神経回路モデルを用いた音声模倣モデルによる音声バブリングと母音獲得過程シミュレーション

神田 尚† 尾形 哲也† 高橋 徹† 駒谷 和範† 奥乃 博†

†京都大学大学院情報学研究科 知能情報学専攻

1. はじめに

人間の言語獲得過程は認知発達における謎の一つとされている. 幼児は親の音声を模倣することで母国語音素を獲得することができ, この音素獲得プロセスを明らかにすることは, 言語獲得基盤を解明する上で重要と言える.

本稿では, バブリング学習による母音獲得モデルを提案する. 我々は, 幼児が不明瞭音を複数含むランダムバブリングにより母音を獲得するという仮説を立てる. 従来多くの母音獲得モデルは, 学習バブリング音が単一母音であった. しかし, 単一母音のみを介した母子間インタラクションにより, 幼児が母音獲得を行うとは考えにくい.

本研究の目標は, 母子間音声模倣インタラクションによる幼児の自己組織的母音カテゴリ形成の再現である. これを実現するため, 我々は 1. 声道物理モデル Maeda モデル [1] を用いたランダムバブリング, 2. 神経回路モデル Recurrent Neural Network with Parametric Bias (RNNPB) [2] の予測誤差に基づくバブリング音の分節化, 3. RNNPB の音声模倣精度に基づく追加学習データの取捨選択, を導入した.

2. 声道物理モデル: Maeda モデル

我々は, 音声合成器として Maeda モデルを用いた. このモデルは, 母音生成時において撮影された構音器官の形状について主成分分析を行った結果得られた 7 つの構音パラメータ‡により構音器官を表現している. 本稿では, 母音のみを扱うため, 母音発声時に一定値をとる Larynx position は扱わない. 表 1 に, Maeda モデルの各母音の第 1·2 フォルマント (F1-F2) を示す.

音声合成器には他にも, PARCOR [3] や STRAIGHT [4] などがあるが, Maeda モデルが人間の解剖学的知見に基づいている点で, 本研究の人間の言語獲得機構の解明に妥当であると言える.

表 1: Maeda モデルの第 1·2 フォルマント.

	/a/	/i/	/u/	/e/	/o/
F1	667	234	269	401	500
F2	1214	2161	924	1894	902

3. RNNPB による時系列データの分節化

3.1 神経回路モデル: RNNPB

RNNPB は, 現状態を入力とし次状態を出力とする予測器である (図 1 参照). RNNPB は再帰結合を持ち, 非線形時系列パターンを学習可能である. さらに, PB 層と呼ばれる入力層を持ち, PB 層の内部値 (PB 値) の変更により複数時系列パターンを生成可能である. また, RNNPB を認識器として用いることで, 希望する時系列データを生成するような PB 値が得られる.

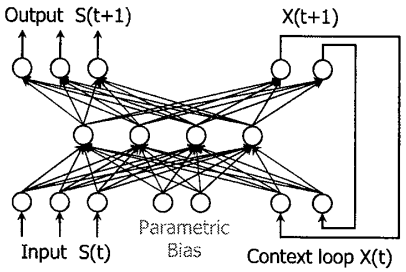


図 1: Recurrent Neural Network with Parametric Bias

3.2 時系列データの複数シーケンスへの分節化手法

本稿では, 安定した予測が可能な区間を単一シーケンスとして分節化を行う. そこで, 次の分節化手法を行った.

i) **初期化**: 入力時系列を分節数に応じて均等に分割.
ii) **学習**: RNNPB の結合重みと各区間の PB 値を更新.
iii) **誤差算出**: 各区間の予測値から予測誤差平均を算出.
iv) **区間更新**: 誤差が隣接区間のものより大きければ区間幅を減少, 小さければ区間幅を増加.
v) 学習が収束するまで ii) から iv) を繰り返す.

単一の PB 値で予測可能なシーケンス (= 単一のダイナミクス) の予測誤差は小さくなる. 一方, 複数ダイナミクスから成るシーケンスでは予測誤差が大きくなる. そのため, 上記の手順により, 区間境界がダイナミクスの境界に近付き, 力学構造に基づく分節化が可能となる.

4. 母子間音声模倣インタラクションモデル

我々が提案した音声模倣モデル [5] に追加学習フェーズを導入し, 母子間インタラクションを構築した. インタラクション手順を以下に示す.

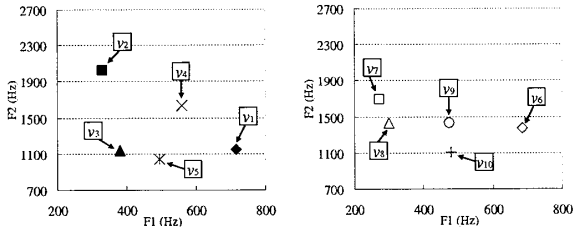
1) **学習 (バブリングによる自己発声経験の獲得)**: 構音パラメータと生成されるバブリング音の音響パラメータのダイナミクスを RNNPB により学習する.
2) **認識 (親の音声を聴取)**: 人間の音声から得られる音響パラメータのみを学習済み RNNPB に入力し, 対応する PB 値を求める.
3) **生成 (親の音声を模倣)**: 2) で求めた PB 値から RNNPB により計算される構音パラメータを求め, Maeda モデルにより模倣音声を生成する.
4) **追加学習 (自己経験から模倣可能な音声・構音動作を学習)**: 模倣音声についてそれぞれ RNNPB の予測エラーを求め, それら平均値よりも低くなった音声・構音ダイナミクスを学習済みの RNNPB に追加学習させる.
5) 2) から 4) を繰り返す.

5. 母音カテゴリ形成シミュレーション実験

本実験では, 4 章で述べたインタラクション手順に基づき, ランダムバブリング学習による母音カテゴリ形成過程をシミュレートした.

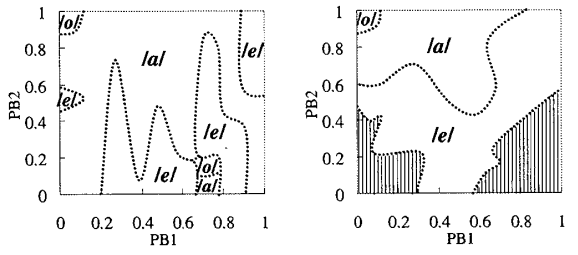
Simulation of Babbling and Vowel Acquisition based on Vocal Imitation Model using Recurrent Neural Network: *Hisashi Kanda, Tetsuya Ogata, Toru Takahashi, Kazunori Komatani, and Hiroshi G. Okuno (Kyoto Univ.)*

‡1. Jaw position, 2. Tongue dorsal position, 3. Tongue dorsal shape, 4. Tongue tip position, 5. Lip opening, 6. Lip protrusion, 7. Larynx position



(a) 学習セット 1　(b) 学習セット 2

図 2: F1-F2 空間におけるバブリング音



(a) 学習セット 1　(b) 学習セット 2

図 3: ランダムバブリング学習後の PB 空間解析結果

5.1 実験条件

学習フェーズでは, バブリング音として, ランダム設定した構音パラメータによる生成音声から 10 種類選択し, 1 セット 5 種類の学習セットを 2 つ用意した (セット 1: /v_1/ ～ /v_5/, セット 2: /v_6/ ～ /v_{10}/). 図 2 に各学習セットの F1-F2 を示す. 各セット毎に, 3 つのバブリング音を含む 45step (30msec/step) の学習データを 10 種類ずつ作成し, 2 つの異なる RNNPB (RNNPB-1, 2) に学習させた (分節数は 8). RNNPB への入力次元は, STRAIGHT スペクトル [4] により求めた基本周波数に影響を受けない MFCC 特徴量 (フィルタバンク:12 次元, 窓幅:250msec, シフト幅:100msec) のうち 5 次元の音響パラメータと, 各音響信号に対する構音パラメータのうち Larynx position 以外の 6 次元を使用し, これらを 0 から 1 に正規化した 11 次元である. RNNPB のニューロン数は, 入出力層: 11, 中間層: 40, 文脈層: 5, PB 層: 2 に設定した.

認識フェーズでは, 2 話者 (話者 1, 2) の 3 連続母音発話 45 種類 (表 2 参照) を用い, 生成・追加学習フェーズでは, 各 RNNPB に対し 1 話者ずつ行った.

5.2 実験結果・考察

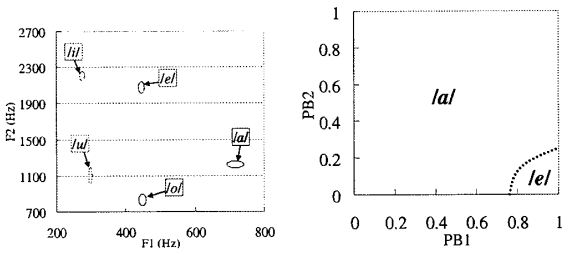
5.2.1 バブリング学習における母音自己組織化への影響

図 2 から, 学習セット 1 は F2 に対する分布が広く, 学習セット 2 は狭いことが確認できる. これは学習初期条件として, 学習セット 1 が舌の上下運動幅が大きく, 学習セット 2 は小さいことを表している.

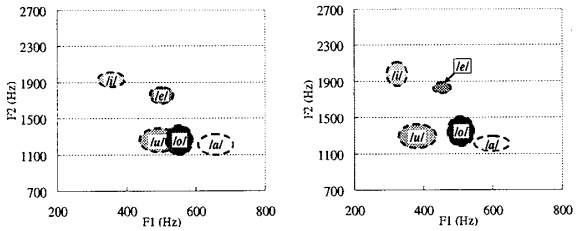
図 3 はランダムバブリング学習後の PB 空間解析結果である. 各 PB 値の生成音が表 1 のどの母音に対応するかを示しており, 図 3(b) の縦線部分は無音空間を表す. 図 3(a), 3(b) から, RNNPB-1, 2 のどちらの場合も, /a/ の母音空間に広がりを持っており, 実際の幼児が生後 1 年間において発話回数の多い母音と対応している [6]. また,

表 2: 人間が発話した認識フェーズで用いた 3 連続母音.

/aeo/	/aeu/	/aia/	/aie/	/aio/	/aiu/	/aoa/	/aou/	/aue/
/eai/	/eia/	/eiu/	/eoa/	/eoe/	/eoi/	/eou/	/eua/	/eue/
/iae/	/iai/	/ieo/	/ioa/	/ioe/	/iua/	/iue/	/iui/	/iuo/
/oae/	/oai/	/oao/	/oau/	/oei/	/oeo/	/oiu/	/oue/	/oui/
/uai/	/uao/	/uea/	/uei/	/ueo/	/ueu/	/uio/	/uiu/	/uoa/



(a) 話者 1 の平均フォルマント　(b) 追加学習後の PB 空間解析結果



(c) 追加学習前の模倣音 F1-F2 分布 (d) 追加学習後の模倣音 F1-F2 分布

図 4: 学習セット 1 における話者 1 に対する追加学習結果

RNNPB-2 について, PB 空間上の無音空間分布も広く, 話者 1, 2 のどちらの音声に対しても模倣能力が確認できなかった. これは, 舌の運動学習における可動域が小さい場合に母音構造の形成が不利となることを示唆している.

5.2.2 追加学習による音素構造の変化

模倣能力を有する RNNPB-1 を用いて各話者の追加学習を行った. 図 4 に, 話者 1 に対する追加学習結果を示している. それぞれ, 図 4(a) が話者 1 の F1-F2 分布, 図 4(b) が追加学習後の PB 空間解析結果, 図 4(c), 4(d) が追加学習前後の模倣音声中の母音区間に対応する F1-F2 分布である. 模倣音声中の母音対応区間は, 分節化手法による予測の安定した区間を割り当てた.

図 4(a) と図 4(c), 4(d) をそれぞれ比べると, 追加学習後の模倣音声が人間の F1-F2 に近付いていることが確認できる. さらに, 図 4(b) を図 3(a) と比較すると, 追加学習により母音構造の収束する可能性を示唆している. これは, 話者 2 についても同様の結果が得られた.

6. おわりに

本稿では, バブリング学習に基づく追加学習母音獲得モデルを構築し, シミュレーションによる母音カテゴリ自己組織化の検証を行った. 本実験では, 1 回の追加学習についての検証のみにとどまっている. 今後の課題として, 追加学習の反復による母音カテゴリ収束の可能性を示し, 音声模倣能力の発達過程の再現を目指す.

謝辞 本研究は科研費学術創成研究, 基盤研究 (S), 及びグローバル COE の支援を受けた.

参考文献

[1] S. Maeda, "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model", Speech production and speech modeling, Kluwer Academic Publishers, pp.131–149, 1990.

[2] J. Tani and M. Ito, "Self-Organization of Behavioral Primitives as Multiple Attractor Dynamics A Robot Experiment", IEEE Trans SMC Part A: Systems and Humans, Vol.33, No.4, pp.481–488, 2003.

[3] 板倉, "偏自己相関関数による音声分析合成系", 音講論集, 1969.

[4] H. Kawahara et al., "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction", Speech Communication, Vol.27, pp.187–207, 1999.

[5] 神田, 尾形, 駒谷, 奥乃, "RNNPB による音響模倣・分節化を用いた音素獲得モデルの提案", 第 70 回情処全大, 2008.

[6] 石塚, 麦谷, 天野, "F1-F2 平面上における成人母音からの距離に基づく幼児母音の月齢変化", 音講論集, 2005.