



映像理解の認知的メカニズム

著者	杉尾 武志
雑誌名	文化情報学
巻	6
号	1
ページ	21-30
発行年	2011-03-10
権利	同志社大学文化情報学会
URL	http://doi.org/10.14988/pa.2017.0000013110

映像理解の認知的メカニズム

杉尾 武志

映像理解に関する認知心理学的研究の中で、基本的な出来事を単位とした形で映像をセグメンテーションするプロセスが注目されている。ある程度の幅をもった時間内の変化と定義される出来事をベースとして環境における情報が知覚されていることは、人間の行動を理解する上で有用な手がかりとなりうる。Newton (1973) による単純なボタン押し課題によるセグメンテーションに関する先行研究において、映像内の人物や物体の運動といった知覚的变化だけでなく、因果関係や目標の変化といった概念的变化もセグメンテーションを生起させる要因に含まれることが明らかにされてきた。現代の文化において映像は中心的な位置を占めており、映像理解の認知的メカニズムの解明は、人間と文化の関わりを考える上で重要なトピックとなっている。

1. はじめに

映画やテレビに代表されるように、映像は現代の文化において中心的な位置を占めている。広辞苑（第六版）によれば、映像は「映し出された物の姿や形」と定義される。この「映し出された」という点を重視するならば、写真や絵画も映像に含まれることになるが、一般的には動画像と同じ意味を持った用語として扱われる。本稿は、映像理解の認知的メカニズムに関する認知心理学的な研究を紹介し、メカニズムの中心的なプロセスとしての出来事 (event) を単位とした映像のセグメンテーション (segmentation) の役割を明らかにすることを目的としている。

ゲシュタルト心理学者であったRudolf Arnheimは、映画が現実とは異なっている点の一つに時間的および空間的に不連続であることを挙げた (Arnheim, 1957)。Arnheimは、こうした現実との相違点を強調することにより、映画は単に現実世界を映し出すだけの装置ではなく、芸術的な価値を持つことができると考えた。

時間的および空間的に不連続であることは、映像の表現能力という点で非常に重要な性質である。不連続であることによって、例えば2時間といった限られた時間内で、さまざまな場所や時間で起きたいくつもの出来事を表現することが可能になる。ただし、単に時間および空間次元で異なる映

像クリップを順番に並べただけでは、全体として何かしらの物語を表現することは困難である。日常的に映像を見る機会がある生活を送っている人間は、暗黙のうちに映像に関する「約束事」を文化的なルールとして習得している。こうした「約束事」にしたがって、映像内の不連続性を認知的に補間することで連続的なものとして解釈し、描き出されている物語を理解している。つまり、映像における物理的な不連続性は、必ずしも認知的にも不連続とみなされる訳ではない。人間が映像を見る際に、どのような事前知識や「約束事」を用いて映像の不連続性を補い、どのような情報に基づいて映像が描き出す物語を理解しているのかといった問題は、映像文化を考える上で重要なトピックであると言える。

2. 出来事とは何か

2.1. 出来事とアフォーダンス

環境の中で生活を送っていくためには、自分の周囲にどのような物体があるのかを正しく知る必要がある。自分の周囲にどのような物体が存在しているかに関する知識に基づいて、自分がどのような行為をこれから実行することができるかを定めることができる。例えば、「缶詰が食べたい」と思ったとしても、周囲に缶切が無ければ缶詰を食べることはできない。しかし、仮に鋭い金属片

のようなものが近くにあれば、その金属片を用いて缶詰を何とか開けることができるかもしれない。このように、人間が何かしらの目的を持って行為を実行しようとする際には、その物体がどのような意味的なカテゴリーに分類されるかではなく、どのような行為を可能にしてくれる存在であるのかが重要な情報となる。Gibsonはこのような物体の性質をアフォーダンス (affordance) と呼んだ (Gibson, 1979)。アフォーダンスという概念は、人間が環境とのインタラクションをどのようなメカニズムに基づいて計画し、遂行しているのかを考える上で大きな役割を果たしてきた。

その物体を用いて、ある特定の行為を実行することが可能であるかどうかを判断するためには、ある程度の幅を持った時間内で、その物体の機能について認知的なシミュレーションを行う必要がある。動きを排除した写真のように「凍りついた」視覚情報のみからでは、アフォーダンス的な情報を獲得することはできない。そもそも、普段の生活において「凍りついた」環境は存在し得ない。例えば、秋の京都において、紅葉した葉や枝はわずかであっても風に吹かれて揺れ続けているし、鴨川の水も止まったままではなく、常に流れ続けている。日常のどのような場面をとってみても、そこには何かしらの動きが常に存在している。このような場面内で起きている出来事は、一定の時間および空間の範囲内で生じる何かしらの変化を表していると定義することができる (中村, 2007)。Gibsonは、出来事として以下の3種類を挙げた (Gibson, 1979) :

- (1) レイアウトの変化 : 何かしらの力が物理的に与えられることで、環境における物体の面の形が変わること。
- (2) 色やテクスチャの変化 : 物体を構成する物質の組成が変化することによって、物体の面の色やテクスチャが変わること。
- (3) 面の形成と消失 : 物質の状態変化によって面が形成されたり消失したりすること。

これらの分類は、いずれも直接観察することが可能な物理的な変化に基づいている。例えば、秋に木々の葉が赤く色づくのは(2)に挙げた色の変化であり、やかんに入れた水を沸騰させた結果、水が注ぎ口から水蒸気となって出て行くのは(3)に挙げた面の消失に分類することができる。いずれの場合も何かしらの変化を表しているが、その変化が始まってから終わるまでの時間は大きく異なっ

ている。一般的に、「出来事」と呼ばれる変化は、神経インパルスの生起のように非常に短い時間幅で生じるものから、恒星の誕生から死までといった非常に長い時間をかけて起きるものまで様々なものが含まれる。本稿では、映像として人間が扱うことができる時間幅において生じる出来事 (具体的には、数秒から数十分) に限定して考察を行なう。数秒から数十分といった時間幅で生じる出来事は、他者が行う行為を理解する際に特に重要となる。例えば、ボクサーのパンチによって壁がへこんだ、という出来事に対して、因果的に関係した別の出来事と認知的に連合することにより、そのボクサーの心的状態について推定することが可能になる (例、「～といった理由で彼はあんなことをしたんだ」)。それに対して、長い時間をかけて壁の土がはがれおちていき、窪みが出来たといった出来事の場合は、人間の認知的な活動とは直接的には無関係な自然現象である。このように時間幅の長さを限定して出来事を扱うことは、認知心理学的には必要な前提となる。

出来事を知覚して理解するためには、連続的な感覚情報に対して、出来事の始まりがどの時点であり、終わりがどの時点であるかを明確に定めることができる必要がある。そのためには、映像理解の認知的メカニズムにおいて、連続的な感覚情報を離散的な「まとまり」に分割を行うプロセスが存在しなければならない。例えば、やかんの中で、水が水蒸気に状態変化する「沸騰」という出来事が起きている際には、どの時点が「沸騰」の始まりであり、どの時点が「沸騰」の終わりであるのかが明確に知覚され、その間の時間に認知された情報が「沸騰」という出来事であると理解される。同時に、その出来事が生じている空間の範囲も明確に示される必要がある。例えば、「沸騰」がやかんの中で起きているのか、それとも熱帯魚がいる水槽の中で起きているのかは、同じ「沸騰」であっても意味が大きく異なり、その出来事を知覚後にすべき行為も大きく異なる。このように、連続的な感覚情報を、時間および空間の二つの次元で明確な境界を持った「まとまり」として離散的に扱うことは、「そこでどのような出来事が起きているのか」を理解する上で基本的なプロセスである。

その他に、他者との間のコミュニケーションのような社会的場面においても人間は出来事を単位として認知的な分割 (セグメンテーション) を行っ

ている。例えば、一日に起きたことを自分の家族に話す場面を考えてみる。普通は、その日に起きたことをいくつかのまとまりに区切って、時系列的に正しい順序でそれらを話していくことになる。これらのまとまりは、それぞれ意味的なカテゴリーに分類された形で言及される（「電車に乗って京都駅にでかけた」、「昼は天下一品で食べた」など）。このように発話においてみられるセグメンテーションが、理解や評価においてどの程度まで本質的な役割を果たしているのかについては多くの議論がなされている（Schwan & Garsoffky, 2008）。しかし、先行研究から、映像理解が段階的かつ階層的に進行することはすでに明らかにされており、このことは、映像理解の認知的メカニズムにとってセグメンテーションは基本的なプロセスの一つであることを意味している。さらに、セグメンテーションの粒度（grain）を意図的に変えることにより、連続的、多次元的に変化する環境からどのようなレベルの情報を抽出するかを変えることができることも明らかにされている。こうした能動的な働きは、理解にとって不可欠である。

2.2. 出来事と状況モデル

文章や談話理解に関する認知心理学的研究から、文章や相手が話した内容を理解するために、音韻レベルやテキスト（命題）レベルだけでなく、その文章や内容が表している状況に関するモデルが構築されることが明らかにされている。ZwaanとRadvanskyは、状況モデル内に表象されている特徴次元として、時間、空間、因果関係、意図、行為主体の五つを挙げている（Zwaan & Radvansky, 1998）。このような状況モデルが構築されることにより、文章や発話の行間を読んだり、登場人物や話者の心的状態を推定したりすることができるようになる。さらに、テキストとして外的に表象された内容を理解するためには、ある時点より前に表象された内容との因果的な関係を正しくモニターしている必要がある。どの程度時間的に先行した内容まで表象できるかは、リーディングスパンのようなワーキングメモリの容量の個人差とも関係があると考えられる（Daneman & Carpenter, 1980）。

このような状況モデルは、テキストだけでなく映像として描かれている物語を理解する際にも形成されていると考えられる。具体的な手順は以下のように考えられる。まず観察対象である出来事

について、状況モデル内の特徴次元ごとに索引（インデックス）が付加される（例、空間次元＝「平安神宮」、時間次元＝「大晦日」、など）。この時、ワーキングメモリ内の出来事と特定の特徴次元において索引が共有されていれば、出来事間の連合が長期記憶内に形成される（つまり、これらの出来事が全体としてより大きな出来事を形成していると判断される）。このような特徴次元において索引が共有されているかを判断するプロセスが繰り返し実行されることにより、状況モデルが逐次更新され、最終的なモデルの形成に至ると仮定される。最終的な状況モデルは、映像として表現された状況に関する情報を内的に表象したものとなっている（どのような人物が登場するのか、登場する人物達はそれぞれどのような目標を持って行動しているのか、など）。

ワーキングメモリ内に形成された映像の状況モデルは、その内容を理解する上で重要な役割を果たしている。例えば、映像から直接知覚することができない情報を補間したり、この先にどういった出来事が生じるのかを予測したりすることは、いずれも状況モデルに基づいて行われると考えられる。世の中における「話し上手」と呼ばれる人たちは、自分の話を聞いている相手が適切な状況モデルを形成できるように、話の中の出来事同士の間関係を明瞭な形に整理する能力が優れていると言える（階層的な関係を明示する、など）。

2.3. 出来事と行為

日常生活において人間が会う出来事は他者の行為そのものであることが多い（例、「〇〇さんがドアを開けた」）。その一方で、他者の行為ではない出来事も頻繁に観察している（例、「風に吹かれてドアが開いてしまった」）。このように、出来事と行為は必ずしも同じ概念ではない（Zacks & Tversky, 2001）。

さらに、行為はあくまでも観察者とは無関係に環境内で起きるのに対して、出来事は観察者が知覚することでその生起が判断される。つまり、出来事は行為ではないが、行為はその行為を観察する人にとって出来事として認知される可能性がある。観察者が行為のどのような側面に注意を向けているかによって、行為者の意図とは異なる形で、出来事が認知されることもしばしば起こる。

このように、基礎的な行為をベースとして出来事は認知的に生成される。出来事とは何であるか

を考える上で、この点は重要であり、時間的および空間的な範囲を限定して情報を抽出しても、その情報は出来事とみなされない可能性がある。例えば、焼き鳥の串を回しながら焼くという行為を考えてみる。この行為は、焼き鳥を回しているという出来事と、焼き鳥を焼くという出来事の二つに分けて認知的に表象することができる。このことは、この行為について「焼き鳥を回す」という点に一切言及することなく「焼き鳥を焼く」という出来事について述べるのが可能であることから明らかである。次節以降で紹介する、映像知覚の際に出来事単位でのセグメンテーションが認知的に行われていることを示した認知心理学的研究に関しても、出来事が必ずしも行為と等しくないことはよく念頭においた上で解釈していく必要がある。

3. 認知心理学的な研究事例

3.1. 顕在的な反応によるセグメンテーションの測定

人間が映像を見ている時に、どのような出来事の構造が知覚されているかを明らかにするための実験手続きとして代表的なものがNewtonsonによる方法である (Newtonson, 1973)。Newtonsonによる方法はきわめて単純であり、以降の多くの研究において用いられてきた。

Newtonsonは5分の無音ビデオを実験参加者に見せた。ビデオの内容は、男子大学生が質問紙に几帳面に回答している場面から始まり、開始から2分後に質問紙を机の上において立ち上がり、自分のコートからタバコの箱を取り出して、マッチを探して見つけた後でタバコに火を付け、火が消えたマッチをゴミ箱に投げ捨てた後、ゴミ箱に入らなかったマッチを歩いて拾いに行き、再びゴミ箱に入れるといったものであった。実験参加者には、このビデオが他の部屋で別の実験を行っている参加者の様子を映したものであると説明された。その後、人間が他者の行動を分類する際に用いられている行動の単位を明らかにするために、自然で意味があると考えられる最も小さなまとまりの境界時点でボタンを押すことでマークを付けて欲しいと実験参加者に教示した。別の実験群に対しては最も大きなまとまりと判断される境界時点でボタン押しを行うように教示した (つまり、セグメンテーションの粒度が異なる2条件が設け

られた)。ここで境界は、ある行動のまとまりが終わって、別のまとまりが始まったと判断できる時点と定義された。ボタンは実験参加者に渡されたイベントレコーダに付いており、押した時点の正確な時間を記録することができた (標準的な映像が1秒間に30フレームであることを考えると、ボタン押しの時間精度も同程度であることが望ましい)。Windows上で実施する実験の場合は、Windows APIとして提供される高分解能パフォーマンスカウンタを用いることで実装することができる)。

こうした教示はかなり曖昧なものであったが、結果として得られた出来事間の境界の位置に関して、参加者間および参加者内で高い一致がみられた。さらに、セグメンテーションの粒度が異なる条件間で、まとまりの大きさは有意に異なっていたことから、実験参加者が意識的にセグメンテーションの細かさを調整できていたことが明らかにされた (「大きなまとまりに分けてください」と教示された場合は、「小さなまとまりに分けてください」と教示された場合よりも実際により大きなまとまりに安定した形で分けることができた)。

こうした手続きによって、映像全体でのセグメンテーションの回数 (頻度) と映像内でのセグメンテーションの位置を測ることができる。Newtonson (1973) においては、さらにビデオに登場した男性の性格についての評定と評定に対する確信度も答えるように求められた。興味深いことに、セグメンテーションの粒度が細かい条件の場合、評定に対する確信度が高くなっていた。Newtonsonは、セグメンテーションの頻度は観察者が映像からどの程度の量の情報が抽出しているかを反映していると考えた。できるだけ小さなまとまりにセグメンテーションを行った条件では、相対的に観察した出来事についてより多くの情報を含んだ心的表象が形成されたとみなすことができる。その結果、より多くの情報を含んだ表象に基づいて評定を行うため、評定に対する確信度が高くなったと説明することができる。

実験参加者全体としての境界時点は、以下の手順にしたがって求められた：

- (1) 映像を一定の区間に分割する。どの程度の時間幅を基準として分割するかは映像によって異なるが、比較対象となる映像の種類によって決められることが多い。Newtonson (1973) は、2秒間隔で分割を行った。

(2) 区間ごとに、その区間が境界であると判断された頻度の観察者間での平均を求める。

(3) (2) で求めた平均が一定の値を上回る場合に、その区間を境界であると決定する（多くは、平均値よりも1.65標準偏差分上回った場合に境界であると判断される）。

Newtonの実験手続きによるセグメンテーション結果が高い信頼性と妥当性を示したことは、人間が出来事単位でセグメンテーションを行う際に、状況モデルに含まれる複数の特徴次元の変化に対して意識的に注意を向けていたことを意味している。「自然で意味がある」といった曖昧さを残した教示であっても信頼性が高いセグメンテーションが可能であるという実験結果は、何かしらの共通した状況モデルが形成されており、セグメンテーションの基盤となっていることを示している。

セグメンテーションと同時に、または事後的に出来事ごとの内容を言語的に報告することもしばしば行われる。客観的なボタン押しの時間と言語的報告の両方を合わせて分析することにより、セグメンテーションにおいてどのような判断基準が用いられたのかを明らかにすることができる。しかしながら、同時に言語化をさせる場合には、言語化プロセスそのものがセグメンテーションに影響する可能性があるため、結果の解釈に注意が必要となる。

3.2. セグメンテーションの粒度と階層関係

Newton (1973) の結果から、教示によってどの程度の細かさで実験参加者がセグメンテーションを行うかを実験的に操作できることが明らかにされた。このような方法を用いて得られた、異なる粒度におけるセグメンテーションの結果間には、どのような関係があるものであろうか。細かくセグメンテーションを実行する場合も、粗くセグメンテーションを実行する場合も、最終的にはその映像を理解する目的のためであることから、両者が全く無関係に独立に行われる可能性は低い。Zacksらは両者の間で階層関係が成立している可能性を実験的に検討した (Zacks, Tversky, & Iyer, 2001)。具体的には、両粒度条件間での境界時点の一致度を、偶然に一致する度合いを考慮した形で求めることで、階層関係の有無を明らかにした。

1980年代以降、認知心理学において見ている物体が何であるかを判断する際にどのような心的表象が関わっているかについて多くの研究が進めら

れてきた。基本的な実験パラダイムにおいては、CGなどを用いて新奇な物体を作成し（例えばワイヤーフレーム状の物体がよく用いられた）、特定の視点から見たときの物体画像を実験参加者に提示した後、別の視点から提示した物体画像が同じ物体であるかどうかをできるだけ速くかつ正確に回答することが実験参加者に求められた。その結果、反応時間や正答率がテスト視点と学習視点間の距離（視点の違い）に比例するかどうかを検討された。比例を示す結果は、脳内で見え方のずれを補正するようなプロセスが働いていることを示唆すると考えられた（視点依存のプロセス）。こうした場合、物体は2次的に絵のような形式で心的に表象されていると考えられた。

しかし、必ずしも絵のような形式でなくても、視点依存的な実験結果が得られることが理論的に明らかにされた (Hummel & Stankiewicz, 1996)。物体が部分とそれらの関係に基づいた構造として表象されていると仮定した場合であっても、視点依存的な実験結果は説明可能である。例えば、自動車はドア、窓、エンジン、タイヤ、座席といった複数の部分から構成されている（細かく考えれば、標準的な自動車は約2万点の部品から構成されているとされる）。これらの部分はさらに細かい部分に分けることができる。シートの場合、座部、シートベルト、背もたれといった形で分けることが可能である。このように部分は階層的な関係を持っており、こうした関係は物体を識別する際に有用な手がかりとなることが多い。MarrとNishiharaが提唱した一般化円錐 (generalized cylinder) 理論は、こうした部分間の構造が物体認識において用いられることをふまえたものである (Marr & Nishihara, 1978)。

映像が出来事にセグメンテーションする際にも、粒度間で階層関係が存在することが期待される。例えば、「自動販売機で缶コーヒーを買う」といった場面を考えてみる。できるだけ大きなまとまりになるようにセグメンテーションを行うように教示された場合に、「財布から小銭を出す」と「缶コーヒーを買う」とセグメンテーションされたとしても、さらに「ポケットから財布を取り出す」、「財布の小銭が入っている箇所を開ける」、「小銭を選別する」、「小銭を取り出す」といった形により細かい水準でのセグメンテーションを行うことが可能である。行為表象が階層構造を有していることは、神経生理学および計算論的な研究からも

支持される (Botvinick, 2008を参照)。

Zacksらが行った実験では、親近度、出現頻度、手順に関する知識が異なる4種類の行動を記録した映像を用いて、同じ参加者が異なる粒度でセグメンテーションを行った際に、粒度条件ごとに求められる出来事間の境界位置がどの程度階層的な関係を有しているかが検討された (Zacks, Tversky, & Iyer, 2001)。粒度条件間での境界位置の一致度を、映像を1秒間隔で分割した後、その範囲内における重複に基づいて離散的に計算した場合と、連続的に計算した場合で比較を行った。離散的、連続的のいずれの場合もほぼ同様の結果が得られた。このことは、ある程度の時間幅を単位としてみなすことが、セグメンテーションの遂行成績を定量化する際に影響しないことを示している。ただし、どの程度の時間幅を1単位とみなすかは映像の内容およびどういった映像同士で比較を行うかによって異なるため、予備的な検討が必要となる。図1に粒度間での一致度を求める際の基本的な考え方を図式化した。まず、実験参加者ごとに、両粒度条件で境界が同定される確率を求めることができる。もし両条件間で階層的な関係が無く、境界位置の判断がそれぞれ独立になされるとするならば、細かい粒度条件で境界位置が同定される確率と、粗い粒度条件で同定される確率は独立であると仮定することができる。こうした仮定に基づいて、階層的な関係がないと仮定される場合に重複が生じる確率 (つまり、偶然両条件で一致する確率) を求めることができる。実験結果と両条件が階層関係にないと仮定した場合に求められる期待値間で対応のあるt検定を行った結果、両条件で有意に重複がみられることが明らかにされた。

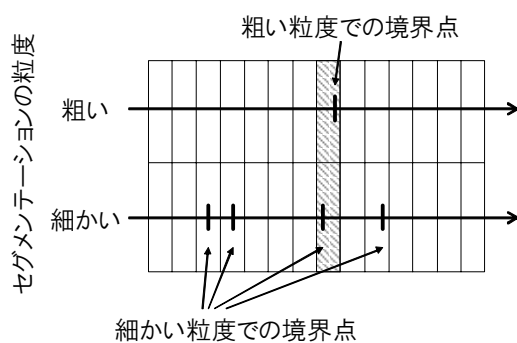


図1. 一致度を求めるための基本的な考え方 (Zacks, Tversky, & Iyer, 2001のFigure 2を改変)

3.3. セグメンテーションの要因

Newton (1973) による出来事間の境界位置を直接的に同定させる方法から、観察者が出来事を単位とした形で映像のセグメンテーションを行っており、出来事同士の階層的な関係が知覚されていることが明らかにされた。しかし、こうした出来事の構造についての知覚が、どのような要因に基づいて生じるのかを検討するためには、元の映像がどのような特徴を持っているのかをあらかじめコーディングしておく必要がある。

文章理解における出来事索引モデルを拡張したものとして、Zacksらは出来事セグメンテーション理論 (EST: event segmentation theory) を提唱した (Zacks et al., 2007)。ESTにおいては、観察者はワーキングメモリ内に出来事に関するモデルを形成すると仮定される (出来事モデル: event model)。このモデルは、映像から得ることができる情報の知覚を誘導する役割を担っている。例えば、映像を見ている際には、映像内でどのような出来事が次に生じるかについてワーキングメモリ内のモデルに基づいた予測がなされる。この予測が実際に起きた出来事と矛盾しない限りにおいて、ワーキングメモリ内の出来事モデルは維持される。出来事モデルは、状況を表すいくつかの特徴次元から構成されており、それらを用いて映像内で何が起きているのかが理解される。

一方で、予測が上手くいかなかった場合は、ワーキングメモリ内の出来事モデルは更新される。更新時には一時的に輸入される感覚情報に対してより多くの注意が向けられる (つまり、ボトムアップの処理が優勢になる)。更新後は、再び感覚情報を出来事として予測的に知覚していくと考えられる (トップダウン的処理が優勢になる)。このように、モデルの更新時に予測エラーが一時的に増加することが、出来事間の境界として主観的に経験されると考えられる。つまり、事前知識からは予測できない事態に直面したときに、「何か自分が知らない新しい出来事が起きている」と判断され、その時点が境界点と同定される (図2)。

何が次に起きるかが予測できない時点が、出来事間での境界として同定されるならば、具体的にどのような要因によって予測が上手くいなくなるのであろうか。一般的に、安定した状態よりも変化の生起の方が予測は難しい。環境において生じる特徴の変化は、予測の失敗をもたらす最も有力な候補である。これらの特徴の変化は、知覚的

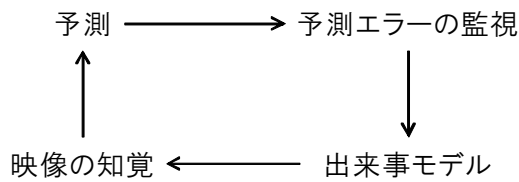


図2. 出来事モデル理論の枠組み
(Zacks et al., 2007のFigure 1を改変)

な変化と概念的な変化の二つに分類することができる。まず知覚的な変化とは、対象の動き（運動）によるものである。脳における視覚情報処理は、その特徴ごとにモジュール化されていることが神経生理学的に明らかにされている。例えば、色に関する情報は一次視覚野（V1）からV4と呼ばれる高次視覚野の一部を経由するとされるのに対して、視覚対象の運動情報は、V5またはMT+と呼ばれる領域を経由する。Zacksらは、fMRIを用いた脳活動のイメージング研究から、出来事間の境界時点でMT+を中心とした領域の活動が一時的に増加することを明らかにした（Zacks, Braver, et al., 2001）。この研究では、参加者は最初は映像を単に観察し、2度目は同じ映像に対してNewtsonの方法により出来事間の境界と判断できる時点で渡された手元のボタンを押して反応した。2度目の映像観察時に得られた境界時点のデータに基づいて、最初に映像を観察した時の脳活動が分析された。このような手順によって、教示によって意図的にセグメンテーションが実行されたのではなく、映像の運動情報が自動的に映像のセグメンテーションをもたらしたことを示すことができた。

その一方で、概念的な変化には、映像において描かれている物語内で登場人物がどういった目標で行為を行っているのかに関する変化などが含まれる。状況モデルを構成する五つの特徴次元は、すべて概念的変化を反映している。

Zacksらは、あらかじめ映像内に含まれる概念的变化をコーディングした後に、粒度が細かい条件と粗い条件のそれぞれについてセグメンテーション結果との関係を検討した（Zacks, Spear, & Reynolds, 2009）。映画などの物語を描いた映像に対するセグメンテーションを考える際には、カットなどの構造的な手がかりとの比較が必要不可欠である。物語の映像化においては、連続的に撮影された映像（ショット）を編集によって上手く繋ぎ合わせる必要がある。いかに編集による繋ぎ目（カット）を観客に意識させないようにするか、

という点は極めて重要であり、そのために多くの技法が生み出されてきた。しかし、カットが不連続な視覚情報を観察者に与えることには変わりがない。もし知覚的な変化が映像のセグメンテーションを必ず引き起こすならば、カットの映像内における位置と出来事間の境界位置は一致するはずである。Zacksらは、映像を5秒間隔で分割し、それぞれの区間においてカットが生じている場合と生じていない場合でセグメンテーションに与える影響をロジスティック回帰分析により検討した。

Zacksらは、実験において*The Red Balloon*（邦題「赤い風船」）という1956年に発表されたフランス映画を刺激材料の映像として用いていた。*The Red Balloon*は、会話は少ないのに対して状況変化が多く、さらに時系列的な不連続が無いことから、日常場面におけるさまざまな活動の知覚と共通した認知的メカニズムの働きを検討することができると期待された。つまり、Zacksらの研究は、必ずしも映像に特化した認知的メカニズムを検討したものではなかった。

映像のコーディングは、フレーム単位で状況変化があったかどうかを評定者が確認することで行われた。コーディングのカテゴリーは出来事索引モデルを基に7種類が設定された（Zwaan & Radvansky, 1998）。5秒間隔でそれぞれのカテゴリーに該当する状況変化が映像に含まれるかについて判断がなされ、少なくとも1回の変化があればその時間間隔に該当するカテゴリーの状況変化があるとされた。7種類のカテゴリーは以下の通りであった：

- (1)空間変化：映像内に登場する人物が動く方向を変えたり、カメラの位置が変わったりする場合（例、動き始める、動きを止める、突然方向を変えるなど）。映画技法の中に「トラック・スルー・ソリッド」と呼ばれるカメラがドアなどの硬い物質をそのまま突き抜けているかのように見せるものがあるが、その場合もフレームの前後で空間変化が生じているとして分類された。
- (2)時間変化：カットの前後でフレーム同士が時間的に連続していない場合は、時間変化が生じたとして分類された。時間変化は空間変化と必ず同期している。そのため、それぞれを別個の独立変数として映像のセグメンテーションの生起を予測するために用いることはできない（多重共線性が生じるため）。

- (3) 物体変化：映像内に登場する人物が、シーン内で物体との相互作用を変えた場合は、物体変化として分類された。例えば、テーブルの物を手に取るとか、すでに手に持っていた物を使い始める、といった場合が物体変化に相当している。
- (4) 人物変化：映像内で描かれている対象に対するフォーカスが特定の人物（複数でも構わない）に向けられており、かつそれらが先行するフレームとは異なる場合に、人物変化が生じたとして分類された。例えば、主人公の少年が追っかけられているシーンにおいて、カット後に主人公の少年から追っかけている少年達にフォーカスが移動した場合は人物変化としてコーディングされた。ただし、人物とはあくまでも「生物とみなすことができる存在」という意味であり、犬や猫、*The Red Balloon*の中で登場するあたかも生きていたかのようにふるまう風船も含まれている。
- (5) 人物間の相互作用の変化：文章のようなテキストには必ずしも明示されないが、物語の映像においてしばしば重要となる変化が人物間の相互作用の変化である。例えば、登場する人物同士が離れていた位置から歩み寄って手を握る、といった場面では物理的な相互作用に変化があったとみなすことができる。その他に、テーブルに向かい合って黙っていた二人が、話し始めるといった場面の場合は抽象的な相互作用に変化があったとみなすことができる。言うまでもなく、これら2種類の相互作用の変化が両方とも含まれる場合もある。
- (6) 因果関係の変化：映像内の先行するフレームにおいて描かれていた状況から、後続するフレームの活動が記述できない場合、因果関係に変化が生じたとして分類された。多くのシーンの変化においては因果関係の変化も同時に生じていることが多い。
- (7) 目標変化：あるフレームにおける登場人物が行っている行為の目標が、先行するフレームにおける目標と異なる場合は、目標変化として分類された。クロス・カットのように、同じ時間に異なる場所で起きている別々の事象を交互につないでいる場合は、目標変化の系列とみなすことができる。目標変化をコーディングする場合は、その行為の目標が大局的な目標であるのかどうかについてあらかじめ考えておく必要が

ある。

映像における上記の7種類の変化とカットの有無に関して5秒間隔でコーディングがなされた。一般に、このようなコーディング作業は2人の評定者が別々に実施し、コーディング結果間の一致度は、たまたま一致する程度を考慮した値であるCohenの κ (カッパ) 係数として定量化されることが多い(3人以上の評定者間での一致度はFleissの κ 係数によって検討されることが多い)。評定者間でコーディング結果が異なる場合は、各不一致項目について話し合いによっていずれかの評定値が修正される。

Zacksらは、コーディング結果を基に、状況変化の数が多ければ多いほどセグメンテーションが実行される確率が高くなるのか、そして個々の5秒間隔におけるセグメンテーションの生起が、各状況変化によってどの程度予測可能であるのかを検討した。前者は状況変化の効果が累積するかどうかを検討したとみなすことができる。

その結果、粒度の違いに関わらず、状況変化が生じた次元が多ければ多いほど、セグメンテーションが生起する確率が高くなることが明らかにされた。さらに、少なくとも細かい粒度でセグメンテーションを行った条件においては、状況変化によってセグメンテーションの生起が予測できることが示された。単にテキストとして物語を読む場合とは異なり、映像には多くの視覚的および(サウンドトラックなどの)聴覚的手がかりが存在している。これらの手がかりは、映像理解において重要な役割を果たしていると考えられ、今後さらに検討すべき問題であると言える。

3.4. 出来事構造の知覚と映像理解

これまで紹介してきた実験結果から、人間は映像を知覚する際に、知覚的および概念的な変化に基づいて映像をより基本的な出来事にセグメンテーションし、それらの出来事間の関係に基づいた構造を用いて次にどのような出来事が生じるかを予測していることが明らかにされた。

映像理解の文脈以外でも、予測メカニズムは対象物の理解において基礎的な役割を担っている。Bar (2007) は、脳が類推により予測生成を行っているとする理論的スキーマを提唱し、このスキーマが視覚的物体認識などの困難な課題を遂行する上で有効であることを明らかにしている。

本稿では紹介できなかったが、出来事間の境界

位置の近辺に含まれる情報についての記憶成績が他の位置の情報よりも優れていることや、出来事間の境界位置に対応するフレームのみを抽出したスライドショーを作成した場合、境界位置以外のフレームを抽出したスライドショーよりも、元の映像の内容が理解しやすいことなども実験的に明らかにされてきた。今後は、記憶や理解といった認知的側面だけでなく、特定の映像のどのような内容が情緒的な反応をもたらすのかについても認知科学的な観点からの検討が必要であると考えられる。さらに、映像には視覚と聴覚の両方の情報が含まれることから、視聴覚的な同期や、いわゆる「吹き替え」が映像理解にどのような影響を及ぼしているのかも興味深いテーマである。近年は日本語吹き替えによる洋画の劇場公開が多くなってきたが、これまでの字幕を用いてスクリーンの端に日本語訳を示す公開方式と、映像理解という観点においてどのような違いがあるのかは未だ検討されていない。

映像に対する認知的なセグメンテーションが、実験参加者が境界であると意識できる時点でボタンを押すといった極めて単純な課題により明らかに出来ることが示されたことにより、認知心理学的な映像理解の研究は大きく前進した。今後も映像の知覚、記憶、推論といった認知的段階におけるメカニズムを明らかにすることを目的とした研究が多く進められることが期待される。例えば、HassonらはfMRIを用いて映画鑑賞中の脳活動の計測を行い、各領域における信号変化率の実験参加者間での相関が極めて高くなることを明らかにした (Hasson et al., 2008)。さらに、同時に眼球運動についても計測を行った場合、水平方向および垂直方向に対する眼球位置についても実験参加者間での相関が高くなっていた。実験参加者は特定の教示を受けておらず、映し出された映画を自由に鑑賞していた。にもかかわらず、脳活動や眼球運動に関して参加者間で高い相関がみられたことは、出来事構造がより明確化される形で編集された映像は、その映像を観察する人間の知覚に関して何かしらの制約が含まれることを示唆している。

4. おわりに —文化と映像理解—

現代社会において映像は極めて身近な存在になりつつある。従来からの映画やテレビだけではな

く、インターネットを通じた動画配信技術の進歩や、高精細度テレビ放送や3Dテレビなどのより多くの情報量を含んだ映像の普及は、人間と映像の関わりをさらに変えていく可能性を秘めている。

こうした形で普段見ている映像のほとんどが何かしらの形で編集されたものである。編集されることにより、映像におけるフレーム間の空間的および時間的な関係は変化するにも関わらず、人間は映像内に描かれている出来事を容易に理解することができる。こうした能力は、映像を見る機会が非常に多い文化の中で生きていく間に無意識的に学習していくものであると考えられる (Schwan & Ildirar, 2010)。人間の認知的メカニズムがどの程度まで文化的に受け継がれたもので、どの程度までは生得的な知覚や記憶に関する能力を反映しているのかについては、未だ多くの問題点が残されている。映像の解釈および予測における文化的要因に関する認知心理学的検討が今後さらに行われることが期待される。

参考文献

- Arnheim, R. (1957). *Film as Art*. Berkeley: University of California Press.
- Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, 11, 280-289.
- Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences*, 12, 201-208.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.
- Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. J. (2008). Neurocinematics: The neuroscience of film. *Projections*, 2, 1-26.
- Hummel, J. E. & Stankiewicz, B. J. (1996). An architecture for rapid, hierarchical structural description. In T. Inui & J. McClelland (Eds.), *Attention and Performance XVI: Information Integration in Perception and Communication* (pp. 93-121). Cambridge, MA: MIT Press.
- Marr, D. & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three dimensional structure. *Proceedings of the Royal Society of London B*, 200, 269-294.

- Newtonson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, 28, 28-38.
- 中村浩 (2007). 事象知覚 大山 正・今井省吾・和氣展二・菊池 正 (編) 新編 感覚・知覚心理学ハンドブックPart 2 誠信書房 pp. 301-328.
- Schwan, S., & Garsoffky, B. (2008). The role of segmentation for perception and understanding of events. In T. Shipley, & J. Zacks (Eds.), *Understanding Events: From Perception to Action* (pp. 391-414). Oxford, et al.: Oxford University Press.
- Schwan, S., & Ildirar, S. (2010). Watching film for the first time: How adult viewers interpret perceptual discontinuities in film. *Psychological Science*, 21, 970-976.
- Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Buckner, R. L., & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, 4, 651-655.
- Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, 138, 307-327.
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, 133, 273-293.
- Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, 127, 3-21.
- Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, 130, 29-58.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162-185.