

第II章

ユニバーサル コミュニケーション 基盤技術



木俵 豊
岩澤昭一郎
大井隆太郎
翠 輝久・水上悦雄
内山将夫
吉田俊介
金 京淑・村上陽平
隅田英一郎・柏岡秀紀
風間淳一
安藤広志
岩爪道昭
荒川佳樹

コミュニケーションの壁を越えるための ユニバーサルコミュニケーション基盤技術



木俣 豊 (きだわら ゆたか)

ユニバーサルコミュニケーション研究所
研究所長

民間企業勤務を経て、2001年独立行政法人通信総合研究所(現 NICT)入所、2006年内閣府出向、2007年知識処理グループリーダーを経て2011年よりユニバーサルコミュニケーション研究所長。ユビキタスコンピューティング、情報分析の研究などに従事。趣味はドライブや写真撮影など。

「本稿では、NICT ユニバーサルコミュニケーション研究所で研究開発を進めている、言語の壁、情報の量と質の壁、距離や臨場感の壁などを越えるためのユニバーサルコミュニケーション基盤技術の概要を説明します。」

はじめに

ブロードバンドネットワーク技術やユビキタスコンピューティング技術によって、いつでもどこでも高速なネットワークの恩恵を享受できるようになりました。また、スマートフォンの普及によって、必要な ICT サービスを即座に利用できるようになっています。このように、情報通信インフラは充実しつつありますが、人と人がコミュニケーションするためには、まだまだたくさんの壁があります。

例えば、外国人とコミュニケーションする場合には「言葉の壁」があります。インターネットを通じて情報を利用する際には、大量の情報によって、信頼性が高く価値の高い情報の発見を阻む

「情報の量と質の壁」があります。さらには遠隔地の友人と簡単にテレビ電話をすることができませんが、限られた 2 次元のディスプレイによる「距離や臨場感の壁」があります。これらの壁を越えるには、自動翻訳技術や情報分析技術、超臨場感通信技術などのユニバーサルコミュニケーション基盤技術の研究開発が必要不可欠です。

ユニバーサルコミュニケーション技術

NICT ユニバーサルコミュニケーション研究所では、「言語の壁」を越えるための多言語音声翻訳や「情報の量と質の壁」を越えるための情報分析技術、「距離や臨場感の壁」を越えるための

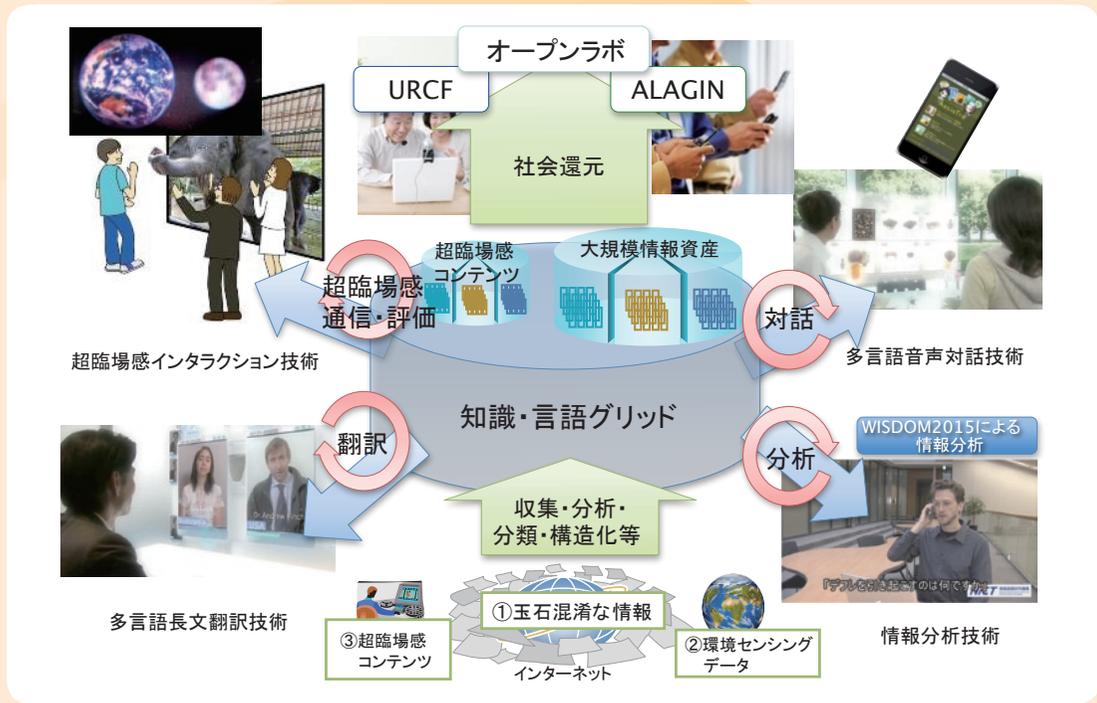


図 ユニバーサルコミュニケーション基盤技術の概念図

超臨場感通信技術を研究・開発しています。また、これらの技術を活用するために新世代ネットワークを活用した知識・言語グリッドと呼ばれる新たな情報活用基盤上で大量の Web 情報やセンサ情報などから構築された大規模高度情報資産を作り上げます。さらには、情報資産から「知」の情報を取り出し、多様な情報サービスへと利用できる仕組みを構築します。

この情報活用基盤によって、言語処理や知識処理の専門的な知識を持たないベンチャー企業などでも、高度な知識処理技術を用いた情報システムが容易に構築できるようになります。これらの技術を用いて既に多言語音声翻訳システム VoiceTra4U-M、音声対話システム AssisTra、情報分析システム WISDOM や一休などを開発しています。また、多感覚インタラクションシステム、臨場感あふれる映像を表示できる200インチ裸眼立体ディスプレイ及び究極の立体ディスプレイである電子ホログラフィも開発しています。これらのシステムを実現するためには、着実な研究開発が必要不可欠です。私たちは、高度な自然言語処理技術や音声処

理技術、多言語機械翻訳技術、大規模情報管理技術、情報検索技術、サービスコンピューティング技術、画像符号化圧縮伝送技術、電子ホログラフィ技術など多岐にわたる分野の基礎研究を行っています。

おわりに

これらの研究開発の成果は、単なる学術的な成果だけに終わらず、実用的に使えるレベルを目指すことで、他に類を見ない世界最先端の「社会で使ってもらえる」成果となっています。例えば、多言語音声翻訳技術などは、既に商用サービスに組み込まれて利用されており、一般の人々の生活に深く関わっています。このような社会への成果展開を目指した共同研究・開発のために、高度言語情報融合フォーラム(ALAGIN)と、超臨場感コミュニケーション産学官フォーラム(URCF)を設立しており、産学官での協力体制も進んでいます。要素技術の更なる高度化のみならず成果の社会展開を加速させながらユニバーサルコミュニケーション領域における国際的 COE 研究拠点を目指します。

世界初！200インチ 裸眼3Dディスプレイによる 自然な立体視の実現

岩澤 昭一郎 (いわさわ しょういちろう)

ユニバーサルコミュニケーション研究所
超臨場感映像研究室 主任研究員

大学院博士課程修了後、民間研究機関等を経て2006年からNICTにて大画面裸眼立体表示技術の研究開発に従事。小学校時代からの強度近視+乱視に近頃は激しい疲れ目も加わって、業務の画質調整時に機器が悪いのか、はたまた自分の目のせいなのか判断に迷うことがしばしば。趣味はものすごく偏った料理。

「3Dメガネなしでいろいろな視点から見られる立体表示技術を用いて、実寸大の自動車が収まる大きな画面サイズを持つ投射式ディスプレイの試作に成功しました。」



● 自然な 3D の実現

私は立体映像技術に関する仕事に取り組んでいます。研究開発の究極の目標は、映像通信技術によって通信相手先の状況がまるで自分の目前にあるかのように感じられて、これまでになくスムーズにコミュニケーションのできる環境を作り出すことです。その際、相手側に提示する映像の大きさ・精細さ・そして立体感の3つが特に大切だと考えています。

みなさんの多くは 3D テレビといえば専用のメガネをかけて楽しむもの、とあっていらっしゃることでしょう。映画館やテーマパークの 3D 映像でもやはり同じではないでしょうか。3D メガネを必要とする方式では右目と左目のそれぞれに対応する別の映像を見せることによって、立体的に見られるようになっていきます。このとき表示されるのは右目用と左目用の 2 種類の映像のみで、どの場所から誰が見ても同じ映像しか見られません。またテレビを見るのにわざわざ 3D メガネをかけなければいけないというのは多くの人々が負担に感じるのではないのでしょうか。たとえば街頭で通りがかりの人に立体的な広告を見せたいといった用途に 3D メガネが必要とあってはそもそも無理があります。

● 三次元的に見る

私たちの脳は外界を立体的に把握するためにいろいろな手がかりを利用しています。なかでも両目に映る見え方の違い(左右の目は離れたところに付いていますから物の見え方もわずかに違い、これを両眼視差といいます。)やどこから見るかによって物の見え方も変化する(運動視差と呼びます。)現象は、立体感つまり奥行きを知るためにとりわけ重要な情報として脳内で扱われることがわかっています。3D メガネ方式のように左目用と右目用の 2 種類の静止画像を見ている人は

両眼視差を手がかりとして奥行きを感じることはできますが、運動視差は生じないため手がかりとしては用いられることはありません。

3D メガネなしでより自然な立体像を見ることができるようになれば、初めに書いたようによりよいコミュニケーション環境の実現に大いに役立つと考えて、私たちは超多視点(色々な角度から映像が見られる)かつ裸眼に対応した大型画面の立体映像表示技術の開発に取り組ましました。その成果として完成させた試作ディスプレイは 3D メガネなしで、画面の大きさは 200 型(対角の長さが 200 インチ = 約 5 m)、表示される映像としてフルハイビジョンの情報量をもち、両眼視差に加えて滑らかな運動視差をも示せるために奥行きをより自然に感じることができる、といった数々の特色を備えています。既に販売されている裸眼(3D メガネのいらぬ) 3D 対応薄型テレビではいずれの特徴もまだ実現されていません。

● 超臨場感の要素

フルハイビジョン画質、すなわち高精細な映像が見えるということは臨場感を高める 1 つの要素になるというのはあえてここで説明をしなくても、みなさんが地上波デジタル放送対応テレビに買い替えられたときのことを思い出していただければ納得していただけるものと思います。

次に画面の大きさですが、画面サイズが大きということは単に迫力があるというだけでなく、自動車や人などいろいろな対象の全体像を実寸大で見ることができることを意味しています。つまり本物と同じ大きさで見ることができるため、より高いレベルの現実感を感じるようになります。50 型前後が主流の薄型テレビではこうはいきません。さらには多くの人々が一度に見ることもできますからこの点でも大画面というのは欠かせない特徴です。

さらに運動視差があることで、横に移動しながら見ると隠れていた面が見えるなどより自然な立体感を堪能できます。例としてショールームで本物の車を見ているとしましょう。左右に動いて車体を見てみるとヘッドライトが見えたり、室内インテリアで今まで隠れていた部分がどんな造りになっているのかわかるでしょう。ボディ部分に現れる光沢の出かたで表面の形を知ることができます。これらはみな運動視差によってもたらされる情報なのですが、私たちの表示技術でも近い感覚を味わえます。図1は実際に試作したディスプレイの画面を撮ったものです。車を見てください。正面から見ると開いているドアに隠れて見えないステアリングハンドルが右の端からは見えています。

● 超多視点で魅せる立体像

それではどのように立体映像を実現しているのか、ごく簡単に説明してみましょう。試作したディスプレイ装置は大きく分けて2つの要素(図2)、特殊なスクリーン光学系と、私たちがプロジェ

クタレイと呼んでいる投射光学系を組み合わせたものです。以下、図3を使って説明します。スクリーンは透過型で横方向は光を拡げず縦方向のみ光を散乱させる特殊な拡散面と、各視点を作るための集光レンズの2層から成っています。プロジェクタレイはプロジェクタユニットを多数並べたもので、全てのプロジェクタユニットから投射された光はスクリーンの拡散面で一度像を結びます。このとき全投射画像において対応する画素同士がぴたりと重なるように各プロジェクタユニットの投射レンズを調整しています。投射光が前述のスクリーンを通過した後は横方向にはほとんど拡がらず上下にのみ大きく拡がり、上から見たときにある場所に集まるように進みます。光が集まってくる場所は視点、すなわち映像が見られる地点となります。視点は1つのプロジェクタユニットに対して1ヶ所作られ、プロジェクタユニットを横に少しずつずらして並べているため同様に視点も横に並びます。隣り合う視点の間隔は約2cm(日本人の成人の両眼間隔のおよそ3分の1)と非常に密になっています。各視点にはその位置に応じて少しずつ違った



図1 実際に見える画面の例(見る場所を変えれば異なる角度の映像が見える)

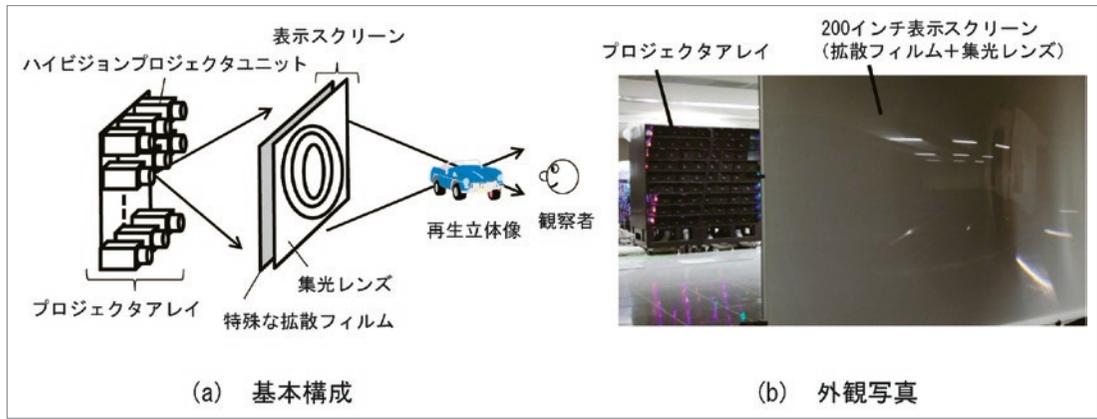


図2 開発した200インチ裸眼立体ディスプレイの構成

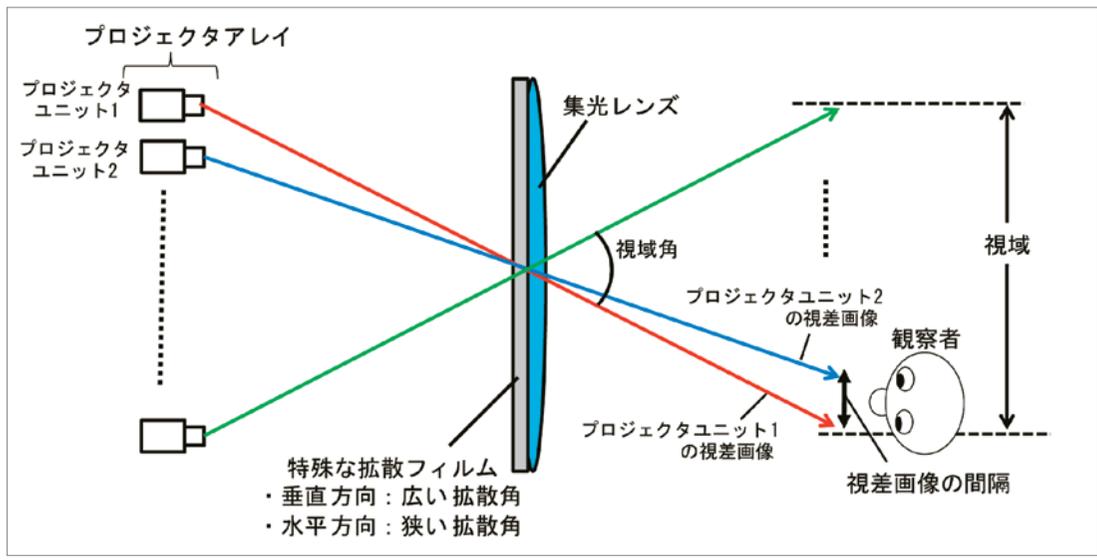


図3 立体表示原理(上から見た図)

角度の映像を屈けているため、(一定の距離離れた)両眼で見たときそれぞれの目は異なる角度の映像を受け取ることになり、両眼視差により立体像を見ることができます。さらに頭を左か右に少し動かすだけで両眼には別の角度の映像がそれぞれ入ってきますから、見ている人は同時に運動視差による奥行きをも感じることができます。これにより両眼視差のみの場合に比べて自然に立体像を見ることができるのです。

● 今後の展開

すでに始めているのですが実写、すなわちカメラを使って超多視点の映像を撮影し、ネットワークを通じて伝送できるよう最適な画像圧縮方式の開発を急いでいます。さらには実験室内から出て実際に公共の往来に試作ディスプレイ装置を長期間設置して大勢の方々に見ていただきながら実証的に評価していくことを計画しています。

電子ホログラフィ

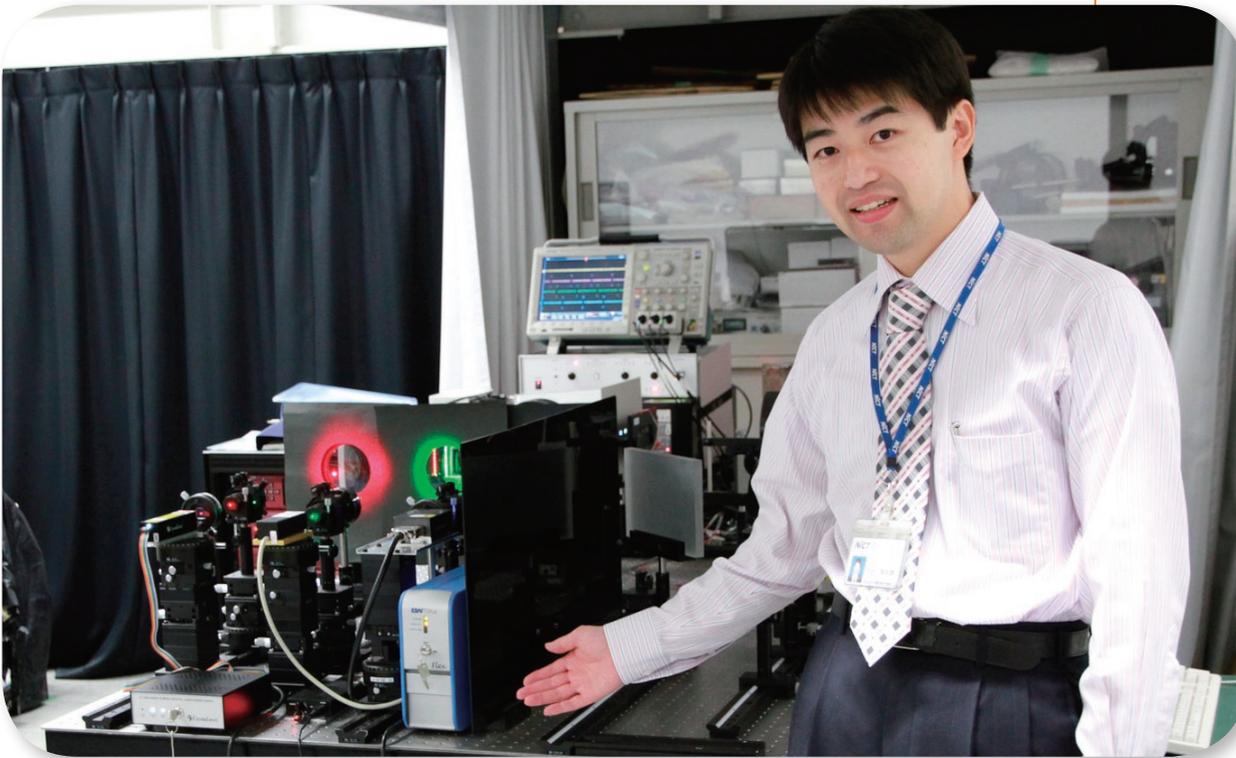
—近未来の表示技術—

大井 隆太郎 (おおい りゅうたろう)

ユニバーサルコミュニケーション研究所
超臨場感映像研究室 主任研究員

大学院博士課程修了後、放送局の研究所を経て、2006年からNICTにて電子ホログラフィの研究に従事。専門は3次元映像技術、イメージセンサ技術。趣味は水泳と飲み会。

「ホログラフィは3次元映像の記録・再生方法として最も理想的です。写真技術(静止画)として発達したこの技術を、撮影から表示まで全て電子化するのが電子ホログラフィの目標です。」



● 「電子ホログラフィ」とは？

読者の方は、ホログラムという言葉を知ったことがあると思います。例えばSF映画の「スター・ウォーズ」や「トータル・リコール」など、近未来の世界を表現する時に空中に立体で人物などを表示するシーンが出てきますが、ホログラムはこうした映画などに出てくるような3次元映像を記録したデータや媒体のことをいいます。

ホログラフィは、正確にはこの3次元映像を記録した媒体であるホログラムの製造技術のことです。ホログラフィには、ホログラムを実写する撮影技術、記録技術、および再生技術が含まれます。私たちは実写の立体映像を電子的にホログラム化し、伝送、記録、再生するまでの全てについて研究をしており、これを「電子ホログラフィ」と呼んでいます。

つまり、これまでの静止画のホログラフィ技術を動画に、しかも電子的に実現するものです。この技術により究極の3次元映像によるコミュニケーションの実現を目指しています。

● 通常の3Dテレビとホログラフィの違い

最近是一般家庭にも3Dテレビが普及しています。家電量販店では多数の3Dテレビを見ることが出来ます。市販の3Dテレビは、左目用、右

目用の少しだけ異なる左右2つの視点、つまり視差を与えた状態で被写体を同時に撮影し、その2つの映像素材を人の左右両眼にそれぞれ個別に見せることで、人間の脳内で立体を知覚させています。つまり人間の脳が持つ特性を利用した原理になっています。この方法は二眼立体と呼ばれ、映画のように画面の正面からじっと動かずに映像を見る場合には優れた方式で広く普及していますが、片目をつぶって見てみると立体的に見ることはできないという欠点があります。

また、画面に対して左側から見ている観客と、右側から見ている観客とでは、物の見え方が本来は異なるはずですが、もとの情報が左右の2視点分しかないため、どの席に座っても正面から画面を見ているときの立体映像しか見ることはできません。

立体視を実現する方法にはいくつかの種類があります(表1)。通常の3Dテレビの原理である二眼立体は、人間の視覚機能に頼った方式といえます。多眼立体は、レンチキュラーレンズ*1などを用いることで、見る位置が変化した場合にそれに応じた映像を見せることが可能です。ホログラフィでは表1の要素全てを記録再生できるため人間の目にとっては実物を見ているときと完全に同じ状態となり、ホログラフィは人の視覚機能に頼らない理想的な立体表示法といえるのです。さらに、例えば一眼レフカメラなどでホログ

表1 3次元映像記録・再生方法の比較

	両眼視差	運動視差	ピント調節(浅い)	ピント調節(深い)
二眼立体 (眼鏡式3Dテレビなど)	○			
多眼立体 (200インチディスプレイや iVisiOnなど)	○	○		
光線再生	○	○	○	
ホログラフィ	○	○	○	○

ラフィ像を撮影すれば、現実空間と同様に正確なピント合わせも可能です。

● 回折と干渉によってシャープな再生像を得る

電子ホログラフィの基本原理について説明する前に、まず2次元のテレビから考えてみましょう。通常のカラーテレビは被写体に反射した光の強さ(輝度)と色(波長)の情報を記録して、再生することができます。これに対してホログラフィでは、これら情報に加えてさらに、ホログラム面を通過する光の方向、つまり光の位相についても記録再生することで、立体映像を表示できます。

図1はこれを模式的に示したものです。撮影時は物体から反射した光である物体光 O (位相と振幅を持つ、ベクトル量)と既知の参照光 R の干渉縞 I (ホログラム、スカラー量)を記録します。再生時はこのホログラムに参照光 R と同じ光をあてて、物体光 O と位相・振幅が等しい光をふたたび発生させます。

二眼立体とホログラフィの間にある技術として、光線再生という技術があり、そこでは多数のレンズを使うことで光の方向をコントロールしています。レンズで位相を変調するので、どうしてもレンズの大きさ程度のポケ感が生じることや、レンズアレイから離れた位置での結像特性が急速に劣化する欠点があります。これに対しホログ

ラフィでは光の回折・干渉を用いて光の方向をコントロールするため、デメリットとしては光の波長程度の微小な画素を持つ特別な素子が必要になる事が挙げられますが、ホログラム面から遠く離れた位置でも非常にシャープな結像が得られるメリットがあります。

私たちの日常生活では光を光線として考えますが、光には波としての性質もあることはご存知かと思います。光の波長程度に細かい格子パターン(縞)が光が通過すると、光の一部は進行方向を変えられます。縞が粗いと少なく曲げられ、縞が細かいほどより大きく曲げられます。ホログラム全体に渡ってこのような粗い縞や細かい縞を予め書き込むことで、何も無い空中に立体像を再現できます。この縞は図1のように光学的に作ることもでき、またコンピュータ内で光の伝搬をシミュレートして、計算により作り出すこともできます。NICTでは実写のホログラム生成を行っており、例えば距離情報を取得できるカメラで撮られた3次元情報を、コンピュータ計算でホログラム化し、開発したホログラム表示装置で3次元像を立体再生するというシステムの研究をしています。

● 実験室レベルの巨大な表示システムは必要なの？

現状では電子ホログラフィ再生装置*2が大きいいため、サイズをコンパクトにすることがまず課題に

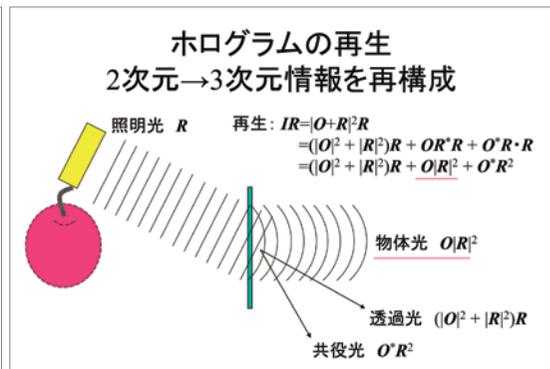
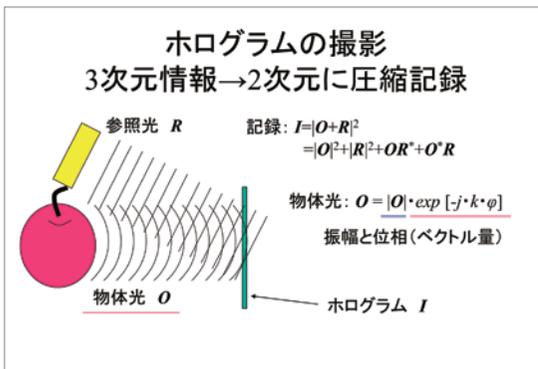


図1 ホログラムの記録と再生



(a) 近い: 自動車のタイヤの付近に焦点



(b) 遠い: 植木の葉の付近に焦点

写真1 電子ホログラフィ再生装置で再生した実写ホログラム像

再生像(a)、(b)はともに1枚の電子ホログラフィから再構成された立体像を再度カメラで撮影して得た画像。輝度、色に加えて、距離も正確に再現。

挙げられます。2010年度に報道発表した電子ホログラフィの表示装置では世界初の技術として視域角15度で対角4cmサイズのカラー動画表示を実現しました。その中で使用している光変調素子は、3,300万画素のものが合計3枚で構成されており、カラー動画表示を実現するためには、素子の周辺に必要な駆動装置やメモリなども大きくなります。また、装置が大きい理由のひとつとして、ホログラムから原理的に発生する妨害光、つまりノイズを除去するために光学的なフィルタ処理を行っている部分のサイズが大きいことが挙げられますが、今後の変調素子技術が進歩するのに従ってコンパクトなものになる見込みが十分にあります。現状で本当に必要なホログラム自体の部分は20cm角程度です。写真1は我々の電子ホログラフィ装置で再生された映像の例です。実際には動画像(ムービー)として再生されますが、像の大きさは高さ約2cm程度で、今後、表示サイズを大きくすることも課題です。

● おわりに

今回の電子ホログラフィの成果は、今ただちに実用化されるというものではなく、研究開発の初期段階にあり、まだまだ研究が必要です。テレビ

がモノクロからカラーへ変わり、カラーからハイビジョンに進歩してきたように、その先の立体映像への進化も必ずやってくるでしょう。現在人類が知り得る立体表示の方法の中で、究極といえる電子ホログラフィによって社会生活をより豊かなものにすることができると信じています。

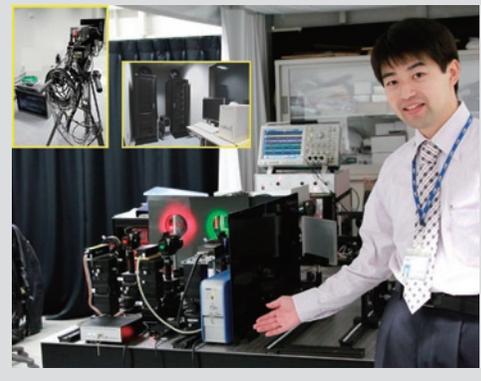
用語解説

*1 レンチキュラーレンズ

かまぼこ形の微細な半円筒レンズを多数並べて作ったシート状のレンズのことで、液晶画面や印刷物の表面に貼り付けて使う。半円筒レンズの働きで光の進行方向を変え、見る位置に応じて液晶画面や印刷物の異なる画素を観察できるようにするためのレンズのこと。

*2 電子ホログラフィ再生装置

上部黄色の枠内は距離カメラと演算用PCクラスター



声できくと、声で答える。 観光案内アプリ AssisTra

—自然な音声で簡単に情報を取得できるシステムの実現を目指して—

「音声コミュニケーション研究室の技術を用いたスマートフォン用の音声対話システム AssisTra を、Siri しゃべってコンシェルに先駆けて、2011年6月に公開しました。」



翠輝久 (みす てるひさ)

ユニバーサルコミュニケーション研究所
音声コミュニケーション研究室 研究員

人と人、人とコンピュータのコミュニケーションに関心を持ち、音声対話の研究をしています。

水上悦雄 (みずかみ えつお)

ユニバーサルコミュニケーション研究所
音声コミュニケーション研究室 主任研究員

コミュニケーションにおける相互調整の性質に関心があり、対話の評価研究をしています。趣味は映画(特にSF)観賞、公園探検と称する娘たちとの散歩、昆虫探索、車でのぶらり遠出旅行。

はじめに

NICTユニバーサルコミュニケーション研究所音声コミュニケーション研究室では、わずらわしい操作を覚えなくとも、その人にとって自然なコミュニケーションの手法で、容易に情報システムを利用できる社会の実現を目指して、研究を進めています。中でも私たちは、人に話しかけるような、自然な音声による要求を受け付け、その意図を理解・推測することによって、適切な情報を提示する、高精度対話処理技術を研究しています。これまでの研究成果の実証実験および実データ収集を目的として、観光案内 iPhone 用アプリ「AssisTra」を2011年6月にリリースしました。本稿では、AssisTraの中心機能である、『はんなのガイド 京都編』で利用されている音声対話処理技術について解説します。

『はんなのガイド 京都編』とは？

ユーザの自然な音声による要求を受け付け、音声と画面で、その要求に答える「音声対話システム」です。図1の例のような音声対話をすることができ、ユーザは京都の観光スポットやレストランなど観光に役立つ様々な情報を調べることができます。

音声対話処理技術

一般に音声対話システムは、図2のような構成をしており、大きく分けて、音声認識、音声言語理解、対話制御、応答文生成、音声合成の5つのモジュール(要素技術)で構成されます。『はんなのガイド 京都編』に用いられているこれらのモジュールは、すべて当研究室で開発したものです。以下では、これらのモジュールについて概説します。



図1 『はんなのガイド 京都編』対話例

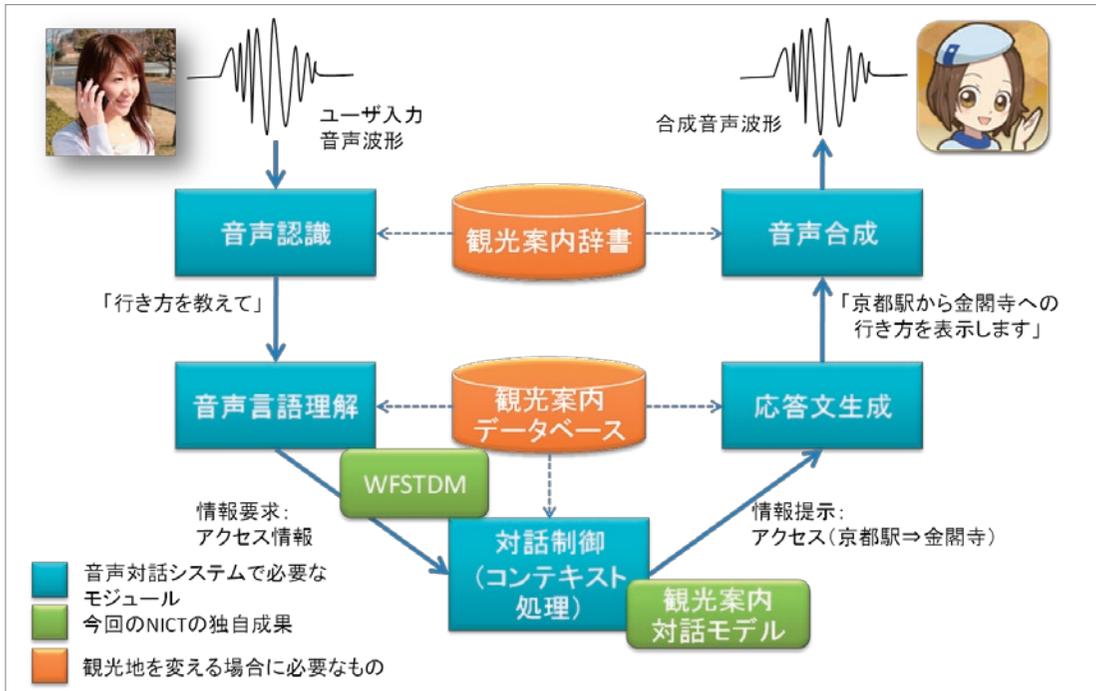


図2 音声対話システム構成図

・音声認識・音声合成・応答文生成

音声認識・音声合成は、隠れマルコフモデル*に基づく統計的手法を利用しており、大量の音声データを学習に用いることで、自然で連続的な音声発話文を認識し、人の発話音声に近い合成音を作成することができます。後述の大量の観光案内対話データを利用して、観光案内用に特化したモデルを作成することにより、高い音声認識率と、ユーザに話しかけるような自然な合成音声を実現しています。さらに、応答文生成で利用するテキストとして、プロのガイドの発話内容をもとに、桜、紅葉など様々な観点からの観光スポットの説明文を整備しました。

・音声言語理解

人間の自然発話には、ユーザや状況によって様々な言い回しが存在します。たとえば、「観光スポットへのバスを利用したアクセス方法」が知りたいと考えている場合を考えると、図3の例をはじめとして、多種多様な言い回しが

存在します。このような発話の意図を解釈することは人にとっては難しいことではありませんが、コンピュータがこれらの発話を理解するためには、これらの表現を同一のシンボル(コンピュータが処理可能な言葉)に変換する必要があります。これが音声言語理解の役割です。

この機能を実現するためには、ユーザが実際に使用する表現を収集するとともに、高精度な音声言語理解アルゴリズムを研究・開発することが重要になります。会話の中で実際に利用される言い回しを収集するために、私たちはプロの観光ガイドと旅行者の模擬会話を150時間300対話収録しました。これは、現在収集されている単一状況での音声対話データとしては世界的にも大規模なものです。さらに、プロトタイプ音声対話システムを構築して、被験者実験を行い、実際のシステム利用を想定した状況での発話表現を収集しました。これらのデータをもとに、私たちの研究室で独自に開発した音声言語理解・対話制御フレームワークである『重み付き有限状

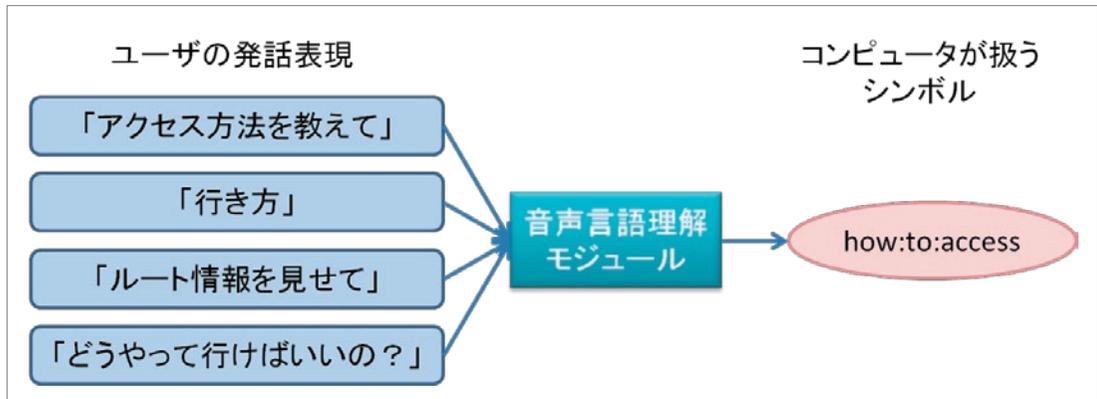


図3 言語音声理解の例

態トランスデューサ対話制御機構(WFSTDM: Weighted Finite-State Transducer-based Dialog Manager)』を用いて、WFST 表現による音声言語理解モデルを作成することで、高速かつ高精度な音声言語理解を実現しています。

・対話制御

まったく同じ発話が入力された場合でも、状況や発話履歴に応じて発話に含まれるユーザの意図が異なる場合があります。たとえば、「アクセス方法を教えて」という入力があった場合には、直前の対話内容に基づいて「どこから、どこまで、どのような交通手段で」などの情報を補完する必要があります。これらの発話に隠れた意図を適切に補って応答内容を決めることが対話処理の役割です。

このような対話履歴処理は、対話システムが利用される状況や、ユーザがシステムを使う目的に対する依存性が高いものです。そこで、ユーザの実際の利用状況に近い、前述の大規模対話データをもとに観光対話用の履歴処理モデルを作成し、対話履歴を適切に処理しています。

● おわりに

今回、アプリを公開し、収集されたログデータを分析していますが、システムの応答の精度はま

だ十分ではありません。人間の発話や意図の種類・言い回しのバリエーションが 150 時間程度の学習データではカバーしきれないほど多様で複雑なものであり、コンピュータが人の意図を正確に理解するためには、より大きな対話データを収集するとともに、音声言語理解や対話履歴処理の精度の改善が必要であることが分かりました。今後はシステム運用により収集した発話データを追加して各モジュールのモデルを再構築するとともに、より柔軟に発話を理解し対話を制御するアルゴリズムの研究を進めていきます。また、システムの利用の拡大を目指して、訪日観光支援に利用できるように『はんなのガイド 京都編』の英語版を 2012 年 3 月に無料公開しました。さらに、チケット予約や、コールセンター業務など、実世界で必要とされている様々なタスクを扱う音声対話システムを構築し、対話処理技術の実用性を証明していきたいと考えています。

用語解説

* 隠れマルコフモデル
 観測される記号列(音声認識の場合、音声の特徴量)が、直前の m 個の記号から決定されるマルコフ過程であると仮定し、それを出力するような状態遷移系列が非決定的である(隠れている)とする確率モデルです。音声認識の場合、状態遷移確率などのパラメータが、大量のデータから学習され、最も高い確率で記号列を出力するような単語列や音素列が認識結果となります。

身近になった多言語自動翻訳

— 半世紀以上の研究開発を経て実用化されつつある技術 —

内山 将夫 (うちやま まさお)

ユニバーサルコミュニケーション研究所
多言語翻訳研究室 主任研究員

自動翻訳の研究を10年くらいしています。自動翻訳の性能をあげるにより、世の中が住みやすくなることに貢献したいと思っています。家族にわかる研究をするように努めています。

「自動翻訳は、長年の研究開発を経て実用化されつつあります。NICTの自動翻訳エンジンは、対訳データから自動構築可能です。この研究成果を社会に還元したいと思っています。」



● 身近になった多言語自動翻訳

自動翻訳の研究は、1940年代から始まりました。日本でも、1980年代に産学官で盛んに研究開発されて、たくさんの商用自動翻訳システムが開発されました。そして、現在では、インターネットのポータルサイトなどで、自動翻訳が無料で提供されていますし、商用の自動翻訳システムの提供も盛んです。

自動翻訳システムの種類には、基本的には、人手で記述した規則に基づいたシステムと、大量の対訳テキストから自動的に翻訳規則を学習するコーパスベースの自動翻訳システムの2種類があります。コーパスというのは、大量のテキストからなるデータベースのことです。

これまでは、一般的に利用されている自動翻訳システムは、人手で記述した規則に基づくシ

テムでしたが、最近では、コーパスベースの自動翻訳システムの性能も向上しています。NICTが研究している自動翻訳システムは、コーパスベースの自動翻訳システムです。その研究開発の基盤には、大きく分けると、ユーザ、言語資源、アルゴリズムの3点があります。

● ユーザ

自動翻訳システムの目的は、ユーザの役に立つことです。そのため、ユーザは、自動翻訳システムの研究において、もっとも尊重する必要があります。

NICTでは、旅行会話専用の自動音声翻訳システムとしてVoiceTraを開発しています(p.140-143参照)。また、eコマース用の自動翻訳エンジンとして、日本最大級のアプリサイト



図1 みんなの翻訳 (http://trans-aid.jp)

の、韓国サイトについて、日韓自動翻訳により商品説明文を韓国語に訳すサービスを提供しています。この自動翻訳エンジンのための対訳データを作成するために、株式会社バオバブとの共同研究により「留学生ネットワーク@みんなの翻訳」(<https://en.ecom.trans-aid.jp/>)を開発し、留学生のアルバイトにより効率的に対訳データを作成しました。また、ボランティアによる人手翻訳を支援するために東京大学図書館情報学研究室と共同で「みんなの翻訳」(<http://trans-aid.jp>)を運営しています(図1)。

このように、NICTでは、最新の研究成果を一般に利用していただくことにより、研究成果を社会に還元すると同時に、そのフィードバックを研究開発に役立てています。

言語資源

最も重要な言語資源は、対訳テキストです。統計的自動翻訳では、分野を限定した場合で十数万文、分野を限定しない場合には1,000万文以

上の対訳文が翻訳エンジンの訓練に必要です。図2は、eコマース分野における、訓練に利用した対訳文数と翻訳精度の関係を示しています。なお、AとBは商用の翻訳エンジンですが、これらは対訳文での訓練はしていないため、一定の精度です。

NICTでは、異なる言語の文書から自動的に文と文の対訳を作成する技術を開発しています。そして、この技術を利用することにより、日本と米国に同時出願された特許文書から1,000万文以上の大規模な日英対訳コーパスを自動作成しました。今後は、このコーパスを利用して、日英の特許翻訳や類似文検索などのサービスを開発する予定です。

また、NICTでは、新聞記事から25万文規模の日英対訳コーパスも作成しており、この対訳コーパスは、ライセンス契約により第三者も利用可能です。このコーパスは英辞郎に採用予定ですので、たとえば、「英辞郎 on the Web」でNICTが提供した対訳文が検索可能となります。

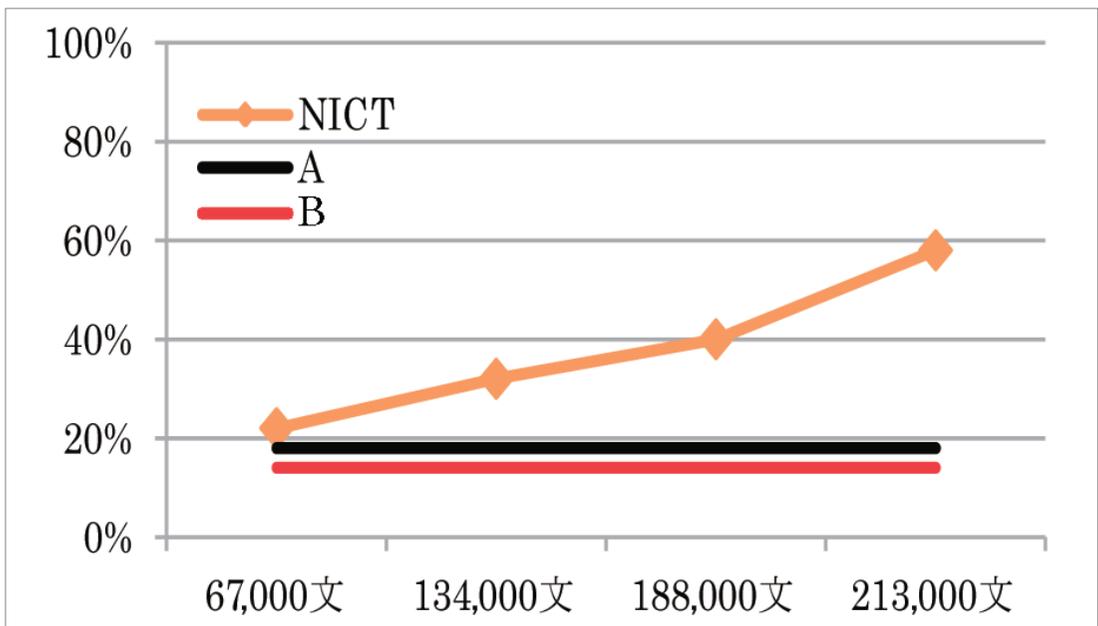


図2 日英翻訳の性能改善

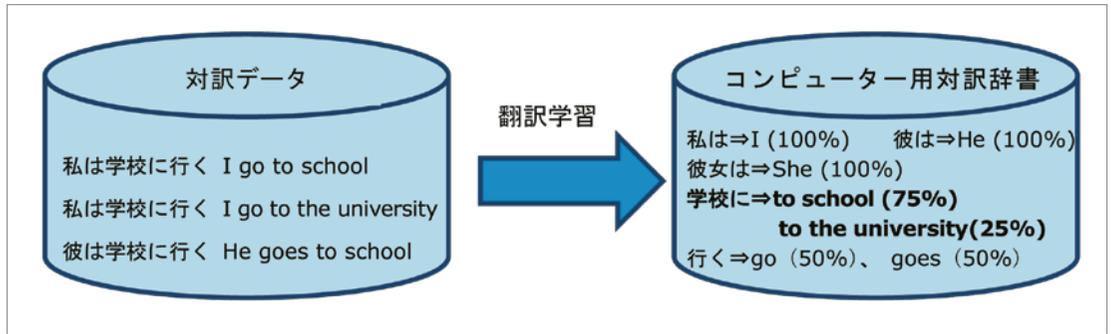


図3 コンピューター用対訳辞書の作成

● アルゴリズム

コーパスベースの自動翻訳は、1980年代に、長尾真氏により提案されました。その後、1990年代に、IBMにより、対訳コーパスから自動的に対訳辞書や翻訳規則を推定するアルゴリズムが開発されました。2000年代には、自動翻訳を(不完全ながらも)自動的に評価する方法が開発されました。現在では、英日や日英などの言語構造が離れた言語対の翻訳も、ある程度の精度で自動翻訳ができるようになりました。

コーパスベースの自動翻訳では、対訳コーパスから自動的に翻訳辞書を構築します(図3)。この翻訳辞書は、たとえば、「私」と「I」が対訳関係にあることを記述しています。さらに、通常の辞書と異なる点として、どのくらいの確率でこれらが対応関係にあるかも記述しています。更に、単語だけではなく、フレーズの対訳関係も大量に格納しています。

この辞書を使って日本語を英語に翻訳する方法の概略は、まず、日本語文をフレーズに分割します。そして、そのフレーズを英語に翻訳して、最後に、翻訳したフレーズを並べ替えて英語の語順にします。ここで、もちろん、個々の日本語のフレーズには複数の英語のフレーズが対応しますし、日本語文をフレーズに分割する方法も多量にあります。したがって、自動翻訳が出力可能な英語文の数は無数にあります。この無数の翻訳候

補の中から、前述の対訳関係の確率等を利用して、最適な英文を選択します。

もっとも、実際に行われている方法は、もっと複雑です。たとえば、英日自動翻訳の場合には、英語と日本語の構造差が大きいので、あらかじめ入力文の英語を構造解析して、日本語の語順に近くなるように英語の語順を変更してから、上述の方法で翻訳したりします。

自動翻訳の研究においては、これまで、ほぼ10年に1回の割合で、ブレークスルーとなる研究が起きていますので、ここ数年で、次のブレークスルーが起きるのではないかと思います。私たちは、そのブレークスルーをNICTから起こすように研究しています。

fVisiOn: 何も無いテーブルの上に浮かぶ3D映像の作り方

—メガネなしで360度から見えるテーブル型3Dディスプレイの研究—

「テーブルを囲んだコミュニケーションのような自然な情報伝達環境を目指し、メガネなしで観察できる3D映像をテーブル上のそこにあるかのように提示する技術を紹介します。」

吉田 俊介 (よしだ しゅんすけ)

ユニバーサルコミュニケーション研究所
超臨場感映像研究室 主任研究員

大学院修了後、通信・放送機構(TAO)研究員、国際電気通信基礎技術研究所(ATR)研究員を経て、2006年よりNICT主任研究員。博士(学術)。人とコンピュータをつなぐ境界面(インタフェース)の研究者。「未来を作っています」と言えるハカセになるべく精進中。

● テーブルトップを介したコミュニケーション

コミュニケーションにはいろいろな形がありますが、この研究が対象としているのはテーブルの周りに集う人々のコミュニケーションです。テーブルの上(テーブルトップ)は様々な作業をみんなで共同して進めるのに適した空間です。書類や模型を並べる場所として使うことができますし、それらをみんなで共有して書き込んだり修正したりしながら議論を進めることができます。これをコンピュータで支援し、テーブルトップに表示されたデジタルな書類を扱えたり、3D映像の模型を修正できたりすれば、その場に集まった人同士だけでなく、データのやりとりで遠隔地間でもテーブルトップを介したコミュニケーションができるようになります。



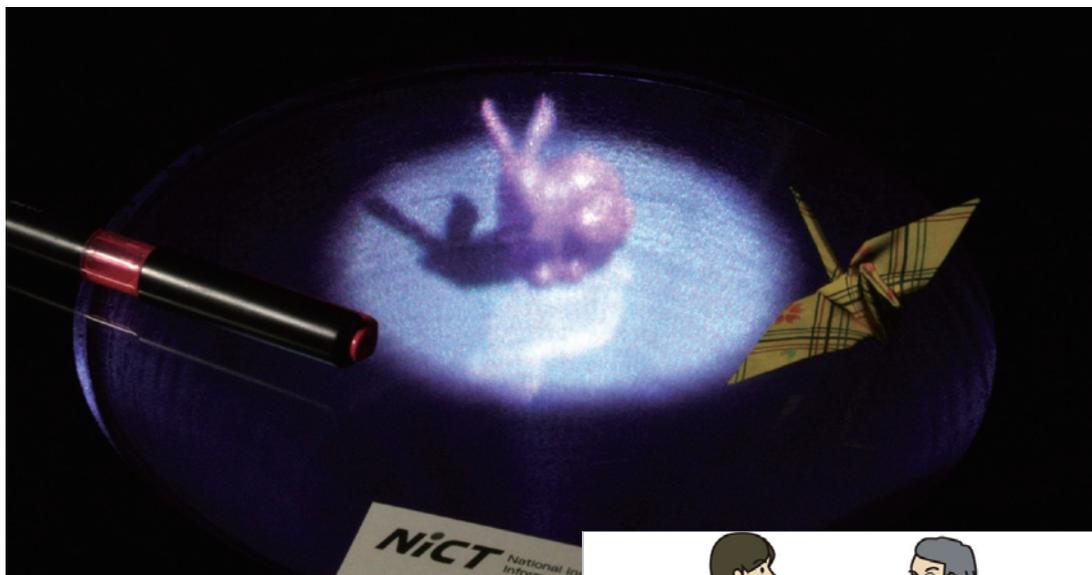


図1 テーブル型メガネなし3Dディスプレイ [fVisiOn]
 上 試作したfVisiOnによる3D映像。中央にウサギの3D映像と、
 周りに折り鶴やペンなどが置かれたテーブルトップ
 右下 テーブルトップを介したコミュニケーションの例



● テーブルトップに求められる 3D 映像

「fVisiOn (エフ・ビジョン)」と名付けたテーブル型 3D ディスプレイの研究は、テーブルトップにて 3D 映像をみんなで自然に共有するためにはどのようにすればよいのかという新しい着想より生まれました。

テーブルに置かれた模型がそうであるように、テーブルトップに表示された 3D 映像は、いろいろな方向から観察すると違った見え方をしなければいけません。しかし一般的な 3D ディスプレイの技術は、テレビのように正面側から 3D 映像を見ることが想定されていて、得られる立体感は奥行きのみで、3D 映像の横や後ろが見えるものではありませんでした。全周 360° から観察可能な 3D ディスプレイ技術も提案されていますが、それらはテーブルに置かれたガラスケースの中に表示するような仕組みであり、その表示装置がテーブルトップでの自由な作業を邪魔してしまうことが問題でした。また、より

自然なコミュニケーションを達成するためには、特別なメガネをかけることなく、何人でも同時に 3D 映像を観察できることが望ましいと言えます。

fVisiOn で提案する方式では、何も無い平らなテーブルの上に高さのある 3D 映像を浮かび上がらせて再生することができます。テーブルの周囲にいる人々は、特別なメガネを使うことなく、何人でも同時に周囲 360° からそれぞれの視点に応じた 3D 映像を観察できます。テーブルトップには作業の邪魔となる表示装置が一切ないので、従来と同じように 3D 映像の隣で書類を交わしたり、模型をそばに置いたりすることもできます(図 1)。

fVisiOn を実現する技術

現実世界の物体は、両目が左右に離れているので、それぞれの眼には少しずつ違う見え方で写ります。この見え方の差が立体を感じる要因のひとつです。fVisiOnでは、円状に並べた多数のプロジェクタを使って様々な方向へ向かう光線群を大量に作り出し、それらの進み方をうまく制御する光学素子を使うことによって、見る方向で見え方が変わる映像をテーブルトップに表示します。これにより、立体的な映像として両目で知覚することができます(図2)。

fVisiOnの研究では、テーブルトップに適した(作業の邪魔にならない、斜め上からの観察に対応する、特別なメガネがいらぬ、みんなで使える)新しい3D映像の再生技術の考案に加え、それを実現するための技術開発に困難が伴いました。特に再生原理を実現する光学素子の作製が

難しかったのですが、すり鉢状のアクリル円錐に糸状のレンズを巻くという工夫で、必要な光学性能を得ることに成功しました。

現在の試作機では、テーブルトップから5cmほど飛び出した3D映像を周囲から観察できます。例えば、3D映像のウサギでは、頭側から見る人と尻尾側の人では見え方が異なり、テーブルに落ちた影もウサギの模型がそこあるかのように見え方が変わります(図3、4)。静止面だけではなく動画も再生可能で、実物の模型ではできない動きのある情報提示が可能であることもfVisiOnの利点のひとつです。

今後の展望

試作機はまだ生まれただけの状態であり、3D映像の品質は今後さらに改良を加えていきます。用意できたプロジェクタの数の制限(100台強)か

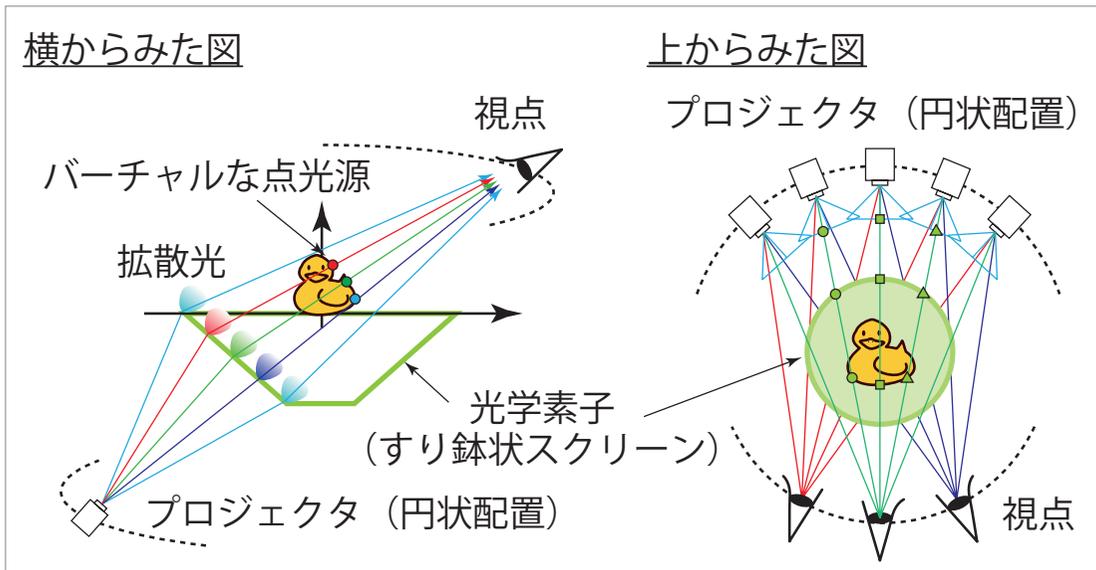


図2 fVisiOnにおける3D映像を再現する原理

- ・ 横から見た図のように、プロジェクタから投射された光線を、光学素子は垂直方向には拡散して、テーブルの周囲、斜め上方向の視点へ光を向ける。
- ・ 一方、光学素子は水平方向には光線を拡散させずにそのまま直進させる。
- ・ そのため、テーブルの周囲のある視点では、複数のプロジェクタから放たれた映像の一部(スリット状)がそれぞれ横に連なって1つの映像として観察される。
- ・ 別の視点では、それぞれ別の一部が連なった映像が見えるため、観察方向ごとに異なる映像を見せることができる。
- ・ この原理によって、観察方向に応じたそれぞれ異なる見え方が再現されるので、両目で見たときに3D映像として知覚される。

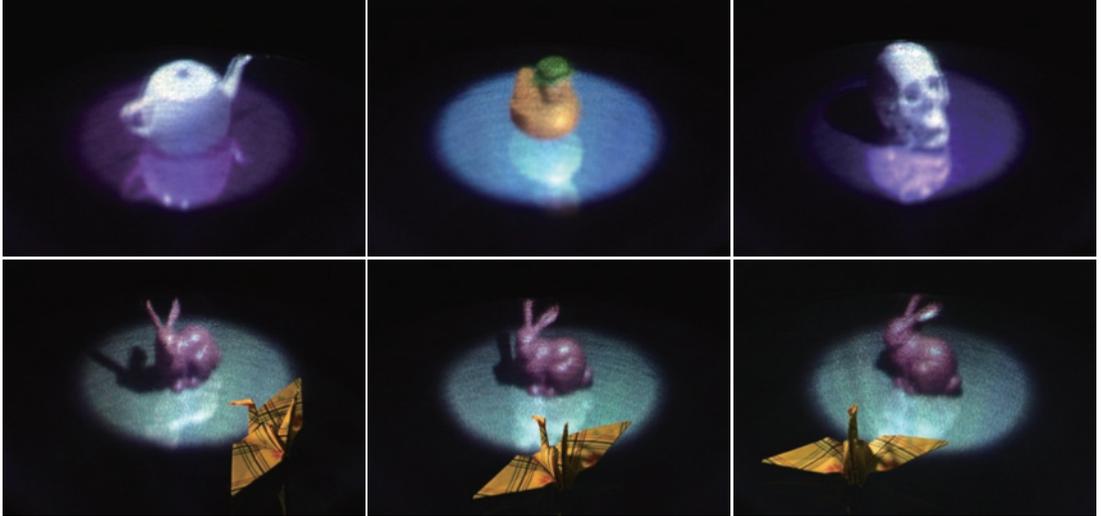


図3 fVisiOnで再生された3D映像の写真

上段: 左からティーポット、おもちゃのアヒル、頭蓋骨

下段: 3D映像のウサギと手前に置いた実物の折り鶴を異なる角度から撮影

ら観察範囲も今は 130° 程度ですが、360° からの観察へ拡張可能なことは原理的に確認できました。3D 映像の全周化は次の試作で試みたいと思います。

fVisiOn は従来からのテーブルトップ作業に親和性の高い 3D 映像技術です。これまでに述べたようなテーブルを介した議論や作業といった産業用途だけではなく、平面ではわかりにくい身体の構造を立体的に表現すること

で、お医者さんたちの手術の事前検討や患者さんとのコミュニケーションなど医療の場面でも役立つでしょう。また、提案技術は斜め上からの観察に最適化されていますので、3D 映像の地図を使った防災訓練や、交通管制などにも有効です。さらには、家族みんなで楽しめる 3D 映像のテーブルゲームや、将来的に大型化ができれば 3D 映像のサッカースタジアムといったエンタテインメントへの応用も広がります。

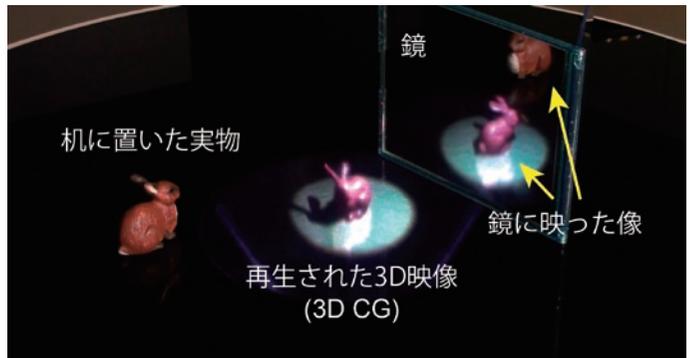


図4 机の上に鏡を置いた場合の写真

鏡に映り込んで見える像が、実物と同じように、3D映像も反対側から見た様子になっていることがわかります。

普段の生活で利用しているテーブルにさりげなく 3D 映像を加える、それが fVisiOn の目指す究極の形です。

実世界とクラウドをつなぐ ICTインフラ

「クラウド上の多種多様な情報サービスと、実世界の大規模なセンシングデータを連携させることで、実世界に有用な情報を提供するシステムの構築を可能にするICTインフラを研究開発しています。」



金京淑 (キム キョンスク)

ユニバーサルコミュニケーション研究所
情報利活用基盤研究室 研究員

2007年韓国釜山大学自然科学研究科電子計算学博士取得。2007年11月よりNICT知識創成コミュニケーション研究センター知識処理グループ有期研究員、2010年8月に同研究員を経て、2011年より現職。時空間データベース、時空間データマイニングに興味を持つ。日本データベース学会会員。

村上陽平 (むらかみ ようへい)

ユニバーサルコミュニケーション研究所
情報利活用基盤研究室 主任研究員

2006年京都大学大学院社会情報学専攻博士課程修了。博士(情報学)。電子情報通信学会サービスコンピューティング研究専門委員会委員長を務める。

● 研究の背景

近年、インターネットを介してサーバ上のプログラム処理を情報サービスとして提供するクラウドが普及しています。クラウドは、サーバのCPUやハードディスクといったハードウェアの提供から、サーバ上で動作しているプログラムやデータといったソフトウェアの提供まで多岐にわたっています。このようなクラウド環境を用いて、実世界の情報をセンサーやモバイル端末上のアプリケーションから収集して、多様な情報サービスと適切に組み合わせることで実世界に有用な情報を提供するサイバー・フィジカル・システムが注目されています。我々の研究室では、このような実世界とクラウドを繋ぐサイバー・フィジカル・システムを実現するためのICTインフラの研究開発に取り組んでいます。さらに、このICTインフラを用いて、実世界の多種多様なBig Dataを収集し、その関係性を分析して配信することで実世界の Awareness を高めるサイバー・フィジカル・データクラウドの実現を目指しています。このICTインフラは、それぞれの大規模データや情報サービスのギャップを埋め、ユーザの要求に応じて適切に連携させる基盤ソフトウェア「ZODIAC (Zero-gap Orchestration for Data Intensive Actionable Collaboration)」と、ユーザに情報サービスを効率的に届けるネットワーク「Service-Controlled Networking」から構成されます。

● 異種・異分野 Big Data の検索

図1に示すZODIACの大規模データ管理技術では、様々な研究機関が持つ幅広い分野の科学

データ、社会を反映した新聞記事データや20億ページのWebコンテンツ、実世界の環境の観測データ(降水量、風速、震度等)など近年重要性が増しているBig Data間の相関関係を発見することで、分野間のギャップを越えて、様々な出来事や現象に関連するデータを横断的に検索・統合することを可能にします。たとえば、単純な時空間(位置・時間)の相関関係を使い、台風や地震などの自然現象の観測データと、新聞記事やWebコンテンツなどの社会の観測データを横断的に繋ぐことで、自然現象が人々の日常生活にどのような影響を与えたかをより詳しく、分かりやすく伝えることができます。

● 多様な情報サービスの連携

図1に示すZODIACの情報サービス連携技術では、大規模データ管理技術によって提供される検索サービスとデータ分析サービスを組み合わせることでクラウド上の新しい情報サー



図1 ZODIACの役割

ビスを実現し、災害など刻々と変化する状況を分析するアプリケーションの構築を容易にします。情報サービスの連携では、情報サービス間で交換するデータの形式や、情報サービスの利用手順にギャップが生じます。これらのギャップは情報サービスが多様化し、環境の変化が激しくなるにつれてより顕著になります。情報サービス連携技術は、データ間の依存関係に基づいてデータ形式の変換を行い、環境の変化に基づいて情報サービスの切り替えを行うことで、これらのギャップを越えて情報サービスの連携を可能にします。たとえば、降水量や河川の水位、交通状況などの様々な環境センサーサービスと、Twitterなどのソーシャルサービス、翻訳サービスを連携させることで、災害状況やユーザーに合わせてアナウンス内容や言語、伝達方法をリアルタイムに自動的に切り替えることができます。

情報サービスを届けるネットワーク

図2に示す情報サービスによるネットワーク制御技術(Service-controlled networking)は、クラウド上の情報サービスの連携プロセスに連動してネットワークの構成を柔軟に変更します。特に、データの大規模化・情報サービスの複雑化による新しい要求に応えるために、従来のインターネットよりも高い処理能力と拡張性を備えた「新世代ネットワーク」を活かし、膨大なデータを処理したり、多種多様な情報サービスを組み合わせたネットワーク構成をプログラムで調整することができます。これにより、災害などによりネットワークが混み合ったり、遮断されたりといった既存システムで想定していない事態が発生した場合でも、ユーザーが素早くデータを収集し蓄積できるように専用の連携ネットワークをオンデマンドに構成することができるようになります。

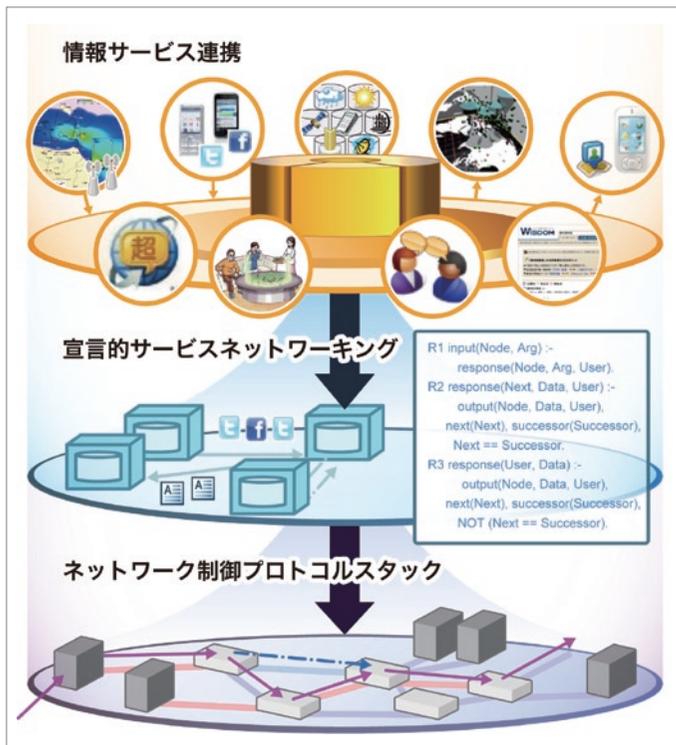


図2 情報サービスによるネットワーク制御技術

今後の展望

この ICT インフラによりクラウド上の情報サービスと Big Data の効率的な連携が可能になることで、実世界の観測データから実世界の状況を多面的に分析するアプリケーションの開発が促進されることが期待されます。たとえば、Stream Concordance(ストリームコンコダンス)は、Twitter上の意見を分析するアプリケーションです(図3)。Twitter上に流れるツイートを意見分析サービスで肯定、否定に分類し、ツイート内に出現するキーワードで整列させることで、社会の中でそのトピックがどのように語られているかをリアルタイムで把握することがで

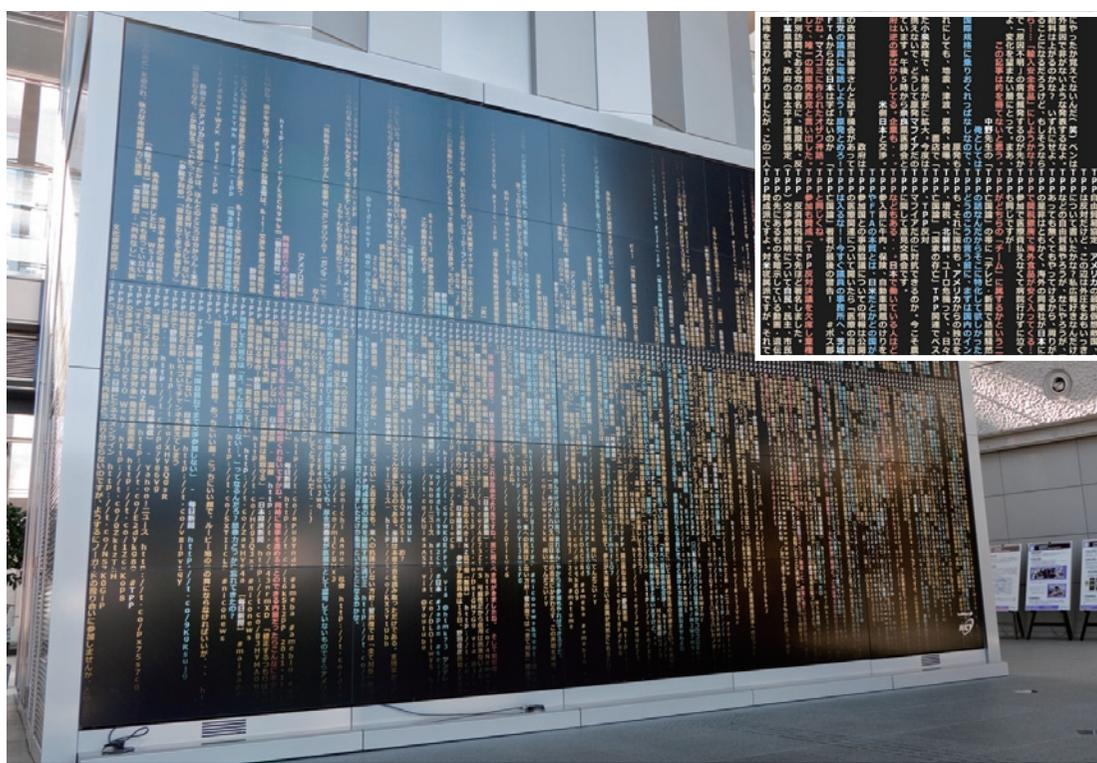


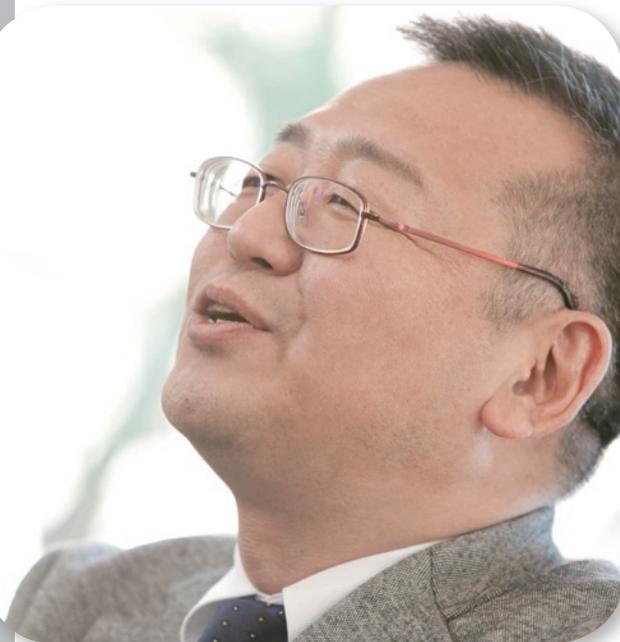
図3 Stream Concordanceを超高精細大画面に写し出した例（“TPP”を含むツイートを整列）

きます。このようなアプリケーションの開発を加速させるためにも、現在、世界中の科学データの共有を図っている国際科学会議(ICSU)による世界科学データシステムのWDS(World Data System)と連携して、利用できるデータや情報サービスの拡充を進めています。また今後はこのICTインフラを用いて「知識・言語グリッド」と呼ばれるユーザ参加型のテストベッドを公開することで、参加者によるデータや情報サービスの共有を可能にし、互いの情報サービスの利活用を促進していく予定です。

世界中の人々が母国語で外国人と対話できる多言語音声翻訳技術

—スマートフォンに話しかければ自動的に通訳するソフトウェア「VoiceTra」—

世界中の人々が母国語で外国人と対話できる多言語音声翻訳技術 隅田英一郎・柏岡秀紀



隅田 英一郎 (すみた えいいちろう)

ユニバーサルコミュニケーション研究所
多言語翻訳研究室 特別招へい研究員・室長

京大、IBM、ATR、NICTと居場所を変えつつ、もう四半世紀も自動翻訳をやっているロートルですが、今が一番楽しいですね。研究の進展が凄く速く、技術移転も好調だからです。自動翻訳は、アカデミックでもビジネスでも、かつてない熱いステージに達しました。一方、文脈処理や同時通訳等の大物課題もしっかり残っています。次の四半世紀もワクワクものです。皆さん、一緒に楽しみませんか？



「スマートフォンの33台に1台にダウンロードされ利用されたNICTの多言語音声翻訳アプリの技術と評価、今後の展開について説明します。」

柏岡 秀紀 (かしおか ひでき)

ユニバーサルコミュニケーション研究所
音声コミュニケーション研究室 室長

コンピュータと自然に会話できる世界を作ることに関われればと思います。音声翻訳、音声対話の研究をしています。大学を卒業した頃は、携帯電話は、珍しいものだったのに、今ではみんな持っているのが当たり前となり、想像以上に音声翻訳、音声対話が身近なものになりつつあります。突拍子もないことでも、すぐに当たり前になるかも、そんな研究がしたいと思っています。

言葉の壁はボーダーレス社会において大きな課題です。例えば、政府の『新成長戦略』*1では「訪日外国人を2020年までに現在の3倍の2,500万人(経済波及効果10兆円、新規雇用56万人)にする」としていますが、公共交通、宿泊施設、飲食店での外国語対応の遅れが、訪日外国人の最大の不満となっています。

このような「言葉の壁」を克服するため、NICTでは、多言語音声翻訳ソフトウェアの研究・開発を進めています。その成果として、音声翻訳ソフトウェア VoiceTra*2をスマートフォン用に公開しました。無償でダウンロードできるこのアプリケーションを使えば、例えば、図1と図2の組み合わせで示したような日本語と英語の対話がで

きます。電話をかける時のようにスマートフォンを耳元に近づけると短時間振動するので、これを合図に音声を入力すると、翻訳結果が音声で返ってきます。図1の1番目の窓はシステムが認識した(聞き取った)結果、3番目の窓が翻訳結果です。2番目の窓は、「逆翻訳」(翻訳文を元の言語に逆に翻訳する)の結果で、これを見て正しく翻訳できたかどうかを確認できます。VoiceTraは2010年8月に公開し、2012年3月時点で累計60万件を超えるダウンロード数を記録しています。日本人の200人に1人が利用者であり、日本のスマートフォンの33台に1台にダウンロードされた計算になり、音声翻訳技術を多数の方に知っていただくことができました。さらに、後述するように、民間事業者と事業化が始まり、NICT技術が社会に還元された代表例の1つになっています。

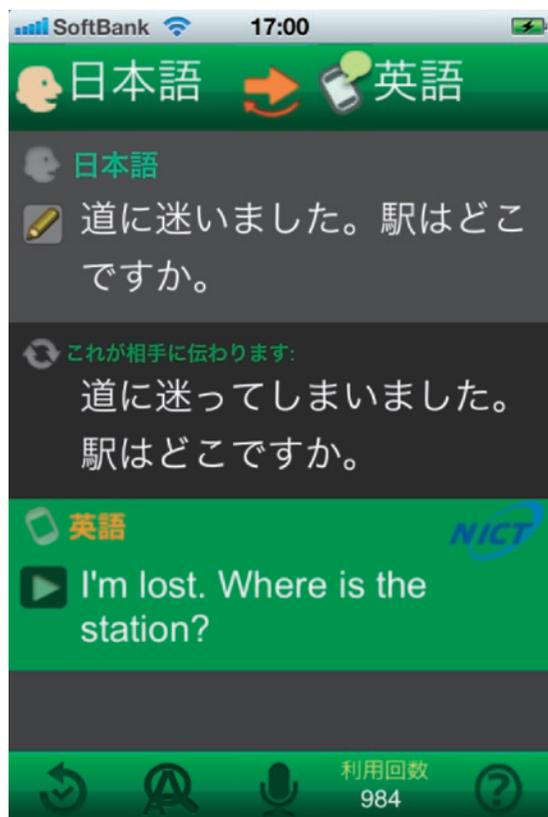


図1 例(日英翻訳)

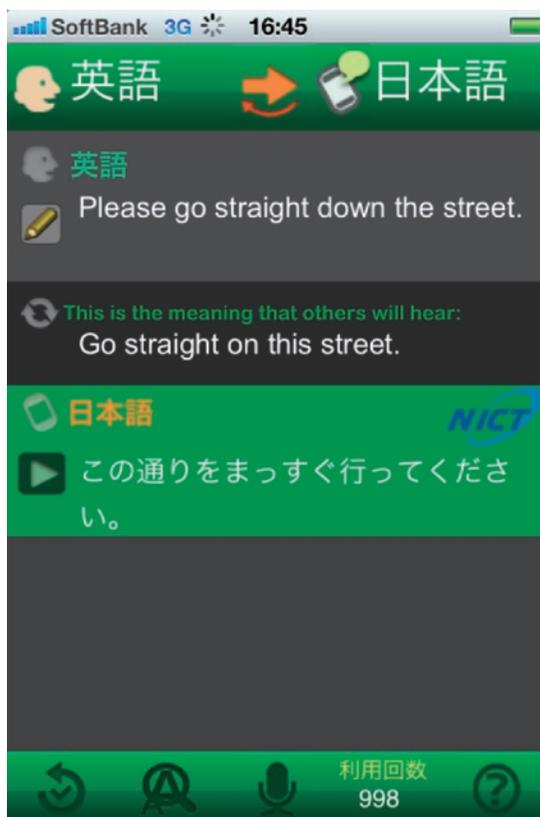


図2 左図の応答例(英日翻訳)

VoiceTra の技術

●「基本の」音声翻訳技術

図3は、日本語音声認識が日本語文章となり、さらに英語文章に翻訳され、英語音声に合成される例を表しています。音声認識モジュールで、多くの話者の音声データから構成された音響のモデル(モデルは音声の要素である音素ごとに構成)と、入力音声との照合が行われて、音素列に変換されます。次に、この音素列は、かな漢字で表記される単語列確率(言語モデルと呼ぶ)を最大化するように変換されます。この変換では、日本語の大量のテキストから学習された、3つ組の単語列の生起確率をもとに、適切な単語列を求めます。これをさらに翻訳モジュールで、日本語の単語列が対応する英語の適切な単語の選択、および語順の入れ替えが行われます。日本語の単語列に対応する英語の単語列を選択するために、日本

語と英語の対訳文から学習された翻訳モデルを用います。次に、語順を英語に合わせるため、大量の英語のテキストから学習される3つ組の単語列の生起確率から英語として適切な単語列を求め、それを音声合成部へ送ります。音声合成部では、まず、英語の単語列にあわせて発音、イントネーションを推定します。次に、それにあう波形を、大量の音声から学習された音声特徴量に合わせてフレームと呼ばれる時間単位で作り、それらを接続して音声合成を行います。

図の下方にある大規模コーパス(日本語のデータや対訳文や英語の音声のデータなど)を基盤にして、そこから自動的にシステムを構築するため、コーパスベースの技術と呼ばれます。

● ネットワーク型の「音声翻訳技術」と実用化

さらに、前述の基本技術を実用化するために無線通信を活用したネットワーク型にしました。単一のコン

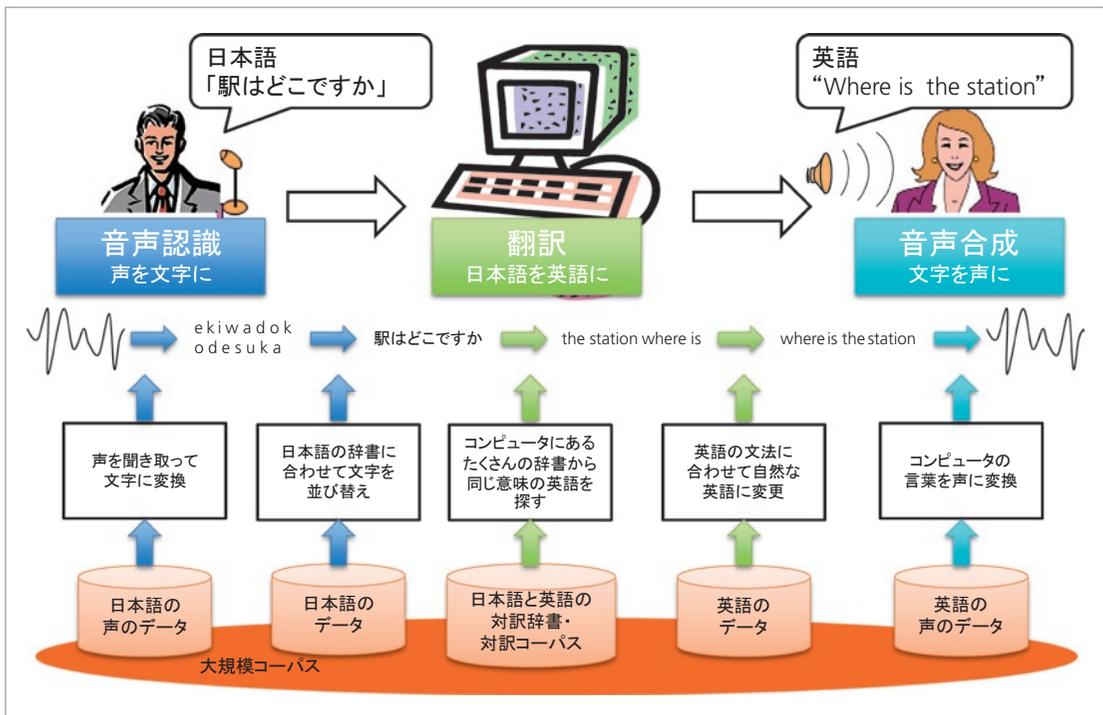


図3 音声翻訳の概略

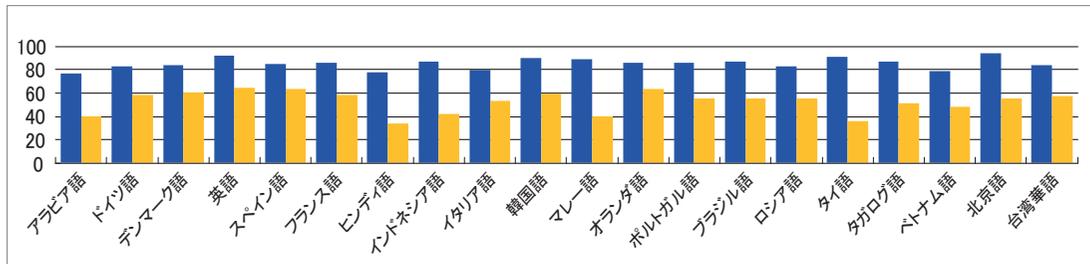


図4 翻訳率の比較 (広く利用されているソフトウェア (黄色) とNICTのソフトウェア (紺色) と比較。縦軸が日本語への翻訳率。横軸が翻訳元の言語)

コンピュータ内に閉じた実装では得られない可搬性、言語や語彙の拡張性、自律的な性能改善など特有の効果で、実用化を加速することができました。①利用者端末を100g程度に軽くできるので可搬性が高くなり実用性が増しました。②サーバはハードの制限がほぼないので、言語や語彙の拡張性は大きく、利用データに基づく自律的な性能改善が可能になりました。実際に、VoiceTraの利用データの一部を利用し音声認識の改善を行ったところ、対象言語により差はありますが、5%から10%、精度が向上できました。

NICTと成田国際空港株式会社(NAA)は2010年10月4日から2011年2月25日まで、商用化検証実験を実施しました。成田国際空港に関連する固有名詞(エアライン名、観光地名、駅名、商品名等)1,600件を追加し、従来、語彙の不足から「穴のカウンターは何処ですか?」と誤認識されていた音声も「ANAのカウンターは何処ですか?」と正しく認識・翻訳が可能となりました。NAAは、ネットワーク型の「音声翻訳技術」が外国人との「言葉の壁」解消のソリューションとなると判断し事業化に着手し、2011年12月末にアプリケーションを旅行者のスマートフォンにダウンロードするサービスを開始しました。VoiceTraは本件を含め4社に技術移転されました。

● 翻訳ソフトウェアの性能

VoiceTraは旅行会話を対象としていますが、その翻訳能力としては、おおよそTOEIC600点の人に相当します。VoiceTraの特徴は、多言語対応であると同時に高品質な点にあります。図4のグラフは、日本語への他の20言語からの翻訳について、広く利用されている多言語ソフトウェア(黄色で表示)とNICTのソフトウェア(紺色で表示)と、翻訳率(翻訳者が評価した意味が通じる率)で比較したものです。

● 音声翻訳研究の今後

音声翻訳技術は1986年に基礎研究が開始されましたが、VoiceTraは、同技術がついに実用化に至った、大きな成果の1つです。しかしながら、現在の音声翻訳技術には、長い文に対応できないこと、文脈を理解できないこと、などの課題があります。NICTは、これらの困難な研究課題に取り組み、ニュースや会議の同時通訳という次の大きな夢の実現を目指しています。

*1 <http://www.kantei.go.jp/jp/sinseichousenryaku/>
*2 <http://mastar.jp/translation/index.html>

情報分析技術

—大量の情報の中から価値ある情報を見つける—



風間 淳一 (かざま じゅんいち)

ユニバーサルコミュニケーション研究所
情報分析研究室 主任研究員

2004年に東京大学大学院を修了し、博士(情報理工学)を取得しました。その後、北陸先端科学技術大学院大学の助教を経て、2008年からNICTに勤務しています。専門は自然言語処理で、機械学習や確率モデルの自然言語処理への応用の研究を行ってきました。プライベートでは(最近はあまり行くことができませんが)釣りをしたり、けいはんな周辺の色々な名所を巡ったりしています。また、最近、家庭菜園を始めて、北陸先端大時代に知って好きになった加賀野菜の「金時草」を植えたりしています。

「大量の情報の中から価値ある情報を見つけるために私の所属する情報分析研究室で研究開発を行っている様々な技術、特に質問応答システム『一休』について紹介します。」

● はじめに

近年のインターネットの発達と、そこでの情報発信や検索の技術の発展には目覚ましいものがあります。しかし、現状で十分満足でしょうか。今でも、Web 検索の結果はページが単純にランキングされてリストアップされるだけで、検索する人が元々欲しかった情報を手に入れるためには、リストアップされたページを順に見ていくしか方法がありません。私の所属する情報分析研究室では、このような現状に満足せず、大量の情報の中から価値ある情報を見つけるための様々な技術の研究開発を行っています。

● 質問応答システム『一休』

質問応答とは、ユーザが「脳梗塞を防ぐのは何ですか?」のような質問文に対して、答えの候補を列挙する技術です。通常の Web 検索と大きく異なる点は、入力が文であることと、結果が「答え」の端的な列挙であるということです。このことから分かるように、通常の Web 検索よりも一段高度で、「価値ある情報の発見」という私たちの目的にかなった技術です。実は、質問応答の研究は古くから行われてきましたが、つい最近までは実用には届かない難しい技術ととらえられてきました。しかし、最近の技術の進歩で、かなり実用に近いところまで来ています。Web 上の大量のデータや、Wikipedia などのある程度組織化されたデータから、必要な知識を自動で抽出する技術が大きく進歩したためです。

我々の研究室でも、これまで開発してきた様々な技術を応用して、『一休』という質問応答システムを開発しています。このシステムの特徴は、Web に書かれていることであればおおよそどんなトピックに関する質問でも回答を返すことができるということです。音声入力機能を備え、スマートフォン上で動作します。技術的に

は、「言い換え認識」という技術が鍵となっています。私たち人間が使う言語では、同じ意味を表すために様々な表現を使うことができます。様々な言い方をされている所から回答を抽出するために、この言い換え認識の技術が重要となります。また、「推論」といういわば「計算機に考えさせる」技術を使って、Web に直接的に書かれていない回答を導くことも可能です。

ちなみに、我々がいるユニバーサルコミュニケーション研究所からほど近い京都府京田辺市にはとんちで有名な一休和尚(一休さん)が晩年を過ごしたお寺が有ります。『一休』という名前は、一休和尚のようにどんな難問でも答えられるようにという気持ちを込めてつけられました。

図 1 は、『一休』による質問応答の例を示した画面です。「脳梗塞を防ぐものは何ですか?」という質問に対して、「ダークチョコレート」「ナットウキナーゼ」「アスピリン」などの回答がされています。実は、「ダークチョコレート」という回答は、『一休』が用いている約 6 億ページの Web データの中にははっきりと簡単な形では書かれていません。『一休』は、Web 上の別々のところに書かれていた「ダークチョコレートは、動脈硬化を防止する働きがある。」「心臓病や脳梗塞などの原因になる動脈硬化は...」という 2 つの情報を自分で考えて組み合わせることで「ダークチョコレートは脳梗塞を防ぐ」という回答を導き出したのです。最近、アメリカの有名なクイズ番組で質問応答システムが人間のチャンピオンに勝ったというニュースがありました。それでも、そこで用いられたシステムには、『一休』のような進んだ推論の機能はありませんでした。『一休』は、この点で、一步進んだ質問応答システムと言うことができます。我々は、人間が行うような推論、つまり機械に考えさせる技術が、次世代の情報システムにおいて特に重要になると考えて、研究開発を進めています。

私も含めて、研究室の研究者は、開発のために日々『一休』のシステムを試していますが、通常の Web 検索では到底見つけることができない意外で価値ある情報を見つけ、それを自分の生活に実際に活かすという経験をしています。例えば、「味噌汁に何を入れますか?」という質問に対して「チーズ」を見つけて試してみたりといったことです。もう少し真面目な例を挙げると、「デフレの原因は何ですか」という質問に対して、『一休』は、大きな利益を上げているある大企業を意外な答えとして挙げました。Web でさらに調べると、その企業が莫大な利益を内部で貯金し社会に還元しないためであるという根拠を見つけることができ、またその後、雑誌に同主旨の記事が発表されました。これらの例で重要なことは、味噌汁の具としてチーズを、デフレの原因としてその企業を自ら

思いつくことはその分野の素人ではかなり難しいということです。もしかしたら、専門家でも難しいかもしれません。

このように見てくると、現在の通常のインターネット環境では、いかに多くのことがまだ可能になっていないかが分かります。私たちは『一休』の例で示したような、価値ある情報の発見を効率的に行うことができる技術を開発し、普及させたいと考えています。これにより、社会生活をさらに大きく進歩・進展させることができると考えています。

● 耐災害情報技術

2011年3月11日に発生した東日本大震災では、Twitterをはじめとする情報技術の可能性が強く認識された一方で、大量に飛び交う情

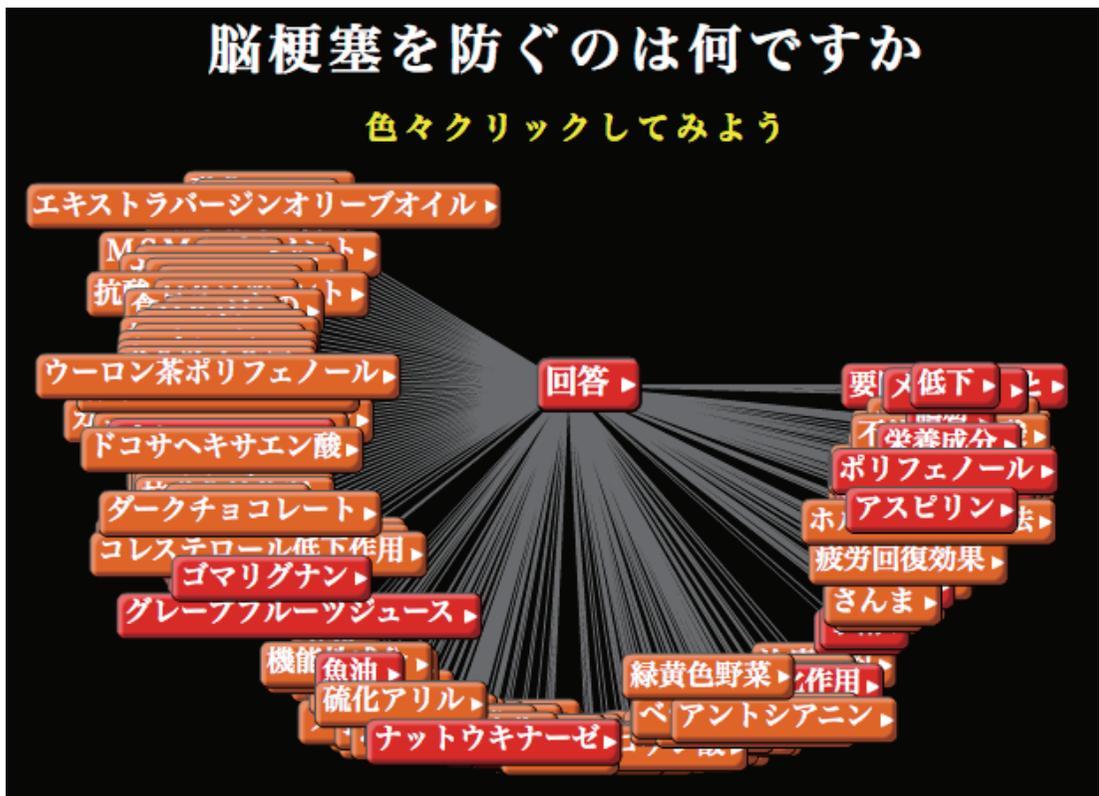


図1 「一休」による質問応答の例

報から状況を正確に把握し、救援や復興に役立てることがいかに難しいかを痛感させられました。結局のところ、私たちには「必要な情報を必要な人に届ける」技術の準備が十分にできていなかったのです。それが、大災害という極限状態になって露呈しました。あらためて、私たちが目標としてきた技術が必要とされていると感じています。情報分析研究室では、「宮城県で孤立しているのはどこですか?」といった災害時の質問に答えることが可能なシステムや、様々な事実や意見を分析して、多様な観点を考慮した冷静な判断を助けるシステムなどの開発を進めていきます(図 2)。

● おわりに

今、日本は原発、デフレなど難しい問題が山積みですが、そうした問題の解決の道筋はいくら Web 検索を使ったり、新聞を読んだりしてもなかなか分からないというのが多くの人の実感ではないでしょうか。現状の研究開発はまだ、そのレベルには達していませんが、将来的には、色々な人が書いている解決の道筋を可能な限り見つけ、分かりやすい形でまとめるといった技術にも繋がって行きたいと思っています。

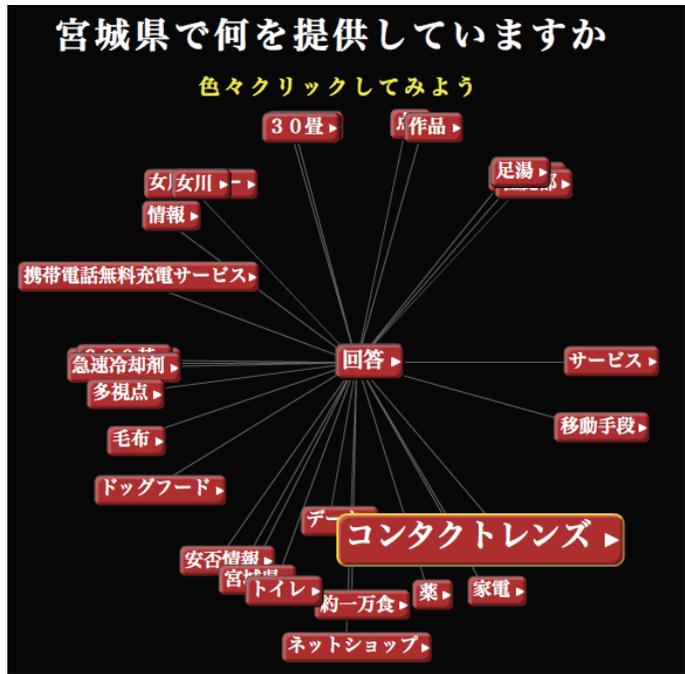


図2 「耐災害一休」プロトタイプによる質疑応答の例

五感を伝える 多感覚統合・評価技術

—自然でリアルな遠隔コミュニケーションを目指して—

安藤 広志 (あんどう ひろし)

ユニバーサルコミュニケーション研究所
多感覚・評価研究室 室長

京都大学にて学部(物理学)・修士課程(実験心理学)を修了後、渡米しMIT 脳・認知科学科にてPh.D(計算神経科学)を取得、その後、国際電気通信基礎技術研究所(ATR)に入社、主任研究員、研究室長を経て、2006年よりNICTの超臨場感プロジェクトに参画。心理物理学、認知脳科学、多感覚インタフェース等の研究に従事。趣味は、自然の風景写真やスナップ写真を撮ること。

「五感による自然でリアルな情報通信の実現を目指して、多感覚情報の統合提示技術、人の知覚認知メカニズムに基づく臨場感の定量的・客観的評価技術の開発を進めています。」

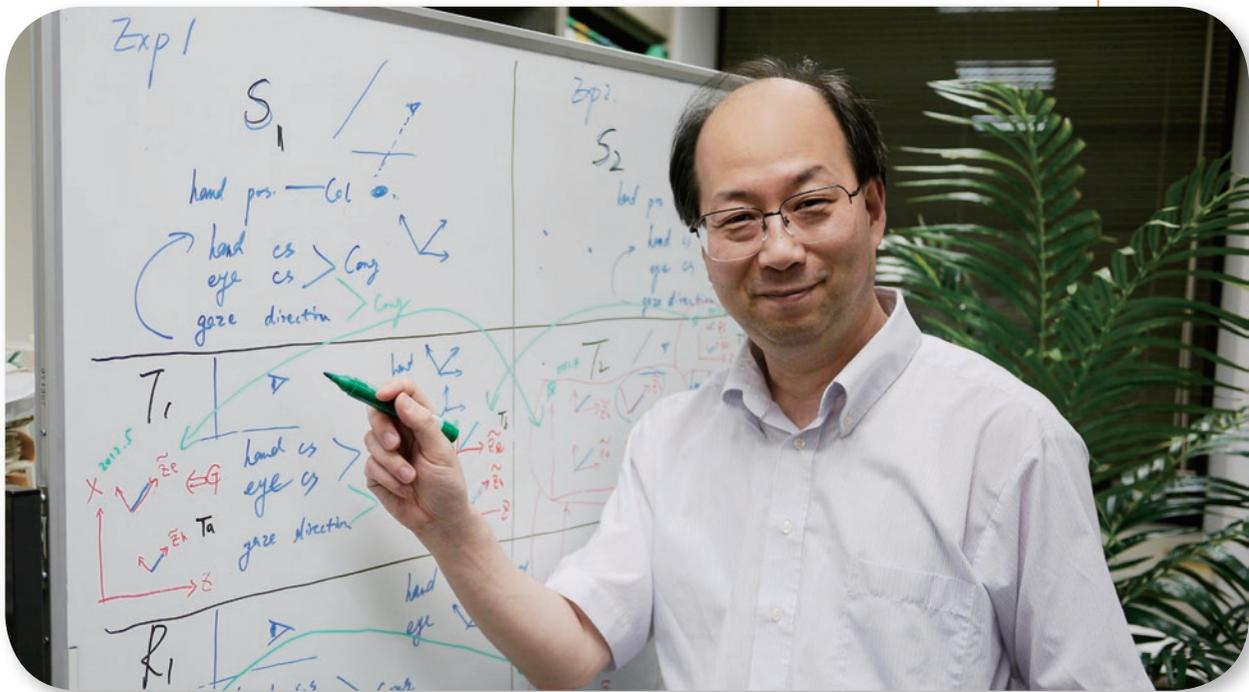




図1 多感覚インタラクションシステム (MSenS)

● 研究の狙い

私たちは、視覚、聴覚、触覚(体性感覚)、嗅覚、味覚といった五感の感覚器官を通して、外界で何が起きているのかを瞬時に理解しています。しかし、現在の情報通信技術では、立体映像、立体音響、感触、香り等の五感情報を統合し人に自然に伝えることはできず、遠隔地にいる人とも実際の対面(face-to-face)のような自然なコミュニケーションを行うことはできません。

多感覚・評価研究室では、このような現在の情報通信の限界に挑み、実世界でのコミュニケーションのように、五感による自然でリアルな情報通信を可能にするための多感覚技術の研究開発にチャレンジしています。また、多感覚の情報を人に違和感や不快感を与えずに、人にとって快適で心地よい情報を伝えるために、人が感じる臨場感を定量的、客観的に捉える測定・評価技術を開発し、人の知覚認知メカニズムを明らかにする研究を行っています。

● 多感覚インタラクション技術

図1は、立体映像、立体音響、感触、香りといった多感覚情報を統合してユーザの動作に合わせて提示する多感覚インタラクションシステム(MSenS)を示しています。このシステムを用いると、立体で提示されたバーチャルな物体映像をあたかも実物体のように自由に操れて、表面の凹凸・ざらつきや柔らかさといった、触れた時の感触も自然に伝わります。このような感触は、あらかじめ実物の表面構造を3次元デジタル化により計測し、物体のモデルを計算機に構築・保存しておき、物体の立体映像を力覚提示装置のペンで触れると、ペン先の位置をセンシングして、モータが3次元的な力をペンに発生させることで生成されます。

また、物体に触れると、その部分の材質(金属、木、皮等)特有の接触音が、触り方(こする、叩く等)に応じてリアルタイムで生成されます。人が音場

を感じる仕組み(音の頭部伝達関数)に基づき、頭の外に拡がる立体音をヘッドホンで聞くこともできます。さらに、私たちが開発した超小型香り噴射装置「マイクロ・アロマ・シューター」を用いて、異なる種類の香りを映像や音・感触とともに提示することができます。香りは、一旦空中に広く拡散してしまうと排除するのが困難になるため、香りの噴出を時空間制御して、香りを含む少量の空気を鼻にめがけて短時間だけ提示しています。

このような多感覚技術を用いて、通常は触れることのできない貴重な文化財(国の重要文化財「海獣葡萄鏡」、正倉院宝物「銀薫爐」等)や科学データ(小惑星イトカワ等)の五感体験など、さまざまな多感覚コンテンツの制作に取り組んできました。今後は、さらに、遠隔地の五感情報をリアルタイムで伝える多感覚通信技術を開発し、遠

隔会議や遠隔医療、災害時の遠隔操作等に役立てていきたいと考えています。

多感覚情報の知覚認知メカニズムと評価技術

一方、多感覚情報を人に安心・安全に伝えるためには、人の知覚認知メカニズムを明らかにし、多感覚情報が人に与える効果をできるだけ定量的に測定する必要があります。私たちは、眼鏡を必要とする3Dテレビの視聴が人の疲労に与える影響を調べる心理・生理評価実験を被験者500名に対して大規模に実施しました(図2)。また、人の知覚認知を定量的・客観的に測定・評価する新しい技術の開発も進めています。

その1つの方法は、心理物理測定による人の知覚の定量的評価です。これは、感覚刺激の物理



図2 眼鏡あり3D映像の心理・生理評価実験

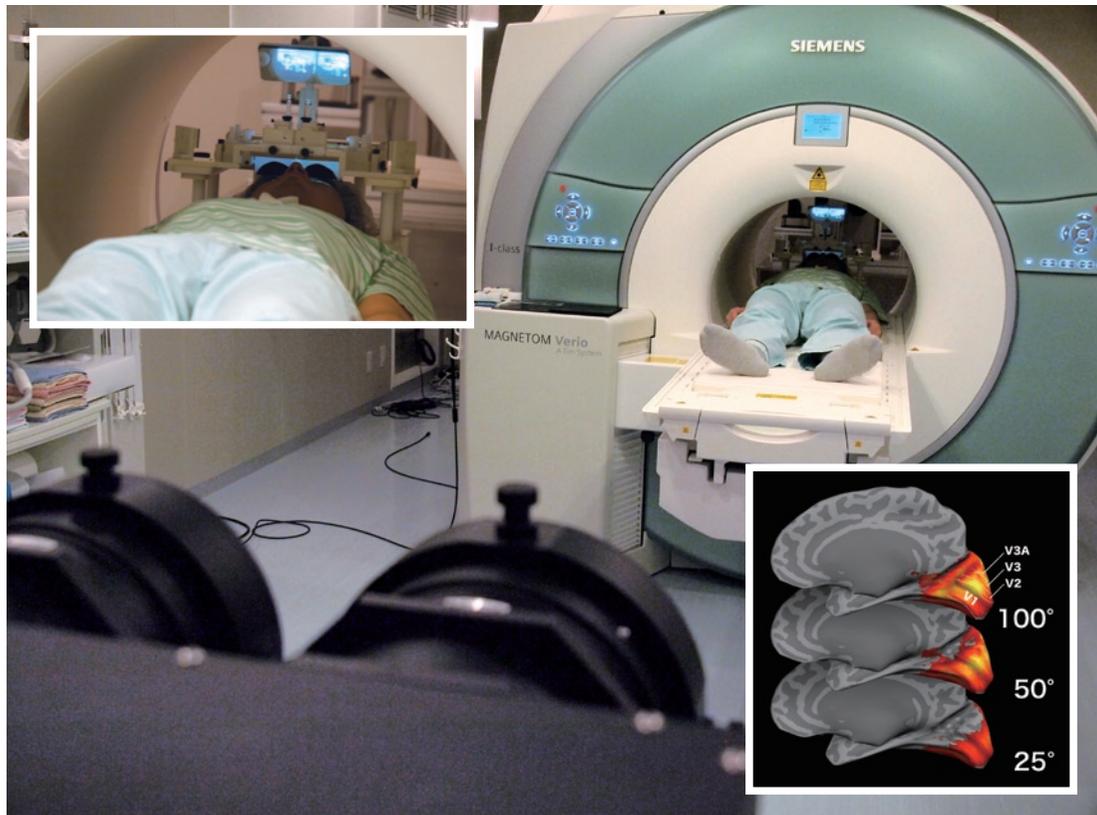


図3 超広視野立体映像提示装置を用いたfMRI脳活動評価実験

パラメータを厳密に制御し、物理情報と知覚反応との対応関係を定式化しようという手法です。私たちは、人が感じる物体の質感の定量的測定を試み、立体や動きの情報が物体表面の光沢知覚に大きく影響することを明らかにしました。

人の知覚認知を定量的に捉えるもう1つの方法は、脳活動計測です。心理物理測定では、刺激と反応の対応だけから知覚認知の仕組みを探ることになりますが、脳活動計測では、直接、脳の中で生じている現象を捉えることができます。私たちは、光沢感が脳の中で生じる仕組みをfMRI（機能的磁気共鳴撮像法）により明らかにしてきました。また、MRI装置の高磁場・狭空間でも使用できる超広視野立体映像提示装置の開発に世界で初めて成功し、水平視野角100°の立体映像が人に与える包囲感・没入感の脳内メカニズムの解明を進めています(図3)。

このような人の知覚認知を定量的・客観的に評価する技術を開発することにより、人に最適に多感覚情報を伝えるための技術要件を明らかにし、多感覚技術利用に関するガイドライン策定や国際標準化に寄与していきたいと考えています。

大量の情報を蓄え 研究を支える情報基盤技術

ービッグデータ時代のI/O指向ハイパフォーマンスコンピューティングー

岩爪 道昭 (いわづめ みちあき)

ユニバーサルコミュニケーション研究所
統括

奈良先端科学技術大学院大学修了後、日本学術振興会特別研究員、近畿大学理工学部助手、理化学研究所脳科学総合研究センター研究員を経て、2005年 NICT 入所。内閣府(総合科学技術会議事務局)出向等を経て、2011年より現職。知識ベースシステム、Web 情報処理、オントロジー工学などの研究に従事。博士(工学)。趣味は、ジャズギター、ルアーフィッシング、キャンピング、アウトドアスポーツ全般。最近は、1歳半の娘と戯れることと、メタボ解消に始めたランニングに夢中。現在の目標は、フルマラソン完走。

「ユニバーサルコミュニケーションの実現に向けて、40 億ページ以上の Web アーカイブをはじめとする大量の情報を収集、蓄積可能な大規模情報基盤の研究開発について紹介します。」



新型データセンターを背景に、研究所長命で編成された企画室共通基盤チームのメンバーとともに。
左から、原口弘志専門研究員、田中康司有期技術員、岩爪道昭統括、泥谷誠専門研究員、岩瀬高博特別研究員、藤井秀明専門研究員

● 背景

近年、ライフログ、インターネット・オブ・シングスと呼ばれるネットワークに接続されたセンサー、自動車等あらゆるモノから非定型かつ大量のデータがリアルタイムに生成されるようになってきました。このようなデータは、短期間に数億～数百億エントリー、ペタバイト級のデータ量に達することもあり「ビッグデータ」と呼ばれています。ビッグデータの集積、有効活用は、2012年1月のダボス会議^{*1}においても、議題に取り上げられるなど、ICT分野にとどまらず、グローバルな新しいビジネスサービスやイノベーションのフロンティアとして経済界においても期待と関心が高まっています。

NICTユニバーサルコミュニケーション研究所では、ユニバーサルコミュニケーションの実現に向けて、多言語音声技術や高度な意味処理に基づく情報分析技術等の基盤技術の研究開発に取り組んでいます。これらの技術では、大規模なテキストデータを収集し、大規模なコーパス^{*2}や各種の言語資源^{*3}を整備することが不可欠です。私は、インターネット上の大量の情報を効率よく集め、蓄え、用途に応じて高速に取り出すための情報基盤技術の研究開発に取り組んでいます。具体的には、(1)超高性能クロールリング技術、(2)大規模分散データストア技術、(3)大規模計算機基盤構築・管理技術の実証的な研究開発を行っています。

● 超高性能クロールリング技術

Web クローラは、ネットワーク上の情報を自動的に収集するソフトウェアの一種です。その実行原理は、Web 文書のリンクを順次

たどるだけの極めて単純なものです。ネット上にはスパムページや重複サイト等が多数存在します。また、ほとんど更新されない情報もあれば、ニュースのように頻繁に更新されるものもあります。一方でネットワーク環境や取得先のWebサイトに過度な負荷をかけないように十二分な配慮も不可欠です。

当研究所では、第2期中期計画(平成18～22年度)において開発された情報分析システムWISDOMの情報基盤として、Webクローラを開発、運用してきました。同クローラでは、1日に最大約1,000万ページのWeb文書を収集することができ、約10億ページのWeb文書を集積しています。第3期中期計画では、研究所全体の研究開発資源として活用可能な40億ページ以上のWebアーカイブの構築に向けて、質の高い情報を効率的に収集するためのクローラ制御の高度化、計算機やネットワーク環境に応じて柔軟にスケールする非同期的な並列分散収集機能の実現を目指しています(図1)。

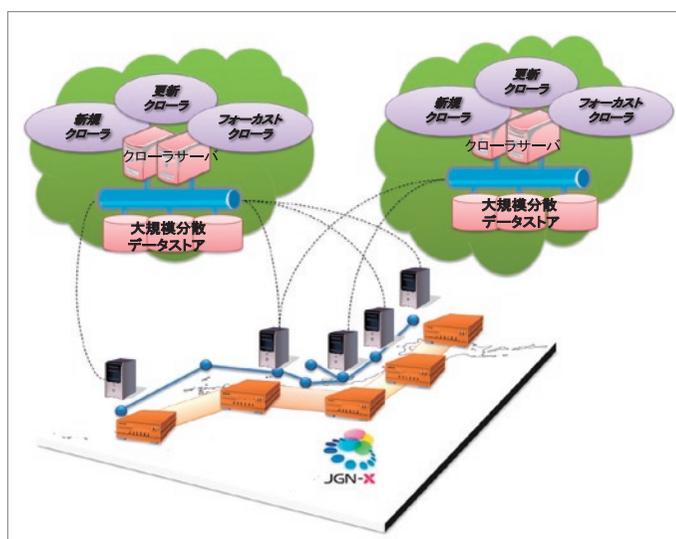


図1 並列分散クロールリング環境

大規模分散データストア技術 (分散キー・バリュー・ストア技術)

数億～数百億エントリー、ペタバイト級のデータ量に及び大量かつ非定型なデータを研究開発やイノベーションのために戦略的に活用するためには、超高速に蓄え、必要に応じて取り出す仕組みが不可欠になります。しかし、従来の関係データベースシステム(以下、RDBMS)は、このような用途には必ずしも適しておらず、仮に実現しようとするとハード、ソフトともに高いコストが伴います。そこで、近年NoSQLと呼ばれる新しいデータベース技術が注目されています。NoSQLは、RDBMSのように関係モデルに基づく固定的なデータ構造ではなく、データや計算機資源の増加に応じてスケールアウト^{*4}しやすいシンプルなデータ構造とシステムアーキテクチャを採用しています。NoSQLは、そのデータ構造や検索方式によってドキュメント指向型、カラム指向型、キー・バリュー型などのタイプがありますが、私の研究では、非定型なネット上の収集、蓄積により適した、分散キー・バリュー・ストア(以下、分散kvs)を採用しています。分散kvsは、データを格納する「データノード」、多重化されたデータノードを管理するとともに、分散kvsへアクセスするインタフェースを提供する「マスターノード」、アプリケーションから分散kvsにアクセスするための「クライアント」から構成されます(図2)。これによりデータや計算機資源の増加に応じた柔軟なデータノードの追加や自動障害復旧が可能となるほか、取扱うデータのサイズ

によっては、データを全てメモリ上に格納することで、毎秒数万～数十万件のデータストリームにも対応可能な超高速なインメモリデータベース^{*5}も実現可能です。

現在、私は、分散kvsのオープンソース・ソフトウェアの1つであるokuyamaの開発者、岩瀬高博氏((株)神戸デジタル・ラボ)と共同で、ビッグデータ時代に対応した大規模データストア技術とそれに基づく大規模Webアーカイブ構築の研究プロジェクトに取り組んでいます。

大規模計算機基盤構築・管理技術～I/O指向ハイパフォーマンスコンピューティング

近年、コンピュータのコモディティ化^{*6}が急速に進んでおり低価格かつハイスpekな計算機資源を利用できるようになってきました。では、多数のCPUや大容量のメモリ、ストレージを搭載した計算機をただ沢山並べれば首尾よく問題は解決するのでしょうか。

問題はそれほど単純ではありません。スパコン

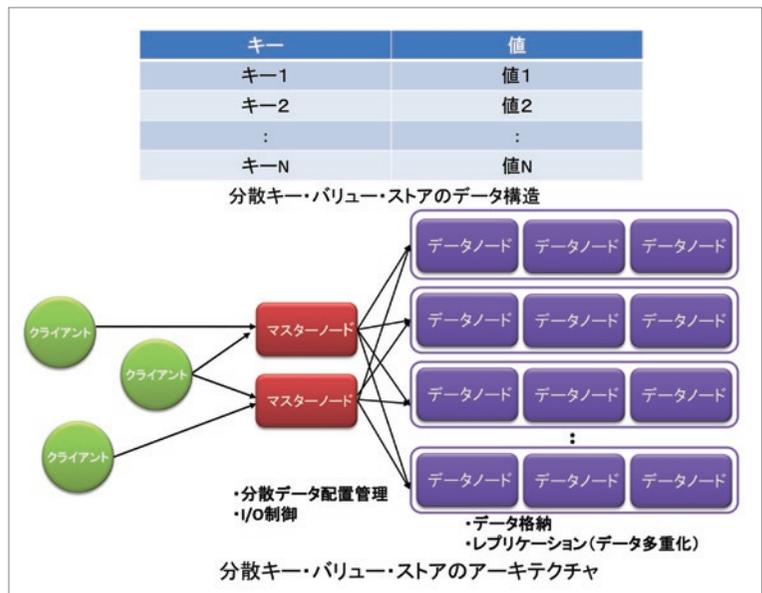


図2 分散キー・バリュー・ストア概要

に代表されるハイパフォーマンスコンピューティングの分野では、浮動小数点演算をいかに高速に実行するかということに主眼が置かれていました。しかし、ビッグデータ時代の到来によって、その様相は一変し、いかに大量のデータを取り扱えるかが重要になっています。

例えば、前述の情報分析システム WISDOM では、1 つの Web 情報はたかだか 100 キロバイト程度の小さ

いものですが、これを数億～数十億件のオーダーで取り扱います。この際に、最も問題になるのは、演算速度ではなく、ストレージ上に存在する大量のデータへの書き込み / 読み出し (I/O) の速度です。特にハードディスクは、どんなに最新の高性能のものでも、メモリと比較すると $10^{-5} \sim 10^{-6}$ の速度になるため、データ I/O の遅延 (レイテンシー) がボトルネックとなり、情報システム全体のスループットが上がりません。

膨大なデータを効率よく活用するには、コンピュータの持つポテンシャルを最大限に引き出すために、I/O のレイテンシーを極力最小化するためのハードウェア構成はもちろん、分散ファイルシステム、分散処理フレームワークなどのミドルウェアも一体となったシステム構築技術が不可欠です。また大規模な計算機基盤を少人数・少コストで運用するための管理技術も非常に重要です (図 3)。

● 今後の展望

今後ますます重要性が増していくアジアに関する情報を 40 億ページ規模で集積化し、当研究所内外の研究開発及び成果の社会還元役に

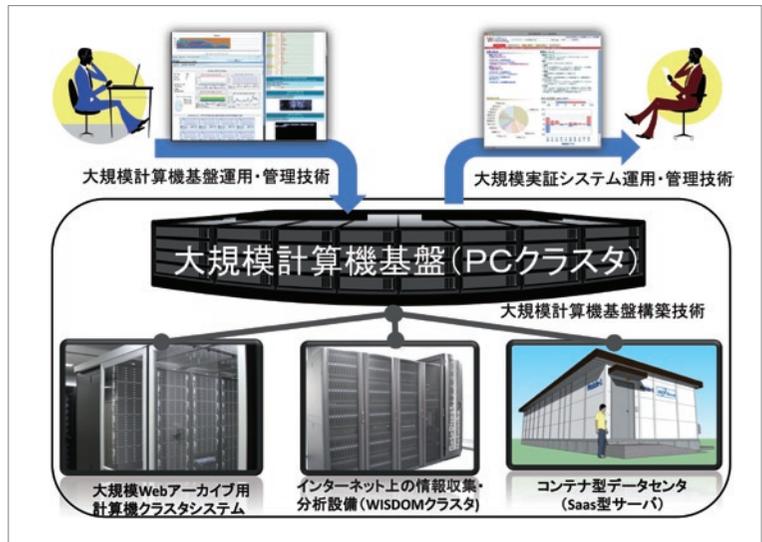


図3 大規模計算機基盤構築・管理技術

立つ情報のハブの実現を目指します。また、これらの実証的研究開発を通じて、ビッグデータ時代の新しいコンピューティング分野を開拓し、ユニバーサルコミュニケーション及び新世代ネットワークの実現に貢献していきたいと考えています。

用語解説

*1 ダボス会議

毎年1月末にスイスのダボスで開催される「世界経済フォーラム」の年次総会。世界中の知識人、経済・企業のトップ、政治家、学者、ジャーナリスト等が一堂に会し、世界が直面する重大な問題について議論する。

*2 コーパス

日本語や英語などの自然言語の文書を大量に集めた言語データのこと。

*3 言語資源

自然言語処理を研究する際に利用する辞書などの資源のこと。

*4 スケールアウト

サーバの数を増やすことで、サーバ群全体の処理能力を向上させること。

*5 インメモリデータベース

データストレージを主にメインメモリ上で行うデータベース管理システムのこと。

*6 コモディティ化

商品がメーカー、機能、品質等において差や違いがはっきりしなくなること。

人の視覚限界に迫る映像技術

—大容量映像データの伝送を可能とする映像伝送技術—

荒川 佳樹 (あらかわ よしき)

ユニバーサルコミュニケーション研究所
企画室 専門推進員

1980年早稲田大学大学院理工学研究科機械工学修了。
同年松下電器産業(株)(現パナソニック(株))入社。
1990年郵政省通信総合研究所(現NICT)入所。幾何・
画像・映像情報の研究に従事。工学博士。趣味は森
林浴とサボテン(霸王樹)。現在、「第三の眼」に興味を
持っています。

「21世紀の通信は『分身通信』(身体
性拡張 / 3次元遠隔操作通信)です。
その実現の第一歩として、超高精細
3次元映像技術とその伝送技術を
研究開発しています。」





図1 4K超高精細映像プロジェクタ(左)と4Kカメラ(右)(どちらも世界1号機)

● 人工知能と人工視覚

21世紀における人類の夢の1つは「人工知能」です。その中核となるのが、「人工視覚」の実現です。そして、この人工視覚を実現する第一歩として、「人の視覚限界に迫る超高精細3次元映像(画像)技術」を研究開発し完成することが、まず重要であると考えています。

● 4K 超高精細映像技術

1997年から、現行のHDTV(ハイビジョン)を超える超高精細映像技術とその伝送技術の研究開発に取り組んできました。NICTと日本ビクター(株)(現(株)JVCケンウッド、以下JVCと呼ぶ)は共同研究を行い、HDTVの4倍の解像度である800万画素(水平3,840×垂直2,048画素)を実現した4K超高精細映像プロジェクタ(2001年、図1左)および4K超高精細映像カメラ(2002年、図1右)を、いずれも世界に先駆けて世界1号機として完成しました。現行のハイビジョン映像は、水平1,920×垂直1,080画素です。このハイビジョン映像の4倍(4画面分)である3,840×

2,048または2,160画素の解像度を持つ映像を、水平解像度が $3,840 \div 4,000 = 4K$ 画素であることから、4K2Kまたは4K映像と呼んでいます。

● 分身通信と4K 超高精細ロボットビジョン

また、ネットワークを介して分身を実現する分身通信(身体性拡張通信、3次元遠隔操作通信)に関する研究開発を行ってきました。図2に示すように、人の手が持つ機能に近い触覚付き5本指ハンド(両手)を開発し、「分身ロボット」のプロトタイプを2003年に完成しました。

一方、JVCは、4Kカメラを当初(図1右)と比べて、体積比および重量比で約1/20に小型化かつ軽量化することに成功しました。図2上部に示すように、2005年に、この小型4Kカメラを、分身ロボットの頭部に、ロボットの眼として搭載しました。人の視力に一段と迫るロボットビジョンを実現しました。図3に示すように、この4Kロボットビジョンが撮像する超高精細映像を用いて、分身ロボットを遠隔から高臨場遠隔操作できるようになりました。

● マルチチャンネル映像伝送システム

現在、並列 PC で構成されるマルチチャンネル映像伝送システム(オールソフトウェアコーデック)を研究開発しています。この伝送システムは、図 4 に示すように、多数の HD 映像(多チャンネル映像)を同時 / 同期伝送することができます。さらに、このシステムでは、超マルチチャンネル映像同期伝送を実現していますので、例えば、200 視点を有する超多視点 3D 映像(メガネなし裸眼 3D 映像)を同期して伝送することも可能です。

最近の成果として、伝送ソフトウェアを PC マルチコアテクノロジーに最適化し、PC1台(1対)で、4K 超高精細映像をソフトウェアのみで伝送することを可能にしました。これにより、超高精細映像の伝送が手軽に行えるようになりました。一般家庭に 4K 映像が普及するのにもそう遠くないと考えています。

● 4K3D 超高精細 3 次元映像技術とその伝送

2010 年に、図 4 右上に示すように、NICT と JVC

は共同で、2 台の小型 4K カメラを用いて、4K3D 超高精細 3 次元映像カメラを開発しました。カメラを極薄型化することにより、人の眼の間隔 65mm にかなり近いカメラ間隔 70mm を実現しました。これにより、眼に自然な 3D 映像を撮影することが可能になりました。さらに一段と、人の視覚に近づくことができました。

2010 年 7 月 6 日(火)には、図 4 に示すように、NICT 鹿島宇宙技術センター(茨城県)と NICT ユニバーサルコミュニケーション研究所(京都府)を JGN2plus で接続し、マルチチャンネル映像伝送システム 8 チャンネルを用いて、4K3D のライブ映像を世界で初めて伝送上映することに成功しました。

さらに、2010 年 11 月 2 日(火)には、このマルチチャンネル映像伝送システムと超高速インターネット衛星「きずな」(WINDS、使用帯域 400Mbps)を用いて、4K3D 映像のライブ伝送にも成功しました。図 5 に示すように、大極殿(奈良平城遷都 1300 年祭会場)の 4K3D ライブ映像を、WINDS 車載局 → WINDS 衛星 → NICT 鹿島宇宙技術センター地上局 → JGN2plus → けいはんなプラザの伝送経路で、世界で初めて衛星通信することに成功しました。



図2 分身ロボット



図3 4Kロボットビジョンの撮像映像 (3D操作グローブを用いて遠隔操作)



図4 JGNを用いた4K3D映像ライブ伝送実験 (2010/7/6)

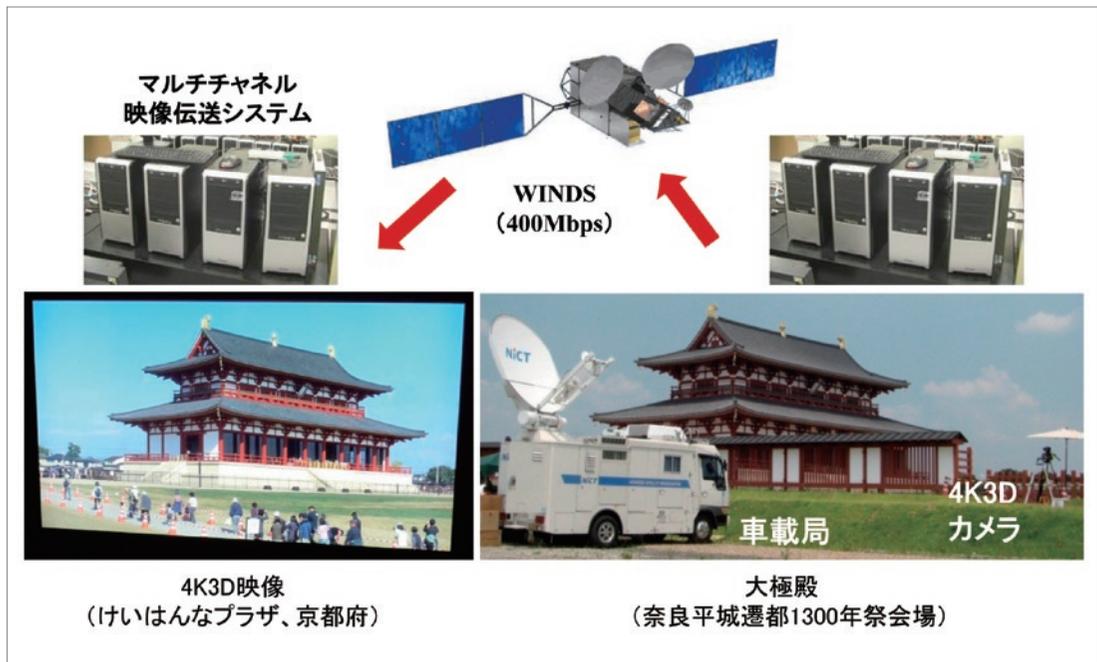


図5 WINDS衛星を用いた4K3D映像ライブ伝送実験 (2010/11/2)