

## 超高臨場ライブ体験を実現するILEの 日本初の国内標準について

NTTサービスエボリューション研究所 ながお じろう  
長尾 慈郎



### 1. はじめに

一般社団法人情報通信技術委員会 (TTC) マルチメディア応用専門委員会にて、2020年2月に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」の標準制定を行った。本稿では、この標準の概要について説明する。

JT-H430.2はITU-T勧告 H.430.2をダウンストリームしたTTC標準であり、ILEの基本アーキテクチャと一般的な役割モデルを規定するとともに、ILEの機能を実現するための候補技術を紹介している。具体的には、波面合成技術による超高臨場音声、超高解像度のビデオ統合、メディア伝送、任意背景のオブジェクト抽出について紹介している。以下、JT-H430.2の内容について詳しく解説する。

### 2. ILEとは

テレビ中継技術の発達によって、現代の我々はたとえ地球の裏側の出来事であってもほとんどタイムラグなしに見たり聴いたりすることができるという、大昔の人々には想像もできなかったような体験を享受している。しかしながら出来事の現場は、未だテレビでは再現しきれない部分が残っている。それは実物の大きさであったり、その場になければ感じられないようなリアルな音や振動や匂いのような体験であったり、平面で映し出された映像では見ることができない物体の側面や裏側の状態であったりする。

一方近年、注目度の高い音楽コンサートやスポーツイベントの多くがリアルタイムに中継されて会場外でパブリックビューイングされるようになり、本会場と同時に遠隔の視聴会場においても多数の観客がイベントを観覧し、お互いに感動を共有することが当たり前のように行われるようになってきた。ここでさらに、表示されている映像や音声を3次元的に再現したり、さらには振動や熱気といった環境情報も伝送・再現することができれば、観客の臨場感をより高めた「超高臨場感」を実現することができる。

こうした超高臨場感を実現するシステムやサービスは、従来よりaugmented reality (AR)、virtual reality (VR) (最近では両者及びmixed reality (MR) も含めてXRと呼ばれることもある) 等の技術を用いて提供されることが多

い。その他にも、映像音声の伝送技術や立体音響技術等が利用されており、そのうちのいくつかは標準化されている。また、サービスはリアルタイムなものだけでなく、蓄積型のものも多く提供されている。

XR技術はヘッドマウントディスプレイ (HMD) などの装置を頭部に装着することが多いが、観客どうしのコミュニケーションがしにくくなり感動の共有の妨げとなったり、装置を装着することによる違和感や、いわゆるVR酔いなどの影響により臨場感が低下するといった問題がある。

ILEはHMD等を装着する方法も含むが、HMD等を装着しない方法も複数人に同時に超高臨場感を提供するための重要なアプローチととらえている。またILEの特徴であるリアルタイム伝送は蓄積型のシステムと異なり、結果を知ってからでは観覧の楽しさが損なわれてしまうスポーツ競技等のような、イベントとの同時性に起因する感動を体験できる利点がある。リアルタイム性を確保するためには、異なるシステム間でのフォーマット変換等の負荷を軽減するとともに、遠隔地までのコンテンツ伝送の接続性を確保する必要がある。

本会場側システムと視聴会場側システムを国際的に世界のどこにいてもリアルタイムで相互接続可能とし、多くの観客に同時に超高臨場感を提供することがILEの目的であり、これを実現するための国際標準がITU-T H.430シリーズである。

### 3. ILEの国際標準化

ITU-T H.430シリーズは2016年からITU-T SG16 Q8で議論されており、H.430.1ではILEの定義と要求条件が規定されている。H.430.1において、ILEは次のように定義されている。

#### ILEの定義

あたかも視聴会場の視聴者が実際の本会場に入り込み、目の前でイベントを観覧したかのように感動が湧き、本会場の視聴者と感動を共有する観覧体験であり、センシング、メディア処理、メディア伝送、メディア同期、メディア表現等のマルチメディア技術の組み合わせで提供される高いリアリティによって引き起こされる感覚



H.430.1には、表で示すILEの要求条件が規定されている。要求条件には、再現された情景が目の前で起きているかのように感じるための映像の実物大表示や音像定位、人物等を視聴会場で再現するための物体のリアルタイム抽出、映像音声その他の情報を時間ずれなくリアルタイムに会場間で伝送するための同期伝送などが盛り込まれている。

#### 4. 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)

JT-H.430.2は、ITU-T勧告H.430.1 (3章) で規定されている要求条件を実現するためのILEの基本アーキテクチャと一般的な役割モデルを規定するとともに、ILEの機能を実現するための候補技術を紹介している。

#### 4.1 ILEの基本アーキテクチャ

図1にILEの基本アーキテクチャを示す。

アーキテクチャは以下の要素から構成されている。

- 環境情報を取得する機能
- メディア同期伝送
- 伝送レイヤ
- ILEアプリケーション
- 表示機能

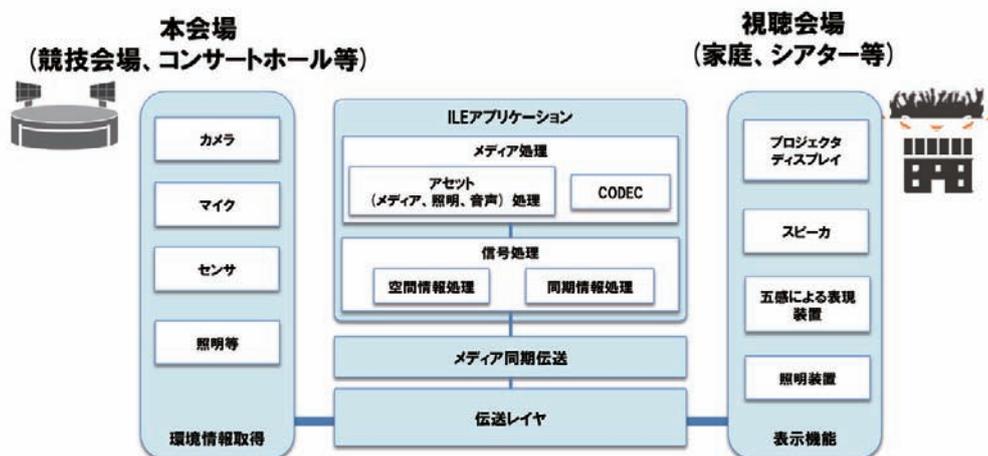
以下で各要素について解説する。

##### 4.1.1 環境情報を取得する機能

本会場の映像、音声、人物の位置、照明状態等を、カメラ、マイク、センサ等で取得する。

■表. ILEの要求条件

要求条件	内容	必須/推奨/オプション
1	実物大の表示	推奨
2	音声方向の再現	必須
3	照明等の効果による雰囲気再現	推奨
4	視聴会場の端末の性能・3次元位置に合わせた、空間環境の再現	必須
5	複数資産 (映像、ビデオ、位置情報、構成情報等) の同期表現	必須
6	拡張情報の表示機能	オプション
7	物体のリアルタイム抽出	必須
8	空間情報の計測	推奨
9	映像、音声、照明、空間、構成情報の同期伝送	必須
10	映像、音声、照明、空間、構成情報の同期データの蓄積	オプション
11	再構築のための、複数メディア (映像、音声、照明、空間情報等) の処理	必須
12	リアルな音像定位	推奨
13	ワイド映像生成	オプション



■図1. ILEシステムの基本アーキテクチャ



## (a) カメラ

高解像度なカメラを使用し、立体的な映像のための複数のカメラによる撮影や、映像への空間的な位置情報の付加、等を行う。

## (b) マイク

音声の方向再構成のための複数のマイクでの取音、音声への各オブジェクトの空間的位置情報の付加を行う。

## (c) センサ

高臨場感を提供するために、本会場の温度や湿度、振動等の環境情報を取得する。また、競技者などのオブジェクトの3次元位置計測や、オブジェクト追跡を行う。

## (d) 特殊効果及びその他

視聴会場にて照明などの特殊効果や舞台演出を再現するために、これらの情報を取得する。

### 4.1.2 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。視聴会場で環境情報をメディアと同期して表示するためには、環境情報を扱うための規定が必要である。

### 4.1.3 伝送レイヤ

コンテンツ保護やネットワーク遅延を考慮しつつ、本会場から視聴会場にメディアを伝送する。

### 4.1.4 ILEアプリケーション

CODECにより、コンテンツのエンコード・デコード処理を行う。映像から人物像を抽出したり、複数の映像を統合

してより臨場感の高い映像を生成したりする。また表示機能での環境再現に向けてビデオ、音声、照明、センサ情報などのコンテンツを同期させる。

### 4.1.5 表示機能

視聴会場において、映像の表示、音声方向の再現、照明の再現などを行う。これらを視聴会場で同期して提供することで、視聴者は超高臨場感を得られる。

#### (a) プロジェクタとディスプレイ

会場のサイズや表示装置のサイズ、位置等は一般に本会場と視聴会場とは異なる。この違いを吸収するために、本会場から伝送されるこれらの情報をもとに視聴会場に適した表示方法を再構成する。

#### (b) スピーカ

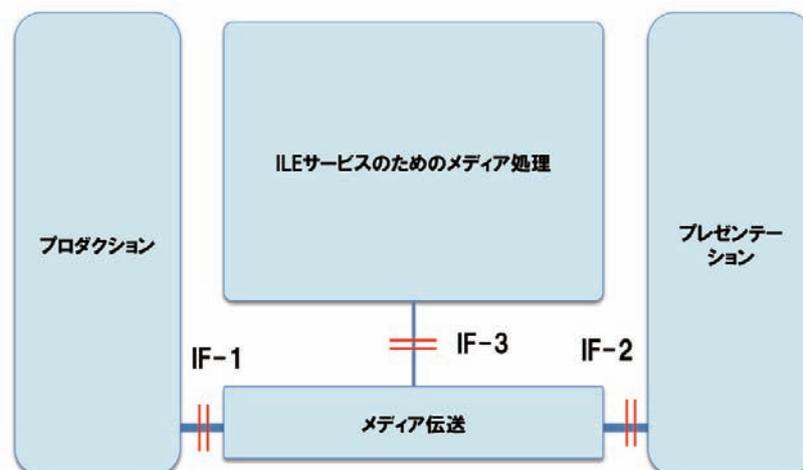
イベント会場で抽出したオブジェクトの空間情報や音声データを利用することで、ディスプレイに表示されたオブジェクトと音声の方向を合わせて視聴会場での音声の方向を再構成して提示する。

#### (c) 五感による表現装置

本会場から伝送された照明情報を含む特殊効果や舞台演出の情報を利用して、視聴会場において本会場の照明や演出を再構成する。その他、本会場のセンサで取得した様々なデータを活用し、視聴会場で本会場の環境を高臨場に再現する。

## 4.2 ILEサービスを提供する一般的役割モデル

基本的なアーキテクチャを元にしてインタフェースを規定した一般的な役割モデルを図2に示す。このモデルは大き



■ 図2. ILEサービス提供のための一般的役割モデル



く以下の4つの役割と3つのインタフェースからなる。

#### 〈役割〉

- 本会場での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

#### 〈インタフェース〉

- IF1：プロダクションとメディア伝送間  
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のやり取りを規定する。
- IF2：メディア伝送とプレゼンテーション間  
各視聴会場用に設計されたアセット情報のやり取りを規定する。
- IF3：メディア伝送とメディア処理間  
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のやり取りを規定する。

### 4.3 ILE機能の候補技術

#### 4.3.1 任意背景リアルタイム被写体抽出機能

擬似3次元映像のような高臨場感映像を視聴会場で提示するためには、競技者や演技者等の被写体を、本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。ILEサービスでは、スポーツの試合や音楽コンサート等といった、クロマキーが利用できない実際のイベント映像に対して、リアルタイムの被写体抽出が必要である。本技術はこの要件を満たすILE機能の候補技術である。

#### 4.3.2 メディア同期伝送機能

ILEサービスの主要な特徴の1つに、映像や音声に加えて、物体の位置等の空間情報や照明制御情報、舞台演出情報等の複数のライブストリームコンテンツを同期伝送することが挙げられる。ISO/IEC 23008-1 MPEG Media Transport

(MMT) は、これらのメディア情報を同期伝送するための情報ストリームに利用可能である。なお、本機能を実現するための、ILEのためのMMTプロファイルはITU-T H.430.4で規定されている。

#### 4.3.3 超高解像度のためのサラウンド映像合成

従来の16:9のテレビでは実現困難な、超広角映像による高臨場感を実現するための技術である。複数の高解像度カメラを使って隣接する視野の映像を取得し、連結、合成することで超広角映像を生成する。本技術は、実時間で高解像度映像を大量高速処理することを可能とする技術革新によって実現された。

#### 4.3.4 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スピーカアレイは、臨場感を強化可能なILE機能の候補技術の1つである。WFSの典型的な実現例としては、スピーカを直線状に配置する方法がある。

## 5. おわりに

本稿では、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」の標準制定について概説した。本標準は、超高臨場なライブ中継を実現するためのアーキテクチャを規定しており、ILEサービス展開の基盤となると期待される。

今後も他のILE関連ITU-T勧告のダウンストリームや、セミナー等の開催を通してILE関連標準の普及促進を図るとともに、VR等の関連技術の取込みを検討するなど、他の標準化機関と連携して勧告の拡充を進めていく。

なお、本稿は2020年4月にTTCより刊行されたTTCレポートVol.35 No.1 pp.7-10「標準類制定状況 マルチメディア応用専門委員会」の記事を本誌向けに加筆・修正したものである。