

**TTC標準**  
Standard

JT-H430.2

超高臨場ライブ体験  
(ILE: Immersive Live Experience) :  
アーキテクチャフレームワーク

Immersive Live Experience:  
Architectural framework

第 1.0 版

2020 年 2 月 20 日制定

一般社団法人  
情報通信技術委員会

THE TELECOMMUNICATION TECHNOLOGY COMMITTEE



本書は、一般社団法人情報通信技術委員会が著作権を保有しています。  
内容の一部又は全部を一般社団法人情報通信技術委員会の許諾を得ることなく複製、転載、改変、転用及びネットワーク上での送信、配布を行うことを禁止します。

## 目次

<参考>.....	4
1. 規定範囲.....	5
2. 参考文献.....	5
3. 定義.....	5
3.1 他の標準にて定義された用語.....	5
3.1.1 超高臨場ライブ体験 (ILE: Immersive Live Experience) [ITU-T H.430.1].....	5
3.2 本標準にて定義する用語.....	5
4. 略称.....	5
5. 慣例.....	6
6. ILEの基本アーキテクチャ.....	6
6.1 ILEシステムの基本アーキテクチャ.....	6
6.2 本会場での環境情報取得機能.....	7
6.3 視聴会場での表示機能.....	8
6.4 メディア処理の機能.....	9
6.5 伝送部分の機能.....	9
7. ILE機能の候補技術.....	9
7.1 波面合成技術を用いた超高臨場音声.....	9
7.2 超高解像度のサラウンド表示のためのビデオ結合.....	10
7.3 メディア伝送機能.....	10
7.4 任意背景の実時間オブジェクト抽出機能.....	10
8. ILEサービスを提供する一般的役割モデル.....	10
Appendix I Examples of candidate technologies for functions of ILE.....	13
I.1 Implementation of WFS.....	13
I.2 Implementation of super high-definition video stitching.....	14
I.3 Implementation of media transport functions.....	14
I.4 Implementation of arbitrary background real-time object extraction function.....	15
文献一覧.....	17

## <参考>

### 1. 国際勧告などとの関連

本標準は超高臨場ライブ体験（ILE: Immersive Live Experience）のアーキテクチャフレームワークについて規定しており、2018年7月にITU-T SG16において発行されたITU-T勧告H.430.2に準拠している。

### 2. 上記勧告などに対する追加項目など

#### 2.1 オプション選択項目

なし

#### 2.2 ナショナルマター決定項目

なし

#### 2.3 その他

なし

#### 2.4 原勧告との章立て構成比較表

章立てに変更なし

### 3. 改版の履歴

版数	発行日	改版内容
第1版	2020年2月20日	制定

### 4. 工業所有権

本標準に関わる「工業所有権の実施の権利に係る確認書」の提出状況は、TTC ホームページでご覧になれます。

### 5. その他

(1) 参照している勧告、標準など

TTC 標準

ITU-T 標準 H.430.1, H.430.2, H.430.3, H.430.4

ISO/IEC 標準

### 6. 標準作成部門

マルチメディア応用専門委員会 ILE-SWG

## 1. 規定範囲

本標準は、超高臨場ライブ体験（ILE: Immersive Live Experience）の基本的な機能アーキテクチャを含むアーキテクチャフレームワークを規定する。また、ILE に要求される機能について、いくつかの候補技術を示す。また、本会場におけるイベントプロモータ、伝送部分のメディア伝送事業者、視聴会場におけるパブリックビューイングやライブビューイングプロモータから成る一般的な役割モデルを定義する。

本標準の既定範囲は以下を含む。

- ・ ILE の基本的アーキテクチャ
- ・ ILE に要求される候補技術
- ・ ILE サービスを提供する一般的な役割モデル

## 2. 参考文献

以下の ITU-T 勧告およびその他の参考文献には、規定条項が含まれており、本標準の本文で参照することによって、本標準の規定条項を構成することになる。出版の時点では、表示されている版が有効である。これら全ての標準や勧告とその他の参考文献は、改定される可能性があるため、本標準の利用者は、以下に示された標準、勧告および参考文献の最新版の適用可能性を確認することを推奨する。現在有効である ITU-T 勧告リストは定期的に発行されている。

- [ITU-T H.430.1] Recommendation ITU-T H.430.1 (2018), Requirements for immersive live experience (ILE) services.
- [ITU-T H.430.2] Recommendation ITU-T H.430.2 (2018), Architectural framework for immersive live experience (ILE) service.
- [ITU-T H.430.3] Recommendation ITU-T H.430.3 (2018), Service scenario of immersive live experience (ILE).
- [ITU-T H.430.4] Recommendation ITU-T H.430.4 (2019), Service configuration, media transport protocols and signaling information of MMT for Immersive Live Experience (ILE) systems.

## 3. 定義

この標準では以下の用語を定義する。

### 3.1 他の標準にて定義された用語

この標準では、他の標準にて定義された以下の用語を用いる。

#### 3.1.1 超高臨場ライブ体験（ILE: Immersive Live Experience） [ITU-T H.430.1]

センサ情報収集、メディア処理、メディア伝送、メディア同期、メディア表示などのマルチメディア技術の組み合わせで実現された高臨場感により、あたかも遠隔会場の観客が実際のイベント会場に入り、観客の目の前で実際のイベントを見ているかのように、イベント会場と遠隔会場の両方の観客の感動を刺激する共感視聴体験。

### 3.2 本標準にて定義する用語

この標準で新たに定義する用語はない。

## 4. 略称

本標準では、下記の略称を使用している。

ILE Immersive Live Experience

CODEC Coding and decoding

MPEG Moving Picture Expert Group

MMT MPEG Media Transport

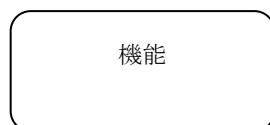
WFS Wave Field Synthesis

DMX Digital Multiplex

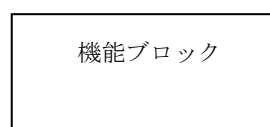
## 5. 慣例

本標準では、

- 「求められる」は、厳密に従うべき要求条件を示す。本標準へのコンフォーマンスを訴求するには、この要求条件からの逸脱は許されない。
- 「推奨される」は、推奨されるが無条件に「求められる」のではない要求条件を示す。従って、この要求条件はコンフォーマンスの訴求のために必ずしも示す必要はない。
- 「オプションで対応する」は、「推奨される」のではなく容認可能な任意の要求条件を示す。ベンダがそのオプションを（選択できるように）実装したり、その特徴がネットワークオペレータやサービスプロバイダから任意の方法で提供されたりすることを意図するものではない。むしろ、ベンダが任意にこの特徴を提供するかもしれないし、本標準に従ってコンフォーマンスを訴求するために用いるかもしれないことを意味する。
- 「機能」は、機能要素の集まりで定義される。本標準では、以下のシンボルで表現される。



- 「機能ブロック」は、この標準の中での記述における詳細レベルでは、これより細かく分割されない機能要素の集合として定義される。以下のシンボルで表現される。



注 - 将来、他の標準でこれらの機能ブロックをさらに分割することがある。

「機能」と「機能ブロック」の境界線と「機能」と「機能ブロック」の関係線は実線もしくは破線で描かれる。実線は機能要素や関係が要求されることを意味し、破線は任意の機能要素や関係であることを意味する。

## 6. ILE の基本アーキテクチャ [ITU-T H.430.2]

### 6.1 ILE システムの基本アーキテクチャ

ライブでのコンテンツ配信を含む ILE サービスを提供するため、アーキテクチャには環境情報を取得する機能、表示機能、ILE アプリケーション、同期メディア伝送、伝送レイヤが必要となる。ILE システムの基本アーキテクチャを図 1 に示す。本会場においてカメラ、マイク、照明情報を有するセンサで取得された環境情報を統合したコンテンツは、同期して視聴会場に伝送される。このコンテンツは、メディア処理ブロックのアセット処

理と CODEC により処理され、信号処理ブロックで空間情報処理と同期情報処理により信号が再構成される。これらのブロックは、ILE アプリケーションに配置される。視聴会場では、配信されたコンテンツは表示装置で再構成され、実際と同じ大きさかほぼ同じ大きさでプロジェクタかディスプレイに表示される。音声は、音声の方向を再構成される。また、受信した照明情報から照明が再構成される。これらの情報を視聴会場で同期して再生することで、視聴者は高臨場感を得られる。

超高臨場ライブ体験（ILE: Immersive Live Experience）のサービスシナリオは、[ITU-T H.430.3]にて規定されている。

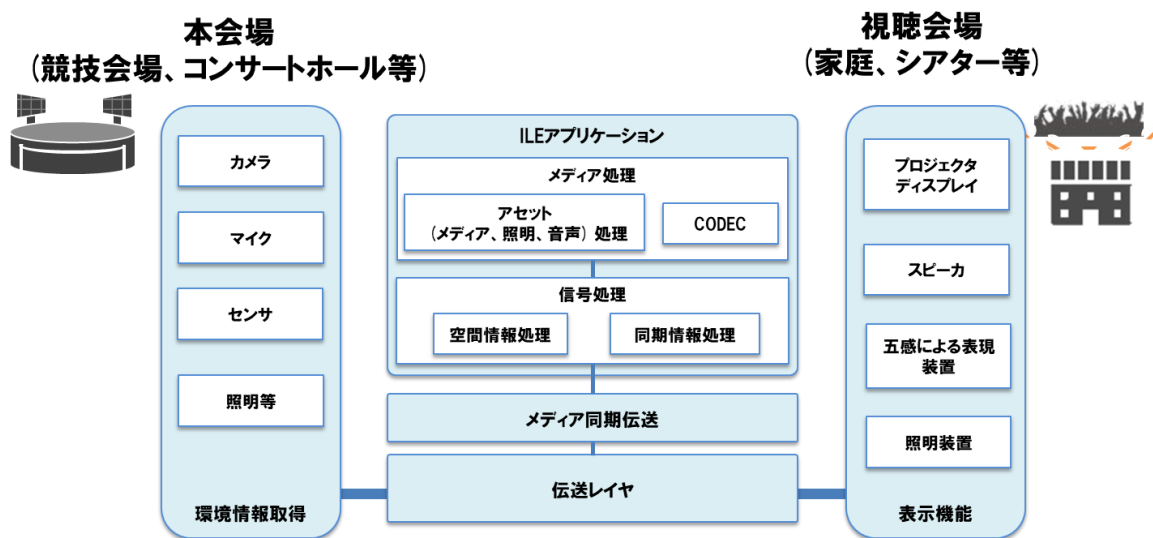


図 1 : ILE システムの基本アーキテクチャ

## 6.2 本会場での環境情報取得機能

### (1) カメラ

視聴会場での擬似 3 次元映像の表示のためには、実物大での表示を考慮し、複数のカメラによる立体的な映像を取得すること、また、取得した映像に空間的な位置情報を追加することが求められる。使用するカメラはできるだけ高解像度にすべきである。

### (2) マイク

視聴会場で表示されたオブジェクトから音声聞こえるよう音声の方向を再構成するためには、本会場において複数のマイクを利用してオブジェクトからの音声を取得し、取得した音声に各オブジェクトの空間的位置情報を追加することが求められる。音声方向は、視聴会場のディスプレイに表示されたオブジェクトに合わせるべきである。

### (3) センサ

視聴会場で視聴者に高臨場感を提供するため、映像や音声に加えて温度や湿度のような環境情報を提供することが望ましい。このような環境情報はいくつかの種類センサにより取得する。例えば、本会場の振動を伝送し、センサにより取得した様々な環境情報により視聴会場で振動を再構成できる可能性がある。

センサ情報の他の利用方法として、視聴会場でより現実的なオブジェクト表示のため、本会場において競技者などのオブジェクトの 3 次元位置情報を計測することや、奥行きセンサ等のいくつかのセンサを組み合わせることでオブジェクトを追跡することが求められる。

### (4) 特殊効果およびその他

通常、音楽コンサートなどの多くのイベントでは、照明などの特殊効果や舞台演出が使われる。パブリック

ビューイングやライブビューイングの会場は、たいていの場合、音楽コンサート会場やスポーツスタジアムのようなイベント会場とは異なっている。視聴会場よりイベント会場に似せた再構成するために、照明情報などの特殊効果や舞台演出の情報は視聴会場の視聴者に高臨場感を増加させることに有効である。

### 6.3 視聴会場での表示機能

#### (1) プロジェクタとディスプレイ

視聴会場には、劇場型、オープン型、アリーナ型のようにいくつかの種類がある。これらの種類を図2に示す。

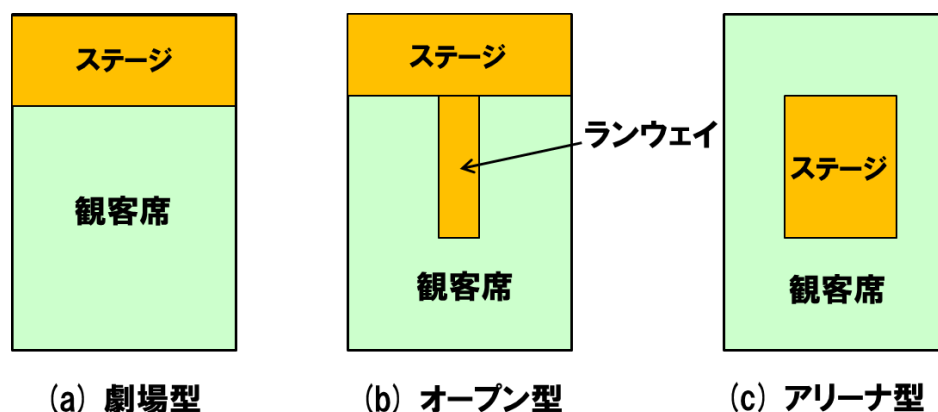


図2 視聴会場の種類

#### a) 劇場型

多くの劇場では、視聴者は観客席の前に設置された1つのステージかスクリーンを視聴する。図2(a)に示すように、ステージは劇場の1つの面に配置され、観客席はステージと反対側に設置される。これを劇場型と呼ぶ。

多くの視聴会場は、ホールの大きさや観客席数などイベント会場とは完全に異なる環境となる。また、表示装置はたくさんの異なる仕様（解像度や大きさ）を有する様々な種類がある。装置の空間的な位置（垂直・水平方向の位置や傾き）は、視聴会場毎に異なっている。

視聴会場においてオブジェクトやイベント会場の環境を再構成するため、端末装置の能力や空間的な位置情報を含む空間環境情報が必要となる。さらに、実物大のオブジェクト表示には、床面からの高さや隣接する表示装置からの幅などの空間位置も要求される。

#### b) オープン型

多くの音楽コンサートホールでは、図2(b)に示すように1つまたは複数のステージやランウェイが存在する。このため、観客はそれぞれ異なる方向を見ることになる。これをオープン型と呼ぶ。劇場型に1つまたは複数のランウェイを有するもの、劇場型の真ん中に1つまたは複数のアイランド型のステージを有するものなど、オープン型は多くの種類がある。

視聴会場でイベント会場の環境を再構成するために、イベント会場で抽出されたオブジェクトや環境映像の空間的な位置情報を提供するための、多層型プロジェクションや三次元的な表示のための複数スクリーン表示などを考慮することが求められる。

#### c) アリーナ型

アリーナ型は1つのフィールドと劇場の中心に据えられた1つのステージを有する。観客は、ステージを囲む全ての角度からステージを見ることができる（図2(c)に示すようにステージ上のオブジェクトを360度どの角度からも見られる）。

アリーナ型視聴会場のステージを囲む多くの視聴者に対し360度どの角度でも三次元のオブジェクト視聴を提供するため、擬似三次元表示システムにはオブジェクトの位置を扱うことが求められる。三次元空間プロジェ



クシオンは、HMDを用いずに奥行方向へのなめらかなオブジェクトの移動ができる三次元オブジェクト表示を可能とする。

#### (2) スピーカ

視聴会場での音声の方向を再構成するため、イベント会場で抽出したオブジェクトの空間情報や音声データを利用することで、ディスプレイに表示されたオブジェクトと音声の方向を合わせる必要がある。

#### (3) 五感による表現装置

視聴会場の視聴者は、最大限イベント会場の環境を体感したいと考えるため、イベント会場のセンサで取得した様々なデータを活用して、視聴会場を再構成するべきである。

また、イベント会場から送られてきた照明情報を含む特殊効果や舞台演出の情報は、イベントの雰囲気視聴者が体感できるよう、視聴会場においてイベント会場の空間を再構成することを可能とする。視聴会場ですらに臨場感を高めるため、視聴会場の照明装置はイベント会場の照明装置と同様なものを用い、イベント会場から送られた照明情報を使うべきである。

### 6.4 メディア処理の機能

#### (1) メディア処理

メディア処理ブロックのアセットの処理機能は、ビデオと音声等のコンテンツを抽出、再構成することにより、メディアフレームを含むアセットを作成する。メディア処理ブロックの CODEC は、アセット情報をエンコード及びデコードする。

#### (2) 信号処理

信号処理ブロックの空間情報機能は、視聴会場のディスプレイの物理的な大きさと位置から空間情報を作成する。同期情報処理は、ビデオ、音声、照明、センサ情報などのコンテンツを同期させるため、全てのコンテンツ情報の時刻情報を再形成する。

### 6.5 伝送部分の機能

#### (1) メディア同期伝送

メディア同期伝送機能は、いくつかのメディアを同期して配信する機能である。ILE サービスの提供、特にライブでの擬似三次元コンテンツの表示と音声方向の再現をするために、いくつかの拡張を施した MPEG Media Transport (MMT) のようなメディア同期伝送技術と端末装置側での表示機能を検討する必要がある。ISO/IEC JTC 1/SC 29/WG 11 (MPEG) で標準化された MMT は、アセット情報の複数のメディアを伝送する能力を持っているが、空間情報のような環境情報を伝送するための仕様を規定する必要がある。そこで、端末装置で環境情報をメディアと同期して表示するために、環境情報を扱うための MMT の利用方法 (プロファイル) を規定する必要がある。[ITU-T H.430.4]

#### (2) 伝送レイヤ

伝送レイヤは、イベント会場から視聴会場にメディアを伝送する。コンテンツ保護の安全性とネットワークの遅延を考慮すると、このレイヤは NGN や IMT-2020 によって実現すべきである。

## 7. ILE 機能の候補技術

### 7.1 波面合成音響技術を用いた超高臨場音声

音場の再構成技術の 1 つに波面合成音響技術 (Wave Field Synthesis : WFS) がある。音響サラウンドシステムのための WFS のスピーカレイは、臨場感を拡大するための一つの方法である。WFS は、二次音源を用いることにより、任意の音場を再構成することが目的である。WFS の典型的な実現例は、平面的な音場の再構成のために線形に配列した二次音源を使うものである。

## 7.2 超高解像度のためのサラウンド映像合成

超広角ビデオによる高臨場感を実現するためには、従来の 16:9 のテレビは ILE サービスには不十分である。そのため、水平面に並べた複数の 4K カメラを使ってビデオを取得するための新しいアーキテクチャが求められる。図 3 は、5 台の 4K カメラを使った実時間画面統合システムの例である。実時間での 4K の画像データの大量高速処理のため、データは分割され、前のフレームからの位置合わせ情報を待たずにフレームの連続処理を可能とする機構が使われ、これによりフリッカを抑制する。こうした技術的革新により、4K 60p の実時間処理が可能となる。

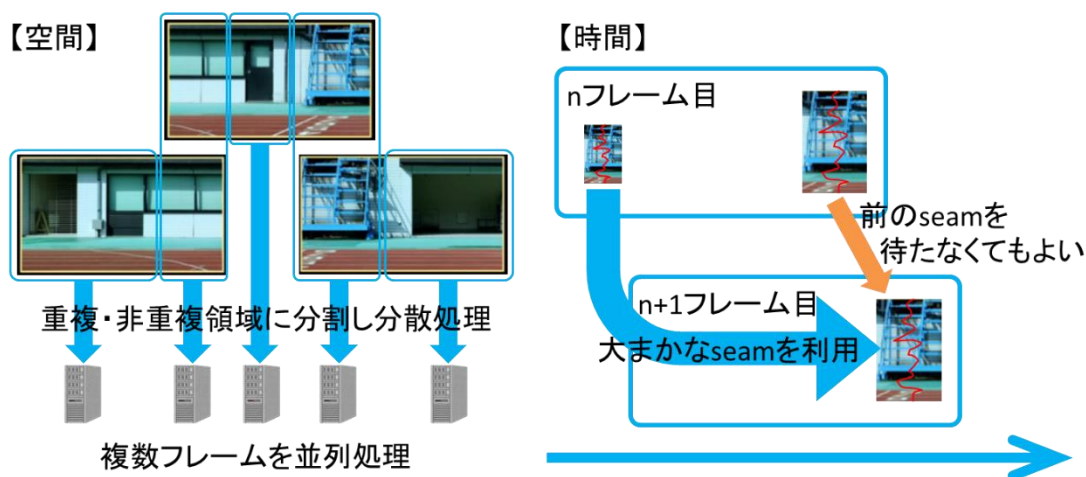


図 3 サラウンド映像合成

## 7.3 メディア同期伝送機能

ILE サービスの主要な特徴の一つに、空間情報のデータストリームや他の舞台演出情報を伴うビデオや音声のような複数のライブストリームコンテンツを伝送することが挙げられる。MMT[b ISO/IEC 23008-1]は、これらのメディア情報を同期して伝送するための情報ストリームに使える。[ITU-T H.430.4]

## 7.4 任意背景リアルタイム被写体抽出機能

擬似三次元映像のような高臨場感映像を視聴会場で再構成するため、競技者や演技者等の被写体を、イベント会場で取得したビデオからリアルタイムに抽出する必要がある。このため、通常はクロマキーと呼ばれる特別な写真技術が利用される。クロマキーでは、青や緑の単色のスクリーンを被写体の背後に配置し、被写体と背景スクリーンとの色の違いによりオブジェクトを抽出する。しかし、この技術はスポーツの試合や音楽コンサートの実際のイベントでは、背景が単色ではないため使うことができない。ILE サービスでは、試合会場やステージでも被写体の背景を含むビデオから、リアルタイムに任意背景の被写体抽出が必要となる。

## 8. ILE サービスを提供する一般的役割モデル

この節では、ILE サービスを提供するための一般的な役割モデルを記載する。図 4 に、基本的なアーキテクチャを元にし、いくつかのインタフェースを有する一般的な役割モデルを示す。このモデルは、以下の 4 つの主要な役割から成る。

- コンテンツプロバイダもしくはコンテンツ製作者であるプロダクション
- ILE サービスのためのメディア処理
- メディア伝送
- 視聴会場でのプレゼンテーション

3つのインタフェースを有する。

- IF1 : プロダクションとメディア伝送間
- IF2 : メディア伝送とプレゼンテーション間
- IF3 : メディア伝送とメディア処理間

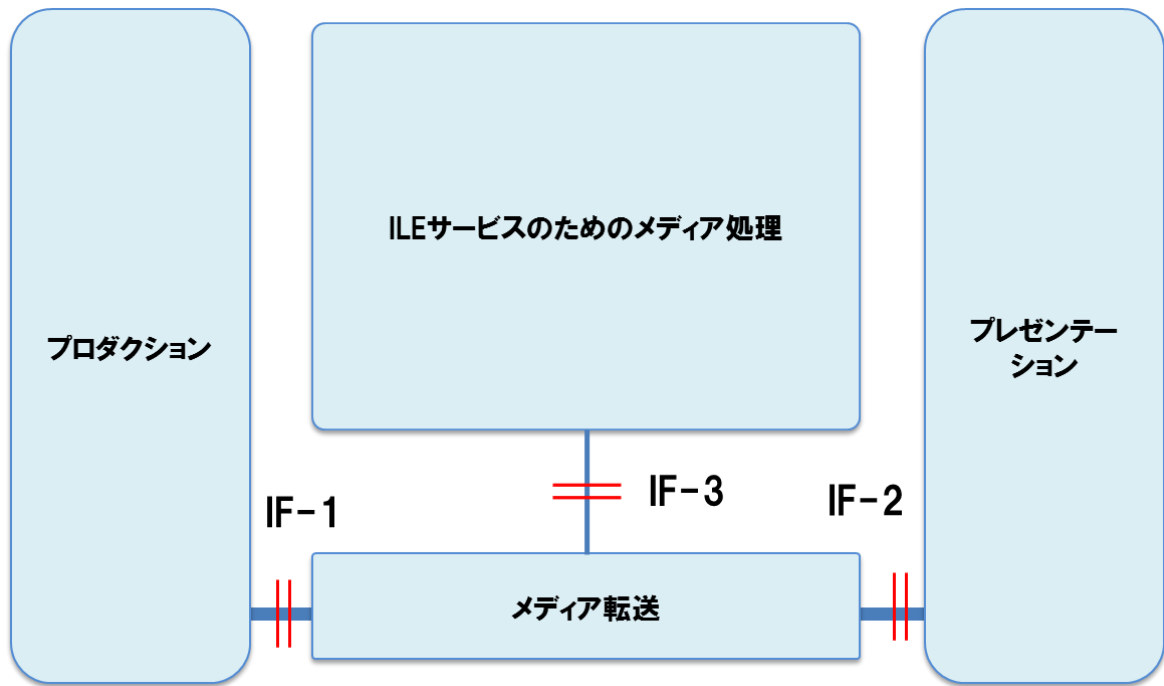


図4 ILE サービス提供のための一般的役割モデル

**IF1**

このインタフェースは、ビデオ、音声、照明、空間情報のストリーミングを含むソースコンテンツ情報として定義される。

**IF2**

このインタフェースは、各視聴会場用に設計されたアセット情報として定義される。

**IF3**

このインタフェースは、ソースコンテンツ情報とアセット情報を含むメディア処理情報として定義される。

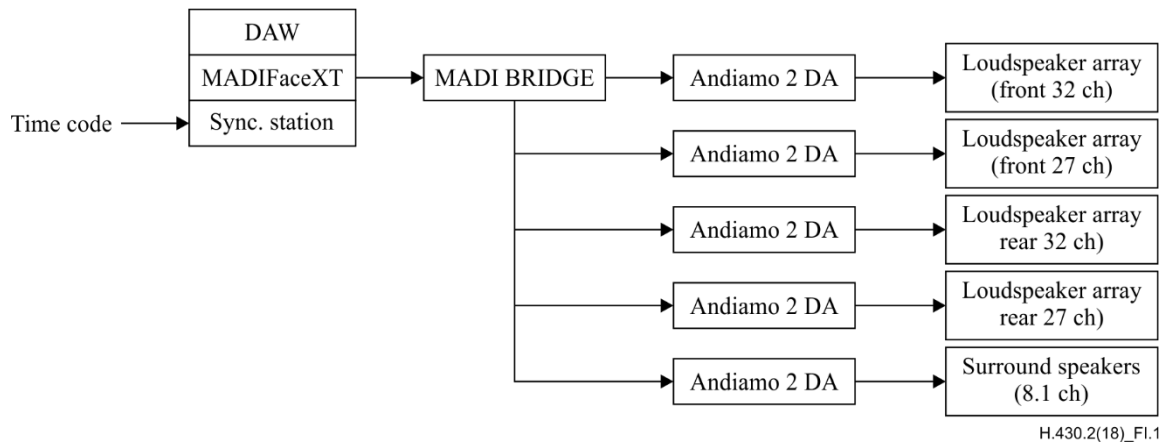
## Appendix I

### Examples of candidate technologies for functions of ILE

(This appendix does not form an integral part of this Recommendation.)

#### I.1 Implementation of WFS

Figure I.1 shows a block diagram of one implementation of WFS for an ILE system [b-Tsutsumi]. Figure I.2 shows an implementation of a front array for audio spatialization installed at the bottom of a stage (the box surrounded by the dotted line) and a rear loudspeaker array for audio spatialization special effects is shown in Figure I.3. In this example, there are eight loudspeakers and one sub-woofer for surround sound reproduction. All the loudspeakers were connected to each other with Ethernet in a daisy chain configuration and controlled by digital audio network through Ethernet (DANTE) protocol. However, audio signals were fed through XLR cables from interfaces (Andiamo 2 XT by DirectOut Technologies).



**Figure I.1 – Block diagram of immersive sound reproduction system**



**Figure I.2 – Front array embedded at the bottom of the stage (box with dotted line)**



H.430.2(18)\_F1.3

**Figure I.3 – Rear loudspeaker array**

A linear distribution of secondary sources was used along the x-axis for reproducing sound fields of focused sources defined at  $(x, y) = (0.0, 1.0)$ , emitting monochromatic signals of 1 kHz. Assuming sound velocity as 343 m/s, the distance between reference lines and the x-axis was set to 3 m and the distance between loudspeakers was set to 0.075 m. Fifty-nine loudspeakers along the x-axis were used as secondary sources.

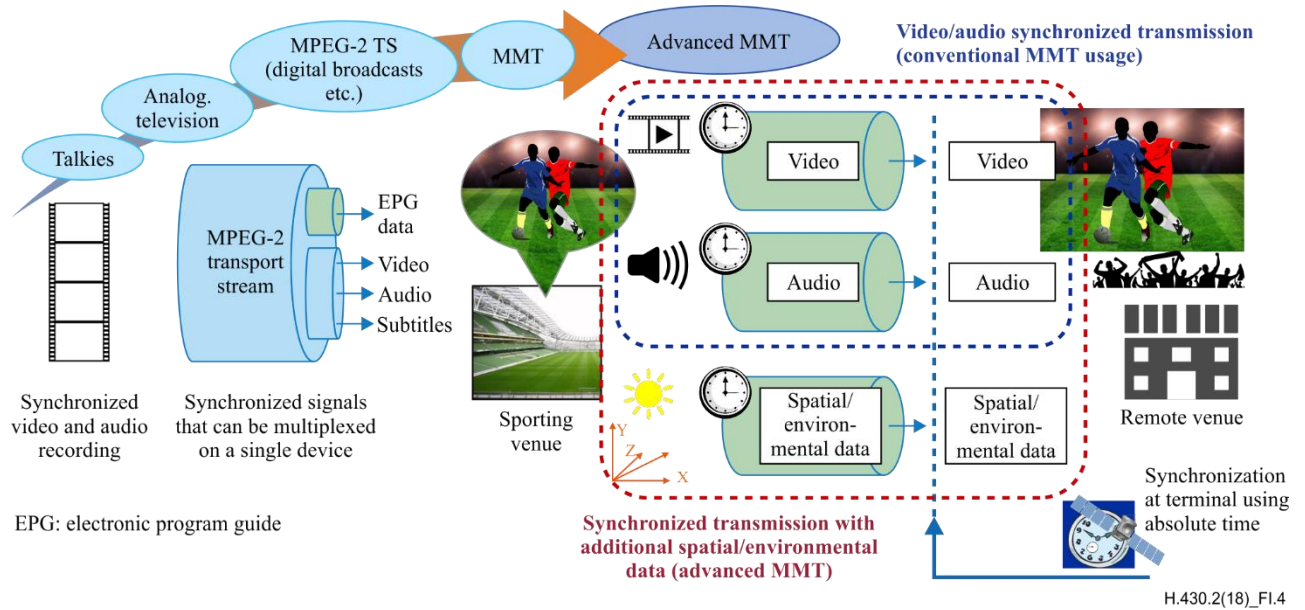
Since the sound field was assumed to be 1.1 m in height, i.e. the height of the ears of the sitting audience members, all the secondary sources were arranged in one line at the height of the loudspeakers (0.9 m).

#### I.2 Implementation of super high-definition video stitching

Details of an implementation of super high-definition video stitching are shown in [b-Akutsu].

#### I.3 Implementation of media transport functions

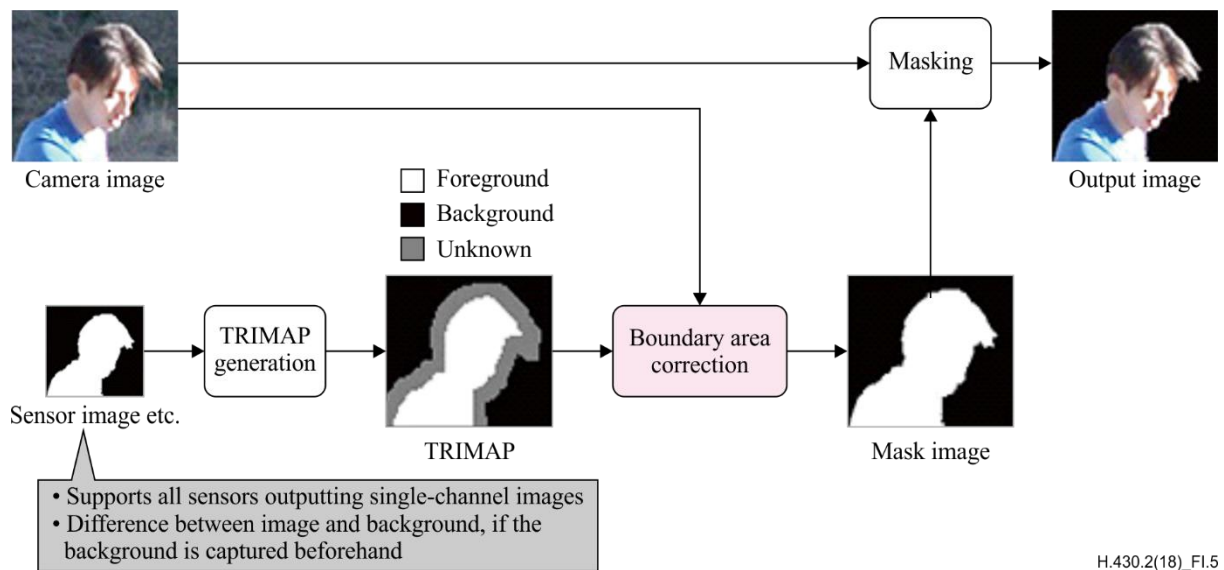
In order to synchronize and transport video and audio combined with spatial information, MMT, an optimized protocol for media synchronization, can be utilized with definitions for MMT signalling to describe 3D information such as the size, position and direction of MMT assets, see Figure I.4. This technology makes it possible to correlate physical spatial parameters such as the size and position of the display device with asset data (frame pixel data) so that the space can be reconstructed with high realism at the destination at the intended size. In addition, transmission of the digital multiplex (DMX) signals commonly used in production to control stage lighting and audio devices together with the MMT assets enables realistic presentations that accurately synchronize remote stage equipment with the media [b-Akutsu].



**Figure I.4 – Media synchronization for realistic sensation (Advanced MMT)**

#### I.4 Implementation of arbitrary background real-time object extraction function

An overview of the actual implementation of an arbitrary background object extraction process is shown in Figure I.5 [b-Nagata]. The object extraction uses sensors in addition to the imaging camera to capture the object. This enables constraints on object extraction to be eliminated so that the range of applications can be expanded as much as possible in the future.



**Figure I.5 – Overview of arbitrary background object extraction process**

Processing can be divided broadly into two stages. In the first stage, a sensing device and the background image (only if it can be captured ahead of time) are used to roughly identify the object area. In this stage, data (a trimap) are generated and pixels are labelled into three categories: foreground, background and unknown. Then, pixels in the trimap in the region labelled 'unknown' (for which the foreground and background could not be distinguished during the rough identification stage and is assumed to contain the object boundary) are rapidly further classified as foreground and background and corrections are applied to the boundary.

In the second stage, some objects targeted for extraction have an extremely complex boundary, for example target objects are with untargeted objects in background. The boundaries of target objects must be represented in fine detail to avoid

degrading the sense of realism. In such cases, pixels in the unknown region are classified as foreground or background, but they are also assigned a transparency value called an  $\alpha$  (alpha) value, which provides a more natural-looking boundary. Further work on the arbitrary background real-time object extraction technology is required to increase speed, robustness and detail.



## 文献一覧

- [b-ITU-T H.430.1] Recommendation ITU-T H.430.1 (2018), Requirements for immersive live experience (ILE) services.
- [b-ITU-T H.430.3] Recommendation ITU-T H.430.3 (2018), Service scenario of immersive live experience (ILE).
- [b-ISO/IEC 23008-1] ISO/IEC 28008-1:2014, Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 1: MPEG media transport (MMT).
- [b-Akutsu] 阿久津明人, 小野朗, 高田英明, 外村喜秀, 井元麻衣子 (2016), 2020 Public Viewing-イマーシブテレプレゼンス技術「Kirari!」, NTT技術ジャーナル.
- [b-Nagata] 長田秀信, 宮下広夢, 柿沼弘員, 山口真理子 (2017), 任意背景リアルタイム被写体抽出技術, NTT技術ジャーナル
- [b-Tsutsumi] 堤公孝, 高田英明 (2017), 客席まで飛び出す音響を実現する波面合成音響技術, NTT技術ジャーナル.