

NGBF Standard

차세대방송표준포럼표준(국문표준)

NGBF-STD-020

제정일: 2017년 1월 06일

하이브리드 오디오 파일 포맷

Hybrid Audio File Format

표준초안 검토 위원회	디지털라디오분과위원회				
표준안 심의 위원회	운영위원회				
	성명	소 속	직위	위원회 및 직위	표준번호
표준(과제) 제안	정영호	ETRI	책임	-	
표준 초안 작성자	이석진	경기대	교수	-	NGBF-STD-0xx
	정영호	ETRI	책임	-	
	서상원	ETRI	연구원	-	

본 문서에 대한 저작권은 차세대방송표준포럼에 있으며, 차세대방송표준포럼과 사전 협의 없이 이 문서의 전체 또는 일부를 상업적 목적으로 복제 또는 배포해서는 안 됩니다.

본 표준 발간 이전에 접수된 지식재산권 확약서 정보는 본 표준의 ‘부록(지식재산권 확약서 정보)’에 명시하고 있으며, 이후 접수된 지식재산권 확약서는 차세대방송표준포럼 웹사이트에서 확인할 수 있습니다.

본 표준과 관련하여 접수된 확약서 외의 지식재산권이 존재할 수 있습니다.

발행인 : 차세대방송표준포럼 의장

발행처 : 차세대방송표준포럼

06130, 서울특별시 강남구 테헤란로 7길 22 신관 1108호

Tel : 02-568-3556, Fax : 02-568-3557

발행일 : 2017.01.xx

서 문

1 표준의 목적

본 표준의 목적은 채널 및 객체, 장면 오디오로 구성된 하이브리드 오디오 신호를 다양한 청취 환경에 적합하도록 렌더링하여 개인화된 인터랙션과 몰입감을 제공하기 위한 하이브리드 오디오 파일 포맷을 정의한다.

2 주요 내용 요약

본 표준은 하이브리드 오디오 콘텐츠의 저작 및 재생에 필요한 메타데이터 모델 및 파일 구조에 대해 규정한다. 메타데이터 모델은 Content와 Format의 두 부분으로 구성되며, Content 부분은 오디오가 담고 있는 내용에 관련된 정보를, Format 부분은 렌더링에 필요한 기술적인 정보를 정의한다. 파일 구조는 4 GB 이상의 오디오 데이터 및 이와 관련된 메타데이터를 저장할 수 있도록 정의한다.

3 인용 표준과의 비교

3.1 인용 표준과의 관련성

본 표준은 ITU-R BS.2076 표준을 기반으로 하여, 국내 하이브리드 오디오 콘텐츠의 저작 및 재생 환경에 적합하도록 인용 표준의 세부내용을 3.2의 비교표와 같이 재구성한다.

3.2 인용 표준과 본 표준의 비교표

NGBF-STD-0xx	ITU-R BS.2076(2015)	비고
5.1 개요	1. Introduction	—
—	2. Background	삭제
5.2 구조	3. Description of the Model	—
—	4. Standard formats	삭제
5.3 메타데이터	5. ADM elements	수정
—	6. Use of IDs	삭제
—	7. <chna> chunk	삭제
5.4 좌표계	8. Coordinate System	수정

–	9. Object-based parameter descriptions	삭제
---	--	----

Preface

1 Purpose

The purpose of the standard is to define a hybrid audio file format, which provides personalized interaction and immersion in a variety of listening environments, by rendering the hybrid audio signal of channel-, object-, and scene-based audio.

2 Summary

The standard specifies the metadata model and the file structure required for authoring and playback of hybrid audio content. The metadata model is divided into two parts: the Content part, which defines audio content information, and the Format part, which defines rendering information. The file structure is defined to store audio data over 4 GB and related metadata.

3 Relationship to Reference Standards

3.1 Relevance to the Reference Standard

Based on ITU-R BS.2076, the details of the reference standard is reorganized as followed table on section 3.2 to be suitable for domestic authoring and playback environment of hybrid audio contents.

3.2 Comparison of the Reference and the Quoted Standard

NGBF-STD-0xx	ITU-R BS.2076(2015)	Remarks
5.1 Overview	1. Introduction	–
–	2. Background	Deleted
5.2 Structure	3. Description of the Model	–
–	4. Standard formats	Deleted
5.3 Metadata Element	5. ADM elements	Modified

–	6. Use of IDs	Deleted
–	7. <chna> chunk	Deleted
5.4 Coordinate System	8. Coordinate System	Modified
–	9. Object-based parameter descriptions	Deleted

목 차

1 적용 범위	1
2 인용 표준	1
3 용어 정의	1
4 약어	1
5 하이브리드 오디오 메타데이터 모델	2
5.1 개요	2
5.2 구조	2
5.3 메타데이터 요소	3
5.4 좌표계	19
6 하이브리드 오디오 파일 구조	19
부록 I -1 지식재산권 확약서 정보	20
I -2 시험인증 관련 사항	21
I -3 본 표준의 연계(family) 표준	22
I -4 참고 문헌	23
I -5 영문표준 해설서	24
I -6 표준의 이력	25
부록 II 하이브리드 오디오 메타데이터 예시	26

하이브리드 오디오 파일 포맷

Hybrid Audio File Format

1 적용 범위

본 표준은 채널 및 객체, 장면 오디오로 구성된 하이브리드 오디오 콘텐츠의 저작 및 재생에 필요한 파일 포맷을 정의하며, 이는 메타데이터 모델 및 파일 구조로 구성된다. 본 표준에서 정의하는 메타데이터 모델은 Content와 Format의 두 부분으로 구성되며, Content 부분은 오디오가 담고 있는 내용에 관련된 정보를, Format 부분은 렌더링에 필요한 기술적인 정보를 포함한다. 파일 구조는 4 GB 이상의 오디오 데이터 및 이와 관련된 메타데이터를 저장할 수 있도록 정의한다.

2 인용 표준

ITU-R BS.2076 (2015), Audio Definition Model

ITU-R BS.2088 (2015), Long-form file format for the international exchange of audio programme materials with metadata

3 용어 정의

3.1 채널 기반 오디오

각 트랙이 출력 스피커에 바로 대응되는 형태의 오디오 신호로, 별도의 렌더러를 거치지 않고 스피커로 출력되는 형태를 가지는 오디오 데이터 및 포맷을 의미한다.

3.2 객체 기반 오디오

각 트랙이 오디오 프로그램 상의 객체에 대응되는 형태의 오디오 신호로, 각 객체 기반 오디오를 스피커 출력 신호로 변환하는 렌더러를 필요로 한다.

3.3 장면 기반 오디오

고차 앰비소닉(Higher Order Ambisonics)과 같이 오디오 음장을 독립적인 컴포넌트로 표현한 오디오 신호를 의미한다.

4 약어

dBTP	dB True-Peak
HOA	Higher Order Ambisonics
LFE	Low Frequency Effects
LKFS	Loudness, K-weighted, relative to Full Scale
LU	Loudness Unit
LUFS	Loudness Units relative to Full Scale
PCM	Pulse Code Modulation

5 하이브리드 오디오 메타데이터 모델

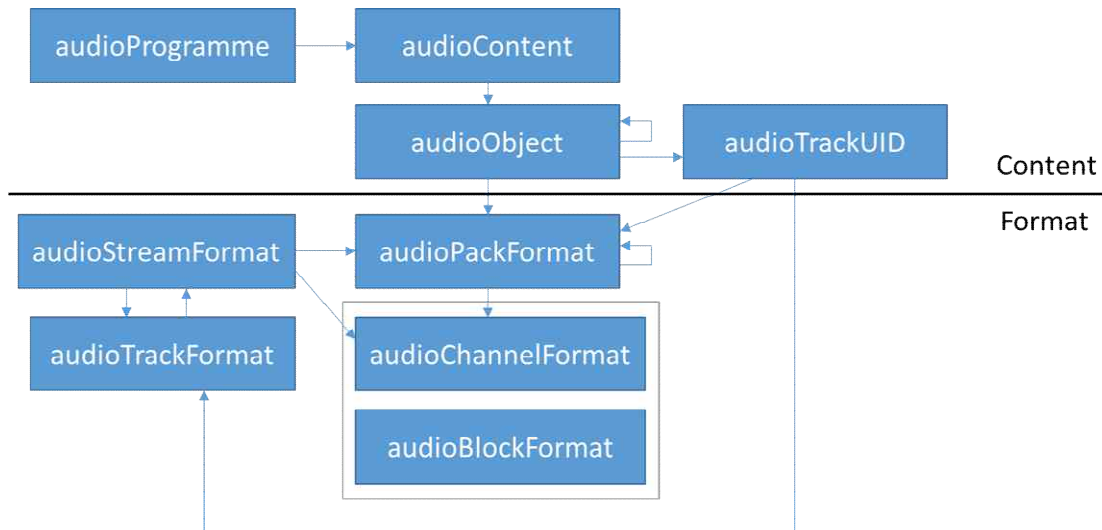
5.1 개요

고품질 실감 음향을 제공하기 위해 방송 및 영화관의 오디오 시스템은 보다 복잡하고 다양해지고 있으며, 이에 따라 오디오 포맷 또한 다양해지는 추세이다. 기존 채널 기반 오디오는 이와 같이 발전하는 오디오 시스템을 충실히 지원할 수 없으며, 이를 위해 객체 기반 오디오 및 장면 기반 오디오 등 여러 오디오 시스템이 발전되고 있다.

이와 같이 다양한 오디오 시스템을 모두 만족시키기 위한 메타데이터 모델은 다양한 시스템 상의 오디오를 모두 정의할 수 있어야 한다. 본 표준에서는 이러한 목적을 달성할 수 있는 하이브리드 오디오 메타데이터 모델을 기술한다.

5.2 구조

본 표준에서 기술하는 하이브리드 오디오 메타데이터 모델은 (그림 5-1)과 같은 구조를 가지고 있으며, Content와 Format의 두 부분으로 이루어져 있다. Content 부분은 오디오 프로그램의 언어라던가 라우드니스 정보 등 오디오가 담고 있는 내용에 관련된 정보를, Format 부분은 디코딩 혹은 렌더링에 필요한 기술적인 정보를 규정한다. 하이브리드 오디오 메타데이터 모델은 XML을 기반으로 표현된다.



(그림 5-1) 하이브리드 오디오 메타데이터 모델 구조

5.3 메타데이터 요소

5.3.1 audioTrackFormat

audioTrackFormat은 저장 장치 상에 존재하는 하나의 트랙에 대응되는 요소이다. 이는 렌더러가 오디오 신호를 디코딩할 때 필요한 데이터 포맷에 대한 정보를 기술하는 데에 사용된다. 여러 트랙 정보가 동시에 디코딩될 때, 이 요소는 audioStreamFormat에 의해 참조된다.

일반적인 PCM 오디오의 경우, audioStreamFormat과 audioTrackFormat은 동일한 역할을 하기 때문에 하나의 audioStreamFormat은 대응되는 하나의 audioTrackFormat을 참조한다. 반면에 인코딩된 오디오 신호의 경우, 하나의 audioStreamFormat이 여러 개의 audioTrackFormat을 참고하는 경우도 있다.

5.3.1.1 Attributes

Attribute	상세	예시
audioTrackFormatID	트랙 식별자	AT_00010001_01
audioTrackFormatName	트랙 이름	PCM_FrontLeft
formatLabel	포맷 식별자	0001
formatDefinition	포맷 설명	PCM

5.3.1.2 Sub-elements

Element	상세	예시	개수
audioStreamFormatIDRef	참조하는 audioStreamFormat의 식별자	AT_00010001_01	0 혹은 1

5.3.2 audioStreamFormat

스트림은 채널 기반의 신호를 디코딩하거나 객체 기반의 오디오 신호를 렌더링 할 때 필요한 트랙의 묶음이며, audioStreamFormat은 이러한 스트림에 대한 정보를 기술한다. 또한 이 요소는 audioTrackFormat과 audioChannelFormat (혹은 audioPackFormat) 사이의 관계를 설정하는 역할을 한다. 기본적으로 PCM이 아닌 인코딩된 신호를 다룰 때에 주로 사용된다.

5.3.2.1 Attributes

Attribute	상세	예시
audioStreamFormatID	스트림 식별자	AS_00010001
audioStreamFormatName	스트림 이름	PCM_FrontLeft
formatLabel	포맷 식별자	0001
formatDefinition	포맷 설명	PCM

5.3.2.2 Sub-elements

Element	상세	예시
audioChannelFormatIDRef	참조하는 audioChannelFormat 식별자	AC_00010001
audioPackFormatIDRef	참조하는 audioPackFormat 식별자	AP_00010003
audioTrackFormatIDRef	참조하는 audioTrackFormat 식별자	AT_00010001_01

5.3.3 audioChannelFormat

audioChannelFormat은 채널 기반 신호의 경우에는 채널 관련 정보를, 객체 기반 신호의 경우에는 하나의 객체 관련 정보 등을 기술하는 요소이다. 이 요소는 시간 축에서 여러 개의 audioBlockFormat으로 구성된다.

5.3.3.1 Attributes

Attribute	상세	예시
audioChannelFormatID	채널 식별자	AC_00010001
audioChannelFormatName	채널 이름	FrontLeft
typeLabel	채널 유형 식별자	0001
typeDefinition	채널 유형 설명	DirectSpeakers

typeDefinition은 오디오 신호의 종류를 기술하며, 본 속성에 따라 audioBlockFormat의 파라미터가 결정된다. 다음과 같은 5개의 typeDefinition이 정의된다.

typeDefinition	typeLabel	상세
DirectSpeakers	0001	채널 기반 신호 (스피커에 직접적으로 전송되는 경우)
Matrix	0002	채널 기반 신호 (채널 신호가 믹싱되어 스피커로 전송되는 경우)
Objects	0003	객체 기반 신호
HOA	0004	장면 기반 신호
Binaural	0005	바이노럴 신호

5.3.3.2 Sub-elements

Element	상세	Attributes	개수
audioBlockFormat	채널 신호의 시간축 블록	5.3.4 절 참고	1...*
frequency	오디오 신호의 저주파 혹은 고주파 임계 주파수	typeDefinition = "lowPass" 혹은 "highPass"	0...2

frequency 파라미터는 오디오 신호의 주파수 범위를 기술하는 요소이다. 이는 저역통과 혹은 고역통과 범위로 설정될 수 있으며, 이를 조합하여 대역통과 범위로의 설정이 가능하다. 가장 일반적인 사용 예로 LFE 채널 신호를 기술할 때 저역통과 주파수를 설정하는 경우를 들 수 있다.

5.3.4 audioBlockFormat

audioBlockFormat은 audioChannelFormat 샘플 중 특정 시간 구간 내에서 고정된 파라미터를 가지는 오디오 신호를 기술한다. 예를 들어, 객체 신호 중 음상의 위치가 고정된 구간이 하나의 audioBlockFormat으로 표현될 수 있다.

5.3.4.1 Attributes

Attribute	상세	예시
audioBlockFormatID	블록 식별자	AB_00010001_00000001
rTime	블록의 시작 시간 (상위 audioObject의 시작 시간을 기준으로 한 상대적인 값)	00:00:00.00000
duration	블록의 지속 시간	00:00:05.00000

5.3.4.2 Sub-elements

Sub-element는 audioChannelFormat의 typeDefinition (혹은 typeLabel)에 따라 달라진다.

5.3.4.2.1 typeDefinition0이 “DirectSpeaker”인 경우

Element	Attribute	Bound Attribute	상세	단위	예시	개수
speakerLabel		-	스피커 위치에 대한 고유 이름	-	M-30	0...*
position	coordinate="azimuth"		스피커 위치 (수평각)	도	-30.0	1
position	coordinate="azimuth"	max	위치 최대값 (수평각)	도	-22.5	0 혹은 1
position	coordinate="azimuth"	min	위치 최소값 (수평각)	도	-30.0	0 혹은 1
position	coordinate="elevation"		스피커 위치 (수직각)	도	0.0	1
position	coordinate="elevation"	max	위치 최대값 (수직각)	도	5.0	0 혹은 1
position	coordinate="elevation"	min	위치 최소값 (수직각)	도	0.0	0 혹은 1
position	coordinate="distance"		스피커 위치 (거리)	1로 정규화된 값	1.0	0 혹은 1

position	coordinate="distance"	max	위치 최대값 (거리)	1로 정규화된 값	0.9	0 혹은 1
position	coordinate="distance"	min	위치 최소값 (거리)	1로 정규화된 값	0.8	0 혹은 1
position	screenEdgeLock		스피커 위치를 화면 가장자리로 정의	left, right, top, bottom	left	0 혹은 1

screenEdgeLock 속성은 스피커의 위치를 화면의 크기에 연동되도록 정의하며, coordinate="azimuth" 혹은 coordinate="elevation"과 함께 사용된다. 이 때 화면이 존재하지 않거나 혹은 화면의 크기가 제공되지 않는 경우를 대비하여, screenEdgeLock이 설정된 경우에도 스피커 위치는 반드시 명시되어야 한다.

coordinate="distance"인 요소의 경우 정규화된 거리를 사용하며, 이 때 정규화의 기준 거리는 audioPackFormat에 기술된 값을 사용한다.

5.3.4.2.2 typeDefinition0이 "Matrix"인 경우

Element	Attribute	Bound Attribute	상세	단위	예시	개수
speakerLabel		-	스피커 위치에 대한 고유 이름	-	M-30	0...*
position	coordinate="azimuth"		스피커 위치 (수평각)	도	-30.0	1
position	coordinate="azimuth"	max	위치 최대값 (수평각)	도	-22.5	0 혹은 1
position	coordinate="azimuth"	min	위치 최소값 (수평각)	도	-30.0	0 혹은 1
position	coordinate="elevation"		스피커 위치 (수직각)	도	0.0	1
position	coordinate="elevation"	max	위치 최대값 (수직각)	도	5.0	0 혹은 1
position	coordinate="elevation"	min	위치 최소값 (수직각)	도	0.0	0 혹은 1
position	coordinate="distance"		스피커 위치 (거리)	1로 정규화된 값	1.0	0 혹은 1
position	coordinate="distance"	max	위치 최대값 (거리)	1로 정규화된 값	0.9	0 혹은 1
position	coordinate="distance"	min	위치 최소값 (거리)	1로 정규화된 값	0.8	0 혹은 1

position	screenEdgeLock		스피커 위치를 화면 가장자리로 정의	left, right, top, bottom	left	0 혹은 1
----------	----------------	--	---------------------	-----------------------------------	------	--------

Element	Sub-elements	상세	개수
matrix	coefficient	각 채널에 적용되는 계수를 담고 있는 요소	1

Element	Attribute	상세	단위	예시	개수
coefficient	gain	어느 한 채널에 곱해질 크기 이득 계수		-0.5	1...*
coefficient	phase	어느 한 채널에 적용될 위상 지연 계수	도	90	1...*
coefficient		참조하는 채널 식별자		AC_00010001	1...*

5.3.4.2.3 typeDefinition01 “Objects”인 경우

Element	Attribute	상세	단위	예시	개수	기본값
position	coordinate = “azimuth”	객체 위치 (수평각)	도 (-180~180)	-22.5	1	
position	coordinate = “elevation”	객체 위치 (수직각)	도 (-90~90)	5.0	1	
position	coordinate = “distance”	객체 위치 (거리)	1로 정규화 (-1~1)	0.9	0 혹은 1	1.0
width		객체의 수평방향 크기	도	45	0 혹은 1	0.0
height		객체의 수직방향 크기	도	20	0 혹은 1	0.0
depth		객체의 거리방향 크기	1로 정규화	0.2	0 혹은 1	0.0
gain		객체 신호에 적용될 이득	선형값 (dB 아님)	0.5	0 혹은 1	1.0
diffuse		객체 신호의 확산감	0.0~1.0			
jumpPosition	interpolation Length	1: 보간하지 않음 0: 보간함	0/1	1	0 혹은 1	0
screenRef		1: 스크린 연동된 위치 0: 스크린과 무관한 위치	0/1	0	0 혹은 1	0

priority	method = “panning”	렌더링 방법의 우선순위 (패닝)	1~3의 정수	1	0 혹은 1	1
priority	method = “binaural”	렌더링 방법의 우선순위 (바이노럴)	1~3의 정수	2	0 혹은 1	2
priority	method = “transaural”	렌더링 방법의 우선순위 (트랜스오럴)	1~3의 정수	3	0 혹은 1	3

객체 신호의 경우, 음향객체 제어를 위하여 다음과 같이 디스플레이 상의 스크린 맵 정보를 포함한다.

Element	Attribute	상세	단위	예시	개수	기본값
screenmap	available	1: 스크린맵 사용 가능 0: 스크린맵 사용 불가	0 혹은 1	1	0 혹은 1	0

Element	Sub-element	Attribute	상세	단위	예시	개수	기본값
screenmap	position	x	마커 중심의 가로 위치 (스크린 좌측 기준)	가로 해상도로 정규화된 픽셀	0.3	1	-
screenmap	position	y	마커 중심의 세로 위치 (스크린 하단 기준)	세로 해상도로 정규화된 픽셀	0.2	1	-
screenmap	size	mode	-	-	-	1	-

size 요소의 경우 mode 속성에 따라 포함하고 있는 하위요소의 종류가 다르며, mode 속성은 0, 1, 2, 3의 값을 가진다. mode 속성이 0인 경우는 하위요소를 포함하지 않으며, 기타 mode에 따른 하위요소는 다음 표와 같다.

Element	Sub-element	상세	단위	예시	개수	비고
size	radius	스크린맵 마커(원형) 의 반지름	가로 해상도로 정규화된 픽셀	0.05	1	mode="1"일 때 사용
size	width	스크린맵 마커(사각 형)의 가로 길이	가로 해상도로 정규화된 픽셀	0.02	1	mode="2"일 때 사용
size	height	스크린맵 마커(사각 형)의 세로 길이	세로 해상도로 정규화된 픽셀	0.03	1	mode="2"일 때 사용
size	point	스크린맵 마커(윤곽)			1...*	mode="3"일 때 사용

		의 꼭지점				다음 표의 속성 참고
--	--	-------	--	--	--	-------------

위의 하위요소 중 mode="3"일 때 사용하는 point 요소의 속성은 다음과 같다.

Element	Attribute	상세	단위	예시
point	x	꼭지점의 가로 좌표 (스크린 좌측 기준)	가로 해상도 로 정규화된 픽셀	0.3
point	y	꼭지점의 세로 좌표 (스크린 하단 기준)	세로 해상도 로 정규화된 픽셀	0.4
point	id	꼭지점의 순서 (1부터 윤곽선 순서대로)		1

스크린맵 정보의 각 mode별 예시는 다음과 같다.

```

<screenmap available="1">
  <position x="0.3" y="0.2"></position>
  <!-- mode="0"인 경우의 예시 -->
  <size mode="0"/>
  <!-- mode="1"인 경우의 예시 -->
  <size mode="1">
    <radius>0.05</radius>
  </size>
  <!-- mode="2"인 경우의 예시 -->
  <size mode="2">
    <width>0.02</width>
    <height>0.03</height>
  </size>
  <!-- mode="3"인 경우의 예시 -->
  <size mode="3">
    <point x="0.3" y="0.4" id="1"/>
    <point x="0.35" y="0.45" id="2"/>
    <point x="0.37" y="0.45" id="3"/>
  </size>
</screenmap>

```

</size> </screenmap>

5.3.4.2.4 typeDefinition이 “HOA”인 경우

Sub-element	상세	자료형	예시	개수
equation	HOA 성분을 정의하는 수식	문자열		0 혹은 1
degree	HOA 성분의 degree	정수	1	0 혹은 1
order	HOA 성분의 order	정수	1	0 혹은 1

5.3.4.2.5 typeDefinition이 “Binaural”인 경우

바이노럴 오디오의 경우 audioBlockFormat을 위한 별도의 메타데이터가 필요하지 않다.

5.3.5 audioPackFormat

audioPackFormat은 서로 연관되어 있는 여러 개의 audioChannelFormat을 하나의 그룹으로 묶는 역할을 한다. 예를 들면, 채널 기반 포맷의 “stereo” 혹은 “5.1” 등이 이에 해당하며, audioPackFormat은 네스팅 구조를 가지고 있어서 다른 audioPackFormat을 참조할 수 있다. typeDefinition과 typeLabel은 해당 audioPackFormat이 참조하고 있는 audioChannelFormat과 일치해야 한다.

5.3.5.1 Attributes

Attribute	상세	예시
audioPackFormatID	팩 식별자	AP_00010001
audioPackFormatName	팩 이름	stereo
typeLabel	채널 유형 식별자	0001
typeDefinition	채널 유형 설명	DirectSpeakers

5.3.5.2 Sub-elements

Element	상세	예시	개수
audioChannelFormatIDRef	참조하는 audioChannelFormat의 식별자	AC_00010001	0...*
audioPackFormatIDRef	참조하는 audioPackFormat의 식별자	AP_00010002	0...*
absoluteDistance	정규화에 사용되는 기준 거리 (단위: 미터)	4.5	0 혹은 1

여기서 absoluteDistance 요소는 audioBlockFormat의 정규화된 거리를 정의할 때 기준이 되는 값이다.

5.3.6 audioObject

audioObject는 Content와 Format, 그리고 trackUID를 연결하는 역할을 한다. 또한 audioObject는 네스팅 구조를 가지고 있어서 다른 audioObject를 참조할 수 있다.

5.3.6.1 Attributes

Attribute	상세	예시
audioObjectID	오브젝트 식별자	AO_1001
audioObjectName	오브젝트 이름	dialogue_stereo
start	오브젝트 시작 시간 (프로그램 시작 시간 기준)	00:00:00.00000
duration	오브젝트의 지속 시간	00:02:00.00000
dialogue	0: 음성을 포함하지 않은 경우 1: 음성만 있는 경우 2: 음성과 비음성 오디오가 함께 포함되어 있는 경우	0
interact	1: 사용자 인터랙션이 가능한 경우 0: 사용자 인터랙션이 불가능한 경우	1
disableDucking	1: 오디오 덕킹을 허용하지 않음 0: 오디오 덕킹을 허용	0

5.3.6.2 Sub-elements

Element	상세	예시
audioPackIDRef	참조하는 audioPackFormat의 식별자	AP_00010001
audioObjectIDRef	참조하는 audioObject의 식별자	AO_1002
audioComplementaryObjectIDRef	상보적 관계에 있는 audioObject의 식별자 (예를 들어, 서로 다른 언어로 되어 있는 음성 신호 등)	AO_1003
audioTrackUIDRef	참조하는 audioTrackUID	ATU_00000001
audioObjectInteraction	사용자 인터랙션에 대한 정보 기술 (5.3.6.4절 참고)	

5.3.6.3 audioComplementaryObjectIDRef

audioComplementaryObjectIDRef는 자신과 상보적인 관계에 있는 다른 audioObject를 참조한다. 예를 들어, 한국어와 영어로 더빙된 오디오 프로그램이 있다고 가정할 때, 한국어 음성과 영어 음성은 동시에 재생되지 않으므로 두 음성 신호는 상보적 관계에 있다. 한국어 음성의 식별자가 AO_1001이고 영어 음성의 식별자가 AO_1003일 때, AO_1001의 하위요소인 audioComplementaryObjectIDRef는 AO_1003을 참조한다.

5.3.6.4 audioObjectInteraction

Attribute	상세	예시
onOffInteract	1: 사용자가 해당 오브젝트를 활성화/비활성화 할 수 있음 0: 사용자가 해당 오브젝트를 활성화/비활성화 할 수 없음	0
gainInteract	1: 사용자가 해당 오브젝트의 음량을 조절할 수 있음 0: 사용자가 해당 오브젝트의 음량을 조절할 수 없음	0
positionInteract	1: 사용자가 해당 오브젝트의 위치를 조절할 수 있음 0: 사용자가 해당 오브젝트의 위치를 조절할 수 없음	0

5.3.7 audioContent

audioContent는 배경음악 혹은 음성 등과 같은 프로그램의 한 구성요소의 내용을 기술하는 요소이며, audioObject를 참조하고 있다. 또한 해당 요소는 라우드니스 메타데이터를 포함한다. audioContentLanguage는 ISO 636-1에 정의된 2자리

언어 코드를 사용하여 표기한다.

5.3.7.1 Attributes

Attribute	상세	예시
audioContentID	콘텐츠 식별자	ACO_1001
audioContentName	콘텐츠 이름	Music
audioContentLanguage	콘텐츠 언어	en

5.3.7.2 Sub-elements

Element	상세	예시
audioObjectIDRef	참조하는 audioObject 식별자	AO_1001
loudnessMetadata	5.3.7.4절 참고	
dialogue	0: 음성을 포함하지 않은 경우 1: 음성만 있는 경우 2: 음성과 비음성 오디오가 함께 포함되어 있는 경우	0

5.3.7.3 dialogue

dialogue는 해당 콘텐츠의 내용을 기술하는 요소이다. 기본적으로 dialogue 요소는 0 (음성 없음), 1 (음성만 존재), 2 (음성과 비음성이 섞여 있음)의 값을 가지며, 각 값에 따라 아래와 같은 속성이 정의된다.

dialogue의 값	Attribute	상세	예시
0	nonDialogueContentKind	콘텐츠 종류 (아래 표 참고)	0
1	dialogueContentKind	콘텐츠 종류 (아래 표 참고)	0
2	mixedContentKind	콘텐츠 종류 (아래 표 참고)	0

세부적인 콘텐츠 종류에 따른 각 속성값은 아래 표와 같다.

nonDialogueContentKind	상세
------------------------	----

0	정의되지 않음
1	음악
2	이펙트

dialogueContentKind	상세
0	정의되지 않음
1	(스토리 상의) 대사
2	해설 (voiceover)
3	음성자막
4	시각장애인을 위한 음성 묘사
5	논평 (commentary)
6	비상시 (emergency) 음성

mixedContentKind	상세
0	정의되지 않음
1	완성된 통합 신호 (complete main)
2	믹싱된 신호
3	청각 장애인 (hearing impaired)을 위한 오디오

5.3.7.4 loudnessMetadata

Attribute	상세	예시
loudnessMethod	라우드니스 계산 방법	“BS.1770”
loudnessRecType	라우드니스 보정 시 준수된 권고안	“R128”
loudnessCorrectionType	라우드니스 보정 종류 (실시간 혹은 파일기반)	“File-based”

Element	상세	단위	예시
integratedLoudness	평균 라우드니스 값	LKFS/LUFS	-23.0

loudnessRange	라우드니스 범위	LU	10.0
maxTruePeak	최대 피크값	dBTP	-2.3
maxMomentary	최대 순간 라우드니스 값	LKFS/LUFS	-19.0
maxShortTerm	최대 단기 라우드니스 값	LKFS/LUFS	-21.2
dialogueLoudness	음성 신호의 평균 라우드니스 값	LKFS/LUFS	-24.0

5.3.8 audioProgramme

audioProgramme은 전체 오디오 프로그램을 구성하는 audioContent의 집합을 기술한다. 또한 audioProgramme은 비디오 타임코드에 맞추어 전체 프로그램의 기준이 되는 시작 및 종료 시간을 기술하며, 라우드니스 메타데이터를 포함한다.

5.3.8.1 Attributes

Attribute	상세	예시
audioProgrammeID	프로그램 식별자	APR_1001
audioProgrammeName	프로그램 이름	
audioProgrammeLanguage	프로그램 내 음성의 언어	en
start	프로그램의 시작 타임코드 (비디오 타임코드 기준)	00:00:10.00000
end	프로그램의 종료 타임코드 (비디오 타임코드 기준)	00:10:00.00000

5.3.8.2 Sub-elements

Element	상세	예시
audioContentIDRef	참조하는 audioContent 식별자	ACO_1001
loudnessMetadata	5.3.8.4절 참고	
audioProgrammeReferenceScreen	프로그램 제작 시 기준으로 정의했던 스크린의 크기 (5.3.8.3절 참고)	

5.3.8.3 audioProgrammeReferenceScreen

audioProgrammeReferenceScreen은 프로그램 제작 시 사용했거나 혹은 기준으

로 삼았던 스크린의 위치 및 크기를 정의하는 요소이다.

Attribute	상세	예시
aspectRatio	스크린의 화면 비율	1.78, 1.6

Element	Coordinate Attribute	상세	단위	예시
screenCenterPosition	azimuth	스크린 중앙의 위치 (수평각)	도	+30.0
	elevation	스크린 중앙의 위치 (수직각)	도	-15.0
	distance	스크린 중앙의 위치 (거리)	0.0 ~ 1.0	1.0
screendWidth	azimuth	스크린의 너비 (수평각)	도	+58.0

5.3.8.4 loudnessMetadata

Attribute	상세	예시
loudnessMethod	라우드니스 계산 방법	"BS.1770"
loudnessRecType	라우드니스 보정 시 준수된 권고안	"R128"
loudnessCorrectionType	라우드니스 보정 종류 (실시간 혹은 파일기반)	"File-based"

Element	상세	단위	예시
integratedLoudness	평균 라우드니스 값	LKFS/LUFS	-23.0
loudnessRange	라우드니스 범위	LU	10.0
maxTruePeak	최대 피크값	dbTP	-2.3
maxMomentary	최대 순간 라우드니스 값	LKFS/LUFS	-19.0
maxShortTerm	최대 단기 라우드니스 값	LKFS/LUFS	-21.2
dialogueLoudness	음성 신호의 평균 라우드니스 값	LKFS/LUFS	-24.0

5.3.9 audioTrackUID

audioTrackUID는 하나의 트랙 혹은 에셋(asset)을 단일 식별하는 요소이다. 즉,

하나의 audioTrackUID는 파일 혹은 레코딩된 신호의 한 개 트랙 혹은 에셋에 대응되며, 양자화 비트수 및 샘플링 주파수 정보를 포함한다.

5.3.9.1 Attributes

Attribute	상세	예시
UID	트랙의 단일 식별자 값	ATU_00000001
sampleRate	샘플링 주파수 (단위: Hz)	48000
bitDepth	양자화 비트수	24

5.3.9.2 Sub-elements

Element	상세	예시
audioTrackFormatIDRef	참조하는 audioTrackFormat의 식별자	AT_00010001_01
audioPackFormatIDRef	참조하는 audioPackFormat의 식별자	AP_00010002

5.3.10 audioFormatExtended

audioFormatExtended는 모든 메타데이터 요소를 포함하는 상위 요소이다.

5.3.10.1 Sub-elements

Element	상세
audioProgramme	전체 오디오 프로그램에 대한 기술
audioContent	프로그램 내의 오디오 콘텐츠들에 대한 기술
audioObject	오디오 트랙과 그 포맷을 연결하는 요소
audioPackformat	서로 연관되어 있는 채널 그룹에 대한 기술
audioChannelFormat	오디오 채널에 대한 기술
audioStreamFormat	오디오 스트림(트랙의 그룹)에 대한 기술
audioTrackFormat	오디오 트랙에 대한 기술
audioTrackUID	실제 오디오 트랙에 대한 단일 식별자

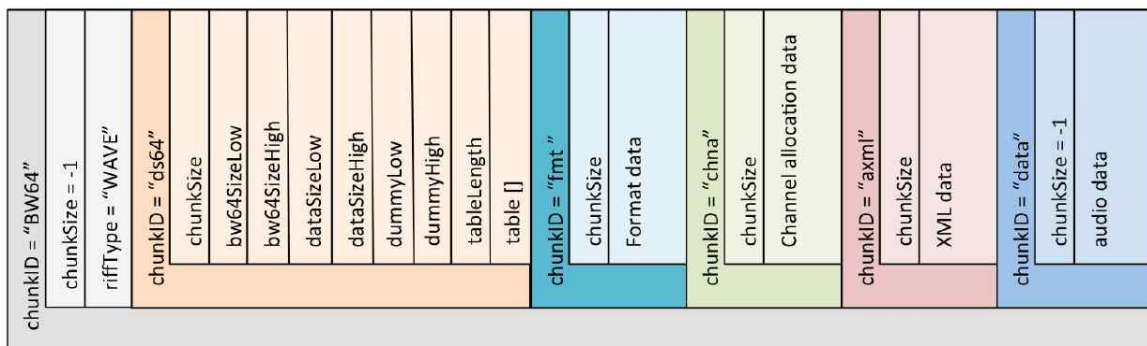
5.4. 좌표계

본 표준에서는 구면 좌표계를 기준으로 사용한다. 구면 좌표계의 원점, 수평각도, 수직각도는 다음과 같은 기준에 따라 정의한다.

- 원점: 시스템의 스위트 스팟(sweet spot)을 기준으로 한다. 만약, 스위트 스팟이 정의되지 않는 시스템이라면, 청취 공간의 중앙을 원점으로 한다.
- 수평각도: 정면을 기준으로 하는 수평면 상의 각도이며, 위에서 보았을 때 반시계 방향을 양(+)의 방향으로 정의한다.
- 수직각도: 수평면을 기준으로 하는 수직면 상의 각도이며, 위쪽 방향을 양(+)의 방향으로 정의한다.

6 하이브리드 오디오 파일 구조

본 표준에 기술된 하이브리드 오디오 파일은 (그림 6-1)과 같은 long-form 오디오 파일 구조 (ITU-R BS.2088)에 따라 구성되며, XML로 기술된 하이브리드 오디오 메타데이터는 “axml” 청크에 포함된다.



(그림 6-1) Long-form 오디오 파일 구조

부 록 1-1

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

지식재산권 요약서 정보

본 표준의 ‘지적 재산권 요약서’ 제출 현황은 차세대방송표준포럼 웹사이트에서 확인할 수 있다.

※ 상기 기재된 지식재산권 요약서 이외에도 본 표준이 발간된 후 접수된 요약서가 있을 수 있으니, 차세대방송표준포럼 웹사이트에서 확인하시기 바랍니다.

부 록 1-2

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

시험인증 관련 사항

1-2.1 시험인증 대상 여부

해당사항 없음

1-2.2 시험표준 제정 현황

해당사항 없음

부 록 1-3

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

본 표준의 연계(family) 표준

해당 사항 없음

부 록 | -4

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

참고

해당사항 없음

부 록 1-5

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

영문표준 해설서

해당 사항 없음

부 록 1-6

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

표준의 이력

판수	채택일	표준번호	내용	담당 위원회
제1판	2017.1.06	제정 NGBF-STD-020	-	디지털라디오분과 위원회

부 록 II

하이브리드 오디오 메타데이터 예시

다음은 본 표준에 따라 XML 파일 형식으로 기술된 하이브리드 오디오 메타데이터에 대한 예시이며, 해당 오디오 신호는 스테레오 형식의 채널 기반 오디오(2채널)와 1개의 객체 기반 오디오로 구성되어 있다.

```
<?xml version='1.0' encoding='utf-8'?>
<audioModel xmlns="urn:metadata:audio">
<coreMetadata>
<format>
<audioFormatExtended>

    <!-- ##### -->
    <!-- PACKS -->
    <!-- ##### -->

    <audioPackFormat audioPackFormatID="AP_00010002"
audioPackFormatName="Stereo" typeLabel="0001" typeDefinition="DirectSpeakers">
        <audioChannelFormatIDRef>AC_00010001</audioChannelFormatIDRef>
        <audioChannelFormatIDRef>AC_00010002</audioChannelFormatIDRef>
    </audioPackFormat>

    <audioPackFormat audioPackFormatID="AP_00031001" audioPackFormatName="Car"
typeLabel="0003" typeDefinition="Objects">
        <audioChannelFormatIDRef>AC_00031001</audioChannelFormatIDRef>
    </audioPackFormat>

    <!-- ##### -->
    <!-- CHANNELS -->
    <!-- ##### -->

    <audioChannelFormat audioChannelFormatID="AC_00010001"
audioChannelFormatName="FrontLeft" typeLabel="0001" typeDefinition="DirectSpeakers">
        <audioBlockFormat audioBlockFormatID="AB_00010001_00000001">
            <speakerLabel>M+30</speakerLabel>
            <position coordinate="azimuth">30.0</position>
```

```

    <position coordinate="elevation">0.0</position>
    <position coordinate="distance">1.0</position>
  </audioBlockFormat>
</audioChannelFormat>

<audioChannelFormat audioChannelFormatID="AC_00010002"
audioChannelFormatName="FrontRight" typeLabel="0001" typeDefinition="DirectSpeakers">
  <audioBlockFormat audioBlockFormatID="AB_00010002_00000001">
    <speakerLabel>M-30</speakerLabel>
    <position coordinate="azimuth">-30.0</position>
    <position coordinate="elevation">0.0</position>
    <position coordinate="distance">1.0</position>
  </audioBlockFormat>
</audioChannelFormat>

<audioChannelFormat audioChannelFormatID="AC_00031001"
audioChannelFormatName="Car1" typeLabel="0003" typeDefinition="Objects">
  <audioBlockFormat audioBlockFormatID="AB_00031001_00000001"
rtime="00:00:00.00000" duration="00:00:05.00000">
    <position coordinate="azimuth">-22.5</position>
    <position coordinate="elevation">5.0</position>
    <position coordinate="distance">1.0</position>
    <priority method="panning">1</priority>
    <priority method="transaural">2</priority>
    <priority method="binaural">3</priority>
    <screenmap available="1">
      <position x="0.4" y="0.2"/>
      <size mode="1">
        <radius>0.05</radius>
      </size>
    </screenmap>
  </audioBlockFormat>
  <audioBlockFormat audioBlockFormatID="AB_00031001_00000002"
rtime="00:00:05.00000" duration="00:00:10.00000">
    <position coordinate="azimuth">-24.5</position>
    <position coordinate="elevation">6.0</position>
    <position coordinate="distance">0.9</position>
    <priority method="panning">1</priority>
    <priority method="transaural">2</priority>

```

```

    <priority method="binaural">3</priority>
    <screenmap available="1">
        <position x="0.35" y="0.2"/>
        <size mode="1">
            <radius>0.05</radius>
        </size>
    </screenmap>
</audioBlockFormat>
<audioBlockFormat audioBlockFormatID="AB_00031001_00000003"
rtime="00:00:15.00000" duration="00:00:20.00000">
    <position coordinate="azimuth">-26.5</position>
    <position coordinate="elevation">7.0</position>
    <position coordinate="distance">0.8</position>
    <priority method="panning">1</priority>
    <priority method="transaural">2</priority>
    <priority method="binaural">3</priority>
    <screenmap available="1">
        <position x="0.3" y="0.2"/>
        <size mode="1">
            <radius>0.05</radius>
        </size>
    </screenmap>
</audioBlockFormat>
</audioChannelFormat>

<!-- ##### -->
<!-- STREAMS -->
<!-- ##### -->

<audioStreamFormat audioStreamFormatID="AS_00010001"
audioStreamFormatName="PCM_FrontLeft" formatLabel="0001" formatDefinition="PCM">
    <audioChannelFormatIDRef>AC_00010001</audioChannelFormatIDRef>
    <audioTrackFormatIDRef>AT_00010001_01</audioTrackFormatIDRef>
</audioStreamFormat>

<audioStreamFormat audioStreamFormatID="AS_00010002"
audioStreamFormatName="PCM_FrontRight" formatLabel="0001" formatDefinition="PCM">
    <audioChannelFormatIDRef>AC_00010002</audioChannelFormatIDRef>
    <audioTrackFormatIDRef>AT_00010002_01</audioTrackFormatIDRef>

```

```

</audioStreamFormat>

<audioStreamFormat audioStreamFormatID="AS_00031001"
audioStreamFormatName="PCM_Car1" formatLabel="0001" formatDefinition="PCM">
    <audioChannelFormatIDRef>AC_00031001</audioChannelFormatIDRef>
    <audioTrackFormatIDRef>AT_00031001_01</audioTrackFormatIDRef>
</audioStreamFormat>

<!-- ##### -->
<!-- AUDIO TRACKS -->
<!-- ##### -->

<audioTrackFormat audioTrackFormatID="AT_00010001_01"
audioTrackFormatName="PCM_FrontLeft" formatLabel="0001" formatDefinition="PCM">
    <audioStreamFormatIDRef>AS_00010001</audioStreamFormatIDRef>
</audioTrackFormat>

<audioTrackFormat audioTrackFormatID="AT_00010002_01"
audioTrackFormatName="PCM_FrontRight" formatLabel="0001" formatDefinition="PCM">
    <audioStreamFormatIDRef>AS_00010002</audioStreamFormatIDRef>
</audioTrackFormat>

<audioTrackFormat audioTrackFormatID="AT_00031001_01"
audioTrackFormatName="PCM_Car1" formatLabel="0001" formatDefinition="PCM">
    <audioStreamFormatIDRef>AS_00031001</audioStreamFormatIDRef>
</audioTrackFormat>

<!-- ##### -->
<!-- PROGRAMMES -->
<!-- ##### -->

<audioProgramme audioProgrammeID="APR_1001"
audioProgrammeName="Documentary">
    <audioContentIDRef>ACO_1001</audioContentIDRef>
    <audioContentIDRef>ACO_1002</audioContentIDRef>
</audioProgramme>

<!-- ##### -->
<!-- CONTENTS -->

```

```

<!-- ##### -->

<audioContent audioContentID="ACO_1001" audioContentName="Beds">
  <audioObjectIDRef>AO_1001</audioObjectIDRef>
  <loudnessMetadata loudnessMethod="BS.1770" loudnessRecType="R128"
loudnessCorrectionType="file-based">
    <integratedLoudness>-28.0</integratedLoudness>
  </loudnessMetadata>
</audioContent>

<audioContent audioContentID="ACO_1002" audioContentName="PointSources">
  <audioObjectIDRef>AO_1002</audioObjectIDRef>
  <loudnessMetadata loudnessMethod="BS.1770" loudnessRecType="R128"
loudnessCorrectionType="file-based">
    <integratedLoudness>-23.0</integratedLoudness>
  </loudnessMetadata>
</audioContent>

<!-- ##### -->
<!-- OBJECTS -->
<!-- ##### -->

<audioObject audioObjectID="AO_1001" audioObjectName="Music"
start="00:00:00.00">
  <audioPackFormatIDRef>AP_00010002</audioPackFormatIDRef>
  <audioTrackUIDRef>ATU_00000001</audioTrackUIDRef>
  <audioTrackUIDRef>ATU_00000002</audioTrackUIDRef>
</audioObject>

<audioObject audioObjectID="AO_1002" audioObjectName="Car" start="00:00:00.00">
  <audioPackFormatIDRef>AP_00031001</audioPackFormatIDRef>
  <audioTrackUIDRef>ATU_00000003</audioTrackUIDRef>
  <audioObjectInteraction positionInteract="1"/>
</audioObject>

<!-- ##### -->
<!-- AUDIO TRACK UIDs -->
<!-- ##### -->

```

```
<audioTrackUID UID="ATU_00000001">
  <audioTrackFormatIDRef>AT_00010001_01</audioTrackFormatIDRef>
  <audioPackFormatIDRef>AP_00010002</audioPackFormatIDRef>
</audioTrackUID>

<audioTrackUID UID="ATU_00000002">
  <audioTrackFormatIDRef>AT_00010002_01</audioTrackFormatIDRef>
  <audioPackFormatIDRef>AP_00010002</audioPackFormatIDRef>
</audioTrackUID>

<audioTrackUID UID="ATU_00000003">
  <audioTrackFormatIDRef>AT_00031001_01</audioTrackFormatIDRef>
  <audioPackFormatIDRef>AP_00031001</audioPackFormatIDRef>
</audioTrackUID>

</audioFormatExtended>
</format>
</coreMetadata>
</audioModel>
```

앞에서 기술된 하이브리드 오디오 메타데이터 예시의 요소들간 관계를 나타내는 구성도는 다음과 같다.

