

乳児母音発達における親の期待の誘導的効果： 発達シミュレーション及び音声模倣実験による検討

Guiding effect of caregiver's anticipation on vowel development: a computer simulation and a human subject experiment

石原 尚[†], 若狭 みゆき[†], 吉川 雄一郎[‡], 浅田 稔^{‡†}
Hisashi Ishihara, Miyuki Wakasa, Yuichiro Yoshikawa, Minoru Asada

[†]大阪大学大学院工学研究科, [‡]JST ERATO 浅田共創知能システムプロジェクト
Graduate School of Engineering, Osaka University

Asada Synergistic Intelligence Project, ERATO, JST

{hisashi.ishihara, miyuki.wakasa, asada}@ams.eng.osaka-u.ac.jp, yoshikawa@jeap.org

Abstract

The mechanism of caregiver's imitation is modeled by conjecturing caregiver's biases of interpreting infant's utterances as closer to prototypical vowels and precedent own ones. The computer simulation of caregiver-infant interaction of mutual imitation is conducted to investigate roles of these tendencies on infant vowel development. Simulation results indicate the latter bias guides infant's associative learning between his/her articulations and contingent replies by the caregiver, and the former bias stabilizes his/her learning. The latter bias was tested in an experiment with human subjects. These results imply guiding effect of caregiver's anticipation for infant development.

Keywords — Constructive model, Vowel development, Caregiver's bias in imitation

1. はじめに

乳児が親と共通の母音体系をいかにして獲得していくかという問いは、発達心理学の研究者の関心を集めてきた重要なテーマの一つである。初めは言語的特徴を有さない乳児の知覚と発声は、生後一年をかけて徐々に母国語に特化していく[1, 2, 3]。この親への適応の過程には、非常に多くの要因が絡み合っている。例えば、知覚と構音は各々独立に発達するのではなく、相互に依存し合っ

て発達している可能性が指摘されている[4, 5, 6]。また、親との社会的な関わり合いの重要性も多く、研究者が主張するところである[7, 8, 9, 10, 11]。このように、母音発達の仕組みの詳細を知るためには、乳児の知覚と構音の発達の個体内相互作用に加えて親との社会的相互作用も含めて考えることが求められる[12]。

このような複雑な相互作用の中で成立する乳児発達の過程を理解する上では計算論的な取り組みが有効であると考えられ[13, 14]、母音発達を説明する計算モデルを構築する試みも行われてきた。

しかし、それらの多くは知覚の発達[15, 16, 17]、もしくは構音の発達[18, 19, 20]を独立に扱っている。また、いくつかの研究では知覚と構音の間のマッピングの発達が取り扱われてはいるものの、親と乳児が音響的に同じバリエーションの声を生成可能であることが仮定されており[21, 22, 23]、乳児の生成可能音が物理的に親とは異なる[24]という母音獲得における本来の難しさが扱われていない。この問題を解く上で、親に模倣される経験が重要な役割を果たしていることを示唆する構成的研究もあるが[25, 26]、これらの研究で想定された親は単に音声の対応関係を教えるだけの存在として扱われるにとどまっている。

親は乳児を意図や信念を持つ存在として捉え、未熟な行為をその観点から理解しようとする傾向(maternal mind-mindedness[27])を有し、これが乳児の発達因となっている可能性が議論されている([28]に詳しい)。また、親が乳児を模倣しようとして関わる場面においても、親は乳児の行為を単に再現するだけでなく、乳児の内的状態を反映するように行為を強調するなど、自身の主観的な解釈を加えた上で模倣することが乳児の社会的能力の発達を導いている、との主張もなされている[29, 30]。しかし、親は乳児の発声をどのように解釈し、そのことがどのように親の模倣に影響し、その結果母音発達が誘導されるのか、その詳細な仕組みは明らかでない。

そこで本研究では、親と乳児の相互模倣やりとりを通じた母音発達を再現する計算機シミュレーションを行い、親の解釈のどのような歪み(バイアス)が模倣を通じて乳児の母音発達を導き得るのかを検討する。そして、この検討の結果誘導に寄与すると考えられたバイアスが実際に存在するかを検証する被験者実験を行い、シミュレーションの妥当性を確認する。

2. 相互模倣を通じた母音発達モデル

2.1 概要

本研究で想定している発達は、乳児が親の模倣を試み、親もそれに解釈を加えた上で模倣を返すという交互の模倣を通じて、乳児が親の母国語母音を獲得するというものである。まず、親と乳児の模倣及び乳児の学習の計算モデルを構築した。模倣モデルは大きく分けて、相手の音声の知覚の表象と自身の発話のための構音表象、そしてそれらを結合する知覚構音写像（以後単に写像と呼ぶ）の三つの要素で構成される。

この模倣モデルは親と乳児それぞれに用意されるが、以下の二つの点で異なる。一つは学習可能性である。親の写像はやりとりを通じて変化しないが、乳児の写像は親による模倣フィードバックを利用して逐次修正される。もう一つは、親の模倣過程における二つのバイアスの存在である。一つは、乳児の発話をより母国語母音らしい音として解釈するバイアスであり、これは母国語母音の典型例（/a//i//u//e//o/のような母音プロトタイプ）を中心とした知覚バイアスである知覚のマグネット効果[31]である。ただし、音声を構音する際にも自身の普段発声し慣れた母音へのバイアスが生じると考えられるため、これらを模倣におけるプロトタイプへのバイアスとしてまとめて扱い、以後知覚構音マグネットと呼ぶ。もう一つのバイアスは、乳児の応答をより精度のよい模倣として解釈するというものである。これは、相互模倣の最中に親は乳児に模倣されることを期待しており、その期待に沿うように乳児の応答を解釈するという仮定に基づくものである。以下ではこれを自己鏡映バイアスと呼ぶ。

提案相互模倣モデルの概念図を図1に示す。 t 回目のやりとりにおいて、乳児は構音運動 $\mathbf{a}'(t) \in \mathbb{R}^{N_a}$ により母音音声 $\mathbf{s}'(t) \in \mathbb{R}^{N_s}$ を発声する。親はそれを知覚した結果を写像 $f: \mathbb{R}^{N_s} \rightarrow \mathbb{R}^{N_a}$ を入力し、その出力結果 $\mathbf{a}(t) \in \mathbb{R}^{N_a}$ を構音運動とし、 $\mathbf{s}'(t)$ の模倣として音声 $\mathbf{s}(t) \in \mathbb{R}^{N_s}$ を発声する。乳児は親の模倣 $\mathbf{s}(t)$ を知覚し、自身の構音運動 $\mathbf{a}'(t)$ とそれに随伴した親の応答 $\mathbf{s}(t)$ の対応情報に基づいて写像 f' の最適パラメータを最尤推定し修正する。そして、修正した写像を用いて親の発声 $\mathbf{s}(t)$ を構音 $\mathbf{a}'(t+1)$ により模倣することを試みる。

2.2 模倣メカニズム

乳児の発話 $\mathbf{s}'(t)$ に対する親の模倣 $\mathbf{a}(t)$ は、

$$\mathbf{a}(t) = f(\mathbf{s}_b(t); \mathbf{p}_i, \lambda), \quad (1)$$

$$\mathbf{s}_b(t) = (1 - \eta)\mathbf{s}'(t) + \eta\mathbf{s}_a(t), \quad (2)$$

$$\mathbf{s}_a(t) = f^*(\mathbf{a}(t-1); \mathbf{p}_i^*), \quad (3)$$

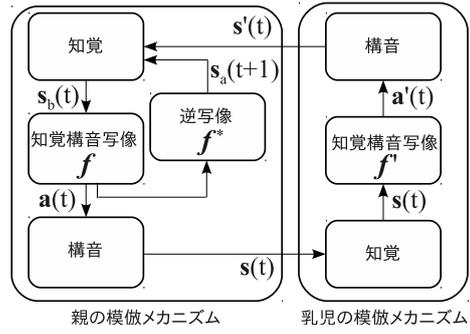


図1 親と乳児の模倣メカニズムにより構成される提案モデル

のようにモデル化した。ここで、 $\mathbf{p}_i \in \mathbb{R}^{N_a}$ は母音プロトタイプを発声するための構音運動であり（本研究ではこれをプロトタイプと呼ぶ）、 $\lambda (0 \leq \lambda \leq 1)$ は知覚構音マグネットの強さを決定するパラメータである。また、 $f^*: \mathbb{R}^{N_a} \rightarrow \mathbb{R}^{N_s}$ は写像 f とは反対に構音表象の入力を知覚表象に出力する写像（以下ではこれを逆写像と呼ぶ）であり、 $\mathbf{p}_i^* \in \mathbb{R}^{N_s}$ は、写像 f によりプロトタイプ \mathbf{p}_i へと写像される音声である。すなわち、乳児の発声の中で親が自身のプロトタイプと同じ音韻として知覚する音声であり、乳児はこの音声を発声ようになることが本研究での母音発達の目標である。そこで以後この音声を、乳児が獲得することを期待されている母音、すなわち期待母音と呼ぶ。

親自身の構音 $\mathbf{a}(t)$ は逆写像 f^* に入力され、乳児の応答に対する予期 $\mathbf{s}_a(t+1)$ へと変換される。すなわち、親は自身の逆写像を乳児の模倣をシミュレートするのに使用する。この乳児の発声の予期は実際の乳児の発声と混合率 $\eta (0 < \eta \leq 1)$ で混合される。これが自己鏡映バイアスであり、 η がその強さを決定するパラメータである。このバイアスにより歪んだ知覚 $\mathbf{s}_b(t)$ は写像 f で構音 $\mathbf{a}(t)$ に変換されるが、その出力はこの写像の過程でプロトタイプ $\mathbf{p}_i (i = 1, \dots, M)$ に収斂度 λ で引き寄せられる。この収斂が知覚構音マグネットである。このように、親の模倣モデルにおける知覚構音マグネットと自己鏡映バイアスの強さは、パラメータ λ と η の設定により調整が可能である。

2.3 知覚構音写像

写像及び逆写像は非線形関数近似器の一つである正規化ガウス関数ネットワーク（Normalized Gaussian network: NGnet[32, 33]）でモデル化した。NGnetは M 個の線形写像の組み合わせで非線形関数を近似するものであり、線形写像の組み合わせの重みを決定するための M 個のガウス関数 $g_i (i = 1, \dots, M)$ が入力空間に配置されている。ガウス関

数と線形写像は1対1の対応関係を持っており、入力と各々のガウス関数の中心との距離によって、そのガウス関数に対応する線形写像の重ね合わせる際の重みの大きさが決定される。すなわち、親の写像 f のNGnetの場合、入力である知覚 $\mathbf{s}_b(t)$ は、

$$\mathbf{a}(t) = f(\mathbf{s}_b(t); \mathbf{p}_i, \lambda) \quad (4)$$

$$= \sum_{i=1}^M \frac{g_i(\mathbf{s}_b(t); \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}{\sum_{j=1}^M g_j(\mathbf{s}_b(t); \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \tilde{\mathbf{W}}_i \tilde{\mathbf{s}}_b(t), \quad (5)$$

のようにして構音 $\mathbf{a}(t)$ に変換される。ここで、 $\boldsymbol{\mu}_i \in \mathbb{R}^{N_s}$ と $\boldsymbol{\Sigma}_i \in \mathbb{R}^{N_s \times N_s}$ はそれぞれ i 番目のガウス関数の中心と分散共分散行列である。また、 λ は線形回帰行列 $\tilde{\mathbf{W}}_i \in \mathbb{R}^{N_a \times (N_s+1)}$ の固有値を $(1-\lambda)$ に決定するパラメータである。 $\tilde{\mathbf{s}}_b(t) \equiv [\mathbf{s}_b^T(t), 1]^T \in \mathbb{R}^{N_s+1}$ は $\mathbf{s}_b(t)$ の拡大行列である。

図2は、 $M=1$ のNGnetにより知覚構音マグネットがどのようにモデル化され、また λ の設定によってバイアスの強さがどのように調節されるかを示している。横軸が1次元の入力空間であり、いくつかの入力がNGnetのガウス関数の中心ベクトル周りに正規分布状に与えられ、それぞれの入力が線形回帰行列 $\tilde{\mathbf{W}}_1$ で1次元の出力空間に写像されるとする。この時、出力データの分布は線形回帰行列の固有値（この場合は線形写像の傾き）の大きさによって異なった結果となる。すなわち、固有値 $(1-\lambda)$ が小さいほど、出力データはガウス関数中心の像 $\tilde{\mathbf{W}}_1 \tilde{\boldsymbol{\mu}}_1$ を中心に小さく収束する。ここで、ガウス関数中心の像をプロトタイプとして、またガウス関数中心を期待母音としてみなす、すなわち、 $\mathbf{p}_i \equiv \tilde{\mathbf{W}}_i \tilde{\boldsymbol{\mu}}_i$ 、 $\mathbf{p}_i^* \equiv \boldsymbol{\mu}_i$ と定義すると、出力分布の収束を知覚構音マグネットとしてみなすことができる。

親の逆写像もNGnetでモデル化するが、写像とは入出力関係が逆になっている。すなわち、写像は \mathbf{p}_i^* を \mathbf{p}_i に写すが、逆写像は \mathbf{p}_i を \mathbf{p}_i^* に写す。

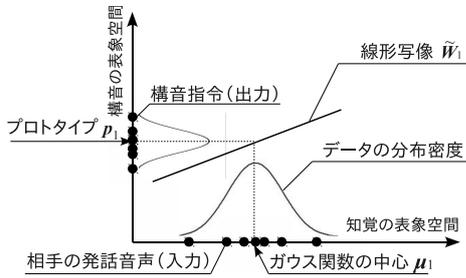


図2 $M=1$ のNGnetによる写像における知覚構音マグネットの表現

2.4 乳児の学習メカニズム

乳児の写像 $f|_t$ もNGnetでモデル化する。乳児はやりとりの T ステップ目において、直前 n ステ

ップ分の自身の構音と親の応答の組の履歴に基づいて写像のパラメータを修正する。すなわち乳児が行うことは、同時分布 $(\mathbf{a}'(t), \mathbf{s}(t)|t = T - n + 1, \dots, T)$ をうまく表現するようなNGnetのパラメータ $\{\boldsymbol{\mu}'_i(t), \boldsymbol{\Sigma}'_i(t), \mathbf{W}'_i(t) | i = 1, \dots, M\}$ を推定することである。この推定には確率モデルのパラメータの最尤推定法の一つであるEMアルゴリズム[32, 34]を用いる。

2.5 シミュレーション設定

2.5.1 状況設定

親は乳児を常に模倣するが、一方の乳児は基本的には親の模倣を試みるものの、何度かは模倣以外の発話をする。具体的には、やりとり開始後 n ステップが過ぎるまでは初期プロトタイプの周辺の音をランダムに選び発話する。これは、学習前のある程度網羅的にデータを集めておくことで学習が局所解に落ち込むのを避けるためである。また学習が始まった後も5ステップに一度、学習中のプロトタイプの一つをランダムに選び、発話するようにした。これは互いに模倣しあうだけでは同じ発話を繰り返す状態にすぐに収束してしまい、学習が進まなくなってしまうためである。やりとりの総ステップ数は $T_L = 5000$ とし、 $n = 500$ とした。

母音音声の次元は $N_s = 2$ とした。これは実際の母音が第一フォルマント及び第二フォルマントという2次元母音特徴量でよく特徴づけられることが知られているためである。親と乳児の発声可能な母音のバリエーション（発話可能領域）は図3のように、実際の親と乳児の発話可能領域の報告結果[3, 35]に基づき、2次元母音特徴量空間で異なる領域を占めるように設定した。

シミュレーションの簡単化のため、構音 a （あるいは a' ）とその結果生成される音声 s （あるいは s' ）は同じ2次元ベクトルで表されることを仮定した。つまり、乳児が自分の構音とその結果生成される音の関係をどのように学んでいるのかという問題は今回は扱わない。また、乳児の写像 $f|_t$ の初期パラメータは基本的にはランダムに設定するが、ガウス関数の中心が親の母音プロトタイプのある程度近辺に収まった状態になるようにした。すなわち、乳児は親と自身の声の対応関係は知らないが、親の母音プロトタイプに関する知識はある程度有していると仮定した。

2.6 親の模倣モデルの設定

親の母音プロトタイプ $\mathbf{p}_i (i = 1, \dots, M)$ の配置とその数 M は、親の母語が日本語であることを想定して設定した。すなわち、図3のように親の

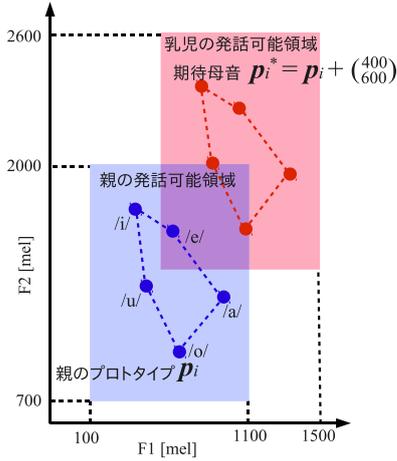


図3 実際の母子の領域の違いを考慮した発話可能領域の設定

プロトタイプの数 $M = 5$ とし、その配置も日本語五母音のものに従った。具体的には、 $\mathbf{p}_1 = [700, 1200]^T$, $\mathbf{p}_2 = [400, 1700]^T$, $\mathbf{p}_3 = [400, 1300]^T$, $\mathbf{p}_4 = [600, 1500]^T$, $\mathbf{p}_5 = [500, 1000]^T$ とし、期待母音 \mathbf{p}_i^* は、

$$\mathbf{p}_i^* = \mathbf{p}_i + \begin{pmatrix} 400 \\ 600 \end{pmatrix} \quad (i = 1, \dots, M). \quad (6)$$

とした。

親の写像 f のパラメータは、上記の設定を満たすように決定した。すなわち、

$$\boldsymbol{\mu}_i = \mathbf{p}_i^* \quad (i = 1, \dots, M), \quad (7)$$

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} 3600 & 0 \\ 0 & 3600 \end{pmatrix} \quad (i = 1, \dots, M), \quad (8)$$

$$\tilde{\mathbf{W}}_i = ((1 - \lambda)\mathbf{I}, \mathbf{p}_i - (1 - \lambda)\boldsymbol{\mu}_i) \quad (i = 1, \dots, M, 0.0 \leq \lambda < 1.0). \quad (9)$$

とした。親の逆写像 f' も同様に、

$$\boldsymbol{\mu}_i^* = \mathbf{p}_i \quad (i = 1, \dots, M), \quad (10)$$

$$\boldsymbol{\Sigma}_i^* = \begin{pmatrix} 3600 & 0 \\ 0 & 3600 \end{pmatrix} \quad (i = 1, \dots, M), \quad (11)$$

$$\tilde{\mathbf{W}}_i^* = (\mathbf{I}, \mathbf{p}_i^* - \mathbf{p}_i) \quad (i = 1, \dots, M). \quad (12)$$

とした。

2.7 結果

親の模倣に含まれる二つのバイアスが乳児の母音発達にどのように寄与するのかを調べるため、自己鏡映バイアスと知覚構音マグネットの強さを決めるパラメータである η と λ の値を様々に変えて相互模倣シミュレーションを行い、学習の結果とバイアスの強さの関係を調べた。

2.8 乳児発話の母語への特化

親が両バイアスを有している場合に($\eta = 0.5, \lambda = 0.6$)、乳児の発話とプロトタイプがどのように変遷したかを検討するため、母音特徴量空間における(1)乳児の各発声から最も近い期待母音までの距離、(2)親の各発声から最も近い親自身のプロトタイプまでの距離、(3)各期待母音から最も近い乳児のプロトタイプまでの平均距離、をステップごとに計算し、その変遷をみた。図4のグラフは上記三つの距離の500ステップごとの移動平均を表している。(1)及び(3)のグラフはステップを経るにつれた減少の傾向を示しており、これは、このバイアスの条件においては乳児の発声とプロトタイプが期待母音に一致するように、すなわち発話が母国語に特化するよう遷移したことを示している。

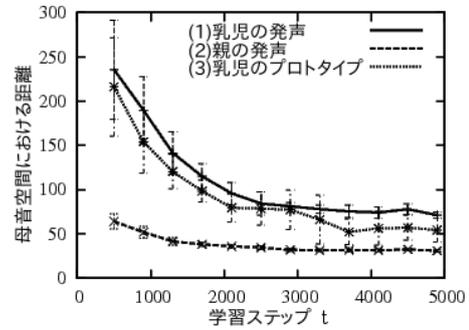


図4 親が両バイアスを有する場合における乳児発話及び乳児プロトタイプの各々からの期待母音までの距離の変遷(5回のシミュレーションの結果を平均)

2.9 バイアスによる結果の変化

図5は二つのバイアスの有無の組み合わせの各条件下における母子の発話やりとりと学習結果の例を示しており、各図における条件の違いは、

- (a) 両バイアスとも有り($\eta = 0.5, \lambda = 0.6$),
- (b) 自己鏡映バイアスのみ有り($\eta = 0.5, \lambda = 0.0$),
- (c) 知覚構音マグネットのみ有り($\eta = 0.0, \lambda = 0.6$),
- (d) 両バイアスとも無し($\eta = 0.0, \lambda = 0.0$),

というものである。これらの図の各々は、学習終盤1000ステップにおける発話やりとりの様子を示しており、母音空間において赤の点(青の点)の各々が各ステップの乳児の発話(親の発話)を表している。赤色(青色)の五角形の頂点が期待母音(親のプロトタイプ)を、黒の点が学習終了時の乳児のプロトタイプを表している。これらの図は、親のバイアスが乳児の発話の変遷に強く影響を与えていることを示している。例えば、条件(a)と(c)で

は、他の(b)と(d)の条件下の結果と比較して、乳児の発声はより小さなクラスターを形成している様子が伺える。

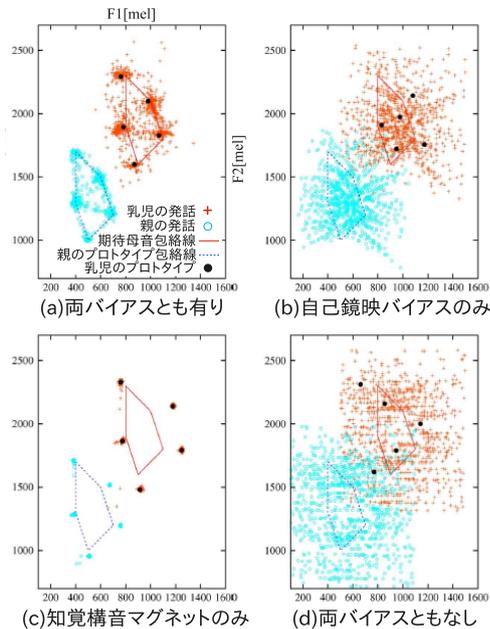


図5 バイアスの有無の異なる各条件下における学習終盤での母子の発話やりとりと学習結果の例

2.10 考察

2.11 知覚構音マグネットの収斂効果

親が知覚構音マグネットを有する図5(a)及び(c)の条件下において、発話が複数のクラスターを形成する傾向が見られた。実際の乳児の発達においても、発話が複数のクラスターに分離していく過程を経ることが報告されており[36, 3], この傾向に親の知覚構音マグネットが影響している可能性が示唆される。

この傾向をさらに検討するため、知覚構音マグネットの強さを様々に変えた複数の条件下でシミュレーションを行った。図6は、知覚構音マグネットの強さ λ (横軸)と終盤の1000ステップにおける乳児自身のプロトタイプへの発話の収斂度(縦軸)の関係を見たものである。この図は、知覚構音マグネットが強いほど母音クラスターがより収斂する傾向にあることを示している。

これはおそらく知覚構音マグネットの働きによって親の発話クラスターが乳児の発話クラスターと比較してより収斂したものになるため、それを模倣した乳児の発話が以前よりも収斂したものになるためだと考えられる。ただし、図5(c)の例で分かるように、乳児の発話クラスターの収斂の中心が常に期待母音となるわけではない。なぜなら、親の発話クラスターが一定の位置を中心に収斂していくだ

けでは、乳児は学習前にランダムに配置されたガウス関数の中心(プロトタイプ)の位置はそのままに、線形写像関数のパラメータのみ修正させれば済むからである。すなわち、親の知覚構音マグネットは乳児の発話を収斂させ、プロトタイプをその位置に安定化させる働きをするが、期待母音に誘導する効果は持たないと考えられる。

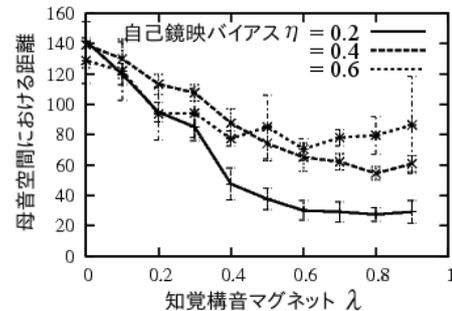


図6 知覚構音マグネットの強さと乳児発話収斂度の関係(5回のシミュレーション結果の平均)

2.12 自己鏡映バイアスの誘導効果

ここであらためて図5の結果を見てみると、条件(c)と比較して条件(a)は発話クラスターの中心及びプロトタイプが期待母音とよく一致しているようである。すなわち、親の自己鏡映バイアスが乳児の発話を期待母音へと誘導している可能性が示唆される。ただし、この効果は条件(a)と(b)の結果を比較すると明らかなように、知覚構音マグネットの影響も受けていると考えられる。そこで両バイアスの強さを様々に変えてシミュレーションを行い、自己鏡映バイアスと知覚構音マグネットの強さのバランスと誘導効果の関係について検討した。

図7は、自己鏡映バイアス(縦軸)と知覚構音マグネット(横軸)の強さの組み合わせと、各組み合わせの条件下における学習終了時の乳児のプロトタイプと期待母音の一致の度合い(各期待母音から最も近いプロトタイプまでの距離の平均値)を示している。赤色が濃い条件ほど、一致の度合いが高かったことを表しており、これは両バイアスの強さのバランスがとれている場合に誘導効果が発揮されることを示唆している。

これは以下のように説明されうる。まず、どちらのバイアスも無ければ、親は単に乳児の発声を正確に再現するだけなので乳児の発話パターンは初期の状態から変化しない。しかし知覚構音マグネットがあると、親の発話は収斂し始め、それにつられて乳児の発話もクラスターを形成し始める。ただし、前項で述べたようにその中心位置は初期位置で安定であり、そこから動かない。ところが、

ここで自己鏡映バイアスがあると、親は乳児の発声を違う音として解釈して模倣するため、親の発話の分布が変動する。そのため、乳児はマップの線形写像だけでなくガウス関数のパラメータも修正することを迫られ、乳児のプロトタイプは別の場所に移動することになる。移動の結果プロトタイプが期待母音に近づくと、乳児の模倣の精度も上がっていくために自己鏡映バイアスの影響が弱まり、親の発話分布の変動も徐々に少なくなっていく。すなわち自己鏡映バイアスは、乳児のプロトタイプと期待母音の一致度に応じて乳児のプロトタイプを変動させる効果を持つと考えられる。

この効果を知覚構音マグネットの持つ安定化効果が上回った場合、プロトタイプが十分に期待母音まで移動する前にそこで学習が収束してしまう。しかし逆に下回った場合は安定化させることができず、プロトタイプはふらふらと移動を続けてしまう。安定化効果と変動効果のバランスがとれたときに、プロトタイプは期待母音まで誘導され、そこで安定化するのだと考えられる。

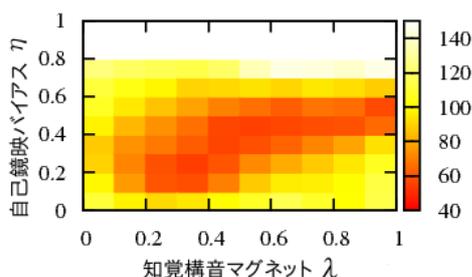


図7 バイアスの様々な組み合わせ条件下における乳児プロトタイプと期待母音の一致度の比較

3. 音声模倣実験

前節のシミュレーションでは、相手の発話をより模倣らしく解釈するという自己鏡映バイアスが乳児の母音発達の誘導に寄与する可能性が示唆された。しかしながら、このバイアスの存在を直接的に支持する報告はない。このバイアスは、相手が自分を模倣してくるという期待により生じると考えられる。そこで、模倣されることへの期待の程度を統制した被験者に合成音声を連続的に模倣させ、そのような模倣やりとりの中で自己鏡映バイアスが生じるかを検証する。本実験ではその指標として、被験者自身の発声の変化幅に注目した。なぜなら、期待の方に歪められて音声が発話される自己鏡映バイアスが存在するならば、被験者は合成音声をその直前に自身が発声した音声により似た音声として解釈するため、その解釈に基づいた模倣も、その直前に被験者自身が発した音声により似たものとなる、すなわち発声の変化幅が小さくなると考えられるためである。

3.1 方法

3.1.1 被験者

日本語を母語とする19歳から29歳までの健聴者30名(男性女性各15名ずつ)が実験に参加した。被験者は音声模倣調査実験への参加者として外部機関により集められ、音声模倣の仕方を調査するという以外に実験の詳細は知らされなかった。被験者は、実験中に模倣されることへの期待を強く持つような教示を受ける実験群と、期待を持たないような教示を受ける統制群に分けられた。実験群は男性8名女性7名の計15名で、統制群は男性7名女性8名の計15名であった。

模倣されることへの期待の程度に関して群間で差をつけるため、各群の被験者への実験説明に差をつけた。具体的には、まずどちらの群に対しても、「ヘッドフォンから流れる音声を聞き、その音声を模倣していただきます。マイクに向かってできるだけ似るように(第三者が聞いて同じ音声に聞こえるように)その音声の真似をしてください」と教示したが、実験群に対してのみこれに続けて、「あまり上手くはありませんが、やりとりの中でコンピュータがあなたの声をできる限り真似しようと試みるので、あなたもできるだけ似るように音声の真似をしてください」と教示した。

3.1.2 刺激

実験で被験者に提示する音声は、多様な音声処理が可能なソフトウェアであるPraat[37]を用いて人の声を変成させたものを使用した。具体的には、人のサンプル音声から音源成分を抽出し、この音源に対して、指定したフォルマント周波数を持つように設計したフィルタをかけることで、様々な母音音声を合成した。フォルマントとは音声の共鳴周波数のことであり、母音はそれらのうち最も周波数の低い二つの共鳴周波数(第一、第二フォルマント、F1、F2とも表記される)でよく特徴づけられる[35]。またサンプル音声として、成人男性が/a/を発声した時の音声の基本周波数を変化させたものを用いた。基本周波数は音の高さを決定づける成分であり、今回の実験では、男女の声の高さの平均値に近い170Hzに設定した。音声の変成には多少の時間を要するため、実験に先立ち、F1-F2空間上で日本人の成人男女ともに発声可能であると思われる母音領域内(各10名の日本人成人男女の日本語プロトタイプの発声をサンプリングした事前調査の結果に基づく)の100Hz毎の格子点となる148種の音声を予め作成しておいた。この音声群を以後音声セットと呼ぶ。

3.1.3 手順

被験者は防音室内で実験概要説明書以外何も置かれていない机に就き、前項で示した様な教示を受けた後、模倣の練習を数回行った上で実験を受けた。実験が開始されると1秒間の合成音声で6秒おきに次々と提示され、被験者は音声を聞き終わってから次の音声の流れ始めるまでの間に聞いた音声を模倣することを求められた。合成音声は30音提示されると音声の提示は自動的に一時停止し、ここまでを1セットとして1分の休憩を挟んで3セット計90音が被験者に提示された。合成音声は基本的にはどちらの群に対してもランダムな順で提示されるが、どちらの群についても教示内容と大きく異なる状況にならないよう、提示する音声は直前に提示した音声からF1-F2空間上で300-500Hzの変化幅に収まるように音声セットからランダムに選択された。合成音声器はこのようにランダムで発声するが、模倣されることへの期待を持った実験群の被験者は統制群と比較してより模倣されたと感じる事が予想される。

3.1.4 計測

被験者はマイク付きヘッドフォン (ELECOM MS-HS58V) を装着し、音声刺激の提示と被験者の発声の録音にはこれを用いた。また条件統制の確認のために実験終了後、各セットにおいてどのくらいうまく模倣できたか (質問1)、相手に模倣の意図を感じていたか (質問2)、模倣されたと感じたか (質問3)、コミュニケーションを取り合ったと感じたか (質問4)、を5段階で問うアンケートに回答させた。

3.1.5 解析

音声学の知識を持つ実験者が録音された被験者の各発声についてフォルマント成分がよく表れている区間を切り出し、その区間におけるF1及びF2成分の各々の平均値をPraatを用いて求めた。そして、各セットにおける被験者自身の発声の、F1-F2空間上での変化幅の平均値を計算した。

3.2 結果

3.2.1 アンケート

上記4つの質問に対する被験者の各セットごとの回答の結果を図8に示す。これらのグラフは、各群の被験者の各セットにおけるアンケート得点 (縦軸) の分布の中央値、下側四分位、上側四分位、最大値、最小値、及び外れ値を示しており、グラフ

において、アンケート得点が高いほど問われた内容の実感の程度が高かったことを示している。

模倣がうまくできたと思った程度を問う質問1、及びコミュニケーションをとったと感じた程度を問う質問4に関しては、全セットに関して群間にそれほど大きな違いは見られなかった (すべてのセットで $p > 0.1$)。また、模倣がうまくできなかったと感じた被験者が少なからずいたようであった。

一方で、相手に模倣の意図があるように思っていた程度 (すなわち、模倣されることを期待していた程度) を問う質問2、及び模倣されていると思っていた程度を問う質問3に関しては、想定された群間の差がある程度みられるようである。ただし、本実験の独立成分である期待の程度についての被験者の内省報告である質問2に対し、第一セットで実験群であるにも関わらず、期待をほとんどしていなかった被験者もあり、有意差はなかった ($p > 0.1$)。また、全ての質問に関して得点の分散は比較的大きいものの、セットを経るにつれ得点の上昇する傾向が見られた。

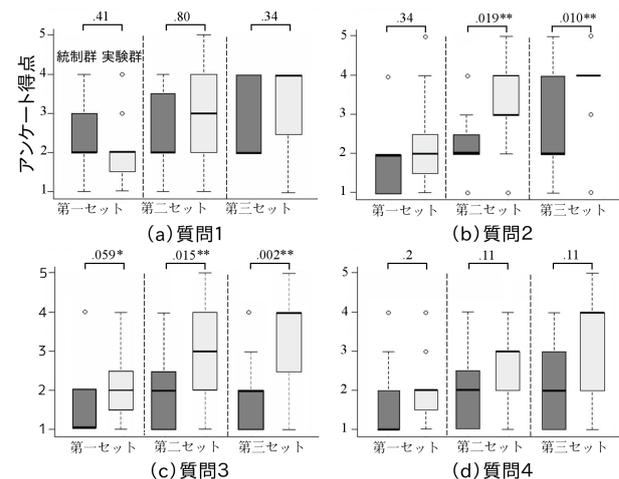


図8 各質問に対するアンケート得点の群間比較 (破線の両端は分布の最大値及び最小値を、ボックスの両端は上側及び下側四分位を、太く強調された横棒が中央値の値を示している。各グラフ上部の値は群間の分布の差のMann-Whitney U testによる検定結果の p 値である。 $p < 0.1$ の場合は数値に*を、 $p < 0.05$ の場合は**を付した。)

3.2.2 発声変化幅

アンケートの集計結果を受け、発声の変化幅の群間比較に先立って、条件統制のために以下のようなデータ処理を行った。

本実験では模倣されることを期待するような教示を行った実験群は実際に期待の程度が高く、また反対にそのような教示をしなかった統制群は期待の程度が低いことが仮定されていた。しかし、

期待の程度を問うアンケートの質問2に対して、実験群であるにも関わらず模倣されることをほぼ期待していなかった者や、統制群であるにも関わらず模倣されることを過剰に期待していた者がいた。そこで、このような被験者は上記のような仮定を満たさなかった事例として解析対象からはずした。具体的には、アンケートの質問2に対して、1点をつけた実験群の被験者と5点をつけた統制群の被験者を、そのように回答したセットの解析対象からはずした。結果的に第一セットで実験群4名、第二セットで実験群1名、第三セットで実験群1名と統制群1名を解析対象から除外した。この結果、解析の対象とする被験者の質問2に対する回答の得点は図9に示すように群間にさらに差があった（全てのセットにおいて $p < 0.05$ ）。

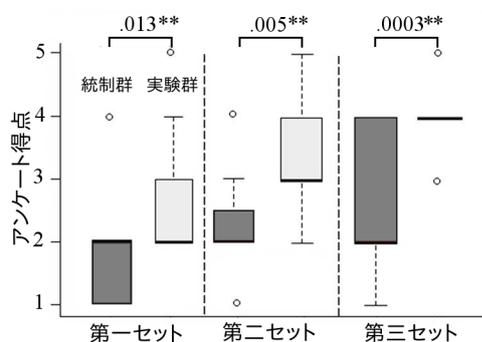


図9 模倣されることに対する期待の程度を問う質問（質問2）に対する回答結果の群間比較

また、被験者には合成音声をできるだけ似るように模倣するように教示したが、うまく模倣できなかった被験者もいるようであった。そこで、各セットにおいて模倣をうまくできなかったと申告した被験者、すなわちアンケートの質問1に対して1点をつけた両群の被験者を、そのように回答したセットの解析対象からはずした。結果的に第一セットでさらに実験群2名、統制群3名、第二セットでさらに実験群2名、統制群1名、第三セットでさらに実験群1名を解析対象から除外した。

このようなデータの除外を行った上で、各セットにおける被験者の平均発声変化幅の分布の中央値を群間で比較した結果のグラフを図10に示す。第一及び第二セットでは我々の予想したように、発声変化幅は実験群の方が小さい傾向にあった（第一セットでは $p < 0.1$ 、第二セットでは $p < 0.05$ ）。これは、統制群に比べて模倣される期待の程度の高かった実験群の被験者が、提示された音声を直前の被験者自身の発声とより似た音として知覚した結果であると考えられ、自己鏡映バイアスの存在を支持するものであるといえる。ただし、第三セットでは群間に差があるとは言えなかった（ $p > 0.1$ ）。

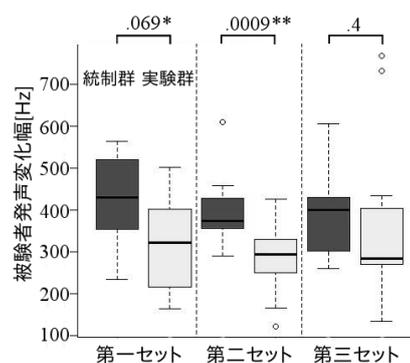


図10 被験者発声変化幅の群間比較（検定はMann-Whitney U testによる）

3.3 考察

3.3.1 教示による期待の効果

今回の実験では、実験群の被験者のみに合成音声器に模倣の意図があるように思わせる教示を実験前に行うことで、模倣されることへの期待の程度に関して群間で差がつくようにした。この効果はアンケートの質問2の得点結果に現れており、期待の程度は群間によく差がついていたと言えるであろう。今回の実験での独立変数はこの期待の程度であり、発声変化幅の群間の差は期待の程度の差により生じた、すなわち自己鏡映バイアスは期待によって生じたと考えられる。

期待の程度はセットを経るごとに両群ともに高まっていく傾向が見られた。模倣されたように思った程度を問う質問3の得点もセット毎に高まっている傾向にあることから、合成音声の模倣意図の教示の有無に関わらず、交互発話やりとりをする中で実際に模倣されている、と感じた経験を経るにつれて模倣されることへの期待も徐々に高まっていったと考えられる。統制群の発声変化幅がセットを経るにしたがって小さくなっていることも期待の程度と自己鏡映バイアスの関連を支持する結果である。ただし、実験の進行に伴う結果の変化を説明するためには、実験への馴化や集中力低下の影響など、今回の実験では測定しなかった要素も含めてさらに解析する必要があるだろう。

3.3.2 期待の誘発因

今回の実験設定では条件統制のために相手の模倣の意図に対する期待は実験前の教示により持たせた。しかしながら、親が乳児と模倣やりとりをするときには当然このような教示は与えられない。乳児を意図を持った存在であるとみなす傾向が親の mind-mindedness という性質の一つとして考えられており、模倣の場面で乳児に模倣の意図があるように捉え、模倣を期待することは mind-mindedness

の一例であると考えられる。このような傾向はどのような要因から起こっているのだろうか。今回教示で抱かせた期待が、親が自然に乳児に抱く期待と本質的に同じものであるかを議論するためには、期待を誘発する要因について考えていく必要がある。

アンケートの結果から、ヘッドフォンから聞こえてくる合成音声であっても、交互発話を続けるに従って期待する程度（質問2）及びコミュニケーションをとっていると感じる程度（質問3）が上昇していく傾向が見られた。このことは、相手とのやり取りの経験が少なからず相手に意図を帰属する程度に影響していることを示していると考えられる。ただし、今回は上昇したとはいえ、その程度はそれほど高いものではなく、また個人差も大きかった。本実験では被験者はヘッドフォンから聞こえる合成音声と対話をしたが、その時に何と向かい合っているのか、ということは期待の程度を決定するとても大きな要因であるように思われる。例えば、乳児が発話をするCGがディスプレイに表示されていたり、乳児の外見を持つロボットが発話するなどしていたら、被験者はより強く相手の存在を感じ、意図の帰属も強くなるであろう。これについては今後の実験で検討していきたい。

3.3.3 自己鏡映バイアスの生起因

今回の実験結果は、相手の行動を自分の期待していたように解釈する、という傾向を示唆するものであるが、このバイアスはなぜ生じるのだろうか。関連する知見として、ミラーニューロンシステムが挙げられる。これは、他者の行為の観察した時と自身がその行為を行った時のどちらの場合にも賦活する脳神経系[38, 39, 40]であり、他者が行ったことを自分が行ったように感じる、という人の模倣や共感能力の神経基盤として注目されている[41, 42, 43]。このような、ある意味で自他を区別しないような認知基盤があるとすれば、他者を自己に投影する共感の例とは逆に、自身が行ったことを他者が行ったように感じる、というような自己を他者に投影することも起こっていると考えられる。これは無理のない推測であろう。このような投影は、知覚を歪めるように作用するであろうし、またこのような投影があるからこそ、人は他者に模倣を期待することも考えられる。すなわち、相手に模倣を期待することも、模倣されたように解釈することも、この投影の影響であるとの推察ができる。これは推測の域をでない仮説であるが、これを検証する取り組みは前項で述べた内容と関連が深く、なぜ母親はそもそも乳児に期待するのか、という問いに迫る興味深い課題である。

4. おわりに

本研究では、親が乳児の発話をより母国語母音らしい音として（知覚構音マグネット）、またより精度のよい模倣として（自己鏡映バイアス）解釈するというバイアスを持つと仮定し、そのようなバイアスを考慮した模倣メカニズムのモデルを提案した。そして、この模倣モデルを用いて母子間相互模倣の計算機シミュレーションを行ったところ、それらのバイアスがバランスを取り合って働くことが乳児の母音発達を促す結果が得られた。この結果から、自己鏡映バイアスは乳児の知覚構音写像をよりよい方へ誘導する働きを持ち、知覚構音マグネットは乳児の学習を収束させる働きを持つことが示唆された。さらに、シミュレーションで仮定した自己鏡映バイアスが存在するかを検証する実験を行った。被験者が音声合成器に模倣を期待する程度を統制し、交互に模倣させあつたところ、より高い期待を持つように統制した被験者群ほど、自身の直前の発声とより似たように発声する傾向が見られた。この結果は、相手の発声をより模倣らしく知覚する自己鏡映バイアスの存在を支持するものである。

乳児の心的側面に注目するmind-mindednessと呼ばれる親の傾向により、親は乳児の発達状況に比べてより高い発達程度の心的帰属をすることが乳児の発達を促す可能性が議論されている（[44]に詳しい）。模倣場面でも、親が乳児の内面状態に注目し、誇張して模倣することが乳児の発達を促すという主張もなされている[29, 30]。今回の結果は、上記のような親の関与の仕方の蓋然性を、音声模倣を例にとり、構成的に示したものと考えられる。すなわち、親が乳児の模倣の意図を読み取る傾向が模倣の期待を親に生じさせ、それが乳児の発声を歪めて知覚させ（自己鏡映バイアス）、その歪みを含んだ模倣が乳児にフィードバックされることが乳児の母音発達のための知覚と構音の連合学習において誘導的に働く、という発達の仕組みの存在を示唆するものである。

しかしながら、このような仕組みがどの程度実際の乳児の発達に寄与しているかは明らかでない。親による模倣が乳児の母音発達に重要であるかどうかの議論は観察研究においてもなされており[45]、親の模倣によらない母音発達の可能性を示唆する報告もある[7]。今回の結果は、親の模倣による母音発達という説を支持するものの、この議論に決着をつけるものではなく、今後はバイアスによる誘導が実際に乳児にも起こっているのかを検証する必要がある。

謝辞

本研究は一部に特別研究員奨励費の助成を受けた。

参考文献

- [1] J. F. Werker and R. C. Tees. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, Vol. 25, pp. 121–133, 2002.
- [2] B. Cheour, R. Ceponiene, A. Lehtokoski, A. Luuk, J. Allik, K. Alho, and R. Näätänen. Development of language-specific phoneme representation in the infant brain. *Nat. Neur.*, Vol. 1, pp. 351–353, 1998.
- [3] P. K. Kuhl and A. N. Meltzoff. Infant vocalizations in response to speech: Vocal imitation and developmental change. *J. of Acoustic Society of America*, Vol. 100, pp. 2415–2438, 1996.
- [4] M. M. Vihman and S. Nakai. Experimental evidence for an effect of vocal experience on infant speech perception. *Proc. of the 15th Int. Cong. of Phonetic Science*, pp. 1017–1020, 2003.
- [5] F. J. K. v. Beinum, C. J. Clement, and I. v. d. Dikkenberg-Pot. Babbling and the lack of auditory speech perception: a matter of coordination? *Developmental Science*, Vol. 4:1, pp. 61–70, 2001.
- [6] D. K. Oller and R. E. Eilers. The role of audition in infant babbling. *Child Development*, Vol. 59, pp. 441–449, 1988.
- [7] M. H. Goldstein and J. A. Schwade. Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, Vol. 19:5, pp. 515–523, 2008.
- [8] J. F. Werker, F. Pons, C. Dietrich, S. Kajikawa, L. Fais, and S. Amano. Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, Vol. 103, pp. 147–162, 2007.
- [9] H.-M. Liu, P. K. Kuhl, and F.-M. Tsao. An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, Vol. 6:3, pp. F1–F10, 2003.
- [10] N. Masataka. Early ontogeny of vocal behavior of Japanese infants in response to maternal speech. *Child Development*, Vol. 63, pp. 1177–1185, 1992.
- [11] K. Bloom, A. Russell, and K. Wassenberg. Turn taking affects the quality of infant vocalizations. *Journal of Child Language*, Vol. 14, pp. 211–227, 1987.
- [12] P. K. Kuhl, B. T. Conboy, S. Coffey-Corina, D. Padden, M. Rivera-Gaxiola, and T. Nelson. Learning as a pathway to language: new data and native language magnet theory expanded. *Philosophical Transactions of the Royal Society B*, Vol. 363, pp. 979–1000, 2008.
- [13] 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. 1:1, pp. 12–34, 2009.
- [14] J. Weng. Developmental robotics: theory and experiments. *International Journal of Humanoid Robotics*, Vol. 1:2, pp. 199–236, 2004.
- [15] B. M. Lake, G. K. Vallabha, and J. L. McClelland. Modeling unsupervised perceptual category learning. *Aut. Ment. Dev.*, Vol. 1:1, pp. 35–43, 2009.
- [16] B. McMurray, R. N. Aslin, and J. C. Tascano. Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science*, Vol. 12:3, pp. 369–378, 2009.
- [17] G. K. Vallabha, J. L. McClelland, F. Pons, J. F. Werker, and S. Amano. Unsupervised learning of vowel categories from infant-directed speech. *Proc. of National Academy of Sciences USA*, Vol. 104:33, pp. 13273–13278, 2007.
- [18] H. Kanda, T. Ogata, K. Komatani, and H. G. Okuno. Segmenting acoustic signal with articulatory movement using recurrent neural network for phoneme acquisition. *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 1712–1717, 2008.
- [19] F. H. Guenther and J. S. Perkell. A neural model of speech production and its application to studies of the role of auditory feedback in speech. In *Speech Motor Control in Normal and Disordered Speech*, pp. 29–49, 2004.
- [20] G. Westermann and E. R. Miranda. A new model of sensorimotor coupling in the development of speech. *Brain and Language*, Vol. 89:2, pp. 393–400, 2004.
- [21] P.-Y. Oudeyer. The self-organization of speech sounds. *J. of Theoretical Biology*, Vol. 233:3, pp. 435–449, 2005.
- [22] I. S. Howard and M. A. Huckvale. Learning to control an articulatory synthesizer by imitating real speech. *ZAS Papers in Linguistics*, Vol. 40, pp. 63–78, 2005.
- [23] B. de Boer. Self organization in vowel systems. *Journal of Phonetics*, Vol. 28:4, pp. 441–465, 2000.
- [24] H. K. Vorperian and R. D. Kent. Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *J. of Speech, Language, and Hearing Research*, Vol. 50, pp. 1510–1545, 2007.
- [25] K. Miura, M. Asada, and Y. Yoshikawa. Unconscious anchoring in maternal imitation that helps finding the correspondence of caregiver's vowel categories. *Advanced Robotics*, Vol. 21, pp. 1583–1600, 2007.
- [26] Y. Yoshikawa, J. Koga, M. Asada, and K. Hosoda. A constructivist approach to infants' vowel acquisition through mother-infant interaction. *Connection Science*, Vol. 15:4, pp. 245–258, 2003.
- [27] E. Meins. *Security of attachment and social development of cognition*. Psychology press, 1997.
- [28] 篠原 郁子. 乳児を持つ母親における mind-mindedness 測定方法の開発. *心理学研究*, Vol. 77:3, pp. 244–252, 2006.
- [29] P. Rochat. *The Infant's World*. Harvard University Press, 2001.
- [30] D. N. Stern. *The interpersonal world of the infant*. Karnac, 1998.
- [31] P. K. Kuhl. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, Vol. 50, pp. 93–107, 1991.
- [32] Masa-aki Sato and Shin Ishii. On-line EM algorithm for the normalized gaussian network. *Neural Computation*, Vol. 12, pp. 407–432, 2000.
- [33] J. Moody and C. Darken. Fast learning in networks of locally-tuned processing units. *Neural Computation*, Vol. 1:2, pp. 281–294, 1989.
- [34] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *J. of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, pp. 1–38, 1977.
- [35] G. E. Peterson and H. L. Barney. Control methods used in a study of the vowels. *Journal of Acoustic Society of America*, Vol. 24, pp. 175–184, 1952.
- [36] K. Ishizuka, R. Mugitani, H. Kato, and S. Amano. Longitudinal developmental changes in spectral peaks of vowels produced by Japanese infant. *Acou. Soc. of America*, Vol. 121:4, pp. 2272–2282, 2007.
- [37] P. Boersma. Praat, a system for doing phonetics by computer. *Glott Int.*, Vol. 5:9, pp. 341–345, 2001.
- [38] G. Rizzolatti and L. Craighero. The mirror-neuron system. *Annual Review of Neuroscience*, Vol. 27, pp. 169–192, 2004.
- [39] V. Gallese, L. Fadiga, L. Fogassi, and G. Rizzolatti. Action recognition in the premotor cortex. *Brain*, Vol. 119:2, pp. 593–609, 1996.
- [40] G. Rizzolatti and M. A. Arbib. Language within our grasp. *Trends in Neuroscience*, Vol. 21:5, pp. 188–194, 1998.
- [41] G. Rizzolatti and L. Craighero. *Neurobiology of Human Values*, chapter Mirror neuron: a neurological approach to empathy, pp. 107–123. Springer-Verlag Berlin Heidelberg, 2005.
- [42] 村田 哲. 模倣の神経回路と自他の区別. *バイオメカニズム学会誌*, Vol. 29:1, pp. 14–19, 2005.
- [43] G. Rizzolatti, L. Fogassi, and L. Gallese. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, Vol. 2, pp. 661–670, 2001.
- [44] 篠原 郁子. 子どもの心的理解の発達を支えるものとは—養育者の感受性及び mind-mindedness の役割. *京都大学大学院教育学研究科紀要*, Vol. 51, pp. 357–370, 2005.
- [45] G. Markova and M. Legerstee. Contingency, imitation, and affect sharing: Foundations of infants' social awareness. *Developmental Psychology*, Vol. 42:1, pp. 132–141, 2006.