高臨場観戦を盛り上げる映像音響技術

三上 弾†1 /國田 豊
鎌本 優†2 /志水 信哉
丹羽 健太†1 /木下 慶介

NTTメディアインテリジェンス研究所†1
NTTコミュニケーション科学基礎研究所†2

スポーツイベントでは会場での観戦者よりはるかに多くの人がTV，インターネットなどを介して遠隔地から観戦・視聴します．近年，個人の嗜好・視聴スタイルが多様化し，視聴者が各々好みの方法で観戦を楽しむことが求められています．本稿では，現実に近い視聴体験の再現（高臨場感）と同時に，現実を超える体験の提供（超高臨場感）という2つの観点から，臨場感の高い視聴の提供を可能とするための映像音響技術に関するNTTの取り組みを紹介します．

高臨場感とは

「臨場感」の辞書的な意味は，「あたかもその場所にいるような感覚」です．スポーツ観戦で求められる高い臨場感はそれだけでしょうか？　臨場感には大きく分けて2つの側面があります．第1は，あたかもその場にいるような感覚，すなわち高臨場感です．第2は，その場では分からないことまでも分かる感覚，言うなれば「超」高臨場感です．スポーツ観戦では，両側面の高臨場感が求められています．家にいながら，あたかもスタンドさらにはフィールド内にいるかのように観戦するという体験は多くの人が求めるものです．一方で，スタンドや既存のTVでは見ることができない映像（選手目線での映像など），聞くことができない音（選手どうしの会話など）による高臨場感も多くの視聴者が求める体験であり，観客席での視聴体験もより豊かなものにします．本稿では，2つの高臨場感を実現するためにNTTが取り組んでいる要素技術のいくつかを紹介します．

全天球映像向けインタラクティブ配信技術

近年，ヘッドマウントディスプレイ（HMD）や360度に近い画角で撮影できるカメラの安価な製品が発表され，これまで一部の専門家や愛好家向けであったバーチャルリアリティ視聴が普及に向けて活況を呈してきました．NTTメディアインテリジェンス研究所では，ユーザが好みに応じて好きな領域を視聴できるインタラクティブパノラマ配信技術[(1)]の研究を進め，このような視聴スタイルに適合した「全天球映像向けインタラクティブ配信技術」を研究しています．本技術では，**図1**に示すとおり，全天球（360度）で広範囲に撮影された映像をいくつかの領域に分割し個別に高品質エンコード後，ユーザが見ている方向に応じた

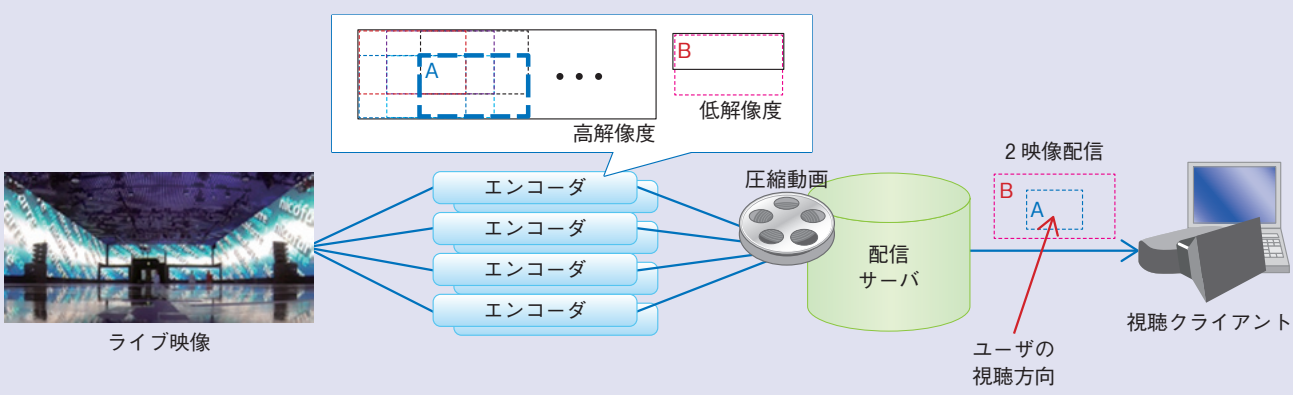

図1　全天球映像向けインタラクティブ配信技術の全体構成

領域の高品質映像を選択配信します．見ている方向の領域のみが高品質に視聴できることから，全天球映像全域を高品質に配信するよりも限られた帯域で配信できるようになります．

このように，全天球向けの配信技術ではインタラクティブパノラマ配信技術で培った選択配信技術を利用していますが，以下の特徴によりユーザの視聴体験は大きく異なります．

① 広視野：視野を覆う映像により，利用者は空間に入り込んだような感覚（没入感）を得ることができます．人間の視野は中心部ほど空間解像度が高く，周辺部の視野の空間解像度は低いことが知られています．この特徴を利用し，中心部の視野のみ部分的に高解像度で伝送・提示することで，限られた伝送帯域で高い臨場感をユーザに提供することが可能になります．

② 頭部追従性：ユーザの頭部の動きをHMDに搭載された加速度センサや位置センサにより検出し，動きに応じた映像を両眼に提示するので，あたかも空間を見回しているように感じることができます．タブレットなどを利用した従来の視聴と比べ，ユーザは視聴する部分の選択を意識する必要がないので，より直感的な視聴が可能になります．

これまでのゲームやアミューズメントパークでのアトラクションで体験できる空間は，コンピュータグラフィクスで作成したものが多数でした．しかし，ここで紹介した技術により，音楽ライブなど，実写の映像ならではの臨場感を配信できることを確認してきました．本技術をスポーツ観戦に用いることで，スタジアムの観客席の興奮した雰囲気や，フィールドに入り込んだかのような迫力をユーザに伝送できることを期待しています．

ロスレス音響符号化技術

現在，ポータブルオーディオプレーヤや地上デジタル放送で用いられているMP3やAAC（Advanced Audio Coding）などの音響符号化は，伝送帯域や記憶容量の制約の下，ある程度の品質を保ったまま圧縮をすることができる技術であり広く普及しています．一方，高い臨場感を実現するにあたっては原音を忠実に伝送することも求められています．NTTではMPEG-4 ALS（Audio Lossless Coding）の標準化に参画し，ロスレス音響符号化の普及に努めてきました[(2)]．

ロスレス音響符号化は圧縮しても原音の波形が完全に復元できるため，ネットワークリソースの無駄遣いをさけつつ，一切劣化の起きない音質を保証したまま伝送することが可能です．実際，NTT未来ねっと研究所と協力して開発した映像・ロスレス音響符号化装置はNTT西日本などによる高臨場感音響ライブ配信トライアルでも使用され，配信先でも本会場と同様の拍手や声援が自然に観客から沸き起こるなど，従来のライブ配信を超える会場との一体感を創出することができたと報告されています[(3), (4)]．このような高音質化の流れは4K・8K放送にも影響を与えました．2014年の春に総務省が実施した超高精細度テレビジョン放送システムへの意見募集に対し，半数近くが音質向上に関するものであり，その中でもロスレス音響符号化の利用を求めるものが多くを占めました[(5)]．その結果，2014年の夏にロスレス音響符号化MPEG-4 ALSは4K・8K放送にも使われ得る方式として総務省の省令告示に掲載され，ARIB標準（ARIB STD-B32）として規格化されました．

このように，臨場感を向上させるために音質の向上を求める声が大きくなってきています．私たちもそのような要望にこたえるために，タブレット端末やセットトップボックスへの実装，電波帯域の有効利用の実証実験などを進めてきました．今後，ロスレス音響符号化の普及により，TV放送やコンテンツ配信の臨場感が向上することが期待されます．また，後述するズームアップマイク技術で収録した音声を，原音のままロスレス音響符号化により圧縮することで効率良く伝送し，届いた原音を受聴環境に応じた残響制御により再生することで，臨場感の高

いコンテンツを楽しむことができるようになる日もすぐそこまで来ています．

自由視点映像配信・符号化技術

自由視点映像とはシーンを撮影したカメラの位置や向きに関係なく，好きな位置や向きからのカットを視聴できる映像です．通常はカメラを設置できない位置からの映像，例えばサッカー映像における選手やボール目線の映像を提供することで，これまでの映像では得られなかった臨場感のある映像体験の実現を目指した技術です．

自由視点映像はシーンをさまざま位置や向きから同時に撮影した多視点映像を用いて生成します．撮影に必要なカメラの台数は，視点の自由度や生成する映像の品質に依存しますが，一般的に非常に多くのカメラが必要とされています．しかし，そのような多数のカメラによる撮影や，それら大量の映像データの蓄積・伝送を実現することは困難です．より少ない映像データを用いて自由視点映像を実現する方法として，映像に加えて，カメラから被写体までの距離を表現したデプスマップを用いる方法が知られています．ここでは，多視点からの映像とデプスマップを用いた自由視点映像について私たちの取り組みを紹介します．

近年のセンサ技術の発展により，デプスカメラやレンジスキャナなどによりデプスマップを直接取得することが可能になってきています．しかしそれらセンサを用いて取得されるデプスマップは空間解像度が低く，多数のノイズを含みます．そのため，そこから生成可能な自由視点映像の品質は高くありません．私たちはこの問題に対して，映像とデプスマップ間の相関や視点間におけるデプスマップの整合性を利用したノイズ除去処理やデプスマップのアップサンプル処理を開発しました．さらに，これらをGPU実装することで，デプスセンサで取得された多視点映像とデプスマップからのリアルタイム自由視点映像合成を実現しています．

多視点映像とデプスマップは自由視点映像をコンパクトに表現した映像データですが，通常の映像と比較するとそのデータ量は膨大です．そのため，自由視点映像を実際に配信するためには効率的な圧縮符号化技術が必要不可欠です．これまでに私たちは，視点合成予測やパレットベース予測など，自由視点映像のための符号化技術を多数開発してきました．視点合成予測は，自由視点映像合成技術を符号化に応用した技術であり，すでに符号化済みの別の視点の映像とデプスマップとを用いて予測画像を合成することで，効率的な視点間予測を実現します．パレットベース予測は，同一の被写体内では値の変化が少なく，異なる被写体間では値が大きく異なるというデプスマップの特徴を利用した予測画像生成手法であり，高精度な予測を実現するだけでなく，デプスマップの符号化による自由視点映像合成の性能劣化を防ぐこともできる手法です．これら開発した手法は，最新の映像符号化国際標準規格HEVCの拡張規格3D-HEVCに採用されています[(6)]．

自由視点映像の実現には，上記紹介した技術に加え，撮像から表示・ユーザインタフェースに至るまで多くの技術が必要となります．また，現在のデプスマップを利用した自由視点映像合成技術には視点移動の自由度や合成映像の品質に限界があります．今後，スポーツイベントなどにおいてより広大な空間を対象とした，あたかも競技空間に降り立ったような映像体験を提供する自由視点映像の実現に向けてさらなる技術開発を進める予定です．

ズームアップマイク技術

通信・放送を通じたスポーツ観戦を盛り上げるために，ユーザがあたかもフィールドの中にいるかのような映像・音声を生成するための技術開発が進んでいます．ズームアップマイクは，そうしたコンテンツをつくるうえで必要となる要素技術で，遠くの音をクリアに収音することを可能にします．

この技術は「カメラで遠方をズームするように，遠方の音をクリアに収音できないのはなぜなのか」というふとした疑問から研究がスタートしました．これまで録りたくても録れなかった音をクリアに収音することができる

としたら，高臨場感あふれるコンテンツをつくるための重要なツールになるでしょう．また，通信・放送を通じたスポーツ観戦において，ユーザがあたかもフィールドの中にいるかのような音声を提供することが将来的にできるようになるかもしれません．

ズームアップマイクは，大きく2つの技術で構成されています（**図2**）．

⑴　音源群を詳細に解析するための受音系設計技術

遠方にある音源群を分離して解析するために，どのような音を複数のマイクロホンで収音したら良いのかについて基本原理[7]を確立しました．観測信号から得られる音源群の情報量を定義し，それを最大化するための観測信号の性質を明らかにしました．図2に示す受音系は，確立した原理に則って実装した受音系です．12枚のパラボラ反射板の前に96本のマイクロホンを設置するという条件下で，最適な受音系を構築しました．

⑵　出力音質と雑音抑圧性能を両立する信号処理技術

受音系から出力された多観測信号を用いて，ねらった位置にある音だけをクリアに収音するための信号処理系を構築しました．マイクロホン間に生じる位相・振幅差を利用した指向制御技術だけでなく，雑音の出力パワーを最大で1万分の1まで低減するスペクトルフィルタ生成技術[8]を組み合わせることにより，出力音質と高い雑音抑圧性能を両立した信号処理技術を確立しました．

これまでに，録りたい音をクリアに収音するための原理が確立しつつあり，実験環境では，20 m離れた任意の位置にある音源をクリアに収音できることを確認しています．今後は実フィールドでの適用可能性について探っていきます．また，少ないマイクロホンでもクリアに収音するための技術改良を進めるとともに，映像分野の研究者や通信・放送といった分野で強みを持つ企業や大学とのコラボレーションを積極的に推進したいと考えています．

残響除去・制御技術

スポーツの臨場感において歓声は非常に重要な要素の1つです．歓声に包まれることで大きな臨場感を得ることができる一方で，歓声を抑えることで落ち着いたより分析的な視聴が可能になる可能性もあります．NTTでは臨場感のコントロールに大きな役割を果たす残響除去・制御にも取り組んできています．本技術はすでに収音された

図2　ズームアップマイクのシステム構成

コンサートを主な対象とした技術ですが併せて紹介します．

私たちの身の周りには，過去の素晴らしい演奏・楽曲を収めた名盤（CDやレコードなど）がたくさんあります．しかし，これらの名盤をステレオ再生したときに当時の収録音場（演奏会場）でその音楽を聴いているかのような臨場感が蘇るかといえば，必ずしもそうではありません．この理由の一端は，再生時にその演奏を収録したときと同じ音響環境を再現することが困難であるからです．

私たちがコンサートホールの観客席で音楽を聴いているとき，私たちの耳には大きく分けて2種類の音が舞台から到来します．1つは直接音で，これは観客席の前にある舞台から私たちの耳にまっすぐ届く音成分です．もう1つは残響で，これは舞台から放たれた音が壁や天井に反射して四方八方から到来する音成分です．CDなどには一般的に，観客席位置付近で収録されたような音が記録されており，直接音と残響は混ざって収録されています．このため通常のステレオ再生では収録時と同じ音響環境を再現することはできません．

私たちが開発した音信号に含まれる直接音・残響成分を分離する世界初の技術「残響制御」を用い，音楽信号を直接音と残響に分離し，直接音成分をサラウンド再生環境のフロントスピーカ，残響成分をフロント・リア両方のスピーカからそれぞれ再生すれば，演奏の収録時に類似した音響環境を再現でき，臨場感が蘇ります(9)．現在までに，有名海外アーティストの過去音源のサラウンド化や民生用オーディオ製品に応用され好評を博しています．今後は，放送分野などへの普及を目指すとともに残響制御処理の精度向上を目指した基礎研究を進めていきます．

今後の展開

本稿では高臨場観戦を盛り上げるためのいくつかの要素技術を紹介しました．高臨場観戦は映像と音響の両者に関し，撮影収音，符号化・配信，加工，視聴システムを含むさまざまな要素が関係する複雑な課題です．今後もなお一層NTT研究所の幅広い研究を融合させて遠隔地および会場での体験をより臨場感の高いものにしていきたいと考えています．

■参考文献

(1) 田中・越智：“4Kライブ映像インタラクティブ配信技術，”NTT技術ジャーナル，Vol.26，No.2，pp.59-62．2014.
(2) 特集：“高品質ロスレス・オーディオ符号化技術と展開，”NTT技術ジャーナル，Vol.20，No.2，pp.6-25，2008.
(3) 山根・山下・鎌谷・森崎・光成・尾本：“高臨場感音響ライブ配信トライアル，”NTT技術ジャーナル，Vol.23，No.7，pp.20-24，2011.
(4) Y. Kamamoto, N. Harada, T. Moriya, S. Kim, T. Yamaguchi, M. Ogawara, and T. Fujii: “Multichannel Audio Transmission over IP Network by MPEG-4 ALS and Audio Rate Oriented Adaptive Bit-rate Video Codec,” NTT Technical Review, Vol.11, No.7, pp.1-8, 2013.
(5) http://www.soumu.go.jp/main_content/000283104.pdf
(6) 志水：“JCT-3Vにおける3次元映像符号化方式の標準化動向，”映像情報メディア学会誌，Vol.67，No.7，pp.557-561，2013.
(7) K. Niwa, Y. Hioka, K. Furuya, and Y. Haneda: “Diffused Sensing for Sharp Directive Beamforming,” IEEE Transactions on Audio, Speech and Language Processing, Vol.21, No.11, pp.2346-2355, 2013.
(8) K. Niwa, Y. Hioka, and K. Kobayashi: “Post-filter design for speech enhancement in various noisy environments,” IWAENC2014, pp.35-39, Juan-les-Pins, France, Sept. 2014.
(9) 木下：“音声をよりクリアに，音楽をより豊かに――残響制御が切り拓く「音」の世界，”NTT技術ジャーナル，Vol.26，No.9，pp.20-22，2014.

（上段左から）鎌本　優/ 木下　慶介/ 國田　豊
（下段左から）志水　信哉/ 丹羽　健太/ 三上　弾

NTTではスポーツ・コンサートをはじめとするさまざまなイベントを高い臨場感で楽しむことを可能とする研究に，撮影・収音，伝送から合成・再現に至るまで幅広く取り組んでいます．これらの技術で遠隔地でも，また会場でもより魅力的な視聴体験を楽しんでいただきたいと思います．

◆問い合わせ先
NTTメディアインテリジェンス研究所
画像メディアプロジェクト
TEL　046-859-5179
FAX　046-855-1062
E-mail　mikami.dan@lab.ntt.co.jp