学術・技術論文

音声模倣と語彙獲得の共発達のための主観的整合機構に基づく対応学習

笹本勇輝[*1]　吉川雄一郎[*2]　浅田稔[*1]

Associative learning based on subjective consistency: modeling co-development of vocal imitation and lexicon acquisition

Yuki Sasamoto[*1], Yuichiro Yoshikawa[*2] *and* Minoru Asada[*1]

This paper presents a learning mechanism that enables co-development of vocal imitation and lexicon acquisition by integrating multimodal information based on subjective consistency. Infant-caregiver interaction is often assumed in modeling infant development. A caregiver is basically assumed to react to an infant in a teaching manner, for example, imitating the learner's voice and labelling an object that it is looking at. However, such tendency is not always expected. Subjective consistency is introduced to judge whether to believe the observed experiences (external input) as a reliable signal for learning. The learning mechanism estimates the outputs of one layer by combining that of other layers and an external input. Based on the proposed mechanism, a simulated infant robot learns associations from its caregiver's phonemes, its own phonemes and objects to co-develop the vocal imitation and lexicon acquisition. The result of computer simulation indicates that the proposed mechanism realized mappings for vocal imitation and lexicon acquisition even when a caregiver does not always react to a learner in a teaching manner.

Key Words: vocal imitation, lexical acquisition, mutual facilitation, subjective consistency

1. はじめに

人間社会への導入が期待されるロボットにとって，人間とのコミュニケーション能力は重要である．特に音声言語は，人にとって最も自然で馴染みのあるコミュニケーション手段の一つであり，それが可能な会話ロボットの開発が進められている．しかし，工学的な観点から設計製作されたロボットの多くは，技術的な限界も含めて，人間と同等レベルにはほど遠い．これは，音声信号解析，音声合成などの個別の技術課題に加えて，言語という極めて困難かつ大きな研究課題を内包しているためと考えられる．これは，人間自身が，いかにして，そのような能力を獲得できたかというミステリーを脳神経科学，認知科学，心理学，言語学など多くの分野も共有していることを意味している．

このような背景に対し，まずは人間の初期，すなわち乳幼児がいかにして，言語を始めとする様々な認知能力を獲得するかの課題に注目し，それをロボットを発達させる課題を扱うことを通じて，構成論的に迫ると同時に，発達する人工物の設計論の確立を目指した認知発達ロボティクスと呼ばれるアプローチが注目されている [1] 認知発達ロボティクスの従来研究では，音声模倣や語彙といった音声言語コミュニケーションの基礎となる機能の発達が個別に研究されている．音声模倣の従来研究では，音声情報と構音運動の対応学習 [2]〜[4] が，語彙に関しては，視覚情報と物体ラベルの対応学習 [5]〜[7] が，モデル化されてきた．しかしながら，実際の乳児は，これら音声模倣と語彙を同時に発達させてい発達メカニズムについて議論する必要があると考えられる．

そこで本研究では，どのようなメカニズムにより音声模倣と語彙獲得の相互促進的な学習が可能であるかを検討することを通じて，これらの共発達過程の構成的理解を目指す．これまでの研究では，乳児の音声模倣のための対応学習の手がかりとして，養育者が乳児の発話を高頻度で模倣するという傾向 [8] [9] が仮定されることが多かった．しかしながら，実場面での養育者と乳児のインタラクションを観察した実験結果から，そのような養育者の模倣の傾向は非常に低いことが報告されている [10] [11]．そのため，従来研究で考えられているような同期して観測される乳児自身の音声と養育者の音声とを結びつける単純な対応学習だけでは，音声模倣のための対応学習は容易ではないと考えられる．これに対し，乳児がある程度語彙を獲得しており，養育者の音声と語彙との対応及び乳児自身の音声と語彙との対応が分かると仮定すると，語彙を介することで，間接的に養育者の音声と乳児自身の音声との対応が想起できると考えられる．そ

原稿受付
[*1]大阪大学大学院工学研究科
[*2]大阪大学大学院基礎工学研究科
[*1]Graduate School of Engineering, Osaka University
[*2]Graduate School of Engineering Science, Osaka University

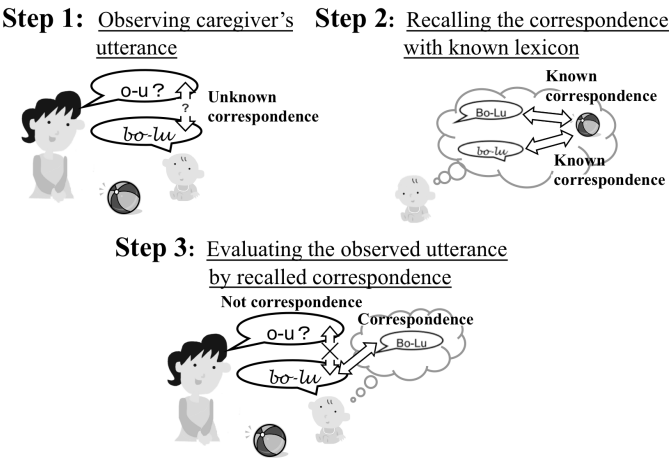


Fig. 1 An example scene when the learning for vocal imitation is facilitated by acquired lexicon

のため，聴取した養育者の音声が発話した乳児自身の音声と対応するものであるかを乳児が直接知らない場合であっても，語彙を介して想起することで，その判断が可能になると考えられる (Fig.1 参照)．これにより，乳児が発話した音声に対して，それに対応しない音声を養育者が発話した場合でも，乳児が単純にそれらの対応を結ぶといった誤りが抑制され，音声模倣のための対応の学習が促進される．このような既知の対応を利用した促進は，語彙の学習においても同様に当てはまる．すなわち，乳児が，聴取した養育者の音声（ラベル）と注目している物体との対応が正しいものであるかを知らない場合でも，養育者の音声と乳児自身の音声の対応及び物体と乳児自身の音声の対応を知っていれば，それが対応するものであるか判断できる．近年の脳科学において，音声処理のための聴覚，語彙概念，構音運動，のそれぞれの表象を繋ぐマッピングが示唆されている [12] ことからも，音声模倣と語彙の学習は相互のマッピングを利用し合いながら発達していくことが察せられるが，それがどのように相互促進的に形成されるかは明らかではない．

そこで本研究では，音声模倣と語彙の共発達のための対応学習に焦点をあて，これを，養育者の音声，物体，ロボット自身の音声の 3 つの表象の相互マッピングの学習過程としてモデル化する．以下では，まず最初に，乳幼児の認知発達に対する知見の概略を示し，その上で，本研究の位置づけを 2 節で明らかにする．3 節では，想定する母子相互作用と学習の具体的なメカニズムを示し，4 節では，シミュレーション結果に基づき，手法の妥当性を評価する．そして，最後に実験結果を考察し，本研究の限界と今後の展望に関して議論する．

2. 乳幼児の認知発達と主観的整合性による共発達

人の乳児は，8ヶ月頃から聴取した語彙への理解を示し，12ヶ月頃から自身も語彙を発し始める [13]．また，ほぼ時を同じくして，8ヶ月頃には，言語の基本単位と呼べる母音の模倣を示すようになり，さらに 14ヶ月頃には，それらが連なった複数母音についても模倣できるようになる [14]．このような語彙と音声模倣の能力は，これらが出現する時期が重複していることや，どちらも基本的には音韻の聴取と発声を必要とすることから，互いに影響を及ぼし合いながら発達していくものと察せられる．また，模倣の経験がその後の語彙の発達を促進すること [15] や語彙の知識が模倣に必要な音韻対比の形成を可能とすること [16] が示唆されていることからも，これらの能力は相互促進的に発達していると考えられるが，どのようなメカニズムで，それらの相互促進的発達が可能となるのかについての理解は十分ではない．

このような乳児の発達に関する問いに対して，これまで発達心理学の分野において盛んに研究されてきた．しかしながら，倫理的な問題あるいは言葉の通じない乳児を扱っているため，実験の統制が容易ではなく，発達様相の記述には及んでもその過程の裏にあるメカニズムを理解するには限界があった．そこで，認知発達ロボティクスのアプローチで，この問題に取り組むが，その基本的な考え方は以下である：音声模倣の発達を計算論的に扱うためには，連続的に聴取した音声情報を音韻列として認識できるかのカテゴリ化の問題，その音韻と構音運動との対応付けのマッピングの問題に取り組む必要がある．同様に，語彙の発達では，視覚情報を一つの対象（物体）として認識できるかのカテゴリ化の問題，その対象と音声ラベルとの対応付けのマッピングの問題に取り組む必要がある．カテゴリ化およびマッピングの課題では，ともに，教師信号の選択が大きな問題となる．これは本稿の実験で確認していくように，信号の主観的整合度と呼ぶ指標を導入することで，取扱いが可能である．本稿では，主観的整合度の有効性の検証に焦点を当てるため，カテゴリ化の課題は取り扱わず，以後では，マッピング課題に集中して議論する．

ここで導入する信号の主観的整合度とは，ある事物をある方法で捉えた信号が，他の複数の方法で捉えた信号とどれほど一致しているかを表すものである．本研究では，これを観測信号や，獲得されたマッピングから想起される信号に対して適用し，各信号がどの程度正しい対応関係を表しているかを示すものとみなして、対応関係の学習に用いることを考える．つまり観測された信号をそのまま対応関係の教師信号とするのではなく、それまでに学習された対応関係やそれらの推移的な関係から、観測されるべき信号を想起し、観測信号と複数の想起信号の中に一貫して現れる信号を教師信号とみなす．これにより，対応しない信号が観測される場合，あるいはマッピングの学習が途上である場合，観測信号と想起信号はお互いに異なる信号となり，いずれの主観的整合度も低くなるため，それらの平均的な信号が教師信号となる．この平均的な信号に従い，徐々にマッピングの学習が進み，入力に対して一貫した信号がマッピングから想起されるようになると，対応しない信号が観測された場合に，観測信号と想起信号との主観的整合度が低くなることで，教師信号と見なされにくくなっていくと期待される．従って，複数のマッピングの学習は相互促進的に進んでいくことになると期待される．

3. 音声模倣と語彙獲得の共発達メカニズム

3.1 問題設定

本研究では，発達心理学の知見を参考にし，以下を想定する．

- **言語的音声模倣：発声時の構音運動とその後聞こえてきた**

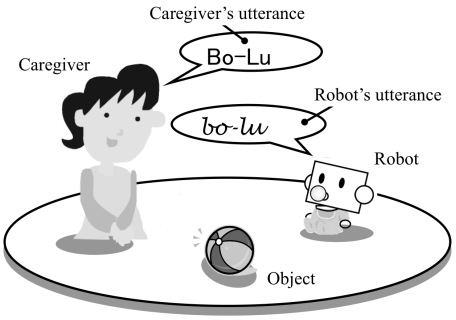


Fig. 2 Assumed environment of caregiver-robot inreraction

言葉との対応学習

乳児は，構音運動の未熟さや養育者とは異なる構音器官を持つことから，養育者と音響的に同一の音声を発声することは困難である．これに対して，乳児は，養育者に乳児自身の発話を模倣される経験を通じて，音声模倣の学習をしていると考えられている [17]．すなわち，既に乳児の発声に対応する言葉を知っている養育者が，乳児が発した音声をある一つの言葉として言語的に模倣して発声することで，乳児は，その時の構音運動と聞こえてきた発声とを対応付けることができ，音声模倣の能力を獲得していると考えられる．

- **語彙獲得：物体とそのラベルとの対応学習**

  乳児の語彙の学習にとって，養育者から物体を提示され，また同時にその名前を聞かされる経験は重要であるとされている．[18]，養育者が，環境中にある多くの物体の中から，乳児が注目している物体を特定し，また同時にそのラベルを発話することで，乳児はそれらを対応付け，語彙を獲得していると考えられる．

- **常に理想的な教師ではない養育者**

  乳児に応じる養育者の行動を分析した研究 [10] [11] が示しているように，養育者が乳児に対して，乳児の発した音声を模倣したり，物体とそのラベルを同時に提示したりする教師的な行動で応じることは，現実世界では頻度が低い．

本研究では，上記の状況を想定して，Fig.2 に示すロボットと養育者と物体が存在する環境を考える．ロボット，養育者の順に交互に行動し，養育者の行動が終了した時点までを 1 ステップとする．各ステップで，ロボットは養育者と物体のどちらかを見る．その時，同時に発声するかしないかを選択する．養育者は，ロボットの行動に対して，音声提示，物体提示，物体呼称の 3 つの何れかの行動で応答する．ただし，養育者の行動が常にはロボットの行動に対応するものであるとは限らないことを考慮し，それぞれの行動の結果は確率的に定まるとする．各行動の詳細は以下の通りである．

音声提示: ロボットの発声に応じて発声する．ただし，養育者がロボットの発声を模倣する言葉 (対応する言葉) を発する場合と，ロボットの発声と関係なく言葉 (この場合は，ロボットの発声とは対応しない他の言葉) を発する場合もあることを考慮して，養育者の発した言葉がロボットの発声の模倣である確率を p_V とする．

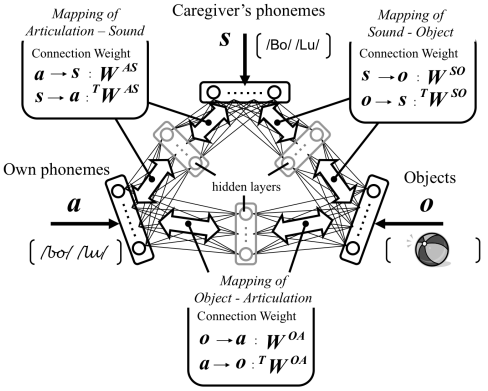


Fig. 3 Mutually associated multimodal mapping model

物体提示: ロボットの発声に応じて物体を見せる．ただし，養育者がロボットの発声に対応する物体を提示するだけでなく，ロボットの発声に関係なく物体を提示する場合もあることを考慮して，養育者の提示した物体がロボットの発声した物体に対応している確率を p_S とする．

物体呼称: ロボットに物体の名前を教える，すなわち，物体を見せて，その名前を発声する．あるいは，ロボットが見ている物体の名前を発声する．ただし，養育者が物体の名前を教えるため以外に，ロボットに対して言葉を発する場合もあることを考慮して，養育者の発した言葉が物体の名前を教えるためのものである確率を p_D とする．

ロボットは，このような養育者とのインタラクションを通じて，ロボット自身の音声と養育者の音声の対応（音声模倣の能力），養育者の音声と物体及び物体とロボット自身の音声の対応 (語彙理解及び生成の能力) を学習する．ただし，養育者は，行動ごとに定められた確率 (p_V, p_S, p_D) に応じてロボットに対応を示し，ロボットは，その確率を事前に知ることはなく，養育者の行動を観測する．これらの確率は，養育者がロボットに対応を与える，つまり，ロボットに対して教示的に振る舞う確率を表し，以後総称して，教示率と呼ぶ．

本研究では，教示率を変えて母子相互作用シミュレーションを行うことで，養育者が必ずしもロボットにとって理想的な教師であるとは限らない状況を想定する．そして，そのような状況下でも，提案する主観的整合機構に基づいた対応学習により，ロボットが音声模倣と語彙獲得のための対応を学習可能であることを示す．つまり，養育者がロボットに対してより教示的である状況では，それをより信頼して対応を学習し，養育者がロボットに対して全く教示的でない状況では，他のマッピングから想起されるものを信頼して学習する相互促進的学習を実現する．

3.2 観測信号ベクトル

前節の母子相互作用によって，ロボットは以下の 3 種の信号を観測する．

(1) ロボット自身の発声によって形成される，ロボットの音韻系列ベクトル ($\boldsymbol{a} \in \Re^{M_i}$)

(2) 養育者の発声を観測することで形成される，養育者の音韻系列ベクトル ($\boldsymbol{s} \in \Re^{M_c}$)

(3) 物体を観測することで形成される，物体ベクトル ($\boldsymbol{o} \in \Re^{N}$)

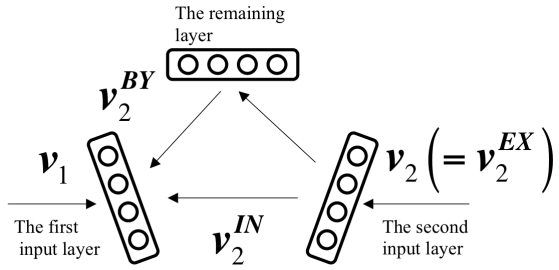


Fig. 4 Notations for learning rule

例えば，ロボットが音韻系列ベクトル $\boldsymbol{a_c}$ による発声を行い，養育者が音声提示を行う場合，養育者は確率 p_V で音韻系列ベクトル $\boldsymbol{s_c}$ による発声を，確率 $1-p_V$ で $\boldsymbol{s_{\bar{c}}}$ による発声を行い，これらがネットワークに入力される．ここで，$\boldsymbol{a_i}$，$\boldsymbol{s_i}$ は i 番目の物体のラベルに対する音韻系列ベクトルであり，$\bar{c}$ は，c 番目以外の物体のラベルを表す．

3.3 マッピングの学習

本研究では，マッピングの学習則として，相互想起型ボルツマンマシンの学習則 [19] を適用する．相互想起型ボルツマンマシンは，確率的ニューラルネットワークの一つであり，観測される二つの変数間の確率的関係を学習することができる．そのため，本研究で想定する入力に対して出力が一意に定まらない非決定的な状況での対応学習手法として有効である．

ボルツマンマシンでは，ある入力 $\boldsymbol{x}$ に対する出力 $\boldsymbol{y}$，及びその逆方向の想起は，共通の結合強度行列 $\boldsymbol{W}$ を用いて，

$$\Pr\left(y_m=1|\boldsymbol{W},\boldsymbol{x}\right)=\frac{1}{1+\exp\left(-\sum_n w_{nm}x_n\right)},\quad(1)$$

$$\Pr\left(x_n=1|^{\boldsymbol{T}}\boldsymbol{W},\boldsymbol{y}\right)=\frac{1}{1+\exp\left(-\sum_m w_{nm}y_m\right)},\quad(2)$$

の確率に従って実行される．ただし，y_m は $\boldsymbol{y}$ の m 番目の要素であり，x_n は $\boldsymbol{x}$ の n 番目の要素である．w_{nm} は入力層の n 番目のノードと出力層の m 番目のノードの結合の強さを表し，値が大きいほど，それらのノードが対応していることを表す．

本研究では，前節で説明した 3 種の観測から，マッピングを学習することを想定し，Fig.3 に示すように異なる 3 つの表象が相互に結合したニューラルネットワークモデルを考える．このモデルでは，3 種のベクトルの各要素は 3 つの異なる層のそれぞれ，M_i, M_c, N 個のノードに対応付けられ，ロボットはそれらノード同士の結合強度を要素とする以下の 3 種の行列を学習する．

(1) ロボット自身の音韻と養育者の音韻との対応を表す，構音-聴覚マッピング $\boldsymbol{W}^{AS}$
(2) 養育者の音韻系列と物体との対応を表す，聴覚-単語（語彙）マッピング $\boldsymbol{W}^{SO}$
(3) 物体とロボット自身の音韻系列との対応を表す，単語（語彙）-構音マッピング $\boldsymbol{W}^{OA}$

ここで，Fig.4 に示すように， 3 つの層のうち何れかに入力があり，続いて別の層に入力があった場合について考える．初めに入力があった層 (以下，第 1 入力層) への入力が $\boldsymbol{v_1}$ であり，次に入力あった層 (以下，第 2 入力層) への入力が $\boldsymbol{v_2}$ であったとき，それら二つの入力の対応関係は以下の手順に従って学習される (詳細な式の導出は付録 A を参照されたい)．

(1) 第 1 入力層及び第 2 入力層をそれぞれ，$\boldsymbol{v_1}$，$\boldsymbol{v_2}$ で固定した状態で，マッピング (式 (2)) により，隠れ層の状態を更新する．全てのノードの状態更新を行った時点を 1 サイクルとし，十分なサイクル数で状態を更新する．そして，各結合に関して，それらを繋ぐノードが同時に 1 になる頻度を計算する．ノード n と ノード m が 結合強度 w_{nm} で結合している場合，それらが同時に 1 になる頻度を K_{nm} として計算する．上記の操作を 2 度繰り返し行い，K_{nm} を計算する．
(2) 上記の計算を，今度は，第 1 入力層のみ $\boldsymbol{v_1}$ で固定した状態で行う．同様に各結合に関して，それらを繋ぐノードが同時に 1 になる頻度を計算する．ノード n と ノード m が 結合強度 w_{nm} で結合している場合，それらが同時に 1 になる頻度を ${}^1K'_{nm}$ として計算する．また，両方向の対応を学習するため，第 2 入力層のみ $\boldsymbol{v_2}$ で固定した状態でも同様に計算し，ノード n と ノード m が同時に 1 になる頻度を ${}^2K'_{nm}$ とする．そして，入力層を固定した状態で，ノード n と ノード m が同時に 1 になる頻度 K'_{nm} を，第 1 及び第 2 入力層を固定した状態で計算した，${}^1K'_{nm}$，${}^2K'_{nm}$ の和として，以下のように計算する．

$$K'_{nm}={}^1K'_{nm}+{}^2K'_{nm}.\quad(3)$$

(3) ノード n と ノード m の結合強度 w_{nm} を上記計算で求めた K_{nm}， K'_{nm} により，以下のように更新する．

$$w_{nm}=w_{nm}+\alpha\left(K_{nm}-K'_{nm}\right).\quad(4)$$

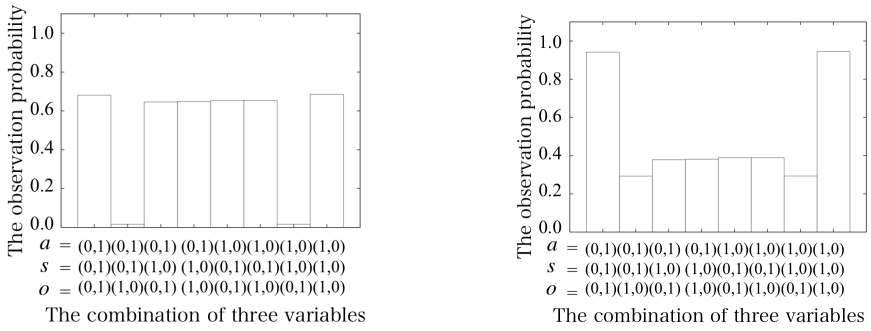
ここで，α は学習係数である．

上記の手順を繰り返し行い，結合強度を更新していくことで，二つの入力間の対応がとれるようにマッピングを学習する．例えば，ロボットが発声したのち，養育者が発声した場合，$\boldsymbol{v_1}$ はロボット自身の音声 $\boldsymbol{a}$，$\boldsymbol{v_2}$ は観測される養育者の音声 $\boldsymbol{s}$ となり，ロボットが養育者の発声を模倣できるようになるためには，聞こえてきた養育者の音声 $\boldsymbol{s}$ からそれに対応するロボット自身の音声 $\boldsymbol{a}$ をマッピングを介して正しく推定できるように，結合強度行列 $\boldsymbol{W}^{\boldsymbol{AS}}$ を更新していくことが必要となる．

3.4 主観的整合機構：複数信号の整合度に基づく統合

前節の学習則により，例えば，ロボットの発声を養育者が高頻度で模倣する場合，ロボットはロボット自身の音声とそれに対応する養育者の音声の関係を学習することができる．しかし，養育者が高頻度で模倣ではない発声をロボットに行う場合は，そのような単純な対応学習のみでは，ロボットは誤った対応関係を学習する問題がある．

これに対して，Fig.3 に示す学習モデルでは，複数のマッピングが相互に結合しており，学習時に養育者から与えられる信号以外にマッピングを介して想起される信号も同時に利用することができる．そのため，マッピングが成熟していれば，そこから想起される信号を学習に利用できる，すなわち，養育者からの信号ではなく，想起される信号を対応する信号とみなして



(a) Without integrated signal (b) With integrated signal

Fig. 5 Observation probability with respect to combinations of three input variables (a) without integrated signal generated by proposed mechanism and (b) with it

学習することができる．そのため，養育者が高頻度で対応しない信号をロボットに与える場合であっても，ロボットが誤った対応を学習する問題を抑制できると考えられる．しかしながら，それをロボット自身が判断するとなれば，ロボットが主観的に観測・計算しうる形で，その判断の仕組みが構成される必要がある．本節では，そのような複数の信号を基に主観的に対応する信号を生成するための整合度に基づいた統合手法を提案する．

前節と同様に，3 つの層のうち何れかに入力があり，続いて別の層に入力があった場合について考える．この時，第 2 入力層への入力 $\boldsymbol{v_2^{EX}}(=\boldsymbol{v_2})$ (以下，外部信号) ，第 1 入力層 から第 2 入力層への直接のマッピングを介して想起される信号 $\boldsymbol{v_2^{IN}}$ (以下，直接予測信号) 及び 第 1 入力層から 残っているもう一つの層を介した第 2 入力層への間接のマッピングを介して想起される信号 $\boldsymbol{v_2^{BY}}$ (以下，間接予測信号) を統合した統合信号 $\boldsymbol{v_2'}$ を，次式のように計算する．

$$\boldsymbol{v_2'} = f(\boldsymbol{v_2^{EX}}, \boldsymbol{v_2^{IN}}, \boldsymbol{v_2^{BY}}) = \lambda_{EX}\boldsymbol{v_2^{EX}} + \lambda_{IN}\boldsymbol{v_2^{IN}} + \lambda_{BY}\boldsymbol{v_2^{BY}}. \tag{5}$$

ここで，$\lambda_n\,(n \in \{EX, IN, BY\})$ は 外部信号，直接予測信号，間接予測信号，それぞれの主観的整合度を表し，

$$\lambda_n = \frac{\exp\left(-e_n/\sigma^2\right)}{\sum_{m\in\{EX,IN,BY\}} \exp\left(-e_m/\sigma^2\right)}, \tag{6}$$

と計算される．ここで，σ は e_n に対する感度パラメータである．また，e_n は信号 $\boldsymbol{v_2^n}$ が他の二つの信号と比べて，どれ程かけ離れたものであるかを表し，信号間の距離の和として，

$$e_n = \sum_{l\notin n} \left\| \boldsymbol{v_2^n} - \boldsymbol{v_2^l} \right\|, \tag{7}$$

と計算される．式 (7) より，信号間の距離の和を計算し，それに基づき，式 (5) を用いて統合することで，信号間の近さに応じて統合信号を求める．これにより，ある一つの信号が他の二つの信号と近ければ，統合時にその信号が反映される．逆に，ある一つの信号が他の二つの信号と遠ければ，統合時にその信号が反映されないように統合信号が計算される．提案手法では，外部信号，直接予測信号，間接予測信号，それぞれについて，主観的整合度を計算し，式 (5) で重み付け統合した信号を，入力とみなしてマッピングを学習する．すなわち，前節の学習則の手順 (1)，(2) において，第 2 入力層を外部からの入力 $\boldsymbol{v_2^{EX}}$ ではなく，統合信号 $\boldsymbol{v_2'}$ で固定して結合強度を更新する．

以上のように，外部信号だけでなくマッピングを介して想起される信号も含めて統合し，それを対応する信号とみなして学習することで，養育者の行動だけに依存して学習することを防ぐことができる．これにより，養育者が正しい対応をあまり示さない場合に誤った対応を学習する問題を抑制できると考えられる．また，それらを相互の近さによって重み付け統合することで，より一貫している信号がより正しい対応を示す信号であるとロボット自身が主観的に判断できる．

4. シミュレーション実験

提案手法の有効性を確かめるために，計算機シミュレーションを用いて以下の 3 つの実験を実施した．

予備実験: 簡単な状況において，提案手法により，誤った対応が与えられる場合に，それを抑制できるかを確認する．

実験 1: 様々な教示率で養育者がロボットに応答する状況を想定し，提案手法の頑健性及び限界を検証する．

実験 2: 乳児に応じる養育者の行動を分析した実験 [10] から，実場面では，養育者が乳児に対して模倣やラベル付けを行う割合は非常に低いものであることが示されている．そこで，そのように養育者が乳児にほとんど対応を与えない状況においても，提案手法により，相互促進的な対応の学習が可能であるか検証する．

4.1 基本設定

乳児は，まず聴取した音声を母国語の音韻としてある程度認識できるようになり [20]，その後，模倣できる [14] ように発達していると考えられる．すなわち，乳児は，聴取した音声から既にある程度排他的に音韻を認識し，それと乳児自身が発話した音韻との対応がとれるように音声模倣能力を発達させていると考えられる．そこで，ここでは簡単のため，ロボットと養育者の発声はお互いが共通に持ついくつかの音韻（モーラ）で構成されるものとした．具体的には，$\boldsymbol{a}$，$\boldsymbol{s}$ は，それぞれが M 種類あるうちのどのモーラで構成されているかを表わすベクトルであるとした．例えば，ロボットの発声が $/a_2\ a_8/$ のように，2 番目のモーラと 8 番目のモーラの組合せで構成される音であっ

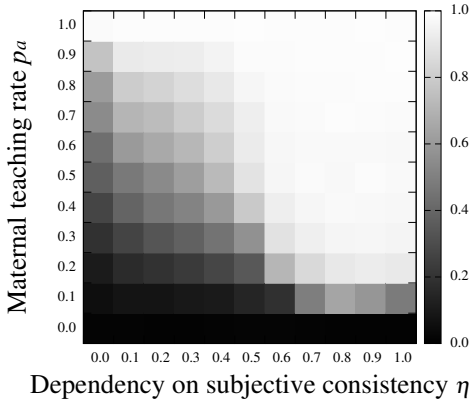


Fig. 6 Average probability of predicting corresponding vector by acquired mappings until 100,000 step with respect to dependency on subjective consistency(η) and maternal teaching rate (p_a)

た場合，$\boldsymbol{a}$ は，2 番目と 8 番目の要素が 1，それ以外の要素が 0 である M 次元ベクトルとなる．また，乳児は語彙の獲得以前から，ある程度物体の認識が可能である [21] ことが報告されている．そこで，物体は，N 種類のどの物体について注目しているかを表わすベクトル $\boldsymbol{o}$ で表記できるとした．例えば，ロボットが i 番目の物体に注目している場合，$\boldsymbol{o}$ は i 番目の要素が 1，それ以外が 0 である N 次元ベクトルとなる．これらのベクトルは，それぞれの表象に入力される特徴ベクトルに相当する．例えば，養育者の音韻系列ベクトル $\boldsymbol{s}$ は，聴取した人の音声データを音声認識器を用いて対応するモーラ毎に分解し，ベクトル化 (あるモーラが含まれていれば 1，含まれていなければ 0) したものに相当する．

予備実験では，主観的整合機構により，どのように誤った対応学習が抑制されるについて注目するため，簡単な状況を想定する．具体的には，環境中の物体及びラベルの数は，$N = 2$ 個とし，すべてのラベルを構成するためのモーラ数も $M = 2$ 個とした．また，すべての変数が同じ値 ($\boldsymbol{a} = \boldsymbol{s} = \boldsymbol{o} = [0, 1]^T$ 及び $\boldsymbol{a} = \boldsymbol{s} = \boldsymbol{o} = [1, 0]^T$) となる場合を正しい対応とした．

実験 1，2 では，より複雑な状況として，実際の乳児が置かれている状況を想定して，2009 年 2 月 22 日時点で goo ベビー [22] に記載されていた乳児が 10ヶ月から 18ヶ月までに獲得する語彙を実際の母子相互作用場面において頻出する語彙とし，その中の名詞単語をシミュレーションで使用するデータとした．抽出したデータから，環境中の物体及びラベルの数は，$N = 39$ 個であり，すべてのラベルを構成するためのモーラ数は $M = 37$ 個であった．また，実験 1，2 では，提案手法の有効性を確かめるために，観測された信号を対応する信号とみなす，通常の相互想起型のボルツマンマシンの学習手法 (3.3 節参照) と，主観的整合度に基づき統合した信号を対応する信号とみなす，提案手法 (3.4 節参照) を比較する．具体的には，後者に従う程度を表すパラメータ η を導入し，対応の学習のための信号 $\boldsymbol{v}'$ を，

$$\boldsymbol{v}' = (1-\eta)\boldsymbol{v}^{EX} + \eta(\lambda_{EX}\boldsymbol{v}^{EX} + \lambda_{IN}\boldsymbol{v}^{IN} + \lambda_{BY}\boldsymbol{v}^{BY}), \quad (8)$$

と決定する．すなわち，η の値が大きいほど，提案手法に頼る割合が増え，低いほど，通常の学習手法に頼る割合が増えることを意味する．

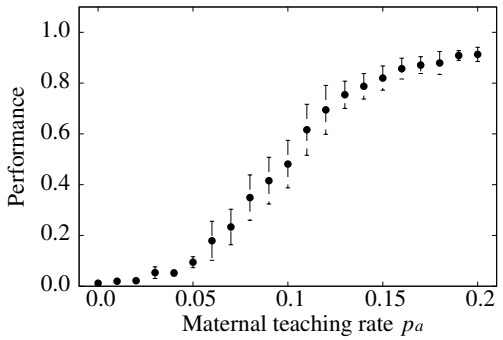


Fig. 7 Average probability of predicting corresponding vector by acquired mappings until 100,000 step with respect to low maternal teaching rate (p_a) under the proposed method ($\eta = 1.0$)

4.2 予備実験: 主観的整合機構による誤った対応学習の抑制効果の確認

提案手法によって，誤った対応が与えられる場合に，それを抑制できるかを確認するために，あえて誤った対応が与えられる状況を想定して実験した．具体的には，教示率を $p_V = 0.05$, $p_S = 1.0$, $p_D = 1.0$ とし，100,000 ステップの母子相互作用シミュレーションを実施した．Fig.5 に，100,000 ステップの学習中に観測された 3 つの変数の組み合わせとそれを教師信号として学習に用いた頻度を示す．Fig.5(a) は，入力された信号をそのまま教師信号として用いた場合であり，(b) は，提案手法により統合した信号を教師信号として用いた場合である．本実験では，$\boldsymbol{a} = \boldsymbol{s} = [0, 1]^T$ 及び $\boldsymbol{a} = \boldsymbol{s} = [1, 0]^T$ を正しい対応としているので，それらの観測頻度が，3 つの変数の他の組み合わせに比べて相対的に高い場合，観測の共起関係のみから，それら正しい対応を学習できる．しかしながら，$p_V = 0.05$ としたため，$\boldsymbol{a}$ と $\boldsymbol{s}$ の間の正しい対応 ($\boldsymbol{a} = \boldsymbol{s} = [0, 1]^T$ 及び $\boldsymbol{a} = \boldsymbol{s} = [1, 0]^T$) が入力される割合がチャンスレベルよりも非常に少く，結果として，Fig.5(a) に示すように，3 つの変数の正しい対応 ($\boldsymbol{a} = \boldsymbol{s} = \boldsymbol{o} = [0, 1]^T$ 及び $\boldsymbol{a} = \boldsymbol{s} = \boldsymbol{o} = [1, 0]^T$) を教師信号として用いる頻度が他の組み合わせの観測頻度とほぼ同程度となっている．そのため，単純に入力の共起関係をみるだけでは，正しい対応の学習が困難である．これに対して，提案手法によって統合した信号を教師信号とすることで，Fig.5(b) に示すように，正しい対応を教師信号とする頻度が，他の組み合わせよりも相対的に高くなっていることがわかる．この結果から，提案手法により，誤った対応が与えられる場合は，それを教師信号として用いないようにするという抑制が可能となることが確認された．

4.3 実験 1: 主観的整合機構による促進効果の検証

4.3.1 教示率毎の学習パフォーマンス

対応学習の目的においては間違った対応となる信号が観測される状況であっても，提案手法により頑健に対応の学習が可能であるかを確かめるために，養育者の教示率 p_V, p_S, p_D を同じ値 p_a に固定し，p_a を 0.0 から 1.0 まで 0.1 ずつ変化させ

たそれぞれの場合について，η を 0.0 から 1.0 まで 0.1 ずつ変化させ，100,000 ステップの母子相互作用シミュレーションを 10 回実施した．Fig.6 は，p_a と η の違いによる，対応学習のパフォーマンスの違いを示している．但し，明暗は 100,000 ステップ経過時点での学習パフォーマンスを表しており，明るい程高く，暗い程低いことを表す．パフォーマンスは，入力ベクトルを一通り入力することで測定する．具体的には，39 通りの可能な入力ベクトルをそれぞれ各表象に入力し，マッピングを介して想起されるベクトルが 39 通りの可能な出力ベクトルの中で正しく対応するものに最も近くなった場合の割合の，各マッピングに関する平均値として評価した．また，提案手法における各パラメータは経験的に 学習係数 $\alpha = 0.2$，主観的整合度の感度 $\sigma = 1.0$ とした．Fig.6 より，教示率が高い場合 ($p_a = 1.0$)，すなわち，外部信号として直接対応を示す信号が観測される場合では，通常の学習手法のみでも高いパフォーマンスが得られていることがわかる．一方，教示率が比較的低い場合 ($0.2 <= p_a < 1.0$)，すなわち，外部信号として対応を示さない信号も観測される場合では，主観的整合度に基づく統合に従う程度 (η) が高いほど，正しい対応学習が可能になっていることがわかり，提案手法の有効性が確認できる．しかしながら，教示率がかなり低い場合 ($p_a < 0.2$) では，たとえ主観的整合度に基づく統合に従う程度が高かったとしても，最終パフォーマンスは低く，$\eta = 0.8$ 付近でのパフォーマンスが最も高い．これは，教示率がかなり低い場合では，誤った対応を経験する割合が多くなり，複数のマッピングが誤った対応を学習する割合も増えるため，主観的整合度が高いことで，他のマッピングもその誤った対応で固定される結果であると考えられる．

4.3.2 極端に低い教示率に対する提案手法の効果

教示率が低い場合での提案手法の効果を詳細に分析するために，$\eta = 1.0$ と固定し，p_a を 0.0 から 0.2 まで 0.01 ずつ変化させた場合の学習パフォーマンスを Fig.7 に示す．これより，p_a が低くなるにつれ，パフォーマンスが急激に下がっていることがわかる．$0.05 < p_a < 0.15$ の範囲では，分散が大きく，場合によっては，ある程度高いパフォーマンスが得られているものの，$p_a <= 0.05$ の範囲では，パフォーマンはゼロ付近に留まっている．これは，主観的整合度に基づく統合では，一貫した信号であれば，それがたとえ誤った対応を示すものであっても，信頼してマッピングの学習が進められるためであると考えられる．つまり，提案手法により，学習者は，パフォーマンスの良し悪しとは関係なく，あるマッピングが一貫した信号を想起していれば，その信号を正しい対応を示すものであると主観的に判断し学習する．しかしながら，その一貫した信号が誤った対応を示すものであった場合，その誤った対応を示す信号によって他のマッピングの学習が拘束される．そして，最終的には，対応を何も学習しないのではなく，すべてのマッピングで一貫した対応ではあるものの，誤った対応を学習してしまい，結果として，最終的なパフォーマンスが低くなっている．そのため，養育者が対応を与える割合がかなり低い場合 ($p_a < 0.2$) では，提案手法のような主観的整合度に基づく統合により，すべてのマッピングで一貫した対応を学習するものの，それが正しい対応となる保証はなく，注意する必要がある．

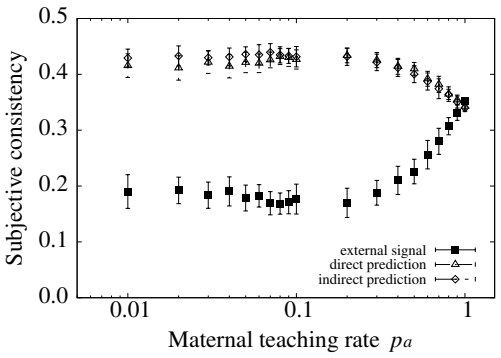


Fig. 8 Average of average subjective consistencies among different three mappings during final 100 steps of learning with respect to maternal teaching rate p_a: subjective consistency for external signal (filled squares with solid line), direct prediction (blank triagle with broken line), and indirect prediction (blank diamond with dash-dotted line).

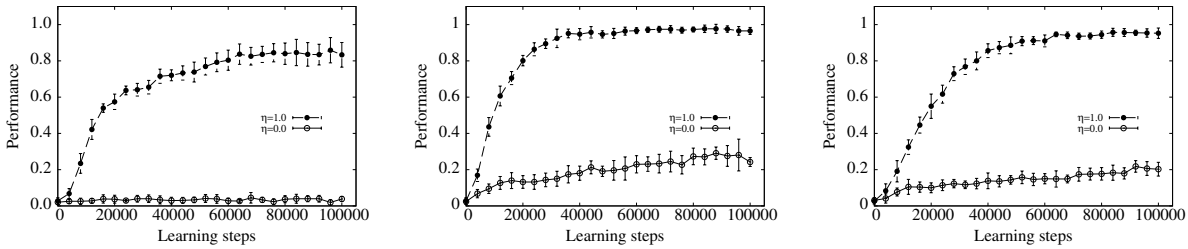
4.3.3 教示率に対する学習の適応性

提案手法により，教示率に対してどのように適応的に学習が可能となったかを見るために，提案手法にのみ頼って学習した場合 ($\eta = 1.0$) の教示率と学習終了時 (99,900 〜 100,000 ステップ) における各信号に対する主観的整合度の関係を Fig.8 に示す．Fig.8 から分かるように，$0.2 <= p_a < 1.0$ の範囲では，教示率が低くなるにつれ，外部信号に対する主観的整合度のみが下がっていることが見てとれる．すなわち提案手法により，外部信号が常に対応する信号ではない場合には，それが対応する信号としてあまり反映されないようにすることで，教示率によらない頑健な対応学習が可能になっていると考えられる．しかしながら，教示率がかなり低い場合 ($p_a < 0.2$) では，外部信号がほとんど対応を示すものではないにも関わらず，その主観的整合度が $p_a = 0.2$ の場合と同程度であることがわかる．これが 4.3.2 節で説明した，対応を何も学習しないのではなく，誤った対応を学習する原因であると考えられる．

4.4 実験 2: 実環境での養育者の応答を模した状況での検証

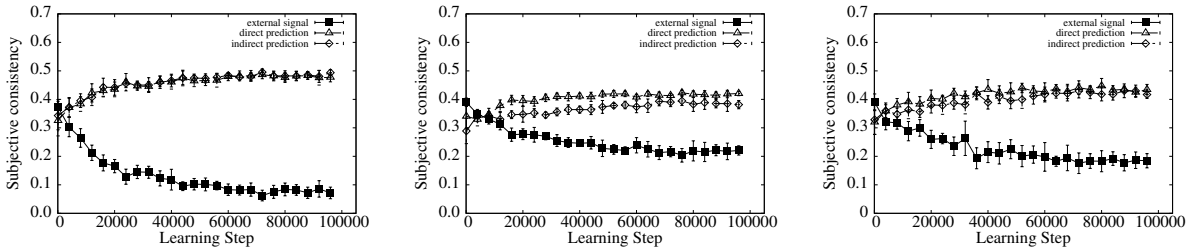
前節の実験では，3 種の教示率について同じであると仮定して，学習パフォーマンスを検証した．しかし，実際の乳児に応じる養育者の行動を分析した実験 [10] によれば，養育者の模倣の頻度は非常に低く約 5 %程度，ラベル付けについては約 35 %程度と，行動によってその頻度が異なる．そこで，この観察実験結果に倣い，p_V を 0.05，p_S と p_D を 0.35 とした場合について，100,000 ステップの母子相互作用シミュレーションを 10 回実施した．η は 1.0 と 0.0 の 2 通りに設定し，パフォーマンスは前節と同様の方法により評価した．

各マッピングの学習パフォーマンスの遷移を Fig.9 (a) (b) (c) に示す．通常の学習手法に従って対応を学習した場合 ($\eta = 0.0$)，語彙の学習 ((b), (c)) に関しては，教示率程度の想起が可能になるまでにとどまっており，模倣の学習 ((a)) に関しては，ほとんど対応を学習出来ていない．一方，提案手法に従って対応を学習した場合 ($\eta = 1.0$)，養育者の模倣をほとんど経験できない状況であっても，構音-聴覚マッピングの最終パフォーマン



(a) Articulation-Sound Mapping　(b) Sound-Word Mapping　(c) Word-Articulation Mapping

Fig. 9 Average transitions of learning performance of (a) articulation-sound mapping, (b) sound-word one, and (c) word-articulation one with the proposed subjective consistency ($\eta = 1.0$) and without it ($\eta = 0.0$)



(a) Articulation-Sound Mapping　(b) Sound-Word Mapping　(c) Word-Articulation Mapping

Fig. 10 Average transition of subjective consistency calculating in learning (a) articulation-sound mapping, (b) sound-word one , (c) word-articulation one: subjective consistency for external signal (filled squares with solid line), direct prediction (blank triagle with broken line), and indirect prediction (blank diamond with dash-dotted line)

スが維持されていることがわかる．

$\eta = 1.0$ の場合の，各信号に対する主観的整合度の 100 ステップ毎の移動平均値の遷移 (Fig.10 (a) (b) (c)) をみてみると，外部信号に対する主観的整合度が学習が進むにつれて減少していくこと，すなわち養育者から与えられる外部信号をあまり信頼しないように学習が進んでいることがわかり，これにより，提案手法では高いパフォーマンスが維持されている．また，模倣の学習 (Fig.10(a)) は，語彙の学習 (Fig.10(b)(c)) に比べ，間接予測信号をより信頼して，つまり語彙を介した想起をより利用して進んでいることがわかり (Fig.10 (a) ◇の一点鎖線)，模倣の学習において，より対応を経験できる語彙の学習の経験が利用され，相互促進的な学習が実現されていることがわかる．この結果は，他のいずれかのマッピングについての教示率が非常に低い場合についても再現されると考えられる．すなわち，ロボットに対応を与える養育者の行動に偏りがあり，ロボットがあるマッピングにおいてはほとんど対応を学習できない状況であっても，提案手法により，他のマッピングの学習経験を利用することで，相互促進的に対応の学習が可能となる．

5. 考　　察

5.1 学習手法としての位置づけ

機械学習の一つに半教師有り学習 [23] と呼ばれる学習手法がある．これは，データの識別器の学習問題において，データベースとしてラベル付けされたデータとともにラベル付けされないデータも与えられる時に，いかにして，高い汎化性能を得るかに注目した研究である [24] [25]．Nigam et al. や Ando and Zhang は，文書のカテゴリを識別する問題において，部分的にしかそのカテゴリのラベルが付与されていない文書のデータベースが与えられた時に，残りのラベル付けされていないデータについては，与えらえている部分的な対応関係から推定したラベルを用いる方法を提案している．これに対して本研究は，教師データが完全でないことを想定している点で類似しているが，問題として，間違ってラベル付けされたデータが与えられてしまう，ということを想定している点が異なる．

また，本研究では，音声模倣と語彙の二つの機能を同時に学習するためのメカニズムを提案したが，そのような複数機能の同時学習メカニズムは，これまでいくつか提案されている．Wolpert and Kawato が提案している MOSAIC [26] では，複数の学習モジュールの予測性能を評価し，環境の状態や文脈に対して適応的にモジュール割当と学習の重み付けを行うことで，各モジュールの効率の良い学習が実現されている．これに対して，提案手法では，学習すべきモジュールの割り当てには着目せず，音声模倣と語彙のそれぞれのモジュールの学習において，いかにして他のモジュールと相互促進的に学習を進められるかということに着目している点で異なると考えられる．また，Uchibe and Doya が提案している CLIS [27] では，複数モジュールがある中で，各モジュールの状態価値関数に従って学習すべきモジュールを選択することで，単純なモジュールから学習が進み，さら

にその学習における経験が複雑なモジュールの学習の手助けとなり，複雑なモジュールのみで学習させた場合よりも効率の良い学習が実現されている．提案手法も，一つのマッピングの学習に他のマッピングの学習経験を利用している点では同じである．しかし，CLIS では，状況に応じて適応的に選択されるモジュールによって得られた信号を，他のモジュールの学習に利用するのに対して，提案手法では，複数のモジュールによって得られた信号の競合的な統合によって学習に利用する信号を生成している点で異なると考えられる．

しかし，提案手法は，上記の半教師有り学習や複数機能の同時学習の手法を否定するものではなく，それらの手法とは相補的に機能しうると考えられ，これらと組み合わせた取り組みを行うことは今後の課題である．

5.2 乳児の発達との関連性

5.2.1 問題設定の考察

本研究では，乳児の音声模倣と語彙の発達における，マッピングとカテゴリ化の 2 つの課題のうち，特にマッピングの課題に取り組んだ．そのため，本来連続的な信号である養育者の音声を，乳児が離散的な表象として認識可能であることや養育者と乳児が共通のカテゴリを有していること，など，実際の発達途上の乳児が備えているものを超える能力を仮定してしまっていた．しかし本来は，そのような認識能力の発達、すなわちカテゴリ化を同時に考慮すべきである．音声の音圧や長さが，乳児が，聴取した音声から単語を切り出すためには重要であることが指摘されている [28]．つまり，音声の連続的な情報は，音声を一つの音韻或いは単語カテゴリとして認識することに影響すると考えられ，これはカテゴリ化の課題においては重要である．また，本研究では，養育者と乳児のカテゴリの数を同一としていたが，実際の養育者は，未熟な乳児に比べて圧倒的に多い音韻やラベルのカテゴリを有していると考えられる．さらに，そのように圧倒的に多いカテゴリ集合を持つ養育者の応答は，乳児にとっては，観測する刺激の変化として現れてくるものであり，そのような変化が，乳児のカテゴリ化の学習に影響すると考えられる．このような問題は，基本的には本研究で扱った教師信号の選択の問題と捉えられる．例えば，音声模倣のためのカテゴリ化の課題では，乳児が連続的に生成した構音運動パターンのどの部分が，その後聴取した養育者の連続音声パターンのどの部分に対応するかなどの時間的な対応を見つけることが重要であり，これによって音声を音韻系列やラベルなどのカテゴリとして認識できるようになると考えられる．そのため，カテゴリ化の課題に対して，本研究で提案した主観的整合機構のアイデアを適用し，音声模倣と語彙の共発達過程において，どのように相互促進的に各表象のカテゴリ化が可能であるかを検証することが，今後の課題である．

また，本研究では，音声模倣と語彙獲得に注目して共発達を議論したが，他の機能が共発達に及ぼす影響も考慮する必要がある．実際の乳児では，ある種の反射的な行動が生誕直後には備わっていることや [29] [30]，親の視線の理解 [31] を発達させていくことが知られている．これらの機能によって，一見乳児が親を模倣しているように親に感じさせることで，親の模倣行動を促進したり，乳児が親の視線方向を推定できるようになることで，教示対象を特定する手がかりが得られることが考えられる．従って，これらの発達に関する学習メカニズム [7] [32] と統合することで，より精緻な共発達モデルを得ることが今後の課題となる．

5.2.2 実験結果からの考察

実験 1 において，養育者の教示率が 0.2 以上の場合は，提案手法により音声模倣と語彙の発達のための対応学習が可能となるが，それ以下の教示率では，誤った学習をしてしまうことがわかった．ここで提案手法が誤った学習をしてしまうとき，対応していないものを対応しているとみなした対応学習が起こったと考えられた．これは発達障害の乳児や，発達途上の健常の乳児が，ときおり誤った対応音韻の転換や言葉の誤用を一貫して示すことと対応する現象であると考察される．しかしながら本稿のシミュレーションは，実際の母子相互作用を単純化し，養育者が固定の教示率で応じ続けること，また一定数の物体が環境に置かれていることを前提としている点に注意が必要である．一方、実際の養育者の応答は，乳児の学習の様子や相互作用の履歴によって変わりうるものであり，また，環境中の物体も乳児の発育に応じて変化し，その結果，養育者の教示率も変移するものであることが予想される．従って，これら環境側の変遷も考慮し，より精緻な問題設定の下でのシミュレーションを実施していくことが今後の課題となる．

実験 2 の結果から，提案手法は，実場面で見られるような，養育者が乳児にほとんど模倣を示さない状況においても，語彙の学習によって獲得した対応を利用することで，音声模倣のための対応の学習を可能とすることが示された．これは，言い換えれば，外部から与えられる音声信号を，一度語彙的なものへと歪ませて認識し，対応を学習すべきかを決定することで，音声模倣のための対応の学習が可能となった，と捉えることができる．一方，乳児は，発達初期では聴取した音声の音韻的な違いを識別可能であるが，語彙の学習が始まるに連れ，次第に音韻的に異なる音声でも同じ一つの語彙として識別するようになる [33]．そして，再び音韻的な違いを識別するようになるのは，さらに月齢が過ぎてからのようである [34]．本研究の実験結果から，このような乳児の聴取した音声に対する音韻的・語彙的な識別の切り替えは，音声模倣と語彙の共発達過程において，相互の学習を相互促進的に利用し合う結果として起こる可能性が示唆される．すなわち，初期の音韻的な識別が語彙の学習を促進することで，語彙的な識別を可能とし，語彙的な識別がさらに音韻の学習を促進することで，再び音韻的な識別を可能とする結果として起こると考えられる．今後は，提案手法の主観的整合度の遷移から，どのように識別能力の変化が引き起こされるのかを検証し，音声模倣や語彙の共発達過程の理解を深めることが必要である．

6. お わ り に

本研究では，音声模倣と語彙の共発達過程の構成的理解を目指し，音声模倣や語彙の発達に必要となる対応学習において相互促進的な学習を可能とする主観的整合度に基づく統合手法を提案した．また，ロボットと養育者とのインタラクション場面を想定した計算機シミュレーションを実施し，養育者の教示率

に対する提案手法の学習の頑健性を確認した．また，実場面の乳児と養育者のインタラクションで見られるような，養育者が乳児に対してほとんど模倣を示さない状況においても，提案手法により，語彙のための対応の学習を利用することで，正しく対応の学習が可能となることを確認し，音声模倣と語彙の共発達のための相互促進的な対応学習が実現された．

より実世界に則した母子相互作用であるほど，モデルの妥当性の検証に繋がることが期待されるが，人の振る舞いの多様性を無視することが困難となる．本研究で実施した実験結果から，提案手法は，そのような多様な振る舞いを経験し，一意に対応を見出すことが難しい状況においても対応学習を可能にすることが期待され，前節で述べた課題に取り組み，モデルを精緻化していくことで，計算機シミュレーションに閉じることが多かった従来の認知発達ロボティクスにおける社会知能のモデル化に対する取り組みを実世界で検証するための重要な一歩を導くと期待される．

参 考 文 献

[1] Minoru Asada, Koh Hosoda, Yasuo Kuniyoshi, Hiroshi Ishiguro, Toshio Inui, Yuichiro Yoshikawa, Masaki Ogino, and Chisato Yoshida. Cognitive developmental robotics: a survey. *IEEE Transactions on Autonomous Mental Development*, Vol. 1, No. 1, pp. 12–34, 2009.

[2] Frank H. Guenther, Michelle Hampson, and Dave Johnson. A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, Vol. 105, pp. 611–633, 1998.

[3] Hisashi Kanda, Tetsuya Ogata, Kazunori Komatani, and Hiroshi G. Okuno. Vocal imitation using physical vocal tract model. In *Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1846–1851, 2007.

[4] Katsushi Miura, Yuichiro Yoshikawa, and Minoru Asada. Unconscious anchoring in maternal imitation that helps finding the correspondence of caregiver's vowel categories. *Advanced Robotics*, Vol. 21, pp. 1583–1600, 2007.

[5] Deb K. Roy and Alex P. Pentland. Learning words from sights and sounds: a computational model. *Cognitive Science*, Vol. 26, No. 1, pp. 113–146, 2002.

[6] 菊池匡晃, 荻野正樹, 浅田稔. 顕著性に基づくロボットの能動的語彙獲得. 日本ロボット学会誌, Vol. 26, No. 3, pp. 261–270, 2008.

[7] 中野吏, 吉川雄一郎, 浅田稔, 石黒浩. 相互排他性原理に基づくマルチモーダル共同注意. 日本ロボット学会誌, Vol. 27, No. 7, pp. 814–822, 2009.

[8] Theano Kokkinaki and Giannis Kugiumutzakis. Basic aspects of vocal imitation in infant-parent interaction during the first 6 months. *Journal of reproductive and infant psychology*, Vol. 18, No. 3, pp. 173–187, 2000.

[9] Martha Pelaez-Nogueras, Jacob L. Gewirtz, and Michael M. Markham. Infant vocalizations are conditioned both by maternal imitation and motherese speech. *Infant Behavior and Development*, Vol. 19, No. 1, p. 670, 1996.

[10] Catherine S.Tamis-LeMonda, Marc H.Bornstein, and Lisa Baumwell. Maternal responsiveness and children's achievement of language milestones. *Child Development*, Vol. 72, No. 3, pp. 748–767, 2001.

[11] Julie Gros-Luis, Meredith J.West, Michael H.Goldstein, and Andrew P.King. Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, Vol. 30, No. 6, pp. 509–516, 2006.

[12] Gregory Hickok and David Poeppel. The cortical organization of speech processing. *Nature Reviews*, Vol. 8, pp. 393–402, 2007.

[13] Elizabeth Bates, Philip S. Dale, and Donna Thal. *The Handbook of Child Language*, chapter 4: Individual Differences and their Implications for Theories of Language Development, pp. 96–151. Blackwell Publishing, 1995.

[14] Susan S. Jones. Imitation in infancy the development of mimicry. *Psychological Science*, Vol. 18, pp. 593–599, 2007.

[15] Elise Frank Masur and Doreen L.Eichorst. Infants' spontaneous imitation of novel versus familiar words: Relations to observational and maternal report measures of their lexicons. *Merrill-Palmer Quarterly*, Vol. 48, No. 4, pp. 405–426, 2002.

[16] Janet F.Werker and Suzanne Curtin. Primir: A developmental framework of infant speech processing. *Language Learning and Development*, Vol. 1, No. 2, pp. 197–234, 2005.

[17] Susan S.Jones. Infants learn to imitate by being imitated. In *Proceedings of IEEE/RSJ International Conference on Development and Learning*, 2006.

[18] Lakshmi J. Gogate, Laura H. Bolzani, and Eugene A. Betancourt. Attention to maternal multimodal naming by 6- to 8-month-old infants and learning of word-object relations. *Infancy*, Vol. 9, No. 3, pp. 259–288, 2006.

[19] David H.Ackley, Geoffrey E. Hinton, and Terrence J.Sejnowski. A learning algorithm for boltzmann machines. *Cognitive Science*, Vol. 9, pp. 147–169, 1985.

[20] Patricia K.Kuhl. Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, Vol. 5, No. 11, pp. 831–843, 2004.

[21] Barbara A.Younger and Leslie B.Cohen. Infant perception of correlations among attributes. *Child Development*, Vol. 54, No. 4, pp. 858–867, 1983.

[22] http://baby.goo.ne.jp.

[23] Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien, editors. *Semi-supervised Learning*, chapter 1: Introduction to Semi-Supervised Learning, pp. 2–12. MIT Press, 2006.

[24] Kamal Nigam, Andrew McCallum, Sebastian Thrun, and Tom Mitchell. Learning to classify text from labeled and unlabeled documents. In *Proceedings of the 15th National Conference on Artificial Intelligence*, pp. 792–799, 1998.

[25] Rie K.Ando and Tong Zhang. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, Vol. 6, pp. 1817–1853, 2005.

[26] Daniel M. Wolpert and Mitsuo Kawato. Multiple paired forward and inverse models for motor control. *Neural Networks*, Vol. 11, No. 7-8, pp. 1317–1329, 1998.

[27] Eiji Uchibe and Kenji Doya. Competitive-cooperative-concurrent reinforcement learning with importance sampling. In *Proceedings of International Conference on Simulation of Adaptive Behavior: From Animals and Animates*, pp. 287–296, 2004.

[28] 佐藤久美子, 梶川祥世, 坂本清恵, 松本博文. 日本語母語乳児の文中からの単語切り出しにおけるアクセントと音素配列の役割. 音声研究, Vol. 2007, No. 3, pp. 38–47, 11.

[29] Andrew N. Meltzoff and M. Keith Moore. Newborn infants imitate adult facial gestures. *Child Development*, Vol. 54, No. 3, pp. 702–709, 1983.

[30] Xin Chen, Tricia Striano, and Hannes Rakoczy. Auditory-oral matching behavior in newborns. *Developmental Science*, Vol. 7, No. 1, pp. 42–47, 2004.

[31] Dare A. Baldwin. Infants' contribution to the achievement of joint reference. *Child Development*, Vol. 62, No. 5, pp. 875–890, 1991.

[32] Andrew N. Meltzoff and M. Keith Moore. Explaining facial imitation: A theoretical model. *Early Development and Parenting*, Vol. 6, pp. 179–192, 1997.

[33] Christine L.Stager and Janet F.Werker. Infants listen for more phonetic detail in speech perception than in word-learning

tasks. *Nature*, Vol. 388, pp. 381–382, 1997.
[34] Janet F.Werker and Christopher E.Fennell. *Waving a Lexicon*, chapter From listening to sounds to listening to words: Early steps in word learning, pp. 79–109. MIT Press, 2004.

付録 A. 結合強度の更新式の導出

相互想起型ボルツマンマシン [19] では，入力 V_1 とその出力 V_2 が与えられたとき，ネットワークを介して想起される信号の分布 $P'(\boldsymbol{V_2} \mid \boldsymbol{V_1})$ を，可能な限り入力信号の分布 $P(\boldsymbol{V_2} \mid \boldsymbol{V_1})$ に近づけるように学習が行われ，その評価関数 G は以下のように表される．

$$G=\sum_{V_1,V_2} P(\boldsymbol{V_1},\boldsymbol{V_2})\left\{ log \frac{P(\boldsymbol{V_2}|\boldsymbol{V_1})}{P'(\boldsymbol{V_2}|\boldsymbol{V_1})} \right\}. \quad \text{(A.1)}$$

ここで，右辺第一項は，V_1 から V_2 へ想起した場合の信号の分布と，実際の入力の分布との近さを表し，G の値が小さい程，2つの分布が近いことを表す．そのため，最急降下法により，この評価関数 G をネットワークの結合強度 w_{nm} で偏微分した量を用い，以下のように結合強度が更新される．

$$w_{nm} = w_{nm} - \varepsilon \frac{\partial G}{\partial w_{nm}}. \quad \text{(A.2)}$$

式 (A.2) の右辺第二項のエネルギー関数の偏微分は次のように表される．

$$\frac{\partial G}{\partial w_{nm}} = -\frac{1}{T}\Big\{ \sum_H P'(\boldsymbol{H}|\boldsymbol{V_1},\boldsymbol{V_2})^1u_n{}^1u_m + - \sum_{V_2',H} P'(\boldsymbol{V_2'},\boldsymbol{H'}|\boldsymbol{V_1})^{1'}u_n'{}^{1'}u_m' \Big\}. \quad \text{(A.3)}$$

ここで，H は隠れ層における状態を，n, m は結合する2つのノード番号を，${}^1u_n$ はノード n の状態 (0, 1 の 2 値) を表す．${}^1u_n$ と ${}^{1'}u_n$ は同じ変数であるが，$\sum$ によって総和を計算する時に，別々に動かすことを意味するために左上の添字 (^{1})，(${}^{1'}$) を付与している

ボルツマンマシンでは，入力層，隠れ層，出力層の状態を，ネットワークを介して繰り返し更新することで，$\frac{\partial G'}{\partial w_{nm}}$ を計算する．すなわち，式 (A.3) における，右辺第一項の $\sum_H P'(\boldsymbol{H} \mid \boldsymbol{V_1},\boldsymbol{V_2})^1u_n{}^1u_m$ は，ネット ワークの入出力を V_1，V_2 で固定して，隠れ層の状態を繰り返し更新した時の，結合する2つのノードの状態 ${}^1u_n$ と ${}^1u_m$ が同時に 1 になる頻度 k_{nm} として計算される．また，右辺第二項の $\sum_{V_2',H} P'(\boldsymbol{V_2'},\boldsymbol{H'} \mid \boldsymbol{V_1})^{1'}u_n'{}^{1'}u_m'$ は，ネットワークの入力を V_1 で固定し，隠れ層と出力層の状態を繰り返し更新した時の，結合する2つのノードの状態 ${}^1u_n'$ と ${}^1u_m'$ が同時に 1 になる頻度 k_{nm}' として計算される．これより，係数部をまとめて $\alpha = -\frac{1}{T}\varepsilon$ とすると，結合荷重の更新式は最終的に以下のように表される．

$$w_{nm} = w_{nm} + \alpha\left(k_{nm} - k'_{nm}\right). \quad \text{(A.4)}$$

これに対して，本研究では，双方向の想起を考慮するために，入出力関係を入れ替えて，ネットワークを介して逆方向に想起される信号の分布 $P'(\boldsymbol{V_1} \mid \boldsymbol{V_2})$ を，可能な限りその出力信号の分布 $P(\boldsymbol{V_1} \mid \boldsymbol{V_2})$ に近づけることも考える．すなわち，式 (A.1) に，逆方向の項も追加した以下の評価関数 G' を考える．

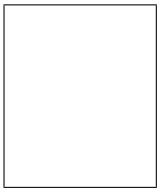
$$G'=\sum_{V_1,V_2} P(\boldsymbol{V_1},\boldsymbol{V_2})\left\{ log \frac{P(\boldsymbol{V_2}|\boldsymbol{V_1})}{P'(\boldsymbol{V_2}|\boldsymbol{V_1})} + log \frac{P(\boldsymbol{V_1}|\boldsymbol{V_2}))}{P'(\boldsymbol{V_1}|\boldsymbol{V_2})} \right\}. \quad \text{(A.5)}$$

ここで，新しく加えた右辺第二項は，V_2 から V_1 へ想起した場合の信号の分布と，実際の出力の分布との近さを表す．式 (A.1) と同様に，G' の値が小さい程，2つの分布が近いことを表す．式 (A.3) と同様に展開すると，エネルギー関数の偏微分は次のように表される．

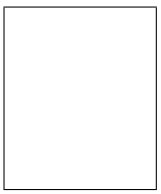
$$\frac{\partial G'}{\partial w_{nm}} = -\frac{1}{T}\Big\{ \sum_H P'(\boldsymbol{H}|\boldsymbol{V_1},\boldsymbol{V_2})^1u_n{}^1u_m + \sum_H P'(\boldsymbol{H}|\boldsymbol{V_2},\boldsymbol{V_1})^2u_n{}^2u_m - \sum_{V_2',H} P'(\boldsymbol{V_2'},\boldsymbol{H'}|\boldsymbol{V_1})^{1'}u_n'{}^{1'}u_m' - \sum_{V_1',H} P'(\boldsymbol{V_1'},\boldsymbol{H'}|\boldsymbol{V_2})^{2'}u_n{}^{2'}u_m \Big\}. \quad \text{(A.6)}$$

式 (A.6) の右辺第一，第二項は，式 (A.3) の右辺第一項に $\sum_H P'(\boldsymbol{H} \mid \boldsymbol{V_2},\boldsymbol{V_1})^2u_n{}^2u_m$ を足したものであり，これは，ネット ワークの入出力を V_1，V_2 で固定して，隠れ層の状態のみを繰り返し更新した時の，結合する2つのノードの状態 ${}^1u_n$ と ${}^1u_m$ が同時に 1 になる頻度を2度 (2度目は，式の上では ${}^2u_n$ と ${}^2u_m$ と表現している) 繰り返して計算することを表す．また，右辺第三，第四項は，式 (A.3) の右辺第二項に $\sum_{V_1',H} P'(\boldsymbol{V_1'},\boldsymbol{H'} \mid \boldsymbol{V_2})^{2'}u_n{}^{2'}u_m$ を足したものであり，これは，ネットワークの入力を V_1 で固定し，隠れ層と出力層の状態を繰り返し更新した時の，結合する2つのノードの状態 ${}^1u_n'$ と ${}^1u_m'$ が同時に 1 になる頻度と，出力を V_2 で固定し，隠れ層と入力層の状態を繰り返し更新した時の，結合する2つのノードの状態 ${}^2u_n'$ と ${}^2u_m'$ が同時に 1 になる頻度の和として計算される．これら，右辺第一，第二項として計算される頻度を K_{nm}，右辺第三，第四項として計算される頻度を K'_{nm} とすると，最終的な結合強度の更新式は，以下のように表される．

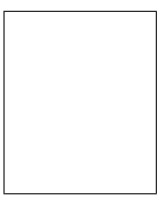
$$w_{nm} = w_{nm} + \alpha\left(K_{nm} - K'_{nm}\right). \quad \text{(A.7)}$$



笹本 勇輝（Yuki Sasamoto）
2010 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前期課程修了．同年同大学博士後期課程に入学し現在に至る．認知発達ロボティクスの研究に従事． （日本ロボット学会学生会員）



吉川 雄一郎（Yuichiro Yoshikawa）
2005 年大阪大学大学院工学研究科知能・機能創成工学専攻終了．博士（工学）．同年 ATR 知能ロボティクス研究所研究員．2006 年 JST ERATO 浅田共創知能システムプロジェクト研究員．2010 年より大阪大学大学院基礎工学研究科講師となり現在に至る．認知発達ロボティクスの研究に従事．
（日本ロボット学会正会員）



浅田 稔（Minoru Asada）
1982 年大阪大学大学院基礎工学研究科後期課程修了．1989 年大阪大学工学部助教授．1995 年同教授．1997 年大阪大学大学院工学研究科知能・機能創成工学専攻教授．工学博士 (大阪大学) となり現在に至る．この間，1986 年から 1 年間米国メリーランド大学客員研究員．1989 年，情報処理学会研究賞，1992 年，IEEE/RSJ IROS'92 Best Paper Award．1996 年日本ロボット学会論文賞，1997 年人工知能学会研究奨励賞，1999 年日本機

械学会ロボティクス・メカトロニクス部門貢献賞，2001 年文部科学大臣賞・科学技術普及啓発功績者賞，2001 年日本機械学会ロボティクス・メカトロニクス部門賞：学術業績賞，2006 年科学技術政策研究所科学技術への顕著な貢献 in 2006 ナイスステップな研究者「イノベーション部門」，2007 (財) 大川情報通信基金大川出版賞，2008 年 2008 グッドデザイン賞，2009 年日本ロボット学会論文賞それぞれ受賞．博士 (工学)．電子情報通信学会，情報処理学会，人工知能学会，日本機械学会 (フェロー)，計測自動制御学会，システム制御情報学会，日本赤ちゃん学会 (理事)，IEEE RAS, CS,SMC societies などの会員．NPO RoboCup 日本委員会理事，RoboCup 国際委員前プレジデント．（日本ロボット学会正会員）