



ITU-T H.430シリーズ活用事例：東京2020オリンピックのバドミントン競技の超高臨場ライブ伝送技術実証

NTT人間情報研究所

なが お じろう
長尾 慈郎は せ がわ けいすけ
長谷川 馨亮む とう まこと
武藤 誠なみかわ だい ち
並河 大地こん や せい い ち
紺谷 精一

1. はじめに

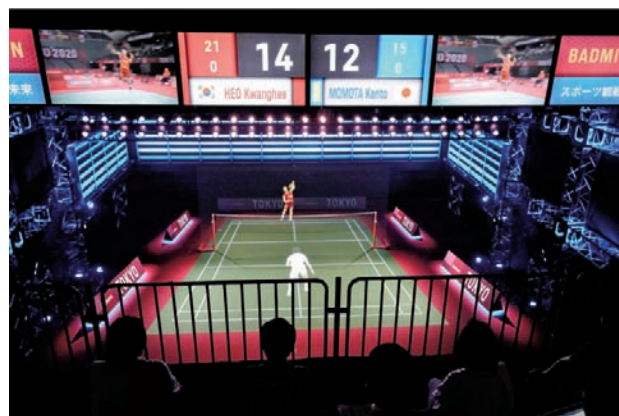
ITU-T SG16 Q8では、スポーツや音楽コンサートなどのイベントを遠隔地へリアルタイムに伝送し、立体的な実物大表示や音響再現等を駆使して、あたかも会場で観戦しているかのような超高臨場感を実現するImmersive Live Experience (ILE：超高臨場ライブ体験) の国際標準化を進めてきた。本稿では、ILE関連の国際標準であるITU-T H.430シリーズ勧告に準拠したシステムにより、東京2020オリンピックのバドミントン競技を対象にNTTが行った技術実証について解説する。

2. バドミントンの超高臨場ライブビューイング技術実証

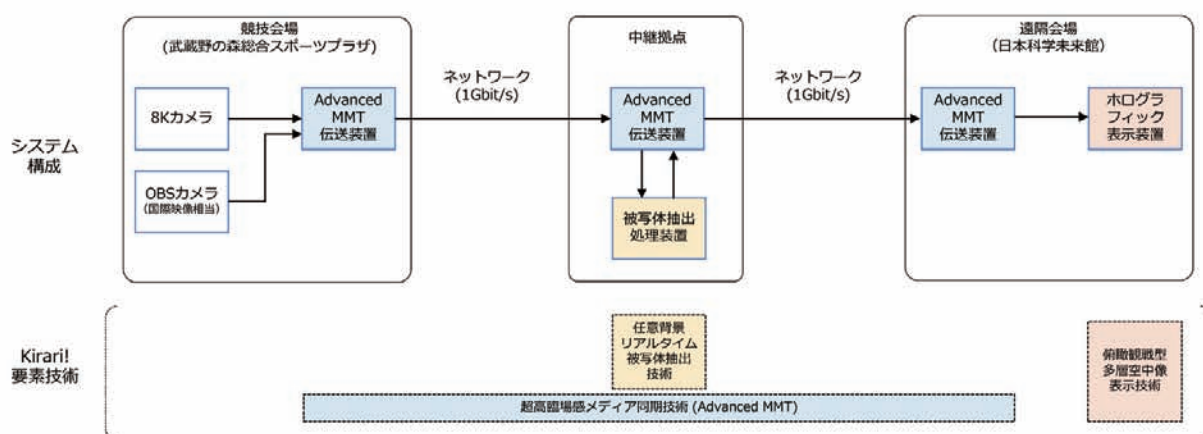
NTTは、東京2020組織委員会が主導する「TOKYO 2020 未来のスポーツ観戦プロジェクト」において、超高臨場感通信技術Kirari!の技術協力を行った。本プロジェクトは、日本科学未来館の観戦者があたかも競技会場で観戦しているかのような臨場感と一体感を提供することを目指した次世代臨場感テクノロジー実証である。本プロジェクトでは武蔵野の森総合スポーツプラザで行われた東京2020オリンピックのバドミントン競技を対象とし、その様子を遠隔会場（日本科学未来館）へホログラフィックに映像伝送した。

本プロジェクトでは当初、一般参加者を招いてのライブビューイング実証を計画していたが、新型コロナウイルス感染拡大予防の観点から中止とした。そこで実施形態を変更し、2021年7月30日～31日、日本科学未来館において、メディア向けに「スポーツ観戦の未来～次世代臨場感テクノロジー実証プログラム～」を公開した（図1）。

本技術実証の概要を、システム構成に沿って説明する。図2は全体のシステム構成図である。



■図1. 東京2020バドミントン競技の超高臨場ライブビューイング技術実証の様子



■図2. 全体のシステム構成



ホログラフィック表示に用いるための映像は、競技会場である武蔵野の森総合スポーツプラザの観客席に設置した8Kカメラで撮影した。また、ホログラフィックな映像と併せて通常の2D映像も表示するため、競技会場ではOBS (Olympic Broadcasting Services) 提供の映像、音声も取得した。

競技会場で撮影、取得した複数の音声、映像は「超高臨場感メディア同期技術 (Advanced MMT)」を用いて、後述の被写体抽出作業を行う中継拠点と、遠隔会場である日本科学未来館へ同期伝送した。伝送には1Gbit/sのネットワークを用いた。

中継拠点では「任意背景リアルタイム被写体抽出技術」を用いて、試合会場から伝送された映像から選手やシャトルの映像のみを抽出した。抽出後の映像もAdvanced MMTにより遠隔会場へ同期伝送した。

遠隔会場 (図3) では、競技会場そのものを再現するために実物大のコートと約100席の観客席を設置した。表示装置として、観客から見て手前側と奥側の2層それぞれで、4K映像をホログラフィックに表示可能な投影装置を使用した。中継会場から伝送された選手とシャトルの抽出映像を、「俯瞰観戦型多層空中像表示技術」により手前側のコートにいる選手はネットの手前側に、奥側のコートにいる選手はネットの奥側に、シャトルは動きに合わせて投影位置を

切り替えながら投影した。これにより、リアルなコートに二人の選手が降り立ったかのような空間を実現した。

ここまでで述べた「Advanced MMT」、「任意背景リアルタイム被写体抽出技術」、「俯瞰観戦型多層空中像表示技術」は、それぞれ超高臨場感通信技術Kirari!を構成する技術要素であり、後述のITU-T H.430シリーズに準拠している。以下で、各技術要素の概要をITU-T H.430シリーズと関連付けながら解説する。

2.1 任意背景リアルタイム被写体抽出技術

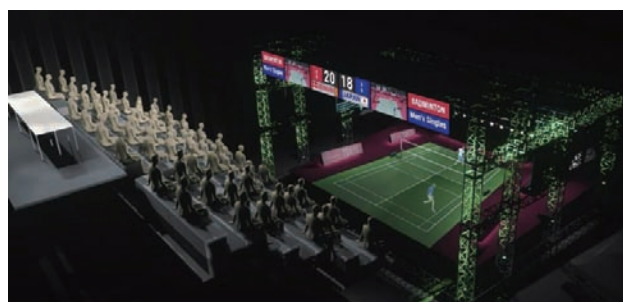
「任意背景リアルタイム被写体抽出技術」はKirari!を構成する技術の一つであり、後述のILEアプリケーションに相当する。本技術は、入力映像の中から特定の被写体に相当する映像のみを抽出する技術である^[1] (図4 (a))。

通常、映像の中から被写体のみを抽出する場合は、グリーンバックやブルーバックなどの背景を用意し、クロマキーによりその背景色を消す手法がとられる。本技術では、特別な背景環境を用意することなく、特定の被写体をリアルタイムに抽出することが可能となる。これにより、競技会場で撮影した映像からリアルタイムに選手やシャトルの映像のみを抽出することができる。

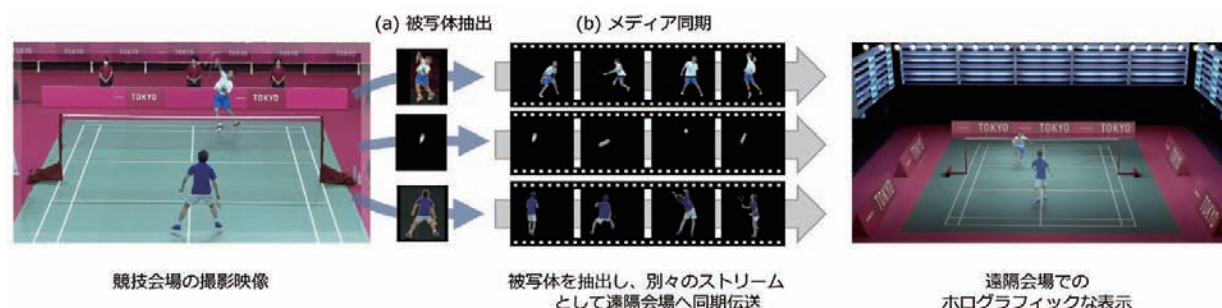
今回、本技術をバドミントン競技へ適応させるにあたり、以下の5点について技術改善を行った。

(1) コートの手前と奥の選手の個別抽出

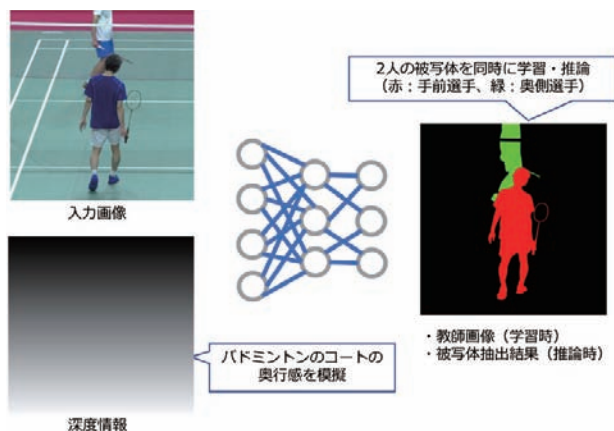
従来技術では、手前と奥の選手を分けて抽出することができなかった。そこで、バドミントンコートの空間を模擬した深度情報を入力し、手前と奥の選手について同時に学習・推論できる深層学習モデルを考案した。これにより、バドミントンのようにコートの手前と奥で選手が分かれる競技においても、個別に選手を抽出することを可能とした (図5)。



■図3. 遠隔会場の投影装置



■図4. Kirari!の被写体抽出、メディア同期の流れ



■図5. コートの手前と奥の選手の個別抽出の仕組み

(2) 選手映像のフレームレート・解像度向上

従来技術によるリアルタイムの被写体抽出処理は、4K・30fpsまでの入力映像にしか対応していなかった。30fpsでは、スマッシュ等の速い動きに追従できず、途切れ途切れの映像になってしまう。そこで、被写体抽出システムの実装を見直し、処理する映像フレームの多重化や計算リソースの平滑化などの工夫を行うことで、4K・60fpsのリアルタイム被写体抽出処理を可能とした。

また、スペースの制約上、試合会場には1台しかカメラを設置できないが、4Kカメラ1台での撮影では十分な選手像の解像度が得られず、等身大サイズの空中像として投影すると粗さが目立ってしまう。そこで解像度向上のため、8Kカメラで撮影した映像から、専用装置により手前側のコート、奥側のコートに対応する領域をそれぞれ4K解像度で切り出して、被写体抽出システムへの入力映像とした。

これらの改善により、高解像度で被写体の粗さが目立たず、かつ動きが滑らかな映像表現を実現できた。

(3) 微小かつ高速移動するシャトルの安定的な抽出

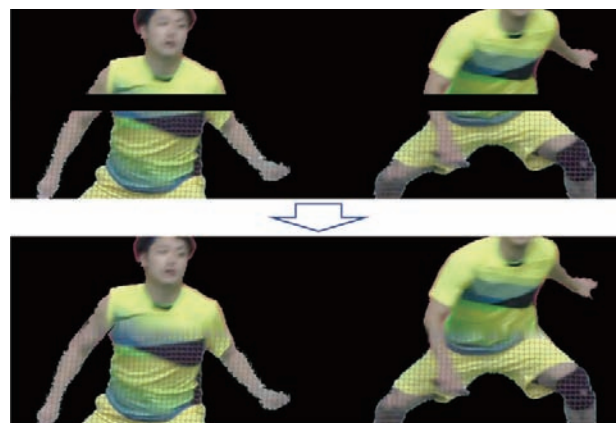
映像中に映るシャトルは非常に小さくかつ高速に移動するため、これまでの画像認識手法ではノイズが多く、ちらつきが目立ったり、シャトルの軌跡が途切れがちになったりと、十分な認識精度が得られなかった。今回、シャトル専用の被写体抽出方式を開発し、正確なシャトル位置の検出と、様々な形に変形するシャトルの高精度な抽出に成功した。

シャトル位置の検出には、映像に映り込むシャトルと同様の小さな物体（客席のガイド照明など）の影響を排除するため、畳み込みニューラルネットワーク（CNN）を用いて、シャトルの位置情報及び移動情報を同時に学習させる方式を考案した。検出されたシャトル位置・形状と、次フレー

ムにおけるシャトル予測情報（位置・形状・モーションブラーの程度）でフィルタリングして再現率を高め、背景差分法によりシャトルのみを抽出した。この手法により、観戦に用いることができるレベルのシャトル抽出品質を達成した。

(4) 選手映像の欠落部分の自動補完

図6のように、奥の選手がネットと重なった場合、選手像に黒い帯状の欠落が発生する。欠落部分の上下の色情報から欠落部分の色を推定・補完することにより、欠落のない映像をつくり出した。



■図6. ネットの黒帯状部分の推定

(5) 選手の影の自動生成

実際の会場で現れる選手の影を遠隔会場にも再現することで、より自然な選手像をつくり出すことができる。選手の抽出結果を基に、映像から影を検出した。この際、ジャンプの高さや競技会場の照明条件から、選手の影が存在し得る範囲を定めることにより高速化を図っている。

2.2 超高臨場感メディア同期技術

「超高臨場感メディア同期技術（Advanced MMT）」は、メディア伝送規格MMT（MPEG Media Transport）をNTTがITU-T H.430.4に基づき拡張した技術である。本技術では、映像、音声、照明情報といった複数の連続するデータ（ストリーム）を、時刻的な同期を保ったまま伝送できる。本技術実証では、競技会場で撮影した競技映像、音声、競技映像から選手だけを抽出した映像、シャトルだけを抽出した映像、OBSから提供された映像など、複数のストリームを本技術によって同期伝送し、遠隔会場に必要なデータを適切なタイミングで表示することで、臨場感の高い空間表現を実現した（図4（b））。

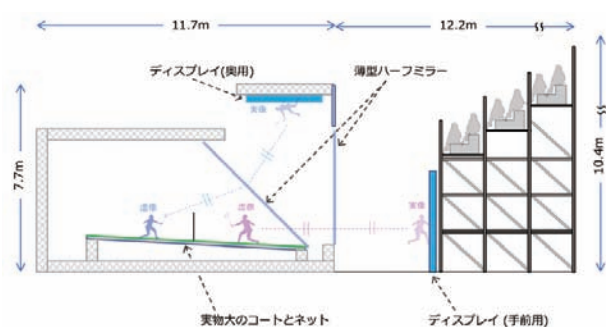


2.3 俯瞰観戦型多層空中像表示技術

後述のITU-T H.430.5^[4]におけるバドミントン用の表示環境の例では、手前側の選手のみをペッパーズゴーストの原理によりホログラフィックに表示し、奥側の選手は通常の2D表示としているが、本技術実証では手前側、奥側の両方の選手をホログラフィックに表示した。それを可能にしたのが「俯瞰観戦型多層空中像表示技術」である。

本技術では、実際の競技会場に近い高さの観客席を設置し、そこから見下ろす状況下において、手前側と奥側の選手がネットを挟んでリアルに見えるように、コート、ネット、2枚のハーフミラー、LEDディスプレイ、プロジェクタなどの配置を最適化している（図7）。これにより手前側と奥側の選手をそれぞれ正しい位置でホログラフィックに表示できるようになり、あたかも競技会場の観客席から実際のコートを見下ろしているような感覚を体験できる。

しかし、本技術では2つのハーフミラーを用いているため、その間を行き来するシャトルが不連続な動きに見えてしまう課題が発生する。この課題は、2つの層が物理的に分離しているという構造的な問題に起因するため、完全に解決することはできない。そこで、コート側面映像や打球音の解析結果と手動操作を組み合わせることでシャトルの打球タイミングを判定し、より自然に見えるタイミングでシャトルの表示層を切り替えた。これにより、シャトルの動きがダイナミックかつ自然に見えるようになった。



■図7. 装置のイメージと空中像表示の例

2.5 評価・成果・結果など

本技術実証は、2021年7月30日～31日の2日間実施し、東京2020オリンピックバドミントン競技の男子ダブルス、女子シングルス、男子シングルのライブ中継や予選の録画映像を、報道関係者やその他関係者に体験いただいた。体験者からは、「実物大の大きさで見ると、臨場感が全く違う」、「試合中、思わず会場にいるときと同じような歓声

をあげてしまった」など、高い臨場感を体験いただけたという感想を多く得られた。

被写体抽出については、8K・60fpsで撮影した映像からのリアルタイムでの抽出処理に成功した。選手同士が重なった場合に被写体が一部欠けるなど、抽出精度面での課題は残ったものの、おおむね観戦に堪え得る精度の抽出結果は得られた。シャトル表示については、端の方の座席など視聴場所によっては不連続が目立つため、改善が必要となる。

通信性能としては、競技会場から遠隔会場までのエンドツーエンドの遅延時間は、合計2800ms以下であった。表に内訳を示す。MMT伝送遅延には、ネットワークの伝送遅延である、競技会場～中継拠点の約1ms、中継拠点～遠隔会場の約0.1msも含まれる。被写体抽出処理の遅延には、処理時間のゆらぎ等を考慮したバッファも含まれるが、それを除いた抽出処理自体の遅延は400ms以下である。

■表. 遅延時間の内訳

撮影系	17ms以下（1frame以下）
MMT伝送	1600ms（4K 59.94p ITU-R BT.709、エンコードレート40Mbit/s、FEC5%）
被写体抽出処理	1000ms（入出力バッファ込み）
表示系	～120ms（30fps換算で4frames）

3. 本システムで利用したITU-T勧告H.430シリーズ

ITU-T SG16 Q8では、これまでにILEに関連する国際標準としてITU-T H.430.1^[5]からH.430.5まで5つの勧告を制定してきた。本章では5勧告について本システムとの関係がある部分を中心に解説する。

3.1 ITU-T H.430.1（要求条件）

ITU-T 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を刺激され、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合せで提供される高いリアリティーによって引き起こされる感覚

定義にもあるとおり、ILEは、スポーツ競技会場などの本会場側システムから、本会場から離れたリモート会場に設置された視聴会場側システムへ映像、音声、位置情報等



をリアルタイムで伝送し、多くの観客に同時に超高臨場感を提供することを基本コンセプトとしている。これを実現するためには、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、伝送された情報を基に情景が目の前で起きているかのように感じさせるための映像の実物大表示や位置の再現等、いくつかの要求条件を満たす必要がある。ITU-T H.430.1はILEが満たすべき要求条件も規定しており、今回の技術実証では全必須要件及び推奨要求条件1（実物大表示）を満たしている。

3.2 ITU-T H.430.2^[2] (アーキテクチャフレームワーク)

ITU-T H.430.2は、ILEの基本アーキテクチャを規定している。図8にILEの基本アーキテクチャを示すとともに、H.430シリーズの各勧告と基本アーキテクチャとの関係を示す。アーキテクチャは大きく以下の要素から構成されている。

・環境情報取得

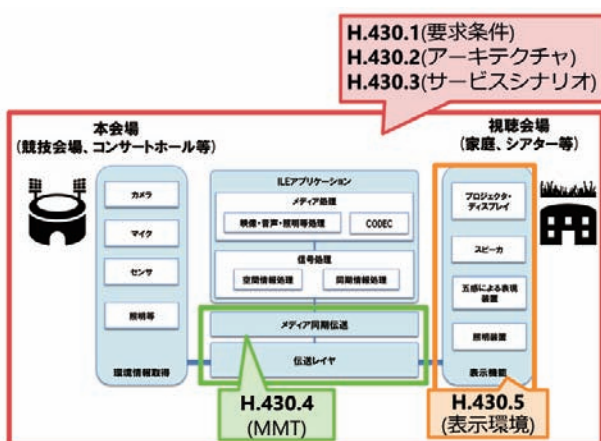
本技術実証では、武蔵野の森総合スポーツプラザでの競技映像撮影、OBS (Olympic Broadcasting Services) 提供の映像、音声の取得がこれにあたる。

・伝送レイヤ及び同期メディア伝送

ITU-T H.430.4^[3]で規定されており、本技術実証では超高臨場感メディア同期技術 (Advanced MMT) として実装された。

・ILEアプリケーション

本技術実証では「任意背景リアルタイム被写体抽出技術」により、シャトルのリアルタイム抽出など非常に高度な映像処理が行われている。



■ 図8. ILEシステムの基本アーキテクチャ及びH.430シリーズの各勧告との関係

・表示機能

ITU-T H.430.5で規定されており、本技術実証に関連する表示環境の参照モデルの詳細は3.5節にて解説する。

3.3 ITU-T H.430.3^[6] (サービスシナリオ)

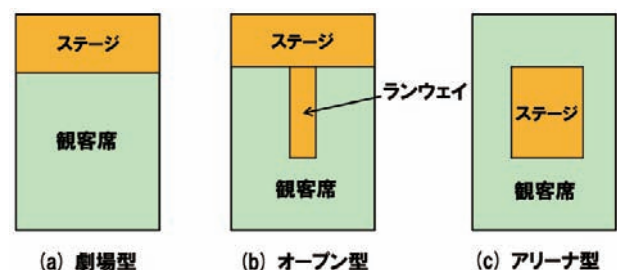
本勧告は、ILEの要件やフレームワークの理解を促進するために、ILEで実現されるサービスシナリオ及びユースケースを記載している。サービスシナリオの一つにスポーツのリアルタイム伝送があり、本技術実証はこのサービスシナリオの1実装形態と捉えることができる。

3.4 ITU-T H.430.4 (MMTプロファイル)

視聴会場で競技等の情景をリアルタイムに再構築して超高臨場感を実現するためには、映像、音声に加え、オブジェクトの位置（空間情報）や演出のための照明情報をリアルタイムに同期伝送する必要がある。ITU-T H.430.4ではMMTでILEの情報を伝送するため、オブジェクトの空間情報及び照明情報を対象として、サービス構成、メディア伝送プロトコル、シグナリング情報を規定している。本技術実証では、競技会場で撮影した競技映像や音声と、その競技映像から選手だけやシャトルだけを抽出した映像等を伝送し、表示環境で時刻同期して表示するために活用されている。

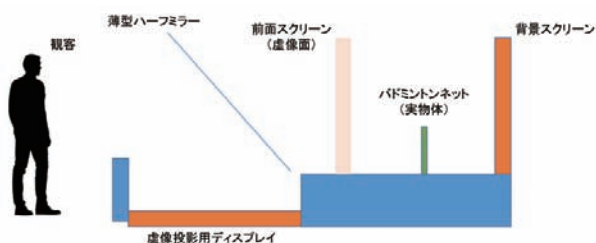
3.5 ITU-T H.430.5 (表示環境)

図9にITU-T H.430.5で規定されている表示環境の分類を示す。本技術実証は、ステージには複雑な構成の表示装置が使用されているものの、ステージと観客が相対する配置であるので、図9 (a) の劇場型に分類される。

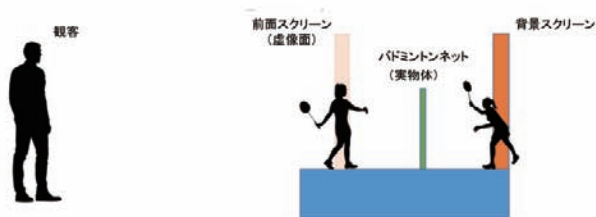


■ 図9. 表示環境の分類

図10に劇場型の実装ガイドライン記載の構成例 (a) と、観客からの見え方 (b) を示す。図10に示すように、観客から見ると背景スクリーンに表示された選手映像はそのまま



(a) 表示装置の構成例



(b) 観客から見た構成例の見え方

■ 図10. 劇場型表示環境構成例

見えると同時に、虚像投影用ディスプレイに表示された選手の映像は薄型ハーフミラーで反射され、あたかも虚像面に選手が立っているかのように見える。薄型ハーフミラーは透明であるから、虚像面の選手以外の部分は透過してバドミントンネット（実物体）と背景スクリーンが見える。観客からはこれらが同時に見えるので、あたかも二人の選手がバドミントンネットを挟んで対戦しているかのような感覚を体験できる。ここではITU-T H.430.5記載の構成例に従って説明したが、本技術実証ではさらに複雑に設計された、背景にも薄型ハーフミラーを使用する構成が使用されている。

4. おわりに

本稿では、ITU-T SG16 Q8が制定を進めてきたITU-T H.430シリーズを解説するとともに、本勧告シリーズに準拠してシステムが構築され、2021年に実施された技術実証について述べた。

NTTは「TOKYO 2020 未来のスポーツ観戦プロジェクト」において、超高臨場感通信技術Kirari!を用いてH.430シリーズに準拠した技術実証のシステムを構築し、あたかも競技会場で観戦しているかのような、臨場感の高いバドミントンの観戦体験を実現した。今後は、抽出性能向上等の技術的改善、他の競技やエンタメなど他分野への展開などを検討するとともに、Kirari!をはじめとしたリモートワー

ルドを具現化するための研究開発を進めていく。

ITU-T H.430シリーズはこれまでに、要求条件、フレームワーク、サービスシナリオを規定する基本3勧告に加え、国際接続に不可欠なILE特有の情報を映像・音声と共に同期伝送する通信プロトコルを規定する勧告と、超高臨場感を実現する表示環境について規定する勧告が完成している。これら5勧告によりILEを構成するための基本的な規定は整ったと考えられるが、ITU-T SG16 Q8ではさらに高い臨場感を実現するために、引き続きILEに関連する技術の国際標準化を進める予定である。例えば、視覚、聴覚だけでなく、触覚や嗅覚に関する情報を伝送し再現することも臨場感をさらに高めると考えられ、今後の標準化が期待される。

※NTTは、東京2020ゴールドパートナー（通信サービス）です。

謝辞

本技術実証の実現に向け、共に様々な課題を克服してきていただいた東京2020組織委員会イノベーション推進室の皆様、主催者として先導いただいた日本科学未来館の皆様、そして、技術的に支えていただいたパートナー企業の皆様に感謝します。

参考文献

- [1] 柿沼・長尾・宮下・外村・長田・日高：“機械学習を用いた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, Vol.30, No.10, pp.16-20, 2018.
- [2] ITU-T H.430.2, “Architectural framework for immersive live experience (ILE) services,” <https://www.itu.int/rec/T-REC-H.430.2>
- [3] ITU-T H.430.4, “Service configuration, media transport protocols, signalling information of MPEG media transport for immersive live experience (ILE) systems,” <https://www.itu.int/rec/T-REC-H.430.4>
- [4] ITU-T H.430.5, “Reference models for immersive live experience (ILE) presentation environments,” <https://www.itu.int/rec/T-REC-H.430.5>
- [5] ITU-T H.430.1, “Requirements for immersive live experience (ILE) services,” <https://www.itu.int/rec/T-REC-H.430.1>
- [6] ITU-T H.430.3, “Service scenario of immersive live experience (ILE),” <https://www.itu.int/rec/T-REC-H.430.3>