of values is scaled before quantization.

207. **Scalefactor band:** A set of spectral coefficients which are scaled by one scalefactor.

208. **Scalefactor index:** A numerical code for a scalefactor.

209. **Scheduler:** The component of MPEG-4 Structured Audio that describes the mapping from control instructions to sound synthesis using the specified synthesis techniques. The scheduler description provides normative bounds on event-dispatch times and responses.

210. **Scope :** The code within which access to a particular variable name is allowed.

211. **Score:** A description in some format of the sequence of control parameters needed to generate a desired music composition or sound scene. In MPEG-4 Structured Audio, scores are described in SASL and/or MIDI.

212. **Score time:** The time at which an event happens in the score, measured in beats. Score time is mapped to absolute time by the current tempo.

213. **Semantics:** The rules describing what a particular instruction or bitstream element should do. Most aspects of bitstream and SAOL semantics are normative in MPEG-4.

214. **Send statement:** A statement in the global block that describes how to pass a bus on to an effect instrument for post-processing.

215. **Sequence rules:** The set of rules, both default and explicit, given in the global block that define in what order to execute instrument instantiations during an orchestra cycle.

216. **SIAQ**: Scalable Inverse AAC Quantization Module.

217. **Side information:** Information in the bitstream necessary for controlling the decoder.

218. **Signal variable:** A unit of memory, labelled with a name, that holds intermediate processing results. Each signal variable in MPEG-4 Structured Audio is instantaneously representable by a 32-bit floating point value.

219. **Sinusoidal synthesis:** A method to obtain a time domain waveform by a sum of amplitude modulated sinusoidal waveforms.

220. **Spatialisation:** The process of creating special sounds that a listener perceives as emanating from a particular direction.

221. **Spectral coefficients:** Discrete frequency domain data output from the analysis filterbank.

222. **Spectral envelope:** A set of harmonic magnitudes.

223. **Speed control:** A functionality to control the speed of the synthesized speech signal without changing its pitch or phonemes.

224. **Spreading function:** A function that describes the frequency spread of masking effects.

225. **SQ:** Scalar Quantization.

226. **State space:** A set of variable-value associations that define the current computational state of an instrument instantiation or opcode call. All the "current values" of the variables in an instrument or opcode call.

227. **Statement:** "One line" of a SAOL orchestra.

228. **Stereo-irrelevant:** A portion of a stereophonic audio signal which does not contribute to spatial perception.

229. **Structured audio:** Sound-description methods that make use of high-level models of sound generation and control. Typically involving synthesis description, structured audio techniques allow for ultra-low bitrate description of complex, high-quality sounds.

230. **Stuffing (bits); stuffing (bytes):** Code-words that may be inserted at particular locations in the coded bitstream that are discarded in the decoding process. Their purpose is to increase the bitrate of the stream which would otherwise be lower than the desired bitrate.

231. **Surround channel:** An audio presentation channel added to the front channels (L and R or L, R, and C) to enhance the spatial perception.

232. **Symbol:** A sequence of characters in a SAOL program, or a symbol token in a MPEG-4 Structured Audio bitstream, that represents a variable name, instrument name, opcode name, table name, bus name, etc.

233. **Symbol table:** In an MPEG-4 Structured Audio bitstream, a sequence of data that allows the tokenised representation of SAOL and SASL code to be converted back to a readable textual representation. The symbol table is an optional component.

234. **Symbolic constant:** A floating-point value explicitly represented as a sequence of characters in a textual SAOL orchestra, or as a token in a bitstream.

235. **Syncword:** A code embedded in audio transport bitstreams that identifies the start of a transport frame.

236. **Syntax:** The rules describing what a particular instruction or bitstream element should look like. All aspects of bitstream and SAOL syntax are normative in MPEG-4.

237. **Syntax error:** The condition that results when a bitstream element does not comply with its governing rules of syntax.

238. **Synthesis:** The process of creating sound based on algorithmic descriptions.

239. **Synthesis filterbank:** Filterbank in the decoder that reconstructs a PCM audio signal from subband samples.

240. **Synthetic Sound:** Sound created through synthesis.

241. **Tempo:** The scaling parameter that specifies the relationship between score time and absolute time. A tempo of 60 beats per minute means that the score time measured in beats is equivalent to the absolute time measured in seconds; higher numbers correspond to faster tempi, so that 120 beats per minute is twice as fast.

242. **Terminal:** The "client side" of an MPEG transaction; whatever hardware and software are necessary in a particular implementation to allow the capabilities described in this document.

243. **Termination:** The process of destroying an instrument instantiation when it is no longer needed.

244. **Text-to-speech synthesizer :** A device producing synthesized speech according to the input sentence character strings.

245. **Timbre:** The combined features of a sound that allow a listener to recognise such aspects as the type of instrument, manner of performance, manner of sound generation, etc. Those aspects of sound that distinguish sounds equivalent in pitch and loudness.

246. **TNS:** Temporal Noise Shaping

247. **Token:** A lexical element of a SAOL orchestra a keyword, punctuation mark, symbol name, or symbolic constant.

248. **Tokenisation:** The process of converting a orchestra in textual SAOL format into a bitstream representation consisting of a stream of tokens.