

# 実世界情報システムプロジェクト

## ～ 視聴覚研究グループ ～

安藤 繁、小野 順貴、小室 孝、嵯峨山 茂樹、篠田 裕之、並木 明夫 (システム情報学専攻)、  
広瀬 啓吉 (電子情報学専攻)、石川 正俊 (理事/副学長/創造情報学専攻)、  
眞溪 歩 (新領域創成科学研究科)、井野 秀一、伊福部 達 (先端技術開発センター)

### 概要

実世界情報システムプロジェクトの中の「視聴覚グループ」について活動経過を報告する。人間に代わって働き、人間と対話し、人間に見せて聞かせる機械あるいは環境を構成するために、音響、音声、聴覚、視覚、触覚情報などの物理量・感覚量のセンシング、認識と理解、制御と呈示の技術を開発する。

### 1 本グループの方向性

20世紀の情報学は、コンピュータやネットワークやソフトウェアなどの情報世界に閉じた情報学を中心にしてきた。しかし、21世紀の情報学は、人間が生活する実世界に深く関るようになるだろう。そのためには、情報世界と実世界(物理世界)をつなぐ技術要素が極めて重要になる。

実世界の情報処理においては、実世界から情報を獲得することは不可欠の研究分野である。例えば、自律動作するロボットのような機械を考えると、視覚・聴覚・触覚などの知覚および認識機能を持つことが必須である。さらにそれが、人間とコミュニケーションをする能力を持つためには、音声による対話の能力が必要である。

近年進歩した人間型ロボットの運動能力は、人間の骨格・筋肉・小脳に相当するものであるのに対し、本グループは人間の感覚器・大脳に相当する能力を開発するものであり、今後のロボットの重要な技術と捉えることもできる。

### 2 本グループのメンバー構成

現時点では、本COEメンバーの中で、11名のメンバーが6つの研究ユニットとしてこのグループに属して活動している。以下に代表的な専門分野とともに示す。

- 嵯峨山 茂樹 教授 [本グループ主査]、小野 順貴 講師 (システム情報学専攻)  
音声認識・対話、音響・音楽情報処理、文字認識、ヒューマンインタフェース
- 安藤 繁 教授、篠田 裕之 助教授 (システム情報学専攻)  
視覚、聴覚情報のセンシング、音源分離、触覚センサ、二次元通信
- 石川 正俊 教授 (創造情報学専攻)、並木 明夫 講師、小室 孝 講師 (システム情報学専攻)  
ビジョンチップ、センサフュージョン、高速ロボットハンド、ビジョンサーボ
- 広瀬 啓吉 教授 (電子情報工学専攻)  
音声情報処理、音声合成
- 眞溪 歩 助教授 (新領域創成科学研究科 複雑理工学専攻)  
計測信号処理、脳インタフェース
- 伊福部 達 教授、井野 秀一 助教授 (先端技術開発センター)  
福祉工学

この他に、特任助手1名(酒向 慎司)と、RA(Research Assistant)1名(米田 隆一 (システム情報学専攻 博士課程学生))が、主に本グループにおいて活動した。

### 3 本グループの統合研究構想

本グループは、高いレベルの特徴ある個々の研究の集合体でもあるが、同時に本COEプログラムにおいてそれらの技術を融合して具体的なシステムとして実現し、個々の技術の有用性を立証することも目標としている。

図1に示すのは、博物館内を案内するシステムである。音声対話擬人化エージェントを搭載した大型のディ

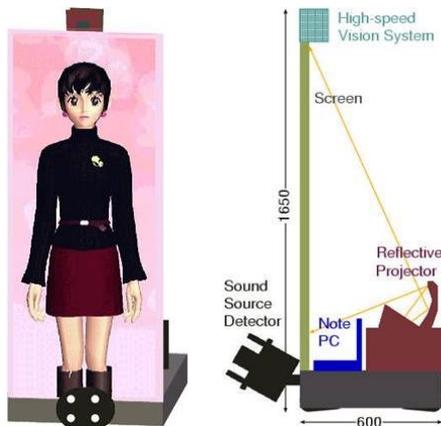


図 1: 案内エージェントロボット構想

スプレイが自走機能を持ち、ユーザと対話をする。この機器や環境に埋め込まれたマイクロフォンアレイによる音響処理、音源定位、高速視覚機能により、来客を先導したり追従したりする。本グループでは、この統合構想に向けて、設計、必要な要素技術の開発、必要部品の購入とシステムの組み立てなどを継続的に進めている。

#### 4 本グループの本年度の研究成果

本 COE プログラムでは、21 世紀の新しい情報学を構築するためには、人間のまわりに遍在するヒューマノイド、エージェント、ユビキタスデバイスが、人間とともに生き、人間を支える情報システム（実世界情報システム）の構築が必要であると考えている。その中で、実世界情報システムプロジェクトは、その実現を目指して、実世界情報学を展開することを念頭において、人と新しいインタラクションをする知的環境の構築を進めている。

これらについて、各研究ユニット（研究活動単位）からの報告は、本報告書の後の節に掲載する。

##### 音声認識・音声言語対話・音楽・文字認識の研究

音声対話擬人化エージェントの統合処理に関して、全方向移動台車に搭載された投影型ディスプレイにより擬人化エージェントを表示し、ユーザと対話をするシステムを開発している。従来開発してきた Galatea ツールキットの改良と、各種センサ情報との統合を行った。

音源分離、残響や雑音に頑健な音声認識に関しては、擬人化エージェントとの対話のために、実環境における音声認識の性能向上を目指し、残響に対する音響モ



図 2: SmartHeadBoy の概観

デルの適応、音声信号の性質を用いた音源分離信号処理の研究を行った。

音楽信号処理、音楽情報の理解に関しては、音楽演奏情報を対象とした自動採譜のためのリズムとテンポの同時認識、多重音の基本周波数と時刻推定、調の認識、与えられた旋律への自動和声付けや対位法に基づく対旋律の生成などの研究を行った。

また、確率文脈自由文法に基づく手書き数式の認識で高い性能を得た。

##### 音源定位・音源分離に関する研究

音源定位センサシステム “Smart Head Boy” に新たな機能を加え、音源数・音源方向・距離を認識し、動的に音源方向へカメラを向け、顔認証と連携できおるようになった。顔認証側の利点として、話者の顔を画像中心にもってこることが可能で、音環境認識システム側の利点として、話者を特定して、固有の音声辞書を選択・利用できる。

音ユニット部分は、時空間勾配法を用いた音源定位を用い、観測音の総和、空間微分  $(x,y)$ 、時間微分の計 4 信号の相関から、音源方向  $(x,y)$  成分と距離を算出する。1 秒間に 200 回定位し、1 音源に対しては方向と、真正面以外の距離の定位が、等距離 2 音源に対しては平面が定位が可能である。また音声認識・音声合成として、Microsoft Speech SDK を利用した。

画像ユニット部分には、顔検出（東芝製 SmartFace-Pro）を用い、目・鼻・顔位置を特定して顔認証（東芝製 SmartFacePro）を 1:N 認証（不特定ユーザーに対する認証）と 1:1 認証（特定ユーザーに対する認証）を行うことができる。



図 3: 顔の特徴部分の抽出



図 4: 開発したビジョンチップシステム

### 高速ビジョンシステムの研究

1秒間に1000フレームの画像処理能力を持つ高速ビジョンシステムを複数台用いて人間の動きや意図を人間が行動を終了する前にあらかじめ察知するようなセンシングシステムを開発した。人を超える性能を提示するデモンストレーションを実現することで、システムの有効性を目に見える形で示すことを目標とする。

実環境で人とロボットがインタラクトする場合、システムが人間と直接的に触れ合うのでビジョンなどのセンサによって人間の動きを素早く観測し、正確かつ安全にロボットを制御する必要がある。

一方、我々はこれまでに、視覚情報を実時間に取得・処理する視覚情報処理デバイスとして、イメージセンサの画素毎に処理回路を取り付けたビジョンチップおよびその技術を応用した高速ビジョンシステムの開発を行っており、それを用いて従来の視覚センサでは不可能な1msの高速視覚フィードバックを実現している。視覚情報は実世界情報の中でも特に情報量が多く、人間の活動の基本となるものであり、これを実時間に取得・処理することが、情報システムの高度化において重要であるといえる。

本研究では、この1秒間に1000フレームの画像処理能力を持つ高速ビジョンシステムを複数台用いて人間の動きや意図を人間が行動を終了する前にあらかじめ察知するようなセンシングシステムを開発した。

具体的には、視覚によって微少な人間の行動を高速に検出するアルゴリズムおよび視覚特徴量から人間の意図を知るための学習アルゴリズムを開発する。さらに、人と対戦するゲームなどにおいて人を超える性能を提示するデモンストレーションを実現することで、システムの有効性を目に見える形で示すことを目標とする。

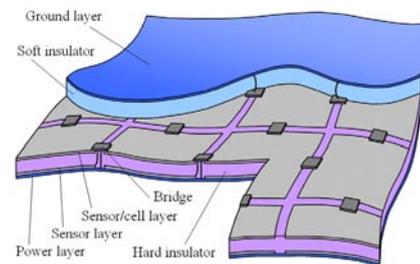


図 5: セルブリッジシステムによる人工皮膚の構造図

### 二次元通信による柔軟体インターフェースの開発

多数の素子を様々な素材の表面に集積するための基礎技術として二次元信号伝送技術を開発している。配線など1次元の媒体による通信、無線など3次元空間を媒体とする通信に対し、二次元面内を無指向伝播する電磁エネルギーによって信号を伝送する。本年度はそのアプリケーションとして、柔軟体に多数の筋電計測用電極を配置するリストバンド型インターフェースを検討した。また類似のコンセプトをもつ技術としてセルブリッジシステムを提案し、それを用いた柔軟な触覚センサシートの開発を行なった。触覚素子となる導電性サイトの境界に容量計測回路を実装した信号転送素子を接続し、計測したデータをマルチホップ伝送する。各素子は接触力と接触面積を計測する。(図5, 6参照)

### 音声言語情報処理における韻律の処理の研究

音声言語情報処理における韻律の処理に重点を置いて研究を進め、以下のような成果を達成した。

生成過程モデルに基づく基本周波数パターンのコーパス生成: 生成過程モデルの指令をコーパスベ-

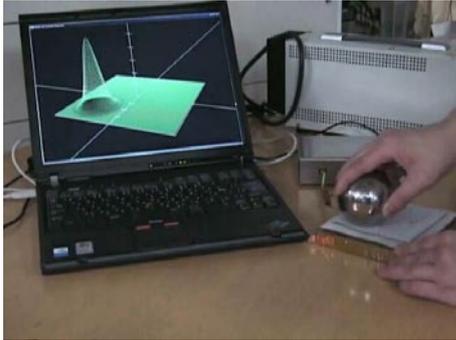


図 6: セルブリッジシステムによる人工皮膚の試作システムの評価実験の様子

手法で推定し、基本周波数パターンを合成する手法を構築し、感情音声の合成を行った。特に、文節単位での感情のレベルを考慮することで、より良好な感情表現が可能になることを合成音声の評価実験を行って示した。現在は、声優等のプロの感情表現を捉え、それを任意の話者で実現する手法の開発を進めている。

自発発話音声認識におけるフィルターの検出：2パス構成の音声認識エンジンである Julius の第2パスでのフィルター仮説を韻律の観点から検証し、仮説が正しいと考えられる場合には、言語尤度を増加する手法を開発した。自発発話音声コーパスについて実験を行い、フィルターでないと誤って判定された結果を訂正する一方、フィルターであると正しく判定された結果については悪影響を与えないことを示した。

道案内システムにおける概念音声合成の実現：応答内容を連文節単位の LISP 表現とすることにより、応答文の柔軟性の高い手法を開発した。LISP 表現で統語構造、焦点情報を保持するため、音声合成で、それを適切に反映した韻律制御が可能となる。道案内システムを構築して、効果を実証した。

この他、2段階手法による標準中国語基本周波数パターンの生成、ニューラルネットワークと音調核モデルを用いた標準中国語の音調型識別について成果をあげた。

信号処理に関連して、劣条件下における複数音源の分離手法を開発した。Hilbert スペクトル、Empirical Mode Decomposition の導入により従来手法よりも高性能を達成した。特にステレオ混合音源からの音源分離に関しは、音源が動く場合にも有効な手法を開発した。



図 7: BCI の実験風景

### 計測信号処理の研究

複雑理工学専攻真溪研究室では計測信号処理の研究を行っているが、具体的には、Brain-Computer Interface(BCI)、言語処理に関する脳機能計測、注意・集中に関する脳機能計測を行った。図7にその実験風景を示す。

### 身体機能を支援する情報バリアフリー機器の開発

「読む」「聞く」「話す」というコミュニケーションに関わる身体機能に障害のある人たちの日常生活を支援する情報バリアフリー機器の研究開発に取り組んでいる。具体的な研究対象は、(1) 視覚障害がある場合でも晴眼者のように Web 情報の斜め読み操作をパソコンで素早くできるように支援する「触覚ジョグダイアル (TAJODA) インタフェース」、(2) 手話通訳によるコミュニケーションが得意でない難聴者たちにも会議や講義での情報伝達を文字で保障するための「リアルタイム音声字幕システム」、(3) 喉頭癌などにより声帯を失った人たちの発声を支援する電気人工喉頭のユーザビリティ向上を目指した「ハンズフリー型人工喉頭」である。今年度はこれらのシステムの試作研究およびユーザ評価を実施し、実用化および製品化に向けた成果を得た。

## 5 本グループのその他の活動

以上の他の本グループのCOE活動として、年間を通じて大学院演習科目「実世界情報システム講究」を実施した。