長岡佐知
三輪研究室 M2

The role of perspective in event segmentation

Khena M. Swallow, Jovan T. Kemp, Ayse Candan Simsek, Cognition 177, August 2018, pp. 249-262.

Introduction

- ✧ 経験を出来事に分けるプロセスを出来事分離（event segmentation）と呼ぶ(DuBrow & Davachi, 2013; Kurby &Zacks, 2008)
- ✧ このプロセスは様々な認知機能の重要な役割を持つ(Baird & Baldwin, 2001; Buchsbaum, Griffiths, Plunkett, Gopnik,& Baldwin, 2015),
- ✧ しかし、連続する経験を意味ある出来事に分離するプロセスがどういう情報に影響されているのか明確ではない

- ✧ 本研究は視覚的情報と情報内容がどうセグメンテーションに影響するのかを調査する
- ✧ 同じ行動（内容）が役者の視点（一人称視点）もしくは観察者の視点（第三者視点）で見ることによって出来事の分離の仕方が変わるのかを検討する
- ✧ これらの視点は視覚的特徴や役者の目標や感情へのアクセスが異なる(Jackson, Meltzoff, & Decety, 2006; Lamm, Batson, & Decety, 2007; Libby & Eibach, 2011; Nigro & Neisser, 1983; Storms, 1973; Taylor & Fiske, 1975; Vogeley & Fink, 2003; Vogt, Taylor, & Hopkins, 2003)

- ✧ Event segmentation の測り方 : 被験者に他者の行動を見せて (基本的に映像で) 聞く。被験者に行動を観察してもらう時、意味のある出来事が終わり新たな出来事が始まったと観察者が思った時に、ボタンを押して出来事の境界を示してもらう(Newtson, 1973)
- ✧ 曖昧なタスクだが、観察者の出す出来事の境界は観察者同士で意見が一致することが多く、信頼できる(Newtson, 1973; Speer, Swallow, & Zacks, 2003)

- ✧ Event segmentation には主にトップダウンモデルとボトムアップモデルがある
  - ➢ ボトムアップモデル：出来事の分離は視覚的情報などの知覚に影響されている
  - ➢ トップダウンモデル：出来事の分離は目的や知識などの解釈に影響されている

- ✧ ほとんどの event segmentation 研究が第三者視点映像を使用している
- ✧ 一人称視点と第三者視点は知覚的情報のボトムアップと知識や解釈トップダウンの両方に影響する
  - ➢ 一人称視点：

- ✧ 頭の動きでブレが多く、視覚的情報の変化が多い。
- ✧ 役者の近くにある物に焦点が当たりやすくなり、細かいディテールに気付きやすくなる(Borghi,Flumini, Natraj, & Wheaton, 2012; Jackson et al., 2006; Roche & Chainay, 2013; Vogt et al., 2003).
- ✧ 逆に、役者の姿勢や見た目と場所についての情報が少ない
  ⇒場面の空間的情報と内容が制限される(Henderson, Larson, & Zhu, 2008)

➢ 一人称視点はより具体的な描写を持ち、行動の仕方に焦点があたる
➢ 第三者視点はより抽象的な描写を持ち、役者の目的に焦点があたる
(Libby & Eibach, 2011)

✧ つまり一人称視点と第三者視点は異なる観念化方法を促す
➢ 一人称視点はより具体的なプロセス、第三者視点はより抽象的なプロセス

本研究について

✧ 二つの実験を通して event segmentation が一人称視点と第三者視点で異なるかどうか検討した
✧ ６つのアクティビティをヘッドマウントカメラ（一人称視点カメラ）と添え付けのカメラ（第三者視点カメラ）で録画した
✧ 被験者が異なる粒度で映像を分離した。そのデータを使い視点が出来事の分離にどう影響したのかを検証した
✧ ３つの仮説
➢ 視覚的情報従属仮説：もし分離が役者の姿勢や視覚的変化などのビデオの視覚的情報に影響されているのであれば、第三者視点と一人称視点の間に分離の差が見られる
➢ 内容従属仮説 : もし分離が映像の内容に基づくのであれば、第三者視点と一人称視点に分離の差は見られない
➢ 焦点変調仮説 : 視覚的情報が内容に影響する仮説。第三者視点より一人称視点ではもっと具体的な特徴に焦点を当てやすいため、一人称視点の映像は第三者視点より細かく分離される
✧ 実験でこれらの仮説を検証した

Methods

✧ Participants
➢ Cornell 大学の 72 人大学生（女性 51 名、男性 21 名、平均年齢=20.19、SD=1.86）
✧ Videos
➢ 役者（女性４名、男性４名）が６つの日常的なアクティビティを行っている映像を録画した

- ✧ それぞれのアクティビティは一人称視点と第三者視点で録画された
- ➢ 一人称視点：Go Pro Hero 3+, Black Edition をヘッドストラップで役者の頭に装着し、録画した
- ➢ 第三者視点：Go Pro Hero 4, Silver Edition を三脚に設置し。録画した
- ➢ 練習用の７つ目のアクティビティも録画されたがこのデータは分析に使わなかった

✧ Video feature coding

- ➢ 映像の視覚的特徴はグレイスケールに変換された後、５フレームごとに計算された
- ➢ 視覚的特徴
  - ✧ Luminance:フレームの平均的明度
  - ✧ Clutter:フレームごとのエッジと定義されるピクセルの比率(MatLab の edge detection algorithm の内の Laplacian of Gaussion method を使用し計算)
  - ✧ Visual activity index (VAI):前フレームとフレームの間のピクセルの変化
    - ✧ 全く同じ画像ならば VAI は 0
  - ✧ Optical flow:映像の動きの量
  - ✧ VAI と Optical flow の間に強い相関（r=.897without regard to perspective)
  - ✧ Touch onset: 手が物を触った時符号化
  - ✧ Touch offset:手が物から離れ、手と物の間に背景が見えた時符号化

✧ Procedure and design

- ➢ 被験者に日常的アクティビティを行っている人の映像を見てもらい、意味のある出来事が自然に終わり新たな出来事が始まった時ボタンを押すようにと教示した
- ➢ 全被験者は分離タスク課題の練習を行った
- ➢ 被験者は一つの視点の３つの映像を視聴した後、もう一つの視点の３つの映像を視聴した（全映像は別アクティビティ）
- ➢ アクティビティと視点の順番は被験者ごとにカウンターバランスを取った
- ➢ 分離教示は被験者間で異なった
  - ✧ Fine segmentation:出来事をできるだけ細かく見分けてもらう
  - ✧ Coarse segmentation:出来事をできるだけ大まかに見分けてもらう
  - ✧ Neutral segmentation:出来事の大きさについての教示なし
  - ✧ 被験者の出来事の分離率が一定の範囲になるまで練習タスクを行わせた

Results

✧ Visual features

- ➢ 第三者視点映像に比べて一人称視点映像の方が暗く、散かりが低かった
- ➢ 一人称視点の映像の方が visual activity, optical, flow, visible touch onset が多かった

- Clutter, VAI, optical flow は一人称視点と第三者視点の間に有意差あり、Luminance は有意傾向(Table 1.)

✧ Segmentation Rate

- 被験者が粒度の教示通り(fine, coarse, neutral)、出来事を分離したのかマニピュレーションチェックした
  - ✧ 被験者間 ANOVA によると、一分間でボタンが押された数と粒度間に有意差あり $F(2, 69)=23.34$, $p < .001$, $\eta^2_p=.402$.
  - ✧ 予測通り、被験者は fine 教示で一番多く境界を示し、neutral 教示で中くらいの境界を示し、coarse 教示で一番少なく境界を示した（Table 2）
- より着目したいのは出来事の分離率が視点によって異なったのか
  - ✧ 被験者のデータは統計の検定力を上げるために z-score に変換され(Bush, Hess, & Wolford, 1993)、視点と粒度を独立変数として ANOVA で計算された(Fig. 2a)
  - ✧ 分離率は一人称視点映像の方が第三者視点映像より高かった $F(1, 69)=5.86$, $p=.018$, $\eta^2_p =.078$.
  - ✧ 視点の効果は neutral grain では逆の数値を示していたが、視点×粒度のインタラクションに有意差はなかった $F(2, 69)=2.20$, $p=.116$.

✧ Boundary agreement

- 出来事と出来事の境界を示す場所が視点ごとに異なるのか検証した
- 仮説
  - ✧ 視覚的情報従属仮説 : 出来事分離は視覚的情報に影響されるなら、第三者視点と一人称視点の視覚的情報が異なるため視点間の境界の場所が異なる
  - ✧ 内容従属仮説 : 出来事分離は内容に影響されるなら、両視点同じ内容を持つため、視点間の境界の場所は変わらない
  - ✧ 焦点変調仮説 : 内容的従属仮説と同じで視点間の境界の場所は変わらない。粒度のみに影響される
- 被験者個人のデータとグループ全体のデータとの類似度を計算し、同じ視点で見たグループと違う視点で見たグループのデータを比較
- 内容従属仮説と焦点変調仮説と整合し、個人とグループのデータの類似度に有意な差は見られなかった $F(1, 69)=0.14$, $p=.713$
- 視点×グループ内外×粒度のインタラクションも見られなかった
- 粒度の主効果あり $F(2, 69)=49.19$, $p < .001$, $\eta^2_p =.588$
- 結果：被験者は一人称視点と第三者視点で似たような境界を示した

✧ The relationship between visual features and segmentation

- 視覚的特徴と分離の関係が視点間に異なるのか検証した
  - ✧ 焦点変調仮説を証明するために、視覚的情報と境界識別の関係は視点によっ

て変わるのかを見る必要がある

- 予測
  - 内容従属仮説：第三者視点映像と一人称視点映像の視覚的特徴が適度に相関するため(Table 1)、出来事分離と視覚的特徴の関係は視点ごとに適度に異なる
  - 焦点変調仮説：出来事分離は第三者視点映像より一人称視点映像の視覚的特徴に影響される
  - 視覚的情報従属仮説：どちらの視点でも分離は視覚的変化に同じくらい影響される
- 回帰モデルを個人の分離データに当てはめた
  - モデルは VAI, touch onset, touch offset とそれらのインタラクションを予測因子として使った
  - 一致度は Penalized log likelihood ratio statistic (PLR)で指標された
- 先行研究と整合し、分離の関数として視覚的特徴を用いた回帰モデルはチャンスレベルよりデータに一致した（Table 3）
- 重要な点としてモデルが第三者視点映像より一人称視点映像により一致し、視点の主効果が見られた(Fig. 3)
- 粒度が下がるほど一致度は下がる傾向が見られた(Table 3)

Discussion

- 被験者が一人称視点映像と第三者視点映像で示した出来事に差は見られなかった
  - 内容従属仮説と焦点変調仮説を支持
- 被験者は高粒度条件で第三者視点映像より一人称視点映像を見たとき、より多く分離を行った。回帰モデルは第三者視点映像より一人称視点映像のデータに当てはまった
  - 焦点変調仮説をより支持

Experiment 2

- 実験２は実験１のレプリケーションを主に行った

General Discussion

- アクティビティを見る時の視点を変えることによって、観察者が得る視覚的特徴が変化する
- 本研究の一人称視点映像と第三者視点映像は散かり、視覚的動き、や touch onset and offset が異なった
- 予想通り、一人称視点映像は第三者視点映像より移り変わりが多く、視点間の映像に強い相関はなかった
- 興味深いのは視覚的情報が出来事の分離に影響がほとんどないことだ

✧ 視覚的情報の変化に対して分離はロバストということが見られた
  - ➢ 視覚的情報従属仮説を否定する
✧ 出来事の分離のメカニズムは柔軟に視覚的情報を用いて行われる
✧ 役者の体が直視できなくても、出来事分離に影響はない

✧ 視点の視覚的情報の違い以外に、一人称視点と第三者視点はナラティブ理解力や記憶に影響する(Borghi, Glenberg, & Kaschak, 2004; Brunyé, Ditman, Mahoney, Augustyn, & Taylor, 2009; Libby & Eibach, 2011; Mcisaac &Eich, 2002; Nigro & Neisser, 1983; Storms, 1973; Taylor & Fiske, 1975).

Fig. 1. Frames from the twelve videos, which show six activities from both the first- and third-person perspectives.

Table 1

Mean, standard deviation (in parentheses), and correlations of visual attributes for each activity (N = 6) recorded from first- and third-person perspectives.

	Luminance	Clutter	Flow	VAI	Onset	Offset
First-person	151 (13)	.014 (.001)	.388 (.099)	.206 (.064)	.263 (.094)	.225 (.072)
Third-person	160 (6.9)	.018 (.002)	.026 (.013)	.006 (.004)	.244 (.120)	.231 (.092)
Correlation	.056 (.148)	.140 (.181)	.117 (.135)	.335 (.109)	.568 (.104)	.595 (.107)

Table 2

Means and standard deviations of the number of button presses per minute in each segmentation condition of Experiments 1 and 2.

	Fine	Neutral	Coarse
Experiment 1	10.70 (5.21)	6.01 (5.46)	2.09 (0.87)
Experiment 2	15.59 (6.85)	–	2.85 (1.73)

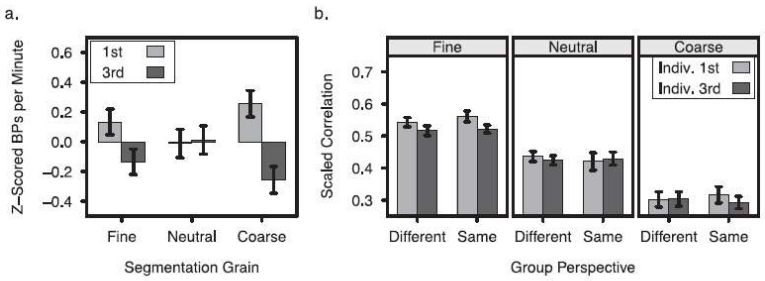

Fig. 2. Segmentation rate (a), indexed by z-scored button presses per minute, and individual-group agreement (b), indexed by the scaled individual-group correlation, in Experiment 1. Error bars indicate ± 1 standard error of the mean.

Table 3
The effects of visual features (VAI, touch onsets, and touch offsets) on the likelihood of a button press and their modulation by perspective and grain in Experiment 1 (N = 72) and Experiment 2 (N = 24).

Exp	Effect	Overall		Perspective			Grain			Persp. × Grain		
		Obs.	*95% CI*	*F*	*p*	η^2_p	*F*	*p*	η^2_p	*F*	*p*	η^2_p
Exp. 1												
	PLR	**18.554**	5.249–8.146	**6.575**	.016	.087	3.057	.057	.081	0.721	.471	.020
	VAI	**0.173**	−0.113 to 0.095	**7.564**	.010	.099	**9.792**	.000	.221	0.802	.432	.023
	Onset	**0.482**	−0.243 to 0.188	0.310	.575	.004	0.208	.817	.006	3.029	.054	.081
	Offset	**0.430**	−0.247 to 0.216	2.552	.103	.036	2.648	.071	.071	0.403	.674	.012
	VAI × Onset	−0.110	−0.183 to 0.236	**4.037**	.026	.055	0.247	.846	.007	0.675	.564	.019
	VAI × Offset	0.036	−0.246 to 0.327	1.898	.183	.027	0.118	.909	.003	0.142	.864	.004
	Onset × Offset	**−0.496**	−0.397 to 0.416	1.091	.280	.016	0.690	.530	.020	2.583	.078	.070
	3-Way	0.031	−0.648 to 0.397	2.529	.090	.035	0.251	.779	.007	1.509	.196	.042
Exp. 2												
	PLR	**16.152**	4.878–7.967	**4.540**	.041	.165	**34.409**	.000	.599	3.400	.075	.129
	VAI	**0.147**	−0.117 to 0.130	**5.337**	.035	.188	**4.806**	.019	.173	1.697	.203	.069
	Onset	**0.484**	−0.283 to 0.259	1.229	.241	.051	2.358	.133	.093	3.423	.073	.130
	Offset	**0.578**	−0.276 to 0.274	0.013	.907	.001	**6.853**	.013	.230	0.102	.752	.004
	VAI × Onset	0.040	−0.292 to 0.312	2.064	.162	.082	2.170	.139	.086	0.017	.886	.001
	VAI × Offset	0.382	−0.486 to 0.524	**8.466**	.014	.269	1.344	.276	.055	0.083	.786	.004
	Onset × Offset	**−0.625**	−0.415 to 0.487	0.034	.850	.001	**8.497**	.009	.270	0.040	.844	.002
	3-Way	−0.423	−0.819 to 0.650	0.899	.358	.038	0.206	.665	.009	0.606	.425	.026

Note: Obs.: Observed value. The observed value for the overall test are averaged across perspectives and grains. 95% CI: interval that captured 95% of the statistics in a simulation of expected values under the null hypothesis. PLR: Penalized Likelihood Ratio for the full model. For Experiment 1, perspective *F* degrees of freedom = 1, 69, grain and perspective × grain interaction *F* degrees of freedom = 2, 69. For Experiment 2, *F* degrees of freedom = 1, 23. Values in bold are unexpected under the null model with $p < .05$.

Table 4
Mean and standard deviation (in parentheses) of the logistic regression coefficients from model fits to individual segmentation data in Experiment 1.

Effect	Fine Grain		Neutral Grain		Coarse Grain	
	First	Third	First	Third	First	Third
VAI	0.140	0.041	0.247	0.150	0.242	0.216
	(0.150)	(0.141)	(0.198)	(0.108)	(0.217)	(0.124)
Onset	0.677	0.329	0.323	0.533	0.523	0.506
	(0.436)	(0.477)	(0.903)	(0.58)	(0.805)	(0.500)
Offset	0.199	0.296	0.567	0.631	0.327	0.562
	(0.699)	(0.399)	(0.712)	(0.445)	(0.819)	(0.632)
VAI × Onset	−0.171	−0.026	−0.261	−0.022	−0.108	−0.069
	(0.236)	(0.34)	(0.775)	(0.214)	(0.336)	(0.231)
VAI × Offset	0.009	0.098	−0.009	0.035	−0.020	0.106
	(0.428)	(0.297)	(0.379)	(0.304)	(0.570)	(0.375)
Onset × Offset	−0.487	−0.22	−0.396	−0.746	−0.411	−0.715
	(0.656)	(0.651)	(1.164)	(0.704)	(1.116)	(0.936)
3-Way	0.032	−0.074	−0.133	0.184	−0.107	0.282
	(0.613)	(0.589)	(1.037)	(0.641)	(0.669)	(0.864)

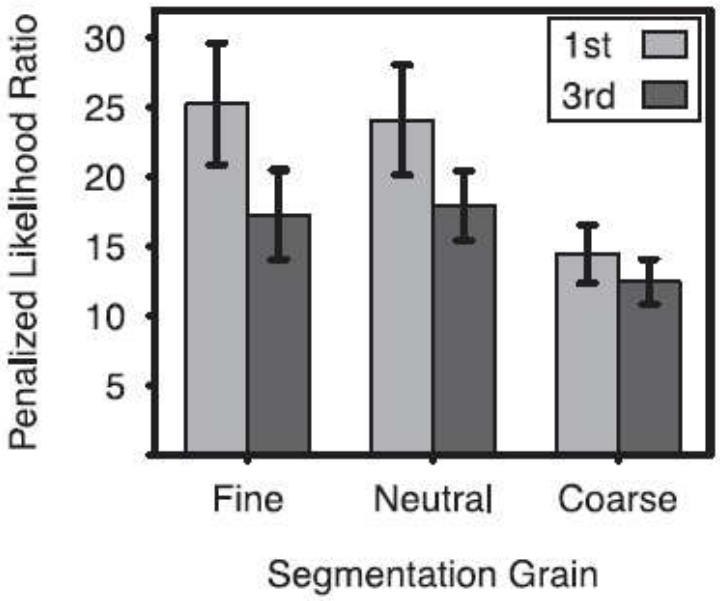

Fig. 3. Penalized likelihood ratios of logistic regression models fit to button presses across perspectives and grains in Experiment 1. Error bars indicate ± 1 standard error of the mean.