

小型音声対話モジュールの開発

佐藤 幹 岩沢 透 杉山 昭彦 高野 陽介 *藤田 善弘

NEC 共通基盤ソフトウェア研究所, *NEC 企業ソリューション企画本部

小型音声対話モジュールの開発について報告する。本モジュールは、実環境における音声対話に必要な機能として、マイクの指向性を人の発話した方向に向けるための音源方向検出機能、マイク入力信号に混入する周囲雑音成分の消去を行う雑音消去機能、入力音声を確認するための音声認識機能、認識結果等をユーザに伝えるための音声合成機能を備える。NECにおいて研究開発中のパーソナルロボット PaPeRo に搭載されるこれらの機能を、携帯電話用アプリケーションプロセッサを利用することにより、様々な機器への組み込みを容易とする小型モジュール上で実現している。

Implementation of a Compact Speech Dialogue Module

Miki SATO, Toru IWASAWA, Akihiko SUGIYAMA, Yosuke TAKANO, *Yoshihiro FUJITA

NEC Common Platform Software Research Laboratories, *NEC Enterprise Solutions Planning Division

This paper presents development of a compact speech dialogue module. This module consists of direction of arrival estimation (DOA), noise cancellation (NC), speech recognition (SR), and text-to-speech conversion (TTS). DOA is essential to adjust the microphone direction toward the speaker. NC is useful to reduce undesirable influence by ambient noise and interference. SR recognizes the input speech and TTS conveys the recognition result to the user. These functions, equipped with in PaPeRo, are implemented on an application processor that was developed primarily for mobile phones.

1. はじめに

近年、ロボットやカーナビ等の様々な機器のインターフェースとして、音声対話機能が注目されている。実環境でスムーズな音声対話を実現するために、様々な雑音が存在する環境に対処できる耐雑音性能が求められる。耐雑音性能を向上させるために、指向性マイクが広く用いられる。マイクの指向性を音声の到来方向に一致させるために、音源方向を推定することが重要である[1]。マイクの指向性だけで抑圧できない雑音や妨害信号は、雑音消去処理によって影響を軽減することが有効である[2]。

一方、実環境での音声対話の困難さは、音声対話機能を搭載する端末の開発にとって障害となっている。そこで、実環境での音声対話に必要な機能を小型・低消費電力・低コストの組込用モジュールとして提供することにより、開発の困難度を低減することができる。

我々は、パーソナルロボット PaPeRo[3]を用いて、実環境を対象とした音声対話技術の研究開発を行っている。この技術を踏まえ、かつ小型・低消費電力・低コストを達成するために、携帯電話用アプリケーションプロセッサを利用して、PaPeRo における音声対話機能と同等の機能を搭載する音声対話モジュールの開発を行った。本稿では、音声対話モジュールの機能、ハードウェア、及びその実装形態について報告する。

2. 音声対話モジュールの機能

音声対話モジュールの全体構成を図1に示す。本モジュールは、実環境における音声対話の基本機能として、音源方向検出機能、雑音消去機能、音声認識機能、音声合成機能を有する。これらの機能は、RT (robot technology) ミドルウェアに対応した RT コンポーネントとして動作し[4]、ネットワーク接続された外部アプリケーションから利用することが可能である。

音源方向検出機能は、3または4つのマイクの入力信号を用いて、音源方向の推定を行う。本機能は、近

接音場を想定した方向推定[1]を行うことにより、音源とマイクが同一平面上になく十分に距離が離れていない場合も、高精度な方向検出を可能とする。

雑音消去機能は、音声用及び雑音用の2つのマイクの入力信号を用いて、音声側入力信号に混入する雑音成分を推定することにより、雑音の消去を行う。本機能は、雑音側入力信号に混入する音声成分の推定も行うことにより、低歪の雑音消去が可能である[2]。

音声認識機能は、音声認識辞書に登録した単語を対象とした音声認識処理を行う。音声認識辞書は、SRG S(Speech Recognition Grammar Specification)[5] 準拠のXML形式で記述される。本機能は、不要雑音を認識単語と誤認識させないためのリジェクション機能[6]も備える。

音声合成機能は、入力テキストの読上げを行う。本機能は、組込用途向けのコンパクトかつ高音質な音声合成エンジンにより、合成発話音声の生成を行う。

3. 音声対話モジュールのハードウェア

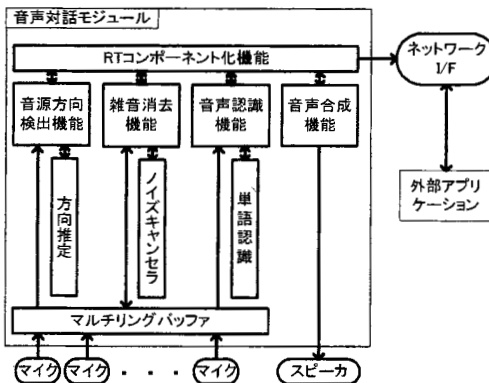


図1 音声対話モジュールの全体構成

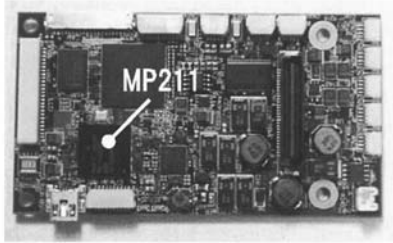


図2 音声対話モジュールの外観

表1 音声対話モジュールの仕様

項目	仕様
CPU	ARM9(192MHz) × 3, DSP(SPXK6 192MHz) × 1
メモリ	128MB, +128MB (拡張ボード使用時) フラッシュ 64MB
音声 I/F	マイク入力 2ch × 2 系統 +16ch (AUDIO ボード使用時) スピーカ出力 2ch
画像 I/F	カメラ入力 × 2 系統、ビデオ出力、LCD 出力
その他 I/F	USB, LAN, IrDA, GPIO CF Card (拡張ボード使用時)
サイズ	55mm × 100mm × 32mm (AUDIO ボード、拡張ボード使用時)

音声対話モジュールの外観を図2に、仕様を表1に示す。本モジュールは、NEC エレクトロニクス社製携帯電話用アプリケーションプロセッサ MP211 (ARM9 × 3, DSP × 1 で構成) を搭載し、Linux OS によって動作する。音声入力インターフェースは、メインボード上に 2 チャンネル同期マイク入力 2 系統を備え、拡張用 AUDIO ボードにより、16 チャンネル同期マイク入力を増設することが可能である。その他には、スピーカ出力 × 2、カメラ入力 × 2、LCD 出力、USB、LAN などの周辺インターフェースを備え、拡張ボードにより、メモ리카ードの使用が可能である。

4. 基本機能の実装形態

音声対話モジュールの基本機能は、MP211 上の ARM9(192MHz) 1 個と DSP(192MHz) 1 個により実現している。ARM-DSP 間の基本機能分割を図3に示す。音声認識機能、音声合成機能、RT コンポーネント化機能は ARM 上で動作する。音源方向検出機能は、ARM 上で動作する音源方向検出制御部と、DSP 上で動作する方向推定コア部から構成され、3 または 4 マイクの入力に対応できる。雑音消去機能は、ARM 上で動作する雑音消去制御部と、DSP 上で動作するノイズキャンセラコア部から構成される。

マイクに入力された信号は、11025kHz でサンプリングされ、ARM 用メモリに配置されたマルチリングバッファに書き込まれる。音源方向検出制御部は、外部から動作開始命令を受けると、マイク数(3 または 4)に対応した入力信号をマルチリングバッファから取り出し、ARM-DSP 間の共有メモリにコピーする。方向推定コア部は、入力信号を共有メモリから取得し、FFT、白色化、符号化を行い、マイクペア毎の相互相関

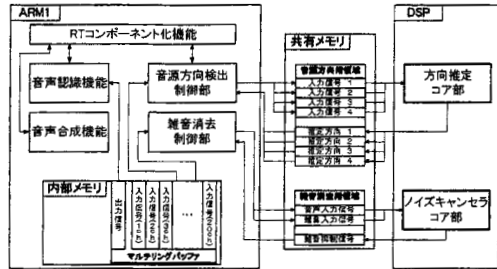


図3 ARM-DSP 間の基本機能分割

から推定方向を求める。マイクペア数に対応した相関値は、共有メモリを介して ARM に伝達される。音源方向検出制御部は、相関値で表された複数の方向推定値を個別に最適化した後、最終方向決定処理[1]によって音源方向を決定する。方向推定コア部の方向推定値計算は 1024 サンプル毎、総演算量は 35MIPS である。

雑音消去制御部は、音声用マイクと雑音用マイクの入力信号を、共有メモリを介してノイズキャンセラコア部に伝達する。ノイズキャンセラコア部は、音声側入力に混入する雑音成分と、雑音側入力に混入する音声成分を、それぞれ個別の適応フィルタで推定し、推定結果を各信号から差し引くことにより雑音消去信号を生成する。雑音成分用と音声成分用の適応フィルタは、それぞれに専用のサブ適応フィルタを用いて推定した信号対雑音比でステップサイズを制御し、大きな消去量と小さな音声歪を両立する。雑音消去信号は、共有メモリ経由で雑音消去制御部に伝達される。ノイズキャンセラコア部と共有メモリ間の入出力信号転送は 512 サンプル毎に行う。総演算量は 78MIPS である。

5. おわりに

実環境で動作する音声対話機能を搭載した小型音声対話モジュールの機能、ハードウェア、及び実装形態について報告した。本モジュールは、現在、NEC を含む 5 機関のロボットに搭載され、機能検証中である。今後は、基本機能の安定性とモジュールの汎用性を向上させ、音声対話インターフェースの普及促進と知的対話の研究推進に貢献する予定である。

本研究の一部は、NEDO の次世代ロボット共通基盤開発プロジェクトの一環として行った。

参考文献

- [1] M. Sato, A. Sugiyama, O. Hoshuyama, N. Yamashita, Y. Fujita, "Near-Field Sound-Source Localization Based on a Signed Binary Code", IEICE Trans. Fundamentals, Vol.E88-A, No.8, pp.2078-2086, Aug. 2005.
- [2] M. Sato, A. Sugiyama, S. Ohnaka, "An Adaptive Noise Canceller with Low Signal-Distortion Based on Variable Stepsize Subfilters for Human-Robot Communication", IEICE Trans. Fundamentals, Vol.E88-A, No.8, pp.2055-2061, Aug. 2005.
- [3] Y. Fujita, "Personal Robot PaPeRo," J. of Robotics and Mechatronics, Vol.14, No.1, pp.60-63, Jan. 2002.
- [4] 高野, 宇田, 石田, 西沢, 藤田, "ロボットプラットフォーム RoboStudio による RT コンポーネントの実現", Proc. RoboMec2007, 1P1-C02, May. 2007.
- [5] <http://www.w3.org/TR/speech-grammar/>.
- [6] 岩沢, "パーソナルロボット PaPeRo の音声認識インターフェース", 人工知能学会 AI チャレンジ研究会, Vol.13, pp.17-23, Jun. 2001.