

ロボット聴覚用オープンソースソフトウェア HARK の展開

中臺 一博 ((株) ホンダ・リサーチ・インスティチュート・ジャパン/
東京工業大学) 奥乃 博 (京都大学)

概要 ロボット聴覚用のオープンソースソフトウェアとして研究開発を行っている HARK (HRI-JP Audition for Robots with Kyoto Univ.) の展開について説明する。HARK は複数のマイクロフォン (マイクロフォンアレイ) からの入力をもとに、音源定位、音源分離、さらに分離音声の認識までをサポートするソフトウェアであり、GUI プログラミング環境上で様々なモジュールを配置・接続することにより、形状やマイクロフォンレイアウトが異なるロボットに対応させたり、用途に合わせたロボット聴覚システムを構築したりすることができる。本稿では、HARK の設計指針を解説し、HARK を用いて構築したシステムの応用例、HARK の展開も併せて報告する。

1. ロボット聴覚ソフトウェアとは

ロボット聴覚は自然な人・ロボットインタラクションを実現するために、ロボット自身に装着されたマイクロフォンを用いて、音を聞き分ける機能を実現することを目指した研究領域である。ロボット聴覚は研究領域としての歴史が浅いこともあり[1]、ビジョン研究における OpenCV のように、ロボット聴覚に必要な機能のセットをオープンソースとしてリリースし、世の中に広く利用してもらおうというアプローチはほとんど見られなかった。

我々は、これまでの研究成果の集大成として、同時発話を聞き分けるロボット聴覚ソフトウェア HARK (HRI-JP Audition for Robots with Kyoto Univ., hark は listen を意味する中世英語) を『聴覚の OpenCV』を目指すべく、2008 年から研究用にオープンソースソフトウェアとして、無償公開¹を開始した[2,3]。また、10 人の訴えを同時に聞き分けたといわれる聖徳太子の逸話になぞらえ『世界中のロボットを聖徳太子に』にすべく展開活動を行っている。これまで、国内外の大学・研究機関・企業から多数ダウンロードがあり、後述(5 章)の例を含め、国内外で HARK を通じた協力関係も構築されている。こうした活動により、着実に HARK のノウハウが蓄積されている。また、要素技術や応用技術の国際学会での継続的な発表や講習会といった情報発信を通じて、国際的な認知・賛同を得た HARK コミュニティが徐々に形成されつつあり、これまでのところ、オープンソース化の恩恵を十分に受けていると考えている。

HARK では、ロボットの耳を実現する上で音環境理解

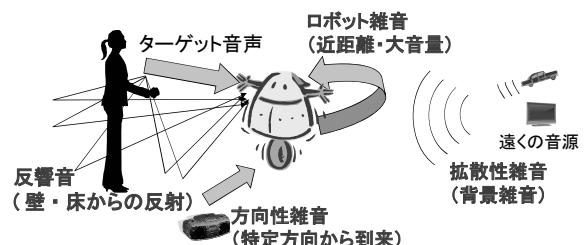


図 1 ロボットを取り巻く 4 種類の雑音

(Computational Auditory Scene Analysis) が重要であると捉え、その主要課題である音源定位 (Sound Source Localization), 音源分離 (Sound Source Separation), 音声強調 (Speech Enhancement), 分離音声の音声認識 (Automatic Speech Recognition) を最低限提供すべき機能として、開発している。

以下、第 2 章で HARK の技術課題、第 3 章で設計思想について概説する。第 4 章で HARK の概要を述べる。第 5 章で HARK の応用、国内外での展開について紹介し、第 6 章で今後の開発予定とまとめを述べる。

2. ロボット聴覚の技術課題

近年の音声認識技術の進展は目覚ましく、少々の雑音下であっても口元にマイクロフォンがあれば、実用的なレベルでの音声認識が達成されるようになってきている。例えば、Google Talk は、取得した音声信号をサーバに送り、音声認識を行い、所望の情報検索を行っている。一方、ロボット聴覚の場合は、ハンズフリー音声認識あるいは遠隔発話音声認識 (Distant-Talking Speech Recognition) と同様に、雑音の影響が顕著になり、音声認識アルゴリズムやモデルによる対応だけでは実用レベルでの性能は望めない。音響信号のパワーが距離の二乗

¹ <http://winnie.kuis.kyoto-u.ac.jp/HARK/>

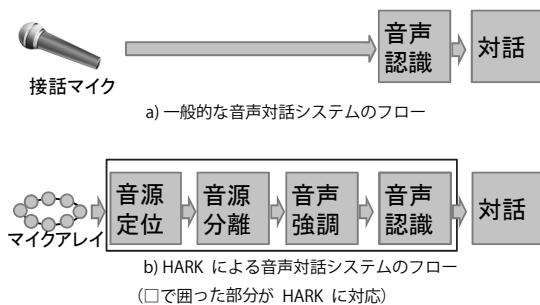


図2 一般的なシステムとHARK の違い

に反比例して減衰する逆二乗則の影響を受けるだけでなく、様々な雑音が混入して、目的音声の信号対雑音(S/N)比が著しく劣化するからである。一般的なロボットの環境では、次の4種類の雑音(図1参照)に対処する必要がある。

1. 方向性雑音：特定の方向から到来する雑音、TVや他の人の声など。
2. 拡散性雑音：暗騒音や遠くから到来する雑音など方向性が失われた雑音、一般に背景雑音。
3. 残響：壁や天井、床などに反射して時間遅れを伴って様々な方向から到来する雑音、上の2つを加法性雑音というのに対し、乗法性雑音ともいう。
4. ロボット雑音：ロボット特有の雑音、ロボットのファンやモータなど、目的音源より近いところから到来する雑音、方向性と拡散性両方の側面を持つ。

最後のロボット自身の雑音という部分が、ハンズフリー音声認識や遠隔発話音声認識にない課題である。この4種類の雑音への対応は、ロボットに限らず、TVなどの情報家電にマイクロフォンが埋め込まれた場合にも対処すべき技術的課題である。

3. HARK の設計思想

3.1 HARK での雑音対策の方針

同時発話や割込み発話に加えて雑音問題を解決するためにHARKではマイクロフォンアレイを導入し、これを用いた音源定位、音源分離、音声強調といった処理を音声認識の前処理としている。このような音声認識の前処理を行うHARKベースの音声対話システムと口元にマイクロフォンがあるような一般的な音声対話システムとの違いを図2に示す。なお、一般的な音声認識システムにも雑音除去や音声強調が組み込まれており、雑音レベルが小さい場合は有効である。

方向性雑音は、音源が空間的にスペースに配置される

という仮定を満たす場合²、方向情報を利用した音源分離により除去できる。例えば、20度以上離れた複数の話者が同時に話しても、同時発話数がマイク数よりも少なければ、それらをすべて分離認識することが理論上は可能となる[2]。拡散性雑音は、方向情報を使用した分離手法では除去できないので、分離音に、さらに音声強調手法を適用して除去する。残響は、時間遅れが20~30ms程度までの初期反射と、それ以上のいわゆる後期残響に分けて考える。前者は音声認識の処理単位であるフレーム(一般的にフレーム長は25ms程度)内に影響が留まるため、各フレームごとに音声認識特徴量を抽出する際に対応できるのに対し、後者は難しい問題であり、HARKでも現時点では、部分的な解決(音声強調時に簡単な時間減衰モデルを導入)しか行っていない。ロボット雑音については、方向性成分を音源分離で、拡散性の部分を音声強調で除去し、さらに音源定位で定位できないことがあるパワーの小さい定常雑音でも除去できるよう特定方向の音源を常に抑圧し続ける処理を音源分離に導入している。

3.2 音響処理アルゴリズムの選択

ロボット聴覚では、音源定位データを基にした音源分離、音声強調をして、分離した音声に対して音声認識を行うことが多い(図2参照)。音源定位、音源分離、音声強調といった処理に対する信号処理アルゴリズムは非常に多くのものが提案されている。処理アルゴリズムは、何らかの前提条件の下で設計・開発されているので、その分野のプロでないと、どのアルゴリズムがよいのかなかなか分からぬ。また、純粹に理論だけに焦点を当てているものもあるので、適材適所での使い方が難しい。

HARKではこれまでに得られた開発者やユーザのノウハウに基づいて最も性能の良かったアルゴリズムのみを選別して提供している。つまり、HARKでは、開発した信号処理アルゴリズムをすべて提供するのではなく、主に後述する応用例・展開例へのHARK適用を通じて得られた最適なアルゴリズムだけを選びすぐって提供する、ある意味で、尖ったシステム構成となっている。このため、音声強調などモジュールによっては、例えば、リリースごとに、チューニングが大変だがピーク性能は高い、逆にチューニングは簡単だがピーク性能が期待したほど上がらないという特徴が現れることがある。

² HARKでは、ロボットから見て2つの音源が20度以上離れている場合、空間的にスペースであるとみなせる。

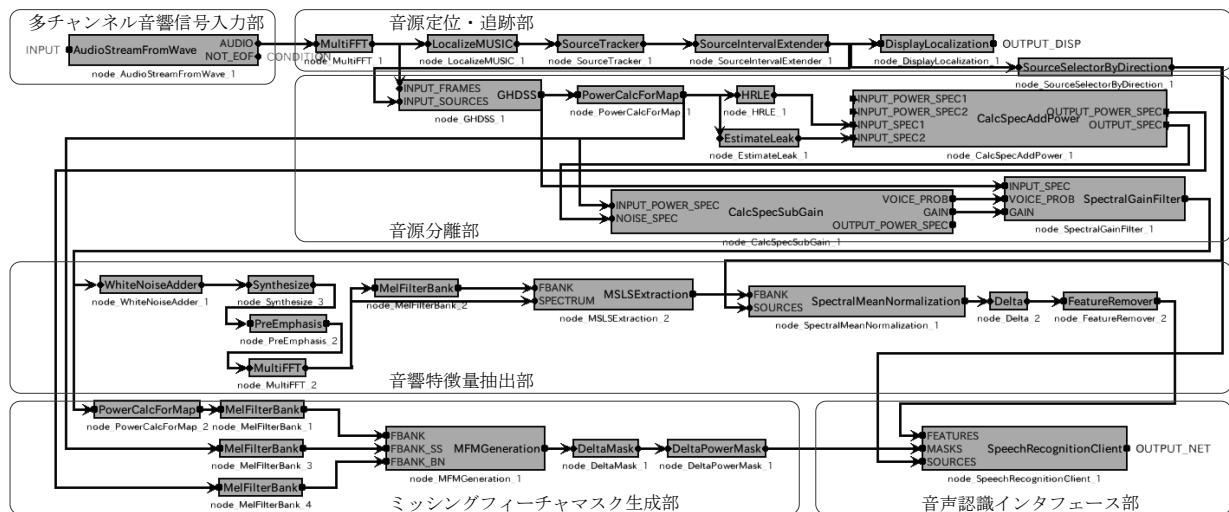


図 3 HARK の GUI プログラミング環境画面（モジュール接続例、モジュールネットワーク）

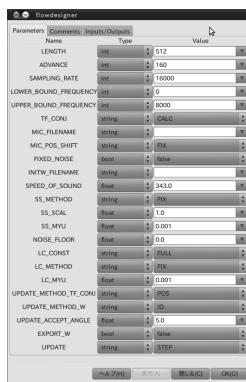


図 4 モジュールプロパティ設定画面例

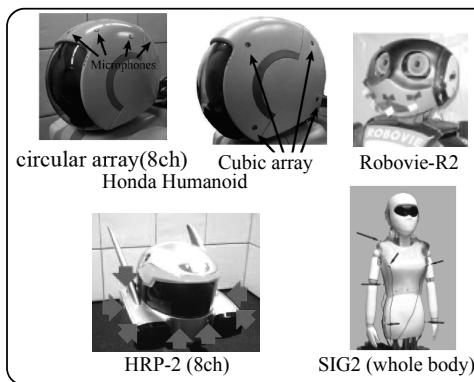


図 5 HARK で検証したロボットとマイクロフォンアレイ



図 6 HARK で利用可能なマルチチャネル A/D

3.3 HARK のシステムとしての設計方針

ロボット聴覚ソフトウェア HARK の設計思想を以下にまとめるとする。

1. GUI ベースのフレキシブルなシステムカスタマイズ機能
2. 入力から音源定位・音源分離・音声強調・音声認識までの機能単位のモジュールとそれらの総合機能の提供
3. マルチチャネル A/D 装置への対応
4. 実時間処理

これらの設計思想の下に、2008 年 4 月に HARK 0.1.7 を研究用にオープンソースとして公開³し、開発者自身での改良、ユーザからのフィードバックの反映、バグフィックス、ドキュメントの充実などを通じて 2009 年 11 月に HARK 1.0.0 プレリリースを 2010 年 11 月に、HARK 1.0.0 確定版リリースを行った。209 ページのマニ

ュアルと 159 ページのクックブックと呼ばれる Q&A 集が公開されており、3 月末には英語版も公開予定である。

HARK は、LinuxOS (Ubuntu 10.04)上で稼働し、ミドルウェアとして Flowdesigner [4]を利用している。現在、Linux 系の OS のみサポートしているが、HARK チュートリアルを国内で 3 回、韓国で行った時のアンケートから、Windows 対応への要求が高いので、現在、Windows 系 OS のサポート計画も進行中である。

4. HARK の概要

4.1 GUI インタフェース

ロボット聴覚では、音源定位、音源分離を中心としたさまざまな処理モジュールを組み合わせて所望のタスクを実現する。この時にモジュールの接続に GUI が使えるように設計をした。具体的には、Flowdesigner というミドルウェアを用いて HARK システムを構築している。図 3 に、Flowdesigner 上での典型的な HARK のモジュール接続例（ファイル入力によりマルチチャネル音響信号を取得、音源定位・音源分離を行い、分離音から生成した音

³ ライセンスとしては商用化もライセンシングで対応可能。

表 1 HARK 提供モジュールリスト

モジュール	
カテゴリ	説明
音声入出力	マイク・ファイルから音を取得・格納など3モジュール
音源定位	音源定位、音源追跡、定位情報の入出力など7モジュール
音源分離	音源分離、音声強調、背景雑音・リーケ推定など8モジュール
音声認識	音声認識特微量抽出・送出・格納、ミッシングフィーチャマスク(MFM)生成、ミッシングフィーチャ理論ベース音声認識など15モジュール
その他	チャネル選択、マルチチャネルFFT、白色雑音追加、ログ出力など13モジュール
非モジュール	
データ生成ツール	データ可視化・各種設定ファイル作成(harktool)

声特徴量と音声特徴量の信頼度に応じて音声特徴量をマスキングするミッシングフィーチャマスク[7]を、音声認識に送出)である。ロボット形状、マイク配置といった各ロボットに固有データは、それらが記述されたコンフィグファイルを作成し、このファイル名を各モジュールの属性値にGUIを用いて設定するだけで、容易に構成変更が可能なよう実装している。オリジナルのFlowdesignerを用いても、HARKの利用は可能であるが、設定属性数が多いモジュールでは、属性設定画面が煩雑になるため、属性設定画面の階層化やリスト選択機能をFlowdesignerに追加した(図4参照)。これに加え、メモリリーク等のバグ対処や他の操作性向上を含めた形で、完全アッパーコンパチブルの改良版Flowdesignerも公開している⁴。

4.2 機能単位のモジュールとそれらの統合機能の提供

ロボットマイクロフォンからの入力、マルチチャネル信号処理による音源定位、音源分離、音声強調、分離音認識といったロボット聴覚の主要機能に対して、機能単位のモジュールを提供することによって、ユーザの利便性を向上させる試みを行っている。例えば、音源定位機能を一つのモジュールとして提供することにより、行列演算、ピークサーチといった処理粒度の小さいモジュールをたくさん組み合わせて音源定位機能を構築する手間を省いている。また、前述したように、同じ機能を持ったモジュールを数多く提供するのではなく、経験に基づき厳選したアルゴリズムのみを提供することにより、モジュールが煩雑にならないよう工夫している。これにより、HARKをツールとして利用するユーザ、ロボット聴覚研究者に利用するユーザの双方にメリットがある。前者は、ロボット聴覚システム構築が簡便になり、構築時のヒューマンエラーが低減できる。また、後者は、機能単位の研究が多く、モジュールを一つ置き換えるだけで、

独自のモジュールを同じ土俵で評価することができる。HARKでどのようなモジュールを提供しているかを表1にまとめる(詳細リストは[2]を参照)。音声入出力は、4.3節で紹介するデバイスに対するI/Oを含むモジュール群である。音源定位は、雑音ロバスト性が高いことで定評のある適応ビームフォーミングの一種であるMUSIC(MUltiple SIgnal Classification)法を中心としたモジュール群である。音源分離は、音声強調を含むモジュール群であり、音源分離には、GHDSS-AS(Geometric High-order Decorrelation-based Source Separation with Adaptive Stepsize)[5]、音声強調にはHRLE(Histogram-based Recursive Level Estimation)[6]が提供されている。いずれもこれまでのHARKの適用の中で得られた問題点に対応するために我々が独自開発してきた手法である。音声認識は、基本的な特微量抽出や音声認識エンジンを提供するモジュール群であるが、音源分離で生じる歪みに対処するために、ミッシングフィーチャ理論[7]を取り入れている。また、HARKで用いる各種データを作成・可視化するためharktoolと呼ばれるツールを提供している。

統合機能の提供については、図3に示すような典型的なモジュール接続例(Flowdesignerのネットワーク)を数例、HARKのモジュール群と同時に公開している。提供している接続例を利用すれば、一部のパラメータをチューニングするだけで、様々なロボットへ適用することが容易である。また、チューニングの方法については、クリックブックに詳細が記述されている。

4.3 マルチチャネル入力装置のサポート

HARKではロボットに搭載した複数のマイク(マイクロフォンアレイ)を用いて処理を行う。マイクの設置例(図5)では、いずれも8チャネルのマイクロフォンアレイを搭載しているが、HARKでは、任意のチャネル数のマイクロフォンアレイが利用可能である。マルチチャネルA/D装置は、下記の3種類をサポートしている(図6参照)。

- ALSA(Advanced Linux Sound Architecture)をサポートするA/D装置、例えば、RME社製Hammerfall

⁴ Flowdesignerオリジナルは、<http://Flowdesigner.sourceforge.net/>、0.9.0ベース改良版は、<http://winnie.kuis.kyoto-u.ac.jp/HARK/>からそれぞれダウンロードできる。

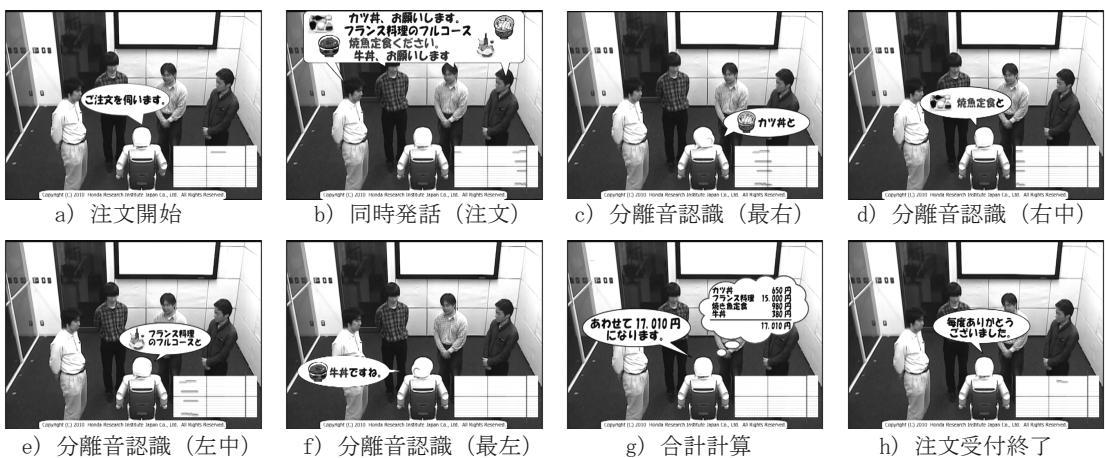


図 7 料理注文タスク（4 話者同時発話）

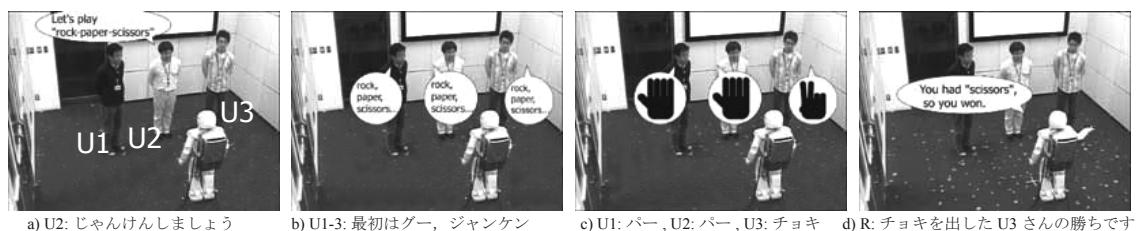


図 8 ロジャンケン判定タスク



図 9 会話シーンのアーカイブと可視化:

音源定位結果 : b)-d)下, 音声認識結果 : b)-d)右, 分離音のインタラクティブな再生 : c)

- シリーズなど.
- ・ 東京エレクトロンデバイス社製 16 チャネル A/D ボード TD-BD-16ADUSB (USB 接続)⁵
 - ・ システムインフロンティア社製マルチチャネル A/D RASP シリーズ (無線/有線 LAN 接続)
- マイクは、安価なピンマイクで構わないが、ゲイン不足解消のため、プリアンプがあった方がよい。TD-BD-16ADUSB や RASP は、プリアンプおよび、プラグインパワー対応の電源供給機能を有しているので、使い勝手がよい。また、RASP シリーズに対応した MEMS マイクロフォンも利用可能である。MEMS マイクロフォンは、マイクロフォン間の個体差が小さい、動作温度範

囲が広い、比較的周波数応答がフラットであることなどから、ロボットへの応用には適しているといえる。

4.4 実時間処理

実時間処理は、音を通じたインタラクションや挙動を扱うロボットシステムには不可欠で本質的な処理である。HARK で用いているアルゴリズムの多くは、一般的な周波数領域信号処理の枠組みを利用しているため、フレーム単位で処理が行われる。音声認識処理ではフレームのシフト長が 10ms 程度であることが一般的であるため、提供しているネットワークでは、一つのフレームに対する全モジュールの平均処理時間が 10ms 以下になることを確認⁶した上で公開を行っている。ロボットシステム

⁵ TD-BD-8CSUSB 用もあるがドライバが古く、推奨しない。

⁶ Intel Core2Duo クラスの CPU で検証を行っている。

における実時間処理では、システムのレスポンス時間に直結する処理遅れも、重要な要素である。処理遅れは、アルゴリズム的に避けられない部分も多いが、HARK では、極力フレームをまたがって行う処理を削減し、提供しているネットワークでは音声認識の前処理部（音源定位から音声認識で用いる特徴量の抽出まで）の処理遅れを 700-800ms 程度に抑えている。処理遅れに直結する属性値は、後からチューニング可能であり、その方法はドキュメントに詳解されている。

5. HARK の応用と展開

5.1 HARK の応用例

HARK の応用として、4人の同時料理注文を聞き分けるロボット（図 7）、ロジヤンケン判定を行うロボット（図 8）、さらには、HARK で処理したデータを可視化するシステム（図 9）を開発してきた。

図 7 では、ロボットが4名のお客に対して料理の注文を聞き（図 7a），4名の客が一斉に料理を注文する。図中右下のグラフにはロボットから見た発話の方向（縦軸）と時間（横軸）が示されており、4名が同時に発話をしたことがわかる（図 7b）。ロボットは頭部に搭載した8チャネルのマイクロフォンアレイを用いて、4名同時に発話の混合音声信号を収音し、話者数と各話者の方向を推定した後、各話者の音声を分離抽出する。さらに、分離抽出した音声の認識を行い、お客様にそれぞれに注文内容を確認する（図 7c-f）。注文の総額を計算し、お客様に告げ（図 7g），注文タスクを終了する（図 7h）。

図 8 では、ユーザーの「ジャンケンしましょう」という発話をトリガーにして、ジャンケンの勝敗を判定するタスクを開始する（図 8a）。3名のユーザーそれぞれがグー、チョキ、パーのいずれかを選択し、同時に発話を（図 8b, c）。このロジヤンケンの同時発話混合音声を、料理注文タスクと同様、頭部に設置したマイクロフォンアレイで収音し、音源定位（人数推定を含む）、音源分離、音声強調、分離音の音声認識といった処理を行い、誰が勝ったか、もしくはあいこだったかを判定する（図 8d）。

図 9 は、音環境シーンをマイクロフォンアレイ（図中央の黒い球体状のもの、7 本のマイクを搭載）を用いて収録し（図 9a），オフラインで HARK を用いてアーカイブしておく。これにより、発話内容、発話方向といった情報を時間同期をとって可視化することができる（図 9b）。また、特定の方向から到来する音だけを選択的に再生することができる（図 9c）。また、インターラクティブに発話イベントを指定して再生することもできる（図 9d）。

これらの応用例は、各ロボット固有のコンフィギュレーションファイルを用意しておけば、あとは GUI 上のカスタマイズのみで他のロボットへの移植が可能である。実際に HRP-2, Robovie-R2 といった複数種類のロボットに移植した実績がある。HARK 導入以前は、ファイル経由ベースの処理を用いており、実話者全員が話し終えてから認識終了まで、約 7.9 秒を要していたが、HARK の使用により、応答が約 1.9 秒に短縮された⁷。応答が速いため、全員の注文終了後、直ちにロボットがそれぞれの注文を復唱し、合計金額を答えるように感じられる。HARK の詳細な性能評価については、文献[2,5,6,7]を参照されたい。

5.2 海外への展開：Texai への応用

2010 年の 3 月に、米国 Willow Garage 社からの招聘を受けて、彼らのテレプレゼンスロボット Texai に、HARK の移植を行った[9]。Texai は、遠隔地にいるユーザ（遠隔ユーザ）が遠隔地から、物理的なボディをもったエージェントとして会議参加や、室内を動き回ってチャットなどを行うためのロボットであり、商用化も予定されている。我々が訪問した時点では、遠隔ユーザからはだれが話しているかわからない、周囲の騒音が大きく、聞きたい人の声が聞き取りづらいといった問題を抱えていた。

そこで、定位情報の可視化、音源分離方向を制御する GUI の構築を通じて、遠隔ユーザが音源方向をカメラ映像上で指定し、特定方向の音源の音だけを聞く機能を新たに実現した。図 10 に HARK を用いた Texai による多人数遠隔インタラクション例を示す。我々が使用したところ、指定した音声を明瞭に聞き取れるようになったと感じられた⁸。具体的な HARK の移植は以下の工程で行った。

1. Texai へのマイクロフォン搭載、インパルス応答の測定及び HARK の移植、
2. Texai を制御している ROS (Robot Operating System) と HARK 間のインターフェースモジュールの実装。

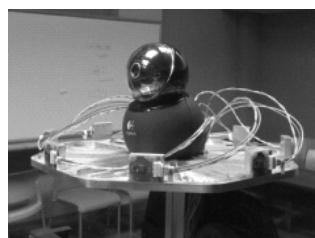
図 11a) に最初に設置したマイクロフォンの設置状況を示す。このロボットを講義室と大食堂に設置、それぞれ 5 度間隔でインパルス応答を測定し、音源定位の性能を測定した。次に、見えた、さらには、マイクロフォン間の音圧差を強調し、音源定位・分離性能を向上させるために Texai に頭を付けることを提案し、付近の店を何件か回り、最終的に、大きさや加工のし易さ、丈夫さを考

⁷ デモは <http://winnie.kuis.kyoto-u.ac.jp/SIG/>

⁸ <http://www.willowgarage.com/blog/2010/03/25/hark-texai>
動画は <http://www.youtube.com/watch?v=xpjPun7Owxg>



図 10 HARK を用いた Texai による多人数遠隔インタラクション：遠隔 Texai（中央）を通じて、遠隔操作者が 2 人の話者と、1 台の Texai とインタラクションを行う。なお、場所はカリフォルニア州であるが、左側の Texai はインディアナ州から遠隔操作中。



a) 初期バージョン



b) 改良版（竹製サラダボール利用）

図 11 Texai の頭部、いずれも 8 個の MEMS マイクロフォンを円状に設置

慮して、雑貨店で見つけた竹製サラダボールを選択した。最初に付けたものとほぼ同じ直径になる辺りに MEMS マイクロフォンを設置した（図 11b）。実は、両者の音源定位性能はそれほど変わらないことが後に判明した。これは一例であるが、理論的に妥当であっても、実際に影響がほとんど現れないことがある。理論研究が多い音響信号処理の分野ではこうしたノウハウ的な知見は、あまり重視されないことが多い。HARK では、こうしたノウハウを取り入れ、プログラムやパラメータのチューニング方法を公開している。このようなノウハウを蓄積する機会が多く得られるのもオープンソース化の利点であろう。

構築した GUI を図 12 に示す。Texai 自身の斜め下の全方位の画像の中央から出ている矢印が、話者の音源方向である。矢印の長さは音量を表している。図中では 3 名の話者がしゃべっていることが分かる。Texai のもう 1 つのカメラの画像が右下に、リモートオペレーターの画像が左下に示されている。図中の円弧は、フィルタで音を通過させる範囲を示す。この円弧内にある方位から届いた音のみが、リモートオペレータに送られる。GUI とリ

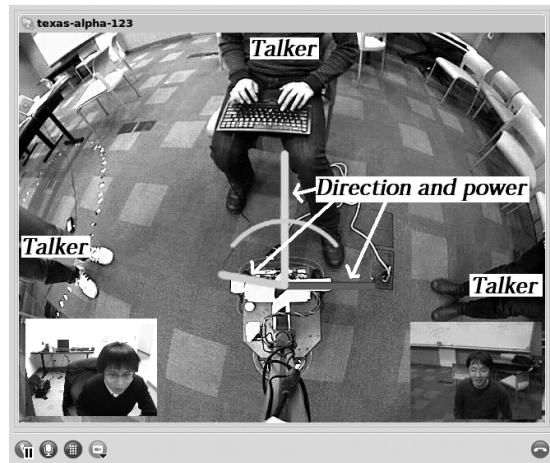


図 12 Texai を通じて、remote operator に見える画面

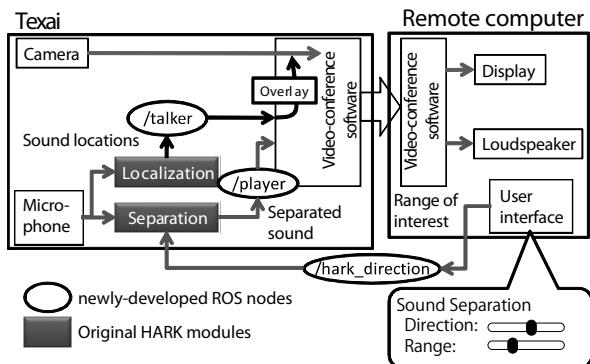


図 13 Texai への HARK の組込方法

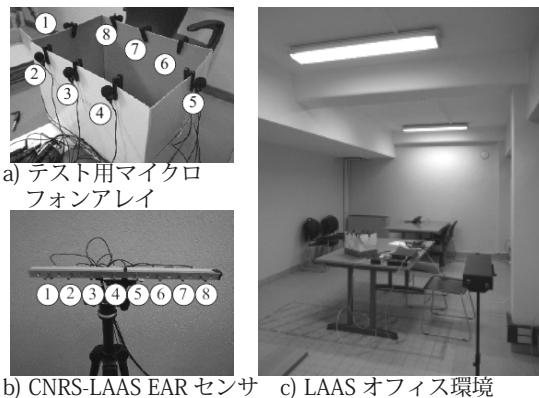


図 14 CNRS-LAAS での HARK 展開

モートオペレータ用の操作コマンド群はすべて ROS モジュールとして実装されているため、図 13 のように HARK と ROS のインターフェースモジュールを定位、分離、フィルタ制御用に 3 つ作成した（図中の○で囲まれたモジュール）。これら一連の作業は頭部の加工、インパルス応答の測定、予備実験、GUI と操作コマンド群の設計を含めて、教員 3 名を含めた計 7 名で目標の 1 週間内で終了できた。HARK や ROS の高いモジュール性が、生産性向上に寄与したと考える。

5.3 普及に向けた国内外の活動

HARK の普及に向けた活動として、ソフトウェア公開以来、国内では、年1回のペースで HARK 講習会の開催を無料で行っている。また、海外でもこれまで、韓国の KIST (Korean Institute of Science and Technology) に招待され講習会を、IEEE の国際学会である Humanoids 2009 ではサテライトイベントとして HARK チュートリアルを開催した。これらの講習会は HARK をより多くの人に認知してもらい、普及を図るという意味で重要である以上に、その場で得られる生の声を参考に今後の研究開発に役立てていくためのフィードバックを得る場という意味で大きな役割を果たしている。

また、2010年11月から12月にかけ、1か月間、フランス CNRS-LAAS (Centre National de la Recherche Scientifique - Laboratoire d'Analyse et d'Architecture des Systèmes) にて学生2名が HARK の HRP-2 へのポーティング作業を行った。図14a) に示すテスト用マイクロフォンアレイで、HARK の動作テスト、および、CNRS-LAAS で研究開発中の EAR センサ (図14b)) を HARK で利用するための音入力インターフェース部の作成を行った。実際に LAAS のオフィス環境 (図14c)) で HARK の音源定位が動作することを確認した。今後、LAAS と協力して評価実験などを行っていく予定である。

6. まとめと今後に向けて

本稿では、ロボット聴覚オープンソースソフトウェア HARK の概要を報告した。HARK は、音環境理解の基本機能である音源定位、音源分離、音声強調、分離音の音声認識をモジュールとして提供し、これらを GUI ベースのプログラミング環境で組み合わせてロボット聴覚の統合システムを構築することができる。また、実際の HARK の利用例として、ロボットの耳への応用・国内外での展開について概説した。

現状で HARK を展開する上での課題の一つは、マイクロフォンアレイのハードウェアの価格であろう。現状、数十万円程度のコストがかかるが、1~2万程度で購入可能なマイクロフォンアレイデバイスが登場すれば、HARK の爆発的な普及も期待できる。Xbox 360 の Kinect のような廉価なマイクロフォンアレイに期待を寄せている（このデバイスは4マイク 1.5万円程度）。

HARK に実装されている信号処理手法は、ロボットに限らず実時間システム構築の際には有効である。本稿がロボット聴覚研究者・開発者の裾野を広げるだけではなく、ロボット聴覚以外の研究者・開発者の目に止まり、様々な場面に展開されるきっかけになれば幸いである。

アンケートにご協力ください。

https://www.ipjs.or.jp/15dp/enquete/enq_dp0202.html

謝辞：ロボット聴覚の研究を共同で推進してきた京都大学 奥乃・尾形研究室の HARK チームの皆さん、HRI-JP の皆さんに感謝します。

参考文献

- 1) K. Nakadai, T. Lourens, H.G. Okuno, H. Kitano: Active Audition for Humanoid, Proc. of 17th National Conference on Artificial Intelligence (AAAI-2000), pp. 832-839, AAAI (2000).
- 2) K. Nakadai, H.G. Okuno, H. Nakajima, Y. Hasegawa, H. Tsujino: Design and Implementation of Robot Audition System "HARK", Advanced Robotics, vol.24, pp.739-761 (2010).
- 3) 奥乃博, 中臺一博: ロボット聴覚オープンソフトウェア HARK, 特集「ロボット聴覚」, 日本ロボット学会誌, Vol.28, No.1 (Jan. 2010) pp.6-9.
- 4) C. Côté, et al.: Code Reusability Tools for Programming Mobile Robots, IEEE/RSJ Int'l Conf. on Robots and Intelligent Systems (IROS 2004), pp.1820-1825 (2004).
- 5) H. Nakajima, K. Nakadai, Y. Hasegawa, H. Tsujino: Blind Source Separation with Parameter-Free Adaptive Step-Size Method for Robot Audition, IEEE trans. Audio, Speech, and Language Processing, vol. 18, no. 6, pp. 1476-1484 (2010)
- 6) H. Nakajima, G. Ince, K. Nakadai, Y. Hasegawa: An Easily-Configurable Robot Audition System Using Histogram-Based Recursive Level Estimation, IEEE/RSJ Int'l Conf. on Robots and Intelligent Systems (IROS 2010), pp.958-963, (2010)
- 7) S. Yamamoto, J.-M. Valin, K. Nakadai, T. Ogata, and H. G Okuno. Enhanced Robot Speech Recognition Based on Microphone Array Source Separation and Missing Feature Theory. IEEE Int'l Conf. on Robotics and Automation (ICRA 2005), pp.1427-1482 (2005).
- 8) T. Mizumoto, K. Nakadai, T. Yoshida, R. Takeda, T. Otsuka, T. Takahashi, and H. G Okuno. Design and Implementation of Selectable Sound Separation on a Texai Telepresence System Using HARK. IEEE Int'l Conf. on Robotics and Automation (ICRA 2011), accepted (2011).

中臺 一博 (非会員)

E-mail: nakadai@jp.honda-ri.com

(株) ホンダ・リサーチ・インスティチュート・ジャパン、プリンシパル・リサーチャ、東京工業大学大学院情報理工学研究科連携教授、ならびに早稲田大学大学院創造理工学研究科客員教授兼務。博士(工学)。1993年東京大学工学部電気工学科卒。1995年同大学院情報工学専攻修了。NTT, NTT コムウェア, JST を経て、現職。ロボット聴覚、実時間情報統合、音環境理解の研究に従事。

奥乃 博 (正会員)

E-mail: okuno@i.kyoto-u.ac.jp

京都大学大学院情報学研究科教授、博士(工学)。1972年東京大学教養学部基礎科学科卒。NTT 研究所、JST、東京理科大学を経て、現職。ロボット聴覚、音環境理解、音楽情報処理、人工知能研究に従事。

投稿受付：2010年10月27日

採録決定：2011年2月16日

編集担当：前田 章 (日立製作所)