

母音と色の感覚間協応における音響特徴と音韻カテゴリーの影響

Effects of Acoustic Features and Phoneme Categories on Vowel–Color Crossmodal Correspondence

原島 小也可^{†,‡}, 浅野 倫子[†]
Sayaka Harashima, Michiko Asano

[†] 東京大学, [‡] 日本学術振興会
The University of Tokyo, Japan Society for the Promotion of Science
harasima@l.u-tokyo.ac.jp

概要

日本語母語話者を対象に、母音に感じられる色を選択する課題を行った。その結果、母音の音響特徴（フォルマント）に加えて音韻カテゴリーが色選択に強く影響すること、過半数の参加者で音と色の空間構造に同型性が見られること、マッピングが構造的に同型な人ほどマッピングが時間的に安定していることが明らかになった。この結果は、母音と色の感覚間協応が音韻カテゴリーという知覚者の認知構造を反映する形で形成されている可能性を示唆する。

キーワード: 感覚間協応, 共感覚, カテゴリー, 同型投射 (isomorphism)

1. 序論

1.1 感覚間協応

感覚間協応とは、異なる感覚モダリティに属する刺激の属性や次元間に恣意的ではない結びつきが見出される現象である (Spence, 2011)。例えば、高いピッチの音は、小さなサイズ、高い位置、明るい色の視覚刺激と結びつけられやすいことが示されている。

1.2 母音と色の感覚間協応

感覚間協応には多様な組み合わせがあり、母音と色の間にも協応関係が存在することが報告されている。Moos et al. (2014) は、母音に含まれる音響特徴（フォルマント）が選択される色に影響を与えることを明らかにした。具体的には、前舌・高母音 (e.g., /i/) には高明度で黄・緑がかった色が、後舌・中低母音 (e.g., /o/, /a/) には低明度で青・赤みがかった色が対応づけられた。

さらに、Cuskley et al. (2019) はオランダ人を対象に大規模なオンライン実験を実施し、Moos et al. (2014) の結果が概ね再現されたことを報告するとともに、線形混合モデルを用いた分析を行い、色を予測するモデルとして音響特徴（フォルマント）に基づくモデルと音韻カテゴリーに基づくモデルを構築して両モデルを比較した。その結果、音響特徴に加え、/a/や/i/といった音韻カテゴリーが色のより強力な予測因子であることを明らかにした。このことは、母音と色の協応では、ある音声がどの

母音のカテゴリーに属するものとして主観的に知覚されるのかという聴取者の音韻表象が重要な役割を果たしていることを意味する。加えて、Cuskley et al. (2019) は、多くの参加者において、F1（第一フォルマント）- F2（第二フォルマント）音響空間内で近い（遠い）母音の組み合わせが、CIE L*a*b*色空間内で近い（遠い）色の組み合わせに結びつけられたことを報告し、母音空間の特徴間の関係性が構造的に同型な形 (isomorphism) をもって色空間に写像されている可能性を提案した。このことは、感覚間協応が、あるモダリティの特徴間の関係性が別のモダリティの特徴間の関係性に同型投射されるように形成される可能性を示唆している。

音韻表象の役割を検証することは、母音と色の感覚間協応の基盤となる認知メカニズムを明らかにするうえで重要である。仮にこの協応が母音の音響的特性 (e.g., フォルマント周波数) に主として依拠するのであれば、本現象は知覚レベルでの感覚特徴間の構造的対応に起因する現象として捉えることができる。一方で、音韻カテゴリーは特定の文化や環境において学習されるものであり、文化や個人ごとに音韻カテゴリーの境界や構成が異なる。音声がどの音韻カテゴリーとして知覚されるかという、聴取者の言語的な表象が協応に影響を及ぼしているとすれば、感覚間協応は単なる知覚的な対応にとどまらず、言語経験や文化的背景によって構築される高次の認知表象の影響を受けていることを示唆する。この観点から、異なる文化圏において音韻表象が選択される色に及ぼす影響を検討することは、感覚間協応の文化的可塑性を評価するうえで理論的に意義深い。

1.3 本研究の目的

本研究は、Cuskley et al. (2019) を追試・拡張し、母音と色の感覚間協応がモダリティ間で構造的な同型な形をもって形成されているかを検証することを目的に実施された。基本的にはCuskley et al. (2019) の手続きを踏襲し、比較可能な設計での再現的検討を行ったが、部分

的に手続きに修正を加えた。具体的には、Cuskley et al. (2019) の研究では、音声刺激として母音音響空間に点在する、モーフィング刺激を含む 16 の音声を使用されていたのに対し、本研究では日本人にとって自然な日本語の母音音声を使用し、同様のマッピング傾向が見られるか検討した。さらに、Cuskley et al. (2019) は、各音声刺激に対してすべての参加者が同一の音韻カテゴリを知覚しているという仮定のもとに分析を行っていた。この仮定は、音韻知覚に個人差がある可能性を無視しており、モデルの精度に限界をもたらす。したがって、本研究では、各参加者が各音声刺激をどの音韻カテゴリとして知覚したかを個別に取得し、その主観的カテゴリ情報を統計モデルに組み込んだ。これにより、音韻カテゴリが色選択に与える影響を、より実際の知覚に即して厳密に分析することが可能となった。

2. 方法

2.1 参加者

視聴覚が正常な日本語母語話者 246 名 ($M_{age}=27.7$ 歳, $SD_{age}=7.9$ 歳, 男性 116 名, 女性 127 名, 回答を希望しない 3 名) が実験に参加した。

2.2 聴覚刺激

ATR デジタル音声データベースセット A (Kurematsu et al., 1990) に収録されている、50 代男性アナウンサーが発声した日本語母音 5 種を使用した。加えて、母音空間において隣接する母音同士をモーフィングして中間母音 5 種を作成し、さらに将来的な実験のために追加で 2 種のモーフィング母音も作成した。したがって、音声刺激は計 12 種類であった (図 1a)。

2.3 手続き

オンライン上で実験を実施した。参加者は、聴覚刺激に対して、最も似合うと感じる色をカラーピッカーを

用いて選択し、どうしても色を選択できない場合にはチェックボックス「色を選択しない」を選択した (色選択課題)。各聴覚刺激はランダムに 3 回呈示されたが、これは母音と色の対応付けがどの程度時間的に安定しているかを評価するためであった (後述)。その後、参加者は、各音声刺激に対して何という音韻カテゴリを知覚したか、「あ、い、う、え、お」の中から選択した (評定課題)。最後に、アンケート課題を行い、共感覚の保有の有無等を尋ねた。

3. 結果

3.1 スクリーニング

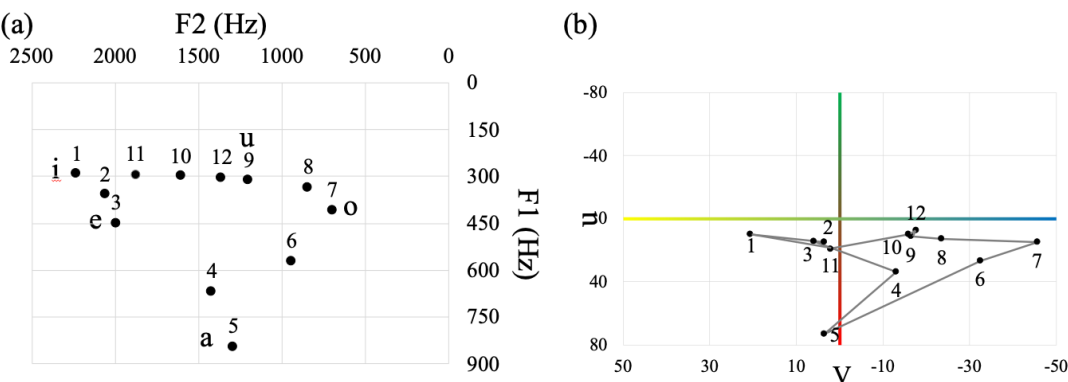
分析の前にデータのスクリーニングを行った。半数以上の試行で「色を選択しない」を選択した参加者等、11 名のデータを以降の分析から除外した。さらに、アンケート課題で「音声に色を感じる」と報告し、母音と色の対応づけが時間的に安定していた者 (基準は Rothen et al. (2013) を参照) 6 名を共感覚者とみなし、以降の分析から除外した。したがって、計 229 名の非共感覚者のデータが以降の分析の対象となった。

3.2 マッピング様態

Cuskley et al. (2019) にない、色選択課題で選択された全ての色について、RGB 値を CIE L*u*v*値に変換して分析した。図 1b に示す通り、F2 が高い母音 (e.g., /i/) には高明度で黄・緑みのある色が、F2 が低い母音 (e.g., /o/) には低明度で青みがかった色が、F1 が高い母音 (例: /a/) には赤みがかった色が選択された。これらは Cuskley et al. (2019) と一致する結果である。ただし、色の分布は赤み (u+) および青み (v-) 方向において、Cuskley et al. (2019) よりも約 2~2.5 倍程度拡張していた。

3.3 音響特徴と音韻カテゴリが選択色に及ぼす影響

図 1. (a) 音声刺激。(b) 刺激に選択された色の u*v*値。



各音声刺激に対する選択色の L^* (明-暗軸), u^* (緑-赤軸), v^* (青-黄軸) の各次元の座標値について、線形混合モデルを用いて音響特徴モデル ($F1$, $F2$ を固定効果とする) と音韻カテゴリモデル (評定課題で取得した音韻カテゴリを固定効果とする) を構築し、どちらのモデルが選択色の予測力・説明力に優れるか、モデルを比較して検討した。両モデル共通のランダム効果は参加者と呈示順序 (セット数) とした。モデルは R の lmer 関数 (lme4 package) を用いて推定され、変数選択は buildmer package を用いた後退法により行った。AIC, marginal R^2 , conditional R^2 に基づく比較の結果, Cuskley et al. (2019) と同様に、いずれの色次元においても音響特徴モデルよりも音韻カテゴリモデルの方が予測力・説明力に優れていることが明らかになった (表 1)。

3.4 音と色のモダリティ間の構造的同型性

Cuskley et al. (2019) にならい、参加者ごとに、マッピングの構造的同型性の程度 (構造化スコア、音響特徴が似た (似ていない) 母音ほど、似た (似ていない) 色と対応付けられる度合い) を算出した。構造化スコアの算出方法を以下に説明する。まず、すべての音のペア間の距離 ($F1$ - $F2$ 座標上のユークリッド距離) を算出した。また、選択されたすべての色のペア間の距離 (CIE $L^*u^*v^*$ 色空間におけるユークリッド距離) を算出した。次に、音の距離行列と色の距離行列の相関係数 r (実測された相関係数 r) を算出した。また、色の距離行列をランダムにシャッフルし、音の距離行列との相関係数を算出する操作を 1 万回繰り返し、ランダムな相関係数の分布を得た。そして、真の相関係数 r がこのランダム分布の中でどの位置にあるかを示す z 得点を算出し、これを母音-色マッピングの構造化スコアと定義した。 z 得点が高いほど、音と色の対応関係が構造的に同型に近いことを意味する。最後に、参加者が音と色をランダムにマッピングしているという帰無仮説を置いた場合に、実際に得られたような相関係数が生じる確率を示す p 値を算出し、 $p < .05$ の場合、偶然に期待され

るよりも統計的に有意に構造化されていると判断した。その結果、約 53% の参加者においてマッピングが偶然に期待されるよりも構造化されていることが明らかになった。また、Cuskley et al. (2019) と同様に、構造化スコアと時間的安定性スコア (値が小さいほどマッピングが時間的に安定していることを示す) の相関は $r = -.41$ ($p < .01$) で、構造化スコアが高いほど、母音と色の対応付けの時間的安定性が高かった。すなわち、母音の特徴間の関係性と、母音に結びつけられる色の特徴間の関係性が構造的に同型である人ほど、母音と色の対応付けが時間的に安定していることが明らかになった。

4. 考察

本研究では、日本語母語話者を対象に、母音と色の対応づけにおける諸特性を検討した。その結果、(1) 母音間での音響特徴 ($F1$, $F2$) の関係性が、母音に対応付けられる色の次元 ($L^*u^*v^*$) に反映されること、(2) 音響特徴そのものよりも、聴取者が知覚した音韻カテゴリが色の選択に対してより強い影響を及ぼすこと、(3) 参加者の過半数で音響空間と色空間の距離構造の間に統計的に有意な同型性 (isomorphism) が見られたこと、(4) マッピングが構造的に同型である人ほど、マッピングが時間的に安定していたことが明らかとなった。

4.1 母音の音響特徴と色の対応づけ

本研究では、Moos et al. (2014) や Cuskley et al. (2019) と同様に、 $F2$ が高く $F1$ が低い前舌・高母音 (e.g., /i/) に高明度で黄・緑がかった色が、 $F2$ が低く $F1$ が高い後舌・中低母音 (e.g., /o/, /a/) に青・赤みがかった色が選択される傾向が確認された。この知見は、母音の音響的特徴と色の間に、感覚モダリティを超えた対応関係が存在するという立場を支持する。一方で、本研究における色の選択分布は、Cuskley et al. (2019) の報告よりも赤-青方向に約 2~2.5 倍拡張していた。著者らは、別の実験において Cuskley et al. (2019) が使用しているものと

表 1. 音響特徴モデルと音韻カテゴリモデルの比較。

	Model type	Marginal R^2	Conditional R^2	dAIC	Model
L^*	音響	0.06	0.29	681.6	$L^* \sim 1 + F1 + F2 + (1 id)$
	音韻カテゴリ	0.12	0.35	–	$L^* \sim 1 + \text{phoneme} + (1 id)$
u^*	音響	0.10	0.18	169.5	$u^* \sim 1 + F1 + F2 + (1 id)$
	音韻カテゴリ	0.11	0.20	–	$u^* \sim 1 + \text{phoneme} + (1 id)$
v^*	音響	0.05	0.16	100.3	$v^* \sim 1 + F1 + F2 + (1 id)$
	音韻カテゴリ	0.06	0.17	–	$v^* \sim 1 + \text{phoneme} + (1 id)$

全く同じ音声刺激を用いて色選択課題を実施したが、その実験でも赤-青方向への色分布の拡張が確認された。この結果は、色の選択範囲の拡張が音声刺激の違いに起因する可能性を排除するものであり、日本語母語話者に特有の母音と色の対応づけ様式が反映された結果である可能性を示唆する。拡張の要因は現時点では明らかではないものの、マッピング様態の詳細においては、文化的・言語的背景に起因する差異が生じ得ることを示す所見であると解釈できる。すなわち、母音と色の対応には基本的な傾向が広く共有されている一方で、文化や言語によって微妙な調整が加えられている可能性がある。

4.2 母音の音韻カテゴリの影響

Cuskley et al. (2019) と同様に、線形混合モデルの比較により、音響特徴モデルよりも音韻カテゴリモデルの方が色の選択に対してより高い予測力・説明力を持つことが示された。この結果は、音の物理的な特徴に加えて、音に対して言語的に対応づけられた音韻カテゴリの方が感じられる色に及ぼす影響が強いことを示している。これは、視聴覚間の対応が、母音の物理的属性だけではなく、カテゴリ的な認知表象に基づいて生じる可能性を示唆する。特に、言語経験によって形成された音韻カテゴリが色と結びつくことで、視聴覚間のマッピングが文化的・個人的に組織化されていくプロセスを示唆している可能性がある。

4.3 モダリティ間の構造的同型性

音空間と色空間の構造的同型性を z 得点として定量化した結果、半数以上の参加者において偶然よりも統計的に有意に構造化されていることが確認された。この結果は、母音の音響特徴間の類似関係が、母音に対応づけられる色の関係性に構造的に同型な形で投射されている（音響的に類似している（いない）母音には類似している（いない）色に対応づけられる）ことを意味する。また、この構造化スコアと時間的安定性の間に中程度の負の相関（ $r = -0.41$ ）が認められたことは、感覚間の対応づけが構造的に同型であるほど、対応づけが時間的に安定していることを示唆する。この結果は、対応関係が構造的に整理されている人ほど、時間が経過しても投射にランダムな変動が生じにくくなるために生じたのかもしれない。

このような構造的同型性の背景には、音韻カテゴリが果たす認知的機能が関与している可能性がある。Goldstone & Hendrickson (2010) によれば、カテゴリには異なるカテゴリに属するもの同士の違いを強調し、同

じカテゴリ内の差異を目立たせなくするという性質があり、感覚空間における特徴同士の関係を組織化する働きを持つ。したがって、音韻カテゴリによって音と色の対応関係もより構造的に組織され、結果として対応づけの時間的安定性が高まる可能性がある。

5. 結論

日本語母語話者においても母音と色の感覚間協応が確認され、母音の音響特徴に加え、音韻カテゴリが対応づけに強く寄与することが示された。さらに、過半数の参加者において対応関係にある母音と色の空間構造に統計的に有意な同型性が確認され、構造化の程度が高い人ほど対応づけが時間的に安定していた。母音と色の感覚間協応では、感覚モダリティ間の構造化された写像関係に音韻表象が寄与している可能性がある。このことは、感覚間協応には言語的・文化的経験を通じて構造化される認知的現象として捉えられるタイプのものが存在することを示唆しており、今後は対応づけの文化差や個人差も含めた多層的な理解が求められるだろう。

6. 引用文献

- Cuskley, C., Dingemanse, M., Kirby, S., & van Leeuwen, T. M. (2019). Cross-modal associations and synesthesia: Categorical perception and structure in vowel-color mappings in a large online sample. *Behavior Research Methods*, 51(4), 1651–1675.
- Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *WIREs Cognitive Science*, 1(1), 69–78.
- Kurematsu, A., Takeda, K., Sagisaka, Y., Katagiri, S., Kuwabara, H., & Shikano, K. (1990). ATR Japanese speech database as a tool of speech recognition and synthesis. *Speech Communication*, 9(4), 357–363.
- Moos, A., Smith, R., Miller, S. R., & Simmons, D. R. (2014). Cross-modal associations in synaesthesia: Vowel colours in the ear of the beholder. *i-Perception*, 5(2), 132–142.
- Rothen, N., Seth, A. K., Witzel, C., & Ward, J. (2013a). Diagnosing synaesthesia with online colour pickers: Maximising sensitivity and specificity. *Journal of Neuroscience Methods*, 215, 156–160.
- Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Attention, Perception & Psychophysics*, 73(4), 971–995.