

「実世界情報システムプロジェクト」

—音響研究グループ システム情報第三研究室—

安藤 繁, 篠田裕之, 小野順貴, 栗原 徹

実環境・実時間音環境認識とその応用

当グループのテーマは、実環境で十分に実用に供し得る性能をもった音源定位と音源分離システムの実現、およびそのヒューマンインターフェースへの適用にある。これまで、大きく分けて二つの方向での研究開発を進めており、ひとつはヤドリバエの鼓膜を模倣した新たなMEMS音響センシングデバイスの開発であり、もうひとつは視覚センサと聴覚センサを任意に顔面と視線の向きを変えられる頭部に搭載したセンサアクチュエータ統合型のヒューマンインターフェースセンサ“SmartHead”の開発である。

具体的テーマと本年度の進展

前者については、本年度は、豊田中央研究所との共同によるMEMSデバイス開発を進め、量産化可能な最適な半導体微細加工プロセスの基礎データに関して大きな成果を得た（東京大学との共同特許出願中）。しかし、最終段階での円盤型振動板のリリースで十分な平面度をもったチップを完成するには至っていない。この点の詰めが最終年度である来年度の主要なテーマとなる。このセンサを実環境で使用する場合、十分な音源定位能力と分離能力を長期にわたって安定して確保するために、センサに動作時の自律的な校正能力を埋め込むことが重要となる。この点について、数学的な方法論として広範で柔軟で安定した収束が保証される凸射影法を導入する研究を進め、理論解析と基礎実験で大きな進展を得た。このアルゴリズムの実際の音響センサへの実装が来年度のテーマとなる。

後者の“SmartHead”センサの開発については、従来の初期知覚レベル機能の実現として得られていた音環境認識や視覚的環境認識とその運動機能との統合に加えて、新たに高次知覚機能との統合として顔認識と音声認識の機能、並びに周囲のセンサや知識ベースとのセンサネットワーク機能を実装し、上記の初期知覚機能と密に結合する新たなアーキテクチャを提案し基礎実験を行うことに成功した。未知な実環境で動作するシステムでは、対象に関する知識をどれだけよく獲得できるかが、その高度で柔軟な機能の実現の鍵となる。このために、

ID情報の付与と獲得が重要であることは近年非常に注目されることとなり、ユニークID、ユビキタスID、RF-ID、携帯電話のID機能の活用などの研究が盛んに行われている。本研究で顔認識をセンサに導入したのは、顔の認識は、それが十分な性能で行われるならば対象人物のIDの獲得に他ならず、これがセンサの初期レベルの処理にも対象に関する知識の導入として有効に機能するに相違ないと考えたからである。以下、これらの研究について具体的成果の概要を紹介する。



(a) SmartHead



(b) 顔発見追従と顔認識

図1. “SmartHead”センサ。正方形の頂点上に4つのマイクロフォンが配置され、中央には2台のカメラと、1台のUSBカメラが搭載されている。可動基台にマウントされており、音源定位と動体検出により上下左右に首を振ることが可能となっている。また両眼の輻輳機能と微分両眼視法により注視対象の立体感を得る。

顔認証をキーとするヒューマンインターフェースセンサ
今回開発したシステムは、従来より本研究室で開発されてきた“SmartHead”に、顔検出認証機能、肌色検出機能、音源分離機能、音声認識機能を追加し、それらを

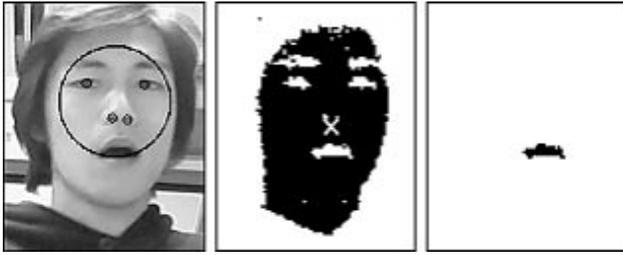


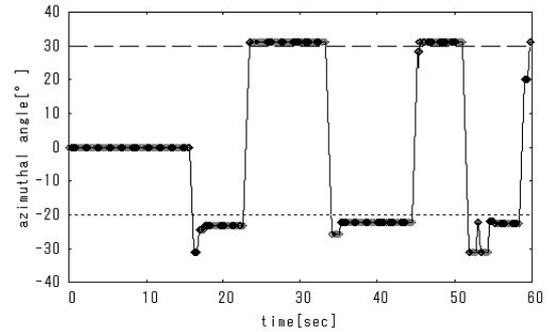
図 2. 顔認識と同時に行われる口検出。口の動きがマイクロフォンで定位された音の強弱と相関をもつかが話者特定の情報となる。

連携させたものである。これらの機能は、関連するセンサに応じて4つのセクションに分類することができる。そのセクションとは、1) 音源定位機能と音声認識機能を持つ音セクション、2) 顔検出・認証機能を持つ画像セクション、3) 顔方向を制御するモーターセクション、4) 任意の環境において、上記の3セクションを自律的なセンサとして連携させるための制御セクションである。音セクションは、概して人間の口や目に相当し、画像セクションは目に相当、モーターセクションは頭・首に相当する。これらを統括する制御セクションは、人間の知能に相当する。

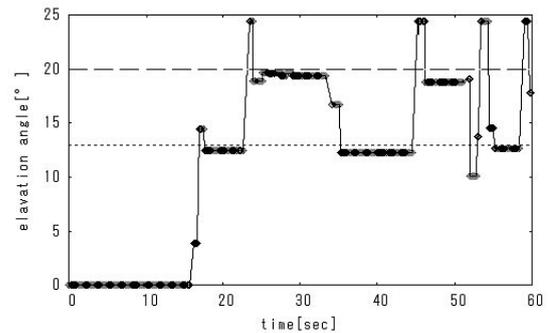
画像セクション — 顔検出追従センサと顔認証センサから成る。顔検出・追従センサでは、顔検出(東芝製 SmartFacePro)による顔の中心位置・瞳位置・鼻孔位置と、肌色検出による肌色領域の重心・面積を求める。

音セクション — 音源定位センサ、音源分離、音声認識・合成から成る。音源定位センサでは、時空間勾配法を用いて1秒間に200回の音源定位を行うことができる。音声認識・合成では Microsoft Speech SDK を用いる。

センサ間の連携 — 肌色と音により顔の検出・定位を行なうことで、対象者の顔を画像中心に持ってこれるという利点を得られ、音声認識をするに当たっては個人を特定することで音声認識辞書という個人固有のモデルを利用できる利点を得られる。口・口面積検出を行なうことで、画像内に映る複数の顔から、話者の顔を特定することができる。カメラ-顔間距離検出を行なうことで、顔の三次元位置を特定することが可能となる。また距離検出中に得られる情報と個人固有の顔モデルを用いることで、視線方向を計測することができる。顔の三次元位置、顔の向き、視線方向を合わせることで、人間の注目する方向を知ることができ、対話する人間の応答をセンサの環境把握機能の一部として取り込むことが可能となる。また、これらによって得られた正確な位置情報を用いて音源分離を行なうことによって、2話者が同時に話しかけてきた場合や定常な雑音が存在する場合、



(a) 方位角の時間経緯



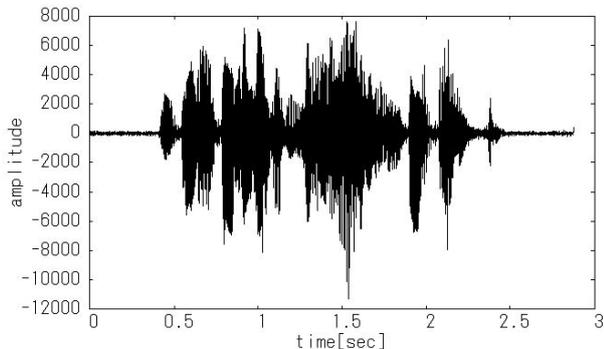
(b) 上下角の時間経緯



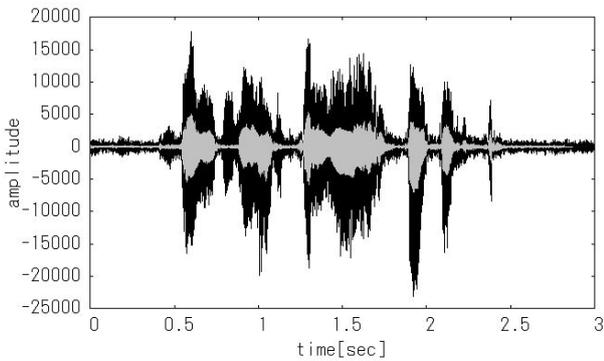
(c) 対象人物 1

(d) 対象人物 2

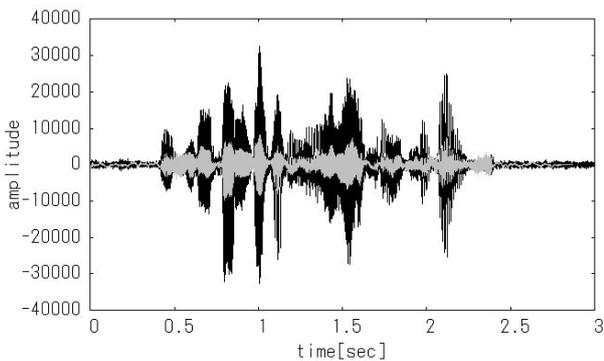
図 3. “SmartHead” の前に立った二人に対する顔と口の交互の定位。最初に一人目を発見して定位した顔認識した後、二人目が現れてそちらに顔を向けて顔認識を行った例。その後は話しかけるのに応じて、話者の方向に顔の向き変えている。



(a) 混合音声信号の例



(b) 音源 1 の原音と分離音



(c) 音源 2 の原音と分離音

図 4. 顔認証と口検出を用いた 2 人の話者の音源分離の例。音源 1 は男性話者、音源 2 は女性話者。今後、マイクホン数を増やし周波数特性の動作時自律校正を十分に行うことにより、格段の性能向上が見込まれる。

目的音声を抽出することが可能となる。

図 2, 図 3, 図 4 に通常のオフィス環境（武田先端知ビル COE 研究室）における実験例を示す。

二次元通信による柔軟体インタフェースの開発

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。その応用として、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った。

一つの機器やシステムに組み込まれるセンサの数は、MEMS 技術の進展も伴いながら著しく増大してきている。例えば標準的な自動車内にはすでに 100 個以上のセンサが組み込まれているが、安全性、快適性を含む自動車の性能向上を追及する限り、より多くのセンサを実装しようとする傾向は今後も続くと考えられる。MEMS 技術によって、個々のセンサを小さく安価に製造できるようになってきた昨今において、センシングにおける次なる重要課題は、乗り物の内装・外装をはじめ、部屋や什器類、着衣など、種々の素材に微小なセンサを実装し、有機的に結合させる手法を確立することであると考える。このような多くのセンサを結合する手段として現在有力視されているのは無線技術であろう。センサや回路配線を印刷工程の中で一体成型する技術なども有力な方向性ではあるが、様々な対象物に対して共通して適用可能な実装技術として、多くの技術者が想定し始めているのは無線技術であると思われる。しかし法的規制を含む帯域確保の問題、各素子への電源供給の問題等があり、例えば「一つの部屋に 1,000 個の素子を実装し、高速に動作させる」実装を実現するのは容易ではないと考えられる。

上記のような状況に対して我々が提案する通信物理層が、「二次元通信」である。ケーブルなど 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送る。具体的には種々の柔軟材料・低コスト素材で作製可能な二次元通信層を、壁、床、机などの表面や、衣服などの布状シートとして作りこんでおく。その表面に専用のインタフェースを近接させると、電磁波は通信シートの内部および近傍を伝播し、通信シートに近接している他の通信素子と通信をする、というものである。空中に放射するアンテナよりも小型のインタフェースが実現可能であり、空中に放射される電磁波はわずかであるため、マイクロ波による数 10 mW オーダーの電力供給も容易に達成できる。このような信号伝

送技術は、ユビキタス通信媒体として幅広く利用可能なものであるが、本年度は特にそれを柔軟体への高密度センサ素子実装に応用した。

REFERENCES

- [1] 王, 小野, 安藤, “咀嚼動作を模倣した食感評価音響センサシステム,” 電気学会全国大会講演論文集, pp.195, 2004.
- [2] 齋藤, 小野, 安藤, “音源位置情報を利用した音源分離の理論と実験,” 電気学会全国大会講演論文集, pp.196-197, 2004.
- [3] 王, 小野, 安藤, 佐野, “音による食感定量化のためのセンサシステムに関する研究,” 日本音響学会秋季研究発表会, vol.I, pp.493-494, 2004.
- [4] 小野, 岡田, 安藤, “正則フィルタバンクを用いた音声信号の時間周波数分析,” 日本音響学会秋季研究発表会, vol.I, pp.529-530, 2004.
- [5] 小野, 藤田, 千條, 安藤, “時空間勾配法に基づく実時間3次元音源定位センサ,” 日本音響学会秋季研究発表会, vol.I, pp.607-608, 2004.
- [6] 千條, 藤田, 小野, 安藤, “音源位置情報に基づく最適音源分離システム,” 日本音響学会秋季研究発表会, vol.I, pp.627-628, 2004.
- [7] 王, 小野, 安藤, 佐野, “咀嚼動作を模倣した食感評価音響センサシステム,” 日本食品科学工学会第51回大会講演集, pp.65, 2004.
- [8] N. Ono, T. Shimizu, T. Kurihara, and S. Ando, “Real-time 3D Imager Based Using Spatio-Temporal Phase Unwrapping,” Technical Digest of the 21st Sensor Symposium, pp.437-440, Kyoto, 2004.
- [9] T. Shimizu, T. Kurihara, N. Ono, K. Kitagawa, and S. Ando, “Real-time En-face Optical Coherence Tomography Using Correlation Image Sensor,” Proc. 21th Sensor Symp., pp.441-444, Tokyo, 2004.
- [10] 小野, 牧野, 王, 佐野, 安藤, “咀嚼動作を模倣した食感評価音響センサシステム,” フードサイエンス, 2005.
- [11] Y. Nagai, T. Kurihara, N. Ono, and S. Ando, “Multipole marker and correlation image sensor for position and pose estimation in sensor networks,” Proc. Int. Conf. Solid State Sensors, Actuators, and Microsystems (Transducers'05), pp.474-478, Seoul, 2005.
- [12] Y. Nagai, T. Kurihara, N. Ono, and S. Ando, “Sensing and networking with correlation image sensor, modulated light source, and multipole polarizing reflector,” Proc. Int. Workshop Networked Sensing Systems (INSS 2005), pp.58-63, San Diego, 2005.
- [13] S. Ando and N. Ono, “Electric Circuit Analogy and Optimal Algorithm of the Cooperative Calibration and Synchronization Problem in Sensor Networks,” Proc. Int. Workshop Networked Sensing Systems (INSS 2005), pp.204-209, San Diego, 2005.
- [14] 安藤, 来海, 小野, 栗原, “時間相関イメージセンサを用いる新しい実時間映像化技術,” ケミカルエンジニアリング, vol.50, no.7, pp.552-559, 2005.
- [15] Akimasa Okada, Yasutoshi Makino and Hiroyuki Shinoda, “Cell Bridge: A Signal Transmission Element for Constructing High Density Sensor Networks,” Proc. of INSS2005 (Second Int. Workshop on Networked Sensing Systems), San Diego, pp.180-185, June, 2005.
- [16] Takayuki Hoshi and Hiroyuki Shinoda, “Tactile Sensing Using Nonlinear Elasticity,” Proc. SICE Annual Conf. 2005, pp. 2978-2981, 2005.
- [17] Yasutoshi Makino, Naoya Asamura and Hiroyuki Shinoda, “Multi Primitive Tactile Display Based on Suction Pressure Control,” Proc. IEEE 12th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, (Haptic Symposium 2004), pp. 90-96, 2004.
- [18] 安藤, 田村, 戸辺, 南, “センサネットワーク技術,” 東京電機大学出版局 (2005)
- [19] 小野, 安藤, “食品のテクスチャー評価におけるシステムインテグレーション,” 計測と制御, vol.44, 2005.
- [20] 安藤 繁, “ネットワークセンシングの現状と展望 第1回センサのネットワーク化の目的・技術要素・発展方向・研究課題,” 計測と制御, vol.43, no.7, 2004.