

言語獲得における母音範疇の形成過程のシミュレーション

宮澤 幸希 白勢 彩子 菊池 英明

早稲田大学人間科学学術院 〒359-1122 埼玉県所沢市三ヶ島 2-579-15

E-mail: m-kouki@moegi.waseda.jp

あらまし 本研究は、ヒトが母国語の母音の音韻体系を獲得する過程を、神経回路網モデルによってシミュレーションすることを目的としている。シミュレーションは自己組織化マップに、単独の日本語話者の F1・F2 の値を入力することで行った。従来より、言語獲得において生得的な機構がどの程度存在し、学習にどのような影響を与えているのかという問題がある。この点に対するアプローチとして、本研究では初期制約を与えたモデルと、制約を与えないモデルによる学習を行い、その結果を比較した。実験の結果、学習の初期段階では、生得的な制約を仮定したモデルのほうが、日本語の母音範疇に対して高い正解率を示し、生得的な機構が自己組織的なメカニズムと言語経験の補助的な役割を果たしている可能性が示された。

キーワード 言語獲得 母音 音声知覚 ニューラルネットワーク 自己組織化

Simulation of the process of learning vowel categories in language acquisition

MIYAZAWA Kouki SHIROSE Ayako and KIKUCHI Hideaki

Faculty of Human Sciences, Waseda University, Saitama, 350-1192, Japan

E-mail: m-kouki@moegi.waseda.jp

Abstract In order to address a problem whether language system is acquired by innate mechanism or by language experience, this study examines how neural network model, *Self-Organizing Maps* (SOM) model, organizes a learning process of Japanese vowel category. As a key to considering this problem, two models were proposed here: one was a "biased" model toward universal vowel category and another was a "nonbiased" model. Our results demonstrated that the biased model showed a high accuracy for Japanese vowel category at the initial stage of the learning process. This result suggested the possibility that innate mechanism played a supporting role in language acquisition.

Keywords Language acquisition, Vowels, Speech perception, Neural network, Self organization

1. はじめに

健常なヒトは、言語のある環境で生活すれば、明示的な教示なしで母国語の音韻体系を獲得できる。獲得した音韻体系は話者や環境の変化に対して頑強であり、既存の音声言語理解技術にはない、人間の言語能力の優れた特性である。しかし、この能力が学習によって獲得されるのか、言語に特化した生得的な機構が存在するのかに関しては、明らかになっていない点が多い。

そこで本研究では、ヒトが母国語の母音の音韻体系を獲得する原理を解明するため、言語獲得過程のモデル化とシミュレーションによる検討を行う。

2. 先行研究

2.1. 母国語音韻体系の形成過程

子音の同定に必要な音響的特徴を連続的に変化させて聞き取ると、連続的には知覚されず、変化が一定の範囲内であれば同一の子音に知覚される、カテゴリ知覚（範疇化知覚）がある[1]。生後数日～6ヶ月の乳児が、母国語に存在しない子音に対してもカテゴリ知

覚を示すことが報告されている[2][3]。すなわち、生後間もない乳児は、言語間に共通した子音を知覚することができる。その後、生後7～8ヶ月で母国語の子音範疇が獲得され、母国語にない範疇は消失することが分かっている[2]。例えば日本語ではこの時期を過ぎると、子音 /r/, /l/ は一つの範疇として知覚されるようになる。

母音でも、類似した知覚様式が報告されているが、子音に比べると範疇の境界が明確ではないため、カテゴリ知覚と区別して「マグネット効果」とよばれる[3]。子音と同様に、生後間もない乳児は母国語にない母音の対比を弁別できることが報告されているが、明確な範疇をもつかどうかは明らかでない。また、子音より早く、生後6ヶ月程度で母国語の母音範疇が獲得されるという報告がある[2]。

乳児は話者適応に関しても優れた能力を示す。ヒトは異なる話者が発声した母音を容易に同定することができるが、生後6ヶ月の乳児も、可変性によらず正しい母音を高確率で弁別できることが明らかになっている[2]。前述したカテゴリ知覚のメカニズムは、この可

変性を標準化・範疇化する際に効果的に機能している可能性が指摘されている[3]。

乳児が生後約半年間で聴取する音声は、限られた人数の限定された語彙でありながら、前述のように乳児は母国語の音韻体系を獲得できる(「刺激の貧困」問題)。これを説明する立場として、生得的なメカニズムが仮定されている。Kuhl&Millerの研究によると、ヒト以外の哺乳類にも、言語間に共通な子音に対するカテゴリ知覚の能力があることが示されている[4]。これは、言語の知覚能力が生得的に備わっているとする立場からは、哺乳類の聴覚系には、元来特定の音響的特徴に対してカテゴリカルに応答する特性があり、ヒトの言語の音韻体系はこの生物学的制約を利用する形で生まれたと説明される。聴覚神経経路には、音信号の急激なスペクトル変化に応答するニューロンがあるという報告があるため[5]、この過程で子音が抽出されている可能性がある。母音範疇の形成過程に関しては、子音ほどには生得的な制約を受けていない。すなわち、学習の影響がより大きい可能性がある。しかし、何に対して、どの程度生得的な機構が影響しているのかは明らかになっていない。

言語の生得性に対し、学習による獲得過程を主張する立場もある。この考えによると、乳児は周囲の成人が発話する音素の統計分布を母国語の音韻体系の学習に利用している可能性がある[3]。両者の考えは、近年では対立するものではなく、生得的な過程と学習経験により言語が形成されるとする立場が主である。しかしながら、生得的な過程と学習経験がどのような関係で言語獲得の過程に役割を担っているのかについては、明確ではない。そこで、これらの知見をふまえて、母国語の音韻体系の獲得メカニズム、特に生得的機構と学習の影響を考察するため、我々はこれまで子音範疇の獲得について調べてきた。

2.2. 子音範疇の獲得モデル

我々は、コホネンの Self-Organizing Maps(以下 SOM, 詳細は次章)が、生得的な機構をもたない聴覚野モデルとして適当であると仮定して、これまでに有声・無声子音範疇の形成過程のシミュレーションを行った[6]。SOM に成人英語話者が発声する子音/d/, /t/の有声開始時間の頻度分布に基づく値を入力して学習を行った結果、SOM が分類したクラスタの境界は、ヒトの有声・無声子音の範疇の知覚境界とほぼ一致した。

この結果だけでは、人間が生得的な制約なしで母国語の子音範疇を獲得できるということは主張できないが、特殊な言語処理機構をもたない神経回路モデルと、音声の統計分布のみを用いた実験で子音範疇の再現ができたことから、言語獲得における自己組織的な機構と言語経験による学習の重要性を指摘した。

3. 母音範疇の獲得モデル

子音範疇の獲得モデルの検討により、母音範疇の獲得モデルを作成し、シミュレーションによる検証実験を行い、母国語の獲得過程をより詳細に考察する。

3.1. 実験の目的

本実験では、母国語の母音範疇の以下のパラメータが、学習によって形成される過程を検証する。

- ・ 母音範疇の数
- ・ 母音範疇の境界(正解率)
- ・ どの程度の学習量で母音体系が形成されるか

生得的な制約と学習の関係を解明するために、生得的な制約をもたないモデルと、制約を仮定した(後述)モデルで検証実験を行い、結果を比較する。さらに、発達初期の話者適応を視野に入れた予備的な検討として、別話者音声で評価する実験も行う。

乳児の言語環境を模擬するために、乳児が最初に親の音声を聞くと仮定して、単独の日本人話者の F1, F2 の値を入力とした。なお、F1, F2 を入力としたのは、生得的な制約として導入した母音体系([7]による、詳細は後述)のデータが F1, F2 で表現されているためである。フォルマントを使用することの妥当性を厳密に検討したわけではないが、メルケプストラムとフォルマントを比較する予備的な分析を行ったところ、結果に明確な差は見られなかった。この点について、引き続きの検討課題としたい。

3.2. 実験の手続

フォルマント解析:

日本語話し言葉コーパス(CSJ)に収録されている講演再朗読音声 R00M0036(60代男性, 東京生まれ)を学習実験の対象とした。フレーム長 20ms, フレームシフト長 10ms で解析し、線形予測分析によって F1, F2, (F2-F1) を求め、メルスケールに変換した。続いて過去 2 フレーム分の値を合わせて、9次元の学習用データセットを作成した。3 フレームをセットにして入力としたのは、1 フレームで学習を行うよりも正解率が約 10% 向上したためであり、ヒトの聴覚系において時間情報がどのように符号化されているのかを厳密にモデル化したわけではないので、これについては議論の余地があると考えている。F1 が 10Hz 以下もしくは 1500Hz 以上、F2 が 3000Hz 以上、F1, F2 が前フレームより 500Hz 以上変化していた場合、エラーデータと判断して削除した。

学習結果の評価用には、CSJ に付属している各音素の開始タグに基づいて各母音の定常部 20ms を取り出して同様の解析を行い、母音ごとに 300 個の音素ラベル(正解情報)つきデータセットを作成した。Figure 1-1 に学習用・評価用のデータセットを例示する。

また、別話者音声による評価用に、CSJ の講演再朗

読音声 R00F0028(60代女性, 東京生まれ)と R00M0134 (70代男性, 神奈川生まれ)を分析し,母音ごとに300個の音素ラベルつきデータセットを作成した. Figure 1-2 に別話者音声の評価用データを例示する.

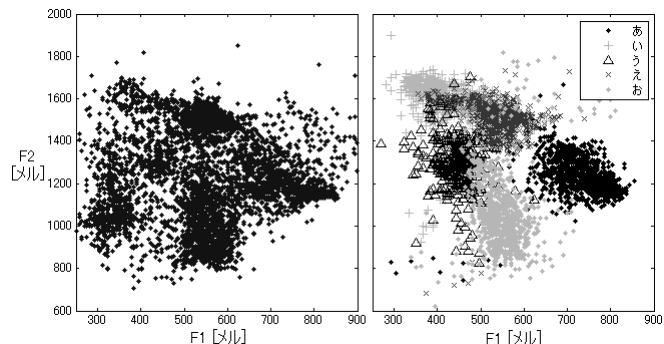


Figure 1-1 学習用データ (左), 評価用データ (右)

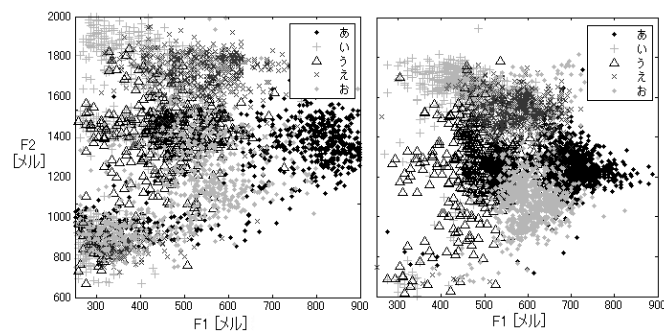


Figure 1-2 別話者評価用データ (左:女性, 右:男性)

シミュレーション:

本実験では,学習モデルとして SOM を使用する. SOM は,大脳の感覚皮質における情報処理を参考に考案された自己組織化ニューラルネットワークの一種である[8]. SOM は主に視覚野の機能研究に用いられてきたが,一次視覚野と一次聴覚野の構造はほぼ同じであることが分かっている[9]ため,聴覚の機能研究に応用することも可能であると思われる.

実験によって, SOM には,

- ・ 入力値の類似度,出現頻度分布を自動的に(教師信号や報酬なしに)抽出し分類する
- ・ 一度も学習していない値でも,どのカテゴリに属するか分類することができる
- ・ カテゴリ数が未知の入力信号でも,後述する手法[10]によってカテゴリ数を推定できる

といった特徴があることを確認できたため,明示的な教示なしで母国語の音韻体系が形成される過程を検証する手がかりになると考える.

SOM による実験手続の詳細を述べる.本研究では,可能な限り言語に特化した機構を除いたモデルと,生得的な機構を仮定したモデルを作成し,それぞれのモ

デルと同じ母国語の統計的分布を与えて学習実験を行い,結果の相違点を分析することで,母音体系の獲得過程における生得的機構と学習の影響を明らかにすることを考えた.

可能な限り言語に特化した機構を除いたモデルとして,SOMのノードを全て同一の重みベクトルに初期化した,16ノードの1次元SOMを使用する(制約なし条件).学習は学習用データの中から,毎回(Learning Stepごとに)ランダムに選択して入力する.

生得的な機構を仮定したモデルとして,あらかじめ各ノードに言語間に共通した母音のバイアスをおけることとした,16ノードの1次元SOMを使用する(制約あり条件).具体的には,文献[7]で提唱されている分散収束理論に基づいて計算された母音体系(Figure 2)のF1,F2の値とその差分値を入力として,あらかじめSOMの学習を10エポック(310回)分行っておく.これは,生後間もない乳児が示すとされる,言語間に共通した母音の弁別能力が,一次聴覚野の神経細胞の結合状態として生得的に備わっていると仮定したといえる.制約あり条件では,初期制約の学習を終えた後,続いて制約なし条件と同様の学習を行う.

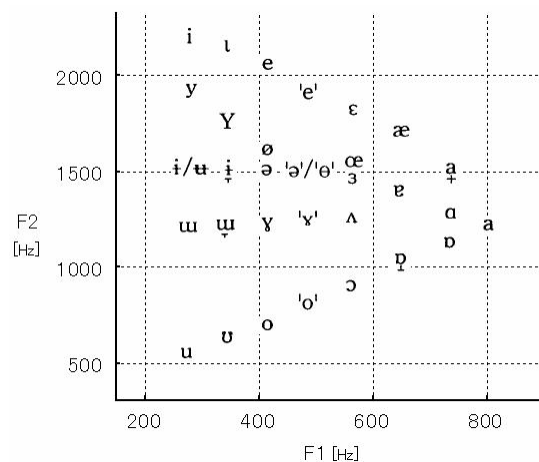


Figure 2 分散収束理論に基づく母音体系 ([7]による)

3.3. 評価方法

評価用データセットを使用して,SOMが獲得した母音範疇の正解率を求める.具体的には,学習済みSOMに評価用データセットを入力して,それぞれのニューロンと入力データの対応関係を調べ,後述の手法によりカテゴリを統合し,SOMが獲得したカテゴリの数と境界を求める.続いて,正解情報に基づく母音範疇とSOMのカテゴリを比較して,その一致率を正解率とした(Figure 3).

本稿でのSOMの学習の実例と,学習済みSOMのカテゴリ統合の手法を述べる.例として,分散が等しく

平均が異なる3種類の正規分布を分類する場合を考える。入力がどのような分布をとるか不明で、何種類のカテゴリに分類できるのかも未知であるとする。母音数は言語によって異なり、乳児は母国語に応じた母音数を学習している。この点を仮定すると、SOMのニューロンの数は十分に多くとる必要がある。SOMによる学習の結果をFigure 4に示す。SOMでは、カテゴリ間の境界付近にニューロンが割り当てられ独立したカテゴリになるが[10], Figure 4の結果は、ヒトのマグネット効果(境界付近の母音も典型的な母音範疇として知覚される)を再現できているとはいえない。

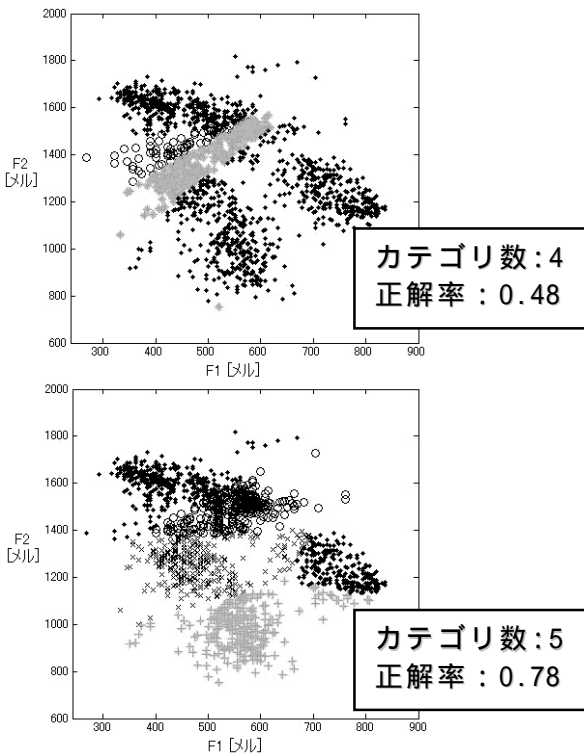


Figure 3 正解率(評価用データとの一致度)の例

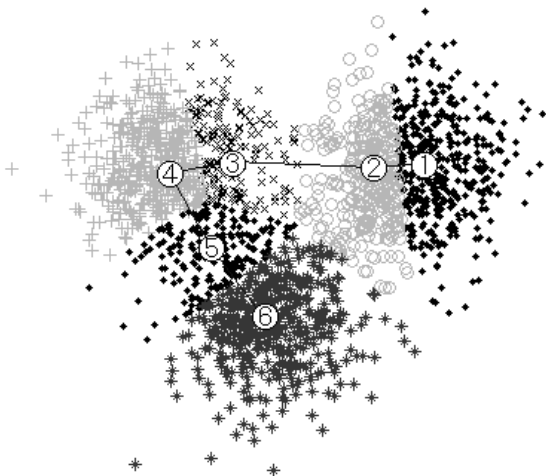


Figure 4 6ノードの1次元SOMによる分類結果

そこで、学習結果をもとに、類似したカテゴリを同一のカテゴリに統合して扱えるよう、枠組みを考える必要がある。ニューロンが1列に配置された1次元SOMについて、以下のような有効な手法が提案されており、本稿ではこれに基づいて評価を行った[10]。

- 1) 学習を終えたSOMのそれぞれのニューロンに対応した入力データの数を調べ、入力データのカテゴリ集積度を表すヒストグラムを作る。
- 2) ヒストグラムのピークピッキングを行い、各カテゴリの中心となるニューロンを求める。
- 3) カテゴリの中心以外のニューロンの重みをゼロにして、入力データとの対応関係を再計算する。

Figure 5に、この手法を適用後のSOMの例を示す。上図がヒストグラム、下図が再計算の結果である。カテゴリが統合され、3つの正規分布を抽出できている。

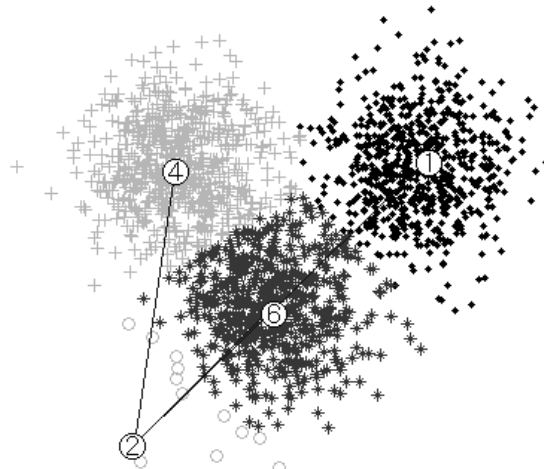
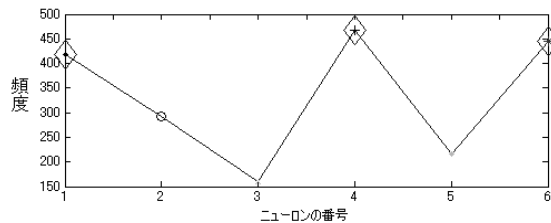


Figure 5 カテゴリの統合([10]の提案手法による)

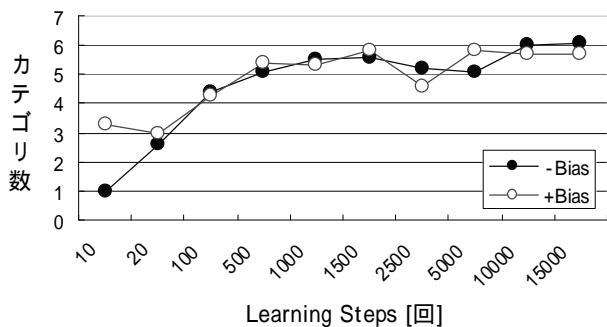
なお、生物学的により妥当性があると思われるカテゴリ統合の手法として、SOMに入力頻度に基づいた母音範疇化のアルゴリズムを導入する手法[11]が報告されており、今後検討したい。

3.4. 実験結果

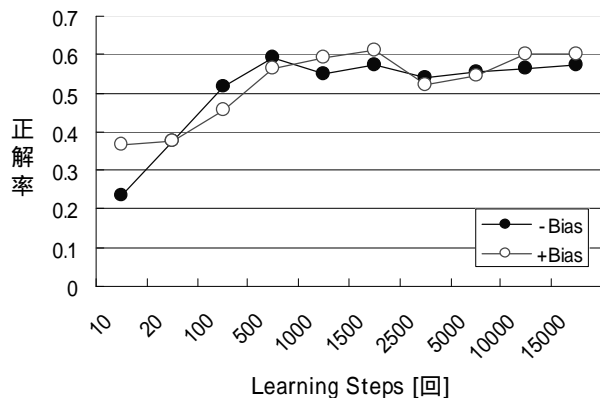
制約なし(-Bias)条件, 制約あり(+Bias)条件それぞれについて、学習の進行による母音範疇の数と正解率の変化を示す。Figure 6の(a)は、横軸を学習量(Learning Steps, 学習データの入力回数), 縦軸をカテゴリ数としてプロットしたものである。(b)は、横

軸を学習量, 縦軸を正解率としてプロットしたものである。(c)は, 別話者による評価の結果をプロットしたものである。例えば, 「-Bias 別話者男性」とは, 制約なし条件の学習結果を, 異なる男性話者のデータで評価したときの正解率を示す。なお, カテゴリ数・正解率の各値は, それぞれ同条件で 10 回ずつ学習を行った結果の平均である。

(a) カテゴリ数



(b) 正解率



(c) 別話者による正解率

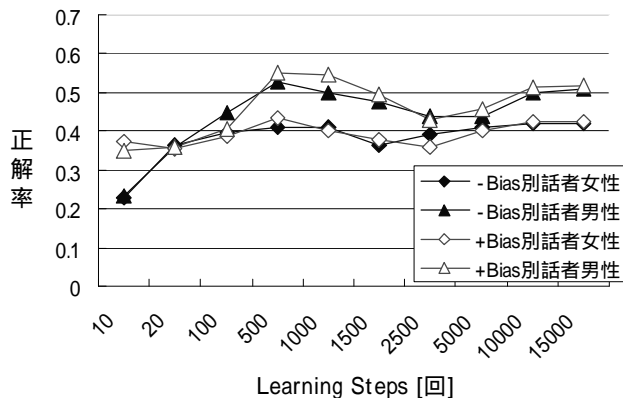


Figure 6 SOM の分類結果

まず, 制約ありの条件となしの条件とを比較する。Figure 6 の(a)カテゴリ数, (b)正解率の結果を見ると, 学習の初期段階 (Learning Steps = 10 ~ 100) では, 制約あり条件のほうが正解率は高く, また制約あり条件のほうが, 日本語の母音クラス数「5」により早く近づいているといえる。学習が進行するにつれて (Learning Steps = 100 ~), 制約なし条件と制約あり条件の差は少なくなっている。

結果を別話者で評価した結果について, Figure 6 の (b)正解率, (c)別話者による正解率の結果によると, 別話者が異性 (女性) の場合は, 学習量に応じた変動は明確ではないが, 別話者が同性 (男性) の場合は, 学習による正解率の向上が見られた。この点について, 他の話者音声での検討も行ったところ, 学習データと評価データの組み合わせにより, 結果に相違が見られることがわかった。今後, より詳細に検討したい。

4. まとめ

本稿では, 母国語の母音体系の獲得過程における生得的機構と学習の影響を明らかにすることを目的として, 初期制約を与えたモデルと, 制約を与えないモデルによる学習実験を行った。その結果, 学習の初期段階では, 初期制約を与えたモデルのほうが日本語の母音範疇に対する高い正解率を示すことが分かった。これは, 乳児が短期間で新たな音韻の範疇を獲得できたとする先行研究[3]の知見を支持する結果であると考えられる。

本実験の結果だけでは, 人間が言語獲得の際に生得的な機構を使用しているということは主張できないが, 言語獲得の初期段階において, 生得的な仕組みが効果的に働いている可能性が示唆されたといえる。ただし, 生得的な制約の生物学的妥当性や脳内での表現形式, 母音体系の学習に必要な知識を乳児が生得的に持っているのか, 学習によって獲得したのかに関しては議論の余地がある。また, 制約にどの程度の重みを与えるのが妥当であるのかに関しては, 乳児の発達研究の知見も考慮してより詳しく検討したい。

本研究では, 過去 3 フレーム分 (40ms に相当) の入力データの学習において, もっとも高い正解率が得られた。40ms の入力 が生物学的に, また言語の情報処理の観点から見て妥当であるかどうかについても検討が必要である。

別話者音声による評価実験の結果, 話者による結果の相違が見られたことについては, 今後, 話者の音声やモデルの再検討などにより改めて検討を加えていきたい。現段階では, この結果は, 学習の過程において, 学習を搬化させやすい音声とそうではない音声があることを示唆しているとも捉えることができ, 従来指摘

される非母語音声学習者の学習過程に類似の結果であるとの解釈を考えている。

今後の課題として、獲得した音韻情報の統合による音節の学習、音節の統合による単語の学習、単語の意味概念の学習と意味体系の学習を統一したモデルの構築を目指している。現在、SOMの多層化によってこれを再現することを試みており、継続して検証したい。

文 献

- [1] 重野純, 音の世界の心理学, ナカニシヤ出版, 2003.
- [2] Peter W. Jusczyk, The Discovery of Spoken Language, MIT Press, 2000.
- [3] Patricia K. Kuhl, Early language acquisition: Cracking the speech code, Nature Reviews (Neuroscience), 5, pp.831-843, 2004.
- [4] Patricia K. Kuhl, Miller, J. D., Speech Perception by the Chinchilla: Voiced-Voiceless Distinction in Alveolar Plosive Consonants, Science, 1975.
- [5] 内藤泰, 一次聴神経による語音の符号化, 耳鼻臨床, 1993.
- [6] 宮澤幸希, 白勢彩子, 菊池英明, 音声による音カテゴリーカル知覚の習得モデルに関する研究, 人工知能学会全国大会, 2E3-1 (CD-ROM), 2006.
- [7] Boe Louis-Jean et al., 言語における母音体系の普遍的性質, 日本音響学会誌, 58(7), 450-458, 2002.
- [8] T. Kohonen, 自己組織化マップ, シュプリンガー・フェアラー東京, 2005.
- [9] Jeffrey L Elman, et al, 認知発達と生得性, 井立出版, 1998.
- [10] 寺島幹彦, 白谷文行, 山本公明, 自己組織化特徴マップ上のデータ密度ヒストグラムを用いた教師なしクラスタ分類法, 電子情報通信学会論文誌, 1996.
- [11] Guenther Frank H, Bohland Jason W, 音カテゴリーの学習 - ニューラルモデルとそれを支持する実験結果 -, 日本音響学会誌, 58(7), 441-449, 2002.