INTERNATIONAL STANDARD

ISO/IEC 14496-3

Fourth edition
2009-09-01

AMENDMENT 5
2015-08-01

Information technology — Coding of audio-visual objects —

Part 3: **Audio**

AMENDMENT 5: Support for Dynamic Range Control, New Levels for ALS Simple Profile, and Audio Synchronization

Technologies de l'information — Codage des objets audiovisuels —

Partie 3: Codage audio

AMENDEMENT 5: Aide pour le contrôle de plage dynamique, nouveaux niveaux pour profil simple ALS et synchronisation audio

Reference number
ISO/IEC 14496-3:2009/Amd.5:2015(E)

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

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.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

Information technology — Coding of audio-visual objects —

Part 3:
Audio

AMENDMENT 5: Support for Dynamic Range Control, New Levels for ALS Simple Profile, and Audio Synchronization

1 Changes to the text of ISO/IEC 14496-3:2009

After 0.3.8.4, add:

0.3.9 Audio Synchronization Tool

The audio synchronization tool provides capability of synchronizing multiple contents in multiple devices. Synchronization is done by using audio features (fingerprint) extracted from the content. Neither common clock covering the multiple devices nor way to exchange time-stamps between the devices is required.

In the cover page of Part 3: Audio, replace:

This part of ISO/IEC 14496 contains twelve subparts:

with

This part of ISO/IEC 14496 contains thirteen subparts:

In the cover page of Part 3: Audio, add:

Subpart 13: Audio Synchronization

after

Subpart 12: Scalable lossless coding

In 1.3 Terms and Definitions, add:

1.3.z **Audio Sync**: Audio feature for synchronization

and increase the index-number of subsequent entries

In 1.5.1.1 Audio object type definition, amend Table 1.1 with the updates in the table below:

Object type ID	Audio object type	Gain control	[...]	Remark
0	Null			
[..]	[...]			
43	SAOC			
44	LD MPEG Surround			
45	SAOC-DE			
46	Audio Sync			
47 to 95	(reserved)			

After 1.5.1.2.40 add the following new subclauses:

1.5.1.2.41 Audio Sync object type

The Audio Sync object type conveys audio feature for multiple media stream synchronization (see ISO/IEC 14496-3 Subpart 13) in the MPEG-4 Audio framework.

In 1.5.2.1 (Profiles), Table 1.3 (Audio Profiles definition), add:

Object type ID	Audio object type	...
...	...	...
43	SAOC	
44	LD MPEG Surround	
45	SAOC-DE	
46	Audio Sync	

In 1.5.2.3 (Levels within the profiles), replace Table 1.13B and notes with:

— **Levels for the ALS Simple Profile**

Table 1.13B — Level for the ALS Simple Profile

Level	Max. number of channels	Max. sampling rate [kHz]	Max. word length [bit]	Max. number of samples per frame	Max. prediction order	Max. BS* stages	Max. MCC** stages
1	2	48	16	4096	15	3	1
2	2	48	24	4096	15	3	1
3	6	48	16	4096	15	3	1
4	6	48	24	4096	15	3	1

* BS: Block switching, ** MCC: Multi-channel coding

The BGMC tool and the RLS-LMS tool are not permitted. Floating-point audio data is not supported.

Insert the following new entries into Table 1.14 "audioProfileLevelIndication values" and adapt the "reserved for ISO use" range accordingly:

0x58	SAOC Dialogue Enhancement Profile	L1
0x59	SAOC Dialogue Enhancement Profile	L2
0x5A	ALS Simple Profile	L2
0x5B	ALS Simple Profile	L3
0x5C	ALS Simple Profile	L4
0x5D to 0x7F	reserved for ISO use	—

In 1.6.2.1 extend Table 1.15 "AudioSpecificConfig()"as follows:

Table 1.15 — Syntax of AudioSpecificConfig()

Syntax	No. of bits	Mnemonic
AudioSpecificConfig()		
{		
…		
switch (audioObjectType) {		
case 1:		
case 2:		
…		
case 43:		
saocPresentFlag = 1:		
saocPayloadEmbedding	**1**	**uimsbf**
SaocSpecificConfig():		
break;		
case 44:		
ldmpsPresentFlag = 1:		
ldsacPayloadEmbedding	**1**	**uimsbf**
LDSpatialSpecificConfig():		
break;		
case 45:		
saocDepresentFlag = 1:		
saocDePayloadEmbedding	**1**	**uimsbf**
SaocDeSpecificConfig():		
break;		
case 46:		
AudioSyncFeatureSpecificConfig():		
break;		
default:		
/* reserved */		
}		
}		

After 1.6.2.1.20 add the new subclause as follows:

1.6.2.1.21 AudioSyncFeatureSpecificConfig

Defined in ISO/IEC 14496-3 Subpart 13.

In 1.6.2.2.1 extend Table 1.17 "Audio Object Types" as follows:

Table 1.17 — Audio Object Types

Object type ID	Audio object type	Definition of elementary stream payloads and detailed syntax	Mapping of audio payloads to access units and elementary streams
0	NULL		
…			
43	SAOC	ISO/IEC 23003-2	
44	LD MPEG Surround	ISO/IEC 23003-2	
45	SAOC-DE	ISO/IEC 23003-2:2010/Amd.3	
46	Audio Sync	ISO/IEC 14496-3 Subpart 13	

In Table 4.57 add:

Table 4.57 — Syntax of extension_payload()

Syntax	No. of bits	Mnemonic
extension_payload(cnt)		
{		
extension_type;	**4**	**uimsbf**
align = 4;		
switch(extension_type) {		
case EXT_DYNAMIC_RANGE:		
return dynamic_range_info();		
case EXT_UNI_DRC:		
return uniDrc();		
case EXT_SAC_DATA:		
return sac_extension_data(cnt);		
case EXT_SAOC_DATA:		
return saoc_extension_data(cnt);		
case EXT_LDSAC_DATA:		
return ldsac_extension_data(cnt);		
case EXT_SBR_DATA:		
return sbr_extension_data(id_aac, 0);		Note 1
case EXT_SBR_DATA_CRC:		
return sbr_extension_data(id_aac, 1);		Note 1
case EXT_SAOC_DE_DATA:		
return saoc_de_extension_data(cnt);		
case EXT_DATA_LENGTH:		
…		

In Table 4.121 add:

Table 4.121 — Values of the extension_type field

1. Symbol	2. Value of extension_type	3. Purpose
EXT_FILL	'0000'	bitstream payload filler
EXT_FILL_DATA	'0001'	bitstream payload data as filler
EXT_DATA_ELEMENT	'0010'	data element
EXT_DATA_LENGTH	'0011'	container with explicit length for extension_payload()
EXT_UNI_DRC	'0100'	Unified dynamic range control
EXT_LDSAC_DATA	'1001'	LD MPEG Surround
EXT_SAOC_DATA	'1010'	SAOC
EXT_DYNAMIC_RANGE	'1011'	dynamic range control
EXT_SAC_DATA	'1100'	MPEG Surround
EXT_SBR_DATA	'1101'	SBR enhancement
EXT_SBR_DATA_CRC	'1110'	SBR enhancement with CRC
EXT_SAOC_DE_DATA	'1111'	SAOC-DE
-	all other values	Reserved: These values can be used for a further extension of the syntax in a compatible way.
Note: Extension payloads of the type EXT_FILL or EXT_FILL_DATA have to be added to the bitstream payload if the total bits for all audio data together with all additional data are lower than the minimum allowed number of bits in this frame necessary to reach the target bitrate. Those extension payloads are avoided under normal conditions and free bits are used to fill up the bit reservoir. Those extension payloads are written only if the bit reservoir is full.		

In 4.5.14.1.1 Data elements, replace:

Table AMD4.7 - Definition of downmix procedure

stereo_downmix_mode	downmix procedure
0	Lo/Ro
1	Lt/Rt

with:

Table AMD4.7 - Definition of downmix procedure

stereo_downmix_mode	downmix procedure
0	Lo/Ro
1	Lo/Ro or Lt/Rt

In 4.5.2.14.2 "Decoding Process", rename the headline of 4.5.2.14.2.1

4.5.2.14.2.1 Downmixing from 5.1 to Stereo

as

4.5.2.14.2.1 Downmixing from 5.1 to Stereo/Mono

Immediately after this headline add a new subclause headline:

4.5.2.14.2.1.1 Downmixing to Stereo

In 4.5.14.2.1.1 Downmixing to stereo, replace:

if **stereo_downmix_mode** is 0,

$$L' = L + C \times b + Ls \times a + LFE \times c$$

$$R' = R + C \times b + Rs \times a + LFE \times c$$

else if **stereo_downmix_mode** is 1,

$$L' = L + C \times b - (Ls + Rs) \times a + LFE \times c$$

$$R' = R + C \times b + (Ls + Rs) \times a + LFE \times c$$

where **surround_mix_level**, "a" and **center_mix_level**, "b" are shown as "Multiplication factor" in Table AMD4.8. C, L, R, Ls, Rs are the source signals and L' and R' are the derived stereo signals. LFE channels should be omitted from the mixdown (i.e. c is equal to zero) if **ext_downmixing_lfe_level_status** is "0". If **ext_downmixing_lfe_level_status** is "1", the LFE mix level "c" shall be derived as shown in Table AMD4.9.

with:

if **stereo_downmix_mode** is 0,

$$Lo = L + C \times b + Ls \times a + LFE \times c$$

$$Ro = R + C \times b + Rs \times a + LFE \times c$$

else if **stereo_downmix_mode** is 1,

$$Lo = L + C \times b + Ls \times a + LFE \times c$$

$$Ro = R + C \times b + Rs \times a + LFE \times c$$

or

$$Lt = L + C \times b - (Ls + Rs) \times a + LFE \times c$$

$$Rt = R + C \times b + (Ls + Rs) \times a + LFE \times c$$

where **surround_mix_level**, "a" and **center_mix_level**, "b" are shown as "Multiplication factor" in Table AMD4.8. C, L, R, Ls, Rs are the source signals and Lo/Ro or Lt/Rt are the derived stereo signals.

If stereo_downmix_mode is "0", the decoder should apply a downmix by obtaining Lo and Ro. If stereo_downmix_mode is "1", the decoder may obtain Lt and Rt as an alternative to Lo and Ro.

LFE channels should be omitted from the mixdown (i.e. c is equal to zero) if **ext_downmixing_lfe_level_status** is "0". If **ext_downmixing_lfe_level_status** is "1", the LFE mix level "c" shall be derived as shown in Table AMD4.9.

Further, after Table AMD4.9, insert the following subclause:

4.5.2.14.2.1.2 Downmixing to Mono

$$M' = L + R + 2 \times C \times b + (Ls + Rs) \times a + 2 \times LFE \times c$$

where **surround_mix_level**, "a" and **center_mix_level**, "b" are shown as "Multiplication factor" in Table AMD4.8. C, L, R, Ls, Rs are the source signals and M' is the derived mono signal. LFE channels should be omitted from the mixdown (i.e. c is equal to zero) if **ext_downmixing_lfe_level_status** is "0". If **ext_downmixing_lfe_level_status** is "1", the LFE mix level "c" shall be derived as shown in Table AMD4.9.

In 4.5.2.14.2.5 after " Table AMD4.12: Default values after synchronization" add:

In addition the "actual compression value" shall be set to 1.0 (0 dB).

Add new section 4.5.2.16 immediately before 4.5.3 with the following text:

4.5.2.16 Unified Dynamic Range Control

The DRC tool specified in ISO/IEC 23003-4 is supported. The corresponding data is carried in an extension payload with the type EXT_UNI_DRC. The DRC tool is operated in regular delay mode and the DRC frame size has the same duration as the AAC frame size.

The time resolution of the DRC tool is specified by *deltaTmin* in units of the audio sample interval. It is calculated as specified in 23003-4. Specific values are provided here as examples based on the following formula:

$$\mathrm{deltaTmin} = 2^{M}.$$

The applicable exponent *M* is found by looking up the audio sample rate range that fulfils:

$$f_{s,min} \le f_s < f_{s,max}$$

Table — AMD5.1 — Lookup table for the exponent *M*

f_s,min [Hz]	f_s,max [Hz]	*M*
8 000	16 000	3
16 000	32 000	4
32 000	64 000	5
64 000	128 000	6

Given the codec frame size N_{Codec}, the DRC frame size in units of DRC samples at a rate of *deltaTmin* is:

$$N_{DRC} = N_{Codec}\, 2^{-M}.$$

For AAC, the DRC tool of 23003-4 offers mandatory decoding capability of up to four DRC subbands using the time-domain DRC filter bank. Optionally, more DRC subbands can be supported by replacing the time-domain DRC filter bank by a uniform 64-band QMF analysis and synthesis filter bank, such as the one defined for HE-AAC. DRC sets that contain more than four DRC subbands must contain gain sequences that are all aligned with the QMF domain.

For HE-AAC and HE-AACv2 decoders the DRC gains are applied to the sub-bands of the QMF domain immediately before the synthesis filter bank.

The drcLocation parameter shall be encoded according to Table AMD5.2.

Table — AMD5.2 — Encoding of drcLocation parameter

drcLocation n	Payload
1	uniDrc() (see ISO/IEC 23003-4)
2	dyn_rng_sgn[i] / dyn_rng_ctl[i] in dynamic_range_info() (see 4.5.2.7)
3	compression_value in MPEG4_ancillary_data() (defined in ISO/IEC 14496-3:2009/Amd.4:2013)
4	*reserved*

In 4.B add new subclause

4.B.22 Features of MPEG-D Part 4: Dynamic Range Control

See ISO/IEC 23003-4 (23003-4:2015, Annex D)

In 1.2 Normative References add:

ISO/IEC 23003-4, "Information technology — MPEG audio technologies — Part 4: Dynamic Range Control"

After Subpart 12, as a new subpart, add:

Subpart 13: Audio Synchronization

13.1 Scope

This subpart of ISO/IEC 14496-3 describes the Audio Synchronization algorithm. An example of the applications using the audio synchronization scheme is a "second screen" application where the 2nd screen content is automatically synchronized to the 1st screen content. In this scenario, no common clock covering the 1st and 2nd screen devices is required, nor an exchange of time-stamps between the devices. Synchronization of the contents between the devices is done by using audio features extracted from the 1st screen content.

For example, the 1st screen content is distributed over existing broadcast system, and the 2nd screen content is distributed over IP network. The audio feature stream of the 1st screen content is sent to the 2nd screen together with the 2nd screen audio/video content over the IP network. In the 2nd screen device, the audio of the 1st screen content is also captured by a microphone and its feature is extracted. The extracted feature from the microphone input and received feature from IP network is compared and the time difference is computed. This time difference is used to align the 2nd screen audio/video content to the 1st screen content. One of the greatest benefits of this approach is that there is no need to modify the transmitter/receiver system of main media stream (for 1st screen).

Figure 13.1 shows the overview of an Audio Synchronization system describing how the system synchronizes two input audio signals. Audio Signal #1 is to be broadcasted as the 1st screen content and Audio Signal #2 is an audio of the 1st screen content captured by a microphone of the 2nd screen device. The system consists of an Audio Feature Extraction tool and an Audio Feature Similarity Calculation tool. The Audio Feature Extraction tool generates audio feature for synchronization from a time domain audio signal. The Audio Feature Similarity Calculation tool compares two audio feature streams to find time difference between the audio signals.

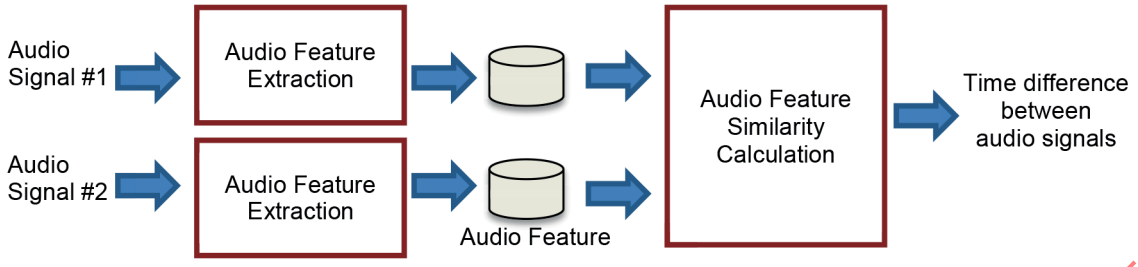

Figure 13.1 — Audio synchronization system

13.2 Definitions

audio feature: coded binary digit sequence extracted from audio signal for audio synchronization

13.3 Bitstream syntax (Normative)

Table 13.1 — Syntax of AudioSyncFeatureSpecificConfig()

Syntax	No. of bits	Mnemonic
AudioSyncFeatureSpecifcConfig()		
{		
audio_sync_feature_type	**4**	**uimsbf**
switch (audio_sync_feature_type) {		
case 0:		
audio_sync_feature_frame_length_index	**4**	**uimsbf**
audio_sync_feature_time_resolution_index	**4**	**uimsbf**
audio_sync_number_of_streams_index	**4**	**uimsbf**
Reserved	**16**	**uimsbf**
break;		
default:		
break;		
}		
}		

Table 13.2 — Syntax of AudioSyncFeatureFrame()

Syntax	No. of bits	Mnemonic
AudioSyncFeatureFrame()		
{		
switch (audio_sync_feature_type) {		
case 0:		
for (i = 0; i < audio_sync_number_of_streams_index+1; i++) {		
for (j=0; j<audio_sync_feature_frame_length; j++) {		
audio_sync_feature	**1**	**uimsbf**
}		
}		
break;		
default:		
break;		
}		
}		

13.4 Semantics (Normative)

Data Elements:

audio_sync_feature_type A four bit field indicating type of audio feature

Table 13.3 — audio_sync_feature_type

audio_sync_feature_type	Description
0	feature type 0
1..15	reserved

audio_sync_feature_frame_length_index A four bit field indicating the bit-length of the feature for a single frame (**audio_sync_feature_frame_length**). The value of **audio_sync_feature_frame_length** is set to the value of the corresponding entry in Table 13.4.

Table 13.4 — audio_sync_feature_frame_length

audio_sync_feature_frame_length_index	Value
0	128
1..15	reserved

audio_sync_feature_time_resolution_index A four bit field indicating the time resolution in milliseconds of the feature (**audio_sync_feature_time_resolution**). The value of **audio_sync_feature_time_resolution** is set to the value of the corresponding entry in Table 13.5.

Table 13.5 — audio_sync_feature_time_resolution

audio_sync_feature_time_resolution_index	Value [milliseconds]
0	32
1	8
2..15	reserved

audio_sync_number_of_streams_index A four bit field indicating the number of audio feature of main media stream conveyed in multiplexed data stream for sub device.

audio_sync_feature The binary feature for audio synchronization for a single frame.

13.5 Audio Feature Extraction Tool (Normative)

This chapter describes the feature extraction algorithm for the feature type 0 (audio_sync_feature_frame_length_index = 0).

13.5.1 Overview

The block diagram of Audio Feature Extraction tool of Figure 13.2 shows how the audio feature is extracted from a time domain audio signal.

First of all, the sampling rate of the input audio signal is converted to 8kHz and divided into audio frames in time domain. For each audio frame, pre-emphasis filter is applied to emphasis high frequency, then band pass filtering is applied in order to split the audio signals into 5 equally spaced frequency bands in log frequency domain.

Then, auto-correlation within each sub-band is calculated and each of the auto-correlation is normalized by maximum peak of the auto-correlation within the sub-band. The normalized auto-correlations obtained from the sub-bands with strong pitch component are summed together to obtain a single integrated auto-correlation values for each time frame .

The lag values which give peaks in the integrated autocorrelation values are detected. The integrated auto-correlation values are converted to audio features represented with binary data based on the detected peak position. The rate of the binary data (Audio Feature Frame Rate) is converted to that for transmission (**audio_sync_feature_time_resolution**) to obtain **audio_sync_feature**

The series of the above process is repeated while the input audio signal is available.

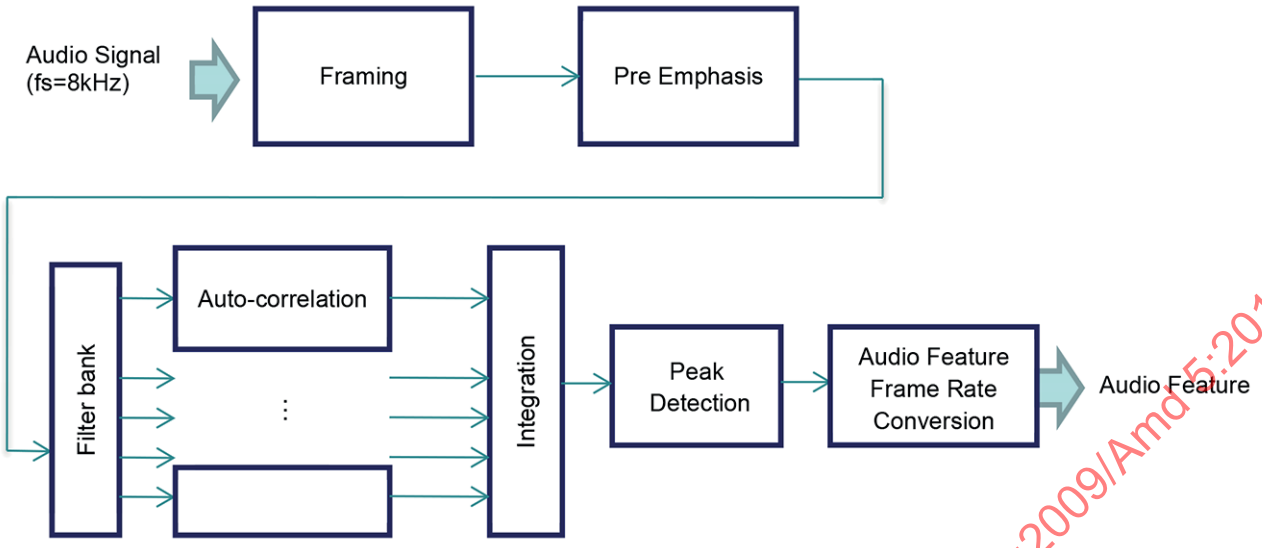

Figure 13.2 — Block diagram of Audio Feature Extraction Tool

13.5.2 Framing

The input audio signal is divided into audio frames with parameters listed in Table 13.6. The input audio signal is then analyzed to generate Audio Features for each input frame. The output frame interval is determined by **audio_sync_feature_time_resolution** (Table 13.5)**.** If the input frame interval and the output frame interval are different, the frame rate of audio feature is converted downward so that it becomes identical to the output frame interval. Table 13.3 shows the audio framing structure of Audio Feature Extraction tool in case of **audio_sync_feature_time_resolution** is 32 msec. In this case the input audio feature is converted downward by a factor of 4 in order to generate output Audio Features.

Table 13.6 — Parameter for framing

Parameter	Value
sampling frequency	8 kHz
input frame length	32 msec (256 samples)
input frame interval	8 msec (64 samples)
window	Hamming Window (256 samples)
output frame interval	**audio_sync_feature_time_resolution** (see Table 13.5)

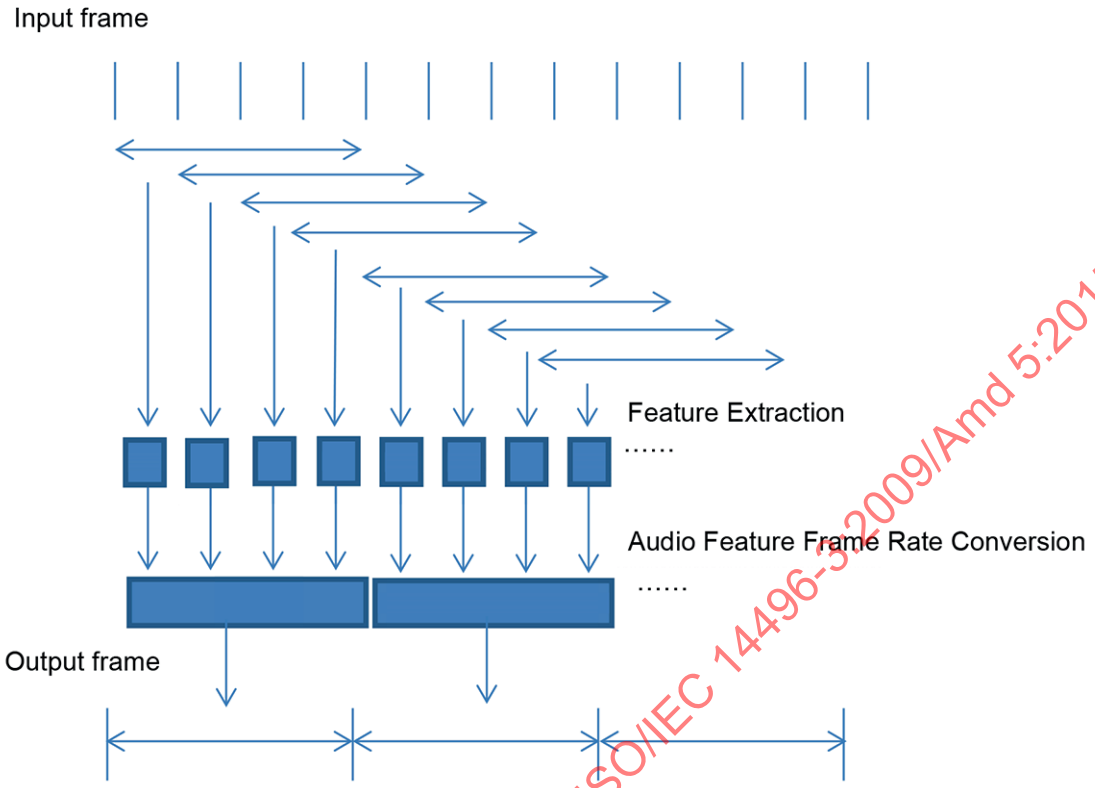

Figure 13.3 — Audio framing structure of Audio Feature Extraction tool

13.5.3 Pre Emphasis

A pre-emphasis filter is applied for the input audio signal to emphasize the high frequency components using:

$$y(n) = x(n) - 0.97 \cdot x(n-1)$$

13.5.4 Filter Bank

Band pass filters are used to split input audio signal into 5 different frequency bands. The band-pass filtering is performed using FIR filters with 129 taps for each sub-band. The coefficients of the filters are listed in Table 13.A.1.

13.5.5 Auto-correlation

Auto-correlation is calculated for each frequency band using:

$$ACF_m(k) = \sum_{n=0}^{N-1} x_m(n) \cdot x_m(n+k), \quad 0 \le k < K, \quad 0 \le m < M$$

where N denotes input frame length, m denotes the index of frequency band, k denotes the index of lag for autocorrelation, K denotes the order of auto-correlation and is set to 128, and n denotes the index of the input audio signal. M denotes the number of frequency bands and is set 5.

The auto-correlation is normalized by $ACF_m(0)$ and normalized auto-correlation function $NACF_m(k)$ is derived as:

$$NACF_m(k) = \frac{ACF_m(k)}{ACF_m(0)}, \quad 0 \le k < K, \quad 0 \le m < M$$

There are K autocorrelation coefficients.

13.5.6 Confidence Measure

For each frequency band m, confidence measure CM_m is calculated based on the auto-correlation value.

$$CM_m = \max_{10 \le k \le K-1} NACF_m(k), \quad 0 \le m < M$$

Where 10 is a lag value at the beginning of the search area, and K-1 is the one at the end of the search area.

13.5.7 Integration

The normalized auto-correlation function $NACF_m(k)$ values derived from each sub-band are summed together into a single integrated auto-correlation function.

$$ACF_{integrated}(k) = \frac{\sum_{m=0}^{N_b-1} NACF_m(k) \cdot W_m}{\sum_{m=0}^{N_b-1} W_m}, \quad 0 \le k < K$$

where W_m is defined as following

$$W_m = \begin{cases} 0, & CM_m < T \\ 1, & CM_m \ge T \end{cases}$$

where T denotes a threshold for the confidence measure and is set to 0.3. In case that $\sum_{m=0}^{N_b-1} W_m$ is zero then $ACF_{integrated}(k)$ is set to zero.

13.5.8 Peak Detection

The integrated auto-correlation function is converted into a 128-bit length feature vector $\boldsymbol{f}(k)$ ($0 \le k \le K - 1$) and each bit position corresponds to the lag of the auto-correlation function. If the auto-correlation function at the lag of k is a positive peak, the k-th bit of the feature vector is set to 1. Otherwise, the k-th bit of the feature vector is set to 0. Picking a positive peak is done as shown below:

Threshold Calculation:

$$t(k) = \frac{1}{\min(k+N, K-1) - \max(0, k-N) + 1} \sum_{l=\max(0,k-N)}^{l=\min(k+N,K-1)} ACF_{integrated}(k), \quad 1 < k < K-1$$

where N denotes the order for the moving average, and is set to 10.

$$q(k) = \begin{cases} 1, & ACF_{integrated}(k) > t(k) + h \\ 0, & otherwise \end{cases}, \quad 1 \le k < K-1$$

where h denotes the offset for the threshold and is set to 0.1.

Pick up peak candidate:

$$p(k)=\begin{cases}1, & ACF_{integrated}(k)>ACF_{integrated}(k-1) \\ & ACF_{integrated}(k)>ACF_{integrated}(k+1) \\ 0, & otherwise\end{cases}, \quad 1\le k<K-1$$

In order to find the peak positions, the following operation is performed.

$$\boldsymbol{f}(0)=0$$

$$\boldsymbol{f}(k)=\begin{cases}1, & p(k)=1, q(k)=1 \\ 0, & otherwise\end{cases}, \quad 1\le k<K-1$$

$$\boldsymbol{f}(127)=0$$

13.5.9 Audio Feature Frame Rate Conversion

The audio feature is calculated every 8 msec, on the other hand the output Audio Feature frame rate may differ according to **audio_sync_feature_time_resolution** (see 13.5.2). If **audio_sync_feature_time_resolution** is larger than 8 msec, for example 32 msec, the Audio Feature is converted downward in time by a factor of 4. Let $\boldsymbol{f}_i$ be audio feature calculated from i-th frame of input audio signal. The dimension of the feature vectors is 128 and each element is 0 or 1. The conversion is performed by the following operation:

$$\bar{\boldsymbol{f}}_j(k)=\boldsymbol{f}_{4j}(k)\vee\boldsymbol{f}_{4j+1}(k)\vee\boldsymbol{f}_{4j+2}(k)\vee\boldsymbol{f}_{4j+3}(k), \quad 0\le j<\mathrm{N}_f/4, \quad 0\le k<128$$

where N_f is the number of audio frames of Audio Signal. The audio feature $\bar{\boldsymbol{f}}_j(k)\,(0\le k<128)$ is the j-th output of the Audio Feature Extraction tool. This corresponds to **audio_sync_feature** conveyed in AudioSyncFeaureFrame(). (See Table 13.2) Figure 13.4 illustrates how the conversion works for excerpt of Audio feature $\boldsymbol{f}_i(k)$ and $\bar{\boldsymbol{f}}_j(k)$.

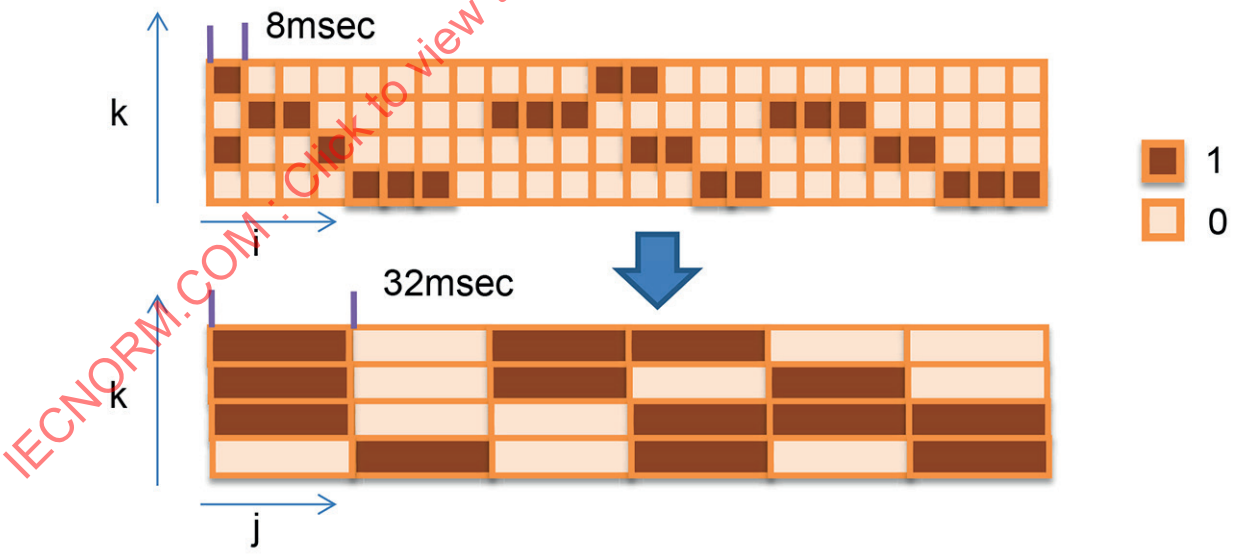

Figure 13.4 — Audio feature frame rate conversion (downward)

Annex 13.A

(normative)

Tables

13.A.1 Filterbank Coefficient

Index	Band0	Band1	Band2	Band3	Band4
1	-2.2646310E-05	3.6961410E-05	3.1100770E-04	4.2873920E-04	-5.2505410E-04
2	-1.0648430E-05	1.2090870E-04	5.1869620E-04	-3.2939290E-04	2.9676110E-04
3	1.4011680E-05	2.3692150E-04	5.8116440E-04	-8.2971190E-04	1.5824390E-04
4	5.3137000E-05	3.7609210E-04	4.0719320E-04	-3.5076500E-04	2.8374260E-04
5	1.0907140E-04	5.2520690E-04	4.0565840E-19	3.2601100E-04	-8.5287150E-04
6	1.8455350E-04	6.6628850E-04	-5.2602910E-04	2.4818210E-04	4.0760090E-04
7	2.8249230E-04	7.7680260E-04	-9.7686940E-04	-3.8347340E-05	2.5647240E-05
8	4.0567390E-04	8.3093040E-04	-1.1559410E-03	4.2583860E-04	7.2409030E-04
9	5.5641240E-04	8.0212520E-04	-9.5977570E-04	9.3771390E-04	-1.2746420E-03
10	7.3616290E-04	6.6689330E-04	-4.4618200E-04	1.0735680E-04	4.4152940E-04
11	9.4511900E-04	4.0940930E-04	1.7003920E-04	-1.3617380E-03	-6.7248780E-05
12	1.1818200E-03	2.6265320E-05	6.0963480E-04	-1.3226280E-03	1.2739500E-03
13	1.4427910E-03	-4.6956320E-04	6.7705340E-04	1.6563670E-04	-1.6256100E-03
14	1.7222520E-03	-1.0465180E-03	3.8891350E-04	7.4213930E-04	1.5994920E-04
15	2.0119010E-03	-1.6548770E-03	2.5663470E-18	-1.1208780E-18	-2.5600970E-18
16	2.3008240E-03	-2.2303320E-03	-1.1350150E-04	3.0887600E-04	1.7611800E-03
17	2.5755220E-03	-2.7007400E-03	3.2486950E-04	1.9362530E-03	-1.5234900E-03
18	2.8200850E-03	-2.9955360E-03	1.2788460E-03	1.5565730E-03	-7.8145500E-04
19	3.0165200E-03	-3.0566390E-03	2.3286760E-03	-1.7468130E-03	4.0516120E-04
20	3.1452280E-03	-2.8492760E-03	2.8202340E-03	-3.5714360E-03	1.8864180E-03
21	3.1856230E-03	-2.3708450E-03	2.1948110E-03	-1.1685650E-03	-5.0682600E-04
22	3.1168900E-03	-1.6559820E-03	3.4242630E-04	1.4584790E-03	-2.6411560E-03
23	2.9188440E-03	-7.7637170E-04	-2.2255240E-03	5.9017240E-04	1.2131400E-03
24	2.5728730E-03	1.6552440E-04	-4.5219660E-03	-2.2817720E-04	1.4104450E-03
25	2.0629280E-03	1.0487100E-03	-5.5244010E-03	2.7575310E-03	1.7029770E-03
26	1.3765120E-03	1.7507670E-03	-4.7090260E-03	4.6948600E-03	-5.3214960E-03
27	5.0564490E-04	2.1677450E-03	-2.3886320E-03	-2.1033160E-04	2.1583970E-03
28	-5.5226550E-04	2.2339940E-03	3.6950870E-04	-6.6033940E-03	3.8229950E-04
29	-1.7936700E-03	1.9386980E-03	2.2609010E-03	-5.1855440E-03	4.9334180E-03
30	-3.2082920E-03	1.3359280E-03	2.4604450E-03	1.1805520E-03	-8.1512190E-03
31	-4.7787500E-03	5.4567240E-04	1.1503070E-03	2.2514570E-03	2.5101210E-03
32	-6.4804100E-03	-2.5568310E-04	-4.7791950E-04	-7.8306030E-04	-7.3016560E-04
33	-8.2815050E-03	-8.5544590E-04	-8.2787120E-04	2.2059570E-03	8.4363220E-03
34	-1.0143520E-02	-1.0337050E-03	1.1515000E-03	8.9101690E-03	-9.8607390E-03
35	-1.2021880E-02	-5.9823080E-04	5.1746620E-03	5.3665670E-03	1.1420120E-03
36	-1.3866860E-02	5.7928760E-04	9.4790380E-03	-7.9999370E-03	-1.1252640E-03

Index	Band0	Band1	Band2	Band3	Band4
37	-1.5624810E-02	2.5309190E-03	1.1529230E-02	-1.2739490E-02	1.1014030E-02
38	-1.7239590E-02	5.1694150E-03	9.3053490E-03	-2.7789020E-03	-8.8219550E-03
39	-1.8654160E-02	8.2790220E-03	2.5411700E-03	4.5587760E-03	-3.1506310E-03
40	-1.9812350E-02	1.1523390E-02	-6.7911570E-03	-6.0343350E-18	-1.2059670E-17
41	-2.0660700E-02	1.4471670E-02	-1.5135520E-02	-4.3949620E-04	1.1397290E-02
42	-2.1150330E-02	1.6641630E-02	-1.8952410E-02	1.2137550E-02	-3.4999310E-03
43	-2.1238770E-02	1.7555760E-02	-1.6521650E-02	1.6407400E-02	-1.1163200E-02
44	-2.0891630E-02	1.6804670E-02	-8.9905750E-03	-3.3137190E-03	2.9479390E-03
45	-2.0084190E-02	1.4109880E-02	1.5222830E-17	-2.2929810E-02	8.7518770E-03
46	-1.8802700E-02	9.3779470E-03	6.0845780E-03	-1.4640670E-02	7.0718910E-03
47	-1.7045360E-02	2.7377980E-03	6.5369450E-03	4.8060670E-03	-2.2804190E-02
48	-1.4823020E-02	-5.4450960E-03	1.9954830E-03	3.9645650E-03	6.9892610E-03
49	-1.2159450E-02	-1.4581830E-02	-3.5005280E-03	-4.5925870E-03	3.1088670E-03
50	-9.0913050E-03	-2.3901380E-02	-4.4330750E-03	1.1493490E-02	2.3098030E-02
51	-5.6675390E-03	-3.2517530E-02	3.0776180E-03	3.4269030E-02	-3.6842540E-02
52	-1.9485360E-03	-3.9515430E-02	1.8488380E-02	1.5480620E-02	9.8473050E-03
53	1.9952040E-03	-4.4048970E-02	3.5961580E-02	-3.2942710E-02	-4.4817370E-03
54	6.0847240E-03	-4.5438180E-02	4.6214510E-02	-4.3694920E-02	4.4617790E-02
55	1.0234580E-02	-4.3255230E-02	4.0599040E-02	-5.5416710E-03	-5.1033200E-02
56	1.4355060E-02	-3.7388270E-02	1.5702400E-02	1.3771300E-02	6.6073280E-03
57	1.8354580E-02	-2.8074520E-02	-2.3748350E-02	-8.4429930E-03	-1.2176890E-02
58	2.2142240E-02	-1.5897160E-02	-6.5361940E-02	2.0105420E-03	7.4737960E-02
59	2.5630330E-02	-1.7446850E-03	-9.3201770E-02	7.3622490E-02	-6.2634400E-02
60	2.8736850E-02	1.3264400E-02	-9.4113690E-02	9.0906890E-02	-1.7235540E-02
61	3.1387860E-02	2.7884200E-02	-6.3648010E-02	-3.9194780E-02	-1.7895170E-02
62	3.3519640E-02	4.0861130E-02	-8.9013790E-03	-1.8728450E-01	1.4615060E-01
63	3.5080580E-02	5.1057600E-02	5.3226390E-02	-1.3339780E-01	-6.9168850E-02
64	3.6032670E-02	5.7566460E-02	1.0193650E-01	1.0315500E-01	-2.7241490E-01
65	3.6352650E-02	5.9803310E-02	1.2033160E-01	2.4025630E-01	4.8015420E-01
66	3.6032670E-02	5.7566460E-02	1.0193650E-01	1.0315500E-01	-2.7241490E-01
67	3.5080580E-02	5.1057600E-02	5.3226390E-02	-1.3339780E-01	-6.9168850E-02
68	3.3519640E-02	4.0861130E-02	-8.9013790E-03	-1.8728450E-01	1.4615060E-01
69	3.1387860E-02	2.7884200E-02	-6.3648010E-02	-3.9194780E-02	-1.7895170E-02
70	2.8736850E-02	1.3264400E-02	-9.4113690E-02	9.0906890E-02	-1.7235540E-02
71	2.5630330E-02	-1.7446850E-03	-9.3201770E-02	7.3622490E-02	-6.2634400E-02
72	2.2142240E-02	-1.5897160E-02	-6.5361940E-02	2.0105420E-03	7.4737960E-02
73	1.8354580E-02	-2.8074520E-02	-2.3748350E-02	-8.4429930E-03	-1.2176890E-02
74	1.4355060E-02	-3.7388270E-02	1.5702400E-02	1.3771300E-02	6.6073280E-03
75	1.0234580E-02	-4.3255230E-02	4.0599040E-02	-5.5416710E-03	-5.1033200E-02
76	6.0847240E-03	-4.5438180E-02	4.6214510E-02	-4.3694920E-02	4.4617790E-02
77	1.9952040E-03	-4.4048970E-02	3.5961580E-02	-3.2942710E-02	-4.4817370E-03
78	-1.9485360E-03	-3.9515430E-02	1.8488380E-02	1.5480620E-02	9.8473050E-03

Index	Band0	Band1	Band2	Band3	Band4
79	-5.6675390E-03	-3.2517530E-02	3.0776180E-03	3.4269030E-02	-3.6842540E-02
80	-9.0913050E-03	-2.3901380E-02	-4.4330750E-03	1.1493490E-02	2.3098030E-02
81	-1.2159450E-02	-1.4581830E-02	-3.5005280E-03	-4.5925870E-03	3.1088670E-03
82	-1.4823020E-02	-5.4450960E-03	1.9954830E-03	3.9645650E-03	6.9892610E-03
83	-1.7045360E-02	2.7377980E-03	6.5369450E-03	4.8060670E-03	-2.2804190E-02
84	-1.8802700E-02	9.3779470E-03	6.0845780E-03	-1.4640670E-02	7.0718910E-03
85	-2.0084190E-02	1.4109880E-02	1.5222830E-17	-2.2929810E-02	8.7518770E-03
86	-2.0891630E-02	1.6804670E-02	-8.9905750E-03	-3.3137190E-03	2.9479390E-03
87	-2.1238770E-02	1.7555760E-02	-1.6521650E-02	1.6407400E-02	-1.1163200E-02
88	-2.1150330E-02	1.6641630E-02	-1.8952410E-02	1.2137550E-02	-3.4999310E-03
89	-2.0660700E-02	1.4471670E-02	-1.5135520E-02	-4.3949620E-04	1.1397290E-02
90	-1.9812350E-02	1.1523390E-02	-6.7911570E-03	-6.0343350E-18	-1.2059670E-17
91	-1.8654160E-02	8.2790220E-03	2.5411700E-03	4.5587760E-03	-3.1506310E-03
92	-1.7239590E-02	5.1694150E-03	9.3053490E-03	-2.7789020E-03	-8.8219550E-03
93	-1.5624810E-02	2.5309190E-03	1.1529230E-02	-1.2739490E-02	1.1014030E-02
94	-1.3866860E-02	5.7928760E-04	9.4790380E-03	-7.9999370E-03	-1.1252640E-03
95	-1.2021880E-02	-5.9823080E-04	5.1746620E-03	5.3665670E-03	1.1420120E-03
96	-1.0143520E-02	-1.0337050E-03	1.1515000E-03	8.9101690E-03	-9.8607390E-03
97	-8.2815050E-03	-8.5544590E-04	-8.2787120E-04	2.2059570E-03	8.4363220E-03
98	-6.4804100E-03	-2.5568310E-04	-4.7791950E-04	-7.8306030E-04	-7.3016560E-04
99	-4.7787500E-03	5.4567240E-04	1.1503070E-03	2.2514570E-03	2.5101210E-03
100	-3.2082920E-03	1.3359280E-03	2.4604450E-03	1.1805520E-03	-8.1512190E-03
101	-1.7936700E-03	1.9386980E-03	2.2609010E-03	-5.1855440E-03	4.9334180E-03
102	-5.5226550E-04	2.2339940E-03	3.6950870E-04	-6.6033940E-03	3.8229950E-04
103	5.0564490E-04	2.1677450E-03	-2.3886320E-03	-2.1033160E-04	2.1583970E-03
104	1.3765120E-03	1.7507670E-03	-4.7090260E-03	4.6948600E-03	-5.3214960E-03
105	2.0629280E-03	1.0487100E-03	-5.5244010E-03	2.7575310E-03	1.7029770E-03
106	2.5728730E-03	1.6552440E-04	-4.5219660E-03	-2.2817720E-04	1.4104450E-03
107	2.9188440E-03	-7.7637170E-04	-2.2255240E-03	5.9017240E-04	1.2131400E-03
108	3.1168900E-03	-1.6559820E-03	3.4242630E-04	1.4584790E-03	-2.6411560E-03
109	3.1856230E-03	-2.3708450E-03	2.1948110E-03	-1.1685650E-03	-5.0682600E-04
110	3.1452280E-03	-2.8492760E-03	2.8202340E-03	-3.5714360E-03	1.8864180E-03
111	3.0165200E-03	-3.0566390E-03	2.3286760E-03	-1.7468130E-03	4.0516120E-04
112	2.8200850E-03	-2.9955360E-03	1.2788460E-03	1.5565730E-03	-7.8145500E-04
113	2.5755220E-03	-2.7007400E-03	3.2486950E-04	1.9362530E-03	-1.5234900E-03
114	2.3008240E-03	-2.2303320E-03	-1.1350150E-04	3.0887600E-04	1.7611800E-03
115	2.0119010E-03	-1.6548770E-03	2.5663470E-18	-1.1208780E-18	-2.5600970E-18
116	1.7222520E-03	-1.0465180E-03	3.8891350E-04	7.4213930E-04	1.5994920E-04
117	1.4427910E-03	-4.6956320E-04	6.7705340E-04	1.6563670E-04	-1.6256100E-03
118	1.1818200E-03	2.6265320E-05	6.0963480E-04	-1.3226280E-03	1.2739500E-03
119	9.4511900E-04	4.0940930E-04	1.7003920E-04	-1.3617380E-03	-6.7248780E-05
120	7.3616290E-04	6.6689330E-04	-4.4618200E-04	1.0735680E-04	4.4152940E-04