

---

---

**Information technology — Multimedia  
content description interface —**

**Part 4:  
Audio**

**AMENDMENT 2: High-level descriptors**

iTeh STANDARD PREVIEW  
(standards.iteh.ai)

*Technologies de l'information — Interface de description du contenu  
multimédia —*

*ISO/IEC 15938-4:2002/Amd 2:2006*

*Partie 4: Audio*

[https://standards.iteh.ai/catalog/standards/sist/4662c125-224d-4e64-a41e-](https://standards.iteh.ai/catalog/standards/sist/4662c125-224d-4e64-a41e-b07f7d67207d/iso-15938-4-2002-amd-2-2006)

[b07f7d67207d/iso-15938-4-2002-amd-2-2006](https://standards.iteh.ai/catalog/standards/sist/4662c125-224d-4e64-a41e-b07f7d67207d/iso-15938-4-2002-amd-2-2006)  
**AMENDEMENT 2: Descripteurs de haut niveau**

**PDF disclaimer**

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

[ISO/IEC 15938-4:2002/Amd 2:2006](https://standards.iteh.ai/catalog/standards/sist/4662c125-224d-4e64-a41e-b07f7d673fe8/iso-iec-15938-4-2002-amd-2-2006)

<https://standards.iteh.ai/catalog/standards/sist/4662c125-224d-4e64-a41e-b07f7d673fe8/iso-iec-15938-4-2002-amd-2-2006>

© ISO/IEC 2006

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

Published in Switzerland

## 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.

Amendment 2 to ISO/IEC 15938-4:2002 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

## iTeh STANDARD PREVIEW (standards.iteh.ai)

[ISO/IEC 15938-4:2002/Amd 2:2006](https://standards.iteh.ai/catalog/standards/sist/4662c125-224d-4e64-a41e-b07f7d673fe8/iso-iec-15938-4-2002-amd-2-2006)

<https://standards.iteh.ai/catalog/standards/sist/4662c125-224d-4e64-a41e-b07f7d673fe8/iso-iec-15938-4-2002-amd-2-2006>

## **iTeh STANDARD PREVIEW** **(standards.iteh.ai)**

ISO/IEC 15938-4:2002/Amd 2:2006

<https://standards.iteh.ai/catalog/standards/sist/4662c125-224d-4e64-a41e-b07f7d673fe8/iso-iec-15938-4-2002-amd-2-2006>

# Information technology — Multimedia content description interface —

## Part 4: Audio

### AMENDMENT 2: High-level descriptors

Remove subclauses 5.2.3. and add following subclauses:

#### 5.2.3 SeriesOfScalarType

This descriptor represents a series of scalars, at full resolution or scaled. Use this type within descriptor definitions to represent a series of feature values.

##### 5.2.3.1 Syntax

```

<!-- ##### (standards.iteh.ai) ##### -->
<!-- Definition of SeriesOfScalar datatype -->
<!-- ##### -->
<complexType name="SeriesOfScalarType">
  <complexContent>
    <extension base="mpeg7:ScalableSeriesType">
      <sequence>
        <element name="Raw" type="mpeg7:floatVector" minOccurs="0"/>
        <element name="Min" type="mpeg7:floatVector" minOccurs="0"/>
        <element name="Max" type="mpeg7:floatVector" minOccurs="0"/>
        <element name="Mean" type="mpeg7:floatVector" minOccurs="0"/>
        <element name="Random" type="mpeg7:floatVector" minOccurs="0"/>
        <element name="First" type="mpeg7:floatVector" minOccurs="0"/>
        <element name="Last" type="mpeg7:floatVector" minOccurs="0"/>
        <element name="Variance" type="mpeg7:floatVector" minOccurs="0"/>
        <element name="Weight" type="mpeg7:floatVector" minOccurs="0"/>
        <element name="LogBase" type="float" default="1.0" minOccurs="0" />
      </sequence>
    </extension>
  </complexContent>
</complexType>

```

##### 5.2.3.2 Semantics

Name	Definition
SeriesOfScalarType	A representation of a series of scalar values of a feature.
Raw	Series of unscaled samples (full resolution). Use only if scaling is absent to indicate the entire series.

Min	Series of minima of groups of samples. The value of <code>numOfElements</code> shall equal the length of the vector. This element shall be absent or empty if the <code>Raw</code> element is present.
Max	Series of maxima of groups of samples. The value of <code>numOfElements</code> shall equal the length of the vector. This element shall be absent or empty if the <code>Raw</code> element is present.
Mean	Series of means of groups of samples. The value of <code>numOfElements</code> shall equal the length of the vector. This element shall be absent or empty if the <code>Raw</code> element is present.
Random	Downsampled series (one sample selected at random from each group of samples). The value of <code>numOfElements</code> shall equal the length of the vector. This element shall be absent or empty if the <code>Raw</code> element is present.
First	Downsampled series (first sample selected from each group of samples). The value of <code>numOfElements</code> shall equal the length of the vector. This element shall be absent or empty if the <code>Raw</code> element is present.
Last	Downsampled series (last sample selected from each group of samples). The value of <code>numOfElements</code> shall equal the length of the vector. This element shall be absent or empty if the <code>Raw</code> element is present.
Variance	Series of variances of groups of samples. The value of <code>numOfElements</code> shall equal the length of the vector. This element shall be absent or empty if the <code>Raw</code> element is present. Mean must be present in order for <code>Variance</code> to be present.
Weight	Optional series of weights. Contrary to other fields, these do not represent values of the descriptor itself, but rather auxiliary weights to control scaling (see below). The value of <code>numOfElements</code> shall equal the length of the vector.
LogBase	In the case, its value is different to the default value, a logarithm has to be performed on the input data, before calculating any series (mean, variance...). The value is the base of a logarithm that is performed on the input data. Note that the value of <code>LogBase</code> must be greater than 0.

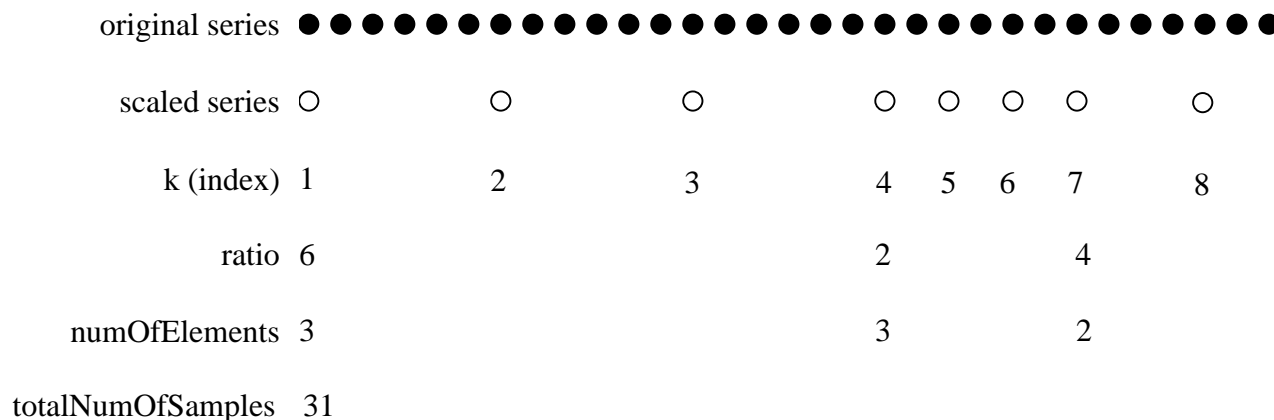
Note: Data of a full resolution series (`ratio` = 1) are stored in the `Raw` field. Accompanying zero-sized fields (such as `Mean`) indicate how the series may be scaled, if the need for scaling arises. The data are then stored in the scaled field(s) and the `Raw` field disappears.

In the case, that the value of `LogBase` is different from its default value, a logarithm must be performed on any input data before series calculation. In case, it is equal to the default value, no logarithm must be performed on the input data. The following formula shows the rule for this calculation. *Base* contains the base of the logarithm and is defined in `LogBase`. In case the logarithmic calculation is invalid (for e.g.  $\log 0$ ) or the calculated output is smaller than  $-1.0e^2$  the output value is fixed to  $-1.0e^2$ .

$$outputValue = \log_{base}(inputValue)$$

Scalable Series allow data to be stored at reduced resolution, according to a number of possible scaling operations. The allowable operations are those that are *scalable* in the following sense. Suppose the original series is scaled by a scale ratio of *P*, and this scaled series is then rescaled by a factor of *Q*. The result is the same as if the original series had been scaled by a scale ratio of  $N = PQ$ .

Figure AMD2.1 illustrates the scalability property. This scaled series can be derived indifferently from the original series by applying the scaling operation with the `ratios` shown, or from the scaled Series of Figure AMD2.1 by applying the appropriate rescaling operation. The result is identical. Scaling operations are chosen among those for which this property can be enforced.



**Figure AMD2.1 — An illustration of the scalability property**

If the scaling operations are used, they shall be computed as follows.

Name	Definition	Definition if Weight present
Min	$m_k = \min_{i=1+(k-1)N}^{kN} x_i$	Ignore samples with zero weight. If all have zero weight, set to zero by convention.
Max	$M_k = \max_{i=1+(k-1)N}^{kN} x_i$	Ignore samples with zero weight. If all have zero weight, set to zero by convention.
Mean	$\bar{x}_k = (1/N) \sum_{i=1+(k-1)N}^{kN} x_i$	$\bar{x}_k = \sum_{i=1+(k-1)N}^{kN} w_i x_i / \sum_{i=1+(k-1)N}^{kN} w_i$ If all samples have zero weight, set to zero by convention.
Random	choose at random among N samples	Choose at random with probabilities proportional to weights. If all samples have zero weight, set to zero by convention.
First	choose the first of N samples	Choose first non-zero-weight sample. If all samples have zero weight, set to zero by convention.
Last	choose the last of N samples	Choose last non-zero-weight sample. If all samples have zero weight, set to zero by convention.
Variance	$z_k = (1/N) \sum_{i=1+(k-1)N}^{kN} (x_i - \bar{x}_k)^2$ $= (1/N) \sum_{i=1+(k-1)N}^{kN} x_i^2 - \bar{x}_k^2$	$z_k = \sum_{i=1+(k-1)N}^{kN} w_i (x_i - \bar{x}_k)^2 / \sum_{i=1+(k-1)N}^{kN} w_i$ If all samples have zero weight, set to zero by convention.
Weight	$\bar{w}_k = (1/N) \sum_{i=1+(k-1)N}^{kN} w_i$	

In these formulae,  $k$  is an index in the scaled series, and  $i$  an index in the original series.  $N$  is the number of samples summarized by each scaled sample. In case  $\logBase$  is not equal to the default value,  $X$  is the logarithm of the input data, otherwise the raw input data. The formula for `Variance` differs from the standard formula for unbiased variance by the presence of  $N$  rather than  $N - 1$ . Unbiased variance is easy to derive from it. If the `Weight` field is present, the terms of all sums are weighted.

Replace subclause 5.2.5 with the following:

## 5.2.5 SeriesOfVectorType

This descriptor represents a series of vectors.

### 5.2.5.1 Syntax

```
<!-- ##### -->
<!-- Definition of SeriesOfVector datatype -->
<!-- ##### -->
<complexType name="SeriesOfVectorType">
  <complexContent>
    <extension base="mpeg7:ScalableSeriesType">
      <sequence>
        <element name="Raw" type="mpeg7:FloatMatrixType" minOccurs="0"/>
        <element name="Min" type="mpeg7:FloatMatrixType" minOccurs="0"/>
        <element name="Max" type="mpeg7:FloatMatrixType" minOccurs="0"/>
        <element name="Mean" type="mpeg7:FloatMatrixType" minOccurs="0"/>
        <element name="Random" type="mpeg7:FloatMatrixType" minOccurs="0"/>
        <element name="First" type="mpeg7:FloatMatrixType" minOccurs="0"/>
        <element name="Last" type="mpeg7:FloatMatrixType" minOccurs="0"/>
        <element name="Variance" type="mpeg7:FloatMatrixType" minOccurs="0"/>
        <element name="LogBase" type="float" default="1.0" minOccurs="0" />
        <element name="Covariance" type="mpeg7:FloatMatrixType"
          minOccurs="0"/>
        <element name="VarianceSummed" type="mpeg7:floatVector"
          minOccurs="0"/>
        <element name="MaxSqDist" type="mpeg7:floatVector" minOccurs="0"/>
        <element name="Weight" type="mpeg7:floatVector" minOccurs="0"/>
      </sequence>
      <attribute name="vectorSize" type="positiveInteger" default="1"/>
    </extension>
  </complexContent>
</complexType>
```

### 5.2.5.2 Semantics

Name	Definition
<code>SeriesOfVectorType</code>	A type for scaled series of vectors.
<code>Raw</code>	Series of unscaled samples (full resolution). Use only if <code>ratio=1</code> for the entire series.
<code>Min</code>	Series of minima of groups of samples. Number of rows must equal <code>numOfElements</code> , number of columns must equal <code>vectorSize</code> . This element must be absent or empty if the element <code>Raw</code> is present.
<code>Max</code>	Series of maxima of groups of samples. Number of rows must equal <code>numOfElements</code> , number of columns must equal <code>vectorSize</code> . This element must be absent or empty if the element <code>Raw</code> is present.



Name	Definition
Mean	Series of means of groups of samples. Number of rows must equal <code>numOfElements</code> , number of columns must equal <code>vectorSize</code> . This element must be absent or empty if the element <code>Raw</code> is present.
Random	Downsampled series (one sample selected at random from each group of samples). Number of rows must equal <code>numOfElements</code> , number of columns must equal <code>vectorSize</code> . This element must be absent or empty if the element <code>Raw</code> is present.
First	Downsampled series (first sample selected from each group of samples). Number of rows must equal <code>numOfElements</code> , number of columns must equal <code>vectorSize</code> . This element must be absent or empty if the element <code>Raw</code> is present.
Last	Downsampled series (last sample selected from each group of samples). Number of rows must equal <code>numOfElements</code> , number of columns must equal <code>vectorSize</code> . This element must be absent or empty if the element <code>Raw</code> is present.
Variance	Series of variance vectors of groups of vector samples. Number of rows must equal <code>numOfElements</code> , number of columns must equal <code>vectorSize</code> . This element must be absent or empty if the element <code>Raw</code> is present. <code>Mean</code> must be present in order for <code>Variance</code> to be present.
LogBase	Base of a logarithm that is performed on the input data. If the value is equal the default value, no logarithm is performed. Note that the value of <code>LogBase</code> must be greater than 0.
Covariance	Series of covariance matrices of groups of vector samples. This is a three-dimensional matrix. Number of rows must equal <code>numOfElements</code> , number of columns and number of pages must both equal <code>vectorSize</code> . This element must be absent or empty if the element <code>Raw</code> is present. <code>Mean</code> must be present in order for <code>Covariance</code> to be present.
VarianceSummed	Series of summed variance coefficients of groups of samples. Size of the vector must equal <code>numOfElements</code> . This element must be absent or empty if the element <code>Raw</code> is present. <code>Mean</code> must be present in order for <code>VarianceSummed</code> to be present.
MaxSqDist	Maximum Squared Distance (MSD). Series of coefficients representing an upper bound of the distance between groups of samples and their mean. Size of array must equal <code>numOfElements</code> . This element must be absent or empty if the element <code>Raw</code> is present. If <code>MaxSqDist</code> is present, <code>Mean</code> must also be present.
Weight	Optional series of weights. Weights control downsampling of other fields (see explanation for <code>SeriesOfScalars</code> ). Size of array must equal <code>numOfElements</code> .
<code>vectorSize</code>	The number of elements of each vector within the series.

Most of the above operations are straightforward extensions of operations previously defined in section 5.2.3.2 for series of scalars, applied uniformly to each dimension of the vectors. Operations that are specific to vectors are defined here:

Name	Definition	Definition if weight present
Covariance	$\sigma_k^{jj'} = \frac{1}{N} \sum_{i=1+(k-1)N}^{kN} (x_i^j - \bar{x}^j)(x_i^{j'} - \bar{x}^{j'})$	$\sigma_k^{jj'} = \frac{\sum_{i=1+(k-1)N}^{kN} w_i (x_i^j - \bar{x}^j)(x_i^{j'} - \bar{x}^{j'})}{\sum_{i=1+(k-1)N}^{kN} w_i}$
VarianceSummed	$z_k = (1/N) \sum_{j=1}^D \sum_{i=1+(k-1)N}^{kN} (x_i^j - \bar{x}_i^j)^2$	$z_k = \frac{\sum_{j=1}^D \sum_{i=1+(k-1)N}^{kN} w_i (x_i^j - \bar{x}_i^j)^2}{\sum_{i=1+(k-1)N}^{kN} w_i}$ If all samples have zero weight, set to zero by convention.
MaxSqDist	$MSD_k = \max_{i=1+(k-1)N}^{kN} \ x_i - \bar{x}_k\ ^2$	Ignore samples with zero weight. If all samples have zero weight, set to zero by convention

In these formulae,  $k$  is an index in the scaled series and  $i$  an index in the original series.  $N$  is the number of vectors summarized by each scaled vector.  $D$  is the size of each vector and  $j$  is an index into each vector.

In the case, that the value of  $LogBase$  is different from the default value, a logarithm must be performed on any input data before series calculation. In case, it is equal to the default value, no logarithm must be performed on the input data. The following formula shows the rule for this calculation.  $Base$  contains the base of the logarithm and is defined in  $LogBase$ . In case the logarithmic calculation is invalid (for e.g.  $\log 0$ ) or the calculated output is smaller than  $-1.0e^2$  the output value is fixed to  $-1.0e^2$ .

$$outputValue = \log_{base}(inputValue)$$

In case  $logBase$  is equal to the default value,  $\bar{x}_i^j$  is the mean of  $N$  samples and  $X$  are the raw input data.

The various variance/covariance options offer a choice of several cost/performance tradeoffs for the representation of variability.

Add at the end of subclause 6.8.3.3.3:

## 6.9 Rhythmic Pattern

### 6.9.1 RhythmicBaseType

This base descriptor contains a description of one single rhythmic pattern. The pattern is represented in a way that the parts of the bar are sorted in order of their importance. This is based on the fact, that the importance decreases with the order of the prime index. Regarding any further classification or matching of rhythmic patterns, this representation allows a setting of several grades in the resolution. Therefore for every pattern the most compact representation can be provided, resulting in an efficient comparison of the patterns and minimal memory needed for the storage.

### 6.9.1.1 Syntax

```

<!-- ##### -->
<!-- Definition of RhythmicBaseD -->
<!-- ##### -->
<complexType name="RhythmicBaseType">
  <complexContent>
    <extension base="mpeg7:AudioDType">
      <sequence>
        <element name="PrimeIndex" type="mpeg7:integerVector" minOccurs="1"
          maxOccurs="1"/>
        <element name="Velocity" type="mpeg7:integerVector" minOccurs="1"
          maxOccurs="1"/>
      </sequence>
    </extension>
  </complexContent>
</complexType>

```

### 6.9.1.2 Semantics

Name	Definition
RhythmicBaseType	The RhythmicBaseType contains elements of a single rhythmic pattern with different degrees of resolution.
PrimeIndex	The Integer vector indicating the initial index of the rhythmic pattern.
Velocity	The Integer vector indicating the velocity of the elements.

<https://standards.iteh.ai/catalog/standards/sist/4662c125-224d-4e64-a41e-b07f7d673fe8/iso-iec-15938-4-2002-amd-2-2006>

### 6.9.1.3 Usage

The RhythmicBaseType contains elements of a single rhythmic pattern with different degrees of resolution (i.e. on any different hierarchic levels). A representation of a rhythmic pattern requires indexing of the rhythmic grid with respect to the rhythmic significance of the grid position.

The calculation of each *PrimeIndex* of the example pattern may be done in the following manner:

1. Vector of prime factorization of the top part of the meter:

$\text{nomVec} = \{ \text{nom}_1 \dots \text{nom}_k \}$  sorted by size with the largest value first

2. Vector of prime factorization of the number of divisions per beat (tick):

$\text{mtVec} = \{ \text{mt}_1 \dots \text{mt}_k \}$

3. calculation of the prime indices from the grid positions:

```

patternLength = product(nomVec) * product(mtVec)
                ( product: multiply each value within the vector)
vector primeVec() ( initialized with the size of patternLength )
primeVec() = 0    ( set all Elements of the vector to 0 )
primeIndex = 1
primeProduct = 1
count = 0

for i=1 : length(nomVec)
{

```

```

primeProduct *= nomVec(i)
while (count < patternLength)
{
    if (((count/(patternLength/primeProduct)) modulo 1 ) == 0)
    {
        if (primeVec[count]==0)
        {
            primeVec[count] = primeIndex;
            primeIndex++;
        }
    }
    count++;
}
}
for i=1 : length(mtVec)
{
    primeProduct *= mtVec (i)
    while (count < patternLength)
    {
        if (((count/(patternLength/primeProduct)) modulo 1 ) == 0)
        {
            if (primeVec[count]==0)
            {
                primeVec[count] = primeIndex;
                primeIndex++;
            }
        }
        count++;
    }
}

```

**iTeh STANDARD PREVIEW**  
(standards.iteh.ai)

<https://standards.iteh.ai/catalog/standards/sist/4662c125-224d-4e64-a41e-b07d0751c810/iso-iec-15938-4-2002-amd-2-2006>

The successive prime factorization of the nominator and the micro time is necessary, because a joint prime factorization of the maximum number of elements can lead to comparisons of patterns with different time signature (but same length) in cases of reduced rhythmic resolution.

**Example 1:** meter = 4/4; micro time = 2; resulting size = 4 \* 2 = 8 ;

The following table shows a rhythmic pattern with a binary feeling notated as commonly done in a score-like representation:

part of the bar	1	1+	2	2+	3	3+	4	4+
grid position	1	2	3	4	5	6	7	8
prime index	1	5	3	6	2	7	4	8
velocity	100	0	112	0	150	68	120	0

No elements will be applied for any part of the bar with velocity equal to zero:

prime index	1	3	2	7	4
velocity	100	112	150	68	120

According to the ascending order of the prime indexes the elements will be rearranged, resulting in the final representation:

prime index	1	2	3	4	7
velocity	100	150	112	120	68

**Example 2:** meter = 4/4; micro time = 3; resulting size =  $4 * 3 = 12$  ;

The following table shows a rhythmic pattern with a ternary feeling notated as commonly done in a score-like representation:

part of the bar	1	1+	1++	2	2+	2++	3	3+	3++	4	4+	4++
grid position	1	2	3	4	5	6	7	8	9	10	11	12
prime index	1	5	6	3	7	8	2	9	10	4	11	12
velocity	180	0	100	200	0	99	190	0	97	205	0	101

No elements will be applied for any part of the bar with velocity equal to zero.

prime index	1	6	3	8	2	10	4	12
velocity	180	100	200	99	190	97	205	101

According to the ascending order of the prime indexes the elements will be rearranged, resulting in the final representation:

grid position	1	2	3	4	6	8	10	12
velocity	180	190	200	205	100	99	97	101

The following example demonstrates how two representations exhibiting different grades of resolution can be easily compared to each other. Only the number of elements of the shorter representation is taken into account. It is advantageous to compare only patterns with similar meter.

Pattern 1: meter: 4/4;

grid position	1	2	3	4	7
velocity	100	150	112	120	68

Pattern2: meter: 4/4;

prime index	1	2
velocity	120	145

The examples demonstrate how the use of a specific order allows the specification of a rhythmic hierarchy without any additional information.

## 6.9.2 AudioRhythmicPatternDS

The AudioRhythmicPatternType provides a more comprehensive description of a rhythmical structure of a whole song. The internal structure of the representation is dependent on the underlying rhythmical structure of the pattern that has been defined in RhythmicPatternType. Additionally to the rhythmic information of the pattern, this descriptor contains meter, instrument information, number of recurrences and segments as well.

### 6.9.2.1 Syntax of AudioRhythmicPatternType

```
<complexType name="AudioRhythmicPatternDS">
  <complexContent>
    <extension base="mpeg7:AudioDSType">
      <sequence>
        <element name="SinglePattern">
          <complexType>
            <sequence>
              <element name="Instrument" type="mpeg7:CreationToolType"/>
              <element name="Recurrences">
                <simpleType>
                  <restriction base="integer">
                    <minInclusive value="0"/>
                  </restriction>
                </simpleType>
              </element>
              <element name="RhythmPattern" type="mpeg7:RhythmicBaseType"/>
              <element name="Meter" type="mpeg7:MeterType"/>
              <element name="AudioSegment" type="mpeg7:AudioSegmentType"/>
            </sequence>
          </complexType>
        </element>
      </sequence>
    </extension>
  </complexContent>
</complexType>
```

<https://standards.iteh.ai/catalog/standards/sist/4662c125-224d-4e64-a41e-b07f7d673fe8/iso-iec-15938-4-2002-amd-2-2006>  
 (standards.iteh.ai)

### 6.9.2.2 Semantics

Name	Definition
AudioRhythmicPatternDS	A description scheme providing a compact and efficient representation of rhythmical patterns.
SinglePattern	Element that describes one single rhythmic pattern.
Instrument	Describes the devices/procedure and settings used for the creation of the metadata, such as the tools used to extract the metadata or the extraction parameters. Instrument is of type CreationToolType. Musical instruments should be only drum instruments and they must have been previously defined.
Recurrences	The number of recurrences of the same pattern.
RhythmPattern	The rhythmical pattern of an audio signal of <i>PatternType</i> data.
Meter	The meter of the music pattern.
AudioSegment	The time when the pattern starts and the duration of it.

### 6.9.2.3 Instantiation requirements

In order to guarantee a proper instantiation of this description scheme, the following requirements have to be fulfilled:

- The number of elements of *Velocity* and *PrimeIndex*, as provided by *RhythmicBaseType* must be equal and not 0.

### 6.9.2.4 Usage

The *AudioRhythmicPattern* DS aggregates rhythmic information from different fields of the song. Single patterns have been specified in *RhythmicBaseType*. To define a drum instrument that plays the actual pattern the attribute *Instrument* has been introduced. When using the *CreationToolType* for specifying an instrument it must be assured, that only drum instruments have been used. *Recurrences* describes the number of times the same pattern is played consecutively. To describe the currently played pattern *RhythmPattern* must be used. The start of the rhythmic pattern and its length (including the number of recurrences) must be indicated by *AudioSegment*.

The first step of extracting rhythmic patterns from a polyphonic music signal would be to transcribe percussive instruments. A commonly used technique is template matching by performing a differentiation, a half-way-rectification and a Principal Component Analysis on the input data to find spectral characteristics of un-pitched percussive instruments and to transcribe the actual drum tracks.

After transcribing the input data and obtaining the drum tracks, the actual pattern could be calculated. At first, the audio signal might be segmented into similar and characteristic regions using a self-similarity method. The segmentation is motivated by the assumption, that within each region not more than one representative drum pattern occurs, and that the rhythmic features are nearly invariant. Subsequently, the temporal positions of the events are quantized on a tatum grid. The term tatum grid refers to the pulse series on the lowest metric level. Tatum period and phase is computed by means of a two-way mismatch error procedure. The pattern length might be estimated by searching for the prominent periodicity in the quantized score with periods equal to an integer multiple of the bar length. The periodicity function is obtained by calculating a similarity measure between the signal and its time shifted version. The similarity between two score representations is calculated as weighted sum of the number of simultaneously occurring notes and rests in the score. The pattern is calculated by means of a histogram representation measuring the occurrence of notes on each metrical position within the pattern for each instrument. By comparing the histogram values with an arbitrary threshold the pattern elements are chosen as frequently occurring notes.

### 6.9.2.5 Applications

#### 6.9.2.5.1 Automatic Retrieval and Recommendation

When using rhythmic patterns to refer to a particular musical style or genre it is possible to query for musical content that is also characterized by one or more representative rhythmic patterns. This mechanism can serve as a search criterion in applications proposing a number of musical titles belonging to a particular musical style or genre or sounding similar to a title being proposed by the user as a "reference sound".

## 6.10 Chord Pattern

### 6.10.1 ChordBaseType

This descriptor contains a description of a single musical chord. A chord is a musical structure formed when at least two tones having different notes are played simultaneously. If the default ISO/IEC 15938-4 scale descriptor is used, numbers 0 to 11 indicate the half tones in the scale. The representation of the *ChordBaseType* element consists of a *RelativeChordNumber*, which represents the base note of a chord relative to the active key, the type of triad, and additional tones. For example, a *RelativeChordNumber* of 4, in the Key of C, represents a base note of E because it is four half tones higher than the base key.