# 音声相互模倣過程を収束に導くマグネット効果

三浦勝司 [1)2)], 吉川雄一郎 [1)] and 浅田稔 [1)2)]

Katsushi MIURA[1)2)], Yuichiro Yoshikawa[1)] and Minoru ASADA[1)2)]

[1)] 科学技術振興機構 ERATO 浅田共創知能システムプロジェクト

[2)] 大阪大学大学院　工学研究科

[1)]Asada Synergistic Intelligence Project, ERATO, JST

[2)]Graduate School of Eng., Osaka Univ., 2-1 Yamada-oka, Suita, Osaka 565-0871 Japan

{miura,yoshikawa,asada}@jeap.org

## Abstract

Human-robot communication is expected to be realized by the usual means for human-human communication. However, it is difficult for the robot to directly copy the human's means due to the difference of bodies. As argued in the issue of imitation with dissimilar bodies, what kinds of representation could correspond between human and robots is one of fundamental issues. The previous work [1] has hypothesized that mutual imitation of voice between the robot and the caregiver leads robot vowels to be more natural but the underlying mechanism has not been deeply argued. This paper focuses on two types of the magnet effects, the perceptual magnet effect and what we call *the articulatory magnet effect* as the underlying mechanism of mutual imitation. Toward the design principle of the robot behavior through mutual-imitation, we examine these magnet effects in the experiment of imitation of the vocalizing robot with human subjects.

## 1　はじめに

ヒューマノイドロボットは音声などのように，人が人とコミュニケーションをする際に用いる様式を用いて，人との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，ロボットと人の身体構造や運動能力は異なるため，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困難である．従ってロボットの行動は，それが人にどのように解釈されるかを考慮して構成されるべきである．一方，人の乳児は未発達のため，親の発声をそっくりそのままコピーすることは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音韻様の発声を獲得することができるが，その発達メカニズムは明らかでない．従って，この乳児の発達メカニズムをモデル化 [2] することは，母音を発声できるロボットの実現だけでなく，人の乳児の言語獲得に至る認知発達過程の理解にも関連した非常に興味深い課題といえる．

インタラクションを通じた母音の獲得の従来研究として，複数の発話エージェントが知覚の自己組織化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 [3, 4]．しかし，これらの研究ではエージェント同士が同じ身体構造を持つことが仮定されており，乳児と母親のような身体構造が異なるエージェント同士がどのように母音を共有するかについては扱われていない．

身体構造の異なるエージェント同士が母音を共有する問題を扱った従来研究に，母親の模倣が乳児の発声を促し [5]，乳児の母音様の発声が母親の模倣を促す [6] という 2 つの知見に基づく母子間インタラクションモデルがある [7]．この研究では，発話ロボットの発声を教示者が母音でオウム返しすることで，ロボットが母音を獲得可能であることを示した．さらに，Miura et al. [1] は人が発声困難な音を模倣するとき，実際に模倣で返すべき音よりも無意識のうちに自身の母音よりの発声を返すとの仮説を基に，ロボットと人が互いの発声を模倣し合うことでロボットの発声を明瞭な母音に導くことができることを示した．しかし，人がなぜ無意識のうちに自身の母音よりの発声を返すのかについて十分に議論されていなかった．

人は自身の知覚する音を実際の音よりも自身の言語環境における特徴的な音素である母音や子音に似た音として知覚することが知られている．この現象は知覚のマグネット効果と呼ばれている．本論文では，相互模倣において人が無意識的に自身の母音よりの発声を返してしまう原因として，この知覚のマグネット効果に加え，我々が構音のマグネット効果と呼ぶ現象に注目する．これは人の発声が実際に出そうとした発声よりも自身の構音機構の制御や運動能力によって母音や子音に近い発声になる現

象である．

この構音のマグネット効果を示すための実験として，ロボットの発声を聞いたときに，人はどのように知覚し，模倣するかを検証する 2 種類の実験を行う．一つはロボットの発声を被験者が模倣する実験であり，もう一つはロボットの発声を日本語 5 母音のどれに聞こえたか被験者が判定する実験である．次節では相互模倣におけるマグネット効果についての仮説を説明する．そして，本研究で使用する発話ロボットについて紹介した後，実験および実験結果の解析について述べる．

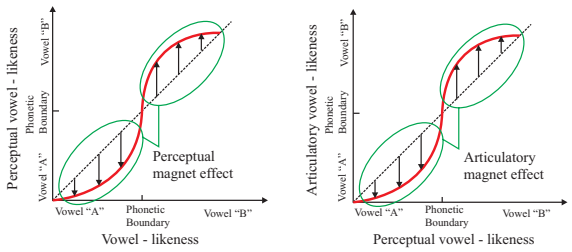
## 2　相互模倣におけるマグネット効果の仮説

発話の相互模倣を扱った先行研究 [1] では，人が模倣困難なロボットの発声を模倣しようとすると，無意識のうちに実際に発声すべき音よりも自身の母音に似た発声をしてしまうとの仮説が立てられている．そして，教示者と発話ロボットとの相互模倣を通じて発話ロボットに母音を獲得させることにより，仮説どおり教示者が母音よりの模倣をすることを示した．しかし，なぜ教示者が母音様の音声で模倣してしまうのかについての議論は尽くされていない．本研究では，この無意識的な母音様の模倣二間して知覚のマグネット効果と我々の提案する構音のマグネット効果の観点から考える．

知覚のマグネット効果とは，人の知覚する音が実際の音よりも自身の言語環境における特徴的な音素である母音や子音に似た音として知覚される現象である [8] (Figure 1 (a) 参照)．人はこの知覚のマグネット効果によって，他者の母音様の発声を実際よりも母音に近い音として知覚するため，模倣音声も実際より母音に近い音になる (Figure 1 (b) 参照)．一方，我々が構音のマグネット効果と呼ぶもうひとつのマグネット効果とは，人の発声する音が実際に知覚した音よりも自身の構音機構の制御や運動能力の拘束によって無意識のうちに自身の母音や子音に近い音になる現象である．従って，人は他者の発声を模倣するときに知覚と構音の 2 つのマグネット効果によって無意識的に自身の母音や子音に似た音を発声することになるため，相互模倣によってロボットの発声を人の母音に導くことが可能であると考えられる．

## 3　発話ロボット

本研究で用いる発話ロボット (Figure 2) は先行研究[7] [9] に習い，ソースフィルタ理論[10] に基づいて設計されている．この発話ロボットはエアコンプレッサ，人工声帯，声道，唇を備えており，母子間インタラクションのモデル化を目的として作成されている (Figure 3 参照)．コンプレッサから供給された空気はチューブを通して人工声帯を振動させ，音源になる．生成された音源は中空のシリコン製声道内を通過することで，声道形状に応じた共鳴周波数




(a) Perperception　　(b) Articulation

Figure 1: The shift of perceptual/articulatory vowel-likeness by the perceptual/articulatory magnet effect

を持つ音声として産出される．

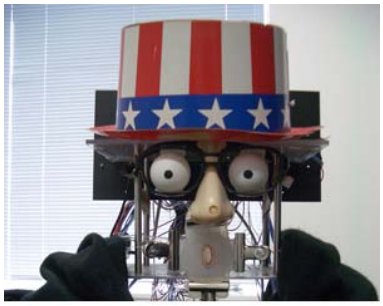


Figure 2: The vocalizing robot

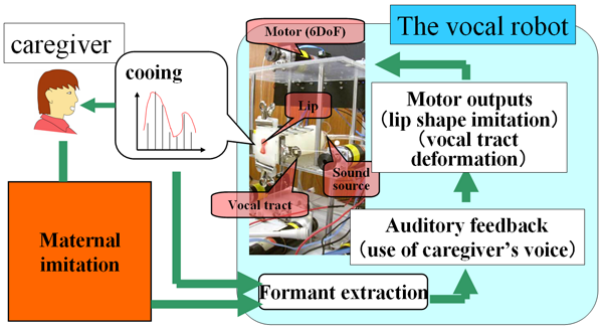



Figure 3: Vocalizing robot to model mother-infant interaction

ロボットは声道形状と唇形状をそれぞれ 4 つのモータを用いて変形させることにより，生成される音声の共鳴周波数を制御可能である．空気の流れはバルブの開閉でコントロールされており，モータとバルブのコントローラはホストコンピュータからの指令によって制御されている．ホストコンピュータはマイクから信号を受け取り，母音認識の特徴量でよく知られるフォルマント [11] を抽出する．

### 3.1 構音能力

声道形状の変形に用いる 4 つのモータの出力をそれぞれ 0.0(無変形) から 1.0(変形量最大) までの 5 段階に量子化し，唇の形状を人の母音 /a/，/u/，/e/ の発声時の唇の形状に似せた 3 種類に設定した．従って，ロボットが取りうる声道部と口唇部の形状は全部で 1875 通り ($3 \times 5^4$) である．ここで /i/，/o/ の口唇形状を除外したのは，それぞれ/e/，/u/ の口唇形状との間にフォルマントの分布の差がなかったためである．

Figure 4 は横軸を第 1 フォルマント，縦軸を第 2 フォルマントとするフォルマント空間上にロボットが 1 秒間発声したフォルマントの平均値をプロットした結果である．以降フォルマント空間上の位置ベクトルをフォルマントベクトルと呼ぶ．また，人とロボットのフォルマントの分布を比較するため，日本人の男女 7 人が発声した日本語母音 /a/, /i/, /u/, /e/, /o/ のフォルマントの平均も Figure4 に示す．Figure 4 から，ロボットのフォルマントと人のフォルマントの分布は重なりあっていないことがわかる．すなわち，ロボットも人も互いの発声のフォルマントベクトルを再現することはできないことを示しているといえる．

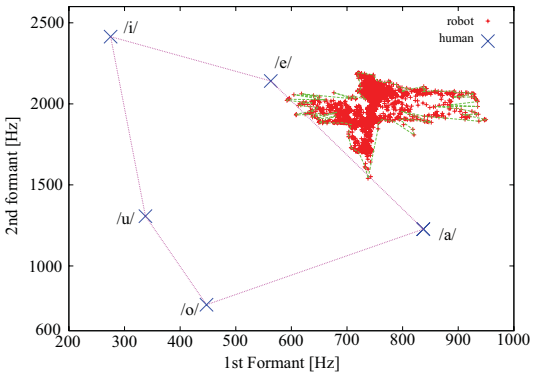
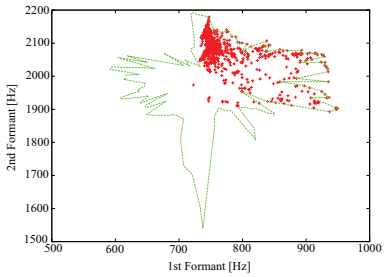
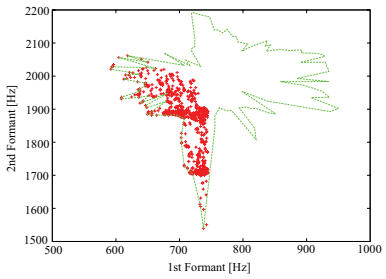



Figure 4: The distributions of the 1st and the 2nd formants of utterances by a human and the robot.

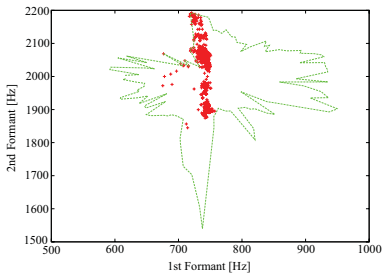
Figure 5 は母音 /a/, /u/, /e/ を発声するときの人の唇の形状 (Figure 5 (d),(e),(f)) とその母音に対応するロボットの唇の形状 (Figure 5 (g),(h),(i)) を示してある．Table 1 はそのときの唇のモータ出力である．さらに，その対応する唇の形状でロボットが発声したときのフォルマントの分布 (Figure 5 (a),(b),(c)) が示されているおり，口唇形状の違いによって，ロボットが発声するフォルマントの分布領域は異なることがわかる．この分布位置は人の母音のフォルマントの相対的位置関係 (Figure4 参照) に似ており，ロボットに人の口唇形状を模倣させることによって，口唇形状に対応した母音をロボットに発声させやすくなると考えられる．



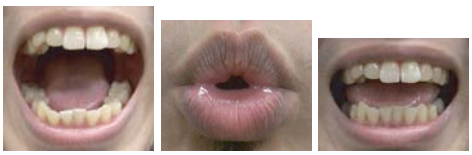

(a) Formant distribution for the lip shape /a/




(b) Formant distribution for the lip shape /u/




(c) Formant distribution for the lip shape /e/



(d) /a/ (e) /u/ (f) /e/

(g) /a/ (h) /u/ (i) /e/

Figure 5: Formant distribution of the utterances, the lip shape of the human model and the mapped shape onto the robot lip for three vowels, /a/, /u/, /e/

Table 1: Relation between robot's lip shape and motor outputs

| Deformation | /a/ | /u/ | /e/ |
|---|---|---|---|
| vertical direction | 1.0 | 0.0 | 0.0 |
| horizontal direction | 1.0 | 0.0 | 1.0 |

## 4 実験

マグネット効果が人の知覚や模倣にどのような影響を与えるかを検証するために 2 種類の実験を行った．一つ目の実験では，目隠しをした被験者にロボットの発声を模倣させることで，知覚と構音の 2 つマグネット効果の影響について調べる．もうひとつの実験では，目隠しをした被験者にロボットの発声が日本語の 5 母音のうちのどれに聞こえたか判定させ，そのときのロボットの発声の母音らしさを評価させることで，知覚のマグネット効果の影響について調べる．実験の被験者には大学院在学中の 10 人の被験者を選び，実験の試行順は被験者ごとに入れ替えた．この 2 つの実験の結果を比較することで，相互模倣インタラクションにおけるマグネット効果の影響について議論する．

### 4.1 刺激

２章で議論したマグネット効果から，模倣者が普段発声する母音にどの程度近い音声で模倣するかは呈示される音声の母音らしさの程度のシグモイド関数で近似可能であることが予想される (Figure 1 参照)．そこで，母音らしさの程度が異なる様々なロボットの発声を被験者に模倣させ，これらの関係を観察することを考える．

ただし，ロボットが発声可能な音の中で，どのようなフォルマントベクトルを持つ音が最も母音らしいかは不明であるため，最も母音らしい音に対応するフォルマントベクトル $\boldsymbol{r}^{/v/}$ を以下のように操作的に定義する．すなわち，日本人の男女 7 人が発声した日本語母音のフォルマントの平均 $\boldsymbol{h}^{/v/}$ とその日本語の 5 母音のフォルマント空間上での重心 $\boldsymbol{h}_c$ を用いて $\boldsymbol{r}^{/v/} = (\boldsymbol{h}^{/v/} - \boldsymbol{h}_c) + \boldsymbol{r}_c$ のように与える (Figure 6 参照)．ここで $\boldsymbol{r}_c$ はロボットが構音可能なフォルマントの分布 (Figure 4 参照) の重心である．

そして，各母音についてロボットが発声する母音らしさを変えた 5 つのフォルマントベクトル $\boldsymbol{r}_i^{/v/} (i = 1, \cdots, 5)$ をロボットが構音可能なフォルマントの分布の重心 $\boldsymbol{r}_c$ と最も母音らしい発声 $\boldsymbol{r}^{/v/}$ を用いて次式のように定めた．

$$\boldsymbol{r}_i^{/v/} = \boldsymbol{r}_c + \frac{i}{5}\alpha(\boldsymbol{r}^{/v/} - \boldsymbol{r}_c), \quad (i = 1, \cdots, 5), \qquad (1)$$

ここで $\alpha$ は $\boldsymbol{r}_i^{/v/}$ がロボットの構音可能な領域内に収まるようにするためのスケーリング係数である．これにより，実験では，各母音につき 5 通り，合計 15 通りのロボットの発声が模倣の対象として被験者に提示される．

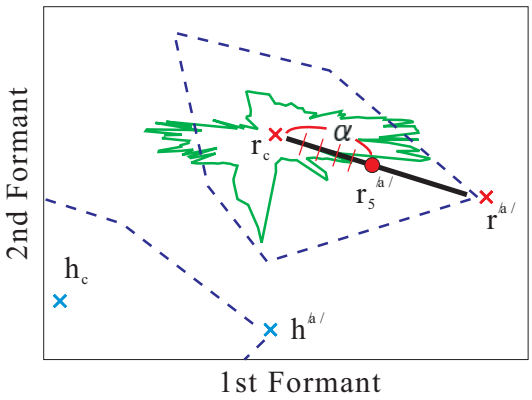



Figure 6: The formant vectors of the most "vowel-like" sound and the test sounds in the case of a vowel /a/. Note that this figure is schematic.

ロボットには実験の各試行ごとに 2 つの音を選択させる．ただし，一つ目の音に $\boldsymbol{r}_i^{/v/}$ が選ばれた場合，連続して発声される次の音は $\boldsymbol{r}_i^{/v'/}, (/v'/ \neq /v/)$ となるように選択させる．従って，発声される連続音の組み合わせの場合の数は，母音の選び方 ${}_3P_2$ 通りに母音らしさの選び方 ${}_5P_1$ 通りで合計 30 通りとなる．それぞれの音の組み合わせはランダムで 1 度だけ被験者に聞かされる．

発声する音が決まった後，ロボットはあらかじめ準備しておいたフォルマントと唇や声道形状を決定するモータ出力とのマッピングを利用して発声する．

### 4.2 実験手順

**音声模倣実験**　被験者に"ロボットの発声に対して第 3 者が同じ音だと知覚できるように模倣してください"と説明した後，被験者にロボットの 2 つの母音の連続発声を聞かせ，それぞれの音を模倣させた．これを 1 試行とし全部で 30 試行繰り返した．また，模倣実験の開始前に被験者に日本語の 5 母音を発声させ，フォルマントベクトルを抽出した．この抽出した日本語の 5 母音のフォルマントと被験者が模倣発声したフォルマントベクトルを比較することで，模倣時の知覚のマグネット効果と構音のマグネット効果を調べた．

**母音判別実験**　被験者に目隠しした状態でロボットの 2 つの母音の連続発声を聞かせ，日本語の 5 母音のうちのどれに聞こえたかと，そのときの自身の判断に対する自信の度合いを 5 段階で評価させた．そのときの評価値は '1' (適当に選んだ), '3' (なんとなくそう聞こえた), '5' (確信を持ってその母音だといえる) であり，'2'，'4' はそれぞれの中間の値である．これを 1 試行とし 30 試行繰り返した．そして，ロボットが発声した母音と被験者が答えた母音とが一致したときの評価値によって母音判別時の知覚のマグネット効果を調べた．

### 4.3 結果

まず初めに，我々が設定したロボットの母音らしさが適切であったかを検討する．設定した母音らしさが適切であるならば，ロボットの母音らしさが増すに連れ，被験者の模倣音声が母音に近付いていかねばならない．

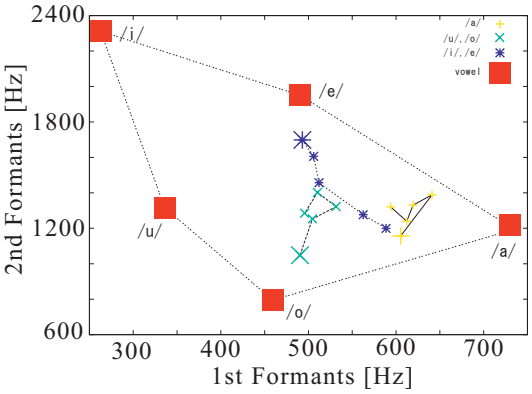



Figure 7: The average of formant vectors of subjects imitation and the average formant vectors.

Figure 7 はロボットの発声 $\boldsymbol{r}_i^{/v/}$ (/v/ = {/a/, /u/, /e/}，$i = 1, \cdots, 5$) に対して被験者が模倣したときのフォルマントである．ただし，＋は /v/ = /a/，×は /v/ = /u/，*は /v/ = /e/ を示しており，$i = 5$ のみ大きなプロットとして $i = 1, \cdots, 5$ までを線で結んである．また，■を頂点とする五角形は被験者 10 人の母音のフォルマントベクトルの平均である．Figure 7 より，ロボットの母音らしさ $i$ の増加に合わせて被験者の模倣がロボットの発声と同じ母音に近付いているのは，ロボットが /e/ を発声したときのみであり，ロボットの母音 /a/，/u/ に設定した母音らしさは不適切であったと考えられる．従って，ロボットの母音らしさの適切であると考えられる /e/ の結果のみについて考察する．

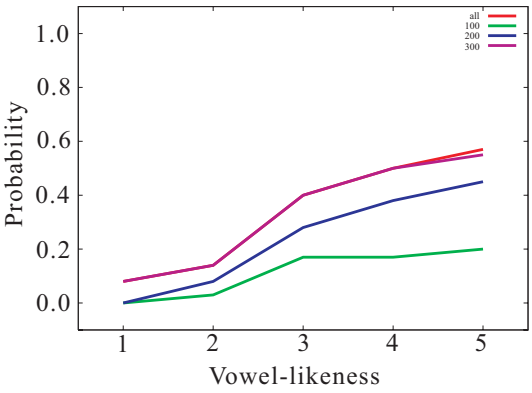



Figure 8: Probabilities in which the subjects replied with their vowels of which lip shapes were corresponded to the robot's one. Note that vowel-lbikeness corresponds to $i$ in equation (1).

本実験で使用した発話ロボットでは構音可能領域が狭いため，スケーリング係数 $\alpha$ により $\boldsymbol{r}_i^{/i/}$，$\boldsymbol{r}_i^{/o/}$ がそれぞれ $\boldsymbol{r}_i^{/e/}$，$\boldsymbol{r}_i^{/u/}$ とほぼ同じ値となった．このようにフォルマントベクトルがほぼ同じである場合，人は母音の違いを判別できないと考えられるため，ロボットの発声 /u/，/e/ に対応する被験者の発声はそれぞれ /u/，/o/ と /i/，/e/ であるとする．

Figure 8 はロボットの発声に対して被験者が /i/ または/e/ の母音で模倣した確率がロボットの発声の母音らしさによってどのように変化するかを示している．この確率は人が感じる音の差を低周波，高周波に限らず一定の間隔であらわせるように周波数を変換した値である MEL を閾値とし，各被験者の平均値で表される．Figure 8 は，模倣音声と母音との差が MEL 空間上で 100 以内，200 以内，300 以内，制限なしの 4 つを閾値として用いた結果である．

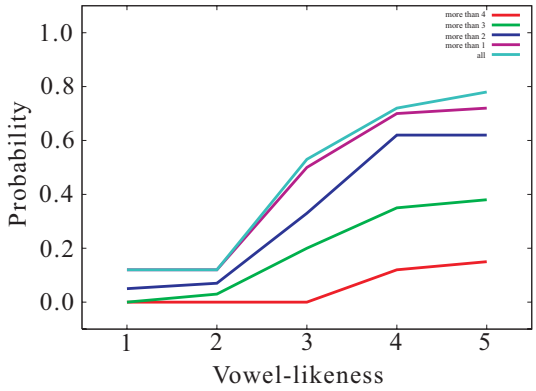



Figure 9: Probabilities in which the subjects confidentially categorized the heard sound of which lip shapes were corresponded to the robot's one. Note that vowel-likeness corresponds to $i$ in equation (1).

一方，Figure 9 はロボットの発声に対して被験者が /i/ または/e/の母音であると答えた確率がロボットの発声の母音らしさによってどのように変化するかを示している．この確率は被験者の評価点を閾値とし，各被験者の平均値で表される．Figure 9 は評価値が 5 点以上，4 点以上，3 点以上，2 点以上，1 点以上の 5 種類の結果であ

ある母音らしさ i でのロボットの発声 4 回に対し，被験者が母音 /i/ または /e/ と答えたときの回数を被験者 10 人で平均化したものである．また，各線は評価値による閾値が 5 点以上，4 点以上，3 点以上，2 点以上，1 点以上の 5 つを閾値として用いた結果である．

発声の必要がない母音の判別では知覚のマグネット効果のみが働くと考えられる．一方，模倣する場合にはロボットの発声を聞くときに知覚のマグネット効果が，さらにその知覚した音を発声する際に構音のマグネット効果が働くと考えられる．そこで，Figure 8，9 を比較することで構音のマグネット効果が模倣時にどのような働きを

しているか考察する．Figure 8，9 はどちらもシグモイド関数に似たデータの変化を示していることがわかる．これは，ロボットの発声が母音様に近付くと，マグネット効果によって急激に母音であると知覚，模倣しやすくなることをあらわしている．

ここで，いつマグネット効果が現れ始めたのかを調べるため，各データの縦軸の最大値のと最小値の中間となる地点を通過するときの母音らしさを計算することでマグネット効果の現れるタイミングをあらわした．ただし，各母音らしさは線形で補間し，被験者による差や個人の応答のばらきをの影響を防ぐため，それぞれの実験の各条件における平均値で計算した．結果，母音らしさが模倣時は各条件での平均値で 2.65，判別のときは各条件での平均値で 3.04 のときに中間点を通過した．これは模倣時のほうがマグネット効果が早くあらわれると考えられることを示している．つまり，人はロボットの発声を模倣するときに，知覚のマグネット効果だけでなく構音のマグネット効果が加わることで，より母音に近い音を発声すると考えられる．

## 5 結論

本論文では，前研究で仮定した相互模倣が母音に収束するメカニズムの原因となる知覚のマグネット効果と，構音のマグネット効果の 2 つのマグネット効果について議論した．被験者がロボットの発声を模倣，またはどの母音であるか判別する実験より，知覚のマグネット効果だけではなく構音のマグネット効果によって被験者が母音様の発声で模倣することを示した．

## 参考文献


[1] 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.

[2] Minoru Asada, Karl F. MacDorman, Hiroshi Ishiguro, and Yasuo Kuniyoshi. Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous System*, 37:185–193, 2001.

[3] B. de Boer. Self-organization in vowel systems. *Journal of Phonetics*, 28:441–465, 2000.

[4] 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)*, pages 406–416, 2002.

[5] M. Peláez-Nogueras, J. L. Gewirtz, and M. M. Markham. Infant vocalizations are confitioned both by maternal imitation and motherese speech. *Infant behavior and development*, 19:670, 1996.

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

[7] Yuichiro Yoshikawa, Minoru Asada, Koh Hosoda, and Junpei Koga. A constructivist approach to infants' vowel acquisition through mother-infant interaction. *Connection Science*, 15(4):245–258, Dec 2003.

[8] Patricia K. Kuhl. *Plasticity of development*, chapter 5 Perception, cognition, and the ontogenetic and phylogenetic emergence of human speech., pages 73–106. MIT Press, 1991.

[9] 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*, pages 3858–3863, 2002.

[10] Philip Rubin and Eric Vatikiotis-Bateson. *Animal Acoustic Communication*, chapter 8 Measuring and modeling speech production. Springer-Verlag, 1998.

[11] R. K. Potter and J. C. Steinberg. Toward the specification of speech. *Journal of the Acoustical Society of America*, 22:807–820, 1950.