

シーンの瞬時的カテゴリ認識における 上位レベルの優位性の検討

坂野 逸紀

京都大学大学院人間・環境学研究科

齋木 潤

京都大学大学院人間・環境学研究科

We investigated how fast superordinate and basic categorization was achieved in a go/no-go visual categorization task. Stimuli were extracted from a database of Oliva & Torralba's (2001) 2 superordinate and 8 basic categories. In Experiment 1, Natural category was defined as a set of Coast and Mountain, while Man-made as Highway and City Center. In Experiment 2, Highway images were replaced with Street images. Subjects were asked to press a key if briefly presented stimuli (40ms) were targets as quickly and as accurately as possible. In Experiment 1, reaction times of superordinate category classification were not faster than basic-level category. Meanwhile, In Experiment 2, speed advantage for superordinate categories over basic categories appeared. These results suggest that superordinate categorization is not faster than basic level categorization when membership within a superordinate category are widely distant, and that it is unlikely that scene classification is sequential processing on which either of category levels is processed "first".

Keywords: scene perception, rapid categorization, natural image, similarity.

問題・目的

ある山の写真を見せられたとき、多くの人はその写真を「山」という言葉で表現する。「自然」や「吉田山」と呼んでも誤りではないが、そのような抽象的、あるいは具体的な表現を用いることはあまりない。このように、あるシーンが属する複数のカテゴリレベルのうち、中間的な抽象度を持った表現が選好的に使われることが知られている。これは基本レベルと呼ばれ、より抽象的な上位レベル、具体的な下位レベルとは区別される。近年、カテゴリ判断に要する時間からカテゴリ認識の視覚的処理を調べた研究では、上位レベルの方が基本レベルより早く判断できることが示された(Joubert et al., 2007)。この知見のシンプルな解釈の1つは、上位レベルのカテゴリ認識が基本レベルに先だっ「最初に」達成されるというものである。しかしながら、この解釈は非直感的な予想をもたらす。基本レベル間の類似性は一定ではない。例えば、先行研究ではCity CenterとHighwayの類似性はCity CenterとStreetの類似性よりも低いことが示されている(e.g., Loschky & Larson, 2008)。類似性の低いもの同士の間は高いもの同士よりも早く、正確に行えるはずだが、もし先の解釈が正しいならば、基本レベルの判断がどれだけ区別の容易なカテゴリ間で行われていようと、上位レベルの反応時間における優位性は保たれることになる。本研究ではgo/no-go課題を用いて、シーンのカテゴリ認識が基本レベルのカテゴリ間の類似性構造に依存せず、常に上位レベルから基本レベルへと順に達成されているのかを検討した。

方法

協力者 実験1では18名、実験2では20名の大学生および大学院生が参加した。全員が裸眼ないしは矯正

した上での正常な視力を有していた。実験2では、教示を正確に理解していなかった2名が分析から除外された。

刺激 Oliva & Torralba(2001)のデータベースより自然画像を各基本レベルから114枚ずつ選択した(Figure 1)。そのうち各カテゴリから6枚ずつが練習でのみ用いられた。上位レベルであるMan-madeを構成するカテゴリとして、実験1ではHighwayとCity Center、実験2ではStreetとCity Centerを用いた。一方で、Naturalを構成するカテゴリは両実験ともにCoastとMountainで固定した。画像は256階調のグレイスケールに変換後、平均輝度とRMSコントラストの平均化が行われた。

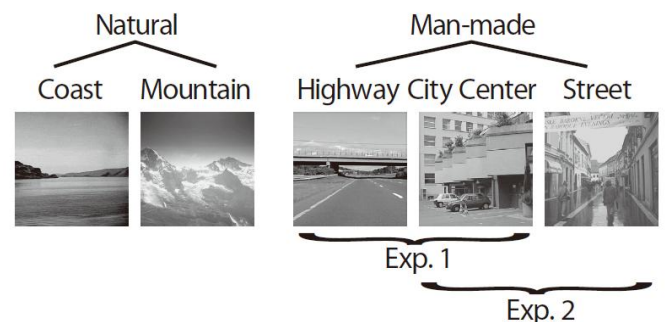


Figure 1. Examples of the stimuli. Highway images were used only in Experiment 1, while Street images in Experiment 2.

手続き 協力者は、注視点を800ms注視した後に、瞬間呈示(40ms)された画像が事前にキューとして指定されたカテゴリに属していた場合、できるだけ早く、正確にキー押しで回答することを求められた。回答に

は 2000ms の時間的制約があり、この間に協力者がキーを押さなかった場合は指定されたカテゴリに属しないと判断したものとみなされた。キューは上位レベルの2つと基本レベルの4つ、計6つが用意された。上位レベルがキューとなった際は、キューに該当しない基本レベルカテゴリから均等にディストラクタ刺激が選択された。一方、基本レベルがキューとなった際は、キューと上位レベルを共有する/しない他の基本レベルカテゴリが1つずつ選ばれ、そこから均等にディストラクタ刺激が選択された。回答後、注視点が再び呈示され、以降同じ手続きが繰り返された。1ブロックは36試行であり、キューごとに2ブロック連続、計12ブロックが行われた。キューは各ブロックの始めにのみ呈示された。

結果

実験1においては、先行研究で指摘されていた上位レベルの優位性、すなわち"Natural"や"Man-made"の判断が基本レベルでの判断よりも早く、あるいは正確に為されるような結果は認められなかった(Figure 2)。ある上位レベルとそれに従属する基本レベル、つまり"Natural"と"Coast", "Natural"と"Mountain", "Man-made"と"Highway", "Man-made"と"City Center"の計4組に対してWilcoxonの符号付き順位和検定を行ったところ、どの組においても感度(A'), 平均反応時間ともに有意差はみられなかった(Holm法により有意水準を調整、全体の水準は.05)。また、全協力者をプールした反応時間の分布を10msごとのビンで区切り、ヒット数がフォールスアラーム数を初めて有意に上回るビンを最小反応時間と定義した上で、その値を検討したところ、上位レベルの反応は基本レベルの反応より0~30ms遅いという結果が得られた。

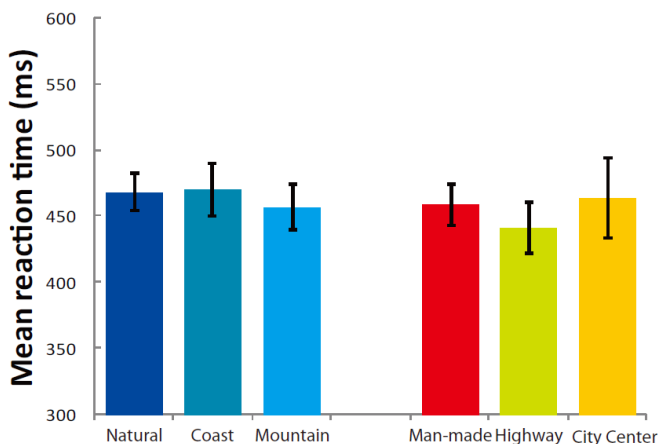


Figure 2. Mean reaction times in Experiment 1.

実験2においては、上位レベルの判断は基本レベルに対し有意に早く、正確に為された(Figure 3)。実験1と同様に4組の比較対に対して検定を行ったところ、"Natural"と"Mountain"の間を除き、どの組においても感度、平均反応時間ともに有意差がみられた。"Natural"と"Mountain"の組では、感度では有意差は見

られなかったものの、平均反応時間は"Natural"条件の方が早かった。また、最小反応時間を求めたところ、上位レベルの反応は基本レベルの反応より0~60ms早いという結果が得られた。

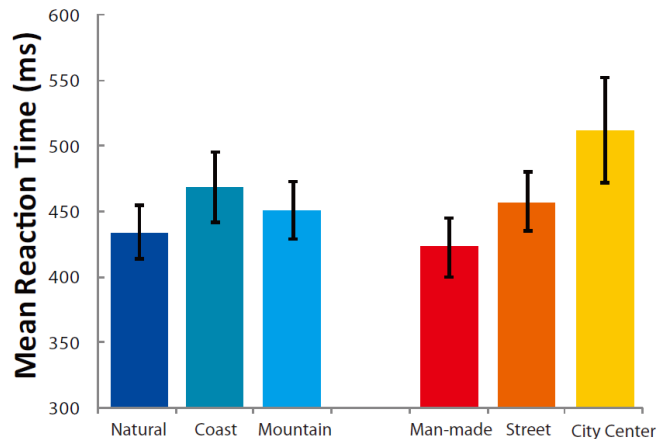


Figure 3. Mean reaction times in Experiment 2.

考察

2つの実験から、NaturalやMan-madeという上位レベルは、より下位の基本レベルよりも常に早く、正確に判断されるものではないことが分かった。このことは、シーンのカテゴリ認識が必ずしも上位レベルから、続けて基本レベルへと段階的に達成されていくものではないことを示唆している。むしろ、本研究の結果はカテゴリ間の類似性構造をもとに解釈するのが妥当と思われる。Murphy & Brownell(1985)は、最も早く判断されるカテゴリレベルはカテゴリ間の非類似性とカテゴリ内の類似性のトレードオフによって決まることを示している。実験1と実験2の結果の違いは、この類似性構造を操作することで、トレードオフが最適化されるレベルが変動したことによるものと解釈できる。

引用文献

- Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. 2007 Processing scene context: Fast categorization and object interference. *Vision Research*, 47, 3286-3297.
- Loschky, L. C., & Larson, A. M. 2008 Localized information is necessary for scene categorization, including the Natural/Man-made distinction. *Journal of Vision*, 8(1):4, 1-9
- Murphy, G. L., & Brownell, H. H. 1985 Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 11, 70-84.
- Oliva, A., & Torralba, A. 2001 Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42, 145-175.