Ecma International

Standard ECMA-407

1st Edition / June 2014

Scalable Sparse Spatial Sound System (S5) – Base S5 Coding

Standard

Rue du Rhône 114 CH-1204 Geneva T: +41 22 849 6000 F: +41 22 849 6001

Contents

Page

Introduction

S5 denotes a scalable multichannel coding system for spatial audio data compression, which can be applied to provide 3D audio experience with little overhead. Such system may incorporate a wide range of state-of-the-art audio codecs and can be applied to provide 3D audio experience. By using an audio codec, which may offer encapsulation capacity for external data, S5 data may be carried within the audio coder stream with little overhead and maintain a compatible bit stream syntax.

This Standard specifies the base S5 encoder and decoder in terms of configuration data, downmix, inverse coding parameter data and upmix. It provides reference and guidance on how to incorporate further components to form a scalable multichannel coding system for audio data compression.

The base S5 codec achieves data compression of multichannel audio information by mapping the audio information on to a downmix signal and to sparse spatial data, which refers to the parameter values of a mathematical model to reconstruct localization and ambiance. A specific method, denoted as inverse coding, is used for upmixing from the audio downmix and its associated parameter values.

Compressing the downmix audio by a state-of-the-art audio codec will further increase the coding efficiency of S5. An overview is given on how the base S5 encoder/decoder may be extended by incorporation of an audio codec and other components; however, the components themselves and their interfaces are not specified.

Such specific S5 codecs are subject to separate standards, which share the base S5 coding standard as their common basis.

This Ecma Standard has been adopted by the General Assembly of June 2014.

Scalable Sparse Spatial Sound System (S5) – Base S5 Coding

1 Scope

This Standard specifies the base S5 encoder and decoder in terms of configuration data, downmix, inverse coding parameter data and upmix. In addition it provides reference and guidance on how to incorporate further components to form a scalable multichannel coding system for audio data compression.

2 Conformance

Conformant base S5 encoders generate dataflows as specified in Clauses 7, 8, and 9. Conformant base S5 decoders generate the upmix as specified in Clause 10 by processing dataflows as specified in Clauses 7, 8, and 9.

3 Normative references

ISO/IEC 23001-8, *Information technology -- MPEG systems technologies -- Part 8: Coding-independent code points*

IETF RFC 5234, *Augmented BNF for syntax specifications: ABNF*

4 Terms, definitions and acronyms

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

4.1
downmix
reduced number of audio channels from an input signal

4.2
upmix
increased number of audio channels from a downmix

4.3
base S5 encoder
encoding unit providing the downmix, the inverse coding parameter data and the upmix configuration

4.4
base S5 decoder
decoding unit providing the upmix based on the downmix, the inverse coding parameter data and the upmix configuration

4.5
base audio codec
audio codec component providing lossless or lossy compression and decompression of the downmix

4.6
loudness
perceived level of an audio programme

4.7
Mid (M) signal
non-directional input signal to a Mid-Side (MS) decoder

4.8
Side (S) signal
directional input signal to a Mid-Side (MS) decoder

4.9
uimsbf
unsigned integer, most significant bit first

4.10
Q format
fixed point binary format for fractional numbers, where the number of fractional bits and the number of integer bits is specified

4.11
uqmsbf
unsigned Q format most significant bit first

NOTE This Standard uses for **uqmsbf** the Q format notation Qm.n, where “m” designates the number of bits of the integer part and “n” denotes the number of bits of the fractional portion to the right of the binary point. The width “w” of the corresponding bitfield is w = m + n bits. The value range covers 0 to $2^{m} - 2^{-n}$ with a constant resolution of 2^{-n}. To convert a number from unsigned Q format to a decimal number take the Q bitfield as an integer and multiply it by 2^{-n}.

4.12
sqmsbf
signed Q format most significant bit first

NOTE This Standard uses for **sqmsbf** the 2’s complement with Q format notation Qm.n, where “m” designates the number of bits of the integer part without the sign bit and “n” the number of bits of the fractional portion to the right of the binary point. The width “w” of the corresponding bitfield is w = m + n + 1 bits, which includes the sign bit as most significant bit. The value range covers 2^{-m} to $2^{m} - 2^{-n}$ with a constant resolution of 2^{-n}. To convert a number from signed Q format to a decimal number take the Q bitfield as an 2’s complement integer and multiply it by 2^{-n}.

5 S5 Overview

S5 denotes a scalable multichannel coding system for spatial audio data compression, which can be applied to provide 3D audio experience with little overhead. Such a system may incorporate a wide range of state-of-the-art audio codecs and can be applied to provide a 3D audio experience. By using an audio codec, which may offer encapsulation capacity for external data, S5 data may be carried within the base audio coder stream with little overhead and maintain a compatible bit stream syntax.

The system of an S5 codec can be determined by the functional block diagrams of the S5 encoder, as depicted in Figure 1, and of the S5 decoder, as depicted in Figure 2. An S5 encoder shall at least consist of a base S5 encoder; and a S5 decoder shall at least consist of a base S5 decoder.

The base S5 encoder shall achieve compression of multichannel audio information by downmixing the f-channel signal to g channels and shall produce sparse spatial data, which is a parametric encoding of a mathematical model to reconstruct from the downmix an upmix having the localization and ambiance approaching that of the original signal. A specific method, denoted as inverse coding (see Clause 6), is used to construct an upmix of h channels from the audio downmix and its associated spatial data.

Compressing the downmix audio by a state-of-the-art base audio coder can further increase the coding efficiency of S5. The various bitstreams produced by the functional units of an S5 encoder may be encapsulated into a single bitstream by the functional unit 'Multiplexer' (see Annex F).

Ancillary data may be conveyed from the S5 encoder to the S5 decoder and may be used to encapsulate data other than coding parameters, for example, loudness parameters, which may be used to adjust the perceived level of audio signals. For the loudness parameters, see Annex D.

This Standard specifies the base S5 encoder/decoder and their interfaces only. All other components and their interfaces are not specified. Such specific S5 codecs are subject to separate standards. As the base S5 encoder/decoder is agnostic to the other system components, the base S5 coding standard shall represent the common base for all S5 specific standards.

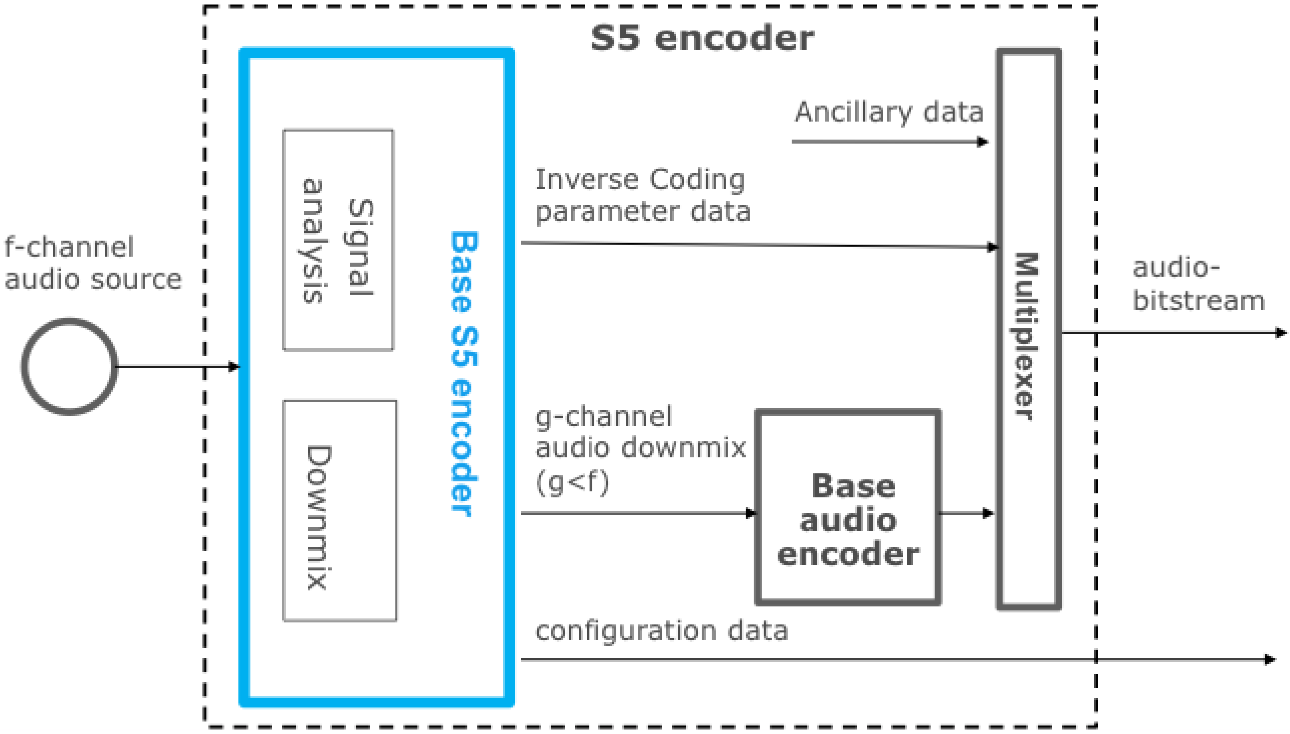

Figure 1 — Functional block diagram of the S5 encoder

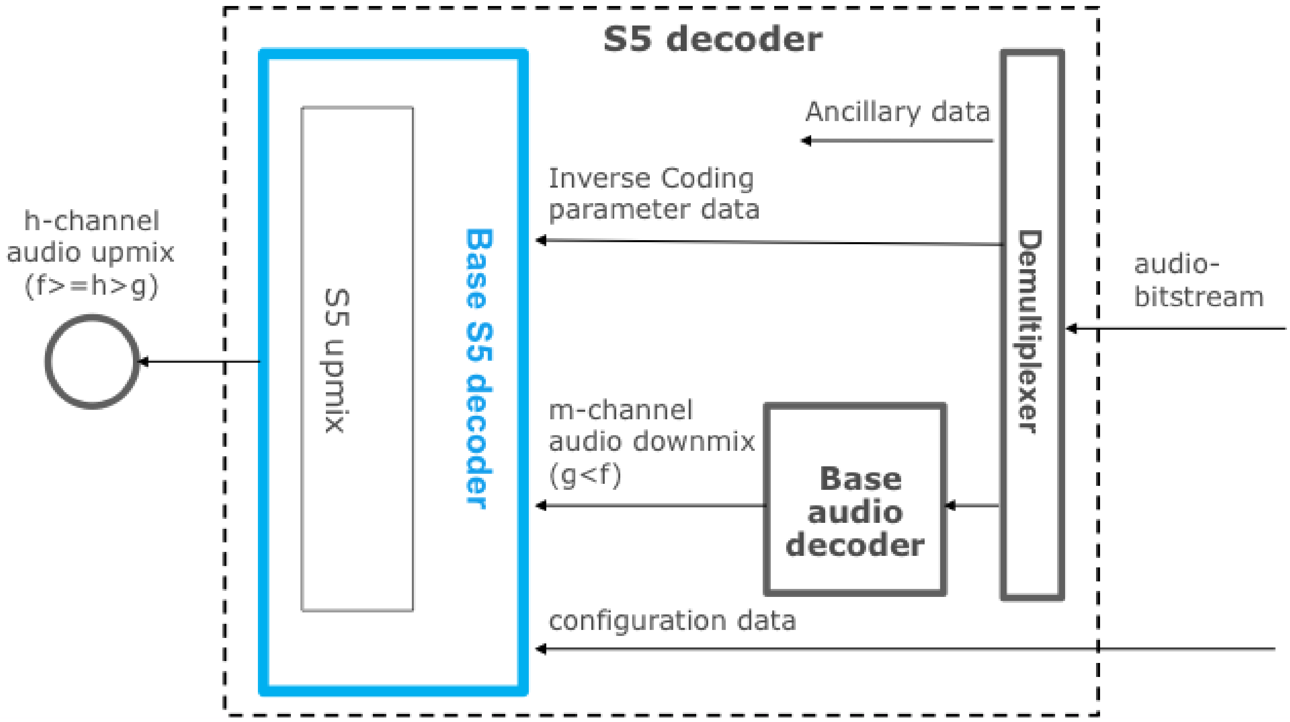

Figure 2 — Functional block-diagram of the S5 decoder

The subsequent clauses of this Standard specify the syntax of data streams by using Augmented Backus Naur Form (ABNF) as is defined in IETF RFC 5234. In addition to this notation, the code of the data stream elements is denoted by the format and the length of their bit fields. Note, that syntax and final encoding of a data stream are strictly separated. For the same syntax of a data stream, an external encoding e.g. by a multiplexer may vary according to the constraints of the storage or transmission environment. Examples are byte alignment or error protection. However, external encoding details are beyond the scope of this Standard and are subject to specific S5 standards or other specifications.

6 Inverse Coding

Inverse coding denotes a mathematical method for upmixing a channel-based audio signal while preserving to a high degree the localization and ambiance information of the audio source. Inverse coding is based on the spatial representation of a left and a right signal by a real-valued composite signal, the mid (M) signal, and a real-valued differential signal, the side (S) signal. A mid-side (MS) decoder maps without information loss the samples of the MS signals on to the left and the right channel. The mapping follows the equations below:

$$\mathrm{Left} = (\mathrm{M} + \mathrm{S}) * \frac{1}{\sqrt{2}}$$

$$\mathrm{Right} = (\mathrm{M} - \mathrm{S}) * \frac{1}{\sqrt{2}}$$

Inverse coding assumes that the S signal can be approximated by processing the M signal with two specific gains $P_\propto$, P_β and two specific delays $L'_\propto$, L'_β.

Figure 4 depicts inverse coding as a signal processing unit. The corresponding functions $L'_\propto$, L'_β, $P_\propto$, P_β and their parameters refer to Table 1:

Table 1 — Formulae of inverse coding gains and delays

Delay L'_α	$L'_\alpha = L_\alpha * \tau = \left(-\frac{f(\alpha)}{2\sin\alpha} + \sqrt{\frac{f^2(\alpha)}{4\sin^2\alpha} + f^2(\varphi) - \frac{f(\alpha) * f(\varphi) * \sin\varphi}{\sin\alpha}}\right) * \tau$
Delay L'_β	$L'_\beta = L_\beta * \tau = \left(-\frac{f(\beta)}{2\sin\beta} + \sqrt{\frac{f^2(\beta)}{4\sin^2\beta} + f^2(\varphi) + \frac{f(\beta) * f(\varphi) * \sin\varphi}{\sin\beta}}\right) * \tau$
Gain P_α	$P_\alpha = \frac{f^2(\alpha)}{4\sin^2\alpha} + f^2(\varphi) - \frac{f(\alpha) * f(\varphi) * \sin\varphi}{\sin\alpha}$
Gain P_β	$P_\beta = \frac{f^2(\beta)}{4\sin^2\beta} + f^2(\varphi) + \frac{f(\beta) * f(\varphi) * \sin\varphi}{\sin\beta}$

The discriminant relationship of these gains and delays induces sound source separation for sound sources, even at the same frequency. It relies on an inverse problem solution, which takes into account the directivity pattern together with angular assumptions with regard to the main angle of incidence and a left opening angle α and a right opening angle β. These parameters of such discriminant relationship define the overall sound stage of the resulting MS signal.

Figure 3 depicts this relationship for an M signal showing a cardioid polar pattern.

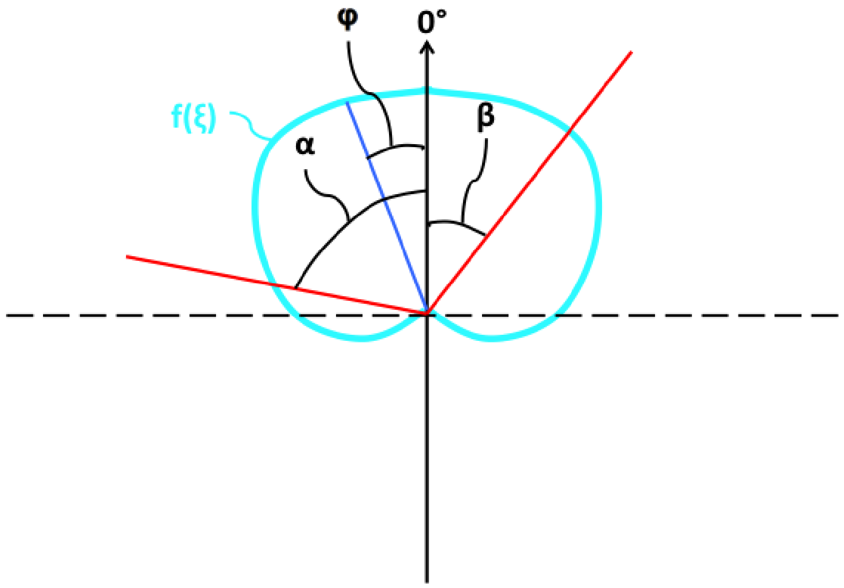

Figure 3 — Angular assumptions and directivity pattern with inverse coding

The following inverse coding parameters are applied:

the directivity pattern f(ξ)
is the ascertained polar diagram with

$$f(\xi) = 1 - \frac{n}{2} + \frac{n}{2} * \sin \xi$$

with $0 \leq \xi < 2\pi$ and $0 \leq n \leq 2$

the ascertained main angle of incidence φ
between sound source and the polar main axis of the directivity pattern (such polar axis corresponding with a microphone's main axis), with $-\pi/2 \leq \varphi \leq \pi/2$

the stipulated left opening angle α
adjoining the polar main axis of the directivity pattern on the left, with, $0 < \alpha \leq \pi/2$.
For a positive main angle of incidence φ, the condition $\varphi \leq \alpha$ shall be satisfied.

the stipulated right opening angle β
adjoining the polar main axis of the directivity pattern on the right, with $0 < \beta \leq \pi/2$.
For a negative main angle of incidence φ, the condition $| \varphi | \leq \beta$ shall be satisfied.

the time scaling factor τ
for generating a S-signal, with $0.029s \leq \tau \leq 0.146s$

the side signal ratio gain λ
to control the S-signal level, with, $0 \leq \lambda \leq 1$, leading to signals that may be seamlessly varied between a degree of correlation of –1 and +1.

These parameters are applied for inverse coding as shown by the signal processing circuit of Figure 4:

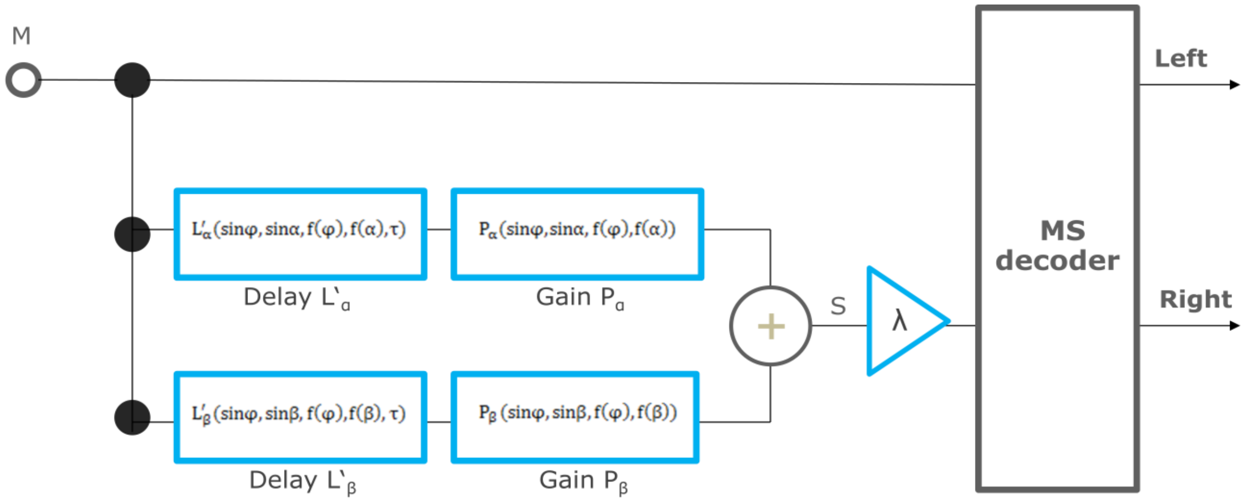

Figure 4 — Functional block diagram for an inverse coding function

According to Figure 4, the samples of left and right channels shall be derived from the previous equations as given in Table 2.

Table 2 — Formulae of inverse coding functions

$$\text{Left} = \frac{1}{\sqrt{2}} * (M + \lambda * (P_\alpha * \text{delay}(M, L'_\alpha) + P_\beta * \text{delay}(M, L'_\beta)))$$

$$\text{Right} = \frac{1}{\sqrt{2}} * (M - \lambda * (P_\alpha * \text{delay}(M, L'_\alpha) + P_\beta * \text{delay}(M, L'_\beta)))$$

Altogether, the inverse coding parameters should approximate a given spatial signal in the best way, as determined by the base S5 encoder. This means: delay(M, L'_α**)** and delay(M, L'_β), denoting the delays L'_α and L'_β, as applied to the M signal, shall be implemented in such way, which best approximates the values for their real-valued equivalent L'_α and L'_β. As the inverse coding parameters rely on an "inverse problem" as formulated by V. Ambartsumian [2], they may not be determined in a linear way by the base S5 encoder. Plural sets of parameter values may thus achieve the same desired perceptual spatial effect.

For instance, as may be learned from Figure 4, there is a trade-off between the side signal ratio gain λ and the gains P_α and P_β (which essentially relies on the stipulated left opening angle α and the stipulated right opening angle β, which will influence the overall choice of parameters). The base S5 encoder may therefore pick a specific inverse coding parameter set due to its algorithmic structure - without significant perceptual change of the overall system.

The function and parameter names used in this description of inverse coding are assigned to the S5 syntax elements of this Standard according to Table 3:

Table 3 — Assignment of inverse coding functions and parameters to S5 syntax elements

S5 syntax element	Inverse coding function / parameter
S5Left	Equation ‘Left’
S5Right	Equation ‘Right’
S5Directivity	Directivity pattern f(ξ), represented by parameter n
S5Incidence	Main angle of incidence φ
S5Alpha	Left opening angle α
S5Beta	Right opening angle β
S5Scale	Time scaling factor τ
S5Side	Side signal ratio gain λ

For multichannel upmixing, inverse coding considers a downmix channel or a previously processed signal as M signal and uses the equations ‘Left’ or ‘Right’ of Table 3 to generate an upmix. However, this coding principle may be applied in multiple ways. As the sum of the left and right signals creates the M signal, the system is robust with regard to colouration (and to comb filter effects, in case anew downmixing should take place).

7 Configuration Data

The configuration data shall consist of the downmix configuration ID (**S5DownmixConfig),** the output channel configuration ID (**S5ChannelConfig**) and a group of data elements which specify the upmix configuration (**S5UpmixConfig**).

7.1 Syntax of Configuration Data (S5Config)

Table 4 — Syntax of S5Config

Syntax	No. of bits	Format
S5Config =		
“S5ConfigID”	**4**	**uimsbf**
“S5SyncTagWindow”	**10**	**uimsbf**
“S5 SyncTagAccuracy”	**2**	**uimsbf**
“S5DownmixConfig”	**7**	**uimsbf**
“S5ChannelConfig”	**7**	**uimsbf**
S5UpmixConfig		

7.2 Configuration Identifier (S5ConfigID)

Each instance of **S5Config** is represented by its unique identifier **S5ConfigID**.This identifier is referred by S5 data streams to link data elements or group of data elements to its corresponding configuration data.

The value range covers decimal numbers from 0 to 15.

7.3 Window Size for the Calculation of Synchronization Tags (S5SyncTagWindow)

The value of **S5SyncTagWindow** sets the number of consecutive audio samples used for the calculation of synchronisation tags as specified in 8.2.

The value range covers decimal numbers from 64 to 1024.

7.4 Accuracy of the Calculation of Synchronization Tags (S5SyncTagAccuracy)

The value of **S5SyncTagAccuracy** sets the number of bits used in 8.2 for the integer representation of the energy values.

Table 5 — Code assignment for S5SyncTagAccuracy

"**00**"	**16 bit**
"**01**"	**24 bit**
"10"	**32 bit**
"11"	reserved

7.5 Downmix Configuration (S5DownmixConfig)

The **S5DownmixConfig** gives the channel configuration, which may be used to render the downmix channels on to loudspeaker positions whilst ignoring **S5ChannelConfig** and **S5UpmixConfig.**

The downmix configuration shall be identified by a 'Channel Configuration' value, as specified in Annex A.

7.6 Output Channel Configuration (S5ChannelConfig)

The **S5ChannelConfig** describes the reference loudspeaker arrangement for which the base S5 decoder upmix shall be intended.

The output channel configuration shall be identified by a 'Channel Configuration' value, as specified in Annex A.

7.7 Upmix Configuration (S5UpmixConfig)

The data elements of **S5UpmixConfig** instruct the base S5 decoder on how to combine the inverse coding functions for generating the upmix samples from the downmix. This is accomplished for each output channel by a mathematical expression using Polish Notation.

The **S5UpmixConfig** syntax shall adhere to the specification in Annex B.

8 Inverse Coding Parameter Data

The inverse coding parameter data shall comprise the data elements **S5ConfigID**, **S5SyncTag, S5SyncTag-1, S5SyncTag-2,** and **S5ParameterSetCount,** which may be followed by **S5ParameterSetID**, **S5ParameterSetType** and at least one group of data elements denoted as **S5ParameterSet**.

The inverse coding parameter set data shall be conveyed whenever initially determined or changed by the base S5 encoder. Synchronisation information for temporal synchronisation of the S5 base encoder with the S5 base decoder may be conveyed more frequently.

An inverse coding parameter data stream starts always with a unique identifier value which links the parameter data to the corresponding configuration data. This element **S5ConfigID** is specified in Clause 7.

8.1 Syntax of Inverse Coding Parameter Data (S5InvCodeData)

The inverse coding parameter data bitstream from the base S5 encoder to the base S5 decoder shall correspond to the syntax as specified in Table 6:

Table 6 — Syntax of S5InvCodeData

Syntax	No. of bits	Format
S5InvCodeData =		
“S5ConfigID”	**4**	**uimsbf**
“S5SyncTag”	**8**	**uimsbf**
“S5SyncTag-1”	**8**	**uimsbf**
“S5SyncTag-2”	**8**	**uimsbf**
“S5ParameterSetCount”	**8**	**uimsbf**
0*(		
“S5ParameterSetID”	**10**	**uimsbf**
“S5ParameterSetType”	**3**	**uimsbf**
S5ParameterSet		
)		

8.2 Synchronization Elements (S5SyncTag, S5SyncTag-1, S5SyncTag-2)

The inverse coding parameter data shall be applied to the downmix according to the synchronization element **S5SyncTag**, The value of the synchronization element **S5SyncTag** allows to re-establish synchronization between the downmix stream and the inverse coding parameter data inside the base S5 decoder. The bit field **S5SyncTag** shall denote the energy signature of the downmix in a time granularity corresponding to the duration of such non-overlapping window of **WindowLength** samples as set by the value of **S5SyncTagWindow** in the configuration data stream (7.3). The energy signature **S5SyncTag** shall be derived from the downmix by the following process:

For each of the downmix channels, the energy of its time domain signal over **WindowLength** samples is computed by accumulating the squared sample value of the DC-removed signal (can be calculated by subtracting the arithmetic mean of the signal).

The time domain values are integer values of type **uimsbf** where the number of bits is set in the configuration data stream by **S5SyncTagAccurancy** (7.4).

The sum of the energy of all downmix channels **TotalEnergy** is normalized by **WindowLength** and converted into a logarithmic dB representation.

$$\mathbf{dBTotalEnergy} = 10 * \log10(\max\frac{\mathbf{TotalEnergy} * 1024}{\mathbf{WindowLength} + 1}, 1)$$

and is quantized into an 8 bit unsigned integer by determining min(**dBTotalEnergy**, 255).

Whereas **S5SyncTag** refers to the window where the parameter set data should be applied, **S5SyncTag-1 and S5SyncTag-2** relate to its past two windows and shall use the same rules as given for **S5SyncTag.**

8.3 Number of Parameter Sets (S5ParameterSetCount)

The value of **S5ParameterSetCount** indicates the number of occurrences of the subsequent group of data elements including **S5ParameterSetID**, **S5ParameterType** and **S5ParameterSet.** The value range of **S5ParameterSetCount** covers decimal numbers from 0 to 255.

If **S5ParameterSetCount** is set to **“0”,** no parameter set data is conveyed.

8.4 Inverse Coding Parameter Data Set ID (S5ParameterSetID)

S5ParameterSetID carries an integer, which is used within all parameter data to label a parameter set with a unique address. This address must be consistent with the settings in **S5UpmixConfig (**see Annex B). It can take values from 0 to 1023.

8.5 Parameter Data Set Type (S5ParameterSetType)

S5ParameterSetType shall indicate by a 3-bit prefix code the type of content of **S5ParameterSet**. The 3-bit binary codes shall adhere to Table 7:

Table 7 — Code assignment for S5ParameterSetType

"**000**"	indicates an **S5AbrParameterSet**
"**001**"	indicates an **S5FullParameterSet**
"**010**" to "**111**"	are reserved for assignment by specific S5 standards

8.6 Inverse Coding Parameter Data Set (S5ParameterSet)

This Standard specifies two groups of inverse coding parameter data sets, the abbreviated inverse coding parameter data set **S5AbrParameterSet** and the full inverse coding parameter set **S5FullParameterSet**. In addition a placeholder for parameter sets, which are not specified in this Standard, is provided as a reserved item. This allows specific S5 standards to incorporate additional parameter sets, which are tailored for their needs. The occurrence of such a parameter set shall be indicated by setting **S5ParameterSetType** to one of the reserved codes of Table 7. The syntax of **S5ParameterSet** shall adhere to Table 8.

Table 8 — Syntax of S5ParameterSet

Syntax	No. of bits	Format
S5ParameterSet =		
"**S5SwapLeftRight"**	**1**	**uimsbf**
S5AbrParameterSet		
/S5FullParameterSet		
/"Reserved"		

S5SwapLeftRight instructs the decoder to swap the inverse coding equations S5Left and S5Right for this specific parameter set if its value is set to "1".

The syntax of the abbreviated inverse coding parameter data set **S5AbrParameterSet** shall adhere to Table 9.

Table 9 — Abbreviated inverse coding parameter data set

Syntax	No. of bits	Format
S5AbrParameterSet =		
"**S5Directivity"**	**Q1.7**	**uqmsbf**
"**S5Alpha"**	**Q1.9**	**uqmsbf**
"**S5Scale"**	**Q1.9**	**uqmsbf**
"**S5Side"**	**Q1.10**	**uqmsbf**

The syntax of the full inverse coding parameter data set **S5FullParameterSet** shall adhere to Table 10.

Table 10 — Full inverse coding parameter data set

Syntax	No. of bits	Format
S5FullParameterSet =		
"**S5Directivity"**	**Q1.7**	**uqmsbf**
"**S5Incidence"**	**Q1.10**	**sqm**sbf
"**S5Alpha"**	**Q1.9**	**uqmsbf**
"**S5Beta"**	**Q1.9**	**uqmsbf**
"**S5Scale"**	**Q1.9**	**uqmsbf**
"**S5Side"**	**Q1.10**	**uqmsbf**

9 Downmix

The base S5 encoder shall generate the g-channel downmix from the f-channel input signal according to the channel configuration number, which has been assigned to **S5DownmixConfig**.

10 Upmix

The base S5 decoder shall create the upmix from the downmix in accordance with the configuration data and the inverse coding parameter data streams.

10.1 Synchronization of Inverse Coding Parameter Data

A synchronization tag shall be calculated from the downmix by the base S5 decoder following the same procedure as given in cause 7.2 for **S5SyncTag**. In case this synchronization tag matches with the conveyed **S5SyncTag**, the associated inverse coding parameter data sets shall be applied to all samples of the corresponding downmix channels following the matching window. To verify the correctness of the synchronisation result, the preceding windows may be checked in addition by using the corresponding synchronization tags **S5SyncTag-1** or **S5SyncTag-2**.

10.2 Expanding of S5AbrParameterSet

For use with an inverse coding function, **S5AbrParameterSet** shall be appended with **S5Incidence** and **S5Beta** according to Table 11.

Table 11 — Values of S5Incidence and S5Beta in S5AbrParameterSet

S5Incidence	= 0
S5Beta	= **S5Alpha**

10.3 Default Values of Inverse Coding Parameter Data

In case of missing or erroneous inverse coding parameter data at the base S5 decoder, the default parameter values in Table 12 shall apply for any inverse coding function:

Table 12 — Default values of inverse coding parameter data

Element of **S5ParameterSet**	Default value
S5Directivity	0.00
S5Incidence	0.000
S5Alpha	0.349
S5Beta	0.349
S5Scale	0.100
S5Side	0.2500

10.4 Default values of S5UpmixConfig

In case an **S5UpmixConfig** is not available to the base S5 decoder, the default **S5UpmixConfig,** as given by Table 13, shall apply. In a first step all downmix channels shall sum up in order to form a composite signal. In a second step, the composite signal shall be normalized to 0dB. To reflect the channel configuration of this signal, **S5DownmixConfig** shall be set to "1", **S5ChannelConfig** shall be set to "5", and the inverse coding parameter data shall be set to the default values given in Table 12. The upmix is based on the downmix signal "C" and delivers the centre front speaker (C_{Upmix}), the left, right front speakers (L_{Upmix}, R_{Upmix}), and the left, right surround speakers (Ls_{Upmix}, Rs_{Upmix}).

Table 13 specifies the binary code of the default **S5UpmixConfig**. For better reading the corresponding syntactical elements and the upmix expressions are provided in additional columns.

Table 13 — Default binary coding, syntactical description and upmix formulae for S5UpmixConfig

Binary coding of **S5UpmixConfig**	Syntactical description of **S5UpmixConfig**	Upmix expressions	Upmix expressions in Polish Notation
0000101	**S5ChannelCount**		
0000010 **00000011** **010** **110** **001** **0001011010100001** **101** **0000010**	**S5OutputChannelPos** **S5ExpressionCount** **S5OperatorType** **S5OperandType** **S5NumberType** "0.7071" **S5OperandType** **S5InputChannelPos**	C_{Upmix} = 0.7071 * C	* 0.7071 C
0000000 **00000001** **100** **000** **0000000000** **0000010**	**S5OutputChannelPos** **S5ExpressionCount** **S5OperandType** **S5FunctionType (S5Left)** **S5ParameterSetID** **S5InputChannelPos**	L_{Upmix} = **S5Left**(C)	**S5Left** C
0000001 **00000001** **100** **001** **0000000000** **0000010**	**S5OutputChannelPos** **S5ExpressionCount** **S5OperandType** **S5FunctionType (S5Right)** **S5ParameterSetID** **S5InputChannelPos**	R_{Upmix} = **S5Right**(C)	**S5Right** C
0000100 **00000011** **010** **110** **001** **0001011010100001** **100** **000** **0000000000** **0000010**	**S5OutputChannelPos** **S5ExpressionCount** **S5OperatorType** **S5OperandType** **S5NumberType** "0.7071" **S5OperandType** **S5FunctionType (S5Left)** **S5ParameterSetID** **S5InputChannelPos**	Ls_{Upmix} = 0.7071 * **S5Left**(C)	* 0.7071 **S5Left** C
0000101 **00000011** **010** **110** **001** **0001011010100001** **100** **001** **0000000000** **0000010**	**S5OutputChannelPos** **S5ExpressionCount** **S5OperatorType** **S5OperandType** **S5NumberType** "0.7071" **S5OperandType** **S5FunctionType (S5Right)** **S5ParameterSetID** **S5InputChannelPos**	Rs_{Upmix} = 0.7071 * **S5Right**(C)	* 0.7071 **S5Right** C

Annex A
(normative)

Channel Positions and Configurations

The output channel position, the loudspeaker position, the channel configurations from "1" to "31" and the channel to speaker mapping shall adhere to Tables 7 and 8 in ISO/IEC 23001-8 *Information technology -- MPEG systems technologies -- Part 8: Coding-independent code points*.

Table 7 indicates the loudspeaker position in the 3D environment of the listener as used in Annex B by **S5OutputChannelPos** and **S5InputChannel**. In order to ease the understanding of loudspeaker positions, Table 7 also contains loudspeaker positions according to IEC 62574.

Table 8 specifies the channel configuration, as is used in Clause 8 by **S5DownmixConfig** and **S5UpmixConfig**. This defines the number of audio channels and their associated loudspeaker positions. The name, abbreviation, and general position of each loudspeaker can be deduced from Table 7 (and Figure 9).

For the purpose of this Standard the following normative extensions apply:

- 'Output Channel Position' values from "32" to "127" are reserved.
- 'Channel Configuration' values from "15" to "63" are reserved.
- 'Channel Configuration' values from "64" to "126" shall adhere to http://www.ecma-international.org/standards/ecma-407/registry.htm.
- 'Channel Configuration' value "127" indicates a non-renderable downmix.

Tables C.1, C.2 and Figure C.1 of Annex C in this Standard repeat the information in Table 7 and 8 and Figure 9 of ISO/IEC 23001-8 *Information technology -- MPEG systems technologies -- Part 8: Coding-independent code points.* This is for the convenience of the reader, and is for information only.

Annex B
(normative)

Syntax for S5UpmixConfig

Each upmix configuration shall be denoted by **S5UpmixConfig**. **S5UpmixConfig** instructs as part of the configuration data (see Clause 7) to the base S5 decoder on how to combine the inverse coding functions for generating the upmix samples from the downmix. This is accomplished for each output channel by a mathematical expression represented in Polish Notation.

In the following, the syntax of the coded representation is specified. An example on the generation of **S5UpmixConfig** is available in 10.4. The syntax of **S5UpmixConfig** shall adhere to the specification in Table B.1.

Table B.1 — Syntax of S5UpmixConfig

Syntax	No. of bits	Format
S5UpmixConfig =		
“S5ChannelCount”	**7**	**uimsbf**
1* (		
“**S5OutputChannelPos”**	**7**	**uimsbf**
“**S5ExpressionCount”**	**8**	**uimsbf**
1* (		
“**S5OperatorType”**	**3**	**uimsbf**
/ **S5Operand**		
)		
)		

S5ChannelCount carries the total number of output channels needed for a channel configuration according to the value of **S5ChannelConfig.** Its integer value sets the number of occurrences of the subsequent group of data elements. **S5ChannelCount** values cover the range from 1 to 127.

S5OutputChannelPos shall show the index number of the output channel as defined in ‘Channel Position’ of Annex A. **S5OutputChannelPos** covers the range from 0 to 127.

S5ExpressionCount carries the total number of operators and operands required to specify the upmix expression. Its integer value indicates the number of occurrences of the subsequent group of data elements. **S5ConfigCount** values cover the range from 1 to 255.

S5OperatorType denotes the required mathematical operation in the upmix expression by a 3-bit code. The binary codes shall adhere to the specification in Table B.2.

Table B.2 — Codes used by S5OperatorType

“**000**”	indicates addition
“**001**”	indicates subtraction
”**010**”	indicates multiplication
”**011**”	indicates division

S5Operand specifies the different types of an operand. The syntax is given in Table B.3.

Table B.3 — Syntax of S5Operand

Syntax	No. of bits	Format
S5Operand =		
“**S5OperandType”**	**3**	**uimsbf**
(		
S5Function		
/ **S5Sample**		
/ S5Number		
)		

S5OperandType is a 3-bit prefix code to indicate the type of the operand. Table B.4 specifies the assignment.

Table B.4 — Codes used by S5OperandType

“**100**”	indicates type “Function”
“**101**”	indicates type “Sample”
”**110**”	indicates type “Number”
“**111**”	reserved

S5Function indicates the use of an inverse coding function. The syntax to apply is given in Table B.5.

Table B.5 — Syntax of S5Function

Syntax	No. of bits	Format
S5Function =		
“**S5FunctionType”**	**3**	**uimsbf**
“**S5 ParameterSetID”**	**10**	**uimsbf**
“S5InputChannelPos”	**7**	**uimsbf**

S5FunctionType indicates the inverse coding function to apply by a 3 bits code according Table B.6.

Table B.6 — Codes used by S5FunctionType

“**000**”	indicates Inverse coding function **S5Left**
“**001**”	indicates Inverse coding function **S5Right**
“**010**” to “**100**”	reserved for use by external specifications

S5ParameterSetID corresponds to an integer, which is used as a unique address of the parameter sets within a parameter data stream. The values of **S5ParameterSetID** cover the range from 0 to 1023.

S5InputChannelPos shall show the index number of the input channel as defined in ‘Channel Position’ of Annex A. The values of **S5InputChannelPos** cover the range from 0 to 127.

S5Sample corresponds to an audio sample value, which is taken as input according to the sanctification in Table B.7.

Table B.7 — Syntax of S5Sample

Syntax	No. of bits	Format
S5Sample =		
“**S5InputChannelPos”**	**7**	**uimsbf**

S5Number denotes a fixed-point integer, which is taken as an operand for the computation of the upmix. Its syntax is given in Table B.8.

Table B.8 — Syntax of S5Number

Syntax	No. of bits	Format
S5Number =		
“**S5NumberType”**	**3**	**uimsbf**
“**Format1”**	**16**	**uimsbf**
/ **“Format2”**	**Q3.13**	**uqmsbf**
/ **“Format3”**	**Q3.12**	**sqmsbf**
/ “**Reserved”**		

S5NumberType indicates the format of the value provided by **S5Number** by a 3-bit prefix code according to Table B.9.

Table B.9 — Codes used by S5NumberType

“**000**”	indicates use of Format1
“001”	indicates use of Format2
“**010**”	indicates use of Format3
“**100**” to “**111**”	reserved for future extensions

Annex C
(informative)

Channel Configuration and Position Tables

Table C.1 illustrates the 'Output Channel Position', the speaker abbreviation and the loudspeaker position - indicating the loudspeaker position in the 3D environment of the listener. In order to ease the understanding of loudspeaker positions, Table C.1 also contains loudspeaker positions according to IEC 62574, which are listed here for information. Table C.1 gives informative guidance; however, is no normative reference with regard to ISO/IEC 23001-8 *Information technology -- MPEG systems technologies -- Part 8: Coding-independent code points.*

Table C.1 — Mapping of Channel position to loudspeaker position

Output Channel Position	**Loudspeaker position**		**Loudspeaker position according to IEC 62574 :2011**	
	Abbr.	**Name**	**Abbr.**	**Name**
0	L	Left front	FL	Front left
1	R	Right front	FR	Front right
2	C	Centre front	FC	Front centre
3	LFE	Low frequency enhancement	LFE1	Low frequency effects-1
4	Ls	Left surround	LS	Left surround
5	Rs	Right surround	RS	Right surround
6	Lc	Left front centre	FLc	Front left centre
7	Rc	Right front centre	FRc	Front right centre
8	Lsr	Rear surround left	BL	Back left
9	Rsr	Rear surround right	BR	Back right
10	Cs	Rear centre	BC	Back centre
11	Lsd	Left surround direct	LSd	Left surround direct
12	Rsd	Right surround direct	RSd	Right surround direct
13	Lss	Left side surround	SL	Side left
14	Rss	Right side surround	SR	Side right
15	Lw	Left wide front	FLw	Front left wide
16	Rw	Right wide front	FRw	Front right wide
17	Lv	Left front vertical height	TpFL	Top front left
18	Rv	Right front vertical height	TpFR	Top front right
19	Cv	Centre front vertical height	TpFC	Top front centre
20	Lvr	Left surround vertical height rear	TpBL	Top back left
21	Rvr	Right surround vertical height rear	TpBR	Top back right
22	Cvr	Centre vertical height rear	TpBC	Top back centre
23	Lvss	Left vertical height side surround	TpSiL	Top side left
24	Rvss	Right vertical height side surround	TpSiR	Top side right
25	Ts	Top centre surround	TpC	Top centre
26	LFE2	Low frequency enhancement 2	LFE2	Low frequency effects-2
27	Lb	Left front vertical bottom	BtFL	Bottom front left
28	Rb	Right front vertical bottom	BtFR	Bottom front right
29	Cb	Centre front vertical bottom	BtFC	Bottom front centre
30	Lvs	Left vertical height surround	TpLS	Top left surround
31	Rvs	Right vertical height surround	TpRS	Top right surround
32-127		Reserved		Reserved

Figure C.1 below informatively illustrates the loudspeaker position in the 3D environment relative to the listener, with each labelled with an abbreviation from Table C.1. Loudspeakers lying on the innermost box are in the bottom level, those on the middle box are in the middle level and those on the outermost box are in the top level. The circles labelled Ts represent the listener position:

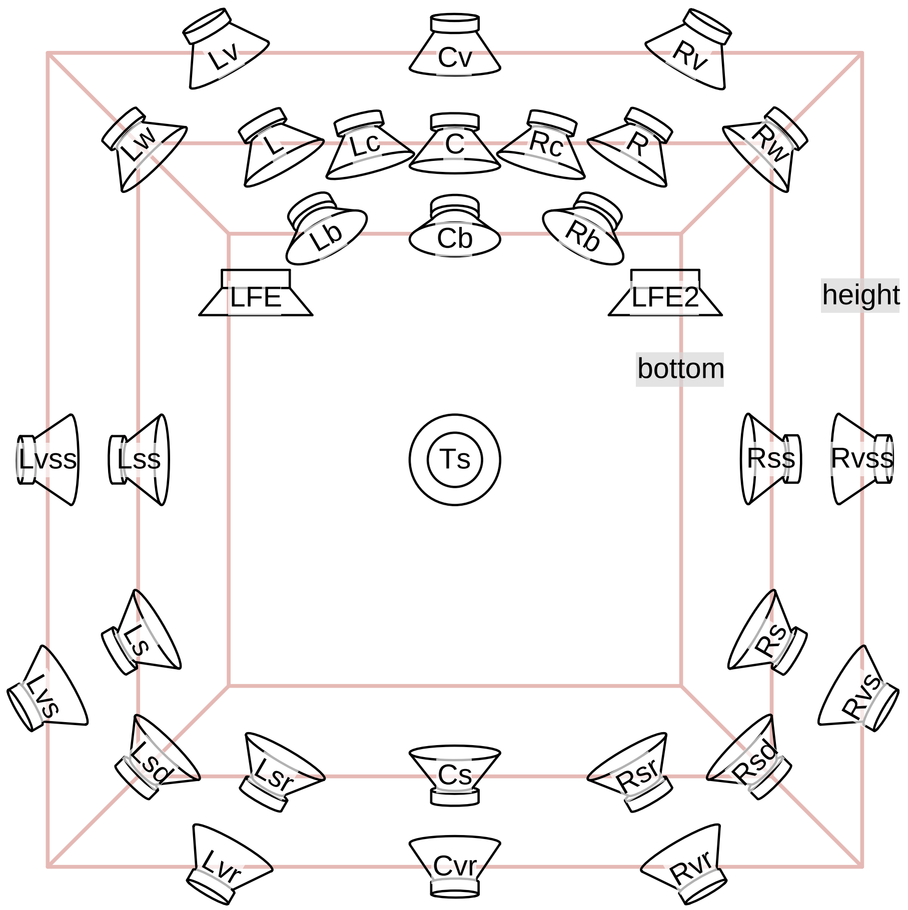

Figure C.1 — Loudspeaker position in the 3D environment

Table C.2 provides information about 'Channel Configuration', the assigned speaker positions, and their abbreviations in accordance with Figure C.1. In addition it includes the notation for indicating the involved number of audio channels and their associated loudspeaker positions. Note that Table C.2 gives informative guidance only and is no normative reference with regard to ISO/IEC 23001-8 *Information technology -- MPEG systems technologies -- Part 8: Coding-independent code points.*

Table C.2 — Channel configuration, channel to speaker mapping, and speaker abbreviation

Channel Configuration	Channel to speaker mapping	Speaker abbreviation	"Front/ Surr. LFE" notation
1	centre front speaker	C	1/0.0
2	left, right front speakers	L, R	2/0.0
3	centre front speaker, left, right front speakers	C L, R	3/0.0
4	centre front speaker, left, right front speakers, rear centre speaker	C L, R Cs	3/1.0
5	centre front speaker, left, right front speakers, left surround, right surround speakers	C L, R Ls, Rs	3/2.0
6	centre front speaker, left, right front speakers, left surround, right surround speakers, low frequency enhancement speaker	C L, R Ls, Rs LFE	3/2.1
7	centre front speaker, left, right front centre speakers, left, right front speakers, left surround, right surround speakers, low frequency enhancement speaker	C Lc, Rc L, R Ls, Rs LFE	5/2.1
8	channel1 channel2	N.A. N.A.	1+1
9	left, right front speakers, rear centre speaker	L, R Cs	2/1.0
10	left, right front speaker, left surround, right surround speakers,	L, R Ls, Rs	2/2.0
11	centre front speaker, left, right front speakers, left surround, right surround speakers, rear centre speaker, low frequency enhancement speaker	C L, R Ls, Rs Cs LFE	3/3.1
12	centre front speaker, left, right front speakers, left surround, right surround speakers, rear surround left, right speakers, low frequency enhancement speaker	C L, R Ls, Rs Lsr, Rsr LFE	3/4.1
13	centre front speaker, left, right front centre speakers, left, right front speakers, left, right side surround speakers, rear left, right surround speakers, rear centre speaker, left front low frequency enhancement speaker, right front low frequency enhancement speaker, centre front vertical height speaker, left, right front vertical height speakers, left, right vertical height side surround speakers, top centre surround speaker, left, right surround vertical height rear speakers, centre vertical height rear speaker, centre front vertical bottom speaker, left, right front vertical bottom speakers	C Lc, Rc L, R Lss, Rss Lsr, Rsr Cs LFE LFE2 Cv Lv, Rv Lvss, Rvss Ts Lvr, Rvr Cvr Cb Lb, Rb	11/11.2
14	centre front speaker, left, right front speakers, left surround, right surround speakers, low frequency enhancement speaker, left, right front vertical height speakers	C L, R Ls, Rs LFE Lv, Rv	5/2.1
15-63	Reserved	Reserved	-
64-126		See NOTE	
127		Non-renderable downmix	

NOTE See registry at http://www.ecma-international.org/standards/ecma-407/registry.htm

Annex D
(informative)

Loudness Adjustment

The subjective loudness of audio programmes, as perceived by the listener, should be described by four measurements according to ITU-R BS.1770-3:2012 *Algorithms to measure audio programme loudness and true-peak audio level,* namely

Programme Loudness
the integrated loudness of the whole programme

Max True Peak
the maximum peak value measured after 4x oversampling

Max Momentary Loudness
the maximum loudness value for one block of 400ms

Max Short Term Loudness
the maximum loudness value for a 3 seconds sliding window

and one optional additional measurement according to EBU Tech3342 *Loudness Range, a descriptor to supplement loudness normalisation in accordance with EBU R-128,* namely

Loudness Range
the measure that reflects the overall dynamic range of the programme.

The upmix generated by the base S5 decoder may be adjusted according to these measurements, which may be conveyed as ancillary data from the S5 encoder to the S5 decoder together with any descriptive data, which may go along with the loudness measurements.

Annex E
(informative)

Multiplexing

The various bitstreams produced by the functional elements of an S5 encoder may be either transmitted separately or be multiplexed into a single logical bitstream. The corresponding optional functional elements of the S5 encoder in Figure 1 and S5 decoder in Figure 2 are referred to as 'Multiplexer' and 'Demultiplexer'.

A common way of multiplexing bitstreams (audio and video) with other data into a single transport stream may be the use of container formats, e.g. Ogg (see S. Pfeiffer. The Ogg Encapsulation Format Version 0, IETF *RFC 3533*, May 2003).

Such formats generate high-level media codec streams, which provide framing, error protection and random access structure.

If the S5 data should be transported within the data stream of a audio codec, another approach may be more appropriate: some audio-coding standards are built upon a 'core and extension' architecture, which provides a single, unified bit stream syntax for the core and extension technology to encapsulate external data. Examples are the MPEG coders HE-AAC (see ISO/IEC 14496-3 *Information technology – Coding of audio-visual objects – Part 3: Audio)* and USAC (see SO/IEC 23003-3 *Information technology -- MPEG audio technologies -- Part 3: Unified speech and audio coding)*. These mechanisms can be used as a container to carry S5 data within the data stream of the base audio codec with little overhead while maintaining compatible bit stream syntax.

Bibliography

[1] C. Par. Two Undiscovered Treasures for Ground-breaking 3D Audio Coding Technology at Lowest Bitrates: Inverse Problems and Invariant Theory. Proceedings of 27th VDT International Convention, 11/2012, ISBN 978-3-9812830-3-7.

[2] R. V. Ambartsumian, ed. A Life in Astrophysics: Selected Papers of Viktor Ambartsumian. Allerton Press, 1998.