

## ”光”新世代ビジョンを支える研究開発の動向

# コミュニケーション・コラボレーション技術

### 概要

NTTの掲げる「“光”新世代ビジョン」は、「光による本格的なブロードバンド・ユビキタス時代」を展望し、「ユーザビリティに優れたレゾナントなコミュニケーションネットワーク環境の実現と、これを活用する多彩で豊富な新サービス、ビジネス機会の創出」を目指している。NTT研究所では、レゾナントコミュニケーションネットワーク（RENAネットワーク）上に展開される、よりリアルで自然なコミュニケーション環境を実現する基盤技術の研究開発を進めている。本稿では、これらのコミュニティや企業における円滑なコミュニケーションをサポートするコミュニケーション・コラボレーション技術について紹介する。

### はじめに

わが国のインターネット利用人口は2002年末には6900万人に上り、国民の2人に1人がインターネットを利用している。このうち約3割にあたる2000万人が既にブロードバンドユーザーである。2003年3月現在、世界で最も低廉・高速なブロードバンドを提供し、今後5年間でブロードバンド利用人口は6000万人に達す

ると見込まれている。このようなブロードバンド化の進展に伴い、E-mailやWWWを牽引役として普及してきたインターネットは、コンテンツ配信、E-Learning、ネット

ワークゲームなど、より多くの場面で日常的に利用され始めている。

通信といえば電話の時代から、今日では、コンテンツ配信系のデータ通信が通信の多くを占めつつあるものの、電話やE-mailに代表されるように、通信の大きな役割は、人と人とのコミュニケーションを支援することにある。家族の語らいや仕事の打ち合わせなど、対人コミュニケーションは日常のあらゆる場面で行われている。人と人とのコミュニケーションは社会活動の基盤であり、通信にとって依然重要なコンテンツといえる。

しかしながら、電話やE-mailのような従来のコミュニケーションメディアは、対面対話とは異なった面でその有用性を捉えることはできるものの、基本的にナローバンドの上に成り立っているため、対面対話が本



日本電信電話株式会社  
NTTサイバースペース研究所  
主任研究員  
八木 貴史



日本電信電話株式会社  
NTTサイバースペース研究所  
主任研究員  
小林 稔

来持っているリアリティやアフォーダンス（周囲環境から得られる価値）の多くを失ってしまっている。そのため簡単な用件を伝えることはできても、より深い議論を行ったり、感情を伝えたりすることは困難である。

NTT研究所では、これまで高度なコミュニケーション・コラボレーション環境の実現を目指し研究開発を進めてきた。パーソナルマルチメディア通信会議システムPMTTC（Personal Multimedia-multipoint TeleConference System）は、汎用のワークステーションを用いて在席で資料などを共有しながらの多地点テレビ会議を実現した。TeamWork Stationは個人と協同の作業空間をシームレスに統合することで、自由度の高い作業空間を提供した。ClearBoardは対話空間と描画空間をシームレスに統合することで、意思

疎通感の高い協同描画空間を実現した。また、インタースペースは三次元の共有仮想空間を構築し、その空間を通して多様なコミュニケーションを可能にした。これらの取組みは多くの可能性を示唆してきたが、これまでの通信環境やコンピューティングパワーの下では、コンセプトの可視化にとどまり、実用に供するのは困難であった。

光によるレゾナントコミュニケーション環境では、これまでに培われてきた技術やコンセプトをベースに発展させた技術をRENAネットワーク上に展開することで、従来メディアの失ってしまったリアリティやアフォーダンスの伝達を実現する。これにより、よりリアルで豊かなコミュニケーションが可能になる。

本稿では、初めに高品質なコミュニケーション環境実現の基盤となるメディアの符号化および処理技術について解説する。そして、高度なコラボレーション環境を提供する多地点コラボレーション技術、コミュニケーションの可能性を広げる「場」の通信技術について紹介する。

### メディア符号化技術

#### 1. 映像符号化技術

映像の符号化技術は、ITU-T (International Telecommunications Union - Telecommunication Standardization Sector) やISO (International Organization for Standardization) から国際標準として多くの方式が勧告化されている。ISDN時代

のテレビ電話・会議用のITU-T H.261に始まり、Video CD等蓄積用途のMPEG-1、通信からデジタル放送、DVDをカバーする汎用高品質符号化MPEG-2、H.263をベースにさらに高圧縮・高機能化を図ったMPEG-4などが標準化され、既に多くの製品やサービスに利用されている。

#### (1) MPEG-2符号化技術

MPEG-2は、デジタル放送やDVDでの利用が有名であるが、蓄積メディアを対象に勧告化されたMPEG-1とは異なり、放送、通信、蓄積などの幅広い要求条件を考慮して勧告化された方式であり、その応用範囲は広い。“光”のさらなる普及により、MPEG-2は超高品質版のレゾナントコミュニケーションを実現する上で重要なツールとなる。

NTT研究所が開発した標準TVを越える高画質対応のコンシューマ向けCODEC LSI (開発コード名: ISIL) は、480/60P (プログレッシブ) の高画質映像のエンコード (圧縮) /デコード (伸張) の同時処理 (全二重) を、世界で初めて1チップ上で実現する。1.5W以下という低消費電力化と低コスト化を併せて実現しており、幅広いコンシューマ向け機器への組み込みが可能である。パソコンはもちろん、デジタルビデオカメラや各種モバイル機器に搭載でき、ユビキタスな環境での双方向・高品質ビジュアルコミュ

ニケーションが可能となる (図1)。

NTT研究所は、本格化するHDTV時代に向け、HDTV映像の圧縮・伸張処理の分野においても精力的な研究開発を進めており、従来、専用チップを複数個用いて実現していたMPEG-2準拠HDTV CODECの1チップ化に世界で初めて成功した (開発コード名: VASA)。VASAは、デジタルシネマ、ステレオやマルチアングルTV等、将来の高臨場感映像通信の分野への活用が期待され、本格的な超高品質・超高臨場感レゾナントコミュニケーションの実現に寄与すると考えている。

NTT研究所では、ハードウェアのみならず、ソフトウェアコーデックの実現にも力を注いでいる。これまでに光ブロードバンドネットワークとパソコンだけで、リアルタイムかつ双方向に標準TV並みの高品質映像伝送を実現するMPEG-2リアルタイムソフトウェアコーデック (開発コード名: RISCA: Real-time and Interactive Software-based Codec Architecture) を開発している。これにより、従来必要とされていた高価な専用ハードウェアと専用

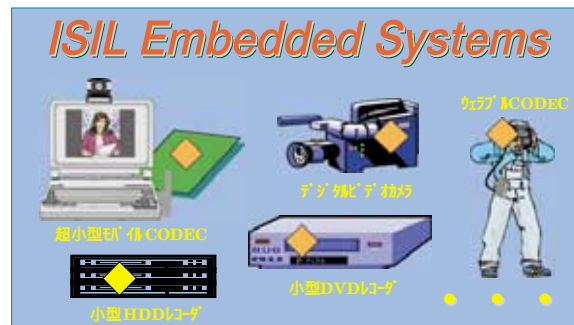


図1 ISILの適用例

回線を必要とせず、標準TV並みの高品質映像コミュニケーションを手軽に利用できるようになる。

## (2) MPEG-4 符号化技術

MPEG-4は、高圧縮・多機能を狙った符号化技術で、現在もさらに進化を続けている。パソコン間や第3世代携帯電話(NTTドコモのFOMA)間のテレビ電話に用いられるシンプルプロファイルをベースに、高品質化のため、両方向予測、1/4画素精度の動き補償、カメラのパニングやズームリングを補償するグローバル動き補償等を加えたアドバンストシンプルプロファイルがある。グローバル動き補償は、NTT研究所が提案し、採用されたツールで、シンプルプロファイルとの比較では、パニングやズームリング等を伴うシーンで最大2倍程度の符号量削減効果がある。また、多様な速度のネットワークに動的に適合させるためのスケラビリティを加えたストリーミングプロファイルや、放送素材を扱うスタジオプロファイル、CG/アニメ系のプロファイル等、バリエーションは多岐にわたっている。さらに、最近では、さらなる高圧縮を目指して、JVT (Joint Video Team: ITUとISOのジョイント)にて検討が進められてきたJVT CODECがある。ITU-TではH.264として、ISOではMPEG-4 Part 10として2003年に標準化された。

現在のパソコンの処理能力では、シンプルプロファイルが高々アドバンストシンプルプロファイルまでが

リアルタイム処理可能な範囲であるが、今後は、パソコンの高速化と符号化アルゴリズムの進化により、リアルタイム処理を保ちつつ、さらなる高品質化が図れると考えられる。

## 2. 音声符号化技術

### (1) MPEG-4 オーディオ符号化技術

MPEG オーディオは、映像の符号化に付随する規格として標準化が進められてきた。MP3としてインターネット上での音楽配信サービスや携帯型音楽プレーヤー等に広く用いられているMPEG-1オーディオLayer3方式や、衛星デジタル放送や地上波デジタル放送にも用いられるMPEG-2 AAC (Advanced Audio Coding)方式が有名である。MPEG-4では、双方向通信や無線通信等の多様な利用形態を想定して標準化が進められ、その結果、従来のMPEGで対応してきた広い周波数帯域の楽音符号化のみならず、CELP (Code Excited Linear Prediction)を含む音声符号化、さらには、テキスト音声合成も統一的に扱えるものとなった。

NTT研究所は、MPEG-4オーディオの標準化作業に大きく貢献し、NTT提案のTwinVQ方式も採用されるに至った。TwinVQ方式はベクトル量子化技術を基本とした楽音用符号化であり、低ビットレート時も品質劣化が少なく、64kbit/s以下の従来の音声符号化の領域でも高い品質を実現している。さらに、圧縮ビット列が固定ビット配分であるため誤り耐性が高く、ランダムアクセス

が容易であるなど、従来の楽音符号化にはない特徴を持つ。

### (2) MPEG-4 オーディオの応用技術

NTT研究所では、MPEG-4オーディオ準拠の楽音符号化を応用したロスレススケラブル符号化方式“Shalls”を開発した。ロスレス符号化とは、原信号と1ビットたりとも異なる信号を複合化できる完全に可逆変換可能な符号化方式であり、スケラブル符号化とは、圧縮ビット列が多段構成になっており、その1段のみ、または数段まとめて復号して復号信号を得ることができる符号化方式である。ロスレススケラブル符号化とはその両方の特徴を兼ね備え、全階層の圧縮ビット列を用いて復号するとロスレス符号化となる符号化方式である。

ロスレススケラブル符号化“Shalls”の圧縮ビット列は、MPEG-4オーディオが基本階層であり、これに原信号との誤差のロスレス圧縮ビット列を順に積み上げた階層構造である。したがって、最低のビットレートとする時はMPEG-4オーディオ圧縮ビット列のみを用い、中間のビットレートに対応する時はロスレス圧縮ビット列を順に積み上げて用いる。情報量が約1400kbit/sである44.1kHzサンプリング16bitのステレオ音源に対して、ロスレス圧縮ビット列全てを用いた場合で平均約700kbit/sとなる。16~80kbit/sの高圧縮MPEG-4オーディオから、約700kbit/sまでほぼ連続的にビットレートを選択することができる。これ



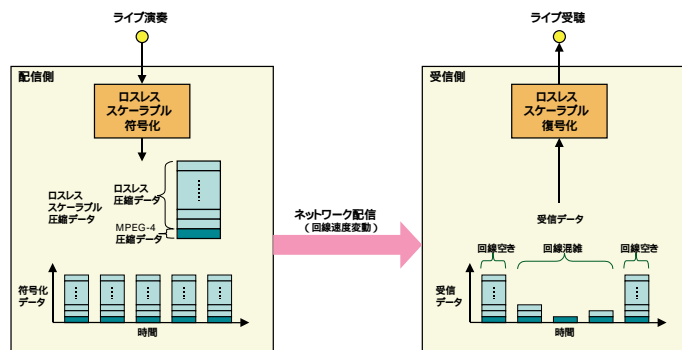


図2 ロスレススケラブル符号化の適用例

により、帯域変動を伴う無線から高品質な光ネットワークまで様々な環境で、音切れや不快な雑音が発生しない質の高い快適なサービスが可能となる(図2)。

## メディア処理技術

### 1. エコーキャンセラ技術

レゾナントコミュニケーションにおいては、高品質な音響通信はもちろん、あたかも同一空間内で話し合うような違和感のない会議や、リアルな対話の実現は重要な課題である。このようなサービスを実現するためには、原則的に片方向である放送や配信系サービスとは異なり、スピーカから出た音がマイクに回り込むことによって生じるエコーの問題を解決する必要がある。

電話機のようなハンドセットや、近年パソコンで利用されているヘッドセットを用いればエコーの問題は回避できるが、より自然で自由度の高いコミュニケーションを実現するには、室内等に配置されたマイクとスピーカを利用するハンズフリー型のコミュニケーションが望ましい。

### (1) エコーキャンセラ構成技術

エコーキャンセラの構成を図3に示す。エコーキャンセラの原理は、適応フィルタの特性を、スピーカ・マイク間の音響伝達特性と同じになるように推定し、マイク收音信号から差し引くことでエコーを消去するというものである。

エコーキャンセラの基本性能として、エコーを消去するまでの所要時間がある。NTT研究所は、音響伝達特性の推定を高速に行うES射影アルゴリズムを開発している。この方式は、従来の代表的なアルゴリズムである学習同定法と比べ、約4倍の速度で推定を行うことが可能であり、エコー消去時間を大幅に短縮している。また、ステレオ等の多チャンネル型のエコーキャンセラにおいても、各チャンネルの信号間の相関を除去した上で、音響伝達特性を推定するステレオ射

影アルゴリズムを開発し、従来の1チャンネル型エコーキャンセラを単純に多チャンネルに拡張したものと比べ、推定時間を5分の1に短縮している。

空間内にスピーカから再生された音以外の音(雑音)が存在すると、十分なエコーの消去が行えなかったり、雑音により送出する主音が聞き難くなるという問題が生じる。これは、空調音やOA機器の雑音のあるオフィスや、走行音のある自動車内など一般的な環境で起こりえる問題である。そこで、NTT研究所では、マイクで收音した信号を周波数領域に変換し、雑音および雑音に埋もれてこれまで消去できなかったエコーを抑圧するノイズ・エコー同時抑圧技術“Noer”を開発し、前述した音響伝達特性推定技術と組み合わせることで広範囲な環境における快適なコミュニケーションを可能としている。

### 2. ゾーン分離收音技術

NTT研究所は、複数の音源が混在した空間から、ある特定のゾーン内で発せられた音を抽出して收音す

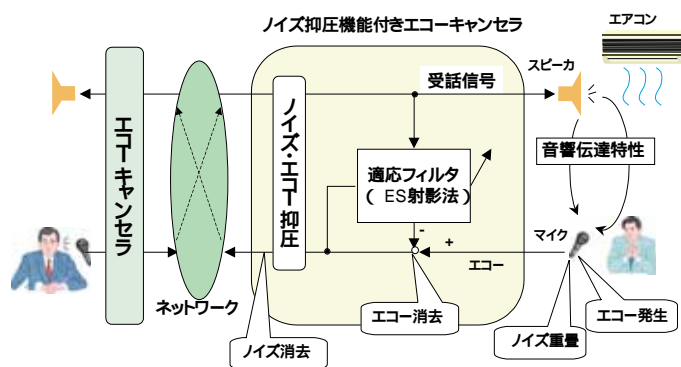


図3 音響エコーキャンセラの構成

るゾーン分離收音技術“SAFIA”を開発した。

従来は指向性マイク1本を用い、一定方向の音を收音する方法が取られていたが、単一マイクだけでは、目的とする方向以外からの音を十分に抑圧することができなかった。SAFIAは、複数の指向性マイクを用いて各マイクへの到達レベルや位相差から、目的とするゾーンの音をクリアに收音する。従来の指向性マイクと比べて20dB以上の分離收音性能を発揮する。

この技術を用いれば、例えば、室内を收音するコミュニケーションゾーンと收音しないプライベートゾーンに分け、テレビ会議をコミュニケーションゾーン内で行うことで、出席者の発言のみを伝達し、プライベートゾーンの人の声や物音は伝達しないということも可能になる。

### 多地点コラボレーション技術

井戸端会議からビジネスの会議まで、Face-to-Faceの環境では、3人以上が参加するコミュニケーションは日常的に存在する。遠隔地においても、Face-to-Faceと同様なコミュニケーションを実現するためには、多地点を接続してのコミュニケーション環境の提供が必須である。

従来、多地点間の会議や打合せには、交換機の付加サービスとしての電話会議や、テレビ会議端末を多地点接続装置で接続した多地点テレビ会議システムが用いられてきた。これらのシステムは、固定のISDN回

線や専用端末を必要とし、会議の場所が固定され、会議資料は事前にFAX等で配布するなど、必ずしも使い勝手の良いものではなかった。一方、パソコンの高性能化、常時接続ネットワークの低廉化とブロードバンド化など、パソコンを利用した多地点間コラボレーションの環境も整いつつある。

NTT研究所が開発した多地点コラボレーション支援システム“NetOfficeHIKARI”は、パソコンとIPネットワークがある環境なら、オフィス、外出先、自宅を問わず、あらゆる場所からコラボレーションワークへの参加を可能にする。最大32地点間の映像・音声通信によるコミュニケーションに加え、Microsoft PowerPointファイルやWebページを全地点で共有できる資料共有機能により、参加者全員が同じ資料を供覧しながら遠隔会議や遠隔講義、ネット上の講演会等を行うことができる。パソコン上に資料を表示し、マーキングなどの描画も共有できるので、打合せや講義の効率は格段に向上する。

NetOfficeHIKARIの大きな特長は、Windowsアプリケーションの画面と操作情報を多地点間で共有するアプリケーション共有機能を実装したことである。これにより、任意のWindowsアプリケーションを多地点間で共有して操作することができる。Microsoft Excelを共有して



図4 NetOfficeHIKARIの利用シーン

いる例を図4に示す。たとえば、東京、大阪、名古屋等の各支店の売り上げを各地点が個々に入力し、その場でグラフ化するという使い方が可能になる。

NetOfficeHIKARIでは、GUIはWebアプリケーションとして作成することができるので、会議、講義、講演など、サービスに応じて柔軟に画面をカスタマイズすることが可能である。図4は、数名(最大32名)の参加者が一つのバーチャルルームに会する画面構成である。よりリアルな対面型のサービスを想定し、大画面の高品質映像によるテレビ電話的な使い方も可能である。

NetOfficeHIKARIは、企業内・企業間利用ではテレビ会議や遠隔協調作業、企業の顧客向け利用では遠隔講義やビジュアルコールセンタ、マス向けにはテレビ電話やビデオチャットといった様々な使い方が可能である。NTT研究所は、NetOfficeHIKARIのような多地点コラボレーション支援システムは、1対1から

多地点間をカバーするレゾナントコミュニケーションの一実現形態と考  
えており、さらなる高品質化、使い  
勝手の向上に向けて研究開発を進め  
ている。

### 「場」の通信技術

コンピュータグラフィックス  
(CG) やバーチャルリアリティ  
(VR) 技術の発展に伴い、これらを  
人と人のコミュニケーション手段と  
して応用する取組みがなされている。  
急速に普及しつつあるブロード  
バンドネットワークを活かした高臨  
場感メディア通信により、遠隔地に  
いる人同士が、あたかも同じ場所  
にいるかのような感覚を持ってコミュ  
ニケーションすることができるよう  
になれば、我々のワークスタイルや  
ライフスタイルは大きく変わり、よ  
り豊かな環境が実現する。

前述のNetOfficeHIKARIは、現在  
のパソコンで実現し得る現実的な解  
の一つであるが、NTT研究所は、よ  
り本格的な「場」の通信システムを

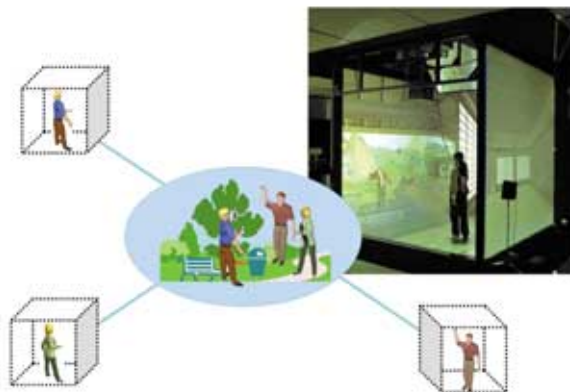


図5 仮想空間による場の共有とGAVAの概観

目指した研究も  
進めている。以  
降、「場」の通  
信システム  
“GAVA  
(Generation  
and Accelerati  
on environment  
for Virtual and  
Augmented  
reality commu  
nication)” を中  
心とした取組みを紹介する。

### 1. 「場」の通信システムGAVA

「場」の通信の目指すところは、  
場の雰囲気も含めた「場」の総合的  
な共有である。空間的に離れた人同  
士が、あたかも同じ場所にいるよう  
な感覚で交流でき、協調して作業で  
きる環境の実現を目指している(図  
5)。NTT研究所が開発した「場」  
の通信システムGAVAは、こうした  
次世代コミュニケーションの形態  
を、没入型仮想空間表示技術CAVE  
(Cave Automatic Virtual Environ

ment: イリノイ大学  
で開発された仮想空  
間表示システム) 等  
のVR技術を用いて可  
視化することで、そ  
の効果や技術課題を  
整理し、次世代のコ  
ミュニケーションの  
あるべき姿を探索す  
ることを狙っている。

図5にGAVAの概  
観を示す。CGでモデ

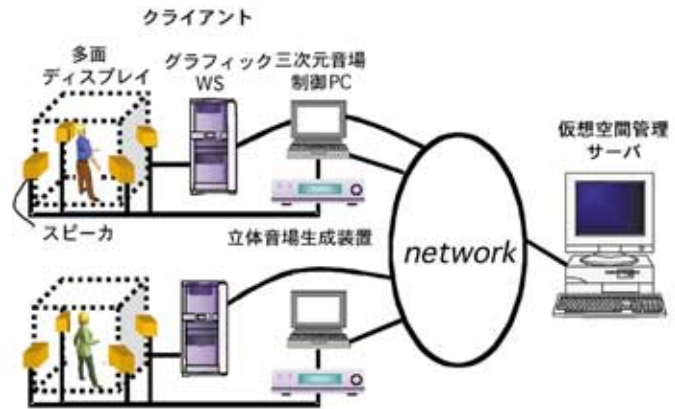


図6 GAVAのシステム構成

リングされた日本庭園に入り、ネッ  
トワークを介してCG表示された相  
手の分身(アバタ)とコミュニケー  
ションしている様子を示している。  
没入型の表示装置に実物大表示する  
ことで、相手との距離感等も含め、  
現実空間に近い形のコミュニケーシ  
ョンの実現を狙っている。

システム構成を図6に示す。CAVE  
に音響機器やユーザーの動作を入力  
するセンサ装置を加えたクライアント  
端末装置は、ネットワークを介し  
て仮想空間管理サーバに接続される。  
仮想空間管理サーバは、仮想空間の  
構成情報と、クライアント端末から  
送られたユーザーの位置情報等を管  
理し、各端末の要求に応じて配信す  
る。各クライアント端末では、サー  
バから受け取った他のユーザーの位  
置情報等をもとに、他のユーザーの  
姿をCGアバタとして表示する。

### 2. GAVAによるジェスチャ・表情の伝達

コミュニケーションでは、ジェス  
チャや仕草などの情報も重要な役割



を果たす。GAVAでは、ユーザーの動作を磁気位置センサで計測し、仮想共有空間中のアバタの動作に反映させることで、ジェスチャの伝達を可能にしている。人間の全関節の動きを計測するには多数のセンサが必要となるが、多くのセンサを身体に装着することは非現実的である。そこで、ロボットの制御などに用いられる逆運動学を用いて、両手首と背中、頭部に取り付けられた4個のセンサ情報から、身体全体の動きを推定している。また、用途によっては、ユーザーの動作を推定することで、装着センサ数をさらに減らすことも可能である。たとえば、GAVAを用いたバーチャルゴルフゲームでは、リアルタイムにセンサ情報を反映させるリアルタイム制御と、予め用意されたパターンに沿った動作を行うパターン制御とを切り替えながらゲームを進行させる状態制御を行っている。

身体のジェスチャだけでなく、視線や口の動きなどの表情もコミュニケーションでは重要である。GAVAでは、小型の視線検出装置により利用者の視線方向を検出し仮想空間中のアバタの目の動きに反映させ、また、マイクに入力される音圧を口の動きに反映させることで、表情豊かなアバタ



図7 表情付きアバタ

表現を可能にしている(図7)。

### 3.GAVAの今後の発展形態

GAVAでは、前述のような視覚的な表現だけでなく、立体音場技術により環境音やユーザーの声を空間的に提示することで、発話者の特定や、発話者に対する笑いや拍手等、場の雰囲気共有できる環境の実現を目指している。また、本格的な「場」の通信の実現には、視覚・聴覚の他に、触覚や力覚の共有も重要な要素である。NTT研究所では、物体の存在や動きを表現するために、風圧を利用して力覚を表現するインタフェースにも取り組んでいる。本インタフェースをGAVAと組み合わせ、仮想空間中で跳ねるカップをハンマーで叩くゲームの形で実現した例を図8に示す。ユーザーは、風を受け止める風受容器を先端に取り付けた棒をハンマーの代わりに握る。風受容器の動きを位置センサで計測し、仮想空間中のカップとの衝突に合わせて床に置かれたノズルから空気を噴出することで、物体との衝突感を表現する。風を風受容器で受け止めることにより、風圧を「力」として感じさせることを実現している。

風圧による力覚提示では微妙な触覚を提示することは困難であるが、ユーザーを器具などで拘束しない特長がある。ユーザーを拘束しない特質は、将来これらの技術が生活の中に融け



図8 風圧による力覚提示インタフェース

込み日常の道具として利用される場合に、ヒューマンインタフェースの重要な設計要件になると考えられる。

## おわりに

レゾナントコミュニケーション環境の実現に向けた上位レイヤの取り組みとして、コミュニティや企業における円滑なコミュニケーションをサポートするコミュニケーション・コラボレーション技術について紹介した。RENAネットワーク上にこれらの技術を展開することで、まさに人と人との思いが共鳴する豊かなコミュニケーション環境が実現することが期待される。

### 問い合わせ先

NTTサイバースペース研究所

八木 貴史

TEL 046-859-4501

E-mail:yagi.takashi@lab.ntt.co.jp