F-014

無意識的引き込み模倣により母音を獲得する模倣学習モデル

A Learning Model of Acquiring Vowel Categories Based on Unconscious Anchoring in Maternal Imitation

大段 智広† 鈴木 輝彦‡ 太原 育夫‡
Odan Tomohiro Teruhiko Suzuki Ikuo Tahara

1. はじめに

乳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であるといえる．近年では母子間における相互模倣・共同注視による知能獲得・学習が注目され，研究が行われている [1][2]．本稿では，生後間もない乳幼児の母音獲得能力に注目し，母子間における親の模倣音声から母音のカテゴリを生成する模倣学習モデルを提案する．

2. 人間の言語獲得能力

標準日本語において母音は 5 種類であるが，アラビア語では 3 種類であるなど，母音の種類数は環境 (言語) に依存する．人間は乳幼児のとき，学ぶ環境によって適切な母音を獲得し，言語に応じて母音の種類数もまた適切に獲得する．また，人が音声を知覚する際，カテゴリ知覚と呼ばれる知覚現象が確認できる．たとえば，発声音を/i/から/e/へと徐々に変えて聞かせると，途中で/i/と聞こえていた音がある時点 (カテゴリ境界) を境にして/e/としか聞こえなくなる．初期の知覚実験による検証では，カテゴリ知覚は音声のみ，それも子音のみで生じるとされたが，母音でも条件によってはカテゴリ知覚が生じること，音声以外の刺激でも生じることなどが分かってきている [3]．

3. 模倣による母音獲得

乳幼児が周りの成人と同じ言葉が話せるようになるためには，乳幼児自身と母親 (養育者) の 2 者間において模倣学習を行うことで言語を獲得していると考えられる．その要因として，以下の 4 つが挙げられる．

- 母親による高頻度の乳児の音声模倣 [4]
- 母親の模倣による乳児の音韻様の発話の促進 [5]
- 乳幼児の馴化による学習の切り替え

  乳幼児は親の行動を模倣し，それを繰り返すことで学習を行っていることが知られているが，同じ刺激を受け続けると，やがてそれに飽きてしまい，注意を他の事に向ける [6]．
- 無意識的引き込み模倣

  教示者がある母音を聴取し，模倣する際の発声が無意識のうちに教示者自身の母音に引き込まれる現象をいう．これは乳幼児が聞いた音素を自身の母音の音素に対応付けて音声知覚していると考えられ，教示者の無意識的引き込み模倣の検証，母音発見などについての研究が行われている [7]．

さらに，乳児は優れた聴覚を持っており，成人でも聞き分けることのできないような微妙な発話の違いまで聞き分けることが可能であると言われている [3]．また，乳幼児は発達の非常に早い段階から，事物の間に類似性を認め，類似したものを一つのまとまりとしてカテゴリを形成する能力が備わっている [8]．したがって，生後まもない段階の乳幼児は，自身の発声を母音などの音に認識する前に，無意識的引き込み模倣によって得られた親の音声から母音カテゴリを形成することによって親の音声から母音獲得を行っていると考えられる．

4. 模倣学習モデル

4.1 本モデルの目的

本稿の目的は，上記の観点から模倣により母音獲得を行うモデルを作成することで母音音素獲得プロセスを明らかにすることである．乳幼児の発話を聴いた親は乳幼児の声 (発話音) を模倣して発話するという一連の行動を繰り返していき，乳幼児は親の声 (模倣音) を聴き，自身の声と比較を行うことで，母音音素獲得の初期に親の母音カテゴリを獲得していくと考えられる．そこで，この母音音素獲得プロセスを説明するために模倣学習モデルを作成し，検証を行う．

4.2 母音とフォルマント

音声は，声帯振動によって生成された音源波 (喉頭原音) が声道で共鳴することで形成される．この声帯音源波が基本周波数 (ピッチ) と呼ばれ，声道つまり咽頭喉頭および唇，舌，歯，顎，頬で構成される口腔，さらに鼻腔，副鼻腔で共鳴することにより特定帯域ごとに倍音が増幅される．この増幅された成分の塊もしくはピークをフォルマントと呼ぶ．周波数の低い順に第 1 フォルマント，第 2 フォルマント，…という様に数字を当てて呼び，それぞれ F1,F2 と表記する．また，母音においては F1,F2 は定常な値を取り続けるため，その値によって母音を表現することが可能である．

4.3 本モデルの概要

本モデルでは，教示者はモデルにより生成された発話音を聞く．そして，その教示者が無意識的引き込み模倣によって発声した音声から F1,F2 を抽出し，発話音の F1,F2 と比較処理することで，母音カテゴリを 1 つも持たない状態 (初期状態) から母音カテゴリの生成と母音カテゴリの修正を繰り返し，母音カテゴリを形成する．これによって，各母音における教示者の模倣音をフォルマント周波数の近い音声同士にまとめ，母音カテゴリの領域を徐々に拡張することで母音カテゴリの獲得とその領域の学習が期待できる．

†東京理科大学大学院理工学研究科情報科学専攻
‡東京理科大学理工学部情報科学科

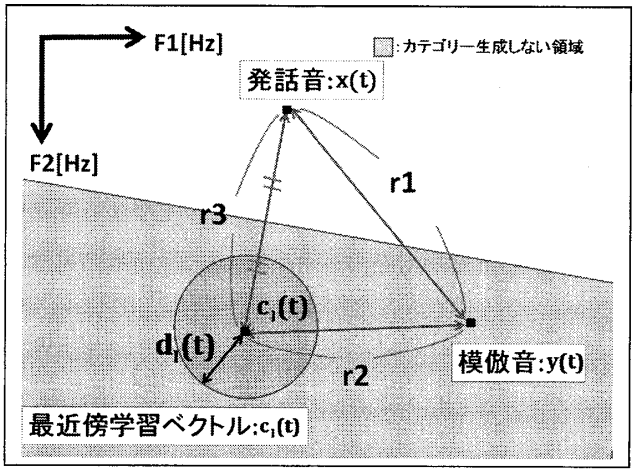

図 1: 模倣音, 発話音, 最近傍学習ベクトルの距離関係

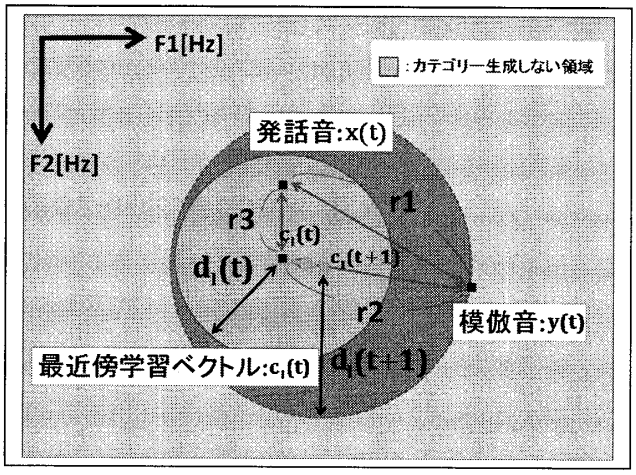

図 2: 最近傍母音カテゴリ (学習ベクトル) の修正

4.4 模倣学習手順

本モデルにおける模倣学習手順を以下に示し, 母音カテゴリにおける処理について説明する.

1) 乳幼児の発話 発話音の生成
本モデルの発話音は F1,F2 を可変パラメータとするフォルマント合成音とする. 生成される合成音は限定された一定の範囲内で発話する. また, 乳幼児の馴化による学習の切り替えを考慮し, 同じ発話を最大模倣回数 α_{max} 回繰り返していれば, 新たにランダム発話を行う.

2) 親の模倣発話 教示者の無意識的引き込み模倣による母音発話
本モデルの発話音を聴取した教示者はその音声を模倣発話する. 教示者の模倣音は無意識のうちに教示者自身の母音に引き込まれて発話されることで, 本モデルは教示者の母音カテゴリからの母音発話を得る.

3) 模倣学習 母音カテゴリの生成, 最近傍母音カテゴリの修正
教示者の模倣音から F1,F2 を抽出し, 発話音と比較することで母音カテゴリの生成, 領域拡張を行う.

4) 1) に戻る

4.4.1 母音カテゴリの生成

本モデルでは, 人間の母音発話での F1,F2 は定常な値を取り続けるが, 発話をする際は声道など身体の調音§ 器官の形状を変化させることによって音声を生成しているために, 同じ人間が母音発話を複数回繰り返し, 発話をすると各フォルマント周波数の近い音声同士にバラついて存在する. したがって,1 つの母音カテゴリを 1 つのクラスタとして考え, クラスタリングの手法を母音カテゴリの生成に用いる. これによって, 乳幼児のカテゴリ形成能力として類似する音声同士を 1 つの集まりとしてまとめ, 母音カテゴリとその領域を得ることができる. しかし, 一般的なクラスタリングの手法では, あらかじめクラスタ数を指定することが多い. 本モデルは, 母音カテゴリを 1 つも持たない状態から母音カテゴリを 1 つずつ獲得することを表すため, 母音カテゴリ数を単純に指定することができない. そこで, 時間毎に与えられた音声データを F1-F2 二次元空間上で表現し, 図 1 のように, 乳幼児自身の声 (発話音) $x(t)$, 親の声 (模倣音) $y(t)$, 最近傍母音カテゴリ (学習ベクトル), $c_l(t)$ の三者間の距離関係から母音カテゴリの生成を行うか否かの逐次判定を行う. 図 1 において,

r1: 発話音 $x(t)$ と模倣音 $y(t)$ の距離

r2: 模倣音 $y(t)$ と最近傍母音カテゴリ (学習ベクトル)$c_l(t)$ の距離

r3: 発話音 $x(t)$ と最近傍母音カテゴリ (学習ベクトル)$c_l(t)$ の距離

$c_l(t)$: 時刻 t における l 番目に生成された最近傍母音カテゴリ (学習ベクトル)

$d_l(t)$: 過去に割り当てられた音声データ $y(k)$ と最近傍母音カテゴリ (学習ベクトル)$c_l(k)$ の最大距離 $(k = 1, ..., t-1)$

とする.

母音カテゴリを生成しない領域は図 1, 図 2 のような 2 つの場合に分けられる. 母音カテゴリ生成を行うか否かの場合分けを以下の式で表す.

$$\begin{cases} New\ Category & (No\ Category) \\ New\ Category & (d_l(t) < r3\ and\ r1 < r2) \\ Not\ New\ Category & (d_l(t) > r3) \\ Not\ New\ Category & (d_l(t) < r3\ and\ r1 > r2) \end{cases}$$

図 1 は,$d_l(t) < r3$ である場合を示し $r1$ と $r2$ の大小関係によって判定を行う. 図 2 は,$d_l(t) > r3$ である場合を示

§声帯の動きによって生じた気流を, 音声器官が音声に変えることをいう.

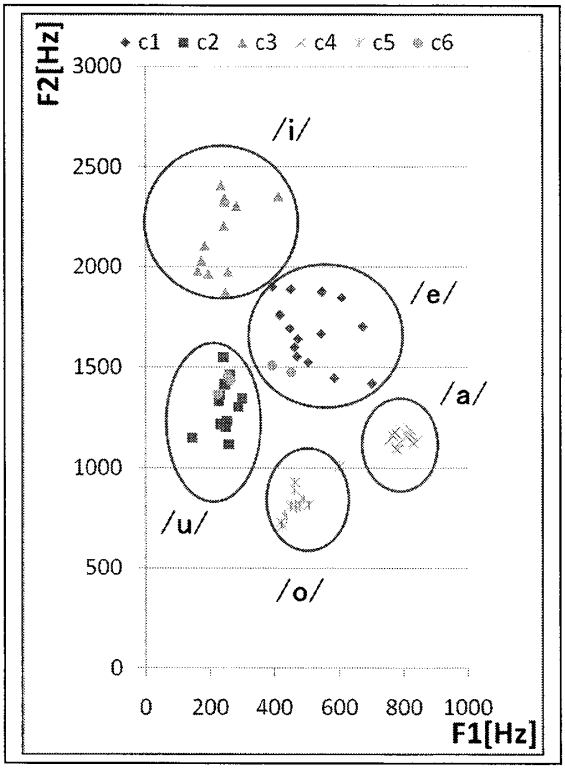

図 3: 模倣音に対する母音カテゴリの分布 (発話数:計 75 回)

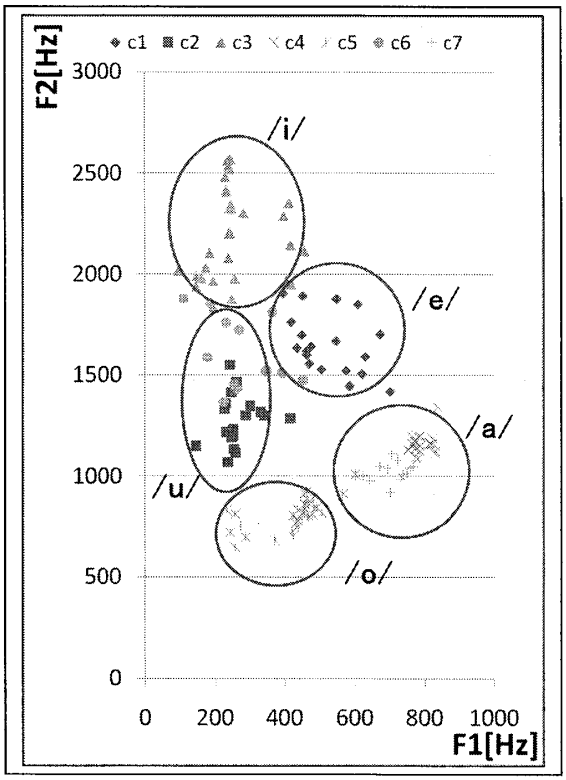

図 4: 模倣音に対する母音カテゴリの分布 (発話数:計 150 回)

し, このとき模倣音 $y(t)$ は最近傍学習ベクトルに必ず割り当てられる. また, 新しい母音カテゴリを生成する場合は与えられた模倣音 $y(t)$ を母音カテゴリの初期値として与える.

4.4.2 最近傍母音カテゴリ (学習ベクトル) の修正

与えられたデータ $y(t):(t=1,2,...)$ に対して割り当てられた最近傍学習ベクトル, または母音カテゴリ中心点 $c_l(t+1)$ をオンラインクラスタリングの LVQ[9] の手法を用いて以下のように値の修正を行う.

$$\begin{aligned} c_l(t+1) &= c_l(t)+\beta(t)[y(t)-c_l(t)] \\ c_k(t+1) &= c_k(t), k \neq l \end{aligned}$$

ここで, 模倣回数を α とし, 学習率 $\beta(\alpha)=\frac{\text{定数}}{\alpha}$ とする. また, 母音カテゴリの中心点 c_l を修正した際に, その中心点 $c_l(t+1)$ と模倣音 $y(t)$ の距離 J を計算し, 過去に割り当てられた音声データ $y(k):(k=1,...,t-1)$ との最近傍学習ベクトル $c_l(k):(k=1,...,t-1)$ の最大距離 $d_l(t)$ と比較し以下の式で距離 $d_l(t+1)$ を更新する (図 2).

$$J=\sum ||y(t)-c_l(t+1)||^2$$

$$d_l(t+1)=\begin{cases} J & (d_l(t)<J) \\ d_l(t) & (d_l(t)>J) \end{cases}$$

5. 実験

5.1 実験の目的

本モデルにおいて, 母音カテゴリを 1 つも持たない状態から本モデルからの発話音によって得られた教示者の母音/a/,/i/,/u/,/e/,/o/の模倣発話から 5 つの母音カテゴリが得られるか否かの検証を行う.

5.2 実験の条件

実験者は F1,F2 を可変パラメータとするフォルマント合成音に対して, イヤホンで 2 秒間聞いた後, その音を模倣する. これを 1 回とし, 同じ合成音で計 3 回行うことを 1 セットとする. 聞かせる音をランダムに 1 セット毎変えて,25 セット計 75 回,50 セット計 150 回繰り返す.

ランダム音声は The Snack Sound Toolkit[10], 声道物理モデル Maeda[11] での 5 つの母音平均フォルマント (表 1) から半径 200[Hz] の円内で発話させた. また, 合成音の生成と分析には The Snack Sound Toolkit を用い, 音声の録音にはサンプリング周波数は 8k[Hz], 量子化 bit は 16[bit] で行った. カテゴリ生成条件には,最大模倣回数 :$\alpha_{max}=3$, 学習率 :$\beta(\alpha)=\frac{0.3}{\alpha}$ とし, 日本語話者 1 人の各 150 個の音声データ使ってそれぞれ 25 セット計 75 回, 50 セット計 150 回発話分の学習を行う.

5.3 実験結果

図 3 に初期状態から計 75 回発話分の学習を行った結果を示す. 母音カテゴリは c1,...,c6 の 6 つが生成され, 図 4 では図 3 の状態から追加で 75 回発話分 (計 150 回) の学習を行った結果で, 母音カテゴリは新たに 1 つ生成さ

表 1: 発話させた合成音声の中心 第 1, 第 2 フォルマント

	/a/	/i/	/u/
F1[Hz]	710,667	280,234	310,310
F2[Hz]	1100,1214	2250,2161	1400,870
	/e/	/o/	
F1[Hz]	550,401	590,500	
F2[Hz]	1770,1894	880,902	

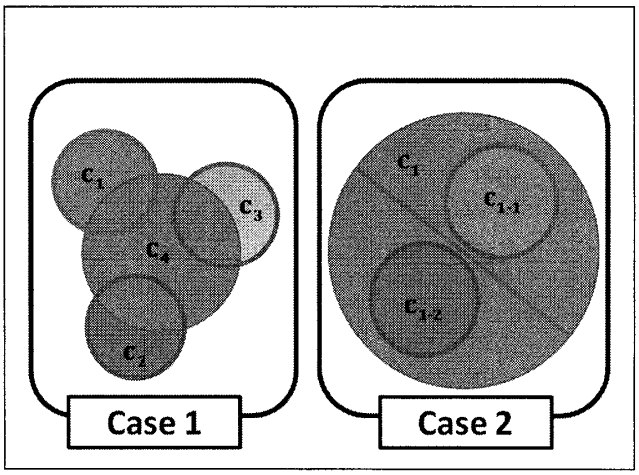

図 5: 実験結果から得られた特殊な母音カテゴリ

れ, 全部で 7 つになった. すなわち, 教示者の 5 つの母音/a/,/i/,/u/,/e/,/o/を模倣した音声から生成された母音カテゴリは 5 つより多くなった.

5.4 考察

5.4.1 母音カテゴリ数

図 3 と図 4 において生成された母音カテゴリ c1～c5 はそれぞれの母音として生成されている. したがって, 教示者の各母音カテゴリ 5 つを得ることができ, 本モデルは母音音素獲得プロセスをよく表したといえる. しかし, 教示者の各母音カテゴリ以外に c6,c7 の 2 つの母音カテゴリが別に生成された. これは, 母音の音声については成人であれば, それぞれの母音, たとえば,/a/と/e/の中間音を正確に知覚し, 発話できることにより教示者の無意識的引き込み効果が弱くなったことが第一の要因として考えられる. しかし,5 つの母音として知覚される模倣音から 5 つ以上の母音を得られたことは妥当な結果といえる. なぜならば, 得られた母音カテゴリはいずれも, 複数の母音に対する中間音として知覚されるか (図 5:Case 1), ある 1 つの母音に対して,F1-F2 二次元空間上で複数に分割された母音カテゴリとして分布しているからである (図 5:Case 2). この結果は,5 つの母音から 5 つ以下の母音カテゴリしか得られない場合とは異なり, 今後の乳幼児の言語獲得においては特に問題なく, むしろ獲得が容易になると考えられる.

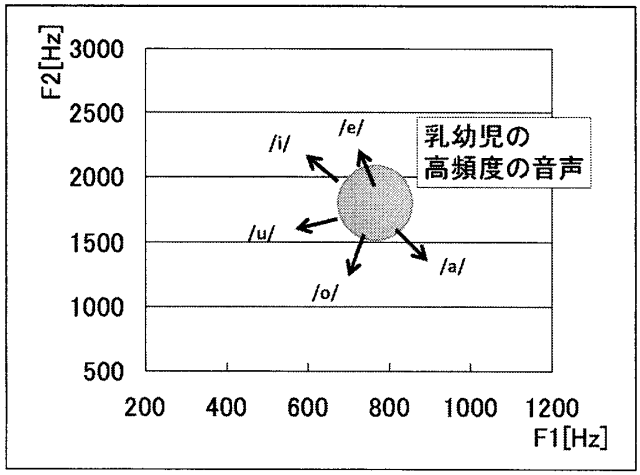

図 6: 乳幼児の音声における高頻度発声領域

5.4.2 母音カテゴリの統合

実験結果では, 生成された母音カテゴリ数はいずれも 5 つより多かった. しかし, 標準日本語の話者における母音は 5 つである. 本モデルでは, 与えられた音声を 1 つずつ処理していくことで母音カテゴリの生成および修正を行い, 教示者の各母音カテゴリ 5 つを獲得した. しかし, このままでは本モデルの母音カテゴリは増え続けてしまうので, いずれ母音カテゴリの統合を行う必要がある. また, 図 4 では, 母音/u/の模倣音から生成された母音カテゴリ 2 つの内の 1 つは母音/o/に分類された. このような母音カテゴリに対して統合を行うためには, 今まで得られた音声を再び一定量にまとめて処理するか, 自身の喉の形状や舌の位置などの調音器官の構造や調音運動を考慮することで統合していることが考えられる.

5.4.3 乳幼児の身体拘束条件

本モデルでは, 親の声 (模倣音), 乳幼児自身の声 (発話音), 最近傍母音カテゴリ (学習ベクトル) の三者間の距離関係から母音カテゴリの生成を行うかどうかの判定を行っている. また, この実験においては発話音を成人の平均フォルマントの中心点から 200[Hz] の円内からランダムに発声させた音声として与えている. そのため, 発話音に対して割り当てた最大距離円内に入らないかつ親の音声が極端に近い状態で模倣が行われてしまった場合にも母音カテゴリが生成されてしまう. しかし, 乳幼児音声は, 生後まもない頃では咽頭, 下顎等を上手く使うことができず, 意図する音声を発話することが難しいと考えられる. そのため, 親が模倣した音声に対して乳幼児が意図する発話音は高頻度である 1 点に集中しがちになる [12]. また, 成長するにつれて咽頭の急速な広がり・喉頭の急速な低下, 下顎等の発達に伴う口の可動範囲の拡大など発声器官の発達によって母音カテゴリの修正および発声領域の拡大を行っていると考えられる. 生後まもない乳幼児の限定された発声領域は親のどの母音カテゴリに対しても重複して認識されるような領域が存在する. したがって, その条件を本モデルに適応する場合には図

6 のように方向を考慮することでも同様に母音カテゴリが生成されると思われる.

6. おわりに

本稿では, 獲得した母音カテゴリは乳幼児が発話する母音カテゴリではなく, その前段階として, 乳幼児が自身の非常に優れた聴覚を利用し, 親の音声から母音カテゴリを形成することで, 親の母音カテゴリ獲得を行っているのではないかという考えから母音カテゴリの獲得を行うモデルを提案し, 検証を行った. 結果として, 母音獲得においては聴覚における自分以外の他者の母音カテゴリを認識することで, 乳幼児はその獲得した母音カテゴリを自身の声に対する母音カテゴリに対応づけていることが考えられる. また, 乳幼児の音声知覚および言語獲得においては乳幼児自身の身体的特徴とその発達を更に考慮することでモデルの表現能力の向上の可能性が考えられる.

参考文献

[1] 神田尚, 尾形哲也, 高橋徹, 駒谷和憲, 奥乃博, “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,” 第 71 回情報処理学会講演論文集, No.2,pp.133–134, 2009.

[2] 浅田 稔, “認知発達ロボティクスによる赤ちゃん学の試み,” ベビーサイエンス, Vol.4, pp.2–27, 2004.

[3] ジャック・ライアルズ, 音声知覚の基礎, KAIBUNDO, pp.39–110,2003.

[4] Masataka,N. and Bloom,K., “Accoustic properties that determine adult's preference for 3-month-old infant vocalization,” Infant Behavior and Development, Vol.17, pp.461–464, 1994.

[5] Pelaez-Nogueras,M. and Gewirtz,J.L. and Markham,M.M., “Infant vocalizations are conditioned both by maternal imitation and motherese speech,” Infant behavior and development, Vol.19, p.670, 1996.

[6] Eimas,P.D. and Siqueland,E.R. and Jusczyk,P. and Vigorito,J., “Speech Perception in infants,” Science, Vol.171, pp303–306, 1971.

[7] 吉川雄一郎, 三浦勝司, 浅田稔, “教示者の無意識的引き込み模倣に基づく母音カテゴリの発見,” ロボティクス・メカトロニクス講演会 2007, Vol.CD-ROM, 1A2-L07, 2007.

[8] 須藤珠水, 茂木健一郎, “言語獲得期における語意学習とカテゴリー認知のメカニズム”, 信学技報 SIS2004-4 Vol.104 No.144, 電子情報通信学会, pp.17–22, 2004

[9] 宮本定明, クラスター分析入門 - ファジィクラスタリングの理論と応用, 森北出版, 1999.

[10] The Snack Sound Toolkit http://www.speech.kth.se/snack/

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

[12] 石塚健太郎, 麦谷綾子, 天野成昭, “乳幼児の母音に対する周波数ピークの縦断的分析,” 日本音響学会講演論文集, 2-2-7, pp.335–336, 2005.