ロボット聴覚オープンソースソフトウェアHARKの紹介

中臺 一博[1,2]，　奥乃 博[3]
1(株) ホンダ・リサーチ・インスティチュート・ジャパン　2 東京工業大学 大学院 情報理工学研究科
3 早稲田大学 理工学術院

Introduction to Robot Audition Open Source Software HARK

Kazuhiro Nakadai[1,2], Hiroshi G. Okuno[3].
1 Honda Research Institute Japan Co., Ltd. 2 Tokyo Institute of Technology, 3 Waseda Univ.

Abstract— "HARK" is open source software for robot audition, and it is freely downloadable for research purpose. Since sound source localization, sound source separation and recognition of separated speech are basic and necessary for robot audition, HARK provides such functions with modular architecture. HARK also offers web-based GUI environment called HARK Designer so that users can easily and flexibly build their own robot audition software working in real time. This paper overviews HARK and introduces our continuous activities to deploy HARK.

1. はじめに

近年，災害地用ロボットや自動走行車など様々な研究がロボット分野で行われるようになってきた．また，欧米を中心に研究成果をもとにしたスタートアップが盛んに行われるなど商業化の動きも活発である．こうした中で，ロボットの聴覚機能は，人ロボット音声コミュニケーションに限らず，災害地での人の発見，走行の安全確保，異音検出といった用途でも有用であるにもかかわらず，一部を除いて，盛んに研究開発が行われているとは言い難い．

ロボット聴覚は，ロボットに装着されたマイクロホンを用いて，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り [1)]，HARK (HRI-JP Audition for Robots with Kyoto Univ., hark は listen を意味する中世英語，「はーく」と発音します) [1] は，その 10 年以上に渡る研究成果を共有するために研究開発されたソフトウェアである [2)]．HARK を用いれば，信号処理や音声処理に対する十分な知識がない研究者でも，容易に各自のシステムに組み込むことができるようになる．これにより，研究分野の裾野の拡大を図るとともに，ユーザからのフィードバックによるシステムの安定化を同時に狙っている．

以降では，現在公開中の HARK 2.0 (2014 年 11 月 19 日に 2.1 をリリース予定）をベースに，HARK の機能を紹介するとともに，2008 年にオープンソース化を行って以来，我々が続けてきた HARK の普及活動を紹介する．

2. HARK の概要

HARK の開発は，主に 1) 机上だけでなく，実ロボットにそのまま搭載して利用可能であること，2) 信号処理や音声処理に対する十分な知識がない人でもできるだけ手間をかけず利用できることの 2 点に重きを置いて行われている．

[1] http://www.hark.jp/

一点目に対しては，FlowDesigner [3)] に含まれる batchflow を用いることにより，モジュラー構造を取りつつも，モジュール間のオーバヘッドを最小に保つ取り組みがなされている．具体的には，各モジュールは共有ライブラリとして実装され，実行時にダイナミックリンクされるため，モジュール間通信は，単なる関数コールとして実現できる仕組みになっている．このため，ネットワーク通信ベースの RTM, ROS, CORBA といった汎用的なミドルウェアと比べてオーバヘッドが小さいという特長がある．また，マイクの本数やレイアウトもそれぞれのロボットに応じて変更可能である．ただし，マイクのキャリブレーションは別途必要になるため，事前の音響測定作業，もしくはマイクの位置計測作業が必要となる．HARK では，音響測定用のツールとして Wios, 計測データを基にキャリブレーションを行うツールとして harktool を提供している．独自のマイクアレイを構築する場合には，マイクに加えて，マルチチャネルに対応した A/D 機器も必要である．これについては，ALSA や Direct X, ASIO をサポートしているデバイスであれば，基本的にそのまま利用可能である．

一般に信号処理では，オフライン処理の方が性能が高いため，多くの実装はオフライン処理のみをサポートする形でプログラミングされている．しかし，ロボットでは，オンラインかつ実時間で動作できることが要件であるため，すべてのモジュールは，オンラインで実行が可能なように拡張されている．また，音源定位や音源分離といった処理は，環境の変化に対応できるように適応的に環境の変化に追従できるような処理を可能としている．もちろん，ファイルに一旦格納したデータに対してバッチ的に実行したり，適応的な処理を行わないオフライン処理を行ったりすることも，できるように設計されている．このように，オフライン処理とオンライン処理が同じ実行環境で実現できることにより，大きく作業効率が向上する．

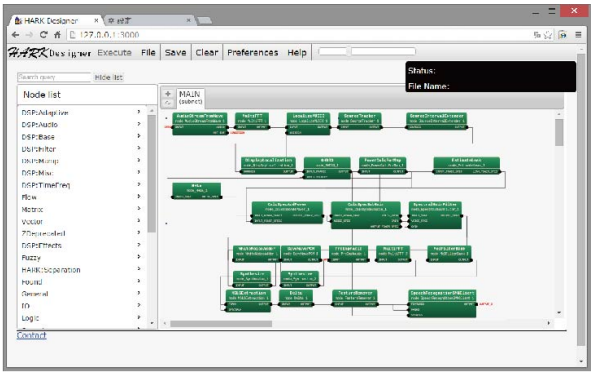


Fig.1 HARK の GUI 環境: HARK Designer

実際にロボットにそのまま搭載して利用する場合，各自で構築した既存システムと接続したいという要望があるだろう．このようなニーズにこたえるため，ロボットではデファクトスタンダードのミドルウェアである ROS とのシームレスな接続が行えるインタフェースを用意している．さらに，比較的簡単に独自のモジュールを開発したいという要望にも応えるため，C/C++ を使わず，python でモジュールが作成できる機能も用意している．

２点目に関しては，GUI ベースのプログラミング環境を提供していることがまず挙げられる．図 1 に示す HARK Designer は web ベースの GUI 環境であり，このためＯＳに依存することなく，ほぼ同じルックアンドフィールでプログラミングを行うことができる．各機能はモジュール化されており、一つの箱 (右側のパネル）として表示される．この箱を GUI 画面上に配置したのち，箱と箱の間を線でつなげるという作業を行うことでプログラミングを行うことができる．各機能の設定は，箱をダブルクリックすることで現れる設定画面を通じて行うことができる．このように，機能をカプセル化して，必要最小限の表示のみを行うことで，一覧性を向上している．各機能の設定については，必ず調整が必要な設定も一部あるものの，ほとんどの設定はデフォルトのままで問題なく利用することができる．このように，調整すべき設定の数を最小化することによりユーザの負担を軽減することは，HARK の設計の指針である．実際には，最適に設定を調整するためにはノウハウが必要であり，ユーザにとっては，ここが大きなバリアとなっている．この問題に対しては，後述する無料の講習会を開催している．また，HARK ドキュメントおよびクックブックという HARK の使い方について記述した 300 頁を超えるドキュメントを日本語および英語の両方で公開している．さらに，ヘルプデスクも用意しており，メールによる個別の対応も行っている．

インストールが容易になるように務めることも開発の際に注意を行っている大きなポイントである．Linux は apt-get を用いて，また，Windows では専用のインストーラを通じて簡単にインストールができるようになっている．実際にはすべてを一つのバイナリとして配布できるとさらに容易なインストールが可能になると考えられるが，サードパーティのライブラリを利用する際など，ライセンスの違いから別パッケージとして配布しなければならない場合もあり，扱いに苦労する点の一つである．

実際に使用する際には，前述のようにマイクロホンアレイを自分で構築することもできるが，多くのユーザは市販の入手しやすいマイクロホンアレイを使っているようである．Kinect for Windows (4 マイク) , Playstation Eye (4 マイク), Microcone (7 マイク)[2], SiF 社クラゲ君 (8 マイク) は，比較的入手しやすいマイクロホンアレイであり，HARK これらのマイクとの接続をサポートしている．また，これらのマイクアレイはキャリブレーション用のファイルも配布している．このため，これらのマイクアレイを用いる場合には，ユーザは前述の測定作業を行うことなく，実時間かつオンラインでマイクロホンアレイ処理を利用できる．一般にマイクアレイ処理の性能はマイクロホン数との相関が高いので，マイクロホン数の多いアレイを選ぶ方がよいが，逆に計算コストと金額は高くなるため，これらのバランスを考えて選択する必要がある．

3. HARK の主要機能

HARK 2.0 では，表 1 に示すパッケージを提供している．ROS, Python, OpenCV といった一般的によく使われるライブラリや言語をサポートしている．このうち，HARK 本体に含まれる音源定位，音源分離，音声認識の 3 つの主要機能について，以下に述べる．

3·1 音源定位

MUltiple SIgnal Classification (MUSIC) 法をベースにした音源定位手法を提供している．MUSIC 法は，固有値分解に基づく手法であり，一般的なビームフォーマと比べて，音源方向のピークが鋭く出やすいことから雑音に頑健である．しかし，アルゴリズム上，雑音レベルが目的音の音量レベルよりも大きくなると目的音ではなく，雑音を定位してしまうという問題があり，HARK ではこれを解決する手法として，一般固有値展開に基づく GEVD-MUSIC 法, および，一般特異値展開に基づく GSVD-MUSIC 法を提供している．これらの手法は，雑音相関行列と呼ばれる雑音に関する知識を用いることで，雑音のレベルが極めて高い場合でも目的音源の定位を可能としている．さらに雑音が動的に変化する場合に対応するために，雑音相関行列を逐次的に推定する iGEVD-MUSIC 法, iGSVD-MUSIC 法も提供している．実際にクアドロコプタに搭載したマイクアレイを用いて，プロペラ音がある場合でも音源定位が可能であることを示している [4].

[2]2014 年 9 月現在，買収に伴い出荷停止中

Table 1 HARK パッケージリスト

パッケージ名	内容
HARK	HARK 本体のモジュール群
JuliusMFT	音声認識
HARKDesigner	HARK GUI 環境
HARK-ROS	HARK と ROS のインタフェース
HARK-Python	Python 用インタフェース
HARK-OpenCV	OpenCV とのインタフェース
HARK-Kinect	Kinect とのインタフェース
HARK-MUSIC	音楽処理
HARK-Binaural	両耳聴処理
Wios	収録ツール
harktool4	マイクキャリブレーションツール
HARK-For-Windows	Windows パッケージ

Table 2 HARK-SSS で提供する音源分離

適応アルゴリズム なし (BF:ビーム フォーマの略)	遅延和 BF (DS-BF) 死角型 BF (NULL-BF) 最小ノルム重み付き BF (WDS-BF) 不定項最小二乗誤差 BF (ILSE-BF) [6]
雑音情報を 陽に利用	最尤 BF（ML-BF） [7, 8] SN 比最大 BF (MNSR-BF) [9]
線形拘束付き 最小分散 (LCMV)	ベース型 (LCMV-BF) [10] Griffith-Jim 型 (GJ-BF) [11]
線形拘束付き ブラインド分離	幾何的音源分離 (GSS) [12] 線形拘束付き独立性に基づく分離 (GICA) [13] 拘束付き高次相関に基づく分離 (GHDSS) [5]

3·2 音源分離

音源分離法として，Geometric High-order Decorrelation Source Separation (GHDSS) [5] を提供している．GHDSS は，ビームフォーミングとブラインド分離のハイブリッド型の音源分離手法である．また，移動音源など音響環境が動的に変化する場合でも，これに追従できるよう適応ステップサイズ法を用いた拡張 (GHDSS-AS 法) を行っている．一般的に，GHDSS-AS 法は実環境でも高い分離性能を示しており，ロボットによる同時発話認識をはじめとした音源分離のデモはこの手法を用いて構築している．

これまでに，様々な音源分離アルゴリズムが発表されており，それぞれが異なる特徴を持っている．このため，場合によっては GHDSS-AS 法以外の手法の方が有利な場合もあろう．そこで，HARK では，HARK-SSS というパッケージを用意して，GHDSS-AS 法を含めて，表 2 に示すように 11 種類の代表的な音源分離手法を提供している．実装が可能なものには，適応ステップサイズ法を用いた拡張も行っている．

3·3 音声認識

音声認識には，Julius をベースにして拡張した音声認識エンジン MFT-Julius を提供している．MFT-Julius は，音源分離や音声強調処理で生じる歪みに対処するため，認識時に歪みをマスクして性能向上を図るミッシングフィーチャ理論を導入した実装である[3]．また，音響特徴量への歪みの影響を最小限にとどめることができる特徴量として，メルスケール対数スペクトル特徴量 (Mel Scale Log Spectrum, MSLS) [14] を提供している．HARK では，音声認識で一般的に使用されるメル周波数ケプストラム係数 (Mel-Frequency Cepstrum Coefficient, MFCC) も提供しているが，MFCC では，スペクトル歪みが特徴量全体に拡散するため，MSLS の方がマイクアレイ処理との相性がよい．一話者発話では，S/N が-3 dB 程度でも，高精度な認識が可能であることを確認している [2]．

[3]実際には東京工業大学旧古井研が公開していた実装をさらに拡張したものである

Table 3 HARK のリリースと講習会のリスト

Apr., 2008：初リリース（0.1.7） 第 1 回講習会：2008/11/17 京都大学, 第 2 回： 2008/12/5 韓国ソウル KIST
Nov., 2009：1.0.0 プレリリース 第 3 回： 2009/1 1 /7 慶應義塾大学日吉, 第 4 回： 2009/12/7 仏パリ UPMC
Nov., 2010：メジャーバージョンアップ（1.0.0） 音源分離の高性能化，ドキュメントの充実 第 5 回講習会： 2010 年 11 月 25 日 京都大学
Feb., 2012：バージョンアップ（1.1） 音源分離の高性能化，64bit 対応，ROS 対応 第 6 回講習会： 2012/2/29 仏パリ UPMC, 第 7 回： 2012/3/9 名古屋大学
Mar., 2013 ：バージョンアップ（1.2） Kinect, PSEye 対応 第 8 回講習会： 2013/3/19 京都大学
Oct., 2013 ：バージョンアップ（1.9.9） Windows & HarkDesigner α版 第 9 回講習会： 2013/10/2 仏ツールーズ CNRS-LAAS
Dec., 2013 ：メジャーバージョンアップ（2.0） Windows & HarkDesigner 対応 第 10 回講習会： 2013/12/5 早稲田大学
Nov., 2014 ：バージョンアップ予定（2.1） 自己雑音抑圧対応 第 11 回講習会： 2014/11/19 早稲田大学

3·4 HARK によるロボット聴覚のデモ

HARK を用いて構築したロボット聴覚のデモを紹介する．

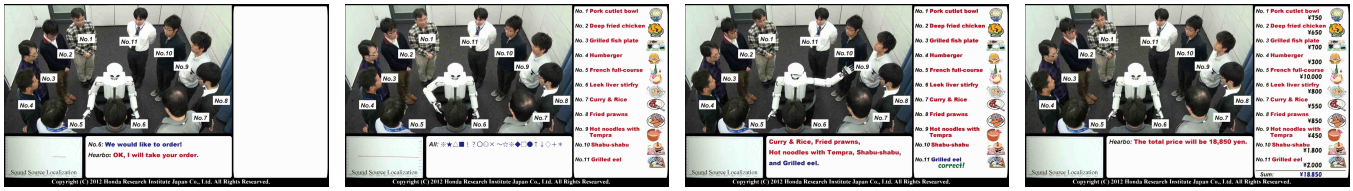
- ロボットによる 11 人の同時料理注文聞き分け (図 2)
- タブレットを用いた多言語コミュニケーション支援 (図 3)

前者は，同時に 11 人が発声した料理注文に対して，頭部に搭載した 16 本のマイクを用いて，GHDSS-AS による音源分離，分離音の音声認識を行い，各自の注文を確認した後，合計の金額を告げるというデモである．

後者はタブレットの周囲に装着した 8 本のマイクを用いて，まず，音源定位を行い，各話者の方向を認識する．各話者の発話は，音声認識を行った後に，それぞれの母国語に翻訳される．最終的に翻訳されたテキストを，認識した話者の方向に表示する．これにより，各話者に見やすい方向でテキストを提示できる．なお，認識や翻訳は大語彙の認識に対応したクラウドサービスを利用している．

4. HARK の普及活動

HARK のリリースとそれに伴って行ってきた講習会のリストを表 3 に示す．ほぼ毎年，ソフトウェアのアップデートを行っており，同時に国内外での講習会を開催して



a) 注文開始　b) 11 人が同時に注文　c) 各声を分離認識し確認　d) 合計金額の提示

Fig.2 ロボットによる 11 人の同時料理注文聞き分け

a) 母国語の異なる 4 名の会話 (日・英・中・仏語)　b) 日本語発話（緑）の認識・翻訳，話者向き表示　c) 英語発話（赤）を同様に処理　d) 会話が進んだところ 話者位置変化に追従し表示

Fig.3 タブレットを用いた多言語コミュニケーション支援

いる．講習会は，毎回 50 名程度の募集を行い，ほぼ満席となる．研究用途のソフトウェアという位置づけの割に企業からの参加が比較的多いことも特徴である．2014 年度は，講習会に加えてハッカソンの開催も企画しており，普及に向けた活動をもう一段加速したいと考えている．

海外への展開も積極的に行っている．2010 年の 3 月には，米国 Willow Garage 社からの招聘を受けて，テレプレゼンスロボット Texai に，HARK の移植活動を行った[15]．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室内を動き回ってチャットなどを行うために開発されたロボットであるが，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．ロボット頭部の加工，マイクのキャリブレーション，予備実験，GUI と操作コマンド群の設計・実装を，教員 3 名を含めた計 7 名でなんとか目標の 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．

また，2010 年 11 月から 12 月にかけ，1 か月間，フランス CNRS-LAAS 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．HARK の動作テスト，および，CNRS-LAAS で研究開発中の Ear Sensor と呼ばれるマイクアレイを HARK で利用するための音入力インタフェース部の作成を行った．その後，LAAS とは，共同研究プロジェクト（BINNAHR）[4]にも発展し，HARK の展開活動のよい成功例となった．

[4]http://projects.laas.fr/BINAAHR/BINAAHR/Welcome.html

5. おわりに

本稿では，ロボット聴覚研究の成果として，2008 年から研究用途に一般公開を行っているオープンソースソフトウェア HARK の概要を 2014 年 9 月現在公開中の最新版である HARK 2.0 をベースに紹介した．また，公開を開始して以来，行ってきた継続的な更新と講習会等による展開活動についても報告した．是非，みなさんも HARK をお使いいただき，ロボットにおける音の重要性を再認識していただくとともに，忌憚のないコメントをヘルプデスクまでいただければ幸いである．

謝辞

中村圭佑氏，水本武志氏をはじめとした，HRI-JP，京大，東工大の HARK 開発チームの各メンバに感謝する．

参考文献

1) K. Nakadai *et al.* Active Audition for Humanoid, *AAAI-2000*, pp. 832-839.
2) K. Nakadai *et al.* Design and Implementation of Robot Audition System “HARK”, *Advanced Robotics*, vol.24, pp.739-761 (2010).
3) C. Côté *et al.* Code reusability tools for programming mobile robots. *IEEE IROS 2004*, pp. 1820–1825.
4) T. Ohata *et al*, Improvement in Outdoor Sound Source Detection Using a Quadrotor-Embedded Microphone Array, *IEEE IROS 2014*.
5) H. Nakajima *et al.*, Blind Source Separation with Prameter-Free Adaptive Step-Size Method for Robot Audition, *IEEE Trans. ASLP*, 18(6), pp. 1476-1484.
6) 中島他, 不定項を用いた任意配置マイクロホンによるビームフォーミング, 2002 年秋季研究発表会講演論文集, pp.527-528, 2002, ASJ
7) V.A.N. Barroso and J.M.F. Moura, Maximum likelihood beamforming in the presence of outliers, *IEEE ICASSP-91*, pp. 1409 - 1412, 1991.
8) M.L. Seltzer *et al*, A Bayesian Framework for Spectrographic Mask Estimation for Missing Feature Speech Recognition, *Speech Communication*, 43(4), pp. 379-393, 2004.
9) R.A. Monzingo, and T.W. Miller, Introduction to adaptive arrays, SciTech Publishing, 1980
10) O.L. Frost, An algorithm for linearly constrained adaptive array processing, *Proceedings of the IEEE*, 60(8), pp.926–935, 1972.

11) L.J. Griffiths, and C.W. Jim, An alternative approach to linearly constrained adaptive beamforming, *IEEE Trans. on Antennas and Propagation*, 30(1), pp.27–34, 1982.
12) L.C. Parra and C.V. Alvino, Geometric source separation: Mergin convolutive source separation with geometric beamforming, *IEEE Trans. on Speech and Audio Processing*, 10(6), pp. 352–362, 2002
13) M. Knaak *et al.*, Geometrically Constrained Independent Component Analysis, *IEEE Trans. on ASLP*, 15(2), pp.715–726, 2007.
14) S. Yamamoto *al.*, Enhanced robot speech recognition based on microphone array source separation and missing feature theory, *IEEE/RAS ICRA 2005*, pp.1427-1482.
15) T. Mizumoto *et al*, Design and implementation of selectable sound separation on the Texai telepresence system using HARK, *IEEE/RAS ICRA-2011*, pp.2130–2137.