

人の本来の活動を邪魔しない自然なやり取りによる情報提供を実現するナチュラルコミュニケーション技術

NTTサービスエボリューション研究所では、バーチャルな人やモノに命がふきこまれ、本当に目の前に存在するかのように感じられる、高い臨場感体験の提供をめざし研究を進めています。本稿では、現実と見まがうほどの存在感を放つ人やモノ、あるいは空間そのものを創出し、遅延を感じさせない自然なインタラクションを通じてこれらを現実世界とシームレスにつなぐことで、自然な体験をユーザに価値として提供するためのナチュラルコミュニケーション技術の取り組みを紹介します。

さが た あつし さの たかし ひだか こうた
嵯峨田 淳 / 佐野 卓 / 日高 浩太

さとう たかし ふかつ しんじ むこうち たかふみ
佐藤 隆 / 深津 真二 / 向内 隆文

なが た ひでのぶ
長田 秀信

NTTサービスエボリューション研究所

現実空間と仮想空間の交差が生み出す「ナチュラル」な体験創出に向けて

私たちはこれまでに、離れた場所にいる人やモノの情報を伝送し、それらをまるごと遠隔地で再現することで、あたかも被写体が私たちの目の前に存在しているかのような体験の提供をめざしてきました^{(1),(2)}。高度に智能化されたアプリケーションやサービスが人々の生活に浸透し、VR (Virtual Reality)/AR (Augmented Reality) がより身近になりつつある現在、私たちは視聴者に提供する体験の幅をさらに広げ、また、より自然 (ナチュラル) な体験を提供したいと考えています。このとき、どこかにある人やモノの情報を忠実に再現する、というだけにとどまらず、投影される映像や音声を通じて実物と見まがう存在感までもが感じられ、いわば命が吹き込まれたかのような被写体をつくり出し、現実を超越した体験を価値として創出することをめざします。

このような体験を提供するために、どのような観点が重要となるのかを考察します。対象を正確に表現すること

は大変価値のあることですが、あえて変形や誇張を加えることで、よりインパクトある表現となる場合があります。例えば、葛飾北斎は、富嶽三十六景「神奈川沖浪裏」において、波がまさに動いている様を静止画で表現しています。波に動きを与えた独自の表現によって、写実的な絵画や写真で見えるものよりも、波の力を感じる方もいらっしゃるのではないのでしょうか。ある人が頭でイメージする大波の情景が、写真で見える波のそれよりも遥かに力強いものであったならば、北斎の絵画は現実を超越したリアルな体験をもたらしていると考えられます。

■Society 5.0

情報社会の到来は、現実空間のさまざまなものをデジタル化して利活用することを可能にし、本物を電子的に精緻に表現する取り組みを一般化しました。一方、内閣府では、狩猟社会 (Society 1.0)、農耕社会 (Society 2.0)、工業社会 (Society 3.0)、情報社会 (Society 4.0) に続く、新たな社会 Society 5.0 を、サイバー空間 (仮想空間) とフィジカル空間 (現実空間) を高度に融合させたシステムにより、

経済発展と社会的課題の解決を両立する、人間中心の社会、と定義しています⁽³⁾。すなわち、Society 5.0にかんがみれば、現実空間に存在する実物を精緻に模倣することに加え、現実と仮想との融合によって、現実を超越することが期待されているといえます。

■敵対的生成ネットワークGAN

現実と見まがう存在感を視聴者が感じ取ることができるような被写体映像の創出、あるいは本物を超越したような被写体を表現する、という観点においてポイントとなる技術としては、近年研究が進んでいる敵対的生成ネットワークGAN (Generative Adversarial Networks) が挙げられます⁽⁴⁾。この技術により、実写と見まがうような品質のCG画像の生成が現実になりつつあるほか、特徴を変換する処理により、1つの入力画像の特徴を反映した、全く別の画像を生成することができるようになりつつあります。これらの技術によれば、例えば人をセンシングして、動物の姿を生成するようなタスクの処理が可能となります。歌舞伎の演目の1つである「連獅子」では、谷を駆け上がる子獅子と見守る親獅子の演技を

通して、親子の情愛が表現される場面がありますが、ここで歌舞伎役者の動きをセンシングし、その結果から真の獅子の姿を描き出すことができれば、舞台を見ている視聴者が想像するイメージを描き出すことができます。このとき、獅子を写実的にライオンで表現しても効果的ですが、歌川広重の描いた「獅子の児落し」が現実にも目の前で起こったことのように体感することができれば、それは現実を超えた体験といえるかもしれません。

また、画像処理技術によって生成した情報を提示した際に、それらに命がふきこまれているかのように視聴者が感じるためには、生成物がセンシング対象を模倣したものではないと、ユーザに体感いただくことが重要です。模倣だと認識されると、生成された被写体への感情移入が困難となります。大変チャレンジングな課題ですが、生成された被写体が自律的に動作しているような視聴体感につなげるための技術が必要です。現在、骨格に基づいて人の姿勢を予測する研究が始まっており、あらかじめ学習された動作や、ある程度形の決まった、再現性のある動作を対象に、少し未来の姿勢を推測することができます。私たちは、センシング対象の被写体を自然かつ自律的に動作させ、例えば実在の人物とのインタラクションにおいて、処理の遅れなどを全く気にすることのない自然なやり取りの実現をめざします。

■ ナチュラルコミュニケーション技術

これらの実現に向け、私たちはナチュラルコミュニケーション技術の研究開発を推進します。ナチュラルコ

ミュニケーション技術は、次の5つから構成されています。

- ① 現実とみまがう空間・被写体などを自在に創出し、現実を超越し人間がイメージしているものを生み出す「空間創出技術」
- ② 伝送や処理等における物理的遅延を減らすだけでなく、人が感じる遅延による違和感といった感覚的遅延までもゼロにする「ゼロ・レイテンシメディア技術」
- ③ 2Dと3D表示をどちらも自然に見ることができる「2D・3D映像表示技術」
- ④ 現実と仮想空間の自然なインタラクションを実現する「ナチュラルインタラクションのための提示技術」
- ⑤ 人への直接的かつ自然に働きかける、五感+ α を伝送・提示するための新たな取り組み

空間創出技術

空間創出技術は、得られたセンシングデータ以上のデータを過去の同様のシーンや人物の形状を考慮することで

推定する技術です。私たちはこれまでに、深層学習を用いて2D映像から3D空間情報（CGモデル）をリアルタイムに生成する技術を開発してきました（図1）。本技術は、ニコニコ超会議2019「超歌舞伎 Supported by NTT今昔饗宴千本桜」⁽⁵⁾で実際に利用しており、2D映像には3D情報は内包されていませんが、過去の同様のシーンを基にしてリアルタイムに3D情報を生成し、超歌舞伎の登場人物を別のCGキャラクタへと“変身”させることで、超歌舞伎ならではの新たな演出を可能としました。

今後は、3D空間情報（CGモデル）と被写体抽出技術等を組み合わせることで、別の実写映像へと“変容”することにも取り組んでいくことを考えています。

ゼロ・レイテンシメディア技術

遠隔地への通信や、VR/ARなどによる仮想世界とのインタラクションにおいて、伝送や処理等における物理的遅延は自然なインタラクション実現における大きな課題となっています。こ



図1 リアルタイム3D情報生成

れまで、この物理的遅延を減らす努力を行い、ある程度実用化が進んでいますが、光の速さをもってしても遅延を物理的にゼロにすることはできません。そこで私たちは、自然なインタラクションを実現するために、物理的遅延をゼロにするだけでなく、人が感じる遅延による違和感をなくし、感覚的遅延までもゼロにする技術が必要になると考え、「ゼロ・レイテンシメディア技術」の研究に取り組んでいます。

具体的には、周囲の状況や行動パターンなど、さまざまな情報から感覚的遅延のメカニズムを解明し、遅延から生じる違和感を感じさせない、より自然な予測技術の構築、並びに、人の脳内予測のメカニズムを解明し、人の脳の内部で予測している世界との感覚的遅延の解消に向けた研究開発を進めていきます。

2D・3D映像表示技術

NTTは人の視覚特性を利用し、メガネをかけると3D画像を、メガネを外すと鮮明な2D画像を楽しめるステレオ映像生成技術「HiddenStereo」を開発しました⁽⁶⁾。本技術では、左右での視差を生み出す“視差誘導パターン”を1枚の2D画像に加算・減算することで左右画像を生成します。この視差誘導パターンは合成すると互いに打ち消し合い、裸眼で見ると元の2D画像しか見えなくなります。

視差誘導パターンを生成するためには、2D画像の奥行き情報が必要になります。例えば、ステレオ画像の場合はエピポーラ幾何等により奥行き情報が得られますが、撮影には多大な稼働と工夫が必要になります。一方で、単

眼カメラで撮影された通常の2D画像の場合は各画素の奥行き情報は得られないため、3次元CG空間に写像することで、奥行き情報を手動で作成するなどに対応しています。

私たちは、単眼カメラで撮影された2D映像を対象に、深層学習モデルを用いた奥行き推定、背景差分・フレーム間差分によるオブジェクト抽出、深層学習モデルを用いたインスタンスセグメンテーションにより、「HiddenStereo」の自動生成化、およびシステム化に取り組んでいます。

ナチュラルインタラクションのための情報提示技術

ナチュラルで臨場感のある情報提示技術として、360度テーブルトップ型裸眼3D映像表示技術（裸眼3D）（図2）と波面合成音響技術（波面合成）の研究開発を進めています。

裸眼3Dは、円形に配置された複数のプロジェクタと「空間結像アイリス面型光学スクリーン」という特殊なスクリーンを組み合わせることで、3Dメガネを使わずに、テーブル上のスクリーン上に立体像を見せることができる技術です⁽⁷⁾。直径120 cmの大型スクリーンの実現と、光学リニアブレンディングによって、従来技術より4分の1～10分の1の少ないプロジェクタ数であっても、滑らかな視点移動が可能なことを特徴としています。

波面合成は、複数のスピーカを用いて、単なる音の強弱や方向のみならず音場そのものを再現する技術です。これまでに直線状に並べたスピーカによって、歌舞伎の観客席まで音が飛び出す演出や、ゴルフボール競技のコー

ト上での音の再現などを行ってきました⁽⁸⁾。最近では、多重極スピーカアレイという複数の小型スピーカを格子状に密に並べた構成を採用することで、音の放射方向の制御を従来より小さなスピーカアレイで実現しました^{(9),(10)}。

これらの技術では、HMDや3Dメガネ、サラウンドヘッドホンなどの特別な装置を身につけなくても臨場感を享受できることが、ナチュラルなコミュニケーションを実現するうえで重要です。つまり、ユーザを取り巻く環境が高度化することで、ユーザに負担をかけずに自然な臨場感再生が可能になるのです。

応用分野として、スポーツやコンサートをよりリアルに体験するといったエンタテインメント分野は当然考えられます。それ以外にも、ビジネス分野においては、遠隔会議を「画面と音声の共有」から「空間の共有」に進化させるのに役立つでしょう。会議の隣席に、デジタルツインが投影された遠隔参加者がいて、ヒソヒソ話や筆談ができるようになるのも、夢でなくなるかもしれません。

家庭内においても、高齢者の方向にだけ音量や周波数特性を最適化した音声を再生するTVを実現できるでしょう。いわば部屋やTVが補聴器代わりとなり、家族そろってTVを見ても高齢者に合わせて音量を大きくする必要がなくなります。ヘッドホンを使わなくても、子ども部屋の方向には音が漏れないように制御して、こっそりTVを見ることも可能になるでしょう。

人への直接的かつ自然に働きかける、五感+αを伝送・提示する技術の確立に向けた新たな取り組み

情報を提示する手段としての映像や音の進化はめざましく、精細さの向上だけでなく、立体感なども体験することができるようになってきています。しかし、それ以外の感覚の活用は進んでいるとはいえません。さまざまな情報提示や体験を自然に行うには視覚、聴覚以外の感覚も積極的に活用する必要があります。臨場感の高い体験を生み出すサービスを実現するためには、視覚と聴覚だけでなく、触覚、嗅覚、さらには味覚にも訴える必要があります。また、情報提示という観点で考えた場合、視覚、聴覚といった重要な感覚器官を占有する手段は、必ずしもさ

りげない、自然な情報の伝達手段とはいえません。そのような意味でも、視覚、聴覚以外の手段、さらにはそれらを複合的に組み合わせた手段で情報を伝達できると、いついかなる場所でも自然に情報を受け取ることができるようになります。

現在は、皮膚感覚に情報を提示する研究開発に取り組んでいます(図3)。温度を利用した、視聴覚を邪魔しない情報提示技術「Thermal Bit Display」は、熱電素子を埋め込んだ指輪を唇に押し当てたときだけ情報を得ることができるデバイスです。プッシュ型通知と異なり、知りたいときにだけ情報を確認することができ、しかも視覚や聴覚を利用しないため、さりげなく情報を確認することができるという特徴が

あります。

また、渦輪(いわゆる空気砲で発生させることができる空気の動き)を利用した刺激による新しいリアリティの表現にも取り組んでいます。刺激の時間差を制御することによって、何かが自分の近くを通過していくかのような感覚を疑似的に生み出す技術の研究にも取り組んでいます。刺激の場所を含めた制御により通過以外の感覚を生み出したり、他の感覚、例えば音と組み合わせた刺激の提示により、音源の移動のリアリティをさらに向上させることができると考えています。

これからも、視覚、聴覚以外の感覚の伝送により、よりナチュラルに体験や情報を受け取ることができる技術の

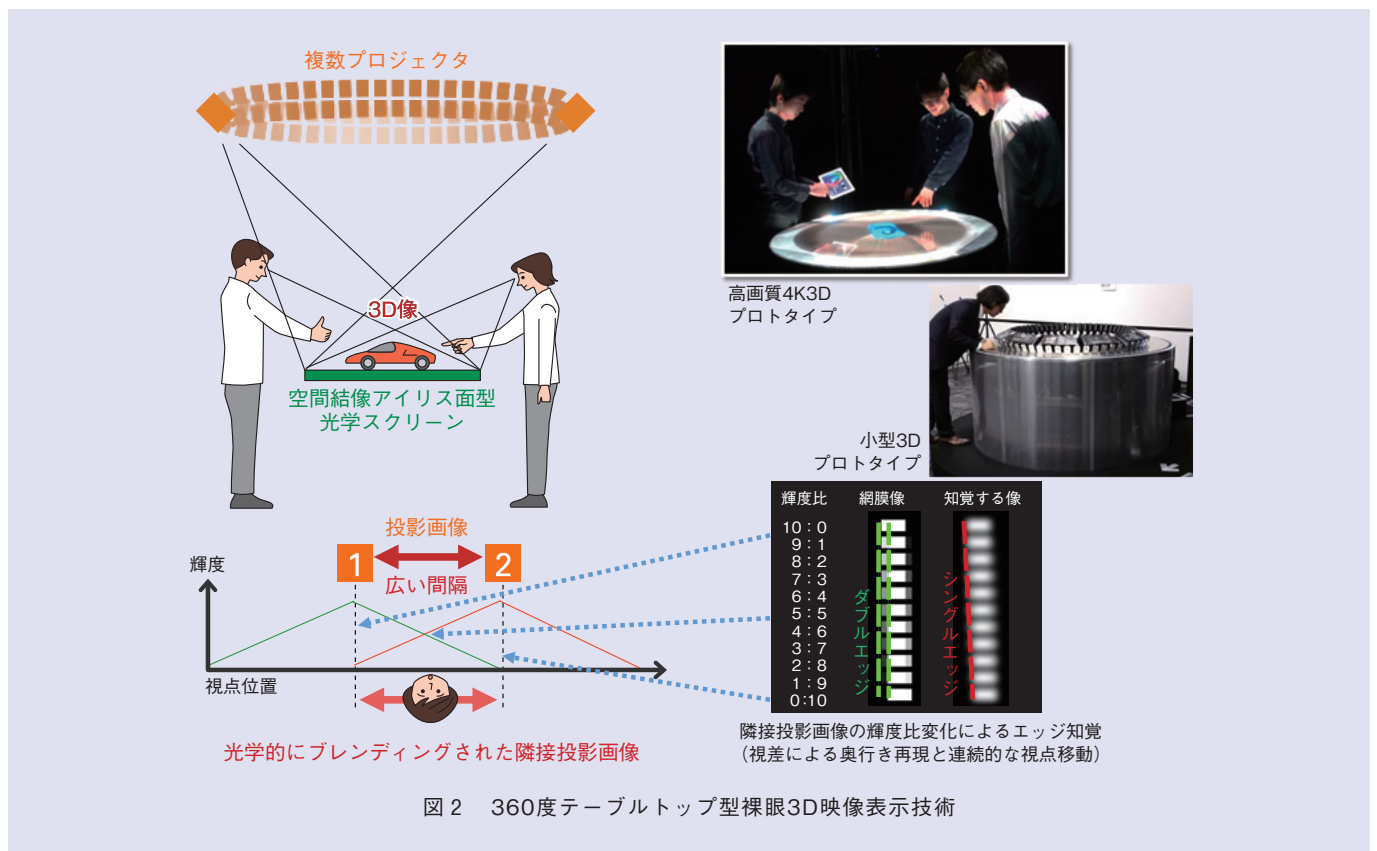




図3 皮膚感覚への情報提示技術

確立に向け、研究開発を進めていきます。

おわりに

現実空間と仮想空間の間で、情報のやり取りを含む自在なインタラクションを実現させ、現実と仮想空間を自在に交差するためのナチュラルコミュニケーション技術の研究開発内容について述べました。仮想空間はミラーワールド⁽¹¹⁾として述べられることがありますが、単に電子化された世界ではなく、現実空間と“交差”することで、もう1つの世界“パラレルワールド”のような役割を果たすと考えています。理想的には、上手に歌えることや魅力的に踊れるなど、それぞれのユーザがなりたいた姿をパラレルワールドで

体感することのみならず、現実世界にフィードバックすることが望ましいと考えます。この“交差”をテーマの1つとして、私たちは研究開発を進めていきます。

■参考文献

- (1) 長田・宮下・柿沼・山口：“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.29, No.10, pp.33-37, 2017.
- (2) 長尾・宮下・佐野・長谷川・井阪：“Kirari! for Arena：奥行感のある4方向イベント観覧体験の創造,” 日本画像学会誌, Vol. 58, No. 3, pp. 306-315, 2019.
- (3) https://www8.cao.go.jp/cstp/society5_0/index.html
- (4) I. J. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D. W. Farley, S. Ozair, A. Courville, and Y. Bengio: “Generative Adversarial Networks,” Proc. of NIPS2014, pp. 2672-2680, Motreal, Canada, Dec. 2014.
- (5) <https://www.ntt.co.jp/news2019/1903/190325b.html>
- (6) T. Fukiage, T. Kawabe, and S. Nishida: “Hiding of Phase-Based Stereo Disparity for Ghost-Free Viewing Without Glasses,” ACM Transactions on Graphics, Vol.36, No.4,

- pp.147: 1-17, Los Angeles, U.S.A., July 2017.
- (7) 巻口・高田：“360度テーブルトップ型裸眼3D映像表示技術,” NTT技術ジャーナル, Vol. 30, No.10, pp.25-29, 2018.
 - (8) 堤・高田：“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, Vol.29, No.10, pp.24-28, 2017.
 - (9) K. Tsutsumi, K. Imaizumi, A. Nakadaira, and Y. Haneda: “Analytical Method to Convert Circular Harmonic Expansion Coefficients for Sound Field Synthesis by Using Multipole Loudspeaker Array,” Proc. of EUSIPCO 2019, A Coruña, Spain, Sept. 2019.
 - (10) K. Imaizumi, K. Tsutsumi, A. Nakadaira, and Y. Haneda: “Analytical Method of 2.5 D Exterior Sound Field Synthesis by Using Multipole Loudspeaker Array,” Proc. Of WASPAA 2019, New York, U.S.A., Oct. 2019.
 - (11) <https://wired.jp/special/2019/mirrorworld-next-big-platform>



(後列左から) 長田 秀信/ 嵯峨田 淳/
日高 浩太
(前列左から) 深津 真二/ 佐藤 隆/
向内 隆文/ 佐野 卓

視覚・聴覚のみならず五感のすべてを活用し、仮想世界のヒト・モノとシームレスにつながることで、これまでできなかった新たな体験を提供する。そんなナチュラルコミュニケーションの実現に向けこれからも研究開発に取り組んでいきます。

◆問い合わせ先

NTTサービスエボリューション研究所
ナチュラルコミュニケーションプロジェクト
TEL 046-859-3901
E-mail ev-journal-pb-ml@hco.ntt.co.jp