



6DoF 音響サービスのための発音源メタデータの記述法

Description method of audio metadata about sound source for 6DoF audio services

長谷川知美¹⁾, 杉本岳大¹⁾, 久保弘樹¹⁾, 大出訓史¹⁾, 岩崎泰士¹⁾, 大久保洋幸²⁾
Tomomi HASEGAWA, Takehiro SUGIMOTO, Hiroki KUBO, Satoshi OODE, Taishi IWASAKI and
Hiroyuki OOKUBO

1) 日本放送協会 (〒157-8510 東京都世田谷区砧 1-10-11, hasegawa.t-co@nhk.or.jp)

2) NHK 財団 (〒157-8510 東京都世田谷区砧 1-10-11)

概要: イマーシブメディア用にユーザの six-degrees-of-freedom (6DoF) に対応したオブジェクトベース音響のメタデータの記述法を検討している。互換性の高い、標準メタデータの記述法の策定を目的として、three-degrees-of-freedom (3DoF) 対応の ITU-R 勧告の Audio Definition Model を基にリスナの位置に拠らない 6DoF コンテンツ空間上の発音源の位置、および放射パターン、発音源の向き、距離減衰特性に関する拡張記述法を開発した。

キーワード: 6DoF, イマーシブ, メタデータ, オブジェクトベース音響

1. はじめに

近年、コンテンツ空間をユーザが自由に移動できる 6DoF にも対応したイマーシブなメディア (以下、イマーシブメディア) の開発が進んでいる[1][2]。放送業界においても 6DoF に対応した音響サービスの検討が始まっており、我々も制作者への調査を通じて 6DoF 音響コンテンツの制作要件を抽出した[3]。

リスナの位置変化に応じて音を再生するためには、音声信号を再生するための位置等の情報を音響メタデータとして記述するオブジェクトベース音響の技術が適している。そこで、6DoF 音響コンテンツの制作要件を満たす音響メタデータの要件の抽出を行った[4]。

6DoF に対応した技術革新はゲーム業界では目覚ましいが[5]、プラットフォームに特化したコンテンツであるとともに、非公開な技術を用いている場合も多く、互換性が乏しい。放送業界においては海外や他の放送局とコンテンツを交換することもあり、高い互換性が求められる。そこで様々なプラットフォームへ互換運用が可能な国際標準となる 6DoF 音響サービスに用いる音響メタデータ (以下、6DoF 音響メタデータ) を規定することを目指して、記述法の開発に着手した。

オブジェクトベース音響で用いる高い互換性を特徴とする音響メタデータとして、ITU-R (国際電気通信連合の無線通信部門)が勧告した Audio Definition Model (以下、

ADM) [6]があるものの、3DoF までしか対応していない。本稿では、ADM を 6DoF に拡張するため、リスナの位置に拠らない発音源の絶対位置と、方向毎に異なる周波数特性をもつ発音源をリスナが様々な方向から視聴するための放射パターンと発音源の向き、距離減衰特性の記述法について検討した。2章で拡張のベースとなる ADM の記述法について紹介し、3章で提案する拡張記述法について述べる。

2. ADM における位置に関する記述法

ADM は、eXtensible Markup Language (XML) 形式による音響メタデータの記述法である。コンテンツ内の音声オブジェクトに対して、その音声オブジェクトを構成する音声信号の再生位置や構成等を要素 (以下、Element) と属性 (以下、Attribute) を用いて規定している。ここで位置に関する記述法について紹介する。

オブジェクトの位置はリスナの位置を基準として図 1、2 で示すように右手系の極座標またはデカルト座標で記述する。極座標ではリスナの位置を中心としてオブジェクトの方角 (azimuth)、仰角 (elevation)、動径 (distance) を記述する。動径はリスナの位置を中心とした再生するスピーカまでの距離を 1 と正規化した値である。デカルト座標ではリスナ位置を原点としてスピーカが設置される立方体の四隅を 1 または-1 と正規化した左右方向 (X)、前後

方向 (Y), 上下方向 (Z) の値を記述する. ADM ではリスナの位置を座標軸の原点としているため, 6DoF のようにリスナが自由に動くコンテンツ空間におけるオブジェクトの絶対的な位置は明示されない.

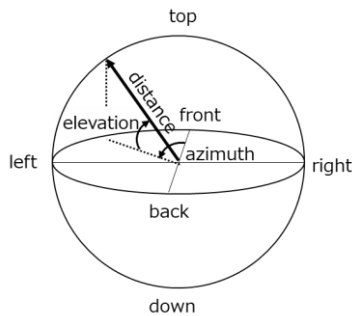


図 1: ADM における極座標

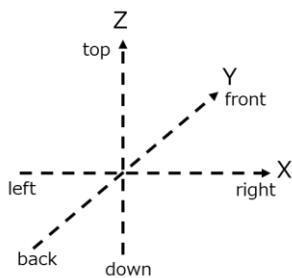


図 2: ADM におけるデカルト座標

3. 6DoF 対応の音響メタデータの記述法

3.1 リスナと発音源の位置

リスナが 6DoF コンテンツ空間内を自由に動くため, リスナの位置に拠らない発音源の絶対的な位置をグローバル座標で記述する必要がある. また, 6DoF コンテンツ空間上を動くあるオブジェクトを中心として別のオブジェクトが周囲を回る場合など, オブジェクト間に依存関係がある場合がある. そのような場合では, 複数のオブジェクトの動きを考慮した絶対的な位置を記述するよりも, あるオブジェクトを基準としたローカル座標系を導入した方がコンテンツ制作者による記述が容易になる. オブジェクトの位置を規定する座標系の原点 (referencePosition) を別のオブジェクトとしたい場合は audioObjectIDRef という Element に基準にしたいオブジェクトの識別子 (audioObjectID) を記述し, リスナの位置を原点としたい場合は listenerObjectIDRef という Element にリスナの識別子 (listenerObjectID) を記述し, また, ある位置を原点としたい場合は referencePosition の Sub-element に position を追加する. position という Element に空間の特定位置を原点とした位置あるいは referencePosition で指定した原点からの相対位置を記述する. 表 1-1 に提案するオブジェクトの位置に関する Element と Sub-element を表し, 表 1-2, 1-3 にはそれぞれデカルト座標と極座標における表 1-1 の position の Attribute を示す. 従来の ADM に追加した Element と Attribute を太い赤字で示している. ここで, 従

来の ADM でも使用されている position は, オブジェクトを構成する音声信号毎に時系列情報を記述可能な Element, audioBlockFormat の Sub-element (以下, audioBlockFormat の Sub-element) である. 従来の単位 (units) をデフォルトとすることで, 単位の記述が省略されている場合はデフォルト値を使用して position の記述について後方互換性を保つことができる.

表 1-1: 提案するオブジェクトの位置に関する Element と Sub-element

| Element | Sub-element |
|--------------------------|----------------------------|
| referencePosition | audioObjectIDRef |
| | listenerObjectIDRef |
| | position |
| position | |

表 1-2: デカルト座標における position の Attribute

| Attribute | Value |
|------------|---|
| coordinate | "X" |
| | "Y" |
| | "Z" |
| units | "meters" /"Relative Units" (default) (for coordinate="X","Y","Z") |

表 1-3: 極座標における position の Attribute

| Attribute | Value |
|------------|--|
| coordinate | "azimuth" |
| | "elevation" |
| | "distance" |
| units | "degrees" (for coordinate="azimuth","elevation") |
| | "meters" /"relative distance value"(default) (for coordinate="distance") |

3.2 放射パターン

人声や楽器などの発音源から放射される音は方向ごとに異なる周波数特性を持つため, 6DoF コンテンツ空間上でリスナが発音源を様々な方向から視聴する際には方向に応じた周波数特性を再現することでリアルな音を表現することができる (図 3). ここで周波数ごとの放射パターンデータ (以下, 放射パターン) の記述および付与方法を検討した.

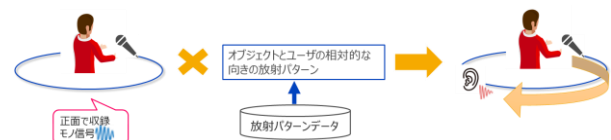


図 3: 放射パターンの使用例

3.2.1 放射パターンの記述

放射パターンは方向(azimuth, elevation), 周波数別のゲイン情報で構成される. 記述は大きく分けて 2 種類考えられる. 1 種類目は方向, 周波数別のゲイン情報をそのまま記述する方法であり, 2 種類目は記述量の削減を考

慮してそれらをパラメトリックに記述する方法である。

放射パターンに関する Element として `directivity` を定めて、放射パターンを構成する各データをその Sub-element, `directivityPattern` とする。表 2-1~2-6 に放射パターンをそのまま記述する場合とパラメトリックに記述する場合の Element と Attribute を表す。

表 2-1: 放射パターンに関する Element と Sub-element

| Element | Sub-element |
|-------------|---|
| directivity | directivityPattern |
| | functionType (0:none(default) 1:sphericalfunction) |

表 2-2: `directivityPattern` の Sub-element

(左 : `functionType` が 0, 右 : `functionType` が 1 の場合)

| Sub-element | Sub-element |
|-------------|----------------------|
| direction | directionalParameter |
| frequency | frequency |
| gain | coefficient |

表 2-3: `direction` の Attribute

| Attribute | Value |
|------------|---|
| coordinate | "azimuth" |
| | "elevation" |
| units | "degrees" (for coordinate="azimuth","elevation") |

表 2-4: `directionalParameter` の Attribute

| Attribute | Value |
|-----------|----------|
| parameter | "order" |
| | "degree" |

表 2-5: `frequency` の Attribute

| Attribute | Value |
|-----------|-------|
| units | "Hz" |

表 2-6: `gain` の Attribute

| Attribute | Value |
|-----------|----------------|
| units | "linear"/ "dB" |

2 種類目のパラメトリックな記述法のうち、球面調和関数を用いる方法では、放射パターンデータそのものではなく、式(1)で示す球面調和展開の展開係数 $A_{nm}(\omega)$ を記述する。ここで係数 $A_{nm}(\omega)$ は $Y_n^m(\theta, \varphi)$ を球面調和関数として音源を取り囲む球面上の音圧分布を球面調和展開した時の展開係数である。 θ は方位角、 φ は仰角、 m は位数(degree)、 n は次数(order)、 ω は周波数を表す。レンダーラは係数に基づいて重みづけをした球面調和関数の線形和で放射パ

ターンデータ (式 1 の `directivityData`) を生成して用いる。

例えば、密に測定した精緻なデータをそのまま使用する場合は 1 種類目の記述法を用いて、球面調和展開の展開次数を打ち切ることでデータ量を削減し 1 種類目の方法よりは精緻でなくても十分な場合は 2 種類目の記述法を用いるというような使い分けが考えられる。

$$directivityData(\theta, \varphi, \omega) = \sum_{n=0}^N \sum_{m=-n}^n A_{nm}(\omega) Y_n^m(\theta, \varphi) \quad (1)$$

3.2.2 放射パターンの付与の記述

オブジェクトに放射パターンを付与するために `directivity` の Attribute に `directivityID` を追加し、放射パターンの識別子を記述する。さらに `audioBlockFormat` の Sub-element に `directivityIDRef` を追加し、`directivityID` を記述する。

3.3 発音源の向き

放射パターンを持つ発音源をリスナが様々な方向から視聴するためには、3.2 節で述べた放射パターンや発音源とリスナの位置に加えて発音源の向きの情報も必要である。そこで `audioBlockFormat` の Sub-element に向きに関する `orientation` を追加する。表 3 に `orientation` の Attribute を示す。"yaw", "roll", "pitch" はオブジェクトの位置を原点に 6DoF コンテンツ空間上のそれぞれ Z 軸, Y 軸, X 軸回りの反時計回りの回転角度を表している。ただし、`position` でローカル座標系を用いる場合はオブジェクトの位置を原点に基準となる向きは `referencePosition` で指定されたオブジェクトやリスナのローカル座標系における Z 軸, Y 軸, X 軸と同じ方向とする。

表 3: `orientation` の Attribute

| Attribute | Value |
|-----------|--|
| rotation | "yaw" |
| | "roll" |
| | "pitch" |
| units | "degrees" (for rotation="yaw","roll","pitch") |

4. 距離減衰特性

自由音場において、点音源から発せられたエネルギーは距離の二乗に反比例して減衰することが知られている[7]。6DoF コンテンツ空間においても同様に発音源の位置からリスナ位置までの距離に基づいた距離減衰特性を定義式(2)の通りに付与できるように、距離減衰特性に関する Element として `distanceAttenuation` を定めた。表 4 に距離減衰特性に関する Element と Sub-element を示す。ここで、 s は発音源の音声信号、 $\hat{s}(r)$ は発音源から距離 r メートルだけ離れたリスナ位置における音声信号、 k は振幅の正規化係数 (normalizationCoefficient)、 q は減衰定数 (attenuationConstant) である。 k は発音源の音声信号がどの距離を基準として距離減衰特性を計算するかの基準距離

に相当する. q の値を変えることで減衰カーブを変えることができ, 1 の場合は点音源を表す. 1 より小さい場合は減衰が緩やかになり 1 より大きい場合は急減衰し, 制作者の意図, 演出によって設定が可能である.

$$\hat{s}(r) = s \cdot \left(\frac{k}{r}\right)^q \quad (2)$$

表 4: 距離減衰特性に関する Element と Sub-element

| Element | Sub-element |
|---------------------|--------------------------|
| distanceAttenuation | normalizationCoefficient |
| | attenuationConstant |

放射パターンと同様に, 定めた距離減衰特性については distanceAttenuation の Attribute として distanceAttenuationID に距離減衰特性の識別子を記述する. さらに audioBlockFormat の sub-Element に distanceAttenuationIDRef を追加し, 付与したい distanceattenuationID を記述する.

5. コンテンツの記述例

提案した記述法を用いた 6DoF コンテンツの記述例の一部を図 4 に表す. 従来の記述法からの変更点は太い赤字で

記す.

6. むすび

本稿では音響メタデータ ADM を拡張して 6DoF に対応させるためにリスナと発音源の位置, 放射パターンと発音源の向き, および距離減衰特性に関するメタデータの記述法を開発し, その記述例を示した.

参考文献

- [1] 長尾慈郎・今中秀郎, 杉本岳大, 谷本正幸, 青木秀一: より自由でより没入感の高いイマージブメディア, 情報処理学会誌, vol.62, No.5, 2021.
- [2] ISO/IEC 23090-4: Immersive Audio (規格化中)
- [3] 杉本岳大, 長谷川知美: 6DoF 音響コンテンツの制作要件, 映像情報メディア学会年次大会, 2023.
- [4] 杉本岳大, 長谷川知美, 大出訓史: 6DoF 音響サービスに用いる音響メタデータの要求条件, 日本バーチャルリアリティ学会大会 論文集, 2023.
- [5] CEDEC2022, <https://cedec.cesa.or.jp>, 2022.
- [6] Rec. ITU-R BS.2076-2: Audio Definition Model, 2019.
- [7] 伊藤毅: 音響工学原論, コロナ社, 1957.

```
<audioChannelFormat audioChannelFormatID="AC_00031001" audioChannelFormatName="D1g" typeDefinition="Objects" typeLabel="0003">
  <audioBlockFormat audioBlockFormatID="AB_00031001_00000001" rtime="00:00:00.00000" duration="00:00:19.03500">
    <position coordinate="X" Units="meters">0.50</position>
    <position coordinate="Y" Units="meters">0</position>
    <position coordinate="Z" Units="meters">1.5</position>
    <orientation rotation="yaw" Units="degrees">90</orientation>
    <distanceAttenuationIDRef>DA_1001</distanceAttenuationIDRef>
    <directivityIDIDRef>DI_1001</directivityIDIDRef>
  </audioBlockFormat>
</audioChannelFormat>
<audioTrackUID UID="ATU_00000001" bitDepth="24" sampleRate="48000">
  <audioPackFormatIDRef>AP_00031001</audioPackFormatIDRef>
  <audioChannelFormatIDRef>AC_00031001</audioChannelFormatIDRef>
</audioTrackUID>
<acousticProperties>
  <directivity directivityID="DI_1001">
    <directivityPattern>
      <direction coordinate="azimuth" Units="degrees">90</direction>
      <frequency Units="Hz">1000</frequency>
      <gain Units="linear">0.5</gain>
    </directivityPattern>
    <directivityPattern>
      <direction coordinate="azimuth" Units="degrees">90</direction>
      <frequency Units="Hz">4000</frequency>
      <gain Units="linear">0.3</gain>
    </directivityPattern>
  </directivity>
  <distanceAttenuation distanceAttenuationID="DA_1001">
    <attenuationConstant>1</attenuationConstant>
    <normalizationCoefficient>0.3</normalizationCoefficient>
  </distanceAttenuation>
</acousticProperties>
```

図 4: 6DoF コンテンツの記述例