

# 社会情報基盤のための音声・映像の知的統合

末永 康仁 名古屋大学大学院情報科学研究科 [suenaga@is.nagoya-u.ac.jp](mailto:suenaga@is.nagoya-u.ac.jp)  
 間瀬 健二 名古屋大学情報連携基盤センター [mase@itc.nagoya-u.ac.jp](mailto:mase@itc.nagoya-u.ac.jp)

名古屋大学情報系COE「知的メディア統合」研究拠点(拠点リーダー:情報科学研究科教授 末永康仁)は、文部科学省21世紀COEプログラムにおいて、平成14年度に情報・電気・電子分野で採択された20拠点の中でも数少ない情報系の拠点である。ここでは本拠点の概要とこれまでの活動状況を紹介する。本拠点では、情報化社会における「耳」と「目」と「頭脳」の実現を目指し、「聞く技術」(音声メディア処理)と「見る技術」(映像メディア処理)の知的な統合によるメディア情報処理の実証的研究を推進中である。

## 「耳」と「目」を知的に統合する名古屋大学情報系COE

科学技術の目覚ましい進展とともに、我々人間の住む社会のさまざまな局面において情報の重要性はますます高まってきている。名古屋大学情報系COE「知的メディア統合」研究拠点(正式拠点名:「社会情報基盤のための音声・映像の知的統合」<sup>☆1</sup>, 英文略称:「Intelligent Media Integration (IMI)」, 拠点リーダー:末永康仁 情報科学研究科教授)は、文部科学省21世紀COEプログラムにおいて、平成14年度に情報・電気・電子分野で採択された全国20拠点の1つであり、情報系の拠点である。これまでに多くの実績を有する著名で経験豊かなベテラン研究者に加え、さまざまなオリジナルの研究成果を意欲的に世界に発信し続けている新進気鋭の若手研究者を全国の大学および産業界から積極的に集め、強力な研究教育拠点を構成している。事業推進担当者の1/3以上が企業経験者である。また、ほぼ全員が海外の一流研究機関での滞在研究経験を有し、世界の研究者たちと活発な交流を行っている。

☆1 <http://imi.coe.nagoya-u.ac.jp/>

本拠点では、社会情報基盤における「耳」と「目」および「頭脳」の実現を目指し、「聞く技術」(音声メディア処理)と「見る技術」(映像メディア処理)の知的な統合によるメディア情報処理の実証的研究を推進する。大規模な実世界データベースの構築とそれを用いた国際コンペティションの実施により、メディア情報処理技術の実証的研究を行い、メディア情報処理理論の質的な高度化を図ることを目指している。

本拠点は、平成15年度に新設された名古屋大学大学院情報科学研究科を中心に、工学研究科、情報連携基盤センター、情報メディア教育センターの教員で構成されている。とりわけ情報科学研究科のメディア科学専攻を本拠点形成の核として、音声・画像メディア処理研究のエキスパートを集中的に集めた。また、関連する専攻、センター等の組織に所属する情報各分野の専門家をメンバーに加えて図-1および表-1に示す強力な推進体制をとっている。本拠点では、上記事業推進担当者20名を含む総勢約40名の研究者が協力して意欲的な研究教育プログラムを推進中である。研究者の専門分野は音声・音響、画像・映像、認知、情報基礎、ソフトウェア、データベースなどの広い分野にまたがっている。

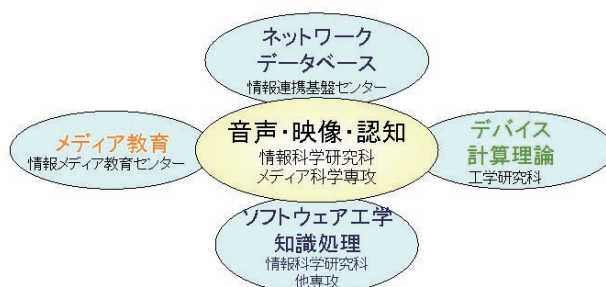


図-1 名古屋大学情報系COE推進体制

氏名	所属	専門分野
末永 康仁	(メ)	画像処理
阿草 清滋	(情)	ソフトウェア工学
谷本 正幸	(電)	画像通信工学
渡邊 豊英	(社)	知識情報処理
横井 茂樹	(社)	メディア情報学
石井 健一郎	(社)	パターン認識
坂部 俊樹	(情)	計算機言語論
平田 富夫	(計)	アルゴリズム
大西 昇	(メ)	生体情報工学
村瀬 洋	(メ)	画像処理
間瀬 健二	(基C)	インタフェース
吉川 正俊	(基C)	データベース学
武田 一哉	(メ)	音声情報処理
長尾 確	(メC)	メディア情報学
山里 敬也	(メC)	通信理論
河口 信夫	(基C)	ネットワーク
森 健策	(メ)	画像情報処理
三輪 和久	(メ)	認知科学
藤井 俊彰	(電)	画像情報処理
工藤 博章	(メ)	音声音響信号処理

メ： 情報科学研究科・メディア科学専攻  
 社： 情報科学研究科・社会システム情報学専攻  
 情： 情報科学研究科・情報システム学専攻  
 計： 情報科学研究科・計算機数理科学専攻  
 電： 工学研究科・電子情報システム専攻  
 メC： 情報メディア教育センター  
 基C： 情報連携基盤センター

表-1 名古屋大学・情報系COE事業推進担当者

### 知的メディア統合

メディア情報処理技術は、コンピュータと人間を結ぶ優れたインタフェースの実現に不可欠であるという意味で社会情報基盤の根幹技術であり、その真の高度化は、社会生活の利便性や快適性、安全性の向上に直結する。過去数10年間、コンピュータは速度、容量ともに目覚ましい性能向上を遂げてきており、今後もまだまださらなる向上が見込まれている。しかし、コンピュータの性能向上のみによって音声処理、映像処理の問題がすべて解決されるわけではない。

音声、映像ともに、同じ物理空間をさまざまな音声マイクや映像カメラで計測することによって入手できるものである。従来、音声と映像は一般には別々に処理されることが多かったが、今後は特に、それらを人間活動に効果的に利用するという観点から両者を知的かつ統一的に扱うことがきわめて重要となってくる。音声、映像を単に物理空間から得られる信号として扱うだけでなく、人間活動に密接にかかわる重要なメディアとして知的な扱いをするための賢い仕組みの解明と構築が求められている。しっかりとした基本原理、理論を構築するとともに、それを実世界のデータに適用し、効力を検証することが必要である。我々は図-2に示すように、映像信号

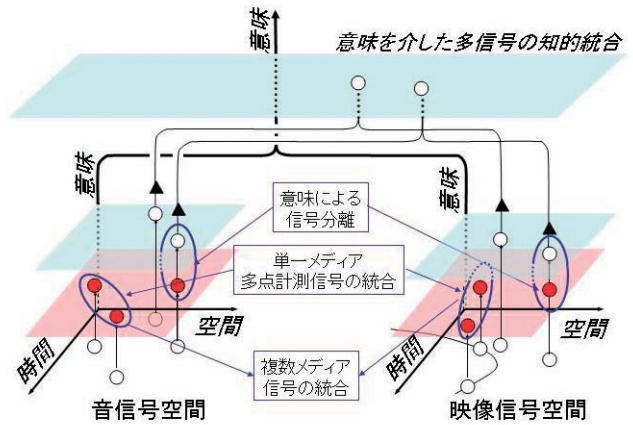


図-2 知的メディア統合

と音声信号が人間の社会活動などによって意味づけされて統合されるというモデルを仮定して、意味を介した各種信号の統合方法を模索している。

### 実証的アプローチ

#### —大規模DBの収集と共同利用—

我々は当初より音声・映像メディア処理の研究においては実世界の音声・映像データを利用することがきわめて重要であると考えている。特に我々人間自身および人間を含む環境からなる実世界を多元的に計測記録することを可能とするため、多数の音声マイクと映像カメラを利用した計測系を構成し、これを用いて実証的な研究を進めてきている。この計測系を用いて音声・映像の大規模な実世界データを収集し、共通課題を設けて他研究機関も参加可能な国際的競争評価を実施することを目指している。データの共同利用や人的交流を通じ国内外の研究機関との連携を図り、音声と映像の実世界データベースを用いた実証的な研究を推進し、国際的ベンチマーク（競争的技術評価）を実施することによって、世界中から有益な情報を収集するとともに、実世界データベースの使用を通じて有益な情報を世界中に発信する計画である。

まず、音声メディア処理研究を核とした計測系として、マイクロフォンをはじめとする各種センサを搭載した計測実験車を用いて、データの収集を行っている。このデータは、カーナビゲーションシステムなど車載情報システムのユーザインタフェースや音声認識システムの設計に有効に利用される。現在、「車内音声・映像信号コーパスおよび評価環境（CENSREC-3およびAURORA-J-AV）」と呼ばれるデータベースを構築中であり、一部は情報処理学会のWGを通じて配布を開始している。次章

にその詳細を述べる。

さらに、映像メディア処理研究を核とした計測系として、自由視点テレビの研究成果をもとに、2003年度までに高品質音声入力24チャンネル、実時間高解像度映像入力3チャンネルの完全同期入力可能な音声・映像入力装置のプロトタイプを開発した。2004年度はこのプロトタイプ構築の経験をもとに、映像入力および音声入力のユニットを専用ボード化し、装置全体を小型で使いやすいものとした改良型を作成した。その結果、高品質音声入力200チャンネル、実時間高解像度映像入力100チャンネルの完全同期入力可能なシステムのハードウェアを実現した。現在、このハードウェアの性能を存分に発揮するソフトウェアの開発を含む各種調整を進めている段階である。「自由視点テレビプロジェクト」の章では、自由視点テレビプロジェクトおよび音声・映像入力装置の紹介と、映像100チャンネルによるデータ収集の準備状況を紹介する。

さらに、続く章では多次元信号の統合に基づく医用画像処理技術を紹介する。医用画像処理は、映像を中心とした研究が進められているが、CTやビデオなど各種モダリティの統合や知的画像診断のための研究が行われている。本稿では、手術・検査ナビゲーションシステムの研究成果を紹介するとともに、ナビゲーションにおける多点映像・音声観測による高度ユーザインタフェース開発への展望を述べる。

### 実世界音声・映像信号コーパスと評価基盤

本拠点の目的の1つに、実世界を対象とした大規模な実測信号を収集記録するとともに、それをさまざまな研究目的に共用可能な形式に整備して、実世界の観測信号を対象とした多様な研究の評価基盤を構築することがある。ここでは、そのような実世界データベースの収集と評価環境の整備の一例として、情報処理学会雑音下音声認識評価ワーキンググループと連携して整備を進めている、2種類の車内音声・映像信号コーパスおよびその評価環境（CENSREC-3, AURORA-J-AV）について概説する。

#### CENSREC-3

CENSREC-3のデータ収集には、名古屋大学統合音響情報研究拠点CIAIRで整備された、多目的信号計測実験車<sup>1), 2)</sup>を用いた。当該実験車は、走行車内の音声・映像を、音声16チャンネル、画像3チャンネルまで同時収録する機能を備えた汎用実験車であり、音声・映像信号のほか、アクセル、ブレーキ、ハンドル、などの運転行動や、車速、エンジン回転数などの車両情報を同期して

取り込む機能が備えられている。

収録した評価データの一覧を、表-2に示す。3種類の走行速度（アイドリング、低速（市街地）走行、高速走行）と、6種類の車内環境（通常走行、ハザードOn, エアコン（Low）, エアコン（High）, オーディオOn, 窓開）を組み合わせた16種類の環境で行っている<sup>1)</sup>。評価データの発話者数は18名（男性8名、女性10名）である。

一方、学習データの収録は、アイドリング、低速走行の条件で行い、車内環境は通常走行のみである<sup>2)</sup>。学習データの発話者は、293名（男性202名、女性91）である。音素バランス文の収録は、アイドリング時には発話内容が記された原稿を読み上げることにより行う。しかし、走行時では原稿を見て読み上げを行うことは不可能であるので、ヘッドセットマイクロホンのイヤホンから発話内容を指示することにより、発話を行う。またこの際、発話者が指示された1文を記憶して発話することは困難であるので、1文をいくつかの文節に区切って指示を行う。

#### AURORA-J-AV

音声情報と映像情報を統合したバイモーダル処理を高度化するためには、両信号を実環境下で同期して収集したデータベースが必要となる。しかし、そのような信号計測は容易ではなく、これまでに作成されたデータベースは実験室環境下で収録されたものは小規模な収集にとどまっており、実環境下で収録されたコーパスは多様な研究に共用することを意図したものではない。本拠点では、音声・映像を統合した知的メディア処理技術の評価基盤として多様な研究に利用可能な実世界信号データベースの構築を進めている。大規模かつ実環境下で音声・映像の同期収録を行うことを目的としており、日本語音声認識の標準評価基盤であるAURORA-2J, AURORA-3Jを拡張するかたちで作成を進めていることから、AURORA-2J-AV, AURORA-3J-AVと呼ばれている。

AURORA-2J-AVは、雑音環境下連続日本語数字音声認識タスクの共通評価フレームワークAURORA-2Jに、発話中の被験者の顔を撮影したカラー映像と近赤外映像を加えて収録を行ったデータベースである。発話セットはAURORA-2Jのタスクである1～7桁の連続数字をそのまま採用した。被験者数は学習セットが41名（男性19名、女性22名）、評価セットが49名（男性23名、女性26名）である。被験者の男女比ができるだけ等しくなるように、また20代から50代の幅広い年齢層を対象とした。

AURORA-3J-Aは、実環境（自動車内）日本語連続数字とコマンド語のデータベースAURORA-3Jの仕様に則



走行速度	車内環境
アイドリング	通常走行, ハザードOn, エアコン(Low), エアコン(High), オーディオOn, 窓開
低速走行	通常走行, エアコン(Low), エアコン(High), オーディオOn, 窓開
高速走行	通常走行, エアコン(Low), エアコン(High), オーディオOn, 窓開

表-2 評価データの収録環境



図-3 自動車内音声映像同期信号コーパスのビュー画面

り、自動車運転中の話者の発話音声に被験者の顔を撮影したカラー映像と近赤外映像を加えて収録したものである(図-3)。AURORA-3J-AVの発話タスクは日本語連続数字である。発話セットは10種類、それぞれの発話セットは1桁の独立数字が4つ、10桁数字が4つ、16桁数字が1つで構成される。10桁数字や16桁数字は、3桁ないし4桁に区切って発話を促す。このような9つの数字列を、走行条件2種類(アイドリング、低速走行)と、車内環境条件(ノーマル、オーディオON、ハザードON、エアコンON)の組合せからなる7つの収録条件で繰り返し収録する。CENSREC-3データベースではアイドリング、低速走行、高速走行の3つの走行条件で収録が行われているが、AURORA-3J-AVでは高速走行を除き、アイドリングと市街地走行のみとした。また、各発話セットで年齢層、男女比のバランスをとるように被験者を配分した。AURORA-2J-AVと異なり、運転中のタスクが含まれるために発話内容を視覚的に提示することはできない。そこで、被験者の片耳にイヤフォンを装着させ、音声により発話内容を指示する。被験者は、イヤフォンから聞こえてきた数字列をリピートするだけでよい。

自動車内で収録された顔映像を利用するためには、夜間は可視光の照明が利用できないため、近赤外照明による近赤外映像が用いられる。AURORA-J-AVでは通常

図-4 自由視点テレビシステム(FTV)  
—多視点カメラとPCクラスタによる並列信号処理—

のカラー映像に近赤外映像を加え、2系統で画像の収録を行った。

CENSREC-3は情報処理学会のWGを通じてすでに配布が開始されており、今後実環境下での音声認識の評価基盤として、幅広い利用が期待される。AURORA-2/3J-AVについても、データ収集をほぼ終了し、現在データベース化を進めるとともに、読唇による音声認識など音声・映像の統合信号処理の研究が進められている。今後は、本拠点において開発された100地点規模の多点信号収集設備を用いて大規模実世界メディア信号のコーパスの構築を進め、音声・映像空間信号処理技術の研究基盤を構築することで、知的メディア統合のブレイクスルーを目指す。

### 自由視点テレビプロジェクト

本プロジェクトでは、ユーザが自由に視点を変えられる自由視点テレビ(Free Viewpoint TV, FTV)の研究を行っている。これは、ユーザがあたかもその場にいるかのように、自由に視点を変えて遠隔地の情景を見ることができ、究極のテレビである。

自由視点テレビは、多数のカメラを用いて映像を同時に取得し、信号処理によって自由な視点の映像を生成する。この自由視点映像を生成する原理は「光線空間法」<sup>3)</sup>に基づいている。光線空間法とは、空間を伝播する光線の情報を用いて3次元シーンの情報を記述する方法である。この手法はリアルな自由視点映像を作り出せる特長がある反面、非常に細かい間隔でカメラを配置する必要がある。自由視点テレビシステムでは、粗い間隔で配置されたカメラの映像を元に、信号処理によって密な光線情報を補間・生成することにより、これを実現している。

図-4に自由視点テレビシステムを示す。水槽を16台のテレビカメラを用いて撮影し、16台のPCクラスタを

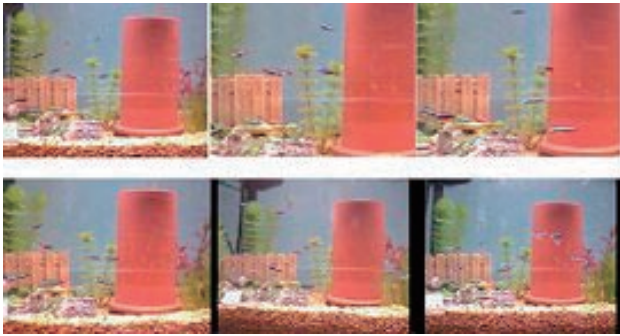


図-5 FTVで生成された自由視点映像

用いてリアルタイムで自由視点映像を生成している。このように入力・処理・自由視点映像表示の一連の処理をリアルタイムで行うことに世界で初めて成功した。これまでも、コンピュータビジョンの技術を用いて3次元形状モデルを生成し、自由な視点の映像を生成する方式が提案されているが、自由視点テレビでは、図-5のように小さな魚や藻、気泡など形状のモデル化が困難なものまで表現できているのが特徴である。

自由視点テレビの研究成果の上に、我々は多点映像音声の同期入力システムを開発している。カメラリンクインタフェースを有する高解像度カメラ（1392×1040画素、カラー、Bayer方式）を用いてフレーム周波数29.4118Hzで映像をサンプリングできる。また、音声をはじめとする多様な信号の多チャンネル入力用として、映像入力に完全に同期した、分解能12bit、最大96KHzでサンプリングできるアナログ入力を備えている。2003年度には初期プロトタイプを完成させた。これは物理的にかなり大きな装置であるが、性能上は画期的なものであり、動作確認と基礎データの取得に成功している。特に、GPSを使って、遠隔のモジュールを高精度で同期させてサンプリングできるという特長を持っている。引き続き、この装置の小型化に着手し、2004年度には専用ボードを開発し、1ユニットあたり、映像1チャンネル、音声2（または4）チャンネルを入力・記録ができるシステムを開発した（図-6）。これまでにこのユニットを100セット導入しており、現在100眼+200耳による高精度同期データ収集の体制が整っている。また、100セットを同時に動作させて、データ収録を行うインタフェースソフトウェアも開発中である。これらをベースに、音声と映像の多点収録データからの任意視聴点鑑賞を可能にする理論的枠組みの研究と、知的統合を可能にする手法の研究開発を進めていく。

自由視点テレビは、映像の国際標準を定める機関であるMPEG(Moving Picture Experts Group)会議において最も挑戦的な課題として認められ、国際的にも高く評



図-6 映像・音声記録ユニット(画像1チャンネル, 音声4チャンネル)

価されている<sup>4)</sup>。我々の2年半にわたる活動の結果、多視点画像符号化の標準化活動がMPEGにおいて開始されることになった。現在、この100点記録システムを用いて、MPEG標準化のための自由視点テレビテスト画像を作成し提出する準備を他機関と協力して進めている。今後は、任意の地点での視聴覚情報を再現できるシステムへと発展させていく予定である。

### 多次元信号の統合に基づく高度医用画像処理技術の開発

医学の分野において人体内部の様子を撮影するイメージング装置の発展は目覚ましく、最近では多検出器型CTの登場により短時間で高精細な人体3次元画像の撮影が可能となっている。1患者あたり500枚以上のスライス画像(人体の輪切り画像)が撮影される場合もあり、計算機を用いた画像診断支援を必要不可欠のものとしている。一方、外科手術においても、手術中における人体・手術器具位置をリアルタイムにセンシングし、手術前に撮影した高精細の3次元画像と組み合わせることで医師を誘導する手術ナビゲーションシステムが実用化されてきている。また、複合現実感を用いた手術ナビゲーションシステムなども興味深い研究対象の1つである。セグメンテーション、変形モデル、コンピュータグラフィクス、レジストレーションなどメディア処理における興味深いトピックスがこの分野には包含される。また、処理すべき対象に関しての各種メディアの統合といった観点から見た場合でも、CT、MRI、ビデオ(内視鏡、超音波画像)といった「各種モダリティ間の統合」、知的画像診断を目指した「解剖学的知識と画像との意味的統合」、手術中におけるナビゲーションシステムの操作、手術中



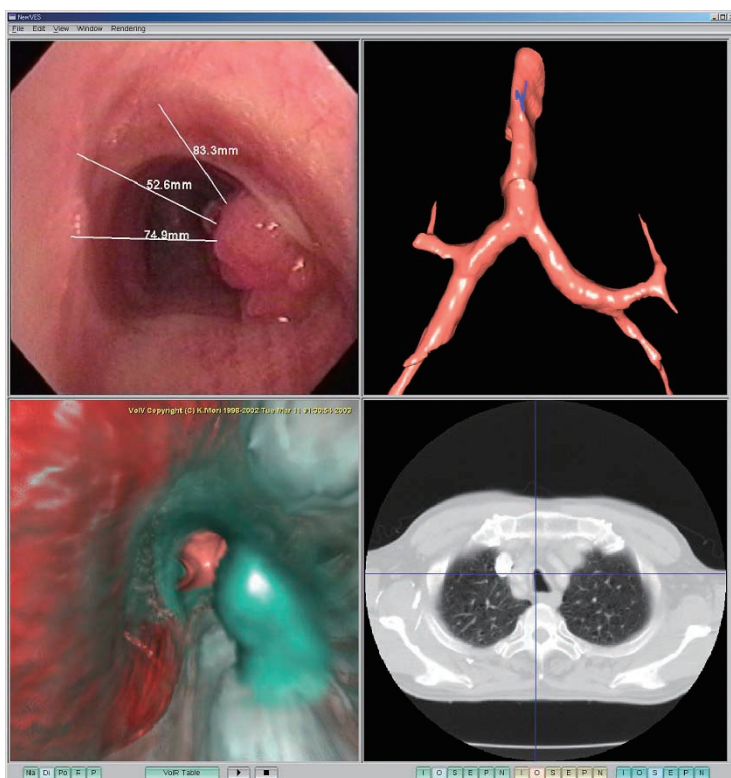


図-7 気管支鏡ナビゲーションシステム概観

ロボットの操作といった医用画像処理システムにおける「ユーザインタフェースとしての各種メディアの統合」などがあげられる<sup>5)</sup>。本研究プロジェクトでは画像誘導下手術・画像診断におけるメディア処理の必要性に着目し、高次元医用画像から構築される仮想化人体と実際の人体(実人体)の統合、医師と医用画像処理システムとの間のユーザインタフェースとしての音声・映像の利用を試みている。以下、研究内容の具体例として気管支鏡ナビゲーションシステムを取り上げる。

手術・検査ナビゲーションシステムは、手術前に撮影された3次元画像を地図として用い、手術中において何らかのセンシングデバイスにより取得された術具の3次元位置情報を組み合わせることで、医師を誘導するシステムのことである。気管支鏡・大腸鏡といった軟性内視鏡に対する検査・手術ナビゲーションを考えた場合、現在内視鏡により観察している部位の検査・手術前画像への表示、臓器壁面下に存在する重要臓器(大動脈など)の表示、目的とする部位までの誘導、などの機能が必須である。仮想化内視鏡システムはこれらの情報をあらかじめ入力されるCT像に基づいて提示することが可能であり、気管支鏡・大腸鏡といった軟性実内視鏡と仮想化内視鏡を融合することができれば(=実人体と仮想化人体の統合)、よりの確な情報を提供可能な軟性内視鏡ナビゲーションシステムが実現できる。

気管支鏡ナビゲーションシステムを考えた場合、気管

支内部の場所的制約からその先端に位置センサを取り付け可能なセンサは存在するが、呼吸による内視鏡位置の変動に影響され、静止したCT像上で現在の観察位置を求めることは難しい。それで位置センシングデバイスの出力と画像同士の位置合わせに基づいた気管支内視鏡ナビゲーションシステムの開発を精力的に進めている。ここでは、実内視鏡ビデオの各フレーム画像と3次元CT像から生成される仮想化内視鏡画像との間の画像間類似度を計算し、画像間類似度が最も高くなるような仮想化内視鏡システムにおける視点位置・視線方向を求める<sup>6)</sup>。探索の初期値として位置センサの出力を用いている。この処理を連続的に実行することにより、実内視鏡カメラの動きが推定可能となる。いったん、実内視鏡カメラの動きが推定できれば、仮想化内視鏡システムが生成するさまざまな情報を実内視鏡画像上に重畳表示することが可能となる。気管支鏡ナビゲーションシステムの画面例を図-7に示す。気管支の場合3次元胸部CT像から抽出された気管支領域の各枝に解剖学的名称を対応付ける手法と組み合わせることによって、実内視鏡の観察時に解剖学的名称を表示することも可能となる。また、手術検査中の医師は検査器具の操作に両手を用いているため、音声、ジェスチャを利用したユーザインタフェースの導入により、使い勝手の良いナビゲーションシステムの構築が可能となる。診断支援や手術支援のための各種映像、音声、触覚等の知的統合も今後の重要な課題である。

## 世界で活躍する若手の育成

社会情報基盤のための音声・映像の知的統合を探求する本拠点において、若手の育成は研究教育プログラムのきわめて重要な目的の1つである。大学院博士後期課程の大学院生をRA (Research Assistant) として雇用し、実世界データの収集・処理システムの構築や評価実験の実施などの具体的プロジェクトを通じた多彩な研究教育プログラムを設け、これを通じて世界レベルで活躍する若手研究者の育成を図ってきている。

本当の意味で世界の一流レベルで活躍できる若手を育成するためには、しっかりとした基礎能力に裏打ちされた学際性の涵養がきわめて重要である。工学、情報科学、認知心理学の広い分野にまたがる研究教育プログラムを実施し、各人がそれぞれの専門分野を深く理解すると同時に、他専門分野に関しても幅広い知識を習得し、異なる分野の問題の間に潜む共通性、相似性、双対性などを把握する能力を養うことを目指している。

毎月大学院生のみによって計画と実施運営がなされるセミナー (IMIセミナー)、優れた企業関係者に講師を務めていただく非定期の産学連携先端技術セミナーを継続実施し、分野の垣根を越えた若手の育成に努めてきている。

また、若手の国際会議発表支援や外国人研究者の招聘を含むさまざまな国際交流推進、ソフトウェア技術・英語論文作成技術・プレゼンテーション技術等の習得のためのセミナー実施、先端的ソフトウェアの配布、本拠点内での公募プロジェクトによる競争的研究支援によるインセンティブ向上など、従来の枠にとらわれない意欲的な研究教育プログラムを推進中である。

COE主催の国際シンポジウムの実施(2回)などは当然のこととして、海外の大学との相互交流として、学生・研究者の短期滞在(数カ月単位)のプログラムを準備している。

本研究拠点では、他研究機関の若手研究者の当拠点における短期滞在研究や、当拠点のCOE研究員(ポスドク)の採用についても積極的に進めており、国内、国外を問わず、優れた人材を広く求めている。いろいろな意味での垣根を越えた真の協力は本当に良いものを生み出すきっかけになり得るはずであり、当拠点のメンバ同、一貫して努力を続けたいと考えている。

## ポストCOE

21世紀COEプログラムは文部科学省の大ヒットである。大学間での競争による切磋琢磨、縦割りにりがち

な学内講座間のテーマ連携の推進、比較的自由度の高い資金援助による新しい教育研究プログラム運営への挑戦など、良い点は多数ある。ポスドク研究員の雇用を促進して研究者市場の流動性向上にも貢献していよう。しかし一方で、成果主義に陥って、見かけの数値を出すためのイベントを無駄に多発する愚は避けねばならない。若手の研究者を疲弊させては本末転倒となってしまうからである。本拠点で、我々は、データベースを収集するシステム・環境を開発し、大規模データベースを収集して評価するという具体的課題を挙げて、着実に研究開発を進めることを主眼としている。また、研究教育拠点の育成・成長には5年間のプログラムでは十分といえない。大学はインキュベータとしてのCOEプログラムをバネに新しい展開をする責務を負っており、我々は一層の努力を続けていく。一方、国も、真の国力増強のため、長期的観点から拠点成長を強力に支援するプログラムを開発・実施していただきたい。国と大学の真の協力が重要である。

**謝辞** 本プログラムを実施するにあたり、本COEの事業推進担当者ほか関係者のご協力に深謝するとともに、今後とも内外の多くの皆様のご指導、ご支援をお願い申し上げます。本稿のまとめに際し、武田一哉、森健策、藤井俊彰の各氏にご協力をいただいた。日頃のCOE運営への積極的な貢献と合わせて深く感謝する。実世界音声・映像コーパスの設計・配布は、情報処理学会音声言語情報処理研究会雑音下音声認識評価ワーキンググループと協力して進めている。WG委員各位に感謝する。

## 参考文献

- 1) 武田一哉他: 走行状況別車内音声データベースとその予備評価, 音講論集, 3-P-10, pp.185-186 (Mar. 2002).
- 2) Takeda, K. et al.: Construction and Evaluation of a Large in-car Speech Corpus, IEICE Transactions on Information and Systems, Vol. E88-D, No.3 (Mar. 2005). (to appear)
- 3) Fujii, T. and Tanimoto, M.: Free-Viewpoint TV System Based on Ray-Space Representation, SPIE ITCOM Vol.4864-22, pp.175-189 (Aug. 2002).
- 4) Tanimoto, M. and Fujii, T.: FTV (Free Viewpoint Television): Achievements and Challenge, ISO/IEC JTC1/SC29/WG11 M11259, (Oct. 2004).
- 5) 森 健策: 仮想化内視鏡と手術支援画像生成, 日本ロボット学会誌, Vol.22, No.4, pp.455-460 (May 2004).
- 6) Nagao, J. et al.: Fast and Accurate Bronchoscope Tracking using Image Registration and Motion Prediction, Proceedings of 7th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2004), pp.551-558 (Sep. 2004).

(平成17年3月4日受付)