

# 実世界情報システムプロジェクト ～ 視聴覚研究グループ ～

グループ主査: 嵯峨山茂樹

情報理工学系研究科システム情報学専攻

## 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

本グループでは、視聴覚情報などの人間情報処理を中心に、信号処理、パターン情報処理、画像処理、音響処理などの分野で先進的な研究を進めることを目的としている。

本グループは「視聴覚グループ」と称して研究教育活動を行なうが、これは略称・仮称であり、以後に述べるように広く感覚量や物理量を情報処理の対象として捉えて、技術開発を行なうものである。英語名称を Neo-Cybernetics Group と称する。これは、N. Wiener によって提唱されたサイバネティクスに対し、現代では Cyberspace などの語のように cyber- に仮想空間や仮想現実などの意味も付与されていることから、より情報論的な新サイバネティクスと言う意味で Neo-Cybernetics (NC) と称することにした。研究内容は多岐に渡っており、人間の視聴触覚の知覚と認識から、行動制御、感覚行動統合までを含む。

人間は、実世界の情報に囲まれてその中で生きている。いわゆる五感と呼ばれる視覚・聴覚・味

覚・嗅覚・触覚や温感、湿感、力覚などの各情報を取り込み、認識理解し、それに基づいて思考行動する。これに関して、研究の方向は二つに分かれる。一方は、このような人間と同様に実世界情報を知覚・認識・理解する能力を機械に与えることである。もう一方は、このような情報チャネルと能力を持つ人間に対して機械から情報を呈示し、あるいは機械自ら行動制御することである。これらにより、人間と対話する機械や環境、人間に代わって仕事をする機械、人間のための生活空間で自律的に機能する機械などが生み出される。

実世界情報処理プロジェクトでは、未来の人間の生活および作業環境において、人間と対話・協調する機械・環境を提供する、そのための技術開発研究を目標の一つとして目指している。その中で、上記のような物理量・感覚量のセンシング、認識と理解、制御と呈示などは、このような機械の将来像において極めて重要な意味を持つ。

## 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、7名のメンバーが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。

- 安藤 繁 教授 (システム情報学専攻)  
視覚, 聴覚情報のセンシング
- 石川 正俊 教授 (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
- 嵯峨山 茂樹 教授 (システム情報学専攻)  
音声認識・対話, 音楽情報処理, 文字認識

- 篠田 裕之 助教授 (システム情報学専攻)  
触覚センサ、二次元通信
- 橋本 浩一 助教授 (システム情報学専攻)  
ビジョンサーボ、ロボットハンド
- 眞溪 歩 助教授 (新領域創成科学研究科 複雑理工学専攻)  
多チャンネル信号処理
- 峯松 信明 助教授 (電子情報工学専攻)  
音声情報処理、音声合成

この他に、RA(Research Assistant) 1名:

- 栗原 徹 (システム情報学専攻 博士課程学生)

が、本グループで活動した。

### 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にしている。個々の技術の組合せから生じるシステム構想は数多くあるが、以下にその一例を示す。

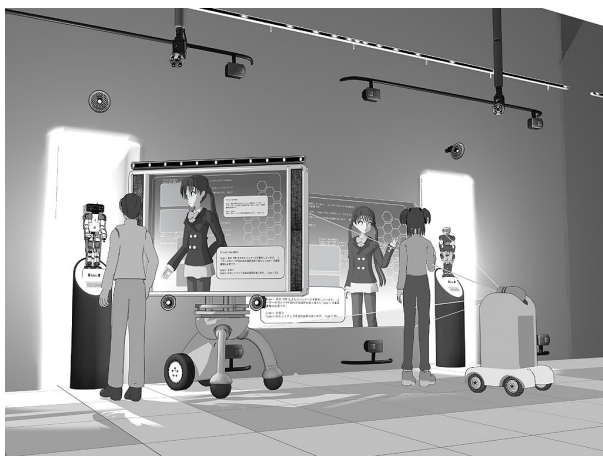


図 1: 案内ロボット構想 (融合システム例)

上の図に示すのは、たとえば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、構想の議論、実現法の検討、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開始などを進めた。

### 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、本報告書の後野節に掲載する。

音声認識・音声言語対話・音楽の研究

- (1) 音声対話擬人化エージェントの身体動作の精密化

擬人化エージェントの表現力向上を目指して、従来開発してきたツールキット Galatea の各モジュールと協調して動作し、表情・個性・身体動作の表現が行なえるエージェント表示システムを構築した。また、エージェントの心的状態を視線移動などの挙動に結びつける力学モデルの構築を目指して、基本的な定式化と予備的な検討を行った。

- (2) 残響や雑音に頑健な音声認識に関する研究

擬人化エージェントとの対話を前提とし、特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デルの適応や、マイクロホンアレイを用いた信号処理の高度化に取り組んだ。

### (3) 音楽情報の理解と作成支援

人間が実際に行った演奏情報を対象とした自動採譜のためのリズム認識、人間が音楽として鑑賞している多重音の音響信号を対象とした採譜支援のための基本周波数の分析、旋律から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。

## 音源定位・音源分離に関する研究

本グループでは、実世界で活用可能な音源定位センサと音源分離システムの実現を目指している。

### (1) 四耳型音源定位システムの開発

新たに全デジタル処理により、5msの時間窓で近接4点音場の時空間勾配を観測し、波面の法線方向から音源の方位を、法線方向での波の振幅減衰率から音源までの距離を検出するシステムを構築した。複数音源に対しては、音場の時空間勾配に関する確率モデルに従い、パルゼン推定とノンパラメトリックカルマンフィルタによって経時的に最も明瞭な音源を選択できる。

### (2) 音源分離システムの開発

複数の音源の中から着目する1個の音源と排除したい1個の音源を選択し、それらの方向と距離の情報を与えて、一方を他方から分離する最適フィルタを設計する原理を考案した。本方法は、環境の変化に適応しやすく、上位システムによる機動的な音源選択が可能な実用的なシステムを可能にする。また分離された音の周波数特性が広い範囲でフラットという利点をもつため、後段の音声認識の性能向上も可能にする。現在は、数値実験による諸性能の確認と、オフラインの音源定位・音源分離実験の段階にある。

### (3) マイクロ音源定位センサの開発

ヤドリバエの鼓膜を模倣することにより、従来のマイクロフォンアレイ技術では困難であったマイクロ音源定位センサの実現を進めている。本年度は、Si表面微細加工を利用し、直径1~3mmのセンサを試作し、ジンバル構造体の形成と基本的な感度を確認した。

## 触覚情報処理の研究

### (1) 触覚センサと通信素子の開発

皮膚感覚のセンサと、その感覚を忠実に再現できる提示装置の開発を軸として研究を進めている。本年度は、人工皮膚に埋め込まれた触覚素子から信号を読み出すために、2次元通信層を介して信号を伝達する技術を提案した。通信プロトコルを確立し、通信素子LSIの試作・動作確認までを行った。

### (2) 触覚提示法の提案・開発

これと並行し、マルチプリミティブ触覚提示法を提案した。触覚受容器単体では応力の正負を識別できないことを利用し、吸引圧で手掌部に触感を提示するデバイスを開発した。

## 超高速視覚センサに関する研究

### (1) 超並列ビジョンチップの新しい画素構造と制御構造の開発

高速な視覚処理を実現する超並列ビジョンチップの画素回路のために、新たに周辺画素の連結機能を取り入れたアーキテクチャを開発し、処理の柔軟性を拡大した。また、これに対して最適化された新たな制御アーキテクチャを開発した。

### (2) 新しいビジョンチップに対応した視覚処理システムの実装

(1)で開発した新しいアーキテクチャに基づいた実時間視覚処理システムの実装を行った。ここでは、さまざまな周辺技術とのシステム

統合を視野に入れ、高い拡張性・相互接続性と、実用的な性能・サイズの実現を兼ね備える実装を採用した。

### (3) ソフトウェアによる画素レベル A-D 変換手法の開発

さまざまな環境への柔軟な適応を実現するため、入射光の明るさの A-D 変換をソフトウェアにより制御する手法を開発し、(1) および (2) で開発した視覚処理システム上にこれを実装した。

## 音声合成・音声学の研究

### (1) 音声に内在する音響的普遍構造の解明

音声コミュニケーションに潜むある数学的「からくり」を解き明かした。スペクトル時系列となる音声事象を一端分布系列に変換し、その分布群が成す構造として音声事象を捉えると、その構造は音声コミュニケーションにおいて一切歪まず（回転及びシフトのみによって）話し手から聞き手に達することが数学的に示された。従来の音声情報処理では、話者・収録機器・伝送機器・再生機器・聴取者らに起因する音響特徴の変化を全て「歪み」として捉える方法論が展開されてきたが、提案分析手法は、これら「歪み」を表現する次元を消失させた形で定義される音声の物理表象である。

この音声の音響的普遍構造を利用した発音教育支援の研究も行った。

### (2) 音声のパラ言語・非言語情報の表出と受容

音声の韻律的特徴に着眼し、基本周波数生成過程モデルに基づくパラメータ抽出、感情音声合成に必要な韻律制御を実装した。これらは音声対話システム構築におけるパラ言語・非言語情報の表出には不可欠な要素である。一方これら情報の受容に関しては、音声の音響的普遍構造のサイズ及び構造的歪みに着眼して情報抽出を行なう方法の初期検討を

行なった。その結果、人間の知覚結果と非常に相関の高い情報抽出が可能であることが示された。

## 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活動として主なものを以下に挙げる。

- 年間を通じて大学院演習科目「実世界情報システム講究」の実施
- 9月8,9日の国際シンポジウムにおけるセッション「ネオ・サイバネティクス」の講演:

- ”Anthropomorphic Agent as an Integrating Platform of Audio-Visual Information” (嵯峨山 茂樹)
- ”Localizing Sound Sources from Spatio-Temporal Gradients” (安藤 繁)
- ”Sensor Fusion Architecture for High-Speed Visual Servo” (橋本 浩一)

および見学会対応:

- 嵯峨山・守谷研究室：音声対話擬人化エージェント、音楽情報処理
- 安藤・篠田研究室：生物模倣型音源定位センサと視覚・聴覚・触覚センサ
- 石川・橋本研究室：高速ビジョンと高速グラスピングシステム

## 6 結び

昨年度に開始された21世紀COE「情報科学技術戦略コア」のもとで実世界情報システムプロジェクト中の一グループとして活動した「視聴覚研究グループ」の状況を報告した。いままでの研究の蓄積の上に各種の進展を見ることができた。今後は、導入設備を生かし、大きな統合目標に向かって一層の研究を進め、融合研究を通して個々の実世界情報処理技術を身近に役立つ先進技術として提供する方向性を打ち出したい。