

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

Finding the Correspondence of Caregiver's Vowel Categories Based on Unconscious Anchoring in Maternal Imitation

学 三浦 勝司 (科学技術振興機構 ERATO, 阪大院)
正 吉川 雄一郎 (科学技術振興機構 ERATO)
正 浅田 稔 (科学技術振興機構 ERATO, 阪大院)

Katsushi Miura¹⁾²⁾, Yuichiro Yoshikawa¹⁾, and Minoru Asada¹⁾²⁾

¹⁾Asada Synergistic Intelligence Project, ERATO, JST (www.jeap.org)

²⁾Graduate School of Engineering, Osaka University
{miura,asada,yoshikawa}@jeap.org

It is a formidable issue how robots can show the behaviors corresponding to human's ones since the dissimilar body. As a simple case for such correspondence problem, this paper presents a robot that acquires vowels through the interaction with its caregiver. We focus on the roles of maternal imitation since it could instruct the correspondence of the sounds. Furthermore, we suppose that it causes "unconscious anchoring" in which the caregiver's imitated voice is performed unconsciously, closely to one of his/her own vowels, and helps to leading robot's utterances to be more vowel-like. We experimented under the assumptions in which the robot knows the caregiver's vowels and the rough estimate of mapping between the caregiver's utterance and the robot's one. Through the experiments with a caregiver, we show that a robot acquires more vowel-like utterances than one without such a caregiver even when it is given different types of mapping functions.

Key Words: maternal imitation, unconscious anchoring, dissimilar body

1 緒言

ヒューマノイドロボットは音声のような、人同士のコミュニケーションで用いられる手段を使って人とコミュニケーションを実現することが期待される。しかし、ロボットと人では身体構造が異なるため、人のコミュニケーション行動をロボットがそのまま再現することは一般に容易ではない。一方、人の乳児は親と同じ発声ができないにもかかわらず、親とのインタラクションを通じて言語を獲得していると言われている。このような母子間インタラクションのように、人とのインタラクションを利用してロボットに言語を獲得させる試みは適応的なロボットの実現だけでなく乳児の発達過程の理解を目指す認知発達ロボティクスの観点からも非常に興味深い問題である。

言語獲得の初歩に母音の獲得がある。母音獲得の従来研究として、複数の発話エージェントが知覚を自己組織化することによって共通の母音を獲得するモデルが提案されている^{1, 2)}。しかし、これらの研究ではエージェント同士が同じ構音能力を持っていると仮定しており、エージェント同士で構音能力が異なることがエージェントの母音獲得にどのような影響を与えるかは検討されていない。

人工声帯を備えることで実際に発声可能なロボット(例^{3, 4)})を使用することは構音能力の異なる他者同士での模倣の問題を解決するための一つの方法である。このような発話ロボットを用い、Yoshikawa et al.⁵⁾は乳児の母音獲得に関する母子間インタラクションモデルを提案した。Yoshikawa et al.のモデルでは、母親の模倣が乳児の発声を促す⁶⁾、乳児の母音様のクレーニングが母親の模倣を促す⁷⁾という2つの観測結果に基づいている。そして、ロボットの発声に対し教示者が母親的模倣を行うことで、ロボットに母音を獲得させることが可能であることを示した。しかし、このモデルではロボットは多くの母音の候補を獲得することができたが、どれが最も母音らしい、人にとって認識しやすい発声かを選定すること

はできなかった。

我々は教示者による母親的模倣がロボットの発声をただ真似するだけでなく、ロボットの発声を母音らしいものへと引き付ける力があるのではないかと考えている。これまで我々は教示者がロボットの発声を模倣することによってロボットが明瞭な母音を獲得できることを示してきた。しかし、前研究⁸⁾では教示者の発声とロボットの発声との対応関係をロボットに与えていたが、構音能力が異なるもの同士の模倣において正確な対応関係を事前にロボットに与えることは困難である。そこで本研究では、ロボットに大まかな対応関係を与えた場合でも教示者の模倣によってロボットに明瞭な母音を獲得させることができることを示す。

ロボットの母音学習にはロボットの発声に対する教示者の母親的模倣を利用する。このとき、教示者はロボットと構音能力が異なるためロボットの発声をそのまま再現することはできない。このような場合、教示者の模倣は無意識のうち自身に自身の母音に引き込まれた発声となると考えられる。そのため、ロボットは教示者の母親的模倣によってロボット自身の発声と教示者の発声との対応関係を修正し、教示者の無意識的引き込みによってロボットの発声は徐々に母音らしい発声へと導かれると考えられる。次節では、この無意識的引き込みの概念を利用したロボットの母音獲得メカニズムについて説明する。その後、ロボットの学習メカニズム、実験設定、実験結果について順に述べていく。

2 無意識的引き込みを利用した母音獲得

教示者とロボットによるインタラクションの概要を Fig. 1 に示す。インタラクションでは、

R: ロボットは日本語の母音を発声しようと試み、

C: 教示者はロボットの発声を模倣しようと試みる。

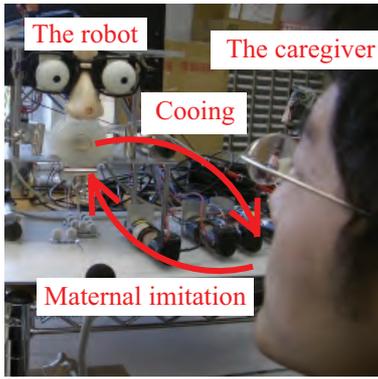


Fig.1 An interaction model between an imitative caregiver and a vocal robot

ロボットは教示者の模倣を聞くことで、自身の発声が教示者にどのように聞こえたかを予測する。

ロボットはインタラクションを通じて教示者と共通の母音を学習する。このとき、ロボットと人の構音機構が異なるため、互いに相手と同じ音響的特徴を持つ発声を実現することはできない。にもかかわらず、人はロボットの発声と自身の発声に対応付けて模倣することが可能である⁵⁾。このような対応付けをロボットに正確に与えることは非常に困難なことである。そこで我々は教示者の発声とロボットの発声との対応関係を表す粗い変換式をロボットに与え、ロボットと人とのインタラクションを実現した。ここでは本研究ではロボットに獲得目標とする母音のフォルマント¹と、それに対応する教示者の母音のフォルマントをあらかじめ与え、インタラクションを利用してロボットの目標母音のフォルマントを更新することとした。

以上の設定より、ロボットは教示者の模倣に対して粗い変換式を用いることで、ロボット自身の発声と目標母音との差がどの程度かを知ることができる。このときの人の模倣は構音機構の違いによってロボットと同じ発声ができないため、無意識のうちに自身の持つ知覚のカテゴリーや身体の拘束によって自分の母音に似た音を発声してしまうと考えられる。ロボットがこのような実際よりも母音に近い模倣を参照することにより、ロボットの発声も母音らしい音に徐々に引き寄せられていくと考えられる。我々はこのような人の模倣の特性が音声だけに限らず視覚等の他のモダリティでも起こりうると考えており、ロボットに行動を獲得させるための新たな手法として期待している。

3 学習メカニズム

ロボットは教示者の母親的模倣を通して母音を学習する。学習過程において、フォルマントで定義されたロボットの目標母音は教示者とのインタラクションを通じて更新される。本節ではロボットに与える粗い変換式とロボットの目標母音の更新方法について説明する。

3.1 変換式

人の母音は母音判別によく利用されるフォルマントによって分類される。Fig. 2 は男女1名ずつが母音を発声したときのフォルマントを示しており、日本語母音がフォルマント空間上で五角形を形成していることがわかる。

我々はフォルマント空間上の五角形が母音の獲得に興味を持つと考え、ロボットが構音可能な領域に教示者と

¹人の母音はフォルマントと呼ばれる音響特徴量によって分類可能である⁹⁾とされている。

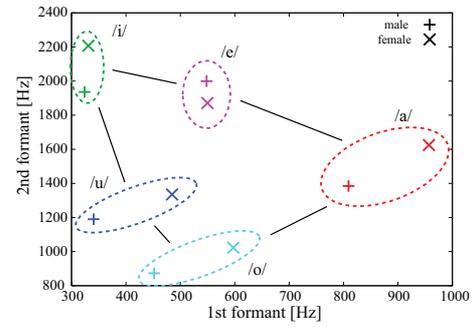


Fig.2 A sample distribution of human vowels in the formant space

相似な五角形を作成し、ロボットの目標母音に設定した。そのため、人の発声をロボットの構音可能な領域へと対応付ける変換式は線形変換となる。つまり、教示者の母音 $h^{/v/}$ ($/v/ = /a/, /i/, /u/, /e/, \text{ or } /o/$) はスケール係数 α 、回転角 θ の回転行列 $R(\theta)$ 、補正值 s を持つ変換式 g によって $h^{/v/}$ へと変換される。

$$h^{/v/} = g(h^{/v/}; \alpha, \theta, s) \equiv r_c + \alpha R(\theta)(h^{/v/} - h_c) + s \quad (1)$$

ここで h_c, r_c はそれぞれ教示者とロボットの構音可能領域の重心である。

3.2 ロボットの目標母音の更新

次にロボットの目標母音の更新方法を説明する。まず初めに、ロボットは $r_d^{/v/}$ を発声する。ただし、 $/v/$ はロボットが発声を試みた母音を示す。この $r_d^{/v/}$ に対し人は模倣音 h を返す。ロボットは人の模倣 h と人の母音 $h^{/v/}$ との差に基づき目標母音 $r_d^{/v/}$ を更新する。この過程を以下に式で表す。

1. k 回目のインタラクションで、ロボットはある母音 $/v/$ に対応する $r_d^{/v/}(k)$ を発声する。
2. 教示者は模倣音 $h(k)$ を発声する。
3. 教示者の模倣と母音のベクトルの差 $\Delta h = h^{/v/} - h(k)$ を変換式 g を用いてロボットの構音可能な領域上に変換する。変換されたベクトルの差を用いて次式のように $r_d^{/v/}(k)$ を更新する。

$$r_d^{/v/}(k+1) = r_d^{/v/}(k) + g(\Delta h) \quad (2)$$

Fig. 3 はこの更新過程の概要を示している。

4. $k+1$ 回目のインタラクションで、ロボットが $r_d^{/v/}(k+1)$ を発声する。

4 実験

教示者による母親的模倣によってロボットが母音らしい発声を獲得することを示すための実験を行った。実験には Fig. 4 (a) ~ (d) に示す 4 種類の変換式を用いた。

4.1 発話ロボット

Fig. 5 は実験で使用する発話ロボットの発話メカニズムを示している。ロボットの音源はコンプレッサから流れ込んだ空気がロボットの人工声帯を振動させることで生成される。この音源が中空のシリコン製声道を通過することで、声道形状に合わせた共鳴周波数を持ったロボットの発声となる。この声道と唇はそれぞれ 4 つずつのモー

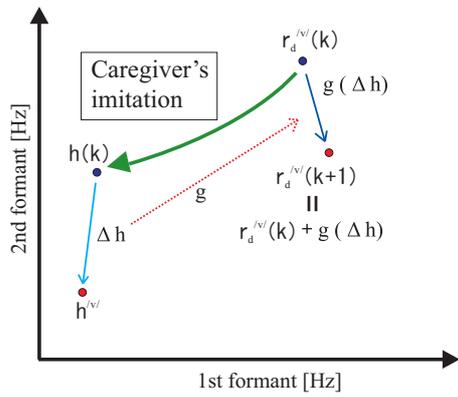


Fig.3 Updating process of the prototype vector of a vowel (/v/) category $r_d^{(v)}$

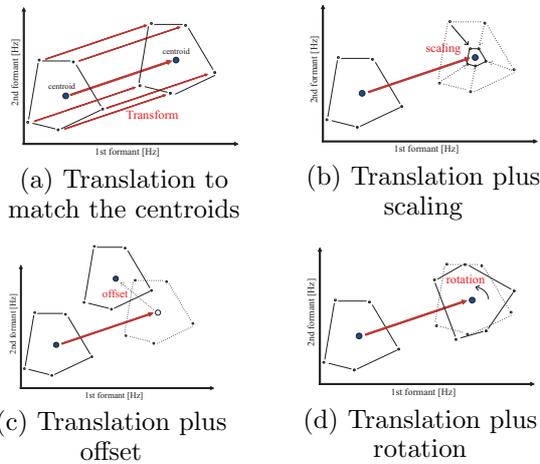


Fig.4 Four examples of the mapping functions from the region of generable sounds by the caregiver to one by the robot

タで制御されている。ただし、唇は鉛直・水平方向の開閉のみであるため、ロボットの持つ自由度は6である。

このロボットのモータ出力をそれぞれ0~1までの5段階に量子化し、構音能力を調べた。その結果がFig. 6である。ただし、ロボットの唇の形状を人の母音発声時の形状に似せて発声しており (Table 1 参照), Fig. 6の各色は母音の種類を示している。母音獲得の実験では、ロボットはFig. 6のデータを用いて目的の音を発声する。

4.2 実験設定

全実験で教示者は同一人物とし、4つの変換式を用いてロボットに母音を学習させた。実験中のインタラクションの回数は各母音ごとに順番に20回ずつとし、教示者には第3者が聞いた時にロボットの声を真似していると判断できるように模倣させた。また、各実験の試行回数は5回ずつとし、解析には獲得した母音の平均値を用いた。そして、母親的模倣の効果を検証するため、学習開始時に

Table 1 The motor commands to form the lip shapes to resemble human's ones in vocalizing vowels

Motor output	/a/	/i/	/u/	/e/	/o/
vertical direction	1.0	0.0	0.0	0.5	0.5
horizontal direction	1.0	1.0	0.0	1.0	0.0

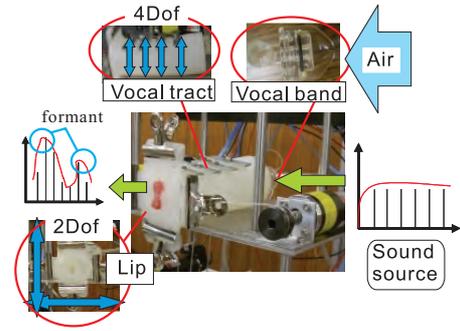


Fig.5 The articulatory system of the vocal robot

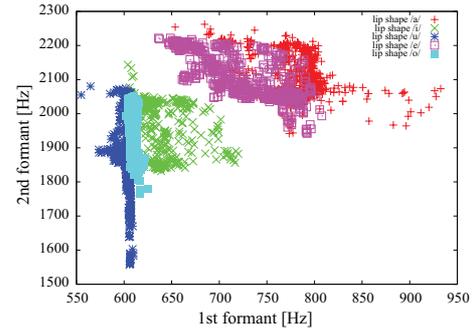


Fig.6 The distribution of the robot utterances in the formant space

与えたロボットの目標母音に最も近い音をFig. 6のデータから選び、ロボットに母音として獲得させた場合との比較を行った。以下に4つの変換式を説明する。

平行移動：教示者の構音領域の重心からロボットの構音領域の重心へ五角形を平行移動： $\alpha=1.0, \mathbf{R}(0), s=(0, 0)$ (Fig. 4 (a) 参照)。

サイズ：平行移動とサイズ変更： $\alpha=0.24, \mathbf{R}(0), s=(0, 0)$ (Fig. 4 (b) 参照)。

補正：平行移動と補正： $\alpha=1.0, \mathbf{R}(0), s=(-100, 200)$ (Fig. 4 (c) 参照)。

回転：平行移動と回転： $\alpha=1.0, \mathbf{R}(30), s=(0, 0)$ (Fig. 4 (d) 参照)。

4.3 実験結果

フォルマント空間上で獲得した母音の母音らしさを判別することは困難であるため、2種類の実験方法でロボットが獲得した母音の母音らしさを15人の被験者に比較させた。比較実験では、ロボットに同じ変換式で獲得した母音を「あいうえお」の順番で連続して発声させた。ただし、どちらの実験方法で獲得した母音を先に聞かせるかはランダムである。被験者には2つの「あいうえお」を聞き比べさせ、日本語母音として聞き取りやすかった方の発声を選択させた。

Fig. 7は15人の被験者が教示者による母親的模倣あり(左側の赤い棒グラフ)またはなし(右側の青い棒グラフ)で獲得したロボットの母音が聞き取りやすいと判断した割合の平均値である。この結果に対し、母親的模倣によって獲得した母音が母音らしいと判断した割合がチャンスレベルと比較して有意な差があるかどうかをT検定によって判定した。その結果、変換式が平行移動の

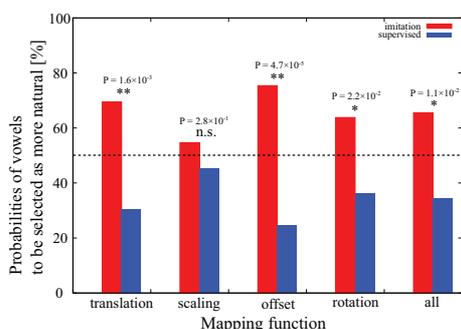


Fig.7 The probabilities of vowels to be selected as more natural: the red, left-side bars indicates the probabilities of ones acquired by the proposed imitative learning while the blue, right-side bars indicates ones acquired by the supervised learning with four types of rough estimation of the mapping

とき ($p = 1.6 \times 10^{-3}$), 補正のとき ($p = 4.7 \times 10^{-5}$), 回転のとき ($p = 2.2 \times 10^{-2}$)となり, 有意な差が得られた. 変換式がサイズのときは ($p = 2.8 \times 10^{-1}$)で有意な差は得られなかったが, 全体の平均で ($p = 1.1 \times 10^{-2}$)と有意な差が得られた. このことから, 教示者の母親的模倣によってロボットは母音らしい発声を獲得できたといえる.

5 結論および討論

本稿では, 我々は教示者の無意識的引き込み模倣のアイデアに基づいたロボットの母音学習手法を提案した. そして, 実験により教示者の模倣によってロボットが母音らしい発声を獲得することができることを示した. この無意識的引き込みの利用はロボットに人間らしい行動を獲得させるための新しい手法として期待されるため, 今後音声以外のモダリティについても同様のことが示されるべきである.

模倣における基本的な問題の一つに他者の行動を自身の行動に変換するための変換式の設定があり, 本研究ではその変換式にフォルマント空間上でのアフィン変換を使用している. 実験で使用したパラメータは4種類だけであり, 他のパラメータについて検討していない. そのため, あまりにも変換式のパラメータが間違っている場合, 学習に不具合が生じる可能性はある. しかし, 母音カテゴリの対応付けが干渉しあわない限り, どのようなパラメータを選んでも働く我々は考えており, どのようにしてこれを保証するかは今後の課題と考えている.

本研究では, 変換式のパラメータが正確でないにもかかわらず実験中に修正されることはない. それゆえ, 目標母音の更新と同時にこの変換式のパラメータを修正することは今後の課題である. 乳児はクーイングを行う時期にこの変換式の修正を行っていると考えられており, この推論を支持するような従来研究^{10, 11)}も存在している. しかし, 乳児は単純に母音のみを学習しているのではなく連続発声や子音なども同時に学習している. さらに, 母親も単純に乳児の発声を模倣するだけではない. このような環境でどのように我々の研究を進展させていくかは非常に興味深い課題である.

無意識的引き込みは人が模倣するときの一般的な行動であり, 本研究の枠組みは視覚等の他のモダリティにも適用できると我々は考えている. この無意識的引き込みが一般的なものであることを示すことがもう一つの我々の目的である. また, マルチモーダルな感覚では, マガーク効果¹²⁾などのように互いの感覚が干渉しあう事が起

きるため, 視覚と音声の双方の無意識的引き込みが影響する. 本研究ではこれらの影響を分離することはできていないが, これらの影響を分離して検証することは重要な課題であると考えている.

参考文献

- [1] B. de Boer. Self-organization in vowel systems. *Journal of Phonetics*, Vol. 28, pp. 441–465, 2000.
- [2] P.-Y. Oudeyer. Phonemic coding might result from sensory-motor coupling dynamics. In *Proceedings of the 7th international conference on simulation of adaptive behavior (SAB02)*, pp. 406–416, 2002.
- [3] Kotaro Fukui, Kazufumi Nishikawa, Shunsuke Ikeo, Eiji Shintaku, Kentaro Takada, Hideaki Takanobu, Masaaki Honda, and Atsuo Takanishi. Development of a talking robot with vocal cords and lips having human-like biological structures. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2526–2531, Augst 2005.
- [4] T. Higashimoto and H. Sawada. Speech production by a mechanical model construction of a vocal tract and its control by neural network. In *Proc. of the 2002 IEEE Intl. Conf. on Robotics & Automation*, pp. 3858–3863, 2002.
- [5] Yuichiro Yoshikawa, Minoru Asada, Koh Hosoda, and Junpei Koga. A constructivist approach to infants' vowel acquisition through mother-infant interaction. *Connection Science*, Vol. 15, No. 4, pp. 245–258, Dec 2003.
- [6] M. Peláez-Nogueras, J. L. Gewirtz, and M. M. Markham. Infant vocalizations are conditioned both by maternal imitation and motherese speech. *Infant behavior and development*, Vol. 19, p. 670, 1996.
- [7] N. Masataka and K. Bloom. Acoustic properties that determine adult's preference for 3-month-old infant vocalization. *Infant Behavior and Development*, Vol. 17, pp. 461–464, 1994.
- [8] Katsushi Miura, Minoru Asada, Koh Hosoda, and Yuichiro Yoshikawa. Vowel acquisition based on visual and auditory mutual imitation in mother-infant interaction. In *The 5th International Conference on Development and Learning (ICDL'06)*, 2006.
- [9] R. K. Potter and J. C. Steinberg. Toward the specification of speech. *Journal of the Acoustical Society of America*, Vol. 22, pp. 807–820, 1950.
- [10] Anthony J. DeCasper and Melanie J. Spence. Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, Vol. 9, pp. 133–150, 1986.
- [11] Patricia K. Kuhl. Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, Vol. 66, pp. 1668–1679, 1979.
- [12] Harry McGurk and Jrohn MacDonald. Hearing lips and seeing voices. *Nature*, Vol. 72, pp. 746–748, 1976.