

## 人位置・姿勢検出と音声対話によるユーザ対移動ロボット間 インタラクション・システムの試作開発

○佐々木 智典<sup>\*1)</sup>

### ■キーワード 人検出、点群データ処理、音声認識、音声合成

1. 運動と音声による人間-移動ロボット間インタラクション実施について試作
2. 並行動作するソフトウェアモジュールの RT ミドルウェアによる統合
3. 移動ロボットによる人間の追従運動タスクの実施

### ■はじめに

人間同士がコミュニケーションを行うときには、視覚、視覚刺激（ジェスチャ動作、歩行運動）、聴覚、聴覚刺激（音響、音声言語）が同時並行に活用されている。人間とロボットの間でも同様にコミュニケーションを行えば、協調作業の実施に有益である。本稿では、移動ロボットが人間の後ろに追従する運動タスクを題材として試作開発したインタラクション・システムについて述べる。

### ■取り組み内容

インタラクション・システムは、所定のシナリオに従うコミュニケーションを前提としている。シナリオは初期状態、追従状態等の場面に対応する状態により構成される。シナリオを実現する状態遷移はユーザの運動（歩行、ジェスチャ）、発話に応じて進行する。図1にインタラクション・システムの構成を示す。図中の各ブロックはハード/ソフトウェアのモジュールに対応し、並行動作する。各モジュール間の通信には産業技術総合研究所において開発された RT ミドルウェア<sup>[1]</sup> を利用している。

音声認識、音声合成には情報通信研究機構において開発された rospeech<sup>[2]</sup> を利用している。rospeech は、音声認識の前処理として必要な発話区間検出・雑音抑制信号処理機能も提供する。音声認識及び音声合成は、インターネット上のサーバにおいて実行されるが、ユーザプログラム側は rospeech のインタフェースを利用することにより、直接通信の管理を行わずに処理結果を利用できる。rospeech 内部ではサーバとのデータの送受信に HTTP (Hyper Text Transfer Protocol) 及び JSON (JavaScript Object Notation) が使われる。

rospeech は RTM と異なるミドルウェアである ROS を利用している。このため、両者の接続ができるように、ROS 及び RTM のコンポーネントの両方として機能するモジュールを構築した。両者の処理ループは別々のスレッドにおいて実行され、両者の間ではキューを介してデータの受け渡しを行う。

ユーザの運動の検出は、深度画像センサによって行われる。深度画像センサ (ASUS Xtion を利用) は周辺物体上の 3 次元点群データを取得する。取得された点群データに基づき人間に相当する点群の抽出、人体位置・姿勢の推定が行われる。ロボット近傍のユーザ候補点群及びシナリオ上の質問への回答から、追従対象が決定される。追従対象の位置を基に追従のための運動計画が行われ、速度指令が生成される。この速度指令により車輪ロボットベースの制御が行われる。

### ■今後の課題

現状のシステムでは音源が単一であり、雑音も比較的小さいことを前提としているが、現実の環境はより複雑である。人間同士が会話するような距離・状況下でコミュニケーションを行うには、音源位置推定やロボットからの発話音の考慮などの音環境理解<sup>[3]</sup>、発話の途中に割り込み話し出すような状況への対応が必要となる。今後、これらと点群データに基づくユーザ位置推定との統合を検討する。

### 参考文献

- [1] 安藤慶昭, 神徳徹雄, 末廣尚士, 計測自動制御学会誌, Vol.52, pp.778-781 (2013)  
 [2] 杉浦孔明, 堀智織, 是津耕司, 信学技報, Vol.113, pp.7-10 (2013), <http://rospeech.org/top-ja/>  
 [3] 奥乃博, 中臺一博, 水本武志, 電子情報通信学会誌, Vol.95, pp.401-404 (2012)

\*1) ロボット開発セクター

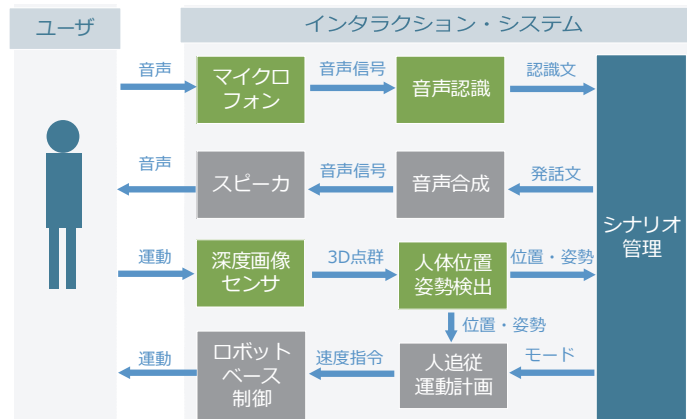


図1. インタラクション・システムの構成