視聴者の注意指標に基づいた短縮映像の生成方法

○豊沢 聡、河合 隆史（早稲田大学 国際情報通信センター）

A method for video digest generation based on a viewer's attention measure

Satoshi Toyosawa, Takashi Kawai

(Global Information and Telecommunication Institute, Waseda University)

1. はじめに

増加の一途を辿る映像コンテンツを効率的に検索、視聴できるようにするには、映像の短縮化技術が欠かせない。短縮映像は、端的には元映像を代表する映像区間を選択することで生成されるが、これには、重み付け値やアノテーションなどの指標を各映像区間に割り当てる必要がある。選択指標には、テキスト、音声ピッチや音量、色構成、ボールや人物などのオブジェクトの動きといった映像そのものから得られる物理的な特徴量を用いるのが一般的であるが、視聴者の主観的な印象を重視し、視聴時の情動反応を用いる手法も近年注目を集めている。視聴時の反応は、主として視聴者の生体反応から得られる。例えば、皮膚電位反応から驚きを誘起した場面¹⁾や、顔の筋電から笑いを誘った場面²⁾を抽出する方法などが研究されている。

筆者らは、視聴者の興味関心を呼び起こしたり覚醒させた、すなわち注意資源の集中を要求した映像区間（以下、注意区間）を集めた短縮映像の生成方法を検討している。注意の指標として利用できる生体反応はいくつかあるが、映像視聴体験を損なわないという制約条件から、感圧センサを介して体動から間接的に測定できる、心拍動およびそこから得られる心拍変動を利用する。すでに報告したように、覚醒的な映像刺激呈示時には心拍動が低下するという知見に基づいた方法では、主観的な選択に対して平均 27% のヒット率が得られている³⁾。本稿ではこれに続き、注意指標である副交感神経系の活動を利用した短縮映像の生成方法について報告する。

2. 注意区間の選択方法

防衛反応の2段階モデル⁴⁾によれば、刺激が呈示された初期の注意的な段階では副交感神経の賦活と心拍動の低下が、次の行動的な段階では交感神経の賦活と心拍動の上昇が現れる。この生理的な反応は、初期段階については注意資源を感覚器官に集中させることで周囲の状況を把握する

ため、次段については逃走や接近などの身体活動をサポートするために発生するとされている。ただし、映像視聴のように実際の身体活動が必要でない場合は、恐怖症に関係するような特殊な映像を除き、次段へ移行することはないとされる。そこで、注意区間を検出するには、視聴時の副交感神経の活動に注目すればよい。副交感神経の活動は心拍変動の高周波成分に反映するとされるので、これを指標値として映像区間に割り当てればよい。

具体的には、まず映像時間と同期して取得した心拍動データをフーリエ変換することで、心拍変動の高周波成分（HF: High Frequency component）の帯域である0.15～0.40 Hzのパワーを得る。続いて、このパワーのカーブのピークと一致した映像区間を注意区間とし、そこに選択指標値としてピークと前の谷との差分値を割り当てる。ピークと一致しない映像区間には値0を割り当てる。映像はどのような基準で区間分割してもよいが、ここでは映像の最小単位であるショットを念頭に置いている。短縮映像は、必要な映像長になるまで指標値が大きいショット順に選択し、これらを元映像と同じ時系列順に連結することで生成する。

3. 評価実験

本稿の提案する心拍変動の高周波成分を用いた注意区間選択に基づく短縮映像生成方法を評価する目的で、被験者4名に3種類の刺激映像を呈示し、それぞれの被験者について、本方法が選択したショットと被験者が主観的に選択したショットとの一致率（ヒット率）を調査する評価実験を行った。

刺激映像には、快不快の度合いが中庸で覚醒度が低いと判断した野鳥のドキュメンタリー（以下、A）、快適さの度合いが高く覚醒的なゴールシーンを2つ含むサッカーゲーム（B）、不快さの度合いが高く覚醒的で残虐シーンを含む短編映像（C）の3種類を用いた。いずれも6分半程度のカラー映像で、平均60ショットで構成されている。映像には背景音楽、背景音などの音響は含まれる

図1 刺激映像別のヒット率

図2 被験者別のヒット率

が、音声や文字は含まれていない。

心拍動データは、被験者の胸部に取り付けた R-R 間隔を記録できるスポーツ心拍計 (Polar S810i) から取得した。刺激映像は2分の安静期間後に、被験者毎に順序を無作為化して呈示した。呈示には、被験者から約70 cm 離れた 19" LCD カラーディスプレイを用いた。また、音量は刺激映像間で同量で適切なものとなるよう実験者が調節した。そして、得られた心拍動データから前述の方法で10ショット (約1分半) を抽出した短縮映像 (以下、HF版) を被験者毎に作成した。

主観的な10ショットで構成される短縮映像は、心拍動取得実験後に被験者に再度刺激映像を視聴させることで得た。再視聴時には、被験者に印象深さ、重要さ、そして興奮の度合いの高かった10ショットをそれぞれ選ばせた上で、多くの評価を集めた10ショットを選択した (以下、主観版)。再視聴中は、被験者は一時停止をしてもよいとした。1回の再視聴で10ショットが得られなければ再度視聴させ、多い場合は後に削除させた。2度目以降の再視聴では早送りしてもよいとした。

ヒット率は、主観版中の10ショットに見出せる、HF版のショットの数から算出した。なお、全ショットからランダムに10ショットを選択した場合に主観版と一致するショットの数は1.7ショット (ヒット率17%) である。

4. 結果と考察

全被験者、全刺激映像の平均ヒット率は29%で、ランダムに主観版と一致する確率17%との間に有意差がみられた ($p < .05$)。刺激映像別と被験者別の平均ヒット率をそれぞれ図1と図2に示した。刺激映像間では有意差がみられた ($p < .05$) が、

被験者間では有意差はみられなかった。

ヒット率29%は、1つの生理指標に基づいた情動判定 (4~6種の情動から分類するので、ランダムヒット率は17~20%) のヒット率が8%~50%であるので、平均的なヒット率であるといえる。このヒット率はまた、前回報告の心拍動低下に基づいた方法の平均ヒット率と同程度である³⁾。刺激映像間に差異がない点も同じで、どちらの方法でも各種の映像に対応できるという示唆が得られた。しかし、心拍動低下方法ではヒット率が有意に他よりも低い被験者があり、ばらつきが大きかったのに対し、本方法では有意差はみられなかったところから、本方法の方が視聴者依存性が少ないといえる。一般に、生体反応を用いた情動判定は複数の生体指標を用いることでヒット率を向上させることが出来るとされるので、今後は、前回の心拍動低下方法と本方法の効果的な組み合わせ方法を検討していく予定である。

参考文献

- 1) J. Healey and R. W. Picard: StartleCam: a cybernetic wearable camera, Proc. 2nd Int. Symposium on Wearable Computing, 42-49, 1998.
- 2) Y. Shibata, et al.: Weight DS extension for affect-based content characterization, ISO/IEC JTC1/SC29/WG11, MPEG 88/M5481, 1999.
- 3) 豊沢 聡, 河合隆史: 「心拍動を利用した個人的な印象メモとしての短縮映像」, 人間工学会関東支部第37回大会, 89-90, 2007.
- 4) A. Ohman, et al.: Cognition and the autonomic nervous system, in "Handbook of Psychophysiology" (Cacioppo, et al. Eds.), Cambridge University Press, 2000.