# 主観的コンシステンシーに基づく模倣と語彙の共発達

○笹本勇輝 (JST ERATO，阪大)，吉川雄一郎 (JST ERATO)，
浅田稔 (JST ERATO，阪大)

## Selective integration based on subjective consistency facilitates simultaneous development of vocal imitation and lexicon acquisition

Yuki Sasamoto (JST ERATO, Osaka Univ.), Yuichiro Yoshikawa (JST ERATO),
Minoru Asada (JST ERATO, Osaka Univ.)

**Abstract**— This paper presents a method for simultaneous development of vocal imitation and lexicon acquisition through mutually associative cross-modal mapping. Subjective consistency is introduced to determine the value of a certain layer by combining with one from another layer and a directly receiving external input. In the computer simulation, the proposed method is applied to learn mappings among representations of caregiver's phonemes, those of own phonemes, and those of objects. It turns out that the proposed mechanism enables correct mappings even when the caregiver does not always give correct examples.

**Key Words:** Subjective consistency, vocal imitation, lexical acquisition, mutually associative cross-modal mapping

## 1. はじめに

人が他者とコミュニケーションする上で重要な能力の一つに言語能力がある．言語によって他者とコミュニケーションするためには，他者の言葉を理解し，同じように言葉を生成できる必要がある．人の乳児は 12ヶ月頃には初語がみられはじめ [1]，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 [2]．一方，語彙の範疇に入らないようなものでも，単母音のように単純な音声については 6ヶ月頃から，複母音では 14ヶ月頃から模倣を示すようになることが報告されている [3]．これらの発達は，時期的に重複しているようであり，また機能的にも相互に影響し合うと考えられる．すなわち，模倣ができることで，初めて聞いた言葉でもすぐに発話できたり，語彙を知っていることで，聞き取りにくいような言葉でも正しく模倣ができる等の形で，それぞれの発達が互いに促進し合っていると考えられるが，どのようなメカニズムで，そのような発達の相互作用が可能になっているのだろうか？

近年，このような発達の仕組みについての問いに対し，ロボットなど統制可能な計算モデルを用い，構成論的にその仕組みに迫るアプローチが注目されている [4]．従来の構成的研究では，語彙獲得過程は視覚情報と物体ラベルの相関学習としてモデル化されることが多かった [5]．一方，模倣発達に関しては，聞こえてくる音声と自身の構音運動の関係を統計学習過程としてモデル化する研究 [6] やそれを導く養育者の振る舞いの性質に注目した研究がある [7]．しかし，語彙獲得と視線追従能力の発達過程の相互作用を取り扱った研究 [8] はあるものの，従来の構成的研究では，語彙獲得と模倣発達は別々に研究され，これらの相互作用の仕組みについては十分に議論されていない．

語彙獲得と模倣発達を同時に考えるということは，それぞれの発達において学習される対応関係を相互に利用できるようになることにつながる．例えば，模倣に必要となる自分と相手の音声表象間の連関学習を考える場合，相手が常に自分の声を模倣しているとは限らないため，自分の音声と対応する相手の音声が得られるとは限らない．しかし自分の音声表象と語彙の表象間の対応及び語彙の表象と相手の音声表象間の対応がわかっていれば，これらにより，自分の音声から相手の音声を推定し，聴取した相手の音声が対応するものであるかの判断がある程度できると考えられる．しかし，対応関係が未学習である場合など，これらを利用した判断が常には正しいとは限らず，どの程度信頼するかを状況によって決められる仕組みが必要である．

そこで本研究では，模倣と語彙の共発達を，乳児自身の発話，養育者の発話，及び注目物体の 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し，それぞれのマッピングを，マッピング同士の主観的コンシステンシーに基づいて選択的に利用することで相補的に学習される手法を提案する．本稿では，その学習メカニズムと主観的コンシステンシーに基づく統合手法について説明し，それらの妥当性を検証するために行った計算機シミュレーションの結果について述べる．

## 2. 仮定

Fig.1 のようなロボットと養育者と物体が存在する環境を想定する．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．そして，同時に発話を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，模倣，提示，教示の 3 つの行動の何れかを選択する．養育者は，無意識的に，あるいはロボットが知りたがっていることを汲み取ることで，模倣，提示，教示，の行動を選択していると考えられる．本研究ではその戦略を，恣意的に Fig.2 のような確率モデルであると想定する，すなわち，養育者の行動は，ロボットが自分のことを見て

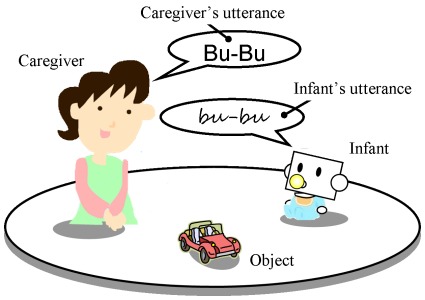



**Fig.**1 Assumed environment of caregiver-infant inreraction

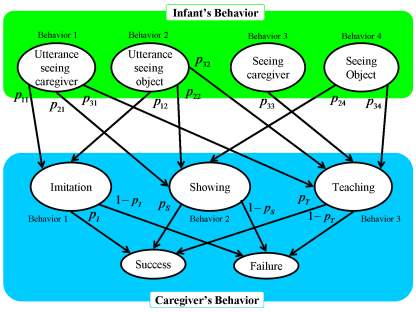



**Fig.**2 Response rule of the caregiver

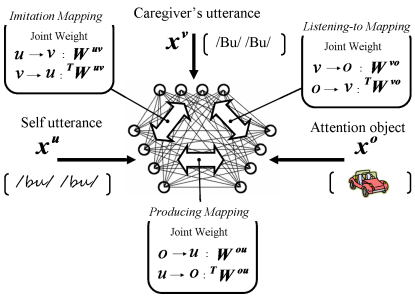



**Fig.**3 Mutually associative cross-modal mapping

いるか，また言葉を発しているか，の状況ごとに事前に定められた確率 $(p_{11},p_{21},p_{31},p_{12},p_{22},p_{32},p_{33},p_{24},p_{34})$ に応じて，模倣，提示，教示，の何れかから選択される．さらに，ロボットの発話が未熟であることや，養育者の行動が常にはロボットの行動に対応しないことを考慮し，各行動ごとに定められた成功率 $(p_I,p_S,p_T)$ によって，その成否が決定されるとする．

上述のやりとりを通じてロボットは，Fig.3 のような，ロボットの発話，養育者の発話，および注目物体の 3 つの表象が相互に結合するマッピングを学習する．各表象は，ベクトル $\boldsymbol{x}^u \in \Re^{M_i}$，$\boldsymbol{x}^v \in \Re^{M_c}$，および $\boldsymbol{x}^o \in \Re^N$ として表記され，ロボットは観測からこれらを形成できるとする．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，ロボットは各層同士の相互結合マッピング，すなわち，ノード間の結合強度を要素とする行列 $\boldsymbol{W}^{uv}$，$\boldsymbol{W}^{vo}$，$\boldsymbol{W}^{ou}$ を学習する．ここで，$\boldsymbol{W}^{uv}$ は自身の発話と養育者の発話との対応 (模倣マッピング)，$\boldsymbol{W}^{vo}$ は養育者の発話と物体との対応 (語彙聴取マッピング)，$\boldsymbol{W}^{ou}$ は自身の発話と物体との対応 (語彙生成マッピング) を表す．

## 3. 共発達メカニズム

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．ここで，はじめに入力があった層の ID を $i$，次に入力があった層の ID を $j$，残りの層の ID を $k$ とする．$i$ 層への入力が $\boldsymbol{x}^i$ であったとき，ロボットは，$j$ 層への入力 $\boldsymbol{x}^j$ の予測ベクトル $\boldsymbol{x}^{ij}$ の $m$ 番目の要素 $x^{ij}_m$ を

$$P\left(x^{ij}_m|\boldsymbol{W}^{ij}\right) = \frac{1}{1+\exp\left(-\sum_n W^{ij}_{nm}x^i_n\right)} \quad (1)$$

の確率に従って生成する．ここで，$\boldsymbol{W}^{ij}$ は $i$ 層から $j$ 層への結合強度であり，$W^{ij}_{nm}$ は $\boldsymbol{W}^{ij}$ の $n$ 行 $m$ 列の要素である．ロボットが養育者の発話を模倣するため，また養育者が発話したラベルに対応する物体を特定するため，あるいは注目物体のラベルを発話するためには，各マッピングの結合強度行列を正しく学習する必要がある．例えば，ロボットが養育者の発話を模倣するためには，養育者の発話から自身の発話及び自身の発話から養育者の発話を正しく予測できていることが必要であり，養育者の発話から予測される自身の発話の分布 $P\left(x^{ij}_m|\boldsymbol{W}^{ij}\right)$ が，養育者を模倣した時の自身の発話の分布となるような結合強度 $\boldsymbol{W}^{ij}$ を求めることが必要となる．

式 (1) のようなマッピングの出力の確率がロジスティック関数で表されるモデルは，ロジスティック信念ネットと呼ばれ，Contrastive Divergence 学習と呼ばれる効率的な学習アルゴリズムが提案されている [9] [10]．本研究では，それらの研究における隠れ変数に対して，入力によってバイアスをかけることができるため，結合強度の更新則を次式のように拡張する．

$$\Delta W^{ij}_{nm} = \varepsilon\left(\langle x^{i,0}_n x^j_m\rangle_{P^{i,0}P^j} - \langle x^{i,1}_n x^j_m\rangle_{P^{i,1}_{\boldsymbol{W}^{ij}}P^j}\right) \quad (2)$$

ここで，$\langle\cdot\rangle_P$ は確率分布 $P$ に関する期待値を表し，$x^{i,0}_n$，$x^j_m$ はそれぞれ $i$ 層への入力 $\boldsymbol{x}^{i,0}$ の $n$ 番目の要素，$j$ 層への入力 $\boldsymbol{x}^j$ の $m$ 番目の要素であり，$P^{i,0}$ $P^j$ はこれらの同時確率分布である．例えば，ロボットが発話したのち，養育者がこれに模倣する場合，$\boldsymbol{x}^{i,0}$ は自らの発話の表象，$\boldsymbol{x}^j$ は養育者の発話の表象となる．また，$x^{i,1}_n$ は，$j$ 層への入力 $\boldsymbol{x}^j$ から結合強度行列 ${}^T\boldsymbol{W}^{ij}$ を用いて再構築される $i$ 層への予測 $\boldsymbol{x}^{i,1}$ の $n$ 番目の要素であり，$P^{i,1}_{\boldsymbol{W}^{ij}}$ $P^j$ は $\boldsymbol{x}^{i,1}$ と $\boldsymbol{x}^j$ の同時確率分布である．

各ステップにおいて，式 (2) による更新量を，入出力を入れ替えたものについても計算し，

$$\boldsymbol{W}^{ij} = \boldsymbol{W}^{ij} + \left(\Delta\boldsymbol{W}^{ij} + {}^T\Delta\boldsymbol{W}^{ji}\right) \quad (3)$$

のように更新する．ただし，式 (2) 中の期待値計算は各ステップのサンプリング値で代用した．

### 3·1 主観的コンシステンシーに基づく統合

養育者が乳児の発話の模倣や提示行動を常に行う場合，養育者の発話や物体と正しく対応する自身の発話の分布を学習することができるが，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，誤った分布を学習する恐れがある．

これに対して，提案手法では，マッピング同士が相互に結合しており，他のマッピングからの予測も学習に使用できるため，どのマッピングがより信頼できる出力をしているのかを判断し，その出力を学習に反映させることで，誤った対応関係の学習を防ぐことができると考えられる．しかしながら，マッピングの信頼度を学習者自身が判断するとなれば，学習者が主観的に観測・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．そこで，本節では，マッピングの信頼性を主観的に判断するためのコンシステンシーに基づいた統合手法を提案する．

はじめに入力があった層の ID が $i$，次に入力があった層の ID が $j$ の場合を考える．この時，$j$ 層への入力 $\boldsymbol{x}^j$ (以下，外部入力)，$i$ 層 から $j$ 層への直接のマッピングからの予測 $\boldsymbol{x}^{ij}$ 及び $i$ 層から $k$ 層を介した $j$ 層へ

の間接のマッピングからの予測 $\boldsymbol{x}^{kj}$ (以下，内部入力) を統合した $\hat{\boldsymbol{x}}^{j}$ を，次式のように表す．

$$\hat{\boldsymbol{x}}^{j} = f(\boldsymbol{x}^{i}, \boldsymbol{x}^{j}, \boldsymbol{x}^{kj}) = \lambda_i \boldsymbol{x}^{i} + \lambda_{ij} \boldsymbol{x}^{ij} + \lambda_{kj} \boldsymbol{x}^{kj} \quad (4)$$

ここで，$\lambda_n$ は 外部又は内部入力 $\boldsymbol{x}^{n}$ のコンシステンシーの度合いを表し，

$$\lambda_n = \frac{\exp\left(-\prod_{l,l\notin n} \frac{\left\|\boldsymbol{x}^{n} - \boldsymbol{x}^{l}\right\|}{\sigma^2}\right)}{\sum_{m,m\in\{i,ij,kj\}} \exp\left(-\prod_{o,o\notin m} \frac{\left\|\boldsymbol{x}^{m} - \boldsymbol{x}^{o}\right\|}{\sigma^2}\right)} \quad (5)$$

と計算される．ここで，$\sigma$ はコンシステンシーに対する感度パラメータである．式 (5) より，外部又は内部入力 $\boldsymbol{x}^{n}$ が他の入力と近い値である程より信頼できるものであると判断し，その入力の重みであるコンシステンシーの度合い $\lambda_n$ が大きな値となるように計算される．

以上のように，外部入力だけでなく内部入力として自身のマッピングからの予測も考慮して，主観的に入力と捉えることで，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してしまう問題を回避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，他の信号との間で矛盾のより少ない予測又は入力をより信頼できると判断し，それを $j$ 層への入力として選択することができる．統合を利用する場合の結合強度の更新則は以下のようになる．

$$\Delta W^{ij}_{nm} = \varepsilon \left( \langle x^{i,0}_n \hat{x}^{j,0}_m \rangle_{P^{i,0} P^{j,0}_{\boldsymbol{W}^{ij}}} - \langle x^{i,1}_n \hat{x}^{j,1}_m \rangle_{P^{i,1}_{\boldsymbol{W}^{ij}} P^{j,1}_{\boldsymbol{W}^{ij}}} \right) \quad (6)$$

ここで，$\hat{x}^{j,0}_m$, は　コンシステンシーに基づく統合により予測された $j$ 層への入力 $\hat{\boldsymbol{x}}^{j,0} = f(\boldsymbol{x}^{j,0}, \boldsymbol{x}^{ij,0}, \boldsymbol{x}^{kj,0})$ の $m$ 番目の要素であり，$P^{i,0}$ $P^{j,0}_{\boldsymbol{W}^{ij}}$ は $\boldsymbol{x}^{i,0}$ と $\hat{\boldsymbol{x}}^{j,0}$ の同時確率分布である．また $x^{i,1}_n$ は，$\hat{\boldsymbol{x}}^{j,0}$ から再構成された $i$ 層への入力の予測 $\boldsymbol{x}^{i,1} = f(\hat{\boldsymbol{x}}^{ji,0})$ の $n$ 番目の要素であり，$\hat{x}^{j,1}_m$ はコンシステンシーに基づく統合により予測された $j$ 層への入力 $\hat{\boldsymbol{x}}^{j,1} = f(\boldsymbol{x}^{j,0}, \boldsymbol{x}^{ij,1}, \boldsymbol{x}^{kj,1})$ の $m$ 番目の要素であり，$P^{i,1}_{\boldsymbol{W}^{ij}}$ $P^{j,1}_{\boldsymbol{W}^{ij}}$ は $\boldsymbol{x}^{i,1}$ と $\hat{\boldsymbol{x}}^{j,1}$ の同時確率分布である．統合を行う場合でも，期待値を各ステップにおけるサンプリング値で代用し，式 (3) と同様の計算により更新量を決定する．

## 4. 共発達シミュレーション

提案手法の妥当性を確かめるために，計算機シミュレーションを実施した．実験では，ロボットの発話に対して養育者が模倣ではない発話や対応しない物体を提示するような誤った入力が与えられる場合でも頑健に対応関係を学習可能であるかを検証した．また，養育者がロボットの発話をまったく模倣しない場合でも，養育者の提示や教示から学習される対応関係を利用することで，模倣に関する対応関係を学習可能であるかを検証した．

### 4·1 実験設定

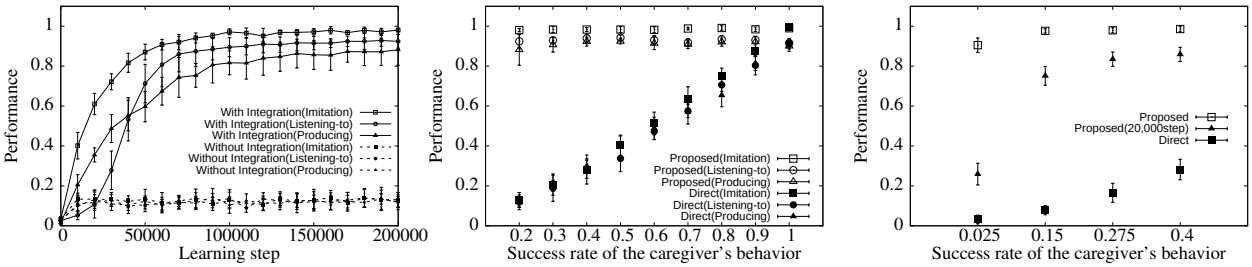
ここでは簡単のため，ロボットと養育者の発話はお互いが共通に持ついくつかのモーラで構成されるものとする．すなわち，$\boldsymbol{x}^{u}$，$\boldsymbol{x}^{v}$ は，それぞれが $M$ 種類あるうちのどのモーラで構成されているかを表わすベクトルであるとした．例えば，ロボットの発話が $/m_8$ $m_2/$ のように，8 番目のモーラと 2 番目のモーラの組合せで構成される音であった場合，$\boldsymbol{x}^{u}$ は，2 番目と 8 番目の要素が 1，それ以外の要素が 0 であるベクトルとなる．また，物体は，$N$ 種類のどの物体について注目しているかを表わすベクトル $\boldsymbol{x}^{o}$ 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，$\boldsymbol{x}^{o}$ は $i$ 番目の要素が 1，それ以外が 0 であるベクトルとなる．

ロボットは，2 章で説明した 4 つの行動の中からステップ毎の行動をランダムに選択し，養育者は，ロボットが自分を見て発話している場合には模倣を，物体を見て発話している場合には提示を，自分を見ているまたは物体を見ているだけの場合は教示を必ず行うように設定した．すなわち，$p_{11}$，$p_{22}$，$p_{33}$，$p_{34}$ を 1 とし，それ以外を 0 とした．ただし，2 章で説明したように，養育者は正しい模倣を $p_I$，提示を $p_S$，教示を $p_T$ の割合でしか行わない．養育者が正しい模倣や提示，教示を行わない場合，ロボットにはロボットの発話や注意に関係なく，環境中の物体がランダムに選ばれ，そのラベルの発話や物体の提示が行われる．また，環境中の物体及びラベルの数は $N = 39$ 個とし，ロボットと養育者は $M = 37$ 個のモーラを持っているものとする[1]．また，提案手法における各パラメータは経験的に $\varepsilon = 0.2$，$\sigma = 1.0$ とした．提案手法 (式 (6) に従う方法) のパフォーマンスを，統合を利用しないで与えられる入力のみから対応関係を学習する方法 (式 (2) に従う方法，以下直接手法) と比較することで，提案手法の有効性を評価する．

### 4·2 実験結果

まず，養育者の間違いに対する提案手法の頑健性を確かめるために，養育者の正しい行動の割合 $p_I = p_S = p_T$ を 0.2 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，200,000 回の母子相互作用シミュレーションを 10 回実施した．$p_I = p_S = p_T = 0.2$ の場合の各マッピングの学習パフォーマンスの 10 回の試行に関する平均値の遷移を Fig.4(a) に示す．パフォーマンスは，39 通りの可能な入力ベクトルをそれぞれの層に入力した時に各マッピングによって計算される予測ベクトルが 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とした．すなわち，パフォーマンスが 1 に近い程，より正しい対応関係を学習していることを示す．直接手法 (破線) に比べ，提案手法 (実線) の方が，すべてのマッピングの学習で，高いパフォーマンスを示している．養育者の正しい行動の割合と各マッピングの学習の最終的なパフォーマンスの関係を Fig.4(b) に示す．Fig.4(b) より，養育者が正しく模倣や提示，教示を行う条件では，直接手法 (黒四角，黒丸，黒三角のポイント) でもすべてのマッピングで高いパフォーマンスを示している．しかし，養育者の正しい行動の割合が減少するにつれ，パフォーマンスが減少し，誤った対応関係を学習している．それに

[1] 2009 年 2 月 22 日時点で goo ベビー (http://baby.goo.ne.jp) 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の中から乳児と養育者以外の対象の名前を表す名詞単語を抽出し，それをモーラ毎に離散化させたデータを学習データとして使用した．




(a) Transition of performance of each mapping ($p_I = p_S = p_T = 0.2$)

(b) Final performance of each mapping with respect to the success rate of caregiver's behavior

(c) Final and intermediate (20,000 step) performance of each mapping with respect to the success rate of caregiver's imitation ($p_S = p_T = 0.4$)

**Fig.**4 Experimental results

対して，提案手法 (白四角，白丸，白三角のポイント) では，養育者の正しい行動の割合が減少しても，頑健に正しい対応関係を学習していることがわかる．

次に，提案手法により養育者がロボットの発話をまったく模倣しない場合でも模倣マッピングの学習が可能であるかを確かめるために，$p_S = p_T = 0.4$ と固定し，$p_I$ を 0.025 から 0.4 まで 0.125 ずつ変化させたそれぞれの場合について，200,000 回の母子相互作用シミュレーションを 10 回実施した．$p_I$ と模倣マッピングの最終的なパフォーマンスの関係 (Fig.4(c) 参照) より，養育者がまったく模倣しない状況下，すなわち $p_I$ がチャンスレベルである 0.025 に設定された場合でも，提案手法 (白四角のポイント) では，高いパフォーマンスを示していることがわかる．また，提案手法による学習の途中段階 (20,000 ステップ時点) でのパフォーマンス (黒三角のポイント) から，養育者が模倣を行う割合が増加するにつれ，学習が加速されていることがわかる．

## 5. おわりに

本研究では，模倣と語彙の共発達メカニズムの構成を目指し，語彙と模倣に関わる表象が相互に結合したモデルを考え，主観的コンシステンシーに基づく統合を利用した学習メカニズムを提案した．提案メカニズムにより，養育者がロボットの行動に対して常に正しく対応する行動をしない状況やまったく模倣しない状況であっても，自身のマッピングからの予測も考慮して選択的に統合が行われることで，語彙や模倣能力の獲得に必要な対応関係の学習が可能であることを計算機シミュレーションにより確認した．

本稿では，語彙や模倣に関わる表象として，自身の発話の表象，養育者の発話の表象，注目物体の表象を用い，それぞれの表象に与えられる情報をすでに範疇化されたベクトルとして扱った．しかし適応性や発達モデルとしての妥当性の観点から，これらがセンサ信号や運動指令からどのように範疇化可能であるのかも考慮する必要がある．また，本稿の実験では，ロボットの行動がランダムに決定され，養育者はそれに完全に応じるように設定して計算機シミュレーションを実施した．しかし，実際の乳児は，そのようにランダムに行動しているわけではなく，養育者の行動や自身の前の行動，又は学習の習熟度などから能動的に自身の行動を決定していると考えられる．本研究で提案した，主観的コンシステンシーによる各マッピングの信頼度の評価は，そのまま学習の習熟度の評価へ応用できると考えられる．このように，実際の乳児が置かれている状況を想定しモデルを修正していくことで，より妥当性のあるモデルへと発展させていくことが今後の課題となる．

## 参考文献


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

[2] 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. *Journal of Child Language*, Vol. 33, pp. 1–29, 2006.

[3] S.S.Jones. Imitation in infancy the development of mimicry. *Psychological Science*, Vol. 18, No. 7, pp. 593–599, 2007.

[4] M.Asada, K.Hosoda, Y.Kuniyoshi, H.Ishiguro, T.Inui, Y.Yoshikawa, M.Ogino, and C.Yoshida. Cognitive developmental robotics: a survey. *IEEE Transactions on Autonomous Mental Development*, Vol. 1, No. 1, pp. 12–34, 2009.

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

[6] H.Kanda, T.Ogata, K.Komatani, and H.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.

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

[8] Y.Yoshikawa, T.Nakano, H.Ishiguro, and M.Asada. Multimodal joint attention through cross facilitative learning based on $\mu x$ principle. In *Proceedings of the 7th International Conference on Development and Learning*, pp. 226–231, 2008.

[9] G.E.Hinton. Training products of experts by minimizing contrastive divergence. *Neural Computation*, Vol. 14, No. 8, pp. 1771–1800, 2002.

[10] G.E.Hinton, S.Osindero, and Y.Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, Vol. 18, pp. 1527–1800, 2006.