

## 第11章

# 結論

本研究では、将来、人間と共生するロボットに求められるロボット聴覚について論じ、アクティブオーディションの提案、また、これを積極的に利用したロボット聴覚処理のモデル化を行った。実際に、アクティブオーディションを利用した動作時のノイズキャンセル、音源定位、音源分離、分離音の音声認識について、工学的な実現・評価を行った。さらに、パーソナリティを導入した非言語インタラクションについても検討し自然なインタラクションを実現した。

2章では、まず、ロボット聴覚という研究の定義、ロボットの聴覚機能実現に向けた課題の整理を行った。次に、研究の方向性をより明確にするために、必要なマイクロホンの本数について議論を行い、2本のマイクロホンが必要であるという考えを述べた。そして、2本のマイクロホンでの処理を前提に、実環境ロボット聴覚を実現するためには、体を動かして聴くというアクティブオーディションが鍵となることを述べた。また、アクティブオーディションを実現するためには、混合音が扱えること、動きながら聞く機構、未知環境や実環境での音の知覚、画像処理などの他の処理との統合、実時間処理といった課題が本質的であることを、その関連研究を交えて議論した。さらに、ロボット聴覚の応用として従来のヒューマンロボットインタラクションより、自由度が高く、自然なインタラクションの実現について議論を行った。

3章では、ロボット聴覚の実現に向けた課題を考慮して、アクティブオーディションを効果的に利用できるロボット聴覚システムの全体像を示した。同時に、ロボット聴覚システムを適用するテストベッドとして開発したヒューマノイドロボット SIG についても紹介し、その諸元、ラピッドプロトタイピングによる製作過程を説明した。

4章では、アクティブオーディションによるアクティブな動作は、知覚をロバストにし、インタラクションを豊かにする有効な手段である反面、モータノイズが聴覚処理を困難してしまうという問題を扱った。このような問題に対しては、ロボットが音響的な身体性を獲得できるように外装の内外にマイクを設置し、外装内部での共鳴に対応するために外装

の音響測定を行うことが鍵であることを示し、外装の音響効果を利用した新しいノイズキャンセル手法を提案した。一方で、実環境での連続的な音源定位を行うために聴覚エピソード幾何を提案した。聴覚エピソード幾何は、一般に両耳聴の音源定位に利用されている HRTF を利用しないため、実環境・未知環境への適用が難しいという HRTF の抱えている問題を解決できることを示した。外装の音響効果を利用したノイズキャンセルと聴覚エピソード幾何によって、ある程度の反響がある部屋で、“*stop-perceive-act*” 原理に従わない、動作を行いながらの連続的な純音の定位を可能にした。そして、音源方向を向くというアクティブな動作は、聴覚処理による音源定位の精度を向上させること、および視覚による定位情報を利用すれば、より正確な定位が期待できることを示し、アクティブオーディションの有効性を示した。

5 章では、より一般的な環境でのロバストな動作を目指し、実時間・実環境での純音を含めた調波構造を持った複数音源の定位・追跡を扱った。正確に調波構造を抽出するためのピーク抽出に関しては、スペクトルサブトラクション法に基づく、高速で周波数解像度の高い手法を提案した。音源定位・追跡には、4 章で述べた手法を拡張し、IPD, IID, 調波構造といった複数の聴覚的な手がかりを Dempster-Shafer 理論を用いて統合するモデルを提案した。実際に、実験を通じて、聴覚情報の統合により、部分的に歪んでしまった音響信号の定位が可能になることを示し、聴覚処理のロバスト性を高めるために情報統合が本質的であることを示した。また、提案手法では、最大同時音源数については、各音源の角度差が 30 度以上という条件で、3 から 4 音源であることを定量的に明らかにした。ただし、定位は水平方向、かつ前方向に限られ、上下方向への対応やアクティブな動作を用いた前後問題の解決は、今後の課題である。

6 章では、音源定位・追跡処理を顔認識・定位を行う視覚処理と統合して、実環境でロバストに複数人物を実時間で追跡できるシステムをヒューマノイド SIG 上に実現した。システム内の聴覚情報、視覚情報はその時間的な流れを考慮できるストリームとして表現され、視聴覚ストリームがその距離に応じてアソシエーションを行うメカニズムを導入することで視聴覚統合を行った。このような視聴覚統合により、聴覚の曖昧性を解消するだけでなく、視覚の視野の狭さ、オクルージョンといった視覚情報の曖昧性も解決できることを示した。さらに、提案した情報統合法のスケラビリティを示すために、実時間複数人物追跡システムをステレオビジョン、話者同定といった処理によって得られるモダリティと階層的に統合できるように拡張した。これにより、ストリームベースのシンボリックな情報統合はモダリティ数のスケールアップが容易であることを示すと同時に、モダリティ追加によりロバスト性が向上できることも併せて示した。つまり、各々の処理は精度が低くても、それらを統合することにより、ロバスト性を高めることができるという情報統合の本質的な有効性を示すことができた。

7章では、アクティブ方向通過型フィルタを用いた音源分離を扱った。アクティブ方向通過型フィルタは、特定の方向の音を抽出するフィルタであり、マイクロホンの本数以上の音源を実時間で分離抽出することが可能である。また、本研究で聴覚中心窩と呼ぶ、音源定位・分離性能が正面方向で高く、横方向に行くに従い、低くなるという現象を積極的に利用できることもアクティブ方向通過型フィルタの特徴である。つまり、フィルタ通過帯域を正面方向では狭く、横方向では広がるようアクティブに制御するようになっており、注意を傾ける場合には、対象音源に最も感度の高いフィルタの正面方向を向けるアクティブな動作を利用する。アクティブ方向通過型フィルタの有効性は、ある程度の反響のある部屋で3話者同時発話の分離性能はノイズ除去率で9dB程度であること、最大分離音源数は3から4音源程度であること、対象物の方向を向くというアクティブな動作による音源分離向上によって示すことができた。通過帯域のアクティブな制御、対象物の方向を向くというアクティブな動作が音源分離を向上させることから、音源分離でも、アクティブな動作は知覚向上において本質的であることつまりアクティブオーディションの有効性を示すことができた。

8章では、自然なヒューマンロボットインタラクションを実現するためのロボット聴覚システムの応用として、アクティブ方向通過型フィルタで分離した分離音声の孤立単語認識を扱った。分離音声の認識では、音源方向、話者ごとに音響モデルを作成し、複数の音声認識を用いた方法を提案した。入力に対して、複数の音響モデルを用いた音声認識処理を並列に行い、その結果を予め測定した各音響モデルの音声認識率に基づき、確率ベースの統合を行う手法を提案した。また、この統合手法は顔認識から得られた名前情報とも統合が可能な枠組みを備えている。これにより、同時3話者発話の音声認識を達成した。また、音声認識においても、音源方向を向くというアクティブな動作を行ったり、顔認識と組み合わせたりすることによって認識率が向上することを示すことにより、提案統合手法の有効性を示した。つまり、音声認識におけるアクティブオーディションの有効性、視聴覚統合の有効性を示すことができた。

9章では、ロボット聴覚システムの応用として、パーソナリティを取り入れた注意制御を扱った。パーソナリティには心理学で用いられるインターパーソナル理論を用い、friendly, dominant, hostile といったパーソナリティを実現し、受付やコンパニオンなど様々なケースに適用した。実際に人間とのインタラクションでは、非言語でパッシブなインタラクションであっても、話者の方向を向くことによって、人々の興味を喚起できる非言語 Eliza ともいうべき現象が見受けられた。音源方向を向くという動作は、ロボット聴覚の向上だけでなく、人間とのフレンドリーなインタラクションを実現したり、テレグジスタンスによる会議では、相手の注意を向けさせたりという意味でも重要であることを示すことができた。ただし、詳細な評価は今後の課題である。

10 章では、実用化に向けた今後の課題について議論した。特に提案したロボット聴覚システムの汎用性、環境に対するロバスト性、システムの物理的な大きさといった点について議論を行った。また、実際に他のロボットへの適用を通じて、システムの汎用性も示した。

2 章の冒頭で引用した節からも、聴覚機能は人間生活にとって極めて重要であり、また、ロボティクスという面からも、AI における実環境・実時間処理という面からも、課題が多い分野であるため、より一層の研究が必要である。

本研究では、実環境での音響処理の課題をまとめ、これを解決するべく、システムの工学的な実現を通して、アクティブな動作が実環境での聴覚処理向上に本質的であること示した。

本研究で示した内容は、ロボットに限らず、実環境での聴覚処理に有効である。本研究が、ロボットに特化しない、一般的な実環境における音響処理の新しい枠組を打ち立てるための一助となれば幸いである。

しかし、考察でも述べたように、実環境処理の課題は多い。これらの課題を解決し、将来、本研究の応用として、いくつかの音を同時に聞くことができるカクテル・パーティー・ロボットまたは「聖徳太子」ロボットや人間と共生し、制約のない環境で、人間と言語・非言語のインタラクションを通じてコミュニケーションを行うロボットなどが実現できると信じている。

# 謝辞

本研究を進めるにあたって、京都大学情報学研究科知能情報学専攻の奥乃 博教授からは、北野共生システムプロジェクトへの参加の機会を与えていただき、また、ロボット聴覚にというテーマに取り組むに当たり、多くの御指導、御助言をいただきました。心から感謝の意を表します。

東京大学情報理工学系研究科長の田中英彦教授からは、博士論文提出に際し、快く主査を引き受けていただいただけでなく、学生時代から今日に至るまでの多くの御指導、御助言に対し、心から感謝の意を表します。

副査の安藤繁教授、石川正俊教授、稲葉雅幸教授、坂井修一教授には、審査時に貴重なご意見をいただいただけでなく、学会活動等でも、ご指導ご鞭撻をいただいたことに感謝いたします。

科学技術振興事業団 ERATO 北野プロジェクトの北野宏明博士には、私が本プロジェクトに参加した 1999 年 7 月以来、研究をバックアップしていただいたことに感謝します。

元北野プロジェクトの研究者 Dr. Tino Lourens (現 HRI 研究員) には、共同研究者として、視覚情報処理に関する様々な意見、時には、音響処理についても有意義な意見をいただき感謝します。

元北野プロジェクトの技術員 日台健一氏 (現 SONY DCL) には、共同研究者として、視覚処理のプログラミングを始め、様々なサポートをしていただいたことに感謝します。特に本研究の顔認識・ステレオビジョンについては、日台氏の協力無くしては成し得ませんでした。

元北野プロジェクトの研究者 松井龍哉氏 (現 Flower Robotics, Inc.) については、SIG のデザイナーであり、非常に美しいロボットを提供していただいたことを始め、プレゼンテーション資料作成など様々な面でお世話になったことに感謝します。

元北野プロジェクトの研究者 宮下敬宏博士 (現 ATR 研究員) には、制御理論を始め、SIG2 の機構設計など様々な面でお世話になったことに感謝します。

元北野プロジェクトの技術員 山崎文敬氏 (現 大阪大学) には、研究に関する有意義な議論、および、研究以外でも北野共生プロジェクトでの様々なサポートに感謝いたします。

東京工業大学 松浦 大輔氏, および, 慶応大学 鈴木 紀章氏については, 研究のサポート, および, 有意義な議論をしていただいたことに感謝します.

大阪大学 石黒 浩教授には, 視覚と聴覚の統合に関して, 様々なご意見をいただいたことに感謝いたします.

京都大学 河原 達也助教授, 南條 浩輝氏, 秋田 祐哉氏には, 話者同定, 音声認識について様々なご意見や議論をいただきましたことに感謝します.

東京理科大学 溝口 博助教授には, 視覚処理について様々なご意見をいただいたことに感謝いたします.

豊橋科学技術大学の中川 聖教授には, 音源分離に対してご意見をいただいたことに感謝します.

株式会社 SGI 五十嵐宏樹氏, 早稲田大学 鈴木 健氏, 法政大学 森田 良二氏, 奥乃 知飛氏, 種村 裕介氏には, 研究に対するサポートをいただいたことに感謝します.

Flower Robotics, Inc. の星野氏, 中三川氏, 樫村氏には, ロボットのデザインからビデオや資料の作成に至るまで, 様々なサポートをいただき感謝いたします.

また, 科学技術振興事業団 ERATO 北野プロジェクトの方々には, 日ごろから様々な面において研究のバックアップをしていただいたことに感謝します.

NTT CS 研究所の柏野 邦夫博士には, 学生時代より研究に対して, 大変お世話になりました. 深い感謝の意を表します.

日東紡音響エンジニアリングの高島氏, 茂田氏には音響測定について様々なご意見をいただきました. 感謝いたします.

株式会社サイバネットの秋山氏には, 音響シミュレーションで大変お世話になりました. 感謝いたします.

合資会社エヌメディアールの竹内 啓祐氏には, 主にプロジェクトのネットワーク管理の面でお世話になりました. 感謝いたします.

産業総合研究所 小池汎平博士には, 本プロジェクトへ参加する際に推薦状をいただきました. 感謝いたします.

田中英彦・坂井修一研究室の皆様には, 私が卒業後も変わりなくお付き合いいただいたことに感謝いたします.

私の前職である NTT, および NTT コムウェアの人々には, 退社後も貴重なご意見をいただきました. 感謝いたします.

ここには, 挙げられませんが, いつも私に元気や活力を与えてくださった多くの友人たちに感謝いたします.

また, いままで自分の意見を尊重し, 見守っていただいた両親に感謝いたします. 最後に, 妻の紀江といつも私に元気を与えてくれる娘, 彩貴に感謝します.

# 発表文献

## 本研究に直接関係する文献一覧

### [ジャーナル論文] 5 件

1. Kazuhiro Nakadai, Ken-ichi Hidai, Hiroshi G. Okuno, Hiroshi Mizoguchi and Hiroaki Kitano: Real-Time Auditory and Visual Multiple-Speaker Tracking For Human-Robot Interaction, *Journal of Robotics and Mechatronics*, pp.479-489, JSME, 2002.
2. Hiroshi G. Okuno, Kazuhiro Nakadai, Ken'ichi Hidai, Hiroshi Mizoguchi, Hiroaki Kitano: Human-Robot Non-Verbal Interaction Empowered by Real-Time Auditory and Visual Multiple-Talker Tracking, *Advanced Robotics*, RSJ, 2002 (印刷中).
3. Hiroshi G. Okuno, Kazuhiro Nakadai, Lourens Tino, Hiroaki Kitano: Sound and Visual Tracking for Humanoid Robot, *Applied Intelligence*, Kluwer Academic Publishers (印刷中).
4. 中臺 一博, 日台 健一, 溝口 博, 奥乃 博, 北野 宏明: ヒューマノイドを対象にした視聴覚統合による実時間人物追跡 - アクティブオーディションと顔認識の統合 -, *日本ロボット学会誌*, vol.21, no.6, 2003
5. 中臺 一博, 奥乃 博, 北野 宏明: ヒューマノイドにおける聴覚機能の課題とアクティブオーディションによる音源定位, *人工知能学会論文誌*, Vol. 18, No. 2-F, pp.104-113, 2003 (印刷中)

### [国際学会] 29 件

1. Kazuhiro Nakadai, Hiroshi G. Okuno and Hiroaki Kitano: Robot Recognizes Three Simultaneous Speech By Active Audition, *Proc. of IEEE International Conference on Robotics and Automation (ICRA 2003)*, Taipei, Taiwan, 2003 (accepted)

2. Hiroshi G. Okuno, Kazuhiro Nakadai and Hiroaki Kitano: Realizing Personality in Audio-Visually Triggered Non-verbal Behaviors, Proc. of IEEE International Conference on Robotics and Automation (ICRA 2003), Taipei, Taiwan, 2003 (accepted)
3. Hiroshi G. Okuno, Kazuhiro Nakadai and Hiroaki Kitano: Design and Implementation of Personality of Humanoids in Human Humanoid Non-verbal Interaction, Proc. of 19th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE-2003), 2003 (accepted)
4. Kazuhiro Nakadai, Hiroshi G. Okuno, Hiroaki Kitano: Auditory Fovea Based Speech Separation and Its Application to Dialog System, Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2002), pp.1314-1319, IEEE, Lausanne, Swiss, Oct. 2002.
5. Kazuhiro Nakadai, Hiroshi G. Okuno, Hiroaki Kitano: Auditory Fovea Based Speech Enhancement and Its Application to Human-Robot Dialog System Proc. of 7th International Conference on Spoken Language Processing (ICSLP-2002), pp.1817-1820, Denver, USA, Sep. 2002.
6. Hiroshi G. Okuno, Kazuhiro Nakadai, Hiroaki Kitano: Real-time Sound Source Localization and Separation for Robot Audition Proc. of 7th International Conference on Spoken Language Processing (ICSLP-2002), pp.193-196, Denver, USA, Sep. 2002.
7. Hiroshi G. Okuno, Kazuhiro Nakadai, Hiroaki Kitano: Realizing Audio-Visually triggered ELIZA-like non-verbal Behaviors. Seventh Pacific Rim International Conference on Artificial Intelligence(PRICAI-2002), pp.552-562, Tokyo(NII), Japan, Aug. 2002.
8. Kazuhiro Nakadai, Hiroshi G. Okuno, Hiroaki Kitano: Exploiting Auditory Fovea in Humanoid-Human Interaction Proc. of the Eighteenth National Conference on Artificial Intelligence (AAAI-2002), pp.431-438, Edmonton, Canada, Aug. 2002.
9. Hiroshi G. Okuno, Kazuhiro Nakadai, Hiroaki Kitano: Non-verbal ELIZA-like Human Behaviors in Human-Robot interaction through Real-Time Auditory and Visual Multiple-Talker Tracking The Third International Cognitive Robotics Workshop (CogRob2002), pp.59-65, Edmonton, Canada, Jul. 2002.
10. Hiroshi G. Okuno, Kazuhiro Nakadai, Hiroaki Kitano: Social Interaction of Humanoid Robot Based on Audio-Visual Tracking Proc. of 18th International



- Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE-2002), Lecture Notes in Artificial Intelligence, Springer-Verlag. Cairns, Australia, June 2002.
11. Kazuhiro Nakadai, Ken-ichi Hidai, Hiroshi G. Okuno, Hiroaki Kitano: Real-Time Speaker Localization and Speech Separation by Audio-Visual Integration Proc. of IEEE International Conference on Robotics and Automation (ICRA 2002), pp.1043-1049, Washington D.C., May. 2002.
  12. Hiroshi G. Okuno, Kazuhiro Nakadai, Lourens, Tino, Hiroaki Kitano: Sound and Visual Tracking by Active Audition, Q. Jin, J. Li, N. Zhang, J. Cheng, C.Yu, S. Noguchi (Eds.) Enabling Society with Information Technology, pp.174-185, Springer-Verlag, Tokyo, Jan. 2002.
  13. Kazuhiro Nakadai, Ken-ichi Hidai, Hiroshi G. Okuno, Hiroaki Kitano: Real-Time Active Human Tracking by Hierarchical Integration of Audition and Vision Proc. of Second IEEE-RAS International Conference on Humanoid Robots (Humanoids2001), pp.91-98, Tokyo, Nov. 2001.
  14. Kazuhiro Nakadai, Ken-ichi Hidai, Hiroshi G. Okuno, Hiroaki Kitano: Epipolar Geometry Based Sound Localization and Extraction for Humanoid Audition Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2001), IEEE, Maui, Hawaii, Oct. 2001.
  15. Hiroshi G. Okuno, Kazuhiro Nakadai, Tino Lourens, Hiroaki Kitano: Human-Robot Interaction Through Real-Time Auditory and Visual Multiple-Talker Tracking Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2001), IEEE, Maui, Hawaii, Oct. 2001. (**BEST PAPER Nomination Finalist**)
  16. Hiroshi G. Okuno, Kazuhiro Nakadai, Tino Lourens, Hiroaki Kitano: Separating three simultaneous speeches with two microphones by integrating auditory and visual processing. Proc. of European Conference on Speech Processing (Eurospeech 2001), pp.2643-2646, Aalborg, Denmark, Sep. 2001.
  17. Kazuhiro Nakadai, Ken-ichi Hidai, Hiroshi Mizoguchi, Hiroshi G. Okuno, Hiroaki Kitano: Real-Time Multiple Speaker Tracking by Multi-Modal Integration for Mobile Robots Proc. of European Conference on Speech Processing (Eurospeech 2001), pp.1193-1196, Aalborg, Denmark, Sep. 2001.
  18. Kazuhiro Nakadai, Ken-ichi Hidai, Hiroshi Mizoguchi, Hiroshi G. Okuno, Hiroaki Kitano: Real-Time Auditory and Visual Multiple-Object Tracking for Robots. Proc. of 17th International Joint Conference on Artificial Intelligence

- (IJCAI-01), 1425-1432, Seattle, Aug. 2001. (電気通信普及財団奨励賞受賞)
19. Tino Lourens, Kazuhiro Nakadai, Hiroshi G. Okuno, Hiroaki Kitano: Detection of Oriented Repetitive Alternating Patterns in Color Images - A Computational Model of Monkey Grating Cells. Proc. of Sixth International Work-Conference on Artificial and Natural Neural Networks (IWANN2001), Lecture Notes in Artificial Intelligence, No.2084, 95-107, Springer-Verlag. Granada, Spain, June 2001.
  20. Hiroshi G. Okuno, Kazuhiro Nakadai, Lourens, Tino, Hiroaki Kitano: Sound and Visual Tracking for Humanoid Robot. Proc. of 17th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE-2001), Lecture Notes in Artificial Intelligence, No. 2070, Springer-Verlag. Budapest, Hungary, June 2001. (**BEST PAPER AWARD**)
  21. Ian Frank, Kumiko Ishii-Tanaka, Hiroshi G. Okuno, Jun-ichi Akita, Yukiko Nakagawa, K. Maeda, Kazuhiro Nakadai, and Hiroaki Kitano: And The Fans are Going Wild! SIG plus MIKE. RoboCup 2000: Robot Soccer World Cup IV, Lecture Notes in Artificial Intelligence No.2019, 139-148, Springer-Verlag, May 2001.
  22. Hiroshi G. Okuno, Kazuhiro Nakadai, Tino Lourens, Hiroaki Kitano: Sound and Visual Tracking for Humanoid, Proc. of 2000 International Conference on Information Society in the 21st Century: Emerging Technologies and New Challenges (IS2000), 254-261, Aizu-Wakamatsu, Nov. 2000. (**BEST PAPER AWARD**)
  23. Kazuhiro Nakadai, Tatsuya Matsui, Hiroshi G. Okuno, Hiroaki Kitano: Active Audition System and Humanoid Exterior Design. Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2000), 1453-1461, Takamatsu, Nov. 2000.
  24. Hiroaki Kitano, Hiroshi G. Okuno, Kazuhiro Nakadai, Theo Sabische, Tatsuya Matsui, Design and Architecture of SIG the Humanoid: An Experimental Platform for Integrating Perception in RoboCup Humanoid Challenge. Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2000), 181-190, Takamatsu, Nov. 2000.
  25. Kazuhiro Nakadai, Hiroshi G. Okuno, Tino Lourens, Hiroaki Kitano. Humanoid Active Audition System. Proc. of First IEEE-RAS International Conference on Humanoid Robots (Humanoids2000), Cambridge, Sep. 2000.

26. Tino Lourens, Kazuhiro Nakadai, Hiroshi G. Okuno, Hiroaki Kitano: Selective Attention by Integration of Vision and Audition. Proc. of First IEEE-RAS International Conference on Humanoid Robots (Humanoids2000), Cambridge, Sep. 2000.
27. Kazuhiro Nakadai, Tino Lourens, Hiroshi G. Okuno, Hiroaki Kitano: Humanoid Active Audition System Improved by The Cover Acoustics. In Mizoguchi, R. and Slaney, J. (eds) PRICAI-2000 Topics in Artificial Intelligence (Sixth Pacific Rim International Conference on Artificial Intelligence), 544-554, Lecture Notes in Artificial Intelligence No. 1886, Springer-Verlag, Melbourne, Aug. 2000.
28. Ian Frank, Kumiko Ishii-Tanaka, Hiroshi G. Okuno, Kazuhiro Nakadai, Yukiko Nakagawa, K. Maeda, Hiroaki Kitano: And The Fans are Going Wild! SIG plus MIKE. Proc. of the Fourth Workshop on RoboCup (RoboCup-2000), 267-276, RoboCup, Melbourne, Aug. 2000.
29. Kazuhiro Nakadai, Tino Lourens, Hiroshi G. Okuno, Hiroaki Kitano: Active Audition for Humanoid. Proc. of the Seventeenth National Conference on Artificial Intelligence (AAAI-2000), 832-839, Austin, Aug. 2000

[研究会, Workshop, 解説, 全国大会] 20 件

1. 鈴木 紀章, 中臺一博, 天野 英晴, 奥乃 博, 北野宏明: 方向通過型音源分離フィルタへの FPGA の応用. 信学技報, VLSI 設計技術研究会 VLD2002-143, pp.130-145, 電子情報通信学会, Jan. 2003.
2. 中臺一博, 奥乃 博, 北野宏明: アクティブオーディションによる複数音源の定位・分離・認識. 第 16 回 AI チャレンジ研究会, SIG-Challenge-0216-4, pp.25-32, 人工知能学会, Nov. 2002.
3. 中臺一博, 日台健一, 奥乃 博, 北野宏明: アクティブな視聴覚統合を用いた実時間人物追跡ヒューマノイド SIG. 知能と複雑系研究会 2001-ICS-126, pp.37-42, 情報処理学会, Oct. 2001.
4. 中臺一博, 日台健一, 奥乃 博, 北野宏明: 視聴覚情報の階層的統合による実時間アクティブ人物追跡. 第 13 回 AI チャレンジ研究会, SIG-Challenge-0113-5, pp.35-42, 人工知能学会, Mar. 2001.
5. 中臺一博, 日台健一, 溝口 博, 奥乃 博, 北野宏明: 顔認識とアクティブオーディションを利用した実時間人物追跡, 第 11 回 SIG-Challenge 研究会, pp.27-34, 日本人工知能学会, 2001 (日本人工知能学会研究奨励賞受賞)
6. 奥乃 博, 中臺一博: ロボット聴覚の課題と現状. 信学技報, NLC2001-60, SP2001-

- 95, pp.69-74, 電子情報通信学会, Dec. 2001.
7. Nakadai, K., Okuno, H.G., and Kitano, H.: A Method Of Peak Extraction And Its Evaluation For Humanoid 第 7 回 AI チャレンジ研究会, SIG-Challenge-99-7, pp.53-60. 人工知能学会, Nov. 1999.
  8. Okuno, H.G., Nakadai, K., and Kitano, H.: Research Issues of Humanoid Audition 第 7 回 AI チャレンジ研究会, SIG-Challenge-99-7, pp.61-66. 人工知能学会, Nov. 1999.
  9. 奥乃 博, 京田 耕司, 中臺 一博, 北野 宏明: Beowulf クラスクラスタ ERATO-1 のチューニングと評価, 「ハイパーフォーマンスコンピューティングのアーキテクチャと性能評価」に関する北海道ワークショップ (HOKKE-2000), pp.119-124, 情報処理学会, 計算機アーキテクチャ 137-21, ハイパーフォーマンスコンピューティング 80-21, 千歳科学技術大学, Mar, 2000.
  10. 奥乃 博, 中臺 一博, 北野 宏明: ロボットの耳は 2 つで十分か, 日本音響学会誌 小特集 - なぜ耳は二つあるか? -, vol. 58, no.3, pp.205-210, 日本音響学会, 2002.
  11. 中臺 一博, 奥乃 博, 北野 宏明: アクティブオーディションに基づくヒューマン・ロボットインタラクション, 第 3 回システムインテグレーション (SI2002), 計測自動制御学会 Dec, 2002.
  12. 中臺 一博, 松浦 大輔, 宮下 敬宏, 奥乃 博, 北野 宏明: ロボット聴覚実現に向けて - ヒューマノイドロボット SIG2 の開発 - 日本ロボット学会学術講演回, Oct. 2002.
  13. 中臺 一博, 松浦 大輔, 奥乃 博, 北野 宏明: 聴覚中心窩を利用した方向通過型フィルタによる音源定位・分離日本ロボット学会学術講演回, Oct. 2002.
  14. 松浦 大輔, 中臺 一博, 奥乃 博, 北野 宏明: サポートベクターマシンを用いた話者トラッキング注意制御日本ロボット学会学術講演回, Oct. 2002.
  15. 中臺一博, 日台健一, 奥乃 博, 北野宏明: 視聴覚のストリームベース統合による実時間人物追跡システム. 日本ロボット学会学術講演回, Sep. 2001.
  16. 日台健一, 中臺一博, 奥乃 博, 北野宏明: ステレオ視による実時間人物追跡システムの高精度化. 日本ロボット学会学術講演回, Sep. 2001.
  17. 中臺 一博, 奥乃 博, 北野 宏明: ヒューマノイド聴覚機能のためのアクティブオーディションの提案. 日本ロボット学会学術講演回, 1C31, 草津, Sep. 2000.
  18. 中臺 一博, 奥乃 博, 北野 宏明: ロボットの外装の音響効果を利用したアクティブオーディションシステム. 日本ロボット学会学術講演回, 1C32, 草津, Sep. 2000.
  19. 中川 友紀子, 中臺 一博, 奥乃 博, 北野 宏明: 動画像と音響を統合したインタラクティブロボットの動的環境での制御. 日本ロボット学会学術講演回, 1C36, 113-116, 草津, Sep. 2000.
  20. 人間主体の知的情報技術に関する調査研究 IV, (財) 日本情報処理開発協会 先端情

報技術研究所, Mar. 2003.

[書籍] 1 件

1. 北野宏明監訳, 奥乃 博, 諸橋 峰雄, 京田 耕司, 中臺 一博 訳: 『PC クラスタ構築法 - Linux クラスタベオウルフ』産業図書, March, 2001.

[受賞] 6 件

1. IS-2000 Best Paper Award (Oct. 2000)
2. 人工知能学会研究奨励賞 (May 2001)
3. International Society of Applied Intelligence, IEA/AIE-2001 Best Paper Award (1st Prize) (Jun. 2001)
4. 電気通信普及財団研究奨励賞 (Mar. 2002)
5. International Conference on Intelligent Robots and Systems (IROS 2001) BEST PAPER Nomination Finalist (Oct. 2002)
6. (財) 船井情報科学振興財団 船井情報科学振興賞 (Mar. 2003)

[特許]

- 国内特許 8 件
  1. 特願平 11-341240 ロボット聴覚装置
  2. 特願 2000-173915 ロボット聴覚装置及びロボット聴覚システム
  3. 特願 2001-067846 ロボット視聴覚システム
  4. 特願 2001-067847 ロボット視聴覚システム
  5. 特願 2001-067848 ロボット視聴覚システム
  6. 特願 2001-067849 ロボット視聴覚システム
  7. 特願 2001-107696 鳴き声による音声ガイダンスシステム
  8. 特願 2002-56670 ロボット視聴覚システム
- PCT 特許 5 件 (ERATO 内整理番号)
  1. E05506 PCT (特願平 11-341240)
  2. E05511 PCT (特願 2000-173915)
  3. E05515 PCT (特願 2001-067846,067847,067848,067849)
  4. E05516 PCT (特願 2001-107696)
  5. E05519 PCT (特願 2002-56670)

[マスメディア] 11 件

1. 日経産業新聞, 2000-03-10, 「視聴覚備え動き自然高性能ロボット開発」

2. 産経新聞, 2000-05-16, 「聞き分けのいいロボット登場」
3. 日経産業新聞, 2000-11-29, 「ロボットに賢い耳」
4. siliconindia, 2001-08-04, - News "Japan researchers develop receptionist robot-paper"
5. 日経新聞 夕刊一面, 2001-08-04, 「聞き上手ロボット」
6. 日経メカニカル, 2001 8 No.563 pp.32 「視聴覚を統合して話者を追う」
7. TV Tokyo, 2002-01-05, "Navigator 21"
8. MX TV, 2002-01-25, "Galileo Channel"
9. NHK BS-1, 2002-06-30, "What's on Japan"
10. 朝日新聞西日本版 夕刊一面, 2002-10-3, 「聖徳太子ロボット」
11. Newton, 2003 4 月号, 「生命科学とロボットの接点を探る - 北野共生システムプロジェクトを訪ねて -」

## 本研究に直接関係しない文献一覧 (1999 年 7 月以前)

### [学位論文]

1. 中臺一博: 音楽音響信号を対象とする音源分離システムの研究 ~音モデルに基づくアプローチ~, 卒業論文, 東京大学工学部, 1993.
2. 中臺一博: 音楽演奏における単音の分離抽出とその音楽情景分析システムへの応用, 修士論文, 東京大学工学系研究科, 1995.

### [ジャーナル論文] 2 件

1. 柏野邦夫, 中臺一博, 木下智義, 田中英彦: 音楽情景分析の処理モデル OPTIMA における単音の認識, 電子情報通信学会論文誌 (D-II), Vol.J79-D-II, No.11, pp.1751-1761, 1996.
2. 柏野 邦夫, 木下 智義, 中臺 一博, 田中 英彦「音楽情景分析の処理モデル OPTIMA における和音の認識」電子情報通信学会論文誌, Vol.J79-D-II, No.11, pp.1762-1770, 1996.

### [国際学会] 2 件

1. Kunio Kashino, Kazuhiro Nakadai, Tomoyoshi Kinoshita and Hidehiko Tanaka: Organization of Hierarchical Perceptual Sounds, Proc. of the 14th Int. Joint Conf. on Artificial Intelligence (IJCAI-95), vol.1, pp.158-164, 1995.

2. Kunio Kashino, Kazuhiro Nakadai, Tomoyoshi Kinoshita and Hidehiko Tanaka: Application of Bayesian Probability Network to Music Scene Analysis Workshop, pp.52-59,1995.

[研究会, Whorkshop, 全国大会] 15 件

1. 中臺一博, 柏野邦夫, 田中英彦: “音楽音響信号を対象とする音源分離システム～音モデルに基づくアプローチ～”, 情処研報 SIGMUS 1-1 (1993).
2. K.Nakadai, H., K.Kashino: “Parallel Pattern Matching in Sound Source Separation System on AP1000”, in *Proc. of the Third Parallel Computing Workshop*, pp. P2-R, Fujitsu Parallel Computing Reseach Facilities Fujitsu Laboratories LTD. (1994).
3. 中臺一博, 柏野邦夫, 田中英彦: “音源分離システムにおけるパターン照合モジュールの動的負荷分散を用いた並列実装”, SWoPP'94, 情報処理学会 94-AI-95-7, pp. 59-60 (1994).
4. 柏野邦夫, 中臺一博, 田中英彦: “音楽音響信号から単音記号列を生成するシステム OPTIMA の全体像”, 情処研報 94, Vol. 71, pp. 57-64 (1994).
5. 渡辺浩史, 中臺一博, 佐藤幸男, 坂口全志, 芦川宏利: ATM ノードの連続高速呼設定処理におけるシステムリソースの保護, CSPY98-151, 1999.
6. 中臺一博, 柏野邦夫, 田中英彦: “音源分離システムにおける音モデルの自動獲得 I: 音モデルの抽象化”, 第 47 回情処全大,1R-9 (1993).
7. 柏野邦夫, 中臺一博, 田中英彦: “音源分離システムにおける音モデルの自動獲得 II:ボトムアップ処理と音モデルに基づく処理との協調”, 第 47 回情処全大,1R-10 (1993).
8. 中臺一博, 柏野邦夫, 田中英彦: “音源分離システムにおけるパターン照合モジュールの並列実装と評価”, 第 48 回情処全大,4T-6 (1994).
9. 中臺一博, 柏野邦夫, 田中英彦: “音楽単音記号列生成システムの処理モデル OPTIMA における単音仮説生成処理”, 第 49 回情処全大,1B-4 (1994).
10. 柏野邦夫, 中臺一博, 田中英彦: “OPTIMA: 音楽音響信号から単音記号列を生成するシステムの処理モデル”, 第 49 回情処全大,1B-3 (1994).
11. 中臺一博, 柏野邦夫, 木下智義, 田中英彦: “音楽情景分析の処理モデル OPTIMA における統計的単音仮説生成処理”, 第 50 回情処全大,6D-4 (1995).
12. 柏野邦夫, 中臺一博, 木下智義, 田中英彦: “音楽情景分析の処理モデル OPTIMA の実装”, 第 50 回情処全大,6D-2 (1995).
13. 中臺一博, 柏野邦夫, 木下智義, 田中英彦: “処理モデル OPTIMA に基づく音楽情景分析の構築 I:各処理モジュールの実現と評価”, 音響学会 (1995).

14. 柏野邦夫, 中臺一博, 木下智義, 田中英彦: “処理モデル OPTIMA に基づく音楽情景分析の構築 II:情報統合の実現と評価”, 音響学会 (1995).
15. 木下智義, 柏野邦夫, 中臺一博, 田中英彦: “音楽情景分析の処理モデル OPTIMA におけるシーン情報の抽出と利用”, 第 50 回情処全大,6D-3 (1995).



## 参考文献

- [1] P. Aarabi and S. Zaky. Robust sound localization using multi-source audiovisual information fusion. *Information Fusion*, Vol. 2, No. 3, pp. 209–223, 2001.
- [2] 安部素嗣, 安藤繁. 共有 FM-AM の時間周波数統合に基づく聴覚情景解析 (I) – Lagrange 微分特微量とその周波数軸統合 –. 電子情報通信学会論文誌, Vol. J83-D-II, No. 2, pp. 458–467, 2000.
- [3] 安部素嗣, 安藤繁. 共有 FM-AM の時間周波数統合に基づく聴覚情景解析 (II) – 最適な時間軸統合とストリーム音の再合成 –. 電子情報通信学会論文誌, Vol. J83-D-II, No. 2, pp. 468–477, 2000.
- [4] 秋田祐哉, 河原達也. 会議音声の自動アーカイブ化システム. 電子情報通信学会技術研究報告 (NLC2000-37/SP2000-85), 12 2000.
- [5] Y. Aloimonos, I. Weiss, and A. Bandyopadhyay. Active vision. *International Journal of Computer Vision*, 1987.
- [6] Y. Aloimonos(Ed). *Active Perception*. Lawrence Erlbaum Associates, Publisher, NJ., 1993.
- [7] S. Amari and J. F. Cardoso. Blind source separation – semiparametric statistical approach. *IEEE Trans. on Signal Processing*, Vol. 45, No. 11, pp. 2692–2700, 1997.
- [8] S. Amari, A. Cichocki, and H. Yang. A new learning algorithm for blind signal separation, in Eds. David S. Touretzky, Michakel C. Mozer and Michael E. Hasselmo. *Advances in Neural Information Processing Systems*, Vol. 8, pp. 757–763, 1996.
- [9] S. Ando. An autonomous three-dimensional vision sensor with ears. In *IAPR Workshop on Machine Vision Applications(MVA-94)*, pp. 417–422, 1994.
- [10] S. Ando. An autonomous three-dimensional vision sensor with ears. *IEICE Transactions on Information and Systems*, Vol. E78–D, No. 12, pp. 1621–1629, 1995.

- [11] M. Aoki, M. Okamoto, S. Aoki, H. Matusi, T. Sakurai, and Y. Kaneda. Sound source segregation based on estimating incident angle of each frequency component of input signals acquired by multiple microphones. *Acoust. Sci. and Tech.*, Vol. 22, No. 2, pp. 149–157, 2001.
- [12] 浅野太, 鈴木陽一, 曾根敏夫. 適応マイクロホンアレイを用いた補聴器のための雑音抑圧手法に関する一考察. 電子情報通信学会技術研究報告, 第 EA90-68 巻, 1990.
- [13] F. Asano, M. Goto, K. Itou, and H. Asoh. Real-time sound source localization and separation system and its application to automatic speech recognition. In *Proceedings of International Conference on Speech Processing (Eurospeech 2001)*, pp. 1013–1016. ESCA, Sep. 2001.
- [14] Isaac Asimov. *I, robot*. Fawcett Publications, 1950.
- [15] H. Asoh, S. Hayamizu, I. Hara, Y. Motomura, S. Akaho, and T. Matsui. Socially embedded learning of the office-conversant mobile robot *jijo-2*. In *Proceedings of 15th International Joint Conference on Artificial Intelligence (IJCAI-97)*, Vol. 1, pp. 880–885. AAAI, 1997.
- [16] 東貞男, 打越聰, 城戸健一. 交差形円弧状配列マイクロホンシステムの実用化に関する実験的検討. 日本音響学会誌, Vol. 46, No. 2, pp. 111–120, 1990.
- [17] R. Bajcsy. Active perception. *Proceedings of the IEEE*, Vol. 76, No. 8, pp. 996–1006, 1988.
- [18] J. Barker, M. Cooke, and P. Green. Robust asr based on clean speech models: An evaluation of missing data techniques for connected digit recognition in noise. In *Proc. of 7th European Conference on Speech Communication Technology (EUROSPEECH-01)*, Vol. 1, pp. 213–216. ESCA, 2001.
- [19] J. Bates. The role of emotion in believable agents. *Communications of the ACM*, Vol. 37, No. 7, pp. 122–125, 1994.
- [20] J. Blauert. *Spatial Hearing*. The MIT Press, 1999.
- [21] S. F. Boll. A spectral subtraction algorithm for suppression of acoustic noise in speech. In *Proceedings of 1979 International Conference on Acoustics, Speech, and Signal Processing (ICASSP-79)*, pp. 200–203. IEEE, 1979.
- [22] C. Breazeal. Emotive qualities in robot speech. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2001)*, pp. 1389–1394. IEEE, 2001.
- [23] C. Breazeal and B. Scassellati. A context-dependent attention system for a social robot. In *Proc. of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI-99)*, pp. 1146–1151, 1999.

- [24] A. S. Bregman. *Auditory Scene Analysis*. The MIT Press, MA., 1990.
- [25] R. A. Brooks, C. Breazeal, R. Irie, C. C. Kemp, M. Marjanovic, B. Scassellati, and M. M. Williamson. Alternative essences of intelligence. In *Proceedings of 15th National Conference on Artificial Intelligence (AAAI-98)*, pp. 961–968. AAAI, 1998.
- [26] R. Brooks, C. Breazeal, M. Marjanovic, B. Scassellati, and M. Williamson. The cog project: Building a humanoid robot. In C.L. Nehaniv, editor, *Computation for metaphors, analogy, and agents*, pp. 52–87. Spriver-Verlag, 1999.
- [27] G. J. Brown. *Computational auditory scene analysis: A representational approach*. University of Sheffield, 1992.
- [28] J. Cassell. More than just another pretty face: Embodied conversational interface agents. *Communications of the ACM*, Vol. 43, No. 4, pp. 70–78, 2000.
- [29] S. Cavaco and J. Hallam. A biologically plausible acoustic azimuth estimation system. In *Proceedings of IJCAI-99 Workshop on Computational Auditory Scene Analysis (CASA '99)*, pp. 78–87. IJCAI, 1999.
- [30] E. C. Cherry. Some experiments on the recognition of speech, with one and with two ears. *Journal of Acoustic Society of America*, Vol. 25, pp. 975–979, 1953.
- [31] P. Comon. Independent component analysis, a new concept? *Signal Processing*, Vol. 36, No. 3, pp. 287–314, 1994.
- [32] M. P. Cooke, G. J. Brown, M. Crawford, and P. Green. Computational auditory scene analysis: Listening to several things at once. *Endeavour*, Vol. 17, No. 4, pp. 186–190, 1993.
- [33] A.P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, Vol. 38, pp. 325–339, 1967.
- [34] O. D. Faugeras. *Three Dimensional Computer Vision: A Geometric Viewpoint*. The MIT Press, MA., 1993.
- [35] J.L. Flanagan, D.A. Berkley, G.W. Elko, J.E. West, and M.M. Sondhi. Autodirective microphone systems. *Acustica*, Vol. 73, No. 2, pp. 58–71, 1991.
- [36] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 1972.
- [37] 福留公利, 大久保忠俊, 河原一彦. 球の回折情報を利用した音源の方向推定に関する研究 (MUSIC アルゴリズムの適用). 電子情報通信学会技術研究報告, 第 EA91-36 巻, 1991.
- [38] J.E. Greenberg. *A real time adaptive beamforming hearing aid*. Massachusetts

- Institute of Technology, 1989.
- [39] L.J. Griffiths and C.W. Jim. An alternative approach to linearly constrained adaptive beamforming. *IEEE Transactions on Antennas and Propagation*, Vol. AP-30, No. 8, pp. 27–34, 1982.
- [40] E. T. Hall. *The Hidden Dimension*. Anchor books doubleday, 1966.
- [41] 浜田 晴夫 *et al.* 特集: 静かで快適な生活を求めて – 最近のアクティブコントロール技術–. 情報処理, Vol. 40, No. 1, pp. 1–31, 1999.
- [42] B. Hayes-Roth, G. Ball, C. Lisetti, R. Picard, and A. Stern. Affect and emotion in the user interface. In *Proceedings of 1998 International Conference on Intelligent User Interfaces*, pp. 91–96. ACM, 1998.
- [43] J. Hershey, H. Ishiguro, and J. R. Movellan. Audio vision: Using audio-visual synchrony to locate sounds. In *Neural Information Processing Systems*, Vol. 12, pp. 813 – 819. MIT Press, 2000.
- [44] K. Hidai, H. Mizoguchi, K. Hiraoka, M. Tanaka, T. Shigehara, and T. Mishima. Robust face detection against brightness fluctuation and size variation. In *Proc. of IEEE/RAS Int. Conf. on Intelligent Robots and Systems (IROS-2000)*, pp. 1397–1384. IEEE, 2000.
- [45] K. Hiraoka, S. Yoshizawa, K. Hidai, M. Hamahira, H. Mizoguchi, and T. Mishima. Convergence analysis of online linear discriminant analysis. In *Proc. of IEEE/INNS/ENNS Int. Joint Conference on Neural Networks*, pp. III-387–391. IEEE, 2000.
- [46] 本谷秀堅, 来海暁, 安藤繁. 固視微動型イメージセンサとその応用. 研究報告「コンピュータビジョンとイメージメディア」, Vol. No.118-002, , 1999.
- [47] J. Huang, N. Ohnishi, and N. Sugie. Building ears for robots: sound localization and separation. *Artificial Life and Robotics*, Vol. 1, No. 4, pp. 157–163, 1997.
- [48] 池田徹志, 石黒浩, 浅田稔. 相互情報量最大化に基づく異種センサの信号の適応的統合. 第 19 回ロボット学会学術講演会, 第 2F23 巻, 2001.
- [49] M. Z. Ikram and D. R. Morgan. A multiresolution approach to blind separation of speech signals in a reverberant environment. In *Proceedings of 2001 International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2001)*, pp. 2757–2760. IEEE, 2001.
- [50] 今井聖. 情報・電子入門シリーズ 16 音声認識. 共立出版, 1995.
- [51] M. Inaba, T. Ninomiya, Y. Hoshino, K. Nagasaka, S. Kagami, and H. Inoue. A remote-brained full-body humanoid with multisensor imaging system of binocular viewerearswrist force and tactile sensor suit. In *Proceedings of the 1997*

- IEEE Int. Conf. on Robotics and Automation*, pp. 2497–2502, 1997.
- [52] R. E. Irie. Multimodal sensory integration for localization in a humanoid robot. In *Proceedings of the Second IJCAI Workshop on Computational Auditory Scene Analysis (CASA '97)*, pp. 54–58. IJCAI, 1997.
- [53] T. Irino and R.D. Patterson. A time-domain, level-dependent auditory filter: The gammachirp. *J. Acoust. Soc. Am.*, Vol. 101, pp. 412–419, 1997.
- [54] T. Irino and R.D. Patterson. Stabilised wavelet mellin transform: An auditory strategy for normalising sound-source size. In *Proc. of 6th European Conference on Speech Communication Technology (EuroSpeech-99)*, pp. 1899–1902. ESCA, 1999.
- [55] 石塚満. Dempster & Shafer の確率理論. 電子通信学会誌, Vol. 66, No. 9, pp. 900–903, 1983.
- [56] L.A. Jeffress. A place theory of sound localization. *Journal of Comparative Physiology, Psychology*, Vol. 41, pp. 35–39, 1948.
- [57] C. Jutten and J. Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, Vol. 24, No. 1, pp. 1–10, 1991.
- [58] S. Kagami, K. Okada, M. Inaba, and H. Inoue. Real-time 3d optical flow generation system. In *Proc. of Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems (MFI'99)*, pp. 237–242, 1999.
- [59] 金森文郎, 茨木悟, 古川博之, 斉藤浩, 西川清. 2次元デジタルフィルタを用いた超指向性マイクロホン. 電子情報通信学会技術研究報告, 第 EA91-84 巻, 1992.
- [60] Y. Kaneda and J. Ohga. Adaptive microphone-array system for noise reduction. *IEEE Transactions on Acoustics Speech Signal Processing*, Vol. ASSP-34, No. 6, pp. 1391–1400, 1986.
- [61] 柏野邦夫, 中臺一博, 木下智義, 田中英彦. 音楽情景分析の処理モデル OPTIMA における単音の認識. 電子情報通信学会論文誌 (D-II), Vol. J79-D-II, No. 11, pp. 1751–1761, 1996.
- [62] 柏野邦夫, 木下智義, 中臺一博, 田中英彦. 音楽情景分析の処理モデル OPTIMA における和音の認識. 電子情報通信学会論文誌 (D-II), Vol. J79-D-II, No. 11, pp. 1762–1770, 1996.
- [63] K. Kashino, K. Nakadai, T. Kinoshita, and H. Tanaka. Application of bayesian probability network to music scene analysis. In *Working Notes of the IJCAI-95 Computational Auditory Scene Analysis Workshop*, pp. 52–59. AAAI, 1995.
- [64] D.J. Kiesler. The 1982 interpersonal circle: A taxonomy for complementarity

- in human transactions. *Psychological Review*, Vol. 90, pp. 185–214, 1993.
- [65] H. Y. Kim, F. Asano, Y. Suzuki, and T. Sone. Speech enhancement based on short time spectral amplitude estimation with two-channel beamformer. *IEICE Trans. fundamentals*, Vol. E79-A, No. 12, 1996.
- [66] 来海暁, 安藤繁. ノンパラメトリックカルマンフィルタを用いた両眼視による多層立体再構成. 電子情報通信学会論文誌 D-II, Vol. J84-D-II, No. 7, pp. 1328–1338, 2001.
- [67] Tetsuro Kitahara, Masataka Goto, and Hiroshi G. Okuno. Musical instrument identification based on f0-dependent multivariate normal distribution. In *Proc. of the 2003 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP-2003)*, 2003.
- [68] W.N. Klarquist and A.C. Bovik. Fovea: A foveated vergent active stereo vision system for dynamic 3-dimensional scene recovery. *RA*, Vol. 14, No. 5, pp. 755–770, October 1998.
- [69] 久野義徳. アクティブビジョン—歴史と展望—. 人工知能学会誌, Vol. 10, No. 4, pp. 493–499, 1995.
- [70] T. Kurata, D. Chang, and S. Hashimoto. Multimedia sensing system for robot. In *Proceedings of 4th IEEE International Workshop on Robot and Human Communication (RO-MAN '95)*, pp. 83–88. IEEE, 1995.
- [71] T. Lourens, K. Nakadai, H. G. Okuno, and H. Kitano. Selective attention by integration of vision and audition. In *Proceedings of First IEEE-RAS International Conference on Humanoid Robot (Humanoid-2000)*. IEEE/RSJ, 2000.
- [72] A. B. Loyall and J. Bates. Personality-rich believable agents that use language. In *Proceedings of the First International Conference on Autonomous Agents*, pp. 106–113, 1997.
- [73] J. C. Makous and J. C. Middlebrooks. Two-dimensional sound localization by human listeners. *Journal of the Acoustical Society of America*, Vol. 87, No. 5, pp. 2188–2200, 1990.
- [74] Y. Matsusaka, T. Tojo, S. Kuota, K. Furukawa, D. Tamiya, K. Hayata, Y. Nakano, and T. Kobayashi. Multi-person conversation via multi-modal interface — a robot who communicates with multi-user. In *Proc. of 6th European Conference on Speech Communication Technology (EUROSPEECH-99)*, pp. 1723–1726. ESCA, 1999.
- [75] H. McGurk and J. MacDonald. Hearing lips and seeing voices. In *Nature*, Vol. 264, pp. 746–748, 1976.

- [76] H. Miwa, T. Okuchi, H. Takanobu, and A. Takanishi. Development of a new human-like head robot we-4. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2002)*, pp. 2443–2448. IEEE, 2002.
- [77] B.C.J. ムーア (大串他訳). 聴覚心理学概論. 誠信書房, 1994.
- [78] N. Murata and S. Ikeda. An on-line algorithm for blind source separation on speech signals. In *Proceedings of 1998 International Symposium on Nonlinear Theory and its Applications*, pp. 923–927, 1998.
- [79] K. Nakadai, K. Hidai, H. Mizoguchi, H. G. Okuno, and H. Kitano. Real-time auditory and visual multiple-object tracking for robots. In *Proc. of the 17th Int. Joint Conf. on Artificial Intelligence (IJCAI-01)*, pp. 1424–1432. MIT Press, 2001.
- [80] K. Nakadai, H. G. Okuno, and H. Kitano. Exploiting auditory fovea in humanoid-human interaction. In *Proceedings of 18th National Conference on Artificial Intelligence (AAAI-2002)*, pp. 431–438. AAAI, 2002.
- [81] Y. Nakagawa, H. G. Okuno, and H. Kitano. Using vision to improve sound source separation. In *Proc. of 16th National Conference on Artificial Intelligence (AAAI-99)*, pp. 768–775. AAAI, 1999.
- [82] T. Nakatani and H. G. Okuno. Sound ontology for computational auditory scene analysis. In *Proceedings of 15th National Conference on Artificial Intelligence (AAAI-98)*, pp. 1004–1010. AAAI, 1998.
- [83] T. Nakatani and H. G. Okuno. Harmonic sound stream segregation using localization and its application to speech stream segregation. *Speech Communication*, Vol. 27, No. 3-4, pp. 209–222, 1999.
- [84] T. Nakatani, H. G. Okuno, and T. Kawabata. Auditory stream segregation in auditory scene analysis with a multi-agent system. In *Proceedings of 12th National Conference on Artificial Intelligence (AAAI-94)*, pp. 100–107. AAAI, 1994.
- [85] T. Nakatani, H. G. Okuno, and T. Kawabata. Residue-driven architecture for computational auditory scene analysis. In *Proceedings of 14th International Joint Conference on Artificial Intelligence (IJCAI-95)*, Vol. 1, pp. 165–172. AAAI, 1995.
- [86] 中谷智広, 奥乃博, 川端豪. 音環境理解のためのマルチエージェントによる調波構造ストリームの分離. 人工知能学会誌, Vol. 10, No. 2, pp. 232–241, Mar. 1995.
- [87] P. A. Nelson and S. J. Elliott. *Active Control of Sound*. ACADEMIC PRESS,

- 1992.
- [88] 西隆司, 三上淳一, 井上友幸, 古川宣一, 清水寧, 川上福司. 多方向同時集音装置 (マルチビームアレーマイクロホン). 電子情報通信学会技術研究報告, 第 EA88-65 巻, 1988.
- [89] 西一樹, 安藤繁. 定 Q 楕形フィルタとその時間/周波数特性—ピッチ変動誤差に頑健な調波信号推定のためのフィルタ特性—. 電子情報通信学会論文誌 A, Vol. J83-A, No. 2, 2000.
- [90] R. Nishimura, T. Uchida, A. Lee, H. Saruwatari, and K. Shikano. Aska: Receptionist robot with speech dialogue system. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2002)*, pp. 1308–1313. IEEE, 2002.
- [91] T. Nishiura, M. Nakamura, A. Lee, H. Saruwatari, and K. Shikano. Talker tracking display on autonomous mobile robot with a moving microphone array. In *Proceedings of the 2002 International Conference on Auditory Display (ICAD 2002)*, 2002.
- [92] 岡田慧, 加賀美聡, 稲葉雅幸, 井上博允. PC による高速対応点探索に基づくロボット搭載可能な実時間視差画像・フロー生成法と実現. 日本ロボット学会誌, Vol. 18, No. 6, pp. 138 – 143, 2000.
- [93] H. G. Okuno, T. Nakatani, and T. Kawabata. Interfacing sound stream segregation to speech recognition systems — preliminary results of listening to several things at the same time. In *Proceedings of 13th National Conference on Artificial Intelligence (AAAI-96)*, pp. 1082–1089. AAAI, 1996.
- [94] 奥乃博. 音環境理解 — 混合音の認識を目指して. 情報処理, Vol. 40, No. 10, pp. 1096–1101, Oct. 2000.
- [95] H.G. Okuno, S. Ikeda, and T. Nakatani. Combining independent component analysis and sound stream segregation. In *Proc. of IJCAI-99 Workshop on Computational Auditory Scene Analysis (CASA'99)*, pp. 92–98. IJCAI, 1999.
- [96] H.G. Okuno, K. Nakadai, T. Lourens, and H. Kitano. Separating three simultaneous speeches with two microphones by integrating auditory and visual processing. In *Proc. of European Conf. on Speech Processing (Eurospeech 2001)*. ESCA, 2001.
- [97] H.G. Okuno, Kazuhiro Nakadai, Tino Lourens, and Hiroaki Kitano. Separating three simultaneous speeches with two microphones by integrating auditory and visual processing. In *Proceedings of International Conference on Speech Processing (Eurospeech 2001)*, pp. 2643–2646. ESCA, Sep. 2001.



- [98] M. Okutomi. Stereo vision. In T. Matsuyama, Y. Kuno, and J. Imiya, editors, *Computer Vision*, pp. 123–137. New Technology Communications, 1998.
- [99] 小野順貴, 安藤繁. 基本周波数によらない調波間干渉を生じるフィルタバンクの理論. *日本音響学会誌*, Vol. 57, No. 10, pp. 637–648, 2001.
- [100] T. Ono, M. Imai, and H. Ishiguro. A model of embodied communications with gestures between humans and robots. In *Proceedings of Twenty-third Annual Meeting of the Cognitive Science Society (CogSci2001)*, pp. 732–737. AAAI, 2000.
- [101] 大賀寿郎, 山崎芳男, 金田豊. 音響システムとデジタル処理. 電子情報通信学会, 1995.
- [102] H.E. Pashler. *The Psychology of Attention*. The MIT Press, MA., 1997.
- [103] S. Perrett and W. Noble. The effect of head rotations on vertical plane sound localization. *Journal of the Acoustical Society of America*, Vol. 102, pp. 2325–2332, 1997.
- [104] B. Reeves and C. Nass. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press, Cambridge, UK, 1996.
- [105] Philippe Renevey, Rolf Vetter, and Jens Kraus. Robust speech recognition using missing feature theory and vector quantization. In *Proc. of 7th European Conference on Speech Communication Technology (EUROSPEECH-01)*, Vol. 2, pp. 1107–1110. ESCA, 2001.
- [106] D. Rosenthal and H. G. Okuno, editors. *Computational Auditory Scene Analysis*. Lawrence Erlbaum Associates, Mahwah, New Jersey, 1998.
- [107] S. Rougeaux and Y. Kuniyoshi. Robust real-time tracking on an active vision head. In *Proc. of IEEE/RAS Int. Conf. on Intelligent Robots and Systems (IROS-97)*, pp. 873–879. IEEE, 1997.
- [108] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and K. Fujimura. The intelligent asimo: System overview and integration. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2002)*, pp. 2478–2483, 2002.
- [109] H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura. Speech enhancement using nonlinear microphone array based on complementary beamforming. *IEICE Trans. fundamentals*, Vol. E82-A, No. 8, 1999.
- [110] G. Schuller and G. Pollak. Disproportionate frequency representation in the inferior colliculus of horseshoe bats: evidence for an “acoustic fovea”. In *J.*

- Comp. Physiol. A*, Vol. 132, pp. 47–54, 1979.
- [111] 瀬名秀明. 明日のロボット. 文藝春秋, Oct. 2002.
- [112] M. Slaney, D. Naar, and R. F. Lyon. Auditory model inversion for sound separation. In *Proceedings of 1994 International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, pp. 77–80, 1994.
- [113] D. E. Sturim, M. S. Brandstein, and H. F. Silverman. Tracking multiple talkers using microphone-array measurements. In *Proceedings of 1997 International Conference on Acoustics, Speech, and Signal Processing (ICASSP-97)*. IEEE, 1997.
- [114] 高西淳夫, 升川聡, 森寧, 小川太郎. 人間形聴覚ロボットの研究. 第 11 回日本ロボット学会学術講演会, pp. 793–796. *Int. Soc. Opt. Eng.*, 1994.
- [115] A. Takanishi, S. Masukawa, Y. Mori, and T. Ogawa. Development of an anthropomorphic auditory robot that localizes a sound direction (*in japanese*). *Bulletin of the Centre for Informatics*, Vol. 20, pp. 24–32, 1995.
- [116] 田中穂積. 言語理解と行動制御に関する研究 (課題番号 12NP9201). 平成 12 年度科学研究費補助金 (創成的基礎研究), Mar. 2001.
- [117] K. Tanaka, M. Abe, and S. Ando. A novel mechanical cochlea “fishbone” with dual sensor/actuator characteristics. *IEEE Transactions on Mechatronics*, Vol. 3, No. 2, pp. 98–105, 1998.
- [118] 植松尚, 加藤正晴, 柏野牧夫, 平原達也. 頭外音像定位における自発的な頭部回転の影響. 日本音響学会講演論文集, pp. 501–502, 2001.
- [119] T. Usagawa, K. Sakai, and M. Ebata. Frequency domain binaural model as the front end of speech recognition system. In *Proceedings of International Conference on Spoken Language Processing (ICSLP-1998)*. ISCA, 1998.
- [120] S. Waldherr, S. Thrun, R. Romero, and D. Margaritis. Template-based recognition of pose and motion gestures on a mobile robot. In *Proceedings of 15th National Conference on Artificial Intelligence (AAAI-98)*, pp. 977–982. AAAI, 1998.
- [121] F. Wang, Y. Takeuchi, N. Ohnishi, and N. Sugie. A mobile robot with active localization and discrimination of a sound source. *Journal of Robotic Society of Japan*, Vol. 15, No. 2, pp. 61–67, 1997.
- [122] J. Weizenbaum. Eliza – a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, Vol. 9, No. 1, pp. 36–45, 1966.
- [123] J.M. Wolfe, K. R. Cave, and S.L. Franzel. Guided search: An alternative to the

- feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 15, No. 3, pp. 419–433, 1989.
- [124] 河原英紀. 聴覚の情景分析が生んだ高品質 VOCODER: STRAIGHT. 日本音響学会誌, Vol. 54, No. 7, pp. 521–526, 1998.
- [125] 鹿野清宏, 伊藤克亘, 河原達也, 武田一哉, 山本幹雄. 音声認識システム. オーム社, 2001.
- [126] 浅井泰行, 中島弘道, 山村毅, 黄捷, 大西昇. 運動を介した視聴覚による物体定位能力の獲得. *IEICE Trans. D-II*, Vol. j83-D-II, No. 7, pp. 1676–1684, 2000.
- [127] 柏野邦夫. 音楽音響信号を対象とする聴覚的情景分析に関する研究. PhD thesis, 東京大学大学院工学系研究科, 1994.