

音声模倣発達過程の理解に向けた認知発達ロボティクスの試み

吉川雄一郎, 浅田稔 (JST ERATO 浅田共創知能システムプロジェクト)

要旨 ヒトの乳児は周りの大人とのどのような相互作用を通じて、またどのような仕組みで、大人が話す言葉を獲得するのか。本稿では、この問題に対して、従来の観察に基づくアプローチを補うことが期待されている認知発達ロボティクスでの取り組みを取り上げる。はじめに、親との相互模倣を通じて乳児が母音を獲得していく過程を構成する研究について紹介し、親が乳児を模倣することの役割と仕組みについて議論する。次に、乳児に対する物の提示や物の名前などの働きかけを含む、より自然な養育者の振る舞いのもとで音声模倣および語彙を獲得する過程を構成する研究を紹介し、これらの共発達を可能にする仕組みについて議論する。

キーワード 音声模倣, 語彙獲得, 相互模倣, 知覚運動バイアス, 自己模倣バイアス, 共発達

1 はじめに

音声言語によるコミュニケーションを実現するためには、そこで交わされる言葉の単位系、すなわち対話者同士が音韻系や語彙を共有していること、あるいはその実感が必要がある。乳児は生後、音韻知覚能力を徐々に母国語に特化させていくこと [1]、これに並行して発話できる声を大人のそれに徐々に近づけていくこと [2] が知られている。またその収斂を待たず、乳児とその養育者の間に模倣的なやりとり [3, 4] や言葉によるやり取り [5] が見られ始め、養育者との相互作用が乳児の音韻発達に重要な役割を果たしていると考えられる [6]。では、養育者とのどのような相互作用、そして学習メカニズムが言葉の共有化を可能にしているのだろうか。

そのような問いは人の発達を知るという観点のみならず、人と関わる知的人工物を設計するという観

点からも非常に興味深い。しかし従来の発達心理学における観察に基づくアプローチでは、統制の困難さのため、長期にわたる発達のしくみを取り扱うことは一般に容易ではない。これに対し、認知発達ロボティクス [7] と呼ばれる分野では、そのような乳児の発達過程を構成的に理解することを目指し、音声模倣ができるようになるロボットを実現する研究が進められている。本稿では、筆者らのグループで進めてきた音韻と語彙の共有化、すなわち音声相互模倣発達の構成に関する研究について紹介し、議論する。

2 言葉の共有化の構成的課題

声を聴取する乳児の能力は、生誕直後は言語圏によらないユニバーサルなものであるが、徐々に乳児が属する言語圏の声を聴取するのに特化していく [1]。また、はじめは非音韻様の発話しかできないが、咽頭蓋の沈下 [8] に従って、徐々に養育者の声に適應していく [2]。これらの過程を経て、乳児は8か月になる頃に単母音を、14か月になる頃に連続する母音の模倣を示すようになると言われている [4]。この模倣発達の初期のマイルストーンに達するためには、自分の声と自分の声を産出するための構音運動の関連付けることだけでなく、生後の周りの大人とのやりとりの中から、親の声と自分の声の対応を見出すこと（音韻の共有化）が求められる。

またこの母音を模倣するようになる時期に前後して、乳児は大人が理解できる言葉を発しはじめる [5]（語彙の共有化）。親の声と自分の声の対応づけができることで、言葉の獲得が促進される可能性がある。一方、不完全にでも獲得された乳児の言葉は、寛容な解釈を伴った親の模倣を誘発し、逆に両者の声の対応の鮮鋭化も促進される可能性がある。これらが真ならば、これらを相互促進する重層的な共有化（音韻と語彙の共有化の共発達）のしくみを持つことが求められる。

〒 565-0871 大阪府吹田市山田丘2-1 大阪大学大学院工学研究科 FRC-1 401
吉川雄一郎
E-mail: yoshikawa.yuichiro@gmail.com

2.1 相互模倣による音韻共有化モデル

従来の構成的研究では、音韻共有化の発達には自身の発話運動と生成される音声の関係の知覚運動学習課題として取り扱われることが多かった [9, 10]. しかし実際の音韻共有化の発達過程では、発達する乳児（あるいはロボット）と、その発達の目標あるいは導き手となる養育者が、それぞれ異なる身体で声を発し、聞き取らざるを得ないことが無視できない。人の音声は、声帯に呼気が送り込まれることにより生じた空気の振動が、声道を通過する際に共鳴することによって産出されるため、その音響特性を表すフォルマント（共鳴周波数）は声道がどのように形作られているかによって決まる。従って、身体の違いのため、両者が発することができる声の音響特徴の範囲は異なり（図 1 参照）、養育者の声と物理的に同じ音響特性を再現しようとしても、それが産出可能範囲内のものであるとは限らない。

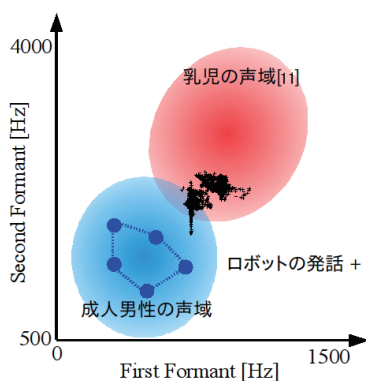


図 1 声域の比較：日本人成人男性 10 人による日本語母音発話の分布（五角形）、乳児の声域（文献 [11] の図から適応）、発話ロボット [12] の声域（十字の点）

また人は体内にある鼓膜の振動として音を聴取するため、相手の声についてはその気導音のみが知覚されるのに対し、自身の声については気導音と体内を伝わってきた骨導音が合わさった形で知覚される。従って、自分自身にとっては相手と物理的に同じに知覚される声を産出できたとしても、それから骨導音が除かれた気導音を聴取する他者にとっては、同じと知覚される音になっているとは限らない。従って、**物理的な音響特徴そのものを模倣の指標とするのではなく、乳児は何かに対応するのかをどのように発見可能かが重要な問いとなる。**

2.1.1 養育者が乳児を模倣することの意味

乳児が何を模倣すべきかを発見するのに、何がヒントになりうるのかを生後 6 か月までの親子のインタラクションの観察から考える。初期の親子インタラクションにおいては、養育者側からの模倣がより頻繁に見られることが報告されている [3]. また、乳児の発声が音韻様であるときに母親は高頻度で乳児の音声を模倣すること [13] や、母親の模倣が乳児の音韻様の発声の頻度を高めること [14] などが報告されている。これらのことから、乳児の音韻発達において養育者に自身の発話を模倣されることが重要な役割を果たしていることが示唆されるが、その具体的な機能については明らかでない。これに対し、我々の研究グループでは養育者による模倣に二つの役割があるとの仮説を提案している（図 2）。一つは、乳児の行動に対応する養育者の行動、すなわち身体構造の違いを吸収する対応づけを正しく学習するための正解例を乳児に例示する働きである。そしてもう一つは、養育者が乳児の発声を再現する時に身体構造の違いのため完全には再現できず、無意識のうちに自身の普段よくする行動、つまり母音の発話に置き換えた模倣をすることで、乳児の発話カテゴリを母音に誘導する働きである。

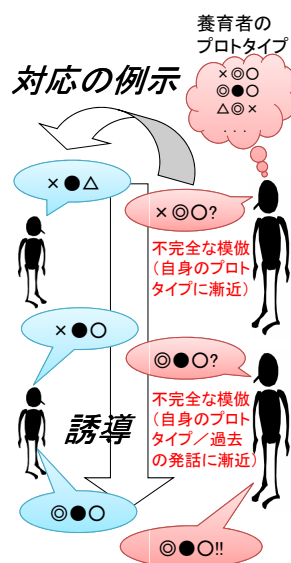


図 2 養育者の模倣の二つの役割

この考えに基づき Miura et al. [12] は、発話ロボット（図 3）を試作し、その発話を養育者役の実験者に模倣させることを通じて、人が同じと思える母音の発話をロボットに獲得させる問題に取り組んだ。先行研究 [15] を参考に試作したロボットは、人工声帯に圧縮空気を流入することで基本周波数の音源を生

成し、シリコン製の人工声道でこれを共鳴させることで音声を産出し、人工声道の喉・顎・舌にあたる筒の部分と唇にあたる開口部を電動モータにより変形させることで、様々なフォルマントの音を調音することができる。

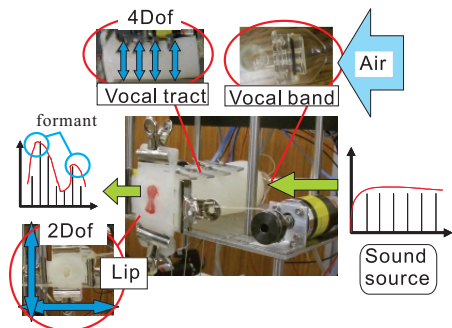


図 3 発話ロボット：Burpy

ロボットと人それぞれが生成可能な音声のフォルマントの分布範囲をおおよそ対応づける一次変換、および人の口の形を大まかに再現する視覚的な模倣メカニズムにより、ロボットは観測した人の母音発話を、人によっては母音と聞き取れる程度に模倣することができる。しかし、どのような変形により生成される音が、より母音様であると人に思わせるものであるかは不明であった。Miura et al.[12] は、これをフォルマント空間上の発話プロトタイプを適応させる問題と捉え、人の発話音声とその人の典型的な発話音声とどのように異なるかを基に、発話プロトタイプを移動させる学習メカニズムを提案した。ロボットの発話が人に模倣されることを仮定したインタラクション実験を実施し、実際にロボットの発話する母音をより明瞭なものへと誘導できたことが報告されている。

2.1.2 養育者の模倣メカニズム

Miura et al.[12] の研究では、本来未知であるはずの身体構造の対応付け、すなわち人とロボットの声域をおおまかに対応付ける写像が与えられていた。また、人の模倣が誘導的な働きを持つことが議論されていたが、どのようなしくみで人がそのような模倣を示すのかについての議論は十分ではなかった。模倣一般に言われているように、人は何かを模倣する際、感覚器からの信号をそのまま模写し、行動するわけではなく、なんらかの知覚運動バイアスがかかると考えられる。Ishihara et al.[16] はさらに相互作用がフィードバックを増幅する自己模倣バイアスと

いう概念を提案し、計算機シミュレーションによって、二つのバイアス (図.4) が乳児の、身体構造の対応付けの学習と並行した音韻共有化過程に及ぼす効果を検討している。

前者の知覚運動バイアスは、次の事前実験から判明した。シンセサイザによって生成された音声を被験者に聞かせ、単純な模倣を依頼しても、模倣音声は、被験者自身の母音の周りに集中することが明らかになった。すなわち人は、無意識のうちに他者の模倣音声を自身の音声に引き込んでいることがわかった。これは、知覚範疇によるマグネット効果 [17] により、連続的に分布している音を離散的な音カテゴリー (この場合、自身の各母音のいずれか) として聞こえてくること、これに加え発声する場合も、発声器官そのものの拘束を反映して範疇化されたモータコマンドにバイアスされることで、その効果が増幅されると考えられる。さらに相互模倣の過程で、曖昧な音声も、予測により、期待する音声に聞こえると仮定し、これが後者の自己模倣バイアスである。

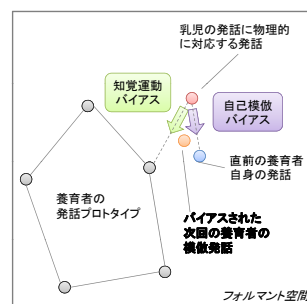


図 4 養育者の知覚運動バイアスと自己模倣バイアス

Oudeyer は、知覚マグネット効果を考慮した音韻認識メカニズムを持つエージェントの集団が、相互に模倣するやりとりを通じて、集団の中で音韻が共有化されていく過程を計算機上でシミュレートした [18]。しかし、エージェント同士は同じ構音機構を持つこと、各エージェントは代わる代わる様々なエージェントと模倣相互作用をすることが想定されている。人の乳児が置かれている状況はやりとりの相手が身体構造の異なる何人かの大人にある程度限られており、これについて直接シミュレートしたものではなかった。Ishihara et al.[16] はこの問題に対して、養育者との相互模倣を通じて、自身の声域と養育者の声域の対応付けを推定しながら、人が母音と聞き取れる発話の獲得を可能にする養育者の模倣メカニズムをモデル化することを試みている。

彼らは、計算機上で養育者役のエージェントと乳児役のエージェントを相互に模倣させた。図.5に、二

つのバイアスの強さを変えて相互作用させたときの、養育者と乳児が発声した発話音声のフォルマントの分布を示す。赤の分布が乳児の模倣音声であり、5つの黒点は相互作用終了時の乳児の5つの母音プロトタイプ的位置を示す。青の分布は養育者の模倣音声である。図5(a)は、両方のバイアスが存在し、相互模倣を通じて、互いに正しい母音位置に収束している様子を示す。図5(b)では、知覚運動バイアスにより収束しているものの、正しい位置ではない。図5(c)と(d)では、自己模倣バイアスのみが作用している場合と二つのバイアスがともに無い状態で、ばらけた音を互いに適当に模倣しあっている。以上から、模倣における二つのバイアスがバランスした状態で両方存在していることが、乳児の音韻の共有化を誘導している可能性が示唆される。

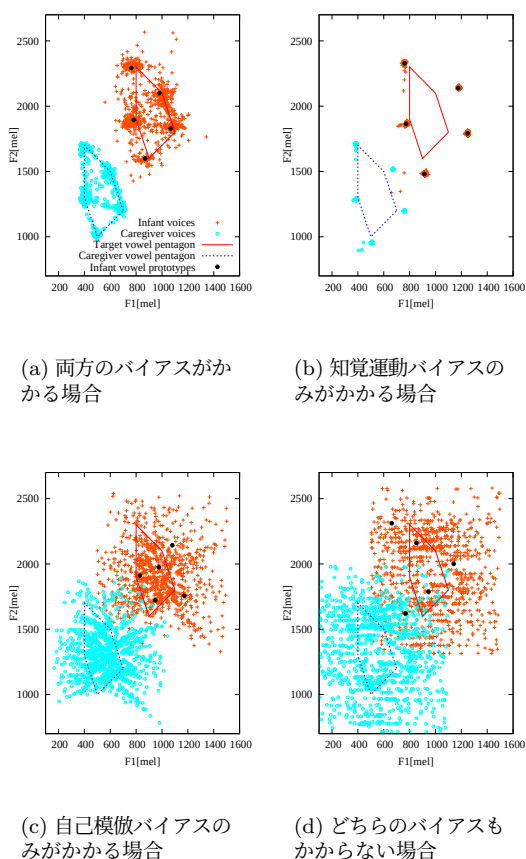


図5 知覚運動バイアスと自己模倣バイアスの違いによる相互模倣発話分布の違い

また最近では、さらに乳児の構音可能領域が徐々に拡大していくことを考慮したシミュレーションを通じて、乳児に対する養育者の発話の分布が大人に対するそれよりも拡大するというマザリーズ[19]が、乳児の生成する音声に対する養育者の期待と関連す

るという仮説が提唱されている[20]。実際の人間による相互模倣を通じた検証、神経科学的基盤の考慮による妥当性に関する検証が、今後欠かせない課題である。

2.2 語彙と音韻の相互拘束的共有化

従来の構成的研究では、音韻に関する共有化過程は単独で扱われることが多く、語彙に関する共有化過程の研究[21, 22]とは別々に研究されることが多かった。しかし、人の乳児が初語を示しはじめ[5]、名詞や動詞など種々の品詞の語彙を理解、利用するようになっていく過程[23]は、生後1年前後の乳児が音声模倣を示し始める時期[4]と同期しているようであり、双方の発達がどのように影響しあうのか、またそれはどのようなメカニズムによって起こるのか、は興味深い問いである。

養育者は、環境中の物体を乳児に見せながらそのラベルを教示したり、乳児の発話に近いラベルのものを乳児に提示したりすることも多いと考えられる。従って、従来の構成的研究のように養育者を乳児の発話を模倣するだけの存在とみなすのではなく、これら全てを含む、より自然な振る舞いを示す存在であると想定する必要がある。Sasamoto et al. [24]は、音声に関する二つの処理経路を示唆する近年の脳科学の知見[25]を参考に、乳児が自身の発話、養育者の発話、物体、それぞれについて表象する層が相互に結合するマッピングを持つと考え、このような養育者との相互作用を通じて音声模倣（自身と養育者の発話の表象のマッピング）と語彙（自身の発話と物体、養育者の発話と物体の表象のマッピング）の発達の相互作用のモデル化に取り組んでいる。より自然な仮定として、養育者が必ずしも乳児の発話に正しく対応する行動をしない場合、例えば、乳児が不明瞭な発話を行うため養育者が正しく模倣できない場合が想定され、従来の構成的研究で想定されていたような単純に対応付ける学習方法では、誤ったマッピングを学習する恐れがある。

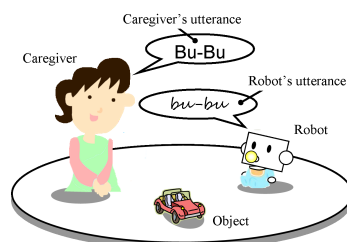


図6 想定するインタラクション

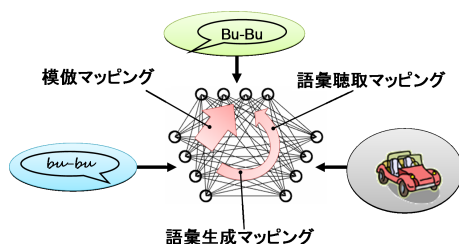


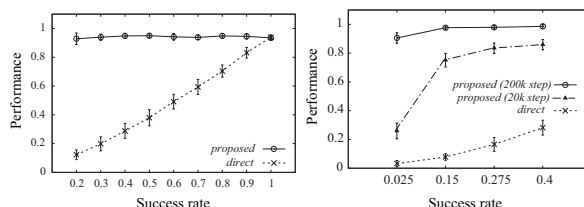
図 7 相互連結型クロスモーダルマップ

これに対し彼らは、任意の二つの層の間のマッピングを学習する際に、学習途上のマッピングによる予測と照らし合わせることで、観測した事象を対応の正解例とみなしてよいかを判断させる方法を提案した。具体的には、観測した事象を、これから学習するマッピングによる予測、さらに他のマッピングを連結したものによる予測と照らし合わせ、他とよく一致するものを対応すべきものとみなさせる。これにより、他のマッピングを学習した経験を利用することで、養育者の行動だけに依存した学習を防ぎ、上述した誤った対応関係を学習する問題を回避できると期待される。

これを検証するため、goo baby[26] を参考に、10 か月から 18 か月までに乳児が獲得するとされる語彙の中から名詞単語を抽出し、選ばれた 39 個の単語のやりとりを想定した計算機シミュレーションを実施した。各ステップで、ロボットはランダムに行動を決定し、養育者は、ロボットの発話の模倣、教示、提示のいずれかの行動を選択する。養育者の行動の恣意性や乳児の発話の未熟さを考慮し、これらの養育者の行動の成功率、すなわち乳児が対応する事象を観測できる確率を変え、200,000 回の相互作用の後の提案手法の学習達成度（39 個の物体に対して対応する表象が正しく想起できる確率）を評価した（図 8(a)）。養育者の応答を対応するデータであるとみなしてそのまま学習する方法（点線）では、養育者が正しく模倣や提示、教示を行う状況でしか学習が出来ないのに対し、提案手法（実線）では、養育者の正しい行動の割合が減少しても、頑強に正しい対応関係を学習できていることがわかる。また興味深いことに、養育者の模倣の成功率をチャンスレベル（この場合約 0.025）まで下げたとしても、教示と提示の成功率がある程度（この場合それぞれ 0.4）であれば、教示と提示により獲得されるマッピングを用いることで、模倣のマッピングを獲得可能であった（図 8(b)）。

彼らの方法のような複数経路の情報統合により、実際に乳児が曝されているような疎にしか対応の情報

が与えられない状況での音韻と語彙の共発達が実現されている可能性が示唆される。この実験では、成功率という表現で構音や知覚の未熟さが考慮されていたものの、実際には、その範疇化はこれらの発達と並行して起こると考えられ、モデルのさらなる拡張が必要である。



(a) 養育者の成功率が一樣に変化する場合

(b) 養育者の模倣の成功率だけが変化する場合

図 8 養育者の行動の対応確率と学習後の達成度

3 終わりに

本稿では、乳児の音韻および語彙の共有化の発達過程に対する構成論的アプローチとして、最初に音韻の共有化を誘導する養育者の模倣メカニズムについて、次に音韻と語彙の共有化の共発達を可能にする乳児の学習メカニズムについての一連の研究を紹介した。このように、乳児だけでなく養育者、あるいは模倣だけでなく語彙に関するマッピングというように、乳児の発達を支える複数の要素の間の相互作用をモデル化を進めていくことで、実際の乳児に対しては倫理的には許されない操作的な統制を施した計算機上の実験を実施し、発達過程の理解に対して洞察を得る発達の動的モデルを得ることが期待される。これらの研究を統合し、音韻と語彙の共発達における養育者の誘導的役割をモデル化すること、さらに人との相互作用を通じて音声模倣および語彙を獲得するロボットを実現することを通じ、モデルの妥当性の検証、修正していくことが今後の課題となる。

謝辞 本稿を作成するにあたり、細田耕准教授（JST/阪大）には日頃から議論を通じて貴重なご意見を頂いた。また三浦勝司君、石原尚君、笹本勇輝君をはじめとする大阪大学大学院の大学院生諸君には、本稿で紹介した研究の推進に献身的に協力して頂いた。彼らの貢献なしに本稿は作成しえず、深く感謝する。

参考文献

- [1] L. Polka and J. Weker: Developmental changes in perception of nonnative vowel contrasts, *Journal of Experimental Psychology: Human Perception and Performance*, 20(2), 421–435, (1994).
- [2] H. Vorperian and R. Kent: Vowel acoustic space development in children: A synthesis of acoustic and anatomic data, *J. Speech Lang. Hear. Res.*, 50, 1510–1545, (2007).
- [3] T. Kokkinaki and G. Kugiumutzakis: Basic aspects of vocal imitation in infant-parent interaction during the first 6 months, *J. of reproductive and infant psychology*, 18, 173–187, (2000).
- [4] S. Jones: Imitation in infancy - the development of mimicry, *Psycho. Sci.*, 18(7), 593–599, (2007).
- [5] E. Bates, P. Dale, and D. Thal: Individual Differences and their Implications for Theories of Language Development, In *Fletcher and Whinney (Eds): The Handbook of Child Language*, 96–151. Blackwell, (1995).
- [6] P. Kuhl, F. Tsao, and H. Liu: Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning, *Proc. Nat. Acad. Sci.*, 100, 9096–9101, (2003).
- [7] M. Asada, K. Hosoda, Y. Kuniyoshi, H. Ishiguro, T. Inui, Y. Yoshikawa, M. Ogino, and C. Yoshida: Cognitive developmental robotics: a survey, *IEEE Trans. on Autonomous Mental Development*, 1(1), 12–34, (2009).
- [8] C. Sasaki, P. Levine, J. Laitman, and S. Crelin: Postnatal developmental descent of the epiglottis in man, *Archives of Otolaryngology*, 103, 169–171, (1977).
- [9] F. Guenther, S. Ghosh, and J. Tourville: Neural modeling and imaging of the cortical interactions underlying syllable production, *Brain and Language*, 96, 280–301, (2006).
- [10] H. Kanda, T. Ogata, K. Komatani, and H. G. Okuno: Vocal imitation using physical vocal tract model, In *Proc. of 2007 IEEE/RSJ Intl. Conf. on Intelligent Robots and Sys.*, 1846–1851, (2007).
- [11] P. Kuhl and A. Meltzoff: Infant vocalizations in response to speech: Vocal imitation and developmental change, *J. Acoust. Soc. of Am.*, 100, 2415–2438, (1996).
- [12] K. Miura, Y. Yoshikawa, and M. Asada: Unconscious anchoring in maternal imitation that helps finding the correspondence of caregiver's vowel categories, *Adv. Robotics*, 21, 1583–1600, (2007).
- [13] N. Masataka and K. Bloom: Acoustic properties that determine adult's preference for 3-month-old infant vocalization, *Infant Behavior and Development*, 17, 461–464, (1994).
- [14] M. Pélaez-Nogueras, J. L. Gewirtz, and M. M. Markham: Infant vocalizations are conditioned both by maternal imitation and motherese speech, *Infant Behavior and Development*, 19, 670, (1996).
- [15] T. Higashimoto and H. Sawada: Speech production by a mechanical model construction of a vocal tract and its control by neural network, In *Proc. of the 2002 IEEE/RAS Intl. Conf. on Robotics & Automation*, 3858–3863, (2002).
- [16] H. Ishihara, Y. Yoshikawa, K. Miura, and M. Asada: Caregiver's sensorimotor magnets lead infant's vowel acquisition through auto mirroring, In *Proc. of the 7th IEEE Intl. Conf. on Development and Learning*, (2008).
- [17] P. K. Kuhl: Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not, *Perception & Psychophysics*, 50, 93–107, (1991).
- [18] P. Oudeyer: The self-organization of speech sounds, *J. Theoret. Bio.*, 233(3), 435–449, (2005).
- [19] P. Kuhl, J. Andruski, I. Chistovich, L. Chistovich, E. Kozhenikova, V. Ryskina, E. Stolyarova, U. Sundberg, and F. Lacerda: Cross-language analysis of phonetic units in language addressed to infant, *Science*, 277, 684–686, (1997).
- [20] H. Ishihara, Y. Yoshikawa, and M. Asada: Caregiver's auto-mirroring and infant's articulatory development enable sharing vowels, In *Intl. Conf. on Epigenetic Robotics*, (2009)(submitted).
- [21] D. Roy and A. Pentland: Learning words from sights and sounds: a computational model, *Cognitive Sci.*, 26, 113–146, (2002).
- [22] C. Yu, L. Smith, Krystal, A. Klein, and R. Shiffrin: Hypothesis testing and associative learning in cross-situational word learning: Are they one and the same? In *Proc. of the 29th Annual Conf. of the Cognitive Sci. Society*, (2007).
- [23] T. Ogura, P. Dale, Y. Yamashita, and T. Murase: The use of nouns and verbs by Japanese children and their caregivers in book-reading and toy-playing contexts, *J. Child Lang.*, 33, 1–29, (2006).
- [24] Y. Sasamoto, Y. Yoshikawa, and M. Asada: Selective integration based on subjective consistency facilitates simultaneous development of vocal imitation and lexicon acquisition, In *Intl. Conf. on Epigenetic Robotics*, (2009)(submitted).
- [25] G. Hickok and D. Poeppel: The cortical organization of speech processing, *Nature Reviews*, 8, 393–402, (2007).
- [26] <http://baby.goo.ne.jp>.

吉川雄一郎（よしかわゆういちろう）

2005 年大阪大学大学院工学研究科知能・機能創成工学専攻博士後期課程修了。在学中、日本学術振興会特別研究員 (DC2)。同年株式会社国際電気通信基礎技術研究所知能ロボティクス研究所研究員。2006 年 4 月 JST ERATO 浅田共創知能システムプロジェクト研究員。コミュニケーションロボット、認知発達ロボティクスの研究に従事。博士 (工学)、日本ロボット学会会員、日本認知科学会会員

浅田稔（あさだみのる）

大阪大学大学院工学研究科知能・機能創成工学専攻教授。JST ERATO 浅田共創知能システムプロジェクト研究員。IEEE フェロー。日本赤ちゃん学会理事。日本ロボット学会などの会員